

Problemes resolts de probabilitat i inferència aplicats a les ciències socials

Pablo Juan Verdoy
Modesto Joaquín Beltrán
María José Peris

Problemes resolts de probabilitat i inferència aplicats a les ciències socials

Pablo Juan Verdoy
Modesto Joaquín Beltrán
María José Peris



DEPARTAMENT DE MATEMÀTIQUES

■ Codis d'assignatura RA10, RL0906

UNIVERSITAT
JAUME I

Edita: Publicacions de la Universitat Jaume I. Servei de Comunicació i Publicacions
Campus del Riu Sec. Edifici Rectorat i Serveis Centrals. 12071 Castelló de la Plana
<http://www.tenda.uji.es> e-mail: publicacions@uji.es

Col·lecció Sapientia, 95
www.sapientia.uji.es
Primera edició, 2015

ISBN: 978-84-697-1551-2



Publicacions de la Universitat Jaume I és una editorial membre de l'UNE, cosa que en garanteix la difusió de les obres en els àmbits nacional i internacional. www.une.es



Aquest text està subjecte a una llicència Reconeixement-NoComercial-CompartirIgual de Creative Commons, que permet copiar, distribuir i comunicar públicament l'obra sempre que especifique l'autor i el nom de la publicació i sense objectius comercials, i també permet crear obres derivades, sempre que siguin distribuïdes amb aquesta mateixa llicència.
<http://creativecommons.org/licenses/by-nc-sa/2.5/es/deed.ca>

Índex

Introducció.	5
Unitat 1. Introducció a la probabilitat i a la probabilitat condicionada.	7
Introducció teòrica.	9
Objectius.	32
Enunciats.	33
Ajudes.	44
Solucions.	51
Unitat 2. Variables aleatòries discretes. Models de probabilitat discrets . . .	79
Introducció teòrica.	81
Objectius.	95
Enunciats.	96
Ajudes.	100
Solucions.	111
Unitat 3. Variables aleatòries contínues. Models de probabilitat continus.	
Teorema central del límit.	135
Introducció teòrica.	137
Objectius.	162
Enunciats.	163
Ajudes.	168
Solucions.	184
Unitat 4. Introducció a la inferència 1: intervals de confiança.	201
Introducció teòrica.	203
Objectius.	232
Enunciats.	233
Ajudes.	239
Solucions.	249
Unitat 5. Introducció a la inferència 2: contrast d'hipòtesis.	267
Introducció teòrica.	269
Objectius.	284
Enunciats.	285
Ajudes.	289
Solucions.	295
Bibliografia.	307

Introducció

El present llibre de problemes es pot considerar com el segon dels dos complements del manual *Introducció a l'estadística aplicada a les ciències socials* de la col·lecció «Sapientia» del servei de publicacions de l'UJI, el qual consta fonamentalment de continguts teòrics i, per tant, amb l'apartat de problemes en un segon pla. Amb aquest nou text, basat quasi exclusivament en problemes resolts, es completa part del manual teòric i es facilita a l'estudiant una ferramenta excel·lent per a consolidar l'aprenentatge dels seus continguts.

Els problemes compten amb ajudes, l'última de les quals és la seua resolució completa. És a dir, cadascun dels problemes té dos tipus d'ajudes, que no són més que una breu informació que pot facilitar a l'estudiant l'ardu treball de resoldre el problema. Les ajudes de tipus 1 són una mera orientació que tenen com a objecte manifestar els continguts que s'han de consultar per a poder resoldre el problema. Les ajudes de tipus 2 donen prou més informació que les primeres. Així, en moltes ajudes d'aquest tipus es mostra part de la resolució de l'exercici. Finalment, en la resolució del problema es mostren fil per randa els continguts estadístics que s'utilitzen i nombrosos comentaris que permeten intuir la resolució de problemes similars.

A més a més, els problemes estan classificats per objectius, ja que d'aquesta manera l'estudiant sap en cada moment quins continguts s'estan treballant i, per tant, pot consultar el manual teòric per tal de revisar aquelles qüestions en les que trobe dificultats.

Per una altra part, aquest manual està dividit en unitats que fan referència des de la probabilitat fins a la inferència. Cada unitat està dividida en quatre blocs: En el primer es proposen els enunciats dels problemes classificats per objectius; la segona part proporciona únicament les ajudes de tipus 1; en el tercer bloc les ajudes de tipus 2 (el fet que per a un mateix problema no es troben els dos tipus d'ajudes conjuntament té la pretensió que l'estudiant realitze la consulta detallada de les ajudes, reforçant la idea de pensar abans de consultar); en la darrera part es mostren les resolucions completes dels problemes, les quals estan repletes de comentaris, gràfics i diagrames que en faciliten la comprensió.

UNITAT 1

Introducció a la probabilitat i a la probabilitat condicionada

Introducció teòrica

La teoria de la probabilitat té el seu origen en els jocs d'atzar, que es van convertir en una activitat molt comuna en la França del segle XVII. En aquest tipus de jocs (cartes, daus...) s'apostaven grans quantitats de diners, d'ací l'interès per predir-ne els resultats.

Un ric jugador professional parisenc de l'època, Antoine Gombaud, el Chevalier de Méré, va plantejar un problema referent al joc d'atzar anomenat *points* a un dels matemàtics coetanis d'ell, Blais Pascal. En aquest joc es compten els punts guanyadors en una tirada de daus i el jugador que siga el primer a guanyar un cert nombre de punts és el vencedor i s'endú els diners de l'aposta.

Segons compten, Gombaud s'havia trobat jugant a *points* amb un jugador més experimentat quan, a causa d'un compromís, es veren obligats a deixar el joc a mig fer. El problema que es plantejà aleshores fou què es podia fer amb els diners que havien apostat. La solució simple hauria estat donar tots els diners al competidor amb més punts, però Gombaud demanà a Pascal si hi havia una manera més justa de dividir els diners. A Pascal li demanaren, doncs, que calculara la probabilitat que li quedava a cada jugador de guanyar si hagués continuat el joc i partint de la base que tots dos jugadors tenien les mateixes probabilitats de guanyar els punts que quedaven en joc. Els diners de l'aposta serien dividits segons aquest càlcul de probabilitats.



Així, els primers estudis de probabilitat en aquella època se centraren en qüestions com l'anterior o de l'estil: per què en el joc consistent a llançar un dau dues vegades i sumar les dues puntuacions és més senzill obtenir un vuit que un quatre?, o per què és més senzill obtenir una cara en llançar dues vegades una moneda que obtenir-ne dues en llançar la moneda quatre vegades, si pareix clar que ha de resultar igual de senzill si raonem en termes de proporcionalitat?

Abans del segle XVII, les lleis de probabilitat eren definides per la intuïció i per l'experiència dels jugadors, però Pascal inicià un intercanvi epistolar amb Fermat a fi d'esbrinar les regles matemàtiques que descrigueren la probabilitat. Tres segles més tard, Bertrand Russell comentaria aquest aparent oxímoron: com es pot parlar de les lleis de la probabilitat? No és la probabilitat l'antítesi de la llei?

Els problemes de probabilitat de vegades són controvertits, perquè la resposta matemàtica, la verdadera resposta, sol ser contrària al que la intuïció sol suggerir. Per exemple, un dels problemes més contraintuïtius que hi ha es refereix a la probabilitat de celebrar l'aniversari el mateix dia que alguna altra persona. Si imaginem una festa amb 23 persones, no sembla molt probable que dues persones qualsevol complisquen anys el mateix dia. Amb 23 persones i 365 dies per triar, la intuïció s'apropa a la idea que ningú comparteix la seua data d'aniversari. Si se'ns demanara posar una xifra a aquesta probabilitat, molts conjecturàriem una probabilitat de potser un 0,1. Doncs bé, la resposta real és més del 0,5. Això vol dir que, posades en la balança de les probabilitats, és més probable trobar dues persones en la festa que compartisquen data d'aniversari que no trobar-ne cap.

No és d'estranyar, doncs, que els matemàtics de l'època es deixaren seduir per aquest tipus de problemes que en moltes ocasions semblen anar en contra de la intuïció. De fet, Fermat i Pascal descobriren les regles essencials que governen tots els jocs de probabilitat i que poden ser utilitzades pels jugadors per a definir el joc perfecte i les estratègies de les seues apostes. Més encara, aquestes lleis de probabilitat han trobat aplicacions en un gran nombre de situacions, que van des de les especulacions en el mercat de valors fins a les estimacions de la probabilitat d'un accident nuclear. Pascal¹ estava convençut fins i tot que podia utilitzar les seues teories per a justificar la creença en Déu. Va afirmar que «l'excitació que un jugador sent quan fa una aposta és igual a la quantitat que pot guanyar multiplicada per la probabilitat de guanyar-la». Sostenia que el possible premi de la felicitat eterna té un valor infinit i que la probabilitat d'arribar al cel amb un vida virtuosa, per més petita que aquesta siga, és certament finita. Per tant, segons la definició de Pascal, la religió era un joc d'excitació infinita que valia la pena jugar, perquè multiplicar un premi infinit per una probabilitat finita dóna un resultat infinit.

La correspondència epistolar que Pascal va mantindre amb Pierre de Fermat i altres grans matemàtics de l'època originà la teoria de la probabilitat i féu que aquesta passara de ser una mera col·lecció de problemes aïllats sobre jocs a constituir, amb el temps, una part molt important de les matemàtiques.

Atzar i probabilitat

En essència, el concepte inherent als jocs tractats en la introducció és bàsicament un: l'atzar.

Què s'entén per atzar i per experiment aleatori?

Per a entendre completament aquest concepte és convenient comparar-lo amb la seua idea oposada: la *necessitat*. Així, els fenòmens governats per la necessitat són aquells en què una mateixa causa determina, inevitablement, un efecte. Per exemple:

1. Simon Singh (1997): *Fermat's Enigma*.

- Una pedra cau si res no la suporta.
- Si un líquid es calfa prou, s'evapora.
- Si un nombre enter és divisible per 4, també ho és per 2.
- Si se sumen els angles d'un triangle, el resultat és 180.

Aquestes afirmacions poden considerar-se casos de necessitat, ja que mai s'ha observat que una pedra sense suport no caiga o és sabut que si un nombre és múltiple de 4 també ho és de 2. D'altra banda, els exemples anteriors mostren que la necessitat pot ser empírica (observacions) o lògica (basada en axiomes).

Com a conclusió, es pot asseverar que en els fenòmens dominats per la necessitat, quan ocorren certes causes, l'efecte està determinat, per això a aquests fenòmens se'ls denomina deterministes.

Enfront d'aquests fenòmens se'n perceben d'altres que atribuïm a l'atzar:

- El resultat obtingut en llançar una moneda sobre una taula.
- El resultat de llançar un dau.
- El nombre de telefonades que rep una centraleta.
- El resultat obtingut en un sorteig d'una cistella de Nadal.

En aquest tipus d'experiments, si s'anomenen esdeveniments als possibles resultats del fenomen (per exemple, cara i creu són els esdeveniments de l'experiment llançar una moneda), un esdeveniment A_1 pot ocórrer quan es donen una sèrie de condicions, però sota aquestes mateixes condicions pot ocórrer qualsevol altre esdeveniment: A_2, A_3, \dots, A_n , dels possibles. És a dir, no pot preveure's quin d'ells succeirà. En aquest cas es diu que l'ocurrència A_1 ha estat causada per l'atzar i que el fenomen és aleatori.

Per exemple, en l'experiment llançar un dau els possibles esdeveniments són: eixir un 1, eixir un 2, eixir un 3, eixir un 4, eixir un 5 i eixir un 6. Si el resultat del llançament ha sigut 6, és evident que si es realitza l'experiment novament el resultat podria ser 6 o qualsevol altre resultat dels possibles: 1, 2, 3, 4 o 5. És per això que es diu que el resultat ha sigut conseqüència de l'atzar i que l'experiment es denomina aleatori.

Què s'entén per probabilitat?

La necessitat està lligada a la certesa o seguretat; de fet, s'afirma que:

- És segur que una pedra sense suport caiga.
- És cert que la terra es mou en l'espai.

De manera semblant, l'atzar s'associa amb la noció de probabilitat; d'ací que es diga:

- És probable que pugue la borsa.
- És probable que un cert equip de futbol guanye la lliga.

En aquest sentit, la probabilitat expressa intuïtivament bé una apreciació de la facilitat que s'atribueix que ocorregui un esdeveniment o bé una mesura del grau de creença o versemblança en l'aparició d'un esdeveniment.

A més, hi ha una tendència en l'ésser humà, més o menys conscient, a pensar que un esdeveniment és més versemblant que un altre, i per tant un desig de mesurar eixa creença:

- Si s'extrau una bola d'una urna en què hi ha 3 boles blanques i 1 negra, és més probable que s'obtinga una bola blanca.
- És més probable que es consumisca més energia un divendres que un dissabte.

Com es veu, és relativament senzill comparar la versemblança de dos esdeveniments si es coneix la naturalesa de l'experiment aleatori.

D'altra banda, cal tenir present que *l'objectiu fonamental* és comparar tots els possibles esdeveniments o resultats d'un experiment segons la seua versemblança (certesa de que ocorreguen). El mètode més senzill consisteix a assignar un nombre a cada un d'aquests possibles resultats, els quals han de permetre realitzar l'esmentada comparació. Es cerca doncs, una llei a la qual se sotmeta l'atzar però que respecte el seu caràcter imprevisible.

Evidentment, d'ara endavant es considerarà que els experiments són aleatoris, ja que si foren deterministes no tindria sentit parlar de probabilitat.

Concepte de probabilitat basat en la freqüència

El mode més intuïtiu d'assignar un nombre a cada esdeveniment consisteix a realitzar l'experiment successives vegades i estudiar els resultats de les observacions. Així, per exemple, en llançar una moneda 100 vegades s'han obtingut els resultats següents:

C, C, +, +, +, C, C, C, +, C, C, C, C, +, +, C, C, +, +, +, +, C, +, C, +, C, C, +,
 C, +, C, +, +, C, C, C, C
 +, +, +, +, +, C, C, +, +, C, +, +, C, +, C, +, +, +, C, C, C, C, C, +, C, C, C, +,
 +, +, C, C, C, +, C, C, C
 C, C, C, +, C, C, +, C, +, +, +, +, C, C, +, C, +, +, C, +, +, C, C, C, +, +.

Els dos gràfics següents (fig. 1) reflecteixen la diferència existent entre el nombre de cares i el de creus en cada llançament. En ells s'observa l'aparició de ratxes de cares i creus, les quals provoquen que aquesta diferència no s'estabilitze entorn del 0, com pareixeria indicar la intuïció.

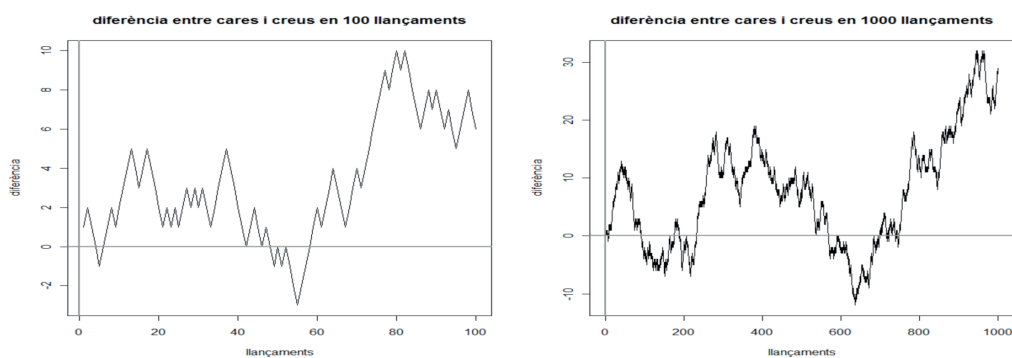


Fig. 1

Vol dir açò que la nostra intuïció és errònia? És més probable que resulten més cares que creus en l'experiment? Si se centra l'atenció en el gràfic (fig. 2), reflex de la freqüència relativa del nombre de cares en cada llançament, s'observa que a partir del llançament número 200 aproximadament (en el gràfic dels 1000 llançaments) la freqüència relativa del resultat «eixer cara» s'estabilitza entorn del 0,5. És més, pareix clar que, a mesura que es va reiterant l'experiment amb nous llançaments, la freqüència relativa s'aproxima ràpidament a un valor ideal de 0,5.

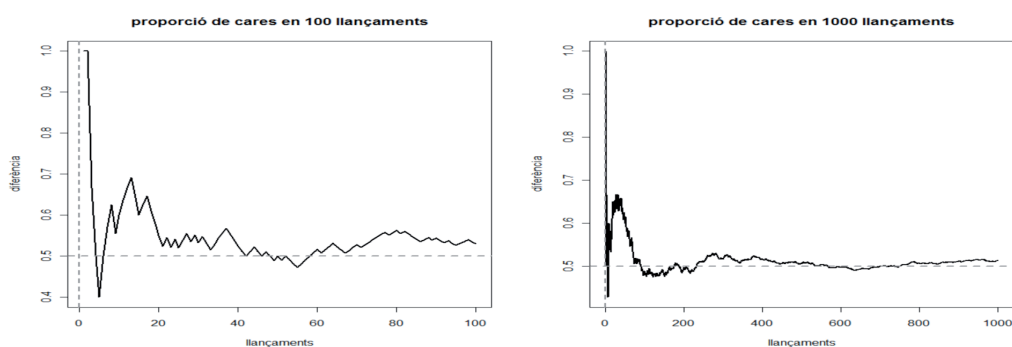


Fig. 2

Cap a quin valor ideal s'aproximarà la freqüència relativa de l'esdeveniment «eixer creu»? Òbviament, a 0,5. D'altra banda, l'aparició de ratxes demostra que els resultats de l'experiment són totalment imprevisibles, encara que les freqüències relatives s'acosten a un valor ideal de 0,5.

Tenint en compte les distintes consideracions anteriors, s'obté una primera noció de probabilitat com a mesura de versemblança d'un esdeveniment. Així, es pot definir la probabilitat d'un esdeveniment A1 (en avant $P(A1)$) com el valor ideal a què s'acosta la freqüència relativa de l'esdeveniment A1 quan es realitza l'experiment un gran nombre de vegades.

Per ser la probabilitat de A1 una freqüència relativa, es té $0 \leq P(A1) \leq 1$

Aquest fet queda ratificat per la llei dels grans nombres demostrada pel cèlebre matemàtic Jacob Bernoulli i que diu així: «La freqüència relativa d'un succés

tendeix a estabilitzar-se entorn d'un nombre a mesura que el nombre de proves de l'experiment creix indefinidament».

A l'exemple $P(\text{eixir cara}) = \text{freqüència relativa}(\text{eixir cara})$, en moltes repeticions de l'experiment i com s'ha vist al gràfic, la freqüència és 0,5.

Cal notar que, segons el tractament que se li ha donat al concepte, la probabilitat d'eixir cara o eixir creu és una conseqüència lògica d'una llei empírica que podria enunciar-se de la manera següent: «Si es llança una moneda nombroses vegades, la freqüència relativa de l'esdeveniment “eixir cara” s'aproxima a un valor denominat $P(\text{eixir cara}) = 0,5$ ».

Per tant, $P(\text{eixir cara})$ i $P(\text{eixir creu})$ passen a ser dos qualitats objectives més de la moneda, com la massa, el color, etc., ja que, de la mateixa manera que no es pot afirmar que la moneda no és sotmesa a l'acció de la gravetat en algun moment, tampoc es pot afirmar que si es repeteix l'experiment de llançar la moneda infinites vegades la freqüència d'eixir cara no s'acoste a 0,5.

Concepte subjectivista de probabilitat

En la interpretació de probabilitat com una freqüència és necessari repetir l'experiment un considerable nombre de vegades per a saber a quin número s'acosta la freqüència relativa de l'esdeveniment que s'està estudiant. No obstant, en la majoria de les ocasions és convenient conèixer-ne la probabilitat abans que succeeixca per no poder-se realitzar l'experiment més d'una vegada (per exemple, calcular al probabilitat que el nostre bitllet de loteria siga el premiat).

És per això que sorgeix la idea d'assignar un nombre a la probabilitat d'un esdeveniment segons la creença que es tinga de que ocorrega. Cal notar que el nombre assignat ha d'estar comprés entre 0 i 1 per a que siga compatible amb la definició basada en la freqüència de probabilitat comentada anteriorment.

Per exemple, davant de l'experiment llançar una moneda, es pot assignar $P(\text{Eixir cara}) = 0,5$ i $P(\text{Eixir creu}) = 0,5$, ja que es té la creença que els dos esdeveniments poden donar-se amb la mateixa facilitat.

És clar que les assignacions de probabilitats han de realitzar-se de manera que les dues interpretacions de la probabilitat coincidisquen; és a dir, l'assignació realitzada a priori de la probabilitat d'un esdeveniment ha de coincidir amb la probabilitat a posteriori. Una vegada assignades les probabilitats als esdeveniments (tant si és pel mètode empíric com si és pel subjectiu) les diferents deduccions que es realitzen segueixen un procediment lògic.

Assignació de probabilitats

Quan s'han d'assignar probabilitats als esdeveniments lligats a un experiment aleatori, és crucial entendre que el grau de versemblança depèn de l'evidència de la qual es disposa sobre el fenomen.

Espai mostral i successos

Als epígrafs anteriors, davant d'un experiment s'han observat intuïtivament esdeveniments o successos simples i compostos. A continuació es formalitzaran aquests conceptes, així com les relacions que poden establir-se entre ells.

Definicions

Experiments aleatoris. Es diu que un experiment és aleatori, estocàstic o estadístic quan, podent-se repetir indefinidament en anàlogues condicions, és impossible predir-ne el resultat, encara que se'n coneixen les condicions inicials. En un experiment aleatori no es coneix el resultat fins que s'ha realitzat la prova. S'anomena *prova* a cada realització d'un experiment.

Espai mostral. El conjunt de tots els resultats possibles a què pot donar lloc un experiment aleatori s'anomena espai mostral. Sol representar-se per E o Ω i es diu que és *finit* si el nombre de resultats possibles és finit.

Succés. Donat un experiment aleatori, l'espai mostral del qual és E , s'anomena succés a cada un dels subconjunts de E .

Es distingeixen els següents tipus de successos:

- Succés simple o elemental: només consta d'un element.
- Succés compost: consta de dos o més elements.
- Succés impossible: és el que mai pot realitzar-se (ve determinat pel conjunt buit, \emptyset).
- Succés segur: és el que sempre es compleix (ve determinat pel conjunt total, E).
- Successos disjunts o mútuament excloents: aquells successos A i B que no poden realitzar-se conjuntament, $A \cap B = \emptyset$.

NOTA

Tenint en compte que els successos són subconjunts de E , es pot aplicar la teoria general de conjunts. Concretament les unions, interseccions, diferències i complementarietats entre conjunts.

Operacions amb successos

En tot moment es considerarà $E =$ espai mostral.

No obstant, abans de començar amb la definició de les diferents operacions, és necessari aclarir els conceptes d'igualtat i d'inclusió.

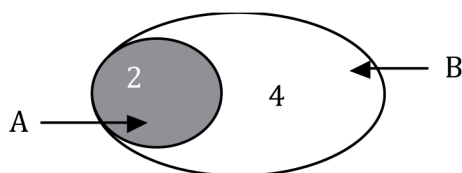
Igualtat

Es diu que els successos A i B són iguals ($A = B$) quan l'ocurrència de A implica l'ocurrència de B i viceversa. És a dir, els dos successos estan formats pels mateixos successos elementals.

Inclusió

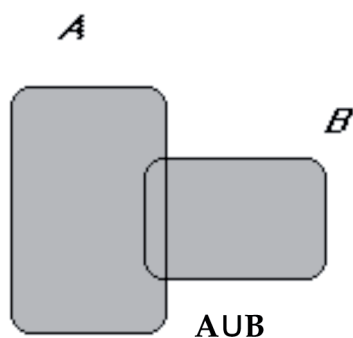
Es diu que el succés A està inclòs en el succés B ($A \subseteq B$) quan l'ocurrència de A implica l'ocurrència de B . És a dir, els successos elementals d' A estan tots també en B .

Gràficament:



Unió de successos (\cup)

Donats dos successos qualsevol A i B , el succés $A \cup B$ (A unió B) és un succés que es verifica quan es verifica A , quan es verifica B , o bé quan es verifiquen els dos al mateix temps.



Propietats de la unió

Siguen A, B i C tres successos qualsevol de l'experiment, llavors es compleix:

$$A \cup (B \cup C) = (A \cup B) \cup C$$

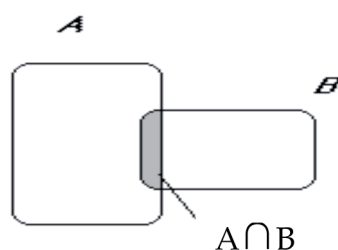
$$A \cup B = B \cup A$$

$$A \cup E = E$$

$$A \cup \emptyset = A$$

Intersecció de successos (\cap)

Donats dos successos qualsevol, A i B, el succés $A \cap B$ (A intersecció B) és un succés que es verifica quan ocorre el succés A i el succés B simultàniament.



Propietats de la intersecció

Siguen A, B i C tres successos qualsevol de l'experiment, llavors es compleix:

$$A \cap (B \cap C) = (A \cap B) \cap C$$

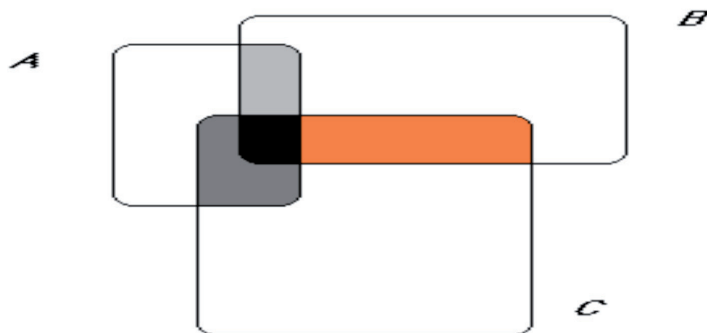
$$A \cap B = B \cap A$$

$$A \cap E = A$$

$$A \cap \emptyset = \emptyset$$

Propietats distributives de la unió y de la intersecció

Si es consideren tres successos qualsevol, A, B i C, llavors es compleix:

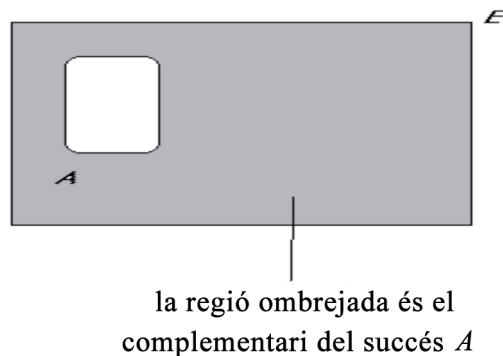


- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Succés contrari o complementari

Donat un succés A , es denomina succés contrari o succés complementari de A , escrit \bar{A} , al succés que només ocorre quan no succeix A .

La regió amb ombra és el succés complementari d' A .



Propietats del complementari

- $\bar{\emptyset} = E$ y $\bar{E} = \emptyset$
- $A \cup \bar{A} = E$ y $A \cap \bar{A} = \emptyset$
- $\overline{\bar{A}} = A$
- $\overline{A \cup B} = \bar{A} \cap \bar{B}$ y $\overline{A \cap B} = \bar{A} \cup \bar{B}$ (Lleis de Morgan)

Successos independents

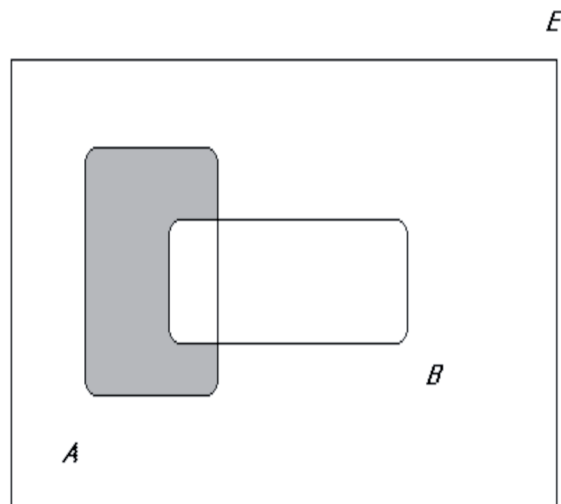
Dos successos són independents quan l'ocurrència d'un d'ells *no* influeix en l'ocurrència de l'altre.

NOTA

Dos successos *disjunts* (aquells que no tenen cap element en comú) no són independents, ja que si un ocorreix, *és segur* que l'altre no ho farà. Per tant, l'ocurrència d'un influeix en l'ocurrència de l'altre.

Diferència de dos successos

Donats els successos A i B , es denomina succés A menys B i s'escriu $A - B$ al succés en què es verifica A però no B , és a dir, que el succés $A - B$ està format pels successos elementals que estan en A i no estan en B . És evident que $A - B = \bar{B}$.



Concepte de probabilitat. Definició i propietats

Quan s'ha introduït el concepte de probabilitat en els primers punts del tema, s'han distingit dues interpretacions de la mateixa: la basada en la freqüència (en la qual s'ha definit la probabilitat d'un succés com el valor al qual s'acosta la freqüència relativa d'aquest esdeveniment en realitzar l'experiment un nombre molt gran de vegades) i la subjectiva (en què la probabilitat d'un succés s'assigna segons la informació que es coneix de l'experiment).

És evident que els conceptes de probabilitat més formal o clàssic i el basat en la freqüència han de coincidir, perquè la probabilitat d'un succés ha de ser un únic valor.

És el moment ara de formalitzar el concepte de probabilitat. Per fer-ho, es definiran els axiomes que determinen el concepte clàssic de probabilitat, basant-se en tres propietats fonamentals que compleix la probabilitat definida per les freqüències. Així, es considera E l'espai mostral d'un experiment aleatori i A i B dos successos qualsevol de l'experiment.

Cal recordar que:

$P(A)$ = freqüència relativa de A =

$$fr(A) = \frac{\text{n}^\circ \text{ de ocurrencies de } A}{\text{n}^\circ \text{ de vegades que s'ha realitzat l'experiment}} = \frac{n_A}{n}$$

quan n tendeix a l'infinit.

Es denota per:

n = nombre de vegades que s'ha realitzat l'experiment
 n_A = nombre de vegades que ha ocorregut A
 n_B = nombre de vegades que ha ocorregut B

Amb aquesta notació, passem a comentar les tres propietats:

Propietat 1

Per a qualsevol succés A , es té $0 \leq fr(A) \leq 1$

És evident, doncs, que $0 \leq n_A$ i $n > 0$. Per això $0 \leq \frac{n_A}{n}$. Per un altre costat, és evident que $n_A \leq n$ i conseqüentment, $\frac{n_A}{n} \leq 1$.

Unint les dues desigualtats, s'obté $0 \leq \frac{n_A}{n} \leq 1$, que demostra la propietat.

Propietat 2

$$fr(E) = 1$$

Es evident, ja que E és el succés segur, per ser E l'espai mostral.

Propietat 3

Si A y B son dos successos disjunts, $A \cap B = \emptyset$, aleshores es compleix:

$$fr(A \cup B) = fr(A) + fr(B)$$

$$fr(A \cup B) =$$

$$\frac{\text{nombre de vegades que ocorre } A \text{ o } B}{\text{número de vegades que es realitza l'experiment}} =$$

$$\frac{\text{nombre de vegades de } A + \text{nombre de vegades de } B - \text{nombre de vegades dels dos}}{\text{nombre de vegades que es realitza l'experiment}} =$$

$$\frac{n_A + n_B - 0}{n} = \frac{n_A}{n} + \frac{n_B}{n} = fr(A) + fr(B)$$

Basant-se en aquestes propietats, es pot definir ja el concepte més formal i clàssic de la probabilitat.

Definició clàssica o axiomàtica de probabilitat

Donat E l'espai mostral d'un experiment aleatori, una probabilitat en E és qualsevol funció que assigne a cada succés A de l'experiment un nombre $P(A)$, de manera que es complisquen els axiomes següents:

Axioma 1: la probabilitat del succés segur val 1; $P(E) = 1$.

Axioma 2: la probabilitat de qualsevol altre succés A és no negativa:
 $P(A) \geq 0$.

Axioma 3: la probabilitat de la unió de dos successos mútuament excloents, A i B , és la suma de les seues probabilitats.

És a dir, si $A \cap B = \emptyset$, llavors $P(A \cup B) = P(A) + P(B)$

Aquest darrer axioma pot generalitzar-se del mode següent:

La probabilitat de la unió d'un conjunt de successos mútuament excloents és igual a la suma de les seues probabilitats.

$$P(\cup A_i) = \sum P(A_i) = P(A_1) + P(A_2) + \dots + \dots$$

Propietats de la probabilitat

D'aquests axiomes es poden deduir una sèrie de propietats:

Propietat 1

Si A_1, A_2, \dots, A_n són successos disjunts dos a dos amb $n > 2$

(o siga, $A_i \cap A_j = \emptyset$ amb $i \neq j$), llavors:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

A més a més, si $B = A_1 \cup A_2 \cup \dots \cup A_n \rightarrow P(B) = P(A_1) + P(A_2) + \dots + P(A_n)$

És immediata per l'axioma 3, ja que el nombre de successos que s'han pres és n (un nombre finit) i, com la propietat es compleix per a dos successos i per a una quantitat infinita numerable, es compleix també per a una quantitat finita.

Propietat 2

$P(\bar{A}) = 1 - P(A)$, sent A un succés qualsevol.

Nota: \bar{A} és el complementari de A

$$A \cup \bar{A} = E \longrightarrow P(A \cup \bar{A}) = P(E) = 1$$

$$\text{I com } A \cap \bar{A} = \emptyset \longrightarrow \text{axioma 3} \longrightarrow P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

$$\text{D'ambdós conseqüències, } P(A) + P(\bar{A}) = 1 \longrightarrow P(\bar{A}) = 1 - P(A)$$

Propietat 3

$$P(\emptyset) = 0$$

$$\emptyset = \bar{E} \longrightarrow P(\emptyset) = P(\bar{E})$$

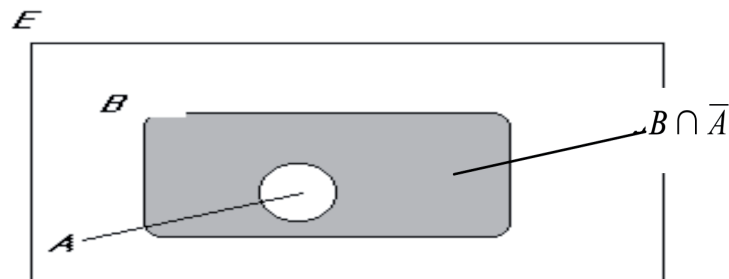
$$\text{Per la propietat 2, } P(\bar{E}) = 1 - P(E) = 1 - 1 = 0$$

$$\text{Per tant, } P(\emptyset) = 0$$

Propietat 4

Donats dos successos A i B en què $A \subseteq B \longrightarrow P(A) \leq P(B)$.

Observant el gràfic es dedueix $B = A \cup (B \cap \bar{A})$



A més, $A \cap (B \cap \bar{A}) = \emptyset \longrightarrow A$ i $(B \cap \bar{A})$ són disjunts

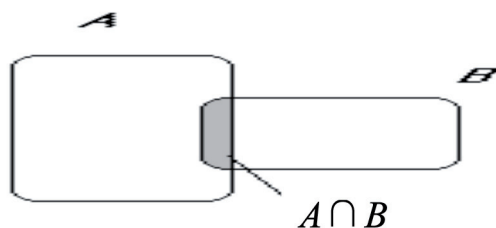
Per tant, per l'axioma 3: $P(B) = P(A) + P(B \cap \bar{A})$

Com també, per l'axioma 1, $P(B \cap \bar{A}) \geq 0 \longrightarrow P(B) \geq P(A)$

Propietat 5

Per a tot A i B que siguin de E , $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Intuïtivament:



En fer $A \cap B$, es pren dos vegades $A \cap B$, després per a calcular el que es vol s'ha de restar una vegada $A \cap B$.

Amb més formalitat, tenint present que:

$$A \cup B = A \cup (B \cap \bar{A}) \text{ i que } B = (A \cap B) \cup (B \cap \bar{A}),$$

$$\text{s'obté d'una banda } P(A \cup B) = P(A \cup (B \cap \bar{A})) = P(A) + P(B \cap \bar{A})$$

$$\text{i d'una altra banda } P(B) = P((A \cap B) \cup (B \cap \bar{A})) = P(A \cap B) + P(B \cap \bar{A}).$$

$$\text{Restant ambdues expressions, s'obté } P(A \cup B) - P(B) = P(A) - P(A \cap B)$$

Corol·lari

Donats dos successos A i B en què $B \subseteq A \rightarrow P(A - B) = P(A) - P(B)$,

tenint en compte que $P(A - B) = P(A \cap \bar{B})$ i la propietat 5, s'obté que:

$$P(A \cap \bar{B}) = P(A) + P(\bar{B}) - P(A \cup \bar{B})$$

i com que $B \subseteq A \rightarrow \bar{A} \subseteq \bar{B}$, llavors l'espai mostral E compleix:

$$E = A \cup \bar{A} \subseteq A \cup \bar{B} \subseteq E \rightarrow A \cup \bar{B} = E \rightarrow P(A \cup \bar{B}) = 1.$$

$$P(A \cap \bar{B}) = P(A) + P(\bar{B}) - P(A \cup \bar{B}) = P(A) + (1 - P(B)) - 1,$$

per tant es té que $P(A - B) = P(A) - P(B)$.

Espais mostrals finits. Regla de Laplace

S'anomenen *espais mostrals finits* els espais mostrals que provenen d'experiments per als quals només hi ha un nombre finit de resultats possibles:

$$\text{així, } E = \{w_1, w_2, \dots, w_n\}$$

En un experiment aleatori amb un espai mostral finit, una distribució de probabilitat s'especifica assignant una probabilitat p_i a cada resultat w_i que pertany a E : $p_i = P(\{w_i\})$. Ha de complir-se:

$$\begin{array}{l} a) p_i \geq 0 \\ b) P(E) = 1 \longrightarrow \sum_{i=1}^n p_i = p_1 + \dots + p_n = 1, \end{array}$$

ja que els esdeveniments $\{w_1, w_2, \dots, w_n\}$ són elementals i, per tant, disjunts.

En aquestes condicions, si $A = \{w_{i1}, w_{i2}, \dots, w_{ir}\}$, es té $P(A) = p_{i1} + \dots + p_{ir}$.

S'anomenen *espais mostrals simples* als espais mostrals finits en què tots els resultats són equiprobables (tenen la mateixa probabilitat). Si $E = \{w_1, w_2, \dots, w_n\}$,

$$\text{llavors } P(\{w_i\}) = \frac{1}{n}, \quad i = 1, \dots, n$$

En aquests espais mostrals simples, donat un succés $A = \{w_1, w_2, \dots, w_k\}$ amb $k < n$, es té:

$$P(A) = P(w_1 \cup w_2 \cup \dots \cup w_k) = P(w_1) + P(w_2) + \dots + P(w_k) =$$

$$\frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n} = k \cdot \frac{1}{n} = \frac{\text{n}^\circ \text{ successos elementals de } A}{\text{n}^\circ \text{ successos elementals de } E}$$

Aquest fet justifica la fórmula de Laplace:

Donat un experiment aleatori on tots els resultats (esdeveniments simples) tenen la mateixa probabilitat de produir-se, si A és un succés possible de l'experiment, es compleix:

$$P(A) = \frac{\text{casos favorables a } A}{\text{casos possibles}}$$

On els *casos possibles* són tots els resultats possibles de l'experiment (el nombre total d'esdeveniments simples) i els *casos favorables a A* és el nombre de resultats que compleixen A (nombre d'esdeveniments simples que compleixen A).

Teorema de la probabilitat total. Teorema de Bayes

Probabilitat condicionada

En una carrera de cavalls en què en participen 15, un jugador ha apostat al cavall número 2. El jugador rep dues hores després de l'aposta la informació que el número del cavall guanyador és parell. Per una altra part, un segon jugador rep la mateixa informació abans d'apostar i decideix fer-ho també al cavall número 2.

Si tots dos han decidit el número del cavall aleatòriament, quin dels dos jugadors té una major probabilitat de guanyar? És evident que la informació de la qual gaudeix el segon jugador l'afavoreix a l'hora de prendre la decisió.

1r jugador

Segons diu l'enunciat, *a priori* tots els cavalls tenen la mateixa probabilitat de guanyar i, en conseqüència, es pot aplicar la regla de Laplace a l'hora de calcular la probabilitat d'encert del cavall guanyador, així:

- Casos possibles: 15 (ja que són 15 cavalls els possibles guanyadors)
- Casos favorables: 1 $\rightarrow P(\text{guanyar jug. 1}) = P(\text{guanyar el cavall 2}) = \frac{1}{15}$

2n jugador

El segon jugador gaudeix d'una informació privilegiada abans d'apostar. Sap que el cavall guanyador duu un nombre parell. Aplicant la regla de Laplace:

- Casos possibles: 7 (ja que són 7 cavalls els possibles guanyadors, perquè per la informació rebuda els possibles guanyadors són els cavalls 2, 4, 6, 8, 10, 12, 14)
- Casos favorables: 1 $\rightarrow P(\text{guanyar jug. 2}) = P(\text{guanyar el cavall 2}) = \frac{1}{7}$

Així doncs, el jugador 2 té, abans de començar la carrera, més probabilitat d'encertar el cavall guanyador que el jugador 1.

Aquest exemple posa de manifest la importància de conèixer informació sobre els resultats de l'experiment a l'hora d'establir les probabilitats. En aquest exemple, si anomenem A a l'esdeveniment {guanyar el cavall núm. 2} i B a l'esdeveniment {guanyar un cavall amb núm. parell}; la probabilitat del succés A no és la mateixa quan se sap que passa B quan no se sap.

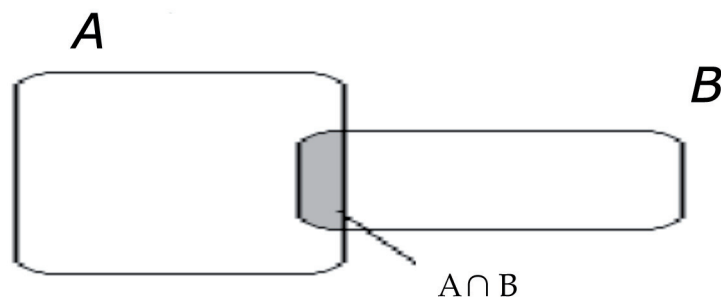
D'aquesta manera el primer cas és: $P(A) = \frac{1}{15}$

i en el segon cas és: $P(A \text{ sabent que passa B}) = P(A/B) = \frac{1}{7}$.

El càlcul de probabilitats en aquest tipus de situacions s'anomena càlcul de probabilitats condicionades, perquè l'ocurrència o no d'uns esdeveniments influeix en les probabilitats dels altres.

Definició de probabilitat condicionada

Donats dos successos A i B de l'espai mostral E, s'anomena probabilitat de A condicionada a B i s'escriu $P(A/B)$ a la probabilitat que ocorregui el succés A considerant que abans ha ocorregut el succés B.



Si ha ocorregut B, es tindrà un nou espai mostral, $E_B = E \cap B$, ja que l'ocurrència de qualsevol altre succés ha de tindre en compte que B haja ocorregut, i així:

$$P(A/B) = \frac{\text{núm. casos favorables en } A \cap B}{\text{núm. casos favorables en } B} = \frac{\text{núm. casos favorables en } A \cap B}{\frac{\text{n}^\circ \text{ casos possibles en } E}{\text{n}^\circ \text{ casos possibles en } B}} = \frac{P(A \cap B)}{P(B)}$$

És a dir: $P(A/B) = \frac{P(A \cap B)}{P(B)}$

Nota

En ocasions, la fórmula $P(A/B)$ trobada anteriorment no és necessària perquè és molt senzill calcular l'espai mostral E_B . Tanmateix, en la gran majoria dels casos, el fet de conèixer E_B és realment complicat.

Probabilitat composta (teorema del producte)

Del concepte de probabilitat condicional se'n dedueixen algunes conseqüències. La primera d'elles és el teorema del producte. El següent exemple l'introdueix.

De la definició de la probabilitat condicionada es pot deduir que si E és un espai mostral, donats dos successos A i B d' E de manera que $P(A) > 0$ i $P(B) > 0$, es compleix:

$$\begin{aligned}P(A \cap B) &= P(A / B) \cdot P(B) \\P(B \cap A) &= P(B / A) \cdot P(A)\end{aligned}$$

Açò és així, perquè per la definició de la probabilitat condicionada

$$P(A / B) = \frac{P(A \cap B)}{P(B)} \longrightarrow P(B \cap A) = P(A/B) \cdot P(B).$$

Anàlogament, $P(B/A) = \frac{P(A \cap B)}{P(A)} \longrightarrow P(B \cap A) = P(B/A) \cdot P(A)$

Teorema de la probabilitat total

Una altra conseqüència de la probabilitat condicionada és el teorema que seguidament es presenta i que permet calcular la probabilitat d'un esdeveniment si es coneix una «divisió» del succés i les seues probabilitats.

Donat un espai mostral E i un conjunt de successos $A_1, A_2, \dots, A_n \in \wp(E)$ i partició de E , és a dir:

- $\bigcup_{i=1}^n A_i = E$
- $A_i \cap A_j = \emptyset \quad \forall i \neq j$

Si es coneixen les probabilitats condicionades $P(B/A_i), \forall i$

$$\text{es té } P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B/A_i) \cdot P(A_i)$$

Per a comprovar-ho, es considera que la següent figura és l'espai mostral E .

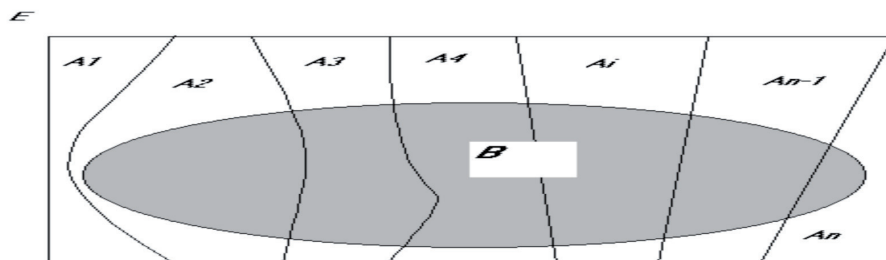


Fig. 3

B es pot expressar: $B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup \dots \cup (B \cap A_i) \cup \dots \cup (B \cap A_n)$

Com que tots els A_i són disjunts, tots els $(B \cap A_i)$ també ho són, llavors:

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + \dots + P(B \cap A_n)$$

I aplicant el teorema del producte tenim:

$$\begin{aligned} P(B) &= P(B/A_1) \cdot P(A_1) + P(B/A_2) \cdot P(A_2) + \dots + P(B/A_n) \cdot P(A_n) \\ &= \sum_{i=1}^n P(B/A_i) \cdot P(A_i) \end{aligned}$$

Teorema de Bayes

Una conseqüència del teorema anterior és el teorema de Bayes, que permet calcular probabilitats condicionades.

Siga E un espai mostral i un conjunt de successos, $\{A_i\} \in \wp(E)$, de manera que:

- $\bigcup_i A_i \{A_i\} = E$
- $A_i \cap A_j = \emptyset \forall i \neq j$
- Són coneguts $P(A_i)$ per a qualsevol i , i $P(A_i) > 0$
- Siga B un succés tal que $P(B) > 0$ i del qual es coneixen $P(B/A_i) \forall_i$

Llavors:

$$P(A_i/B) = \frac{P(B/A_i) \cdot P(A_i)}{P(B/A_1) \cdot P(A_1) + P(B/A_2) \cdot P(A_2) + \dots + P(B/A_n) \cdot P(A_n)}$$

És a dir:

$$P(A_i/B) = \frac{P(B/A_i) \cdot P(A_i)}{\sum_{j=1}^n P(B/A_j) \cdot P(A_j)}$$

Per a comprovar-ho, cal emprar el teorema anterior i la definició de probabilitat condicionada.

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)}$$

Si en el numerador s'aplica el teorema del producte i en el denominador la probabilitat total, queda:

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(A)} = \frac{P(B/A_i) \cdot P(A_i)}{\sum_{i=1}^n P(B/A_i) \cdot P(A_i)}$$

Independència de successos

A la nota de l'apartat 6.5.1 anterior s'ha introduït el concepte d'esdeveniments independents. En aquest epígraf es formalitzarà el concepte i es donaran un conjunt de propietats, sent sense dubte la més interessant el teorema de caracterització. Aquest teorema confirma que les conjectures esmentades en la nota anterior són completament certes.

Definició

Es diu que dos successos A i B són *estocàsticament independents* si $P(A/B) = P(A)$, és a dir, que el fet que ocorregui el succés B no influeix en absolut en l'ocurrència del succés A. El següent teorema caracteritza aquest tipus d'esdeveniments, és a dir, si dos successos el compleixen seran independents i, si són independents, llavors han de complir el teorema.

Teorema de caracterització

Dos successos A i B són independents si, i només si, $P(A \cap B) = P(A) \cdot P(B)$.

Demostració:

Se sap que $P(A \cap B) = P(A/B) \cdot P(B)$. Com que se suposa que A i B són independents, es dedueix que $P(A/B) = P(A)$.

Unint ambdues coses, $P(A \cap B) = P(A) \cdot P(B)$.

Ara se suposa que $P(A \cap B) = P(A) \cdot P(B)$. Com que $P(A \cap B) = P(A/B) \cdot P(B)$, substituïm: $P(A) \cdot P(B) = P(A/B) \cdot P(B)$. Per tant, $P(A) = P(A/B)$. Així doncs, els successos A i B són independents.

Nota

Una de les qüestions que més sol confondre els estudiants d'estadística pel que fa a la independència de successos és la següent: són independents els esdeveniments disjunts?

La resposta és clara: NO. Dos esdeveniments disjunts són aquells que tenen intersecció buida. És a dir, si un ocorre l'altre NO pot fer-ho (Un depèn de l'altre.) Formalment també és evident. Si A i B són dos esdeveniments disjunts, tots dos diferents de \emptyset , llavors la $P(A) > 0$ i $P(B) > 0 \rightarrow P(A) \cdot P(B) > 0$. Tanmateix, $P(A \cap B) = P(\emptyset) = 0 \rightarrow$ pel teorema de caracterització A i B no són independents.

Nota

Hi ha experiments que estan formats per la repetició d'experiments de caràcter més simple –com per exemple llançar dues vegades una moneda– i sobre els que es pot fer l'anàlisi de dues maneres: d'una banda es pot considerar com un únic experiment i d'una altra, com una mena de concatenació dels experiments simples que el formen.

Així, un exemple molt simple és l'esmentat anteriorment, que consisteix a llançar dues vegades una moneda equilibrada. L'espai mostral és $E = \{(C, C); (C, +); (+, C); (+, +)\}$ i si l'anàlitzem com un únic experiment obtenim que tots els esdeveniments simples són equiprobables per ser la moneda equilibrada. Llavors:

$$P(C, C) = P(C, +) = P(+, C) = P(+, +) = \frac{1}{4}$$

Fent l'anàlisi com si foren dos experiments, en cadascun d'ells $P(C) = P(+)= \frac{1}{2}$ i l'espai mostral es pot calcular combinant cada resultat del primer experiment simple amb cada resultat del segon: $E = \{(C_1, C_2); (C_1, +_2); (+_1, C_2); (+_1, +_2)\}$.² A més a més, les probabilitats dels elements de l'espai mostral es poden calcular emprant la regla del producte de probabilitats:

$$P(C, C) = P(C_1) \cdot P(C_2/C_1) = P(C_1) \cdot P(C_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}; \text{ per independència de } C_2 \text{ i } C_1.$$

$$P(+, C) = P(+_1) \cdot P(C_2/+_1) = P(+_1) \cdot P(C_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}; \text{ per independència de } C_2 \text{ i } +_1.$$

$$P(C, +) = P(C_1) \cdot P(+_2/C_1) = P(C_1) \cdot P(+_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}; \text{ per independència de } +_2 \text{ i } C_1.$$

$$P(+, +) = P(+_1) \cdot P(+_2/+_1) = P(+_1) \cdot P(+_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}; \text{ per independència de } +_2 \text{ i } +_1.$$

Fent l'anàlisi d'aquesta manera s'observa que, més que d'esdeveniments independents, es pot parlar d'experiments independents. I si això és així, els resultats d'un són independents dels resultats dels altres. A més a més, emprant aquesta òptica, el càlcul de les probabilitats és més senzill.

2. Els subíndexs indiquen l'ordre temporal en què ocorren els successos. A més a més, cal notar que el parell ordenat (A, B) representa $A \cap B$, per això cal emprar els subíndexs.

Per exemple, si considerem l'experiment que consisteix en llançar tres vegades una moneda equilibrada i calcular la probabilitat d'obtenir (C, C, +), l'anàlisi canvia segons que es realitzi d'una manera o d'una altra:

Forma 1

Cal calcular l'espai mostral. Si no és possible, almenys cal saber el nombre d'esdeveniments simples que el formen per a poder aplicar la regla de Laplace. En aquest cas, l'espai mostral:

$$E = \{(C, C, C); (C, C, +); (C, +, C); (C, +, +); (+, +, +); (+, +, C); (+, C, +); (+, C, C)\}$$

$$P(C, C, +) = \frac{\text{casos favorables}}{\text{casos possibles}} = \frac{1}{8}$$

Forma 2

Com que els tres experiments simples són independents, llavors:

$$P(C, C, +) = P(C_1) \cdot P(C_2) \cdot P(+_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

Si en lloc de considerar tres llançaments se'n consideren trenta, la dificultat per a calcular l'espai mostral augmentaria i caldria recórrer a la combinatòria per a esbrinar el nombre de casos possibles i el de casos favorables. Seria, doncs, més recomanable considerar l'experiment com una concatenació de trenta experiments independents simples, on en cadascun d'ells la probabilitat d'obtenir cara en cada llançament és constant i no depèn del resultat obtingut en el llançament anterior.

La pregunta que hom ara ens podem fer és: aquest mètode d'anàlisi únicament es pot realitzar en experiments independents? L'exemple següent respon aquesta qüestió.

Propietats de la independència estocàstica

Si A i B són independents $\longrightarrow \bar{A}$ i B també ho són

Si A i B són independents $\longrightarrow A$ i \bar{B} també ho són

Si A i B són independents $\longrightarrow \bar{A}$ i \bar{B} també ho són

Si hi ha implicació entre A i B \longrightarrow No hi ha independència

(llevat que $A = E$ o que $B = E$)

Objectius

Els problemes han de permetre que els alumnes assoleixen els objectius didàctics següents:

- 1a) Distingir entre fenòmens deterministes i aleatoris.
- 1b) Identificar els fenòmens governats per l'atzar i comprendre'n el concepte intuïtiu de probabilitat.
- 1c) Saber emprar la teoria de conjunts per a representar esdeveniments o successos i les seues relacions.
- 1d) Calcular probabilitats senzilles emprant la regla de Laplace.
- 1e) Calcular probabilitats en què cal aplicar els principis axiomàtics de la teoria de la probabilitat.
- 1f) Diferenciar les situacions on la regla de Laplace no es pot calcular i saber calcular les probabilitats en aquests casos.
- 1g) Diferenciar la probabilitat condicionada de la no condicionada.
- 1h) Saber aplicar la regla del producte; resoldre problemes on convinga aplicar-la.
- 1i) Aplicar el teorema de la probabilitat total i el teorema de Bayes en les situacions que ho requerisquen.

La taula següent ens mostra com estan distribuïts els objectius segons els exercicis:

Objectiu Exercici	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>
1	X	X							
2		X	X						
3			X						
4				X					
5				X					
6					X				
7					X				
8						X			
9						X			
10							X		
11							X		
12							X		
13								X	
14								X	
15									X
16									X
17									X

Enunciats

-
- 1a) Distingir entre fenòmens deterministes i aleatoris.
 - 1b) Identificar els fenòmens governats per l'atzar i comprendre'n el concepte intuïtiu de probabilitat.

Exercici 1

Digues quines de les següents situacions són governades per l'atzar i quines no:

- a) El resultat obtingut en llançar un dau.
- b) Els ingressos d'una societat en un mes, producte de les aportacions mensuals dels seus socis.
- c) Els resultats obtinguts en una travessa.
- d) Nombre de cares obtingudes en llançar dues-centes vegades una moneda.
- e) Nombre de peces defectuoses que s'obtenen en extraure'n 20 d'una capsa en què el 60 % són defectuoses.
- f) Nombre de peces defectuoses que s'obtenen en extraure'n 20 d'una capsa en què el 99,9 % són defectuoses.
- g) Nombre de cridades telefòniques que rep una centraleta en 2 hores on, normalment, se'n reben 50 per hora.
- h) Distància d'un planeta del sistema solar a la Terra en un instant de temps determinat.

-
- 1b) Identificar els fenòmens governats per l'atzar i comprendre'n el concepte intuïtiu de probabilitat.
 - 1c) Saber emprar la teoria de conjunts per representar esdeveniments o successos i les seues relacions.

Exercici 2

Es considera l'experiment de llançar dos daus consecutivament i anotar-ne els resultats:

- a) Troba els diferents resultats de l'experiment (espai mostral).
- b) Digues els esdeveniments elementals que formen l'esdeveniment compost: A = la suma de les xifres és parell. Calcula el seu complementari \overline{A} .
- c) Digues els esdeveniments elementals que formen l'esdeveniment compost: B = la suma de les xifres és senar. Calcula el seu complementari \overline{B} .

- d) Digues els esdeveniments elementals que formen l'esdeveniment compost: $C =$ la suma de les xifres és major que 5. Calcula el seu complementari \bar{C} .
- e) Calcula l'esdeveniment $A - C$.
- f) Calcula l'esdeveniment $A \cup C$ i $A \cap C$.
- g) Calcula l'esdeveniment $(A \cup C) \cup (A \cap \bar{C})$.
- h) Calcula l'esdeveniment $(A \cap C) \cup (A \cap \bar{C})$.

1c) Saber emprar la teoria de conjunts per representar esdeveniments o successos i les seues relacions.

Exercici 3

La taula següent mostra els països, la superfície, la població i la densitat de població de la Unió Europea l'any 2007. Si es considera l'experiment triar un país dels que formen la Unió Europea, respon a les següents qüestions:

- a) Identifica l'espai mostral de l'experiment.
- b) Identifica els següents esdeveniments:
1. $A = \{\text{Tindre una densitat major que la mitjana aritmètica de les densitats dels països que formen la UE}\}$.
 2. $B = \{\text{Tindre una població major que la mitjana aritmètica de les poblacions dels països que formen la UE}\}$.
 3. $C = \{\text{Tindre una superfície major que la mitjana aritmètica de les poblacions dels països que formen la UE}\}$.
- c) Identifica els esdeveniments següents:
1. $(A \cup B)$, $(A \cup C)$ i $(B \cup C)$.
 2. $(A \cap B)$, $(A \cap C)$ i $(B \cap C)$.
- d) És cert A implica B ? I que B implica A ?
- e) Troba un esdeveniment que impliqui B i un que impliqui A .
- f) Calcula els esdeveniments $A \setminus B$, $C \setminus B$, $B \setminus C$.

Espanya y la UE de los 27			
	Superficie (km ²)	Población 2006 (miles)	Densidad de población*
UE-27	4.309.289	492.852,4	114,4
Alemania	357.030	82.438,0	230,9
Austria	83.871	8.265,9	98,6
Bélgica	30.528	10.511,4	344,3
Bulgaria	111.002	7.718,8	69,5
Chipre	5.695	766,4	134,6
Dinamarca	43.098	5.427,5	125,9
Eslovaquia	49.034	5.389,2	109,9
Eslovenia	20.273	2.003,4	98,8
Espanya	505.990	43.758,3	86,5
Estonia	43.698	1.344,7	30,8
Finlandia	338.145	5.255,6	15,5
Francia	543.965	62.886,2	115,6
Grecia	131.957	11.125,2	84,3
Hungria	93.030	10.076,6	108,3
Irlanda	69.797	4.209,0	60,3
Italia	301.336	58.751,7	195,0
Letonia	64.589	2.294,6	35,5
Lituania	65.300	3.403,3	52,1
Luxemburgo	2.586	459,5	177,7
Malta	315	404,3	1.283,5
Països Baixos	41.543	16.334,2	393,2
Polonia	312.685	38.157,1	122,0
Portugal	92.117	10.569,6	114,7
Reino Unido	243.069	60.393,1	248,5
República Checa	78.868	10.251,1	130,0
Rumania	238.391	21.610,2	90,7
Suecia	441.370	9.047,8	20,5

*Densidad de población: habitantes por km² Fuente: Eurostat

Font: INE

1d) Calcular probabilitats senzilles emprant la regla de Laplace.

Exercici 4

En una empresa de Castelló treballen 200 persones. 80 saben parlar anglés, 110 saben francès i 60, alemany. 50 persones saben parlar francès i anglés, 35 saben parlar francès i alemany i 40 saben parlar anglés i alemany. A més a més, 30 treballadors saben parlar els tres idiomes.

En aquestes condicions, contesta:

- Quantes persones saben parlar únicament anglés?
- Calcula la probabilitat que en escollir una persona a l'atzar de l'empresa únicament sàpia parlar anglés.
- Calcula la probabilitat que en escollir una persona a l'atzar de l'empresa no sàpia parlar cap dels tres idiomes.

1d) Calcular probabilitats senzilles emprant la regla de Laplace.

Exercici 5

En un poble d'una província espanyola es va realitzar una enquesta per a conèixer l'acceptació de tres productes (A, B i C) als 450 veïns que formen el cens del poble. Així, hi havia 250 persones que comprarien el producte A, 250 persones que comprarien el producte B i 250 persones que comprarien el C. A més a més, n'hi havia 150 que comprarien els productes A i B, 160 que comprarien els productes A i C i 140 que comprarien els productes B i C. Per últim, 100 persones digueren que comprarien els tres productes.

- a) Si s'escull una persona completament a l'atzar de les persones empadronades al poble, quina és la probabilitat que no estigués interessada en comprar cap dels tres productes?
- b) Si s'escull una persona completament a l'atzar de les persones empadronades al poble, quina és la probabilitat que estigués interessada en comprar el producte A o el producte B?

1e) Calcular probabilitats en què cal aplicar els principis axiomàtics de la teoria de la probabilitat.

Exercici 6

La taula següent mostra els adults segons la relació amb l'activitat en 2004.

- a) Calcula la probabilitat que en escollir una persona adulta a l'atzar siga assalariada.
- b) Calcula la probabilitat que en escollir una persona adulta a l'atzar estiga Inactiva, sense calcular el nombre de persones inactives.
- c) Calcula la probabilitat que en escollir una persona a l'atzar siga assalariada o inactiva.
- d) Calcula la probabilitat que en escollir una persona a l'atzar siga inactiva o jubilada.
- e) Calcula la probabilitat que en escollir una persona a l'atzar siga inactiva i jubilada.

	Número de personas (miles)
TOTAL ADULTOS	36.187,6
Relación con la actividad	
ACTIVOS	19.067,4
Ocupados	17.139,3
Asalariados	
Empresarios	2.880,0
Parados	1.928,0
INACTIVOS	
Jubilados	5.477,6
Otros inactivos	11.573,4
NO CONSTA	69,2

Font: Anuari 2007 INE

-
- 1e) Calcular probabilitats en què cal aplicar els principis axiomàtics de la teoria de la probabilitat.

Exercici 7

Es desitja cobrir un lloc de treball en una empresa. A la prova es presenta un 55 % de ciutadans espanyols i un 65 % de ciutadans americans. Sabent que a la prova es presenten persones amb dues nacionalitats, calcula la probabilitat que, si fos l'atzar qui decidira, la persona que guanyés el lloc de treball no tingués doble nacionalitat.

-
- 1f) Diferenciar les situacions on la regla de Laplace no es pot calcular i saber calcular les probabilitats en aquests casos.

Exercici 8

La taula mostra l'alumnat estranger classificat pel tipus d'ensenyament a Espanya al llarg del curs 2004/2005. Aquest tipus d'ensenyament es poden classificar en tres grups atenent a l'edat de l'alumne i al grau de maduresa. Així, el grup 1 està format pels ensenyaments d'educació infantil, educació primària i educació especial. El grup 2 per l'educació secundària obligatòria i pel batxillerat. El grup 3, per la resta d'ensenyaments. Es considera l'experiment seleccionar aleatòriament un alumne estranger resident en Espanya l'any 04/05 i preguntar-li a quin grup pertany.

- a) Quin és l'espai mostral d'aquest experiment? Els esdeveniments elementals tenen la mateixa probabilitat de donar-se?
- b) Calcula la probabilitat de cada esdeveniment elemental.
- c) Quina relació hi ha entre la probabilitat del succés cursar ESO i el succés pertànyer al grup 2?

	2004-05
TOTAL	459.291
Educación Infantil/ Preescolar	85.799
Educación Primaria/ E.G.B	198.165
Educación Especial	1.572
E.S.O	124.714
Bachilleratos	19.160
Formación Profesional	19.255
Enseñanzas Artísticas	1.816
Enseñanzas de Idiomas	8.795
Enseñanzas Deportivas	15
No consta enseñanza (Enseñanzas Régimen General)	0

Fuente de información: Ministerio de Educación y Ciencia.

Font: Anuari 2007, INE

- 1f) Diferenciar les situacions on la regla de Laplace no es pot calcular i saber calcular les probabilitats en aquests casos.

Exercici 9

En una empresa, en el departament de recursos humans hi ha tres seccions. En la primera secció treballen 2 homes i 3 dones, en la segona secció hi ha 2 dones i 1 home i en la tercera secció treballen 2 homes. L'empresa ha decidit nomenar una persona d'aquelles que formen part del personal de recursos humans per a acomiadar un conjunt de treballadors. Si l'elecció és completament aleatòria, calcula:

- a) La probabilitat que la persona seleccionada siga un home. Calcula la probabilitat que la persona seleccionada siga una dona.
- b) La probabilitat que la persona escollida pertanga a cadascuna de les seccions.
- c) Digues els esdeveniments elementals de l'experiment «secció a què pertany la persona encarregada de comunicar els acomiadaments» i calcula'n les probabilitats.

1g) Diferenciar la probabilitat condicionada de la no condicionada.

Exercici 10

La taula següent mostra els adults segons la seua relació amb l'activitat en 2004.

	Número de personas (miles)
TOTAL ADULTOS	36.187,6
Relación con la actividad	
ACTIVOS	19.067,4
Ocupados	17.139,3
Asalariados	14.259,3
Empresarios	2.880,0
Parados	1.928,0
INACTIVOS	17.051,0
Jubilados	5.477,6
Otros inactivos	11.573,4
NO CONSTA	69,2

Font: Anuari 2007, INE

- Calcula la probabilitat que una persona adulta escollida a l'atzar de les que apareixen en la taula siga empresària.
- Calcula la probabilitat que una persona adulta escollida a l'atzar siga empresària si se sap que està ocupada.
- Calcula la probabilitat que una persona adulta escollida a l'atzar d'entre les persones actives siga empresària.
- Fes un diagrama d'arbre en què el node inicial siguen les persones ocupades.

1g) Diferenciar la probabilitat condicionada de la no condicionada.

Exercici 11

Tenint en compte les dades mostrades en la taula següent, calcula:

- La probabilitat que una empresa escollida a l'atzar d'entre les de Catalunya i la Comunitat Valenciana siga de Castelló.

- b) La probabilitat que una empresa escollida a l'atzar siga d'Alacant si se sap que l'empresa és de la Comunitat Valenciana.
- c) La probabilitat que una empresa escollida a l'atzar siga catalana i de Lleida.
- d) Si se sap que l'empresa és catalana, quina és la probabilitat que siga de Lleida?

Número de Empresas. 2006	Empleo y paro. 2006			
	Ocupados (miles)	Tasa de actividad (%)	Tasa de paro (%)	
578.340	3.418,7	62,2	6,6	Cataluña
444.410	2.532,2	62,0	6,7	Barcelona
47.169	347,2	66,3	6,6	Girona
33.956	187,6	58,3	6,3	Lleida
52.805	351,8	62,0	6,3	Tarragona
348.692	2.152,3	59,6	8,4	Comunitat Valenciana
133.016	776,4	58,9	9,6	Alicante/Alacant
39.749	259,6	60,0	6,1	Castellón/Castelló
175.927	1.116,3	60,1	8,0	Valencia/València

Font: Anuari 2007, INE

- 1g) Diferenciar la probabilitat condicionada de la no condicionada.

Exercici 12

La taula següent mostra un extracte dels resultats nacionals d'activitat, treball i atur l'any 2006. Les dades vénen donades en percentatges. A més a més, se sap que 17.412 milers de persones formen la població activa, de les quals 14.689 estan ocupades.

Actius per sectors econòmics, 2006

PORCENTAJE	
Agricultura, silvicultura, caza y pesca	4,80
Industria (sin construcción)	15,93
Construcción	12,53
Servicios	63,36
No clasificables	3,39

Ocupats per sectors econòmics, 2006

PORCENTAJE	
Agricultura, silvicultura, caza y pesca	4,78
Industria (sin construcción)	16,67
Construcción	12,88
Servicios	65,67

Aturats per sectors econòmics, 2006

Agricultura, silvicultura, caza y pesca	4,94
Industria (sin construcción)	7,96
Construcción	8,80
Servicios	38,47
No clasificables	39,83

Font: Anuari 2007, INE

S'escull una persona a l'atzar de les persones actives:

- Calcula la probabilitat que una persona activa estiga ocupada i la probabilitat que una persona activa estiga a l'atur.
- Calcula la probabilitat que una persona es dedique al sector de la construcció si se sap que forma part de la població aturada.
- Calcula la probabilitat que una persona activa estiga aturada i que es dedique al sector de la construcció.
- Quina relació i quines diferències hi ha entre els dos apartats anteriors?

1h) Saber aplicar la regla del producte; resoldre problemes on convinga aplicar-la.

Exercici 13

Una urna conté 10 boles verdes, 5 boles roges i 8 boles negres. S'extreuen 3 boles aleatòriament de l'urna. Calcula la probabilitat d'obtenir la seqüència verda, roja, negra en els següents casos:

- La bola extreta no es retorna a l'urna.
- La bola extreta es retorna a l'urna.

- 1h) Saber aplicar la regla del producte; resoldre problemes on convinga aplicar-la.

Exercici 14

La taula següent mostra un extracte de la taula d'empreses segons la seua condició jurídica per estrats d'assalariats l'any 2006.

	Condición jurídica			Personas físicas
	Total	Sociedades anónimas	Sociedades de responsabilidad limitada	
TOTAL	3.174.393	117.441	992.658	1.791.126
Sin asalariados	1.616.883	25.286	267.857	1.202.750
De 1 a 2 asalariados	881.748	19.908	328.565	448.155
De 3 a 5 asalariados	328.820	13.592	175.633	105.500
De 6 a 9 asalariados	154.635	12.880	101.321	25.635
De 10 a 19 asalariados	105.470	16.958	72.051	7.591
De 20 a 49 asalariados	58.725	16.425	35.485	1.495
De 50 a 99 asalariados	15.087	6.059	7.005	0
De 100 a 199 asalariados	7.357	3.325	3.023	0
De 200 a 499 asalariados	3.908	1.994	1.342	0
De 500 a 999 asalariados	999	574	253	0
De 1.000 a 4.999 asalariados	658	372	118	0
De 5.000 o más asalariados	93	68	5	0

Font: Anuari 2007, INE

- a) Calcula la probabilitat d'escollir dues empreses a l'atzar i que una siga una societat anònima i l'altra siga una persona física.
- b) Calcula la probabilitat que, en escollir dues empreses a l'atzar d'aquelles que tenen més de 5.000 assalariats, una siga SA i l'altra, SL.

- 1i) Aplicar el teorema de la probabilitat total i el teorema de Bayes en les situacions que ho requerisquen.

Exercici 15

En una ciutat castellonenca hi ha un 45 % de ciutadans de Castelló, un 35 % de ciutadans espanyols però que no són de Castelló i la resta són de la Unió Europea. Se sap que un 20 % dels ciutadans de Castelló tenen accions en borsa i un 15 % dels ciutadans espanyols també les tenen, mentre que únicament un 4 % dels ciutadans de la Unió Europea també tenen accions.

Calcula molt acuradament la probabilitat que, en escollir una persona a l'atzar de les que viuen en l'esmentada ciutat, tinga accions en borsa. Justifica l'ús de les fórmules que utilitzes.

-
- 1i) Aplicar el teorema de la probabilitat total i el teorema de Bayes en les situacions que ho requerisquen.

Exercici 16

Una empresa té dues seccions de fabricació del seu producte. La secció 1 produeix el 60 % dels productes que l'empresa exporta a l'estranger. El 40 % restant el produeix la secció 2. D'altra banda, l'empresa coneix que la secció 1 elabora un 4 % de productes defectuosos i la secció 2 n'elabora un 8 %. Un client estranger ha emés una queixa perquè el producte que ha rebut és defectuós. Calcula la probabilitat que haja estat produït per la secció 1.

-
- 1i) Aplicar el teorema de la probabilitat total i el teorema de Bayes en les situacions que ho requerisquen.

Exercici 17

La taula següent mostra dades referents a la Comunitat Valenciana.

	Población a 1 de enero de 2006		
	Total	Mujeres (%)	Extranjeros (%)
Comunitat Valenciana	4.806.908		
Alicante/Alacant	1.783.555	49,9	20,1
Castellón/Castelló	559.761	49,7	13,9
Valencia/València	2.463.592	50,5	9,4

Font: INE

- a) Calcula la probabilitat que en escollir una persona a l'atzar de la Comunitat Valenciana siga dona.
- b) S'ha escollit una persona a l'atzar de la Comunitat Valenciana i ha resultat ser home. Calcula la probabilitat que siga de Castelló.

Ajudes

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de resoldre els exercicis i problemes. És convenient no fer un ús excessiu d'aquestes ajudes, és a dir, que abans d'emprar l'ajuda cal pensar el problema almenys durant uns 10-15 minuts. Després es consultarà l'ajuda de tipus 1 i s'intentarà resoldre l'exercici amb aquesta ajuda. Si no és possible resoldre'l, llavors es consultarà l'ajuda de tipus 2; en darrer terme es consultarà la solució.

Ajudes de tipus 1

Exercici 2

Escriu diferents possibilitats d'una manera ordenada, per exemple: (1, 1); (1, 2), ...

Exercici 3

En primer lloc, cal calcular les mitjanes aritmètiques de la densitat, la població i la superfície. Després cal comparar-les amb les respectives de cada país. D'aquesta manera s'obtenen els països que formaran cada esdeveniment: A, B o C.

Exercici 4

Representa els esdeveniments gràficament. Fixa't en les interseccions a l'hora de fer la representació gràfica.

Exercici 5

Reviseu l'exercici 4, car cal fer el mateix.

Exercici 6

En aquest exercici cal, en primer lloc, identificar l'espai mostral i les diferents particions d'aquest que es troben reflectides a la taula. Així, l'espai mostral està format per totes les persones adultes: 36.187,6. D'altra banda, estar actiu o, estar inactiu i que no conste l'estat d'activitat és una partició de l'espai mostral.

Exercici 7

Cal identificar l'espai mostral i la partició d'aquest. Recordeu que el fet que $0,55 + 0,65 > 1$ implica necessàriament que hi haja persones amb dues nacionalitats, és a dir, aspirants que són espanyols i americans.

Exercici 8

Cada esdeveniment elemental de l'espai mostral està format pel mateix nombre d'estudiants estrangers? La resposta a aquesta pregunta permet respondre a la pregunta sobre si els esdeveniments elementals són igualment probables.

Exercici 9

Quin és l'espai mostral de l'experiment? Quantes persones són dones? Quantes persones pertanyen a cada secció? La resposta a aquestes preguntes permetrà resoldre els exercicis.

Exercici 10

Els apartats b i c són probabilitats condicionades. Es poden fer els càlculs demanats utilitzant la definició de probabilitat condicionada o calculant la probabilitat directament.

Exercici 11

Els apartats b i c són probabilitats condicionades. Es poden fer els càlculs demanats utilitzant la definició de probabilitat condicionada o calculant la probabilitat directament.

Exercici 12

Els apartats *b* i *c* són una probabilitat condicionada i una probabilitat d'una intersecció.

Exercici 13

En aquest exercici es realitzen tres experiments consecutius: extraure tres boles d'una urna. En els dos apartats es demana el mateix: la probabilitat que es done verda la bola, roja la segona bola i negra la tercera bola. Una manera d'analitzar el problema és mitjançant un diagrama d'arbre on les branques són probabilitats condicionades.

Exercici 14

- a) Reviseu l'ajuda de tipus 1 de l'exercici 13.
- b) Cal tenir en comte que la probabilitat de les empreses possibles de ser escollides és baixa en aquest cas.

Exercici 15

El problema demana la probabilitat de que, en escollir una persona a l'atzar de les que viuen en la ciutat, tinga accions en borsa. És evident, no obstant, que a l'hora de fer el càlcul probabilístic cal tenir en compte d'on és cada persona, ja que no és el mateix que la persona escollida siga de Castelló, de la UE o espanyola. Feu un dibuix per aclarir l'enunciat.

Exercici 16

En aquest exercici es parteix d'un esdeveniment que ja ha passat: un producte dels exportats a l'estranger ha resultat ser defectuós. Es demana la probabilitat que aquest producte haja estat fabricat per la secció 1. Realitzar un gràfic pot ajudar a aclarir l'enunciat.

Exercici 17

A l'apartat *a* convindria fer un diagrama d'arbre. Per a fer-lo, busca en primer lloc una partició de l'espai mostral.

A l'apartat *b* cal tenir en compte que el problema ja ens dóna una informació: la persona escollida ha estat home.

Ajudes de tipus 2

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de resoldre els exercicis i problemes, després d'haver consultat l'ajuda de tipus 1.

Exercici 2

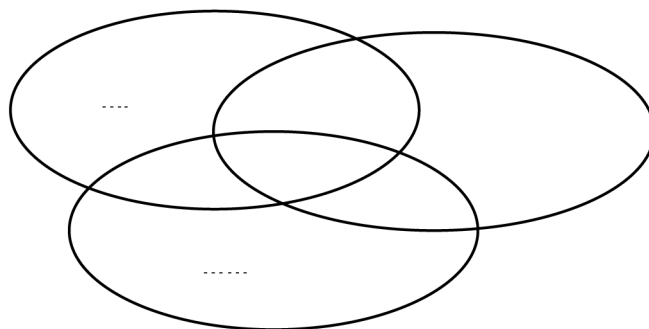
Reviseu l'ajuda de tipus 1.

Exercici 3

Simplement s'han d'anar aplicant les definicions dels conceptes d'unió, intersecció, complementari, etc. En tot cas, consulteu els exercicis anteriors.

Exercici 4

Representeu els successos de la següent manera:



Exercici 5

Reviseu l'ajuda de l'exercici anterior.

Exercici 6

A l'apartat *b* cal aplicar els axiomes de probabilitat: la probabilitat que se satisfaci l'espai mostral E és 1, i la probabilitat d'esdeveniments disjunts és igual a la suma de les probabilitats de cadascun dels esdeveniments. Reviseu l'ajuda de tipus 1.

Exercici 7

Cal relacionar la probabilitat de la unió amb la probabilitat de la intersecció dels esdeveniments ser espanyol i ser americà.

Exercici 8

Per a respondre a l'apartat *c* cal buscar un experiment on cursar ESO i pertànyer al grup 2 siguin esdeveniments.

Exercici 9

Reviseu l'ajuda de tipus 1.

Exercici 10

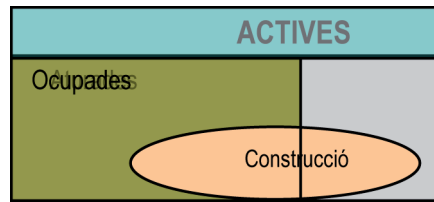
Reviseu l'ajuda de tipus 1.

Exercici 11

Reviseu l'ajuda de tipus 1.

Exercici 12

El gràfic següent és un resum de la informació pel que fa a les persones que treballen en la construcció i estan actives.



Exercici 13

Cal aplicar la definició de probabilitat condicionada de l'esdeveniment següent:

$$P(V1 \cap R2 \cap N3) = P(N3/V1 \cap R2) \cdot P(V1 \cap R2) = \\ P(N3/V1 \cap R2) \cdot P(R2/V1) \cdot P(V1)$$

On:

V1 = primera bola extreta és verda.

R2 = segona bola extreta és roja.

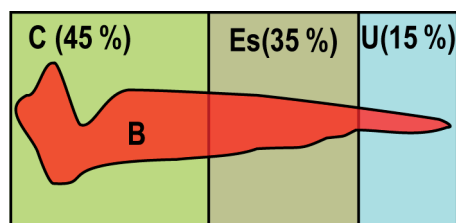
N3 = tercera bola extreta és negra.

Exercici 14

Reviseu l'ajuda de tipus 1.

Exercici 15

Un dibuix que representa l'enunciat del problema pot ser:



On:

C = la persona escollida és de Castelló.

Es = la persona escollida és espanyola, no de Castelló.

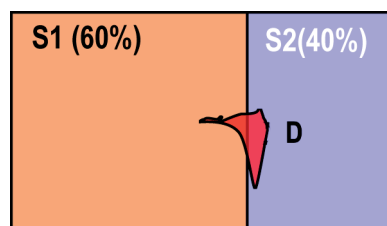
U = la persona escollida és de la UE.

B = la persona escollida té accions en borsa.

Per a calcular la probabilitat de B cal emprar les probabilitats de $B \cap C$, $B \cap Es$ i $B \cap U$.

Exercici 16

Un dibuix que representa l'enunciat del problema pot ser:



On:

S1 = el producte prové de la secció 1.

S2 = el producte prové de la secció 2.

D = el producte és defectuós.

L'exercici mostra, intuïtivament, el quocient:

$$\frac{\text{S1}}{\text{D}} : \frac{\text{S2}}{\text{D}}$$

que formalment és: $P(S1/D)$.

Llavors, cal aplicar el teorema adequat que permeti obtenir la probabilitat que es busca.

Exercici 17

Reviseu l'ajuda de tipus 1. També l'ajuda de tipus 1 de l'exercici 16.

Solucions

Exercici 1

Digues quines de les següents situacions són governades per l'atzar i quines no:

- a) El resultat obtingut en llançar un dau.
- b) Els ingressos d'una societat en un mes, producte de les aportacions mensuals dels seus socis.
- c) Els resultats obtinguts en una travessa.
- d) Nombre de cares obtingudes en llançar dues-centes vegades una moneda.
- e) Nombre de peces defectuoses que s'obtenen en extraure'n 20 d'una capsa en què el 60 % són defectuoses.
- f) Nombre de peces defectuoses que s'obtenen en extraure'n 20 d'una capsa en què el 99,9 % són defectuoses.
- g) Nombre de cridades telefòniques que rep una centraleta en 2 hores on, normalment, se'n reben 50 per hora.
- h) Distància d'un planeta del sistema solar a la Terra en un instant de temps determinat.

Solució

- a) És clarament un experiment governat per l'atzar, ja que no se'n pot saber el resultat amb total certesa.
- b) És un fenomen determinista; cada soci paga una quota i al final de cada mes la societat té uns ingressos determinats.
- c) És un experiment aleatori. No es poden predir amb total exactitud els signes de la travessa.
- d) És aleatori, ja que en cada llançament no es pot saber si sortirà cara o creu.
- e) És aleatori, perquè encara que siga més probable extraure peces defectuoses, és possible extraure'n de no defectuoses.
- f) Està governat per l'atzar per la mateixa raó que l'apartat anterior. En aquest cas és més probable traure més peces defectuoses que en l'apartat anterior –n'hi ha més a la capsa–, però pot ser que no se'n traga cap.
- g) És aleatori perquè, encara que normalment se'n reben 50, aquest fet no és segur i, en conseqüència, no és possible predir amb exactitud el nombre de cridades.
- h) És determinat, ja que la posició dels planetes està regulada per lleis físiques.

Exercici 2

Es considera l'experiment de llançar dos daus consecutivament i anotar-ne els resultats:

- Troba els diferents resultats de l'experiment (espai mostral).
- Digues els esdeveniments elementals que formen l'esdeveniment compost:
 $A =$ la suma de les xifres és parell. Calcula el seu complementari \bar{A} .
- Digues els esdeveniments elementals que formen l'esdeveniment compost:
 $B =$ la suma de les xifres és senar. Calcula el seu complementari \bar{B} .
- Digues els esdeveniments elementals que formen l'esdeveniment compost:
 $C =$ la suma de les xifres és major que 5. Calcula el seu complementari \bar{C} .
- Calcula l'esdeveniment $A - C$.
- Calcula l'esdeveniment $A \cup C$ i $A \cap C$.
- Calcula l'esdeveniment $(A \cup C) \cup (A \cap C)$.
- Calcula l'esdeveniment $(A \cap C) \cup (A \cap \bar{C})$.

Solució

Es considera l'experiment de llançar dos daus consecutivament i anotar-ne els resultats:

- L'espai mostral està format pels possibles resultats de l'experiment. Si en aquest cas s'utilitza una notació en forma de tupla, en què la primera xifra fa referència al primer resultat del dau i la segona xifra al segon espai formal, és:

$$E = \{(1, 1); (1, 2); (1, 3); (1, 4); (1, 5); (1, 6); \\ (2, 1); (2, 2); (2, 3); (2, 4); (2, 5); (2, 6); \\ (3, 1); (3, 2); (3, 3); (3, 4); (3, 5); (3, 6); \\ (4, 1); (4, 2); (4, 3); (4, 4); (4, 5); (4, 6); \\ (5, 1); (5, 2); (5, 3); (5, 4); (5, 5); (5, 6); \\ (6, 1); (6, 2); (6, 3); (6, 4); (6, 5); (6, 6)\}$$

- L'esdeveniment $A =$ la suma de les seues xifres és parell, és el mateix que l'esdeveniment «les dues xifres són les dues parells» o «les dues senars». Aleshores:

$$A = \{(1, 1); (1, 3); (1, 5); (2, 2); (2, 4); (2, 6); (3, 1); (3, 3); \\ (3, 5); (4, 2); (4, 4); (4, 6); (5, 1); (5, 3); (5, 5); (6, 2); \\ (6, 4); (6, 6)\}$$

Evidentment, el complementari de A , està format per tots aquells esdeveniments de l'espai mostral E que no estan en A :

$$\bar{A} = E - A = \{(1, 2); (1, 4); (1, 6); (2, 1); (2, 3); (2, 5); (3, 2); \\ (3, 4); (3, 6); (4, 1); (4, 3); (4, 5); (5, 2); (5, 4); \\ (5, 6); (6, 1); (6, 3); (6, 5)\}$$

- c) És evident que els esdeveniments B i \bar{A} són iguals, ja que la suma de dos nombres enters pot ser únicament o parell o senar. Com que A és el succés «la suma de les xifres és parell», llavors el complementari és B .

Així doncs:

$$B = \bar{A} \text{ i } \bar{B} = A$$

- d) Digues els esdeveniments elementals que formen l'esdeveniment compost: $C =$ la suma de les xifres és major que 5. Calcula el seu complementari \bar{C} .

El succés $C = \{(1, 5); (1, 6); (2, 4); (2, 5); (2, 6); (3, 3); (3, 4); (3, 5); (3, 6); (4, 2); (4, 3); (4, 4); (4, 5); (4, 6); (5, 1); (5, 2); (5, 3); (5, 4); (5, 5); (5, 6); (6, 1); (6, 2); (6, 3); (6, 4); (6, 5); (6, 6)\}$

El succés $\bar{C} = \{(1, 1); (1, 2); (1, 3); (1, 4); (2, 1); (2, 2); (2, 3); (3, 1); (3, 2); (4, 1)\}$

- e) $A - C$ està format pels esdeveniments que estan en A i no estan en C . És a dir:

$A - C = \{(1, 1); (1, 3); (1, 5); (2, 2); (2, 4); (2, 6); (3, 1); (3, 3); (3, 5); (4, 2); (4, 4); (4, 6); (5, 1); (5, 3); (5, 5); (6, 2); (6, 4); (6, 6)\} - \{(1, 5); (1, 6); (2, 4); (2, 5); (2, 6); (3, 3); (3, 4); (3, 5); (3, 6); (4, 2); (4, 3); (4, 4); (4, 5); (4, 6); (5, 1); (5, 2); (5, 3); (5, 4); (5, 5); (5, 6); (6, 1); (6, 2); (6, 3); (6, 4); (6, 5); (6, 6)\} = \{(1, 1); (1, 3); (2, 2); (3, 1)\}$

$$A - C = \{(1, 1); (1, 3); (2, 2); (3, 1)\}$$

- f) L'esdeveniment $A \cup C$ està format pels esdeveniments que estan en A o bé estan en C . Llavors:

$A \cup C = \{(1, 1); (1, 3); (2, 2); (3, 1); (1, 5); (1, 6); (2, 4); (2, 5); (2, 6); (3, 3); (3, 4); (3, 5); (3, 6); (4, 2); (4, 3); (4, 4); (4, 5); (4, 6); (5, 1); (5, 2); (5, 3); (5, 4); (5, 5); (5, 6); (6, 1); (6, 2); (6, 3); (6, 4); (6, 5); (6, 6)\}$

L'esdeveniment $A \cap C$ està format pels esdeveniments que estan en A que també estan en C . Llavors:

$A \cap C = \{(1, 5); (2, 4); (2, 6); (3, 3); (3, 5); (4, 2); (4, 4); (4, 6); (5, 1); (5, 3); (5, 5); (6, 2); (6, 4); (6, 6)\}$

- g) Com que $(A \cap C)$ és un esdeveniment que està inclòs en $(A \cup C)$, llavors:

$$(A \cup C) \cup (A \cap C) = (A \cup C)$$

h) Com que a la vegada $A \cap C$ està format pels esdeveniments elementals que estan en A i en C , i $A \cap \bar{C}$ està format pels esdeveniments elementals que estan en A i no en C , aleshores $(A \cap C) \cup (A \cap \bar{C})$ està format pels esdeveniments elementals que estan en A i estan o no estan en C és a dir, està format pels esdeveniments elementals que estan en A .

Exercici 3

La taula següent mostra els països, la superfície, la població i la densitat de població de la Unió Europea l'any 2007. Si es considera l'experiment triar un país dels que formen la Unió Europea, respon a les següents qüestions:

- Identifica l'espai mostral de l'experiment.
- Identifica els següents esdeveniments:
 - $A = \{\text{Tindre una densitat major que la mitjana aritmètica de les densitats dels països que formen la UE}\}$.
 - $B = \{\text{Tindre una població major que la mitjana aritmètica de les poblacions dels països que formen la UE}\}$.
 - $C = \{\text{Tindre una superfície major que la mitjana aritmètica de les poblacions dels països que formen la UE}\}$.
- Identifica els esdeveniments següents:
 - $(A \cup B)$, $(A \cup C)$ i $(B \cup C)$.
 - $(A \cap B)$, $(A \cap C)$ i $(B \cap C)$.
- És cert A implica B ? I que B implica A ?
- Troba un esdeveniment que impliqui B i un que impliqui A .
- Calcula els esdeveniments $A \setminus B$, $C \setminus B$, $B \setminus C$.

Espanya y la UE de los 27			
	Superficie (km ²)	Población 2006 (miles)	Densidad de población*
UE-27	4.309.289	492.852,4	114,4
Alemania	357.030	82.438,0	230,9
Austria	83.871	8.265,9	98,6
Bélgica	30.528	10.511,4	344,3
Bulgaria	111.002	7.718,8	69,5
Chipre	5.695	766,4	134,6
Dinamarca	43.098	5.427,5	125,9
Eslovaquia	49.034	5.389,2	109,9
Eslovenia	20.273	2.003,4	98,8
Espanya	505.990	43.758,3	86,5
Estonia	43.698	1.344,7	30,8
Finlandia	338.145	5.255,6	15,5
Francia	543.965	62.886,2	115,6
Grecia	131.957	11.125,2	84,3
Hungria	93.030	10.076,6	108,3
Irlanda	69.797	4.209,0	60,3
Italia	301.336	58.751,7	195,0
Letonia	64.589	2.294,6	35,5
Lituania	65.300	3.403,3	52,1
Luxemburgo	2.586	459,5	177,7
Malta	315	404,3	1.283,5
Països Baixos	41.543	16.334,2	393,2
Polonia	312.685	38.157,1	122,0
Portugal	92.117	10.569,6	114,7
Reino Unido	243.069	60.393,1	248,5
República Checa	78.868	10.251,1	130,0
Rumania	238.391	21.610,2	90,7
Suecia	441.370	9.047,8	20,5

* Densidad de población: habitantes por km² Fuente: Eurostat

Font: INE

Solució

- a) L'espai mostral de l'experiment està format pel conjunt de països que formen la UE, ja que l'experiment consisteix en escollir-ne un.

Per la qual cosa, $E = \{\text{Alemanya, Àustria, Bèlgica, Bulgària, Xipre, Dinamarca, Eslovàquia, Eslovènia, Espanya, Estònia, Finlàndia, França, Grècia, Hongria, Irlanda, Itàlia, Letònia, Lituània, Luxemburg, Malta, Països Baixos, Polònia, Portugal, Regne Unit, República Txeca, Romania, Suècia}\}$

b)

1. $A = \{\text{Tindre una densitat major que la mitjana aritmètica de les densitats dels països que formen la UE}\}$.
2. $B = \{\text{Tindre una població major que la mitjana aritmètica de les poblacions dels països que formen la UE}\}$.
3. $C = \{\text{Tindre una superfície major que la mitjana aritmètica de les poblacions dels països que formen la UE}\}$.

En primer lloc, cal calcular les mitjanes aritmètiques de cada variable: població, superfície i densitat. La taula següent mostra, a més d'aquests càlculs, els països que formen cada esdeveniment:

PAÏSOS UE	SUPERF.	POBLACIÓ	DENSITAT
Finlàndia	338.145,0	5.255,6	15,5
Suècia	441.370,0	9.047,8	20,5
Estònia	43.698,0	1.344,7	30,8
Letònia	64.589,0	2.294,6	35,5
Lituània	65.300,0	3.403,3	52,1
Irlanda	69.797,0	4.209,0	60,3
Bulgària	111.002,0	7.718,8	69,5
Grècia	131.957,0	11.125,2	84,3
Espanya	505.990,0	43.758,3	86,5
Romania	238.391,0	21.610,2	90,7
Àustria	83.871,0	8.265,9	98,6
Eslovènia	20.273,0	2.003,4	98,8
Hongria	93.030,0	10.076,6	108,3
Eslovaquia	49.034,0	5.389,2	109,9
Portugal	92.117,0	10.569,6	114,7
França	543.965,0	62.886,2	115,6
Polònia	312.685,0	38.157,1	122,0
Dinamarca	43.098,0	5.427,5	125,9
República Txeca	78.868,0	10.251,1	130,0
Xipre	5.695,0	766,4	134,6
Luxemburg	2.586,0	459,5	177,7
Itàlia	301.336,0	58.751,7	195,0
Alemanya	357.030,0	82.438,0	230,9

Regne Unit	243.069,0	60.393,1	248,5
Bèlgica	30.528,0	10.511,4	344,3
Països Baixos	41.543,0	16.334,2	393,2
Malta	315,0	404,3	1.283,5
	C	B	A
Mitjanes Aritmètiques	159.603,0	18.253,8	169,5

En conseqüència:

$A = \{\text{Luxemburg, Itàlia, Alemanya, Regne Unit, Bèlgica, Països Baixos, Malta}\}$

$B = \{\text{Espanya, Romania, França, Polònia, Itàlia, Alemanya, Regne Unit}\}$

$C = \{\text{Finlàndia, Suècia, Espanya, Romania, França, Polònia, Itàlia, Alemanya, Regne Unit}\}$

c)

1. $(A \cup B)$, $(A \cup C)$ i $(B \cup C)$.

$(A \cup B) = \{\text{Luxemburg, Itàlia, Alemanya, Regne Unit, Bèlgica, Països Baixos, Malta}\} \cup \{\text{Espanya, Romania, França, Polònia, Itàlia, Alemanya, Regne Unit}\} = \{\text{Luxemburg, Itàlia, Alemanya, Regne Unit, Bèlgica, Països Baixos, Malta, Espanya, Romania, França, Polònia}\}$

$(A \cup C) = \{\text{Luxemburg, Itàlia, Alemanya, Regne Unit, Bèlgica, Països Baixos, Malta}\} \cup \{\text{Finlàndia, Suècia, Espanya, Romania, França, Polònia, Itàlia, Alemanya, Regne Unit}\} = \{\text{Luxemburg, Itàlia, Alemanya, Regne Unit, Bèlgica, Països Baixos, Malta, Finlàndia, Suècia, Espanya, Romania, França}\}$

$(B \cup C) = \{\text{Espanya, Romania, França, Polònia, Itàlia, Alemanya, Regne Unit}\} \cup \{\text{Finlàndia, Suècia, Espanya, Romania, França, Polònia, Itàlia, Alemanya, Regne Unit}\} = \{\text{Finlàndia, Suècia, Espanya, Romania, França, Polònia, Itàlia, Alemanya, Regne Unit}\} = C$

2. $(A \cap B)$, $(A \cap C)$ i $(B \cap C)$.

$A \cap B = \{\text{Itàlia, Alemanya, Regne Unit}\}$

$A \cap C = \{\text{Itàlia, Alemanya, Regne Unit}\}$

$B \cap C = \{\text{Espanya, Romania, França, Polònia, Itàlia, Alemanya, Regne Unit}\} = B$

d) A no implica B perquè no està inclòs en B. És a dir, Pot donar-se A i no donar-se B (per exemple si es dona l'esdeveniment elemental Malta).

Per la mateixa raó, B no està inclòs dintre de A i B no implica A.

e) L'esdeveniment $H = \{\text{Itàlia, Alemanya}\}$ implica B i també implica A. Si es compleix H, llavors també es compleix en A i B.

No obstant, l'esdeveniment $L = \{\text{Malta, Bèlgica}\}$ implica A però no B.

f) $A \setminus B = \{\text{Luxemburg, Bèlgica, Països Baixos, Malta}\}$
 $C \setminus B = \{\text{Finlàndia, Suècia}\}$
 $B \setminus C = \{\emptyset\}$

Exercici 4

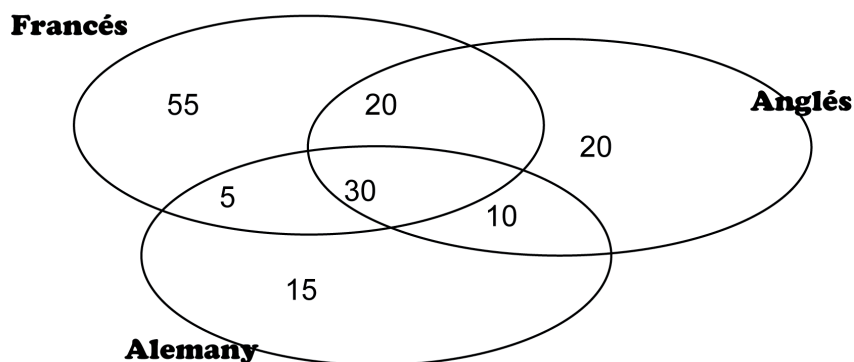
En una empresa de Castelló treballen 200 persones. 80 saben parlar anglés, 110 saben francès i 60, alemany. 50 persones saben parlar francès i anglés, 35 saben parlar francès i alemany i 40 saben parlar anglés i alemany. A més a més, 30 treballadors saben parlar els tres idiomes.

En aquestes condicions, contesta:

- Quantes persones saben parlar únicament anglés?
- Calcula la probabilitat que en escollir una persona a l'atzar de l'empresa únicament sàpia parlar anglés.
- Calcula la probabilitat que en escollir una persona a l'atzar de l'empresa no sàpia parlar cap dels tres idiomes.

Solució

a) Realitzant un gràfic semblant a l'emprat en el problema 12:



Hi ha 20 persones que únicament saben parlar anglés.

b) Emprant la nomenclatura de Laplace:

$$\left. \begin{array}{l} \text{Casos possibles} \rightarrow 200 \\ \text{Casos favorables} \rightarrow 20 \end{array} \right\} \Rightarrow P(\text{únicament anglés}) = \frac{20}{200} = 0,10$$

- c) Comptant totes les persones que saben parlar almenys 1 idioma, és a dir, sumant tots els nombres que hi ha dintre del gràfic, $55 + 5 + 30 + 20 + 10 + 20 + 15 = 155$,

s'obté $P(\text{saber al menys 1 idioma}) = \frac{155}{200}$. Per tant, utilitzant les propietats:

$$P(\text{no saber cap idioma}) = 1 - P(\text{saber al menys 1 idioma}) = 1 - \frac{155}{200} = \frac{45}{200} = 0,225$$

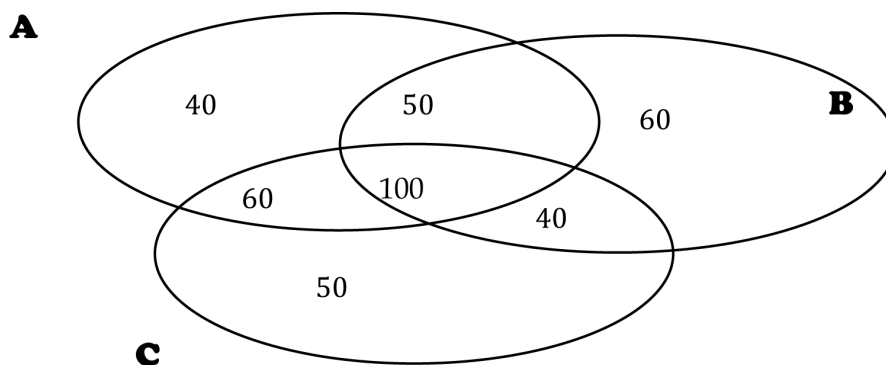
Exercici 5

En un poble d'una província espanyola es va realitzar una enquesta per a conèixer l'acceptació de tres productes (A, B i C) als 450 veïns que formen el cens del poble. Així, hi havia 250 persones que comprarien el producte A, 250 persones que comprarien el producte B i 250 persones que comprarien el C. A més a més, n'hi havia 150 que comprarien els productes A i B, 160 que comprarien els productes A i C i 140 que comprarien els productes B i C. Per últim, 100 persones digueren que comprarien els tres productes.

- a) Si s'escull una persona completament a l'atzar de les persones empadronades al poble, quina és la probabilitat que no estigués interessada en comprar cap dels tres productes?
- b) Si s'escull una persona completament a l'atzar de les persones empadronades al poble, quina és la probabilitat que estigués interessada en comprar el producte A o el producte B?

Solució

- a) Realitzant un gràfic semblant a l'emprat en el problema 12:



Hi ha 400 persones que comprarien almenys un producte. Aleshores, n'hi ha 50 que no en comprarien cap $\rightarrow P(\text{no comprar-ne cap}) = \frac{50}{450} = \frac{1}{9}$

b) En aquest exercici cal aplicar les propietats de la probabilitat. Així:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{250}{450} + \frac{250}{450} - \frac{150}{450} = \frac{350}{450} = \frac{7}{9} = 0,77777$$

Exercici 6

La taula següent mostra els adults segons la relació amb l'activitat en 2004.

- Calcula la probabilitat que en escollir una persona adulta a l'atzar siga assalariada.
- Calcula la probabilitat que en escollir una persona adulta a l'atzar estiga Inactiva, sense calcular el nombre de persones inactives.
- Calcula la probabilitat que en escollir una persona a l'atzar siga assalariada o inactiva.
- Calcula la probabilitat que en escollir una persona a l'atzar siga inactiva o jubilada.
- Calcula la probabilitat que en escollir una persona a l'atzar siga inactiva i jubilada.

	Número de personas (miles)
TOTAL ADULTOS	36.187,6
Relación con la actividad	
ACTIVOS	19.067,4
Ocupados	17.139,3
Asalariados	██████████
Empresarios	2.880,0
Parados	1.928,0
INACTIVOS	██████████
Jubilados	5.477,6
Otros inactivos	11.573,4
NO CONSTA	69,2

Font: Anuari 2007 INE

Solució

- Les persones ocupades es divideixen en la taula en assalariades i empresàries. Com que hi ha 17.139,3 milers d'ocupats, dels quals 2.880 milers són empresaris, llavors hi ha 14.259,3 milers.

$$Aleshores, P(\text{escollir una persona assalariada}) = \frac{14259,3}{36187,6} = 0,394$$

b) Calcula la probabilitat que en escollir una persona adulta a l'atzar estiga inactiva, sense calcular el nombre de persones inactives.

Calcularem en primer lloc les probabilitats següents:

- $P(\text{escollir una persona activa}) = \frac{19067,4}{36187,6} = 0,527$
- $P(\text{escollir una persona de qui no conste activitat}) = \frac{69,2}{36187,6} = 0,00191$

Per una altra part, en la taula una persona pot estar activa, inactiva o que no hi conste la seua activitat. Aquests tres esdeveniments són disjunts i la seua unió forma tot l'espai mostral (en aquest cas les persones adultes d'Espanya l'any 2004). Aplicant els axiomes de probabilitat:

$$P(\{\text{escollir una persona activa}\} \cup \{\text{escollir una persona activa}\} \cup \{\text{escollir una persona de qui no conste activitat}\}) = P(E) = 1$$

Com que els esdeveniments són disjunts:

$$1 = P(\{\text{escollir una persona activa}\} \cup \{\text{escollir una persona que inactiva}\} \cup \{\text{escollir una persona de qui no conste activitat}\}) =$$

$$P(\{\text{escollir una persona que activa}\}) + P(\{\text{escollir una persona que inactiva}\}) + P(\{\text{escollir una persona de qui no conste activitat}\}) =$$

$$0,527 + 0,00191 + P(\{\text{Escollir una persona estiga inactiva}\})$$

$$\text{És a dir: } P(\{\text{escollir una persona inactiva}\}) = 1 - 0,527 - 0,00191 = 0,47109$$

c) Denotarem per A = escollir una persona a l'atzar i que siga assalariada.
 I = escollir una persona a l'atzar i que siga inactiva.

$$\text{El problema demana } P(A \cup I) = P(A) + P(I) - P(A \cap I) = 0,394 + 0,47109 - 0 = 0,865.$$

d) Denotarem per J = escollir una persona a l'atzar i que siga jubilada.
 I = escollir una persona a l'atzar i que siga inactiva.

$$\text{El problema demana } P(J \cup I) = P(J) + P(I) - P(J \cap I).$$

Com que totes les persones jubilades són inactives, llavors $J \cap I = J$ i, per tant, $P(J \cup I) = P(J) + P(I) - P(J) = P(I) = 0,47109.$

e) Denotarem per J = escollir una persona a l'atzar i que siga jubilada.
 I = escollir una persona a l'atzar i que siga inactiva.

El problema demana $P(J \cap I)$.

Com que totes les persones jubilades són inactives, llavors $J \cap I = J$ i, per tant,

$$P(J \cap I) = P(J) = \frac{5477,6}{36187,6} = 0,1514$$

Exercici 7

Es desitja cobrir un lloc de treball en una empresa. A la prova es presenten un 55 % de ciutadans espanyols i un 65 % de ciutadans americans. Sabent que a la prova es presenten persones amb les dues nacionalitats, calcula la probabilitat que, si fos l'atzar qui decidira, la persona que guanyés el lloc de treball no tingués doble nacionalitat.

Solució

Denotarem per Es = ser espanyol; i de l'enunciat $P(Es) = 0,55$.

A = ser americà; i de l'enunciat $P(A) = 0,65$.

Es demana $P(\overline{Es}) = 1 - P(Es)$. Com que tots els aspirants són espanyols, americans o tenen les dues nacionalitats, llavors $P(Es \cup A) = 1$.

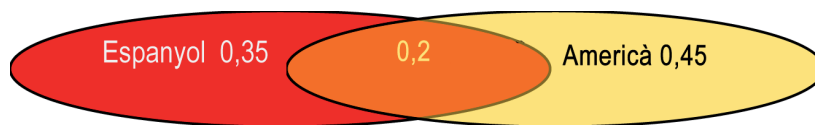
Però, per una altra part, $P(Es \cup A) = P(Es) + P(A) - P(Es \cap A) \rightarrow$

$$1 = 0,55 + 0,65 - P(Es \cap A).$$

$$\text{Per tant, } P(Es \cap A) = 0,55 + 0,65 - 1 = 0,2.$$

És a dir, la probabilitat de doble nacionalitat és 0,2.

Per tant, $P(\overline{Es \cap A}) = 1 - P(Es \cap A) = 1 - 0,2 = 0,8$.



Exercici 8

La taula mostra l'alumnat estranger classificat pel tipus d'ensenyament a Espanya al llarg del curs 2004/2005. Aquest tipus d'ensenyament es poden classificar en tres grups atenent a l'edat de l'alumne i al grau de maduresa. Així, el grup 1 està format pels ensenyaments d'educació infantil, educació primària i educació especial. El grup 2 per l'educació secundària obligatòria i pel batxillerat. El grup 3, per la resta d'ensenyaments. Es considera l'experiment seleccionar aleatòriament

un alumne estranger resident en Espanya l'any 04/05 i preguntar-li a quin grup pertany.

- Quin és l'espai mostral d'aquest experiment? Els esdeveniments elementals tenen la mateixa probabilitat de donar-se?
- Calcula la probabilitat de cada esdeveniment elemental.
- Quina relació hi ha entre la probabilitat del succés cursar ESO i el succés pertànyer al grup 2?

	2004-05
TOTAL	459.291
Educación Infantil/ Preescolar	85.799
Educación Primaria/ E.G.B	198.165
Educación Especial	1.572
E.S.O	124.714
Bachilleratos	19.160
Formación Profesional	19.255
Enseñanzas Artísticas	1.816
Enseñanzas de Idiomas	8.795
Enseñanzas Deportivas	15
No consta enseñanza (Enseñanzas Régimen General)	0

Fuente de información: Ministerio de Educación y Ciencia.

Font: Anuari 2007, INE

Solució

- Denotarem per G_1 = pertànyer al grup 1.
 G_2 = pertànyer al grup 2.
 G_3 = pertànyer al grup 3.

L'espai mostral d'aquest experiment és $\{G_1, G_2, G_3\}$.

Evidentment, els successos elementals no tenen la mateixa probabilitat de donar-se, ja que no hi ha els mateixos alumnes en cada grup. Així doncs:

$$b) P(G_1) = \frac{283.536}{459.291} = 0,6217; P(G_2) = \frac{143.874}{459.291} = 0,3133; P(G_3) = \frac{29.881}{459.291} = 0,065$$

- L'esdeveniment cursar ESO és un esdeveniment elemental de l'experiment «seleccionar aleatòriament un alumne estranger resident en Espanya l'any 04/05 i preguntar-li quin ensenyament està realitzant». I l'esdeveniment pertànyer al grup 2 es pot considerar un esdeveniment compost d'aquest

experiment. De manera que, si denotem per $ESO = \text{ cursar } ESO$, llavors es compleix $ESO \rightarrow G2$ (ESO està inclòs en $G2$).

Exercici 9

En una empresa, en el departament de recursos humans hi ha tres seccions. En la primera secció treballen 2 homes i 3 dones, en la segona secció hi ha 2 dones i 1 home i en la tercera secció treballen 2 homes. L'empresa ha decidit nomenar una persona d'aquelles que formen part del personal de recursos humans per a acomiadar un conjunt de treballadors. Si l'elecció és completament aleatòria, calcula:

- La probabilitat que la persona seleccionada siga un home. Calcula la probabilitat que la persona seleccionada siga una dona.
- La probabilitat que la persona escollida pertanga a cadascuna de les seccions.
- Digues els esdeveniments elementals de l'experiment «secció a què pertany la persona encarregada de comunicar els acomiadaments» i calcula'n les probabilitats.

Solució

- a) Aquest apartat és molt simple:

$$P(\text{ser home}) = \frac{5}{10} = \frac{1}{2} \quad \text{i} \quad P(\text{ser dona}) = \frac{5}{10} = \frac{1}{2}$$

- b) De la mateixa manera que en el cas anterior:

$$P(\text{secció 1}) = \frac{5}{10} = \frac{1}{2} \quad P(\text{secció 2}) = \frac{3}{10} \quad P(\text{secció 3}) = \frac{2}{10} = \frac{1}{5}$$

- c) Els esdeveniments elementals d'aquest experiment, i que per tant formen l'espai mostral, són:

$$E = \{\text{secció 1, secció 2, secció 3}\}$$

Les probabilitats han sigut calculades a l'apartat anterior. Com es pot comprovar, aquest segon experiment no és equiprobable.

Exercici 10

La taula següent mostra els adults segons la seua relació amb l'activitat en 2004.

	Número de personas (miles)
TOTAL ADULTOS	36.187,6
Relación con la actividad	
ACTIVOS	19.067,4
Ocupados	17.139,3
Asalariados	14.259,3
Empresarios	2.880,0
Parados	1.928,0
INACTIVOS	17.051,0
Jubilados	5.477,6
Otros inactivos	11.573,4
NO CONSTA	69,2

Font: Anuari 2007, INE

- Calcula la probabilitat que una persona adulta escollida a l'atzar de les que apareixen en la taula siga empresària.
- Calcula la probabilitat que una persona adulta escollida a l'atzar siga empresària si se sap que està ocupada.
- Calcula la probabilitat que una persona adulta escollida a l'atzar d'entre les persones actives siga empresària.
- Fes un diagrama d'arbre en què el node inicial siguen les persones ocupades.

Solució

- Denominarem E_m = ser empresària. El problema demana la probabilitat que en escollir una persona a l'atzar d'entre els 36.187,6 milers que formen el total d'adults en Espanya l'any 2004, siga empresària.

$$\text{Així, aplicant Laplace: } P(E_m) = \frac{2880,0}{36187,6} = 0,0796$$

- Denominem O = estar ocupada/at. El problema demana la probabilitat d'escollir una persona empresària si se sap que està ocupada. És a dir, cal calcular $P(E_m/O)$.

Hi ha dues maneres de fer aquest càlcul:

1r mètode

Com que hi ha 17.193,3 milers d'ocupats i d'aquests, hi ha 2880,0 milers d'empresaris, llavors $P(\text{Em}/O) = \frac{2880}{17139,3} = 0,167$.

Cal notar que, en aquest cas, si E és l'espai mostral de l'experiment «escollir una persona a l'atzar de les adultes en Espanya l'any 2004», s'ha considerat un nou experiment, «escollir una persona a l'atzar de les adultes i ocupades en Espanya l'any 2004», i un nou espai mostral format per $E \cap O$.

2n mètode

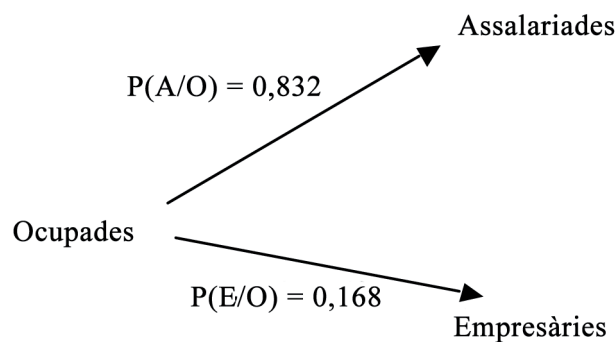
Aquest mètode consisteix a aplicar la definició de probabilitat condicionada:

$$P(\text{Em}/O) = \frac{P(\text{Em} \cap O)}{P(O)} = \frac{P(\text{Em})}{P(O)} = \frac{\frac{2880,0}{36187,6}}{\frac{17139,3}{36187,6}} = \frac{2880,0}{17139,3} = 0,167$$

c) Denominarem A = persona activa, i fem el mateix que en el cas anterior:

$$P(\text{Em}/A) = \frac{2880}{19067,4} = 0,151$$

Fes un diagrama d'arbre en què el node inicial siguin les persones ocupades.



Exercici 11

Tenint en compte les dades mostrades en la taula següent, calcula:

- La probabilitat que una empresa escollida a l'atzar d'entre les de Catalunya i la Comunitat Valenciana siga de Castelló.
- La probabilitat que una empresa escollida a l'atzar siga d'Alacant si se sap que l'empresa és de la Comunitat Valenciana.
- La probabilitat que una empresa escollida a l'atzar siga catalana i de Lleida.
- Si se sap que l'empresa és catalana, quina és la probabilitat que siga de Lleida?

Número de Empresas. 2006	Empleo y paro. 2006			
	Ocupados (miles)	Tasa de actividad (%)	Tasa de paro (%)	
578.340	3.418,7	62,2	6,6	Cataluña
444.410	2.532,2	62,0	6,7	Barcelona
47.169	347,2	66,3	6,6	Girona
33.956	187,6	58,3	6,3	Lleida
52.805	351,8	62,0	6,3	Tarragona
348.692	2.152,3	59,6	8,4	Comunitat Valenciana
133.016	776,4	58,9	9,6	Alicante/Alacant
39.749	259,6	60,0	6,1	Castellón/Castelló
175.927	1.116,3	60,1	8,0	Valencia/València

Font: Anuari 2007, INE

Solució

- Denominem C = ser de Castelló. La probabilitat és molt senzilla de calcular:

$$P(C) = \frac{39749}{972032} = 0,041$$

- Denotem A = ser d'Alacant, i CV = ser de la Comunitat Valenciana.

$$\text{El problema demana } P(A|CV) = \frac{P(A \cap CV)}{P(CV)} = \frac{P(A)}{P(O)} = \frac{\frac{133016}{972032}}{\frac{348692}{972032}} = \frac{133016}{348692} = 0,38$$

- Denotem Ca = ser de Catalunya, i LL = ser de Lleida

$$\text{El problema demana } P(Ca \cap LL) = P(Ca|LL) \cdot P(LL) = 1 \cdot P(LL) = \frac{33956}{972032} = 0,034.$$

Fent el càlcul d'una manera diferent:

$$P(\text{Ca} \cap \text{LL}) = P(\text{LL}/\text{Ca}) \cdot P(\text{Ca}) = \frac{33956}{578340} \frac{578340}{972032} = \frac{33956}{972032} = 0,034$$

Però, com que $\text{LL} \rightarrow \text{Ca}$ ($\text{LL} \cap \text{Ca}$), s'hagués pogut fer el càlcul directament:

$$P(\text{Ca} \cap \text{LL}) = P(\text{LL}) = \frac{33956}{972032} = 0,034$$

d) En aquest cas demana una probabilitat condicionada. Cal fer el càlcul directe emprant l'espai mostral $E \cap \text{Ca}$ (Reviseu l'exercici 26).

$$P(\text{LL}/\text{Ca}) = \frac{33956}{578430} = 0,0587$$

Exercici 12

La taula següent mostra un extracte dels resultats nacionals d'activitat, treball i atur l'any 2006. Les dades vénen donades en percentatges. A més a més, se sap que 17.412 milers de persones formen la població activa, de les quals 14.689 estan ocupades.

Actius per sectors econòmics, 2006

PORCENTAJE	
Agricultura, silvicultura, caza y pesca	4,80
Industria (sin construcción)	15,93
Construcción	12,53
Servicios	63,36
No clasificables	3,39

Ocupats per sectors econòmics, 2006

PORCENTAJE	
Agricultura, silvicultura, caza y pesca	4,78
Industria (sin construcción)	16,67
Construcción	12,88
Servicios	65,67

Aturats per sectors econòmics, 2006

Agricultura, silvicultura, caza y pesca	4,94
Industria (sin construcción)	7,96
Construcción	8,80
Servicios	38,47
No clasificables	39,83

Font: Anuari 2007, INE

S'escull una persona a l'atzar de les persones actives:

- Calcula la probabilitat que una persona activa estiga ocupada i la probabilitat que una persona activa estiga a l'atur.
- Calcula la probabilitat que una persona es dedique al sector de la construcció si se sap que forma part de la població aturada.
- Calcula la probabilitat que una persona activa estiga aturada i que es dedique al sector de la construcció.
- Quina relació i quines diferències hi ha entre els dos apartats anteriors?

Solució

- La població activa està formada per 17.412 milers de persones, de les quals n'hi ha 14.689 milers ocupades i 2.723 milers a l'atur. En conseqüència:

$$P(\text{estar ocupat}) = \frac{14689}{17412} = 0,844 \text{ i } P(\text{estar a l'atur}) = \frac{2723}{17412} = 0,156$$

- L'exercici demana una probabilitat condicionada. Així la informació que es coneix és que la persona està aturada. Llavors, es busca:

$$P(\text{dedicar-se sector construcció} / \text{la persona està aturada}) = 0,088$$

Cal notar que en aquest cas es consideren únicament les persones que estan aturades i, d'aquestes, es busquen aquelles que es dediquen al sector de la construcció.

- Es demana la probabilitat següent:

$$P(\text{la persona estiga aturada} \cap \text{es dedique al sector de la construcció}) =$$

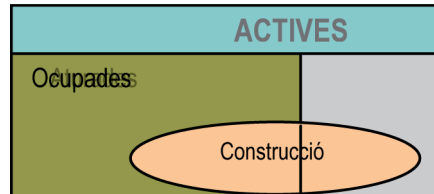
$$P(\text{es dedique al sector de la construcció} / \text{la persona estiga aturada}) \cdot P(\text{la persona estiga aturada}) = 0,088 \cdot 0,156 = 0,0137$$


La igualtat és conseqüència de la definició de probabilitat condicionada.


- A l'apartat *b* la persona és escollida únicament d'aquelles que estan a l'atur (de color gris) i es calcula la probabilitat que es dedique a la construcció

(zona rosa de dins la secció d'el·lipse que està dins de les persones aturades), mentre que en l'apartat c la persona s'escull d'aquelles que estan actives (tot el rectangle) i es calcula la probabilitat que estiga aturat i es dedique a la construcció (zona rosa de dins la secció d'el·lipse que està dins de les persones aturades).

Gràficament:



En el primer cas b :  _____

En el segon cas c :  _____

Exercici 13

Una urna conté 10 boles verdes, 5 boles roges i 8 boles negres. S'extreuen 3 boles aleatòriament de l'urna. Calcula la probabilitat d'obtenir la seqüència verda, roja, negra en els següents casos:

- La bola extreta no es retorna a l'urna.
- La bola extreta es retorna a l'urna.

Solució

- Definim $V1$ = primera bola extreta és verda.
 $R2$ = segona bola extreta és roja.
 $N3$ = tercera bola extreta és negra.

El problema demana $P(V1 \cap R2 \cap N3)$.

Aplicant la definició de probabilitat condicionada, aquesta probabilitat es calcula:

$$P(V1 \cap R2 \cap N3) = P(N3/V1 \cap R2) \cdot P(V1 \cap R2) = \\ P(N3/V1 \cap R2) \cdot P(R2/V1) \cdot P(V1) =$$

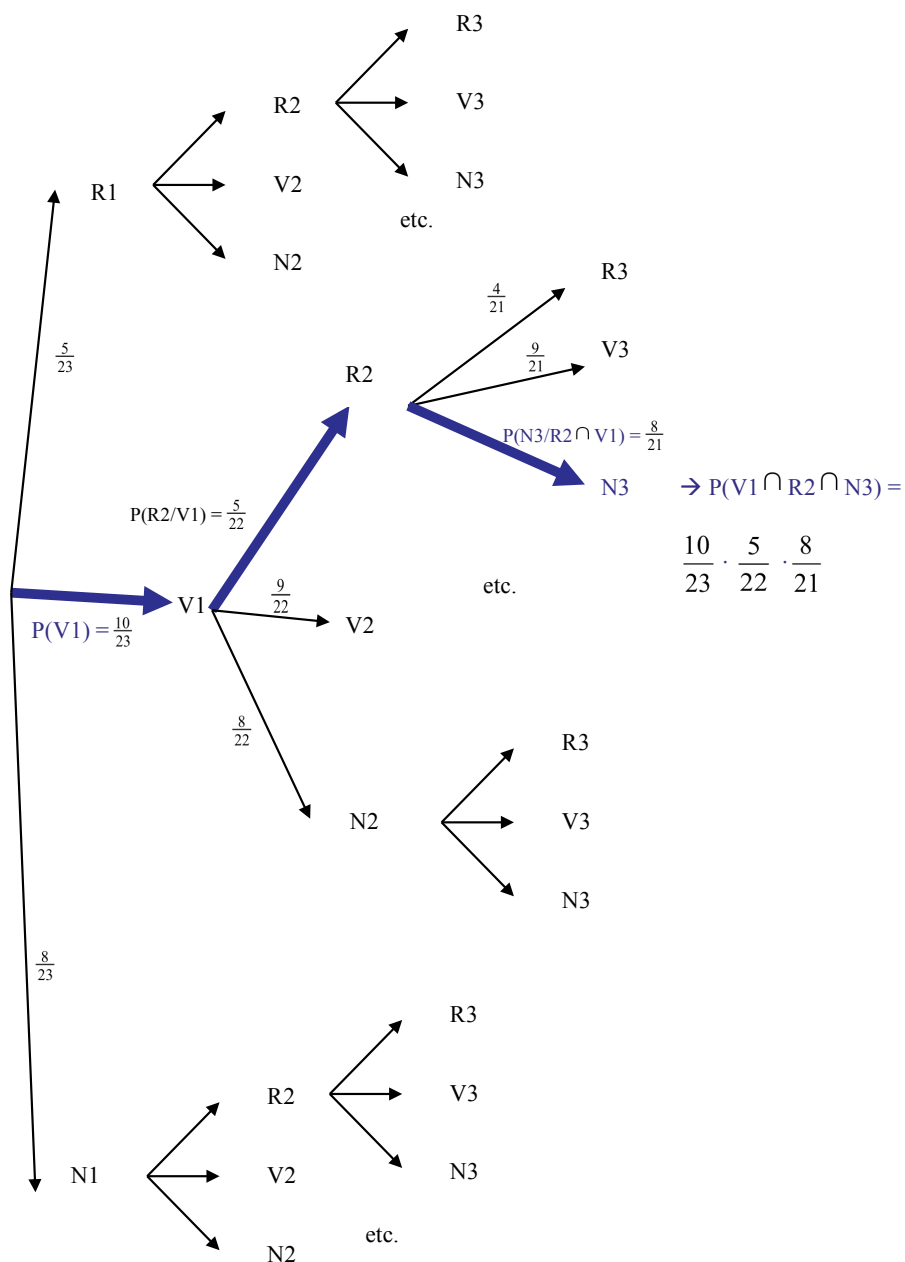
$$1. P(V1) = \frac{10}{23}, \text{ ja que en l'urna hi ha 23 boles, de les quals 10 són verdes.}$$

2. $P(R2/V1) = \frac{5}{22}$, ja que en l'urna queden 22 boles –se n'ha extret una verda–, de les quals 5 són roges.

3. $P(N3/V1 \cap R2) = \frac{8}{21}$, ja que en l'urna queden 21 boles –se n'han extret una verda i una roja–, de les quals 8 són negres.

Això fa: $\frac{10}{23} \cdot \frac{5}{22} \cdot \frac{8}{21} = 0,0376$.

Aquest exercici es podria fer també en un diagrama d'arbre, el qual aclareix notablement la informació:



El diagrama mostra el «camí» que permet obtenir la probabilitat demanada emprant el diagrama d'arbre.

- b) L'única diferència d'aquest exercici respecte de l'anterior és que la bola extreta es retorna a l'urna. Aquest fet implica que cada vegada que s'extreu una bola la composició de l'urna no varia. Per tant:

$$P(V1 \cap R2 \cap N3) = P(N3/V1 \cap R2) \cdot P(V1 \cap R2) = \\ P(N3/V1 \cap R2) \cdot P(R2/V1) \cdot P(V1) =$$

1. $P(V1) = \frac{10}{23}$, ja que en l'urna hi ha 23 boles, de les quals 10 són Verdes.
2. $P(R2/V1) = \frac{5}{23}$, ja que en l'urna queden 23 boles, de les quals 5 són roges.
3. $P(N3/V1 \cap R2) = \frac{8}{23}$, ja que en l'urna queden 23 boles, de les quals 8 són negres.

$$\frac{10}{23} \cdot \frac{5}{23} \cdot \frac{8}{23} = 0,0328.$$

El diagrama d'arbre que sortiria és semblant a l'anterior, però canviarien les probabilitats.

Exercici 14

La taula següent mostra un extracte de la taula d'empreses segons la seua condició jurídica per estrats d'assalariats l'any 2006.

	Condición jurídica			Personas físicas
	Total	Sociedades anónimas	Sociedades de responsabilidad limitada	
TOTAL	3.174.393	117.441	992.658	1.791.126
Sin asalariados	1.616.883	25.286	267.857	1.202.750
De 1 a 2 asalariados	881.748	19.908	328.565	448.155
De 3 a 5 asalariados	328.820	13.592	175.633	105.500
De 6 a 9 asalariados	154.635	12.880	101.321	25.635
De 10 a 19 asalariados	105.470	16.958	72.051	7.591
De 20 a 49 asalariados	58.725	16.425	35.485	1.495
De 50 a 99 asalariados	15.087	6.059	7.005	0
De 100 a 199 asalariados	7.357	3.325	3.023	0
De 200 a 499 asalariados	3.906	1.994	1.342	0
De 500 a 999 asalariados	999	574	253	0
De 1.000 a 4.999 asalariados	658	372	118	0
De 5.000 o más asalariados	93	68	5	0

Font: Anuari 2007, INE

- a) Calcula la probabilitat d'escollir dues empreses a l'atzar i que una siga una societat anònima i l'altra siga una persona física.
- b) Calcula la probabilitat que, en escollir dues empreses a l'atzar d'aquelles que tenen més de 5.000 assalariats, una siga SA i l'altra, SL.

Solució

- a) Denotem SA_i = la i-èsima empresa escollida és una SA.
 PF_i = la i-èsima empresa escollida és una persona física.

Es demana la $P(\text{una empresa siga SA} \cap \text{una altra empresa siga PF}) =$
 $P(SA_1 \cap PF_2 \cup PF_1 \cap SA_2) = P(SA_1 \cap PF_2) + P(PF_1 \cap SA_2) =$

Els esdeveniments són tots disjunts, és a dir, $SA_1 \cap PF_2$ és disjunt amb $SA_2 \cap PF_1$.

D'altra banda, com que el nombre total d'empreses d'entre les que se n'escolliran dos és molt gran (3.174.393), llavors es pot considerar que les empreses escollides en una selecció no es tenen en compte en les següents seleccions. Per tant, $P(SA_1 \cap PF_2) = P(PF_1 \cap SA_2) = P(PF_1) \cdot P(SA_1) =$

$$= 2 \cdot \frac{117441}{3174393} \cdot \frac{1791126}{3174393} = 2 \cdot 0,03699 \cdot 0,56424 = 0,0417$$

- b) Denotem SA_i = la i-èsima empresa escollida és una SA.
 SL_i = la i-èsima empresa escollida és una persona física.

Es demana la $P(\text{una empresa siga SA} \cap \text{una altra empresa siga SL}) =$

$P(SA_1 \cap SL_2 \cup SL_1 \cap SA_2) = P(SA_1 \cap SL_2) + P(SL_1 \cap SA_2) =$

Els esdeveniments són tots disjunts, és a dir $SA_1 \cap SL_2$ és disjunt amb $SA_2 \cap SL_1$.

D'altra banda, en aquest cas el nombre total d'empreses que formen la població de l'exercici, empreses d'entre les quals se n'escolliran dos, és petit (93), així és que no es pot suposar el mateix que a l'apartat anterior, és a dir, no es pot suposar que la independència dels escolliments. Així:

$$P(SA_1 \cap SL_2) = P(SL_2/SA_1) \cdot P(SA_1) = \frac{5}{92} \cdot \frac{68}{93} = \frac{85}{2139}$$

$$P(SL_1 \cap SA_2) = P(SA_2/SL_1) \cdot P(SL_1) = \frac{68}{92} \cdot \frac{5}{93} = \frac{85}{2139}$$

$$\text{Això fa: } \frac{85}{2139} + \frac{85}{2139} = \frac{10}{129} = 0,0775.$$

NOTA

Si en aquest cas s'hagués suposat la independència, aleshores:

$$P(SA1 \cap SL2 \cup SL1 \cap SA2) = 2 \cdot P(SA1) \cdot P(SL1) = \frac{5}{93} \cdot \frac{68}{93} = 0,0786218.$$

L'error seria inferior a 0,01, no seria molta la diferència.

Exercici 15

En una ciutat castellanenca hi ha un 45 % de ciutadans de Castelló, un 35 % de ciutadans espanyols però que no són de Castelló i la resta són de la Unió Europea. Se sap que un 20 % dels ciutadans de Castelló tenen accions en borsa i un 15 % dels ciutadans espanyols també les tenen, mentre que únicament un 4 % dels ciutadans de la Unió Europea també tenen accions.

Calcula molt acuradament la probabilitat que en escollir una persona a l'atzar de les que viuen en l'esmentada ciutat, tinga accions en borsa. Justifica l'ús de les fórmules que utilitzes.

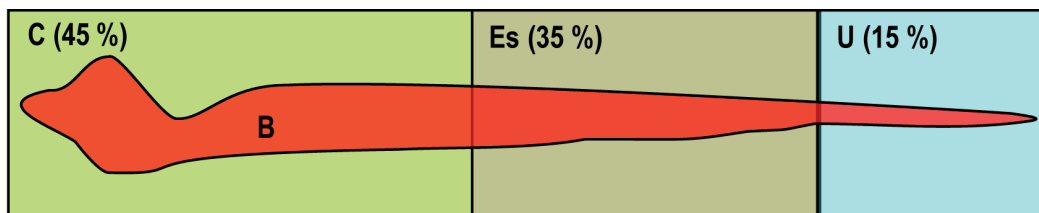
Denotem per C = la persona escollida és de Castelló.

Es = la persona escollida és espanyola, però no de Castelló.

U = la persona escollida és de la UE.

B = la persona escollida té accions en borsa.

El problema demana la probabilitat que una persona, en escollir-la a l'atzar d'entre les que viuen en la ciutat, tinga accions en borsa. És evident, no obstant, que a l'hora de realitzar el càlcul probabilístic cal tenir en compte d'on és cada persona, ja que no és el mateix que la persona escollida siga de Castelló que de la UE o espanyola. El següent dibuix esquemàtic ho manifesta gràficament:



La diferent amplària de la representació de l'esdeveniment B correspon al fet que hi ha un 20 % de les persones de Castelló que tenen accions, un 15 % dels espanyols que tenen accions i tan sols un 4 % dels de la UE que en tenen. És a dir:

$$P(B/C) = 0,2$$

$$P(B/Es) = 0,15$$

$$P(B/U) = 0,04$$

Una vegada analitzat amb cura l'enunciat del problema, és evident que per a calcular el que es demana cal aplicar el teorema de la probabilitat total, ja que, com s'observa al gràfic, la probabilitat del succés B es pot obtenir sumant les probabilitats $B \cap C$, $B \cap Es$ i $B \cap U$.

Cal recordar que per aplicar el teorema de la probabilitat total s'han de conèixer les probabilitats condicionades d'un esdeveniment vers una partició de l'espai mostral i les probabilitats de cada esdeveniment que forma la partició.

En el nostre cas, l'espai mostral E està format per tots els ciutadans de la ciutat Castelló i la partició està formada pels esdeveniments C, Es i U, llavors:

- $E = C \cup Es \cup U$
- $Es \cap C = \emptyset, U \cap Es = \emptyset; U \cap C = \emptyset$

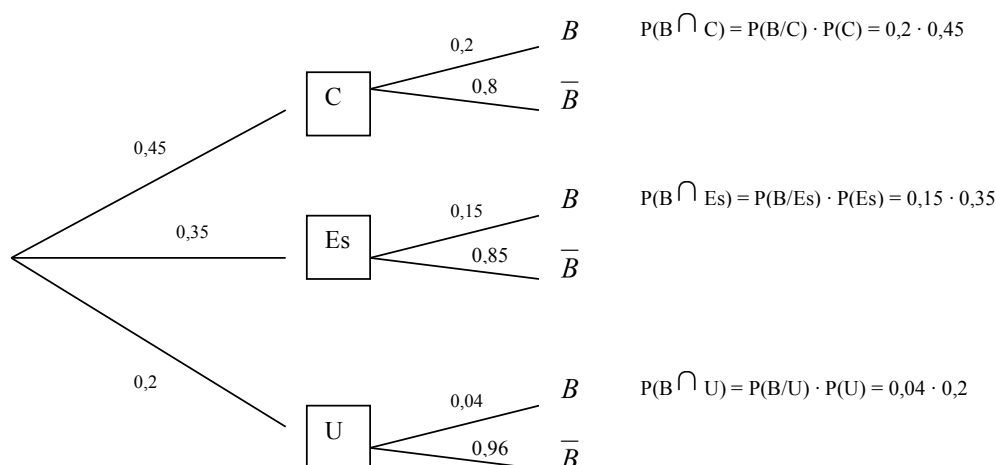
Per tant, aplicant el teorema:

$$P(B) = P(B \cap C \cup B \cap Es \cup B \cap U) = P(B \cap C) + P(B \cap Es) + P(B \cap U) = P(B/C) \cdot P(C) + P(B/Es) \cdot P(Es) + P(B/U) \cdot P(U) = 0,2 \cdot 0,45 + 0,15 \cdot 0,35 + 0,2 \cdot 0,04 = 0,1505.$$

NOTA

Aquest problema, una vegada s'ha observat la necessitat d'emprar el teorema de la probabilitat total –s'observa una possible partició de l'espai mostral i es coneixen les probabilitats condicionades vers aquesta partició–, també es pot raonar emprant un arbre. De fet, sol ser la tècnica més habitual a l'hora de fer aquest tipus d'exercicis.

Així:



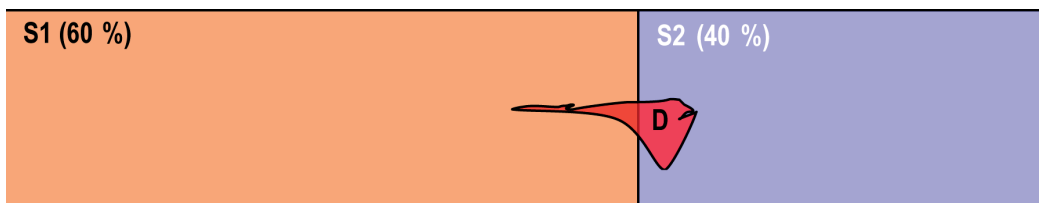
Exercici 16

Una empresa té dues seccions de fabricació del seu producte. La secció 1 produeix el 60 % dels productes que exporta l'empresa a l'estranger. El 40 % restant el produeix la secció 2. D'altra banda, l'empresa coneix que la secció 1 elabora un 4 % de productes defectuosos i la secció 2 n'elabora 8 %. Un client estranger ha emés una queixa perquè el producte que ha rebut és defectuós. Calcula la probabilitat que haja estat produït per la secció 1.

En aquest exercici es parteix d'un esdeveniment que ja ha passat: un producte dels exportats a l'estranger ha resultat ser defectuós. Es demana la probabilitat que aquest producte haja estat fabricat per la secció 1. És evident que la secció 1 fabrica més productes que la 2, però també és cert que en produeix menys de defectuosos.

Denotem per $S1$ = el producte prové de la secció 1.
 $S2$ = el producte prové de la secció 2.
 D = el producte és defectuós.

Amb aquesta notació, el gràfic següent resumeix les característiques més rellevants de l'enunciat:



A més a més que es dedueix que $P(D/S1) = 0,04$ i que $P(D/S2) = 0,08$.

L'enunciat demana la probabilitat de que, si s'ha donat un producte defectuós, aquest s'haja produït a la secció 1. Intuïtivament, cal fer el següent quocient:

$$\frac{\text{[taronja]} \text{ i } \text{[vermell]}}{\text{[vermell]}} \text{ i formalment es busca } P(S1/D) = \frac{P(S1 \cap D)}{P(D)}.$$

És a dir, el problema és una clara aplicació del teorema de Bayes. A més a més, es coneix una partició de l'espai mostral $E = \{\text{peces fabricades per l'empresa que són exportades a l'estranger}\}$ formada per $S1$ i $S2$:

- $E = S1 \cup S2$
- $S1 \cap S2 = \emptyset$

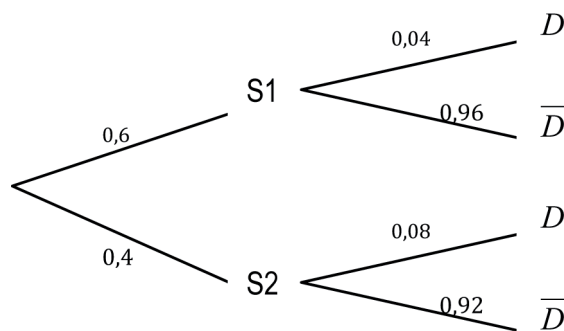
Per tant, aplicant el teorema de Bayes:

$$P(S1/D) = \frac{P(S1 \cap D)}{P(D)} = \frac{P(D/S1) \cdot P(S1)}{P(S1 \cap D) + P(S2 \cap D)} = \frac{P(D/S1) \cdot P(S1)}{P(D/S1) \cdot P(S1) + P(D/S2) \cdot P(S2)}$$

$$= \frac{0,04 \cdot 0,6}{0,04 \cdot 0,6 + 0,08 \cdot 0,4} = 0,4286$$

NOTA

Per a calcular la $P(D)$ s'ha aplicat el teorema de la probabilitat total. Cal dir que en aquest cas també és possible realitzar el diagrama d'arbre del problema, el qual facilita la comprensió de l'enunciat.



Exercici 17

La taula següent mostra dades referents a la Comunitat Valenciana.

	Población a 1 de enero de 2006		
	Total	Mujeres (%)	Extranjeros (%)
Comunitat Valenciana	4.806.908		
Alicante/Alacant	1.783.555	49,9	20,1
Castellón/Castelló	559.761	49,7	13,9
Valencia/València	2.463.592	50,5	9,4

Font: INE

- Calcula la probabilitat que en escollir una persona a l'atzar de la Comunitat Valenciana siga dona.
- S'ha escollit una persona a l'atzar de la Comunitat Valenciana i ha resultat ser home. Calcula la probabilitat que siga de Castelló.

Solució

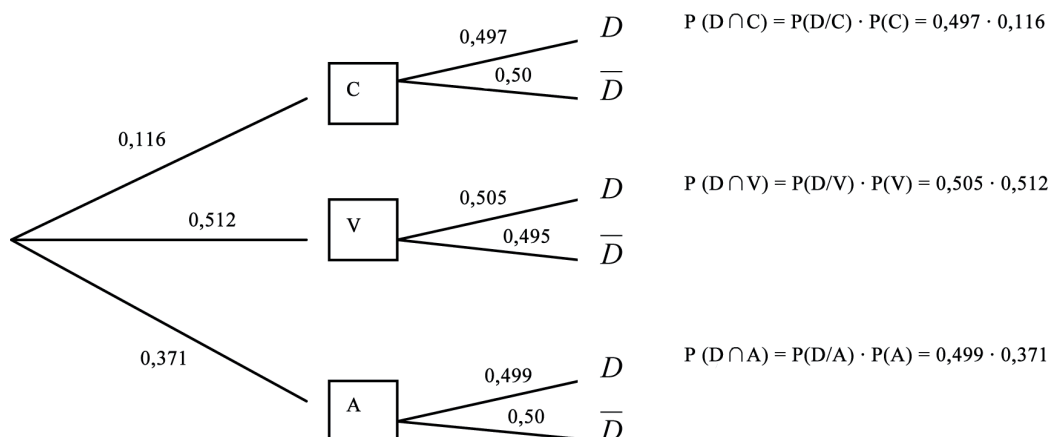
- a) Denotem per C = la persona escollida és de Castelló.
 A = la persona escollida és de València.
 V = la persona escollida és d'Alacant.
 D = la persona escollida és dona.

De la taula es poden conèixer les probabilitats dels tres esdeveniments primers:

$$P(C) = \frac{559761}{4806908} = 0,1164 \quad P(V) = \frac{2463592}{4806908} = 0,5125 \quad P(A) = \frac{1783555}{4806908} = 0,3711$$

L'enunciat demana $P(D)$, que directament no es pot calcular. Però, com que es coneixen les probabilitats condicionades $P(D/A) = 0,499$, $P(D/C) = 0,497$ i $P(D/V) = 0,505$, i els esdeveniments $\{V, A, C\}$ són una partició de l'espai mostral $E = \{\text{escollir una persona a l'atzar de la CV}\}$, llavors es pot aplicar el teorema de la probabilitat total.

Fem en primer lloc l'arbre per a aclarir l'enunciat:



D'aquí s'obté:

$$P(D) = P(D/C) \cdot P(C) + P(D/A) \cdot P(A) + P(D/V) \cdot P(V) = 0,497 \cdot 0,116 + 0,505 \cdot 0,512 + 0,499 \cdot 0,371 = 0,501$$

- b) En aquest cas se sap que s'ha escollit un home, aquest esdeveniment ha succeït. Es demana la probabilitat que aquest siga de Castelló. Formalment, es demana la $P(C/\bar{D})$.

Aplicant la definició de probabilitat condicionada: $P(C/\bar{D}) = \frac{P(C \cap \bar{D})}{P(\bar{D})}$.

Però pel que s'ha comentat a l'apartat anterior es pot aplicar el teorema de Bayes. Així:

$$P(C/\bar{D}) = \frac{P(C \cap \bar{D})}{P(\bar{D})} = \frac{P(\bar{D}/C) \cdot P(C)}{P(\bar{D})} = \frac{0,503 \cdot 0,116}{0,499} = 0,1169$$

D'on $P(\bar{D}) = 1 - P(D) = 1 - 0,501 = 0,499$.

UNITAT 2

Variables aleatòries discretes. Models de probabilitat discrets

Introducció teòrica

En la unitat anterior s'han tractat diverses qüestions relatives a la probabilitat. Així, donat un experiment aleatori, es determinaven les probabilitats de l'ocurrència de cadascun dels esdeveniments. Per a fer-ne els càlculs, calia conèixer les característiques bàsiques de l'experiment i aplicar els axiomes i les propietats de la probabilitat.

El coneixement d'estudis probabilístics pel que fa a experiments semblants als que s'expliquen, permetia intuir els càlculs de les probabilitats de l'experiment en qüestió. Per exemple, els experiments «llançar una moneda equilibrada i observar el resultat i extreure una bola d'una urna amb tres boles, dues verdes i una roja, i observar-ne el color de la bola extreta» són molt semblants, però el que s'ha estudiat al capítol anterior tan sols es poden utilitzar els càlculs de probabilitats del primer com un exemple per a realitzar els càlculs en el segon.

D'altra banda, és evident que els dos experiments tenen característiques molt semblants des del punt de vista de la probabilitat:

- Són aleatoris.
- Tenen dos resultats possibles (C,+) i (Verd, Roig)
- Les probabilitats de cada succés elemental són $P(\text{succés 1}) = p$, i $P(\text{succés 2}) = 1 - p$

$$P(C) = \frac{1}{2}, \text{ i } P(+) = 1 - \frac{1}{2} = \frac{1}{2}; P(\text{verd}) = \frac{2}{3}, \text{ i } P(\text{roig}) = 1 - \frac{2}{3} = \frac{1}{3}.$$

Així, els dos experiments segueixen un mateix «patró» en la distribució de probabilitats vers els successos elementals –coneguda la probabilitat de l'esdeveniment elemental, la probabilitat de l'altre és 1 menys la probabilitat del primer–, és a dir, tots dos segueixen el mateix model de probabilitat. Cal dir que existeixen molts experiments que no segueixen aquest patró, com per exemple llançar un dau equilibrat i observar-ne el resultat.

Al llarg del capítol s'estudiaran conceptes com la variable aleatòria, la funció de probabilitat, la funció de quantia, la funció de distribució, la distribució de probabilitat, etc., els quals permeten assignar cada experiment a un model de probabilitat. També es presentaran els principals models de probabilitat: el de Bernouilli, el binomial, el de Poisson, l'uniforme, l'exponencial i el normal. Per a finalitzar s'introduirà el teorema central del límit com una de les aplicacions més importants de la distribució normal.

De l'experiment al model

Hi ha nombrosos experiments aleatoris que tenen característiques intrínseques comunes des del punt de vista de la probabilitat, malgrat que els seus resultats i posteriors conseqüències siguin diferents.

Es presenten dos experiments:

Exp. 1: llançar 15 vegades una moneda equilibrada i calcular la probabilitat d'obtindre 8 cares en els 15 llançaments.

Exp. 2: triar a l'atzar 8 persones d'una mostra de 15 persones i calcular la probabilitat que les 8 siguin consumidores d'un determinat producte.

S'observa que en ambdós casos es realitzen 15 experiments consecutius i independents els uns dels altres. El resultat de cadascun d'aquests 15 experiments té dues úniques possibles opcions: en el primer, «Cara» o «Creu» i en el segon, «ser consumidor del producte» o «no ser consumidor el producte». A més, de la mateixa manera que la probabilitat d'obtindre cara en el primer experiment és 0,5 (la moneda és equilibrada), en el segon experiment és segur que ha d'existir una probabilitat p que una persona triada a l'atzar de la mostra siga consumidora del producte.

Per tant, es pot dir que ambdós experiments segueixen un mateix patró, ambdós poden *modelar-se per la mateixa distribució de probabilitat*. Explicarem el significat d'açò últim per mitjà d'un exemple:

Variabls aleatòries. Estudi de la seua distribució

Allò que es pretén amb l'ús de variables aleatòries és transformar l'estudi dels experiments per facilitar-ne la comprensió i la recerca del model de probabilitat que segueix. Així, bàsicament, les variables aleatòries són funcions o relacions que assignen un nombre real a cada esdeveniment elemental d'un experiment aleatori. Aquest fet provoca una equivalència entre l'espai mostral i el conjunt numèric producte de les assignacions.

Així doncs, s'anomena *variable aleatòria* aquella funció que assigna un nombre a cada esdeveniment elemental de l'espai mostral. És a dir, si E és l'espai mostral, X és una variable aleatòria si és una funció que associa a cada esdeveniment elemental s de E un nombre $X(s)$. Al conjunt format per tots els possibles valors que pren la variable aleatòria se l'anomena *rang* (R) de la variable aleatòria.

D'altra banda, donada una variable aleatòria X i un subconjunt R_x del rang de X , la probabilitat de R_x ve donada per la probabilitat que es done el conjunt format pels esdeveniments elementals A , els quals tenen per imatge de X , R_x , és a dir, $X(A) = R_x$. Els següents exemples aclareixen aquestes qüestions.

Tipus de variables aleatòries

De la mateixa manera que ocorria en les variables estadístiques estudiades en el bloc referent a estadística descriptiva, les variables aleatòries es poden classificar en discretes o contínues.

Les variables aleatòries discretes poden assumir tan sols certs valors, amb freqüència nombres enters, i resulten fonamentalment del recompte. La variable «resultat obtingut en llançar un dau» és discreta i els possibles valors que pot prendre són $\{1, 2, 3, 4, 5 \text{ i } 6\}$. També el nombre de cridades que rep una família en una hora, el nombre de vaixells que arriben a un port o el nombre de clients que tenen les empreses ceràmiques són exemples de variables aleatòries discretes.

Variabls aleatòries discretes

Com s'ha definit amb anterioritat, una variable aleatòria és discreta si únicament pot prendre un número finit de valors reals o, com a màxim, un conjunt numerable. Per tant, en aquest tipus de distribucions té sentit calcular la probabilitat de que la variable aleatòria prenga un valor concret. Intuïtivament pareix clar, perquè en l'assignació de probabilitats s'ha de repartir 1 entre un conjunt finit o numerable de valors.

Funció de probabilitat

En les variables aleatòries discretes es defineix una funció, denominada *funció de probabilitat*, que associa a cada valor de la variable aleatòria la seua probabilitat. Per una altra part, com que se sol representar per X (en majúscules) la variable aleatòria i per x (en minúscules) el valor que pren, es pot definir la funció de probabilitat amb aquesta notació: $f(x) = P(X = x)$.

NOTA

Cal notar que les distribucions de probabilitat discretes es representen de la mateixa manera que les variables estadístiques en què les dades no s'agrupen. En tots dos casos s'empra un diagrama de barres. En les variables estadístiques es representa la freqüència relativa (o l'absoluta) i en les variables aleatòries discretes, la probabilitat.

Funció de distribució

La funció de distribució d'una variable aleatòria té el mateix significat que la freqüència relativa acumulada de les distribucions estadístiques descriptives. Així, donada una distribució estadística discreta X en què a és un valor numèric, es defineix la funció de distribució $F(a)$ com la probabilitat que la variable aleatòria X no prenga un valor superior a a . És a dir:

$$F(a) = P(X \leq a) = \sum_{x \leq a} P(x) = \sum_{x \leq a} f(x)$$

Propietats

- La funció de distribució es caracteritza per ser creixent, és a dir: si $x_1 \geq x_2 \rightarrow F(x_1) \geq F(x_2)$.
- $F(-\infty) = P(X \leq -\infty) = 0$, ja que cap valor de la variable aleatòria pot ser menor que menys infinit.
- $F(+\infty) = P(X \leq +\infty) = 1$, ja que tots els valors de la variable aleatòria són menors que més infinit.
- $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \sum_{x > a}^b f(x)$.

Per exemple, si una variable aleatòria pren els valors $\{1, 2, 3, 5\}$, llavors $P(2 < X \leq 5) = F(5) - F(2) = P(X = 5) + P(X = 4) + P(X = 3)$.

Esperança i variància d'una variable aleatòria discreta

De la mateixa manera que ocorria amb les variables estadístiques, en què es definien estadístics per a caracteritzar i, en certa manera, resumir la informació de les dades, en les variables aleatòries també és possible definir estadístics amb el mateix propòsit. En aquest cas sintetitzen la informació de la distribució de probabilitat de les variables aleatòries que representen.

Valor esperat o esperança matemàtica

Donada X una variable aleatòria, l'esperança matemàtica pot considerar-se com el valor esperat al realitzar l'experiment, el valor central de la distribució de probabilitats. Es calcula multiplicant cada valor de la variable aleatòria per la seua probabilitat.

Així, si X és una variable aleatòria discreta que pren els valors $\{x_1, x_2, \dots, x_n\}$ i $f(x)$ és la funció de probabilitat associada, l'esperança de X :

$$\mu = E(X) = x_1 \cdot f(x_1) + x_2 \cdot f(x_2) + x_3 \cdot f(x_3) + \dots + x_n \cdot f(x_n) = \sum_{i=1}^n x_i \cdot f(x_i)$$

De la definició del valor esperat també se'n pot deduir una altra definició equivalent. Així, l'esperança d'una variable aleatòria discreta és la mitjana ponderada de tots els possibles resultats on els pesos són les probabilitats respectives dels resultats.

Variància

Donada X una variable aleatòria, en termes de probabilitat la variància pot considerar-se com una mesura de la dispersió dels valors respecte del valor esperat. És a dir, conceptualment és el mateix que el que es calcula en l'estadística descriptiva. Es calcula cercant la mitjana aritmètica del quadrat de les desviacions respecte de l'esperança.

Així, si X és una variable aleatòria discreta que pren els valors $\{x_1, x_2, \dots, x_n\}$ i $f(x)$ és la funció de probabilitat, llavors:

$$\begin{aligned} \text{Var}(X) \\ = \sigma^2 = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + \dots + (x_n - \mu)^2 f(x_n) = \sum_{i=1}^n (x_i - \mu)^2 f(x_i) \end{aligned}$$

NOTA

Es pot demostrar matemàticament que $\text{Var}(X) = E(X^2) - E(X)^2$.

A més a més, de la mateixa manera que en les variables estadístiques, es calcula la desviació típica de la distribució de probabilitat com $\sigma = \sqrt{\sigma^2}$.

Distribució conjunta de dues variables aleatòries

És d'interès l'estudi conjunt de dues variables aleatòries que introduïrem en aquest apartat. Per exemple, pot ser de molta utilitat calcular la probabilitat que es venguen quatre mil productes que fabrica una determinada empresa a un preu de 25 € la unitat o conèixer la probabilitat que un treballador d'un país té d'arribar a guanyar un sou determinat si domina quatre idiomes.

Així, si es considera un experiment aleatori i a cada esdeveniment elemental se li associen dos valors numèrics (x, y) el que s'obté és una *variable aleatòria Bidimensional* (X, Y) . És evident que aquesta associació ha de tenir sentit, és a dir, no és coherent associar la mateixa característica al primer valor (x) que al segon (y) ,

ja que en aquest cas únicament es necessitaria una variable aleatòria. En conseqüència, els valors que prenen les x i les y representen aspectes diferents, i llavors es pot parlar de dues variables aleatòries, X i Y .

De la mateixa manera que per a una variable aleatòria, la probabilitat de cada parell (x, y) ve determinada per la probabilitat de l'esdeveniment que representa. És a dir $P(x, y) = P(X = x, Y = y)$.

En el concepte de variable aleatòria bidimensional s'ha considerat que ambdues variables que la conformen són discretes. No obstant, aquest fet no sempre és així, poden ser contínues, tal com apareixerà en apartats posteriors.

D'altra banda, com ja s'ha comentat, cada una d'aquestes variables té entitat per ella mateixa i per tant es pot fer l'estudi de la variable aleatòria de la mateixa manera que s'ha vist per a les variables aleatòries unidimensionals. A aquests tipus de variables se les anomena variables aleatòries marginals, per analogia amb les variables estadístiques.

En algunes ocasions és possible trobar variables aleatòries bidimensionals en què una component siga contínua i l'altra, discreta; per exemple, si (X, Y) representa l'edat de les persones i el nombre d'empreses on ha treballat. Cal dir, però, que en aquest text no es tractaran aquest tipus de variables.

Per una altra part, tal com ocorre amb les variables aleatòries unidimensionals, si els resultats de l'experiment són numèrics, llavors les associacions entre nombres i esdeveniments elementals són directes.

Binomial discreta

Es considera (X, Y) una variable aleatòria bidimensional discreta. A cada possible resultat (x_i, y_j) se li associa un nombre $p(x_i, y_j)$ que representa $P(X = x_i, Y = y_j)$ de manera que satisfaga els axiomes de probabilitat (d'una manera semblant a les variables unidimensionals):

- $p(x_i, y_j) \geq 0$ per a tots els parells (x_i, y_j)
- $\sum_{x_i} \sum_{y_j} p(x_i, y_j) = 1$

A la funció p definida per a tots els valors (x_i, y_j) del recorregut de (X, Y) se l'anomena funció de probabilitat de (X, Y) i determina la distribució de probabilitat de la variable. A més a més, per a calcular la probabilitat d'un subconjunt A del recorregut, cal sumar les probabilitats de cadascun dels valors que el formen:

$$P(A) = \sum_{(x_i, y_j) \in A} p(x_i, y_j)$$

Cada probabilitat pot representar-se mitjançant uns eixos tridimensionals on la base la constitueixen els valors de les variables X i Y i l'altura, la funció de probabilitat p. Cal remarcar que aquest gràfic és molt semblant al que es va comentar en el capítol cinquè referent a les variables estadístiques bidimensionals.

Aquesta definició es pot generalitzar per a un nombre indeterminat de variables aleatòries. Així, es pot parlar de variables aleatòries n-dimensionals (X_1, \dots, X_n) i de la seua funció de probabilitat conjunta $p(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$, la qual també complix els axiomes de probabilitat:

- $p(x_1, \dots, x_n) \geq 0$ per a tots els parells (x_1, \dots, x_n) de R de (X_1, \dots, X_n)
- $\sum_{x_1} \dots \sum_{x_n} p(x_1, \dots, x_n) = 1$

Exemple

Una empresa està intentant introduir un producte en Noia, una ciutat ubicada en la província de la Corunya. Per fer-ho, ha encarregat a una empresa que estudie com es relaciona el nombre de fills menors de 10 anys que tenen les parelles amb la predisposició a comprar un determinat producte. S'ha arribat als següents resultats:

- El 10 % de les parelles tenen 0 fills i estan disposades a comprar el producte. El 25 % de les parelles tenen 0 fills i no estan disposades a fer-ho.
- El 20 % de les parelles tenen 1 fill i estan disposades a comprar el producte. El 25 % de les parelles tenen 1 fill i no estan disposades a fer-ho.
- El 8 % de les parelles tenen 2 fills i estan disposades a comprar el producte. El 9 % de les parelles tenen 2 fills i no estan disposades a fer-ho.
- El 2 % de les parelles tenen 3 fills i estan disposades a fer-ho. L'1 % de les parelles tenen 3 fills i no estan disposades a comprar el producte.

Si seleccionem una parella a l'atzar de la ciutat, quina és la probabilitat de que tinga dos fills i estiga disposada a comprar el producte?

Per a contestar a la pregunta, en primer lloc es definirà la variable aleatòria conjunta i la seua distribució de probabilitat.

Així, l'espai mostral d'aquest experiment està format per totes les parelles del tipus $(x, \text{sí})$ o (x, no) , on x representa el nombre de fills (0, 1, 2 i 3) que tenen les parelles. És clar que la primera component representa una variable aleatòria $X = \text{nombre de fills de la parella}$, i la segona component indica si la família està interessada en la compra del producte. Com que aquesta segona component no és una variable aleatòria, cal associar nombres a cada possible valor. Així, es pot associar el nombre 1 a «estar interessat en comprar el producte» i 0 a «no estar-ne interessat», llavors s'obté la variable aleatòria bidimensional (X, Y) , la qual té per rang $\{(0, 1); (1, 1); (2, 1); (3, 1); (0, 0); (1, 0); (2, 0); (3, 0)\}$.

La distribució de probabilitats, tenint en compte els percentatges que apareixen abans, és la que apareix a la taula següent:

		X = nombre de fills per família			
		0	1	2	3
Y = decisió de compra	0	0,25	0,25	0,09	0,01
	1	0,1	0,2	0,08	0,02

Taula 1

Per una altra banda, el que es demana és la probabilitat de que tinga dos fills i estiga disposada a comprar el producte, la qual cosa és equivalent a:

$$P(2, 1) = P(X = 2, Y = 1) = 0,08.$$

Models de probabilitat discrets: les distribucions Bernoulli, binomial, hipergeomètrica i Poisson

En els epígrafs anteriors s'han considerat les característiques més importants de les distribucions de probabilitat discretes des d'un punt de vista general. En aquest epígraf es tractaran alguns dels models de probabilitat discreta unidimensional més importants. Aquests són els models Bernoulli, binomial, hipergeomètric i Poisson.

La distribució de Bernoulli

La distribució Bernoulli és una peça clau per a comprendre la distribució binomial, ja que aquesta última es pot considerar com una *suma* de distribucions Bernoulli. Un experiment aleatori es distribueix segons un model Bernoulli si l'experiment té únicament dos possibles resultats mútuament excloents. A aquests resultats se'ls sol denominar «èxit» i «fracàs». A més, si p és la probabilitat d'èxit, aleshores $1 - p$ és la probabilitat de fracàs. Si ara es defineix la variable aleatòria X de manera que prenga el valor 1 si es dona èxit i 0 si es dona fracàs, llavors la distribució queda completament determinada.

Notació del model

Si es denota per X la variable aleatòria, X pot prendre els valors $\{1, 0\}$: $X \approx \text{Be}(p)$ on p = robabilitat d'èxit. Cal remarcar que p és el paràmetre de la distribució.

Funció de probabilitat d'una distribució Bernouilli

Si se suposa que una variable aleatòria $X \approx \text{Be}(p)$, llavors: la funció de probabilitat és òbviament $f(1) = p$; $f(0) = 1 - p$.

Esperança i variància d'una distribució Bernouilli

Utilitzant les definicions d'esperança i variància de les distribucions discretes, juntament amb la definició de la funció de probabilitat de la variable Bernouilli, es pot calcular fàcilment l'esperança i la variància. Així, si $X \approx \text{Be}(p)$, llavors:

$$E(X) = p$$

$$\text{Var}(X) = p(1 - p)$$

La distribució binomial

Una vegada explicats els conceptes més bàsics dels models de probabilitat de les distribucions discretes, així com el model Bernouilli, cal presentar el model de probabilitat discret més important. Es tracta del model binomial. En aquest apartat es mostraran les característiques que han de complir els experiments per a poder modelar-se mitjançant aquest model. També es mostraran la funció de probabilitat del model i els estadístics més importants.

Un experiment es distribueix segons un model binomial si compleix les següents característiques:

- Consisteix a realitzar n repeticions independents d'un experiment Bernouilli.
- Aquest segon experiment, com s'ha explicat anteriorment, té dos possibles resultats (Èxit/Fracàs), i en cada repetició la probabilitat d'èxit (p) no varia, és constant (Els experiments Bernouilli són independents).
- L'objectiu de l'experiment és *comptar el nombre d'èxits* obtinguts en les n repeticions.

Notació del model

Si es denota per X la variable aleatòria X = nombre d'èxits obtinguts en els n llançaments, llavors es representa per $X \approx \text{Bi}(n, p)$ on n = nombre de realitzacions, i p = probabilitat d'èxit.

Funció de probabilitat d'una distribució binomial

Si se suposa que una variable aleatòria $X \approx \text{Bi}(n, p)$ llavors:

X pot prendre els valors $\{0, 1, 2, \dots, n\}$, ja que X és el nombre d'èxits varia entre 0 i n .

Es pot demostrar matemàticament que la funció de probabilitat és:

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, 3, \dots, n \\ 0 & \text{la resta} \end{cases}$$

La representació gràfica, com ja s'havia comentat, és un diagrama de barres (fig. 4).

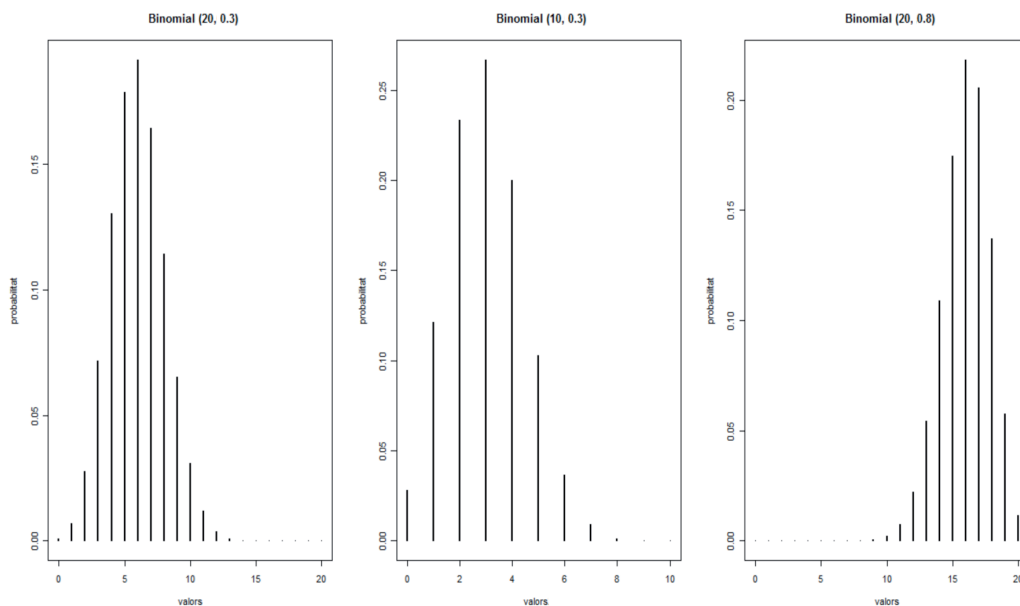


Fig. 4

Esperança i variància d'una distribució binomial

Utilitzant les definicions d'esperança i variància de les distribucions discretes, junt a la definició de la funció de probabilitat de la variable binomial, es pot demostrar matemàticament que l'esperança i la variància són:

$$E(X) = np$$

$$\text{Var}(X) = np(1-p)$$

La distribució hipergeomètrica

Un dels supòsits de la distribució binomial és que la probabilitat d'èxit en cada experiment Bernoulli és constant. Si aquest supòsit no s'exigeix, llavors no es pot considerar una distribució binomial. Si aquest és l'únic supòsit que deixa de complir-se, aleshores la distribució de probabilitats s'anomena hipergeomètrica.

Notació del model

Si es denota per X la variable aleatòria, amb $X =$ nombre d'èxits, llavors aquest model es representa per $X \approx \text{HG}(n, m, r)$ on:

$n =$ nombre total d'elements.

$m =$ nombre d'elements del subconjunt.

$r =$ nombre d'elements del total n que són èxits.

o que compleixen la condició.

Funció de probabilitat d'una distribució hipergeomètrica

Si se suposa que una variable aleatòria $X \approx \text{HG}(n, m, r)$, llavors:

X pot prendre els valors: $\{0, 1, 2, \dots, r\}$, ja que X és el nombre d'èxits i varia entre 0 i r .

Es pot demostrar matemàticament que la funció de probabilitat és:

$$P(X = k) = f(x) = \begin{cases} \frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}} & \text{per a } k = 1, \dots, r \text{ i } r \geq k \\ 0 & \text{la resta} \end{cases}$$

La representació gràfica és un diagrama de barres (fig. 5).

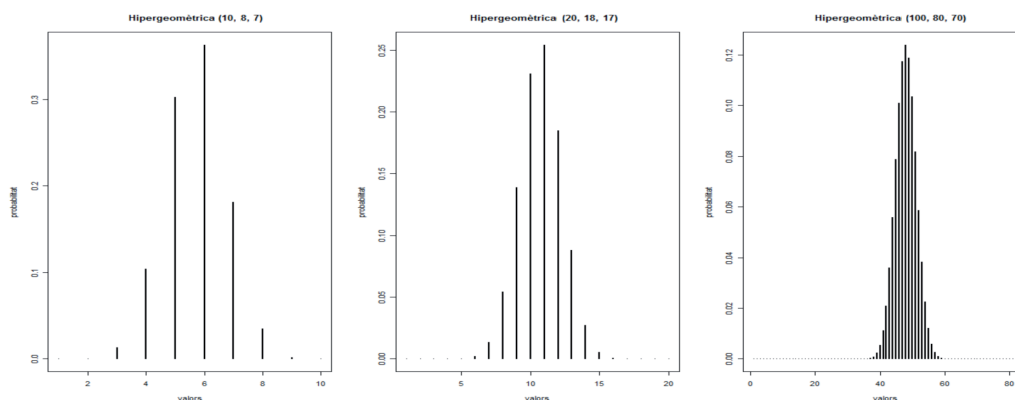


Fig. 5

Esperança i variància d'una distribució hipergeomètrica

Utilitzant les definicions d'esperança i variància de les distribucions discretes, junt a la definició de la funció de probabilitat de la variable hipergeomètrica, es pot demostrar matemàticament que l'esperança i la variància són:

$$E(X) = \frac{m \cdot r}{n}$$

$$\text{Var}(X) = \frac{n-m}{n-1} \cdot \frac{m \cdot r}{n} \cdot \left(1 - \frac{r}{n}\right)$$

La distribució Poisson

La distribució de Poisson fou proposta per primera vegada per Siméon Poisson (1781-1840) en un llibre publicat en 1837. La distribució Poisson és una important distribució de probabilitat discreta de nombroses aplicacions, com poden ser el nombre d'errades d'un gran sistema informàtic, el nombre de vaixells que arriben a un port en un període de sis hores, el nombre de camions de repartiment que arriben a un magatzem en una hora, el nombre de defectes d'una peça metàl·lica, el nombre de clients que arriben a una caixa en un supermercat en un interval concret de temps, etc.

Així, es pot emprar la distribució de Poisson en les distribucions que es caracteritzen per ser el nombre d'ocurrències o èxits d'un succés en un interval continu donat (com el temps, la superfície o la longitud). En resum, la distribució Poisson mesura la probabilitat d'un esdeveniment aleatori en algun interval continu.

Són necessaris tres supòsits per a poder aplicar la distribució Poisson. En primer lloc se suposa que l'interval està dividit en una gran quantitat de subinterval·ls de la mateixa amplitud i en els quals la probabilitat que es done una ocurrència és ínfima. Tenint en compte aquesta consideració, els supòsits són:

- La probabilitat que ocorregui un esdeveniment és constant en tots els subinterval·ls.
- No pot haver més d'una ocurrència en cada subinterval.
- Les ocurrències són independents; és a dir, les ocurrències en interval·ls que no se solapen són independents entre si.

Notació del model

Si es denota per X a la variable aleatòria $X =$ nombre d'ocurrències en un interval, llavors es representa per $X \approx \text{Po}(\mu)$, on $\mu =$ mitjana d'esdeveniments per unitat de temps o espai.

Funció de probabilitat d'una distribució Poisson

Si es denota per X la variable aleatòria, llavors X pot prendre els valors $\{0, 1, 2, 3, 4, \dots\}$ i la funció de probabilitat de la distribució Poisson es representa mitjançant un diagrama de barres (fig. 3), que s'expressa com:

$$P(X = x) = f(x) = \frac{\mu^x \cdot e^{-\mu}}{x!}$$

On x és el nombre de vegades que ocorre l'esdeveniment.

μ és la Mitjana d'esdeveniments per unitat de temps o espai.

$X \approx \text{Po}(\mu)$.

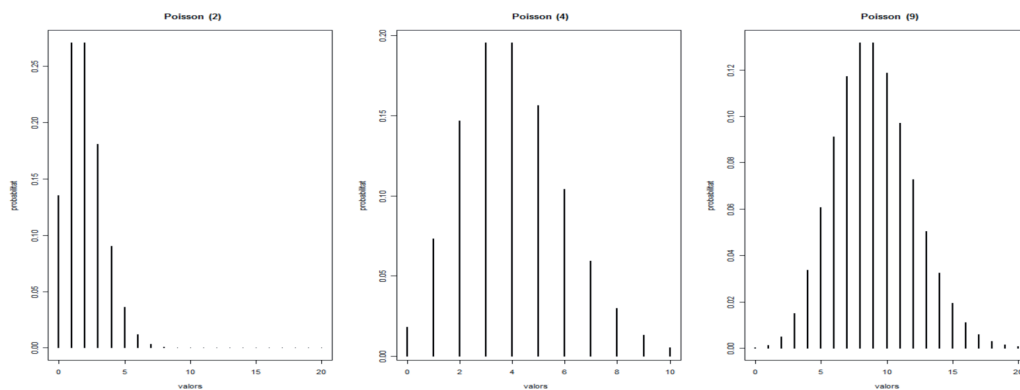


Fig. 6

Esperança i variància d'una distribució Poisson

Utilitzant les definicions d'Esperança i variància de les distribucions discretes, junt a la definició de la funció de probabilitat de la variable Poisson, es pot demostrar matemàticament que l'esperança i la variància són:

$$E(X) = \mu$$

$$\text{Var}(X) = \mu$$

Aproximació Poisson de la binomial

La distribució de probabilitats Poisson està molt relacionada amb la distribució binomial. De fet, quan el nombre de repeticions de l'experiment Bernoulli és molt elevat i la probabilitat d'èxit és reduïda, es pot aproximar una distribució per l'altra.

Així doncs, en els casos en què el nombre de repeticions n i la probabilitat d'èxit p compleixen $np \leq 7$, pot aproximar-se una distribució $\text{Bi}(n, p)$ per una distribució $\text{Po}(np)$. És a dir, per una distribució Poisson de paràmetre np . La figura reflexa aquesta relació.

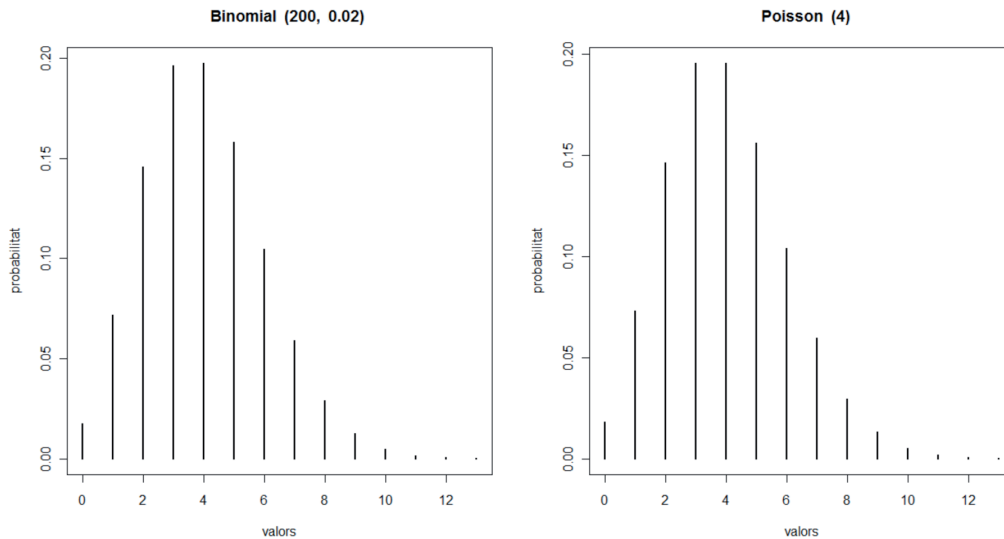


Fig. 7

Suma de variables Poisson

La distribució Poisson compleix la mateixa propietat que les distribucions binomials pel que fa a la suma de distribucions Poisson independents, és a dir:

$$\left. \begin{array}{l}
 \text{Si } X_1 \approx Po(\lambda_1) \text{ i} \\
 X_2 \approx Po(\lambda_2) \\
 \\
 X_1 \text{ i } X_2 \text{ són} \\
 \text{independents}
 \end{array} \right\} \Rightarrow X_1 + X_2 \approx Po(\lambda_1 + \lambda_2)$$

Així, per exemple, si el nombre de trucades de telèfon en una hora segueix una distribució Poisson de mitjana 12 i el nombre de cridades en la mateixa hora del dia següent segueix una distribució Poisson de 8 cridades de mitjana, llavors el nombre de cridades entre les dues hores segueix una distribució Poisson de mitjana 20.

Formalment: $X_1 \approx Po(12)$ i $X_2 \approx Po(8) \rightarrow X_1 + X_2 \approx Po(20)$.

Objectius

Els problemes han de permetre que els alumnes assolisquen els objectius didàctics següents:

- 2a) Reconèixer característiques comunes en els experiments aleatoris senzills.
- 2b) Conèixer el concepte de variable aleatòria i de model de probabilitat.
- 2c) Conèixer els trets més característics de les distribucions de probabilitat d'una variable discreta: funció de probabilitat, funcions de distribució, esperança matemàtica, variància, percentils, etc.
- 2d) Identificar els experiments que es modelen mitjançant una distribució discreta.
- 2e) Conèixer i saber fer càlculs amb les distribucions discretes de Bernoulli, binomial, hipergeomètrica i de Poisson.
- 2f) Saber reconèixer i aplicar la llei dels successos rars per transformar el model binomial en Poisson.

La taula següent ens mostra com estan distribuïts els objectius segons els exercicis:

Objectiu Exercici	2a	2b	2c	2d	2e	2f
1	X	X	X			
2	X	X	X			
3	X	X	X			
4	X	X	X			
5	X	X	X			
6			X	X	X	
7			X	X	X	
8			X	X	X	
9			X	X	X	
10			X	X	X	X

Enunciats

-
- 2a) Reconèixer característiques comunes en els experiments aleatoris senzills.
- 2b) Conèixer el concepte de variable aleatòria i de model de probabilitat.
- 2c) Conèixer els trets més característics de les distribucions de probabilitat: funció de probabilitat, funcions de distribució, esperança matemàtica, variància, percentils, etc.

Exercici 1

Hem fet un estudi de la distribució del personal en els diferents departaments d'un servei públic de 8 treballadors i després de fer el recompte hem obtingut aquestes dades:

nre. de treballadors	1	2	3	4	5	6	7	8
percentatge (%)	3	9	8	24	26	15	5	10

- a) Troba l'expressió de les funcions de probabilitat i distribució que corresponen a aquestes dades.
- b) Calcula l'esperança i la variància d'aquesta distribució.
- c) Calcula les següents probabilitats:
- I. La probabilitat de que un departament tinga menys de 4 treballadors.
 - II. La probabilitat de que un departament tinga entre 5 i 7 treballadors, ambdós inclosos.
 - III. La probabilitat de que un departament tinga més de 6 treballadors.
- d) Calcula el nombre mínim de treballadors que té el 75 % dels departaments.

Exercici 2

Imagina que tirem dos daus sobre la taula i anotem la resta dels resultats que obtenim.

- a) Quin creus que serà el resultat menys probable? Si tingueres que apostar per un únic resultat, quina hauria de ser la teua aposta per jugar amb major probabilitat d'èxit?
- b) Calcula tot l'espai mostral i construeix la funció de probabilitat i la funció de distribució.
- c) Calcula la probabilitat de traure menys de 2, igual o més de 4, entre 2 i 5. Realitza aquests càlculs utilitzant les dues funcions de l'apartat anterior i trau les teues conclusions sobre la seua idoneïtat.

Exercici 3

Per conèixer el nivell d'estudis en certa llengua estrangera dels treballadors d'una empresa, hem fet una equivalència adjudicant una escala de l'1 al 5 als certificats de nivell que han presentat als seus currículums, i hem obtingut aquesta funció de distribució dels 5 nivells que hem establert:

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ 0.1 & \text{si } 1 \leq x < 2 \\ 0.2 & \text{si } 2 \leq x < 3 \\ 0.5 & \text{si } 3 \leq x < 4 \\ 0.9 & \text{si } 4 \leq x < 5 \\ 1 & \text{si } 5 \leq x \end{cases}$$

- a) Calcula l'esperança i la variància d'aquesta distribució. Interpreta els resultats.
- b) Calcula la probabilitat de que si triem un treballador a l'atzar d'aquesta empresa tinga:
 - i. Un nivell 3.
 - ii. Un nivell superior a 2.
 - iii. Un nivell igual o superior a 2.
 - iv. Un nivell inferior a 3.
 - v. Un nivell major que 2, però igual o menor que 4.

NOTA: Fixa't amb la conveniència de treballar amb la funció de probabilitat o amb la funció de distribució.
- c) Calcula el nivell mínim que podem assegurar que té el 60 % dels treballadors per a oferir els nostres productes a l'exterior.
- d) Calcula el nivell màxim que podríem assegurar que té el 40 % dels treballadors per a demanar un pla de formació.

Exercici 4

Imagina que llancem tres monedes a l'aire i comptem el nombre de cares que han eixit.

- a) Calcula la probabilitat de cadascun dels possibles valors de la variable X i defineix la funció de probabilitat corresponent.
- b) Calcula la probabilitat d'obtenir 2 cares, almenys 2 cares, i més de 2 cares en un llançament.
- c) Calcula la probabilitat d'obtenir un nombre de cares entre 0 i 2, ambdós inclusos.
- d) Calcula la funció de distribució, expressa-la i representa-la.
- e) Calcula el nombre de cares que s'obté com a màxim en el 75 % dels llançaments.
- f) Calcula l'esperança i la variància i interpreta-les.

Exercici 5

Se està controlant el ritme de treball que porta un servei d'atenció al públic. S'estima que es poden atendre 12 persones cada hora. Del recompte que hem fet amb les dades del mes passat, hem obtingut la següent funció de distribució:

$$F(x) = \begin{cases} 0 & \text{si } x < 10 \\ 0,35 & \text{si } 10 \leq x < 11 \\ 0,55 & \text{si } 11 \leq x < 12 \\ 0,65 & \text{si } 12 \leq x < 13 \\ 0,85 & \text{si } 13 \leq x < 14 \\ 1 & \text{si } x \geq 14 \end{cases}$$

- Calcula la probabilitat de que en una hora siguin ateses 9 persones; menys de 12 persones; més de 11, però menys de 13.
- Calcula la funció de probabilitat de la variable X .
- Calcula les probabilitats de l'apartat *a* utilitzant la funció de probabilitats.
- Calcula el nombre mínim de persones que podem assegurar que han estat ateses el 80 % de les hores. Calcula la mediana.
- Calcula l'esperança i la variància d'aquesta distribució i interpreta-les.

-
- Conèixer el concepte de variable aleatòria i de model de probabilitat.
 - Conèixer els trets més característics de les distribucions de probabilitat: funció de probabilitat, funcions de distribució, esperança matemàtica, variància, percentils, etc.
 - Identificar els experiments que es modelen mitjançant una distribució Discreta.
 - Conèixer i saber fer càlculs amb les distribucions discretes de Bernoulli, binomial, hipergeomètrica i de Poisson.

Exercici 6

En aquest moment, el percentatge d'atur entre els llicenciats menors de 29 anys a Espanya és del 23 %¹. Considerem un grup de 5 joves llicenciats menors de 29 anys.

- Calcula la probabilitat de que 3 d'ells estiguen aturats.
- Calcula la probabilitat de que més de 2 joves estiguen treballant.

- c) Troba la funció de probabilitat de cada valor possible de la variable $X =$ nombre de joves treballant en el grup de 5 llicenciats de l'enunciat.
- d) Troba el model de probabilitat que podem ajustar a aquest plantejament.
- e) Escriu l'expressió algebraica d'aquesta funció per a un grup de n joves.
- f) Si ara considerem un grup de 20 joves ($n = 20$), calcula l'esperança i la variància.

Exercici 7

Sabem que a la nostra empresa el nombre de treballadors que desenvolupen tasques directives és del 15 %. Considerem que en cada departament hi ha 10 treballadors.

- a) Calcula la probabilitat de que en un departament qualsevol hi haja 2 directius; menys de 4; més de 6; 4 com a molt.
- b) Calcula el nombre mitjà de directius en els departaments i la desviació típica corresponent. Interpreta els resultats.
- c) Quants directius podem assegurar que tenen com a molt el 80 % dels departaments?
- d) Calcula la probabilitat de que hi haja 8 treballadors no directius.

Imaginem ara que els departaments tingueren 50 treballadors.

- e) Calcula la probabilitat de que hi haja 15 directius.

Exercici 8

En el servei de teleassistència de 24 hores d'un centre de benestar social comarcal, s'han comptabilitzat les demandes dels usuaris, que de forma aleatòria han estat de 5 incidències per hora, al llarg de tot el dia.

- a) Calcula la probabilitat de que es reben 3, 4, 5 i 6 incidències en una hora.
- b) Calcula la mitjana d'incidències en un dia.
- c) Calcula el nombre d'incidències que s'atenen, com a molt, en el 80 % de les hores.

Exercici 9

Estem preparant unes oposicions que tenen 50 temes a l'apartat teòric i sóc conscient de que en aquests moments en podria defensar amb èxit 30. En la convocatòria, s'explicita que el tribunal extraurà 5 boles, cadascuna corresponent a una unitat

del temari. Es tria un d'aquests temes i se'n fa el desenvolupament teòric. Calcula les següents probabilitats:

- a) Podria defensar amb èxit tan sols un tema de les boles extretes.
- b) Podria defensar amb èxit, com a molt, dos temes de les boles extretes.
- c) Podria defensar amb èxit els cinc temes que han eixit.
- d) No podria defensar cap tema.
- e) Calcula la mitjana de temes que podré defensar en cada extracció de boles i la seua variabilitat.

-
- 2b) Conèixer el concepte de variable aleatòria i de model de probabilitat.
 - 2c) Conèixer els trets més característics de les distribucions de probabilitat: funció de probabilitat, funcions de distribució, esperança matemàtica, variància, percentils, etc.
 - 2d) Identificar els experiments que es modelen mitjançant una distribució discreta.
 - 2e) Conèixer i saber fer càlculs amb les distribucions discretes de Bernoulli, binomial, hipergeomètrica i de Poisson.
 - 2f) Saber reconèixer i aplicar la llei dels successos rars per transformar el model binomial en Poisson.

Exercici 10

A la premsa del passat mes d'octubre de 2013 es van publicar aquestes dades: «un 10 % de les llars espanyoles tenen tots els seus membres a l'atur».¹ Si en una aula de primària d'un col·legi públic tenim 30 alumnes (suposem que cadascun pertany a una llar diferent), es podria considerar una mostra representativa de la societat, i així, calcula:

- a) La mitjana d'alumnes d'aquesta classe les famílies dels quals pateixen aquesta situació.
- b) La probabilitat de necessitar entre 2 i 5 beques de menjador en aquesta classe per als alumnes que estan en aquesta situació.

1. <http://www.rtve.es/noticias/20121026/10-hogares-espanoles-tiene-todos-miembros-paro/571737.shtml>

Ajudes

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de realitzar els exercicis i problemes. És convenient no fer un abús excessiu d'aquestes ajudes, és a dir, abans d'emprar l'ajuda cal pensar el problema almenys durant uns 10-15 minuts. Després es consultarà l'ajuda de tipus 1 i s'intentarà resoldre l'exercici amb aquesta ajuda. Si no és possible resoldre'l, llavors es consultarà l'ajuda de tipus 2; i en darrer terme la solució.

Ajudes de tipus 1

Exercici 1

Per a trobar la funció de probabilitat, cal recordar que a cada valor de la variable $\text{Dom}(f) = \{1, 2, 3, 4, 5, 6, 7, 8\}$ li fem correspondre la seua probabilitat, que es pot extraure de l'enunciat directament convertint els percentatges de la taula (en tant per cent) a l'escala de probabilitats (en tant per un).

Una vegada trobem $f(x) = P(X = x)$, caldrà construir $F(x) = P(X \leq x)$; amb aquestes expressions ja es poden abordar tots els apartats.

Recordeu que per a calcular l'esperança i la variància podreu trobar-ne les seues expressions al resum teòric.

Exercici 2

Per començar aquest exercici seria convenient que, per trobar l'espai mostral i les probabilitats dels successos elementals, omplirem aquesta taula completament:

—	1	2	3	4	5	6
1	0	1	2		4	5
2	1	0				4
3	2	1	0		2	3
4	3		1			
5					0	1
6	5					0

Després, fa el recompte de les probabilitats de cadascun dels valors que hem obtingut per construir les funcions $f(x) = P(X = x)$, i $F(x) = P(X \leq x)$; amb aquestes expressions ja es poden abordar tots els apartats.

Exercici 3

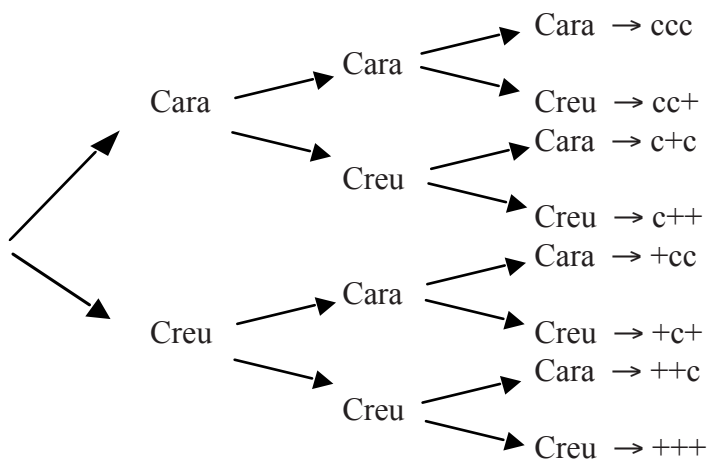
Per a calcular l'esperança i la variància d'una distribució necessitem trobar la funció de probabilitats $f(x)$, la qual extraurem de la funció de distribució $F(x)$ que tenim a l'enunciat. Utilitzarem la següent igualtat i l'aplicarem a cadascun dels valors de la variable $X = \{1, 2, 3, 4, 5\}$.

Per obtenir la funció de probabilitats $f(x_i) = F(x_i) - F(x_{i-1})$, que resulta del concepte o definició de la funció de distribució, ja que aquesta acumula els valors de la funció de probabilitat dels valors de la variable anteriors: $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$.

Recordeu que per a calcular l'esperança i la variància podreu trobar-ne les seues expressions al resum teòric.

Exercici 4

Per començar caldrà trobar l'espai mostral i la funció de probabilitats: Podem ajudar-nos fent el recompte sobre el següent diagrama d'arbre; definirem $X =$ nombre de cares obtingudes en el llançament de tres monedes.



$E = \{ccc, cc+, c+c, c++, +cc, +c+, ++c, +++\}$.

En aquesta taula, hem definit la *funció de probabilitat* de la variable X , que és una variable discreta.

x_i	0	1	2	3
$f(x_i) = P(X = x_i)$	1/8	3/8	3/8	1/8

Amb ella es poden abordar la resta dels apartats.

Exercici 5

Per calcular les probabilitats que ens demanen a l'apartat a utilitzarem la funció de distribució $F(x)$ que tenim a l'enunciat.

En un apartat ens demanen obtenir la funció de probabilitat $f(x_i) = F(x_i) - F(x_{i-1})$, que resulta del concepte o definició de la funció de distribució ja que aquesta acumula els valors de la funció de probabilitat dels valors de la variable anteriors:
 $F(x) = (X \leq x) = \sum_{x_i \leq x} f(x_i)$.

Recordeu que per a calcular l'esperança i la variància podreu trobar-ne les seues expressions al resum teòric.

Exercici 6

Per a plantejar-nos el problema, definim la variable X = nombre de joves aturats. És fàcil comprendre que, si considerem tan sols si un jove està aturat o no, es pot considerar un model de Bernouilli, on p = probabilitat d'estar aturat = 0,23.

Quan considerem un grup de 5 joves i cadascun d'ells pot estar o no aturat, passem a considerar un model binomial de paràmetres $n = 5$ i $p = 0,23$.

Recordeu que per a calcular l'esperança i la variància podreu trobar-ne les seues expressions al resum teòric.

Exercici 7

Per a plantejar-nos el problema, definim la variable X = nombre de directius en un departament de 10 treballadors. És fàcil comprendre que, si considerem tan sols si un treballador és directiu o no, es pot considerar un model de Bernouilli, on p = probabilitat de ser directiu = 0,15.

Quan considerem un departament de 10 treballadors i cadascun d'ells pot ser directiu o no, passem a considerar un model binomial de paràmetres $n = 10$ i $p = 0,15$.

Per a calcular probabilitats caldrà ajudar-vos del resum teòric, on podeu trobar l'expressió algebraica de la funció de probabilitats i substituir els valors dels paràmetres n i p pels del nostre exercici.

Recordeu que per a calcular l'esperança i la variància podreu trobar-ne les seues expressions al resum teòric.

Si el departament canvia el nombre de treballadors a 50, caldrà fer nous càlculs amb valors dels paràmetres diferents.

Exercici 8

Per a plantejar-nos el problema començarem per definir la variable X = nombre d'assistències ateses per hora. Sabem que la mitjana és de 5 incidències per hora i suposarem que aquesta demanda és proporcional a l'interval de temps contemplat. Podem considerar que la variable X es distribueix com una variable Poisson de mitjana $\lambda = 5$.

Cal trobar la funció de probabilitats d'aquest model i substituir el valor del paràmetre. Podem trobar-ne l'expressió algebraica en el resum teòric.

Per fer l'apartat b cal adaptar el paràmetre a l'interval d'un dia.

Per calcular el percentil cal trobar els valors de la funció de distribució $F(x)$.

Exercici 9

Per a plantejar-nos el problema començarem per definir la variable:

X = nre. d'unitats que puc defensar amb èxit entre els temes que ixen en agafar 5 boles a l'atzar que corresponen a 50 unitats que hi ha al temari. Considerarem que tan sols puc defensar 30 dels 50 temes possibles.

Podem identificar l'experiment com una variable hipergeomètrica i identificar els paràmetres del model: N , K , n .

Així: $X \rightarrow \text{Hiper}(N = 50, n = 5, K = 30)$.

Per respondre a les qüestions plantejades, caldrà trobar al resum teòric les expressions algebraiques generals de la funció de probabilitats, l'esperança i la variància.

Exercici 10

Per a plantejar-nos el problema començarem per definir la variable aleatòria: X = nre. de llars que tenen tots els seus membres a l'atur o alumnes que pateixen a casa aquesta situació, suposant que cada alumne pertany a una llar diferent i que a classe hi ha 30 alumnes.

Aquesta descripció i el coneixement de la dada de que la probabilitat de que una llar triada a l'atzar tinga tots els seus membres a l'atur és del 10 %, ens fan considerar que la nostra variable s'ajusta a un model de probabilitat binomial amb paràmetres $n = 30$ i $p = 0,10$.

Ajudes de tipus 2

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de resoldre els exercicis i problemes, després d'haver consultat l'ajuda de tipus 1.

Exercici 1

En general, per fer aquests exercicis et serà pràctic consultar aquesta taula:

<i>Ens demanen</i>	<i>Necessitarem conèixer</i>
$E[X], \text{Var}[X]$	$f(x)$
Probabilitats	$f(x), F(x)$ segons els casos
Percentils	$F(x)$

Si considerem aquestes funcions de probabilitats i funcions de distribucions, podrem resoldre fàcilment la resta d'apartats:

x_i	1	2	3	4	5	6	7	8
$f(x_i)$	0,03	0,09	0,08	0,24	0,26	0,15	0,05	0,10

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 1 \\ 0.03 & \text{si } 1 \leq x < 2 \\ 0.12 & \text{si } 2 \leq x < 3 \\ 0.20 & \text{si } 3 \leq x < 4 \\ 0.44 & \text{si } 4 \leq x < 5 \\ 0.70 & \text{si } 5 \leq x < 6 \\ 0.85 & \text{si } 6 \leq x < 7 \\ 0.90 & \text{si } 7 \leq x < 8 \\ 1 & \text{si } x \geq 8 \end{cases}$$

Pots comprovar també que els resultats són:

$$E[X] = \sum_{i=1}^n x_i \cdot f(x_i) = 4.76 \quad \text{Var}[X] = E[X^2] - E[X]^2 = 78.024$$

Per a calcular les probabilitats en una variable discreta serà més fàcil utilitzar la $f(x)$ dels valors concrets que compleixen l'enunciat, fixant-se especialment en els valors dels extrems, que estaran inclosos o no.

Exercici 2

Una vegada ja tenim la taula completada i fet el recompte de les probabilitats de cadascun dels successos elementals, podem obtenir la funció de probabilitat:

x_i	0	1	2	3	4	5
$f(x_i)$	1/6	5/18	2/9	1/6	1/9	1/18

I amb aquesta, per trobar l'expressió de la funció de distribució $F(x)$, recordem la seua definició: $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$. Així, anirem acumulant els valors de la $f(x_i)$ i podrem expressar-la:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/6 & \text{si } 0 \leq x < 1 \\ 4/9 & \text{si } 1 \leq x < 2 \\ 2/3 & \text{si } 2 \leq x < 3 \\ 5/6 & \text{si } 3 \leq x < 4 \\ 17/18 & \text{si } 4 \leq x < 5 \\ 1 & \text{si } x \geq 5 \end{cases}$$

Per a calcular les probabilitats en una variable discreta podem utilitzar qualsevol de les dues funcions; caldria saber, en cada enunciat, quina és la funció més adient.

Exercici 3

Una vegada ja hem obtingut la funció de probabilitats $f(x)$, els valors de la qual figuren a la següent taula,

x_i	1	2	3	4	5
$f(x_i)$	0.1	0.1	0.3	0.4	0.1

Podrem calcular l'esperança i la variància. Els seus valors, com pots comprovar, són:

$E[X] = 3,3$ el nivell mitjà de coneixement dels treballadors de l'empresa.

$\text{Var}\{X\} = E[X^2] - E[X]^2 = 12,1 - 3,3^2 = 1,21$ nivell², és a dir: $E[X^2] = \sum_{i=1}^n x_i^2 f(x_i) = 12,1$

Si volem traure la *desviació típica* $= \sigma = \sqrt{\text{VAR}[X]} = \sqrt{1,21} = 1,1$ nivell,

Per a calcular les diferents probabilitats dels apartats següents, podem utilitzar qualsevol de les dues funcions treballades en els exercicis anteriors.

Per a calcular el nivell mínim que podem assegurar que té el 60 % dels treballadors a l'hora d'oferir els nostres productes a l'exterior, expressarem aquesta frase en termes de probabilitat recordant el concepte de percentil. Si x_0 és el valor de la variable (nivell de llengua) que ens demanen, llavors:

$$P(X \geq x_0) = 0,60 \Leftrightarrow P(X \leq x_0) = 0,40$$

És a dir, que cal trobar el percentil P_{40} d'aquesta distribució. Ho resoldrem utilitzant la funció de distribució, on cal trobar $F(x_0) = 0,40$.

Per a calcular el nivell màxim que podríem assegurar que té el 40 % dels treballadors a l'hora de demanar un pla de formació, farem un raonament semblant al de l'apartat anterior: $P(X \leq x_0) = 0,40$ i podem veure que la solució és la mateixa, només canvia la matisació de la conclusió: el 40 % dels treballadors tenen un nivell inferior o igual a 3.

Exercici 4

A les ajudes de tipus 1 ja hem indicat els valors de la funció de probabilitats. Per a fer els següents apartats, considerarem la definició de la funció de distribució $F(x) = P(X \leq x)$.

L'apartat *e* ens demana un percentil i les expressions de l'esperança i la variància podem trobar-les al resum teòric.

Exercici 5

Per a calcular la funció de probabilitats serà suficient calcular els «increments» de la funció de distribució en cadascun dels valors de la variable. Per a això podem ajudar-nos de la taula següent:

x_i	10	11	12	13	14	
$F(x_i)$	0,35	0,55	0,65	0,85	1	
$f(x_i)$	0,35	0,20	0,10	0,20	0,15	$\sum f(x_i) = 1$

Amb aquesta taula podem abordar l'exercici, i per calcular l'esperança i la variància tan sols cal recordar-ne les expressions, substituir els valors dels paràmetres convenientment i comprovar que els resultats són:

$$E[X] = n \cdot p = 11,6 \qquad \text{Var}[X] = n \cdot p \cdot (1 - p) = 2,62$$

Exercici 6

Podrem utilitzar directament l'expressió algebraica de la teoria, on tal sols cal identificar els valors de n i p .

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, 3, \dots, n \\ 0 & \text{la resta} \end{cases}$$

En els diferents apartats, $p = 0,23$, i en els primers apartats el valor de $n = 5$, mentre que en els darrers, $n = 20$.

Per a calcular l'esperança i la variància del model binomial tan sols cal recordar-ne les expressions, substituir els valors dels paràmetres convenientment i comprovar que els resultats són:

$$E[X] = n \cdot p = 4,6 \qquad \text{Var}[X] = n \cdot p \cdot (1 - p) = 3,542$$

Exercici 7

Podrem utilitzar directament l'expressió algebraica de la teoria, on tal sols cal identificar els valors de n i p .

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, 3, \dots, n \\ 0 & \text{la resta} \end{cases}$$

En els diferents apartats, $p = 0,15$ i en els primers apartats el valor de $n = 10$, mentre que en els darrers, $n = 50$.

Per a calcular l'esperança i la variància del model binomial tan sols cal recordar-ne les expressions, substituir els valors dels paràmetres convenientment i comprovar que en els departaments de 10 treballadors els resultats són:

$$E[X] = n \cdot p = 1,5 \qquad \text{Var}[X] = n \cdot p \cdot (1 - p) = 1,275$$

Exercici 8

Podrem utilitzar directament l'expressió algebraica de la teoria, on tal sols cal identificar el valor del paràmetre. La funció de probabilitat de la distribució Poisson s'expressa com:

$$P(X = x) = f(x) = \frac{\mu^x \cdot e^{-\mu}}{x!}$$

Per calcular la mitjana, considerem que l'esperança és $E(X) = \mu$.

Per a calcular P_{80} , cal trobar els valors de la funció de distribució, sumant i acumulant els valors de la funció de probabilitats.

Exercici 9

Ja hem definit la variable aleatòria: $X = \text{nre. d'unitats que puc defensar amb èxit entre els temes que ixen en agafar 5 boles a l'atzar que corresponen a 50 unitats que hi ha al temari. L'hem identificada com una variable hipergeomètrica, els paràmetres de la qual són:$

$N = 50$ temes que hi ha a l'oposició i que es corresponen amb 50 boles en total.

$n = 5$ temes que s'extrauen sense reposició i que es corresponen amb 5 temes possibles a desenvolupar triats a l'atzar.
 $K = 30$ temes que conec i puc defensar amb èxit entre els 50 que hi ha.

Així: $X \rightarrow$ Hiper ($N = 50, n = 5, K = 30$).

Recordeu que $f(x) = \frac{\binom{K}{x} \cdot \binom{N-K}{n-x}}{\binom{N}{n}}$ $E[X] = n \frac{K}{N}$ i $Var[X] = n \frac{K}{N} \left(1 - \frac{K}{N}\right) \frac{N-n}{N-1}$

Exercici 10

Cal definir la variable aleatòria: $X = nre$, de llars que tenen tots els seus membres a l'atur o alumnes que pateixen a casa aquesta situació, suposant que cada alumne pertany a una llar diferent i que a classe hi ha 30 alumnes.

Aquesta descripció i el coneixement de la dada de que la probabilitat de que una llar triada a l'atzar tinga tots els seus membres a l'atur és del 10 %, ens fan considerar que la nostra variable s'ajusta a un model de probabilitat binomial amb paràmetres $n = 30$ i $p = 0,10$.

- a) Per a calcular la mitjana recordarem que al model binomial $E[X] = n \cdot p$.
- b) Per a calcular les probabilitats que ens demanen, aplicarem la llei dels successos rars i veurem l'aproximació que podem fer del model binomial mitjançant el model Poisson. Cal consultar a la teoria les condicions que caldrà que complisquen els paràmetres n i p del model binomial per poder fer la transformació a una Poisson amb paràmetre $\lambda = n \cdot p$.

Solucions

Exercici 1

Hem fet un estudi de la distribució del personal en els diferents departaments d'un servei públic de 8 o menys treballadors i després de fer el recompte hem obtingut aquestes dades:

nre. treballadors	1	2	3	4	5	6	7	8
percentatge (%)	3	9	8	24	26	15	5	10

- Troba l'expressió de les funcions de probabilitat i distribució que corresponen a aquestes dades.
- Calcula l'esperança, la variància i desviació típica d'aquesta distribució.
- Calcula les següents probabilitats:
 - La probabilitat de que un departament tinga menys de 4 treballadors.
 - La probabilitat de que un departament tinga entre 5 i 7 treballadors, ambdós inclosos.
 - La probabilitat de que un departament tinga més de 6 treballadors.
- Calcula el nombre mínim de treballadors que té el 75 % dels departaments.

Solució

- Per trobar l'expressió de la funció de probabilitats recordarem que $f(x) = P(X = x)$ que l'expressem en termes de freqüència relativa $0 \leq f(x) \leq 1$ i que es pot comprovar que $\sum_{i=1}^n x_i^2 f(x_i) = 1$.

Com que la taula de les dades ve donada en tants per cent, serà suficient amb convertir-la en tants per u.

En aquest problema, la variable aleatòria serà $X =$ nre. de treballadors en els departaments menuts (menors de 9 treballadors) del servei públic estudiat. Així, podem expressar la funció de distribució com una taula com la que figura a continuació:

x_i	1	2	3	4	5	6	7	8
$f(x_i)$	0,03	0,09	0,08	0,24	0,26	0,15	0,05	0,10

O també la podem expressar com una funció que tan sols té com a $\text{Dom}(f) = \{1, 2, 3, 4, 5, 6, 7, 8\}$ i podem classificar-la com una «gràfica de punts». Així:

$$f(x) = P(X = x) = \begin{cases} 0,03 & \text{si } x = 1 \\ 0,09 & \text{si } x = 2 \\ 0,08 & \text{si } x = 3 \\ 0,24 & \text{si } x = 4 \\ 0,26 & \text{si } x = 5 \\ 0,15 & \text{si } x = 6 \\ 0,05 & \text{si } x = 7 \\ 0,10 & \text{si } x = 8 \\ 0 & \text{en qualsevol altre valor} \end{cases}$$

Per a trobar l'expressió de la funció de distribució $F(x)$ recordem la seua definició: $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$. Així, anem acumulant els valors de $f(x_i)$ i podem expressar-la:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 1 \\ 0,03 & \text{si } 1 \leq x < 2 \\ 0,12 & \text{si } 2 \leq x < 3 \\ 0,20 & \text{si } 3 \leq x < 4 \\ 0,44 & \text{si } 4 \leq x < 5 \\ 0,70 & \text{si } 5 \leq x < 6 \\ 0,85 & \text{si } 6 \leq x < 7 \\ 0,90 & \text{si } 7 \leq x < 8 \\ 1 & \text{si } x \geq 8 \end{cases}$$

b) Per a trobar l'esperança i la variància de la distribució necessitem treballar amb la funció de distribució que hem calculat a l'apartat anterior.

Comencem pel càlcul de l'esperança: $E[X] = \sum_{i=1}^n x_i \cdot f(x_i)$.

$$E[X] = 1 \cdot 0,03 + 2 \cdot 0,09 + 3 \cdot 0,08 + 4 \cdot 0,24 + 5 \cdot 0,26 + 6 \cdot 0,15 + 7 \cdot 0,05 + 8 \cdot 0,10 = 4,76 \text{ treballadors.}$$

És a dir, la mitjana del nombre de treballadors per departament és de 4,76.

I per a calcular la variància utilitzarem la propietat que es pot veure al capítol de la teoria: $\text{Var}\{X\} = E[X^2] - E[X]^2$.

$$\text{Var}\{X\} = E[X^2] - E[X]^2 = 25,7 - 4,76^2 = 7,8024 \text{ treballadors}^2$$

$$E[X^2] = \sum_{i=1}^n x_i^2 f(x_i) = 1^2 \cdot 0,03 + 2^2 \cdot 0,09 + 3^2 \cdot 0,08 + 4^2 \cdot 0,24 + 5^2 \cdot 0,26 + 6^2 \cdot 0,15 + 7^2 \cdot 0,05 + 8^2 \cdot 0,10 = 25,7$$

Aquesta variància ens permet calcular la *desviació típica* com a mesura de la dispersió respecte de la mitjana de les dades, en la mateixa magnitud que la variable, que facilitarà la seua interpretació.

$$\text{Desviació típica} = \sigma = \sqrt{\text{VAR}[X]} = \sqrt{7,8024} = 2,79 \text{ treballadors.}$$

Podem interpretar que entre el nombre de treballadors dels departaments i l'esperança calculada (4,76) hi ha una diferència mitjana de 2,79 treballadors.

c) Per a calcular les probabilitats que ens demanen podem utilitzar tant la funció de probabilitat $f(x)$, com la funció de distribució $F(x)$. En aquesta solució ho farem amb les dues possibilitats i el lector anirà descobrint en cada cas quina li pot convindre més i extraure les seues conclusions, tot i que en aquest exercici ja disposem de les dues funcions per l'apartat a.

I. Calcula la probabilitat de que un departament tinga menys de 4 treballadors.

$$P(X < 4) = F(X = 3) = f(1) + f(2) + f(3) = 0,20$$

II. Calcula la probabilitat de que un departament tinga entre 5 i 7 treballadors, ambdós inclosos.

$$P(5 < X < 7) = F(7) - F(4) = 0,90 - 0,44 = 0,46$$

$$P(5 < X < 7) = f(5) + f(6) + f(7) = 0,26 + 0,15 + 0,05 = 0,46$$

III. Calcula la probabilitat de que un departament tinga més de 6 treballadors.

$$P(X > 6) = 1 - P(X \leq 6) = 1 - F(6) = 1 - 0,85 = 0,15$$

$$P(X > 6) = f(7) + f(8) = 0,05 + 0,10 = 0,15$$

d) Per a calcular el nombre mínim (x_0) de treballadors que té el 75 % dels departaments, ens plantejarem la següent expressió, que ens recorda el concepte de percentil de l'estadística descriptiva.

$P(X \geq x_0) = 0,75$, és a dir, estem buscant el percentil (P_{25}) que equival a $P(X < x_0) = 0,25$, i per a calcular aquest valor (x_0) ens fixarem en els valors de la funció $F(x)$. Si mirem la seua expressió, podem afirmar que:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 1 \\ 0.03 & \text{si } 1 \leq x < 2 \\ 0.12 & \text{si } 2 \leq x < 3 \\ 0.20 & \text{si } 3 \leq x < 4 \\ 0.44 & \text{si } 4 \leq x < 5 \\ 0.70 & \text{si } 5 \leq x < 6 \\ 0.85 & \text{si } 6 \leq x < 7 \\ 0.90 & \text{si } 7 \leq x < 8 \\ 1 & \text{si } x \geq 8 \end{cases}$$

$$P(X \leq 3) = 0,20 \rightarrow P(X > 3) = 0,80$$

$$P(X \leq 4) = 0,44 \rightarrow P(X > 4) = 0,56$$

Així doncs, podem concloure que el nombre mínim de treballadors que té el 75 % dels departaments és 3.

NOTA

Com a primera conclusió de l'exercici podem afirmar que en una variable aleatòria discreta caldrà relacionar:

<i>Ens demanen</i>	<i>Necessitarem conèixer</i>
$E[X], \text{Var}[X]$	$f(x)$
Probabilitats	$f(x), F(x)$ segons els casos
Percentils	$F(x)$

Per tant, un objectiu a treballar quan es coneix la funció de probabilitats és l'obtenció de la funció de distribució, i a l'inrevés.

Exercici 2

Imagina que tirem dos daus sobre la taula i anotem la resta dels resultats que obtenim.

- Quin creus que serà el resultat menys probable? Si tingueres que apostar per un únic resultat, quina hauria de ser la teua aposta per jugar amb major probabilitat d'èxit?
- Calcula tot l'espai mostral i construeix la funció de probabilitat i la funció de distribució.
- Calcula la probabilitat de traure menys de 2, igual o més de 4, entre 2 i 5. Realitza aquests càlculs utilitzant les dues funcions de l'apartat anterior i trau les teues conclusions sobre la seua idoneïtat.

Solució

- La variable aleatòria és $X = \text{valor absolut de la resta dels valors de les cares si tirem dos daus.}$

Començarem per trobar l'espai mostral a la taula següent taula i adjudicarem les probabilitats a cadascun dels successos elementals:

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Si mirem els possibles resultats a les caselles en blanc, l'aposta més favorable és el 1, ja que és el resultat més freqüent.

b) L'espai mostral és $E = \{0, 1, 2, 3, 4, 5\}$ correspon als possibles valors absoluts de les restes; les seues probabilitats són:

$$P(X = 0) = 6/36 = 1/6 \quad P(X = 1) = 10/36 = 5/18 \quad P(X = 2) = 8/36 = 2/9$$

$$P(X = 3) = 6/36 = 1/6 \quad P(X = 4) = 4/36 = 1/9 \quad P(X = 5) = 2/36 = 1/18$$

La funció de probabilitats serà:

x_i	0	1	2	3	4	5
$f(x_i)$	1/6	5/18	2/9	1/6	1/9	1/18

També la podem expressar com una funció que tan sols té com a $\text{Dom}(f) = \{0, 1, 2, 3, 4, 5\}$ i podem classificar-la com una «gràfica de punts». Així:

$$f(x) = P(X = x) = \begin{cases} 1/6 & \text{si } x = 0 \\ 5/18 & \text{si } x = 1 \\ 2/9 & \text{si } x = 2 \\ 1/6 & \text{si } x = 3 \\ 1/9 & \text{si } x = 4 \\ 1/18 & \text{si } x = 5 \\ 0 & \text{en qualsevol altre valor} \end{cases}$$

Per a trobar l'expressió de la funció de distribució $F(x)$ recordem la seua definició: $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$. Així, anem acumulant els valors de $f(x_i)$ i podem expressar-la:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/6 & \text{si } 0 \leq x < 1 \\ 4/9 & \text{si } 1 \leq x < 2 \\ 2/3 & \text{si } 2 \leq x < 3 \\ 5/6 & \text{si } 3 \leq x < 4 \\ 17/18 & \text{si } 4 \leq x < 5 \\ 1 & \text{si } x \geq 5 \end{cases}$$

c) Per a calcular les probabilitats dels diferents apartats abordarem els càlculs; paral·lelament, utilitzem la funció de probabilitats i la funció de distribució. Així:

- I. $P(X < 2) = F(1) = 4/9$
 $P(X < 2) = f(0) + f(1) = 1/6 + 5/18 = 4/9$
- II. $P(X \geq 4) = 1 - P(X < 4) = 1 - F(3) = 1 - 5/6 = 1/6$
 $P(X \geq 4) = f(4) + f(5) = 1/9 + 1/18 = 1/6$
- III. $P(2 < X < 5) = F(4) - F(2) = 17/18 - 2/3 = 5/18$
 $P(2 < X < 5) = f(3) + f(4) = 1/6 + 1/9 = 5/18$

NOTA

Podem concloure que, per a calcular probabilitats en les variables discretes, quan utilitzem la funció de distribució $F(x)$ cal anar amb compte amb els valors dels extrems que ens demanen per a poder agafar correctament el valor de la $F(x)$ adient, mentre que el sumatori dels valors de les $f(x_i)$ sembla un raonament més simple.

Exercici 3

Per a conèixer el nivell d'estudis de certa llengua estrangera dels treballadors d'una empresa, hem fet una equivalència adjudicant una escala de l'1 al 5 als certificats de nivell que han presentat als seus currículums i hem obtingut aquesta funció de distribució dels 5 nivells que hem establert:

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ 0,1 & \text{si } 1 \leq x < 2 \\ 0,2 & \text{si } 2 \leq x < 3 \\ 0,5 & \text{si } 3 \leq x < 4 \\ 0,9 & \text{si } 4 \leq x < 5 \\ 1 & \text{si } 5 \leq x \end{cases}$$

- a) Calcula l'esperança i la variància d'aquesta distribució. Interpreta els resultats.
- b) Calcula la probabilitat de que si triem un treballador a l'atzar d'aquesta empresa tinga:
- i. Un nivell 3.
 - ii. Un nivell superior a 2.
 - iii. Un nivell igual o superior a 2.
 - iv. Un nivell inferior a 3.
 - v. Un nivell major que 2, però igual o menor que 4.
- NOTA: Fixa't amb la conveniència de treballar amb la funció de probabilitat o amb la funció de distribució.
- c) Calcula el nivell mínim que podem assegurar que tenen el 60 % dels treballadors per a oferir els nostres productes a l'exterior.
- d) Calcula el nivell màxim que podríem assegurar que tenen el 40 % dels treballadors per a demanar un pla de formació.

Solució

- a) Per a calcular l'esperança i la variància d'una distribució necessitem trobar la funció de probabilitats $f(x)$ que extraurem de la funció de distribució $F(x)$ que tenim a l'enunciat. Utilitzarem la següent igualtat que aplicarem a cadascun dels valors de la variable $X = \{1, 2, 3, 4, 5\}$.

$f(x_i) = F(x_i) - F(x_{i-1})$ resulta del concepte o definició de la funció de distribució, ja que aquesta acumula els valors de la funció de probabilitat dels valors de la variable anteriors $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$.

Així:

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ 0,1 & \text{si } 1 \leq x < 2 \\ 0,2 & \text{si } 2 \leq x < 3 \\ 0,5 & \text{si } 3 \leq x < 4 \\ 0,9 & \text{si } 4 \leq x < 5 \\ 1 & \text{si } 5 \leq x \end{cases}$$

$$\begin{aligned} f(1) &= F(1) - F(0) = 0,1 \\ f(2) &= F(2) - F(1) = 0,2 - 0,1 = 0,1 \\ f(3) &= F(3) - F(2) = 0,5 - 0,2 = 0,3 \\ f(4) &= F(4) - F(3) = 0,9 - 0,5 = 0,4 \\ f(5) &= F(5) - F(4) = 1 - 0,9 = 0,1 \end{aligned}$$

Això ens permet interpretar que el 10 % dels treballadors té els certificats de nivell 1, 2 i 5, respectivament, que el 30 % té el nivell 3 i que el 40 % té el nivell 4.

Potser és útil expressar la funció de probabilitats com a taula:

x_i	1	2	3	4	5
$f(x_i)$	0,1	0,1	0,3	0,4	0,1

Per a calcular l'esperança: comencem pel càlcul de l'esperança: $E[X] = \sum_{i=1}^n x_i \cdot f(x_i)$.

$E[X] = 1 \cdot 0,1 + 2 \cdot 0,1 + 3 \cdot 0,3 + 4 \cdot 0,4 + 5 \cdot 0,1 = 3,3$ nivell mitjà de coneixement dels treballadors de l'empresa.

Per a calcular la variància, utilitzarem la propietat que es pot veure al capítol de la teoria: $\text{Var}\{X\} = E[X^2] - E[X]^2$.

$$\text{Var}\{X\} = E[X^2] - E[X]^2 = 12,1 - 3,3^2 = 1,21 \text{ nivell}^2$$

$$E[X^2] = \sum_{i=1}^n x_i^2 f(x_i) =$$

$$1^2 \cdot 0,1 + 2^2 \cdot 0,1 + 3^2 \cdot 0,3 + 4^2 \cdot 0,4 + 5^2 \cdot 0,1 = 12,1$$

Si volem traure la *desviació típica*, escriurem $\sigma = \sqrt{\text{Var}[X]} = \sqrt{1,21} = 1,1$ nivell, que és la mitjana de la diferència (en valor absolut) entre el nivell certificat de llengua d'un treballador qualsevol i el nivell mitjà de l'empresa (esperança).

b) Per a calcular la probabilitat de que si triem un treballador a l'atzar d'aquesta empresa tinga:

i. Un nivell 3:

$$P(X = 3) = f(3) = 0,3$$

ii. Un nivell superior a 2:

$$P(X > 2) = f(3) + f(4) + f(5) = 1 - [f(1) + f(2)] = 1 - (0,1 + 0,1) = 0,8$$

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F(2) = 1 - 0,2 = 0,8$$

iii. Un nivell igual o superior a 2:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - F(1) = 1 - 0,1 = 0,9$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - f(1) = 1 - 0,1 = 0,9$$

iv. Un nivell inferior a 3:

$$P(X < 3) = f(1) + f(2) = 0,1 + 0,1 = 0,2$$

$$P(X < 3) = F(2) = 0,2$$

v. Un nivell major de 2, però igual o menys que 4:

$$P(2 < X \leq 4) = F(4) - F(2) = 0,9 - 0,2 = 0,7$$

$$P(2 < X \leq 4) = f(3) + f(4) = 0,3 + 0,4 = 0,7$$

c) Per a calcular el nivell mínim que podem assegurar que té el 60 % dels treballadors per a oferir els nostres productes a l'exterior, expressarem aquesta frase en termes de probabilitat recordant el concepte de percentil. Si x_0 és el valor de la variable (nivell de llengua) que ens demanen, llavors:

$$P(X \geq x_0) = 0,60 \Leftrightarrow P(X \leq x_0) = 0,40$$

és a dir, cal trobar el percentil P_{40} d'aquesta distribució. Si utilitzem la funció de distribució, cal trobar $F(x_0) = 0,40$.

Mirem a l'enunciat els valors de la funció $F(x)$: $F(2) = 0,2$; $F(3) = 0,5$; ... Per la qual cosa direm que $x_0 = 3$ i podem concloure que el 60 % dels treballadors tenen un nivell de la llengua superior a 3.

- d) Per a calcular el nivell màxim que podríem assegurar que té el 40 % dels treballadors per a demanar un pla de formació, farem un raonament semblant al de l'apartat anterior: $P(X \leq x_0) = 0,40$, així podrem veure que la solució és la mateixa, únicament canvia la matisació de la conclusió: el 40 % dels treballadors tenen un nivell inferior o igual a 3.

Exercici 4

Imagina que llancem tres monedes a l'aire i comptem el nombre de cares que han eixit.

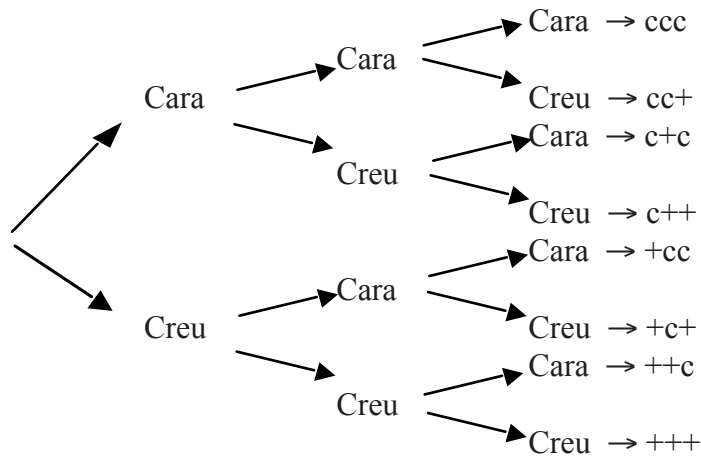
- Calcula la probabilitat de cadascun dels possibles valors de la variable X i defineix la funció de probabilitat corresponent.
- Calcula la probabilitat d'obtenir 2 cares, almenys 2 cares, i més de 2 cares en un llançament.
- Calcula la probabilitat d'obtenir un nombre de cares entre 0 i 2, ambdós inclosos.
- Calcula la funció de distribució, expressa-la i representa-la.
- Calcula el nombre de cares que s'obté com a màxim en el 75 % dels llançaments.
- Calcula l'esperança i la variància i interpreta-les.

Solució

Imagina que llancem tres monedes a l'aire i comptem el nombre de cares que han eixit.

- Calcula la probabilitat de cadascun dels possibles valors de la variable X i defineix la funció de probabilitat corresponent.

Per a trobar l'espai mostral ens ajudarem del diagrama d'arbre que farem a continuació i adjudicarem les probabilitats dels possibles valors de la variable, la qual definirem com $X =$ nombre de cares obtingudes en el llançament de tres monedes.



$E = \{ccc, cc+, c+c, c++, +cc, +c+, ++c, +++\}$.

$X =$ nombre de cares obtingudes en el llançament de tres monedes.

x_i	0	1	2	3
$f(x_i) = P(X = x_i)$	1/8	3/8	3/8	1/8

En aquesta taula, hem definit la *funció de probabilitat* de la variable X , que és una variable discreta.

b) Per a calcular la probabilitat d'obtenir 2 cares, almenys 2 cares, i més de 2 cares en un llançament, utilitzarem la funció anterior:

$$P(X = 2) = f(2) = 3/8$$

$$P(X \geq 2) = f(2) + f(3) = 3/8 + 1/8 = 4/8 = 1/2$$

$$P(X > 2) = f(3) = 1/8$$

c) Per a calcular la probabilitat d'obtenir un nombre de cares entre 0 i 2, ambdós inclosos:

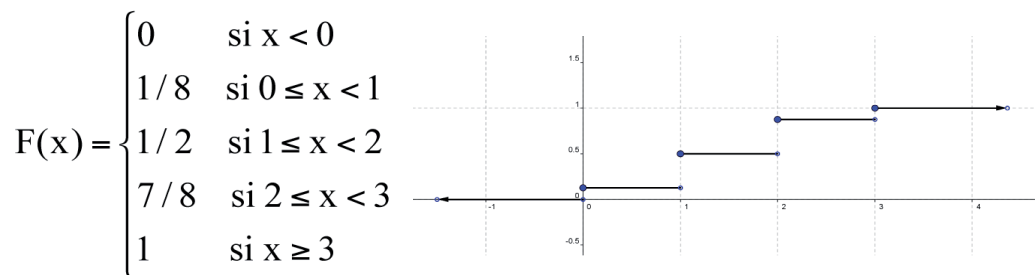
$$P(0 \leq X \leq 2) = f(0) + f(1) + f(2) = 1/8 + 3/8 + 3/8 = 7/8$$

d) Calcula la funció de distribució, expressa-la i representa-la.

Per a calcular la funció de distribució, $F(x) = P(X \leq x)$, ens ajudarem de la tercera fila de la taula següent, on acumulem les probabilitats dels diferents valors de la variable:

x_i	0	1	2	3
$f(x_i) = P(X = x_i)$	1/8	3/8	3/8	1/8
$F(x_i)$	1/8	4/8=1/2	7/8	1

Però, com que la funció de distribució ha de ser definida per a qualsevol valor $x \in \mathbb{R}$, la definirem com una «funció d'escala» definida a trossos:



e) Calcula el nombre de cares que s'obté com a màxim en el 75 % dels llançaments.

Per a respondre a aquesta qüestió, caldrà calcular els valors de les fraccions en decimals i decidir el que corresponga per trobar el valor de la variable que ens demanen:

x_i	0	1	2	3
$f(x_i) = P(X = x_i)$	1/8	3/8	3/8	1/8
$F(x_i)$	1/8 = 0,125	1/2 = 0,5	7/8 = 0,875	1

Podem dir que el valor que ens demanen és 2, perquè en el 87,5 % de les ocasions eixiran, com a molt, 2 cares, però en el 50 % dels casos s'obté com a màxim 1 cara.

f) Per a calcular l'esperança i la variància, vegem les operacions:

$$E[X] = \sum_i x_i \cdot f(x_i) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2} = 1,5 \text{ cares}$$

$$\text{Var}[X] = E[X^2] - E[X]^2 = 3 - 1,5^2 = 0,75$$

↓

$$E[X^2] = \sum_i x_i^2 \cdot f(x_i) = 0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8} = 3 \text{ cares}^2$$

$$\sigma = \sqrt{0,75} = 0,87 \text{ cares}$$

La mitjana de cares que s'obtenen en llançar tres monedes és de 1,5 cares, amb una desviació típica de 0,87.

Exercici 5

S'està controlant el ritme de treball que porta un servei d'atenció al públic. S'estima que es poden atendre 12 persones cada hora. Del recompte que hem fet amb les dades del mes passat, hem obtingut la següent funció de distribució:

$$F(x) = \begin{cases} 0 & \text{si } x < 10 \\ 0,35 & \text{si } 10 \leq x < 11 \\ 0,55 & \text{si } 11 \leq x < 12 \\ 0,65 & \text{si } 12 \leq x < 13 \\ 0,85 & \text{si } 13 \leq x < 14 \\ 1 & \text{si } x \geq 14 \end{cases}$$

- Calcula la probabilitat de que en una hora siguin ateses 9 persones; menys de 12 persones; més de 11, però menys de 13.
- Calcula la funció de probabilitat de la variable X.
- Calcula les probabilitats de l'apartat a utilitzant la funció de probabilitats.
- Calcula el nombre mínim de persones que podem assegurar que han estat ateses el 80 % de les hores. Calcula'n la mediana.
- Calcula l'esperança i la variància d'aquesta distribució i interpreta-les.

Solució

- Per a respondre a les probabilitats que ens demanen amb la funció de distribució de l'enunciat caldrà fer especial atenció als extrems per a distingir si estan o no inclosos, ja que es tracta d'una variable discreta. Així, fixant-se en la funció:

$$F(x) = \begin{cases} 0 & \text{si } x < 10 \\ 0,35 & \text{si } 10 \leq x < 11 \\ 0,55 & \text{si } 11 \leq x < 12 \\ 0,65 & \text{si } 12 \leq x < 13 \\ 0,85 & \text{si } 13 \leq x < 14 \\ 1 & \text{si } x \geq 14 \end{cases}$$

$$P(X = 9) = 0$$

$$P(X > 13) = 1 - P(X \leq 13) = 1 - F(13) = 1 - 0,85 = 0,15$$

$$P(X < 12) = F(11) = 0,55$$

$$P(11 < X < 13) = F(12) - F(11) = 0,65 - 0,55 = 0,10$$

- Per calcular la funció de probabilitats serà suficient amb calcular els «increments» de la funció de distribució en cadascun dels valors de la variable. Per a això podem ajudar-nos de la taula següent:

x_i	10	11	12	13	14	
$F(x_i)$	0,35	0,55	0,65	0,85	1	
$f(x_i)$	0,35	0,20	0,10	0,20	0,15	$\sum f(x_i) = 1$

c) Calcula les probabilitats de l'apartat *a* utilitzant la funció de probabilitats.

$$P(X = 9) = f(9) = 0$$

$$P(X > 13) = f(14) = 0,15$$

$$P(X < 12) = f(10) + f(11) = 0,35 + 0,20 = 0,55$$

$$P(11 < X < 13) = f(12) = 0,10$$

d) Calcula el nombre mínim de persones que podem assegurar que han estat ateses el 80 % de les hores. Calcula'n la mediana.

Per a calcular el P_{80} , cal trobar el valor de la variable del qual podem afirmar que el 80 % dels valors són menors que ell, per la qual cosa ens basarem en la $F(x)$. Vegem les cel·les de la taula anterior que corresponen als diferents valors de la variable i la seua interpretació:

$F(12) = 0,65 \rightarrow$ En el 65 % de les hores s'han atés 12 persones o menys.

$F(13) = 0,85 \rightarrow$ En el 85 % de de les hores s'han atés 13 persones o menys.

Però l'enunciat ens indica «almenys», per la qual cosa caldrà trobar el P_{20} :

$$P(X \geq x_0) = 1 - P(X \leq x_0) = 0,80 \rightarrow P(X \leq x_0) = 1 - 0,80 = 0,20$$

I el valor més proper en la taula que es correspon és:

$F(10) = 0,35 \rightarrow$ En el 35 % de les hores s'han atés com a molt 10 persones, del que podem deduir que en el 65 % almenys 11 persones se n'han atés.

$F(9) = 0 \rightarrow$ En el 0 % de les hores s'han atés 9 persones o menys, del que podem deduir que en el 100 % de les hores almenys 10 persones se n'han atés.

Podem concloure, així, que en el 80 % de les hores s'atenen almenys 10 persones, que és el que ens pregunten a l'enunciat.

NOTA

En les qüestions que fan referència a percentils, cal fixar-se en la redacció dels enunciats i no tan sols en el percentatge que ens donen.

e) Calcula l'esperança i la variància d'aquesta distribució i interpreta-les.

$$\bullet E[X] = \sum_i x_i \cdot f(x_i) =$$

$$10 \cdot 0,35 + 11 \cdot 0,20 + 12 \cdot 0,10 + 13 \cdot 0,20 + 14 \cdot 0,15 = 11,6 \text{ persones.}$$

- $\text{Var}[X] = E[X^2] - E[X]^2 = 3 - 1,5^2 = 0,75 \rightarrow$

$$E[X]^2 = \sum_i x_i^2 \cdot f(x_i)$$

$$10 \cdot 0,35^2 + 11 \cdot 0,20^2 + 12 \cdot 0,10^2 + 13 \cdot 0,20^2 + 14 \cdot 0,15^2 = 11,6 \text{ persones}^2$$

- $\sigma = \sqrt{2,62} = 1,62 \text{ persones}$

La interpretació que podem fer-ne és que la mitjana de les persones ateses per hora és de 11,6, amb una variabilitat mitjana de 1,62 persones. Aquesta segona dada es pot entendre com la mitjana de les desviacions del nombre de persones ateses cada hora, tant per excés com per defecte, sobre 11,6 persones, que és la mitjana global.

Exercici 6

En aquest moment, el percentatge d'atur entre els llicenciats menors de 29 anys a Espanya és del 23 %¹. Considerem un grup de 5 joves llicenciats menors de 29 anys.

- Calcula la probabilitat de que 3 d'ells estiguen aturats.
- Calcula la probabilitat de que més de 2 joves estiguen treballant.
- Troba la funció de probabilitat de cada valor possible de la variable $X =$ nombre de joves treballant en el grup de 5 llicenciats de l'enunciat.
- Troba el model de probabilitat que podem ajustar a aquest plantejament.
- Escriu l'expressió algebraica d'aquesta funció per a un grup de n joves.
- Si ara considerem un grup de 20 joves ($n = 20$), calcula l'esperança i la variància.

Solució

- Per a plantejar-nos el problema definim la variable $X =$ nombre de joves aturats.

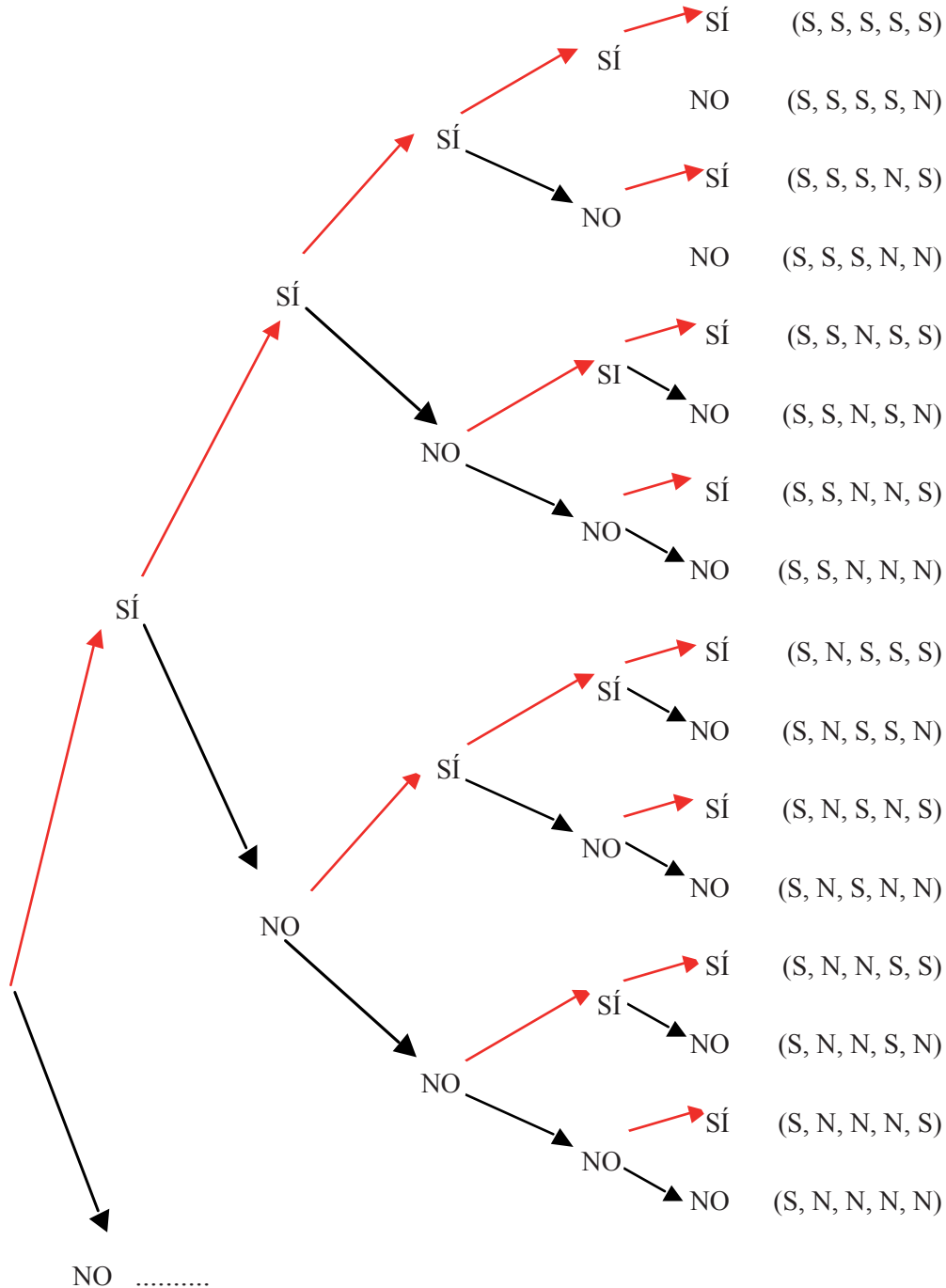
És fàcil comprendre que, si considerem tan sols si un jove està aturat o no, es pot considerar un model de Bernouilli, on $p =$ probabilitat d'estar aturat $= 0,23$.

Quan considerem un grup de 5 joves i cadascun d'ells pot estar o no aturat, passem a considerar un model binomial de paràmetres $n = 5$ i $p = 0,23$.

Si tenim clar el significat de la funció de distribució, podem utilitzar directament l'expressió algebraica de la teoria, on tal sols cal identificar els valors de n i p :

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, 3, \dots, n \\ 0 & \text{la resta} \end{cases}$$

Ara bé, podem fer un bon esquema que ens ajudarà a identificar aquest model i facilitarà la identificació dels problemes amb aquest model, sempre i quant es puga abordar com una disjuntiva (èxit o fracàs) per a cada element de la selecció. Suposem que davant cada individu, ens podem preguntar: està a l'atur? I com a resposta, es evident que tan sols tenim dues possibilitats: SÍ o NO i a més a més, sabem que $P(SÍ) = 0.23$ i $P(NO) = 0.77$.



Aquí tan sols hem fet la meitat del diagrama d'arbre; vos proposem que comproveu els càlculs següents fent l'altra meitat la que començaria amb un NO en la primera proposta.

Per a calcular les probabilitats cal que fem un recompte de tots els casos que corresponen a cada valor de X que necessitem obtenir. En aquest diagrama podrem trobar totes les possibilitats:

- Si $X = 0$ aturats:
 $\{N, N, N, N, N\}$
 $f(0) = 1 \cdot 0,77^5 = 0,2707$
- Si $X = 1$ aturat:
 $\{S, N, N, N, N\}, \{N, S, N, N, N\}, \{N, N, S, N, N\}, \{N, N, N, S, N\},$
 $\{N, N, N, N, S\}$
 $f(1) = 5 \cdot 0,23 \cdot 0,77^4 = 0,4043$
- Si $X = 2$:
 $\{S, S, N, N, N\}, \{S, N, S, N, N\}, \{S, N, N, S, N\}, \{S, N, N, N, S\},$
 $\{N, S, S, N, N\}, \{N, S, N, S, N\}, \{N, S, N, N, S\}, \{N, N, S, S, N\},$
 $\{N, N, S, N, S\}, \{N, N, N, S, S\}$
 $f(2) = 10 \cdot 0,23^2 \cdot 0,77^3 = 0,2415$
- Si $X = 3$:
 $\{S, S, S, N, N\}, \{S, S, N, S, N\}, \{S, S, N, N, S\}, \{N, S, S, S, N\},$
 $\{N, S, S, N, S\}, \{N, N, S, S, S\}, \{S, N, S, S, N\}, \{N, S, N, S, S\},$
 $\{S, N, N, S, S\}, \{S, N, S, N, S\}$
 $f(3) = 10 \cdot 0,23^3 \cdot 0,77^2 = 0,0721$
- Si $X = 4$:
 $\{N, S, S, S, S\}, \{S, N, S, S, S\}, \{S, S, N, S, S\}, \{S, S, S, N, S\},$
 $\{S, S, S, S, N\}$
 $f(4) = 5 \cdot 0,23^4 \cdot 0,77 = 0,0108$
- Si $X = 5$:
 $\{S, S, S, S, S\}$
 $f(5) = 1 \cdot 0,23^5 = 0,0006$

Així, hem calculat la funció de probabilitats que ens demanen a l'apartat *c*.

x_i	0	1	2	3	4	5
$f(x_i)$	0,2707	0,4043	0,2415	0,0721	0,0108	0,0006

Això ens permetrà respondre amb facilitat els apartats anteriors:

- a) Per a calcular la probabilitat de que 3 d'ells estiguen aturats:
 $P(X = 3) = f(3) = 0,0721$
- b) Per a calcular la probabilitat de que estiguen treballant més de 2 joves:
 $P(X > 2) = f(3) + f(4) + f(5) = 0,0721 + 0,0108 + 0,0006 = 0,0835$

c) La funció de probabilitats ja l'hem feta per anticipat.

d) Troba el model de probabilitat que podem ajustar a aquest plantejament.

Aquest problema compleix les característiques teòriques del model binomial amb $n = 5$ intents (cada persona pot estar o no aturada) i el paràmetre $p =$ probabilitat d'estar aturat $= 0,23$.

e) La funció de probabilitat per a un grup de n joves respon a l'expressió algebraica que podem trobar al resum teòric:

$$f(x) = \begin{cases} \binom{n}{x} 0,23^x (1-0,23)^{n-x} & \text{si } x = 0, 1, 2, 3, \dots, n \\ 0 & \text{la resta} \end{cases}$$

On hem substituït $p = 0,23$.

f) Si ara considerem un grup de 20 joves, per a calcular l'esperança i la variància canviarem al model binomial de paràmetres $n = 20$ i $p = 0,23$.

Per calcular-les tan sols necessitem identificar els valors dels paràmetres del model. Així:

$E[X] = n \cdot p = 20 \cdot 0,23 = 4,6$, aquesta és la mitjana de joves aturats en un grup de 20 llicenciats.

$\text{Var}[X] = n \cdot p \cdot (1 - p) = 20 \cdot 0,23 \cdot 0,77 = 3,542$

Exercici 7

Sabem que a la nostra empresa el nombre de treballadors que desenvolupen tasques directives és del 15 %. Considerem que en cada departament hi ha 10 treballadors.

- Calcula la probabilitat que en un departament qualsevol hi haja 2 directius; menys de 4; més de 6; 4 com a molt.
- Calcula el nombre mitjà de directius en els departaments i la desviació típica corresponent. Interpreta els resultats.
- Quants directius podem assegurar que tenen com a molt el 80 % dels departaments?
- Calcula la probabilitat de que hi haja 8 treballadors no directius.

Imaginem ara que els departaments tingueren 50 treballadors.

- Calcula la probabilitat de que hi haja 15 directius.

Solució

Per a plantejar el problema començarem per definir la variable a estudiar: $X =$ nre. de directius en un departament de 10 treballadors.

Es tracta d'una variable discreta i podem deduir que es distribuirà ajustant-se a un model binomial de paràmetres $n = 10$ treballadors (que corresponen a les 10 proves de Bernouilli independents) i $p = 0,15$ (que correspon a la probabilitat d'èxit = probabilitat de que un treballador triat a l'atzar siga directiu). Així:

$$X \rightarrow \text{Bi}(n = 10, p = 0,15)$$

- a) Calcula la probabilitat de que en un departament qualsevol hi haja 2 directius; menys de 4; més de 6; 4 com a molt.

Per a treballar amb variables que corresponen a models coneguts podem treballar amb l'expressió algebraica de la funció de probabilitats que apareix al resum de teoria o crear taules on estan els valors ja calculats per als valors dels paràmetres més freqüents. Així, substituïm $n = 10$ i $p = 0,15$ en l'expressió general i obtenim la probabilitat $P(X = x) = f(x)$ per a cada valor de x .

$$f(x) = \begin{cases} \binom{10}{x} 0,15^x (1-0,15)^{10-x} & \text{si } x = 0, 1, 2, 3, \dots, 10 \\ 0 & \text{la resta} \end{cases}$$

Per a calcular probabilitats acumulades podem ajudar-nos de les taules o d'un paquet estadístic. A continuació hem inserit un fragment d'una taula de la distribució binomial on podrem consultar els valors que necessitem:

$$P(X = 2) = \binom{10}{2} 0,15^2 (1-0,15)^8 = 45 \cdot 0,15^2 \cdot 0,85^8 = 0,2759$$

Tabla I Distribución B

n	p	.01	.05	.10	.15	.20	.25
9	0	.9135	.6302	.3874	.2316	.1342	.0751
	1	.0830	.2985	.3874	.3679	.3020	.2253
	2	.0034	.0629	.1722	.2597	.3020	.3003
	3	.0001	.0077	.0446	.1069	.1762	.2336
	4	.0000	.0006	.0074	.0283	.0661	.1168
	5	.0000	.0000	.0008	.0050	.0165	.0389
	6	.0000	.0000	.0001	.0006	.0028	.0087
	7	.0000	.0000	.0000	.0000	.0003	.0012
	8	.0000	.0000	.0000	.0000	.0000	.0001
	9	.0000	.0000	.0000	.0000	.0000	.0000
10	0	.9044	.5987	.3487	.1969	.1074	.0563
	1	.0914	.3151	.3874	.3474	.2684	.1877
	2	.0042	.0746	.1937	.2759	.3020	.2816
	3	.0001	.0105	.0574	.1298	.2013	.2503
	4	.0000	.0010	.0112	.0401	.0881	.1460
	5	.0000	.0001	.0015	.0085	.0264	.0584
	6	.0000	.0000	.0001	.0012	.0055	.0162
	7	.0000	.0000	.0000	.0001	.0008	.0031
	8	.0000	.0000	.0000	.0000	.0001	.0004
	9	.0000	.0000	.0000	.0000	.0000	.0000
10	.0000	.0000	.0000	.0000	.0000	.0000	

$$P(X \leq 4) = f(0) + f(1) + f(2) + f(3) + f(4) = 0,1969 + 0,3474 + 0,2759 + 0,1298 + 0,0401 = 0,990$$

També podem valorar si és més convenient, segons els casos, abordar les probabilitats del complementari:

$$P(X \leq 4) = 1 - P(X > 4)$$

$$P(X < 4) = f(0) + f(1) + f(2) + f(3) = 0,1969 + 0,3474 + 0,2759 + 0,1298 = 0,95$$

$$P(X > 6) = f(7) + f(8) + f(9) + f(10) = 0,0001$$

Quan obtenim valors tan propers a 0, podem interpretar que la proposta té probabilitat quasi nul·la, si més no és gairebé impossible.

b) Per a calcular el nombre mitjà de directius en els departaments i la desviació típica corresponent, calcularem l'esperança i la variància de la distribució. Ara bé, per tractar-se del model binomial, podem fer ús de la teoria i serà suficient substituir els valors dels paràmetres en les expressions corresponents:

$$E[X] = np = 10 \cdot 0,15 = 1,5 \text{ directius.}$$

$$\text{Var}[X] = npq = 10 \cdot 0,15 \cdot 0,85 = 1,275 \text{ directius}^2$$

$$\sigma = \sqrt{\text{Var}} = \sqrt{1,275} = 1,13 \text{ directius}$$

Podem interpretar que la mitjana del nombre de directius en aquests departaments (de 10 treballadors) és de 1,5.

La desviació mitjana ens indica la mitjana, en valor absolut, de les diferències entre el nombre de directius de cada departament i l'anterior mitjana (1,5 treballadors), que és de 1,13 directius.

c) Per a esbrinar quants directius podem assegurar que té com a molt el 80 % dels departaments, relacionem la qüestió amb el càlcul de percentils. Cal calcular la funció de distribució $F(x) = P(X \leq x)$ per als diferents valors de la variable fins a trobar el primer valor de la variable x per al qual se supere $F(x) = 0,80$.

$$F(0) = f(0) = 0,1969$$

$$F(1) = f(0) + f(1) = 0,1969 + 0,3474 = 0,5443$$

$$F(2) = f(0) + f(1) + f(2) = 0,1969 + 0,3474 + 0,2759 = 0,8202$$

Del darrer valor que hem obtingut es dedueix que en el 82 % dels departaments hi ha com a molt 2 directius.

Així, podem concloure i assegurar que el 80 % dels departaments tindrà 2 directius o menys.

d) Per a calcular la probabilitat que hi haja 8 treballadors no directius, tan sols cal relacionar que aquesta circumstància siga equivalent a calcular la probabilitat de que hi haja 2 directius, la qual cosa ja hem calculat en l'apartat *a*.

$$P(X = 2) = f(2) = 0,2759$$

Imaginem ara que els departaments tingueren 50 treballadors.

e) Calcula la probabilitat de que hi haja 15 directius.

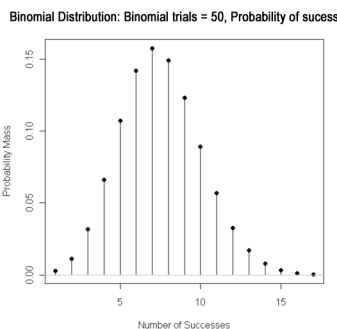
Si els departaments tingueren 50 treballadors caldria canviar el paràmetre $n = 50$ en el model binomial. $X = nre.$ de directius en els departaments de 50 treballadors $\rightarrow Bi(n = 50, p = 0,15)$.

$$P(X = 15) = \binom{50}{15} 0,15^{15} (1 - 0,15)^{35} = 2,2508 \cdot 10^{12} \cdot 0,15^{15} \cdot 0,85^{35} = 0,0033$$

Si ens demanaren alguna probabilitat, com ara $P(X \leq 20)$, caldria ajudar-nos d'un paquet estadístic per comoditat de càlcul, ja que aquests sumatoris ja resultarien una tasca farragosa.

Altres possibilitats que hem desenvolupat als textos teòrics són les transformacions del model binomial al Poisson o al normal quan els paràmetres compleixen les condicions que ens indiquen els teoremes de cada cas.

A continuació podem veure la gràfica de la funció de probabilitats del model binomial de $n = 50$ i $p = 0,15$.



Exercici 8

En el servei de teleassistència de 24 hores d'un centre de benestar social comarcal, s'han comptabilitzat les demandes dels usuaris, que de forma aleatòria han estat de 5 incidències per hora, al llarg de tot el dia.

- Calcula la probabilitat de que es reben 3, 4, 5 i 6 incidències en una hora.
- Calcula la mitjana d'incidències en un dia.
- Calcula el nombre d'incidències que s'atenen, com a molt, en el 80 % de les hores.

Solució

Per a plantejar el problema començarem per definir la variable aleatòria. $X =$ nre. d'incidències cada hora \rightarrow $Po(\lambda = 5)$.

- Per a calcular la probabilitat de que es reben 3, 4, 5 i 6 incidències en una hora cal substituir el valor de λ en l'expressió algebraica de la funció de probabilitats o trobar-ne els valors en alguna taula:

$$\text{Considerem que } P(X = x) = f(x) = \frac{\mu^x \cdot e^{-\mu}}{x!} = \frac{5^x \cdot e^{-5}}{x!}$$

$$P(X = 3) = f(3) = 0,1404$$

$$P(X = 4) = f(4) = 0,1755$$

$$P(X = 5) = f(5) = 0,1755$$

$$P(X = 6) = f(6) = 0,1462$$

Si consultem en alguna taula de les que estan publicades, podrem trobar els valors següents per als valors de $X = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ i 10 .

5.0	.0067	.0337	.0842	.1404	.1755	.1755	.1462	.1044	.0653	.0363	.0181	.0082	.0034
-----	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

- b) Per a calcular la mitjana d'incidències en un dia, considerarem que l'esperança és $E[Y] = \lambda = 120$, ja que $\lambda = 5 \cdot 24 = 120$, perquè caldrà considerar l'adaptació de la mitjana proporcionalment a l'interval d'un dia.
- c) Per a calcular el nombre d'incidències que s'atenen, com a molt, en el 80 % dels dies, necessitem trobar els valors de la funció de distribució: $F(y) = P(Y \leq y)$, acumulant els valors de la funció de probabilitats. Necessitem calcular $P(Y \leq a) = 0,80$.

$$F(6) = 0,616$$

$$F(7) = 0,7622$$

$$F(8) = 0,8666$$

Podem interpretar, dels resultats obtinguts, que en el 80 % de les hores s'atenen 8 incidències com a molt.

Exercici 9

Estem preparant unes oposicions que tenen 50 temes a l'apartat teòric i sóc conscient de que en aquests moments en podria defensar amb èxit 30. En la convocatòria, s'explicita que el tribunal extraurà 5 boles, cadascuna corresponent a una unitat del temari. Es tria un d'aquests temes i se'n fa el desenvolupament teòric. Calcula les següents probabilitats:

- Podria defensar amb èxit tan sols un tema de les boles extretes.
- Podria defensar amb èxit com a molt, dos temes de les boles extretes.
- Podria defensar amb èxit els cinc temes que han eixit.
- No podria defensar cap tema.
- Calcula la mitjana de temes que podré defensar en cada extracció de boles i la seua variabilitat.

Solució

Per a plantejar el problema començarem per definir la variable aleatòria: $X =$ nre. d'unitats que puc defensar amb èxit entre els temes que ixen en agafar 5 boles a l'atzar que corresponen a 50 unitats que hi ha al temari.

Podem identificar l'experiment com una variable hipergeomètrica, els paràmetres de la qual són:

$N = 50$ temes que hi ha a l'oposició i que es corresponen amb 50 boles en total.

$n = 5$ temes que s'extrauen sense reposició i que es corresponen amb 5 temes possibles a desenvolupar triats a l'atzar.

$K = 30$ temes que conec i puc defensar amb èxit entre els 50 que hi ha.

Així: $X \rightarrow$ Hiper ($N = 50, n = 5, K = 30$).

a) Per a calcular la probabilitat de conèixer tan sols un tema, calculem:

$$P(X = 1) = f(1) = \frac{\binom{30}{1} \cdot \binom{20}{4}}{\binom{50}{5}} = \frac{30 \cdot 4845}{2118760} = 0,06860$$

b) Per a calcular la probabilitat de que puga defensar amb èxit, com a molt, dos temes de les boles extretes:

$$P(X \leq 2) = f(0) + f(1) + f(2) = 0,30997$$

Considerem:

$$f(0) = \frac{\binom{30}{0} \cdot \binom{20}{5}}{\binom{50}{5}} = \frac{1 \cdot 15504}{2118760} = 0,00732$$

$$f(1) = \frac{\binom{30}{1} \cdot \binom{20}{4}}{\binom{50}{5}} = \frac{30 \cdot 4845}{2118760} = 0,06860$$

$$f(2) = \frac{\binom{30}{2} \cdot \binom{20}{3}}{\binom{50}{5}} = \frac{435 \cdot 1140}{2118760} = 0,23405$$

c) Per a calcular la probabilitat de que puga defensar amb èxit els temes que corresponen a les 5 boles extretes:

$$P(X = 5) = f(5) = \frac{\binom{30}{5} \cdot \binom{20}{0}}{\binom{50}{5}} = \frac{142506 \cdot 1}{2118760} = 0,06726$$

- d) Per a calcular la probabilitat de que no puga defensar amb èxit cap tema es correspon amb la probabilitat de que cap bola extreta corresponga acap dels 30 temes que se sap l'aspirant. Així:

$$P(X = 0 = f(0)) = \frac{\binom{30}{0} \cdot \binom{20}{5}}{\binom{50}{5}} = \frac{1 \cdot 15504}{2118760} = 0,00732$$

- e) Per a calcular la mitjana de temes que podré defensar en cada extracció de boles i la seua variabilitat, consultarem un text teòric en què podrem trobar que: l'esperança és $E[X] = n \frac{K}{N} = 5 \frac{30}{50} = 3$ temes que podré defensar de mitjana. I la seua variància és $Var[X] = n \frac{K}{N} \left(1 - \frac{K}{N}\right) \frac{N-n}{N-1} = 5 \frac{30}{50} \left(1 - \frac{30}{50}\right) \frac{45}{49} = 1,1020$ temes.²

Exercici 10

A la premsa del passat mes d'octubre de 2013 es van publicar aquestes dades: «un 10 % de les llars espanyoles tenen tots els seus membres a l'atur».² Si en una aula de primària d'un col·legi públic tenim 30 alumnes (suposem que cadascun pertany a una llar diferent), es podria considerar una mostra representativa de la societat, i així, calcula:

- La mitjana d'alumnes d'aquesta classe les famílies dels quals pateixen aquesta situació.
- La probabilitat de necessitar entre 2 i 5 beques de menjador en aquesta classe per als alumnes que estan en aquesta situació.

Solució

Per a plantejar el problema començarem per definir la variable aleatòria: $X = nre.$ de llars que tenen tots els seus membres a l'atur o alumnes que pateixen a casa aquesta situació, suposant que cada alumne pertany a una llar diferent i que a classe hi ha 30 alumnes.

Aquesta descripció i el coneixement de la dada de que la probabilitat de que una llar triada a l'atzar tinga tots els seus membres a l'atur és del 10 %, ens fan considerar que la nostra variable s'ajusta a un model de probabilitat binomial amb paràmetres $n = 30$ i $p = 0,10$.

2. <http://www.rtve.es/noticias/20121026/10-hogares-espanoles-tiene-todos-miembros-paro/571737.shtml>

Aquests valors dels paràmetres ens permeten veure que es poden fer els càlculs mitjançant l'aproximació que permet calcular una funció de probabilitat d'un model binomial (n, p) amb la funció de probabilitat del model Poisson, on el seu paràmetre $\lambda = n \cdot p$.

Cal veure quines condicions són aconsellables i compleixen els paràmetres n i p de la binomial perquè es pugui fer aquesta aproximació.

- a) Per a calcular la mitjana, recordarem que al model binomial $E[X] = n \cdot p = 30 \cdot 0,10 = 3$, que alhora és el paràmetre $\lambda = 3$ de la distribució Poisson que utilitzarem a l'apartat següent.
- b) Per a calcular la probabilitat de necessitar entre 2 i 5 beques de menjador en aquesta classe per als alumnes que estan en aquesta situació abordarem el problema amb els dos models i comprovarem que l'aproximació és possible i les ventages del canvi a Poisson, sobretot si tenim alguna taula per consultar:

Si $X \rightarrow \text{Bi}(n = 30, p = 0,10)$

$$f(2) = \binom{30}{2} 0,10^2 \cdot 0,90^{28} = 0,2277$$

$$f(3) = \binom{30}{3} 0,10^3 \cdot 0,90^{27} = 0,2361$$

$$f(4) = \binom{30}{4} 0,10^4 \cdot 0,90^{26} = 0,1771$$

$$f(5) = \binom{30}{5} 0,10^5 \cdot 0,90^{25} = 0,1023$$

Si $X \rightarrow \text{Po}(\lambda = 3)$

$$f(2) = \frac{3^2 \cdot e^{-3}}{2!} = 0,2240$$

$$f(3) = \frac{3^3 \cdot e^{-3}}{3!} = 0,2240$$

$$f(4) = \frac{3^4 \cdot e^{-3}}{4!} = 0,1680$$

$$f(5) = \frac{3^5 \cdot e^{-3}}{5!} = 0,1008$$

Es pot comprovar que els resultats són propers, però aquesta aproximació és encara millor si són més exigents en els valors dels paràmetres n i p que permeten fer aquest canvi. (lleï dels successos rars).

Així, $P(2 \leq X \leq 5) = f(2) + f(3) + f(4) + f(5) = 0,7432$ (model binomial).

Així, $P(2 \leq X \leq 5) = f(2) + f(3) + f(4) + f(5) = 0,7168$ (model Poisson).

UNITAT 3

Variables aleatòries contínues. Models de probabilitat continus. Teorema central del límit

Introducció teòrica

Les *variables aleatòries contínues* poden prendre qualsevol valor dintre d'un rang donat. La variable «temps que tarda un alumne a realitzar un examen amb un límit de temps d'una hora» és contínua i els valors que pot prendre són tots aquells compresos entre 0 i 60 minuts. Així 30,252 minuts, 30,253 minuts o 30,2527 minuts són valors possibles que pot prendre la variable. Altres exemples són el pes de les persones matriculades en els gimnasos, l'edat exacta dels assistents habituals als congressos de matemàtiques o les tones que carreguen els camions d'una companyia de transport.

NOTA

Al conjunt de valors que pot prendre una variable aleatòria X se'l denomina rang de X .

Variables aleatòries contínues

Tal com s'ha comentat en el primer epígraf, una variable aleatòria és contínua si pot prendre qualsevol dels infinits valors compresos en un interval de la recta real. Per exemple, si es considera la variable X = altura d'una persona, X pot prendre qualsevol valor de l'interval $(1,85, 1,86)$ —els valors 1,8501; 1,80502; 1,8053....; 1,85001; 1,85002; 1,85011; 1,85012 en són uns exemples—. Com pot intuir-se amb facilitat, pareix que existixisquen infinits valors entre 1,85 i 1,86.

Conseqüentment, no té sentit calcular la probabilitat de que X prenga un valor concret, perquè n'hi ha infinits i aleshores la probabilitat és 0. Tanmateix, si que en té el càlcul de la probabilitat de que X prenga valors dins d'un interval $[a, b]$, és a dir, $P(X \in [a, b]) = P(a \leq X \leq b)$.

Ara bé, si en les distribucions discretes es definia la funció de probabilitat com la que *permetia el càlcul de la probabilitat de que X siga igual a un valor concret* $f(x) = P(X = x)$, en les variables aleatòries contínues no es pot definir una funció d'aquest tipus. Es defineix, en canvi, per a cada model, una funció denominada *funció de densitat de probabilitat*, $f(t)$, la qual *permet calcular la probabilitat de que la variable aleatòria prenga valors dins d'un interval*.

El concepte utilitzat per al càlcul de la probabilitat és el d'àrea. Concretament, es defineix la probabilitat de que X prenga valors dintre de l'interval $[a, b]$, com s' l'àrea compresa entre la funció de densitat, l'eix horitzontal de les abscisses i els punts a i b . Al gràfic (fig. 8), $P(X \in [0, 5, 1]) = \text{àrea ombrejada} = \int_{0,5}^1 f(t)dt$ (ja que en terminologia matemàtica l'àrea es representa mitjançant la integral).

funció de densitat

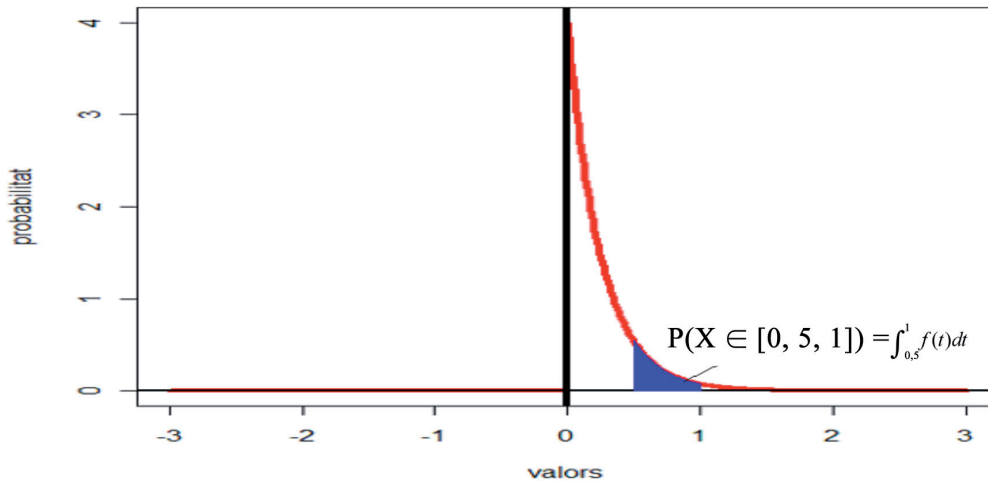


Fig. 8

NOTA

Cal recordar que, en les variables estadístiques en què s'agrupaven les dades en intervals, la representació gràfica de les dades es realitzava mitjançant els histogrames, les àrees dels quals representaven la freqüència de l'interval.

En les variables aleatòries contínues es representa la probabilitat emprant el mateix concepte d'àrea. En aquest cas, però, cal conèixer la funció de densitat de probabilitat.

Com s'ha comentat abans, amb la funció de densitat no és possible calcular directament la probabilitat, tanmateix sí que permet calcular-la.

D'altra banda, com que la funció de densitat permet determinar la distribució de probabilitats d'un experiment aleatori, és necessari que complisca amb els axiomes de probabilitat; els quals es tradueixen en:

- $f(x) \geq 0$ (la funció és positiva perquè l'àrea es positiva; fig. 9, esquerra).
- $\int_{-\infty}^{+\infty} f(t)dt = 1$ (l'àrea total ha de ser 1; fig. 9, dreta).

A més a més, si $A = [a, b]$, $P(X \in A) = P(X \in [a, b]) = P(a \leq X \leq b) = \int_a^b f(t)dt$

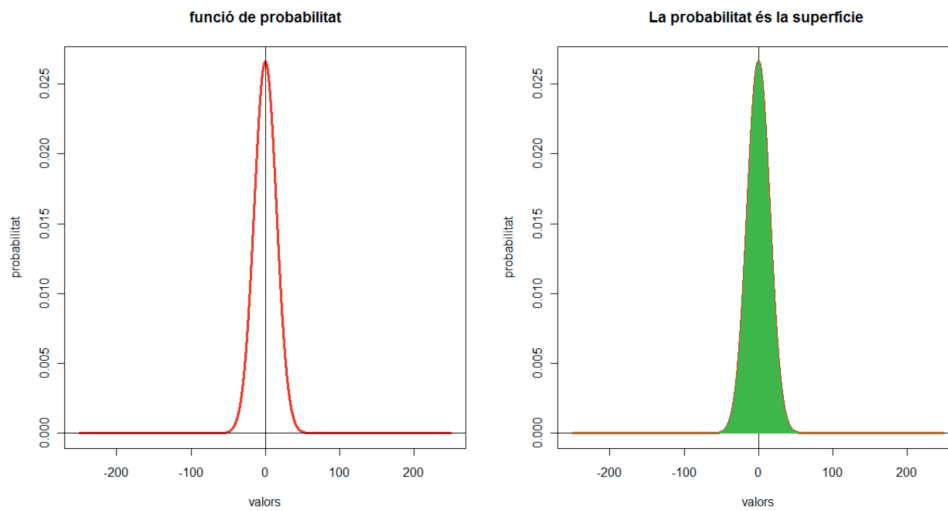


Fig. 9

Funció de distribució

La funció de distribució d'una variable aleatòria contínua té el mateix significat que la de la funció de la variable aleatòria discreta. El que varia és la manera de calcular la probabilitat demanada. Així, donada una distribució estadística discreta X en què a és un valor numèric, es defineix la funció de distribució $F(a)$ com la probabilitat de que la variable aleatòria X no prengui un valor superior a a . És a dir:

$$F(a) = P(X \leq a) = \int_{-\infty}^a f(t) dt$$

El gràfic següent (fig. 10) mostra la funció de densitat i la funció de distribució per a una distribució de probabilitat contínua. Com s'observa, la superfície marcada en el gràfic de la funció de densitat $P(X \leq 100)$ és el valor que li correspon a 100 en el gràfic de la funció de distribució.

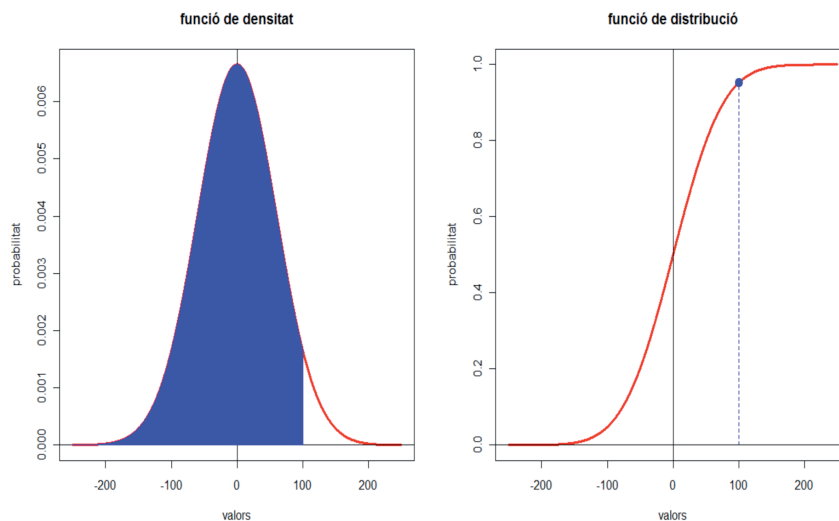


Fig. 10

Propietats

- La funció de distribució es caracteritza per ser creixent, és a dir, si $x_1 \geq x_2 \rightarrow F(x_1) \geq F(x_2)$.
- $F(-\infty) = P(X \leq -\infty) = 0$, ja que cap valor de la variable aleatòria pot ser menor que menys infinit.
- $F(+\infty) = P(X \leq +\infty) = 1$, ja que tots els valors de la variable aleatòria són menors que més infinit.

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx = \int_a^b f(x)dx$$

Les dues primeres propietats es poden comprovar intuïtivament en els gràfics anteriors. Aquesta darrera propietat és molt important i de molta utilitat a l'hora de fer càlculs de probabilitats. Els següents gràfics en faciliten la comprensió (fig. 11).

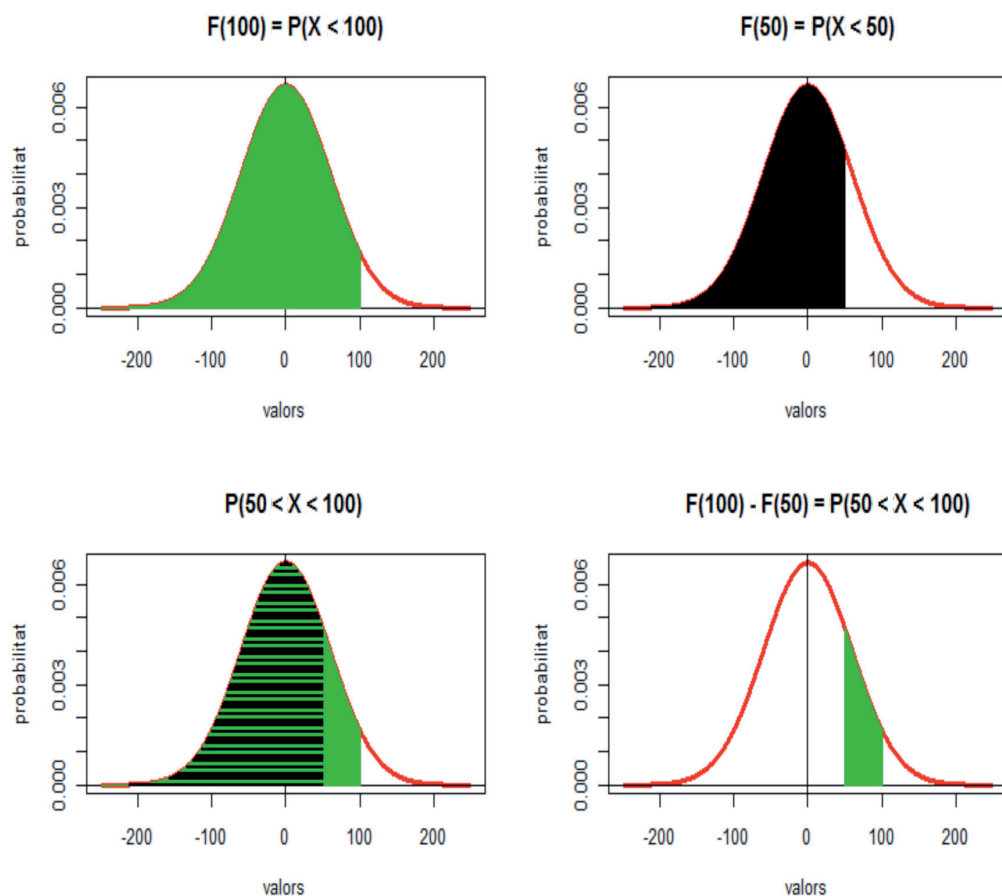


Fig. 11

Esperança i variància d'una variable aleatòria contínua

De la mateixa manera que ocorria amb les variables aleatòries discretes, es poden definir els estadístics d'esperança matemàtica i variància per a les variables contínues. Cal dir que malgrat tindre significats equivalents en tots dos tipus de variables aleatòries, en les variables contínues són en molts casos aquests paràmetres els que caracteritzen completament la distribució, fins al punt, de ser necessaris per a poder definir la funció de densitat de probabilitat.

Valor esperat o esperança matemàtica

Si X és una variable aleatòria contínua i $f(t)$ és la funció de quantia associada, l'esperança de X es calcula:

$$\mu = E(X) = \int_{-\infty}^{\infty} t \cdot f(t) dt$$

Variància

Si X és una variable aleatòria contínua i $f(t)$ la funció de quantia que hi té asso-

ciada, la variància de X es calcula: $Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (t - \mu)^2 f(t) dt$, o l'expressió equivalent: $Var(X) = E(X^2) - E(X)^2$.

Variables aleatòries bidimensionals contínues

També introduïrem les variables aleatòries bidimensionals contínues. Es considera (X, Y) una variable aleatòria bidimensional contínua que pren tots els valors possibles dintre d'una regió R de dues dimensions (R és un subconjunt de un interval de dues dimensions). Al conjunt format per tots els possibles valors que pot prendre la variable aleatòria se l'anomena recorregut o rang.

Pel mateix raonament que en les variables aleatòries unidimensionals, en les bidimensionals no té sentit parlar de la probabilitat de que la variable aleatòria prenga un valor concret (un parell concret (x_i, y_j)), de fet, aquesta probabilitat és 0. Per contra, és evident que hi haurà intervals bidimensionals més probables que altres i per això sorgeix la necessitat de definir una funció $f(x, y)$ que ho reflectisca. A aquesta funció se la denomina *funció de densitat de probabilitat conjunta* f , i permet calcular la probabilitat d'un interval bidimensional mitjançant el càlcul integral que, si en considerar una dimensió representa l'àrea, en considerar-ne dues representa el volum.

D'altra banda, aquesta funció ha de complir les exigències definides en els axiomes de probabilitat:

$$a) f(x, y) \geq 0 \text{ per a tots els parells } (x, y) \text{ de } R \text{ de } (X, Y)$$

$$b) \iint_R f(x, y) dx dy = 1$$

La segona condició significa que el volum total, el qual representa la probabilitat total, és 1.

Per una altra part, és possible construir el gràfic que representa la funció de densitat de probabilitat conjunta. Així, a cada parell (x, y) del recorregut se li associa a l'eix vertical una altura igual al valor $f(x, y)$. La representació gràfica és obviament una superfície, ja que tant X com Y són contínues. Noteu que el volum que encercla la superfície i el pla XY es calcula segons la integral doble abans esmentada i és 1 (segona exigència dels axiomes de probabilitat). A més a més, per a calcular la probabilitat d'un subconjunt A del recorregut cal trobar el volum encerclat entre A i la funció de densitat. Per a fer-ho s'ha de resoldre la integral doble sobre l'esmentat subconjunt:

$$P(A) = \iint_A f(x, y) dx dy$$

De la mateixa manera que en les variables discretes, la definició de variable aleatòria bidimensional contínua es pot generalitzar per a un nombre indeterminat de variables aleatòries contínues. Així, es pot parlar de variables aleatòries n -dimensionals (X_1, \dots, X_n) contínues i la seua funció de densitat de probabilitat conjunta $f(x_1, \dots, x_n)$, la qual pren valors majors en aquelles regions més probables i compleix els axiomes de probabilitat:

$$a) f(x_1, \dots, x_n) \geq 0 \text{ per a tots els parells } (x_1, \dots, x_n) \text{ de } R \text{ de } (X_1, \dots, X_n)$$

$$b) \int \dots \int f(x_1, \dots, x_n) dx_1 \dots dx_n = 1$$

On R és el rang de la variable aleatòria multidimensional contínua.

Exemple

Un equip d'investigació intenta conèixer la població d'una determinada regió, atès que han descobert que una gran part de la seua població té uns ingressos econòmics alts. Després de molt de temps d'investigació han arribat a la conclusió que les variables que influeixen en l'èxit econòmic són el coeficient intel·lectual i la cultura dels individus. Així és que han decidit estudiar aquestes dues variables en la població més exhaustivament. El primer que han és ou «matematitzar» les dues característiques que estaven considerant. És a dir, han aconseguit assignar a cada persona un nombre de l'interval $[0, 1]$ per a determinar el seu coeficient intel·lectual i un nombre de l'interval $[0, 2]$ per a determinar la seua cultura. Així

mateix, designant per X el coeficient intel·lectual i Y el grau de cultura, han hipotitzat una funció de densitat de probabilitat conjunta per a la variable (X, Y) . Aquesta funció és la següent:

$$f(x, y) = x^2 + \frac{x \cdot y}{3}$$

Variabes aleatòries bidimensionals independents

En el capítol de probabilitat es va introduir el concepte d'esdeveniments independents i el d'experiments independents. Així, es deia que dos successos són independents quan l'ocurrència d'un no influeix en l'ocurrència de l'altre. És a dir, A i B són dos successos independents si $P(A/B) = P(A)$, o segons el teorema de caracterització, A i B són dos successos independents si $P(A \cap B) = P(A) \cdot P(B)$.

És evident que aquest concepte d'independència també té la seua «versió» en les variables aleatòries, ja que aquestes no són res més que una altra manera d'expressar els esdeveniments. Així, si l'experiment consisteix a anotar d'una banda el nombre de trucades de telèfon que rep una centraleta al llarg de les 5 hores de feina del matí d'una companyia en Madrid –nombre que anomenarem X –, i de l'altra fer el mateix recompte però d'una centraleta en Barcelona –nombre que anomenarem Y –, les variables X i Y són intuïtivament independents. El nombre de telefonades que es reben en un lloc és independent del de les que es reben en l'altre. Com a conseqüència, la probabilitat de que en Madrid reben x_i trucades, sabent que en Barcelona es reben y_j trucades, és la probabilitat que en Madrid se'n reben x_i . Per tant, $P(X = x_i/Y = y_j) = P(X = x_i)$ per a qualsevol parell de valors (x_i, y_j) que prenen. Aquesta expressió també es pot escriure segons la definició de probabilitat condicionada com: $P(X = x_i, Y = y_j) = P(X = x_i \cap Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$. Més formalment: si (X, Y) és una *variable aleatòria bidimensional discreta*, es diu que X i Y són independents si $P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$ per a qualsevol parell (x_i, y_j) . Emprant una notació amb funcions de probabilitat, la igualtat de probabilitats es tradueix en:

$$p(x_i, y_j) = p_x(x_i) \cdot p_y(y_j)$$

On p és la funció de probabilitat conjunta i p_x i p_y les funcions de probabilitat marginals.

De manera anàloga, donada una *variable aleatòria bidimensional contínua* (X, Y) , es diu que X i Y són independents si $f(x, y) = F_x(x) \cdot f_y(y)$.

On f és la funció de densitat conjunta i f_x i f_y les funcions de densitat marginals.

Cal notar que aquestes definicions es poden generalitzar per a qualsevol nombre de variables aleatòries, així:

- Donada (X_1, \dots, X_n) una variable aleatòria n-dimensional discreta, les variables aleatòries X_1, \dots, X_n són independents si la funció de probabilitat conjunta és igual al producte de les funcions de probabilitat marginals:

$$p(x_1, \dots, x_n) = p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n)$$

- Anàlogament, donada (X_1, \dots, X_n) una variable aleatòria n-dimensional contínua, les variables aleatòries X_1, \dots, X_n són independents si la funció de densitat de probabilitat conjunta és igual al producte de les funcions de densitat de probabilitat marginals:

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$$

Cal dir que en les variables aleatòries bidimensionals es poden definir conceptes semblants als que es defineixen en les variables estadístiques bidimensionals, com són la covariància i la correlació lineal, per a conèixer el grau de relació existent entre les variables. No obstant, aquest estudi queda fora dels objectius d'aquest text.

Exemple

Un inversor té accions en dues empreses espanyoles. El rendiment de les accions en tant percent de cada empresa són dues variables aleatòries: X i Y. La distribució de probabilitat conjunta és la que apareix en la taula:

Rendiment de X	Rendiment de Y			
	5 %	8 %	12 %	
$p_X(x_i)$				
5 %	0,1	0,2	0,2	0,5
10 %	0,04	0,08	0,08	0,2
12 %	0,06	0,12	0,12	0,3
$p_Y(y_j)$	0,2	0,4	0,4	1

Són les variables X i Y independents?

Les variables són independents si per a tots els possibles valors de la variable aleatòria bidimensional (x_i, y_j) es compleix que $p(x_i, y_j) = p_X(x_i) \cdot p_Y(y_j)$:

- $p(5, 5) = 0,1$ i $p_X(5) \cdot p_Y(5) = 0,5 \cdot 0,2 = 0,1 \rightarrow p(5, 5) = p_X(5) \cdot p_Y(5)$
- $p(5, 8) = 0,2$ i $p_X(5) \cdot p_Y(8) = 0,5 \cdot 0,4 = 0,2 \rightarrow p(5, 8) = p_X(5) \cdot p_Y(8)$
- $p(5, 12) = 0,2$ i $p_X(5) \cdot p_Y(12) = 0,5 \cdot 0,4 = 0,2 \rightarrow p(5, 12) = p_X(5) \cdot p_Y(12)$

- $p(10, 5) = 0,04$ i $P_X(10) \cdot P_Y(5) = 0,2 \cdot 0,2 = 0,04 \rightarrow p(10, 5) = P_X(10) \cdot P_Y(5)$
- $p(10, 8) = 0,08$ i $P_X(10) \cdot P_Y(8) = 0,2 \cdot 0,4 = 0,08 \rightarrow p(10, 8) = P_X(10) \cdot P_Y(8)$
- $p(10, 12) = 0,08$ i $P_X(10) \cdot P_Y(12) = 0,2 \cdot 0,4 = 0,08 \rightarrow p(10, 12) = P_X(10) \cdot P_Y(12)$
- $p(12, 5) = 0,06$ i $P_X(12) \cdot P_Y(5) = 0,3 \cdot 0,2 = 0,06 \rightarrow p(12, 5) = P_X(12) \cdot P_Y(5)$
- $p(12, 8) = 0,12$ i $P_X(12) \cdot P_Y(8) = 0,3 \cdot 0,4 = 0,12 \rightarrow p(12, 8) = P_X(12) \cdot P_Y(8)$
- $p(12, 12) = 0,12$ i $P_X(12) \cdot P_Y(12) = 0,3 \cdot 0,4 = 0,12 \rightarrow p(12, 12) = P_X(12) \cdot P_Y(12)$

Com a conseqüència, les variables X i Y són independents.

Combinació lineal de variables aleatòries independents

En nombroses ocasions resulta molt interessant l'estudi de les funcions de variables aleatòries (Kenett i altres, 2003; Rao, 1985; Ritchey, 2008; Rohatgi, 2004). Per exemple, si s'extrau a l'atzar una parella d'una gran població de parelles que treballen i es vol saber els ingressos conjunts de la parella. En aquest cas, considerant les variables X = ingressos de l'home i Y = ingressos de la dona, els ingressos totals, que anomenarem S, és el resultat d'una funció de les dues variables i es compleix que $S = X + Y$. O bé, si se'n vol saber la retenció de l'IRPF i se sap que a l'home li retenen un 16 % i a la dona un 17 %, llavors la retenció serà de $W = 0,16 \cdot X + 0,18 \cdot Y$.

És evident que les variables unidimensionals resultants de les funcions de variables aleatòries també són variables aleatòries (tant S com W ho són). Conseqüentment, tenen una distribució de probabilitat que depèn de les variables de les quals són funcions. També els estadístics relacionats, com l'esperança i variància de la nova variable, depenen de les esperances i variàncies de les variables que els originen. Així, a l'exemple, tant la distribució de probabilitat com l'esperança i la variància de S i W depenen de les distribucions i de les esperances i variàncies de X i Y.

Aquestes qüestions poden estendre's per a més de dues variables. És a dir, donades n variables aleatòries, X_1, \dots, X_n , i Z una funció d'elles ($Z = f(X_1, \dots, X_n)$), la distribució de probabilitat de Z depèn de les distribucions de probabilitat de cada variable X_1, \dots, X_n . Així mateix, l'esperança i la variància de Z depèn de les esperances i variàncies de X_1, \dots, X_n . El següent exemple aclareix aquests aspectes.

Encara que al paràgraf anterior s'ha considerat una funció qualsevol, en el text únicament es tractaran funcions lineals, atès que aquest és un manual introductori. Cal recordar que les funcions lineals són aquelles en què les variables aleatòries no es multipliquen entre si, ni apareixen elevades a una potència. Si X_1, X_2 i X_3 són tres variables aleatòries, les variables $Z = 2 \cdot X_1 + X_2 + X_3$, $M = \frac{1}{2} X_1 - 3 \cdot$

$X_2 + X_3 - 8$ són dos exemples de variables aleatòries construïdes com a funcions lineals de X_1 , X_2 i X_3 . També se sol dir que les variables Z i M són combinació lineal de les variables X_1 , X_2 i X_3 .

Exemple

Es consideren X i Y dues variables aleatòries que representen el nombre previst en grans inversions que tenen les dues empreses per a l'any vinent. La distribució de probabilitat de cadascuna és la següent:

X(primera empresa)	0	1	2	Y(segona empresa)	0	1	2	3
$p(x_i)$	0,2	0,5	0,3	$q(y_i)$	0,1	0,3	0,5	0,1

Les dues empreses estan pensant en fusionar-se i per això volen saber com es distribueix el nombre total de grans projectes que durien a terme les dues empreses l'any següent. Per la informació que es maneja, les dues variables poden considerar-se independents.

La variable aleatòria de la qual es demana la distribució de probabilitat és la suma de les dues variables X i Y , puix que aquestes són el nombre de grans projectes que té prevists cada empresa per a l'any següent. Es defineix doncs, $S = X + Y$. A més a més, aquestes variables poden considerar-se independents, ja que les dues empreses, en principi, no basen les seues decisions en el coneixement del que fa l'altra. Llavors, el rang de S és: $\{0, 1, 2, 3, 4, 5\}$ i la seua funció de probabilitat és:

- $p(0) = P(S = 0) = P(X = 0, Y = 0) = p(0) q(0) = 0,2 \cdot 0,1 = 0,02$
- $p(1) = P(S = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 0) = p(0) q(1) + p(1) q(0) = 0,2 \cdot 0,3 + 0,5 \cdot 0,1 = 0,11$
- $p(2) = P(S = 2) = P(X = 2, Y = 0) + P(X = 1, Y = 1) + P(X = 0, Y = 2) = p(2) q(0) + p(1) q(1) + p(0) q(2) = 0,3 \cdot 0,1 + 0,5 \cdot 0,3 + 0,2 \cdot 0,5 = 0,28$
- $p(3) = P(S = 3) = P(X = 2, Y = 1) + P(X = 1, Y = 2) + P(X = 0, Y = 3) = p(2) q(1) + p(1) q(2) + p(0) q(3) = 0,3 \cdot 0,3 + 0,5 \cdot 0,5 + 0,2 \cdot 0,1 = 0,36$
- $p(4) = P(S = 4) = P(X = 1, Y = 3) + P(X = 2, Y = 2) = p(1) q(3) + p(2) q(2) = 0,5 \cdot 0,1 + 0,3 \cdot 0,5 = 0,2$
- $p(5) = P(S = 5) = P(X = 2, Y = 3) = p(2) q(3) = 0,3 \cdot 0,1 = 0,03$

El que hom ara es pot preguntar és com es poden calcular la mitjana i la variància de variables construïdes com a combinació lineal d'altres. Cal dir que el càlcul d'aquests estadístics depèn del grau de relació existent entre les variables, en

concret de si estan correlacionades o no. Això no obstant, com que es consideren variables aleatòries independents, tant el càlcul com les propietats que tot seguit es mostren es redueixen notablement.

Per a dues variables:

Es consideren dues variables aleatòries independents X_1 i X_2 . Es considera $Z = a \cdot X_1 + b \cdot X_2 + K$ on a , b i K són tres nombres distints de zero. Llavors:

1. $E(Z) = a \cdot E(X_1) + b \cdot E(X_2) + K$. En particular $E(X_1 + X_2) = E(X_1) + E(X_2)$
2. $Var(Z) = a^2 \cdot Var(X_1) + b^2 \cdot Var(X_2)$. En particular $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$.

Per a més de dues variables:

Es consideren les variables aleatòries independents X_1, \dots, X_n . Es considera $Z = a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_n X_n + K$ on a_1, a_2, \dots, a_n i K són nombres distints de zero. Llavors:

1. $E(Z) = a_1 \cdot E(X_1) + a_2 \cdot E(X_2) + \dots + a_n E(X_n) + K$
2. $Var(Z) = a_1^2 \cdot Var(X_1) + a_2^2 \cdot Var(X_2) + \dots + a_n^2 Var(X_n)$

Covariància i correlació en variables bidimensionals

Com a continuació i finalització d'aquest apartat, es presenten els conceptes de covariància i coeficient de correlació en variables aleatòries bidimensionals.

La covariància es pot calcular, en cas de tindre dues variables aleatòries X i Y , com: $Cov(X, Y) = E(X, Y) - E(X)E(Y)$

Mentre que la correlació amb aquestes dues variables es pot calcular així:

$$\rho = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$$

Models de probabilitat continu: les distribucions uniforme i exponencial

En els epígrafs anteriors s'han considerat les característiques més importants de les distribucions de probabilitat discreta des d'un punt de vista general. En aquest epígraf es tractaran alguns dels models de probabilitat contínua unidimensional més importants. Aquests són els models uniforme i exponencial. El model normal serà tractat en un epígraf específic per la seua importància.

La Distribució Uniforme

Una variable aleatòria X segueix una distribució uniforme a l'interval $[a, b]$ i es representa per $U[a, b]$ si és contínua, i tots els valors de l'interval tenen les mateixes possibilitats de ser presos per la variable aleatòria.

Notació del model

Si es denota per X la variable aleatòria:

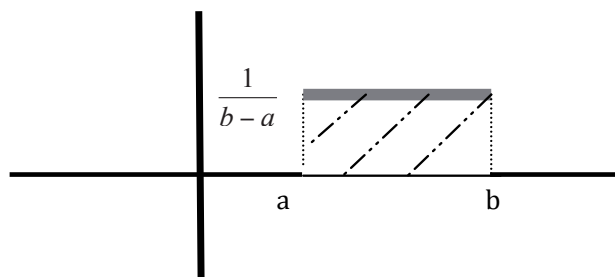
$X =$ valor escollit aleatòriament de l'interval $[a, b]$. Es representa per $X \approx U[a, b]$.

Funció de densitat d'una distribució $U[a, b]$

La funció de densitat és:

$$f(t) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq t \leq b \\ 0 & \text{si } t < a \text{ o } t > b \end{cases}$$

I la representació gràfica és:



A més a més, donats x_1 i x_2 dos valors de l'interval $[a, b]$, es pot demostrar mitjançant un càlcul integral que:

$$P(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{b - a}$$

$$\text{Com a conseqüència, } P(X \leq x_1) = P(a \leq X \leq x_1) = \frac{x_1 - a}{b - a}$$

Esperança i Variància d'una distribució U[a, b]

Es pot demostrar que, si una variable aleatòria es distribuïx uniformement a l'interval [a,b] llavors:

$$E(X) = \mu = \frac{a+b}{2}$$

$$\text{Var}(X) = \sigma^2 = \frac{(b-a)^2}{12}$$

La distribució exponencial

La distribució exponencial se sol emprar per a resoldre problemes de llistes d'espera o cues. D'alguna manera, es pot considerar com complementària de la distribució Poisson, ja que si aquesta calcula la probabilitat que ocorreguen un nombre d'esdeveniments en un interval de temps o espai en el qual se'n produeixen una mitjana de μ , la distribució exponencial calcula la probabilitat del temps que ha de transcórrer entre dues ocurrències si el nombre d'ocurrències per unitat de temps és μ .

Notació del model

Si es denota per X la variable aleatòria:

X = temps que passa entre dues ocurrències, es representa per $X \approx \text{Exp}(\mu)$, on μ = mitjana d'esdeveniments per unitat de temps o espai.

Funció de densitat d'una distribució Exp(μ)

La funció de densitat és:

$$f(t) = \begin{cases} \mu \cdot e^{-\mu t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

On t és el nombre d'unitats de temps fins la següent ocurrència. μ és el nombre mitjà d'ocurrències per unitat de temps.

Es pot demostrar matemàticament que la funció de distribució és $F(t) = 1 - e^{-\mu t}$. La representació gràfica d'ambdues funcions és la següent:

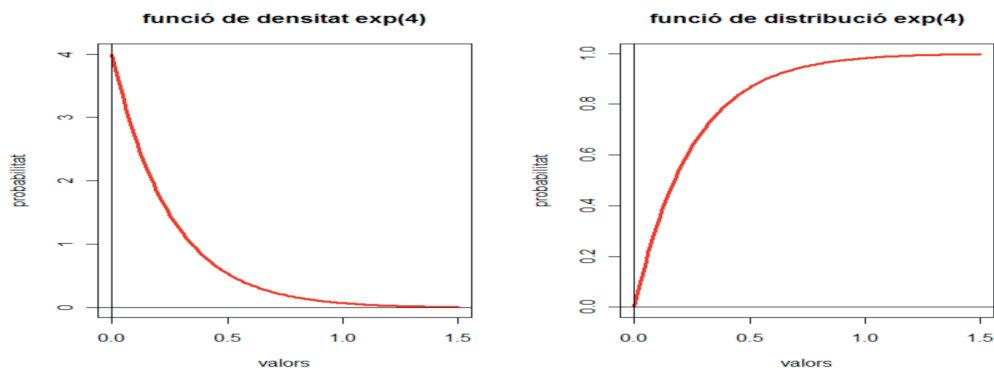


Fig. 12

Esperança i variància d'una distribució $Exp(\mu)$

Es pot demostrar que, si una variable aleatòria es distribueix exponencialment de paràmetre μ , llavors:

$$E(X) = \frac{1}{\mu}$$

$$\text{Var}(X) = \sigma^2 = \frac{1}{\mu^2}$$

Distribució normal. El teorema central del límit

La distribució normal és, sense dubte, la més important de totes les distribucions de probabilitat del càlcul de probabilitats i de l'estadística. Hi ha tres raons principals:

- *Les propietats matemàtiques.* És d'enorme importància en la inferència estadística.
- *L'aplicació.* Un gran nombre de fenòmens reals es poden modelar mitjançant la distribució normal. Per exemple, la distribució d'altures, pesos, distàncies i altres variables que són divisibles fins a l'infinit.
- *El teorema central del límit.* La distribució normal serveix per a aproximar la suma i la mitjana de qualsevol altre tipus de distribucions.

Cal dir que la distribució normal necessita dos paràmetres per a poder ser definida. Així, és necessari conèixer la mitjana o esperança i la desviació típica per a poder calcular el model. Aquest fet és en moltes ocasions impossible, i es fa necessari l'ús d'estimacions puntuals d'aquests paràmetres o d'altres tècniques d'inferència.

No obstant, en definir-se la distribució normal des d'un punt de vista teòric es consideraran coneguts tots dos paràmetres.

Definició

Una variable aleatòria X segueix una distribució Normal amb paràmetres μ i σ ($-\infty < \mu < +\infty$ i $\sigma > 0$), i es denota per $X \sim N(\mu, \sigma)$, si la *funció de densitat de probabilitat* és:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]} \text{ per a } -\infty < x < \infty$$

Per ser contínua, la probabilitat de que la variable aleatòria X estiga compresa entre dos valors x_1 i x_2 es calcula mitjançant el càlcul integral:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{(t-\mu)^2}{2\sigma^2}\right]} dt$$

No obstant, existeix un procediment que es desenvoluparà més endavant i que permet el càlcul d'aquestes probabilitats emprant una taula i, en conseqüència, sense resoldre la integral. La representació gràfica de la funció de densitat té forma de campana. La figura 13 és la funció de densitat de una distribució normal de paràmetres (10, 5).

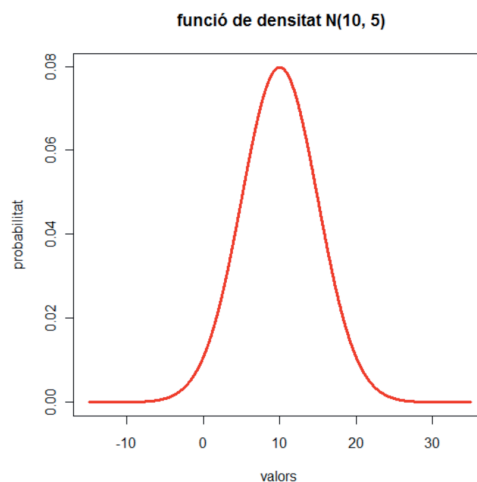


Fig. 13

Esperança i variància d'una distribució normal

Mitjançant càlculs matemàtics, es pot demostrar que: $E(X) = \mu$ i que $\text{Var}(X) = \sigma^2$, per tant la desviació típica de la distribució és σ . És a dir, els paràmetres que defineixen una distribució de probabilitat normal són l'esperança (μ) i la desviació típica (σ).

Representació gràfica. Interpretació dels paràmetres

Es mostra tot seguit la representació gràfica de la funció de densitat d'una distribució normal d'esperança 0 i desviació típica 1 (fig. 14).

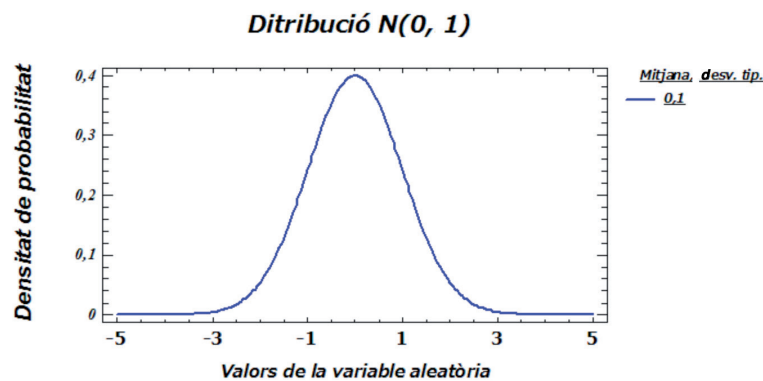


Fig. 14

Com es pot comprovar, la funció de densitat té forma de campana (se l'anomena campana de Gauss) i és *simètrica respecte del valor de l'esperança* (en aquest cas 0). També es pot comprovar que la densitat de probabilitat decreix a tots dos costats (les cues) i és molt més gran en els valors centrals, propers a l'esperança 0. Aquest fet significa que els valors centrals tenen molta més probabilitat de donar-se que els valors que es troben en les cues.

Què passa si en una distribució normal augmenta o disminueix la desviació típica? Això es comprovarà també mitjançant els gràfics de les respectives funcions de densitat de probabilitat.

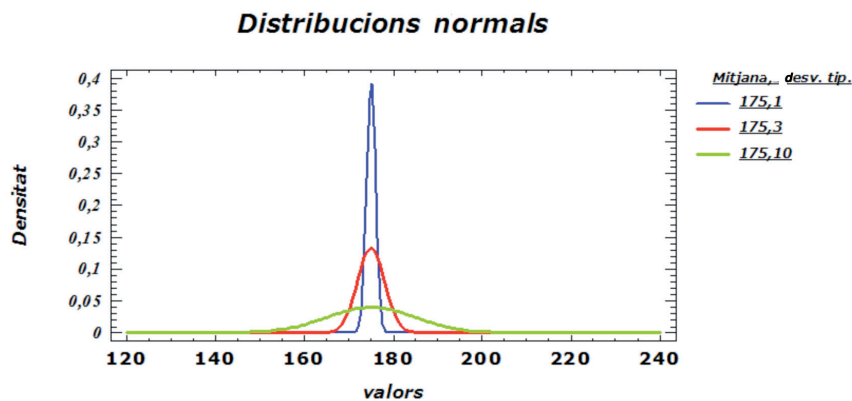


Fig. 15

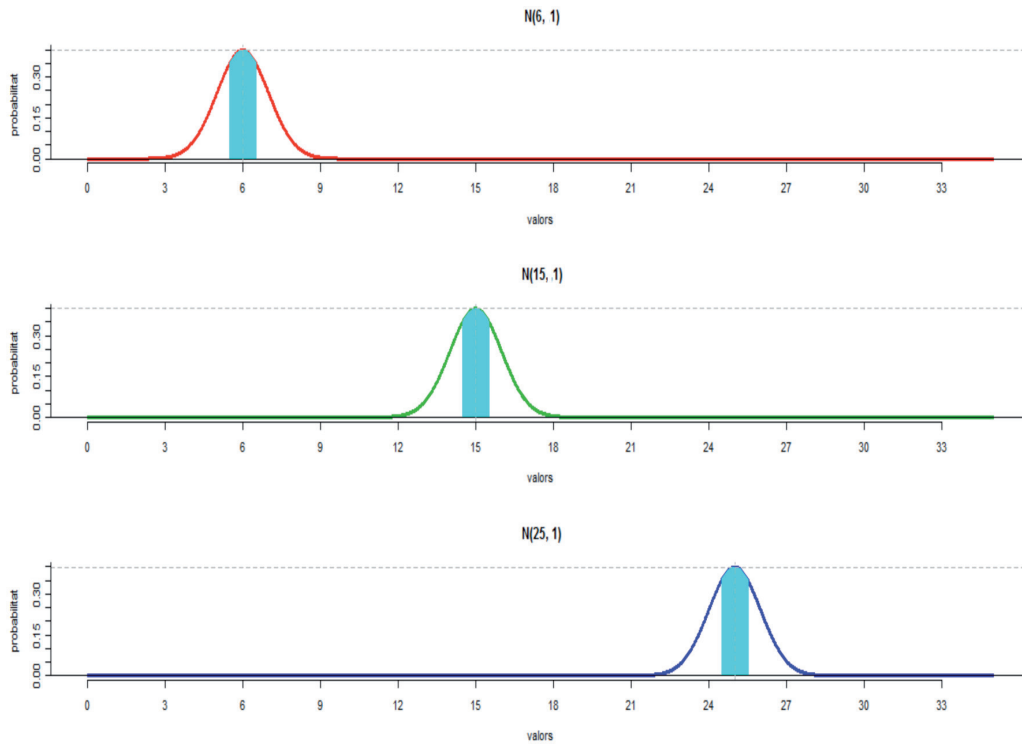


Fig. 16

Distribucions normals

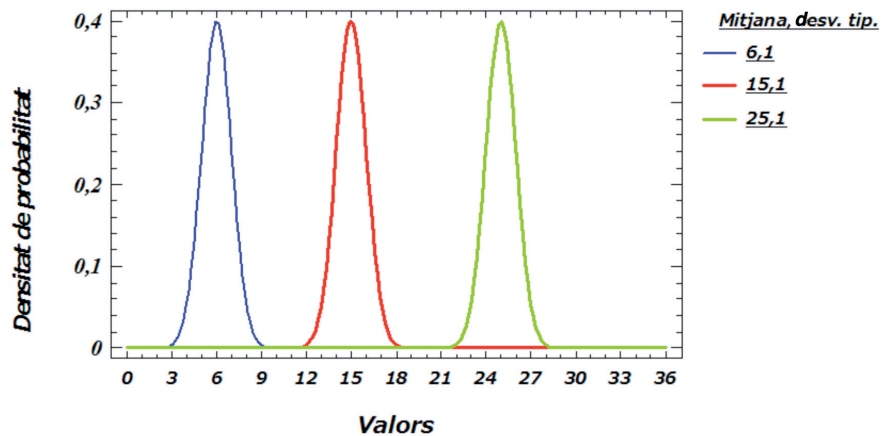


Fig. 17

La representació gràfica de totes les distribucions és la mateixa (fig. 17). Únicament varia el punt on se centren. Aquest punt és, evidentment, l'esperança de cadascuna de les distribucions: 6 per a la primera, 15 per a la segona i 25 per a la tercera.

Això si aquest fet significa que els valors que pren la variable aleatòria es distribueixen de la mateixa manera entorn de l'esperança.

El teorema que tot seguit es presenta permet realitzar els càlculs de probabilitats de variables aleatòries que es distribueixen segons una distribució normal sense haver de realitzar els càlculs integrals. Així doncs, el teorema següent relaciona distribucions $N(\mu, \sigma)$ amb distribucions $N(0, 1)$.

Teorema de les transformacions lineals

Siga $X \approx N(\mu, \sigma)$ i siga $Y = aX + b$, on a i b són constants. Aleshores $Y \approx N(a\mu + b, a \cdot \sigma)$.

Corol·lari

Si $X \approx N(m, s)$, llavors la nova variable aleatòria $Z = \frac{X - \mu}{\sigma} \approx N(0, 1)$.

Distribució normal tipificada

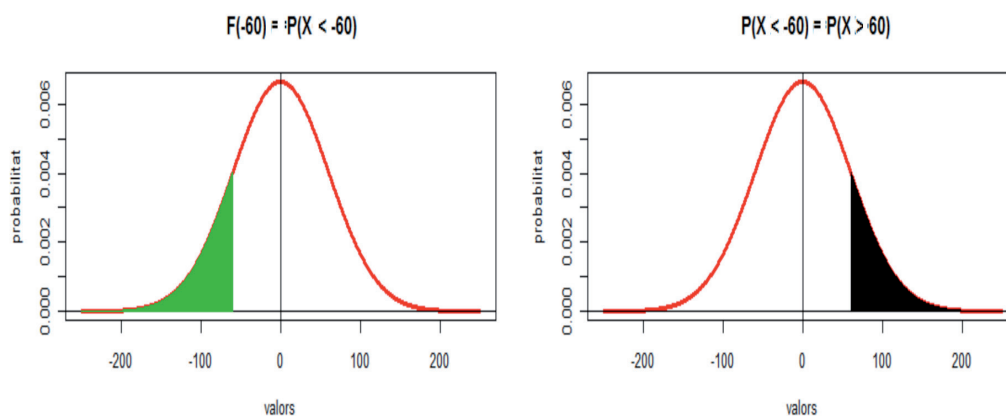
La distribució normal amb esperança o mitjana 0 i variància 1 es denomina distribució normal tipificada: $Z \approx N(0, 1)$. La funció de distribució d'aquesta variable aleatòria es denota per Φ , on $\Phi(z) = P(Z \leq z)$, i s'utilitza per al càlcul de probabilitats amb taules, ja que les probabilitats acumulades per a Z estan tabulades. També es pot utilitzar *software* informàtic. Al llarg del text s'emprarà indistintament $\Phi(z)$ i $P(Z \leq z)$.

Important: Per ser la distribució simètrica, es compleix (fig. 12):

$$P(Z \leq z) = P(Z > -z) = 1 - P(Z \leq -z)$$

$$\text{Així, } \Phi(-z) = 1 - \Phi(z)$$

És important familiaritzar-se amb el càlcul de probabilitats utilitzant la taula de la distribució $N(0, 1)$, ja que tots els càlculs de probabilitats es realitzen emprant aquestes taules.



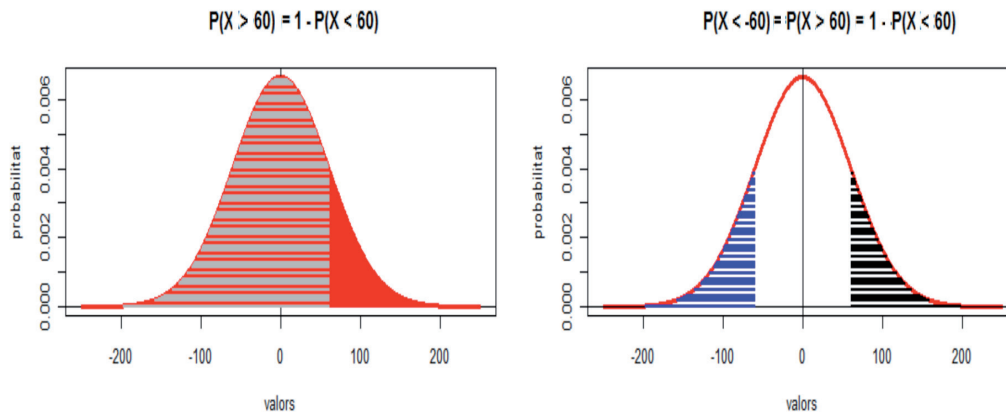


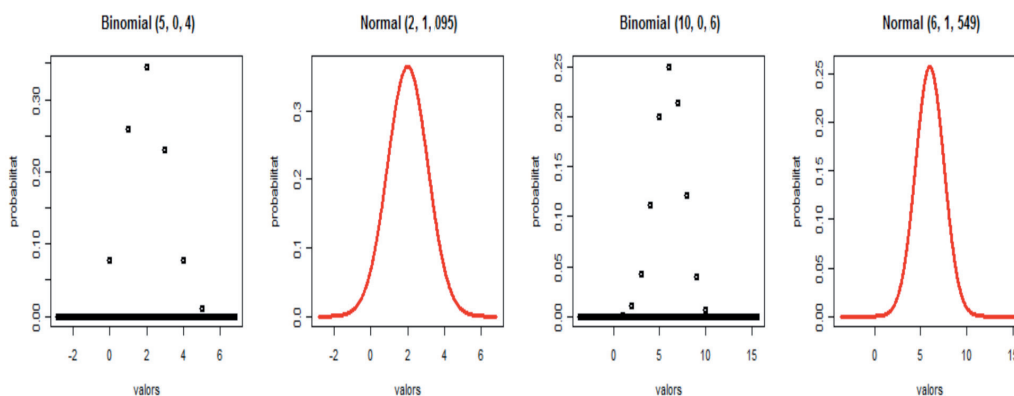
Fig. 18

La distribució normal com una aproximació a la distribució binomial

Ja s'ha vist que el càlcul de probabilitats de la binomial (n, p) pot aproximar-se per la distribució de Poisson ($n \cdot p$) si n i p compleixen unes determinades condicions. D'un mode semblant, la distribució binomial pot aproximar-se per una distribució normal.

Així doncs, si una distribució X és $Bi(n, p)$ i n és suficientment gran, es pot emprar l'aproximació normal en què $\mu = n \cdot p$ i $\sigma^2 = n \cdot p \cdot (1 - p)$. Aquesta aproximació és molt acurada quan $n \cdot p \geq 5$ i $n \cdot (1 - p) \geq 5$.

Gràficament (fig. 19), sembla bastant clar que, com més gran és el valor de n , més s'assembla la distribució binomial a una normal.



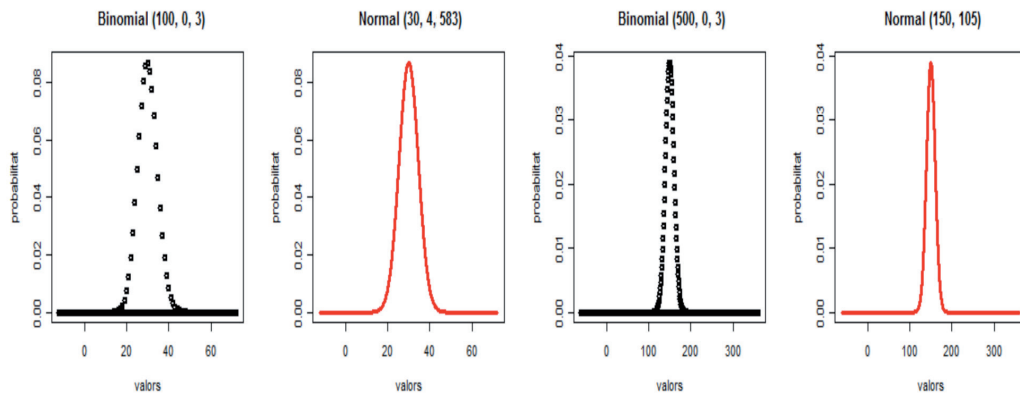


Fig. 19

Com s'observa en els gràfics anteriors, en augmentar el nombre d'experiments, n , la distribució va assemblant-se cada vegada més a una distribució normal. Cal tindre en compte, però, que en ser la distribució binomial una distribució discreta i la normal, una distribució contínua, a l'hora de fer les aproximacions és necessari aplicar una «correcció». Així, si es denota X = distribució binomial i X' = distribució normal que l'aproxima, llavors la correcció consisteix a assumir que $P(X = a) = P(a - 0,5 \leq X' \leq a + 0,5)$.

Al gràfic (fig. 20) s'observa que en una distribució Binomial (30, 0, 6) la $P(X = 17)$ és la mateixa que en la seua aproximació per la normal $N(18, 7, 2)$, $P(17 - 0,5 \leq X' \leq 17 + 0,5)$.

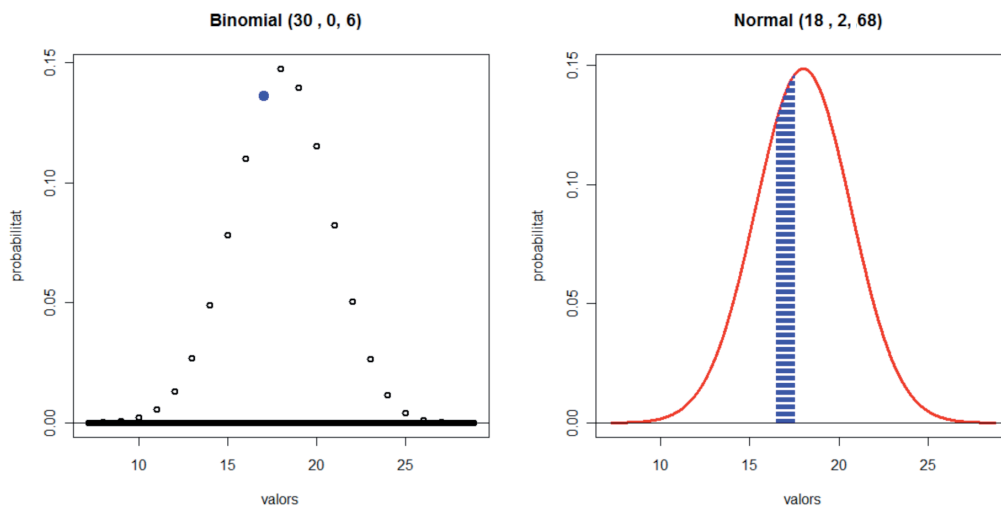


Fig. 20

Numèricament, es pot calcular emprant *software* estadístic:

Valors (x_i)	Binomial $P(X = x_i)$	Normal $P(x_i - 0,5 < X < x_i + 0,5)$	Diferència
15	7.831221e-02	7.969387e-02	1.381658e-03
16	1.101265e-01	1.123280e-01	2.201428e-03
17	1.360387e-01	1.380144e-01	1.975729e-03
18	1.473752e-01	1.478211e-01	4.458439e-04
19	1.396186e-01	1.380144e-01	-1.604236e-03
20	1.151854e-01	1.123280e-01	-2.857403e-03
21	8.227527e-02	7.969387e-02	-2.581401e-03
22	5.048710e-02	4.928696e-02	-1.200133e-03
23	2.634109e-02	2.657081e-02	2.297196e-04
24	1.152423e-02	1.248641e-02	9.621778e-04
25	4.148722e-03	5.114733e-03	9.660103e-04
26	1.196747e-03	1.826220e-03	6.294734e-04
27	2.659437e-04	5.683541e-04	3.024103e-04
28	4.274096e-05	1.541737e-04	1.114327e-04
29	4.421478e-06	3.645155e-05	3.203007e-05
30	2.210739e-07	7.511459e-06	7.290385e-06
31	0.000000e+00	1.349033e-06	1.349033e-06
32	0.000000e+00	2.111533e-07	2.111533e-07
33	0.000000e+00	2.880289e-08	2.880289e-08
34	0.000000e+00	3.423917e-09	3.423917e-09
35	0.000000e+00	3.546865e-10	3.546865e-10
36	0.000000e+00	3.201728e-11	3.201728e-11

Valors (x_i)	Binomial $P(X = x_i)$	Normal $P(x_i - 0,5 < X < x_i + 0,5)$	Diferència
37	0.000000e+00	2.518430e-12	2.518430e-12
38	0.000000e+00	1.726397e-13	1.726397e-13
39	0.000000e+00	1.032507e-14	1.032507e-14
40	0.000000e+00	5.551115e-16	5.551115e-16
41	0.000000e+00	0.000000e+00	0.000000e+00
42	0.000000e+00	0.000000e+00	0.000000e+00
43	0.000000e+00	0.000000e+00	0.000000e+00
44	0.000000e+00	0.000000e+00	0.000000e+00
45	0.000000e+00	0.000000e+00	0.000000e+00

Com s'observa a la taula, els valors de la binomial i de la normal amb correcció s'ajusten molt. El gràfic següent (fig. 21) mostra com els valors de les diferències poden considerar-se gairebé menyspreables per estar molt pròxims al zero.

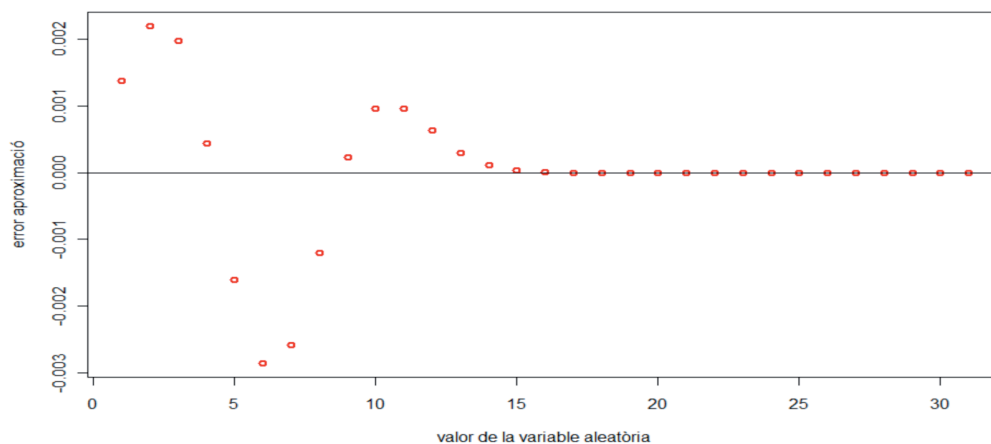


Fig. 21

La distribució normal també pot emprar-se per a aproximar la distribució de Poisson quan el paràmetre és molt gran. La mitjana i la desviació típica de la distribució normal són «heretades» de la Poisson. El gràfic següent (fig. 22) mostra les funcions de distribució de les tres variables aleatòries en un dels casos en què podria fer-se l'aproximació.

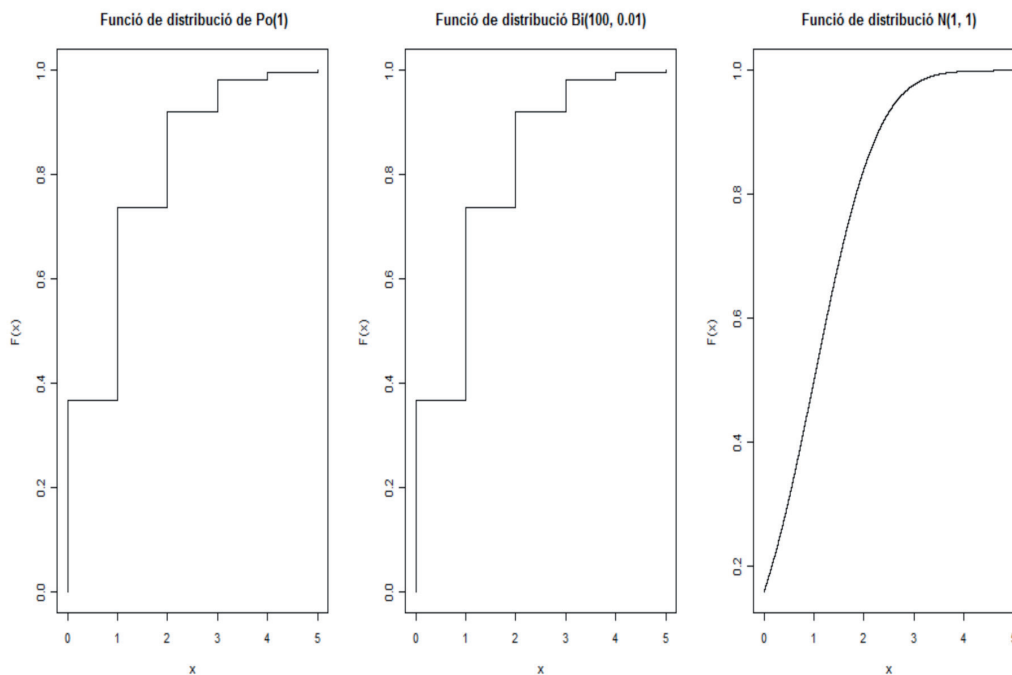


Fig. 22

Teorema central del límit

Les distribucions binomial i de Poisson poden ser aproximades per la distribució normal, com ja s'ha comentat al llarg del text. Analitzant amb més detall les dues distribucions, s'observa que la primera es pot considerar com una suma de variables aleatòries independents Bernoulli amb probabilitat d'èxit constant, i la segona, com una suma de distribucions de Poisson independents (tan sols cal dividir l'interval de temps o espai en intervals més petits i aplicar la propietat de la suma de distribucions Poisson). Per tant, d'alguna manera la distribució normal ha estat emprada per a aproximar distribucions suma de variables aleatòries independents. Però aquest «privilegi» tan sols el gaudeixen per les distribucions binomial i de Poisson? La resposta és no, i la raó, el teorema central del límit.

Així doncs, el teorema central del límit estudia el comportament de la suma (o la mitjana) de variables aleatòries. És a dir, el teorema central del límit afirma que la distribució d'una suma molt gran de variables aleatòries s'aproxima a una distribució normal. D'aquesta manera, per una banda aporta a l'estadística un resultat crucial per a l'estudi de la distribució asimptòtica de moltes variables aleatòries que, com es comprovarà en els capítols posteriors, serà de gran utilitat per als contrastos d'hipòtesis i els intervals de confiança. Per una altra banda, el teorema central del límit proporciona una explicació teòrica fonamentada en un fenomen habitual en els experiments reals: les variables presenten en moltes ocasions una distribució empírica aproximadament normal.

Ara bé, com es tradueix en termes de les matemàtiques aquesta «aproximació» a una distribució normal? La resposta ve donada pel concepte de convergència estadística d'una successió de variables aleatòries, que breument es passarà a desenvolupar.

Convergència estadística

La convergència estadística és un concepte complex que no es pot reduir a una única tipologia. Així, donada una successió de variables aleatòries X_1, X_2, \dots, X_n amb les corresponents funcions de distribució F_1, F_2, \dots, F_n , es pot dir que la successió convergeix (s'aproxima més i més en augmentar n) a una variable aleatòria X en probabilitat, en distribució o llei, quasi segurament, etc. En aquest manual tan sols es comentaran les dues primeres.

Convergència en probabilitat

Una successió de variables aleatòries X_1, X_2, \dots, X_n convergeix en probabilitat a una variable aleatòria X i es representa per $X_n \xrightarrow{p} X$ si per a cada $\varepsilon > 0$ es compleix que:

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$$

És a dir, a mesura que es van prenent variables aleatòries amb n més i més gran, la probabilitat de que siguin diferents de la variable aleatòria X es va fent cada cop més menuda.

Convergència en distribució

Una successió de variables aleatòries X_1, X_2, \dots, X_n amb funcions de distribució F_1, F_2, \dots, F_n convergeix en distribució o llei a una variable aleatòria X amb funció de distribució F , i es representa per $X_n \xrightarrow{p} X$ per a cada x en què $F(x)$ és contínua:

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$$

És a dir, a mesura que es van prenent variables aleatòries amb n més gran, les funcions matemàtiques que representen les funcions de distribucions $F_n(x)$ i $F(x)$ són cada vegada més i més semblants.

Es pot demostrar que la convergència en probabilitat implica la convergència en distribució, encara que l'asseveració oposada no és sempre certa.

Teorema central del límit

Així doncs, tenint en compte els conceptes de convergència en probabilitat i en distribució esmentats, i que, com qualsevol teorema important de les matemàtiques, s'ha anat construint al llarg de la història i existeix més d'una versió d'aquest teorema, depenent de les exigències de les hipòtesis, un dels enunciats més comuns del teorema central del límit és el que tot seguit es presenta.

Si X_1, X_2, \dots, X_n variables aleatòries independents i idènticament distribuïdes amb esperança μ i variància σ^2 , i si n és suficientment gran, la variable aleatòria $X_1 + X_2 + \dots + X_n$ convergeix en distribució a un model normal d'esperança $n \cdot \mu$ i variància $n \cdot \sigma^2$. És a dir:

$$X_1 + X_2 + \dots + X_n \approx N(n \cdot \mu, \sqrt{n} \cdot \sigma)$$

Cal remarcar que aquest teorema també admet la seua versió equivalent:

$$\frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Objectius

Els problemes han de permetre que els alumnes assolisquen els objectius didàctics següents:

- 3a) Conèixer els trets més característics de les distribucions de probabilitat d'una variable contínua: funció de densitat, funció de distribució, esperança matemàtica, variància, percentils, etc.
- 3b) Identificar els experiments que es modelen mitjançant una distribució contínua.
- 3c) Conèixer i saber fer càlculs amb les distribucions contínues uniforme, exponencial i normal.
- 3d) Estudiar la relació entre els models de Poisson i exponencial.
- 3e) Saber reconèixer i aplicar el teorema central del límit i aplicar-lo per transformar altres models en una distribució normal.

La taula següent ens mostra com estan distribuïts els objectius segons els exercicis:

Objectiu Exercici	3a	3b	3c	3d	3e
1	x				
2	x				
3	x	x	x		
4	x	x	x		
5	x	x	x	x	
6	x	x	x		
7	x	x	x		x
8	x	x	x		x

Enunciats

-
- 3a) Conèixer els trets més característics de les distribucions de probabilitat d'una variable contínua: funció de densitat, funció de distribució, esperança matemàtica, variància, percentils, etc.

Exercici 1

Considerem una variable X que expressa el nombre de comandes (en unitats de milers) ateses diàriament per una empresa de missatgeria a nivell nacional. Suposem que la funció següent $f(x)$ és la seua funció de densitat que ens mostra la seua distribució.

$$f(x) = \begin{cases} cx^2 & \text{si } 1 \leq x \leq 3 \\ 0 & \text{en la resta} \end{cases}$$

- a) Calcula el valor del paràmetre c .
- b) Creus que aquesta empresa ha atès 300 comandes algun dia? Justifica la teua resposta.
- c) Calcula la probabilitat que un dia qualsevol atenga entre 1.500 i 2.000 comandes, menys de 3.500, més de 3.000 i exactament 1.400.
- d) Calcula la funció de distribució $F(x)$ i respon un altra vegada a les qüestions de l'apartat c amb aquesta funció. Representa-la.
- e) Calcula la mitjana o esperança, la variància i la desviació típica. Interpreta-les.

Exercici 2

Per a justificar l'eliminació de cert servei de transport urbà en els trajectes de l'any s'han anotat els percentatges d'ocupació de tots els viatges realitzats. Així, la variable $X =$ percentatge d'ocupació (en tant per u), es distribueix amb la següent funció de distribució:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 2x - x^2 & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

- a) Podries calcular la probabilitat de que un trajecte presente menys del 10 % d'ocupació, entre el 20 i 40 %, o més del 80 %?
- b) Calcula la funció de densitat de la variable X .

- c) Calcula la mediana, la mitjana i la desviació típica d'aquesta distribució. Interpreta-les.
- d) Calcula el percentatge d'ocupació que té, com a molt, el 95 % dels trajectes.

-
- 3a) Conèixer els trets més característics de les distribucions de probabilitat d'una variable contínua: funció de densitat, funció de distribució, esperança matemàtica, variància, percentils, etc.
- 3b) Identificar els experiments que es modelen mitjançant una distribució contínua.
- 3c) Conèixer i saber fer càlculs amb la distribució contínua uniforme.

Exercici 3

Estem fent un estudi entre les empreses d'un cert sector que van vendre beneficis l'any passat. S'observa una distribució uniforme en el sector que va del 4 % al 7 % de beneficis que volem estudiar amb més profunditat.

- a) Si considerem que una empresa amb beneficis inferiors al 6,7 % no pot assegurar el seu manteniment de plantilles, calcula la proporció d'empreses que podem estimar, dins d'aquest sector crític que estem analitzant, que està en risc d'acomiadar treballadors.
- b) Calcula les probabilitats $P(X \leq 5, 2)$, $P(4, 5 \leq X \leq 5, 2)$.
- c) Calcula l'esperança, la variància i la desviació típica. Interpreta-les.

-
- 3a) Conèixer els trets més característics de les distribucions de probabilitat d'una variable contínua: funció de densitat, funció de distribució, esperança matemàtica, variància, percentils, etc.
- 3b) Identificar els experiments que es modelen mitjançant una distribució contínua.
- 3c) Conèixer i saber fer càlculs amb la distribució contínua exponencial.

Exercici 4

Volem planificar l'assignació de tasques en un cert departament d'atenció al públic. Per tractar-se majorment de demandes de documentació, hem estudiat que el

temps d'atenció de cadascuna de les peticions segueix una distribució exponencial amb una mitjana de 8 minuts.

- a) Calcula la probabilitat de que una atenció qualsevol necessite entre 5 i 12 minuts.
- b) Calcula el temps mínim que podem assegurar que necessitem en el 90 % de les atencions.
- c) Calcula la probabilitat de que una consulta que ja porta 5 minuts atenent-se necessite, encara, 7 minuts més.
- d) Calcula la probabilitat de que una tasca dure exactament 8 minuts.

-
- 3a) Conèixer els trets més característics de les distribucions de probabilitat d'una variable contínua: funció de densitat, funció de distribució, esperança matemàtica, variància, percentils, etc.
 - 3b) Identificar els experiments que es modelen mitjançant una distribució contínua.
 - 3c) Conèixer i saber fer càlculs amb la distribució contínua exponencial.
 - 3d) Estudiar la relació entre els models de Poisson i exponencial.

Exercici 5

En un servei d'atenció en carretera de 24 hores, la mitjana de serveis que requereixen l'atenció amb grua és de 4 incidències per hora.

Considerem la variable que estudia el temps transcorregut entre dues eixides de la grua en el servei.

- a) Calcula el temps mitjà d'espera entre dos serveis que requereixen grua. Calcula'n també la desviació típica.
- b) Calcula la probabilitat de que la grua espere entre 10 i 20 minuts fins a rebre un nou avís.
- c) Calcula la probabilitat de que no hi haja demanda del servei en un interval de més de 30 minuts.

-
- 3a) Conèixer els trets més característics de les distribucions de probabilitat d'una variable contínua: funció de densitat, funció de distribució, esperança matemàtica, variància, percentils, etc.
 - 3b) Identificar els experiments que es modelen mitjançant una distribució contínua.
 - 3c) Conèixer i saber fer càlculs amb la distribució contínua normal.

Exercici 6

En un departament d'estadística s'està dissenyant un qüestionari que haurà d'implementar-se en una mostra gran. Per a preveure les necessitats de temps i personal que necessitem, hem passat el qüestionari a 200 individus seleccionats a l'atzar i hem anotat el temps que han emprat per a omplir-lo. S'observa que els resultats d'aquest mostratge inicial s'ajusten a un model de distribució normal de mitjana 12 i desviació típica de 0,1 minuts.

- a) Per a classificar els temps de resposta, denominarem «estàndard» aquelles que han estat d'una duració que es desvien de la mitjana com a molt, en 0,2 minuts. Calcula la proporció de respostes estàndard que hi haurà. Si la mostra real que ompli el qüestionari dissenyat és de 2.000 individus, calcula el nombre de respostes d'aquest tipus que puguem esperar.
- b) Calcula la probabilitat de que un individu necessite menys de 10,5 minuts o més de 13 minuts per a omplir el qüestionari.

-
- 3a) Conèixer els trets més característics de les distribucions de probabilitat d'una variable contínua: funció de densitat, funció de distribució, esperança matemàtica, variància, percentils, etc.
 - 3b) Identificar els experiments que es modelen mitjançant una distribució contínua.
 - 3c) Conèixer i saber fer càlculs amb la distribució contínua normal.
 - 3e) Saber reconèixer i aplicar el teorema central del límit i aplicar-lo per a transformar altres models en una distribució normal.

Exercici 7

En certa classe d'estudiants d'estadística hi ha 80 alumnes, dels quals pensem que el 55 % ja coneix suficientment el programari de les pràctiques. Si en aquest moment plantegem un exercici per a resoldre'l utilitzant l'esmentat programari, calcula la probabilitat de que, al menys, el resolga el 75 % dels alumnes.

Exercici 8

En certa empresa de venda *on line* estan rebent una mitjana de 30 comandes per hora. Se suposa que ho fan de manera aleatòria, proporcional al temps que transcorre i amb les característiques que permeten assignar el model de Poisson a aquesta distribució.

- a) Calcula la probabilitat de que en certa hora arriben entre 20 i 40 comandes.
- b) Si volem atendre tots els encàrrecs que arriben a aquesta hora, calcula els treballadores lliures que necessitem en una hora per atendre el 99 % de les comandes (suposem que un treballador pot despatxar 3 comandes per hora).

Ajudes

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de resoldre els exercicis i problemes. És convenient no fer un ús excessiu d'aquestes ajudes, és a dir, que abans d'emprar l'ajuda cal pensar el problema almenys durant uns 10 – 15 minuts. Després es consultarà l'ajuda de tipus 1 i s'intentarà resoldre l'exercici amb aquesta ajuda. Si no és possible resoldre'l, llavors es consultarà l'ajuda de tipus 2; en darrer terme es consultarà la solució.

Ajudes de tipus 1

Per a resoldre qualsevol exercici comencem per definir la variable i fixar-nos en la informació que el problema dona de la seua distribució. Podem tindre la funció de densitat $f(x)$ com al problema 1, la funció de distribució $F(x)$ com al problema 2 o el model al qual es pot ajustar la distribució, amb la qual cosa coneixem totes les característiques. En aquest cas, de la comprensió de l'enunciat cal obtindre els valors dels paràmetres del model que es proposa.

Exercici 1

- El que es necessita per a resoldre aquest apartat és conèixer la condició de que l'àrea tancada per sota de la corba de la funció de densitat és 1.
- Aquest apartat pretén que es faça una interpretació del rang de la variable i les mesures en unitats de milers.
- Com que ja tenim la funció de densitat, caldrà calcular algunes probabilitats mitjançant les integrals corresponents.
- Ens demanen la funció de distribució, aplicada per integració, de la seua definició $F(x) = P(X \leq x)$ per a després utilitzar-la en el càlcul de probabilitats.
- Per a calcular la mitjana o esperança, la variància i la desviació típica, cal aplicar les definicions i la funció de densitat del nostre exercici.

$$E[X] = \int_1^3 x \cdot f(x) dx$$

$$\text{Var}[X] = E[X^2] - E[X]^2 \text{ on } E[X^2] = \int_1^3 x^2 \cdot f(x) dx \text{ i } \sigma = \sqrt{\text{VAR}[X]}$$

Exercici 2

Considerarem la variable $X =$ percentatge d'ocupació (en tant per u) en els trajectes del darrer any.

- Per tractar-se d'una variable contínua, el recompte de probabilitats és fàcil de fer a partir de la $F(x)$ de l'enunciat.
- Per a obtenir la funció de densitat, derivem la funció de distribució.
- Per a calcular la mitjana o esperança, la variància i la desviació típica, cal aplicar les definicions i la funció de densitat del nostre exercici.

$$E[X] = \int_1^3 x \cdot f(x) dx$$

$$\text{Var}[X] = E[X^2] - E[X]^2 \text{ on } E[X^2] = \int_1^3 x^2 \cdot f(x) dx \text{ i } \sigma = \sqrt{\text{VAR}[X]}$$

Podeu vore l'apartat del problema anterior, on es demanen aquestes mateixes característiques.

Recordem que la mediana és el percentil 50.

- Calcula el percentatge d'ocupació que té, com a molt, el 95 % dels trajectes.

Ens demanen un percentil, caldrà decidir si interessa el P_5 o el P_{95} . Vegem:

$$F(x_0) = P(X \leq x_0) = 0,95$$

Exercici 3

Aquest és un exemple de distribució uniforme.

- Pel que es diu a l'enunciat, la funció de densitat tan sols ens mostra la distribució de les empreses del sector que tenen uns beneficis d'entre el 4 % i el 7 % i que seran l'objectiu del nostre treball. Podem escriure'n la funció de densitat i observar que és una funció constant, en el rang de la variable:

$$f(x) = \begin{cases} \frac{1}{7-4} & \text{si } 4 \leq x \leq 7 \\ 0 & \text{en la resta} \end{cases}$$

Per tant, podem calcular la probabilitat com l'àrea d'un rectangle.

Per a calcular probabilitats integrem la $f(x)$ anterior en l'interval que necessitem i en calculem el valor per Barrow.

Així $P(X \leq 6,7) = 1 - P(X > 6,7) = \dots$

$$\text{És a dir, } P(X > 6,7) = \int_{6,7}^7 \frac{1}{3} dx = \dots$$

- b) Abordarem aquest apartat per a calcular les probabilitats $P(X \leq 5, 2)$, $P(4,5 \leq X \leq 5,2)$ de la mateixa manera que l'anterior.
- c) Per a calcular l'esperança, variància i desviació típica, caldrà anar a la teoria i consultar les expressions algebraiques d'aquestes en funció del model. Són expressions que depenen dels valors dels paràmetres i que tan sols cal substituir convenientment amb els valors concrets de cada problema.

Exercici 4

Aquest és un exemple de distribució exponencial on caldrà concretar-ne el valor del paràmetre, el qual extraïem de la mitjana:

$$E[X] = \frac{1}{\beta} = 8 \rightarrow \beta = \frac{1}{8}$$

L'expressió de la funció de densitat serà $f(x) = \frac{1}{8} e^{-\frac{1}{8}x}$ si $x > 0$, i amb aquesta podem abordar els càlculs mitjançant integració.

- a) Calcula la probabilitat de que una atenció qualsevol necessite entre 5 i 12 minuts.

$$P(5 \leq X \leq 12) = \int_5^{12} \frac{1}{8} e^{-\frac{1}{8}x} dx$$

- b) Calcula el temps mínim que podem assegurar que necessitarem en el 90 % de les atencions.

Aquesta qüestió requereix el càlcul d'un percentil. Recordeu que és convenient no precipitar-se i distingir, amb la lectura atenta de l'enunciat, si ens demanen el P_{10} o, pel contrari, ens demanen el P_{90} .

- c) Calcula la probabilitat de que una consulta que ja porta 5 minuts atenent-se necessite, encara, 7 minuts més.

Ens demanen el càlcul d'una probabilitat condicionada: $P\left(\frac{X > 12}{X > 5}\right)$

d) Calcula la probabilitat de que una tasca dure exactament 8 minuts.

Recordem que per a fer el tractament de la variable X amb el model d'una contínua, escrivim $P(X = 8) = 0$.

Exercici 5

En aquest exercici volem treballar la relació que existeix entre la variable Poisson, que és el model de distribució del nombre d'esdeveniments en un cert període de temps o espai (si es compleixen les altres característiques del fenomen, és clar), i la distribució exponencial, que és el model que s'ajusta a la variable que mesura el temps que transcorre entre dos esdeveniments consecutius de la variable Poisson, abans esmentada.

Així, al nostre problema veiem que Y (nombre de cridades al servei de grues) es pot ajustar a una distribució Poisson de mitjana 4 hores: $Y \rightarrow \text{Po}(\lambda = 4)$.

Així doncs, la variable $X =$ temps (en hores) que transcorre entre dos serveis consecutius de la grua, es distribueix com una variable exponencial de mitjana $E[X] = \frac{1}{\lambda} = 0,25$ hores, per això $X =$ temps entre dos serveis $\rightarrow \text{Exp}(\beta = \lambda = 4)$.

- a) Per a calcular l'esperança, variància i desviació típica, caldrà anar a la teoria i consultar les expressions algebraiques del model exponencial de paràmetre ($\beta = \lambda = 4$).
- b) Calcula la probabilitat de que una grua espere entre 10 i 20 minuts fins a rebre un nou avís.

Cal transformar els valors de la variable «temps» a hores perquè la mitjana del nombre d'esdeveniments en la Poisson era de 4 serveis per hora.

Plantejarem el càlcul de la probabilitat a partir de la funció de distribució (si està al nostre abast) o per integració:

$$P(0,17 < X < 0,33) = F(0,33) - F(0,17) = \int_{0,17}^{0,33} 4e^{-4x} dx$$

- c) Calcula la probabilitat de que no hi haja demanda del servei en un interval de més de 30 minuts.

$$P(X > 0,5) = ?$$

Exercici 6

Aquest és un exemple de distribució normal on caldrà concretar-ne els valors dels paràmetres, els quals extraiem de l'enunciat: $X \rightarrow N(m = 12, s = 0,1)$.

Recorda que per a calcular les probabilitats caldrà transformar qualsevol distribució normal en la $Z \rightarrow N(m = 0, s = 1)$ i resoldre aquestes probabilitats mitjançant la tipificació, que és una transformació que permet calcular una àrea equivalent sota la nova funció de densitat amb l'ajuda de les taules o el programari habitual. No es podrà calcular per integració, tal com es pot comprendre en observar l'expressió algebraica de la funció de densitat d'aquest model.

- a) Per a calcular la proporció de respostes estàndard considerem que l'interval de temps que es correspon és $P(11,8 < X < 12,2)$.
- b) Calcula la probabilitat de que un individu necessite menys de 10,5 minuts o més de 13 minuts per a omplir el qüestionari.

$P(X < 10,5)$ i $P(X > 13)$ donaran resultats sorprenents, ja que, encara que no estiguen massa allunyats de la mitjana (12 minuts), sí que són valors segurs o gens probables perquè la desviació típica també és menuda.

Exercici 7

Aquest exercici és un exemple d'aplicació del teorema central del límit, el qual permet aproximar el càlcul de probabilitats de la distribució binomial per a certs valors dels paràmetres ($np > 5$, $n(1 - p) > 5$) mitjançant el càlcul de probabilitats del model normal. Ara bé, cal comprovar que les hipòtesis es complisquen i calcular els valors dels paràmetres de la funció normal.

Per a aproximar el model binomial amb model normal:

$$\left. \begin{array}{l} X \rightarrow \text{Bin}(n, p) \\ np > 5, n(1 - p) > 5 \end{array} \right\} \Rightarrow X \approx N(\mu = np, \sigma^2 = np(1 - p))$$

Ens demanen $P(X \geq 60) = 1 - P(X < 60)$. Recordeu la tipificació.

Exercici 8

Aquest exercici és un exemple d'aplicació del teorema central del límit, el qual permet aproximar el càlcul de probabilitats de la distribució Poisson quan el valor del paràmetre $\lambda > 10$ amb el càlcul de probabilitats del model normal. Ara bé, cal comprovar que les hipòtesis es complisquen i calcular els valors dels paràmetres de la funció normal.

Considerarem la variable $X = \text{nre. de comandes rebudes en una hora}$, de la qual sabem que té de mitjana 30 i que es distribueix com una variable discreta Poisson. $X \rightarrow \text{Po}(\lambda = 30)$.

Per aproximar el model Poisson pel model normal:

$$\left. \begin{array}{l} X \rightarrow \text{Po}(\lambda) \\ \lambda > 10 \end{array} \right\} \Rightarrow X \approx N(\mu = \lambda, \sigma^2 = \lambda)$$

En el nostre cas, podem fer l'aproximació per una normal, ja que el paràmetre compleix la condició.

- a) Per a calcular la probabilitat de que en certa hora arriben entre 20 i 40 comandes:

$P(20 < X < 40)$. Recordeu la tipificació.

- b) Si volem atendre en una hora tots els encàrrecs que arriben en el 99 % de les ocasions, cal calcular quantes comandes ens poden arribar. És a dir, calcularem el percentil P_{99} de la variable normal tipificada que trobarem en les taules o en el programari que utilitzem habitualment.

$$\Phi(2,33) = 0,9901$$

A continuació plantejarem l'equació corresponent.

Darrerament, caldrà considerar que els treballadors tan sols poden atendre 3 comandes per hora, per la qual cosa, per a calcular el nombre de treballadors que necessitem, dividirem el nombre de comandes que cal atendre entre 3.

Ajudes de tipus 2

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de resoldre els exercicis i problema, després d'haver consultat l'ajuda de tipus 1.

Exercici 1

- a) Suposem que la funció següent $f(x)$ és la funció de densitat d'aquesta distribució:

$$f(x) = \begin{cases} cx^2 & \text{si } 1 \leq x \leq 3 \\ 0 & \text{en la resta} \end{cases}$$

Per a calcular el valor del paràmetre c cal aplicar la condició de que l'àrea tancada per sota de la corba de la funció de densitat és 1:

$$\int_1^3 cx^2 dx = c \left[\frac{x^3}{3} \right]_1^3 = 1 \rightarrow c = \frac{3}{26}$$

Així doncs, la funció de densitat queda:

$$f(x) = \begin{cases} \frac{3}{26}x^2 & \text{si } 1 \leq x \leq 3 \\ 0 & \text{en la resta} \end{cases}$$

- b) Creus que aquesta empresa ha atés 300 comandes algun dia? Justifica la teua resposta.

Pensa el valor de 300 comandes, que correspondria a $x = 0,3$, ja que s'expressen en milers, i fixa't en els valors del rang de la variable X (possibles valors de la variable)

- c) Calcula la probabilitat de que un dia qualsevol atenga entre 1.500 i 2.000 comandes, menys de 3.500, més de 3.000 i exactament 1.400.

$$P(1,5 \leq X \leq 2) = \int_{1,5}^2 \frac{3}{26}x^2 dx = 0,1779$$

$$P(X \leq 3,5) = ?$$

$$P(X > 3) = ?$$

$$P(X = 1,4) = ?$$

- d) Calcula la funció de distribució $F(x)$ i respon una altra vagada a les qüestions de l'apartat c amb aquesta funció. Representa-la.

Per a calcular la funció de distribució $F(x)$ per a qualsevol valor de x , cal que fem un canvi de variable en l'expressió de l'integrand:

$$F(x) = P(X \leq x) = \int_1^x \frac{3}{26} t^2 dt = \frac{3}{26} \left[\frac{t^3}{3} \right]_1^x = \frac{3}{26} \cdot \left[\frac{x^3}{3} - \frac{1}{3} \right] = \frac{x^3 - 1}{26}$$

Per la qual cosa, la seua expressió quedarà:

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{x^3 - 1}{26} & \text{si } 1 \leq x \leq 3 \\ 1 & \text{si } x > 3 \end{cases}$$

Amb aquesta funció podrem calcular:

$$P(1,5 \leq X \leq 2) = F(2) - F(1,5) = 0,1779$$

$$P(X \leq 3,5) = F(3,5)$$

$$P(X > 3) = 1 - F(3)$$

$$P(X = 1,4) = 0$$

e) Calcula la mitjana o esperança, la variància i la desviació típica. Interpreta-les.

Per a calcular l'esperança i la variància, vegem les integrals a plantejar i resoldre per Barrow.

$$E[X] = \int_1^3 x \cdot f(x) dx = \dots = 2,31$$

$$\text{Var}[X] = E[X^2] - E[X]^2 = 5,5846 - 2,31^2 = 0,2485$$

$$\text{Així doncs: } E[X^2] = \int_1^3 x^2 \cdot f(x) dx = \int_1^3 x^2 \cdot \frac{3}{26} x^2 dx = \dots$$

Exercici 2

Definim la variable X = percentatge d'ocupació (en tant per u). Es distribueix amb la següent funció de distribució:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 2x - x^2 & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

- a) Podries calcular la probabilitat de que un trajecte presente menys del 10 % d'ocupació, entre el 20 i 40 %, o més del 80 %?

Per tractar-se d'una *variable contínua*, el recompte de probabilitats és fàcil de fer a partir de la $F(x)$ de l'enunciat. Així:

$$P(X < 0,10) = F(0,10)$$

$$P(0,20 < X < 0,40) = F(0,40) - F(0,20)$$

$$P(X > 0,80) = 1 - P(X < 0,80) = ?$$

- b) Calcula la funció de densitat de la variable X .

Per a calcular la funció de densitat $f(x)$ serà suficient amb derivar la funció de distribució $F(x)$ en cadascun dels diferents intervals de la definició:

$$f(x) = \begin{cases} 2 - 2x & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en la resta} \end{cases}$$

- c) Calcula la mediana, la mitjana i la desviació típica d'aquesta distribució. Interpreta-les.

Per a calcular la *mediana* considerem que és un percentil, més concretament $P_{.50}$.

A tal fi, cal trobar el valor de x al qual correspon $F(x) = P(X \leq x) = 0,50$.

Trobarem que la mediana és $x_2 = 0,29$, és a dir, la meitat dels trajectes té menys del 29 % d'ocupació i només el 50 % dels trajectes té un major nivell d'ocupació.

Per a calcular la *mitjana*:

$E[X] = \int_0^1 x \cdot f(x) dx = \dots = \frac{1}{3}$. La mitjana del nivell d'ocupació dels trajectes és del 33 %.

Per a calcular la *desviació típica*:

$$\text{Var}[X] = E[X^2] - E[X]^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18} = 0,06$$

$$E[X^2] = \int_0^1 x^2 \cdot f(x) dx = \dots = \frac{1}{6}$$

$$\sigma = \sqrt{\frac{1}{18}} = 0,24$$

d) Calcula el percentatge d'ocupació que té, com a molt, el 95 % dels trajectes.

Ens demanen un percentil; caldrà decidir si interessa el P_5 o el P_{95} . Vegem:

$$F(x_0) = P(X \leq x_0) = 0,95$$

Plantejarem la següent equació i la resolldrem considerant els possibles valors del rang de la variable: $2x - x^2 = 0,95$.

Exercici 3

Aquest exercici, com es planteja a l'enunciat, és un exemple de model de distribució uniforme on tan sols ens interessa la distribució de les empreses els beneficis de les quals estan dintre de l'interval que va del 4 % al 7 %. Per la qual cosa els valors dels paràmetres són $a = 4$ i $b = 7$.

La funció de densitat ens mostrarà la distribució de les empreses del sector que tenen uns beneficis dintre d'aquest interval. Podem escriure la funció de densitat i observar que és una funció constant en el rang de la variable:

$$f(x) = \begin{cases} \frac{1}{7-4} & \text{si } 4 \leq x \leq 7 \\ 0 & \text{en la resta} \end{cases}$$

Per tant, podem calcular la probabilitat com l'àrea d'un rectangle.

a) Per a conèixer la proporció d'empreses amb beneficis inferiors al 6,7 %, farem els càlculs per integrals aplicant aquest procediment per a calcular l'àrea de la zona baix la corba de $f(x)$, que és la funció de densitat.

$$P(X \leq 6,7) = 1 - P(X > 6,7) = 1 - 0,1 = 0,9$$

$$P(X > 6,7) = \int_{6,7}^7 \frac{1}{3} dx = \dots = 0,1$$

b) Calcula les probabilitats $P(X \leq 5,2)$, $P(4,5 \leq X \leq 5,2)$.

$$P(X \leq 5,2) = \int_4^{5,2} f(x)dx = \dots = 0,4$$

$$P(4,5 \leq X \leq 5,2) = \int_{4,5}^{5,2} f(x)dx = \dots = 0,2333$$

c) Calcula l'esperança, la variància i la desviació típica. Interpreta-les.

Podem utilitzar aquestes expressions que podem trobar a la teoria de cada model de distribució:

$$E[X] = \frac{a+b}{2} \quad \text{Var}[X] = \frac{(b-a)^2}{12} \quad \sigma = \sqrt{\text{Var}[X]} =$$

Exercici 4

Definim la variable X = temps d'atenció dedicat a cadascuna de les peticions, i com que sabem que es distribueix com una variable exponencial, llavors coneixem l'expressió de la seua funció de densitat. Tan sols caldrà concretar el valor del seu paràmetre, el qual extraem de la mitjana:

$$E[X] = \frac{1}{\beta} = 8 \rightarrow \beta = \frac{1}{8}. \text{ Així:}$$

$$X \rightarrow \text{Exp}\left(\beta = \frac{1}{8}\right) \text{ i la seua funció de densitat és: } f(x) = \frac{1}{8}e^{-\frac{1}{8}x} \text{ si } x > 0$$

a) Calcula la probabilitat de que una atenció qualsevol necessite entre 5 i 12 minuts.

$$P(5 \leq X \leq 12) = \int_5^{12} \frac{1}{8}e^{-\frac{1}{8}x} dx = \dots = 0,3121$$

b) Calcula el temps mínim que podem assegurar que necessitarem en el 90 % de les atencions.

Aquesta qüestió requereix el càlcul d'un percentil. Caldrà decidir si ens demanen el P_{10} o si, pel contrari, ens demanen el P_{90} .

En el nostre cas, denotarem per «a» la duració mínima, en minuts, que necessitem trobar.

$P(X < a) = 0,90 \Leftrightarrow P(X < a) = 0,10$, per la qual cosa plantejarem la següent integral i l'equació per trobar el P_{10} :

$$P(X < a) = \int_0^a \frac{1}{8}e^{-\frac{1}{8}x} dx = \dots = 0,10$$

Després resoldrem l'equació exponencial amb l'ajut dels logaritmes:

$$1 - e^{-\frac{a}{8}} = 0,10.$$

- c) Calcula la probabilitat de que una consulta que ja porta 5 minuts atenent-se, necessite, encara, 7 minuts més.

Ens demanen el càlcul d'una probabilitat condicionada:

$$P\left(X > 12 \middle/ X > 5\right) = \frac{P(X > 12)}{P(X > 5)} = \frac{1 - P(X \leq 12)}{1 - P(X \leq 5)}$$

Calculem per separat les probabilitats del numerador i el denominador:

$$\bullet P(X \leq 12) = \int_0^{12} \frac{1}{8} e^{-\frac{1}{8}x} dx = \dots = 0,7769$$

$$\bullet P(X \leq 5) = \int_0^5 \frac{1}{8} e^{-\frac{1}{8}x} dx = \dots = 0,4647$$

Així:

$$P\left(X > 12 \middle/ X > 5\right) = \frac{P(X > 12)}{P(X > 5)} = \frac{1 - P(X \leq 12)}{1 - P(X \leq 5)} = 0,4168$$

Podem comprovar una propietat que es compleix en les probabilitats condicionades d'aquest model de distribució (exponencial) calculant la probabilitat de que la duració siga major de 7 minuts directament, com si el compliment previ de la condició no afectara la probabilitat del que ocorrerà a partir d'aquest moment:

$$P(X > 7) = 1 - P(X \leq 7) = 1 - 0,5831 = 0,4168$$

$$P(X \leq 7) = \int_0^7 \frac{1}{8} e^{-\frac{1}{8}x} dx = \dots = 0,5831$$

és a dir, en el nostre cas, $P\left(X > 12 \middle/ X > 5\right) = P(X > 7)$, i en general podem enunciar-la així:

$$P\left(X > s + t \middle/ X > s\right) = P(X > t)$$

- d) Calcula la probabilitat de que una tasca dure exactament 8 minuts.

Recordem que X és una contínua i, per tant, $P(X = 8) = 0$, ja que calculem probabilitats amb el concepte d'àrees tancades sota la corba d'un interval.

Exercici 5

Aquest exercici treballa la relació que existeix entre una situació que es pot ajustar al model de distribució Poisson i la distribució exponencial que es pot ajustar a la variable que mesura el temps que transcorre entre dos esdeveniments consecutius que es distribueixen com una Poisson.

Vegem aquesta relació en el nostre cas i com cal adjudicar els paràmetres de cada model, sempre atents a les unitats de mesura.

Vegem que Y (nombre de cridades al servei de grues) es pot ajustar a una distribució Poisson de mitjana 4 hores: $Y \rightarrow \text{Po}(\lambda = 4)$.

Així doncs, la variable $X = \text{temps (en hores)}$ que transcorre entre dos serveis consecutius, es distribueix com una variable exponencial de mitjana $E[X] = \frac{1}{\lambda} = 0,25$ hores per això $X = \text{temps entre dos serveis: Exp}(\lambda = 4)$.

- a) Calcula el temps mitjà d'espera entre dos serveis que requereixen grua. Calcula també la desviació típica.

$$E[X] = \frac{1}{\lambda} \quad \text{Var}[X] = \frac{1}{\lambda^2}$$

- b) Calcula la probabilitat de que una grua espere entre 10 i 20 minuts fins a rebre un nou avís.

Cal transformar els valors de la variable «temps» a hores, perquè la mitjana del nombre d'esdeveniments en la Poisson era de 4 serveis per hora.

$$\text{Així: } 10 \text{ min} = 0,17 \text{ hores} \qquad 20 \text{ min} = 0,33 \text{ hores}$$

$$P(0,17 < X < 0,33) = F(0,33) - F(0,17) = \dots = 0,2395$$

També pot plantejar-se a partir de la integral amb la funció de densitat:

$$P(0,17 < X < 0,33) = \int_{0,17}^{0,33} 4e^{-4x} dx = -\int_0^7 -4e^{-4x} dx = \dots = 0,2395$$

- c) Calcula la probabilitat de que no hi haja demanda del servei en un interval de més de 30 minuts.

Recordem: 30 min = 0,5 hores.

$$P(X > 0,5) = 1 - P(X < 0,5) = \dots = 0,1353$$

Exercici 6

Ja ens diu l'enunciat del problema que la variable $X =$ temps necessitat per a omplir el qüestionari, s'ajusta a una distribució normal i ens diu el valor dels paràmetres. Així:

$$X \rightarrow N(\mu = 12, \sigma = 0,1)$$

- a) Per a classificar els temps de resposta, denominarem estàndard aquells que han estat d'una duració que es desvia de la mitjana en 0,2 minuts com a molt. Així, per a calcular la seua proporció, notarem que l'interval de temps que correspon a les respostes estàndard és:

$P(11,8 < X < 12,2)$; tipifiquem:

$$P\left(\frac{11,8-12}{0,1} < \frac{X-\mu}{\sigma} < \frac{12,2-12}{0,1}\right) = P(-2 < Z < 2) = \varphi(2) - \varphi(-2) = \dots = 0,9546$$

- b) Per calcular la probabilitat de que un individu necessite menys de 10,5 minuts o més de 13 minuts per a omplir el qüestionari, plantegem les probabilitats:

$$P(X < 10,5); \text{ tipifiquem: } P(Z < -15) = \varphi(-15) = 1 - \varphi(15) = 0$$

$$P(X > 13) = 1 - P(X < 13) = 1 - 1 = 0; \text{ per tant: } P(X < 13) =$$

$$P\left(\frac{X-\mu}{\sigma} < \frac{13-12}{0,1}\right) = \dots = \varphi(10) = 1$$

Exercici 7

Per la teoria ja coneixem una aplicació del teorema central del límit que permet aproximar el càlcul de probabilitats de la distribució binomial per a certs valors dels paràmetres ($np > 5$, $n(1-p) > 5$) mitjançant el càlcul de probabilitats del model normal. Ara bé, cal comprovar que les hipòtesis es complisquen i calcular els valors dels paràmetres de la funció normal.

Per a aproximar el model binomial amb el model normal:

$$\left. \begin{array}{l} X \rightarrow \text{Bin}(n,p) \\ np > 5, n(1-p) > 5 \end{array} \right\} \Rightarrow X \approx N(\mu = np, \sigma^2 = np(1-p))$$

En el nostre cas, $X =$ nombre d'alumnes que han resolt el problema.

$$\left. \begin{array}{l} X \rightarrow \text{Bin}(n = 80, p = 0,55) \\ np = 80 \cdot 0,55 = 44 > 5, n(1-p) = 80 \cdot 0,45 = 36 > 5 \end{array} \right\} \Rightarrow X \approx N(\mu = 44, \sigma^2 = 19,8)$$

Per a resoldre la qüestió plantejada, calculem el 75 % de 80, és a dir, 60 alumnes.

$P(X < 60)$; tipificant: 0,9998

$P(X \geq 60) = 1 - P(X < 60) = 1 - 0,9998 = 0,0002$

NOTA

Quan aproximem una variable discreta per un altre model d'una variable contínua, seria més exacte fer la correcció per continuïtat, la qual consisteix a considerar que el valor de $f(x_0)$ en la variable discreta és equivalent, en la contínua, a l'àrea tancada baix la corba de $f(x)$ en l'interval $(x_0 - 0,5, x_0 + 0,5)$.

Exercici 8

Considerarem la variable $X =$ nombre de comandes rebudes en una hora, de la qual sabem que té de mitjana 30 i que es distribueix com una variable discreta Poisson. Així doncs: $X \rightarrow \text{Po}(\lambda = 30)$.

Per la teoria ja coneixem una aplicació del teorema central del límit que permet aproximar el càlcul de probabilitats de la distribució Poisson quan el valor del paràmetre $\lambda > 10$ mitjançant el càlcul de probabilitats del model normal. Ara bé, cal comprovar que les hipòtesis es compleixen i calcular els valors dels paràmetres de la funció normal.

Per a aproximar el model Poisson pel model normal:

$$\left. \begin{array}{l} X \rightarrow \text{Po}(\lambda) \\ \lambda > 10 \end{array} \right\} \Rightarrow X \approx N(\mu = \lambda, \sigma^2 = \lambda)$$

En el nostre cas podem fer l'aproximació per una normal, ja que el paràmetre compleix la condició:

$$\left. \begin{array}{l} X \rightarrow \text{Po}(\lambda = 30) \\ \lambda = 30 > 10 \end{array} \right\} \Rightarrow X \approx N(\mu = \lambda = 30, \sigma^2 = \lambda = 30)$$

$$\text{Així, } P(20 < X < 40) = P\left(\frac{20-30}{\sqrt{30}} < \frac{X-\mu}{\sigma} < \frac{40-30}{\sqrt{30}}\right) = \dots = 0,9328$$

En aquest cas, igual que en el problema anterior, seria convenient fer una correcció per continuïtat.

- b) Si volem atendre tots els encàrrecs que arriben en hora en el 99 % de les ocasions, cal calcular quantes comandes ens poden arribar. És a dir, calcularem el percentil P_{99} .

Denotem per «a» el nombre de comandes que compleixen $P(X \leq a) = 0,99$, és a dir, el percentil 99 de la variable normal tipificada que trobarem en les taules o en el programari que utilitzem habitualment.

$$\Phi(2,33) = 0,9901$$

$$P(X \leq a) = P\left(\frac{X - \mu}{\sigma} < \frac{a - 30}{\sqrt{30}}\right) = P\left(Z < \frac{a - 30}{\sqrt{30}}\right) = \Phi\left(\frac{a - 30}{\sqrt{30}}\right) = 0,99$$

Per la qual cosa, podem plantejar la igualtat $\frac{a - 30}{\sqrt{30}} = 2,33$.

Resoldrem l'equació per a trobar el valor de $a = 42,76 \approx 43$ comandes.

Així, afirmarem que en el 99 % de les hores es reben, com a molt, 43 encàrrecs i necessitarem $43/3 \approx 15$ treballadors per a atendre el servei, segons les condicions de l'enunciat.

Solucions

Exercici 1

Considerem una variable X que expressa el nombre de comandes (en unitats de milers) ateses diàriament per una empresa de missatgeria a nivell nacional. Suposem que la funció següent $f(x)$ és la seua funció de densitat que ens mostra la seua distribució.

$$f(x) = \begin{cases} cx^2 & \text{si } 1 \leq x \leq 3 \\ 0 & \text{en la resta} \end{cases}$$

- Calcula el valor del paràmetre c .
- Creus que aquesta empresa ha atés 300 comandes algun dia? Justifica la teua resposta.
- Calcula la probabilitat de que un dia qualsevol atenga entre 1.500 i 2.000 comandes, menys de 3.500, més de 3.000 i exactament 1.400.
- Calcula la funció de distribució $F(x)$ i respon una altra vegada a les qüestions de l'apartat c amb aquesta funció. Representa-la.
- Calcula la mitjana o esperança, la variància i la desviació típica. Interpreta-les.

Solució

- a) Suposem que la funció següent $f(x)$ és la funció de densitat d'aquesta distribució:

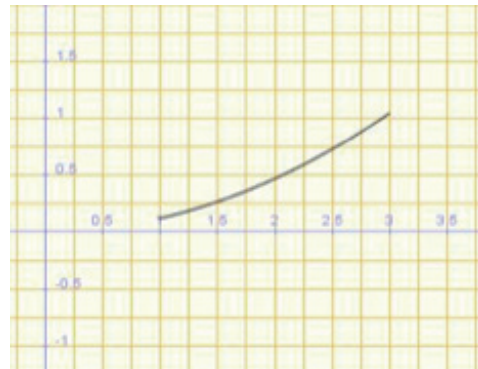
$$f(x) = \begin{cases} cx^2 & \text{si } 1 \leq x \leq 3 \\ 0 & \text{en la resta} \end{cases}$$

Per a calcular el valor del paràmetre « c » cal aplicar la condició de que l'àrea tancada per sota de la corba de la funció de densitat és 1.

$$\int_1^3 cx^2 dx = c \left[\frac{x^3}{3} \right]_1^3 = c \left[\frac{3^3}{3} - \frac{1^3}{3} \right] = c \cdot \left(9 - \frac{1}{3} \right) = c \cdot \frac{26}{3} = 1 \rightarrow c = \frac{3}{26}$$

Així doncs, la funció de densitat queda:

$$f(x) = \begin{cases} \frac{3}{26}x^2 & \text{si } 1 \leq x \leq 3 \\ 0 & \text{en la resta} \end{cases}$$



b) No atindrà mai 300 comandes, que correspondria a $x = 0,3$, ja que se expressen en milers, perquè $f(0,3) = 0$. El rang de X (possibles valors de la variable) és l'interval $(1, 3)$, per la qual cosa podem afirmar que l'empresa sempre atén entre 1.000 i 3.000 comandes. Si fem la gràfica de la funció de densitat, a més a més, podem veure que el més probable és facturar 3.000 comandes.

$$c) P(1,5 \leq X \leq 2) = \int_{1,5}^2 \frac{3}{26}x^2 dx = \frac{3}{26} \left[\frac{x^3}{3} \right]_{1,5}^2 = \frac{3}{26} \cdot \left[\frac{2^3}{3} - \frac{1,5^3}{3} \right] = 0,1779$$

$$P(X \leq 3,5) = 1 \text{ (sempre)}$$

$$P(X > 3) = 0 \text{ (mai, perquè està fora del rang)}$$

$$P(X = 1,4) = 0 \text{ (per ser la probabilitat d'un valor, no d'un interval).}$$

d) Per a calcular la funció de distribució $F(x)$, per a qualsevol valor de x , cal que fem un canvi de variable en l'expressió de l'integrand:

$$F(x) = P(X \leq x) = \int_1^x \frac{3}{26}t^2 dt = \frac{3}{26} \left[\frac{t^3}{3} \right]_1^x = \frac{3}{26} \cdot \left[\frac{x^3}{3} - \frac{1}{3} \right] = \frac{x^3 - 1}{26}$$

Per la qual cosa, la seua expressió quedarà:

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{x^3 - 1}{26} & \text{si } 1 \leq x \leq 3 \\ 1 & \text{si } x > 3 \end{cases}$$



Amb aquesta funció podrem calcular:

$$P(1,5 \leq X \leq 2) = F(2) - F(1,5) = \left(\frac{2^3 - 1}{26}\right) - \left(\frac{1,5^3 - 1}{26}\right) = 0,1779$$

$$P(X \leq 3,5) = F(3,5) = 1$$

$$P(X > 3) = 1 - F(3) = 1 - 1 = 0$$

$$P(X = 1,4) = 0 \quad (\text{per ser una funció continua})$$

e) Per a calcular l'esperança i la variància, vegem les integrals a plantejar i resoldre per Barrow:

$$E[X] = \int_1^3 x \cdot f(x) dx = \int_1^3 x \frac{3}{26} x^2 dx = \int_1^3 \frac{3}{26} x^3 dx = \frac{3}{26} \left[\frac{x^4}{4} \right]_1^3 = \frac{3}{26} \left(\frac{81}{4} - \frac{1}{4} \right) = 2,31$$

$$\text{Var}[X] = E[X^2] - E[X]^2 = 5,5846 - 2,31^2 = 0,2485$$

Així doncs,

$$E[X^2] = \int_1^3 x^2 \cdot f(x) dx = \int_1^3 x^2 \frac{3}{26} x^2 dx = \int_1^3 \frac{3}{26} x^4 dx = \frac{3}{26} \left[\frac{x^5}{5} \right]_1^3 =$$

$$= \frac{3}{26} \left(\frac{243}{5} - \frac{1}{5} \right) = 5,5846$$

$$\sigma = \sqrt{0,2485} = 0,4985$$

Amb aquestes dades, podem interpretar que la mitjana és atendre 2.310 comandes diàriament, amb una desviació típica de 499 comandes, la qual cosa voldrà dir que aquesta és la mitjana de les desviacions del nombre de comandes ateses cada dia, i l'anterior és esperança de 2.310 comandes.

Si trobem el coeficient de variació, podem interpretar aquesta dispersió en termes relatius o de percentatges:

$$CV = \frac{\sigma}{|\mu|} = \frac{0,4985}{2,31} = 0,22 \rightarrow 22 \% \text{ de variabilitat}$$

Exercici 2

Per a justificar l'eliminació de cert servei de transport urbà en els trajectes de l'any s'han anotat els percentatges d'ocupació de tots els viatges realitzats. Així, la variable $X =$ percentatge d'ocupació (en tant per u), es distribueix amb la següent funció de distribució:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 2x - x^2 & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

- Podries calcular la probabilitat de que un trajecte presente menys del 10 % d'ocupació, entre el 20 i 40 %, o més del 80 %?
- Calcula la funció de densitat de la variable X .
- Calcula la mediana, la mitjana i la desviació típica d'aquesta distribució. Interpreta-les.
- Calcula el percentatge d'ocupació que té, com a molt, el 95 % dels trajectes.

Solució

- a) Considerarem que la variable $X =$ percentatge d'ocupació (en tant per u) en els trajectes del darrer any. Per tractar-se d'una *variable contínua*, el recompte de probabilitats és fàcil de fer a partir de la $F(x)$ de l'enunciat. Així:

$$P(X < 0,10) = F(0,10) = 2 \cdot 0,10 - 0,10^2 = 0,19$$

$$P(0,20 < X < 0,40) = F(0,40) - F(0,20) = (2 \cdot 0,40 - 0,40^2) - (2 \cdot 0,20 - 0,20^2) = 0,28$$

$$P(X > 0,80) = 1 - P(X < 0,80) = 1 - (2 \cdot 0,80 - 0,80^2) = 1 - 0,96 = 0,04$$

- b) Per a calcular la funció de densitat $f(x)$ serà suficient amb derivar la funció de distribució $F(x)$ en cadascun dels diferents intervals de definició:

$$f(x) = \begin{cases} 2 - 2x & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en la resta} \end{cases}$$

- c) Per calcular la *mediana* considerem que és un percentil, més concretament P_{50} .

A tal fi, cal trobar el valor de x al qual correspon $F(x) = P(X \leq x) = 0,50$.

Plantejant l'equació:

$$2x - x^2 = 0,5 \rightarrow x^2 - 2x + 0,5 = 0 \rightarrow x_1 = 1,71, x_2 = 0,29$$

Com que $x_1 = 1,71$ i està fora del rang de la variable, la descartem com a solució possible.

Direm que la mediana és $x_2 = 0,29$, és a dir, la meitat dels trajectes té menys del 29 % d'ocupació i tan sols el 50 % dels trajectes té un major nivell d'ocupació.

Per a calcular la *mitjana*:

$$E[X] = \int_0^1 x \cdot f(x) dx = \int_0^1 x(2 - 2x) dx = \int_0^1 2x - 2x^2 dx = 2 \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_0^1 = 2 \left(\frac{1}{2} - \frac{1}{3} \right) = \frac{1}{3} = 0,33$$

La mitjana del nivell d'ocupació dels trajectes és del 33 %.

Per a calcular la *desviació típica*:

$$\text{Var}[X] = E[X^2] - E[X]^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18} = 0,06$$

$$E[X^2] = \int_0^1 x^2 \cdot f(x) dx = \int_0^1 x^2(2 - 2x) dx = 2 \int_0^1 x^2 - x^3 dx = 2 \left[\frac{x^3}{3} - \frac{x^4}{4} \right]_0^1 = 2 \left(\frac{1}{3} - \frac{1}{4} \right) = \frac{1}{6}$$

$$\sigma = \sqrt{\frac{1}{18}} = 0,24$$

d) Ens demanen un percentil; caldrà decidir si interessa el P_5 o el P_{95} . Vegem:

$$F(x_0) = P(X \leq x_0) = 0,95$$

Plantejarem la següent equació i la resoldrem:

$$2x - x^2 = 0,95 \rightarrow x^2 - 2x + 0,95 = 0 \rightarrow x_1 = 1,22, x_2 = 0,78$$

Com que $x_1 = 1,22$ i està fora del rang de la variable, la descartem.

Agafarem $x_2 = 0,78$ i podrem interpretar que el 95 % dels trajectes té menys d'un 78 % d'ocupació.

Exercici 3

Estem fent un estudi entre les empreses d'un cert sector que van vendre beneficis l'any passat. S'observa una distribució uniforme en el sector que va del 4 % al 7 % de beneficis que volem estudiar amb més profunditat.

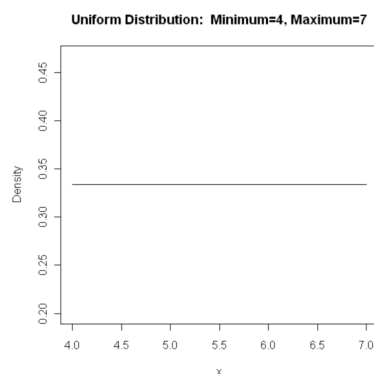
- Si considerem que una empresa amb beneficis inferiors al 6,7 % no pot assegurar el seu manteniment de plantilles, calcula la proporció d'empreses que podem estimar, dins d'aquest sector crític que estem analitzant, que està en risc d'acomiarar treballadors.
- Calcula les probabilitats $P(X \leq 5,2)$, $P(4,5 \leq X \leq 5,2)$.
- Calcula l'esperança, la variància i la desviació típica. Interpreta-les.

Solució

Pel que es diu a l'enunciat, la funció de densitat tan sols ens mostra la distribució de les empreses del sector que tenen uns beneficis d'entre el 4 % i el 7 % i que seran l'objectiu del nostre treball. Podem escriure'n la funció de densitat i observar que és una funció constant en el rang de la variable:

$$f(x) = \begin{cases} \frac{1}{7-4} & \text{si } 4 \leq x \leq 7 \\ 0 & \text{en la resta} \end{cases}$$

Per tant, podem calcular la probabilitat com l'àrea d'un rectangle. En aquest enunciat ens han indicat, expressament, que el model de distribució és uniforme, però també ens haurien pogut dir que es tractava d'una distribució constant a l'interval $[4, 8]$ amb tots els valors que agafa la variable equiprobables.



$$a) P(X \leq 6,7) = 1 - P(X > 6,7) = 1 - 0,1 = 0,9$$

$$P(X > 6,7) = \int_{6,7}^7 \frac{1}{3} dx = \frac{1}{3}(7 - 6,7) = \frac{1}{3} \cdot 0,3 = 0,1$$

Així, podem interpretar que la proporció d'empreses que podem estimar dins d'aquest sector crític que estem analitzant, que està en risc d'acomiadar treballadors, és del 90 %.

$$b) P(X \leq 5,2) = \int_4^{5,2} \frac{1}{3} dx = \frac{1}{3}(5,2 - 4) = \frac{1}{3} \cdot 1,2 = 0,4$$

$$P(4,5 \leq X \leq 5,2) = \int_{4,5}^{5,2} \frac{1}{3} dx = \frac{1}{3}(5,2 - 4,5) = \frac{1}{3} \cdot 0,7 = 0,2333$$

c) Podem utilitzar les expressions d'aquestes que trobarem a la teoria de cada model de distribució:

$$E[X] = \frac{a+b}{2} = \frac{4+7}{2} = 5,5 \%$$

$$\text{Var}[X] = \frac{(b-a)^2}{12} = \frac{(7-4)^2}{12} = \frac{9}{12} = \frac{3}{4} \text{ i } \sigma = \sqrt{\text{Var}[X]} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2} = 0,87$$

Que en el sector voldrà dir que en aquest grup d'empreses la mitjana dels beneficis és del 5,5 %, amb una variabilitat de 0,87 %.

Exercici 4

Volem planificar l'assignació de tasques en un cert departament d'atenció al públic. Per tractar-se majorment de demandes de documentació, hem estudiat que el temps d'atenció de cadascuna de les peticions segueix una distribució exponencial amb una mitjana de 8 minuts.

- Calcula la probabilitat de que una atenció qualsevol necessite entre 5 i 12 minuts.
- Calcula el temps mínim que podem assegurar que necessitarem en el 90 % de les atencions.
- Calcula la probabilitat de que una consulta que ja porta 5 minuts atenent-se necessite, encara, 7 minuts més.
- Calcula la probabilitat de que una tasca dure exactament 8 minuts.

Solució

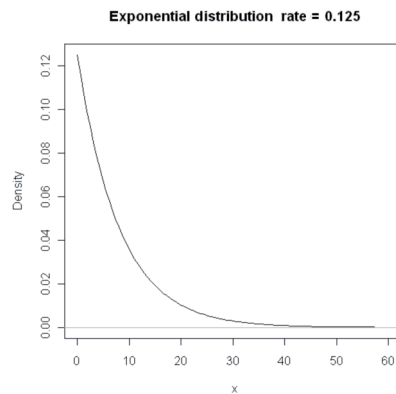
Com diu l'enunciat, la variable $X =$ temps d'atenció dedicat a cadascuna de les peticions, com que ens diu que es distribueix com una variable exponencial, coneixem l'expressió de la seua funció de densitat. Tan sols caldrà concretar el valor del seu paràmetre, el qual extraïem de la mitjana:

$$E[X] = \frac{1}{\beta} = 8 \rightarrow \beta = \frac{1}{8}. \text{ Així:}$$

$X \rightarrow \text{Exp} \left(\beta = \frac{1}{8} \right)$ i la seua funció de densitat és:

$$f(x) = \frac{1}{8} e^{-\frac{1}{8}x} \text{ si } x > 0$$

Podem representar la funció de densitat per a que ens facilite la comprensió del fenomen i les interpretacions:



a) $P(5 \leq X \leq 12) =$

$$\int_5^{12} \frac{1}{8} e^{-\frac{1}{8}x} dx = -\int_5^{12} -\frac{1}{8} e^{-\frac{1}{8}x} dx = -\left[e^{-\frac{1}{8}x} \right]_5^{12} = -\left(e^{-\frac{12}{8}} - e^{-\frac{5}{8}} \right) = 0,3121$$

Podem interpretar que el 31 % de les atencions han requerit entre 5 i 12 minuts.

b) Aquesta qüestió requereix el càlcul d'un percentil. Recordeu que és convenient no precipitar-se i distingir amb la lectura atenta de l'enunciat si ens demanen el P_{10} o si pel contrari ens demanen el P_{90} .

En el nostre cas, denotarem per «a» la duració mínima, en minuts, que necessitem trobar.

$P(X \geq a) = 0,90 \Leftrightarrow P(X < a) = 0,10$, per la qual cosa plantejarem la següent integral i l'equació per a trobar el P_{10} :

$P(X < a) =$

$$\int_0^a \frac{1}{8} e^{-\frac{1}{8}x} dx = -\int_0^a -\frac{1}{8} e^{-\frac{1}{8}x} dx = -\left[e^{-\frac{1}{8}x} \right]_0^a = -\left(e^{-\frac{a}{8}} - e^0 \right) = 1 - e^{-\frac{a}{8}} = 0,10$$

I resoldrem l'equació:

$$1 - e^{-\frac{a}{8}} = 0,10 \rightarrow 0,90 = e^{-\frac{a}{8}} \rightarrow -\frac{a}{8} = \ln 0,90 \rightarrow$$

$$a = -8 \cdot \ln 0,90 = 0,8429 \text{ minuts} = 50 \text{ segons.}$$

Podem assegurar que el 90 % de les atencions tenen una duració superior a 50 segons.

c) Ens demanen el càlcul d'una probabilitat condicionada:

$$P\left(X > 12 \middle/ X > 5\right) = \frac{P(X > 12)}{P(X > 5)} = \frac{1 - P(X \leq 12)}{1 - P(X \leq 5)}$$

Calculem per separat les probabilitats del numerador i el denominador:

• $P(X \leq 12) =$

$$\int_0^{12} \frac{1}{8} e^{-\frac{1}{8}x} dx = -\int_0^{12} -\frac{1}{8} e^{-\frac{1}{8}x} dx = -\left[e^{-\frac{1}{8}x} \right]_0^{12} = -\left(e^{-\frac{12}{8}} - e^0 \right) = 1 - e^{-\frac{12}{8}} = 0,7769$$

• $P(X \leq 5) =$

$$\int_0^5 \frac{1}{8} e^{-\frac{1}{8}x} dx = -\int_0^5 -\frac{1}{8} e^{-\frac{1}{8}x} dx = -\left[e^{-\frac{1}{8}x} \right]_0^5 = -\left(e^{-\frac{5}{8}} - e^0 \right) = 1 - e^{-\frac{5}{8}} = 0,4647$$

Així:

$$P\left(X > 12 \middle/ X > 5\right) = \frac{P(X > 12)}{P(X > 5)} = \frac{1 - P(X \leq 12)}{1 - P(X \leq 5)} = \frac{1 - 0,7769}{1 - 0,4647} = 0,4168$$

Podem comprovar una propietat que es compleix en les probabilitats condicionades d'aquest model de distribució (exponencial) calculant la probabilitat que la duració siga major de 7 minuts directament, com si el compliment previ de la condició no afectara a la probabilitat del que ocorrerà a partir d'aquest moment:

$$P(X > 7) = 1 - P(X \leq 7) = 1 - 0,5831 = 0,4168$$

↓

$$P(X \leq 7) =$$

$$\int_0^7 \frac{1}{8} e^{-\frac{1}{8}x} dx = -\int_0^7 -\frac{1}{8} e^{-\frac{1}{8}x} dx = -\left[e^{-\frac{1}{8}x} \right]_0^7 = -\left(e^{-\frac{7}{8}} - e^0 \right) = 1 - e^{-\frac{7}{8}} = 0,5831$$

És a dir, en el nostre cas, $P\left(X > 12 \middle/ X > 5\right) = P(X > 7)$, i en general podem enunciar-la així:

$$P\left(X > s + t \middle/ X > s\right) = P(X > t)$$

NOTA

També podem calcular aquestes probabilitats a partir de l'expressió de la funció de distribució del model exponencial, la qual podem consultar a la teoria o obtenir-la aplicant la seua definició:

$$F(x) = \int_0^x \beta e^{-\beta t} dt = \begin{cases} 1 - e^{-\beta x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

d) Calcula la probabilitat que una tasca duri exactament 8 minuts.

Recordem, que per fer el tractament de la variable X , amb el model d'una continua, $P(X = 8) = 0$, ja que calculem probabilitats amb el concepte d'àrees tancades sota la corba en un interval.

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

Exercici 5

En un servei d'atenció en carretera de 24 hores, la mitjana de serveis que requereixen l'atenció amb grua és de 4 incidències per hora.

Considerem la variable que estudia el temps transcorregut entre dues eixides de la grua en el servei.

- Calcula el temps mitjà d'espera entre dos serveis que requereixen grua. Calcula'n també la desviació típica.
- Calcula la probabilitat de que la grua espere entre 10 i 20 minuts fins a rebre un nou avís.
- Calcula la probabilitat de que no hi haja demanda del servei en un interval de més de 30 minuts.

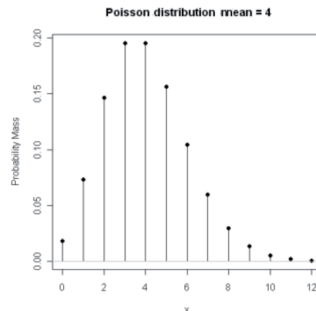
Solució

Aquest exercici treballa la relació que existeix entre una situació que es pot ajustar al model de distribució Poisson (cridades de telèfon a una central cada hora, nombre de pacients atesos en urgències en un hospital cada dia, nombre de *mails* publicitaris que reps cada hora, nombre d'accidents que es produeixen cada mes en un «punt negre» de certa carretera, etc.) i la distribució exponencial que es pot

ajustar a la variable que mesura el temps que transcorre entre dos esdeveniments consecutius en qualsevol de les situacions anteriors.

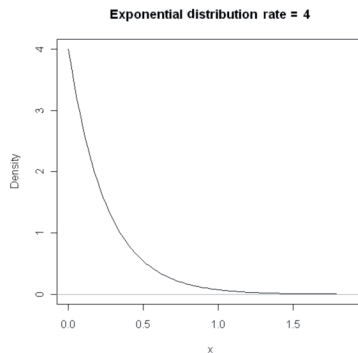
Vegem aquesta relació en el nostre cas i com cal adjudicar els paràmetres de cada model, sempre atents a les unitats de mesura.

Vegem que Y (nombre de cridades al servei de grues) es pot ajustar a una distribució Poisson de mitjana 4 hores: $Y \rightarrow \text{Po}(\lambda = 4)$.



Així doncs, la variable $X =$ temps (en hores) que transcorre entre dos serveis consecutius, es distribueix com una variable exponencial de mitjana $E[X] = \frac{1}{\lambda} = 0,25$ hores, per això $X =$ temps entre dos serveis: ($\beta = \lambda = 4$).

La gràfica de la funció de densitat ens pot ajudar a interpretar els resultats:



$$a) E[X] = \frac{1}{\lambda} = 0,25 \text{ hores} \quad \text{Var}[X] = \frac{1}{\beta^2} = 0,0625 \text{ hores}^2 \quad \sigma = 0,25 \text{ hores}$$

Podem interpretar que el temps mitjà que transcorre entre dos serveis de grua consecutius és de 0,25 hores = 15 minuts i que la seua variabilitat és també de 15 minuts de mitjana.

b) Cal transformar els valors de la variable «temps» a hores, perquè la mitjana del nombre d'esdeveniments en la Poisson era de 4 serveis per hora.

Així: 10 min = 0,17 hores

20 min = 0,33 hores

$$P(0,17 < X < 0,33) = F(0,33) - F(0,17) = (1 - e^{-4 \cdot 0,33}) - (1 - e^{-4 \cdot 0,17}) = 0,2395$$

També pot plantejar-se a partir de la integral amb la funció de densitat:

$$\begin{aligned} P(0,17 < X < 0,33) &= \\ &= \int_{0,17}^{0,33} 4e^{-4x} dx = -\int_0^7 -4e^{-4x} dx = -\left[e^{-4x} \right]_{0,17}^{0,33} = -(e^{-4 \cdot 0,33} - e^{-4 \cdot 0,17}) = 0,2395 \end{aligned}$$

Podem afirmar que quasi el 24 % de les ocasions el temps que transcorre entre dues cridades consecutives a la grua està dintre del interval de 10 a 20 minuts.

c) Recordem que 30 min = 0,5 hores.

$$P(X > 0,5) = 1 - P(X < 0,5) = 1 - F(0,5) = 1 - (1 - e^{-4 \cdot 0,5}) = 0,1353$$

És a dir, que una espera entre serveis superior a 30 minuts tan sols ocorre en el 13,5 % de les ocasions.

Exercici 6

En un departament d'estadística s'està dissenyant un qüestionari que haurà d'implementar-se en una mostra gran. Per a preveure les necessitats de temps i personal que necessitem, hem passat el qüestionari a 200 individus seleccionats a l'atzar i hem anotat el temps que han emprat per a omplir-lo. S'observa que els resultats d'aquest mostratge inicial s'ajusten a un model de distribució normal de mitjana 12 i desviació típica de 0,1 minuts.

- Per a classificar els temps de resposta, denominarem «estàndard» aquelles que han estat d'una duració que es desvien de la mitjana com a molt, en 0,2 minuts. Calcula la proporció de respostes estàndard que hi haurà. Si la mostra real que ompli el qüestionari dissenyat és de 2.000 individus, calcula el nombre de respostes d'aquest tipus que puguem esperar.
- Calcula la probabilitat de que un individu necessite menys de 10,5 minuts o més de 13 minuts per a omplir el qüestionari.

Solució

Ja ens diu l'enunciat del problema que la variable $X =$ temps necessitat per a omplir el qüestionari, s'ajusta a una distribució normal i ens diu el valor dels paràmetres. Així:

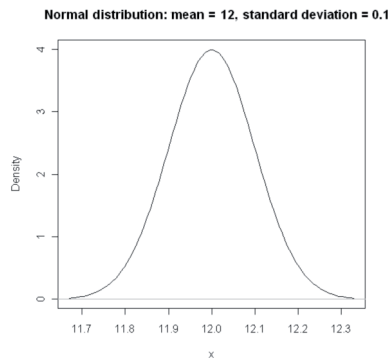
$$X \rightarrow N(\mu = 12, \sigma = 0,1)$$

a) Considerem que l'interval de temps que correspon a les respostes estàndard és:

$$P(11,8 < X < 12,2) =$$

$$= P\left(\frac{11,8-12}{0,1} < \frac{X-\mu}{\sigma} < \frac{12,2-12}{0,1}\right) = P(-2 < Z < 2) = \varphi(2) - \varphi(-2) =$$

$$= \varphi(2) - (1 - \varphi(2)) = 0,9773 - (1 - 0,9773) = 0,9546$$



b) $P(X < 10,5) =$

$$P\left(\frac{X-\mu}{\sigma} < \frac{10,5-12}{0,1}\right) = P(Z < -15) = \varphi(-15) = 1 - \varphi(15) = 1 - 1 = 0$$

$$P(X > 13) = 1 - P(X < 13) = 1 - 1 = 0$$

$$P(X < 13) = P\left(\frac{X-\mu}{\sigma} < \frac{13-12}{0,1}\right) = P(Z < 10) = \varphi(10) = 1$$

Podem advertir que quan fem la tipificació sovint resulten valors de la variable $Z \rightarrow N(\mu = 0, s = 1)$ que estan tan allunyats de la mitjana que els assignem probabilitats 0, o 1, per aproximació, ja que habitualment treballem amb 4 o 5 xifres decimals.

Exercici 7

En certa classe d'estudiants d'estadística hi ha 80 alumnes, dels quals pensem que el 55 % ja coneix suficientment el programari de les pràctiques. Si en aquest moment plantegem un exercici per a resoldre'l utilitzant l'esmentat programari, calcula la probabilitat de que, al menys, el resolga el 75 % dels alumnes.

Solució

Per la teoria ja coneixem una aplicació del teorema central del límit que permet aproximar el càlcul de probabilitats de la distribució binomial per a certs valors dels paràmetres ($np > 5$, $n(1-p) > 5$) mitjançant el càlcul de probabilitats del model normal. Ara bé, cal comprovar que les hipòtesis es complisquen i calcular els valors dels paràmetres de la funció normal.

Per a aproximar el model binomial amb el model normal:

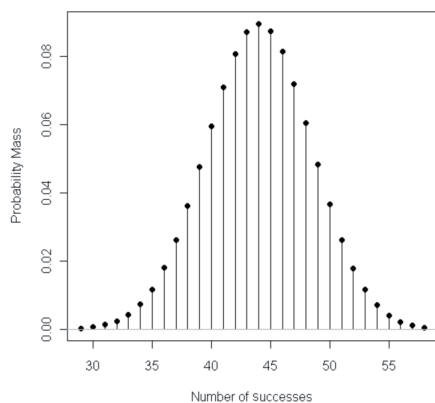
$$\left. \begin{array}{l} X \rightarrow \text{Bin}(n,p) \\ np > 5, n(1-p) > 5 \end{array} \right\} \Rightarrow X \approx N(\mu = np, \sigma^2 = np(1-p))$$

En el nostre cas, $X =$ nombre d'alumnes que han resolt el problema.

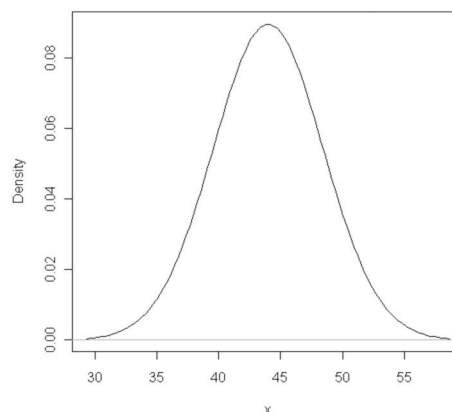
$$\left. \begin{array}{l} X \rightarrow \text{Bin}(n = 80, p = 0,55) \\ np = 80 \cdot 0,55 = 44 > 5, n(1-p) = 80 \cdot 0,45 = 36 > 5 \end{array} \right\} \Rightarrow X \approx N(\mu = 44, \sigma^2 = 19,8)$$

Podem veure i comparar les gràfiques de la funció $f(x)$ d'ambdós models:

Binomial distribution binomial trials = 80, probability of success



Normal distribution mean = 44, standard deviation=4.45



Per a resoldre la qüestió plantejada, calculem el 75 % de 80, és a dir, 60 alumnes.

$$P(X < 60) = P\left(\frac{X - \mu}{\sigma} < \frac{60 - 44}{\sqrt{19,8}}\right) = P(Z < 3,6) = \Phi(3,6) = 0,9998$$

$$P(X \geq 60) = 1 - P(X < 60) = 1 - 0,9998 = 0,0002$$

Quasi impossible, com es pot veure a la gràfica superior.

NOTA

Quan aproximem una variable discreta per un altre model d'una variable contínua, seria més exacte fer la correcció per continuïtat, la queal consisteix a considerar que el valor de $f(x_0)$ en la variable discreta és equivalent, en la contínua, a l'àrea tancada baix la corba de $f(x)$ en l'interval $(x_0 - 0,5, x_0 + 0,5)$. Així, estaria millor plantejat de la següent manera:

$$P(X < 60) = P\left(\frac{X - \mu}{\sigma} < \frac{59,5 - 44}{\sqrt{19,8}}\right)$$

$$P(X \leq 60) = P\left(\frac{X - \mu}{\sigma} < \frac{60,5 - 44}{\sqrt{19,8}}\right)$$

Malgrat aquesta explicació, en aquest treball considerem que no cal abordar amb profunditat aquest detall.

Exercici 8

En certa empresa de venda *on line* estan rebent una mitjana de 30 comandes per hora. Se suposa que ho fan de manera aleatòria, proporcional al temps que transcorre i amb les característiques que permeten assignar el model de Poisson a aquesta distribució.

- Calcula la probabilitat de que en certa hora arriben entre 20 i 40 comandes.
- Si volem atendre tots els encàrrecs que arriben a aquesta hora, calcula els treballadores lliures que necessitem en una hora per atendre el 99 % de les comandes (suposem que un treballador pot despatxar 3 comandes per hora).

Solució

- Considerarem la variable $X =$ nombre de comandes rebudes en una hora, de la qual sabem que té de mitjana 30 i que es distribueix com una variable discreta Poisson. Així doncs: $X \rightarrow \text{Po}(\lambda = 30)$.

Per la teoria ja coneixem una aplicació del teorema central del límit que permet aproximar el càlcul de probabilitats de la distribució Poisson quan el valor del paràmetre $l > 10$ mitjançant el càlcul de probabilitats del model normal. Ara bé, cal comprovar que les hipòtesis es complisquen i calcular els valors dels paràmetres de la funció normal.

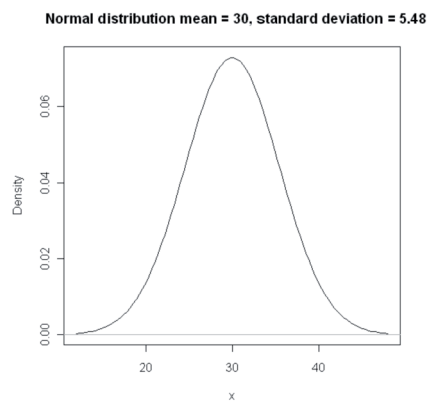
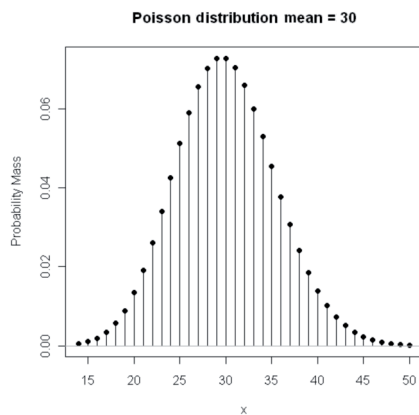
Per a aproximar el model Poisson amb el model normal:

$$\left. \begin{array}{l} X \rightarrow \text{Po}(\lambda) \\ \lambda > 10 \end{array} \right\} \Rightarrow X \approx N(\mu = \lambda, \sigma^2 = \lambda)$$

En el nostre cas, podem fer l'aproximació per una normal, ja que el paràmetre compleix la condició:

$$\left. \begin{array}{l} X \rightarrow \text{Po}(\lambda = 30) \\ \lambda = 30 > 10 \end{array} \right\} \Rightarrow X \approx N(\mu = \lambda = 30, \sigma^2 = \lambda = 30)$$

Si comparem les gràfiques, veurem les seues distribucions:



Així, $P(20 < X < 40) =$

$$\begin{aligned} &= P\left(\frac{20-30}{\sqrt{30}} < \frac{X-\mu}{\sigma} < \frac{40-30}{\sqrt{30}}\right) = P(-1,83 < Z < 1,83) = \varphi(1,83) - \varphi(-1,83) = \\ &= \varphi(1,83) - (1 - \varphi(1,83)) = 0,9664 - (1 - 0,9664) = 0,9328 \end{aligned}$$

En el 93 % de les hores s'atén un nombre de comandes comprés entre 20 i 40.

En aquest cas, igual que en el problema anterior, seria convenient fer una correcció per continuïtat (vegeu nota al final del problema anterior).

- b) Denotem per «a» el nombre de comandes que compleixen $P(X \leq a) = 0,99$, és a dir, el percentil 99 de la variable normal tipificada i que trobarem en les taules o en el programari que utilitzem habitualment.

$$\varphi(2,33) = 0,9901$$

$$P(X \leq a) =$$

$$P\left(\frac{X-\mu}{\sigma} < \frac{a-30}{\sqrt{30}}\right) = P\left(Z < \frac{a-30}{\sqrt{30}}\right) = \Phi\left(\frac{a-30}{\sqrt{30}}\right) = 0,99$$

Per la qual cosa, podem plantejar la igualtat:

$$\frac{a-30}{\sqrt{30}} = 2,33 \rightarrow a = 30 + 2,33 \cdot \sqrt{30} = 42,76 \approx 43 \text{ comandes.}$$

Així, afirmarem que en el 99 % de les hores es reben, com a molt, 43 encàrrecs i que necessitarem $43/3 \approx 15$ treballadors per a atendre el servei, segons les condicions de l'enunciat.

UNITAT 4

Introducció a la inferència 1: intervalls de confiança

Introducció teòrica

Tot estudi estadístic té per objectiu analitzar una determinada característica en una població. Seria ideal poder observar i mesurar aquesta característica en tots els individus, encara que sovint aquest plantejament global no és possible.

Imaginem, per exemple, els següents estudis estadístics:

- Es vol conèixer l'alçada mitjana de la població major d'edat de la nostra ciutat. Requereix un temps i unes despeses econòmiques excessives prendre la mesura a tots els ciutadans.
- Per a fer un control de qualitat del procés de fabricació en una factoria de mistos, caldria conèixer-ne la proporció de productes defectuosos. Seria un desgavell provar cadascun dels elements, ja que es destruiria la producció.

En ambdós casos cal estudiar únicament una part de la població, la qual ha de ser seleccionada amb criteris adequats per tal d'assegurar-se que es tracta d'una *mostra representativa* de la població total i de les seues particularitats internes. Aquests criteris es presentaran en el següent apartat.

En la pràctica és molt freqüent haver de recórrer a una mostra per tal d'inferir dades d'una població per alguns dels motius següents:

- La població és excessivament nombrosa; tota la població major d'edat de la nostra ciutat.
- La població és molt difícil o impossible de controlar; el nombre d'individus de certa espècie d'aus en un determinat parc natural.
- El procés de mesurament és destructiu; la comprovació de mistos defectuosos en una producció.
- Es desigen conèixer els resultats ràpidament i es tardaria massa temps en prendre les mesures de tots; el sondeig d'opinió electoral.

Població i mostra. Tipus de mostreig

Població o *Univers* és el conjunt de tots els individus objecte del nostre estudi. *Mostra* és un subconjunt extret de la població; el seu estudi serveix per a inferir característiques de tota la població.

L'elecció de la mostra es diu *mostreig*. Un mètode molt eficaç per aconseguir una mostra representativa és que aquesta haja estat triada aleatòriament, és a dir, a l'atzar. Per contra, si l'elecció és subjectiva, els prejudicis de qui fa l'elecció es projecten en el resultat de la mostra, la qual reflectirà el que aquesta persona creu que és la realitat (*mostreig opinàtic*) i no s'aconseguiran mostres representatives.

Es diu que un mostreig és aleatori quan tots els individus de la mostra s'elegeixen a l'atzar, de manera que tots els individus de la població tenen, *a priori*, la mateixa probabilitat de ser triats.

Al llarg d'aquest epígraf se suposarà que es té una població de N elements de la qual es vol extraure una mostra de n elements. Es comentarà a continuació com es pot realitzar el mostreig perquè proporcione mostres representatives.

Mostreig aleatori simple

És el tipus de mostreig més senzill i en el qual es basen tots els altres.

Descripció	Exemple
<ul style="list-style-type: none"> – Numerem els elements de la població de l'1 a N. – Efectuem un sorteig per a seleccionar els n elements que ha de contenir la mostra. 	<ul style="list-style-type: none"> – Numerem els alumnes seguint l'ordre alfabètic dels seus cognoms. – Amb la calculadora, anatem una seqüència de nombres aleatoris de dues xifres i agafarem els primers 5 nombres menors o iguals que 36. <p style="text-align: center;">91, 31, 21, 56, 48, 10, 89, 34, 11</p> <p>La mostra està formada pels alumnes que corresponen amb els nombres 31, 21, 10, 34 i 11.</p>

Aquest sorteig pot fer-se de moltes maneres: extraure paperetes d'un caixó, extraure boles numerades d'una urna, etc. També podem ajudar-nos de la generació de nombres aleatoris fent servir la calculadora (tecla $RAN\#$) o l'ordinador.

Si després de cada extracció l'individu seleccionat de la població pot tornar a ser-ho, es diu que és un mostreig aleatori *amb reemplaçament*. En cas contrari, és un mostreig aleatori *sense reemplaçament*.

Mostreig aleatori sistemàtic

La tria dels elements assegura una forma uniforme d'escollir els elements al llarg de la llista de la població.

Descripció	Exemple
<ul style="list-style-type: none"> – Numerem els elements de la població de l'1 a N. – Triem per sorteig un individu qualsevol N_0. – Calculem k, l'enter més proper al <i>coeficient de d'elevació</i> $\frac{N}{n}$. – Seleccionem l'individu N_0 i els següents de k en k a partir d'ell. Cal tindre en compte que en sobrepassar N hem de continuar amb el començament de la llista. 	<ul style="list-style-type: none"> – Numerem els alumnes seguint l'ordre alfabètic dels seus cognoms. – Obtenim per sorteig un número. Imaginem que és el 21. – Trobem el <i>coeficient de d'elevació</i> $\frac{N}{n} = \frac{36}{5} = 7,2 \rightarrow k = 7$ – Comencem per 21 i anem triant la mostra, agafant els elements de 7 en 7, i presentem amb detall l'estratègia en arribar a 36. $[21] \xrightarrow{+7} [28] \xrightarrow{+7} [35] \xrightarrow{+7} [36, 1, 2, 3, 4, 5] \xrightarrow{+7} [6] \xrightarrow{+7} [13]$ <p>La mostra està formada pels alumnes que corresponen als nombres 21, 28, 35, 6 i 13.</p>

Aquest procediment es pot realitzar sempre que en la numeració dels individus de la població no sospitem cap regularitat. No seria adient agafar els dies en periodicitat setmanal, mensual, etc., per a fer un estudi relacionat amb el trànsit.

Mostreig aleatori estratificat

El *mostreig aleatori estratificat* proporcional pressuposa que la població està formada per grups diferenciats que es denominem estrats. En el cas de l'*afixació proporcional* aquests grups estan representats a la mostra en la mateixa proporció numèrica que ho estan a la població.

Així, es considera que la població de N individus està formada per diferents estrats de grandària N_1, N_2, \dots, N_t que complixen la condició: $N_1 + N_2 + \dots + N_t = N$. Cal trobar el nombre d'individus n_1, n_2, \dots, n_t que han de ser seleccionats per mostreig aleatori de cadascun dels estrats, de manera que $n_1 + n_2 + \dots + n_t = n$. Per a calcular-los plantegem la proporció:

$$\frac{N_1}{n_1} = \frac{N_2}{n_2} = \dots = \frac{N_t}{n_t} = \frac{N}{n}$$

Descripció	Exemple
<ul style="list-style-type: none"> – Identifiquem els estrats en que dividirem la població i calculem N_1, N_2, \dots, N_t. – Calculem la grandària de les mostres en els estrats n_1, n_2, \dots, n_t mitjançant la proporció: $\frac{N_1}{n_1} = \frac{N_2}{n_2} = \dots = \frac{N_t}{n_t} = \frac{N}{n}$ – En cada mostra de cada estrat es trien els individus per mostreig aleatori. 	<ul style="list-style-type: none"> – Imaginem que els alumnes estan diferenciats en 21 xiques i 15 xics: $N_1 = 21, N_2 = 15$ – Per a calcular n_1, n_2 resollem les proporcions $\frac{N_i}{n_i} = \frac{N}{n}$ i obtenim: $n_1 = 3, n_2 = 2$ – La mostra està formada per 3 xiques i 2 xics que triarem dintre de cada grup mitjançant qualsevol dels procediments anteriors.

Hi ha altres mètodes d'afixació no proporcional en el mostreig estratificat, però no seran abordats per tindre aquest text una caràcter introductori.

Mostreig aleatori per conglomerats

El *mostreig aleatori per conglomerats* pressuposa que la població està formada per grups diferenciats que s'anomenen conglomerats. Es considera que qualsevol d'ells és una mostra representativa de la població, puix que en cadascun d'ells està proporcionalment reflectida la diversitat de les seues característiques.

Podem utilitzar com a mostra qualsevol conglomerat. Per exemple, si estem fent un treball d'estudi de mercat sobre l'acceptació de cert producte genèric en una urbanització d'habitatges que es molt homogènia (tots els habitatges tenen el mateix preu, es van construir simultàniament, els seus habitants tenen costums professionals i poder adquisitiu semblants, etc.), podem considerar que cada carrer d'aquesta urbanització és un conglomerat. Així, es pot escollir aleatòriament un carrer per a fer l'enquesta.

Per concloure, es pot dir que els *estrats* són grups internament homogenis i que entre tots reflecteixen l'heterogeneïtat de la població, mentre que els *conglomerats* són homogenis entre si, tot i que dintre de cadascun d'ells està representada l'heterogeneïtat de la població.

Inferència. Paràmetres i estadístics

Una vegada seleccionats els elements de la mostra, es procedeix a fer un treball d'estadística descriptiva calculant els paràmetres que es necessiten de la mostra (mitjana aritmètica, variància, etc.). Basant-se en aquests resultats, es pot obtenir informació i conclusions de la característica de la població que es vol conèixer. Aquest darrer treball és l'objectiu de la *inferència estadística*.

Així doncs, la inferència estadística pot definir-se com la part de l'estadística que té per objecte el desenvolupament de tècniques que permeten conèixer o comprovar el valor dels paràmetres d'una població, a partir de les dades obtingudes d'una petita part extreta d'ella que hem denominat mostra.

La fiabilitat d'aquestes deduccions es mesura en termes probabilístics, és a dir, tota afirmació en inferència va acompanyada de la seua probabilitat d'encert. Per la qual cosa, es pot veure que s'arriba a les conclusions de la inferència aplicant la teoria de la probabilitat de la qual ja se n'han vist alguns resultats als temes anteriors.

La inferència estadística té dues grans branques:

- *Estadística inductiva o estimació*, l'objecte de la qual és estimar el valor dels paràmetres de la població. Pot ser:
 - Estimació puntual: que desenvoluparem als darrers apartats d'aquest tema.
 - Estimació per intervals: que desenvoluparem al tema 9.
- *Contrast d'hipòtesis*, l'objecte de la qual és comprovar, mitjançant mètodes matemàtics, hipòtesis realitzades sobre el valor d'algun paràmetre de la població. Desenvoluparem aquesta branca al tema 10.

Reprenem els exemples plantejats en començar el tema:

- Volem intuir el valor de la mitjana de l'alçada de la població de la nostra ciutat, és a dir, la mitjana poblacional, que denotarem per μ . A tal fi, triarem un grup de persones que reflecteixen la diversitat de la població. És, per tant, una mostra representativa. Calcularem la mitjana de l'alçada d'aquests individus, és a dir, calcularem la mitjana mostral, que denotarem per \bar{x} .
- En una factoria, podem esbrinar la proporció de mistos defectuosos en la producció total, és a dir, la proporció poblacional, que denotarem per p , seleccionant un nombre prefixat d'objectes que extraurem de la cadena de producció i que conformaran la mostra. D'entre aquests mistos, comptarem la proporció de defectuosos, que serà la proporció mostral que denotarem per p .

No obstant això, cal remarcar que \bar{x} no té exactament el valor de μ i que p no coincideix exactament amb el valor de p .

Així, quan treballem amb mostres, caldrà diferenciar aquests paràmetres calculats amb els individus seleccionats, denominats *paràmetres estadístics* o simplement *estadístics*, dels paràmetres reals corresponents a la població, que denotem com a *paràmetres poblacionals* o simplement *paràmetres*, els quals volem esbrinar mitjançant les tècniques de la inferència estudiarem.

Així, en general denotarem per θ el paràmetre a estimar d'una població i que serà segons el cas, la seua mitjana m , la seua variància σ^2 , o en el cas de poblacions de Bernoulli, la p . Hem comentat que per a trobar aquest valor triarem una mostra representativa de grandària n , amb aquestes dades, calcularem la seua mitjana aritmètica \bar{x} , la seua variància S_x^2 , o la proporció p en el cas d'una població Bernoulli, segons siga el valor a estimar.

En general, les mesures estudiades en l'estadística descriptiva dels primers temes són exemples d'estadístics (si els considerem com a funcions), ja que els seus valors depenen de les dades obtingudes a les mostres. Aquest paràmetre que s'obté a partir del càlcul amb els valors coneguts d'una mostra és al que direm estadístic i el denotarem en aquest tema, de manera genèrica, per la lletra T .

Intentarem explicar ara que aquests estadístics són variables aleatòries, ja que els valors mostrals també ho són. Per exemple, imaginem totes les mostres de 5 elements que podem obtenir de la població que ens havíem plantejat l'exemple 1, on volíem conèixer l'alçada de la població dels majors d'edat de la nostra ciutat.

Així, podrem dir que una *mostra aleatòria simple* de grandària n és un conjunt de n variables aleatòries $X_1, X_2, X_3, \dots, X_n$, on cada X_i representa el valor observat en la i -èsima extracció i podrà, doncs, agafar qualsevol valor de la població. Totes aquestes variables X_i tenen la mateixa distribució de probabilitat, és a dir, tenen la mateixa funció de probabilitat (en el cas de les discretes) o de densitat (en el cas de les contínues) que anomenarem *distribució de la població*. A més a més, és clar que aquestes variables són *independents*.

Com que les variables que componen la mostra són aleatòries, qualsevol estadístic calculat amb les dades de la mostra també és una variable aleatòria. Per tant, un primer pas en la inferència estadística consisteix a analitzar les distribucions de probabilitat dels estadístics per a saber com de fiables són els resultats que obtenim basant-nos en ells. Les distribucions de probabilitat dels estadístics s'anomenen distribucions en el mostreig.

Models de distribució de probabilitat en el mostreig

Al llarg del tema arribarem a definir les distribucions dels estadístics amb els quals volem abordar la inferència en aquest treball i podrem conèixer el model de distribució que s'ajusta a la mitjana mostral, a la proporció mostral, a la diferència de mitjanes, al quocient de variàncies, etc.

Primerament, cal introduir altres models de distribucions de probabilitat i afegir-los als que ja coneixem (distribució binomial, Poisson, normal, etc.), puix que els necessitem per al nostre propòsit per tractar-se de models que apareixen relacionats amb les distribucions dels estadístics de mostres aleatòries de poblacions normals.

Models de distribució de probabilitat en el mostreig

Xi-quadrat

Considerem Z_1, Z_2, \dots, Z_n un conjunt de variables aleatòries independents que es distribueixen segons el model normal de mitjana 0 i desviació típica 1, és a dir, $Z_i \rightarrow N(0, 1)$, i definim la variable $X = \sum_{i=1}^n Z_i^2$. Direm que la variable X té una distribució Xi-quadrat amb n graus de llibertat i ho denotarem així: $X \rightarrow \chi_n^2$.

D'aquesta distribució podem afirmar que $E[X] = n$ i que $\text{VAR}[X] = 2n$.

En aquesta figura (fig. 23), com a exemple, mostrarem la gràfica de la seua funció de densitat per $n = 10$ graus de llibertat.

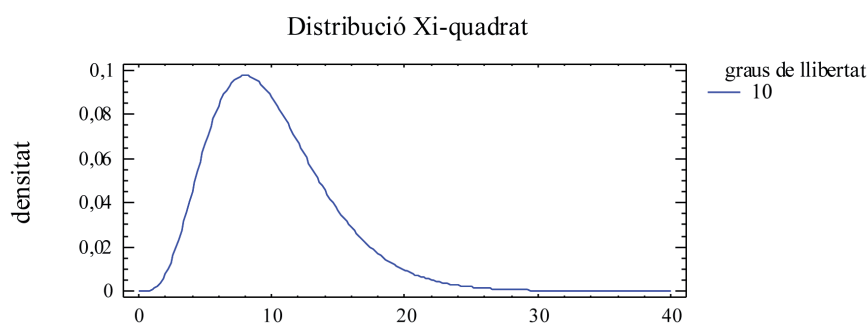


Fig. 23

Observem que aquesta funció no és simètrica. Aquest model de distribució apareixerà en les estimacions de la variància de la població, com podrem veure més endavant. A més a més, coneixerem els valors de les seues funcions de distribució $F(x)$ mitjançant taules o programes informàtics d'estadística.

t de student

Considerem X i Z dues variables aleatòries independents que es distribueixen respectivament com una Xi-quadrat amb n graus de llibertat i com una variable normal tipificada, és a dir, $X \rightarrow \chi_n^2$, i $Z_i \rightarrow N(0, 1)$, i definim la variable $T = \frac{Z}{\sqrt{\frac{X}{n}}}$. Direm

que la variable T es distribueix com una variable t de student amb n graus de llibertat i ho denotarem així $T \rightarrow t_n$.

D'aquesta distribució es pot afirmar que $E[X] = 0$.

En aquesta figura (fig. 24), com a exemple, mostrarem de la gràfica de la seua funció de densitat per $n = 10$ graus de llibertat.

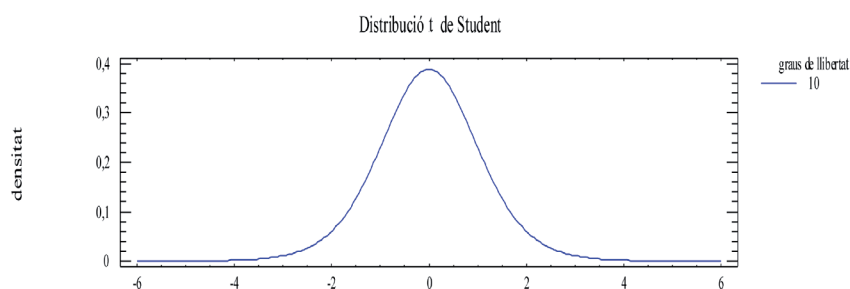


Fig. 24

Com es pot advertir a la figura 24, la forma general d'aquesta funció de densitat és similar a la d'una distribució normal; ambdues tenen forma de campana i són simètriques respecte de la mitjana. De la mateixa manera que la normal tipificada, la variable T té una mitjana igual a 0. No obstant això, la seua variància depèn del paràmetre que hem anomenat graus de llibertat. La variància de la variable t és major que 1, però s'apropa a 1 quan n augmenta. De fet, és possible demostrar que la distribució t amb n graus de llibertat tendeix a la distribució normal tipificada quan $n \rightarrow \infty$.

A les fórmules que utilitzarem en el capítol següent en el càlcul dels intervals podrem operar $P(T \leq x) = 1 - P(T \leq -x)$ per la simetria d'aquesta funció de densitat. Aquesta distribució t de Student apareixerà a les distribucions de la mitjana i de la diferència de mitjanes quan la variància poblacional és desconeguda, i coneixerem els valors de les seues funcions de distribució $F(x)$ mitjançant taules o programes informàtics d'estadística.

F de Fisher-Snedecor

Considerem X_1 i X_2 dues variables aleatòries independents que es distribueixen segons el model de χ^2 amb n i m graus de llibertat respectivament, és a dir,

$$X_1 \rightarrow \chi_n^2, \quad X_2 \rightarrow \chi_m^2, \quad \text{i definim la variable } F = \frac{\frac{X_1}{n}}{\frac{X_2}{m}}.$$

Direm que F es distribueix com una variable F de Fisher-Snedecor amb n i m graus de llibertat i ho denotarem així: $F \rightarrow f_{n,m}$.

A la següent figura (fig. 25) es mostra un exemple de la seua funció de densitat per a $n = 10$ i $m = 5$ graus de llibertat.

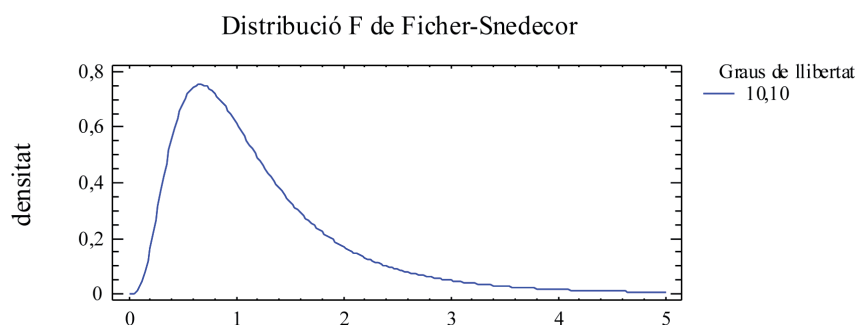


Fig. 25

Aquest model de distribució ens apareixerà a les estimacions del quocient de variàncies. Per al càlcul dels valors de la seua funció de distribució cal fer ús de la relació $P(f_{n,m} \leq x) = 1 - P(f_{m,n} \leq \frac{1}{x})$ quan s'empren les taules per a trobar-los.

Models de distribució de probabilitat d'alguns estadístics

Models per a una mostra

En aquest apartat considerarem X_1, X_2, \dots, X_n una mostra aleatòria d'una població que es distribueix amb qualsevol model de distribució amb mitjana m i variància σ^2 .

NOTA

Al llarg del capítol es parlarà de distribució de probabilitat de la població, paràmetre de la població, etc., en comptes de distribució de la variable que s'està estudiant sobre tota la població, paràmetre relatiu a la variable aleatòria que s'està estudiant sobre tota la població, etc. Aquest abús de notació es realitza per motius obvis.

Distribució de la mitjana mostral (coneguda σ^2 poblacional)

Suposem que volem estimar el valor del paràmetre m d'una població. A tal fi, agafaríem una mostra de n elements i calcularíem la seua mitjana mostral.

Considerem la variable aleatòria $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$, que assigna a cada mostra de grandària n el valor de la seua mitjana. Aquesta variable aleatòria es diu *mitjana mostral* i la distribució que segueix es diu *distribució mostral de les mitjanes*. Es pot demostrar que la mitjana d'aquesta variable aleatòria $E[\bar{X}] = \mu$, és a dir, la seua esperança matemàtica és igual a la mitjana de la població.

- La desviació típica de la variable aleatòria \bar{X} és: $\sqrt{VAR[\bar{X}]} = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. On σ és la desviació típica poblacional i n la grandària de la mostra. D'aquesta propietat es dedueix que les mitjanes mostrals estan més disperses respecte del valor de μ si n és menuda i més propers al valor que volem estimar si augmenta la grandària de la mostra.
- La distribució de la variable aleatòria \bar{X} , que anomenem distribució mostral de les mitjanes, segueix una distribució normal ($\bar{X} \rightarrow N(\mu_{\bar{X}}, \sigma_{\bar{X}})$) sempre que considerem un mostreig aleatori simple i que la mostra aleatòria tinga una distribució normal. Si no es dóna aquesta darrera exigència, podem aproximar la distribució mostral de les mitjanes per una distribució normal si n és gran. A més a més, pel teorema central del límit, aquesta aproximació és millor com més gran és n . Tanmateix, en el cas que la població *no segueisca una distribució normal i n siga menut*, \bar{X} segueix una distribució t de Student amb n graus de llibertat (ho veurem amb detall al proper apartat).

NOTA

Com que el text és clarament de caràcter docent i introductori, podem considerar que la mostra té una grandària qualificable de «gran» si $n > 30$.

Per una altra part, cal recordar que el fet que $\bar{X} \rightarrow N(\mu_{\bar{X}}, \sigma_{\bar{X}})$ és equivalent a afirmar que la variable $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ es distribueix com una normal tipificada, és a dir, $Z_i \rightarrow N(0, 1)$.

Com a conseqüència d'aquests resultats podrem tenir:

- Control de les mitjanes mostrals.

En una població de mitjana μ i desviació típica σ , ens disposem a extraure una mostra de grandària n . Abans de fer-ho, sabem la distribució de les mitjanes \bar{X} i amb això podrem esbrinar la probabilitat de que la mitjana d'una mostra estiga en un cert interval de valors.

- Control de la suma dels valors d'una mostra.

Podrem calcular la probabilitat de que la suma dels valors d'una mostra estiga dintre de cert interval, ja que $\sum_{i=1}^n X_i = n\bar{X}$ i la distribució d'aquesta darrera és coneguda.

- Inferir la mitjana de la població a partir de la mitjana d'una mostra.

Aquesta és l'aplicació més important i la que ens havíem plantejat des del començament del tema. En aquesta unitat parlarem de l'estimació puntual i a la propera unitat parlarem de l'estimació per intervals amb un cert grau de certesa que denominarem nivell de confiança i que indicarà, en termes de probabilitat, la certesa dels nostres resultats.

Distribució de la mitjana mostral (si no és coneguda σ^2 poblacional)

Quan no es coneix el valor de la σ^2 poblacional, cal emprar una altra distribució per a abordar un estadístic que explique el comportament de la mitjana mostral, sense que aquest paràmetre aparega. Així, s'utilitza la següent expressió que es defineix com a quocient d'una variable normal tipificada i l'arrel quadrada d'una variable Xi-quadrat amb $n-1$ graus de llibertat.

Si denotem per \bar{X} la variable aleatòria mitjana mostral de grandària n extreta

d'una població normal de mitjana μ i variància σ^2 , i per $S = S_{X_{n-1}}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$

la quasi variància mostral, es pot demostrar que la variable aleatòria $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$

es distribueix segons una distribució t de Student amb $n-1$ graus de llibertat. És a dir, que $T \rightarrow t_{n-1}$.

Aquest resultat és més general que l'anterior, ja que no necessitem conèixer el valor de la variància poblacional s^2 . Tanmateix, és més restrictiu perquè implica el supòsit d'una població normal. Malgrat tot, aquesta restricció no és tan severa com sembla. Els estudis han demostrat que la distribució de la variable aleatòria

$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ s'apropa molt a una distribució t de Student, encara que considerem

mostres que no provenen d'una població normal. En la pràctica serà suficient poder assegurar-nos de que la població de la qual fem el mostreig tinga una distribució en forma de campana no massa esbiaixada.

Distribució d'una proporció

En una determinada població, la proporció d'individus que posseeixen una característica determinada la denotem per p . Per a esbrinar aquest valor es consideren totes les possibles mostres de grandària n que es poden extraure d'aquesta població. En cadascuna de les mostres hi haurà una proporció p_1, p_2, \dots , d'individus de la població que posseeixen aquesta característica. Estudiarem com es distribueixen tots els possibles valors de p .

Si denotem per X el nombre d'individus de la mostra que té una determinada característica, podem afirmar, pels coneixements de temes anteriors, que $X \rightarrow \text{Bi}(n, p)$; també vam veure que si es compleix que $np \geq 5$ i que $nq \geq 5$ (on òbviament $q = 1 - p$), podem transformar una variable binomial en una normal i assegurar que $X \rightarrow N(np, \sqrt{npq})$

Si denotem per $p = \frac{\text{nombre d'individus amb la característica determinada}}{\text{nombre d'individus de la mostra}} = \frac{X}{n}$

La distribució de p , també anomenada distribució mostral de la proporció serà la mateixa que la de X , però amb els paràmetres de mitjana i desviació típica dividits per n . Per tant:

$$p \rightarrow N\left(\frac{np}{n}; \frac{\sqrt{npq}}{n}\right) = N\left(p; \sqrt{\frac{pq}{n}}\right)$$

Aquesta aproximació és millor com més gran és la mostra i més propera és p a 0,5. Així, si les condicions inicials ($np \geq 5$ i $nq \geq 5$) es compleixen, podem considerar que tenim una bona aproximació. Si aquest no fóra el cas, caldria augmentar la grandària de la mostra.

Distribució de la variància

No podem trobar un model de distribució que ens permeti conèixer directament la distribució de l'estadístic de la variància mostral, però sí que podem conèixer la distribució d'aquesta variable que definirem, que permetrà conèixer el valor de la variància de la població basant-nos en la variància de la mostra.

Si denotem per $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ la quasi-variància d'una mostra aleatòria de grandària n extreta d'una població normal de variància s^2 , es pot demostrar que la variable aleatòria $X = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$ es distribueix com una variable Xi-quadrat amb $n-1$ graus de llibertat, és a dir, $X \rightarrow \chi_{n-1}^2$.

Recordem dels apartats anteriors que la variable Xi-quadrat l'havíem definida com la suma de variables normals tipificades elevades al quadrat $X = \sum_{i=1}^n Z_i^2$ i que aquesta estructura podem trobar-la a l'expressió de la nostra variable.

Models per a dues mostres

Per començar, cal diferenciar dos possibles circumstàncies pel que fa a les composicions de les mostres. Així, cal saber si estem amb el cas de mostres independents de dues poblacions (alumnes de dos grups diferents, treballadors de dos seccions diferents, homes o dones per a estudiar algun comportament lligat o no amb el gènere, etc.) o si es tracta de mostres relacionades o aparellades (si estudiem el desgast en la roda dreta i esquerra d'un vehicle; la duració d'un material de sola d'un calcer esportiu per al qual s'ha dissenyat un model, proporcionant a cada individu; una sabata de cada peu amb cada material, les mesures de la pressió arterial preses en cada malalt abans i després de seguir un tractament farmacològic; etc.).

Suposem ara que volem comparar dues variables de dues poblacions. Necessitem, doncs, triar dues mostres, una de cadascuna d'elles. A més, suposem que cada mostra és independent. En cas de considerar-les relacionades o emparellades s'especificarà.

En els apartats següents denotarem per X_1, X_2, \dots, X_n una mostra aleatòria de grandària n extreta d'una població que es distribueix amb qualsevol model de distribució amb mitjana μ_X i variància σ_X^2 i denotarem Y_1, Y_2, \dots, Y_m una mostra aleatòria de grandària m extreta d'una altra població que es distribueix amb qualsevol model de distribució amb mitjana μ_Y i variància σ_Y^2 .

Diferència de mitjanes

Sovint comparem les mitjanes de dues poblacions per a vore si el rendiment mitjà és millor en una població que en l'altra. Així, podríem comparar els resultats d'una prova d'estadística en un grup d'estudiants del torn del matí amb els del grup de la vesprada. A tal fi, tal com estem treballant al llarg d'aquest tema, caldrà triar una mostra representativa de les proves dels alumnes de cadascun dels grups i traure la mitjana mostral de cadascuna d'elles. Imaginem que una mitjana és unes dècimes més alta que l'altra. Amb els plantejaments que vorem a continuació podrem decidir si aquesta diferència entre les mitjanes és realment significativa o és causada per l'atzar.

A continuació presentem les distribucions de tres variables que ens permetran fer aquesta comparació en tres circumstàncies que cal considerar a l'hora de comparar les mitjanes:

- Mostres independents amb variàncies poblacionals conegudes.

- Mostres independents amb variàncies poblacionals no conegudes, però iguals.
- Mostres dependents de dades relacionades ($n = m$, necessàriament).

Distribució de la diferència de les mitjanes (si coneixem σ_X^2 i σ_Y^2)

Considerarem les següents variables aleatòries, $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ i $\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_m}{m}$, que assignen a cada mostra el valor de la seua mitjana i definim a partir d'aquestes la variable aleatòria $\bar{X} - \bar{Y}$. Es pot demostrar que aquesta variable es distribueix com una variable normal que té per mitjana la resta de les mitjanes mostrals, $\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$, i per variància la suma de les variàncies de les mitjanes mostrals, $\sigma_{\bar{X}-\bar{Y}}^2 = \sigma_{\bar{X}}^2 + \sigma_{\bar{Y}}^2 = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$. Així,

$$\bar{X} - \bar{Y} \rightarrow N\left(\mu_X - \mu_Y, \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)$$

Una expressió equivalent a l'anterior obtinguda en tipificar seria:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \rightarrow N(0, 1)$$

Es pot demostrar que aquest resultat és vàlid si les poblacions de les quals hem extret les mostres es distribueixen com una variable normal. No obstant això, el resultat també és compleix quan treballem amb mostres suficientment grans que ens permeten aplicar el teorema central del límit i aproximar σ_X^2 i σ_Y^2 per les respectives quasi-variàncies. Als exemples d'aquest manual considerarem adient l'aproximació quan n i m són majors que 30.

Distribució de la diferència de les mitjanes (si no coneixem σ_X^2 i σ_Y^2)

Quan desconeguem les variàncies poblacionals no podrem aplicar la distribució de la variable que hem presentat a l'apartat anterior. No obstant, podem emprar-ne una altra sempre que puguem assumir que les variàncies són iguals, encara que siguin desconegudes.

Així doncs, caldrà fer, primerament, un treball d'inferència sobre el quocient de les variàncies de la població amb l'objecte de saber si les variàncies poden considerar-se iguals. Aquest treball d'inferència serà tractat més endavant. Així mateix,

tan sols quan siga possible inferir l'esmentada igualtat de variàncies, podrem aplicar el resultat que tot seguit desenvoluparem.

Si denotem per $S_{X_{n-1}}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ la quasi variància de la mostra aleatòria de grandària n extreta de la primera població i per $S_{Y_{m-1}}^2 = \frac{\sum_{i=1}^m (y_i - \bar{y})^2}{m-1}$ la quasi variància de la mostra de grandària m de la segona població, es pot demostrar que la variable

aleatòria $T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{(n-1)S_{X_{n-1}}^2 + (m-1)S_{Y_{m-1}}^2}{n+m-2} \sqrt{\frac{1}{n} + \frac{1}{m}}}}$ es distribueix segons el model d'una variable t de Student amb $n + m - 2$ graus de llibertat. És a dir, $T \rightarrow t_{n+m-2}$.

NOTA

Aplicarem aquest resultat quan n i m siguin menors que 30 i les mostres siguin independents.

Distribució de la diferència de mitjanes amb mostres relacionades

En alguns casos pot ser interessant comparar les mitjanes de dues poblacions que estan relacionades, com per exemple la mitjana del pes d'uns individus abans i després de fer un tractament per a aprimar-se, les diferències d'uns indicadors de qualitat abans i després de fer una millora en els materials o el procés de producció, etc.

Sovint és interessant, en el disseny de l'experiment, considerar aquesta relació. Per exemple, si volem conèixer la difèrença de la qualitat d'un material per a pneumàtics, és un bon experiment posar en la roda dreta el material A i en l'esquerra, el material B per a que una mostra aleatòria de vehicles formen part de l'experiment. Així, ens assegurem de que els materials estiguen sotmesos a les mateixes condicions de conducció, com ara la temperatura, el terreny, la forma de conducció, etc.

Són molts, doncs, els experiments que ens portaran a fer aquesta comparació de mitjanes amb mostres que podem anomenar relacionades o aparellades.

Per a dur a terme aquesta distribució, crearem una variable, que podem anomenar diferència i que denotarem per D , que pren com a valors la diferència de cada parell. Així, calcularem $D_i = X_i - Y_i$ i, amb aquestes dades, calcularem $S_{D_{n-1}}$, que serà la quasi-desviació típica d'aquests valors.

Si definim la variable $T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\frac{S_{D_{n-1}}}{\sqrt{n}}}$, es pot demostrar que té una distri-

bució que segueix el model d'una variable t de Student amb $n - 1$ graus de lliber-
tat. Es a dir, $T \rightarrow t_{n-1}$.

Distribució de la diferència de proporcions

En alguns casos ens podem plantejar comparar les proporcions d'individus de dues poblacions que tinguen alguna característica comuna. Ens pot interessar comparar el percentatge de productes defectuosos en dos processos de producció, el percentatge de votants d'un determinat partit polític en dues eleccions consecutives, el percentatge d'afectats de certa malaltia en dues comunitats, etc. En tots aquests exemples, nosaltres calcularem la proporció dels individus en cadascuna de les mostres i voldrem inferir si les diferències entre els valors que s'obtenen en les dues mostres són significatives, així com també estan al voltant d'un determinat valor o tan sols estan causades per un procés d'atzar per tractar-se de mostres triades aleatòriament i que poden diferir mínimament en els seus valors concrets.

En aquest apartat denotarem per X_1, X_2, \dots, X_n una mostra aleatòria de grandària n extreta d'una població en la qual la proporció d'individus que té certa característica serà denotada per p_X . A la proporció que obtenim amb aquests valors de la mostra la denotarem per p_x .

Igualment, referirem per Y_1, Y_2, \dots, Y_m la mostra aleatòria de grandària m extreta de una altra població que té una proporció que denotarem per p_Y . Al valor de la proporció calculat amb les dades d'aquesta mostra el denotarem per p_y .

La distribució de p_x , també anomenada distribució mostral de la proporció, ja l'hem vista anteriorment en analitzar el cas d'una mostra i correspon a una distribució normal amb els paràmetres següents:

$p_x \rightarrow N\left(p_x; \sqrt{\frac{p_x(1-p_x)}{n}}\right)$, i si la transformem per tipificar-la:

$$\frac{p_x - p_x}{\sqrt{\frac{p_x(1-p_x)}{n}}} \rightarrow N(0; 1)$$

Paral·lelament, obtenim una expressió semblant de la distribució de p_Y i de la seua tipificació: $\frac{p_Y - p_Y}{\sqrt{\frac{p_Y(1-p_Y)}{m}}} \rightarrow N(0; 1)$.

Si ens plantegem la distribució de la seua diferència, $p_x - p_y$, és la diferència de dues distribucions normals tipificades, per la qual cosa la seua distribució serà una normal que té per mitjana la resta de les mitjanes i per variància la suma de les variàncies de cadascuna. Així, coneixerem la distribució de la variable, que prenem com a valors la diferència de les proporcions de les mostres extretes de les dues poblacions.

Definim la variable:

$$Z = \frac{(p_x - p_Y) - (p_x - p_Y)}{\sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_Y(1-p_Y)}{m}}} \rightarrow N(0; 1)$$

Es pot demostrar que es distribueix segons una distribució normal tipificada de mitjana 0 i variància 1 quan la grandària de les mostres es gran, la qual cosa permet transformar una variable binomial en una normal.

Distribució del quocient de variàncies

En algunes situacions necessitem comparar les variàncies de dues poblacions. Per exemple, aquest procés és necessari per a calcular l'estimació de la diferència de mitjanes de dues poblacions, on cal comprovar que les seues variàncies poblacionals poden ser desconegudes però iguals.

Considerarem la mostra X_1, X_2, \dots, X_n una mostra aleatòria de grandària n extreta d'una població amb variància σ_x^2 i Y_1, Y_2, \dots, Y_m una mostra aleatòria de grandària m extreta d'una altra població amb variància σ_y^2 . Les quasi-variàncies de cadascuna de les mostres les denotem per $S_{X_{n-1}}^2$ i per $S_{Y_{m-1}}^2$, respectivament.

Per a comparar les variàncies de les dues poblacions cal estudiar la distribució de la variable següent, que és l'estadístic que ens permet comparar per quocient, ja que d'aquesta expressió podrem obtenir una distribució coneguda que ja vam veure a l'apartat anterior. Recordem que el model de la variable F de Fisher-Snedecor és un model que s'ajusta al quocient de dues variables que presenten una distribució d'una variable Xi-quadrat.

A tal fi, definim la variable aleatòria $F = \frac{(n-1)S_{X_{n-1}}^2}{\sigma_x^2 \cdot n} \cdot \frac{\sigma_y^2 \cdot m}{(m-1)S_{Y_{m-1}}^2}$ i que podem demostrar que té

la distribució de la variable F de Fisher-Snedecor amb $n - 1$ i $m - 1$ graus de llibertat, respectivament.

Recordem que aquest model el vam definir com una distribució, la qual és el quocient de dues variables que es distribueixen com a χ^2 de n i m graus de llibertat.

NOTA

Cal fixar-se que, en aquest apartat que ara acaba, els exemples no aborden cap problema d'inferència, tal com havíem plantejat a l'inici del tema, on ja hem definit que l'objectiu és poder esbrinar el valor d'un paràmetre poblacional desconegut q , basant-nos en un paràmetre mostral T , els quals calculem amb les dades reals dels valors d'una mostra aleatòria.

Els exemples que hem vist tan sols pretenen mostrar que les variables que hem definit en cada apartat tenen una distribució determinada que ens permet plantejar-nos el càlcul de probabilitats i conèixer les seues funcions de densitat, la simetria d'aquestes, etc.

Estimació puntual

Al llarg el tema hem comentat que l'objectiu de la inferència estadística és esbrinar el valor d'un paràmetre poblacional q basant-nos en un estadístic T que calculem amb les dades d'una mostra.

Pel desenvolupament dels apartats anteriors, ja sabem tots que \bar{x} no té exactament el valor de μ , ni p coincideix exactament amb el valor de p , etc. Podríem dir el mateix de les altres mesures estadístiques que podem calcular amb les dades de la mostra per a obtenir el valor corresponent de la població.

Aquest procediment és el que definim com *estimació puntual*. Afirmarem que el valor concret d'un estadístic T , és aquell que s'apropa al valor que volem esbrinar de la població q , però no podem controlar el grau de fiabilitat de la nostra estimació.

Ara bé, el que sí que podem fer és estudiar una sèrie de propietats que seria desitjable que compliren les distribucions dels estadístics, ja que aquestes són conegudes. Així, entre els diferents estadístics T_1, T_2, \dots, T_n que ens podem plantejar per a esbrinar un determinat paràmetre q , triarem el més idoni.

Per exemple, si desitgem calcular l'alçada de la població que ens havíem plantejat al primer exemple del tema, podrem triar mostres de 5 individus i plantejar-nos diferents alternatives:

- Calcular la mitjana de cada mostra \bar{X} .
- Calcular la semisuma del valor més gran i el més petit de la mostra $\frac{X_{\max} + X_{\min}}{2}$.
- Donar com a estimació el valor central de la mostra, etc.

Qualsevol d'aquestes estratègies podria ser un estimador de la mitjana m , ara bé, és evident que, a cop d'ull, no ens semblen totes igual de fiables.

Un altre problema clàssic és decidir entre la variància o la quasi-variància mostrals com dos estimadors de la variància de la població.

En aquest apartat estudiarem tres propietats que són, en certa manera, una mesura de la «qualitat» d'un estadístic com a estimador. Aquestes són el biaix, la consistència i la eficiència. Notem que encara que utilitzem T per a denotar l'estimador, es pot interpretar també per T_n , on el subíndex n denota la grandària de la mostra.

Biaix

Direm que un estimador T és un estimador *centrat* o *no esbiaixat* de q si $E[T] = \theta$. És a dir, la distribució d'aquest estimador està centrada en el paràmetre que volem estimar.

Consistència

Direm que un estimador $T = T_n$ és un estimador *consistent* de q si la successió d'estimadors T_1, T_2, \dots, T_n convergeix en probabilitat a q . És a dir, si tal com vam definir en la introducció del teorema central del límit, per a cada $\varepsilon > 0$ es compleix que $\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$.

Com a conseqüència d'aquesta definició es pot comprovar que la successió d'estimadors compleix que $\lim_{n \rightarrow \infty} E[T] = \theta$ i que $\lim_{n \rightarrow \infty} \text{VAR}[T] = 0$.

Per tant, un estimador consistent compleix que, quan augmenta la grandària de la mostra, l'estimació millora perquè els valors de l'estadístic estan centrats en q i la dispersió disminueix gradualment.

Eficiència

Si considerem dos estimadors T_1 i T_2 del paràmetre q , direm que T_1 és més eficient que T_2 si $\text{VAR}[T_1] < \text{VAR}[T_2]$, és a dir, els valors de l'estimador T_1 estan menys dispersos que els de T_2 respecte del valor q .

Una imatge de les distribucions d'uns estimadors hipotètics pot ajudar-nos a il·lustrar aquests conceptes:

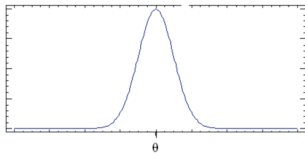


Fig. a

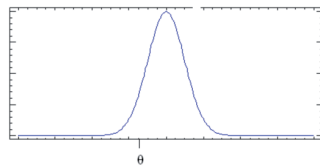


Fig. b

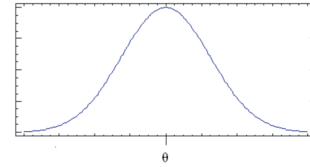


Fig. c

Podem afirmar que els estimadors de les figures a i c són no esbiaixats de q , mentre que el de la figura b no ho és. També es pot veure que l'estimador de a és més eficient que el de c.

Per a trobar estimadors que complisquen aquestes propietats, caldria explicar el mètode de màxima versemblança, el qual ens ho asseguraria. Però aquest estudi queda fora de l'objectiu del nostre treball.

Intervals de confiança

Si considerem una població la distribució de la qual és coneguda però de la que en desconeixem algun paràmetre, podem estimar el valor d'aquest paràmetre a partir d'una mostra representativa mitjançant un procés anomenat *estimació paramètrica*.

A la unitat anterior ja s'ha vist ja l'estimació puntual i en aquesta unitat estudiarem l'estimació per intervals de confiança.

Per exemple, si estimem que el percentatge de vots d'un determinat partit polític és del 48 % estarem realitzant una estimació puntual, mentre que si afirmem que l'estimació de vots es troba entre el 46 % i el 59 %, amb un nivell de confiança del 95 %, estarem considerant una estimació per intervals de confiança.

Intervals de confiança

Quan ens plantegem estimar mitjançant intervals de confiança pretenem trobar un interval en el qual podem afirmar que el valor del paràmetre de la població que volem conèixer, i que podem denotar genèricament com θ hi està dins amb una alta probabilitat.

En primer lloc, cal fixar aquesta probabilitat que anomenem *nivell de confiança* i que denotem per $1 - \alpha$. Al valor α se l'anomena *nivell de significació*.

Si $I = (a, b)$ és l'interval que busquem, cal que $P(\theta \in I) = 1 - \alpha$, i es diu que I és l'interval de confiança per al paràmetre θ amb un nivell de confiança $1 - \alpha$.

Paràmetres poblacionals i mostrals

El paràmetre a estimar d'una població que hem anomenat θ , serà, segons el cas, la seua mitjana μ , la seua variància σ^2 , o, en el cas de poblacions de Bernouilli, la proporció p que ens indica la probabilitat d'èxit, com ja vam vore al tema de les distribucions.

Per a trobar aquest interval, començarem el treball per triar una mostra representativa de la població d'una grandària n . Amb aquestes dades, en calculem la mitjana \bar{x} , la quasi-variància S_{n-1}^2 , o la proporció p (en el cas d'una població Bernouilli).

Donat que els valors d'aquests paràmetres calculats a partir de les mostres (o la transformació adient per a cada cas, com hem vist al tema anterior) segueixen una distribució coneguda, podem aplicar la teoria de la probabilitat i, amb l'ajut de taules o programes estadístics, trobar els extrems de l'interval que necessitem en cada cas.

Errada, grandària de la mostra i nivell de confiança

Per a començar a treballar un problema d'estimació cal fixar prèviament tres factors que condicionaran el nostre treball i la validesa del resultat:

- Nivell de confiança
- Errada de la solució
- Grandària de la mostra

Una vegada calculat l'interval de confiança podem mesurar l'errada de la nostra estimació. Així, si $\theta \in I = (a, b)$, podem mesurar l'errada amb l'expressió $E = \frac{b-a}{2}$. Com és obvi, aquesta errada depèn de l'amplitud de l'interval i també del nivell de confiança que demanem en la nostra estimació, puix que per al càlcul de l'interval es té en compte aquest nivell.

Si volem que aquesta errada disminuïska, cal rebaixar el nivell de confiança. Recordem que el *nivell de confiança* $1 - \alpha$ és la probabilitat de que el paràmetre θ a trobar estiga dintre de l'interval de confiança, i en certa manera podem considerar-la com una mesura de garantia de la certesa de la nostra afirmació. En conseqüència, és desitjable que el nivell de confiança siga un nombre proper a 1.

El tercer factor que juga en aquest equilibri és la *grandària* de la mostra n , ja que l'amplitud de l'interval –i conseqüentment l'errada– disminueixen en augmentar-ne la grandària. Així doncs, sempre podem fixar *a priori* dos dels factors a controlar i obtindre el tercer en funció d'aquells.

Per una altra part, en un experiment d'aquestes característiques (control de qualitat d'un producte, previsions electorals, investigacions de biologia, d'ecologia, medicina, etc.) en la vida real, la manipulació dels elements de la mostra té un cost econòmic i de temps, per la qual cosa cal trobar un equilibri entre aquests tres factors que hem anomenat per a trobar una solució satisfactòria per a la seua seguretat, exactitud i cost per a tots.

En els exemples que desenvoluparem al llarg del tema veurem aquesta relació aplicada en cada cas concret.

Intervals de confiança per als paràmetres d'una població

Per a la utilització de les següents fórmules, en les quals intentarem estimar un interval de confiança per a un paràmetre poblacional, partirem de l'existència d'una mostra de grandària n de la qual en coneixem, mitjançant els càlculs de l'estadística descriptiva que hem estudiat als primers temes, la mitjana aritmètica \bar{x} i la quasi-desviació típica S_{n-1} . Aquests sempre són coneguts o podem calcular-los a partir dels elements de la mostra.

Respecte dels paràmetres poblacionals a estimar, denotarem per μ la mitjana aritmètica i per σ la desviació típica.

En el cas d'una població Bernouilli, denotarem per p la proporció obtinguda a partir dels valors de la mostra i per p el paràmetre a estimar de la població.

En tots els casos denotarem per $1 - \alpha$ el nivell de confiança amb el qual decidirem treballar en cada cas. Els percentatges més habituals són del 90 %, 95 % i 99 %.

Així mateix, denotarem per $z_{1-\frac{\alpha}{2}}$ el valor x de la variable normal tipificada, al qual correspon $P(Z \leq x) = 1 - \frac{\alpha}{2}$. Aquests valors els podrem trobar a les taules que anirem especificant en cadascun dels casos o també amb l'ajut de qualsevol programa estadístic (Statgraphics, SPSS, R, etc.).

Amb el cas de les variables que necessiten la utilització d'un subíndex en la seua notació per a expressar els graus de llibertat (Xi-quadrat, t de Student, F de Fisher-Snedecor), passarem a indicar aquesta probabilitat amb $1 - \frac{\alpha}{2}$ o altres semblants, amb un subíndex davant de la lletra que ens indica la variable. Així, denotarem per $t_{1-\frac{\alpha}{2}, n-1}$ el valor x de la variable t de Student amb $n-1$ graus de llibertat, al qual correspon la probabilitat $P(t_{n-1} \leq x) = 1 - \frac{\alpha}{2}$, i raonarem de la mateixa manera en el cas de la variable Xi-quadrat amb $n-1$ graus de llibertat, χ^2_{n-1} , que denotarem per $\chi^2_{1-\frac{\alpha}{2}, n-1}$.

Aquests aspectes es tractaran amb més detall a la resolució dels exemples de cadascun dels casos.

Intervals per a les mitjanes

En aquest apartat es definiran els intervals de confiança per a les mitjanes poblacionals i per a la diferència de mitjanes (si es tenen en compte dues poblacions), segons la informació de la qual es disposa de la població. En cada subapartat es presentaran l'interval de confiança i un exemple d'aplicació.

Interval per a la mitjana d'una població normal amb desviació típica poblacional coneguda σ

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Demostrem aquest interval com a exemple i la resta dels casos es poden traure seguint el mateix procés a partir de les distribucions dels estimadors que hem vist al tema anterior.

Vam veure que la distribució de la mitjana mostral \bar{x} és una distribució normal de mitjana μ i desviació típica $\frac{\sigma}{\sqrt{n}}$. Per tant, si tipifiquem, s'obté que la variable $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ es distribueix com una variable normal estandaritzada. Així $Z \rightarrow N(0, 1)$.

Per una altra part, si s'observa la gràfica d'aquesta distribució, es pot afirmar que els valors centrals d'aquesta variable estan en un interval que té com a extrems uns percentils que depenen del percentatge que volem considerar. Denotem per $1 - \alpha$ aquesta probabilitat.

Com podem veure a la gràfica següent (fig. 27), $z_{\frac{\alpha}{2}}$ i $z_{1-\frac{\alpha}{2}}$ seran aquests percentils,

del quals podem dir que $P\left(Z \leq z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ i que $P\left(Z \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$.

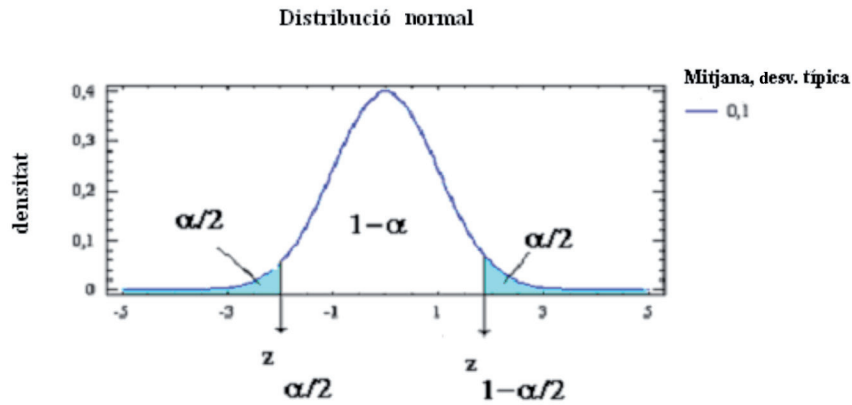


Fig. 27

Per la simetria de la distribució, es compleix que $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$, per la qual cosa es pot afirmar que $P\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$, i substituint l'expressió de la variable Z en aquest interval tenim:

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

En realitat, si \bar{x} es una mitjana mostral, com que es un valor de la variable \bar{X} sabem que almenys en un $(1 - \alpha)$ % dels casos es compleix que:

$$-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\frac{\alpha}{2}}$$

Aïllem pas a pas el paràmetre μ en aquesta desigualtat:

$$\begin{aligned} -z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} &\leq \bar{x} - \mu \leq z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \\ -\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} &\leq -\mu \leq -\bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \\ \bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} &\leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Així, l'interval de confiança de la mitjana amb un nivell de confiança del $(1 - \alpha)$ % és:

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

NOTA

Si la grandària de la mostra n és suficientment gran, podem utilitzar la fórmula d'aquest apartat, considerant que la desviació típica poblacional s , encara que fóra desconeguda, coincideix amb el valor de la quasi-desviació típica de la mostra S_{n-1} . Per tractar-se d'un manual de docència, trobarem adient aquesta consideració per a mostres de grandària n majors que 30.

Interval per a la mitjana d'una població normal amb desviació típica poblacional desconeguda

L'interval de confiança per a la mitjana en aquest cas és:

$$\mu \in \left(\bar{x} - t_{n-1, 1-\frac{\alpha}{2}} \frac{S_{n-1}}{\sqrt{n}}, \bar{x} + t_{n-1, 1-\frac{\alpha}{2}} \frac{S_{n-1}}{\sqrt{n}} \right)$$

Altres Intervals

Interval per a la proporció d'una població Bernoulli

L'interval de confiança per a la proporció en aquest cas és:

$$p \in \left(p - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}, p + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right)$$

Interval per a la variància d'una població normal

L'interval de confiança per a la variància en aquest cas és:

$$\sigma^2 \in \left(\frac{(n-1)S_{n-1}^2}{\frac{\alpha}{2} \chi_{n-1}^2}, \frac{(n-1)S_{n-1}^2}{1-\frac{\alpha}{2} \chi_{n-1}^2} \right)$$

Intervals de confiança per als paràmetres de dues poblacions

Fins ara hem abordat estimacions dels paràmetres d'una població i hem pogut calcular intervals dintre dels quals es troba el valor del paràmetre a estimar (mitjana, proporció, variància) amb una probabilitat que anomenem nivell de confiança i que pren valors, generalment, per damunt del 90 %.

En aquest apartat compararem els paràmetres de dues poblacions. Per a dur a terme el treball necessitem les dades de dues mostres diferents, extretes de cadascuna de les respectives poblacions, ja que volem inferir conclusions al respecte dels valors dels seus paràmetres poblacionals.

Cal detallar que en aquests casos no podem estimar els valors dels paràmetres de les poblacions; més bé podem comparar-los i concloure quin dels dos és major i en quina mesura. És a dir, podrem estimar l'interval en el qual es «mouen» els valors de la seua diferència i, en els cas concret de la comparació de variàncies, podem estimar l'interval corresponent als valors dels seus quocients. Aquesta circumstància ens permetrà estimar quin és major i fins a quin punt és significativa la diferència.

Per el que fa a la notació, hem utilitzat el subíndex X per als valors mostrals i poblacionals d'una de les poblacions a considerar i el subíndex Y per a l'altra.

Cal observar que en alguns casos les dades de les dues poblacions estaran referides a la mateixa variable a estudiar i voldrem extraure conclusions de la magnitud de les diferències en el valor dels paràmetres a comparar. Per exemple, si volem estimar si les qualificacions mitjanes de l'assignatura d'estadística del grup del matí i les del grup de la vesprada són significativament diferents, podrem denotar per \bar{x} la nota mitjana dels alumnes del matí que hem seleccionat aleatòriament per a la mostra i per S_x^2 la variància d'aquestes dades. Denotarem per \bar{y} i per S_y^2 els mateixos paràmetres referits a les qualificacions dels estudiants triats del grup de la vesprada. És clar que les dades a treballar corresponen també a la variable «qualificacions en l'assignatura d'estadística».

En altres casos voldrem comparar les diferències en els paràmetres de dues variables amb les dades extretes dels individus d'una mateixa població. Per exemple, podríem estudiar si hi ha més dispersió en les dades de l'alçada o en les del pes dels nounats d'un hospital. En aquest cas, podríem parlar de la comparació de variàncies de la variable X, que seria el pes, i la variable Y, que seria l'alçada de cada xiquet.

Si seguim amb la qüestió de la notació, mantenim el paral·lelisme amb els valors de les mitjanes μ_x i μ_y per a cadascuna de les poblacions i els símbols σ_x^2 i σ_y^2 per a denotar les variàncies.

En el cas d'una població Bernoulli denotem per p_x i p_y les proporcions obtingudes a partir dels valors de les mostres i per ρ_x i ρ_y les proporcions poblacionals a estimar de les poblacions respectives.

Quant a les notacions dels valors de les variables, totes coincideixen amb els de l'apartat anterior, tret de la variable F de Fisher-Snedecor amb $n-1$ i $m-1$ graus de llibertat que denotarem per $f_{(n-1), (m-1)}$. Cal recordar que l'ordre d'aquests subíndex no és commutatiu.

La resta d'aspectes són iguals als tractats a l'apartat anterior i que anirem comentant mitjançant els exemples de cada cas.

És important ressaltar en aquest apartat que ara estudiem una certa particularitat a l'hora d'interpretar els resultats de l'interval trobat (ens hi referirem amb més profunditat en cadascun dels casos). Així mateix, seria adient fixar-se en el quocient de variàncies per la diferència amb la resta dels intervals en aquest aspecte.

Intervals per a la diferència de mitjanes

Per a començar abordarem el cas de la diferència de mitjanes i caldrà saber si estem amb el cas de mostres independents de dues poblacions o si es tracta de mostres relacionades o aparellades.

En el primer cas, amb les mostres independents, no cal que la grandària de les dues mostres coincidisca. No obstant, és obvi que en el cas de les mostres aparellades les grandàries de les dues mostres han de ser iguals, puix que a cada individu li corresponen un parell d'observacions que permeten definir una variable a treballar anomenada diferència, D . Aquesta nova variable és la diferència de les dues observacions de cada element, així: $d_i = x_i - y_i$.

Interval per a la diferència de mitjanes de poblacions independents amb variàncies poblacionals conegudes

L'interval de confiança per a la diferència de mitjanes en aquest cas és:

$$\mu_x - \mu_y \in \left(\bar{x} - \bar{y} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}, \bar{x} - \bar{y} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \right)$$

Interval per a la diferència de mitjanes de poblacions independents amb variàncies poblacionals desconegudes, però iguals

L'interval de confiança per a la diferència de mitjanes en aquest cas és:

$$\mu_x - \mu_y \in \left(\bar{x} - \bar{y} - z_{1-\frac{\alpha}{2}} t_{n+m-2} \sqrt{\frac{(n-1)S_{x_{n-1}}^2 + (m-1)S_{y_{m-1}}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}, \right. \\ \left. \bar{x} - \bar{y} + z_{1-\frac{\alpha}{2}} t_{n+m-2} \sqrt{\frac{(n-1)S_{x_{n-1}}^2 + (m-1)S_{y_{m-1}}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

Aplicarem aquesta fórmula quan no coneguem les variàncies poblacionals σ_x^2 ni σ_y^2 , però podem afirmar que són iguals. Si no fos el cas, podríem fer un interval per

a estimar la igualtat entre les dues variàncies, com es vorà un poc més endavant en aquest mateix apartat. Si la interpretació ens permet estimar que són iguals continuarem calculant aquesta diferència de mitjanes.

Interval per a la diferència de mitjanes de dues poblacions amb mostres relacionades

Per a aquest cas cal definir prèviament la variable $D = X - Y$ i denotarem per $S_{D_{n-1}}^2$ la seua quasi-variància mostral, on $d_i = x_i - y_i$. La grandària de les dues mostres coincidiria necessàriament i la denotarem per n . L'interval de confiança per a la diferència de mitjanes en aquest cas és:

$$\mu_x - \mu_y \in \left(\bar{x} - \bar{y} - t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S_{D_{n-1}}^2}{n}}, \bar{x} - \bar{y} + t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\frac{S_{D_{n-1}}^2}{n}} \right)$$

Altres intervals de confiança

Interval per a la diferència de proporcions de dues poblacions Bernoulli

L'interval de confiança per a la diferència de proporcions en aquest cas és:

$$p_x - p_y \in \left(p_x - p_y - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}}, p_x - p_y + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}} \right)$$

Interval per al quocient de dues variàncies de dues poblacions normals

L'interval de confiança per al quocient de variàncies en aquest cas és:

$$\frac{\sigma_x^2}{\sigma_y^2} \in \left(\frac{\frac{(n-1)S_{X_{n-1}}^2}{n}}{\frac{(m-1)S_{Y_{n-1}}^2}{m}} \cdot \frac{1}{f_{1-\frac{\alpha}{2}, (n-1), (m-1)}}, \frac{\frac{(n-1)S_{X_{n-1}}^2}{n}}{\frac{(m-1)S_{Y_{n-1}}^2}{m}} \cdot \frac{1}{f_{\frac{\alpha}{2}, (n-1), (m-1)}} \right)$$

Errada i grandària de la mostra

Ja havíem comentat a la introducció que quan es fa un treball d'inferència hi ha tres factors que determinen la possible solució. Recordem-los:

- Nivell de confiança
- Errada de la solució
- Grandària de la mostra

Als exemples que hem treballat al llarg del desenvolupament del tema hem vist que el nivell de confiança expressa la probabilitat que l'estimació siga correcta; Determina els valors que hem trobat a les taules de les funcions de distribució de les diferents variables treballades i és costum utilitzar valors que van del 90 % al 99 %; com més gran l'agafem, l'amplitud de l'interval creix.

Seria desitjable que l'interval de la solució fóra el més ajustat possible als valors que volem inferir. L'expressió que tenim de l'errada és aquesta desviació del valor central. Així, direm que l'errada és $E = \frac{b-a}{2}$ i que, per tant, està directament relacionada amb l'amplitud de l'interval.

El tercer factor a considerar és la grandària de la mostra. Els resultats són més exactes com més gran és aquesta. Així doncs, podem plantejar-nos la qüestió a l'inrevés per a fer un treball que satisfaga les nostres necessitats.

Objectius

Els problemes han de permetre que els alumnes assolisquen els objectius didàctics següents:

- 4a) Conèixer el procés de la inferència estadística.
- 4b) Conèixer i saber diferenciar els conceptes de població i mostra.
- 4c) Conèixer els principals tipus de mostreig.
- 4d) Diferenciar paràmetre i estadístic.
- 4e) Comprendre el concepte de distribució en el mostreig d'un estadístic.
- 4f) Conèixer el concepte d'estimació puntual.
- 4g) Estudiar i interpretar els intervals de confiança per a estimar els paràmetres poblacionals d'una i dues poblacions.
- 4h) Saber interpretar els resultats dels càlculs per a poder inferir conclusions.
- 4i) Saber calcular la grandària de la mostra necessària d'acord amb les exigències de l'errada que es pretenen en les estimacions.

La taula següent ens mostra com estan distribuïts els objectius segons els exercicis:

<i>Objectiu</i> Exercici	4a	4b	4c	4d	4e	4f	4g	4h	4i
1			X	X	X				
2	X	X		X			X	X	
3							X	X	
4	X	X					X	X	
5	X	X				X			
6	X	X		X			X	X	X
7	X	X					X	X	
8	X						X	X	
9		X					X	X	X
10	X	X	X				X	X	

Enunciats

-
- 4c) Conèixer els principals tipus de mostreig.
 - 4d) Diferenciar paràmetre i estadístic.
 - 4e) Comprendre el concepte de distribució en el mostreig d'un estadístic.

Exercici 1

Tenim les variables aleatòries X (que es distribueix com una t de Student, amb 30 graus de llibertat) Y (distribuïda com una χ^2 amb 10 graus de llibertat) i Z distribuïda com una F de Fisher-Snedecor amb graus de llibertat 5 y 10. Sabent que treballem al 99 %, calcula la probabilitat per a cadascun dels següents casos:

$$P(\text{v.a.} = 4)$$

$$P(\text{v.a.} \leq 2)$$

$$P(\text{v.a.} \geq 3,3)$$

$$P(3 \leq \text{v.a.} \leq 5)$$

-
- 4a) Conèixer el procés de la Inferència estadística.
 - 4b) Conèixer i saber diferenciar els conceptes de població i mostra.
 - 4d) Diferenciar paràmetre i estadístic.
 - 4g) Estudiar i interpretar els intervals de confiança per a estimar els paràmetres poblacionals d'una i dues poblacions.
 - 4h) Saber interpretar els resultats dels càlculs per a poder inferir conclusions.

Exercici 2

Els paquets de farina envasats per una certa màquina tenen $\mu = 500\text{g}$ i $\sigma = 35\text{g}$. Els paquets s'embalen en caixes de 100 unitats.

- a) Calcula la probabilitat de que la mitjana dels pesos dels paquets d'una caixa siga menor que 495g.
- b) Calcula la probabilitat de que una caixa de 100 paquets pese més de 51 kg.

- c) A més, un altre productor de farina assegura que amb una sobrecàrrega del 20 % els sacs es prepararien en una mitjana de 12,40 minuts. Per a comprovar aquesta afirmació agafem 20 sacs, els sotmetem a una sobrecàrrega del 20 % i calculem els temps que van tardar a preparar-se; vam obtenir una mitjana de 10,63 minuts i una quasi-desviació típica de 2,48 minuts. Suposarem que les dades provenen d'una distribució normal. Podem asseverar o refutar la informació del productor?

4g) Estudiar i interpretar els intervals de confiança per a estimar els paràmetres poblacionals d'una i dues poblacions.

4h) Saber interpretar els resultats dels càlculs per a poder inferir conclusions.

Exercici 3

Suposem que el percentatge d'empreses de la Comunitat Valenciana que té només un treballador és del 20 %. Considerem una mostra de 1.000 empreses i volem esbrinar quina és la probabilitat de que almenys el 21 % d'aquestes empreses tinga només un treballador.

4a) Conèixer el procés de la inferència estadística.

4b) Conèixer i saber diferenciar els conceptes de població i mostra.

4g) Estudiar i interpretar els intervals de confiança per a estimar els paràmetres poblacionals d'una i dues poblacions.

4h) Saber interpretar els resultats dels càlculs per a poder inferir conclusions.

Exercici 4

Volem comprovar l'eficàcia d'una màquina classificadora de documents. Per a que un document es classifique dins del tipus A, cal que la seua longitud siga de mitjana 7,25 cm, i per a classificar el document de tipus B, 5,5 cm. Suposem que les desviacions típiques d'aquestes mesures de les mides dels documents són iguals en els dos tipus. Hem triat 12 documents que la màquina ha mesurat com a tipus A amb una desviació típica en la mida de 0,98 cm, i hem triat 15 documents de tipus B i hem calculat una desviació típica de 1 cm. Calca la probabilitat de que les mitjanes de la mida dels documents d'aquestes mostres tinguen una diferència major que 1,25 cm de longitud.

-
- 4a) Conèixer el procés de la inferència estadística.
 - 4b) Conèixer i saber diferenciar els conceptes de població i mostra.
 - 4f) Conèixer el concepte d'estimació puntual.
-

Exercici 5

Estudia les propietats que compleixen els estimadors S_X^2 , variància mostral, i $S_{X_{n-1}}^2$, quasi-variància mostral, com a estimadors del paràmetre σ^2 (si són esbiaixats o consistents i quin és més eficient).

-
- 4a) Conèixer el procés de la inferència estadística.
 - 4b) Conèixer i saber diferenciar els conceptes de població i mostra.
 - 4d) Diferenciar paràmetre i estadístic.
 - 4g) Estudiar i interpretar els intervals de confiança per a estimar els paràmetres poblacionals d'una i dues poblacions.
 - 4h) Saber interpretar els resultats dels càlculs per a poder inferir conclusions.
 - 4i) Saber calcular la grandària de la mostra necessària d'acord a les exigències de l'errada que es pretenen en les estimacions.
-

Exercici 6

Volem fer un estudi del nombre d'hores que els estudiants d'una certa Universitat passen davant l'ordinador. Definim X (nombre d'hores de son diàries) com la variable a estudiar. Sabem que aquesta es distribueix seguint un model normal de mitjana desconeguda, però per altres treballs podem considerar que la seua desviació típica és coneguda i el seu valor és 3. Per a estimar aquest valor de m caldrà triar una mostra; per a tal fi, farem una enquesta a 25 alumnes triats aleatòriament i, després de preguntar-los pel nombre d'hores que passen davant l'ordinador al dia, obtenim una mitjana de 7 hores entre les que gasten per a estudiar i les hores que estan en les xarxes socials i jugant. Aquesta dada la denotarem per $\bar{x} = 7$ i cal distingir que aquest és un paràmetre mostral, ja que l'hem calculat amb les dades obtingudes en l'enquesta dels individus de la mostra. Treballarem amb un nivell de confiança del 95 %. Si després de vore el resultat es creu que té errades l'estudi, com es podria millorar?

També volem confirmar el funcionament dels ordinadors que utilitzen per a comprovar que és cert que aguanten una hora davant d'ells. Per a dur a terme el treball vam seleccionar 20 alumnes triats aleatòriament i vam calcular la mitjana aritmètica del temps, que és $\bar{x} = 0,995$ hores, i la seua desviació típica, que és $S = 0,005$ hores (en aquest cas treballem al 99 %). També se sol·licita fer una estimació de la variància dels resultats per a vore l'homogeneïtat de l'estudi.

-
- 4a) Conèixer el procés de la inferència estadística.
 - 4b) Conèixer i saber diferenciar els conceptes de població i mostra.
 - 4g) Estudiar i interpretar els intervals de confiança per a estimar els paràmetres poblacionals d'una i dues poblacions.
 - 4h) Saber interpretar els resultats dels càlculs per a poder inferir conclusions.

Exercici 7

Volem comparar l'eficiència de dues empreses del sector ceràmic en l'àmbit internacional. El que volem estudiar és el temps que tarden a enviar el material per mar, a través de contenidors, als clients, després de que aquests haguen fet una encomanda de material. Sabem que el nombre de dies que tarden a arribar els enviaments de l'empresa A segueix una distribució normal de la qual coneixem la desviació típica de 25 dies i que el de l'empresa B també segueix una distribució normal amb desviació típica de 30 dies.

Per a estudiar la situació, anotem en 10 enviaments de l'empresa A un temps mitjà de 80 dies, mentre que en la mostra dels 15 enviaments de l'empresa B el temps mitjà és de 75 dies.

En el cas que treballem en dos empreses del mateix sector, C i D, a què volem mesurar el control de qualitat en la fabricació de les peces enviades, volem comparar si dos processos de producció són equivalents i si mantenen en el seu procés els mateixos estàndards de qualitat. Considerarem que les variàncies de les dues poblacions són iguals. Per a fer la comparació tenim les següents dades de mostres de cada empresa:

Empresa C: 10, 9, 7, 6, 12, 3, 7, 9, 10, 6.

Empresa D: 12, 8, 5, 11, 9, 10, 13, 7, 12, 9, 8, 13.

Amb aquesta informació volem estimar, mitjançant un interval de confiança, si la qualitat mitjana és la mateixa en els dos processos de fabricació. Treballarem amb un nivell de confiança del 95 %.

Finalment, volem estudiar si la variabilitat d'aquestes dues empreses és la mateixa.

-
- 4a) Conèixer el procés de la inferència estadística.
 - 4g) Estudiar i interpretar els intervals de confiança per a estimar els paràmetres poblacionals d'una i dues poblacions.
 - 4h) Saber interpretar els resultats dels càlculs per a poder inferir conclusions.

Exercici 8

Es pot estudiar l'eficàcia de dues impressores pel número de fulls amb defectes després d'imprimir? Per a comprovar-ho, hem seleccionat aleatòriament una mostra de 200 treballs impresos amb la impressora A, dels quals 15 eren defectuosos, i 250 treballs de la impressora B, dels quals n'han eixit 16 de defectuosos.

-
- 4b) Conèixer i saber diferenciar els conceptes de població i mostra.
 - 4g) Estudiar i interpretar els intervals de confiança per a estimar els paràmetres poblacionals d'una i dues poblacions.
 - 4h) Saber interpretar els resultats dels càlculs per a poder inferir conclusions.
 - 4i) Saber calcular la grandària de la mostra necessària d'acord a les exigències de l'errada que es pretenen en les estimacions.

Exercici 9

Estudiem l'errada en la solució d'un problema en què treballem la inferència sobre la mitjana poblacional. En la classe d'estadística tenim 25 alumnes amb una nota mitjana de 7 en el curs anterior i amb una desviació típica 3 (agafem aquesta dada com si fóra de tota la població d'estudi). Volem l'interval de confiança d'aquestes dades al 95 %. A més a més, si l'amplitud de l'interval és molt gran (major d'un punt), volem saber el nombre d'estudiants per a no tindre tanta variació en les dades.

-
- 4a) Conèixer el procés de la inferència estadística.
 - 4b) Conèixer i saber diferenciar els conceptes de població i mostra.
 - 4c) Conèixer els principals tipus de mostreig.
 - 4g) Estudiar i interpretar els intervals de confiança per a estimar els paràmetres poblacionals d'una i dues poblacions.
 - 4h) Saber interpretar els resultats dels càlculs per a poder inferir conclusions.

Exercici 10

Per als treballadors d'una empresa que tenia problemes de rendiment es va proposar una millora que consistia a deixar més temps de descans en les parades. Per a avaluar l'eficàcia d'aquest plantejament, s'ha observat l'índex de millora plantejat per la gestió de l'empresa a 8 treballadors d'una secció. Aquest índex té una escala de 0 a 3. Els resultats es detallen a continuació:

Abans: 1, 2; 1, 3; 1, 5; 1, 4; 1, 7; 1, 9; 1, 4; 1, 2.

Després: 1, 4; 1, 7; 1, 5; 1, 3; 2, 0; 2, 1; 1, 7; 1, 6.

Calculem un interval de confiança al 95 % per vore si ha donat resultat la mesura adoptada i estraure'n conclusions.

Ajudes

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de resoldre els exercicis i problemes. És convenient no fer un ús excessiu d'aquestes ajudes, és a dir, que abans d'emprar l'ajuda cal pensar el problema almenys durant uns 10 – 15 minuts. Després es consultarà l'ajuda de tipus 1 i s'intentarà resoldre l'exercici amb aquesta ajuda. Si no és possible resoldre'l, llavors es consultarà l'ajuda de tipus 2; en darrer terme es consultarà la solució.

Ajudes de tipus 1

Exercici 1

Consulteu la introducció teòrica, en concret l'apartat de models de distribució de probabilitat en el mostreig d'aquest tema.

Exercici 2

El que es necessita per a resoldre aquest exercici és, primerament, vore sobre quin paràmetre es fan les preguntes. Clarament és sobre la mitjana i les probabilitats.

També és important, en aquest tipus de problemes, vore quins paràmetres coneixem, sobretot si la variància o la desviació típica és coneguda, per a saber a quin tipus d'inferència ens referim.

Exercici 3

Consulteu la introducció teòrica, en concret l'apartat de distribució d'una proporció i com es desenvolupa en el cas que es considere que ve d'una distribució normal.

Exercici 4

Recorda que les dades donades a l'enunciat són les següents (si entenem que els valors de l'enunciat corresponen a les quasi-desviacions típiques):

Categoria A: $n = 12$	$\mu_X = 7,25$	$S_{X_{n-1}}^2 = 0,98^2 = 0,9604$
Categoria B: $m = 15$	$\mu_Y = 5,5$	$S_{Y_{n-1}}^2 = 1^2 = 1$

Amb açò, consulteu la introducció teòrica en l'apartat de comparació de intervals d'confiança per als paràmetres de dues poblacions.

Exercici 5

Per a poder fer aquest exercici cal conèixer els conceptes de biaix, consistència i eficiència d'un estadístic como a estimador. Per aquest motiu, repassa l'apartat d'estimació puntual dins de la inferència de paràmetres. A més a més, recorda que cal fer l'estimació d'un paràmetre poblacional a partir de paràmetres mostrals.

Exercici 6

Aquest exercici té com a primer problema vore què és tot el que realment s'ens demana. Vegem un llistat de coses que s'ens demanen:

- Primerament, l'estimació del paràmetre mitjana poblacional, coneguda la seua desviació típica.
- A partir de l'interval calculat, vore l'errada que es comet.
- Canviar la mostra si l'errada és molt gran i calcular de nou l'interval.
- També volem estimar el temps d'una hora seguida davant l'ordinador, per la qual cosa utilitzarem l'interval per a la mitjana μ poblacional, però cal observar que no coneixem el valor de la desviació típica de la població.
- Finalment, hem d'estimar un interval per a la variància σ^2 amb un nivell de confiança del 95 %.

Exercici 7

En aquest cas queda clar que trobem, en cada apartat del problema, una primera part on ens demanen l'interval de confiança per a la mitjana amb variància coneguda de dos mostres; un segon apartat en què ens pregunten el mateix amb variància desconeguda, però igual per a les dos empreses; finalment, ens demanen que comparem la variabilitat en les dues empreses del cas anterior, per la qual cosa hem de buscar el quocient de variàncies.

A més a més, les fórmules principals per a cada cas són:

$$a) \mu_x - \mu_y \in \left(\bar{x} - \bar{y} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}, \bar{x} - \bar{y} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \right)$$

$$b) \mu_x - \mu_y \in \left(\bar{x} - \bar{y} - z_{1-\frac{\alpha}{2}} t_{n+m-2} \sqrt{\frac{(n-1)S_{x_{n-1}}^2 + (m-1)S_{y_{m-1}}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}, \right. \\ \left. \bar{x} - \bar{y} + z_{1-\frac{\alpha}{2}} t_{n+m-2} \sqrt{\frac{(n-1)S_{x_{n-1}}^2 + (m-1)S_{y_{m-1}}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

$$c) \frac{\sigma_X}{\sigma_Y} \in \left(\frac{\frac{(n-1)S_{x_{n-1}}^2}{n}}{\frac{(m-1)S_{y_{m-1}}^2}{m}} \cdot \frac{1}{z_{1-\frac{\alpha}{2}} f_{(n-1), (m-1)}}, \frac{\frac{(n-1)S_{x_{n-1}}^2}{n}}{\frac{(m-1)S_{y_{m-1}}^2}{m}} \cdot \frac{1}{z_{\frac{\alpha}{2}} f_{(n-1), (m-1)}} \right)$$

Exercici 8

En aquest cas, el major problema és vore sobre quin paràmetre hem de fer inferència. Com que no s'ens pregunta sobre mitjanes, sinó sobre eficàcia, ens referirem a la proporció de defectes sobre el total de la producció.

També pot ser un problema vore les proporcions a utilitzar, que en aquest cas seran:

$$\text{Per a la impressora A: } p_X = \frac{15}{200} = 0,075.$$

$$\text{Per a la impressora B: } p_Y = \frac{16}{250} = 0,064.$$

Exercici 9

En aquest cas ens centrem, primerament, en conèixer el problema i plantejar quina és la fórmula a utilitzar per a cada apartat:

$$a) \mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

$$b) E = \frac{b-a}{2} = \frac{\left(\bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)}{2}$$

Exercici 10

En aquest exercici queda clar que comparem dues mostres: les dades d'abans i les de després de la millora en la productivitat. Però hi ha un element que diferencia aquest problema dels anteriors pel que fa la comparació de mostres: que són dades relacionades, per la qual cosa serà necessari calcular la diferència entre les dues mostres.

Com que són dades relacionades, necessitarem calcular els valors de la variable $D = X - Y$ restant els valors emparellats, tal com es veu a la taula següent:

X = valors abans de la mesura	1,2	1,3	1,5	1,4	1,7	1,9	1,4	1,2
Y = valors després de la mesura	1,4	1,7	1,5	1,3	2	2,1	1,7	1,6
D = X - Y	-0,2	-0,4	0	0,1	-0,3	-0,2	-0,3	-0,4

Ajudes de tipus 2

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de resoldre els exercicis i problemes, després d'haver consultat l'ajuda de tipus 1.

Exercici 1

A més a més de l'ajuda teòrica, cal conèixer el funcionament de les taules de les distintes distribucions que se'ns demanen en aquest problema.

Respecte de la taula t de Student, buscarem en la primera columna el nombre corresponent als graus de llibertat, en aquest cas 30, i en la primera fila el valor del quantil corresponent, en aquest cas 0,99. Respecte de la taula Xi-quadrat, de la mateixa forma que en el cas anterior, buscarem en la primera columna el valor dels graus de llibertat, en aquest cas 10, i en la primera fila el valor del quantil corresponent.

Per contra, en les taules de la F de Fisher-Snedecor, el primer pas és buscar la taula corresponent al quantil que es pregunta; després, en la primera fila, el primer grau de llibertat i, en la primera columna, el segon grau de llibertat.

Per als tres casos, cal tenir en compte que, al ser distribucions contínues, no podem calcular una probabilitat igual a un valor, en aquest cas serà 0.

Exercici 2

Si no tens prou amb la informació de l'ajuda de tipus 1, busca en les taules, formularis o en la teoria i trobaràs les fórmules necessàries per a cada apartat, que són:

- a) Les mitjanes es distribueixen en aquest problema de la forma següent: $\bar{X} \rightarrow N(500; 3,5)$, i la forma de la de la probabilitat a calcular és:

$$P(\bar{X} < 495) = P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{495 - 500}{3,5}\right)$$

- b) Per al segon cas passem a una distribució: $\sum_{i=1}^n X_i \rightarrow N(50000; 350)$.

$$P\left(\sum_{i=1}^n X_i > 51000\right) = P\left(\frac{\sum_{i=1}^n X_i - 50000}{350} > \frac{51000 - 50000}{350}\right)$$

- c) En aquest cas, el valor de la variable T, segons les nostres dades poblacionals i mostrals, serà:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Exercici 3

En aquest cas, la variable aleatòria p, que indica les proporcions mostrals, té una distribució normal.

$$p \rightarrow N\left(\frac{np}{n}; \frac{\sqrt{npq}}{n}\right)$$

A més a més, el càlcul de les probabilitats es pot fer tipificació perquè ve d'una distribució normal.

Exercici 4

Encara que amb l'ajuda de tipus 1 ja es podria resoldre el problema, si es revisen les fórmules es comprovarà que l'estadístic que s'ha d'utilitzar és:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{(n-1)S_{X_{n-1}}^2 + (m-1)S_{Y_{m-1}}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Les dades necessàries ja s'han comentat en la primera ajuda.

Exercici 5

Per a veure si un estimador és esbiaixat, hem de calcular l'esperança i la variància i veure que es complez: $E[X_i] = \mu$ i $Var[X_i] = \sigma^2$.

A més a més, recorda les fórmules dels dos possibles estimadors:

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2}{n-1} - \bar{X}_n^2$$
$$S_{X_{n-1}}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{n}{n-1} \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{n}{n-1} S_{X_1}^2$$

Pel que fa a la consistència a partir dels límits:

Del primer estimador: $\lim_{n \rightarrow \infty} E[S_X^2] \mid \lim_{n \rightarrow \infty} Var[S_X^2]$.

Del segon estimador: $\lim_{n \rightarrow \infty} E[S_{X_{n-1}}^2] \mid \lim_{n \rightarrow \infty} Var[S_{X_{n-1}}^2]$.

Per a garantir l'eficiència dels estimadors, ho farem mitjançant el quocient de les variàncies.

Exercici 6

Anem a veure què és necessari per a cada cosa que es demana:

- Per a l'estimació del paràmetre «mitjana poblacional», coneguda la seua desviació típica, utilitzarem la següent fórmula:

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

També cal recordar el càlcul de $1 - \frac{\alpha}{2}$. $1 - \alpha$ és el nivell de confiança que establím *a priori* per fer el treball

$$1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow \frac{\alpha}{2} = 0,025 \rightarrow 1 - \frac{\alpha}{2} = 0,975$$

- L'errada que es comet la calcularem a partir de la següent fórmula:

$$E = \frac{b - a}{2}, \text{ on } b \text{ i } a \text{ són els límits de l'interval.}$$

- Per a canviar la mostra si l'errada és molt gran i calcular de nou l'interval, només cal pensar amb un canvi de n , com per exemple passar a 30 alumnes. L'altra opció és canviar el nivell de confiança, com ara passar al 90 %.

També volem estimar el temps d'una hora sense pauses davant l'ordinador. En aquest cas, com que no coneixem la desviació típica poblacional, utilitzarem:

$$\mu \in \left(\bar{x} - t_{1-\frac{\alpha}{2}, n-1} \frac{S_{n-1}}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2}, n-1} \frac{S_{n-1}}{\sqrt{n}} \right)$$

- En l'últim apartat, per a calcular la variància:

$$\sigma^2 \in \left(\frac{(n-1)S_{n-1}^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}, \frac{(n-1)S_{n-1}^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \right)$$

I per a poder comparar aquest paràmetre amb el valor mostral de l'enunciat, cal calcular l'interval de la desviació típica.

Exercici 7

No només és important conèixer quines són les fórmules que utilitzarem en cada cas, també ho és saber interpretar-ne els resultats.

Tant per al cas a com per al b : en tots els intervals en què ens plantegem la diferència de paràmetres cal considerar tres possibilitats per a interpretar la solució.

Suposem que l'interval solució és $\mu_x - \mu_y \in (a, b)$.

- Si $0 \in (a, b)$, interpretarem que la diferència entre els paràmetres no és significativa i podrem estimar que $\mu_x \approx \mu_y$.
- Si $a > 0$ i $b > 0$, interpretarem que la diferència és positiva en una alta probabilitat, per la qual cosa podrem estimar que $\mu_x - \mu_y > 0 \rightarrow \mu_x > \mu_y$.
- Si $a < 0$ i $b < 0$, interpretarem que la diferència és negativa en una alta probabilitat, per la qual cosa podrem estimar que $\mu_x - \mu_y < 0 \rightarrow \mu_x < \mu_y$.

En els dos darrers casos, els valors dels intervals ens donaran una idea aproximada dels valors al voltant dels quals oscil·la la diferència entre els paràmetres.

En el cas *c*, els valors dels extrems de l'interval ens permeten inferir que les variàncies de les poblacions de les quals provenen les mostres són iguals, és a dir, que els diferents productes de cadascuna de les línies de fabricació que comparem presenten el mateix comportament en la dispersió dels valors de qualitat respecte de la mitjana de cadascuna d'elles.

Aquesta inferència està basada en el fet que el valor 1 és dintre de l'interval calculat, ja que com que estem comparant les variàncies mitjançant la seua ràtio, o quocient, els valors dels extrems ens diuen que els quocients estan al voltant de la unitat, i aquesta solució permet inferir igualtat entre les dues mesures comparades per divisió.

En general, quan comparem les variàncies de dues poblacions mitjançant l'expressió de l'interval que hem mostrat en aquest apartat, podem arribar a tres possibles resultats, dels quals en comentem a continuació la interpretació, sempre tenint en compte que es tracta de l'anàlisi d'uns quocients.

Si considerem que l'interval solució és $\frac{\sigma_x^2}{\sigma_y^2} \in (a, b)$:

- Si $1 \in (a, b)$, interpretarem que la diferència entre els paràmetres no es significativa i podrem estimar que $\sigma_x^2 \approx \sigma_y^2$.
- Si $a > 1$ i $b > 1$, interpretarem que el quocient és major que 1 en una alta probabilitat, per la qual cosa podrem estimar que $\frac{\sigma_x^2}{\sigma_y^2} > 1 \rightarrow \sigma_x^2 > \sigma_y^2$.
- Si $a < 1$ i $b < 1$, interpretarem que el quocient és menor que 1 en una alta probabilitat, per la qual cosa podrem estimar que $\frac{\sigma_x^2}{\sigma_y^2} < 1 \rightarrow \sigma_x^2 < \sigma_y^2$.

En els dos darrers casos, els valors dels intervals ens donaran una idea aproximada dels valors al voltant dels quals oscil·la el quocient entre els paràmetres.

Exercici 8

Amb les dades de les proporcions, la fórmula a utilitzar serà:

$$p_x - p_y \in \left(p_x - p_y - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}}, p_x - p_y + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}} \right)$$

A més a més, el problema no ens especifica el nivell de confiança per a treballar. Per defecte podem treballar al 95 % i, com que treballarem amb un nivell de confiança del 95 %, podem calcular el valor de $1 - \frac{\alpha}{2}$, ja que:

$$1 - \alpha = 0,95 \quad \rightarrow \quad \alpha = 0,05 \quad \rightarrow \quad \frac{\alpha}{2} = 0,025 \quad \rightarrow \quad 1 - \frac{\alpha}{2} = 0,975$$

En la taula de la distribució acumulada de la variable normal tipificada busquem el valor de la variable x , tal que, $P(Z \leq x) = 0,975$. Aquest valor és $x = 1,96$, el qual denotem per $z_{1-\frac{\alpha}{2}} = 1,96$.

Exercici 9

La següent ajuda per al problema se centra en el segon apartat, ja que es diferencia en altres casos en en sentit que ens proposen quina ha de ser l'amplitud de l'errada, que en aquest cas serà com a màxim d'un punt. Per açò, el desenvolupament de la fórmula serà:

$$E = \frac{b-a}{2} = \frac{\left(\bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)}{2} = \frac{2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}{2} = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1,96 \frac{3}{\sqrt{n}} = 0,5$$

D'on traurem el valor de n que es demana.

Exercici 10

A més de ser mostres relacionades, hem de proposar un nivell de confiança, el qual pot ser del 95 %.

$$1 - \alpha = 0,95 \quad \rightarrow \quad \alpha = 0,05 \quad \rightarrow \quad \frac{\alpha}{2} = 0,025 \quad \rightarrow \quad 1 - \frac{\alpha}{2} = 0,975$$

A més a més, la fórmula per a aquest cas, diferent dels anteriors, serà:

$$\mu_x - \mu_y \in \left(\bar{x} - \bar{y} - t_{n-1} \sqrt{\frac{S_{D_{n-1}}^2}{n}}, \bar{x} - \bar{y} + t_{n-1} \sqrt{\frac{S_{D_{n-1}}^2}{n}} \right)$$

Cal vore que s'ha de calcular tot sobre les variables i la diferència entre les dues, ja que són relacionades (variable D).

Solucions

Exercici 1

Tenim les variables aleatòries X (que es distribueix com una t de Student, amb 30 graus de llibertat) Y (distribuïda com una Xi-quadrat amb 10 graus de llibertat) i Z distribuïda com una F de Fischer-Snedecor amb graus de llibertat 5 y 10. Sabent que treballem al 99 %, calcula la probabilitat per a cadascun dels següents casos:

$$P(\text{v.a.} = 4)$$

$$P(\text{v.a.} \leq 2)$$

$$P(\text{v.a.} \geq 3,3)$$

$$P(3 \leq \text{v.a.} \leq 5)$$

Solució

Els resultats, després de consultar les taules i fer els càlculs pertinents, són:

Per al cas de la t de Student:

$$P(X = 4) = 0$$

$$P(X \leq 2) = 0,9727$$

$$P(X \geq 3,3) = 1 - P(X \leq 3,3) = 0,00124$$

$$P(3 \leq X \leq 5) = P(X \leq 5) - P(X \leq 3) = 0,999988 - 0,9973 = 0,002688$$

Per al cas de la Xi-quadrat:

$$P(Y = 4) = 0$$

$$P(Y \leq 2) = 0,9963$$

$$P(Y \geq 3,3) = 1 - P(Y \leq 3,3) = 1 - 0,9735 = 0,02654$$

$$P(3 \leq Y \leq 5) = 0,0902$$

Per al cas de la F-Fisher-Snedecor:

$$P(Y = 4) = 0$$

$$P(Y \leq 2) = 0,1641$$

$$P(Y \geq 3,3) = 1 - P(Y \leq 3,3) = 0,9489$$

$$P(3 \leq Y \leq 5) = P(Y \leq 5) - P(Y \leq 3) = 0,0507$$

Exercici 2

Els paquets de farina envasats per una certa màquina tenen $\mu = 500\text{g}$ i $\sigma = 35\text{g}$. Els paquets s'embalen en caixes de 100 unitats.

- Calcula la probabilitat de que la mitjana dels pesos dels paquets d'una caixa siga menor que 495g.
- Calcula la probabilitat de que una caixa de 100 paquets pese més de 51 kg.
- A més, un altre productor de farina assegura que amb una sobrecàrrega del 20 % els sacs es prepararien en una mitjana de 12,40 minuts. Per a comprovar aquesta afirmació agafem 20 sacs, els sotmetem a una sobrecàrrega del 20 % i calculem els temps que van tardar a preparar-se; vam obtenir una mitjana de 10,63 minuts i una quasi-desviació típica de 2,48 minuts. Suposarem que les dades provenen d'una distribució normal. Podem asseverar o refutar la informació del productor?

Solució

- Comencem per identificar la totalitat de paquets de farina envasats per la màquina com la població de mitjana $\mu = 500\text{ g}$ i desviació típica $\sigma = 35\text{ g}$. Cada caixa serà una mostra de grandària $n = 100$.

Les mitjanes dels pesos dels paquets d'una caixa es distribueixen com una variable normal de mitjana $\mu = 500\text{ g}$ i desviació típica $\frac{\sigma}{\sqrt{n}} = \frac{35}{\sqrt{100}} = 3,5$. Així doncs, $\bar{X} \rightarrow N(500; 3,5)$.

Per a calcular la probabilitat de que la mitjana dels pesos dels paquets d'una caixa siga menor que 495 g, transformarem la variable \bar{X} mitjançant una tipificació i acudirem a un paquet informàtic per a trobar el valor de la variable en la funció de distribució de la variable normal tipificada Z .

$$\begin{aligned} P(\bar{X} < 495) &= P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{495 - 500}{3,5}\right) = P(Z < -1,43) = \phi(-1,43) = 1 - \phi(1,43) = \\ &= 1 - 0,9236 = 0,0764 \end{aligned}$$

- Per a calcular la probabilitat de que una caixa de 100 paquets pese més de 51 kg, considerarem la variable $\sum_{i=1}^n X_i$, que també es distribueix com una normal, però de mitjana $n\mu$ i de desviació típica $n\frac{\sigma}{\sqrt{n}} = \sigma\sqrt{n}$. Per tant, $\sum_{i=1}^n X_i \rightarrow N(50000; 350)$.

$$P\left(\sum_{i=1}^n X_i > 51000\right) = P\left(\frac{\sum_{i=1}^n X_i - 50000}{350} > \frac{51000 - 50000}{350}\right) = P(Z > 2,86) =$$

$$= 1 - P(Z \leq 2,86) = 1 - \phi(2,86) = 1 - 0,9979 = 0,0021$$

c) Calculem el valor de la variable T amb les nostres dades poblacionals i mostrals, així:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{10,63 - 12,40}{2,48/\sqrt{20}} = -3,19$$

El resultat serà un valor de la distribució t de student amb 19 graus de llibertat. Si l'afirmació és certa, seria lògic que el nostre valor estigués, per exemple, dintre del 90 % dels valors centrals d'aquesta distribució. Així, calculem els valors a i b, els quals compleixen:

$P(t \leq a) = 0,05$ i $P(t \geq b) = 0,05$, és a dir: $P(t \leq b) = 0,95$. Si el valor $t = -3,19$ està fora d'aquest interval, direm que la nostra prova no corrobora l'afirmació del productor.

Amb l'ajut d'un programa d'estadística calculem $a = -1,73$ i, per simetria, $b = +1,73$, per la qual cosa el valor que hem obtingut està fora d'aquest interval i ens porta a refutar la informació prèvia que tenim del valor de μ .

Exercici 3

Suposem que el percentatge d'empreses de la Comunitat Valenciana que té només un treballador és del 20 %. Considerem una mostra de 1.000 empreses i volem esbrinar quina és la probabilitat de que almenys el 21 % d'aquestes empreses tinga només un treballador.

Solució

En aquest cas, $n = 1000$ i $p = 0,20$, per tant, la variable aleatòria que indica les proporcions mostrals p té una distribució normal.

$$p \rightarrow N\left(\frac{np}{n}; \sqrt{\frac{npq}{n}}\right) = N\left(p; \sqrt{\frac{pq}{n}}\right) = N\left(0,2; \sqrt{\frac{0,2 \cdot 0,8}{1000}}\right) = N(0,2; 0,0126)$$

Amb aquesta distribució, la probabilitat que ens plantegem serà:

$$P(p \geq 0,21) = P\left(\frac{p - \mu}{\sigma} \geq \frac{0,21 - 0,20}{0,0126}\right) = P(Z \geq 0,79) = 1 - P(Z \leq 0,79) = 0,2148$$

Podem considerar que és una bona aproximació, ja que $p \cdot n = 0,2 \cdot 1000 = 20 > 5$ i $q \cdot n = 0,8 \cdot 1000 = 800 > 5$.

Exercici 4

Volem comprovar l'eficàcia d'una màquina classificadora de documents. Per a que un document es classifiqui dins del tipus A, cal que la seua longitud siga de mitjana 7,25 cm, i per a classificar el document de tipus B, 5,5 cm. Suposem que les desviacions típiques d'aquestes mesures de les mides dels documents són iguals en els dos tipus. Hem triat 12 documents que la màquina ha mesurat com a tipus A amb una desviació típica en la mida de 0,98 cm, i hem triat 15 documents de tipus B i hem calculat una desviació típica de 1 cm. Calca la probabilitat de que les mitjanes de la mida dels documents d'aquestes mostres tinguen una diferència major que 1,25 cm de longitud.

Solució

Les dades donades a l'enunciat són les següents, si entenem que els valors de l'enunciat corresponen a les quasi-desviacions típiques:

$$\text{Categoria A: } n = 12 \quad \mu_X = 7,25 \quad S_{X_{n-1}}^2 = 0,98^2 = 0,9604$$

$$\text{Categoria B: } m = 15 \quad \mu_Y = 5,5 \quad S_{Y_{m-1}}^2 = 1^2 = 1$$

Podem veure la distribució de la variable T si substituïm els valors de la mostra i la diferència entre les mitjanes de les poblacions corresponents. Així:

$$\begin{aligned} T &= \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{(n-1)S_{X_{n-1}}^2 + (m-1)S_{Y_{m-1}}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{1,25 - (7,25 - 5,5)}{\sqrt{\frac{11 \cdot 0,9604 + 14 \cdot 1}{25}} \sqrt{\frac{1}{12} + \frac{1}{15}}} = \frac{-0,5}{\sqrt{0,9826} \cdot \sqrt{0,15}} = \\ &= \frac{-0,5}{0,9912 \cdot 0,3873} = \frac{-0,5}{0,3839} = -1,3024 \end{aligned}$$

La probabilitat de que la diferència de les mitjanes siga superior a 1,25 equival a calcular la probabilitat de que aquest estadístic tinga valors majors que $-1,3024$. Així, calcularem, amb l'ajut d'un programa o de les taules, aquesta probabilitat en una distribució t de Student amb 25 graus de llibertat.

$$P(\bar{X} - \bar{Y} > 1,25) = P(T > -1,3024) = 0,8977$$

Exercici 5

Estudia les propietats que compleixen els estimadors S_X^2 , variància mostral, i $S_{X_{n-1}}^2$, quasi-variància mostral, com a estimadors del paràmetre σ^2 (si són esbiaixats o consistents i quin és més eficient).

Solució

a) S_X^2 és un estimador esbiaixat i $S_{X_{n-1}}^2$ és un estimador no esbiaixat de s^2 .

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}_n^2 \text{ amb } E[X_i] = \mu \text{ i } Var[X_i] = \sigma^2$$
$$E[S_X^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right] = E\left[\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2\right] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] - E[\bar{X}^2]$$

D'altra banda:

$$\begin{cases} \overbrace{Var[X_i]}^{\sigma^2} = E[X_i^2] - \underbrace{E[X_i]^2}_{\mu^2} \Rightarrow E[X_i^2] = \sigma^2 + \mu^2 \\ \overbrace{Var[\bar{X}]}^{\frac{\sigma^2}{n}} = E[\bar{X}^2] - \underbrace{E[\bar{X}]^2}_{\mu^2} \Rightarrow E[\bar{X}^2] = \frac{\sigma^2}{n} + \mu^2 \end{cases}$$

Substituïm aquests resultats en l'expressió anterior:

$$E[S_X^2] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] - E[\bar{X}^2] = \frac{1}{n} \sum_{i=1}^n (\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) = \frac{n(\sigma^2 + \mu^2)}{n} - \left(\frac{\sigma^2}{n} + \mu^2\right) =$$
$$= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2$$

Així, podem concloure que $E[S_X^2] = \frac{n-1}{n} \sigma^2$.

Estudiem ara, l'esperança de l'altre estimador:

$$S_{X_{n-1}}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{n}{n-1} \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{n}{n-1} S_X^2$$

$$E[S_{X_{n-1}}^2] = E\left[\frac{n}{n-1} S_X^2\right] = \frac{n}{n-1} E[S_X^2] = \frac{n}{n-1} \frac{n-1}{n} \sigma^2 = \sigma^2$$

Hem demostrat, doncs, que S_X^2 és un estimador esbiaixat i que $S_{X_{n-1}}^2$ és un estimador no esbiaixat de σ^2 .

b) Per a provar la consistència dels estimadors S_X^2 i $S_{X_{n-1}}^2$ considerarem la distribució de l'estimador següent (teorema de Cochran), el qual ja hem vist a l'apartat 5 d'aquest mateix tema.

$\frac{nS_X^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \rightarrow \chi_{n-1}^2$, i com que coneixem l'esperança i la variància d'aquesta distribució, podem concloure:

$$E\left[\frac{nS_X^2}{\sigma^2}\right] = n-1 \Leftrightarrow E[S_X^2] = \frac{n-1}{n} \sigma^2$$

$$Var\left[\frac{nS_X^2}{\sigma^2}\right] = 2(n-1) \Leftrightarrow \frac{n^2}{\sigma^4} Var[S_X^2] = 2(n-1) \Leftrightarrow Var[S_X^2] = \frac{2(n-1)}{n^2} \sigma^4$$

b.1. Podem demostrar fàcilment, a partir dels resultats anteriors, que la variància poblacional S_X^2 és un estimador consistent de σ^2 :

$$\lim_{n \rightarrow \infty} E[S_X^2] = \lim_{n \rightarrow \infty} E\left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}\right] = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2$$

$$\lim_{n \rightarrow \infty} Var[S_X^2] = \lim_{n \rightarrow \infty} Var\left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}\right] = \lim_{n \rightarrow \infty} \frac{2(n-1)}{n^2} \sigma^4 = 0$$

b.2. Provem que la quasi-variància $S_{X_{n-1}}^2$ és també un estimador consistent de σ^2 .
A tal fi, comencem per demostrar el valor de la seua variància a partir de la de l'estimador S_X^2 .

$$Var[S_{X_{n-1}}^2] = Var\left[\frac{n}{n-1} S_X^2\right] = \left(\frac{n}{n-1}\right)^2 Var[S_X^2] = \left(\frac{n}{n-1}\right)^2 \cdot \frac{2(n-1)}{n^2} \cdot \sigma^4 = \frac{2}{(n-1)} \sigma^4$$

$$\lim_{n \rightarrow \infty} E[S_{X_{n-1}}^2] = \lim_{n \rightarrow \infty} \sigma^2 = \sigma^2$$

$$\lim_{n \rightarrow \infty} Var[S_{X_{n-1}}^2] = \lim_{n \rightarrow \infty} \frac{2}{(n-1)} \sigma^4 = 0$$

c) Comparem l'eficiència dels estimadors $S_{X_{n-1}}^2$ i S_X^2 , mitjançant el quocient de les variàncies:

$$\text{Var}[S_X^2] = \frac{2(n-1)}{n^2} \sigma^4 \qquad \text{Var}[S_{X_{n-1}}^2] = \frac{2}{(n-1)} \sigma^4$$

$$\frac{\text{Var}[S_X^2]}{\text{Var}[S_{X_{n-1}}^2]} = \frac{\frac{2(n-1)}{n^2} \sigma^4}{\frac{2}{(n-1)} \sigma^4} = \frac{2(n-1)(n-1)}{2n^2} = \frac{(n-1)^2}{n^2} < 1 \Rightarrow \text{Var}[S_X^2] < \text{Var}[S_{X_{n-1}}^2]$$

Podem concloure que l'estimador S_X^2 és més eficient que $S_{X_{n-1}}^2$.

Com a conseqüència d'aquestes propietats, podem veure que les expressions algebraïques d'estimadors i estadístics, així com les fórmules dels intervals de confiança del tema següent, estan en funció de l'expressió de la quasi-variància. Per a transformar un estimador en un altre tan sols cal recordar la relació que hi ha entre

ells. $S_{X_{n-1}}^2 = \frac{n}{n-1} S_X^2$; de fet és freqüent trobar bibliografia que utilitza una o l'altra a criteri de l'autor.

Exercici 6

Volem fer un estudi del nombre d'hores que els estudiants d'una certa Universitat passen davant l'ordinador. Definim X (nombre d'hores de son diàries) com la variable a estudiar. Sabem que aquesta es distribueix seguint un model normal de mitjana desconeguda, però per altres treballs podem considerar que la seua desviació típica és coneguda i el seu valor és 3. Per a estimar aquest valor de m caldrà triar una mostra; per a tal fi, farem una enquesta a 25 alumnes triats aleatòriament i, després de preguntar-los pel nombre d'hores que passen davant l'ordinador al dia, obtenim una mitjana de 7 hores entre les que gasten per a estudiar i les hores que estan en les xarxes socials i jugant. Aquesta dada la denotarem per $\bar{x} = 7$ i cal distingir que aquest és un paràmetre mostral, ja que l'hem calculat amb les dades obtingudes en l'enquesta dels individus de la mostra. Treballarem amb un nivell de confiança del 95 %. Si després de veure el resultat es creu que té errades l'estudi, com es podria millorar?

També volem confirmar el funcionament dels ordinadors que utilitzen per a comprovar que és cert que aguanten una hora davant d'ells. Per a dur a terme el treball vam seleccionar 20 alumnes triats aleatòriament i vam calcular la mitjana aritmètica del temps, que és $\bar{x} = 0,995$ hores, i la seua desviació típica, que és $S = 0,005$ hores (en aquest cas treballem al 99 %). També se sol·licita fer una estimació de la variància dels resultats per a veure l'homogeneïtat de l'estudi.

Solució

Així doncs, per tractar-se de l'estimació del paràmetre mitjana poblacional, coneguda la seua desviació típica ($\sigma = 3$), utilitzarem:

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Cal insistir en el fet que aquesta desviació és un paràmetre poblacional que ja coneixem per treballs anteriors i no obtingut a partir de les dades de l'enquesta de la mostra. Així: podem observar que coneixem tots els valors a substituir llevat del valor $z_{1-\frac{\alpha}{2}}$. Per calcular-lo començarem per deduir fàcilment el valor de $1 - \frac{\alpha}{2}$, ja que, com hem dit abans, $1 - \alpha$ és el nivell de confiança que establim *a priori* per a fer el treball.

$$1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow \frac{\alpha}{2} = 0,025 \rightarrow 1 - \frac{\alpha}{2} = 0,975$$

Tanmateix, busquem el valor de la variable x tal que $P(Z \leq x) = 0,975$. És en realitat el valor d'un percentil que podem obtenir de la taula de la funció de distribució de la variable normal tipificada. Aquest valor és $x = 1,96$, per la qual cosa, segons

la notació d'aquest tema, $z_{1-\frac{\alpha}{2}} = 1,96$, o també $z_{0,975} = 1,96$.

Ara, substituïrem tots els valors en la fórmula per a obtenir l'interval:

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = \left(7 - 1,96 \frac{3}{\sqrt{25}}, 7 + 1,96 \frac{3}{\sqrt{25}} \right) = (5,824, 8,176)$$

Podem concloure que, amb una probabilitat del 95 %, la mitjana del nombre d'hores davant l'ordinador que passen alumnes de la Universitat està dintre d'aquest interval.

Donat que l'amplitud de l'interval és $8,176 - 5,824 = 2,352$ hores, l'errada en la nostra estimació és de $E = \frac{b-a}{2} = \frac{2,352}{2} = 1,176$ hores.

Si aquesta errada la considerem excessiva, podem comprovar la diferència si la nostra mostra augmentara a 30 individus. Suposarem que el valor de la mitjana mostral hagués estat la mateixa: $\bar{x} = 7$.

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = \left(7 - 1,96 \frac{3}{\sqrt{30}}, 7 + 1,96 \frac{3}{\sqrt{30}} \right) = (5,926, 8,074)$$

També podem observar l'efecte que es produeix en l'amplitud de l'interval si rebaixem el nivell de confiança al 90 % per exemple, i mantenim la grandària de la mostra en $n = 25$ per a contrastar l'efecte d'aquest darrer canvi.

En aquests cas caldrà tornar a buscar a les taules el valor de $z_{1-\frac{\alpha}{2}}$.

$$1 - \alpha = 0,9 \rightarrow \alpha = 0,1 \rightarrow \frac{\alpha}{2} = 0,05 \rightarrow 1 - \frac{\alpha}{2} = 0,95$$

Així, buscarem aquesta probabilitat a la taula de la distribució acumulada de la variable normal tipificada i obtindrem $z_{1-\frac{\alpha}{2}} = 1,65$, la qual cosa ens dóna l'interval:

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = \left(7 - 1,65 \frac{3}{\sqrt{25}}, 7 + 1,65 \frac{3}{\sqrt{25}} \right) = (6,01, 7,99)$$

Com es pot observar, aquest interval és més menut i la resposta és més ajustada, però les conclusions han perdut nivell de confiança.

També volem estimar el temps d'una hora seguida davant l'ordinador, per la qual cosa utilitzarem l'interval per a la mitjana μ poblacional, però cal observar que no coneixem el valor de la desviació típica de la població. També cal considerar que la grandària de la mostra, $n = 20$, és inferior a 30 i no podem estimar aquesta desviació σ basant-nos en el valor de la quasi-desviació típica mostral S_{n-1} . En aquest cas, doncs, empremem la fórmula següent:

$$\mu \in \left(\bar{x} - t_{1-\frac{\alpha}{2}, n-1} \frac{S_{n-1}}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2}, n-1} \frac{S_{n-1}}{\sqrt{n}} \right)$$

Treballarem amb el 99 % de nivell de confiança en les nostres estimacions.

Com podem veure, cal calcular la quasi-desviació típica S_{n-1} a partir dels valor que coneixem de $S = 0,005$ i $n = 20$.

$$S_{n-1} = \sqrt{\frac{n}{n-1}} S = \sqrt{\frac{20}{19}} 0,005 = 0,005130$$

També cal esbrinar el valor de $t_{1-\frac{\alpha}{2}, n-1}$ amb la taula de la variable t de Student o

amb l'ajut d'algun programa. En primer lloc, calcularem el valor de $1 - \frac{\alpha}{2}$, ja que

$$1 - \alpha = 0,99 \rightarrow \alpha = 0,01 \rightarrow \frac{\alpha}{2} = 0,005 \rightarrow 1 - \frac{\alpha}{2} = 0,995$$

En la taula de la distribució acumulada de la variable t de Student amb 19 graus de llibertat, trobem el valor de la variable x, tal que $P(t_{19} \leq x) = 0,995$. Aquest valor és $x = 2,861$, per la qual cosa, utilitzant la notació en aquest tema, ${}_{1-\frac{\alpha}{2}}t_{n-1} = {}_{0,995}t_{19} = 2,861$.

Substituïm aquests valors en la fórmula per a obtenir:

$$\mu \in \left(\bar{x} - {}_{1-\frac{\alpha}{2}}t_{n-1} \frac{S_{n-1}}{\sqrt{n}}, \bar{x} + {}_{1-\frac{\alpha}{2}}t_{n-1} \frac{S_{n-1}}{\sqrt{n}} \right) = \left(0,995 - 2,861 \frac{0,005130}{\sqrt{20}}; 0,995 + 2,861 \frac{0,005130}{\sqrt{20}} \right) = (0,9917, 0,9983)$$

Aquest interval ens indica que els estudiants no estan realment una hora davant l'ordinador, sinó que estan menys temps.

Respecte de la variància, considerarem els mateixos valors mostrals: la grandària de la mostra $n = 20$, la mitjana aritmètica dels pesos, que és $\bar{x} = 0,995$ hores, i la seua desviació típica $S = 0,005$ hores.

Estimarem un interval per a la variància σ^2 amb un nivell de confiança del 95 %, per a la qual cosa utilitzarem la fórmula d'abans.

Necessitem calcular la quasi-variància de la mostra i la podem obtenir a partir de la variància de la mostra:

$$S_{n-1}^2 = \frac{n}{n-1} S^2 = \frac{20}{19} 0,000025 = 0,00002632$$

En aquest cas el nivell de confiança escollit és del 95 %, és a dir, $1 - \alpha = 0,95$.

$$1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow \frac{\alpha}{2} = 0,025 \rightarrow 1 - \frac{\alpha}{2} = 0,975$$

En la taula de la distribució acumulada de la variable Xi-quadrat amb 19 graus de llibertat, trobem els valors de ${}_{1-\frac{\alpha}{2}}\chi_{n-1}^2 = {}_{0,975}\chi_{19}^2$ i de ${}_{\frac{\alpha}{2}}\chi_{n-1}^2 = {}_{0,025}\chi_{19}^2$, ja que, en no ser simètrica la variable, obtenim dos valors diferents en el valor absolut per a cada probabilitat.

En el nostre cas obtenim ${}_{0,975}\chi_{19}^2 = 32,8523$ i per a ${}_{0,025}\chi_{19}^2 = 8,90652$. Si ara substituïm tots aquests valors en la fórmula de l'interval, obtenim:

$$\sigma^2 \in \left(\frac{(n-1)S_{n-1}^2}{{}_{1-\frac{\alpha}{2}}\chi_{n-1}^2}, \frac{(n-1)S_{n-1}^2}{{}_{\frac{\alpha}{2}}\chi_{n-1}^2} \right) = \left(\frac{19 \cdot 0,00002632}{32,8523}, \frac{19 \cdot 0,00002632}{8,90652} \right) = (0,00001522, 0,00005614)$$

Per a poder comparar aquest paràmetre amb el valor mostral de l'enunciat, podem calcular l'interval per a estimar la desviació típica simplement calculant l'arrel quadrada dels extrems de l'interval anterior.

$$\sigma \in (\sqrt{0,00001522}, \sqrt{0,00005614}) = (0,003901, 0,000749)$$

Així, podem estimar, amb un nivell de confiança del 95 % que la desviació típica.

Exercici 7

Volem comparar l'eficiència de dues empreses del sector ceràmic en l'àmbit internacional. El que volem estudiar és el temps que tarden a enviar el material per mar, a través de contenidors, als clients, després de que aquests haguen fet una encomanda de material. Sabem que el nombre de dies que tarden a arribar els enviaments de l'empresa A segueix una distribució normal de la qual coneixem la desviació típica de 25 dies i que el de l'empresa B també segueix una distribució normal amb desviació típica de 30 dies.

Per a estudiar la situació, anotem en 10 enviaments de l'empresa A un temps mitjà de 80 dies, mentre que en la mostra dels 15 enviaments de l'empresa B el temps mitjà és de 75 dies.

En el cas que treballem en dos empreses del mateix sector, C i D, a què volem mesurar el control de qualitat en la fabricació de les peces enviades, volem comparar si dos processos de producció són equivalents i si mantenen en el seu procés els mateixos estàndards de qualitat. Considerarem que les variàncies de les dues poblacions són iguals. Per a fer la comparació tenim les següents dades de mostres de cada empresa:

Empresa C: 10, 9, 7, 6, 12, 3, 7, 9, 10, 6.

Empresa D: 12, 8, 5, 11, 9, 10, 13, 7, 12, 9, 8, 13.

Amb aquesta informació volem estimar, mitjançant un interval de confiança, si la qualitat mitjana és la mateixa en els dos processos de fabricació. Treballarem amb un nivell de confiança del 95 %.

Finalment, volem estudiar si la variabilitat d'aquestes dues empreses és la mateixa.

Solució

- a) Per al cas en què coneixem les variàncies, estimarem amb un interval al 99 % de nivell de confiança quina empresa té una mitjana inferior utilitzant la fórmula que hem presentat anteriorment:

$$\mu_x - \mu_y \in \left(\bar{x} - \bar{y} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}, \bar{x} - \bar{y} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \right)$$

Per a la població A podem anotar el paràmetre de la població $\sigma_x^2 = 25^2 = 625$, i amb les dades de la mostra podem dir que $n = 10$ i que $\bar{x} = 80$.

Per a la població B podem anotar el paràmetre de la població $\sigma_y^2 = 30^2 = 900$, i amb les dades de la mostra podem dir que $m = 15$ i que $\bar{y} = 75$.

Com que treballem amb un nivell de confiança del 99 %, podem calcular $1 - \alpha = 0,99$ i calcularem el valor de $1 - \frac{\alpha}{2}$, ja que:

$$1 - \alpha = 0,99 \rightarrow \alpha = 0,01 \rightarrow \frac{\alpha}{2} = 0,005 \rightarrow 1 - \frac{\alpha}{2} = 0,995$$

En la taula de la distribució acumulada de la variable normal tipificada busquem el valor de la variable x , tal que $P(Z \leq x) = 0,995$. Aquest valor és $x = 2,57583$, per la qual cosa, utilitzant la notació en aquest tema, $z_{1-\frac{\alpha}{2}} = z_{0,995} = 2,57583$

Substituïm aquestes dades en la fórmula i obtenim:

$$\mu_x - \mu_y \in \left(\bar{x} - \bar{y} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}, \bar{x} - \bar{y} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \right) =$$

$$\left(80 - 75 - 2,576 \sqrt{\frac{625}{10} + \frac{900}{15}}, 80 - 75 + 2,576 \sqrt{\frac{625}{10} + \frac{900}{15}} \right) = (-23,511, 33,511)$$

En aquest cas, el 0 pertany a l'interval, per la qual cosa interpretarem que la diferència entre els paràmetres no és significativa, i podem estimar que $\mu_x \approx \mu_y$.

- b) En aquesta segona part, per a dur a terme els càlculs, identificarem els valors dels paràmetres de cadascuna de les mostres, ja que, com diu l'enunciat, suposarem que $\sigma_x^2 = \sigma_y^2$.

De la mostra de la població X anotarem $n = 10$, $\bar{x} = 7,9$, $S_{x_{n-1}}^2 = 6,77$.

De la mostra de la població Y anotarem $m = 12$, $\bar{y} = 9,75$, $S_{y_{n-1}}^2 = 6,39$.

En les taules podem trobar la dada que ens falta: $t_{1-\frac{\alpha}{2}, n+m-2}$ i, com que treballem amb un nivell de confiança del 95 %, tenim que $1 - \alpha = 0,95$ i a partir d'això el valor de $1 - \frac{\alpha}{2}$ és:

$$1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow \frac{\alpha}{2} = 0,025 \rightarrow 1 - \frac{\alpha}{2} = 0,975$$

Així, cal calcular, amb les taules de la distribució acumulada de la variable t de Student amb 20 graus de llibertat, el valor de la variable ${}_{1-\frac{\alpha}{2}}t_{n+m-2} = {}_{0,975}t_{20} = 2,086$ i substituïrem aquests valors en la fórmula:

$$\begin{aligned} \mu_x - \mu_y \in & \left(\bar{x} - \bar{y} - {}_{1-\frac{\alpha}{2}}t_{n+m-2} \sqrt{\frac{(n-1)S_{x_{n-1}}^2 + (m-1)S_{y_{m-1}}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}, \right. \\ & \left. \bar{x} - \bar{y} + {}_{1-\frac{\alpha}{2}}t_{n+m-2} \sqrt{\frac{(n-1)S_{x_{n-1}}^2 + (m-1)S_{y_{m-1}}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}} \right) = \\ & \left(7,9 - 9,75 - 2,086 \sqrt{\frac{9 \cdot 6,77 + 11 \cdot 6,39}{20}} \sqrt{\frac{1}{10} + \frac{1}{12}}, \right. \\ & \left. 7,9 - 9,75 + 2,086 \sqrt{\frac{9 \cdot 6,77 + 11 \cdot 6,39}{20}} \sqrt{\frac{1}{10} + \frac{1}{12}} \right) = \\ & = (-4,138, 0,438) \end{aligned}$$

Podem concloure que, com que el 0 està dintre d'aquest interval, la diferència entre les mitjanes no és significativa i que les dues línies de producció tenen un nivell de qualitat semblant, la qual cosa podem afirmar amb un nivell de confiança del 95 %.

- c) Treballarem amb les dades de l'apartat anterior y amb un 95 % de nivell de confiança.

Cal comentar que si utilitzem taules cal trobar el primer valor en funció del segon, el qual podem trobar a les taules que solen publicar-se, aplicant la propietat de la funció de distribució de la variable F de Fisher-Snedecor que expliquem.

Així, amb les taules o amb un programa estadístic, trobem el valor del percentil ${}_{1-\frac{\alpha}{2}}f_{(n-1),(m-1)} = {}_{0,975}f_{9,11} = 3,5879$ i per a calcular l'altre valor farem ús de

la propietat següent: $\frac{\alpha}{2}f_{(n-1),(m-1)} = \frac{1}{{}_{1-\frac{\alpha}{2}}f_{(m-1),(n-1)}}$ que aplicada al nostre cas

$$\text{seria } {}_{0,025}f_{9,11} = \frac{1}{{}_{0,975}f_{11,9}} = \frac{1}{3,91207} = 0,255619.$$

Si substituïm tots aquests valors a l'expressió de l'interval, obtindrem com a resultat:

$$\frac{\sigma_X}{\sigma_Y} \in \left(\frac{\frac{(n-1)S_{X_{n-1}}^2}{n} \cdot \frac{1}{1 - \frac{\alpha}{2} f_{(n-1), (m-1)}}}{\frac{(m-1)S_{Y_{n-1}}^2}{m}}, \frac{\frac{(n-1)S_{X_{n-1}}^2}{n} \cdot \frac{1}{\frac{\alpha}{2} f_{(n-1), (m-1)}}}{\frac{(m-1)S_{Y_{n-1}}^2}{m}} \right) =$$

$$\left(\frac{\frac{9}{10} 6,77}{\frac{11}{12} 6,39} \cdot \frac{1}{3,5879}, \frac{\frac{9}{10} 6,77}{\frac{11}{12} 6,39} \cdot \frac{1}{0,255619} \right) = (0,2899, 4,0694)$$

Considerem que l'interval solució és $\frac{\sigma_X^2}{\sigma_Y^2} \in (a, b)$. Si $1 \in (a, b)$, interpretarem que la diferència entre els paràmetres no és significativa i podrem estimar que $\sigma_X^2 \approx \sigma_Y^2$

Exercici 8

Es pot estudiar l'eficàcia de dues impressores pel número de fulls amb defectes després d'imprimir? Per a comprovar-ho, hem seleccionat aleatòriament una mostra de 200 treballs impresos amb la impressora A, dels quals 15 eren defectuosos, i 250 treballs de la impressora B, dels quals n'han eixit 16 de defectuosos.

Solució

Si volem treballar amb un nivell de confiança del 95 %, calcularem l'interval anterior amb les dades del nostre problema. En primer lloc, calculem les dades que necessitem de la mostra de la impressora A:

$$n = 200 \quad p_X = \frac{15}{200} = 0,075$$

i les dades de la mostra de la impressora B:

$$m = 250 \quad p_Y = \frac{16}{250} = 0,064$$

Com que treballarem amb un nivell de confiança del 95 %, podem calcular el valor

de $1 - \frac{\alpha}{2}$, ja que:

$$1 - \alpha = 0,95 \quad \rightarrow \quad \alpha = 0,05 \quad \rightarrow \quad \frac{\alpha}{2} = 0,025 \quad \rightarrow \quad 1 - \frac{\alpha}{2} = 0,975$$

En la taula de la distribució acumulada de la variable normal tipificada busquem el valor de la variable x , tal que $P(Z \leq x) = 0,975$. Aquest valor és $x = 1,96$, denotem

per $z_{1-\frac{\alpha}{2}} = 1,96$. Substituïm aquestes dades a la fórmula de l'interval i obtenim:

$$p_x - p_y \in \left(p_x - p_y - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}}, p_x - p_y + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}} \right) =$$

$$\left(0,075 - 0,064 - 1,96 \sqrt{\frac{0,075 \cdot 0,925}{200} + \frac{0,064 \cdot 0,936}{250}}, 0,075 - 0,064 + 1,96 \sqrt{\frac{0,075 \cdot 0,925}{200} + \frac{0,064 \cdot 0,936}{250}} \right) =$$

$$(-0,036, 0,058)$$

La conclusió que podem extraure d'aquest interval és que, si treballem amb un nivell de confiança del 95 %, la diferència entre les proporcions mostrals que hem obtingut no és significativa; podem estimar que la proporció de peces defectuoses en la producció de les dues màquines és la mateixa. Aquesta afirmació la basem en el fet que el valor 0 està dintre de l'interval, ja que, per tractar-se d'una diferència, té les mateixes conclusions que els casos que hem treballat abans de la diferència de mitjanes.

Exercici 9

Estudiem l'errada en la solució d'un problema en què treballem la inferència sobre la mitjana poblacional. En la classe d'estadística tenim 25 alumnes amb una nota mitjana de 7 en el curs anterior amb una desviació típica 3 (agafem aquesta dada com si fóra de tota la població d'estudi).

- Volem l'interval de confiança d'aquestes dades al 95 %.
- A més a més, si l'amplitud de l'interval és molt gran (major d'un punt), volem saber el nombre d'estudiants per a no tindre tanta variació en les dades.

Solució

- Primer calculem l'interval de la mitjana per a les dades. Com que tenim 25 alumnes i coneixem la mitjana mostral i la desviació típica poblacional al 95 %:

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = \left(7 - 1,96 \frac{3}{\sqrt{25}}, 7 + 1,96 \frac{3}{\sqrt{25}} \right) = (5,824, 8,176)$$

Podem concloure que, amb una probabilitat del 95 %, la mitjana de la nota dels alumnes d'estadística està dintre d'aquest interval.

b) Donat que l'amplitud de l'interval és $8,176 - 5,824 = 2,352$ hores, l'errada en la nostra estimació és de $E = \frac{b-a}{2} = \frac{2,352}{2} = 1,176$.

Si nosaltres volem acotar aquesta errada i que siga de 0,5 hores, podem esbrinar *a priori* la grandària de la mostra amb la que hem de treballar i així coneixerem el nombre d'individus a què cal fer l'enquesta. Així:

$$E = \frac{b-a}{2} = \frac{\left(\bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) - \left(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)}{2} = \frac{2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}{2} = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1,96 \frac{3}{\sqrt{n}} = 0,5$$

D'on podem aïllar el valor de n , $\sqrt{n} = \frac{1,96 \cdot 3}{0,5} = 11,76 \rightarrow n = 11,76^2 = 138,30$,

és a dir, amb 139 alumnes podem extraure l'interval amb l'errada que ens havíem plantejat.

Exercici 10

Per als treballadors d'una empresa que tenia problemes de rendiment es va proposar una millora que consistia a deixar més temps de descans en les parades. Per a avaluar l'eficàcia d'aquest plantejament, s'ha observat l'índex de millora plantejat per la gestió de l'empresa a 8 treballadors d'una secció. Aquest índex té una escala de 0 a 3. Els resultats es detallen a continuació:

Abans: 1, 2; 1, 3; 1, 5; 1, 4; 1, 7; 1, 9; 1, 4; 1, 2.

Després: 1, 4; 1, 7; 1, 5; 1, 3; 2, 0; 2, 1; 1, 7; 1, 6.

Calculem un interval de confiança al 95 % per vore si ha donat resultat la mesura adoptada i extraure'n conclusions.

Solució

Com que són dades relacionades, necessitarem calcular els valors de la variable $D = X - Y$ restant els valors emparellats, tal com es veu a la taula següent:

X = valors abans de la mesura	1,2	1,3	1,5	1,4	1,7	1,9	1,4	1,2
Y = valors després de la mesura	1,4	1,7	1,5	1,3	2	2,1	1,7	1,6
D = X - Y	-0,2	-0,4	0	0,1	-0,3	-0,2	-0,3	-0,4

Amb la calculadora esbrinem els valors següents de les mostres de les variables X, Y: $n = 8$, $\bar{x} = 1,45$, $\bar{y} = 1,6625$; amb les dades de la variable D calculem $S_{D_{n-1}}^2 = 0,0327$.

Com que treballem amb un nivell de confiança del 95 %, llavors:

$$1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow \frac{\alpha}{2} = 0,025 \rightarrow 1 - \frac{\alpha}{2} = 0,975$$

Així, cal calcular en les taules de la distribució acumulada de la variable t de Student ${}_{1-\frac{\alpha}{2}}t_{n-1} = {}_{0,975}t_7 = 2,3646$. Amb la resta de les dades conegudes substituïm aquests valors i s'obté l'interval:

$$\begin{aligned} \mu_x - \mu_y \in & \left(\bar{x} - \bar{y} - {}_{1-\frac{\alpha}{2}}t_{n-1} \sqrt{\frac{S_{D_{n-1}}^2}{n}}, \bar{x} - \bar{y} + {}_{1-\frac{\alpha}{2}}t_{n-1} \sqrt{\frac{S_{D_{n-1}}^2}{n}} \right) = \\ & \left(1,45 - 1,6625 - 2,3646 \sqrt{\frac{0,0327}{8}}, 1,45 - 1,6625 + 2,3646 \sqrt{\frac{0,0327}{8}} \right) = (-0,3637, -0,0613) \end{aligned}$$

Com que els dos valors dels extrems de l'interval són negatius, podem inferir que $\mu_x - \mu_y < 0 \rightarrow \mu_x < \mu_y$, per la qual cosa interpretarem que la mitjana de X és menor que Y de les enquestes de satisfacció és més gran en les que han estat realitzades després de la mesura establerta per a satisfer els clients. Podem estimar que la mesura d'eliminar comissions sí que ha aconseguit l'objectiu per al qual havia estat dissenyada.

UNITAT 5

Introducció a la inferència 2: contrast d'hipòtesis

Introducció teòrica

Es pot definir el Contrast d'Hipòtesis com la part de la inferència estadística que té com a objectiu comprovar, mitjançant mètodes matemàtics, hipòtesis realitzades sobre el valor d'algun paràmetre d'una o diverses poblacions.

Comencem plantejant-nos algunes situacions:

- Un fabricant de piles afirma que la durada mitjana de les seues piles és com a mínim de 53 hores i la seua desviació típica és de 4 hores. Rebutjaríem aquesta afirmació en el cas de que una pila durara 48 hores? I si la mitjana de la durada de 100 piles fóra de 50 hores? I si aquesta mitjana fóra de 56 hores?
- Tenim un dau que suposem correcte. El llancem 100 vegades i obtenim 25 vegades el número 5. Podem dir que el dau es correcte? O cal verificar que no està trucat en funció dels resultats?
- Suposem dues empreses que produeixen les piles anteriors. Al llarg dels anys, aquestes piles han tingut una durada similar, però en l'actualitat una segona empresa afirma que les seues piles duren més perquè ha fet una millora en la producció. Podem creure la seua afirmació?

En la pràctica és freqüent trobar-nos davant situacions com aquestes en les quals es precis prendre decisions sobre hipòtesis estadístiques senzilles relatives als paràmetres d'una població o de dues. És per això que aquesta branca de la inferència també s'anomena *teoria de la decisió*.

És clar que en prendre una decisió estadística mai estarem completament segurs d'encertar. Al llarg del tema vorem com prendre aquesta decisió i com minimitzar el risc d'enganyar-nos.

Un *test estadístic* és un procediment per a extraure, a partir d'una mostra aleatòria i representativa, conclusions que permeten acceptar o rebutjar una hipòtesi prèviament emesa sobre el valor d'un paràmetre desconegut d'una població.

Si volem conèixer la veracitat d'una informació, que considerarem que és una hipòtesi sobre aquesta, la contrastarem amb la informació que traurem d'una mostra. Si ambdues informacions coincideixen dins d'un marge que creiem admissible, mantindrem que la hipòtesi inicial era certa; en cas contrari, la rebutjarem i plantejarem noves hipòtesis que expliquen les dades observades.

Bàsicament, podríem comparar aquests mètodes amb un judici en què, en principi, se suposa la innocència del acusat (hipòtesi nul·la) i es tracta d'aportar proves per a rebutjar aquesta hipòtesi. La innocència serà rebatuda quan les proves ho demostrin amb un alt grau de fiabilitat.

Hipòtesi nul·la i hipòtesi alternativa

Començarem plantejant-nos les dues primeres situacions que hem presentat en començar el tema. En ambdós exemples hi ha una *hipòtesi de partida* i uns resultats obtinguts a partir de les dades d'una mostra que difereixen de la hipòtesi. Ens preguntem si aquesta diferència és atribuïble a l'atzar o si estem davant d'una hipòtesi de partida que era falsa.

Al primer exemple és evident que la primera qüestió estaria fora de lloc. No ens semblarà raonable que el fet que un element contradiga una generalització ens porte a pensar que estem davant d'una informació falsa. Sempre refutarem la hipòtesi inicial sobre la població obtenint dades reals d'una mostra representativa d'aquesta. Analtzarem la segona qüestió, on tenim una mostra de 100 piles que tenen una durada mitjana de 50 hores.

	Cas 1: piles	Cas 2: dau
Hipòtesi	El fabricant afirma $\mu = 53, \sigma = 4$	Si el dau és correcte, la proporció de cincs és $p = 1/6 = 0,167$
Resultats de la mostra	$\bar{X} = 50$	$p = 0,25$
Interrogant	La diferència observada pot atribuir-se a l'atzar? Podem suposar raonablement que la mostra ha estat extreta de la població sobre la qual hem fet la hipòtesi?	

Als dos exemples hem considerat una hipòtesi que en principi admetem com a vàlida i que desitgem contrastar. S'anomena *hipòtesi nul·la* i es designa per H_0 i cal definir la hipòtesi contrària que anomenem *hipòtesi alternativa* i que es designa per H_1 , la qual serà admesa quan el resultat del contrast siga rebutjar H_0 .

	Cas 1: piles	Cas 2: dau
Hipòtesis	$\begin{cases} H_0: \mu = 53 \\ H_1: \mu \neq 53 \end{cases}$	$\begin{cases} H_0: p = 0,167 \\ H_1: p \neq 0,167 \end{cases}$

En aquest tema abordarem contrastos paramètrics en què la hipòtesi nul·la tracta el valor d'un paràmetre de la població. Aquests contrastos poden tenir dos tipus d'hipòtesis:

- Hipòtesis simples: especifiquen un únic valor del paràmetre. Exemple: $\theta = \theta_0$.

- Hipòtesis compostes: especifiquen un interval de valors. Exemple: $\theta < \theta_0$.

Nosaltres realitzarem contrastos en els quals la hipòtesi nul·la serà simple i l'alternativa serà composta.

En principi, podem plantejar dos tipus d'hipòtesis alternatives:

- *Contrast bilateral* $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$ que plantejarem en general si no tenim informació de quina de les dues possibilitats portarà a que H_0 siga falsa.
- *Contrast unilateral* $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$ per la dreta, o bé $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$ per l'esquerra que plantejarem si sabem en quina possibilitat està la probabilitat major de que la hipòtesi nul·la siga falsa o si no ens interessa que siga falsa en l'altra direcció (si volem comprovar que un tractament millora un cert indicador de malaltia, donem per fet que és evident que sabem que no l'empitjora).

Tipus d'errada

En prendre qualsevol decisió, com a conseqüència de l'aplicació d'un test estadístic, pot ocórrer que:

1. Acceptem H_0 i aquesta és vertadera.
2. Acceptem H_0 i aquesta és falsa.
3. Rebutgem H_0 i aquesta és vertadera.
4. Rebutgem H_0 i aquesta és falsa.

H_0	acceptem	rebutgem
vertadera	No cometem errada	Errada de tipus I
falsa	Errada de tipus II	No cometem errada

En els casos 1 i 4 estem prenent una decisió encertada. En els casos 2 i 3 estem prenent una decisió equivocada, és a dir, cometem una errada.

Quan rebutgem H_0 i aquesta és veritat, direm que cometem una errada de tipus I.
Quan acceptem H_0 i aquesta és falsa, direm que cometem una errada de tipus II.

La probabilitat de cometre una errada de tipus I es denota per α i es denomina *nivell de significació* del test d'hipòtesis. Aquest nivell de significació és una probabilitat menuda (0,1; 0,05; 0,01) que habitualment tria l'investigador.

La probabilitat de cometre una errada de tipus II es denota per β . El número $1 - \beta$ es diu *potència* del test d'hipòtesis. Aquesta disminueix quan augmenta la grandària de la mostra.

Un bon test cal dissenyar-lo minimitzant les probabilitats de cometre errades. Però no és fàcil reduir els dos tipus d'errades simultàniament. En general, quan disminueix un tipus d'errada, augmenta l'altre, si no és que augmentem la grandària de la mostra.

Per a tindre un bon test d'hipòtesi plantejarem l'elecció adient de la hipòtesi nul·la, de tal manera que l'errada de tipus I siga la de «pitjors conseqüències». Per exemple, si estem estudiant si un nou fàrmac produeix o no efectes secundaris en els pacients, seria correcte plantejar-ho així:

H_0 : El fàrmac produeix efectes secundaris.

H_1 : El fàrmac no produeix efectes secundaris.

Això ho fem perquè és més greu concloure que el fàrmac no produeix efectes secundaris, si és que els produeix (errada de tipus I), que decidir que els produeix i no ser veritat (errada de tipus II).

Disseny d'un test d'hipòtesis

Desenvoluparem el procés complet per a realitzar un test d'hipòtesis. En detallarem cadascuna de les fases i utilitzarem l'exemple 1 per a anar concretant cada concepte.

1. Identificació de les hipòtesis i les dades

Aquest apartat ja està suficientment desenvolupat a l'apartat anterior.

En el cas de les piles volem comprovar que la mitjana de la duració de les piles és de 53 hores. També coneixem que $\sigma = 4$. Per a contrastar aquesta afirmació triarem una mostra de 100 piles i obtindre el valor de la mitjana mostral $\bar{X} = 50$.

Així, plantejarem les hipòtesis: $\begin{cases} H_0: \mu = 53 \\ H_1: \mu \neq 53 \end{cases}$. Podem concloure doncs, que farem un contrast bilateral sobre el valor de la mitjana.

2. Triar un estadístic de contrast T

Aquest és un estimador del paràmetre poblacional del qual volem refutar-ne el valor, i del qual coneixem la distribució. Suposem que es tracta d'estimadors no esbiaixats respecte del paràmetre del qual es planteja la hipòtesi nul·la.

En cadascun dels casos que anirem plantejant en els apartats següents, presentarem l'estadístic adient i la seua distribució (ja explicada als apartats corresponents del tema 8). Abordarem un exemple de cada cas com als temes anteriors.

A l'exemple 1 de les piles utilitzarem com a estadístic de contrast la variable $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$, de la qual coneixem que es distribueix seguint el model d'una normal tipificada.

3. Determinació de la regió crítica o de rebuig i de la regió d'acceptació

Si la hipòtesi nul·la és certa, l'estadístic de contrast és una variable de la qual en coneixem la distribució. Ens plantejarem esbrinar quins seran els valors més probables d'aquest i quins seran molt improbables.

Busquem un interval, dintre del qual serà molt probable que queden inclosos la majoria dels valors possibles de l'estadístic, que anomenarem regió d'acceptació. Se l'anomena així perquè, si el valor de l'estadístic mostrat està dintre d'ell, podem confirmar que la hipòtesi nul·la és certa. La informació extreta de la mostra no contradueix l'afirmació inicial que donàvem per vàlida respecte del valor del paràmetre poblacional.

Pel contrari, si aquest valor de l'estadístic calculat amb les dades de la mostra cau fora de l'interval de la regió d'acceptació, és a dir, en la regió crítica o de rebuig, rebutjarem la hipòtesi nul·la. Aleshores podem plantejar-nos una disjuntiva:

1. El valor trobat és molt improbable, però gràcies a l'atzar la nostra mostra aleatòria ha donat un valor molt allunyat dels valors més probables.
2. El valor extret de la realitat està molt lluny dels valors més probables de la distribució (que podem calcular perquè la coneixem), per la qual cosa podem concloure que la distribució que havíem donat com a vàlida no és correcta.

A tal fi, definirem el nivell de significació α , el qual és la probabilitat que triem per a definir l'amplitud de la regió d'acceptació, ja que considerarem que la probabilitat de que un estadístic tinga un valor dintre d'aquest interval és de $1 - \alpha$. Aquesta elecció de α també ens defineix la regió de rebuig R , ja que considerarem que $P(T \in R / \theta = \theta_0) = \alpha$. És evident que ha de ser menuda; els valors més habituals són 0,10, 0,05 i 0,01.

Podem comentar que el nivell de significació és la probabilitat de cometre una errada de tipus I, és a dir, de rebutjar la hipòtesi nul·la quan aquesta és certa, per tractar-se de la probabilitat que hem adjudicat a la situació α plantejada abans en la disjuntiva.

A l'exemple de les piles, la hipòtesis nul·la estava plantejada sota els supòsits que la mitjana poblacional de la duració de les piles és de 53 hores i que la seua desviació típica és de 4 hores. Ara bé, la mitjana de la mostra de 100 piles és de 50 hores.

Amb aquestes dades calcularem la zona d'acceptació i de rebuig que assignarem al nostre contrast amb l'expressió de l'estadístic $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ el qual es distribueix com

una variable normal tipificada. Treballarem amb un nivell de significació $\alpha = 0,05$.

Calcularem els extrems de l'interval (a, b) de la regió d'acceptació, considerant que la variable Z és simètrica i que hem plantejat un contrast bilateral.

$$P(a \leq Z \leq b) = 0,95 \rightarrow P(Z \leq b) = 0,975$$

Per a trobar $b = Z_{0,975} = 1,96$ utilitzarem taules o programes estadístics.

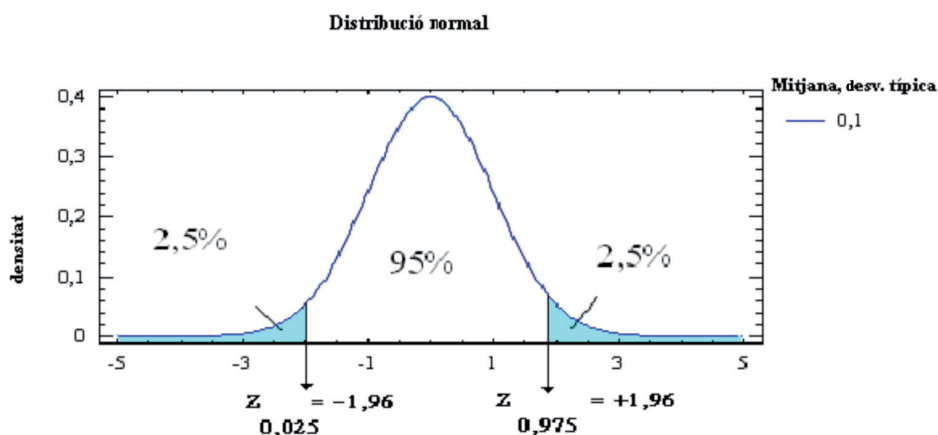


Fig. 26. Contrast bilateral

Així, la regió d'acceptació serà l'interval $(-1,96; +1,96)$ i la resta de valors reals seran la regió crítica, es pot observar en el gràfic de la figura 26.

En general, quan es tracta d'un contrast bilateral, la regió de rebuig està repartida en dos intervals, a l'esquerra i a la dreta de l'interval que defineix la zona central determinada per la regió d'acceptació.

En el contrast unilateral, la regió d'acceptació queda a un lateral i la regió de rebuig és un únic interval que queda a la seua dreta o esquerra, segons el cas. És per aquesta raó que es diu contrast unilateral o d'una cua, front al bilateral o de dues cues.

4. Càlcul del valor de l'estadístic i elaboració de les conclusions

Amb les dades dels paràmetres que donem per veritables i amb els que hem definit la hipòtesi nul·la, i amb els valors mostrals extrets de la mostra representativa triada adequadament, calculem el valor de l'estadístic en cada cas.

Al nostre exemple, l'estadístic val
$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{50 - 53}{\frac{4}{\sqrt{100}}} = \frac{-3}{0,4} = -7,5.$$

Si el valor de l'estadístic «cau» dins de la regió d'acceptació, acceptem la hipòtesi nul·la i la considerem veritat. Si, pel contrari, cau dins de la regió crítica, rebutjarem la hipòtesi nul·la i la suposarem falsa.

En el nostre exemple, és evident que el valor de l'estadístic està fora de la regió d'acceptació i, en aquest cas, molt allunyat. Conseqüentment, rebutjarem la hipòtesi nul·la que afirmava que la duració mitjana de les piles és de 53 hores.

Contrastos d'hipòtesis per als paràmetres d'una població

Per a construir aquests tests cal utilitzar les distribucions dels estadístics que vam veure al tema 8. Seguirem la mateixa notació i necessitarem les mateixes condicions per a aplicar-la que les estudiades en tots els apartats d'estimació.

Considerarem X_1, X_2, \dots, X_n una mostra aleatòria d'una població que es distribueix amb qualsevol model de distribució amb mitjana μ i variància σ^2 .

Explicarem el procés de cada cas amb el desenvolupament d'un exemple. En cada apartat proposarem els tres tipus de contrast: bilateral, unilateral per la dreta i unilateral per l'esquerra. Abordarem les tres possibilitats, desenvolupant les tres opcions en els diferents exemples, de manera que quedaran prou aclarides les semblances i diferències entre elles.

Com que la inferència mitjançant l'estimació per intervals i el contrast d'hipòtesis estan molts relacionats, en cada cas plantejarem els mateixos exemples del tema anterior, però donant-los un plantejament adient a la teoria de la decisió.

Contrast d'hipòtesis per a mitjanes

Contrast d'hipòtesis per a la mitjana mostral (coneguda σ^2 poblacional)

1. Estadístic: $T = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$, que es distribueix $T \rightarrow N(0, 1)$.
2. Bilateral:
$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \quad R = \{T : |T| \geq z_{1-\frac{\alpha}{2}}\}$$
3. Unilateral per la dreta:
$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \quad R = \{T : T \geq z_{1-\alpha}\}$$
4. Unilateral per l'esquerra:
$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \quad R = \{T : T \leq -z_{1-\alpha}\}$$

Contrast d'hipòtesis per a la mitjana mostral (si no es coneix la σ^2 poblacional)

Si l'estadístic és $T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$, podem afirmar que $T \rightarrow t_{n-1}$, és a dir, que la variable T té una distribució t de Student amb $n - 1$ graus de llibertat . Recordem que S fa referència a la quasi-variància mostral.

1. Bilateral
$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases} \quad R = \{T : |T| \geq t_{1-\frac{\alpha}{2}, n-1}\}$$
2. Unilateral per la dreta
$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \quad R = \{T : T \geq t_{1-\alpha, n-1}\}$$
3. Unilateral per l'esquerra
$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \quad R = \{T : T \leq -t_{1-\alpha, n-1}\}$$

Altres contrastos

Contrast d'hipòtesis per a la proporció d'una població Bernouilli

Donat l'estadístic $T = \frac{p - p_0}{\sqrt{\frac{pq}{n}}}$, que es distribueix $T \rightarrow N(0, 1)$, denotarem per p la proporció en la població i per p la proporció en la mostra de grandària n .

1. Bilateral $\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases} \quad R = \{T : |T| \geq z_{1-\frac{\alpha}{2}}\}$
2. Unilateral per la dreta $\begin{cases} H_0: p = p_0 \\ H_1: p > p_0 \end{cases} \quad R = \{T : T \geq z_{1-\alpha}\}$
3. Unilateral per l'esquerra $\begin{cases} H_0: p = p_0 \\ H_1: p < p_0 \end{cases} \quad R = \{T : T \leq -z_{1-\alpha}\}$

Contrast d'hipòtesis per a la variància d'una població normal

Considerem l'estadístic $T = \frac{(n-1)S^2}{\sigma^2}$. Si denotem per $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ la quasi-variància d'una mostra aleatòria de grandària n , extreta d'una població normal de variància σ^2 , definirem com a estadístic la variable T com

$T = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$, que es distribueix com una variable Xi-quadrat amb $n - 1$ graus de llibertat, és a dir, $T \rightarrow \chi_{n-1}^2$.

1. Bilateral $\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases} \quad R = \left\{T : T \leq \frac{\alpha}{2} \chi_{n-1}^2 \text{ o bé } T \geq \frac{1-\alpha}{2} \chi_{n-1}^2\right\}$
2. Unilateral per la dreta $\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 > \sigma_0^2 \end{cases} \quad R = \{T : T \geq (1-\alpha) \chi_{n-1}^2\}$
3. Unilateral per l'esquerra $\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 < \sigma_0^2 \end{cases} \quad R = \{T : T \geq \alpha \chi_{n-1}^2\}$

Contrast d'hipòtesis per als paràmetres de dues poblacions

A l'apartat anterior hem vist els contrastos d'hipòtesis dels paràmetres d'una població per a decidir si aquests tenen uns valors determinats *a priori*.

En aquest apartat farem els contrastos sobre els paràmetres de dues poblacions. Per a refutar o confirmar la hipòtesi de partida necessitarem les dades de dues mostres diferents extretes cadascuna d'elles de les respectives poblacions, ja que volem inferir conclusions respecte dels valors dels seus paràmetres poblacionals.

Recordem que en aquests casos no podrem decidir sobre els valors respectius dels paràmetres de les poblacions, sinó que podrem comparar-los per diferència (en el cas de mitjanes i proporcions) o per quocient (en el cas de les variàncies) i concloure quin dels dos és major i en quina mesura.

Podem recordar l'exemple de la tercera situació presentada a la introducció d'aquest tema: «suposem dues empreses que produeixen piles. Al llarg dels anys, aquestes piles han tingut una durada similar, però en l'actualitat una segona empresa afirma que les seues piles duren més perquè ha fet una millora en la producció. Podem creure la seua afirmació?».

Amb el contrast d'hipòtesis podem fer-nos plantejaments respecte de les mitjanes de la duració de les piles en cada empresa com ara:

- Duraran el mateix nombre d'hores les piles de les dues empreses? És a dir, són iguals les mitjanes de A i B?
- Duraran les piles de l'empresa B més que les de A tal com es presenta? És a dir, la mitjana de B és major que la de A?
- Duraran les piles de l'empresa B 5 hores més que les de l'empresa A? És a dir, la diferència de les mitjanes de A i B és de 5 hores?

Per a respondre a aquestes qüestions caldrà triar una mostra representativa de cadascuna de les produccions de les empreses A i B de certa grandària n i m , les quals, en principi, poden ser diferents. Calcularem les mesures que necessitem en cadascuna d'elles i analitzarem si aquestes mesures ens fan dubtar o no, en termes de probabilitat, de la versemblança de les nostres hipòtesis inicials.

Recordem la notació: X_1, X_2, \dots, X_n és una mostra aleatòria de grandària n extreta d'una població i Y_1, Y_2, \dots, Y_m la mostra aleatòria de grandària m extreta d'una altra població. Mantenim el paral·lelisme amb els valors de les mitjanes μ_x, μ_y per a cadascuna de les poblacions i els símbols σ_x^2, σ_y^2 per a denotar-ne les variàncies.

Els valors de les mitjanes mostrals són \bar{X} i \bar{Y} . Les quasi-variàncies mostrals són $S_{X_{n-1}}^2$ i $S_{Y_{m-1}}^2$.

En el cas d'una població Bernouilli, denotem per p_X, p_Y les proporcions obtingudes a partir dels valors de les mostres i per p_x, p_y les proporcions poblacionals a estimar de les poblacions respectives. Quant a les notacions de les variables i les expressions dels percentils, són les que ja hem utilitzat als temes anteriors.

La resta d'aspectes del disseny d'un test són iguals que els tractats a l'apartat anterior, tot i que anirem comentant-los mitjançant l'exemple de cada cas.

Contrast d'hipòtesis per a la diferència de mitjanes

Totes les consideracions que hem fet a la introducció de la comparació de mitjanes en l'estimació per intervals al tema anterior, pel que fa a la possibilitat de treballar amb mostres independents o emparellades, es tornaria a plantejar en aquest apartat. Per la qual cosa, també ací presentem els tres casos que hem desenvolupant en els tres temes d'inferència.

- Diferència de mitjanes amb mostres independents i variàncies poblacionals conegudes.
- Diferència de mitjanes amb mostres independents i variàncies poblacionals desconegudes, però que podem suposar iguals.
- Diferència de mitjanes amb mostres dependents i/o emparellades.

És interessant que vegem com plantejaríem les hipòtesis nul·la i alternativa en els casos de la introducció.

- Duraran el mateix nombre d'hores les piles de les dues empreses? És a dir, són iguals les mitjanes de A i B?

$$\begin{cases} H_0: \mu_X - \mu_Y = 0 \\ H_1: \mu_X - \mu_Y \neq 0 \end{cases}$$

- Duraran les piles de l'empresa B més que les de A tal com es presenta? És a dir, la mitjana de B és major que la de A?

$$\begin{cases} H_0: \mu_X - \mu_Y = 0 \\ H_1: \mu_X - \mu_Y < 0 \end{cases}$$

Duraran les piles de l'empresa B 5 hores més que les de l'empresa A? És a dir, la diferència de les mitjanes de A i B és de 5 hores?

$$\begin{cases} H_0: \mu_X - \mu_Y = -5 \\ H_1: \mu_X - \mu_Y \neq -5 \end{cases}$$

Contrast d'hipòtesis per a la diferència de mitjanes de poblacions independents amb variàncies poblacionals conegudes

Considerem l'estadístic $T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$ que es distribueix $T \rightarrow N(0, 1)$.

1. Bilateral $\left\{ \begin{array}{l} H_0: \mu_X - \mu_Y = d_0 \\ H_1: \mu_X - \mu_Y \neq d_0 \end{array} \right. \quad R = \left\{ T : |T| \geq z_{1-\frac{\alpha}{2}} \right\}$
2. Unilateral per la dreta $\left\{ \begin{array}{l} H_0: \mu_X - \mu_Y = d_0 \\ H_1: \mu_X - \mu_Y > d_0 \end{array} \right. \quad R = \left\{ T : T \geq z_{1-\alpha} \right\}$
3. Unilateral per l'esquerra $\left\{ \begin{array}{l} H_0: \mu_X - \mu_Y = d_0 \\ H_1: \mu_X - \mu_Y < d_0 \end{array} \right. \quad R = \left\{ T : T \leq -z_{1-\alpha} \right\}$

Contrast d'hipòtesis per a la diferència de poblacions independents amb variàncies poblacionals desconegudes, però iguals

Considerem l'estadístic en què denotem per $S_{X_{n-1}}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ la quasi-variància de la mostra aleatòria de grandària n extreta de la primera població i per

$S_{Y_{m-1}}^2 = \frac{\sum_{i=1}^m (y_i - \bar{y})^2}{m-1}$ la quasi-variància de la mostra de grandària m de la segona població. Definim la següent variable: $T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{(n-1)S_{X_{n-1}}^2 + (m-1)S_{Y_{m-1}}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}}$, que

és l'estadístic que es distribueix segons el model d'una variable t de Student amb $n + m - 2$ graus de llibertat. És a dir, $T \rightarrow t_{n+m-2}$.

1. Bilateral $\left\{ \begin{array}{l} H_0: \mu_X - \mu_Y = d_0 \\ H_1: \mu_X - \mu_Y \neq d_0 \end{array} \right. \quad R = \left\{ T : |T| \geq z_{1-\frac{\alpha}{2}} t_{n+m-2} \right\}$
2. Unilateral per la dreta $\left\{ \begin{array}{l} H_0: \mu_X - \mu_Y = d_0 \\ H_1: \mu_X - \mu_Y > d_0 \end{array} \right. \quad R = \left\{ T : T \geq z_{1-\alpha} t_{n+m-2} \right\}$
3. Unilateral per l'esquerra $\left\{ \begin{array}{l} H_0: \mu_X - \mu_Y = d_0 \\ H_1: \mu_X - \mu_Y < d_0 \end{array} \right. \quad R = \left\{ T : T \leq -z_{1-\alpha} t_{n+m-2} \right\}$

Aplicarem aquesta fórmula quan no coneguem les variàncies poblacionals σ_x^2 ni σ_y^2 , però que puguem afirmar que són iguals. Si no fos el cas, podem fer un contrast per a estimar la igualtat entre les dues variàncies, com es vorà un poc més endavant en aquest mateix apartat, i, si la interpretació ens permet estimar que són iguals, continuarem calculant aquesta diferència de mitjanes.

Contrast d'hipòtesis per a la diferència de mitjanes de dues poblacions amb mostres relacionades

Aquest contrast l'aplicarem, com ja es va explicar amb els apartats d'estimació dels temes anteriors, quan vulguem comparar la mitjana de dues poblacions que estan relacionades i les mostres, de les quals estiguen formades per parells (x_i, y_i) .

Tenim l'estadístic $T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\frac{S_{D_{n-1}}}{\sqrt{n}}}$, del qual sabem que té una distribució

que segueix el model d'una variable t de Student amb $n - 1$ graus de llibertat. És a dir, $T \rightarrow t_{n-1}$.

Recordem que en aquest cas cal definir prèviament la variable $D = X - Y$ i que denotarem per $S_{D_{n-1}}$ la seua quasi- desviació típica mostral, on $d_i = x_i - y_i$. La grandària de les dues mostres coincidirà necessàriament i serà n .

- | | | |
|------------------------------|---|--|
| 1. Bilateral | $\begin{cases} H_0: \mu_X - \mu_Y = d_0 \\ H_1: \mu_X - \mu_Y \neq d_0 \end{cases}$ | $R = \{T : T \geq t_{n-1, \frac{1-\alpha}{2}}\}$ |
| 2. Unilateral per la dreta | $\begin{cases} H_0: \mu_X - \mu_Y = d_0 \\ H_1: \mu_X - \mu_Y > d_0 \end{cases}$ | $R = \{T : T \geq t_{n-1, 1-\alpha}\}$ |
| 3. Unilateral per l'esquerra | $\begin{cases} H_0: \mu_X - \mu_Y = d_0 \\ H_1: \mu_X - \mu_Y < d_0 \end{cases}$ | $R = \{T : T \leq -t_{n-1, 1-\alpha}\}$ |

Així, cal calcular, amb les taules de la distribució acumulada de la variable t de Student amb 7 graus de llibertat, el valor de la variable $t_{n-1, 1-\alpha} = t_{7, 0,95} = 1,8946$, i després substituïrem aquest valor en la fórmula de la regió crítica R.

$$R = \{T : T \leq -t_{n-1, 1-\alpha}\} = \{T : T \leq -1,8946\} = (-\infty, -1,8946)$$

És evident que $T \in R$. La informació de les mostres contradiu la hipòtesi inicial i ens porta a rebutjar-la. Com a conseqüència, podem dir que la mesura sí que ha reportat una millora en el grau de satisfacció dels clients, tal com ho indicava la nostra hipòtesi alternativa.

Altres contrastos d'hipòtesis

Contrast d'hipòtesi per a la diferència de proporcions de dues poblacions Bernoulli

Tenim l'estadístic $T = \frac{(p_x - p_y) - (p_X - p_Y)}{\sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}}}$ i sabem que T es distribueix com una normal tipificada.

Recordem que denotem per p_X, p_Y les proporcions obtingudes a partir dels valors de les mostres i per ρ_X, ρ_Y les proporcions poblacionals de la diferència de les quals farem el contrast.

- | | | |
|------------------------------|---|---|
| 1. Bilateral | $\begin{cases} H_0: \rho_X - \rho_Y = d_0 \\ H_1: \rho_X - \rho_Y \neq d_0 \end{cases}$ | $R = \{T : T \geq z_{1-\frac{\alpha}{2}}\}$ |
| 2. Unilateral per la dreta | $\begin{cases} H_0: \rho_X - \rho_Y = d_0 \\ H_1: \rho_X - \rho_Y > d_0 \end{cases}$ | $R = \{T : T \geq z_{1-\alpha}\}$ |
| 3. Unilateral per l'esquerra | $\begin{cases} H_0: \rho_X - \rho_Y = d_0 \\ H_1: \rho_X - \rho_Y < d_0 \end{cases}$ | $R = \{T : T \leq -z_{1-\alpha}\}$ |

Contrast d'hipòtesi per al quocient de dues variàncies de dues poblacions normals

Considerem l'estadístic: $T = \frac{\frac{(n-1)S_{X_{n-1}}^2}{\sigma_X^2 \cdot n}}{\frac{(m-1)S_{Y_{m-1}}^2}{\sigma_Y^2 \cdot m}}$, del qual ja sabem que es distribueix

com una variable F de Fisher-Snedecor amb $n - 1$ i $m - 1$ graus de llibertat, respectivament.

- | | | |
|----------------------------|---|--|
| 1. Bilateral | $\begin{cases} H_0: \frac{\sigma_X^2}{\sigma_Y^2} = \sigma_0^2 \\ H_1: \frac{\sigma_X^2}{\sigma_Y^2} \neq \sigma_0^2 \end{cases}$ | $R = \left\{ T : T \leq \frac{\alpha}{2} f_{n-1, m-1} \text{ o bé } T \geq \frac{1-\alpha}{2} f_{n-1, m-1} \right\}$ |
| 2. Unilateral per la dreta | $\begin{cases} H_0: \frac{\sigma_X^2}{\sigma_Y^2} = \sigma_0^2 \\ H_1: \frac{\sigma_X^2}{\sigma_Y^2} > \sigma_0^2 \end{cases}$ | $R = \{T : T \geq f_{1-\alpha, n-1, m-1}\}$ |

$$3. \text{ Unilateral per l'esquerra } \begin{cases} H_0: \frac{\sigma_X^2}{\sigma_Y^2} = \sigma_0^2 \\ H_1: \frac{\sigma_X^2}{\sigma_Y^2} < \sigma_0^2 \end{cases} \quad R = \{T : T \leq \alpha f_{n-1, m-1}\}$$

NOTA

Arribats a aquest punt del desenvolupament del tema, volem remarcar que el contrast bilateral es pot abordar també des del tema d'interval de confiança.

Si cal comprovar que un paràmetre de la població té un cert valor $\theta = \theta_0$ (hipòtesi nul·la) o $\theta \neq \theta_0$ (hipòtesi alternativa) amb un nivell de significació α , serà equivalent a veure si a l'interval de confiança corresponent a aquest paràmetre, tal que $P(\theta \in (a, b)) = 1 - \alpha$, podem observar que $\theta_0 \in (a, b)$.

No es pot dir el mateix amb els contrastos unilaterals per la dreta o esquerra. Si es comparen els exemples dels temes 9 i 10, es pot comprovar que majoritàriament hem utilitzat els mateixos exemples per a afavorir el plantejament paral·lel de les dues tècniques d'inferència. Convidem el lector a llegir-los i analitzar-los comparant-ne la tècnica i els resultats.

P-valor o valor crític

A hores d'ara ja coneixem que el resultat d'un contrast d'hipòtesis depèn del nivell de significació que triem per a realitzar-lo. En algunes situacions el valor de l'estadístic queda prop dels valors frontera de la regió crítica o de rebuig, per la qual cosa resulta interessant conèixer quin seria el nivell de significació que, amb aquest valor de l'estadístic, ens faria rebutjar la hipòtesi nul·la.

Per la qual cosa, és raonable i freqüent adjuntar a la decisió de rebutjar H_0 el que direm p-valor o valor crític.

Definirem p-valor o valor crític al grau mínim de significació amb què es pot rebutjar la hipòtesi nul·la amb el valor de l'estadístic T que hem calculat amb les dades del plantejament del test. És a dir, que si a fóra major que el p-valor, rebutjaríem la hipòtesi de partida i si fos menor, l'acceptaríem.

Cal diferenciar els tres tipus de contrast:

- | | |
|------------------------------|--|
| 1. Bilateral | $p = 2 \min\{P(T \geq t), P(T \leq t)\}$ |
| 2. Unilateral per la dreta | $p = \{P(T \geq t)\}$ |
| 3. Unilateral per l'esquerra | $p = \{P(T \leq t)\}$ |

Objectius

Els problemes han de permetre que els alumnes assolisquen els objectius didàctics següents:

- 5a) Conèixer les fases i assimilar la nomenclatura dels contrastos.
- 5b) Saber interpretar els resultats dels contrastos per a poder inferir conclusions.
- 5c) Conèixer els contrastos d'alguns paràmetres d'una població.
- 5d) Conèixer els contrastos per a comparar alguns paràmetres de dues poblacions.

La taula següent ens mostra com estan distribuïts els objectius segons els exercicis:

<i>Objectiu</i> Exercici	<i>5a</i>	<i>5b</i>	<i>5c</i>	<i>5d</i>
1	X	X	X	
2	X	X	X	
3	X	X		
4		X	X	X
5		X	X	X
6		X		X
7			X	X
8		X	X	

Enunciats

-
- 5a) Conèixer les fases i assimilar la nomenclatura dels contrastos.
 - 5b) Saber interpretar els resultats dels contrastos per a poder inferir conclusions.
 - 5c) Conèixer els contrastos d'alguns paràmetres d'una població.

Exercici 1

Ens hem proposat analitzar les hores que estudien els alumnes d'estadística del grau de Relacions Laborals diàriament. Sabem que per estudis anteriors, que la mitjana del nombre d'hores de son és de 7,5 hores per setmana amb una desviació típica de 3 hores. Corrobora aquesta afirmació. Com a dades, triem una mostra de 25 estudiants i en calculem la mitjana mostral, que és de $\bar{x} = 7$.

-
- 5a) Conèixer les fases i assimilar la nomenclatura dels contrastos.
 - 5b) Saber interpretar els resultats dels contrastos per a poder inferir conclusions.
 - 5c) Conèixer els contrastos d'alguns paràmetres d'una població.

Exercici 2

Estem comprovant el funcionament d'una màquina que ens prepara el café en l'empresa. Se suposa que els cafés han de tindre 1 g de sucre aproximadament si el demanem a la meitat de dolçor, però quan es desajusta sospitem que aquest pes augmenta, perquè es veu que ix més sucre del que és necessari. Alguns treballadors es queixen d'aquest fet, ja que prefereixen el café més amarg. Per a comprovar aquest procés, triem al llarg de la jornada 20 cafés per a controlar-ne els pesos. Hem obtingut $\bar{x} = 0,995$ g i $S_{n-1} = 0,00513$. Estudia si els empleats tenen motiu per a demanar el canvi de la cafetera.

-
- 5a) Conèixer les fases i assimilar la nomenclatura dels contrastos.
 - 5b) Saber interpretar els resultats dels contrastos per a poder inferir conclusions.
-

Exercici 3

Estàvem comprovant el funcionament d'una màquina que ens prepara el café en l'empresa, però ara farem una estimació de la variància dels pesos del sucre que aporta la màquina expendedora de cafés. Considerarem que aquesta funciona correctament si la desviació típica del peso és de 0,005 g, és a dir, la seua variància és 0,000025 g².

Agafarem una mostra de 20 gots de café i prendrem els següents valors mostrals del pes del sucre: la grandària de la mostra $n = 20$, la mitjana aritmètica dels pesos, que és $\bar{x} = 0,995\text{g}$, i la seua desviació típica $S = 0,005\text{ g}$. El que volem saber és si aquest valor de la variància és correcte o no.

-
- 5b) Saber interpretar els resultats dels contrastos per a poder inferir conclusions.
 - 5c) Conèixer els contrastos d'alguns paràmetres d'una població.
 - 5d) Conèixer els contrastos per a comparar alguns paràmetres de dues poblacions.
-

Exercici 4

Volem comparar l'eficiència de dues empreses del sector ceràmic atenent al nombre d'hores que tardem en rebre les seues remeses per a elaborar els nostres productes i enviar-los a altres països. Creiem que el funcionament de les dues és semblant si en comparem les mitjanes i coneixem les desviacions típiques de l'empresa A (25 hores) i B (30 hores). Les dues es comporten com a variables normals. Per a confirmar la nostra creença, anotem en 10 enviaments de l'empresa A un temps mitjà de 80 hores, mentre que en la mostra dels 15 enviaments de l'empresa B el temps mitjà és de 75 hores.

-
- 5b) Saber interpretar els resultats dels contrastos per a poder inferir conclusions.
 - 5c) Conèixer els contrastos d'alguns paràmetres d'una població.
 - 5d) Conèixer els contrastos per a comparar alguns paràmetres de dues poblacions.

Exercici 5

Una empresa té dues plantes, de les quals es volen comparar els resultats dels tests d'eficiència recollits a través de les dades dels clients. Considerarem que les variàncies de les dues poblacions són iguals ($\sigma_x^2 = \sigma_y^2$). Sabem que la planta B és millor que la planta A i se suposa que aquesta es pot concretar almenys en 1 punt més en la mitjana de les puntuacions del test que utilitzem com a indicador. Per a fer el nostre treball, agafem les enquestes realitzades de cada planta i obtindrem aquests resultats: de la mostra de l'enquesta en A, anotarem $n = 10$, $\bar{x} = 7,9$, $S_{x_{n-1}}^2 = 6,77$; de la mostra de l'enquesta en B, anotarem $m = 12$, $\bar{y} = 9,75$, $S_{y_{n-1}}^2 = 6,39$. Es demana fer la comparació amb un nivell de significació del 10 %.

A més a més, per a millorar el grau de satisfacció dels clients, es comparen també amb una planta que està en un altre país on els preus són més baixos i suposem que, en aquest cas, ens donarà una millora en el grau de satisfacció dels clients.

Per a avaluar l'eficàcia d'aquesta decisió, s'ha passat una enquesta a 8 clients dissenyada per a esbrinar el grau mitjà de satisfacció de l'empresa X i l'empresa Y. Com que les formes de treballar són diferents, l'escala ha canviat: ara és de 0 a 3. Amb les seues respostes s'obtenen els resultats següents de les mostres de les empreses X i Y: $n = 8$, $\bar{x} = 1,45$, $\bar{y} = 1,6625$, i amb les dades de la variable diferència (D), calculem $S_{D_{n-1}}^2 = 0,0327$ (en aquest segon cas treballarem al 5 %).

-
- 5b) Saber interpretar els resultats dels contrastos per a poder inferir conclusions.
 - 5d) Conèixer els contrastos per a comparar alguns paràmetres de dues poblacions.

Exercici 6

Després de treballar amb dos ordinadors diferents, el portàtil i el de taula, volem comparar la qualitat de les dues màquines i veure si són igual d'eficients, considerant la proporció de les vegades que hem hagut de reiniciar-los. Per a realitzar el nostre treball seleccionem aleatòriament una mostra de 200 dies de treball del portàtil, en els quals hem hagut de reiniciar-lo 15 vegades, i 250 dies de l'ordinador de taula, en els quals hem tingut problemes 16 dies.

-
- 5c) Conèixer els contrastos d'alguns paràmetres d'una població.
 - 5d) Conèixer els contrastos per a comparar alguns paràmetres de dues poblacions.
-

Exercici 7

Què passaria si allò que estem comparant en dues empreses, equips o qualsevol altra cosa no fóra homogeni? En aquest cas estudiarem si el contrast de comparació de mitjanes del problema 5 és o no homogeni. Recordem que en aquell cas ja hem comentat que es tracta de dues mostres, les poblacions de les quals necessitem pressuposar que tenen les variàncies iguals.

-
- 5b) Saber interpretar els resultats dels contrastos per a poder inferir conclusions.
 - 5c) Conèixer els contrastos d'alguns paràmetres d'una població.
-

Exercici 8

En una empresa, dos treballadors fan torns per a anar a recollir els documents a classificar. Com que en les hores fora de la feina juguen molt als daus, han pensat que estaria bé jugar-s'ho tots els dies als daus. Però el que passa és que quasi sempre li toca al mateix company. Aquest pensa que els daus estan trucats i, per vore si és cert, fa un mostreig i obté que, fent 100 tirades, la proporció de traure un cinc és de 0,25 (sabem que aquesta proporció és de 0,167, si el dau està bé). Analitza si el dau estarà trucat utilitzant el p-valor.

Ajudes

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de resoldre els exercicis i problemes. És convenient no fer un ús excessiu d'aquestes ajudes, és a dir, que abans d'emprar l'ajuda cal pensar el problema almenys durant uns 10 – 15 minuts. Després es consultarà l'ajuda de tipus 1 i s'intentarà resoldre l'exercici amb aquesta ajuda. Si no és possible resoldre'l, llavors es consultarà l'ajuda de tipus 2; en darrer terme es consultarà la solució.

Ajudes de tipus 1

Exercici 1

Encara que no ho diga el problema, el que es demana és un contrast d'hipòtesis, que en aquest cas serà per a la mitjana (poblacional). Repasa els apunts de l'apartat del contrast d'hipòtesis per a una mostra.

Recorda també que, encara que no ho diga el problema, és necessària la introducció d'un nivell de significació, que en aquest cas pot ser del 95 %.

Exercici 2

En aquest cas també cal estudiar els apunts sobre el contrast d'hipòtesis de la mitjana, encara que canvia la possible alternativa. Cal també proposar un nivell de significació, que en aquest cas podria ser de l'1 % per a diferenciar-lo de l'exercici anterior.

Exercici 3

En aquest cas hem de treballar amb l'expressió: $T = \frac{(n-1)S^2}{\sigma^2}$.

Encara que utilitzarem la quasi-variància de la mostra, la qual podem obtenir a partir de la variància de la mostra: $S_{n-1}^2 = \frac{n}{n-1}S^2$

Exercici 4

Encara que parega que no es demana res, el que volem és comparar dues empreses, però a diferència del tema d'interval, el que volem fer ara és un contrast d'hipòtesis per a veure si la diferència de les empreses és 0. Amb tot açò, veiem que també es podria plantejar com un exercici d'interval.

En aquest cas, cal estudiar la teoria de l'apartat de contrastos d'hipòtesis de comparació o diferència de mitjanes.

També cal comentar que, com en altres casos, podem utilitzar un nivell de confiança del 99 %.

Exercici 5

Com a primera ajuda, el que cal destacar d'aquest problema és que té dos parts: un primer on es comparen dues plantes amb els seus valors mitjans d'eficiència i variàncies iguals, i un segon on es treballa amb valors de diferències de les variàncies. A més a més, en els dos casos treballarem amb contrastos d'hipòtesis unilaterals.

Exercici 6

En aquest cas, cal recordar que, encara que comparem dos equips informàtics, no ens referim a la mitjana de dies treballats, sinó a la comparació de proporcions del bon funcionament en cada cas.

Amb les dades obtenim que:

$$p_X = \frac{15}{200} = 0,075 \text{ per al cas del portàtil.}$$

$$p_Y = \frac{16}{250} = 0,064 \text{ per al cas de l'ordinador de taula.}$$

Exercici 7

El que se'ns demana és que dissenyem un test bilateral amb un nivell de significació del 5 %, per exemple amb:

$$\left\{ \begin{array}{l} H_0: \frac{\sigma_X^2}{\sigma_Y^2} = 1 \\ H_1: \frac{\sigma_X^2}{\sigma_Y^2} \neq 1 \end{array} \right. \quad R = \left\{ T : T \leq \frac{\alpha f_{n-1, m-1}}{2} \text{ o bé } T \geq \frac{1-\alpha f_{n-1, m-1}}{2} \right\}$$

L'estadístic serà:

$$T = \frac{\frac{(n-1)S_{X_{n-1}}^2}{\sigma_X^2 \cdot n}}{\frac{(m-1)S_{Y_{n-1}}^2}{\sigma_Y^2 \cdot m}}$$

Exercici 8

El que cal plantejar és el contrast $\left\{ \begin{array}{l} H_0: p = 0,167 \\ H_1: p \neq 0,167 \end{array} \right.$ amb l'estadístic $T = \frac{p - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$

Ajudes de tipus 2

En aquest apartat es presentaran les ajudes que cal emprar en cas de ser necessari a l'hora de resoldre els exercicis i problemes, després d'haver consultat l'ajuda de tipus 1.

Exercici 1

Sabent que és un contrast d'hipòtesis de la mitjana, com que la mitjana mostral és menor que el valor poblacional, proposarem una alternativa de mitjana menor que 7,5.

Amb aquestes dades i l'ús de les taules de la distribució normal d'exercicis anteriors, es pot solucionar el problema.

Exercici 2

El contrast que cal presentar és $\begin{cases} H_0: \mu = 1000 \\ H_1: \mu > 1000 \end{cases}$ amb una regió de rebuig $R = \{T: T \geq t_{1-\alpha, n-1}\}$.

L'estadístic de contrast es pot aconseguir amb la fórmula $T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$

Exercici 3

El contrast d'hipòtesis serà $\begin{cases} H_0: \sigma^2 = 0,000025 \\ H_1: \sigma^2 \neq 0,000025 \end{cases}$ amb la regió de rebuig

$R = \left\{ T : T \leq \frac{\alpha}{2} \chi^2_{n-1} \text{ o bé } T \geq \frac{1-\alpha}{2} \chi^2_{n-1} \right\}$.

Cal tindre en compte el càlcul dels valors per a la regió de rebuig i saber que per a aquest cas seran:

$$\frac{1-\alpha}{2} \chi^2_{n-1} = 0,975 \chi^2_{19} \quad \text{i} \quad \frac{\alpha}{2} \chi^2_{n-1} = 0,025 \chi^2_{19}$$

Si el valor està en la regió de rebuig, acceptarem la hipòtesi alternativa; en cas contrari, no contradirem l'afirmació inicial que diu que la variància dels pesos del sucre pot ser de 0,000025 g.

Exercici 4

El contrast per a la hipòtesi alternativa serà unilateral o bilateral? En aquest cas serà:

$$\begin{cases} H_0: \mu_X - \mu_Y = 0 \\ H_1: \mu_X - \mu_Y \neq 0 \end{cases}$$

La regió de rebuig d'aquest contrast $R = \{T: |T| \geq z_{\frac{1-\alpha}{2}}\}$, l'estadístic a calcular serà:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

Exercici 5

Pel que fa a la primera part, el contrast serà:

$$\begin{cases} H_0: \mu_X - \mu_Y = -1 \\ H_1: \mu_X - \mu_Y < -1 \end{cases} \quad R = \{T : T \leq -_{1-\alpha} t_{n+m-2}\}$$

Es pot veure que la diferència de les mitjanes és -1 .

L'estadístic a utilitzar per a aquest cas serà:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{(n-1)S_{x_{n-1}}^2 + (m-1)S_{y_{m-1}}^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Per a la segona part, com que comparem dos mostres i volem veure si són o no iguals, escriurem:

$$\begin{cases} H_0: \mu_X - \mu_Y = 0 \\ H_1: \mu_X - \mu_Y < 0 \end{cases} \quad R = \{T : T \leq -_{1-\alpha} t_{n-1}\}$$

Amb l'estadístic:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\frac{S_{D_{n-1}}}{\sqrt{n}}}$$

Exercici 6

Com que saben les proporcions, és clar que el test bilateral que es demana és:

$$\begin{cases} H_0: p_X - p_Y = 0 \\ H_1: p_X - p_Y \neq 0 \end{cases} \quad R = \{T : |T| \geq z_{1-\frac{\alpha}{2}}\}$$

Amb l'expressió de l'estadístic:

$$T = \frac{(p_x - p_y) - (p_X - p_Y)}{\sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_Y(1-p_Y)}{m}}}$$

Exercici 7

Cal recordar, per a no tindre problemes en els càlculs, que treballem amb taules de la funció de distribució d'aquesta variable F de Fisher- Snedecor.

Així, amb les taules o amb un programa estadístic, trobem el valor del percentil

$${}_{1-\frac{\alpha}{2}}f_{(n-1), (m-1)} = {}_{0,975}f_{9,11} = 3,5879.$$

Per calcular l'altre valor farem ús de la propietat següent:

$$\frac{\alpha}{2}f_{(n-1), (m-1)} = \frac{1}{{}_{1-\frac{\alpha}{2}}f_{(m-1), (n-1)}}$$

Exercici 8

Cal calcular, amb els valors de la funció de distribució de la variable Z, la probabilitat $P(Z \leq -1,9169)$ per tractar-se d'una funció simètrica.

A més a més, l'expressió que utilitzarem per a treballar amb el p-valor en aquest cas serà $p = 2 \min\{P(T \geq t), P(T \leq t)\}$.

El valor es comparà amb el nivell de significació, que en aquest cas és del 5 %.

Solucions

Exercici 1

Ens hem proposat analitzar les hores que estudien els alumnes d'estadística del grau de Relacions Laborals diàriament. Sabem que per estudis anteriors, que la mitjana del nombre d'hores de son és de 7,5 hores per setmana amb una desviació típica de 3 hores. Corrobora aquesta afirmació. Com a dades, triem una mostra de 25 estudiants i en calculem la mitjana mostral, que és de $\bar{x} = 7$.

Solució

Plantejarem un contrast unilateral per l'esquerra, ja que les sospites són que la mitjana pot haver disminuït. Treballarem amb un nivell de significació del 5 %.

$$\text{Contrast d'hipòtesi: } \begin{cases} H_0: \mu = 7,5 \\ H_1: \mu < 7,5 \end{cases} \quad \text{Regió de rebuig: } R = \{T : T \leq -z_{1-\alpha}\}$$

Calculem l'estadístic amb les dades mostrals i poblacionals de l'enunciat:

$$T = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{7 - 7,5}{\frac{3}{\sqrt{25}}} = \frac{-0,5}{0,6} = -0,8333$$

Per a trobar la regió crítica o de rebuig, calcularem el valor de $1 - \alpha$ i el percentil corresponent de la variable Z normal tipificada:

$$\alpha = 0,05 \quad \rightarrow \quad 1 - \alpha = 0,95$$

Per a trobar el valor de $z_{1-\alpha}$, calcularem el valor de x , tal que $P(Z \leq x) = 0,95$, amb l'ajuda de les taules o d'un programa informàtic. Així $z_{1-\alpha} = 1,64$.

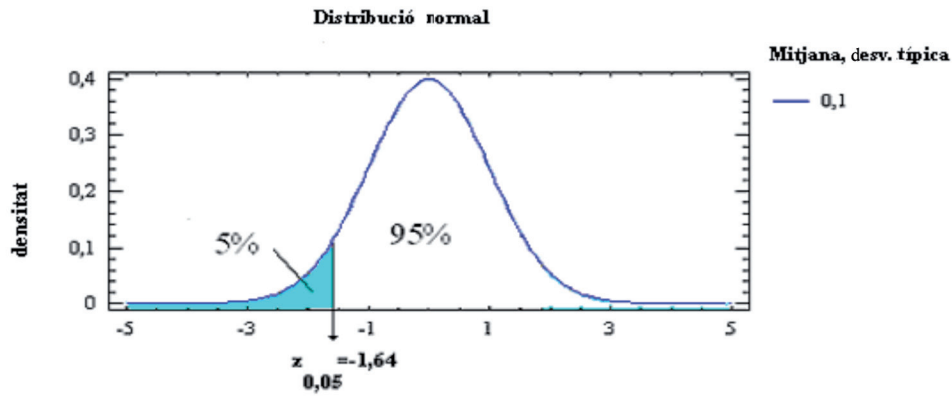


Fig. 27. Contrast unilateral per l'esquerra

$$R = \{T : T \leq -z_{1-\alpha}\} = \{T : T \leq -1,64\} = (-\infty, -1,64)$$

Podem comprovar que $Z = -0,8333 \notin R$, per tant no rebutjarem la hipòtesis nul·la i podrem donar per bona l'afirmació que les hores de son dels universitaris té una mitjana de 7,5 hores. Es pot observar gràficament en la figura anterior.

Exercici 2

Estem comprovant el funcionament d'una màquina que ens prepara el cafè en l'empresa. Se suposa que els cafès han de tindre 1 g de sucre aproximadament si el demanem a la meitat de dolçor, però quan es desajusta sospitem que aquest pes augmenta, perquè es veu que ix més sucre del que és necessari. Alguns treballadors es queixen d'aquest fet, ja que prefereixen el cafè més amarg. Per a comprovar aquest procés, triem al llarg de la jornada 20 cafès per a controlar-ne els pesos. Hem obtingut $\bar{x} = 0,995$ g i $S_{n-1} = 0,00513$. Estudia si els empleats tenen motiu per a demanar el canvi de la cafetera.

Solució

El que hem de fer és un contrast unilateral per la dreta amb un nivell de significació de l'1 % (aquesta dada la suposarem nosaltres).

Plantejarem les hipòtesis nul·la i alternativa i la regió de rebuig:

$$\begin{cases} H_0: \mu = 1000 \\ H_1: \mu > 1000 \end{cases} \quad R = \{T : T \geq t_{n-1, 1-\alpha}\}$$

Calculem l'estadístic amb les dades mostrals i poblacionals de l'enunciat:

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{0,995 - 1,000}{\frac{0,00513}{\sqrt{20}}} = -4,3588$$

Per a trobar la regió crítica o de rebuig, calcularem el valor de $1-\alpha$ i el percentil corresponent de la variable t de Student amb 19 graus de llibertat.

$$\alpha = 0,01 \rightarrow 1 - \alpha = 0,99$$

Per a trobar el valor de ${}_{1-\alpha}t_{n-1}$, calcularem el valor de x , tal que, $P(T \leq x) = 0,99$, amb ajut de les taules o d'un programa informàtic. Així, ${}_{1-\alpha}t_{n-1} = {}_{0,99}t_{19} = 2,54$.

$$R = \{T : T \geq {}_{1-\alpha}t_{n-1}\} = \{T : T \geq 2,54\} = (2,54, +\infty)$$

Podem comprovar que el valor de nostre estadístic és $T = -4,3588 \notin R$, això ens permet donar per veritable la hipòtesi nul·la i acceptar que la màquina funciona correctament. Es pot observar gràficament en la figura anterior.

Exercici 3

Estàvem comprovant el funcionament d'una màquina que ens prepara el café en l'empresa, però ara farem una estimació de la variància dels pesos del sucre que aporta la màquina expendedora de cafés. Considerarem que aquesta funciona correctament si la desviació típica del peso és de 0,005 g, és a dir, la seua variància és 0,000025 g².

Agafarem una mostra de 20 gots de café i prendrem els següents valors mostrals del pes del sucre: la grandària de la mostra $n = 20$, la mitjana aritmètica dels pesos, que és $\bar{x} = 0,995$ g, i la seua desviació típica $S = 0,005$ g. El que volem saber és si aquest valor de la variància és correcte o no.

Solució

Necessitem calcular la quasi-variància de la mostra i la podem obtenir a partir de la seua variància:

$$S_{n-1}^2 = \frac{n}{n-1} S^2 = \frac{20}{19} 0,000025 = 0,000026316$$

Realitzarem un contrast bilateral per a la variància σ^2 amb un nivell de significació del 5 %, la regió crítica R del qual és:

$$\left\{ \begin{array}{l} H_0: \sigma^2 = 0,000025 \\ H_1: \sigma^2 \neq 0,000025 \end{array} \right. \quad R = \left\{ T : T \leq \frac{\alpha}{2} \chi_{n-1}^2 \text{ o bé } T \geq \frac{1-\alpha}{2} \chi_{n-1}^2 \right\}$$

Utilitzarem l'expressió de l'estadístic anterior per a substituir els valors de la població i de la mostra:

$$T = \frac{(n-1)S^2}{\sigma^2} = \frac{19 \cdot 0,000026316}{0,000025} = 20$$

En aquest cas, el nivell de significació escollit és del 5 %, per tant:

$$1 - \alpha = 0,95 \quad \rightarrow \quad \alpha = 0,05 \quad \rightarrow \quad \frac{\alpha}{2} = 0,025 \quad \rightarrow \quad 1 - \frac{\alpha}{2} = 0,975$$

Per a determinar la regió crítica necessitem cercar els valors de la variable $\frac{1-\alpha}{2} \chi_{n-1}^2 = 0,975 \chi_{19}^2$ i $\frac{\alpha}{2} \chi_{n-1}^2 = 0,025 \chi_{19}^2$ en la funció acumulada de la distribució Xi-quadrat amb 19 graus de llibertat, ja que, en no ser simètrica la variable, obtenim resultats diferents en el valor absolut per a cada probabilitat.

En el nostre cas obtenim $0,975 \chi_{19}^2 = 32,8523$ i $0,025 \chi_{19}^2 = 8,90652$. Els substituïm en la fórmula de la regió de rebuig i tenim:

$$R = \left\{ T : T \leq \frac{\alpha}{2} \chi_{n-1}^2 \text{ o bé } T \geq \frac{1-\alpha}{2} \chi_{n-1}^2 \right\} = \left\{ T : T \leq 8,91 \text{ o bé } T \geq 32,85 \right\} = (-\infty; 8,91) \cup (32,85; +\infty)$$

És obvi que el valor del nostre paràmetre està fora d'aquesta regió, i per tant, donem com a vàlida la hipòtesi nul·la: considerarem que la màquina funciona correctament, perquè el valor de la variància extret de la mostra no contradueix l'afirmació inicial que la variància dels pesos és de 0,000025 g.

Exercici 4

Volem comparar l'eficiència de dues empreses del sector ceràmic atenent al nombre d'hores que tardem en rebre les seues remeses per a elaborar els nostres productes i enviar-los a altres països. Creiem que el funcionament de les dues és semblant si en comparem les mitjanes i coneixem les desviacions típiques de l'empresa A (25 hores) i B (30 hores). Les dues es comporten com a variables normals. Per a confirmar la nostra creença, anotem en 10 enviaments de l'empresa A un

temps mitjà de 80 hores, mentre que en la mostra dels 15 enviaments de l'empresa B el temps mitjà és de 75 hores.

Solució

Dissenyarem un test de contrast bilateral. Definim l'estadístic T i decidim treballar amb un nivell de significació de l'1 % per a decidir la regió de rebuig. Així:

$$\begin{cases} H_0: \mu_X - \mu_Y = 0 \\ H_1: \mu_X - \mu_Y \neq 0 \end{cases} \quad R = \{T : |T| \geq z_{1-\frac{\alpha}{2}}\}$$

Calculem l'estadístic amb les dades mostrals i poblacionals de l'enunciat, del qual sabem que es distribueix com una variable normal tipificada.

Per a la població A podem anotar el paràmetre de la població $\sigma_x^2 = 25^2 = 625$ i amb les dades de la mostra, anem $n = 10$, $\bar{x} = 80$.

Per a la població B podem anotar el paràmetre de la població $\sigma_y^2 = 30^2 = 900$, i amb les dades de la mostra, anem $m = 15$, $\bar{y} = 75$.

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} = \frac{(80 - 75) - 0}{\sqrt{\frac{625}{10} + \frac{900}{15}}} = \frac{5}{11,0680} = 0,4518$$

Per a trobar la regió crítica o de rebuig amb un nivell de significació de l'1 %, calculem el valor de $1 - \frac{\alpha}{2}$ i el percentil corresponent de la variable Z normal tipificada.

$$\alpha = 0,01 \quad \rightarrow \quad \frac{\alpha}{2} = 0,005 \quad \rightarrow \quad 1 - \frac{\alpha}{2} = 0,995$$

En la taula de la distribució acumulada de la variable normal tipificada trobem el valor de la variable x , tal que $P(Z \leq x) = 0,995$. Aquest valor és $x = 2,57583$, per la qual cosa, utilitzant la notació d'aquest tema $z_{1-\frac{\alpha}{2}} = z_{0,995} = 2,58$, el substituïm en l'expressió de R i tenim:

$$R = \{T : |T| \geq z_{1-\frac{\alpha}{2}}\} = \{T : |T| \geq 2,58\} = (-\infty; -2,58) \cup (2,58; +\infty)$$

$T = 0,4518 \notin R$, per tant, no rebutjarem la hipòtesis nul·la i podem donar per bona l'afirmació que les dues empreses tenen les mitjanes iguals i ens semblen igualment eficaces.

Exercici 5

Una empresa té dues plantes, de les quals es volen comparar els resultats dels tests d'eficiència recollits a través de les dades dels clients. Considerarem que les variàncies de les dues poblacions són iguals ($\sigma_x^2 = \sigma_y^2$). Sabem que la planta B és millor que la planta A i se suposa que aquesta es pot concretar almenys en 1 punt més en la mitjana de les puntuacions del test que utilitzem com a indicador. Per a fer el nostre treball, agafem les enquestes realitzades de cada planta i obtindrem aquests resultats: de la mostra de l'enquesta en A, anotarem $n = 10$, $\bar{x} = 7,9$, $S_{x_{n-1}}^2 = 6,77$; de la mostra de l'enquesta en B, anotarem $m = 12$, $\bar{y} = 9,75$, $S_{y_{m-1}}^2 = 6,39$. Es demana fer la comparació amb un nivell de significació del 10 %.

A més a més, per a millorar el grau de satisfacció dels clients, es comparen també amb una planta que està en un altre país on els preus són més baixos i suposem que, en aquest cas, ens donarà una millora en el grau de satisfacció dels clients.

Per a avaluar l'eficàcia d'aquesta decisió, s'ha passat una enquesta a 8 clients dissenyada per a esbrinar el grau mitjà de satisfacció de l'empresa X i l'empresa Y. Com que les formes de treballar són diferents, l'escala ha canviat: ara és de 0 a 3. Amb les seues respostes s'obtenen els resultats següents de les mostres de les empreses X i Y: $n = 8$, $\bar{x} = 1,45$, $\bar{y} = 1,6625$, i amb les dades de la variable diferència (D), calculem $S_{D_{n-1}}^2 = 0,0327$ (en aquest segon cas treballarem al 5 %).

Solució

Per a la primera part:

$$\begin{cases} H_0: \mu_X - \mu_Y = -1 \\ H_1: \mu_X - \mu_Y < -1 \end{cases} \quad R = \{T : T \leq -_{1-\alpha} t_{n+m-2}\}$$

Calculem l'estadístic amb les dades mostrals i poblacionals de l'enunciat, del qual sabem que es distribueix com una variable t de Student amb 20 graus de llibertat.

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{(n-1)S_{x_{n-1}}^2 + (m-1)S_{y_{m-1}}^2}{n+m-2} \sqrt{\frac{1}{n} + \frac{1}{m}}}} = \frac{(7,9 - 9,75) - (-1)}{\sqrt{\frac{9 \cdot 6,77 + 11 \cdot 6,39}{20} \sqrt{\frac{1}{10} + \frac{1}{12}}}} = \frac{-0,85}{1,0967} = -0,7750$$

Per a conèixer la regió de rebuig, cal esbrinar els valors dels extrems d'aquesta. Cal trobar el valor de $_{1-\alpha} t_{n+m-2}$ utilitzant un nivell de significació del 10 %.

$$\alpha = 0,10 \quad \rightarrow \quad 1 - \alpha = 0,90$$

Així, cal calcular, amb les taules de la distribució acumulada de la variable t de Student amb 20 graus de llibertat, el valor de la variable ${}_{1-\alpha}t_{n+m-2} = {}_{0,90}t_{20} = 1,3253$ i substituir aquest valor en la fórmula de la regió crítica R .

$$R = \{T : T \leq -{}_{1-\alpha}t_{n+m-2}\} = \{T : T \leq -{}_{0,90}t_{20}\} = (-\infty, -1,3253)$$

Com podem veure, $T = -0,7750 \notin R$, per la qual cosa no podem refutar la hipòtesi nul·la i direm que efectivament hi ha un punt de diferència a favor de l'empresa B en la mitjana de les puntuacions de l'indicador de qualitat, tot i que la informació de les mostres no permet inferir que la diferència siga major.

Per a la segona part, dissenyarem un test unilateral per l'esquerra (sembla obvi descartar que pugui empitjorar com a conseqüència de la mesura) amb un grau de significació del 5 %.

$$\begin{cases} H_0: \mu_X - \mu_Y = 0 \\ H_1: \mu_X - \mu_Y < 0 \end{cases} \quad R = \{T : T \leq -{}_{1-\alpha}t_{n-1}\}$$

Calculem l'estadístic amb les dades mostrals i poblacionals de l'enunciat, del qual sabem que es distribueix com una variable t de Student amb 7 graus de llibertat.

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\frac{S_{D_{n-1}}}{\sqrt{n}}} = \frac{(1,45 - 1,6625) - 0}{\frac{0,0327}{\sqrt{8}}} = \frac{-0,2125}{0,0116} = -18,3190$$

Per a conèixer la regió de rebuig, cal esbrinar els valors dels extrems d'aquesta. Cal trobar el valor de ${}_{1-\alpha}t_{n-1}$ utilitzant un nivell de significació del 5 %.

$$\alpha = 0,05 \rightarrow 1 - \alpha = 0,95$$

Així, cal calcular, amb les taules de la distribució acumulada de la variable t de Student amb 7 graus de llibertat, el valor de la variable ${}_{1-\alpha}t_{n-1} = {}_{0,95}t_7 = 1,8946$, i després substituir aquest valor en la fórmula de la regió crítica R .

$$R = \{T : T \leq -{}_{1-\alpha}t_{n-1}\} = \{T : T \leq -{}_{0,95}t_7\} = (-\infty, -1,8946)$$

És evident que $T \in R$. La informació de les mostres contradiu la hipòtesi inicial i ens porta a rebutjar-la. Com a conseqüència, podem dir que la mesura sí que ha reportat una millora en el grau de satisfacció dels clients de l'empresa a l'estranger, tal com ho indicava la nostra hipòtesi alternativa.

Exercici 6

Després de treballar amb dos ordinadors diferents, el portàtil i el de taula, volem comparar la qualitat de les dues màquines i vore si són igual d'eficients, considerant la proporció de les vegades que hem hagut de reiniciar-los. Per a realitzar el nostre treball seleccionem aleatòriament una mostra de 200 dies de treball del portàtil, en els quals hem hagut de reiniciar-lo 15 vegades, i 250 dies de l'ordinador de taula, en els quals hem tingut problemes 16 dies.

Solució

$$\begin{array}{lll} \text{Anotem aquestes dades mostrals:} & n = 200 & p_X = \frac{15}{200} = 0,075 \\ & m = 250 & p_Y = \frac{16}{250} = 0,064 \end{array}$$

Dissenyarem un test bilateral amb un nivell de significació del 5 % per a la diferència de proporcions.

$$\begin{cases} H_0: p_X - p_Y = 0 \\ H_1: p_X - p_Y \neq 0 \end{cases} \quad R = \{T : |T| \geq z_{1-\frac{\alpha}{2}}\}$$

Després substituïrem les dades en l'expressió de l'estadístic:

$$T = \frac{(p_x - p_y) - (p_X - p_Y)}{\sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}}} = \frac{(0,075 - 0,064) - 0}{\sqrt{\frac{0,075 \cdot 0,925}{200} + \frac{0,064 \cdot 0,936}{250}}} = \frac{0,011}{0,0242} = 0,4545$$

Com que treballarem amb un nivell de confiança del 5 %, podem calcular el valor de $1 - \frac{\alpha}{2}$ ja que $\alpha = 0,05 \rightarrow \frac{\alpha}{2} = 0,025 \rightarrow 1 - \frac{\alpha}{2} = 0,975$. En la taula de la distribució acumulada de la variable normal tipificada busquem el valor de la variable x , tal que $P(Z \leq x) = 0,975$. Aquest valor és $x = 1,96$, per la qual cosa $z_{1-\frac{\alpha}{2}} = 1,96$. Substituïm aquest valor en l'expressió de la regió de rebuig i obtenim:

$$R = \{T : |T| \geq z_{1-\frac{\alpha}{2}}\} = \{T : |T| \geq 1,96\} = (-\infty, -1,96) \cup (1,96; +\infty)$$

Per tant, $T = 0,45 \notin R$. La conclusió que podem extraure d'aquest resultat és que, treballant amb un nivell de significació del 5 %, la diferència entre els dos equips, amb les proporcions mostrals que hem obtingut, no és significativa, i podem estimar que la proporció de dies en què no ha funcionat un i l'altre és la mateixa.

Exercici 7

Què passaria si allò que estem comparant en dues empreses, equips o qualsevol altra cosa no fóra homogeni? En aquest cas estudiarem si el contrast de comparació de mitjanes del problema 5 és o no homogeni. Recordem que en aquell cas ja hem comentat que es tracta de dues mostres, les poblacions de les quals necessitem pressuposar que tenen les variàncies iguals.

Solució

Per a dur a terme els càlculs, identificarem els valors dels paràmetres de cadascuna de les mostres:

De la mostra de la població X, anotarem $n = 10$, $\bar{x} = 7,9$, $S_{X_{n-1}}^2 = 6,77$.

De la mostra de la població Y, anotarem $m = 12$, $\bar{y} = 9,75$, $S_{Y_{n-1}}^2 = 6,39$.

Dissenyarem un test bilateral amb un nivell de significació del 5 %.

$$\left\{ \begin{array}{l} H_0: \frac{\sigma_X^2}{\sigma_Y^2} = 1 \\ H_1: \frac{\sigma_X^2}{\sigma_Y^2} \neq 1 \end{array} \right. \quad R = \left\{ T : T \leq \frac{\alpha}{2} f_{n-1, m-1} \text{ o bé } T \geq \frac{1-\alpha}{2} f_{n-1, m-1} \right\}$$

Comencem per substituir els valors anteriors en l'estadístic:

$$T = \frac{\frac{(n-1)S_{X_{n-1}}^2}{\sigma_X^2 \cdot n}}{\frac{(m-1)S_{Y_{n-1}}^2}{\sigma_Y^2 \cdot m}} = \frac{\sigma_Y^2 \cdot m(n-1)S_{X_{n-1}}^2}{\sigma_X^2 \cdot n(m-1)S_{Y_{n-1}}^2} = \frac{12 \cdot 9 \cdot 6,77}{10 \cdot 11 \cdot 6,39} = \frac{731,16}{702,9} = 1,0402$$

Treballarem amb un nivell de significació del 5 %, per la qual cosa escriurem:

$$\alpha = 0,05 \quad \rightarrow \quad \frac{\alpha}{2} = 0,025 \quad \rightarrow \quad 1 - \frac{\alpha}{2} = 0,975$$

Buscarem en la taula de la distribució acumulada de la variable F de Fisher-Snedecor els percentils que corresponen al valors de $\frac{\alpha}{2} = 0,025$ i $1 - \frac{\alpha}{2} = 0,975$. Cal comentar que, si utilitzem taules, calcularem el primer valor en funció del segon, el qual sí que podrem trobar a les taules que solen publicar-se, aplicant la propietat que expliquem a continuació de la funció de distribució d'aquesta variable F de Fisher-Snedecor.

Així, amb les taules o amb un programa estadístic, trobem el valor del percentil $1-\frac{\alpha}{2} f_{(n-1), (m-1)} = 0,975 f_{9,11} = 3,5879$. Per calcular l'altre valor farem ús de la propietat

següent: $\frac{\alpha}{2} f_{(n-1), (m-1)} = \frac{1}{1-\frac{\alpha}{2} f_{(m-1), (n-1)}}$, que aplicada al nostre cas seria:

$$0,025 f_{9,11} = \frac{1}{0,975 f_{11,9}} = \frac{1}{3,91207} = 0,255619$$

Si substituïm tots aquests valors a l'expressió de la regió de rebuig, obtindrem com a resultat:

$$\begin{aligned} R &= \left\{ T : T \leq \frac{\alpha}{2} f_{n-1, m-1} \text{ o bé } T \geq \frac{1-\frac{\alpha}{2} f_{n-1, m-1}}{\alpha} \right\} = \left\{ T : T \leq 0,025 f_{9,11} \text{ o bé } T \geq 0,975 f_{9,11} \right\} = \\ &= \left\{ T : T \leq 0,2556 \text{ o bé } T \geq 3,5879 \right\} = (-\infty, 0,2556) \cup (3,5879, +\infty) \end{aligned}$$

El valor de $T = 1,040 \notin R$ permet inferir que la hipòtesi nul·la és veritable i que, consegüentment, les variàncies de les poblacions de les quals provenen les mostres són iguals, és a dir, que es presenta el mateix comportament en la dispersió dels valors de qualitat respecte de la mitjana de cadascuna de les dues mostres.

Exercici 8

En una empresa, dos treballadors fan torns per a anar a recollir els documents a classificar. Com que en les hores fora de la feina juguen molt als daus, han pensat que estaria bé jugar-s'ho tots els dies als daus. Però el que passa és que quasi sempre li toca al mateix company. Aquest pensa que els daus estan trucats i, per veure si és cert, fa un mostreig i obté que, fent 100 tirades, la proporció de traure un cinc és de 0,25 (sabem que aquesta proporció és de 0,167, si el dau està bé). Analitza si el dau estarà trucat utilitzant el p-valor.

Solució

Recordem l'exemple 3: Si el dau és correcte, la proporció de vegades de que isca el 5 és $p = \frac{1}{6} = 0,167$, mentre que amb el nostre dau hem fet 100 tirades i la proporció en la mostra és $p = 0,25$. Ens plantejarem que el dau està trucat.

Plantejarem les hipòtesis nul·la i alternativa d'un contrast bilateral:

$$\begin{cases} H_0: p = 0,167 \\ H_1: p \neq 0,167 \end{cases}$$

Calculem l'estadístic amb les dades mostrals i poblacionals de l'enunciat, sabent que aquest es distribueix com una normal tipificada.

$$T = \frac{p - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0,167 - 0,25}{\sqrt{\frac{0,25 \cdot 0,75}{100}}} = \frac{-0,083}{0,0433} = -1,9169$$

Cal calcular, en els valors de la funció de distribució de la variable Z , la probabilitat $P(Z \leq -1,9169) = 0,0277$ per tractar-se d'una funció simètrica.

$$\text{Així, } p = 2 \min\{P(T \geq t), P(T \leq t)\} = 2 \cdot 0,0277 = 0,0554$$

Per la qual cosa, si treballem amb un nivell de significació del 5 %, per ser menor que el p-valor, acceptariem la hipòtesi nul·la (com passava a l'exemple 3), tot i que per a valors superiors a 5,5 % de nivell de significació la hipòtesi nul·la hauria estat rebutjada i hauriem conclòs que el dau estara trucat.

Bibliografia

- BARBANCHO, A. G. (1992): *Estadística elemental moderna*, Ariel.
- BELTRAN, M. J. i PERIS, M. J. (2011): *Introducció a l'estadística aplicada a les ciències socials*, Universitat Jaume I, Castelló de la Plana.
- BIOSCA, A. i altres (1999): *Matemáticas aplicadas a las ciencias sociales II*, Edebé, Barcelona.
- BRUNET, I. i altres (2000): *Les tècniques d'investigació social i la seva aplicació*, Universitat Rovira i Virgili, Tarragona.
- COLERA, J. i altres (2003): *Matemáticas aplicadas a las ciencias sociales*, Anaya, Madrid.
- CORREA, J. C. i GONZÁLEZ, N. (2002): *Gráficos en R*, Universidad Nacional, Medellín.
- ESCUDELLER VALLÉS, R.: *Métodos estadísticos aplicados a la economía*, Ariel.
- FERNÁNDEZ CUESTA, C. i FUENTES GARCÍA, F. (1994): *Curso de estadística descriptiva. Teoría y práctica*, Ariel.
- GRACIA, F. i altres (1997): *Problemas de probabilidad y estadística*, Tilde, València.
- IBÁÑEZ, M. V. i SIMÓ, A. (2002): *Apuntes de estadística para ciencias empresariales*, Universitat Jaume I, Castelló de la Plana.
- KAZMIER, L. (1998): *Estadística aplicada a la administración y a la economía*, McGraw-Hill.
- KENETT, R. S. i ZACKS, S. (2003): *Statistical Inference*, Courier Dover Publications.
- MARTÍN, GUZMÁN, M. P. i MARTÍN PLIEGO, J. (1991): *Curso básico de estadística económica*, Libros AC.
- MARTÍN-PLIEGO LÓPEZ, L. i RUIZ-MAYA PÉREZ, L. (2005): *Fundamentos de inferencia estadística*, Thomson.
- MARTÍN PLIEGO, F. (1994): *Introducción a la estadística económica y empresarial*, Libros AC.
- MEYER, P. L. (1986): *Probabilidad y aplicaciones estadísticas*, Addison-Wesley.
- MONTEAGUDO, M. F. i PAZ, J. (2003): *Matemáticas aplicadas a las ciencias sociales II*, Luis Vives, Saragossa.
- MONTERO LORENZO, J. M. (2003): *Estadística para relaciones laborales*, Libros AC.
- NEWBOLD, P. i altres (2007): *Estadística para administración y economía*, Prentice Hill.
- RAO, C. R. (1985): *Inferencia estadística lineal*, Servicio Central de Publicaciones, País Basc.
- RITCHEY, F. J. (2008): *Estadística para las ciencias sociales*, McGraw-Hill.
- ROHATGI, V. K. (2004): *Estadística industrial moderna*, ITP Latin America.
- TOMEU PERUCHA, V. i UÑA JUÁREZ, I.: *Diez lecciones de estadística descriptiva (curso teórico-práctico)*, Libros AC.
- TRIOLA, M. F. (2000): *Estadística elemental*, Pearson Educations.
- SANZ, J. A. i altres (1996): *Problemas de estadística descriptiva empresarial*, Ariel.
- SPIEGEL, M. R.: *Estadística*, McGraw-Hill.

- VENABLES, W. N. i SMITH, D. M. (2008): *An introduction to R*, The R Development Core Team.
- WEBSTER, A. L. (2000): *Estadística aplicada a los negocios y a la economía*, McGraw-Hill.
- WONNACOT, T. H. i WONNACOT, R. J. (1996): *Introducción a la estadística*, Limusa Noriega Editores.
- ZAIATS, V. i altres (1999): *Probabilitat i estadística. Exercicis 1*, Eumo Editorial.

Webs

www.ine.es

www.ub.edu

www.ub.edu/stat/GrupsInnovació/Statmedia

www.bioestadistica.uma.es

www.bioestadistica.uma.es/libro

www.monografias.com

www.rtve.es

<http://www.rtve.es/noticias/20131024/numero-hogares-todos-miembros-paro-jo-13400-entre-julio-septiembre/775020.shtml>

