Estadística I
(*Probabilitat*)

Xavier Vilà
Universitat Autònoma de Barcelona

Curs 2007-2008

Universitat Autònoma de Barcelona

Índex

Capítol 1

Introducció

Qualsevol teoria científica aspira a explicar fenòmens que observem en el món real mitjançant un model matemàtic que incorpori tots els factors que influeixen en el seu comportament. L'objectiu principal és entendre el seu funcionament i fer prediccions que s'ajustin el màxim possible al que observem a la realitat. Per exemple, la Teoria de la gravitació universal de Newton explica com i perquè la Terra gira al voltant del sol dins una òrbita constant i a una velocitat determinada. Mitjançant aquesta teoria no només podem explicar perquè la Terra gira al voltant del sol, també podem fer prediccions de quina serà la seva posició en qualsevol moment. Fenòmens com el moviment de la Terra s'anomen *deterministes* perquè, en les mateixes condicions, el seu comportament està determinat, és a dir, és sempre el mateix.

De vegades, però, els fenòmens que es volen estudiar no tenen aquesta característica de *determinisme*, són fenòmens que no sempre tenen el mateix comportament. Per exemple, quan llancem un dau no sempre observem el mateix número, o quan modifiquem el tipus d'interès a una economia no sempre s'observa el mateix impacte sobre el producte interior brut. Aquests tipus de fenòmens s'anomenen *estocàstics* o *aleatoris* i tenen la característica de que el seu comportament està principalment determinat per *l'atzar*[1]. En la majoria d'aquests casos les eines matemàtiques tradicionals (àlgebra, càlcul, etc) que serveixen per a construir teories sobre fenòmens *deterministes* no són suficients i calen eines addicionals, són les eines *probabilístiques*

Tot i que hi ha evidència de que la humanitat ja participava en *jocs d'atzar* fa milers d'anys[2], l'estudi formal[3] de l'atzar no va començar fins els segles XVI (Pacioni, Cardano, Tartaglia) i XVII (Pascal, Fermat), desenvolupant-se de forma unificada i rigorosa als segles XVIII (Bernoulli, Bayes) i XIX (Laplace, Lagrange, De Morgan, Boole). El segle XX, gràcies principalment als treballs de Kolmogorov, representa l'establiment de la Teoria moderna de la probabilitat, què és la que se segueix en aquestes notes

---

[1] Molts científics opinen que, en realitat, l'atzar no existeix. Defensen que l'atzar és una entelèquia creada per l'home per tal de poder donar explicació a fenòmens *deterministes* extremadament complexes, determinats per un número molt gran de factors. Pretendre construir teories matemàtiques que contemplin tots aquests factors i les seves interrelacions seria totalment inviable

[2] Hi ha pintures a les piràmides d'Egipte que daten de l'any 3500 a.c.que mostren jocs d'atzar. Al museu de les runes d'Empúries es poden observar els daus que feien servir els grecs i els romans.

[3] Les civilitzacions antigues atribuïen aquests resultats *atzarosos* a la *divina providència*

Capítol 2

Estadística Descriptiva

2.1 L'Estadística: Població i Mostra, Variables i Observacions

De forma simplificada, es pot definir l'estadística de la següent manera

Definició 2.1.1 *L'Estadística és un conjunt de tècniques l'objectiu de les quals és entendre i treure conclusions sobre un **fenomen** (o fenòmens) concret en un **lloc i moment determinats** a partir d'unes **dades***

Per exemple, podem estudiar **l'atur** a **Catalunya al 2007** a partir de l**'Enquesta de Població Activa**, o l**a intenció de vot a les properes eleccions generals** al parlament de l'estat espanyol a partir de **les enquestes publicades als diaris**

Tècnicament, el "fenomen" a estudiar s'anomena **variable**, mentre que el "lloc i moment determinats" són la **població**. De la mateixa manera, les "dades" també s'acostumen a anomenar **observacions** de la **mostra**. D'aquesta manera,

Definició 2.1.2 *La **Població** és el conjunt d'elements que s'estudia i sobre el qual es vol treure una conclusió amb respecte d'alguna característica (o variable) seva*

Definició 2.1.3 *La **mostra** és un subconjunt de la població utilitzat per a recollir informació i treure conclusions sobre la mateixa*

Definició 2.1.4 *Reben el nom de **Variables** aquelles característiques o fenòmens que es volen estudiar de la població.*

Definició 2.1.5 *Reben el nom de **Observacions** els valors concrets de les variables considerades obtinguts a partir de la mostra*

Per tant, en el primer dels exemples anteriors, la **població** és la *població activa de Catalunya al 2007* i les **observacions** són les que s'obtenen a la **mostra** anomenada *Enquesta de població activa*. La **variable** a estudiar és, clarament, l'atur. A l'altre exemple, la **població** són tots els ciutadans de l'estat espanyol de més de 18 anys, la

PARTIDO	RESULTADO ELECCIONES MARZO 2004		INTENCION DE VOTO DECIDIDO				
			30/05/07	12/06/07	04/07/07	29/08/07	13/09/07
	%	ESC.	%	%	%	%	%
PSOE	42,64	164	44,0	43,0	45,0	44,0	**44,0**
P.P.	37,64	148	38,0	40,0	38,5	39,5	**38,0**
IU - ICV	4,96	5	5,0	5,0	5,0	3,5	**5,0**
CiU	3,24	10	2,5	3,0	4,5	4,5	**3,5**
PNV	1,63	7	1,5	1,5	1,5	1,3	**1,2**
ERC	2,54	8	1,5	2,0	1,7	1,9	**1,8**
OTROS	7,35	8	7,5	5,5	3,8	5,3	**6,5**
TOTAL	100,00	350	100,0	100,0	100,0	100,0	**100,0**

Figura 2.1: Intenció de Vot (Font: El pulsómetro-Cadena Ser-13/09/2007)

mostra és l'*enquesta publicada*, les **observacions** són les dades que en ella es recullen i la **variable** d'interès és la *intenció de vot*. (Veure figura 2.1)

Així, de forma més acurada podem re-definir l'estadística de la següent manera

Definició 2.1.6 *L'Estadística és un conjunt de tècniques l'objectiu de les quals és entendre i treure conclusions sobre una **variable** (o variables) concreta en una **població** a partir de les **observacions** d'una **mostra***

D'altra banda, per tal de poder realitzar una anàlisi estadística a partir de les dades que configuren la mostra fent servir les eines que veurem, cal ordenar i presentar de forma clara, simplificada i entenedora aquesta informació. D'això s'ocupa la *Estadística Descriptiva*

Definició 2.1.7 *L'Estadística Descriptiva tracta de l'estudi d'un conjunt de dades referides a una o més variables d'una població i de la seva ordenació per tal de facilitar la seva comprensió i utilitat*

Un cop aquesta informació esta ordenada de forma que resulti útil a l'investigador, es poden utilitzar tècniques de "Inferència Estadística", què utilitzen eines probabilístiques, per tal d'estendre les conclusions que s'obtinguin sobre la *mostra* al conjunt de la *població* a estudi

2.2 Distribucions de Freqüències

2.2.1 Tipus de Variables

Com a d'altres branques de les matemàtiques, les variables acostumen a denotar-se amb les últimes lletres de l'alfabet (en majúscules), X, Y, Z. Si es consideren moltes variables a la vegada, resulta més convenient utilitzar subíndexs, $X_1, X_2, \ldots, X_n$.

Les diferents observacions de cada variable que s'obtenen d'una mostra sovint es denoten amb subíndexs de la variable corresponent i amb minúscules. Així, si tenim 3 observacions de la variable X_3, les denotarem x_{31}, x_{32}, x_{33}.

Segons la seva naturalesa, trobarem que hi ha variables de dos tipus diferent:

1. **Variables Qualitatives** (o categòriques) Són variables que no és poden mesurar numèricament. Cada observació s'associa a un número (o lletra)

2. **Variables Quantitatives** (o mesurables) Són variables que es poden mesurar numèricament. Poden ser de dos tipus

   (a) *Contínues* Poden prendre qualsevol valor dins d'un interval

   (b) *Discretes* Poden prendre qualsevol valor de una llista finita (o numerable) de valors[1]

Exemple 2.2.1 *Consideren la taula 2.1, què resumeix una enquesta realitzada a 10 famílies de Cerdanyola*

Aquí tenim 5 variables

X_1*: Components de la família*
X_2*: Professió del cap de família*
X_3*: Ingressos mensuals*
X_4*: Despesa Telefònica mensual (sense quotes fixes)*
X_5*: ADSL (Si=1, No=0)*

De cada una d'aquestes variables tenim 10 observacions

$$\{x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{110}\} = \{2, 3, 4, 5, 8, 2, 4, 5, 7, 2\}$$
$$\{x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{210}\} = \{3, 3, 6, 3, 1, 5, 3, 1, 4, 1\}$$

i així successivament amb totes les variables

Amb respecte del seu tipus tenim:

- $X_1 \to$ *Quantitativa Discreta*

$$X_1 \in \{1, 2, 3, 4, 5, 6, \ldots\}$$

- $X_2 \to$ *Qualitativa*

$$X_2 \in \{1, 2, 3, 4, 5, 6\}, \textit{ on}$$

1 *Empresari*
2 *Personal Qualificat*
3 *Personal No Qualificat*
4 *Professional Liberal*
5 *Ensenyants*
6 *Atur*

- $X_3 \to$ *Quantitativa Contínua*

[1] En la pràctica, són com variables qualitatives

	X_1	X_2	X_3	X_4	X_5
Família	No. de Components	Professió cap de família	Ingressos mensuals	Despesa telefònica	ADSL (Si/No)
1	2	3	592.18	12.06	0
2	3	3	743.83	14.88	0
3	4	6	527.18	12.35	0
4	5	3	1090.47	18.92	0
5	8	1	2744.23	26.5	1
6	2	5	902.71	15.89	1
7	4	3	888.26	14.44	0
8	5	1	1588.76	15.38	1
9	7	4	3069.2	24.19	1
10	2	1	707.72	13.72	0

Taula 2.1: Famílies de Cerdanyola

$$X_3\,[0,\infty]$$

- $X_4 \rightarrow$ *Quantitativa Contínua*

$$X_4\,[0,\infty]$$

- $X_5 \rightarrow$ *Qualitativa*

$$X_5 \in \{0,1\}, \textit{ on}$$

0 *No té ADSL*

1 *Té ADSL*

2.2.2 Distribucions de Freqüències

Les observacions de una variable "per se" donen poca informació d'utilitat, sobre tot quan el seu número és elevat. Cal classificar-les de manera que resultin informatives. Això es fa de forma diferent depenent de si es tracta d'una variable *Quantitativa Contínua* o de una variable *Quantitativa Discreta o Qualitativa.* El què es tracta, en definitiva, és de conèixer quantes vegades es repeteix cada observació (quanta gent està aturada, quanta gent votarà al PP, etc) o d'altre tipus de resum.

2.2.2.1 Distribució de Freqüències de variables qualitatives o quantitatives discretes

En el cas de variables qualitatives o quantitatives discretes (es tracten de la mateixa manera), el que es fa és mirar quants valors diferents ha pres aquesta variable i després comptar quantes vegades ha pres cada un d'aquests valors.

En el exemple anterior, $X_1,\ X_2,\ i\ X_5$ poden ser, doncs, tractades de la mateixa manera.

Agafem X_1com a exemple i recordem la llista d'observacions que tenim

$$\{x_{11},x_{12},x_{13},x_{14},x_{15},x_{16},x_{17},x_{18},x_{19},x_{110}\}=\{2,3,4,5,8,2,4,5,7,2\}$$

Veiem doncs que tenim 10 observacions la variable X_1, però que no momes pren 6 valors diferents, que es representen amb la lletra Y. Així, la llista d'aquests valors és

$$\{y_{11},y_{12},y_{13},y_{14},y_{15},y_{16}\}=\{2,3,4,5,7,8\}$$

Podem, a partir d'aquí, construir una *Distribució de Freqüències* què consisteix en veure com la freqüència amb la que es repeteix cada valor es distribueix entre totes les observacions. Podem definir, doncs, els conceptes de *Freqüència absoluta* i de *Freqüència relativa*.

Definició 2.2.2 *S'anomena* ***Freqüència Absoluta*** (n_i) *de un determinat valor* y_i *al número de vegades que aquest valor es repeteix entre les observacions*

Definició 2.2.3 *S'anomena **Freqüència Relativa** (f_i) (o proporció) d'un determinat valor y_i al percentatge de vegades (proporcions) que aquest valor apareix entre les observacions*

Clarament, si n si és el total de observacions que tenim,

$$f_i = \frac{n_i}{n}$$

Finalment, podem calcular[2] les *Freqüències Acumulades* (Absolutes (N_i) i Relatives (F_i)) simplement sumant, per a cada valor, totes les freqüències (absolutes (n_i) i relatives (f_i)) de valor menor incloent el valor del qual estem calculant la freqüència acumulada. Així, aquestes *freqüències acumulades* s'interpreten de la següent manera:

$N_i \rightarrow$ *número* d'observacions de la variable que són $\leq y_i$

$F_i \rightarrow$ *proporció* d'observacions de la variable que són $\leq y_i$

La taula 2.2 recull totes les freqüències de la variable X_1. Observar, per exemple, les interpretacions següents:

$n_3 = 2 \rightarrow$ Hi ha 2 observacions de la variable que són $= y_3 = 4$

$f_3 = 0.2 \rightarrow$ El 20% de les observacions de la variable són $= y_3 = 4$

$N_3 = 6 \rightarrow$ Hi ha 6 observacions de la variable que són $\leq y_3 = 4$

$F_3 = 0.6 \rightarrow$ El 60% de les observacions de la variable són $\leq y_3 = 4$

Les diferents *freqüències* d'una variable sempre compleixen les següents **propietats**, on k representa el número de valors diferents que pren la variable[3]

(i) $0 \leq n_i \leq n; \quad 0 \leq N_i \leq n$

(ii) $0 \leq f_i \leq 1; \quad 0 \leq F_i \leq 1$

(iii) $\sum_{i=1}^{k} n_i = n$

(iv) $\sum_{i=1}^{k} f_i = 1$

(v) $n_1 = N_1 \leq N_2 \leq \cdots \leq N_k = n$

(vi) $f_1 \leq F_1 \leq F_2 \leq \cdots \leq F_k = 1$

[2] No té sentit fer-ho per a *variables qualitatives* ja que aquest càlcul implica que l'ordre dels valors Y_i és un ordre natural i que té sentit

[3] En el exemple anterior, $k = 6$.

Valor de X_1	Freqüència Absoluta (n_i)	Freqüència Relativa $(f_i = \frac{n_i}{n})$	Freqüència Absoluta Acumulada (N_i)	Freqüència Relativa Acumulada $(F_i = \frac{N_i}{n})$
2	3	0.3	3	0.3
3	1	0.1	4	0.4
4	2	0.2	6	0.6
5	2	0.2	8	0.8
7	1	0.1	9	0.9
8	1	0.1	10	1

Taula 2.2: Taula de Freqüències de la variable X_1

2.2.2.2 Distribució de Freqüències de variables quantitatives contínues

Una variables contínua (la *renda*, la *temperatura*, etc) pot prendre tants valor diferents que una taula de freqüències (o el què és el mateix, una *distribució de freqüències*) afegiria molt poca informació. El més probable és que cada observació tindria un valor diferent i, per tant, totes les *freqüències absolutes* serien igual a 1.

Per a resoldre això el que es fa és agrupar les dades en intervals, que s'anomenen *Intervals de classe* (o *classes*). El procediment per a construir-los és el següent:

1. **Rang** (Diferència entre el valor més gran i el més petit)
   Si $y_1, y_2, \ldots, y_k$ són els diferents valors que observem de la variable **ordenats** de més petit a més gran, el *rang* és

   $$\textit{rang} = y_k - y_1$$

   Per exemple, si considerem la variable X_3de la taula2.1, llavors el *rang* és

   $$\text{RANG} = 3069.2 - 527.18 = 2542.02$$

   Aquesta "longitud total" és la que cal dividir en intervals

2. **Longitud de l'Interval** (Amplada de cada interval)
   Un cop tenim el *rang*, cal dividir aquesta longitud entre els intervals. Això dependrà del número d'intervals desitjats. Aquest número es decideix en funció del problema que es tracti. Si decidim que el número d'intervals és I, aleshores la **longitud de cada interval** (l) serà igual a la longitud total (el *rang*) dividida pel número de intervals[4].

   $$l = \frac{rang}{I}$$

   Per exemple, en el cas de la variable X_3 de la taula2.1 (Ingressos mensuals) si volem 8 intervals farem:

   $$l = \frac{2542.02}{8} = 317.75$$

3. **Construcció dels intervals**
   Un cop tenim determinada la longitud l de cada interval, la seva construcció es realitza de forma iterada a partir del valor més petit observat de la variable, és a dir, el valor $y_1$si la llista de valors $y_1, \ldots, y_k$ està ordenada de més petit a més gran. Així, el primer interval serà$[y_1, y_1 + l)$, el segon interval començarà on acaba el primer i tindrà també una longitud l $[y_1 + l, (y_1 + l) + l)$, i així successivament ...

A la taula 2.3 hi ha la distribució de freqüències de la variable X_3 obtinguda a partir de la informació de la taula 2.1. Com pot observar-se, a més de la descripció dels intervals i de la seva longitud, també apareix (tercera columna), L'anomenada *marca de classe*. Aquesta *marca de classe* coincideix amb el **punt mig** de l'interval i , d'alguna manera,

[4]També és podria fer que cada interval tingues una longitud diferent (l_i). En aquest cas, el que s'ha de complir és que $\sum_{i=1}^{k} l_i = rang$

Interval	Longitud l_i	Marca de Classe c_i	Freq. Absoluta n_i	Freq. Relativa f_i	Freq. Abs. Acumulada N_i	Freq. Rel. Acumulada F_i
[527.18,844.93)	317.75	686.06	4	0.4	4	0.4
[844.93,1162.69)	317.75	1003.81	3	0.3	7	0.7
[1162.69,1480.44)	317.75	1321.56	0	0	7	0.7
[1480.44,1798.19)	317.75	1639.31	1	0.1	8	0.8
[1798.19,2115.94)	317.75	1957.07	0	0	8	0.8
[2115.94,2433.7)	317.75	2274.82	0	0	8	0.8
[2433.7,2751.45)	317.75	2592.57	1	0.1	9	0.9
[2751.45,3069.2]	317.75	2910.32	1	0.1	10	1
			10	1		

Taula 2.3: Taula de Freqüències de la variable X_3

"representa" l'interval. Més endavant veurem com la *marca de classe* s'utilitza per a fer alguns càlculs sobre la variable en qüestió.

En aquesta mateixa taula 2.3 apareix el càlcul de les freqüències corresponents a aquesta variable. La seva obtenció és similar a la realitzada en el cas de *variables discretes*, només que en aquest cas, per a calcular la freqüència absoluta de cada interval, cal comptar quantes observacions de la variable són a dins l'interval.

2.3 Mesures de Centralització

Són mesures (*estadístics*) que tracten de resumir la informació que proporciona les observacions de la (les) variables(s) amb uns valors descriptius. Les tres *mesures centrals* principals són: La *mitjana*, la *mediana*, i la *moda*. Les dues primeres tracten de trobar un "valor central" que representa els valors obtinguts de la variable i al voltant del qual oscil·len les observacions. La darrera, la *moda*, simplement identifica el valor que més cops apareix entre les observacions de les variables.

2.3.1 La mitjana aritmètica

Representa un valor central al voltant del qual oscil·len les observacions de les variables, constituint el centre de gravetat de la distribució. El seu càlcul és senzill. Si tenim n observacions de la variable X, $\{x_1, x_2, \ldots, x_n\}$, la *mitjana aritmètica* (denotada $\bar{X}$) és:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Si les dades sobre la variable X en qüestió venen recollides en una *taula de freqüències* (no tenim les observacions en "brut"), aleshores la mitjana s'obté a partir de les freqüències absolutes n_i de la següent manera:

$$\bar{X} = \frac{1}{n}\sum_{i=}^{k} n_i y_i$$

o també a partir de les freqüències relatives f_i,

$$\bar{X} = \sum_{i=1}^{k} f_i y_i$$

on $\{y_1, y_2, \ldots, y_k\}$ són els k diferents valors que s'han observat de la variable.

En el cas de *variables contínues*, si les dades estan agrupades en intervals, aleshores el càlcul de la mitjana només es pot fer de forma <u>aproximada</u> de la següent manera:

$$\bar{X} \approx \frac{1}{n}\sum_{i=1}^{I} n_i c_i \quad \text{ó} \quad \bar{X} \approx \sum_{i=1}^{I} f_i c_i$$

on $\{c_1, c_2, \ldots, c_I\}$ són les "marques de classe" dels I diferents intervals.

Exemple 2.3.1 ***[Variable discreta]*** *En el cas de la variable X_1 de la taula 2.1, podem calcular la mitjana directament a partir de les 10 observacions de la variable*

$$\bar{X}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1i} = \frac{1}{10}(2+3+4+5+8+2+4+5+7+2) = 4.2$$

Si en comptes de les "dades en brut" de la taula 2.1 només tinguéssim la informació resumida en la taula de freqüències 2.2 podríem fer el càlcul a partir d'aquestes,

$$\bar{X}_1 = \frac{1}{n}\sum_{i=1}^{k} n_i y_{1i} = \frac{1}{10}(3\cdot 2+1\cdot 3+2\cdot 4+2\cdot 5+1\cdot 7+1\cdot 8) = 4.2$$

o bé mitjançant les freqüències relatives

$$\bar{X}_1 = \sum_{i=1}^{k} f_i y_{1i} = (0.3\cdot 2+0.1\cdot 3+0.2\cdot 4+0.2\cdot 5+0.1\cdot 7+0.1\cdot 8)$$

Exemple 2.3.2 ***[Variable contínua]*** *En el cas de la variable X_3 de la taula 2.1, podem calcular la mitjana directament a partir de les 10 observacions de la variable*

$$\bar{X}_3 = \frac{1}{n}\sum_{i=1}^{n} x_{3i} = \frac{1}{10}(592.18+743.83+\cdots+707.72) = 1285.45$$

Si en comptes de les "dades en brut" de la taula 2.1 només tinguéssim la informació resumida en la taula de freqüències 2.2 podríem fer el càlcul a partir d'aquestes, però només de forma aproximada fent servir les "marques de classe" dels intervals:

$$\bar{X}_3 \approx \frac{1}{n}\sum_{i=1}^{k} n_i c_{3i} = \frac{1}{10}(4\cdot 686.06+3\cdot 1003.81+\cdots+1\cdot 2910.32) = 1289.79$$

Si ho féssim a partir de les freqüències relatives obtindríem el mateix resultat,

$$\bar{X}_3 \approx \frac{1}{n}\sum_{i=1}^{k} n_i c_{3i} = (0.4\cdot 686.06+0.3\cdot 1003.81+\cdots+0.1\cdot 2910.32) = 1289.79$$

Com es pot comprovar, el valor de la mitjana $\bar{X}$ obtingut del càlcul a partir de les dades agrupades (fent servir les freqüències) (1289.79) és només aproximadament igual al valor real (1285.45) de la mitjana

La *mitjana aritmètica* $\bar{X}$ té les següents propietats què en ocasion resulten d'utilitat

1. Sigui a una constant qualsevol i $\{x_1, x_2, \ldots, x_n\}$ les n observacions de la variable X. Aleshores, la *mitjana* del resultat de multiplicar les observacions per la constant $(\{a\cdot x_1, a\cdot x_2, \ldots, a\cdot x_n\})$ és $a\cdot\bar{X}$
2. Siguin X i Y dues variables diferents. Aleshores, $\overline{X+Y} = \bar{X}+\bar{Y}$
3. Siguin $\{x_1, x_2, \ldots, x_n\}$ les n observacions de la variable X. Aleshores,

$$\sum_{i=1}^{n}(x_i - \bar{X}) = 0$$

4. La *mitjana aritmètica* sempre existeix, és única, i està molt afectada pels valors extrems

És important entendre que la *mitjana mostral* $\bar{X}$ és la mitjana calculada a partir de les observacions de la mostra i, per tant, és només una aproximació al valor autèntic de la *mitjana poblacional* (o mitjana real de la població) què normalment es denota amb el símbol μ. Per exemple, el fet de que en el exemple 2.3.1 hem calculat la mitjana de X_1 i ens ha donat $\bar{X}_1 = 4.2$ no vol dir que el tamany mig de una família en la població de referència sigui 4.2, si no que aquest tamany mig real és aproximadament de 4.2.

2.3.2 La Mediana

Representa un valor central situat al mig de les observacions de la variable, de forma que el 50% de les observacions són iguals o superiors a aquest valor i el 50% són iguals o inferiors.

Així, donat un conjunt ordenat d'observacions, la *mitjana* (M) és el valor que és més gran que no més de la meitat de les observacions i, a la vegada, més petit que no més de la meitat de les observacions.

El càlcul de la *mediana* es fa de forma diferent depenent de si el tamany de la mostra (n) és un número parell o senar.

Exemple 2.3.3 ***[Número parell observacions]*** *Si considerem les observacions de la variable $\bar{X}_1$ tenim*

$$\{x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{110}\} = \{2, 3, 4, 5, 8, 2, 4, 5, 7, 2\}$$

el primer que caldria fer és ordenar-les de més petita més gran:

$$[x_{11}, x_{16}, x_{1,10}, x_{12}, x_{13}, x_{17}, x_{14}, x_{18}, x_{19}, x_{15}\} = \{2, 2, 2, 3, 4, 5, 5, 5, 7, 8\}$$

En aquest cas, donat que el número d'observacions és parell, cap d'elles està just al mig de la llista (què seria la mediana). Notar que tant el $x_{13} = 4$ com $x_{17} = 5$ compleixen la condició de mediana:

Pel que fa a $x_{13} = 4$

- *"El 50% de les observacions (és a dir, 5 observacions) són més grans o iguals a x_{13}"*
  Efectivament, les 5 observacions $\{x_{17}, x_{14}, x_{18}, x_{19}, x_{15}\}$són més grans o iguals que x_{13}
- *"El 50% de les observacions (és a dir, 5 observacions) són més petites o iguals a x_{13}"*
  Efectivament, les 5 observacions $\{x_{11}, x_{16}, x_{1,10}, x_{12}, x_{13}\}$són més grans o iguals que x_{13}
  Per tant, x_{13} compleix la condició per a ser mediana *de les observacions de la mostra*

Pel que fa a $x_{17} = 5$

- *"El 50% de les observacions (és a dir, 5 observacions) són més grans o iguals a x_{15}"*
  Efectivament, les 5 observacions $\{x_{17}, x_{14}, x_{18}, x_{19}, x_{15}\}$són més grans o iguals que x_{17}
- *"El 50% de les observacions (és a dir, 5 observacions) són més petites o iguals a x_{17}"*
  Efectivament, les 5 observacions $\{x_{11}, x_{16}, x_{1,10}, x_{12}, x_{13}\}$són més grans o iguals que

Per tant, x_{17} compleix la condició per a ser mediana *de les observacions de la mostra*

En aquests casos hi ha dues opcions que s'accepten com a igual de vàlides:

1. *Les* medianes *són $x_{13} = 4$ i $x_{17} = 5$*
2. *La mediana és la* mitjana *d'aquests dos valors*

$$M = \frac{4+5}{2} = 4.5$$

Exemple 2.3.4 ***[Número senar de dades]*** *Considerar les observacions següents corresponents a una mostra de tamany $n = 7$ obtingudes d'una determinada població amb respecte de la variable X*

$$\{x_1, x_2, \ldots, x_7\} = \{3, 1, 4, 3, 2, 5, 1\}$$

Si les ordenem de més petita a més gran tindrem:

$$\{x_2, x_7, x_5, x_1, x_4, x_3, x_6\} = \{1, 1, 2, 3, 3, 4, 5\}$$

En aquesta cas està clar que l'observació que està al "mig" és $x_1 = 3$.

Si la informació que tenim és en forma de "taula de freqüències" (no disposem de les dades en brut), aleshores la *mediana* es pot trobar fent servir la *Freqüència Absoluta Acumulada* de la següent manera:

1. Primer determinen per quin valor (que denotarem y_m) es compleix que

$$N_{m-1} \leq \frac{n}{2} < N_m$$

2. Aleshores verifiquem el següent:

   (a) si $N_{m-1} < \frac{n}{2}$ aleshores direm que la *mediana* és $M = y_m$

   (b) Si $N_{m-1} = \frac{n}{2}$ aleshores tant el valor $y_m$com el valor y_{m-1} són *medianes* o bé, de la mateixa manera que hem fet abans en el cas del número parell de dades, podem dir que la *mediana* és la mitjana aritmètica d'aquests dos valors

$$M = \frac{y_m + y_{m-1}}{2}$$

Exemple 2.3.5 *En el cas de la variable X_1 tenim que $n = 10$ i, per tant, $\frac{n}{2} = 5$. Així, la mediana està entre els valors*

	Valor de X_1	*Freq. Abs. Acumulada* N_i
Y_2	*3*	4
Y_3	*4*	6

Per tant, com que,

$$4 < \frac{n}{2} < 6$$

podem concloure que la mediana és $Y_3 = 4$

Si es tracta de una variable contínua, la *mediana* s'obté de la mateixa manera, però mirant els *Intervals de classe* en comptes dels *valors*. En aquest cas es parla de la *classe mediana*. Per exemple, en el cas de la variable X_3, la *mediana* està a la *classe* c_2, $[844.93, 1162.69)$.

2.3.3 La Moda

La *moda* correspon al valor (o valors) que és més freqüent entre els què estan al seu voltant. La *moda absoluta* el valor que més es repeteix de tota la mostra, el més típic.

Així, pot existir més d'una moda, només una, o cap (si tots els valors śon igual de freqüents).

Formalment, *moda* és qualsevol valor y_q que compleix

$$n_{q-1} \leq n_q \quad \text{i} \quad n_q \geq n_{q+1}$$

Exemple 2.3.6 *En el cas de la variable X_1, tenim que hi ha 3 modes, $X_1 = 2$, $X_3 = 4$ i $X_4 = 5$. La* moda absoluta *és el valor $X_1 = 2$*

Si es tracta de una variable contínua, la *moda* s'obté de la mateixa manera, però mirant els *Intervals de classe* en comptes dels *valors*. En aquest cas es parla de la *classe m*odal. Per exemple, en el cas de la variable X_3, la *moda* està a les *classes* amb marques c_1, c_4, c_7, i c_8

2.4 Mesures de Dispersió

Són mesures (*estadístics*) que tracten de mesurar "com de diferents" són les observacions de la (les) variables(s) amb respecte d'alguna mesura de centralització, normalment la mitjana.

2.4.1 L'Error Quadràtic Mig

Si proposem el valor v com a *representant* de la mostra (valor central), està clar que estarem cometent errors

Observació	Valor central	Error
x_1	v	$e_1 = (x_1 - v)$
x_2	v	$e_2 = (x_2 - v)$
$\vdots$	$\vdots$	$\vdots$
x_n	v	$e_n = (x_n - v)$
	Error Total	$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(x_i - v)$

Però comptar d'aquesta manera

$$E.T.(v) = \sum_{i=1}^{n}(x_i - v)$$

l'*Error Total* comès en prendre v com a mesura representant de les observacions de la mostra pot portar a conclusions errònies. Efectivament, mentre alguns dels errors e_i poden ser positius (quan $x_i > e_i$), d'altres seran negatius (quan $x_i < e_i$). Aleshores, en fer la suma de tots els erros e_i els "positius" poden anul·lar-se amb els "negatius" i al final quedar que l'error total és zero o proper a zero quan això no és veritat.

Per tal d'evitar que errors positius s'anul·lin amb errors negatius, el que es fa és elevar tots els errors al quadrat (e_i^2) de manera que tots siguin "positius". Això és el que s'anomena *Error Quadràtic*. Si aquest *Error Quadràtic* el dividim pel número d'observacions (n), és a dir, fem el promig, tindrem l'*Error Quadràtic Mig* comès en prendre el valor v com a mesura central representant de les n observacions de la mostra.

$$E.Q.M.(v) = \frac{1}{n}\sum_{i=1}^{n}(x_i - v)^2$$

2.4.2 La Variància

Donat un conjunt de n observacions, $\{x_1, x_2, \ldots, x_n\}$, la *Variància mostral* és l'Error Quadràtic Mig comès quan agafem la mitjana aritmètica $\bar{X}$ com a mesura central

$$Varincia = E.Q.M.(\bar{X})$$

El símbol utilitzat per a denotar la variància mostral és, normalment, S^2. Per tant, donades les observacions $\{x_1, x_2, \ldots, x_n\}$, la *Variància mostral* (S^2) és:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

Per raons que s'estudiaran més endavant[5], moltes vegades es fa servir la següent fórmula quan es vol calcular la variància mostral

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

La variància medeix el *grau de dispersió* de les dades amb respecta de la mitjana $\bar{X}$. A igualtat de mitjanes, quant més gran sigui la variància, més dispersió hi haurà entre les observacions.

Exemple 2.4.1 *En el cas de la variable* X_3 *(Ingressos mensuals), donat que* $\bar{X}_1 = 1285.45$*, tindrem que:*

$$\begin{aligned} S_1^2 &= \frac{1}{10}((592.18-1285.45)^2 + (743.83-1285.45)^2 + \cdots + (707.72-1285.45)^2) = 742678.8 \\ \hat{\sigma}_1^2 &= \frac{1}{9}((592.18-1285.45)^2 + (743.83-1285.45)^2 + \cdots + (707.72-1285.45)^2) = 825199.04 \end{aligned}$$

Una propietat important de la *variància* és que, per a qualsevol mesura central v es compleix que

$$S^2 = E.Q.M.(\bar{X}) \leq E.Q.M.(v)$$

És a dir, la mitjana mostral $\bar{X}$ és el valor central que **minimitza** l'error quadràtic mig.

2.4.3 La Desviació Típica

Donat un conjunt de n observacions, $\{x_1, x_2, \ldots, x_n\}$, a *Desviació Típica* (Desviació Típica, Desviació Estàndard, Error Típic i Error Estàndard són tots sinònims), què denotem amb S, és l'arrel quadrada de la variància

$$S = \sqrt{S^2}$$

El que la Desviació Típica fa amb respecte de la variància és "retornar" les unitats de mesura (que s'havien elevat al quadrat en calcular la variància) a la seva escala inicial.

Si en comptes de fer servir la fórmula S^2fem servir la $\hat{\sigma}^2$, aleshores la Desviació Típica és denota amb $\hat{\sigma}$. Així doncs,

$$\begin{aligned} S &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2} = \sqrt{S^2} \\ \hat{\sigma} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2} = \sqrt{\hat{\sigma}^2} \end{aligned}$$

Exemple 2.4.2 *En el cas de la variable* X_3 *(Ingressos mensuals), tindrem que:*

$$\begin{aligned} S_1 &= \sqrt{S^2} = \sqrt{742678.8} = 861.79 \\ \hat{\sigma}_1 &= \sqrt{\hat{\sigma}_1^2} = \sqrt{825199.04} = 908.4 \end{aligned}$$

[5] La fórmula proposada té *biaix*, com es veurà a Estadística II

2.4.4 El Coeficient de Variació

Imaginem (situació *a*) que tenim dues observacions mesurades en Kg. $\{100, 200\}$. Clarament, la mitjana és

$$\bar{X}_a = \frac{100 + 200}{2} = 150$$

i la variància i desviació típica són

$$\begin{aligned} S_a^2 &= \frac{(100 - 150)^2 + (200 - 150)^2}{2} = 2500 \\ S_a &= \sqrt{2500} = 50 \end{aligned}$$

Suposem ara (situació *b)* que en comptes de donar-nos la informació en Kg ens la donen en grams $\{100000, 200000\}$. Ara la mitjana és

$$\bar{X}_b = \frac{100000 + 200000}{2} = 150000$$

i la variància i desviació típica són

$$\begin{aligned} S_b^2 &= \frac{(100000 - 150000)^2 + (200000 - 150000)^2}{2} = 2500000000 \\ S_b &= \sqrt{2500} = 50000 \end{aligned}$$

Donat que S_b és molt més gran que S_a, la primera temptació seria concloure que

> "Les observacions de la situació *b* són <u>molt més disperses</u> que les de la situació *a* donat que $S_b > S_a$"

i això és clarament fals donat que les dues situacions corresponen a les mateixes observacions, nomes que mesurades en <u>unitats diferents</u>.

Per tal d'evitar aquest tipus d'error d'interpretació, sovint es calcula l'anomenat *Coeficient de Variació*, V, que tracta de mesurar la dispersió de cada serie d'observacions però de forma relativa a les seves unitats de mesura (o la seva magnitud).

$$V = \frac{S}{\bar{X}}$$

Així, tornem a mirar l'exemple anterior trobarem que:

$$\begin{aligned} V_a &= \frac{S_a}{\bar{X}_a} = \frac{50}{150} = \frac{2}{3} \\ V_b &= \frac{S_b}{\bar{X}_b} = \frac{50000}{150000} = \frac{2}{3} \end{aligned}$$

Com veiem, el *Coeficient de Variació* és el mateix.

Capítol 3

Teoria de la Probabilitat.

La *Teoria de la Probabilitat* es un conjunt d'eines per a tractar i estudiar fenòmens *aleatoris* o *estocàstics*. Qualsevol experiment o fenòmen que satisfaci els tres requisits següents és un experiment aleatori.

1. Es coneixen prèviament els resultats possibles de l'experiment,
2. És impossible predir el resultat abans de realitzar l'experiment,
3. En successives repeticions de l'experiment amb les mateixes condicions inicials es donen resultats diferents.

L'esquema bàsic de la Teoria moderna de la Probabilitat està caracteritzat per tres conceptes:

- *L'Espai mostral* Ω,
- *La* σ*-Àlgebra d'esdeveniments* $\mathcal{A}$, i
- *La Funció de Probabilitat* sobre els esdeveniments P.

3.1 Conjunt de resultats possibles o Espai mostral (Ω).

La col·lecció de tots els resultats possibles d'un experiment forma el Conjunt de Resultats o **Espai mostral** de l'experiment.

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$$

Exemple 3.1.1 *Llençar un dau*

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Aquest és un espai mostral ***finit****.*

Exemple 3.1.2 *Llençar una moneda fins que surti una cara*

$$\Omega = \{c, +c, ++c, +++c, ++++c, +++++c, \ldots\}$$

Aquest és un espai mostral ***infinit numerable****.*

Exemple 3.1.3 *Triar un número entre 0 i 1 a l'atzar*

$$\Omega = \{x \in \Re / x \in [0,1]\}$$

Aquest és un espai mostral infinit no numerable o ***continu****.*

3.2 Esdeveniments i σ-Àlgebra d'esdeveniments ($\mathcal{A}$).

3.2.1 Esdeveniments.

Notar que el tipus de qüestions que ens poden interessar d'un experiment no sempre estan recollides directament a l'espai mostral.

Exemple 3.2.1 *Llençar un dau. La qüestió "que surti un número parell" no és un element d'* Ω*, és una combinació d'elements d'* Ω*.*

Definim aleshores el concepte de **esdeveniment**, que és més ampli que el de **resultat** (recordar que un resultat és un element de l'espai mostral).

Definició 3.2.2 *Donat un experiment aleatori, qualsevol cosa que ens pugui interessar és un esdeveniment.*

Exemple 3.2.3 *Llançar un dau.* $A =$*"El resultat és un número parell"*

Notar que qualsevol resultat (qualsevol element de l'espai mostral) és un esdeveniment (que s'anomena **esdeveniment elemental**), però **no** qualsevol esdeveniment és un resultat. **Resultat** i **esdeveniment** són conceptes diferents. Un esdeveniment és més "ric" que un resultat.

Una definició més formal (més matemàtica) d'esdeveniment és la següent:

Definició 3.2.4 *Donat un experiment i el seu espai mostral* Ω*, un* ***esdeveniment*** *és qualsevol subconjunt d'* Ω*.*

Exemple 3.2.5 *Llençar un dau. Els quatre conjunts que apareixen a continuació corresponen als esdeveniments "Surt un número parell", "Surt menys de 3", "No surt 6" i "Surt 1" respectivament (notar que l'esdeveniment "Surt 1" és un esdeveniment elemental):*

$$A = \{2,4,6\}$$

$$B = \{1,2\}$$

$$C = \{1,2,3,4,5\}$$

$$D = \{1\}$$

3.2.2 σ-Àlgebra d'esdeveniments ($\mathcal{A}$).

Definició 3.2.6 *Donat un experiment aleatori, el conjunt de tots els esdeveniments, és a dir, el conjunt de tots els subconjunts d' Ω s'anomena σ-**Àlgebra d'esdeveniments** i es denota amb la lletra $\mathcal{A}$.*

Exemple 3.2.7 *Llançar un dau*

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$\mathcal{A} = \{\Omega, , \{2, 4, 6\}, \{1, 2\}, \{1, 2, 3, 4, 5\}, \{1\}, \ldots, etc, \ldots\}$$

és a dir

$$\mathcal{A} = \{A/A \subseteq \Omega\}$$

En el llenguatge de Teoria de Conjunts, $\mathcal{A}$ és el **conjunt de parts d'** Ω o **conjunt potència d'** Ω, $\mathcal{P}(\Omega)$.

Definició 3.2.8 *Direm que un esdeveniment **ha ocorregut** o **s'ha donat** si el resultat de l'experiment així ho indica, és a dir, si el resultat de l'experiment ha estat ω_j, l'esdeveniment A **ha ocorregut** o **s'ha donat** si $\omega_j \in A$.*

Exemple 3.2.9 *En l'exemple 2.2.5, si el resultat obtingut és 4, tindrem que*

A s'ha donat perquè $4 \in A$.

B no s'ha donat perquè $4 \notin B$.

C s'ha donat perquè $4 \in C$.

D no s'ha donat perquè $4 \notin D$.

3.2.3 Tipus d'esdeveniments.

Siguin A i $B \in \mathcal{A}$ dos esdeveniments qualsevol

(i) A i B són **diferents** si hi ha un resultat possible que faci que es doni un esdeveniment i l'altre no. Exemple: $A = \{2, 4, 6\}$ i $B = \{1, 2\}$

(ii) A i B són **iguals** si no són diferents. Exemple: $A = \{2, 4, 6\}$ i $B = \{2, 4, 6\}$

(iii) A i B són **contraris o complementaris** si quan un es dona l'altre no es dona i quan un no es dona l'altre si que es dona. Exemple: $A = \{2, 4, 6\}$ i $B = \{1, 3, 5\}$

(iv) A i B són **mútuament excloents, disjunts o incompatibles** si o es poden donar els dos a la vegada, és a dir, si quan un es dona l'altre no es dona. Exemple: $A = \{2, 4, 6\}$ i $B = \{1, 3\}$

(v) A **està inclòs** en B si sempre que es dona A també es dona B. Exemple: $A = \{2, 4, 6\}$ i $B = \{2, 3, 4, 6\}$

3.2.4 Operacions amb esdeveniments

(i) Complementació.- $\bar{A}$ és l'esdeveniment que es dona si i només si no es dona A.

(ii) Unió.- $A \cup B$ és l'esdeveniment que es dona si i només si es dona A o es dona B o tots dos a la vegada.

(iii) Intersecció.- $A \cap B$ és l'esdeveniment que es dona si i només si es donen A i B a la vegada.

(iv) Diferència.- $A \backslash B$ és l'esdeveniment que es dona si i només si es dona A i no es dona B.

3.2.5 Propietats dels esdeveniments.

(i) Distributiva.-

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

(ii) Lleis de Morgan.-

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

3.3 Probabilitat.

El que ara ens interessa es trobar una funció (o formula) que assigni a cada esdeveniment la probabilitat de que aquest esdeveniment es doni.

A més de la definició axiomàtica de probabilitat que hi ha a continuació, també són d'interès els enfocs freqüencialista i clàssic o de Laplace.

3.3.1 Definició axiomàtica de probabilitat.

Definició 3.3.1 *Sigui Ω un espai mostral i $\mathcal{A}$ la seva σ-àlgebra d'esdeveniments associada. Una **probabilitat** és una funció $P : \mathcal{A} \rightarrow \Re$ que satisfà els tres axiomes següents:*

(i) $\forall A \in \mathcal{A} \quad P(A) \geq 0$.

(ii) $P(\Omega) = 1$.

(iii) Si $A, B \in \mathcal{A}$ i $A \cap B = \emptyset$, aleshores $P(A \cup B) = P(A) + P(B)$.

Exemple 3.3.2 *La definició clàssica (o de Laplace) de probabilitat*

$$P(A) = \frac{n(A)}{n(\Omega)}$$

3.3.2 Propietats de les probabilitats.

Les 6 propietats que a continuació es relacionen es deriven dels 3 axiomes de les probabilitats.

(1) $P(\emptyset) = 0$

(2) $P(\bar{A}) = 1 - P(A)$

(3) $A \subseteq B \Rightarrow P(A) \leq P(B)$

(4) $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$ si $A_i \cap A_j = \emptyset$ per a $i \neq j$

(5) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(6) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

3.3.3 Independència Estocàstica.

Definició 3.3.3 *Dos esdeveniments $A, B \in \mathcal{A}$ són **estocàsticament independents** o, simplement, **independents** si:*

$$P(A \cap B) = P(A) \cdot P(B)$$

Definició 3.3.4 *Una col·lecció de n esdeveniments $A_1, \ldots, A_n \in \mathcal{A}$ són **estocàsticament independents** o, simplement, **independents** si:*

$$P(A \cap A_2 \cap \ldots \cap A_n) = P(A) \cdot P(A_2) \cdot \ldots \cdot P(A_n)$$

3.3.4 Probabilitat Condicionada.

Definició 3.3.5 *Sigui Ω l'espai mostral associat a un experiment aleatori i sigui $\mathcal{A}$ la corresponent σ-àlgebra d'esdeveniments. Sigui $B \in \mathcal{A}$ un esdeveniment qualsevol tal que $P(B) > 0$. Aleshores, per a qualsevol esdeveniment $A \in \mathcal{A}$, anomenarem **probabilitat de A condicionada a B** a l'expressió:*

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

Exemple 3.3.6 *Llençar un dau.*

$$A = \{6\}$$
$$B = \{2, 4, 6\}$$

Sabem que $P(A) = \frac{1}{6}$ i que

$$P(B) = P(\{2,4,6\}) = P(\{2\} \cup \{4\} \cup \{6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{3}{6} = \frac{1}{2}$$

Aleshores:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{6\})}{P(\{2,4,6\})} = \frac{1/6}{3/6} = \frac{1}{3}$$

és la probabilitat de que surti 6 donat que ha sortit un número parell.

3.3.5 Teorema de la probabilitat total.

Sigui $\mathcal{A}$ una σ-àlgebra d'esdeveniments i $A_1, A_2, \ldots, A_n$ una col·lecció d'esdeveniments qualsevol que satisfà:

(i) $A_i \cap A_j = \emptyset$ per a $i \neq j$.

(ii) $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$

> **((i)** i (ii) indiquen que $A_1, A_2, \ldots, A_n$ formen una **partició** d'Ω)

(iii) $P(A_i) > 0 \quad \forall A_i \quad (i = 1, \ldots, n)$

Aleshores, per a qualsevol $B \in \mathcal{A}$:

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

3.3.6 Teorema de Bayes

Sigui $\mathcal{A}$ una σ-àlgebra d'esdeveniments i $A_1, A_2, \ldots, A_n$ una col·lecció d'esdeveniments qualsevol que satisfà:

(i) $A_i \cap A_j = \emptyset$ per a $i \neq j$.

(ii) $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$

> **((i)** i (ii) indiquen que $A_1, A_2, \ldots, A_n$ formen una **partició** d'Ω)

(iii) $P(A_i) > 0 \quad \forall A_i \quad (i = 1, \ldots, n)$

Aleshores, per a qualsevol $B \in \mathcal{A}$:

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

3.3.7 Combinatòria: càlcul de probabilitats.

(i) [SI TENIM EN COMPTE L'ORDRE]

Permutacions: De n objectes, fer llistes ordenades de r objectes (sense reposició)

$$P(n, r) = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+1) = \frac{n!}{(n-r)!}$$

Exemple 3.3.7 *Si en una classe hi ha 5 nenes, de quantes maneres podem formar cues de 3 nens ?*

$$P(5, 3) = 5 \cdot 4 \cdot 3 = 60$$

Si les llistes es fan amb reposició, el resultat de les permutacions és:

$$P(n,r) = n^r$$

(ii) [SI NO TENIM EN COMPTE L'ORDRE]

Combinacions: De n objectes, fer grups de r objectes (sense reposició)

$$C(n,r) = \binom{n}{r} = \frac{n!}{r! \cdot (n-r)!}$$

Exemple 3.3.8 *Si en una classe hi ha 5 nenes, de quantes maneres podem formar grups de 3 nens ?*

$$C(5,3) = \binom{5}{3} = \frac{5!}{3! \cdot 2!} = 10$$

(iii) [SI TENIM EN COMPTE L'ORDRE I HI HA ELEMENTS QUE ES REPETEIXEN]

Permutacions amb repetició: De n objectes, dels quals hi ha n_1 objectes iguals, n_2 objectes iguals, fins a n_k objectes iguals $(n_1 + \ldots + n_k = n)$, fer llistes ordenades de r objectes (sense reposició)

$$PR(n,r) = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$$

Exemple 3.3.9 *De quantes maneres podem ordenar 6 llibres si hi ha 3 de vermells, 3 de grocs i 3 de negres ?*

$$PR(6;3,3,3) = \frac{6!}{3! \cdot 3! \cdot 3!} = 20$$

Capítol 4

Variables aleatòries discretes unidimensionals.

4.1 Introducció.

4.1.1 Definició de variable aleatòria.

Definició 4.1.1 *Una **variable aleatòria** X és una funció que associa a cada possible resultat (ω) de l'espai mostral (Ω) un número real.*

$$\begin{array}{rcl} X:\Omega & \rightarrow & \Re \\ \omega & \rightarrow & X(\omega) \end{array}$$

4.1.2 Valors que pren la variable aleatòria.

Donat el conjunt Ω de possibles resultats de l'experiment, és fàcil trobar el conjunt de valors que pren la variable aleatòria X:

$$\Omega = \{\omega_1, \omega_2, \ldots\}$$

$$X \in \{x_1, x_2, \ldots\}$$

Definició 4.1.2 *Una variable aleatòria X es distribueix de forma **discreta** (o té una distribució discreta) si nomes hi ha un conjunt finit o numerable de valors que pren amb probabilitat estrictament positiva.*

Exemple 4.1.3 *Llencem dos daus*

$$\Omega = \{(1,1),(1,2),\ldots,(2,1),(2,2),\ldots,(6,5),(6,6)\}$$

Sigui X la variable "Suma dels dos daus". És a dir:

$$X(1,1) = 2, X(1,2) = 3, X(1,3) = 4, \ldots, etc.$$

Clarament, en aquest cas:

$$X \in \{2,3,4,5,6,7,8,9,10,11,12\} \rightarrow \textit{valors que pren } X$$

4.1.3 Probabilitat de cada un dels valors que pren X.

Un cop tenim la llista de valors que pren la variable X, és fàcil trobar la probabilitat $p(x)$ de cada un d'aquests valors a partir de les probabilitats (que ja coneixem o podem calcular) dels elements d' Ω que corresponen a cada un dels valors de X. Concretament:

$$p(x) = P(X = x) = P(\{\omega_i \in \Omega / X(\omega_i) = x\})$$

Exemple 4.1.4 *En el cas de l'exemple 2.1.3*

$$p(2) = P(X = 2) = P(\{\omega_i \in \Omega / X(\omega_i) = 2\}) = P(\{(1,1)\}) = \frac{1}{36}$$

$$p(3) = P(X = 3) = P(\{\omega_i \in \Omega / X(\omega_i) = 3\}) = P(\{(1,2),(2,1)\}) = \frac{2}{36}$$

$$\vdots$$

$$p(7) = P(X = 7) = P(\{\omega_i \in \Omega / X(\omega_i) = 7\}) =$$

$$= P(\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}) = \frac{6}{36}$$

etc ...

4.1.4 Funció de massa de probabilitat de una variable aleatòria.

Un cop tenim la llista de valors que pren la variable aleatòria X i les probabilitats d'aquests valors, ja podem oblidar-nos de l'experiment original a partir del qual hem construït la variable. Definim a continuació la **Funció de massa de probabilitat**, que descriu completament la variable X (ens dona tota la informació que necessitem sobre X).

Definició 4.1.5 *Donada una variable aleatòria X i les probabilitats dels seus valors, la* ***Funció de massa de probabilitat*** *f_X de X es defineix com a:*

$$f_X : \Re \rightarrow \Re$$

$$\text{tal que}$$

$$f_X(x) = p(x) \quad \text{si} \quad x \quad \text{es un dels valors que pren} X$$

$$f_X(x) = 0 \quad \text{altrament}$$

Notar que nomes amb aquesta funció ja tenim tota la informació que ens cal sobre la variable, no ens fa falta ni la llista de valors que la variable pren ni els resultats possibles de l'experiment. Aquesta funció **caracteritza** completament la variable.

Exemple 4.1.6 *La funció*

$$f_X(x) = \frac{6 - |7 - x|}{36} \quad si \quad 2 \leq x \leq 12$$

$$f_X(x) = 0 \quad altrament$$

és la funció de massa de probabilitat de l'exemple 2.1.4

4.2 Moments i generatrius de moments.

4.2.1 Esperança.

Definició 4.2.1 ***L'esperança*** *(o valor esperat) de una variable aleatòria X es defineix com a:*

$$E(X) = \sum_x p(x) \cdot x$$

L'esperança té les següents propietats:

(i) $E(a) = a$

(ii) $E(aX) = aE(x)$

(iii) $E(X+Y) = E(X) + E(Y)$

(iv) $E(XY) = E(X) \cdot E(Y)$ si X i Y són independents[1]

4.2.2 Variància.

Definició 4.2.2 *La* ***variància*** *de una variable aleatòria X es defineix com a:*

$$V(X) = \sum_x p(x)(x - E(X))^2 = E(X - E(X))^2$$

La variància té les següents propietats:

(i) $V(a) = 0$

(ii) $V(aX) = a^2 V(X)$

(iii) $V(X+Y) = V(X) + V(Y)$ si X i Y són independents

(iv) $V(X+a) = V(X)$

(v) $V(X) \geq 0$

Una fórmula diferent i més pràctica per a calcular la variància és:

$$V(X) = E(X^2) - (E(X))^2$$

[1] Dues variables aleatòries X i Y són independents si $P(\{X = x\} \cap \{Y = y\}) = p(x) \cdot p(y)$

4.2.3 Coeficient d'asimetria.

Definició 4.2.3 *El **coeficient d'asimetria** de la variable aleatòria X es defineix com a:*

$$CA(X) = \frac{E(X - E(X))^3}{\sigma^3}$$

4.2.4 Funció generatriu de moments.

Definició 4.2.4 *La **Funció generatriu de moments** de la variable aleatòria X es defineix com a:*

$$\Psi(t) = E(e^{tX}) = \sum_x p(x)e^{tx}$$

D'aquesta funció és important recordar que:

$$\Psi'(0) = E(X), \Psi''(0) = E(X^2), \ldots, \Psi^n(0) = E(X^n)$$

4.2.5 Funció generatriu de cumulants.

Definició 4.2.5 *La **Funció generatriu de cumulants** de la variable aleatòria X es defineix com a:*

$$a(t) = \ln \Psi(t)$$

D'aquesta funció és important recordar que:

$$a'(0) = E(X), a''(0) = V(X), \ldots, a^n(0) = E((X - E(X))^n)$$

4.3 Principals distribucions discretes.

4.3.1 Distribució de Bernouilli.

Definició 4.3.1 *Una variable aleatòria X es distribueix segons una distribució de Bernouilli si nomes pot prendre els valors 1 i 0 amb probabilitats p i $1 - p$ respectivament.*

En aquest cas tenim:

- $E(X) = p$
- $V(X) = p(1 - p)$
- $\Psi(t) = 1 - p(1 - e^t)$

Aquesta distribució correspon al cas senzill d'un experiment que pot resultar en 1 (èxit) amb probabilitat p ó 0 amb probabilitat $1 - p$

4.3.2 Distribució Binomial

Definició 4.3.2 *Una variable aleatòria X es distribueix segons una binomial amb paràmetres n i p ($X \sim B(n,p)$) si nomes pot prendre els valors 0,1,2,... amb probabilitats donades per l'expressió:*

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

En aquest cas tenim:

- $E(X) = np$
- $V(X) = np(1-p)$
- $\Psi(t) = (pe^t + (1-p))^n$

Aquesta variable X compta el número d'èxits en n repeticions d'un experiment que pot resultar en èxit amb probabilitat p o no resultar en èxit amb probabilitat $(1-p)$. Per exemple, llençar una moneda 10 vegades ($n = 10$) i comptar el número de cares ($p = 0,5$).

4.3.3 Distribució Geomètrica.

Definició 4.3.3 *Una variable aleatòria X es distribueix segons una distribució **geomètrica** si nomes pot prendre els valors 1,2,... amb probabilitats donades per l'expressió:*

$$P(X = x) = p(1-p)^{x-1}$$

En aquest cas tenim:

- $E(X) = \frac{1}{p}$
- $V(X) = \frac{(1-p)}{p^2}$
- $\Psi(t) = \frac{p}{(1-p)} \cdot \frac{(1-p)e^t}{(1-(1-p)e^t)}$

Aquesta variable X compta el número de vegades que cal repetir un experiment que pot resultar en èxit amb probabilitat p per tal d'obtenir el primer èxit. Per exemple, quantes vegades hem de llençar una moneda fins a obtenir la primera cara.

4.3.4 Distribució de Poisson.

Definició 4.3.4 *Una variable aleatòria X es distribueix segons una distribució de* ***Poisson*** *amb paràmetre λ ($X \sim P(\lambda)$)si nomes pot prendre els valors 0,1,2,. . . amb probabilitats donades per l'expressió:*

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

En aquest cas tenim:

- $E(X) = \lambda$
- $V(X) = \lambda$
- $\Psi(t) = e^{\lambda(e^t-1)}$

Aquesta variable X és com una Binomial però sense saber quantes repeticions hi ha de l'experiment (o bé amb infinites repeticions). És a dir, X compta quantes vegades es donarà "èxit" en un determinat interval de temps. Per exemple, quantes trucades tindrem un divendres a la tarda si la mitjana és de 1,3 trucades ($\lambda = 1,3$).

Capítol 5

Variables aleatòries discretes bidimensionals

Una variable aleatòria bidimensional no és res més (ni res menys) que dues variables aleatòries unidimensionals considerades conjuntament. Per tant, tot el que s'ha fet amb variables unidimensionals serà la base de les bidimensionals.

A continuació es resumeixen alguns dels conceptes que es poden definir a partir de l'utilització de variables aleatòries bidimensionals.

5.1 Distribució conjunta

Correspon a la probabilitat de que les dues variables tinguin conjuntament els valors que se indiquen (funció de massa de probabilitat conjunta).

$$f_{XY}(x,y) = P(\{X = x\} \cap \{Y = y\})$$

5.2 Distribucions marginals

Corresponen a la funció de massa de probabilitat de cada una de les dues variables aïlladament.

$$f_X(x) = P(\{X = x\}) = \sum_y f_{XY}(x,y)$$

$$f_Y(y) = P(\{Y = y\}) = \sum_x f_{XY}(x,y)$$

5.3 Probabilitat i esperança condicionades.

Aquests dos conceptes es corresponen als conceptes de probabilitat i esperança que s'han vist abans, però incorporant informació sobre el valor obtingut per una de les variables.

Probabilitat de Y condicionada a que $X = x$

$$f_{Y/X=x}(y) = \frac{f_{XY}(x,y)}{f_X(x)}$$

Esperança de Y condicionada a que $X = x$

$$E(Y/X) = \sum_y f_{Y/X}(y)y$$

5.4 Covariància i Coeficient de correlació.

Aquests dos estadístics donen informació sobre el grau de dependència entre les dues variables.

Covariància

$$COV(X,Y) = E(X - E(X))(Y - E(Y)) = E(XY) - E(X)E(Y)$$

Coeficient de correlació

$$\rho = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$$

Dues variables independents tenen una covariància de zero. En cas de que no siguin independents, tindrem que:

$$V(X+Y) = V(X) + V(Y) + 2COV(X,Y)$$

$$V(X-Y) = V(X) + V(Y) - 2COV(X,Y)$$

5.5 Desigualtat de Chebyschev.

Aquesta desigualtat diu que:

$$P[|X - E(X)| \leq k\sigma] \geq 1 - \frac{1}{k^2}$$

5.6 Esperança iterada

La *llei de l'esperança iterada* diu que:

$$E(E(Y/X)) = E(Y)$$

Capítol 6

Variables aleatòries contínues unidimensionals.

Definició 6.0.1 *Una variable aleatòria X és **contínua** si el seu rang, és a dir, el conjunt de valors que la variable pren amb probabilitat positiva, és un conjunt infinit no numerable.*

Exemple 6.0.2 *Triar un número aleatòriament entre 0 i 1.*

6.1 Funció de densitat de probabilitat (f.d.p).

Definició 6.1.1 *Donada una variable aleatòria contínua X, la **Funció de densitat de probabilitat** f_X de X es defineix com a:*

$$f_X : \Re \to \Re$$
$$\text{tal que}$$
$$p[a \leq X \leq b] = \int_a^b f_x(x)dx$$

Notar que, al igual que passava en el cas de les distribuciions discretes, la funció de densitat de probabilitat **caracteritza** completament la variable, és a dir, ens dona tona la informació que cal sobre la variable X.

Exemple 6.1.2 *Sigui X una variable aleatòria amb funció de densitat donada per:*

$$f_X(x) = \begin{cases} 2x & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$$

Aleshores,

$$p[0 \leq X \leq 0.5] = \int_0^{0.5} 2xdx =$$
$$= \left. \frac{2x^2}{2} \right|_0^{0.5} = 0.5^2 = 0.25$$

6.2 Moments i generatriu de moments de variables aleatòries contínues.

6.2.1 Esperança.

Definició 6.2.1 ***L'esperança*** *(o valor esperat) de una variable aleatòria contínua X es defineix com a:*

$$E(X) = \int_{x \in \Re} f_X(x) \cdot x dx$$

L'esperança té les mateixes propietats que en el cas discret:

(i) $E(a) = a$

(ii) $E(aX) = aE(x)$

(iii) $E(X+Y) = E(X) + E(Y)$

(iv) $E(XY) = E(X) \cdot E(Y)$ si X i Y són independents[1]

Recordar que l'esperança es pot calcular no només de la variable aleatòria X sinó també de qualsevol transformació contínua g. Així:

$$E(g(X)) = \int_{x \in \Re} f_X(x) \cdot g(x) dx$$

6.2.2 Variància.

Definició 6.2.2 *La **variància** de una variable aleatòria contínua X es defineix com a:*

$$V(X) = \int_{x \in \Re} f_X(x)(x - E(X))^2 dx = E(X - E(X))^2$$

La variància té les mateixes propietats que en el cas discret:

(i) $V(a) = 0$

(ii) $V(aX) = a^2 V(X)$

(iii) $V(X+Y) = V(X) + V(Y)$ si X i Y són independents

(iv) $V(X+a) = V(X)$

(v) $V(X) \geq 0$

Una fórmula diferent i més pràctica per a calcular la variància és:

$$V(X) = E(X^2) - (E(X))^2$$

[1] Dues variables aleatòries X i Y són independents si $P(\{X = x\} \cap \{Y = y\}) = p(x) \cdot p(y)$

6.2.3 Coeficient d'asimetria.

Definició 6.2.3 *El **coeficient d'asimetria** de la variable aleatòria contínua X es defineix com a:*

$$CA(X) = \frac{E(X - E(X))^3}{\sigma^3}$$

6.2.4 Funció generatriu de moments.

Definició 6.2.4 *La **Funció generatriu de moments** de la variable aleatòria contínua X es defineix com a:*

$$\Psi(t) = E(e^{tX}) = \int_{x\in\Re} f_X(x) \cdot e^{tx} dx$$

D'aquesta funció és important recordar que:

$$\Psi'(0) = E(X), \Psi''(0) = E(X^2), \dots, \Psi^n(0) = E(X^n)$$

6.3 Funció de distribució.

Definició 6.3.1 *Donada una variable aleatòria contínua X, la **Funció de distribució** F_X de X es defineix com a:*

$$F_X : \Re \rightarrow \Re$$

$$\text{tal que}$$

$$F_X(x) = p[X \leq x] = \int_{-\infty}^{x} f_X(t)dt$$

La funció de distribució té les següents propietats:

(i) $0 \leq F_X(x) \leq 1$

(ii) $\lim_{x\to\infty} F_X(x) = 1; \quad \lim_{x\to-\infty} F_X(x) = 0$

(iii) $x_1 \leq x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$

(iv) $p[a \leq X \leq b] = F_X(b) - F_X(a)$

(v) $F'_X(x) = f_X(x)$ (propietat important).

Exemple 6.3.2 *Sigui X una variable aleatòria amb funció de densitat donada per:*

$$f_X(x) = \begin{cases} 2x & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$$

Aleshores, la seva funció de distribució $F_X(x)$ és:

$$F_X(x) = \int_0^x 2t dt = t^2\Big|_0^x = x^2$$

És a dir:

$$F_X(x) = \begin{cases} 0 & x \in (-\infty, 0) \\ x^2 & x \in [0,1) \\ 1 & x \in [1, \infty) \end{cases}$$

6.4 Percentils, quartils i mediana.

Definició 6.4.1 *El* ***percentil*** *p de la distribució d'una variable aleatòria X és aquell valor x_p tal que el p% de la probabilitat està a la seva esquerra, és a dir:*

$$p[X \leq x_p] = F_X(x_p) = \frac{p}{100}$$

Definició 6.4.2 *Els* ***quartils*** *primer (Q_1), segon (Q_2) i tercer (Q_3) són aquells valors que fan que el 25%, 50% i 75% respectivament de la probabilitat estigui a la seva esquerra, és a dir:*

$$\begin{aligned} p[X \leq Q_1] = F_X(Q_1) = \frac{25}{100} \quad &\Rightarrow \quad Q_1 = x_{25} \\ p[X \leq Q_2] = F_X(Q_2) = \frac{50}{100} \quad &\Rightarrow \quad Q_2 = x_{50} \\ p[X \leq Q_3] = F_X(Q_3) = \frac{75}{100} \quad &\Rightarrow \quad Q_3 = x_{75} \end{aligned}$$

Definició 6.4.3 *La* ***mediana*** *de la distribució d'una variable aleatòria contínua és aquell valor que fa que el 50% de la probabilitat estigui a la seva esquerra, és a dir, coincideix amb el segon quartil.*

$$\text{Mediana} = Q_2$$

Exemple 6.4.4 *En la distribució de l'exemple 4.3.1, tenim que:*

$$Q_1 = 0.5 \quad \textit{ja que } p[X \leq Q_1] = F_X(0.5) = 0.5^2 = \frac{25}{100}$$

$$Q_2 = 0.707 \quad \textit{ja que } p[X \leq Q_2] = F_X(0.707) = 0.707^2 = \frac{50}{100}$$

$$Q_3 = 0.866 \quad \textit{ja que } p[X \leq Q_3] = F_X(0.866) = 0.866^2 = \frac{75}{100}$$

6.5 Principals distribucions contínues.

6.5.1 Distribució uniforme.

Definició 6.5.1 *Una variable aleatòria X es distribueix segons una* ***uniforme*** *en el interval $[a, b]$ ($X \sim U[a, b]$) si la seva funció de densitat de probabilitat (f.d.p) és:*

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}$$

En aquest cas tenim:

- $E(X) = \frac{b+a}{2}$
- $V(X) = \frac{(b-a)^2}{12}$

6.5.2 Distribució exponencial.

Definició 6.5.2 *Una variable aleatòria X es distribueix segons una* ***exponencial*** *de paràmetre λ si la seva funció de densitat de probabilitat (f.d.p) és:*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \in [0, \infty) \\ 0 & x \notin [0, \infty) \end{cases}$$

En aquest cas tenim:

- $E(X) = \frac{1}{\lambda}$
- $V(X) = \frac{1}{\lambda^2}$

6.5.3 Distribució Normal.

Definició 6.5.3 *Una variable aleatòria X es distribueix segons una* ***Normal*** *de paràmetres μ i σ ($X \sim N(\mu, \sigma^2)$) si la seva funció de densitat de probabilitat (f.d.p) és:*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

En aquest cas tenim:

- $E(X) = \mu$
- $V(X) = \sigma^2$

6.5.4 Distribució Normal Estàndard (o tipificada).

Definició 6.5.4 *Una variable aleatòria X es distribueix segons una **Normal Estàndard** $(X \sim N(0,1))$ si la seva funció de densitat de probabilitat (f.d.p) és:*

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty$$

En aquest cas tenim:

- $E(X) = 0$
- $V(X) = 1$

Resultat: Sigui X una variable aleatòria normal $X \sim N(\mu, \sigma^2)$. Aleshores, la variable Z definida com segueix es distribueix com una normal estàndard:

$$Z = \frac{x-\mu}{\sigma} \sim N(0,1)$$

6.5.5 Distribució log normal.

Definició 6.5.5 *Una variable aleatòria X es distribueix segons una **log normal** de paràmetres μ i σ si la variable $Y = \ln X$ és una Normal de paràmetres μ i σ^2. La funció de densitat de probabilitat (f.d.p) d'una variable log normal és:*

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}(\frac{\ln x-\mu}{\sigma})^2} \qquad -\infty < x < \infty$$

En aquest cas tenim:

- $E(X) = e^{\mu+\frac{\sigma^2}{2}}$
- $V(X) = e^{\sigma^2}(e^{\sigma^2}-1)e^{2\mu}$

6.5.6 Distribució khi-quadrat amb n graus de llibertat.

Definició 6.5.6 *Una variable aleatòria X es distribueix segons una khi-**quadrat** amb n graus de llibertat $(X \sim \chi^2_n)$ si és una suma de n Normals estàndard $(X = X_1 + X_2 + \ldots + X_n, \quad X_i \sim N(0,1))$. La funció de densitat de probabilitat (f.d.p) d'una variable khi-quadrat és:*

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}(\frac{\ln x-\mu}{\sigma})^2} \qquad -\infty < x < \infty$$

En aquest cas tenim:

- $E(X) = n$
- $V(X) = 2n$

6.5.7 Distribució t de Student amb n graus de llibertat.

Definició 6.5.7 *Una variable aleatòria X es distribueix segons una t **de Student** amb n graus de llibertat $(X \sim t_n)$ si és el quocient de una Normal estàndard i l'arrel quadrada d'una khi-quadrat amb n graus de llibertat dividida per n ($X = \frac{Z}{\sqrt{(Y/n)}}$ with $Z \sim N(0,1)$ and $Y \sim \chi^2_n$). La funció de densitat de probabilitat (f.d.p) d'una variable t de Student és:*

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}(\frac{\ln x-\mu}{\sigma})^2} \qquad -\infty < x < \infty$$

En aquest cas tenim:

- $E(X) = 0$
- $V(X) = \frac{n}{n-2} \quad (n > 2)$

6.6 Aproximacions a la Normal de la Binomial i la Poisson.

Tant la distribució discreta *binomial* com la *Poisson* poden ser aproximades mitjançant l'us de la distribució *normal*. És a dir, podem calcular la probabilitat de que una variable que es distribueix com una *binomial* (o una *Poisson*) assoleixi uns determinats valors a través de l'us d'una variable *normal* (amb les avantatges de còmput que això implica en poder utilitzar les taules). Concretament:

$B(n,p)$	s'aproxima per una $Poisson(\lambda)$	si $np = \lambda \geq 1$ i $p < 0.1$
$B(n,p)$	s'aproxima per una $N(\mu,\sigma^2)$	si $np = \mu,\ npq = \sigma^2$ i $npq > 5$
$Poisson(\lambda)$	s'aproxima per una $N(\mu,\sigma^2)$	si $\lambda = \mu,\ \lambda = \sigma^2$ i $\lambda > 5$

6.7 Transformacions de variables aleatòries contínues.

Veurem ara com es distribueix (és a dir, quina és la funció de densitat) la variable aleatòria que resulta d'aplicar una *transformació contínua i monòtona*[2] $h(\cdot)$ a una variable aleatòria contínua la distribució (funció de densitat o de distribució) de la qual és coneguda. Per exemple, podem trobar quina és la funció de densitat d'una *exponencial* elevada al quadrat o de l'arrel quadrada d'una *normal*, etc.

En general, la variable aleatòria Y que resulta d'aplicar la *transformació contínua i monòtona* $h(\cdot)$ a la variable aleatòria X es denota amb:

$$Y = h(X)$$

Per exemple:

$$Y = X^2 \qquad \text{En aquest cas, } h(\cdot) = x^2$$

[2]*També es podria operar amb transformacions no monòtones, però les complicacions augmenten considerablement*

La funció de densitat $(f_Y(y))$ d'aquesta nova variable s'obté a partir de la formula:

$$f_Y(y) = f_X(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right|$$

Si convenim que a l'expressió $Y = h(X)$ podem "aïllar" X en funció de Y, obtindrem que $X = h^{-1}(Y)$.

Per tant, la fórmula anterior també pot expressar-se com:

$$f_Y(y) = f_X(h^{-1}(y)) \left| \frac{dx}{dy} \right|$$

Exemple 6.7.1 *Sigui X una variable aleatòria contínua distribuïda uniformement en l'interval* $[0, 1]$. *Sigui Y una altra variable donada per l'expressió:*

$$Y = -\frac{1}{\lambda} \ln(1 - X)$$

Trobar la funció de densitat de Y.

En aquest cas tenim que $h(x) = -\frac{1}{\lambda} \ln(1 - x)$. Per tant, $h^{-1}(y) = x = 1 - e^{-\lambda y}$. Aleshores, $\frac{dx}{dy} = \lambda e^{-\lambda y}$. Per tant, d'acord amb la fórmula,

$$f_Y(y) = f_X(h^{-1}(y)) \left| \frac{dx}{dy} \right| = 1\lambda e^{-\lambda y}$$

$$f_Y(y) = \lambda e^{-\lambda y}$$

És a dir, Y és una exponencial.

6.8 Convergència de variables aleatòries.

Es tracta ara d'estudiar quin serà el comportament i característiques principals si el que considerem no és una variable aleatòria aïllada sinó una col·lecció (com més gran millor) de variables estocàsticament independents però relacionades d'alguna forma. Dos són els resultats importants en aquest terreny. El primer, o *Teorema Central del Límit* estableix que la mitjana d'aquesta col·lecció de variables es distribuirà, sigui quina sigui la distribució de les variables en qüestió, com una variable *normal*. El segon resultat, o *Llei dels grans números*, especifica que la probabilitat de que aquesta mitjana prengui un valor diferent del valor esperat de cada una de les variables de la col·lecció és molt propera a zero, i serà tant més propera a zero quant més gran sigui la col·lecció de variables.

6.8.1 Teorema Central del Límit.

Theorem 6.8.1 *Siguin $\{X_1, \ldots, X_n\}$ una col·leció de variables aleatòries independents amb la mateixa esperança i variància $E(X_i) = \mu$, $V(X_i) = \sigma^2$. Sigui $\bar{X}$ la mitjana d'aquestes n variables, $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Aleshores, la variable*

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

segueix una distribució $N(0, 1)$

6.8.2 Llei dels grans números.

Theorem 6.8.2 *Siguin* $\{X_1, \ldots, X_n\}$ *una col·lecció de variables aleatòries independents amb la mateixa esperança i variància* $E(X_i) = \mu$, $V(X_i) = \sigma^2$. *Sigui* $\bar{X}$ *la mitjana d'aquestes* n *variables,* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. *Aleshores, per a qualsevol* $\epsilon > 0$,

$$p(|\bar{X} - \mu| > \epsilon) \to 0 \quad \text{quan } n \to \infty$$

El primer teorema diu que el que resultat de fer la mitjana de n variables aleatòries és una altra variables aleatòria amb la particularitat de que quantes més variables fem servir per a calcular la mitjana, més se sembla aquesta mitjana a una variable que es distribueix segons una *normal*. El segon resultat estableix que la probabilitat de que aquesta mitjana sigui diferent de l'esperança tendeix a zero a mesura que augmentem el número de variables utilitzades en el càlcul de la mitjana.

Capítol 7

Variables aleatòries contínues bidimensionals.

Bàsicament, els mateixos conceptes estudiats en el cas de variables discretes bidimensionals s'estenen ara al cas continu.

7.1 Distribució conjunta.

Per a dues variables (X, Y) que es distribueixen conjuntament tindrem la seva *funció de densitat conjunta* $f(x, y)$ i la seva *funció de distribució conjunta* $F(x, y)$, on:

$$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x,y)dxdy = p(X \leq x \ i \ Y \leq y)$$

De forma equivalent,

$$p(a \leq X \leq b \ i \ c \leq Y \leq d) = \int_{c}^{d} \int_{a}^{b} f(x,y)dxdy$$

Exemple 7.1.1 *Considerem al següent funció de densitat conjunta:*

$$f(x,y) = \begin{cases} k(x+y) & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$$

(1) Quina ha de ser k per tal de que f(x,y) sigui una funció de densitat ?

$$k \int_{0}^{1} \int_{0}^{1} (x+y)dxdy = 1 \Rightarrow k \int_{0}^{1} (\frac{1}{2} + y)dy = 1$$

$$k(\frac{1}{2} + \frac{1}{2}) = 1 \Rightarrow k = 1$$

(2) Trobar $p(X \leq 0.5, Y \leq 0.5)$

$$p(X \leq 0.5, Y \leq 0.5) = \int_{0}^{0.5} \int_{0}^{0.5} (x+y)dxdy = \int_{0}^{0.5} (\frac{0.25}{2} + 0,5y)dy = \frac{1}{16} + \frac{1}{16} = \frac{1}{8}$$

(3) Trobar $p(X \geq Y)$

$$p(X \geq Y) = \int_0^1 \int_0^x (x+y)dydx = \int_0^1 (x^2 + \frac{x^2}{2})dx = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}$$

(4) Trobar $p(X+Y \leq 1)$

$$p(X+Y \leq 1) = \int_0^1 \int_0^{1-x} (x+y)dydx = \int_0^1 (x(1-x) + \frac{(1-x)^2}{2})dx = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$$

7.2 Distribucions marginals.

Corresponen a la funció de densitat de probabilitat de cada una de les dues variables aïlladament.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

7.3 Densitat i esperança condicionades.

A l'igual que en el cas discret, podem incorporar informació coneguda sobre una de les variables en l'altra variable.

La funció de densitat de X condicionada a Y és

$$f_{X/Y}(x) = \frac{f(x,y)}{f_Y(y)}$$

Exemple 7.3.1 *En el cas de l'exemple anterior tindríem:*

$$f_{X/Y}(x) = \frac{x+y}{y+\frac{1}{2}}$$

L'esperança de X *condicionada a* Y *és:*

$$E(X/Y) = \int_{-\infty}^{\infty} f_{X/Y}(x)dx$$

7.4 Covariància i Coeficient de correlació.

Aquests dos estadístics donen informació sobre el grau de dependència entre les dues variables.

Covariància

$$COV(X,Y) = E(X-E(X))(Y-E(Y)) = E(XY) - E(X)E(Y)$$

Coeficient de correlació

$$\rho = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$$

Dues variables independents tenen una covariància de zero. En cas de que no siguin independents, tindrem que:

$$V(X+Y) = V(X) + V(Y) + 2COV(X,Y)$$

$$V(X-Y) = V(X) + V(Y) - 2COV(X,Y)$$

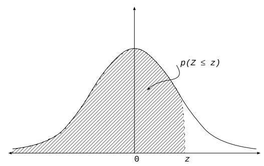
Part I

Appendix

Apèndix A

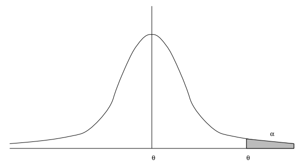
Taules Estadístiques

A.1 Distribució Normal Estàndard



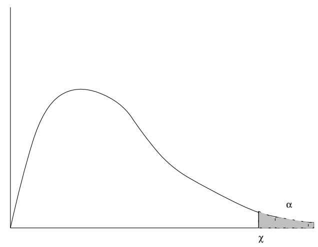
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

A.2 Distribució $t - student$



n	α 0.90	0.95	0.975	0.990	0.995	0.999	0.9995
1	3.078	6.314	12.706	31.821	63.657	318.302	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.598
3	1.638	2.353	3.182	4.541	5.841	10.215	12.941
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.894	6.859
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.405
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.611	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	1.290	1.660	1.984	2.364	2.626	3.174	3.391
200	1.286	1.653	1.972	2.345	2.601	3.131	3.340
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

A.3 Distribució χ^2



	α								
n	0.005	0.01	0.025	0.05	0.950	0.975	0.990	0.995	0.999
1.0	$0.0^4 393$	$0.0^3 157$	$0.0^3 982$	0.00393	3.841	5.024	6.635	7.879	10.828
2.0	0.010003	0.02010	0.05064	0.1026	5.991	7.378	9.210	10.597	13.816
3.0	0.07172	0.1148	0.2158	0.3518	7.815	9.348	11.345	12.838	16.266
4.0	0.2070	0.2971	0.4844	0.7107	9.488	11.143	13.277	14.860	18.467
5.0	0.4117	0.5543	0.8312	1.145	11.070	12.832	15.086	16.750	20.515
6.0	0.6757	0.8721	1.237	1.635	12.592	14.449	16.812	18.548	22.458
7.0	0.9893	1.239	1.690	2.167	14.067	16.013	18.475	20.278	24.322
8.0	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	26.124
9.0	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	27.877
10.0	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	29.588
11.0	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	31.264
12.0	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	32.909
13.0	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	34.528
14.0	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	36.123
15.0	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	37.697
16.0	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267	39.252
17.0	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718	40.790
18.0	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156	42.312
19.0	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582	43.820
20.0	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997	45.315
21.0	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401	46.797
22.0	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796	48.268
23.0	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181	49.728
24.0	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.559	51.179
25.0	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928	52.620
26.0	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290	54.052
27.0	11.808	12.879	14.573	16.151	40.113	43.195	46.963	49.645	55.476
28.0	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993	56.892
29.0	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336	58.301
30.0	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672	59.703
40.0	20.707	22.164	24.433	26.509	55.758	59.342	63.691	66.766	73.402
50.0	27.991	29.707	32.357	34.764	67.505	71.420	76.154	79.490	86.661
60.0	35.534	37.485	40.482	43.188	79.082	83.298	88.379	91.952	99.607
70.0	43.275	45.442	48.758	51.739	90.531	95.023	100.425	104.215	112.317
80.0	51.172	53.540	57.153	60.391	101.879	106.629	112.329	116.321	124.839
90.0	59.196	61.754	65.647	69.126	113.145	118.136	124.116	128.299	137.208
100.0	67.328	70.065	74.222	77.929	124.342	129.561	135.807	140.169	149.449

A.4 Distribució F de Snedecor

Aquesta taula correspon a una distribució F de Snedecor amb ν_1 graus de llibertat en el numerador i ν_2 graus de llibertat en el denominador

Els valors tabulats corresponen a la probabilitat:

$$p(F_{v_1,v_2} \leq F) = \alpha$$

tal i com aquesta figura representa:

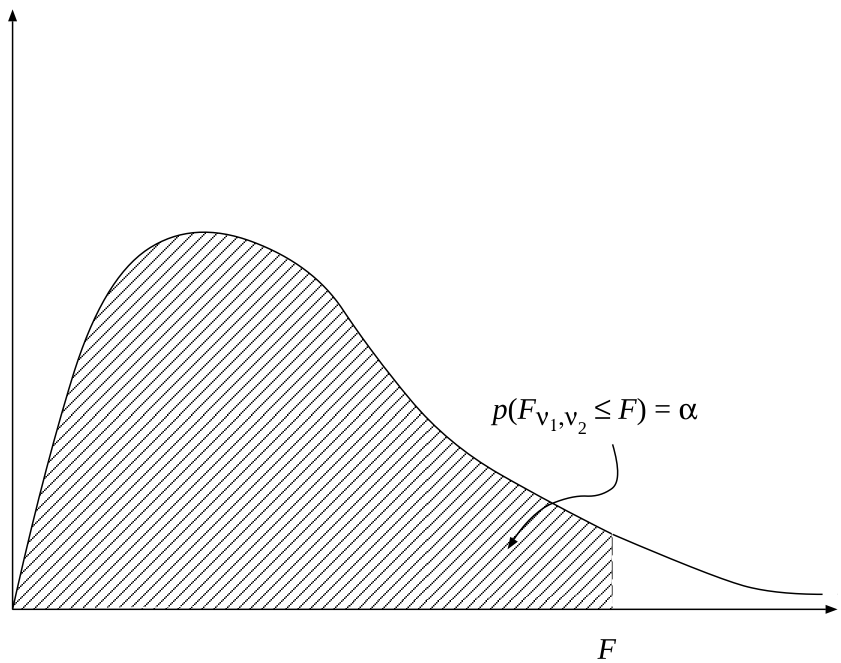


Figura A.1: Distribució F de Snedecor

En aquesta taula es donen els valors per a $\alpha = 0.95$, $\alpha = 0.975$ i $\alpha = 0.99$. En cada casella, el primer número correspon al valor per a $\alpha = 0.95$, i els altres per a 0.975 i 0.99 respectivament

		ν_1									
		1	2	3	4	5	6	7	8		
	1	161 648 4052	199 799 5000	216 864 5403	225 900 5625	230 922 5764	234 937 5859	237 948 5928	239 957 5981	1	
	2	18.51 38.51 98.50	19.00 39.00 99.00	19.16 39.17 99.17	19.25 39.25 99.25	19.30 39.30 99.30	19.33 39.33 99.33	19.35 39.36 99.36	19.37 39.37 99.37	2	
	3	10.13 17.44 34.12	9.55 16.04 30.82	9.28 15.44 29.46	9.12 15.10 28.71	9.01 14.88 28.24	8.94 14.73 27.91	8.89 14.62 27.67	8.85 14.54 27.49	3	
	4	7.71 12.22 21.20	6.94 10.65 18.00	6.59 9.98 16.69	6.39 9.60 15.98	6.26 9.36 15.52	6.16 9.20 15.21	6.09 9.07 14.98	6.04 8.98 14.80	4	
	5	6.61 10.01 16.26	5.79 8.43 13.27	5.41 7.76 12.06	5.19 7.39 11.39	5.05 7.15 10.97	4.95 6.98 10.67	4.88 6.85 10.46	4.82 6.76 10.29	5	
	6	5.99 8.81 13.75	5.14 7.26 10.92	4.76 6.60 9.78	4.53 6.23 9.15	4.39 5.99 8.75	4.28 5.82 8.47	4.21 5.70 8.26	4.15 5.60 8.10	6	
	7	5.59 8.07 12.25	4.74 6.54 9.55	4.35 5.89 8.45	4.12 5.52 7.85	3.97 5.29 7.46	3.87 5.12 7.19	3.79 4.99 6.99	3.73 4.90 6.84	7	
ν_2	8	5.32 7.57 11.26	4.46 6.06 8.65	4.07 5.42 7.59	3.84 5.05 7.01	3.69 4.82 6.63	3.58 4.65 6.37	3.50 4.53 6.18	3.44 4.43 6.03	8	ν_2
	9	5.12 7.21 10.56	4.26 5.71 8.02	3.86 5.08 6.99	3.63 4.72 6.42	3.48 4.48 6.06	3.37 4.32 5.80	3.29 4.20 5.61	3.23 4.10 5.47	9	
	10	4.96 6.94 10.04	4.10 5.46 7.56	3.71 4.83 6.55	3.48 4.47 5.99	3.33 4.24 5.64	3.22 4.07 5.39	3.14 3.95 5.20	3.07 3.85 5.06	10	
	12	4.75 6.55 9.33	3.89 5.10 6.93	3.49 4.47 5.95	3.26 4.12 5.41	3.11 3.89 5.06	3.00 3.73 4.82	2.91 3.61 4.64	2.85 3.51 4.50	12	
	15	4.54 6.20 8.68	3.68 4.77 6.36	3.29 4.15 5.42	3.06 3.80 4.89	2.90 3.58 4.56	2.79 3.41 4.32	2.71 3.29 4.14	2.64 3.20 4.00	15	
	20	4.35 5.87 8.10	3.49 4.46 5.85	3.10 3.86 4.94	2.87 3.51 4.43	2.71 3.29 4.10	2.60 3.13 3.87	2.51 3.01 3.70	2.45 2.91 3.56	20	
	25	4.24 5.69 7.77	3.39 4.29 5.57	2.99 3.69 4.68	2.76 3.35 4.18	2.60 3.13 3.85	2.49 2.97 3.63	2.40 2.85 3.46	2.34 2.75 3.32	25	
	50	4.03 5.34 7.17	3.18 3.97 5.06	2.79 3.39 4.20	2.56 3.05 3.72	2.40 2.83 3.41	2.29 2.67 3.19	2.20 2.55 3.02	2.13 2.46 2.89	50	
	100	3.94 5.18 6.90	3.09 3.83 4.82	2.70 3.25 3.98	2.46 2.92 3.51	2.31 2.70 3.21	2.19 2.54 2.99	2.10 2.42 2.82	2.03 2.32 2.69	100	

		ν_1									
		9	10	12	15	20	25	50	100		
		241	242	244	246	248	249	252	253		
	1	963	969	977	985	993	998	1008	1013	1	
		6022	6056	6106	6157	6209	6240	6303	6334		
		19.38	19.40	19.41	19.43	19.45	19.46	19.48	19.49		
	2	39.39	39.40	39.41	39.43	39.45	39.46	39.48	39.49	2	
		99.39	99.40	99.42	99.43	99.45	99.46	99.48	99.49		
		8.81	8.79	8.74	8.70	8.66	8.63	8.58	8.55		
	3	14.47	14.42	14.34	14.25	14.17	14.12	14.01	13.96	3	
		27.35	27.23	27.05	26.87	26.69	26.58	26.35	26.24		
		6.00	5.96	5.91	5.86	5.80	5.77	5.70	5.66		
	4	8.90	8.84	8.75	8.66	8.56	8.50	8.38	8.32	4	
		14.66	14.55	14.37	14.20	14.02	13.91	13.69	13.58		
		4.77	4.74	4.68	4.62	4.56	4.52	4.44	4.41		
	5	6.68	6.62	6.52	6.43	6.33	6.27	6.14	6.08	5	
		10.16	10.05	9.89	9.72	9.55	9.45	9.24	9.13		
		4.10	4.06	4.00	3.94	3.87	3.83	3.75	3.71		
	6	5.52	5.46	5.37	5.27	5.17	5.11	4.98	4.92	6	
		7.98	7.87	7.72	7.56	7.40	7.30	7.09	6.99		
		3.68	3.64	3.57	3.51	3.44	3.40	3.32	3.27		
	7	4.82	4.76	4.67	4.57	4.47	4.40	4.28	4.21	7	
		6.72	6.62	6.47	6.31	6.16	6.06	5.86	5.75		
		3.39	3.35	3.28	3.22	3.15	3.11	3.02	2.97		
ν_2	8	4.36	4.30	4.20	4.10	4.00	3.94	3.81	3.74	8	ν_2
		5.91	5.81	5.67	5.52	5.36	5.26	5.07	4.96		
		3.18	3.14	3.07	3.01	2.94	2.89	2.80	2.76		
	9	4.03	3.96	3.87	3.77	3.67	3.60	3.47	3.40	9	
		5.35	5.26	5.11	4.96	4.81	4.71	4.52	4.41		
		3.02	2.98	2.91	2.85	2.77	2.73	2.64	2.59		
	10	3.78	3.72	3.62	3.52	3.42	3.35	3.22	3.15	10	
		4.94	4.85	4.71	4.56	4.41	4.31	4.12	4.01		
		2.80	2.75	2.69	2.62	2.54	2.50	2.40	2.35		
	12	3.44	3.37	3.28	3.18	3.07	3.01	2.87	2.80	12	
		4.39	4.30	4.16	4.01	3.86	3.76	3.57	3.47		
		2.59	2.54	2.48	2.40	2.33	2.28	2.18	2.12		
	15	3.12	3.06	2.96	2.86	2.76	2.69	2.55	2.47	15	
		3.89	3.80	3.67	3.52	3.37	3.28	3.08	2.98		
		2.39	2.35	2.28	2.20	2.12	2.07	1.97	1.91		
	20	2.84	2.77	2.68	2.57	2.46	2.40	2.25	2.17	20	
		3.46	3.37	3.23	3.09	2.94	2.84	2.64	2.54		
		2.28	2.24	2.16	2.09	2.01	1.96	1.84	1.78		
	25	2.68	2.61	2.51	2.41	2.30	2.23	2.08	2.00	25	
		3.22	3.13	2.99	2.85	2.70	2.60	2.40	2.29		
		2.07	2.03	1.95	1.87	1.78	1.73	1.60	1.52		
	50	2.38	2.32	2.22	2.11	1.99	1.92	1.75	1.66	50	
		2.78	2.70	2.56	2.42	2.27	2.17	1.95	1.82		
		1.97	1.93	1.85	1.77	1.68	1.62	1.48	1.39		
	100	2.24	2.18	2.08	1.97	1.85	1.77	1.59	1.48	10	
		2.59	2.50	2.37	2.22	2.07	1.97	1.74	1.60		