

Probabilitat i Estadística

Definicions i resultats

apunts per a l'assignatura
del grau de telecomunicacions de l'ETSETB

Xavier Gràcia

`xavier.gracia@upc.edu`

<https://web.mat.upc.edu/xavier.gracia/pie/>

Departament de Matemàtiques
Escola d'Enginyeria de Telecomunicació de Barcelona (ETSETB)
Universitat Politècnica de Catalunya

juny 2024



© Xavier Gràcia Sabaté 2024

Document subjecte a la llicència de Reconeixement –
NoComercial – CompartirIgual 4.0 Internacional

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Sumari

1	Conceptes bàsics de probabilitat	5
1.1	Experiments aleatoris i esdeveniments aleatoris	5
1.2	Àlgebra d'esdeveniments	6
1.3	Espais de probabilitat	7
1.4	Espais de probabilitat discrets	8
1.5	Esdeveniments independents	9
1.6	Producte d'espais de probabilitat	10
1.7	Probabilitat condicionada	11
1.8	Fórmula de la probabilitat total i teorema de Bayes	12
2	Variables aleatòries	13
2.1	La σ -àlgebra de Borel	13
2.2	Variables aleatòries	13
2.3	Funció de distribució d'una variable aleatòria	14
2.4	Variables aleatòries discretes	15
2.5	Exemples importants de variables aleatòries discretes	16
2.6	Variables aleatòries contínues	18
2.7	Exemples importants de variables aleatòries contínues	19
2.8	Variables aleatòries mixtes	21
2.9	El teorema de Moivre–Laplace	21
2.10	Funcions de distribució i densitat condicionades	22
2.11	Funcions d'una variable aleatòria	23
2.12	Esperança	24
2.13	Variància i moments	26
2.14	Paràmetres de les distribucions més usuals	27
2.15	Desigualtats de Markov i Txebixev	27
2.16	Llei dels grans nombres	29
3	Variables aleatòries multidimensionals	31
3.1	Variables aleatòries multidimensionals	31
3.2	Cas continu: funció de densitat	32
3.3	Cas discret: funció de massa de probabilitat	33
3.4	Independència de variables aleatòries	34
3.5	Condicionament	34
3.6	Funcions de diverses variables aleatòries. Suma de variables aleatòries	36
3.7	Transformacions i canvis de variable	36
3.8	Esperança d'una funció de variables aleatòries	37
3.9	Covariància i correlació	37
3.10	Variable gaussiana bidimensional	38
3.11	Variables aleatòries n -dimensionals	39

3.12	Variables gaussianes n -dimensionals	41
3.13	Estimació de variables aleatòries	42
3.14	Convergència de variables aleatòries	43
4	Estadística	47
4.1	Més paràmetres de les variables aleatòries	47
4.2	Funció característica d'una variable aleatòria	49
4.3	Algunes distribucions associades a la normal	50
4.4	Poblacions i mostres	51
4.5	Estadístics	52
4.6	Estimadors	54
4.7	Mètodes d'estimació: mètode dels moments	55
4.8	Mètodes d'estimació: mètode de la màxima versemblança	56
4.9	Informació de Fisher i fita de Cramér-Rao	57
4.10	Intervals de confiança	58
4.11	Intervals de confiança per a alguns paràmetres destacables	59
4.12	Hipòtesis estadístiques i tests d'hipòtesis	62
4.13	Significació i potència d'un test d'hipòtesi	63
4.14	Valor p d'un test d'hipòtesi	64
4.15	Tests d'hipòtesi sobre paràmetres interessants	66
A	Complements	67
A.1	Elements de combinatòria	67
A.2	Funcions especials	70
Índex	terminològic	73

Prefaci

Amb algunes diferències, aquests apunts venen a ser l'esquelet de l'assignatura de Probabilitat i Estadística del Grau d'Enginyeria de Telecomunicació de l'Escola de Telecom de Barcelona, de la UPC, que estic impartint recentment. No hi ha demostracions, ni gairebé exemples; només les definicions i els resultats, d'importància diversa. Aquestes notes no substitueixen els apunts, ni els llibres, ni, encara menys, el treball personal.

El text sempre està obert a millores, així que us agrairé que em feu arribar correccions i suggeriments.

Les *definicions* hi apareixen indicades així. La resta d'enunciats són proposicions, teoremes, corollaris, ... Alguns comentaris i resultats auxiliars també hi apareixen, amb una lletra més petita.

1 Conceptes bàsics de probabilitat

1.1 Experiments aleatoris i esdeveniments aleatoris

(1.1.1) Un *experiment aleatori* és un experiment $\mathcal{E}$ que es pot repetir tantes vegades com es vulgui en les mateixes condicions.

Cada repetició d'un experiment aleatori s'anomena *prova*, i se n'obté un *resultat*.

(1.1.2) L'*espai mostral* és el conjunt Ω dels possibles resultats d'un experiment aleatori.

(1.1.3) Donat un experiment aleatori $\mathcal{E}$, un *esdeveniment aleatori* és una proposició lògica $\mathcal{S}[\omega]$ referida al possible resultat ω d'una prova; la realització de la prova determina si $\mathcal{S}$ és cert o fals (si l'esdeveniment ha ocorregut o no).

(1.1.4) L'esdeveniment aleatori $\mathcal{S}$ s'identifica a un subconjunt $S \subset \Omega$, a saber,

$$S = \{\omega \in \Omega \mid \mathcal{S}[\omega] \text{ és cert}\}.$$

Amb aquesta identificació, el conjunt de tots els possibles esdeveniments aleatoris s'identifica amb el conjunt $\mathcal{P}(\Omega)$ de parts de Ω .

(1.1.5) Les operacions lògiques efectuable amb les proposicions referides als elements de Ω es corresponen amb les operacions de l'àlgebra de Boole efectuable amb els subconjunts de Ω . Tenim una taula de correspondències:

proposicions	subconjunts
disjunció $\mathcal{R} \text{ o } \mathcal{S}$	unió $R \cup S$
conjunció $\mathcal{R} \cap \mathcal{S}$	intersecció $R \text{ i } S$
negació $\text{no } \mathcal{R}$	complementari $\bar{R} \equiv R^c := \Omega - R$
implicació $\mathcal{R} \implies \mathcal{S}$	inclusió $R \subset S$
equivalència $\mathcal{R} \iff \mathcal{S}$	igualtat $R = S$
$\mathcal{R}$ però no $\mathcal{S}$	diferència $R - S = R \cap \bar{S}$

(1.1.6) Hi ha algunes denominacions especials per a esdeveniments, o parells d'esdeveniments, que també tenen la seva correspondència en termes de subconjunts:

proposicions	subconjunts
esdeveniment impossible (sempre fals)	$\emptyset$
esdeveniment segur (sempre cert)	Ω
esdeveniments incompatibles: quan $\mathcal{R}$ i $\mathcal{S}$ és impossible	subconjunts disjunts: $R \cap S = \emptyset$

- (1.1.7) En vista d'aquesta equivalència, considerarem els esdeveniments tant com a proposicions o com a subconjunts, segons ens resulti més convenient.

1.2 Àlgebra d'esdeveniments

Encara que hem identificat esdeveniments aleatoris $\mathcal{S}$ amb subconjunts S de l'espai mostral Ω , per motius tant teòrics com pràctics cal considerar només certs subconjunts $S \subset \Omega$.

- (1.2.1) Una **àlgebra de conjunts** o **àlgebra d'esdeveniments**¹ d'un conjunt Ω és un subconjunt $\mathcal{F} \subset \mathcal{P}(\Omega)$ no buit i estable per complementari i unions finites:

- (i) $\mathcal{F} \neq \emptyset$;
- (ii) $A \in \mathcal{F} \implies \bar{A} \in \mathcal{F}$;
- (iii) $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$.

- (1.2.2) Si se satisfan les condicions anteriors, també se satisfan:

- (iv) $\emptyset, \Omega \in \mathcal{F}$;
- (v) $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$.

- (1.2.3) En termes més tècnics, podríem dir que una àlgebra de conjunts és una *subàlgebra de Boole* de $\mathcal{P}(\Omega)$.

- (1.2.4) Tanmateix, per al desenvolupament de la teoria de la probabilitat no n'hi ha prou amb requerir la propietat (iii) per a unions finites; cal considerar unions *numerables*.

- (1.2.5) Una **σ -àlgebra** (de conjunts, o d'esdeveniments) és una àlgebra de conjunts on en lloc de la propietat (iii) imposen que $\mathcal{F}$ sigui estable per unions *numerables*:

¹ Aquest terme usat més habitualment en el context de la teoria de la probabilitat.

$$(iii') \quad A_i \in \mathcal{F} \quad (i \in \mathbf{N}) \implies \bigcup_{i \in \mathbf{N}} A_i \in \mathcal{F}.$$

(1.2.6) En tal cas, $\mathcal{F}$ també és estable per interseccions numerables:

$$(v') \quad A_i \in \mathcal{F} \quad (i \in \mathbf{N}) \implies \bigcap_{i \in \mathbf{N}} A_i \in \mathcal{F}.$$

(1.2.7) Evidentment, una σ -àlgebra és una àlgebra; el recíproc és fals. No hi ha diferència entre els dos conceptes quan Ω és finit.

(1.2.8) Sigui Ω un conjunt, i $\mathcal{C} \subset \mathcal{P}(\Omega)$ un conjunt arbitrari de parts de Ω . La **σ -àlgebra generada per $\mathcal{C}$** és la σ -àlgebra més petita que conté els elements de $\mathcal{C}$. A vegades es representa per $\sigma(\mathcal{C})$. És la intersecció de totes les σ -àlgebres que contenen $\mathcal{C}$.

1.3 Espais de probabilitat

(1.3.1) Un **espai mesurable**, o **espai probabilitzable**, és un parell $(\Omega, \mathcal{F})$ on

- Ω és un conjunt no-buit;
- $\mathcal{F}$ és una σ -àlgebra en Ω .

(1.3.2) *Axiomes de Kolmogórov, 1933*

Una **mesura de probabilitat** (o *funció de probabilitat* o simplement *probabilitat*) en $(\Omega, \mathcal{F})$ és una funció $P: \mathcal{F} \rightarrow \mathbf{R}$ que satisfà les propietats següents:

- (i) Normalització: $P(\Omega) = 1$;
- (ii) Positivitat: $P(A) \geq 0$ per a tot $A \in \mathcal{F}$;
- (iii) σ -additivitat: si $A_i \in \mathcal{F}$ ($i \in \mathbf{N}$), amb $A_i \cap A_j = \emptyset$ per a $i \neq j$, aleshores

$$P\left(\bigcup_{i \in \mathbf{N}} A_i\right) = \sum_{i \in \mathbf{N}} P(A_i).$$

En altres paraules, P és una «mesura positiva normalitzada».

La terna $(\Omega, \mathcal{F}, P)$ és un **espai de probabilitat** amb *espai mostral* Ω i conjunt d'esdeveniments $\mathcal{F}$.

(1.3.3) En el cas que Ω sigui finit, $\mathcal{F}$ és una àlgebra d'esdeveniments, i la σ -additivitat es redueix a l'additivitat:

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

(1.3.4) Una mesura de probabilitat satisfà també les propietats següents:

- (iv) $P(\emptyset) = 0$;

- (v) Complement: $P(\bar{A}) = 1 - P(A)$;
- (vi) $P(A) \leq 1$;
- (vii) Monotonia: $A \subset B \implies P(A) \leq P(B)$.
- (viii) $P(A - B) = P(A) - P(A \cap B)$;
- (ix) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(1.3.5) *Fórmula d'inclusió-exclusió*

$$P(A_1 \cup \dots \cup A_n) = \sum_i P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n).$$

(1.3.6) *Lemes de convergència monòtona*

Sigui A_i ($i \in \mathbf{N}$) una successió *creixent* d'esdeveniments de $\mathcal{F}$ ($A_i \subset A_{i+1}$).
Aleshores

$$P(\cup_i A_i) = \lim_i P(A_i) = \sup_i P(A_i).$$

Sigui B_i ($i \in \mathbf{N}$) una successió *decreixent* d'esdeveniments de $\mathcal{F}$ ($B_i \supset B_{i+1}$). Aleshores

$$P(\cap_i B_i) = \lim_i P(B_i) = \inf_i P(B_i).$$

1.4 Espais de probabilitat discrets

(1.4.1) Sigui Ω un espai mostral *numerable* (finit o infinit). El conjunt $\mathcal{P}(\Omega)$ de totes les parts de Ω és una σ -àlgebra, anomenada **discreta**.

(1.4.2) Els **esdeveniments elementals** són els subconjunts de Ω formats per un sol element, $\{\omega\}$.

Tot esdeveniment $A \subset \Omega$ és la unió (òbviament disjunta i numerable) dels esdeveniments elementals que conté.

(1.4.3) Si a més tenim una mesura de probabilitat $P: \mathcal{P}(\Omega) \rightarrow \mathbf{R}$, direm que $(\Omega, \mathcal{P}(\Omega), P)$ és un **espai de probabilitat discret**.

Els esdeveniments elementals tenen associades unes **probabilitats elementals** que escriurem simplement

$$P(\omega) = P(\{\omega\}).$$

Igualment escriurem amb la mateixa lletra P la funció

$$P: \Omega \rightarrow \mathbf{R}, \quad P(\omega) = P(\{\omega\}),$$

i l'anomenarem **funció de massa de probabilitat**, o simplement **funció de probabilitat**.

(1.4.4) Suposem que Ω és finit, $\Omega = \{\omega_1, \dots, \omega_n\}$, i escrivim $p_i = P(\omega_i)$, Aleshores les probabilitats elementals satisfan

- $p_i \geq 0$,
- $\sum_i p_i = 1$.

Les probabilitats elementals determinen la probabilitat de qualsevol esdeveniment, ja que

$$P(\{\omega_{i_1}, \dots, \omega_{i_r}\}) = p_{i_1} + \dots + p_{i_r}.$$

(1.4.5) *Esdeveniments equiprobables*

Es diu que els esdeveniments elementals són **equiprobables** quan les probabilitats elementals associades són totes iguals. Si $|\Omega| = n$, aleshores cada esdeveniment elemental té probabilitat $1/n$.

Si $A \subset \Omega$ té r elements, $P(A) = \frac{r}{n}$, és a dir,

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

(«casos favorables entre casos possibles»).

(1.4.6) En el cas que Ω sigui infinit numerable les probabilitats elementals també determinen totes les probabilitats. Si $A \subset \Omega$, també tenim

$$P(A) = \sum_{\omega \in A} P(\omega).$$

Ordenant, si calgués, els elements de A , aquesta suma infinita es pot expressar com una sèrie, però no hi ha cap qüestió de convergència ja que la sèrie és de termes positius i és convergent perquè la suma de totes les probabilitats elementals val 1.

Notem que en el cas infinit els esdeveniments elementals mai *no* poden ser equiprobables.

1.5 Esdeveniments independents

En tot el que segueix estem en un espai de probabilitat $(\Omega, \mathcal{F}, P)$.

(1.5.1) Dos esdeveniments aleatoris $A, B \subset \Omega$ es diuen **independents** quan

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

(1.5.2) Aquesta és la definició des del punt de vista teòric. A la pràctica la independència ve donada per les condicions físiques en què es realitza l'experiment, i la fórmula anterior sovint s'utilitza per a *definir* les probabilitats.

- (1.5.3) Una família finita d'esdeveniments $A_1, \dots, A_n$ es diu **independent** quan, per a tot subconjunt no-buit $I \subset \{1, \dots, n\}$,

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

- (1.5.4) Aquesta noció es pot definir també recursivament: els n esdeveniments són independents quan $P(A_1 \cap \dots \cap A_n) = P(A_1) \dots P(A_n)$ i tot subconjunt de k d'aquests esdeveniments, amb $2 \leq k < n$, és independent.

- (1.5.5) El nombre de condicions a verificar és el nombre de parts $I \subset \{1, \dots, n\}$ de $k \geq 2$ elements, o sigui $2^n - (n + 1)$.

En particular, tres esdeveniments A, B, C són independents si i

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

$$P(A \cap B) = P(A) P(B), \quad P(B \cap C) = P(B) P(C), \quad P(C \cap A) = P(C) P(A).$$

La independència dos a dos no assegura la independència de tot el conjunt.

1.6 Producte d'espais de probabilitat

- (1.6.1) Considerem dos experiments aleatoris $\mathcal{E}_1$ i $\mathcal{E}_2$ descrits per respectius espais de probabilitat $(\Omega_i, \mathcal{F}_i, P_i)$.

Si realitzem els dos experiments de manera *independent* (és a dir, cap dels dos no afecta l'altre), aquesta realització es pot considerar com un nou experiment $\mathcal{E}$, els resultats del qual són parells $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$.

Podem assignar una probabilitat P a certs esdeveniments de $\mathcal{E}$ de manera que $P(A_1 \times \Omega_2) = P_1(A_1)$, $P(\Omega_1 \times A_2) = P_2(A_2)$. Si considerem que obtenir $\omega_1 \in A_1$ és un esdeveniment independent d'obtenir $\omega_2 \in A_2$, conclouem que la probabilitat que $(\omega_1, \omega_2) \in A_1 \times A_2$ ha de ser $P(A_1 \times A_2) = P_1(A_1) P_2(A_2)$. Aquesta fórmula porta a definir el concepte de mesura de probabilitat producte.

- (1.6.2) Siguin $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ dos espais mesurables. Aleshores es defineix l'**espai mesurable producte** com el parell $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, essent

- $\Omega_1 \times \Omega_2$, espai mostral producte;
- $\mathcal{F}_1 \otimes \mathcal{F}_2$, σ -àlgebra en $\Omega_1 \times \Omega_2$ generada pels productes cartesianes $A_1 \times A_2$, amb $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$.

- (1.6.3) Suposem a més que $P_1: \mathcal{F}_1 \rightarrow \mathbf{R}$, $P_2: \mathcal{F}_2 \rightarrow \mathbf{R}$ són mesures de probabilitat. Aleshores es pot provar que existeix una única mesura de probabilitat

$$P_1 \times P_2: \mathcal{F}_1 \otimes \mathcal{F}_2 \rightarrow \mathbf{R}$$

definida pel fet que, si $A_1 \in \mathcal{F}_1$ i $A_2 \in \mathcal{F}_2$,

$$(P_1 \times P_2)(A_1 \times A_2) = P_1(A_1) P_2(A_2).$$

S'anomena **mesura de probabilitat producte** de P_1 i P_2 .

La terna $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \times P_2)$ s'anomena **espai de probabilitat producte**.

(1.6.4) Suposem per exemple que els espais de probabilitat anteriors són discrets. Aleshores l'espai de probabilitat producte $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P)$ també és discret, i la funció de massa de probabilitat és $P = P_1 \otimes P_2 : \Omega_1 \times \Omega_2 \rightarrow \mathbf{R}$, definida per $P(\omega_1, \omega_2) = P_1(\omega_1) P_2(\omega_2)$.

(1.6.5) De manera anàloga es pot considerar la realització independent de n experiments $\mathcal{E}_1, \dots, \mathcal{E}_n$, que es descriuria amb el producte dels n respectius espais de probabilitat: $P(A_1 \times \dots \times A_n) = P_1(A_1) \cdots P_n(A_n)$.

En particular, els n experiments poden ser iguals (*proves repetides*).

1.7 Probabilitat condicionada

(1.7.1) Siguin A, B esdeveniments aleatoris, amb $P(B) \neq 0$. La **probabilitat de A condicionada a (que hagi ocorregut) B** és

$$P(A | B) \equiv P_B(A) := \frac{P(A \cap B)}{P(B)}.$$

(1.7.2) La interpretació de la probabilitat condicionada $P(A|B)$ és que l'esdeveniment A tenia una *probabilitat a priori* $P(A)$. El fet de saber que s'ha verificat l'esdeveniment B canvia la probabilitat de A , que passa a tenir una *probabilitat a posteriori* $P(A|B)$.

(1.7.3) A i B són independents sii $P(A|B) = P(A)$.

(1.7.4) Això ens dona una interpretació de la independència: A és independent de B quan el fet que s'hagi verificat B no canvia la probabilitat que es verifiqui A .

(1.7.5) Sigui $(\Omega, \mathcal{F}, P)$ un espai de probabilitat, i sigui $B \in \mathcal{F}$ un esdeveniment amb $P(B) \neq 0$. Aleshores la probabilitat condicionada $P_B : \mathcal{F} \rightarrow \mathbf{R}$ també és una mesura de probabilitat en $(\Omega, \mathcal{F})$.

(1.7.6) Una altra interpretació de la probabilitat condicionada és que ens permet reduir l'espai mostral. Sabent que ha ocorregut B , l'espai mostral, que era Ω , es redueix a B . Un esdeveniment $A \subset \Omega$ passa a ser un esdeveniment $A \cap B \subset B$. El factor $\frac{1}{P(B)}$ normalitza la nova probabilitat: igual com era $P(\Omega) = 1$, $P(B|B) = 1$.

$$(1.7.7) \quad P(\bar{A}|B) = 1 - P(A|B).$$

1.8 Fórmula de la probabilitat total i teorema de Bayes

(1.8.1) *Fórmula de la probabilitat total*

Sigui $(\Omega, \mathcal{F}, P)$ un espai de probabilitat. Sigui $(A_i)_{i \in I}$ una família numerable (finita o infinita) d'esdeveniments que formen una *partició* de Ω .

Per a tot esdeveniment B , es compleix la **fórmula de la probabilitat total**:

$$P(B) = \sum_i P(B \cap A_i) = \sum_i P(B|A_i) P(A_i).$$

(1.8.2) En el darrer membre d'aquesta fórmula, observem que la probabilitat condicionada $P(B|A_i)$ no està definida si $P(A_i) = 0$. Tanmateix, aquest hipotètic valor és irrellevant, ja que dins la relació $P(B \cap A_i) = P(B|A_i) P(A_i)$ tant $P(A_i)$ com $P(B \cap A_i)$ són zero.

(1.8.3) Si A, B són esdeveniments, $P(A \cap B) = P(A|B) P(B) = P(B|A) P(A)$.

(1.8.4) *Teorema de Bayes*

Sigui $(\Omega, \mathcal{F}, P)$ un espai de probabilitat, i $B \in \mathcal{F}$ un esdeveniment amb $P(B) \neq 0$.

Si A és un esdeveniment tal que coneixem $P(B|A)$, aleshores

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}.$$

Suposem que $A_1, \dots, A_n$ és una partició de Ω formada per conjunts de $\mathcal{F}$. Aleshores

$$P(A_i|B) = \frac{P(B|A_i) P(A_i)}{\sum_k P(B|A_k) P(A_k)}.$$

Ambdues igualtats es coneixen com a **fórmula de Bayes**.

(1.8.5) Un cas particular és el de la partició formada per un esdeveniment A i el seu complementari $\bar{A}$:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B|A) P(A) + P(B|\bar{A}) P(\bar{A})}.$$

2 Variables aleatòries

2.1 La σ -àlgebra de Borel

- (2.1.1) Quan l'espai mostral no és numerable, no podem estudiar la probabilitat a partir d'esdeveniments elementals. En aquest cas cal definir σ -àlgebres apropiades. Els casos més importants són els de $\Omega = \mathbf{R}$, $\Omega = \mathbf{R}^n$, o més generalment un subconjunt de $\mathbf{R}^n$.
- (2.1.2) La **σ -àlgebra de Borel** de $\mathbf{R}$ és la σ -àlgebra $\mathcal{B}$ generada pels intervals oberts de $\mathbf{R}$.
Els elements de $\mathcal{B}$ s'anomenen conjunts **borelians**.
- (2.1.3) La σ -àlgebra de Borel $\mathcal{B}$ conté tots els intervals (oberts, semioberts, tancats; fitats o no fitats), així com tots els conjunts oberts, i els tancats, els conjunts numerables, etc.
- (2.1.4) $\mathcal{B}$ també és la σ -àlgebra generada pels intervals $] -\infty, a]$, amb $a \in \mathbf{R}$.

2.2 Variables aleatòries

- (2.2.1) Sigui $(\Omega, \mathcal{F})$ un espai probabilitzable. Una **variable aleatòria** en $(\Omega, \mathcal{F})$ és una «aplicació mesurable» de $(\Omega, \mathcal{F})$ en $(\mathbf{R}, \mathcal{B})$, on $\mathcal{B}$ és la σ -àlgebra de Borel de $\mathbf{R}$.
És a dir, una aplicació

$$X: \Omega \rightarrow \mathbf{R}$$
tal que, per a tot borelià $B \subset \mathbf{R}$, la seva antiimatge $X^{-1}(B) \subset \Omega$ és un esdeveniment de $\mathcal{F}$.
Recordem que $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}$.
- (2.2.2) Com que la σ -àlgebra de Borel està generada per les semirectes $] -\infty, a]$ ($a \in \mathbf{R}$), per a comprovar que X és mesurable basta comprovar que

$$\forall a \in \mathbf{R} \quad X^{-1}(] -\infty, a]) \in \mathcal{F}.$$
- (2.2.3) **Variable aleatòria constant**
Donat qualsevol $c \in \mathbf{R}$, la funció constant $X = c$ és una variable aleatòria.
- (2.2.4) **Variable indicadora**
Sigui $E \subset \Omega$ qualsevol esdeveniment. La funció indicadora de E , que a vegades es representa per $\mathbb{1}_E$:

$\mathbb{1}_E: \Omega \rightarrow \mathbf{R}$, $\mathbb{1}_E(\omega) = 1$ si $\omega \in E$, $\mathbb{1}_E(\omega) = 0$ si $\omega \notin E$,
és una variable aleatòria.

(2.2.5) Si $(\Omega, \mathcal{F})$ és discret aleshores tota funció $X: \Omega \rightarrow \mathbf{R}$ és una variable aleatòria.

(2.2.6) Suposem ara que $(\Omega, \mathcal{F}, P)$ és un espai de probabilitat. Si $X: \Omega \rightarrow \mathbf{R}$ és una variable aleatòria, per a cada borelià $B \subset \mathbf{R}$ podem calcular

$$P_X(B) \equiv P(X \in B) := P(X^{-1}(B)) = P(\{\omega \in \Omega \mid X(\omega) \in B\}).$$

La notació $P(X \in B)$ és informal però la usarem sovint; la llegirem dient que és la probabilitat que, en realitzar l'experiment, el valor obtingut pertanyi a B .

Un cas particular important és aquell on B és un interval. Per exemple, escriurem $P(a \leq X \leq b) \equiv P(X \in [a, b])$.

(2.2.7) $P_X: \mathcal{B} \rightarrow \mathbf{R}$ és una mesura de probabilitat en $(\mathbf{R}, \mathcal{B})$.

(2.2.8) Es diu que P_X és la **lei de probabilitat**, o **distribució de probabilitat**, de la variable aleatòria X .

La llei P_X conté una informació parcial de P i X : ens informa de la probabilitat dels esdeveniments vinculats als valors de X .

2.3 Funció de distribució d'una variable aleatòria

En aquesta secció $(\Omega, \mathcal{F}, P)$ és un espai de probabilitat i $X: \Omega \rightarrow \mathbf{R}$ una variable aleatòria.

(2.3.1) Com que $\mathcal{B}$ està generada pels intervals $]-\infty, x]$ ($x \in \mathbf{R}$), la llei P_X de X està determinada per les probabilitats

$$P_X(]-\infty, x]) \equiv P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}).$$

(2.3.2) La **funció de distribució** de X és la funció

$$F_X: \mathbf{R} \rightarrow \mathbf{R}, \quad F_X(x) = P(X \leq x).$$

(2.3.3) Si $a < b$, $P(a < X \leq b) = F_X(b) - F_X(a)$.

(2.3.4) La funció de distribució satisfà

- $0 \leq F_X(x) \leq 1$.
- F_X és creixent.

(2.3.5) *Límits de la funció de distribució*

$$(i) \quad \lim_{x \rightarrow -\infty} F_X(x) = 0, \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

(ii) F_X és contínua per la dreta: $F_X(a) = F_X(a^+)$.

(iii) $F_X(a) - F_X(a^-) = P(X=a)$.

(On hem escrit el límit per la dreta $F(a^+) = \lim_{x \rightarrow a} F(x)$, i anàlogament $F(a^-)$.)

Per tant, F_X és contínua sii tots els «esdeveniments elementals» $X=a$ tenen probabilitat nul·la.

(2.3.6) Recíprocament, es pot provar que una funció $F: \mathbf{R} \rightarrow \mathbf{R}$ és la funció de distribució d'una variable aleatòria sii és creixent, satisfà els límits $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$, i és contínua per la dreta.

(2.3.7) La funció de distribució permet expressar fàcilment les probabilitats dels esdeveniments $X \in I$, amb $I \subset \mathbf{R}$ interval. Per exemple,

$$\begin{aligned} P(a < X < b) &= F_X(b^-) - F_X(a), & P(a \leq X \leq b) &= F_X(b) - F_X(a^-), \\ P(X < a) &= F_X(a^-) & P(X > b) &= 1 - F_X(b), & P(X \geq b) &= 1 - F_X(b^-). \end{aligned}$$

2.4 Variables aleatòries discretes

(2.4.1) Una variable aleatòria $X: \Omega \rightarrow \mathbf{R}$ es diu **discreta** si el seu conjunt de valors $\Omega_X := X(\Omega) \subset \mathbf{R}$ és numerable.

Això es compleix, per exemple, sempre que Ω sigui numerable.

Les variables constants i les variables indicadores són discretes.

(2.4.2) En tal cas, la llei de X defineix una probabilitat discreta en el conjunt Ω_X , que queda determinada per la funció de massa de probabilitat

$$P_X: \Omega_X \rightarrow \mathbf{R}, \quad P_X(x) \equiv P(X=x),$$

per als $x \in \Omega_X$. Si $J \subset \mathbf{R}$,

$$P(X \in J) = \sum_{y \in J \cap \Omega_X} P(X=y),$$

i la funció de distribució és

$$F_X(x) = \sum_{\substack{y \in \Omega_X \\ y \leq x}} P(X=y).$$

(2.4.3) *Suposem* que podem ordenar els elements de Ω_X de forma creixent:

$$\Omega_X = \{x_i \mid i \geq 1\}, \quad x_1 < x_2 < x_3 < \dots$$

i que aquests tenen probabilitats elementals $p_i = P(X=x_i)$. Aleshores F_X és una funció esglaonada que es pot expressar en termes de la funció esglaonada

de Heaviside² $\theta = \mathbb{1}_{[0,+\infty[}$ com

$$F_X(x) = \sum_{i \geq 1} p_i \theta(x - x_i).$$

La suposició que hem fet no es compleix sempre, però sí en els casos importants.

2.5 Exemples importants de variables aleatòries discretes

És habitual fer servir una notació de l'estil $X \sim \text{Nom}(param)$ per a indicar que una variable aleatòria segueix una llei de tipus **Nom** amb paràmetres *param*.

(2.5.1) *Llei de Bernoulli*

$X \sim \text{Bern}(p)$, on $0 < p < 1$.

Ve caracteritzada per

$$\Omega_X = \{0, 1\}, \quad P_X(1) = p, \quad P_X(0) = 1-p.$$

La seva funció de distribució es pot expressar

$$F_X(x) = (1-p) \theta(x) + p \theta(x-1).$$

(2.5.2) Un **experiment de Bernoulli** és un experiment aleatori que pot tenir dos resultats, que podem anomenar *èxit* i *fracàs*. Si la probabilitat d'èxit és p ($0 < p < 1$), la probabilitat de fracàs és $1-p$.

A partir d'un experiment de Bernoulli podem construir una variable aleatòria de Bernoulli, assignant els valors 1 o 0 segons els dos possibles resultats (èxit o fracàs).

(2.5.3) Si $\mathcal{E}$ és un experiment qualsevol i $E \subset \Omega$ és un esdeveniment *fixat* referit a $\mathcal{E}$, podem definir un experiment de Bernoulli assignant a un $\omega \in \Omega$ l'èxit si $\omega \in E$, el fracàs en cas contrari.

La variable indicadora de E , $X = \mathbb{1}_E$, és una variable de Bernoulli. Està definida per

$$X(\omega) = 1 \text{ si } \omega \in E \quad (E \text{ es verifica}),$$

$$X(\omega) = 0 \text{ si } \omega \notin E \quad (E \text{ no es verifica}),$$

amb $P_X(1) = p = P(E)$.

(2.5.4) *Llei binomial*

$X \sim \text{Binom}(n, p)$, on $n \in \mathbb{N}^*$ i $0 < p < 1$.

² Altres notacions habituals per a la funció de Heaviside són u , H , etc.

Ve caracteritzada per

$$\Omega_X = \{0, 1, \dots, n\}, \quad P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Aquesta variable s'obté a partir de la repetició n vegades de manera independent d'un experiment de Bernoulli de paràmetre p ; X designa el nombre d'èxits obtinguts.

(2.5.5) **Llei geomètrica**

$X \sim \text{Geom}(p)$, on $0 < p < 1$.

Ve caracteritzada per

$$\Omega_X = \mathbf{N} - \{0\}, \quad P_X(k) = (1-p)^{k-1} p.$$

El valor de la funció de distribució en els enters $n \geq 1$ és

$$F_X(n) = 1 - (1-p)^n.$$

X és el nombre de vegades que s'ha de repetir un experiment de Bernoulli de paràmetre p fins assolir l'èxit.

(2.5.6) **Llei de Poisson**

$X \sim \text{Poiss}(\alpha)$, on $\alpha > 0$.

Ve caracteritzada per

$$\Omega_X = \mathbf{N}, \quad P_X(k) = e^{-\alpha} \frac{\alpha^k}{k!}.$$

La llei de Poisson de paràmetre α es pot obtenir com a límit de la llei binomial de paràmetres $(n, p = \alpha/n)$ quan $n \rightarrow \infty$.

L'aproximació $P_{\text{Binom}}(k; n, p) \approx P_{\text{Poiss}}(k; \alpha=np)$ és bona si $n \geq 20$ i $p \leq 0.05$.

(2.5.7) La llei de Poisson modela el nombre de vegades que ocorre un cert esdeveniment en un interval fixat de temps, essent aquestes ocurrències independents; llavors el nombre mitjà d'aquestes ocurrències és el paràmetre α . Si el nombre d'ocurrències en un interval $[0, T]$ és $\text{Poiss}(\alpha)$, el nombre d'ocurrències en un interval $[0, cT]$ és $\text{Poiss}(c\alpha)$, per a $c > 0$.

(2.5.8) **Llei binomial negativa**

$X \sim \text{BinomNeg}(r, p)$ on $0 < p < 1$ i $r \in \mathbf{N}^*$.

Ve caracteritzada per

$$\Omega_X = \{r, r+1, \dots\}, \quad P_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}.$$

X és el nombre de vegades que cal repetir de manera independent un experiment de Bernoulli de paràmetre p fins a assolir r èxits.

Evidentment, quan $r = 1$ és una llei geomètrica.

(2.5.9) **Llei hipergeomètrica**

$X \sim \text{Hipergeom}(N, K, n)$, on $N, K, n \in \mathbf{N}$, amb $0 \leq K \leq N$, $0 \leq n \leq N$.

Ve caracteritzada per

$$\Omega_X = \{k \mid \max(0, n+K-N) \leq k \leq \min(n, K)\}, \quad P_X(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}.$$

Suposem que tenim una urna amb N boles, de les quals K posseeixen una certa propietat, i n'extraiem n boles (sense reposició). La variable aleatòria X designa quantes de les boles extretes posseeixen la propietat esmentada.

Si reposéssim cada bola extreta aleshores les extraccions serien independents, i X seria una variable Binomial($n, K/N$).

2.6 Variables aleatòries contínues

(2.6.1) Una variable aleatòria X es diu **contínua** quan la seva funció de distribució F es pot expressar com

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi,$$

on $f_X: \mathbf{R} \rightarrow [0, +\infty[$ és una funció positiva integrable anomenada **funció de densitat (de probabilitat)** de X .

(2.6.2) En aquest cas, la funció de distribució F_X és contínua.

De fet, per ser més precisos, F_X és «absolutament contínua».

Hi ha variables aleatòries amb funció de distribució contínua que tanmateix *no* admeten funció densitat; són les *variables aleatòries singulars*, que no considerarem.³

(2.6.3) A efectes de la integral, podem canviar el valor de f_X en un conjunt de mesura nul·la. Tanmateix, si F_X és de classe C^1 a trossos, prendrem $f_X = F'_X$ en els punts on F_X és diferenciable.

(2.6.4) Si X té densitat f_X , aleshores:

- $\int_{\mathbf{R}} f_X(x) dx = 1$;
- $P(X=a) = 0$;
- $P(a \leq X \leq b) = \int_a^b f_X(x) dx$.

³ Hi ha textos que anomenen *contínues* les variables X amb F_X contínua, i *absolutament contínues* les que tenen una funció de densitat, f_X . No necessitem fer aquesta distinció.

(2.6.5) Recíprocament, es pot provar que qualsevol funció positiva $f: \mathbf{R} \rightarrow [0, +\infty[$ i integrable amb $\int_{\mathbf{R}} f(x) dx = 1$ és la funció densitat d'una variable aleatòria.

(2.6.6) Si f és contínua en x , i Δx és petit, $P(x \leq X \leq x + \Delta x) \approx f(x)\Delta x$, de manera que $f(x)\Delta x$ és un «element de probabilitat».

(2.6.7) Una variable aleatòria discreta amb $\Omega_X = \{x_i\}_{i \geq 1}$, $x_1 < x_2 < \dots$, i probabilitats elementals $p_i = P(X=x_i)$, admet una funció densitat en el sentit de distribucions, que es pot escriure en termes de la delta de Dirac com

$$f_X(x) = \sum_i p_i \delta(x-x_i).$$

2.7 Exemples importants de variables aleatòries contínues

(2.7.1) **Distribució uniforme**

$X \sim \text{Unif}(a, b)$ on $a < b$.

El seu recorregut és $\Omega_X = [a, b]$. Funció densitat:

$$f = \frac{1}{b-a} \mathbb{1}_{[a,b]}.$$

Funció de distribució:

$$F(x) = 0 \text{ si } x \leq a, \quad F(x) = \frac{x-a}{b-a} \text{ si } a \leq x \leq b, \quad F(x) = 1 \text{ si } x \geq b.$$

Si $J \subset [a, b]$ és un interval, $P(X \in J) = \frac{\text{long}(J)}{b-a}$.

(2.7.2) **Distribució exponencial**

$X \sim \text{Exp}(\lambda)$ on $\lambda > 0$.

El seu recorregut és $\Omega_X = [0, +\infty[$. Funció densitat:

$$f(x) = \lambda e^{-\lambda x} \theta(x).$$

Funció de distribució:

$$F(x) = 0 \text{ si } x < 0, \quad F(x) = 1 - e^{-\lambda x} \text{ si } x \geq 0.$$

(2.7.3) La llei exponencial es pot considerar com un límit continu de la llei geomètrica. En efecte, considerem un interval de temps $\Delta > 0$. Suposem que en cada interval de temps $[(k-1)\Delta, k\Delta]$ ($k \geq 1$) es pot produir un esdeveniment amb probabilitat $p = \lambda\Delta$. El moment $X > 0$ de la primera ocurrència de l'esdeveniment és descrit per una llei geomètrica: $P(X > k\Delta) = (1 - \lambda\Delta)^k$, o bé $P(X > t) = (1 - \lambda\Delta)^{t/\Delta}$. Quan $\Delta \rightarrow 0$ el segon membre tendeix a $e^{-\lambda t}$, que correspon a la llei exponencial.

(2.7.4) *Manca de memòria*

Sigui X una variable aleatòria contínua i positiva. Diem que X és **sense memòria** quan

$$\forall t, s \geq 0 \quad \mathbf{P}(X > t+s \mid X > t) = \mathbf{P}(X > s).$$

Interpretació d'aquesta propietat: si X descriu el temps que dura un procés, si a temps t encara no ha acabat, la probabilitat que duri un temps addicional s no depèn de t .

Una variable exponencial descriu per exemple la durada d'un sistema sense desgast.

(2.7.5) Sigui X una variable aleatòria contínua i positiva. X és sense memòria sii és una variable exponencial.

Es pot fer un raonament similar si X és discreta i $\Omega_X = \mathbf{N}$. Aleshores X és una variable geomètrica.

(2.7.6) **Distribució gaussiana o normal**

$X \sim \mathbf{N}(m, \sigma)$ on $m \in \mathbf{R}$, $\sigma > 0$.

El seu recorregut és $\Omega_X = \mathbf{R}$. Funció densitat:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Funció de distribució:

$$F(x) = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x-m}{\sqrt{2}\sigma}\right) \right),$$

on erf designa la funció d'error (vegeu apèndix).

(2.7.7) Una funció **gaussiana** és l'exponencial d'un polinomi de grau 2 amb coeficient principal negatiu. Així doncs, la densitat d'una variable gaussiana ve descrita, amb la normalització adequada, per una funció gaussiana. La seva gràfica s'anomena **campana de Gauss**.(2.7.8) Una **variable normal estàndard** és una variable normal amb paràmetres $m = 0$ i $\sigma = 1$.

En aquest cas les fórmules anteriors s'escriuen

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad F(x) = \frac{1}{2} \left(1 + \operatorname{erf}(x/\sqrt{2}) \right).$$

(2.7.9) **Distribució de Cauchy**

$X \sim \text{Cauchy}(\alpha)$ on $\alpha > 0$.

El seu recorregut és $\Omega_X = \mathbf{R}$. Funció densitat:

$$f(x) = \frac{\alpha/\pi}{x^2 + \alpha^2}.$$

Funció de distribució:

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x}{\alpha}.$$

2.8 Variables aleatòries mixtes

(2.8.1) Hi ha variables aleatòries que tenen un comportament alhora continu i discret. De fet, pel *teorema de descomposició de Lebesgue*, tota llei de probabilitat en $\mathbf{R}$ es pot escriure com la suma d'una part discreta, una part contínua i una part singular.

(2.8.2) El cas més habitual on ens trobarem una llei ***mixta*** és aquell on la funció de distribució de X és contínua a trossos i de classe C^1 a trossos, però no és ni constant a trossos (cas en què X seria discreta) ni contínua (cas en què X seria contínua).

Això sovint passa quan el valor de X s'obté d'un procés en dues etapes: a partir d'una variable contínua amb probabilitat α (on $0 < \alpha < 1$), o d'una de discreta amb probabilitat $1-\alpha$. Llavors la funció de distribució pren la forma

$$F = \alpha F_c + (1-\alpha) F_d,$$

amb F_c, F_d les funcions de distribució de les variables contínua i discreta, respectivament.

Podem descriure aquesta situació amb una funció densitat $f = \alpha f_c + (1-\alpha)f_d$, on f_c és la densitat de la variable contínua i f_d és la densitat de la variable discreta expressada en termes de deltes de Dirac.

2.9 El teorema de Moivre–Laplace

El teorema de Moivre–Laplace és un cas particular del *teorema del límit central* aplicat al cas d'una variable de Bernoulli.

(2.9.1) ***Teorema de Moivre–Laplace***

Sigui $X \sim \text{Binom}(n, p)$, i $\hat{X} \sim N(m=np, \sigma=\sqrt{npq})$ (essent $q = 1-p$).

Aleshores

$$P(X=k) \sim f_{\hat{X}}(k)$$

quan $n \rightarrow \infty$, i $k \approx np$.

- (2.9.2) Això significa que quan n és prou gran (diguem $np(1-p) \geq 10$) la binomial es pot aproximar per una gaussiana. Com que X és discreta i $\hat{X}$ és contínua, farem les aproximacions següents, per a k natural:

$$\begin{aligned} X = k &\longleftrightarrow \hat{X} \in \left[k - \frac{1}{2}, k + \frac{1}{2} \right], \\ P(k_1 \leq k \leq k_2) &\approx P\left(k_1 - \frac{1}{2} \leq \hat{X} \leq k_2 + \frac{1}{2}\right), \\ P(X \leq k) &\approx P\left(\hat{X} \leq k + \frac{1}{2}\right) = F_{\hat{X}}\left(k + \frac{1}{2}\right). \end{aligned}$$

2.10 Funcions de distribució i densitat condicionades

Considerem un espai de probabilitat. Si E és un esdeveniment amb $P(E) \neq 0$, havíem definit el concepte de probabilitat condicionada a E .

- (2.10.1) Si X és una variable aleatòria podem considerar les probabilitats condicionades a E , i doncs la **funció de distribució condicionada**

$$F_X(x | E) := P(X \leq x | E) = \frac{P(X \leq x \text{ i } E)}{P(E)}.$$

- (2.10.2) Si X és contínua amb funció de densitat $f_X = F'_X$ aleshores $F_X(\cdot | E)$ té com a densitat la **funció de densitat condicionada**

$$f_X(x | E) = F'_X(x | E),$$

si la derivada existeix. Llavors, si $A \subset \mathbf{R}$ és un borelià,

$$P(X \in A | E) = \int_A dx f_X(x | E).$$

Anàlogament, si X és discreta la **funció de probabilitat condicionada** és

$$P_X(x | E) = \frac{P(X=x \text{ i } E)}{P(E)}.$$

- (2.10.3) Suposem que els esdeveniments E_i formen una partició de Ω . Podem aplicar la fórmula de la probabilitat total a la llei P_X i als esdeveniments $X \leq x$. Llavors

$$\begin{aligned} F_X(x) &= \sum_i F_X(x | E_i), \\ f_X(x) &= \sum_i f_X(x | E_i). \end{aligned}$$

- (2.10.4) Versió contínua de la fórmula de la probabilitat total:

Si E és un esdeveniment,

$$P(E) = \int_{-\infty}^{+\infty} P(E \mid X=x) f_X(x) dx,$$

on s'ha definit $P(E \mid X=x) = \lim_{\Delta x \rightarrow 0} P(E \mid x \leq X \leq x+\Delta x)$.

(2.10.5) Versió contínua de la fórmula de Bayes:

$$f_X(x \mid E) = \frac{P(E \mid X=x) f_X(x)}{\int_{-\infty}^{+\infty} P(E \mid X=x') f_X(x') dx'}.$$

(2.10.6) Considerem ara un esdeveniment E de la forma $X \in B$, amb $B \subset \mathbf{R}$ un borelià tal que $P(X \in B) \neq 0$.

Si X és contínua amb densitat f_X , aleshores la densitat condicionada és

$$f_X(x \mid X \in B) = \begin{cases} f_X(x)/P_X(B) & \text{si } x \in B, \\ 0 & \text{altrament.} \end{cases}$$

Si X és discreta la llei de X condicionada a $X \in B$ també és discreta, i la funció de probabilitat condicionada ve donada per

$$P_X(x \mid X \in B) = \begin{cases} P_X(x)/P_X(B) & \text{si } x \in B, \\ 0 & \text{altrament.} \end{cases}$$

2.11 Funcions d'una variable aleatòria

(2.11.1) Sigui $(\Omega, \mathcal{F}, P)$ un espai de probabilitat, i $X: \Omega \rightarrow \mathbf{R}$ una variable aleatòria. Sigui $g: \mathbf{R} \rightarrow \mathbf{R}$ una funció **mesurable Borel**, és a dir, tal que, per a tot borelià $B \subset \mathbf{R}$, $g^{-1}(B) \subset \mathbf{R}$ també és un borelià. Aleshores $g(X) \equiv g \circ X: \Omega \rightarrow \mathbf{R}$ també és una variable aleatòria.

L'interès de considerar $g(X)$ pot ser variat. És possible que un experiment ens doni X però per algun motiu només puguem mesurar $g(X)$. O que per algun motiu ens interressi fer un canvi de variable $Y = g(X)$.

(2.11.2) *Cas d'una variable aleatòria discreta*

Si X és discreta $\Omega_X = X(\Omega)$ és numerable. Aleshores $Y = g(X)$ també és discreta, i pren valors $\Omega_Y = Y(\Omega) = g(\Omega_X)$.

La funció de massa de probabilitat corresponent a Y és

$$P_Y(y) = \sum_{x \in g^{-1}(y)} P_X(x).$$

(2.11.3) *Cas general*

Si X és contínua llavors $Y = g(X)$ pot ser discreta (per exemple si g pren un nombre finit de valors), contínua (per exemple si Ω_X és un interval no

degenerat i g és una bijecció contínua amb un altre interval), pot ser mixta, etc.

En qualsevol cas, la funció de distribució de Y ve donada per

$$F_Y(y) = P(g(X) \leq y) = P_X(G_y),$$

on

$$G_y = g^{-1}([-\infty, y]) = \{x \mid g(x) \leq y\}.$$

En alguns casos es pot expressar la probabilitat de G_y de manera explícita.

Per exemple:

- Si g és estrictament creixent i $y = g(x)$ llavors $G_y =]-\infty, x]$, i $F_Y(y) = F_X(x)$.
- Si g és estrictament decreixent i $y = g(x)$ llavors $G_y = [x, +\infty[$, i $F_Y(y) = 1 - F_X(x^-)$.

Quan g no és estrictament monòtona la discussió és més complicada.

(2.11.4) Transformació de la funció densitat

Si X és contínua amb densitat f_X i g compleix certes condicions, aleshores $Y = g(X)$ és contínua i es pot calcular la densitat f_Y a partir de f_X .

El cas més senzill és aquell on $g: \Omega_X \rightarrow \Omega_Y$ és un difeomorfisme de classe C^1 (creixent o decreixent) entre dos intervals. Llavors

$$f_Y = (f_X \circ g^{-1}) |Dg^{-1}| = (f_X \circ g^{-1}) \frac{1}{|Dg \circ g^{-1}|}.$$

Una expressió una mica més complicada es té quan g no és globalment injectiva però, llevat d'un nombre finit de punts, Ω_X és la unió d'un nombre finit d'intervals on g és un difeomorfisme amb la imatge. Llavors

$$f_Y(y) = \sum_{\substack{x \\ g(x)=y}} f_X(x) \frac{1}{|g'(x)|},$$

on la suma s'estén a totes les antiimatges x d'una certa y (i per tant $f_Y(y)$ val zero en les y que no pertanyin a la imatge de g).

2.12 Esperança

(2.12.1) Si (Ω, P) és un espai de probabilitat discret i $X: \Omega \rightarrow \mathbf{R}$ una variable aleatòria, es defineix l'*esperança* de X com el valor mitjà dels valors que pren, ponderats per les respectives probabilitats:

$$E[X] = \sum_{\omega \in \Omega} P(\omega) X(\omega).$$

Una reinterpretació d'aquesta suma porta a la definició general d'esperança.

(2.12.2) *Esperança*

L'**esperança** (o *valor esperat*, *valor mitjà*, *mitjana*) de X és el nombre $E[X] \equiv \mathbb{E}[X] \equiv m_X$ definit de la manera següent:

- Si X és discreta,

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} P_X(x) x.$$
- Si X és contínua,

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} dx f_X(x) x.$$

En tots dos casos se suposa que la convergència és absoluta.

(2.12.3) Propietats de l'esperança:

- Esperança d'una variable constant: $\mathbb{E}[1] = 1$.
- Esperança de la variable indicadora d'un esdeveniment $A \subset \Omega$:
 $\mathbb{E}[\mathbb{1}_A] = P(A)$.
- Linealitat: $\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$.

(2.12.4) Una variable aleatòria $X: \Omega \rightarrow \mathbf{R}$ es diu **positiva** quan pren valors positius, o sigui $X(\Omega) \subset [0, +\infty[$. Ho escriurem, com és habitual, $X \geq 0$.

(2.12.5) Si $X \geq 0$, llavors $\mathbb{E}[X] \geq 0$.

(2.12.6) Si Y és una variable aleatòria contínua i positiva, $\mathbb{E}[Y] = \int_0^{+\infty} dy P(Y > y)$.

(2.12.7) **Teorema de l'esperança**

Sigui X contínua amb densitat f_X , i sigui g una funció mesurable. Aleshores

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} dx f_X(x) g(x),$$

suposant que la integral sigui absolutament convergent.

Si X és discreta,

$$\mathbb{E}(g(X)) = \sum_{x \in \Omega_X} P_X(x) g(x),$$

suposant que la suma sigui absolutament convergent.

(2.12.8) *Esperança condicionada*

Donada una variable aleatòria X i un esdeveniment A , teníem la funció de distribució condicionada $F_X(\cdot|A)$, i les funcions de massa i de densitat condicionades si X és discreta o contínua. A partir d'elles podem definir l'**esperança condicionada**:

- Si X és discreta,

$$\mathbb{E}[X|A] = \sum_{x \in \Omega_X} P_X(x|A) x.$$

- Si X és contínua,

$$\mathbb{E}[X|A] = \int_{-\infty}^{+\infty} dx f_X(x|A) x.$$

L'esperança es pot calcular amb una fórmula de la probabilitat total. Si A_i són esdeveniments que formen una partició de l'espai mostral,

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X|A_i] P(A_i).$$

2.13 Variància i moments

(2.13.1) Variància

Sigui $X: \Omega \rightarrow \mathbf{R}$ una variable aleatòria. La **variància** de X és el nombre $V[X] \equiv \text{Var}[X] \equiv \sigma_X^2$ definit per

$$V[X] = \mathbb{E}[(X - m_X)^2],$$

on $m_X = \mathbb{E}[X]$, suposant que ambdues esperances existeixin.

(2.13.2) Propietats de la variància:

- $V[X] \geq 0$ (positivitat)
- $V[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
- $V[\alpha X] = \alpha^2 V[X]$ si $\alpha \in \mathbf{R}$.
- $V[X+c] = V[X]$ si $c \in \mathbf{R}$.

(2.13.3) Desviació estàndard

La **desviació estàndard** (o *desviació típica*) de X és

$$\sigma_X = V[X]^{1/2}.$$

Evidentment, $\sigma_{\alpha X} = |\alpha| \sigma_X$.

(2.13.4) Sigui X una variable aleatòria amb mitjana m i desviació estàndard $\sigma > 0$.

La **variable aleatòria normalitzada** de X és

$$\tilde{X} = \frac{X - m}{\sigma}$$

(2.13.5) La variable normalitzada satisfà $\mathbb{E}[\tilde{X}] = 0$, $V[\tilde{X}] = 1$.

(2.13.6) Suposem que la variable X té dimensions físiques, per exemple de longitud, L. La funció de distribució és adimensional, però la funció densitat té dimensió L^{-1} , l'esperança té dimensió L, la variància té dimensió L^2 , i la desviació estàndard dimensió L. La variable normalitzada és adimensional.

- (2.13.7) *Moments d'una variable aleatòria*
Si $n \in \mathbf{N}$, el **moment n -èsim** d'una variable aleatòria X es defineix per
$$m_n = \mathbf{E}[X^n] .$$

Si X és discreta, $m_n = \sum_x \mathbf{P}_X(x) x^n$.
Si és contínua, $m_n = \int_{-\infty}^{+\infty} dx \, \mathbf{f}_X(x) x^n$.
 $m_0 = 1$, $\mathbf{E}[X] = m_1$, $\mathbf{V}[X] = m_2 - m_1^2$.
- (2.13.8) El **moment central n -èsim** es defineix per
$$\mu_n = \mathbf{E}[(X - m_X)^n] .$$

 $\mu_0 = 1$, $\mu_1 = 0$, $\mu_2 = \mathbf{V}[X]$.

2.14 Paràmetres de les distribucions més usuals

- (2.14.1) En aquestes taules donem la mitjana m i la variància σ^2 de diversos tipus de variables aleatòries.

variable	m	σ^2
Bern(p)	p	$p(1-p)$
Binom(n, p)	np	$np(1-p)$
Geom(p)	$1/p$	$(1-p)/p^2$
Poiss(α)	α	α
Hipergeom(N, K, n)	nK/N	$\frac{nK(N-K)(N-n)}{N^2(N-1)}$

variable	m	σ^2
Unif(a, b)	$(a + b)/2$	$(b - a)^2/12$
Exp(λ)	$1/\lambda$	$1/\lambda^2$
N(m, σ)	m	σ^2
Cauchy(α)	no existeix	no existeix

2.15 Desigualtats de Markov i Txebixev

- (2.15.1) Un esdeveniment $A \subset \Omega$ d'un espai de probabilitat es diu *gairebé segur* quan $\mathbf{P}(A) = 1$.

Equival a afirmar que $P(\bar{A}) = 0$.

Òbviament l'esdeveniment segur Ω és gairebé segur.

- (2.15.2) Aquest adjectiu l'aplicarem en particular a una propietat qualsevol d'una variable aleatòria, quan es compleix llevat d'un conjunt de probabilitat nul·la.

Per exemple, una variable aleatòria X és **constant gairebé segurament** quan existeix un nombre c tal que $P(X=c) = 1$. Això es pot formular de manera equivalent dient que

- $F_X(x) = 0$ per a cada $x < c$, i $F_X(x) = 1$ per a cada $x \geq c$.

- (2.15.3) De manera semblant, X és **positiva gairebé segurament** quan és positiva (≥ 0) per a tot ω llevat d'un conjunt de probabilitat nul·la. Això es pot expressar amb qualsevol d'aquestes condicions, que són equivalents:

- $P(X \geq 0) = 1$;
- $P(X < 0) = 0$;
- $F_X(x) = 0$ per a cada $x < 0$.

En el cas que X sigui discreta això significa que $P_X(x) = 0$ per a tot $x < 0$. En el cas continu, que X admet una funció de densitat tal que $f_X(x) = 0$ per a cada $x < 0$.

Si X és positiva llavors és positiva gairebé segurament, però el recíproc pot no complir-se, ja que X podria prendre valors < 0 en un conjunt de probabilitat zero.

- (2.15.4) **Desigualtat de Markov**

Sigui X una variable aleatòria positiva. Per a qualsevol nombre $a > 0$,

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

- (2.15.5) Suposem que $E[X] > 0$. Aleshores la desigualtat de Markov es pot expressar com $P(X \geq cE[X]) \leq \frac{1}{c}$.

- (2.15.6) **Desigualtat de Třebixev**

Sigui X una variable aleatòria amb esperança m i variància σ^2 finites. Per a qualsevol nombre $a > 0$,

$$P(|X - m| \geq a) \leq \frac{\sigma^2}{a^2}.$$

- (2.15.7) Suposem que $\sigma \neq 0$. Llavors, canviant a per k/σ podem escriure també

$$P(|X - m| \geq k\sigma) \leq 1/k^2.$$

Això expressa que la probabilitat que X s'allunyi k desviacions estàndard de la seva mitjana és baixa (menor que $1/k^2$) quan k és gran.

- (2.15.8) La desigualtat de Txebyev és un instrument bàsicament teòric, ja que, essent vàlida per a qualsevol variable aleatòria, la fita que dona no és gaire bona.
- (2.15.9) Tanmateix, hi ha variables aleatòries on la desigualtat es compleix com una igualtat (la fita és ajustada).
- (2.15.10) Una variable aleatòria constant $X = c$ té esperança c i variància 0. Recíprocament, si una variable aleatòria X té variància 0, llavors X és constant gairebé segurament.

2.16 Llei dels grans nombres

- (2.16.1) Les *lleis dels grans nombres* descriuen el resultat de realitzar el mateix experiment un gran nombre de vegades: la mitjana dels resultats (numèrics) obtinguts s'acosta a l'esperança a mesura que es fan més proves.
- (2.16.2) ***Llei feble dels grans nombres per a un experiment de Bernouilli***
Repetim n vegades un experiment de Bernouilli, i sigui X el nombre total d'èxits. Per a tot $\varepsilon > 0$.

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X}{n} - p\right| \leq \varepsilon\right) = 1.$$

3 Variables aleatòries multidimensionals

3.1 Variables aleatòries multidimensionals

(3.1.1) La **σ -àlgebra de Borel** de $\mathbf{R}^n$ és la σ -àlgebra $\mathcal{B}_n$ generada pels rectangles oberts $I_1 \times I_2 \times \dots \times I_n \subset \mathbf{R}^n$.

També està generada pels conjunts $]-\infty, a_1] \times \dots \times]-\infty, a_n] \subset \mathbf{R}^n$.

(3.1.2) Sigui $(\Omega, \mathcal{F}, P)$ un espai de probabilitat. Una **variable aleatòria n -dimensional**, o **vector aleatori**, és una aplicació

$$\mathbf{X}: \Omega \rightarrow \mathbf{R}^n$$

«mesurable», és a dir, tal que, per a tot borelià $B \subset \mathbf{R}^n$, el conjunt $\mathbf{X}^{-1}(B) = \{\omega \in \Omega \mid \mathbf{X}(\omega) \in B\}$ és un esdeveniment de $\mathcal{F}$.

(3.1.3) $\mathbf{X} = (X_1, \dots, X_n)$ és una variable aleatòria n -dimensional sii $X_1, \dots, X_n$ són variables aleatòries unidimensionals.

(3.1.4) Considerem una variable aleatòria n -dimensional $\mathbf{X}: \Omega \rightarrow \mathbf{R}^n$. Si $B \subset \mathbf{R}^n$ és un borelià, podem definir

$$P_{\mathbf{X}}(B) \equiv P(\mathbf{X} \in B) := P(\mathbf{X}^{-1}(B)) = P\left(\{\omega \in \Omega \mid \mathbf{X}(\omega) \in B\}\right).$$

(3.1.5) $P_{\mathbf{X}}: \mathcal{B}_n \rightarrow \mathbf{R}$ és una mesura de probabilitat en $(\mathbf{R}^n, \mathcal{B}_n)$.

Per simplicitat, a partir d'ara farem la major part de l'exposició amb $n = 2$.

(3.1.6) Sigui $(X, Y): \Omega \rightarrow \mathbf{R}^2$ una variable aleatòria bidimensional. Si $B_1, B_2 \subset \mathbf{R}$ són borelians, $B_1 \times B_2 \subset \mathbf{R}^2$ és borelià, i

$$(X, Y)^{-1}(B_1 \times B_2) = \{\omega \mid X(\omega) \in B_1 \text{ i } Y(\omega) \in B_2\} = X^{-1}(B_1) \cap Y^{-1}(B_2).$$

(3.1.7) La **funció de distribució conjunta** de (X, Y) és la funció que representarem per $F_{XY}: \mathbf{R}^2 \rightarrow \mathbf{R}$ i està definida per

$$F_{XY}(x, y) := P_{XY}(]-\infty, x] \times]-\infty, y]) \equiv P(X \leq x, Y \leq y).$$

(3.1.8) Propietats de la funció de distribució conjunta:

- i. $0 \leq F_{XY}(x, y) \leq 1$.
- ii. $F_{XY}(x, y)$ és creixent tant respecte a x com a y .
- iii. $\lim_{x \rightarrow -\infty} F_{XY}(x, b) = 0$, $\lim_{y \rightarrow -\infty} F_{XY}(a, y) = 0$.
- iv. $\lim_{(x, y) \rightarrow (+\infty, +\infty)} F_{XY}(x, y) = 1$.
- v. $F_{XY}(\cdot, b)$ i $F_{XY}(a, \cdot)$ són contínues per la dreta.

$$(3.1.9) \quad P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \\ = F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1).$$

(3.1.10) Les **funcions de distribució marginals** de (X, Y) són les funcions de distribució de X i de Y .

(3.1.11) Les funcions de distribució marginals es poden recuperar a partir de la funció de distribució conjunta:

$$F_X(x) = \lim_{y \rightarrow +\infty} F_{XY}(x, y), \quad F_Y(y) = \lim_{x \rightarrow +\infty} F_{XY}(x, y).$$

3.2 Cas continu: funció de densitat

(3.2.1) La variable aleatòria bidimensional (X, Y) es diu **contínua**, o les variables X, Y es diuen **conjuntament contínues**, quan la funció de distribució conjunta es pot expressar com

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y dx' dy' f_{XY}(x', y'),$$

on $f_{XY}: \mathbf{R}^2 \rightarrow \mathbf{R}$ és una funció integrable positiva, anomenada **funció de densitat conjunta**.

(3.2.2) La funció densitat satisfà:

$$\text{i. } \iint_{\mathbf{R}^2} dx dy f_{XY}(x, y) = 1.$$

$$\text{ii. Per a tot borelià } B \subset \mathbf{R}^2, \\ P((X, Y) \in B) = \iint_B dx dy f_{XY}(x, y).$$

(3.2.3) Si F_{XY} és de classe C^2 , aleshores

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}}{\partial x \partial y}(x, y).$$

(3.2.4) Les **funcions de densitat marginals** de (X, Y) són les funcions de densitat f_X, f_Y de les variables unidimensionals X i Y .

(3.2.5) Les funcions de densitat marginals es poden recuperar a partir de la funció de densitat conjunta:

$$f_X(x) = \int_{\mathbf{R}} dy f_{XY}(x, y), \quad f_Y(y) = \int_{\mathbf{R}} dx f_{XY}(x, y).$$

(3.2.6) *Distribució uniforme*

Sigui $B \subset \mathbf{R}^2$ un borelià amb mesura (àrea) finita i no nul·la. La **distribució uniforme** sobre B és aquella que té funció densitat constant sobre B

i nul·la fora de B . Per tant,

$$f = \frac{1}{\text{area}(B)} \mathbb{1}_B.$$

(3.2.7) Distribució normal bidimensional

La **distribució normal bidimensional** té funció densitat

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right)\right)$$

essent els paràmetres $m_1, m_2 \in \mathbf{R}$, $\sigma_1, \sigma_2 \in]0, +\infty[$, $-1 < \rho < 1$.

3.3 Cas discret: funció de massa de probabilitat

(3.3.1) Suposem que $X, Y: \Omega \rightarrow \mathbf{R}$ són variables aleatòries discretes; tenen conjunts de valors Ω_X, Ω_Y numerables. Aleshores el vector aleatori $(X, Y): \Omega \rightarrow \mathbf{R}^2$ pren els seus valors en el conjunt numerable $\Omega_X \times \Omega_Y \subset \mathbf{R}^2$

La **funció de massa de probabilitat conjunta** de (X, Y) és la funció $P_{XY}: \Omega_X \times \Omega_Y \rightarrow \mathbf{R}$

$$P_{XY}(x, y) = P(X=x, Y=y) \equiv P(X^{-1}(x) \cap Y^{-1}(y)).$$

(3.3.2) Amb ella es pot calcular la probabilitat de qualsevol esdeveniment. Si $B \subset \Omega_X \times \Omega_Y$,

$$P_{XY}(B) = \sum_{(x,y) \in B} P_{XY}(x, y).$$

(3.3.3) Les **funcions de massa de probabilitat marginals** són les corresponents a X i Y separatament, P_X i P_Y .

Es poden obtenir a partir de la funció de probabilitat conjunta:

$$P_X(x) = \sum_{y \in \Omega_Y} P_{XY}(x, y), \quad P_Y(y) = \sum_{x \in \Omega_X} P_{XY}(x, y).$$

(3.3.4) Distribució trinomial

Considerem tres esdeveniments E_1, E_2, E_3 que formen una partició de l'espai mostral, amb probabilitats p_1, p_2, p_3 . Es repeteix l'experiment n vegades, i es denota per X_i el nombre d'èxits de l'esdeveniment E_i . La **distribució trinomial** és la llei de probabilitat del vector aleatori (X_1, X_2, X_3) .

La seva funció de massa de probabilitat és

$$P(X_1 = n_1, X_2 = n_2, X_3 = n_3) = \frac{n!}{n_1! n_2! n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3},$$

per als nombres naturals n_i tals que $n_1 + n_2 + n_3 = n$.

De manera anàloga es defineixen les distribucions multinomials. Les seves distribucions marginals són multinomials d'ordre inferior.

3.4 Independència de variables aleatòries

- (3.4.1) Siguin $X, Y: \Omega \rightarrow \mathbf{R}$ variables aleatòries. Es diu que X i Y són **independents** quan, per a tot parell de borelians $A, B \subset \mathbf{R}$, els esdeveniments $X^{-1}(A)$, $Y^{-1}(B)$ són independents:

$$P(X \in A \text{ i } Y \in B) = P(X \in A) P(Y \in B).$$

- (3.4.2) Si X, Y són variables aleatòries independents, i g, h són funcions mesurables Borel, aleshores $g(X)$, $h(Y)$ també són variables independents.

- (3.4.3) X i Y són independents sii la funció de distribució conjunta satisfà

$$F_{XY}(x, y) = F_X(x) F_Y(y).$$

- (3.4.4) Suposem que X i Y són conjuntament contínues, amb densitat conjunta f_{XY} . Aleshores X i Y són independents sii

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

gairebé pertot.

- (3.4.5) Ara suposem que X, Y són discretes. Aleshores són independents sii la funció de massa de probabilitat satisfà

$$P(X=x, Y=y) = P(X=x) P(Y=y).$$

- (3.4.6) Podem expressar la noció d'independència en termes del producte d'espais de probabilitat. X i Y són independents sii

$$(\mathbf{R}^2, \mathcal{B}_2, P_{XY}) = (\mathbf{R}, \mathcal{B}_1, P_X) \times (\mathbf{R}, \mathcal{B}_1, P_Y),$$

on el segon membre denota el producte d'espais de probabilitat, de manera que $P_{XY} = P_X \otimes P_Y$.

En general no és així: si X i Y no són independents, les mesures de probabilitat P_X, P_Y (o les funcions de distribució F_X, F_Y) no són suficients per conèixer P_{XY} .

3.5 Condicionament

- (3.5.1) Siguin $X, Y: \Omega \rightarrow \mathbf{R}$ variables aleatòries. Podem definir la **funció de distribució de X condicionada a $Y = y$** com la probabilitat condicionada

$$F_{X|Y}(x|y) \equiv F_X(x|Y=y) := P(X \leq x | Y=y),$$

que està definida en un parell de casos:

- quan Y és una variable discreta i $P(Y=y) \neq 0$;
- quan Y és una variable contínua amb densitat > 0 al voltant de y , llavors $P(X \leq x | Y=y) := \lim_{\Delta y \rightarrow 0} P(X \leq x | y \leq Y \leq y + \Delta y)$.

(3.5.2) En el cas que la variable bidimensional (X, Y) sigui contínua amb funció densitat conjunta f_{XY} , la funció de distribució condicionada es pot calcular com

$$F_X(x | Y=y) = \frac{\int_{-\infty}^x dx' f_{XY}(x', y)}{f_Y(y)},$$

de manera que la distribució condicionada admet una **funció de densitat condicionada**

$$f_{X|Y}(x|y) \equiv f_X(x | Y=y) := \frac{f_{XY}(x, y)}{f_Y(y)},$$

definida per a les y al voltant de les quals $f_Y > 0$.

(3.5.3) $f_X(\cdot | Y=y)$ és una densitat de probabilitat.

(3.5.4) X i Y són independents sii

$$f_X(x | Y=y) = f_X(x).$$

(3.5.5) *Esperança condicionada*

Les fórmules per al càlcul de l'esperança condicionada $E[X|E]$ poden aplicar-se en particular al cas d'un esdeveniment de la forma $Y = y$ quan la distribució condicionada a $Y = y$ està definida. Així obtenim l'**esperança de X condicionada a $Y=y$** , que representem per $E[X | Y=y]$.

Aquest és un nombre que depèn del valor y que hagi pres la variable aleatòria Y . Per tant, *podem considerar que és una variable aleatòria* funció de Y ; representem-la per $E[X|Y]$.

(3.5.6) En el cas que X i Y siguin conjuntament contínues l'esperança condicionada es calcula amb la fórmula

$$E[X | Y=y] = \int_{-\infty}^{+\infty} dx f_X(x | Y=y) x.$$

En el cas que siguin discretes,

$$E[X | Y=y] = \sum_{x \in \Omega_X} P(X=x | Y=y) x.$$

(3.5.7) Si $E[X]$ i $E[X|Y]$ estan definides,

$$E[X] = E[E[X|Y]].$$

En el cas que Y sigui contínua això és $E[X] = \int_{\mathbf{R}} dy f_Y(y) E[X|Y=y]$.

En el cas discret, $E[X] = \sum_{y \in \Omega_Y} P(Y=y) E[X|Y=y]$.

3.6 Funcions de diverses variables aleatòries. Suma de variables aleatòries

(3.6.1) Sigui $(X, Y): \Omega \rightarrow \mathbf{R}^2$ una variable aleatòria bidimensional. Si $g: \mathbf{R}^2 \rightarrow \mathbf{R}$ és una funció *mesurable Borel*, podem definir la variable aleatòria

$$Z = g(X, Y) \equiv g \circ (X, Y).$$

(3.6.2) Donat $z \in \mathbf{R}$, sigui $G_z = \{(x, y) \in \mathbf{R}^2 \mid g(x, y) \leq z\}$. Aleshores la funció de distribució de Z és

$$F_Z(z) = P(g(X, Y) \leq z) = P((X, Y) \in G_z).$$

En el cas que (X, Y) sigui contínua, amb funció de densitat conjunta f_{XY} obtenim

$$F_Z(z) = \iint_{G_z} dx dy f_{XY}(x, y).$$

(3.6.3) En particular, prenent $g(x, y) = x + y$, la **suma de variables aleatòries** $X + Y$ és una variable aleatòria.

(3.6.4) Si X, Y són variables aleatòries conjuntament contínues, $Z = X + Y$ és contínua amb densitat

$$f_Z(z) = \int_{\mathbf{R}} dx f_{XY}(x, z-x).$$

Si a més X, Y són *independents*, llavors

$$f_{X+Y} = f_X * f_Y$$

(**teorema de convolució**).

(3.6.5) Si $X_1 \sim \text{Pois}(\alpha_1)$, $X_2 \sim \text{Pois}(\alpha_2)$ són independents, llavors $X_1 + X_2 \sim \text{Pois}(\alpha_1 + \alpha_2)$.

Si $X_1 \sim N(m_1, \sigma_1^2)$, $X_2 \sim N(m_2, \sigma_2^2)$ són independents, aleshores $c_1 X_1 + c_2 X_2 \sim N(c_1 m_1 + c_2 m_2, \sqrt{c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2})$.

3.7 Transformacions i canvis de variable

(3.7.1) Sigui $\mathbf{X} = (X_1, X_2): \Omega \rightarrow \mathbf{R}^2$ una variable aleatòria bidimensional. Sigui $g: U \rightarrow V$ una aplicació mesurable Borel entre subconjunts de $\mathbf{R}^2$, amb

$\Omega_{(X_1, X_2)} \subset U$. Amb ella podem construir una nova variable aleatòria bidimensional $\mathbf{Y} = g \circ \mathbf{X}$, o

$$(Y_1, Y_2) = (g_1(X_1, X_2), g_2(X_1, X_2)).$$

- (3.7.2) Suposem que g és un difeomorfisme entre conjunts oberts de $\mathbf{R}^2$. Si (X_1, X_2) és contínua amb densitat $\mathbf{f}_{\mathbf{X}}$, aleshores (Y_1, Y_2) és contínua, i admet com a densitat la funció $\mathbf{f}_{\mathbf{Y}}$

$$\mathbf{f}_{\mathbf{Y}}(y) = \mathbf{f}_{\mathbf{X}}(g^{-1}(y)) |\det Jg^{-1}(y)|.$$

- (3.7.3) Un cas més general és aquell en què g no és injectiva. Sota certes condicions, en tal cas, al segon membre cal sumar les contribucions de les diverses antiimatges de y per g :

$$\mathbf{f}_{\mathbf{Y}}(y) = \sum_{\substack{x \\ g(x)=y}} \frac{\mathbf{f}_{\mathbf{X}}(x)}{|\det Jg(x)|}.$$

3.8 Esperança d'una funció de variables aleatòries

- (3.8.1) Sigui $(X, Y): \Omega \rightarrow \mathbf{R}^2$ una variable aleatòria bidimensional contínua, amb densitat conjunta $\mathbf{f}_{XY}$, i considerem una funció $Z = g(X, Y)$. Aleshores

$$\mathbf{E}[g(X, Y)] = \int_{\mathbf{R}^2} dx dy \mathbf{f}_{XY}(x, y) g(x, y).$$

Una expressió similar s'obté quan X i Y són discretes.

- (3.8.2) En particular,

$$\mathbf{E}[\alpha X + \beta Y] = \alpha \mathbf{E}[X] + \beta \mathbf{E}[Y].$$

- (3.8.3) Suposem que X i Y són *independents*. Aleshores $\mathbf{E}[h(X)k(Y)] = \mathbf{E}[h(X)] \mathbf{E}[k(Y)]$, i en particular $\mathbf{E}[XY] = \mathbf{E}[X] \mathbf{E}[Y]$.

3.9 Covariància i correlació

- (3.9.1) Siguin X, Y variables aleatòries amb esperances m_X, m_Y . La **covariància** de X i Y és

$$\mathbf{C}[X, Y] := \mathbf{E}[(X - m_X)(Y - m_Y)],$$

suposant que existeixi.

- (3.9.2) Propietats:

- i. $\mathbf{C}[X, Y] = \mathbf{E}[XY] - \mathbf{E}[X] \mathbf{E}[Y]$.
- ii. $\mathbf{C}[X, Y] = \mathbf{C}[Y, X]$.

- iii. $C[X, X] = V[X]$.
- iv. $C[\alpha X, Y] = \alpha C[X, Y]$.
- v. $C[X_1 + X_2, Y] = C[X_1, Y] + C[X_2, Y]$.

$$(3.9.3) \quad V[X + Y] = V[X] + V[Y] + 2C[X, Y].$$

(3.9.4) Suposem que $\sigma_X, \sigma_Y \neq 0$. El **coeficient de correlació** de X i Y és

$$\rho_{XY} := \frac{C[X, Y]}{\sigma_X \sigma_Y}.$$

$$(3.9.5) \quad |\rho_{XY}| \leq 1.$$

(3.9.6) Els **moments conjunts** de X i Y són

$$m_{jk} = E[X^j Y^k].$$

En particular, $m_{00} = 1$, $m_{10} = m_X$, $m_{01} = m_Y$.

(3.9.7) Els **moments centrals conjunts** de X i Y són

$$\mu_{jk} = E[(X - m_X)^j (Y - m_Y)^k].$$

En particular, $\mu_{00} = 1$, $\mu_{10} = \mu_{01} = 0$, $\mu_{20} = \sigma_X^2$, $\mu_{02} = \sigma_Y^2$, $\mu_{11} = C[X, Y]$.

(3.9.8) Algunes d'aquestes dades es poden organitzar en forma de

- **vector d'esperances** $\mathbf{m} = \begin{pmatrix} m_X \\ m_Y \end{pmatrix}$,
- **matriu de covariàncies** $C = \begin{pmatrix} C[X, X] & C[X, Y] \\ C[Y, X] & C[Y, Y] \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_Y \sigma_X & \sigma_Y^2 \end{pmatrix}$.

(3.9.9) Dues variables aleatòries X, Y es diuen **incorrelades** quan $\rho_{XY} = 0$.

Equivalentment, quan $C[X, Y] = 0$, o $E[XY] = E[X]E[Y]$.

(3.9.10) X i Y es diuen **ortogonals**⁴ quan $E[XY] = 0$. En tal cas escriurem $X \perp Y$.

(3.9.11) Tenim les relacions següents:

- i. X, Y incorrelades $\iff (X - m_X) \perp (Y - m_Y)$.
- ii. X, Y incorrelades $\iff V[X+Y] = V[X] + V[Y]$.
- iii. X, Y independents $\implies X, Y$ incorrelades.

3.10 Variable gaussiana bidimensional

(3.10.1) Una variable gaussiana bidimensional (X, Y) està definida per la funció densitat a $\mathbf{R}^2$

$$f(x, y) =$$

⁴ Vegeu (3.13.2).

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right)\right)$$

que depèn de cinc paràmetres: les mitjanes $m_1, m_2 \in \mathbf{R}$, les desviacions estàndard $\sigma_1, \sigma_2 > 0$, i la correlació $-1 < \rho < 1$.

(3.10.2) Les seves densitats marginals són gaussianes unidimensionals:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-m_1)^2}{\sigma_1^2}\right), \quad f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(y-m_2)^2}{\sigma_2^2}\right).$$

(3.10.3) La distribució de la variable X condicionada a $Y = y$ és una distribució normal amb esperança $m_1 + \rho\frac{\sigma_1}{\sigma_2}(y - m_2)$ i variància $\sigma_1^2(1 - \rho^2)$.

(Anàlogament Y condicionada a $X = x$.)

(3.10.4) $\rho_{XY} = \rho$.

(3.10.5) Si (X, Y) és una variable gaussiana bidimensional, X i Y són independents sii $\rho_{XY} = 0$.

(3.10.6) A partir de la gaussiana bidimensional (X, Y) podem construir les variables normals normalitzades $\frac{X-m_X}{\sigma_X}, \frac{Y-m_Y}{\sigma_Y}$, i amb elles la suma $Z = \frac{X-m_X}{\sigma_X} + \frac{Y-m_Y}{\sigma_Y}$ i la diferència $T = \frac{X-m_X}{\sigma_X} - \frac{Y-m_Y}{\sigma_Y}$. Aquestes són normals amb esperança 0 i variàncies $2(1 + \rho)$, $2(1 - \rho)$, respectivament. Z i T són incorrelades.

Les podem normalitzar fent

$$U = \frac{1}{\sqrt{2(1+\rho)}} \left(\frac{X-m_X}{\sigma_X} + \frac{Y-m_Y}{\sigma_Y} \right), \quad V = \frac{1}{\sqrt{2(1-\rho)}} \left(\frac{X-m_X}{\sigma_X} - \frac{Y-m_Y}{\sigma_Y} \right).$$

U, V són variables normals estàndard independents.

3.11 Variables aleatòries n -dimensionals

Recordem que una aplicació $\mathbf{X} = (X_1, \dots, X_n): \Omega \rightarrow \mathbf{R}^n$ és una variable aleatòria n -dimensional sii les seves components X_i són variables aleatòries unidimensionals.

En aquesta secció escriurem les expressions generals que en seccions anteriors hem escrit per al cas bidimensional.

(3.11.1) Si $B \subset \mathbf{R}^n$ és un borelià, podem definir

$$P_{\mathbf{X}}(B) \equiv P(\mathbf{X} \in B) := P(\mathbf{X}^{-1}(B)).$$

$P_{\mathbf{X}}: \mathcal{B}_n \rightarrow \mathbf{R}$ és una probabilitat en $(\mathbf{R}^n, \mathcal{B}_n)$.

(3.11.2) La **funció de distribució conjunta** de $\mathbf{X}$ és la funció que representarem per $F_{\mathbf{X}} \equiv F_{X_1 \dots X_n}: \mathbf{R}^n \rightarrow \mathbf{R}$, definida per

$$F_{\mathbf{X}}(x_1, \dots, x_n) := P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

- (3.11.3) Les **funcions de distribució marginals** són cadascuna de les funcions de distribució de les X_i , i s'obtenen fent $\rightarrow +\infty$ la resta de variables dins de la funció de distribució conjunta.

Més generalment, podem considerar funcions de distribució «intermèdies», com ara $F_{X_1 X_2}(x_1, x_2) = F_{\mathbf{X}}(x_1, x_2, +\infty, \dots, +\infty)$ (els $+\infty$ corresponen a fer el límit de les variables respectives).

- (3.11.4) La variable multidimensional $\mathbf{X}$ es diu **continua** quan la funció de distribució conjunta es pot expressar com

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} dx'_1 \dots dx'_n f_{\mathbf{X}}(x'_1, \dots, x'_n),$$

on $f_{\mathbf{X}} \equiv f_{X_1 \dots X_n} : \mathbf{R}^n \rightarrow \mathbf{R}$ és una funció integrable positiva, anomenada **funció de densitat conjunta**.

Llavors, per a tot borelià $B \subset \mathbf{R}^n$,

$$P(\mathbf{X} \in B) = \int_B d^n x f_{\mathbf{X}}(x_1, \dots, x_n).$$

- (3.11.5) Si $F_{\mathbf{X}}$ és de classe C^n , aleshores

$$f_{\mathbf{X}} = \frac{\partial^n F_{\mathbf{X}}}{\partial x_1 \dots \partial x_n}.$$

- (3.11.6) Les **funcions de densitat marginals** són les funcions de densitat corresponents a les distribucions de probabilitat marginals.

Es poden recuperar a partir de la funció de densitat conjunta integrant a tot $\mathbf{R}$ les altres variables. Per exemple

$$f_{X_1}(x_1) = \int_{\mathbf{R}} dx_2 \dots \int_{\mathbf{R}} dx_n f_{\mathbf{X}}(x_1, x_2, \dots, x_n).$$

Expressions similars s'apliquen als casos intermedis, per exemple

$$f_{X_1 X_2}(x_1, x_2) = \int_{\mathbf{R}} dx_3 \dots \int_{\mathbf{R}} dx_n f_{\mathbf{X}}(x_1, x_2, \dots, x_n).$$

- (3.11.7) Les variables aleatòries $X_1, \dots, X_n$ es diuen conjuntament **independents** quan, donats borelians $B_i \subset \mathbf{R}$ qualssevol, els esdeveniments $X_i \in B_i$ són independents. Això equival a afirmar que

$$F_{\mathbf{X}}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n).$$

En el cas continu això també equival a

$$f_{\mathbf{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n)$$

gairebé pertot.

- (3.11.8) Les **densitats condicionades** es defineixen dividint la densitat conjunta per les densitats marginals. Per exemple,

$$f_{X_1 \dots X_k}(x_1, \dots, x_k | X_{k+1}=x_{k+1}, \dots, X_n=x_n) = \frac{f_{X_1 \dots X_n}(x_1, \dots, x_n)}{f_{X_{k+1} \dots X_n}(x_{k+1}, \dots, x_n)}.$$

El primer membre a vegades s'escriu $f(x_1, \dots, x_k | x_{k+1}, \dots, x_n)$ de manera despreocupada. Amb aquesta notació observem que

$$f(x_1, \dots, x_n) = f(x_1 | x_2, \dots, x_n) f(x_2 | x_3, \dots, x_n) \dots f(x_{n-1} | x_n) f(x_n).$$

(3.11.9) Les expressions per a l'esperança d'una funció de n variables aleatòries, o per a la densitat d'una transformació d'una variable aleatòria multidimensional, són essencialment les mateixes que per a dues variables.

(3.11.10) Considerant $\mathbf{X}$ com un vector columna, el **vector d'esperances** es defineix com

$$m_{\mathbf{X}} = E[\mathbf{X}] = (m_{X_1} \dots m_{X_n})^{\top},$$

i la **matriu de covariàncies**

$$K_{\mathbf{X}} = E[(\mathbf{X} - m_{\mathbf{X}})(\mathbf{X} - m_{\mathbf{X}})^{\top}];$$

té components $k_{ij} = E[(X_i - m_{X_i})(X_j - m_{X_j})]$.

Observem que $k_{ij} = C[X_i, X_j]$.

(3.11.11) La matriu $K_{\mathbf{X}}$ és simètrica i semidefinida positiva.

$K_{\mathbf{X}}$ és definida positiva sii les variables $X_i - m_{X_i}$ són linealment independents gairebé segurament.

(3.11.12) Sigui A un matriu $m \times n$ constant, $\mathbf{Y} = A\mathbf{X}$. Aleshores

$$m_{\mathbf{Y}} = A m_{\mathbf{X}}, \quad K_{\mathbf{Y}} = A K_{\mathbf{X}} A^{\top}.$$

3.12 Variables gaussianes n -dimensionals

(3.12.1) La **distribució normal n -dimensional** té funció densitat en $\mathbf{R}^n$

$$f(\mathbf{x}) = \frac{(\det A)^{1/2}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^{\top} A (\mathbf{x} - \mathbf{m})\right)$$

essent els paràmetres $\mathbf{m} \in \mathbf{R}^n$ un vector i $A \in M_n(\mathbf{R})$ una matriu simètrica definida positiva.

Una variable aleatòria n -dimensional $\mathbf{X} = (X_1, \dots, X_n)$ es diu **gaussiana** o **normal** quan té la densitat anterior. Els paràmetres de la densitat es relacionen amb els paràmetres de $\mathbf{X}$ per

$$\mathbf{m} = m_{\mathbf{X}}, \quad A = K_{\mathbf{X}}^{-1}.$$

Totes les distribucions marginals són també normals, així com totes les distribucions condicionades.

Les variables X_i són independents sii són variables incorrelades sii A és una matriu diagonal.

3.13 Estimació de variables aleatòries

(3.13.1) L'objectiu de l'estimació és aproximar una variable aleatòria Y per una variable aleatòria $\hat{Y}$ pertanyent a una família predefinida. Aquesta $\hat{Y}$ s'anomena **estimador** de Y .

(3.13.2) El concepte d'*aproximació* requereix d'algun tipus de *distància*. El problema es pot plantejar en els termes següents.

El conjunt de les variables aleatòries definides en $(\Omega, \mathcal{F}, \mathbb{P})$ d'esperança i variància finites (o sigui, amb moments de segon ordre) forma un *espai de Hilbert* real (que podem denotar per simplicitat $L^2(\Omega)$), amb el producte escalar

$$(X|Y) := \mathbb{E}[XY].$$

(Tècnicament cal identificar dues variables $X \sim X'$ quan són iguals gairebé segurament.) Aquest producte escalar defineix la norma euclidiana $\|Z\|_2 \equiv \|Z\| = \mathbb{E}[Z^2]^{1/2}$, i la distància euclidiana corresponents.

Llavors utilitzarem com a criteri d'aproximació de Y per $\hat{Y}$ minimitzar l'**error quadràtic mitjà**,

$$\|Y - \hat{Y}\|^2 = \mathbb{E}[(Y - \hat{Y})^2].$$

(3.13.3) *Estimació per una constant*

La millor estimació de Y per $\hat{Y} = c$ ve donada per $\hat{Y} = \mathbb{E}[Y]$.

L'error quadràtic mitjà de l'estimació és $\mathbb{V}[Y]$.

(3.13.4) *Estimació no lineal*

Sigui X una variable aleatòria. La millor estimació de Y per $\hat{Y} = g(X)$ ve donada per la funció $g(x) = \mathbb{E}[Y|X=x]$.

La gràfica $y = g(x)$ s'anomena *corba de regressió*.

(3.13.5) En l'estimació no lineal podem considerar els casos extrems:

- Quan X, Y són independents $g(x) = \mathbb{E}[Y]$ és constant, de manera que la informació donada per X és irrellevant.
- Quan $Y = h(X)$, $g(x) = h(x)$. Òbviament X determina Y completament.

Interessen els casos intermedis, on X pot aportar informació rellevant, però no tota.

(3.13.6) *Estimació lineal*

Sigui X una variable aleatòria. La millor estimació de Y per $\hat{Y} = aX + b$ (a, b constants) ve donada per

$$a = \rho_{XY} \frac{\sigma_Y}{\sigma_X}, \quad b = m_Y - a m_X.$$

L'error quadràtic mitjà és $\sigma_Y^2(1 - \rho_{XY}^2)$.

La corba de regressió ara és una *recta de regressió*, $y = ax + b$.

(3.13.7) A priori pot semblar que l'estimació lineal és menys precisa que la no lineal. Tanmateix, l'estimació lineal té l'avantatge que només requereix conèixer els paràmetres bàsics de les variables, així com el coeficient de correlació.

Si les variables (X, Y) són conjuntament normals, llavors l'aproximació no lineal coincideix amb la lineal.

(3.13.8) *Estimació lineal per diverses variables aleatòries*

Siguin $X_1, \dots, X_n$ variables aleatòries. En aquest cas busquem un estimador de la forma

$$\hat{Y} = a_1 X_1 + \dots + a_n X_n.$$

Les variables $X_1, \dots, X_n$ formen un subespai de dimensió finita $\langle X_1, \dots, X_n \rangle$ dins de $L^2(\Omega)$. Com sabem, la millor aproximació $\hat{Y}$ de Y per un element d'aquest subespai és precisament la seva projecció ortogonal, que ve caracteritzada pel *principi d'ortogonalitat*:

$$(Y - \hat{Y} | X_j) = 0 \quad (j = 1..n)$$

Això dona un sistema lineal per als coeficients a_i :

$$\sum_i E[X_j X_i] a_i = E[Y X_j] \quad (j = 1..n)$$

que té una única solució quan la matriu és invertible, cosa que passa quan les X_i són linealment independents.

L'error quadràtic mitjà corresponent a aquesta solució és $\|Y - \hat{Y}\|^2 = E[Y^2] - E[\hat{Y}^2]$.

(3.13.9) A vegades s'anomena *estimació lineal no homogènia* una del tipus $\hat{Y} = a_1 X_1 + \dots + a_n X_n + b$, però en realitat és una estimació lineal amb $X_{n+1} = 1$.

3.14 Convergència de variables aleatòries

En aquesta secció X i X_n ($n \in \mathbb{N}$) són variables aleatòries unidimensionals definides en un espai mostral Ω .

Hi ha diverses nocions de convergència de la successió (X_n) . Les descriurem succintament.

(3.14.1) **Convergència en distribució**

$X_n \xrightarrow{d} X$ quan, en tot punt x on F_X és contínua, $\lim_n F_{X_n}(x) = F_X(x)$.

També s'anomena *convergència feble*.

(3.14.2) **Convergència en probabilitat**

$X_n \xrightarrow{P} X$ quan, per a tot $\varepsilon > 0$, $\lim_n P(|X_n - X| > \varepsilon) = 0$.

El límit també es pot expressar com $\lim_n P(|X_n - X| \leq \varepsilon) = 1$.

(3.14.3) **Convergència gairebé segura**

$X_n \xrightarrow{gs} X$ quan $P(\{\omega \in \Omega \mid \lim_n X_n(\omega) = X(\omega)\}) = 1$.

També s'anomena *convergència forta*.

(3.14.4) **Convergència en la norma L^r ($r \geq 1$)**

$X_n \xrightarrow{L^r} X$ quan $\lim_n E[|X - X_n|^r] = 0$.

Els casos més importants són $r = 1$ (convergència en mitjana) i $r = 2$ (convergència en mitjana quadràtica).

(3.14.5) Les relacions més senzilles entre aquests tipus de convergència són:

conv. gairebé segura $\implies$ conv. en probabilitat $\implies$ conv. en distribució

conv. $L^r \implies$ conv. en probabilitat

si $r > s \geq 1$, conv. $L^r \implies$ conv. L^s

(3.14.6) Si $X_1, \dots, X_n$ són variables aleatòries, la seva **mitjana mostral**

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n},$$

també és una variable aleatòria. Si les X_i tenen esperances i variàncies finites,

$$E[\bar{X}_n] = \frac{\sum_i E[X_i]}{n}, \quad V[\bar{X}_n] = \frac{1}{n^2} \sum_i V[X_i] + \frac{2}{n^2} \sum_{i < j} C[X_i, X_j].$$

Si totes les variables són *independents i idènticament distribuïdes* (iid), amb esperança μ i variància σ^2 , llavors

$$E[\bar{X}_n] = \mu, \quad V[\bar{X}_n] = \frac{\sigma^2}{n}.$$

La **mitjana mostral normalitzada** és la variable normalitzada de la mitjana mostral. En el cas iid és

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_i \frac{X_i - \mu}{\sigma}.$$

(3.14.7) **Llei feble dels grans nombres (teorema de Khintxin)**

Sigui $(X_i)_{i \geq 1}$ una successió de variables aleatòries independents i idènticament distribuïdes, amb esperança finita μ . Aleshores la successió de mitjanes mostrals $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ convergeix en probabilitat cap a l'esperança μ :

$$\text{per a tot } \varepsilon > 0, \lim_n \mathbf{P} \left(|\bar{X}_n - \mu| > \varepsilon \right) = 0.$$

Hi ha altres versions d'aquesta llei que no requereixen que les X_i tinguin mateixa distribució, però llavors requereixen hipòtesis addicionals.

(3.14.8) La **llei forta dels grans nombres**, en la versió provada per Kolmogorov, té el mateix enunciat que la llei feble anterior, però estableix la convergència gairebé segura, que és una noció més forta.

També aquí hi ha altres versions de la llei que, amb hipòtesis addicionals, no requereixen que les X_i tinguin mateixa distribució.

En situacions més generals, doncs, pot donar-se la convergència en probabilitat sense que es doni la convergència gairebé segura.

(3.14.9) **Teorema del límit central (Laplace; Lindeberg–Lévy)**

Siguin $(X_i)_{i \geq 1}$ variables aleatòries independents idènticament distribuïdes, amb esperança μ i variància $\sigma^2 \neq 0$ finites. Sigui $\bar{X}_n$ la mitjana mostral n -èsima de les X_i . Aleshores

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathbf{N}(0, 1).$$

Això significa que, per a cada $z \in \mathbf{R}$, $\lim_n \mathbf{P} \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z \right) = \Phi(z)$, on

$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z dt e^{-t^2/2}$ és la funció de distribució d'una variable normal estàndard.

4 Estadística

4.1 Més paràmetres de les variables aleatòries

Si X és una variable aleatòria, ja en coneixem l'esperança $\mu = E[X]$, que és un paràmetre de centralització, i la variància $\sigma^2 = V[X]$, que és un paràmetre de dispersió. Hi ha altres paràmetres interessants que descrivim tot seguit.

(4.1.1) *Asimetria*

El **coeficient d'asimetria** de X és

$$\kappa_3 = \frac{E[(X - \mu)^3]}{\sigma^3}.$$

Si la densitat de X és simètrica al voltant de μ la asimetria val zero.

(4.1.2) *Curtosi*

El **coeficient de curtosi** de X és

$$\kappa_4 = \frac{E[(X - \mu)^4]}{\sigma^4}.$$

Un valor gran significa que les «cues» de la distribució són pesants.

Com que una variable normal té curtosi 3, a vegades aquest valor es pren com a referència i es diu que $\kappa_4 - 3$ és l'**excés de curtosi**.

Com és obvi la definició d'aquests paràmetres requereix l'existència dels moments fins a ordre 3 o 4, i que la variància no sigui zero.

(4.1.3) *Moda*

Una **moda** de X és un valor x on la funció densitat de X té un màxim local. Les variables més usals tenen una única moda, però una variable pot tenir diverses modes.

Una definició semblant es pot donar en el cas discret.

(4.1.4) *Mediana*

La mediana és un nombre que divideix $\mathbf{R}$ en dues regions equiprobables.

En general, la **mediana** d'una variable aleatòria X és un nombre m tal que $P(X \leq m) \geq 1/2$ i $P(X \geq m) \geq 1/2$.

En el cas d'una variable *contínua* això significa que m és solució de

$$F_X(m) = 1/2.$$

Aquesta equació té sovint una única solució. Si el conjunt de solucions és un interval, es pot prendre per conveni el seu punt mig com a mediana.

(4.1.5) *Quantils*

Els ***k*-quantils** són nombres que divideixen $\mathbf{R}$ en k regions equiprobables; també s'anomenen *quantils* les regions obtingudes d'aquesta manera.

Els casos més comuns són $k = 2$ (mediana), $k = 4$ (quartils), $k = 10$ (decils), $k = 100$ (percentils).

(4.1.6) *La funció quantil*

Considerem el cas d'una variable aleatòria contínua, amb funció de distribució $F_X: \mathbf{R} \rightarrow [0, 1]$; recordem que és contínua i creixent. Excloent els valors de probabilitat 0 i 1, l'aplicació $F_X: S \rightarrow]0, 1[$, definida en un interval obert $S \subset \mathbf{R}$ apropiat, és a més suprajectiva.

Si aquesta funció és estrictament creixent llavors és bijectiva. La seva inversa és la ***funció quantil*** $Q_X:]0, 1[\rightarrow S \subset \mathbf{R}$. Per definició,

$$Q_X(F_X(x)) = x, \quad F_X(Q_X(p)) = p.$$

Hi ha una definició general de Q_X , vàlida per a qualsevol F_X ; ometem donar-ne els detalls.

- (4.1.7) Si disposem de la funció quantil aleshores és fàcil expressar els quantils. Per exemple, els quartils són els valors $Q_1 = Q_X(1/4)$, $Q_2 = Q_X(2/4)$ (mediana) i $Q_3 = Q_X(3/4)$.

S'anomena ***recorregut interquartílic*** la diferència $IQR = Q_3 - Q_1$.

Els ***valors atípics*** o ***outliers*** d'una distribució són els que estan més allunyats dels valors centrals. Per conveni es diu que X és un valor atípic quan $X < Q_1 - \frac{3}{2} IQR$ o $X > Q_3 + \frac{3}{2} IQR$.

- (4.1.8) Sigui X una variable aleatòria contínua amb funció de distribució F_X estrictament creixent, de manera que està definida la funció quantil $Q_X:]0, 1[\rightarrow S$. Si U és una variable del tipus $\text{Unif}(0, 1)$, aleshores la variable

$$Q_X(U)$$

té la mateixa distribució que X .

Per tant, qualsevol distribució que compleixi les condicions esmentades es pot generar de manera simple a partir d'una distribució uniforme.

4.2 Funció característica d'una variable aleatòria

(4.2.1) Sigui X una variable aleatòria (real). Per a cada $\omega \in \mathbf{R}$, la funció $e^{i\omega X} = \cos \omega X + i \sin \omega X$ és una *variable aleatòria complexa*.

(4.2.2) La **funció característica** de X és la funció $C_X: \mathbf{R} \rightarrow \mathbf{C}$ definida per

$$C_X(\omega) = E[e^{i\omega X}] .$$

(4.2.3) En el cas que X sigui una variable aleatòria contínua, pel teorema de l'esperança

$$C_X(\omega) = \int_{\mathbf{R}} dx f_X(x) e^{i\omega x} ,$$

que és, llevat d'un signe i un factor 2π , la transformada de Fourier de la funció de densitat.

Com que f_X és integrable, $C_X(\omega)$ existeix per a tota $\omega \in \mathbf{R}$. La funció C_X és contínua.

(4.2.4) Si $\mathbf{X} = (X_1, \dots, X_n)$ és una variable aleatòria n -dimensional es defineix $C_{\mathbf{X}}: \mathbf{R}^n \rightarrow \mathbf{C}$ com

$$C_{\mathbf{X}}(\omega_1, \dots, \omega_n) = E[\exp(i(\omega_1 X_1 + \dots + \omega_n X_n))] ,$$

que en el cas continu és essencialment una transformada de Fourier n -dimensional

$$C_{\mathbf{X}}(\omega_1, \dots, \omega_n) = \int_{\mathbf{R}^n} dx_1 \dots dx_n f_{\mathbf{X}}(x_1, \dots, x_n) e^{i(\omega_1 x_1 + \dots + \omega_n x_n)} .$$

(4.2.5) Igual que passa amb la transformació de Fourier, a partir de la funció característica es pot reconstruir, sota hipòtesis adequades, la funció de densitat.

(4.2.6) Algunes propietats:

- i. $C_X(0) = 1$.
- ii. $|C_X(\omega)| \leq 1$.
- iii. El moment k -èsim de X , si existeix, es pot calcular derivant C_X :

$$E[X^k] = (-i)^k \frac{d^k C_X}{d\omega^k}(0) .$$

- iv. Les variables $X_1, \dots, X_n$ són independents sii $C_{\mathbf{X}}(\omega_1, \dots, \omega_n) = C_{X_1}(\omega_1) \dots C_{X_n}(\omega_n)$.

- v. Si $Z = X_1 + \dots + X_n$, on les X_i són independents,

$$C_Z(\omega) = C_{X_1}(\omega) \dots C_{X_n}(\omega)$$

(teorema de convolució).

(4.2.7) *Funció característica d'una variable normal*

Si $X \sim N(\mu, \sigma)$,

$$C_X(\omega) = \exp\left(i\mu\omega - \frac{1}{2}\sigma^2\omega^2\right).$$

4.3 Algunes distribucions associades a la normal

(4.3.1) *Distribució gamma*

$X \sim \text{Gamma}(\alpha, \beta)$, amb paràmetres $\alpha > 0$ (*forma*) i $\beta > 0$ (*escala*).

Està definida per la densitat

$$f_X(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0.$$

Observem que $\text{Gamma}(1, \beta) \sim \text{Exp}(1/\beta)$.

(4.3.2) Moments:

$$E[X^k] = \beta^k \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)}.$$

En particular

$$E[X] = \alpha\beta, \quad V[X] = \alpha\beta^2.$$

Funció característica:

$$C_X(\omega) = \frac{1}{(1 - i\beta\omega)^\alpha}.$$

(4.3.3) Si $X \sim \text{Gamma}(\alpha_1, \beta)$, $Y \sim \text{Gamma}(\alpha_2, \beta)$ són independents,

$$X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta).$$

La suma de n variables $\text{Exp}(\lambda)$ independents és $\text{Gamma}(n, 1/\lambda)$.

(4.3.4) Siguin X, Y independents, i $Z = X + Y$. Si $Z \sim \text{Gamma}(\alpha_Z, \beta)$ i $Y \sim \text{Gamma}(\alpha_Y, \beta)$, llavors $X \sim \text{Gamma}(\alpha_Z - \alpha_Y, \beta)$.

(4.3.5) Si $X \sim N(\mu, \sigma)$, llavors $\left(\frac{X - \mu}{\sigma}\right)^2 \sim \text{Gamma}\left(\frac{1}{2}, 2\right)$.

(4.3.6) *Distribució khi quadrat*

$X \sim \chi_n^2$, amb paràmetre $n \in \mathbb{N}^*$ (*graus de llibertat*).

Una variable aleatòria amb aquesta distribució s'obté fent la suma $X = Z_1^2 + \dots + Z_n^2$ on les Z_i són $N(0, 1)$ independents. Per tant

$$\chi_n^2 \sim \text{Gamma}(n/2, 2),$$

$$f_X(x) = \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2}, \quad x > 0.$$

(4.3.7) Els moments són

$$\begin{aligned} E[X^k] &= \frac{(n+2k-2)!!}{(n-2)!!}; \\ E[X] &= n, \quad V[X] = 2n. \end{aligned}$$

(4.3.8) Quan n és gran ($n \geq 50$) la distribució χ_n^2 es pot aproximar bé per una $N(n, \sqrt{2n})$.

(4.3.9) **Distribució t de Student**

$T \sim t_n$, $n \in \mathbf{N}^*$ (*graus de llibertat*).

Una variable T de Student s'obté prenent $X \sim \chi_n^2$ i $Y \sim N(0, 1)$ independents, i definint

$$T = \frac{Y}{\sqrt{X/n}}.$$

La seva densitat és la funció parella

$$f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n} \Gamma(\frac{n}{2})} \frac{1}{(1 + \frac{t^2}{n})^{\frac{n+1}{2}}}.$$

Quan $n = 1$ obtenim $t_1 \sim \text{Cauchy}(1)$.

(4.3.10) Quan $t \rightarrow \pm\infty$, $f_T(t) \sim \frac{\text{const}}{|t|^{n+1}}$, per tant T només té moments d'ordre $k < n$.

(4.3.11) La gràfica de la densitat f_T és semblant a la de la normal estàndard però amb cues més pesants.

Quan $n \rightarrow \infty$ tenim el límit $\lim_{n \rightarrow \infty} f_{t_n}(t) = f_{N(0,1)}(t)$.

4.4 Poblacions i mostres

(4.4.1) Una **població** és un fragment de la realitat format per un conjunt nombrós d'elements (o individus) amb unes característiques que admeten una descripció probabilística. També pot ser el model matemàtic que descriu la població real (en aquest cas, pot ser finita o infinita).

Aquestes característiques poden ser no numèriques, però ens centrarem en el cas *numèric*.

(4.4.2) Els objectius de l'**estadística inferencial** són trobar el model probabilístic que descriu una població (amb els seus paràmetres) i verificar si el model proposat és correcte.

- (4.4.3) Una **mostra** d'una població és un subconjunt finit de la població que és objecte d'observació. La **mida** de la mostra és el seu nombre d'elements. En una **mostra aleatòria simple** aquest subconjunt es tria a l'atzar, i de manera que totes les mostres de la mateixa mida tinguin la mateixa probabilitat de ser seleccionada.
- (4.4.4) La mostra té associats uns valors $x_1, \dots, x_n$ de la característica numèrica considerada (dades). Podem considerar que aquests valors són el resultat d'avaluar n variables aleatòries $X_1, \dots, X_n$, totes elles amb la mateixa distribució de probabilitat. Si es fes el mostratge amb reposició (la qual cosa vol dir que eventualment es podria repetir un individu) llavors les X_i serien **independents i idènticament distribuïdes** (iid). Aquest és el marc teòric més simple i dins del qual treballarem. En realitat, les X_i no poden ser independents, ja que quan es trien elements d'una població finita es perd la independència. Tanmateix, si la població és prou gran aquest efecte no es notarà.
- (4.4.5) Si en lloc d'una característica numèrica en recollim diverses, aleshores utilitzarem variables aleatòries multidimensionals.

4.5 Estadístics

- (4.5.1) Suposem que un cert paràmetre de la població ve descrit per una variable aleatòria X . Considerem una mostra $(X_1, \dots, X_n)$ amb les X_i variables aleatòries independents idènticament distribuïdes, amb la mateixa distribució que X . Un **estadístic** és una magnitud que es calcula a partir dels valors de la mostra. Per tant, és una variable aleatòria de la forma
- $$Z = \varphi(X_1, \dots, X_n).$$
- Un estadístic es pot utilitzar per estimar un paràmetre d'una població, descriure les dades d'una mostra, o contrastar una hipòtesi. Els estadístics més importants són la mitjana mostral i la variància mostral.

- (4.5.2) *Mitjana mostral*

La **mitjana mostral** és

$$\bar{X}_n \equiv \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

(4.5.3) Si X té esperança μ i variància σ^2 ,

$$\mathbb{E}[\bar{X}_n] = \mu, \quad \mathbb{V}[\bar{X}_n] = \frac{\sigma^2}{n}.$$

(4.5.4) Si la població és normal $N(\mu, \sigma)$, llavors

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Si la població no és normal però n és gran ($n \geq 30$) llavors el teorema del límit central ens diu que $\bar{X}_n$ es comporta com una variable normal $N(\mu, \sigma/\sqrt{n})$.

(4.5.5) *Variància mostral*

La **variància mostral esbiaixada** és

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

Atenció: hi ha diferent terminologia i notació per a les variàncies mostrals!

(4.5.6) Si X té esperança μ , variància σ^2 i curtosi κ_4 ,

$$\mathbb{E}[\tilde{S}^2] = \frac{n-1}{n} \sigma^2, \quad \mathbb{V}[\tilde{S}^2] = \sigma^4 \frac{n-1}{n^2} \left(\kappa_4 - 1 + \frac{3-\kappa_4}{n} \right).$$

(4.5.7) La **variància mostral corregida**, o simplement **variància mostral**, és

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

Aquesta definició aplica un factor $n/(n-1)$ (correcció de Bessel) a la variància mostral sense corregir, de manera que la nova variància mostral té esperança σ^2 :

$$\mathbb{E}[S^2] = \sigma^2, \quad \mathbb{V}[S^2] = \frac{\sigma^4}{n-1} \left(\kappa_4 - 1 + \frac{3-\kappa_4}{n} \right).$$

(4.5.8) La **desviació mostral** és l'arrel quadrada de la variància mostral: $S = \sqrt{S^2}$.

(4.5.9) Siguin $X_1, \dots, X_n$ variables independents idènticament distribuïdes $N(\mu, \sigma)$. Aleshores:

- i. $\bar{X}_n \sim N(\mu, \sigma/\sqrt{n})$.
- ii. $\bar{X}_n$ i S^2 són independents.
- iii. $(n-1) \frac{S^2}{\sigma^2} \sim \chi_{n-1}^2$.
- iv. $\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t_{n-1}$.

(4.5.10) *Moments mostrals*

El **moment mostral** k -èsim es defineix per

$$M_k = \frac{1}{n} \sum_i X_i^k.$$

Si $E[X^k] = \mu_k$, aleshores

$$E[M_k] = \mu_k, \quad V[M_k] = \frac{\mu_{2k} - \mu_k^2}{n}.$$

De manera similar es poden definir els moments mostrals centrats.

(4.5.11) *Quantils mostrals*

Els quantils mostrals s'obtenen ordenant els valors X_i en ordre creixent, i tallant-los en el punt adequat.

El cas més senzill és el de la **mediana mostral**, a vegades representada per $\tilde{X}_n$.

Si n és imparell, és el valor en la posició $(n+1)/2$; si n és parell, es pren la mitjana dels valors en les posicions $n/2$ i $n/2 + 1$.

Els quantils són útils quan la variable no té esperança, i són més robustos davant la presència de dades anòmales.

(4.5.12) *Mínim i màxim mostrals*

Són els valors extrems dels X_i .

4.6 Estimadors

(4.6.1) Considerem una població amb unes característiques modelades per una variable aleatòria X . Aquest model pot tenir associats uns paràmetres, o més generalment quantitats que depenguin d'aquests paràmetres. Sigui θ una d'aquestes quantitats.

Un **estimador** de θ és un estadístic $\hat{\theta}(X_1, \dots, X_n)$ usat per a *estimar* el valor de θ a partir d'una mostra $(X_1, \dots, X_n)$.

El valor que pren $\hat{\theta}$ sobre una mostra concreta $(x_1, \dots, x_n)$ és una **estimació** de θ .

(4.6.2) El **biaix** d'un estimador $\hat{\theta}$ de θ es defineix com

$$B_{\hat{\theta}} = E[\hat{\theta} - \theta].$$

Un estimador es diu **no esbiaixat**, o **centrat**, quan no té biaix:

$$E[\hat{\theta}] = \theta.$$

(4.6.3) L'**error quadràtic mitjà** de l'estimador $\hat{\theta}$ és

$$\text{MSE}_{\hat{\theta}} = \text{E} \left[(\hat{\theta} - \theta)^2 \right].$$

$$(4.6.4) \quad \text{E} \left[(\hat{\theta} - \theta)^2 \right] = \text{V}[\hat{\theta}] + \text{B}_{\hat{\theta}}^2.$$

(4.6.5) Un estimador es diu **consistent** quan

$$\lim_n \text{E} \left[(\hat{\theta} - \theta)^2 \right] = 0.$$

(4.6.6) Per tant l'estimador és consistent si $\lim_n \text{V}[\hat{\theta}] = 0$ i $\lim_n \text{B}_{\hat{\theta}}^2 = 0$.

Per a estimadors no esbiaixats la consistència equival a $\lim_n \text{V}[\hat{\theta}] = 0$.

(4.6.7) Si $\lim_n \text{E} \left[(\hat{\theta} - \theta)^2 \right] = 0$ aleshores, per a cada $\varepsilon > 0$, $\lim_n \text{P} \left(|\hat{\theta} - \theta| < \varepsilon \right) = 1$.

(4.6.8) Donats dos estimadors de θ , diem que és **més eficient** aquell que té menor $\text{E} \left[(\hat{\theta} - \theta)^2 \right]$.

(4.6.9) La mitjana mostral és un estimador no esbiaixat i consistent de l'esperança de la població.

La variància mostral *corregida* és un estimador no esbiaixat i consistent de la variància de la població.

4.7 Mètodes d'estimació: mètode dels moments

(4.7.1) Se suposa que una característica de la població està modelada per una variable aleatòria X que depèn de k paràmetres $\theta = (\theta_1, \dots, \theta_k)$.

Suposem també que els moments m_ℓ d'ordres $1, \dots, k$ de X es poden calcular. Aquests moments òbviament depenen de θ :

$$\begin{aligned} m_1 &= \text{E}(X) = g_1(\theta_1, \dots, \theta_k), \\ m_2 &= \text{E}(X^2) = g_2(\theta_1, \dots, \theta_k), \\ &\vdots \\ m_k &= \text{E}(X^k) = g_k(\theta_1, \dots, \theta_k). \end{aligned}$$

(4.7.2) Suposem d'altra banda que tenim una mostra de mida n $(X_1, \dots, X_n)$. Amb ella podem calcular els moments mostrals d'ordre ℓ , amb $1 \leq \ell \leq k$:

$$\hat{m}_\ell = \frac{1}{n} \sum_{i=1}^n X_i^\ell.$$

- (4.7.3) El mètode dels moments consisteix a igualar uns amb altres, de manera que s'obté el sistema de k equacions

$$g_\ell(\hat{\theta}_1, \dots, \hat{\theta}_k) = \frac{1}{n} \sum_{i=1}^n X_i^\ell \quad (1 \leq \ell \leq k);$$

aïllar-ne $\hat{\theta}_1, \dots, \hat{\theta}_k$ com a funcions de $(X_1, \dots, X_n)$, i prendre $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ com a estimador de $(\theta_1, \dots, \theta_k)$.

4.8 Mètodes d'estimació: mètode de la màxima versemblança

- (4.8.1) Considerem un model estadístic dependent de k paràmetres $\theta = (\theta_1, \dots, \theta_k) \in \Theta \subset \mathbf{R}^k$, on la densitat de probabilitat d'obtenir un resultat $\mathbf{x} = (x_1, \dots, x_n) \in \mathbf{R}^n$ ve donada per $f(\mathbf{x}; \theta) = f(x_1, \dots, x_n; \theta_1, \dots, \theta_k)$. Avaluant aquesta expressió en les dades $\mathbf{x}$ obtingudes obtenim una funció de θ , anomenada **versemblança**. A vegades es representa per $\mathcal{L}(\theta; \mathbf{x}) = f(\mathbf{x}; \theta)$.

- (4.8.2) Per a cada valor fixat de $\mathbf{x}$, és possible que hi hagi un valor $\hat{\theta}(\mathbf{x})$ de $\theta \in \Theta$ que faci **màxima** la versemblança. D'aquesta manera obtenim una funció $\hat{\theta}: \mathbf{R}^n \rightarrow \Theta$, i, si es compleixen els requisits tècnics adequats, obtenim un estimador

$$\hat{\theta}(X_1, \dots, X_n)$$

de θ . És l'**estimador de màxima versemblança** de θ .

- (4.8.3) L'existència de $\hat{\theta}(\mathbf{x})$ està assegurada per exemple quan Θ és compacte i la funció (de θ) $\mathcal{L}(\cdot; \mathbf{x})$ és contínua.

Si Θ és obert i $\mathcal{L}(\cdot; \mathbf{x})$ és diferenciable, una condició necessària perquè $\hat{\theta}(\mathbf{x})$ sigui un màxim de la versemblança és que en sigui un *punt crític*:

$$\frac{\partial}{\partial \theta_r} f(\mathbf{x}; \theta) = 0 \quad (1 \leq r \leq k) \quad \text{quan } \theta = \hat{\theta}(\mathbf{x}).$$

- (4.8.4) Maximitzar la funció $\mathcal{L}(\cdot; \mathbf{x})$ equival a maximitzar el seu logaritme. A vegades s'anomena **log-versemblança** la funció

$$\ell(\theta; \mathbf{x}) = \log f(\mathbf{x}; \theta),$$

definida allà on la densitat sigui estrictament positiva. Llavors el valor θ de màxima versemblança ha de satisfer les *equacions de versemblança*

$$\frac{\partial}{\partial \theta_r} \ell(\theta; \mathbf{x}) = 0 \quad (1 \leq r \leq k).$$

Recordem que una condició suficient per assegurar que s'ateny un màxim local és que la hessiana

$$\left(\frac{\partial^2 \ell(\theta; \mathbf{x})}{\partial \theta_r \partial \theta_s} \right)$$

sigui *definida negativa* en el punt crític.

- (4.8.5) Suposem (com és habitual) que la mostra $\mathbf{x}$ està modelada per n variables independents idènticament distribuïdes $X_1, \dots, X_n$, totes elles amb densitat $f(x; \theta)$. Aleshores $f(\mathbf{x}; \theta) = f(x_1; \theta) \dots f(x_n; \theta)$, de manera que

$$\ell(\theta; \mathbf{x}) = \sum_{i=1}^n \log f(x_i; \theta).$$

4.9 Informació de Fisher i fita de Cramér-Rao

- (4.9.1) Considerem un model estadístic dependent (per simplicitat) d'un sol paràmetre θ i amb funció log-versemblança $\ell(\theta; \mathbf{x}) = \log f(\mathbf{x}; \theta)$. La seva derivada també és una variable aleatòria $\frac{\partial \ell}{\partial \theta}(\theta; X_1, \dots, X_n)$.

- (4.9.2) Si $G = g(X_1, \dots, X_n; \theta)$ aleshores

$$\mathbb{E} \left[G \frac{\partial \ell}{\partial \theta} \right] = \frac{\partial}{\partial \theta} \mathbb{E}[G] - \mathbb{E} \left[\frac{\partial G}{\partial \theta} \right].$$

- (4.9.3) $\mathbb{E} \left[\frac{\partial \ell}{\partial \theta} \right] = 0.$ $\mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right] = \mathbb{E} \left[-\frac{\partial^2 \ell}{\partial \theta^2} \right].$

- (4.9.4) Es defineix la **informació de Fisher** com

$$I_F(\theta) = \mathbb{E} \left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right].$$

- (4.9.5) $\mathbb{V} \left[\frac{\partial \ell}{\partial \theta} \right] = I_F.$

- (4.9.6) **Desigualtat de Cramér-Rao**

Si $G(X_1, \dots, X_n),$

$$\mathbb{V}[G] \geq \frac{\left(\frac{\partial}{\partial \theta} \mathbb{E}[G] \right)^2}{I_F}.$$

- (4.9.7) **Fita de Cramér-Rao**

Sigui $\hat{\theta}$ un estimador de θ , amb biaix $B_{\hat{\theta}}$. Aleshores

$$\mathbb{V}[\hat{\theta}] \geq \frac{(1 + \partial B_{\hat{\theta}} / \partial \theta)^2}{I_F(\theta)}.$$

Si $\hat{\theta}$ no té biaix,

$$V[\hat{\theta}] \geq \frac{1}{I_F(\theta)}.$$

- (4.9.8) Un estimador no esbiaixat es diu **eficient** quan aquesta desigualtat és una igualtat (la variància és la mínima possible).

4.10 Intervals de confiança

En els apartats anteriors hem estudiat l'*estimació puntual*: a partir d'una mostra aleatòria $(X_1, \dots, X_n)$ obteníem un valor estimat $\hat{\theta}(\mathbf{x})$ d'un paràmetre desconegut θ . L'*estimació intervalar* dona com a resultat $\theta \in C_{\mathbf{x}}$, on $C_{\mathbf{x}} \subset \Theta \subset \mathbf{R}$ és un interval estimat a partir de la mostra. El que es perd en precisió es guanya en *confiança*.

- (4.10.1) Considerem una mostra $\mathbf{X} = (X_1, \dots, X_n)$ de mida n d'una població, on les X_i són variables iid corresponents a una distribució de probabilitat dependent d'un cert paràmetre θ (i eventualment d'altres paràmetres).

Un **interval de confiança** per a θ , amb **nivell de confiança** c , és un interval $[L(\mathbf{X}), U(\mathbf{X})]$, dependent de les dades mostrals, tal que

$$P_{\theta} \left(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X}) \right) = c,$$

per a qualsevol valor de θ .

Notem que θ no és una variable aleatòria, sinó un paràmetre concret (per bé que desconegut). Els extrems de l'interval sí que són variables aleatòries.

En algunes ocasions es busca un interval estimat semiinfinit, ja sigui $]-\infty, U(\mathbf{X})]$, o $[L(\mathbf{X}), +\infty[$.

- (4.10.2) Valors típics de confiança que es prenen: $c = 0.9$, $c = 0.95$, $c = 0.99$.

En lloc de c s'utilitza sovint $1 - \alpha$. Llavors $\alpha = 0'1, 0'05, 0'01$.

- (4.10.3) Cada realització concreta $\mathbf{x}$ de la mostra dona lloc a un **interval estimat** $[L(\mathbf{x}), U(\mathbf{x})]$ concret. Els extrems de l'interval a vegades es diuen **límits de confiança**.

La interpretació de l'interval de confiança és que, en una proporció c de vegades que s'adquireix la mostra, l'interval de confiança obtingut conté el valor del paràmetre θ .

En canvi, seria confús afirmar que «hi ha una probabilitat c que el valor real de θ estigui dins de l'interval de confiança obtingut».

- (4.10.4) El procediment més habitual per a construir intervals de confiança consisteix a construir una nova variable aleatòria $W = W(\mathbf{X}, \theta)$ amb distribució de probabilitat coneguda i *independent de* θ , de manera que podem calcular fàcilment

$$P_{\theta} (w_1 \leq W(\mathbf{X}, \theta) \leq w_2) .$$

Si trobem w_1, w_2 tals que aquesta probabilitat valgui c , llavors el conjunt

$$C_{\mathbf{X}} = \{\theta \mid w_1 \leq W(\mathbf{X}, \theta) \leq w_2\}$$

és un conjunt de confiança per a θ amb nivell de confiança c ; ens els casos més favorables serà un interval.

- (4.10.5) L'interval $[w_1, w_2]$ on la probabilitat de W és c no és de cap manera únic; per tant, tampoc no ho serà l'interval de confiança. Es pot triar amb el criteri de que sigui petit, o que sigui simètric.

Per exemple, suposant que W tingui funció quantil Q ben definida, podem prendre

$$w_1 = Q\left(\frac{1-c}{2}\right), \quad w_2 = Q\left(\frac{1+c}{2}\right) .$$

(O sigui, $w_1 = Q(\alpha/2)$, $w_2 = Q(1 - \alpha/2)$.)

Treballant amb les desigualtats $w_1 \leq W(\mathbf{X}, \theta) \leq w_2$ s'aïlla (si es pot!) θ , i s'intenta obtenir l'interval de confiança de la forma $L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})$.

4.11 Intervals de confiança per a alguns paràmetres destacables

Intervals de confiança per la mitjana si es coneix la variància

- (4.11.1) *Població normal*

Suposem que la població és normal, $X \sim N(\mu, \sigma)$, amb σ coneguda. Aleshores $\bar{X}_n \sim N(\mu, \sigma/\sqrt{n})$. i

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) .$$

Fixat un nivell de confiança c , l'equació

$$P(-z_o \leq Z \leq z_o) = c$$

determina

$$z_o = Q_Z\left(\frac{1+c}{2}\right) ,$$

amb Q_Z la funció quantil de la distribució normal estàndard.

La relació $|Z| \leq z_o$ equival a

$$\bar{X}_n - z_o \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_o \frac{\sigma}{\sqrt{n}} .$$

Per tant, en una proporció c de mostres, l'interval estimat

$$\left[\bar{x}_n - z_o \frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_o \frac{\sigma}{\sqrt{n}} \right]$$

conté la mitjana μ .

En el cas de $c = 0'95$, $z_o = Q(0'975) = 1'959964 \dots \approx 1'96$.

Altres valors d'interès s'obtenen d'aquesta taula:

alguns valors crítics de la normal estàndard, $p = F_Z(z)$, $z = Q_Z(p)$								
p	0.75	0.90	0.95	0.975	0.99	0.995	0.9975	0.999
z	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090

(4.11.2) Població no normal

Si la població no és normal però la mostra és prou gran ($n \geq 30$) aleshores pel teorema del límit central $\bar{X}_n$ es comporta com una $N(\mu, \sigma/\sqrt{n})$ i se segueix el mateix procediment.

Intervals de confiança per a la mitjana si no es coneix la variància

(4.11.3) Població normal

Suposem que la població és normal, $X \sim N(\mu, \sigma)$, amb μ, σ desconegudes. En aquest cas estimem la variància σ^2 amb la variància mostral corregida S_n^2 . Com sabem,

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}.$$

Aquesta distribució és coneguda, i donat un nivell de confiança c podem determinar $t_o = Q_{t_{n-1}}\left(\frac{1+c}{2}\right)$ de manera que $P(-t_o \leq T \leq t_o) = c$.

Llavors l'interval de confiança per a la mitjana μ és

$$\left[\bar{X}_n - t_o \frac{S_n}{\sqrt{n}}, \bar{X}_n + t_o \frac{S_n}{\sqrt{n}} \right].$$

Recordem que quan n és gran T tendeix a una $N(0, 1)$.

(4.11.4) Població no normal

Per a mostres petites no hi ha resultats generals.

Per a mostres grans es pot usar el mètode d'**agregació de dades**: es divideix la mostra (de mida n) en n_1 blocs de mida $b \approx 30$ ($n = bn_1$). Amb això podem considerar que tenim una mostra de mida n_1 , on cada element és la mitjana d'un bloc de mida b . Aquestes mitjanes tenen l'esperança original i, pel teorema del límit central, es comporten com variables normals, de manera que s'hi pot aplicar el procediment descrit just abans.

Intervals de confiança per a proporcions

(4.11.5) Volem saber quina proporció p d'una població posseeix una certa propietat. Que un individu la posseeixi segueix una llei de Bernoulli de paràmetre p . Volem estimar p a partir d'una mostra aleatòria.

Si considerem una mostra aleatòria $(X_1, \dots, X_n)$, el nombre d'individus que posseeix la propietat és

$$F = X_1 + \dots + X_n \sim \text{Binomial}(n, p)$$

(freqüència absoluta). Podem usar com a estimador de p la freqüència relativa, que és la mitjana mostral de les X_i :

$$\hat{p} = \frac{F}{n} = \bar{X}_n.$$

D'acord amb el teorema de Moivre–Laplace, si n , np , $n(1-p)$ són prou grans podem aproximar F per una distribució $N(np, \sqrt{np(1-p)})$. Aleshores $\hat{p}$ és aproximadament una $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$. Per tant

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

és aproximadament una $N(0, 1)$.

El valor $z_o = Q_Z\left(\frac{1+c}{2}\right)$ satisfà que $P(-z_o \leq Z \leq z_o) = c$. Encara que a partir de les desigualtats es pot donar una expressió exacta del paràmetre p , és més pràctic usar una expressió aproximada, vàlida perquè n és gran. Amb ella els límits de confiança (aproximat) per a la proporció p són

$$\hat{p} \pm z_o \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Intervals de confiança per a la variància

(4.11.6) *Població normal*

Suposem que la població és normal, $X \sim N(\mu, \sigma)$, amb μ, σ desconegudes. Estimem la variància σ^2 amb la variància mostral corregida S_n^2 . Ara tenim

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2.$$

La distribució χ^2 no és simètrica. Definim $x_1 = Q\left(\frac{1-c}{2}\right)$, $x_2 = Q\left(\frac{1+c}{2}\right)$, essent Q la funció quantil de la distribució χ_{n-1}^2 . Amb això

$$P\left(x_1 \leq \frac{n-1}{\sigma^2} S_n^2 \leq x_2\right) = c.$$

Les desigualtats equivalen a

$$\frac{\sqrt{n-1}}{\sqrt{x_2}} S_n \leq \sigma \leq \frac{\sqrt{n-1}}{\sqrt{x_1}} S_n ,$$

que és l'interval de confiança buscat.

4.12 Hipòtesis estadístiques i tests d'hipòtesis

(4.12.1) Una **hipòtesi estadística** és un enunciat relatiu als paràmetres d'una població.

(4.12.2) Un **test d'hipòtesi** és un mètode d'inferència estadística que pretén determinar la validesa d'una hipòtesi estadística a partir d'una *mostra* de la població.

(4.12.3) S'utilitza la terminologia següent.

- **Hipòtesi simple** Afirma que el valor θ del paràmetre (o paràmetres) de la població pren un valor concret, θ_0 .
- **Hipòtesi composta** Afirma que el valor θ del paràmetre pertany a un cert subconjunt Θ_0 . Hi ha dos subcasos d'interès especial (quan $\Theta_0 \subset \mathbf{R}$):
 - **Hipòtesi unilateral**
 - * **Cua dreta** $\theta > \theta_0$.
 - * **Cua esquerra** $\theta < \theta_0$.
 - **Hipòtesi bilateral** $\theta \neq \theta_0$.

(4.12.4) De manera més concreta, en un test d'hipòtesi es vol confrontar dues hipòtesis:

- **Hipòtesi nul·la:** $H_0: \theta \in \Theta_0$ (tot i que acostuma a ser una hipòtesi simple);
- **Hipòtesi alternativa:** $H_1: \theta \notin \Theta_0$.

El paper d'aquestes hipòtesis no és simètric. Rebutjar la hipòtesi nul·la i doncs acceptar l'alternativa requereix un cost i ha d'estar justificat amb una evidència suficient.

(4.12.5) El test d'hipòtesi divideix l'espai mostral en dues regions complementàries, $\Omega = \Omega_0 \cup \Omega_1$.

- **Regió d'acceptació:** si $\mathbf{x} \in \Omega_0$, no rebutgem la hipòtesi nul·la H_0 .
- **Regió de rebuig**, o **regió crítica:** si $\mathbf{x} \in \Omega_1$, rebutgem la hipòtesi nul·la H_0 i doncs ens acollim a la hipòtesi alternativa H_1 .

4.13 Significació i potència d'un test d'hipòtesi

(4.13.1) La bondat d'un test d'hipòtesi depèn de seleccionar correctament la regió d'acceptació (i doncs la de rebuig). A tal fi pot ser útil examinar la probabilitat de prendre una decisió incorrecta rebutjant, o no rebutjant, la hipòtesi nul·la.

(4.13.2) Hi ha dos tipus d'error a considerar:

- **Error de primera espècie**, o de tipus I: rebutjar H_0 tot i ser certa.
- **Error de segona espècie**, o de tipus II: no rebutjar H_0 tot i ser falsa.

	no rebutjar H_0	rebutjar H_0
H_0 certa	correcte!	error tipus I
H_0 falsa	error tipus II	correcte!

(4.13.3) Suposem per simplicitat que H_0 és una hipòtesi simple $\theta = \theta_0$, de manera que H_1 és la hipòtesi $\theta \neq \theta_0$, o $\theta \in \Theta_1 = \Theta - \{\theta_0\}$. Siguin Ω_0 i Ω_1 les regions d'acceptació i de rebuig.

L'error de primera espècie té una probabilitat

$$P_{\theta_0}(\Omega_1),$$

mentre que l'error de segona espècie té una probabilitat

$$P_{\theta}(\Omega_0) \quad \text{per a } \theta \in \Theta_1.$$

Aquests errors es poden expressar en termes de la **funció de potència** del test: $\Phi: \Theta \rightarrow [0, 1]$ definida per

$$\Phi(\theta) = P_{\theta}(\Omega_1).$$

Llavors els errors de primera i segona espècie són $\Phi(\theta_0)$ i $1 - \Phi(\theta)$ (per a $\theta \in \Theta_1$), respectivament.

(4.13.4) Idealment voldríem minimitzar els dos errors però habitualment disminuir-ne un implica augmentar l'altre.

Es planteja el problema de manera que l'error de primera espècie sigui el de pitjors conseqüències (principi de Neyman).

El **nivell de significació** d'un test d'hipòtesi, sovint denotat per α , és el valor màxim que admetrem per a l'error de primera espècie.

Llavors es fixa un nivell de significació α , i entre tots els tests amb aquest

nivell de significació se’n busca un que minimitzi l’error de segona espècie (és a dir, que maximitzi la potència).

- (4.13.5) Alguns valors típics que s’adopten per al nivell de significació són $\alpha = 0'1, 0'05, 0'01$.
- (4.13.6) Habitualment no es treballa amb un estimador $\hat{\theta}$ de θ , sinó amb una variable Z amb distribució coneguda.
- (4.13.7) Podem expressar la regió d’acceptació en termes dels valors de $\hat{\theta}$, o potser millor de Z . Llavors la forma d’aquesta regió depèn de la hipòtesi alternativa considerada (bilateral o unilateral) i els valors z que separen la regió d’acceptació de la regió de rebuig s’anomenen *valors crítics*.
Suposant que H_0 és $\theta = \theta_0$, tenim aquesta taula:

hipòtesi alternativa H_1	regió d’acceptació A_0
bilateral: $\theta \neq \theta_0$	$[z_{\alpha_1}, z_{\alpha_2}]$, on $P(Z < z_{\alpha_1}) = \frac{\alpha}{2} = P(Z > z_{\alpha_2})$
cua dreta: $\theta > \theta_0$	$] -\infty, z_\alpha]$, on $P(Z \geq z_\alpha) = \alpha$
cua esquerra: $\theta < \theta_0$	$[z_\alpha, +\infty[$, on $P(Z \leq z_\alpha) = \alpha$

4.14 Valor p d’un test d’hipòtesi

- (4.14.1) El resultat d’un test d’hipòtesi depèn del nivell de significació α escollit; no ens diu si el valor observat de l’estadístic és proper a la frontera que separa la regió d’acceptació i la regió de rebuig. Un procediment alternatiu consisteix a no prefixar α , sinó mirar com el resultat del test depèn d’ella.
- (4.14.2) Si t_{obs} és el valor de l’estadístic de test T obtingut en realitzar una prova, i es pren aquest valor com a punt de separació entre la regió d’acceptació i la regió de rebuig, llavors el valor p és el nivell de significació corresponent.
- (4.14.3) En altres paraules, en un test d’hipòtesi, el **valor p** (o p-valor, o valor de probabilitat) és la probabilitat d’obtenir, suposant certa la hipòtesi nul·la, un resultat de l’estadístic de test *tan o més extrem que* t_{obs} . Per tant el valor p és, segons el cas:
 - Cua dreta: $P_{\theta_0}(T \geq t_{\text{obs}})$;
 - Cua esquerra: $P_{\theta_0}(T \leq t_{\text{obs}})$;

- Test bilateral: $2 \min \left(P_{\theta_0}(T \leq t_{\text{obs}}), P_{\theta_0}(T \geq t_{\text{obs}}) \right)$.

En el cas que T tingui una distribució contínua i simètrica aquests nombres són, respectivament,

- $1 - F(t_{\text{obs}})$,
- $F(t_{\text{obs}})$,
- $2 F(-|t_{\text{obs}}|)$,

essent F la funció de distribució de T quan $\theta = \theta_0$.

(4.14.4) Amb el valor p calculat, i fixat un nivell de significació α , el test d’hipòtesi pren aquesta forma:

- No rebutgem H_0 si el valor p és $> \alpha$;
- Rebutgem H_0 si el valor p és $< \alpha$.

(4.14.5) Una interpretació habitual del resultat del valor p és la següent:

- $0 \leq p \leq 0'01$: rebutgem H_0 .
- $0'01 < p < 0'10$: zona grisa, cal ser prudent. És preferible repetir el test, amb mostres més grans. Si no hi ha altre remei que prendre una decisió, podríem adoptar $p = 0'05$ com a frontera entre rebuig i no rebuig.
- $0'10 \leq p \leq 1$: no rebutgem H_0 .

De tota manera, l’ús del valor p és controvertit, i no hauria de ser l’únic argument usat en un test d’hipòtesi.

(4.14.6) *Comparació entre test d’hipòtesi clàssic i valor p*

En aquestes dues taules resumim i comparem els procediments explicats:

test clàssic	valor p
formular hipòtesis nul·la H_0 i alternativa H_1	
decidir l’estadístic de test T	
seleccionar nivell de significació α	
establir regions d’acceptació i de rebuig per als valors de T	
fer una observació t_{obs} de T	
rebutjar H_0 si t_{obs} pertany a la regió de rebuig, altrament no rebutjar-la	calcular el valor p corresponent a t_{obs} i decidir a partir del resultat

test de $H_0: \theta = \theta_0$; probabilitats calculades sota H_0 ; estadístic de test T		
hip. alternativa H_1	regió d'acceptació A_0	valor p
bilateral: $\theta \neq \theta_0$	$[t_1, t_2]$, on $P(T < t_1) = \frac{\alpha}{2} = P(T > t_2)$	$2 \min\left(P(T \leq t_{\text{obs}}), P(T \geq t_{\text{obs}})\right)$
cua dreta: $\theta > \theta_0$	$] -\infty, t_\alpha]$, on $P(T > t_\alpha) = \alpha$	$P(T \geq t_{\text{obs}})$
cua esquerra: $\theta < \theta_0$	$[t_\alpha, +\infty[$, on $P(T < t_\alpha) = \alpha$	$P(T \leq t_{\text{obs}})$

4.15 Tests d'hipòtesi sobre paràmetres interessants

(4.15.1) En el cas de poblacions normals podem contrastar els valors dels paràmetres mitjançant els estadístics de test següents:

hipòtesi	estadístic	nom
$\mu = \mu_0$ (i variància coneguda)	$Z = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$	test Z
$\mu = \mu_0$ (i variància desconeguda)	$T = \frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} \sim t_{n-1}$	test T
$\sigma^2 = \sigma_0^2$	$\chi^2 = \frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi_{n-1}^2$	test χ^2

A Complements

A.1 Elements de combinatòria

En moltes situacions el càlcul de probabilitats requereix comptar el nombre d'elements d'algun conjunt finit. La **combinatòria** és la branca de la matemàtica discreta que estudia l'enumeració, construcció i existència de *configuracions* que satisfan certes propietats prescrites.

Aritmètica bàsica

(A.1.1) Si A és un conjunt finit, el nombre d'elements de A és un nombre natural anomenat **cardinal**, que es representa per $\text{Card}(A) \equiv |A|$.

La propietat bàsica dels cardinals és que $|A| = |B|$ sii existeix una bijecció entre els conjunts A i B .

(A.1.2) Les operacions amb cardinals es defineixen, de fet, a partir d'operacions amb conjunts:

- $|A \cup B| = |A| + |B|$ si A i B són conjunts *disjunts*.
- $|A \times B| = |A| |B|$.
- $|\mathcal{F}(A, B)| = |B|^{|A|}$, essent $\mathcal{F}(A, B)$ el conjunt d'aplicacions de A en B .

(A.1.3) Si $S \subset A$, $|A - S| = |A| - |S|$.

(A.1.4) Donada una part $S \subset A$, es defineix la seva **funció indicadora** (o *funció característica*) com la funció $\mathbb{1}_S: A \rightarrow \{0,1\}$ definida per

$$\mathbb{1}_S(x) = \begin{cases} 1 & \text{si } x \in S \\ 0 & \text{si } x \notin S \end{cases}$$

Recíprocament, qualsevol funció $f: A \rightarrow \{0,1\}$ defineix un subconjunt $f^{-1}(1) = \{x \in A \mid f(x) = 1\} \subset A$.

Tenim doncs una bijecció

$$\mathcal{P}(A) \longrightarrow \mathcal{F}(A, \{0,1\}), \quad S \mapsto \mathbb{1}_S,$$

entre el conjunt de parts de A i el conjunt de funcions definides en A amb valors en $\{0,1\}$. Per tant,

- $|\mathcal{P}(A)| = 2^{|A|}$.

(A.1.5) *Principi d'inclusió-exclusió*

$$|A_1 \cup \dots \cup A_n| =$$

$$\sum_i |A_i| - \sum_{i < j} |A_i \cap A_j| + \sum_{i < j < k} |A_i \cap A_j \cap A_k| - \dots + (-1)^{n+1} |A_1 \cap \dots \cap A_n|.$$

(A.1.6) Casos particulars:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Seleccions ordenades

(A.1.7) *Variacions*

Les **variacions de m elements presos de n en n** són les aplicacions injectives $\{1, \dots, n\} \rightarrow A$, essent $|A| = m$.

També podem interpretar-les com les cadenes de longitud n construïdes amb un alfabet A de m lletres sense repetir-ne cap.

$$V_m^n = \frac{m!}{(m-n)!}.$$

(A.1.8) *Permutacions*

Les **permutacions de m elements** són les aplicacions bijectives $\{1, \dots, m\} \rightarrow A$, essent $|A| = m$.

Éa a dir, són totes les possibles ordenacions dels m elements de A .

$$P_m = V_m^m = m!.$$

(A.1.9) *Variacions amb repetició*

Les **variacions amb repetició de m elements presos de n en n** són les aplicacions $\{1, \dots, n\} \rightarrow A$, essent $|A| = m$.

També podem interpretar-les com les cadenes de longitud n construïdes amb un alfabet A de m símbols, amb possibles repeticions.

$$VR_m^n = m^n.$$

(A.1.10) *Permutacions amb repetició*

Les **permutacions amb repetició** del conjunt $\{a_1, a_2, \dots, a_m\}$ són ordenacions dels elements de A en successions de n elements de manera que a_1 aparegui n_1 vegades, a_2 aparegui n_2 vegades, etc; òbviament, $n_1 + n_2 + \dots + n_m = n$.

$$PR_n^{n_1 \dots n_m} = \frac{n!}{n_1! n_2! \dots n_m!}.$$

Seleccions no ordenades

(A.1.11) *Combinacions*

Les **combinacions de m elements presos de n en n** són els subconjunts de n elements que hi ha en un conjunt A de m elements.

$$C_m^n \equiv \binom{m}{n} = \frac{m!}{n!(m-n)!}.$$

També es corresponen amb les aplicacions $A \rightarrow \{0, 1\}$.

(A.1.12) Suposem que el conjunt A està ordenat o, més simplement, que $A = \{1, 2, 3, \dots, m\}$. Aleshores una aplicació $A \rightarrow \{0, 1\}$ és una seqüència de longitud m , i C_m^n és el nombre de les seqüències amb n 1's i $m-n$ 0's.

(A.1.13) Encara suposant que $A = \{1, 2, 3, \dots, m\}$, podem especificar un subconjunt $S \subset A$ citant els seus n elements de manera creixent. Així veiem que les combinacions de m elements presos de n en n també es corresponen amb les aplicacions estrictament creixents $\{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, m\}$.

(A.1.14) *Propietats dels coeficients binomials*

$$\binom{m}{0} = \binom{m}{m} = 1; \quad \binom{m}{n} = 0 \text{ si } n > m.$$

$$\binom{m}{n} = \binom{m}{m-n}.$$

$$\binom{m+1}{n} = \binom{m}{n} + \binom{m}{n-1}.$$

$$\sum_{n=0}^m \binom{m}{n} = 2^m.$$

$$(A.1.15) \quad PR_m^{n_1 \dots n_m} = \binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \dots \binom{n-n_1-\dots-n_{m-1}}{n_m}.$$

(A.1.16) *Combinacions amb repetició*

Les **combinacions amb repetició de m elements presos de n en n** són seleccions (sense ordre) de n elements, amb possibles repeticions, d'un conjunt A de m elements.

$$CR_m^n = C_{m+n-1}^n = \binom{m+n-1}{n}.$$

(A.1.17) També es poden interpretar com les maneres de distribuir n boles idèntiques en m caixes diferents.

Si $A = \{1, 2, 3, \dots, m\}$, una tal distribució es pot codificar amb una seqüència de n 0's i $m-1$ 1's, on els 0 codifiquen les boles i els 1 els canvis

de caixa.

- (A.1.18) Alternativament, i suposant encara que $A = \{1, 2, 3, \dots, m\}$, les combinacions amb repetició es corresponen amb les aplicacions creixents (no cal que ho siguin estrictament) $\{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, m\}$, ja que podem citar els elements de la combinació de manera creixent (i amb repeticions).

A.2 Funcions especials

- (A.2.1) La **funció gamma** és una funció que en els nombres naturals coincideix amb el factorial (desplaçat). En principi es pot definir per als nombres reals $x > 0$ com la integral

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

- (A.2.2) La funció gamma compleix la relació de recurrència

$$\Gamma(x+1) = x \Gamma(x).$$

Quan $n \in \mathbf{N}$ tenim

$$\Gamma(n+1) = n!.$$

Fora dels enters un valor important és

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi},$$

a partir del qual es poden obtenir els valors de Γ en els nombres semienters.

- (A.2.3) Quan x tendeix a infinit la funció gamma creix més ràpidament que qualsevol exponencial. La **fórmula de Stirling** afirma que

$$\Gamma(x+1) \sim \sqrt{2\pi x} \left(\frac{x}{e}\right)^x \quad \text{quan } x \rightarrow +\infty.$$

(infinites equivalents).

En el cas de x natural tenim

$$n! \sim \sqrt{2\pi n} e^{-n} n^n,$$

expressió d'interès quan volem aproximar coeficients binomials, per exemple.

- (A.2.4) La **funció beta** és la funció de dues variables reals estrictament positives (en principi) definida per

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt.$$

Està estretament relacionada amb la funció gamma ja que

$$B(x, y) = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x + y)}.$$

(A.2.5) La **funció d'error** és la funció de variable real definida per

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

El factor del davant s'ha escollit de manera que $\lim_{x \rightarrow +\infty} \operatorname{erf}(x) = 1$. Semblantment $\lim_{x \rightarrow -\infty} \operatorname{erf}(x) = -1$, ja que erf és una funció imparella.

La funció d'error és infinitament diferenciable, ja que la seva derivada és

$$\operatorname{erf}'(x) = \frac{2}{\sqrt{\pi}} e^{-x^2}.$$

Índex terminològic

- σ -àlgebra (1.2.5)
- σ -àlgebra de Borel (2.1.2)
- σ -àlgebra de Borel de $\mathbf{R}^n$ (3.1.1)
- σ -àlgebra discreta (1.4.1)
- σ -àlgebra generada (1.2.8)
- agregació de dades (4.11.4)
- àlgebra de conjunts (1.2.1)
- àlgebra d'esdeveniments (1.2.1)
- biaix (4.6.2)
- coeficient d'asimetria (4.1.1)
- coeficient de correlació (3.9.4)
- coeficient de curtosi (4.1.2)
- conjunt borelià (2.1.2)
- covariància (3.9.1)
- desigualtat de Cramér-Rao (4.9.6)
- desigualtat de Markov (2.15.4)
- desigualtat de Txebyxev (2.15.6)
- desviació estàndard (2.13.3)
- desviació mostral (4.5.8)
- distribució gamma (4.3.1)
- distribució khi quadrat (4.3.6)
- distribució normal n -dimensional (3.12.1)
- distribució normal bidimensional (3.2.7)
- distribució normal estàndard (2.7.8)
- distribució t de Student (4.3.9)
- distribució trinomial (3.3.4)
- distribució uniforme bidimensional (3.2.6)
- error de primera espècie (4.13.2)
- error de segona espècie (4.13.2)
- esdeveniment aleatori (1.1.3)
- esdeveniment elemental (1.4.2)
- esdeveniment gairebé segur (2.15.1)
- esdeveniments equiprobables (1.4.5)
- esdeveniments independents (1.5.1)
- esdeveniments independents (1.5.3)
- espai de probabilitat (1.3.2)
- espai de probabilitat discret (1.4.3)
- espai de probabilitat producte (1.6.3)
- espai mesurable (1.3.1)
- espai mesurable producte (1.6.2)
- espai mostral (1.1.2)
- espai probabilitzable (1.3.1)
- esperança (2.12.2)
- esperança condicionada (2.12.8)
- esperança condicionada (3.5.5)
- estimador (3.13.1)
- estimador (4.6.1)
- estimador consistent (4.6.5)
- estimador de màxima versemblança (4.8.2)
- estimador eficient (4.9.8)
- estimador més eficient (4.6.8)
- estimador no esbiaixat (4.6.2)
- excés de curtosi (4.1.2)
- experiment aleatori (1.1.1)
- fitxa de Cramér-Rao (4.9.7)
- fórmula de Bayes (1.8.4)
- fórmula de la probabilitat total (1.8.1)
- fórmula de Stirling (A.2.3)
- fórmula d'inclusió-exclusió (1.3.5)
- funció beta (A.2.4)
- funció característica (4.2.2)
- funció de densitat (2.6.1)
- funció de densitat condicionada (2.10.2)
- funció de densitat condicionada (3.5.2)
- funció de densitat conjunta (3.11.4)
- funció de densitat conjunta (3.2.1)
- funció de densitat marginal (3.2.4)
- funció de distribució (2.3.2)
- funció de distribució condicionada (2.10.1)
- funció de distribució condicionada (3.5.1)
- funció de distribució conjunta (3.11.2)

- funció de distribució conjunta (3.1.7)
- funció de massa de probabilitat (1.4.3)
- funció de massa de probabilitat conjunta (3.3.1)
- funció de potència d'un test (4.13.3)
- funció d'error (A.2.5)
- funció gamma (A.2.1)
- funció mesurable Borel (2.11.1)
- funcions de densitat marginals (3.11.6)
- funcions de distribució marginals (3.1.10)
- funcions de distribució marginals (3.11.3)
- funcions de massa de probabilitat marginals (3.3.3)
- funció quantil (4.1.6)
- hipòtesi alternativa (4.12.4)
- hipòtesi bilateral (4.12.3)
- hipòtesi composta (4.12.3)
- hipòtesi de cua dreta (4.12.3)
- hipòtesi de cua esquerra (4.12.3)
- hipòtesi estadística (4.12.1)
- hipòtesi nul·la (4.12.4)
- hipòtesi simple (4.12.3)
- hipòtesi unilateral (4.12.3)
- informació de Fisher (4.9.4)
- interval de confiança (4.10.1)
- lleï binomial (2.5.4)
- lleï binomial negativa (2.5.8)
- lleï de Bernoulli (2.5.1)
- lleï de Cauchy (2.7.9)
- lleï de Poisson (2.5.6)
- lleï de probabilitat d'una variable aleatòria (2.2.8)
- lleï exponencial (2.7.2)
- lleï feble dels grans nombres (2.16.2)
- lleï gaussiana (2.7.6)
- lleï geomètrica (2.5.5)
- lleï hipergeomètrica (2.5.9)
- lleï uniforme (2.7.1)
- log-versemblança (4.8.4)
- matriu de covariàncies (3.11.10)
- mediana (4.1.4)
- mediana mostral (4.5.11)
- mesura de probabilitat (1.3.2)
- mesura de probabilitat producte (1.6.3)
- mitjana mostral (3.14.6)
- mitjana mostral (4.5.2)
- moda (4.1.3)
- moment n -èsim (2.13.7)
- moment central n -èsim (2.13.8)
- moment mostral (4.5.10)
- moments centrals conjunts (3.9.7)
- moments conjunts (3.9.6)
- mostra aleatòria simple (4.4.3)
- nivell de confiança (4.10.1)
- nivell de significació (4.13.4)
- outlier (4.1.7)
- població (4.4.1)
- probabilitat condicionada (1.7.1)
- probabilitat elemental (1.4.3)
- quantil (4.1.5)
- recorregut interquartílic (4.1.7)
- regió d'acceptació (4.12.5)
- regió de rebuig (4.12.5)
- teorema de convolució (3.6.4)
- teorema de l'esperança (2.12.7)
- teorema de Moivre–Laplace (2.9.1)
- test d'hipòtesi (4.12.2)
- valor atípic (4.1.7)
- valor p (4.14.3)
- variable aleatòria (2.2.1)
- variable aleatòria contínua (2.6.1)
- variable aleatòria contínua multidimensional (3.11.4)
- variable aleatòria contínua multidimensional (3.2.1)
- variable aleatòria discreta (2.4.1)
- variable aleatòria mixta (2.8.2)
- variable aleatòria mixta (2.8.2)
- variable aleatòria multidimensional (3.1.2)

variable aleatòria normalitzada
(2.13.4)
variable aleatòria positiva (2.12.4)
variable aleatòria positiva gairebé
segurament (2.15.3)
variable aleatòria sense memòria
(2.7.4)
variable indicadora (2.2.4)
variables aleatòries independents

(3.4.1)
variables aleatòries ortogonals
(3.9.10)
variables incorrelades (3.9.9)
variància (2.13.1)
variància mostral corregida (4.5.7)
variància mostral esbiaixada (4.5.5)
vector d'esperances (3.11.10)