Apèndixs

A. Quadre amb algunes distribucions

Distribució	Paràmetres	Funció de densitat	Mitjana	Variança	Funció característica
Bernoulli	$0 \le p \le 1$, $q = 1-p$	$p^k\,(1-p)^{1-k}$, $k = 0{,}1$	p	$p\,(1-p)$	$q + p\,\mathrm{e}^{i\,t}$
Binomial	$0 \le p \le 1$, $q = 1-p$	$\binom{n}{k} p^k\,(1-p)^{n-k}$, $k = 0{,}1,\ldots n$	$n\,p$	$n\,p\,(1-p)$	$(q + p\,\mathrm{e}^{i\,t})^n$
Geomètrica	$0 \le p \le 1$, $q = 1-p$	$p\,(1-p)^k$, $k \ge 0$	$\dfrac{1-p}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{p}{1-q\,\mathrm{e}^{i\,t}}$
Binomial negativa	$r > 0$, $0 \le p \le 1$, $q = 1-p$	$\binom{k+r-1}{k} p^r\,q^k$, $k \ge 0$	$r\,\dfrac{1-p}{p}$	$r\,\dfrac{1-p}{p^2}$	$\left(\dfrac{p}{1-q\,\mathrm{e}^{i\,t}}\right)^r$
Poisson	$\lambda > 0$	$\dfrac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}$, $k \ge 0$	λ	λ	$\exp\left\{\lambda\,(\mathrm{e}^{i\,t}-1)\right\}$
Normal $N(\mu,\sigma)$	$\mu \in \mathbb{R}$, $\sigma > 0$	$\dfrac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left\{\dfrac{-(x-\mu)^2}{2\,\sigma^2}\right\}$, $x \in \mathbb{R}$	μ	σ^2	$\exp\left\{\mu\,i\,t - \dfrac{t^2\,\sigma^2}{2}\right\}$
Uniforme	$a < b$	$\dfrac{1}{b-a}$, $a < x < b$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{\mathrm{e}^{i\,t\,b} - \mathrm{e}^{i\,t\,a}}{i\,t\,(b-a)}$
Exponencial	α	$\alpha\,\mathrm{e}^{-\alpha\,x}$, $x > 0$	$\dfrac{1}{\alpha}$	$\dfrac{1}{\alpha^2}$	$\left(1 - \dfrac{i\,t}{\alpha}\right)^{-1}$
Gamma $\gamma(n,\alpha)$	$\alpha > 0$, $n > 0$	$\dfrac{\alpha^n\,x^{n-1}\,\mathrm{e}^{-\alpha\,x}}{\Gamma(n)}$, $x > 0$	$\dfrac{n}{\alpha}$	$\dfrac{n}{\alpha^2}$	$\left(1 - \dfrac{i\,t}{\alpha}\right)^{-n}$
Beta $\beta(p,q)$	$p > 0$, $q > 0$	$\dfrac{x^{p-1}\,(1-x)^{q-1}}{\mathrm{B}(p{,}q)}$, $0 < x < 1$	$\dfrac{p}{p+q}$	$\dfrac{p\,q}{(p+q)^2(p+q+1)}$	
Cauchy	$a > 0$, $b \in \mathbb{R}$	$\dfrac{a}{\pi\,[a^2 + (x-b)^2]}$, $x \in \mathbb{R}$	No existeix	No existeix	$\mathrm{e}^{i\,t\,b - a\,\lvert t\rvert}$

$$\Gamma(x) = \int_0^\infty e^{-t}\,t^{x-1}\,dt,\ \Gamma(n) = (n-1)! \quad \mathrm{B}(p{,}q) = \int_0^1 t^{p-1}\,(1-t)^{q-1}\,dt,\ \mathrm{B}(p{,}q) = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}$$

Apunts de probabilitat de variables aleatòries contínues

Llorenç Cerdà-Alabern
llorenc@ac.upc.edu
Barcelona, setembre de 2012.

Índex

Prefaci

Vaig escriure aquests apunts mentre preparava l'assignatura *cálculo de probabilidades 2* del curs 2011-12 de la carrera de matemàtiques de la UNED. Els apunts estan basats fonamentalment en l'excel·lent llibre de l'assignatura [1]. El contingut no és rigorós ni complert, i només hi ha demostracions senzilles orientades a entendre o ajudar a memoritzar les relacions que demostren.

Capítol 1

Model de probabilitat

1.1 Àlgebra dels successos

Esdeveniment (*outcome*) ω és un dels possibles resultats d'un experiment aleatori. El conjunt de tots els esdeveniment és l'espai mostral (*sample space*) Ω. Per exemple, en el llançament d'un dau $\Omega = \{1.2.3.4.5.6\}$ i els possibles esdeveniments són $\omega = 1{,}2{,}3{,}4{,}5{,}6$. Ω pot tenir un nombre ∞ d'esdeveniments, per exemple, coordenades que resulten de llançar un dard en una diana.

Succés (*event*) A és qualsevol subconjunt d'esdeveniments de l'espai mostral Ω . Per exemple, en el llançament d'un dau podria ser $A = \{\text{és parell}\} = \{2{,}4{,}6\}$, o en el llançament d'un dard podria ser $A = \{\text{el dard cau en el cercle de puntuació 20}\}$.

Lògica dels successos

- Un esdeveniment ω realitza el succés A si $\omega \in A$. $\emptyset$ és el succés impossible, doncs no el pot realitzar cap esdeveniment Ω és el succés segur, doncs qualsevol esdeveniment $\omega \in \Omega$.
- El complement d'un succés, A^c, és un altre succés format per tots els esdeveniments que no realitzen A: $A^c = \{\omega \in \Omega \mid \omega \notin A\}$.
- Un succés B implica un succés A, $B \subset A$, quan tots els esdeveniments que realitzen B també realitzen A: $\omega \in B \Rightarrow \omega \in A$.
- Diferència de successos: quan $B \subset A$, es fa servir la notació $A \setminus B$ per referir-se a tots els esdeveniments

que realitzen A però no B, és a dir: $A \setminus B = A \cap B^c$. Clarament $A^c = \Omega \setminus A$.

- Un esdeveniment ω realitza el succés $A \cup B$ si ω està en un o els dos subconjunts A i B.
- Un esdeveniment ω realitza el succés $A \cap B$ si ω està en ambdós subconjunts A i B.
- Dos successos són disjunts si $A \cap B = \emptyset$.
- La seqüència de successos $\{A_n\}_{n\geq 1}$ és una partició d'Ω si tots els successos $A_1, A_2, \cdots$ són mútuament disjunts i, a més, la seva unió és exhaustiva: $\cup_{n=1}^{\infty} A_n = \Omega$.
- Lleis de Morgan (*De Morgan's laws*): per a qualsevols successos $A,B \subseteq \Omega$:

$$(A \cap B)^c = A^c \cup B^c$$
$$(A \cup B)^c = A^c \cap B^c$$

La teoria de la probabilitat estudia l'assignació d'un número real (una probabilitat) a cada succés d'Ω. Quan Ω està format per un nombre finit d'esdeveniments (per exemple el llançament d'un dau), qualsevol subconjunt d'Ω es pot considerar un succés.

En el cas del llançament d'un dard el model probabilístic és més complex, doncs els conjunts involucren ∞ esdeveniments (tots els punts de l'àrea inclosa en el cercle de puntuació 20). Per caracteritzar rigorosament aquests models **continus** es va desenvolupar la **teoria de la mesura**. L'ingredient fonamental en aquesta teoria és el concepte de **longitud** per a conjunts en $\mathbb{R}$ (que denotarem per λ), i que es generalitza als conjunts en $\mathbb{R}^n$ (λ_n).

Longitud La longitud, λ, és una mesura que podem assignar als conjunts de $\mathbb{R}$ que té les següents propietats:

1. En un segment A d'extrems a, b (independentment si és obert, tancat o semi-obert) val $\lambda(A) = |b - a|$.
2. Si A_1 i A_2 són disjunts: $\lambda(A_1 \cup A_2) = \lambda(A_1) + \lambda(A_2)$.
3. Si $\{A_n\}_{n\geq 1}$ és una successió creixent i $A = \cup_{n=1}^{\infty} A_n$, aleshores $\lambda(A) = \lim_{n\to\infty} \lambda(A_n)$.
4. Si $A_2 \subset A_1$, i $\lambda(A_2) \neq \infty$, aleshores $\lambda(A_1 - A_2) = \lambda(A_1) - \lambda(A_2)$.

En els models probabilístics continus el conjunts de successos als que s'assigna una probabilitat no sempre és la col·lecció de tots els subconjunts possibles d'Ω. Per a poder fer un model consistent cal introduir certes restriccions als subconjunts d'Ω que s'associaran als possibles successos, i als que s'assignarà una probabilitat. En concret, els subconjunts han de formar una σ-àlgebra, que es defineix a continuació. Per això, també s'anomena **σ-àlgebra de successos** (*σ-algebra of events*).

σ-àlgebra Una família $\mathcal{F}$ de subconjunts d'Ω és una σ-àlgebra de subconjunts d'Ω si satisfà els axiomes:
- Per a qualsevol conjunt $A \subset \mathcal{F}$, el conjunt $A^c \subset \mathcal{F}$.
- La unió de qualsevol col·lecció numerable de conjunts $A_1, A_2, \cdots$ de $\mathcal{F}$ està en $\mathcal{F}$: $\cup_{n=1}^{\infty} A_n \subseteq \mathcal{F}$.

σ-àlgebra engendrada per una família de subconjunts $\mathcal{C}$ és la intersecció de totes les σ-àlgebra que contenen a $\mathcal{C}$:

$$\sigma(\mathcal{C}) = \cap_{\mathcal{F}\supset\mathcal{C}} \mathcal{F}$$

Sigui $\mathcal{I}$ la família formada per tots els intervals de $\mathbb{R}$. La σ-àlgebra engendrada per $\mathcal{I}$, $\sigma(\mathcal{I})$, s'anomena **σ-àlgebra de Borel en** $\mathbb{R}$ i sol denotar-se per $\mathbb{B}$. Anàlogament, $\mathcal{I}^2$ denota la família de tots els rectangles en $\mathbb{R}^2$, i $\mathbb{B}^2$ la σ-àlgebra engendrada.

σ-àlgebra producte Siguin $\mathcal{F}_1, \mathcal{F}_2$ σ-àlgebra dels conjunts Ω_1, Ω_2 respectivament. La família de subconjunts de $\Omega_1 \times \Omega_2$:

$$\mathcal{F}_1 \times \mathcal{F}_2 = \{A \times B \mid A \in \mathcal{F}_1, B \in \mathcal{F}_2\} \tag{1.1}$$

engendra una σ-àlgebra anomenada σ-àlgebra producte de $\mathcal{F}_1, \mathcal{F}_2$ i sol indicar-se per: $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$. Notar que $\mathbb{B}^2 = \mathbb{B} \otimes \mathbb{B}$.

Proposició 1.1 Donats dos conjunts $A,B \in \mathbb{B}$, el conjunt $A \times B \in \mathbb{B}^2$ verifica:

$$\lambda(A \times B) = \lambda(A)\, \lambda(B).$$

En $\mathbb{R}^k$ es generalitza el volum d'un paral·lelepípede $A^n \in \mathbb{B}^n$ com el producte de les seves arestes: $\lambda_n(A^n) = \prod_{i=1}^{n} |b_i - a_i|$. A més, per $A^r \in \mathbb{B}^r$, $B^s \in \mathbb{B}^s$, $\lambda_{r+s}(A^r \times B^s) = \lambda_r(A^r)\, \lambda_s(B^s)$.

1.2 Espais de probabilitat

Espai mesurable és la tupla $(\Omega, \mathcal{F})$ on Ω és un conjunt qualsevol i $\mathcal{F}$ és una σ-àlgebra de subconjunts d'Ω.

Propietats

1. $X_n \xrightarrow{\text{q.s.}} X \Rightarrow X_n \xrightarrow{\text{P}} X$.
2. $X_n \xrightarrow{\text{P}} X \Rightarrow X_n \xrightarrow{\text{d}} X$.
3. $X_n \xrightarrow{\mathcal{L}^p} X \Rightarrow X_n \xrightarrow{\mathcal{L}^q} X$ per $q < p$, i $X_n \xrightarrow{\text{P}} X$.
4. Una successió de distribucions F_n **convergeix dèbilment** cap a una distribució F (i s'indica $F_n \xrightarrow{\text{d}} F$) si $F_n(x) \to F(x)$, $\forall x \in \mathbb{R}$ on $F(x)$ és contínua. $F_n \xrightarrow{\text{d}} F \Rightarrow X_n \xrightarrow{\text{d}} X$.
5. Si $X_n \xrightarrow{\text{q.s.}} X$, $Y_n \xrightarrow{\text{q.s.}} Y$ i $g : \mathbb{R}^2 \to \mathbb{R}$ és contínua: $g(X_n,Y_n) \xrightarrow{\text{q.s.}} g(X,Y)$. Igualment per la convergència en probabilitat.
6. Si $X_n \xrightarrow{\text{d}} X$ i f és contínua, aleshores $f(X_n) \xrightarrow{\text{d}} f(X)$.
7. $F_n \xrightarrow{\text{d}} F \Rightarrow \psi_n \to \psi$.

9.1 Teorema central del límit

Sigui $\{X_j\}$ una successió de VA iid amb variança finita σ^2 i mitjana μ. Definim la suma parcial: $S_n = X_1 + \cdots X_n$, $\mathrm{E}[S_n] = n\,\mu$, $\sigma^2(S_n) = n\,\sigma^2$. Aleshores, la variable tipificada

$$Z_n = \frac{S_n - n\,\mu}{\sigma\sqrt{n}} \xrightarrow{\text{d}} Z \sim N(0,1)$$

De forma que per n gran es pot aproximar S_n per $N(n\,\mu, \sigma\sqrt{n})$. Quan S_n és discreta és millor l'aproximació:

$$P(S_n = a) \approx \phi\Big(\frac{a + 0.5 - n\,\mu}{\sigma\sqrt{n}}\Big) - \phi\Big(\frac{a - 0.5 - n\,\mu}{\sigma\sqrt{n}}\Big)$$

on $\phi(x)$ és la funció de distribució de $N(0,1)$.

Cas general Sigui $\{X_j\}$ una successió de VA independents amb $\mu_j = \mathrm{E}[X_j]$ i $\sigma_j^2 = \mathrm{E}[(X_j - \mu_j)^2]$. $S_n = X_1 + \cdots X_n$, $\mathrm{E}[S_n] = \mu_1 + \cdots \mu_n$, $\sigma_n^2 = \sigma^2(S_n) = \sigma_1^2 + \cdots \sigma_n^2$.

Proposició 9.1 (Lyapunov) Si $\mathrm{E}[|X_j - \mu_j|^3]$ són finits:

$$\frac{1}{s_n^3} \sum_{j=1}^{n} \mathrm{E}[|X_j - \mu_j|^3] \to 0 \Rightarrow \frac{S_n - m_n}{\sigma_n} \xrightarrow{\text{d}} Z \sim N(0,1)$$

9.2 Lleis dels grans números

Llei dèbil Una successió $\{X_j\}$ compleix la llei respecte d'una successió numèrica $b_n > 0 \to \infty$ si

$$\frac{S_n - \mathrm{E}[S_n]}{b_n} \xrightarrow{\text{P}} 0$$

Si $b_n = n$ la condició anterior és:

$$S_n/n - \mathrm{E}[S_n/n] \xrightarrow{\text{P}} 0$$

Llei forta En les condicions anteriors:

$$\frac{S_n - \mathrm{E}[S_n]}{b_n} \xrightarrow{\text{q.s.}} 0$$

Propietats

1. Llei dèbil:

$$\frac{\sigma^2(S_n)}{b_n^2} \to 0 \Rightarrow \frac{S_n - \mathrm{E}[S_n]}{b_n} \xrightarrow{\mathcal{L}^2} 0$$

2. Llei forta: Si $\{X_j\}$ són variables independents

$$\sum_{j=1}^{\infty} \frac{\sigma_j^2}{b_j^2} < \infty \Rightarrow \frac{S_n - \mathrm{E}[S_n]}{b_n} \xrightarrow{\text{q.s.}} 0$$

Bibliografia

[1] Ricardo Vélez Ibarrola. *Cálculo de probabilidades 2*. Ediciones Académicas, S.A., 2004.

6. Inversió:
$$f(x)=\frac{1}{2\pi}\int_{\infty}^{\infty}\mathrm{e}^{-i\,t\,x}\,\psi(t)\,dt$$

7. En una distribució k-dimensional la independència implica que $\psi(t_1,\cdots t_k)=\psi(t_1)\cdots\psi(t_k)$.

Capítol 8
Distribució normal k-dimensional

$$N_k(\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{1}{\sqrt{(2\pi)^k\,|\Sigma|}}\,\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}}$$

Propietats

1. Si $\boldsymbol{X}\sim N_K(\boldsymbol{\mu},\boldsymbol{\Sigma})$ ($\sim$ vol dir "distribuït com") i $\boldsymbol{Y}=\boldsymbol{X}\,\boldsymbol{B}$, on $\boldsymbol{B}$ és una matriu $k\times m$, aleshores $\boldsymbol{Y}\sim N_m(\boldsymbol{\mu}\,\boldsymbol{B},\boldsymbol{B}^{\mathrm{T}}\,\boldsymbol{\Sigma}\,\boldsymbol{B})$.
2. Si $\boldsymbol{X}=(\boldsymbol{X}_1,\boldsymbol{X}_2)\sim N_K(\boldsymbol{\mu},\boldsymbol{\Sigma})$ amb
$$\boldsymbol{\mu}=(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\quad \boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{1,1}, & \boldsymbol{\Sigma}_{1,2}\\ \boldsymbol{\Sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2}\end{bmatrix}$$
aleshores $\boldsymbol{X}_1\sim N_r(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{1,1})$ i $\boldsymbol{X}_2\sim N_{k-r}(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{2,2})$.
3. En les condicions anteriors $\boldsymbol{X}_1,\boldsymbol{X}_2$ són independents si i només si $\boldsymbol{\Sigma}_{1,2}=\boldsymbol{0}$.
4. En les condicions anteriors
$$F(\boldsymbol{X}_1|\boldsymbol{X}_2=\boldsymbol{x}_2)=N_r(\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c)$$
on:
$$\boldsymbol{\mu}_c=\boldsymbol{\mu}_1+(\boldsymbol{x}_2-\boldsymbol{\mu}_2)\,\boldsymbol{\Sigma}_{2,2}^{-1}\,\boldsymbol{\Sigma}_{2,1}$$
$$\boldsymbol{\Sigma}_c=\boldsymbol{\Sigma}_{1,1}-\boldsymbol{\Sigma}_{1,2}\,\boldsymbol{\Sigma}_{2,2}\,\boldsymbol{\Sigma}_{2,1}$$

Exemple 8.1 Distribució normal k-dimensional[1]
Donat $(X,Y)\sim N_K(\boldsymbol{\mu},\boldsymbol{\Sigma})$ amb $\boldsymbol{\mu}=(0,0)$, $\boldsymbol{\Sigma}=\begin{bmatrix}2 & \sqrt{2}\\ \sqrt{2} & 3\end{bmatrix}$, determinar la combinació lineal (U,V) de (X,Y) tal que U i V siguin iid[2] amb distribució N(0,1).

Solució Sigui $(U,V)=(X,Y)\,\boldsymbol{C}$. Aleshores $\boldsymbol{\Sigma}_{(U,V)}=\boldsymbol{C}^{\mathrm{T}}\,\boldsymbol{\Sigma}_{(X,Y)}\,\boldsymbol{C}$, on és possible trobar la matriu ortogonal $\boldsymbol{C}$ tal que $\boldsymbol{\Sigma}_{(U,V)}=\boldsymbol{C}^{\mathrm{T}}\,\boldsymbol{\Sigma}_{(X,Y)}\,\boldsymbol{C}=\boldsymbol{\Lambda}$ (veure la secció 6.3, pàg. 13). Els autovalors de $\boldsymbol{\Sigma}_{(X,Y)}$ són $\lambda_1=1$, $\lambda_2=4$. De $\boldsymbol{\Sigma}_{(X,Y)}\,\boldsymbol{C}=\boldsymbol{C}\,\boldsymbol{\Lambda}=\begin{bmatrix}\lambda_i\,C_1 & \lambda_2\,C_2 & \cdots\end{bmatrix}$, on C_i són les columnes de $\boldsymbol{C}$, tenim que $\boldsymbol{C}$ es pot calcular com els autovectors per la dreta de mòdul 1 (doncs ha de ser $\boldsymbol{C}\,\boldsymbol{C}^{\mathrm{T}}=\mathbf{I}$): $(\boldsymbol{\Sigma}_{(X,Y)}-\lambda_i\mathbf{I})\,\boldsymbol{x}_i=0, \boldsymbol{x}_i\,\boldsymbol{x}_i^{\mathrm{T}}=1$. Al ser $(\boldsymbol{\Sigma}_{(X,Y)}-\lambda_i\,\mathbf{I})$ una matriu 2×2 de determinant nul, basta agafar la primera fila, $\begin{bmatrix}2-\lambda_i & \sqrt{2}\end{bmatrix}$:
$$\lambda_1=1\Rightarrow\begin{bmatrix}1 & \sqrt{2}\end{bmatrix}\,\boldsymbol{x}_1=0\Rightarrow\boldsymbol{x}_1=\begin{bmatrix}\sqrt{2} & -1\end{bmatrix}\frac{1}{\sqrt{3}}$$
$$\lambda_2=4\Rightarrow\begin{bmatrix}-2 & \sqrt{2}\end{bmatrix}\,\boldsymbol{x}_1=0\Rightarrow\boldsymbol{x}_2=\begin{bmatrix}\sqrt{2} & 2\end{bmatrix}\frac{1}{\sqrt{6}}$$
d'on: $\boldsymbol{C}=\begin{bmatrix}\sqrt{2/3} & 1/\sqrt{3}\\ -1/\sqrt{3} & \sqrt{2/3}\end{bmatrix}$. Amb el canvi $(U,V)=(X,Y)\,\boldsymbol{C}$, U,V seran independents amb variances λ_1, λ_2, respectivament. Perquè les variances siguin 1 hem de multiplicar per $\boldsymbol{C}\,\boldsymbol{\Lambda}^{-1/2}$ (doncs així serà $(\boldsymbol{C}\,\boldsymbol{\Lambda}^{-1/2})^{\mathrm{T}}\,\boldsymbol{\Sigma}_{(X,Y)}\,(\boldsymbol{C}\,\boldsymbol{\Lambda}^{-1/2})=\mathbf{I}$). $\boldsymbol{C}\,\boldsymbol{\Lambda}^{-1/2}=\begin{bmatrix}1/\sqrt{\lambda_i}\,C_1 & 1/\sqrt{\lambda_2}\,C_2 & \cdots\end{bmatrix}$, on C_i són les columnes de $\boldsymbol{C}$. Per tant, la combinació lineal demanada és
$$(U,V)=(X,Y)\begin{bmatrix}\sqrt{2/3} & \sqrt{3}/6\\ -1/\sqrt{3} & 1/\sqrt{6}\end{bmatrix}\quad\blacksquare$$

Capítol 9
Convergència

Puntual És molt estricta i rarament es fa servir.
$$\{\omega\in\Omega\mid X_n(\omega)\to X\}=\{X_n\to X\}=\Omega$$

Convergència quasi-segura $X_n\xrightarrow{\text{q.s.}}X$
$$P(\{\omega\in\Omega\mid X_n(\omega)\to X\})=P(\{X_n\to X\})=1$$

Convergència en probabilitat $X_n\xrightarrow{\mathrm{P}}X$
$$\lim_{n\to\infty}P(|X_n-X|>\epsilon\})=0,\forall\epsilon>0$$

Convergència en distribució $X_n\xrightarrow{\mathrm{d}}X$
$$F_n(x)\to F(x),\forall x\in\mathbb{R}$$

Convergència en mitjana d'ordre p $X_n\xrightarrow{\mathcal{L}^p}X$
$$\lim_{n\to\infty}\mathrm{E}[|X_n-X|^p]=0$$

[1]Enunciat del primer problema de l'examen de l'UNED de juny 2009, segona setmana.
[2]iid: Independents i idènticament distribuïdes.

Probabilitat en un espai mesurable $(\Omega,\mathcal{F})$ és una aplicació $\mathrm{P}:\mathcal{F}\to\mathbb{R}$ que verifica:
- $\mathrm{P}(A)\geq 0$ per $\forall A\in\mathcal{F}$.
- (σ -additivitat) Per qualsevol col·lecció numerable de successos disjunts $\{A_n\}\subset\mathcal{F}$:
$$\mathrm{P}(\cup A_n)=\sum_n\mathrm{P}(A_n)$$
- $\mathrm{P}(\Omega)=1$.

Aquestes tres condicions s'anomenen **Axiomàtica de Kolmogorov**.

Espai de probabilitat és la terna $(\Omega,\mathcal{F},\mathrm{P})$ on P és una probabilitat en l'espai mesurable $(\Omega,\mathcal{F})$.

Independència Els successos $\{A_i\}_{i\in I}$ són independents si:
$$\mathrm{P}(\cap_{i\in F}A_i)=\prod_{i\in F}\mathrm{P}(A_i)$$
per qualsevol subconjunt finit $F\subset I$.

Probabilitat condicionada Si $\mathrm{P}(B)>0$, aleshores:
$$\mathrm{P}(A\mid B)=\frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}$$

Distribució uniforme és fa servir aquest nom per designar l'assignació de probabilitat més senzilla, que consisteix en assignar a cada conjunt $B\subset\Omega\in\mathbb{B}$, una probabilitat proporcional a la seva longitud:
$$\mathrm{P}(B)=\lambda(B)/\lambda(\Omega). \tag{1.2}$$

Més genèricament, per a qualsevol conjunt $B\in\mathbb{B}$, s'assigna una probabilitat proporcional a la seva intersecció amb $\Omega\in\mathbb{B}$:
$$\mathrm{P}(B)=\lambda(B\cap\Omega)/\lambda(\Omega). \tag{1.3}$$

En $\mathbb{R}^k$ es generalitza assignant a cada conjunt $V\in\mathbb{B}^n$, una probabilitat proporcional a la intersecció del seu volum amb $\Omega\in\mathbb{B}^n$ (àrea en $\mathbb{R}^2$):
$$\mathrm{P}(V)=\lambda_n(V\cap\Omega)/\lambda_n(\Omega). \tag{1.4}$$

Proposició 1.2 Si $A\in\Omega^r$, $B\in\Omega^s$, on $\Omega^r\in\mathbb{B}^r$, $\Omega^s\in\mathbb{B}^s$ són dos intervals triats a l'atzar i independentment un de l'altra, aleshores el conjunt $A\times B\in\mathbb{B}^r\times\mathbb{B}^s$ té una probabilitat distribuïda uniformement en $\Omega^r\times\Omega^s$.

Demostració. Tenim que
$$\mathrm{P}(A)=\lambda(A)/\lambda(\Omega^r)=\lambda(A\cap\Omega^r)/\lambda(\Omega^r),$$
$$\mathrm{P}(B)=\lambda(B)/\lambda(\Omega^s)=\lambda(B\cap\Omega^s)/\lambda(\Omega^s).$$
Per independència:
$$\mathrm{P}(A\times B)=\frac{\lambda(A\cap\Omega^r)\,\lambda(B\cap\Omega^s)}{\lambda(\Omega^r)\,\lambda(\Omega^s)}.$$
Aplicant la proposició 1.1, pàg. 2:
$$\mathrm{P}(A\times B)=\frac{\lambda((A\cap\Omega^r)\times(B\cap\Omega^s))}{\lambda(\Omega^r\times\Omega^s)}=\frac{\lambda((A\times B)\cap(\Omega^r\times\Omega^s))}{\lambda(\Omega^r\times\Omega^s)}\quad\square$$

En particular, el resultat també és vàlid si A i B es redueixen a un sòl punt. Per exemple, triar dos punts (X,Y) a l'atzar independentment un de l'altra en l'interval $[0,1]$, és equivalent a triar un punt de coordenades (X,Y) aleatòriament en el quadrat $[0,1]\times[0,1]$.

1.3 Propietats

Fórmula d'inclusió-exclusió
$$\mathrm{P}(\cup_{i=1}^nA_i)=\sum_{i=1}^n\mathrm{P}(A_i)-\sum_{1\leq i<j\leq n}\mathrm{P}(A_i\cap A_j)+\sum_{1\leq i<j<k\leq n}\mathrm{P}(A_i\cap A_j\cap A_k)+\cdots+(-1)^{n-1}\,\mathrm{P}(A_i\cap\ldots\cap A_k)$$

Continuïtat de seqüències monòtones de successos
1. Si $\{A_n\}_{n\geq 1}$ és una seqüència creixent de successos, és a dir, $A_n\subset A_{n+1}$, aleshores:
$$\mathrm{P}(\cup_{n=1}^\infty A_n)=\lim_{n\to\infty}\mathrm{P}(A_n)$$
2. Si $\{B_n\}_{n\geq 1}$ és una seqüència decreixent de successos, és a dir, $B_n\supset B_{n+1}$, aleshores:
$$\mathrm{P}(\cap_{n=1}^\infty B_n)=\lim_{n\to\infty}\mathrm{P}(B_n)$$

Demostració.
1. Podem fer la partició del succés A_n:
$$A_n=A_1\cup(A_2\setminus A_1)\cup\cdots(A_n\setminus A_{n-1})$$
d'on:
$$\cup_{k=1}^\infty A_k=A_1\cup(A_2\setminus A_1)\cup(A_3\setminus A_2)\cup\cdots$$

per tant:

$$\mathrm{P}(\cup_{k=1}^{\infty} A_k) = \mathrm{P}(A_1) + \sum_{j=2}^{\infty} \mathrm{P}(A_j \setminus A_{j-1}) =$$
$$\lim_{n\to\infty} \left\{ \mathrm{P}(A_1) + \sum_{j=2}^{n} \mathrm{P}(A_j \setminus A_{j-1}) \right\} =$$
$$\lim_{n\to\infty} \mathrm{P}(A_1 \cup (A_2 \setminus A_1) \cup \cdots (A_n \setminus A_{n-1})) =$$
$$\lim_{n\to\infty} \mathrm{P}(A_n)$$

2. Basta escriure:

$$\mathrm{P}(\cap_{n=1}^{\infty} B_n) = 1 - \mathrm{P}((\cap_{n=1}^{\infty} B_n)^c) = 1 - \mathrm{P}(\cup_{n=1}^{\infty} B_n^c)$$

i, aplicant el resultat anterior:

$$\mathrm{P}(\cup_{n=1}^{\infty} B_n^c) = \lim_{n\to\infty} \mathrm{P}(B_n^c) = 1 - \lim_{n\to\infty} \mathrm{P}(B_n) \quad \square$$

Límits inferior i superior d'una successió Donada una seqüència de successos $\{A_n\}_{n\geq 1}$, estem interessats en determinar la probabilitat de que la seqüència es realitza un nombre infinit de vegades. Per això analitzarem les seqüències decreixent $\{D_m\}_{m\geq 1}$ i creixent $\{C_m\}_{m\geq 1}$ definides com:

$$D_m = \cap_{n=m}^{\infty} A_n$$
$$C_m = \cup_{n=m}^{\infty} A_n.$$

A_n es realitzarà un nombre infinit de vegades si per algun m hi ha un o més esdeveniments que pertanyen a tots els successos A_n. El succés format per aquests esdeveniments ve donat per:

$$\liminf_n A_n = \cup_{m=1}^{\infty} D_m = \cup_{m=1}^{\infty} \cap_{n=m}^{\infty} A_n.$$

Una condició menys restrictiva és que per a cada $m \geq 1$ es realitzi algun A_n amb $n \geq m$. El succés format per aquests esdeveniments ve donat per:

$$\limsup_n A_n = \cap_{m=1}^{\infty} C_m = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} A_n.$$

Sempre és

$$\liminf_n A_n \subseteq \limsup_n A_n.$$

Si $\liminf_n A_n = \limsup_n A_n$, aleshores es diu que la successió és convergent, i es designa simplement per $\lim_n A_n$.

- En general:
$$\mathrm{P}(\liminf_n A_n) \leq \liminf_n \mathrm{P}(A_n) \leq \mathrm{P}(\limsup_n A_n) \leq \limsup_n \mathrm{P}(A_n)$$
- Si A_n és convergent:
$$\mathrm{P}(\lim_n A_n) = \lim_n \mathrm{P}(A_n)$$

Lemes de Borel-Cantelli

1. Si la serie $\sum_{n=1}^{\infty} \mathrm{P}(A_n)$ és convergent, aleshores:
$$\mathrm{P}(\limsup_n A_n) = 0$$
2. Si $\{A_n\}$ és una successió de successos independents tal que la sèrie $\sum_{n=1}^{\infty} \mathrm{P}(A_n)$ és divergent, aleshores:
$$\mathrm{P}(\limsup_n A_n) = 1$$

Per tant, en una successió de successos independents, la probabilitat de hi hagi infinits successos $\{A_n\}$ només pot valer 0 o 1, segons $\sum_{n=1}^{\infty} \mathrm{P}(A_n)$ sigui convergent o no.

Demostració.

1. La successió $B_n = \cup_{n=m}^{\infty} A_n$ és decreixent cap a $\limsup_n A_n$, per tant:
$$\mathrm{P}(\limsup_n A_n) = \lim_{m\to\infty} \mathrm{P}(\cup_{n=m}^{\infty} A_n) \leq \lim_{m\to\infty} \mathrm{P}(A_n)$$
i l'últim terme és zero si la sèrie és convergent. $\square$

Regla de la multiplicació Si $\mathrm{P}(\cap_{i=1}^{n-1} A_i) > 0$, aleshores:

$$\mathrm{P}(\cap_{i=1}^{n} A_i) = \mathrm{P}(A_1)\,\mathrm{P}(A_2|A_1)\times$$
$$\mathrm{P}(A_3|A_2 \cap A_1) \cdots \mathrm{P}(A_n \mid \cap_{i=1}^{n-1} A_i)$$

Fórmula de les probabilitats totals Si B_n és una partició numerable d'Ω amb $\mathrm{P}(B_n) > 0$, aleshores:

$$\mathrm{P}(A) = \sum_n \mathrm{P}(B_n)\,\mathrm{P}(A \mid B_n)$$

Fórmula de Bayes Si $\mathrm{P}(B) > 0$, aleshores:

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A)\,\mathrm{P}(B \mid A)}{\mathrm{P}(B)}$$

Capítol 2
Probabilitats en $\mathbb{R}$

2.1 Variables aleatòries

Són funcions X que assignen un valor numèric a cada esdeveniment $\omega \in \Omega$. Formalment, una variable aleatòria (VA) definida en un espai de probabilitat $(\Omega,\mathcal{F},\mathrm{P})$ és una funció $X : \Omega \to \mathbb{R}$ tal que $X^{-1}(B) \in \mathcal{F}$ per a cada $B \in \mathbb{B}$. $X^{-1}(B)$ s'anomena la **contra-imatge** de B, i és el succés format per els esdeveniments: $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}$. Normalment s'indica només $X^{-1}(B) = \{X \in B\}$.

6.3 Transformacions lineals

Sigui la VA k-dimensional $\boldsymbol{X}$ amb vector de mitjanes $\boldsymbol{\mu}$ i matriu de covariances $\boldsymbol{\Sigma}$. Si $\boldsymbol{C}$ és una matriu $k \times m$, aleshores $\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{C}$ és una VA m-dimensional amb:

$$\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{C} \Rightarrow \begin{cases} \mathrm{E}[\boldsymbol{Y}] = \boldsymbol{\mu}\,\boldsymbol{C} \\ \boldsymbol{\Sigma}_Y = \boldsymbol{C}^{\mathrm{T}}\,\boldsymbol{\Sigma}\,\boldsymbol{C} \end{cases}$$

Notar que si s'agafa $\boldsymbol{C}$ tal que diagonalitza $\boldsymbol{\Sigma}$: $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{C} = \boldsymbol{\Lambda}$, aleshores $\boldsymbol{\Sigma}_Y = \boldsymbol{C}^{\mathrm{T}}\,\boldsymbol{\Sigma}\,\boldsymbol{C} = \boldsymbol{\Lambda}$. Per tant, les VA Y_i estaran incorrelades. Si el rang de $\boldsymbol{\Sigma}$ és $r < k$, aleshores $\boldsymbol{\Lambda}$ tindrà $k - r$ elements de la diagonal nuls. Per tant, hi haurà $k - r$ variables Y_i iguals a la mitjana $\mathrm{E}[Y_i] = (\boldsymbol{\mu}\,\boldsymbol{C})_i$ amb probabilitat 1. Com que $\boldsymbol{X} = \boldsymbol{Y}\,\boldsymbol{C}$, serà:

$$X_i = \sum_{j=1}^{r} c_{i,j}\,Y_j + \sum_{j=r+1}^{k} c_{i,j}\,\mathrm{E}[Y_j]$$

Per tant, la distribució serà singular i concentrada en un hiperplà de dimensió $r < k$.

El determinant $|\boldsymbol{\Sigma}|$ s'anomena **variança generalitzada** de $\boldsymbol{X}$. Si és nul, $\boldsymbol{X}$ no té dispersió k-dimensional (doncs la probabilitat es concentra en un hiperplà de dimensió inferior a k). Altrament, $|\boldsymbol{\Sigma}| = |\boldsymbol{\Lambda}|$ proporciona el producte de les variances de les VA incorrelades que resulten de $\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{C}$, amb $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{C} = \boldsymbol{\Lambda}$.

6.4 Regressió bidimensional

Consisteix en realitzar la millor previsió de X_2 conegut que $X_1 = x_1$, $\hat{x}_2 = g(x_1)$. L'**error residual** de l'estimació és $\eta_2 = X_2 - \hat{x}_2$ El criteri sòl ser minimitzar $\mathrm{E}[\eta_2^2]$.

Recta de regressió L'estimació és $\hat{x}_2 = a\,x_1 + b$. La recta que minimitza l'error és:

$$x_2 - \mu_2 = \frac{\sigma_{1,2}}{\sigma_1^2}(x_1 - \mu_1)$$
$$= \frac{\mathrm{Cov}(X_1,X_2)}{\sigma^2(X_1)}(x_1 - \mu_1) \quad (6.1)$$

O també:

$$\frac{x_2 - \mu_2}{\sigma_2} = \frac{\sigma_{1,2}}{\sigma_1\,\sigma_2}\,\frac{x_1 - \mu_1}{\sigma_1} = \rho\,\frac{x_1 - \mu_1}{\sigma_1} \quad (6.2)$$

En termes de les variables tipificades $Z_1 = \frac{X_1 - \mu_1}{\sigma_1}$ $Z_2 = \frac{X_2 - \mu_2}{\sigma_2}$, l'equació (6.2) és:

$$z_2 = \rho\,z_1.$$

L'**error residual** de l'estimació $\eta_2 = X_2 - \hat{x}_2$ té mitjana 0 i **variança residual**

$$\mathrm{E}[\eta_2^2] = \sigma_2^2\,(1 - \rho^2)$$

Per simetria la regressió de X_1 sobre X_2 ve donada per la recta $x_1 - \mu_1 = \frac{\sigma_{1,2}}{\sigma_2^2}(x_2 - \mu_2)$. Ambdues rectes es tallen en (μ_1,μ_2)

Corbes de regressió Es busca la funció $h(x_1)$ que minimitza la variança residual:

$$\mathrm{E}[(X_2 - h(X_1))^2] = \int_{\mathbb{R}} \mathrm{E}[(X_2 - h(x_1))^2 | X_1 = x_1]\,F(dx_1)$$

que representa la dispersió de $X_2|X_1$ al voltant de $h(x_1)$, que és mínima quan $h(x_1)$ és igual a la seva mitjana. Per tant, la millor funció h és

$$\mu_2(x_1) = \mathrm{E}[X_2|X_1 = x_1]$$

L'error residual també té mitjana 0. La seva variança residual és:

$$\mathrm{E}[(X_2 - \mu_2(X1))^2] = \sigma_2^2\,(1 - \vartheta_2^2)$$

on es defineix la **raó de correlació** de X_2 sobre X_1

$$\vartheta_2 = \frac{\sqrt{\mathrm{E}[(\mu_2(X1) - \mu_2)^2]}}{\sigma_2} = \frac{\sqrt{\mathrm{E}[\mu_2(X1)^2] - \mu_2^2}}{\sigma_2}$$

Capítol 7
Funció característica

En cas de VA no negatives és convenient la transformada de Laplace:

$$g(t) = \mathrm{E}[\mathrm{e}^{-t\,X}] = \int_0^{\infty} \mathrm{e}^{-t\,x}\,f(x)\,dx$$

Per signe arbitrari és millor la transformada de Fourier[1]:

$$\psi(t) = \mathrm{E}[\mathrm{e}^{i\,t\,X}] = \int_{-\infty}^{\infty} \mathrm{e}^{i\,t\,x}\,f(x)\,dx$$

Propietats

1. $|\psi(t)| \leq \psi(0) = 1$.
2. $\psi(-t) = \psi(t)^*$ (* representa el conjugat).
3. La FC de $a\,X + b$ és $\psi(a\,t)\,\mathrm{e}^{i\,t\,b}$.
4. La FC de $X_1 + X_2$ és el producte $\psi_1(t)\,\psi_2(t)$.
5. La FC és real si i només si la distribució és simètrica ($f(x) = f(-x)$).

[1] En enginyeria es sol definir la transformada de Fourier amb signe canviat: $\mathcal{F}[f(x)] = \int_{-\infty}^{\infty} \mathrm{e}^{-i\,t\,x}\,f(x)\,dx$

Covariança

$$\mathrm{Cov}(X_1,X_2) = \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)]$$

Propietats de la variança

1. $\sigma^2(a\,X+b) = a^2\,\sigma^2(X)$
2. $\sigma^2(X_1+X_2) = \sigma^2(X_1)+\sigma^2(X_2)+2\,\mathrm{Cov}(X_1,X_2)$
3. Si X_1,X_2 són independents:
$$\mathrm{Cov}(X_1,X_2) = 0$$
$$\sigma^2(X_1+X_2) = \sigma^2(X_1)+\sigma^2(X_2)$$
4. $$\mu = \arg\min_a \mathrm{E}[(X-a)^2]$$
$$\sigma^2 = \min_a \mathrm{E}[(X-a)^2]$$
5. $\mathrm{Cov}(X_1,X_2) = \mathrm{E}[X_1\,X_2]-\mu_1\,\mu_2$
6. $\mathrm{Cov}(X_1,X_2) = \mathrm{Cov}(X_2,X_1)$
7. Desigualtat d'Schwarz
$$|\mathrm{Cov}(X_1,X_2)| \leq \sigma(X_1)\,\sigma(X_2)$$

Coeficient de correlació

$$\rho(X_1,X_2) = \frac{\mathrm{Cov}(X_1,X_2)}{\sigma(X_1)\,\sigma(X_2)}$$

on $-1 \leq \rho(X_1,X_2) \leq 1$. Si són independents $\rho(X_1,X_2)=0$. Si $\rho(X_1,X_2)=0$ s'anomenen **incorrelades**. Per tant, si són independents són incorrelades, però el contrari no és cert.

Coeficient d'asimetria *skewness coefficient*

$$\gamma_3 = \frac{\mu_3}{\sigma^3} = \frac{\mathrm{E}[(X-\mu)^3]}{\sigma^3}$$

Coeficient d'apuntament o curtosis mesura en que la probabilitat està concentrada al voltant de μ

$$\gamma_4 = \frac{\mu_4}{\sigma^4} - 3$$

on s'agafa com a referència la distribució normal, per a la que $\gamma_4 = 0$.

Variable tipificada

$$Z = \frac{X-\mu}{\sigma} = \frac{X-\mathrm{E}[X]}{\sqrt{\mathrm{E}[(X-\mu)^2]}}$$

que té: $\mathrm{E}[Z]=0$, $\sigma(Z)=1$.

Moda correspon al punt on la probabilitat és màxima. Dóna idea del valor més probable.

Mediana és un punt M (pot no ser únic) que verifica

$$F(M^-) \leq 1/2$$
$$F(M) \geq 1/2$$

Es compleix $M = \arg\min_a \mathrm{E}[|X-a|]$.

Quantil d'ordre p és un punt c_p (pot no ser únic) que verifica

$$F(c_p^-) \leq p$$
$$F(c_p) \geq p$$

La mediana és $M = c_{1/2}$. $c_{1/4}, c_{3/4}$ s'anomenen **quartils**. Notar que si F és continua, l'interval $I = [c_{1/4}, c_{3/4}]$ té probabilitat 1/2.

6.2 En $\mathbb{R}^k$

Moment respecte l'origen d'ordre $r>0$

$$\alpha_{r_1,\cdots r_k} = \mathrm{E}[X_1^r \cdots X_k^{r_k}]$$

on $r_1+\cdots r_k = r$.

Moments centrals d'ordre $r>0$

$$\mu_{r_1,\cdots r_k} = \mathrm{E}[(X_1-\mu)^{r_1}\cdots(X_k-\mu)^{r_k}]$$

on $r_1+\cdots r_k = r$. Només tenen interès els d'ordre 2. Notació:

$$\begin{aligned}
\text{Variança}\quad & \sigma_{i,i} = \sigma_i^2 = \sigma^2(X_i) = \mathrm{E}[(X-\mu_i)^2] \\
\text{Covariança}\quad & \sigma_{i,j} = \mathrm{Cov}(X_i,X_j) = \mathrm{E}[(X-\mu_i)(X-\mu_j)] \\
\text{Des. típica}\quad & \sigma_i = \sqrt{\sigma_{i,i}} = \sigma(X_i) \\
\text{Matriu de Cov.}\quad & \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{1,1} & \cdots & \sigma_{1,k} \\ \vdots & \vdots & \vdots \\ \sigma_{k,1} & \cdots & \sigma_{k,k} \end{bmatrix}
\end{aligned}$$

La matriu de covariances és quadrada simètrica i semidefinida positiva, per tant, es pot diagonalitzar en la forma $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{C} = \boldsymbol{\Lambda}$ on $\boldsymbol{\Lambda}$ és la matriu diagonal amb els autovalors de $\boldsymbol{\Sigma}$, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\cdots\lambda_k)$, i $\boldsymbol{C}$ és una matriu ortogonal, és a dir: $\boldsymbol{C}^{-1} = \boldsymbol{C}^{\mathrm{T}}$ (veure l'exemple 8.1, pàg. 14).

Indicador Qualsevol succés A té associat la variable aleatòria I_A, anomenat indicador de A, definida com:

$$I_A = \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{si } \omega \notin A \end{cases} \tag{2.1}$$

2.2 Funció de distribució

La mesura de probabilitat $\mathrm{P}_X(B) = \mathrm{P}\{X \in B\}$, $\forall B \in \mathbb{B}$, s'anomena la distribució de la VA X i $(\mathbb{R},\mathbb{B},\mathrm{P}_X)$ l'espai de probabilitat induït per X. La funció

$$F_X(x) = \mathrm{P}_X((-\infty,x]) = \mathrm{P}\{X \leq x\} \tag{2.2}$$

s'anomena **funció de distribució de la variable aleatòria** X (*cumulative distribution function, CDF*).

Una funció de distribució té les següent propietats:

- És una funció $F:\mathbb{R}\to[0,1]$.
- És creixent.
- És continua per la dreta: $F(x) = \lim_{y\downarrow x} F(y)$.
- Verifica $F(-\infty)=0$, $F(\infty)=1$.

$$\begin{aligned}
\mathrm{P}((a,b]) &= \mathrm{P}((-\infty,b]) - \mathrm{P}((-\infty,a]) = F(b)-F(a) \\
\mathrm{P}((a,b)) &= \mathrm{P}((-\infty,b)) - \mathrm{P}((-\infty,a]) = F(b^-)-F(a) \\
\mathrm{P}([a,b]) &= \mathrm{P}((-\infty,b]) - \mathrm{P}((-\infty,a)) = F(b)-F(a^-) \\
\mathrm{P}((a,b]) &= \mathrm{P}((-\infty,b]) - \mathrm{P}((-\infty,a]) = F(b)-F(a) \\
\mathrm{P}([a,b)) &= \mathrm{P}((-\infty,b)) - \mathrm{P}((-\infty,a)) = F(b^-)-F(a^-)
\end{aligned}$$

Així doncs, $\mathrm{P}(\{x\}) = \mathrm{P}([x,x]) = F(x)-F(x^-)$, que implica que els punts on és $\mathrm{P}(x)>0$ són exactament els punts de discontinuïtat de salt (per la dreta) de F.

Una distribució pot estar **concentrada** en un subconjunt $\Omega\subset\mathbb{R}$. En aquest cas $\mathrm{P}(B)=0$ per qualsevol $\{B\in\mathbb{B}\mid B\subset\Omega^c\}$.

Qualsevol distribució es pot expressar com la **mixtura** de dues distribucions: $F(x) = \alpha\,F_c(x)+(1-\alpha)\,F_d(x)$, on $F_c(x)$ és una distribució continua i $F_d(x)$ és una funció de distribució **discreta**, esglaonada amb discontinuïtats en un conjunt numerable de punts.

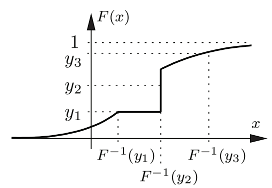

Figura 2.1: Funció de distribució inversa.

Funció de distribució inversa En general una funció de distribució pot tenir punts de discontinuïtat de salt per la dreta, o intervals on és constant (veure la figura 2.1). Així doncs, una funció de distribució pot no ser una funció bijectiva i, per tant, no tenir inversa. Per això es defineix la **funció de distribució inversa** (*Inverse distribution function*) anomenada també **funció quantil** (*quantile function*)[1]:

$$F^{-1}(y) = \min\{x\in\mathbb{R}\mid F(x)\geq y\} \tag{2.3}$$

Notar que F^{-1} coincideix amb la funció inversa de F quan aquesta és bijectiva.

2.3 Funció de densitat

Una VA es diu absolutament contínua si ho és la seva funció de distribució. En aquest cas existeix una **funció de densitat** (*probability density function, PDF*), $f(x)$, tal que per a qualsevol interval $I=\{x\mid a\leq x\leq b\}$:

$$\int_a^b f(x)\,dx = \mathrm{P}(X\in I) = F(b)-F(a) \tag{2.4}$$

Es diu que la funció de densitat està **concentrada** en un interval I quan $f(x)=0$, $x\notin I$. Evidentment:

$$F(x) = \mathrm{P}(X\in(-\infty,x]) = \int_{-\infty}^x f(y)\,dy$$

De les propietats de F és immediat que $f(x)\geq 0$ i $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Les VA absolutament contínues solen caracteritzar-se amb la seva funció de densitat (veure l'apèndix A). Hi ha distribucions continues que no tenen funció de densitat, com ara la distribució de Cantor (veure [1, secció 4.4.3]). Tenen la característica d'estar associades a un conjunt de Borel $L\in\mathbb{B}$ no numerable i de longitud nul·la tal que $\mathrm{P}(x)=0$, $\forall x\in L$, però $\mathrm{P}(L)>0$. S'anomenen **distribucions singulars**.

2.4 Canvis de variable

Si X és una VA, i $g:\mathbb{R}\to\mathbb{R}$ és una funció mesurable respecte $\mathbb{B}$ (és a dir, $g^{-1}(B)\in\mathbb{B}$ per a qualsevol $B\in\mathbb{B}$), aleshores $Y=g(X)$ també és una VA amb

$$F_Y(y) = \mathrm{P}\{Y\leq y\} = \mathrm{P}(g(X)\leq y) = \mathrm{P}\{X\in g^{-1}(-\infty,y]\}$$

que es pot determinar fàcilment a partir de F_X, amb l'única dificultat de determinar els subconjunts donats per $g^{-1}(-\infty,y]$.

[1] En el llibre de l'assignatura [1] es fa servir la notació F^* per denotar la funció de distribució inversa.

Si X és una VA absolutament contínua amb funció de densitat f_X concentrada en I, i $g : I \to \mathbb{R}$ és una funció monòtona i derivada contínua en I, aleshores $Y = g(X)$ és una VA absolutament contínua amb densitat concentrada en $y \in g(I)$:

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|} = f_X(g^{-1}(y))\,|(g^{-1})'(y)| \quad (2.5)$$

Demostració. Si g és estrictament creixent:

$$F_Y(y) = \mathrm{P}(Y \le y) = \mathrm{P}(g(X) \le y) = \mathrm{P}(X \le g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(x)\,dx = \int_{-\infty}^{x} \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}\,dx$$

on s'ha fet el canvi $x = g^{-1}(y)$, d'on $dy = g'(x)\,dx$ i $dx = dy/g'(g^{-1}(y))$. Si g és estrictament decreixent s'obté el mateix amb signe canviat. □

Si g no és monòtona en I, però es pot descompondre en intervals on ho és, aleshores es pot aplicar el resultat 2.5 en cada interval i la densitat de $Y = g(X)$ serà:

$$f_Y(y) = \sum_{x \in g^{-1}(y)} \frac{f_X(x)}{|g'(x)|}. \quad (2.6)$$

Finalment, si g és constant en algun interval U ($g(x \in U) = y_0$), aleshores Y no serà absolutament contínua (tindrà una distribució mixta), amb:

$$F_Y(y_0) = \mathrm{P}(Y = y_0) = \mathrm{P}(X \in U) = \int_U f_X(x)\,dx$$

Generació de variables aleatòries

Proposició 2.1 Si X és una VA absolutament contínua, aleshores la VA $Y = F_X(X)$, $Y \in [0,1]$ té una distribució uniforme.

Demostració.

$$F_Y(y) = \mathrm{P}(Y \le y) = \mathrm{P}(F_X(X) \le y) = \mathrm{P}(X \le F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y,\ y \in [0,1] \quad \square$$

La proposició anterior suggereix el següent mètode per a generar una VA amb distribució F_X qualsevol, i funció de distribució inversa F_X^{-1}:

Proposició 2.2 Si Y és una VA amb distribució uniforme en $Y \in [0,1]$, aleshores la VA $X = F_X^{-1}(Y)$ té la funció de distribució F_X.

Demostració.

$$\mathrm{P}(X \le x) = \mathrm{P}(F_X^{-1}(Y) \le x) = \mathrm{P}(Y \le F_X(x)) = F_Y(F_X(x)) = F_X(x) \quad \square$$

Per tant, per a generar la VA amb distribució F_X basta generar una VA Y uniforme en $[0,1]$, i substituir $X = F_X^{-1}(Y)$.

2.5 Independència

Definició 2.1 Les VA $X_1, \cdots X_n$ definides en un mateix espai de probabilitat $(\Omega, \mathcal{F}, \mathrm{P})$ són independents quan ho són les σ-àlgebra $X_i^{-1}(\mathbb{B}), i = 1, \cdots n$. És a dir, per qualsevol $B_i \in \mathbb{B}, i = 1, \cdots n$:

$$\mathrm{P}(X_1 \in B_1, \cdots X_n \in B_n) = \mathrm{P}(X_1 \in B_1) \cdots \mathrm{P}(X_n \in B_n).$$

Això implica que:

$$\mathrm{P}(X_1 \le x_1, \cdots X_n \le x_n) = F_1(x_1) \cdots F_2(x_n)$$

on F_i són les respectives funcions de distribució.

Proposició 2.3 Siguin X_1 i X_2 VA independents amb funcions de distribució F_1 i F_2 mixtes. F_1 amb densitat i salts $p_1(x)$ en un conjunt numerable D. Si h és una funció contínua en $I = [a,b]$, aleshores:

$$\mathrm{P}(X_1 \in I, X_2 \le h(X_1)) = \int_I F_2(h(x))\,F_1(dx),$$

on es defineix:

$$\int_I F_2(h(x))\,F_1(dx) = \sum_{x \in I \cap D} F_2(h(x))\,p_1(x) + \int_a^b F_2(h(x))\,f_1(x)\,dx. \quad (2.7)$$

Suma de VA independents Amb les mateixes condicions de la proposició 2.3, la VA $Z = X_1 + X_2$ té distribució:

$$\begin{aligned}\mathrm{P}(Z \le z) &= \mathrm{P}(X_1 + X_2 \le z) \\ &= \int_{-\infty}^{\infty} F_2(z-x)\,F_1(dx) \\ &= \int_{-\infty}^{\infty} F_1(z-x)\,F_2(dx) \\ &= F_1 * F_2(z)\end{aligned}$$

on $F_1 * F_2(z)$ s'anomena **convolució** de les funcions F_1, F_2.

Propietats

1. $\mathrm{E}[a\,X + b\,Y] = a\,\mathrm{E}[X] + b\,\mathrm{E}[Y]$.
2. Si $X \ge Y$ aleshores $\mathrm{E}[X] \ge \mathrm{E}[Y]$.
3. $|\mathrm{E}[X]| \le \mathrm{E}[|X|]$.
4. Si X,Y són independents: $\mathrm{E}[X\,Y] = \mathrm{E}[X]\,\mathrm{E}[Y]$.
5. Per qualsevol VA $X \ge 0$

$$\sum_{n=1}^{\infty} P(X \ge n) \le \mathrm{E}[X] \le 1 + \sum_{n=1}^{\infty} P(X \ge n).$$

En particular, si $X \ge 0$ és discreta:

$$\mathrm{E}[X] = \sum_{n=1}^{\infty} P(X \ge n) = \sum_{n=0}^{\infty} P(X > n)$$

i en general, si $X \ge 0$:

$$\mathrm{E}[X] = \int_0^{\infty} (1 - F(X))\,dx$$

6. Desigualtat de Markov. Si $g : [0,\infty) \to [0,\infty)$ és una funció creixent, per qualsevol $c > 0$:

$$P(|X| \ge c) \le \frac{\mathrm{E}[g(|X|)]}{g(c)}.$$

En particular:

$$P(|X - a| \ge c) \le \frac{\mathrm{E}[|X - a|^r]}{c^r}.$$

i la desigualtat de Chebyshev:

$$P(|X - \mu| \ge c) \le \frac{\sigma^2}{c^2}.$$

7. Per $r > 0$, $\mathrm{E}[|X - a|^r] = 0$ si i només si $P(X = a) = 1$.

Esperança condicionada

$$\mathrm{E}[X|A] = \int_{-\infty}^{\infty} x\,F(dx|A).$$

En particular, si I_A és l'indicador de A (veure (2.1)):

$$\mathrm{E}[X|A] = \frac{\mathrm{E}[X\,I_A]}{P(A)}$$

Si A_n és una partició, aleshores $X = \sum_n X\,I_{A_n}$, d'on

$$\mathrm{E}[X] = \sum_n P(A_n)\,\mathrm{E}[X|A_n]$$

Corba de regressió de X sobre Y Si $\boldsymbol{Y}$ és una VA k-dimensional $k \ge 1$, i X és unidimensional, és la corba

$$\mu_X(\boldsymbol{y}) = \mathrm{E}[X \mid \boldsymbol{Y} = \boldsymbol{y}] = \int_{\mathbb{R}} x\,F(dx|\boldsymbol{y})$$

i indica la mesura en que la VA X depèn dels valors observats de $\boldsymbol{Y}$.

En particular, per a la VA $(X,\boldsymbol{Y})$ amb densitat $f(x,\boldsymbol{y})$ i marginals $f_1(x)$, $f_2(\boldsymbol{y})$, es té:

$$\mathrm{E}[X \mid \boldsymbol{Y} = \boldsymbol{y}] = \int_{\mathbb{R}} x\,f(x|\boldsymbol{y})\,dx$$

$$\mathrm{E}[X] = \int_{\mathbb{R}^k} \mathrm{E}[X \mid \boldsymbol{Y} = \boldsymbol{y}]\,f_2(\boldsymbol{y})\,d\boldsymbol{y}$$

on la última equació equival a: $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|\boldsymbol{Y}]]$

Capítol 6
Anàlisi descriptiu

6.1 En ℝ

Moment respecte l'origen d'ordre $r > 0$

$$\alpha_r = \mathrm{E}[X^r] = \int_{\mathbb{R}} x^r\,F(dx)$$

d'on destaca la **mitjana**:

$$\mu = \alpha_1 = \mathrm{E}[X]$$

Moments centrals d'ordre $r > 0$

$$\mu_r = \mathrm{E}[(X - \mu)^r] = \int_{\mathbb{R}} (x - \mu)^r\,F(dx).$$

d'on destaca la **variança**:

$$\sigma^2 = \sigma^2(X) = \mu_2 = \mathrm{E}[(X - \mu)^2]$$

i la **desviació típica**:

$$\sigma = \sigma(X) = \sqrt{\sigma^2} = \sqrt{\mathrm{E}[(X - \mu)^2]}$$

De $(X - \mu)^r = \sum_{i=0}^{r} \binom{i}{r} (-\mu)^i\,X^{r-i}$ resulta

$$\mu_r = \sum_{i=0}^{r-1} \binom{i}{r} (-\mu)^i\,\alpha_{r-i}$$

doncs $\mathrm{E}[X^0] = 0$. D'aquí:

$$\sigma^2 = \alpha_2 - \mu^2 = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \mathrm{E}[X^2] - \mu^2.$$

Coeficient de variació

$$\gamma_2 = \frac{\sigma}{|\mu|}$$

on B és el conjunt:

$$a \le x_1 \le b, h_1(x_1) \le x_2 \le h_2(x_1), \cdots$$
$$g_1(x_1, \cdots x_{k-1}) \le x_k \le g_2(x_1, \cdots x_{k-1})$$

Distribucions singulars És quan la distribució k-dimensional està concentrada en un conjunt no numerable $S \in \mathbb{B}^k$ amb volum k-dimensional nul, i amb $P(\{\boldsymbol{x}\}) = 0, \forall \boldsymbol{x} \in S$. Sovint es poden tractar com distribucions absolutament continues de dimensió inferior a k.

Canvis de variable Si $\boldsymbol{X}$ és una VA k-dimensional absolutament continua i $\boldsymbol{Y} = g(\boldsymbol{X})$, on $g : \mathbb{R} \to \mathbb{R}^k$ és injectiva i diferenciable amb Jacobià J_g, i inversa h amb Jacobià J_h:

$$\begin{aligned} f_Y(\boldsymbol{y}) &= f_X(g^{-1}(\boldsymbol{y}))\,|J_g(g^{-1}(\boldsymbol{y}))|^{-1} \\ &= f_X(h(\boldsymbol{y}))\,|J_h(\boldsymbol{y})| \end{aligned}$$

Distribucions marginals Una distribució k-dimensional té $2^k - 2$ distribucions marginals que s'obtenen quan $i, 0 < i < k$ de les variables tendeixen a ∞. La distribució marginal resultant és $k - i$ dimensional.

Distribució condicional Si $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, on $\boldsymbol{X} \in \mathbb{R}^k, \boldsymbol{X}_1 \in \mathbb{R}^r, \boldsymbol{X}_2 \in \mathbb{R}^{k-r}$, aleshores la distribució r-dimensional de $\boldsymbol{X}_1$ condicionada per $\boldsymbol{X}_2 = \boldsymbol{x}_2, P(\{\boldsymbol{X}_2 = \boldsymbol{x}_2\}) > 0$ és:

$$P(\boldsymbol{X}_1 \le \boldsymbol{x}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2) = \frac{P(\boldsymbol{X}_1 \le \boldsymbol{x}_1, \boldsymbol{X}_2 = \boldsymbol{x}_2)}{P(\boldsymbol{X}_2 = \boldsymbol{x}_2)}.$$

Si $\boldsymbol{X}$ és discreta amb probabilitat $p(\boldsymbol{x}_1, \boldsymbol{x}_2)$ o absolutament continua amb densitat $f(\boldsymbol{x}_1, \boldsymbol{x}_2)$, aleshores, respectivament:

$$p(\boldsymbol{x}_1 | \boldsymbol{x}_2) = \frac{p(\boldsymbol{x}_1, \boldsymbol{x}_2)}{p(\boldsymbol{x}_2)} \tag{4.5}$$
$$f(\boldsymbol{x}_1 | \boldsymbol{x}_2) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f(\boldsymbol{x}_2)} \tag{4.6}$$

Composició de marginals i condicionades La distribució de $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, i la marginal de $\boldsymbol{X}_2$ es poden obtenir, respectivament, com:

$$F(\boldsymbol{x}) = \int_{-\infty}^{\boldsymbol{x}_1} F(\boldsymbol{x}_2 | \boldsymbol{s})\, F_1(d\boldsymbol{s})$$
$$F_2(\boldsymbol{x}_2) = \int_{-\infty}^{\infty} F(\boldsymbol{x}_2 | \boldsymbol{s})\, F_1(d\boldsymbol{s})$$

on F_1 és la distribució marginal de $\boldsymbol{X}_1$.

Si $\boldsymbol{X}$ és discreta o absolutament continua, respectivament:

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2) = p(\boldsymbol{x}_1 | \boldsymbol{x}_2)\, p(\boldsymbol{x}_2) \tag{4.7}$$
$$f(\boldsymbol{x}_1, \boldsymbol{x}_2) = f(\boldsymbol{x}_1 | \boldsymbol{x}_2)\, f(\boldsymbol{x}_2) \tag{4.8}$$

Independència Les components d'una distribució k-dimensional poden ser totalment independents o entre grups. En aquest cas:

$$F(\boldsymbol{x}) = F_1(\boldsymbol{x}_1)\, F_2(\boldsymbol{x}_2) \cdots F_n(\boldsymbol{x}_n)$$
$$f(\boldsymbol{x}) = f_1(\boldsymbol{x}_1)\, f_2(\boldsymbol{x}_2) \cdots f_n(\boldsymbol{x}_n)$$

Capítol 5
Esperança

Definició d'esperança, o mitjana, (*expectation*):

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

suposat que la serie no divergeix per $\{\omega | X(\omega) > 0\}$ ni $\{\omega | X(\omega) < 0\}$. Si divergeixen ambdues, $\mathrm{E}[X]$ no existeix, si divergeix només per $\{\omega | X(\omega) > 0\}$, aleshores $\mathrm{E}[X] = \infty$, altrament $\mathrm{E}[X] = -\infty$. Si X es discreta o absolutament continua, respectivament:

$$\mathrm{E}[X] = \sum_k x_k\, P(X = x_k)$$
$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

En distribucions mixtes:

$$\mathrm{E}[X] = \sum_k x_k\, P(X = x_k) + \int_{-\infty}^{\infty} x\, f(x)\, dx$$

que amb la notació (2.7) és:

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, F(dx). \tag{5.1}$$

Canvi de variable Sigui $Y = g(X)$ on $g : \mathbb{R} \to \mathbb{R}$, aleshores:

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} g(x)\, F(dx) = \mathrm{E}[g(X)].$$

Igualment, si $\boldsymbol{X}$ és una VA k-dimensional i $Y = g(\boldsymbol{X})$ on $g : \mathbb{R}^k \to \mathbb{R}$, aleshores:

$$\mathrm{E}[Y] = \int_{\mathbb{R}^k} g(\boldsymbol{x})\, F(d\boldsymbol{x}).$$

Proposició 2.4 Si la VA $Z = X_1 + X_2$ té distribució absolutament continua, aleshores la seva densitat és:

$$\begin{aligned} f_Z(z) = f_1 * f_2(z) &= \int_{-\infty}^{\infty} f_2(z - x)\, f_1(x)\, dx \\ &= \int_{-\infty}^{\infty} f_1(z - x)\, f_2(x)\, dx \end{aligned}$$

Capítol 3
Probabilitats en $\mathbb{R}^2$

3.1 Variables aleatòries bidimensionals

Definició 3.1 Una VA bidimensional en un espai de probabilitat $(\Omega, \mathcal{F}, \mathrm{P})$ és qualsevol funció $\boldsymbol{X} = (X, Y) : \Omega \to \mathbb{R}^2$ on X, Y són VA definides en $(\Omega, \mathcal{F}, \mathrm{P})$. Això implica que la contra-imatge de qualsevol $B \in \mathbb{B}^2$ verifica $\boldsymbol{X}^{-1}(B) \in \mathcal{F}$. $(\mathbb{R}^2, \mathbb{B}^2, \mathrm{P}_X)$, on $\mathrm{P}_x(B) = \mathrm{P}(\boldsymbol{X} \in B)$, s'anomena espai de probabilitat induït per $\boldsymbol{X}$. La funció de distribució és:

$$F_X(x,y) = \mathrm{P}(X \le x, Y \le y) = \mathrm{P}_x(Q_{x,y})$$

on es defineix el **quadrant** $Q_{x,y} = (-\infty, x] \times (-\infty, y]$.

Notar que el rectangle $(a_1, b_1] \times (a_2, b_2]$, $a_1 < b_1$, $a_2 < b_2$, té probabilitat:

$$\begin{aligned} &\mathrm{P}(\boldsymbol{X} \in (a_1, b_1] \times (a_2, b_2]) = \\ &\mathrm{P}_x(Q_{b_1,b_2}) - \mathrm{P}_x(Q_{a_1,b_2}) - \mathrm{P}_x(Q_{b_1,a_2}) + \mathrm{P}_x(Q_{a_1,a_2}) = \\ &F_X(b_1,b_2) - F_X(a_1,b_2) - F_X(b_1,a_2) + F_X(a_1,a_2) \end{aligned}$$

En general una funció de distribució bidimensional pot tenir 3 components (distribucions *mixtes*):

$$F = \alpha_1 F_D + \alpha_2 F_C + \alpha_3 F_S \tag{3.1}$$

on F_D és una component discreta, F_C és una component absolutament continua i F_S és una singular.

Distribucions discretes Tenen la probabilitat concentrada en un conjunt numerable de punts $D = \{(x,y) \in \mathbb{R}^2 | \mathrm{P}(x,y) > 0\}$

Distribucions absolutament continues Si F és una distribució absolutament contínua, existeix una *funció de densitat* f tal que:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\, \partial y} \tag{3.2}$$

a més, si $B = \{(x,y) \in \mathbb{R}^2 \mid a \le x \le b, h(x) \le y \le g(x)\}$, on $a < b$ i $h(t) \le g(t)$ per $t \in [a,b]$, aleshores (veure la figura 3.1 (a)):

$$\mathrm{P}(B) = \int_a^b \int_{h(x)}^{g(x)} f(x,y)\, dy\, dx \tag{3.3}$$

Anàlogament, si $B = \{(x,y) \in \mathbb{R}^2 \mid a \le y \le b, h(y) \le x \le g(y)\}$, on $a < b$ i $h(t) \le g(t)$ per $t \in [a,b]$, aleshores (veure la figura 3.1 (b)):

$$\mathrm{P}(B) = \int_a^b \int_{h(y)}^{g(y)} f(x,y)\, dx\, dy \tag{3.4}$$

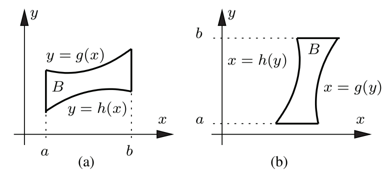

Figura 3.1: Regió d'integració.

Distribucions singulars Anàlogament al cas unidimensional, una distribució singular bidimensional està concentrada en un conjunt $S \in \mathbb{B}^2$ no numerable i d'àrea nul·la tal que $\mathrm{P}(x,y) = 0$, $\forall (x,y) \in S$, però $\mathrm{P}(S) > 0$. A diferència del cas unidimensional, en el que un conjunt $L \in \mathbb{B}$ no numerable i de longitud nul·la tal que $\mathrm{P}(x) = 0$, $\forall x \in L$, però $\mathrm{P}(L) > 0$ és tan difícil de trobar com el conjunt de Cantor, en el cas bidimensional és tan comú com una corba en $\mathbb{R}^2$.

Per especificar la probabilitat en una distribució singular en $\mathbb{R}^2$ concentrada en una corba, basta especificar la corba en forma paramètrica: $S = \{(x,y) \in \mathbb{R}^2 \mid x = h(t),\ y = g(t),\ t \in [a,b]\}$. Assignant una distribució (unidimensional) al paràmetre t, queda especificada la distribució singular en $\mathbb{R}^2$.

Nota respecte les discontinuïtats: Només es produeixen en cas d'haver-hi un segment paral·lel a un eix amb probabilitat no nul·la. Això provocaria una discontinuïtat en F al incloure el segment. Notar que si el segment no és paral·lel a un eix, al incloure una part en un quadrant, la probabilitat inclosa seria proporcional a la longitud que intercepte el quadrant, i no hi hauria discontinuïtat.

3.2 Canvis de variable

Sigui $g : \mathbb{R}^2 \to \mathbb{R}^2$ una funció mesurable. La distribució de la VA bidimensional $(U,V) = g(X,Y)$ es pot determinar com:

$$P(U \le u, V \le v) = P_X(g(X,Y) \in Q_{u,v}) = P_X((X,Y) \in g^{-1}(Q_{u,v})) \quad (3.5)$$

on $Q_{u,v} = (-\infty,u] \times (-\infty,v]$. En la pràctica, aplicar l'equació (3.5) pot tenir la dificultat de determinar el conjunt $g^{-1}(Q_{u,v})$, o calcular la probabilitat que P_X li assigna.

Exemple 3.1 Canvi de variable directe
Suposem que la VA bidimensional (X,Y) té funció de densitat $f(x,y)$ absolutament continua en $x \ge 0, y \ge 0$. Desitgem calcular la distribució de la VA $(U,V) = g(X,Y)$ on:

$$U = X + Y, U \ge 0$$
$$V = \frac{X}{X+Y}, V \in [0,1]$$

Tenim que la regió $U \le u$ ve donada per $x+y \le u$, és a dir, $y \le u-x$. La regió $V \le u$ ve donada per $\frac{X}{X+Y} \le v$, és a dir: $y \ge \frac{1-v}{v}\,x$. Per tant, la funció de distribució de (U,V) vindrà donada per la integració de $f(x,y)$ en el triangle $B = \{(x,y) \mid x \ge 0, y \le u-x, y \ge \frac{1-v}{v}\,x\}$ (veure la figura 3.2).

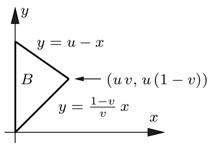

Figura 3.2: Canvi de variable.

Aplicant (3.4):

$$F(u,v) = P(B) = \int_0^{u\,v} \int_{\frac{1-v}{v}x}^{u-x} f(x,y)\,dy\,dx$$

Un cop calculat $F(u,v)$ podem calcular la densitat $f(u,v) = \frac{\partial^2 F(u,v)}{\partial u\,\partial v}$. ■

El càlcul anterior es pot simplificar significativament quan $f_X(x,y)$ és absolutament continua. Si $f_X > 0$ en una regió $\mathcal{R}$ i $g : \mathbb{R}^2 \to \mathbb{R}^2$ és injectiva[1] i diferenciable en $\mathcal{R}$. En aquest cas, $f(u,v)$ es pot calcular directament d'una de les següents maneres:

$$f(u,v) = f_X(g^{-1}(u,v))\,1/|J_g(g^{-1}(u,v))| \quad (3.6)$$
$$f(u,v) = f_X(h(u,v))\,|J_h(u,v)| \quad (3.7)$$

on $h = g^{-1}$, i $|J_g|$, $|J_h|$ són el valor absolut del Jacobià de les funcions g i $h = g^{-1}$, respectivament.

Les dues formes són equivalents doncs $J_h(u,v) = 1/J_g(g^{-1}(u,v))$.

En el cas de que g no sigui injectiva en $\mathcal{R}$, però $\mathcal{R}$ es pugui descompondre en varies subregions $\mathcal{R}_1$, $\mathcal{R}_2$, $\cdots$ on si ho sigui, aleshores les probabilitats en un punt $(u,v) \in g(\mathcal{R})$ acumulen les probabilitats dels punts $g^{-1}(u,v) \cap \mathcal{R}_i$.

Exemple 3.2 Canvi de Variable amb el Jacobià
Aplicant el mètode anterior a l'exemple 3.1 tenim:

$$g = \begin{cases} u = x + y, & u \ge 0 \\ v = \frac{x}{x+y}, & v \in [0,1] \end{cases}$$
$$h = \begin{cases} x = u\,v, & x \ge 0 \\ y = u\,(u-v), & y \ge 0 \end{cases}$$

Aplicant (3.7):

$$J_h = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -u$$

d'on:

$$f(u,v) = f_X(u\,v, u\,(u-v))\,u,\; u \ge 0, v \in [0,1]$$ ■

3.3 Distribucions marginals

La distribució bidimensional té la informació completa sobre les dues VA que la composen. Les distribucions unidimensionals de cadascuna d'elles s'anomenen *distribucions marginals*. En concret:

$$F_X(x) = P(X \le x) = \lim_{y\to\infty} P(X \le x, Y \le y) = \lim_{y\to\infty} F(x,y) = F(x,\infty)$$

anàlogament, $F_Y(y) = F(\infty,y)$. Per a les densitats de VA absolutament contínues:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad (3.8)$$

i anàlogament per f_Y.

[1] No mapeja diferents elements del domini en un mateix element del codomini.

3.4 Distribucions condicionades

Per a una distribució bidimensional absolutament contínua es compleix:

$$f(x,y) = f_X(x)\,f(y|x)$$
$$f(x,y) = f_Y(y)\,f(x|y)$$

Substituint en (3.8):

$$f_X(x) = \int_{-\infty}^{\infty} f_Y(y)\,f(x|y)\,dy \quad (3.9)$$

i anàlogament per f_Y, d'on tenim:

$$f(y|x) = \frac{f_Y(y)\,f(x|y)}{\int_{-\infty}^{\infty} f_Y(y)\,f(x|y)\,dy} \quad (3.10)$$

i anàlogament per $f(x|y)$, que no és més que la versió de la fórmula de Bayes per a VA absolutament contínues.

Proposició 3.1 Sigui F_X una funció de distribució mixta amb funció de probabilitat p_X concentrada en un conjunt numerable D i densitat f_X. Si $F(y|x)$ és una funció de distribució continua, aleshores la distribució bidimensional de (X,Y) és:

$$F(x,y) = \sum_{s\in D, s\le x} p_X(s)\,F(y|s) + \int_{-\infty}^{x} f_X(s)\,F(y|s)\,ds$$

que amb la notació (2.7) és:

$$F(x,y) = \int_{-\infty}^{x} F(y|s)\,F_X(ds). \quad (3.11)$$

3.5 Independència

Si dues VA X,Y són independents:

$$F(x,y) = F_X(x)\,F_Y(y)$$
$$F(x|y) = F_X(x)$$
$$F(y|x) = F_Y(y)$$
$$f(x,y) = f_X(x)\,f_Y(y)$$

A més, si es pot expressar: $f(x,y) = f_1(x)\,f_2(y)$, aleshores també són independents.

Capítol 4
Probabilitat en $\mathbb{R}^k$

funció de distribució k-dimensional Cada quadrant

$$C_{\boldsymbol{x}} = (-\infty,x_1] \times \cdots (-\infty,x_k] \quad (4.1)$$

associat al punt $\boldsymbol{x}$ té la probabilitat:

$$F(\boldsymbol{x}) = F(x_1,\cdots x_k) = P(C_{\boldsymbol{x}}) \quad (4.2)$$

on $F(\boldsymbol{x})$ s'anomena la **funció de distribució k-dimensional** associada a P.

La probabilitat d'un rectangle:

$$I = (x_1,x_1+h_1] \times \cdots (x_k,x_k+h_k]$$

ve donada per:

$$P(I) = \sum_{r=1}^{2^k} (-1)^{\gamma_r} F(\boldsymbol{x} + \boldsymbol{h}_r) \quad (4.3)$$

on $\boldsymbol{h}_r$ recorre els 2^k vectors obtinguts al substituir les coordenades de $\boldsymbol{h} = (h_1,\cdots h_k)$ per zeros de totes les formes possibles, i γ_r és el nombre de zeros de $\boldsymbol{h}_r$.

Distribució discreta Una distribució és discreta si està concentrada en un conjunt $D \subset \mathbb{R}^k$ finit o numerable. La funció:

$$p(\boldsymbol{x}) = p(\{\boldsymbol{x}\}) \quad (4.4)$$

s'anomena funció de probabilitat associada a P, i per $B \in \mathbb{B}^k$, $\boldsymbol{x} \in \mathbb{R}$ es té:

$$P(B) = \sum_{\boldsymbol{x}\in B\cap D} p(\boldsymbol{x})$$
$$F(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in C_{\boldsymbol{x}}\cap D} p(\boldsymbol{y})$$

Exemple 4.1 Distribució multinomial
En una urna hi ha bolles de k colors en proporcions $p_1,\cdots p_k$, $p_1 + \cdots p_k = 1$. La probabilitat d'extreure amb reemplaçament n bolles amb colors $n_1,\cdots n_k$, $n_1 + \cdots n_k = n$ és:

$$p(n_1,\cdots n_k) = \frac{n!}{n_1!\cdots n_k!}\,p_1^{n_1}\cdots p_k^{n_k}$$ ■

Funció de densitat Una distribució és absolutament contínua si existeix una **funció de densitat** (*probability density function, PDF*), $f : \mathbb{R}^k \to \mathbb{R}$, $f(\boldsymbol{x}) \ge 0$ i $\int_{\mathbb{R}^k} f(\boldsymbol{x})\,d\boldsymbol{x} = 1$, tal que per a qualsevol rectangle $I \subset \mathbb{R}^k$:

$$P(I) = \int_I f(\boldsymbol{x})\,d\boldsymbol{x}$$

d'on:

$$F(\boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(\boldsymbol{x})\,dx_k \cdots dx_1$$
$$f(\boldsymbol{x}) = \frac{\partial^k F(\boldsymbol{x})}{\partial x_1 \cdots \partial x_k}$$
$$P(B) = \int_a^b \int_{h_1(x_1)}^{h_2(x_1)} \cdots \int_{g_1(x_1,\cdots x_{k-1})}^{g_2(x_1,\cdots x_{k-1})} f(\boldsymbol{x})\,dx_k \cdots dx_1$$