Algoritmos aleatorizados: Repaso de Probabilidad y Variables Aleatorias

J. M. Almira

Universidad de Murcia

Algoritmos aleatorizados: qué son

Se trata de algoritmos en los que se introduce una cierta componente de azar en alguno de sus pasos. Por ejemplo, en alguna de las instrucciones se lanza una moneda (con cierta probabilidad $0 < p < 1$ de cara) de cuyo resultado depende lo que se va a realizar en la siguiente instrucción del algoritmo.

Probabilidad: un breve repaso

Concepto de Probabilidad. Probabilidad condicionada. Independencia. Teorema de Bayes. Teorema de la Probabilidad total

Concepto de variable aleatoria. Variables aleatorias continuas y discretas. Distribuciones de probabilidad.Ejemplos importantes.

Esperanza, Varianza, covarianza, momentos. Desigualdad de Tchebychev. Ley de los grandes números.

Variables aleatorias independientes. Teorema central del límite. Función generatriz de momentos. Propiedades reproductivas de v.a.

Vectores aleatorios. Distribución conjunta. Distribuciones marginales y condicionadas.

Concepto de Probabilidad -Según Kolmogorov

Álgebra de sucesos

Supongamos que realizamos un experimento aleatorio $\mathcal{E}$ cuyos sucesos elementales quedan descritos por el conjunto no vacío Ω (al que llamamos espacio muestral).
Decimos que $\Sigma \subseteq 2^{\Omega}$ es una σ-álgebra de sucesos si:

- $\Omega \in \Sigma$.
- Si $A \in \Sigma$, entonces $A^c = \Omega \setminus A \in \Sigma$.
- Si $\{A_i\}_{i\in\mathbb{N}} \subseteq \Sigma$, entonces $\bigcup_{i\in\mathbb{N}} A_i \in \Sigma$.

Ejemplo

2^{Ω}, es siempre una σ-álgebra.

Concepto de Probabilidad -Según Kolmogorov

Ejemplo importante

Dada una familia $D \subseteq 2^{\Omega}$, denotamos por $\sigma(D)$ al menor σ-álgebra $\Sigma \subseteq 2^{\Omega}$ que contiene a todos los elementos de D.
Si X es un espacio topológico (por ejemplo, $X = \mathbb{R}^d$), llamamos σ-álgebra de Borel de X a

$$\sigma(\mathcal{T}),$$

donde $\mathcal{T}$ denota el conjunto de los abiertos de X.

Concepto de Probabilidad -Según Kolmogorov

Proposición

Si Σ es una σ-álgebra de sucesos y $\{A_i\}_{i\in\mathbb{N}} \subset \Sigma$, entonces $\bigcap_{i\in\mathbb{N}} A_i \in \Sigma$.

- $\bigcup A_i$ representa el suceso: "sucede alguno de los sucesos A_i"
- $\bigcap A_i$ representa el suceso: "suceden todos los sucesos A_i"

¿Prueba? Idea: Usar las Leyes de Morgan

Concepto de Probabilidad -Según Kolmogorov

Dados $\mathcal{E}$, Ω y Σ como antes. Decimos que $P : \Sigma \to [0, 1]$ es una medida de probabilidad en Ω si

- $P(\emptyset) = 0$, $P(\Omega) = 1$.
- $P(A^c) = 1 - P(A)$
- Si $\{A_i\}_{i=1}^{\infty}$ es una familia numerable de elementos de Σ disjuntos dos a dos, entonces $P(\bigcup_i A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Espacio de Probabilidad

Llamamos espacio de probabilidad a todo triplete (Ω, Σ, P), donde $\Omega \neq \emptyset$, $\Sigma \subseteq 2^{\Omega}$ es una σ-álgebra y $P : \Sigma \to [0, 1]$ es una medida de probabilidad. Si $A \in \Sigma$ decimos que A es un suceso cuya probabilidad es $P(A)$.

Concepto de Probabilidad -Según Kolmogorov

Ejemplo importante nº 1: Espacio de Laplace

Si Ω es un conjunto finito, $\Sigma = 2^{\Omega}$ es una σ-álgebra y

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

define una medida de probabilidad (llamada probabilidad de Laplace).

- Se trata de un ejemplo "trivial" desde el punto de vista "teórico"
- Sin embargo, dado un experimento $\mathcal{E}$ cuyo espacio muestral asociado es finito, puede ser extremadamente complicado obtener una buena representación de Ω en la que sea posible contar. Este es, de hecho, el problema fundamental de la combinatoria.

Concepto de Probabilidad -Según Kolmogorov

Ejemplo importante nº 2: Probabilidades geométricas

Si $\Omega \subseteq \mathbb{R}^d$ es un conjunto medible con medida finita, podemos considerar $\Sigma = \{A \in 2^{\Omega} : |A| < \infty\}$ y

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

define una medida de probabilidad en Ω que generaliza muy bien a la probabilidad de Laplace del ejemplo anterior.

Concepto de Probabilidad -Según Kolmogorov

- Dados dos sucesos A, B, es claro que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

En efecto, $\{A \setminus (A \cap B), A \cap B\}$ es una partición de A en sucesos, por lo que

$$P(A) = P(A \setminus (A \cap B)) + P(A \cap B)$$

Análogamente, $P(B) = P(B \setminus (A \cap B)) + P(A \cap B)$ y $P(A \cup B) = P(A \setminus (A \cap B)) + P(B \setminus (A \cap B)) + P(A \cap B)$, de modo que:

$$\begin{aligned} &P(A \cup B) \\ &= P(A \setminus (A \cap B)) + P(A \cap B) - P(A \cap B) + P(B \setminus (A \cap B)) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

Concepto de Probabilidad -Según Kolmogorov

- Dado (Ω, Σ, P) un espacio de probabilidad y dado un suceso $B \in \Sigma$ tal que $P(B) > 0$, Podemos considerar en B la siguiente σ-álgebra:

$$\Sigma_{|B} = \{A \cap B : A \in \Sigma\} \subseteq 2^B,$$

con lo que $P_{|B} : \Sigma_{|B} \to [0, 1]$ definida por $P_{|B}(A \cap B) = \frac{P(A \cap B)}{P(B)}$ define una medida de probabilidad.

Probabilidad condicionada

Dados dos sucesos A, B, con $P(B) > 0$. La probabilidad de A, dado B, es:

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

Concepto de Probabilidad -Según Kolmogorov

Sucesos independientes

Dados dos sucesos A, B, se dice que son independientes si

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

.

Obsérvese que, en el caso de que $P(B) > 0$, la independencia de ambos sucesos se traduce en la identidad:

$$P(A|B) = P(A)$$

Concepto de Probabilidad -Según Kolmogorov

Familia de sucesos independientes

Dados los sucesos $\{A_1, A_2, \cdots, A_m\}$, se dice que son independientes si, para todo $k \leq m$ y todo vector $(i_1, i_2, \cdots, i_k) \in \{1, \cdots, m\}^k$ con $i_1 < i_2 < \cdots < i_k$, se tiene que

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \prod_{s=1}^{k} P(A_{i_s})$$

.

Además, la familia de sucesos $\{A_i\}_{i=1}^{\infty}$ se dice que es independiente si para todo $m \geq 1$, $\{A_1, A_2, \cdots, A_m\}$ forma una familia de sucesos independientes.

Concepto de Probabilidad -Según Kolmogorov

Teorema (de la Probabilidad Total)

Sea $\{B_1, B_2, \cdots, B_m\}$ *una familia de sucesos disjuntos dos a dos cuya unión es* Ω *(se dice que estos conjuntos forman una partición de* Ω *en sucesos). Sea* A *otro suceso. Entonces*

$$P(A) = \sum_{k=1}^{m} P(A \cap B_k) = \sum_{k=1}^{m} P(A|B_k)P(B_k).$$

(Asumimos, para la segunda igualdad, que $P(B_k) > 0$ *para todo* k*).*

De este modo, a menudo es posible conocer $P(A)$ a partir de su “dependencia” respecto de los sucesos $B_1, \cdots, B_m$.

Concepto de Probabilidad -Según Kolmogorov

Teorema (de Bayes)

Sean $\{A, B\}$ *sucesos con probabilidad positiva. Entonces*

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

(Asumimos, para la segunda igualdad, que $0 < P(B) < 1$*, por lo que también* $0 < P(B^c) < 1$*).*

En general, si la familia $\{B_i\}_{i=1}^{\infty}$ forma una partición de Ω en sucesos de probabilidad positiva, tenemos que:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{k=1}^{\infty} P(A|B_k)P(B_k)}, \text{ para } i = 1, 2, \cdots .$$

Variables aleatorias

Variable aleatoria

Una v.a. es una aplicación $X : \Omega \to \mathbb{R}$.

Asociados a una v.a. podemos describir numerosos sucesos en Ω, como $A = \{w \in \Omega : X(w) \text{ verifica la propiedad } P\}$, lo que conduce al siguiente concepto:

Independencia de v.a.

Las v.a. X, Y son independientes si para todo par de sucesos: A descrito en términos de X y B descrito en términos de Y, se tiene que $P(A \cap B) = P(A)P(B)$. Esta defección se extiende de forma natural a familias de v.a.

Variables aleatorias

Casos importantes

- (Variable discreta finita) $X(\Omega) = \{x_1, x_2, \cdots, x_N\}$. En tal caso nos interesa conocer:
$$p_k = P(X = x_k); 1 \leq k \leq N.$$
- (Variable discreta) $X(\Omega) = \{x_k\}_{k\in\mathbb{N}}$. En tal caso nos interesa conocer:
$$p_k = P(X = x_k); k \in \mathbb{N}$$
- (Variable continua) $X(\Omega)$ tiene interior no vacío (contiene un intervalo con interior no vacío). En tal caso nos interesa conocer:
$$F_X(t) = P(X \leq t); t \in \mathbb{R}$$

Variables aleatorias

- Obviamente, $\sum_k p_k = 1$ y $0 \leq p_k \leq 1$ para todo k.
  (De hecho, cualquier familia p_k con las propiedades anteriores determina la distribución de prob. de una v.a.)
- Para v.a. continuas a menudo podemos escribir:

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f_X(t)dt$$

  para cierta función $f_X : \mathbb{R} \rightarrow [0, \infty)$ (que hace el papel de los p_k's de las v.a. discretas y recibe el nombre de función densidad de probabilidad (fdp) de la v.a. X). Obviamente, $\int_{-\infty}^{\infty} f_X(t)dt = 1$ y conocer f_X determina la distribución de probabilidad de X.

Variables aleatorias

Las v.a. X, Y son independientes si:

- (Caso discreto) Si las v.a. son ambas discretas....:

$$P(X = x_k, Y = y_s) = P(X = x_k)P(Y = y_s)$$

- (Caso continuo) Si las v.a. son ambas continuas....:

$$F_{X,Y}(t,s) = P(X \leq t, Y \leq s) = P(X \leq t)P(Y \leq s) = F_X(t)F_Y(s)$$

Variables aleatorias

Variable aleatoria de Bernoulli

Sea $0 < p < 1$. Entonces $X \sim \mathbf{Be}(p)$ si $X(\Omega) = \{0, 1\}$, $P(X = 1) = p$, $P(X = 0) = q = 1 - p$.

Variable aleatoria Geométrica

Sea $0 < p < 1$. Entonces $X \sim \mathbf{Geom}(p)$ si $X(\Omega) = \mathbb{N}$,

$$P(X = k) = q^{k-1}p,$$

donde $q = 1 - p$.

Variables aleatorias

Variable aleatoria Binomial

Sea $0 < p < 1$ y sea $N \geq 1$. Entonces Entonces $X \sim \textbf{Binom}(N, p)$ si $X(\Omega) = \{0, 1, \cdots, N\}$ y

$$P(X = k) = \binom{N}{k} p^k q^{N-k}, \text{ para } 0 \leq k \leq N$$

donde $q = 1 - p$.

Proposición

Si $X_1, \cdots, X_N \sim \textbf{Be}(p)$ *son v.a. indep., entonces* $S_N = X_1 + \cdots + X_N \sim \textbf{Binom}(N, p)$.

Variables aleatorias

Variable aleatoria de Poisson

Sea $\lambda > 0$. Entonces Entonces $X \sim \mathbf{Poiss}(\lambda)$ si $X(\Omega) = \mathbb{N}$ y

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \text{ para } k \in \mathbb{N}$$

Variables aleatorias

Variable aleatoria uniforme

Entonces $X \sim \mathbf{U}(a,b)$ si $X(\Omega) = [a,b]$ y

$$F_X(t) = P(X \leq t) = \frac{1}{b-a}\int_a^t dx, \text{ para } a \leq t \leq b$$

Equivalentemente,

$$f_X(t) = \frac{1}{b-a}\chi_{[a,b]}(t).$$

Variables aleatorias

Variable aleatoria Normal

Sea $\mu \in \mathbb{R}$, $\sigma > 0$. Entonces Entonces $X \sim \mathbf{N}(\mu, \sigma)$ si $X(\Omega) = \mathbb{R}$ y

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

Equivalentemente,

$$f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(t-\mu)^2}{2\sigma^2}}$$

Variables aleatorias

Variable aleatoria exponencial

Sea $\lambda > 0$. Entonces Entonces $X \sim \mathbf{Exp}(\lambda)$ si $X(\Omega) = \mathbb{R}^+$ y

$$F_X(t) = P(X \leq t) = \int_0^t \lambda e^{-\lambda x} dx, \text{ para } t \geq 0$$

Equivalentemente,

$$f_X(t) = \chi_{[0,\infty)}(t) \lambda e^{-\lambda t}.$$

Variables aleatorias

Variable aleatoria $\chi^2(n)$

Sea $n \in \mathbb{N}$. Entonces Entonces $X \sim \chi^2(n)$ si $X(\Omega) = \mathbb{R}^+$ y

$$F_X(t) = P(X \leq t) = \int_0^t \frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx, \text{ para } t \geq 0$$

Equivalentemente,

$$f_X(t) = \chi_{[0,\infty)}(t) \frac{1}{2^{n/2}\Gamma(n/2)} t^{\frac{n}{2}-1} e^{-\frac{t}{2}}.$$

Nota. Recuérdese que $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ es una función que extiende de forma natural $n!$, pues $\Gamma(n+1) = n!$. De hecho, $\Gamma(1) = 1$ y $\Gamma(z+1) = z\Gamma(z)$.

Variables aleatorias

Variable aleatoria t-Student

Sea $n \in \mathbb{N}$. Entonces $X \sim t(n)$ si $X(\Omega) = \mathbb{R}$ y

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{n\pi}}(1 + \frac{x^2}{n})^{\frac{-(n+1)}{2}} dx, \text{ para } t \in \mathbb{R}$$

Equivalentemente,

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

Variables aleatorias: Propiedades

Esperanza matemática

- Si $X(\Omega) = \{x_k\}_{k\in\Delta}$, entonces $E(X) = \sum_k x_k P(X = x_k)$.
- Si $X(\Omega) = \mathbb{R}$, entonces $E(X) = \int_{\mathbb{R}} xF_X(t)dt$.

Proposición

Se verifican las siguientes propiedades:

- $E(X + Y) = E(X) + E(Y)$
- $E(\alpha X) = \alpha E(X)$
- *Si la v.a. es continua, y $h(x)$ es una función, entonces* $E(h(X)) = \int_{-\infty}^{\infty} h(x)f_X(x)dx.$
- *Si la v.a. X es discreta con $X(\Omega) = \{x_k\}_{k=0}^{\infty}$, entonces* $E(h(X)) = \sum_{k=0}^{\infty} h(x_k)P(X = x_k).$

Variables aleatorias: Propiedades

Varianza y covarianza

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

$$Cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

Proposición

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

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

Proposición

$V(\alpha X) = \alpha^2 V(X)$. *Además, si las v.a.* X, Y *son independientes, entonces* $V(X + Y) = V(X) + V(Y)$ *y* $Cov(X, Y) = 0$.

Variables aleatorias: Propiedades

Teorema (Desigualdad de Tchebychev)

Supongamos que $E(X) = \mu$, $V(X) = \sigma^2$ *son finitas. Entonces*

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

En particular, si $a = k\sigma$,

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

Variables aleatorias: Propiedades

Demostración. Tomemos

$$Y = \begin{cases} a^2 & \text{si } |X-\mu| \geq a \\ 0 & \text{resto de casos} \end{cases}$$

Entonces $Y \leq (X-\mu)^2$ y por tanto $E(Y) \leq E((X-\mu)^2) = \sigma^2$. Ahora bien,

$$E(Y) = a^2 P(Y = a^2) + 0 \cdot P(Y = 0) = a^2 P(|X-\mu| \geq a).$$

Por tanto,

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

Variables aleatorias: Propiedades

Teorema (Ley de los grandes números)

Sea $X_1, X_2, , \cdots$ *una sucesión de v.a. independientes idénticamente distribuidas con esperanza* μ *y varianza* σ^2 *finitas. Sea* $M_n = \frac{X_1+\cdots+X_n}{n}$. *Entonces, para cada* $\varepsilon > 0$*, se tiene que*

$$lim_{n\to\infty} P(|M_n - \mu| \geq \varepsilon) = 0$$

Variables aleatorias: Propiedades

Demostración. Es evidente que $E(M_n) = \mu$ y $V(M_n) = \frac{\sigma^2}{n}$ son ambas finitas. Aplicando la desigualdad de Tchebychev, tendremos que:

$$P(|M_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2} = \frac{1}{n}\frac{\sigma^2}{\varepsilon^2},$$

que tiende a 0 cuando $n \to \infty$.

Teorema central del límite

Teorema

Sea $X_1, X_2, \cdots$, una sucesión de v.a. independientes idénticamente distribuidas, todas ellas con $E(X_i) = \mu < \infty$ y $V(X_i) = \sigma^2 < \infty$. Consideremos la v.a.

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Entonces $Z = \frac{\sqrt{n}(S_n - \mu)}{\sigma}$ converge, en distribución, a una v.a. $N(0,1)$.

En otras palabras,

$$lim_{n\to\infty} F_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx,$$

con convergencia puntual en toda la recta.

Función generatriz de momentos

Función generatriz de momentos de una v.a.

$$M_X(t) = E(e^{tX})$$

Proposición

En condiciones generales,

$$M_X(t) = \sum_{k=0}^{\infty} \frac{E(X^n)}{n!} t^n,$$

de modo que $E(X^n) = M_X^{(n)}(0)$, $n = 0, 1, \cdots$.

Función generatriz de momentos

Teorema

Si $M_X(t) = M_Y(t)$ *para todo* t*, entonces las v.a.* X *e* Y *están idénticamente distribuidas.*

Función generatriz de momentos

Teorema

Si $Y = \alpha X + \beta$, *entonces* $M_Y(t) = e^{\beta t} M_X(\alpha t)$.

Teorema

Sean X, Y *v.a. independientes. Entonces* $M_{X+Y}(t) = M_X(t) M_Y(t)$.

Función generatriz de momentos

Teorema

Sean $X_1, \cdots, X_n$ *v.a. independientes y sea* $Z = X_1 + \cdots + X_n$*. Entonces:*

- *Si* $X_i \sim N(\mu_i, \sigma_i^2)$*,* $i = 1, \cdots, n$*, entonces* $Z \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$
- *Si* $X_i \sim \textbf{Poiss}(\alpha_i)$*,* $i = 1, \cdots, n$*, entonces* $Z \sim \textbf{Poiss}(\sum_{i=1}^n \alpha_i)$
- *Si* $X_i \sim \chi^2(n_i)$*,* $i = 1, \cdots, n$*, entonces*

$$Z \sim \chi^2(\sum_{i=1}^n n_i)$$

- *Si* $X_i \sim N(0,1)$*,* $i = 1, \cdots, n$*, entonces*

$$X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n)$$

Función generatriz de momentos

Teorema

Sean Z, X v.a. independientes. Supongamos además que $Z \sim N(0,1)$ y $X \sim \chi^2(n)$. Entonces

$$T = \frac{Z}{\sqrt{X/n}} \sim t(n).$$

(Este resultado se puede probar con un cálculo directo, sin necesidad de utilizar $M_T(t)$ ...)

Distribución conjunta

$$F_{X_1,\cdots,X_n}(t_1.\cdots,t_n) = \int_{-\infty}^{t_1}\cdots\int_{-\infty}^{t_n} f_{X_1,X_2,\cdots,X_n}(x_1,\cdots,x_n)$$

Nota:

- Si las v.a. son independientes, entonces

$$f_{X_1,X_2,\cdots,X_n} = f_{X_1}f_{X_2}\cdots f_{X_n}$$

Vectores aleatorios

Teorema (Distribución de una suma de v.a. independientes)

Supongamos que X, Y *son v.a. continuas independientes y* $Z = X + Y$. *Entonces*

$$f_Z(t) = (f_X * f_Y)(t) = \int_{-\infty}^{\infty} f_X(s) f_Y(t-s)\,ds.$$

Vectores aleatorios

Demostración: Recordemos el teorema del cambio de variables:

Teorema

Sea $T : W \to \Omega$ un cambio de variables entre abiertos de $\mathbb{R}^d$. Entonces

$$\int_{\Omega} f(x) = \int_{W} f(T(z)) |\det(T'(z))|$$

Vectores aleatorios

Ahora bien, por hipótesis, $f_{X,Y} = f_X f_Y$, de modo que

$$F_{X+Y}(t) = P(X+Y \leq t) = \int_{\{(x,y):x+y\leq t\}} f_X(x) f_Y(y)$$

Si tomamos $T(z,w) = (w, z-w)$, tendremos que $T(\{(z,w) : z \leq t\}) = \{(x,y) : x+y \leq t\}$. Además, $T' = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}$, por lo que $|\det(T')| = 1$. Se sigue que

$$\begin{aligned} \int_{\{(x,y):x+y\leq t\}} f_X(x) f_Y(y) &= \int_{\{(z,w):z\leq t\}} f_X(T(z,w)) f_Y(T(z,w)) \\ &= \int_{\{(z,w):z\leq t\}} f_X(w) f_Y(z-w) \\ &= \int_{-\infty}^{t} \left(\int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\,dw \right) dz \end{aligned}$$

Distribuciones marginales (caso continuo) (el caso discreto es similar)

Sea (X, Y) vector aleatorio continuo con fdp $f_{X,Y}(t,s)$. Entonces podemos definir las distribuciones de probabilidad marginales:

- $P(X \leq t, Y < \infty) = \int_{-\infty}^{t} \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \right) dx$, que nos conduce a la función densidad de probabilidad marginal

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

- Análogamente, $P(X < \infty, Y \leq s) = \int_{-\infty}^{s} \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \right) dy$, que nos conduce a la función densidad de probabilidad marginal

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

Vectores aleatorios

Distribuciones condicionadas (caso continuo) (el caso discreto es similar)

Sea (X,Y) vector aleatorio continuo con fdp $f_{X,Y}(t,s)$ y con fdp's marginales f_X y f_Y. Entonces podemos definir las distribuciones de probabilidad condicionadas:

- $P(X\leq t|Y=y)=\int_{-\infty}^{t}\frac{f_{X,Y}(x,y)}{f_Y(y)}dx$, que nos conduce a la función densidad de probabilidad condicionada

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

- Análogamente, $P(Y\leq s|X=x)=\int_{-\infty}^{s}\frac{f_{X,Y}(x,y)}{f_X(x)}dy$, que nos conduce a la función densidad de probabilidad condicionada

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