Vectores Aleatorios

Cecilia Galarza

Procesos Estocásticos
Facultad de Ingeniería, Universidad de Buenos Aires

1er cuatrimestre 2025

Vector aleatorio

- Un *espacio de probabilidad* queda definido por $(\Omega, \mathcal{F}, \mathbb{P})$
  - Ω: *espacio muestral* que contiene a todos los posibles resultados del experimento
  - $\mathcal{F}$: conjunto de eventos donde cada evento es un subconjunto de Ω
  - $\mathbb{P}$: función de probabilidad.
- En algunos experimentos aleatorios, nos interesa analizar más de una característica del resultado. Por ejemplo
  - Experimento: seleccionar una persona en la calle.
  - Observación: peso, altura, ancho de cintura.
- El resultado del experimento está caracterizado por 3 VA distintas. Resulta conveniente agruparlas en un *Vector Aleatorio* (VeA).

Vector aleatorio

Formalmente, un *vector aleatorio* de dimensión n es una función (medible)

$$\mathbf{X} : \Omega \to \mathbb{R}^n$$
$$\mathbf{X}(\xi) = [X_1(\xi), \ldots, X_n(\xi)]^T.$$

X es un vector aleatorio de dimensión n si X_i, $i = 1, \cdots n$, son todas variables aleatorias.

Función de distribución

Función de distribución multivariable

Cómo caracterizamos un vector de VA o un Vector aleatorio $\mathbf{X} = [X_1, \ ..., \ X_n]^t$? Utilizando la *función de distribución cumulativa*[1].

Sea $\boldsymbol{x} = [x_1, \ ..., \ x_n]^t \in \mathbb{R}^n$ el vector realización de $\mathbf{X}$. Luego,

$$F_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}$$

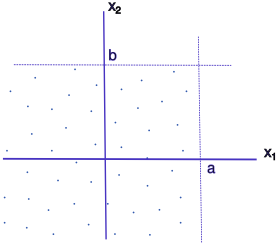
$$F_{\mathbf{X}}(\boldsymbol{x}) := F_{X_1, X_2, ... X_n}(x_1, x_2, ..., x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n).$$

Cuando $n = 1$, recuperamos la definición de la CDF de una VA.

[1] CDF, *Cumulative Distribution Function*

Función de distribución multivariable

$$F_{X_1,X_2}(x_1, x_2) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)$$



$$F_{\mathbf{X}}(\boldsymbol{x}) = \mathbb{P}(\mathbf{X} \leq \boldsymbol{x}).$$

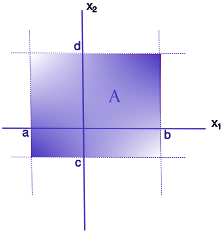
$\mathbf{X} \leq \boldsymbol{x}$ es una desigualdad simultánea en todas las componentes de $\mathbf{X}$.

Propiedades de la función de distribución

- *Monótona en cada uno de sus argumentos.* Si $x_i \leq x_i'$, $i = 1, ..., n$

$$F_{X_1,...,X_n}(x_1, ..., x_n) \leq F_{X_1,...,X_n}(x_1', ..., x_n').$$

- *Probabilidad de rectángulos*



Cuál es la probabilidad de observar **X** en A ?

$$A = \{(x_1, x_2) \in \mathbb{R}^2 : a < x_1 \leq b, c < x_2 \leq d\}$$

$$\mathbb{P}(a < X_1 \leq b, c < X_2 \leq d) = F_{X_1,X_2}(b, d) - F_{X_1,X_2}(a, d) - F_{X_1,X_2}(b, c) + F_{X_1,X_2}(a, c).$$

Propiedades de la función de distribución

- *Límite inferior:* Si algún x_i toma el valor $-\infty$, la CDF se anula.

$$\lim_{x_i \to -\infty} F_{X_1,\dots,X_n}(x_1,\dots,x_i,\dots,x_n) = \lim_{x_i \to -\infty} \mathbb{P}(X_1 \leq x_1,\dots,X_i \leq x_i,\dots,X_n \leq x_n) = 0,$$

- *Límite superior:* Si todos los x_i toma el valor $+\infty$, la CDF llega a 1.

$$\lim_{x_1,x_2,\dots x_n \to +\infty} F_{X_1,\dots,X_n}(x_1,\dots,x_i,\dots,x_n) = \lim_{x_1,x_2,\dots x_n \to +\infty} \mathbb{P}(X_1 \leq x_1,\dots,X_i \leq x_i,\dots,X_n \leq x_n) = 1,$$

Distribuciones marginales

Si algún argumento toma el valor $+\infty$, la CDF se *marginaliza*.

$$
\begin{aligned}
\lim_{x_i \to +\infty} F_{X_1,\dots,X_n}(x_1,\dots,x_i,\dots,x_n) = \\
\lim_{x_i \to +\infty} \mathbb{P}(X_1 \leq x_1,\dots,X_i \leq x_i,\dots,X_n \leq x_n) = \\
\mathbb{P}(X_1 \leq x_1,\dots,X_{i-1} \leq x_{i-1},X_{i+1} \leq x_{i+1},\dots,X_n \leq x_n) \\
= F_{X_1,\dots X_{i-1},X_{i+1},\dots X_n}(x_1,\dots,x_{i-1},x_{i+1}\dots,x_n).
\end{aligned}
$$

$F_{X_1,\dots X_{i-1},X_{i+1},\dots X_n}$ es la distribución marginal con respecto a X_i. Por ejemplo, en el caso bivariable:

$$
F_{X_1,X_2}(\infty,x_2) = F_{X_2}(x_2), \quad F_{X_1,X_2}(x_1,\infty) = F_{X_1}(x_1)
$$

Claramente, $F_{X_1,X_2}(\infty,\infty) = 1$.

Función de densidad

Vector aleatorio continuo

$\mathbf{X}$ es un vector aleatorio continuo si $F_{\mathbf{X}}$ es continua y existen sus derivadas parciales de orden n. Su *función de densidad de probabilidad* [2] resulta:

$$f_{\mathbf{X}}(\boldsymbol{x}) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \frac{\partial^n F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}.$$

[2] PDF, *Probability Density Function*

Función de densidad continua: Propiedades

$\mathbf{X}$ es un VeA continuo. Su PDF, $f_{\mathbf{X}}$, satisface las siguientes propiedades:

- $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \geq 0,\ \forall \boldsymbol{x} \in \mathbb{R}^n$.
- $F_{\mathbf{X}}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{\mathbf{X}}(\mathbf{u})\, d\mathbf{u}$ para $\boldsymbol{x} = [x_1,\ldots,x_n]^t$.

Función de densidad multivariable: Propiedades

- *Probabilidad de una región*

$$\mathbb{P}(\mathbf{X} \in A) = \int_{\mathbf{u} \in A} f_{\mathbf{X}}(\mathbf{u})\, d\mathbf{u}.$$

Claramente, $\int_{\mathbb{R}^n} f_{\mathbf{X}}(\boldsymbol{x}) dx_1 \dots dx_n = \mathbb{P}(\mathbf{X} \in \mathbb{R}^n) = 1$.

- *Probabilidad de un punto.* Si **X** es un VeA continuo, entonces

$$\mathbb{P}(\mathbf{X} = \boldsymbol{x}_0) = 0, \ \forall \boldsymbol{x}_0 \in \mathbb{R}^n.$$

Funciones de densidad de probabilidad marginales

A partir de la PDF $f_{\mathbf{X}}(\boldsymbol{x})$ podemos obtener las PDFs *marginales*, integrando con respecto al resto de las variables. Por ejemplo:

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{X_1,X_2}(x_1, x_2)\, dx_2.$$

$$f_{X_3}(x_3) = \int_{\mathbb{R}^2} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\, dx_1\, dx_2.$$

$$f_{X_1,X_3}(x_1, x_3) = \int_{\mathbb{R}^2} f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4)\, dx_2\, dx_4.$$

Función de masa de probabilidad

Vector aleatorio discreto

$\mathbf{X}$ es un vector aleatorio discreto si $F_{\mathbf{X}}$ es constante por regiones y tiene un número finito o infinito numerable de discontinuidades. Su *función de masa de probabilidad* [3] resulta:

$$p_{\mathbf{X}}(\boldsymbol{x}) = p_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = \mathbb{P}(\mathbf{X} = \boldsymbol{x}),$$

donde $\boldsymbol{x}$ es uno de los puntos de discontinuidad de $F_{\mathbf{X}}$. Llamamos $\mathcal{X}$ al conjunto que contiene a todos los puntos de discontinuidad de $F_{\mathbf{X}}$.

[3] PMF *Probability Mass Function*

Vectores aleatorios discretos y funciones de masa de probabilidad

$\mathbf{X}$ es un VeA discreto. Su PMF, $p_{\mathbf{X}}(\boldsymbol{x})$, satisface las siguientes propiedades:

- $p_{\mathbf{X}}(\boldsymbol{x}) \geq 0, \ \forall \boldsymbol{x} \in \mathcal{X}$.
- $F_{\mathbf{X}}(\boldsymbol{x}) = \sum_{\mathbf{u} \in \mathcal{X} : \mathbf{u} \leq \boldsymbol{x}} p_{\mathbf{X}}(\mathbf{u})$.

Vectores aleatorios discretos y funciones de masa de probabilidad

- *Probabilidad de una región.*

$$\mathbb{P}(\mathbf{X} \in A) = \sum_{\mathbf{u} \in \mathcal{X} \cap A} p_{\mathbf{X}}(\mathbf{u}).$$

Nuevamente, $\sum_{\boldsymbol{x} \in \mathcal{X}} p_{\mathbf{X}}(\boldsymbol{x}) = \mathbb{P}(\mathbf{X} \in \mathbb{R}^n) = 1.$

- *Probabilidad de un punto.* Por definición

$$\mathbb{P}(\mathbf{X} = \boldsymbol{x}) = p_{\mathbf{X}}(\boldsymbol{x}), \ \forall \boldsymbol{x} \in \mathcal{X}.$$

En este caso, $\mathbb{P}(\mathbf{X} = \boldsymbol{x})$ puede ser no nula.

Funciones de masa de probabilidad marginales

De forma análoga al caso continuo, podemos marginalizar la PMF conjunta para obtener las *PMFs marginales*. Por ejemplo:

$$p_{X_1}(x_1) = \sum_{x_2:(x_1,x_2)\in\mathcal{X}} p_{X_1,X_2}(x_1,x_2).$$

$$p_{X_3}(x_3) = \sum_{(x_1,x_2):(x_1,x_2,x_3)\in\mathcal{X}} p_{X_1,X_2,X_3}(x_1,x_2,x_3).$$

$$p_{X_1,X_3}(x_1,x_3) = \sum_{(x_2,x_4):(x_1,x_2,x_3,x_4)\in\mathcal{X}} p_{X_1,X_2,X_3,X_4}(x_1,x_2,x_3,x_4).$$

Función de densidad generalizada

En general, para no tener que diferenciar entre VeA continuo y discreto vamos a considerar una función de densidad generalizada

- Si $\mathbf{X}$ está compuesto por VA continuas, entonces $f_{\mathbf{X}}$ sigue la definición vista.
- Si $\mathbf{X}$ está compuesto por V.A discretas que toman valores en $\boldsymbol{x} = \xi_i$, $i = 1, 2, \cdots$ entonces la función de densidad generalizada toma la siguiente forma

$$f_{\mathbf{X}}(\boldsymbol{x}) = \sum_i p_{\mathbf{X}}(\xi_i)\delta(\boldsymbol{x} - \xi_i)$$

Utilizando esta generalización, podemos trabajar las fórmulas en forma más genérica, sin tener que diferenciar entre la existencia de una función de densidad o no.

Condicionalidad

Distribuciones condicionales

Al analizar varias VA en forma conjunta es interesante introducir el concepto de condicionalidad. Para ello, definimos la CDF condicionada a un evento A que caracteriza el comportamiento de $\mathbf{X}$ cuando sabemos que A ocurrió. Si $\mathbb{P}(A) > 0$, tenemos:

$$F_{\mathbf{X}|A}(\boldsymbol{x}) = \mathbb{P}(\mathbf{X} \leq \boldsymbol{x}|A) = \frac{\mathbb{P}(\mathbf{X} \leq \boldsymbol{x}, A)}{\mathbb{P}(A)}.$$

PDF condicional

Analizamos primero el caso de un VeA continuo. Para hacer el tratamiento más concreto tomamos $n = 3$. Queremos obtener

$$F_{X2,X3|X1=x1}(x_2, x_3|x_1).$$

Consideramos el evento $A = \{x_1 < X_1 \leq x_1 + \Delta\}$. Entonces,

$$\begin{aligned} F_{X2,X3|A}(x_2, x_3|A) &= \mathbb{P}(X_2 \leq x_2, X_3 \leq x_3|x_1 < X_1 \leq x_1 + \Delta) \\ &= \frac{\mathbb{P}(X_2 \leq x_2, X_3 \leq x_3, x_1 < X_1 \leq x_1 + \Delta)}{\mathbb{P}(x_1 < X_1 \leq x_1 + \Delta)}. \end{aligned}$$

Hay que calcular:
$\mathbb{P}(X_2 \leq x_2, X_3 \leq x_3, x_1 < X_1 \leq x_1 + \Delta)$ y $\mathbb{P}(x_1 < X_1 \leq x_1 + \Delta)$.

PDF condicional

$$\begin{aligned}\mathbb{P}(X_1 \in A, X_2 \le x_2, X_3 \le x_3) &= \int_{x_1}^{x_1+\Delta} \int_{-\infty}^{x_2} \int_{-\infty}^{x_3} f_{X_1 X_2 X_3}(u, v, w) du dv dw \\ &= F_{X_1,X_2,X_3}(x_1 + \Delta, x_2, x_3) - F_{X_1,X_2,X_3}(x_1, x_2, x_3)\end{aligned}$$

y

$$\mathbb{P}(X_1 \in A) = \int_{u \in A} f_{X_1}(u) du = \int_{x_1}^{x_1+\Delta} f_{X_1}(u) du = F_{X_1}(x_1 + \Delta) - F_{X_1}(x_1).$$

PDF condicional

Luego,

$$F_{X2,X3|A}(x_2, x_3|A) = \frac{F_{X_1,X_2,X_3}(x_1+\Delta, x_2, x_3) - F_{X_1,X_2,X_3}(x_1, x_2, x_3)}{F_{X_1}(x_1+\Delta) - F_{X_1}(x_1)}$$
$$= \frac{\frac{F_{X_1,X_2,X_3}(x_1+\Delta,x_2,x_3)-F_{X_1,X_2,X_3}(x_1,x_2,x_3)}{\Delta}}{\frac{F_{X_1}(x_1+\Delta)-F_{X_1}(x_1)}{\Delta}}.$$

Usando límites, tenemos

$$F_{X_2,X_3|\{X_1=x_1\}}(x_2, x_3|x_1) =$$
$$\lim_{\Delta\to 0} F_{X_2,X_3|\{x_1\leq X_1\leq x_1+\Delta\}}(x_2, x_3|x_1 \leq X_1 \leq x_1+\Delta) = \frac{\frac{\partial F_{X_1,X_2,X_3}(x_1,x_2,x_3)}{\partial x_1}}{\frac{dF_{X_1}(x_1)}{dx_1}}.$$

PDF condicional

Finalmente, la PDF condicional resulta

$$
\begin{aligned}
f_{X_2,X_3|X_1=x_1}(x_2,x_3|x_1) &= \frac{\partial^2}{\partial x_2 \partial x_3} F_{X_2,X_3|\{X_1=x_1\}}(x_2,x_3|x_1) \\
&= \frac{\partial^2}{\partial x_2 \partial x_3} \frac{\frac{\partial F_{X_1,X_2,X_3}(x_1,x_2,x_3)}{\partial x_1}}{\frac{dF_{X_1}(x_1)}{dx_1}} \\
&= \frac{\frac{\partial^3 F_{X_1,X_2,X_3}(x_1,x_2,x_3)}{\partial x_1 \partial x_2 \partial x_3}}{f_{X_1}(x_1)} = \frac{f_{X_1,X_2,X_3}(x_1,x_2,x_3)}{f_{X_1}(x_1)}.
\end{aligned}
$$

En forma general

PDF condicional

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

PMF condicional

Cuando **X** es un VeA discreto es posible condicionar en un valor específico de una VA. Luego, la *PMF condicional* se define directamente a partir de la definición de probabilidad condicional. Por ejemplo:

$$p_{X_1,X_3|X_2=x_2}(x_1,x_3) = \frac{p_{X_1,X_2,X_3}(x_1,x_2,x_3)}{p_{X_2}(x_2)}.$$

$$p_{X_1|X_2=x_2,X_3=x_3}(x_1) = \frac{p_{X_1,X_2,X_3}(x_1,x_2,x_3)}{p_{X_2,X_3}(x_2,x_3)}.$$

En este caso nuevamente,

PMF condicional

$$p_{\mathbf{X}|\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y}) = \frac{p_{\mathbf{X},\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y})}{p_{\mathbf{Y}}(\boldsymbol{y})}$$

Notación

Vamos a utilizar la siguiente notación en forma equivalente para determinar condicionalidad en distribuciones, densidades, y funciones de masa de probabilidad:

$$F_{\mathbf{X}_1|\mathbf{X}_2=\boldsymbol{x}_2}(\boldsymbol{x}_1) = F_{\mathbf{X}_1}(\boldsymbol{x}_1|\mathbf{X}_2 = \boldsymbol{x}_2) = F(\boldsymbol{x}_1|\boldsymbol{x}_2)$$

$$f_{\mathbf{X}_1|\mathbf{X}_2=\boldsymbol{x}_2}(\boldsymbol{x}_1) = f_{\mathbf{X}_1}(\boldsymbol{x}_1|\mathbf{X}_2 = \boldsymbol{x}_2) = f(\boldsymbol{x}_1|\boldsymbol{x}_2)$$

$$p_{\mathbf{X}_1|\mathbf{X}_2=\boldsymbol{x}_2}(\boldsymbol{x}_1) = p_{\mathbf{X}_1}(\boldsymbol{x}_1|\mathbf{X}_2 = \boldsymbol{x}_2) = p(\boldsymbol{x}_1|\boldsymbol{x}_2)$$

Regla de la cadena

Podemos hallar la PDF y la PMF conjuntas a partir de las condicionadas. Por ejemplo:

$$
\begin{aligned}
& f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \\
& = f_{X_1}(x_1) f_{X_2,X_3}(x_2,x_3|X_1=x_1) = f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3|X_1=x_1,X_2=x_2) \\
& = f_{X_3}(x_3) f_{X_1,X_2}(x_1,x_2|X_3=x_3) = f_{X_3}(x_3) f_{X_2}(x_2) f_{X_1}(x_1|X_2=x_2,X_3=x_3) \\
& = f_{X_2}(x_2) f_{X_1,X_3}(x_1,x_3|X_2=x_2) = f_{X_2}(x_2) f_{X_3}(x_3) f_{X_1}(x_1|X_2=x_2,X_3=x_3)
\end{aligned}
$$

En forma similar, tenemos las PMF conjunta de 3 VAD:

$$
\begin{aligned}
& p_{X_1,X_2,X_3}(x_1,x_2,x_3) = \\
& = p_{X_1}(x_1) p_{X_2,X_3}(x_2,x_3|X_1=x_1) = p_{X_1}(x_1) f_{X_2}(x_2) p_{X_3}(x_3|X_1=x_1,X_2=x_2) \\
& = p_{X_3}(x_3) p_{X_1,X_2}(x_1,x_2|X_3=x_3) = p_{X_3}(x_3) p_{X_2}(x_2) p_{X_1}(x_1|X_2=x_2,X_3=x_3) \\
& = p_{X_2}(x_2) p_{X_1,X_3}(x_1,x_3|X_2=x_2) = p_{X_2}(x_2) p_{X_3}(x_3) p_{X_1}(x_1|X_2=x_2,X_3=x_3)
\end{aligned}
$$

Independencia entre VA

Independencia entre variables aleatorias

Dos VA X e Y son *independientes* si

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) = F_X(x)F_Y(y).$$

- Si X, Y son VA discretas, la condición es equivalente a

$$p_{X,Y}(x,y) = p_X(x)p_Y(y).$$

- Si X, Y son VA continuas la condición es equivalente a

$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

Independencia entre vectores aleatorias

- Las componentes de un vector aleatorio $\mathbf{X}$ son mutuamente independientes si

$$F_{\mathbf{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} F_{X_i}(x_i).$$

- Dos VeAs $\mathbf{X} \in \mathbb{R}^n$ e $\mathbf{Y} \in \mathbb{R}^m$ son independientes si

$$F_{\mathbf{X},\mathbf{Y}}(\boldsymbol{x}, \boldsymbol{y}) = F_{\mathbf{X}}(\boldsymbol{x}) F_{\mathbf{Y}}(\boldsymbol{y}).$$

Ejercicios

Ejercicio

Sea $Y = X + N$, con X y N variables aleatorias independientes.

1. Demostrar que $f_{Y|X}(y|x) = f_N(y - x)$.
2. Si $X \in \{0, 1\}$ es una variable aleatoria Bernoulli con $\mathbb{P}(X = 0) = p$ y $\mathbb{P}(X = 1) = q = 1 - p$, expresar y representar $f_Y(y)$ y $f_{Y|X}(y|x)$.

Ejercicio

Sea $\mathbf{X} = [X_1, X_2]$ un vector aleatorio continuo. Si $f_{\mathbf{X}}(\boldsymbol{x})$ fuera conocida, cómo haría para calcular las siguientes probabilidades?

- $\mathbb{P}(X_1 - X_2 \leq 2)$
- $\mathbb{P}(\max(X_1, X_2) \leq -1)$
- $\mathbb{P}(\min(|X_1|, |X_2|) \geq 2)$
- $\mathbb{P}(X_1 X_2 \geq 0)$

Ejercicio

Sean X, Y, y Z V.A independientes. Hallar las siguientes probabilidades en función de $F_X(x)$, $F_Y(y)$, $F_Z(z)$

- $\mathbb{P}(|X| \leq 5, Y > 3, Z^2 \leq 2)$
- $\mathbb{P}(\text{máx}(X, Y, Z) \leq -1)$

Ejercicio

Sea $\mathbf{X} = [X_1, X_2]$ un vector aleatorio continuo cuya PDF es

$$F_{\mathbf{X}}(x_1, x_2) = k(x_1 + x_2) \qquad 0 < x_1 < 1 \quad 0 < x_2 < 1$$

.

- Hallar k
- Hallar $F_{\mathbf{X}}$
- Hallar $f_{X_1}(x_1)$ y $f_{X_2}(x_2)$.

Esperanza

Lo último q' se pierde

El operador esperanza

Sea $\mathbf{X} \in \mathbb{R}^n$ un vector aleatorio n-dimensional. La media o esperanza de $\mathbf{X}$ es vector de las esperanzas de sus componentes:

$$\mathbb{E}[\mathbf{X}] = \mu_{\mathbf{X}} = \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix}.$$

El vector de esperanzas es un vector con cada una de sus esperanzas

Demostración

$$\mathbb{E}[\mathbf{X}] = \int_{-\infty}^{\infty} \boldsymbol{x}\, f_{\mathbf{X}}(\boldsymbol{x}) d\boldsymbol{x} = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} f_{X_1 \cdots X_n}(x_1, \cdots x_n) dx_1 \cdots dx_n$$

Marginalizar, me saco de encima todos los que integro (i.e. todos menos x1)

vector de integrales OBVIO

$$= \begin{bmatrix} \int_{-\infty}^{\infty} x_1 dx_1 \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 \cdots X_n}(x_1, x_2 \cdots x_n) dx_2 \cdots dx_n\right] \\ \vdots \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 \cdots X_n}(x_1, \cdots x_{n-1}, x_n) dx_1 \cdots dx_{n-1} \int_{-\infty}^{\infty} x_n dx_n \end{bmatrix}$$

$$= \begin{bmatrix} \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 \\ \vdots \\ \int_{-\infty}^{\infty} x_n f_{X_n}(x_n) dx_n \end{bmatrix}$$

Todo esto para mostrar que lo de arriba NO es una definición, sino una deducción

Pero re

$$= \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix}$$

Chimardo, pero el resultado es fácil

El operador esperanza: Propiedades

♡ → Esto pretende ser un Corazón

- El operador $\mathbb{E}[]$ es lineal, luego $\forall \mathbf{A} \in \mathbb{R}^{m\times n}$ y $\forall \mathbf{b} \in \mathbb{R}^{m}$ tenemos:

$$\mathbb{E}[\mathbf{A}\mathbf{X} + \mathbf{b}] = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{b}.$$

- En forma general, si $\boldsymbol{g}$ es una función $\boldsymbol{g} : \mathbb{R}^{n} \to \mathbb{R}^{m}$, tenemos

Donde $\vec{g}(\vec{X}) = \begin{pmatrix} g_1(\vec{X}) \\ g_2(\vec{X}) \\ \vdots \\ g_m(\vec{X}) \end{pmatrix}$

$$\mathbb{E}[\boldsymbol{g}(\mathbf{X})] = \begin{bmatrix} \mathbb{E}[g_1(\mathbf{X})] \\ \vdots \\ \mathbb{E}[g_m(\mathbf{X})] \end{bmatrix}.$$

La esperanza condicional

Sea un evento A. Luego, definimos el operador $\mathbb{E}[\cdot|A]$

$$\mathbb{E}[\mathbf{X}|A] = \int_{\mathbb{R}^n} \boldsymbol{x} \; f_{\mathbf{X}|A}(\boldsymbol{x}|A) \; d\boldsymbol{x}.$$

Propiedades de la esperanza condicional

- *Esperanza condicional de una constante.* $\mathbb{E}[\boldsymbol{a}|\mathbf{X} = \boldsymbol{x}] = \boldsymbol{a}$.
- *Linealidad.* $\mathbb{E}[a\mathbf{Y} + b\mathbf{Z}|\mathbf{X} = \boldsymbol{x}] = a\mathbb{E}[\mathbf{Y}|\mathbf{X} = \boldsymbol{x}] + b\mathbb{E}[\mathbf{Z}|\mathbf{X} = \boldsymbol{x}]$.
- *Esperanza condicional de una función.*
  $\mathbb{E}[\boldsymbol{g}(\mathbf{X})\boldsymbol{h}(\mathbf{Y})|\mathbf{X} = \boldsymbol{x}] = \boldsymbol{g}(\boldsymbol{x})\mathbb{E}[\boldsymbol{h}(\mathbf{Y})|\mathbf{X} = \boldsymbol{x}]$. → Si $\vec{X}$ es fija ($\vec{X} = \vec{x}$), $g(\vec{X})$ sale como constante ($g(\vec{x})$)

  minúscula
- *Redundancia en condiciones.*
  $\mathbb{E}[\mathbf{Y}|\underbrace{\mathbf{X} = \boldsymbol{x}}, \underbrace{\boldsymbol{g}(\mathbf{X}) = \boldsymbol{g}(\boldsymbol{x})}] = \mathbb{E}[\mathbf{Y}|\mathbf{X} = \boldsymbol{x}]$.

  Significan lo mismo, redundante

Ley de la esperanza total I

Sumamente importante

$\mathbf{X}_1 \in \mathbb{R}^n, \mathbf{X}_2 \in \mathbb{R}^m$

A.K.A. "Ley de suavización", por si eso te dice algo

$$\mathbb{E}[\mathbf{X}_2] = \mathbb{E}\big[\mathbb{E}[\mathbf{X}_2|\mathbf{X}_1]\big] \qquad \text{y} \qquad \mathbb{E}[\mathbf{X}_1] = \mathbb{E}\big[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]\big]$$

Demostración:

¿?

- Consideramos evento $A = \{\mathbf{X}_1 = \boldsymbol{x}_1\}$

también es un vector

$$\mathbb{E}[\mathbf{X}_2|A] = \int \boldsymbol{x}_2 \, f_{\mathbf{X}_2|\mathbf{X}_1}(\boldsymbol{x}_2|\mathbf{X}_1 = \boldsymbol{x}_1) \, d\boldsymbol{x}_2 = g(\boldsymbol{x}_1).$$

- $g(\boldsymbol{x}_1)$ es una función de $\boldsymbol{x}_1$. Si consideramos todas sus posibles realizaciones, $g(\mathbf{X}_1)$ es un VeA de m componentes como $\mathbf{X}_2$.

Ley de la esperanza total II

- Calculando la esperanza de $g(\mathbf{X}_1) = \int \boldsymbol{x}_2 \, f_{\mathbf{X}_2|\mathbf{X}_1}(\boldsymbol{x}_2|\mathbf{X}_1 = \boldsymbol{x}_1) \, d\boldsymbol{x}_2$,

randomizo → Técnica interesante para demostrar cosas q' impliquen condicionalidad

$\vec{g}(\vec{x}_1)$

$$\mathbb{E}[g(\mathbf{X}_1)] = \mathbb{E}\Big[\mathbb{E}[\mathbf{X}_2|\mathbf{X}_1]\Big] = \int_{\mathbb{R}} g(\boldsymbol{x}_1) \, f_{\mathbf{X}_1}(\boldsymbol{x}_1) \, d\boldsymbol{x}_1 =$$

$$= \int\int \vec{x}_2 \cdot f_{X_2|X_1}(\vec{x}_2|\vec{x}_1) \, d\vec{x}_2 \cdot f_{X_1} \cdot d\vec{x}_1 = \int_{\mathbb{R}}\int_{\mathbb{R}} \boldsymbol{x}_2 \; \underbrace{f_{\mathbf{X}_2|\mathbf{X}_1}(\boldsymbol{x}_2|\mathbf{X}_1 = \boldsymbol{x}_1) \; f_{\mathbf{X}_1}(\boldsymbol{x}_1)}_{f_{\mathbf{X}_1,\mathbf{X}_2}(\boldsymbol{x}_1,\boldsymbol{x}_2)} \; d\boldsymbol{x}_1 d\boldsymbol{x}_2$$

$$= \int_{\mathbb{R}} \vec{x}_2 \underbrace{\int_{\mathbb{R}} f_{\mathbf{X}_1,\mathbf{X}_2}(\boldsymbol{x}_1, \boldsymbol{x}_2) \, d\boldsymbol{x}_1}_{f_{\mathbf{X}_2}(\boldsymbol{x}_2)} \, d\boldsymbol{x}_2 = \mathbb{E}[\mathbf{X}_2]$$

De todos modos, parece medio contradictorio. Imponer una condición es asumir que X1 tomó un valor, pero después digo "X1 puede tomar cualquier valor". Flojo de papeles?

$\therefore \; E[\vec{g}(\vec{X}_1)] = E\big[E[\vec{X}_2|\vec{X}_1]\big] = E[\vec{X}_2]$

Más allá de la demostración (*) ¿Qué significa esto?

Esperanza condicional de VA independientes

① $\to E[E[X_2|X_1]] = E[X_2]$

Dada una realización de $\vec{X1}$, $\vec{X2}$ tiene cierta esperanza, $\vec{X2}|\vec{X1} = \vec{x1}$. Si desconozco ese valor de $\vec{X1}$, la esperanza de $\vec{X2}$ no cambia. → ¿No es medio obvio?

"Si no sé cuánto valió $\vec{X}_1$, la esperanza de X_2 no cambia" → y sí, maestro

¿Lo estoy entendiendo bien? ¿consultar? ¿xq es tan importante este teorema?

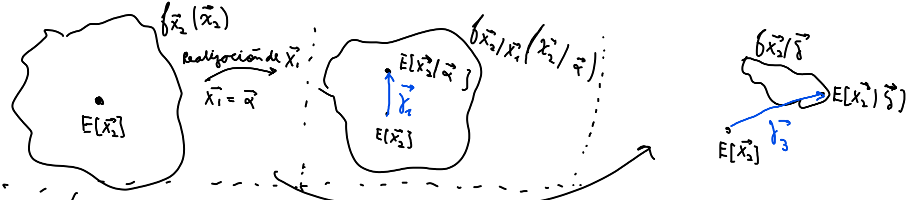
Si **X** y **Y** son independientes, entonces $\mathbb{E}[\mathbf{X}|\mathbf{Y}] = \mathbb{E}[\mathbf{X}]$.

¿Qué es la esperanza?¿El valor más probable? ==> NO!
Es el valor esperado. ¿Y qué cuernos significa eso?
Es una suma de todos los posibles valores de mi variable ponderados por su probabilidad $\to E[x] = \sum_{\forall x} p(x) \cdot x$
De alguna manera, la esperanza representa un centro de masa en el gráfico de p(x).

Esta propiedad es una consecuencia directa de

$$f_{\mathbf{X},\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y}) = f_{\mathbf{X}}(\boldsymbol{x}) f_{\mathbf{Y}}(\boldsymbol{y}) \iff \mathbf{X}, \mathbf{Y} \text{ independientes}$$

Con esta interpretación, el teorema sería: "El C de M de la variable aleatoria g(X1) es el mismo que el de la variable aleatoria X2"
¿Tiene sentido hablar de C de M de una V.A. que es un vector? ==> Sí, es la posición promedio de la función de distribución de ese vector, que puede ser una función de R^m a R. Además, decir C de M y posición promedio es lo mismo.
Entonces, el teorema sería: "Dada una realización de X1, la posición promedio de X2 cambia. Si yo no sé cuál es el valor de X1, ¿cuál es la posición promedio de la posición promedio de X2? ==> Resulta que es exactamente igual a la posición promedio de X2!!!!. Ahora el teorema no es tan obvio.



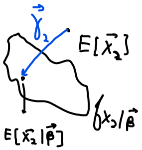
otra realización $\vec{X}_1 = \vec{\beta}$

Una tercera realización: $\vec{X}_1 = \vec{\delta}$

Ahora te pregunto, ¿Dónde caen en promedio las nuevas esperanzas $E[\vec{X}_2|\vec{\alpha}], E[\vec{X}_2, \vec{\beta}]$ y $E[\vec{X}_2, \vec{\delta}]$?

Matriz de covarianza

↳ Caen en $E[\vec{X}_2]$!!!



o también: $E[\vec{\gamma}] = 0$!!! → Dou

Creo q' caché

Sumar una dimensión al problema ayuda a visualizar, aunque tal vez no sea tan difícil verlo el R1.
Ver más allá del resultado

Si voy a hacer esto por cada teorema me quedo peladísimo

31/3==> Además, puede que no conozca la distribución de X2, pero sí conozca la distribución de X2|X1 = x1

Repaso varianza y correlación

Si X es una V.A de media es μ_X, su *varianza* está definida como

$$\mathbb{V}[X] = \sigma_X^2 = \mathbb{E}[(X - \mu_X)^2].$$

Centro la variable. Quiero ver cómo se comporta alrededor de la media

Tenemos que

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mu_X^2.$$

Cuadrado del binomio

Momentos cruzados entre variables aleatorias

Sean X e Y dos V.A. con $\mathbb{E}[X] = \mu_X$ y $\mathbb{E}[Y] = \mu_Y$.

- Momento cruzado central o *covarianza cruzada* ¿y esto q' significa?

Multiplico ambas variables

$$C_{X,Y} = \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right]$$

Pensar dsp. si no llego a terminar el video

- Momento cruzado o *correlación cruzada* (sin centrar)

$$R_{X,Y} = \mathbb{E}\left[XY\right]$$

Tenemos que

$$C_{X,Y} = R_{X,Y} - \mu_X \mu_Y.$$

→ desarrollar el producto

Matriz de covarianza

Sea $\mathbf{X} \in \mathbb{R}^n$ un VeA. La **Matriz de Covarianza** de $\mathbf{X}$ es

La esperanza de la matriz es la matriz de las esperanzas

$$\underbrace{\mathbf{Cov}(\mathbf{X})}_{\in \mathbb{R}^{n \times n}} = C_{\mathbf{X}} = \mathbb{E}[\underbrace{(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^t}]$$

donde

$$\mathbf{X} - \mu_{\mathbf{X}} = \begin{bmatrix} X_1 - \mu_{X_1} \\ \vdots \\ X_n - \mu_{X_n} \end{bmatrix}$$

$$\begin{bmatrix} X_1 - \mu_{X_1} \\ X_2 - \mu_{X_2} \\ \vdots \\ X_n - \mu_{X_n} \end{bmatrix} \cdot [X_1 - \mu_1, X_2 - \mu_2 \cdots]$$

$$(\vec{X} - \mu_{\vec{X}}) \cdot (\vec{X} - \mu_{\vec{X}})^T \in \mathbb{R}^{n \times n}$$

Armado de la matriz de covarianza

$$C_{\mathbf{X}} = \mathbb{E}\left\{ \begin{bmatrix} X_1 - \mu_{X_1} \\ \vdots \\ X_n - \mu_{X_n} \end{bmatrix} \begin{bmatrix} X_1 - \mu_{X_1} & \cdots & X_n - \mu_{X_n} \end{bmatrix} \right\}$$

Hay q' quererla a esta matriz

$$= \mathbb{E}\left\{ \begin{bmatrix} (X_1 - \mu_{X_1})^2 & \cdots & (X_1 - \mu_{X_1})(X_n - \mu_{X_n}) \\ & \ddots & \\ (X_n - \mu_{X_n})(X_1 - \mu_{X_1}) & \cdots & (X_n - \mu_{X_n})^2 \end{bmatrix} \right\}$$

$$= \begin{bmatrix} \underbrace{\mathbb{E}[(X_1 - \mu_{X_1})^2]}_{\mathbb{V}[X_1]} & \cdots & \underbrace{\mathbb{E}[(X_1 - \mu_{X_1})(X_n - \mu_{X_n})]}_{Cov[X_1, X_n]} \\ & \ddots & \\ Cov[X_n, X_1] & \cdots & \mathbb{V}[X_n] \end{bmatrix}$$

Covarianza cruzada

Covarianza cruzada

Varianzas en la diagonal principal

Construcción de $C_{\mathbf{X}}$:

- Diagonal principal: $\mathbb{V}[X_1], \cdots, \mathbb{V}[X_n]$
- Término (i, j): $Cov[X_i, X_j]$

Matriz de correlación

Como la covarianza, pero sin centrar

Sea $\mathbf{X} in \in \mathbb{R}^n$ un VeA. La **Matriz de correlación** de **X** es:

Si la media del Ve.A. es cero, la matriz de correlación y la de covarianza es la misma

$$R_{\mathbf{X}} = \mathbb{E}[\mathbf{X}\mathbf{X}^t]$$

Como la matriz de covarianzas, pero sin centrar

Considerando la PDF generalizada $f_{\mathbf{X}}$ para incluir tanto el caso continuo como el discreto, tenemos:

$$R_{\mathbf{X}} = \int_{\mathbb{R}^n} \boldsymbol{x}\boldsymbol{x}^t f_{\mathbf{X}}(\boldsymbol{x}) d\boldsymbol{x} = \begin{bmatrix} \mathbb{E}[X_1^2] & \mathbb{E}[X_1 X_2] & \cdots & \mathbb{E}[X_1 X_n] \\ \mathbb{E}[X_2 X_1] & \mathbb{E}[X_2^2] & \cdots & \mathbb{E}[X_2 X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[X_n X_1] & \mathbb{E}[X_n X_2] & \cdots & \mathbb{E}[X_n^2] \end{bmatrix}.$$

Propiedades de $C_{\mathbf{X}}$ y $R_{\mathbf{X}}$

- Matrices cuadradas, $R_{\mathbf{X}}, C_{\mathbf{X}} \in \mathbb{R}^{n\times n}$
- Matrices simétricas: $R_{\mathbf{X}} = R_{\mathbf{X}}^t$, $C_{\mathbf{X}} = C_{\mathbf{X}}^t$
  - $\mathbb{E}[X_i X_j] = \mathbb{E}[X_j X_i]$ y en forma similar, $Cov_{X_i X_j} = Cov_{X_j X_i}$
  - Elemento (i, j) = elemento (j, i).

Nos gustan las simetrias no?

Propiedades de $C_{\mathbf{X}}$ y $R_{\mathbf{X}}$

- Relación entre $C_{\mathbf{X}}$ y $R_{\mathbf{X}}$

¿Formas cuadráticas?

$$C_{\mathbf{X}} = \mathbb{E}\left[(\mathbf{X}-\mu_{\mathbf{X}})(\mathbf{X}-\mu_{\mathbf{X}})^t\right]$$
$$= \mathbb{E}[\mathbf{X}\mathbf{X}^t - \mathbf{X}\mu_{\mathbf{X}}^t - \mu_{\mathbf{X}}\mathbf{X}^t + \mu_{\mathbf{X}}\mu_{\mathbf{X}}^t] = R_{\mathbf{X}} - \mu_{\mathbf{X}}\mu_{\mathbf{X}}^t.$$

$$C_{\mathbf{X}} = R_{\mathbf{X}} - \mu_{\mathbf{X}}\mu_{\mathbf{X}}^t.$$ DoU

- Matrices semidefinidas positivas:

donde $f(\cdot)$ es una función lineal de $\mathbb{R}^n \to \mathbb{R}^m$

$= f(\vec{a})$ $a^t \cdot a \to$ producto interno

$$\forall \mathbf{a} \in \mathbb{R}^n: \quad \mathbf{a}^t (R_{\mathbf{X}}\mathbf{a}) \geq 0 \quad \text{y} \quad \mathbf{a}^t C_{\mathbf{X}}\mathbf{a} \geq 0.$$

¿Cuádricas?

MUY IMPORTANTE

Si yo tomo la imagen de $\vec{a}$ a través de Rx y le hago el P.I. con $\vec{a}$, el resultado es mayor a cero, el ángulo entre a y Rx*a no es mayor a 90°. Ponele. ¿Qué cuernos es un angulo entre 0 y 90° en 7 dimensiones??? ¿Qué significa esta propiedad?

$C_{\mathbf{X}} \geq 0$ y $R_{\mathbf{X}} \geq 0$: Demostración

- Armamos una nueva V.A. Y combinando linealmente las componentes de $\mathbf{X}$.

¿?

- $Y = \sum_{i=1}^{n} a_i X_i = \boldsymbol{a}^t \mathbf{X}$, donde $\boldsymbol{a} \in \mathbb{R}^n$ es un vector determinístico.
- Claramente $0 \leq \mathbb{E}[Y^2]$ para cualquier valor de $\mathbf{a}$. Por otro lado,

Xq un real al cuadrado siempre es mayor a cero. ¿Y los complejos? ¿Será por eso lo de "semi" definida?

$$\mathbb{E}[Y^2] = \mathbb{E}[\mathbf{a}^t \mathbf{X}\mathbf{X}^t \mathbf{a}] = \boldsymbol{a}^t \boldsymbol{R}_{\mathbf{X}} \boldsymbol{a} \geq 0$$

→ NANA NA

Cine, CINEMA PARADISO

- Haciendo el mismo razonamiento con $Z = \sum_{i=1}^{n} a_i (X_i - \mu_i) = \boldsymbol{a}^t (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})$ obtenemos

$$\mathbb{E}[Z^2] = \boldsymbol{a}^t \boldsymbol{C}_{\mathbf{X}} \boldsymbol{a} \geq 0.$$

- Ambas desigualdades se cumplen para cualquier $\boldsymbol{a} \in \mathbb{R}^n$, por ende $\boldsymbol{R}_{\mathbf{X}} \geq 0$ y $\boldsymbol{C}_{\mathbf{X}} \geq 0$.

Propiedades de $R_{\mathbf{X}}$ y $C_{\mathbf{X}}$

Delta del hijo de remil puta de Kronecker

$\mathbb{E}[X_i^2]$

- Si $R_{\mathbf{X}}$ es una matriz diagonal, $\mathbb{E}[X_i X_j] = \delta_{i,j}$ y decimos que las componentes de **X** son **ortogonales** de a dos.
- Si $C_{\mathbf{X}}$ es una matriz diagonal, $C_{X_i X_j} = \delta_{i,j}\mathbb{V}[X_i]$, decimos que las componentes de **X** están **descorrelacionadas** de a dos.

"Definiciones sueltas" → Es decir ¿Qué cuernos significa esto?
Pensar. ¿Quién definió esto así? ¿Cuándo? ¿Por qué, con qué fin?

Propiedades de $C_{\mathbf{X}}$: VeA degenerado

Qué sucede si $C_{\mathbf{X}}$ es una matriz singular?

- Recordemos que cuando una V.A X tiene varianza nula

$$\sigma_X^2 = 0 \iff \mathbb{P}[X = \mu_X] = 1 \rightarrow \text{i.e. no hay dispersión}$$

  La V.A se comporta como una constante igual a su media.

- Para el caso n-dimensional tenemos el siguiente resultado:

Teorema

Sea $\mathbf{X}$ un VeA. Su matriz de covarianza $\boldsymbol{C}_{\mathbf{X}}$ es singular si y solo si existe un vector $\boldsymbol{v} \neq \mathbf{0}$ tal que

→ q' tiene determinante cero
→ q' no tiene inversa
(o q' tiene un A.V. nulo)

PI: $<v, \vec{X} - \vec{\mu}_X>$

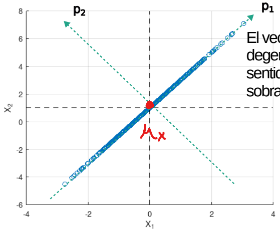
$$\mathbb{P}\left[\boldsymbol{v}^t(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) = 0\right] = 1.$$

Esto significa: Todas las observaciones de X centradas en la media mu van a ser ortogonales a v. Mi Ve.A. solo se da en una dirección. Loco, no?

En este caso decimos que $\mathbf{X}$ es un VeA *degenerado*.

Propiedades de $C_{\mathbf{X}}$: VeA degenerado

- Un modo de interpretar este resultado es ver que si $\mathbf{X}$ es un VeA degenerado, entonces existe una dirección fija ($\boldsymbol{v}$) que va a ser ortogonal a todas las realizaciones ($\boldsymbol{x} - \mu_{\mathbf{X}}$). Es decir que con probabilidad 1 NO vamos a observar realizaciones de $\mathbf{X}$ que se encuentren en el espacio generado por $\boldsymbol{v}$ desplazado en $\mu_{\mathbf{X}}$.
- Es decir, hay componentes del VeA redundantes que se pueden determinar a partir de las demás componentes.



El vector aleatorio se degenera en el sentido de que le sobra una dimensión

No hay realizaciones en dirección $\mathbf{p}_2$. Si se proyecta $\mathbf{X}$ sobre $\mathbf{p}_1$, se describe el experimento con una sola VA, $Z = \mathbf{X}^t\mathbf{p}_1$

Lo aleatorio no son las dos componentes de X, si no que lo aleatorio es r, un escalar, donde: $\vec{X} = r \cdot \vec{p_1}$ Es decir, me puedo ahorrar una dimensión!!!

VeA degenerados: demostración del teorema

- *Prueba ida.* $C_{\mathbf{X}}$ singular entonces $\exists \boldsymbol{v} \neq \mathbf{0}$: $\mathbb{P}(\boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}}) = 0) = 1$.
  $C_{\mathbf{X}}$ singular implica que $C_{\mathbf{X}}$ existe $\boldsymbol{v} \neq \mathbf{0}$ tal que $\boldsymbol{C}_{\mathbf{X}} \boldsymbol{v} = \mathbf{0}$. → autovalor nulo
  Entonces $\boldsymbol{v}^t \boldsymbol{C}_{\mathbf{X}} \boldsymbol{v} = 0$. Ahora definimos $Y = \boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}})$. Para cualquier $a > 0$, la cota de Markov establece que → Parecido a diapo 55 $Y \in \mathbb{R}$

  Y esto se cumple $\forall a > 0$ ←
  $$\mathbb{P}(Y^2 \geq a) \leq \frac{\mathbb{E}[Y^2]}{a} = \frac{\boldsymbol{v}^t \boldsymbol{C}_{\mathbf{X}} \boldsymbol{v}}{a} = 0,$$
  ¿Y si a=0? ¿Qué dice Markov? → ¿soy un hincha pelotas?

  Luego, $\mathbb{P}(Y^2 = 0) = 1$, y por ende, $\mathbb{P}(\boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}}) = 0) = 1$.
- *Prueba vuelta.* $\exists \boldsymbol{v} \neq \mathbf{0}$: $\mathbb{P}(\boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}}) = 0) = 1$ luego, $C_{\mathbf{X}}$ es singular.
  $Y = \boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}}) \in \mathbb{R}$ es una V.A. de media nula. Vimos que si $\mathbb{P}[Y = 0] = 1$ implica que su varianza es nula, es decir
  $$\sigma_Y^2 = \mathbb{E}[Y^2] = \boldsymbol{v}^t \boldsymbol{C}_{\mathbf{X}} \boldsymbol{v} = 0.$$
  Esto implica que $\boldsymbol{C}_{\mathbf{X}}$ es singular.

□

Transformación afín de $\mathbf{X}$

Sea $\mathbf{X}$ un VeA con covarianza $C_{\mathbf{X}}$. Consideramos una transformación afín formada a partir de la matriz $A \in \mathbb{R}^{m\times n}$ y el vector $\mathbf{b} \in \mathbb{R}^m$,

No lineal

$$\mathbf{Y} = A\mathbf{X} + \mathbf{b}.$$

Entonces

$$\mathbb{E}[\mathbf{Y}] = \mathbb{E}[A\mathbf{X} + \mathbf{b}] = A\mathbb{E}[\mathbf{X}] + \mathbf{b}$$

$$\begin{aligned}
\mathbf{Cov}(\mathbf{Y}) &= \mathbb{E}\left\{[\mathbf{Y} - \mathbb{E}[\mathbf{Y}]][\mathbf{Y} - \mathbb{E}[\mathbf{Y}]]^t\right\} \\
&= \mathbb{E}\left\{[A\mathbf{X} \cancel{+\mathbf{b}} - A\mathbb{E}[\mathbf{X}] \cancel{-\mathbf{b}}][A\mathbf{X} \cancel{+\mathbf{b}} - A\mathbb{E}[\mathbf{X}] \cancel{-\mathbf{b}}]^t\right\} \\
&= \mathbb{E}\left\{[A(\mathbf{X} - \mathbb{E}[\mathbf{X}])]\,[\underline{A}(\mathbf{X} - \mathbb{E}[\mathbf{X}])]^t\right\} \\
&= \mathbb{E}\left\{[A(\mathbf{X} - \mathbb{E}[\mathbf{X}])][(\mathbf{X} - \mathbb{E}[\mathbf{X}])]^t A^t\right\} \\
&= A\mathbb{E}\left\{[\mathbf{X} - \mathbb{E}[\mathbf{X}]][\mathbf{X} - \mathbb{E}[\mathbf{X}]]^t\right\} A^t = AC_{\mathbf{X}}A^t
\end{aligned}$$

Muy conveniente para el desarrollo de la trama

Transformación afín (recap importante!)

$\mathbf{Y} = A\mathbf{X} + \mathbf{b}$

$$\mathbb{E}[\mathbf{Y}] = \mu_{\mathbf{Y}} = A\mathbb{E}[\mathbf{X}] + \mathbf{b}$$

$$\mathbf{Cov}(\mathbf{Y}) = C_{\mathbf{Y}} = AC_{\mathbf{X}}A^t$$

Observaciones:

- **b** desplaza la media
- **b** no afecta $C_{\mathbf{Y}}$ pero SÍ afecta $R_{\mathbf{Y}}$

Xq la Cx se centra en la media.
Trabaja al rededor de la media

Relación entre correlación e independencia

- X e Y están *descorrelacionadas* si y sólo si $C_{X,Y} = 0$. En este caso, $R_{X,Y} = \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
- X e Y son *ortogonales* si y sólo si $R_{X,Y} = 0$. En este caso, $C_{X,Y} = -\mathbb{E}[X]\mathbb{E}[Y]$.
- X e Y son independientes si y sólo si $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

Que la covarianza sea nula NO implica que las variables sean independientes. Como serían dos variables dependientes y con covarianza distinta de cero?

- X , Y independientes $\Longrightarrow$ X , Y descorrelacionadas.
- X , Y independientes $\not\Longleftarrow$ X , Y descorrelacionadas. !

Desigualdad de Cauchy-Schwarz

¿Normas? ¿Ángulos? ¿Distancias?

Sean X e Y dos VA. Luego,

¿cuál es la norma de una V.A?

¿cuál es la base de todas las V.A? ¿Tienen sentido estas preguntas?

$$\mathbb{E}^2[XY] \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

Esta expresión es la *desigualdad de Cauchy-Schwarz*. Una formulación equivalente considera los momentos centrados, resultando

$$C^2_{X,Y} \leq \sigma^2_X \sigma^2_Y.$$

20/3 → Hasta acá llegó el amor

31/03 → ARRANCAMOS

Transformaciones de vectores aleatorios

Transformaciones de vectores aleatorios

Sea $\mathbf{X}, \mathbf{Y}$ dos VeAs tal que $\mathbf{Y} = \boldsymbol{g}(\mathbf{X})$ donde

$$\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m.$$

La función de distribución de $\mathbf{Y}$ es

$$F_{\mathbf{Y}}(\boldsymbol{y}) = \mathbb{P}(\mathbf{Y} \leq \boldsymbol{y}) = \mathbb{P}(\boldsymbol{g}(\mathbf{X}) \leq \boldsymbol{y}) = \int_{R_{\boldsymbol{y}}} f_{\mathbf{X}}(\boldsymbol{x}) d\boldsymbol{x},$$

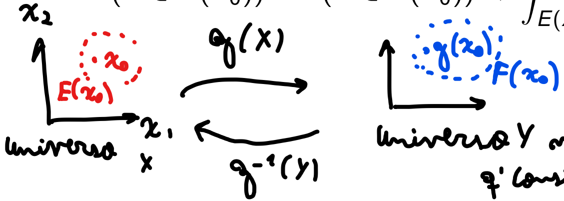
donde $\overbrace{R_{\boldsymbol{y}} = \{\boldsymbol{x} : \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{y}\}}^{\text{EVENTO EQUIVALENTE}}$. Podemos obtener la función de densidad de $\mathbf{Y}$ (posiblemente generalizada) como

$$f_{\mathbf{Y}}(\boldsymbol{y}) = \frac{\partial^m F_{\mathbf{Y}}(\boldsymbol{y})}{\partial y_1 \dots \partial y_m}.$$

Transformación biyectiva y suave

- Supongamos $n = m$ y $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^n$ una función inversible
- Luego, $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{y}$ tiene una única solución.
- Sea $E(\boldsymbol{x}_0)$ un entorno alrededor de $\boldsymbol{x}_0$ y $F(\boldsymbol{x}_0) = \{\boldsymbol{g}(\boldsymbol{x}) : \boldsymbol{x} \in E(\boldsymbol{x}_0)\}$ el entorno alrededor de $\boldsymbol{g}(\boldsymbol{x}_0)$.
- Para satisfacer la equivalencia de eventos, planteamos:

$$\mathbb{P}(\mathbf{X} \in E(\boldsymbol{x}_0)) = \mathbb{P}(\mathbf{Y} \in F(\boldsymbol{x}_0)) \Rightarrow \int_{E(\boldsymbol{x}_0)} f_{\mathbf{X}}(\boldsymbol{x}) d\boldsymbol{x} = \int_{F(\boldsymbol{x}_0)} f_{\mathbf{Y}}(\boldsymbol{y}) d\boldsymbol{y}.$$



Si $y_0 = g(x)$ tuviera n soluciones, tendría n entornos $E(x_0)$, q' tendría q' considerar en el evento equi.

Transformación biyectiva y suave

- Considerando un cambio de variable en la integral, obtenemos la siguiente relación entre las PDFs conjuntas:

$$f_{\mathbf{Y}}(\boldsymbol{y}) = \frac{f_{\mathbf{X}}(\boldsymbol{g}^{-1}(\boldsymbol{y}))}{|\det(J(\boldsymbol{g}^{-1}(\boldsymbol{y})))|},$$

donde J es el Jacobiano de la transformación dado por

¿y si $\nexists \frac{\partial g}{\partial x}$?

¿De qué me disfrazo? → No rompas las pelotas

$$J(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\boldsymbol{x}) & \dots & \frac{\partial g_1}{\partial x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \vdots \\ \frac{\partial g_n}{\partial x_1}(\boldsymbol{x}) & \dots & \frac{\partial g_n}{\partial x_n}(\boldsymbol{x}) \end{bmatrix}.$$

Transformación general suave

- Si $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^n$ es una función suave pero no inversible, dado un $\boldsymbol{y}$ la ecuación $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{y}$ tiene en general varias soluciones $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. → LO DE ANTES
- En este caso, hay que sumar las contribuciones de cada una de las soluciones. Luego,

$$f_{\mathbf{Y}}(\boldsymbol{y}) = \sum_{i=1}^{k} \frac{f_{\mathbf{X}}(\boldsymbol{x}_i)}{|\det(J(\boldsymbol{x}_i))|},$$

donde $J(\boldsymbol{x}_i)$ es el Jacobiano de la transformación dado por

$$J(\boldsymbol{x}_i) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\boldsymbol{x}_i) & \ldots & \frac{\partial g_1}{\partial x_n}(\boldsymbol{x}_i) \\ \vdots & \vdots & \vdots \\ \frac{\partial g_n}{\partial x_1}(\boldsymbol{x}_i) & \ldots & \frac{\partial g_n}{\partial x_n}(\boldsymbol{x}_i) \end{bmatrix}.$$

Ejemplo: Magnitud y ángulo de vector aleatorio I

- Sea $\mathbf{X} = [X_1, X_2]^t$ un vector aleatorio. Nos interesa hallar la PDF conjunta de la representación del vector en coordenadas polares:

$$\mathbf{Y} = \begin{bmatrix} R \\ \Theta \end{bmatrix} = \begin{bmatrix} \sqrt{X_1^2 + X_2^2} \\ \arctan\left(\frac{X_2}{X_1}\right) \end{bmatrix}.$$

- Como el cambio de coordenadas cartersianas a polares es inversible, la transformación $\mathbf{Y} = \boldsymbol{g}(\mathbf{X})$ es inversible. La transformación inversa es

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} R\cos(\Theta) \\ R\sin(\Theta) \end{bmatrix} = \boldsymbol{g}^{-1}(\mathbf{Y}).$$

Ejemplo: Magnitud y ángulo de vector aleatorio II

- Primero calculamos la matriz Jacobiana y su determinante:

$$J(x_1, x_2) = \begin{bmatrix} \frac{x_1}{\sqrt{x_1^2+x_2^2}} & \frac{x_2}{\sqrt{x_1^2+x_2^2}} \\ \frac{-x_2}{x_1^2+x_2^2} & \frac{x_1}{x_1^2+x_2^2} \end{bmatrix}$$

$$\Rightarrow |\det(J(x_1, x_2))| = \frac{x_1^2 + x_2^2}{(x_1^2 + x_2^2)^{3/2}} = \frac{1}{r}.$$

- Aplicando la fórmula del método, tenemos que

$$f_{R,\Theta}(r, \theta) = r\ f_{X_1,X_2}(r\cos(\theta), r\sin(\theta)), \qquad r \geq 0, \theta \in [0, 2\pi).$$

Transformación general suave

El método anterior también se puede aplicar a una transformación $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$ con $m \neq n$ agregando variables auxiliares X_i o Y_i de forma tal que se obtenga una nueva transformación $\boldsymbol{h} : \mathbb{R}^p \to \mathbb{R}^p$ donde $p = \max(n, m)$.

Transformación lineal o afín (caso biyectivo)

$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b} \qquad \det(\mathbf{A}) \neq 0$

$$f_{\mathbf{Y}}(\boldsymbol{y}) = \frac{f_{\mathbf{X}}(\mathbf{A}^{-1}(\boldsymbol{y} - \mathbf{b}))}{|\det(\mathbf{A})|}$$

Suma de V.A independientes

Compensa para que $f_Y(y)$ integre 1

Sean $X_1, \cdots X_n$ independientes

Si $Y = \sum_{i=1}^{n} a_i X_i = \boldsymbol{a}^t \mathbf{X}$,

$$f_Y(y) = \frac{1}{|a_1|} f_{X_1}\left(\frac{y}{a_1}\right) * \ldots * \frac{1}{|a_n|} f_{X_n}\left(\frac{y}{a_n}\right).$$

WTF

Make love, not convolution

Simulación de V.A.

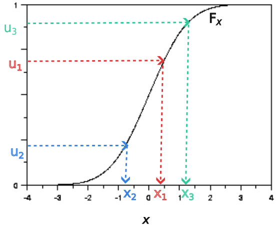
- Todo paquete numérico tiene un algoritmo que genera muestras de una V.A uniforme. Es un generador *pseudo aleatorio*
- Pero cómo generamos muestras de una V.A. no uniforme?
- Es posible utilizar una transformación aplicada a una V.A. $U \sim \mathcal{U}[0,1]$.

Simulación de V.A.: método de inversión

Montecarlo por inversión

Sea X una V.A. continua con distribución $F_X(x)$ y $U \sim \mathcal{U}[0,1]$.

Obtenemos una realización de X utilizando una realización de $F^{-1}(U)$.



Muestreo de $\mathcal{U}[0,1]$: $[u_1, u_2, u_3]$

$\downarrow$

Realizaciones de X: $[x_1, x_2, x_3]$

Método muy sencillo cuando se conoce F_X

Simulación de V.A.: método de inversión

Demostración:

- X continua $\Longrightarrow F_X()$ estrictamente creciente. Entonces existe F_X^{-1}

$$F_X^{-1} : (0,1) \to \mathbb{R}$$

$$\forall u, 0 < u < 1 \quad , \quad F_X^{-1}(u) = x.$$

- Recordemos que si $U \sim \mathcal{U}[0,1]$, entonces $F_U(u) = u$.
- Luego,

$$\mathbb{P}\left[F_X^{-1}(U) \leq x\right] = \mathbb{P}(U \leq F_X(x)) = F_U(F_X(x)) = F_X(x)$$

Es decir, las realizaciones de $F_X^{-1}(U)$ están distribuidas como X.

LOL

Ejercicio

- **Ejercicio 1**: Demuestre que si $U \sim \mathcal{U}[0,1]$, y $\lambda > 0$ luego,

Hacerlo

$$-\frac{1}{\lambda}\log(U) \sim Exp(\lambda)$$

- **Ejercicio 2**: *Algoritmo de Inversión para distribuciones discretas*

  Sea X una V.A discreta con PMF $p_X(k)$, $k = 0, 1, \cdots$ y distribución $F_X(j) = \sum_{k=0}^{j} p_X(k)$. Considere el siguiente pseudocódigo

  1. $j := -1$
  2. Generar $U \sim \mathcal{U}[0,1]$
  3. Repetir $j := j + 1$ hasta que $U < F_X(j)$
  4. Devolver j

  Demuestre que la salida del algoritmo sigue la distribución $F_X(x)$.