

Resumen Estadística III

Universidad Nacional de Córdoba - Facultad de Ciencias Económicas

Año 2017

Índice general

1. Unidad 1. Espacios de probabilidad	6
1.1. Espacio muestral	6
1.2. Eventos Aleatorios	6
1.3. Álgebra	6
1.4. σ -álgebra	7
1.5. σ -álgebra generada	7
1.6. σ -álgebra de Borel	8
1.7. Probabilidad finitamente aditiva	8
1.8. Medida de Probabilidad	9
1.9. Espacio de Probabilidad	9
1.9.1. Propiedades de P	9
1.10. Sucesiones de eventos	10
1.11. Continuidad de la Probabilidad	10
1.11.1. Continuidad en el vacío	10
1.12. Probabilidad condicional	10
1.13. Teorema de la Probabilidad Compuesta	11
1.13.1. Partición del espacio muestral	11
1.14. Independencia estocástica.	11
1.14.1. Independencia colectiva	12
2. Unidad 2. Variables Aleatorias.	13
2.1. Variable Aleatoria	13
2.1.1. Función medible	14
2.2. Función de Distribución.	15
2.2.1. Propiedades de la Función de Distribución	15
2.3. Tipos de variables aleatorias	16

2.3.1.	Variables aleatorias discretas	16
2.3.2.	Variable aleatoria absolutamente continua	17
2.3.3.	Variable aleatoria mixta	17
2.4.	Cálculo de probabilidades	18
2.4.1.	V.a. discreta	18
2.4.2.	V.a. continua (con densidad)	18
3.	Unidad 3. Vectores aleatorios	19
3.1.	Función de vectores aleatorios	20
3.1.1.	Función medible	20
3.2.	Función de Distribución de un vector aleatorio	21
3.2.1.	Propiedades de la función de distribución de un vector aleatorio	21
3.2.2.	Función de distribución marginal	23
3.3.	Vector aleatorio discreto	23
3.3.1.	Función de probabilidad conjunta de un subvector	26
3.4.	Vectores aleatorios absolutamente continuos	27
3.4.1.	Función de distribución de un subvector abs. continuo . .	30
3.5.	Distribución condicional	31
3.5.1.	Caso discreto	33
3.5.2.	Caso continuo	34
3.6.	Independencia de variables aleatorias	35
4.	Unidad 4. Distribución de funciones de vectores aleatorios	38
4.1.	Método de la función de distribución	39
4.2.	Método del Jacobiano/ de la transformación	43
4.2.1.	Aplicaciones del método del Jacobiano	49
4.3.	Distribuciones en el muestreo	49
4.3.1.	Estadístico	50
4.3.2.	Distribuciones en el muestreo en poblaciones normales .	50
4.3.3.	Distribución <i>t de student</i>	51
4.3.4.	Distribución F de Snedecor	51
4.4.	Estadísticos de orden	51
4.4.1.	Función de distribución marginal de los e.o.	52
4.4.2.	Distribución conjunta de los e.o.	53

5. Unidad 5. Momentos de variables aleatorias.	55
5.0.1. Integral de Riemann	55
5.0.2. Integral de Riemann - Stieltjes	56
5.0.3. Propiedades de la Esperanza	59
5.1. Momentos de las variables aleatorias.	60
5.1.1. Propiedades de la Varianza	61
5.2. Función Característica	61
5.2.1. Números complejos	62
5.2.2. Propiedades de la función característica	64
5.3. Momentos condicionados	64
5.3.1. Aplicación. Parámetros aleatorios.	68
6. Unidad 6. Teoría asintótica	71
6.1. Modos de convergencia	72
6.1.1. Convergencia casi cierta	72
6.1.2. Convergencia en Probabilidad	72
6.1.3. Convergencia en media cuadrática	73
6.1.4. Convergencia en distribución.	73
6.2. Distribución asintótica	74
6.3. Dos desigualdades útiles en teoría asintótica	76
6.3.1. Desigualdad de Markov	76
6.3.2. Desigualdad de Tchebycheff	76
6.4. Ley débil de los Grandes Números (varianza finita)	76
6.5. Unicidad	76
6.6. Relaciones entre los modos de convergencia	77
6.7. Teorema de Helly - Bray	77
6.8. Teorema de continuidad de Lévy	78
6.9. Ley débil de los grandes números	78
6.10. Teorema Central del Límite	79
6.11. Teorema de Slutsky	80
6.12. Método delta	80
7. Unidad 7. Introducción a la Teoría de la Estimación Puntual	82
7.1. Familias paramétricas de funciones de densidad	82
7.1.1. Modelo paramétrico	82

7.2.	Estimación puntual	83
7.2.1.	Estadístico	83
7.3.	Error cuadrático medio	83
7.3.1.	Sesgo	84
7.3.2.	Eficiencia relativa	84
7.4.	Estimadores EIO o IMVU	84
7.5.	Estadísticos suficientes	85
7.5.1.	Criterio de Factorización de Neyman	85
7.5.2.	Suficiencia minimal	86
7.6.	Teorema de Rao - Blackwell	86
7.7.	Complejitud	90
7.8.	Teorema de Lehmann - Scheffe sobre IMVU	91
7.9.	Propiedades asintóticas	91
7.9.1.	Consistencia	92
7.9.2.	Estimadores CAN	92
8.	Unidad 8. Métodos de estimación.	93
8.1.	Método de Máxima Verosimilitud	93
8.1.1.	Función de verosimilitud	93
8.1.2.	Estimador máximo verosímil. (EMV)	94
8.1.3.	Cómputo del EMV	94
8.1.4.	Propiedades de los EMV	97
8.1.5.	Distribución asintótica del EMV	97
8.2.	Eficiencia asintótica: Cota de Rao - Cramer	100

Prefacio

Este resumen ha sido escrito en su totalidad por Ana Paula Satorres y -desinteresadamente- transcripto a L^AT_EX por Agustín Cugno a partir de las bibliografía y apuntes vigentes al año 2017.

Cada referencia en negrita con números romanos entre paréntesis es una demostración que se puede ver en el resumen de demostraciones.

Capítulo 1

Unidad 1. Espacios de probabilidad

1.1. Espacio muestral

Se define como Ω a todos los resultados posibles del experimento. Este conjunto Ω se denomina *espacio muestral*. Ej: tirar un dado $\rightarrow \Omega = \{1, 2, 3, 4, 5, 6\}$

1.2. Eventos Aleatorios

Un *evento aleatorio* es un subconjunto de Ω . Ej: $A =$ el número es par: $A = \{2, 4, 6\}$. Si simbolizamos $\omega \in \Omega$ cómo el resultado del experimento y ocurre A , ej. $\omega = 2$, ω es favorable a A . Solo son aleatorios los eventos en \mathcal{A}

1.3. Álgebra

Una Álgebra (\mathcal{A}) es una colección no vacía de subconjuntos de Ω , que representa la colección de eventos a los que queremos asignarles probabilidad. La clase \mathcal{A} satisface:

- 1) $\Omega \in \mathcal{A}$
- 2) Si $A \in \mathcal{A} \rightarrow A^c \in \mathcal{A}$
- 3) Si $A \in \mathcal{A}$ y $B \in \mathcal{A} \rightarrow A \cup B \in \mathcal{A}$

1.4. σ -álgebra

Una σ -álgebra (\mathcal{F}) es una colección de subconjuntos de Ω que satisface:

- 1) $\Omega \in \mathcal{F}$
- 2) Si $A \in \mathcal{F} \rightarrow A^c \in \mathcal{F}$
- 3) Si $A_1, A_2, \dots, A_n \in \mathcal{F}$ entonces $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Es decir, es una colección de subconjuntos de Ω no vacía que es cerrada bajo las operaciones de tomar complemento y efectuar uniones infinitas numerables. Estas propiedades garantizan que también sea cerrada para todas las operaciones usuales entre conjuntos (unión, intersección, complemento, diferencia, diferencia simétrica). Es decir, al realizar estas operaciones se obtiene algo que $\in \mathcal{F}$.

(I) Otra definición de σ -álgebra: Sea \mathcal{F} una σ -álgebra, entonces:

- 1') $\emptyset \in \mathcal{F}$
- 2') Si $A_1, A_2, \dots, A_n \in \mathcal{F}$ entonces $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$
- 3') Si $A, B \in \mathcal{F}$ entonces $A - B \in \mathcal{F}$ y $A \triangle B \in \mathcal{F}$

Siempre existe una σ -álgebra de subconjuntos de Ω , cualquiera que esta sea, pero no única.

La σ -álgebra mas chica: $\{\emptyset, \Omega\}$

La σ -álgebra mas grande: $2^\Omega = \#(2^\Omega) = 2^{\#\Omega}$

Operaciones entre σ -álgebra.

Dada \mathcal{F}_1 y \mathcal{F}_2 , dos σ -álgebra de un mismo espacio muestral Ω :

- (II) - La $\mathcal{F}_1 \cap \mathcal{F}_2$ siempre es una σ -álgebra.
- La $\mathcal{F}_1 \cup \mathcal{F}_2$ no siempre es una σ -álgebra

1.5. σ -álgebra generada

Dado un conjunto Ω y una familia \mathcal{C} de subconjuntos de Ω , existe una σ -álgebra sobre Ω denotado por $\sigma(\mathcal{C})$, tal que:

- 1) $\mathcal{C} \subset \sigma(\mathcal{C})$
- 2) Si \mathcal{F} es σ -álgebra sobre Ω tal que $\mathcal{C} \subset \mathcal{F}$, entonces $\mathcal{C} \subset \sigma(\mathcal{F})$

Se dice entonces que $\sigma(\mathcal{C})$ es la σ -álgebra sobre Ω generada por \mathcal{C} . Si definimos

$\Sigma = \{\mathcal{F} : \mathcal{F} \text{ es una } \sigma\text{-álgebra sobre } \Omega \text{ y } \mathcal{C} \subset \mathcal{F}\}$

(III) Definimos a

$$\sigma(\mathcal{C}) \equiv \bigcap_{\mathcal{F} \in \Sigma} \mathcal{F}$$

como la intersección entre todas las σ -álgebras que contienen a \mathcal{C} . Se puede demostrar que $\sigma(\mathcal{C})$ es una σ -álgebra (3 axiomas), y es la más pequeña que contiene a \mathcal{C}

1.6. σ -álgebra de Borel

La σ -álgebra de Borel (\mathcal{B}) es la σ -álgebra sobre \mathbb{R} generada por los conjuntos de la forma $A_x = (-\infty, b]$, para todo $b \in \mathbb{R}$.

Un conjunto $B \in \mathcal{B}$ se denomina *boreliano*.

Es decir, $\mathcal{B} = \sigma\{(-\infty, b] : b \in \mathbb{R}\}$

(IV) Para cualesquiera números reales $a \leq b$, los intervalos $[a, b]$, (a, ∞) , $(-\infty, b)$, $[a, b)$, $(a, b]$, (a, b) , $\{b\}$ son todos elementos de $\mathcal{B}(\mathbb{R})$

(V) Además, se puede demostrar que las siguientes σ -álgebras son todas idénticas a \mathcal{B} :

- | | |
|----------------------------------|--|
| 1) $\sigma\{(a, b) : a \leq b\}$ | 4) $\sigma\{[a, b] : a \leq b\}$ |
| 2) $\sigma\{(a, b] : a \leq b\}$ | 5) $\sigma\{(a, \infty] : a \in \mathbb{R}\}$ |
| 3) $\sigma\{[a, b) : a \leq b\}$ | 6) $\sigma\{(-\infty, b] : b \in \mathbb{R}\}$ |

Proposiciones sobre las σ -álgebras generadas

- 1) Sean \mathcal{C}_1 y \mathcal{C}_2 dos colecciones de subconjuntos de Ω tales que $\mathcal{C}_1 \subseteq \mathcal{C}_2$. Entonces $\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$
- 2) Si \mathcal{F} es una σ -álgebra, entonces $\sigma(\mathcal{F}) = \mathcal{F}$

1.7. Probabilidad finitamente aditiva

Una Probabilidad finitamente aditiva es una función P definida sobre cierta σ -álgebra \mathcal{F} de subconjuntos de Ω que satisface (Axiomas de Kolmogorov):

- 1) Si $A \in \mathcal{F} \rightarrow P(A) \geq 0$
- 2) $P(\Omega) = 1$
- 3) Si $A_1, \dots, A_n \in \mathcal{F}$ son disjuntos ($A_j \cap A_k = \emptyset$ para $i \neq j$) entonces

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

Si al axioma 3 se lo reemplaza por:

3') Si $A_1, \dots, A_n \in \mathcal{F}$ son disjuntos ($A_j \cap A_k = \emptyset$ para $i \neq j$) entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \rightarrow \text{se denomina } \sigma\text{-aditividad}$$

1.8. Medida de Probabilidad

Una medida de Probabilidad es una función P definida sobre cierta σ -álgebra \mathcal{F} de subconjuntos de Ω que satisface los axiomas 1, 2 y 3'. ($P : \mathcal{F} \rightarrow [0, 1]$)

1.9. Espacio de Probabilidad

Un Espacio de Probabilidad es una terna (Ω, \mathcal{F}, P) donde:

- 1) Ω es un conjunto no vacío.
- 2) \mathcal{F} es una σ -álgebra de subconjuntos de Ω
- 3) P es una medida de probabilidad cuyo dominio es \mathcal{F}

1.9.1. Propiedades de P

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

(VII) 2) $P(A^c) = 1 - P(A)$

3) Si $A \subseteq B$, entonces $P(B - A) = P(B) - P(A)$

(VIII) 4) Si $A \subseteq B$, entonces $P(A) \leq P(B) \rightarrow P$ es una función monótona.

(IX) 5) $0 \leq P(A) \leq 1$

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

(X) 7) $P(A \cup B) \leq P(A) + P(B) \rightarrow$ (XI) $P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$ (subaditividad finita)

(XII) 8) $P(A) + P(B) \leq 2$

(XIII) 9) $P(A \cap B) \leq P(B)$

1.10. Sucesiones de eventos

Sea $\{A_n\} : n \in \mathbb{N}$ una sucesión monótona de eventos:

- 1) Si $A_1 \subseteq A_2 \subseteq \dots$ (creciente), entonces $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$
- 2) Si $A_1 \supseteq A_2 \supseteq \dots$ (decreciente), entonces $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$

1.11. Continuidad de la Probabilidad

Las medidas de Probabilidad son funciones continuas, y por lo tanto:

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\lim_{n \rightarrow \infty} A_n\right)$$

1.11.1. Continuidad en el vacío

Si tenemos $\{A_n\}_{n \geq 1}$ una sucesión que decrece hacia el vacío ($A_n \downarrow \emptyset$), tal que $A_n \supset A_{n+1} \forall n \in \mathbb{N}$ y $\bigcap_{n \geq 1} A_n = \emptyset$, entonces $P(A_n) \rightarrow 0$ cuando $n \rightarrow \infty$. Esto es así porque al ser P una función continua, $\lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n)$ y dado que $\lim_{n \rightarrow \infty} A_n = \emptyset$, entonces: **(XIV)**

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\lim_{n \rightarrow \infty} A_n\right) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(\emptyset) = 0$$

(LXXIII) Propiedad. Si $A_n \downarrow A$, entonces $P(A_n) \downarrow P(A)$

1.12. Probabilidad condicional

Sea (Ω, \mathcal{F}, P) un espacio de probabilidad y sean A y B dos eventos tal que $P(B) > 0$, la probabilidad *condicional* de A dado B se define cómo:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \forall A, B \in \mathcal{F}$$

Cuando $P(B) = 0$, la probabilidad condicional de A dado B es indefinida, o puede ser arbitrariamente definida.

(XV) Si $B \in \mathcal{F} : P(B) > 0$, $P(\cdot|B)$ es una medida de probabilidad ya que satisface los 3 axiomas de Kolmogorov:

$$1) P(A|B) \geq 0 \quad \forall A \in \mathcal{F}$$

$$2) P(\Omega|B) = 1$$

$$3) \text{ Si } A_1, A_2, \dots, A_n \in \mathcal{F} \text{ tal que los conjuntos son disjuntos, entonces: } P\left(\bigcup_{i=1}^{\infty} A_i | B\right) = \sum_{i=1}^{\infty} P(A_i | B)$$

De la definición de probabilidad condicional se obtiene que:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A), \text{ valida también cuando } P(B) = 0$$

1.13. Teorema de la Probabilidad Compuesta

Sea (Ω, \mathcal{F}, P) un espacio de probabilidad:

$$i) P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \quad \forall A, B \in \mathcal{F}$$

$$(XVI) \quad P\left(\bigcap_{i=1}^n A_i\right) = P(A_i | \bigcap_{n=k}^{i-1} A_k) \quad \forall A_1, A_2, \dots, A_n \in \mathcal{F}$$

1.13.1. Partición del espacio muestral

$A_1, A_2, \dots \in \mathcal{F}$ conforman una partición del espacio muestral si y solo si son mutuamente excluyentes y $\cup A_i = \Omega$

Suponemos una secuencia de A_i finita o numerable.

$$\forall B \in \mathcal{F} \text{ se tiene que } B = \bigcup_{i \in I} (A_i \cap B)$$

$$\text{Esto es así porque } B \subset \Omega, \quad B = \Omega \cap B = \left(\bigcup_{i \in I} A_i\right) \cap B = \bigcup_{i \in I} (A_i \cap B)$$

Entonces $P(B) = \sum P(A_i \cap B) = \sum P(A_i)P(B|A_i) \rightarrow$ Teorema de la Probabilidad Total.

De aquí, se deriva el *Teorema de Bayes*, que permite calcular $P(A_i|B)$, a partir de $P(B|A_i)$ y de $P(A_i)$

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)} \quad \rightarrow \quad \text{Fórmula de Bayes}$$

1.14. Independencia estocástica.

Dos eventos A y B son estocásticamente independientes si y solo si

$$P(A \cap B) = P(A)P(B) \rightarrow \text{independientes 2 a 2}$$

Los eventos independientes son aquellos para los cuales la probabilidad condicional es igual a la no condicional. Es decir, el hecho de que haya ocurrido uno de los eventos no modifica la probabilidad de ocurrencia del otro.

$$P(B|A) = P(B) \text{ dado que } P(B|A) = \frac{P(B \cap A)}{P(A)} \rightarrow P(A \cap B) = P(A)P(B)$$

Propiedades

-Si $A \cap B = \emptyset$ no son independientes, porque si ocurre uno, sé que el otro no.

(XVII) - Si $P(A) = 0$ entonces A es independiente cualquier otro evento.

(XVIII) - Si A y B son independientes, A y B^c , A^c y B, A^c y B^c también lo son.

1.14.1. Independencia colectiva

Los eventos $A_1, \dots, A_n \in \mathcal{F}$ son estocásticamente independientes de manera colectiva si cumplen:

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j) \quad i \neq j \quad \text{Independencia de a pares} \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k) \quad i \neq j \neq k \\ &\vdots \\ P(A_1 \cap \dots \cap A_n) &= P(A_1)P(A_2)\dots P(A_n) \end{aligned}$$

Es decir, una colección infinita de eventos es independiente si cualquier subcolección finita lo es.

Capítulo 2

Unidad 2. Variables Aleatorias.

2.1. Variable Aleatoria

Una *variable aleatoria* representa una traducción de cada uno de los elementos del espacio muestral en número reales. Mediante una v.a. uno puede considerar que el posible experimento aleatorio en cuestión no produce como resultado elementos de Ω , sino de \mathbb{R} .

Una variable aleatoria en un espacio de probabilidad (Ω, \mathcal{F}, P) es una función real definida en Ω , tal que $[X \leq x]$ es un evento aleatorio para todo $x \in \mathbb{R}$. Es decir:

$$X : \Omega \rightarrow \mathbb{R} \text{ es v.a. sii } [X \leq x] \in \mathcal{F} \quad \forall x \in \mathbb{R}$$

Decir $[X \leq x]$ es igual que decir $X^{-1}B$ o $X \in B$, siendo B el boreliano $(-\infty, x]$. Por lo tanto, una definición equivalente de v.a. es:

$$X : \Omega \rightarrow \mathbb{R} : X^{-1}(B) \in \mathcal{F} \quad \forall B = (-\infty, x]$$

siendo $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$

Esta condición se conoce como *medibilidad*.

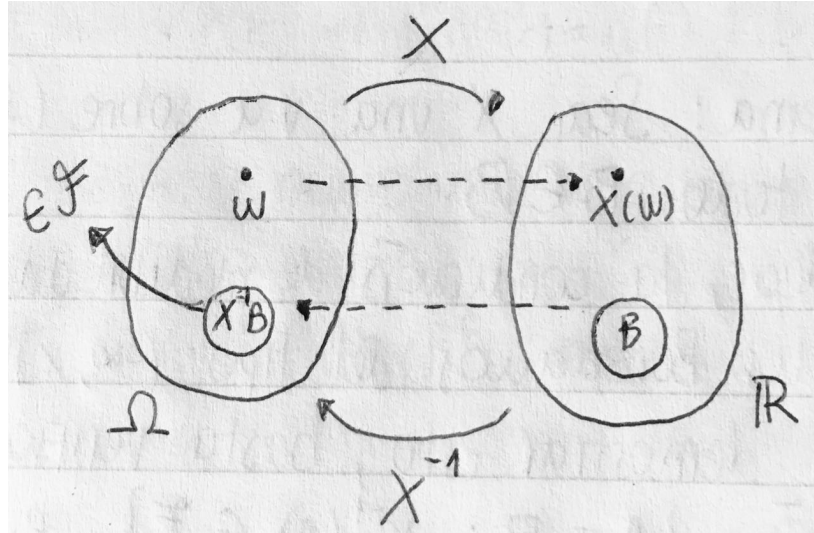
Teorema 1 Sea X una v.a. sobre (Ω, \mathcal{F}, P) , entonces vale que $X^{-1}(B) \in \mathcal{F}$ para todo $B \in \mathcal{B}$.

Es decir, la condición de medibilidad puede ser comprobada no solo con los borelianos del tipo $(-\infty, x]$, sino para cualquier boreliano. (XIX)

¿Por qué en la definición de v.a. se le pide a $X : \Omega \rightarrow \mathbb{R}$ esa condición de medibilidad?

Recordemos que P es una medida de probabilidad definida sobre el *espacio medible* (Ω, \mathcal{F}) . Si X es una variable aleatoria, entonces podemos trasladar la medida de probabilidad al espacio medible $(\mathbb{R}, \mathcal{B})$ del siguiente modo: Si B es un boreliano, definimos $P_X(B) = P(X^{-1}B)$, lo cual es posible pues el conjunto $X^{-1}B$ es un elemento de \mathcal{F} , dominio de definición de P .

(XXVI) Se puede probar que la función $P_X : \mathcal{B} \rightarrow [0, 1]$ resulta ser una medida de probabilidad y es llamada *medida de probabilidad inducida por la variable aleatoria*. En resumen:



(XX) Ejemplo de v.a. $\rightarrow X = c$ (función constante)

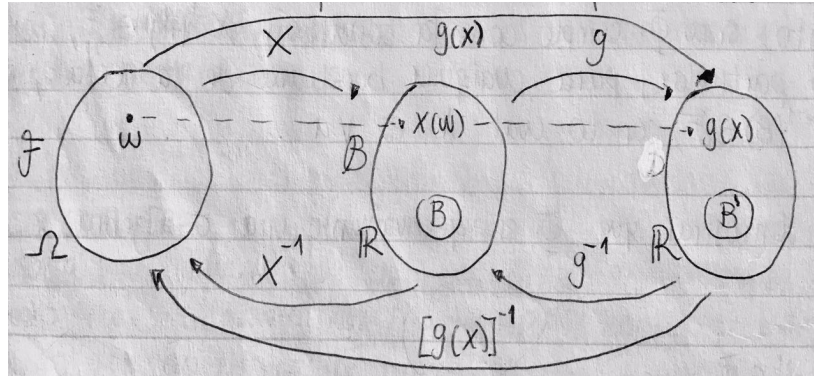
2.1.1. Función medible

Una función $g : \mathbb{R} \rightarrow \mathbb{R}$ que satisface:

$$g^{-1}((-\infty, x]) \in \mathcal{B} \quad \forall x \in \mathbb{R}$$

se denomina función *medible Borel*

En el marco de un espacio de probabilidad con $\Omega = \mathbb{R}$ y $\mathcal{F} = \mathcal{B}$, decir que g es medible Borel equivaldría a decir que $g(X)$ es una variable aleatoria. Es decir, el requisito es el mismo pero en distintos espacios.



Si $X^{-1}(B) \in \mathcal{F} \rightarrow X$ es variable aleatoria.

Si $g^{-1}(B) \in \mathcal{B} \rightarrow g$ es medible borel.

Si $g : \mathbb{R} \rightarrow \mathbb{R}$ es medible y $X : \Omega \rightarrow \mathbb{R}$ es v.a. entonces $g(X) : \Omega \rightarrow \mathbb{R}$ también es una variable aleatoria.

Ejemplos de funciones medibles:

- Funciones monótonas
- Si B es boreliano, su función indicadora (1_B)
- Funciones continuas
- etc.

2.2. Función de Distribución.

Sea X una v.a. Se define la *Función de Distribución* asociada a X como la función $F_X : \mathbb{R} \rightarrow [0, 1]$ dada por:

$$F_X(x) \equiv P_X((-\infty, x]) \equiv P(X^{-1}((-\infty, x]) \equiv P(X \in (-\infty, x]) \equiv P[X \leq x]$$

La función de distribución queda definida por la medida de probabilidad inducida por la v.a. También se la conoce como función de probabilidad acumulada.

2.2.1. Propiedades de la Función de Distribución

(XXI) Sea X una v.a. definida sobre (Ω, \mathcal{F}, P) y sea F_X su función de distribución, entonces se verifica:

- 1) F_X es monótona no decreciente, es decir: Si $x_1 < x_2$ entonces: $F_X(x_1) \leq F_X(x_2)$
- 2) $\lim_{x \rightarrow \infty} F_X(x) = 1$

3) $\lim_{x \rightarrow -\infty} F_X(x) = 0$

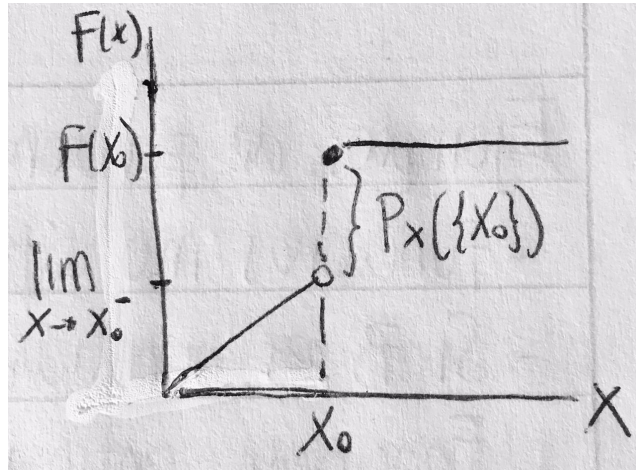
4) $F_X(x)$ es continua a derecha para todo $x \in \mathbb{R}$, es decir: $F_X(x^+) = F_X(x)$

Si una función $F: \mathbb{R} \rightarrow [0, 1]$ satisface las cuatro propiedades anteriores, es llamada función de distribución y existe una v.a. X tal que $F = F_X$. Una función de distribución puede corresponder a varias v.a. en el mismo espacio de probabilidad. (Ω, \mathcal{F}, P)

De las propiedades se deduce que para todo $x_0 \in \mathbb{R}$ se tiene que:

$$\lim_{x \rightarrow x_0^-} F(x) = F(x_0) - P_X(\{x_0\}) = P(X \leq x_0) - P(X = x_0)$$

Si F_X es continua a izquierda $\rightarrow P_X(\{x_0\}) = 0$



2.3. Tipos de variables aleatorias

2.3.1. Variables aleatorias discretas

La v.a. X es *discreta* si asume un número finito o numerable de valores, es decir, si existe un conjunto finito o numerable $\{x_1, x_2, \dots\} \subset \mathbb{R}$ tal que $X(\omega) \in \{x_1, x_2, \dots\}$ para todo $\omega \in \Omega$.

Su función de distribución F_X es una función constante por pedazos, donde x_1, x_2, \dots son los puntos de *discontinuidad* de F_X . En cada uno de estos puntos, el tamaño de la discontinuidad es $P(X = x_i) = F(x_i) - F(x_i^-) > 0$.

A la función $f(x)$ que indica estos incrementos se le llama *función de probabilidad* de $X \rightarrow p(x_i) = P(X = x_i)$, $i = 1, 2, \dots$

La función de distribución se reconstruye de la forma:

$$F_X(x) = \sum_{u \leq x} p(u)$$

2.3.2. Variable aleatoria absolutamente continua

La v.a. X es *absolutamente continua* si existe una función $f(x) \geq 0$, integrable Lebesgue, tal que para cada $x \in \mathbb{R}$ se cumple:

$$F_X(x) = \int_{-\infty}^x f(t) dt$$

La función f es la *función de densidad* de X y $F_X(x)$ su función de distribución. **(XXII)** $F_X(x)$ cumple las 4 propiedades de una función de distribución, por lo tanto, $f(x) \geq 0$ es densidad de una v.a. sii:

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Si X es absolutamente continua entonces $f(x) = F'_X(x) = \frac{\partial F(x)}{\partial x}$ c.t.p.

Además, si X es continua, $F(x)$ es continua.

¿Cómo sabemos si X tiene función de densidad?

X tiene densidad si F_X es:

1) Continua

2) Derivable por partes: F_X es derivable en el interior de un conjunto finito o numerable de intervalos cerrados cuya unión es \mathbb{R} .

Si se cumple 1 y 2, $F'_X(x) = f(x)$ en casi todo punto. (c.t.p.)

Si hay un punto de F_X en el cual la función no es derivable, $f(x)$ en ese punto se define arbitrariamente.

2.3.3. Variable aleatoria mixta

Son aquellas v.a. que no son discretas ni continuas, sino una mezcla de ambos tipos.

La función de distribución de una v.a. mixta se puede descomponer en dos partes $\longrightarrow F = F_d + F_{ac}$

2.4. Cálculo de probabilidades

¿Cómo calculamos $P[X \in B], B \in \mathcal{B}$? Primero hay que distinguir si X es v.a. discreta o continua.

2.4.1. V.a. discreta

Rango: Se define al rango como el conjunto imagen de la v.a., es decir, es el conjunto de puntos de discontinuidad de la función de distribución. $X : \Omega \rightarrow \mathbb{R}$ es v.a. discreta su rango (R_X):

$$R_X = \{x \in \mathbb{R} : P_X(\{x\}) > 0\} = \{x_1, x_2, \dots\}$$

Como consecuencia de la definición:

$$\text{(XXIII)} \quad P_X(R_X) = 1$$

$$\text{(XXIV)} \quad P(X \in B) = P_X(B) = P_X(B \cap R_X) \quad \forall B \in \mathcal{B}$$

$$P(X \in B) = \sum_{x \in B \cap R_X} p(x)$$

2.4.2. V.a. continua (con densidad)

El *soporte* de X se define como todos los valores de x para los cuales la densidad es positiva

$$S_X = \{x \in \mathbb{R} : f(x) > 0\}$$

Sea X una v.a. continua con densidad $f(x)$, entonces: **(XXV)**

$$P(X \in B) = P_X(B) = \int_{B \cap S_X} f(x) dx \quad \forall B \in \mathcal{B}$$

Capítulo 3

Unidad 3. Vectores aleatorios

Un *vector aleatorio* es un vector $\tilde{X} = (X_1, X_2, \dots, X_k)$ cuyos componentes son variables aleatorias definidas todas en el mismo espacio de probabilidad. Si \tilde{X} es un vector aleatorio de dimensión k , entonces puede ser interpretado como una función del tipo $\tilde{X} : \Omega \rightarrow \mathbb{R}^k$

Es decir, ahora se toma un $\omega \in \Omega$ y al aplicar $\tilde{X}(\omega)$, obtenemos: $\tilde{X}(\omega) = [X_1(\omega), X_2(\omega), \dots, X_k(\omega)]$, un vector con k características numéricas.

Cuando hablábamos de v.a., llegábamos a \mathbb{R} , cuya σ -álgebra asociada era \mathcal{B} , la engendrada por intervalos del tipo $[-\infty, x]$

Ahora, como llegamos a \mathbb{R}^k , ya no tenemos intervalos sino *hiper-rectángulos*, que surgen del producto cartesiano de distintos intervalos y que generan la nueva σ -álgebra asociada a $\mathbb{R}^k \implies \mathcal{B}^k$

Es decir, si \tilde{X} tiene k dimensiones: $\times_{i=1}^k (-\infty, x_i] \implies$ hiper-rectángulos, y $\mathcal{B}^k = \sigma\{\times_{i=1}^k (-\infty, x_i]\}$

Teorema 2 Sea \tilde{X} un vector aleatorio, $\tilde{X}(\omega) = [X_1(\omega), X_2(\omega), \dots, X_k(\omega)]$, para todo $\tilde{x} = (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ se tendrá que: **(XXVII)**

$$\begin{aligned}\tilde{X}^{-1}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]) &\in \mathcal{F} \text{ o} \\ \tilde{X}^{-1}(\times_{i=1}^k (-\infty, x_i]) &\in \mathcal{F} \text{ o} \\ \tilde{X}^{-1}(B) &\in \mathcal{F}, \forall B \in \mathcal{B}^k \text{ o} \\ [\tilde{X} \in B] &\in \mathcal{F}, \forall B \in \mathcal{B}^k\end{aligned}$$

Este requisito técnico es necesario para poder trasladar la medida de probabilidad $P_{\tilde{X}}$ al espacio de probabilidad original.

Dado un espacio de probabilidad (Ω, \mathcal{F}, P) y un vector aleatorio $\tilde{X} = (X_1, X_2, \dots, X_k)$ se puede definir un nuevo espacio de probabilidad $(\mathbb{R}^k, \mathcal{B}^k, P_{\tilde{X}})$ donde dado $B \in \mathcal{B}^k$, se define:

$$P_{\tilde{X}}(B) = P(\tilde{X}^{-1}(B)) = P[\tilde{X} \in B]$$

Entonces, $(\Omega, \mathcal{F}, P) \xrightarrow{\tilde{X}} (\mathbb{R}^k, \mathcal{B}^k, P_{\tilde{X}})$

3.1. Función de vectores aleatorios

3.1.1. Función medible

Una función $g : \mathbb{R}^k \rightarrow \mathbb{R}$ es medible Borel si para todo $x \in \mathbb{R}$

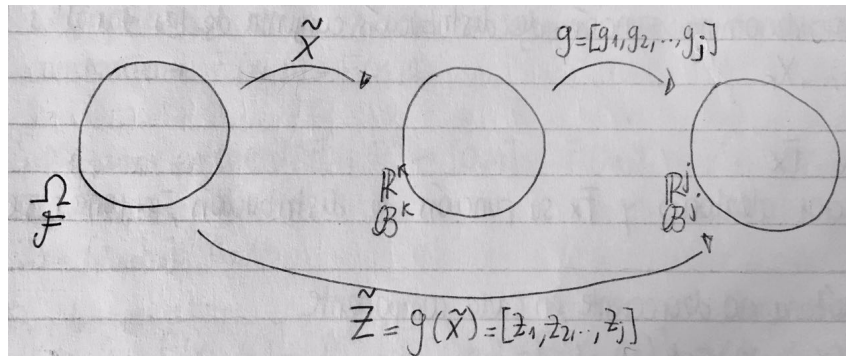
$$g^{-1}((-\infty, x]) \in \mathcal{B}^k$$

Una función medible, entonces, puede interpretarse como una v.a. en el espacio $(\mathbb{R}^k, \mathcal{B}^k)$, ya que el requisito técnico es el mismo. En particular, si g es continua, entonces es medible.

Teorema 3 Sea $\tilde{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio sobre un espacio de probabilidad (Ω, \mathcal{F}, P) y $g : \mathbb{R}^k \rightarrow \mathbb{R}$ una función medible, entonces $g(\tilde{X}) : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria.

La función g no sólo puede llevar elementos de \mathbb{R}^k a \mathbb{R} , sino que también puede ir de \mathbb{R}^k a \mathbb{R}^j , con lo cual g debería ser un vector con j funciones, que va a ser *vectorial medible* si las j funciones son medibles, con lo cual el teorema anterior se puede generalizar

Teorema 4 Sea $\tilde{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio sobre (Ω, \mathcal{F}, P) y $g : \mathbb{R}^k \rightarrow \mathbb{R}^j$ una función vectorial medible, entonces $\tilde{Z} = g(\tilde{X})$ es un vector aleatorio, donde $\tilde{Z} = (Z_1, Z_2, \dots, Z_j)$



Si $\tilde{X}^{-1}(B) \in \mathcal{F} \quad \forall B \in \mathcal{B}^k \implies \tilde{X}$ es vector aleatorio.

Si $g_i^{-1}(B) \in \mathcal{B}^k \quad \forall B \in \mathcal{B}^j \implies g$ es vectorial medible. (cada g_i es medible borel)

Si $\tilde{X} : \Omega \rightarrow \mathbb{R}^k$ es vector aleatorio y $g : \mathbb{R}^k \rightarrow \mathbb{R}^j$ es vectorial medible, entonces $\tilde{Z} = g(\tilde{X}) : \Omega \rightarrow \mathbb{R}^j$ es vector aleatorio.

3.2. Función de Distribución de un vector aleatorio

Sea $\tilde{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio, se define la función de distribución asociada a \tilde{X} como la función $F_X : \mathbb{R}^k \rightarrow [0, 1]$ por:

$$\begin{aligned}
 F_{\tilde{X}}(X_1, X_2, \dots, X_k) &= P_{\tilde{X}}(\times_{i=1}^k (-\infty, x_i]) \\
 &= P_{\tilde{X}}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]) \\
 &= P(\tilde{X}^{-1}(\times_{i=1}^k (-\infty, x_i])) \\
 &= P[\tilde{X} \in (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k)] \\
 &= P\left(\bigcap_{i=1}^k \{\omega \in \Omega : X_i(\omega) \leq x_i\}\right) \\
 &= P(\{\omega \in \Omega : X_1(\omega) \in (-\infty, x_1], \dots, X_k(\omega) \in (-\infty, x_k]\}) \\
 &= P\left(\bigcap_{i=1}^k [X_i \in (-\infty, x_i]]\right) \\
 &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_k \leq x_k)
 \end{aligned}$$

F_X también se denomina función de distribución *conjunta* de las variables aleatorias X_1, \dots, X_k

3.2.1. Propiedades de la función de distribución de un vector aleatorio

Sea \tilde{X} un vector aleatorio y F_X su función de distribución. Entonces, se verifica:

(XXVIII) 1) F_X es monótona no decreciente en cada componente.

(XXIX) 2) $\lim_{\forall i \ x_i \rightarrow \infty} F_X(x_1, \dots, x_k) = 1$

(XXX) 3) $\lim_{x_i \rightarrow -\infty} F_X(x_1, \dots, x_k) = 0$

4) F_X es continua a derecha en todo punto de \mathbb{R}^k (y en cada componente)

5) $\Delta_{I_1}^1 \dots \Delta_{I_k}^k F_X(x_1, \dots, x_k) \geq 0 \quad \forall I_j = (a_j, b_j), a_j < b_j, j = 1, \dots, k$

La propiedad 5) se traduce en solicitar que la probabilidad de que el vector $\tilde{X} = (X_1, \dots, X_k)$ tome valores en el hiper-rectángulo $(a_1, b_1) \times (a_2, b_2) \times \dots \times (a_k, b_k)$ sea no negativo, que es lo mismo que la probabilidad del evento $(a_1 < X_1 \leq b_1, \dots, a_k < X_k \leq b_k)$.

Cuando hablábamos de v.a., esta propiedad era consecuencia de las otras cuatro y por lo tanto no se hacía explícita: $P(a < X \leq b) = F_X(b) - F_X(a) \geq 0$. Pero ahora si es necesaria su explicitación y cumplimiento al definir F_X

El operador diferencia (Δ) actúa de la siguiente manera:

Supongamos una $F_X : \mathbb{R} \rightarrow \mathbb{R}$

$\Delta_{(a,b]} F(x) = F(b) - F(a)$, es decir, es la diferencia de la función valuada en el extremo superior y el inferior.

Si ahora tenemos $F_{(X_1, X_2)} : \mathbb{R}^2 \rightarrow \mathbb{R}$, debemos indicarle a Δ sobre qué argumento va a trabajar.

$\Delta_{(a,b]}^1 F(x_1, x_2)$ calcula diferencias sobre el primer argumento dejando el segundo constante $\rightarrow F(b, x_1) - F(a, x_1)$.

$\Delta_{(a,b]}^2 F(x_1, x_2)$ calcula diferencias sobre el segundo argumento dejando el primero constante $\rightarrow F(x_1, b) - F(x_1, a)$.

Supongamos entonces $F_X : \mathbb{R}^2 \rightarrow [0, 1] = F(x_1, x_2)$

$I_1 = (a_1, b_1) \quad a_1 < b_1$

$I_2 = (a_2, b_2) \quad a_2 < b_2$

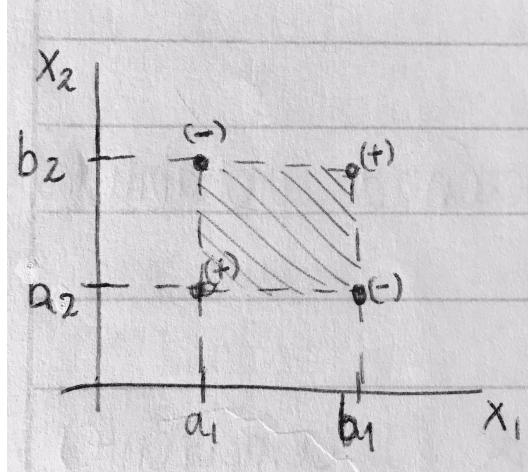
$$\Delta_{I_1}^1 [\Delta_{I_2}^2 F(x_1, x_2)] = \Delta_{I_1}^1 [F(x_1, b_2) - F(x_1, a_2)]$$

Como ahora tengo una F de x_1 , aplico Δ a x_1

$$= [F(b_1, b_2) - F(b_1, a_2)] - [F(a_1, b_2) - F(a_1, a_2)]$$

$$= F(a_1, a_2) + F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2)$$

y eso debe ser ≥ 0



Teorema 5 Si $F : \mathbb{R}^k \rightarrow [0, 1]$ que satisface las 5 propiedades, entonces existe un vector aleatorio \tilde{X} tal que $F_X = F$

Es decir, se garantiza la existencia de un espacio de probabilidad (Ω, \mathcal{F}, P) en donde se encuentra un vector \tilde{X} cuya función de distribución es esa F .

3.2.2. Función de distribución marginal

Dado un vector aleatorio $\tilde{X} = (X_1, \dots, X_k)$ con función de distribución $F_{\tilde{X}}$, se puede encontrar la función de distribución marginal de la variable X_i , haciendo que el resto de las variables tiendan a infinito. Es decir **(XXIX)**:

$$F_{X_i}(x_i) = \lim_{\forall x_j \rightarrow \infty} F_{\tilde{X}}(\tilde{x}) \quad i \neq j$$

De esta forma también se puede encontrar la función de distribución de un sub-vector a partir de la $F_{\tilde{X}}$ del vector original, haciendo tender a infinito las variables que no forman parte de ese subvector.

3.3. Vector aleatorio discreto

Sea $\tilde{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio, se dice que es discreto sii para cada $i = 1, 2, \dots, k$, X_i es una variable aleatoria discreta, todas definidas en el mismo (Ω, \mathcal{F}, P)

Esto implica que cada una de las X_i tenga un rango que sea un conjunto finito o infinito numerable.

La función $p_X : \mathbb{R}^k \rightarrow [0, 1]$ definida por:

$$p_{\tilde{X}}(\tilde{x}) = P_{\tilde{X}}(\{\tilde{x}\}) = P(\tilde{X} = \tilde{x})$$

(siendo $\tilde{x} = \{x_1, x_2, \dots, x_k\} : X_i = x_i$)

se denomina función de probabilidad de \tilde{X} o función de probabilidad conjunta de las variables X_i

Dado que en un vector aleatorio discreto cada una de sus componentes es una v.a. discreta, podemos hablar del rango de la v.a. discreta que ocupa la posición i en el vector, siendo R_{X_i} :

$$R_{X_i} = \{x_i \in \mathbb{R} : P_{X_i}(x_i) > 0\}$$

En base a cada uno de los rangos de las X_i variables, podemos definir un conjunto $R_{\tilde{X}}^* = R_{X_1} \times \dots \times R_{X_k}$. Dado que cada R_{X_i} es un subconjunto de \mathbb{R} entonces $R_{\tilde{X}}^* \subset \mathbb{R}^k$

Dado que la cardinalidad de cada R_{X_i} es finita o numerable, la cardinalidad de $R_{\tilde{X}}^*$ también lo es, y va a ser el producto cartesiano de la cardinalidad de cada uno de los R_{X_i}

Por definición de rango sabemos que $P_{X_i}(R_{X_i}) = 1$. Si ahora aplicamos la medida de probabilidad inducida por el vector aleatorio a $R_{\tilde{X}}^*$ también es igual a 1. $\rightarrow P_{\tilde{X}}(R_{\tilde{X}}^*) = 1$

Esto no implica que $R_{\tilde{X}}^*$ sea el rango del vector aleatorio, pero $R_{\tilde{X}} \subset R_{\tilde{X}}^*$

Se define $R_{\tilde{X}}^*$ para calcular probabilidades porque es más fácil que hacerlo sobre $R_{\tilde{X}}$

Un ejemplo en \mathbb{R}^2

$$\tilde{X} = (X_1, X_2)$$

$$P(\tilde{X} = (0, 0)) = P(\tilde{X} = (1, 1)) = 1/2$$

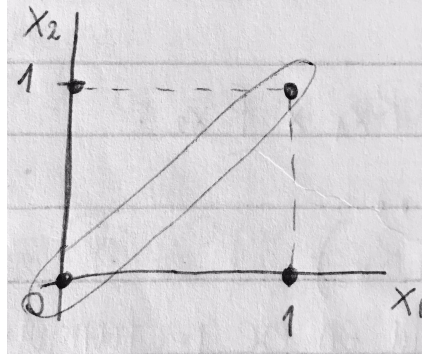
$$R_{X_1} = \{0, 1\} = R_{X_2}$$

$$R_{\tilde{X}} = \{(0, 0), (1, 1)\}$$

$$R_{\tilde{X}}^* = R_{X_1} \times R_{X_2} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

(XXXI) Claramente $R_{\tilde{X}} \subset R_{\tilde{X}}^*$

Gráficamente:



$R_{\tilde{X}}^*$ son los cuatro puntos que forman la retícula. $R_{\tilde{X}}$ es únicamente $\{(0, 0), (1, 1)\}$. Para calcular probabilidades se debe realizar un doble sumatorio en este caso (\mathbb{R}^2) y gráficamente se puede asimilar moviéndose sobre una retícula rectangular. Esa es la ventaja de $R_{\tilde{X}}^*$. Dado que la probabilidad en los puntos que están en $R_{\tilde{X}}^*$ pero no en $R_{\tilde{X}}$ es cero, el cálculo de probabilidad no se ve afectado por esos puntos que sobran.

Entonces, de la definición de $R_{\tilde{X}}$ se deduce que:

$$p_X(x) = \begin{cases} > 0 & \text{si } x \in R_X \\ 0 & \text{si } x \notin R_X \end{cases}$$

De manera similar al caso de una v.a., para todo $B \in \mathcal{B}^k$, se tiene que:

$$P(\tilde{X} \in B) = P_{\tilde{X}}(B) = \sum_{x \in (B \cap R_X)} p_{\tilde{X}}(\tilde{x})$$

Dado que:

$$B = B_1 \times B_2 \times \dots \times B_k \subset \mathcal{B}^k \quad \text{y}$$

$$R_{\tilde{X}}^* = R_{X_1} \times R_{X_2} \times \dots \times R_{X_k}$$

$$\begin{aligned} B \cap R_{\tilde{X}}^* &= (B_1 \times \dots \times B_k) \cap (R_{X_1} \times \dots \times R_{X_k}) \\ &= (B_1 \cap R_{X_1}) \times \dots \times (B_k \cap R_{X_k}) \end{aligned}$$

Por lo tanto:

$$P(\tilde{X} \in B) = P_{\tilde{X}}(B \cap R_{\tilde{X}}) = \sum_{X_1 \in (B_1 \cap R_{X_1})} \dots \sum_{X_k \in (B_k \cap R_{X_k})} p_{\tilde{X}}(x_1, \dots, x_k)$$

3.3.1. Función de probabilidad conjunta de un subvector

Dado $\tilde{X} = (X_1, X_2, \dots, X_h, X_{h+1}, \dots, X_k)$ con $P_{\tilde{X}}(\tilde{x})$ conocida, y dado $\tilde{X}' = (X_1, X_2, \dots, X_h)$, se quiere obtener su función de probabilidad, $p(\tilde{x}')$

$$\begin{aligned} p(\tilde{x}') &= P(\tilde{X}' = \tilde{x}') \\ &= P(\{\omega \in \Omega : X_1(\omega) = x_1, \dots, X_h(\omega) = x_h\}) \end{aligned}$$

Este conjunto no va cambiar si agregamos, sobre la misma notación, restricciones que no estén activas o sean redundantes (se cumplan siempre)

$$p(\tilde{x}') = P(\{\omega \in \Omega : X_1(\omega) = x_1, \dots, X_h(\omega) = x_h, X_{h+1}(\omega) \in \mathbb{R}, \dots, X_k(\omega) \in \mathbb{R}\})$$

Dado que las X_i son funciones que van de $\Omega \rightarrow \mathbb{R}$, siempre se va a cumplir que $X_i(\omega) \in \mathbb{R}$, y en particular para X_{h+1}, \dots, X_k . Por lo tanto este conjunto no cambia.

Esto se hace porque queremos expresar a $p(\tilde{x}')$ en terminos de $p(\tilde{x})$. Pero como $p(\tilde{x})$ involucra k -uplas y no a h -uplas, se debe establecer la equivalencia entre eventos en los que solo intervienen las h variables del vector X' , con eventos que involucren a todas las v.a. del vector original.

Por lo tanto, volviendo a la notación anterior, se tiene:

$$p_{X'}(x') = P(X_1 = x_1, \dots, X_h = x_h, X_{h+1} \in \mathbb{R}, \dots, X_k \in \mathbb{R})$$

Esto se puede expresar como un conjunto en términos del vector original:

$$p_{X'}(x') = P(\tilde{X} \in \underbrace{\{x_1\} \times \dots \times \{x_h\} \times \mathbb{R} \times \dots \times \mathbb{R}}_{B \in \mathcal{B}^k})$$

Para calcular probabilidades sobre un evento debemos calcular la intersección del evento con el rango del vector aleatorio, es decir $B \cap R_{\tilde{X}}$, y realizar el sumatorio de $p_{\tilde{X}}$ sobre esa intersección:

$$p_{X'}(x') = \sum_{B \cap R_X} p_X(X_1, X_2, \dots, X_k) = \sum_{B \cap R_{\tilde{X}}^*} p_X(X_1, X_2, \dots, X_k) \quad (1)$$

Sabemos que:

$$B \cap R_{\tilde{X}}^* = (B_1 \cap R_{X_1}) \times \dots \times (B_k \cap R_{X_k})$$

Dado que B_1, \dots, B_h son intervalos degenerados, la intersección va a ser el punto $\{x_1\}, \dots, \{x_h\}$, mientras que B_{h+1}, \dots, B_k son los reales, por lo que la intersección con el rango de cada variable va a ser ese rango: $R_{X_{h+1}}, \dots, R_{X_k}$.

El sumatorio (1) puede realizarse como:

$$p_{X'}(x') = \sum_{X_1 \in (B_1 \cap R_{X_1})} \dots \sum_{X_k \in (B_k \cap R_{X_k})} p_{\tilde{X}}(x_1, \dots, x_k)$$

Es decir, para calcular la función de probabilidad de un subvector a la partir de la función de probabilidad del vector original, se debe sumar las probabilidades de todas aquellas variables que no aparecen en el nuevo vector, sobre todo su rango.

Cuando trabajamos con la función de distribución, la operación relevante para "hacer desaparecer" variables era el límite. Ahora, cuando trabajamos con la función de probabilidad, la operación relevante es la sumatoria.

3.4. Vectores aleatorios absolutamente continuos

A diferencia de los vectores aleatorios discretos, la condición de que las X_i v.a. que conforman el vector sean continuas no es suficiente. Es necesaria la existencia de la función de densidad.

El vector aleatorio $\tilde{X} = (X_1, \dots, X_k)$ es absolutamente continuo si existe una función, integrable Lebesgue, $f : \mathbb{R}^k \rightarrow \mathbb{R} > 0$ tal que:

$$F_X(x_1, \dots, x_k) = \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_{k-1}} \dots \int_{-\infty}^{x_1} f(\mathbf{t}) d\mathbf{t}$$

$\forall (x_1, \dots, x_k) \in \mathbb{R}^k$, donde $\mathbf{t} = (t_1, \dots, t_k)$ y $d\mathbf{t} = (dt_1, \dots, dt_k)$ La función f se denomina función de densidad de \tilde{X} o función de densidad conjunta de las X_i

Para obtener la función de distribución de un vector aleatorio k -dimensional, se debe integrar k veces la densidad.

Para todo boreliano $\in \mathcal{B}^k$ se cumple que:

$$P(\tilde{X} \in B) = P_{\tilde{X}}(B) = \int_B \cdots \int f_{\tilde{X}}(x_1, \dots, x_k) dx_1, \dots, dx_k$$

como esa $f_{\tilde{X}}$ tiene una una indicadora sobre el soporte,

integro sobre $B \cap S_{\tilde{X}}$ y desaparece la indicadora

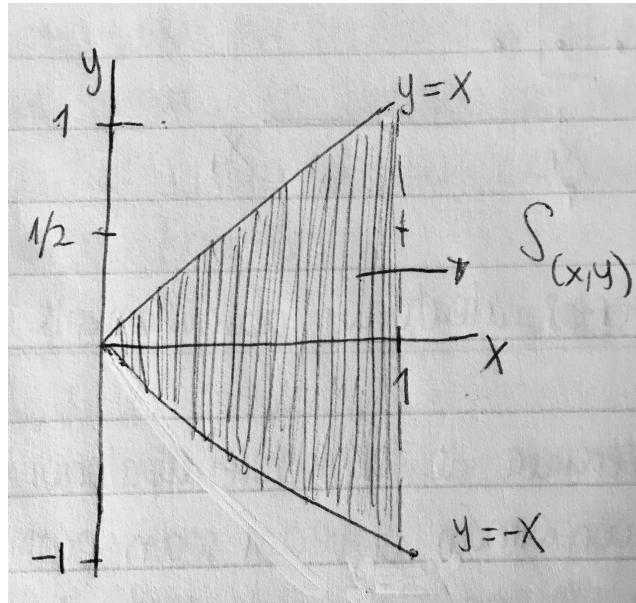
$$= \int_{B \cap S_{\tilde{X}}} \cdots \int f_{\tilde{X}}(x_1, \dots, x_k) dx_1, \dots, dx_k$$

Para resolver esa integral múltiple empiezo de adentro para afuera, y cada vez que integro con respecto a una variable X_i , la misma tiene que desaparecer. Así, cuando haya integrado k veces, no queda ninguna variable sino un número entre $[0, 1]$

Ejemplo. $\tilde{X} : \Omega \rightarrow \mathbb{R}^2$

$\tilde{X} = (X, Y)$ absolutamente continuo.

$$f_{(X,Y)}(x, y) = \begin{cases} 1 & 0 < |y| < x < 1 \\ 0 & \text{c.c.} \end{cases}$$



$$S_{(X,Y)} = \{(x, y) \in \mathbb{R}^2 : 0 < |y| < x < 1\}$$

Para que $f(x, y)$ sea densidad, la integral sobre todo el soporte debe ser

igual a 1. Lo cual se cumple en este caso:

$$\int_0^1 \int_{-x}^x dy \, dx = \int_0^1 y|_{-x}^x = \int_0^1 (x + x) dx = 2 \cdot \frac{x^2}{2} \Big|_0^1 = 1. \quad \checkmark$$

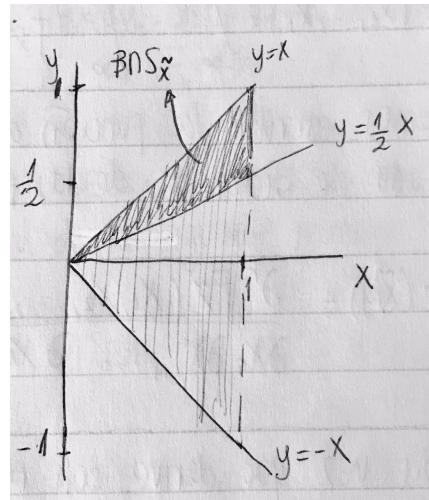
Para definir los límites de integración primero se debe elegir el orden de integración. Si primero integro y (integral de adentro), debo restringir esa variable (en este caso varía entre $-x$ y x), y en la segunda integral deajo que x se mueva libremente entre 0 y 1.

Para calcular probabilidades:

$$P(y \geq \tfrac{1}{2}x) = P(\underbrace{\{x, y\} \in \mathbb{R}^2 : y \geq \tfrac{1}{2}x}_{B \in \mathcal{B}^2})$$

Para llevar a cabo la integración debo realizar $B \cap S_{\tilde{X}}$

$$P(y \geq \tfrac{1}{2}x) = \int_0^1 \int_{(1/2)x}^x dy \, dx$$



De nuevo, primero restrinjo y , que como se ve gráficamente varía entre $\frac{1}{2}x$ y x , y deajo que x se mueva libremente entre 0 y 1.

$$\begin{aligned} P\left(y \geq \frac{1}{2}x\right) &= \int_0^1 y|_{(1/2)x}^x dx \\ &= \int_0^1 \left(x - \frac{1}{2}x\right) dx \\ &= \int_0^1 \frac{1}{2}x \, dx = \frac{1}{2} \frac{x^2}{2} \Big|_0^1 = \frac{1}{4} \end{aligned}$$

Teorema 6 Por Teorema Fundamental del Cálculo se da que:

$$\text{Sea } F(x) = \int_a^x g(t) \, dt, \text{ entonces } F'(x) = g(x)$$

Particularmente, para el caso de una v.a. continua, el teorema se aplica:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \rightarrow \frac{\partial F_X}{\partial x} = f_X(x)$$

Para $\tilde{X} = (X_1, \dots, X_k)$, un vector aleatorio absolutamente continuo, dado que:

$$F_{\tilde{X}}(x_1, \dots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f_{\tilde{X}}(t) dt_k \dots dt_1,$$

podemos encontrar la función de densidad del vector como la derivada parcial cruzada de orden k de la función de distribución:

$$f_{\tilde{X}}(\tilde{x}) = \frac{\partial^k F_{\tilde{X}}(x_1, \dots, x_k)}{\partial x_k \dots \partial x_1}$$

Cada vez que derive con respecto a una variable, va a desaparecer la integral y el diferencial asociados a esa variable, quedando una función de la variable.

3.4.1. Función de distribución de un subvector abs. continuo

Dado $\tilde{X} = (X_1, \dots, X_h, X_{h+1}, \dots, X_k)$, si se quiere encontrar la función de distribución de un subvector $X' = (X_1, \dots, X_h)$, se debe tomar límite de la función de distribución del vector original haciendo tender a ∞ a las variables que no aparecen en el subvector. Es decir:

$$F_{X'}(x_1, \dots, x_h) = \lim_{x_{h+1}, \dots, x_k \rightarrow \infty} F_{\tilde{X}}(X_1, \dots, X_k)$$

Dado que para vectores aleatorios absolutamente continuos la función de distribución es la integral múltiple de la densidad, entonces:

$$F_{X'}(x_1, \dots, x_h) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_h} \overbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}^{\{x_{h+1} \dots x_k\} \rightarrow \infty} f_{\tilde{X}}(\tilde{t}) dt_k \dots dt_{h+1} dt_h \dots dt_1$$

Como a las variables x_{h+1} hasta x_k se las integra sobre todos los reales, toda la parte de esa expresión que involucraba a las variables desaparece porque se las reemplaza por números. Lo que queda, entonces, es una expresión que depende de las variables x_1 hasta x_h . $\rightarrow F_{X'}(x_1, \dots, x_h)$.

Si el vector es absolutamente continuo, el subvector también lo será porque se puede encontrar su función de densidad de la siguiente manera:

$$\frac{\partial^h F_{X'}}{\partial x_h \dots \partial x_2 \partial x_1} = f_{X'}(x_1, \dots, x_h) \rightarrow \text{densidad del subvector } X'$$

Cuando derivo respecto a una variable una función que involucra una integral sobre esa variable, esa integral desaparece de la función, por ejemplo:

$$\frac{\partial F_{X'}(x_1, \dots, x_h)}{\partial x_1} \Rightarrow \int_{-\infty}^{x_1'} \dots \int_{-\infty}^{x_h} \dots \int_{-\infty}^{\infty} f_{\tilde{X}}(t) dt_k \dots dt_h \dots \cancel{dt_1}$$

Entonces, cuando derive $F_{X'}$ con respecto a las primeras h variables, van a quedar solamente las integrales para x_{h+1} en adelante sobre todo su soporte. Es decir, puedo encontrar la densidad de un subvector o la densidad marginal de una variable integrando la función de densidad del vector original sobre todo su soporte a las variables que no se desean en el subvector. (En el caso de buscar la densidad marginal de una variable, se integra el resto de las variables).

Ejemplo en \mathbb{R}^2

$$\begin{aligned} \tilde{X} &= (X, Y), \quad f_{X,Y}(x, y) \\ f_X(x) &= \int_{S_Y} f_{X,Y}(x, y) dy \\ f_Y(y) &= \int_{S_X} f_{X,Y}(x, y) dx \end{aligned}$$

3.5. Distribución condicional

Dado el vector n dimensional

$\tilde{X} = (X_1, \dots, X_m, X_{m+1}, \dots, X_n)$, cuyo espacio de probabilidad se conoce.

Supongamos que se tiene interés en asignar probabilidad al evento $[X_1, \dots, X_m] \in C$, dado que $[X_{m+1}, \dots, X_n] \in D$, en el sentido que D ocurrirá con certeza.

La clave para definir una función de probabilidad / densidad condicional consiste en establecer una equivalencia entre eventos que involucran al vector aleatorio m -dimensional (X_1, \dots, X_m) y al vector aleatorio $(n - m)$ -dimensional (X_{m+1}, \dots, X_n) , y eventos para el vector aleatorio n -dimensional (X_1, \dots, X_n)

Ejemplo en \mathbb{R}^2

$$\tilde{X} = (X_1, X_2), \quad P(X_1 \in C) \text{ dado que } X_2 \in D?$$

$$\begin{aligned} [X_1 \in C] &= \{\omega \in \Omega : X_1(\omega) \in C\} : C \in \mathcal{B} \\ &= \{\omega \in \Omega : X_1(\omega) \in C, X_2(\omega) \in \mathbb{R}\} \rightarrow \text{El evento no cambia al agregar} \\ &\quad \text{restricciones redundantes} \\ A &= \{(X_1, X_2) \in \mathbb{R}^2 : X_1 \in C, X_2 \in \mathbb{R}\} \end{aligned}$$

Escribimos el evento de esa manera para hacer entrar al evento las realizaciones del vector aleatorio y no solo de la variable

$$\begin{aligned} [X_2 \in D] &= \{\omega \in \Omega : X_2(\omega) \in D\} : D \in \mathcal{B} \\ &= \{\omega \in \Omega : X_1(\omega) \in \mathbb{R}, X_2(\omega) \in D\} \\ B &= \{(X_1, X_2) \in \mathbb{R}^2 : X_1 \in \mathbb{R}, X_2 \in D\} \end{aligned}$$

Entonces, la probabilidad de que $X_1 \in C$ dado que $X_2 \in D$ se puede definir como:

$$P_{X_1|X_2}(C|D) = P(X_1 \in C|X_2 \in D) = P_{\tilde{X}}(A|B) = \frac{P_{\tilde{X}}(A \cap B)}{P_{\tilde{X}}(B)}$$

La $P_{\tilde{X}}(A|B)$ es la probabilidad inducida por el vector aleatorio porque está trabajando con borelianos de \mathbb{R}^2 .

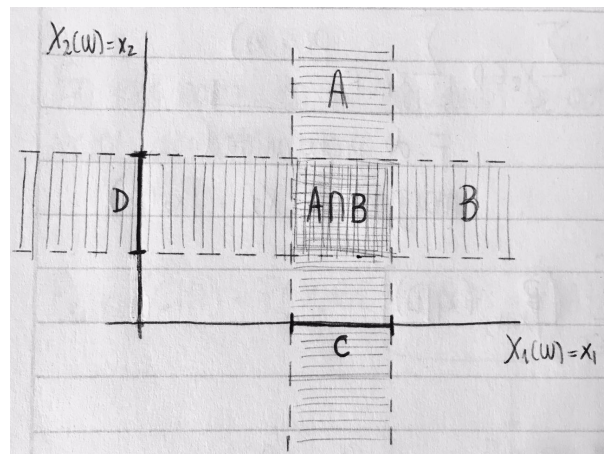
Gráficamente:

$$C, D \rightarrow B \in \mathcal{B}$$

$$A, B \rightarrow B \in \mathcal{B}^2$$

$$(A \cap B) = C \times D$$

$$A \cap B = \{\omega \in \Omega : X_1(\omega) \in C, X_2(\omega) \in D\}$$



En términos geométricos, estoy calculando el volumen relativo de $A \cap B$, en relación a B , que sirve para normalizar (como siempre $P(B) \geq P(A \cap B)$, no corro riesgo de que la probabilidad sea mayor que 1).

3.5.1. Caso discreto

Como para cualquier vector aleatorio discreto, la probabilidad se obtiene sumando:

$$P_{X_1|X_2}(C|D) = \frac{\sum_{x_1, x_2 \in (A \cap B)} p(x_1, x_2)}{\sum_{x_1, x_2 \in B} p(x_1, x_2)}$$

Dado que $B = R_{X_1} \times D$ es un producto cartesiano y $(A \cap B)$ también lo es, la sumatoria puede ser expresada como una doble sumatoria. Es decir:

$$\begin{aligned} \sum_{x_1, x_2 \in (A \cap B)} &= \sum_{x_1, x_2 \in (C \times D)} = \sum_{x_1 \in C} \sum_{x_2 \in D} \\ \sum_{x_1, x_2 \in B} &= \sum_{x_1, x_2 \in R_{X_1} \times D} = \sum_{x_1 \in R_{X_1}} \sum_{x_2 \in D} \end{aligned}$$

Entonces:

$$\begin{aligned} P_{X_1|X_2}(C|D) &= \frac{\sum_{x_1 \in C} \sum_{x_2 \in D} p(x_1, x_2)}{\sum_{x_1 \in R_{X_1}} \sum_{x_2 \in D} p(x_1, x_2)} = \frac{\sum_{x_1 \in C} \sum_{x_2 \in D} p(x_1, x_2)}{\sum_{x_2 \in D} \underbrace{\sum_{x_1 \in R_{X_1}} p(x_1, x_2)}_{\text{función de probabilidad marginal de } X_2}} \\ P_{X_1|X_2}(C|D) &= \sum_{X_1 \in C} \left[\frac{\sum_{x_2 \in D} p(x_1, x_2)}{\sum_{x_2 \in D} p_{X_2}(x_2)} \right] = \sum_{X_1 \in C} \underbrace{\left[p_{X_1|X_2}(X_1|D) \right]}_{\text{f. de probabilidad condicionada}} \end{aligned}$$

Esa función de probabilidad condicionada satisface todas las propiedades de una función de probabilidad de un vector aleatorio discreto.

Para calcular una probabilidad para un vector aleatorio discreto se debe sumar sobre la función de probabilidad del mismo. Como ahora se quiere calcular una probabilidad condicionada, se debe sumar sobre una función de probabilidad condicionada.

La función de probabilidad condicionada es una función de D , es decir del evento con el cual estoy condicionado. Si cambia el evento, cambia todo el sumatorio que se realiza para D .

Si $D = \{x_2\}$ el sumatorio de $x_2 \in D$ no se va a realizar porque solo hay un punto, y la probabilidad condicional se representará por $p_{X_1|X_2}(x_1|x_2)$ con lo cual:

$$p_{X_1|X_2}(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)}$$

3.5.2. Caso continuo

En este caso, las probabilidades se obtienen integrando sobre la densidad del vector aleatorio:

$$P_{X_1|X_2}(C|D) = P(X_1 \in C | X_2 \in D) = P_{\tilde{X}}(A|B) = \frac{\int \int_{A \cap B} f(x_1, x_2) dx_1 dx_2}{\int \int_B f(x_1, x_2) dx_1 dx_2}$$

Dado que $(A \cap B) = C \times D$ y $B = \mathbb{R} \times D$, eso se puede expresar como:

$$P_{X_1|X_2}(C|D) = \frac{\int_C \int_D f(x_1, x_2) dx_2 dx_1}{\int_{\mathbb{R}} \int_D f(x_1, x_2) dx_2 dx_1} = \int_C \underbrace{\left[\frac{\int_D f(x_1, x_2) dx_2}{\int_D f_{X_2}(x_2) dx_2} \right]}_{\text{f. de densidad condicionada}} dx_1$$

Esa función de densidad condicional de X_1 dado $X_2 \in D$ se denota: $f_{X_1|X_2}(x_1|D)$ y cumple con todas las propiedades de una densidad.

Para obtener una probabilidad de un vector aleatorio $\tilde{X} \in B$, se integra la densidad en ese B . Como ahora se quiere obtener una probabilidad condicionada, se integra una densidad condicionada.

¿Qué pasa si $D = \{b\}$? En ese caso $f_{X_1|X_2}(x_1|D) = 0/0$, con lo cual $p_{X_1|X_2}(C|D)$ no está definida.

La forma de obtener esa probabilidad es por medio de un límite. Se arma un intervalo alrededor de $\{b\}$ de amplitud de 2ε , entonces:

$$p_{X_1|X_2}(C|b) = \lim_{\varepsilon \rightarrow 0^+} p_{X_1|X_2}(C|[b-\varepsilon, b+\varepsilon])$$

$$p_{X_1|X_2}(C|b) = \lim_{\varepsilon \rightarrow 0^+} \left[\frac{\int_C \int_{[b-\varepsilon, b+\varepsilon]} f(x_1, x_2) dx_2 dx_1}{\int_{[b-\varepsilon, b+\varepsilon]} f_{X_2}(x_2) dx_2} \right]$$

Si existe $\varepsilon > 0$ tal que $f_{X_2}(x_2)$ y $f(x_1, x_2)$ son continuas en x_2 para todo $x_2 \in [b-\varepsilon, b+\varepsilon]$ y se cumple que $f_{X_2}(b) > 0$, puede aplicarse el *Teorema del Valor Medio* para integrales.

Este teorema nos dice que una integral puede aproximarse por un rectángulo de base $(b-a)$ y altura μ , que es una media entre el mayor y menor punto de la función en el intervalo $[a, b]$.

$$\int_a^b f(x) dx = \mu(b-a)$$

Con lo cual:

$$p_{X_1|X_2}(C|b) = \lim_{\varepsilon \rightarrow 0^+} \left[\frac{2\varepsilon \int_C f(x_1, x_2^*) dx_1}{2\varepsilon f_{X_2}(x_2^o)} \right]$$

Donde 2ε es la amplitud del intervalo y x_2^* y x_2° son puntos en $[b - \varepsilon, b + \varepsilon]$ y donde x_2^* depende también del valor x_1

Cuando $\varepsilon \rightarrow 0^+$, $[b - \varepsilon, b + \varepsilon] \rightarrow \{b\}$, por lo cual en el límite $x_2^* = x_2^\circ = b$.

Así, el valor de la probabilidad condicional es:

$$p_{X_1|X_2}(C|b) = \int_C \frac{f(x_1, b)}{f_{X_2}(b)} dx_1$$

y la densidad condicional es:

$$f_{X_1|X_2} = \frac{f(x_1, b)}{f_{X_2}(b)}$$

3.6. Independencia de variables aleatorias

Sean X_1, X_2, \dots, X_k v.a. definidas en el mismo (Ω, \mathcal{F}, P) , de modo que $\tilde{X} = (X_1, \dots, X_k)$ es un vector aleatorio en el mismo espacio de probabilidad.

Las v.a. X_1, X_2, \dots, X_k son colectivamente independientes sii:

$$P(X_1 \in B_1, \dots, X_k \in B_k) = \prod_{i=1}^k P(X_i \in B_i) \quad \forall B_i \in \mathcal{B}$$

Es decir:

$$P\left(\bigcap_{i=1}^k \{\omega \in \Omega : X_i(\omega) \in B_i\}\right) = P\left(\bigcap_{i=1}^k X_i \in B_i\right) = \prod_{i=1}^k P(X_i \in B_i)$$

(XXXII) Además, cualquier subfamilia de una familia de v.a. independientes está formada por v.a. independientes. Es decir:

Si X_1, X_2, \dots, X_k son v.a. independientes, entonces:

$$P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2)$$

y funciones de familias disjuntas de las X_i también lo son.

Criterio general de independencia **(XXXIII)**

1) Si X_1, X_2, \dots, X_k son independientes, entonces:

$$F_X(x_1, x_2, \dots, x_k) = \prod_{i=1}^k F_{X_i}(x_i)$$

$$\forall (x_1, \dots, x_k) \in \mathbb{R}^k$$

2) Recíprocamente, si existen funciones F_1, F_2, \dots, F_k tales que:

$$\lim_{x_i \rightarrow \infty} F_i(x_i) = 1$$

y

$$F_X(X_1, X_2, \dots, X_k) = \prod_{i=1}^k F_i(x_i) \quad \forall (X_1, X_2, \dots, x_k) \in \mathbb{R}^k$$

Entonces X_1, X_2, \dots, X_k son independientes y $F_i = F_{X_i}$, es decir que esas funciones son efectivamente las funciones de distribución marginal de las k variables. El punto 1 del criterio general no es muy útil ya que normalmente se parte sin el conocimiento de la independencia de las variables, y es justamente lo que se intenta demostrar. El punto 2, en cambio, es útil para demostrar la independencia y básicamente nos dice:

Dada la función de distribución conjunta de $X_1, X_2, \dots, X_k = F_{\tilde{X}}(x_1, \dots, x_k)$, si la misma se puede factorizar y expresar como k funciones que solo dependan de una variable, y si cada una de esas funciones tiende a 1 cuando la variable de la cual depende tiende a infinito, entonces las k variables son independientes y dichas funciones son su respectiva función de distribución.

(XXXIV) Criterio en el caso absolutamente continuo

1) Si X_1, X_2, \dots, X_k son v.a. continuas independientes y poseen densidades f_{X_1}, \dots, f_{X_k} , entonces la función:

$$f(x_1, \dots, x_k) = \prod_{i=1}^k f_{X_i}(x_i) \quad \forall (x_1, \dots, x_k) \in \mathbb{R}^k$$

es la densidad conjunta de las v.a. X_1, X_2, \dots, X_k .

2) Recíprocamente, si X_1, X_2, \dots, X_k tienen densidad conjunta $f_X(x)$ y satisfacen:

$$f_X(x_1, \dots, x_k) = \prod_{i=1}^k f_i(x_i) \quad \forall (x_1, \dots, x_k) \in \mathbb{R}^k$$

donde $f_i(x) \geq 0 \quad \forall x \in \mathbb{R}$ y

$$\int_{-\infty}^{\infty} f_i(x) dx = 1 \quad i = 1, 2, \dots, k$$

Entonces X_1, X_2, \dots, X_k son independientes y f_i es la densidad de X_i , donde $i = 1, \dots, k$.

Dado que tanto la función de densidad conjunta como las marginales llevan una indicadora del soporte, si las v.a. son independientes, el soporte del vector debe poder expresarse como el producto del soporte de cada v.a.

Criterio en el caso discreto.

Sean X_1, X_2, \dots, X_k variables aleatorias discretas definidas sobre el mismo (Ω, \mathcal{F}, P) , las mismas son independientes sii:

$$p_X(x) = \prod_{i=1}^k p_{X_i}(x_i) \quad \forall x = (x_1, \dots, x_k) \in \mathbb{R}^k$$

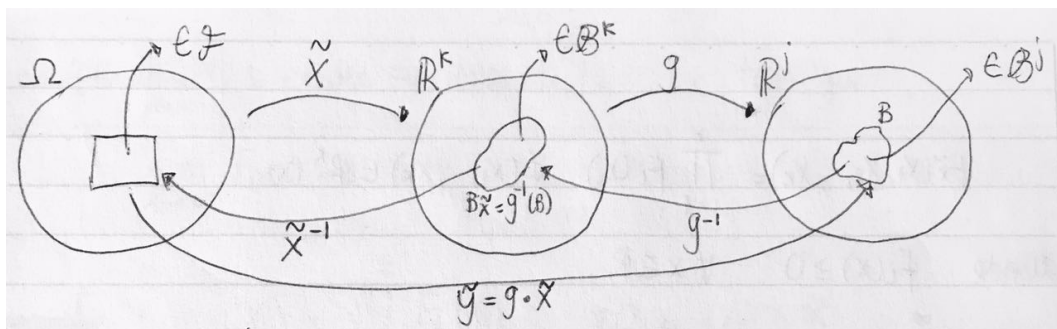
Capítulo 4

Unidad 4. Distribución de funciones de vectores aleatorios

Sea $\tilde{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio en (Ω, \mathcal{F}, P) , $\tilde{X} : \Omega \rightarrow \mathbb{R}^k$ con distribución $p_{\tilde{X}}$ conocida.

Y sea $g : \mathbb{R}^k \rightarrow \mathbb{R}^j$ una función vectorial medible.

Entonces $g \circ \tilde{X}$ es vector aleatorio. El problema ahora está en encontrar su distribución.



Como $g : \mathbb{R}^k \rightarrow \mathbb{R}^j$ es medible $\rightarrow g^{-1}(B) \in \mathcal{B}^k$

Como $\tilde{X} : \Omega \rightarrow \mathbb{R}^k$ es vector aleatorio $\rightarrow \tilde{X}^{-1} \in \mathcal{F}$

$\tilde{Y} = g(\tilde{X})$ es vector aleatorio. ($\tilde{Y} : \Omega \rightarrow \mathbb{R}^j$)

Queremos obtener $P_Y(B) = P(\tilde{Y} \in B)$

El evento $[\tilde{Y} \in B]$ tiene un equivalente en términos de X .

$[\tilde{Y} \in B] = \{\omega \in \Omega : \tilde{Y}(\omega) \in B\}$, pero esos ω que son mapeados a B por \tilde{Y} son

los mismos que son mapeados a $g^{-1}(B)$ por \tilde{X} . Por lo tanto:

$$[\tilde{Y} \in B] = [\tilde{X} \in g^{-1}(B)] = \{\omega \in \Omega : \tilde{X}(\omega) \in g^{-1}(B)\}$$

$$\text{Entonces: } P_Y(B) = P[Y \in B] = P[\tilde{X} \in g^{-1}(B)] = P_{\tilde{X}}(\underbrace{g^{-1}(B)}_{\in \mathcal{B}^k})$$

4.1. Método de la función de distribución

Si $g : \mathbb{R}^k \rightarrow \mathbb{R}^j$, $Y = g \circ X$ es v.a. y $B = (-\infty, y]$, $y \in \mathbb{R}$.

$$P_Y(B) = P(Y \in (-\infty, y]) = P[\tilde{X} \in g^{-1}((-\infty, y])] = F_Y(y)$$

Otra forma de verlo:

$$F_Y(y) = P(Y \leq y) = P(\underbrace{g(X_1, \dots, X_k) \leq y}_{B_y}) = P(X \in B_y) = P_X(B_y)$$

$$B_y = \{(X_1, \dots, X_k) : g(X_1, \dots, X_k) \leq y\}$$

Ejemplo para $k = 1$, $j = 1$

$X \sim U[0, 1]$ v.a.

$f_X(x) = 1 \cdot 1_A(x)$ donde $A = [0, 1]$

$$F_X(x) = \begin{cases} 1 & x \geq 1 \\ x & 1 > x \geq 0 \\ 0 & x < 0 \end{cases}$$

Dado $Y = -\ln(X)$, se quiere encontrar $F_Y(y)$ y $f_Y(y)$

Primero veamos entre qué valores varía Y . Para eso se debe valorar la función en los valores extremos de la v.a. X .

$-\ln([0, 1])$, como $\ln(0)$ no existe, abrimos el intervalo y valuamos en $(0, 1]$, con lo cual:

$$-\ln((0, 1]) \rightarrow (-\infty, 0] \rightarrow \text{Este es el intervalo para el cual } f_Y(y) > 0$$

Por lo tanto ya sabemos que $F_Y(y) = 0$ si $y < 0$

Ahora debemos encontrar $F_Y(y)$ si $y \geq 0$

$$F_Y(y) = P(Y \leq y)$$

$$\left. \begin{aligned} &= P(-\ln(x) \leq y) \\ &= P(\ln(x) \geq y) \quad (1) \\ &= P(x \geq e^{-y}) \quad (2) \end{aligned} \right\} \begin{aligned} &[-\ln(x) \leq y] = \{\omega \in \Omega : -\ln(X(\omega)) \leq y\} \\ &\text{El conjunto no cambia si se llevan a cabo operaciones} \\ &\text{inyectivas sobre la restricción. En este caso, en (1) se} \\ &\text{multiplica ambos miembros por } -1 \text{ y en (2)} \\ &\text{se toma anti-logaritmo.} \end{aligned}$$

$$F_Y(y) = 1 - P(X < e^{-y})$$

Como X es continua, $P(X = e^{-y}) = 0$, por lo tanto es lo mismo usar $<$ o \leq .

Como necesitamos llegar a la expresión de $F_X(x)$ para usarla para determinar $F_Y(y)$, entonces reemplazamos:

$$\begin{aligned} F_Y(y) &= 1 - P(X \leq e^{-y}) \\ &= 1 - F_X(e^{-y}) \end{aligned}$$

Como $y \in [0, \infty]$, $e^{-y} \in [0, 1]$, por lo tanto debemos utilizar la forma funcional de X cuando $X \in (0, 1] \rightarrow F_X(x) = X$, $F_x(e^{-y}) = e^{-y}$.

$$F_Y(y) = 1 - e^{-y}$$

Entonces:

$$\begin{aligned} F_Y(y) &= \begin{cases} 0 & \text{si } y < 0 \\ 1 - e^{-y} & \text{si } y \geq 0 \end{cases} \\ f_Y(y) &= \frac{\partial F_Y(y)}{\partial y} = e^{-y} \mathbf{1}_{[0, \infty)}(y) \end{aligned}$$

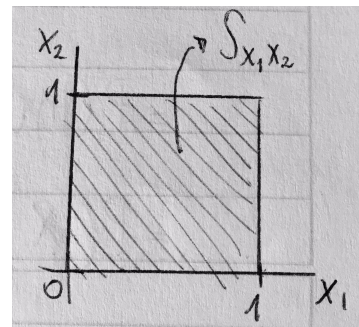
Ejemplo para $k = 2, j = 1$

X_1, X_2 v.a.i.i.d (variables aleatorias independientes idénticamente distribuidas)

$$X_1 \sim U[0, 1], \quad X_2 \sim U[0, 1] \longrightarrow f_{X_i}(x_i) = \mathbf{1}_{[0, 1]}(x_i) \text{ para } i = 1, 2.$$

Como son independientes: $f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \mathbf{1}_{[0, 1] \times [0, 1]}(x_1, x_2) \\ &= \mathbf{1}_{[0, 1]^2}(x_1, x_2) \end{aligned}$$



Dada $Y = g(X_1, X_2) = X_1 + X_2$, debemos encontrar $F_Y(y)$.

Como g es continua, es medible. Como X_1, X_2 son v.a., (X_1, X_2) es vector aleatorio. Entonces $Y = g(X_1, X_2)$ es variable aleatoria.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(Y \in (-\infty, y)) \\ &= P(Y \in B_y) \end{aligned}$$

Como tenemos información de la probabilidad del vector (X_1, X_2) , debemos expresar esa probabilidad en términos del mismo, para lo cual se calcula la *preimagen* de B_y :

$$\begin{aligned} g^{-1}(B_y) &= \{(X_1, X_2) : g(X_1, X_2) \in B_y\} \\ &= \{(X_1, X_2) : X_1 + X_2 \leq y\} = B_{\tilde{X}} \end{aligned}$$

$$\begin{aligned} F_Y(y) &= P[(X_1, X_2) \in B_{\tilde{X}}] \\ &= P(X_1 + X_2 \leq y) \end{aligned}$$

Para calcular una probabilidad de v.a. continuas se debe integrar la función de densidad conjunta sobre la intersección del boreliano relevante y el soporte del vector.

Además, como X_1 y X_2 varían entre 0 y 1, $Y = X_1 + X_2$ varía entre 0 y 2.

Por lo tanto, sabemos que:

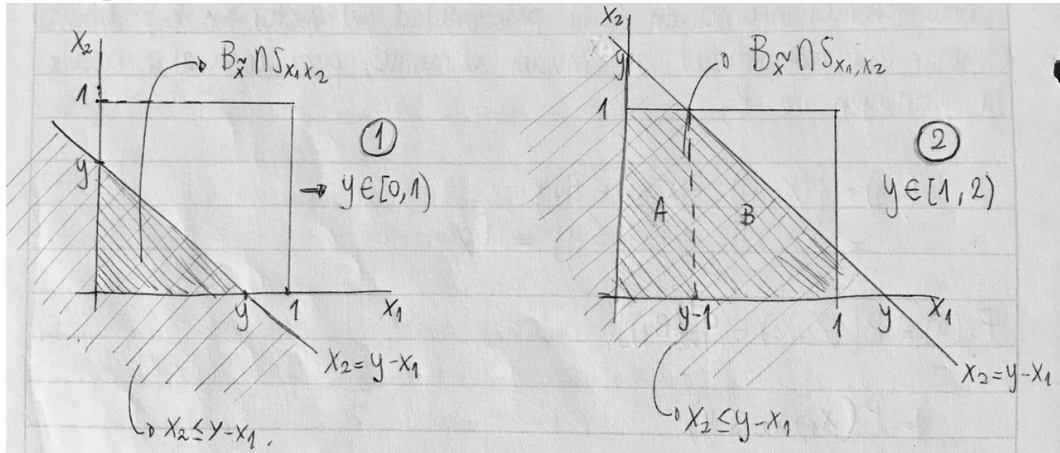
$$F_Y(y) = \begin{cases} 0 & y < 0 \\ ? & y \in [0, 2) \longrightarrow \text{debemos encontrar esa forma funcional} \\ 1 & y \geq 2 \end{cases}$$

Primero debemos graficar en el plano x_1, x_2 el boreliano en cuestión e intersectarlo con el soporte de (X_1, X_2) .

$$P(X_1 + X_2 \leq y) = P(X_2 \leq y - X_1)$$

Debemos graficar la recta $x_2 = y - x_1$ en el plano. Pero esa recta va a cambiar de posición según cual sea el valor de y ; y al cambiar de posición también cambia la forma geométrica del soporte el cual integrar y por lo tanto, los límites de integración.

En particular, vamos a tener dos casos:



Caso 1. $y \in [0, 1)$

$$\begin{aligned} F_Y(y) &= \int_0^y \int_0^{y-x_1} 1 \, dx_2 \, dx_1 \\ &= \frac{y^2}{2} 1_{[0,1)} \end{aligned}$$

Caso 2. $y \in [1, 2)$

$$\begin{aligned} F_Y(y) &= \underbrace{\int_0^{y-1} \int_0^1 1 \, dx_2 \, dx_1}_A + \underbrace{\int_{y-1}^1 \int_0^{y-x_1} 1 \, dx_2 \, dx_1}_B \\ F_Y(y) &= -1 + 2y - \frac{1}{2}y^2 1_{[1,2)}(y) \end{aligned}$$

Por lo tanto:

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2}y^2 & y \in [0, 1) \\ -1 + 2y - \frac{1}{2}y^2 & y \in [1, 2) \\ 1 & y \geq 2 \end{cases}$$

$$f_Y(y) = \frac{\partial F_Y(y)}{\partial y} = \begin{cases} y & y \in [0, 1) \\ 2 - y & y \in [1, 2) \\ 0 & c.c. \end{cases}$$

4.2. Método del Jacobiano/ de la transformación

Sea $g: \mathbb{R}^k \rightarrow \mathbb{R}^k$

Sea $G_0 \subset \mathbb{R}^k$ y $G \subset \mathbb{R}^k$

y sea g una biyección entre G_0 y G , dónde:

$$g(X_1, \dots, X_k) = (g_1(X_1, \dots, X_k), \dots, g_k(X_1, \dots, X_k)) = (Y_1, \dots, Y_k)$$

Como g es biyectiva, existe la función inversa $h \equiv g^{-1}$ en $G \rightarrow G_0$.

Si las funciones g nos permiten expresar a Y_i como funciones de X_i , la inversa h permite expresar a las X_i como funciones de las Y_i :

$$X_1 = h_1(Y_1, \dots, Y_k), \dots, X_k = h_k(Y_1, \dots, Y_k)$$

Dichas funciones existen y las k^2 derivadas parciales $\frac{\partial X_i}{\partial y_j}$ también existen y son continuas en G .

Con esas derivadas parciales, se construye la Matriz Jacobiana (MJ)

$$\text{MJ} = \begin{bmatrix} \frac{\partial X_1}{\partial y_1} & \cdots & \frac{\partial X_1}{\partial y_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial X_k}{\partial y_1} & \cdots & \frac{\partial X_k}{\partial y_k} \end{bmatrix}$$

Se define el jacobiano $j(x, y)$ como el determinante de la matriz jacobiana:

$j(x, y) = |\text{MJ}|$; el mismo no se debe anular para ningún $y \in G$

$$\boxed{f_Y(y) = f_X(h(y)) |j(x, y)| 1_G(y)}$$

La región G_0 es tal que $P(X \in G_0) = 1$

Si el soporte de X es un abierto, entonces $G_0 = S_X$

Si el soporte de X es un cerrado, entonces $G_0 = S_X - \text{frontera}$; (la probabilidad en la frontera es cero, entonces la probabilidad en $G_0 = \text{en } S_X$)

La región G se halla con los rangos de variación de las variables X_i reemplazando las X por su igual $h(y)$.

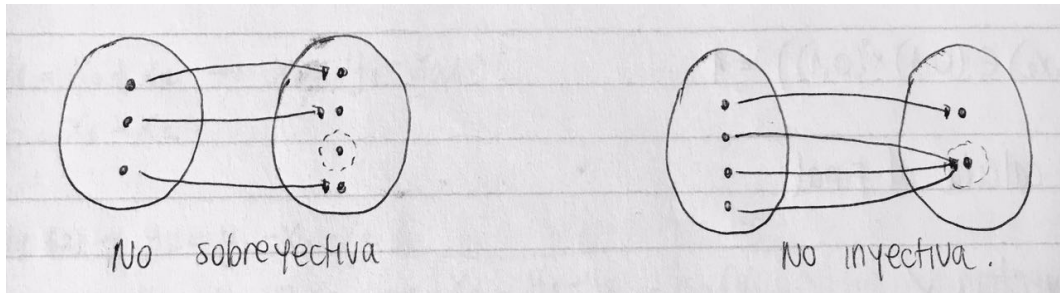
Antes de llevar a cabo el método se deben verificar los siguientes supuestos:

- 1) $g: \mathbb{R}^k \rightarrow \mathbb{R}^k$
- 2) $G_0 \subset \mathbb{R}^k$ y $G \subset \mathbb{R}^k$ abiertos.

3) g biyectiva (sobreyectiva e inyectiva).

Sobreyectividad \rightarrow a cada elemento de la imagen le corresponde, al menos, un elemento del dominio.

Inyectividad \rightarrow a cada elemento de la imagen le corresponde solo un elemento del dominio. Es decir cada imagen tiene solo una preimagen.



4) Las k^2 derivadas parciales existen y son continuas.

5) El jacobiano no se anula para ningún $y \in G$

Hay dos supuestos que, si no se cumplen, pueden ser salvados:

1) Si $g: \mathbb{R}^k \rightarrow \mathbb{R}^j$ para $k \neq j$

si $k < j$ no se puede hacer nada

si $k > j$ se puede salvar.

2) Si la función $g: \mathbb{R}^k \rightarrow \mathbb{R}^k$ no es 1 a 1 (no inyectiva).

Ejemplo.

Sean X_1, X_2 v.a.i.i.d, $X_1 \sim U[0, 1] \rightarrow f_{X_1, X_2} = 1 \cdot 1_{[0,1]^2}(x_1, x_2)$

Y sea $\tilde{Y} = (X_1 + X_2, X_1 - X_2) = (Y_1, Y_2)$, se quiere encontrar $F_{\tilde{Y}}(\tilde{y})$.

Antes de aplicar el método se deben verificar los supuestos:

1) $g: \mathbb{R}^k \rightarrow \mathbb{R}^k \checkmark$

2) $G_0 \subset \mathbb{R}^k$ y $G \subset \mathbb{R}^k$ abiertos.

$Sop_{(X_1, X_2)} = [0, 1] \times [0, 1] \rightarrow$ cerrado en \mathbb{R}^2

$G_0 = \overset{\circ}{Sop}_{(X_1, X_2)} = (0, 1) \times (0, 1) \checkmark$

$P((X_1, X_2) \in (0, 1) \times (0, 1)) = 1$

G se calcula al final.

3) g biyectiva ✓

a) Sobreyectividad ✓

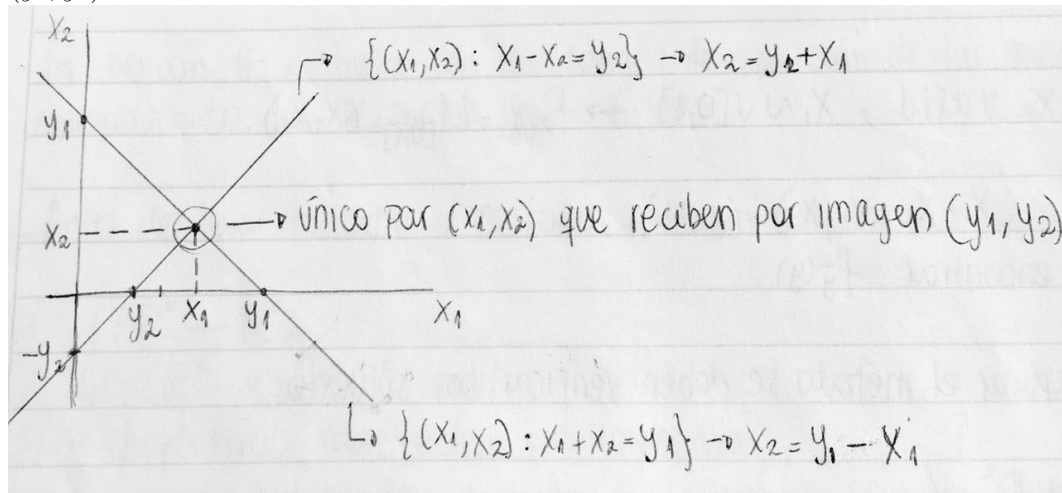
$$P(Y \in G) = P(X \in h(G)) = P(X \in G_0) = 1.$$

Por lo tanto, la función g es sobreyectiva porque el conjunto de llegada va a contener a \tilde{Y} con probabilidad 1.

b) Inyectividad ✓

Se debe calcular la cardinalidad del conjunto preimagen para cualquier y de la llegada. Si esa cardinalidad es 1, la función es inyectiva.

$$(y_1, y_2) \in G$$



$$g^{-1}(Y_1, Y_2) = \{(X_1, X_2) : X_1 + X_2 = Y_1, X_1 - X_2 = Y_2\}$$

$$\#g^{-1}(Y_1, Y_2) = 1 \longrightarrow g \text{ inyectiva.}$$

Ahora se debe expresar a X_1, X_2 como funciones de Y_1, Y_2 .

$$(1) y_1 = x_1 + x_2 \rightarrow x_1 = y_1 - x_2 \quad (3)$$

$$(2) y_2 = x_1 - x_2$$

$$(3) \text{ en } (2) \rightarrow y_2 = y_1 - x_2 - x_2$$

$$y_2 = y_1 - 2x_2 \rightarrow x_2 = \frac{y_1 - y_2}{2} = h_2(y_1, y_2) \quad (4)$$

$$(4) \text{ en } (3) \rightarrow x_1 = y_1 - \frac{y_1 - y_2}{2} \rightarrow x_1 = \frac{y_1 + y_2}{2} = h_1(y_1, y_2)$$

4) Las 4 derivadas parciales deben existir y ser continuas ✓

Ahora armamos la Matriz jacobiana

$$MJ = \begin{bmatrix} \frac{\partial X_1}{\partial y_1} & \frac{\partial X_1}{\partial y_2} \\ \frac{\partial X_2}{\partial y_1} & \frac{\partial X_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Y calculamos el jacobiano: $j((x_1, x_2), (y_1, y_2)) = |MJ| = -\frac{1}{2}$

5) El Jacobiano no se debe anular para ningún $y \in G$ ✓

Valuamos $f_{X_1, X_1}(x_1, x_2)$ en h_1, h_2 :

$$f_{X_1, X_1}(h_1(y), h_2(y)) = 1 \quad (\text{porque no hay } X \text{ para reemplazar})$$

Aplicamos la fórmula:

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(h_1, h_2) |j(y_1, y_2)| 1_G(y_1, y_2)$$

$$f_{Y_1, Y_2}(y_1, y_2) = 1 \cdot \frac{1}{2} \cdot 1_G(y_1, y_2)$$

Nos falta solamente calcular la región G ; para eso debemos usar los rangos de variación de X_1 y X_2 :

$$0 \leq X_1 \leq 1$$

$$0 \leq \frac{y_1 + y_2}{2} \leq 1$$

$$0 \leq y_1 + y_2 \leq 2$$

$$\longrightarrow y_1 + y_2 = 0 \longrightarrow y_2 = -y_1$$

$$\longrightarrow y_1 + y_2 = 2 \longrightarrow y_2 = 2 - y_1$$

$$0 \leq X_2 \leq 1$$

$$0 \leq \frac{y_1 - y_2}{2} \leq 1$$

$$0 \leq y_1 - y_2 \leq 2$$

$$\longrightarrow y_1 - y_2 = 0 \longrightarrow y_2 = y_1$$

$$\longrightarrow y_1 - y_2 = 2 \longrightarrow y_2 = y_1 - 2$$

Trabajamos con las igualdades para poder graficar.

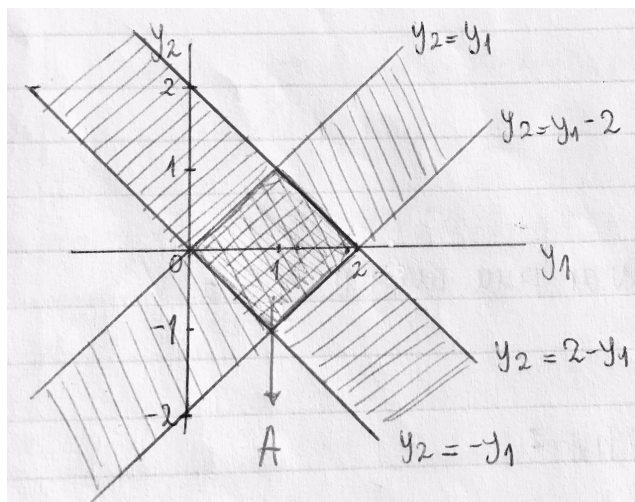
Desigualdades

$$y_2 \geq -y_1$$

$$y_2 \leq 2 - y_1$$

$$y_2 \geq y_1$$

$$y_2 \leq y_1 - 2$$



La región $G = \overset{\circ}{A} \rightarrow$ abierto ✓

Casos especiales

1) Si $g : \mathbb{R}^k \rightarrow \mathbb{R}^j$ con $k \neq j$

i) Si $j > k$ no se puede hacer nada.

Como G_0 es un conjunto de dimensión k , $g(G_0)$ también será un conjunto de dimensión k , pero en \mathbb{R}^j . Como $j > k$, va a faltar variación en por lo menos una componente, y por lo tanto al integrar la función de densidad en $g(G_0)$, esta integral dará 0.

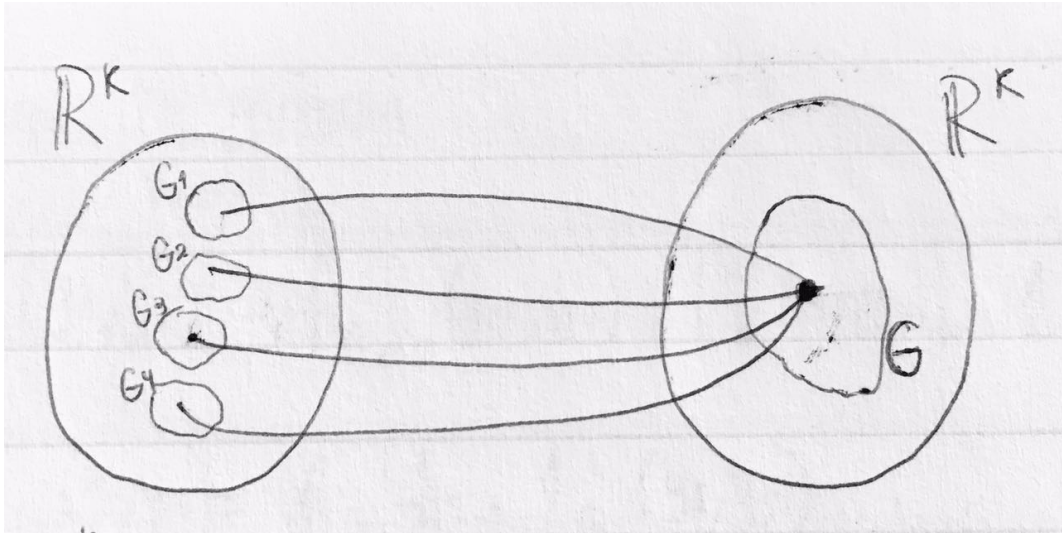
Sin embargo, $P_Y(g(G_0)) = 1$. Por lo tanto Y no puede ser un vector continuo (en cualquier vector continuo la integral de la densidad sobre un conjunto debe ser igual a la probabilidad) y no se puede aplicar el método del Jacobiano.

ii) Si $j < k$ se puede aplicar el método siempre y cuando se complete a g a través de la definición de variables auxiliares $Y_{j+1} = g_{j+1}(\tilde{X}) \dots Y_k = g_K(\tilde{X})$, es decir equiparar la dimensión de \tilde{Y} con la de \tilde{X} .

Una vez empleada g , se podrá obtener la densidad conjunta de $Y_1, \dots, Y_j, \dots, Y_k$ por el método del Jacobiano y luego encontrar la densidad marginal de Y_1, \dots, Y_j integrando sobre todo el soporte de las variables auxiliares, Y_{j+1}, \dots, Y_k . Las variables auxiliares que se eligen deben ser lo mas sencillas posibles y en general serán del tipo $Y_j = X_i$.

2) Si $g : \mathbb{R}^k \rightarrow \mathbb{R}^k$ no es inyectiva (no es 1 a 1).

El método del jacobiano puede aplicarse en este caso si se restringe a g a un número finito de regiones abiertas en donde g sea inyectiva, cuya unión contenga el valor de \tilde{X} con probabilidad 1.



$$P(\tilde{X} \in \bigcup_{i=1}^n G_i) = 1$$

En cada región G_i , la función g debe ser 1 a 1:

$g|_{G_i}$ (g restringida a la región G_i) es biunívoca entre G_i y G , $i = 1, \dots, n$.

Como $g|_{G_i}$ es biyectiva va a existir su inversa, $h^{(i)}$

$$h^{(i)} \equiv g^{-1}|_{G_i} : G \rightarrow G_i$$

Cada $h^{(i)}$ va a tener un $j_i(x, y)$

La densidad conjunta de Y ahora será:

$$f_Y(y) = \begin{cases} \sum_{i=1}^n f_X(h^{(i)}(y)) |j_i(x, y)| & \text{si } y \in G \\ 0 & \text{si } y \notin G \end{cases}$$

4.2.1. Aplicaciones del método del Jacobiano

Sean $Z_i \sim N(0, 1)$ v.a.i.i.d $i = 1, \dots, n$

$$(XXXV) Y = \sum_{i=1}^n Z_i^2 \sim \chi_n^2$$

Se puede probar por el método del Jacobiano que la suma de n variables aleatorias normales estándar al cuadrado es una χ^2 con n grados de libertad, que en realidad es una *Gamma* con $\alpha = n/2$ y $\beta = 2$. La prueba es por inducción.

4.3. Distribuciones en el muestreo

De manera frecuente, los datos recolectados de un experimento consisten en varias observaciones de una variable de interés. Un modelo que se utiliza para describir estas situaciones es el *muestreo aleatorio*.

Muestra aleatoria

Las v.a. X_1, \dots, X_n son llamadas una muestra aleatoria de tamaño n de una población $f(x)$ si son v.a. independientes y si la f. de probabilidad o densidad de cada X_i es la misma $f(x)$. Se puede decir que X_1, \dots, X_n son v.a.i.i.d con f.p. o f.d. $f(x)$

X_i representa la i -ésima observación muestral.

x_i representa el valor observado (realización) de la i -ésima observación muestral ($x_i = X_i(\omega)$)

Cada observación es sobre la misma característica de interés.

Como las X_i son v.a.i.i.d:

$$\begin{aligned} f(x_1, \dots, x_n) &= \prod_{i=1}^n f_{X_i}(x_i) \longrightarrow \text{por independencia} \\ &= \prod_{i=1}^n f_X(x_i) \longrightarrow \text{por idénticamente distribuidas} \end{aligned}$$

Para que aplique la definición de modelo de muestreo aleatorio, la muestra debe provenir de una población infinita o finita con reemplazo.

Cuando una muestra es seleccionada generalmente se calcula un valor que resuma los datos. Este valor puede representarse matemáticamente como una

función $T(x_1, \dots, x_n)$, que es una variable o vector aleatorio $Y = T(X_1, \dots, X_n)$. Como la distribución de Y se obtiene a partir de las variables en la m.a., se denomina distribución muestral de Y . Diferencia entre las distribuciones de X_i e Y : F de X_i : población; F de Y : muestra.

4.3.1. Estadístico

Sea X_1, \dots, X_n una muestra aleatoria tamaño n y sea $T(x_1, \dots, x_n)$ una función a valores reales o una función vectorial cuyo dominio incluye el espacio muestral de (X_1, \dots, X_n) . Entonces la variable o vector aleatorio $Y = T(X_1, \dots, X_n)$ se denomina *estadístico* y su distribución *distribución muestral* de Y .

La única restricción a los estadísticos es que no pueden ser función de un parámetro (valores desconocidos), y además deben ser medibles para ser v.a.

4.3.2. Distribuciones en el muestreo en poblaciones normales

Si X_1, \dots, X_n es una m.a. con distribución $N(\mu, \sigma^2)$

$$\underbrace{(X_1, \dots, X_n)}_{\text{v.a.i.i.d.} \sim N} \xrightarrow{T} \left(\underbrace{\frac{\sum_{i=1}^n X_i}{n}}_{\bar{X}}, \underbrace{\frac{\sum (X_i - \bar{X}_n)^2}{n-1}}_{S^2} \right)$$

Entonces:

- 1) \bar{X} y S^2 son v.a. independientes.
- 2) \bar{X} tiene distribución normal $(\mu, \frac{\sigma^2}{n})$
- 3) $\frac{(n-1)S^2}{\sigma^2}$ tiene distribución χ^2 con $n-1$ grados de libertad.

Todo eso se puede demostrar con el método del jacobiano.

4.3.3. Distribución *t de student*

Sean U y V v.a. independientes, donde $U \sim N(0, 1)$ y $V \sim \chi_p^2$. Entonces la v.a. $\frac{U}{\sqrt{\frac{V}{p}}}$ se distribuye *t de student* con p grados de libertad.

La densidad de T es: **(XXXVI)**

$$f_T(t) = \frac{\Gamma((p+1)/2)}{\Gamma(p/2)} \frac{1}{\sqrt{p\pi}} \frac{1}{(1 + (t^2/p))^{\frac{p+1}{2}}} \mathbf{1}_{\mathbb{R}}(t)$$

4.3.4. Distribución F de Snedecor

Sean U y V v.a. independientes, donde $U \sim \chi_p^2$ y $V \sim \chi_q^2$. Entonces la v.a. $F = \frac{U/p}{V/q}$ tiene distribución F de Snedecor con p y q grados de libertad. La densidad de F es: **(XXXIX)**

$$f_F(x) = \frac{\Gamma((p+q)/2)}{\Gamma(p/2) \Gamma(q/2)} \left(\frac{p}{q}\right)^{p/2} \frac{x^{(p/2)-1}}{[1 + (p/q)x]^{(p+q)/2}} \mathbf{1}_{(0,\infty)}(x)$$

4.4. Estadísticos de orden

Sean X_1, \dots, X_n una m.a. de una población con función de distribución F y sea X una v.a. genérica con esta distribución.

Los estadísticos de orden son ordenamientos de las observaciones muestrales, por ejemplo: la observación mas grande, mas pequeña, central, etc.

Para $k = 1, 2, \dots, n$ sea $X_{(k)}$ = el k -ésimo más pequeño de X_1, \dots, X_n ($X_{(1)}, X_{(2)}, \dots, X_{(n)}$) se denominan *estadísticos de orden* y $X_{(k)}$ es la k -ésima variable de orden.

Hay tantos estadísticos de orden como tamaño de muestra.

Los e.o. se obtienen de la muestra original a través de permutaciones de modo que las observaciones muestrales son ubicadas en orden creciente:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

Por lo tanto:

$$X_{(1)} = \min\{x_1, \dots, x_n\}$$

$$X_{(n)} = \max\{x_1, \dots, x_n\}$$

Rango muestral $\rightarrow R = X_{(n)} - X_{(1)}$

Mediana muestral $\rightarrow M$

$$M = \begin{cases} X_{((n+1)/2)} & \text{si } n \text{ es impar} \\ \frac{X_{(n/2)} + X_{(n/2+1)}}{2} & \text{si } n \text{ es par} \end{cases}$$

La mediana muestral es un número tal que aproximadamente la mitad de las observaciones son menores que M y la otra mitad, mayores. El es 50-ésimo percentil muestral.

4.4.1. Función de distribución marginal de los e.o.

Dado que los e.o. son funciones de la muestra, afirmaciones probabilísticas respecto a ellos pueden ser calculadas en términos de las probabilidades de X_1, \dots, X_n (las observaciones muestrales).

La función de distribución marginal del k -ésimo e.o. es:

$$F_{X_{(k)}}(x) = \frac{\Gamma(n+1)}{\Gamma(k) \Gamma(n+1-k)} \int_0^{F(x)} y^{k-1} (1-y)^{n-k} dy$$
$$F_{X_{(k)}}(x) = \frac{n!}{(k-1)! (n-k)!} \int_0^{F(x)} y^{k-1} (1-y)^{n-k} dy$$

Sabemos por el Teorema Fundamental del Cálculo que:

$$h(x) = \int_0^x f(t) dt \rightarrow \frac{\partial h(x)}{\partial x} = f(x)$$

Pero ahora nuestra h es una función compuesta:

$$h(F(x)) = \int_0^{F(x)} f(t) dt = F_{X_{(k)}}(x)$$
$$\frac{\partial h(F(x))}{\partial x} = f_{X_{(k)}}(x) = \frac{\partial F_{X_{(k)}}(x)}{\partial x}$$

Al ser una función compuesta aplicamos la regla de la cadena:

$$\frac{\partial h(F(x))}{\partial x} = h'(F(x)) F'(x)$$
$$f_{X_{(k)}}(x) = h'(F(x)) f_X(x)$$

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)! (n-k)!} \underbrace{[F_X(x)]^{k-1} [1 - F_X(x)]^{n-k}}_{\text{valuado en } F_X(x)} \underbrace{f_X(x)}_{\text{densidad poblacional}} \rightarrow \text{densidad del e.o.k}$$

(XXXVII) Si $k = n$ (máximo)

$$f_{X_{(n)}}(x) = n [F_X(x)]^{n-1} f_X(x)$$

(XXXVIII) Si $k = 1$ (mínimo)

$$f_{X_{(1)}}(x) = n [1 - F_X(x)]^{n-1} f_X(x)$$

4.4.2. Distribución conjunta de los e.o.

Sea \tilde{X} vector abs. continuo y X_1, \dots, X_n v.a.i.i.d

Partimos de la densidad conjunta de las observaciones muestrales:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

Consideremos la transformación

$$(X_1, \dots, X_n) \longrightarrow (X_{(1)}, \dots, X_{(n)})$$

Queremos encontrar $f_{X_{(1)}, \dots, X_{(n)}}(y_1, y_2, \dots, y_n)$, cuyo soporte es $\{y_1 < y_2 < \dots < y_n\}$

Como los estadísticos de orden son simplemente un reordenamiento de las observaciones muestrales, esta transformación puede ser escrita cómo:

$$\begin{pmatrix} X_{(1)} \\ X_{(2)} \\ \vdots \\ X_{(n)} \end{pmatrix} = P \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \text{ donde } P \text{ es una matriz de permutación}$$

Sin embargo, la aplicación no es inyectiva, ya que existen $n!$ resultados diferentes que generan los mismos estadísticos de orden. Por ejemplo, en una muestra

de tamaño 3, podemos obtener:

$$\left. \begin{array}{l} X_1 < X_2 < X_3 \\ X_1 < X_3 < X_2 \\ X_2 < X_1 < X_3 \\ X_2 < X_3 < X_1 \\ X_3 < X_1 < X_2 \\ X_3 < X_2 < X_1 \end{array} \right\} 6 = 3! \text{ resultados que van a arrojar los mismos estadísticos de orden}$$

Por lo tanto se debe particionar el espacio \mathbb{R}^n en $n!$ partes del mismo tipo, de manera que la aplicación restringida a cada parte sea inyectiva. Aplicando la fórmula del Jacobiano:

$$f_{X_{(1)}, \dots, X_{(n)}} = \sum_{i=1}^n f_{X_1, \dots, X_n}(x_{1i}(y), \dots, x_{ni}(y)) |j_i|$$

Dado que existen $n!$ ordenamientos que arrojan los mismos resultados e.o., existen por lo tanto $n!$ matrices de permutación, y $n!$ jacobianos. Pero dado que el determinante de una matriz de permutación es 1 o -1 , $|j_i| = 1 \forall i = 1, \dots, n!$.

Además, como cada estadístico de orden va a depender solo de una de las observaciones muestrales, cada $X_{ki}(y)$ va a depender de un solo y_j . y por lo tanto:

$$f_{X_1, \dots, X_n}(x_{1i}(y), \dots, x_{ni}(y)) = \prod_{i=1}^n f(y_k) \quad i = 1, \dots, n!$$

Es decir, se multiplica la densidad original f evaluada en los puntos $X_{ki}(y)$ con $k = 1, 2, \dots, n$, es decir en los puntos y_1, y_2, \dots, y_n pero en un orden diferente. La densidad conjunta de los e.o. es, por lo tanto, la suma de $n!$ términos $\prod_{i=1}^n f(y_k)$ 1. Llegamos así al siguiente resultado:

$$f_{X_{(1)}, \dots, X_{(n)}}(y_1, \dots, y_n) = \begin{cases} n! \prod_{k=1}^n f(y_k) & \text{si } y_1 < y_2 < \dots < y_n \\ 0 & \text{c.c.} \end{cases}$$

Aunque las X_i sean independientes, los e.o. nunca lo son, a causa de esa relación de orden $(y_1 < y_2 < \dots < y_n)$

Capítulo 5

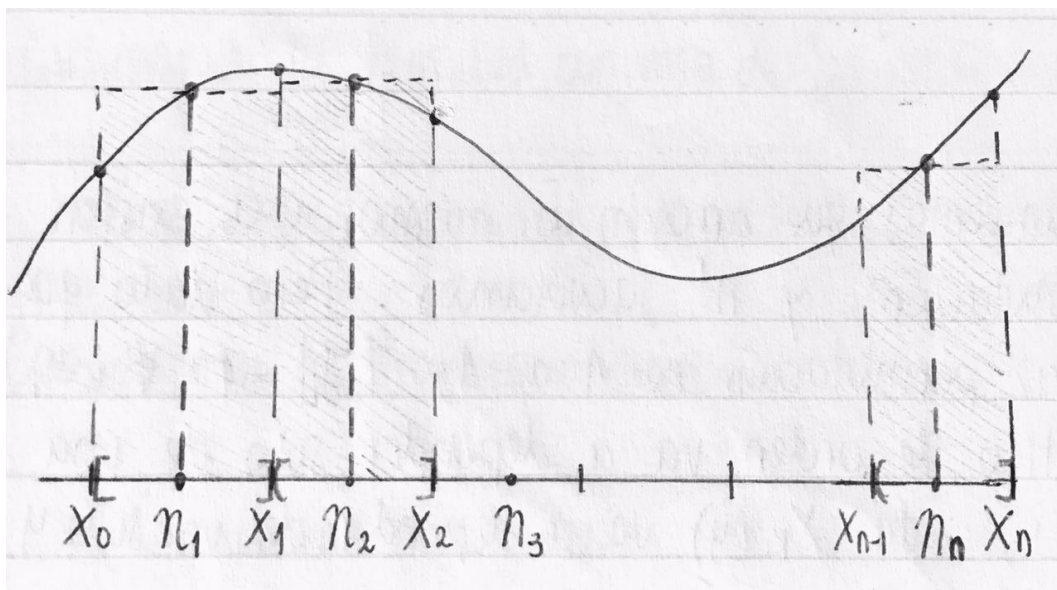
Unidad 5. Momentos de variables aleatorias.

5.0.1. Integral de Riemann

Sea $g : [a, b] \rightarrow \mathbb{R}$

Consideramos una partición de $[a, b]$ que llamaremos $\pi = \{x_0, \dots, x_n\}$, tal que $a = x_0 < x_1 < \dots < x_n = b$

Sea $\eta = \{\eta_i\}_{1 \leq i \leq n}$ una colección de puntos tal que $\eta_i \in (x_{i-1}, x_i]$ para $i = 1, \dots, n$ se denomina selección en π



Se define la *Suma de Riemann* como:

$$S_a^b(\pi, \eta, g) = \sum_{i=1}^n g(\eta_i)(x_i - x_{i-1})$$

Es decir, es la suma de n rectángulos y representa una aproximación del área entre la función y el eje de las abscisas y entre a y b .

La forma de mejorar esa aproximación es aumentando la cantidad de puntos que contiene la partición.

Se define la *norma* de la partición como la mayor longitud de los intervalos:

$$\|\pi\| = \max_{1 \leq i \leq n} \{x_i - x_{i-1}\}$$

Definición:

g es integrable Riemann sobre $[a, b]$ con valor $I = \int_a^b g = \int_a^b g(x) dx$, sii para cada $\varepsilon > 0$ existe un $\delta > 0$ tal que si $\|\pi\| < \delta$, entonces:

$$|S_a^b(\pi, \eta, g) - I| < \varepsilon$$

5.0.2. Integral de Riemann - Stieltjes

Dadas las funciones g y F definidas sobre $[a, b]$, se define la *Suma de Riemann-Stieltjes* asociada a la partición π y la selección η de π por:

$$S_a^b(\pi, \eta, g, F) = \sum_{i=1}^n g(\eta_i)[F(x_i) - F(x_{i-1})]$$

Existe la integral de Riemann - Stieltjes sobre $[a, b]$ con valor $I = \int_a^b g dF = \int_a^b g(x) dF(x)$ sii para cada $\varepsilon > 0$ existe un $\delta > 0$ tal que si $\|\pi\| < \delta$, entonces:

$$|S_a^b(\pi, \eta, g, F) - I| < \varepsilon$$

La integral de Riemann es un caso particular de la de R-S.

Es suficiente para la existencia de la integral de R-S que g sea continua en $[a, b]$ y F monótona en $[a, b]$

Resultados principales

Si $g(x) = X$ y $F(x) = F_X(x)$

Sea X v.a. discreta con p_X y $\sum_{X \in R_X} |x| p_X(x) < \infty$

$$\int_a^b g(x) dF(x) = \int_a^b x dF_X(x) = \sum_{x \in R_X \cap [a,b]} x p_X(x)$$

Si definimos $a = -\infty$ y $b = \infty$

$$\int_{-\infty}^{\infty} x dF_X(x) = \sum_{x \in R_X} x p(x) \equiv E(X) \Rightarrow \text{Esperanza de una v.a. discreta}$$

Sea X v.a. continua:

$$\int_{-\infty}^{\infty} x dF_X(x) = \int_{-\infty}^{\infty} x f_X(x) dx \equiv E(X) \Rightarrow \text{Esperanza de una v.a. continua}$$

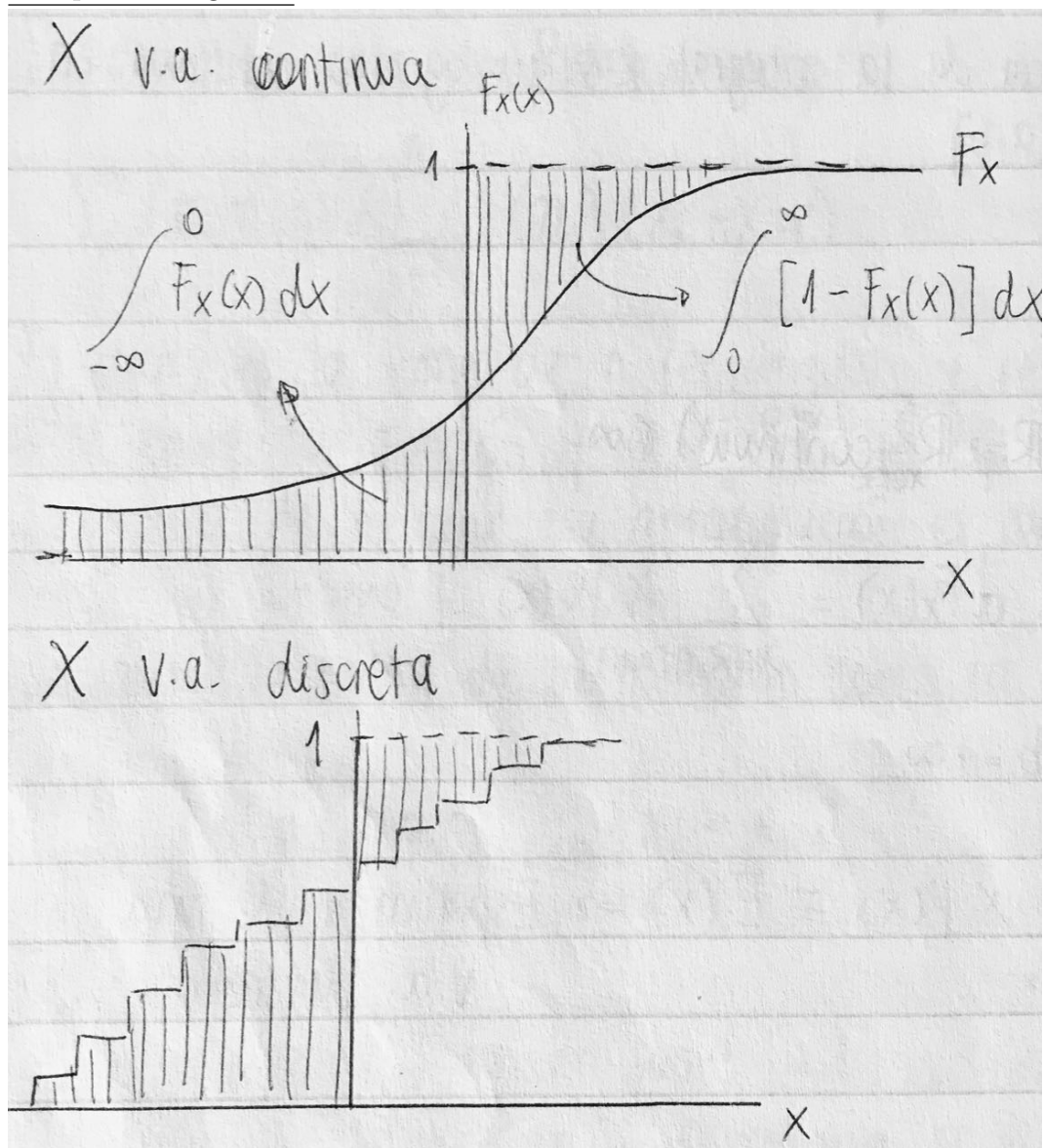
Teorema 7 Si $\int |x| dF(x) < \infty$, entonces:

$$\left. \begin{array}{l} 1) \lim_{x \rightarrow \infty} x(1 - F_X(x)) = 0 \\ 2) \lim_{x \rightarrow -\infty} xF_X(x) = 0 \end{array} \right\} \begin{array}{l} \text{Condición suficiente para que} \\ E(X) \text{ sea finita.} \end{array}$$

Teorema 8 Si $\int |x| dF(x) < \infty$, entonces: **(XL)**

$$EX = \int_0^{\infty} (1 - F_X(x)) dx - \int_{-\infty}^0 F_X(x) dx$$

Interpretación gráfica



Sea \tilde{X} un vector aleatorio de k componentes y $g: \mathbb{R}^k \rightarrow \mathbb{R}$, el siguiente teorema nos dice como calcular la esperanza de $Y = g(X)$ sin necesidad de obtener primero $f_Y(y)$

Caso discreto

Sea \tilde{X} un vector aleatorio discreto de dimensión k y sea $g: \mathbb{R}^k \rightarrow \mathbb{R}$ una función continua. Definase $Y = g(\tilde{X})$, entonces:

$$EY = \sum_{x \in R_X} g(x) p_X(x)$$

Ejemplo

$\tilde{X} = (X, Y)$ vector aleatorio discreto

$$Z = X + Y = g(X, Y)$$

$$EZ = \sum_{(x,y) \in R_{X,Y}} (x + y) p_{(x,y)}(x, y)$$

Si $g(X, Y) = X$

$$EX = \sum_{(x,y) \in R_{X,Y}} x p_{(x,y)}(x, y)$$

Por lo tanto, no es necesario encontrar las funciones de probabilidad marginal de las variables del vector para calcular su esperanza.

Caso continuo

Sea \tilde{X} un vector aleatorio absolutamente continuo de dimensión k con función de densidad $f_{\tilde{X}}$ y sea $g: \mathbb{R}^k \rightarrow \mathbb{R}$ una función continua. Definase $Y = g(\tilde{X})$, entonces:

$$EY = \underbrace{\int \cdots \int}_{Sop_{X,Y}} g(x) f_{\tilde{X}}(x) dx$$

5.0.3. Propiedades de la Esperanza

- (XLI) 1) Si $X = c$, es decir, $X(\omega) = c \ \forall \omega \in \Omega$, entonces $EX = c$
- (XLII) 2) Si $X \leq Y$, es decir, $X(\omega) \leq Y(\omega) \ \forall \omega \in \Omega$, entonces $EX \leq EY$
- (XLIII) 3) Si EX existe, entonces $E(aX + b) = aEX + b \ \forall a, b \in \mathbb{R}$
- (XLIV) 4) si EX y EY existen, entonces $E(aX + bY) = aEX + bEY$
- (XLV) 5) Desigualdad de Jensen.

Sea Ψ una función convexa definida en \mathbb{R} . Si EX es finita, entonces:

$$E\Psi(x) \geq \Psi(EX)$$

Si Ψ es cóncava, entonces $E\Psi(x) \leq \Psi(EX)$

- (XLVI) 6) Si X y Y son v.a. independientes con esperanza finita, entonces:

$$EXY = EXEY$$

5.1. Momentos de las variables aleatorias.

Sea $g: \mathbb{R} \rightarrow \mathbb{R}$

$$g(x) = X^k \quad k \in \mathbb{N}$$

EX^k se denomina *momento de orden k* de la v.a. X y simbolizaremos como

$$EX^k = \mu'_k$$

Si estamos interesados en calcular los momentos respecto a la media, $E(X - \mu'_1)^k$, los simbolizaremos con μ_k

Si una v.a. tiene momento de orden r , en el sentido que $E|X^r| < \infty$, entonces existen todos los momentos de orden k , para $k \leq r$

En particular, podemos decir que:

$$\mu_1 = 0$$

$$\mu_2 = \mu'_2 - (\mu'_1)^2 = VX$$

Además, cualquier momento centrado en la esperanza se puede expresar como combinación lineal de momentos centrados en el origen.

$$\mu_k = E(X - \mu'_1)^k$$

Aplicamos el Binomio de Newton

$$\begin{aligned} &= E \left(\sum_{i=1}^k \binom{n}{i} X^i (\mu'_1)^{k-i} \right) \\ &= \sum_{i=1}^k \binom{k}{i} EX^i (\mu'_1)^{k-i} \\ \mu_k &= \sum_{i=1}^k \binom{k}{i} \mu'_i (\mu'_1)^{k-i} \end{aligned}$$

5.1.1. Propiedades de la Varianza

1) Si $X(\omega) = c \quad \forall \omega \in \Omega$, entonces $VX = 0$

2) $V(aX) = a^2 VX$

(XLVII) 3) $VX = \min_{\{a \in \mathbb{R}\}} E(X - a)^2$ y $a^* = EX$

(XLVIII) 4) La varianza no es lineal respecto de la suma:

$$V(X + Y) = VX + VY + 2E[(X - EX)(Y - EY)]$$

Donde $E[(X - EX)(Y - EY)] = \text{covarianza de } X \text{ e } Y$. Operando se obtiene:

$Cov(X, Y) = EXY - EXEY \rightarrow$ es una medida de dependencia lineal de las v.a.

La $Cov(X, Y) = 0$ cuando $EXY = EXEY$, es decir, si X e Y son independientes, y por lo tanto se cumple la linealidad de la varianza.

Puede darse el caso que $Cov(X, Y) = 0$ sin ser independientes, por el hecho de que simplemente las variables no estén relacionadas.

Definida la covarianza, se define el *coeficiente de correlación* entre X e Y como:

$$\rho(x, y) = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$$

donde $\sigma_i = \sqrt{VI}$.

Utilizando la versión de la *desigualdad de Cauchy - Schwartz* para v.a.:

$$[EXY]^2 \leq EX^2 EY^2$$

se puede concluir que $\rho(x, y) \in [-1, 1]$

5.2. Función Característica

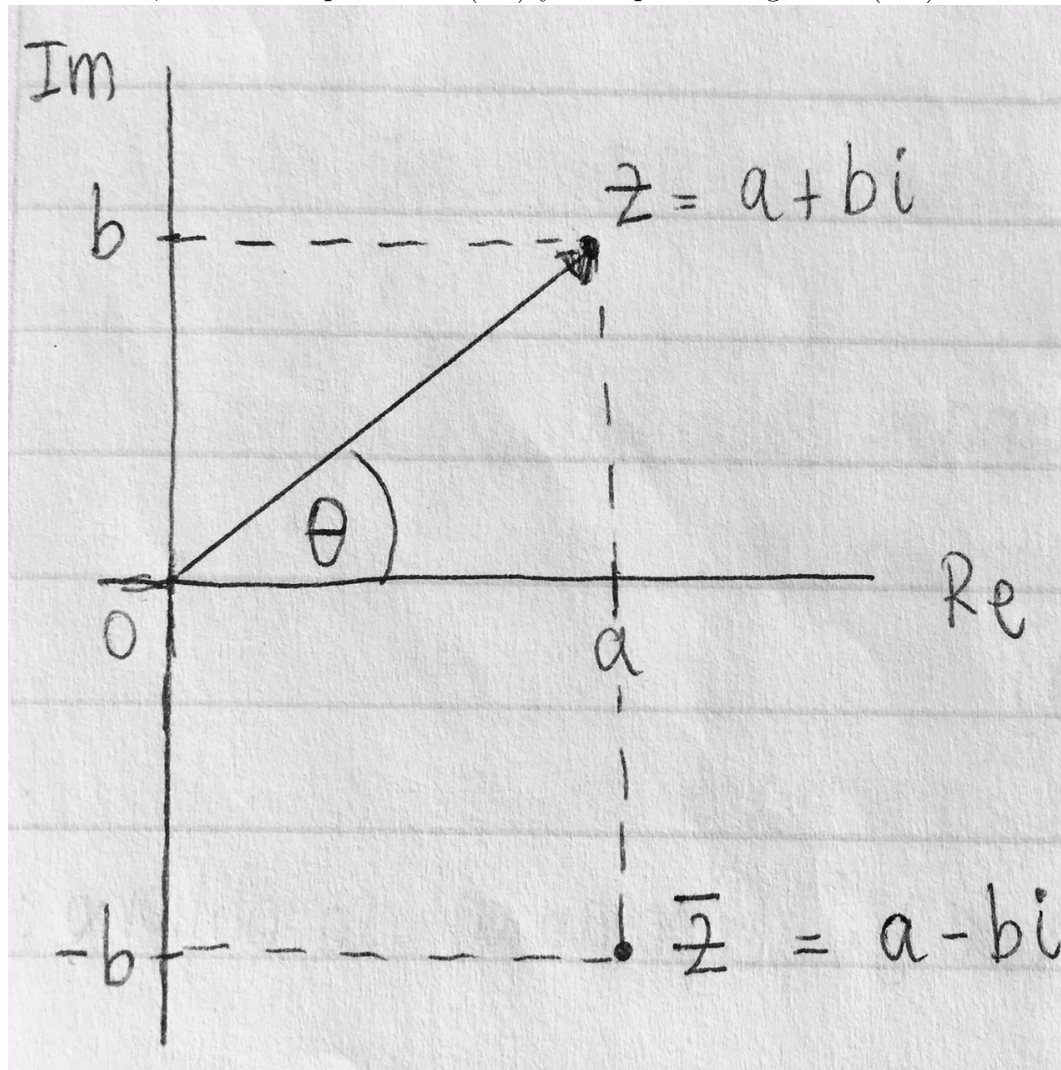
Es una función que nos permite encontrar los momentos de la variable sin necesidad de integrar $x^k f(x) dx$, simplemente derivando dicha función.

La f.c. está definida sobre los complejos.

5.2.1. Números complejos

En general, un número complejo tiene la forma:

$Z = a + bi$, donde a = parte real (Re) y bi = parte imaginaria (Im)



Se puede representar a Z por medio de un segmento orientado, y como todo segmento orientado, se puede definir a partir de su longitud (módulo) y su orientación (ángulo).

Sea $|Z|$ el módulo de Z y θ el ángulo del triángulo rectángulo

$$|Z| = \sqrt{a^2 + b^2} \text{ (por Teorema de Pitágoras)}$$

$$\operatorname{sen} \theta = \frac{b}{|Z|} \rightarrow b = |Z| \operatorname{sen} \theta$$

$$\operatorname{cos} \theta = \frac{a}{|Z|} \rightarrow a = |Z| \operatorname{cos} \theta$$

$$Z = a + bi = |Z| \cos \theta + i |Z| \sen \theta$$

$$Z = |Z| (\cos \theta + i \sen \theta)$$

Tomando módulo a ambos miembros

$$|Z| = | |Z| (\cos \theta + i \sen \theta) |$$

$$|Z| = |Z| |(\cos \theta + i \sen \theta)|$$

Donde $|(\cos \theta + i \sen \theta)| = 1$.

Consideremos las siguientes funciones:

$F(\theta)$	$F(0)$	$F'(0)$	$F''(0)$	$F'''(0)$
$\cos \theta$	$\cos(0) = 1$	$-\sen(0) = 0$	$-\cos(0) = -1$	$\sen(0) = 0$
$i \sen \theta$	$i \sen(0) = 0$	$i \cos(0) = i$	$-i \sen(0) = 0$	$-i \cos(0) = -i$
$e^{i\theta}$	$e^{i0} = 1$	$ie^{i0} = i$	$i^2 e^{i0} = i^2 = -1$	$-ie^{i0} = -i$

Si uno realiza la expansión de $e^{i\theta}$ por serie de Mc Laurin, los términos pares de esa expansión se corresponden con la expansión de $\cos \theta$ y los impares con los de $i \sen \theta$.

Por lo tanto, uno puede expresar la función $e^{i\theta} = \cos \theta + i \sen \theta$.

Entonces, ahora podemos expresar a $|Z|$ como:

$$Z = |Z| e^{i\theta} \implies |e^{i\theta}| = 1 \text{ (número complejo acotado)}$$

Volviendo a la función característica:

Sean X y Y v.a. en (Ω, \mathcal{F}, P) , entonces $Z = X + iY$ se denomina v.a. compleja. Así, Z es una función definida sobre Ω que asume valores complejos, con $Z(\omega) = X(\omega) + iY(\omega) \quad \forall \omega \in \Omega$

Si EX y EY son finitas, $EZ = EX + iEY$

Definición

Sea X una v.a. La función característica de X es la función $\varphi: \mathbb{R} \rightarrow \mathbb{C}$, definida por:

$$\varphi_X(t) = E e^{itX}$$

$$\varphi_X(t) = \int e^{itx} dF_X(x)$$

Dijimos que $e^{itx} = \cos(tx) + i \sen(tx)$. Entonces e^{itx} siempre va a tener esperanza finita porque $E e^{itx} = E[\cos(tx)] + i E[\sen(tx)]$, y tanto $E(\cos)$ y $E(\sen)$

son finitas.

La función de distribución determina la función característica: si X e Y están idénticamente distribuidas, entonces $\varphi_X = \varphi_Y$

Si X v.a. discreta, entonces: $\varphi_X(t) = \sum_{x \in R_X} e^{itx} p_X(x)$

Si X v.a. continua, entonces: $\varphi_X(t) = \int_{\text{Sup}_X} e^{itx} f_X(x) dx$

5.2.2. Propiedades de la función característica

(XLIX) 1) $|\varphi_X(t)| \leq 1 \quad \forall t \in \mathbb{R}$

(L) 2) $\varphi_X(0) = 1$

(LI) 3) $\overline{\varphi_X(t)} = \varphi_X(-t)$

4) $\varphi_X(t)$ es continua

(LII) 5) Si X y Y son v.a. independientes, entonces: $\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t)$

(LIII) 6) Si $Y = aX + b$, entonces $\varphi_Y(t) = e^{itb} \varphi_X(at)$

(LIV) 7) Si $E|X|^n < \infty$, entonces φ_X posee n derivadas continuas con respecto

a t y se cumple que:

$$i) \varphi_X^k(t) = \int (ix)^k e^{itx} dF_X(x)$$

$$ii) \varphi_X^k(0) = i^k EX^k$$

iii) φ_X admite representación por polinomio de McLaurin de orden n

$$\varphi_X(t) = 1 + \sum_{k=1}^n \frac{(it)^k}{k!} EX^k + o(|t|^n)$$

5.3. Momentos condicionados

Caso discreto

Sea Y una v.a. y \tilde{X} un vector aleatorio k -dimensional, ambos discretos. La esperanza condicional de Y condicionada a $X = x$ se define como la esperanza de Y utilizando como función de probabilidad a $p_{Y|X}(y|x)$. Es decir:

$$E(Y|X = x) = \sum_{y \in R_Y} y p_{Y|X}(y|x)$$

Ese valor representa la esperanza de la v.a. Y cuando se sabe que el vector \tilde{X} ha tomado el valor x .

Como trabajamos con $P(Y|X = x) \Rightarrow \frac{p_{X,Y}(x, y)}{p_X(x)} = p_{Y|X}(y|x)$

Simolicemos $g(X) = E(Y|X = x)$, luego $g : R_X \rightarrow \mathbb{R}$. Eso si ya se ha observado la realización de $X = x$

Si el experimento se lleva a cabo sin conocer el valor que asume X , entonces $g(X) = E(Y|X)$, se denomina esperanza de Y condicionada a X , y es una variable(y por lo tanto se le puede tomar esperanza).

Teorema 9 (LV) Si Y es una v.a. con esperanza finita, entonces se cumple que:

$$E(E(Y|X)) = EY$$

Teorema 10 (LVI). Si Y es una v.a. con esperanza finita, y es independiente de X , entonces:

$$E(Y|X) = EY$$

Teorema 11 Sean \tilde{X} e \tilde{Y} vectores aleatorios discretos de dimensión k y j y sea h una función real continua definida sobre \mathbb{R}^{k+j} . Definimos la v.a. discreta $Z = h(x, y)$ y suponemos que tiene esperanza finita. Entonces, para todo $x \in R_X$ se tiene que:

$$E(Z|X = x) = \sum_{y \in R_Y} h(x, y) p_{Y|X}(y|x)$$

Como X está fijo $X = x$, entonces Z termina siendo una función solo de y . $Z = h(y)$, por eso es que el sumatorio se realiza sobre el rango de Y y se utiliza $p_{Y|X}(y|x)$.

El resultado anterior se utiliza para demostrar que la esperanza condicionada es también un operador lineal.

Si Y_1 e Y_2 son v.a. discretas con esperanza finita, y \tilde{X} un vector aleatorio discreto, entonces:

$$E(\underbrace{c_1 Y_1 + c_2 Y_2}_{h(y_1, y_2)} | X = x) = c_1 E(Y_1 | X = x) + c_2 E(Y_2 | X = x)$$

También puede demostrarse que:

(LVII) Si $P(Y \geq 0) = 1$, entonces $E(Y|X = x) \geq 0$

$E(Y^2|X = x) \geq [E(Y|X = x)]^2 \rightarrow$ Desigualdad de Jensen para esperanzas condicionadas.

Caso continuo Todo lo visto para el caso discreto se aplica al caso continuo, reemplazando los sumatorios por integrales y utilizando la función de densidad condicional en vez de la f. de probabilidad condicional.

Así:

$$E(Z|X = x) = \underbrace{\int \cdots \int}_{Sop_Y} h(x, y); f_{Y|X}(y|x) dy$$

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

Además, la probabilidad de que $Y \in B$, con $B \in \mathcal{B}^j$, condicional a $X = x$, simbolizada $P_{Y|X}(B|X = x)$ es igual a:

$$P_{Y|X}(B|X = x) = \int \cdots \int_B f_{Y|X}(y|x) dy$$

Podemos expresar una probabilidad como la esperanza de un tipo de v.a. y de esa manera se puede asimilar cualquier esperanza condicionada como una probabilidad condicionada: y por lo tanto el teorema de la esperanza condicionada se puede aplicar a una probabilidad condicionada para obtener una probabilidad condicional.

El tipo de variable que se debe definir es del tipo *Bernoulli*.

$$X \sim Be(p)$$

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$

$$EX = 0(1 - p) + 1(p) = p$$

Supongamos que estamos interesados en el evento:

$$P(X = A)$$

Definimos una v.a. $Y \sim Be$

$$Y = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

$$\begin{aligned} EY &= 0 P_Y(0) + 1 P_Y(1) = P_Y(1) \\ &= P(x \in A) \end{aligned}$$

Por lo tanto, se puede expresar cualquier probabilidad como una esperanza de una Bernoulli cuyo éxito sea el evento cuya probabilidad se quiere calcular.

Además, cualquier probabilidad condicional se puede también expresar como una esperanza.

Si el evento que nos interesa es:

$$P(Y \in B|X = x)$$

Definimos

$$Z = h(x, y) = \begin{cases} 1 & \text{si } y \in B \\ 0 & \text{si } y \notin B \end{cases}$$

$$Z = 1_B(y)$$

Sabemos que

$$\begin{aligned} E(Z|X = x) &= \int \cdots \int h(x, y) f_{Y|X}(y|x) dy \\ &= \int \cdots \int 1_B(y) f_{Y|X}(y|x) dy \\ &= \int \cdots \int_B f_{Y|X}(y|x) dy \end{aligned}$$

Además, sabemos que:

$$\begin{aligned} P(Y \in B|X = x) &= \int_b \frac{f(x, y)}{f_X(x)} dy = \int_B f_{Y|X}(y|x) dy \\ &= E(Z|X = x) \end{aligned}$$

Además, dado que:

$$E(E(Y|X)) = EY$$

tenemos que:

$$\begin{aligned} E(\underbrace{E(Z|X = x)}) &= E(\underbrace{1_B(y)}) \\ EP(Y \in B|X = x) &= P(Y \in B) \end{aligned}$$

Por lo tanto, la esperanza de una probabilidad condicionada es una probabilidad sin condicionar.

Definición

Sea \tilde{X} un vector aleatorio de k componentes e Y una v.a. con varianza finita.

Entonces, la varianza de Y condicional a $\tilde{X} = \tilde{x}$ se define como:

$$V(Y|\tilde{X} = \tilde{x}) = E((Y - E(Y|\tilde{X} = \tilde{x}))^2|\tilde{X} = \tilde{x})$$

Ese valor representa la varianza de la v.a. Y cuando se sabe que el vector \tilde{X} ha tomado el valor \tilde{x}

Definimos ahora una v.a. $V(Y|\tilde{X})$ que llamamos varianza de Y condicional a \tilde{X} :

$$V(Y|\tilde{X}) = E((Y - E(Y|\tilde{X}))^2|\tilde{X}) = q(\tilde{X})$$

Operando sobre el segundo miembro el cuadrado de un binomio, se tiene:

$$V(Y|\tilde{X}) = E(Y^2|\tilde{X}) - E^2(Y|\tilde{X})$$

Teorema 12 (LXXII) Sean X e Y v.a. y g una función. Entonces:

$$E(Y - g(X))^2 = EV(Y|X) + E(E(Y|X) - g(X))^2$$

Entonces:

$$VY = EV(Y|X) + VE(Y|X)$$

5.3.1. Aplicación. Parámetros aleatorios.

Supongamos que la cantidad de glóbulos rojos por milímetro cúbico, X , tiene distribución *Poisson* cuyo parámetro depende de características del individuo. Es decir, no solo X es una v.a., sino también el parámetro de su distribución.

Para una persona elegida al azar, se toma el valor del parámetro M como una v.a. tal que dado $M = m$, se tiene $X \sim Po(m)$

$$P(X = x|M = m) = \frac{e^{-m} m^x}{x!} 1_{\{0,1,2,\dots\}}(x)$$

El enfoque de la aplicación es bayesiano, ya que esta corriente considera que todas las fuentes de incertidumbre tienen que tener una distribución. Ahora, tenemos dos fuentes de incertidumbre: X y M .

Nosotros queremos saber la probabilidad de que la v.a. X asuma determinado valor, independientemente de las características de la unidad muestral observada. Es decir, se quiere encontrar $P(X = x)$.

Para ello también debemos conocer la distribución del parámetro. En este

ejemplo se propone $M \sim Exp(1)$. Lo cual tiene sentido ya que el parámetro de la *Poisson* $\in \mathbb{R}_{++}$, y la distribución exponencial tiene soporte en \mathbb{R}_{++} . Sabemos que toda esperanza de una probabilidad condicionada es una probabilidad sin condicionar. Por lo tanto:

$$P(X = x) = E \underbrace{P(X = x|M = m)}_{g(m)} = \int_{Sop_M} P(X = x|M = m) f_M(m) dm$$

Dado que $M \sim Exp(1) \Rightarrow f_M(m) = e^{-m} 1_{(0,\infty)}(m)$

Por lo tanto

$$P(X = x) = \int_0^\infty \frac{e^{-m} m^x}{x!} e^{-m} dm$$

Este procedimiento aplicando el teorema de la esperanza condicionada se realiza porque la naturaleza de la variable X y del parámetro M no es la misma: X es discreta y M es continua.

Si X y M tuvieran la misma naturaleza, por ejemplo, ambas continuas, el procedimiento para hallar $f_X(x)$ sería mucho mas simple. Sabemos que:

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

Para hallar la densidad marginal $f_X(x)$:

$$f_X(x) = \int_{Sop_Y} f_{x,y}(x,y) dy = \int f_{X|Y}(x|y) f_Y(y) dy$$

Eso no se puede hacer, entonces se utiliza la esperanza *iterada*. Volviendo al ejemplo: $P(X = x) = \int_0^\infty \frac{e^{-m} m^x}{x!} e^{-m} dm$

Podemos manipular la expresión para llegar a una *Gamma*

$$\frac{e^m m^x}{x!} e^m = \frac{m^{(x+1)-1} e^{-2m}}{\Gamma(x+1)}$$

Que podría ser una *Gamma* con $\alpha = x + 1$ y $\beta = 1/2$ pero falta el término $(1/2)^{k+1}$ en el denominador.

Multiplicamos y dividimos por 2^{k+1}

$$P(X = x) = \int_0^\infty \frac{m^{(x+1)-1} e^{-2m}}{\Gamma(x+1)} \frac{2^{k+1}}{2^{k+1}} dm = \frac{1}{2^{k+1}} \int_0^\infty \underbrace{\frac{m^{(x+1)-1} e^{-2m}}{\Gamma(x+1)}}_{=1} 2^{k+1} dm$$

Entonces: $P(X = x) = \frac{1}{2^{k+1}} = \frac{1}{2} \left(\frac{1}{2}\right)^k \Rightarrow X \sim Geo(1/2)$

También se pueden obtener los momentos de X sin conocer su distribución:

Dado que $X|M = m \sim Po(m)$, entonces $E(X|M) = M$ y $V(X|M) = M$

Como $EX = EE(X|M) = EM = 1$ (Ya que $M \sim Exp(1)$)

Como $VX = EV(X|M) + VE(X|M) = EM + VM = 1 + 1 = 2$

Capítulo 6

Unidad 6. Teoría asintótica

Vamos a exponer resultados relacionados con las características probabilísticas de sucesiones de v.a. y de funciones de vectores aleatorios de n componentes, cuando n es grande. En particular, ciertos tipos de funciones de vectores aleatorios, digamos $Y_n = g(X_1, \dots, X_n)$, pueden converger de diferentes maneras a una constante o la distribución de $g(X_n)$ puede aproximarse a una distribución límite cuando $n \rightarrow \infty$

Existen al menos tres motivos para estudiar el comportamiento asintótico de $g(X_n)$:

- 1) Funciones del tipo $g(X_n)$ serán utilizadas en procedimientos de estimación puntual y test de hipótesis y n simbolizará el número de observaciones asociadas al experimento analizado. Para ello es necesario determinar la distribución de $g(X_n)$.
- 2) La teoría asintótica proporciona métodos para obtener aproximaciones de la distribución de $g(X_n)$ cuando n es suficientemente grande
- 3) La T.A. proporciona el fundamento principal del uso dominante de la distribución normal en el análisis estadístico.

En el contexto de la teoría de la probabilidad existen varias nociones de convergencia. Algunos de ellos son:

6.1. Modos de convergencia

6.1.1. Convergencia casi cierta

$\{X_n\}$ converge casi ciertamente (c.c.) a la v.a. X cuando $n \rightarrow \infty$ sii

$$P(\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega) \text{ cuando } n \rightarrow \infty\}) = 1$$

Es decir, en la convergencia casi cierta se permite que para algunos valores de ω , la sucesión numérica $X_1(\omega), X_2(\omega), \dots$ pueda no converger. Sin embargo, el subconjunto de Ω en donde esto suceda debe tener probabilidad cero. La convergencia c.c. es convergencia puntual con probabilidad 1.

Simbolizaremos la convergencia casi cierta como $X_n \xrightarrow{\text{c.c.}} X$

6.1.2. Convergencia en Probabilidad

$\{X_n\}$ converge en Probabilidad a la v.a. X cuando $n \rightarrow \infty$ sii

$$\forall \varepsilon > 0 \quad P(|X_n - X| > \varepsilon) \rightarrow 0 \text{ cuando } n \rightarrow \infty, \text{ o}$$

$$\lim_{n \rightarrow \infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0$$

Tenemos otra noción de proximidad pero más débil. Esta convergencia no dice nada respecto a la convergencia puntual.

Para valores de n suficientemente grandes, las variables X_n y X son iguales con probabilidad bien alta.

Simbolizaremos esta convergencia con $X_n \xrightarrow{p} X$ o $p\lim X_n = X$

Ejemplo:

Dada una sucesión $\{X_n\}_{n=1}^\infty$ tal que puede asumir dos valores: 1 o n

$$P(X_n = 1) = 1 - 1/n$$

$$P(X_n = n) = 1/n$$

Y dada la v.a. $X(\omega) = 1 \forall \omega \in \Omega$, probar que $X_n \xrightarrow{p} 1$

Para ello se debe demostrar que $P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0$

Nuestros ω ahora son: $\{\omega \in \Omega : |X_n(\omega) - 1| > \varepsilon\}$

$$\{\omega \in \Omega : -\varepsilon > X_n(\omega) - 1 > \varepsilon\}$$

$$\{\omega \in \Omega : \underbrace{1 - \varepsilon}_{\text{nunca se da}} > X_n(\omega) > \underbrace{\varepsilon + 1}_{\text{La unica realizacion de } X_n(\omega) > 1 \text{ es } n}\}$$

Por lo tanto, el evento es en realidad:

$$\{\omega \in \Omega : X_n(\omega) = n\}$$

y la probabilidad de ese evento:

$$P(\{\omega \in \Omega : X_n(\omega) = n\}) = P(X_n = n) = \frac{1}{n}$$

Tomamos límite: $\lim_{n \rightarrow \infty} P(|X_n - 1| > \varepsilon) = \lim_{n \rightarrow \infty} P(X_n = n) = \lim_{n \rightarrow \infty} \frac{1}{n} = 0$ QED.

Las sucesiones de números reales se pueden asimilar como sucesiones de v.a. degeneradas. Es decir, considerando $X_n(\omega) = a_n$ y $X(\omega) = a$, entonces si $a_n \rightarrow a$ se puede decir que $a_n \xrightarrow{p} a$

6.1.3. Convergencia en media cuadrática

$\{X_n\}$ converge en media cuadrática a la v.a. X cuando $n \rightarrow \infty$ sii

$$E(X_n - X)^2 \rightarrow 0 \text{ cuando } n \rightarrow \infty \text{ o}$$

$$\lim_{n \rightarrow \infty} E(X_n - X)^2 = 0$$

Como $E(X_n - X)^2$ puede interpretarse como la distancia al cuadrado esperada entre las realizaciones de X_n y las de X , la c.m.c. implica que las realizaciones de X_n y las de X están cerca unas de otra cuando n es grande y arbitrariamente cerca cuando $n \rightarrow \infty$

Simbolizaremos la c.m.c. escribiendo $X_n \xrightarrow{m} X$

6.1.4. Convergencia en distribución.

$\{X_n\}$ converge en distribución a la v.a. X cuando $n \rightarrow \infty$ sii

$$F_{X_n}(x) \rightarrow F_X(x) \text{ o}$$

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad \forall x \in C(F_X)$$

donde $C(F_X) = \{x : F_X(x) \text{ es continua en } x\}$

La c.d. es la única noción de convergencia que no exige que X_n y X estén definidas sobre el mismo espacio de probabilidad.

Simbolizaremos esta convergencia con $X_n \xrightarrow{d} X$. También se puede decir que F_{X_n} converge *débilmente* a F_X . F_X es la distribución límite.

La utilidad del concepto de c.d. radica en establecer una aproximación a la

verdadera función de distribución de X_n cuando n es suficientemente grande. Puede suceder que la v.a. asociada a F_X sea degenerada, en este caso la sucesión de v.a. converge en distribución a una constante. Si las X_n no son degeneradas y convergen a una X degenerada, entonces F_X es inadecuada para aproximar ciertas características de la distribución de X_n .

El concepto de aproximar la distribución de X_n , para n grande, puede generalizarse para incluir casos donde la sucesión $\{X_n\}$ no tiene distribución límite o la d.l. es degenerada. Se introduce, para ello, el concepto de *distribución asintótica*.

6.2. Distribución asintótica

Sea la sucesión de v.a. $\{Z_n\}$ definidas por $Z_n = g(X_n, \theta_n)$ donde $X_n \xrightarrow{d} X$ para una v.a. X no degenerada y $\{\theta_n\}$ es una sucesión de números reales, vectores o parámetros. Entonces la distribución asintótica de Z_n está dada por la distribución de $g(X, \theta_n)$, simbolizada por $Z_n \overset{a}{\sim} g(X, \theta_n)$ que significa que Z_n está distribuida asintóticamente como $g(X, \theta_n)$

Ejemplo:

Dada la sucesión $X_n = \sqrt{n}(Z_n - k)$

$$X_n \xrightarrow{d} X \sim N(0, \sigma^2)$$

Despejando Z_n tenemos:

$$Z_n = \frac{X_n}{\sqrt{n}} + k \equiv g(X_n, \theta_n) \text{ donde } \sqrt{n} \text{ es la parte } \theta_n \text{ de } g$$

Vamos a ver mas adelante que la distribución límite de Z_n es degenerada, por eso utilizamos la distribución asintótica.

$$g(X, \theta_n) = \frac{X}{\sqrt{n}} + k \Rightarrow g(X, \theta_n) \sim N \text{ (se puede probar por jacobiano)}$$

$$\left. \begin{aligned} Eg(X, \theta_n) &= \frac{EX}{\sqrt{n}} + k = k \\ Vg(X, \theta_n) &= V\left(\frac{X}{\sqrt{n}} + k\right) = \frac{VX}{n} = \frac{\sigma^2}{n} \end{aligned} \right\} \begin{aligned} &g(X, \theta_n) \sim N(k, \sigma^2/n) \\ &\text{por lo tanto} \\ &Z_n \overset{a}{\sim} N(k, \sigma^2/n) \end{aligned}$$

En el caso de v.a. discretas no negativas que toman valores enteros o v.a. continuas, la convergencia de la sucesión de las funciones de densidad es suficiente

para establecer la convergencia en distribución.

Teorema 13 Sea X_n una sucesión de v.a. discretas a valores enteros no negativos o continuas, y sea f_n la sucesión asociada de funciones de probabilidad o densidad, respectivamente. Supongamos que existe una f de probabilidad o densidad f tal que $f_n(x) \rightarrow f(x)$ para todo x , excepto, quizás, para un conjunto de puntos A tal que $P_X(A) = 0$ en el caso continuo, donde $X \sim f$.

Entonces, $X_n \xrightarrow{d} X$ y f se denomina densidad límite de X_n

Ejemplo (caso continuo)

Dada la sucesión $X_n \sim N(0, 1)$ y la sucesión $Z_n = (3 + n^{-1})X_n + 2n/(n-1) \sim N(\text{por jacobiano})$

$$\left. \begin{aligned} EZ_n &= (3 + n^{-1}) \underbrace{EX_n}_{=0} + \frac{2n}{n-1} = \frac{2n}{n-1} \\ VZ_n &= (3 + n^{-1})^2 \underbrace{VX_n}_{=1} = (3 + n^{-1})^2 \end{aligned} \right\} Z_n \sim N\left(\frac{2n}{n-1}, (3 + n^{-1})^2\right)$$

Entonces:

$$f_{Z_n}(z) = \frac{1}{\sqrt{2\pi}(3 + n^{-1})} \exp \left[-\frac{1}{2} \frac{(z - 2n/(n-1))^2}{(3 + n^{-1})^2} \right]$$

Queremos saber a dónde converge Z_n en distribución. Para eso se toma límite a la densidad de Z_n para $n \rightarrow \infty$ y se observa si la densidad resultante es una conocida:

$$\lim_{n \rightarrow \infty} f_{Z_n}(z) = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}(3 + n^{-1})} \exp \left[-\frac{1}{2} \frac{(z - 2n/(n-1))^2}{(3 + n^{-1})^2} \right]$$

Aplicando propiedades de límite:

$$\lim_{n \rightarrow \infty} f_{Z_n}(z) = \frac{1}{\sqrt{2\pi}(3 + \lim_{n \rightarrow \infty} n^{-1})} \exp \left[-\frac{1}{2} \frac{(z - \lim_{n \rightarrow \infty} 2n/(n-1))^2}{(3 + \lim_{n \rightarrow \infty} n^{-1})^2} \right]$$

Analicemos los términos afectados por el límite:

$$\{n^{-1}\} = \left\{ \frac{1}{n} \right\} \xrightarrow{n \rightarrow \infty} 0$$

$$\lim_{n \rightarrow \infty} \frac{2n}{n-1} = 2$$

Por lo tanto:

$$\lim_{n \rightarrow \infty} f_{Z_n}(z) = \frac{1}{\sqrt{2\pi}3} \exp \left[-\frac{1}{2} \frac{(z-2)^2}{3^2} \right] \implies \text{es una normal con } \mu = 2, \sigma^2 = 9$$

$$Z_n \xrightarrow{d} Z \sim N(2, 9)$$

6.3. Dos desigualdades útiles en teoría asintótica

6.3.1. Desigualdad de Markov

(LVIII). Si X es una v.a. que toma valores no negativos, entonces, para cualquier $a > 0$ se verifica:

$$P(X > a) \leq \frac{EX}{a}$$

6.3.2. Desigualdad de Tchebycheff

(LIX) Si X es una v.a. con esperanza μ y varianza σ^2 entonces para cualquier $\varepsilon > 0$ se tiene que:

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

6.4. Ley débil de los Grandes Números (varianza finita)

Sea $\{X_n\}$ una sucesión de v.a.i.i.d con media μ y varianza σ^2 finita y definase: $S_n = \sum_{i=1}^n X_i$ con $n \geq 1$. Entonces: (LX)

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0 \quad \forall \varepsilon > 0$$

Esto equivale a decir que:

$$\frac{S_n}{n} \equiv \bar{X} \xrightarrow{p} \mu$$

Que significa que cuando la muestra sea muy grande, la media muestral va a acercarse mucho a μ con probabilidad bien alta.

6.5. Unicidad

La v.a. límite está definida de manera única en el siguiente sentido:

Si $X_n \xrightarrow{m,p,d} X$ y $X_n \xrightarrow{m,p,d} Y$, entonces $P(X = Y) = 1$.

Es decir, la v.a. límite es única

6.6. Relaciones entre los modos de convergencia

- 1) Si $X_n \xrightarrow{c.c.} X \implies X_n \xrightarrow{p} X$
(LXI) 2) Si $X_n \xrightarrow{m} X \implies X_n \xrightarrow{p} X$
3) Si $X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X$
(LXII) 4) Si $X_n \xrightarrow{p} c \iff X_n \xrightarrow{d} \delta(c)$

Donde c es una constante y $\delta(c)$ es la distribución de un punto en c

Es decir: $X \sim \delta(c)$ sii $X(\omega) = c \quad \forall \omega \in \Omega$

Además, la convergencia en media cuadrática se puede demostrar de la siguiente manera:

$X_n \xrightarrow{m}$ sii:

- a) $EX_n \rightarrow EX$
b) $VX_n \rightarrow VX$
c) $Cov(X_n, X) \rightarrow VX$

En particular, dada una v.a. degenerada $X = \delta(c)$, donde:

$$EX = c, VX = 0, Cov(X_n, x) = 0$$

Entonces si

$$EX_n \rightarrow c \text{ y } VX_n \rightarrow 0, X_n \xrightarrow{m} \delta(c) \text{ y también } X_n \xrightarrow{d,p} \delta(c)$$

6.7. Teorema de Helly - Bray

Teorema 14 (Helly - Bray) Sean $F_X, F_{X_1}, F_{X_2}, \dots$ funciones de distribución. Si F_{X_n} converge débilmente a F_X , entonces:

$$\int g(x) dF_{X_n}(x) \rightarrow \int g(x) dF_X(x)$$

Para toda g continua y acotada

Corolario:

Dada $g(x) = e^{itX}$ que es continua y acotada: $|e^{itX}| = 1$, entonces:

$$Ee^{itX_n} \equiv \varphi_{X_n}(t) \rightarrow \varphi_X(t) \equiv Ee^{itX} \quad \forall t \in \mathbb{R}$$

La recíproca de Helly - Bray también es válida. Por lo tanto, basta la convergencia de las funciones características asociadas a las funciones de distribución.

6.8. Teorema de continuidad de Lévy

Teorema 15 (de continuidad) Sean $F_X, F_{X_1}, F_{X_2}, \dots$ funciones de distribución y $\varphi_{X_1}, \varphi_{X_2}, \dots$ sus f.c. respectivamente. Si φ_{X_n} converge puntualmente a φ y φ es continua en 0, entonces:

- i) Existe una función de distribución F tal que $F_{X_n} \rightarrow F$ débilmente
- ii) φ es la f.c. de F

Los dos teoremas anteriores implican que $X_n \xrightarrow{d} X$ si $\varphi_{X_n} \rightarrow \varphi$; aunque el Teorema de Lévy es mas fuerte porque dice que el límite de una sucesión de f.c. es f.c. a condición de que la función límite sea continua en 0.

Corolario 1.

Sean X_1, X_2, \dots v.a. Si $\varphi_{X_n}(t) \rightarrow \varphi(t) \quad \forall t \in \mathbb{R}$ y φ es continua en 0, entonces φ es f.c. de alguna v.a. X digamos $\varphi = \varphi_X$ y $X_n \xrightarrow{d} X$

Corolario 2.

Sean X_1, X_2, \dots v.a. Si $\varphi_{X_n}(t) \rightarrow e^{\frac{-t^2}{2}} \quad \forall t \in \mathbb{R}$ entonces $X_n \xrightarrow{d} N(0, 1)$

Corolario 3.

Sean X_1, X_2, \dots v.a. Si $\varphi_{X_n}(t) \rightarrow e^{itc} \quad \forall t \in \mathbb{R}$, entonces $X_n \xrightarrow{p,d} c$

6.9. Ley débil de los grandes números

(LXIII) Sean X_1, X_2, \dots v.a.i.i.d con esperanza finita μ y definase: $S_n = X_1 + X_2 + \dots + X_n$ con $n \geq 1$. Entonces:

$$\overline{X_n} = \frac{S_n}{n} \xrightarrow{p} \mu$$

6.10. Teorema Central del Límite

(LXIV) Sean X_1, X_2, \dots v.a.i.i.d con esperanza y varianza finitas μ y σ^2 y definase $S_n = X_1 + X_2 + \dots + X_n$ con $n \geq 1$. Entonces:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0, 1)$$

Que es lo mismo que:

$$\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

Teorema 16 Sean X_1, X_2, \dots y Y_1, Y_2, \dots sucesiones de v.a. tal que:

$$X_n \xrightarrow{m,p} X \text{ y } Y_n \xrightarrow{m,p} Y, \text{ entonces } X_n + Y_n \xrightarrow{m,p} X + Y$$

Además, si X_n e Y_n son independientes para todo n , y X e Y son independientes:

$$X_n \xrightarrow{d} X \text{ y } Y_n \xrightarrow{d} Y \implies X_n + Y_n \xrightarrow{d} X + Y$$

También se va cumplir que:

$$X_n Y_n \xrightarrow{p} XY \text{ y } \frac{X_n}{Y_n} \xrightarrow{p} \frac{X}{Y} \text{ (siempre que } P(Y_n = 0) = P(Y = 0) = 0 \text{)}$$

Además, la convergencia en probabilidad y en distribución es preservada por funciones continuas:

Sea $\{X_n\}$ una sucesión de v.a. tal que $X_n \xrightarrow{p,d} X$ y sea la v.a. $g(X)$ definida por una función g que es continua, excepto, quizás, en un conjunto de probabilidad 0 por la distribución de X . Entonces:

$$g(X_n) \xrightarrow{p,d} g(X)$$

- Si X es una v.a. degenerada ($X \sim \delta(a)$), g debe ser continua en a .
- Si g no está definida en toda la recta, la proposición es válida si existe un abierto $A \subset \mathbb{R}$ con g continua en A y $P(X \in A) = 1$

Por ejemplo:

Dada una sucesión $X_n \rightarrow X \sim N(0, 1)$ y dada $g(X) = X^2$. Entonces:

$$g(X_n) = X_n^2 \rightarrow g(X) = X^2 \sim \chi_1^2$$

6.11. Teorema de Slutsky

Sean $\{X_n\}$ y $\{Y_n\}$ tales que $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{p} c$, donde c es una constante:

- i) $X_n + Y_n \xrightarrow{d} X + c$
- ii) $X_n Y_n \xrightarrow{d} X - c$
- iii) $Y_n X_n \xrightarrow{d} cX$
- iv) $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$ si $c \neq 0$ y $P(Y_n \neq 0) = 1$

6.12. Método delta

Cuando vimos la convergencia asintótica dijimos que la misma se utiliza cuando la distribución límite de una sucesión o no existe o es degenerada. Y vimos el ejemplo:

$$X_n \xrightarrow{d} X$$

$$X_n = \sqrt{n}(Z_n - k) \quad \text{y} \quad X \sim N(0, \sigma^2)$$

$$\begin{aligned} \text{Entonces } Z_n &= \frac{X_n}{\sqrt{n}} + k = g(X_n, \theta_n) \\ Z &= \frac{X}{\sqrt{n}} + k = g(X, \theta_n) \end{aligned}$$

$$\text{y } g(X, \theta_n) \sim N(k, \frac{\sigma^2}{n}) \implies Z_n \overset{d}{\approx} N(k, \frac{\sigma^2}{n})$$

¿Por qué usabámos la distribución asintótica? Porque, efectivamente, la distribución límite de Z_n es degenerada. Eso se sabe aplicando Slutsky. Hemos dicho que la convergencia de sucesiones de números reales pueden ser vistas como convergencias en probabilidad donde tanto la sucesión como la constante a a la que converge se consideran v.a. degeneradas.

Por lo tanto si:

$$\frac{1}{\sqrt{n}} \rightarrow 0 \implies \frac{1}{\sqrt{n}} \xrightarrow{p} 0$$

Aplicamos ahora el Teorema de Slutsky:

$$\text{Como } \sqrt{n}(Z_n - k) \xrightarrow{d} N(0, \sigma^2)$$

Entonces:

$$\frac{1}{\sqrt{n}} \sqrt{n}(Z_n - k) \xrightarrow{d} \underbrace{0 \ N(0, \sigma^2)}_{\delta(0)}$$

Por lo tanto

$$(Z_n - k) \xrightarrow{d} \delta(0) \iff (Z_n - k) \xrightarrow{p} 0$$

Además, dada la constante k :

$$k \rightarrow k \implies k \xrightarrow{p} k$$

Entonces, aplicando propiedades:

$$(Z_n - k) + k \xrightarrow{p} 0 + k$$

Y por lo tanto $Z_n \xrightarrow{p} k \sim \delta(k)$

Ahora si en lugar de la constante k tenemos μ , llegamos al *método delta*:

Sean $X_1, X_2 \dots$ v.a. tales que $\sqrt{n}(X_n - \mu) \xrightarrow{d} N(0, \sigma^2)$. Si $g(x)$ es una función derivable en μ , entonces

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{d} N(0, \sigma^2(g'(\mu))^2)$$

Ejemplo Dada $g(x) = x^2$ entonces $g'(x) = 2x$ es continua en μ ✓

$$\begin{aligned} \sqrt{n}(X_n^2 - \mu^2) &\xrightarrow{d} N(0, \sigma^2(2\mu)^2) \\ &\xrightarrow{d} N(0, \sigma^2\mu^2 4) \end{aligned}$$

Capítulo 7

Unidad 7. Introducción a la Teoría de la Estimación Puntual

7.1. Familias paramétricas de funciones de densidad

Muchas familias de distribuciones como la normal, $N(\mu, \sigma^2)$ o la *Poisson*, $P(\lambda)$, dependen de un número finito de parámetros y salvo que éstos se conozcan de antemano, deben ser estimados para conocer aproximadamente la distribución de probabilidad.

Es por esto que ya no se trabaja con una densidad sino con una familia paramétrica de funciones de densidad que contendrán un vector de incógnitas que tomarán valores dentro de un conjunto de valores admisibles llamado *espacio paramétrico*.

Un miembro específico de la familia de densidades estará asociado con un valor específico que asuman los parámetros.

Se usará la notación $f(\tilde{x}; \theta)$, donde x representa elementos del dominio y θ los elementos del espacio paramétrico (Θ)

7.1.1. Modelo paramétrico

Dada una muestra $\tilde{X} = (X_1, \dots, X_n)$, un modelo paramétrico para la misma está constituida por una forma funcional $f(\tilde{x}; \theta)$ para la densidad muestral de \tilde{X} , junto con el espacio paramétrico Θ que define el conjunto de candidatos

potenciales para la verdadera densidad muestral de \tilde{X} como $\{f(\tilde{x}; \theta) : \theta \in \Theta\}$.

El investigador puede elegir la parametrización del modelo que le resulte más útil (una reparametrización es una transformación invertible de θ), siempre y cuando los parámetros sean *identificables*:

Sea $\{f(\tilde{x}; \theta) : \theta \in \Theta\}$ un modelo para la muestra \tilde{X} , el vector de parámetros θ es identificable sii $\forall \theta_1, \theta_2 \in \Theta$, $f(\tilde{x}; \theta_1) \neq f(\tilde{x}; \theta_2)$ si $\theta_1 \neq \theta_2$, es decir, asignan probabilidades diferentes al menos a un evento de \tilde{X}

7.2. Estimación puntual

Supongamos que se ha observado un vector muestra $\tilde{X} = (X_1, \dots, X_n)$ de cuya distribución solo se conoce que pertenece a una familia paramétrica $\mathcal{F} = \{f(\tilde{x}; \theta) : \theta = (\theta_1, \dots, \theta_p) \in \Theta\}$. Supongamos que interese conocer aproximadamente $q(\theta)$, donde $q(\theta)$ es una función de Θ en \mathbb{R} . La única información que se tiene sobre θ es el vector \tilde{X} , por lo que cualquier estimación que se haga sobre θ deberá estar basada en \tilde{X} .

Un estimador puntual de $q(\theta)$ será cualquier estadístico $\delta(\tilde{X})$.

7.2.1. Estadístico

Un estadístico $\delta(\tilde{X})$ es una función $T(\tilde{X})$ medible, que toma valores en un espacio euclideo de dimensión finita.

Si $\delta(\tilde{X})$ es usado para estimar $q(\theta)$ se llama estimador puntual de $q(\theta)$.

Dado que existen infinitas funciones que pueden actuar como estimadores de una característica de interés, es necesario especificar propiedades deseables que deberían satisfacer estas funciones, a los fines de escoger el estimador más conveniente.

7.3. Error cuadrático medio

El *ECM* de un estimador δ de $q(\theta)$ se define como:

$$ECM_{\theta}(\delta) = E_{\theta}(\delta(\tilde{X}) - q(\theta))^2$$

El *ECM* explica el grado de dispersión de la distribución muestral de δ , así como el grado en el cual la tendencia central de δ se desvía de $q(\theta)$

7.3.1. Sesgo

El sesgo de un estimador $\delta(\tilde{X})$ de $q(\theta)$ se define como:

$$b_\theta(\delta) = E_\theta(\delta(\tilde{X}) - q(\theta)) = E_\theta(\delta(\tilde{X})) - q(\theta) \quad \forall \theta \in \Theta$$

Si:

- $b_\theta(\delta) > 0 \longrightarrow \delta(\tilde{X})$ sobreestima a $q(\theta)$
- $b_\theta(\delta) < 0 \longrightarrow \delta(\tilde{X})$ subestima a $q(\theta)$
- $b_\theta(\delta) = 0 \longrightarrow \delta(\tilde{X})$ es insesgado de $q(\theta)$

Trabajando sobre el segundo miembro del ECM , se llega a que: **(LXVII)**

$$ECM_\theta(\delta) = V\delta(\tilde{X}) + (b_\theta(\delta))^2$$

Si el estimador es insesgado, entonces:

$$ECM_\theta(\delta) = V\delta(\tilde{X})$$

El ECM sirve para dar una noción de eficiencia:

7.3.2. Eficiencia relativa

Sean δ_1 y δ_2 estimadores de $q(\theta)$, la eficiencia relativa de δ_1 respecto a δ_2 es:

$$ER_\theta(\delta_1, \delta_2) = \frac{ECM_\theta(\delta_2)}{ECM_\theta(\delta_1)} \quad \forall \theta \in \Theta$$

Si $ER_\theta(\delta_1, \delta_2) \geq 1 \quad \forall \theta \in \Theta$ y > 1 para algún $\theta \in \Theta$, entonces δ_1 es relativamente mas eficiente que δ_2 , y en ese caso se dice que δ_2 resulta inadmisibles para estimar $q(\theta)$

7.4. Estimadores EIO o IMVU

Si restringimos la búsqueda de estimadores a la clase de los insesgados, con frecuencia se podrá encontrar el mejor de dichos estimadores EIO o IMVU (Inssegado de mínima varianza uniforme).

Se dirá que $\delta(\tilde{X})$ es un estimador insesgado de mínima varianza para $q(\theta)$, uniformemente en $\theta \in \Theta$ (IMVU) si:

- i) $\delta(\tilde{X})$ es insesgado para $q(\theta)$
- ii) Dado cualquier otro estimador insesgado para $q(\theta)$, $\delta^*(\tilde{X})$, se cumple que $V_\theta \delta(\tilde{X}) \leq V_\theta \delta^*(\tilde{X}) \quad \forall \theta \in \Theta$ (uniformemente).

7.5. Estadísticos suficientes

Consideremos un vector aleatorio \tilde{X} cuya distribución pertenece a una familia $\mathcal{F} = \{f(\tilde{x}; \theta) : \theta \in \Theta \subset \mathbb{R}^k\}$. El vector \tilde{X} interesa en cuánto nos provee información sobre el verdadero valor de θ . Puede ocurrir que una parte de la información contenida en \tilde{X} carezca de interés para el conocimiento de θ , y por consiguiente convenga eliminarla, simplificando así la información disponible. Al realizar esta simplificación, eliminando de \tilde{X} toda la información irrelevante, se obtendrá otro vector T , seguramente de dimensión menor a n . Llamamos estadístico a cualquier función medible $T = t(\tilde{X})$

Si t no es biunívoca, del conocimiento de T no se podrá reconstruir el valor de \tilde{X} , por lo que T solo conservará una parte de la información que hay en \tilde{X} . El estadístico T será *suficiente* cuando conserve toda la información relevante para el conocimiento de θ .

Definición: Sea \tilde{X} un vector aleatorio de dimensión n con distribución $f(\tilde{x}; \theta) : \theta \in \Theta$, se dice que un estadístico $T = t(\tilde{X})$ es suficiente para θ si la distribución de X condicional a $T = t$ es independiente de θ para todo t .

Esto puede interpretarse como afirmando que una vez conocido el valor t de T , podemos olvidarnos del valor de \tilde{X} , ya que en T está toda la información que \tilde{X} tiene sobre θ .

El problema que se presenta es que para aplicar este método debemos primero presentar un candidato a estadístico suficiente, no nos permite identificarlos. Para eso, existe el siguiente teorema:

7.5.1. Criterio de Factorización de Neyman

Teorema 17 (de factorización) Sea \tilde{X} un vector aleatorio de dimensión n cuya distribución es $f(\tilde{x}; \theta) : \theta \in \Theta$, el estadístico $T = t(\tilde{X})$ es suficiente para

θ sii existen funciones g y h tales que:

$$f(\tilde{x}; \theta) = g(T(\tilde{x}), \theta)h(\tilde{x})$$

Teorema 18 Si $T = T(\tilde{X})$ es un estadístico suficiente para θ y m es una función biunívoca de T , entonces el estadístico $T^* = m(T)$ también es suficiente para θ

Como en cualquier problema, hay, de hecho, muchos estadísticos suficientes, un estadístico que permita lograr la mayor reducción de los datos mientras retiene toda la información sobre θ sería considerado preferible.

7.5.2. Suficiencia minimal

Sea \tilde{X} un vector aleatorio de dimensión n , cuya distribución es $f(\tilde{x}; \theta) : \theta \in \Theta$, un estadístico suficiente $T(\tilde{X})$ es *suficiente minimal* para θ sii dado cualquier otro estadístico suficiente $U(\tilde{X})$ para θ , $T(\tilde{X})$ es función de $U(\tilde{X})$, es decir que existe una función H tal que $T = H(U)$

Teorema 19 (De Lehmann-Scheffe sobre s.m.) Sea $f(\tilde{x}; \theta)$ la densidad de la muestra \tilde{X} , suponga que existe una función $T(\tilde{X})$ tal que, para 2 puntos muestrales, x e y (dos muestras), el cociente $\frac{f(x; \theta)}{f(y; \theta)}$ es constante como función de θ . (Es decir $\frac{\partial f(x; \theta)/f(y; \theta)}{\partial \theta} = 0$) sii $T(x) = T(y)$. Entonces, $T(\tilde{X})$ es un estimador suficiente minimal para θ .

7.6. Teorema de Rao - Blackwell

(LXXI) Supongamos que \tilde{X} es un vector aleatorio que pertenece a la familia $\mathcal{F} = \{f(\tilde{x}; \theta) : \theta \in \Theta\}$ y que $T(\tilde{X})$ es un estadístico suficiente para θ .

De acuerdo al concepto intuitivo que tenemos de estadístico suficiente, para estimar una función $q(\theta)$ deberán bastar estimadores que dependan solo de T , ya que en T está toda la información que \tilde{X} contiene sobre el parámetro θ . Esto es justamente lo que afirma el teorema:

Teorema 20 (de Rao - Blackwell) Sea \tilde{X} un vector aleatorio cuya distribución pertenece a una familia $f(\tilde{x}; \theta) : \theta \in \Theta$. Sea T un estadístico suficiente

para θ y $\delta(\tilde{X})$ un estimador para $q(\theta)$. Definamos un nuevo estimador:

$$\delta^*(T) \equiv E(\delta(\tilde{X})|T)$$

Entonces se tiene que:

$$i) ECM_{\theta}(\delta^*) \leq ECM_{\theta}(\delta) \quad \forall \theta \in \Theta$$

$$ii) \text{ La igualdad en (i) se satisface si y sólo si } P(\delta^*(T) = \delta(\tilde{X})) = 1 \quad \forall \theta \in \Theta$$

iii) Si $\delta(\tilde{X})$ es insesgado, entonces $\delta^*(T)$ también lo es.

Ejemplo

Dada $X \sim Exp(\theta)$, queremos estimar $q(\theta) = E_{\theta}X = \theta$

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta} 1_{(0, \infty)}(x) \quad \theta \in \Theta = \mathbb{R}_{++}$$

Supongamos una m.a. de tamaño 3 : X_1, X_2, X_3 v.a.i.i.d.

1) Demostrar que $\delta(X_1, X_2, X_3) = X_1$ es insesgado de $q(\theta)$

2) Aplicar Rao - Blackwell

1) se debe probar que

$$E_{\theta}\delta(X_1, X_2, X_3) = E_{\theta}X_1 = q(\theta) = \theta$$

$$\begin{aligned} E_{\theta}X_1 &= \int x_1 f(x_1; \theta) dx_1 \\ &= \int_0^{\infty} x_1 \frac{1}{\theta} e^{-x_1/\theta} dx_1 \end{aligned}$$

Aplicamos integración por partes:

$$\begin{aligned} \begin{cases} u = x_1 & dv = \frac{1}{\theta} e^{-x_1/\theta} dx_1 \\ du = dx_1 & v = -e^{-x_1/\theta} \end{cases} \\ &= -x_1 e^{-x_1/\theta} \Big|_0^{\infty} - \int_0^{\infty} -e^{-x_1/\theta} dx_1 \\ &= \frac{-x_1}{e^{x_1/\theta}} \Big|_0^{\infty} - \theta e^{-x_1/\theta} \Big|_0^{\infty} \\ &= \left(\lim_{x \rightarrow \infty} \frac{-x_1}{e^{x_1/\theta}} \right) - \left(\lim_{x \rightarrow \infty} \frac{\theta}{e^{x_1/\theta}} - \theta \right) \end{aligned}$$

Aplicamos la regla de L'Hopital

$$= \lim_{x \rightarrow \infty} \frac{-1}{1/\theta e^{x_1/\theta}} + \theta = \theta$$

Por lo tanto, $\delta(\tilde{X}) = X_1$ es insesgado

2) Para aplicar Rao - Blackwell

a) Encontrar estadístico suficiente en este modelo: CFN

b) Encontrar $f(\delta|t; \theta)$

c) Calcular $\delta^*(T) = E(\delta|T)$

d) Comprobar propiedades de $\delta^*(T)$

a) Criterio de factorización:

La densidad conjunta de la muestra es:

$$\begin{aligned} f(x_1, x_2, x_3; \theta) &= \prod_{i=1}^n f(x_i; \theta) \Rightarrow \text{por ser las } X_i \text{ independientes} \\ &= \prod_{i=1}^3 \frac{1}{\theta} e^{-x_i/\theta} 1_{\mathbb{R}_{++}}(x_i) \\ &= \underbrace{\frac{1}{\theta^3} e^{-\sum_{i=1}^3 x_i/\theta}}_{g(\theta, t)} \underbrace{1_{\mathbb{R}_{++}}(x_1, x_2, x_3)}_{h(x)} \end{aligned}$$

Donde $t = \sum_{i=1}^3 x_i = x_1 + x_2 + x_3$

Por lo tanto, $T(\tilde{X}) = \sum_{i=1}^3 X_i$ es suficiente para θ

b) Encontrar $f(\delta|t; \theta) = \frac{f(\delta, t; \theta)}{f_T(t; \theta)}$

b.i) $f(\delta, t; \theta) = f(x_1, x_1 + x_2 + x_3; \theta)$. Para encontrar esa densidad aplicamos el método del jacobiano. $(X_1, X_2, X_3) \rightarrow (Y_1, Y_2)$ donde $Y_1 = X_1 + X_2 + X_3$, $Y_2 = X_1$. Por lo tanto debemos crear una variable auxiliar, $Y_3 \equiv X_2$

$$f(\tilde{x}; \theta) = \frac{1}{\theta} e^{-(x_1+x_2+x_3)/\theta} 1_{\mathbb{R}_{+++}}(\tilde{x})$$

$$\left. \begin{aligned} t \equiv y_1 &= x_1 + x_2 + x_3 \implies y_1 = y_2 + y_3 + x_3 \\ \delta \equiv y_2 &= x_1 \\ y_3 &= x_2 \end{aligned} \right\} \begin{aligned} x_3 &= y_1 - y_2 - y_3 \\ x_1 &= y_2 \\ x_2 &= y_3 \end{aligned}$$

$$\text{Matriz Jacobiana} \equiv \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix}$$

Entonces, $|MJ| = j = 1 \Rightarrow |j| = 1$

Aplicando la fórmula del jacobiano:

$$f(y_1, y_2, y_3; \theta) = \frac{1}{\theta^3} e^{-(y_1)/\theta} 1_{Sop}(y_1, y_2, y_3)$$

donde $Sop = \{(y_1, y_2, y_3) : y_2 + y_3 < y_1 < \infty; y_2, y_3 > 0\}$

Para obtener ahora $f(y_1, y_2; \theta)$:

$$f(y_1, y_2; \theta) = \int_{Sop_{y_3}} f(y_1, y_2, y_3; \theta) dy_3$$

Soporte de y_3

$$y_3 > 0 \quad ; \quad y_2 + y_3 < y_1 \Rightarrow 0 < y_3 < y_1 - y_2$$

$$f(y_1, y_2; \theta) = \int_0^{y_1 - y_2} \frac{1}{\theta^3} e^{-y_1/\theta} dy_3$$

$$f(y_1, y_2; \theta) = \frac{1}{\theta^3} e^{-y_1/\theta} (y_1 - y_2) 1_{\{0 < y_2 < y_1\}}(y_1, y_2)$$

b.ii) $f_T(t; \theta) = f(y_1; \theta)$ Hay dos formas de encontrar esa densidad:

· Una forma es: $f(y_1; \theta) = \int_0^{y_1} f(y_1, y_2; \theta) dy_2$

· Otra forma es ver que $T \equiv Y_1 = X_1 + X_2 + X_3$ es summa de *Gammas* independientes que comparten el $\beta = \theta$. Según teorema: La summa de *Gammas* independientes que comparten el β y tienen distinto α , es otra *Gamma* con el mismo β y $\alpha = \sum_{i=1}^n \alpha_i$

En este caso, $X_i \sim \text{Gamma}(1, \theta)$, entonces $Y_1 = X_1 + X_2 + X_3 \sim \text{Gamma}(3, \theta)$ y su densidad es:

$$f(y_1; \theta) = \frac{y_1^2 e^{-y_1/\theta}}{2 \theta^3} 1_{\mathbb{R}_{++}}(y_1)$$

$$\text{b.iii) } f(\delta|t; \theta) = f(y_2|y_1; \theta) = \frac{f(y_1, y_2; \theta)}{f(y_1; \theta)}$$

$$f(y_2|y_1; \theta) = \frac{\frac{1}{\theta^3} e^{-y_1/\theta} (y_1 - y_2) 1_{\{0 < y_2 < y_1\}}(y_1, y_2)}{\frac{y_1^2 e^{-y_1/\theta}}{2 \theta^3} 1_{\mathbb{R}_{++}}(y_1)}$$

El cociente de indicadoras va a ser igual a aquella mas restrictiva. En este caso, la del numerador:

$$f(y_2|y_1; \theta) = \frac{2}{y_1^2} (y_1 - y_2) 1_{\{0 < y_2 < y_1\}}(y_1, y_2) \text{ con } y_1 \text{ fijo.}$$

c) $\delta^*(T) = E(\delta|T)$

$$\delta^*(T) = \delta^*(Y_1) = \int y_2 f(y_2|y_1) dy_2$$

$$\begin{aligned} \delta^*(T) &= \int_0^{y_1} y_2 \frac{2}{y_1^2} (y_1 - y_2) dy_2 \\ &= \frac{y_1}{3} \end{aligned}$$

Entonces, el estimador Rao-Blackwell es: $\delta^*(T) = \frac{Y_1}{3} = \frac{\sum_{i=1}^n X_i}{3} = \bar{X}$

d) Propiedades:

· Si $\delta(X)$ era insesgado, $\delta^*(T)$ también lo es:

$$E_\theta \delta^*(T) = E_\theta \frac{\sum_{i=1}^3 X_i}{3} = \frac{1}{3} \sum E_\theta X_i = \frac{1}{3} 3\theta = \theta \quad \checkmark$$

· $\delta^*(T)$ es relativamente mas eficiente que $\delta(X)$

$$\begin{aligned} ECM_\theta \delta(X) &= V_\theta \delta = V_\theta X_1 = \theta^2 \\ ECM_\theta \delta^*(T) &= V_\theta \delta^*(T) = V \frac{\sum x_i}{3} = \frac{1}{9} \sum V X_i = \frac{1}{3} \theta^2 \end{aligned}$$

Entonces se cumple que $ECM_\delta \delta^*(T) < ECM_\delta \delta(X) \quad \checkmark$

7.7. Completitud

Sea $\{f(t; \theta) : \theta \in \Theta\}$ una familia de funciones de densidad para el estadístico $T(\tilde{X})$. La familia se denomina completa si $E_\theta g(T) = 0$ para todo $\theta \in \Theta$ implica que $P_\theta(g(T) = 0) = 1$ para todo $\theta \in \Theta$. En este caso también se dice que $T(\tilde{X})$ es un estadístico completo.

La completitud es una propiedad de una familia de densidades, no de una densidad en particular.

La única función g definida sobre el rango de T que satisface $E_{\theta}g(T) = 0$ es la función $g(T) = 0$.

Además, si T es un estadístico suficiente y completo para θ y si existe un estadístico suficiente minimal, entonces T es suficiente minimal; y si se conoce un estadístico suficiente y completo, existe un método para construir estimadores IMVU.

7.8. Teorema de Lehmann - Scheffe sobre IMVU

(LXX) Sea \tilde{X} un vector aleatorio de dimensión n , cuya distribución pertenece a una familia $\mathcal{F} = \{f(\tilde{x}; \theta): \theta \in \Theta\}$. Sea T un estadístico suficiente y completo para θ . Luego, dada una función $q(\theta)$ se tiene que:

- i) Existe a lo sumo un estimador insesgado de $q(\theta)$ basado en T
- ii) Si $\delta(T)$ es un estimador insesgado de $q(\theta)$, entonces $\delta(T)$ es IMVU.
- iii) Si $\delta(\tilde{X})$ es insesgado para $q(\theta)$, luego $\delta^*(T) \equiv E(\delta(\tilde{X})|T)$ es un estimador IMVU para $q(\theta)$

7.9. Propiedades asintóticas

La diferencia conceptual entre las propiedades asintóticas y las basadas en muestras finitas es que las primeras se basan en las distribuciones asintóticas los estimadores.

El problema de la no unicidad de las distribuciones asintóticas plantea un problema para las propiedades asintóticas.

Una forma de evitar la no unicidad consiste en restringir el estudio de las propiedades asintóticas a clases o familias de estimadores para las cuales el problema no se presenta.

Restringiremos el estudio a la clase de estimadores Consistentes Asintóticamente Normales (CAN).

7.9.1. Consistencia

La sucesión de estimadores $\{\delta_n\}$ es consistente para $q(\theta)$ sii

$$\delta_n \xrightarrow{p} q(\theta) \quad \forall \theta \in \Theta$$

También se dice que δ_n es un estimador consistente para $q(\theta)$

7.9.2. Estimadores CAN

La clase de estimadores CAN de $q(\theta)$ se define como la colección de todos los estimadores δ_n de $q(\theta)$ para los cuales:

$$\sqrt{n}(\delta_n - q(\theta)) \xrightarrow{d} N([0], \Sigma_\delta)$$

Donde Σ_δ es la matriz de varianzas y covarianzas (definida positiva) de δ_n que puede depender de θ

En particular, si $q(\theta)$ es escalar, δ_n es CAN sii

$$\sqrt{n}(\delta_n - q(\theta)) \xrightarrow{d} N(0, \sigma^2)$$

La distribución límite de δ_n es efectivamente degenerada en $q(\theta)$. Eso se prueba por Slutsky y es la demostración de la consistencia de δ_n . Como la distribución límite de δ_n es degenerada, se usa la distribución asintótica:

$$\delta_n \overset{a}{\sim} N(q(\theta), \sigma^2/n)$$

Entonces, para analizar las propiedades se usan esas esperanza y varianza. δ_n es asintóticamente insesgado y su *ECM* asintótico es σ^2/n

Capítulo 8

Unidad 8. Métodos de estimación.

8.1. Método de Máxima Verosimilitud

8.1.1. Función de verosimilitud

Sea $f(\tilde{x}; \theta)$ la función de densidad de la muestra de $\tilde{X} = (X_1, \dots, X_n)$, donde $\theta = (\theta_1, \dots, \theta_k)$. Entonces, dado que se observa $X = x$, la función definida por:

$$L(\theta; x) = f(x; \theta)$$

se denomina función de verosimilitud.

A pesar de ser estructuralmente igual a la función de densidad conjunta de la muestra, no es lo mismo porque se interpreta diferente. Ahora la variable son los θ y los parámetros son las x que están fijas.

Cuando la muestra es aleatoria, $L(\theta; \tilde{x}) = \prod_{i=1}^n f(x_i; \theta)$

En el caso discreto, $f(\tilde{x}; \theta) = p(\tilde{x}; \theta)$ representa la probabilidad de observar el vector $\tilde{X} = (X_1, \dots, X_n)$ cuando el parámetro es θ . Es razonable pensar que si hemos observado el vector \tilde{X} , este tendrá alta probabilidad. Luego, se podría estimar θ como el valor que hace máxima $p(\tilde{x}; \theta)$, es decir, la probabilidad de haber observado esa muestra.

8.1.2. Estimador máximo verosímil. (EMV)

Para cada punto muestral \tilde{x} , sea $\hat{\theta} = \hat{\theta}(x)$ el valor de θ para el cual $L(\theta; \tilde{x})$ alcanza su máximo como función de θ , con \tilde{x} fijo, el estimador máximo verosímil de θ es $\hat{\theta}(x)$

8.1.3. Cómputo del EMV

Si Θ es un subconjunto abierto de \mathbb{R}^p , el soporte de $p(\tilde{x}; \theta)$ no depende de θ y $p(\tilde{x}; \theta)$ tiene derivadas parciales respecto a θ , el cómputo del EMV se realiza de la siguiente manera.

Como la función $\ln(u)$ es monótona creciente, maximizar $L(\theta; \tilde{x})$ es equivalente a maximizar $\ln L(\theta; \tilde{x}) \equiv \mathcal{L}(\theta)$

$$\mathcal{L}(\theta) = \ln L(\theta; \tilde{x}) = \ln \prod_{i=1}^n f(x_i; \theta) = \sum_{i=1}^n \underbrace{\ln f(x_i; \theta)}_{l(x_i; \theta)}$$

Luego:

$$\text{CPO} \quad \mathcal{L}'(\theta) = \frac{\partial \sum_{i=1}^n l(x_i; \theta)}{\partial \theta} = \sum_{i=1}^n l'(x_i; \theta) = 0 \implies \hat{\theta}$$

$$\text{CSO} \quad \mathcal{L}''(\theta) = \sum_{i=1}^n l''(x_i; \hat{\theta}) < 0$$

No se debe olvidar que $\hat{\theta}$ debe pertenecer a Θ .

Ejemplos: $X \sim Po(\theta)$, X_1, \dots, X_n v.a.i.i.d. $\theta \in \mathbb{R}_{++}$

$$f(x_i; \theta) = \frac{e^{-\theta} \theta^{x_i}}{x_i!} 1_{\{0,1,\dots\}}(x_i)$$

$$\begin{aligned} l(x_i; \theta) &= \ln f(x_i; \theta) = \ln \left[\frac{e^{-\theta} \theta^{x_i}}{x_i!} 1_{\{0,1,\dots\}}(x_i) \right] \\ &= \ln e^{-\theta} + \ln \theta^{x_i} - \ln x_i! + \underbrace{\ln 1_{\{0,1,\dots\}}(x_i)}_{=1(*)} \end{aligned}$$

(*) es = 1 porque ya observé la muestra, los valores de x tienen que \in soporte

$$= -\theta + x_i \ln \theta - \ln x_i! + 0$$

$$\begin{aligned}
l'(x_i; \theta) &= -1 + \frac{x_i}{\theta} \\
\sum_{i=1}^n l'(x_i; \theta) &= -n + \frac{\sum X_i}{\theta} = 0 \implies \boxed{\hat{\theta} = \frac{\sum X_i}{n} \equiv \bar{X}} \\
l''(x_i; \theta) &= \frac{-x_i}{\theta^2} \\
\sum l''(x_i; \hat{\theta}) &= \frac{-\sum x_i n^2}{[\sum x_i]^2} < 0 \checkmark
\end{aligned}$$

$$X \sim \text{Exp}(\theta), X_1, \dots, X_n \text{ v.a.i.i.d}, \quad \theta \in \mathbb{R}_{++}$$

$$f(x_i; \theta) = \frac{1}{\theta} e^{-x_i/\theta} 1_{\mathbb{R}_{++}}(x_i)$$

$$\begin{aligned}
l(x_i; \theta) &= \ln f(x_i; \theta) = \ln \left[\frac{1}{\theta} e^{-x_i/\theta} 1_{\mathbb{R}_{++}}(x_i) \right] \\
&= \ln \frac{1}{\theta} + \ln e^{-x_i/\theta} + \ln 1_{\mathbb{R}_{++}}(x_i) \\
&= -\ln \theta - \frac{x_i}{\theta}
\end{aligned}$$

$$\begin{aligned}
l'(x_i; \theta) &= \frac{-1}{\theta} + \frac{x_i}{\theta^2} \\
\sum_{i=1}^n l'(x_i; \theta) &= \frac{-n}{\theta} + \frac{\sum x_i}{\theta^2} = 0 \implies \hat{\theta} = \bar{X} \\
l''(x_i; \theta) &= \frac{1}{\theta^2} - \frac{2x_i}{\theta^3} \\
\sum l''(x_i; \theta) &= \frac{\overbrace{n^2 \sum x_i^3}^{>0} \overbrace{(1-n2)}^{<0}}{\underbrace{\sum x_i^5}_{>0}} < 0 \checkmark
\end{aligned}$$

Un caso interesante se da cuando el parámetro está en el soporte de las X , con lo cual no se puede diferenciar. En ese caso hay que analizar la función y ver dónde se presenta el máximo.

$$X \sim U[0, \theta], X_1, \dots, X_n \text{ v.a.i.i.d} \quad \theta \in \mathbb{R}_{++}$$

$$f(x_i; \theta) = \frac{1}{\theta} 1_{[0, \theta]}(x_i)$$

$$L(\theta; \tilde{x}) = \prod_{i=1}^n f(x_i; \theta) = \frac{1}{\theta^n} \prod_{i=1}^n 1_{[0, \theta]}(x_i)$$

Si $\tilde{X} \in [0, \theta]$, $L(\theta; \tilde{x}) = \frac{1}{\theta}$ porque la indicadora vale uno.

Esa función es decreciente en θ , por lo cual para maximizar esa función debería hallar el valor mas pequeño entre $(0, \infty)$.

Sin embargo, no es posible acercarse arbitrariamente a cero ya que si luego de fijar un valor pequeño para $\hat{\theta}$, la máxima de las observaciones muestrales supera ese valor, habrán $x_i \notin [0, \theta]$ y la indicadora se hará cero, con lo cual se minimizaría $L(\theta; \tilde{x})$.

Por lo tanto, el mínimo valor que puede tomar $\hat{\theta}$ es $X_{(n)}$, es decir, la máxima observación muestral.

Sin embargo, el EMV $X_{(n)}$ no cumple con la propiedad de insesgadez:

$$P_{\theta}(X_{(n)} < \theta) = 1 - P_{\theta}(X_{(n)} \geq \theta) = 1 - \underbrace{[P(X_{(n)} = \theta)]}_{=0(\text{continuidad})} + \underbrace{P(X_{(n)} > \theta)}_{=0(\text{soporte})}$$

Por lo tanto, con probabilidad 1, $X_{(n)}$ será menor a θ .

Si, en cambio, la distribución de la variable fuera $X \sim U(0, \theta)$, $X_{(n)}$ ya no puede ser el estimador de θ porque la indicadora se haría cero. Por el *Teorema de Weierstrass*, al ser el dominio un abierto, no se puede garantizar la existencia de un máximo. Por lo tanto, el EMV no existe.

Otro caso de interés se da si suponemos:

$$X \sim U[\theta, \theta + 1]$$

$$f(x_i; \theta) = 1 \cdot 1_{[\theta, \theta+1]}(x_i)$$

$$L(\theta; \tilde{x}) = 1 \prod_{i=1}^n 1_{[\theta, \theta+1]}(x_i) = \begin{cases} 1 & \text{si } \theta \leq x_i \leq \theta + 1 \\ 0 & \text{si } c.c. \end{cases} \quad i = 1, 2, \dots, n$$

Entonces:

$$\left. \begin{array}{l} x_i \geq \theta \iff x_{(1)} \geq \theta \\ x_i \leq \theta + 1 \iff X_{(n)} - 1 \leq \theta \\ x_i - 1 \leq \theta \end{array} \right\} X_{(n)} \leq \theta \leq x_{(1)}$$

La función de verosimilitud, entonces, también se puede escribir como:

$$L(\theta; \tilde{x}) = 1 \cdot 1_{[x_{(n)}-1, x_{(1)}]}(\theta)$$

Se puede observar que esa función siempre vale uno para cualquier valor de $\theta \in [x_{(n)} - 1, x_{(1)}]$, por lo tanto, cualquier valor de θ en ese intervalo servirá para maximizar $L(\theta; \tilde{x})$. Entonces, el EMV no será único.

8.1.4. Propiedades de los EMV

- 1) Insesgadez: $E(\hat{\theta}) = \theta$
- 2) Consistencia: $\hat{\theta} \xrightarrow{p} \theta$
- 3) Suficiencia: Criterio de Factorización de Neyman.
- 4) Invarianza
- 5) Eficiencia Asintótica.

Invarianza de los EMV

Sea $\hat{\theta}$ el EMV de θ y sea g una función biunívoca, entonces se cumple que $g(\hat{\theta})$ es el EMV de $g(\theta)$

8.1.5. Distribución asintótica del EMV

Comencemos analizando el caso en que θ es univariado y X_1, \dots, X_n son v.a.i.i.d.

Impongamos los siguientes supuestos sobre $f(x; \theta)$:

- 1) Θ es un subconjunto abierto de \mathbb{R}
- 2) El soporte de $f(x; \theta)$, que simbolizaremos con A , no depende de θ
- 3) $f(x; \theta)$ es 3 veces diferenciable con respecto a θ para todo $x \in A$.
- 4) $E_{\theta}l'(x_i; \theta) = 0$ y $V_{\theta}l'(x_i; \theta) = I(\theta) \quad \forall \theta \in \Theta$ y donde $I(\theta) > 0$
- 5) $E_{\theta}l''(x_i; \theta) = -J(\theta)$, donde $J(\theta) > 0 \quad \forall \theta \in \Theta$
- 6) $|l'''(x_i; \theta)| \leq M(x)$ para todo $x \in A$ y $\theta \in \Theta$, donde $E_{\theta}M(x_i) < \infty$

Sean X_1, X_2, \dots, X_n v.a.i.i.d con densidad $\{f(\tilde{x}; \theta) : \theta \in \Theta\}$ que satisfacen los supuestos 1 a 6 y además $\hat{\theta}_n \xrightarrow{p} \theta$, es decir $\hat{\theta}$ es consistente, entonces $\hat{\theta}$ es un estimador que pertenece a la clase CAN:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N\left(0, \frac{I(\theta)}{J^2(\theta)}\right)$$

y

$$\hat{\theta} \overset{a}{\approx} \left(\theta, \frac{I(\theta)}{n J^2(\theta)}\right)$$

Observaciones

· Frecuentemente ocurre que $I(\theta) = J(\theta)$, con lo cual:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I(\theta)^{-1})$$

- $I(\theta)$ se denomina información (esperada) sobre θ por unidad de muestra.
- El teorema afirma que, de verificarse los supuestos, el EMV pertenece a la clase CAN, con varianza asintótica $[n I(\theta)]^{-1}$
- Cuando las v.a. son i.i.d, $n I(\theta)$ es la información contenida en la muestra y es igual al opuesto de $E_{\theta}(\mathcal{L}_n'')$
- Como el verdadero valor de θ es desconocido, es necesario estimar $I(\theta)$

Ejemplo

Dada $X \sim Po(\theta)$ y X_1, \dots, X_n v.a.i.i.d.

$$f(x_i; \theta) = \frac{e^{-\theta} \theta^{x_i}}{x_i!} 1_{\{0,1,2,\dots\}}(x_i)$$

Sea $\frac{\sum_{i=1}^n x_i}{n}$ el EMV de θ , se debe demostrar que pertenece a la clase CAN.

Por lo tanto se deben probar los 5 primeros supuestos, asumir el 6 y probar consistencia.

Supuestos 1) En un modelo *Poisson*, el espacio paramétrico $\Theta = \mathbb{R}_{++} = (0, \infty)$. Como $(0, \infty)^{\circ} = (0, \infty)$, entonces Θ es un abierto. ✓

2) $A = \text{Sop}_X = \{0, 1, 2, \dots\} = \mathbb{N} \cup \{0\}$, es independiente de θ ✓

$$3) f(x_i; \theta) = \frac{e^{-\theta} \theta^{x_i}}{x_i!} 1_A(x_i) = \frac{1}{x_i!} e^{-\theta} \theta^{x_i} 1_A(x)$$

Como función de θ , $\frac{1}{x_i!}$ es constante, y luego se tiene el producto de dos funciones de θ , una exponencial y una potencial, ambas infinitamente diferenciables, por lo tanto su producto también lo es. La función es, entonces, infinitamente diferenciable con derivadas continuas. ✓

$$4) f(x_i; \theta) = \frac{e^{-\theta} \theta^{x_i}}{x_i!} 1_A(x_i)$$

$$\ln f(x_i; \theta) \equiv l(x_i; \theta) = -\theta + x_i \ln \theta - \ln x_i! + \ln 1_A(x_i)$$

$$l'(x_i; \theta) = -1 + \frac{x_i}{\theta}$$

Como el experimento no se llevó a cabo, esas realizaciones son en realidad v.a. y $l'(x_i; \theta)$ también, entonces podemos también tomar esperanza y varianza.

$$\begin{aligned} El'(X_i; \theta) &= E(-1 + \frac{X_i}{\theta}) \\ &= -1 + \frac{1}{\theta} EX_i \\ &= -1 + \frac{\theta}{\theta} = 0 \quad \checkmark \end{aligned}$$

$$\begin{aligned} Vl'(X_i; \theta) &= V(-1 + \frac{X_i}{\theta}) \\ &= V(\frac{X_i}{\theta}) = \frac{1}{\theta^2} \theta = \frac{1}{\theta} \equiv I(\theta) > 0 \quad \forall \theta \in \Theta \quad \checkmark \end{aligned}$$

$$5) \quad l''(X_i; \theta) = \frac{-1}{\theta^2} X_i$$

$$El''(X_i; \theta) = E(\frac{-1}{\theta^2} X_i) = \frac{-1}{\theta^2} EX_i = \frac{-1}{\theta^2} \theta = \frac{-1}{\theta} \equiv -J(\theta) \quad \checkmark$$

$J(\theta) = I(\theta) \rightarrow$ Igualdad de información.

Consistencia:

$$\hat{\theta}_n = \frac{\sum X_i}{n} \xrightarrow{p, LDGN} EX_1 = \theta$$

Entonces $\hat{\theta}_n$ es consistente para θ

Observación: La LDGN se usa para probar consistencia solo cuando el estimador se presenta como la suma normalizada de v.a.i.i.d.. Si esto no es así, se usan otros métodos para probar consistencia: (Tchebycheff, convergencia en distribución mediante f.c., ECM, etc)

Habiendose cumplido los supuestos 1-5, la consistencia y asumiendo el 6, entonces se da que:

$$\sqrt{n} \left(\frac{\sum X_i}{n} - \theta \right) \xrightarrow{d} N(0, \theta)$$

Este es el punto de partida del método delta, lo cual implica que la distribución límite de $\hat{\theta}$ es degenerada en θ . Por lo tanto, se usa su distribución asintótica, que es:

$$\hat{\theta} \overset{a}{\sim} \left(\theta, \frac{\theta}{n} \right)$$

Por propiedad de invarianza se tiene que: si $\hat{\theta}$ es EMV de θ , entonces $q(\hat{\theta})$ es el EMV de $q(\theta)$.

En particular, si definimos $q(\theta) = P(X = 0) = e^{-\theta}$, entonces el EMV será:

$$q(\hat{\theta}) = e^{-\hat{\theta}} = e^{-\sum X_i/n}$$

Si $\hat{\theta}$ es CAN, entonces $q(\hat{\theta})$ también será CAN, por lo que, para conocer su distribución asintótica usamos el método delta; que dice que dada una sucesión de v.a.i.i.d. tales que:

$$\sqrt{n}(X_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

y dada una $g(x)$ tal que $g'(x)$ exista, entonces:

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{d} N(0, \sigma^2(g'(\mu))^2)$$

En el ejemplo: $X_n = \frac{\sum X_i}{n}$, $g(X_n) = e^{-\sum X_i/n}$, $g(X_i) = e^{-\theta}$, $g'(x) = -e^{-\theta}$

Entonces:

$$\sqrt{n}(e^{-\sum X_i/n} - e^{-\theta}) \xrightarrow{d} N(0, \theta e^{-2\theta})$$

y

$$e^{-\sum X_i/n} \overset{a}{\sim} N(e^{-\theta}, \theta e^{-2\theta}/n) \quad (\text{Se prueba por jacobiano})$$

8.2. Eficiencia asintótica: Cota de Rao - Cramer

Hemos visto que si se toma como criterio para evaluar a los estimadores al ECM, cuando consideramos los estimadores insesgados, el problema se reduce a encontrar el de menor varianza, que denominamos IMVU.

Si contamos con un estadístico suficiente y completo, podemos encontrar el IMVU mediante la raoblackwellizacion. El problema surge cuando se cuenta con un estadístico suficiente pero no completo.

En estos casos, sería útil contar con una cota inferior para la varianza de todo estimador insesgado de $q(\theta)$, de modo que si algún estimador insesgado la alcanza, entonces será IMVU.

Sabemos que si T y U son v.a. con momento finito de orden 2, entonces por Cauchy - Schwartz:

$$Cov^2(T, U) \leq V(T)V(U)$$

Por lo tanto, si δ es un estimador insesgado de $q(\theta)$, para cualquier v.a. U se tiene que:

$$V_\theta(\delta) \geq \frac{Cov_\theta^2(\delta, U)}{V_\theta(U)}$$

Así expuesta, esta cota no es útil porque depende también de δ , no solo de θ ; pero debemos encontrar una cota para todos los estimadores insesgados de $q(\theta)$, no que cambie según el δ analizado.

Afortunadamente, se puede encontrar una v.a. U tal que $Cov_\theta^2(\delta, U)$ dependa solo de $q(\theta)$.

Comencemos suponiendo que θ es escalar y que la función de densidad conjunta satisface las siguientes exigencias, llamadas *condiciones de regularidad*:

A) El soporte de $f(\tilde{x}; \theta)$, simbolizado A , no depende de θ .

B) $f(\tilde{x}; \theta)$ es diferenciable con respecto a $\theta \quad \forall x \in A$

(**LXVI**) C) $U_\theta(\tilde{X}) = l'(\tilde{x}; \theta)$ y $E_\theta U_\theta(\tilde{X}) = 0$

(**LXVI**) D) Para cualquier $\delta(\tilde{X})$ con esperanza finita para todo θ e insesgado respecto de $q(\theta)$, diferenciable, se cumple que:

$$q'(\theta) = E_\theta[\delta(\tilde{X}) U_\theta(\tilde{X})]$$

Si se cumplen estas condiciones, entonces: (**LXVIII**)

$$V_\theta(\delta) \geq \frac{[q'(\theta)]^2}{I_n(\theta)} \implies \text{Cota de Rao - Cramer}$$

Donde $I_n(\theta) = E_\theta(U_\theta^2) = V_\theta(U_\theta) \quad \text{y} \quad U_\theta = l'(\tilde{x}; \theta)$

· (**LXIX**) Si X_1, \dots, X_n son v.a.i.i.d. con densidad común $f(\tilde{x}; \theta)$, entonces: $I_n(\theta) = nI(\theta)$

· Cuando un estimador alcanza la cota R-C, entonces es IMVU

· En el caso de que $q(\theta) = \theta$, la cota de R-C es $[I_n(\theta)]^{-1}$

· El EMV asintóticamente alcanza la cota de R-C, por lo que es asintóticamente eficiente.