

Probabilidad I

Santiago de Diego
Braulio Valdivielso
Francisco Luque

Agradecimientos

¡Un saludo a mi gente de Tinder, se os quiere!

Resumen

Índice general

Capítulo 1

Conjuntos y funciones σ -aditivas

1.1. Introducción: Espacio medible

Concepto de σ -álgebra

Sea un conjunto Ω , una familia de subconjuntos \mathcal{A} es σ -álgebra sobre Ω si:

1. $\Omega \in \mathcal{A}$
2. Si $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$
3. Si $A_1, A_2, A_3 \dots \in \mathcal{A} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$

Una σ -álgebra es una clase de conjuntos cerrada para el complementario y unión numerable. Existen una serie de propiedades inmediatas derivadas:

- $\emptyset \in \mathcal{A}$
- \mathcal{A} es cerrada para la intersección numerable.

Espacio medible

Al par (Ω, \mathcal{A}) se le llama espacio medible, y a los conjuntos que pertenecen a \mathcal{A} se les denomina conjuntos medibles.

1.2. Límites en sucesiones de conjuntos

1.2.1. Límites en sucesiones monótonas

Si tomamos la relación de inclusión entre conjuntos (\subseteq) como una relación de orden, podemos hablar sin ningún tipo de problema de sucesiones monótonas. En este sentido, una sucesión $\{A_n\}_{n \in \mathbb{N}}$ creciente de conjuntos cumpliría que $A_i \subseteq A_{i+1}, \forall i \in \mathbb{N}$. De forma análoga para sucesión decreciente de conjuntos.

El límite de una sucesión creciente de conjuntos $\{A_n\}_{n \in \mathbb{N}}$ se puede definir

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{k=1}^{\infty} A_k$$

Esta definición aprovecha la relación de orden entre los conjuntos de la sucesión para afirmar que *si un elemento está en el último conjunto de la sucesión, entonces está en todos los anteriores, y por tanto en **todos***, por ello se puede utilizar una intersección numerable (que está perfectamente definida) para formalizar el concepto.

Con una intuición análoga se define el límite de una sucesión decreciente de conjuntos. Si $\{A_n\}_{n \in \mathbb{N}}$ es una sucesión decreciente de conjuntos, entonces:

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{k=1}^{\infty} A_k$$

1.2.2. Límites superiores e inferiores

No solo se puede hablar de límites en sucesiones monótonas. Para definir los límites en sucesiones arbitrarias de conjuntos tenemos que recurrir a los conceptos de límite inferior y superior. La intuición de estos límites superior e inferior pasan por el concepto de las colas de la sucesión.

Dada una sucesión de conjuntos $\{A_n\}_{n \in \mathbb{N}}$ podemos considerar el conjunto

$$B_n = \bigcap_{k=n}^{\infty} A_k$$

y este conjunto contiene aquellos elementos que están en **todos** los A_k para $k \geq n$. Es fácil probar que la sucesión $\{B_n\}_{n \in \mathbb{N}}$ es una sucesión creciente de conjuntos, y por tanto se puede obtener el límite $\lim_{n \rightarrow \infty} B_n$. Informalmente, ese límite es un conjunto que contiene a todos los elementos de A_n que están

en todos los conjuntos A_k a partir de cierto $n \in \mathbb{N}$. Definimos el límite inferior de A_n como

$$\liminf A_n = \lim_{n \rightarrow \infty} B_n = \bigcup_{n \in \mathbb{N}} \bigcap_{k=n}^{\infty} A_n$$

Análogamente, podríamos definir la sucesión de conjuntos $\{C_n\}_{n \in \mathbb{N}}$ como

$$C_n = \bigcup_{k=n}^{\infty} A_n$$

Cada C_n contiene todos los elementos que están presentes en algún A_k para $k \geq n$. Es fácil también ver que la sucesión $\{C_n\}_{n \in \mathbb{N}}$ es una sucesión decreciente de conjuntos, y por tanto su límite también está bien definido. Se puede definir entonces el límite superior de $\{A_n\}_{n \in \mathbb{N}}$ como

$$\limsup A_n = \lim_{n \rightarrow \infty} C_n = \bigcap_{n \in \mathbb{N}} \bigcup_{k=n}^{\infty} A_n$$

Informalmente se puede pensar en este límite superior de A_n como el conjunto de los elementos que están en infinitos conjuntos de la sucesión.

A partir de estas definiciones, es fácil comprobar que

$$\liminf A_n \subseteq \limsup A_n$$

1.2.3. Límite de una sucesión de conjuntos

Diremos que una sucesión de conjuntos tiene límite si su límite inferior y superior coinciden, y el límite tendrá dicho valor. Es decir,

$$\lim A_n = \liminf A_n = \limsup A_n$$

en caso de que límite superior e inferior coincidan.

1.3. Funciones sobre conjuntos

Una vez definidos los conceptos sobre conjuntos con los que vamos a trabajar, pasamos a definir las funciones sobre conjuntos. Vamos entonces a definir lo que es una función de conjunto. Sean los espacios medibles (Ω, \mathcal{A}) y (Ω', \mathcal{A}') , definimos la función:

$$X : \mathcal{A} \rightarrow \mathcal{A}'$$

$$A \longrightarrow X(A)$$

A raíz de esta definición podemos definir también lo que se conoce como función inversa. Dada una función X , la función inversa de X , X^{-1} , asigna a cada conjunto $A' \in \mathcal{A}'$ el conjunto $A \in \mathcal{A}$ tal que $X(A) = A'$. La propiedad básica que cumplen las funciones inversas es que preservan las operaciones e inclusiones de conjuntos.

1.4. Concepto de σ -aditividad

Sea un conjunto Ω y una σ -álgebra \mathcal{A} sobre Ω . Definimos la función de conjunto:

$$\varphi : \mathcal{A} \rightarrow \mathbb{R}$$

$\varphi(A)$ es único

Se dice que φ es aditiva si $\varphi\left(\sum_1^n A_k\right) = \sum_1^n \varphi(A_k)$

Se dice que φ es σ -aditiva si $\varphi\left(\sum_1^\infty A_k\right) = \sum_1^\infty \varphi(A_k)$

Función subaditiva

Sea φ definida como antes. Se dice que φ es **subaditiva** si $\varphi(A \cup B) \leq \varphi(A) + \varphi(B)$

Lema 1.4.1 Si $\exists B : \varphi(B) < \infty \Rightarrow \varphi(\emptyset) = 0$

Demostración 1.4.1 $\varphi(B) = \varphi(B + \emptyset) = \varphi(B) + \varphi(\emptyset) \Rightarrow \varphi(\emptyset) = 0$.
Notese que $\varphi(B) < \infty$ es obligado porque si no $\varphi(\emptyset)$ podría tomar un valor $a \neq 0$ y se mantendría que $\varphi(B) + \varphi(\emptyset) = \infty + a = \infty$.

Teorema 1.4.1 Si φ es finita, σ -aditiva y $\sum \varphi(A_i) < \infty \Rightarrow \sum (|\varphi(A_i)|) < \infty$

Demostración 1.4.2 Tomemos las sucesiones:

Si $\varphi(A_n) \geq 0 \Rightarrow A_n^+ = A_n$ y $A_n^- = \emptyset$

Si $\varphi(A_n) < 0 \Rightarrow A_n^+ = \emptyset$ y $A_n^- = A_n$

Entonces $\varphi(\sum A_n^+) = \sum \varphi(A_n^+)$ y $\varphi(\sum A_n^-) = \sum \varphi(A_n^-)$, ambas finitas.

Sumando ambas cantidades obtenemos que $\sum (|\varphi(A_i)|) < \infty$

■

Teorema 1.4.2

Si φ es σ -aditiva, $\varphi \geq 0$, entonces: φ es no decreciente y subaditiva

Demostración 1.4.3

Veamos que es no decreciente ($A \subseteq B \Rightarrow \varphi(A) \leq \varphi(B)$). Podemos escribir $B = (B \cap A) + (B \cap \bar{A}) = A + (B \cap \bar{A})$. Entonces $\varphi(A) \leq \varphi(A) + \varphi(B \cap \bar{A}) = \varphi(B)$.

Veamos ahora que es subaditiva ($\varphi(\bigcup_j A_j) \leq \sum_j \varphi(A_j)$). Veamos que $\varphi(\bigcup_i A_i) = \varphi(A_1 + A_1^c A_2 + \dots + A_1^c A_2^c \dots A_n + \dots) \stackrel{1}{=} \varphi(A_1) + \varphi(A_1^c A_2) + \dots + \varphi(A_1^c A_2^c \dots A_n) + \dots \leq \sum_i \varphi(A_i)$ donde en 1 usamos que son disjuntos para aplicar la subaditividad. ■

1.5. Continuidad en funciones sobre conjuntos

Una vez introducido el concepto de límite para una sucesión de conjuntos, vamos a tratar de definir la continuidad para funciones de conjunto. Tendremos tres tipos de continuidad, cada uno relacionado con un tipo de sucesiones de conjuntos de las que hemos definido anteriormente. En esta sección trabajaremos con una función $\varphi : \Omega \rightarrow \Omega'$

Diremos que φ es continua por abajo si cumple que, dada una sucesión creciente de elementos $A_n \uparrow A$, se tiene que

$$\lim \varphi(A_n) = \varphi(A)$$

Por otra parte, diremos que φ es continua por arriba si $\exists A_{n_0} : \varphi(A_{n_0}) < \infty$ y cumple que dada una sucesión decreciente de elementos $A_n \downarrow A$, se tiene que

$$\lim \varphi(A_n) = \varphi(A)$$

Por último, diremos que una función es continua si lo es por arriba y por abajo.

Teorema 1.5.1 *Teorema de continuidad para funciones sobre conjuntos*

Sea φ una función σ -aditiva. Entonces, φ es aditiva y continua. Inversamente, si φ es aditiva y, o bien continua por abajo, o finita y continua en \emptyset , entonces φ es σ -aditiva.

Demostración 1.5.1 *Por un lado, sea φ una función σ -aditiva. Entonces es trivialmente aditiva. Ahora, veamos que es continua por abajo y por arriba. Sea $A_n \uparrow A$, entonces:*

$$A = \lim A_n = \bigcup A_n = A_1 + (A_2 - A_1) + (A_3 - A_2) + \dots$$

Unión de conjuntos disjuntos. Por tanto:

$$\varphi(A) = \varphi(\lim A_n) = \lim_{n \rightarrow \infty} \{\varphi(A_1) + \varphi(A_2 - A_1) + \dots + \varphi(A_n - A_{n-1})\} = \lim \varphi(A_n)$$

Veamos la continuidad por arriba. Sea $A_n \downarrow A$, tomamos A_{n_0} tal que $\varphi(A_{n_0})$ es finito. Entonces $A_{n_0} - A_n \uparrow A_{n_0} - A$, y por el apartado anterior tenemos la convergencia desde abajo, por tanto:

$$\varphi(A_{n_0}) - \varphi(A) = \varphi(\lim (A_{n_0} - A_n)) = \lim \varphi(A_{n_0} - A_n) = \varphi(A_{n_0}) - \lim \varphi(A_n)$$

De donde se deduce que $\varphi(A) = \lim \varphi(A_n)$.

Inversamente, sea φ una función aditiva. Si φ es continua por abajo, tenemos

$$\varphi\left(\sum_{k=1}^{\infty} A_k\right) = \varphi\left(\lim \sum_{k=1}^n A_k\right) = \lim \varphi\left(\sum_{k=1}^n A_k\right) = \lim \sum_{k=1}^n \varphi(A_k) = \sum_{k=1}^{\infty} \varphi(A_k)$$

Y por tanto es σ -aditiva. Si es finita y continua en \emptyset , entonces se obtiene la σ -aditividad de:

$$\varphi\left(\sum_{k=1}^{\infty} A_k\right) = \varphi\left(\sum_{k=1}^n A_k\right) + \varphi\left(\sum_{k=n+1}^{\infty} A_k\right) = \sum_{k=1}^n \varphi(A_k) + \varphi\left(\sum_{k=n+1}^{\infty} A_k\right)$$

Y tomando límites tenemos que

$$\varphi\left(\sum_{k=n+1}^{\infty} A_k\right) \rightarrow \varphi(\emptyset) = 0$$

■

Una vez demostrado este teorema, vamos a ver un teorema que nos relaciona las propiedades del supremo e ínfimo de una función σ -aditiva con los conjuntos sobre los que está dicha función definida:

Teorema 1.5.2 *Teorema del supremo e ínfimo*

Sea φ una función σ -aditiva sobre una σ -álgebra \mathcal{A} . Entonces, existen $C, D \in \mathcal{A}$ tales que $\varphi(C) = \sup \varphi$ y $\varphi(D) = \inf \varphi$

Demostración 1.5.2 *Probaremos la existencia del conjunto C . La del conjunto D es análoga. Si $\varphi(A) = \infty$ para algún $A \in \mathcal{A}$, entonces podemos establecer $A = C$ y la demostración del teorema es trivial. Entonces, supongamos que $\varphi < \infty$ y dado que el valor $-\infty$ está excluido, φ es finita.*

Entonces, existe una sucesión $\{A_n\} \subset \mathcal{A}$ tal que $\varphi(A_n) \rightarrow \sup \varphi$. Sea $A = \cup A_n$ y para cada n , consideramos la partición de A en 2^n conjuntos A_{nm} de la forma $\cap_{k=1}^n A'_k$, donde $A'_k = A_k$ o $A - A_k$. Para $n < n'$, cada conjunto A_{nm} es una suma finita de conjuntos $A_{n'm'}$. Sea ahora B_n la suma de los conjuntos A_{nm} para los cuales φ es no negativa. Si no hay ninguno, entonces $B_n = \emptyset$. Entonces, tenemos por un lado que A_n es la suma de algunos de los A_{nm} , y por otro lado, para $n' > n$, cada conjunto $A_{n'm'}$ está dentro de B_n o son disjuntos. Entonces tenemos

$$\varphi(A_n) \leq \varphi(B_n) \leq \varphi(B_n \cup B_{n+1} \cup \dots \cup B_{n'})$$

Tomando $n' \rightarrow +\infty$, se sigue de la continuidad por debajo que

$$\varphi(A_n) \leq \varphi(B_n) \leq \varphi\left(\bigcup_{k=n}^{+\infty} B_k\right)$$

Haciendo tender ahora $n \rightarrow +\infty$ y tomando $C = \lim \bigcup_{k=n}^{+\infty} B_k$ se sigue, por la continuidad por arriba, que $\sup \varphi \leq \varphi(C)$. Pero $\varphi(C) \leq \sup \varphi$, y por tanto se tiene que $\varphi(C) = \sup \varphi$, como queríamos. ■

Capítulo 2

Medidas y probabilidades

2.1. Concepto de medida, media exterior y probabilidad

Una función de conjuntos μ° es una medida si verifica:

- Es σ -aditiva, es decir, $\mu^\circ(\cup A_j) = \sum \mu^\circ(A_j)$
- Es no decreciente $A \subset B \Rightarrow \mu^\circ(A) \leq \mu^\circ(B)$
- $\mu^\circ(\emptyset) = 0$

A la tupla $(\Omega, \mathcal{A}, \mu_{\mathcal{A}})$ se le denomina espacio de medida.

Una **medida exterior** es una función de conjuntos positiva y σ -subaditiva, es decir, no se cumple la primera propiedad. La σ -subaditividad implica que $\mu(A \cup B) \leq \mu(A) + \mu(B)$.

Para que una medida exterior fuera una medida tendría que ser σ -aditiva. Es decir, la falla la primera condición. Sí que es positiva ya que $\mu^\circ(\emptyset) = 0$ y es creciente. Esta medida exterior se aplica a cualquier conjunto. Va a haber unos subconjuntos en los que la función se comporte como si fuera aditiva.

Una vez definido el concepto de medida, vamos a dar ahora el de probabilidad. Una probabilidad \mathcal{P} sobre un espacio medible (Ω, \mathcal{A}) es una medida que además cumple que $\mathcal{P}(\Omega) = 1$. Por tanto, tiene las siguientes propiedades:

$$\begin{aligned} \mathcal{P}(\emptyset) &= 0 \\ \forall A \in \mathcal{A}, 0 &\leq \mathcal{P}(A) \leq 1 \end{aligned}$$

Es una función σ -aditiva

De forma análoga al concepto de espacio de medida, podemos definir ahora el de espacio probabilístico. Un espacio probabilístico es una tupla formada por un conjunto Ω , una σ -álgebra sobre ese conjunto, \mathcal{A} , y una función de probabilidad \mathcal{P} .

Teorema 2.1.1 *Teorema de sucesión*

1. Si $A_n \uparrow A \Rightarrow \mathcal{P}(A_n) \uparrow \mathcal{P}(A)$
2. $\mathcal{P}(\liminf A_n) \leq \liminf \mathcal{P}(A_n)$
3. Si $A_n \downarrow A \Rightarrow \mathcal{P}(A_n) \downarrow \mathcal{P}(A)$
4. $\mathcal{P}(\limsup A_n) \geq \limsup \mathcal{P}(A_n)$
5. Si $A_n \rightarrow A \Rightarrow \mathcal{P}(A_n) \rightarrow \mathcal{P}(A)$

Demostración 2.1.1 1. *Trivial por continuidad de la probabilidad.*

2. $\liminf(A_n) = (\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k)$. Sea $B_n := \cap_{k=n}^{\infty} A_k$. Entonces sabemos que $B_n \uparrow$ y que, por lo tanto $\lim B_n = \cup_n B_n = \liminf(A_n)$.

Por último sabemos que $\mathcal{P}(A_n) \geq \mathcal{P}(B_n)$ dado que una probabilidad cumple ser no decreciente.

De todo esto podemos considerar que $\liminf \mathcal{P}(A_n) \geq \lim \mathcal{P}(B_n) = \mathcal{P}(\lim B_n) = \mathcal{P}(\liminf A_n)$

3. *trivial por continuidad de la probabilidad.*
4. *Análogo 2.*
5. *Se cumple por 2 y 4.*

2.2. Teorema de extensión de Carathéodory

Una vez vistas las definiciones anteriores de medida y medida exterior, vamos a ver un teorema fundamental, que nos servirá para justificar la forma en la que haremos el cálculo de probabilidades a posteriori. Veamos dos lemas antes que nos serán necesarios para la demostración del teorema principal.

Definición: Un conjunto $A \in S(\Omega)$ es μ° -medible si se cumple que $\mu^\circ(D) \geq \mu^\circ(AD) + \mu^\circ(A^cD), \forall D \in S(\Omega)$

Definición: Una extensión exterior μ° de una medida μ se define como:

$$\mu^\circ(A) = \inf \sum_j (A_j), \quad A \subset \cup_j A_j, \quad \text{recubrimiento}$$

Teorema 2.2.1 *La extensión exterior μ° de una medida μ sobre un álgebra es una extensión de μ a una medida exterior*

Demostración 2.2.1 *Primero, veamos que es una extensión. Esto es trivial, dado que $\mu^\circ(A) = \inf \sum \mu(A) = \mu(A)$. Ahora, veamos que es una medida exterior.*

Veamos que es subaditiva, es decir, $\mu^\circ(\cup_j A_j) \leq \sum_j \mu^\circ(A_j)$. Sea entonces, para cada A_j , un recubrimiento $\cup_k A_{kj} \supset A_j$. Entonces:

$$\sum_k \mu(A_{kj}) \leq \frac{\varepsilon}{2^j} + \mu^\circ(A_j)$$

$$\mu^\circ(\cup_j A_j) \leq \sum_{j,k} \mu(A_{kj}) \leq \sum_j \left(\mu^\circ(A_j) + \frac{\varepsilon}{2^j} \right) \leq \sum_j (\mu^\circ(A_j)) + \varepsilon$$

Claramente es no decreciente, ya que si $A \subseteq B$, todo recubrimiento de B lo será también de A , y por tanto, $\mu^\circ(A) \leq \mu^\circ(B)$

Por último, $\mu^\circ(\emptyset) = \mu(\emptyset) = 0$

■

Teorema 2.2.2 *Sea μ° una medida exterior sobre la σ -álgebra \mathcal{A} , y sea \mathcal{A}° la clase de conjuntos μ° -medibles:*

- \mathcal{A}° es un σ -campo
- μ° en \mathcal{A}° es una medida

Demostración 2.2.2 *Primero, veamos que \mathcal{A}° es un álgebra. Sea $A \in \mathcal{A}^\circ$, entonces $A^c \in \mathcal{A}^\circ$ dado que la definición de μ° -medibilidad es simétrica. Sean entonces $A, B \in \mathcal{A}^\circ$. Entonces:*

$$\mu^\circ(D) = \mu^\circ(AD) + \mu^\circ(A^cD) =$$

$$\begin{aligned}
&= \mu^\circ(ABD) + \mu^\circ(AB^cD) + \mu^\circ(A^cBD) + \mu^\circ(A^cB^cD) \geq \\
&\geq \mu^\circ(ABD) + \mu^\circ(AB^cD \cup A^cBD \cup A^cB^cD) = \\
&= \mu^\circ(ABD) + \mu^\circ(AB)^cD
\end{aligned}$$

Dado entonces que \mathcal{A}° es cerrada bajo la operación complementario e intersección finita, lo es para la unión finita, así que nos encontramos ante un álgebra. Veamos ahora que μ° es aditiva en \mathcal{A}° . Sean $A, B \in \mathcal{A}^\circ$ disjuntos. Entonces:

$$\mu^\circ(A + B) = \mu^\circ((A + B)A) + \mu^\circ((A + B)A^c) = \mu^\circ(A) + \mu^\circ(B)$$

Y teniendo que $\mu^\circ(A) \geq \mu^\circ(\emptyset) = 0$, μ° sobre \mathcal{A}° es una función aditiva y positiva. Para completar la prueba, veamos que μ° es σ -aditiva. Sean $A_n \in \mathcal{A}^\circ$, $A = \sum A_n$. Tomamos los conjuntos $B_n = \sum_{k=1}^n A_k \in \mathcal{A}^\circ$. Tenemos que:

$$\mu^\circ(D) \geq \mu^\circ(B_n D) + \mu^\circ(B_n^c D) \geq \sum_{k=1}^n \mu^\circ(A_k D) + \mu^\circ(A^c D)$$

Haciendo $n \rightarrow +\infty$

$$\mu^\circ(D) \geq \sum_{n=1}^{+\infty} \mu^\circ(A_n D) + \mu^\circ(A^c D) \geq \mu^\circ(AD) + \mu^\circ(A^c D)$$

Además queda demostrado que $A \in \mathcal{A}^\circ$, así que tenemos que \mathcal{A}° es un σ -campo. Por último, tomando $D = A$, tenemos que $\mu^\circ(A) \geq \sum \mu^\circ(A_n) \Rightarrow \mu^\circ(A) = \sum \mu^\circ(A_n) \Rightarrow \mu^\circ$ es una medida. ■

Teorema 2.2.3 de extensión de Carathéodory

Dada una medida μ sobre un álgebra \mathcal{C} , existe una extensión μ° sobre la σ -álgebra minimal sobre \mathcal{C} ($\mathcal{A}(\mathcal{C})$). Además, si μ es finita, la extensión es única.

Demostración 2.2.3

$\forall A \in \mathcal{C}, \forall D \in \mathcal{C}, \forall \varepsilon > 0, \exists \{A_j\} \subset \mathcal{C}$ recubrimiento de D tal que

$$\mu^\circ(D) + \varepsilon \geq \sum \mu(A_j) \geq \sum \mu(AA_j) + \sum \mu(A^c A_j) \geq \mu^\circ(AD) + \mu^\circ(A^c D) \Rightarrow$$

$$\xrightarrow{\varepsilon \rightarrow 0} \mu^\circ(D) \geq \mu^\circ(AD) + \mu^\circ(A^c D), A \in \mathcal{A}^\circ \Rightarrow \mathcal{C} \subseteq \mathcal{A}^\circ, \mathcal{A}(\mathcal{C}) \subseteq \mathcal{A}^\circ$$

Por los dos teoremas que hemos demostrado antes, la restricción μ° sobre $\mathcal{A}(\mathcal{C})$ es una extensión de μ a una medida sobre $\mathcal{A}(\mathcal{C})$. Ahora, veamos que es única. Para ello, sean μ_1, μ_2 extensiones de μ . Sea entonces el conjunto $\mathcal{M} = \{A : \mu_1(A) = \mu_2(A)\} \Rightarrow \mathcal{C} \subset \mathcal{M}$. Tenemos entonces que:

$$\Omega \in \mathcal{C} \Rightarrow \Omega \in \mathcal{M} \Rightarrow \mu_1(\Omega) = \mu_2(\Omega) = \mu(\Omega) < \infty \Rightarrow \mu_1, \mu_2 \text{ finitos}$$

$$A \in \mathcal{M} \Rightarrow \mu_1(A^c) = \mu(\Omega) - \mu_1(A) = \mu(\Omega) - \mu_2(A) = \mu_2(A^c) \Rightarrow A^c \in \mathcal{M}$$

$$A, B \in \mathcal{M} \Rightarrow \mu_1(A+B) = \mu_1(A) + \mu_1(B) = \mu_2(A) + \mu_2(B) = \mu_2(A+B) \Rightarrow A+B \in \mathcal{M}$$

Entonces, de momento \mathcal{M} es un campo.

Sean ahora $\{A_n\} \subset \mathcal{M}$ monótona, entonces $\mu_1(\lim A_n) = \lim \mu_1(A_n) = \lim \mu_2(A_n) = \mu_2(\lim A_n) \Rightarrow \mathcal{M}$ cerrada frente al límite de sucesiones monótonas $\Rightarrow \mathcal{M}$ σ -álgebra que contiene a \mathcal{C} , por tanto, contiene a la σ -álgebra minimal sobre \mathcal{C}

$$\mathcal{A}(\mathcal{C}) \subseteq \mathcal{M} \Rightarrow \mu_1 = \mu_2 \text{ en } \mathcal{A}(\mathcal{C})$$

■

Capítulo 3

Funciones medibles

Antes de nada, vamos a puntualizar unas características básicas de la inversa de una función que podrían pasar desapercividas:

Lema 3.0.1 Sea $X : \Omega \rightarrow \overline{\mathbb{R}}$ y $S', S, S_t \subset \mathbb{R} \forall t \in \mathbb{N}$

- $X^{-1}(S' - S) = X^{-1}(S') - X^{-1}(S)$
- $X^{-1}(\cup_t S_t) = \cup_t X^{-1}(S_t)$
- $X^{-1}(\cap_t S_t) = \cap_t X^{-1}(S_t)$

Lema 3.0.2 Sea $X : \Omega \rightarrow \overline{\mathbb{R}}$

- La imagen inversa de un σ -campo sobre \mathbb{R} es un σ -campo
- La imagen inversa de un $\mathcal{B}(\mathcal{C})$ (σ -campo de borel sobre \mathcal{C}) es el mínimo σ -campo sobre $X^{-1}(\mathcal{C})$.
- La clase de todos los conjuntos cuya antiimagen está en un σ -campo es un σ -campo.

Una vez definidos los conceptos sobre espacios de medida y espacios probabilísticos, vamos a aproximarnos al concepto de función medible. Para ello, daremos dos definiciones de función medible, para demostrar más adelante que estas dos definiciones son equivalentes. Empecemos con la definición constructiva de función medible. Dado que nos interesa que el codominio sea \mathbb{R} , trabajaremos con funciones definidas entre un espacio medible (Ω, \mathcal{A}) y $(\mathbb{R}, \mathcal{B})$, donde \mathcal{B} representa la σ -álgebra de Borel.

Primero, se define la función puntual $X : \Omega \rightarrow \mathbb{R}$, que asigna a cada $\omega \in \Omega$ un número $x = X(\omega)$ único. Llamaremos a esta función variable aleatoria. Utilizaremos de aquí en adelante la siguiente notación:

- $[X]$ = dominio de la variable aleatoria X .
- $[X \leq Y] = \{\omega : X(\omega) \leq Y(\omega)\}$
- $[X = x] = \{\omega : X(\omega) = x\}$

Sea entonces la función $X = \sum_j x_j I_{A_j}$, donde A_j son conjuntos medibles y I_{A_j} denota la función indicadora de dicho conjunto. Estas funciones se llaman funciones elementales, y cuando el número de valores distintos que toma la función X es finito, se conocen como funciones simples. Entonces, la definición constructiva de función medible es la que sigue:

Definición constructiva de función medible: Una función es medible si es límite de una sucesión de funciones simples $\{X_n\}$ convergentes.

Esta definición que hemos dado es constructiva, y por tanto, nos será muy útil para la definición constructiva de las integrales. No obstante, para enunciar y demostrar las propiedades de las funciones medibles, suele ser más útil la definición descriptiva siguiente:

Definición descriptiva de función medible: Sea una función $\varphi : \mathcal{A} \rightarrow \mathcal{B}$. Se dice que φ es medible si $\forall B \in \mathcal{B} \Rightarrow \varphi^{-1}(B) \in \mathcal{A}$. Es decir, una función se dice medible si la imagen inversa de todo conjunto medible es medible.

No obstante, se puede dar, a raíz de esta, otra definición más económica:

Para la definición anterior, es suficiente con exigir la medibilidad de las imágenes inversas de los elementos de una subclase α para la cual la σ -álgebra minimal sobre α sea \mathcal{B}

Veamos ahora que las dos definiciones que hemos dado son equivalentes.

Teorema 3.0.1 *Teorema de medibilidad*

Las definiciones constructiva y descriptiva de una función medible son equivalentes, y la clase de funciones medibles es cerrada bajo las operaciones usuales del análisis.

Demostración 3.0.1 *Sean X_n funciones medibles en el sentido descriptivo. Entonces todos los conjuntos de la forma*

$$[\inf X_n < x] = \cup [X_n < x], [-X_n < x] = [X_n > -x]$$

son medibles, y por tanto, las funciones

$$\sup X_n = -\inf(-X_n), \liminf(X_n) = \sup(\inf_{k \geq n} X_k)$$

$$\limsup X_n = -\liminf(-X_n)$$

son medibles en el sentido descriptivo. Por un lado, las funciones de este tipo son claramente cerradas bajo las operaciones de supremo, ínfimo, y límites. Además, toda función simple es medible en el sentido descriptivo, ya que todos los conjuntos $[X \leq x] = \sum_{x_j \leq x} A_j$ son medibles. Entonces, el límite de sucesiones convergentes de funciones simples es medible, y por tanto las funciones medibles en sentido constructivo lo son en sentido descriptivo.

Veamos la otra implicación, es decir, que las funciones medibles en sentido descriptivo lo son en sentido constructivo. Sea una función X medible en sentido descriptivo. Entonces, las funciones

$$X_n = -nI_{[X < -n]} + \sum_{k=-n2^{n-1}+1}^{n2^n} \frac{k-1}{2^n} I_{\left[\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\right]} + nI_{[X \geq n]}, n \in \mathbb{N}$$

son simples. Entonces, dado que

$$|X_n(\omega) - X(\omega)| < \frac{1}{2^n} \quad \text{para} \quad |X(\omega)| < n$$

y

$$X_n(\omega) = \pm n \quad \text{para} \quad X(\omega) = \pm\infty$$

se tiene entonces que $X_n \rightarrow X$ y esto, con lo anterior, prueba la equivalencia de las dos definiciones de función medible. ■

Teorema 3.0.2

Sea $X : \Omega \rightarrow \mathbb{R}$ medible y $g : \mathbb{R} \rightarrow \mathbb{R}$ continua. Entonces, $g(X)$ es medible

Demostración 3.0.2 Si X es elemental, tenemos que $g(X) = g\left(\sum_j x_j I_{A_j}\right) = \sum_j g(x_j) I_{A_j}$ función elemental, y por tanto, medible.

Si X es el límite de funciones elementales:

$$g(X) = g(\lim X_n) = \lim g(X_n) = \lim \sum_j g(x_j) I_{A_j}$$
■

Para este teorema introduciremos el concepto de función boreliana. Esta es una función $g : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ tal que $g^{-1}(B) \in \mathcal{B}, \forall B \in \mathcal{B}$

Teorema 3.0.3

Sean X función medible y g función boreliana. Entonces, $g(X)$ es medible.

Demostración 3.0.3

$$(g(X))^{-1}(B) = X^{-1}(g^{-1}(B)) \in \mathcal{A}$$

■

3.1. Convergencia en probabilidad y convergencia casi segura

Vamos a establecer ahora criterios que nos permitan comparar variables aleatorias. Dado un espacio de probabilidad $(\Omega, \mathcal{A}, \mathcal{P})$, se dice que dos variables aleatorias son equivalentes si:

$$X \mathcal{R} Y \Leftrightarrow \mathcal{P}[X = Y] = 1$$

Análogamente:

$$X(\omega) = Y(\omega) \quad \forall \omega \in \Omega \setminus \Lambda, \quad \mathcal{P}(\Lambda) = 0$$

3.1.1. Definición de convergencia en probabilidad

Si $P[|X_n - X| \geq \varepsilon] \rightarrow 0$, entonces se dice que $X_n \xrightarrow{\mathbb{P}} X$, es decir, X_n converge en probabilidad a X

Lema 3.1.1 $X_n \xrightarrow{\mathcal{P}} X \wedge X_n \xrightarrow{\mathcal{P}} Y \implies X \mathcal{R} Y$

Demostración 3.1.1

$$|X - Y| = |X - X_n - Y + X_n| \leq |X_n - X| + |X_n - Y|$$

$$P[|X - Y| \geq \varepsilon] \leq P[|X_n - X| \geq \frac{\varepsilon}{2}] + P[|X_n - Y| \geq \frac{\varepsilon}{2}] \rightarrow 0$$

Y por tanto, $X \mathcal{R} Y$

■

3.1.2. Definición de convergencia uniforme en u

Se dice que una variable aleatoria X converge uniformemente en u si se cumple que $\mathcal{P}[|X_{n+u} - X_n| \geq \varepsilon] \rightarrow 0$

Lema 3.1.2

La convergencia en probabilidad implica la convergencia uniforme

Demostración 3.1.2

$$\begin{aligned} |X_{n+u} - X_n| &\leq |X_{n+u} - X| + |X - X_n| \Rightarrow \\ \Rightarrow \mathcal{P}[|X_{n+u} - X_n| \geq \varepsilon] &\leq \mathcal{P}\left[|X_{n+u} - X| \geq \frac{\varepsilon}{2}\right] + \mathcal{P}\left[|X - X_n| \geq \frac{\varepsilon}{2}\right] \rightarrow 0 \end{aligned}$$

■

3.1.3. Propiedades

1. $X_n \xrightarrow{\mathcal{P}} X \wedge Y_n \xrightarrow{\mathcal{P}} Y \implies X_n + Y_n \xrightarrow{\mathcal{P}} X + Y$
2. $X_n \xrightarrow{\mathcal{P}} X \implies KX_n \xrightarrow{\mathcal{P}} KX$
3. $X_n \xrightarrow{\mathcal{P}} K$ (es decir, que degenera), entonces $X_n^2 \xrightarrow{\mathcal{P}} K^2$.
Para demostrarlo basta notar que: $X_n^2 - K^2 = (X_n + K)(X_n - K)$
4. $X_n \xrightarrow{\mathcal{P}} a \wedge Y_n \xrightarrow{\mathcal{P}} b \implies X_n Y_n \xrightarrow{\mathcal{P}} a \cdot b$

Para demostrarlo tenemos que notar que:

$$X_n Y_n = \frac{(X_n + Y_n)^2 - (X_n - Y_n)^2}{4} \xrightarrow{\mathcal{P}} \frac{(a + b)^2 - (a - b)^2}{4} = ab$$

5. $\forall n \in \mathbb{N}$ con $X_n \neq 0$, $X_n \xrightarrow{\mathcal{P}} 1 \implies \frac{1}{X_n} \xrightarrow{\mathcal{P}} 1$
6. $\forall n \in \mathbb{N}$ con $Y_n \neq 0$, $X_n \xrightarrow{\mathcal{P}} a \wedge Y_n \xrightarrow{\mathcal{P}} b \neq 0 \implies X_n Y_n^{-1} \xrightarrow{\mathcal{P}} ab^{-1}$
7. $X_n \xrightarrow{\mathcal{P}} X \wedge Y_n \xrightarrow{\mathcal{P}} Y \implies X_n Y_n \xrightarrow{\mathcal{P}} XY$

Para demostrarlo basta notar que: $(X_n - X)(Y_n - Y) \xrightarrow{\mathcal{P}} 0$ y luego $X_n Y_n - XY_n - X_n Y + XY \xrightarrow{\mathcal{P}} 0$ (Por la propiedad siguiente)

8. $X_n \xrightarrow{\mathcal{P}} X$, entonces $YX_n \xrightarrow{\mathcal{P}} YX$

Teorema 3.1.1 Si $X_n \xrightarrow{\mathcal{P}} X$ y $g(\cdot)$ es continua, entonces se cumple que:

$$g(X_n) \xrightarrow{\mathcal{P}} g(X)$$

Demostración 3.1.3 $\forall \varepsilon > 0, \exists k > 0 : \mathcal{P}[|X| \geq k] < \frac{\varepsilon}{2}$. Consideramos el compacto $[-k, k]$, $A = [|X| < k]$, $B_n = [|X_n - X| < \delta]$ y $C = [|g(X_n) - g(X)| < \varepsilon]$. Tenemos que ver si $\mathcal{P}(B_n) \rightarrow 1 \Rightarrow \mathcal{P}(C) \rightarrow 1$:

$$\begin{aligned} AB \subset C &\Rightarrow C^c \subset A^c \cup B^c \Rightarrow \mathcal{P}(C^c) \leq \mathcal{P}(A^c) + \mathcal{P}(B^c) \leq \\ &\leq \mathcal{P}[|X| \geq k] + \mathcal{P}[|X_n - X| \geq \delta] \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

3.1.4. Definición de convergencia casi segura

Una sucesión de variables aleatorias, X_n , converge con probabilidad 1, o de forma casi segura, a una variable aleatoria X (que puede degenerar en una constante K) cuando se cumple que:

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1$$

De esta forma interpretamos que $X_n \xrightarrow{c.s.} X$ cuando la probabilidad de que en el límite la sucesión de variables aleatorias y aquella a la que converge sean iguales es 1.

Teorema 3.1.2

$$X_n \xrightarrow{c.s.} X \implies X_n \xrightarrow{\mathcal{P}} X$$

$$X_n \text{ finitas c.s.}, X_n \xrightarrow{\mathcal{P}} X \implies \exists X_{n_k} : X_{n_k} \xrightarrow{c.s.}_{k \rightarrow \infty} X$$

Demostración 3.1.4

$$0 = \lim_{n \rightarrow \infty} P \bigcup_{m \geq n} [|X_m - X| \geq \epsilon] \geq \lim_{n \rightarrow \infty} P[|X_n - X| \geq \epsilon]$$

■

Lema 3.1.3 Sea \mathbb{X}_n finita. Entonces se tiene que:

$$\mathbb{X}_n \xrightarrow{c.s.} \mathbb{X} \iff \mathbb{X}_n \text{ es de Cauchy}$$

Demostración 3.1.5

\implies

$\exists n_0 : \forall n \geq n_0$ se tiene que:

$$|\mathbb{X}_n(\omega) - \mathbb{X}(\omega)| < \frac{\epsilon}{2} \quad \omega \in A, \quad P(A) = 1$$

$$|\mathbb{X}_n(\omega) - \mathbb{X}_m(\omega)| \leq |\mathbb{X}_n(\omega) - \mathbb{X}(\omega)| + |\mathbb{X}_m(\omega) - \mathbb{X}(\omega)| < \epsilon$$

\Longleftarrow

$\mathbb{X}_n(\omega)$ de Cauchy

$$\lim \mathbb{X}_n(\omega) = \mathbb{X}(\omega) \quad \omega \in A$$

■

Capítulo 4

Esperanza matemática

4.1. Integral de una función de conjunto

Para el cálculo de probabilidades, nos será muy útil el concepto de integral sobre funciones de conjunto. Este concepto nos servirá para calcular lo que se conoce como la esperanza matemática de una variable aleatoria X . En este apartado trabajaremos sobre el espacio de probabilidad $(\Omega, \mathcal{A}, \mathcal{P})$. Comenzaremos definiendo la integral para las funciones simples, para dar luego una definición de integral para funciones no negativas y por último para funciones cualesquiera.

Sea entonces $\{A_k\} \in \mathcal{A}$, tal que $\sum_k A_k = \Omega$, partición medible del espacio.

Sea entonces la función simple $X = \sum_{j=1}^m x_j I_{A_j}$, $x_j \geq 0$. La integral de la función X se define como:

$$\int_{\Omega} X d\mathcal{P} = \sum_{j=1}^m x_j \mathcal{P}_{A_j}$$

Ahora, para cualquier función no negativa X , se define la integral de la función como:

$$\int_{\Omega} X d\mathcal{P} = \lim \int_{\Omega} X_n d\mathcal{P}$$

Donde $\{X_n\} \rightarrow X$. Finalmente, la integral en Ω de una función medible X se define como:

$$\int_{\Omega} X d\mathcal{P} = \int_{\Omega} X^+ d\mathcal{P} - \int_{\Omega} X^- d\mathcal{P}$$

donde $X^+ = XI_{[X \geq 0]}$ y $X^- = -XI_{[X < 0]}$. Si $\int_{\Omega} X d\mathcal{P}$ es finita, es decir, si los dos términos de la diferencia anterior son finitos, entonces se dice que X es integrable en Ω . Ahora, una vez definida la integral, vamos a ver algunas de sus propiedades. Tenemos primero una serie de propiedades relacionadas con la aditividad de la integral.

Teorema 4.1.1 Sean X, Y dos funciones medibles, entonces :

$$\int (X + Y) d\mathcal{P} = \int X d\mathcal{P} + \int Y d\mathcal{P}$$

Demostración 4.1.1 Sean $X = \sum x_j I_{A_j}$ e $Y = \sum y_k I_{B_k}$. Entonces

$$X + Y = \sum_k \sum_j x_j I_{A_j B_k} + \sum_k \sum_j y_k I_{A_j B_k} = \sum_k \sum_j (x_k + y_j) I_{A_j B_k}$$

Si calculamos ahora las integrales:

$$\begin{aligned} \int (X + Y) d\mathcal{P} &= \sum_k + \sum_j (x_j + y_k) \mathcal{P}(A_j B_k) = \\ &= \sum_j x_j \mathcal{P}(A_j) + \sum_k y_k \mathcal{P}(B_k) = \int X d\mathcal{P} + \int Y d\mathcal{P} \end{aligned}$$

■

Corolario 1: $\int_{A+B} X d\mathcal{P} = \int_A X d\mathcal{P} + \int_B X d\mathcal{P}$

Corolario 2: $\int cX d\mathcal{P} = c \int X d\mathcal{P}$

Por último veamos ahora algunas propiedades relacionadas con el orden:

- $X \geq 0 \Rightarrow \int X d\mathcal{P} \geq \int 0 = 0$
- $X \geq Y \Rightarrow \int X d\mathcal{P} \geq \int Y d\mathcal{P}$
- $X \stackrel{c.s.}{=} Y \Rightarrow \int X d\mathcal{P} = \int Y d\mathcal{P}$

4.2. Esperanza matemática

4.2.1. Definición general

Dado un espacio probabilístico (Ω, \mathcal{A}, P) y una v.a. X , se define la esperanza matemática de X como:

$$E(X) = \int_{\Omega} X dP = \int_{\omega \in \Omega} X(\omega) dP(\omega)$$

La esperanza matemática se define de forma distinta si estamos trabajando con una v.a. discreta o continua. En el caso de una v.a. discreta, sea $P(x_i)$ su función de probabilidad. Definimos la esperanza matemática como:

$$E(X) = x_1 P[X = x_1] + \dots + x_n P[X = x_n] = \sum_{i=1}^n x_i P(x_i)$$

En el caso de una v.a. continua, sea $f(x)$ la función de densidad. Definimos la esperanza matemática como:

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

4.2.2. Teorema de la convergencia monótona

Una vez vista la definición de integral y esperanza matemática y algunas de sus propiedades, vamos a enunciar y demostrar un teorema de convergencia que nos será de mucha utilidad para el estudio de variables aleatorias. Veamos su enunciado y demostración:

Teorema 4.2.1

Teorema de la convergencia monótona para funciones medibles no negativas

Sea $\{X_n\} \geq 0$ tal que $X_n \uparrow X$. Entonces, se tiene que $\int X_n \uparrow \int X$

Demostración 4.2.1

Tomamos las sucesiones $X_{kn} \uparrow X_k$. La sucesión $Y_n = \max_{k \leq n} X_{kn}$ es una sucesión de funciones simples no negativas y no decreciente, y además

$$X_{kn} \leq Y_n \leq X_n \quad \int X_{kn} \leq \int Y_n \leq \int X_n$$

Ahora, cuando hacemos $\lim_{n \rightarrow \infty}$, tenemos que

$$X_k \leq \lim Y_n \leq X \quad \int X_k \leq \int \lim Y_n \leq \lim \int X_n$$

Por último, cuando $k \rightarrow \infty$, obtenemos

$$X \leq \lim Y_n \leq X \quad \lim \int X_n \leq \int \lim Y_n \leq \lim \int X_n$$

De donde extraemos que $\lim Y_n = X$ y que $\int X = \lim \int X_n$ ■

4.2.3. Teorema de Fatou-Lebesgue

Veamos ahora un resultado que nos da información sobre el límite superior e inferior de una sucesión de variables aleatorias, en caso de que estas estén acotadas por variables aleatorias integrables

Teorema 4.2.2 *de Fatou-Lebesgue*

Sean Y, Z dos funciones integrables (pueden no mantener su signo) y X_n una sucesión de variables aleatorias, entonces:

1. $\forall n \in \mathbb{N} \ Y \leq X_n \implies E(\liminf X_n) \leq \liminf E(X_n)$
2. $\forall n \in \mathbb{N} \ X_n \leq Z \implies E(\limsup X_n) \geq \limsup E(X_n)$

Demostración 4.2.2 Si $X_n \geq 0$, tenemos que:

$$X_n \geq Y_n = \inf_{k \geq n} X_k \uparrow \liminf X_n \xrightarrow{TCM} \liminf E(X_n) \geq \lim Y_n = E(\liminf X_n)$$

En otro caso:

1. $X_n - Y \geq 0 \implies \liminf E(X_n) - \liminf E(Y) \geq E(\liminf X_n) - E(\liminf Y)$
2. $Z - X_n \geq 0 \implies -\liminf E(-X_n) \geq -E(\liminf -X_n) \implies \limsup E(X_n) \leq E(\limsup X_n)$ ■

4.2.4. Teorema de la convergencia dominada

Teorema 4.2.3 Sea X_n una sucesión de variables aleatorias t.q. $X_n \xrightarrow{\mathcal{P}} X$ o $X_n \xrightarrow{\text{c.s.}} X$. Si $\exists Y$ t.q. $|X_n| \leq Y$. Entonces se tiene que X es integrable y:

$$\lim \int X_n dP = \int X dP$$

Demostración 4.2.3 Por un lado, si tenemos la convergencia casi segura, $|X_n| \leq Y \Rightarrow -Y \leq X_n \leq Y \xrightarrow{F-L} \int X_n \rightarrow \int X$. Por otro, si la convergencia es en probabilidad, nos bastaría con probar que si $Y_n \xrightarrow{\mathcal{P}} 0 \Rightarrow \int Y_n \xrightarrow{\mathcal{P}} 0$. Tomamos entonces $Y_n = |X_n - X|$

$$Y_n \xrightarrow{\mathcal{P}} 0 \Rightarrow \exists \{Y'_n\} \subset \{Y_n\} \text{ t.q. } \int Y'_n \rightarrow \limsup \int Y_n$$

De nuevo, tenemos que $Y'_n \xrightarrow{\mathcal{P}} 0$. Tomando ahora $\{Y''_n\} \subset \{Y'_n\}$ t.q. $Y''_n \xrightarrow{\text{c.s.}} 0 \Rightarrow \int Y''_n \rightarrow 0 \xRightarrow{\text{unic.lim.}} \limsup \int Y_n = 0 \xRightarrow{Y_n \geq 0} \int Y_n \rightarrow 0$ ■

4.2.5. Desigualdades

Vamos a ver algunas desigualdades que nos permitirán a posteriori simplificar el cálculo de características asociadas a variables aleatorias.

Definición: Momento r-ésimo

Dado un espacio probabilístico $(\Omega, \mathcal{A}, \mathcal{P})$ y una variable aleatoria X , se define el momento r-ésimo o de orden r de la variable X a:

$$EX^r = \int X^r d\mathcal{P}$$

Dicho momento existe si $E|X|^r < \infty$

Desigualdad C_r

Lema 4.2.1

$|a + b|^r \leq C_r |a|^r + C_r |b|^r$ con $C_r = 1$ si $r \leq 1$ y $C_r = 2^{r-1}$ si $r > 1$

Entonces se tiene $E|X + Y|^r \leq C_r E|X|^r + C_r E|Y|^r$

Desigualdad de Hölder

$$E|XY| \leq (E|X|^r)^{\frac{1}{r}} (E|Y|^s)^{\frac{1}{s}} \text{ con } r > 1 ; \frac{1}{r} + \frac{1}{s} = 1$$

Caso particular: Desigualdad de Schwarz

$$E|XY| \leq (E|X|^2)^{\frac{1}{2}} (E|Y|^2)^{\frac{1}{2}}$$

Desigualdad de Minkowsky

$$\text{Si } r \geq 1, \text{ entonces } (E|X + X'|^r)^{\frac{1}{r}} \leq (E|X|^r)^{\frac{1}{r}} + (E|X'|^r)^{\frac{1}{r}}$$

Desigualdad básica

Sea X una v.a. y g una función boreliana no negativa en \mathbb{R} , se tiene:

i) Si g es par y no decreciente en $[0, \infty)$, entonces $\forall a > 0$:

$$\frac{Eg(X) - g(a)}{c.s. \sup g(X)} \leq P[|X| \geq a] \leq \frac{Eg(X)}{g(a)}$$

ii) (Desigualdad de Tchevychev)

Si $g(x) = x^2$ y cambiamos la v.a. por $X - EX$, por ser g no decreciente en \mathbb{R} , podemos prescindir del valor absoluto:

$$P[X - EX \geq a] \leq \frac{Var X}{a^2}$$

iii) (Desigualdad de Markov) Si $g(X) = |X|^r$

$$\frac{EX^r - a^r}{c.s. \sup X^r} \leq P[|X| \geq a] \leq \frac{EX^r}{a^r}$$

Vamos a demostrar que en efecto se cumple la desigualdad básica:

Demostración 4.2.4 Definimos $A = [|X| \geq a]$

$$Eg(A) = \int_A g(X) d\mathcal{P} + \int_{A^c} g(X) d\mathcal{P}$$

$$g(a)\mathcal{P}(A) \leq \int_A g(X) d\mathcal{P} \leq c.s. \sup g(X)\mathcal{P}(A)$$

$$0 \leq \int_{A^c} g(X) d\mathcal{P} \leq g(a)$$

$$g(a)\mathcal{P}(A) \leq Eg(X) \leq c.s. \sup g(X)\mathcal{P}(A) + g(a)$$

■

Aplicaciones

1. Usando la desigualdad básica con $g(X) = X^r$ y $E|X_n - X|^r \rightarrow 0 \Rightarrow X_n \xrightarrow{\mathcal{P}} X$
2. $(\Omega, \mathcal{A}, \mathcal{P}) \xrightarrow{\mathcal{P}} (\mathbb{R}, \mathcal{B})$, definimos ahora $\mathcal{P}^X(\mathcal{B}) = \mathcal{P}(X^{-1}(\mathcal{B}))$. \mathcal{P}^X es una probabilidad, ya que $\mathcal{P}^X(\mathbb{R}) = 1$, $\mathcal{P}^X(\mathcal{B}) \geq 0$, $\mathcal{P}^X(\sum B_j) = \sum \mathcal{P}^X(B_j)$, y $EX = \int_{x \in \mathbb{R}} x d\mathcal{P}^X(x)$; $Eg(X) = \int_{y \in \mathbb{R}} g d\mathcal{P}^Y(y) = \int_{x \in \mathbb{R}} g(x) d\mathcal{P}^X(x)$

4.3. Convergencia en r-medida

Una sucesión de variables aleatorias X_n converge en r-medida a X si:

$$E|X_n - X|^r \rightarrow 0$$

Lo notamos como $X_n \xrightarrow{r} X$

Lema 4.3.1 $X_n \xrightarrow{r} X \Rightarrow E|X_n|^r \rightarrow E|X|^r$

Demostración 4.3.1 *Tenemos dos casos, para $r \leq 1$ y $r \geq 1$*

Para $r \leq 1$: $|E|X_n|^r - E|X|^r| \leq E|X_n - X|^r \rightarrow 0$

Para $r \geq 1$: $|(E|X_n|^r)^{\frac{1}{r}} - (E|X|^r)^{\frac{1}{r}}| \leq (E|X_n - X|^r)^{\frac{1}{r}} \rightarrow 0$

4.4. Función de distribución, distribución inducida y función generatriz de momentos**4.4.1. Función de distribución**

Función puntual que se define sobre la recta real, no negativa, no decreciente, continua por la izquierda y que cumple que $F(-\infty) = 0$, $F(+\infty) = 1$

4.4.2. Distribución inducida

Se conoce como distribución inducida por la variable aleatoria X a $(\mathbb{R}, \mathcal{B}, \mathcal{P}^X)$, donde la medida de probabilidad \mathcal{P}^X se define como:

$$\mathcal{P}^X(B) = \mathcal{P}(X^{-1}(B)) = \mathcal{P}[X \in B], \forall B \in \mathcal{B}$$

Tenemos entonces que:

- $\mathcal{P}^X(B) \geq 0, \forall B \in \mathcal{B}$
- $\mathcal{P}^X\left(\sum_j B_j\right) = \mathcal{P}(X^{-1}(\sum_j B_j)) = \mathcal{P}\left(\sum_j X^{-1}(B_j)\right) = \sum_j \mathcal{P}(X^{-1}(B_j)) = \sum_j \mathcal{P}^X(B_j)$
- $EX = \int_{\omega \in \Omega} X(\omega) d\mathcal{P}(\omega) = \int_{x \in \mathbb{R}} x d\mathcal{P}^X(x)$

4.4.3. Función generatriz de momentos

Llamamos función generatriz de momentos de la variable aleatoria X a la función:

$$M_X(t) = E(e^{tX}) \quad t \in \mathbb{R}$$

4.4.4. Función característica

Dada una variable aleatoria X continua, definimos su función característica:

$$\varphi_X(t) = E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx = E(\cos(tX)) + iE(\sin(tX))$$

Capítulo 5

Independencia

5.1. Concepto de independencia

Definición: Se dice que los eventos A_t son independientes si para todo subconjunto finito (t_i, \dots, t_n) se tiene que:

$$P \bigcap_{k=1}^n A_{t_k} = \prod_{k=1}^n P A_{t_k}$$

De hecho, el concepto de independencia es relativo a la familia de clases. Una colección de clases de eventos se dice independiente si seleccionando un evento arbitrario de cada clase dichos eventos son independientes.

Teorema 5.1.1

Se tiene que las subclases de clases independientes son independientes.

Demostración 5.1.1 *La demostración es trivial puesto que si todos los conjuntos de una clase cumplen la propiedad, también la han de cumplir un subconjunto de esta.*

5.2. Suma de variables aleatorias independientes

Teorema de las funciones Borel

Teorema 5.2.1 *Las funciones de Borel $g(X_i)$ de funciones aleatorias independientes son independientes.*

Teorema de extensión

Teorema 5.2.2 *Los σ -campos minimales sobre clases independientes cerradas para la intersección finita son independientes.*

Esto tiene tres aplicaciones fundamentales:

1. Si los eventos A_t son independientes, también lo son los σ campos $(A_t, A_t^c, \emptyset, \Omega)$.
2. Si la imagen inversa \mathcal{C}_t de las clases de todos los intervalos $(-\infty, x_t)$ en los espacios de Borel R_t son independientes, también lo son las imágenes inversas \mathcal{B}_t de los campos de Borel en R_t . Para todo \mathcal{C}_t , éste es cerrado para la intersección finita y \mathcal{B}_t es el σ campo minimal sobre \mathcal{C}_t
3. Sean \mathcal{B}_t campos de eventos y sea T_s un subconjunto de T . El σ campo \mathcal{B}_T compuesto por componentes \mathcal{B}_t , con $t \in T_s$, es el σ campo minimal sobre la clase \mathcal{C} de todas las intersecciones finitas de eventos A_t y contiene a todos sus componentes

Teorema de composición

Teorema 5.2.3 *Los σ -campos compuestos son independientes si y solo si sus σ -subcampos finitamente compuestos son independientes*

5.2.1. Familia de variables aleatorias

Una familia de variables aleatorias $X_{T_s} = \{X_t : t \in T_s\}$ induce σ -campos $\mathcal{B}(X_{T_s})$ de eventos

Teorema de las familias

Teorema 5.2.4 *Las familias de variables aleatorias son independientes si y solo si sus subfamilias finitas son mutuamente independientes*

5.3. Multiplicación de variables aleatorias independientes

Definición: Las variables aleatorias $X_t, t \in T$ son independientes si para cada clase finita $(S_{t_1}, \dots, S_{t_n})$ de conjuntos de Borel en \mathbb{R} ,

$$P \bigcap_{k=1}^n [X_{t_k} \in S_{t_k}] = \prod_{k=1}^n P[X_{t_k} \in S_{t_k}]$$

Lema de la multiplicación

Lema 5.3.1 Si X_1, \dots, X_n son variables aleatorias independientes no negativas, entonces $E \prod_{k=1}^n X_k = \prod_{k=1}^n EX_k$

Demostración 5.3.1 Basta demostrar el resultado para dos variables, dado que el resto se demuestra por inducción. En primer lugar, veamos que se verifica la condición para funciones simples.

Sean $X = \sum_j x_j I_{A_j}, Y = \sum_k y_k I_{B_k}$ variables simples no negativas, donde $A_j = [X = x_j], B_k = [Y = y_k]$.

Por ser X, Y independientes, $P(A_j B_k) = P(A_j)P(B_k) \Rightarrow E(XY) = \sum_{k,j} P(A_j B_k) = \sum_{j,k} P(A_j)P(B_k) = E(X)E(Y)$

Sean ahora X, Y variables aleatorias no negativas independientes. Su independencia implica la de

$$A_{nj} = \left[\frac{j-1}{2^n} \leq X < \frac{j}{2^n} \right], \quad B_{nk} = \left[\frac{k-1}{2^n} \leq Y < \frac{k}{2^n} \right]$$

Finalmente, tendremos la independencia de las variables aleatorias simples:

$$X_n = \sum_{j=1}^{n2^n} \frac{j-1}{2^n} I_{A_{nj}} \quad Y_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{B_{nk}}$$

Tenemos entonces que:

$$0 \leq X_n \uparrow X, \quad 0 \leq Y_n \uparrow Y \implies X_n Y_n \uparrow XY$$

$$E(X_n Y_n) = E(X_n)E(Y_n)$$

De donde se extrae por el teorema de la convergencia monótona que $E(XY) = E(X)E(Y)$

■

Teorema de la multiplicación

Teorema 5.3.1 Sean X_1, \dots, X_n variables aleatorias independientes. Si son integrables, entonces también lo es su producto con:

$$E \prod_{k=1}^n X_k = \prod_{k=1}^n EX_k$$

A la inversa, si su producto es integrable y ninguna de ellas degenera a 0, entonces todas ellas son integrables.

Demostración 5.3.2

\Rightarrow La independencia de X, Y nos da la independencia de X', Y' , donde $X' \in \{X^+, X^-, |X|\}$, $Y' \in \{Y^+, Y^-, |Y|\}$, por lo que $E(X'Y') = E(X')E(Y')$.

Si X, Y son integrables, también lo son X', Y' y por la anterior igualdad también lo es $X'Y'$, lo cual nos da también $|XY|$, y por tanto, XY integrables con:

$$\begin{aligned} E(XY) &= E[(X^+ - X^-)(Y^+ - Y^-)] = E(X^+)E(Y^+) - E(X^+)E(Y^-) \\ &\quad - E(X^-)E(Y^+) + E(X^-)E(Y^-) = E(X)E(Y) < \infty \end{aligned}$$

\Leftarrow Si XY son integrables, entonces $E(|X|)E(|Y|) = E(|XY|)$. Como ni X ni Y degeneran a 0, $E(|X|)$ y $E(|Y|)$ son finitas y, por tanto, X e Y son integrables. ■

Teorema de equivalencia

Teorema 5.3.2 Para cada clase finita de conjuntos de Borel S_t y de puntos $x_t, u_t \in \mathbb{R}$, son equivalentes las tres definiciones siguientes de independencia de variables aleatorias X_t :

1. $P \bigcap_{k=1}^n [X_{tk} \in S_{tk}] = \prod_{k=1}^n P[X_{tk} \in S_{tk}]$
2. $F_{t_1 \dots t_n}(x_{t_1}, \dots, x_{t_n}) = F_{t_1}(x_{t_1}) \dots F_{t_n}(x_{t_n})$
3. $f_{t_1 \dots t_n}(u_{t_1}, \dots, u_{t_n}) = f_{t_1}(u_{t_1}) \dots f_{t_n}(u_{t_n})$

Demostración 5.3.3

i) \Rightarrow ii) Tomamos $S_t =]-\infty, x_t[$ y aplicamos la definición.

5.3. MULTIPLICACIÓN DE VARIABLES ALEATORIAS INDEPENDIENTES 43

ii) \Rightarrow i) $S_t =]-\infty, x_t[, \mathcal{C}_t =]-\infty, x_t[, t \in T$ son independientes. Por el teorema de extensión, los σ -campos minimales sobre ellos son independientes.

i) \Rightarrow iii) Aplicación directa del teorema de la multiplicación.

iii) \Rightarrow ii) Tenemos que $F_{t_1, \dots, t_n}([a_{t_1}, \dots, a_{t_n}; b_{t_1}, \dots, b_{t_n}]) = F_{t_1}([a_{t_1}, \dots, a_{t_n}; b_{t_1}, \dots, b_{t_n}]) F_{t_2}([a_{t_1}, \dots, a_{t_n}; b_{t_1}, \dots, b_{t_n}]) \dots F_{t_n}([a_{t_1}, \dots, a_{t_n}; b_{t_1}, \dots, b_{t_n}])$. Haciendo $a_t \rightarrow -\infty, b \uparrow x_t$ tenemos ii)

■

Lema de Borel-Cantelli

Lema 5.3.2 Lema de Borel-Cantelli

Sea (Ω, \mathcal{A}, P) , $A_n \in \mathcal{A}$. Entonces:

Si A_n son independientes:

- $\sum P(A_n) < \infty \implies P(\limsup A_n) = 0$
- $\sum P(A_n) = \infty \implies P(\limsup A_n) = 1$

Demostración 5.3.4

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} B_n$$

$$P(\limsup A_n) = P \lim B_n = \lim_n P B_n = \lim_n P \bigcup_{k=n}^{\infty} A_k \leq \lim_n \sum_{k=n}^{\infty} P A_k = 0$$

El recíproco no es cierto.

■

Criterio cero-uno de Borel

Criterio 5.3.1

- Si A_n son independientes
 - $\sum P(A_n) < \infty \implies P(\limsup A_n) = 0$
 - $\sum P(A_n) = \infty \implies P(\limsup A_n) = 1$

Demostración 5.3.5

Usamos las sucesiones definidas anteriormente, para el caso 0 queda probado con el lema de Borel-Cantelli, para el caso 1, se tiene:

$$P(\limsup A_n)^c = P(\liminf A_n^c) = P\left(\bigcup_n \bigcap_{k=n} A_k^c\right) = P\left(\lim_n \bigcap_{k=n} A_k^c\right) =$$

$$\lim_n P\left(\bigcap_{k=n} A_k^c\right) = \lim_n \prod_{k=n} (1 - P(A_k)) \leq \lim_n \prod_{k=n} e^{-P(A_k)} = \lim_n e^{-\sum_{k=n} P(A_k)} = 0$$

■

De esta ley se extraen una serie de corolarios importantes:

Corolario 1: Si A_n son independientes:

$$P(\limsup(A_n)) = 0 \Leftrightarrow \sum_{n \in \mathbb{N}} P(A_n) < \infty$$

Corolario 2: Sea X_n una colección de variables aleatorias independientes, entonces;

$$X_n \xrightarrow{c.s.} 0 \Leftrightarrow \sum_n P([X_n \geq \varepsilon]) < \infty$$

Demostración

$$\sum P([|X_n| \geq \varepsilon]) = 0 \Leftrightarrow P(\limsup([|X_n| \geq \varepsilon])) = 0 \Leftrightarrow P\left(\bigcap_n \bigcup_{k=n} [|X_k| \geq \varepsilon]\right) = 0$$

$$\Leftrightarrow P\left(\lim_n \bigcup_{k=n} [|X_k| \geq \varepsilon]\right) = 0 \Leftrightarrow \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n} [|X_k| \geq \varepsilon]\right) = 0 \Leftrightarrow X_n \xrightarrow{c.s.} 0$$

■

Corolario 3:

X_n son independientes y $X_n \xrightarrow{\mathbb{P}} 0 \Leftrightarrow \exists X_{n_k} \subset X_n$ tal que $X_{n_k} \xrightarrow{c.s.} 0$

Demostración

$$\exists X_{n_k} : P\left[|X_{n_k}| \geq \frac{1}{2^k}\right] \leq \frac{1}{2^k}$$

$$\sum_k P\left[|X_{n_k}| \geq \frac{1}{2^k}\right] \leq \sum_k \frac{1}{2^k} = 1 \Leftrightarrow X_{n_k} \xrightarrow{c.s.} 0$$

■

Pasamos ahora a ver otra ley importante en el estudio de la probabilidad, la cual se conoce como la ley 0-1 de Kolmogorov. Para ello, comenzamos definiendo el concepto de σ -campo cola.

Sean X_n independientes, tenemos las siguientes relaciones entre los σ -campos asociados:

$$\begin{aligned}\mathcal{B}(X_n, X_{n+1}) &\supseteq \mathcal{B}(X_n) \\ \mathcal{B}(X_n, X_{n+1}, \dots) &\supseteq \mathcal{B}(X_{n+1}, \dots) \text{ sucesión decreciente} \\ \mathcal{B}(X_n, X_{n+1}, \dots), \mathcal{B}(X_{n+1}, X_{n+2}, \dots), \dots &\downarrow \lim_n \sup \mathcal{B}(X_n) \equiv \mathcal{C}\end{aligned}$$

Dicho límite es un σ -campo, conocido como el σ -campo cola, que será equivalente a \emptyset o Ω .

5.3.1. Ley 0-1 de Kolmogorov

Sea X_n una sucesión de variables aleatorias independientes. Entonces los eventos en la cola tienen o bien probabilidad 0 o 1 y las funciones cola son degeneradas.

Demostración 5.3.6

- i) Por un lado, tenemos que: $\mathcal{C} \subseteq \mathcal{B}(X_{n+1}, \dots)$
 Además, $\mathcal{B}(X_1, \dots, X_n)$ es independiente de $\mathcal{B}(X_{n+1}, X_{n+2}, \dots)$, lo que implica que $\mathcal{C}, \mathcal{B}(X_1, \dots, X_n)$ son independientes. De aquí tenemos que \mathcal{C} es independiente de $\mathcal{B}(X_1, \dots, X_n, X_{n+1}, \dots)$, por lo que \mathcal{C} es independiente consigo mismo, ya que está contenido en el σ -campo que hace medibles a todos. Una vez aquí, tenemos que:

$$P(AA) = P(A)P(A) = P(A) \quad \forall A \in \mathcal{C} \Rightarrow P(A) = \{0, 1\} \quad \forall A \in \mathcal{C}$$

- ii) Los límites de las sucesiones de variables aleatorias independientes son \mathcal{C} -medibles (también para \limsup). Por tanto, es una variable aleatoria que degenera

■

A raíz de esta ley, podemos extraer también un corolario.

Corolario. Sea X_n una sucesión de variables aleatorias independientes, entonces X_n converge c.s. o diverge c.s., igualmente para la serie $\sum_{n \in \mathbb{N}} X_n$

5.4. Convergencia y estabilidad de sumas; centramiento y truncamiento

Esta sección está dedicada a la investigación de las sumas de variables aleatorias independientes y especialmente, a sus propiedades en el límite. Primero

de todo, consideraremos $S_n = \sum_{k=1}^n X_k$ una suma de variables aleatorias independientes X_1, X_2, \dots, X_n

Definición de estabilidad: Si existe una sucesión $\{a_n\}$ y otra $\{b_n\} \rightarrow \infty$ de forma que $\frac{S_n}{b_n} - a_n \xrightarrow{\mathcal{P}} 0$ o $\frac{S_n}{b_n} - a_n \xrightarrow{c.s.} 0$, entonces decimos que la sucesión S_n es estable en probabilidad o c.s.

A continuación presentamos una ley que ha marcado el desarrollo de toda la teoría de la probabilidad.

5.4.1. Ley de Bernoulli de los grandes números

Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas tales que $P[X_n = 1] = p$ y $P[X_n = 0] = q = 1 - p$. Entonces se tiene que:

$$\frac{S_n}{n} - p \xrightarrow{\mathcal{P}} 0$$

El Teorema del Límite Central, que desarrollaremos más adelante, es una consecuencia directa del perfeccionamiento de esta ley por parte de Moivre y Laplace.

5.4.2. Ley fuerte de Borel de los grandes números

Con las hipótesis anteriores de la ley de Bernoulli de los grandes números, entonces:

$$\frac{S_n}{n} - p \xrightarrow{c.s.} 0$$

Teorema 5.4.1 Sea $f(x)$ continua en $[0, 1] \Rightarrow \exists P_n(x)$ tal que $\lim P_n(x) = f(x)$ uniformemente $\forall x \in [0, 1]$.

5.4.3. Centramiento y truncamiento

Diremos que centramos X a c si reemplazamos X por $X - c$. Si X es integrable, entonces podemos centrarla a su esperanza EX de forma que X es reemplazada por $X - EX$. Es decir, diremos que una variable aleatoria está centrada a su media si y solo si su media existe y es igual a 0.

Sea X integrable, llamaremos varianza de X al segundo momento de $X - EX$ (podría ser ∞) y lo denotaremos como $\sigma^2 X$. Por consiguiente:

$$\sigma^2 X = \text{Var}(X) = E(X - EX)^2 = EX^2 - (EX)^2$$

Para todo c finito si centramos una variable en c su varianza no cambia:

$$\sigma^2(X - c) = E(X - c - E(X - c))^2 = E(X - EX)^2$$

Una vez visto el centramiento vamos con el **truncamiento**, útil cuando no existe la media.

Diremos que truncamos X a $c > 0$ (finita) cuando reemplazamos X por X^c si $|X| < c$ o por 0 si $|X| \geq c$, se define la variable X^c como X truncada a c .

Tenemos entonces que si F es el diferencial de X , entonces existen todos los momentos de X^c y son finitos; estos son:

$$EX^c = \int_{|X| < c} x dF, \quad E(X^c)^2 = \int_{|X| < c} x^2 dF, \quad \dots$$

Siempre podremos seleccionar un c suficientemente grande para hacer que $P[X \neq X^c] = P[|X| \geq c]$ sea arbitrariamente pequeño. Además nosotros también podemos seleccionar el c_j suficientemente grande para conseguir $P \bigcup [X_j \neq X_j^{c_j}]$ arbitrariamente pequeño, de tal forma que, dado $\epsilon > 0$, tenemos:

$$P \left(\bigcup [X_j \neq X_j^{c_j}] \right) \leq \sum P[|X_j| \geq c_j] < \frac{\epsilon}{2^j}$$

Por consiguiente, dada una familia numerable de variables aleatorias, podemos hacer corresponder una familia de variables aleatorias que difiere de la primera en un evento de probabilidad pequeña. Además, si estamos interesados principalmente en propiedades en el límite, no necesitamos probabilidades pequeñas arbitrarias.

Definición. Dadas dos sucesiones X_n y X'_n , se dice que son **cola-equivalentes** si difieren casi seguramente en un número finito de elementos. En otras palabras, si $\forall \omega \in \Omega$ existe un número finito $n(\omega)$ tal que para $n \geq n(\omega)$ las dos sucesiones $X_n(\omega)$ y $X'_n(\omega)$ son la misma.

X_n y X'_n son cola-equivalentes si $P(\limsup [X_n \neq X'_n]) = 0$.

Definición. Se dice que X_n y X'_n son **equivalentes en la convergencia** si el conjunto de convergencia de ambos es el mismo c.s.

Lema 5.4.1 *(de equivalencia)*

Si las series $\sum_{n \in \mathbb{N}} P[X_n \neq X'_n]$ convergen, entonces las sucesiones X_n y X'_n son cola-equivalentes, por tanto, las series $\sum_{n \in \mathbb{N}} X_n$ y $\sum_{n \in \mathbb{N}} X'_n$ son equivalentes en convergencia y las series $\frac{S_n}{b_n}$ y $\frac{S'_n}{b_n}$ con $b_n \rightarrow \infty$ convergen al mismo evento y al mismo límite, excluyendo el evento nulo.

Si las variables aleatorias X_n son independientes e integrables, entonces:

$$\sigma^2 S_n = \sum_{k=1}^n \sigma^2 X_k \quad (\text{Igualdad de Bienaymé})$$

Demostración

Sean X_1, X_2, \dots integrables, como el centramiento no modifica la varianza, podemos asumir, cuando computamos varianzas, que esas variables aleatorias están centradas a la media. Entonces:

$$\sigma^2 S_n = ES_n^2 = \sum_{k=1}^n EX_k^2 + \sum_{j,k=1, j \neq k}^n EX_j X_k = \sum_{k=1}^n \sigma^2 X_k$$

como X_j y X_k son independientes, tenemos que:

$$EX_j X_k = EX_j \cdot EX_k = 0$$

■

Entonces obtenemos la **desigualdad de Bienaymé-Tchebychev**:

$$\frac{\sum_{k=1}^n \sigma^2 X_k - \epsilon^2}{c.s.\sup(S_n - ES_n)^2} \leq P[|S_n - ES_n| \geq \epsilon] \leq \frac{\sigma^2(S_n - ES_n)}{\epsilon^2} \leq \frac{1}{\epsilon^2} \sum_{k=1}^n \sigma^2 X_k$$

Esta desigualdad, aplicada a $(S_{n+k} - ES_{n+k}) - (S_n - ES_n)$ y remplazando ϵ por b_n y pasando al límite se tiene el siguiente corolario:

Corolario:

- Si las series $\sum \sigma^2 X_n$ convergen, entonces las series $\sum (X_n - EX_n)$ convergen en probabilidad.

- Si $\frac{1}{b_n^2} \sum_{k=1}^n \sigma^2 X_k \rightarrow 0$, entonces $\frac{S_n - ES_n}{b_n} \xrightarrow{\mathcal{P}} 0$

Demostración 5.4.1 *La demostración pasa por notar que:*

$$P \left[\sum_{k=1}^{n+h} (X_k - EX_k) \geq \epsilon \right] \leq \frac{\sum_{k=1}^{n+h} \sigma^2 X_k}{\epsilon^2} \rightarrow 0$$

Este último corolario es debido a Tchebichev (cuando $b_n = n$). En el caso de Bernoulli, donde $b_n = n$, $EX_n = p$, $\sigma^2 X_n = pq$ se reduce a la ley de Bernoulli de los grandes números.

5.4.4. Ley débil de los grandes números

Sea $\{X_n\}$ una sucesión infinita de variables aleatorias independientes que tienen el mismo valor esperado μ y varianza σ^2 , entonces

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mu$$

5.4.5. Ley fuerte de los grandes números de Kolmogorov

Sea X_n una sucesión de variables aleatorias independientes e idénticamente distribuidas con una ley común $\mathcal{L}(X)$. Entonces:

$$E|X| < \infty \iff \frac{S_n}{n} = \frac{\sum_{k=1}^n X_k}{n} \xrightarrow{c.s.} EX$$

Capítulo 6

El problema central del límite

El problema del límite central en probabilidad es el problema de convergencia de leyes de secuencias de sumas de variables aleatorias.

6.1. Preliminares

Consideramos S_n el número de ocurrencias de un evento con probabilidad p en n intentos idénticos e independientes. Para evitar trivialidades asumimos que $pq \neq 0$, donde $q = 1 - p$. Si X_k denota el indicador del evento en el k -ésimo intento, entonces $S_n = \sum_{k=1}^n X_k$ con $n = 1, 2, \dots$ donde los sumandos son indicadores independientes e idénticamente distribuidos. Este es el caso de Bernoulli; como $EX_k = p$ y $EX_k^2 = p$, entonces $\sigma^2 X_k = p - p^2 = pq$ de donde se sigue que:

$$ES_n = \sum_{k=1}^n EX_k = np \quad , \quad \sigma^2 S_n = \sum_{k=1}^n \sigma^2 X_k = npq$$

El primer teorema del límite central publicado en 1713 dice que $\frac{S_n}{n} \xrightarrow{P} p$

Puliendo este resultado, de Moivre obtuvo el segundo teorema de límites, que en la forma integral dada por Laplace, dice que:

$$P \left[\frac{S_n - np}{\sqrt{npq}} < x \right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy, \quad -\infty \leq x \leq \infty$$

El tercer teorema del límite central fue obtenido por Poisson, quien modificó el caso de Bernoulli asumiendo que la probabilidad $p = p_n$ depende del número de intentos de tal manera que $np_n \rightarrow \lambda > 0$. Por consiguiente, escribiendo ahora X_{nk} y S_{nn} en lugar de X_k y S_n , el caso de Poisson corresponde a una serie de sumas $S_{nn} = \sum_{k=1}^n X_{nk}$ donde, para cada n , los sumandos X_{nk} son indicadores independientes e idénticamente distribuidos con $P[X_{nk} = 1] = \frac{\lambda}{n} + o\left(\frac{1}{n}\right)$

A partir de aquí Poisson probó que para $k = 0, 1, 2, \dots$

$$P[S_{nn} = k] \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

Así nacieron las tres leyes básicas de la teoría de la probabilidad:

1. La ley de degeneración \mathcal{L} de una variable aleatoria degenera a 0 teniendo un punto de incremento solo en $x = 0$
2. La ley normal $N(0, 1)$ de una variable aleatoria normal con su función de densidad definida viene dada por:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-\frac{1}{2}y^2} dy$$

3. La ley de Poisson $\mathcal{P}(\lambda)$ de una variable aleatoria de Poisson con su función de densidad definida viene dada por:

$$F(x) = e^{-\lambda} \sum_{k=0}^x \frac{\lambda^k}{k!}$$

Para comprobar que son válidas, debemos ver que el límite de las distribuciones es justamente la distribución del límite:

1. **Ley degenerada:** Tenemos que ver que: $\varphi_X(t) = E(e^{itX}) = \int_{-\infty}^{\infty} = 1$

$$\varphi_{\frac{S_n - ES_n}{n}}(t) = E(e^{it \frac{S_n - ES_n}{n}}) = E(e^{it \sum_{k=1}^n X_k - EX_k \frac{1}{n}}) = E\left(\prod_{k=1}^n e^{it(X_k - p) \frac{1}{n}}\right) =$$

$$\prod_{k=1}^n E(e^{it \frac{X_k - p}{n}}) = \prod_{k=1}^n (pe^{\frac{itq}{n}} + qe^{\frac{-itp}{n}}) = (p(1 + \frac{ip}{n}t + \dots) + q(1 - \frac{ip}{n}t + \dots))^n =$$

$$((p + q) + o(\frac{1}{n}))^n \rightarrow 1$$

6.2. PRIMER TEOREMA DEL LÍMITE CENTRAL (PROBLEMA CLÁSICO) 53

2. **Ley normal:** Tenemos que ver que: $\varphi_X = E(e^{itX}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{itX} e^{-\frac{x^2}{2}} dx = e^{-\frac{t^2}{2}}$

$$\varphi_{\frac{S_n - E(S_n)}{\sqrt{npq}}} = E(e^{it \frac{S_n - E(S_n)}{\sqrt{npq}}}) = \prod_{k=1}^n E(e^{it \frac{X_k - p}{\sqrt{npq}}}) = \prod_{k=1}^n E(1 + \frac{i(X_k - p)}{\sqrt{npq}} t + \dots) =$$

$$\prod_{k=1}^n (1 + \frac{i^2 pq t^2}{npq} + \dots) = (1 + \frac{-t^2}{2n} + \dots)^n = e^{-\frac{t^2}{2}}$$

3. **Ley de Poisson** $\varphi_X = E(e^{itX}) = \sum_{x=0}^{\infty} e^{itX} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^{it} \lambda)^x}{x!} = e^{\lambda(e^{it} - 1)}$

Mientras que las dos primeras leyes juegan un papel central en el desarrollo de la teoría de la probabilidad, la de Poisson se ha mantenido ignorada durante mucho tiempo.

6.2. Primer teorema del límite central (Problema clásico)

En el caso de Bernouilli:

$$\mathcal{L}\left(\frac{S_n - ES_n}{n}\right) \rightarrow \mathcal{L}(0) \text{ y } \mathcal{L}\left(\frac{S_n - ES_n}{\sigma S_n}\right) \rightarrow N(0, 1)$$

y en el caso de Poisson:

$$\mathcal{L}(S_{nn}) \rightarrow P(\lambda)$$

Demostración 6.2.1

Si consideramos variables aleatorias centradas en media, podemos enunciar el siguiente lema. Pero antes notamos $s_n = \sigma(S_n)$ y además tendremos en cuenta que las variables aleatorias están centradas en media, por lo que $EX_k = 0$ y $ES_n = 0$ a partir de ahora

Lema 6.2.1 *Si los sumandos son independientes, idénticamente distribuidos y centrados en media, entonces $\mathcal{L}\left(\frac{S_n}{n}\right) \rightarrow \mathcal{L}$ y $\mathcal{L}\left(\frac{S_n}{s_n}\right) \rightarrow N(0, 1)$*

Teorema de composición y descomposición

Teorema 6.2.1 *Los tipos normales y degenerados son cerrados bajo composiciones y descomposiciones. Lo mismo también se cumple para cada familia de Leyes de Poisson $\mathcal{P}(\lambda; a, b)$ con el mismo b*

Una vez visto el lema anterior, enunciamos el siguiente teorema:

Teorema 6.2.2 Segunda extensión del problema clásico central del límite

Consideramos $S_n = \sum_{k=1}^n X_k$ y $s_n = \sigma S_n$ donde los sumandos son variables aleatorias independientes centradas en media. Entonces:

1. Si $\frac{1}{n^{1+\delta}} \sum_{k=1}^n E|X_k|^{1+\delta} \rightarrow 0$ para $0 \leq \delta \leq 1$. Entonces:

$$\mathcal{L}\left(\frac{S_n}{n}\right) \rightarrow \mathcal{L}(0)$$

2. Si $\frac{1}{s_n^{2+\delta}} \sum_{k=1}^n E|X_k|^{2+\delta} \rightarrow 0$ para un δ positivo, entonces:

$$\mathcal{L}\left(\frac{S_n}{s_n}\right) \rightarrow N(0, 1)$$

Demostración 6.2.2

Notar que si tomamos $\delta = 1$, el teorema se convierte en el conocido Teorema de Tchevichev. Además también contiene al resultado de Markov:

Resultado de Markov: Si $E|X_k|^{1+\delta} \leq c < \infty$, entonces $\mathcal{L}\left(\frac{S_n}{n}\right) \rightarrow \mathcal{L}(0)$, $\forall \delta > 0$

Lema 6.2.2 Si $S_{nn} = \sum_{k=1}^n X_{nk}$ donde los sumandos son variables aleatorias independientes (centradas en media), entonces:

1. Si $\frac{1}{n^2} \sum_{k=1}^n E|X_{nk}|^2 \rightarrow 0$ entonces $\mathcal{L}\left(\frac{S_{nn}}{n}\right) \rightarrow \delta(0)$
2. Si $\frac{1}{s_{nn}^3} \sum_{k=1}^n E|X_{nk}|^3 \rightarrow 0$ entonces $\mathcal{L}\left(\frac{S_{nn}}{s_{nn}}\right) \rightarrow N(0, 1)$

Lema de equivalencia de leyes

Lema 6.2.3 Si $X_n - X'_n \xrightarrow{\mathbb{P}} 0$ o $P[X_n \neq X'_n] \rightarrow 0$ entonces las sucesiones $\mathcal{L}(X_n)$ y $\mathcal{L}(X'_n)$ son equivalentes.

Una vistos estos resultados, ya estamos en posición de poder dar una solución completa al problema clásico del límite central, pero antes enunciaremos:

Criterio clásico de convergencia degenerada

Se tiene que $\mathcal{L}\left(\frac{S_n}{n}\right) \rightarrow \mathcal{L}(0)$ si y solamente si:

1. $\sum \int_{|x| \geq n} dF_k \rightarrow 0$
2. $\frac{1}{n} \sum \int_{|x| < n} x dF_k \rightarrow 0$
3. $\frac{1}{n^2} \sum \left(\int_{|x| < n} x^2 dF_k - \left(\int_{|x| < n} x dF_k \right)^2 \right) \rightarrow 0$

Criterio clásico de convergencia normal

Se tiene $\mathcal{L}\left(\frac{S_n}{s_n}\right) \rightarrow N(0, 1)$ y $\max_{k \leq n} \frac{\sigma_k}{s_n} \rightarrow 0$ si y solamente si, $\forall \epsilon > 0$:

$$g_n(\epsilon) = \frac{1}{s_n^2} \sum \int_{|x| \geq \epsilon s_n} x^2 dF_k \rightarrow 0$$