TEORÍA DE LA PROBABILIDAD

Esperanza de variables aleatorias

Introducción

Uno de los conceptos básicos de la teoría de la probabilidad es el de esperanza de una variable aleatoria. Su importancia es comparable con la del concepto mismo de probabilidad de un evento. De hecho, el concepto de esperanza y el de probabilidad surgieron en forma paralela cuando, a mediados del siglo $XVII$, se inicia el Cálculo de Probabilidades. Fue Christiaan Huygens quien introdujo este concepto en su libro *Du calcul dans les jeux de hasard*, publicado en el año 1657. En esa época, Blaise Pascal y Pierre de Fermat habían resuelto algunos problemas de probabilidad, con métodos que sentarían las bases para el desarrollo de una nueva disciplina matemática, el Cálculo de Probabilidades. Huygens resolvió, con sus propios métodos, los problemas que antes habían resuelto Pascal y Fermat y algunos otros más.

Uno de los aspectos interesantes de la metodología utilizada por Huygens es que en ningún momento se consideran ahí probabilidades de eventos, todas las soluciones están basadas en el cálculo de esperanzas, lo cual hacía ver ya que este concepto podía tomarse como primario, previo incluso al de probabilidad de un evento, y a partir de él desarrollar la nueva disciplina. La historia no fue de ese modo pues el concepto que prevaleció como primario fue el de probabilidad. Sin embargo, la historia misma mostraría más adelante que esta dualidad de importancia, entre el concepto de esperanza y el de probabilidad, que se dio al inicio del desarrollo de la teoría de la probabilidad como disciplina matemática, tenía fuertes raíces pues al evolucionar el concepto de probabilidad, hasta fusionarse con el de medida en los primeros años del siglo pasado, resultó palpable la estrecha relación entre ambos conceptos, de tal manera que, efectivamente, cualquiera de los dos puede tomarse como punto de partida, quedando inmersos uno dentro del otro. Esto último ya no únicamente dentro del contexto de la teoría de la probabilidad, sino dentro del contexto más amplio de la teoría de la medida, donde el concepto de probabilidad corresponde al de medida y el de esperanza al de integral.

Repaso de definiciones y resultados

1. Un **álgebra** $\mathcal{A}$ es una familia de subconjuntos de un conjunto $\mathbb{E}$, la cual tiene las siguientes propiedades:

a) $\mathbb{E} \in \mathcal{A}$.

b) $\mathcal{A}$ es cerrada bajo complementos.

c) $\mathcal{A}$ es cerrada bajo uniones finitas.

2. Una **σ-álgebra** $\mathcal{E}$ es una familia de subconjuntos de un conjunto $\mathbb{E}$, la cual tiene las siguientes propiedades:

a) $\mathbb{E} \in \mathcal{E}$.

b) $\mathcal{E}$ es cerrada bajo complementos.

c) $\mathcal{E}$ es cerrada bajo uniones infinitas numerables.

3. La **σ-álgebra de Borel en** $\mathbb{R}$ está generada por los intervalos de la forma $(-\infty, x]$, donde $x \in \mathbb{R}$.

4. La **σ-álgebra de Borel en** $\overline{\mathbb{R}}$ está generada por los intervalos de la forma $[-\infty, x]$, donde $x \in \mathbb{R}$.

5. Un **espacio medible** es una pareja $(\mathbb{E}, \mathcal{E})$, donde $\mathbb{E}$ es un conjunto cualquiera y $\mathcal{E}$ es una $\sigma-$algebra de subconjuntos de $\mathbb{E}$.

6. Dados un espacio medible $(\mathbb{E}, \mathcal{E})$, un conjunto cualquiera $\mathbb{F}$ y una función $f : \mathbb{E} \mapsto \mathbb{F}$. Definamos $\mathbb{H} = \{B \subset \mathbb{F} : f^{-1}(B) \in \mathcal{E}\}$, entonces $\mathbb{H}$ es una $\sigma-$algebra de subconjuntos de $\mathbb{F}$.

Este resultado se demuestra fácilmente gracias a que se tienen las siguientes propiedades:

a) $f^{-1}(\mathbb{F}) = \mathbb{E}$.

b) $f^{-1}(B^c) = [f^{-1}(B)]^c$ para cualquier subconjunto B de $\mathbb{F}$.

c) $f^{-1}\left(\cup_{\{\gamma\in\Gamma\}} B_\gamma\right) = \cup_{\{\gamma\in\Gamma\}} f^{-1}(B_\gamma)$ para cualquier familia $\{B_\gamma : \gamma \in \Gamma\}$ de subconjuntos de $\mathbb{F}$.

7. Si $(\mathbb{E}, \mathcal{E})$ y $(\mathbb{F}, \mathcal{F})$ son espacios medibles, una función $f : \mathbb{E} \mapsto \mathbb{F}$ es **medible** si $f^{-1}(B) \in \mathcal{E}$ para cualquier conjunto $B \in \mathcal{F}$.

8. Si $(\mathbb{E}, \mathcal{E})$ y $(\mathbb{F}, \mathcal{F})$ son espacios medible y la $\sigma-$álgebra $\mathcal{F}$ está generada por una familia $\mathcal{G}$ de subconjuntos de $\mathbb{F}$, entonces una función $f : \mathbb{E} \mapsto \mathbb{F}$ es medible si y sólo si $f^{-1}(B) \in \mathcal{E}$ para cualquier conjunto $B \in \mathcal{G}$.

Este resultado se sigue del inciso 6 ya que, siendo $\mathbb{H} = \{B \subset \mathbb{F} : f^{-1}(B) \in \mathcal{E}\}$ una $\sigma-$algebra de subconjuntos de $\mathbb{F}$ y dado que cualquier elemento de $\mathcal{G}$ pertenece a $\mathbb{H}$, entonces $\mathbb{H}$ contiene a la $\sigma-$álgebra generada por $\mathcal{G}$.

9. Una **función boreliana** es una función medible $f : \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$.

10. Una función medible $\varphi : (\mathbb{E}, \mathcal{E}) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es **simple** si tiene la forma $\varphi = \sum_{k=1}^{m} b_k I_{E_k}$, donde $b_1, \ldots, b_m$ son números reales y $E_1, \ldots, E_m$ son elementos de $\mathcal{E}$.

11. Si $(\mathbb{E}, \mathcal{E}) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una función medible simple, su **representación canónica** está dada por $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$, donde $a_1, a_2, \ldots, a_n$ son números reales no nulos distintos y, para $k \in \{1, \ldots, n\}$, $E_k = \{y \in \mathbb{E} : \varphi(y) = a_k\}$.

12. Si $f : (\mathbb{E}, \mathcal{E}) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una función medible no negativa, entonces existe una sucesión no decreciente de funciones medibles simples no negativas $\varphi_n : (\mathbb{E}, \mathcal{E}) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ tales que $\lim_{n\to\infty} \varphi_n(y) = f(y)$ para cualquier $y \in \mathbb{E}$.

13. Si $\mathbb{F}$ es un conjunto cualquiera y $\mathcal{A}$ es un álgebra de subconjuntos de $\mathbb{F}$, se dice que una función no negativa $\mu : \mathcal{A} \mapsto \overline{\mathbb{R}}$ es **finitamente aditiva** si dada cualquier familia finita, $A_1, \ldots, A_n$, de elementos de $\mathcal{A}$, ajenos por parejas, entonces:

$$\mu\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mu(A_k)$$

14. Si $\mathbb{F}$ es un conjunto cualquiera y $\Im$ es una σ-álgebra de subconjuntos de $\mathbb{F}$, se dice que una función no negativa $\mu : \Im \mapsto \overline{\mathbb{R}}$ es **σ-aditiva** si es finitamente aditiva y dada cualquier familia infinita numerable, $A_1, A_2, \ldots$, de elementos de $\Im$, ajenos por parejas, entonces:

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$$

15. Si $\mathbb{F}$ es un conjunto cualquiera y $\mathcal{A}$ es un álgebra de subconjuntos de $\mathbb{F}$, se dice que una función no negativa $\mu : \mathcal{A} \mapsto \overline{\mathbb{R}}$ es **σ-subaditiva**, o que satisface la propiedad de la subaditividad numerable, si dada cualquier colección infinita numerable $A_1, A_2, \ldots$ de elementos de $\mathcal{A}$ tales que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, entonces:

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

16. Si $\mathbb{F}$ es un conjunto cualquiera, $\mathcal{A}$ una álgebra de subconjuntos de $\mathbb{F}$ y $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ una función no negativa y finitamente aditiva, entonces las siguientes propiedades son equivalentes:

i. μ es σ-subaditiva.

ii. μ es σ-aditiva.

iii. Para cualquier sucesión creciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\mathcal{A}$ tales que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, se tiene $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$.

iv. Para cualquier sucesión decreciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\mathcal{A}$ tales que $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$ y $\mu(A_n) < \infty$ para alguna $n \in \mathbb{N}$, se tiene $\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$.

v. Para cualquier sucesión decreciente $(A_n)_{n\in\mathbb{N}}$, de elementos de $\mathcal{A}$ tales que $\bigcap_{n=1}^{\infty} A_n = \emptyset$ y $\mu(A_n) < \infty$ para alguna $n \in \mathbb{N}$,, se tiene $\lim_{n\to\infty} \mu(A_n) = 0$.

17. Si $\mathbb{F}$ es un conjunto cualquiera y $\Im$ es una σ-álgebra de subconjuntos de $\mathbb{F}$, una **medida** sobre $(\mathbb{F}, \Im)$ es una función no negativa $\mu : \Im \mapsto \overline{\mathbb{R}}$ σ-aditiva y tal que $\mu(\emptyset) = 0$.

18. Un **espacio de medida** es una terna $(\mathbb{F}, \Im, \mu)$, donde $\mathbb{F}$ es un conjunto, $\Im$ una σ-álgebra de subconjuntos de $\mathbb{F}$ y $\mu : \Im \mapsto \overline{\mathbb{R}}$ una medida.

19. Un espacio de medida $(\mathbb{F}, \Im, \mu)$ es **completo** si $\Im$ contiene a todos los subconjuntos de los conjuntos de medida μ igual a cero.

20. Si un espacio de medida $(\mathbb{F}, \Im_0, \mu)$ no es completo, se puede completar.

21. Si $\varphi\,(\mathbb{E}, \mathcal{E}, \mu) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una función medible simple no negativa con representación canónica $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$, se define la integral de φ con respecto a la medida μ, $\int_{\mathbb{E}} \varphi d\mu$, de la siguiente manera:

$$\int_{\mathbb{E}} \varphi d\mu = \sum_{k=1}^{n} a_k \mu(E_k)$$

22. Si $f : (\mathbb{E}, \mathcal{E}, \mu) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una función medible no negativa, se define la integral de f, con respecto a la medida μ, $\int_{\mathbb{E}} f d\mu$, de la siguiente manera:

$$\int_{\mathbb{E}} f d\mu = \sup\left\{ \int_{\mathbb{E}} \varphi d\mu : \varphi \text{ es medible simple y } 0 \le \varphi \le f \right\}$$

23. **Teorema de la convergencia monótona**. Si $(f_n)_{n\in\mathbb{N}}$ es una sucesión no decreciente de funciones medibles no negativas, entonces:

$$\int_{\mathbb{E}} \lim_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_{\mathbb{E}} f_n d\mu$$

24. Si $f : (\mathbb{E}, \mathcal{E}, \mu) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una función medible no negativa. Entonces, la función $m : \mathcal{E} \to \overline{\mathbb{R}}$, definida por $m(E) = \int_E f d\mu$, es una medida.

25. **Lema de Fatou**. Si $\{f_n\}_{n\in N}$ es una sucesión de funciones medibles no negativas, entonces:

$$\int_{\mathbb{E}} \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_{\mathbb{E}} f_n d\mu$$

26. Una función medible $f : (\mathbb{E}, \mathcal{E}, \mu) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es **integrable** sobre un conjunto $E \in \mathcal{E}$ si $\int_E |f|\, d\mu < \infty$.

27. Si f es una función medible e integrable sobre un conjunto $E \in \mathcal{E}$, se define su integral sobre E, $\int_E f d\mu$, de la siguiente manera:

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$$

28. **Teorema de la convergencia dominada**. Si $g : (\mathbb{E}, \mathcal{E}, \mu) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una función nedible no negativa, integrable sobre un conjunto medible E, $\{f_n\}_{n\in N}$ una sucesión de funciones medibles tales que $|f_n| \leq g$ y $\lim_{n\to\infty} f_n$ existe excepto a lo más en un conjunto de medida cero y $f : (\mathbb{E}, \mathcal{E}, \mu) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ una función medible tal que $f = \lim_{n\to\infty} f_n$, donde este límite existe, entonces, f es integrable sobre E y:

$$\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu$$

29. Un subconjunto A de $\mathbb{R}$ tiene **medida cero** si dado cualquier número $\varepsilon > 0$, existe una familia finita o infinita numerable de intervalos abiertos, cuya unión contiene al conjunto A, y tales que la suma de sus longitudes es menor que ε.

30. Existe una medida λ definida sobre la σ-álgebra $\mathfrak{L}(\mathbb{R})$ de subconjuntos de $\mathbb{R}$ generada por los intervalos y los conjuntos de medida cero, la cual asigna a cada intervalo su longitud. λ es llamada la **medida de Lebesgue** sobre $\mathbb{R}$ y a los elementos de $\mathfrak{L}(\mathbb{R})$ se les llama **conjuntos Lebesgue medibles**.

31. Un **espacio de probabilidad** es una terna $(\Omega, \Im, P)$, donde Ω es un conjunto, $\Im$ una σ-álgebra de subconjuntos de Ω y P una medida sobre $\Im$ tal que $P(\Omega) = 1$, a la cual llamaremos **medida de probabilidad**. A Ω lo llamaremos el **espacio muestral**, a los elementos de $\Im$ **eventos** y a la medida P de un evento A, la probabilidad de A.

32. Una **variable aleatoria real** es una función medible $X : (\Omega, \Im, P) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$.

33. Si $X : (\Omega, \Im, P) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una variable aleatoria real, la función $\mu_X : \mathfrak{B}\left(\overline{\mathbb{R}}\right) \mapsto \mathbb{R}$ definida por:

$$\mu_X(B) = P[X \in B]$$

es una medida de probabilidad.

34. Si $X : (\Omega, \Im, P) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una variable aleatoria tal que $P[X \in \mathbb{R}] = 1$, se define su **función de distribución** $F_X : \mathbb{R} \to \mathbb{R}$ mediante la relación $F_X(x) = P[X \leq x]$.

35. Si $X : (\Omega, \Im, P) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una variable aleatoria tal que $P[X \in \mathbb{R}] = 1$, su función de distribución F_X satisface las siguientes propiedades:

a) F_X es una función no decreciente y continua por la derecha.

b) $\lim_{x\to\infty} F_X(x) = 1$

c) $\lim_{x\to-\infty} F_X(x) = 0$

d) $F_X(x-) = P[X < x]$ para cualquier $x \in \mathbb{R}$.

Esperanza de variables aleatorias

Recordemos que, si $X : (\Omega, \Im, P) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una variable aleatoria discreta con valores en $\mathbb{R}$, su esperanza se define de la siguiente manera:

$$E[X] = \sum_{\{x\in\mathbb{R}:P[X=x]>0\}} xP[X = x]$$

Esto siempre y cuando la serie $\sum_{\{x\in\mathbb{R}:P[X=x]>0\}} |x|\, P[X = x]$ sea convergente.

Con esta definición, para fines prácticos, es decir, cuando el espacio de probabilidad modela un determinado experimento aleatorio, la esperanza de una variable aleatoria discreta se puede interpretar como un valor aproximado del promedio de los valores que toma cuando el experimento aleatorio se repite muchas veces.

Ahora, si $\varphi : (\Omega, \Im, P) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es una variable aleatoria simple no negativa, con representación canónica $\varphi = \sum_{k=1}^{n} a_k I_{E_k}$, donde $a_1, \ldots, a_n$ son números reales positivos distintos y $E_1, \ldots, E_m$ son elementos de $\Im$, ajenos por parejas, entonces se puede expresar de la forma siguiente:

$$\varphi = \sum_{k=1}^{n} a_k I_{[\varphi=a_k]}$$

Así que su integral, con respecto a la medida de probabilidad P, está dada por la siguiente expresión:

$$\int_\Omega \varphi dP = \sum_{k=1}^n a_k P\left[\varphi = a_k\right]$$

Vemos entonces que una función medible simple medible es una variable aleatoria discreta y se tiene la siguiente relación:

$$E\left[X\right] = \int_\Omega \varphi dP$$

Esta igualdad nos da la pauta para definir, de manera general, la esperanza de cualquier variable aleatoria real como su integral con respecto a la medida P, cuando esta última esté bien definida.

De esta forma, la esperanza de una variable aleatoria real no es más que otro nombre que le damos a la integral de esa variable aleatoria con respecto a P.

En lo que sigue, $(\Omega, \Im, P)$ **será un espacio de probabilidad completo**.

Definición 1. *Si X es una variable aleatoria no negativa, definimos la* ***esperanza*** *de X, $E\left[X\right]$, de la siguiente manera:*

$$E\left[X\right] = \int_\Omega X dP$$

Definición 2. *Diremos que la variable aleatoria X tiene esperanza finita si $E\left[|X|\right] < \infty$. En ese caso definimos su esperanza de la siguiente manera:*

$$E\left[X\right] = E\left[X^+\right] - E\left[X^-\right]$$

La Esperanza de una variable aleatoria tiene las mismas propiedades que la integral, es decir, se tiene la linealidad, el teorema de la convergencia monótona, el lema de Fatou, etcétera.

Recordemos que la función $\mu_X : \mathfrak{B}\left(\overline{\mathbb{R}}\right) \mapsto \mathbb{R}$ definida por:

$$\mu_X\left(B\right) = P\left[X \in B\right]$$

es una medida de probabilidad.

Así que $\left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right), \mu_X\right)$ es un espacio de probabilidad, el cual podemos completar, obteniendo una nueva $\sigma-$álgebra, la cual denotaremos por $\mathfrak{B}_X\left(\overline{\mathbb{R}}\right)$, y una medida definida sobre $\mathfrak{B}_X\left(\overline{\mathbb{R}}\right)$, para la cual no cambiaremos la notación; es decir, la denotaremos por μ_X.

Tenemos así un espacio de probabilidad completo $\left(\overline{\mathbb{R}}, \mathfrak{B}_X\left(\overline{\mathbb{R}}\right), \mu_X\right)$, de manera que podemos definir la integral de las funciones medibles $f : \left(\overline{\mathbb{R}}, \mathfrak{B}_X\left(\overline{\mathbb{R}}\right), \mu_X\right) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ como con cualquier otra medida de probabilidad y los resultados expuestos acerca de la integral son válidos también en este caso.

Para evitar confusiones, utilizaremos el término variable aleatoria únicamente para las funciones medibles $X : (\Omega, \Im, P) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$.

Teorema 1. *Si X es una variable aleatoria, entonces $E\left[f(X)\right] = \int_{\overline{\mathbb{R}}} f d\mu_X$ para cualquier función medible no negativa $f : \left(\overline{\mathbb{R}}, \mathfrak{B}_X\left(\overline{\mathbb{R}}\right)\right) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$.*

Demostración

Supongamos primero que f es una función medible simple no negativa; es decir, de la forma:

$f = \sum_{k=1}^{m} b_k I_{E_k}$, donde $b_1, \ldots, b_m$ son números reales no negativos y $E_1, \ldots, E_m$ son elementos de $\mathfrak{B}_X\left(\overline{\mathbb{R}}\right)$.

Entonces:

$E\left[f(X)\right] = E\left[\sum_{k=1}^{m} b_k I_{E_k}(X)\right] = E\left[\sum_{k=1}^{m} b_k I_{[X \in E_k]}\right]$

$= \sum_{k=1}^{m} b_k P\left[X \in E_k\right] = \sum_{k=1}^{m} b_k \mu_X\left(E_k\right)$

$= \int_{\overline{\mathbb{R}}} \left(\sum_{k=1}^{m} b_k I_{E_k}\right) d\mu_X = \int_{\overline{\mathbb{R}}} f d\mu_X$

Ahora, si $f : \left(\overline{\mathbb{R}}, \mathfrak{B}_X\left(\overline{\mathbb{R}}\right)\right) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ es cualquier función medible no negativa, sea $(\varphi_n)_{n\in\mathbb{N}}$ una sucesión no decreciente de funciones medibles simples no negativas tal que:

$$\text{lím}_{n\to\infty} \varphi_n(y) = f(y) \text{ para cualquier } y \in \overline{\mathbb{R}}$$

Por el teorema de la convergencia monótona, se tiene:

$E\left[f(X)\right] = \text{lím}_{n\to\infty} E\left[\varphi_n(X)\right] = \text{lím}_{n\to\infty} \int_{\overline{\mathbb{R}}} \varphi_n d\mu_X$

$= \int_{\overline{\mathbb{R}}} \left(\text{lím}_{n\to\infty} \varphi_n\right) d\mu_X = \int_{\overline{\mathbb{R}}} f d\mu_X$

■

Corolario 1. *Si X es una variable aleatoria, entonces $E\left[f(X)\right] = \int_{\overline{\mathbb{R}}} f d\mu_X$ para cualquier función integrable $f : \left(\overline{\mathbb{R}}, \mathfrak{B}_X\left(\overline{\mathbb{R}}\right), \mu_X\right) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$.*

Demostración

Por teorema anterior, se tiene:

$$E\left[f^+(X)\right] = \int_{\overline{\mathbb{R}}} f^+ d\mu_X$$

$$E\left[f^{-}(X)\right]=\int_{\overline{\mathbb{R}}} f^{-} d\mu_X$$

Así que:

$$E\left[f(X)\right]=E\left[f^{+}(X)\right]-E\left[f^{-}(X)\right]=\int_{\overline{\mathbb{R}}} f^{+} d\mu_X-\int_{\overline{\mathbb{R}}} f^{-} d\mu_X=\int_{\overline{\mathbb{R}}} f d\mu_X$$

■

Proposición 1. *Para cualquier variable aleatoria X, el conjunto $\{x \in \mathbb{R} : P[X=x]>0\}$ es a lo más infinito numerable.*

Demostración

Para cada $n \in \mathbb{N}$, definamos:

$$E_n=\left\{x \in \mathbb{R} : P[X=x]>\tfrac{1}{n}\right\}$$

E_n no puede tener más de n elementos ya que, de lo contrario, se tendría , $P[X \in E_n]>1$; así que E_n es un conjunto finito.

Además:

$$\{x \in \mathbb{R} : P[X=x]>0\}=\bigcup_{n=1}^{\infty} E_n$$

Así que el conjunto $\{x \in \mathbb{R} : P[X=x]>0\}$ es a lo más infinito numerable.

■

Recapitulando, hasta el momento, tenemos 4 tipos de integrales:

1. Tenemos el espacio de probabilidad $\left(\overline{\mathbb{R}}, \mathfrak{B}_X\left(\overline{\mathbb{R}}\right), \mu_X\right)$, así que vamos a tratar con integrales de la forma $\int_{\overline{\mathbb{R}}} f d\mu_X$, donde f es una función medible $f:\left(\overline{\mathbb{R}}, \mathfrak{B}_X\left(\overline{\mathbb{R}}\right), \mu_X\right) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$.

2. Tenemos un espacio de probabilidad $(\Omega, \Im, P)$, así que vamos a tratar con integrales de la forma $\int_\Omega X dP$, donde X es una variable aleatoria $X:(\Omega, \Im, P) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$.

3. Tenemos un espacio de medida $(\mathbb{R}, \mathfrak{L}(\mathbb{R}), \lambda)$, así que vamos a tratar con integrales de la forma $\int_{\mathbb{R}} f d\lambda$, donde f es una función medible $f:(\mathbb{R}, \mathfrak{L}(\mathbb{R}), \lambda) \mapsto \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$.

4. Si $f:[a,b] \to \mathbb{R}$ es una función Riemann integrable, denotaremos su integral como $(R)\int_a^b f(x)\,dx$.

Si $f:[0,\infty) \to [0,\infty)$ es una función acotada y Riemann integrable en todo intervalo acotado, definimos $(R)\int_0^\infty f(x)\,dx$ como el límite $\lim_{b\to\infty}(R)\int_0^b f(x)\,dx$.

Si $f : (-\infty, 0] \rightarrow [0, \infty)$ es una función acotada y Riemann integrable en todo intervalo acotado, definimos $(R) \int_{-\infty}^{0} f(x)\, dx$ como el límite $\lim_{a \to -\infty} (R) \int_{a}^{0} f(x)\, dx$.

Teorema 2. *Sean X y Y dos variables aleatorias independientes de esperanza finita, entonces XY también tiene esperanza finita y $E[XY] = E[X]\, E[Y]$.*

Demostración

Consideremos primero el caso de dos variables aleatorias simples no negativas independientes, φ y ψ, con representaciones canónicas $\varphi = \sum_{j=1}^{n} a_j I_{F_j}$ y $\psi = \sum_{k=1}^{m} b_k I_{G_k}$, respectivamente.

Como φ y ψ son independientes, cualquier evento F_j ($j \in \{1, 2, \ldots, n\}$) es independiente de cualquier evento G_k ($k \in \{1, 2, \ldots, m\}$).

Así que:

$E[\varphi\psi] = E\left[\left(\sum_{j=1}^{n} a_j I_{F_j}\right)\left(\sum_{k=1}^{m} b_k I_{G_k}\right)\right]$

$= E\left[\sum_{j=1}^{n} \sum_{k=1}^{m} a_j b_k I_{F_j} I_{G_k}\right] = E\left[\sum_{j=1}^{n} \sum_{k=1}^{m} a_j b_k I_{F_j \cap G_k}\right]$

$= \sum_{j=1}^{n} \sum_{k=1}^{m} a_j b_k P(F_j \cap G_k) = \sum_{j=1}^{n} \sum_{k=1}^{m} a_j b_k P(F_j)\, P(G_k)$

$= \left[\sum_{j=1}^{n} a_j P(F_j)\right]\left[\sum_{k=1}^{m} b_k P(G_k)\right] = E[\varphi]\, E[\psi]$

Ahora, para cada $n \in \mathbb{N}$, definamos las variables aleatorias simples no negativas, φ_n y ψ_n, de la siguiente manera:

$$\varphi_n(\omega) = \begin{cases} \sum_{m=1}^{n2^n} \frac{m-1}{2^n} I_{\left\{z \in \Omega : \frac{m-1}{2^n} \leq X(z) < \frac{m}{2^n}\right\}}(\omega) & \text{si } X(\omega) < n \\ n & \text{si } X(\omega) \geq n \end{cases}$$

$$\psi_n(\omega) = \begin{cases} \sum_{m=1}^{n2^n} \frac{m-1}{2^n} I_{\left\{z \in \not\leq : \frac{m-1}{2^n} \leq Y(z) < \frac{m}{2^n}\right\}}(\omega) & \text{si } Y(\omega) < n \\ n & \text{si } Y(\omega) \geq n \end{cases}$$

Como X y Y son independientes, φ_n y ψ_n, también lo son. Además, sabemos que las sucesiones $(\varphi_n)_{n \in \mathbb{N}}$ y $(\psi_n)_{n \in \mathbb{N}}$ son no decrecientes y, para cualquier $\omega \in \Omega$, $\lim_{n \to \infty} \varphi_n(\omega) = X(\omega)$ y $\lim_{n \to \infty} \psi_n(\omega) = Y(\omega)$.

Así, que, por el teorema de la convergencia monótona, se tiene:

$$E[XY] = E[\lim_{n \to \infty} \varphi_n \psi_n] = \lim_{n \to \infty} E[\varphi_n \psi_n] = \lim_{n \to \infty} E[\varphi_n]\, E[\psi_n] = E[X]\, E[Y]$$

Finalmente, si X y Y son dos variables aleatorias independientes cualesquiera, de esperanza finita, se tiene:

$$|XY| = |(X^+ - X^-)(Y^+ - Y^-)| \leq (X^+ + X^-)(Y^+ + Y^-)$$

Como X y Y son independientes y de esperanza finita, $X^+ + X^-$ y $Y^+ + Y^-$ también son independientes y tienen esperanza finita; así que:

$$E[|XY|] \leq E[(X^+ + X^-)(Y^+ + Y^-)] = E[X^+ + X^-]\, E[Y^+ + Y^-] < \infty$$

Por lo tanto, XY tiene esperanza finita y se tiene:

$E[XY] = E[(X^+ - X^-)(Y^+ - Y^-)] = E[X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+]$

$= E[X^+Y^+] + E[X^-Y^-] - E[X^+Y^-] - E[X^-Y^+]$

$= E[X^+]\, E[Y^+] + E[X^-]\, E[Y^-] - E[X^+]\, E[Y^-] - E[X^-]\, E[Y^+]$

$= (E[X^+] - E[X^-])(E[Y^+] - E[Y^-]) = E[X]\, E[Y]$

■

Un razonamiento de inducción permite demostrar el siguiente corolario:

Corolario 2. *Sean $X_1, \ldots, X_n$ n variables aleatorias independientes de esperanza finita, entonces $\prod_{k=1}^{n} X_k$ también tiene esperanza finita y:*

$$E\left[\prod_{k=1}^{n} X_k\right] = \prod_{k=1}^{n} E[X_k]$$

La esperanza como una integral de Riemann

En esta parte vamos a considerar únicamente el caso en que $P[X \in \mathbb{R}] = 1$ y utilizaremos los siguientes cuatro resultados:

a) Una función acotada $f : [a, b] \to \mathbb{R}$ es Riemann integrable si y sólo si el conjunto de puntos donde f es discontinua tiene medida cero.

b) Si $f : [a, b] \to \mathbb{R}$ y $g : [a, b] \to \mathbb{R}$ son dos funciones Riemann integrables y el conjunto $\{x \in [a, b] : f(x) \neq g(x)\}$ es finito o infinito numerable, entonces:

$$R \int_a^b f(x)\, dx = (R) \int_a^b g(x)\, dx$$

c) Si una función acotada $f : [a, b] \to \mathbb{R}$ es Riemann integrable, entonces f es Lebesgue medible e integrable; además:

$$\int_{[a,b]} f d\lambda = (R) \int_a^b f(x)\, dx$$

d) (**Teorema de Fubini**) Si $f : (\Omega \times \mathbb{R}, \Im \times \mathfrak{L}(\mathbb{R})) \to ([0, \infty), \mathfrak{B}([0, \infty)))$, es medible, donde $\Im \times \mathfrak{L}(\mathbb{R})$ es la $\sigma-$álgebra de subconjuntos de $\Omega \times \mathbb{R}$ generada por la familia:

$$\mathcal{A} = \{F \times E \subset \Omega \times \mathbb{R} : F \in \Im \text{ y } E \in \mathfrak{L}(\mathbb{R})\},$$

entonces:

$$\int_\Omega \left(\int_\mathbb{R} f d\lambda\right) dP = \int_\mathbb{R} \left(\int_\Omega f dP\right) d\lambda$$

Lema 1. *Si* $X : (\Omega, \Im) \to \left(\overline{\mathbb{R}}, \mathfrak{B}\left(\overline{\mathbb{R}}\right)\right)$ *es una variable aleatoria no negativa, entonces la función:*

$$I_{[0,X)} : (\Omega \times \mathbb{R}, \Im \times \mathfrak{L}(\mathbb{R})) \to ([0, \infty), \mathfrak{B}([0, \infty)))$$

es medible, donde $\Im \times \mathfrak{L}(\mathbb{R})$ es la $\sigma-$álgebra de subconjuntos de $\Omega \times \mathbb{R}$ generada por la familia $\mathcal{A} = \{F \times E \subset \Omega \times \mathbb{R} : F \in \Im \text{ y } E \in \mathfrak{L}(\mathbb{R})\}$.

Demostración

Consideremos primero una variable aleatoria simple no negativa φ, con representación canónica $\varphi = \sum_{k=1}^m a_k I_{F_k}$. En este caso se tiene:

$$I_{[0,\varphi)} = \sum_{k=1}^m I_{[0,a_k)} I_{F_k} = \sum_{k=1}^m I_{F_k \times [0,a_k)}$$

Así que $I_{[0,\varphi)}$ es medible.

Sea ahora una sucesión $(\varphi_n)_{n\in\mathbb{N}}$ no decreciente de variables aleatorias simples no negativas tal que:

$$\text{lím}_{n\to\infty} \varphi_n(\omega) = X(\omega) \text{ para cualquier } \omega \in \Omega$$

Entonces la sucesión $\left(I_{[0,\varphi_n)}\right)_{n\in\mathbb{N}}$ es no decreciente y se tiene:

$$\text{lím}_{n\to\infty} I_{[0,\varphi_n(\omega))}(y) = I_{[0,X(\omega))}(y) \text{ para cualquier pareja } (\omega, y) \in \Omega \times \mathbb{R}$$

Por lo tanto, $I_{[0,X)}$ es medible. ■

Proposición 2. *Sea* X *una variable aleatoria no negativa, entonces:*

$$E[X] = \int_{[0,\infty)} (1 - F_X)\, d\lambda$$

Demostración

$E[X] = \int_{\Omega} \left(\int_{\mathbb{R}} I_{[0,X)} d\lambda \right) dP = \int_{\mathbb{R}} \left(\int_{\Omega} I_{[0,X)} dP \right) d\lambda$

Obsérvese que la función $I_{[0,X)}$ es una función en dos variables:

$$(\omega, x) \to I_{[0,X(\omega))}(x)$$

Esta función toma el valor 1 si y sólo si $x \in [0, X(\omega))$; es decir, si y sólo si $0 \leq x < X(\omega)$.

Para $x \in [0, \infty)$ fija, $I_{[0,X(\omega))}(x) = I_{(x,\infty)}(X(\omega))$ para cualquier $\omega \in \Omega$; así que:

$$\int_{\Omega} I_{[0,X)}(x) \, dP = \int_{\Omega} I_{(x,\infty)}(X) \, dP = P[X \in (x, \infty)] = P[X > x] = 1 - F_X(x)$$

Como X es no negativa, para $x \in (-\infty, 0)$ fija, $I_{[0,X(\omega))}(x) = 0$ para cualquier $\omega \in \Omega$; así que:

$$\int_{\Omega} I_{[0,X)}(x) \, dP = 0$$

Por lo tanto:

$E[X] = \int_{\mathbb{R}} \left(\int_{\Omega} I_{[0,X)} dP \right) d\lambda = \int_{[0,\infty)} \left(\int_{\Omega} I_{[0,X)} dP \right) d\lambda + \int_{(-\infty,0)} \left(\int_{\Omega} I_{[0,X)} dP \right) d\lambda$

$= \int_{[0,\infty)} \left(\int_{\Omega} I_{[0,X)} dP \right) d\lambda = \int_{[0,\infty)} (1 - F_X) \, d\lambda$

■

Como la función F_X está acotada y el conjunto de sus discontinuidades es a lo más infinito numerable, tanto F_X como $1 - F_X$ son Riemann integrables en cualquier intervalo finito; así que también podemos escribir el resultado anterior en la siguiente forma:

$$E[X] = (R) \int_0^{\infty} [1 - F_X(x)] \, dx$$

Corolario 3. *$E[X^+] = (R) \int_0^{\infty} P[X > x] \, dx$ y $E[X^-] = (R) \int_{-\infty}^{0} P[X < x] \, dx$.*

Demostración

$E[X^+] = (R) \int_0^{\infty} [1 - F_{X^+}(x)] \, dx = (R) \int_0^{\infty} P[X^+ > x] \, dx$

$= (R) \int_0^{\infty} P[X > x] \, dx$

$E[X^-] = (R) \int_0^{\infty} [1 - F_{X^-}(x)] \, dx = (R) \int_0^{\infty} P[X^- > x] \, dx$

$= (R) \int_0^{\infty} P[X < -x] \, dx = (R) \int_{-\infty}^{0} P[X < x] \, dx$

■

Corolario 4. *Una variable aleatoria X tiene esperanza finita si y sólo si las integrales $(R)\int_0^\infty P[X > x]\,dx$ y $(R)\int_{-\infty}^0 P[X < x]\,dx$ convergen.*

Corolario 5. *Si X es una variable aleatoria de esperanza finita, entonces:*

$$E[X] = (R)\int_0^\infty P[X > x]\,dx - (R)\int_{-\infty}^0 P[X < x]\,dx$$

Obsérvese que, como $F_X(x-) = P[X < x]$ y el conjunto de discontinuidades de F_X es a lo más infinito numerable, $P[X < x] = F_X(x)$ excepto a lo más en un conjunto numerable; así que $(R)\int_{-\infty}^0 P[X < x]\,dx = (R)\int_{-\infty}^0 F_X(x)dx$. Así que entonces se tiene el siguiente resultado:

Teorema 3. *Sea X una variable aleatoria, entonces X tiene esperanza finita si y sólo si $(R)\int_0^\infty [1 - F_X(x)]\,dx < \infty$ y $(R)\int_{-\infty}^0 F_X(x)dx < \infty$ y, en ese caso, se tiene:*

$$E[X] = (R)\int_0^\infty [1 - F_X(x)]\,dx - (R)\int_{-\infty}^0 F_X(x)dx$$

Cuando se tiene $\int_0^\infty P[X > x]\,dx = \infty$ y $\int_{-\infty}^0 P[X < x]\,dx < \infty$, se define $E[X] = \infty$, mientras que cuando $\int_0^\infty P[X > x]\,dx < \infty$ y $\int_{-\infty}^0 P[X < x]\,dx = \infty$, se define $E[X] = -\infty$. Cuando ambas integrales sean divergentes, entonces la esperanza de X no está definida.

Varianza y covarianza

Definición 3. *Sea X una variable aleatoria de esperanza finita. Se define la **varianza** de X, $Var(X)$, mediante la relación:*

$$Var(X) = E\left[(X - E(X))^2\right]$$

Definición 4. *A la raíz cuadrada no negativa de la varianza se le llama la **desviación estándar** de X.*

La varianza de una variable aleatoria mide entonces el alejamiento de los valores de X de su esperanza. También se acostumbra decir que la varianza es una medida de la dispersión de los valores de la variable aleatoria.

Definición 5. *Diremos que una variable aleatoria X tiene varianza finita si se cumplen las siguientes dos condiciones:*

1. *X tiene esperanza finita.*

2. $(X - E[X])^2$ *tiene esperanza finita.*

Proposición 3. *Una variable aleatoria X tiene varianza finita si y sólo si X^2 tiene esperanza finita.*

Demostración

Se tiene $X^2 = (X - E[X])^2 + 2XE[X] - (E[X])^2$, así que si X tiene varianza finita, entonces X^2 tiene esperanza finita.

Supongamos ahora que X^2 tiene esperanza finita.

Se tiene $|X| \leq 1 + X^2$ y $(X - E[X])^2 = X^2 - 2XE[X] + (E[X])^2$. De manera que tanto X como $(X - E[X])^2$ tienen esperanza finita. Es decir, X tiene varianza finita.

Proposición 4. *Sea X una variable aleatoria de esperanza finita, entonces:*

$$Var(X) = E[X^2] - (E[X])^2$$

Demostración

Si X no tiene varianza finita entonces X^2 tampoco tiene esperanza finita, así que se cumple la igualdad. Si X tiene varianza finita, entonces X^2 también tiene esperanza finita y se tiene:

$Var(X) = E\left[(X - E(X))^2\right] = E\left[X^2 - 2XE(X) + (E[X])^2\right]$

$= E[X^2] - 2(E[X])^2 + (E[X])^2$

$= E[X^2] - (E[X])^2$

∎

Proposición 5. *Sean X y Y dos variables aleatorias de varianza finita. Entonces, XY tiene esperanza finita.*

Demostración

Para cualquier par de números reales x y y, se tiene $|xy| \leq \frac{1}{2}(x^2 + y^2)$. Así que, $|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2$.

Por lo tanto, XY tiene esperanza finita.

∎

Corolario 6. *Si X y Y son dos variables aleatorias de varianza finita y a y b son dos números reales cualesquiera, entonces $aX + bY$ tiene varianza finita.*

Demostración

$(aX + bY)^2 = a^2X^2 + 2abXY + b^2Y^2$, así que, por las proposiciones 5 y 3, $aX + bY$ tiene varianza finita.

■

Definición 6. *Sean X y Y dos variables aleatorias de varianza finita. Se define la* ***covarianza*** *de X y Y, $Cov(X, Y)$, mediante la relación:*

$$Cov(X, Y) = E\left[(X - E[X])(Y - E[Y])\right] = E[XY] - E[X]E[Y]$$

Proposición 6. *Sean X y Y dos variables aleatorias independientes de varianza finita, entonces $Cov(X, Y) = 0$.*

Demostración

El resultado es inmediato pues X y Y son independientes y tienen esperanza finita, así que $E[XY] = E[X]E[Y]$.

■

El siguiente ejemplo muestra que la covarianza entre dos variables aleatorias puede ser cero sin que éstas sean independientes.

Ejemplo 1. *Sea X una variable aleatoria discreta con función de densidad dada por:*

$$f_X(x) = \begin{cases} \frac{1}{4} & \text{si } x \in \{-1, 1\} \\ \frac{1}{2} & \text{si } x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

y sea Z una variable aleatoria, independiente de X, con distribución uniforme en el conjunto $\{-1, 1\}$. Definamos la variable aleatoria Y de la siguiente manera:

$$Y = \begin{cases} Z & \text{si } X = 0 \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene entonces:

$f_Y(y) = P[Y = y] = P[Y = y, X = 0] + P[Y = y, X \neq 0]$

$= P[Z = y, X = 0] + P[Y = y, X \neq 0]$

$= P[Z = y]P[X = 0] + I_{\{0\}}(y)P[X \neq 0]$

$$= \frac{1}{4}I_{\{-1,1\}}(y) + \frac{1}{2}I_{\{0\}}(y) = \begin{cases} \frac{1}{4} & \text{si } y \in \{-1,1\} \\ \frac{1}{2} & \text{si } y = 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$P[X = x, Y = y]$$

$$= P[X = x, Y = y, X = 0] + P[X = x, Y = y, X \neq 0]$$

$$= I_{\{0\}}(x)P[Z = y, X = 0] + I_{\{0\}}(y)I_{\{-1,1\}}(x)P[X = x]$$

$$= I_{\{0\}}(x)P[Z = y]P[X = 0] + I_{\{0\}}(y)I_{\{-1,1\}}(x)P[X = x]$$

$$= \frac{1}{4}I_{\{0\}}(x)I_{\{-1,1\}}(y) + \frac{1}{4}I_{\{0\}}(y)I_{\{-1,1\}}(x)$$

$$= \begin{cases} \frac{1}{4} & \text{si } x = 0, y \in \{-1,1\} \text{ ó } y = 0, x \in \{-1,1\} \\ 0 & \text{en otro caso} \end{cases}$$

Así que, por ejemplo, $P[X = 1, Y = 1] \neq P[X = 1]P[Y = 1]$, de manera que X y Y no son independientes. Por otro lado, $E[X] = E[Y] = E[XY] = 0$, de manera que $Cov(X, Y) = 0$.

Proposición 7. *Sean X y Y dos variables aleatorias de varianza finita y a y b dos números reales cualesquiera. Entonces $Cov(aX, bY) = abCov(X, Y)$.*

Demostración

$$Cov(aX, bY) = E[aXbY] - E[aX]E[bY] = ab(E[XY] - E[X]E[Y])$$

$$= abCov(X, Y)$$

■

Teorema 4. *Sean $X, X_1, \ldots, X_n$ $n+1$ variables aleatorias de esperanza finita. Entonces:*

1. $Var(X) = 0$ si y sólo si existe una constante c tal que $P[X = c] = 1$.

2. $Var(aX + b) = a^2Var(X)$ para cualquier pareja $a, b \in \mathbb{R}$.

3. Si $X_1, \ldots, X_n$ tienen varianza finita, entonces $\sum_{i=1}^{n} X_i$ también tiene varianza finita y:

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{\{i,j\in\{1,\ldots,n\}:i<j\}} Cov(X_i, X_j)$$

Demostración

1. $Var(X) = 0$ si y sólo si $E\left[(X - E(X))^2\right] = 0$, lo cual ocurre si y sólo si $P\left[(X - E(X))^2 = 0\right] = 1$, es decir, $P[X = E(X)] = 1$.

2. $Var(aX+b) = E\left[(aX - aE[X])^2\right] = a^2 E\left[(X - E[X])^2\right] = a^2 Var(X)$.

3. Que $\sum_{i=1}^{n} X_i$ tiene varianza finita se sigue del corolario 6 y un razonamiento de inducción. Además:

$Var(\sum_{i=1}^{n} X_i) = E\left[(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} E[X_i])^2\right] = E\left[(\sum_{i=1}^{n} (X_i - E[X_i]))^2\right]$

$= \sum_{i=1}^{n} E\left[(X_i - E[X_i])^2\right] + 2\sum_{\{i,j\in\{1,\ldots,n\}:i<j\}} E\left[(X_i - E[X_i])(X_j - E[X_j])\right]$

$= \sum_{i=1}^{n} Var(X_i) + 2\sum_{\{i,j\in\{1,\ldots,n\}:i<j\}} Cov(X_i, X_j)$

■

Corolario 7. *Sean $X_1, \ldots, X_n$ n variables aleatorias independientes y de varianza finita, entonces $\sum_{i=1}^{n} X_i$ también tiene varianza finita y:*

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$

Teorema 5 (Desigualdad de Cauchy-Schwarz). *Sean X y Y dos variables aleatorias cualesquiera, entonces:*

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

Además, si X y Y tienen varianza finita, entonces $|E[XY]| = \sqrt{E[X^2]}\sqrt{E[Y^2]}$ si y sólo si existen constantes a y b tales que por lo menos una de ellas es distinta de cero y $P[aX + bY = 0] = 1$.

Demostración

Si $E[X^2] = \infty$ o $E[Y^2] = \infty$ la desigualdad es obvia.

Supongamos ahora que $E[X^2] < \infty$ y $E[Y^2] < \infty$, es decir, que tanto X como Y tienen varianza finita.

Sea $\alpha = (E[Y^2])^{\frac{1}{2}}$ y $\beta = (E[X^2])^{\frac{1}{2}}$.

Si $\alpha = 0$, se tiene $E[X^2] = 0$, de manera que:

$$P[|XY| = 0] \geq P[X = 0] = P[X^2 = 0] = 1$$

Por lo tanto, $E[|XY|] = 0$. Así que se cumple la desigualdad.

De la misma manera, si $\beta = 0$, entonces $E[|XY|] = 0$. Así que se cumple la desigualdad.

Supongamos ahora que $\alpha > 0$ y $\beta > 0$.

Sabemos que $\alpha|X| - \beta|Y|$ tiene varianza finita y se tiene:

$0 \leq E\left[(\alpha |X| - \beta |Y|)^2\right] = \alpha^2 E[X^2] + \beta^2 E[Y^2] - 2\alpha\beta E[|XY|] = 2\alpha^2\beta^2 - 2\alpha\beta E[|XY|]$.

Así que, $\alpha\beta - E[|XY|] \geq 0$. Es decir, $E[|XY|] \leq \alpha\beta$.

Para la segunda parte, supongamos primero que X y Y tienen varianza finita y que $|E[XY]| = \sqrt{E[X^2]}\sqrt{E[Y^2]}$.

Definiendo, como antes, $\alpha = (E[Y^2])^{\frac{1}{2}}$ y $\beta = (E[X^2])^{\frac{1}{2}}$, se tiene:

Si $\alpha = 0$ y $\beta = 0$, entonces $P[X = 0] = P[Y = 0] = 1$. Por lo tanto $P[X = 0, Y = 0] = 1$. De manera que, tomando en consideración que $P[X = 0, Y = 0] \leq P[X + Y = 0]$, se tiene $P[X + Y = 0] = 1$. Es decir, se tiene el resultado deseado con $a = b = 1$.

Si $\alpha \neq 0$ ó $\beta \neq 0$ se tienen los siguientes dos casos:

Si $E[XY] > 0$, entonces:

$$0 \leq E\left[(\alpha X - \beta Y)^2\right] = 2\alpha^2\beta^2 - 2\alpha\beta E[XY] = 0$$

Así que, $E\left[(\alpha X - \beta Y)^2\right] = 0$, de lo cual se sigue $P[\alpha X - \beta Y = 0] = 1$.

Es decir, se tiene el resultado deseado con $a = \alpha$ y $b = -\beta$.

Si $E[XY] < 0$, entonces:

$$0 \leq E\left[(\alpha X + \beta Y)^2\right] = 2\alpha^2\beta^2 + 2\alpha\beta E[XY] = 0$$

Así que, $E\left[(\alpha X + \beta Y)^2\right] = 0$, de lo cual se sigue $P[\alpha X + \beta Y = 0] = 1$.

Es decir, se tiene el resultado deseado con $a = \alpha$ y $b = \beta$.

Finalmente, supongamos que existen constantes a y b tales que por lo menos una de ellas es distinta de cero y $P[aX + bY = 0] = 1$. Supongamos, por ejemplo, que $a \neq 0$, entonces $P\left[X = -\frac{b}{a}Y\right] = 1$. Así que:

$$(E[XY])^2 = \tfrac{b^2}{a^2}\left(E[Y^2]\right)^2 = E\left[\left(-\tfrac{b}{a}Y\right)^2\right]E[Y^2] = E[X^2]\,E[Y^2]$$

$\blacksquare$

Corolario 8. *Sean X y Y dos variables aleatorias de varianza finita. Entonces:*

$$|Cov(X,Y)| \leq \sqrt{Var(X)}\sqrt{Var(Y)}$$

Además, la igualdad se cumple si y sólo si existen constantes a, b y c tales que a y b no son ambas cero y $P[aX + bY = c] = 1$.

Demostración

Utilizando la proposición 5, se tiene:

$|Cov(X,Y)| = |E[(X - E[X])(Y - E[Y])]| \leq E[|X - E[X]||Y - E[Y]|]$

$\leq \sqrt{E\left[(X - E[X])^2\right]}\sqrt{E\left[(Y - E[Y])^2\right]} = \sqrt{Var(X)}\sqrt{Var(Y)}$

Si la igualdad se cumple, entonces se tiene:

$$|E[(X - E[X])(Y - E[Y])]| = \sqrt{E\left[(X - E[X])^2\right]}\sqrt{E\left[(Y - E[Y])^2\right]}$$

De manera que, nuevamente por la proposición 5, existen constantes a y b tales que no son ambas cero y $P[a(X - E[X]) + b(Y - E[Y]) = 0] = 1$. Es decir:

$$P[aX + bY = c] = 1$$

donde $c = aE[X] + bE[Y]$.

Supongamos ahora que existen constantes a, b y c tales que a y b no son ambas cero y $P[aX + bY = c] = 1$. Entonces $E[aX + bY - c] = 0$, de lo cual se sigue $c = E[aX + bY]$. De manera que se tiene:

$$P[a(X - E[X]) + b(Y - E[Y]) = 0] = 1$$

Así que, por la proposición 5, se tiene:

$$|Cov(X,Y)| = |E[(X - E[X])(Y - E[Y])]| = \sqrt{Var(X)}\sqrt{Var(Y)}$$

■