Probabilidad II
Tema 3: Esperanza

José R. Berrendero

Departamento de Matemáticas
Universidad Autónoma de Madrid

Estructura de este tema

- Esperanza de una variable aleatoria. Principales propiedades.
- Teorema del cambio de espacio de integración.
- Momentos. Desigualdades.
- Esperanza del producto de v.a. independientes.
- Covarianza y correlación.

Esperanza de una variable aleatoria

La **esperanza** (o media o valor esperado) de una v.a. X se define de la siguiente forma:

$$\mathrm{E}(X) := \int_{\Omega} X \, d\mathrm{P}.$$

Obs. La $\mathrm{E}X$ es la integral sobre Ω de la función medible X respecto de la medida de probabilidad P tal y como se estudia en *Teoría de la medida y la integración*.

Obs. Esta definición de esperanza es consistente con la definición que se suele dar en cursos menos avanzados de probabilidad.

Construcción de la esperanza:

1. Caso de v.a. simples.
2. Extensión a v.a. no negativas.
3. Caso general.

Construcción de la esperanza

(1) **v.a. simples**: sea $X = \sum_{i=1}^{k} x_i I_{A_i}$ una v.a. simple. Entonces,

$$\mathrm{E}(X) := \sum_{i=1}^{k} x_i \mathrm{P}(A_i).$$

(2) **v.a. no negativas**: si $X \geq 0$, existe una sucesión de v.a. simples X_n tales que $0 \leq X_n \uparrow X$. Entonces,

$$\mathrm{E}(X) := \lim_{n \to \infty} \mathrm{E}(X_n).$$

(3) **v.a. arbitraria**: en el caso general, $X = X^+ - X^-$, donde $X^+ = \max\{X, 0\}$ y $X^- = \max\{0, -X\}$. Entonces,

$$\mathrm{E}(X) := \mathrm{E}(X^+) - \mathrm{E}(X^-),$$

siempre que $\mathrm{E}(X^+) < \infty$ o $\mathrm{E}(X^-) < \infty$.

Propiedades más importantes de la esperanza

(1) $E(X)$ es finita $\Leftrightarrow E|X| < \infty$. En este caso se dice que X es integrable. Se usa la notación:

$$X \in L_1(P) = L_1 := \{X : \Omega \to \mathbb{R} \text{ v.a., tales que } E|X| < \infty\}.$$

(2) **Linealidad:** Si $X + Y$ tiene esperanza, entonces $E(X + Y) = E(X) + E(Y)$. Si X tiene esperanza, entonces también la tiene cX para todo $c \in \mathbb{R}$, y $E(cX) = cE(X)$.

(3) **Monotonía:** Si $X \geq 0$, entonces $E(X) \geq 0$. Si $X, Y \in L_1$, y $X \leq Y$, entonces $E(X) \leq E(Y)$. En particular, $|E(X)| \leq E(|X|)$.

Propiedades más importantes de la esperanza

(4) **Teorema de la convergencia monótona (TCM)**:

(a) Si $X_n \uparrow X$ y $X_k^- \in L_1$ para algún k. Entonces $X_k, X_{k+1} \ldots$ tienen esperanza y $\mathrm{E}(X_n) \uparrow \mathrm{E}(X)$.

(b) Si $X_n \downarrow X$ y $X_k^+ \in L_1$ para algún k. Entonces $X_k, X_{k+1} \ldots$ tienen esperanza y $\mathrm{E}(X_n) \downarrow \mathrm{E}(X)$.

(5) **Lema de Fatou**:

(a) Si $X_n \leq Y$ para todo n y $\mathrm{E}Y^+ < \infty$, entonces X_n, $\limsup_{n\to\infty} X_n$ tienen esperanza para todo n y

$$\mathrm{E}\left(\limsup_{n\to\infty} X_n\right) \geq \limsup_{n\to\infty} \mathrm{E}X_n.$$

(b) Si $X_n \geq Z$ para todo n y $\mathrm{E}Z^- < \infty$, entonces X_n, $\liminf_{n\to\infty} X_n$ tienen esperanza para todo n y

$$\mathrm{E}\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} \mathrm{E}X_n.$$

Propiedades más importantes de la esperanza

(6) **Teorema de la convergencia dominada (TCD)**: Si $X_n \to X$ y existe $Z \in L_1$ tal que $|X_n| \leq Z$ para todo n, entonces $X_1, \ldots, X_n, \ldots, X$ son integrables y $\mathrm{E}X = \lim_{n\to\infty} \mathrm{E}X_n$.

(7) **Integral sobre un conjunto**: Sea X una v.a. positiva o integrable y $A \in \mathcal{F}$. Se define

$$\int_A X \, d\mathrm{P} = \mathrm{E}(X \cdot I_A) = \int_\Omega X \cdot I_A \, d\mathrm{P}.$$

Una desigualdad de mucha aplicación

Sea X una variable aleatoria y $a, b \in \mathbb{R}$. Se tiene:

$$a\,\mathrm{P}(a \leq X \leq b) \leq \int_{\{a \leq X \leq b\}} X \, dP \leq b\,\mathrm{P}(a \leq X \leq b).$$

Teorema del cambio de espacio de integración

Teorema: Sea X v.a. sobre $(\Omega, \mathcal{F}, \mathrm{P})$ con distribución F_X. Sea $g : \mathbb{R} \to \mathbb{R}$ medible Borel. Si $Y = g(X)$, entonces

$$\mathrm{E}(Y) = \int_{\mathbb{R}} g(x) dF_X(x) = \int_{\mathbb{R}} g d\mathrm{P}_X.$$

En la demostración se usa el **método de escala ascendente**, es decir, se demuestra sucesivamente para

(1) Indicadores, $g = I_B$.

(2) Funciones simples no negativas, $g = \sum_{i=1}^{n} x_i I_{B_i}$. La linealidad es la que se aplica aquí.

(3) Funciones no negativas, $g \geq 0$. Aquí se suele aplicar TCM.

(4) Funciones borelianas arbitrarias g. Usando $g = g^+ - g^-$.

Consecuencias y extensiones

(1) Si X es un vector aleatorio en $\mathbb{R}^n$, la misma demostración permite afirmar $\mathrm{E}[g(X)] = \int_{\mathbb{R}^n} g(x)dF_X(x)$.

(2) Si X es un vector aleatorio absolutamente continuo con función de densidad f, entonces

$$\mathrm{E}[g(X)] = \int_{\mathbb{R}^n} g(x)dF_X(x) = \int_{\mathbb{R}^n} g(x)f(x)dx.$$

(3) Si X es un vector aleatorio discreto con función de probabilidad $p(x)$, entonces

$$\mathrm{E}[g(X)] = \int_{\mathbb{R}^n} g(x)dF_X(x) = \sum_{x} g(x)p(x).$$

Ejemplos

Expresa en función de las densidades las siguientes esperanzas.

Sea X v.a. con densidad $f(x)$.

$\mathrm{E}X^3 =$

$\mathrm{E}e^{tX} =$

Sea (X, Y) v.a. con densidad $f(x, y)$.

$\mathrm{E}\left(\frac{X^2}{X^4+Y^4}\right) =$

Sea X, Y v.a. independientes con densidades $f_1(x)$ y $f_2(y)$.

$\mathrm{E}e^{X+Y} =$

$\mathrm{E}\sin(X + Y) =$

Momentos

Sea $k > 0$ y X una v.a. sobre $(\Omega, \mathcal{F}, \mathrm{P})$.

- Momento de orden k de X: $\mathrm{E}(X^k)$.
- Momento absoluto de orden k de X: $\mathrm{E}|X|^k$.
- Momento central de orden k de X: $\mathrm{E}[(X - E(X))^k]$.
- Momento absoluto central de orden k de X: $\mathrm{E}|X - \mathrm{E}(X)|^k$.
- La varianza de X es su momento central de orden 2:

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{E}[(X - E(X))^2].$$

- A la raíz de la varianza σ se le llama desviación típica.

Observaciones

- Si $\mathrm{E}(X^k)$ es finita para $k > 0$, entonces $\mathrm{E}(X^\alpha)$ es finita para $0 < \alpha < k$.

- Para todo $k > 0$ y $m \in \mathbb{R}$, $\mathrm{E}|X|^k < \infty \Leftrightarrow \mathrm{E}|X - m|^k < \infty$.

- Como consecuencia, si $X \in L_1$, se verifica

$$\mathrm{Var}(X) < \infty \Leftrightarrow X^2 \in L_1 \Leftrightarrow X \in L_2,$$

  donde $L_k := \{X : \Omega \to \mathbb{R} \text{ v.a., tales que } \mathrm{E}|X|^k < \infty\}$.

- Si $X \in L_2$, $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu^2$, donde $\mu = \mathrm{E}(X)$.

Algunas desigualdades importantes

Markov

Sea $X \geq 0$ y $\epsilon > 0$, entonces

$$\mathrm{P}(X \geq \epsilon) \leq \frac{\mathrm{E}(X)}{\epsilon}.$$

Chebychev

Sea $X \in L_1$ y $\epsilon > 0$, entonces

$$\mathrm{P}(|X - \mathrm{E}(X)| \geq \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}.$$

Jensen

Sea $f : \mathbb{R} \to \mathbb{R}$ convexa, $X \in L_1$, $f(X) \in L_1$. Entonces:

$$f[E(X)] \leq \mathrm{E}[f(X)]$$

Algunas desigualdades importantes

Lyapunov

Sea $0 < s < t$, entonces

$$(\mathrm{E}|X|^s)^{1/s} \leq (\mathrm{E}|X|^t)^{1/t}.$$

Hölder

Sea $1 < p < \infty$ y $1 < q < \infty$, tales que $1/p + 1/q = 1$. Si $\mathrm{E}|X|^p < \infty$, $\mathrm{E}|Y|^q < \infty$, se tiene $\mathrm{E}|XY| < \infty$ y

$$\mathrm{E}|XY| \leq (\mathrm{E}|X|^p)^{1/p}(\mathrm{E}|Y|^q)^{1/q}.$$

Minkowski

Sea $1 \leq p < \infty$, $\mathrm{E}|X|^p < \infty$, $\mathrm{E}|Y|^p < \infty$. Entonces,

$$(\mathrm{E}|X+Y|^p)^{1/p} \leq (\mathrm{E}|X|^p)^{1/p} + (\mathrm{E}|Y|^p)^{1/p}.$$

Esperanza del producto de v.a. independientes

Teorema
Sean $X_1, \ldots, X_n$ v.a. independientes sobre $(\Omega, \mathcal{F}, \mathrm{P})$. Si $X_i \geq 0$ para todo $i = 1, \ldots, n$, o si $X_i \in L_1$ para todo $i = 1, \ldots, n$, $\mathrm{E}(X_1 \cdots X_n)$ existe y

$$\mathrm{E}(X_1 \cdots X_n) = \mathrm{E}(X_1) \cdots \mathrm{E}(X_n).$$

Desigualdad de Hoeffding
Sean $X_1, \ldots, X_n$ v.a. independientes sobre $(\Omega, \mathcal{F}, \mathrm{P})$ tales que $\mathrm{P}(a_i \leq X_i \leq b_i) = 1$, para todo $i = 1, \ldots, n$. Sea $S_n := X_1 + \ldots + X_n$. Entonces, para todo $\epsilon > 0$,

$$\mathrm{P}(|S_n - \mathrm{E}(S_n)| > \epsilon) \leq 2 \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Covarianza

Sean $X \in L_2$, $Y \in L_2$. La **covarianza** entre X e Y se define como:

$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - EX)(Y - EY)] = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$$

Propiedades básicas

(1) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$; $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

(2) Para $a, b, c, d \in \mathbb{R}$, se tiene $\mathrm{Cov}(aX + c, bY + d) = ab\mathrm{Cov}(X, Y)$.

(3) **Desigualdad de Cauchy-Schwarz**: si consideramos la relación de equivalencia $X \sim X' \Leftrightarrow \mathrm{P}(X = X') = 1$, $\langle X, Y \rangle_2 := \mathrm{E}(XY)$ es un producto escalar. Por lo tanto:

$$|\mathrm{Cov}(X, Y)| \leq \sigma_X \cdot \sigma_Y.$$

Coeficiente de correlación

El coeficiente de **correlación** entre $X, Y \in L_2$ se define como:

$$\rho(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}.$$

Propiedades básicas

(1) Para $a, b, c, d \in \mathbb{R}$, se tiene $\rho(aX + c, bY + d) = \mathrm{sgn}(ab)\rho(X, Y)$.

(2) **Desigualdad de Cauchy-Schwarz**: $|\rho(X, Y)| \leq 1$. ¿Cuándo hay igualdad?

(3) Si X e Y son v.a. independientes, entonces $\rho(X, Y) = 0$ (variables incorreladas). El recíproco no es cierto.

(4) Sean $X_1, \ldots, X_n \in L_2$. Entonces:

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j).$$

¿Cómo queda la expresión si las variables son incorreladas dos a dos?

Matriz de covarianzas

En las fórmulas siguientes se entiende que la esperanza y la varianza operan sobre cada elemento del vector o matriz correspondientes.

Sea $X = (X_1, \ldots, X_n)'$ un vector aleatorio n-dimensional (en columna) tal que $X_1, \ldots, X_n \in L_2$.

- Su **vector de medias** es $\mu = (\mu_1, \ldots, \mu_n)'$ donde $\mu_i = \mathrm{E}(X_i)$.
- Su **matriz de covarianzas** es Σ, cuya posición (i, j) es $\sigma_{i,j} = \mathrm{Cov}(X_i, X_j)$. Es fácil comprobar

$$\Sigma = \mathrm{E}[(X - \mu)(X - \mu)'] = \mathrm{E}(XX') - \mu\mu',$$

Transformaciones afines: si A es matriz $p \times n$ y $b \in \mathbb{R}^p$,

- $\mathrm{E}(AX + b) = A\mu + b$.
- $\Sigma_{AX+b} = \mathrm{E}[A(X - \mu)(X - \mu)'A'] = A\Sigma_X A'$.