Estadística

Victoria Torroja Rubio

19/1/2026

Índice general

Despacho: 303E
Correo: beatrizg@ucm.es
Evaluación:

- 80 % examen final
- 20 % evaluación continua (en este caso serán trabajos de R)

Bibliografía:

- 'Interferencia estadística' de M.A. Gómez Villegas
- 'Statistical interference' de Casella Berger

Capítulo 1

Estadística descriptiva

Práctica:

- Hacer en R el ejemplo 1.3.2 (lo del colesterol) del libro (página 11). Lo que hay que hacer aparece en el ejercicio 1.9.3.
- Programar el coeficiente de simetría y de curtosis en R.
- Demostrar todas las cosas del teorema de Fisher.

Ejemplo. **Página 44, Ejercicio 1.** Consideramos la población dada por la distribución de Bernuilli $B\left(\frac{1}{2}\right)$. Se consideran todas las m.a.s de tamaño 3, es decir,

$$(X_1, X_2, X_3).$$

Recordamos que

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2, \ \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Tendremos que

$$\overline{X} = \begin{cases} 0 \to \frac{1}{8} \\ \frac{1}{3} \to \frac{3}{8} \\ \frac{2}{3} \to \frac{3}{8} \\ 1 \to \frac{1}{8} \end{cases}.$$

De esta forma tenemos que

$$P_{\overline{X}}(0) = \frac{1}{8}, \quad P_{\overline{X}}\left(\frac{1}{3}\right) = \frac{3}{8}, \text{ etc.}$$

Lo que estamos haciendo es considerar los diferentes casos para X_1, X_2 y X_3 y calcular la media en cada caso. A partir de ahí, estamos tratando a la media como una variable aleatoria discreta.

Ejemplo. Página 44, Ejercicio 2. Una forma de media θ y varianza 1. Se toma una m.a.s de tamaño n. Así, tenemos que $X \sim F_\theta(x) = P_\theta(X \leq x)$. Tendremos que $E_\theta[X] = \theta$ y $V_\theta(X) = 1$. Tenemos que calcular θ para que

$$P_\theta\left(\left|\overline{X} - \theta\right| \leq 0{,}5\right) \geq 0{,}95.$$

Tenemos que aplicar la desigualdad de Chebychev, primer calculamos

$$E_\theta\left[\overline{X}\right] = \frac{1}{n}\sum_{i=1}^{n} E_\theta[X_i] = \theta.$$

Por otro lado, $D_\theta(X) = \sqrt{V_\theta(X)} = 1$, así

$$D_\theta\left(\overline{X}\right) = \sqrt{V_\theta\left(\overline{X}\right)} = \sqrt{V_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)} =? \frac{1}{n}\sum_{i=1}^{n} V_\theta(X_i) = \frac{1}{n} \to 0.$$

Aplicando la desigualdad de Chebychev

$$P_\theta\left(\left|\overline{X} - \theta\right| \geq \frac{K}{n}\right) \leq \frac{1}{K}.$$

Ejemplo. Página 45, Ejercicio 6. Sea una m.a.s de tamaño n calculamos la distribución de la media muestral $\overline{X}$, cuando la población:

(b) $X \sim \gamma(p, a)$. Denotamos $\theta = (p, a) \in \mathbb{R}^+ \times \mathbb{R}^+$ y decimos que $\theta \in \Theta \subset \mathbb{R}^2$. Recordamos que

$$f(x|_\theta) = f_\theta(x) = \frac{a^p x^{p-1} e^{-ax}}{\Gamma(p)},\ x > 0.$$

Hacemos el ejercicio con la ecuación característica:

$$\varphi_{\overline{X}}(t) = E_\theta\left[e^{it\overline{X}}\right] = E_\theta\left[e^{i\frac{t}{n}\sum_{i=1}^{n} X_i}\right] = \prod_{i=1}^{n} \varphi_{X_i}\left(\frac{t}{n}\right) = \varphi_X\left(\frac{t}{n}\right)^n = \left(1 - i\frac{t}{na}\right)^{-pn}.$$

La función característica se corresponde con $\gamma(np, na)$.

(c) Basta con sustituir $\lambda = 1$ en el apartado anterior.

Distribuciones asociadas a la normal

Consideremos que $(X_1, \ldots, X_n)$ es una m.a.s con cada variable independientes y siguen una distribución $X \sim N(0,1)$. Consideremos la variable aleatoria

$$Y := \sum_{i=1}^{n} X_i^2 = X_1^2 + \cdots + X_n^2.$$

Tenemos que

$$\varphi_Y(t) = \prod_{i=1}^{n} \varphi_{X_i^2}(t).$$

En general, si $Z = X^2$,

$$F_Z(z) = P(Z \leq z) = P\left(X^2 \leq z\right) = P\left(-\sqrt{z} \leq X \leq \sqrt{z}\right) = F_X\left(\sqrt{z}\right) - F_X\left(-\sqrt{z}\right).$$

La función de densidad será

$$f_Z(t) = \frac{1}{2\sqrt{z}} f_X\left(\sqrt{z}\right) + \frac{1}{2\sqrt{z}} f_X\left(-\sqrt{t}\right) = \frac{1}{\sqrt{2\pi}} t^{-\frac{1}{2}} e^{-\frac{1}{2}t},\ t > 0.$$

Podemos ver que se trata de la función de densidad de una distribución Gamma, así, $Z \sim \gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. Definimos $\chi_1^2 \sim \gamma\left(\frac{1}{2}, \frac{1}{2}\right)$, el uno viene de que tenemos un grado de libertad (se genera a partir de una normal $N(0,1)$). Por la propiedad de reproductividad de la distribución Gamma, tenemos que $Y \sim \gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. A esta distribución la llamamos $\chi_n^2 \sim \gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ (decimos que tiene n grados de libertad porque viene de n distribuciones normales $N(0,1)$). Recordamos que la función característica de una distribución γ es

$$\varphi(t) = \left(1 - \frac{it}{a}\right)^{-p}.$$

Ejemplo. **Ejercicio 19.** Sea $(X_1, \ldots, X_n)$ una m.a.s distribuida idénticamente a X tal que

$$P(X = x_i) = p_i,\ i \in \mathbb{N}.$$

Sea $\left(X_{(1)}, \ldots, X_{(n)}\right)$ la muestra ordenada de menor a mayor. Demostremos que la distribución para $X_{(k)}$ es discreta y

$$P\left(X_{(1)} \leq x_i\right) = \sum_{j=k}^{n} \binom{n}{j} \left(F(x_i)\right)^j \left(1 - F(x_i)\right)^{n-j}.$$

El resto viene dado en el enunciado. Tenemos que $X_{(1)} = \text{mín}\{X_1, \ldots, X_n\}$ y $X_{(n)} = \text{máx}\{X_1, \ldots, X_n\}$. Tenemos que

$$F_{X_{(n)}}(t) = P\left(X_{(n)} \leq t\right) = P\left(\bigcap_{1 \leq i \leq n} \{X_i \leq t\}\right) = \prod_{i=1}^{n} P(X_i \leq t) = F_X(t)^n.$$

Por otro lado,

$$F_{X_{(1)}}(t) = 1 - P\left(X_{(1)} > t\right) = 1 - P\left(\bigcap_{1 \leq i \leq n} \{X_i > t\}\right) = 1 - \prod_{i=1}^{n} P(X_i > t) = 1 - \left(1 - F_X(t)\right)^n.$$

Ahora, buscamos $F_{X_{(k)}}(t)$ para $k \in \{1, \ldots, n\}$. Tenemos que

$$F_{X_{(k)}}(t) = P\left(X_{(k)} \leq t\right) = \sum_{j=k}^{n} P(\text{'exactamente } j \text{ por debajo de } t\text{'}) = \sum_{j=k}^{n} \binom{n}{j} F_X(t)^j \left(1 - F_X(t)\right)^{n-j}.$$

Finalmente, recordamos que $P_{X_{(k)}}(x_i) = F_{X_{(k)}}(x_i) - F_{X_{(k)}}(x_{i-1})$.

Teorema 1.1. Es cierto que $\sqrt{n}\left(b_2 - \sigma^2\right) \to N\left(0, \sqrt{\beta_4 - \sigma^4}\right)$.

Teorema 1.2 (Teorema de Fisher)**.** Sea $(X_1, \ldots, X_n)$ una m.a.s idénticamente distribuida a $X \sim N(\mu, \sigma)$. Entonces se cumple que

1. $\overline{X} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.
2. $\dfrac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{n-1}$.
3. $\dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$.
4. $\overline{X}$ y $S = \sqrt{\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$ son independientes.

Demostración. Cogemos $Z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ así, tendremos que

$$T = \frac{\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)}{\sigma^2}\frac{S^2}{n-1}}} = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}.$$

En esto último hemos aplicado el punto 2 del teorema de Fisher. Esto es en teoría una demostración del punto 3 del teorema de Fisher. Para demostrar 4 basta con ver que $Cov\left(\overline{X}, S\right) = 0$, por ser $X \sim N(\mu, \sigma)$. □

Definición 1.1. Sea $(X_1, \ldots, X_n)$ una m.a.s (n) idénticamente distribuidas a $N(0,1)$. Tenemos que $\chi^2_{n-1} := X_1^2 + \cdots + X_n^2 \sim \gamma\left(p = \dfrac{n-1}{2}, a = \dfrac{1}{2}\right)$.

Definición 1.2 (Distribución t_k-student)**.** Sea $\{X_i\}_{i\in\mathbb{N}}$ una sucesión de variables independientes idénticamente distribuidas que siguen la distribución $X \sim N(0,1)$. Cogemos las k primeras y definimos

$$Y = \sqrt{\frac{X_1^2 + \cdots + X_k^2}{k}} \sim \sqrt{\frac{\chi^2_k}{k}}.$$

Si $Z = X_{k+1} \sim N(0,1)$, entonces definimos la T de Student como $T = \dfrac{Z}{Y}$.

Capítulo 2

Muestreo

2.1. Modelos estadísticos

Cada fenómeno aleatorio está gobernado por una función de probabilidad. Para estudiar dicho fenómeno, nos centramos en estudiar su distribución de probabilidad subyacente.

Llamaremos **población** a una variable aleatoria $X \sim F$, donde F es una función de distribución no completamente conocida. Nuestro objetivo será obtener información sobre ciertas características de la distribución de X. Para ello, replicaremos el experimento aleatorio (de forma idéntica e independiente) un número de n veces.

Hay varios tipos de inferencia:

- Inferencia frecuentista: utilizamos información objetiva, es decir, datos de la muestra.
- Inferencia bayesiana: se utiliza información subjetiva.

También podemos diferenciar en función del grado de conocimiento sobre F:

- Inferencia paramétrica: se conoce la forma funcional de F, excepto por ciertas características.
- Inferencia no paramétrica: no se asume una forma funcional para F.

Supongamos por ahora que la verdadera función de distribución F de nuestra población X está contenida en una familia $\mathcal{F}$ de funciones de distribución asociadas a X.

Definición 2.1 (Modelo estadístico). Un **modelo estadístico** es una terna $(\chi, \beta_\chi, \mathcal{F})$, donde

- $\chi \subset \mathbb{R}$ es el espacio muestral.
- β_χ es una σ-álgebra de Borel sobre χ.
- $\mathcal{F}$ es un conjunto de funciones de distribución asociadas a medidas de probabilidad P_X definidas sobre (χ, β_χ).

En la inferencia paramétrica asumimos que conocemos la forma funcional de la verdadera función de distribución F de la variable aleatoria de interés X, excepto por k parámetros $\theta = (\theta_1, \ldots, \theta_k)$. En este caso, la familia $\mathcal{F}$ tiene la forma

$$\mathcal{F} = \{F(\cdot;\theta) \ : \ \theta \in \Theta\}.$$

Decimos que $\theta \in \Theta \subset \mathbb{R}^k$, donde Θ es el **espacio paramétrico**. En el caso de inferencia frecuentista consideraremos que θ es un valor fijo y desconocido, por lo que escribiremos $f_\theta(x)$ para designar a la función de masa. Sin embargo, en el planteamiento bayesiano consideraremos que θ es una variable aleatoria, por lo que escribiremos $f(x|\theta)$.

Definición 2.2 (Modelo estadístico paramétrico)**.** Un **modelo estadístico paramétrico** es una terna $(\chi, \beta_\chi, \mathcal{F})$ donde

$$\mathcal{F} = \{F(\cdot;\theta) \ : \ \theta \in \Theta\}.$$

Ejemplo. Consideremos los siguientes ejemplos.

- Estimar la probabilidad de que un producto resulte defectuoso en una cadena de producción: $X \sim Bern(\theta)$, $\theta \in [0,1]$.
- Estimar el tiempo entre llegadas a la cola de un servidor: $X \sim \exp(\lambda)$, $\lambda > 0$.
- Estimar el peso esperado de niños hasta cierta edad en una región determinada: $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

2.2. Estadísticos muestrales

Consideremos una variable aleatoria X a la que llamaremos **universo** o **población** cuya función de distribución F viene dada por algún elemento de una familia $\mathcal{F}$ de distribuciones de probabilidad.

Definición 2.3 (Muestra aleatoria simple (m.a.s.))**.** Dada una población X se llama **muestra aleatoria simple (m.a.s.)** de tamaño n a la repetición de $X_1, \ldots, X_n$ variables aleatorias independientes con distribución igual a la de X.

Observación. El modelo estadístico asociado a nuestra muestra $\overrightarrow{X} = (X_1, \ldots, X_n)$ es $(\chi^n, \beta_{\chi^n}, \mathcal{F}^n)$, donde $\mathcal{F}^n$ es el conjunto de funciones de distribución $F_X : \mathbb{R}^n \to \mathbb{R}$ tales que

$$F_X(x_1, \ldots, x_n) = \prod_{i=1}^{n} F(x_i), \ \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Observación. Por ejemplo, la función de distribución de la muestra $(x_1, \ldots, x_n)$ es

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F(x_i),$$

donde $F(x)$ es la función de distribución de X, que viene dada por un elemento de $\mathcal{F}$.

Observación. Los datos observados $(x_1, \ldots, x_n)$ son un resultado o realización de la m.a.s., $X_1 = x_1, \ldots, X_n = x_n$.

Definición 2.4 (Estadístico muestral). Dada $(X_1, \ldots, X_n)$ una m.a.s. de X, a toda $f : \mathbb{R}^n \to \mathbb{R}^k$ función conocida medible recibe el nombre de **estadístico muestral**.

Ejemplo. Algunos ejemplos de estadíscitos muestrales son la **media muestral** que viene dada como

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i,$$

y la **cuasivarianza muestral**

$$S^2 := \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$$

Definición 2.5 (Momentos centrales). Se llama **momento muestral respecto al origen** de orden k a $a_k := \frac{1}{n} \sum_{i=1}^{n} X_i^k$, con $k > 0$. Análogamente, se llama **momento muestral respecto al centro** de orden k a $b_k := \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^k$, con $k > 0$.

Ejemplo. Como hemos visto antes, tendremos que $a_1 = \overline{X}$.

Observación. Es sencillo ver que

$$b_k = a_k - \binom{k}{1} \overline{x} a_{k-1} + \cdots + (-1)^k \overline{x}^k,$$

es decir, los momentos muestrales respecto al centro se puede calcular en función de los momentos muestrales respecto al origen. En particular,

$$b_2 = a_2 - a_1^2 = a_2 - \overline{X}^2.$$

Proposición 2.1. Dada una m.a.s. de tamaño n de una población X con momento de segundo orden finito [a], se tiene que $E[\overline{X}] = \mu$ y $V\left(\overline{X}\right) = \frac{\sigma^2}{n}$, donde $\mu = E[X]$ y $\sigma^2 = V(X)$.

[a] Esta hipótesis es necesaria para asegurar la existencia de σ^2 y μ.

Demostración. Aplicando las propiedades de la esperanza,

$$E\left[\overline{X}\right] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E\left[X_i\right] = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu.$$

Por otro lado, aplicando las propiedades de la varianza,

$$
\begin{aligned}
V\left(\overline{X}\right) =& E\left[\left(\overline{X}-\mu\right)^2\right] = E\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\mu\right)\right)^2\right] \\
=& \frac{1}{n^2}E\left[\sum_{i=1}^{n}\left(X_i-\mu\right)^2 + 2\sum_{i=1}^{n}\sum_{j=1,j>i}^{n}\left(X_i-\mu\right)\left(X_j-\mu\right)\right] \\
=& \frac{1}{n^2}\left[\sum_{i=1}^{n}V\left(X_i\right) + 2\sum_{i=1}^{n}\sum_{j>i}E\left[X_i-\mu\right]E\left[X_j-\mu\right]\right] = \frac{\sigma^2}{n}.
\end{aligned}
$$

Podemos observar que el segundo sumando se anula. □

2.3. Propiedades asintóticas de los momentos muestrales

Sea $(X_1,\ldots,X_n)$ una m.a.s (n) de X con $\mu = E\left[X\right]$ y $\sigma = \sqrt{V\left(X\right)}$.

Proposición 2.2 (Momentos muestrales respecto al origen)**.** Se cumple que si $n \to \infty$

$$a_k = \frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{c.s.} \alpha_k = E\left[X^k\right].$$

Particularmente,

$$\overline{X} \xrightarrow{c.s.} \mu \quad \left(\text{Ley Fuerte de Kintchine, } \mu < \infty\right).$$

Demostración. Por el teorema de Kintchine, tenemos que para una sucesión de variables aleatorias $\left\{Y_n\right\}_{n\in\mathbb{N}}$ independientes e idénticamente distribuidas con $E\left[Y_i\right] = \mu < \infty$, $\forall n \in \mathbb{N}$, tenemos que $\overline{Y} \xrightarrow{cs} \mu$. El resultado general lo obtenemos tomando $Y_i = X_i^k$. □

Proposición 2.3 (Momentos muestrales respecto a la media)**.** Si una población tiene momentos de orden $k > 0$, el momento muestral respecto al centro de orden k, converge, casi seguro, al momento poblacional respecto al centro de orden k, es decir

$$b_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^k \xrightarrow{cs} \beta_k = E\left[\left(X-\mu\right)^k\right].$$

En particular,

$$\sigma_n^2 \xrightarrow{cs} \sigma^2 \quad \text{y} \quad S_n^2 \xrightarrow{cs} \sigma^2.$$

Demostración. Hacer en casa. □

Proposición 2.4. Si la población tiene momentos de orden $2k$ respecto al origen, entonces se tiene la siguiente convergencia en ley

$$\frac{a_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}}\sqrt{n} \xrightarrow{\mathcal{L}} N(0,1).$$

En particular,

$$\sqrt{n}\frac{\overline{X} - \mu}{\sigma} \xrightarrow{\mathcal{L}} N(0,1) \text{ (Teorema de Lévy-Lindeberg).}$$

Demostración. Basta con aplicar el teorema central del límite, $\{Y_n\}_{n\in\mathbb{N}}$ con v.a.i.i.d con momento de segundo orden y $E[Y_n] = \mu$ y $V(Y_n) = \sigma^2 < \infty$. Entonces, tendremos que

$$\frac{\overline{Y} - \mu}{\sigma}\sqrt{n} \xrightarrow{\mathcal{L}} N(0,1).$$

Para obtener el resultado general podemos considerar $Y_n = X_n^k$ y $a_k = \overline{Y}$, de forma que $\mu = \alpha_k$ y $\sigma^2 = \alpha_{2k} - \alpha_k^2$. □

Proposición 2.5. Si existe el momento poblacional respecto al origen de orden $2k$, entonces

$$\sqrt{n}\frac{b_k - \beta_k}{\sqrt{\beta_{2k} - \beta_k^2 - 2k\beta_{k-1}\beta_{k+1} + k^2\beta_{k-1}^2\beta_2}} \xrightarrow{\mathcal{L}} N(0,1).$$

En particular,

$$\sqrt{n}\frac{\sigma_n^2 - \sigma^2}{\sqrt{\beta_4 - \sigma^4}} \xrightarrow{\mathcal{L}} N(0,1).$$

Teorema 2.1 (Teorema de Slutsky). Si $X_n \xrightarrow{\mathcal{L}} X$ e $Y_n \xrightarrow{P} a$, con $a \in \mathbb{R}$, entonces

1. $Y_n X_n \xrightarrow{\mathcal{L}} aX$.
2. $X_n + Y_n \xrightarrow{\mathcal{L}} X + a$.

Lema 2.1. Si $\{a_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ con $\lim\limits_{n\to\infty} a_n = \infty$ y $a \in \mathbb{R}$ tal que

$$a_n(X_n - a) \xrightarrow{\mathcal{L}} X,$$

entonces para $g : \mathbb{R} \to \mathbb{R}$ con derivada continua y no nula en a se tiene que

$$a_n(g(X_n) - g(a)) \xrightarrow{\mathcal{L}} g'(a)X.$$

Otro estadístico muestral importante es la función de distribución empírica o muestral.

Definición 2.6 (Función de distribución empírica). Dada una realización particular de una muestra $(x_1, \ldots, x_n)$, la **función de distribución empírica** viene dada por

$$F_n^*(x) = \begin{cases} 0, \; x < x_{(1)} \\ \frac{k}{n}, \; x_{(k)} \leq x < x_{(k+1)}, \; k = 1, \ldots, n-1 \\ 1, \; x_{(n)} \leq x \end{cases}.$$

Recordamos que $\left(x_{(1)}, \ldots, x_{(n)}\right)$ es la muestra ordenada de menor a mayor.

2.4. Distribuciones asociadas a la normal

Sea $(X_1, \ldots, X_n)$ una m.a.s. (n) de $X \sim N(\mu, \sigma)$.

Observación (Distribución de la media muestral). Sabemos que $\varphi_X(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2}$. Por tanto, tendremos que

$$\begin{aligned}\varphi_{\overline{X}}(t) =& E\left[e^{it\overline{X}}\right] = E\left[e^{it\frac{1}{n}\sum_{i=1}^n X_i}\right] = E\left[\prod_{i=1}^n e^{it\frac{1}{n}X_i}\right] = \prod_{i=1}^n \varphi_{X_i}\left(\frac{t}{n}\right) \\ =& \varphi_X^n\left(\frac{t}{n}\right) = e^{it\mu - \frac{1}{2}\frac{\sigma^2}{n}t^2}.\end{aligned}$$

Por tanto, tenemos que $\overline{X} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

2.5. Teorema de Fisher

En el archivo de Fisher hay que saber deducir todas las identidades que vienen en el documento.

Apéndice A

Convergencia de variables aleatorias

A.1. Convergencia en probabilidad

Definición A.1 (Convergencia en probabilidad). Dada una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ definidas sobre un mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$, diremos que $\{X_n\}_{n\in\mathbb{N}}$ **converge en probabilidad** hacia la variable aleatoria X, definida en el mismo espacio de probabilidad, si

$$\forall \varepsilon > 0, \ \lim_{n\to\infty} P\left(\omega \in \Omega \ : \ |X_n(\omega) - X(\omega)| > \varepsilon\right) = 0.$$

Observación. Una definición equivalente es

$$\forall \varepsilon > 0, \ \lim_{n\to\infty} P\left(\omega \in \Omega \ : \ |X_n(\omega) - X(\omega)| \leq \varepsilon\right) = 1.$$

Notación. Si una sucesión $\{X_n\}_{n\in\mathbb{N}}$ converge en probabilidad a X lo escribiremos

$$X_n \xrightarrow{P} X.$$

Proposición A.1 (Propiedades). Si $X_n \xrightarrow{P} X$,

- $X_n - X \xrightarrow{P} 0$.
- Si $c \in \mathbb{R}/\{0\}$, $cX_n \xrightarrow{P} cX$.
- Si Y es una variable aleatoria, $X_n Y \xrightarrow{P} XY$.
- Si $g : \mathbb{R} \to \mathbb{R}$ es continua, $g(X_n) \xrightarrow{P} g(X)$.
- Si $k \in \mathbb{R}^+$, $X_n^k \xrightarrow{P} X^k$.

Si $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} Y$,

- $X_n \pm Y_n \xrightarrow{P} X \pm Y$.
- $X_n Y_n \xrightarrow{P} XY$.
- Si $g : \mathbb{R}^2 \to \mathbb{R}$ es real y continua, $g(X_n, Y_n) \xrightarrow{P} g(X, Y)$.
- Si $k, h \in \mathbb{R}^+$, $X_n^k Y_n^h \xrightarrow{P} X^k Y^h$.

Si $X_n \xrightarrow{P} X$ y $X_n \xrightarrow{P} Y$, se tiene que $P(X = Y) = 1$.
Si $X_n \xrightarrow{P} a$ e $Y_n \xrightarrow{P} b$ con $a, b \in \mathbb{R}$,

- $X_n Y_n \xrightarrow{P} ab$.
- $\dfrac{X_n}{Y_n} \xrightarrow{P} \dfrac{a}{b}$, si $b \neq 0$.

A.2. Convergencia casi segura

Definición A.2 (Convergencia casi segura). Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias, definidas sobre un mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Diremos que $\{X_n\}_{n\in\mathbb{N}}$ **converge casi seguro** a la variable aleatoria X, definida en el mismo espacio de probabilidad, si

$$P\left(\omega \in \Omega \ : \ \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

Notación. Si la sucesión $\{X_n\}_{n\in\mathbb{N}}$ converge casi seguro a X, escribimos

$$X_n \xrightarrow{cs} X.$$

Proposición A.2 (Propiedades). Si $X_n \xrightarrow{cs} X$,

- $X_n - X \xrightarrow{cs} 0$.
- Si $c \in \mathbb{R}/\{0\}$, $cX_n \xrightarrow{cs} cX$.
- Si $g : \mathbb{R} \to \mathbb{R}$ es continua, $g(X_n) \xrightarrow{cs} g(X)$.
- Si $k \in \mathbb{R}^+$, $X_n^k \xrightarrow{cs} X^k$.

Si $X_n \xrightarrow{cs} X$ e $Y_n \xrightarrow{cs} Y$,

- $X_n \pm Y_n \xrightarrow{cs} X \pm Y$.
- $X_n Y_n \xrightarrow{cs} XY$.
- Si $g : \mathbb{R}^2 \to \mathbb{R}$ es continua, $g(X_n, Y_n) \xrightarrow{cs} XY$.
- Si $k, h \in \mathbb{R}^+$, $X_n^k Y_n^h \xrightarrow{cs} X^k Y^h$.
- $\dfrac{X_n}{Y_n} \xrightarrow{cs} \dfrac{X}{Y}$, siempre que los cocientes estén bien definidos.

Teorema A.1. Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión estríctamente creciente de variables aleatorias positivas y supongamos que $X_n \xrightarrow{P} 0$, entonces $X_n \xrightarrow{cs} 0$.

Teorema A.2. Si $X_n \xrightarrow{cs} X$, entonces $X_n \xrightarrow{P} X$.

Teorema A.3. Si $\{X_n\}_{n\in\mathbb{N}} \xrightarrow{P} X$, existe $\left\{X_{n_j}\right\}_{j\in\mathbb{N}}$ subsucesión tal que $X_{n_j} \xrightarrow{cs} X$.

A.3. Convergencia en ley

Definición A.3 (Convergencia en ley). Dada una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$, definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$, diremos que $\{X_n\}_{n\in\mathbb{N}}$ **converge en ley** a la variable aleatoria X, definida sobre el mismo espacio de probabilidad, si la correspondiente sucesión de funciones de distribución de las X_n, $\{F_n\}_{n\in\mathbb{N}}$ converge puntualmente hacia la función de distribución F de X en todo punto de continuidad de la función. Es decir,

$$\lim_{n\to\infty} F_n(x) = F(x),$$

para todos los puntos de continuidad de F.

Observación. La convergencia puntual de F_n a F no es suficiente para garantizar la convergencia, es necesario ver que la función límite sea función de distribución.

Notación. Si la sucesión $\{X_n\}_{n\in\mathbb{N}}$ converge en ley a X, lo escribimos

$$X_n \xrightarrow{d} X \quad \text{o} \quad X_n \xrightarrow{\mathcal{L}} X.$$

Proposición A.3 (Propiedades)**.** Si $X_n \xrightarrow{\mathcal{L}} X$,

- $X_n - X \xrightarrow{\mathcal{L}} 0$.
- Si $c \in \mathbb{R}/\{0\}$, $cX_n \xrightarrow{\mathcal{L}} cX$.
- Si $c \in \mathbb{R}$, $X_n + c \xrightarrow{\mathcal{L}} X + c$.
- Si $g : \mathbb{R} \to \mathbb{R}$ es continua, $g(X_n) \xrightarrow{\mathcal{L}} g(X)$.

Si $X_n \xrightarrow{\mathcal{L}} X$ e $Y_n \xrightarrow{P} c$, con $c \in \mathbb{R}$,

- $X_n \pm Y_n \xrightarrow{\mathcal{L}} X \pm c$.
- Si $c \neq 0$, $X_nY_n \xrightarrow{\mathcal{L}} cX$.
- Si $c = 0$, $X_nY_n \xrightarrow{P} 0$.
- Si $c \neq 0$, $\dfrac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \dfrac{X}{c}$.

Teorema A.4. Si $X_n \xrightarrow{P} X$, entonces $X_n \xrightarrow{\mathcal{L}} X$.

Teorema A.5. Si $k \in \mathbb{R}$ y $X_n \xrightarrow{\mathcal{L}} k$, entonces $X_n \xrightarrow{P} X$.

Observación (Convergencia de las funciones características)**.** Sabemos que las funciones características proporcionan otro medio para determinar la distribución de una variable aleatoria. También se cumple que

$$X_n \xrightarrow{\mathcal{L}} X \Rightarrow \lim_{n\to\infty} \varphi_n(t) = \varphi(t), \ \forall t \in \mathbb{R}.$$

Recíprocamente, si $\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$, $\forall t \in \mathbb{R}$ y $\varphi(t)$ es continua en el 0, entonces $X_n \xrightarrow{\mathcal{L}} X$ y $\varphi(t)$ es la función característica de X.

A.4. Convergencia en media de orden r

Definición A.4 (Convergencia en media de orden r)**.** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias, diremos que **converge en media de orden** r a la variable aleatoria X si

$$\lim_{n\to\infty} E\left(|X_n - X|^r\right) = 0,$$

siendo $E\left(|X_n|^r\right)$ y $E\left(|X|^r\right)$ finitas.

Notación. Denotaremos la convergencia en orden r como

$$X_n \xrightarrow{mr} X.$$

Cuando $r = 2$, hablamos de **convergencia en media cuadrática** y escribimos

$$X_n \xrightarrow{mc} X.$$

Observación. Si existe el momento de orden r, sabemos que existen todos los momentos de orden inferior. De esta forma, si $s \leq r$,

$$X_n \xrightarrow{mr} X \Rightarrow X_n \xrightarrow{ms} X.$$

Proposición A.4 (Propiedades)**.** Algunas propiedades son:

- Si $X_n \xrightarrow{mc} X$, entonces $E\left(X_n\right) \xrightarrow{mc} E\left(X\right)$ y $E\left(X_n^2\right) \xrightarrow{mc} E\left(X^2\right)$. De ambas se deduce que $V\left(X_n\right) \xrightarrow{mc} V\left(X\right)$.
- La sucesión $\{X_n\}_{n\in\mathbb{N}}$ converge en media cuadrática a una constante c si y solo si se verifica que $\lim_{n\to\infty} E\left(X_n\right) = c$ y $\lim_{n\to\infty} V\left(X_n\right) = 0$.
- Si $X_n \xrightarrow{mc} X$ e $Y_n \xrightarrow{mc} Y$,
  - $E\left(X_n Y_n\right) \xrightarrow{mc} E\left(XY\right)$.
  - De esta propiedad y la primera se tiene que $\mathrm{Cov}\left(X_n, Y_n\right) \xrightarrow{mc} \mathrm{Cov}\left(X, Y\right)$.

Teorema A.6. Si $X_n \xrightarrow{mc} X$, entonces $X_n \xrightarrow{P} X$.

A.5. Leyes de los grandes números

A.5.1. Ley débil de los grándes números

Teorema A.7 (Ley débil de los grandes números). Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias, definidas sobre un mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Diremos que obedece la ley débil de los grandes números si existen dos sucesiones $\{A_n\}_{n\in\mathbb{N}}, \{B_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ con $B_n > 0$ y $B_n \to \infty$ creciente, tales que

$$\frac{S_n - A_n}{B_n} \xrightarrow{P} 0,\ n \to \infty,$$

donde $S_n = \sum_{i=1}^{n} X_i$.

Teorema A.8 (Teorema de Chebychev). Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes con $E(X_n) < \infty$, $\forall n \in \mathbb{N}$. Supongamos que existe $\gamma > 0$ tal que $V(X_n) \leq \gamma$, $\forall n \in \mathbb{N}$. Entonces,

$$\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0.$$

Teorema A.9 (Teorema de Markov). Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias con $\lim_{n\to\infty} V\left(\overline{X}_n\right) = 0$, donde $\overline{X}_n = \frac{S_n}{n}$. Entonces,

$$\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0.$$

Teorema A.10 (Teorema de Khintchine). Si $\{X_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas de media finita μ, entonces

$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

Teorema A.11 (Teorema de Bernoulli). Si $\{X_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes tales que $X_n \equiv B(1, p)$, $\forall n \in \mathbb{N}$, entonces

$$\frac{S_n}{n} \xrightarrow{P} p.$$

A.5.2. Ley fuerte de los grandes números

Teorema A.12 (Ley fuerte de los grandes números). Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre un mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Diremos que obedece la ley fuerte de los grandes números si existen $\{A_n\}_{n\in\mathbb{N}}, \{B_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ con $B_n > 0$ y $B_n \to \infty$ creciente tales que

$$\frac{S_n - A_n}{B_n} \xrightarrow{cs} 0.$$

Teorema A.13 (Teorema de Kolmogorov). Si $\{X_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes con $E(X_n) = 0$ y $V(X_n) = \sigma_n^2$, $\forall n \in \mathbb{N}$, tales que

$$\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < \infty,$$

entonces

$$\frac{S_n}{n} \xrightarrow{cs} 0.$$

Teorema A.14 (Teorema de Khintchine). Si $\{X_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas de media finita μ, entonces

$$\frac{S_n}{n} \xrightarrow{cs} \mu.$$

A.6. Teorema central del límite

Teorema A.15 (Teorema central del límite). Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre un mismo espaico de probabilidad $(\Omega, \mathcal{A}, P)$, con media y varianzas finitas. Diremos que obedece el teorema central del límite si

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \xrightarrow{\mathcal{L}} N(0,1).$$

Teorema A.16 (Teorema de Moivre). Si $\{X_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución de Bernoulli de parámetro p, entonces

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} N(0,1).$$

Teorema A.17 (Teorema de Lévy-Lindeberg)**.** Si $\{X_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2 finitas, entonces

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} N\left(0,1\right).$$