MDS7101

Estadística: Teoría y Aplicaciones

ESCRIBAS
Naomí Cautivo B.
Máximo Flores Valenzuela

ÍNDICE

1. SEMANA 1: REPASO DE PROBABILIDADES.

- **¿Qué es una probabilidad?** Una probabilidad es una medida de incertidumbre.
- Tiene dos enfoques: frecuentista y bayesiano. Para el frecuentista, la probabilidad es algo inherente a la naturaleza, y su paradigma de cálculo es $\text{casos favorables}/\text{casos totales}$. Para el bayesiano, la probabilidad es un invento del ser humano, y ya no se usa la fórmula anterior.

1.1. NOTACIONES BÁSICAS

En el curso, usaremos Ω para denotar el espacio muestral, ω para los eventos, y $\mathbb{P}$ para la medida de probabilidad, que corresponde a una función que asigna una probabilidad a cualquier evento en $\mathcal{F}$, donde $\mathcal{F}$ es una colección de subconjuntos de Ω, no necesariamente una partición.

1.2. PROPIEDADES BÁSICAS DE $\mathbb{P}$

① La probabilidad del espacio muestral debe ser siempre 1, es decir, $\mathbb{P}(\Omega) = 1$.

② La probabilidad es no negativa, es decir, para cualquier evento $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$.

③ La probabilidad de la unión de eventos disjuntos es la suma de sus probabilidades por separado, es decir, $\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i)$ cuando $\forall i \neq j, A_i \cap A_j = \emptyset$.

1.3. VARIABLES ALEATORIAS

> **Nota**
>
> Por convención, en este curso usaremos letras mayúsculas para denotar las variables aleatorias (en adelante, abreviadas como v. a.).

Son funciones que toman elementos del espacio muestral, y les asigna a cada uno un número real. Podemos definir una v. a. X como $X : \Omega \to \mathbb{R}$. Por ejemplo, sea X el número de caras en el lanzamiento de una moneda no cargada 3 veces, entonces $X = \{0, 1, 2, 3\}$, porque son las distintas cantidades de caras que pueden salir.

1.3.1. VARIABLES ALEATORIAS DISCRETAS

Se dice que X es una v. a. discreta si toma valores de un conjunto finito, o infinito numerable, y además $\forall x$, $\mathbb{P}(X = x) \neq 0$.

1.3.2. VARIABLES ALEATORIAS CONTINUAS

Se dice que X es una v. a. continua si X toma cualquier valor real con probabilidad cero, es decir, $\forall x, \mathbb{P}(X = x) = 0$.

1.3.3. FUNCIONES DE DENSIDAD

Existen dos funciones de densidad que permiten ver el comportamiento de una variable aleatoria.

- PDF: *Probability Density Function* ($f(x)$). Describe cómo se distribuye la probabilidad a lo largo de los posibles valores de la v. a. En específico, $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) \, \mathrm{d}x$.
- CDF: *Cummulative Density Function* ($F(x)$). Acumula la probabilidad desde $-\infty$ hasta un valor x en el dominio. En específico, $F(x) = \mathbb{P}(X \leq x)$.

Estas funciones están directamente relacionadas mediante la fórmula $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$, lo que puede ser observado gráficamente en la Figura 1.

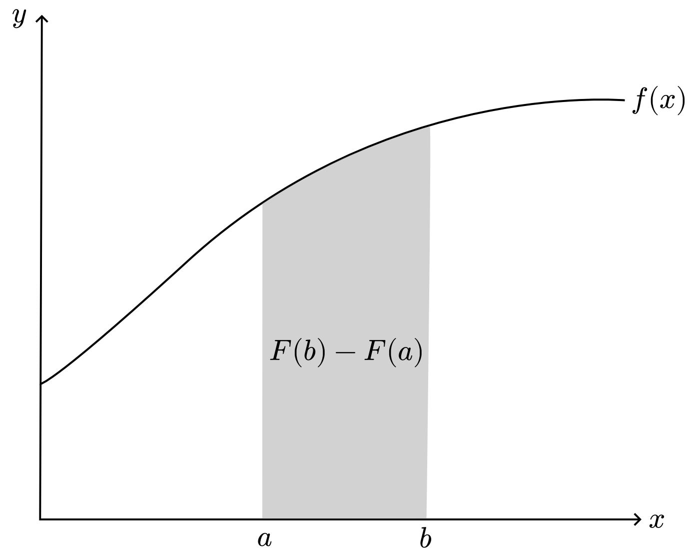

Figura 1: Funciones «PDF» ($f(x)$) y «CDF» ($F(x)$).

Si se conoce F, podemos conocer la probabilidad de un intervalo mediante la siguiente fórmula $\mathbb{P}(a \leq X \leq b) = F(b) - F(a)$.

1.3.4. ESPERANZA DE UNA VARIABLE ALEATORIA

Definimos la esperanza de una variable aleatoria para las v. a. discretas y continuas como:

- X discreta: $\mathbb{E}[X] = \sum_{\Omega} x \cdot \mathbb{P}(X = x)$.
- X continua: $\mathbb{E}[X] = \int_{\mathbb{R}_X} x \cdot f(x)\,\mathrm{d}x$.

También se puede definir como el primer momento de distribución. Los momentos de distribución se definen como $\mathbb{E}[X], \mathbb{E}[X^2], \mathbb{E}[X^3]$, etc.

1.3.5. VARIANZA DE UNA VARIABLE ALEATORIA

Definimos la varianza de una v. a. discreta y continua como:

- X discreta: $\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$.
- X continua: $\mathbb{V}\mathrm{ar}(X) = \int_{\mathbb{R}_X} (X - \mathbb{E}(X))^2 \cdot f(x)\,\mathrm{d}x$.

Con esto mismo podemos definir la desviación estándar de una variable aleatoria, la cual viene a ser la raíz cuadrada de su varianza. Se le conoce también como σ o $\mathrm{STD}(X)$.

1.3.6. ESTANDARIZACIÓN DE UNA VARIABLE ALEATORIA

Sea X una variable aleatoria, se define la variable $Z = (X - \mu)/\sigma$ con $\mu = \mathbb{E}[X]$ y $\sigma = \sqrt{\mathbb{V}\mathrm{ar}(X)}$. Se dice que Z es la estandarización de X, pues cumple $\mathbb{E}[Z] = 0$ y $\mathbb{V}\mathrm{ar}(Z) = 1$.

> ⚠ **Advertencia**
>
> En algunas librerías de programación, la «estandarización» de una v. a. se considera como su «normalización», pero estos términos no son equivalentes.

1.4. DISTRIBUCIONES DISCRETAS

En el curso, veremos principalmente las siguientes distribuciones discretas:

① Bernoulli: $X :=$ lanzamiento de una moneda sólo una vez. Entonces $X \sim \mathrm{Bernoulli}(p)$. Sus valores se definen como:

$$X = \begin{cases} 1 \text{ en el caso de éxito} \\ 0 \text{ en el caso de fracaso} \end{cases}$$

Además, $\mathbb{P}(X = 1) = p$ (probabilidad de éxito) y $\mathbb{P}(X = 0) = 1 - p$ (probabilidad de fracaso). El éxito puede ser, por ejemplo, «obtener cara al lanzar la moneda».

② Binomial: si realizamos el experimento anterior n veces, entonces $X :=$ número de éxitos en n ensayos independientes. Luego, $X \sim \mathrm{Binomial}(p, n)$. La probabilidad asociada a k éxitos es la siguiente:

$$\mathbb{P}(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Además, $\mathbb{E}(X) = np$ y $\mathbb{V}\mathrm{ar}(X) = np \cdot (1 - p)$.

Si p es un vector multivariado $(p_1, p_2, ..., p_n)$, se transforma en una distribución multinomial, denominada $X \sim \mathrm{Multinomial}(p, n)$.

1.5. DISTRIBUCIONES CONTINUAS

① Normal: $X \sim \mathcal{N}(\mu, \sigma^2)$. Su función de densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad x \in \mathbb{R}$$

- Normal estándar: si $X \sim \mathcal{N}(\mu, \sigma^2)$ y $Z = (X - \mu)/\sigma$, entonces $Z \sim \mathcal{N}(0, 1)$.

② «Chi cuadrado» (χ^2): si $Z \sim \mathcal{N}(0, 1)$ entonces:

$$Y = Z^2 \to Y \sim \chi^2_{[1]}$$

donde el subíndice $[1]$ denota los grados de libertad, que es algo que se tratará en las próximas secciones.

③ t-Student: si $Z \sim \mathcal{N}(0, 1)$ e $Y \sim \chi^2_{[n]}$. Entonces definimos t-Student como:

$$t = \frac{Z}{\sqrt{Y/n}} \sim t_{[n]}$$

④ Fischer (F): combinamos dos χ^2 independientes:

$$X_1 \sim \chi^2_{[n_1]} \wedge X_2 \sim \chi^2_{[n_2]} \text{ entonces } F = \frac{X_1/n_1}{X_2/n_2} \sim F_{n_1,n_2}$$

1.6. COVARIANZA DE DOS VARIABLES ALEATORIAS

Medida de cómo en promedio varían linealmente dos variables aleatorias entre sí.

$$\begin{aligned}\mathbb{C}\text{ov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

Si estas variables X, Y son independientes, entonces su covarianza será cero, pues $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ por la propiedad heredada de la esperanza.

> ⚠ **Advertencia**
>
> La implicancia $\mathbb{C}\text{ov}(X,Y) = 0 \Rightarrow X, Y \text{ son independientes}$ es falsa, y es un error muy común asumir que es cierta.

1.7. CORRELACIÓN DE DOS VARIABLES ALEATORIAS.

Es una estandarización de la covarianza, para tener resultados interpretables en el rango $[-1, 1]$. Se calcula de la siguiente forma:

$$\text{Corr}(X,Y) = \frac{\mathbb{C}\text{ov}(X,Y)}{\sqrt{\mathbb{V}\text{ar}(X) \cdot \mathbb{V}\text{ar}(Y)}} = \rho(X,Y)$$

2. SEMANA 2: REPASO DE PROBABILIDADES E INFERENCIA ESTADÍSTICA

- Cuando decimos $\text{Corr}(X,Y) = 0$, quiere decir que no hay información sobre la relación lineal entre X e Y. Esto no quiere decir que X e Y sean independientes, porque pueden tener un tipo de relación no lineal, por ejemplo, cuadrática.
  - Ejemplo: Sea $X \sim \text{U}[-1, 1]$ e $Y = X^2$, con $\text{U}(a, b)$ una distribución uniforme. Como los momentos de una variable Z que distribuye uniformemente en el intervalo (a, b) se calculan mediante la expresión:

    $$\mathbb{E}(Z^n) = \frac{b^{n+1} - a^{n+1}}{(n+1) \cdot (b-a)}$$

    y X es uniforme en el intervalo $[-1, 1]$, entonces su primer momento, $\mathbb{E}(X)$, es nulo. Además, $\mathbb{E}(X^3) = 0$. Esta última expresión nos sirve para deducir la contradicción, pues:

    $$\begin{aligned}\mathbb{C}\text{ov}(X,Y) &= \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y) \\ &= \mathbb{E}(XY) \\ &= \mathbb{E}(X \cdot X^2) \\ &= \mathbb{E}(X^3) = \mathbf{0}\end{aligned}$$

    pero Y sí depende de X, entonces no pueden ser independientes.

2.1. INFERENCIA ESTADÍSTICA

La inferencia estadística es una rama de la estadística que se encarga de hacer predicciones o caracterizaciones sobre una población a partir de una muestra.

Normalmente, habrá una variable $Y \sim f(X)$, con f una función genérica llamada modelo, que encuentra una relación. Y se llama variable endógena, porque depende de X. Será la variable que estudiaremos. Por otro lado, X se llama variable exógena, porque en el mundo ideal no dependen de nada.

Haremos un estudio de X, con una sola variable. Por ejemplo, sea $Y \coloneqq \text{demanda por poleras}$, y $X \coloneqq \text{tallas (estaturas)}$. En Chile, podríamos decir que el promedio de estatura en hombres es $\overline{x}_H = 1.73 \text{ m}$ y en mujeres es $\overline{x}_M = 1.58 \text{ m}$. Diremos que el mínimo es 1 m, y el máximo es 2.5 m.

Podemos decir que las estaturas distribuyen como una variable aleatoria normal, es decir, $X \sim \mathcal{N}(\mu, \sigma^2)$, porque usualmente las concentraciones de estaturas toman esta forma por naturaleza.

2.2. ESTIMADORES

> **Importante**
>
> Para hacer las estimaciones, tomamos muestras aleatorias independientes e idénticamentte distribuidas (en adelante, denotado como i.i.d.). Así, la observación i no depende de la j, y todas vienen de la misma distribución. En el curso trabajaremos sólo con distribuciones i.i.d., salvo que se diga lo contrario.

En el caso anterior, no podemos conocer ni μ ni σ. Como habrán casos donde esto suceda, necesitamos instrumentos que «aproximen» estos valores para poder hacer la inferencia, por ejemplo:

$$\overline{X} = \frac{1}{N} \cdot \sum_{i=1}^{N} X_i$$

- **¿Por qué nos gusta el promedio?** El promedio cumple con propiedades que hacen que sea un buen estimador. Una de ellas se enlista a continuación:
  - *Insesgadez*. Sea $T(X)$ estimador del parámetro θ. $T(X)$ es insesgado si $\mathbb{E}[T(X)] = \theta$. Esto significa que su valor esperado está completamente centrado en el parámetro que estoy buscando. Esta propiedad la cumple el promedio:

$$\begin{aligned}
\mathbb{E}\left(\overline{X}\right) &= \mathbb{E}\left(\frac{1}{N} \cdot \sum_{i=1}^{N} X_i\right) \\
&= \frac{1}{N} \cdot \sum_{i=1}^{N} \mathbb{E}(X_i) \quad (\text{linealidad}) \\
&= \frac{1}{N} \cdot N \cdot \mu = \boldsymbol{\mu} \qquad (X_i \text{ i.i.d.})
\end{aligned}$$

Definimos $\mathbb{V}\mathrm{ar}(T(X))$ como la medida de dispersión del estimador, es decir, qué tan lejos me encuentro del «centro». Para el promedio:

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\overline{X}\right) &= \mathbb{V}\mathrm{ar}\left(\frac{1}{N}\cdot\sum_{i=1}^{N} X_i\right) \\
&= \frac{1}{N^2}\cdot\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{N} X_i\right) \\
&= \frac{1}{N^2}\cdot\sum_{i=1}^{N}\mathbb{V}\mathrm{ar}(X_i) \quad (X_i \text{ i.i.d.}) \\
&= \frac{1}{N^2}\cdot N\cdot\sigma^2 = \boldsymbol{\frac{\sigma^2}{N}}
\end{aligned}
$$

A propósito, queremos que la varianza sea lo más cercana a cero posible, porque esto hace que el estimador esté concentrado en el valor central. Lo malo del resultado obtenido con el promedio, es que si N es muy grande, no podré estimar σ (que sigue siendo desconocido), porque N tiene influencias en el resultado al estar dividiendo.

De esto, nace la necesidad de buscar un estimador insesgado de σ^2. La expresión que toma es la que sigue:

$$
S^2 = \frac{1}{N-1}\cdot\sum_{i=1}^{N}\left(X_i-\overline{X}\right)^2; \quad \mathbb{E}(S^2)=\sigma^2
$$

De esta forma, ya tenemos una estimación de σ^2, por lo tanto, podemos decir que $\mathbb{V}\mathrm{ar}\left(\overline{X}\right) = S^2/N$ con un error $\mathrm{STD}\left(\overline{X}\right) = \sqrt{S^2/N}$.

2.3. INTERVALOS DE CONFIANZA

Se anotan como $\mathrm{IC}\left(\overline{X}\right)$, $\mathrm{CI}\left(\overline{X}\right)$ o $\mathrm{C}\left(\overline{X}\right)$, y corresponden a un rango de valores que con cierta probabilidad contienen al parámetro de interés θ. Lo importante es notar que el parámetro de interés está fijo, lo que varía es justamente el intervalo de confianza.

$$
\mathrm{C}\left(\overline{X}\right) = \overline{X} \pm Z_\alpha\cdot\mathrm{STD}\left(\overline{X}\right)
$$

El valor Z_α es el que escojo para que con «α» nivel de confianza $\mu\in\mathrm{C}(X)$.

$$
\begin{aligned}
\mathbb{P}\left(\mu\in\mathrm{C}\left(\overline{X}\right)\right) &= \mathbb{P}\left(\overline{X}-Z_\alpha\cdot\mathrm{STD}\left(\overline{X}\right)\le\mu\le\overline{X}+Z_\alpha\cdot\mathrm{STD}\left(\overline{X}\right)\right) \\
&= \mathbb{P}\left(-Z_\alpha\le\underbrace{\frac{\overline{X}-\mu}{\mathrm{STD}\left(\overline{X}\right)}}_{\text{estadístico } t}\le Z_\alpha\right)
\end{aligned}
$$

Para fijar la probabilidad de que el parámetro de interés esté en el intervalo de confianza, necesitamos saber cómo distribuye el estadístico t. Vamos a ver algunos ejemplos.

① $X \sim \mathcal{N}(\mu, \sigma^2)$, y supondremos que conocemos σ^2. Entonces $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/N)$ por los cálculos que hicimos anteriormente. Luego,

$$Z \sim \frac{\overline{X} - \mu}{\mathrm{STD}\left(\overline{X}\right)} \sim \mathcal{N}(0,1) \quad \text{(es una normal estandarizada)}$$

Para una normal $\mathcal{N}(0,1)$, el valor de Z_α es aproximadamente 1.96 para una estimación del 95% de confianza para μ (o sea, $\alpha = 1 - 0.95 = 0.05$). Este valor de Z_α varía en función de la probabilidad asociada a la estimación.

② $X \sim \mathcal{N}(\mu, \sigma^2)$, pero no conocemos σ^2. Nuevamente, $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/N)$. Luego, queremos conocer cómo distribuye $Z = \left(\overline{X} - \mu\right)/\sqrt{S^2/N}$. Para esto, necesitamos escribir Z de manera conveniente. Se escribirá de la siguiente forma:

$$Z = \frac{\overline{X} - \mu}{\sqrt{\sigma^2/N}} \Bigg/ \sqrt{\left((N-1) \cdot \frac{S^2}{\sigma^2}\right) \ / \ (N-1)}$$

Ya sabemos que $\left(\overline{X} - \mu\right)/\sqrt{\sigma^2/N} \sim \mathcal{N}(0,1)$. Nos falta estimar el resto. Desarrollando:

$$(N-1) \cdot \frac{S^2}{\sigma^2} = \frac{1}{N-1} \sum_{i=1}^{N} \left[\left(X_i - \overline{X}\right)^2\right] \cdot \frac{N-1}{\sigma^2} = \sum_{i=1}^{N} \left(\frac{X_i - \overline{X}}{\sigma}\right)^2$$

y además, $\left(X_i - \overline{X}\right)/\sigma \sim \mathcal{N}(0,1)$, entonces $(N-1) \cdot S^2/\sigma^2 \sim \chi^2_{[N-1]}$, pues es una suma de normales al cuadrado. Finalmente, y por definición de la variable aleatoria t-Student, Z distribuye $t_{[N-1]}$.

> **Importante**
>
> La suma de variables χ^2 independientes sigue siendo χ^2. Los grados de libertad de la variable resultante son la suma de los grados de libertad de las variables originales.

③ X no distribuye $\mathcal{N}(\mu, \sigma^2)$. Para este caso, es útil emplear una herramienta visual para descartar que su distribución se comporte de forma parecida a una normal. Una manera es usando un *Q-Q Plot* que compara cuantil a cuantil una distribución empírica con una teórica. En este caso, la distribución empírica es X, y la teórica sería una normal.

En la Figura 2 que se muestra a continuación, mientras más cerca esté la línea de puntos azul de la recta, más parecidas son las distribuciones empírica y teórica. Estos roles los toman X y $\mathcal{N}(\mu, \sigma^2)$ respectivamente en el caso que estamos estudiando.

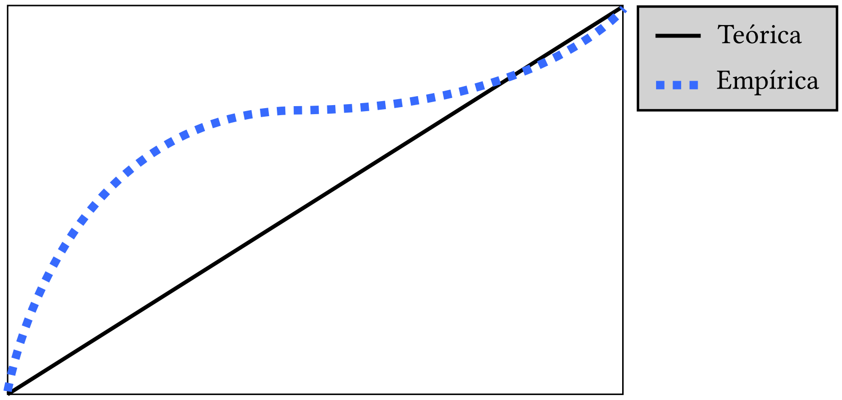

Figura 2: Visualización simplificada de un *Q-Q Plot*.

Si se logra confirmar visualmente que no distribuye normal, debemos buscar otras estrategias para entender la distribución del estadístico *t*. En este punto, introduciremos la teoría asintótica, que se definirá en la siguiente sección (Sección 2.4).

2.4. TEORÍA ASINTÓTICA

2.4.1. CONVERGENCIA EN PROBABILIDAD

Una secuencia de variables aleatorias X_n converge en probabilidad a la variable aleatoria X si para todo $\varepsilon > 0$, su límite cumple lo siguiente:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1$$

Esto se anotará como $X_n \to_{\mathbb{P}} X$ ó $\operatorname{plim}_{n\to\infty} X_n = X$.

> **Nota**
>
> En la mayoría de los *datasets* actuales se tiene que «$n \to \infty$», porque en la estadística clásica, un $n = 30$ ya era considerado muy grande. Esto es porque una t-Student con 30 grados de libertad se empieza a parecer a una normal estándar en distribución.

Basándose en esto se puede definir una nueva propiedad para los estimadores, que extiende la propiedad de insesgadez que se vio en la Sección 2.2:

- *Consistencia*: Un estimador $T(X_n)$ del parámetro θ es consistente si converge en probabilidad al parámetro de interés, es decir:

$$\lim_{n\to\infty} \mathbb{P}(|T(X_n) - \theta| < \varepsilon) = 1$$

2.4.2. EJEMPLOS DE SESGO Y CONSISTENCIA

Un estimador puede ser insesgado y no consistente, o tener otro tipo de variaciones. A continuación, se enlistan ejemplos que dan cuenta de estas variaciones:

① Estimador insesgado e inconsistente:

$$T'(X) = X_1 \;\wedge\; \mathbb{E}(T'(X)) = \mathbb{E}(X_1) = \mu$$

Este estimador de μ es insesgado, porque su esperanza es igual al parámetro estimado, sin embargo, al aumentar la muestra ($n \to \infty$), el valor de $T'(X)$ no cambia, sigue siendo aleatorio e igual a X_1. Al ser un valor aleatorio, esto no se acerca una distancia arbitraria $\varepsilon > 0$ a μ en el límite.

② Estimador sesgado e inconsistente:

$$T''(X) = c \in \mathbb{R}, c \neq \theta$$

En este caso, al ser una constante, el valor esperado es la misma constante (distinta de θ), por lo tanto, cumple ser sesgado. Por otro lado, es inconsistente, ya que la sucesión siempre está concentrada en c, lo que hace imposible que esté centrado en θ, que es lo que se busca con el límite.

③ Estimador sesgado y consistente:

$$S'^2 = \frac{1}{N} \sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2;\ \mathbb{E}(S'^2) = \sigma^2 - \frac{\sigma^2}{N} \neq \sigma^2$$

Este estimador tiene sesgo, porque su valor esperado no es igual al parámetro estimado σ^2. Sin embargo, es consistente, porque converge en probabilidad al parámetro de interés. Esto último se confirma porque el sesgo es $-\sigma^2/N$, que tiende a 0 cuando $N \to \infty$.

2.4.3. CARACTERIZACIÓN DE LA CONSISTENCIA

Si $T(X_n)$ es estimador insesgado de θ, es decir, $E(T(X_n)) = \theta$, y además $\mathbb{V}\mathrm{ar}(T(X_n)) \to 0$ cuando $n \to \infty$, entonces $T(X_n)$ es un estimador consistente de θ. Matemáticamente:

$$T(X_n) \text{ insesgado} \wedge \mathbb{V}\mathrm{ar}(T(X_n)) \to 0 \Longrightarrow T(X_n) \text{ consistente}$$

Por ejemplo, el promedio es un estimador consistente de μ, porque es un estimador insesgado ($E\left(\overline{X}\right) = \mu$), y además $\mathbb{V}\mathrm{ar}\left(\overline{X}\right) = \sigma^2/N \to 0$ cuando $N \to \infty$.

2.4.4. LEY DE LOS GRANDES NÚMEROS

Sea $\{X_i\}_{i \in \mathbb{N}}$ una muestra i.i.d. con $E(X_i) = \mu < \infty$ y $\mathbb{V}\mathrm{ar}(X_i) = \sigma^2 < \infty$ para todo $i \in \mathbb{N}$. La Ley de los Grandes Números (también llamada LGN) establece que:

$$\overline{X_n} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$

es un estimador consistente de μ, es decir, $\overline{X_n} \underset{\mathbb{P}}{\to} \mu$.

2.4.5. CONVERGENCIA EN DISTRIBUCIÓN

Sea X_n es una secuencia de variables aleatorias con $X_n \sim f_n(\cdot)$, y además $X \sim f(\cdot)$. Si para cada x donde $f(x)$ es continua se cumple que $f_n(x) \underset{n \to \infty}{\to} f(x)$ entonces decimos que X_n converge en distribución a X, anotado $X_n \underset{d}{\to} X$.