Muestra aleatoria: v.a. indep. e idénticamente distribuidas

$X_1, \dots, X_n$ es una m.a. de $\mathcal{N}(\mu, \sigma^2)$

$X_1, \dots, X_n$ son v.a.i.i.d. $\mathcal{N}(\mu, \sigma^2)$

$X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Población — parámetros

↓ ↓

Muestra — estadísticos $h(X_1, \dots, X_n)$ [estimadores]

Ejemplo (v.a.i.i.d)

Sea $X_i \sim Exp(\theta)$, $\theta > 0$, denota el tiempo de vida (en meses) de la i-ésima batería de un control remoto que tiene n baterías. La vida de una batería es indep. de las demás. v.a.i.i.d

ⓐ Calcular f.d.p. conjunta de $\underline{X} = (X_1, \dots, X_n)^T$

ⓑ Calcular la proba. de que el control remoto funcione por más de dos meses.

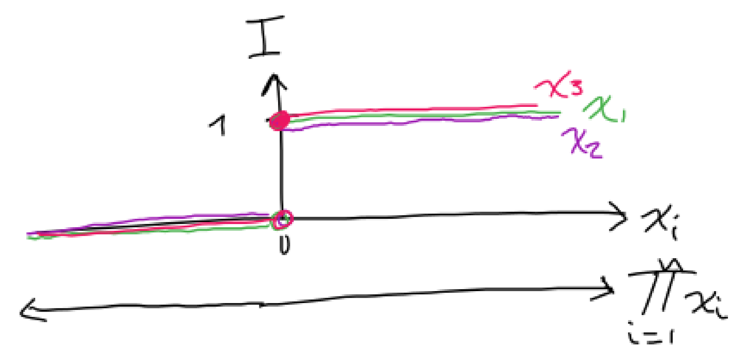
$\xrightarrow{a}$ $f_{X_1,\dots,X_n}(x_1,\dots,x_n) \overset{ind}{=} \prod_{i=1}^{n} f_{X_i}(x_i)$

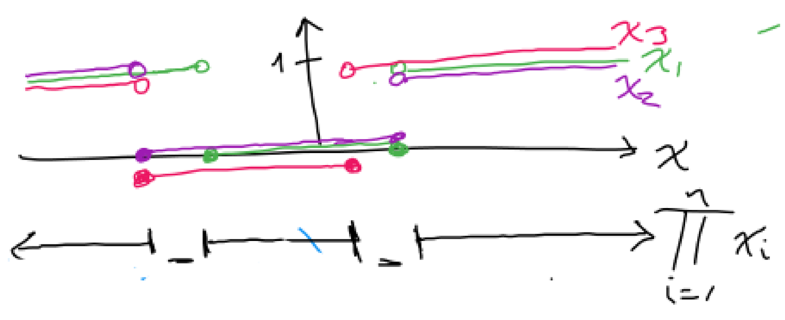
$= \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{1}{\theta}x_i} \, I_{\mathbb{R}_0^+}(x_i)$

$= \left(\frac{1}{\theta}\right)^n e^{\sum_{i=1}^{n}(-\frac{1}{\theta}x_i)} \, I_{\mathbb{R}_0^+}(x_i)$

$= \boxed{\frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i} \, I_{\mathbb{R}_0^+}(x_i)}$

$= \frac{1}{\theta^n} e^{-\frac{1}{\theta} n\bar{x}} \, I_{\mathbb{R}_0^+}(x_i)$





$\xrightarrow{b}$ $P(X_1 \geq 2, X_2 \geq 2, \dots, X_n \geq 2)$

$P(X_1 \leq 2, X_2 \leq 2, \dots, X_n \leq 2)$

$\overset{ind}{=} \prod_{i=1}^{n} P(X_i \leq 2) = \prod_{i=1}^{n} F_{X_i}(2)$

$\overset{*}{=} \prod_{i=1}^{n} \left(1 - e^{-\frac{2}{\theta}}\right) = \left(1 - e^{-\frac{2}{\theta}}\right)^n$

$*$ $X \sim Exp(\theta)$

$F_X(x) = 1 - e^{-\frac{1}{\theta}x}$

$\Rightarrow P(X_1 > 2, \dots, X_n > 2)$

$= 1 - P(X_1 \leq 2, \dots, X_n \leq 2)$

$= 1 - \left(1 - e^{-\frac{2}{\theta}}\right)^n$

Supongamos $\theta = 12$ meses

$n = 2$

$\Rightarrow P(X_1 > 2, X_2 > 2) = 1 - \left(1 - e^{-\frac{1}{6}}\right)^2 = 0.97$

Ejemplo (v.a.i.i.d)

Sean $X_1, \dots, X_n \overset{iid}{\sim} U(0,1)$

ⓐ Encontrar la distribución de muestreo del estadístico

$T(\underline{X}) = \text{máx}\{X_1, \dots, X_n\} = X_{(n)}$ ✓

$\rightarrow F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(\text{máx}\{X_1, \dots, X_n\} \leq x)$

$= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x)$

$\overset{ind}{=} \prod_{i=1}^{n} P(X_i \leq x) = \prod_{i=1}^{n} F_{X_i}(x)$

$\overset{*}{=} \prod_{i=1}^{n} x = x^n = F_{X_{(n)}}(x)$

$*$ $X \sim U(a,b)$

$F_X(x) = \frac{x-a}{b-a}$

$= x$

$\Rightarrow f_{X_{(n)}}(x) = F'_{X_{(n)}}(x) = n x^{n-1}$

(b) ¿Cuál es la proba. de que en un grupo de 15 personas, la calificación más alta sea a lo mucho 6?

$\rightarrow$ X_i = calificación (de 0 a 1) del i-ésimo alumno

$n = 15$ alumnos $\quad X \sim U(0,1)$

$P(X_{(15)} < 0.6) = (0.6)^{15} = 0.00047 \approx 0$

$\therefore$ Si la calificación en un examen se distribuye uniforme, entonces la proba. de que los 15 alumnos reprueban es casi cero.

Def Un estimador es un estadístico que usamos para aproximar un parámetro poblacional

$\rightarrow h(X_1, \dots, X_n) = \hat{\theta}(X_1, \dots, X_n) = \hat{\theta}$

$\hookrightarrow$ estimador para el parámetro θ

$$\mu - \hat{\mu} \qquad \lambda - \hat{\lambda}$$

- ¡Son v.a.!
- Un parámetro puede ser aproximado por más de un estimador.
- Todo estimador es estadístico, pero no todo estadístico es estimador.

Def Una estimación es una realización de un estimador. Es decir que si observamos una muestra tal que

$$X_1 = x_1, \; X_2 = x_2, \; \dots, \; X_n = x_n$$

entonces $h(x_1, x_2, \dots, x_n) = \hat{\theta}(x_1, \dots, x_n) = \hat{\theta}$

$X_1, \dots, X_n$ muestra $\longrightarrow h(X_1, \dots, X_n)$ función de la muestra

$h(X_1, \dots, X_n) \xrightarrow{=} T(X_1, \dots, X_n)$ estadístico

$h(X_1, \dots, X_n) \xrightarrow{=} \hat{\theta}(X_1, \dots, X_n) = \hat{\theta}$ estimador

$h(X_1, \dots, X_n) \xrightarrow{\text{realización de la muestra}} \hat{\theta}(x_1, \dots, x_n) = \hat{\theta}$ estimación

---

X: aleatorio y (hasta el momento) desconocido

x: "realización de la muestra", es decir que X deja de ser desconocida y toma un valor fijo (no aleatorio).

X = estatura

$X \sim N(170, 200 \text{ cm})$

x = estatura de Fulanito

$x = 168$

Ejemplo

X = # de reclamos en un centro de atención telefónica en un día

$X \sim \text{Poisson}(\lambda = 9) \Rightarrow E[X] = 9 = \lambda$

$$\lambda \neq 9, \qquad \lambda = ?$$

Se toma una muestra diaria por dos semanas, es decir

$\underline{X} = (X_1, X_2, \dots, X_{14})$

Como se trata de una m.a., tenemos que $X_i \overset{iid}{\sim} P_0(\lambda)$

Supongamos que no confiamos en que $\lambda = 9$ y queremos, a partir de la muestra, estimar λ.

Para estimar el verdadero valor de λ, consideremos los siguientes tres estimadores

estimadores:

1) $\hat{\lambda}_1(\underline{X}) = \bar{X} = \frac{1}{14}\sum_{i=1}^{14} X_i$

2) $\hat{\lambda}_2(\underline{X}) = \frac{X_{(1)} + X_{(14)}}{2}$ — $X_{(1)} = \min\{X_1, \dots, X_{14}\}$, $X_{(14)} = \max\{X_1, \dots, X_{14}\}$

3) $\hat{\lambda}_3(\underline{X}) = \text{mediana}(\underline{X}) = \frac{X_{(7)} + X_{(8)}}{2}$

Observar que
$$E[\hat{\lambda}_1] = E\left[\tfrac{1}{14}\sum X_i\right] = \tfrac{1}{14}\sum E[X_i] = \tfrac{1}{14}\cdot 14\lambda = \lambda$$

$$E[\hat{\lambda}_2] = E\left[\tfrac{1}{2}\left(X_{(1)} + X_{(14)}\right)\right] = \tfrac{1}{2}\cdot 2\lambda = \lambda$$

$$E[\hat{\lambda}_3] = E\left[\tfrac{1}{2}\left(X_{(7)} + X_{(8)}\right)\right] = \tfrac{1}{2}\cdot 2\lambda = \lambda$$

Supongamos que se observa la muestra

$\underline{x} = (3, 4, 4, 6, 6, 7, 7, 8, 8, 9, 9, 10, 13)$

Considerando esta muestra,

estimaciones:

$\hat{\lambda}_1 = 7.29$

$\hat{\lambda}_2 = 8$

$\hat{\lambda}_3 = 7.5$

¿Cuál de ellas elegimos para estimar λ?

$\hookrightarrow$ (lo veremos más adelante)

$\hookrightarrow$ Lo que sí podemos intuir gracias a $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$ es que el valor real del parámetro poblacional λ es efectivamente menor que 9