Clase 5/06/2023



Recordemos brevemente el resultado exacto de TCL
Supongamos que $\{X_j\}_{j=1}^{n}$ es una sucesión de Gaussianas independientes con distribución $N(\mu,\sigma^2)$ y consideremos el promedio empírico

$$\overline{X}_n = \frac{X_1+\cdots+X_n}{n}$$

Sabemos que $E(\overline{X}_n)=\mu$ $\quad var(X_n) = \sigma^2/n$ lo que implica que $\quad \dfrac{\overline{X}_n-\mu}{\sigma/\sqrt{n}} \stackrel{d}{=} N(0,1)$

Esto se escribe también como $\sqrt{n}\left[\dfrac{\overline{X}_n-\mu}{\sigma}\right]$

luego si $x>0$ podemos calcular sin problemas

$$\underline{P}\left[\sqrt{n}\left|\frac{\overline{X}_n-\mu}{\sigma}\right|>x\right] = 1-\underline{P}\left[\sqrt{n}\left|\frac{\overline{X}_n-\mu}{\sigma}\right|\leq x\right]$$

Concentrémonos en el segundo término

$$\underline{P}\left[\sqrt{n}\left|\frac{\overline{X}_n-\mu}{\sigma}\right|\leq x\right] = \underline{P}\left(-x\leq\sqrt{n}\left[\frac{\overline{X}_n-\mu}{\sigma}\right]\leq x\right]$$

$$= P\left[-\frac{\sigma x}{\sqrt{n}} \leq \overline{X}_n-\mu \leq \frac{\sigma x}{\sqrt{n}}\right] = P\left[\frac{\sigma x}{\sqrt{n}} \geq -\overline{X}_n+\mu \geq -\frac{\sigma x}{\sqrt{n}}\right]$$

$$= P\left[\overline{X}_n-\frac{\sigma x}{\sqrt{n}} \leq \mu \leq \overline{X}_n+\frac{\sigma x}{\sqrt{n}}\right] = \underline{\Phi(x)-\Phi(-x)}$$

$$= 2\Phi(x)-1. \qquad x\geq 0 \qquad \boxed{\Phi(0)=1/2}$$

$$0\leq 2\Phi(x)-1\leq 1$$

De esta forma obtenemos también

$$P\left[\sqrt{n}\left|\frac{\bar{X}_n-\mu}{\sigma}\right|>x\right]=2\left[1-\Phi(x)\right].$$

Una observación importante es que la probabili
de que

$$\mu\in\left[\bar{X}_n-\frac{\sigma x}{\sqrt{n}},\ \bar{X}_n+\frac{\sigma x}{\sqrt{n}}\right]$$

es igual a $2\Phi(x)-1$.

Esto nos permite introducir la noción de intervalo de confianza. Sea $\alpha\in[0,1]$ un número pequeñ tradicionalmente será 0.5, 0.01, etc. y defina $x_\alpha$ el real positivo tal que

$$1-\Phi(x_{\frac{\alpha}{2}})=\alpha/2 \qquad\qquad 1-\alpha/2=\Phi(x_{\frac{\alpha}{2}})$$

Luego $2\Phi(x_{\frac{\alpha}{2}})-1=2-\alpha-1=1-\alpha.$

de esta forma

La probabilidad de que

$$\mu\in\left[\bar{X}_n-\frac{\sigma x_\alpha}{\sqrt{n}},\ \bar{X}_n+\frac{\sigma x_\alpha}{\sqrt{n}}\right]=1-\alpha$$

diremos que con una confianza de $(1-\alpha)\times 100\,\%$ el parámetro estará en ese intervalo.

Podemos entonces introducir la definición de intervalo de confianza ③

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple de $X$ que tiene distribución $p(x;\theta)$ que depende de un parámetro $\theta$.

Dado $\alpha \in (0,1)$, un intervalo de confianza al nivel de confianza $1-\alpha$ para $\theta$ es un intervalo de la forma

$$I_\alpha = I_\alpha(X_1, X_2, \ldots, X_n) = [a(X_1, X_2, \ldots, X_n), b(X_1, X_2, \ldots, X_n)]$$

en donde $a$ y $b$ son tales que

$$\mathbb{P}(\theta \in I_\alpha(X_1, \ldots, X_n)) = 1-\alpha.$$

<u>Observación</u>: ① Notemos que los extremos del intervalo son v.a que dependen de la muestra y son tales que entre ellos está el parámetro $\theta$ con prob. $1-\alpha$

② El nivel de confianza no es la probabilidad de que el parámetro desconocido se encuentre en el intervalo de confianza, sino la probabilidad de que el intervalo de confianza contenga al parámetro desconocido.

③ La definición que acabamos de dar es la de un intervalo exacto. Pero uno se puede encontrar con otros tipos de intervalo.

Así tenemos

$$\begin{cases} 1.\ \text{IdC exacto: es cuando } P(\theta\in I_\alpha(X_1,\dots,X_n)) = 1-\alpha \\ 2.\ \text{IdC aproximado: " " } \lim_{n\to\infty} P(\theta\in I_\alpha(X_1\dots X_n)) = 1-\alpha \\ 3.\ \text{IdC conservado: " " } P(\theta\in I_\alpha(X_1,\dots,X_n)) \geq 1-\alpha. \end{cases}$$

Nuestro desarrollo anterior nos dice que el intervalo de confianza para el parámetro $\mu$ a partir de una muestra $X_1,\dots,X_n$ de $N(\mu,\sigma^2)$ y de nivel de confianza $\underline{\underline{1-\alpha}} \times 100\%$ es.

$$I_\alpha = \left[\overline{X}_n - x_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{X}_n + x_{\alpha/2}\,\sigma/\sqrt{n}\right]$$

Valen dos precisiones:

(a) Cuando $n$ es grande los límites del intervalo se acerca

(b) Para construir este intervalo debemos conocer el verdadero valor del parámetro $\sigma$.

Ahora nos desligaremos de la hipótesis de que las $X_1,\dots,X_n$ son gaussianas. Supongamos en su defecto que $E(X_j)=\mu$ $Var(X_j)=\sigma^2$ y queremos encontrar un intervalo de confianza, esta vez aproximado, para la muestra. No supondremos que $\sigma^2$ es conocido, más bien estimaremos este último parámetro por medio de la varianza empírica.

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$$

Pero tenemos



$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu + \mu - \overline{X}_n)^2$$

desarrollando y simplificando obtenemos

$$= \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \mu)^2 - \frac{n}{n-1} (\mu - \overline{X}_n)^2$$

pero $\frac{1}{n-1} \sum_{j=1}^{n} (X_j - \mu)^2 \xrightarrow{P} \sigma^2$ por la ley de GN.

y también $(\mu - \overline{X}_n)^2 \xrightarrow{P} 0$

luego $S_n^2 \xrightarrow{P} \sigma^2$.

Entonces el Lema de Slutsky dice que si

$$X_n \xrightarrow{d} X \quad \text{y} \quad Y_n \xrightarrow{P} y \text{ (constante)}$$

Entonces $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{y}$.

Así las cosas

Sea entonces $\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{S_n} = \frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\sigma} \frac{\sigma}{S_n}$

entonces $\frac{\sigma}{S_n} \xrightarrow{P} 1$

y por lo tanto $\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{S_n} \longrightarrow N(0,1)$.

Si procedemos como lo hicimos antes para el intervalo exacto, encontramos que

$$I_\alpha = \left[ \overline{X}_n - x_{\alpha/2} \frac{S_n}{\sqrt{n}} , \overline{X}_n + x_{\alpha/2} \frac{S_n}{\sqrt{n}} \right]$$

resulta un intervalo de confianza aproximado para el parámetro $\mu$.

Observemos que todos los ingredientes para construir el intervalo provienen o de la muestra o de la tabla Gaussiana.

Ejemplo de aplicación:

Supongamos que en una encuesta electoral se toma una muestra $n = 2.500$ votantes de los cuales el 31% dicen que votarán por el candidato A (suponemos que hay dos candidatos).

Definimos $X = \begin{cases} 1 & \text{si la persona vota A.} \\ 0 & \text{en otro caso.} \end{cases}$

Queremos estimar el parámetro $\theta$ el porcentaje de todos los votantes que vota a A.

La muestra es $X_1, \text{———}, X_{2500}$

$$\#\text{votantes de A en la muestra} = \sum_{i=1}^{2500} X_i = 775$$

y entonces $\cancel{\ }$ $\overline{X}_{2500} = 0.31$.

Además tenemos

Además sabemos que

$$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2$$

$$= \frac{n}{n-1}\left[\frac{1}{n}\sum_{j=1}^{n} X_j^2 - \overline{X}_n^2\right] \quad \text{pero } X_j^2 = X_j$$

$$= \frac{n}{n-1}\left[\frac{1}{n}\sum_{j=1}^{n} X_j - \overline{X}_n^2\right] = \frac{n}{n-1}\overline{X}_n(1-\overline{X}_n)$$

Entonces el intervalo aproximado de nivel $\alpha$ es

$$I_\alpha = \left[\overline{X}_n \pm x_{\alpha/2}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n-1}}\right]$$

Para $\alpha = 0.05$ $\qquad x_{\frac{0.05}{2}} = 1.96$

$$I_\alpha = \left[0.31 \pm 1.96\sqrt{\frac{0.31(1-0.31)}{2499}}\right].$$

Finalizaremos con el intervalo de confianza para la media para una muestra gaussiana donde no conocemos ni la media ni la varianza

Distribución t- de Student.

Supongamos que tenemos una m.a.s $X_1, X_2 \dots X_n$ de una v.a. Gaussiana $N(\mu, \sigma^2)$

Sea $T_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n}$ ((n-1) grados de libertad) 8

se conoce la distribución de la v.a. se le denomina t- de Student y está tabulada como la de la gaussiana. Entonces:

$$P\left( \left| \frac{\sqrt{n}(\overline{X}_n-\mu)}{S_n} \right| > t \right) = P(|T_n| > t)$$

$$= P(T_n > t) + P(T_n < -t)$$

$$= 1 - F_{t,n-1}(t) + F_{t,n-1}(-t) \quad (*)$$

$F_{t,n-1}(\cdot)$ es la distribución de la $T_n$ de Student con (n-1) grados de libertad. Pero

$1-F_{t,n-1}(t) = F_{t,n-1}(-t)$ ; entonces se tiene

$$P\left( \left| \frac{\sqrt{n}(\overline{X}_n-\mu)}{S_n} \right| > t \right) = 2\left[1-F_{t,n-1}(t)\right]$$

Si fijamos el nivel $\alpha$

$$P\left( |\overline{X}_n - \mu| > \frac{S_n t}{\sqrt{n}} \right) = \alpha \iff F_{t,n-1}(t) = 1 - \alpha/2$$

Si $t_{n-1}(\alpha/2)$ cumple con

$$F_{t_{n-1}}(\alpha/2) = 1 - \alpha/2$$

Con las mismas manipulaciones que hicimos antes

$$|\bar{X}_n - \mu| \le t \frac{S_n}{\sqrt{n}} \Leftrightarrow \mu \in \left[\bar{X}_n - t\frac{S_n}{\sqrt{n}}, \bar{X}_n + t\frac{S_n}{\sqrt{n}}\right]$$

y entonces $I_\alpha = \left[\bar{X}_n - t_{n-1}(\alpha/2)\frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{n-1}(\alpha/2)\frac{S_n}{\sqrt{n}}\right]$

es tal que

$$P(\mu \in I_\alpha) = 1-\alpha.$$

El intervalo

$$I_\alpha \left[\bar{X}_n - t_{n-1}(\alpha/2)\frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{n-1}(\alpha/2)\frac{S_n}{\sqrt{n}}\right]$$

es un intervalo de confianza (exacto) al nivel de confianza $(1-\alpha)$ para $\mu$.

Observación: Se puede ver que el valor crítico $t_{n-1}(\alpha/2)$ es más grande que el correspondiente $x_{\alpha/2}$ para la gaussiana (del caso cuando $\sigma^2$ es conocido).