Laboratorio 4 (Respuestas) - Inferencia Estadística

Convergencia y Teorema Central del Límite

Laboratorista: Héctor Lira Talancón

Ago-Dic 2017

1. Suponga que $X_1, X_2, ...$ es una sucesión de variables aleatorias con $E(X_i) = \mu$, $Var(X_i) = \sigma^2 < \infty$ y $\bar{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Demuestra que $\bar{X}_n \xrightarrow{P} \mu$ (Ley débil de los grandes números).

   RESPUESTA

   Utilizando la desigualdad de Chebyshev tenemos:
   $P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{Var[\bar{X}_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$

   Por lo tanto,
   $$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$$

   Como
   $0 \leq \lim\limits_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon)$ por ser medida de probabilidad y $\lim\limits_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) \leq 0$
   $\Rightarrow \lim\limits_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0.$

2. Suponga que $X_1, ..., X_n$ es una muestra aleatoria de $U(0, \theta)$. Sea $Y = X_{(n)} = \max\{X_1, ..., X_n\}$.
   a) Demuestra que $X_{(n)} \xrightarrow{P} \theta$.

   RESPUESTA

   Sabemos que
   $F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(X_1 \leq x, ..., X_n \leq x) = P(X_1 \leq x) \cdot ... \cdot P(X_n \leq x) = [F_X(x)]^n = (\frac{x}{\theta})^n$ para $0 \leq x \leq \theta$.

   Escrito de otra forma:
   $F_{X_{(n)}}(x) = (\frac{x}{\theta})^n \mathbb{1}_{(0,\theta)}(x) + \mathbb{1}_{[\theta,\infty)}(x)$

   Tomemos
   $\lim\limits_{n\to\infty} P(|X_{(n)} - \theta| \leq \epsilon) = \lim\limits_{n\to\infty} P(-\epsilon \leq X_{(n)} - \theta \leq \epsilon)$
   $= \lim\limits_{n\to\infty} \left[P(X_{(n)} - \theta \leq \epsilon) - P(X_{(n)} - \theta \leq -\epsilon)\right]$
   $= \lim\limits_{n\to\infty} \left[P(X_{(n)} \leq \theta + \epsilon) - P(X_{(n)} \leq \theta - \epsilon)\right]$
   $= \lim\limits_{n\to\infty} \left[1 - P(X_{(n)} \leq \theta - \epsilon)\right]$
   $= 1 - \lim\limits_{n\to\infty} P(X_{(n)} \leq \theta - \epsilon)$

$$= 1 - \lim_{n\to\infty} \left(\frac{\theta - \epsilon}{\theta}\right)^n$$

Como $\left(\frac{\theta-\epsilon}{\theta}\right) < 1$ pues $\epsilon > 0$

$$\Rightarrow \lim_{n\to\infty} \left(\frac{\theta - \epsilon}{\theta}\right)^n = 0$$

$$\Rightarrow \lim_{n\to\infty} P(|X_{(n)} - \theta| \leq \epsilon) = 1$$

b) Demuestra que $W = n(\theta - X_{(n)}) \xrightarrow{d} Exp(\theta)$.

RESPUESTA

Verificando el soporte:

$X_{(n)} \in (0,\theta) \Rightarrow -X_{(n)} \in (-\theta, 0) \Rightarrow \theta - X_{(n)} \in (0,\theta) \Rightarrow n(\theta - X_{(n)}) \in (0, n\theta)$, tomando $n \to \infty \Rightarrow W \in (0,\infty)$

Tomemos la función de distribución acumulada:

$F_W(w) = P(W \leq w) = P(n(\theta - X_{(n)}) \leq w) = P(\theta - X_{(n)} \leq \frac{w}{n}) = P(\theta \leq \frac{w}{n} + X_{(n)}) = P(\theta - \frac{w}{n} \leq X_{(n)}) = 1 - P(X_{(n)} \leq \theta - \frac{w}{n})$

Tomando el límite:

$$\lim_{n\to\infty} F_W(w) = \lim_{n\to\infty} \left[1 - P(X_{(n)} \leq \theta - \frac{w}{n})\right]$$

$$= 1 - \lim_{n\to\infty} \left(P(X_{(n)} \leq \theta - \frac{w}{n})\right)$$

$$= 1 - \lim_{n\to\infty} \left(\frac{\theta - w/n}{\theta}\right)^n$$

$$= 1 - \lim_{n\to\infty} \left(1 - \frac{w/\theta}{n}\right)^n$$

$$= 1 - e^{-w/\theta}$$

Que corresponde a la densidad de una $Exp(\theta)$. Por lo tanto, $W = n(\theta - X_{(n)}) \xrightarrow{d} Exp(\theta)$.

3. Sea $X_1, ..., X_n$ una muestra aleatoria con densidad común f_X, tales que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2 < \infty$. Entonces el Teorema Central del Límite dice:

a) $\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-t^2}{2\sigma^2}} dt$, para toda $x \in \mathbb{R}$

b) $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N(\mu, \frac{\sigma^2}{n})$ si $n \geq 1000$.

c) $\sum_{i=1}^{n} X_i$ tiene una distribución normal para n muy grande.

d) Si $W_n = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)$ con $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, entonces $\lim_{n\to\infty} P(W_n > w) = 1 - P(Z \leq w)$, para toda $w \in \mathbb{R}$, donde $Z \sim N(0,1)$.

RESPUESTA

b) y c) son falsas pues el TCL nos habla sobre una convergencia en distribución cuando $n \to \infty$, no para n fija.

a) es falsa pues $\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$ y la densidad con la que se está calculando la probabilidad corresponde a una $N(0, \sigma^2)$.

Por lo tanto, la respuesta correcta es d).

4. Sea $Y_1, ..., Y_n$ una muestra aleatoria de una población cuya distribución de frecuencias es f_Y tal que $E(Y) = 0$ y $Var(Y) = \sigma^2$. El Teorema Central del Límite dice que:

   A: Si $\bar{Y} = \frac{1}{n} \sum\limits_{i=1}^{n} Y_i$, entonces $\lim\limits_{n\to\infty} P\left(\frac{\bar{Y}}{\sigma/\sqrt{n}} \leq z\right) = \int\limits_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt$

   B: Cuando n es grande, Y_n es aproximadamente normal.

   Entonces,

   a) A y B son falsas.

   b) A es falsa pero B no.

   c) B es falsa pero A no.

   d) Ninguna de las anteriores.

   RESPUESTA

   El TCL nos habla de convergencia en distribución sobre la suma de variables con densidad común. Por lo tanto, A es verdadera.

   El TCL no nos habla de realizaciones particulares como Y_n, por lo tanto, B es falsa.

   La respuesta correcta es c).

5. Considera $Y_1, Y_2, ...$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con varianza finita y $E(Y_i) = \mu$. Sea $\bar{Y} = \frac{1}{n} \sum\limits_{i=1}^{n} Y_i$. Considera el siguiente límite $\lim\limits_{n\to\infty} P(\bar{Y}_n < y)$ con $y > \mu$. ¿Cuál de las siguientes afirmaciones es verdadera?

   a) El límite es cero.

   b) El límite es uno.

   c) El límite está estrictamente entre cero y uno.

   d) El límite no existe.

   RESPUESTA

   Sabemos por la Ley Débil de los Grandes Números que $\bar{Y} \xrightarrow{p} \mu$. Dado que $y > \mu$ podemos decir que $y = \mu + \epsilon$ con $\epsilon > 0$.

   Entonces,

   $\lim\limits_{n\to\infty} P(\bar{Y}_n < y) = \lim\limits_{n\to\infty} P(\bar{Y}_n - \mu < \epsilon) = 1$

   La respuesta correcta es b).

6. Considera $Y_1, Y_2, ...$ una sucesión de variables aleatorias tales que $Y_n \xrightarrow{P} c$ cuando $n \to \infty$ y c una constante. Considera las siguientes afirmaciones:

   A: $F_{Y_n}(c + 0.01) \to 1$ cuando $n \to \infty$.

B: $F_{Y_n}(c - 0.01) \to 1$ cuando $n \to \infty$.

Entonces,

a) A y B son falsas.

b) A es verdadera y B falsa.

c) A y B son verdaderas.

d) Ninguna de las anteriores.

RESPUESTA

Como $Y_n \xrightarrow{P} c$, entonces $Y_n \xrightarrow{d} c$, esto es, $F_{Y_n}(x) \xrightarrow{d} \mathbb{1}_{[c,\infty)}(x)$. Esto implica que

$F_{Y_n}(c + 0.01) \xrightarrow{d} \mathbb{1}_{[c,\infty)}(c + 0.01) = 1$, pero

$F_{Y_n}(c - 0.01) \xrightarrow{d} \mathbb{1}_{[c,\infty)}(c - 0.01) = 0$.

La respuesta correcta es b).

7. Considera $Y_1, Y_2, ...$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con varianza finita y $E(Y_i) = \mu$. Sea $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Considera el siguiente límite $\lim_{n\to\infty} P\Big(\bar{Y}_n < \mu + \frac{y}{\sqrt{n}}\Big)$ con $y < 0$. ¿Cuál de las siguientes afirmaciones es verdadera?

a) El límite es cero.

b) El límite es mayor a 0.5.

c) El límite es menor a 0.5.

d) El límite es uno.

RESPUESTA

$$\lim_{n\to\infty} P\Big(\bar{Y}_n < \mu + \frac{y}{\sqrt{n}}\Big) = \lim_{n\to\infty} P\Big(\bar{Y}_n - \mu < \frac{y}{\sqrt{n}}\Big) = \lim_{n\to\infty} P\Big(\sqrt{n}(\bar{Y}_n - \mu) < y\Big)$$

donde $\sqrt{n}(\bar{Y}_n - \mu) \dot{\sim} N(0, \sigma^2)$ y la aproximación solo es exacta en el límite. Considerando que $y < 0$, obtenemos

$\lim_{n\to\infty} P\Big(\sqrt{n}(\bar{Y}_n - \mu) < y\Big) < 0.5$ pues estamos tomando la probabilidad sobre una normal (que es simétrica) centrada en 0.

La respuesta correcta es c).

8. Sea $X_1, ..., X_n$ una muestra aleatoria de una densidad $Exp(\beta)$. Sea $\hat{\beta} = \bar{X}$, el estimador de máxima verosimilitud.

a) Demuestra que $\frac{\hat{\beta}-\beta}{\beta/\sqrt{n}} \xrightarrow{d} N(0, 1)$ cuando $n \to \infty$.

RESPUESTA

Sabemos que si $\hat{\beta}$ es un EMV, entonces $\hat{\beta} \xrightarrow{d} N(\beta, \frac{v(\beta)}{n})$ con $v(\beta)/n = \text{CICR}$.

Calculemos la Cota Inferior de Crámer-Rao para estimadores insesgados:

$f_X(x|\beta) = \frac{1}{\beta} e^{-x/\beta}$

$\Rightarrow \log f_X(x|\beta) = \log(\frac{1}{\beta}) - x/\beta$

$\Rightarrow \log f_X(x|\beta) = -\log\beta - x/\beta$

$\Rightarrow \frac{\partial}{\partial\beta}\log f_X(x|\beta) = -\frac{1}{\beta} + \frac{x}{\beta^2}$

$\Rightarrow \frac{\partial^2}{\partial\beta^2}\log f_X(x|\beta) = \frac{1}{\beta^2} - \frac{2x}{\beta^3}$

$\Rightarrow E\Big[\frac{\partial^2}{\partial\beta^2}\log f_X(x|\beta)\Big] = \frac{1}{\beta^2} - \frac{2\beta}{\beta^3}$

$= -\frac{1}{\beta^2}$

$\Rightarrow \text{CICR} = \frac{1}{-nE[\frac{\partial^2}{\partial\beta^2}\log f_X(x|\beta)]} = \beta^2/n$

Como $\hat{\beta} \xrightarrow{d} N(\beta, \frac{v(\beta)}{n}) = N(\beta, \beta^2/n)$

$\Rightarrow \frac{\hat{\beta}-\beta}{\beta/\sqrt{n}} \xrightarrow{d} N(0,1)$

b) Demuestra que $\frac{\hat{\beta}-\beta}{\hat{\beta}/\sqrt{n}} \xrightarrow{d} N(0,1)$ cuando $n \to \infty$.

RESPUESTA

Por principio de invarianza, si $\hat{\beta}$ es EMV de β, entonces $\hat{\beta}/n$ es EMV de β/n. Sabemos además que los EMV son estimadores consistenes, esto es, $\hat{\beta}^2/n \xrightarrow{P} \beta^2/n$.

Del inciso anterior sabemos que $\hat{\beta} - \beta \xrightarrow{d} N(0, \beta^2/n)$.

Definimos $g(X,Y) = X/\sqrt{Y}$. Por el teorema de Slutksy y haciendo un abuso de la notación tenemos que

$g(\hat{\beta}-\beta, \hat{\beta}^2/n) = \frac{\hat{\beta}-\beta}{\hat{\beta}/\sqrt{n}} \xrightarrow{d} \frac{N(0,\beta^2/n)}{\sqrt{\beta^2/n}} = N(0,1)$

9. Suponga que $X_1, ..., X_{100}$ son variables aleatorias independientes e idénticamente distribuidas Poisson con valor esperado igual a 1. Encuentre una aproximación para la probabilidad de que la suma sea mayor a 85.

RESPUESTA

Si $E[X_i] = 1 \Rightarrow \lambda = 1$

Sabemos que podemos aproximar $P(X \leq x)$ para $X \sim Po(\lambda)$ con $P(Y \leq y)$ con $Y \sim N(\lambda, \lambda)$, para una λ grande.

Sea $Y = \sum_{i=1}^{100} X_i \Rightarrow Y \sim Po(100)$

$P(Y \leq 85) \approx P(Z \leq \frac{85-100}{10}) = \Phi(-1.5) = 0.93319$

10. a) Sea $Y_n = \frac{\sum_{i=1}^{n} X_i}{n+5}$, para $n = 1, 2, ...$; con X_i's $\sim$ i.i.d. Bernoulli(p). Demuestra que Y_n converge en probabilidad y/o en media cuadrática a p.

RESPUESTA

Primero calculemos lo siguiente:

$E[Y_n] = E[\frac{1}{n+5}\sum_{i=1}^{n} X_i] = \frac{1}{n+5}\sum_{i=1}^{n} E[X_i] = \frac{np}{n+5}$

$Var[Y_n] = Var[\frac{1}{n+5}\sum_{i=1}^{n} X_i] = \frac{1}{(n+5)^2}\sum_{i=1}^{n} Var[X_i] = \frac{np(1-p)}{(n+5)^2}$

Para demostrar convergencia en media cuadrática tomamos

$\lim_{n\to\infty} E[|Y_n - p|^2] = \lim_{n\to\infty} E[(Y_n - p)^2]$

$= \lim_{n\to\infty} \left[E[Y_n^2] - E[2pY_n] + E[p^2]\right]$

$= \lim_{n\to\infty} \left[E[Y_n^2] - 2pE[Y_n] + p^2\right]$

$= \lim_{n\to\infty} \left[Var[Y_n] + E^2[Y_n] - 2pE[Y_n] + p^2\right]$

$= \lim_{n\to\infty} \left[\frac{np(1-p)}{(n+5)^2} + (\frac{np}{n+5})^2 - 2p(\frac{np}{n+5}) + p^2\right]$

$= \lim_{n\to\infty} \frac{np(1-p)}{(n+5)^2} + \lim_{n\to\infty} (\frac{np}{n+5})^2 - \lim_{n\to\infty} 2p\frac{np}{n+5} + \lim_{n\to\infty} p^2$

$= 0 + p^2 \lim_{n\to\infty} (\frac{n}{n+5})^2 - 2p^2 \lim_{n\to\infty} \frac{n}{n+5} + p^2$

$= p^2 \cdot 1 - 2p^2 \cdot 1 + p^2$

$= 2p^2 - 2p^2$

$= 0$

Sabemos que si $Y_n \xrightarrow{L^2} p \Rightarrow Y_n \xrightarrow{P} p$.

b) Sea $\{X_n\}$ una sucesión de variables aleatorias idependientes e idénticamente distribuidas tales que $X_n \sim \text{Poisson}(\lambda)$. Demuestra que $Y_n = e^{-\bar{X}_n}$ converge en probabilidad a $P(X = 0) = e^{-\lambda}$.

RESPUESTA

Primero demostraremos que $\bar{X}_n \xrightarrow{P} \lambda$ y luego, utilizando el Teorema de Mapeo Continuo, se demostrará que $g(\bar{X}_n) = e^{-\bar{X}_n} \xrightarrow{P} g(\lambda) = e^{-\lambda}$:

Calculemos primero

$Var[\bar{X}_n] = Var[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$

Por la desigualdad de Chebyshev tenemos que

$P(|\bar{X}_n - \lambda| \geq \epsilon) \leq \frac{Var[\bar{X}_n]}{\epsilon^2} = \frac{\lambda}{n\epsilon^2}$

$\Rightarrow \lim_{n\to\infty} P(|\bar{X}_n - \lambda| \geq \epsilon) \leq \lim_{n\to\infty} \frac{\lambda}{n\epsilon^2} = 0$

Como $P(|\bar{X}_n - \lambda| \geq \epsilon)$ es una medida de probabilidad $(0 \leq P(\cdot) \leq 1)$

$\Rightarrow P(|\bar{X}_n - \lambda| \geq \epsilon) = 0$

Por lo tanto, $\bar{X}_n \xrightarrow{P} \lambda$,

Definimos $g(X) = e^{-X}$ una función continua. Utilizando el Teorema de Mapeo Continuo, como $\bar{X}_n \xrightarrow{P} \lambda$.
$\Rightarrow g(\bar{X}_n) = e^{-\bar{X}_n} \xrightarrow{P} g(\lambda) = e^{-\lambda}$

11. Analice las siguientes afirmaciones:

A: Sea $X_1, ..., X_n$ una muestra aleatoria con densidad común f_X tales que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2 < \infty$. Defina $W_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$. Si $F_n(w)$ es la función de distribución de la variable aleatoria W_n, entonces $\lim_{n\to\infty} \dfrac{F_n(w)}{\phi(w)} = 1$ donde $\Phi(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w} e^{\frac{-t^2}{2}} dt$.

B: Sea $X_1, ..., X_n$ una muestra aleatoria tales que $X_i \sim U(0, 1)$. Sea $Y_n = \sum_{i=1}^{n} X_i$, entonces $Y_n \dot{\sim} N(\frac{n}{2}, \frac{n}{12})$, donde el símbolo $\dot{\sim}$ significa distribución asintótica.

C: Si X es una varible aleatoria discreta tal que $X \sim Bin(n, p)$, entonces $P(X \leq x) \dot{\sim} \Phi\left(\frac{X-np}{\sqrt{np(1-p)}}\right)$, con $\Phi(\cdot)$ y $\dot{\sim}$ igual que en A y B, respectivamente.

Entonces,
a) Solo A y C son verdaderas.
b) Solo A es verdadera.
c) A, B y C son verdaderas.
d) Solo C es verdadera.

RESPUESTA

A: El TCL nos asegura que W_n converge en distribución a una distribución $N(0, 1)$.

B: El TCL nos asegura que la distribución de Y_n converge a una distribución normal Z con $E[Z] = E[Y_n] = n/2$ y $Var[Z] = Var[Y_n] = n/12$

C: Esta es la definición de la aproximación normal a la distribución binomial. En teoría, habría que aplicar el ajuste de Yate pero es válido el enunciado pues estamos hablando de una distribución aproximada.

Por lo tanto, la respuesta correcta es c).

12. Sean Z_n y Y_n dos sucesiones de variables aleatorias tales que $Z_n \xrightarrow{d} N(0, 1)$ y $Y_n \xrightarrow{P} c, c \neq 0$.
a) Sea $W_n = Y_n Z_n$. Diga cuál es la distribución de W_n.

RESPUESTA

Por el Teorema de Slutsky sabemos que si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{P} c$, entonces $g(X_n, Y_n) \xrightarrow{d} g(X, c)$.

Definimos $g(X, Y) = XY$. Entonces, $W_n = Y_n Z_n \xrightarrow{d} cZ$ con $Z \sim N(0, 1)$.

$E[cZ] = cE[Z] = 0$

$Var[cZ] = c^2 Var[Z] = c^2$

Por lo tanto, $W_n \xrightarrow{d} N(0, c^2)$.

b) ¿Cuál es la distribución límite de $\frac{Z_n}{Y_n}$?

RESPUESTA

Definimos $g(X, Y) = X/Y$. Entonces por el Teorema de Slutsky, $Z_n/Y_n \xrightarrow{d} Z/c$ con $Z \sim N(0, 1)$.

$E[Z/c] = \frac{1}{c}E[Z] = 0$
$Var[Z/c] = \frac{1}{c^2}Var[Z] = \frac{1}{c^2}$

Por lo tanto, $\frac{Z_n}{Y_n} \xrightarrow{d} N(0, 1/c^2)$.

c) Si $X_n \sim F_n(x)$ para cada $n = 1, 2, ...$ con $F_n(x) = (1+\frac{e^{-x}}{n})^{-n}$. Diga cuál es la distribución límite de X_n.

RESPUESTA

Tomando el límite de la distribución y considerando las propiedades de límites:

$$\lim_{n\to\infty} F_n = \lim_{n\to\infty} \left(1 + \frac{e^{-x}}{n}\right)^{-n}$$

$$= \lim_{n\to\infty} \left(\frac{1}{1 + \frac{e^{-x}}{n}}\right)^n$$

$$= \frac{\lim\limits_{n\to\infty} 1}{\lim\limits_{n\to\infty} \left(1 + \frac{e^{-x}}{n}\right)^n}$$

$$= \frac{1}{\lim\limits_{n\to\infty} \left(1 + \frac{e^{-x}}{n}\right)^n}$$

$$= \frac{1}{e^{e^{-x}}}$$

$$= e^{-e^{-x}}$$

$= e^{-\cosh(x)+\sinh(x)}$ (llegar a la expresión anterior es suficiente)

Con
$\cosh(x) = \frac{e^x + e^{-x}}{2}$
$\sinh(x) = \frac{e^x - e^{-x}}{2}$