Soluciones Práctica 6[1]

1. Ejercicio 1

(a)

$$X \sim N\left(50;16\right)$$
$$\bar{X}_{10} \sim N\left(50;\frac{16}{10}\right)$$
$$\bar{X}_{100} \sim N\left(50;\frac{16}{100}\right)$$
$$\bar{X}_{1000} \sim N\left(50;\frac{16}{1000}\right)$$

(b) En caso que no conociéramos la distribución de cada X_i, pero pudiésemos suponer que son todas independientes e idénticamente distribuidas, estaríamos bajo las hipótesis del Teorema Central del Límite y por lo tanto para los casos donde n es grande (100 y 1000) podríamos afirmar que las distribuciones son <u>aproximadamente</u> Normales (con los mismos parámetros obtenidos en el primer inciso)

(c)

$$P\left(45 \leq X \leq 55\right) = 0{,}7887$$
$$P\left(45 \leq \bar{X}_{10} \leq 55\right) = 0{,}9999$$
$$P\left(45 \leq \bar{X}_{100} \leq 55\right) \approx 1$$
$$P\left(45 \leq \bar{X}_{1000} \leq 55\right) \approx 1$$

(d) Si bien las probablidades calculadas para n valiendo 10, 100 y 1000 en este caso resultan muy similares, en verdad ocurre que

$$P\left(45 \leq X \leq 55\right) \leq P\left(45 \leq \bar{X}_{10} \leq 55\right) \leq P\left(45 \leq \bar{X}_{100} \leq 55\right) \leq P\left(45 \leq \bar{X}_{1000} \leq 55\right)$$

lo cual es consistente con la Ley de los Grandes Números. Pues a medida que n aumenta, la LGN nos dice que la distribución de $\bar{X}_n$ se irá concentrando cada vez mas cerca alrededor de $E(X_i) = 50$.

2. Ejercicio 2

Sabemos que $X \sim N(0;1)$ y $\bar{X}_{10} \sim N\left(0;1/10\right)$

(a) $P(X \leq 0) = P(\bar{X}_{10} \leq 0)$ No importa cuanto valga n.

(b) $P(|X-0| \geq 1) \geq P(|\bar{X}_{10}-0| \geq 1)$, y la diferencia se amplía a medida que crece n.

(c) $P(|X-0| \leq 1) \leq P(|\bar{X}_{10}-0| \leq 1)$ y la diferencia se amplía a medida que crece n.

(d) A medida que crece n la distribución de $\bar{X}_n$ se concentra más y más en torno a μ, resultado que se conoce como LGN. Intuitivamente, al tener más información en la muestra aleatoria simple, podemos aproximar de manera más certera el verdadero valor del parámetro poblacional μ (que en este caso vale cero).

3. Ejercicio 3

(a) Consideramos la variable N = "*Cantidad de contribuyentes que consideran importante...*" Basándonos en todo lo que fue ocurriendo en las anteriores resoluciones, es sabido que $N \sim Bi(8;0{,}82)$ y por lo tanto $P(6 \leq N) = 0{,}83918$.
Si quisiéramos usar el TCL, deberíamos calcular $P(0{,}75 \leq X)$ con $X \sim N(0{,}82;0{,}01845)$. Dicha probabilidad vale aproximadamente $P(0{,}75 \leq X) \approx 0{,}6968448$.
Vemos que existe una diferencia notable entre ambas probabilidades, el motivo es que el tamaño de muestra no es lo suficientemente grande.

[1] Algunas soluciones están basadas en soluciones anteriores hechas por Lara Sánchez Peña y Pedro Martínez Bruera.

(b) En este caso $N \sim Bi(80; 0{,}82)$ y $P(60 \leq N) = 0{,}9573982$.
Aplicando el TCL, tenemos $X \sim N(0{,}82; 0{,}001845)$ y $P(0{,}75 \leq X) \approx 0{,}9484144$.
Vemos que en este caso los valores son considerablemente similares.

4. Ejercicio 4

Sean $T_i \sim U(1;7)$ independienetes para $1 \leq i \leq 192$ definidas como

$$T = \text{“}\mathit{Cantidad\ de\ horas\ diarias\ que\ el\ usuario\ i-ésimo\ utiliza\ la\ app}\text{”}$$

(a) Sea $N = X_1 + \ldots + X_{192}$, se nos pide aproximar el valor de $P(N \geq 800)$.
Utilizando el TCL, tendremos que

$$\frac{N}{192} = \frac{X_1 + \ldots + X_{192}}{192} = \bar{X}_{192} \sim N\left(\frac{1+7}{2}; \frac{(7-1)^2/12}{192}\right)$$

Luego

$$P(N \geq 800) = P\left(\bar{X}_{192} \geq \frac{800}{192}\right) = P\left(\bar{X}_{192} \geq 4{,}17\right) \approx 0{,}0869$$

(b) Sea a la tarifa horaria que le permitiría ganar un mínimo de 2000 USD diarios con una confianza del 95 %. Entonces

$$\begin{aligned}
& P\left(\mathit{Ganancia\ diaria} \geq 2000\right) = 0{,}95 \\
\iff & P\left(a \cdot N \geq 2000\right) = 0{,}95 \\
\iff & P\left(X_1 + \ldots + X_{192} \geq \frac{2000}{a}\right) = 0{,}95 \\
\iff & P\left(\bar{X}_{192} \geq \frac{2000}{192 \cdot a}\right) = 0{,}95
\end{aligned}$$

Luego por el TCL y utilizando la app

$$\frac{2000}{192 \cdot a} = 3{,}794 \iff a = \frac{2000}{192 \cdot 3{,}794} \iff a = 2{,}75$$

5. Ejercicio 5

Defininimos la variable $X \sim N(500; \sigma^2)$ como

$$X = \text{“}\mathit{Peso\ de\ un\ paquete\ (en\ gramos)}\text{”}$$

Tenemos $X_1, \ldots, X_{25}$ independientes y sabemos, sin necesidad de hacer uso del TCL, que

$$\bar{X}_{25} \sim N\left(500; \frac{\sigma^2}{n}\right)$$

Queremos un valor de σ tal que $P(490 \leq \bar{X}_{25} \leq 520) = 0{,}99$:

$$\begin{aligned}
& P(490 \leq \bar{X}_{25} \leq 510) = 0{,}99 \\
\iff & P\left(\frac{490-500}{\sigma/5} \leq Z \leq \frac{510-500}{\sigma/5}\right) = 0{,}99 \\
\iff & P\left((-10) \cdot \frac{5}{\sigma} \leq Z \leq 10 \cdot \frac{5}{\sigma}\right) = 0{,}99
\end{aligned}$$

Para despejar σ vamos a usar una propiedad que se desprende de la simetría de la Normal. Sea $a > 0$, entonces

$$P(-a \leq Z \leq a) = p \iff 1 - 2 \cdot F_Z(-a) = p$$

Aplicando este resultado al ejercicio:

$$
\begin{aligned}
& P\left(\frac{-50}{\sigma} \leq Z \leq \frac{50}{\sigma}\right) = 0{,}99 \\
\iff & 1 - 2 \cdot F_Z\left(\frac{-50}{\sigma}\right) = 0{,}99 \\
\iff & F_Z\left(\frac{-50}{\sigma}\right) = 0{,}005 \\
\iff & \frac{-50}{\sigma} = -2{,}575829 \\
\iff & \sigma = \frac{-50}{-2{,}575829} = 19{,}41123
\end{aligned}
$$

6. Ejercicio 6

Si n es la cantidad de reservas aceptadas, consideramos las variables $X_i \sim Bernoulli(0{,}1)$ independientes con $1 \leq i \leq n$ definidas como

$$
X_i = \begin{cases} 1 & \text{Si el pasajero cancela su vuelo.} \\ 0 & \text{Si el pasajero no cancela su vuelo.} \end{cases}
$$

Sea

$$
N = \text{“}Cantidad\ de\ clientes\ que\ cancelan\ su\ vuelo\text{”} = \sum_{i=1}^{n} X_i,
$$

luego el evento “hay clientes con reserva que quedan indigndos” es equivalente a $100 < n - N$ y nos interesa encontrar un n apropiado para que $P(100 < n - N) = P(N < n - 100) < 0{,}01$.
Por el TCL

$$
\frac{N}{n} = \bar{X}_n \sim N\left(0{,}1; \frac{0{,}1 \cdot (1 - 0{,}1)}{n}\right)
$$

y el problema se reduce a encontrar n de manera que

$$
P(N < n - 100) = P\left(\hat{X}_n < 1 - \frac{100}{n}\right) < 0{,}01
$$

Estandarizando

$$
\begin{aligned}
& P\left(\hat{X}_n < \frac{n - 100}{n}\right) < 0{,}01 \\
\iff & P\left(Z < \left(\frac{n - 100}{n} - 0{,}1\right) \cdot \frac{\sqrt{n}}{0{,}3}\right) < 0{,}01 \\
\iff & P\left(Z < \left(\frac{0{,}9 \cdot n - 100}{n}\right) \cdot \frac{\sqrt{n}}{0{,}3}\right) < 0{,}01 \\
\iff & F_Z\left(\frac{3 \cdot n - 1000/3}{\sqrt{n}}\right) < 0{,}01 \\
\iff & \frac{3 \cdot n - 1000/3}{\sqrt{n}} \leq -2{,}326348
\end{aligned}
$$

Realizando la sustitución $\sqrt{n} = z$ y planteando habrá que resolver

$$
3 \cdot z^2 + 2{,}37 \cdot z - \frac{1000}{3} = 0,
$$

aplicando la fórmula resolvente se obtiene $z = 10{,}15$ (la otra raíz se descarta porque es negativa) y consecuentemente $n = 103{,}0225$.
Teniendo en cuenta que el n que buscamos debe ser un número entero y que lo que pretendemos no es que necesariamente valga la igualdad, sino la desigualdad, tomaremos $n = 103$ como respuesta ($n = 104$ no verifica la desigualdad).

7. Ejercicio 7

$$T = \{\text{Galletitas de salvado recibidas}\}$$
$$T = X + Y + Z \; X, Y, Z, \text{ independientes}$$
$$X \sim N(100, 20) \;\; Y = 97 + W \; W \sim Exp(\lambda = 1/3) \;\; Z \sim U[80, 90]$$
$$\mathbf{E}[T] = \mathbf{E}[X + Y + Z] = \mathbf{E}[X] + \mathbf{E}[Y] + \mathbf{E}[Z] = 100 + \mathbf{E}[97 + W] + \frac{90 + 80}{2} = 100 + 100 + 85 = 285$$
$$Var(T) = Var(X + Y + Z) \overset{indep}{=} Var(X) + Var(Y) + Var(Z) =$$
$$20 + Var(W) + \frac{(90 - 80)^2}{12} = 20 + 9 + \frac{25}{3} = \frac{112}{3}$$
$$P(275 < T < 295) = P(-10 < T - \mathbf{E}[T] < 10) = P(|T - \mathbf{E}[T]| < 10)$$

Por la desigualdad de Chebyshev

$$P(|T - \mathbf{E}[T]| > 10) \leq \frac{Var(T)}{10^2} = \frac{28}{75}$$

Entonces

$$1 - P(|T - \mathbf{E}[T]| < 10) \leq \frac{28}{75} \implies P(|T - \mathbf{E}[T]| < 10) \geq \frac{47}{75}$$

8. Ejercicio 8

A

X variable aleatoria con distribución desconocida tal que $\mathbf{E}[X] = 5$ y $Var(X) = 0{,}1$. Utilizando la desigualdad de Chebyshev

$$P(4{,}5 \leq X \leq 5{,}5) = P(|X - \mathbf{E}[X]| < 0{,}5) = 1 - P(|X - \mathbf{E}[X]| > 0{,}5) \geq 1 - \frac{Var(X)}{0{,}5^2} = 0{,}60$$
$$\therefore P(|X - \mathbf{E}[X]| < 0{,}5) \geq 0{,}60$$

B

$X_1, \ldots, X_{10} \overset{iid}{\sim} \mathbf{E}(X) = 5, Var(X) = 0{,}1$. Sea $\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$.

$$\mathbf{E}[\bar{X}] = \mathbf{E}\left[\frac{1}{10}\sum_{i=1}^{10} X_i\right] = \frac{1}{1}\mathbf{E}\left[\sum_{i=1}^{1}\right] \overset{iid}{=} \frac{1}{10}\mathbf{E}[X]\sum_{i=1}^{10} 1 = \mathbf{E}[X]$$
$$Var(\bar{X}) = Var\left(\frac{1}{10}\sum_{i=1}^{10} X_i\right) = \frac{1}{100} Var\left(\sum_{i=1}^{10} X_i\right) \overset{iid}{=} \frac{1}{10} Var(X)$$

Utilizando la desigualdad de Chebyshev para acotar la probabilidad

$$P(4{,}5 \leq \bar{X} \leq 5{,}5) = 1 - P(|\bar{X} - \mathbf{E}[\bar{X}]| > 0{,}5) \geq 1 - \frac{\frac{Var(X)}{10}}{0{,}5^2} = 0{,}96$$
$$\therefore P(|\bar{X} - \mathbf{E}[\bar{X}]| < 0{,}5) \geq 0{,}96$$

C

$X_1, \ldots, X_n \overset{iid}{\sim} \mathbf{E}(X) = 5, Var(X) = 0{,}1$. Sea $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Utilizando la desigualdad de Chebyshev

$$P\left(|\bar{X} - \mathbf{E}[\,\bar{X}\,]| < 0{,}5\right) \geq 1 - \frac{\frac{Var(X)}{n}}{0{,}5^2}$$

$$n \to \infty \implies Var\left(\bar{X}\right) \to 0 \implies P\left(|\bar{X} - \mathbf{E}[X]| < 0{,}5\right) \to 1$$

Funciona no sólo para acotar la probabilidad de $|\bar{X} - \mathbf{E}[\,\bar{X}\,]| < 0{,}5$ sino para cualquier $\varepsilon \in \mathbb{R}_+$. Se verifica que $\bar{X}$ converge en probabilidad a $\mathbf{E}\left[\bar{X}\right]$.

D

$X_1, \ldots X_n$ independientes tal que $\mathbf{E}\left[X\right] = 4$ y $Var\left(X\right) = 9 + \frac{2^i + 1}{2^i} = 10 + \frac{1}{2^i}$ $\forall i \in \mathbb{N}$. Sea $Y_n = \bar{X}_n + e^{5 - \frac{1}{n}}$ Si $\bar{X}_n \xrightarrow{p} a$ entonces $Y_n \xrightarrow{p} a + e^5$. Sin embargo como $X_1, \ldots, X_n$ no son idénticamente distribuidas no se puede aplicar la ley débil de los grandes números.

$$\mathbf{E}\left[\bar{X}\right] = \mathbf{E}\left[X\right]$$

$$Var\left(\bar{X}\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \overset{indep}{=} \frac{1}{n^2}\sum_{i=1}^{n}\left(10 + \frac{1}{2^i}\right) = \frac{10}{n} + \frac{1}{n^2}\frac{1 - 0{,}5^{n+1}}{1 - 0{,}5}$$

Utilizando la desigualdad de Chebyshev

$$P\left(|\bar{X} - \mathbf{E}\left[\bar{X}\right]| > \varepsilon\right) \leq \frac{Var\left(\bar{X}\right)}{\varepsilon^2} = \frac{\frac{10}{n} + \frac{1}{n^2}\frac{1-0{,}5^{n+1}}{1-0{,}5}}{\varepsilon^2}$$

$$Var\left(\bar{X}\right) \xrightarrow{p} 0 \implies \lim_{n\to\infty} P\left(|\bar{X} - \mathbf{E}\left[\bar{X}\right]| > \varepsilon\right) = 0 \implies \bar{X} \xrightarrow{p} \mathbf{E}\left[X\right]$$

$$\therefore Y_n \xrightarrow{p} \mathbf{E}\left[X\right] + e^5 = 4 + e^5$$

9. Ejercicio 9

$X_1, \ldots X_n \overset{iid}{\sim} U[0, \theta]$ $\theta_{MV} = X_{\text{máx}}$. Para encontrar la densidad de $X_{\text{máx}}$

$$F_{X_{\text{máx}}} = P\left(\text{máx}\left\{X_1, \ldots, X_n\right\} \leq x\right) = P\left(X_1 \leq x, \ldots, X_n \leq x\right) \overset{indep}{=} \prod_{i=1}^{n} P\left(X_i \leq x\right)$$

$$\overset{id}{=} P\left(X_i \leq x\right)^n = \left(F\left(X \leq x\right)\right)^n$$

$$f_{X_{\text{máx}}}(x) = \frac{d}{dx}\left(F\left(X \leq x\right)\right)^n = n\left(F\left(X \leq x\right)\right)^{n-1} f_x(x) I_{[0,\theta]}(x) = n\frac{x^{n-1}}{\theta^n} I_{[0,\theta]}(x)$$

A

$$\mathbf{E}\left[X_{\text{máx}}\right] = \int_0^\theta xn\frac{x^{n-1}}{\theta^n}dx = \frac{n}{\theta^n}\int_0^\theta x^n dx = \frac{n}{\theta^n}\frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta$$

$$\mathbf{E}\left[X^2_{\text{máx}}\right] = \int_0^\theta x^2 n\frac{x^{n-1}}{\theta^n}dx = \frac{n}{\theta^n}\int_0^\theta x^{n+1}dx = \frac{n}{\theta^n}\frac{\theta^{n+2}}{n+2} = \frac{n}{n+2}\theta^2$$

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

B

$$\mathbf{E}\left[X_{\text{máx}}\right) \neq \theta \therefore X_{\text{máx}} \text{ sesgado}$$

Un estimador de θ es consistente si y solo si el error cuadrático medio tiende a cero

$$ECM\left[X_{\text{máx}}\right] = sesgo^2 + Var\left(X_{\text{máx}}\right) = \left(-\frac{1}{n+1}\theta\right)^2 + \theta^2 \frac{n}{(n+1)^2(n+2)}$$

$$ECM \to 0 \implies X_{\text{máx}} \xrightarrow{p} \theta \therefore X_{\text{máx}} \text{ consistente}$$

C

$$Z = n\left(\theta - X_{\text{máx}}\right)$$

$$P\left(Z \leq z\right) = P\left(n\left(\theta - X_{max}\right) \leq z\right) = P\left(X \geq \theta - \frac{z}{n}\right) = 1 - F_{X_{\text{máx}}} = 1 - \left(\frac{\theta - \frac{z}{n}}{\theta}\right)^n$$

$$\lim_{n\to\infty} F_Z(z) = \lim_{n\to\infty} 1 - \left(\frac{\theta - \frac{z}{n}}{\theta}\right)^n = 1 - e^{-\frac{1}{\theta}z} \therefore Z \xrightarrow{d} Exp\left(\lambda = \frac{1}{\theta}\right) \therefore Z \xrightarrow{d} Exp\left(\lambda = \frac{1}{\theta}\right)$$

10. Ejercicio 10

$X_1, \ldots, X_n \overset{iid}{\sim} P\left(\lambda = 1\right)$.

A

$$P\left(\sum_{i=1}^{20} X_i > 15\right) \leq \frac{\mathbf{E}\left[\sum_{i=1}^{20} X\right]}{15} = \frac{4}{3}$$

La desigualdad de Markov no aporta información relevante para acotar la probabilidad en este caso.

B

Por el Teorema Central del Límite sabemos que

$$\sqrt{n}\left(\bar{X} - \mathbf{E}[X]\right) \xrightarrow{d} N\left(0, Var(X)\right) \implies \sum_{i=1}^{n} X_i \xrightarrow{d} N\left(n\mathbf{E}\left[X\right], nVar(X)\right)$$

$$n = 60 \implies \sum_{i=1}^{60} X_i \xrightarrow{d} N\left(60, 60\right)$$

$$P\left(\sum_{i=1}^{60} X_i > 15\right) = P\left(Z \geq \frac{15 - \mathbf{E}[X]}{\sqrt{nVar(X)}}\right) = 1 - \Phi\left(\frac{15 - 60}{\sqrt{60}}\right) = 1$$

11. Ejercicio 11

$Z_1, \ldots, Z_n \overset{iid}{\sim} U(-20, 10)$. Por el Teorema Central del Límite sabemos que

$$\sqrt{n}\left(\bar{Z} - \mathbf{E}[Z]\right) \xrightarrow{d} N\left(0, Var(Z)\right) \implies \sum_{i=1}^{n} Z_i \xrightarrow{d} N\left(n\mathbf{E}[Z], nVar(X)\right)$$

$$\mathbf{E}[Z] = -5\, Var(Z) = 75$$

$$P\left(\sum_{i=1}^{900} Z \geq -4470\right) = 1 - \Phi\left(\frac{-4470 + 900 * 5}{\sqrt{75 * 900}}\right) = 0{,}4562$$

12. Ejercicio 12

$X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ $\mathbf{E}[X] = \mu$ $Var(X) = \sigma^2$. $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Por el Teorema Central del Límite sabemos que

$$\sqrt{n}\frac{(\bar{X} - \mathbf{E}[X])}{Var(X)} \xrightarrow{d} N(0,1)$$

Como s^2 es un estimador consitente de la varianza también sabemos que

$$s^2 \xrightarrow{p} Var(X) \implies \frac{Var(X)}{s^2} \xrightarrow{p} 1$$

Esto implica que

$$\sqrt{n}\frac{(\bar{X} - \mathbf{E}[X])}{Var(X)}\frac{Var(X)}{s^2} \xrightarrow{d} N(0,1) \therefore \sqrt{n}\frac{(\bar{X} - \mathbf{E}[X])}{s^2} \xrightarrow{d} N(0,1)$$

13. Ejercicio 13

Recordemos que por la desigualdad de Tchebyshev sabemos que

$$P\left(|X - E(X)| \geq k\sqrt{Var(X)}\right) \leq \frac{1}{k^2}.$$

Es importante recordar para poder utilizar la desigualdad de Tchebyshev solamente se necesita saber de una v.a. X: 2

- su media (valor esperado)
- su varianza y que sea finita

¿Podría mejorarse la cota de Tchebyshev utilizando más información sobre la distribución de X? Es decir si usáramos que $X \sim Exp(\lambda)$, además de que ya sabemos que $E(X) = \frac{1}{\lambda}$ y que $Var(X) = \frac{1}{\lambda^2}$. Veamos:

Si usamos la desigualdad de Tchebyshev sólo ultilizamos que $E(X) = \frac{1}{\lambda}$ y que $Var(X) = \frac{1}{\lambda^2}$:

$$P\left(|X - \frac{1}{\lambda}| \geq \frac{k}{\lambda}\right) = 1 - P\left(|X - \frac{1}{\lambda}| < \frac{k}{\lambda}\right) = 1 - P\left(-\frac{k}{\lambda} < X - \frac{1}{\lambda} < \frac{k}{\lambda}\right).$$

Mientras que si usamos que $X \sim Exp(\lambda)$ (notar que esto implica que sabemos que $E(X) = \frac{1}{\lambda}$ y que $Var(X) = \frac{1}{\lambda^2}$)

$$P\left(\left|X - \frac{1}{\lambda}\right| \geq \frac{k}{\lambda}\right) = 1 - P\left(\underbrace{\frac{1-k}{\lambda}}_{<0} < X < \frac{1+k}{\lambda}\right) = 1 - P\left(0 \leq X < \frac{1+k}{\lambda}\right)$$

Donde la última igualdad vale porque la variable X tiene **soporte** (**toma valores**) en $[0, +\infty)$.

$$P\left(\left|X - \frac{1}{\lambda}\right| \geq \frac{k}{\lambda}\right) = 1 - \int_0^{\frac{1+k}{\lambda}} f_X(x)\,dx = 1 - \left(-e^{-\lambda x}\Big|_0^{\frac{1+k}{\lambda}}\right)$$
$$= 1 - \left(-e^{-(k+1)} + 1\right) = e^{-(k+1)}$$

Notamos que esta cota es mucho más "fina" ("mejor") en el sentido que a medida que k es cada vez mayor, la probabilidad calculada de esta última manera será mucho menor que en el caso de Tchebyshev.Por ejemplo,

si $k = 5$ con Tchebyshev (conociendo la media y varianza de X)

$$P\left(\left|X - \frac{1}{\lambda}\right| \geq \frac{k}{\lambda}\right) \leq \frac{1}{25} = 0{,}04$$

Usando el hecho de que conocemos toda la distribución de la variable aleatoria$X$:

$$P\left(\left|X - \frac{1}{\lambda}\right| \geq \frac{k}{\lambda}\right) = e^{-(5+1)} \approx 0{,}002478$$

si $k = 22$ con Tchebyshev (conociendo la media y varianza de X)

$$P\left(\left|X - \frac{1}{\lambda}\right| \geq \frac{k}{\lambda}\right) \leq \frac{1}{484} \approx 0{,}002066115702479$$

Usando el hecho de que conocemos toda la distribución de la variable aleatoria X:

$$P\left(\left|X - \frac{1}{\lambda}\right| \geq \frac{k}{\lambda}\right) = e^{-(22+1)} \approx 0{,}000000000102619$$

14. Ejercicio 14

Recordamos que si una variable aleatoria $X \sim P(\lambda)$, se tiene que:

- $E(X) = \lambda$
- $Var(X) = \lambda$
- El soporte (los posibles valores que toma la v.a.) de X es $\{0, 1, 2, 3, 4, \cdots\}$

Por lo tanto, no sabemos (por ahora) cómo probar la desigualdad que nos da el enunciado, pero sí sabemos por la desigualdad de Tchebyshev que para todo $k > 0$ vale que

$$P\left(|X - \lambda| \geq k\sqrt{\lambda}\right) \leq \frac{1}{k^2}$$

En particular, tomando $k = \sqrt{\lambda}$ vale que

$$P\left(|X - \lambda| \geq \sqrt{\lambda}\sqrt{\lambda}\right) \leq \frac{1}{\sqrt{\lambda}^2}$$

Es decir,

$$\begin{aligned} P\left(|X - \lambda| \geq \lambda\right) &\leq \frac{1}{\lambda} \\ 1 - P\left(|X - \lambda| < \lambda\right) &\leq \frac{1}{\lambda} \\ 1 - P\left(0 < X < 2\lambda\right) &\leq \frac{1}{\lambda} \end{aligned}$$

Notando que $P\left(0 < X < 2\lambda\right) \leq P\left(0 \leq X < 2\lambda\right) = P(X < 2\lambda)$, tenemos que[2]

$$P(X \geq 2\lambda) = 1 - P(X < 2\lambda) \leq 1 - P\left(0 < X < 2\lambda\right) \leq \frac{1}{\lambda}$$

Como queríamos ver.[3]

[2] La última igualdad vale si recordamos que la variable $X \sim P(\lambda)$ tiene soporte en $\{0, 1, 2, \cdots\}$.
[3] Recordar que si $a \leq b$, entonces vale que $-b \leq -a$ y, por lo tanto $1 - b \leq 1 - a$.

15. Ejercicio 15

Recordamos que si una variable aleatoria $X \sim U([a,b])$, se tiene que:

- $E(X) = \dfrac{a+b}{2}$
- $Var(X) = \dfrac{(b-a)^2}{12}$
- El soporte (los posibles valores que toma la v.a.) de X está compuesto por todos los números entre a y b, $[a,b]$

Recordamos que si una variable aleatoria $X \sim N(\mu, \sigma^2)$, se tiene que:

- $E(X) = \mu$
- $Var(X) = \sigma^2$
- El soporte (los posibles valores que toma la v.a.) de X está compuesto por todos los números reales, es decir $\mathbb{R}$.

Sea $Y = X_1 + X_2 + X_3 + X_4$. Por la linealidad de la esperanza,

$$\begin{aligned} E(Y) &= E(X_1 + X_2 + X_3 + X_4) \\ &= E(X_1) + E(X_2) + E(X_3) + E(X_4) = 4 + 3 + 4 + 3 = 14 \end{aligned}$$

Usando también la independencia de las variables X_i,

$$Var(Y) = Var(X_1 + X_2 + X_3 + X_4) \underbrace{=}_{X_i \text{ indep.}} Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4)$$

$$= \frac{(5-3)^2}{12} + \frac{1}{4} + \frac{1}{4} + \frac{(4-2)^2}{12} = \frac{7}{6}$$

Para el inciso b) nos piden que acotemos

$$P(|Y-14| \geq 2) \underbrace{\leq}_{\text{Tcheb.}} \frac{7/6}{2^2} = 0{,}292\,.$$

16. Ejercicio 16

Teniendo en cuenta que $X \sim P(30)$ (y que por lo tanto $E(X) = Var(X) = 30$), en este ejercicio se nos pide que acotemos inferiormente $P(20 \leq X \leq 40)$, para eso reescribimos esta expresión de manera que podamos usar la desigualdad de Tchebyshev.

$$\begin{aligned} P(20 \leq X \leq 40) &= P(-10 \leq X - 30 \leq 10) = P(|X-30| \leq 10) \\ &= 1 - P(|X-30| > 10) \underbrace{=}_{X \text{ es discreta}} 1 - P(|X-30| \geq 11) \\ &= 1 - P\left(|X-30| \geq \frac{11}{\sqrt{30}} \cdot \sqrt{30}\right) \underbrace{\geq}_{\text{por Tcheby.}} 1 - \frac{1}{\left(\frac{11}{\sqrt{30}}\right)^2} \\ &= 1 - \frac{30}{121} = \frac{91}{121} \geq \frac{91}{121}\,. \end{aligned}$$

17. Ejercicio 17

a) Recordamos que si una variable aleatoria $X_i \sim Be(p)$, se tiene que:

- $E(X_i) = p$.
- $Var(X_i) = p(1-p)$.
- El soporte (los posibles valores que toma la v.a.) de X es $\{0,1\}$.
- Si $X_1, \cdots, X_n \sim Be(p)$ y son **independientes** entonces $S_n = \sum_{i=1}^{n} X_i \sim Bi(n,p)$, es decir que $E(S_n) = np$, $Var(E_n) = np(1-p)$.
- En las mismas condiciones del punto anterior si definimos $\overline{X}_n = \frac{S_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$, vale que $E(\overline{X}_n) = p$ y que $Var(\overline{X}_n) = \frac{p(1-p)}{n}$.[4] En este ejercicio tenemos que acotar

$$P\left(\left|\overline{X}_n - p\right| \geq 0{,}1\right) = P\left(\left|\overline{X}_n - p\right| \geq \underbrace{\frac{0{,}1}{\sqrt{\frac{p(1-p)}{n}}}}_{k} \cdot \sqrt{\frac{p(1-p)}{n}}\right) \leq \frac{1}{k^2} = \frac{\frac{p(1-p)}{n}}{0{,}1^2}$$

$$= \frac{p(1-p)}{0{,}01^2 n} = \frac{100p(1-p)}{n} \underbrace{\leq}_{p(1-p) \leq \frac{1}{4}} \frac{100 \cdot 0{,}25}{n} = \frac{25}{n}$$

b) Para poder garantizar que

$$P\left(\left|\overline{X}_n - p\right| \geq 0{,}1\right) \leq 0{,}1$$

pedimos que

$$\frac{25}{n} \leq 0{,}1 \iff 250 \leq n\,.$$

Es decir que debemos entrevistar al menos 250 personas para poder garantizar que (independientemente del valor verdadero de p) que $P\left(\left|\overline{X}_n - p\right| \geq 0{,}1\right) \leq 0{,}1$.

18. Ejercicio 18

Próximamente

19. Ejercicio 19

Recordamos que si una variable aleatoria $X \sim U([a,b])$, se tiene que:

- $E(X) = \frac{a+b}{2}$
- $Var(X) = \frac{(b-a)^2}{12}$
- El soporte (los posibles valores que toma la v.a.) de X está compuesto por todos los números entre a y b, $[a,b]$

Recordamos que si una variable aleatoria $X \sim N(\mu, \sigma^2)$, se tiene que:

- $E(X) = \mu$
- $Var(X) = \sigma^2$
- El soporte (los posibles valores que toma la v.a.) de X está compuesto por todos los números reales, es decir $\mathbb{R}$.

[4] Notar que la independencia de las X_i sólo se utiliza para calcular la varianza y en el punto anterior también para calcular la distribución de la variable. No obstante, para calcular los valores esperado de S_n y de $\overline{X}_n$ no es necesario utilizar el supuesto de independencia de las variables X_i.

Sea $Y = X_1 + X_2 + X_3 + X_4$. Por la linealidad de la esperanza,

$$\begin{aligned} E(Y) &= E(X_1 + X_2 + X_3 + X_4) \\ &= E(X_1) + E(X_2) + E(X_3) + E(X_4) = 4 + 3 + 4 + 3 = 14 \end{aligned}$$

Usando también la independencia de las variables X_i,

$$\begin{aligned} Var(Y) = Var(X_1 + X_2 + X_3 + X_4) &\underbrace{=}_{X_i \text{ indep.}} Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4) \\ &= \frac{(5-3)^2}{12} + \frac{1}{4} + \frac{1}{4} + \frac{(4-2)^2}{12} = \frac{7}{6} \end{aligned}$$

Para el inciso b) nos piden que acotemos

$$P(|Y - 14| \geq 2) \underbrace{\leq}_{\text{Tcheb.}} \frac{7/6}{2^2} = 0{,}292\,.$$

20. Ejercicio 20

Dadas observaciones independientes $X_1, \dots, X_n$ tales que $E(X_i) = 0$ y $a_i \leq X_i \leq b_i$, la desigualdad de Hoeffding dice que dado $\varepsilon > 0$ fijo se cumple la siguiente desigualdad

$$P\left(\sum_{i=1}^{n} X_i \geq \varepsilon\right) \leq \exp\left(-2\,\frac{\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \tag{1}$$

a) Si $X_1, \dots, X_n \sim Be(p)$ pruebe que para todo $\varepsilon > 0$ se cumple que

$$P(|\overline{X}_n - p| > \varepsilon) \leq 2e^{-2n\varepsilon^2} \tag{2}$$

donde $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$.

b) Muestre que para valores grandes de n la cota de Hoeffding es mejor que la de Tchebyshev.

Solution

a) **Observación**: más adelante esta desigualdad nos será útil para acotar la longitud de intervalos de confianza asintóticos.

Para probar que vale (2) vamos a mostrar que:

- $P(\overline{X}_n - p \geq \varepsilon) \leq e^{-2n\varepsilon^2}$
- $P(p - \overline{X}_n \geq \varepsilon) \leq e^{-2n\varepsilon^2}$

Por lo tanto, eso implicará que vale (1) ya que

$$P(|\overline{X}_n - p| > \varepsilon) = P(\overline{X}_n - p \geq \varepsilon) + P(p - \overline{X}_n \geq \varepsilon) \leq e^{-2n\varepsilon^2} + e^{-2n\varepsilon^2} = 2e^{-2n\varepsilon^2}$$

Para aplicar la desigualdad de Hoeffding necesitamos variables con esperanza igual a cero. Definimos $Y_i = \dfrac{X_i - p}{n}$, tenemos que:

- Las variables Y_i son independientes porque las X_i lo son.
- $E(Y_i) = 0$ porque $E(X_i) = p$ pues $X_i \sim Be(p)$.
- Además, como $0 \leq X_i \leq 1$ tenemos que $\underbrace{-p/n}_{a_i} \leq Y_i \leq \underbrace{(1-p)/n}_{b_i}$.

Entonces $\sum_{i=1}^{n}(b_i-a_i)^2 = n\cdot\frac{1}{n^2} = \frac{1}{n}$.

Se verifican todas las hipótesis para aplicar la desigualdad de Hoeffding:

$$P\left(\overline{X}_n - p > \varepsilon\right) \leq P\left(\overline{X}_n - p \geq \varepsilon\right) = P\left(\sum_{i=1}^{n} Y_i \geq \varepsilon\right) \leq e^{-2n\varepsilon^2} \tag{3}$$

Por otra parte,

$$\begin{aligned} P\left(\overline{X}_n - p < -\varepsilon\right) \leq P\left(\overline{X}_n - p \leq -\varepsilon\right) =& P\left(p - \overline{X}_n \geq \varepsilon\right) \\ =& P\left(-\sum_{i=1}^{n} Y_i \geq \varepsilon\right) \leq e^{-2n\varepsilon^2} \quad [5] \end{aligned} \tag{4}$$

Luego, como valen (3) y (4), se tiene que $P\left(|\overline{X}_n - p| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2}$, como queríamos ver. ■

b) Las cotas son:

- Hoeffding (es una cota uniforme porque no depende del valor del parámetro p): $P\left(\left|\sum_{i=1}^{n} X_i - np\right| \geq n\cdot\varepsilon\right) \leq 2e^{-2n\varepsilon^2}$
- Tchebyshev (es una cota puntual porque depende del valor del parámetro p): $P\left(\left|\sum_{i=1}^{n} X_i - np\right| \geq n\cdot\varepsilon\right) \leq \frac{\left(\sqrt{np(1-p)}\right)^2}{(n\varepsilon)^2} = \frac{p\,(1-p)}{n\varepsilon^2}$[6]

Comparamos las cotas para ε y p fijos:

$$\lim_{n\to\infty} \frac{2e^{-2n\varepsilon^2}}{\frac{p(1-p)}{n\varepsilon^2}} = \lim_{n\to\infty} \frac{2n\varepsilon^2}{e^{2n\varepsilon^2}p(1-p)} \to 0$$

Como el límite es cero, a partir de cierto valor de n la cota de Hoeffding es mejor.

Para el inciso b) nos piden ver que para valores de n grandes, la cota de Hoeffding es 'mejor' que la cota de Tchebyshev. Esta diferencia será más notoria en la medida que p no sea 'demasiado' cercano a 0 ni a 1.

Notar que si $p(1-p) \approx 0$ (es decir que si p fuera 'cercano' a 0 o 1, entonces la cota de Tchebyshev (ya que depende del valor de p) podría llegar a tener una no tan mala performance respecto de la cota de Hoeffding (que no depende de p, es una cota que vale independientemente del valor de p, se conoce como una cota uniforme, justamente porque no depende del valor de p). No obstante, si por ejemplo p representase la población de individuos desempleados sabemos que p no es (lamentablemente) 'cercano' a cero ni (por suerte) similar a uno. Por lo tanto, en un caso así que la cota de Hoeffding será más 'precisa'.

Recordamos que las cotas son:

- Para Hoeffding (es una cota uniforme porque no depende del valor del parámetro p): $P\left(\left|\sum_{i=1}^{n} X_i - np\right| \geq n\cdot\varepsilon\right) \leq 2e^{-2n\varepsilon^2}$

[5] Notar que $P\left(-\sum_{i=1}^{n} Y_i \geq \varepsilon\right) = P\left(\sum_{i=1}^{n} -Y_i \geq \varepsilon\right)$, donde $E(-Y_i) = 0$, $-b_i \leq -Y_i \leq -a_i$. Entonces es cierto que $\sum_{i=1}^{n}(-a_i+b_i)^2 = n\cdot\frac{1}{n^2} = \frac{1}{n}$.

[6] Notar que $p(1-p)$ puede acotarse por $\frac{1}{4}$ y tendríamos una cota uniforme

$$P\left(\left|\sum_{i=1}^{n} X_i - np\right| \geq n\cdot\varepsilon\right) \leq \frac{p(1-p)}{\varepsilon^2 n} \leq \frac{1}{4\varepsilon^2 n}.$$

- Para Tchebyshev (es una cota puntual porque depende del valor del parámetro p): $P\left(\left|\sum_{i=1}^{n} X_i - np\right| \geq n \cdot \varepsilon\right) \leq \frac{\left(\sqrt{np(1-p)}\right)^2}{(n\varepsilon)^2} = \frac{p\,(1-p)}{n\varepsilon^2}$[7]

Veamos cuatro casos (tomando como referencia $\varepsilon = 1$):

- $n = 5$ y $p = 0{,}4$
  - Para Hoeffing: $P\left(\left|\sum_{i=1}^{5} X_i - 2\right| \geq 5\right) \leq 2e^{-10} \approx 0{,}000090799859525$
  - Para Tchebyshev: $P\left(\left|\sum_{i=1}^{5} X_i - 2\right| \geq 5\right) \leq= \frac{0{,}24}{125} = 0{,}00192$
- $n = 50$ y $p = 0{,}4$
  - Para Hoeffing: $P\left(\left|\sum_{i=1}^{50} X_i - 20\right| \geq 50\right) \leq 2e^{-100} \approx 3{,}720075976 \times 10^{-44}$
  - Para Tchebyshev: $P\left(\left|\sum_{i=1}^{50} X_i - 20\right| \geq 50\right) \leq= \frac{0{,}24}{125000} = 0{,}00000192$
- $n = 5$ y $p = 0{,}001$
  - Para Hoeffing: $P\left(\left|\sum_{i=1}^{5} X_i - 0{,}005\right| \geq 5\right) \leq 2e^{-10} \approx 0{,}000090799859525$
  - Para Tchebyshev: $P\left(\left|\sum_{i=1}^{5} X_i - 0{,}005\right| \geq 5\right) \leq= \frac{0{,}000999}{125} = 0{,}00007992$
- $n = 50$ y $p = 0{,}001$
  - Para Hoeffing: $P\left(\left|\sum_{i=1}^{50} X_i - 0{,}05\right| \geq 50\right) \leq 2e^{-100} \approx 3{,}720075976 \times 10^{-44}$
  - Para Tchebyshev: $P\left(\left|\sum_{i=1}^{50} X_i - 0{,}05\right| \geq 50\right) \leq= \frac{0{,}000999}{125000} \approx 0{,}000000007992 = 7{,}992 \times 10^{-9}$

■

21. Ejercicio 21

Basándonos en los datos del problema, aproximamos las probabilidades utilizando el TCL. Para eso, primero necesitamos calcular la media y la varianza de X_i para cada $i = 1, \cdots, 50$. $E(X_i) = \int_0^1 x \cdot 2x\,dx = \frac{2}{3}x^3\Big|_0^1 = \frac{2}{3}$.

Para poder calcular $Var(X_i)$, como $Var(X_i) = E(X_i^2) - (E(X_i))^2$, queda calcular $E(X^2)$.

$E(X_i^2) = \int_0^1 x^2 \cdot 2x\,dx = \frac{1}{2}x^4\Big|_0^1 = \frac{1}{2}$.

Por lo tanto, $Var(X) = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$.

[7] Notar que $p(1-p)$ puede acotarse por $\frac{1}{4}$ y tendríamos una cota uniforme $P\left(\left|\sum_{i=1}^{n} X_i - np\right| \geq n \cdot \varepsilon\right) \leq \frac{p(1-p)}{\varepsilon^2 n} \leq \frac{1}{4\varepsilon^2 n}$.

Como X_i tiene media $\mu = \frac{2}{3}$ y varianza $\sigma^2 = \frac{1}{18} < \infty$ para todo i por el Teorema Central del Límite podemos afirmar que $\sqrt{n}\left(\overline{X}_n - \frac{2}{3}\right) \overset{a}{\sim} N\left(0, \frac{1}{18}\right)$.

Entonces

$$
\begin{aligned}
P\left(S_{50} \leq 30\right) = P\left(\sum_{i=1}^{n} X_i \leq 30\right) &= P\left(\frac{\sum_{i=1}^{n} X_i}{50} \leq \frac{30}{50}\right) = P\left(\frac{\sum_{i=1}^{n} X_i}{50} - \frac{2}{3} \leq \frac{30}{50} - \frac{2}{3}\right) \\
&= P\left(\sqrt{50}\left(\frac{\sum_{i=1}^{n} X_i}{50} - \frac{2}{3}\right) \leq \sqrt{50}\left(\frac{30}{50} - \frac{2}{3}\right)\right) \\
&= P\left(\frac{1}{\sqrt{\frac{1}{18}}}\sqrt{50}\left(\frac{\sum_{i=1}^{n} X_i}{50} - \frac{2}{3}\right) \leq \underbrace{\frac{1}{\sqrt{\frac{1}{18}}}\sqrt{50}\left(\frac{30}{50} - \frac{2}{3}\right)}_{=-2}\right) \\
&= P\left(\frac{\sqrt{50}\left(\overline{X}_{50} - \frac{2}{3}\right)}{\sqrt{\frac{1}{18}}} \leq -2\right) \\
&\approx P(Z \leq -2) = \Phi(-2) \approx 0{,}02275
\end{aligned}
$$

Donde $Z \sim N(0,1)$, Φ es su función de distribución acumulada. Por TCL tenemos que $\frac{\sqrt{50}\left(\overline{X}_{50} - \frac{2}{3}\right)}{\sqrt{\frac{1}{18}}}$ tiene distribución aproximadamente $N(0,1)$.

Ahora bien, para calcular

$$
\begin{aligned}
P\left(\overline{X}_{50} \geq 0{,}7\right) & \\
&= P\left(\overline{X}_{50} - \frac{2}{3} \geq \underbrace{\frac{1}{30}}_{0{,}7 - \frac{2}{3}}\right) \\
&= P\left(\sqrt{50}\left(\overline{X}_{50} - \frac{2}{3}\right) \geq \frac{\sqrt{50}}{30}\right) \\
&= P\left(\frac{\sqrt{50}\left(\overline{X}_{50} - \frac{2}{3}\right)}{\frac{1}{\sqrt{18}}} \geq \frac{\sqrt{50}}{30} \cdot \frac{1}{\frac{1}{\sqrt{18}}}\right) \\
&= P\left(\frac{\sqrt{50}\left(\overline{X}_{50} - \frac{2}{3}\right)}{\frac{1}{\sqrt{18}}} \geq 1\right) \\
&\approx P(Z \geq 1) = 1 - \Phi(1) \approx 0{,}1586
\end{aligned}
$$

Finalmente, para el inciso 2), hallemos a de manera que $P(S_{50} \geq a) \approx 0{,}86$

$$
\begin{aligned}
0{,}86 \approx P(S_{50} \geq a) \\
&= P\left(\frac{S_{50}}{50} \geq \frac{a}{50}\right) \\
&= P\left(\frac{\sqrt{50}\left(\overline{X}_{50} - \frac{2}{3}\right)}{\frac{1}{\sqrt{18}}} \geq \left(\frac{a}{50} - \frac{2}{3}\right)\frac{\sqrt{50}}{\frac{1}{\sqrt{18}}}\right) \\
&\approx P\left(Z \geq \left(\frac{a}{50} - \frac{2}{3}\right)\frac{\sqrt{50}}{\frac{1}{\sqrt{18}}}\right)
\end{aligned}
$$

Por lo tanto, tenemos que

$$1 - \Phi\left(\left(\frac{a}{50} - \frac{2}{3}\right)\frac{\sqrt{50}}{\frac{1}{\sqrt{18}}}\right) \approx 0{,}86$$

$$\Phi\left(\left(\frac{a}{50} - \frac{2}{3}\right)\frac{\sqrt{50}}{\frac{1}{\sqrt{18}}}\right) \approx 0{,}14$$

Utilizando la función inversa de la función de distribución acumulada se tiene que:

$$\left(\frac{a}{50} - \frac{2}{3}\right)\frac{\sqrt{50}}{\frac{1}{\sqrt{18}}} \approx -1{,}080$$

Despejando a se obtiene que $a = \dfrac{473}{15} \approx 31{,}53$

■

22. Ejercicio 22

a)

$$P\left(\left|\overline{X}_{64} - \mu\right| < 1\right) = P\left(-1 < \overline{X}_{64} - \mu < 1\right) = P\left(-\frac{8}{5} < \frac{8}{5}\left(\overline{X}_{64} - \mu\right) < \frac{8}{5}\right)$$

Por el Teorema Central del Límite,

$$P\left(\left|\overline{X}_{64} - \mu\right| < 1\right) \approx \Phi\left(\frac{8}{5}\right) - \Phi\left(-\frac{8}{5}\right) = 0{,}89$$

b)

$$P\left(\left|\overline{X}_{n} - \mu\right| < 1\right) = 1 - P\left(\left|\overline{X}_{n} - \mu\right| \geq 1\right) \geq 1 - Var(\overline{X}_n) = 1 - \frac{25}{n} \geq 0{,}95$$

Despejando n se tiene que $n \geq 500$.

c)

$$P\left(\left|\overline{X}_{n} - \mu\right| < 1\right) = P\left(-\frac{\sqrt{n}}{5} < \frac{\sqrt{n}(\overline{X}_n - \mu)}{5} < \frac{\sqrt{n}}{5}\right)$$

Por el Teorema Central del Límite,

$$P\left(\left|\overline{X}_{n} - \mu\right| < 1\right) \approx \Phi\left(\frac{\sqrt{n}}{5}\right) - \Phi\left(-\frac{\sqrt{n}}{5}\right) = 2\Phi\left(\frac{\sqrt{n}}{5}\right) - 1 = 0{,}95$$

Despejamos,

$$\Phi\left(\frac{\sqrt{n}}{5}\right) = 0{,}975 \Rightarrow \frac{\sqrt{n}}{5} \approx 1{,}96 \Rightarrow \sqrt{n} \approx 9{,}8 \Rightarrow n \approx 96{,}04$$

Por lo tanto, para asegurarnos que la probabilidad $P\left(\left|\overline{X}_n - \mu\right| < 1\right)$ sea aproximadamente 0,95 (o un poco más) n debe ser mayor o igual a 96, notar que con $n = 96$ podemos decir que la probabilidad $P\left(\left|\overline{X}_n - \mu\right| < 1\right) \approx 0{,}95$ pero no podemos garantizar que sea mayor o igual a 0,95. Notar que en el item 3 utilizamos la distribución asintótica de $\overline{X}_n$, con lo cual es razonable que obtengamos un valor de n mucho más pequeño al hallado en 2, donde solamente se utiliza el valor de la media y de la varianza de $\overline{X}_n$.

23. Ejercicio 23

Como la función $\ln(\cdot)$ es estrictamente creciente en el intevalo $[1, +\infty)$ podemos afirmar que

$$P\left(\prod_{i=1}^{100} X_i > e^{55}\right) = P\left(\ln\left(\prod_{i=1}^{100} X_i\right) > 55\right)$$

Luego, utilizando propiedades del logaritmo

$$P\left(\prod_{i=1}^{100} X_i > e^{55}\right) = P\left(\sum_{i=1}^{100} \ln(X_i) > 55\right)$$

Para poder usar el Teorema Central del Límite calculamos $E[\ln(X_i)]$ y $Var[\ln(X_i)]$

$$E(\ln(X_i)) = \int_1^{+\infty} \ln(x)\frac{3}{X_i^4}dx$$

Aplicando integración por partes, donde consideramos $f(x) = \ln(x)$ y $g'(x) = \dfrac{1}{x^4}$

$$E(\ln(X_i)) = -\frac{\ln(x)}{x^3}\bigg|_1^{+\infty} + \int_1^{+\infty} \frac{1}{x^4}dx = \frac{1}{3}$$

Para calcular la varianza, necesitamos conocer:

$$E\left(\ln^2(X_i)\right) = \int_1^{+\infty} \ln^2(x)\frac{3}{x^4}dx$$

De vuelta, integración por partes, donde $f(x) = \ln^2(x)$ y $g'(x) = \dfrac{1}{x^4}$ tenemos que

$$E\left(\ln^2(X_i)\right) = -\frac{\ln^2(x)}{x^3}\bigg|_1^{+\infty} + \frac{2}{3}\int_1^{+\infty} \ln(x)\frac{3}{x^4}dx = 0 + \frac{2}{3}E(\ln(x)) = \frac{2}{9}$$

Con esto ya podemos calcular la $Var\left(\ln(X_i)\right)$

$$Var\left(\ln(X_i)\right) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

Por el Teorema Central del Límite, sabemos que

$$P\left(\left(\frac{\sum_{i=1}^{100}\ln(x)}{100} - \frac{1}{3}\right)\frac{\sqrt{100}}{\sqrt{\frac{1}{9}}} > \left(\frac{55}{100} - \frac{1}{3}\right)\frac{\sqrt{100}}{\sqrt{\frac{1}{9}}}\right) \approx 1 - \Phi(6{,}5) \approx 0$$

■

24. Ejercicio 24

Siendo Y el número total de errores tenemos que

$$P(Y < 90) = P\left(\sum_{i=1}^{100} X_i < 90\right)$$

Como $E[X] = 1$ y $Var[X] = 1$

$$P\left(\frac{\sum_{i=1}^{100} X_i}{100} < \frac{90}{100}\right) = P\left(\sqrt{100}\left(\frac{\sum_{i=1}^{100} X_i}{100} - 1\right) < \sqrt{100}\left(\frac{90}{100} - 1\right)\right)$$

Usando el TCL, podemos decir que

$$P\left(\sqrt{100}\left(\frac{\sum_{i=1}^{100} X_i}{100} - 1\right) < \sqrt{100}\left(\frac{90}{100} - 1\right)\right) \approx \Phi(-1) = 0{,}1586$$

■

25. Ejercicio 25

Tomando el ln de Y_n queda que

$$\ln(Y_n) = \frac{\sum_{i=1}^{n} \ln(X_i)}{n}$$

Luego, como las X_i son iid, al aplicar la LGN tenemos que

$$\ln(Y_n) = \frac{\sum_{i=1}^{n} \ln(X_i)}{n} \xrightarrow{P} E[\ln(X)]$$

Entonces, calculamos $E[\ln(X)]$

$$\begin{aligned} E[\ln(X)] &= \int_0^1 \ln(x) 2x dx \\ &= 2\int_0^1 \ln(x) x dx \end{aligned}$$

Aplicando integración por partes, donde $u = \ln(x)$ y $dv = x$, tenemos que

$$\begin{aligned} E[\ln(X)] &= 2\int_0^1 \ln(x) x dx \\ &= 2\left(\left[\ln(x)\frac{x^2}{2}\right]_0^1 - \int_0^1 \frac{x^2}{2x} dx\right) \\ &= 2\left(\left[\ln(x)\frac{x^2}{2}\right]_0^1 - \frac{1}{2}\left[\frac{x^2}{2}\right]_0^1\right) \\ &= 2\left(-\frac{1}{4}\right) \\ &= -\frac{1}{2} \end{aligned}$$

Entonces decimos que $\ln(Y_n) \xrightarrow{P} -\frac{1}{2}$, por lo tanto $Y_n \xrightarrow{P} e^{-\frac{1}{2}}$.

■

26. Ejercicio 26

Al tratarse de una muestra aleatoria, las X_i son i.i.d. Por la ley de grandes números, se tiene que:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} E(X_1) = 0$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} E\left(X_1^2\right) = 2$$

Por el teorema central del límite

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i = \sqrt{2}\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{\sqrt{2}} \xrightarrow{D} \sqrt{2}Z \sim \mathcal{N}(0,2)$$

Donde $Z \sim \mathcal{N}(0,1)$.

1. Tenemos que

$$V_n = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i}{\frac{1}{n}\sum_{i=1}^{n} X_i^2} \xrightarrow{P} \frac{0}{2} = 0$$

   Como la convergencia en probabilidad implica convergencia en distribución, concluimos que $V_n \xrightarrow{D} 0$

2. Tenemos:

$$W_n = \frac{\sqrt{n}\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i}{\frac{1}{n}\sum_{i=1}^{n} X_i^2} \xrightarrow{D} \frac{\sqrt{2}Z}{2} = \frac{Z}{\sqrt{2}} \sim \mathcal{N}(0,1/2)$$

   El numerador en la segunda igualdad, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{D} \sqrt{2}Z$, donde $Z \sim \mathcal{N}(0,1)$, es decir $\sqrt{2}Z \sim N(0,2)$. Por otro lado, el denominador converge en probabilidad a 2: $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} 2$, por lo que $\frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i^2} \xrightarrow{P} \frac{1}{2}$.

   Por lo tanto, usando el lema de Slutsky probamos el resultado que queríamos mostrar.

3. Tenemos:

$$Y_n = \frac{\sum_{i=1}^{n} X_i}{\sqrt{\sum_{i=1}^{n} X_i^2}} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}} \xrightarrow{D} \frac{\sqrt{2}Z}{\sqrt{2}} = Z \sim \mathcal{N}(0,1)$$

   El numerador en la segunda igualdad, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{D} \sqrt{2}Z$, donde $Z \sim \mathcal{N}(0,1)$, es decir $\sqrt{2}Z \sim N(0,2)$. Por otro lado, el denominador converge en probabilidad a $\sqrt{2}$: usando que $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} 2$ y que $t(x) = \sqrt{x}$ es una función continua, se tiene que $\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2} \xrightarrow{P} \sqrt{2}$ por lo que $\frac{1}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}} \xrightarrow{P} \frac{1}{\sqrt{2}}$.

   Por lo tanto, usando el lema de Slutsky probamos el resultado que queríamos mostrar.

4. Como la función exponencial $r(x) = e^x$ es una función continua y como $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{D} \sqrt{2}Z$ por el teorema central del límite, vale que:

$$Z_n = \exp\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right) \xrightarrow{D} \exp\left(\sqrt{2}Z\right)$$

Como $\sqrt{2}Z \sim \mathcal{N}(0,2)$, y vale que, como $d(x) = \ln(x)$ es una función continua $\ln(Z_n) \xrightarrow{D} \sqrt{2}Z \sim N(0,2)$ concluimos que $Z_n \xrightarrow{D} \exp\left(\sqrt{2}Z\right) \sim \text{lognormal}(0,2)$.

5. Utilizando el inciso anterior y que por la ley de los grandes números vale que $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} E(X_1) = 0$, podemos usar la propiedad 2) del tema de Slutsky, que dice que si $R_n \xrightarrow{D} R$ y $X_n \xrightarrow{P} c$ entonces $R_n + X_n \xrightarrow{D} R + c$.

$$R_n = \exp\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right) + \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{D} \exp\left(\sqrt{2}Z\right) + 0 = \exp\left(\sqrt{2}Z\right) \sim \text{lognormal}(0,2)$$

6. Usando que por la ley de los grandes números vale que $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} E\left(X_1^2\right) = 2$, que por el teorema central del límite vale que

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i = \sqrt{2}\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{\sqrt{2}} \xrightarrow{D} \sqrt{2}Z \sim \mathcal{N}(0,2),$$

y que por la propiedad 1) del lema de Slutsky vale que $D_n \xrightarrow{D} D$ y $U_n \xrightarrow{P} c$ entonces $D_n \cdot U_n \xrightarrow{D} c \cdot D$

$$S_n = \frac{1}{n^{3/2}}\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} X_i^2\right) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) \xrightarrow{D} \left(\sqrt{2}Z\right)2 \sim \mathcal{N}(0,8)$$

■

27. Ejercicio 27

Usamos la notación $X_n \xrightarrow{m.c.} X \Rightarrow X_n \xrightarrow{P} X$ para la convergencia en media cuadrática.

Para el inciso a) queremos ver que $X_n \xrightarrow{m.c.} X \Rightarrow X_n \xrightarrow{P} X$, es decir, que si vale que

$$\lim_{n\to\infty} E\left(|X_n - X|^2\right) = 0$$

Como las variables aleatorias $|X_n - X|^2$ y $|X_n - X|$ son no negativas, la única manera de que el límite del valor esperado de dichas variables aleatorias tienda a 0 es que la probabilidad de que tanto $|X_n - X|^2$ como $|X_n - X|$ sean positivas, -en particular, mayores que cualquier valor $\varepsilon > 0$ o $\varepsilon^2 > 0$- tiene que valer tender a 0.

$$\lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0$$

Para mostrar la implicación, usamos que

$$0 \leq P\left(|X_n - X| \geq \varepsilon\right) = P\left(|X_n - X|^2 \geq \varepsilon^2\right) \underbrace{\leq}_{\text{Markov}} \frac{E\left(|X_n - X|^2\right)}{\varepsilon^2}$$

Por lo tanto, se tiene que

$$0 \leq P\left(|X_n - X| \geq \varepsilon\right) \leq \frac{E\left(|X_n - X|^2\right)}{\varepsilon^2}$$

Tomando límite $\lim_{n\to\infty}$ se tiene que

$$0 \leq \lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) \leq \lim_{n\to\infty} \frac{E\left(|X_n - X|^2\right)}{\varepsilon^2} = 0$$

Por lo que $\lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0$ para cualquier valor de $\varepsilon > 0$ como queríamos ver.

■

Para el inciso b), para ver que $X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{m.c.} X$ tomamos las siguientes variables aleatorias dicotómicas (*di*=dos)

$$X_n = \begin{cases} 0 & \text{con probabilidad } 1 - \frac{1}{n} \\ \sqrt{n} & \text{con probabilidad } \frac{1}{n} \end{cases}$$

Notamos que $E(X_n) = 0 \cdot \left(1 - \frac{1}{n}\right) + \sqrt{n} \cdot \frac{1}{n} = \frac{1}{\sqrt{n}}$, o sea que $\lim_{n\to\infty} E(X_n) = 0$.

Además, $E\left(|X_n - 0|^2\right) = 0^2 \cdot \left(1 - \frac{1}{n}\right) + \sqrt{n}^2 \cdot \frac{1}{n} = 1$ para cualquier n.

Veamos que $X_n \xrightarrow{P} 0$ pero que sin embargo $X_n \not\xrightarrow{m.c.} 0$. [8]

Veamos primero que $\lim_{n\to\infty} P(|X_n - 0| \geq \varepsilon) = 0$ para cualquier $\varepsilon > 0$.

Dado $\varepsilon > 0$ existe n_0 de manera que $\sqrt{n_0} > \varepsilon$. Por lo tanto, **para todo** $n \geq n_0$ **se cumple que**:

$$\begin{aligned} P(|X_n| \geq \varepsilon) &= P(|X_n| \geq \sqrt{n}) \\ &= P(X_n \neq 0) = \frac{1}{n} \end{aligned}$$

Por lo tanto vale que

$$\lim_{n\to\infty} P(|X_n| \geq \varepsilon) = \lim_{n\to\infty} \frac{1}{n} = 0$$

Por lo tanto $X_n \xrightarrow{P} 0$. Ahora bien $E\left(|X_n - 0|^2\right) = 1$ para cualquier n, por lo que $\lim_{n\to\infty} E\left(|X_n - 0|^2\right) = 1 \neq 0$

Por lo tanto, se tiene que $X_n \not\xrightarrow{m.c.} X$.

■

[8] Estamos considerando que X es una v.a. degenerada, es decir que $P(X = 0) = 1$.