Ejercicios Resueltos TP7

Índice

1. Ejercicio 3

Enunciado

Una central telefónica A da servicio a 1800 usuarios de una central cercana B. Sería costoso y extravagante instalar 1800 líneas troncales de A a B. Es suficiente instalar un número N de líneas tan grande que, en condiciones ordinarias, solamente una de cada 100 llamadas en promedio no encuentre inmediatamente una línea troncal disponible. Supóngase que, en la hora más ocupada del día, cada usuario requiere una linea troncal de B durante un promedio de 2 minutos. En un momento fijo de la hora de máximo tráfico, puede suponerse la situación como un conjunto de 1800 ensayos independientes con una probabilidad $p = \frac{1}{30}$ (por lo de 2 min de cada 60 min) en que cada uno se requiere una línea. Determinar el número de líneas N a instalar entre A y B. Explicar claramente el planteo.

Resolución

Variables aleatorias

En este ejercicio, podemos pensar en la siguiente variable aleatoria, pensando en una línea por usuario:

$$S_{1800} = \text{cantidad de usuarios que requieren una línea en la hora más ocupada del día}$$

Datos

Particularmente, podríamos pensar a la variable S_{1800} como una binomial de parámetros 1800 y $p = \frac{1}{30}$. Sin embargo, para usar el Teorema central del límite, se puede pensar a la variable S_{1800} como una suma de variables independientes Bernoulli:

$$S_{1800} = \sum_{i=1}^{1800} X_i, \quad \text{donde } X_i = \begin{cases} 1 & \text{si el } i\text{-ésimo cliente requiere una línea} \\ 0 & \text{en caso contrario} \end{cases}$$

donde $E(X_i) = \frac{1}{30}$ y $\sigma(X_i) = \sqrt{\frac{1}{30} \cdot \frac{29}{30}} = \frac{\sqrt{29}}{30}$.

Por lo tanto, al ser una suma de variables aleatorias independientes idénticamente distribuidas, y al ser 1800 suficientemente grande, el teorema central del límite nos dice que S_{1800} se puede aproximar por una variable con distribución normal con los siguientes parámetros:

$$S_{1800} \overset{(a)}{\sim} \mathcal{N}\left(1800 \cdot E(X_i); \sqrt{1800} \cdot \sigma(X_i)\right) = \mathcal{N}\left(\frac{1800}{30}; \frac{\sqrt{29 \cdot 1800}}{30}\right) = \mathcal{N}\left(60; \sqrt{58}\right) \Rightarrow \frac{S_{1800} - 60}{\sqrt{58}} \overset{(a)}{\sim} \mathcal{N}\left(0; 1\right)$$

Cálculo de N

Se pide que en promedio uno de cada 100 llamados no encuentre una línea disponible. Esto se puede pensar como pedir que la probabilidad de sobrepasar las líneas disponibles sea $\frac{1}{100}$, es decir:

$$P(S_{1800} > N) = \frac{1}{100}$$

Sin embargo, como N debe ser entero, puede no cumplirse esta igualdad y por lo tanto, desde la perspectiva de la compañía el objetivo es **acotar superiormente** la probabilidad de que la demanda sobrepase las líneas disponibles, es decir:

$$P(S_{1800} > N) \leq \frac{1}{100}$$

Seguiremos asumiendo que la cantidad de líneas es suficientemente grande para poder usar TCL, y en caso de que el resultado final no sea consistente con esta hipótesis, usaremos la distribución binomial. Por lo tanto, debemos usar la corrección por continuidad:

$$P(S_{1800} > N) = P(S_{1800} > N + 0{,}5)$$

Vale aclarar que esa es una igualdad ya que el recorrido de la variable S_N está incluido en los números enteros, y que no se incluye un valor de probabilidad positiva al añadir 0.5. Ahora sí, podemos hacer la aproximación:

$$P(S_{1800} > N + 0{,}5) = P\left(\frac{S_{1800} - 60}{\sqrt{58}} > \frac{N + 0{,}5 - 60}{\sqrt{58}}\right) \overset{\text{TCL}}{\approx} 1 - \Phi\left(\frac{N - 59{,}5}{\sqrt{58}}\right)$$

Por lo tanto, podemos buscar N de forma que:

$$1 - \Phi\left(\frac{N - 59{,}5}{\sqrt{58}}\right) \leq 0{,}01 \Rightarrow 0{,}99 \leq \Phi\left(\frac{N - 59{,}5}{\sqrt{58}}\right) \Rightarrow z_{0{,}99} \leq \frac{N - 59{,}5}{\sqrt{58}}$$

Despejando,

$$z_{0{,}99} \cdot \sqrt{58} + 59{,}5 \leq N \Rightarrow 77{,}217 \leq N$$

Por lo tanto, como N debe ser entero, tomamos $N = 78$. Notar que es sensiblemente menor al valor inicial 1800, de forma que se reducen considerablemente los costos asociados a la instalación de líneas.

Verificación

Hoy en día, podemos usar software para verificar estas cuentas de forma exacta, utilizando la distribución binomial de S_{1800}. Por ejemplo, en este caso, podríamos verificar que:

$$P(S_{1800} > 78) = 0{,}0096038 \leq 0{,}01$$

Por lo tanto, se cumple lo pedido, pero más aún, vemos que **no** se cumple lo pedido para $N = 77$

$$P(S_{1800} > 77) = 0{,}013178 \not\leq 0{,}01$$

por lo que podemos inferir que la aproximación es suficientemente precisa.

Variables aleatorias

Datos

2. Ejercicio 14

Enunciado

Se desea estimar la probabilidad p de ocurrencia de cierto suceso a partir de la frecuencia relativa de ocurrencia medida al realizar n repeticiones del experimento en que puede ocurrir el suceso de interés. Se pone como condición que la probabilidad de que el error (diferencia entre p y la frecuencia relativa) no supere el valor 0.05, sea mayor que 0.97. ¿Cuántas pruebas deben realizarse como mínimo? Comparar los resultados obtenidos a partir de usar:

a) la desigualdad de Tchebycheff.

b) la aproximación normal de la distribución binomial.

Resolución

Variables aleatorias

El enunciado hace mención a la frecuencia relativa de ocurrencia del suceso de interés. La pregunta es cómo representarlo en términos de una variable aleatoria. Podríamos definir la frecuencia **absoluta** de ocurrencia del suceso del siguiente modo:

$$X_n = \text{ cantidad de veces que ocurre el suceso de interés en } n \text{ repeticiones del experimento}$$

Por lo tanto, la frecuencia **relativa** vendrá dada por el cociente entre la cantidad absoluta de ocurrencias (X_n) y la cantidad de repeticiones (n):

$$f_R = \frac{X_n}{n}$$

Datos

Al ser independientes las n repeticiones del experimento con la misma probabilidad p de éxito, $X_n \sim \text{Bi}(n, p)$. Por lo tanto, podemos inferir que $E(X_n) = n \cdot p$ y $V(X_n) = n \cdot p \cdot (1-p)$. A partir de estos datos, podemos calcular el valor esperado y varianza de la frecuencia relativa:

- $E(f_R) = E\left(\frac{X_n}{n}\right) = \frac{E(X_n)}{n} = \frac{\cancel{n} \cdot p}{\cancel{n}} = p$
- $V(f_R) = V\left(\frac{X_n}{n}\right) = \frac{V(X_n)}{n^2} = \frac{\cancel{n} \cdot p \cdot (1-p)}{n^{\cancel{2}}} = \frac{p \cdot (1-p)}{n}$

Item a

Recordemos la desigualdad de Tchebychev:

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$$

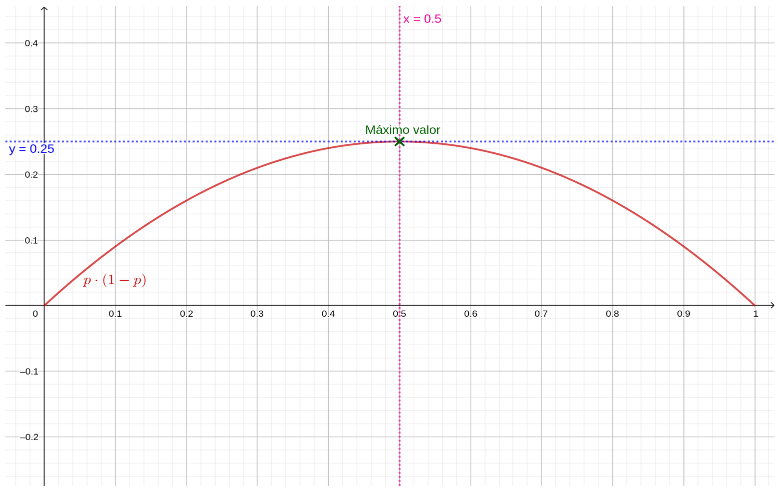
En este caso, nos piden n de forma que:

$$P(|f_R - p| \leq 0{,}05) > 0{,}97 \Leftrightarrow 1 - P(|f_R - p| > 0{,}05) > 0{,}97 \Leftrightarrow P(|f_R - p| > 0{,}05) < 0{,}03$$

Como $E(f_R) = p$, podemos aplicar la desigualdad de Tchebychev con $\varepsilon = 0{,}05$. Por otro lado, a medida que n se hace grande, la probabilidad de que $|f_R - p| = 0{,}05$ se achica (más aún, es cero si $n \cdot (p + 0{,}05)$ y $n \cdot (p - 0{,}05)$ no son enteros) y por lo tanto, podemos considerarlo despreciable:

$$P(|f_R - p| > 0{,}05) = P(|f_R - Ep| \geq 0{,}05) = P(|f_R - E(f_R)| \geq 0{,}05) =\leq \frac{V(f_R)}{0{,}05^2} = \frac{p \cdot (1-p)}{n \cdot 0{,}05^2}$$

Si nos aseguramos que este valor esté por debajo de 0.03, se cumple lo pedido. El problema es que no tenemos un valor para p. Por lo tanto, usaremos una cota superior de la expresión correspondiente a p para sólo lidiar con el tamaño muestral n. Como $p \in (0,1)$, $p(1-p)$ tiene un valor máximo de $\frac{1}{4}$ (cuando $p = \frac{1}{2}$):



Por lo tanto,

$$P(|f_R - p| > 0{,}05) \leq \frac{p \cdot (1-p)}{n \cdot 0{,}05^2} \leq \frac{\frac{1}{4}}{n \cdot 0{,}05^2} = \frac{1}{4n \cdot 0{,}05^2}$$

Es decir, si

$$\frac{1}{4n \cdot 0{,}05^2} < 0{,}03 \Rightarrow P(|f_R - p| > 0{,}05) \leq \frac{1}{4n \cdot 0{,}05^2} < 0{,}03 \Rightarrow P(|f_R - p| > 0{,}05) < 0{,}03$$

Por lo tanto, buscaremos n a partir de la primer inecuación, ya que en ese caso, se cumple lo pedido:

$$\frac{1}{4n \cdot 0{,}05^2} < 0{,}03 \Rightarrow \frac{1}{4 \cdot 0{,}03 \cdot 0{,}05^2} < n \Rightarrow n > 3333{,}3 \Rightarrow n \geq 3334$$

Item b

Como X_n es binomial, si n es suficientemente grande, a través del TCL podemos aproximar su distribución por una normal:

$$X_n \overset{(a)}{\sim} N\left(n \cdot p; \sqrt{n \cdot p \cdot (1-p)}\right) \Rightarrow f_R = \frac{X_n}{n} \overset{(a)}{\sim} N\left(p; \sqrt{\frac{p \cdot (1-p)}{n}}\right) \Rightarrow \frac{f_R - p}{\sqrt{\frac{p \cdot (1-p)}{n}}} \overset{(a)}{\sim} N(0,1)$$

Por lo tanto,

$$P(|f_R - p| \leq 0{,}05) = P(-0{,}05 \leq f_R - p \leq 0{,}05) = P\left(-\frac{0{,}05}{\sqrt{\frac{p\cdot(1-p)}{n}}} \leq \frac{f_R - p}{\sqrt{\frac{p\cdot(1-p)}{n}}} \leq \frac{0{,}05}{\sqrt{\frac{p\cdot(1-p)}{n}}}\right) \approx$$

$$\approx \Phi\left(\frac{0{,}05}{\sqrt{\frac{p\cdot(1-p)}{n}}}\right) - \Phi\left(-\frac{0{,}05}{\sqrt{\frac{p\cdot(1-p)}{n}}}\right) = \Phi\left(\frac{0{,}05\sqrt{n}}{\sqrt{p \cdot (1-p)}}\right) - \left(1 - \Phi\left(\frac{0{,}05\sqrt{n}}{\sqrt{p \cdot (1-p)}}\right)\right) = 2 \cdot \Phi\left(\frac{0{,}05\sqrt{n}}{\sqrt{p \cdot (1-p)}}\right) - 1$$

Luego, para que se cumpla lo pedido, podemos usar la aproximación:

$$P(|f_R - p| \leq 0{,}05) \approx 2 \cdot \Phi\left(\frac{0{,}05\sqrt{n}}{\sqrt{p \cdot (1-p)}}\right) - 1 > 0{,}97 \Leftrightarrow \Phi\left(\frac{0{,}05\sqrt{n}}{\sqrt{p \cdot (1-p)}}\right) > \frac{0{,}97 + 1}{2} = 0{,}985$$

$$\Leftrightarrow \frac{0{,}05\sqrt{n}}{\sqrt{p\cdot(1-p)}} > z_{0{,}985} \Leftrightarrow \sqrt{n} > \frac{z_{0{,}985}\cdot\sqrt{p\cdot(1-p)}}{0{,}05}$$

Nuevamente, como $p\cdot(1-p)\leq\frac{1}{4}$, si se cumple:

$$\sqrt{n} > \frac{z_{0{,}985}\cdot\sqrt{\frac{1}{4}}}{0{,}05} \geq \frac{z_{0{,}985}\cdot\sqrt{p\cdot(1-p)}}{0{,}05} \Rightarrow \sqrt{n} > \frac{z_{0{,}985}\cdot\sqrt{p\cdot(1-p)}}{0{,}05}$$

Por lo tanto, buscamos n de forma que:

$$\sqrt{n} > \frac{z_{0{,}985}\cdot\sqrt{\frac{1}{4}}}{0{,}05} \Rightarrow n > \left(\frac{z_{0{,}985}}{2\cdot 0{,}05}\right)^2 \approx 470{,}93 \Rightarrow n \geq 471$$

3. Ejercicio 17

Enunciado

Se toman muestras independientes de tamaño 10 y 15 de una variable aleatoria con distribución normal de media 20 y varianza 3. ¿Cuál es la probabilidad de que el promedio de las dos muestras se diferencie (en valor absoluto) en más de 0.3?

Resolución

Variables aleatorias

Las muestras independientes provienen de variables idénticamente distribuidas (podemos llamarlas D_i, E_i), es decir, podemos considerar:

$$D_i, E_i \sim N(20,\sqrt{3}),\quad \overline{X}_{10} = \sum_{i=1}^{10}\frac{D_i}{10},\quad \overline{Y}_{15} = \sum_{i=1}^{15}\frac{E_i}{15}$$

Datos

Al ser D_i y E_i normales independientes, sus promedios también son normales. Sólo resta determinar sus parámetros:

- $E(\overline{X}_{10}) = E\left(\sum_{i=1}^{10}\frac{D_i}{10}\right) = \sum_{i=1}^{10}\frac{E(D_i)}{10} = \sum_{i=1}^{10}\frac{20}{10} = \frac{10\cdot 20}{10} = 20$
- $V(\overline{X}_{10}) = V\left(\sum_{i=1}^{10}\frac{D_i}{10}\right) \overset{IND}{=} \sum_{i=1}^{10}V\left(\frac{D_i}{10}\right) = \sum_{i=1}^{10}\frac{V(D_i)}{10^2} = \sum_{i=1}^{10}\frac{3}{100} = \frac{10\cdot 3}{100} = \frac{3}{10} \Rightarrow \sigma(\overline{X}_{10}) = \sqrt{0{,}3}$

Por lo tanto $\overline{X}_{10} \sim N\left(20,\sqrt{\frac{3}{10}}\right)$. Por otro lado:

- $E(\overline{Y}_{15}) = E\left(\sum_{i=1}^{15}\frac{E_i}{10}\right) = \sum_{i=1}^{15}\frac{E(E_i)}{15} = \sum_{i=1}^{15}\frac{20}{15} = \frac{15\cdot 20}{15} = 20$
- $V(\overline{Y}_{15}) = V\left(\sum_{i=1}^{15}\frac{E_i}{15}\right) \overset{IND}{=} \sum_{i=1}^{15}V\left(\frac{E_i}{15}\right) = \sum_{i=1}^{15}\frac{V(E_i)}{15^2} = \sum_{i=1}^{15}\frac{3}{225} = \frac{15\cdot 3}{225} = \frac{1}{5} \Rightarrow \sigma(\overline{Y}_{15}) = \sqrt{0{,}2}$

Luego, $\overline{Y}_{15} \sim N\left(20, \sqrt{\frac{1}{5}}\right)$.

Nos piden la siguiente probabilidad:

$$P\left(\left|\overline{X}_{10} - \overline{Y}_{15}\right| > 0{,}3\right) = 1 - P\left(\left|\overline{X}_{10} - \overline{Y}_{15}\right| \leq 0{,}3\right) = 1 - P\left(-0{,}3 \leq \overline{X}_{10} - \overline{Y}_{15} \leq 0{,}3\right)$$

Por lo tanto, puede ser de utilidad saber la distribución de la resta entre $\overline{X}_{10}$ y $\overline{Y}_{15}$. Nuevamente, al ser independientes y normales, su resta es normal y para terminar de conocer su distribución, necesitamos calcular su media y desvío:

- $E\left(\overline{X}_{10} - \overline{Y}_{15}\right) = E\left(\overline{X}_{10}\right) - E\left(\overline{Y}_{15}\right) = 20 - 20 = 0$
- $V\left(\overline{X}_{10} - \overline{Y}_{15}\right) \overset{IND}{=} V\left(\overline{X}_{10}\right) + V\left(-\overline{Y}_{15}\right) = V\left(\overline{X}_{10}\right) + \cancel{(-1)^2} \cdot V\left(\overline{Y}_{15}\right) = \frac{3}{10} + \frac{1}{5} = \frac{1}{2} \Rightarrow \sigma(\overline{X}_{10} - \overline{Y}_{15}) = \sqrt{\frac{1}{2}}$

Por lo tanto, la probabilidad pedida es:

$$1 - P\left(-0{,}3 \leq \overline{X}_{10} - \overline{Y}_{15} \leq 0{,}3\right) = 1 - P\left(\frac{-0{,}3}{\sqrt{\frac{1}{2}}} \leq \frac{\overline{X}_{10} - \overline{Y}_{15}}{\sqrt{\frac{1}{2}}} \leq \frac{0{,}3}{\sqrt{\frac{1}{2}}}\right) =$$

$$= 1 - \left(\Phi\left(\frac{0{,}3}{\sqrt{\frac{1}{2}}}\right) - \Phi\left(-\frac{0{,}3}{\sqrt{\frac{1}{2}}}\right)\right) \overset{\text{simetría de }\Phi}{=} 1 - \left(\Phi\left(0{,}3 \cdot \sqrt{2}\right) - \left[1 - \Phi\left(0{,}3 \cdot \sqrt{2}\right)\right]\right) =$$

$$= 1 - \left(2 \cdot \Phi\left(0{,}3 \cdot \sqrt{2}\right) - 1\right) = 2 - 2 \cdot \Phi\left(0{,}3 \cdot \sqrt{2}\right) \approx 0{,}67137$$

Otra forma

Otra forma de hacerlo es sin tomar el complemento, es decir:

$$P\left(\left|\overline{X}_{10} - \overline{Y}_{15}\right| > 0{,}3\right) = P(\overline{X}_{10} - \overline{Y}_{15} < -0{,}3) + P(\overline{X}_{10} - \overline{Y}_{15} > 0{,}3)$$

Usando que la diferencia entre $\overline{X}_{10}$ y $\overline{Y}_{15}$ tiene media cero y desvío $\sqrt{\frac{1}{2}}$:

$$P(\overline{X}_{10} - \overline{Y}_{15} < -0{,}3) + P(\overline{X}_{10} - \overline{Y}_{15} > 0{,}3) = \Phi\left(-\frac{0{,}3}{\sqrt{\frac{1}{2}}}\right) + 1 - \Phi\left(\frac{0{,}3}{\sqrt{\frac{1}{2}}}\right) =$$

$$= 1 - \Phi\left(0{,}3 \cdot \sqrt{2}\right) + 1 - \Phi\left(0{,}3 \cdot \sqrt{2}\right) = 2 \cdot (1 - \Phi\left(0{,}3 \cdot \sqrt{2}\right)) \approx 0{,}67137$$