Probabilidades y Estadística (C)

Clase Práctica 15

Matías Data

Departamento de Matemática, FCEyN, Universidad de Buenos Aires

20/10/2020

Introducción

- Hoy veremos aplicaciones del Teorema Central del Límite (TCL).
- Este es uno de los resultados centrales de la teoría de probabilidades.
- Lo que nos dice es que si tenemos variables aleatorias i.i.d. $X_1, X_2, X_3, \dots$ con $\mathbb{E}(X_1) = \mu$ y $\mathbb{V}(X) = \sigma^2$ entonces:

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) = \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow[n\to+\infty]{\mathcal{D}} N(0,1)$$

- Formalmente lo que nos dice la convergencia en distribución es que:

$$P\left(a \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right) \underset{n\to+\infty}{\longrightarrow} P(a \leq Z \leq b) = \phi(b) - \phi(a)$$

para $Z \sim N(0,1)$ y $a, b \in \mathbb{R}$.

Introducción

- Más informalmente, lo que tenemos es que:

$$P\left(a \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right) \overset{\text{TCL}}{\approx} P(a \leq Z \leq b)$$

  para $Z \sim N(0,1)$ y n suficientemente grande.
- De hecho la convergencia que se tiene es uniforme, es decir no depende de a y b, pero en la práctica a partir de que n la aproximación es buena es otro tema, y depende de la distribución de la variable aleatoria.
- El teorema fue originalmente demostrado para el caso de X_i con distribución Bernoulli y $p = 1/2$ por De-Moivre, luego generalizado para todo p por Laplace. El teorema más general enunciado fue demostrado por Liapunov a principios del siglo XX, y luego generalizado aún más.

Ejercicio 1

Ejercicio

En el siglo XIX una compañia de seguros decide ofrecer una nueva póliza de seguros contra incendios. La empresa espera vender 10000 seguros y se sabe que el gasto anual promedio de una casa por incendios es de 10 dólares con un desvio estándar de 50 dólares. La empresa cobra una póliza anual de 10.5 dólares. La empresa tiene además un gasto anual en empleados de 3000 dólares. Si en un año a la empresa sus gastos exceden a sus fondos la empresa cae en bancarrota, por lo tanto la empresa tiene fondos por 6000 dólares para que en un año malo tener plata para pagar las obligaciones. Aproximar la probabilidad de que la empresa caiga en bancarrota en un año.

Ejercicio 1

- Si X_i es el gasto en la casa i-ésima por incendios al año, sabemos que $\mu = 10$ y $\sigma = 50$.
- Luego $S_{10000} = \sum_{i=1}^{10000} X_i$ es el gasto total anual por incendios de la empresa.
- Por otro lado la empresa tiene ingresos por 10.5 dólares por cada póliza al año, lo que da ingresos por 105.000 dólares.
- También tiene gastos por 3000 dólares en sueldos de empleados y tiene fondos por 6000 dólares, con lo cual los fondos totales que dispone en el año son 108.000 dólares.
- Si el gasto en incendios en un año supera este monto la empresa cae en bancarrota.

Ejercicio 1

- Con lo cual calculamos:

$$P(S_{10000} \geq 108000) =$$
$$P\left(\frac{S_{10000} - 10000\mu}{\sqrt{10000}\sigma} \geq \frac{108000 - 10000\mu}{\sqrt{10000}\sigma}\right) \overset{\text{TCL}}{\approx}$$
$$P(Z \geq 1.6) \approx 1 - 0.9452 = 0.0548$$

- Luego la probabilidad de caer en bancarrota es aproximadamente del 5%.

Ejercicio 2

Ejercicio

Para un examen final de Probabilidades y Estadística (C) los profesores deciden tomar en una modalidad Verdadero o Falso que consiste de 20 preguntas.

a) Si se desea que aproximadamente solo un 5% de los alumnos que no estudiaron nada (y por lo tanto responden al azar) aprueben, cuál debe ser el criterio de aprobación?

b) Si un alumno estudio la mitad de los contenidos, aproximar la probabilidad de que apruebe el examen.

Ejercicio 2

a) Sea S el número de respuestas correctas por un alumno que no estudio. Luego $S \sim \mathrm{Bi}(20, 1/2)$.

- Si llamamos c al criterio de aprobación, el número de respuestas correctas que necesita un alumno para aprobar, entonces queremos que $0.05 \approx P(S \geq c)$.
- Como S es discreta tomamos una corrección por continuidad, esto es, usamos que:

$$P(S = k) = P(k - \frac{1}{2} \leq S \leq k + \frac{1}{2})$$

- Luego, calculamos:

$$P(S \geq c) = P(S \geq c - \frac{1}{2}) =$$

$$P\left(\frac{S - 20(1/2)}{\sqrt{20(1/4)}} \geq \frac{c - 21(1/2)}{\sqrt{20(1/4)}}\right) \overset{TCL}{\approx} P\left(Z \geq \frac{c - 10.5}{\sqrt{5}}\right)$$

Ejercicio 2

- Luego queremos que:

$$P\left(Z \geq \frac{c - 10.5}{\sqrt{5}}\right) \approx 0.05$$

- Por lo tanto $\frac{c-10.5}{\sqrt{5}} = z_{0.95} = 1.645$, con lo cual $c = 10.5 + \sqrt{5}(1.645) \approx 14.17$.
- Luego si $c = 15$, es decir si pedimos 15 respuestas correctas como criterio de aprobación, garantizamos que aproximadamente el 95% de los que no estudiaron no aprueben.
- En realidad, con catorce ya nos alcanzaba, ya que la probabilidad exacta para $c = 14$ es ≈ 0.979 (si en lugar de tomar el primer entero más grande tomabamos el entero más cercano, estabamos mejor, aunque esto es menos conservador).

Ejercicio 2

B) Sea X la variable aleatoria que vale uno si el alumno responde correctamente y cero si no, y sea $Y = 1$ si conoce la respuesta a la pregunta.

- Tenemos que $Y \sim \mathrm{Be}(1/2)$, luego:

$$P(X=1) =$$
$$P(X=1 \mid Y=1)\,P(Y=1) + P(X=1 \mid Y=0)\,P(Y=0) =$$
$$1(1/2) + (1/2)(1/2) = 0.75$$

- Luego si S es el número de respuestas correctas de este alumno, entonces $S \sim \mathrm{Bi}(20, 0.75)$.
- Calculamos entonces:

$$P(S \geq 15) = P(S \geq 14.5) =$$
$$P\left(\frac{S - 20(0.75)}{\sqrt{20(0.75)(0.25)}} \geq \frac{14.5 - 15}{\sqrt{15/4}}\right) \overset{TCL}{\approx} P\left(Z \geq 0.26\right) \approx 0.60$$

(la probabilidad exacta es aproximadamente 0.585).

Ejercicio 3

Una panadería vende un pan casero cuyo peso es una variable aleatoria con media $\mu = 1$ kg y desvío estándar $\sigma = 0.1$ kg.

Ejercicio

a) A la panaderia le llega un pedido de 100 kg. El panadero quiere estar bastante seguro de que al enviar una cantidad n de panes, se garantizará cumplir con los 100 kg. Usar el TCL para encontrar un n de forma tal que la probabilidad de que su entrega de panes supere los 100 kg con probabilidad aproximadamente 0.9.

b) Luego de una crisis, un cliente acusa que la panadería redujo el peso de los panes, para esto compro 50 panes en el transcurso de varios días y los fue pesando, y obtuvo un peso promedio de 950 gramos. Qué tan probable es obtener un peso promedio de 950 gramos o menos si los panes son de peso medio 1 kg y desvío estándar $\sigma = 0.1$?

Ejercicio 3

A) Llamemos $X_1, \ldots, X_n$ a las variables aleatorias dadas por el peso en kilogramos de cada pan.

- Sabemos que $\mu = 1$ y $\sigma = 0.1$, y además son independientes.
- Llamemos $S_n = \sum_{i=1}^{n} X_i$ a la suma de los pesos de los N panes a entregar.
- Queremos encontrar un n de forma tal de garantizarnos que este peso sea mayor o igual a 100 kg con probabilidad aproximadamente 0.9, es decir:

$$P(S_n \geq 100) \approx 0.9$$

- Pero para n grande, podemos usar el Teorema Central del Límite. Este nos dice que:

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) = \frac{S_n - n\mu}{\sqrt{n}\sigma} \approx N(0,1)$$

Ejercicio 3

- Luego, usando el TCL podemos aproximar la probabilidad deseada:

$$P(S_n \geq 100) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \geq \frac{100 - n\mu}{\sqrt{n}\sigma}\right) \stackrel{\text{TCL}}{\approx} P\left(Z \geq \frac{100 - n\mu}{\sqrt{n}\sigma}\right) \stackrel{\text{quiero}}{=} 0.9$$

- Para esto necesito que la probabilidad de caer por abajo de $z = \frac{100 - n\mu}{\sqrt{n}\sigma}$ sea 0.1, luego por simetría $z = -z_{0.9}$, es decir es menos el percentil 0.9.

Ejercicio 3

- Pero entonces $\frac{100-n\mu}{\sqrt{n}\sigma} = -z_{0.9} = -1.285$.
- Si llamamos $x = \sqrt{n}$ y reemplazamos $\mu = 1, \sigma = 0.1$ entonces tenemos que resolver la cuadrática:

$$x^2 - 0.1285x - 100 = 0$$

  cuya única solución positiva es $\sqrt{n} = x \approx 10.06446$.
- Por lo tanto si tomamos $n = 102$, nos podemos garantizar una probabilidad aproximada de 0.9 de que superaremos los 100 kg.

b) Ahora queremos calcular la probabilidad de que al comprar 50 panes, el peso promedio obtenido sea menor o igual a 950 gramos.

- Nuevamente, podemos utilizar el TCL para el promedio ahora.

Ejercicio 3

- Calculamos:

$$P(\overline{X}_{50} \leq 0.95) = P\left(\sqrt{50}\left(\frac{\overline{X}_{50} - \mu}{\sigma}\right) \leq \sqrt{50}\left(\frac{0.95 - \mu}{\sigma}\right)\right)$$

$$\overset{\text{TCL}}{\approx} P\left(Z \leq \frac{-\sqrt{50}\,(0.05)}{0.1}\right) = P(Z \leq -3.53) \approx 0.0002$$

- Concluimos que la probabilidad de obtener un peso medio de 950 gramos en 50 panes tiene una probabilidad aproximada de 1 en 5000, con lo cual es muy poco probable observar tal muestra. Esto nos da una evidencia estadística en contra de que el peso medio prometido por el panadero de un kg no es cierto.