Tarea 13 de Modelos Probabilistas Aplicados

Teorema de límite central

5271

8 de diciembre de 2020

1. Introducción

En este documento se presenta las nociones básicas sobre el teorema del limite central, ejemplos y aplicaciones.

2. Teorema del limite central

El teorema del límite central proporciona una aproximación al comportamiento de las sumas de variables aleatorias. El teorema establece que a medida que aumenta el número de variables aleatorias independientes e idénticamente distribuidas con media finita y varianza finita, la distribución de su suma se vuelve cada vez más normal independientemente de la forma de distribución de las variables aleatorias. Es decir, sea $X_1, X_2, ..., X_n$ una secuencia de variables aleatorias mutuamente independientes e idénticamente distribuidas, cada una de las cuales tiene una media finita μ_x y una varianza finita σ_x^2. Sea S_n definido como sigue:

$$S_n = X_1 + X_2 + \ldots + X_n. \tag{1}$$

Ahora, $\mathbb{E}[S_n] = n\mu_x$ y $\sigma_{S_n}^2 = n\sigma_x^2$. Al convertir S_n en una variable aleatoria normal estándar (es decir, media cero y varianza unitaria), se obtiene.

$$Y_n = \frac{S_n - \mathbb{E}[S_n]}{\sigma_{S_n}^2} = \frac{S_n - n\mu_X}{\sigma_X \sqrt{n}}. \tag{2}$$

El teorema del límite central establece que si $F_{Y_n}(y)$ es la función de densidad de γ_n , entonces:

$$\lim_{n\to\infty} F_{Y_n}(y) = \lim_{n\to\infty} \mathbb{P}[Y_n \leq y] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = \Phi(y); \tag{3}$$

Esto significa que $\lim_{n\to\infty} \gamma_n$ se distribuye $\sim N(0,1)$. Por tanto, uno de los roles importantes que juega la distribución normal en estadística es su utilidad como aproximación de otras funciones de distribución de probabilidad. Lo anterior se apoya en el teorema 9.5 del libro "*Introduction to Probability*"[1]:

Para una mejor comprensión de lo anteriormente expuesto se tomará como ejemplo la resolución del ejercicio 11 de la página 355 del mismo libro, que dice:

A tourist in Las Vegas was attracted by a certain gambling game in which the customer stakes 1 dollar on each play; a win then pays the customer 2 dollars plus the return of her stake, although a loss costs her only her stake.Las Vegas insiders, and alert students of probability theory, know that the probability of winning at this game is 1/4. When driven from the tables by hunger, the tourist had played this game 240 times. Assuming that no near miracles happened, about how much poorer was the tourist upon leaving the casino? What is the probability that she lost no money?

Para resolver este ejercicio se realiza el programa 2 en lenguaje R [2] que se muestra a continuación.

```
# Assuming that no near miracles happened, about how much poorer was the tourist upon
    leaving the casino?
valor_e = 0.25*2 + 0.75*(-1)
valor_e_perder = valor_e*240
valor_e_perder

# What is the probability that she lost no money?
n <- 240
p <- 0.25
q <- 1-p
sd <- sqrt(n*p*q)
mean <- -60
normal=as.data.frame(rnorm(n, mean, sd))

png(filename = "ejercicio11.png",width = 2000, height = 1600, res =200)
ggplot(normal, aes(x=`rnorm(n, mean, sd)`)) + scale_x_continuous(breaks=seq(-80, -30,
    5)) + geom_density(alpha=.2, fill="#FF6699")+ theme_bw()+ xlab("Valores simulados")
    +
  ylab("Densididad ")+theme(axis.text = element_text(size = 14))+theme(axis.title =
    element_text(size = 18))
dev.off()
```

Tarea14.R

Como respuesta a la primera pregunta se tiene que $\mathbb{E}[S_n] = n\mu_x = -60$, por lo que si el turista juega 240 veces se espera que pierda 60 dolares. Para la segunda respuesta tenemos la figura 1 de la página 3, que muestra una simulación de 240 veces jugadas con una distribución normal con media $\mu = -60$ y $\sigma^2 = 6{,}708$. en la misma se puede observar que existe cero probabilidad que el jugador gane algún dolar en un total de 240 juegos.

3. Aplicación en prueba de hipótesis

En esta sección un ejemplo de como se utiliza el teorema del limite central para probar hipótesis. Comúnmente para esta fin se utiliza una prueba de χ^2 para determinar si rechazar una distribución de población hipotética (con un número finito de clases)como falso. Aquí se hará esto para cuando la población se descomponga en dos clases (fumadores y no fumadores).

Para el ejemplo plantearemos la hipótesis H_0 que nos dice que el 20 % de los jóvenes en Monterrey fuman, para probar esta hipótesis tomamos datos de una encuesta realizada a 1,888 estudiantes en Monterrey [3] la cual arroja que 18.7 % de los encuestados eran fumadores al momento de la realización de la encuesta.

Sea $X_i = 1$ si el encuestado i dice que es fumador, y sea $X_i = 0$ si el encuestado i dice que no es fumador. Estas X_i son variables aleatorias de Bernoulli independientes. Tenemos que $S_n = X1 + + X_1, 888$. Si la hipótesis de que el 20 % de los jóvenes de Monterrey fuman es correcta, entonces $\mu = \mathbb{E}[X_i] = 0{,}2$,

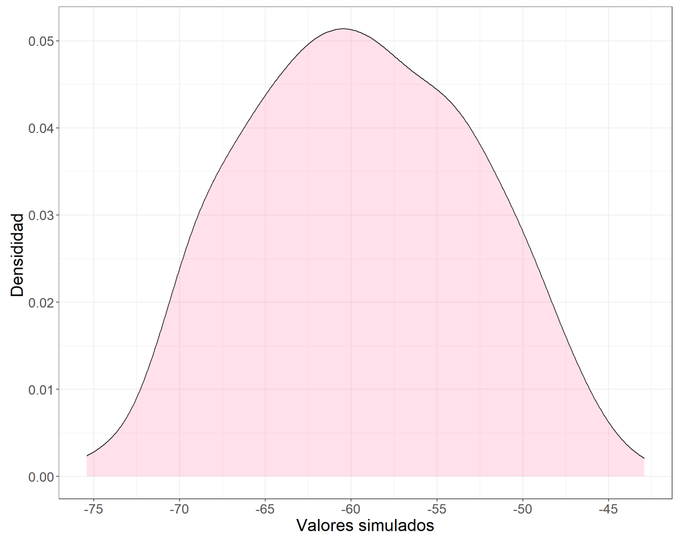

Figura 1: Diagrama de densidad de los resultados para 100 repeticiones del experimento.

y $X_i = (1 -) = 0{,}16$; y entonces, el teorema del límite central nos diría que:

$$\frac{S_{1888} - 378}{\sqrt{1,888 \cdot 0{,}16}}, \tag{4}$$

en el sentido que

$$\mathbb{P}\left(\frac{S_{1888} - 378}{6{,}952} \leq y\right) \approx \Phi(y). \tag{5}$$

Ahora, si $S^*_{1,888}$ es el valor observado $S^1_{1,888}$, y si

$$\gamma = \frac{S_{1888} - 378}{\sqrt{6{,}952}}, \tag{6}$$

entonces, sobre la base del teorema del limite central y la ecuación eq:1 se esperaría que γ es un valor atípico para $\sim N(0,1)$. Específicamente, no se esperaría que γ demasiado grande; es decir no se esperaría que:

$$\mathbb{P}(|N(0,1)||\gamma|) < 0{,}05, \tag{7}$$

si $\mu = 0{,}2$ es la media verdadera. Así se tiene la siguiente prueba estadística: Se fija un $\alpha > 0$, comúnmente se emplea $\alpha = 0{,}05$. Calculando γ como en la ecuación 6, si

$$\mathbb{P}(|N(0,1)||\gamma|) = 2\Phi(-|\gamma|) < \alpha, \tag{8}$$

rechazamos la hipótesis de que el valor medio de X_i es μ, y si esta desigualdad no se satisface, no la rechazamos.

Por lo que se tiene que:

$$\gamma = \frac{353 - 378}{6{,}952} = -3{,}596, \tag{9}$$

y calculando vemos clara mente que

$$2\Phi(-3{,}596) = 2(0{,}49983) = 0{,}9996 > 0{,}05. \tag{10}$$

Por lo tanto no hay suficientes elementos para rechazar H_o, que plantea que el 20 % de los jóvenes en Monterrey fuman. H_0 se acepta con un intervalo de confianza del 95 %.

Referencias

[1] Grinstead, Charles M., Snell, J. Laurie. *Introduction to Probability.* American Mathematical Society, 2006.

[2] R Core Team. R: Un lenguaje y un entorno para la informática estadística, 2020.

[3] Reynales-Shigematsu LM, Valdés-Salgado R, Rodríguez-Bolaños R, Lazcano-Ponce E, Hernández-Ávila M. *Encuesta de Tabaquismo en Jóvenes en México.Análisis descriptivo 2003, 2005, 2006, 2008.* nstituto Nacional de Salud Pública, 2009.