Ejercicios de Probabilidad y Simulación

Tema 3: Variables aleatorias y distribuciones
Juan Cantero Jimenez
2 de diciembre, 2021

2. *Si representamos por μ y σ^2 , respectivamente, la esperanza y varianza de una variable aleatoria X, expresa en términos de estos dos momentos:*

a $E(2X-3) y V(2X-3)$.
b $V(5-X)$.
c $E((X-2)(X+1))$.

Para resolver este ejercicio simplemente se aplican las propiedades de la esperanza y la varianza.

a $E(2X-3) = 2 * E(X) - 3 = 2 * \mu - 3$
$V(2X-3) = 2^2 * V(X) = 4 * \sigma^2$

b $V(5-X) = (-1)^2 * V(X) = V(X) = \sigma^2$

c $E((X-2)(X+1)) = E(X^2 - X - 2) = E(X^2) - E(X) - 2 = E(X^2) - \mu - 2 = V(X) + \mu^2 + \mu - 2 = \sigma^2 + \mu^2 - \mu - 2$

19. *Una compañía de logıstica envıa paquetes de distintos pesos, con una media de 15 kg y una desviación tıpica de 10. Teniendo en cuenta que los paquetes provienen de una gran cantidad de clientes diferentes, es razonable modelizar sus pesos como variables aleatorias independientes. Calcular la probabilidad de que el peso total de 100 paquetes exceda de 1700 kg.*

Sabemos que:

$$X : \textit{pesos de paquetes enviados por compañía logística}$$
$$E(X) = \mu = 15, V(X) = \sigma^2 = 10^2$$

Estamos interesados en modelizar el peso total de 100 paquetes, que será representado por la variable U, así:

$$U : \textit{pesos de } 100 \textit{ paquetes enviados por compañia logística}$$
$$E(U) = E(X_1) + E(X_2) \ ... \ E(X_{100}) = E(X) * 100 = \mu_X * 100 = 1500$$
$$V(U) = V(x_1) + V(X_2) \ ... \ V(X_{100}) = V(X) * 100 = \sigma^2 * 100; \ \sigma_U = \sigma_X * \sqrt{100} = 100$$

Por el teorema central del limite, TCL, podemos asumir que la distribución de la variable U, es normal con media $E(U)$ y desviación típica $V(U)$. Así la probabilidad de que el peso total de 100 paquetes exceda de 1700 kg se puede obtener como:

$$P(U > 1700) = 1 - P(U < 1700)$$

En R esta probabilidad puede calcularse como:

```
n = 100
mu = 15
sigma = 10
mu_u = 15*n
sigma_u = sqrt(n)*sigma
1-pnorm(1700, mu_u,sigma_u)

## [1] 0.02275013
```

Así la probabilidad $P(U > 1700)$ es igual a 0.02275013.
Notese que debido al TCL, la distribución original de X, no afecta al resultado. A continuación se muestra la simulación realizada con una distribución normal y una gamma:

```
#Distribución normal
N <- 10000
data <- matrix(rnorm(N*n, mu, sigma), ncol=n)
data_suma <- apply(data, 1, sum)
mean(data_suma > 1700)

## [1] 0.0224

#Distribución gamma
alpha = mu^2 / (sigma)^2
beta = alpha/mu
N <- 10000
data <- matrix(rgamma(N*n, alpha, beta), ncol=n)
data_suma <- apply(data, 1, sum)
mean(data_suma > 1700)

## [1] 0.0266
```

26.*Sea X una variable aleatoria Normal con media 2 y desviación tıpica 0.1. Sea $Y = exp(X)$, calcula por Monte Carlo, utilizando una muestra de tamaño $N = 10000$, la media y la desviación tıpica de Y . ¿Qué tamaño N serıa necesario para que el intervalo de confianza al 95 % sobre la media tuviera una longitud inferior a una centésima?*

La media y la desviación típica se pueden obtener mediante simulación:

```
set.seed(123)
N <- 10000
x <- rnorm(N, 2, 0.1)
y <- exp(x)
ey = mean(y)
ey

## [1] 7.424239

y2 <- y^2
ey2 <- mean(y2)
```

```r
vy <- ey2-ey^2
vy

## [1] 0.5528891

sdy <- sqrt(vy)
sdy

## [1] 0.7435651

ic_ey <- ey + c(-1, 1)*qt(0.975,N-1)*(sdy/sqrt(N))
ic_ey

## [1] 7.409663 7.438814

ic_vy <- ic_vy <- c( ((N-1)*vy)/qchisq(0.975, N-1), ((N-1)*vy)/qchisq(0.025, N-1) )
ic_vy

## [1] 0.5378778 0.5685408
```

Así obtenemos una media de 7.424239 con un intervalo de confianza al 95% de (7.409663, 7.438814) y una varianza de 0.5528891 con un intervalo de confianza al 95 % de (0.5378778, 0.5685408). De forma análoga la desviación típica obtenida es de 0.7435651.
Para obtener el tamaño de N tal que la longitud del intervalo de confianza sea menor que 0.01 debemos despejar N en:

$$0.01 = 4\frac{s_x}{\sqrt{N}}$$
$$N = (4\frac{s_x}{0.01})^2$$

En R este valor se puede obtener:

```r
ncent <- (4*(sdy/0.01))^2
ncent

## [1] 88462.25
```

Así obtenemos un tamaño de N = 88462. Tamaños muéstrales superiores a N generarán un IC al 95% de tamaño inferior a 0.01
Para comprobar este resultado se pueden simular 10000 medias y comprobar que el 95 % de ellas estén en el intervalo de longitud 0.01 centrado en la media de y.

```r
set.seed(123)
result = replicate(1000, {
N <- ncent
x <- rnorm(N, 2, 0.1)
y <- exp(x)
media <- mean(y)

ey-0.005 < media & media < ey+0.005
})
mean(result)

## [1] 0.892
```

Como se puede observar en el 95% de los experimentos la media se encuentra en el intervalo de longitud 0.01 centrado en la media. Debido a las limitaciones de LaTeX, Overleaf, y R ejecutado en estos, me es imposible simular una muestra de tamaño N = 10000, pero como se puede observar el tamaño es bastante próximo a 0.95 para un N = 1000.