Soluciones a ejercicios seleccionados del Práctico 4.

El siguiente texto es un borrador y está sujeto a correcciones en clase.

Ejercicio 1

El tiempo de vida de cierto componente electrónico es una variable aleatoria Gaussiana con una esperanza de 5000 horas y un desvío estándar de 100 horas. ¿Cuál es la probabilidad que el promedio de vida de 4000 componentes sea menor que 5002 horas?

Solución. Consideremos

$$\sum_{i=1}^{4000} X_i \sim N(4000 \cdot 5000, 4000 \cdot 100^2)$$

donde cada X_i representa el i-ésimo tiempo de vida de un componente electrónico. El promedio de vida es la varible

$$\frac{1}{4000} \sum_{i=1}^{4000} X_i \sim N(5000, \frac{100^2}{4000}) = N(5000, 2.5)$$

Luego,

$$P(\frac{1}{4000} \sum_{i=1}^{4000} X_i < 5012) = P(T < 4000 \cdot 5002)$$

es

```
pnorm(5002, mean=5000, sd=sqrt(2.5))
0.8970484
```

Ejercicio 2

La compilación de un programa de computadora consta de 3 bloques que se procesan de manera secuencial, uno tras otro. Cada bloque lleva un tiempo con distribución exponencial con una media de 5 minutos, y es independiente de los otros bloques.

a) Hallar la esperanza y la varianza del tiempo total de compilación.

b) Calcular la probabilidad de que el programa entero sea compilado en menos de 12 minutos.

Solución.

El tiempo total T es la suma de tres va's independientes con distribución exponencial con $\lambda = 1/5$. Luego, por el Corolario 1,

$$T \sim \text{gamma}(3, 1/5)$$

Así $\mathrm{E}(T) = \alpha/\lambda = 3 \cdot 5 = 15$ minutos y $\mathrm{VAR}(T) = \alpha/\lambda^2 = 3 \cdot 25$ minutos^2.

Y, finalmente

```
pgamma(12, shape=3, scale= (1/(1/5)))
0.4302913
```

Ejercicio 3

La actualización de un determinado paquete de software requiere la instalación de 68 archivos nuevos. Los archivos están instalados consecutivamente. El tiempo de instalación es aleatorio, se puede asumir que tiene distribución normal y, en promedio tarda 15 segundos en instalarse un archivo, con una varianza de 11 seg ^{2}.

a) ¿Cuál es la probabilidad de que todo el paquete se actualice en menos de 12 minutos?

b) Se lanza una nueva versión del paquete. Requiere solo N nuevos archivos para ser instalado, y se afirma que el 95 % de las veces el tiempo de actualización lleva menos de 10 minutos. Dada ésta información, hallar N.

Solución. Por la independencia, la idéntica distribución y la normalidad de cada variable, resulta

$$\sum_{i=1}^{68} X_i \sim N(68 \cdot 15, 68 \cdot 11).$$

La probabilidad buscada en a) es:

```
pnorm(12*60, 68*15, sd=sqrt(68*11))
2.69067e-28
```

Es decir, es aproximadamente cero.

Para b) tenemos que resolver la siguiente ecuación en probabilidad

$$P(Z < \frac{10 \cdot 60 - n \cdot 15}{\sqrt{n \cdot 11}}) = 0.95$$

donde $Z \sim N(0, 1)$.

Como

```
qnorm(0.95)
1.644854
```

entonces el N buscado debe satisfacer la ecuación

$$(\sqrt{N \cdot 11} \cdot 1.64)^2 = (10 \cdot 60 - N \cdot 15)^2$$

ó, equivalentemente,

$$225N^2 - 18030N + 600^2 = 0$$

De aquí se deduce que $N = 38$.

Para obtener esa solución, las raíces de una ecuación cuadrática se puede hallar con la siguiente función en R

```
raices = function(a,b,c){
if (b^2 - 4*a*c >= 0) {
raiz1 = (-b + sqrt(b^2 - 4*a*c))/(2*a)
raiz2 = (-b - sqrt(b^2 - 4*a*c))/(2*a)
}
else {
raiz1 = (-b + sqrt(as.complex(b^2 - 4*a*c)))/(2*a)
raiz2 = (-b - sqrt(as.complex(b^2 - 4*a*c)))/(2*a)
}
c(raiz1, raiz2)
}

> raices(225,-18030,360000)
[1] 42.37703 37.75630
```

y en Python el código es el siguiente

```
import math
import cmath

def raices(a,b,c):
if b**2-4*a*c >= 0: #indentación simple
raiz1 = (-b + math.sqrt(b**2 - 4*a*c))/(2*a) #doble indentación
raiz2 = (-b - math.sqrt(b**2 - 4*a*c))/(2*a) #doble indentación
else: #indentación simple
raiz1 = (-b + cmath.sqrt(complex(b**2 - 4*a*c)))/(2*a) #doble indentación
raiz2 = (-b - cmath.sqrt(complex(b**2 - 4*a*c)))/(2*a) #doble indentación
return [raiz1, raiz2] #indentación simple

  raices(225,-18030,360000)
  [42.377029793488454, 37.75630353984488]
```

Ejercicio 4

Entre todos los chips de computadora producidos por una determinada fábrica, el 6 % son defectuosos. Se selecciona una muestra de 400 chips para su inspección y se asume que el resultado de cada inspección es independiente del otro.

a) Cuál es la probabilidad de que esta muestra contenga entre 20 y 25 chips defectuosos? (incluir los valores 20 y 25).

b) Comparar la probabilidad exacta con la aproximación que provee el teorema central del límite.

Solución.

a) Consideremos la variable X: "número de chips defectuosos en un total de 400". Luego,

$$X \sim binomial(400, 0.06)$$

y así $P(20 \leq X \leq 25)$ es computada en R como

```
pbinom(25, 400, 0.06)-pbinom(19, 400, 0.06)
0.4621308
```

b) Para la aproximación utilizando la distribución normal, consideramos que

$$X = S_{400} = \sum_{i=1}^{400} X_i$$

donde cada X_i tiene una distribución Bernoulli de parámetro $p = 0.06$, y X_i vale 1 o 0 de acuerdo a si el i-ésimo chip es defectuoso o no.

Luego, por el teorema central del límite:

$$\frac{S_{400} - \mathrm{E}(S_{400})}{\sqrt{\mathrm{VAR} S_{400}}} \text{ tiene una distribución aproximadamente } N(0,1). \qquad (1)$$

Por otro lado,

$$\mathrm{E}(S_{400}) = 400 \cdot 0.06 = 24 \text{ y } \mathrm{VAR}(S_{400}) = 400 \cdot 0.06 \cdot (1 - 0.06) = 22.56,$$

Con lo cual, podemos reescribir (1) del siguiente modo:

$$S_{400} \text{ tiene una distribución aproximadamente } N(24, 22.56). \qquad (2)$$

Así $P(20 \leq X \leq 25)$ puede ser aproximada por:

```
pnorm(25,24,sqrt(22.56))-pnorm(20,24,sqrt(22.56))
0.3835246
```

Observamos que la aproximación no es buena.

En cambio si, como explicamos en clase, utilizamos la corrección

```
pnorm(25.5,24,sqrt(22.56))-pnorm(19.5,24,sqrt(22.56))
0.4522133
```

la aproximación mejora notablemente.

Ejercicio 5

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una variable aleatoria $X \sim$Bernoulli(p), $S_n = \sum_{i=1}^{n} X_i$ y considere $Z_n = \frac{S_n - \mathrm{E}(S_n)}{\sqrt{\mathrm{VAR}(S_n)}}$ la suma estandarizada, como en el enunciado del teorema central del límite. Definamos

$$\forall x : F_{n,p}(x) = P(Z_n \leq x)$$

la función de distribución de Z_n.

a) Utilizando R definir una función `fdistBer` que dependa de x, n y p y que compute $F_{n,p}(x)$.

b) Utilizando una secuencia de valores, por ejemplo

```
x<- seq(-4,4,0.01)
```

estudiar la aproximación provista por el teorema central del límite comparando las gráficas de `fdistBer`(x,n,p) y de $\Phi(x)$, para diferentes valores de p y n.