Complementos

Sesión 7

Natalie Julian - www.nataliejulian.com

Estadística UC y Data Scientist en Zippedi Inc.

Práctica 1

a) Sea X la variable aleatoria con la siguiente función de probabilidad:

$$P(X = x) = \begin{cases} 0.69 & si \quad x = 1 \\ 0.076 & si \quad x = 2 \\ 0.073 & si \quad x = 5 \\ 0.161 & si \quad x = 7 \end{cases}$$

Verifique que $P(X = x)$ es una función de probabilidad.

b) Calcule por definición, el valor esperado y la varianza de la variable aleatoria X. Grafique la función de probabilidad y sobreponga en una línea el valor esperado.

c) Se pueden generar muestras con reemplazo utilizando las probabilidades de la siguiente forma:

```
sample(valores, tamañomuestra, replace=T, prob=probabilidades)
```

Genere muestras de tamaño i=1,...,300 con reemplazo de la variable aleatoria X y guarde el promedio en cada iteración. Grafique los promedios y sobreponga el valor esperado. ¿Qué observa?

Práctica 1

Respuestas práctica 1

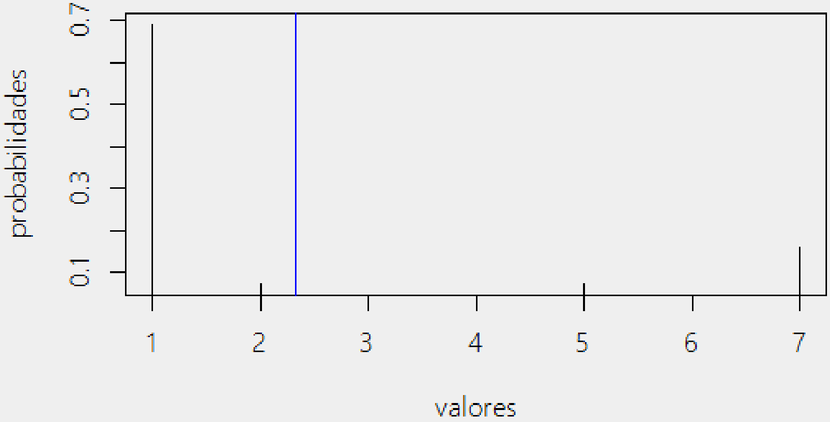
a)

```
#a) ¿Es función de probabilidad?

valores<-c(1,2,5,7) #Definimos los valores que toma X
probabilidades<-c(0.69,0.076,0.073,0.161) #Probabilidades asociadas

probabilidades>=0 #Probabilidades deben ser mayor a 0
[1] TRUE TRUE TRUE TRUE

sum(probabilidades) #Deben sumar 1
[1] 1

#Es función de probabilidad!
```

b)

```
(media<-sum(probabilidades*valores)) #Valor esperado es sumatoria de x*P(X=x)
[1] 2.334

(varianza<-sum((valores-media)^2*probabilidades))
[1] 5.260444

#install.packages("extrafont")
#library(extrafont)

#font_import()
#loadfonts()     #Importa tipos de letra
#fonts()    #Muestra todos los tipos de letra

par(family="Malgun Gothic Semilight")
plot(valores, probabilidades, type="h")
abline(v=media, col="blue")
```

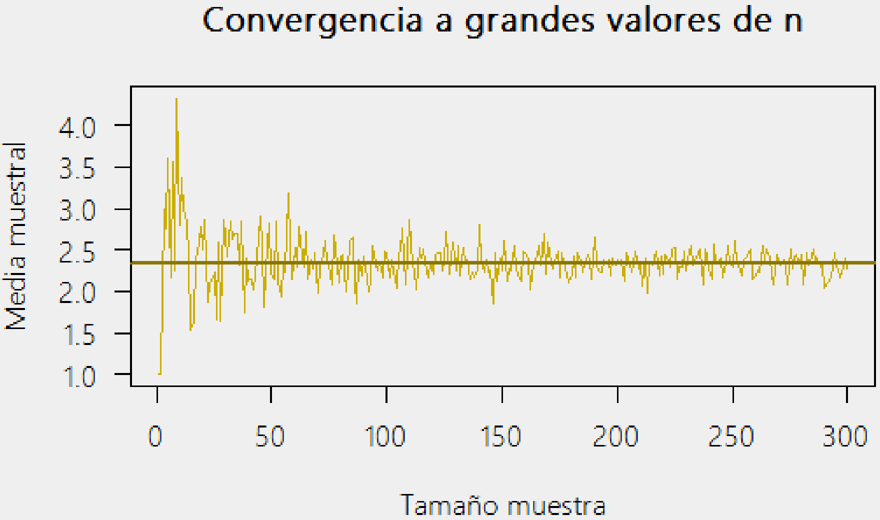
b)



c)

```
medias<-rep(0, 300)

for(i in 1:300){
muestra<-sample(valores, i, replace=T, prob=probabilidades)
medias[i]<-mean(muestra)
}

par(family="Malgun Gothic Semilight")
plot(medias, type="l", main="Convergencia a grandes valores de n", las=1, ylab="Media muestral",
xlab="Tamaño muestra", col="gold3")
abline(h=media, col="gold4", lwd=2)
```

c)



Convergencia a grandes valores de n
Media muestral
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0
50
100
150
200
250
300
Tamaño muestra

Práctica 2

a) Sea X una variable aleatoria con distribución Geométrica de parámetro p. Esta variable aleatoria describe el número de ensayos hasta el primer éxito, en un experimento donde los resultados de éste son éxito o fracaso y la probabilidad de éxito es p:

$$P(X = x) = p(1 - p)^{x-1} \qquad x = 1, 2...$$

Sin embargo, en `R` la función `dgeom()` entrega la siguiente función de probabilidad:

$$P(X = x) = p(1 - p)^{x} \qquad x = 1, 2...$$

Por lo tanto, para obtener la probabilidad de cierto número de ensayos hasta el primer éxito es necesario dividir $\frac{\texttt{dgeom()}}{1-p}$. Realice seis gráficos donde muestre las probabilidades de $x = 1, ..., 40$ utilizando $p < -(0.1, 0.2, 0.4, 0.6, 0.8, 0.95)$ además añada en una línea $E(X) = \frac{1}{p}$ para cada caso. ¿Qué efecto tiene el parámetro p?

Respuestas práctica 2

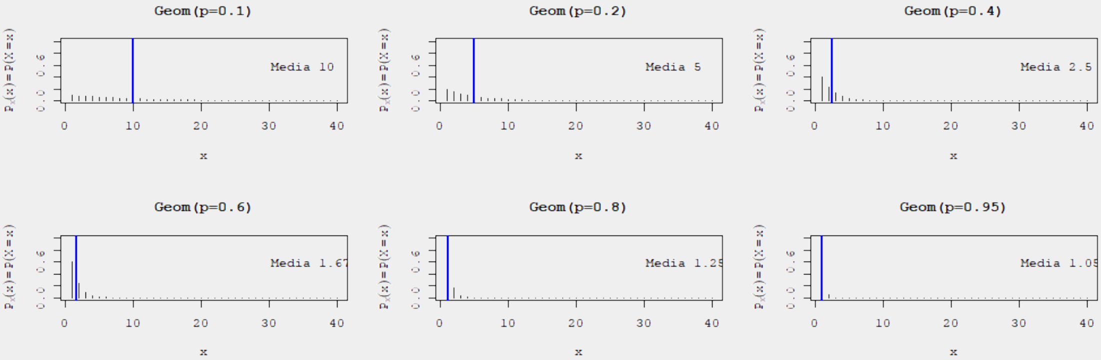
a)

```
#Opción manual

x<-1:40

par(mfrow=c(2,3)) #Ventana de gráficos

plot(x, dgeom(x,0.1)/(1-0.1), type="h", ylim=c(0,1),xlab=expression(x),ylab=expression(P[X](x)==P(X==x)), main="Geom(p=0.1)")
abline(v=1/0.1,lwd=2, col="blue")
legend(20,0.8,"media=10", bty="n")

plot(x, dgeom(x,0.2)/(1-0.2), type="h", ylim=c(0,1),xlab=expression(x),ylab=expression(P[X](x)==P(X==x)), main="Geom(p=0.2)")
abline(v=1/0.2,lwd=2, col="blue")
legend(18,0.8,"media=5", bty="n")

plot(x, dgeom(x,0.4)/(1-0.4), type="h", ylim=c(0,1),xlab=expression(x),ylab=expression(P[X](x)==P(X==x)), main="Geom(p=0.4)")
abline(v=1/0.4,lwd=2, col="blue")
legend(18,0.8,"media=2.5", bty="n")

plot(x, dgeom(x,0.6)/(1-0.6), type="h", ylim=c(0,1),xlab=expression(x),ylab=expression(P[X](x)==P(X==x)), main="Geom(p=0.6)")
abline(v=1/0.6,lwd=2, col="blue")
legend(18,0.8,"media=1.66", bty="n")

plot(x, dgeom(x,0.8)/(1-0.8), type="h", ylim=c(0,1),xlab=expression(x),ylab=expression(P[X](x)==P(X==x)), main="Geom(p=0.8)")
abline(v=1/0.8,lwd=2, col="blue")
legend(18,0.8,"media=1.25", bty="n")

plot(x, dgeom(x,0.95)/(1-0.95), type="h", ylim=c(0,1),xlab=expression(x),ylab=expression(P[X](x)==P(X==x)), main="Geom(p=0.95)")
abline(v=1/0.95,lwd=2, col="blue")
legend(18,0.8,"media=1.05", bty="n")
```

a)

```r
#Manera más eficiente corta con for:

x<-1:40
p<-c(0.1,0.2,0.4,0.6,0.8,0.95)

par(mfrow=c(2,3)) #Ventana de gráficos
#par(mfrow=c(2,3), family="Courier New") #define un tipo de letra

for(i in 1:6){
  plot(x, dgeom(x,p[i])/(1-p[i]), type="h", ylim=c(0,1),xlab=expression(x),ylab=expression(P[X](x)==P(X==x)),
  main=paste("Geom(p=", p[i], ")", sep=""))
  abline(v=1/p[i],lwd=2, col="blue")
  legend(15,0.8,paste("Media", round(1/p[i],2)), bty="n")

}
```

a)



A mayor valor de p menor es el valor esperado de número de ensayos hasta el primer éxito. ¿Por qué? Piensa en la definición de p...