

Maestría en Computo Estadístico
Inferencia Estadística
Tarea 1

22 de agosto de 2020

Enrique Santibáñez Cortés

Repositorio de Git: [Tarea 1, IE](#).

1. La compañía CIE ha desarrollado un nuevo producto. La demanda de tal artículo es desconocida, pero se asume que es una variable aleatoria distribuida uniformemente en $\{0, 1, \dots, N\}$. Los dispositivos deben de ser hechos por adelantado; cada uno vendido produce una ganancia de g pesos y cada uno de los que se queda sin vender produce una pérdida de p pesos. ¿Cuántos de estos artículos tienen que producirse para maximizar la ganancia esperada?

RESPUESTA

Sea X la demanda del nuevo producto, entonces la función de densidad es:

$$f(x) = \begin{cases} \frac{1}{N+1} & \text{para } X \in \{0, 1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

Sea r el número de dispositivos hechos por adelantado, definamos a la función ganancia como:

$$G = \begin{cases} gX - (r - X)p & \text{para } X \leq r \\ gr & \text{para } X > r \end{cases}$$

Ahora podemos calcular la esperanza de G :

$$\begin{aligned} \mathbb{E}[G] &= \mathbb{E}[(gX - (r - X)p) \cdot I_{\{X \leq r\}} + gr \cdot I_{\{X > r\}}] \\ &= \left(g \frac{N+2}{2} - rp + p \frac{N+2}{2} \right) \cdot \mathbb{P}(X \geq r) + gr \cdot \mathbb{P}(X < r) \end{aligned}$$

Ahora, considerando la definición de X :

$$\mathbb{P}(X < r) = \frac{r+1}{N+1}$$

Entonces tenemos que:

$$\mathbb{E}[G] = \left(g \frac{N+2}{2} - rp + p \frac{N+2}{2} \right) \cdot \left[1 - \frac{r+1}{N+1} \right] + gr \cdot \left[\frac{r+1}{N+1} \right]$$

Ahora derivamos con respecto a r para maximizar $\mathbb{E}[G]$.

$$\mathbb{E}'[G] = -p + \frac{2rp - p}{N+1} + \frac{2rg + g}{N+1}.$$

Iguualamos a 0 y despejamos r .

$$-p + \frac{2rp - p}{N+1} + \frac{2rg + g}{N+1} = 0$$

$$2r(p + g) - p + g = p(N + 1)$$

$$r = \frac{p(N+1) - p + g}{2(p+g)}$$

Solo falta demostrar que es un máximo.

Por lo tanto, cuando r es igual a $\frac{p(N+1)-p+g}{2(p+g)}$ la esperanza de la ganancia se maximiza ■.

2. Un conjunto de bits se envían sobre un canal de comunicación en paquetes de 12. Si la probabilidad de que un bit sea corrompido sobre este canal es 0.1 y los errores son independientes, ¿cuál es la probabilidad de que no más de dos bits de un paquete se corrompan? Si 6 paquetes de bits se envían sobre el canal, ¿cuál es la probabilidad de que al menos un paquete contenga 3 o más bits corruptos? Finalmente, si X denota el número de paquetes conteniendo 3 o más bit corruptos, ¿cuál es la probabilidad de que X excederá su media por más de dos desviaciones estándar?

RESPUESTA

Sea Y el número de bits que son corrompidos en este canal dentro de un paquete de 12, entonces podemos decir que, $Y \sim \text{Bin}(12, 0.1)$. Entonces la probabilidad de que no más de dos bits de un paquete se corrompan es

$$\mathbb{P}(Y \leq 2) = \mathbb{P}(Y = 1) + \mathbb{P}(Y = 0) = 0.88913.$$

Ahora si existen 6 paquetes de bits sobre el canal, para determinar la probabilidad de que al menos un paquete contenga 3 o más bits corruptos, primero calculemos la probabilidad de que un paquete contenga 3 o más bits, la cual esta dada como

$$\mathbb{P}(Y \geq 3) = 1 - \mathbb{P}(Y \leq 2) = 1 - 0.88913 = 0.11087.$$

Ahora, sea X el número de paquetes que contienen 3 o más bits corruptos de n paquetes existentes. Por definición podemos decir que $X \sim \text{Bin}(n, p)$, donde p es la probabilidad de que un paquete tenga 3 o más bits corruptos. $X \sim \text{Bin}(6, 0.11087)$, entonces la probabilidad de que al menos un paquete contenga 3 o más bits sería:

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X < 1) = 0.8637243.$$

Ahora, la probabilidad de que X exceda su media por más de dos desviaciones estándar es, primero calculemos la media y la desviación estandar, como X es una distribución binomial tenemos que

$$\begin{aligned}\mathbb{E}(X) &= np = 6 * 0.11087 = 0.66522 \\ \sigma &= \sqrt{\text{var}(X)} = \sqrt{np(1-p)} = \sqrt{6 * 0.11087(1 - 0.11087)} = 0.769069 \\ \mathbb{P}(X > np + 2\sqrt{np(1-p)}) &= \mathbb{P}(X > 0.66522 + 2(0.769069) = \mathbb{P}(X > 2.203358) \quad \blacksquare.\end{aligned}$$

3. Una caja contiene 12 manzanas frescas y 4 que están podridas. Si elige 3 al azar y X denota el número de manzanas frescas que tomó, encuentre la función de densidad de X y su esperanza.

RESPUESTA

Por definición de X , podemos decir que $X \sim \text{Hyper}(n = 3, M = 8, N = 12)$. Por lo que podemos decir que la función de densidad es:

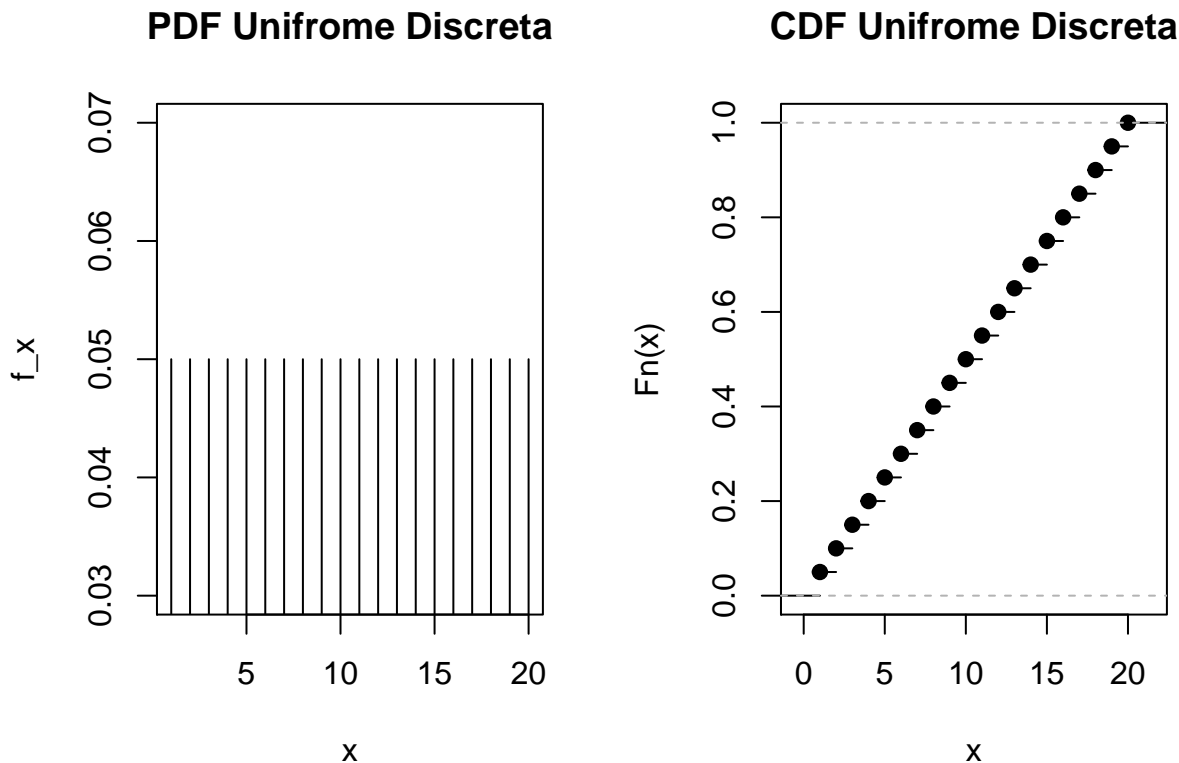
$$f(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

Y su esperanza es $n \frac{M}{N}$. ■.

4. Para el siguiente ejercicio es necesario el programa R.

- a. Escriba un programa en R que reproduzca las gráficas de las funciones de distribución acumulada y de masa de la distribución uniforme que aparecen en las notas del curso. Las gráficas deben verse similares a las figuras de la Figura 1. **RESPUESTA**

```
grafica_pdf_and_cdf_uniforme <- function(n){  
  x <- seq(1:n)  
  f_x <- rep(1/n,n)  
  F_x <- cumsum(f_x)  
  par(mfrow=c(1,2))  
  plot(x, f_x, main="PDF Unifrome Discreta", type="h")  
  plot.ecdf(x, main="CDF Unifrome Discreta")  
}  
grafica_pdf_and_cdf_uniforme(20)
```



- b. Lea en la documentación de R, o en cualquier otra fuente de información confiable, la explicación de la función `sample(x, size, replace=FALSE, prob=NULL)`. (No es necesario entregar algo para este ejercicio).
- c. Usando la función `sample` simule una muestra de tamaño 10 000 de la distribución $U(1, \dots, 10)$. Fijando la semilla en 13 (`set.seed(13)`), muestre los resultados de la simulación en una tabla de frecuencia y calcule la media y la varianza. Sugerencia: Use la función `table`.

RESPUESTA

```
set.seed(13)  
muestra <- sample(10,10000, replace = TRUE)  
table(muestra)
```

```
## muestra
##      1      2      3      4      5      6      7      8      9     10
## 1015 1001  962 1013 1020  982  991  926 1069 1021
```

```
mean(muestra)
```

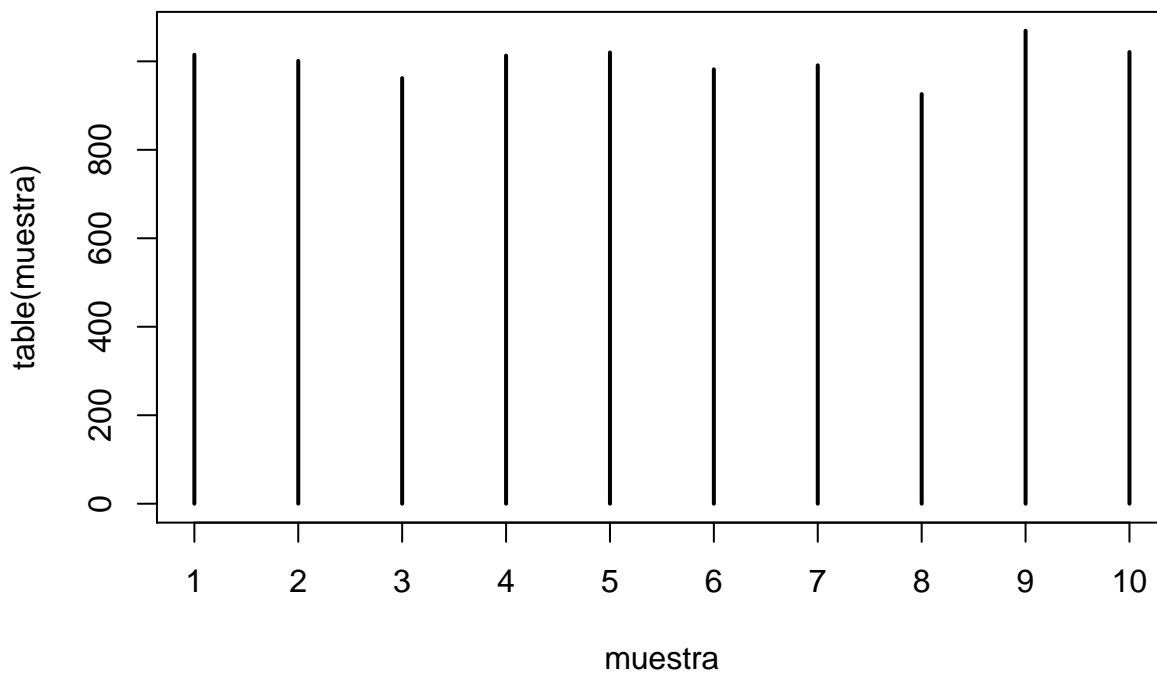
```
## [1] 5.5123
```

```
var(muestra)
```

```
## [1] 8.340283
```

d. Grafique las frecuencias de la simulación anterior.

```
plot(table(muestra))
```



5. Para el siguiente ejercicio también necesitamos R.

- Usando la función `sample`, simule 10 lanzamientos de una moneda equilibrada y cuente el número de águilas que obtiene. Repita este proceso 10^6 veces y muestre sus primeros 3 resultados. Grafique las frecuencias del número de águilas obtenidas en los 10^6 experimentos. También grafique las proporciones del número de águilas obtenidas.

RESPUESTA

Sea 1: obtener una aguil, 0: obtener sol.

```
lanzamientos <- c(0,1)
```

```
sum(sample(lanzamientos, 10, replace = TRUE))
```

```
## [1] 6
```

```
simulacion_monedas <- c()
```

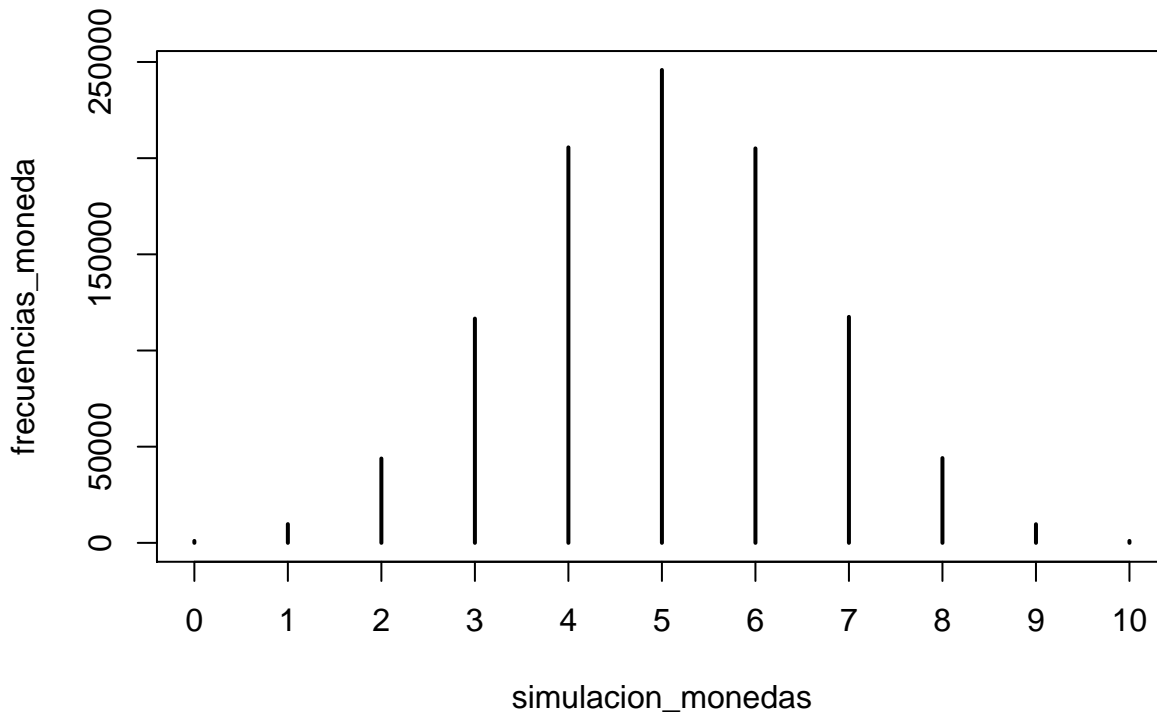
```
for(i in 1:10**6) {
  simulacion_monedas[i] <- sum(sample(lanzamientos, 10, replace = TRUE))
}
```

```

}
simulacion_monedas[1:3]

## [1] 3 6 4
frecuencias_moneda <- table(simulacion_monedas)
plot(frecuencias_moneda)

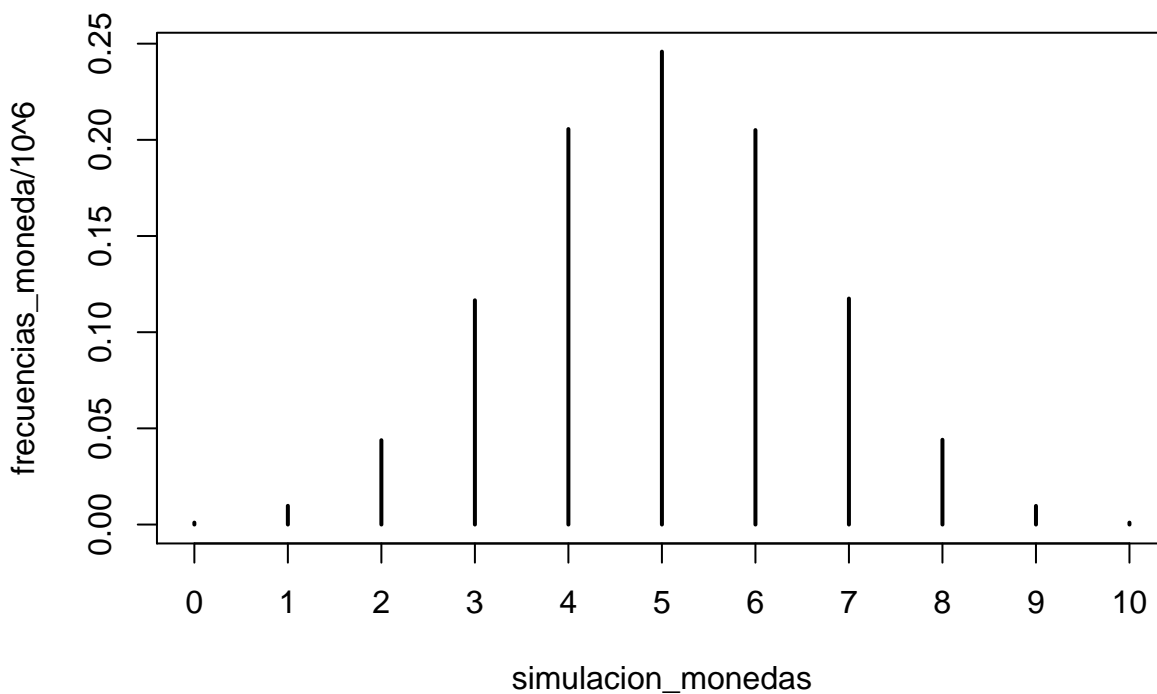
```



```

plot(frecuencias_moneda/10**6)

```

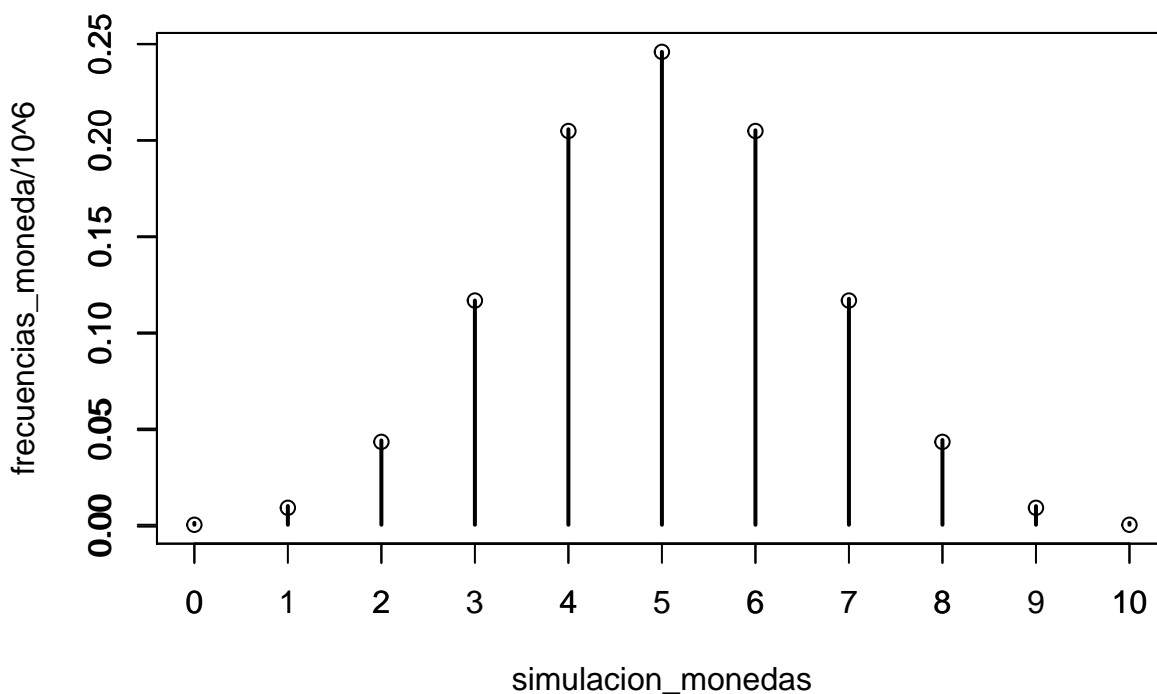


b. Usando la función dbinom grafique la función de masa de una distribución $B(10, 0.5)$ sobre la gráfica

de las proporciones que hizo en el inciso anterior.

RESPUESTA

```
plot(frecuencias_moneda/10**6)
par(new=TRUE)
plot(x=seq(0,10,1),dbinom(x=seq(0,10,1), size = 10,prob = 0.5),xlab = "", ylab = "",)
```



c. Repita los dos incisos anteriores para una moneda desequilibrada que tiene probabilidad $p = 0.3$ de obtener un águila cuando se lanza. ¿Qué observa?

RESPUESTA

```
lanzamientos_moneda_desequilibrada <- c(rep(0,7),rep(1,3))
```

```
sum(sample(lanzamientos_moneda_desequilibrada, 10, replace = TRUE))
```

```
## [1] 2
```

```
simulacion_monedas_desequilibrada <- c()
```

```
for(i in 1:10**6) {
```

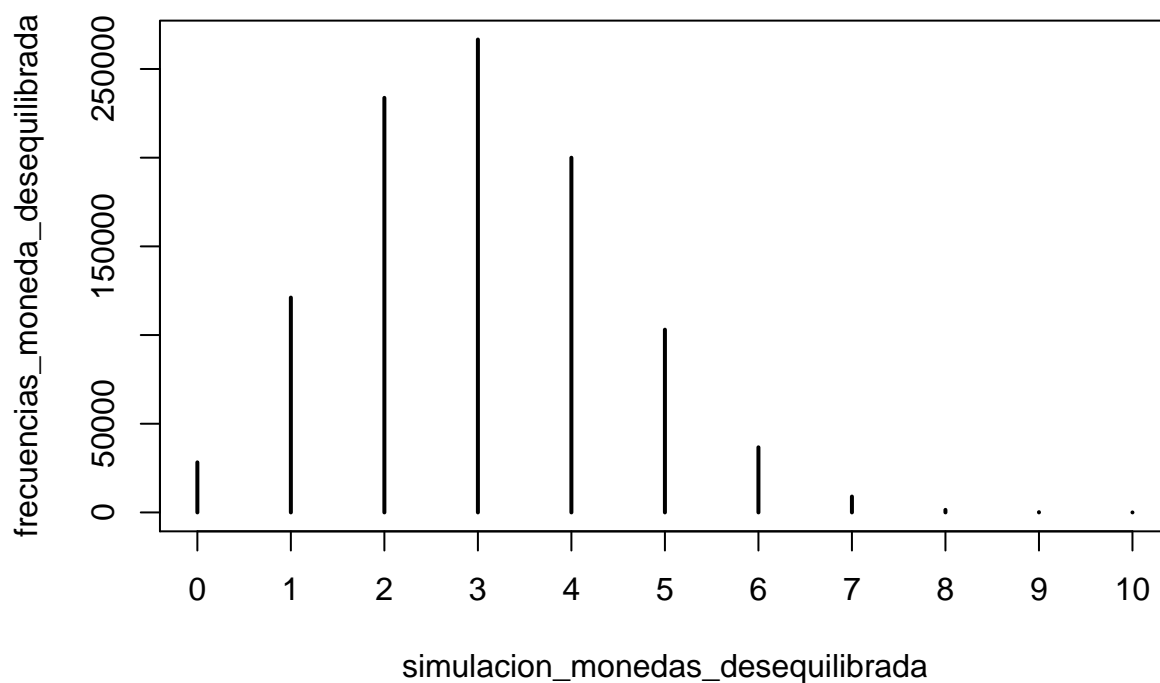
```
simulacion_monedas_desequilibrada[i] <-sum(sample(lanzamientos_moneda_desequilibrada, 10, replace = TRUE))
}
```

```
simulacion_monedas[1:3]
```

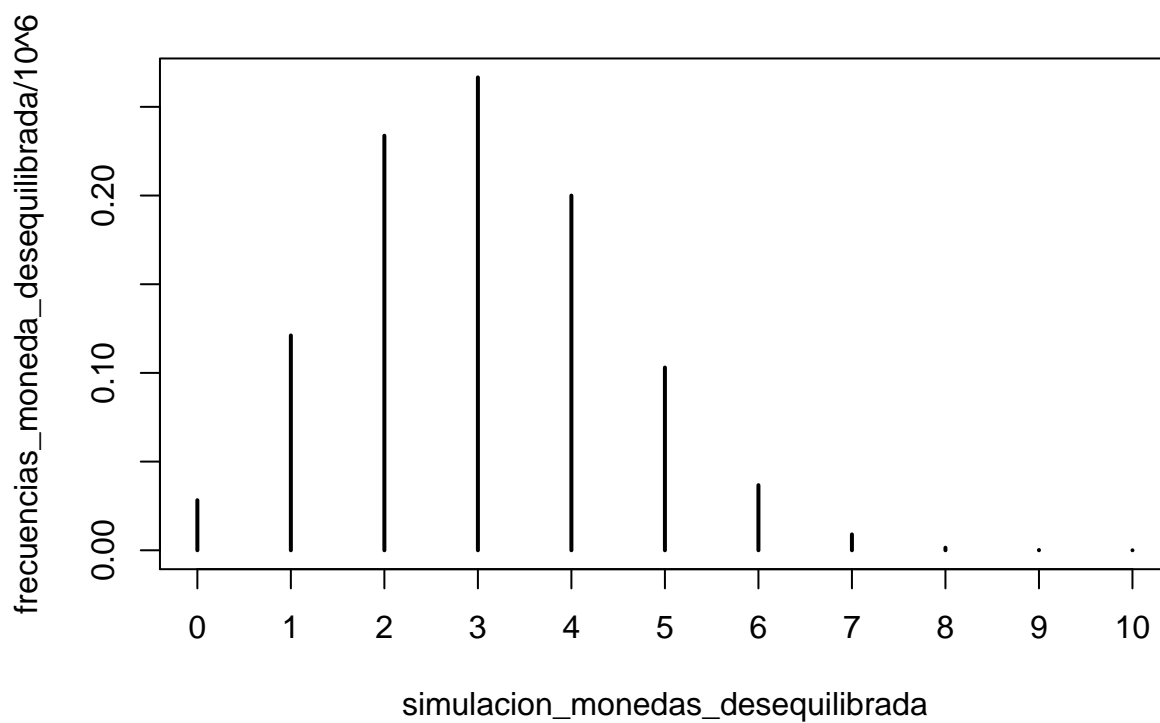
```
## [1] 3 6 4
```

```
frecuencias_moneda_desequilibrada <- table(simulacion_monedas_desequilibrada)
```

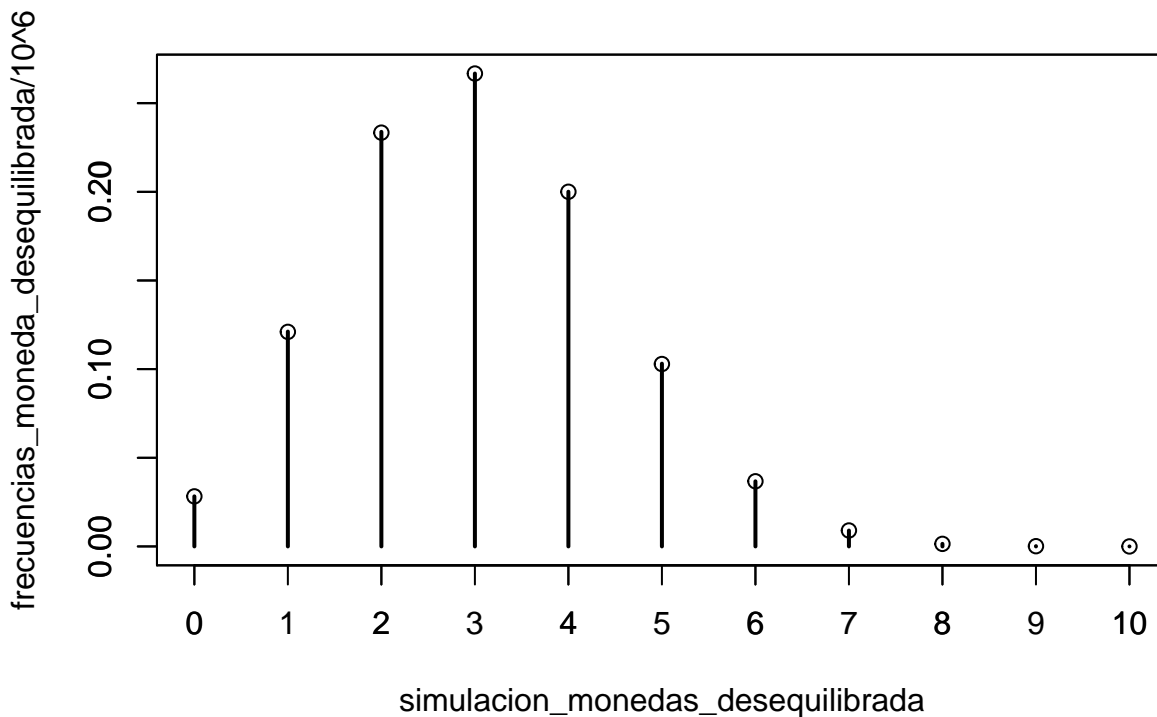
```
plot(frecuencias_moneda_desequilibrada)
```



```
plot(frecuencias_moneda_desequilibrada/10**6)
```



```
plot(frecuencias_moneda_desequilibrada/10**6)
par(new=TRUE)
plot(x=seq(0,10,1), y=dbinom(x=seq(0,10,1), size = 10,prob = 0.3),xlab = "", ylab = "",xlim = c(0
```



6. Una urna contiene 46 bolas grises y 49 bolas blancas. Usando la función `sample` en R, simule la extracción sin reemplazamiento de 20 de estas bolas y cuente el número de bolas grises que obtuvo. Repita este proceso 10^6 veces y grafique las frecuencias de bolas grises obtenidas en cada experimento. ¿Cuál es la probabilidad de que al extraer 20 bolas de la urna 5 de ellas sean grises? También grafique la proporción de bolas grises obtenidas en los experimentos anteriores y sobre esta figura añada la correspondiente función de masa de la distribución Hipergeométrica asociada al experimento total.

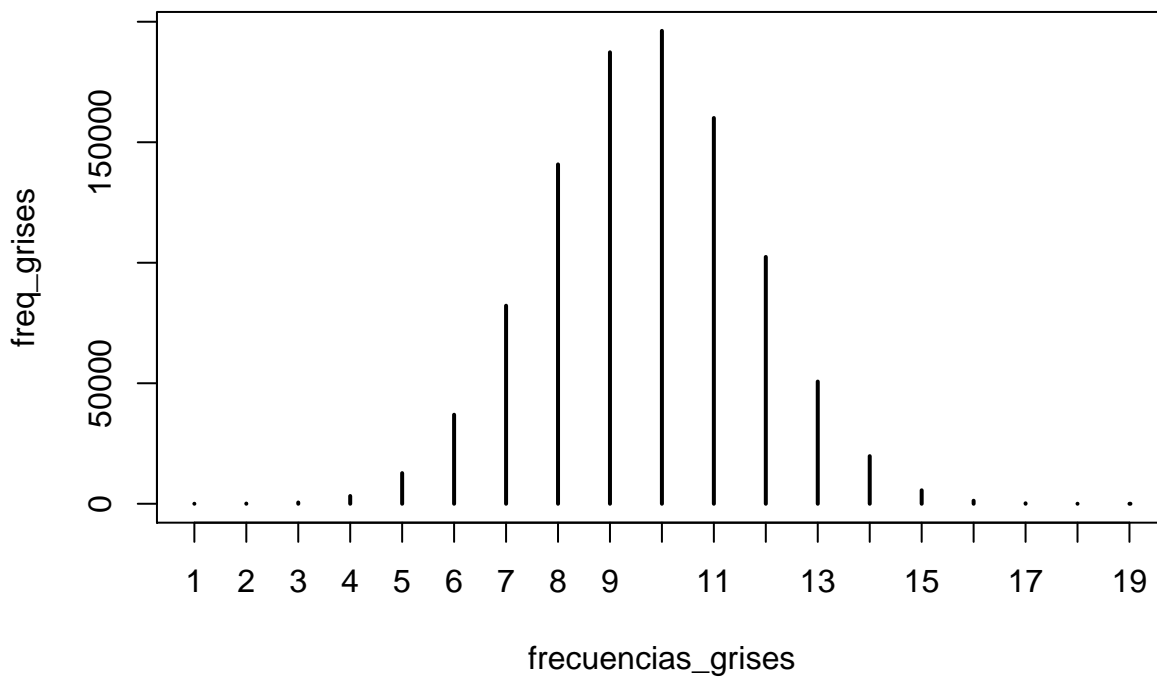
RESPUESTA

Sea 1: una bola gris, y 0: una bola blanca.

```
urna <- c(rep(1,46), rep(0,49))
sum(sample(urna, 20, replace=FALSE))

## [1] 9

frecuencias_grises <- c()
for(i in 1:10**6) {
  frecuencias_grises[i] <- sum(sample(urna, 20, replace=FALSE))
}
freq_grises <- table(frecuencias_grises)
plot(freq_grises)
```

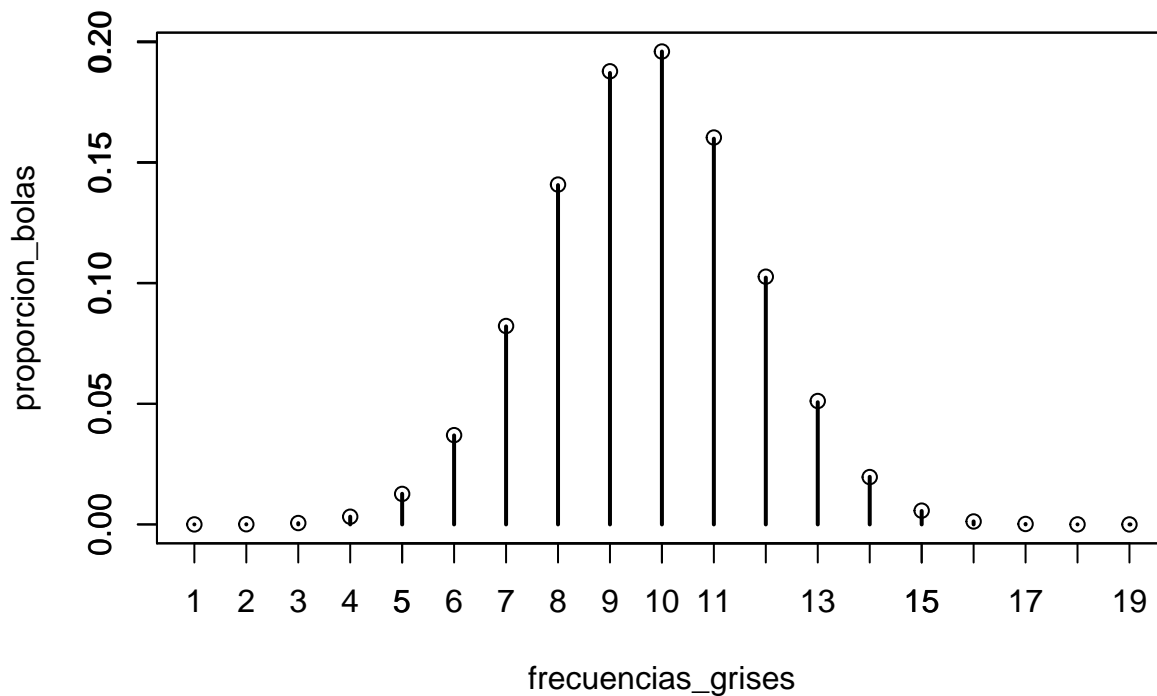
Si X es el número de bolas grises que se obtienen en la extracción sin reemplazamiento de 20 bolas de la urna definida en el problema, entonces podemos decir que $X \sim \text{Hyper}(n = 20, M = 46, N = 95)$. Por lo tanto, la probabilidad de que se extraigan 5 bolas grises es:

$$f(x = 5) = \frac{\binom{46}{5} \binom{49}{15}}{\binom{95}{20}} = 0.01261935.$$

Graficamos la proporción de bolas grises:

```
proporcion_bolas <- freq_grises/10**6

plot(proporcion_bolas)
par(new=TRUE)
plot(x=seq(1,19,1), y=dhyper(x=seq(1,19,1), m = 46, n=49, k=20), xlab = "", ylab="")
```



7. Sea X una variable aleatoria con función de distribución F dada por

$$F(x) = \begin{cases} 0 & \text{para } x < 0 \\ 1/2 & \text{para } 0 \leq x < 1/4 \\ 3/4 & \text{para } 1/4 \leq x < 3/4 \\ 1 & \text{para } 3/4 \leq x \end{cases}$$

Determine la función de probabilidad de X .

RESPUESTA

Recordemos que cuando X es una variable aleatoria, su función de distribución F esta definida como:

- caso discreto

$$F(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f(x_i).$$

- caso continuo

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Donde f es la función de probabilidad de X .

Debido a que $F(x)$ es discontinua en los puntos x_i podemos decir que X es una variable aleatoria discreta.

Ahora ocupando lo anterior, podemos decir que:

- Para $x = 0$,

$$f(x) = \mathbb{P}(X = 0) = F(0) - F(0^-) = 1/2 - 0 = 1/2.$$

- Para $x = 1/4$,

$$f(x) = \mathbb{P}(X = 1/4) = F(1/4) - F(1/4^-) = 3/4 - 1/2 = 1/4.$$

- Para $x = 3/4$,

$$f(x) = \mathbb{P}(X = 3/4) = F(3/4) - F(3/4^-) = 1 - 3/4 = 1/4.$$

Por lo tanto la función de probabilidad de X es:

$$f(x) = \begin{cases} 1/2 & \text{para } x = 0 \\ 1/4 & \text{para } x = 1/4 \\ 1/4 & \text{para } x = 3/4 \end{cases} \quad \blacksquare$$

8. Sea X una variable aleatoria con valores en $[0, 1]$ y función de distribución $F(x) = x^2$. ¿Cuál es la densidad de X ? Calcule las siguientes probabilidades: i) $\mathbb{P}(1/4 \leq X \leq 3/4)$; ii) $\mathbb{P}(X > 1/2)$; iii) $\mathbb{P}(X \leq 3/4 | X > 1/2)$.

RESPUESTA

Debido a que $F(x)$ esta definida como un polinomio, esto implica que $F(x)$ es continua y además que X es una variable aleatoria continua. Ocupando el teorema fundamental del calculo y la definición de $F(x)$ para una variable aleatoria continua tenemos que

Si X es una variable aleatoria continua con función de densidad $f(x)$ y función de distribución acumulada $F(x)$, entonces en cada x en la que existe la derivada $F'(x)$ implica que $F'(x) = f(x)$.

Como $F(x) = x^2$ es un polinomio esto implica que $F'(x)$ en $\forall x$. Por lo tanto, la función la densidad de X es

$$f(x) = F'(x) = 2x \quad I_{x \in [0,1]}.$$

Recordemos que algunas propiedades de la función acumulada:

•

$$F(a) = \mathbb{P}(X < a).$$

•

$$\mathbb{P}(a \leq X \leq b) = F(b) - F(a).$$

•

$$\mathbb{P}(X > a) = 1 - \mathbb{P}(X < a).$$

Ocupando lo anterior tenemos:

i)

$$\mathbb{P}(1/4 \leq X \leq 3/4) = F(3/4) - F(1/4) = (3/4)^2 - (1/4)^2 = 8/16 = 1/2.$$

ii)

$$\mathbb{P}(X > 1/2) = 1 - F(1/2) = 1 - (1/2)^2 = 3/4.$$

iii) Ocupando el teorema de Bayes,

$$\begin{aligned}\mathbb{P}(X \leq 3/4 | X > 1/2) &= \frac{\mathbb{P}(X \leq 3/4 \cap X > 1/2)}{\mathbb{P}(X > 1/2)} = \frac{\mathbb{P}(1/2 < X \leq 3/4)}{1 - \mathbb{P}(X < 1/2)} = \frac{F(3/4) - F(1/2)}{1 - F(1/2)} \\ &= \frac{(3/4)^2 - (1/2)^2}{3/4} = \frac{5/16}{3/4} = 5/12. \quad \blacksquare\end{aligned}$$

9. Un lote muy grande de componentes ha llegado a un distribuidor. Se puede decir que el lote es aceptable solo si la proporción de componentes defectuosos es cuando mucho 0.10. El distribuidor decide seleccionar aleatoriamente 10 componentes y aceptar el lote solo si el número de componentes defectuosos en la muestra es cuando mucho 2.

a) ¿Cuál es la probabilidad de que el lote sea aceptado cuando la proporción real de defectuosos es 0.01, 0.05, 0.10, 0.20, 0.25?

RESPUESTA

Sea X el número de componentes defectuosos en la muestra aleatoria de 10 componentes. Considerando que el lote es muy grande y la proporción p real de defectuosos en el lote, podemos decir que $X \sim \text{Bin}(10, p)$. Entonces la probabilidad de que el lote sea aceptado es

$$\mathbb{P}(X \leq 2).$$

Ahora considerando las diferentes proporciones, tenemos que (ocupando la función `pbinom(q=2, size=10, prob = p)` de software estadístico R.):

- cuando $p = 0.01$, la probabilidad de aceptar el lote es:

$$\begin{aligned}\mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\ &= \binom{10}{0}(0.99)^{10} + \binom{10}{1}(0.99)^9(0.01) + \binom{10}{2}(0.99)^8(0.01)^2 \\ &= 0.9998862.\end{aligned}$$

- cuando $p = 0.05$, la probabilidad de aceptar el lote es:

$$\begin{aligned}\mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\ &= \binom{10}{0}(0.95)^{10} + \binom{10}{1}(0.95)^9(0.05) + \binom{10}{2}(0.95)^8(0.05)^2 \\ &= 0.9884964.\end{aligned}$$

- cuando $p = 0.10$, la probabilidad de aceptar el lote es:

$$\begin{aligned}\mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\ &= \binom{10}{0}(0.90)^{10} + \binom{10}{1}(0.90)^9(0.10) + \binom{10}{2}(0.90)^8(0.10)^2 \\ &= 0.9298092.\end{aligned}$$

- cuando $p = 0.20$, la probabilidad de aceptar el lote es:

$$\begin{aligned}\mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\ &= \binom{10}{0}(0.80)^{10} + \binom{10}{1}(0.80)^9(0.20) + \binom{10}{2}(0.80)^8(0.20)^2 \\ &= 0.6777995.\end{aligned}$$

- cuando $p = 0.25$, la probabilidad de aceptar el lote es:

$$\begin{aligned}\mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\ &= \binom{10}{0}(0.75)^{10} + \binom{10}{1}(0.75)^9(0.25) + \binom{10}{2}(0.75)^8(0.25)^2 \\ &= 0.5255928.\end{aligned}$$

Podemos observar, que cuando la proporción real de componentes defectuosos del lote es más grande esto implica que la probabilidad de aceptar el lote es cada vez menor, es decir, la proporción real es inversamente proporcional a la probabilidad de aceptar el lote. Si observamos la función de distribución acumulada la X cambio el valor p , se puede observar claramente este hecho. ■

10. Sean $G = \{1, 2, 3\}$, $H = \{4, 5, 6\}$. Lanzamos dos dados y sean los eventos A = el primer dado cae en H ; B = el segundo dado cae en H ; C = un dado cae en G y el otro en H ; D = el total es cuatro, E = el total es cinco y F = el total es siete. ¿Cuáles de las siguientes proposiciones son ciertas?
- i) A y F son independientes. ii) A y D son independientes. iii) A y E son independientes. iv) $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. v) A y C son independientes. vi) C y E son independientes. vii) $\mathbb{P}(A \cap C \cap E) = \mathbb{P}(A)\mathbb{P}(C)\mathbb{P}(E)$. viii) A , C y E son independientes. Justifique sus respuestas.

RESPUESTA

Analicemos las proposiciones:

- i) A y F son independientes. **Falso**. Si consideramos la probabilidad:

$$\mathbb{P}(F|A) = .$$

- ii) A y D son independientes. **Falso**

- iii) A y E son independientes. **Falso**.

- iv) $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. **Verdadero**

- v) A y C son independientes. **Falso**.

- vi) C y E son independientes. **Falso**.

- vii) $\mathbb{P}(A \cap C \cap E) = \mathbb{P}(A)\mathbb{P}(C)\mathbb{P}(E)$. **Falso**.

- viii) A , C y E son independientes. **Falso**.

11. **Extra** Solución extra del ejercicio 4 en python.