Maestría en Computo Estadístico
Inferencia Estadística
Tarea 1
25 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 $\{0, 1, \cdots, N\}$. 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 perdida 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:

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

Sea r el número de dispositivos hechos por adelantado, $g > 0$ y $p > 0$,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:

$$\mathbb{E}[G] = \mathbb{E}[[gX - (r - X)p] \cdot I_{\{X \leq r\}} + gr \cdot I_{\{X > r\}}]$$
$$\mathbb{E}[G] = g\mathbb{E}[X \cdot I_{\{X \leq r\}}] - rp\mathbb{E}[\cdot I_{\{X \leq r\}}] + p\mathbb{E}[X \cdot I_{\{X \leq r\}}] + gr\mathbb{E}[\cdot I_{\{X > r\}}]$$

Debido a que $X \sim U(0, N)$, entonces $\mathbb{E}(X \cdot I_{\{X \leq r\}}) = \frac{r(r-1)}{2(N+1)}$, $\mathbb{E}[I_{\{X \leq r\}}] = \mathbb{P}(X < r))$ y $\mathbb{P}(X < r) = \frac{r}{N+1}$. Utilizando lo anterior y algunas propiedades básicas de la esperanza podemos desarrollar la esperanza como

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

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

Simplificando la expresión anterior y derivando $\mathbb{E}[G]$ con respeto a r tenemos

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

Igualando a 0 y despejando r tenemos

$$\begin{aligned} \tfrac{2gr-g}{2(N+1)} + \tfrac{-2rp}{N+1} + \tfrac{2gr-p}{2(N+1)} + g + \tfrac{-2pr}{N+1} &= 0 \\ 2gr - g - 4rp + 2pr - p + 2g(N+1) - 4gr &= 0 \\ -2gr + g - 2rp - p + 2gN &= 0 \\ -2r(p+g) &= -g - 2gN + p \\ r &= \tfrac{g+2gN-p}{2(p+g)} \end{aligned}$$

Ahora usando el criterio de segunda derivada para determinar que el punto crítico encontrado sea un máximo/mínimo, calculemos la segunda derivada de la esperanza de la función de las ganancias:

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

Como $p, g, N > 0$ por como se definieron desde el principio del problema, implica que $\mathbb{E}''[G] < 0$ y que r en $\frac{g+2gN-p}{2(p+g)}$ es un máximo.

Por lo tanto cuando se producen $\frac{g+2gN-p}{2(p+g)}$ artículos la ganancia esperada 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 $Y \sim \ Bin(12, 0.1)$. Por lo que **la probabilidad de que no más de dos bits de un paquete se corrompan es** (ocupando la función pbinom(q=2, size=12, prob = 0.1) en el software estadístico R.):

$$\mathbb{P}(Y \leq 2) = \mathbb{P}(Y = 2) + \mathbb{P}(Y = 1) + \mathbb{P}(Y = 0) = \sum_{y=0}^{2} \binom{12}{y} (0.1)^y (0.9)^{12-y} = 0.88913.$$

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

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

Ahora, sea X el número de paquetes que contienen 3 o más bits corruptos de n paquetes existentes. Por definición podemos decir que $X \sim Bin(n, p)$, donde p es la probabilidad de que un paquete tenga 3 o más bits corruptos. Por lo que para este problema $n = 6$ y $p = 0.1107$ y $X \sim Bin(6, 0.11087)$, entonces **la probabilidad de que al menos un paquete contenga 3 o más bits sería**

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X < 1) = 0.5059265.$$

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

$$\mathbb{E}(X) = np = 6 * 0.11087 = 0.66522.$$

$$\sigma = \sqrt{var(X)} = \sqrt{np(1-p)} = \sqrt{6 * 0.11087(1 - 0.11087)} = 0.769069.$$

Por lo que **la probabilidad de que X exceda su media por más de dos desviaciones estándar es**

$$\mathbb{P}(X > np + 2\sqrt{np(1-p)}) = \mathbb{P}(X > 0.66522 + 2(0.769069) = \mathbb{P}(X > 2.203358).$$

Por ser una variable discreta en los enteros positivos, podemos utilizar que:

$$\mathbb{P}(X > 2.203358) = \mathbb{P}(X \geq 3) = 1 - \mathbb{P}(X \leq 2) = 0.02104176 \ \blacksquare.$$

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 $X \sim Hyper(n=3, M=12, N=16)$. Por lo que podemos decir que la función de densidad es :

$$f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \quad \text{para } x \in \{0,1,2,3\}.$$

Calculemos los valores exactos de la función de densidad (ocupando la función `dhyper(x=c(0,1,2,3), m=12, n=4, k=3)`):

- para $x = 0$,
$$f(0) = \frac{\binom{12}{0}\binom{4}{3}}{\binom{16}{3}} = 0.007142857.$$
- para $x = 1$,
$$f(1) = \frac{\binom{12}{1}\binom{4}{2}}{\binom{16}{3}} = 0.128571429.$$
- para $x = 2$,
$$f(2) = \frac{\binom{12}{2}\binom{4}{1}}{\binom{16}{3}} = 0.471428571.$$
- para $x = 3$,
$$f(3) = \frac{\binom{12}{3}\binom{4}{0}}{\binom{16}{3}} = 0.392857143.$$

En conclusión **la función de densidad de X es**:

$$f(x) = \begin{cases} 0.007142857 & \text{para } x = 0 \\ 0.128571429 & \text{para } x = 1 \\ 0.471428571 & \text{para } x = 2 \\ 0.392857143 & \text{para } x = 3 \\ 0 & \text{en otro caso} \end{cases}$$

Debido a que X se distribuye como una hypergeometrica tenemos que **la esperanza es**

$$\mathbb{E}[X] = \frac{nM}{N} = \frac{3 \cdot 12}{16} = 2.25 \quad \blacksquare.$$

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.

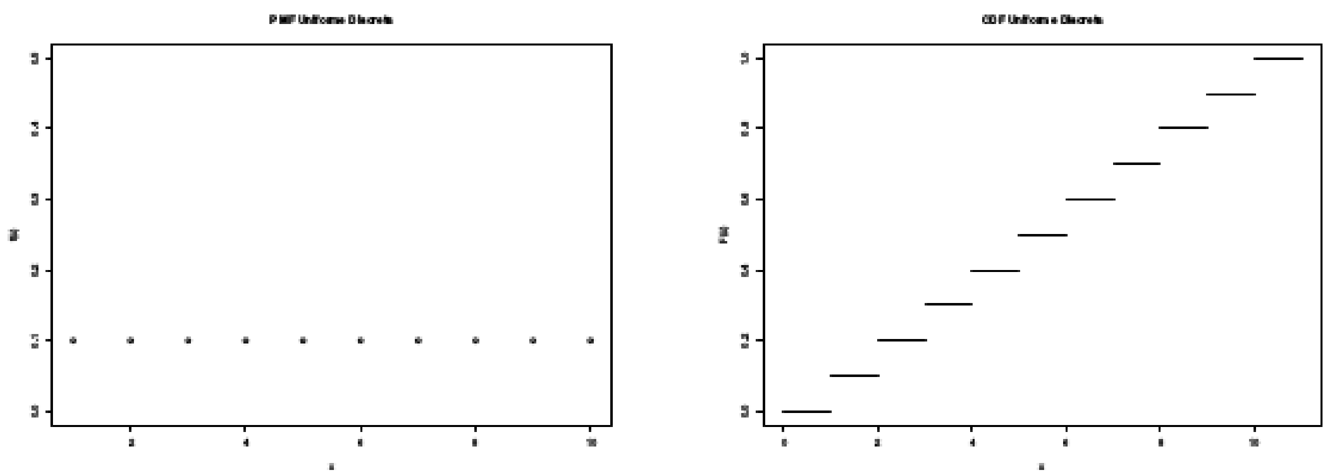


Figura 1: Ejemplo de las funciones de masa y acumulada de la distribución uniforme.

RESPUESTA

Cargamos los paquetes a ocupar:

```
library(tidyverse)
library(gridExtra)
```

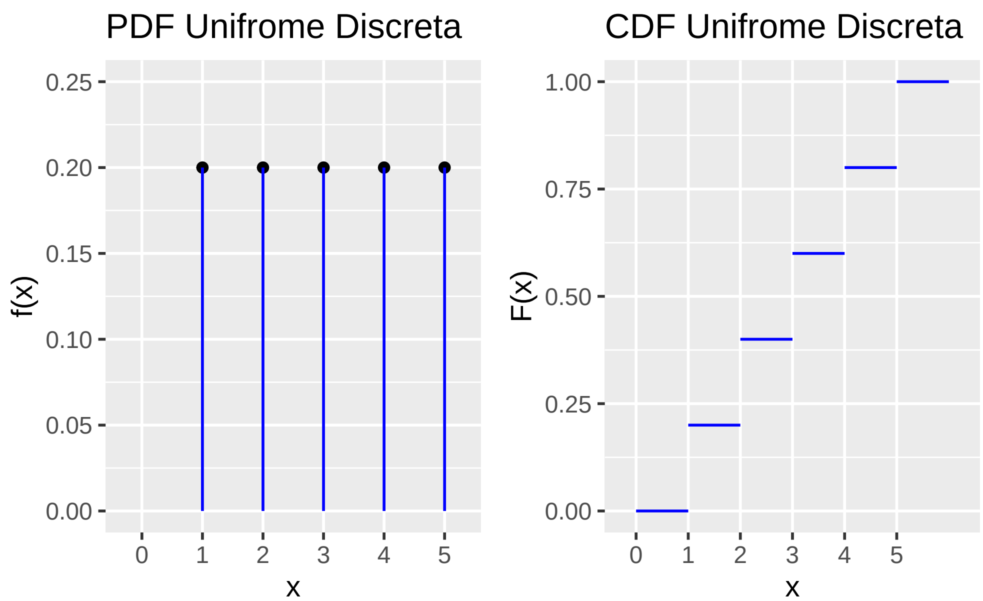
Definimos una función que gráfique la función de probabilidad y la función de probabilidad acumulada para una $U(0,n)$ discreta :

```
grafica_pdf_and_cdf_uniforme <- function(n){ # La función tiene como parametro n.
    data_uniforme <- data_frame(x=seq(1:n)) %>%
      mutate("f_x"=1/n, # Definimos las probabilidades.
             "F_x"= cumsum(f_x)) # Definimos las probabilidades acumaladas.
    data_uniforme[n+1,]<-c(0,NaN,0) # Agregamos el valor inicial

    cdf <- ggplot(data=data_uniforme)+ # CDF
      geom_segment(aes(x=x, xend=x+1, y=F_x, yend=F_x), col="blue")+
      scale_x_discrete(limits=seq(0,n,1))+
      labs(y="F(x)",title="CDF Unifrome Discreta" )+# Graficamos los segementos
      ylim(0,1) # Limites del eje "y".
    pdf <- ggplot(data=data_uniforme)+ # PDF
      geom_point(aes(x,f_x))+
      geom_segment(aes(x=x, xend=x, y=0, yend=f_x), col="blue")+ # Graficamos los segementos.
      scale_x_discrete(limits=seq(0,n,1))+
      ylim(0,1/(n-1))+ # Límites del eje "y".
      labs(x="x", y="f(x)",title = "PDF Unifrome Discreta")
  grid.arrange(pdf, cdf, ncol=2) # Ponemos las gráficas juntas.
  }
```

Mostramos las funciones para una distribución $U(0,5)$ discreta:

```
grafica_pdf_and_cdf_uniforme(5)
```



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 $U(1, ..., 10)$. 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

Simulamos la realización de la muestra de tamaño 10000.

```
set.seed(13) # fijamos la semilla.
muestra <- sample(10, 10000, replace = TRUE) # Muestra aleatoria.
```

La tabla de frecuencia es:

```
table(muestra) # Tabla de Frecuencias
```

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

La media de la muestra es:

```
mean(muestra) # Media de la muestra simulada.
```

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

La varianza de la muestra es:

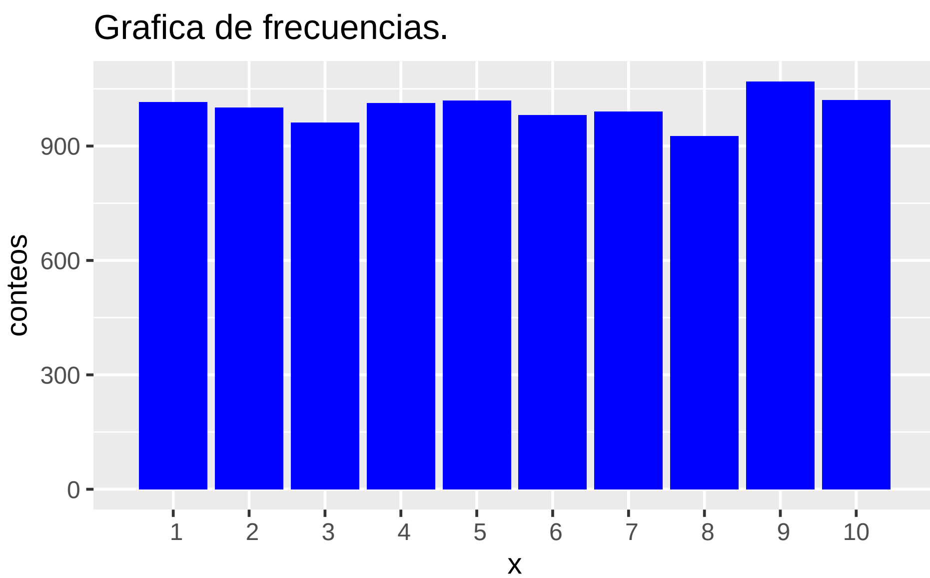
```
var(muestra) # Varianza de la muestra simulada.
```

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

Se calculamos la densidad, la varianza y la media de una $U(N)$ podemos observar que las estimaciones con la simulación anterior se aproxima muy bien a las teoricas.

d. Grafique las frecuencias de la simulación anterior.

```r
ggplot(data=data.frame(x=muestra), aes(x))+
  geom_bar(fill="blue")+
  scale_x_discrete(limits =seq(1,10,1))+
  labs(title="Grafica de frecuencias.", y="conteos")
```



Tiene sentido, es una muy buena aproximación a la probabilidades de la $U(10000)$,

5. Para el siguiente ejercicio también necesitamos R.

a. 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

Consideremos, 1: obtener una aguila, 0: obtener sol.

```r
set.seed(13)
resultados_modena <- c(0,1)
sum(sample(resultados_modena, 10, replace = TRUE))
```

```
## [1] 5
```

Ahora repetimos el proceso 10^6 veces.

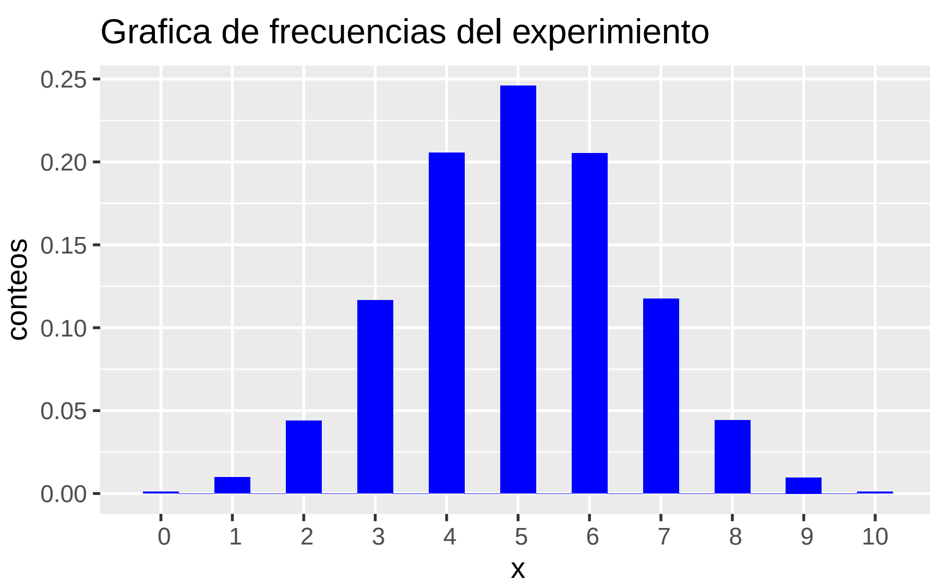
```r
simulacion_moneda_equilibrada <- c() # Inicializamos un vector
for(i in 1:10**6) { # Creamos un ciclo para crear las repeticiones.
 simulacion_moneda_equilibrada[i] <-sum(sample(resultados_modena, 10, replace = TRUE))
}
# Mostramos los primeros 3 resultados.
simulacion_moneda_equilibrada[1:3]
```

```
## [1] 8 5 4
```

Graficamos las frecuencias del número de aguilas obtenidas en cada uno de los experimentos.

```r
ggplot(data=data.frame(x=simulacion_moneda_equilibrada), aes(x))+
  geom_histogram(bins = 11, binwidth=.5, fill="blue")+
  scale_x_discrete(limits=seq(0,10,1))+
  labs(y="conteos", title="Grafica de frecuencias del experimiento")
```


```r
ggplot(data=data.frame(x=simulacion_moneda_equilibrada), aes(x))+
  geom_histogram(bins = 11, binwidth=.5, fill="blue", aes(y=stat(count) / sum(count)))+
  scale_x_discrete(limits=seq(0,10,1))+
  labs(y="conteos", title="Grafica de frecuencias del experimiento")
```

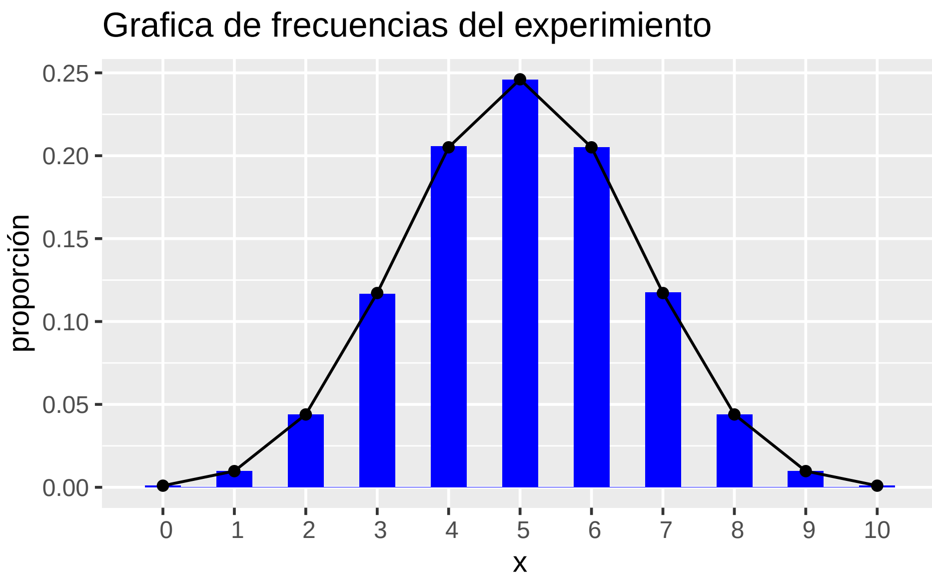


b. Usando la función dbinom grafique la función de masa de una distribución B(10, 0.5) sobre la gráfica de las proporciones que hizo en el inciso anterior.

RESPUESTA

```r
binomial<-data.frame(x=seq(0,10,1), y=dbinom(x=seq(0,10,1), size = 10,prob = 0.5))
ggplot(data=data.frame(x=simulacion_moneda_equilibrada), aes(x))+
```

```r
  geom_histogram(bins = 11, binwidth=.5, fill="blue", aes(y=stat(count)/sum(count)))+
  scale_x_discrete(limits=seq(0,10,1))+
  labs(y="proporción", title="Grafica de frecuencias del experimiento")+
  geom_line(data=binomial, aes(x,y))+
  geom_point(data=binomial, aes(x,y))
```



Podemos observar que el experimento se distribuye como una Binomial con los parametros de arriba.

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

Primero consideremos una moneda desequilibrado con probabilidad de 0.3 de obtener águila.

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

Simulamos 10 lanzamientos de la6 moneda, y contamos el número de águilas en la simulación:

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

```
## [1] 3
```

Ahora realizamos este experimento 10^6 veces.

```r
simulacion_monedas_desequilibrada <- c() #inicializamos el vector de resultados.
for(i in 1:10**6) {
simulacion_monedas_desequilibrada[i] <-sum(sample(lanzamientos_moneda_desequilibrada, 10, replace
4}
```

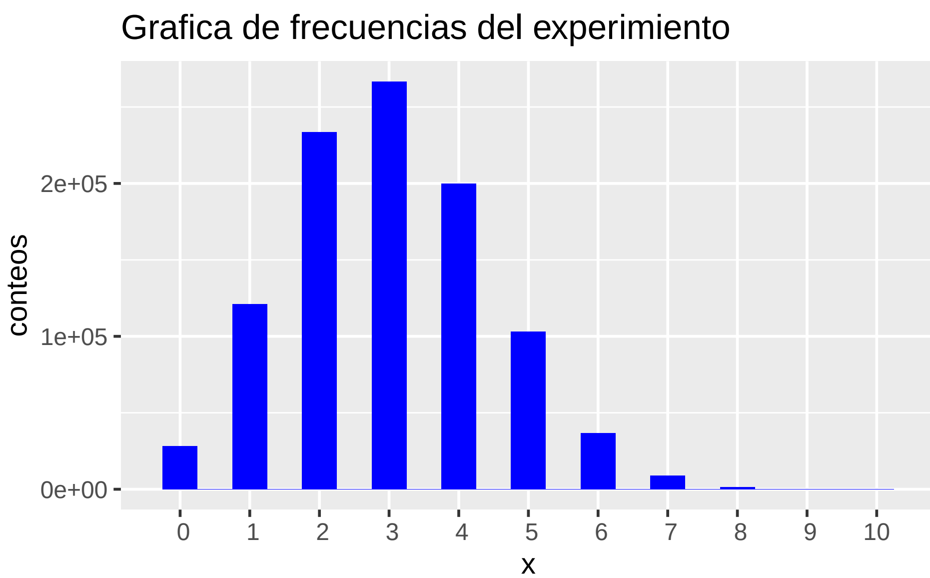
Mostramos los primeros 3 resultados:

```r
simulacion_monedas_desequilibrada[1:3]
```

```
## [1] 3 2 6
```

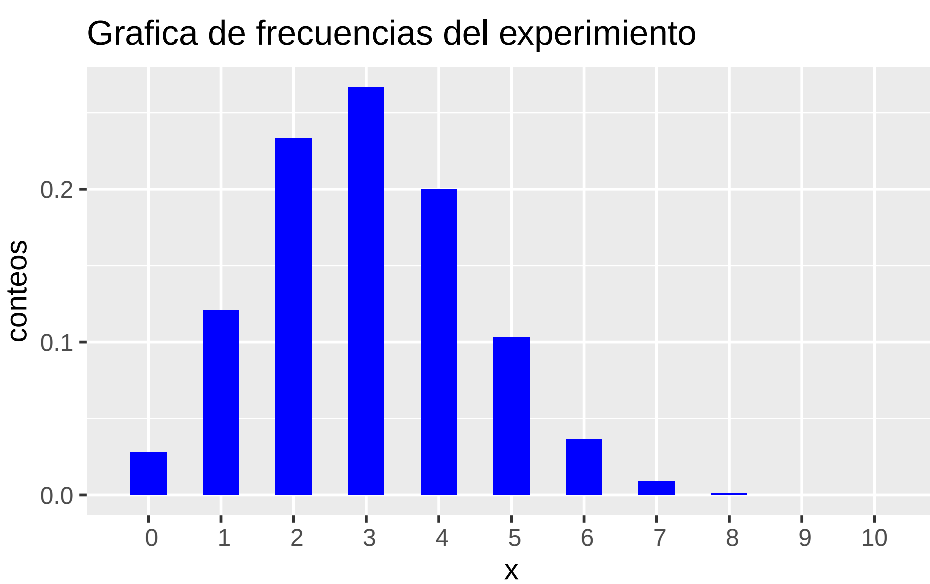
Graficamos las frecuencias del número de aguilas obtenidas en cada uno de los experimentos:

```
ggplot(data=data.frame(x=simulacion_monedas_desequilibrada), aes(x))+
geom_histogram(bins = 11, binwidth=.5, fill="blue")+
scale_x_discrete(limits=seq(0,10,1))+
labs(y="conteos", title="Grafica de frecuencias del experimiento")
```



Graficamos las proporciones del número de águilas obtenidas:

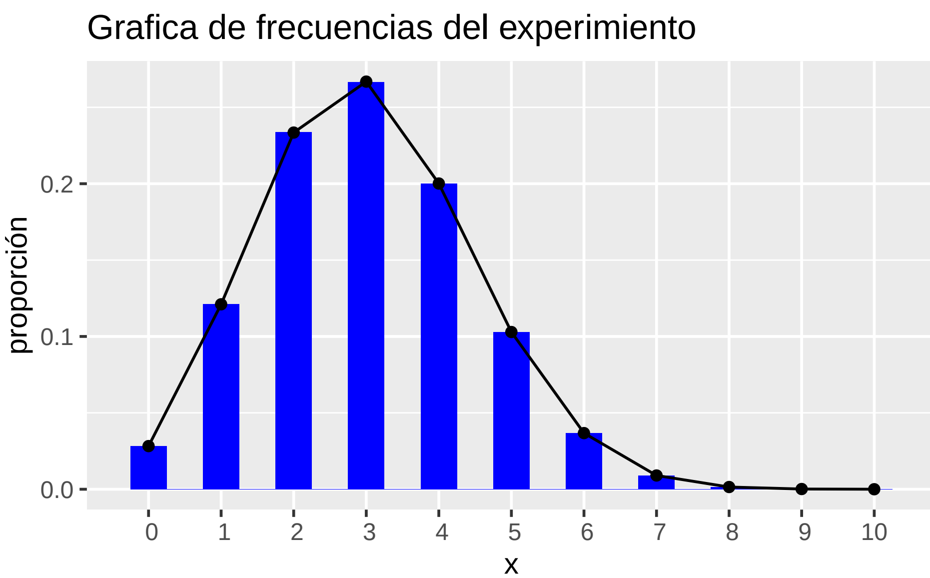
```
ggplot(data=data.frame(x=simulacion_monedas_desequilibrada), aes(x))+
geom_histogram(bins = 11, binwidth=.5, fill="blue", aes(y=stat(count)/sum(count)))+
scale_x_discrete(limits=seq(0,10,1))+
labs(y="conteos", title="Grafica de frecuencias del experimiento")
```



Usando la función dbinom graficamos la función de masa de una distribución $B(10, 0.5)$ sobre la gráfica anterior:

```
binomial<-data.frame(x=seq(0,10,1), y=dbinom(x=seq(0,10,1), size = 10,prob = 0.3))
ggplot(data=data.frame(x=simulacion_monedas_desequilibrada), aes(x))+
geom_histogram(bins = 11, binwidth=.5, fill="blue", aes(y=stat(count)/sum(count)))+
```

```
scale_x_discrete(limits=seq(0,10,1))+
labs(y="proporción", title="Grafica de frecuencias del experimiento")+
geom_line(data=binomial, aes(x,y))+
geom_point(data=binomial, aes(x,y))
```



De igual modo se observa que la distribución $Bin(10, 0.3)$ ajusta muy bien este experimento. Otro punto importante a resaltar es el parametro p, ya que si comparamos con la grafica del experimento considerando a $p = 0.5$ observamos que la grafica esta centrada en 5, pero si observamos esta grafica notamos que esta centrada en 3. Y se observa además como el parametro p influye al lado que esta sesgada la distribución, cuando p menor a 0.5 el sesgo es a la izquierda y si es mayor a 0.5 el sesgo es a la derecha.

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 experiementos anteriores y sobre esta figura añada la correspondiente función de masa de la distristibución Hipergeometrica asociada al experimento total.

RESPUESTA

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

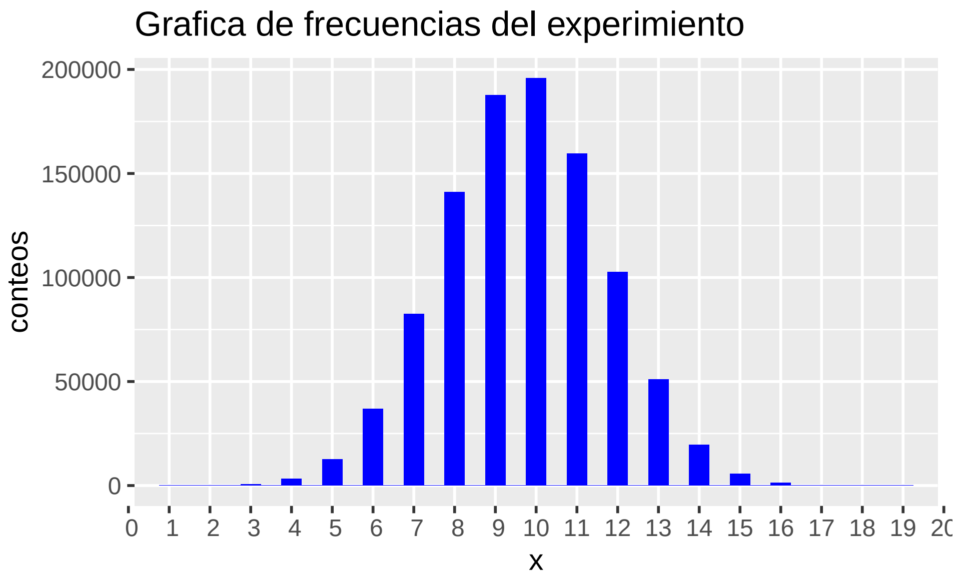
```
set.seed(13)
urna <- c(rep(1,46), rep(0,49)) # Creamos la urna con las bolas.
sum(sample(urna, 20, replace=FALSE)) # Contamos cuantas bolas grises hay en la muestra.
```

```
## [1] 7
```

```
simulacion_bolas <- c() #Inicializamos
for(i in 1:10**6) { # Repetimos el experimento de extraer una muestra.
 simulacion_bolas[i] <-sum(sample(urna, 20, replace=FALSE))
}
```

```
ggplot(data=data.frame(x=simulacion_bolas), aes(x))+
  geom_histogram(bins = 19, binwidth=.5, fill="blue")+
```

```r
  scale_x_discrete(limits=seq(0,20,1))+
  labs(y="conteos", title="Grafica de frecuencias del experimiento")
```

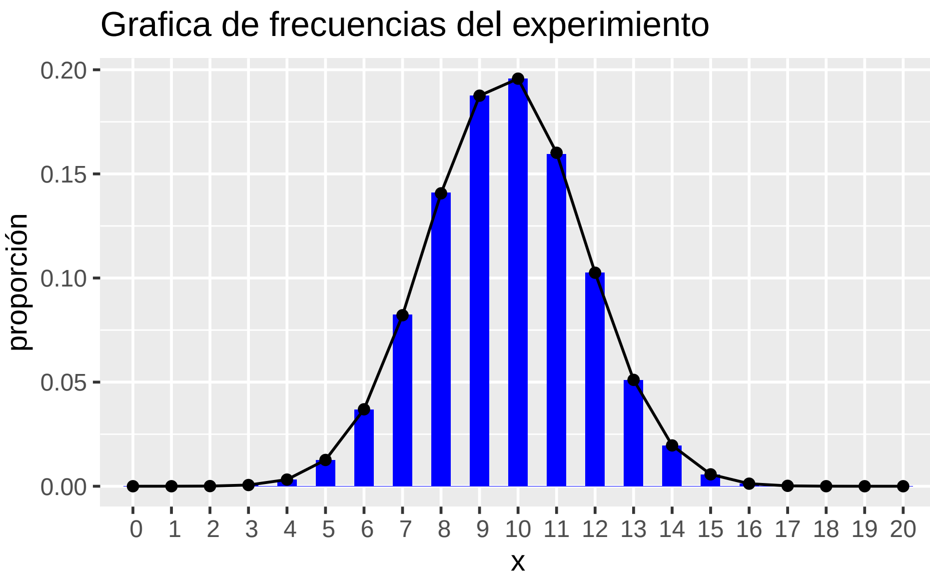


Si X es el número es de bolas grises que se obtienen en la extracción sin reemplazamiento de 20 bolas de la urna definida en el problema, entonces podemos decir que $X \sim Hyper(n = 20, M = 46, N = 95)$. Por lo tanto, la porbabilidad de que se extraigan 5 bolas grises es:

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

Graficamos la proporción de bolas grises:

```r
hyper<-data.frame(x=seq(0,20,1), y=dhyper(x=seq(0,20,1), m = 46, n=49, k=20))
ggplot(data=data.frame(x=simulacion_bolas), aes(x))+
  geom_histogram(bins = 21, binwidth=.5, fill="blue", aes(y=stat(count)/sum(count)))+
  scale_x_discrete(limits=seq(0,20,1))+
  labs(y="proporción", title="Grafica de frecuencias del experimiento")+
  geom_line(data=hyper, aes(x,y))+
  geom_point(data=hyper, aes(x,y))
```



Podemos observar que la distribución Hypergeometrica modela bien este experimento.

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:

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

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 $\exists\ F'(x)\ \forall x$. Por lo tanto, **la función la densidad de X es**

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

   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 que:

   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,

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

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 $X \sim Bin(10, p)$. 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 lotes 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 de X cuando cambia el valor p (ver Figura 3) 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
Para cada una de las proposiciones determinamos que el espacio muestral de los resultados es:

$$\Omega = \left\{ \begin{array}{llllll} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right.$$

Para probar la valides de las proposiciones ocuparemos la siguiente definición de independencia de dos eventos:

Sea A y B dos eventos, entonces son independientes si solo sí

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

i) A y F son independientes. **Verdadera**. Como A es el evento que el primer dado cae en H=$\{4, 5, 6\}$,

$$\mathbb{P}(A) = \frac{\text{Número de veces que sucede A}}{\text{Número total de eventos}} = \frac{18}{36} = \frac{1}{2}.$$

Ahora, como F es el evento en que el total es siete tenemos que

$$\mathbb{P}(F) = \frac{\text{Número de veces que sucede F}}{\text{Número total de eventos}} = \frac{6}{36} = \frac{1}{6}.$$

Y ahora calculemos la probabilidad de la intersección de los eventos A y F:

$$\mathbb{P}(A \cap F) = \frac{\text{Número de veces que sucede A y F}}{\text{Número total de eventos}} = \frac{3}{36} = \frac{1}{12}.$$

Por lo tanto, como

$$\mathbb{P}(A \cap F) = \frac{1}{12} = \frac{1}{2} \cdot \frac{1}{6} = \mathbb{P}(A)\mathbb{P}(F)$$

podemos concluir que A y F son independientes.

ii) A y D son independientes. **Falsa**. En el inciso $i)$ ya obtuvimos $\mathbb{P}(A)$. Ahora, como D es el evento en que el total es cuatro tenemos que

$$\mathbb{P}(D) = \frac{\text{Número de veces que sucede D}}{\text{Número total de eventos}} = \frac{3}{36} = \frac{1}{12}.$$

Y ahora calculemos la probabilidad de la intersección de los eventos A y D:

$$\mathbb{P}(A \cap D) = \frac{\text{Número de veces que sucede A y D}}{\text{Número total de eventos}} = \frac{0}{36} = 0.$$

Por lo tanto, como

$$\mathbb{P}(A \cap D) = 0 \neq \frac{1}{24} = \frac{1}{2} \cdot \frac{1}{12} = \mathbb{P}(A)\mathbb{P}(D)$$

podemos concluir que A y D no son independientes.

iii) A y E son independientes. **Falsa**. En el inciso $i)$ ya obtuvimos $\mathbb{P}(A)$. Ahora, como E es el evento en que el total es cinco tenemos que

$$\mathbb{P}(E) = \frac{\text{Número de veces que sucede E}}{\text{Número total de eventos}} = \frac{4}{36} = \frac{1}{9}.$$

Y ahora calculemos la probabilidad de la intersección de los eventos A y E:

$$\mathbb{P}(A \cap E) = \frac{\text{Número de veces que sucede A y E}}{\text{Número total de eventos}} = \frac{1}{36}.$$

Por lo tanto, como

$$\mathbb{P}(A \cap E) = \frac{1}{36} \neq \frac{1}{18} = \frac{1}{2} \cdot \frac{1}{9} = \mathbb{P}(A)\mathbb{P}(E)$$

podemos concluir que A y D no son independientes.

iv) $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. **Falsa**. En el inciso $i)$ ya obtuvimos $\mathbb{P}(A)$. Ahora, como B es el evento en que el segundo dado cae en H tenemos que

$$\mathbb{P}(B) = \frac{\text{Número de veces que sucede B}}{\text{Número total de eventos}} = \frac{18}{36} = \frac{1}{2}.$$

Como C es el evento en que un dado cae en G y el otro en H tenemos que

$$\mathbb{P}(C) = \frac{\text{Número de veces que sucede C}}{\text{Número total de eventos}} = \frac{18}{36} = \frac{1}{2}.$$

Y ahora calculemos la probabilidad de la intersección de los eventos A, B y C:

$$\mathbb{P}(A \cap B \cap C) = \frac{\text{Número de veces que sucede A, B y C}}{\text{Número total de eventos}} = \frac{0}{36} = 0.$$

Por lo tanto, como

$$\mathbb{P}(A \cap B \cap C) = 0 \neq \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

no se cumple la igualdad.

v) A y C son independientes.**Verdadera**. En el inciso i) y iv) calculamos $\mathbb{P}(A)$ y $\mathbb{P}(C)$ respectivamente. Y ahora calculemos la probabilidad de la intersección de los eventos A y C:

$$\mathbb{P}(A \cap C) = \frac{\text{Número de veces que sucede A y C}}{\text{Número total de eventos}} = \frac{9}{36} = \frac{1}{4}.$$

Por lo tanto, como

$$\mathbb{P}(A \cap C) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A)\mathbb{P}(C)$$

podemos concluir que A y C son independientes.

vi) C y E son independientes.**Verdadera**. En el inciso iii) y iv) calculamos $\mathbb{P}(E)$ y $\mathbb{P}(C)$ respectivamente. Y ahora calculemos la probabilidad de la intersección de los eventos E y C:

$$\mathbb{P}(E \cap C) = \frac{\text{Número de veces que sucede E y C}}{\text{Número total de eventos}} = \frac{2}{36} = \frac{1}{18}.$$

Por lo tanto, como

$$\mathbb{P}(E \cap C) = \frac{1}{18} = \frac{1}{9} \cdot \frac{1}{2} = \mathbb{P}(E)\mathbb{P}(C)$$

podemos concluir que E y C son independientes.

vii) $\mathbb{P}(A \cap C \cap E) = \mathbb{P}(A)\mathbb{P}(C)\mathbb{P}(E)$. **Verdadera**. En el inciso iii) y iv) calculamos $\mathbb{P}(E)$, $\mathbb{P}(C)$ y $\mathbb{P}(A)$ respectivamente. Y ahora calculemos la probabilidad de la intersección de los eventos A, C y E:

$$\mathbb{P}(A \cap C \cap E) = \frac{\text{Número de veces que sucede A, C y E}}{\text{Número total de eventos}} = \frac{1}{36} = \frac{1}{18}.$$

Por lo tanto, como

$$\mathbb{P}(A \cap C \cap E) = \frac{1}{36} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{9} = \mathbb{P}(A)\mathbb{P}(C)\mathbb{P}(E)$$

podemos concluir se cumple la igualdad.

viii) A, C y E son independientes. **Falsa**. Utitlizaremos el siguiente criterio de independencia:

Sea A, C, E eventos son independientes si solo si

$$\begin{aligned}
\mathbb{P}(A \cap C \cap E) &= \mathbb{P}(A)\mathbb{P}(C)\mathbb{P}(E),\\
\mathbb{P}(A \cap C) &= \mathbb{P}(A)\mathbb{P}(C),\\
\mathbb{P}(A \cap E) &= \mathbb{P}(A)\mathbb{P}(E),\\
\mathbb{P}(C \cap E) &= \mathbb{P}(C)\mathbb{P}(E).
\end{aligned}$$

Considerando el resultado del inciso iii), en donde concluimos que $\mathbb{P}(A \cap E) \neq \mathbb{P}(A)\mathbb{P}(E)$. Entonces podemos concluir que A, C y E no son independientes. ■.

En estos ejercicios el inciso i) me pareció interesante, ya que si lo comparamos las independencia con los incisos ii) y iii) lo único que varía es el total y cambia completamente la independencia de los eventos.

Ejercicios de las notas:

1. Sea $X \sim Poisson(\lambda)$, demuestre que $C_X(t) = \log M_x(t), C'_X(0) = \mu, C''_X(0) = \sigma^2$.

RESPUESTA

Recordemos que la función generadora de momentos para $X \sim Poisson(\lambda)$ es

$$M_X(t) = e^{\lambda(e^t - 1)}.$$

Entonces

$$C_X(t) = \log M_x(t) = \log(e^{\lambda(e^t-1)}) = \lambda(e^t - 1).$$

Calculemos la primera y segunda derivada de $C_X(t)$:

$$C'_X(t) = \lambda e^t,$$

$$C''_X(t) = \lambda e^t.$$

Evaluando las derivadas en 0, tenemos que:

$$C'_X(0) = \lambda e^0 = \lambda,$$

$$C''_X(0) = \lambda e^0 = \lambda.$$

Y como $X \sim Poisson(\lambda)$ podemos concluir que:

$$C'_X(0) = \lambda = \mu,$$

$$C''_X(0) = \lambda = \sigma. \quad \blacksquare.$$

2. El factor $\frac{N-n}{N-1}$, cantidad que nos recuerda la finitud de N, es llamada el **factor de correción para la población finita**. ¿Cuál es $\lim_{N\to\infty} \frac{N-n}{N-1}$ con n fija?

   RESPUESTA
   Ocupando algunas propiedades de los límites, tenemos

   $$\lim_{N\to\infty} \frac{N-n}{N-1} = \lim_{N\to\infty} \frac{\frac{N-n}{N}}{\frac{N-1}{N}} = \lim_{N\to\infty} \frac{1-\frac{n}{N}}{1-\frac{1}{N}} = 1.$$

   Esto se puede interpretar como que cuando N es grande el factor correción tiene a 1, es decir, cuando N es grande el factor ya no tiene importancia. $\blacksquare$.