Centro de Investigación en Matemáticas (CIMAT). Unidad Monterrey

Inferencia Estadística (Tarea 2)

Ricardo Cruz Sánchez

10 de septiembre de 2018

1. Ejercicio 1

Una pareja decide tener hijos hasta el nacimiento de la primera niña. Calcule la probabilidad de que tengan más de 4 hijos. Suponga que las probabilidades de tener niño o niña son iguales. ¿Cuál es el tamaño esperado de la familia?

Respuesta: Este problema se puede modelar con una distribución Geometrica. Sea X la variable aleatoria que indica el número de hijos que tienen hasta el nacimiento de la primera niña, entonces, $X \sim Geom(,5)$. Por lo tanto, calcular la probabilidad de que tengan mas de 4 hijos es:

$$\begin{aligned} P(X > 4) &= 1 - P(X \leq 4) \\ &= 1 - (P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)) \\ &= 1 - (,5 + ,5^2 + ,5^3 + ,5^4) \\ &= 1 - (,5 + ,25 + ,125 + ,0625) \\ &= 1 - (0,9375) \\ &= 0,0625 \end{aligned}$$

Por lo tanto, la probabilidad que tengan más de 4 hijos es 0.125

Ahora, el tamaño esperado de la familia es la esperanza de la variable aletoria más los dos padres.

$$
\begin{aligned}
\mathbb{E}(X) &= \frac{1}{p} \\
&= \frac{1}{,5} \\
&= 2
\end{aligned}
$$

Por lo tanto, el tamaño esperado de la familia es 4, 2 padres, 1 hijo y 1 hija.

2. Ejercicio 2

Cuando una máquina no se ajusta adecuadamente tiene una probabilidad 0.15 de producir un artículo defectuoso. Diariamente, la máquina trabaja hasta que se producen 3 artículos defectuosos. Se detiene la máquina y se revisa para ajustarla. ¿Cuál es la probabilidad de que una máquina mal ajustada produzca 5 o más artículos antes de que sea detenida? ¿Cuál es el número promedio de artículos que la máquina producirá antes de ser detenida?

Respuesta: El número de artículos producidos hasta encontrar 3 artículos defectuosos se puede modelar con una variable aleatoria $X \sim BN(3, 0{,}15)$

Por lo tanto, la probabilidad de producir 5 o más artículos hasta que se encuentren 3 defectuosos corresponde a:

$$
\begin{aligned}
P(X \geq 5) &= 1 - P(X < 5) \\
&= 1 - (P(X = 3) + P(X = 4)) \\
&= 1 - (0{,}003375 + 0{,}00860625) \\
&= 0{,}9880187
\end{aligned}
$$

La probabilidad de producir 5 o más artículos, antes de que se detenga la máquina es 0,0988

Ahora, el número promedio de artículos producidos hasta que se presenta el e artículo defectuoso es:

$$
\begin{aligned}
\mathbb{E}(X) &= \frac{r}{p} \\
&= \frac{3}{,15} \\
&= 20
\end{aligned}
$$

Es decir, se espera producir 20 artículos, antes de que se detenga la máquina.

3. Ejercicio 3

Los empleados de una compañía de aislantes son sometidos a pruebas para detectar residuos de asbesto en sus pulmones. Se le ha pedido a la compañía que envíe a tres empleados cuyas pruebas resulten positivas, a un centro médico para realizarles más análisis. Si se sospecha que el 40 % de los empleados tienen residuos de asbesto en sus pulmones, encuentre la probabilidad de que deban ser analizados 10 trabajadores para poder encontrar a 3 con resultado positivo.

Respuesta: El número de trabajadores que deben ser analizados hasta encontrar al 3 trabajador con resultado positivo se puede modelar con la variable aleatoria $X \sim BN(3, ,4)$. Por lo tanto, la probabilidad de que se analicen 10 trabajadores hasta encontrar a 3 con resultado positivo es:

$$P(X = 10) = 0{,}06449725$$

La probabilidad de que deban ser analizados 10 trabajadores hasta encontrar el tercero con resultado positivo es 0,06449

4. Ejercicio 4

Para el siguiente ejercicio es necesario usar R.

1. Considere una moneda desequilibrada que tiene probabilidad p de obtener águila. Usando el comando sample, escriba una función que simule N veces lanzamientos de esta moneda hasta obtener un águila. La función deberá recibir como parámetros a la probabilidad p de obtener águila y al número N de veces que se repite el experimento; y tendrá que regresar un vector de longitud N que contenga el número de lanzamientos hasta obtener un águila en cada uno de los N experimentos.

   Respuesta: El código correspondiente se encuentra en el archivo *codigostarea2.R*

2. Usando la función anterior simule $N = 10^4$ veces una variable aleatoria $Geom(p)$ para $p = 0{,}5; 0{,}1; 0{,}01$ Grafíque las frecuencias normalizadas en color azul. Sobre está última figura empalme en rojo la gráfica de la función de masa correspondiente. ¿Qué observa?

   Respuesta: Las figuras 4.1,4.2 y 4.3 muestra las gráficas solicitadas, las cuales se generaron desde el mismo archivo de R. En las gráficas se puede observar que las proporciones para $p = 0{,}5$ son muy similares a la distribución geométrica asociada, pero, no así para $p = 0{,}1$ y $p = 0{,}01$, en las gráficas correspondientes, se observa como hay valores (barras) que difieren considerablemente de los valores teóricos. Esto se debe a que el tamaño de muestra no es suficiente para las últimos dos casos. Además, conforme p decrece, se observa que la intensidad con la que decae la curva disminuye.

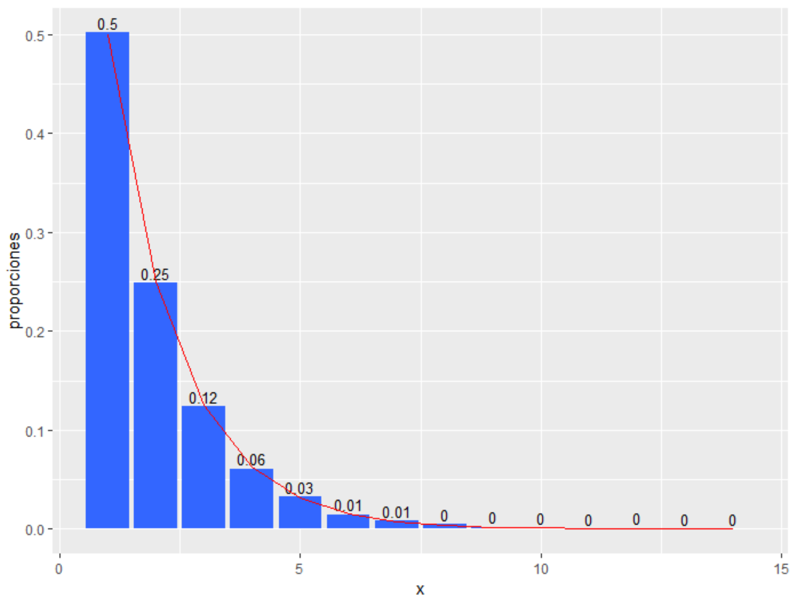

Figura 4.1: $N = 10^4$ $p = 0{,}5$

3. Repita el inciso anterior para $N = 10^6$. Además calcule el promedio y la desviación estándar de las simulaciones que realizó ¿Qué observa?

   Respuesta: La figura 4.4,4.5 y 4.6 muestra las gráficas solicitadas, las cuales se generaron desde el mismo archivo de R. En las gráficas se puede observar que las proporciones para $p = 0{,}5$ prácticamente no cambiaron a comparación del inciso anterior, esto porque el tamaño de muestra ya era suficiente y aumentarlo solo mejoraría la precisión de la aproximación. Mientra que para $p = 0{,}1$ y $p = 0{,}01$, las gráficas mejoraron significativamente y ahora ajustan de manera casi perfecta a la distribución geométrica.

   Para $p = {,}5; {,}1; {,}01$ las medias son 1,997821; 9,995754 y 99,99807 respectivamente, mientras que las varianzas son 1,988252; 89,89035 y 9930,066. Lo anterior reafirma que la simulación fue suficiente, pues los valores de la media y varianza se asemejan mucho a los valores teóricos en cada caso.

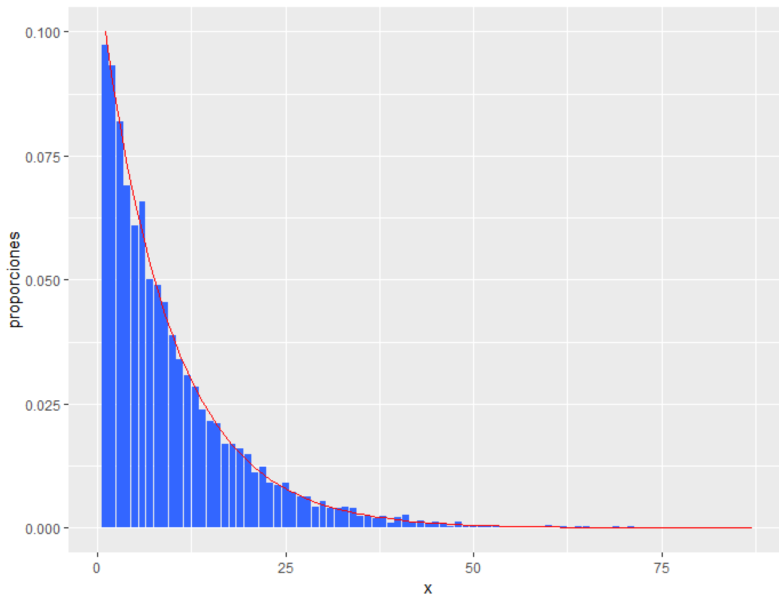

Figura 4.2: $N = 10^4$ $p = 0{,}1$

5. Ejercicio 5

Usando las ideas del inciso anterior escriba una función en R que simule N veces los lanzamientos de moneda hasta obtener r águilas. La función deberá recibir como parámetros a la probabilidad p de obtener águila, al número r de águilas a observar antes de detener el experimento y al número N de veces que se repite el experimento; y tendrá que regresar un vector de longitud N que contenga el número de lanzamientos hasta obtener las r águilas en cada uno de los N experimentos. Grafique las frecuencias normalizadas de los experimentos para $N = 10^6, p = 0,2; 0,1$ y $r = 2; 7$ y compárelos contra la función de masa de la distribución más adecuada para modelar este tipo de experimentos.

Respuesta: El código correspondiente se encuentra en el archivo *codigostarea2.R*. Las gráficas generadas por el código se muestran en las figuras 5.1,5.2,5.3 y 5.4, donde se observa que el tamaño de la muestra es suficiente para ajustar la curva de una distribución binomial negativa. También, se percibe como el parámetro r desplaza a la derecha a la curva a medida que este crece y el parámetro p determina la escala de la curva asociada.

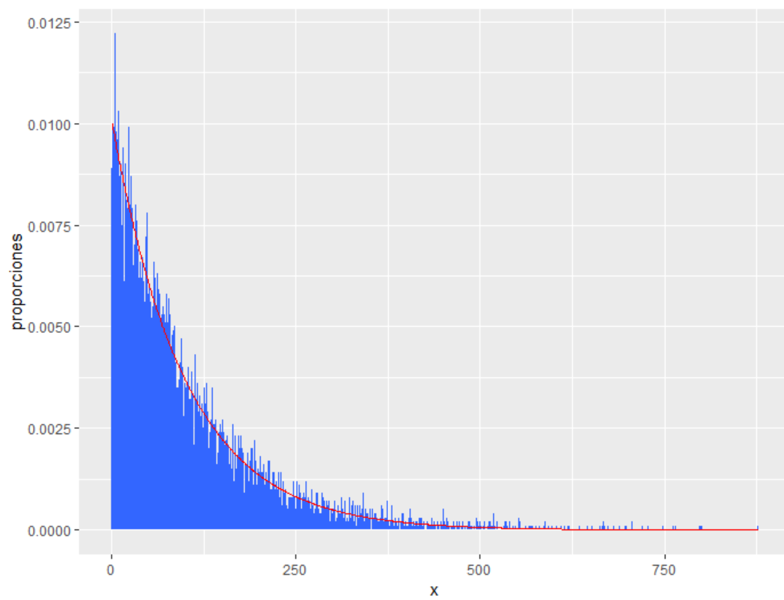

Figura 4.3: $N = 10^4$ $p = 0{,}01$

6. Ejercicio 6

Considera X una v.a. con función de distribución F y función de densidad f, y sea A un intervalo de la línea real $\mathbb{R}$. Definimos la función indicadora $\mathbb{I}_A(x)$:

$$\mathbb{I}_A(x) = \begin{cases} 1 \quad si \quad x \in A \\ \\ 0 \quad e.o.c. \end{cases}$$

Sea $Y = \mathbb{I}_A(X)$. Encuentre una expresión para la distribución acumulada de Y.

Respuesta: Como $\mathbb{I}_A(X)$ toma valores 0 o 1, entonces el soporte de Y es $0, 1$.

Ahora,

$$f_Y(y) = P(Y = y) = P(\mathbb{I}_A(X) = y) \qquad\qquad y = 0, 1$$

Además, se tiene que

$$\mathbb{I}_A(y) = 1 \Leftrightarrow X \in A$$

$$\mathbb{I}_A(y) = 0 \Leftrightarrow X \notin A$$

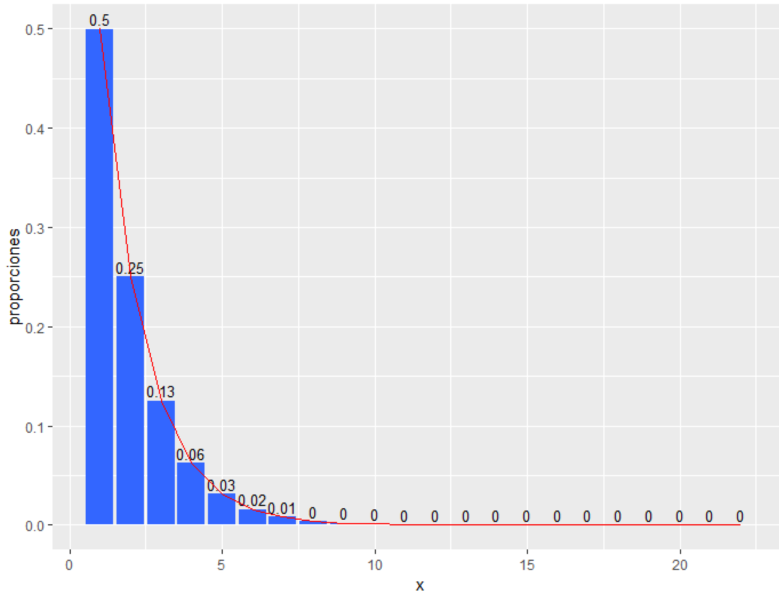

Figura 4.4: $N = 10^6$ $p = 0{,}5$

Sin perdida generalidad, sea $A = [a, b]$, entonces

$$f_Y(y) = P(\mathbb{I}_A(X) = y) = \begin{cases} P(a \leq X \leq b) & si \quad y = 1 \\ P(a \geq X) + P(b \leq X) & si \quad y = 0 \\ 0 & e.o.c. \end{cases}$$

Por lo tanto, la función de masa de probabilidad es (notese que efectivamente suma cero y sus valores están entre 0 y 1:

$$f_Y(y) = \begin{cases} F_X(b) - F_X(a) & si \quad y = 1 \\ F_X(a) + 1 - F_X(b) & si \quad y = 0 \\ 0 & e.o.c. \end{cases}$$

Una vez calculada la función de masa de probabilidad, la función de distribución acumulada es sumar sobre su soporte:

$$F_Y(y) = \begin{cases} 0 & si \quad y < 0 \\ F_X(a) + 1 - F_X(b) & si \quad 0 \leq y < 1 \\ 1 & si \quad y \geq 1 \end{cases}$$

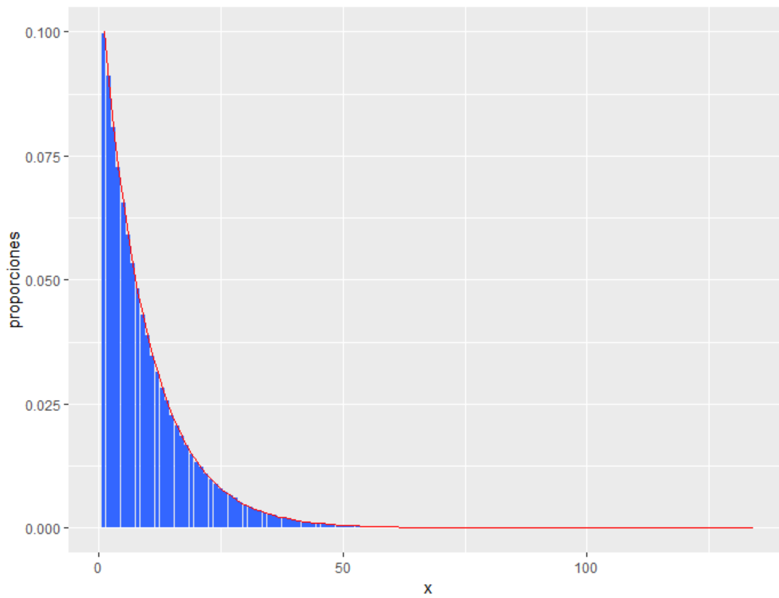

Figura 4.5: $N = 10^6$ $p = 0{,}1$

7. Ejercicio 7.

Entre las más famosas de todas las lluvias de meteoros están las Perseidas, que ocurren cada año a principios de agosto. En algunas áreas, la frecuencia de Perseidas visibles promedian seis por cada cuarto de hora. El modelo de probabilidad que describe a Y, el número de meteoros que una persona ve en un cuarto de hora, tiene la función de probabilidad

$$f_Y(y) = \frac{e^{-6}6^y}{y!} \qquad\qquad y = 0, 1, \ldots$$

Encuentre la probabilidad de que una persona vea en un cuarto de hora determinado al menos la mitad de los meteoros que esperaría ver.

Respuesta: La variable aleatoria Y sigue una distribución $Poisson(6)$. Por lo tanto, el valor esperado es 6 y el problema se reduce a calcular:

$$P(Y \geq \frac{6}{2}) = P(Y \geq 3) = 1 - P(Y \leq 2)$$

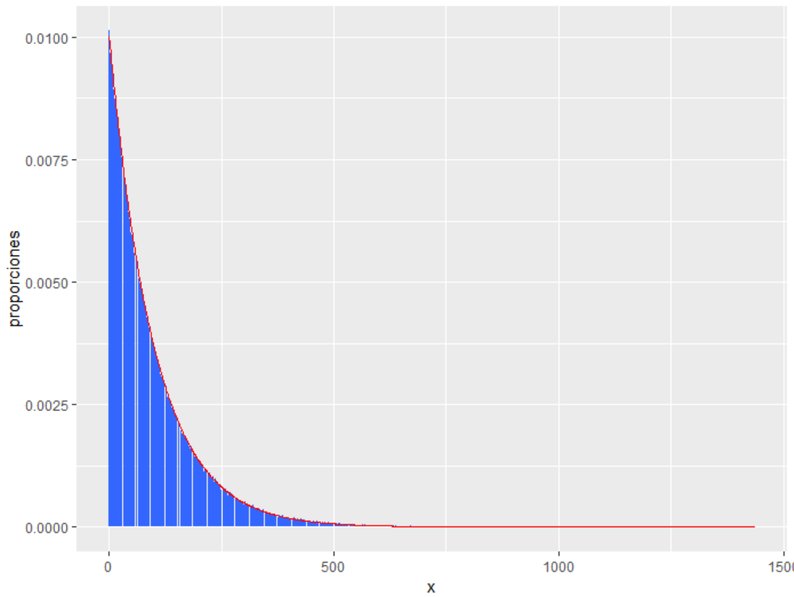

Figura 4.6: $N = 10^6$ $p = 0{,}01$

Desarrollando tenemos que:

$$
\begin{aligned}
1 - P(Y \leq 2) &= 1 - P(Y = 2) - P(Y = 1) - P(Y = 0) \\
&= 1 - \frac{e^{-6}6^2}{2} - \frac{e^{-6}6}{1} - \frac{e^{-6}}{1} \\
&= 1 - 0{,}0619688 \\
&= 0{,}9380312
\end{aligned}
$$

Por lo tanto, la probabilidad de que una persona vea en un cuarto de hora determinado al menos 3 meteoros es 0,938, es decir, es altamente probable que vea 3 o más meteoros.

8. Ejercicio 8.

Las calificaciones de un estudiante de primer semestre en un examen de química se describen por la densidad de probabilidad

$$f_Y(y) = 6y(1 - y) \qquad\qquad 0 \leq y \leq 1$$

donde y representa la proporción de preguntas que el estudiante contesta correctamente. Cualquier calificación menor a 0,4 es reprobatoria. Responda lo siguiente:

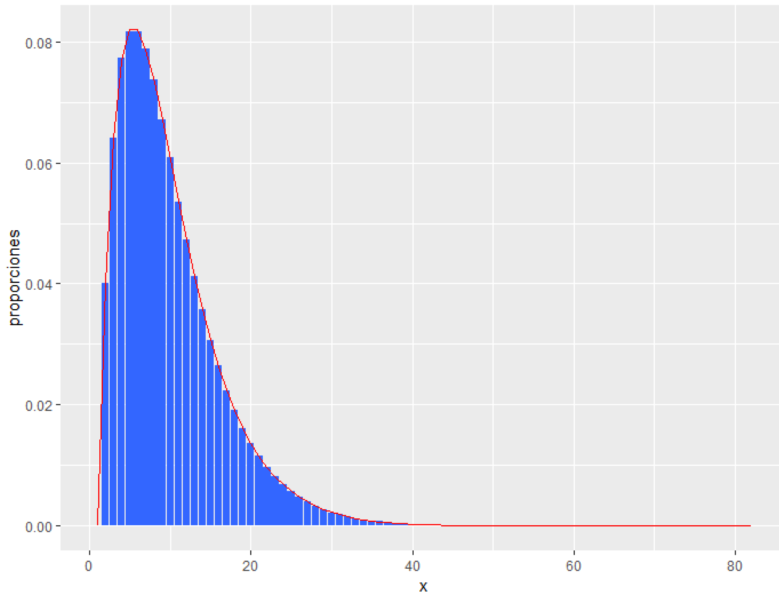

Figura 5.1: $N = 10^6$, $r = 2$, $p = 0{,}2$

1. ¿Cuál es la probabilidad de que un estudiante repruebe?

   Respuesta: La probabilidad de que un alumno repruebe se puede expresar como $P(Y \leq 0{,}4)$. Por lo tanto:

$$\begin{aligned} P(Y \leq 0{,}4) &= \int_0^{0{,}4} 6y(1-y)dy \\ &= \int_0^{0{,}4} 6y - 6y^2 dy \\ &= 3y^2 - 2y^3 \mid_{y=0}^{y=0{,}4} \\ &= 0{,}48 - 0{,}128 \\ &= 0{,}352 \end{aligned}$$

   Por lo tanto, la probabilidad de que un alumno repruebe es de 0,352

2. Si 6 estudiantes toman el examen, ¿cuál es la probabilidad de exactamente 2 reprueben?

   Respuesta: Suponiendo que la calificación de cada alumno es independiente del resto, entonces, $X \sim Bin(6, ,352)$ modela el número de reprobados de un grupo de 6

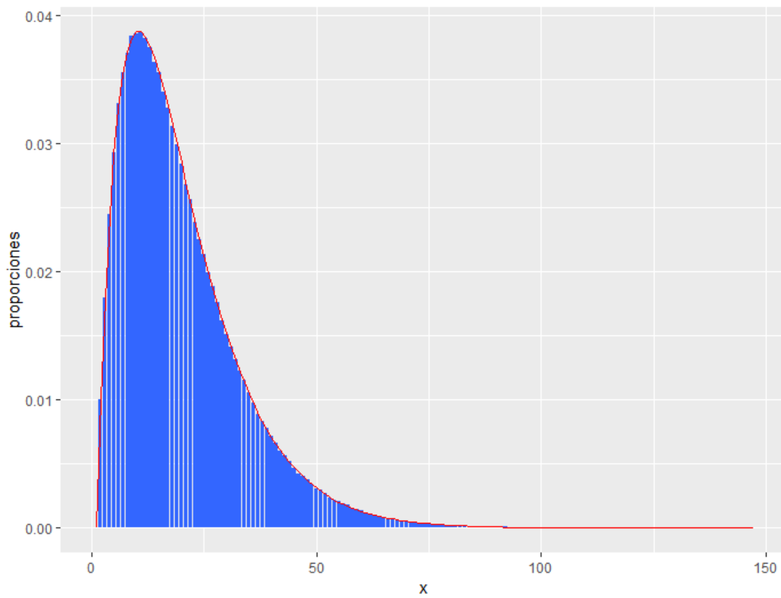

Figura 5.2: $N = 10^6$, $r = 2$, $p = 0{,}1$

alumnos. Nos interesa calcular $P(X = 2)$

$$P(X = 2) = \binom{6}{2} p^2 q^4 = 0{,}3277001$$

Por lo tanto, la probabilidad de que 2 alumnos, de un grupo de 6 repruebe es de 0,327

9. Ejercicio 9.

Escriba una función en R que simule una aproximación al proceso Poisson a partir de las 5 hipótesis que usamos en clase para construir tal proceso. Usando esta función, simule tres trayectorias de un proceso Poisson $\lambda = 2$ sobre el intervalo $[0, 10]$ y grafíquelas. Además, simule 10^4 veces un proceso de Poisson N con $\lambda = \frac{1}{2}$ y hasta el tiempo $t = 1$. Haga un histograma de $N(1)$ en su simulación anterior y compare contra la distribución de Poisson correspondiente.
Hint: Considere el intervalo $[0, T]$ y un número real positivo dt que sea mucho más pequeño que la longitud de $[0, T]$ y que divida dicha longitud, digamos $T = dt = 1000$. Divida el intervalo $[0, T]$ en intervalitos de longitud dt que tengan la forma $(k * dt, (k + 1) * dt]$, $k = 0, 1, 2, \ldots, (T = dt * 1)$. Para cada uno de estos intervalitos simule una v.a. $Bernoulli(\lambda * dt + 10^6)$ y guarde su resultado en un vector del tamaño adecuado.

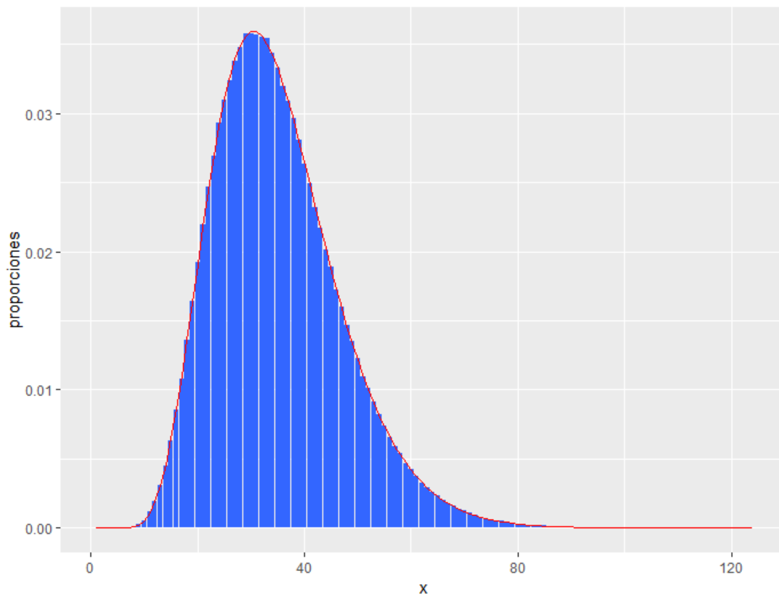

Figura 5.3: $N = 10^6$, $r = 7$, $p = 0{,}2$

Respuesta: Los códigos para generar las gráficas 9.1, 9.2, 9.3 y 9.4 se encuentran en el archivo *codigostarea2.R*. En las primeras 3 gráficas, se observa 3 diferentes trayectorias de un proceso $Poisson(,5)$, donde en promedio el número de eventos es $20 = \lambda * t = 2 * 10$ como se esperaba.

En la cuarta figura se aprecia que el histograma de un proceso Poisson con $\lambda = ,5$ ajusta casi perfectamente con 10^4 simulaciones a una distribución $Poisson(,5)$.

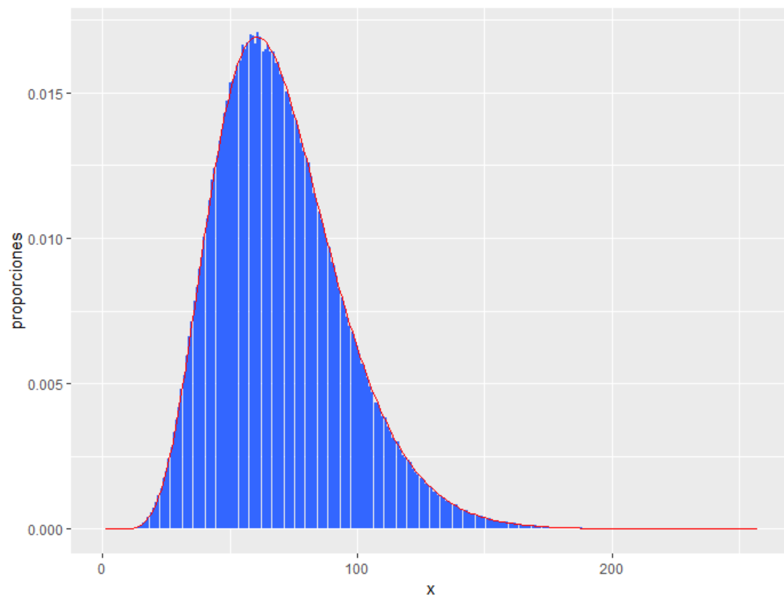

Figura 5.4: $N = 10^6$, $r = 7$, $p = 0{,}1$

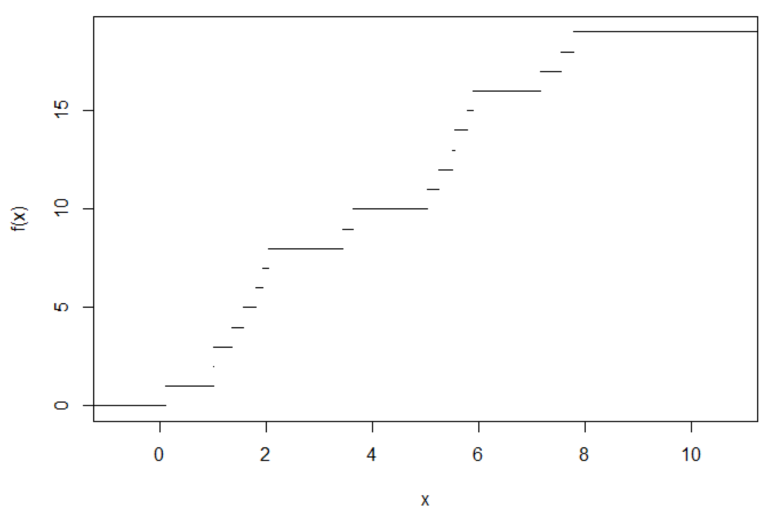

Figura 9.1: Primera trayectoria del proceso $Poisson(2)$

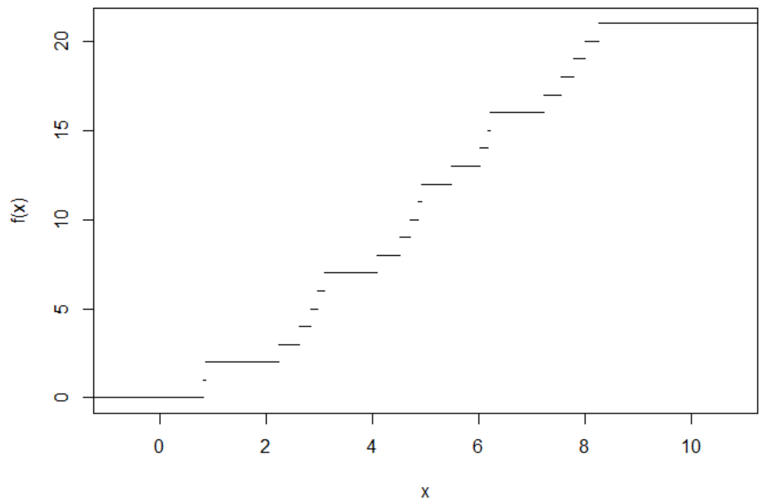

Figura 9.2: Segunda trayectoria del proceso $Poisson(2)$

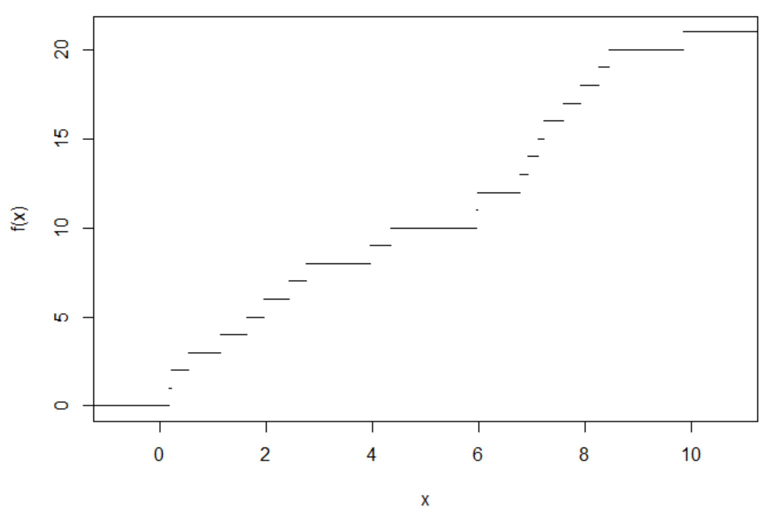

Figura 9.3: Tercera trayectoria del proceso $Poisson(2)$

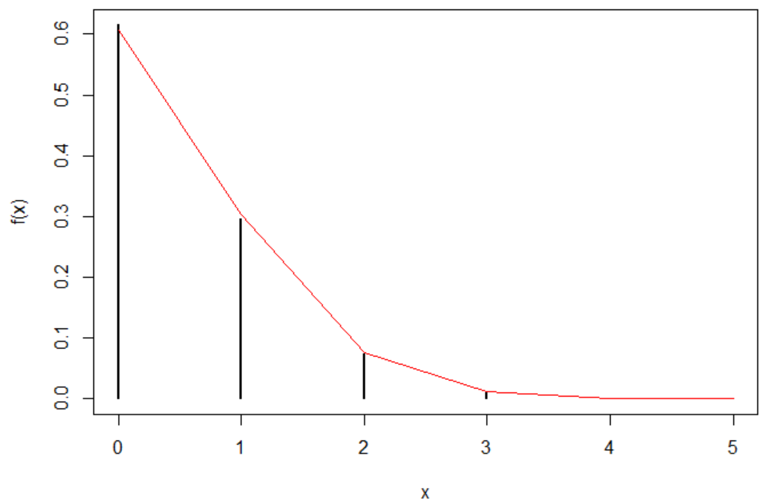

Figura 9.4: Histograma y curva ajustada del proceso $Poisson(,5)$