

CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS (CIMAT).
UNIDAD MONTERREY
INFERENCIA ESTADÍSTICA

Tarea 2

Gustavo Hernández Angeles

8 de septiembre de 2024

1 Problema 1

Una pareja decide tener hijos hasta el nacimiento de su 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?

SOLUCIÓN

Sea $X = \#$ de hijos hasta obtener una niña, la variable aleatoria de nuestro interés. Podemos decir que $X \sim \text{Geom}(p = 0.5)$ ya que **no pondremos restricción** alguna sobre la cantidad de hijos que la pareja pueda tener, buscamos únicamente un “éxito” (una niña), además de suponer una **independencia** entre el sexo de cada hijo.

Buscando la probabilidad de que tengan más de 4 hijos, $P(X > 4)$:

$$\begin{aligned}
 P(X > 4) &= P(X = 5) + P(X = 6) + \dots \\
 &= \sum_{x=5}^{\infty} (0.5)^{x-1} (0.5)^x = \sum_{x=5}^{\infty} (0.5)^x \\
 &= \sum_{x=5}^{\infty} (0.5)^x = \sum_{y=0}^{\infty} (0.5)^{y+5} \quad ; \text{ con } y = x - 5 \\
 &= 0.5^5 \sum_{y=0}^{\infty} (0.5)^y = \frac{0.5^5}{1 - 0.5} = 0.5^4 \quad \text{calculando la serie geométrica con } |r| < 1 \\
 &= 0.0625
 \end{aligned}$$

∴ La probabilidad de que tengan más de 4 hijos es del 6.25 %.

Ahora calcularemos el valor esperado del número de hijos X para obtener el tamaño esperado de la familia. Calculando $E(x)$:

$$\begin{aligned}
 E(X) &= \frac{1}{p} = \frac{1}{0.5} \\
 &= 2
 \end{aligned}$$

Con el número esperado de hijos siendo 2, podemos concluir que el **tamaño esperado de la familia es de 4**, contando a la pareja.

2 Problema 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?

SOLUCIÓN

Sea $X = \#$ de artículos fabricados hasta encontrar 3 defectuosos, la variable aleatoria de nuestro interés. Suponiendo aquí que la probabilidad de que un artículo salga defectuoso se mantiene constante siempre y que no hay fallas sistemáticas, con cada artículo defectuoso independiente, podemos modelar a X con una distribución NB($r = 3, p = 0.15$).

Para la primera pregunta, buscamos la probabilidad de que la máquina produzca 5 o más artículos antes de que sea detenida (encontrar los 3 artículos defectuosos). Entonces buscamos $P(X \geq 5)$:

$$\begin{aligned}
 P(X \geq 5) &= 1 - P(X \leq 4) = 1 - \left[P(X = 3) + P(X = 4) \right] \\
 &= 1 - \left[\binom{3-1}{3-1} (1-0.15)^{3-3} (0.15)^3 + \binom{4-1}{3-1} (1-0.15)^{4-3} (0.15)^3 \right] \\
 &= 1 - \left[\binom{2}{2} (0.85)^0 (0.15)^3 + \binom{3}{2} (0.85) (0.15)^3 \right] \\
 &= 1 - \left[(0.15)^3 + 3(0.85)(0.15)^3 \right] \\
 &\simeq 1 - 0.0119 \\
 &\simeq 0.988
 \end{aligned}$$

∴ La probabilidad de que se produzcan 5 o más artículos antes de detener es de 98.8 %.

Para encontrar el número promedio de artículos que la máquina producirá antes de ser detenida calcularemos $E(X) = r/p$.

$$\begin{aligned}
 E(X) &= r/p = 3/0.15 \\
 &= 20
 \end{aligned}$$

∴ El no. promedio de artículos antes de ser detenida es de 20.

3 Problema 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.

SOLUCIÓN

En este ejercicio estamos suponiendo que la probabilidad de que un trabajador (cualquiera) tenga asbesto en sus pulmones es constante y es del 40 %, por lo que tenemos el parámetro $p = 0.4$. Sea la variable aleatoria $X = \# \text{ de trabajadores hasta encontrar 3 con resultado positivo}$, una v. a. discreta. También sabemos que el asbesto en pulmones no es contagioso, así que supondremos que hay una independencia entre el resultado de cada trabajador. Entonces, buscamos la probabilidad de que se deban analizar a 10 trabajadores “uno por uno” para obtener 3 trabajadores con resultados positivos. Podemos modelar a X mediante una distribución no binomial: $NB(r = 3, p = 0.4)$ y así, obtendremos la probabilidad de que se deban analizar 10 trabajadores para encontrar 3 con resultado positivo evaluando $P(X = 10)$.

$$\begin{aligned} P(X = 10) &= \binom{x-1}{r-1} p^r (1-p)^{x-r} \\ &= \binom{10-1}{3-1} (0.4)^3 (0.6)^{10-3} \\ &\simeq 0.0645 \end{aligned}$$

Por lo tanto

La probabilidad de que se deban analizar 10 trabajadores para poder encontrar 3 con resultado positivo es del 6.45 %.

4 Problema 4

Para el siguiente ejercicio es necesario usar R.

- a) 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 un 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.
- b) Usando la función anterior simule $N = 10^4$ veces una variable aleatoria $\text{Geom}(p)$ para $p = 0.5, 0.1, 0.01$. Grafique las frecuencias normalizadas en color azul. Sobre esta última figura empalme en rojo la gráfica de la función de masa correspondiente. ¿Qué observa?
- c) Repita el inciso anterior para $N = 10^6$. Además, calcule el promedio y la desviación estándar de las simulaciones que realizó. ¿Qué observa?

SOLUCIÓN

Inciso a)

El código se encuentra en el archivo `Problema4.Rmd`, adjunto en la entrega de la tarea.

Inciso b)

Se utilizó la función del inciso anterior para realizar la simulación con $N = 10^4$. Además, ejecutamos la simulación y graficamos los resultados de las proporciones para distintos valores del parámetro p . La gráfica se realizó con la función `plot(...)`, y para empalme de la función de masa de la distribución geométrica correspondiente se utilizó la función `lines(...)` cuyo argumento son los valores de la función de masa (`dgeom(x,prob)`) (cabe destacar que `dgeom()` funciona de forma distinta a la vista en clase; su variable aleatoria representa el número de fracasos hasta obtener un éxito). Por motivos de visualización, modificamos los ticks de cada gráfica con la función `axis()` colocando una secuencia; para el caso de $p = 0.01$ los ticks se acomodan de forma vertical. Decidí no acotar los valores de X para mostrar en su totalidad los resultados de la simulación, por lo que las gráficas pueden verse alargadas.

Cabe destacar que para el caso de $p = 0.01$, tanto en el inciso b) como en el inciso c), decidí empalmar la gráfica de la función de masa en forma de curva, ya que en el modo histograma me generaba una gráfica con errores por los cientos de barras que se acumulaban. Los resultados de este primer inciso se muestran en la figura 4.1.

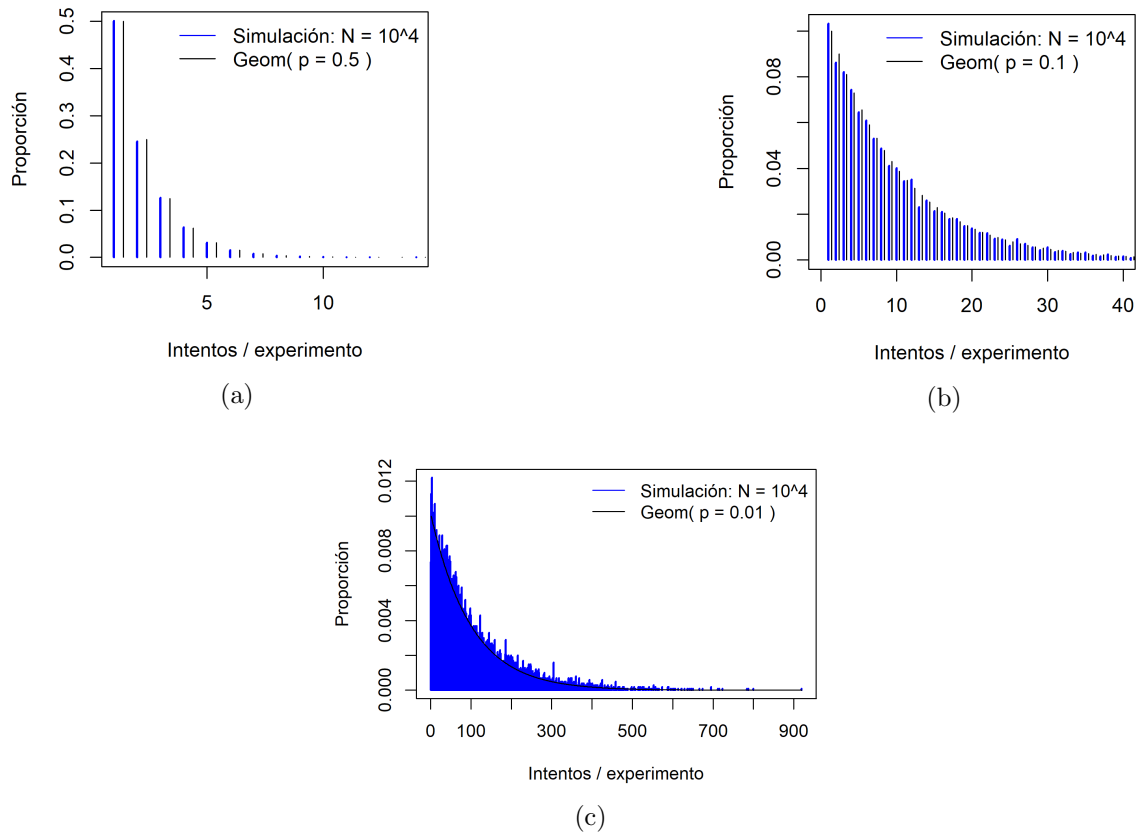


Figura 4.1: Resultados de la simulación con $N = 10^4$ y con distintos valores del parámetro p : (a) 0.5, (b) 0.1 y (c) 0.01.

Las gráficas nos muestran cómo las simulaciones se aproximan a la distribución geométrica correspondiente a los parámetros de cada simulación; en particular y con el valor de $N = 10^4$, vemos que existe menos fluctuaciones entre la simulación y la distribución geométrica cuando p tiene un valor de 0.5, en contraste de un valor pequeño como 0.01 donde muestra muchas más fluctuaciones.

Inciso c)

Se repite el mismo procedimiento que el inciso anterior (Cabe destacar que la realización de las 3 simulaciones me tardó alrededor de 1 hora en total, agradecería retroalimentación sobre el código). Los resultados se muestran en la figura 4.2.

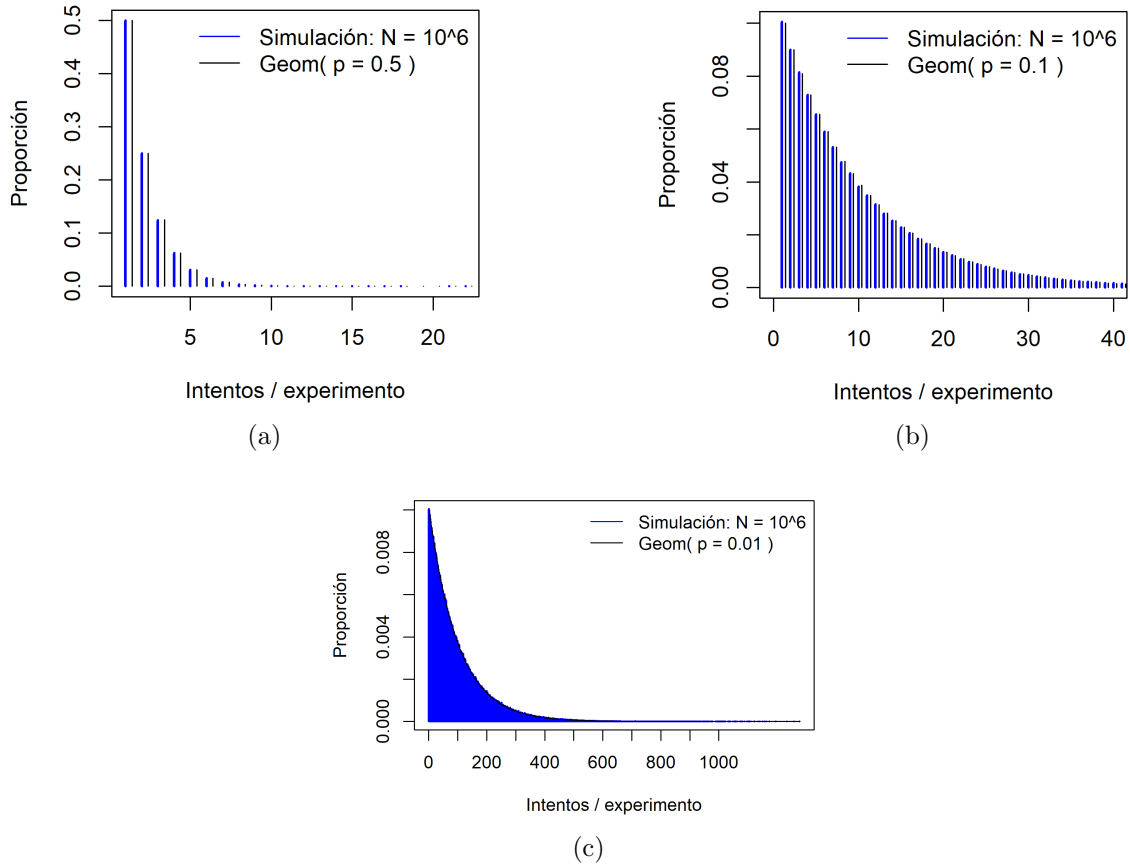


Figura 4.2: Resultados de la simulación con $N = 10^6$ y con distintos valores del parámetro p : (a) 0.5, (b) 0.1 y (c) 0.01.

En estas condiciones, donde aumentamos 100 veces le número de los experimentos, vemos que los resultados de las simulaciones forman representaciones más fieles a cada una de las funciones de masa correspondientes de la distribución geométrica, incluso para valores pequeños de p . A continuación se compararemos ahora los resultados con estadísticos, en particular, la media y la desviación estándar, midiendo el error para cada resultado. Etiquetamos los estadísticos con subíndices que corresponderán a la simulación con $p_1 = 0.5$, $p_2 = 0.1$, $p_3 = 0.01$ y así calculamos para cada simulación con $N = 10^6$:

$$\begin{aligned}
 \mu_1 &\simeq 2.00054 & \delta\mu_1 &= 0.027 \% & ; & \sigma_1 &\simeq 1.4153 & \delta\sigma_1 &= 0.080 \% \\
 \mu_2 &\simeq 9.98745 & \delta\mu_2 &= 0.125 \% & ; & \sigma_2 &\simeq 9.4711 & \delta\sigma_2 &= 0.166 \% \\
 \mu_3 &\simeq 100.075 & \delta\mu_3 &= 0.075 \% & ; & \sigma_3 &\simeq 99.527 & \delta\sigma_3 &= 0.028 \%
 \end{aligned}$$

Por lo que podemos concluir que para un valor **enorme** de N como 10^6 , la simulación dará resultados parecidos a la distribución geométrica con una muy buena precisión.

5 Problema 5

Usando las ideas del inciso anterior escriba una función en R que simule N veces los lanzamientos de una 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.

SOLUCIÓN

Para este problema se realizó la función análoga a la del problema 4 para r éxitos. El cambio que se realizó en la función fue en el ciclo `while` tal que la condición de paro sería hasta que el número de éxitos sea igual o mayor a el parámetro r . Puesto que manejamos el número de éxitos en los experimentos mediante un contador, aumentando de uno en uno por cada éxito, el ciclo pararía hasta que `contador = r`. De esta forma, realizamos las simulaciones para $N = 10^6$ y para los pares $(r, p) = \{(2, 0.2), (7, 0.1)\}$. Además, dado que nuestra simulación puede ser modelada como una suma de la distribución geométrica (comparando con el código del problema 4), la distribución con la que compararemos resultados será la binomial negativa $NB(r, p)$. Los resultados se muestran en la figura 5.1.

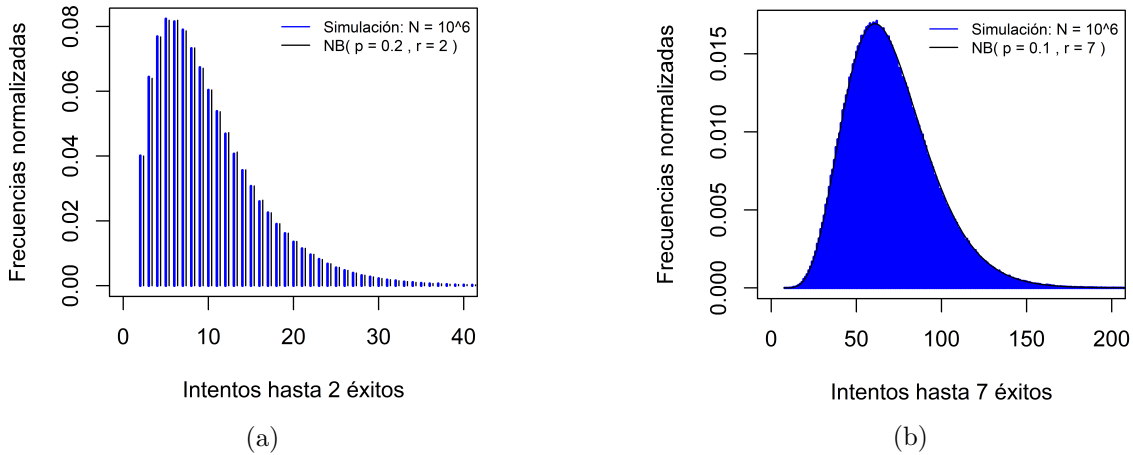


Figura 5.1: Resultados de la simulación con $N = 10^6$ y con distintos valores de los parámetros (p, r) : (a) $(p = 0.2, r = 2)$, (b) $(p = 0.1, r = 7)$.

Se alcanza una alta precisión en los valores de la función de masa de la distribución geométrica mediante la simulación, dados los parámetros correspondientes a la distribución.

6 Problema 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 $1_A(x)$:

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{en otro caso} \end{cases}$$

Sea $Y = 1_A(x)$. Encuentre una expresión para la distribución acumulada de Y .

SOLUCIÓN

Desde la definición de Y , podemos ver que posee únicamente 2 valores posibles: 0 y 1. Y para evaluar cada probabilidad, debemos recurrir a la definición de la función $1_A(x)$; Si $Y = 0$ quiere decir que $x \notin A$, y si $Y = 1$ quiere decir que $x \in A$. Entonces podemos expresar las probabilidades de Y con:

$$\mathbb{P}(Y = y) = \begin{cases} 1 - \mathbb{P}(X = x \in A) & \text{para } y = 0 \\ \mathbb{P}(X = x \in A) & \text{para } y = 1 \\ 0 & \text{en otro caso} \end{cases} \quad (6.1)$$

Pero, ¿cuál es la probabilidad de que $x \in A$? Dada la función de densidad f de X , la función de distribución acumulada F de X y sean $a, b \in \mathbb{R}$ las fronteras de A con $a < b$, la probabilidad $\mathbb{P}(X = x \in A)$:

$$\mathbb{P}(X = x \in A) = \int_a^b f(x)dx = F(b) - F(a) \quad (6.2)$$

Por lo tanto, con ayuda de las ecuaciones anteriores (eq. 6.1, 6.2), y viendo que $\mathbb{P}(Y = 0) + \mathbb{P}(Y = 1) = 1$, podemos definir la función de distribución acumulada G de Y como sigue:

$$G(y) = \begin{cases} 0 & \text{para } y < 0 \\ 1 - (F(b) - F(a)) & \text{para } 0 \leq y < 1 \\ 1 & \text{para } y \geq 1 \end{cases}$$

7 Problema 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

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

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

SOLUCIÓN

La forma de la función de probabilidad es exactamente a una función de masa de una distribución Poisson(λ), donde para nuestro caso $\lambda = 6 \frac{\# \text{ meteoros}}{1/4 \text{ hora}}$. Para saber el número de meteoros que esperaría ver en un cuarto de hora se calcula $E(Y)$.

$$E(Y) = \mu = \lambda = 6$$

Entonces, la probabilidad de que una persona vea en un cuarto de hora al menos la mitad de los meteoros que esperaría ver es $P(Y \geq \mu/2)$:

$$\begin{aligned} P(Y \geq \mu/2) &= P(Y \geq 3) = 1 - P(Y \leq 2) \\ &= 1 - (f_Y(0) + f_Y(1) + f_Y(2)) \\ &= 1 - \left(\frac{e^{-6}6^0}{0!} + \frac{e^{-6}6^1}{1!} + \frac{e^{-6}6^2}{2!} \right) \\ &\simeq 0.938 \end{aligned}$$

La probabilidad de que una persona vea en un cuarto de hora al menos la mitad de los meteoros que esperaría ver es del 93.8 %.

8 Problema 8

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

$$f_y(y) = 6y(1 - y) \quad 0 \leq y \leq 1$$

donde y representa la proporción de preguntas que el estudiante contesta correctamente. Cualquier calificación menor a 0.4 es reprobatoria. Responda lo siguiente:

- ¿Cuál es la probabilidad de que un estudiante repruebe?
- Si 6 estudiantes toman el examen, ¿cuál es la probabilidad de que exactamente 2 reprueben?

SOLUCIÓN

Inciso a)

Tomemos a la probabilidad de que un estudiante repruebe como la probabilidad de que éste obtenga una calificación menor a 0.4. Entonces la probabilidad de que el estudiante repruebe se representa por $P(Y < 0.4)$. Computando la probabilidad

$$\begin{aligned} P(Y < 0.4) &= \int_{-\infty}^{0.4} 6y(1 - y)dy \\ &= \int_{-\infty}^0 6y(1 - y)dy + \int_0^{0.4} 6y(1 - y)dy = \int_0^{0.4} 6y(1 - y)dy \\ &= \int_0^{0.4} (6y - 6y^2)dy = \left(3y^2 - 2y^3\right)\Big|_{y=0}^{y=0.4} \\ &= (0.352) - (0) = 0.352 \end{aligned}$$

La probabilidad de que un estudiante repruebe es del 35.2%.

Inciso b)

Para realizar esto tomaremos ciertos supuestos; la probabilidad de que un alumno repruebe no influye de ninguna manera a la probabilidad de que los demás lo hagan (probabilidades independientes), y también supondremos que la probabilidad de que cada alumno repruebe es constante, es decir, $p = 0.352$ para todos los alumnos. Sea $X = \# \text{ de alumnos reprobados de } 6$, mediante los supuestos que hemos dado podemos establecer que $X \sim \text{Binomial}(n = 6, p = 0.352)$. Entonces, representamos la probabilidad de que exactamente 2 alumnos reprueben por $P(X = 2)$ y

$$\begin{aligned}P(X = 2) &= \binom{6}{2} (0.352)^2 (1 - 0.352)^{6-2} \\&= 15(0.352)^2 (0.648)^4 \\&\simeq 0.328\end{aligned}$$

La probabilidad de que exactamente 2 de 6 alumnos reprueben el examen es del 32.8 %.
--

9 Problema 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 $\lambda = 1/2$ 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 $T/dt = 1000$ veces. Divida el intervalo $[0, T]$ en intervalitos de longitud dt que tengan la forma $(k \cdot dt, (k + 1) \cdot dt]$, con $k = 0, 1, 2, \dots, (T/dt - 1)$. Para cada uno de estos intervalitos simule una v.a. Bernoulli($\lambda \cdot dt + 10^{-6}$) y guarde su resultado en un vector del tamaño adecuado.*

SOLUCIÓN

Los códigos se encuentran en el archivo `Problema9.Rmd`. La figura 9.1 nos muestra tres trayectorias de un proceso de Poisson con los mismos parámetros. Vemos que todas las simulaciones tienden a tener una trayectoria similar a la de una recta nacida en el origen con pendiente $m = 2$, la cual *creo* que se relaciona con la intensidad $\lambda = 2$, ya que representa la media de eventos ocurridos en una unidad de tiempo.

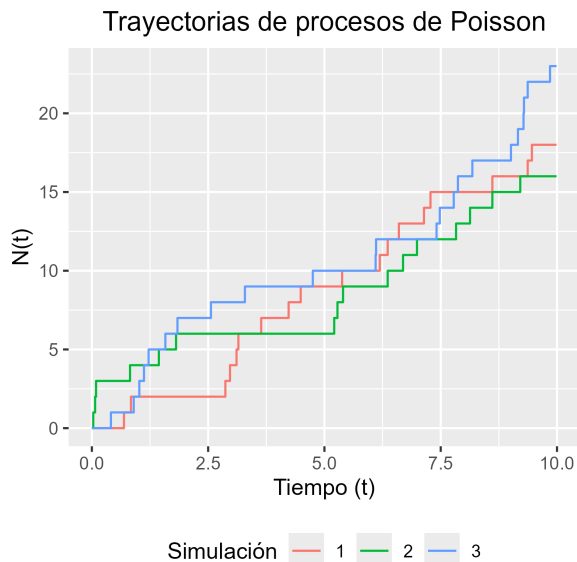


Figura 9.1: Gráfica de 3 trayectorias de un proceso de Poisson con parámetros $\lambda = 2$ hasta un tiempo $t = 10$.

Por otra parte, se realizaron 10^4 simulaciones de un proceso de Poisson con $\lambda = 1/2$ hasta el tiempo $t = 1$. Sacando el dato $N(1)$ de cada simulación como el número total de éxitos en la misma, pudimos realizar una gráfica de las frecuencias normalizadas para su comparación con la función de masa de la distribución Poisson correspondiente. El resultado se muestra en la figura 9.2.

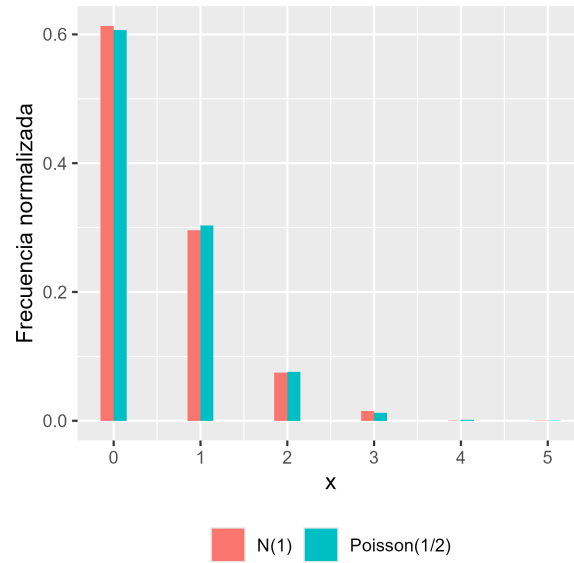


Figura 9.2: Gráfica de las frecuencias normalizadas de $N(1)$ y la función de masa de la distribución de $\text{Poisson}(1/2)$.

Podemos observar la simulación propuesta mediante las 5 hipótesis establecidas en clase sobre $N(t)$ (o X_t) y $p_n(t)$ nos devuelve resultados que coinciden de forma muy aproximada a la distribución de Poisson, es decir, se ve de forma computacional que se cumple el resultado obtenido en clase.

$$p_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad ; \quad n = 0, 1, 2, \dots$$

10 Problema 10

En una oficina de correo los paquetes llegan según un proceso de Poisson de intensidad λ . Hay un costo de almacenamiento de c pesos por paquete y por unidad de tiempo. Los paquetes se acumulan en el local, y se despachan en grupos cada T unidades de tiempo (es decir, se despachan en $T, 2T, 3T, \dots$). Hay un costo por despacho fijo de K pesos (es decir, el costo es independiente del número de paquetes que se despachen).

- a) ¿Cuál es el costo promedio por paquete por almacenamiento en el primer ciclo $[0, T]$?
- b) ¿Cuál es el costo promedio por paquete por almacenamiento y despacho en el primer ciclo?
- c) ¿Cuál es el valor de T que minimiza el costo promedio?

SOLUCIÓN
