UNR Universidad Nacional de Rosario

Universidad Nacional de Rosario

Facultad de Ciencias Exactas, Ingeniería y Agrimensura

Trabajo práctico Final

Probabilidad y Estadística

Autores:
Arroyo, Joaquín | A-4294/3
Montoro, Emiliano | M-6653/2

March 1, 2023

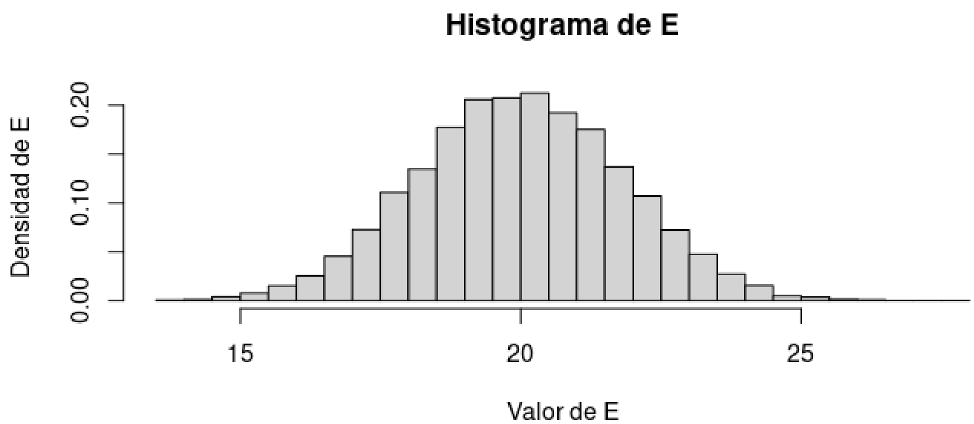
Contents

1 Problema 1

1.1 a

La variable aleatoria E es la suma de 40 variables aleatorias uniformemente distribuidas en el intervalo (0,1), lo que implica que E sigue una distribución uniforme continua en el intervalo (0, 40).

Para calcular la función de densidad de probabilidad de E de forma empírica, podemos generar una gran cantidad de muestras aleatorias y sumarlas para obtener la variable aleatoria E. Luego, podemos crear un histograma de las muestras de E, y dicho histograma aproxima la función de densidad.
A partir de esto, y con el código R, se obtuvo el siguiente gráfico:

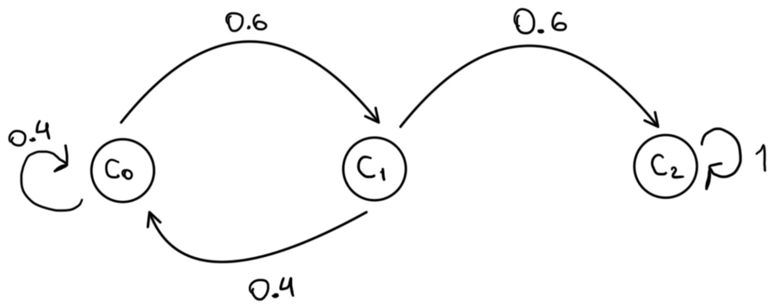


El anterior histograma, puede ser deducido sin realizar ninguna simulación, ya que el teorema central del límite indica que, en condiciones muy generales, si S_n es la suma de n variables aleatorias independientes, entonces la función de distribución de S_n se aproxima bien a una distribución normal, es decir, obtenemos una aproximación de la campana de Gauss. A medida que el n vaya creciendo, el gráfico converge a la campana de Gauss. En nuestro caso S_n es E y $n = 40$.
Vamos a calcular la media y la desviación estandar de E:
$M(E) = M(X_1) + ... + M(X_{40}) = 1/2 + ... + 1/2 = 40/2 = 20$
$V(E) = V(X_1) + ... + V(X_{40}) = 1/12 + ... + 1/12 = 40/12 = 10/3$
$SD(E) = \sqrt{V(E)} = \sqrt{10/3} = 1.82547$
Fue posible calcular la media y la varianza de la anterior manera debido a la independencia de las variables X_i con $i \in [1 : 40]$.

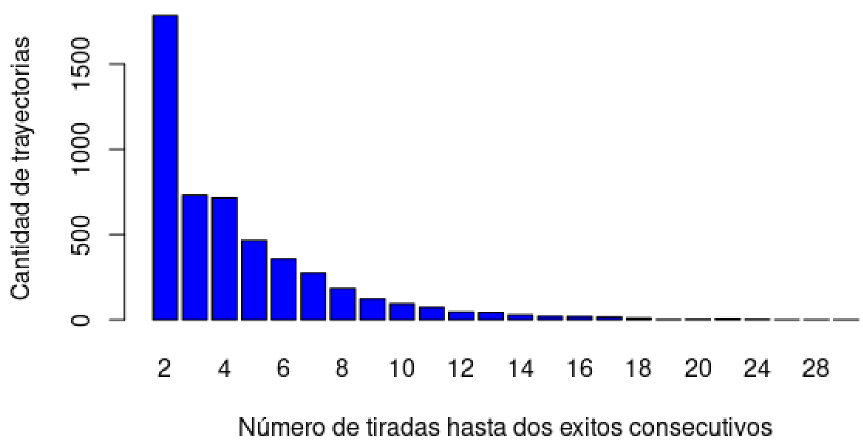
2 Problema 2

2.1 a

Tomando p = 0.6 y k = 2 damos el siguiente grado de una cadena de Markov:



En la cadena de Markov donde cada estado (de $E = C_0, C_1, C_2$) representa la cantidad de veces consecutivas que tuvo éxito la experiencia. Para esta cadena la probabilidad de transición a C_0 es $1 - p$ y la de transición a C_{i+1} es p, excepto para C_2 ya que este es un estado absorbente, entonces, podemos decir que tener 2 éxitos consecutivos es equivalente a llegar a el estado C_2 por primera vez. Con la cantidad de trayectorias en 5000, pudimos ver el siguiente patrón de cantidad de realizaciones de la experiencia antes de obtener k éxitos:



2.2 b

Como usamos un modelo de cadena de Markov para la simulación y el hecho de que C_2 es absorbente, podemos decir que $E(N_2)$ es equivalente a pedir el tiempo esperado de absorción partiendo de C_0. Para calcular esto, tomamos la matriz de transicion de un paso:

$$\begin{pmatrix} 1-p & p & 0 \\ 1-p & 0 & p \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

La cuál, al ser absorbente podemos la podemos descomponer como la siguiente matriz:

$$\begin{pmatrix} Q & R \\ O & I \end{pmatrix} \tag{2}$$

Donde

$$Q = \begin{pmatrix} 1-p & p \\ 1-p & 0 \end{pmatrix}, R = \begin{pmatrix} 0 \\ p \end{pmatrix}, 0 = \begin{pmatrix} 0 & 0 \end{pmatrix}, I = \begin{pmatrix} 1 \end{pmatrix} \tag{3}$$

Con Q podemos calcular

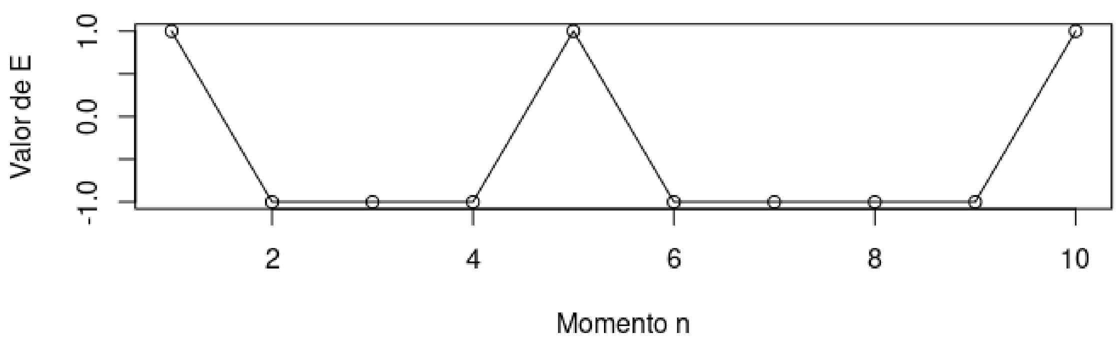
$$F = (I - Q)^{-1} = \begin{pmatrix} 2.78 & 1.67 \\ 1.11 & 1.67 \end{pmatrix} \tag{4}$$

Luego, el tiempo esperado de absoción partiendo de C_0 es $\alpha_0 = F_{00} + F_{01} = 2.78 + 1.67 = 4.45$.
Esto nos dice que partiendo de C_0 esperamos llegar a C_2 en aproximadamente 4.45 pasos.

3 Problema 3

3.1 a

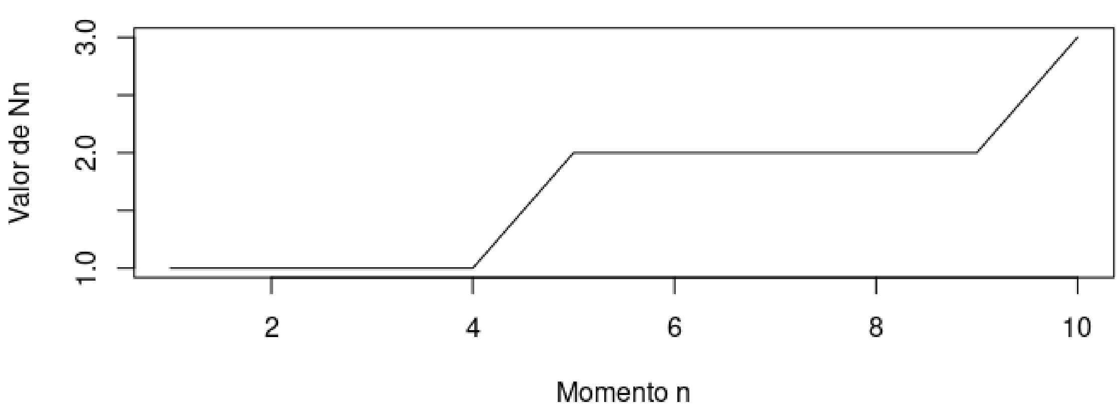
Sea E_n : la señal emitida en el momento n es incorrecta.
La realización del proceso fue realizado con $n \in [1 : 10]$ y $p = 0.3$. Esto arrojó los siguientes resultados:



Esto se asemeja a un proceso de Bernoulli.
Podemos ver que en los momentos 1, 5 y 10 la señal emitida fue incorrecta.

3.2 b

Sea N_n : cantidad de señales incorectas emitidas hasta momento n.
Notar que $N_n = \sum_{i=1}^{n} E_n$
El proceso N_n fue realizado con $n = 10$ y $p = 0.3$, y arrojó los siguientes resultados:



Este proceso es una caminata aleatoria.
Al haber tomado $p = 0.3 < 0.5$ vemos como la caminata no asciende de una forma rápida.
Se puede ver la correspondencia entre el gráfico 1 y el gráfico 2.

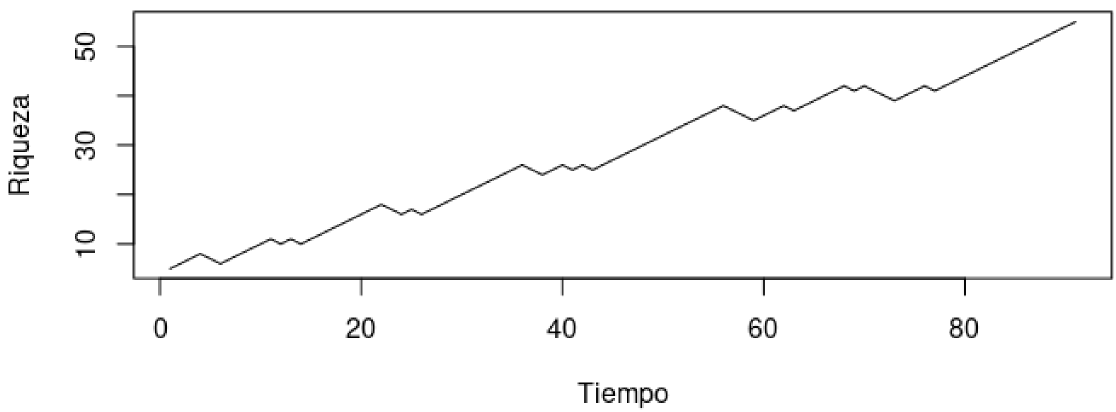
3.3 c

Sabemos que para E_n tenemos $E(E_n) = p$ y $V(E_n) = p(1-p)$, reemplazando, $E(E_n) = 0.3$ y $V(E_n) = 0.21$
Y para N_n tenemos $E(N_n) = \sum_{i=1}^{n} E(E_n) = 10p = 3$ y $V(N_n) = \sum_{i=1}^{n} V(E_n) = 10p(1-p) = 2.1$
Lo anterior es posible debido a que los procesos E_n con $n \in [1:10]$ son independientes.

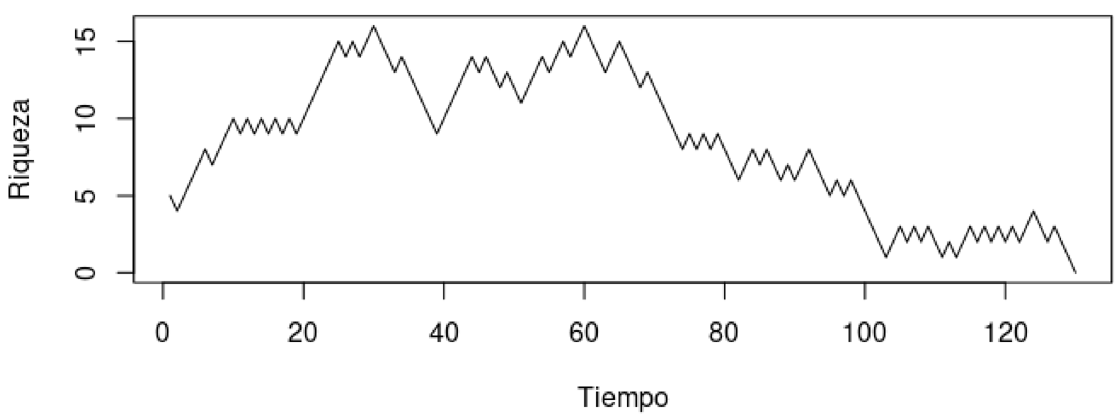
4 Problema 4

4.1 a

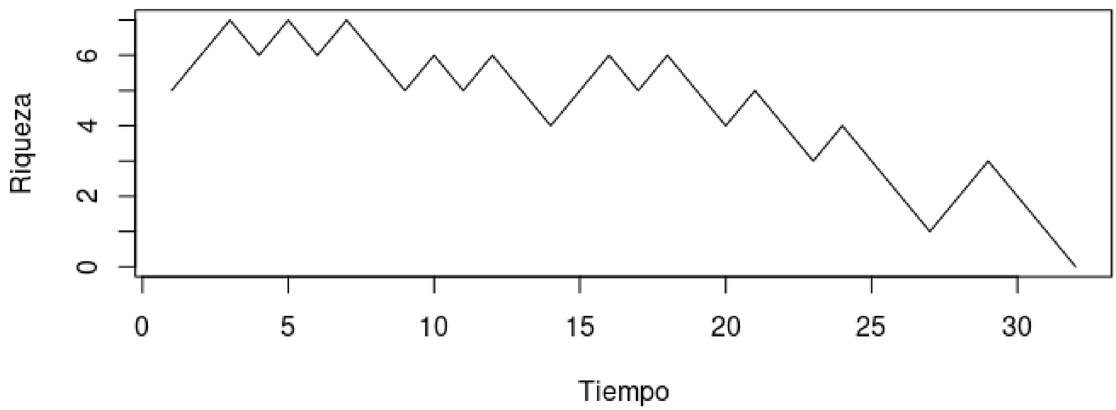
Se adjunta simulación con $k = 5$, $S = 50$, $p = 0.8$



Se adjunta simulación con $k = 5$, $S = 50$, $p = 0.5$



Se adjunta simulación con $k = 5$, $S = 50$, $p = 0.45$



4.2 b

Se puede aproximar la probabilidad de que el jugador llegará a la ruina antes de ganar S dólares a partir de una distribución binomial, calculando la media de una lista que contiene ceros y unos, un cero por cada vez que el experimento falló, es decir, el jugador llegó a la ruina, y uno por cada vez que alcanzo o superó S, luego le restamos a 1 este valor, y obtenemos la probabilidad buscada. Las siguientes probabilidades fueron obtenidas con 100000 simulaciones cada una.

Caso $p = 0.8 \rightarrow probabilidad\ de\ ruina = 0.0009$

Caso $p = 0.5 \rightarrow probabilidad\ de\ ruina = 0.89$

Caso $p = 0.45 \rightarrow probabilidad\ de\ ruina = 0.98$

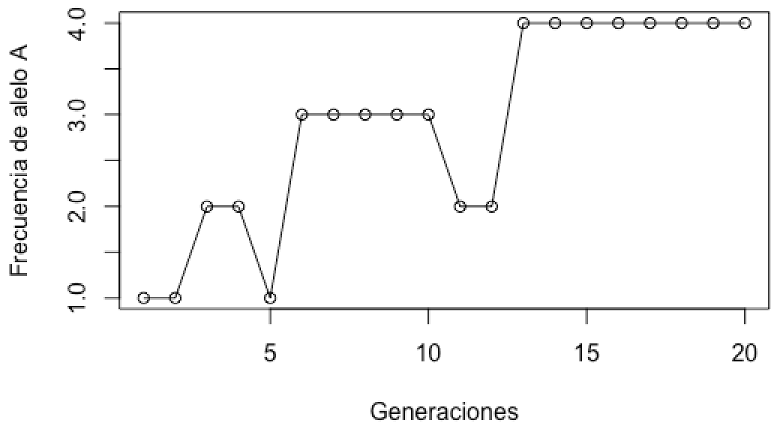
5 Problema 5

Para este problema, tomaremos $k = 4$. A partir de esto, y con la fórmula dada, generamos la siguiente matriz de transición:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 81/256 & 27/64 & 27/128 & 3/64 & 1/256 \\ 1/16 & 1/4 & 3/8 & 1/4 & 1/16 \\ 1/256 & 3/64 & 27/128 & 27/64 & 81/256 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{5}$$

5.1 a

La simulación fue realizada con $k = 4$ y por 20 generaciones, esta misma generó los siguientes resultados:



5.2 b

El valor de $P_{00} = P_{kk} = 1$.

5.3 c

Podemos ver que la cadena no es irreducible, ya que tiene dos estados absorbentes, 0 y k (4 en este caso), por lo tanto, no existe distribución limite.
Entonces aproximamos la probabilidad para cada estado, a partir de potencias altas de la matriz de transición:

$$P^{100} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.25 \\ 0.5 \\ 0.75 \\ 1 \end{pmatrix} \tag{6}$$

Y obtenemos la probabilidad de que la población evolucione al estado k, según el estado en el que se comience.
Claramente, si se comienza desde el estado k, la probabilidad es 1.

6 Problema 6

6.1 a

Para este ejercicio tomamos $E = \{1, 2, 3, 4, 5\}$, donde cada estado representa a una página web del problema.
Para modelar la probabilidad de que el usuario visite una página j desde una página i, podemos utilizar una función:

$$f(i,j) = \begin{cases} 1/n & \text{si i tiene una arista a j} \\ 0 & \text{caso contrario} \end{cases} \tag{7}$$

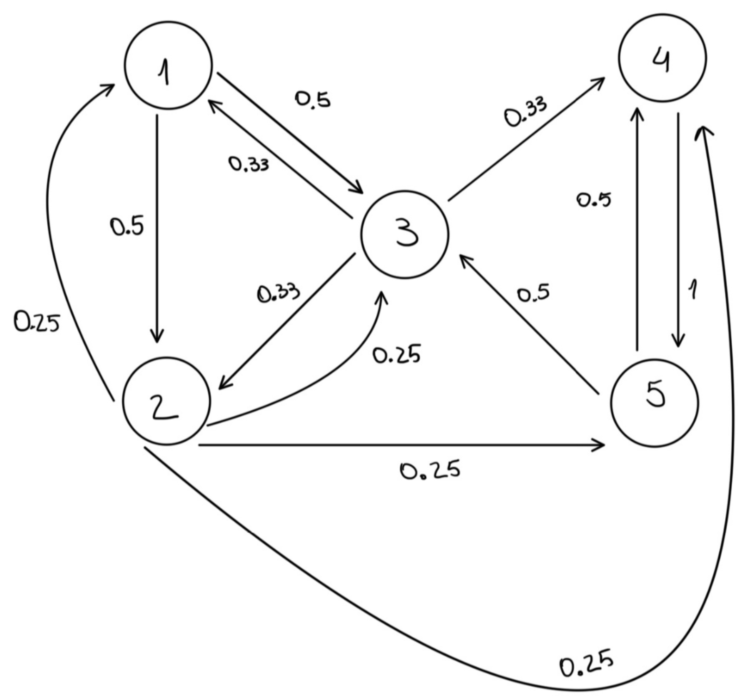
donde n es la cantidad de aristas del nodo i.
Suponemos que el usuario siempre se va a mover a una página diferente, esto para que el al momento de calcular el ranking de páginas, las páginas no dependan de si mismas.
Que en nuestro modelo, el nodo 2 no tenga links hacia ninguna página nos representa un problema, ya que Markov no permite una fila llena de ceros, por lo que la fila que representa a dicho nodo es la siguiente:

$$\begin{pmatrix} 1/4 & 0 & 1/4 & 1/4 & 1/4 \end{pmatrix} \tag{8}$$

ya que de esta forma repartimos la probabilidad equitativamente y para que el estado 2 no sea absorbente.
Para representar la cadena de Markov utilizamos la matriz de transicion de un paso, la cual, en este caso, representa a la probabilidad de que, dada una página, el usuario acceda a otra. Dicha matriz, en este problema particular, es la siguiente:

$$\begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 1/4 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix} \tag{9}$$

y su grafo asociado es el siguiente:



6.2 b

Para calcular la probabilidad de visitar las página i con $i = 1, ..., 5$, lo hacemos usando el teorema fundamental del limite para cadenas de Markov ergódicas, ya que esta cadena (finita) es aperiódica (periodo de todos sus estados igual a 1) e irreducible (es posible ir de cualquier estado a cualquier otro en un número finito de pasos), a partir de esto sabemos que existe una única distribución limite de la cadena, la cual es la siguiente:

$$\begin{pmatrix} 0.1086957 & 0.1304348 & 0.2282609 & 0.25 & 0.2826087 \end{pmatrix} \tag{10}$$

Por lo que el ranking queda de la siguiente manera (mayor a menor probabilidad): $5, 4, 3, 2, 1$.

7 Problema 7

7.1 a

Sea

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 8/3 & 3/13 & 1/13 & 1/13 \\ 1/16 & 3/16 & 3/8 & 1/4 & 1/8 \\ 0 & 1/11 & 4/11 & 5/11 & 1/11 \\ 0 & 1/8 & 1/2 & 1/8 & 1/4 \end{pmatrix} \tag{11}$$

La distribución inicial es

$$\pi_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{12}$$

Luego, calculamos

$$P^7 \approxeq \begin{pmatrix} 0.021 & 0.264 & 0.344 & 0.228 & 0.141 \\ 0.021 & 0.272 & 0.341 & 0.225 & 0.139 \\ 0.021 & 0.265 & 0.344 & 0.228 & 0.141 \\ 0.021 & 0.264 & 0.344 & 0.229 & 0.141 \\ 0.021 & 0.265 & 0.344 & 0.228 & 0.141 \end{pmatrix} \tag{13}$$

A partir de esta matriz, obtenemos la solución, multiplicando la distribución inicial por P^7:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix} P^7 \approxeq \begin{pmatrix} 0.021 & 0.265 & 0.344 & 0.228 & 0.141 \end{pmatrix} \tag{14}$$

Esto es calculado a partir de la propiedad $P(X_n = j) = \pi_0 P^n(j)$, en este caso $j = 4$ y $n = 7$.
Con estos resultados, podemos afirmar que el puerto con menos probabilidades de ser atacado es el puerto 80, con aproximadamente 0.021 de probabilidad; y el que tiene más probabilidades es el puerto 139, con aproximadamente 0.344 de probabilidad.

7.2 b

Sabemos que una cadena de Markov (finita) aperiódica (periodo de todos sus estados igual a 1) y a su vez irreducible (es posible ir de cualquier estado a cualquier otro en un número finito de pasos) es ergódica, con esto y el teorema del limite para cadenas de Markov ergódicas, podemos calcular su distribución límite, la cual es la siguiente:

$$\begin{pmatrix} 0.02146667 & 0.2669333 & 0.3434667 & 0.2273333 & 0.1408 \end{pmatrix} \tag{15}$$

8 Problema 8

Los defectos son un proceso de Poisson que involucra distancias, por lo que:

1 Al inicio tenemos 0 defectos

2 El número de eventos en un intervalo de longitud fija sigue una distribución de Poisson.

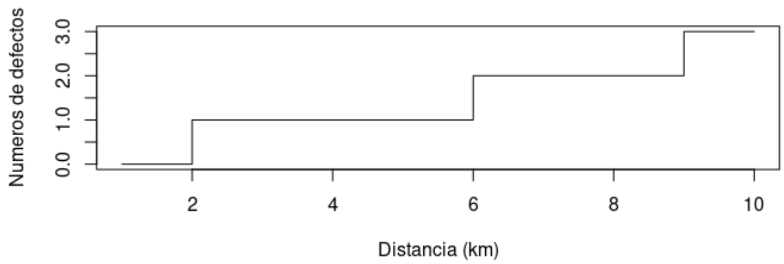
3 La distancia entre dos eventos sucesivos sigue una distribución exponencial con parámetro lambda.

4 No podemos tener dos defectos en un mismo instante.

5 El número de defectos de un intervalo depende de los kilometros recorridos después del último defecto.
A partir de esto, podemos decir que el número de defectos es discreto con aumento 1.

8.1 a y b

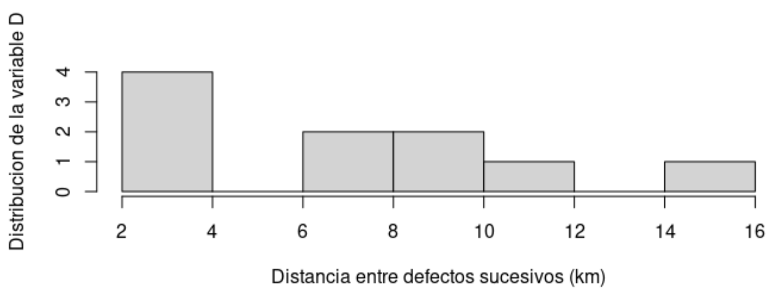
La primer simulación fue realizada con una distribución de Poisson con $\lambda = 0.1$
A partir de dicha simulación, obtuvimos la siguiente gráfica:



Esto nos indica la cantidad de defectos acumulados en cada km.

8.2 c y d

Para esta segunda simulación, se utilizó una distribución exponencial con parámetro igual a 0.1, a partir de dicha simulación, se obtuvieron los siguientes resultados:



Se calculó datos relevantes, como lo son la media y la desviación estandar, las cuales fueron:
$M(D) = 7.5$ y $SD(D) = 4.33$.
Analizando, la media nos indica que en promedio, se espera que haya un defecto cada 7.5km de trayectoria; y la desviación estándar de 4.33km, indica la variabilidad de la distribución de D; mientras más pequeña sea esta desviación, más cerca estaran los valores individuales de D de la media.

9 Anexos

9.1 Problema 1

```
set.seed(123)
n = 40; m = 10000;
X = matrix(runif(n*m, 0, 1), nrow = m)
E = rowSums(X)

hist(E, breaks = 40, main = "Histograma de E",
     xlab = "Valor de E", ylab = "Densidad de E")
```

9.2 Problema 2

```
# Parte A
simulate_Nk <- function(p, k){
  n_successes = 0
  n_trials = 0

  while (n_successes < k) {
    n_trials = n_trials + 1
    if (runif(1) < p) {
      n_successes = n_successes + 1
    } else {
      n_successes = 0
    }
  }
  return(n_trials)
}

n_simulations = 5000
p = 0.6
k = 2
results = replicate(n_simulations, simulate_Nk(p, k))

simulation_results = list(results)

table_results = table(results)
barplot(table_results,
        main = paste(""),
        xlab = "N mero de tiradas hasta dos exitos consecutivos",
        ylab = "Cantidad de trayectorias",
        col = "blue")
```

9.3 Problema 3

```
# Parte A
n = 10; p = 0.3
set.seed(123)

E = ifelse(runif(n) <= p, 1, -1)
replace(E, -1, 0)
Nn = cumsum(E)

plot(E, xlab="Momento n", xlab="Valor de E", ylab="Valor de E")
plot(1:n, Nn, type="l", xlab="Momento n", ylab="Valor de Nn")
```

9.4 Problema 4

```
# Parte A
k = 5; n = 50; p = 0.9; q = 1-p
simulate <- function(){
  sim = k
  while (sum(sim) > 0 && sum(sim) < n+k) {
    if (runif(1) < p) {
      sim = append(sim, 1)
    } else {
      sim = append(sim, -1)
    }
  }
  return(sim)
}
results = simulate()
results = cumsum(results)
plot(results, xlab="Tiempo", ylab="Riqueza", type="l")

# Parte B
1 - mean(replicate(100000, sum(simulate())/(n+k)))
```

9.5 Problema 5

```
# Parte A
k = 4; g = 20;
states = c("0", "1", "2", "3", "4")
P = matrix(ncol = k+1, nrow = k+1)

for (i in 0:k) {
  for (j in 0:k) {
    P[i+1, j+1] = dbinom(j, k, i/k)
  }
}

library(markovchain)
P = new("markovchain", states, TRUE, P)
MM=rmarkovchain(g, P, "0")
plot(MM, type = "o", xlab = "Generaciones", ylab = "Frecuencia de
   alelo A")
# Parte C
P = (P^100) [, k+1]
```

9.6 Problema 6

```
    # Parte
    states = c("1", "2", "3", "4", "5")
    P = matrix(c(0, 1/2, 1/2, 0, 0,
              1/4, 0, 1/4, 1/4, 1/4,
              1/3, 1/3, 0, 1/3, 0,
              0, 0, 0, 0, 1,
              0, 0, 1/2, 1/2, 0),
            nrow = 5, byrow = TRUE)
    library(markovchain)
    P = new("markovchain", states, TRUE, P)
    if (is.irreducible(P) && period(P) == 1) print(steadyStates(P))
```

9.7 Problema 7

```
    states =  c("80", "135", "139", "445", "NA")
    P = matrix(c(0, 0, 0, 0, 1,
              0, 8/13, 3/13, 1/13, 1/13,
              1/16, 3/16, 3/8, 1/4, 1/8,
              0, 1/11, 4/11, 5/11, 1/11,
              0, 1/8, 1/2, 1/8, 1/4),
            nrow = 5, byrow = TRUE)
    library(markovchain)
    P = new("markovchain", states, TRUE, P)
    if (is.irreducible(P) && period(P) == 1) print(steadyStates(P))
```

9.8 Problema 8

```
    lambda = 0.1
    km = 10

    defectos = rpois(km, lambda)
    defectos_km = cumsum(defectos)

    plot(defectos_km, type="s", xlab="Distancia (km)",
         ylab ="Numeros de defectos")

    km = 100
    num_events = rpois(1, km*lambda)
    differences = rexp(num_events, lambda)
    mean(differences)
    sd(differences)

    hist(differences, xlab="Distancia entre defectos sucesivos (km)",
         ylab ="Distribucion de la variable D", main= "")
```