Universidad Nacional de Rosario
Facultad de Ciencias Exactas,
Ingeniería y Agrimensura
Departamento de Ciencias de la
Computación

PROBABILIDAD Y ESTADISTICA

Trabajo Práctico Final

Cassinerio Marcos, Garavano Lautaro

Febrero 2022

Ejercicio 1

Para nuestras simulaciones, elegimos un valor de $p = 0{,}3$, y $k = 3$. Describimos el proceso mediante una cadena de Markov, basándonos en la variable aleatoria C_i que denota la "Cantidad de éxitos consecutivos que se tienen después del i-ésimo experimento". Después de cada experimento, podemos aumentar en 1 nuestra cantidad de éxitos consecutivos hasta ahora, o volver a tener 0 éxitos consecutivos. Este proceso se termina cuando se llega al estado k. El grafo que describe esta cadena se ve de la siguiente manera:

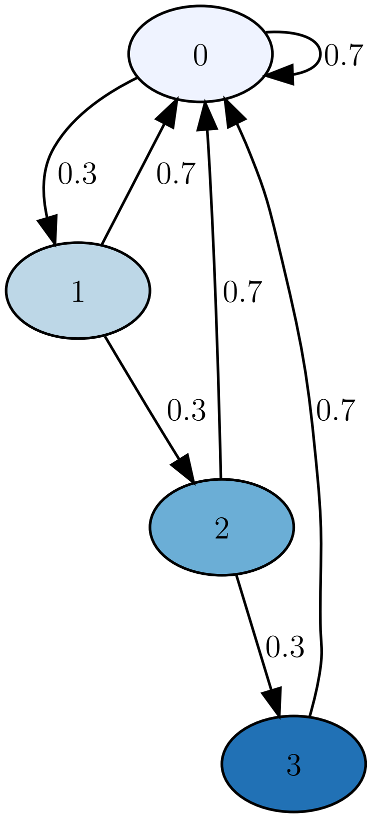

Figura 1: Grafo del Ejercicio 1

Apartado a

Con esta cadena de Markov, simulamos 1000 caminatas aleatorias, empezando desde el estado "0" y terminando en el estado "3". A continuación visualizamos la longitud de las caminatas en un histograma. Incluimos además una tabla mostrando los resultados más precisamente.

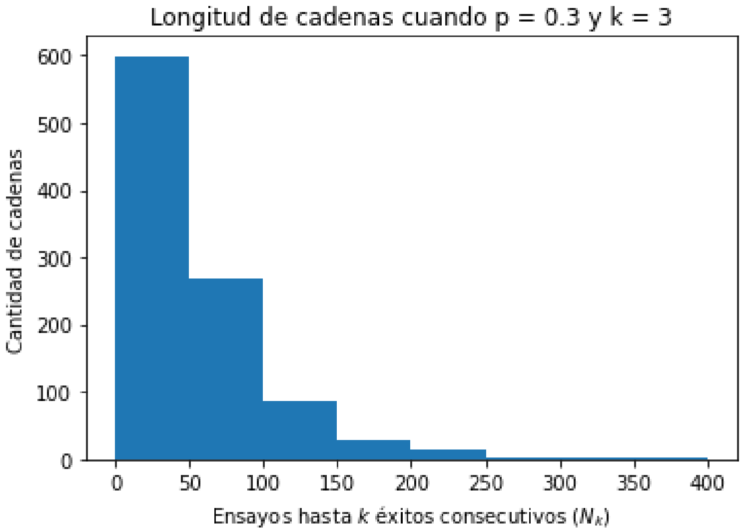

Figura 2: Histograma de las caminatas aleatorias

Ensayos Necesarios	Frecuencia
[0, 50)	599
[50, 100)	268
[100, 150)	87
[150, 200)	28
[200, 250)	14
[250, 300)	2
[300, 350)	1
[350, 400)	1

Tabla 1: Ensayos hasta k éxitos consecutivos

Apartado b

Podemos aproximar la esperanza de N_k calculando la media aritmética de los valores obtenidos para un gran número de cadenas. Con las 1000 cadenas generadas anteriormente, obtuvimos que

$$E(N_k) \simeq 51{,}862$$

Otra manera de aproximarlo es planteando la ecuación de la esperanza de N_k

$$E(N_k) = \sum_{i=1}^{\infty} i P_0(N_k = i)$$

La probabilidad de que N_k sea igual a i no es más que la probabilidad de que, empezando desde el estado 0, lleguemos al estado k por primera vez en i pasos ($F_i(0,k)$).

Reemplazando en la ecuación obtenemos

$$E(N_k) = \sum_{i=1}^{\infty} iF_i(0,k)$$

Sabemos que

$$F_1(i,j) = P(i,j)$$
$$F_n(i,j) = \sum_{b \in E-\{j\}} P(i,b)F_{n-1}(b,j)$$

Más aún, en nuestro caso cada estado i sólo está conectado con $i+1$, con $P(i,i+1) = p$, y con 0, donde $P(i,0) = 1-p$ (Exceptuando el estado k). Entonces

$$F_n(i,k) = pF_{n-1}(i+1,k) + (1-p)F_{n-1}(0,k)$$

Con todo esto podemos calcular recursivamente los términos de esta sumatoria infinita. Con suficientes términos, podemos obtener una buena aproximación de la esperanza de N_k (Gracias a que $\lim_{i\to\infty} P(N_k = i) = 0$)

Con un millón de términos obtenemos que

$$E(N_k) \simeq 51{,}48148148148103$$

Ejercicio 2

Simulamos el proceso D_n con $p = 0{,}75$ como una caminata aleatoria por la cadena dada por el grafo:

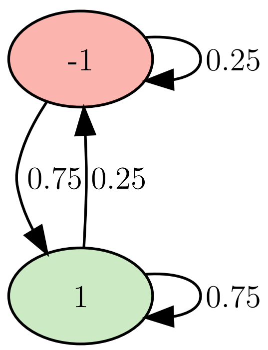

Figura 3: Grafo que representa la cadena de Markov de D_n

Apartado a

Una caminata aleatoria de 50 pasos produjo el siguiente resultado

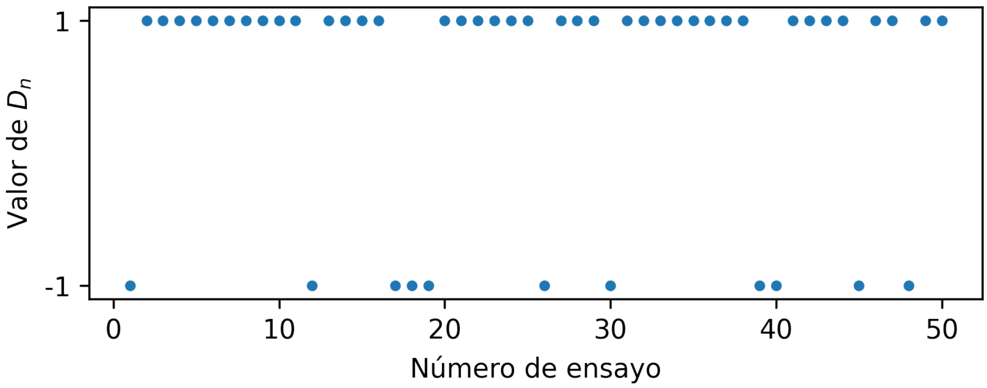

Figura 4: Caminata de 50 pasos

Apartado b

A partir de la simulación creada en el apartado anterior podemos ver la evolución de S_n. Basta con sumar todos los valores de D_n obtenidos hasta el ensayo actual. Formalmente:

$$S_n = \sum_{i=1}^{n} D_i$$

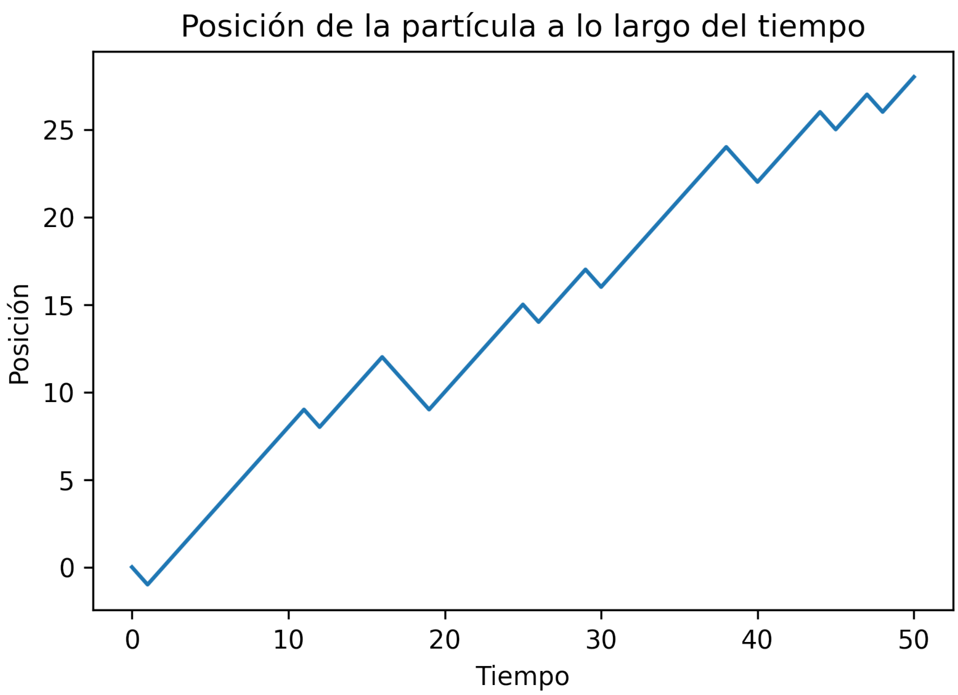

Figura 5: Posición de la partícula

Ejercicio 3

Representaremos el juego con una cadena de Markov con $S + 1$ estados (de 0 a S inclusive). Cada estado i está conectado a $i + 1$ con probabilidad p y a $i - 1$ con probabilidad $1 - p$, salvo por el primer y el último estado, que son absorbentes. Por ejemplo, para $S = 3$ nuestra matriz de transición sería

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

Apartado a

Para nuestras simulaciones elegimos $S = 100$ y $k = 50$. Probamos con 3 valores de p: 0,3, 0,5 y 0,7. A continuación se grafica la evolución del capital del jugador a lo largo del tiempo en cada caso.

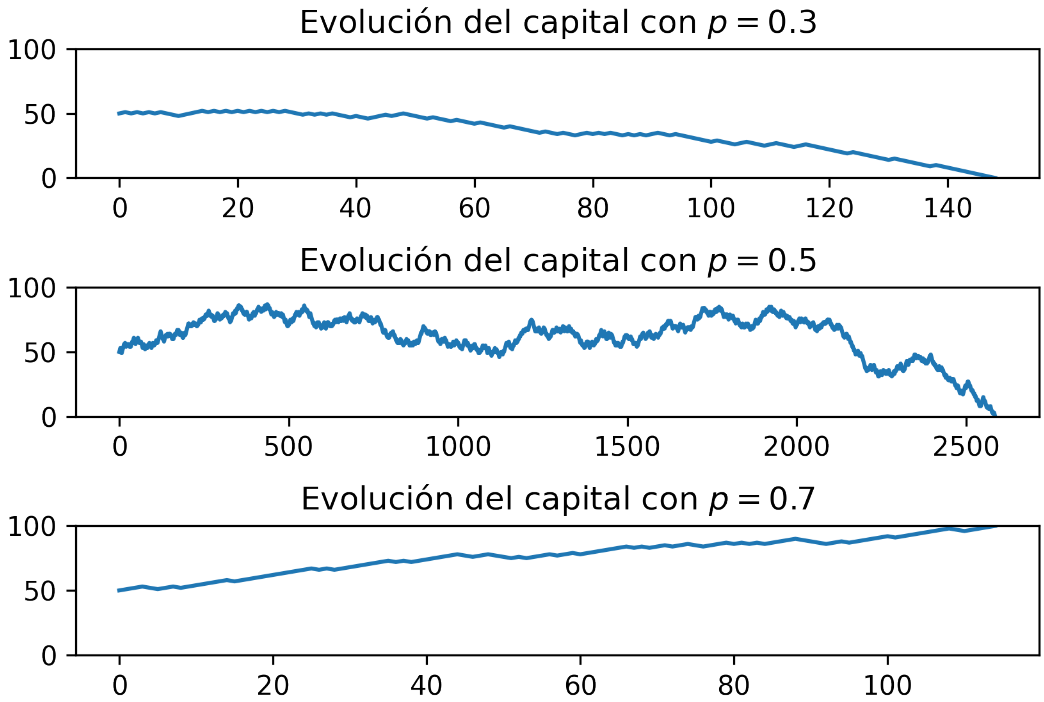

Figura 6: Simulación de la ruina del apostador para distintos valores de p. Notar las diferentes escalas en los ejes. Eje x: Número de jugada. Eje y: Capital actual

Apartado b

Para calcular la probabilidad de ruina, simularemos un gran número de caminatas (1000 para cada valor de p), y calcularemos la frecuencia relativa del evento R: "El jugador llega al estado 0". Al realizar el experimento obtenemos

p	Probabilidad de ruina
0,3	1
0,5	0,525
0,7	0

Tabla 2: Probabilidades de ruina para distintos valores de p

Analíticamente, basándonos en la teoría sabemos que la probabilidad de ruina partiendo de k es $F(k,0)$, es decir, la probabilidad de llegar por primera vez de k a 0 en una cantidad de pasos finita. Si tomásemos $S=3$, la matriz F se vería de la siguiente manera

$$F = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \dfrac{q}{1-pq} & \dfrac{p^2}{1-pq} & pq & p \\ \dfrac{q^2}{1-pq} & \dfrac{p}{1-pq} & q & pq \end{pmatrix}$$

Donde $q=(1-p)$ y k tomaría el valor 1 o 2. Cabe aclarar que el orden de las filas y columnas en F es 0, 3, 1 y 2 (Ya que separamos los elementos de los conjuntos cerrados irreducibles de los estados transitorios).

Ejercicio 4

Para estimar M_n podemos realizar una gran cantidad de ejecuciones del algoritmo y promediar las comparaciones hechas en cada ejecución. Para nuestro experimento inicial tomamos $n = 5$.

Sin pérdida de generalidad usamos permutaciones del conjunto $\{1, 2, 3, 4, 5\}$ como entrada. Corriendo el programa 100000 veces, el número promedio de comparaciones que obtenemos es 7,40486 comparaciones.

Apartado a

De la misma manera que usamos anteriormente, ejecutamos varias veces el algoritmo sobre permutaciones del conjunto $\{1, 2, 3, 4, 5, 6, 7\}$ y obtenemos un promedio de comparaciones igual a 13,48699.

Apartado b

Definimos la variable aleatoria C_n como el número de comparaciones hechas sobre el conjunto de números de longitud n al ejecutar el algoritmo quick-sort.

Intentamos encontrar una fórmula analítica para la cantidad de comparaciones sobre un conjunto de n números. En el algoritmo, lo primero que se hace es elegir al azar un elemento, y luego comparar todos los otros elementos con él. Esto nos garantiza tener al menos $(n - 1)$ comparaciones. Luego, se aplica recursivamente el algoritmo sobre los conjuntos de elementos menores y mayores. Formalmente podríamos decir

$$C_n = (n - 1) + C_L + C_R$$

Donde L y R son la cantidad de elementos menores y mayores al pivot respectivamente.

Un dato importante a notar es que el pivot se elige uniformemente al azar. Como consideramos que todos los elementos son distintos, esto quiere decir que hay una correspondencia biunívoca entre el pivot elegido y la cantidad de elementos menores (y por lo tanto, de la cantidad de mayores). Más aún, esto quiere decir que la cantidad de elementos menores L es una variable aleatoria que también está distribuida uniformemente. L puede tomar cualquier valor entre 0 y $n - 1$ con igual probabilidad. Dado un L, R es simplemente $n - 1 - L$, por lo que podemos decir que

$$C_n = (n - 1) + C_L + C_{n-1-L}$$

Buscamos calcular la cantidad esperada de comparaciones, es decir, $E(C_n)$. Para esto, podemos usar el método de esperanza condicionada basándonos en la variable L. Es decir

$$E(C_n) = E(E(C_n/L)) = \sum_{i=0}^{n-1} E(C_n/L = i) * P(L = i)$$

Podemos expandir esta expresión, y utilizar el hecho de que la distribución de L es uniforme.

$$E(C_n) = \sum_{i=0}^{n-1} (n - 1 + C_i + C_{n-1-i}) * \frac{1}{n}$$
$$E(C_n) = \frac{\sum_{i=0}^{n-1} (n - 1 + C_i + C_{n-1-i})}{n}$$

Podemos separar esta sumatoria y obtener

$$E(C_n) = \frac{n(n-1) + \sum_{i=0}^{n-1} C_i + \sum_{i=0}^{n-1} C_{n-1-i}}{n}$$

Vemos que las dos sumatorias restantes son en realidad la misma. Entonces tenemos

$$E(C_n) = \frac{n(n-1) + 2\sum_{i=0}^{n-1} C_i}{n} = (n-1) + \frac{2}{n} \sum_{i=0}^{n-1} C_i$$

Concluimos que podemos obtener la esperanza de C_n dados los C_i anteriores. Para $i \leq 2$, C_i es un valor fijo, con $C_0 = C_1 = 0$ y $C_2 = 1$. Con estos valores podemos calcular $E(C_3)$. Luego, lo que podemos hacer es utilizar las esperanzas de los casos anteriores para calcular la esperanza actual. Así, por ejemplo para C_4 tenemos

$$E(C_4) = (4-1) + \frac{2}{4}(0 + 0 + 1 + E(C_3))$$

Donde $E(C_3)$ es, a su vez

$$E(C_3) = (3-1) + \frac{2}{3}(0 + 0 + 1) = \frac{8}{3}$$

Reemplazando este valor en la ecuación de $E(C_4)$, obtenemos

$$E(C_4) = 29/6$$

Mediante este procedimiento recursivo calculamos $E(C_7)$ (Es decir M_7). El resultado que obtuvimos es

$$M_7 = \frac{472}{35}$$

Lo cual coincide con los resultados experimentales obtenidos.

Ejercicio 5

Para simular la caminata aleatoria, consideraremos una cadena de Markov cuya matriz de transición se basa en el grafo. Como la probabilidad de pasar de una página a otra que tenga un enlace a ella es igual para todos los enlaces de la primer página, las entradas de una determinada fila i serán 0 o $\frac{1}{e_i}$, donde e_i es la cantidad de enlaces salientes desde el i-ésimo estado.

Apartado a

Ya que el ultimo estado no tiene enlaces saliendo de el, no es posible describir una cadena de Markov debido a que los valores en su fila no sumarían 1. Para solucionar esto consideramos que hay una misma posibilidad de ir hacia cualquiera de las otras paginas, por lo que la matriz de transición de la cadena de Markov quedaría de la siguiente manera:

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \end{pmatrix}$$

Apartado b

La caminata de 100 pasos arrojó los siguientes resultados

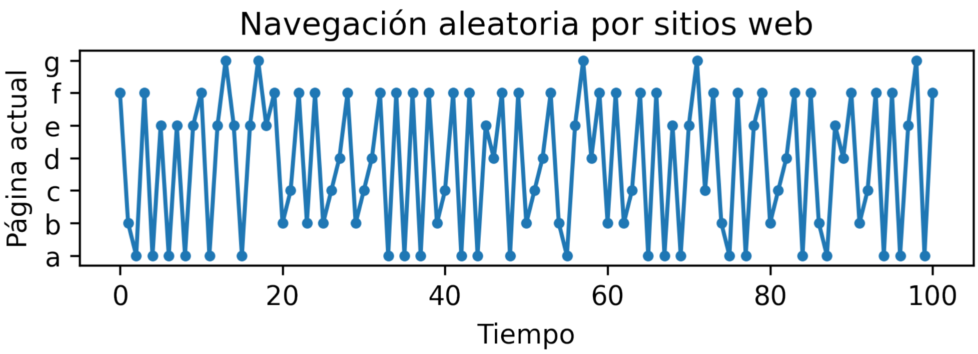

Figura 7: Navegación aleatoria por el WebGraph

Además de esto, agregaremos el porcentaje de tiempo que se pasó en cada página.

Estado	Frecuencia relativa
a	0,23
b	0,14
c	0,09
d	0,07
e	0,14
f	0,29
g	0,05

Tabla 3: Frecuencias relativas de cada estado

Apartado c

Buscamos calcular el rango de cada página, el cual se define como la probabilidad de estar en un determinado estado en algún momento. Esto no es más que la distribución invariante π de nuestra cadena de Markov. Como nuestra cadena es irreducible y aperiódica, podemos encontrar la distribución a partir del sistema de ecuaciones

$$\pi P = \pi$$
$$\sum_{i=1}^{7} \pi_i = 1$$

Resolviendo obtenemos la siguiente distribución:

$$\pi^T = \begin{pmatrix} 0{,}24533333 \\ 0{,}16 \\ 0{,}05866667 \\ 0{,}06666667 \\ 0{,}128 \\ 0{,}30933333 \\ 0{,}032 \end{pmatrix}$$

Ejercicio 6

Apartado a

Para simular el proceso de Poisson, el cual es teóricamente continuo, consideramos intervalos discretos de 1 segundo. En cada uno de esos intervalos, tenemos una probabilidad de $\frac{\lambda}{3600}$ de que llegue un vuelo. Cabe aclarar que esto es simplemente una aproximación de lo que pasaría en el proceso continuo. Sin embargo, podemos asegurar por el lema 3 de los procesos de Poisson que con un intervalo lo suficientemente pequeño la diferencia de este modelo con el proceso continuo se aproxima a 0.

Nuestra simulación discreta produjo el siguiente resultado

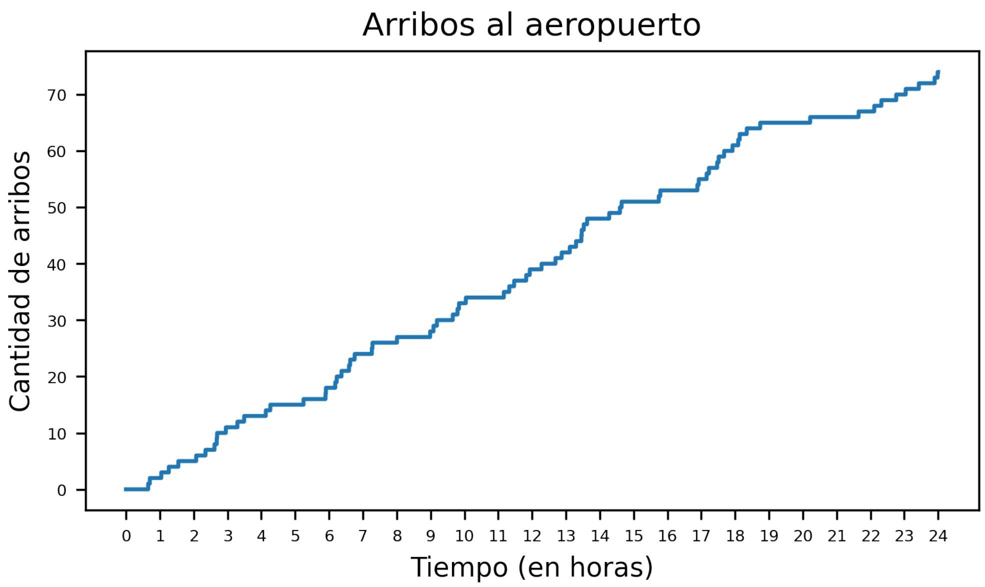

Figura 8: Arribos en 24 horas

El promedio de arribos por hora fue de 3.08, lo cual concuerda con lo que esperaríamos en el proceso continuo real.

Apartado b

Simulamos el proceso de Poisson de la misma manera que en el apartado anterior sólo que esta vez ampliamos la duración de la simulación a una semana. El gráfico de la cantidad de arribos se ve de la siguiente manera:

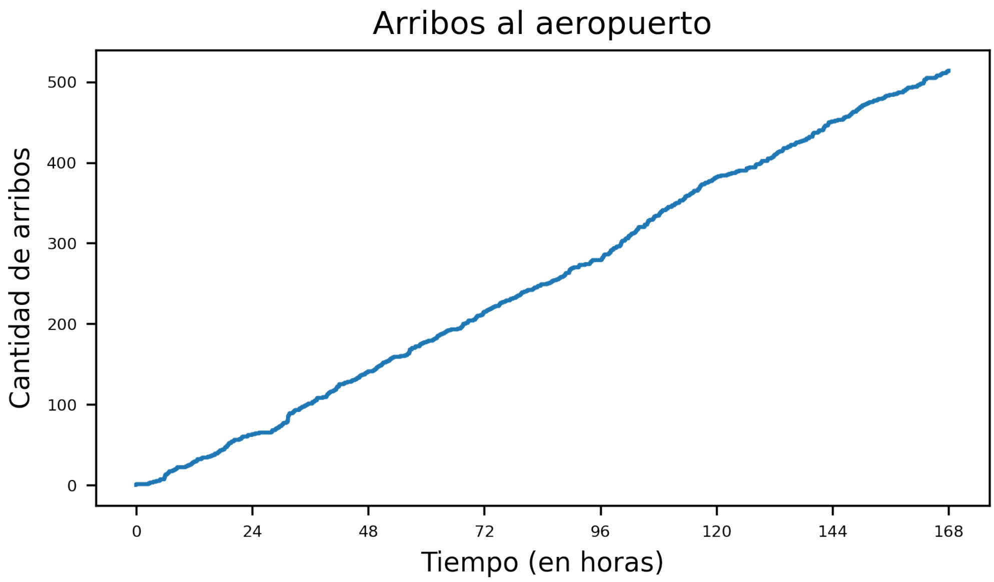

Figura 9: Arribos en una semana

A partir de esta nueva simulación con más datos, calculamos los tiempos entre los arribos. Luego, generamos el siguiente histograma

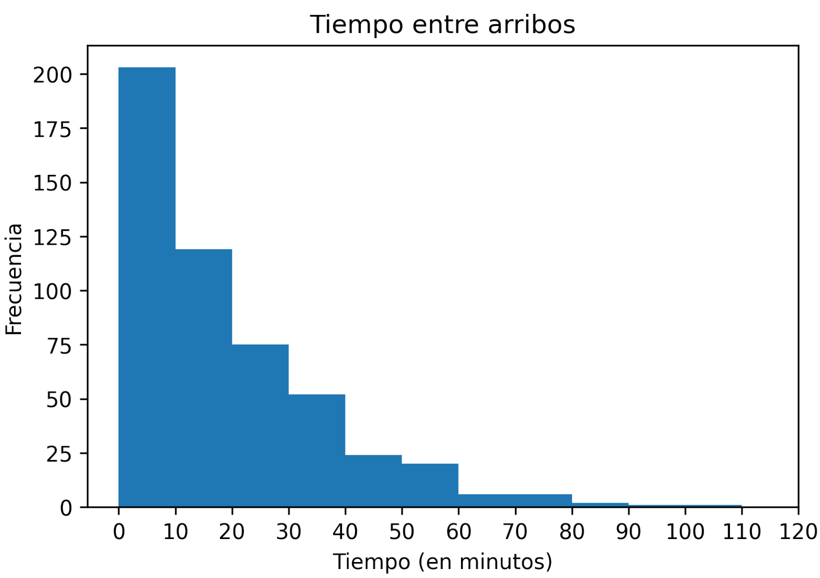

Figura 10: Histograma de los tiempos entre vuelos

Vemos en este histograma que la distribución de esta variable se asemeja a una exponencial. En un proceso de Poisson continuo, efectivamente sería exponencial con parámetro λ.

El promedio de tiempo entre arribos es de 0,3389 horas, que es lo que esperaríamos de la esperanza de una variable con distribución exponencial ($E(T) = \dfrac{1}{\lambda}$, en nuestro caso $\dfrac{1}{3}$).

Ejercicio 7

Apartado a

Tenemos una distribución inicial $\pi_0 = (0, 0, 0, 0, 1)$, esto nos dice que comenzamos nuestro proceso sin ningún ataque. Sabemos que nuestra matriz P representa la evolución de las probabilidades de ataque a cada puerto semana a semana. Por lo tanto, para calcular las probabilidades de ataque en la próxima semana, bastará con multiplicar nuestra distribución π_0 por P. De esta manera conseguimos $\pi_1 = (0, \frac{1}{8}, \frac{1}{2}, \frac{1}{8}, \frac{1}{4})$. Similarmente, para obtener nuestra distribución (las probabilidades de ataque a cada puerto) dentro de 3 semanas, simplemente tendremos que multiplicar nuestra distribución inicial por P 3 veces. Esto es:

$$\pi_3 = \pi_0 * P * P * P = \pi_0 * P^3 = (\frac{885}{36608}, \frac{1268269}{5234944}, \frac{911497}{2617472}, \frac{152189}{654368}, \frac{399807}{2617472})$$

Observando esta nueva distribución, notamos que los estados más y menos probables son “Ataque al puerto 135” y “No hay ataque” respectivamente.

Apartado b

Siendo nuestra cadena de Markov irreducible y aperiódica, nos es posible conseguir la distribución límite a partir de las ecuaciones

$$\pi P = \pi$$
$$\sum_{i=1}^{5} \pi_i = 1$$

Resolviendo obtenemos la siguiente distribución:

$$\pi^T = \begin{pmatrix} 0{,}021 \\ 0{,}267 \\ 0{,}343 \\ 0{,}227 \\ 0{,}141 \end{pmatrix}$$