CAMINATAS ALEATORIAS

GERÓNIMO URIBE BRAVO

Contents

1. La caminata aleatoria simple y simétrica

Una caminata aleatoria es uns sucesión de variables aleatorias reales $(S_n)_{n=0}^{\infty}$ tal que $S_0 = 0$ y las variables aleatorias $(S_i - S_{i-1})_{i=1}^{\infty}$ son independientes e idénticamente distribuidas. Si $X_i = S_i - S_{i-1}$, entonces a la distribución de X_i, que es independiente de i, y a la que daremos el nombre de distribución del salto, gobierna a la caminata aleatoria y las características de esta última se obtienen a través del estudio de dicha distribución. Si X_i tiene una distribución Bernoulli simétrica de parámetro $1/2$, es decir

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2},$$

a la caminata aleatoria se le conoce como caminata aleatoria simple y ciertas distribuciones relacionadas con esta sucesión se encuentran mediante métodos combinatorios; sin embargo, para el caso general, los métodos a utilizar generalmente son más complicados y el objetivo de su estudio es expresar la solución al problema planteado en términos de las caminatas aleatorias a través de la función de distribución. A veces tal solución es muy complicada y no es de utilidad práctica, por lo que además de la solución exacta, generalmente buscamos expresiones aproximadas más sencillas. Este enfoque se puede utilizar aún cuando la solución exacta a nuestro problema nos sea desconocida.

1.1. **Cálculos distribucionales básicos.** Notemos que la evolución de la caminata aleatoria hasta el tiempo n queda determinada por el camino que sigue. Definimos a un camino como un conjunto de puntos

$$\{(i, j_i) : i = 0 \dots n\} \subset \mathbb{N} \times \mathbb{Z},$$

para el cual

$$j_i = j_{i-1} \pm 1,$$

con $i = 2 \dots n$. [1] Notemos que, puesto que los incrementos de la caminata aleatoria son independientes, entonces

$$\mathbb{P}(S_1 = j_1, \dots, S_n = j_n) = \mathbb{P}(X_1 = j_1, X_2 = j_2 - j_1, \dots, X_n = j_n - j_{n-1}) = \frac{1}{2^n}.$$

Esto nos dice que la caminata aleatoria sigue cualquier camino con la misma probabilidad, por lo que el recorrido de la caminata aleatoria en n pasos es uniforme en los caminos de longitud n. Notemos que en particular nos dice que hay 2^n caminos, pues la suma sobre todos los caminos debe ser 1. Esto también lo podemos ver al notar que un camino está determinado por sus incrementos, que son cualquier sucesión de longitud n de $-1, 1$. Claramente hay 2^n de estas sucesiones. Así, para encontrar la probabilidad de que la evolución de la caminata aleatoria simple hasta el tiempo n tenga cierta propiedad, una posibilidad es encontrar la cantidad de caminos para los cuales se cumple dicha propiedad y multiplicar por $1/2^n$. Éste es un método combinatorio.

1.1.1. *Coeficientes binomiales.* Por ejemplo, podemos calcular la probabilidad de que S_n tome el valor k. A este evento lo escribimos como $\{S_n = k\}$. Efectivamente, un camino que comience en cero y termine en k debe subir una cantidad de veces x y bajar $x - k$ veces. Si éste camino tiene longitud n entonces $2x + k = n$, por lo que $x = (n - k)/2$. Así, sólo tenemos que contar cuántos caminos tienen x incrementos hacia arriba. Hay $\begin{pmatrix} n \\ (n-k)/2 \end{pmatrix}$ caminos con estas condiciones y por lo tanto

$$\mathbb{P}(S_n = k) = \frac{\begin{pmatrix} n \\ (n-k)/2 \end{pmatrix}}{2^n}.$$

La anterior expresión es cero si n y k no tienen la misma paridad.

1.1.2. *La fórmula de Stirling y el teorema de de Moivre y Laplace.* Al graficar a la función $\mathbb{P}(S_{2n} = k)$, vemos que se parece a la gráfica de la función $e^{-k^2/2n}/\sqrt{\pi n/2}$. Por ejemplo, podemos utilizar el siguiente código en Octave para hacerlo:

Listing 1. dML.notm

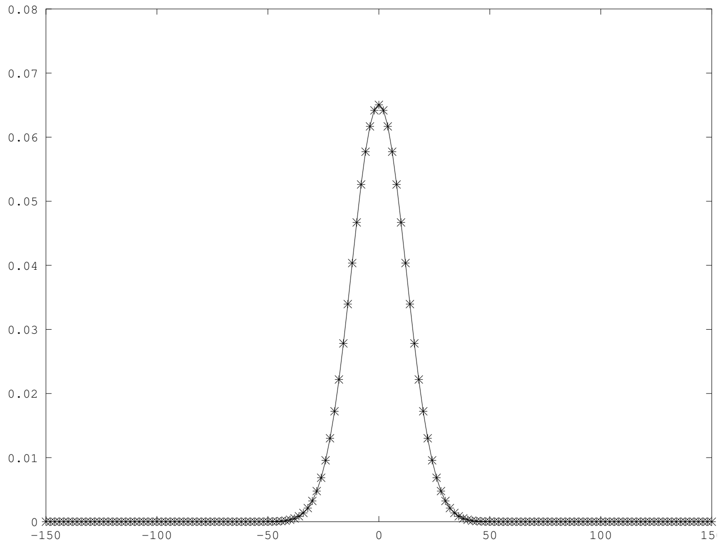
```
1;
# La funcion binom calcula Proba(S_n=k)
# para una caminata aleatoria simple y simetrica.
function p=binom(n,k)
        p=factorial(n)./factorial((n-k)/2)./factorial((n+k)/2)./2^n;
endfunction
n=150;
k=(-n):2:(n);
# Se calculan las probabilidades de terminar en las entradas de k
p=binom(n,k);
# Se grafica junto con la aproximacion.
plot(k,p,'*',k,exp(-k.*k/n/2)/sqrt(pi*n/2),'k')
```

El resultado se puede ver en la Figura 1.1.2.

La aproximación anterior se puede explicar mediante el llamado teorema de de Moivre y Laplace que a su vez admite una prueba basada en la aproximación de Stirling para el factorial. La aproximación de Stirling afirma que

$$n! \sim n^n e^{-n} \sqrt{2\pi n},$$

[1] En la pareja ordenada (i, j_i) perteneciente a un camino, i es el número de pasos y j_i es la posición de la caminata aleatoria.



donde el símbolo $\sim$ significa equivalencia asintótica y significa que el cociente de ambas cantidades converge a 1 conforme $n \to \infty$. A la función $f_G(x) = e^{-x^2/2}/\sqrt{2\pi}$ se le conoce con el nombre de densidad gaussiana y aparece en el teorema de de Moivre y Laplace en la siguiente forma:

Teorema 1 (de de Moivre y Laplace para la caminata aleatoria simple)**.** *Conforme $n, k \to \infty$, con $k = o\big(n^{2/3}\big)$ de la misma paridad que n, se tiene la equivalencia asintótica*

$$\mathbb{P}(S_n = k) \sim e^{-k^2/2n} \frac{\sqrt{2}}{\sqrt{\pi n}}.$$

En particular, si $k/\sqrt{n} \to x$ entonces

$$\mathbb{P}(S_n = k) \sim e^{-x^2/2n} \frac{\sqrt{2}}{\sqrt{\pi n}}.$$

Demostración: Al utilizar la aproximación de Stirling, vemos que si $n, k \to \infty$ entonces

$$\begin{aligned}\binom{n}{(n+k)/2} 2^{-n} &\sim \frac{n^n e^{-n} \sqrt{2\pi n}}{2^n \left(\frac{n+k}{2}\right)^{\left(\frac{n+k}{2}\right)} e^{\frac{n+k}{2}} \sqrt{2\pi \frac{n+k}{2}} \left(\frac{n-k}{2}\right)^{\left(\frac{n-k}{2}\right)} e^{\frac{n-k}{2}} \sqrt{2\pi \frac{n-k}{2}}} \\ &= \frac{\sqrt{2}}{\sqrt{\pi n}} \frac{1}{\left(1 + \frac{k}{n}\right)^{\frac{n+k}{2}} \left(1 - \frac{k}{n}\right)^{\frac{n-k}{2}} \sqrt{(1 + k/n)(1 - k/n)}}.\end{aligned}$$

Si $k/n \to 0$ conforme $n \to \infty$ y definimos $p = k/n$, obtenemos por lo tanto

$$\binom{n}{(n+k)/2} 2^{-n} \sim \frac{\sqrt{2}}{\sqrt{\pi n}} e^{-\frac{n}{2} f(p)}$$

donde $f(p) = (1+p)\log(1+p) + (1-p)\log(1-p)$. Puesto que $f(0) = f'(0) = 0$ y $f''(0) = 2$, se tiene que $f(p) = p^2 + o\big(p^3\big)$ conforme $p \to 0$. Se sigue entonces que

$$\binom{n}{(n+k)/2} \sim \frac{\sqrt{2}}{\sqrt{\pi n}} e^{-\frac{n}{2} p^2} \left(1 + o\big(np^3\big)\right).$$

Recordemos que $np^2/2 = k^2/2n$. Además, si $k = o\big(n^{2/3}\big)$ entonces $np^3 \to 0$, por lo que:

$$\binom{n}{(n+k)/2} \sim \frac{\sqrt{2}}{\sqrt{\pi n}} e^{-k^2/2n}. \qquad \square$$

Al hacer un análisis más preciso (pero no mucho más complicado) con los residuos, podemos ver que de hecho

$$\sup_{k:\,|k|\leq n^{2/3}} \left| \frac{\mathbb{P}(S_n = k)}{e^{-k^2/2n}\sqrt{2}/\sqrt{\pi n}} - 1 \right| \to 0$$

conforme $n \to \infty$.

1.1.3. *Prueba de la aproximación de Stirling.* Ahora utilizaremos un argumento clásico, conocido como método de Laplace, para obtener la aproximación de Stirling. Dicho método se puede consultar en [dB58]. Comenzamos por recordar el valor de la integral

$$I = \int_{-\infty}^{\infty} e^{x^2/2\sigma^2}\, dx.$$

Puesto que

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\left(x^2+y^2\right)/2\sigma^2}\, dx\, dy,$$

un cambio a coordenadas polares ($x = r\cos\theta$ y $y = r\sin\theta$ con $r > 0$ y $0 \leq \theta < 2\pi$) nos da:

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2\sigma^2} r\, dr d\theta = 2\pi \int_0^{\infty} e^{-y}\sigma^2\, dy = 2\pi\sigma^2.$$

Se concluye que $I = \sqrt{2\pi\sigma^2}$.

Ahora, recordemos la representación integral para la función factorial:

$$n! = \int_0^{\infty} x^n e^{-x}\, dx = \int_0^{\infty} e^{-\tilde{h}(x)}\, dx.$$

Ejercicio 1. *Pruebe la representación integral del factorial mediante el siguiente argumento:*

(1) *Defina* $f(n) = \int_0^{\infty} x^n e^{-x}\, dx$ *y pruebe que* $f(0) = 1$.
(2) *Integre por partes para deducir que* $f(n+1) = (n+1)\, f(n)$ *y concluya.*

Para verificar la fórmula de aproximación de Stirling, primero **hacemos** el cambio de variable $x = n\,(1+y)$ para obtener:

$$n! = n^{n+1}e^{-n} \int_0^{\infty} e^{-ny}\,(1+y)^n\; dy = n^{n+1}e^{-n} \int_{-1}^{\infty} e^{-nh(y)}.$$

La idea de este cambio de variable es que el mínimo de la función h se alcance en cero y sea igual a cero. El segundo paso es analizar a la función $\exp -nh(y)$ para n grande. Esta función es 1 cuando $y = 0$ y si $y \neq 0$, tiende a cero. Por lo tanto, su integral debería estar cada vez más localizada alrededor de $y = 0$. Para formalizar esta heurística, **notemos** que $h(y) = y^2 + o\big(y^3\big)$ y que h es convexa pues $h'' > 0$. Puesto que $h(1) > 0$ y $h'(1) = 1/2 > 0$, se sigue que $h(y) \geq h(1) + y/2$. Sea $\delta = n^{-\alpha}$ donde $\alpha \in (1/3, 1/2)$. Veremos que la integral se localiza en $[-\delta, \delta]$. Notemos que $nh(\delta) \sim n^{1-2\alpha}$.

Podemos entonces dividir la integral en lo que pasa en $[-\delta, \delta]$ y en el complemento. En el complemento, notemos que

$$\int_{-1}^{-\delta} e^{-nh(y)}\, dy \leq e^{-nh(-\delta)}\,(1-\delta) = O\Big(e^{-n^{1-2\alpha}}\Big),$$

$$\int_{\delta}^{1} e^{-nh(y)}\, dy \leq e^{-nh(\delta)}\,(1-\delta) = O\Big(e^{-n^{1-2\alpha}}\Big)$$

y

$$\int_1^{\infty} e^{-nh(y)}\, dy \leq e^{-nh(1)} \int_1^{\infty} e^{-ny/2}\, dy = e^{-nh(1)}\frac{2}{n}e^{-n/2}.$$

Por otra parte, en $[-\delta,\delta]$ podemos escribir

$$\begin{aligned}\int_{-\delta}^{\delta} e^{-h(y)}\,dy &= \big(1+o(n\delta^3)\big)\int_{-\delta}^{\delta} e^{-ny^2/2}\,dy\\ &= \big(1+o(n\delta^3)\big)\sqrt{n}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} e^{-z^2/2}\,dz\\ &\sim \sqrt{2\pi/n}.\end{aligned}$$

Se deduce que

$$\int_{-1}^{\infty} e^{-nh(y)}\,dy \sim \int_{-\delta}^{\delta} e^{-nh(y)}\,dy.$$

y por lo tanto

$$n! \sim n^n e^{-n}\sqrt{2\pi n}.$$

1.1.4. *El teorema límite central.* Ahora veremos una consecuencia del Teorema de de Moivre y Laplace conocida como teorema límite central. Es un caso particular de un teorema fundamental en la teoría de la probabilidad.

Teorema 2 (Teorema límite central para la caminata aleatoria simple y simétrica)**.** *Conforme $n\to\infty$,*

$$\mathbb{P}\left(\frac{S_n}{\sqrt{n}} \le x\right) \to \int_{-\infty}^{x} e^{-y^2/2}\frac{1}{\sqrt{2\pi}}\,dy.$$

Otra manera de expresarlo es que $S_n/\sqrt{n}$ tiene una distribución límite conforme $n\to\infty$ y que esta distribución es gaussiana. (Se dice que una variable aleatoria X es gaussiana si $\mathbb{P}(X\le x)=\int_{-\infty}^{x} e^{-y^2/2}\,dy$).

Demostración: Comenzaremos por probar que

$$\mathbb{P}\left(\frac{S_n}{\sqrt{n}} \in (a,b)\right) \to \int_a^b e^{-y^2/2}\,dy$$

para cualesquiera $a<b$. En el intervalo (a,b), al aproximar a la integral en términos de sumas de Riemann, obtenemos

$$\lim_{n\to\infty}\sum_{\substack{\frac{k+n}{2}\in\mathbb{N}\\ a\le k/\sqrt{n}\le b}} e^{-k^2/2n}\frac{2}{\sqrt{2\pi n}} = \int_a^b e^{-x^2/2}\frac{1}{\sqrt{2\pi}}\,dx$$

Por otra parte gracias al teorema de de Moivre y Laplace observamos que

$$\sup_{k:\sqrt{n}a\le k\le\sqrt{n}b}\left|\frac{\mathbb{P}(S_n=k)}{e^{-k^2/2n}\sqrt{2}/\sqrt{\pi n}}-1\right|\to 0,$$

por lo cual

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}\big(S_n/\sqrt{n}\in(a,b)\big) &= \lim_{n\to\infty}\sum_{\substack{\frac{k+n}{2}\in\mathbb{N}\\ a\le k/\sqrt{n}\le b}}\mathbb{P}(S_n=k)\\ &= \lim_{n\to\infty}\sum_{\substack{\frac{k+n}{2}\in\mathbb{N}\\ a\le k/\sqrt{n}\le b}} e^{-k^2/2n}\frac{2}{\sqrt{2\pi n}}\\ &= \int_a^b e^{-x^2/2}\,dx.\end{aligned}$$

Recordemos que $1 = \int_{-\infty}^{\infty} e^{-x^2/2}/\sqrt{2\pi}\, dx$. Por lo tanto, para $x > 0$:

$$\int_{-\infty}^{-x} e^{-y^2/2}/\sqrt{2\pi}\, dy = \frac{1}{2}\left[\int_{-\infty}^{-x} e^{-y^2/2}/\sqrt{2\pi}\, dy + \int_{x}^{\infty} e^{-y^2/2}/\sqrt{2\pi}\, dy\right]$$
$$= \frac{1}{2}\left[1 - \int_{-x}^{x} e^{-y^2/2}/\sqrt{2\pi}\, dy\right]$$

Notemos también que S_n tiene una distribución simétrica, por lo cual, si $x > 0$:

$$\mathbb{P}\big(S_n/\sqrt{n} \leq -x\big) = \frac{1}{2}\left[1 - \mathbb{P}\big(S_n/\sqrt{n} \in (-x, x)\big)\right].$$

Se sigue entonces que para $x > 0$:

$$\mathbb{P}\big(S_n/\sqrt{n} \leq -x\big) = \int_{-\infty}^{x} e^{-y^2/2}/\sqrt{2\pi}\, dy.$$

El caso restante se resuelve notando que si $x > 0$ entonces $\mathbb{P}(S_n/\sqrt{n} \leq x) \sim 1/2 + \mathbb{P}(S_n/\sqrt{x} \in (0, x))$. $\square$

1.2. **La caminata aleatoria Bernoulli y la variable aleatoria geométrica.** Ahora consideraremos un problema en el que aparece una variable aleatoria importante en la probabilidad: la variable aleatoria geométrica. El problema, dividido en dos partes, es sencillo de enunciar: dada una sucesión de volados (con una moneda cargada en la que sale águila con probabilidad $p \in (0, 1)$), ¿Podemos asegurar que eventualmente saldrá un águila? ¿Con qué probabilidad obtenemos exáctamente k soles antes de la primera águila? Probabilísticamente, el modelo que consideramos consta de una sucesión de variables aleatorias $X_1, X_2, \ldots$ que son independientes y tales que $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$. Al evento $\{X_i = 1\}$ lo interpretamos como que sale águila en el i-ésimo volado. Definamos entonces a la cantidad de soles hasta la primera águila como sigue: sean

$$\mathscr{A} = \{n \geq 1 : X_n = 1\} \quad \text{y} \quad T = \begin{cases} \infty & \text{si } \mathscr{A} = \emptyset \\ \min \mathscr{A} & \text{si } \mathscr{A} \neq \emptyset \end{cases}.$$

Entonces, para cada $k = 0, 1, 2, \ldots$:

$$\mathbb{P}(T = k) = \{X_1, \ldots, X_k = 0, X_{k+1} = 1\} = (1 - p)^k\, p.$$

Como $1 = \sum \mathbb{P}(T = k) = \mathbb{P}(T < \infty)$, vemos que $\mathbb{P}(T = \infty) = 0$. Esto responde al primer cuestionamiento: efectivamente, siempre observamos al menos un águila. De hecho, se verifica en cursos de probabilidad que hay una cantidad infinita de águilas con probabilidad uno. Por otra parte, T es lo que se conoce como una variable aleatoria geométrica, caracterizada por tener la propiedad de pérdida de memoria siguiente:

$$\mathbb{P}(T \geq n + m \mid T \geq m) = \mathbb{P}(T \geq n).$$

Ejercicio 2. *Pruebe la afirmación anterior sobre T y pruebe que cualquier variable aleatoria con valores naturales que satisfaga la propiedad de pérdida de memoria es una variable aleatoria geométrica.*

Si $S_n = X_1 + \cdots + X_n$, a la cantidad S_n se le interpreta como la cantidad de águilas en n volados. La interpretación frecuentista de la probabilidad nos dice que S_n/n se debería parecer a p para n grande. Esto puede justificarse a través de la fórmula de Stirling.

Ejercicio 3. *Pruebe que*

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Utilice la aproximación de Stirling, como en el caso de la caminata aleatoria simple y simétrica, para probar que

$$\mathbb{P}(S_n = k) \sim \frac{1}{\sqrt{2\pi np(1-p)}} e^{-k^2/2np(1-p)}$$

si $k - np = O(\sqrt{npq})$. *Sugerencia: ahora será útil definir la función* $f(x) = x \log x/p + (1-x)\log(1-x)/(1-p)$.

Ejercicio 4. *Al escribir*

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!},$$

pruebe que si $p/n \to \lambda$ *conforme* $n \to \infty$ *entonces* $\mathbb{P}(S_n = k) \to e^{-\lambda}\lambda^k/k!$.

1.3. **El problema de la ruina.** Pasaremos ahora a un problema clásico sobre la caminata aleatoria simple, el llamado problema de la ruina. Un jugador tiene un capital inicial de m pesos y se enrola en un juego de volados en el que gana un peso con probabilidad p (y los juegos son independientes). No deja de jugar hasta que pasan dos cosas: acumula un capital objetivo de n pesos o pierde toda su fortuna. Lo que se quiere determinar es la probabilidad de que termine de jugar con n pesos.

Acerca de este juego plantearemos 3 preguntas: ¿Es cierto que eventualmente occuren algunas de las dos posibilidades? En caso afirmativo, ¿Cuál es la probabilidad de que el jugador acumule su capital objetivo? ¿En promedio, cuánto dura el juego? La primera pregunta es la más sencilla y se responde a través de la caminata aleatoria Bernoulli. Consideremos $X_1, X_2, \ldots$ las variables independientes que nos dicen las ganancias sucesivas del jugador (si jugara indefinidamente) Se asume que $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = -1)$ con $p \in (0,1)$ y que las X_i son independientes. Ahora dividimos los volados en bloques de longitud n y notamos que si $X_{kn+1} = \cdots = X_{kn+n} = 1$ entonces el jugador alcanza su capital objetivo. Formalmente, podríamos definir a la ganancia acumulada $S_k = X_1 + \cdots + X_k$ y al instante de ruina

$$\mathscr{R} = \{k \geq 1 : m + S_k \in \{0, n\}\} \quad \text{y} \quad R = \begin{cases} \infty & \mathscr{R} = \emptyset \\ \min \mathscr{R} & \mathscr{R} \neq \emptyset \end{cases}.$$

Entonces notamos que

$$\{X_{kn+1} = \cdots = X_{kn+n} = 1 \text{ para alguna } k \geq 0\} \subset \{R < \infty\}.$$

Por otra parte, los bloques de n volados son independientes entre si, en cada bloque hay puros unos con probabilidad $p^n > 0$ y, por el resultado sobre la caminata aleatoria simple, eventualmente existirá un bloque de n volados con puros unos. Se deduce que eventualmente o el jugador se arruina o alcanza su capital objetivo.

Para la segunda pregunta, notemos que $m + S_R$ es la fortuna del jugador al instante R, que toma los valores 0 ó n. Por lo tanto $m + S_T = n$ si y sólo si el jugador alcanza su capital objetivo. Ahora nos fijaremos en lo que pasa en el primer paso: ya sea $X_1 = 1$ ó $X_1 = -1$. En el primer caso, el jugador alcanza el capital $m+1$ y el problema recomienza con este nuevo capital. Algo similar ocurre cuando $X_1 = -1$, por lo que obtenemos Escribamos

$$q_m = \mathbb{P}(\text{el jugador alcanza su capital objetivo comenzando con capital } m).$$

determinemos ahora el valor de q_m. Hay dos casos sencillos:

$$q_0 = 0 \quad \text{y} \quad q_n = 1.$$

Por otro lado, el análisis de primer paso nos dice que

$$q_m = p q_{m+1} + (1-p) q_{m-1}.$$

Así, la probabilidad de interés queda determinada por una relación de recurrencia con valores de frontera. Afortunadamente, la solución se conoce. En efecto, escribamos la relación de recurrencia en la forma

$$pq_m + qq_m = pq_{m+1} + qq_{m-1} \quad \text{ó} \quad pq_m + qq_m = pq_{m+1} + qq_{m-1}$$

ó inclusive

$$(q_m - q_{m+1}) = (q/p)\,(q_{m-1} - q_m)\,.$$

Se sigue entonces que

$$q_m - q_{m+1} = -q_1\,(q/p)^m$$

y por lo tanto

$$q_m = \begin{cases} q_1 \frac{1-(q/p)^m}{1-(q/p)} & \text{si } q \neq p \\ q_1 m & \text{si } q = p = 1/2 \end{cases}.$$

Al utilizar la igualdad $q_n = 1$ obtenemos finalmente

$$q_m = \begin{cases} \frac{1-(q/p)^m}{1-(q/p)^n} & \text{si } q \neq p \\ m/n & \text{si } q = p \end{cases}.$$

Ejercicio 5. *Determine la duración esperada del juego cuando $p = 1/2$. Dicha duración esperada se define como $v_m = \sum_k k\mathbb{P}(R = k$ con capital inicial $m)$. Pruebe que $v_0 = 0 = v_m$ y que $2v_m = 2 + v_{m+1} + v_{m-1}$. Defina $d_m = v_m - v_{m-1}$ y deduzca la igualdad matricial:*

$$\begin{pmatrix} d_{m+1} \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d_m \\ -2 \end{pmatrix},$$

por lo que

$$\begin{pmatrix} d_{m+1} \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^m \begin{pmatrix} d_1 \\ -2 \end{pmatrix}.$$

1.4. **La persistencia de la (mala) suerte.** Terminaremos nuestro análisis de las caminatas aleatorias simples al estudiar un resultado denominado por Feller de persistencia de la mala suerte. Sea $S_n, n \geq 0$ una caminata aleatoria simple y simétrica. Esto es, $S_n = X_1 + \cdots + X_n$ donde $X_1, X_2, \ldots$ son variables aleatorias independientes que toma los valres -1 y 1 ambos con probabilidad $1/2$. Para $n \geq 1$ fijo, consideremos a la cantidad de tiempo que pasa la caminata aleatoria por encima de cero. (Diremos que la caminata aleatoria pasa una unidad de tiempo por encima de cero en $[i, i+1]$ si $S_i > 0$ ó $S_{i-1} > 0$.) Esta vez, nos conformaremos con comprobar inductivamente el resultado siguiente: Si A_n es la cantidad de tiempo que pasa la caminata por encima de cero en n pasos, veremos que

$$\mathbb{P}(A_{2n} = 2k) = \mathbb{P}(S_{2k} = 0)\,\mathbb{P}\big(S_{2(n-k)} = 0\big)\,. \tag{1}$$

Recordemos las definiciones de f_n y u_n: f_n es la probabilidad de que el primer regreso a cero de la CAS ocurra en el instante n y u_n es la probabilidad de que S_n sea cero. Notemos que $\{S_n = 0\} \subset T \leq n$ y de hecho: $\{S_{2n} = 0\}$ es la unión ajena de los eventos $\{S_{2n} = 0, T = 2k\}$ para $k = 1, 2, \ldots, n$. Así:

$$\mathbb{P}(S_{2n} = 0) = \sum_{k=1}^{n} \mathbb{P}(S_{2n} = 0, T = 2k)\,.$$

Por otra parte: $\mathbb{P}(S_{2n}=0,T=2k)=\mathbb{P}(T=2k)\,\mathbb{P}\big(S_{2(n-k)}=0\big)=f_{2k}u_{2(n-k)}$. En efecto

$$
\begin{aligned}
&\mathbb{P}(S_{2n}=0,T=2k)\\
&=\mathbb{P}\big(X_1+X_2\neq 0,\dots,X_1+\cdots+X_{2(k-1)}\neq 0,X_1+\cdots+X_{2k}=0,X_{2k+1}+\cdots,X_{2k+2(n-k)}=0\big)\\
&=\mathbb{P}\big(X_1+X_2\neq 0,\dots,X_1+\cdots+X_{2(k-1)}\neq 0,X_1+\cdots+X_{2k}=0\big)\,\mathbb{P}\big(X_{2k+1}+\cdots,X_{2k+2(n-k)}=0\big)\\
&=\mathbb{P}(T=2k)\,\mathbb{P}\big(X_1+\cdots,X_{2(n-k)}=0\big)\\
&=\mathbb{P}(T=2k)\,\mathbb{P}\big(S_{2(n-k)}=0\big)\,.
\end{aligned}
$$

Lo anterior se puede escribir en forma compacta como:

$$u_{2n}=\sum_{k=1}^{n}f_{2k}u_{2n}. \tag{2}$$

Ahora sí comprobemos la ecuación (1). Dicha ecuación es claramente cierta si $k=0$ ó $k=n$. En efecto, cuando $k=0$ obtenemos

$$\mathbb{P}(A_{2n}=0)=\mathbb{P}(S_1\leq 0,\dots,S_{2n}\leq 0)=\frac{1}{2}\mathbb{P}(S_1\neq 0,\dots,S_1)\,.$$

Por simetría obtenemos el caso $k=n$.

Supongamos entonces que la ecuación (1) es válida hasta $n-1$ y mostrémosla para n y $1\leq k\leq n$. Puesto que $k\neq 0,n$, entonces $\{A_{2n}=2k\}\subset\{T\leq 2n\}$. Descomponemos entonces al evento de interés respecto del valor de T:

$$\mathbb{P}(A_{2n}=2k)=\sum_{j=1}^{n-1}\mathbb{P}(T=2j,A_{2n}=2k)\,.$$

Luego, notamos que en el evento $T=2j$, la caminata acumula ya sea $2j$ instantes de positividad o ninguno. Así:

$$\mathbb{P}(T=2j,A_{2n}=2k)=\mathbb{P}(T=2j,S_1>0,A_{2n}-A_{2j}=2(k-j))$$

y

$$\mathbb{P}(T=2j,A_{2n}=2k)=\mathbb{P}(T=2j,S_1<0,A_{2n}-A_{2j}=2k)\,.$$

Por otra parte,

$$\mathbb{P}(T=2j,S_1>0,A_{2n}-A_{2j}=2(k-j))=\frac{1}{2}\mathbb{P}(T=2j)\,\mathbb{P}\big(A_{2(n-j)}=2(k-j)\big)\,,$$

como queda claro al notar que todo camino de longitud $2n$ tal que $T=2j,S_1>0$ y $A_{2n}-A_{2j}$ está en correspondencia con dos caminos de longitudes $2j$ y $2(n-j)$. El primero es siempre no negativo y regresa por primera vez a cero en $2j$ y en el segundo la caminata es positiva $2(k-j)$ instantes de tiempo. Análogamente, tenemos la fórmula

$$\mathbb{P}(T=2j,S_1<0,A_{2n}-A_{2j}=2(k-j))=\frac{1}{2}\mathbb{P}(T=2j)\,\mathbb{P}\big(A_{2(n-j)}=2k\big)\,.$$

Por otra parte, notemos que si $A_{2n}=2k$ y $S_1>0$ entonces $T\leq 2k$, mientras que si $A_{2n}=2k$ y $S_1<0$ entonces $T\leq 2(n-k)$.

Al aplicar la hipótesis de inducción y (2), obtenemos:

$$
\begin{aligned}
\mathbb{P}(A_{2n}=2k)&=\sum_{j=1}^{k}\frac{1}{2}f_{2j}u_{2(n-j),2(k-j)}+\sum_{j=1}^{n-k}\frac{1}{2}f_{2j}u_{2(n-j),2k}\\
&=\sum_{j=1}^{k}\frac{1}{2}f_{2j}u_{2(n-k)}u_{2(k-j)}+\sum_{j=1}^{n-k}\frac{1}{2}f_{2j}u_{2(n-j)}u_{2k}\\
&=\frac{1}{2}u_{2(n-k)}u_{2k}+\frac{1}{2}u_{2k}u
\end{aligned}
$$

Ejercicio 6. *Aplique la fórmula de Stirling para deducir que* $\mathbb{P}(A_{2n} = 2k) \sim \frac{1}{\pi\sqrt{x(1-x)}}$ *si* $x \in (0,1)$ *y* $k/n \to x$ *conforme* $n \to \infty$.

References

[dB58] N. G. de Bruijn, *Asymptotic methods in analysis*, Bibliotheca Mathematica. Vol. 4, North-Holland Publishing Co., Amsterdam; P. Noordhoff Ltd., Groningen; Interscience Publishers Inc., New York, 1958. MR 0099564 (20 #6003)

Instituto de Matemáticas, Universidad Nacional Autónoma de México