Miscelánea Matemática **69** (2019) 31-41 SMM

Algunos cálculos sobre el conteo de patrones

Addy Bolívar-Cimé
División Académica de Ciencias Básicas
Universidad Juárez Autónoma de Tabasco
Villahermosa, Tab.
addy.bolivar@gmail.com
y
Aroldo Pérez
División Académica de Ciencias Básicas
Universidad Juárez Autónoma de Tabasco
Villahermosa, Tab.
aroldopz2@gmail.com

1. Introducción

La martingala[1] es una estrategia de apuesta muy popular y frecuentemente utilizada. Esta estrategia consiste en doblar la cantidad apostada cada vez que se pierda, de manera que cuando se gane, se recupere más de lo que se lleve perdido; después de ganar se inicia de nuevo con la cantidad inicialmente apostada. Aparentemente esta estrategia es una estrategia de éxito seguro, lo cual sería el caso (véase [1, ej. 3.6]) si se contara con capital y tiempo ilimitados. Además de que tales condiciones son imposibles, debe considerarse que cada mesa de juego de los casinos tiene un límite máximo de apuesta, que puede ser alcanzado por la suma de nuestras pérdidas en una racha perdedora y hacer entonces, imposible la aplicación de la martingala, es decir, doblar de nuevo la apuesta y recuperar el dinero perdido. En tal contexto, surgen las siguientes preguntas: ¿cuál es la probabilidad de una mala racha que nos produzca pérdidas cuya suma alcance el límite máximo de apuesta y nos impida por lo tanto, volver a doblar la apuesta?, ¿cuál es el número promedio de jugadas para la aparición de cierta racha perdedora?

A manera de ilustración, considere la siguiente situación hipotética: dos personas, digamos, Juan y Pedro, pactan jugar a los volados usando

Palabras clave: Ocurrencia de patrones, tiempos medios, probabilidades de textos.
[1]No confundir con la definición de martingala de la teoría de los procesos estocásticos.

una moneda justa; Juan siempre apostará a «águila» y Pedro a «sol». De esta manera, si la apuesta en un lanzamiento dado es de x pesos y cae águila, Juan recibe de Pedro x pesos; en caso contrario, Pedro recibe de Juan x pesos. Convienen realizar 100 lanzamientos y que la apuesta inicial sea de 1 peso. Pero Juan pide a Pedro que le permita, siempre que pierda, doblar su apuesta; es decir, si en determinado lanzamiento de la moneda Juan pierde 1 peso, en el siguiente lanzamiento apostará 2 pesos, y si vuelve a perder, su siguiente apuesta será de 4 pesos, y así sucesivamente. Pedro acepta, pero con la restricción de que Juan podrá doblar sus apuestas de manera consecutiva (en caso de perder) por solo cinco ocasiones. Además, si en el lanzamiento k-ésimo, con $k < 100$ cae águila, es decir, Juan gana, la apuesta en el lanzamiento $k + 1$ será de nuevo de 1 peso. ¿Quién considera usted que tiene ventaja? De inicio, podríamos empezar por indagar la probabilidad de que en 100 lanzamientos se obtenga al menos una secuencia de al menos cinco soles, y calcular también, el número esperado de lanzamientos de la moneda para obtener, por primera vez, cinco soles consecutivos.

Más generalmente, podríamos estar interesados en la probabilidad de obtener al menos una vez una secuencia de símbolos, o un texto específico, si en un teclado se pulsan teclas al azar un determinado número de veces. El conocido teorema del mono infinito, que afirma que un mono pulsando teclas al azar eventualmente escribirá cualquier texto con probabilidad uno (véase [3], [5]), no nos proporciona la probabilidad exacta de escribir, por ejemplo, la palabra BANANA al pulsar al azar un millón de teclas. Aunado a esto, como se menciona en [6], el cálculo de la probabilidad de aparición de un texto específico dentro de una sucesión finita de símbolos generados aleatoriamente, puede ser utilizado para calcular la probabilidad de obtener una secuencia de ADN específica. Lo anterior es debido a que una molécula de ácido desoxirribonucleico (ADN) es una cadena o secuencia de nucleótidos, cada nucleótido está formado por una de las cuatro bases nitrogenadas adenina (A), citosina (C), guanina (G) y timina (T), las cuales forman «palabras» genéticas de cuatro letras. La ocurrencia de una secuencia específica de nucleótidos en alguna porción de la cadena es el evento de que ocurra un texto específico con los símbolos A, C, G y T.

Debido a todo lo anterior, la probabilidad de aparición de un patrón o texto específico dentro de una sucesión finita de símbolos generados al azar, ha sido de interés desde hace varias décadas. En este trabajo se muestra, utilizando herramienta elemental de probabilidad, cómo obtener una fórmula recursiva para la probabilidad exacta de observar al menos una vez algunos tipos de texto al pulsar al azar teclas un determinado número de veces. Utilizando esta fórmula se calcula también

el número esperado de teclas que se requieren pulsar para obtener por primera vez el texto deseado.

Cabe mencionar que aunque en [6] se obtuvieron fórmulas recursivas en contextos más generales, nuestro aporte, en este trabajo, radica en la manera tan sencilla en que demostramos estas fórmulas en los casos particulares que consideramos. En lo que respecta al cálculo de los tiempos esperados de ocurrencia de textos o patrones, es importante destacar que en [6] no se calculan, mientras que en el presente trabajo obtenemos fórmulas para calcularlos; sin hacer uso de la teoría de cadenas de Markov o martingalas, como es el caso de por ejemplo [2], [5] y [4].

2. Conteo de patrones en la repetición de experimentos

Considere un conjunto formado por m símbolos distintos (podríamos pensar en letras de determinado alfabeto). Suponga un muestreo ordenado con reemplazo de tamaño n de la población de m símbolos; es decir, se seleccionan n símbolos al azar uno tras otro, pero después de cada elección, el símbolo elegido se reintegra al conjunto de manera que ese mismo símbolo puede ser elegido en una elección subsequente. Se supone aquí que todos los símbolos tienen la misma probabilidad, $\frac{1}{m}$ en este caso, de ser elegidos. Sea P cierta «palabra» o «texto» de interés y sea a el número de símbolos (no necesariamente distintos) que conforman a P. Pretendemos en esta sección obtener una fórmula para determinar el número de muestras ordenadas de tamaño n donde aparece la palabra P al menos una vez.

Denotemos por $r(n, m, P)$ al número de muestras ordenadas de tamaño n donde aparece la palabra P, por única vez, en las últimas a posiciones, y por $k(n, m, P)$ al número de muestras donde la palabra P aparece al menos una vez.

Notemos que

$$r(n, m, P) = 0 = k(n, m, P) \quad \text{para todo } 0 \leq n < a, \tag{1}$$

y

$$r(n, m, P) = 1 = k(n, m, P) \quad \text{para } n = a. \tag{2}$$

Teorema 2.1. *Se cumple la relación*

$$k(n, m, P) = r(n, m, P) + mk(n - 1, m, P), \quad n = 1, 2, \ldots.$$

Demostración. La relación es claramente válida para los enteros positivos $n \leq a$. Supongamos pues que $n > a$ y denotemos por B_i al conjunto

de muestras ordenadas en donde aparece la palabra P, por primera vez, de la posición i a la $a+i-1$, $i=1,\ldots,n-a+1$.

Claramente $B_i \bigcap B_j = \emptyset$, $i \neq j$ y

$$k(n,m,P) = \sum_{i=1}^{n-a+1} \operatorname{card} B_i = \sum_{i=1}^{n-a} \operatorname{card} B_i + \operatorname{card} B_{n-a+1},$$

donde $\operatorname{card} B_i$ es la cardinalidad (número de elementos) del conjunto B_i.

Notemos que $\sum_{i=1}^{n-a} \operatorname{card} B_i$ es el número de muestras ordenadas en donde aparece la palabra P al menos una vez antes de la última posición, y que $\operatorname{card} B_{n-a+1}$ es el número de muestras ordenadas en que aparece la palabra P, por única vez, en las últimas a posiciones. Esto es,

$$k(n,m,P) = r(n,m,P) + \sum_{i=1}^{n-a} \operatorname{card} B_i.$$

Si consideramos ahora muestras ordenadas de símbolos de tamaño $n-1$ y definimos B_i' igual que como se definió B_i, para $i=1,\ldots,n-a$, entonces claramente si $\{\mathfrak{b}_1,\ldots,\mathfrak{b}_{n-1}\}$ es una muestra ordenada en B_i', entonces $\{\mathfrak{b}_1,\ldots,\mathfrak{b}_{n-1},\mathfrak{b}\}$ es una muestra ordenada en B_i para todo símbolo $\mathfrak{b}$, $i=1,\ldots,n-a$. De esta manera,

$$\sum_{i=1}^{n-a} \operatorname{card} B_i = m \sum_{i=1}^{n-a} \operatorname{card} B_i' = mk(n-1,m,P).$$

Tenemos así la relación

$$k(n,m,P) = r(n,m,P) + mk(n-1,m,P). \qquad \square$$

Obsérvese que la idea detrás de la demostración del teorema anterior, es que el número de muestras ordenadas de tamaño n donde la palabra P aparece al menos una vez ($k(n,m,P)$) es igual al número de muestras ordenadas de tamaño n donde P aparece por única vez en las últimas a posiciones ($r(n,m,P)$)

													B	A	N	A	N	A

más el número de muestras ordenadas de tamaño n donde P aparece al menos una vez antes de la última posición;

									B	A	N	A	N	A			?

este último número es $mk(n-1,m,P)$, ya que por cada muestra ordenada de tamaño $n-1$ donde aparece al menos una vez P, si agregamos un símbolo al final, obtenemos una muestra ordenada de tamaño n donde P aparece al menos una vez antes de la última posición, y tenemos m formas de elegir la última posición.

En la siguiente proposición damos una fórmula recursiva para el caso en que P consiste de a repeticiones de un mismo símbolo $\mathfrak{b}$. En este caso denotaremos a $k(n, m, P)$ como $k(n, m, a)$.

Proposición 2.1. *Si P consiste de a repeticiones de un mismo símbolo, entonces se cumple la fórmula recursiva*

$$k(n,m,a) = (m-1)\left[m^{n-a-1} - k(n-a-1,m,a)\right] + mk(n-1,m,a),$$

$n = a+1, \ldots,$ *donde m es el número total de símbolos.*

Demostración. Notemos que $m^{n-a-1} - k(n-a-1, m, a)$ es la cardinalidad del conjunto de muestras ordenadas de tamaño $n-a-1$ que no contienen ninguna secuencia de tamaño a del símbolo $\mathfrak{b}$. Si a una de tales muestras le agregamos $a+1$ símbolos después de su última posición, y si la posición $n-a$ es ocupada por un símbolo distinto al $\mathfrak{b}$ y las últimas a posiciones por el símbolo $\mathfrak{b}$, la muestra ordenada resultante pertenecerá al conjunto B_{n-a+1} definido en la demostración del teorema 2.1, que consiste de las muestras ordenadas donde aparece P por primera vez de la posición $n-a+1$ a la n (las últimas a posiciones). Tenemos así, que en este caso, el número de muestras ordenadas de tamaño n donde aparece P por única vez en las últimas a posiciones está dado por

$$r(n,m,a) = (m-1)\left[m^{n-a-1} - k(n-a-1,m,a)\right]$$

y así, el resultado deseado se sigue de la relación recursiva dada en el teorema 2.1. □

Daremos ahora, en la siguiente proposición, una fórmula recursiva para el caso en que P es una secuencia de símbolos donde el primer o último símbolo difiere del resto.

Proposición 2.2. *Si P es una secuencia de a símbolos, donde el primer o último símbolo difiere de los restantes, entonces se cumple la fórmula recursiva*

$$k(n,m,P) = m^{n-a} - k(n-a,m,P) + mk(n-1,m,P), \quad n = a, a+1, \ldots,$$

donde m es el número total de símbolos.

Demostración. Supongamos que el primer símbolo de P difiere del resto. Entonces si la palabra P aparece en las últimas a posiciones, es decir, de la posición $n-a+1$ a la n, entonces independientemente de los símbolos que ocupen las posiciones entre $n-2a+2$ y $n-a$, no se formará la palabra P entre las posiciones $n-2a+2$ y $n-1$. De esto es claro que $m^{n-a} - k(n-a, m, P)$ es, entonces, la cardinalidad del conjunto de muestras ordenadas en las que la palabra P aparece por única vez en las últimas a posiciones, es decir, es $r(n, m, P)$. Por simetría, es claro que esto mismo ocurre cuando el último símbolo de

la palabra P difiere de los restantes. Por lo tanto, el resultado deseado se sigue de nueva cuenta de la relación recursiva dada en el teorema 2.1. □

Note que el teorema 2.1 reduce el problema de encontrar una fórmula para el número de muestras ordenadas de tamaño n donde aparece al menos una vez la palabra P, a encontrar una fórmula para el número de muestras ordenadas de tamaño n donde aparece P por única vez en las últimas a posiciones, $r(n, m, P)$. Es evidente que si se tiene una palabra P para la cual se puede obtener una fórmula para $r(n, m, P)$, la cual puede estar en términos de $k(s, m, P)$ con $s < n$, entonces utilizando el teorema 2.1 se obtiene una fórmula recursiva para calcular $k(n, m, P)$, como sucede con las palabras de las proposiciones 2.1 y 2.2.

3. Probabilidades y tiempos medios para la ocurrencia de algunos patrones

Si P consiste de a repeticiones de un mismo símbolo, y todos los símbolos tienen la misma probabilidad, $\frac{1}{m}$, de ser elegidos, se tiene que la probabilidad de obtener P al menos una vez en un muestreo ordenado con reemplazo de tamaño n es

$$p(n, m, a) \equiv \frac{k(n, m, a)}{m^n}, \tag{3}$$

donde $k(n, m, a)$ puede calcularse mediante la fórmula recursiva de la proposición 2.1; por lo que una fórmula recursiva (fácilmente programable) para $p(n, m, a)$ es la siguiente:

$$p(n, m, a) = \frac{m-1}{m^{a+1}}[1-p(n-a-1, m, a)]+p(n-1, m, a), \quad n = a+1, \ldots. \tag{4}$$

Nótese que por (1) y (2)

$$p(n, m, a) = 0 \quad \text{para todo } 0 \leq n < a, \tag{5}$$

y

$$p(n, m, a) = \frac{1}{m^a} \quad \text{para } n = a. \tag{6}$$

Al número mínimo, T, de elecciones (tamaño de la muestra) necesarias para obtener por vez primera al patrón P, le llamaremos el **tiempo de ocurrencia de** P. En la siguiente proposición se da una fórmula para la esperanza de T.

Proposición 3.1. *Si P consiste de a repeticiones de un mismo símbolo y T es el tiempo de ocurrencia de P, entonces el tiempo medio de*

ocurrencia de P es

$$E(T) = \frac{m(m^a - 1)}{m - 1},$$

donde m es el número total de símbolos.

Demostración. Como T es una variable aleatoria con valores en los enteros positivos, se conoce (véase [3, p. 76, teo. 12.2]) que

$$E(T) = \sum_{n=0}^{\infty} P(T > n).$$

Notemos que

$$P(T > n) = 1 - P(T \leq n) = 1 - p(n, m, a).$$

Luego, por (5) y (6),

$$\begin{aligned} E(T) &= \sum_{n=0}^{\infty} [1 - p(n, m, a)] \\ &= a + 1 - \frac{1}{m^a} + \sum_{n=a+1}^{\infty} [1 - p(n, m, a)]. \end{aligned}$$

Utilizando la fórmula recursiva (4) tenemos que

$$\begin{aligned} E(T) &= a + 1 - \frac{1}{m^a} \\ &+ \sum_{n=a+1}^{\infty} \left[1 - \left\{ \frac{m-1}{m^{a+1}} [1 - p(n - a - 1, m, a)] + p(n - 1, m, a) \right\} \right] \\ &= a + 1 - \frac{1}{m^a} - \frac{m-1}{m^{a+1}} \sum_{n=a+1}^{\infty} [1 - p(n - a - 1, m, a)] \\ &\quad + \sum_{n=a+1}^{\infty} [1 - p(n - 1, m, a)] \\ &= a + 1 - \frac{1}{m^a} - \frac{m-1}{m^{a+1}} \sum_{n=a+1}^{\infty} P(T > n - a - 1) \\ &\quad + \sum_{n=a+1}^{\infty} P(T > n - 1) \\ &= a + 1 - \frac{1}{m^a} - \frac{m-1}{m^{a+1}} \sum_{n=0}^{\infty} P(T > n) + \sum_{n=0}^{\infty} P(T > n) - a \\ &= \frac{m^a - 1}{m^a} - \frac{m-1}{m^{a+1}} E(T) + E(T). \end{aligned}$$

De aquí, se obtiene que

$$E(T) = \frac{\frac{m^a-1}{m^a}}{\frac{m-1}{m^{a+1}}} = \frac{m(m^a-1)}{m-1}. \qquad \square$$

Ahora, si P es una palabra conformada por a símbolos, y su primer o último símbolo difiere de los restantes; y de nuevo todos los símbolos tienen la misma probabilidad, $\frac{1}{m}$, de ser elegidos, se tiene que la probabilidad de obtener P al menos una vez en un muestreo ordenado con reemplazo de tamaño n es

$$p(n, m, P) \equiv \frac{k(n, m, P)}{m^n}, \tag{7}$$

donde $k(n, m, P)$ puede calcularse mediante la fórmula recursiva de la proposición 2.2. Por lo que se tiene la siguiente fórmula recursiva

$$p(n, m, P) = \frac{1}{m^a}[1 - p(n-a, m, P)] + p(n-1, m, P), \qquad n = a, a+1, \ldots. \tag{8}$$

Por (1) y (2) se tiene que

$$p(n, m, P) = 0 \quad \text{para todo } 0 \leq n < a, \tag{9}$$

y

$$p(n, m, P) = \frac{1}{m^a} \quad \text{para } n = a. \tag{10}$$

En la siguiente proposición se da una fórmula para la esperanza del tiempo de ocurrencia de P.

Proposición 3.2. *Si P es una palabra conformada por a símbolos, donde el primer o último símbolo difiere de los restantes y T es el tiempo de ocurrencia de P, entonces el tiempo medio de ocurrencia de P es*

$$E(T) = m^a,$$

donde m es el número total de símbolos.

Demostración. De (9), (10) y la fórmula recursiva (8), se obtiene que (véase el inicio de la demostración de la proposición 3.1)

$$\begin{aligned} E(T) &= \sum_{n=0}^{\infty}[1 - p(n, m, P)] = a + 1 - \frac{1}{m^a} + \sum_{n=a+1}^{\infty}[1 - p(n, m, P)] \\ &= a + 1 - \frac{1}{m^a} \\ &\quad + \sum_{n=a+1}^{\infty}\left[1 - \left\{\frac{1}{m^a}[1 - p(n-a, m, P)] + p(n-1, m, P)\right\}\right] \end{aligned}$$

$$
\begin{aligned}
&= a+1-\frac{1}{m^a}-\frac{1}{m^a}\sum_{n=a+1}^{\infty}[1-p(n-a,m,P)] \\
&\quad +\sum_{n=a+1}^{\infty}[1-p(n-1,m,P)] \\
&= a+1-\frac{1}{m^a}-\frac{1}{m^a}\sum_{n=a+1}^{\infty}P(T>n-a)+\sum_{n=a+1}^{\infty}P(T>n-1) \\
&= a+1-\frac{1}{m^a}-\frac{1}{m^a}\left(\sum_{n=0}^{\infty}P(T>n)-1\right)+\sum_{n=0}^{\infty}P(T>n)-a \\
&= 1-\frac{E(T)}{m^a}+E(T).
\end{aligned}
$$

De lo cual,

$$\frac{E(T)}{m^a}=1,$$

o equivalentemente

$$E(T)=m^a. \qquad \square$$

4. Ejemplos

En nuestro primer ejemplo retomamos el cuestionamiento, hecho en la introducción, sobre cual de los dos apostadores tiene la ventaja.

Ejemplos 4.1. *Se lanza una moneda justa con caras águila y sol (denotadas por a y s, respectivamente) y se considera $P = sssss$. Entonces, la probabilidad de que en una muestra ordenada de tamaño 100 aparezca al menos una vez P es $p(100, 2, 5)$; y tenemos, por (4), la fórmula recursiva*

$$p(n,2,5)=\frac{1}{2^6}[1-p(n-6,2,5)]+p(n-1,2,5), \quad n=6,7,\ldots.$$

Empleando esta fórmula recursiva, puede verse que la probabilidad de obtener al menos una secuencia de al menos 5 soles en 100 lanzamientos es

$$p(100,2,5)\approx 0.81;$$

y por la proposición 3.1, se tiene que el número promedio de lanzamientos requeridos para obtener P por primera vez es

$$\frac{2(2^5-1)}{2-1}=62.$$

Con base en esta información, ¿considera adecuada la estrategia de apuesta de Juan? Note que si aparece una secuencia de cinco soles, Juan perderá en esos cinco lanzamientos

$$1 + 2 + 2^2 + 2^3 + 2^4 = 31 \quad \text{pesos},$$

pérdida difícil de recuperar, considerando que después de esta mala racha de Juan, él deberá reiniciar su estrategia, y podrá entonces, cada vez que gane, recuperar solamente lo que lleve perdido después de su reinicio (siempre y cuando no haya ocurrido otra secuencia de cinco soles) más un peso.

En el caso de al menos una racha de siete soles consecutivos en los 100 *lanzamientos de la moneda, es decir,* $P = sssssss$*, tenemos que la probabilidad es*

$$p(100, 2, 7) \approx 0.317;$$

y el número promedio de lanzamientos requeridos para obtener esta racha es

$$\frac{2(2^7 - 1)}{2 - 1} = 254.$$

Debido a que la probabilidad de obtener, en siete lanzamientos de una moneda justa, siete soles, es de apenas $\frac{1}{2^7} = 0.0078125$*, el hecho de que se requieren en promedio de apenas* 254 *lanzamientos para obtener una racha de este tipo, es muy poco intuitivo.*

Un ejemplo análogo al anterior, surge al considerar el lanzamiento de un dado balanceado.

Ejemplos 4.2. *Se realizan 1000 lanzamientos independientes de un dado justo con caras enumeradas del 1 al 6 y se considera* $P = 2222$*. Entonces, la probabilidad de que aparezca al menos una vez* $P = 2222$ *es* $p(1000, 6, 4)$*; y tenemos por (4), la fórmula recursiva*

$$p(n, 6, 4) = \frac{5}{6^5}\left[1 - p(n - 5, 6, 4)\right] + p(n - 1, 6, 4), \quad n = 5, 6, \ldots.$$

Empleando esta fórmula recursiva, puede verse que la probabilidad de obtener al menos una secuencia de al menos cuatro veces el 2 en 1000 lanzamientos es

$$p(1000, 6, 4) = 0.4743;$$

y por la proposición 3.1, se tiene que el número promedio de lanzamientos requeridos para obtener P *por primera vez es*

$$\frac{6(6^4 - 1)}{6 - 1} = 1554.$$

Por último, consideramos el ejemplo mencionado en la introducción sobre escribir al menos una vez un texto específico al pulsar teclas al azar un determinado número de veces.

Ejemplos 4.3. *Si se tiene un teclado de 26 teclas, entonces mecanografiando al azar, la probabilidad de que una sucesión de letras de tamaño 1000000 contenga al menos una vez la palabra P=BANANA es* $p(1000000, 26, P)$*, la cual, por (8), se obtiene recursivamente mediante la fórmula*

$$p(n, 26, P) = \frac{1}{26^6}[1 - p(n-6, 26, P)] + p(n-1, 26, P), \quad n = 6, 7, \ldots.$$

Empleando esta fórmula recursiva, se obtiene que la probabilidad de obtener al menos una vez la palabra BANANA en 1000000 tecleos al azar es

$$p(1000000, 26, P) = 0.0032;$$

y por la proposición 3.2, se tiene que el número promedio de tecleos requeridos para obtener BANANA por primera vez es

$$26^6 = 308915776.$$

5. Comentarios

En el presente artículo se mostró cómo obtener fórmulas recursivas para las probabilidades de ocurrencia de algunos tipos sencillos de patrones, así como fórmulas para sus tiempos medios de ocurrencia, utilizando herramienta elemental de probabilidad. Se calcularon estas probabilidades y tiempos esperados para algunos ejemplos de interés.

Fórmulas recursivas para el cálculo de probabilidades de ocurrencia de patrones en contextos más generales que los que aquí se consideran, pueden ser consultados en [6]. En [4] se utilizan algunos resultados de la teoría de martingalas para obtener los tiempos medios de ocurrencia de patrones; en [2] y [5] los calculan usando cadenas de Markov.

Bibliografía

[1] Z. Brzeźniak y T. Zastawniak, *Basic stochastic processes: a course through exercises*, Springer, Great Britain, 1999.
[2] C. M. Grinstead y J. L. Snell, *Introduction to probability*, 2.ª ed., American Mathematical Society, USA, 1997.
[3] A. Gut, *Probability: a graduate course*, Springer, USA, 2005.
[4] S.-Y. R. Li, «A martingale approach to the study of occurrence of sequence patterns in repeated experiments», *Annals of Probability*, vol. 8, 1980, 1171–1176.
[5] L. Rincón, «Sobre el problema del mono que escribe caracteres al azar», *Miscelánea Matemática*, núm. 42, 2006, 79–90.
[6] S. J. Schwager, «Run probabilities in sequences of Markov-dependent trials», *J. Amer. Statist. Assoc.*, vol. 78, 1983, 168–175.