Miscelánea Matemática **42** (2006) 79–90 SMM

Sobre el problema del mono que escribe caracteres al azar

Luis Rincón
Departamento de Matemáticas
Facultad de Ciencias, UNAM
Circuito Exterior
Ciudad Universitaria
04510 México D.F.
lars@fciencias.unam.mx

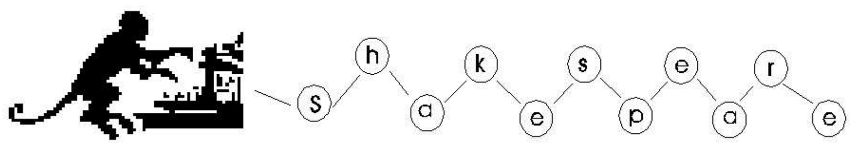
Resumen

Se presentan varias formas de resolver el problema de encontrar la probabilidad de que un mono que escribe caracteres al azar, reproduzca por casualidad y sin error alguno, las obras completas de Shakespeare. Se estima además el tiempo promedio de espera para que tan raro evento ocurra.

1. Planteamiento del problema

El problema es simple. Un mono escribe caracteres completamente al azar en una máquina de escribir. ¿Cuál es la probabilidad de que eventualmente el mono escriba exactamente, y sin ningún error, las obras completas de Shakespeare? La respuesta es uno, y pretendemos convencer al lector de este sorprendente resultado.

Este es un problema teórico contextualizado en una situación real. Se asume idealmente que se tiene el escenario de un mono eterno que escribe permanentemente caracteres al azar en una máquina de escribir que nunca se descompone, y cuya tira de papel nunca termina. Las soluciones que se exponen en esta pequeña nota no aparecen en la mayoría de los textos de probabilidad, pero ciertamente la respuesta a la pregunta planteada no es nueva. El problema aparece resuelto por ejemplo en el excelente libro de probabilidad de Ghahramani [1], como una aplicación de la ley fuerte de los grandes números. Nuestro objetivo es exponer con mayor detalle esta solución, proponer otros métodos de solución, efectuar algunos cálculos numéricos y hacer comentarios respecto a este problema de insospechadas consecuencias.



Mono escribiendo al azar.

2. Solución

Imaginemos entonces que un mono escribe caracteres al azar en una máquina de escribir, y que lo hace de manera continua generando una sucesión lineal de caracteres. Cada uno de los caracteres tiene la misma probabilidad de aparecer y se genera uno independientemente de otro. Sea m el total de caracteres disponibles y sea N el total de caracteres de los que constan las obras completas de Shakespeare. Por el momento no necesitamos especificar los valores exactos de m y N. Segmentamos el arreglo lineal de caracteres generados por el mono en bloques de N caracteres, uno después de otro, y observamos si algún bloque contiene las obras de Shakespeare. Por ejemplo,

$$\underbrace{Xku \cdots aTs}_{N} \; \underbrace{hwW \cdots pzq}_{N} \; Ot \cdots$$

Para cada número natural k defina el evento A_k correspondiente a que el k-ésimo bloque contiene exactamente y sin error alguno las obras completas de Shakespeare. Observe que los eventos A_k son independientes pues los bloques no se sobreponen y además $P(A_k) = (1/m)^N$. Denotaremos por p a tal probabilidad. A continuación se exponen varias formas de encontrar la respuesta a la pregunta inicialmente planteada.

2.1. Solución usando la ley fuerte de los grandes números

La ley de los grandes números es uno de los pilares de la teoría de la probabilidad y una primera versión fue demostrada por primera vez en 1713 por el matemático suizo James Bernoulli en la cuarta parte de su obra *Ars Conjectandi*. La versión fuerte que enunciamos a continuación es adecuada para nuestros propósitos. Sea $X_1, X_2, \ldots$ una sucesión infinita de variables aleatorias independientes todas con la misma distribución con esperanza μ y con cierta varianza finita. Entonces

$$P(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \mu) = 1.$$

Esto quiere decir que la variable aleatoria promedio $\frac{1}{n}\sum_{k=1}^{n} X_k$ converge casi seguramente a la constante μ.

Para la solución del problema planteado podemos definir la variable aleatoria X_k como la función indicadora del evento A_k, es decir,

$$X_k = \begin{cases} 1 & \text{si } A_k \text{ ocurre,} \\ 0 & \text{si } A_k \text{ no ocurre.} \end{cases}$$

Tenemos entonces una sucesión de variables aleatorias $X_1, X_2, \ldots$ independientes con idéntica distribución Bernoulli de parámetro $p = P(A_k) > 0$. En particular la media de cada una de ellas es $E(X_k) = p$. Considere la suma $X_1 + X_2 + \cdots + X_n$. Si para algún valor de n esta suma es positiva significa que alguno de los sumandos es distinto de cero y por lo tanto que el mono ha tenido éxito. Pero esto es justamente lo que garantiza la ley fuerte de los grandes números, pues

$$P(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k \ = \ p) = 1. \tag{1}$$

Es decir, con probabilidad uno la suma de la ecuación (1) es positiva. Esto implica que debe existir al menos un valor de k tal que $X_k > 0$ y por lo tanto $X_k = 1$. Esto significa que en el k-ésimo bloque ¡el mono ha tenido éxito!. Más aún, para que el promedio que aparece en la ecuación (1) sea positivo necesariamente la suma debe ser infinita, y por lo tanto, deben existir una infinidad de valores de k tal que $X_k = 1$. Esto quiere decir que con probabilidad uno ¡el mono escribirá tantas veces como uno desee las obras completas de Shakespeare!.

Observe que el análisis llevado a cabo sólo considera bloques disjuntos y consecutivos de caracteres y que existe la posibilidad de que el mono escriba el texto en un segmento intermedio a los bloques considerados. La probabilidad de un eventual éxito en esas circunstancias es mayor o igual a la calculada, siendo esta última uno, la primera también.

2.2. Solución usando continuidad de la probabilidad

Recordemos brevemente el concepto de continuidad para las medidas de probabilidad. Dada una sucesión infinita de eventos $B_1, B_2, \ldots$ se define el límite superior e inferior como sigue

$$\begin{aligned} \limsup_{n\to\infty} B_n &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} B_k, \\ \text{y} \quad \liminf_{n\to\infty} B_n &= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} B_k. \end{aligned}$$

Cuando sucede que estos dos eventos límite coinciden en el mismo evento B entonces se dice que la sucesión es convergente y se escribe $B = \lim_{n\to\infty} B_n$. En esta circunstancia toda medida de probabilidad es continua en el siguiente sentido

$$P(B) = \lim_{n\to\infty} P(B_n). \tag{2}$$

Esta es una versión del conocido teorema de continuidad de la probabilidad. En particular, no es difícil demostrar que si la sucesión es decreciente, es decir $B_1 \supseteq B_2 \supseteq \cdots$, entonces $\lim_{n\to\infty} B_n = \cap_{n=1}^{\infty} B_n$ y se cumple (2). Usaremos esta propiedad para encontrar nuevamente la respuesta al problema del mono. Considere los eventos A_k definidos como antes. La observación importante es que el evento $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ corresponde a aquel en el cual se verifican una infinidad de A_k. Deseamos entonces demostrar que la probabilidad de tal evento es uno. Observe con cuidado los dos momentos en los que se hace uso de la continuidad de la probabilidad para eventos decrecientes en el siguiente cálculo.

$$\begin{aligned} P(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k) &= \lim_{n\to\infty} P(\bigcup_{k=n}^{\infty} A_k) \\ &= 1 - \lim_{n\to\infty} P(\bigcap_{k=n}^{\infty} A_k^c) \\ &= 1 - \lim_{n\to\infty} \lim_{\eta\to\infty} P(\bigcap_{k=n}^{\eta} A_k^c) \\ &= 1 - \lim_{n\to\infty} \lim_{\eta\to\infty} (1-p)^{\eta-n+1} \\ &= 1. \end{aligned}$$

Es decir, con probabilidad uno se cumple el evento A_k para una infinidad de valores naturales de k, y nuevamente se comprueba que con probabilidad uno ¡el mono tiene una infinidad de éxitos!.

2.3. Solución usando Borel-Cantelli

Veamos ahora una solución más corta. El bien conocido lema de Borel-Cantelli establece lo siguiente: Sea $A_1, A_2, \ldots$ una sucesión de eventos y defina $A = \limsup_{k\to\infty} A_k$. Entonces se tienen los siguientes dos resultados.

a) Si $\sum_{k=1}^{\infty} P(A_k) < \infty$ entonces $P(A) = 0$.

b) Si los eventos son independientes y $\sum_{k=1}^{\infty} P(A_k) = \infty$ entonces $P(A) = 1$.

En nuestro caso los eventos $A_1, A_2, \ldots$ definidos antes son efectivamente independientes y son tales que $P(A_k) = (1/m)^N$. Entonces $\sum_{k=1}^{\infty} P(A_k) = \infty$ y por la segunda parte del lema de Borel-Cantelli,

$$P(\limsup_{k\to\infty} A_k) = 1.$$

Es decir, nuevamente se concluye que con probabilidad uno el mono obtiene una infinidad de éxitos.

2.4. Solución usando cadenas de Markov

Ahora usaremos una cadena de Markov $\{X_n : n = 0, 1, \ldots\}$ para modelar el experimento de generar caracteres al azar con la intención de obtener un escrito específico de N caracteres de longitud. La variable aleatoria X_n representará el número de caracteres correctos obtenidos inmediatamente antes e incluyendo el último momento observado n. Es claro que cada una de estas variables toma valores en el conjunto $\{0, 1, 2, \ldots, N\}$ y dado que los caracteres se generan de manera independiente, el valor de X_{n+1} depende únicamente del valor de X_n y no de los anteriores, es decir, se trata efectivamente de una cadena de Markov. Considerando como antes un conjunto de símbolos de m caracteres se tiene que

$$\begin{aligned} P(X_{n+1} = x + 1 \mid X_n = x) &= \frac{1}{m}, \\ \text{y} \quad P(X_{n+1} = 0 \mid X_n = x) &= \frac{m-1}{m}. \end{aligned}$$

El primer caso corresponde a obtener el caracter correcto al siguiente tiempo $n+1$, y la segunda igualdad refleja la situación lamentable de cometer un error en el siguiente caracter generado cuando ya se habían obtenido x caracteres correctos. Las posibles transiciones de un estado a otro se muestran gráficamente en el siguiente diagrama.

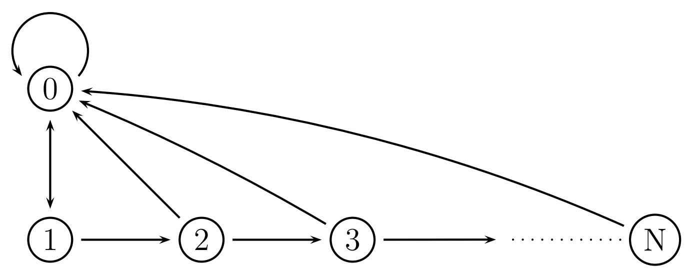


Diagrama para el modelo de Markov.

La matriz de probabilidades de transición en un paso es entonces

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \\ N-1 \\ N \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & & N-1 & N \end{array} \\ \left(\begin{array}{ccccccc} (m-1)/m & 1/m & 0 & 0 & \cdots & 0 & 0 \\ (m-1)/m & 0 & 1/m & 0 & \cdots & 0 & 0 \\ (m-1)/m & 0 & 0 & 1/m & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ (m-1)/m & 0 & 0 & 0 & \cdots & 0 & 1/m \\ (m-1)/m & 1/m & 0 & 0 & \cdots & 0 & 0 \end{array} \right) \end{array}.$$

Se trata de una matriz finita e irreducible pues todos los estados se comunican. Entonces con probabilidad uno la cadena visita cada uno de sus estados una infinidad de veces. En particular, cada vez que la cadena visita el estado N el mono concluye una sucesión exitosa de caracteres, y ello sucederá ¡una infinidad de veces con probabilidad uno!.

3. Tiempos de espera

Una vez que uno se ha repuesto de la incredulidad del resultado, es natural hacerse la siguiente pregunta. ¿Cuánto tiempo le tomará al mono obtener las obras de Shakespeare? Lamentablemente sólo podemos responder a esta pregunta en términos de promedios. Tenemos dos formas de calcular este tiempo medio.

3.1. Tiempos de espera : estimación gruesa

Cada bloque de N caracteres puede considerarse como un ensayo Bernoulli en donde la probabilidad de obtener éxito es p y la probabilidad de obtener fracaso es el número complementario $1-p$. Si T es la variable aleatoria que cuenta el número de ensayos (bloques) que transcurren hasta obtener el primer éxito, entonces T tiene una distribución geométrica con parámetro p, es decir, para $k \geq 1$,

$$P(T = k) = (1-p)^{k-1} p.$$

Es sencillo comprobar que la esperanza de tal variable aleatoria es $1/p$, es decir,

$$E(T) = \frac{1}{p} = m^N.$$

Por lo tanto en promedio se necesitan m^N bloques de N caracteres para obtener las obras de Shakespeare. En términos de caracteres, se necesitan en promedio

$$N m^N \tag{3}$$

caracteres al azar para obtener el texto referido. Observe que esta estimación considera únicamente observaciones de bloques, es decir, secciones sucesivas y ajenas de longitud N, y esto podría parecer inadecuado. Sin embargo la facilidad del cálculo es notable y como veremos más adelante (vea la ecuación (6)), esta estimación está excedida por el factor N. A continuación se presenta un argumento que nos permite hacer una estimación más fidedigna.

3.2. Tiempos de espera : estimación fina

Nuevamente usaremos el modelo de cadena de Markov para calcular el tiempo promedio de espera para obtener un éxito. Esta vez consideraremos la matriz de probabilidades de transición

$$
\begin{array}{c}
\\ 0 \\ 1 \\ 2 \\ \vdots \\ N-1 \\ N
\end{array}
\begin{array}{c}
\begin{array}{ccccccc} 0 & 1 & 2 & 3 & & N-1 & N \end{array} \\
\left(\begin{array}{ccccccc}
(m-1)/m & 1/m & 0 & 0 & \cdots & 0 & 0 \\
(m-1)/m & 0 & 1/m & 0 & \cdots & 0 & 0 \\
(m-1)/m & 0 & 0 & 1/m & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
(m-1)/m & 0 & 0 & 0 & \cdots & 0 & 1/m \\
0 & 0 & 0 & 0 & \cdots & 0 & 1
\end{array} \right)
\end{array}.
$$

Esta matriz y la anterior difieren únicamente en el último renglón. En este caso se considera que el estado N es absorbente, es decir, una vez visitado este estado, la cadena permanece en él para siempre. Este artificio en la dinámica de la cadena es la clave para encontrar el tiempo medio de absorción partiendo de $X_0 = 0$. Sea T el primer momento de la absorción, es decir,

$$T = \text{mín}\{n \geq 0 : X_n = N\}.$$

Cuando el conjunto $\{n \geq 0 : X_n = N\}$ es vacío se puede definir $T = \infty$. Para cada $k = 0, 1, \ldots, N-1$ se define

$$v_k = E(T|X_0 = k).$$

Claramente $v_N = 0$. Nuestro objetivo es encontrar v_0. Usando análisis del primer paso es fácil darse cuenta que v_k satisface la ecuación

$$v_k = 1 + \frac{1}{m}\, v_{k+1} + \frac{m-1}{m}\, v_0,$$

para $k = 0, 1, \ldots, N-1$. El argumento para encontrar esta ecuación es el siguiente. Al efectuar el paso siguiente, el número promedio de pasos para la absorción v_k se incrementa en uno y pueden suceder dos casos: se puede avanzar al estado $k+1$ con probabilidad $1/m$ y reiniciar el conteo a partir de allí, o bien se puede generar un caracter equivocado y reiniciar el conteo desde cero, lo cual sucede con probabilidad $(m-1)/m$. De esta ecuación se obtiene

$$v_0 = \frac{m}{m-1}\,(v_k - \frac{1}{m}\, v_{k+1} - 1). \tag{4}$$

En particular para $k = N-1$,

$$v_{N-1} = 1 + \frac{m-1}{m}\, v_0. \tag{5}$$

Observe que la parte entre paréntesis de (4) debe ser la misma para cualquier valor de $k = 0, 1, \ldots, N-1$. El término para $k = N-1$ es simplemente $v_{N-1} - 1$. Por lo tanto

$$\begin{aligned}
v_0 - \frac{1}{m} v_1 &= v_{N-1} \\
v_1 - \frac{1}{m} v_2 &= v_{N-1} \\
&\vdots \\
v_{N-3} - \frac{1}{m} v_{N-2} &= v_{N-1} \\
v_{N-2} - \frac{1}{m} v_{N-1} &= v_{N-1}.
\end{aligned}$$

Analizando estas ecuaciones en orden inverso se encuentra que

$$\begin{aligned}
v_{N-2} &= v_{N-1}\,(1 + \frac{1}{m}) \\
v_{N-3} &= v_{N-1}\,(1 + \frac{1}{m} + \frac{1}{m^2}) \\
v_{N-4} &= v_{N-1}\,(1 + \frac{1}{m} + \frac{1}{m^2} + \frac{1}{m^3}) \\
&\vdots \\
v_0 &= v_{N-1}\,(1 + \frac{1}{m} + \frac{1}{m^2} + \frac{1}{m^3} + \cdots + \frac{1}{m^{N-1}}).
\end{aligned}$$

Nuestro interés radica en la última ecuación. Usando ésta y (5) se tiene que

$$\begin{aligned}
v_0 &= v_{N-1} \cdot \frac{1 - (\frac{1}{m})^N}{1 - \frac{1}{m}} \\
&= (1 + \frac{m-1}{m} v_0) \cdot \frac{[1 - (\frac{1}{m})^N]}{1 - \frac{1}{m}} \\
&= \frac{m}{m-1}[1 - (\frac{1}{m})^N] + v_0[1 - (\frac{1}{m})^N].
\end{aligned}$$

De donde finalmente se obtiene

$$v_0 = \frac{m}{m-1}(m^N - 1) \approx m^N. \tag{6}$$

Compare esta estimación fina con la estimación gruesa (3). Posiblemente el lector se haya dado cuenta que el modelo usado no refleja perfectamente la situación real. Por ejemplo, cuando el mono ha obtenido un cierto número de caracteres consecutivos correctos y después comete un error, la cadena de Markov toma el valor cero cuando probablemente el caracter equivocado sea nuevamente el primero de la sucesión correcta y por lo tanto el conteo debería regresar a uno. Este tipo de situaciones puede resolverse por ejemplo dando ventaja al mono de los primeros caracteres correctos y solicitándole que escriba los restantes. Dado el gran número de caracteres contenidos en las obras completas de Shakespeare, omitir algunos de ellos no reduce significativamente el tiempo promedio de espera.

4. Cálculos numéricos

Vamos a hacer ahora algunos cálculos numéricos. Suponga que el total de caracteres distintos incluyendo letras mayúsculas y minúsculas, punto, coma, espacio y otros símbolos es $m = 80$. Suponga también que el total de caracteres N en las obras completas de Shakespeare es de seis millones[1]. Estas cantidades revelan que la probabilidad de que el mono tenga éxito en un bloque cualquiera es

$$p = \left(\frac{1}{80}\right)^{6{,}000{,}000}.$$

Al ritmo pausado de imprimir un caracter por segundo, el tiempo promedio de espera, medido en segundos, para obtener éxito es, por (6), aproximadamente igual a

$$80^{6{,}000{,}000}.$$

Como cada año de 365 días tiene $31,536,000$ segundos, el tiempo de espera en años para obtener éxito es

$$\frac{1}{31,536,000} \cdot 80^{6{,}000{,}000} >> 10^{6{,}000{,}000}. \tag{7}$$

La cota inferior sugerida es bastante conservadora pero proporciona una idea del número de años promedio de espera para observar éxito en este experimento. Unos segundos de reflexión revelan que este número de años es desmesuradamente grande. Es el dígito 1 seguido de seis millones de ceros. Compárese por ejemplo con la edad del universo

[1] Una versión de las obras completas contiene 1230 páginas, en donde cada página contiene 65 líneas con un promedio de 80 caracteres por línea. Estos datos arrojan un total de $1230 \times 65 \times 80 = 6,396,000$ caracteres.

que es aproximadamente de 15 billones[2] de años, es decir, $15 \times 10^9 = 15,000,000,000$ años, mientras que la edad de la tierra y del resto del sistema solar es de únicamente 4.5 billones de años, es decir, $4{,}5 \times 10^9 = 4,500,000,000$ años. Estas cantidades son insignificantes cuando se les compara con (7). Así pues, los tiempos de espera para observar un éxito en este experimento parecen ser demasiado largos.

Cualquier incremento razonable en la velocidad de escritura o del número de monos que pudieran ponerse a trabajar a un mismo tiempo, no parece ser suficiente para reducir significativamente el tiempo de espera promedio. No obstante estas consideraciones, pueden encontrarse en internet algunos simuladores[3] por computadora de este experimento. En el fondo creo que existe siempre en las personas una ligera esperanza de que el azar proporcione una sorpresa agradable uno de estos días.

5. Comentarios

El esquema bajo el cual se ha planteado el problema del mono puede fácilmente pensarse en forma general. Suponga que se tiene una sucesión de ensayos independientes en cada uno de los cuales puede observarse un evento A con la misma probabilidad $p > 0$. Nuevamente se define la variable aleatoria X_k que indica si en el k-ésimo ensayo se observa el evento A. La ley fuerte de los grandes números, el lema de Borel-Cantelli, la propiedad de continuidad de las medidas de probabilidad o la teoría de las cadenas de Markov garantizan que con probabilidad uno el evento A ocurrirá eventualmente no sólo una vez sino una infinidad de veces. Lo importante aquí es la condición $p > 0$, y que el experimento aleatorio puede efectuarse bajo las mismas condiciones iniciales tantas veces como sea necesario.

El aspecto dramático en el problema del mono es que la probabilidad p es muy pequeña, pero ciertamente positiva, y el contexto ayuda a pensar que es imposible imaginar que verdaderamente un mono pueda escribir las obras de Shakespeare. Sin embargo hemos probado que con probabilidad uno esto ocurre.

[2]Estos datos siguen el sistema norteamericano de numeración en el cual un billón son mil millones. Esto permite comprobar fácilmente tales cifras con las recientes estimaciones que aparecen en la literatura científica en el tema, predominantemente en inglés.

[3]Escriba por ejemplo las palabras “typing monkey” en algún buscador en internet.

Históricamente este resultado ha sido contextualizado en la obtención de diversos escritos. Thomas Huxley en 1860 sugirió que de forma aleatoria podía obtenerse el Salmo 23, mientras que su sobrino Julian Huxley, tal vez más sensato, repitió la analogía para un texto menos controversial, las obras completas de Shakespeare. Stephen Hawking también señala en su famoso libro *A brief history of time* que dada una manada de monos escribiendo al azar "very ocassionally by pure chance they will type out one of Shakespeare's sonnets". El resultado también ha sido aplicado al caso de obtener una secuencia de ADN al azar, lo cual ha generado muchas controversias. Al respecto prefiero omitir comentario alguno y dejar al lector que reflexione sobre el significado y las posibles consecuencias que esto puede tener en nuestra concepción biológica y religiosa de la vida.

A pesar de los grandes tiempos de espera involucrados en esta interesante aplicación de la probabilidad e imaginando la remota ocurrencia de un evento muy favorable en la vida de cada uno de nosotros, no puedo evitar recordar el conocido proverbio "El que persevera alcanza".

Agradecimientos. Agradezco sinceramente a Gerardo Chávez, a Ana Meda y al revisor anónimo, quienes hicieron una lectura cuidadosa del presente artículo y me permitieron mejorarlo substancialmente.

También me es muy grato mencionar que este trabajo ha sido posible gracias al interés profesional que generosamente Ana Irene Ramírez depositó en él. Le estoy muy agradecido por ello.

Referencias

[1] Ghahramani S. (1999) *Fundamentals of probability* (2nd. edition). Prentice Hall.

[2] Grinstead C. M. y Snell J. L. (1997) *Introduction to probability*. AMS.