Búsqueda y análisis de la estrategia óptima para ganar en el problema de la dote

Javier Cejudo Goñi

Resumen

Esta monografía trata de buscar la solución de un problema que apareció por primera vez en 1960 y que es conocido, entre otros nombres, como el "problema de la dote". En síntesis, consiste en lo siguiente:

Se nos presentan un número determinado de cajas, cada una con una cantidad distinta de dinero. Vamos abriendo una a una las cajas. Cada vez que se abre una caja podemos quedarnos con su contenido o rechazarla y pasar a la siguiente. Si llegásemos de esta manera hasta la última caja, debemos quedarnos con ella. Cabe destacar que no conocemos el máximo valor.

Por supuesto, queremos quedarnos con la mejor caja, para lo cual debemos emplear una estrategia con la mayor probabilidad de éxito posible. Con este objetivo exploraremos el problema y hallaremos su solución a través de un proceso cuidadosamente detallado, en el que ofreceremos entre otras cosas, dos métodos de cálculo de la probabilidad de cada estrategia: uno por medio de sumas y otro por medio de integrales.

Finalmente estableceremos la regla de oro para la toma de decisiones: "La estrategia óptima es dejar pasar el primer e^{-1} % de cajas y elegir después la primera con una cantidad de dinero superior a cualquiera de las cajas que se rechazaron. La probabilidad de ganar siguiendo esta regla es también de e^{-1}".

Índice

1. Introducción

La idea de la presente monografía surgió tras la lectura de un artículo del número 103 de la revista “*The American Mathematical Monthly*”[1] titulado “*Decision Making: A Golden Rule*” escrito por Dimitris A. Sardelis y Theodoros M. Valahas (ver la referencia completa en la bibliografía final).

Elegí este tema por la sorprendente solución que se obtiene y por la posibilidad que ofrece de mostrar una serie de conocimientos adquiridos a lo largo del programa de estudios: integrales definidas, límites, sumas telescópicas, desarrollos en serie de Maclaurin...

Imaginemos que nos proponen el siguiente juego: Se nos presentan, aleatoriamente y de una en una, un número determinado de cajas con una suma de dinero diferente en el interior de cada una. Cuando se abre una caja tenemos dos opciones: cogerla y quedarnos con ella o rechazarla y que nos presenten la siguiente caja. El proceso se repite hasta que aceptemos una caja; en caso de que se nos hayan presentado todas y no hayamos escogido ninguna debemos quedarnos con la última. La pregunta que debemos hacernos es cuál es la mejor opción para conseguir hacernos con la caja con mayor “valor”, es decir, que contiene la mayor cantidad de dinero.

Este problema se planteó por primera vez en el número de febrero de una revista mensual llamada “*Scientific American*”[2] en 1960. Desde entonces, el problema ha sido tratado por sobresalientes estadísticos y expertos en probabilidad que lo han extendido en múltiples direcciones. En sus diversas variantes ha sido conocido como el “problema de la secretaria”, el “problema del prometido”, el “problema del matrimonio” o el “problema de la dote del sultán”, entre otros.

La solución obtenida es extraordinaria: La estrategia óptima se alcanza

[1] La “American Mathematical Monthly” es una revista matemática fundada en 1894 por Benjamin Finkel y en la actualidad es publicada por la Asociación Americana de Matemáticas.

[2] Fundada por Rufus Porter, se publicó por primera vez el 28 de agosto de 1845. Actualmente es una de las revistas científicas más prestigiosas.

cuando el primer 37 % aproximadamente de cajas se rechaza, eligiendo después la primera con un valor mayor que todas las rechazadas. De esta forma, la probabilidad de éxito es también del 37 %, o más exactamente de e^{-1}. Esta regla bien merece ser llamada "*Regla de oro para tomar decisiones*"[3]. El objetivo de la presente monografía es desarrollar el problema para demostrar que la regla es correcta.

2. En torno a las estrategias posibles

Todas las estrategias posibles están entre las dos igualmente efectivas que constituyen la peor opción. Por un lado, podemos elegir la primera caja; por otro lado, podemos escoger la última. La probabilidad de ganar con cualquiera de estas dos estrategias es de $1/N$, siendo N el número total de cajas. Así, el número total de estrategias es también de N, pues desde un principio podemos decidir escoger la primera, la segunda, la tercera, ..., hasta la N-ésima caja. Descartando las dos estrategias extremas por ser las peores, quedan exactamente $N-2$ estrategias no triviales. Denominaremos estrategia S_n a la consistente en dejar pasar n cajas, $1 \leq n \leq N-2$, recordando cuál es el mayor valor que ha salido en las cajas rechazadas, y después de esto elegir la primera que tenga un valor superior a cualquiera de las n cajas que se dejaron pasar.

Nuestro próximo objetivo es elegir la estrategia con mayor probabilidad de éxito entre todas las estrategias S_n posibles.

3. Ejemplos prácticos

Para comprender mejor el problema vamos a comprobar la eficacia de las estrategias cuando el número total de cajas es pequeño, por ejemplo, cuando $N = 3, 4$ ó 5. Calcularemos las probabilidades realizando las tablas con el

[3]Del inglés "A Golden Rule for Decision Making".

total de ordenaciones posibles de los valores de todas las cajas y contando el número de casos en los que gana cada estrategia.

Caso $N = 3$: Para facilitar las cosas los valores de las tres cajas serán 1, 2 y 3 unidades. Hay un total de 3!, es decir 6, ordenaciones posibles y sólo una estrategia no trivial. Observando la Tabla 1, comprobamos que la estrategia S_1 gana en las ordenaciones 2, 3 y 4, es decir, en el 50 % de los casos.

1	1	2	3	
2	1	3	2	S_1
3	2	1	3	S_1
4	2	3	1	S_1
5	3	1	2	
6	3	2	1	

TABLA 1. Ordenaciones y estrategias ganadoras ($N = 3$)

Caso $N = 4$: Con cuatro cajas tenemos 4!, es decir 24, ordenaciones posibles y dos estrategias no triviales. La mejor es S_1, que gana en las ordenaciones 5, 6, 8 y desde la 11 hasta la 18, ambas inclusive. Como se puede ver en la Tabla 2, las probabilidades de éxito de cada una de ellas son $P(S_1) = 11/24$ y $P(S_2) = 10/24$, de manera que la estrategia óptima es S_1.

1	1	2	3	4		
2	1	2	4	3		S_2
3	1	3	2	4		S_2
4	1	3	4	2		S_2
5	1	4	2	3	S_1	
6	1	4	3	2	S_1	
7	2	1	3	4		
8	2	1	4	3	S_1	S_2
9	2	3	1	4		S_2
10	2	3	4	1		S_2
11	2	4	1	3	S_1	
12	2	4	3	1	S_1	

13	3	1	2	4	S_1	S_2
14	3	1	4	2	S_1	S_2
15	3	2	1	4	S_1	S_2
16	3	2	4	1	S_1	S_2
17	3	4	1	2	S_1	
18	3	4	2	1	S_1	
19	4	1	2	3		
20	4	1	3	2		
21	4	2	1	3		
22	4	2	3	1		
23	4	3	1	2		
24	4	3	2	1		

TABLA 2. Ordenaciones y estrategias ganadoras ($N = 4$)

Caso $N = 5$: Hay 5!, es decir 120, ordenaciones y tres estrategias no triviales con las siguientes probabilidades de éxito: $P(S_1) = 50/120$, $P(S_2) = 52/120$ y $P(S_3) = 42/120$. La estrategia óptima es, por tanto, S_2 (consultar la Tabla 3 en la página siguiente).

Una vez estudiados los tres casos anteriores podemos concluir que:

a) las estrategias no triviales aumentan notablemente la probabilidad en comparación con una elección al azar, que la reduce a $1/N$.

b) entre las estrategias posibles unas son más efectivas que otras.

El problema es que la tarea de encontrar la estrategia más efectiva por el método anterior resulta muy afanosa para altos valores de N. De hecho, para $N = 15$ hay 1,307,674,368,000 ordenaciones, de forma que se hace necesario buscar un método de cálculo más liviano para hallar la probabilidad de éxito de las estrategias. Ése es nuestro próximo objetivo.

1	1	2	3	4	5			
2	1	2	3	5	4			S_3
3	1	2	4	3	5			S_3
4	1	2	4	5	3			S_3
5	1	2	5	3	4		S_2	
6	1	2	5	4	3		S_2	
7	1	3	2	4	5			
8	1	3	2	5	4		S_2	S_3
9	1	3	4	2	5			S_3
10	1	3	4	5	2			S_3
11	1	3	5	2	4		S_2	
12	1	3	5	4	2		S_2	
13	1	4	2	3	5	S_2		S_3
14	1	4	2	5	3		S_2	S_3
15	1	4	3	2	5		S_2	S_3
16	1	4	3	5	2		S_2	S_3
17	1	4	5	2	3		S_2	
18	1	4	5	3	2		S_2	
19	1	5	2	3	4	S_1		
20	1	5	2	4	3	S_1		
21	1	5	3	2	4	S_1		
22	1	5	3	4	2	S_1		
23	1	5	4	2	3	S_1		
24	1	5	4	3	2	S_1		
25	2	1	3	4	5			
26	2	1	3	5	4		S_2	S_3
27	2	1	4	3	5		S_2	S_3
28	2	1	4	5	3			S_3
29	2	1	5	3	4			
30	2	1	5	4	3			
31	2	3	1	4	5			
32	2	3	1	5	4		S_2	S_3
33	2	3	4	1	5			S_3
34	2	3	4	5	1			S_3
35	2	3	5	1	4		S_2	
36	2	3	5	4	1		S_2	
37	2	4	1	3	5		S_2	S_3
38	2	4	1	5	3		S_2	S_3
39	2	4	3	1	5		S_2	S_3
40	2	4	3	5	1		S_2	S_3
41	2	4	5	1	3		S_2	
42	2	4	5	3	1		S_2	
43	2	5	1	3	4	S_1		
44	2	5	1	4	3	S_1		
45	2	5	3	1	4	S_1		
46	2	5	3	4	1	S_1		
47	2	5	4	1	3	S_1		
48	2	5	4	3	1	S_1		
49	3	1	2	4	5			
50	3	1	2	5	4	S_1	S_2	S_3
51	3	1	4	2	5			S_3
52	3	1	4	5	2			S_3
53	3	1	5	2	4	S_1	S_2	
54	3	1	5	4	2	S_1	S_2	
55	3	2	1	4	5			
56	3	2	1	5	4	S_1	S_2	S_3
57	3	2	4	1	5			S_3
58	3	2	4	5	1			S_3
59	3	2	5	1	4	S_1	S_2	
60	3	2	5	4	1	S_1	S_2	
61	3	4	1	2	5		S_2	S_3
62	3	4	1	5	2		S_2	S_3
63	3	4	2	1	5		S_2	S_3
64	3	4	2	5	1		S_2	S_3
65	3	4	5	1	2		S_2	
66	3	4	5	2	1		S_2	
67	3	5	1	2	4	S_1		
68	3	5	1	4	2	S_1		
69	3	5	2	1	4	S_1		
70	3	5	2	4	1	S_1		
71	3	5	4	1	2	S_1		
72	3	5	4	2	1	S_1		
73	4	1	2	3	5	S_1	S_2	S_3
74	4	1	2	5	3	S_1	S_2	S_3
75	4	1	3	2	5	S_1	S_2	S_3
76	4	1	3	4	2	S_1	S_2	S_3
77	4	1	5	2	3	S_1	S_2	
78	4	1	5	3	2	S_1	S_2	
79	4	2	1	3	5	S_1	S_2	S_3
80	4	2	1	5	3	S_1	S_2	S_3
81	4	2	3	1	5	S_1	S_2	S_3
82	4	2	3	5	1	S_1	S_2	S_3
83	4	2	5	1	3	S_1	S_2	
84	4	2	5	3	1	S_1	S_2	
85	4	3	1	2	5	S_1	S_2	S_3
86	4	3	1	5	2	S_1	S_2	S_3
87	4	3	2	1	5	S_1	S_2	S_3
88	4	3	2	5	1	S_1	S_2	S_3
89	4	3	5	1	2	S_1	S_2	
90	4	3	5	2	1	S_1	S_2	
91	4	5	1	2	3	S_1		
92	4	5	1	3	2	S_1		
93	4	5	2	1	3	S_1		
94	4	5	2	3	1	S_1		
95	4	5	3	1	2	S_1		
96	4	5	3	2	1	S_1		
97	5	1	2	3	4			
98	5	1	2	4	3			
99	5	1	3	2	4			
100	5	1	3	4	2			
101	5	1	4	2	3			
102	5	1	4	3	2			
103	5	2	1	3	4			
104	5	2	1	4	3			
105	5	2	3	1	4			
106	5	2	3	4	1			
107	5	2	4	1	3			
108	5	2	4	3	1			
109	5	3	1	2	4			
110	5	3	1	4	2			
111	5	3	2	1	4			
112	5	3	2	4	1			
113	5	3	4	1	2			
114	5	3	4	2	1			
115	5	4	1	2	3			
116	5	4	1	3	2			
117	5	4	2	1	3			
118	5	4	2	3	1			
119	5	4	3	1	2			
120	5	4	3	2	1			

Tabla 3. Ordenaciones y estrategias ganadoras ($N = 5$)

4. Probabilidad de éxito de una estrategia

Observando detenidamente las tablas vemos que la estrategia S_n puede fallar en dos aspectos:

1. Puede suceder que la caja con mayor valor se encuentre entre las n primeras que rechazamos, de modo que es imposible que sea elegida. Por ejemplo, para $N = 4$ tenemos dos de estas estrategias: S_1 y S_2. La estrategia S_2 pierde siempre que la caja con el valor 4 (máximo valor) aparece la primera o la segunda. La estrategia S_1 pierde siempre que el valor 4 aparece en primer lugar.

2. Dando por hecho que no está entre las n primeras puede ocurrir que la caja con el mayor valor esté precedida por otra caja con un valor mayor que cualquiera de las n primeras, de manera que sería ésta la que elegiríamos y no la mejor. Por ejemplo, de nuevo para $N = 4$, la estrategia S_2 pierde en las ordenaciones 1 y 7 porque siguiendo la estrategia se ha elegido la caja con el valor 3.

Podemos apuntar que S_n es una estrategia ganadora si y sólo si para algún número k entre $n + 1$ y N tienen lugar los siguientes sucesos a la vez:

(E_k): La mejor caja está en la posición k.

(F_k): Ninguna de las cajas $n + 1, n + 2, \ldots, k - 1$ es mejor que la mejor de las cajas $1, 2, \ldots, n$.

De modo que $P_N(S_n)$ definida como "la probabilidad de ganar de una estrategia S_n para un número N de cajas" es la unión, desde $k = n+1$ hasta $k = N$, de los sucesos E_k y F_k. Estos sucesos son incompatibles por serlo los sucesos E_k entre sí[4]. Teniendo en cuenta que la probabilidad de una unión de sucesos incompatibles (disjuntos) es igual a la suma de las probabilidades

[4]Si la caja está en la posición $k = n + 1$ no puede estar en la posición $k = n + 2$, por ejemplo.

de los sucesos que se unen, tenemos:

$$P_N(S_n) = \sum_{k=n+1}^{N} P(E_k \cap F_k)$$

Ahora, para calcular la probabilidad de la intersección $E_k \cap F_k$ usaremos la fórmula de la probabilidad condicionada:

$$P(E_k \cap F_k) = P(E_k/F_k) \cdot P(F_k)$$

donde $P(E_k/F_k)$ es la probabilidad del suceso E_k condicionado al suceso F_k, es decir, la probabilidad de que ocurra E_k (que la mejor carta esté en la posición k) suponiendo que haya ocurrido F_k (que ninguna de las cajas $n+1, n+2, \ldots, k-1$ haya sido mejor que la mejor de las n primeras). Podemos admitir que el que haya ocurrido F_k no nos da ninguna información del contenido de la caja k-ésima, y que por lo tanto E_k es independiente de F_k:

$$P(E_k/F_k) = P(E_k)$$

De modo que vamos a tener

$$P(E_k \cap F_k) = P(E_k) \cdot P(F_k) = \frac{1}{N} \cdot \frac{n}{k-1}$$

Entonces, obtenemos la expresión

$$P_N(S_n) = \sum_{k=n+1}^{N} \frac{1}{N} \cdot \frac{n}{k-1} = \frac{n}{N} \cdot \sum_{k=n+1}^{N} \frac{1}{k-1} \tag{1}$$

para la probabilidad de que la estrategia S_n sea ganadora para un cierto N.

5. Estrategias óptimas

La expresión general obtenida en el apartado anterior para $P_N(S_n)$ constituye un herramienta muy útil para estudiar la probabilidad de éxito de cualquier estrategia S_n sin necesidad de realizar tablas con las ordenaciones posibles. La Tabla 4 (en la página 10) muestra la probabilidad de éxito para

$N = 1, 2, \ldots, 20$ de las estrategias posibles, con cuatro decimales significativos. Para cada valor de N podemos comprobar cuál es la estrategia que ofrece una mayor probabilidad de ganar, es decir, la mejor para ese número de cajas. Por ejemplo, la mayor probabilidad de ganar con un número de 16 cajas sucede cuando se emplea la estrategia S_6, que ofrece una probabilidad de 0.3881. Pero además también se ve en la tabla para qué valor de N una estrategia concreta S_n resulta óptima. Por ejemplo, la estrategia S_7 da una probabilidad máxima de 0.3850 para $N = 19$.

Esto se resume en los siguientes enunciados:

a) Para cada valor de N hay una estrategia con un probabilidad de éxito máxima, y a su vez,

b) cada estrategia S_n es óptima para un número concreto N de cajas.

Para un número de cajas N fijo, la estrategia óptima S_n corresponde al menor n que hace negativa la diferencia $\Delta P_{(n)} = P_N(S_{n+1}) - P_N(S_n)$ entre las probabilidades de ganar de dos estrategias consecutivas, o bien, al menor n que cumple $P_N(S_{n+1}) < P_N(S_n)$. Tenemos

$$\begin{aligned} \Delta P_{(n)} = P_N(S_{n+1}) - P_N(S_n) &= \frac{n+1}{N} \sum_{k=n+2}^{N} \left(\frac{1}{k-1} \right) - \frac{n}{N} \sum_{k=n+1}^{N} \left(\frac{1}{k-1} \right) \\ &= \frac{1}{N} \left[\sum_{k=n+2}^{N} \left(\frac{1}{k-1} \right) - 1 \right] \end{aligned} \tag{2}$$

Así, la mejor S_n para un N fijo es la que corresponde al menor n que cumple

$$\sum_{k=n+2}^{N} \left(\frac{1}{k-1} \right) = \frac{1}{n+1} + \frac{1}{n+2} + \frac{1}{n+3} + \ldots + \frac{1}{N-1} < 1 \tag{3}$$

Por ejemplo: pongamos que $N = 12$; vamos a encontrar la S_n óptima:

$$\sum_{k=n+2}^{12} \left(\frac{1}{k-1} \right) < 1$$

$$1 + \frac{1}{2} + \frac{1}{3} + \overbrace{\frac{1}{4} + \underbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} + \frac{1}{11}}_{0{,}93}}^{1{,}18}$$

Como se puede apreciar en el ejemplo, $k-1=5$, de donde $k=6$ y $n=4$. Por lo tanto es la estrategia S_4 la que debemos seguir. Así queda demostrado el enunciado a).

Ahora vamos a calcular el número N de cajas para el cual la estrategia S_n (donde n es un número fijo) tiene máxima probabilidad de ganar, que no es otro que el menor N que hace $\Delta P_{(N)} = P_{N+1}(S_n) - P_N(S_n)$ negativo, o bien, el menor N que cumple $P_N(S_{n+1}) < P_N(S_n)$. Ahora tenemos

$$\Delta P_{(N)} = P_{N+1}(S_n) - P_N(S_n) = \frac{n}{N+1}\sum_{k=n+1}^{N+1}\left(\frac{1}{k-1}\right) - \frac{n}{N}\sum_{k=n+1}^{N}\left(\frac{1}{k-1}\right)$$
$$= \frac{n}{N(N+1)}\left[1 - \sum_{k=n+1}^{N-1}\frac{1}{k-1}\right] \tag{4}$$

De ello se deduce que el "mejor" N para una estrategia S_n dada es la menor N para el que se cumple

$$\sum_{k=n+1}^{N}\left(\frac{1}{k-1}\right) = \frac{1}{n} + \frac{1}{n+1} + \frac{1}{n+2} + \ldots + \frac{1}{N-1} > 1 \tag{5}$$

Por ejemplo: pongamos que $n=7$; vamos a encontrar el N para el que S_7 resulta óptima:

$$\sum_{k=n+1}^{N}\left(\frac{1}{k-1}\right) > 1$$

$$\overbrace{\underbrace{\frac{1}{7}+\frac{1}{8}+\frac{1}{9}+\frac{1}{10}+\frac{1}{11}+\frac{1}{12}+\frac{1}{13}+\ldots+\frac{1}{17}}_{0{,}99}+\frac{1}{18}}^{1{,}05}$$

Vemos que $N-1=18$, de donde $N=19$. Así queda demostrado el enunciado b).

Los ejemplos anteriores se pueden comprobar en la Tabla 4 de la página siguiente.

N	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}
1	1.0000																			
2	0.5000	0.5000																		
3	0.3333	0.5000	0.3333																	
4	0.2500	0.4583	0.4167	0.2500																
5	0.2000	0.4167	0.4333	0.3500	0.2000															
6	0.1667	0.3806	0.4278	0.3917	0.3000	0.1667														
7	0.1429	0.3500	0.4143	0.4071	0.3524	0.2619	0.1429													
8	0.1250	0.3241	0.3982	0.4098	0.3798	0.3185	0.2321	0.1250												
9	0.1111	0.3020	0.3817	0.4060	0.3931	0.3525	0.2897	0.2083	0.1111											
10	0.1000	0.2829	0.3658	0.3987	0.3983	0.3728	0.3274	0.2653	0.1889	0.1000										
11	0.0909	0.2663	0.3507	0.3897	0.3984	0.3844	0.3522	0.3048	0.2444	0.1727	0.0909									
12	0.0833	0.2517	0.3366	0.3800	0.3955	0.3902	0.3683	0.3324	0.2847	0.2265	0.1591	0.0833								
13	0.0769	0.2387	0.3236	0.3700	0.3907	0.3923	0.3784	0.3517	0.3141	0.2668	0.2110	0.1474	0.0769							
14	0.0714	0.2272	0.3114	0.3600	0.3848	0.3917	0.3843	0.3651	0.3356	0.2972	0.2508	0.1973	0.1374	0.0714						
15	0.0667	0.2168	0.3002	0.350.	0.3782	0.3894	0.3873	0.3741	0.3513	0.3202	0.2817	0.2366	0.1853	0.1286	0.0667					
16	0.0625	0.2074	0.2898	0.3409	0.3712	0.3859	0.3881	0.3799	0.3627	0.3377	0.3058	0.2676	0.2238	0.1747	0.1208	0.0625				
17	0.0588	0.1989	0.2801	0.3319	0.3641	0.3816	0.3873	0.3832	0.3708	0.3509	0.3246	0.2923	0.2547	0.2122	0.1652	0.1140	0.0588			
18	0.0556	0.1911	0.2711	0.3233	0.3569	0.3767	0.3854	0.3848	0.3763	0.3608	0.3392	0.3120	0.2798	0.2429	0.2018	0.1567	0.1078	0.0556		
19	0.0526	0.1840	0.2626	0.3150	0.3498	0.3715	0.3827	0.3850	0.3799	0.3682	0.3506	0.3278	0.3001	0.2681	0.2321	0.1923	0.1490	0.1023	0.0526	
20	0.0500	0.1774	0.2548	0.3072	0.3429	0.3661	0.3793	0.3842	0.3820	0.3734	0.3594	0.3403	0.3167	0.2889	0.2573	0.2221	0.1836	0.1420	0.0974	0.0500

TABLA 4. Probabilidad de éxito de las estrategias

6. Evaluación geométrica de la probabilidad

En esta sección, primero vamos a deducir un método rápido para encontrar el mejor valor de N para cualquier estrategia S_n y al final damos una fórmula integral aproximada para la probabilidad $P_N(S_n)$, con N y n arbitrarios.

Cualquier término $1/m$ de la suma (5) puede representarse como el área de un rectángulo (ver Figura 1) $iIDd$ de base $(id) = 1$ y altura $(Mm) = 1/m$. M es un punto de la hipérbola $y = 1/x$. El área (LIM) bajo la hipérbola es

$$\begin{aligned} E \equiv (LIM) &= \int_{m-\frac{1}{2}}^{m} \frac{1}{x} dx - \frac{1}{2m} = \Big[\ln x\Big]_{m-\frac{1}{2}}^{m} - \frac{1}{2m} = \ln\left(\frac{m}{m-\frac{1}{2}}\right) - \frac{1}{2m} \\ &= \ln\left(\frac{2m}{2m-1}\right) - \frac{1}{2m} \end{aligned} \tag{6}$$

Análogamente, el área (MDN) sobre la hipérbola es

$$\begin{aligned} \varepsilon \equiv (MDN) &= \frac{1}{2m} - \int_{m}^{m+\frac{1}{2}} \frac{1}{x} dx = \frac{1}{2m} - \Big[\ln x\Big]_{m}^{m+\frac{1}{2}} = \frac{1}{2m} - \ln\left(\frac{m+\frac{1}{2}}{m}\right) \\ &= \frac{1}{2m} - \ln\left(\frac{2m+1}{2m}\right) \end{aligned} \tag{7}$$

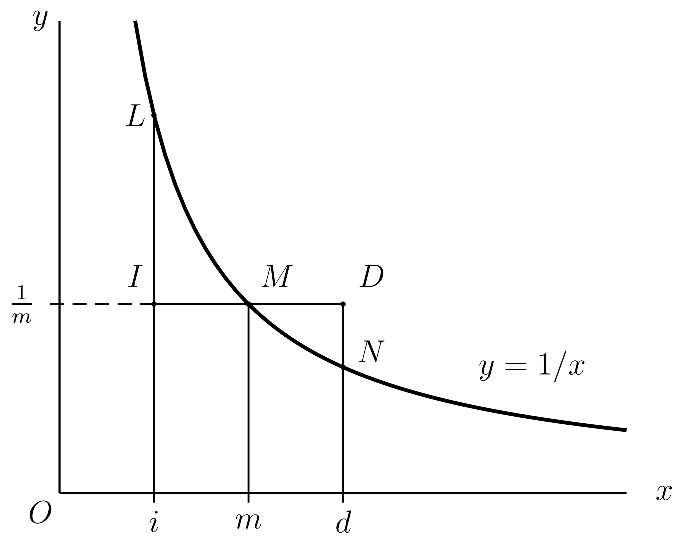

Figura 1

Empezaremos probando que el área (LIM) es mayor que el área (MDN), esto es, $E > \varepsilon$. Se tiene

$$E - \varepsilon = \ln\left(\frac{2m}{2m-1}\right) - \frac{1}{2m} + \ln\left(\frac{2m+1}{2m}\right) - \frac{1}{2m} = \ln\left(\frac{2m+1}{2m-1}\right) - \frac{1}{m}$$

Para agilizar los cálculos vamos a realizar el siguiente cambio de variable:

$$\frac{1}{m} = x$$

Así, tenemos que

$$E - \varepsilon = \ln\left(\frac{2+x}{2-x}\right) - x = \ln\left(\frac{1+\frac{x}{2}}{1-\frac{x}{2}}\right) - x = \ln\left(1+\frac{x}{2}\right) - \ln\left(1-\frac{x}{2}\right) - x$$

y por el desarrollo en serie de Maclaurin de $\ln(1+x)$

$$\begin{aligned} E - \varepsilon = -x &+ \left[+\frac{x}{2} - \frac{\left(+\frac{x}{2}\right)^2}{2} + \frac{\left(+\frac{x}{2}\right)^3}{3} - \frac{\left(+\frac{x}{2}\right)^4}{4} + \frac{\left(+\frac{x}{2}\right)^5}{5} - \frac{\left(+\frac{x}{2}\right)^6}{6} + \frac{\left(+\frac{x}{2}\right)^7}{7} - \dots\right] \\ &- \left[-\frac{x}{2} - \frac{\left(-\frac{x}{2}\right)^2}{2} + \frac{\left(-\frac{x}{2}\right)^3}{3} - \frac{\left(-\frac{x}{2}\right)^4}{4} + \frac{\left(-\frac{x}{2}\right)^5}{5} - \frac{\left(-\frac{x}{2}\right)^6}{6} + \frac{\left(-\frac{x}{2}\right)^7}{7} - \dots\right] \end{aligned}$$

Simplificando adecuadamente la expresión anterior y deshaciendo el cambio de variable obtenemos

$$\begin{aligned} E - \varepsilon &= \frac{x^3}{12} + \frac{x^5}{80} + \frac{x^7}{448} + \dots \\ &= \frac{1}{12m^3} + \frac{1}{80m^5} + \frac{1}{448m^7} + \dots > 0 \end{aligned} \tag{8}$$

La desigualdad anterior es evidente, pues todos los términos tienen signo positivo y $m > 0$. Entonces, tenemos

$$\ln\left(\frac{2m+1}{2m-1}\right) = \int_{m-\frac{1}{2}}^{m+\frac{1}{2}} \left(\frac{1}{x}\right) dx > \frac{1}{m} \tag{9}$$

Sumando la desigualdad anterior para $m = n, n+1, \dots, N-1$, resulta:

$$\int_{n-\frac{1}{2}}^{n+\frac{1}{2}} \left(\frac{1}{x}\right) dx + \int_{n+\frac{1}{2}}^{n+\frac{3}{2}} \left(\frac{1}{x}\right) dx + \dots + \int_{N-\frac{3}{2}}^{N-\frac{1}{2}} \left(\frac{1}{x}\right) dx > \frac{1}{n} + \frac{1}{n+1} + \dots + \frac{1}{N-1} \tag{10}$$

El lado izquierdo de la desigualdad puede englobarse en una sola integral, y obtenemos así la siguiente expresión:

$$\int_{n-\frac{1}{2}}^{N-\frac{1}{2}} \left(\frac{1}{x}\right) dx = \ln\left(\frac{2N-1}{2n-1}\right)$$

Para el n tal que S_n es óptima para un cierto N, el lado derecho de la desigualdad (10) es mayor que uno (ver ecuación(5)). Entonces, N es el menor entero que satisface $\ln\left(\frac{2N-1}{2n-1}\right) > 1$. Por ejemplo: nos dan $n = 7$. Vamos a encontrar el mínimo N tal que

$$\ln\left(\frac{2N-1}{13}\right) > 1\,;\ \frac{2N-1}{13} > e$$

de donde

$$N > \frac{13e+1}{2} = 18{,}1688$$

El menor entero que satisface la desigualdad es $N = 19$, como vimos en la página 9 y como se comprueba en la Tabla 4.

Ahora vamos a despejar la N en la desigualdad:

$$e^{\ln\left(\frac{2N-1}{2n-1}\right)} > e^1 = e$$

y entonces $2N - 1 > e(2n - 1)$, de donde

$$N > e\left(n - \frac{1}{2}\right) + \frac{1}{2} \tag{11}$$

Pero por otra parte, la fórmula (9) permite también estimar de otra manera la correspondiente probabilidad ganadora $P_N(S_n)$, donde N y n son cualesquiera:

$$\begin{aligned} P_N(S_n) \equiv Pr\left(\sum\right) &= \frac{n}{N}\sum_{k=n}^{N-1}\frac{1}{k} < \frac{n}{N}\sum_{k=n}^{N-1}\ln\left(\frac{2k+1}{2k-1}\right) \\ &= \frac{n}{N}\left(\ln(2n+1) - \ln(2n-1) + \ldots + \ln(2N-1) - \ln(2N-3)\right) \\ &= \frac{n}{N}\ln\left(\frac{2N-1}{2n-1}\right) \equiv Pr\left(\int\right) \end{aligned} \tag{12}$$

7. Convergencia del cálculo de la probabilidad

Hemos establecido dos métodos, por sumas (1) y por integrales (12), de cálculo de la probabilidad de ganar de una estrategia determinada S_n, para un cierto valor de N. A continuación vamos a ver qué ocurre con los valores obtenidos por cada uno de los métodos para grandes valores de N. En este sentido, la Tabla 5 muestra:

(i) la pareja de valores (n, N) para grandes N, siendo N el número de cajas para el cual la estrategia S_n tiene máxima probabilidad de ganar.

(ii) el cociente n/N,

(iii) la probabilidad de ganar $P_n(S_n)$ calculada por el método de sumas, denotada como $Pr\left(\sum\right)$,

(iv) la probabilidad de ganar $P_n(S_n)$ calculada por el método de integrales, denotada como $Pr\left(\int\right)$.

Vemos que en la Tabla 5 la probabilidad calculada por (1) y por (12) disminuye conforme aumenta n y que comienzan a coincidir con seis cifras significativas desde la entrada $n = 300$, $N = 815$. Además se observa que en la entrada $n = 2,000,000$, $N = 5,436,563$ ambos métodos coinciden con el cociente n/N en el número 0.367879, una aproximación del número e^{-1}.

Esto podría bastar para afirmar que el cociente n/N y la probabilidad de éxito de una estrategia coinciden en el límite cuando $N \to \infty$. Pero no es difícil probarlo formalmente:

Dada una estrategia S_n, ésta es óptima para el número N mínimo que cumple (11)

$$N > e\left(n - \frac{1}{2}\right) + \frac{1}{2}$$

o bien, multiplicando por 2, que cumple

$$2N > 2en - e + 1$$

n	N	n/N	$\mathbf{Pr}(\sum)$	$\mathbf{Pr}(\int)$
100	271	0.369004	0.369045	0.369046
200	543	0.368324	0.368461	0.368462
300	815	0.368098	0.368267	0.368267
400	1087	0.367985	0.368170	0.368170
500	1359	0.367918	0.368112	0.368112
600	1631	0.367872	0.368073	0.368073
700	1902	0.368034	0.368046	0.368046
800	2174	0.367985	0.368025	0.368025
900	2446	0.367948	0.368009	0.368009
1000	2718	0.367918	0.367996	0.367996
2000	5436	0.367918	0.367938	0.367938
3000	8154	0.367918	0.367918	0.367918
4000	10873	0.367884	0.367909	0.367909
5000	13591	0.367891	0.367903	0.367903
6000	16309	0.367895	0.367899	0.367899
7000	19025	0.367879	0.367896	0.367896
8000	21746	0.367884	0.367894	0.367894
9000	24464	0.367888	0.367892	0.367892
10000	27182	0.367891	0.367891	0.367891
20000	54365	0.367884	0.367885	0.367885
30000	81548	0.367881	0.367883	0.367883
40000	108731	0.367880	0.367882	0.367882
50000	135914	0.367880	0.367882	0.367882
60000	163097	0.367879	0.367881	0.367881
70000	190279	0.367881	0.367881	0.367881
80000	217462	0.367880	0.367881	0.367881
90000	244645	0.367880	0.367881	0.367881
100000	271828	0.367880	0.367881	0.367881
200000	543656	0.367880	0.367880	0.367880
300000	815484	0.367880	0.367880	0.367880
400000	1087312	0.367880	0.367880	0.367880
500000	1359141	0.367879	0.367880	0.367880
600000	1630969	0.367879	0.367880	0.367880
700000	1902797	0.367879	0.367880	0.367880
800000	2174625	0.367880	0.367880	0.367880
900000	2446453	0.367880	0.367880	0.367880
1000000	2718281	0.367880	0.367880	0.367880
2000000	5436563	0.367879	0.367879	0.367879

TABLA 5. Probabilidad de éxito de las estrategias cuando N es grande

El número $N-1$, por ser menor que N, no debe cumplir la condición (11), de modo que

$$2(N-1) \leq 2en - e + 1$$

Uniendo estas desigualdades y dividiéndolas entre $2Ne$

$$\frac{2(N-1)}{2eN} \leq \frac{2en - e + 1}{2eN} < \frac{2N}{2eN}$$

o lo que es lo mismo:

$$\frac{1}{e}\left(1 - \frac{1}{N}\right) \leq \frac{n}{N} - \frac{e-1}{2eN} < \frac{1}{e} \tag{13}$$

Pasando al límite cuando $N \to \infty$ y aplicando la "regla del sandwich" obtenemos

$$\frac{1}{e} = \lim_{N\to\infty} \frac{n}{N} = \frac{1}{e} \tag{14}$$

De esta forma queda demostrado lo que habíamos visto experimentalmente en la Tabla 5.

8. Conclusión

Hechos todos los cálculos necesarios podemos concluir que la regla de oro citada en la introducción es correcta:

La estrategia óptima para ganar en el problema de la dote es dejar pasar el primer e^{-1} % de cajas y elegir después la primera que tenga una cantidad de dinero superior a cualquiera de las cajas que se rechazaron. La probabilidad de ganar siguiendo esta regla es también de e^{-1}.

Cabe destacar que, como he indicado en la introducción, el "problema de la dote" puede extenderse en múltiples direcciones que por razones de espacio no hemos podido analizar en esta monografía, a saber:

a. Se nos permite elegir dos cajas en lugar de sólo una.

b. No conocemos el número total de cajas.

c. Nos conformamos con la segunda mejor caja.

9. Bibliografía

[1] Sardelis, Dimitris A. y Valahas, Theodoros M.: *Decision Making: A Golden Rule*, American Mathematical Monthly, Marzo 1999, pp. 215–226.