El Problema de la Secretaria

Dpto. Dementes Intratables

Estas diapositivas están disponibles en *http://jalon.org*

Esquema

1. Problema de la secretaria clásico
   - Introducción
   - Caso de tres aspirantes
   - Generalización a n >3
2. Variantes y otros problemas similares:
   - Número de aspirantes desconocido.
   - Variante de la recompensa cardinal
   - Elección del segundo mejór ó de más de uno

El Problema de la secretaria es uno de los muchos nombres para un famoso problema de teoría de toma de decisiones y de la parada óptima.

El problema también se conoce bajo el nombre de problema de la dote de la hija del sultán, del pretendiente exigente o el juego del gúgol (= googol = 10^{100}).

Aunque conocido desde mucho antes, aparentemente, su primera publicación no fué hasta Febrero del 1960 cuando el columnista Martin Gardner lo planteó en la revista Scientific American hablando de papeletas que contenían números positivos entre cero y un gúgol ó mas.

Planteamiento del problema clásico:

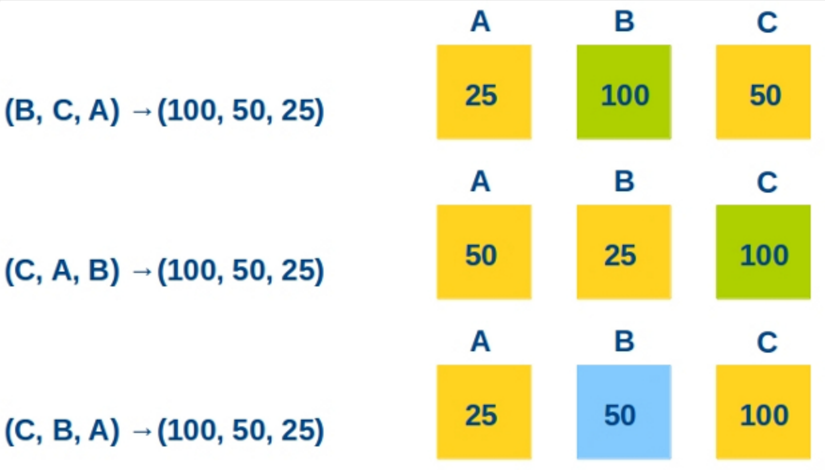
- Hay una única plaza de secretaria disponible.
- Los candidatos se pueden ordenar de mejor a peor sin ambiguedad pero llegan secuencialmente en orden aleatorio.
- Sólo podemos determinar el valor relativo de los candidatos según llegan. No podemos observar sus valores absolutos.
- Nuestro objetivo es elegir el mejor candidato, ninguno otro sirve. (El segundo mejor candidato no tiene mas valor que otro)
- Una vez que se desecha un aspirante, ella se vá para siempre y no se la puede rellamar.
- El número total de candidatos n es conocido de antemano.

Veamoslo para el caso particular n = 3

Para simplificar supongamos que tenemos 3 cajas: A, B y C. Si nos quedamos con una al azar, A, nuestro posibilidades de acertar con el premio máximo es de $1/3$.

Consideremos la siguiente estratégia: abrimos la primera caja (la A) y contamos el dinero que hay, pero la descartamos independientemente de cuánto dinero encierre. Ahora abrimos la segunda caja (la B), si contiene más dinero que la primera, nos quedamos con la segunda, en caso contrario nos quedamos con la tercera ¿En cuántos casos hemos acertado con esta estrategia?

Realicemos un examen viendo todas las posibilidades distintas.



Como vemos, con la estrategia anterior podemos garantizar un éxito del 50 % (ganamos 3 de 6), lo cual es mejor que el 33 % (=1/3) que teníamos si escogemos una caja al azar.

Lo curioso es que esta estrategia se puede aplicar a cualquier número de cajas, por sorprendente que parezca y aunque no se conseguirá siempre un éxito del 50 %, sí que podemos obtener un porcentaje sorprendentemente alto (mayor de 1/3 independientemente del número de cajas).

Si tenemos que escoger entre n cajas, abrimos unas cuantas (digamos k) y las descartamos, pero anotamos de esas k cajas cuánto dinero tenía la que más tenía. A continuación seguimos abriendo las cajas restantes y nos quedamos con la primera que tenga más dinero que el que habíamos anotado como el máximo de las k primeras.

Si ninguna tiene más dinero obviamente nos quedamos con la última. Solo queda por determinar cuánto vale k, es decir, ¿cuántas cajas tenemos que abrir y descartar inicialmente?

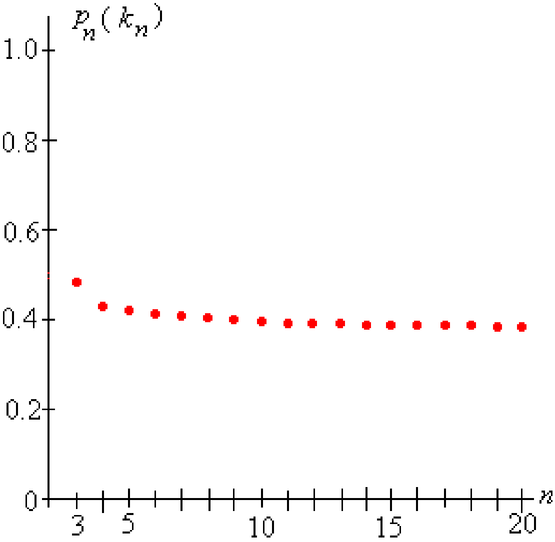
Hemos visto que en el caso de 3 cajas (n=3) k = 1 nos da una probabilidad de éxito de 1/2. Se puede comprobar que si k=0 ó k=2 el método anterior tiene una probabilidad de éxito de 1/3 en ambos casos. Es decir, algo menor que si k = 1, luego el número de cajas a desestimar óptimo es k=1.

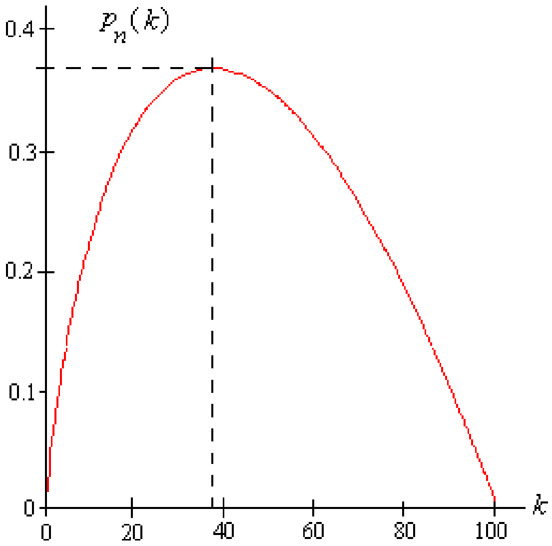
Para el caso de tener cuatro cajas {A,B,C,D} (n=4) se puede listar las 24 permutaciones y comprobar que:
$p_4(k=0) = \frac{6}{24}$, $p_4(k=1) = \frac{11}{24}$, $p_4(k=2) = \frac{10}{24}$, $p_4(k=3) = \frac{6}{24}$
Luego para n=4 el umbral k óptimo es k=1, y su probabilidad es $\frac{11}{24} \simeq 45{,}8\,\%$

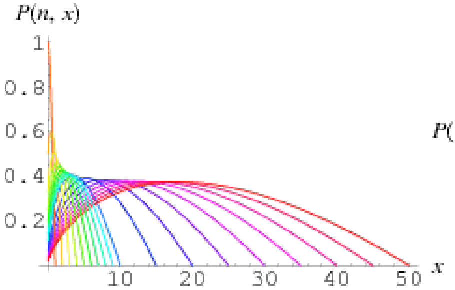
Para cinco cajas {A,B,C,D,E} (n=5) se pueden listar las 120 permutaciones y comprobar que el k óptimo es k=2 y su probabilidad es $\frac{52}{120} \simeq 0{,}433$

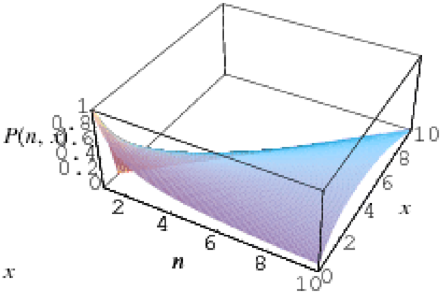
En la siguiente tabla se muestra cuántas cajas tenemos que desestimar dependiendo del número de cajas que tengamos en total para asegurar la mayor probabilidad de éxito:

Número de cajas	3	4	5	6	7	8	9
Cajas desestimadas inicialmente	1	1	2	2	3	3	3
Éxito con estrategia	50%	45,8%	43,3%	42,8%	41,4%	41%	40,6%
Éxito sin estrategia	33%	25%	20%	16,5%	14,3%	12,5%	11,1%


$p_n(k_n)$
1.0
0.8
0.6
0.4
0.2
0
3
5
10
15
20
n


$p_n(k)$
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
k



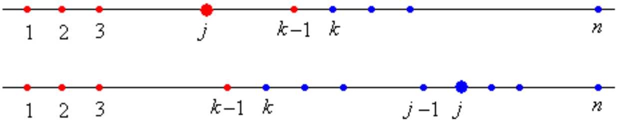


[1]Únicamente en estas dos imágenes *n* denota nuestro *k* y *x* nuestro *n*

Se puede demostrar que la estrategia óptima consiste en descartar los primeros aspirantes hasta un umbral [2] .Sea ahora k el primer aspirante a considerar (k:=k+1).

Para un umbral arbitrario k, la probabilidad de elegir el mejor aspirante es:

$$
\begin{aligned}
P(r) &= \sum_{j=1}^{n} P\left(\text{aspirante } j \text{ es seleccionado} | \text{aspirante } j \text{ es el mejor}\right) \times P\left(\text{aspirante } j \text{ es el mejor}\right) \\
&= \left[\sum_{j=1}^{k-1} 0 \times \frac{1}{n}\right] + \\
&+ \left[\sum_{j=k}^{n} P\left(\begin{array}{l} \text{el mejor de los primeros } j-1 \text{ aspirantes} \\ \text{está en los primeros } k-1 \text{ aspirantes} \end{array} \middle| \text{aspirante } j \text{ es el mejor}\right) \times \frac{1}{n}\right] \\
&= \sum_{j=k}^{n} \frac{k-1}{j-1} \times \frac{1}{n} \quad = \quad \frac{k-1}{n} \sum_{j=k}^{n} \frac{1}{j-1}.
\end{aligned}
$$

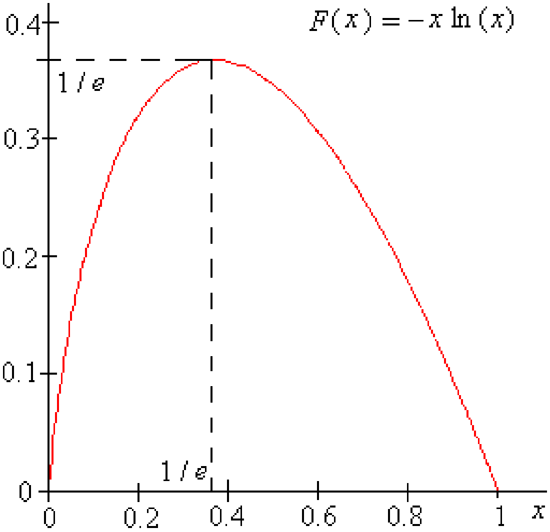


$$P(r) = \quad \frac{k-1}{n} \sum_{j=k}^{n} \frac{1}{j-1}$$

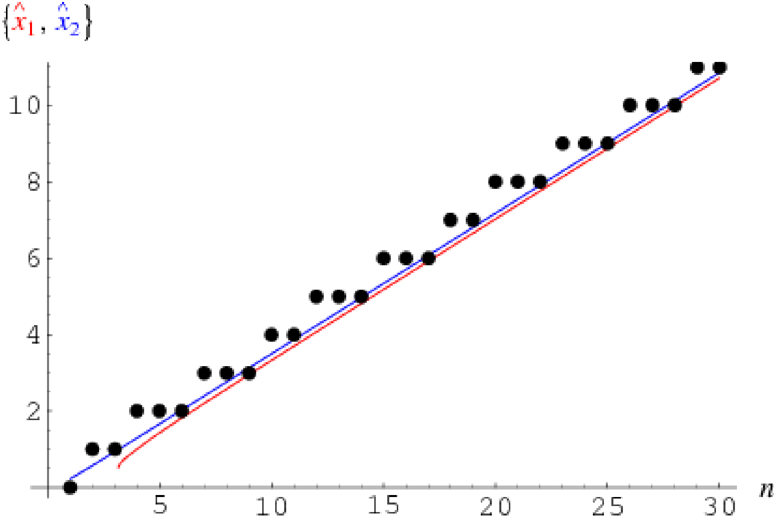
Haciendo n tender a infinito, escribiendo x como el límite de k/n, denotando a i/n por t y a $1/n$ por dt, el sumatorio se puede aproximar por la integral:

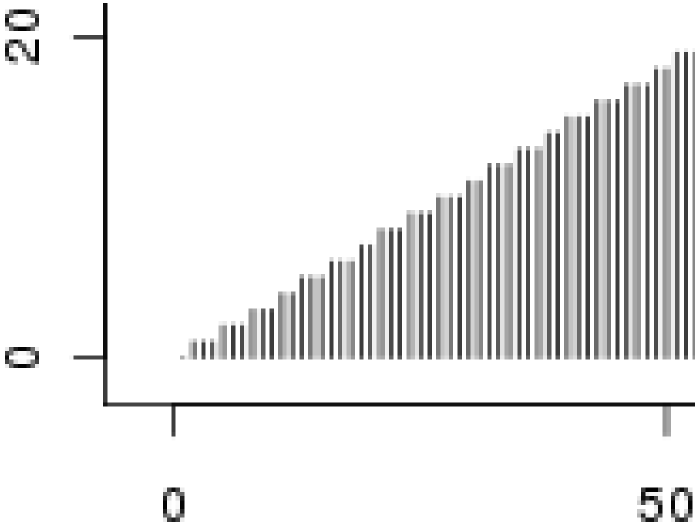
$$P(x) = x \int_x^1 \frac{1}{t}\, dt = -x \log(x).$$

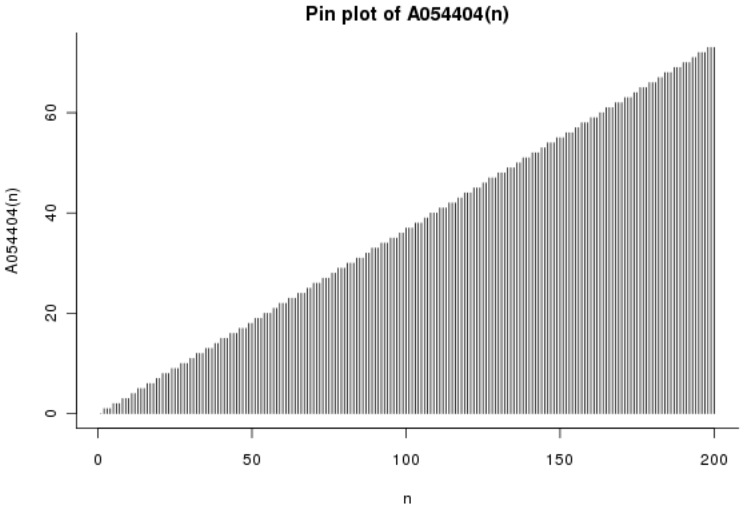
Derivando $P(x)$ respecto a x, posicionandolo en 0, y despejando x, encontramos que el óptimo x es igual a e^{-1}. Entonces, el óptimo umbral tiende a n/e según crece n, y el mejor aspirante es selecionado con probabilidad e^{-1}.

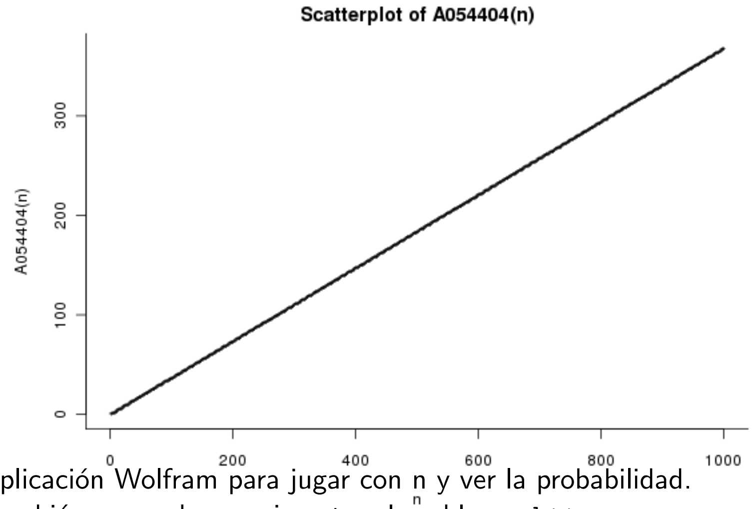

$F(x) = -x\ln(x)$
0.4
0.3
0.2
0.1
0
1 / e
1 / e
0
0.2
0.4
0.6
0.8
1
x

Número de cajas	3	4	5	6	7	8	9
Cajas desestimadas inicialmente	1	1	2	2	3	3	3
N/e	1,1036	1,4715	1,8394	2,2073	2,5752	2,943	3,3109
Éxito con estrategia	50%	45,8%	43,3%	42,8%	41,4%	41%	40,6%
Éxito sin estrategia	33%	25%	20%	16,5%	14,3%	12,5%	11,1%


$\{\hat{x}_1, \hat{x}_2\}$
10
8
6
4
2
5
10
15
20
25
30
n


20
0
0
50


Pin plot of A054404(n)
A054404(n)
0
20
40
60
0
50
100
150
200
n



Aplicación Wolfram para jugar con n y ver la probabilidad.
También se puede experimentar el problema: `http://demonstrations.wolfram.com/TheSecretaryProblem/`

Número de aspirantes desconocido.

Número de aspirantes desconocido.

Si no se conoce el valor de n por adelantado. Una manera de remediarlo es suponer que el número de aspirantes N es una variable aleatoria con una distribución conocida $P(N = k)_{k=1,2,\ldots}$ (Presman and Sonin, 1972).

La esencia del modelo está basada en que los problemas del mundo real se presentan en tiempo real y es más fácil estimar tiempos en los que los eventos ocurren (llegadas de aspirantes) que estimar la distribución del número específico de eventos que ocurriran.

El modelo

Un aspirante debe ser selecionado en un intervalo de tiempo $[0, T]$ de un número desconocido de aspirantes valorados. El objetivo es maximizar la probabilidad de selecionar el mejor. Supongamos que todos los aspirantes llegan independientemente con la misma densidad de tiempo de llegada f en $[0, T]$, y sea F la correspondiente función de distribución del tiempo de llegada:

$$F(t) = \int_0^t f(s)ds, \ 0 \leq t \leq T.$$

The 1/e-law (1984)

Sea τ tal que $F(\tau) = e^{-1}$. Consideremos la estrategia de esperar y observar todo los aspirantes hasta el instante τ y entonces seleccionar, si es posible, el mejor candidato después del instante τ el cual es mejor que todo los anteriores. Entonces, esta estrategia, tiene las siguientes propiedades:
(i) para todo N, hay una probabilidad de éxito de al menos e^{-1}.
(ii) es la única estrategia que garantiza esta probabilidad mínima de éxito e^{-1}, y la cota es óptima.
(iii) seleciona, si hay al menos un aspirante, ninguno en absoluto con probabilidad e^{-1}

Variante de la recompensa cardinal

Para modelar este problema, supongamos que los n aspirantes tiene valores reales"que son variables aleatorias X de una distribución uniforme en [0,1]. Además son i.i.d. Tal y como en el problema clásico, el entrevistador sólo observa si el aspirante es el mejor hasta el momento sin conocer el valor real del aspirante entre todos los aspirantes. Sin embargo, en esta versión su recompensa está dada por el valor real del aspirante selecionado. Por ejemplo, si él selecciona un aspirante cuyo valor real es 0.8, entonces gana 0.8. El objetivo del entrevistador es maximizar el valor esperado del aspirante selecionado.

Puesto que los valores de los aspirantes son extraciones i.i.d de una distribución uniforme en [0,1], el valor esperado de la t-ésima aspirante dado que
$x_t = \text{máx}\{x_1, x_2, \ldots, x_t\}$ está dada por:

$$E_t = E[X_t | I_t = 1] = \frac{t}{t+1}$$

Tal y como en el problema clásico, la política óptima está dada por un umbral, en cual denotaresmos c para este problema, a partir del cual el entrevistador debería empezar a aceptar candidatos.

Bearden (2006) demostró que c es:

$c = \lfloor \sqrt{n} \rfloor$ ó $c = \lceil \sqrt{n} \rceil$ (De hecho, el que esté mas cerca)

Una página web que calcula el umbral: `http://utilitymill.com/utility/Secretary_Problem_Optimizer`

There are at least two variants of the secretary problem that also have simple and elegant solutions.

Elección del segundo mejor

One variant replaces the desire to pick the best with the desire to pick the second-best. Robert J. Vanderbei calls this the "postdoc" problem arguing that the "best" will go to Harvard. For this problem, the probability of success for an even number of applicants is exactly $\frac{0{,}25n^2}{n(n-1)}$.. This probability tends to 1/4 as n tends to infinity illustrating the fact that it is easier to pick the best than the second-best.

Número de elegidos exactamente la mitad de los aspirantes

For a second variant, the number of selections is specified to be greater than one. In other words, the interviewer is not hiring just one secretary but rather is, say, admitting a class of students from an applicant pool. Under the assumption that success is achieved if and only if all the selected candidates are superior to all of the not-selected candidates, it is again a problem that can be solved. It was shown in Vanderbei 1980 that when n is even and the desire is to select exactly half the candidates, the optimal strategy yields a success probability of $\frac{1}{n/2+1}$

Estudios experimentales

Experimental psychologists and economists have studied the decision behavior of actual people in secretary problem situations.[1] In large part, this work has shown that people tend to stop searching too soon. This may be explained, at least in part, by the cost of evaluating candidates. In real world settings, this might suggest that people do not search enough whenever they are faced with problems where the decision alternatives are encountered sequentially.

¿A cuántas personas hay que conocer antes de elegir a la pareja definitiva? La profesora de la Universidad de Sidney, Clio Creswell. En su libro "Mathematics and Sex", se pregunta: ¿a cuántas personas debemos conocer, como mínimo, antes de elegir a la pareja definitiva? La respuesta, según ella, es 12. Es decir, la mejor estrategia consiste en elegir como pareja a la siguiente persona que "mejore" a esos 12 pretendientes.

Bibliografía básica

- J.N. Bearden, *A new secretary problem with rank-based selection and cardinal payoffs*, Journal of Mathematical Psychology 50 (2006) 58–59.
- T. Grams, *Lieses Verehrer oder das „Sekretärinnen-Problem“*, Eine Aufgabe zur stochastischen Simulation und deren analytische Lösung, revisado (Dec, 2008), disponible en `http://www2.hs-fulda.de/~grams/`.
- K. Siegrist, *The Secretary Problem*, `http://www.math.uah.edu/stat/urn/Secretary.html`.
- N. J. A. Sloane, *Sequences A007676/M0869, A007677/M2343, A054404 and A068985* , en "The On-Line Encyclopedia of Integer Sequences."