Pontificia Universidad JAVERIANA Bogotá

UNIVERSITAS SCIENTIARUM
Revista de la Facultad de Ciencias
Vol. 13 N° 1, 5-10

EL PROBLEMA DEL CUMPLEAÑOS, UNA GENERALIZACIÓN

THE BIRTHDAY PROBLEM: AN OVERVIEW

Moisés Aranda[1], Fabio Molina[1], Vladimir Moreno[2]

[1] *Departamento de Matemáticas, Pontificia Universidad Javeriana y Universidad Sergio Arboleda.*
[2] *Departamento de Matemáticas, Facultad de Ciencias, Pontificia Universidad Javeriana*
Cra. 7 Nº 43-82. Bogotá, Colombia
molinaf@javeriana.edu.co

Resumen

El problema del cumpleaños, en el contexto clásico, se resuelve asumiendo una distribución de probabilidades uniforme discreta de los nacimientos. El propósito de este artículo es resolver el mismo problema bajo una distribución de probabilidades discreta arbitraria, y demostrar que bajo la distribución uniforme discreta, la probabilidad de que dos o más personas cumplan años en el mismo día es subestimada.

Palabras clave: combinatoria, probabilidad, distribuciones discretas, muestra aleatoria, Simplex.

Abstract

In the classic context, the birthday problem is solved assuming a discrete uniform probability distribution of births. The main purpose of this paper is to solve the birthday problem through an arbitrary discrete probability distribution and to prove that using discrete uniform distribution, probability of two or more people's birthday at the same date is underestimated.

Key words: combinatorial, probability, discrete distributions, random sample, simplex.

INTRODUCCIÓN

El problema del cumpleaños se presenta de la siguiente manera, ver [2], [3], [4], [5]:
Se supone que a una reunión asisten k personas, las cuales registran el día de su cumpleaños. Se pregunta: ¿cuál es la probabilidad de que dos o más personas, de las k asistentes, cumplan años el mismo día?

A cada fecha de cumpleaños se le asigna un número entero entre 1 y 365, no teniendo en cuenta el año bisiesto, por ejemplo si una persona cumple años el 25 de marzo entonces se le asignará el número 84 = 31 + 28 + 25.

En el siguiente análisis se asume que k < 365, pues en caso contrario se tendrá la solución trivial: probabilidad de que dos o más personas cumplan años el mismo día es igual a 1.

Este problema se resuelve utilizando el complemento del evento definido por "dos o más personas cumplan años el mismo día", es decir, calculando la probabilidad del evento "todos los k asistentes cumplen años en día diferente".

Planteamiento del problema clásico.

Sea $\Omega = \{n \in N: 1 \leq n \leq 365\}$ el espacio muestral, la sigma álgebra es $\Im = 2^{\Omega}$ el conjunto de partes del espacio muestral, y la medida de probabilidad P está definida, sobre los eventos simples unitarios, por.

$$P(\{\omega\}) = \frac{1}{365}, \ \forall \omega \in \Omega.$$

El evento a medir es

$$A = \{(x_1, x_2, \ldots, x_k) \in \Omega^k :$$

$$x_i \neq x_j \ \forall i, j \in \{1, 2, \ldots k\} i \neq j\}$$

Empleando argumentos de conteo, se encuentra que $P(A) = \frac{per(365,k)}{365^k}$, donde $per(365,k) = \frac{365!}{(365-k)!}$ es el número de permutaciones de tamaño k en el conjunto Ω.

De esta manera la probabilidad de que haya dos o más, de los k seleccionados aleatoriamente, que cumplan años el mismo día es

$$P(A^c) = 1 - P(A) = 1 - \frac{per(365,k)}{365^k}.$$

La siguiente tabla muestra los valores de la probabilidad P (A^c) para distintos valores de k.

k	***P(A^c^)***
5	0.027
10	0.117
15	0.253
20	0.411
23	0.507
25	0.569
30	0.706
35	0.814
40	0.891
45	0.941
50	0.970
55	0.986
60	0.994
65	0.998
70	0.999

Una de las preguntas de interés en este problema, ver [3] y [5], es determinar el menor número de asistentes tal que la probabilidad de que haya dos o más con el mismo día de cumpleaños sea por lo menos el cincuenta por ciento. En la tabla se ha resaltado esta probabilidad para un tamaño de 23 asistentes.

PROBLEMA DEL CUMPLEAÑOS BAJO DISTRIBUCIÓN DE PROBABILIDADES GENERAL

Sea X una variable aleatoria con valores en el conjunto finito $A = \{x_1, x_2, ..., x_N\}$, y con función de probabilidad f. Sea $X_1, X_2, ..., X_k$, con $k < N$, variables aleatorias independientes e idénticamente distribuidas con distribución de probabilidades común f.

Proposición 1

La probabilidad p de que las k variables aleatorias $X_1, X_2, \ldots, X_k$, tomen todos sus valores distintos, dos a dos, está dada por:

$$p = k! \sum_{\substack{1 \leq i_1 < i_2 < \ldots < i_k \leq N \\ i_1, i_2, \ldots, i_k \in Z}} f(i_1) f(i_2) \ldots f(i_k)$$

donde Z denota los números enteros positivos.

Demostración

La probabilidad p de que estas k variables aleatorias tomen todos sus valores distintos es

$$p = \sum_{\substack{i_j \neq i_l \quad para \quad j \neq l \\ i_1,\ldots,i_k \in Z}} P\left(X_1 = x_{i_1}, X_2 = x_{i_2}, \ldots, X_k = x_{i_k}\right)$$

Teniendo en cuenta que si I y J son dos permutaciones de (i_1, i_2, …, i_k), donde i_1, i_2, …, i_k son elementos de A distintos dos a dos, y en total hay k! permutaciones de éstos, entonces

$$P\left((X_1, X_2, \ldots, X_k) = I\right) = P\left((X_1, X_2, \ldots, X_k) = J\right)$$

de tal forma que,

$$\text{p=k!} \sum_{\substack{i_1<i_2<\ldots<i_k \\ i_1,\ldots,i_k \in Z}} P\left(X_1 = X_{i_1}, X_2 = X_{i_2}, \ldots, X_k = iX_{i_k}\right),$$

debido a la independencia

$$\text{p=k!} \sum_{\substack{i_1<i_2<\ldots<i_k \\ i_1,\ldots,i_k \in Z}} P\left(X_1 = X_{i_1}\right) \cdot P\left(X_2 = X_{i_2}\right) \cdot \ldots \cdot P\left(X_k = X_{i_k}\right),$$

y finalmente

$$\text{p=k!} \sum_{\substack{i_1<i_2<\ldots<i_k \\ i_1,\ldots,i_k \in Z}} f(i_1) f(i_2) \cdot \ldots \cdot f(i_k)$$

Nota: Observe que el número de sumandos presentes en la sumatoria está dado por $\binom{n}{k}$

En particular, la solución al problema del cumpleaños, con n = 365, será:

$$P(C') = 1 - p;$$

considerando que se tiene una distribución uniforme sobre $C = \{1,2,3,\ldots,365\}$, entonces

$f(i_j) = \dfrac{1}{365}$ para todo $i_j \in C$ y así

$$\text{p=k!} \sum_{\substack{i_1<i_2<\ldots<i_k \\ i_1,\ldots,i_k \in Z}} \frac{1}{365^k}$$

$$k! \binom{365}{k} \frac{1}{365^k} = \frac{Per(365, k)}{365^k}.$$

Proposición 2

Sean $\{x_1, x_2, \ldots, x_N\}$un conjunto finito de números reales positivos, y $k > 1$, un número entero. Entonces

$$\sum_{\substack{1 \le i_1 < i_2 < \ldots < i_k \le N \\ i_1, i_2, \ldots, i_k \in Z}} x_{i_1} x_{i_2} \ldots x_{i_k} \le \binom{N}{k} \left(\frac{x_1 + x_2 + \ldots + x_N}{N} \right)^k$$

Demostración

Los casos extremos k = 1 y k = N se prueban fácilmente.

El caso k = 1 es directo ya que el lado izquierdo de la desigualdad anterior es x_1 + x_2 + ... + x_N, mientras que el lado derecho es

$$\binom{N}{1} \left(\frac{x_1 + x_2 + \ldots + x_N}{N} \right)^1 = x_1 + x_2 + \ldots + x_N$$

El caso k = N se sigue de la desigualdad de media geométrica – media aritmética:

$$\sqrt[N]{x_1 x_2 \ldots x_N} \le \frac{x_1 + x_2 + \ldots + x_N}{N}$$

y de la identidad $\binom{N}{N} = 1$.

Demostración de los casos k: 1< k < N:

Sin pérdida de generalidad se puede asumir que $\dfrac{x_1 + x_2 + \ldots + x_N}{N} = 1$, pues en caso contrario si $\dfrac{x_1 + x_2 + \ldots + x_N}{N} = s$, entonces basta definir las nuevas variables $x_i' = \dfrac{x_i}{s}$ y se tendrá que $\dfrac{x_1' + x_2' + \ldots + x_N'}{N} = 1$, y

$$\sum_{\substack{1\le i_1<i_2<\ldots<i_k\le N\\ i_1,i_2,\ldots,i_k\in Z}} x'_{i_1}x'_{i_2}\ldots x'_{i_k} \le \binom{N}{k}\left(\frac{x'_1+x'_2+\ldots+x'_N}{N}\right)^k$$

Se define la función $g:R_+^N \to R$, mediante

$$g(x_1,x_2,\ldots,x_N)=\sum_{\substack{1\le i_1<i_2<\ldots<i_k\le N\\ i_1,i_2,\ldots,i_k\in Z}} x_{i_1}x_{i_2}\ldots x_{i_k}$$

El dominio de la función es el conjunto cerrado y acotado $S=\{(x_1,x_2,\ldots,x_N)\in R_+^N : x_1+x_2+\ldots+x_N=N\}$ y como g es una función continua entonces g alcanza máximo sobre S.

Sea $z=(z_1,z_2,\ldots,z_N)$ el punto en S donde g alcanza el máximo, el cual resulta único ya que el conjunto S es un simplex.

El máximo de g está en $z=(1,1,\ldots,1)$, en efecto:

Se supone, por el contrario, que existen i y j, $i\neq j$, tales que $z_i\neq z_j$ y se considera el punto $y\in S$, definido por

$$y=\left(z_1,\ldots,\overbrace{\tfrac{z_i+z_j}{2}}^{i-\text{ésima}},\ldots,\overbrace{\tfrac{z_i+z_j}{2}}^{j-\text{ésima}},\ldots,z_N\right),$$ es

decir, tiene las mismas coordenadas del punto z salvo que en las posiciones i–ésima y j–ésima están los promedios de las posiciones i–ésima y j–ésima del punto z.

Entonces $g(y_1,y_2,\ldots,y_N)=\sum_{\substack{1\le i_1<i_2<\ldots<i_k\le N\\ i_1,i_2,\ldots,i_k\in Z}} y_{i_1}y_{i_2}\cdots y_{i_k}=$

$$+\sum_{\substack{1\le i_1<i_2<\ldots<i_k\le N\\ i_1,i_2,\ldots,i_k\in Z,\neq i,\neq j}} y_{i_1}y_{i_2}\cdots y_{i_k}+\sum_{\substack{1\le i_1<i_2<\ldots<i_{k-1}\le N\\ i_1,i_2,\ldots,i_{k-1}\in Z,\neq i,\neq j}} y_{i_1}y_{i_2}\cdots y_{i_{k-1}}+\sum_{\substack{1\le i_1<i_2<\ldots<i_{k-2}\le N\\ i_1,i_2,\ldots,i_{k-2}\in Z,\neq i,\neq j}} y_{i_1}y_{i_2}\cdots y_{i_{k-2}}$$

Luego, $g(y_1,y_2,\ldots,y_N)=$

$$\sum_{\substack{1\le i_1<i_2<\ldots<i_k\le N\\ i_1,i_2,\ldots,i_k\in Z,\neq i,\neq j}} y_{i_1}y_{i_2}\cdots y_{i_k}+\left(\sum_{\substack{1\le i_1<i_2<\ldots<i_{k-1}\le N\\ i_1,i_2,\ldots,i_{k-1}\in Z,\neq i,\neq j}} y_{i_1}y_{i_2}\cdots y_{i_{k-1}}\right)(y_i+y_j)+\left(\sum_{\substack{1\le i_1<i_2<\ldots<i_{k-2}\le N\\ i_1,i_2,\ldots,i_{k-2}\in Z,\neq i,\neq j}} y_{i_1}y_{i_2}\cdots y_{i_{k-2}}\right)(y_iy_j)$$

Entonces por la desigualdad de media geométrica–media aritmética, para dos componentes se tiene que:

$$g(y_1,y_2,\ldots,y_N)=\sum_{\substack{1\le i_1<i_2<\ldots<i_k\le N\\ i_1,i_2,\ldots,i_k\in Z,\neq i,\neq j}} z_{i_1}z_{i_2}\ldots z_{i_k}+\left(\sum_{\substack{1\le i_1<i_2<\ldots<i_{k-1}\le N\\ i_1,i_2,\ldots,i_{k-1}\in Z,\neq i,\neq j}} z_{i_1}z_{i_2}\ldots z_{i_{k-1}}\right)(z_i+z_j)+\left(\sum_{\substack{1\le i_1<i_2<\ldots<i_{k-2}\le N\\ i_1,i_2,\ldots,i_{k-2}\in Z,\neq i,\neq j}} z_{i_1}z_{i_2}\ldots z_{i_{k-2}}\right)\left(\frac{z_i+z_j}{2}\right)^2.$$

$$>g(z_1,z_2,\ldots,z_N)$$

Esta última desigualdad contradice que $z=(z_1,z_2,\ldots,z_N)$ sea el punto donde la función g alcanza máximo.

Por tanto $z_1=z_2=\ldots=z_N$, es decir $z=(1,1,\ldots,1)$, ya que

$$\frac{z_1+z_2+\ldots+z_N}{N}=1.$$

Proposición 3

El supuesto de distribución uniforme discreta de nacimientos para cada día de un año de 365 días subestima la probabilidad de que dos o más personas, de un conjunto de k personas con k < 365, cumplan años el mismo día.

Demostración

Bajo una distribución de probabilidades general, la probabilidad que todas las k personas cumplan años en días distintos es:

$$p=\ k!\sum_{\substack{1\le i_1<i_2<\ldots<i_k\le N\\ i_1,i_2,\ldots,i_k\in Z}} f(i_1)\,f(i_2)\ldots f(i_k)$$

Sea , $x_{i_r}=f(i_r),\ r=1,2,\ldots,k$ entonces de la proposición 2 se tiene que

$$p=\ k!\sum_{\substack{1\le i_1<i_2<\ldots<i_k\le N\\ i_1,i_2,\ldots,i_k\in Z}} f(i_1)\,f(i_2)\ldots f(i_k)\le$$

$$k!\binom{N}{k}\left(\frac{1}{N}\right)^k=\frac{N!}{(N-k)!\,N^k}$$

Esta cota corresponde a la probabilidad dada por la distribución uniforme de nacimientos, así, la probabilidad de tener cumpleaños en días distintos es maximizada por la distribución uniforme que es equivalente a que la distribución uniforme subestima la probabilidad de que dos o más personas, de un conjunto de k personas con k < 365, cumplan años el mismo día.

CONCLUSIONES

- La probabilidad de que dos o más personas cumplan años en el mismo día, bajo el supuesto de distribución uniforme discreta, es subestimada.
- En el caso de una distribución arbitraria de nacimientos el número de individuos que garantizan concurrencia en cumpleaños de por lo menos dos individuos, es a lo más 23, ya que la probabilidad está acotada inferiormente por la correspondiente probabilidad en el caso uniforme.
- Queda la pregunta de si el valor mínimo de 23 individuos, en el caso uniforme, requerido para que por lo menos dos individuos cumplan años el mismo día con probabilidad mayor al 50% se puede mejorar en el sentido de que se necesiten menos individuos en el caso general de una distribución arbitraria de nacimientos.

LITERATURA CITADA

Boyd, S. & Vandenberghe. *Convex optimization*, reprinted, Cambridge. U.K. 2006; 32-35.

Degroot M. *Probabilidad y estadística*. Adisson Wesley. Washington, 1999.

Feller, W. *An introduction to probability theory and its applications*, vol. I, 3rd edition, Wiley and sons, New York, 1968; 33.

PARZEN, E. *Teoría moderna de probabilidad y sus aplicaciones*. Limusa, Noriega Editores. México D.F. 1993; 64.

SHIRYAEV, A. N. *Probability*. Springer GTM 95. 2nd edition. Berlin Heidelberg, 1996; 15.

Recibido: 16-07-07
Aprobado: 28-02-08