

El problema de todos los cumpleaños

Mario Cortina Borja
University College London
 Chiara di Gravio
University of Southampton
 Clive Osmond
University of Southampton

“They say it’s your birthday / It’s my birthday too, yeah”
 John Lennon & Paul McCartney, (*Birthday*)

El problema de cumpleaños

se estudia en cursos elementales de probabilidad y en su forma original, planteada por [vonMises (1964)] puede enunciarse así:

Definición 1. *¿Cuántas personas se necesitan para formar un grupo en el que sea más probable que no que al menos dos de ellas tengan el mismo cumpleaños?*

Una encuesta con esta pregunta generalmente arroja resultados de este estilo:

- 365 (o 366.25 o 366, si el respondiente considera años bisiestos)
- $365 / 2 = 183$
- $365 \times 365 = 133,225$
- ¡Un montón! (o palabras equivalentes dependiendo del nivel cultural del respondiente)

La respuesta correcta es “al menos 23”, como puede mostrarse de varias maneras. Una de las más simples es la siguiente. Si ignoramos el 29 de febrero en años bisiestos y suponemos que los $N = 365$ cumpleaños se reparten independiente y uniformemente entre m personas en un grupo de manera que la probabilidad $D(N, m)$ de que nadie comparta su cumpleaños, es decir de que haya m cumpleaños diferentes en un grupo de m personas, es:

$$D(N, m) = \frac{N(N-1)(N-2) \cdots (N-m+1)}{N^m}.$$

El complemento $1 - D(N, m)$ es la probabilidad de que al menos dos personas compartan el mismo cumpleaños y para $N = 365$ o $N = 366$ esta probabilidad excede $1/2$ por primera vez cuando $m = 23$. Es posible considerar al complemento de la probabilidad $D(N, m)$ como una distribución de probabilidad para la variable aleatoria M con parámetro N y valores $\{0, 1, \dots, N\}$. Si $N = 365$, los menores números de personas en un grupo necesarias para exceder las probabilidades $\{0.5, 0.75, 0.90, 0.95, 0.99, 0.999\}$ de que al menos dos personas tengan el mismo cumpleaños son $\{23, 32, 41, 47, 57, 70\}$ respectivamente.

El supuesto de uniformidad es rara vez válido empíricamente: por ejemplo, en poblaciones humanas contemporáneas el número de nacimientos aumenta en la última semana de septiembre y decrece en fechas donde los hospitales tienen jornadas reducidas, por ejemplo en Navidad, o el 15 y 16 de septiembre en México. El primer caso refleja un incremento en la actividad sexual debido a las extensas vacaciones alrededor de Navidad y año nuevo; el segundo se debe a cambios anuales en la administración de servicios clínicos en poblaciones donde la mayor parte de los nacimientos ocurre en hospitales.

La literatura relacionada a problemas de cumpleaños, a menudo vistos como casos particulares de problemas de ocupación, de asignación, o de urnas, es muy extensa. [Feller (1968)] es la referencia canónica. Aunque está fuera del campo de nuestro artículo cabe mencionar el trabajo de [Diaconis y Holmes (2002)] que desarrolla un enfoque bayesiano sobre el libro de Feller. [Cortina Borja y Haigh (2007)] reseñan el problema original de cumpleaños con distribuciones empíricas así como algunas extensiones y su relación con la teoría de coincidencias desarrollada en el trabajo clásico de [Diaconis y Mosteller (1989)]. [Cortina Borja (2013)] reseña el problema fuerte de cumpleaños, referente a la probabilidad de que todos los integrantes de un grupo de m personas tengan un cumpleaños compartido por al menos una persona más en el grupo – es decir, nadie tiene un cumpleaños solitario.¹

En este artículo presentamos una solución analítica de las probabilidades para

El problema de todos los cumpleaños

enunciado de esta forma:

Definición 2. *¿Cuántas personas se necesitan para formar un grupo en el que sea más probable que no que todos los N cumpleaños aparezcan al menos una vez cada uno?*

Si hay N fechas de cumpleaños disponibles evidentemente la respuesta es $m > N$ y desde luego imaginamos que para $N = 365$ tendremos $m > 23$ pero ¿cuántas personas más? Un problema análogo ha sido estudiado por [Dawkins (1991)] en el marco de problemas de recolección de cupones. [Cortina Borja (2016)] presenta una solución numérica con énfasis en fechas de nacimiento en poblaciones empíricas; su estimación, basada en simulaciones, es 2285 para la mediana y 2980 para el cuantil 90% suponiendo que la distribución de cumpleaños es uniforme con $N = 365$. Ese artículo también modela la distribución para el problema de todos los cumpleaños considerando la distribución empírica de fechas de nacimiento correspondientes a personas nacidas vivas en Inglaterra y Gales entre 1979 y 2014. Para esta población, suponiendo $N = 365$ y $N = 366$ la respuesta para la mediana de M es 2296 y 2435; para el cuantil 90% es 3002 y 3642. En el apéndice 1 presentamos una función en \mathbf{R} para estimar estas cantidades en general.

Es natural preguntarse por

¹La solución de esta variante del problema de cumpleaños resulta en que un grupo de 3064 personas tiene probabilidad mayor que 50% de que todas ellas compartan su cumpleaños con alguien en el grupo; para un grupo de 5000 personas esta probabilidad es 99.45%

Una solución analítica para la probabilidad de observar todos los cumpleaños

y en este artículo esbozamos un argumento basado en recursión. Supongamos que hay N cumpleaños distribuidos uniformemente y sea ${}^N p_{n,m}$ la probabilidad de que haya n cumpleaños distintos en un grupo de m personas. Note que $1 \leq n \leq N$, y $m \geq 1$. Primero consideramos el caso $N = 3$; las probabilidades ${}^3 p_{n,m}$ pueden escribirse utilizando las siguientes relaciones de recursión:

$$\begin{aligned} {}^3 p_{1,m} &= 1/3 {}^3 p_{1,m-1} \\ {}^3 p_{2,m} &= 2/3 {}^3 p_{1,m-1} + 2/3 {}^3 p_{2,m-1} \\ {}^3 p_{3,m} &= 1/3 {}^3 p_{2,m-1} + {}^3 p_{3,m-1} \end{aligned}$$

La Tabla 1 muestra estas probabilidades para $1 \leq m \leq 8$ primero en forma racional, después en forma decimal y finalmente en una forma general que puede ser verificada por inducción. La parte en la tabla correspondiente a ${}^3 p_{m,m}$, muestra las probabilidades de que todos los cumpleaños hayan sido asignados a un grupo de m personas. Estas probabilidades se incrementan cuando m crece y la última columna para ${}^3 p_{3,m}$ en la Tabla 1 muestra la probabilidad acumulada de asignación total, en este caso para $N = 3$.

Esta es la probabilidad del problema de interés y es posible explicarla con un argumento basado en teoría de conjuntos el cual empieza examinando ¿porqué tiene dicha probabilidad una construcción tan elegante? La respuesta está en el principio de inclusión-exclusión a menudo utilizado en teoría de conjuntos e ilustrado en la Figura 1. La unión de conjuntos se expresa en la aplicación sucesiva de la adición y substracción de sus componentes y es este principio el que aplicamos a nuestro problema de la forma siguiente.

De la Figura 1 puede verse que

$$\begin{aligned} \Pr [P_1 \cup P_2 \cup P_3] &= \Pr [P_1] + \Pr [P_2] + \Pr [P_3] \\ &\quad - \Pr [P_1 \cap P_2] - \Pr [P_2 \cap P_3] - \Pr [P_3 \cap P_1] \\ &\quad + \Pr [P_1 \cap P_2 \cap P_3]. \end{aligned}$$

A continuación denotamos con P_i al evento “el cumpleaños i no aparece en el grupo.” Con $N = 3$ cumpleaños posibles y un grupo de tamaño m , tenemos $p_i = \Pr [P_i] = \left(\frac{2}{3}\right)^m$, y hay $\binom{3}{1} = 3$ tales eventos. Note que (a) $P_i \cap P_j$ es el evento “ninguno de los cumpleaños i y j aparecen en el grupo”, (b) que esto ocurre con probabilidad $\left(\frac{1}{3}\right)^m$ y (c) que hay $\binom{3}{2} = 3$ tales eventos. Finalmente, el evento $P_i \cap P_j \cap P_k$ denota “ninguno de los cumpleaños i, j, k aparecen en el grupo”, pero esto es imposible para $m > 0$ y por lo tanto tiene probabilidad 0.

Al combinar los componentes del párrafo anterior obtenemos la probabilidad de que cualquiera de los N cumpleaños no aparezca en un grupo de tamaño m como:

$$\Pr [P_i \cup P_j \cup P_k] = 3(2/3)^m - 3(1/3)^m,$$

Tabla 1: Probabilidades para el número de cumpleaños diferentes en un grupo de m personas – el caso de tres posibles cumpleaños ($N = 3$)

| m | racional | decimal | estructura |
|-----|-----------|---------|---------------------------|
| | | | ${}^3p_{1,m}$ |
| 1 | 1/1 | 1.0000 | $3(1/3)^1$ |
| 2 | 1/3 | 0.3333 | $3(1/3)^2$ |
| 3 | 1/9 | 0.1111 | $3(1/3)^3$ |
| 4 | 1/27 | 0.0370 | $3(1/3)^4$ |
| 5 | 1/81 | 0.0123 | $3(1/3)^5$ |
| 6 | 1/243 | 0.0041 | $3(1/3)^6$ |
| 7 | 1/729 | 0.0014 | $3(1/3)^7$ |
| 8 | 1/2187 | 0.0005 | $3(1/3)^8$ |
| m | | | $3(1/3)^m$ |
| | | | ${}^3p_{2,m}$ |
| 1 | 0/1 | 0.0000 | $3(2/3)^1 - 6(1/3)^1$ |
| 2 | 2/3 | 0.6667 | $3(2/3)^2 - 6(1/3)^2$ |
| 3 | 6/9 | 0.6667 | $3(2/3)^3 - 6(1/3)^3$ |
| 4 | 14/27 | 0.5185 | $3(2/3)^4 - 6(1/3)^4$ |
| 5 | 30/81 | 0.3704 | $3(2/3)^5 - 6(1/3)^5$ |
| 6 | 62/243 | 0.2551 | $3(2/3)^6 - 6(1/3)^6$ |
| 7 | 126/729 | 0.1728 | $3(2/3)^7 - 6(1/3)^7$ |
| 8 | 254/2187 | 0.1161 | $3(2/3)^8 - 6(1/3)^8$ |
| m | | | $3(2/3)^m - 6(1/3)^m$ |
| | | | ${}^3p_{3,m}$ |
| 1 | 0/1 | 0.0000 | $1 - 3(2/3)^1 + 3(1/3)^1$ |
| 2 | 0/3 | 0.0000 | $1 - 3(2/3)^2 + 3(1/3)^2$ |
| 3 | 2/9 | 0.2222 | $1 - 3(2/3)^3 + 3(1/3)^3$ |
| 4 | 12/27 | 0.4444 | $1 - 3(2/3)^4 + 3(1/3)^4$ |
| 5 | 50/81 | 0.6173 | $1 - 3(2/3)^5 + 3(1/3)^5$ |
| 6 | 180/243 | 0.7407 | $1 - 3(2/3)^6 + 3(1/3)^6$ |
| 7 | 602/729 | 0.8258 | $1 - 3(2/3)^7 + 3(1/3)^7$ |
| 8 | 1932/2187 | 0.8834 | $1 - 3(2/3)^8 + 3(1/3)^8$ |
| m | | | $1 - 3(2/3)^m + 3(1/3)^m$ |

pero este evento es el complemento de “todos los cumpleaños aparecen en un grupo de tamaño m ” el cual tiene probabilidad

$${}^3p_{3,m} = 1 - 3(2/3)^m + 3(1/3)^m, \quad (1)$$

como se muestra en la Tabla 1. Usando esta expresión vemos que un grupo con al menos 5 personas resulta en una probabilidad mayor que 0.5 de observar todos los $N = 3$ cumpleaños. Estos cálculos nos llevan a proponer expresiones analíticas para los

Momentos

de esta familia de densidades de probabilidad. Considere la variable aleatoria discreta M con función de probabilidad F_M dependiendo de un parámetro N y con valores en $R_M = \{N, N + 1, \dots, \infty\}$. M representa al número de personas suficiente para observar todos los N cumpleaños al menos una vez cada uno. Si $N = 3$, de la ecuación 1 tenemos:

$$\begin{aligned} f_M(m) &= F_M(m) - F_M(m-1) = -3 \left[\left(\frac{2}{3}\right)^m - \left(\frac{2}{3}\right)^{m-1} \right] + 3 \left[\left(\frac{1}{3}\right)^m - \left(\frac{1}{3}\right)^{m-1} \right] \\ &= \left(\frac{2}{3}\right)^{m-1} - 2 \left(\frac{1}{3}\right)^{m-1} \end{aligned}$$

de manera que el momento de orden $r \geq 1$ es $\sum_{m=3}^{\infty} m^r f_M(m)$. Esta expresión es una combinación lineal de sumas de términos de la forma $m p^{m-1}$, con $p = 1/3$ o $p = 2/3$. Si definimos

$$S_p^{(r)} = \sum_{m=N}^{\infty} m^r p^{m-1} = \sum_{m=3}^{\infty} m^{r-1} \frac{\partial}{\partial p} p^m, \quad (2)$$

obtenemos para $r = 1$

$$\sum_{m=3}^{\infty} p^m = \frac{p^3}{1-p} \quad \text{y} \quad S_p^{(1)} = \frac{\partial}{\partial p} \left(\frac{p^3}{1-p} \right) = \frac{3p^2 - 2p^3}{(1-p)^2},$$

por lo que la media es la combinación lineal $\mu_1 = S_{2/3}^{(1)} - 2S_{1/3}^{(1)}$ de la cual se tiene $\mu_M = \frac{11}{2}$. El caso general puede construirse a partir de la ecuación 2:

$$\begin{aligned} S_p^{(r)} &= \sum_{m=3}^{\infty} m^r p^{m-1} = -p(1 + 2^r p) + \text{Li}_{-r}(p) \\ &= -p(1 + 2^r p) + \sum_{j=0}^r j! \left\{ \begin{matrix} r+1 \\ j+1 \end{matrix} \right\} \left(\frac{p}{1-p} \right)^{j+1} \end{aligned} \quad (3)$$

donde $Li_a(b)$ es la función polilogarítmica expresada en función de números de Stirling de segundo orden; véase por ejemplo [Borwein (2001)].

En particular la varianza puede igualmente obtenerse de principios elementales:

$$\begin{aligned}
 S_p^{(2)} = \mu'_2 &= \sum_{m=3}^{\infty} m^2 p^{n-1} = \sum_{m=3}^{\infty} \{m(m+1)p^{m-1} - mp^{m-1}\} \\
 &= \frac{\partial^2}{\partial p^2} \sum_{m=3}^{\infty} p^{m+1} - \frac{\partial}{\partial p} \sum_{m=3}^{\infty} p^m \\
 &= \frac{\partial^2}{\partial p^2} \left(\frac{p^4}{1-p} \right) - \frac{\partial}{\partial p} \left(\frac{p^3}{1-p} \right) \\
 &= \frac{\partial}{\partial p} \left(\frac{p^3(4-3p)}{(1-p)^2} \right) - \frac{p^2(3-2p)}{(1-p)^2} \\
 &= \frac{12p^2 - 16p^3 + 6p^4}{(1-p)^3} - \frac{p^2(1-p)(3-2p)}{(1-p)^3} \\
 &= \frac{9p^2 - 11p^3 + 4p^4}{(1-p)^3},
 \end{aligned}$$

lo que resulta en $\sigma_M^2 = S_{2/3}^{(2)} - 2S_{1/3}^{(2)} - \mu_1^2 = 37 - 5.5^2 = 6.75$ con desviación estándar $\sigma_M = \frac{3\sqrt{3}}{2} = 2.598$. Omitiendo a M en la notación, los momentos respecto a 0 de orden 3 y 4 son $\mu'_3 = \frac{1237}{4} = 309.25$ y $\mu'_4 = 3196$. Los momentos centrales de orden 3 y 4 se obtienen usando las combinaciones lineales usuales a partir de las funciones $S_p^{(3)}$ y $S_p^{(4)}$ calculadas con la ecuación 3; resultan ser:

$$\mu_3 = 2\mu_1^3 - 3\mu_1\mu'_2 + \mu'_3 = \frac{63}{2} = 31.5 \quad \text{y} \quad \mu_4 = -3\mu_1^4 + 6\mu_1\mu'_2 - 4\mu_1\mu'_3 + \mu'_4 = \frac{5805}{16} = 362.813$$

Finalmente, los coeficientes de asimetría y kurtosis son:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{28}{9\sqrt{3}} = 1.796 \quad \text{y} \quad \gamma_2 = \frac{\mu_4}{\sigma^4} = \frac{215}{27} = 7.962$$

y estos valores son útiles para caracterizar la forma de la distribución de M .

El caso $N = 3$ tiene primariamente un interés expositivo; a continuación analizamos la

Extensión al caso $N = 365$ días

aplicando una vez más el principio de inclusión–exclusión esta vez considerando los días del año como etiquetas de cumpleaños.²

El siguiente resultado general es muy conocido en teoría de conjuntos:

²El término “cumpleaños” no necesariamente se refiere a fecha de nacimiento. Por ejemplo: (a) el cumpleaños oficial del soberano británico se celebra normalmente el segundo sábado de junio y (b) en el hemisferio norte, los cumpleaños de todos los caballos llamados pura sangre es oficialmente el 1º de enero.

Teorema 1. Sean $N \in \mathbb{N}$ y P_i un conjunto finito para $1 \leq i \leq N$. Entonces, si $|\cdot|$ denota cardinalidad:

$$\left| \bigcup_{1 \leq i \leq N} P_i \right| = \sum_{1 \leq i_1 \leq N} |P_{i_1}| - \sum_{1 \leq i_1 \leq i_2 \leq N} |P_{i_1} \cap P_{i_2}| \\ + \sum_{1 \leq i_1 \leq i_2 \leq i_3 \leq N} |P_{i_1} \cap P_{i_2} \cap P_{i_3}| - \dots + (-1)^{N+1} \left| \bigcap_{i=1}^N P_i \right|.$$

Con base en resultados expuestos en [Johnson, Kemp y Kotz (2005)] sobre variables aleatorias para el problema de cumpleaños y en [Charalambides (2005)] sobre la relación entre el teorema anterior y los números de Stirling de segundo orden podemos escribir:

$${}^{365}p_{m,m} = \sum_{k=0}^{365} (-1)^k \binom{365}{k} \left(\frac{365-k}{365} \right)^m \\ = \sum_{k=0}^{365} \binom{365}{k} (365-k)^m \frac{1}{365^m} = \left\{ \begin{matrix} 365 \\ m \end{matrix} \right\} \frac{365!}{365^m}$$

donde $\left\{ \begin{matrix} N \\ m \end{matrix} \right\}$ es el número de Stirling de segundo orden con argumentos N y m , es decir el número de particiones en que es posible distribuir N cumpleaños entre m personas.

Note que $\left\{ \begin{matrix} N \\ N \end{matrix} \right\} = \left\{ \begin{matrix} N \\ 1 \end{matrix} \right\} = 1$. Por ejemplo, si hay 4 objetos, a, b, c, d y se desea distribuirlos en 2 grupos hay 7 particiones, es decir, $\left\{ \begin{matrix} 4 \\ 2 \end{matrix} \right\} = 7$:

$$\{(a), (bcd)\}, \{(b), (acd)\}, \{(c), (abd)\}, \{(d), (abc)\}, \{(ab), (cd)\}, \{(ac), (bd)\}, \{(ad), (bc)\}.$$

El número de particiones puede ser muy grande aún para problemas relativamente pequeños, e.g. el número de particiones de 25 objetos en 5 grupos es $\left\{ \begin{matrix} 25 \\ 5 \end{matrix} \right\} = 2, 436, 684, 974, 110, 751$.

La probabilidad ${}^{365}p_{m,m}$ aparece en función de m en la Figura 2 y es la función de distribución de probabilidad del problema de todos los cumpleaños; la función de densidad de probabilidad correspondiente aparece en la Figura 3. En el apéndice 2 presentamos código en **R** para calcular estas funciones.

La Tabla 2 muestra valores particulares de la función de distribución de la Figura 1 obtenidos con base en la función presentada en el apéndice. De esta forma obtenemos que, por ejemplo, 2287 personas se necesitan para tener una probabilidad mayor que 1/2 de que aparezcan todos los cumpleaños. Note que esta mediana es muy cercana a la solución obtenida por simulación (2285); para el cuantil 90% la respuesta analítica es 2972, también muy cercana al estimador por simulación (2980). La Tabla 3 describe algunas características de la función de densidad de probabilidad.

Finalmente notamos

Figura 1: Principio de inclusión-exclusión ($N = 3$)

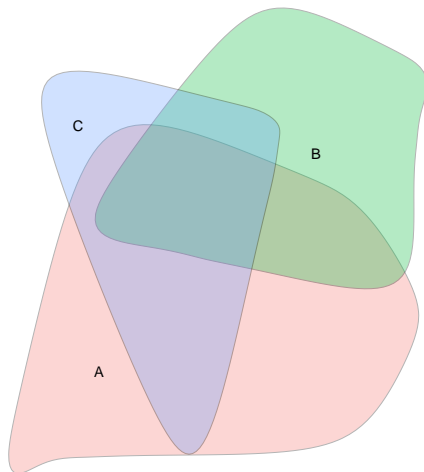


Figura 2: Probabilidad acumulada de observar todos los 365 cumpleaños en un grupo de N personas

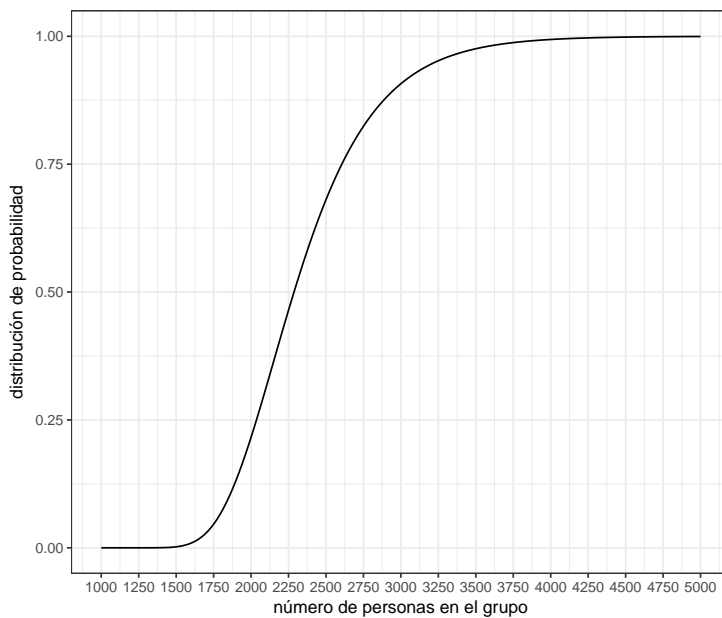
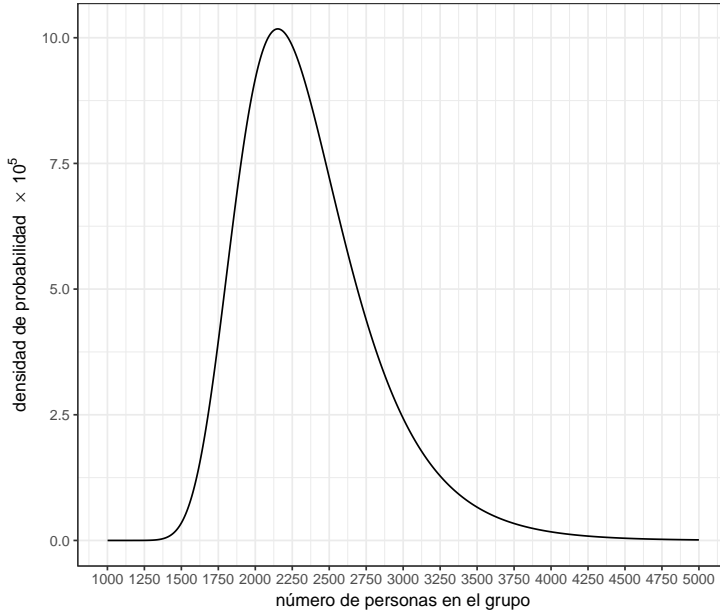


Figura 3: Densidad de probabilidad para el problema de todos los cumpleaños



Una identidad interesante

de los coeficientes binomiales. La probabilidad ${}^N p_{m,m}$ se refiere a que todos los N cumpleaños aparezcan en un grupo de m personas; por lo tanto debe ser 0 si $m < N$, de manera que:

$$\sum_{k=0}^N (-1)^k \binom{N}{k} (N - k)^m = 0$$

para toda $m < N$. Por ejemplo, si $N = 3$ tenemos:

$$\begin{aligned} 1 \times 3^1 - 3 \times 2^1 + 3 \times 1^1 &= \\ 1 \times 3^2 - 3 \times 2^2 + 3 \times 1^2 &= 0, \end{aligned}$$

para $N = 4$:

$$\begin{aligned} 1 \times 4^1 - 4 \times 3^1 + 6 \times 2^1 - 4 \times 1^1 &= \\ 1 \times 4^2 - 4 \times 3^2 + 6 \times 2^2 - 4 \times 1^2 &= \\ 1 \times 4^3 - 4 \times 3^3 + 6 \times 2^3 - 4 \times 1^3 &= 0, \end{aligned}$$

para $N = 5$:

$$\begin{aligned} 1 \times 5^1 - 5 \times 4^1 + 10 \times 3^1 - 10 \times 2^1 + 5 \times 1^1 &= \\ 1 \times 5^2 - 5 \times 4^2 + 10 \times 3^2 - 10 \times 2^2 + 5 \times 1^2 &= \\ 1 \times 5^3 - 5 \times 4^3 + 10 \times 3^3 - 10 \times 2^3 + 5 \times 1^3 &= \\ 1 \times 5^4 - 5 \times 4^4 + 10 \times 3^4 - 10 \times 2^4 + 5 \times 1^4 &= 0, \end{aligned}$$

y así sucesivamente. Esta es una propiedad muy bella de los coeficientes binomiales.

Apéndices

1. Código en **R** para el problema fuerte de cumpleaños, basado en [Cortina Borja (2016)]

```
p_hat<- function(n=365, M=3000, B=10000, emp.prob=rep(1,n)/n)
{
  ### Estima la probabilidad de cubrir todas las celdas D_n = {1,2,...,n}
  ### Genera B simulaciones extrayendo M muestras de D_n
  ### con reemplazo usando la función de distribución especificada
  ### por el vector de probabilidades empíricas emp.prob
  ### MCB, Londres, 02.11.16
  ### Publicado en Significance online, 14.11.16
  ###
  invisible(
    sum(
      apply(
        matrix(
          sample(1:n, M*B, replace=TRUE, prob=emp.prob),
          nrow=B),
          1, function(x){length(unique(x))==n})
        )/B
    )}
}
```

2. Código en **R** para el problema de todos los cumpleaños.

```
library(CryptRndTest) ### para la función Strlng2
library(ggplot2)
library(latex2exp) ### para incorporar LaTeX en expresiones
### MCB, Cambridge, 13.09.17
###
m<- 1000:7000
onto<- function(n=365, m){ invisible(lgamma(n+1) + Strlng2(m,n) - m*log(n))}
onto.V<- Vectorize(onto,'m')
```

2.65704155889863948976832052116954270403668880256364975808182748398572507

```

probs<- exp(onto.V(365,m))
### Figura 2
df<-data.frame(m=m, probs=probs)
p1<- ggplot(df, aes(x=m, y=probs)) + geom_line() +
  xlab('número de personas en el grupo') +
  ylab('distribución de probabilidad') +
  scale_x_continuous(breaks=seq(1000,7000,by=500), limits=c(1000,7000)) +
  theme_bw()
ggsave(filename='Figura2.pdf', plot=p1, device='pdf', dpi=600)
### Figura 3
df<- data.frame(m=m[-1], densidad=10000*diff(probs))
p1<- ggplot(df, aes(x=m, y=densidad)) + geom_line() +
  xlab('número de personas en el grupo') +
  ylab(TeX('densidad de probabilidad  $10^{-5}$ ')) +
  scale_x_continuous(breaks=seq(1000,7000,by=500), limits=c(1000,7000)) +
  theme_bw()
ggsave(filename='Figura3.pdf', plot=p1, device='pdf', dpi=600)

```

Referencias

- [Borwein (2001)] Borwein, J. M., Bradley, D. M., Broadhurst, D. J., Lisonek, P. Special Values of Multidimensional Polylogarithms. *Transactions of the American Mathematical Society* **353**, 907–941, 2001. https://www.researchgate.net/publication/2797416_Special_Values_of_Multidimensional_Polylogarithms
- [Charalambides (2005)] Charalambides, Ch. A. *Combinatorial Methods in Discrete Distributions*. Nueva York: Wiley. 2005.
- [Cortina Borja y Haigh (2007)] Cortina Borja, M., Haigh, J. The birthday problem. *Significance*, **4**, 124–127, 2007. Disponible en <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2007.00246.x/epdf>
- [Cortina Borja (2013)] Cortina Borja, M. The strong birthday problem. *Significance*, **10**, 18–20, 2013. Disponible en <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2013.00705.x/epdf>
- [Cortina Borja (2016)] Cortina Borja, M. A variation of the birthday problem. *Significance online*, 2016. Disponible en <https://www.significancemagazine.com/science/418-ask-a-statistician-a-variation-of-the-birthday-problem-part-2>
- [Dawkins (1991)] Dawkins, B. Siobhan’s Problem: The Coupon Collector Revisited. *The American Statistician*, **45**, 76–82, 1991. <http://www.tandfonline.com/doi/abs/10.1080/00031305.1991.10475772>

- [Diaconis y Holmes (2002)] Diaconis, P., Holmes, S. A Bayesian peek into Feller Volume I. *Sankhyā: The Indian Journal of Statistics*, **64**, 820–841, 2002. Disponible en <https://www.jstor.org/stable/25051431>
- [Diaconis y Mosteller (1989)] Diaconis, P., Mosteller, F. Methods for studying coincidences. *Journal of the American Statistical Association*, **84**, 853–861, 1989. https://www.stat.berkeley.edu/~aldous/157/Papers/diaconis_mosteller.pdf
- [Feller (1968)] Feller, W. *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed.* New York: Wiley, pp. 31-32, 1968.
- [Johnson, Kemp y Kotz (2005)] Johnson, N.L., Kemp, A.W., Kotz, S. *Univariate Discrete Distributions, 3rd edition* New York: Wiley, pp. 440, 2005.
- [vonMises (1964)] Von Mises, R. *Selected Papers of Richard von Mises, vol. 2*, Providence RI: American Mathematical Society. pp. 313-334, 1964.

Tabla 2: Valores particulares para la función de probabilidad mostrada en la Figura 2

| m | probabilidad acumulada |
|------|------------------------|
| 1607 | 0.010084 |
| 1686 | 0.025229 |
| 1760 | 0.050284 |
| 1854 | 0.100315 |
| 2000 | 0.216119 |
| 2287 | 0.500371 |
| 2500 | 0.680499 |
| 2972 | 0.900196 |
| 3000 | 0.906743 |
| 3234 | 0.950081 |
| 3491 | 0.975025 |
| 3828 | 0.990019 |

Tabla 3: Propiedades de la función de densidad de probabilidad mostrada en la Figura 3

| | |
|---------------------|------|
| Media | 2365 |
| Desviación estándar | 465 |
| Mediana | 2287 |
| Moda | 2152 |