Sanders Theater, 1984

Sanders Theater, 1984

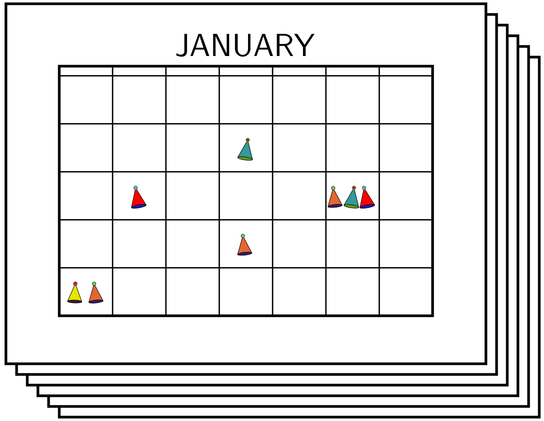
What are the chances?

Four out of 12?

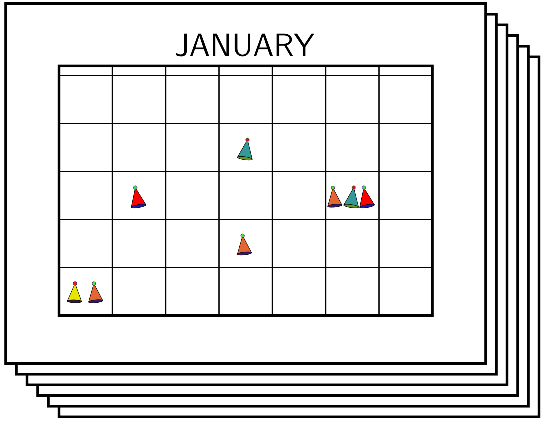
Eight out of 300?

Variation 2: The multiple birthday problem

What is the probability that there is at least one day in the calendar that is the birthday of k or more of the people in a random sample of size n?



Variation 2: The multiple birthday problem



What is the probability that there is at least one day in the calendar that is the birthday of k or more of the people in a random sample of size n?

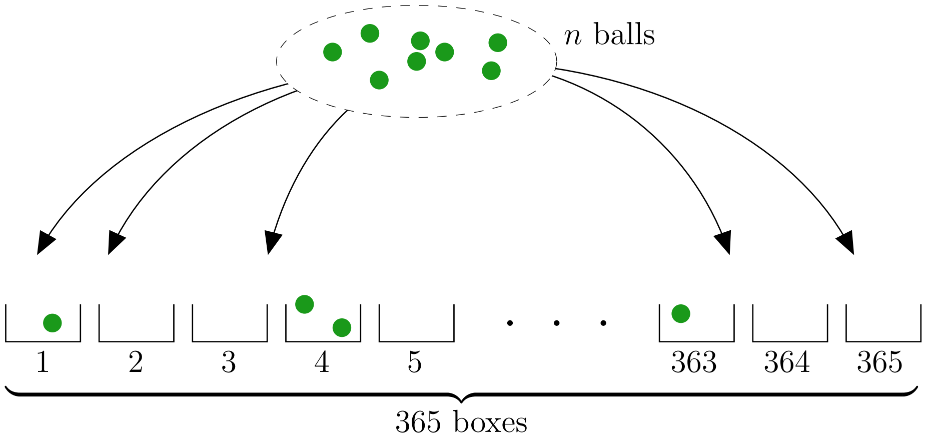
Assumption: Birthdays are uniformly distributed over 365 days.

Solution: Use complementation: The probability of at least one k-way coincidence is

$$\begin{aligned} & 1 - P(\text{no } k\text{-way coincidence}) \\ = {} & 1 - P(\text{each day is the birthday of } k-1 \text{ or fewer people}) \end{aligned}$$

MBP Solution (continued):

For $i = 1, \ldots, 365$, let X_i be the number of people in the random sample whose birthday is day i. Then $(X_1, X_2, \ldots, X_{365})$ is a random vector that follows a multinomial distribution with parameters n and $p_1 = p_2 = \cdots = p_{365} = \frac{1}{365}$.



And we want to find $P(X_i \leq k-1 \text{ for all } i)$.

MPB Solution (continued):

The probability distribution for the vector $(X_1, X_2, \ldots, X_{365})$ is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_{365} = x_{365}) \;=\; \frac{n!}{x_1!\, x_2! \;\cdots\; x_{365}!} \left(\frac{1}{365}\right)^n \text{ if } \sum x_i = n$$

and zero otherwise.

MPB Solution (continued):

The probability distribution for the vector $(X_1, X_2, \ldots, X_{365})$ is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_{365} = x_{365}) = \frac{n!}{x_1!\, x_2! \cdots x_{365}!} \left(\frac{1}{365}\right)^n \text{ if } \sum x_i = n$$

and zero otherwise.

It follows that

$$P(X_i \le k-1 \text{ for all } i) = \frac{n!}{365^n} \sum_{\substack{0 \le x_i \le k-1\ \forall i \\ x_1 + x_2 + \cdots + x_{365} = n}} \frac{1}{x_1!\, x_2! \cdots x_{365}!}$$

So we need to take a sum over all 365-tuples $(x_1, x_2, \ldots, x_{365})$ with

each x_i between 0 and $k-1$ and $x_1 + x_2 + \cdots + x_{365} = n$.

MPB Solution (continued):

Set up a generating function

$$\left.\begin{aligned}&\left(\frac{t^0}{0!}+\frac{t^1}{1!}+\frac{t^2}{2!}+\cdots+\frac{t^{k-1}}{(k-1)!}\right)\\ &\qquad\times\left(\frac{t^0}{0!}+\frac{t^1}{1!}+\frac{t^2}{2!}+\cdots+\frac{t^{k-1}}{(k-1)!}\right)\\ &\qquad\qquad\times\cdots\\ &\qquad\qquad\qquad\times\left(\frac{t^0}{0!}+\frac{t^1}{1!}+\frac{t^2}{2!}+\cdots+\frac{t^{k-1}}{(k-1)!}\right)\end{aligned}\right\}\ \text{365 factors}$$

The coefficient of t^n in this product is the sum of terms $\frac{1}{x_1!}\frac{1}{x_2!}\cdots\frac{1}{x_{365}!}$ (one factor from each "line"), where

- ▷ each x_i is between 0 and $k-1$ inclusive, and
- ▷ the sum of the x_i's is n.

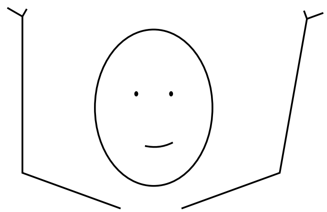
MPB Solution (continued):

Set up a generating function

$$\left.\begin{aligned}&\left(\frac{t^0}{0!}+\frac{t^1}{1!}+\frac{t^2}{2!}+\cdots+\frac{t^{k-1}}{(k-1)!}\right)\\&\qquad\times\left(\frac{t^0}{0!}+\frac{t^1}{1!}+\frac{t^2}{2!}+\cdots+\frac{t^{k-1}}{(k-1)!}\right)\\&\qquad\qquad\times\cdots\\&\qquad\qquad\qquad\times\left(\frac{t^0}{0!}+\frac{t^1}{1!}+\frac{t^2}{2!}+\cdots+\frac{t^{k-1}}{(k-1)!}\right)\end{aligned}\right\}\text{ 365 factors}$$

The coefficient of t^n in this product is the sum of terms $\frac{1}{x_1!}\frac{1}{x_2!}\cdots\frac{1}{x_{365}!}$ (one factor from each "line"), where

- ▷ each x_i is between 0 and $k-1$ inclusive, and
- ▷ the sum of the x_i's is n.



MPB Solution (conclusion):

The probability of at least one k-way birthday coincidence is

$$1 - \frac{n!}{365^n} \; [t^n] \left(\frac{t^0}{0!} + \frac{t^1}{1!} + \frac{t^2}{2!} + \cdots + \frac{t^{k-1}}{(k-1)!} \right)^{365}$$

$\uparrow$

"the coefficient of t^n in …"

MPB Solution (conclusion):

The probability of at least one k-way birthday coincidence is

$$1 - \frac{n!}{365^n} \; [t^n] \; \left(\frac{t^0}{0!} + \frac{t^1}{1!} + \frac{t^2}{2!} + \cdots + \frac{t^{k-1}}{(k-1)!} \right)^{365}$$

$\uparrow$

"the coefficient of t^n in ..."

And the answer is ...

The probability of a 4-way coincidence in a sample of 12 people is 0.0000100024.

The probability of an 8-way coincidence in a sample of 300 people is 0.000844482.

MPB Solution (conclusion):

The probability of at least one k-way birthday coincidence is

$$1 - \frac{n!}{365^n} \; [t^n] \; \left(\frac{t^0}{0!} + \frac{t^1}{1!} + \frac{t^2}{2!} + \cdots + \frac{t^{k-1}}{(k-1)!} \right)^{365}$$

$\uparrow$

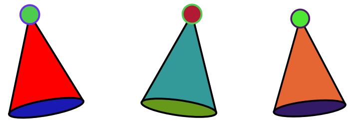
"the coefficient of t^n in ..."

And the answer is ...

The probability of a 4-way coincidence in a sample of 12 people is 0.0000100024.

The probability of an 8-way coincidence in a sample of 300 people is 0.000844482.

this took 40 seconds of CPU time.

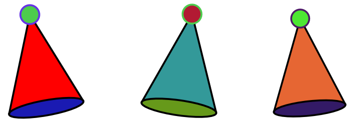


Fun Facts: Did you know . . . ?

If there are n people in a room, then there's a $p\%$ chance that some k of them share a birthday.

k	2	3	4	5	6	7	8	9	10
n for 50%	23	88	187	313	460	623	798	986	1182
n for 90%	41	132	260	413	586	773	973	1182	1400

Fun Facts: Did you know . . . ?



If there are n people in a room, then there's a p% chance that some k of them share a birthday.

k	2	3	4	5	6	7	8	9	10
n for 50%	23	88	187	313	460	623	798	986	1182
n for 90%	41	132	260	413	586	773	973	1182	1400

Fun References:

Richard Arratia, Larry Goldstein, and Louis Gordon, "Poisson approximation and the Chen-Stein method," *Statistical Science* 5(4), 1990.

Persi Diaconis and Frederick Mosteller, "Methods of studying coincidences," *Journal of the American Statistical Association* 84(408), 1989.

Bruce Levin, "A representation for multinomial cumulative distribution functions," *Annals of statistics* 9(5), 1981.

———, "On calculations involving the maximum cell frequency," *Communications in Statistics: Theory and methods* 12(11), 1983.