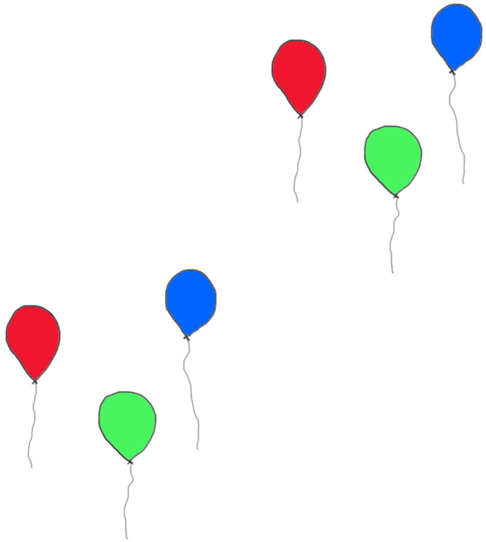
Variation 1: The near-miss birthday problem

Assuming birthdays are uniformly distributed over 365 days, find

$$P\left(\begin{array}{c}\text{at least one pair of birthdays}\\ \text{that are either coincident}\\ \text{or adjacent}\end{array}\right)$$

in a random sample of n people.

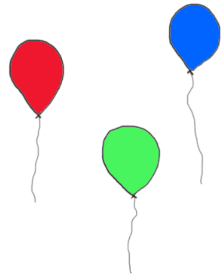


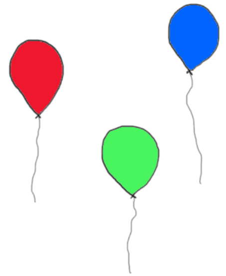
Variation 1: The near-miss birthday problem

Assuming birthdays are uniformly distributed over 365 days, find

$$P\left(\begin{array}{c}\text{at least one pair of birthdays}\\ \text{that are either coincident}\\ \text{or adjacent}\end{array}\right)$$

in a random sample of n people.

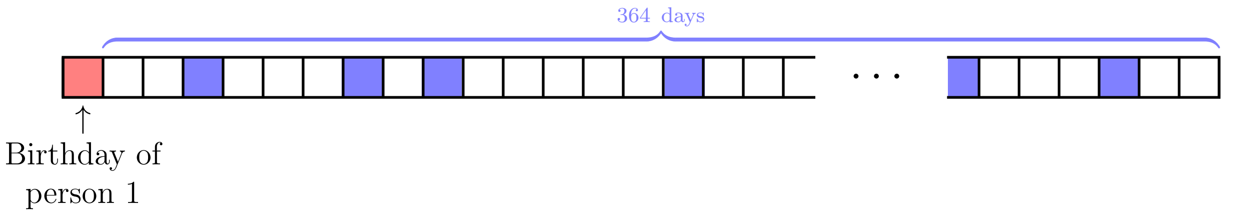




Solution: Use complementation.

$$\begin{aligned} P(\text{at least one near miss}) &= 1 - P(\text{no near misses}) \\ &= 1 - P(n \text{ isolated birthdays}) \end{aligned}$$

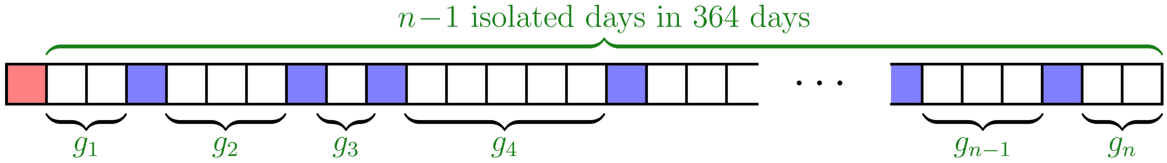
Finding $P(n \text{ isolated birthdays})$ — Strategy



$$P(n \text{ isolated birthdays}) = \frac{\begin{pmatrix}\text{number of ways to place } n-1 \text{ birthdays} \\ \text{in 364 days with no collision} \\ \textit{and} \text{ no two birthdays adjacent}\end{pmatrix}}{\begin{pmatrix}\text{total number of ways to place} \\ n-1 \text{ birthdays in 365 days}\end{pmatrix}}$$

$$= \frac{\begin{pmatrix}\text{number of ways to choose} \\ n-1 \text{ isolated days}\end{pmatrix} \times (n-1)!}{(365)^{n-1}}$$

Choosing a set of $n-1$ isolated days



Every set of $n-1$ isolated days

corresponds to a "gap sequence"

$$g_1,\ g_2,\ \ldots,\ g_n$$

in which $\begin{cases} g_i \geq 1 \text{ for all } i \\ g_1 + g_2 + \cdots + g_n = 365 - n \end{cases}$

How many such gap sequences are there?

Counting gap sequences

Let $a_i = g_i - 1$ to see that every gap sequence

$$g_1,\, g_2,\, \ldots,\, g_n \;:\quad g_i \geq 1\ \forall i,\quad \sum_i g_i = 365 - n$$

corresponds to a sequence

$$a_1,\, a_2,\, \ldots,\, a_n \;:\quad a_i \geq 0\ \forall i,\quad \sum_i a_i = 365 - 2n.$$

So we want to count

sequences of n non-negative integers with fixed sum S.

This is a standard combinatorial "occupancy" problem.

Counting sequences using dots and bars

We can represent each sequence

$$a_1,\ a_2,\ \ldots,\ a_n\ :\quad a_i \geq 0\ \forall i,\quad \sum_i a_i = S$$

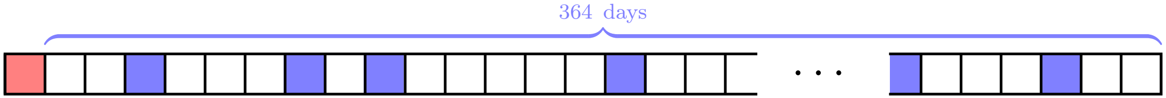
using a row of S dots and $n-1$ bars.

$$\underbrace{\bullet\quad\bullet\quad\bullet}_{a_1}\quad \underset{+}{\Big|}\quad \underbrace{\bullet\quad\bullet}_{a_2}\quad \underset{+}{\Big|}\quad \underbrace{\bullet\quad\bullet\quad\bullet\quad\bullet}_{a_3}\quad \underset{+}{\Big|}\quad \underbrace{\bullet\quad\bullet\quad\bullet}_{a_4}$$

The sequence $a_1,\ a_2,\ \ldots,\ a_n$ is determined by the positions of the $n-1$ bars among these $S+n-1$ objects.

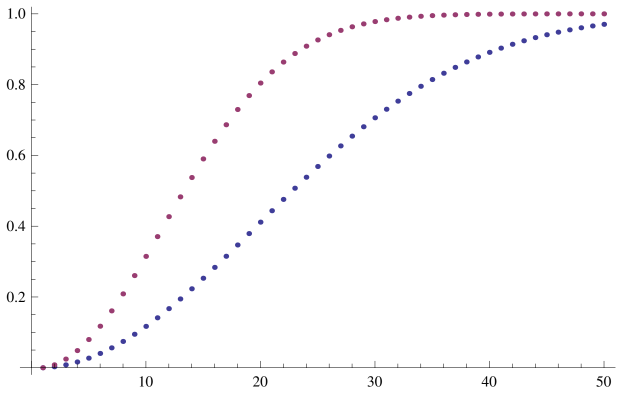
It follows that the number of length-n sequences of non-negative integers with sum S is equal to $\dbinom{S+n-1}{n-1}$.

Back to the birthday problem



$$
\begin{aligned}
P(n \text{ isolated birthdays}) &= \frac{\binom{\text{number of ways to choose}}{n-1 \text{ isolated days}} \times (n-1)!}{(365)^{n-1}} \\
&= \frac{\binom{(365-2n)+n-1}{n-1} \times (n-1)!}{(365)^{n-1}} \\
&= \frac{\binom{364-n}{n-1} \times (n-1)!}{(365)^{n-1}} \\
&= \frac{(364-n)_{n-1}}{(365)^{n-1}}
\end{aligned}
$$

More numbers

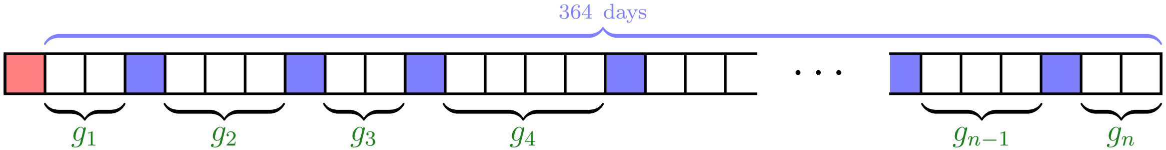


n	P(shared)	P(near miss)
10	0.117	0.314
14	0.223	0.537
15	0.253	0.590
20	0.411	0.804
23	0.507	0.888
24	0.538	0.909
25	0.569	0.926
30	0.706	0.978
35	0.814	0.995
40	0.891	0.999
41	0.903	0.999
45	0.941	1^-
50	0.970	1^-

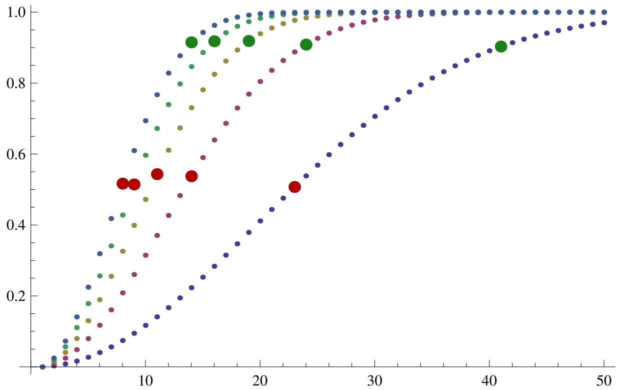
$$P(\text{shared birthday}) = 1 - \frac{(365)_n}{(365)^n}$$

$$P(\text{near miss}) = 1 - \frac{(364-n)_{n-1}}{(365)^{n-1}}$$

A generalization



$$P\left(\begin{array}{c}\text{two birthdays at most}\\ k \text{ days apart}\end{array}\right) = 1 - \frac{(364-kn)_{n-1}}{(365)^{n-1}}$$



	Minimum n for	
k	$P > 0.5$	$P > 0.9$
0	23	41
1	14	24
2	11	19
3	9	16
4	8	14