Monday afternoon, 22nd Sept 2025

Tutorial 1: Probabilities and likelihoods

---

1. The birthday paradox 🎂

A number of k persons meet. Assume that the probability of a person to have his/her birthday is the same for every day of the year. Assume further that the number of days per year is always 365.

a) What is the probability that at least q people have the birthday on January 1st?

Let p be the probability that one person has bday Jan 1st. With 365 days in the year, and each day equally likely, $p = \frac{1}{365}$

For our k people to meet, the the prob of **exactly q** ppl to have this bday is given by the **binomial probability distribution**.

$$\text{Prob}(q \text{ have bday Jan } 1^{st}) = \binom{k}{q} p^q (1-p)^{k-q}$$

Then to find the prob of *at least q ppl* to have this bday, just sum up the probabilities:

$$\text{Prob}(\text{at least } q \text{ ppl have bday Jan } 1^{st}) = \sum_{i=q}^{k} \binom{k}{i} p^i (1-p)^{k-i}$$

b) What is the probability of at least two persons in the meeting have their birthday on the same day?

Consider two groups of events:

- A: None of the k people have the same bday
- B = ¬A: at least 2 ppl share the same bday

Since A and B are two disjoint events that cover the entire sample space Ω, (A+B = Ω), by Kolmogrov's 2nd and 3rd axioms:

P(B) = 1 - P(A)

Plan of attack: We'll calculate P(A) and use that to find what the problem is asking us for, P(B).
Notation: let's denote the events by a k-tuple:

$$E = (X_1, X_2, \ldots, X_k), \quad X_i \in \{1,\ldots,365\}$$

To find a probability of an event E, we can decompose it using the chain rule of probabilities

$$p(E) = p(X_k | X_{k-1}, \ldots, X_1) \;\cdots\; p(X_3 | X_2, X_1) p(X_2 | X_1) p(X_1)$$

A is a composite event, but can still be written as a product of probabilities.
Consider the k people, and let's iterate through the possibilities they have unique bdays:

$$p(A) = p(X_k \neq x_{<k} | x_{<k}) \cdots p(X_3 \neq \{x_1, x_2\} | x_1, x_2)\ p(X_2 | x_1)$$

where $x_{<k} = \{x_1, \ldots, x_{k-1}\}$ (where the last prob is 1 b/c the 1st person will always have a unique bday: $\sum_{x_1} p(X_1 = x_1) = 1$. Solving for these probabilities:

2nd person: $p(X_2 \neq x_1 | x_1) = 1 - p(X_2 = x_1 | x_1) = \dfrac{364}{365}$

3rd person: $p(X_3 \neq \{x_1, x_2\} | x_1, x_2) = \dfrac{363}{365}$

⋮

kth person: $p(X_k \neq \{x_1, x_2, \ldots, x_{k-1}\} | \{x_1, x_2, \ldots x_{k-1}\}) = \dfrac{365 - k + 1}{365}$

Multiplying these probabilities: $p(A) = \prod_{i=1}^{k} \left(\dfrac{365 + 1 - i}{365} \right) = \prod_{i=0}^{k-1} \left(\dfrac{365 - i}{365} \right)$

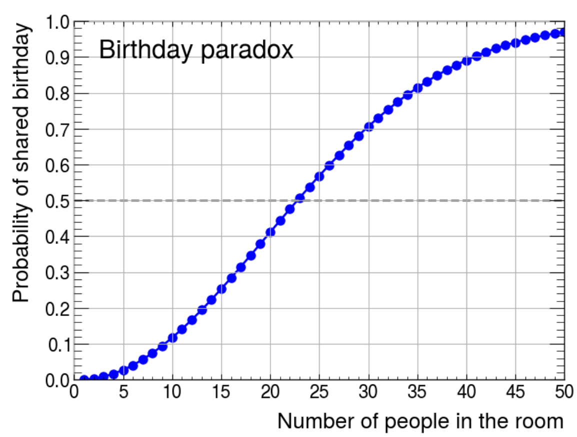
So p(A): prob for all ppl to have independent bdays is

$$p(B) = 1 - p(A) = 1 - \prod_{i=0}^{k-1} \left(\frac{365 - i}{365} \right)$$

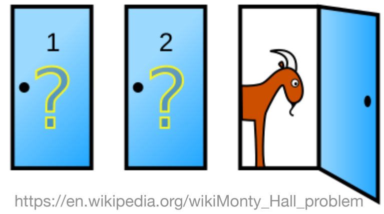
c) For which k is this probability larger than 50%?

Plot the solution from (b) into python (or your fav plotting software), and make the plot !

Once $\boxed{k > 23}$ the prob exceeds 50%.



2. The Monte Hall problem



https://en.wikipedia.org/wikiMonty_Hall_problem

You take part in a gameshow. At one point in the show the host presents you with three doors, each hiding one prize. You get to choose one of the doors and get to keep whatever is behind it. Two of the doors are hiding a goat and one is hiding a sportscar. After you have made your choice the host, who knows which door is hiding the car, opens one of the doors you have not chosen, making sure he is revealing a goat. Now he asks you if you want to stick to your original choice or if you would like to get what is behind the third door. Should you change to the second door (assuming you prefer cars over goats)? Give a formal proof of your answer using Bayes' Theorem.
Hint: Assume the host to play fair and to always reveal a goat behind another door before one chooses.

Let $A \in \{1,2,3\}$ the door the car is behind, where the car is equally likely to be behind any of the 3 doors, $P(A) = 1/3$

Without loss of generality, let's call the door you "pick" door 1.

Then the host will choose a door (not door 1), and again without loss of generality, call the door the host opens door 3, and let B denote this event "host opens door 3"

Goal: Calculate the posterior prob of success if you switch your door, $P(A = 2 \mid B)$.

Soln 1 (with Bayes thm):

Step 1: Enumerate the conditional probabilities $p(B \mid A)$

$p(B \mid A = 1) = 0.5$

^ You have the correct door, the host randomly chose between doors 2 and 3

$p(B \mid A = 2) = 1$

$p(B \mid A = 3) = 0$

^ There was no way the host would have opened door 3 if it had the car

Step 2: Calculate the evidence

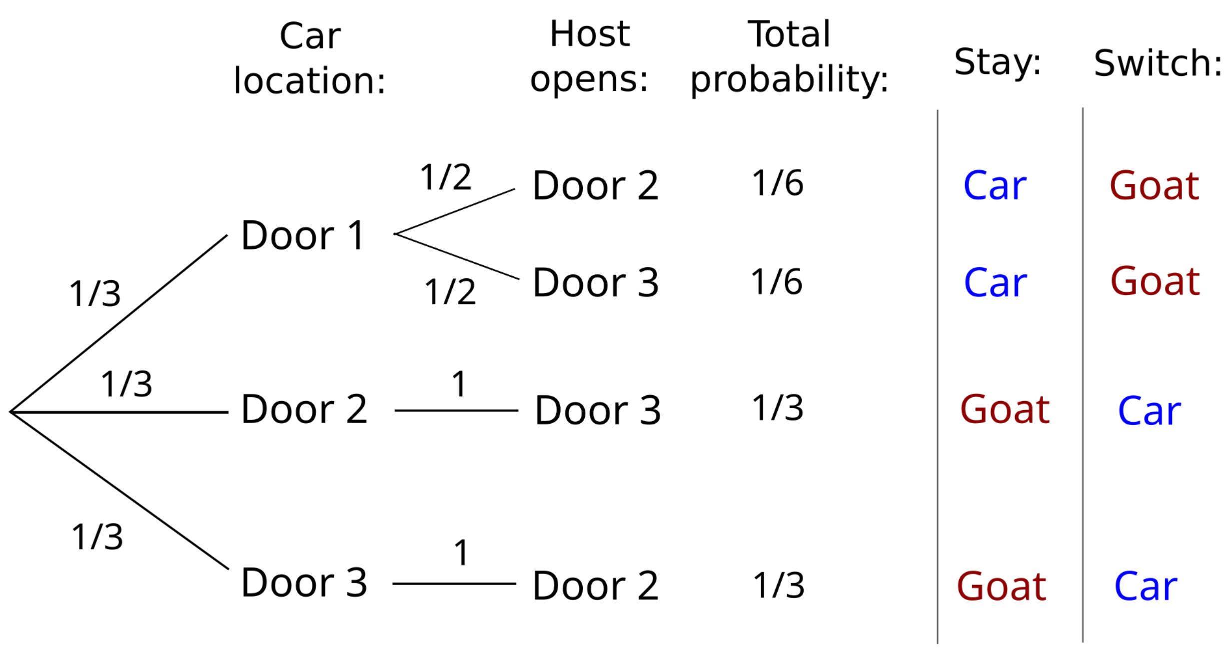
$$p(B) = \sum_a p(A = a, B) = \sum_a p(B \mid a)p(a) = \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2}$$

Step 3: Put the pieces together with Bayes thm

$$p(A=1\,|\,B) = \frac{p(B\,|\,A)p(A)}{p(B)} = \frac{1/3}{1/2} = \frac{2}{3}$$

Soln 2 (Via a probability table):

We can also evaluate all the possibilities via a table



“Tree showing the probability of every possible outcome if the player initially picks Door 1. The sample space consists of exactly four possible outcomes." Source: https://en.wikipedia.org/wiki/Monty_Hall_problem

p(car behind door 2 | host opens door 3)

= prob(car behind door 2 & host opens door 3) / prob(host opens door 3)

$$\frac{1/3 * 1}{1/3 * 1/2 + 1/3 * 1} = \frac{1/3}{1/2} = \frac{2}{3}$$

Soln 3 (Kolmogrov’s axioms of probability):

When you choose a door (say door 1), the probability it has the car behind it is 1/3.

When the host opens door 3 (event B), now we know that door doesn’t have the car!

- The host opening door 3 doesn’t change the probability of the car is behind door 1, $p(A=1\,|\,B) = p(A=1) = 1/3$.
- The probability needs to still sum to 1

$$p(A=1\,|\,B)+p(A=2\,|\,B)+\cancelto{0}{p(A=3\,|\,B)}=1$$

$$\implies p(A=2\,|\,B)=1-p(A=1\,|\,B)=1-\frac{1}{3}=\frac{2}{3}$$

$\therefore$ **Optimal strategy… you should change your door!**

---

3. Poisson probability distribution

Another discrete probability distribtuion that we'll be spending some time with in the class is the Poisson probability distribution:

$$P(k\,|\,\lambda)=\frac{\lambda^k e^{-\lambda}}{k!},\quad k\in\{0,1,2,\ldots\infty\}$$

The following questions give you an opportunity to practice your discrete probability gymnastics, and prove some key properties of this distribution(!)

3a) Prove that the distribution is normalized

$$\sum_{k=0}^{\infty}P(k\,|\,\lambda)=\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}\cdot e^{\lambda}=1\quad\checkmark$$

where we used the definition of exponential we learned in high school maths, $e^x=\sum_{k=0}^{\infty}\frac{x^k}{k!}$.

3b) Prove that the mean is λ

$$\mathbb{E}_P[k]=\sum_{k=0}^{\infty}k\,P(k\,|\,\lambda)=\sum_{k=0}^{\infty}k\frac{\lambda^k e^{-\lambda}}{k!}=\sum_{k=1}^{\infty}k\frac{\lambda^k e^{-\lambda}}{k!}=\sum_{k=1}^{\infty}\frac{\lambda^k e^{-\lambda}}{(k-1)!}$$

where we dropped the first term of the summation which was 0 for $k=0$.
Let $k'=k-1$,

$$\mathbb{E}_P[k]=\sum_{k=1}^{\infty}\frac{\lambda^{k+1}e^{-\lambda}}{(k)!}=\lambda\sum_{k=1}^{\infty}\frac{\lambda^k e^{-\lambda}}{(k)!}=\lambda$$

Note, this ends up being a common trick to solving for the means (and variances) of discrete distributions… try to pull out the extra terms and simplify the sum using the normalization of the probability distribution.

3c) Prove that the variance is λ

Use the handy trick: $\mathrm{Var}(k)=\mathbb{E}[k^2]-\mathbb{E}[k]^2$

And… similar to how we solved in class for the variance of the binomial distribution… we'll use the relation $\mathbb{E}[k(k-1)] = \mathbb{E}[k^2] - \mathbb{E}[k]$ as the "means to the end" of deriving k^2.

$$\mathbb{E}[k(k-1)] = \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=2}^{\infty} k(k-1)\frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=2}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-2)!}$$

where now both the $k = 0$ and $k = 1$ terms don't contribute to the sum.
Let $k' = k - 2$.

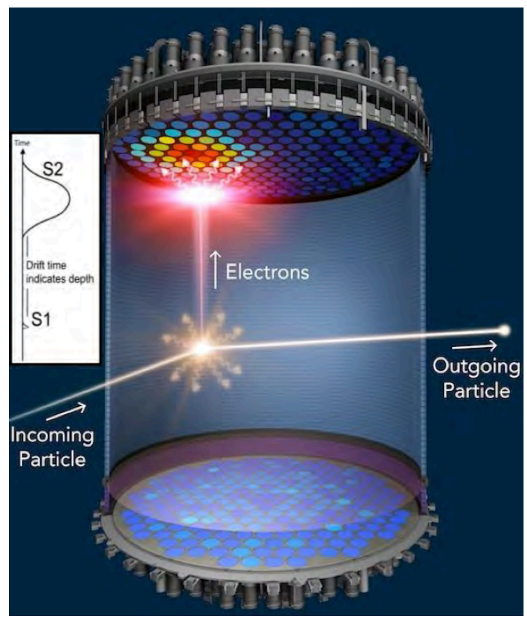
$$\mathbb{E}[k(k-1)] = \sum_{k=0}^{\infty} \frac{\lambda^{k+2} e^{-\lambda}}{k!} = \lambda^2 \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = \lambda^2$$

Great! Let's solve for the variance

$$\mathrm{Var}(k) = \mathbb{E}[k^2] - \mathbb{E}[k]^2 = \mathbb{E}[k(k-1)] + \mathbb{E}[k] - \mathbb{E}[k]^2$$
$$= \lambda^2 + \lambda - \lambda^2 = \lambda \quad \checkmark$$

3d) Imagine you're a pheno PhD student building dark matter models.You just worked through an arduous calculation for your new model, and arrirve at a prediction: you expect $\lambda = 1.5$ dark matter events per year to be observed by the currently running Lux Zeplin (LZ) experiment.

They're already taking data… so you have a prediction to compare against! With some trepedation (and gratitude to you your past self for enrolling in the ODSL Stats block course!)… you set out to calculate some probabilities.



In the LZ experiment, a dark matter particle recoils off a Xenon nucleus and give a light (S1) and charge (S2) signal recorded by photomultiplier tubes at the top of the detector. https://lz.slac.stanford.edu/our-research/lux-zeplin-research

If your model is true, what is the probability that LZ sees 6 or more events in a year?

Use the Poisson distribution to calculate the probabilities!

$$P(k \geq 6 \,|\, \lambda) = 1 - P(k \leq 5 \,|\, \lambda) = 1 - \sum_{k=0}^{5} \frac{\lambda^k e^{-\lambda}}{k!} = 1 - .9955 = 0.45$$

If your model is true, what's the prob LZ didn't see any events in a year?

$$P(k = 0 \,|\, \lambda) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-1.5} = 0.223$$

All hope isn't lost… your model might still be true and LZ just needs to keep taking more data!