Tutorial: Bayesian inference tools

Today

1. A simple analytical example

2. A numerical example using PyMC (notebook)

3. A problem in physics (notebook)

Tomorrow

- MCMC algorithm

Part 1: an example where things are simple

“THE SENDING RATE”

An innovative parameter I have invented to rate climbing performances of my friends and prove them that I m a better climber than them

$$\theta_X = \frac{\text{number of routes I have sent in year X}}{\text{number of routes I have tried in a year X}}$$

So far in 2025 I have tried 3 routes and I have done 2 of them.
Therefore so far $\theta_{2025} \sim 67\,\%$. **It s pretty good!**

However… it is only July, I m going on holidays and I will be climbing much more. So I’d like to know what s my TRUE sending rate. I could have gotten 67% so far in 2025 just out of luck

So what the truth underlying this 67%?

Let s treat θ as a stochastic variable.

A good prior should describe the knowledge about my sending rate I have from the previous years. We will use a **beta distribution**

$$p\left(\theta; \alpha, \beta\right) = Beta\left(\theta; \alpha, \beta\right) = \frac{\Gamma\left(\alpha + \beta\right)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}\left(1-\theta\right)^{\beta-1}$$

- Probability distribution defined on the interval [0,1]
- Model for random variables representing percentages and proportions
- Sport analytics: used to model player performance metrics like shooting accuracy or goal-scoring rates

I can check how many routes I have tried vs how many I have sent since 2022

	N (tried)	**Y (sent)**
2022	20	10
2023	15	5
2024	15	4

$$p\left(\theta;\alpha,\beta\right) = Beta\left(\theta;\alpha,\beta\right)$$
$$= \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}\left(1-\theta\right)^{\beta-1}$$

I can fit these data on a Beta distribution, determining its fixed parameters which best describe my climbing rate from the previous years.

From the data in this table I found $\alpha = 9.02$, $\beta = 15.5$

$$p\left(\theta|D\right) = \frac{p(D|\theta)p\left(\theta\right)}{p\left(D\right)}$$

Likelihood: probability of 2 successes over 3 tries given the sending rate θ.
We can use a binomial distribution

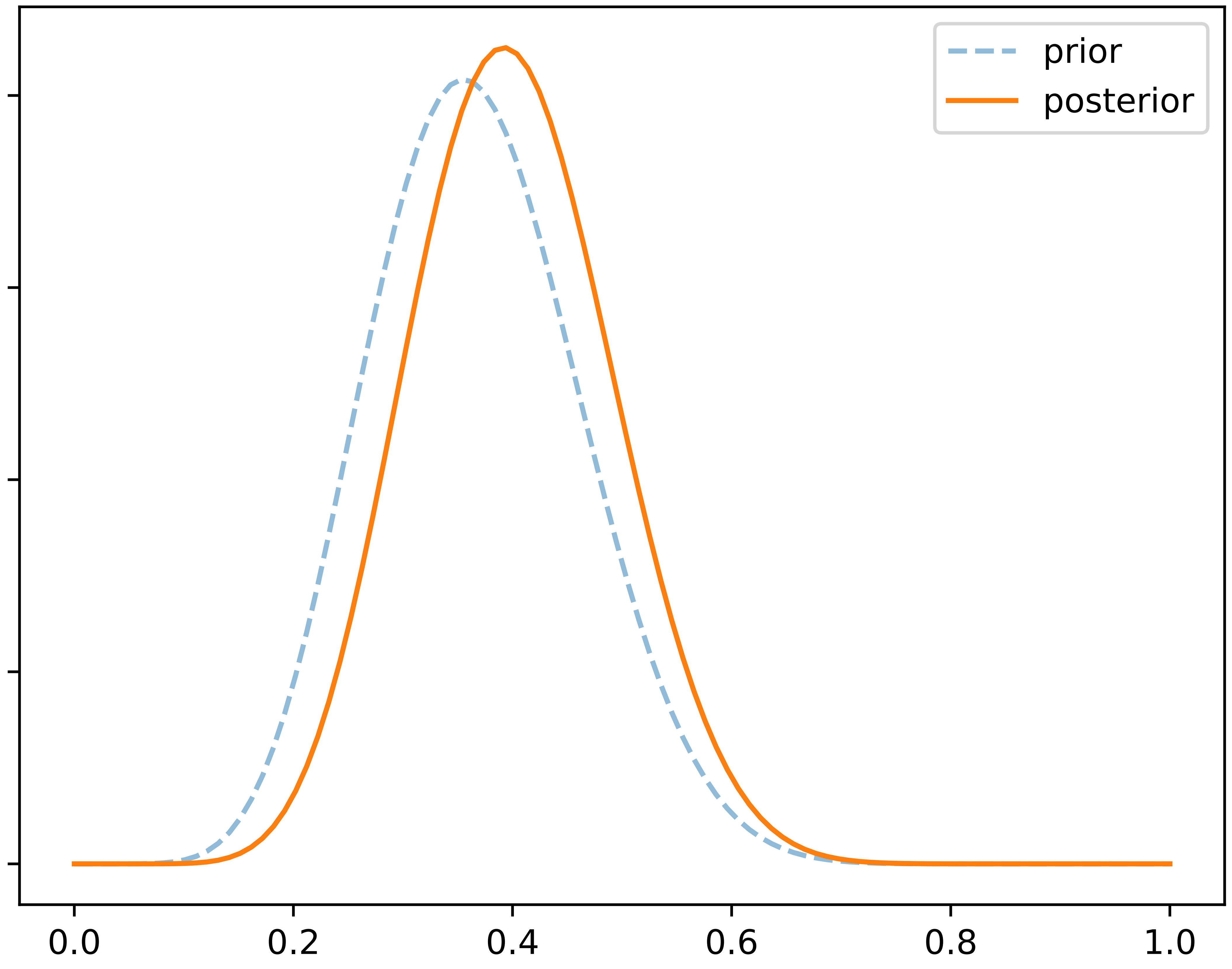
$$p\left(D|\theta\right) = \frac{n!}{y!\left(n-y\right)!}\theta^{y}\left(1-\theta\right)^{n-y} = \text{Binomial}\left(S_n = y\right)$$

Now, it turns out that a Beta times a Binomial is again a Beta.

So we can compute the full posterior analytically

$$p(D \mid \theta)p(\theta) = \text{Binomial}\left(S_n = y\right) Beta\left(\theta; \alpha, \beta\right)$$

$$= Beta\left(\alpha + y, \beta + n - y\right)$$



Given

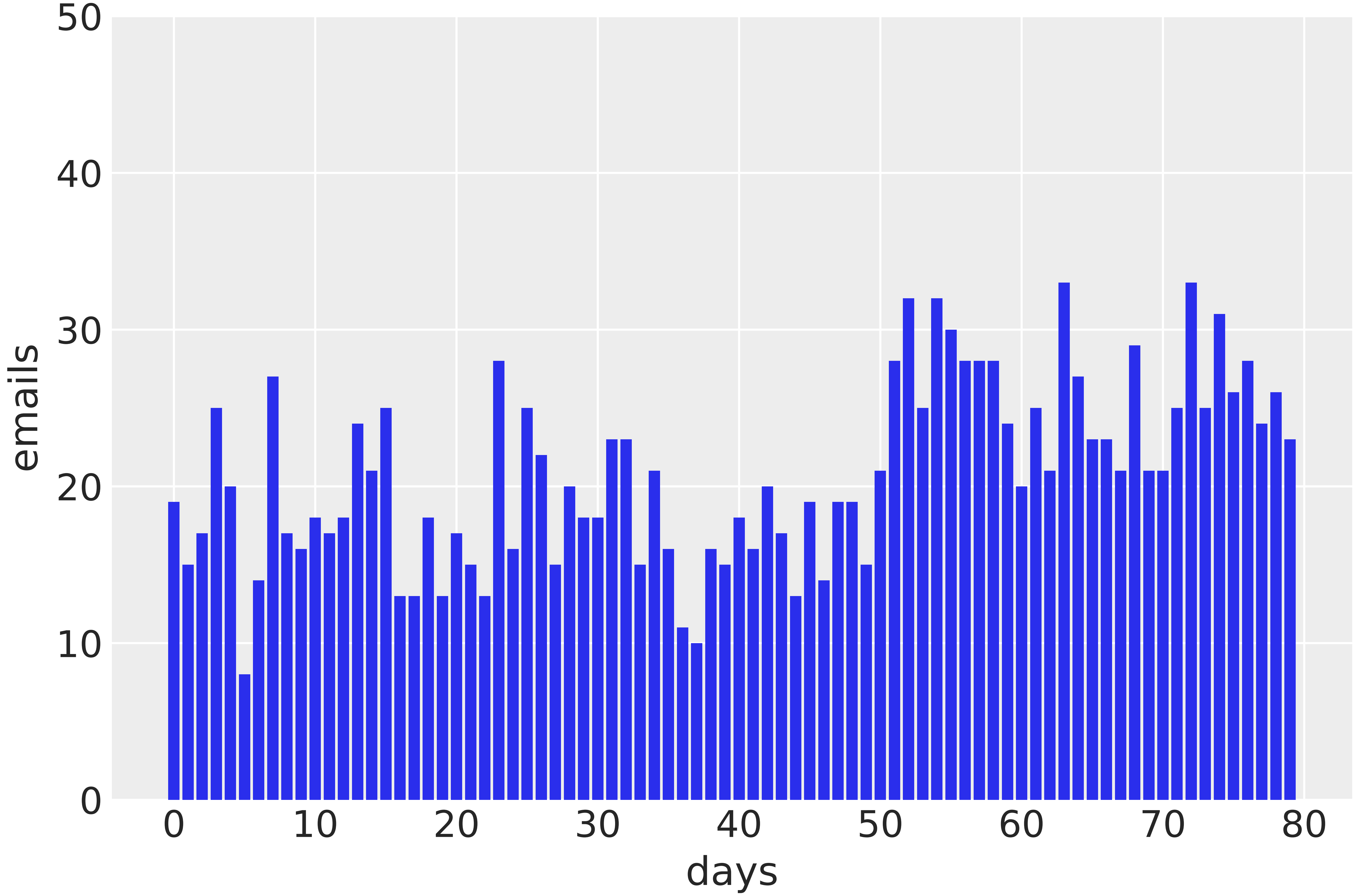
- the data I have observed so far in 2025
- my prior knowledge about my sending rate

I see that probably my real sending rate
has slightly improved, however its probability
distribution peaks at ~ 40%
(quite far from the 67% I ve seen so far in 2025)

:(

Part 2: an example where things are not simple

We want to build a model describing the number of emails I got in the last 80 days



Let s build the model: we can describe the probability distribution for the emails
We get each day with a Poisson distribution

$$p\left(k;\lambda\right) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- Discrete probability distribution expressing the probability of a given number of events in a fixed interval of time (like number of email in a day)

- $\lambda > 0$ represent the expectation value

However, it looks as if at some point we start getting more emails.
Then let s use two distributions with different λ

$$p\left(k;\lambda\right) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{with} \quad \lambda = \begin{cases} \lambda_1 & \text{if} \quad t < \tau \\ \lambda_2 & \text{if} \quad t > \tau \end{cases}$$

So our model has 3 parameters λ_1 , λ_2 , τ that we want to infer from the observed data

Call C_i the email count of i-th day.

Our data is the set of observations $D = \{C_i\,, \quad i = 0\,, \ldots, 80\}$

$$p\left(D\,|\,\lambda_1, \lambda_2, \tau\right) = \prod_i Poisson\left(C_i; \lambda_1, \lambda_2, \tau\right)$$

Define a prior for the unknown variables

$$\lambda_i \sim Exp\left(\alpha\right)$$

$$\tau \sim \text{DiscreteUniform}\left(0{,}80\right)$$

$\alpha = 1/\text{mean}_i\left(C_i\right)$

What s a sensible value for the parameter α ? We could also put a prior on it, but let s keep it simple

$$p\left(\lambda_1, \lambda_2, \tau \,|\, D\right) \propto \prod_i Poisson\left(C_i; \lambda_1, \lambda_2, \tau\right) p\left(\lambda_1, \lambda_2, \tau\right)$$

We need to sample from the posterior : MCMC.
Let s see this on a notebook using PyMC