Introduction to Bayesian statistics

Rosana Zenil-Ferguson
Assistant Professor. University of Kentucky
MOLE at Woods Hole workshop 2026

What is Bayesian statistics?

- Branch of statistics interested in measuring uncertain events and providing probabilistic support to hypotheses.
- Goal: Providing measures of uncertainty and not a point estimate

Two important difference with other statistical inferences

1. Probability is measured as a bet

2. Parameters are unknown but uncertain = random variables with probability distributions

When do we choose Bayesian statistics?

In Biology we frequently choose Bayesian stats because it is computationally powerful and easier to use

The main algorithm that made possible Bayesian stats applications is the MCMC

A good reason to choose Bayesian is to make sure we estimate uncertainty and represent it with a probability distribution

Conditional probability reminder

Given two events A and B the conditional probability of A **given** B

is: $P(A|B) = \frac{P(AB)}{P(B)}$

and if we apply twice the theorem we get

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$P(A|B)$ the posterior distribution is the goal of any Bayesian inference

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

POSTERIOR PROB DISTRIBUTION

We will decompose the posterior distribution

1. P(A) = Prior distribution
2. $P(B|A)$ = Likelihood
3. P(B) = Marginal probability

The experiment: Love is Blind Mexico

We are binge watching this Netflix show: "Love is blind". We are interested in understanding what makes a participant "charming" and potentially "the marrying type".

The experiment: Silvia has a series of blind dates until she finds "the love of her life". However, her personality and charm shapes the outcome of these dates.

Goal: To quantify whether Silvia is still charming after watching the dates.

Will she fall in love with her forever?

Final goal: $P(A|B)$

Is Silvia still charming after watching the show?

$P(A|B)$= The probability that Silvia is charming after observing the outcome of the dates

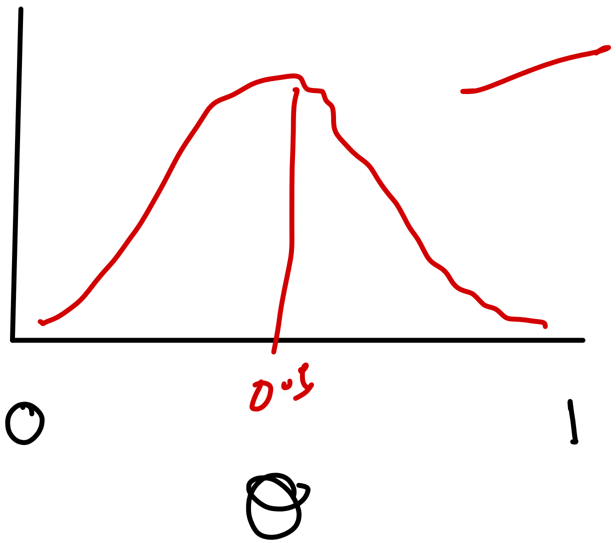
Prior Distribution

As avid reality show watchers, we can define from the beginning our judgment about Silvia's charm

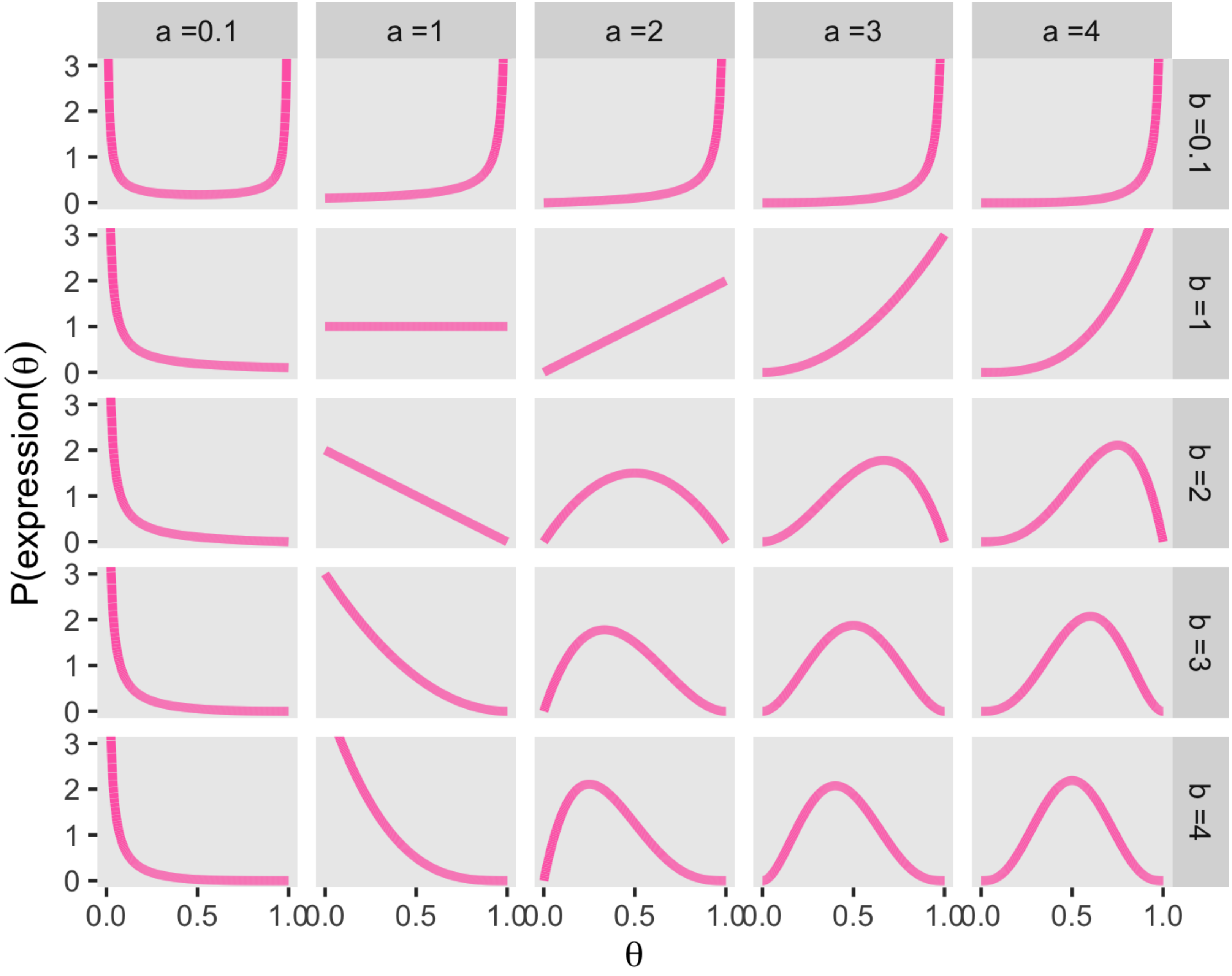
θ = % of Charm of Silvia

θ takes values between $(0,1)$

Subjective belief



Prior distributions are a necessary part of Bayesian stats, they reflect our knowledge our beliefs about the world before we start the experiment. Prior distributions are subjective, and they can be informative or not.



IMPORTANT

In Bayesian statistics parameters are unknown but uncertain and always need a probability distribution

$$P(\theta) \sim \text{Beta}(\alpha, \beta)$$

$$P(\theta) = \cancel{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

ignore because
it does not have θ

Data: Binge watching the series

- Experiment: Silvia has three blind dates with the same person.

$$n = 3$$

Y = Number of dates that went well

$y = 0, 1, 2, 3$

Distribución Binomial

$P(y \mid \theta) \sim \text{Binomial}(3, \theta)$

$=$

$n = 3 \quad p = \theta$
in your magical chart)

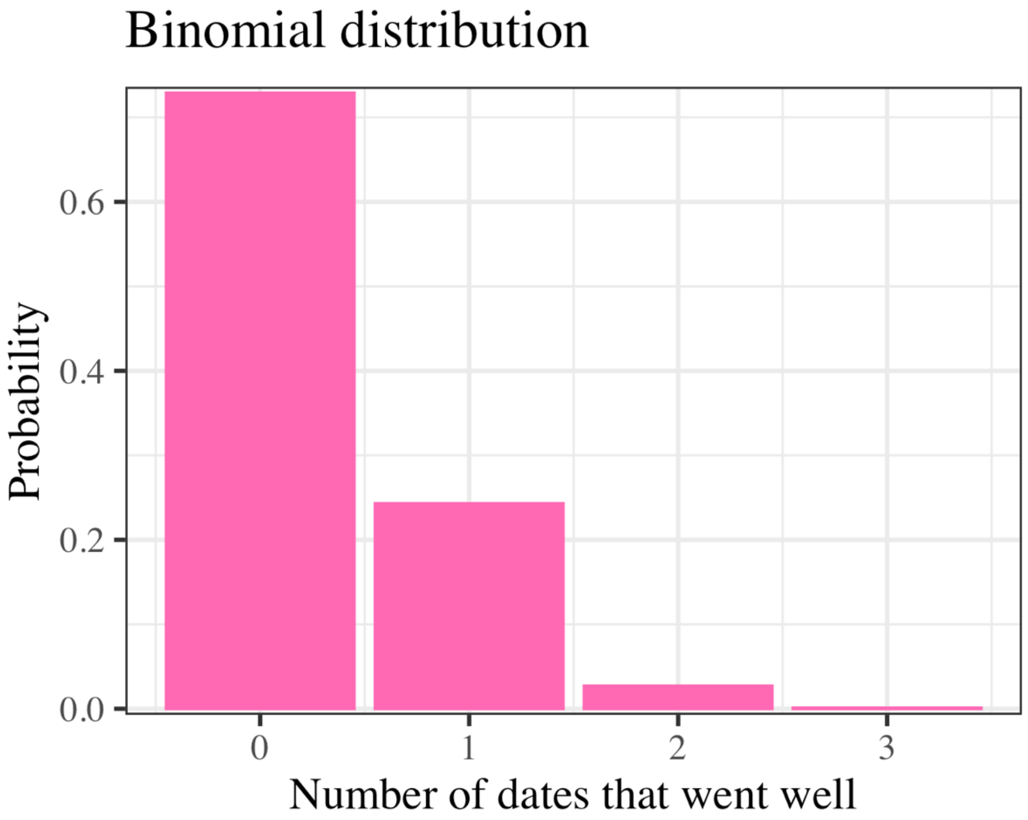
$$p(y \mid \theta) = \binom{3}{y} \theta^{y} (1-\theta)^{3-y}$$

$\boxed{\theta = 0.1}$

$y = 1$

$\cancel{\binom{3}{1}}$ → safe to ignore because it doesn't have θ

combination of 1 out of 3



The likelihood function for 3 dates with the same person

Two dates out of three went well

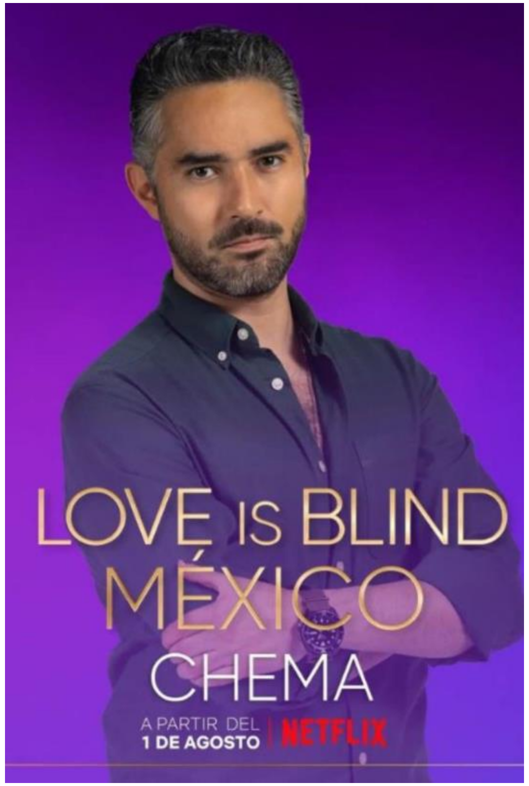
$$P(Y = 2|\theta) = \binom{3}{2}\theta^2(1-\theta)^1$$

If $\theta = 0.1$

then

$$P(Y = 2|\theta = 0.1) =$$

$$\binom{3}{2}(0.1)^2(1-\theta)^1 = 0.0027$$



Exercise: Likelihood of dates with two people each with three dates

$$y_1 = 2 \quad y_2 = 1$$

and $\theta = 0.1$

$$P(y=2 \mid \theta = 0.1)$$
$$\times P(y=1 \mid \theta = 0.1)$$

The likelihood in general

P(Data|Model)

$$P(y_1, y_2|\theta) = \prod_{y_i=1}^{2} P\,(Y = y_i|\theta) \approx \theta^3(1-\theta)^3$$

both men are independent
so we can calculate their
product

Bayesian Statistics: Posterior distribution

In our new language of random variables

$$P(\theta|\boldsymbol{Y}) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

The posterior distribution is proportional to the likelihood times the prior

proportional

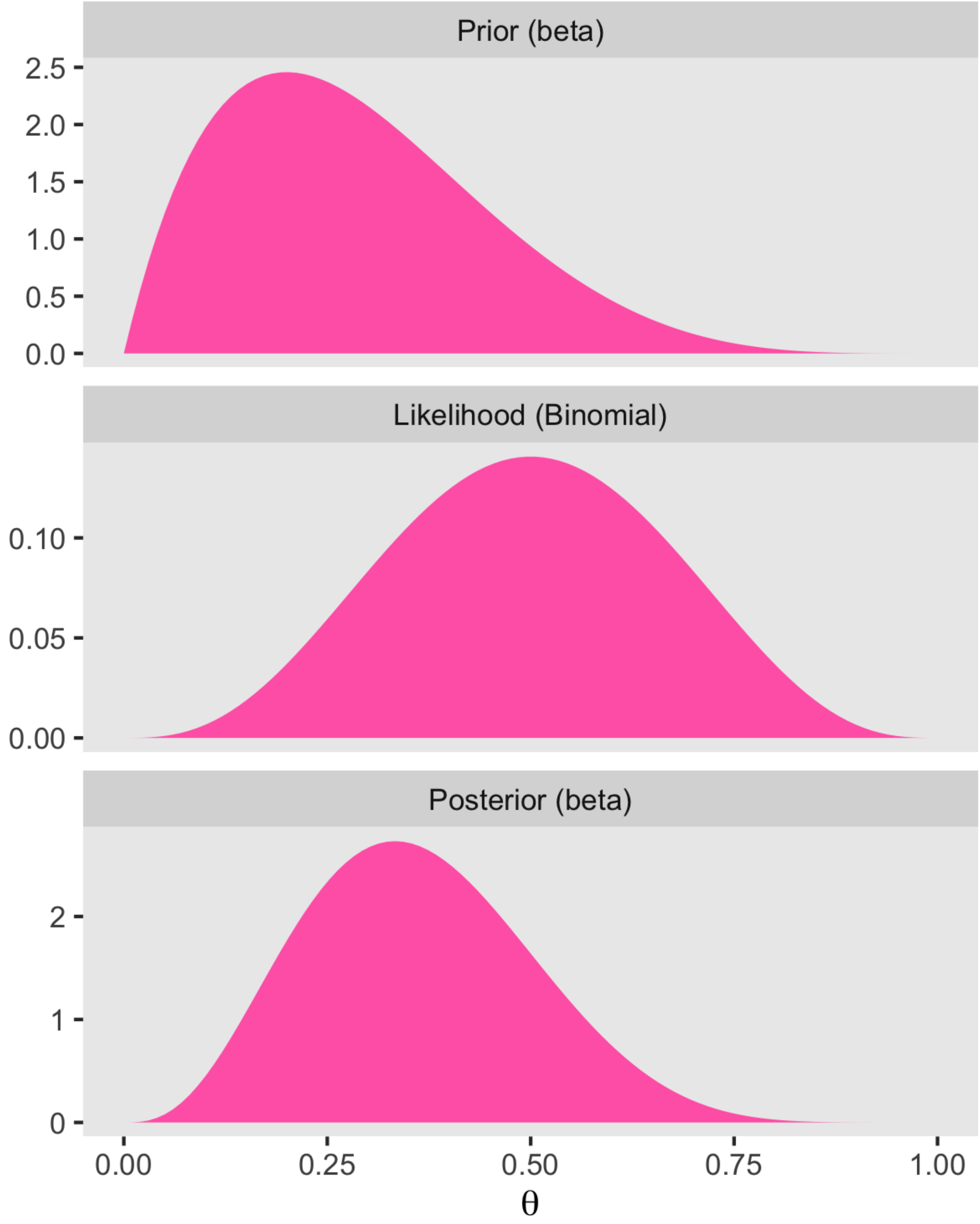
$$P(\theta|\boldsymbol{Y}) \propto P(\mathbf{Y}|\theta) \times P(\theta)$$

prior distribution

likelihood

$$P(\theta|y) \propto \theta^{3}(1-\theta)^{3} \times \theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^{3+\alpha-1}(1-\theta)^{3+\beta-1}$$

Posterior distribution often called an update



Posterior distributions are the combination of our prior beliefs and the likelihood of the data.