

# A Brief Intro to Bayesian Inference



Thomas Bayes (c. 1701 – 1761)


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

# Tossing a Coin, Again

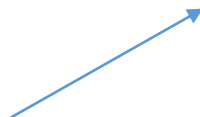
- Suppose the coin came up Heads 65 times and Tails 35 times. Is the coin fair?
- Model:  $P(\text{heads}) = \theta$
- Log-likelihood:  $\log P(\text{data}|\theta) = 65 \log \theta + 35 \log(1 - \theta)$ 
  - Maximized at  $\theta = .65$
- But would you conclude that the coin really is not fair?

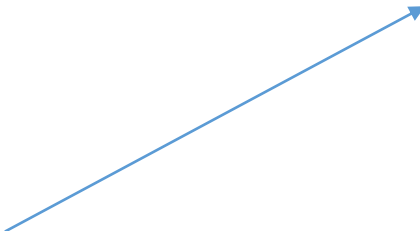
# Prior Distributions

- We can encode our beliefs about what the values of the parameters could be using  $P(\theta)$
- Using Bayes' rule, we have

$$P(\theta = \theta_0 | \text{data}) = \frac{P(\theta = \theta_0, \text{data})}{P(\text{data})} = \frac{P(\text{data} | \theta = \theta_0) P(\theta = \theta_0)}{P(\text{data})}$$


The event that we observed the data



$$= \sum_{\theta_1} P(\text{data} | \theta = \theta_1) P(\theta = \theta_1)$$


# A philosophical aside

- We used notational sleight of hand to sneak in the idea that  $\theta$  is a random variable
  - It could be, if we picked a coin at random from a bunch of coins with different  $\theta$ s
  - But if  $\theta$  is the free-fall acceleration on earth (i.e.,  $g = 9.8$ ), it is definitely not random!
- Discussion

# Maximum a-posteriori (MAP)

- Maximize the *posterior probability* of the parameter:

$$\operatorname{argmax}_{\theta_0} \frac{P(\text{data}|\theta = \theta_0)P(\theta = \theta_0)}{P(\text{data})}$$

$$= \operatorname{argmax}_{\theta_0} P(\text{data}|\theta = \theta_0)P(\theta = \theta_0)$$

$$= \operatorname{argmax}_{\theta_0} \log P(\text{data}|\theta = \theta_0) + \log P(\theta = \theta_0)$$

- The posterior of probability is the product of the prior and the data likelihood
- Represents the *updated* belief about the parameter, given the observed data

## Aside: Bayesian Inference is a Powerful Idea

- You can think about anything like that. You have your prior belief  $P(\theta)$ , and you observe some new data. Now your belief about  $\theta$  *must be* proportional to  $P(\theta)P(data|\theta)$ 
  - But only if you are 100% sure that the likelihood function is correct!
  - Recall that the likelihood function is your model of the world – it represents knowledge about how the data is generated for given values of  $\theta$
  - Where do you get your original prior beliefs anyway?
- Arguably, makes more sense than Maximum Likelihood

# Back to the Coin

- In R, compute the Maximum A Posteriori estimate

# Unicorns

- (time permitting)