Hoff Ch. 1. Introduction

sampling distribution
aka likelihood

prior

$$p(\theta \mid y) = \frac{p(y \mid \theta)\, p(\theta)}{p(y)}$$

posterior distribution

normalizing constant

Sample space

- $\mathcal{Y}$ = sample space
- Y a random variable taking values in $\mathcal{Y}$
- y an observation of Y

Ex. A single coin flip.

- $\mathcal{Y} = \{H, T\}$.

- Y the result of a generic flip, whose outcome is unknown
- Suppose we flip the coin and get heads. Then $y = H$.

Ex. n coin flips

- $\mathcal{Y} = \{H, T\}^n = \{(y_1, \dots, y_n) : y_i \in \{H, T\}\}$

- $n = 3, \quad y = (H, T, T).$

Parameter space

- Θ = parameter space
- $\theta \in \Theta$ a parameter

Ex. Coin flip

- θ = probability of a head (H)

  e.g. $\theta = \frac{1}{2}$ represents a fair coin

- $\Theta = [0,1] = \{\theta \in \mathbb{R} : 0 \le \theta \le 1\}$.
- θ determines the distribution of coin flip Y, i.e.,

$$Y \sim \text{Bernoulli}(\theta)$$

- Let $H=1$ and $T=0$. Then

$$p(y|\theta) = \mathbb{P}(Y=y|\theta) = \begin{cases} \theta & \text{if } y=1 \text{ (heads)} \\ 1-\theta & \text{if } y=0 \text{ (tails)} \end{cases}$$

$$= \theta^{y}(1-\theta)^{1-y}$$

Bayesian analysis of a coin flip

- $p(y|\theta) = \theta^{y}(1-\theta)^{1-y}$
- $p(\theta) = 1$
- $p(y) = \int_0^1 p(y,\theta)\,d\theta = \int_0^1 p(y|\theta)\,d\theta$

$$= \int_0^1 \theta^{y}(1-\theta)^{1-y}\,d\theta = \begin{cases} \int_0^1 \theta\,d\theta & \text{if } y=1 \\ \int_0^1 1-\theta\,d\theta & \text{if } y=0 \end{cases}$$

$$= \frac{1}{2}.$$

- Put it together:

$$p(\theta|y) = \frac{p(y|\theta)\,p(\theta)}{p(y)} = 2\theta^{y}(1-\theta)^{1-y}.$$

Posterior mean

$$\mathbb{E}_{\theta|y}[\theta] = \int_{\Theta} \theta \, p(\theta|y) \, d\theta$$

Coin flip:

$$\mathbb{E}_{\theta|y}[\theta] = \int_0^1 \theta \left(2\theta^{y}(1-\theta)^{1-y} \right) d\theta$$

$$= 2\int_0^1 \theta^{1+y}(1-\theta)^{1-y} \, d\theta$$

$$= \begin{cases} 2\int_0^1 \theta^2 \, d\theta = \frac{2}{3} & \text{if } y = 1 \\ 2\int_0^1 \theta(1-\theta) \, d\theta = \frac{1}{3} & \text{if } y = 0. \end{cases}$$

<u>Summary:</u> If we assume a uniform prior ($p(\theta) = 1$), then the expected value of θ is $\frac{2}{3}$ if we observe a head ($y=1$) and $\frac{1}{3}$ if we observe a tail ($y=0$).

A different prior.

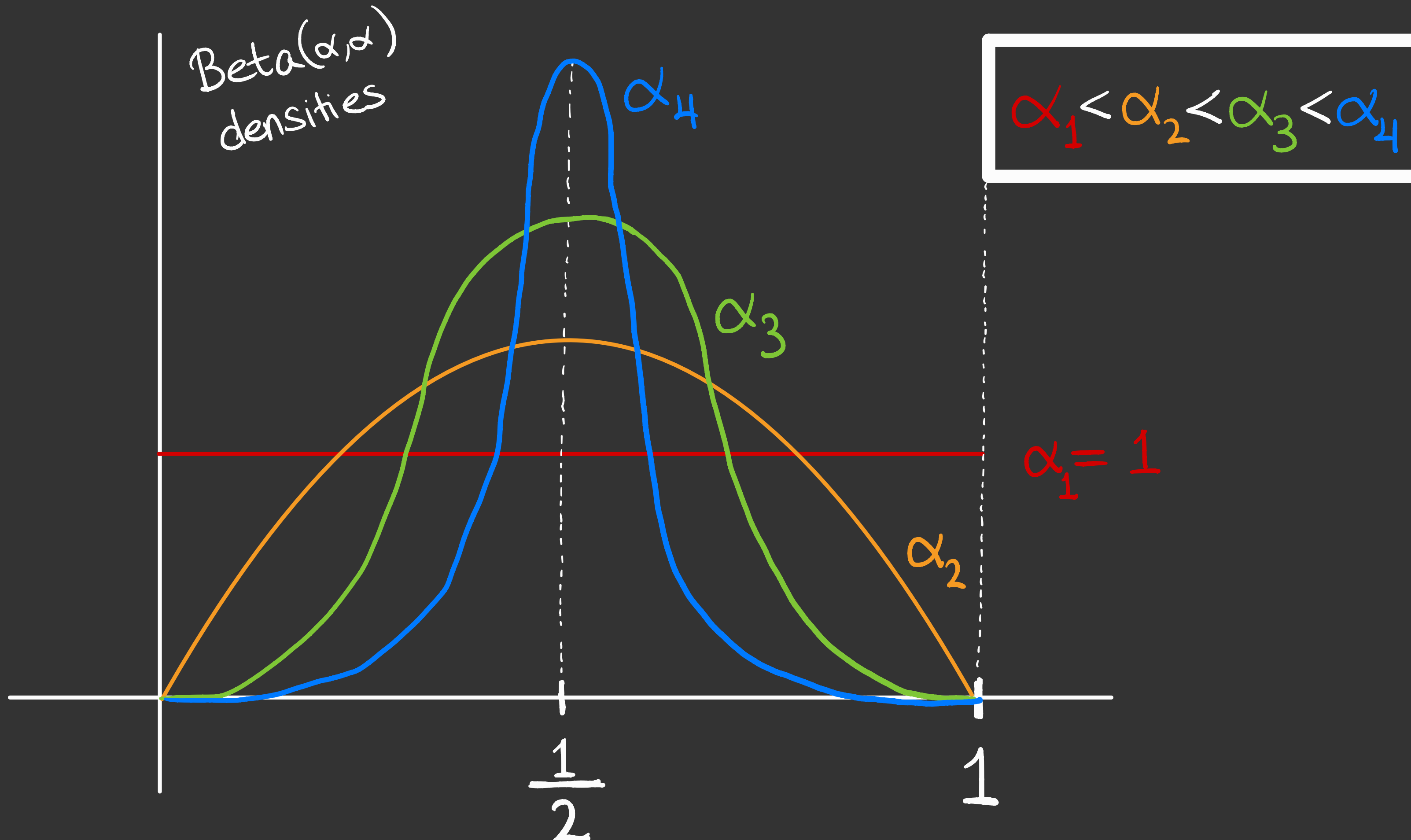
Suppose instead that $p(\theta)$ is $\text{Beta}(\alpha,\alpha)$:

$$p(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\alpha-1}}{B(\alpha)}.$$

$\boxed{\text{* Note that } \alpha = 1 \text{ corresponds to uniform prior, i.e. } p(\theta) = 1.}$

Then

$$p(\theta|y) = \frac{2}{B(\alpha)} \underbrace{\left[\theta^{y}(1-\theta)^{1-y}\right]}_{\text{Likelihood}} \underbrace{\left[\theta^{\alpha-1}(1-\theta)^{\alpha-1}\right]}_{\text{Prior}}$$



Beta(α,α) densities
α_4
α_3
α_2
$\alpha_1 = 1$
$\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$
$\frac{1}{2}$
1

$$= \frac{2}{B(\alpha)} \theta^{y+\alpha-1} (1-\theta)^{\alpha-y+1-1}$$

$$\propto \theta^{y+\alpha-1} (1-\theta)^{\alpha-y+1-1}$$

$$\sim \text{Beta}(y+\alpha,\ \alpha-y+1).$$

$\Rightarrow p(\theta|y)$ is a $\text{Beta}(y+\alpha, \alpha-y+1)$ distribution

Mean of $\text{Beta}(\alpha, \beta)$ is $\dfrac{\alpha}{\alpha+\beta}$

$\Rightarrow$ posterior mean is

$$\mathbb{E}_{\theta|y}[\theta] = \frac{y+\alpha}{2\alpha+1}.$$

- $\alpha = 1 \Rightarrow \mathbb{E}_{\theta|y}[\theta] = \begin{cases} \frac{2}{3} & \text{if } y=1 \\ \frac{1}{3} & \text{if } y=0 \end{cases}$

- $\alpha = 10 \Rightarrow \mathbb{E}_{\theta|y}[\theta] = \begin{cases} \frac{11}{21} \approx 0.52 & \text{if } y = 1 \\ \frac{10}{21} \approx 0.48 & \text{if } y = 0 \end{cases}$

- $\lim_{\alpha \to \infty} \mathbb{E}_{\theta|y}[\theta] = \lim_{\alpha \to \infty} \frac{y + \alpha}{2\alpha + 1} = \frac{1}{2}.$

End Ch. 1