Probability and Statistics

Inference. Bayesian
Frequentist

Module 1: Topic 6

Statistics: Bayesian inference

Konduri Aditya
Department of Computational and Data Sciences
Indian Institute of Science, Bengaluru

Department of Computational and Data Sciences

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

Outline

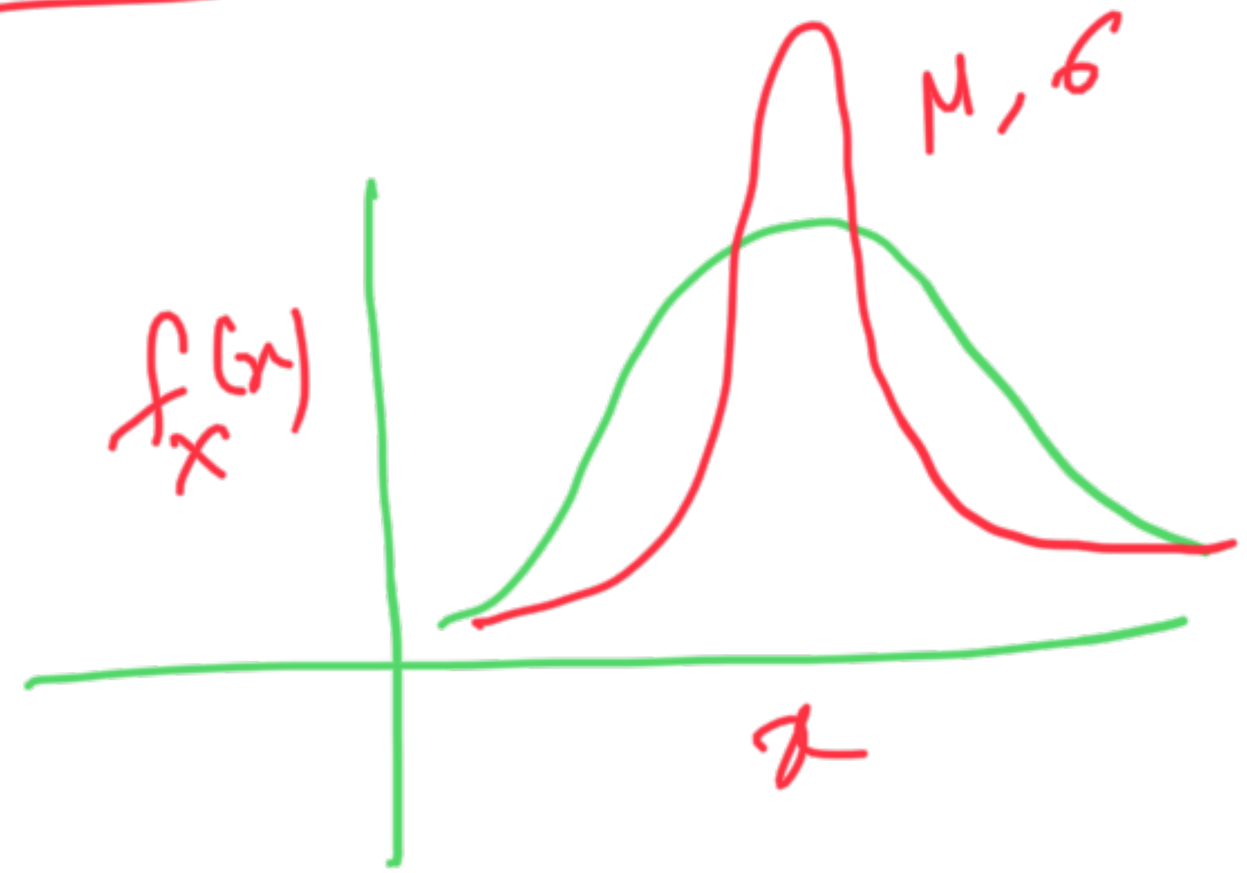
- Parameter estimation
- Maximum likelihood estimate (MLE)
- Recap: Bayes' rule
- Bayesian inference: priors/posteriors
- Example

Parameter estimation

- Suppose there is some data, we are interesting in finding out which distribution is it drawn from

  n, p μ, σ λ

  - Example: Binomial, normal, exponential
- Identifying the distribution also involves finding out its parameters



Parameter estimation

- *Example:* Suppose we want to know the percentage **p** of people for whom an image with an animal looks like a dog.
- Experiment: Ask **n** random people to see the image
- Model:

$X_i \sim Bernoulli(p)$ is whether the i^{th} person says it is a dog

- Data: $x_1, \ldots, x_n$ are the results of the experiment
- Inference: estimate **p** from the data

Probability of that image containing a dog.

Parameters of interest

- *Example:* You ask 100 people to look at the image and 55 say it has a dog. Use this data to estimate **p** the fraction of all people for whom it is a dog image
- So, **p** is the parameter of interest

Likelihood

- For a given value of **p** the probability of getting 55 'successes' is the binomial probability

$$P(55dogs\,|\,p) = {}^{100}C_{55}p^{55}(1-p)^{45}$$

$n=100$
$k=55$

Likelihood: $P(data\,|\,p) = {}^{100}C_{55}p^{55}(1-p)^{45}$

- The likelihood takes the data as fixed and computes the probability of the data for a given **p**

derivative w.r.t p : - Apply chain rule.

$${}^{100}C_{55}\left(55\,p^{54}(1-p)^{45} + p^{55}\,45\,(1-p)^{44}(-1)\right)$$

Maximum likelihood estimate (MLE)

- The maximum likelihood estimate (MLE) is a way to estimate the value of ***a parameter of interest***
- The MLE is the value of **p** that maximises the likelihood

$\frac{dg}{dx} = 0$

How do we find **p**? Use calculus

$$\frac{dP(data\,|\,p)}{dp} = {}^{100}C_{55}(55p^{54}(1-p)^{45} - 45p^{55}(1-p)^{44}) = 0$$

This gives the MLE of $\hat{p} = \frac{55}{100}$

Point estimate for the unknown parameter "p"
↓
Probability of having dog

Log likelihood

- A log function turns multiplication into addition it is often convenient to use the log of the likelihood function
- Log likelihood $= ln(likelihood) = ln(P(data \mid p))$
- Example:

Log likelihood $= ln\left({}^{100}C_{55}\right) + 55ln(p) - 45ln(1-p)$

Un kown.

Problem: coin toss

- A coin is taken from a box containing three coins, which give heads with probability p = 1/3, 1/2, and 2/3. The mystery coin is tossed 80 times, resulting in 49 heads and 31 tails.
- (a) What is the likelihood of this data for each type of coin? Which coin gives the maximum likelihood?
- (b) Now suppose that we have a single coin with unknown probability p of landing heads. Find the likelihood and log likelihood functions given the same data. What is the maximum likelihood estimate for p?

Problem: coin toss

$K=49$ $\quad n=80$

- Data D is 49 heads in 80 tosses
- We have three hypotheses: the coin has probability P=1/3, p=1/2, p=2/3.
- Likelihood function $P(Data\,|\,p)$:

condition: we assume coin to be picked from the box.

$$P(D\,|\,p=1/3) = {}^{80}C_{49}\left(\frac{1}{3}\right)^{49}\left(\frac{2}{3}\right)^{31} = \boxed{6.24e-7}$$

6.24×10^{-7}

$$P(D\,|\,p=1/2) = {}^{80}C_{49}\left(\frac{1}{2}\right)^{49}\left(\frac{1}{2}\right)^{31} = 0.024$$

$$P(D\,|\,p=2/3) = {}^{80}C_{49}\left(\frac{2}{3}\right)^{49}\left(\frac{1}{3}\right)^{31} = 0.082$$

MLE is when p=2/3

Problem: coin toss

- Our hypotheses now allow p to be any value between 0 and 1. So our likelihood function is $P(D\,|\,p = \cancel{1/3}) = {}^{80}C_{49}p^{49}(1-p)^{31}$
- To compute the maximum likelihood over all p, we set the derivative of the log likelihood to 0 and solve for p:

$$\frac{dP(D\,|\,p)}{dp} = \frac{d}{dp} ln\left({}^{80}C_{49}\right) + 49ln(p) - 31ln(1-p) = 0$$

- So our MLE is $\hat{p} = \frac{49}{80}$

→ Likelihood function
↓
→ MLE ($\hat{p}$)
↓
Is a single value. → Point estimate.

Recap: Bayes' rule

- Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space, and assume that $P(A_i) > 0$ for all i. Then, for any event B such that $P(B) > 0$, we have

$$\boxed{P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(B)}} = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + \ldots + P(A_n)P(B \mid A_n)}$$

Bayes' rule: example

- A test of a certain rare disease is assumed to be correct 95% times.
- A random person drawn from a certain population has probability 0.001 of having the disease.
- Given that the person just tested positive, what is the probability of having the disease?
- A: event that the person has disease
- B: event that the test results are positive
- $$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$

→ Probability of the person having the disease given that the test result is positive.

Continuous Bayes' rule

- We represent an unobserved phenomenon by a random variable X with PDF f_X and we make a noisy measurement Y, which is modelled in terms of a conditional PDF $f_{Y|X}$. Once the value of Y is measured, what information does it provide on the unknown value of X?

Prior

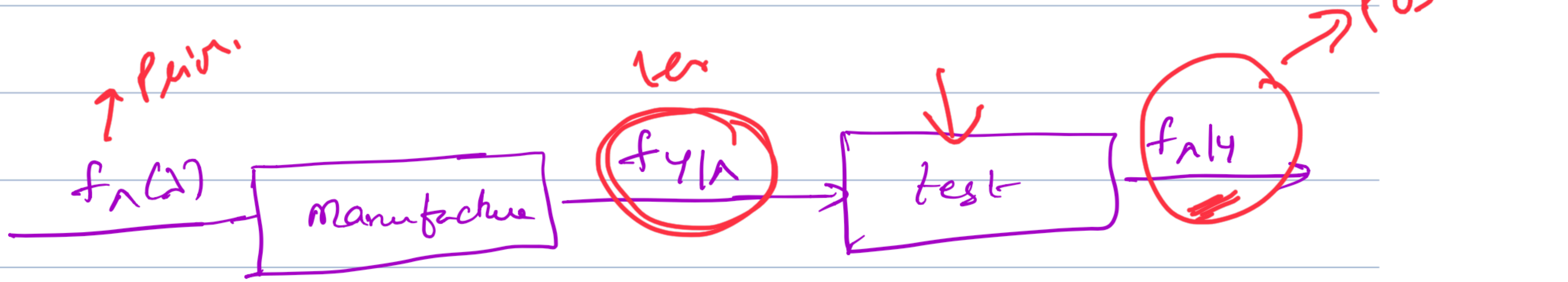
likelihood

- $$f_{X|Y}(x\,|\,y) = \frac{f_X(x) f_{Y|X}(y\,|\,x)}{\int_{-\infty}^{\infty} f_X(t) f_{Y|X}(y\,|\,t)\,dt}$$

Posterior.

Continuous Bayes' rule: example

Uniform: $f_\Lambda(\lambda) = 2$, for $1 \le \lambda \le \frac{3}{2}$



$f_{Y|\Lambda}(y|\lambda) = \lambda e^{-\lambda y}$ for $1 \le \lambda \le 3/2$

$$f_{\Lambda|Y}(\lambda|y) = \frac{f_\Lambda(\lambda)\, f_{Y|\Lambda}(y|\lambda)}{\int_{-\infty}^{\infty} f_\Lambda(t)\, f_{Y|\Lambda}(y|t)\,dt} = \frac{2\lambda e^{-\lambda y}}{\int_{1}^{3/2} 2te^{-ty}\,dt}$$

Bayesian statistical inference

We have a hypothesis

true false

likelihood

Prior.

- Bayes' rule is the key:

$$P(hypothesis\ is\ true \mid data) = \frac{P(data \mid hypothesis\ is\ true)P(hypothesis\ is\ true)}{P(data)}$$

Posterior.

$$P(H \mid D) = \frac{P(D \mid H)P(H)}{P(D)}$$

Priors

Which treatment would you choose?

- Treatment 1: cured 100% of patients in a trial ✓
- Treatment 2: cured 95% of patients in a trial
- Treatment 3: cured 90% of patients in a trial

very loosely.

$P(H)$

Priors

Which treatment would you choose?

- Treatment 1: cured 100% of patients in a trial
- Treatment 2: cured 95% of patients in a trial
- Treatment 3: cured 90% of patients in a trial

Which treatment would you choose?

Data.

- Treatment 1: cured 3 out of 3 patients in a trial
- Treatment 2: cured 19 out of 20 patients treated in a trial
- Standard treatment: cured 90000 out of 100000 patients in clinical practice.

3

Posterior

P(H|D)

Konduri Aditya, CDS, IISc Bengaluru

Discrete priors

Prior.

$P(H+) = 0.005$
$P(H-) = 1 - 0.005$
$= 0.995$

- A certain disease has a prevalence of 0.005
- A screening test has 2% false positives and 1% false negatives
- Suppose a patient is screened and has a positive test

1. Represent this information with a tree and use Bayes' theorem to compute the probabilities the patient does and doesn't have the disease
2. Identify the data, hypotheses, likelihoods, prior probabilities and posterior probabilities

elements of the inference.

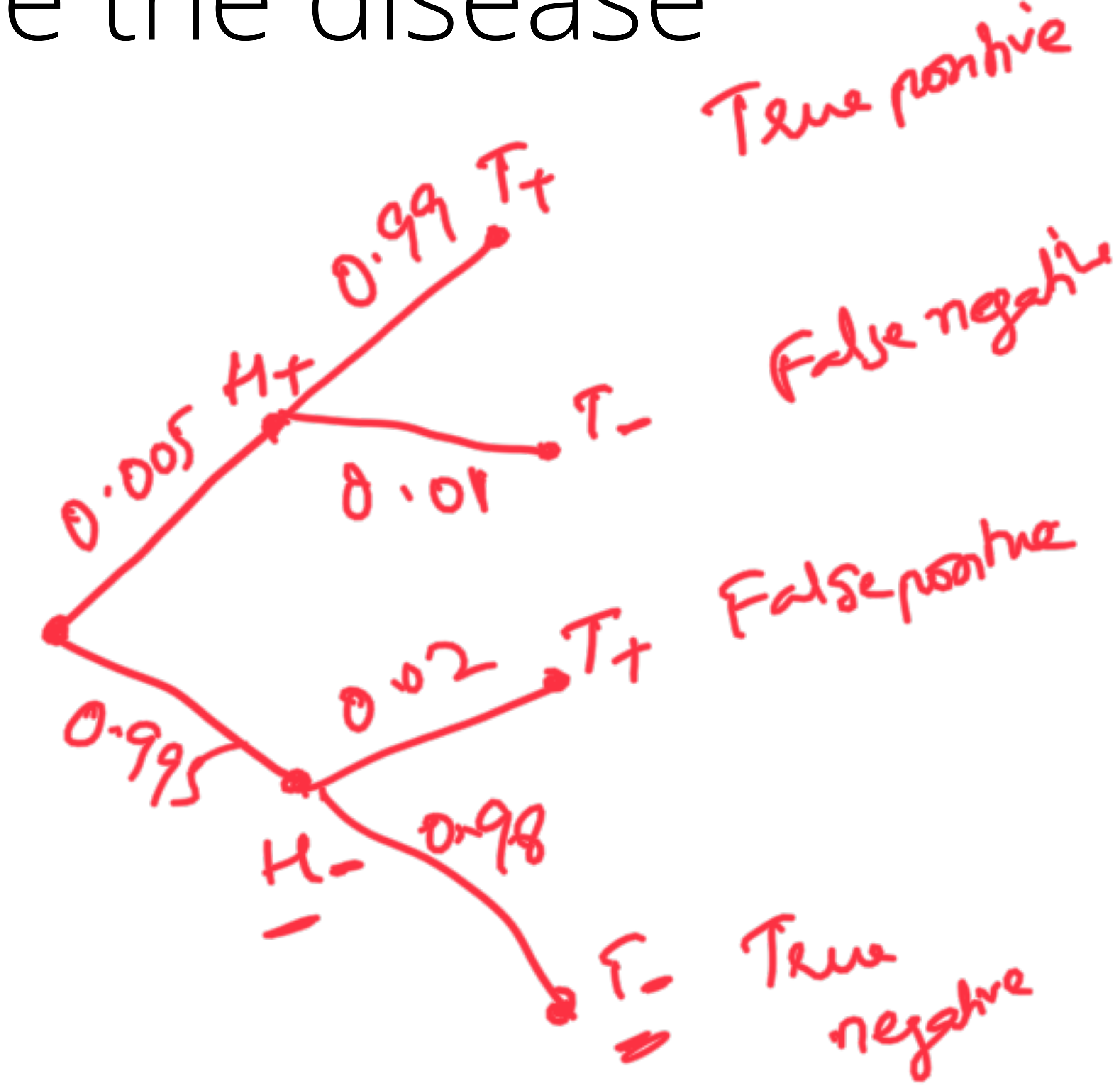
3. Make a full likelihood table containing all hypotheses and possible test data

Discrete priors

1. Represent this information with a tree and use Bayes' theorem to compute the probabilities the patient does and doesn't have the disease

H_+: patient has the disease, H_-: patient does not have the disease

T_+: patient tests positive, T_-: patient tests negative



Baye's rule.

$$P(H_+ \mid T_+) = \frac{P(H_+)P(T_+ \mid H_+)}{P(H_+)P(T_+ \mid H_+) + P(H_-)P(T_+ \mid H_-)} =$$

$$P(H_- \mid T_+) =$$

Discrete priors

Data: The data are the results of the experiment. In this case, the positive test

Hypotheses: The hypotheses are the possible answers to the question being asked. In this case they are H_+ the patient has the disease; H_- they don't

Likelihoods: The likelihood given a hypothesis is the probability of the data given that hypothesis. In this case there are two likelihoods, one for each hypothesis

$$P(T_+ | H_+) = 0.99 \quad and \quad P(T_+ | H_-) = 0.02$$

Likelihood is the probability given the hypothesis

Discrete priors

Prior probabilities of the hypotheses: The priors are the probabilities of the hypotheses prior to collecting data. In this case,

$$P(H_+) = 0.005 \quad and \quad P(H_-) = 0.995$$

Posterior probabilities of the hypotheses: The posteriors are the probabilities of the hypotheses given the data. In this case

→ increase.

$$P(H_+ | T_+) = 0.199 \quad and \quad P(H_- | T_+) = 0.801$$

Bulb example.

Λ → quality parameter

P(Λ = λ) = model wrong some belief.

Priors

$$P(H_+ | T_+) = \frac{P(H_+)P(T_+ | H_+)}{P(T_+)}$$

Testing → D1

Manufacture

P(Λ | D1)

Posterior

Prior

Failure in the equipment and has been replaced.

P(Λ | D2)

Discrete priors

3. Full likelihood table

The table holds likelihoods P(D|H) for every possible hypothesis and data combination

Positive test

Test → Positive
→ negative.

hypothesis $\mathcal{H}$	likelihood $P(\mathcal{D}\vert\mathcal{H})$	
disease state	$P(\mathcal{T}_+\vert\mathcal{H})$	$P(\mathcal{T}_-\vert\mathcal{H})$
$\mathcal{H}_+$	0.99	0.01
$\mathcal{H}_-$	0.02	0.98

negative test.

Beta distribution

- Continuous random variable
- Beta(a,b) has a PDF

$$\boxed{f(\theta) = \frac{(a+b-1)!}{(a-1)!(b-1)!}\theta^{a-1}(1-\theta)^{b-1}}$$

Here $0 \leq \theta \leq 1$ and (a, b) are the shape parameters (values > 0)

Konduri Aditya, CDS, IISc Bengaluru

Beta prior: problem

Parameter

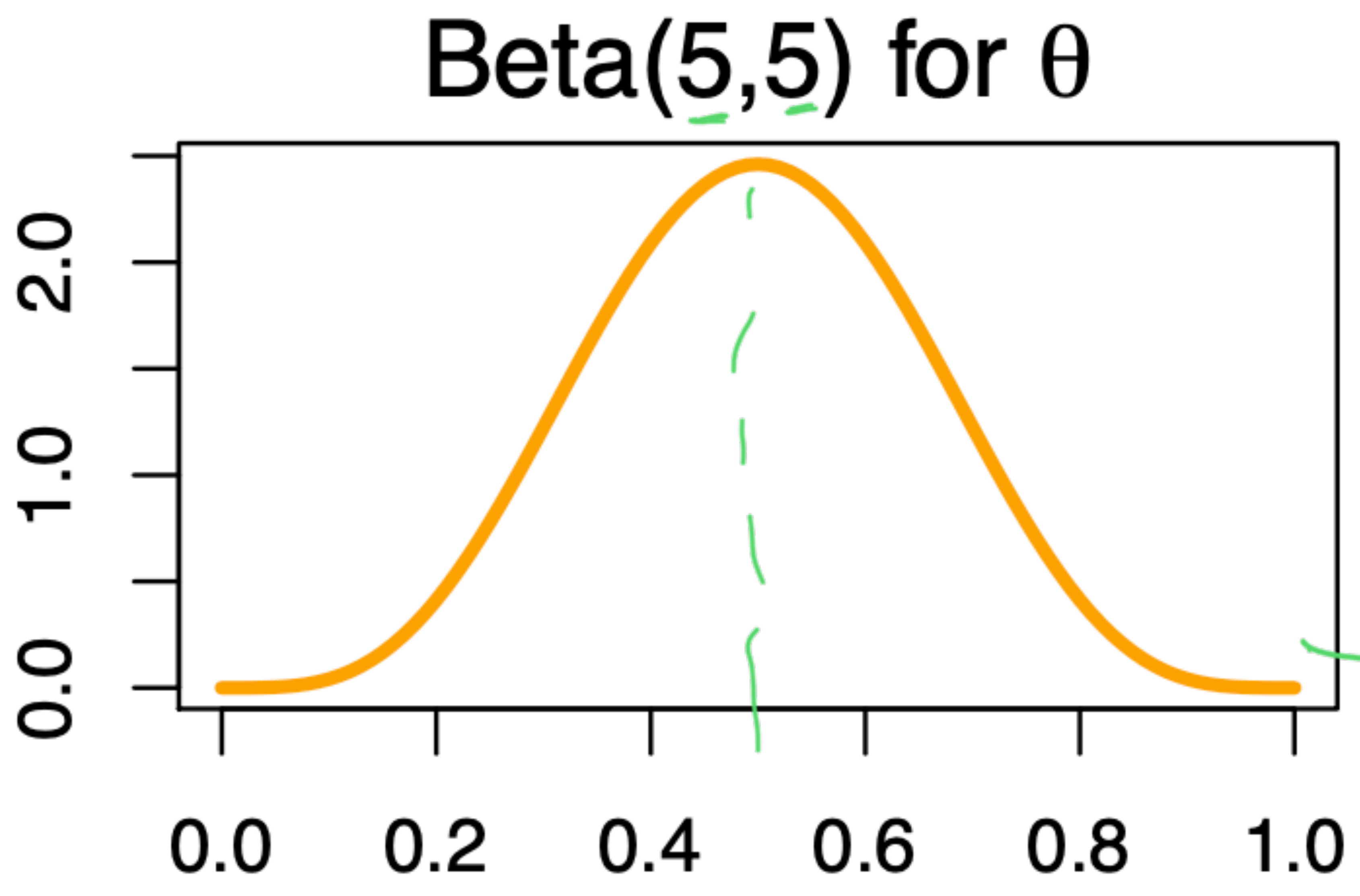
- Suppose you are testing a new medical treatment with unknown probability of success θ. You don't know that θ, but your prior belief is that it's probably not too far from 0.5. You capture this intuition with a beta(5,5) prior on θ

a=b=5

We have assumed a beta distribution for the parameter 'θ'

-

PDF



Prior. $f(\theta)$

Beta prior: problem

- Treatment has prior $f(\theta) = Beta(5,5)$

1. Suppose you test it on 10 patients and have 6 successes. Find the posterior distribution on θ. Identify the type of the posterior distribution

2. Suppose you recorded the order of the results and got SSSFFSSSFF. Find the posterior based on this data

Beta prior: solution

- Prior PDF $f(\theta) = Beta(5,5) = \boxed{\frac{9!}{4!4!}\theta^4(1-\theta)^4 = c_1\theta^4(1-\theta)^4}$
- Hypothesis: θ
- Likelihood: $P(Data\,|\,\theta) = {}^{10}C_6\theta^6(1-\theta)^4$ → Binominal distribution

$n=10$ $k=6$

Prior. likelihood.

- Posterior PDF:

$$f(\theta\,|\,Data) = \boxed{\frac{c_1\theta^4(1-\theta)^4\,.\,{}^{10}C_6\theta^6(1-\theta)^4}{\int_0^1 c_1 t^4(1-t)^4\,.\,{}^{10}C_6 t^6(1-t)^4 dt}} = Beta(11,9)$$

P(Data)

Bayes's rule for the cont. R.V.

Conjugate priors

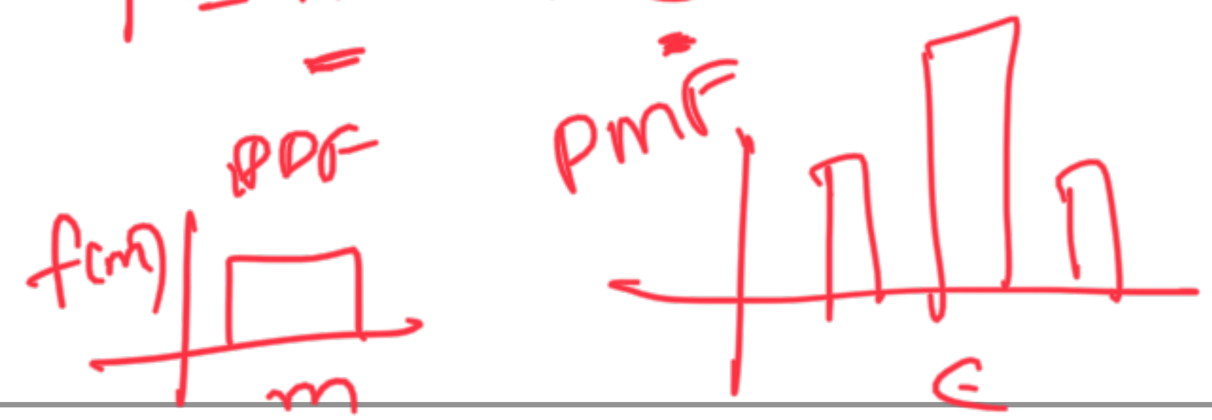
- Prior $f(\theta)d\theta$: beta distribution
- Likelihood: $P(Data\,|\,\theta)$: binomial distribution
- Posterior $f(\theta\,|\,Data)d\theta$: beta distribution
- The beta distribution is called a **conjugate prior** for the binomial likelihood
- What we have are the updated values of **a** and **b.**

Linear regression:

$x \rightarrow$ input

$y \rightarrow$ output

$y = mx + c$



Further reading

- MIT OCW 18:05 course: topics 10a-15b (https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/)
- https://seeing-theory.brown.edu/bayesian-inference/index.html
- https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1
- https://towardsdatascience.com/probability-concepts-explained-bayesian-inference-for-parameter-estimation-90e8930e5348