Image Searching and Modelling using Machine Learning Methods

Part IB Paper 8
Information Engineering Elective

Lecture 3: Bayesian Learning

Zoubin Ghahramani

zoubin@eng.cam.ac.uk

Department of Engineering
University of Cambridge

Easter Term

Bayesian Learning and Binary Data

Imagine we arrive from overseas and start to learn about the weather in Cambridge.

Let $x_n = 1$ mean "rain" on day n, and $x_n = 0$ mean "no rain".

The data set: $\mathcal{D} = \{1, 1, 1\}$ for first three days.

Let θ be the probability of rain (assume that rain is independent across days[1]).

Q: What is the ML estimate of θ? Is this reasonable?

Q: What does the likelihood look like?

Assume a uniform prior on θ: $P(\theta) = 1$
Q: What is the posterior on θ, $P(\theta|\mathcal{D})$?

Q: What is the expected value (mean) of θ given $\mathcal{D}$?

---

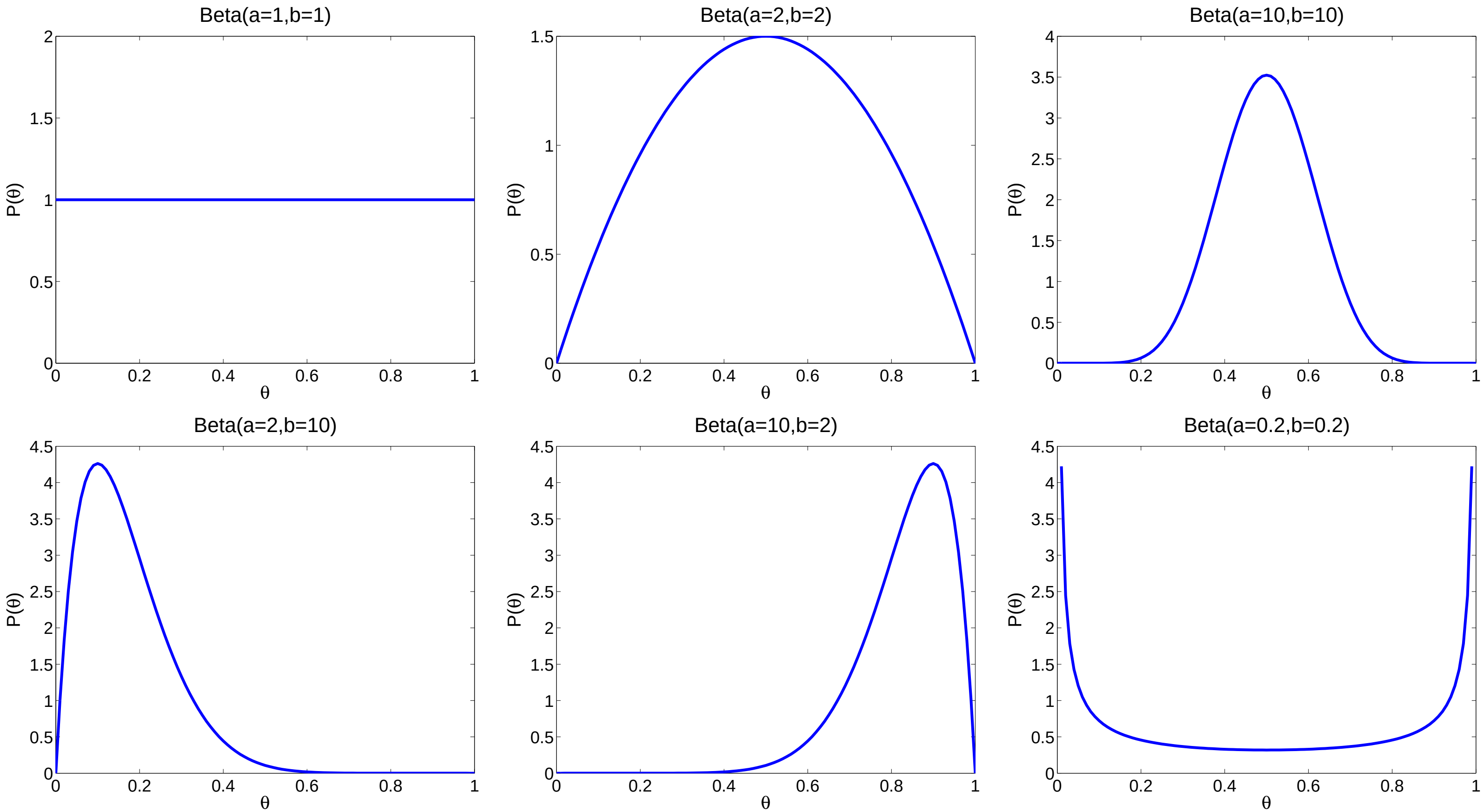
[1] clearly not realistic

The Beta distribution

The Beta distribution is: $$P(\theta|a,b) = \frac{1}{c}\theta^{a-1}(1-\theta)^{b-1}$$

where c is a normalization constant making the distribution integrate to 1 over $[0,1]$.[2]

Q: What does the Beta distribution look like?

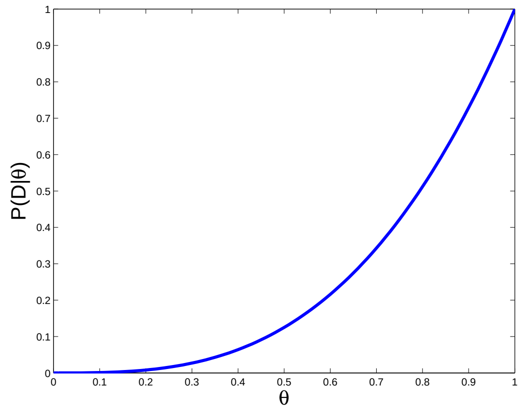


---

[2] $c = \int_0^1 \theta^{a-1}(1-\theta)^{b-1} d\theta$

Bayesian Learning and Binary Data

Q: What does the likelihood look like?



$$P(\mathcal{D}|\theta) = \theta^3$$

Assume a uniform prior on θ: $P(\theta) = 1$

Q: What is the posterior on θ, $P(\theta|\mathcal{D})$?

$$P(\theta|\mathcal{D}) = \theta^3/(1/4) = 4\theta^3$$

Q: What is the expected value (mean) of θ given $\mathcal{D}$?

$$E(\theta|\mathcal{D}) = \int_0^1 \theta P(\theta|\mathcal{D}) d\theta = 4 \int_0^1 \theta^4 d\theta = \frac{4}{5}$$

Q: What is the predicted probability of x given $\mathcal{D}$?

$$P(x = 1|\mathcal{D}) = \int P(x = 1|\theta) P(\theta|\mathcal{D}) d\theta = \int \theta P(\theta|\mathcal{D}) d\theta = \frac{4}{5}$$

The Beta distribution

The Beta distribution is: $P(\theta|a,b) = \frac{1}{c}\theta^{a-1}(1-\theta)^{b-1} = \mathrm{Beta}(a,b)$
where c is a normalization constant making the distribution integrate to 1 over $[0,1]$.

Compare the Beta distribution to the Bernoulli over x:

$$P(x|\theta) = \theta^x(1-\theta)^{(1-x)}$$

Q: What happens if multiply a Beta prior and a Bernoulli likelihood?

$$\begin{aligned} P(\theta|x,a,b) &\propto P(\theta|a,b)P(x|\theta) \\ &= \frac{1}{c}\theta^{a-1}(1-\theta)^{b-1}\theta^x(1-\theta)^{(1-x)} \\ &= \frac{1}{c}\theta^{a+x-1}(1-\theta)^{b+1-x-1} \propto \mathrm{Beta}(a+x, b+1-x) \end{aligned}$$

Beta prior $\times$ Bernoulli likelihood $=$ Beta posterior[3].

[3] This property is called *conjugacy*: Beta is the conjugate prior for the Bernoulli likelihood.

More on the Beta distribution

What is the normalizating constant?

$$P(\theta|a,b) = \frac{1}{c}\theta^{a-1}(1-\theta)^{b-1} = \mathrm{Beta}(a,b)$$

$$c = \int_0^1 \theta^{a-1}(1-\theta)^{b-1} d\theta$$
$$c = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

where

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$$

The Gamma function is a generalization of the factorial function. For integer a:

$$\Gamma(a) = (a-1)!$$

[You will not be examined on the definition of the Gamma function]