Lecture 2B
MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

Today's agenda

Today's lecture

- Review
- Use Bayes' theorem to compute posterior pmf with discrete pmf priors.
- Use Bayes' theorem to compute posterior pfd with continuous pdf priors

Announcements

New office hours starting from today

- Every Wednesday from 12:00-13:00, room MB-B11

Bayesian inference

$Y \sim p(y|\theta)$

- Probability model $p(y \mid \theta)$ depends on a set of parameters θ.
- θ is unknown and we would like to learn about θ.
- Let y be the observed data, assumed to be generated by this probability model, $p(y \mid \theta)$ $Y = y$ (the realisation of Y)
- In Bayesian statistics, we assign probabilities on both the parameters θ and data y

Bayesian inference

- So we start with a probability distribution for the parameters $p(\theta)$, called the prior distribution.
- θ is either discrete or continuous random variable. Hence, $p(\theta)$ is either a pmf or pdf
- The prior is a subjective distribution, based on experimenter's belief, and is formulated before the data y are seen.

Bayesian inference

- Let y be the observed data.
- We then update the prior distribution (pmf or pdf) to a posterior distribution (pmf or pdf) for θ, $p(\theta \mid y)$, using Bayes' theorem

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

  where the observed data enters through the likelihood $p(y \mid \theta)$.
- $p(y)$ is the normalising constant, which is given by

$$p(y) = \int \underbrace{p(\theta')}_{\text{pdf}}\, p(y \mid \theta')\, d\theta' \;\text{ or }\; \sum_{\theta'} \underbrace{p(\theta')}_{\text{pmf}}\, p(y \mid \theta')$$

- $p(y)$ does not depend on θ

θ is continuous $\int p(\theta|y)\, d\theta = 1$

What does it mean?

$$p(\theta \mid y) \propto p(\theta)\, p(y \mid \theta) \qquad (1)$$

Posterior $\propto$ prior $\times$ likelihood

- $p(y \mid \theta)$ is the likelihood and it the probability of data y given the true θ.
- Start with initial beliefs/information about θ, $p(\theta)$ - this is the prior distribution formulated before the data are seen.
- Bayesian updating: Update the prior distribution using the data y, using (1).
- The updated prior, $p(\theta \mid y)$ is called the posterior distribution .
- We base our inferences about θ based on this posterior distribution.

Bayesian updating with discrete data, discrete prior

- parameter θ discrete with values θ_1 and θ_2 and prior pmf $p(\theta)$
- Discrete data x
- Discrete likelihood, $p(x|\theta)$
- Posterior pmf: $p(\theta_1|x)$, $p(\theta_2|x)$

$p(\theta_1|x) = p(\theta = \theta_1 | x)$

Hypothesis	prior	likelihood	Bayes numerator	posterior
θ	$p(\theta)$	$p(x\|\theta)$	$p(x\|\theta)p(\theta)$	$p(\theta\|x)$
θ_1	$p(\theta_1)$	$p(x\|\theta_1)$	$p(x\|\theta_1)\ p(\theta_1)$	$p(\theta_1\|x)$
θ_2	$p(\theta_2)$	$p(x\|\theta_2)$	$p(x\|\theta_2)\ p(\theta_2)$	$p(\theta_2\|x)$
Total	1	NOT SUM TO 1	$p(x)$	1

- Law of total probability: $p(x) = p(x|\theta_1)p(\theta_1) + p(x|\theta_2)p(\theta_2)$.
- Bayes' theorem: $p(\theta_1|x) = \dfrac{p(x|\theta_1)p(\theta_1)}{p(x)}$, $\quad p(\theta_2|x) = \dfrac{p(x|\theta_2)p(\theta_2)}{p(x)}$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{total prob. of data}}.$$

Board Question: Coins

- There are three types of coins which have different probabilities of heads
  - Type A coins are fair, with probability 0.5 of heads.
  - Type B are bent and have probability 0.6 of heads.
  - Type C are bent and have probability 0.9 of heads.

  Suppose I have a drawer containing 5 coins: 2 of type A, 2 of type B, and 1 of type C. I pick a coin at random, and without showing you the coin I flip it once and get heads.

- Make a Bayesian update table and compute the posterior pmf that the chosen coin is each of the three coins.

Solution

Hypothesis. The hypothesis is the probability of heads θ. The value of θ is itself random that takes three values

$$\theta = 0.5 \quad \text{means} \quad \text{coin is Type A}$$
$$\theta = 0.6 \quad \text{means} \quad \text{coin is Type B}$$
$$\theta = 0.9 \quad \text{means} \quad \text{coin is Type C}$$

Prior pmf. Since θ is discrete, it has a prior pmf $p(\theta)$, $\theta \in \{0.5, 0.6, 0.9\}$

$$p(0.5) = P(\theta = 0.5) = \frac{2}{5}$$
$$p(0.6) = P(\theta = 0.6) = \frac{2}{5}$$
$$p(0.9) = P(\theta = 0.9) = \frac{1}{5}$$

Data. We observe heads, $x = 1$

Likelihood. The likelihood $p(x|\theta)$ is the probability of observing heads $x = 1$ given θ. We have 3 likelihood values

$$p(x=1|0.5) = 0.5, \quad p(x=1|\theta=0.6) = 0.6$$
$$p(x=1|\theta=0.9) = 0.9$$

You can think x generated from the Bernoulli pmf with unknown probability of heads θ,

$$x \sim p(x|\theta) \text{ where}$$

$$p(x|\theta) = \theta^x (1-\theta)^{1-x} \qquad x=1 \text{ or } x=0$$

when $x=1$ $\qquad p(x=1|\theta) = \theta$

In our case, $\theta \in \{0.5, 0.6, 0.9\}$

Posterior pmf

These are the probabilites

$p(\theta=0.5|x=1)$

$p(\theta=0.6|x=1)$

$p(\theta=0.9|x=1)$

we want to compute the posterior pmf of θ

Bayesian updating table

hypothesis θ	prior pmf $p(\theta)$	Likelihood $p(x=1|\theta)$	Bayes num $p(x=1|\theta)p(\theta)$	posterior pmf
0.5	$p(0.5)=2/5$	0.5	0.2	0.3226
0.6	$p(0.6)=2/5$	0.6	0.24	0.3871
0.9	$p(0.9)=1/5$	0.9	0.18	0.2903
Total	1		$P(x=1)=0.62$	1

$$p(x=1) = p(x=1|\theta_1)p(\theta_1) + p(x=1|\theta_2)p(\theta_2) + p(x=1|\theta_3)p(\theta_3)$$

$$p(\theta=\theta_1|x=1) = \frac{0.5 \times 2/5}{0.62} = 0.3226$$

- In the previous lecture, we have done Bayesian updating when we had a finite number of hypotheses or a discrete parameter θ e.g.,
  - in the diagnostic example had 2 hypotheses (HIV +ve, HIV -ve),
  - in the coin example we had 3 hypothesis (A, B and C).
- In this topic we will study Bayesian updating where there is a continuous range of hypotheses, i.e., θ is a continuous random variable.
- The Bayesian updating will be essentially the same, based on the Bayes' theorem

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

Examples with continuous parameters

- Suppose we have a medical treatment for a disease than can succeed or fail with probability q. Then q is a continuous quantity between 0 and 1.

- The lifetime of a certain light bulb T is modeled as an exponential distribution $\exp(\lambda)$ with unknown λ. We can assume that λ takes any value greater than 0.

Baysian updating: Discrete likelihoods, continuous priors

- θ : continuous parameter with prior pdf $p(\theta)$ and range $[a, b]$.
- x : random discrete data
- discrete likelihood: $p(x|\theta)$
- posterior pdf: $p(\theta|x)$

- By **Bayes' theorem** we update the prior pdf to a posterior pdf

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int_a^b p(x|\theta)p(\theta)d\theta}.$$

- Law of total probability: $p(x) = \int_a^b p(x|\theta)p(\theta)\, d\theta$.

Baysian updating: Discrete likelihoods, continuous priors

- $p(x)$ does not depend on θ and serves as the normalising constant so that $p(\theta|x)$ is a proper pdf and integrates to 1.
- Hence, we can express Bayes' theorem in the form

$$p(\theta|x) \propto p(x|\theta)p(\theta).$$

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

Bayesian inference

$$p(\theta \mid x) \propto p(\theta)\, p(x \mid \theta)$$

- $p(\theta)$ - initial beliefs/information about θ, the prior pdf.
- $p(x \mid \theta)$ - the likelihood for observed data x with parameters θ.
- Update information about θ using the likelihood.
- The resulting pdf $p(\theta \mid x)$ is called the posterior pdf of θ

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

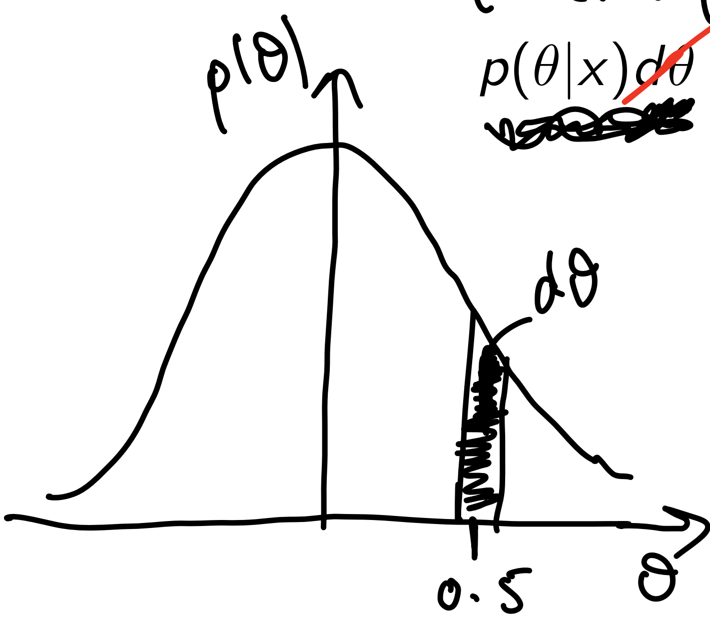
- Sometimes, it is better to use $p(\theta)d\theta$ to work with probabilities instead of densities
  e.g the prior probability that θ is in a small interval of width $d\theta$ around 0.5 if $p(0.5)d\theta$.

- In this case, the Bayes' theorem is

posterior prob

prior prob

$$p(\theta|x)\cancel{d\theta} = \frac{p(x|\theta)p(\theta)d\theta}{p(x)} = \frac{p(x|\theta)p(\theta)\cancel{d\theta}}{\int_a^b p(x|\theta)p(\theta)\, d\theta}.$$



The probability that θ lies in a very small interval of width $d\theta$ around 0.5 is approximately $p(0.5)d\theta$

Bayesian updating: Discrete likelihoods, continuous priors

- θ : continuous parameter with prior pdf $p(\theta)$ and range $[a, b]$.
- x : random discrete data
- likelihood: $p(x|\theta)$

Bayesian updating table

Hypothesis	prior prob	likelihood	Bayes numerator	posterior prob $p(\theta|x)d\theta$
θ	$p(\theta)d\theta$	$p(x|\theta)$	$p(x|\theta)p(\theta)d\theta$	$\frac{p(x|\theta)p(\theta)d\theta}{p(x)}$
Total	$\int_a^b p(\theta)d\theta = 1$		$p(x) = \int_a^b p(x|\theta)p(\theta)d\theta$	1

- The posterior density $p(\theta|x)$ is obtained by removing $d\theta$ from the posterior probability in the table.

$$p(\theta|x)\cancel{d\theta} = \frac{p(x|\theta)p(\theta)\cancel{d\theta}}{p(x)}$$

The posterior density is $p(\theta|x)$

Bayesian updating: Discrete likelihoods, continuous priors

Example: Binomial data, Beta prior

- A biased coin has probability of heads q which is unknown.
- We toss the coin n times and observe k heads (This is my data $x = k$).
- The binomial likelihood for this problem:

$X \sim \text{binom}(n, q)$
$X = k$

$$p(k \mid q) = \binom{n}{k} q^k (1-q)^{n-k} \rightarrow \text{binomial pmf}$$

- For Bayesian inference, we need to specify a prior distribution for q.
- q is a continuous quantity between 0 and 1.
- What is a possible probability distribution for q (or family of distributions)?

Bayesian updating: Discrete likelihoods, continuous priors

Example: Binomial data, Beta prior

- The family of Beta distributions seems a natural choice for a prior distribution for q, since it describes continuous random variables with support on $[0, 1]$.
- If $q \sim \text{Beta}(\alpha, \beta)$, its probability density function is

$$f(q) = \frac{q^{\alpha-1}(1-q)^{\beta-1}}{B(\alpha, \beta)}, \ 0 \leq q \leq 1,$$

where B is the Beta function and α and β are parameters,

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx$$

$$\rightarrow B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$\Gamma(\cdot)$ is the gamma function

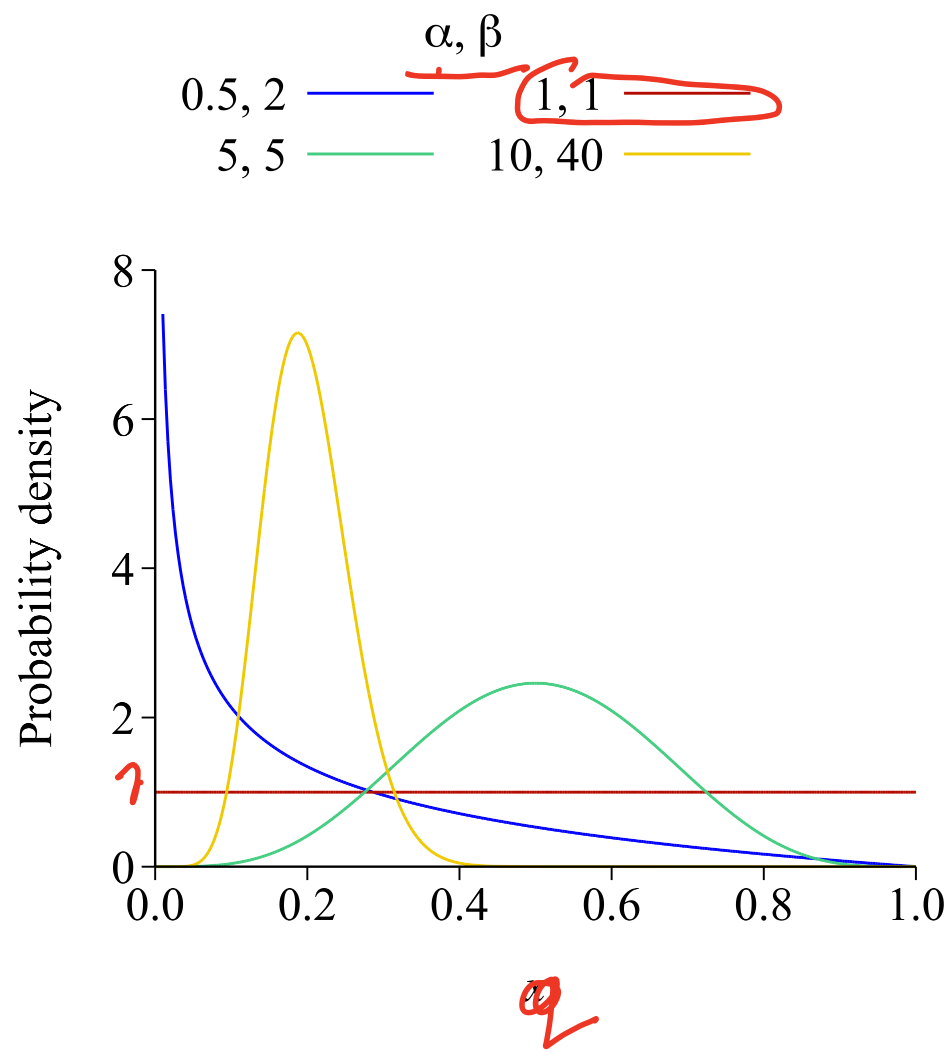
Beta distributions

Beta(1, 1) is the uniform distribution on [0, 1]

- Probability density functions.
- If $q \sim \text{Beta}(\alpha, \beta)$

$$E(q) = \frac{\alpha}{\alpha + \beta}$$

prior mean



Example: Binomial data, Beta prior

- Bayesian updating:

$$\text{posterior} \propto \underline{\text{prior}} \times \underline{\text{likelihood}} \qquad x = k$$

- The posterior distribution $p(q|x)$ is proportional to

$$p(q \mid k) \propto \underbrace{q^{k+\alpha-1}(1-q)^{n-k+\beta-1}}$$

- We recognise this to have the form of a beta distribution, so the posterior is a beta distribution, beta$(k+\alpha, n-k+\beta)$.

- Hence, the normalising constant must be $1/B(k+\alpha, n-k+\beta)$.

Proof: The posterior density, $p(q|x)$, is

$$p(q|x) \propto p(x|q)\, p(q) = \text{Bayes numerator}$$

$$= c_1 \binom{n}{x} q^{x} (1-q)^{n-x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} q^{a-1} (1-q)^{b-1}$$

$$= \underbrace{c_2\, q^{x+a-1} (1-q)^{n-x+b-1}}$$

We recognise this to have the same form with the beta pdf. Hence we can find the value of c_2 that makes $p(q|x)$ a proper density (integrates to 1)

We want

$$B(x+a, n-x+b) \cdot c_2 \int_0^1 \underbrace{\frac{q^{x+a-1} (1-q)^{n-x+b-1}}{B(x+a, n-x+b)}}_{\text{Beta}(x+a,\, n-x+b) \text{ density}} dq = 1$$

$=1$

$$\Rightarrow B(x+a, n-x+b) \cdot c_2 = 1$$

$$\Rightarrow c_2 = \frac{1}{B(x+a, n-x+b)}$$

The posterior density, $p(q|x)$ is $\text{Beta}(x+a, n-x+b)$

$$p(q|x) = \frac{q^{x+a-1} (1-q)^{n-x+b-1}}{B(x+a, n-x+b)}, \quad q \in [0,1]$$

You don't need to compute

$$P(X=x) = \cancel{\int_0^1 p(x|q)\, p(q)\, dq}$$

$$p(q|x) = \frac{p(x|q)\; p(q)}{P(X=x)} = \frac{\text{Bayes numerator}}{\text{Total prob. of heads}}$$

Bayesian updating: Discrete likelihoods, continuous priors

- The actual, normalized pdf is

$$p(q \mid k) = \frac{q^{k+\alpha-1}(1-q)^{n-k+\beta-1}}{B(k+\alpha, n-k+\beta)},$$

the pdf of a Beta$(k+\alpha, n-k+\beta)$ r.v. (**Remember:** the random variable is q and k is fixed).

- Bayesian updating: We update the prior Beta(α, β) to posterior Beta$(k+\alpha, n-k+\beta)$. (For binomial data)

Bayesian updating table: Discrete likelihoods, continuous priors

Example: Binomial data, Beta prior

- $Y \sim$ Binom(n,q), with q unknown
- Continuous hypotheses q in $[0, 1]$.
- Data y $Y=y$
- Prior $p(q)$
- Likelihood $p(y|q)$

Hypothesis	prior prob.	likelihood	Bayes numerator	posterior prob.
q	Beta$(\alpha, \beta)dq$	binomial(n, q)	$cq^{k+\alpha-1}(1-q)^{n-k+\beta-1}dq$	Beta$(k+\alpha, n-k+\beta)dq$
Total	1		$T = \int_0^1 cq^{k+\alpha-1}(1-q)^{n-k+\beta-1}dq$	1

- The posterior density is Beta$(k+\alpha, n-k+\beta)$
- **Note:** We don't need to compute T. Once we know the posterior is of the form $cq^{k+\alpha-1}(1-q)^{n-k+\beta-1}$ we have to find c that makes it a proper density. In this case $c = 1/\text{Beta}(k+\alpha, n-k+\beta)$

Unknown parameters and prior parameters

Remarks:

- We need to distinguish between the parameters we are estimating, which we generally have denoted by θ and the parameters for the prior distribution(s).
- In this binomial example, q is uncertain: we have prior and posterior distributions for q.
- The parameters of the prior distribution, here α and β, are taken as fixed.

$p(q) \sim \text{beta}(\underline{\alpha}, \underline{\beta})$ where α, β are fixed and known

$X \sim \text{binom}(n, q)$ q is unknown

Board question: bent coin

- Bent coin with unknown probability θ of heads
- Prior: $p(\theta) = 2\theta$ on $[0, 1]$
- Data: toss and get heads

- Compute the Bayesian update table.
- Find the posterior pdf to this data.

Solution

Hypothesis is θ, the probability of heads, $\theta \in [0,1]$

Prior pdf. $p(\theta) = 2\theta$, $\theta \in [0,1]$

Data: $x = 1$, in this case we observe heads

Likelihood The likelihood, $p(x=1|\theta)$ is the probability of observing heads given the true θ. Then

$$p(x=1|\theta) = \theta$$

Posterior pdf The posterior density, $p(\theta|x=1)$ is

$$p(\theta|x=1) \propto p(x=1|\theta) \times p(\theta) \rightarrow \text{Bayes numerator}$$

$$= c_1 \theta \times 2\theta$$

$$= c_2 \theta^2$$

We want to find c_2 such that $\int_0^1 p(\theta|x=1) d\theta = 1$

So $c_2 \int_0^1 \theta^2 d\theta = 1 \Leftrightarrow c_2 \left[\frac{\theta^3}{3}\right]_0^1 = 1 \Rightarrow c_2 = 3$

The posterior pdf of θ given the data, $x=1$, is

$$p(\theta|x=1) = \underbrace{3\theta^2}_{\text{Beta}(3,1)}, \quad \theta \in [0,1]$$

Bayesian Table

hypothesis	Prior prob.	likelihood	Bayes numer.	posterior prob
θ	$2\theta\,d\theta$	θ	$c2\theta^2 d\theta$	$3\theta^2 d\theta$
Total	1		$\int_0^1 c2\theta^2 d\theta$	1

Posterior mean

In Bayesian how would you choose a particular value of q?

- A natural estimate for q is the mean of the posterior distribution $p(q|k)$, called the posterior mean.
- For the binomial case with Beta(α, β) prior, the posterior mean is

$$\hat{q}_{\mathrm{B}} = E(q|k) = \frac{k+\alpha}{n+\alpha+\beta}.$$

- The prior distribution has mean $\alpha/(\alpha+\beta)$ which would be our best estimate of q without having observed the data.
- Ignoring the prior, we would estimate q using the maximum likelihood estimate (MLE)

$$\hat{q} = \frac{k}{n}$$

- The Bayes' estimate $\hat{q}_{\mathrm{B}}$ combines all of this information.

If $q \sim \text{Beta}(a, b)$, then $\mathbb{E}(q) = \dfrac{a}{a+b}$ prior mean

If $q \mid x \sim \text{Beta}(x+a, n-x+b)$, $\mathbb{E}(q \mid x) = \dfrac{x+a}{\cancel{x}+a+n-\cancel{x}+b}$

$$= \underbrace{\frac{x+a}{n+a+b}}_{\text{posterior mean}}$$

Posterior mean

- Note that we can rewrite $\hat{q}_b$ as

$$\hat{q}_B = \frac{n}{n+\alpha+\beta}\left(\frac{k}{n}\right) + \frac{\alpha+\beta}{n+\alpha+\beta}\left(\frac{\alpha}{\alpha+\beta}\right)$$

- Thus $\hat{q}_B$ is a linear combination of the prior mean and the MLE, with the weights being determined by n, α and β

$\hat{q}_B \to B$ stands for posterior mean / Bayesian estimate / Bayes estimate

Flat priors

- One important prior is called flat prior or uniform prior.
- A flat prior assumes that every hypothesis is equally probable.

uniform distribution

- For example if q has range $[0, 1]$, then $p(q) = 1$ is a flat prior.
- E.g. a uniform distribution on $[0, 1]$ is Beta$(1, 1)$ $\alpha = 1, \beta = 1$
- So, posterior distribution is Beta$(k + 1, n - k + 1)$
- Posterior mean: $E(q \mid k) = \dfrac{k + 1}{n + 2}$

Board question

- Bent coin with unknown probability θ of heads
- Flat prior: $p(\theta) = 1$ on $[0, 1]$
- Data: toss 27 times and get 15 heads and 12 tails.
- Compute the Bayesian update table.
- Give the integral for the normalising factor but do not compute it. Call its value T and give the posterior pdf in terms of T.

Board question

- A medical treatment has unknown probability θ of success.
- We assume treatment has prior $f(\theta) \sim \text{Beta}(5, 5)$.

1. Suppose you test it on 10 patients and have 6 successes. Find the posterior distribution on θ. Identify the type of the posterior pdf
2. Suppose you recorded the order of the results and got SSSFFSSSFF. Find the posterior based on this new data.