Lecture 3A
MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

Today's agenda

Today's lecture will cover

- Review Bayesian updating with continuous parameters and discrete data.
- Construct a posterior for continuous parameters and continuous data.
- Conjugate priors

Bayesian inference

- Suppose we have data y generated from $p(y \mid \theta)$ where θ is the unknown parameter.
- Start with the prior distribution $p(\theta)$ about θ.
- Likelihood is $p(y \mid \theta)$.
- The resulting probability distribution $p(\theta \mid y)$ is called the posterior distribution.

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

Posterior distribution $\propto$ prior distribution $\times$ likelihood

- Our inferences about θ are based on this posterior distribution.

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.

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

Binomial data/beta prior example

- $Y \sim \text{binom}(n, q)$ with unknown binomial probability of success q.
- We observe $Y = k$ successes in n trials. $k \sim \text{binom}(n, q)$
- The binomial likelihood $p(k|q)$ for this problem is:

$$P(Y=k|q) = p(k \mid q) = \binom{n}{k} q^k (1-q)^{n-k}$$

- Convenient prior distribution for q is $\text{Beta}(\alpha, \beta)$:

$$p(q) = \frac{q^{\alpha-1}(1-q)^{\beta-1}}{B(\alpha, \beta)} \quad , \ q \in [0,1]$$

Binomial data/beta prior example

Bayes numerator

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

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

$c>0$

Hence the posterior distribution is proportional to

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

From this, we can recognise that the posterior distribution $p(q \mid k)$ is the $\text{Beta}(k+\alpha, n-k+\beta)$ distribution.

General $\text{Beta}(a,b)$ pdf: $\quad f(x) = \dfrac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$

We have that the posterior density $p(q|x)$ is

$$p(q|x) = c\, q^{x+a-1} \cdot (1-q)^{n-x+b-1}$$

We see that $p(q|x)$ is the beta density with parameters $x+a$ and $n-x+b$

$$p(q|x) \sim \text{beta}(x+a, n-x+b)$$

The normalising constant c must be

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

The constant c makes the density to integrate 1

Bayesian updating table: Binomial data/beta prior

- Data k generated from $\sim$ Binom(n,q), with q unknown
- Continuous hypotheses q in $[0, 1]$.
- Beta(α, β) prior $p(q)$
- Binomial likelihood $p(k|q)$

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 $c = 1/\text{Beta}(k+\alpha, n-k+\beta)$

Conjugate distributions

- $\alpha + \beta \rightarrow$ initial number of trials

- $\text{Beta}(\alpha, \beta)$ prior distribution for q
- Binomial likelihood $k \sim \text{Bin}(n, q)$
- $\text{Beta}(k + \alpha, n - k + \beta)$ posterior distribution for $q \mid k$

- In this example, we have the same family of distributions for the prior and posterior distribution.
- This is known as a conjugate distribution.
- "The family of Beta distributions is conjugate to the binomial likelihood".

Conjugate distributions

$X \sim \text{Geom}(q)$

Binomial likelihood: $p(k \mid q) = \binom{n}{k} \underline{q^k (1-q)^{n-k}}$ $\quad P(X=x) = \underline{q^x (1-q)}$

Beta prior: $p(q) = \dfrac{q^{\alpha-1}(1-q)^{\beta-1}}{B(\alpha, \beta)}$

- Considered as functions of q, the prior and likelihood have the same functional form as each other (proportional to $q^r(1-q)^s$ for some r, s).
- When we multiply them together, we still have the same form.
- This is what characterises conjugate distributions.

Board question

θ = probability of heads

- Suppose your prior in the bent coin example is Beta$(6, 8)$. You flip the coin 7 times, getting 2 heads and 5 tails. What is the posterior pdf $p(\theta|x)$?

  1. Beta$(2, 5)$
  2. Beta$(3, 6)$
  3. Beta$(6, 8)$
  4. Beta$(8, 13)$

Solution

$\alpha = 6, \quad \beta = 8$

$n = 7, \quad x = 2, \quad n - x = 5$

$p(\theta|x) \sim \text{beta}(8, 13)$

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.

$\theta^6(1-\theta)^4$

Solution

- Prior $p(\theta) = \dfrac{\theta^4(1-\theta)^4}{B(5,5)}$, $\theta \in [0,1]$

- The data $x=6$ successes in 10 patients come from a binomial model. So the likelihood is just the binomial likelihood

$$p(6|\theta) = \underline{\binom{10}{6}} \theta^6 (1-\theta)^4$$

The posterior density, $p(\theta|6)$ is

$$p(\theta|6) \propto \frac{\theta^4(1-\theta)^4}{B(5,5)} \times \binom{10}{6} \theta^6 (1-\theta)^4$$

$$= \underbrace{c\, \theta^{10} (1-\theta)^8}$$

This is the beta$(11,9)$, so the normalising constant is just

$$c = \frac{1}{B(11,9)}$$

The posterior density is $p(\theta|6) = \dfrac{\theta^{10}(1-\theta)^8}{B(11,9)}$.

(b) The answer is again beta$(11, 9)$. The only thing that changes is the binomial coefficient which is just 1.

Bayesian updating: continuous priors, continuous data

We are now ready to do Bayesian updating when both the parameters and the data take continuous values.

- θ continuous parameter
- Prior pdf, $f(\theta)$
- Data: continuous $x \sim f(x|\theta)$
- Likelihood $f(x|\theta)$
- posterior pdf, $f(\theta|x)$
- **Bayesian update table**

Hypothesis	prior prop	likelihood	Bayes numerator	posterior prop $f(x\|\theta)d\theta$
θ	$f(\theta)d\theta$	$f(x\|\theta)$	$f(x\|\theta)\ f(\theta)d\theta$	$\frac{f(x\|\theta)f(\theta)d\theta}{f(x)}$
Total	1		$f(x)$	1

- $f(x) = \int f(x|\theta)\ f(\theta)d\theta$

Normal example, known variance

- $\underbrace{y_1, \ldots, y_n}_{} \overset{iid}{\sim} N(\mu, \sigma^2)$.
- It's simpler if only one parameter is unknown.
- First, consider case where only μ is unknown. σ^2 is known
- Is there a conjugate prior for μ?

Normal example, known variance

- Observed data $y_1, \ldots, y_n \sim N(\mu, \sigma^2)$ with μ unknown and σ^2 known.
- Prior distribution $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$.
- The posterior distribution is

$$\mu \sim \mathcal{N}(\mu_1, \sigma_1^2)$$

where

$$\mu_1 = \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{y}}{\sigma^2}\right) \Big/ \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)$$

$$\sigma_1^2 = 1 \Big/ \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)$$

Proof

Let $y_1, \dots, y_n \sim N(\mu, \sigma^2)$ where σ^2 is known.
The likelihood, $p(y_1, \dots, y_n | \mu)$, is the joint density of $y_1, \dots, y_n$. By independence of $y_1, \dots, y_n$,

$$p(y_1, \dots, y_n | \mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\}$$

$$= \underbrace{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n} \exp\left\{-\sum_{i=1}^{n} \frac{(y_i-\mu)^2}{2\sigma^2}\right\}.$$

$$\propto \exp\left\{-\sum_{i=1}^{n} \frac{(y_i-\mu)^2}{2\sigma^2}\right\}$$

The prior for μ is $N(\mu_0, \sigma_0^2)$

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}$$

We can rewrite the likelihood as

$$\bar{y} = \frac{\sum y_i}{n}$$

$$p(y|\mu) \propto \exp\left\{-\sum_{i=1}^{n} \frac{(y_i^2 - 2y_i\mu + \mu^2)}{2\sigma^2}\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} y_i^2 - 2n\bar{y}\mu + n\mu^2\right)\right\}$$

Normal example, known variance

Normal-normal Bayesian update table

- Data: $x \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 known
- Likelihood: $f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$.
- μ continuous with prior pdf $f(\theta) \sim \mathcal{N}(\mu_0, \sigma_0^2)$
- posterior $f(\mu|x) \sim \mathcal{N}(\mu_1, \sigma_1^2)$

Hypothesis	prior prop	likelihood	Bayes numerator	posterior prop $f(x|\mu)d\mu$
μ	$\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\}d\mu$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$	$c_1 \exp\{-\frac{1}{2\sigma_1^2}(\mu-\mu_1)^2\}d\mu$	$\frac{f(x|\mu)f(\mu)d\mu}{f(x)} = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\{-\frac{1}{2\sigma_1^2}(\mu-\mu_1)^2\}d\mu$
Total	1		$f(x) = \int_{-\infty}^{\infty} c_1 \exp\{-\frac{1}{2\sigma_1^2}(\mu-\mu_1)^2\}d\mu$	1

Normal example, known variance

Normal-normal updating formulas

$$a = \frac{1}{\sigma_0^2}, \quad b = \frac{n}{\sigma^2}, \tag{1}$$

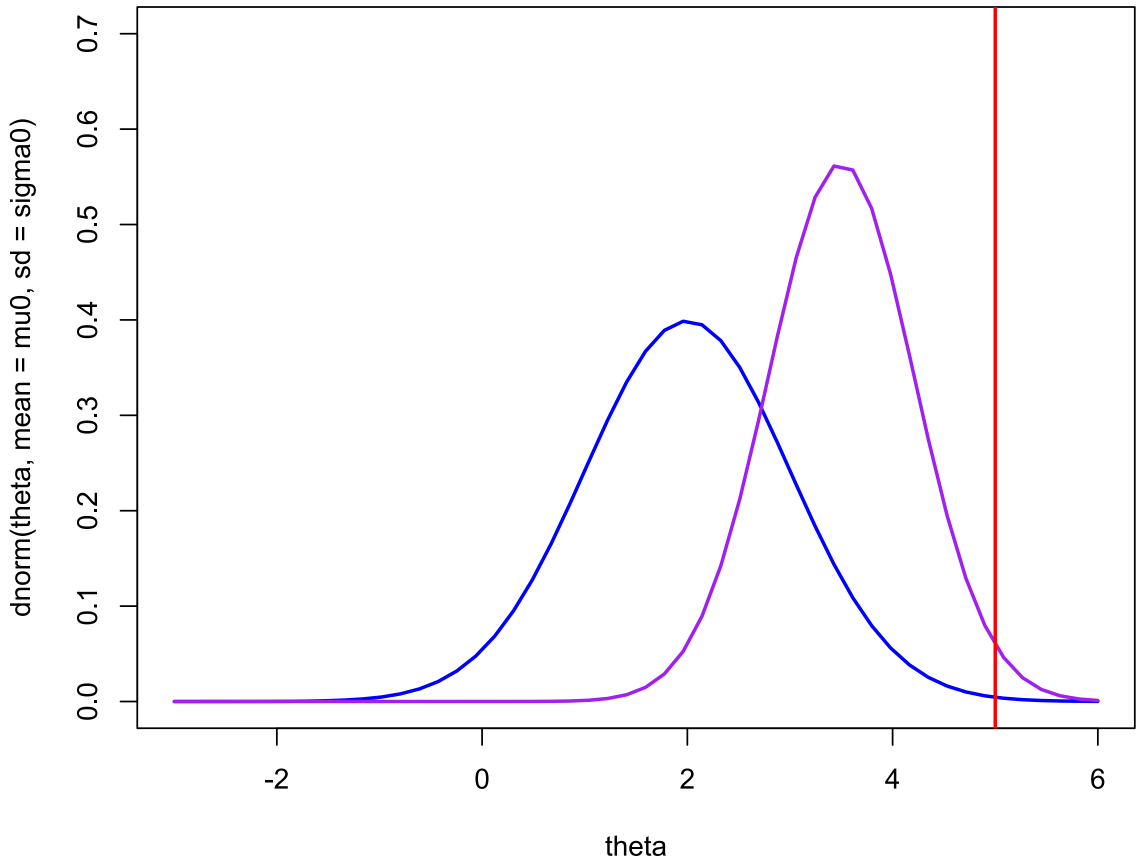
$$\mu_1 = \frac{a\mu_0 + b\bar{y}}{a+b}, \quad \sigma_1^2 = \frac{1}{a+b} \tag{2}$$

- The posterior mean μ_1 is a weighted average of the prior mean μ_0 and sample average $\bar{y}$.
- If n is large then the weight b is large and $\bar{y}$ will have a strong influence on the posterior. In fact if $n \to \infty$, $b/(a+b) \to 1$ and $a/(a+b) \to 0$, so $\mu_1 \to \bar{y}$.
- If σ_0^2 is small then the weight a is large and μ_0 will have a strong influence on the posterior

Board question

- Suppose our data follows a $N(\theta, 1)$ distribution with unknown mean θ.
- Suppose our prior on θ is $N(2, 1)$.
- Suppose we obtain data $x = 5$
- Compute the Bayesian update table and show that the posterior pdf for θ is Normal
- Find the posterior mean and the posterior variance
- Use the updating formulas (1) to find the posterior mean and posterior variance.

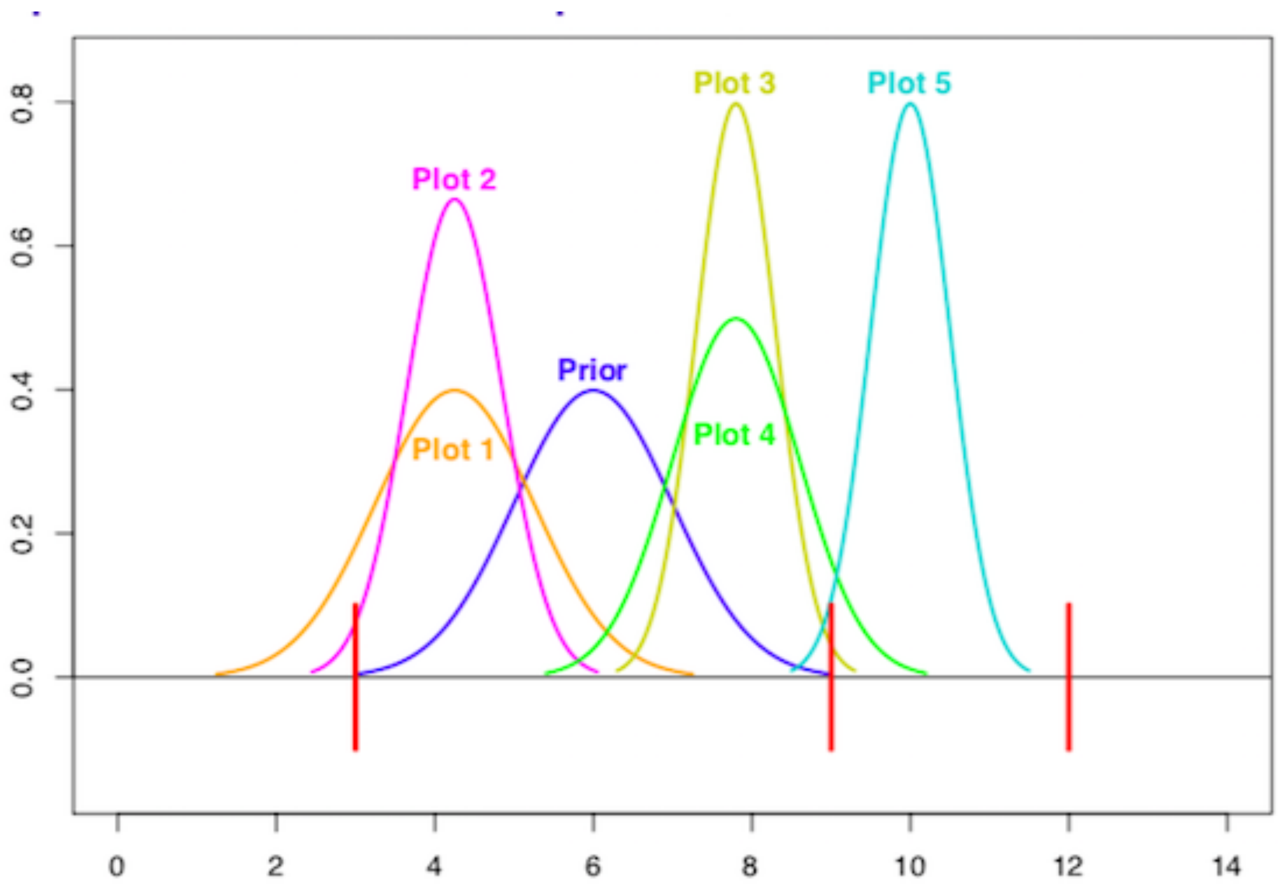
Board question



prior: blue, posterior: purple, $x = 5$ (data).

The posterior mean lies between the data $x = 5$ and the prior mean.

Board question



1. Which plot is the posterior to just the first data value $x = 3$?
2. Which plot is the posterior to all 3 data values, $x = 3$, $x = 9$ and $x = 12$?

Board question

On a basketball team the free throw percentage over all players is a $N(75, 36)$ distribution. In a given year individual players free throw percentage is $N(\theta, 16)$ where θ is their career average.

This season, Sophie Lee made 85 percent of her free throws.

1. What is the posterior expected values of her career percentage θ?

Exponential model

- The time until failure for a type of light bulb is exponentially distributed with parameter λ.
- We observe n bulbs, with failure times $t = t_1, \ldots, t_n$.
- The unknown parameter is λ.
- Can we find a conjugate family of distributions for this likelihood?

Conjugate priors

- A prior is conjugate to a likelihood if the posterior is the same type of distribution as the prior.

	hypothesis	data	prior	likelihood	posterior
Bernoulli/Beta	$\theta \in [0,1]$	x	Beta(α,β)	Bernoulli(θ)	Beta$(\alpha+1,\beta)$ or Beta$(\alpha,\beta+1)$
	θ	$x=1$	$c_1\theta^{\alpha-1}(1-\theta)^{b-1}$	θ	$c_3\theta^{\alpha}(1-\theta)^{\beta-1}$
	θ	$x=0$	$c_1\theta^{\alpha-1}(1-\theta)^{b-1}$	1-θ	$c_3\theta^{\alpha-1}(1-\theta)^{\beta}$
Binomial/Beta	$\theta \in [0,1]$	x	Beta(α,β)	binomial(n,θ)	beta$(\alpha+x,\beta+n-x)$
(fixed n)	θ	x	$c_1\theta^{\alpha-1}(1-\theta)^{b-1}$	$c_2\theta^{x}(1-\theta)^{n-x}$	$c_3\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}$
Normal/Normal	$\theta \in \mathbb{R}$	x	$N(\mu_0,\sigma_0^2)$	$N(\theta,\sigma^2)$	$N(\mu_1,\sigma_1^2)$
(fixed σ^2)	θ	$c_1\exp\{-\frac{1}{2\sigma_0^2}(\theta-\mu_0)^2\}$	x	$c_2\exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$	$c_3\exp\{-\frac{1}{2\sigma_1^2}(\theta-\mu_1)^2\}$

Board question

Which are conjugate priors for the following pairs likelihood/prior?

1. Exponential/Normal
2. Exponential/Gamma
3. Binomial/Normal

Board question

Suppose the prior has been set. Let x_1 and x_1 be two sets of data. Which of the following are true?

- If the likelihoods $f(x_1|\theta)$ and $f(x_2|\theta)$ are the same then they result in the same posterior.
- If x_1 and x_2 result in the same posterior then their likelihood functions are the same.
- If the likelihoods $f(x_1|\theta)$ and $f(x_2|\theta)$ are proportional then they result in the same posterior.
- If two likelihoods functions are proportional then they are equal.