Lecture 3A
MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

Today's agenda

Today's lecture will

- Conjugate priors
- Construct a posterior for continuous parameters and continuous data.

Conjugate prior distributions

- **Definition.** Suppose that we have data y generated from the likelihood function $p(y \mid \theta)$ depending on the unknown parameter θ. Also suppose the prior distribution $p(\theta)$ for θ is one of a family of parameterised distributions. If the posterior distribution for θ, $p(\theta \mid y)$ is in this family, we say the prior is a conjugate prior for the likelihood $p(y \mid \theta)$.

- For example, the beta distribution is a conjugate prior for the binomial likelihood/Bernoulli likelihood.

- Also, the beta distribution is a conjugate prior for the geometric likelihood

Conjugate prior distributions

Binomial likelihood: $p(k \mid q) = \binom{n}{k} q^k (1-q)^{n-k}$

Geometric likelihood: $p(k|q) = q(1-q)^k$.

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

- **Key:** The binomial, the geometric likelihoods and the beta distribution, considered as functions of q, are proportional to $q^r(1-q)^s$ for some $r > 0, s > 0$.
- When we multiply them together, we still have the same form.
- This is what characterises conjugate distributions.

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 $x \sim \underline{f(x|\theta)}$ x = observed data
- Likelihood: $f(x|\theta)$
- posterior pdf, $f(\theta|x) \propto f(\theta)\, f(x \mid \theta)$=prior $\times$ likelihood

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

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

Normal example, known variance

- $y_1, \ldots, y_n \sim N(\mu, \sigma^2)$.
- It's simpler if only one parameter is unknown.
- First, consider case where only μ is unknown.
- 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 $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$, μ_0, σ_0^2 are known.
- The posterior distribution is

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

where

$$\mu_1 = \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{y}}{\sigma^2}\right) \Big/ \left(\underbrace{\frac{1}{\sigma_0^2}}_{=a} + \underbrace{\frac{n}{\sigma^2}}_{=b}\right) \leftarrow \text{posterior mean}$$

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

Proof

We have $y_1, \dots, y_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 known.

Prior, $p(\mu) \sim N(\mu_0, \sigma_0^2)$, μ_0, σ_0^2 known.

The likelihood $p(y_1, \dots, y_n | \mu)$ is just the joint density of $y_1, \dots, y_n$. By independence,

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

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

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

- The sum in the exponential can be expanded as

$$\sum_{i=1}^{n}(y_i-\mu)^2 = \sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

$$= \sum_{i=1}^{n} y_i^2 - 2n\bar{y}\mu + n\mu^2$$

$$= S_2 - 2n\bar{y}\mu + n\mu^2, \text{ where}$$

$$S_2 = \sum_{i=1}^{n} y_i^2, \quad \bar{y} = n^{-1}\sum_{i=1}^{n} y_i \Rightarrow \sum_{i=1}^{n} y_i = n\bar{y}$$

Thus, the likelihood $p(y_1,..,y_n|\mu)$ becomes

$$p(y_1,..,y_n|\mu) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(S_2 - 2n\bar{y}\mu + n\mu^2\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\left(S_2 - 2n\bar{y}\mu + n\mu^2\right)\right\}.$$

The prior, $p(\mu)$, is

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

The posterior, $p(\mu|y_1,,y_n)$, is

posterior $\propto$ prior x likelihood

$= p(\mu) \times p(y_1,..,y_n|\mu)$

$e^{x+y} = e^x \cdot e^y$

$$\text{posterior} \propto \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

$$\times \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(S_2 - 2n\bar{y}\mu + n\mu^2\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma_0^2}\left(\mu^2 - 2\mu\mu_0 + \mu_0^2\right)\right\} \exp\left\{-\frac{1}{2\sigma^2}\left(S_2 - 2n\bar{y}\mu + n\mu^2\right)\right\}$$

$$= \exp\left\{-\underbrace{\frac{S_2}{2\sigma^2}}_{\text{constant}} + \frac{n\bar{y}\mu}{\sigma^2} - \frac{n\mu^2}{2\sigma^2} - \frac{\mu^2}{2\sigma_0^2} + \frac{\mu\mu_0}{\sigma_0^2} - \underbrace{\frac{\mu_0^2}{2\sigma_0^2}}_{\text{constant}}\right\}$$

$$= \exp\left\{-\left(\frac{n}{2\sigma^2} + \frac{1}{2\sigma_0^2}\right)\mu^2 + \left(\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)\mu + C\right\}$$

where $C = -\frac{S_2}{2\sigma^2} - \frac{\mu_0^2}{2\sigma_0^2} \rightarrow$ constant

$$= c_1 \exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right\}$$

$$= c_1 \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\left[\mu^2 - 2\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1}\left(\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right\}$$

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

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

$$= c_1 \exp\left\{-\frac{1}{2\sigma_1^2}\left(\mu^2 - 2\mu_1\mu\right)\right\}$$

$\exp\left(\frac{-1}{2\sigma_1^2}(\mu-\mu_1)^2\right)$

$$= c_1 \exp\left\{-\frac{1}{2\sigma_1^2}\left(\mu^2 - 2\mu_1\mu + \mu_1^2 - \mu_1^2\right)\right\}$$

completing the square

$$= c_1 \exp\left\{-\frac{1}{2\sigma_1^2}\left(\mu^2 - 2\mu_1\mu + \mu_1^2\right)\right\}\underbrace{\exp\left\{+\frac{\mu_1^2}{2\sigma_1^2}\right\}}_{\text{constant}}$$

$$= c_2 \exp\left\{-\frac{1}{2\sigma_1^2}(\mu - \mu_1)^2\right\}$$

We recognize this to be $N(\mu_1, \sigma_1^2)$

The posterior, $p(\mu | y_1, \dots, y_n)$, is $\mathcal{N}(\mu_1, \sigma_1^2)$.

The normalising constant is $\frac{1}{\sqrt{2\pi\sigma_1^2}}$

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 $\bar{y}$. ($\bar{y}$ is the MLE of μ)
- 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

$$\mu_1 = \frac{a}{a+b}\mu_0 + \frac{b}{a+b}\bar{y}$$

set $\quad w = \underline{\frac{a}{a+b}}, \qquad 1-w = \underline{\frac{b}{a+b}}$

$$\mu_1 = w\mu_0 + (1-w)\bar{y} \rightarrow \text{weighted average.}$$

Board question

- Suppose our data follows a $N(\theta, 1)$ distribution with unknown mean θ. $\sigma^2 = 1$
- Suppose our prior on θ is $N(2, 1)$. $\mu_0 = 2, \ \sigma_0^2 = 1$
- Suppose we obtain data $x = 5$ $\sim N(\theta, 1)$ $(n=1)$
- 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 (??) to find the posterior mean and posterior variance.

7-2

$$\left\{ \begin{array}{l} \sigma_0^2 = 1, \ n = 1, \ \sigma^2 = 1 \\ \mu_0 = 2, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{1}\sum_{i=1}^{1} 5 = 5 \end{array} \right\} \text{use 7-2 formulas}$$

Solution

- Our prior is

$$\underline{f(\theta)} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(\theta-2)^2\right\}.$$

Our likelihood is $(x=5)$

$$f(5|\theta) = \frac{1}{\sqrt{2\pi}} \underbrace{\exp\left\{-\frac{1}{2}(5-\theta)^2\right\}}_{f(\underline{x}) \sim N(\theta,1)}$$

The posterior density is

$p(\theta|x=5) \propto$ likelihood $\times$ prior

$$\propto \exp\left\{-\frac{1}{2}(5-\theta)^2\right\}\exp\left\{-\frac{1}{2}(\theta-2)^2\right\}$$

$$= c\exp\left\{-\frac{1}{2}(5-\theta)^2 - \frac{1}{2}(\theta-2)^2\right\}$$

$$= c\exp\left\{-\frac{1}{2}\left(\theta^2 - 4\theta + 4 + 25 - 10\theta + \theta^2\right)\right\}$$

$$= c\exp\left\{-\frac{1}{2}\left(2\theta^2 - 14\theta + 29\right)\right\}$$

$$= c_1\exp\left\{-\left(\theta^2 - 7\theta + 29/2\right)\right\} \qquad \underbrace{c_1\exp\left(-\frac{29}{2}\right)}_{c_2}$$

$$= c_2\exp\left\{-\left(\theta^2 - 7\theta\right)\right\} \leftarrow \text{complete the square} \qquad \left(\theta - \frac{7}{2}\right)^2$$

$$= c_2\exp\left\{-\left(\theta^2 - 2\cdot\frac{7}{2}\theta\right)\right\}$$

Completing the square requires multiplying with $\exp\left(-\frac{49}{4}\right)\exp\left(\frac{49}{4}\right)=1$

$$=c_2\exp\left\{-\left(\theta^2-2\tfrac{7}{2}\theta\right)\right\}\exp\left\{-\tfrac{49}{4}\right\}\exp\left\{\tfrac{49}{4}\right\}$$

$$=c_2\exp\left\{-\left(\theta^2-2\cdot\tfrac{7}{2}\theta+\tfrac{49}{4}\right)\right\}\exp\left\{49/4\right\}$$

$$=c_3\exp\left\{-1\left(\theta-7/2\right)^2\right\}$$

$$\exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2\right\}$$

$$=c_3\exp\left\{-\frac{1}{2(1/2)}\left(\theta-7/2\right)^2\right\}$$

$$\sigma_1^2=\frac{1}{2}$$

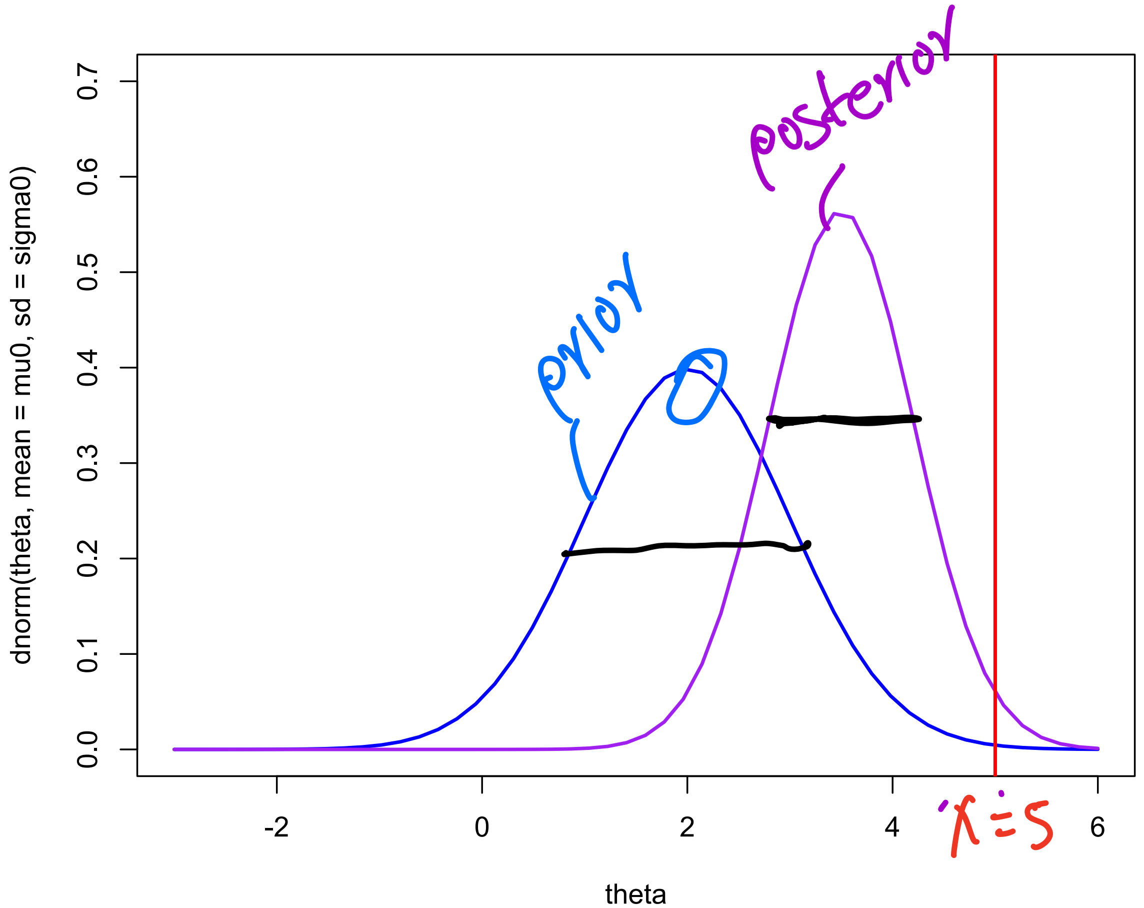
so we recognise this to have the form of the $N\left(\frac{7}{2},\frac{1}{2}\right)$. The normalising constant must be

$$1/\sqrt{2\pi\sigma^2}=\frac{1}{\sqrt{2\pi\frac{1}{2}}}=\frac{1}{\sqrt{\pi}}$$

$$\theta\mid x=5\sim N\left(\frac{7}{2},\frac{1}{2}\right)$$

- $\mu_1=\frac{7}{2}$ (posterior mean)
- $\sigma_1^2=1/2$ (posterior variance)

Board question



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

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

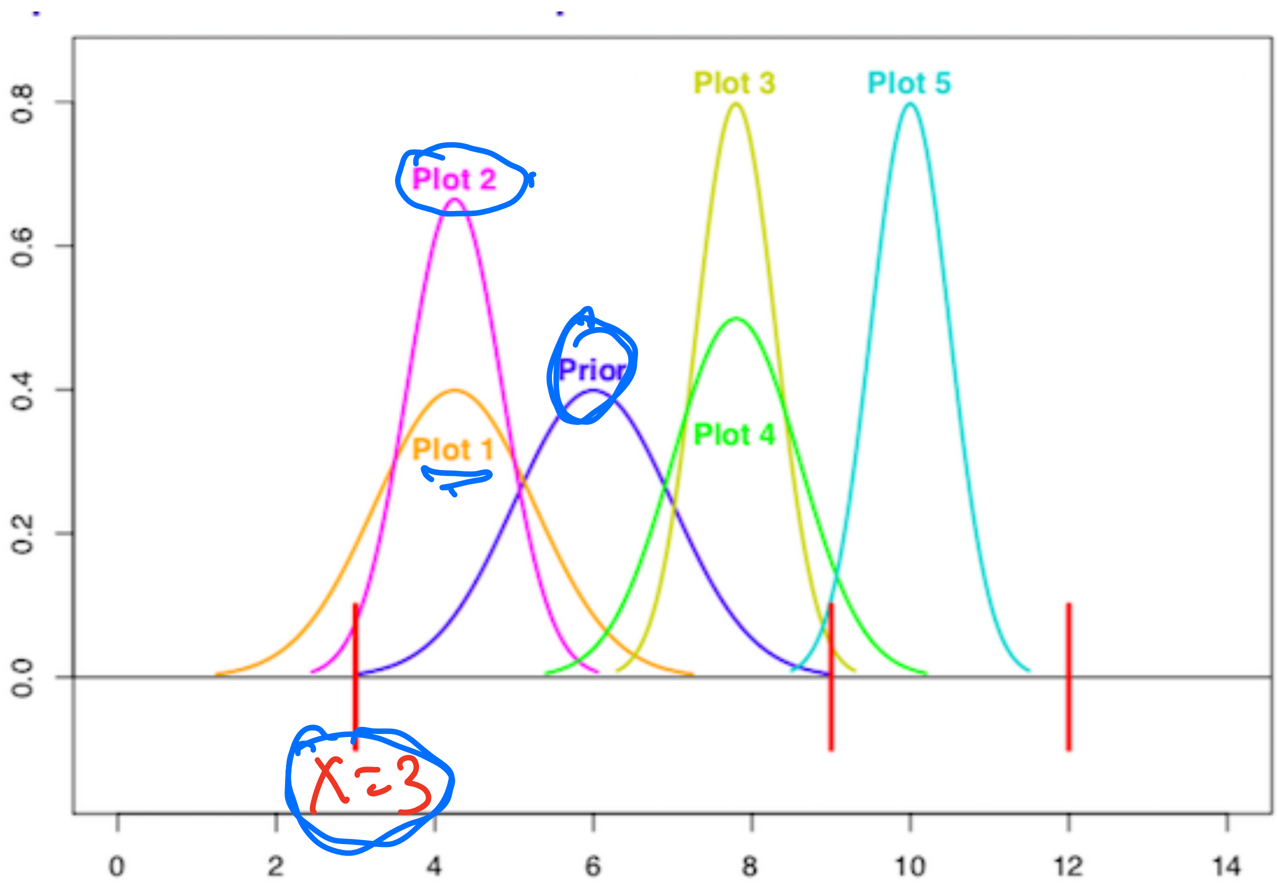
$$\mu_1 = \frac{a}{a+b}\mu_0 + \frac{b}{a+b}\bar{y}$$

$$\bar{y} = 5$$

$$\mu_1 = \frac{a}{a+b}\mu_0 + \frac{b}{a+b}5$$

$$\sigma_1^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \overset{n=1}{=} \frac{1}{\underbrace{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}}} = \sigma_0^2 \underbrace{\frac{\sigma^2}{\sigma^2+\sigma_0^2}}_{<1} < \sigma_0^2$$

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 $x = 12$?

- The posterior density must have its mean between $x=3$ and the prior mean, μ_0. The only possibilities for this are plots 1 and 2.

We also know, for $n=1$

$$\sigma_1^2 = \sigma_0^2 \times \underbrace{\frac{\sigma^2}{\sigma^2+\sigma_0^2}}_{<1} < \sigma_0^2$$

The posterior variance is less than the prior variance. Between plots 1 and 2, only plot 2 has smaller variance than the prior.

- Use the same arguments.

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 $t = t_1, \ldots, t_n$. (iid)
- The unknown parameter is λ.
- Can we find a conjugate family of distributions for this likelihood?

Solution

We have $t_1, \ldots, t_n \overset{iid}{\sim} Exp(\lambda)$, where λ is unknown.
The likelihood of $t_1, \ldots, t_n$ is just the joint density of $t_1, \ldots, t_n$. By independence,

$$p(t_1, \ldots, t_n | \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}$$

$$= \lambda^n e^{-\lambda S} \text{ where}$$

$$S = \sum_{i=1}^{n} t_i$$

A condidate prior distribution for λ using the exponential likelihood is the gamma distribution, Gamma (a, β). If $\lambda \sim$ Gamma(a, β), the density is

$$p(\lambda) = \frac{\beta^a \lambda^{a-1} e^{-\beta\lambda}}{\Gamma(a)}, \quad \lambda > 0$$

The posterior is

$$p(\lambda | t_1, t_n) \propto \text{likelihood} \times \text{prior}$$

$$= C_1 \lambda^n e^{-\lambda S} \times \frac{\beta^a \lambda^{a-1} e^{-\beta\lambda}}{\Gamma(a)}$$

$$= C_2 \lambda^{n+a-1} e^{-(\beta+S)\lambda}$$

We recognise this to have the same form with Gamma $(n+a, S+\beta)$

The normalising constant is just the normalising constant of the gamma density. In this case is

$$\frac{(S+\beta)^{n+a}}{\Gamma(n+a)}$$

$$p(\lambda | t) \sim \text{Gamma}(n+a, S+\beta)$$

The Gamma distribution for λ is a conjugate prior for the exponential 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.