Lecture 3B
MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

Today's agenda

Today's lecture

- Conjugate priors for the normal likelihood
- Update a gamma prior for the inverse of the normal variance given a normal likelihood with known mean and unknown variance.

Normal example, known variance

- $y_1, \ldots, y_n \sim N(\mu, \sigma^2)$, μ is unknown, σ^2 is known
- We saw that a normal prior distribution for μ is conjugate in this case.
- It's conjugate because it results in a posterior in the same family as the prior.

Normal example, known variance

- Observed data $y_1, \ldots, y_n \sim N(\mu, \sigma^2)$.
- Prior distribution $\mu \sim N(\mu_0, \sigma_0^2)$. μ_0, σ_0^2 are known
- The posterior distribution is

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

where

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

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

Normal example, known variance

- If we set $a = \frac{1}{\sigma_0^2}$, $b = \frac{n}{\sigma^2}$, we can rewrite μ_1 and σ_1^2 as

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

- 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.
- 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

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?

Solution

This is a normal/normal conjugate/likelihood pair example.

Parameter of interest is θ, the average free throw percentage.

We assume $\theta \sim N(75, 36)$, the prior distrib.

My data, X is an individual's free throw percentage in a given year

We observed $X = 85$, Sophie's this year free throw percentage, $X \sim N(\theta, 16) \rightarrow$ likelihood

We want to find the posterior mean

By the normal-normal updating formulas (7),

$$a = \frac{1}{\sigma_0^2} = \frac{1}{36}, \quad b = \frac{n}{\sigma^2} = \frac{1}{16}$$

$$\mu_0 = 75, \quad X = 85$$

Thus,

$$\mu_1 = \frac{a\mu_0 + bX}{a+b} = \frac{(1/36)75 + (1/16)85}{1/36 + 1/16}$$

$$= 81.9$$

The posterior variance is

$$\sigma_1^2 = \frac{1}{a+b} = \frac{1}{1/136 + 1/16} = 11.1$$

My posterior density is $N(81.9, 11.1)$ for θ

or $N(82, 11)$

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

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

$$= \frac{\text{Bayes numerator}}{\text{total prob. of data}}$$

$p(\theta|x)$ is a density so it must satisfy:

- $p(\theta|x) \geq 0 \quad \forall\, \theta$

- $\displaystyle\int_\theta p(\theta|x)\, d\theta = 1$

p(x) is the normalising constant because

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

$$= \frac{1}{p(x)} \int \underbrace{p(\theta)\,p(x|\theta)\,d\theta}_{p(x)\ \text{(by definition)}} = \frac{\cancel{p(x)}}{\cancel{p(x)}} = 1$$

because $p(x) = \int p(\theta)\,p(x|\theta)\,d\theta$

For the normal example, we found that

$$\underbrace{p(\theta) \times p(\theta|x)} = \underbrace{c_1 \exp\left(-\frac{(\theta-\mu_1)^2}{2\sigma_1^2}\right)}$$

To be a density we need to calculate $p(x)$ and divide the Bayes numerator with $p(x)$.

You will find,

$$\frac{p(\theta) \times p(\theta|x)}{p(x)} = \frac{c_1 \exp\left(-\frac{(\theta-\mu_1)^2}{2\sigma_1^2}\right)}{p(x)}$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} \underbrace{\exp\left(-\frac{1}{2\sigma_1^2}(\theta-\mu_1)^2\right)}$$

we know that $\theta \sim N(\mu_1, \sigma_1^2)$

Normal example, known mean

- Observed data $y_1, \ldots, y_n \sim N(\mu, \sigma^2)$.
- Suppose that μ is known and σ is unknown.
- It is easier to work with $\tau = 1/\sigma^2$.
- τ is known as the precision.
- τ is the reciprocal of the variance, so large τ means small variance and hence high precision.

Normal example, known mean

- In this case, a gamma prior is conjugate for τ.
- Prior $\tau \sim \text{Gamma}(\alpha, \beta)$.
- Posterior

$$\tau \sim \text{Gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2}\right)$$

μ is known

- The family of gamma distributions is conjugate to the normal likelihood for the normal precision parameter τ, if the mean μ is known.

Proof

The likelihood, $p(y_1,.., y_n|\sigma^2)$, is (by independence)

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

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

Let $\tau = 1/\sigma^2$. Then

$$p(y_1,.., y_n|\tau) = \left(\frac{\tau}{2\pi}\right)^{n/2} \cdot \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}$$

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

$X \sim \text{Gamma}(\alpha, \beta)$

$$f(x) \propto \underbrace{x^{\alpha-1} \exp(-\beta x)}$$

We see that the likelihood has the form of a gamma distribution.

Hence, we can take $\tau \sim$ Gamma(α, β) for the conjugate prior

$$p(\tau) = \frac{\beta^{\alpha} \tau^{\alpha-1} e^{-\beta\tau}}{\underbrace{\Gamma(\alpha)}}, \quad \tau > 0$$

— constant

where α, β are known.

The posterior density is

$p(\tau \mid y_1, \dots, y_n) \propto$ prior $\times$ likelihood

$$= \tau^{\alpha-1} \exp(-\beta\tau) \times \tau^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}$$

$\Rightarrow$

My posterior density, $p(\theta|y_1,..,y_n)$ is

$$p(\theta|y_1,..,y_n) \propto \tau^{\underline{n/2+a-1}} \times \underbrace{\exp\left\{-\underline{\tau}\left[\frac{\sum^n (y_i-\mu)^2}{2}+b\right]\right\}}$$

We recognize this to be the gamma density with parameters

- $\frac{n}{2}+a$
- $\underbrace{\frac{\sum^n (y_i-\mu)^2}{2}+\underline{b}}$

posterior parameters

Normal example, both parameters unknown

- If μ and τ are unknown then there is a bivariate distribution which is conjugate.
- Marginal distribution
$$\tau \sim \text{Gamma}$$
and conditional distribution
$$\mu \mid \tau \sim \text{Normal}.$$
- The joint prior distribution is the product of these two.
- The posterior is of the same form.
- We're not going into details in this module.