

STT 465

Normal Model

Normal Model (I)

⇒ Normal Model

⇒ Likelihood Function

⇒ Maximum Likelihood Estimation

⇒ Bayesian Model: Conditional on the variance

- Model
- Derivation of the posterior distribution
- Posterior mean and posterior variance
- Comparison with MLE (variance, bias & MSE)
- Predictive distribution

Normal Model

=> Conditional distribution of the data: $y_i \stackrel{iid}{\sim} N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2}$

[Note: $y_i = \mu + \varepsilon_i$ $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ => $y_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$]

⇒ For a sample of size n we have:

$$p(y_1, \dots, y_n | \mu, \sigma^2) = N(\mu, \sigma^2) = [2\pi\sigma^2]^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2}$$

⇒ Maximum likelihood estimation (steps):

- Take logarithm,
- Take derivatives with respect to each parameters
- (1st Order Conditions, FOCs) Set each of the derivatives equal to zero
- Solve for the parameters
- Check the sign of 2nd order derivatives.

Normal Model

⇒ The MLEs are (derivation presented in class):

$$\hat{\mu}_{MLE} = \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{SS_y}{n}$$

⇒ Discuss: Bias and Variance of the MLE; approximate 95% CI.

Bayesian Inference (conditional on the variance)

(I) Likelihood

$$p(y_1, \dots, y_n | \mu, \sigma^2) = N(\mu, \sigma^2) = [2\pi\sigma^2]^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2}$$

(II) A conjugate Prior for the Mean

$$\mu \sim N(\mu_0, \sigma_0^2) = [2\pi\sigma_0^2]^{-1/2} e^{-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2}$$

(II) Posterior distribution

$$\begin{aligned} p(\mu | y_1, \dots, y_n, \sigma^2) &\propto p(y_1, \dots, y_n | \mu, \sigma^2) p(\mu) \\ &\propto [2\pi\sigma^2]^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2} \times [2\pi\sigma_0^2]^{-1/2} e^{-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2} \\ &\propto e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2} e^{-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2} \end{aligned}$$

[Combine the two quadratic forms, remove terms that do not involve the mean]

Bayesian Inference (conditional on the variance)

(II) Posterior distribution [cont]

$$p(\mu | y_1, \dots, y_n, \sigma^2) \propto e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2} e^{-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2}$$

[expand the quadratic forms]

$$\propto e^{-\frac{1}{2} \left[\sigma^{-2} \sum_i y_i^2 + \frac{n}{\sigma^2} \mu^2 - 2\mu \frac{n\bar{y}}{\sigma^2} \right]} e^{-\frac{1}{2} \left[\frac{\mu^2}{\sigma_0^2} + \frac{\mu_0^2}{\sigma_0^2} - 2\mu \frac{\mu_0}{\sigma_0^2} \right]}$$

[remove terms that do not involve μ]

$$\propto e^{-\frac{1}{2} \left[\frac{n}{\sigma^2} \mu^2 - 2\mu \frac{n\bar{y}}{\sigma^2} \right]} e^{-\frac{1}{2} \left[\frac{\mu^2}{\sigma_0^2} - 2\mu \frac{\mu_0}{\sigma_0^2} \right]}$$

[combine terms]

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

[re-arrange terms]

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

[re-arrange terms]

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

$$p(\mu | y_1, \dots, y_n, \sigma^2) \propto e^{-\frac{1}{2} C \left[\mu^2 - 2\mu \frac{rhs}{C} \right]}$$

Bayesian Inference (conditional on the variance)

$$\begin{aligned} p(\mu | y_1, \dots, y_n, \sigma^2) &\propto e^{-\frac{1}{2}C[\mu^2 - 2\mu\frac{rhs}{C}]} \quad [\text{multiply by a constant}] \\ &\propto e^{-\frac{1}{2}C[\mu^2 - 2\mu\frac{rhs}{C}]} e^{-\frac{1}{2}C[\frac{rhs}{C}]^2} \quad [combine] \\ &\propto e^{-\frac{1}{2}C[\mu^2 - 2\mu\frac{rhs}{C} + [\frac{rhs}{C}]^2]} \\ &\propto e^{-\frac{1}{2}C\left(\mu^2 - \frac{rhs}{C}\right)^2} \\ &= N\left(\frac{rhs}{C}, C^{-1}\right) \quad [1] \end{aligned}$$

Compare MLE & Bayesian

$$\hat{\mu}_{MLE} = \bar{y}$$

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

- => In this case the MLE is un-biased
- => The Bayesian estimate is a weighted sum of the prior mean and the MLE.
- => The posterior mean 'shrinks' the MLE towards the prior mean.
- => The extent of shrinkage depends on: sampling variance (the variance of the error terms), the prior variance, and sample size. As n increase the Bayesian estimate goes to MLE.

See Example in GitHub (examples_5.md)

Normal Model (conditional on the mean)

- Likelihood (viewed as a function of the variance, for fixed mean)

$$p(y_1, \dots, y_n \mid \mu, \sigma^2) \propto [2\pi\sigma^2]^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2}$$
$$\propto [\sigma^2]^{-n/2} e^{-\frac{SS_y}{2\sigma^2}}$$

- Conjugate prior ('Scaled Inverse Chi-square'; discuss: Gamma, Inverse-Gamma...)

$$p(\sigma^2 \mid DF_0, S_0) = \left[\frac{S_0}{2} \right]^{\frac{DF_0}{2}} \left[\frac{1}{\Gamma(DF_0/2)} \right] [\sigma^2]^{-(DF_0/2+1)} e^{-\frac{S}{2\sigma^2}} \propto [\sigma^2]^{-(DF_0/2+1)} e^{-\frac{S_0}{2\sigma^2}}$$

Normal Model (conditional on the mean)

⇒ Posterior density

$$\begin{aligned} p(\sigma^2 \mid \mu, y_1, \dots, y_n) &\propto [\sigma^2]^{-n/2} e^{-\frac{SS_y}{2\sigma^2}} [\sigma^2]^{-(DF/2+1)} e^{-\frac{S_0}{2\sigma^2}} \\ &\propto [\sigma^2]^{-[(DF_0 + N)/2 + 1]} e^{-\frac{SS_y + S_0}{2\sigma^2}} \\ &= \chi^{-2}(n + DF_0 \ ; \ SS_y + S_0) \end{aligned} \quad [2]$$

Joint inference on the mean and variance

⇒ We have discuss the posterior distributions of:

- The mean given the variance
- The variance given the mean

⇒ The above are called the fully-conditional distributions

$$p(\mu | ELSE) = p(\mu | \sigma^2, y_1, \dots, y_n) \quad [1]$$

$$p(\sigma^2 | ELSE) = p(\sigma^2 | \mu, y_1, \dots, y_n) \quad [2]$$

⇒ Our goal is to draw samples from the joint posterior distribution

$$p(\mu, \sigma^2 | y_1, \dots, y_n) = \frac{p(y_1, \dots, y_n | \mu, \sigma^2) p(\mu, \sigma^2)}{p(y_1, \dots, y_n)} \propto p(y_1, \dots, y_n | \mu, \sigma^2) p(\mu, \sigma^2)$$

Composition Sampling (section 5.3)

⇒ Recall that

$$p(x_1, x_2) = p(x_2 | x_1)p(x_1) = p(x_1 | x_2)p(x_2)$$

⇒ Similarly

$$p(\mu, \sigma^2 | y_1, \dots, y_n) = p(\mu | \sigma^2, y_1, \dots, y_n)p(\sigma^2 | y_1, \dots, y_n)$$

Or

$$p(\mu, \sigma^2 | y_1, \dots, y_n) = p(\sigma^2 | \mu, y_1, \dots, y_n)p(\mu | y_1, \dots, y_n)$$

Composition Sampling:

- Sample on unknown from its marginal posterior distribution
- Sample the other parameter from its fully-conditional distribution

⇒ We have already derived the fully conditionals, for implementing composition sampling we need to find one of the marginal posterior distributions.

Marginal posterior distribution of the variance (section 5.3)

⇒ Joint Posterior Distribution

$$p(\mu, \sigma^2 | y_1, \dots, y_n) = p(\mu | \sigma^2, y_1, \dots, y_n) p(\sigma^2 | y_1, \dots, y_n)$$

⇒ Marginal Posterior Distribution

$$\begin{aligned} p(\sigma^2 | y_1, \dots, y_n) &= \int p(\mu, \sigma^2 | y_1, \dots, y_n) d\mu \\ &\propto \int p(y_1, \dots, y_n | \mu, \sigma^2) p(\sigma^2) p(\mu) d\mu \\ &\propto p(\sigma^2) \int p(y_1, \dots, y_n | \mu, \sigma^2) p(\mu) d\mu \end{aligned}$$

⇒ If, without loss of generality, we set $\sigma_0^2 = \sigma^2 / k_0$

$$p(\sigma^2 | y_1, \dots, y_n) = \chi^{-2}(DF_0 + n ; S_0 + SS_y)$$

Where $SS_y = \sum_i (y_i - \bar{y})^2 + \frac{k_0}{(k_0 + n)} n(\bar{y} - \mu_0)^2$ [compare with fully-conditional dist.]

Composition Sampling (section 5.3)

⇒ Algorithm (composition sampling):

- (i) Sample the error variance from its marginal posterior dist. [3]
- (ii) Sample the mean from the fully conditional dist., [2]
- (iii) Repeat (i) & (ii) B times.

[See examples_5.md in GitHub]

⇒ Intro to Gibbs sampler

- Composition sampling gives IID samples from the joint posterior dist.
- Gibbs sampling
 - (0) Initialize the mean (e.g., set the μ =sample mean)
 - (i) Sample the variance from its fully conditional dist. [2]
 - (ii) Sample the mean from its fully conditional dist. [3]
 - (iii) Repeat (i) and (ii) B times.

The above algorithm renders samples from the joint posterior
(note these are not IID)

[See examples_5.md in GitHub]