STT 465

Normal Model

Normal Model (I)

$\Longrightarrow$ Normal Model

$\Longrightarrow$ Likelihood Function

$\Longrightarrow$ Maximum Likelihood Estimation

$\Longrightarrow$ 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:

$$y_i \overset{iid}{\sim} N\left(\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2}$$

$\Rightarrow$ For a sample of size n we have:

$$p\left(y_1,\ldots,y_n \mid \mu,\sigma^2\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} = \left[2\pi\sigma^2\right]^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$

$\Rightarrow$ Maximum likelihood estimation (steps):

- Take the log,
- Take derivatives with respect to each parameter
- (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):

$$\hat{\mu}_{MLE} = \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}{n} = \frac{SS_y}{n}$$

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

Bayesian Inference (conditional on the variance)

(I) Likelihood

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

(II) A conjugate Prior for the Mean

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

(II) Posterior distribution

$$p\left(\mu \mid y_1,\ldots,y_n,\sigma^2\right)\propto p\left(y_1,\ldots,y_n \mid \mu,\sigma^2\right)p\left(\mu\right)$$

$$\propto\left[2\pi\sigma^2\right]^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\mu\right)^2}\times\left[2\pi\sigma_0^2\right]^{-1/2} e^{-\frac{1}{2\sigma_0^2}\left(\mu-\mu_0\right)^2}$$

$$\propto e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\mu\right)^2} e^{-\frac{1}{2\sigma_0^2}\left(\mu-\mu_0\right)^2}$$

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

Steps required to drive the posterior distribution of the mean

(II) Posterior distribution [cont]

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

[expand the quadratic forms]

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

[remove terms that do not invovle μ]

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

[combine terms]

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

[re-arange terms]

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

[re-arange terms]

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

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

Bayesian Inference (conditional on the variance)

$$p\left(\mu \mid y_1,\dots,y_n,\sigma^2\right) \propto e^{-\frac{1}{2}C\left[\mu^2-2\mu\frac{rhs}{C}\right]} \quad \text{[multiply by a constant]}$$

$$\propto e^{-\frac{1}{2}C\left[\mu^2-2\mu\frac{rhs}{C}\right]} e^{-\frac{1}{2}C\left[\frac{rhs}{C}\right]^2} \quad [combine]$$

$$\propto e^{-\frac{1}{2}C\left[\mu^2-2\mu\frac{rhs}{C}+\left[\frac{rhs}{C}\right]^2\right]}$$

$$\propto e^{-\frac{1}{2}C\left(\mu^2-\frac{rhs}{C}\right)^2}$$

$$= N\left(\frac{rhs}{C}, C^{-1}\right) \qquad [1]$$

Compare MLE & Bayesian

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

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

=> In this case the MLE is unbiased

=> 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)

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

$$\propto \left[\sigma^2\right]^{-n/2} e^{-\frac{SS_y}{2\sigma^2}}$$

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

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

Normal Model (conditional on the mean)

$\Rightarrow$ Posterior density

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

Joint inference on the mean and variance

$\Rightarrow$ We have discuss the posterior distributions of:

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

$\Rightarrow$ The above are called the fully-conditional distributions

$$p\left(\mu \mid ELSE\right) = p\left(\mu \mid \sigma^2, y_1, \ldots, y_n\right) \qquad [1]$$
$$p\left(\sigma^2 \mid ELSE\right) = p\left(\sigma^2 \mid \mu, y_1, \ldots, y_n\right) \qquad [2]$$

$\Rightarrow$ Our goal is to draw samples from the joint posterior distribution

$$p\left(\mu, \sigma^2 \mid y_1, \ldots, y_n\right) = \frac{p\left(y_1, \ldots, y_n \mid \mu, \sigma^2\right) p\left(\mu, \sigma^2\right)}{p\left(y_1, \ldots, y_n\right)} \propto p\left(y_1, \ldots, y_n \mid \mu, \sigma^2\right) p\left(\mu, \sigma^2\right)$$

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

$$p\left(\mu,\sigma^2 \mid y_1,\dots,y_n\right)=p\left(\mu \mid \sigma^2,y_1,\dots,y_n\right)p\left(\sigma^2 \mid y_1,\dots,y_n\right)$$

Or

$$p\left(\mu,\sigma^2 \mid y_1,\dots,y_n\right)=p\left(\sigma^2 \mid \mu,y_1,\dots,y_n\right)p\left(\mu \mid y_1,\dots,y_n\right)$$

Composition Sampling:

- Sample one 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)

$\Longrightarrow$ Joint Posterior Distribution

$$p\left(\mu,\sigma^2 \mid y_1,\ldots,y_n\right) = p\left(\mu \mid \sigma^2, y_1,\ldots,y_n\right) p\left(\sigma^2 \mid y_1,\ldots,y_n\right)$$

$\Longrightarrow$ Marginal Posterior Distribution

$$\begin{aligned} p\left(\sigma^2 \mid y_1,\ldots,y_n\right) &= \int p\left(\mu,\sigma^2 \mid y_1,\ldots,y_n\right) d\mu \\ &\propto \int p\left(y_1,\ldots,y_n \mid \mu,\sigma^2\right) p\left(\sigma^2\right) p(\mu)\, d\mu \\ &\propto p\left(\sigma^2\right) \int p\left(y_1,\ldots,y_n \mid \mu,\sigma^2\right) p(\mu)\, d\mu \end{aligned}$$

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

$$p\left(\sigma^2 \mid y_1,\ldots,y_n\right) = \chi^{-2}\left(DF_0 + n \ ; \ S_0 + SS\right)$$

Where $SS = \sum_i \left(y_i - \bar{y}\right)^2 + \dfrac{k_0}{(k_0+n)} n(\bar{y} - \mu_0)$ [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 mu=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]