01/18/21

- HW#1 : Due 01/20
- HW#2 : Due 01/28 (Fri) 5pm
- Midterm 1 : 02/03 (Th) - not 02/01

♣ Example (contd): Assume that observations, $x_1, \ldots, x_n$ are iid from $\mathrm{N}(\theta, \sigma^2)$, where μ and σ^2 are unknown. Consider $\tilde{\pi}(\theta, \sigma) = 1/\sigma^2$ (not a probability density, i.e., improper prior).

- Find the joint posterior distribution, $\tilde{\pi}(\theta, \sigma^2 \mid \boldsymbol{x})$.

- Find the posterior distributions $\tilde{\pi}_1(\theta \mid \boldsymbol{x}, \sigma^2)$ and $\tilde{\pi}_2(\sigma^2 \mid \boldsymbol{x})$.

- Find the marginal posterior distribution of θ, $\tilde{\pi}(\theta \mid \boldsymbol{x})$.
  ** Also, read Hoff §5.3, BDA §3.2, and Robert §4.4.1-4.4.2.

$$X_i \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$$

$$\tilde{\pi}(\theta, \sigma^2) = \frac{1}{\sigma^2}, \qquad \theta \in \mathbb{R} \quad \& \quad \sigma^2 \in \mathbb{R}^+$$

Likelihood

$$f(\mathbf{x} \mid \theta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\theta)^2}{2\sigma^2}\right)$$

$\ell(\theta, \sigma^2 \mid \mathbf{x})$

$$\propto (\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{S^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right)$$

where $\underbrace{\bar{x} = \sum_{i=1}^{n} x_i/n \quad \& \quad S^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2}_{\text{sufficient statistics}}$

① $$\tilde{\pi}(\theta, \sigma^2 \mid \mathbf{x}) = \tilde{\pi}(\theta, \sigma^2 \mid \bar{x}, S^2)$$

$$\propto f(\bar{x}, S^2 \mid \theta, \sigma^2) \cdot \tilde{\pi}(\theta, \sigma^2)$$

$$\propto (\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{S^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right) \cdot \frac{1}{\sigma^2}$$

* $\underline{m(\mathbf{x})} < \infty \Rightarrow$ posterior distri. exists

② $$\tilde{\pi}_1(\theta \mid \mathbf{x}, \sigma^2) \propto \tilde{\pi}(\theta, \sigma^2 \mid \mathbf{x})$$

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

$$\Rightarrow \quad \theta \mid \underline{x}, \underline{\sigma^2} \sim N\left(\bar{x}, \sigma^2/n\right)$$

$$\tilde{\pi}_2(\sigma^2 \mid \mathbf{x}) = \int_{\mathbb{R}} \tilde{\pi}(\theta, \sigma^2 \mid \mathbf{x}) \, d\theta$$

$$\propto \int_{\mathbb{R}} (\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{S^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right) \cdot \frac{1}{\sigma^2} \, d\theta$$

$(\sigma^2)^{1/2}$

$$= (\sigma^2)^{-\frac{n}{2}-1} \cdot \exp\left(-\frac{S^2}{2\sigma^2}\right) \int_{\mathbb{R}} \exp\left(-\frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right) d\theta \cdot \frac{\sqrt{2\pi\sigma^2/n}}{\sqrt{2\pi\sigma^2/n}}$$

$$\propto (\sigma^2)^{-\left(\frac{n}{2}-\frac{1}{2}\right)-1}$$

$$\rightarrow = -\frac{n}{2}+\frac{1}{2}-1 = -\left(\frac{n}{2}-\frac{1}{2}\right)$$

$$\propto (\sigma^2)^{-\frac{n}{2}-\frac{1}{2}} \cdot \exp\left(-\frac{s^2}{2\sigma^2}\right)$$

$$\Rightarrow \quad \sigma^2 \mid \mathbf{x} \sim IG\left(\overbrace{\frac{n}{2}-\frac{1}{2}}^{=\frac{n-1}{2}}, \frac{s^2}{2}\right)$$

③ $$\tilde{\pi}(\theta \mid \mathbf{x}) = \int_0^\infty \tilde{\pi}(\theta, \sigma^2 \mid \mathbf{x}) \, d\sigma^2$$

$$\propto \int_0^\infty (\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right) \cdot \frac{1}{\sigma^2} \, d\sigma^2$$

$$= \int_0^\infty (\sigma^2)^{-\frac{n}{2}-1} \exp\left(-\frac{1}{\sigma^2} \cdot \frac{1}{2}\left(s^2 + n(\bar{x}-\theta)^2\right)\right) d\sigma^2$$

$x \sim IG(\alpha, \beta)$

$f(x) = \dfrac{\boxed{\beta^\alpha}}{\boxed{\Gamma(\alpha)}} x^{-\alpha-1} \cdot e^{-\beta\theta}$

a kernel for $IG\left(\underbrace{\frac{n}{2}}_{=\alpha}, \underbrace{\frac{1}{2}\left(s^2 + n(\bar{x}-\theta)^2\right)}_{=\beta}\right)$.

$$\propto \left(\frac{1}{2}s^2 + \frac{1}{2}n(\bar{x}-\theta)^2\right)^{-\frac{n}{2}}$$

* t- distr.

$$x \sim t_\nu(\mu, \sigma^2) \Leftrightarrow \frac{x-\mu}{\sigma} \sim t_\nu$$

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \cdot \sqrt{\nu\pi} \cdot \sigma} \left(1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}$$

$$\propto \left(1 + \frac{n-1}{n-1} \frac{n(\bar{x}-\theta)^2}{s^2}\right)^{-\frac{n-1+1}{2}}$$

$\nu = n-1$

$\mu = \bar{x}$

$s^2/n/(n-1) = s^2/\{n(n-1)\}$

$$\left(1 + \frac{1}{n-1} \cdot \frac{(\bar{x}-\theta)^2}{\underbrace{s^2/n/(n-1)}_{=\sigma^2}}\right)^{-\frac{n-1+1}{2}}$$

$\sigma^2 = \underline{s^2/(n(n-1))}$.

$\underbrace{12/3}_{=4}/2 = 2$

$12/(3\cdot 2) = 2$

$$\theta \mid \mathbf{x} \sim t_{n-1}\left(\bar{x}, \frac{s^2}{n(n-1)}\right)$$

♣ Example: Assume that observations, $x_1, \ldots, x_n$ are iid from $\mathrm{N}(\theta, \sigma^2)$, where θ and σ^2 are unknown. Consider

$\bar{x}, s^2$ $\quad x$

$\theta \in \mathbb{R} \rightarrow \theta \in [0, 1]$

$$\pi(\theta, \sigma^2) = \underbrace{\pi_1(\theta \mid \sigma^2)}_{N}\underbrace{\pi_2(\sigma^2)}_{IG},$$

where π_1 is a normal distribution $\mathrm{N}(\mu, \sigma^2/n_0)$ and π_2 is a inverse gamma distribution $\mathrm{IG}(v/2, s_0^2/2)$.

- Find the joint posterior distribution $\pi(\theta, \sigma^2 \mid \boldsymbol{x})$.
- Find the posterior distributions $\underbrace{\pi_1(\theta \mid \boldsymbol{x}, \sigma^2)}_{= N}$ and $\underbrace{\pi_2(\sigma^2 \mid \boldsymbol{x})}_{= IG(\ , \ )}$.
- Find the marginal posterior distribution of θ, $\pi(\theta \mid \cancel{\bar{x}, s^2})$. x $\quad t$
  - ∗ Read Hoff §5.3, BDA §3.3, and Robert §4.4.1-4.4.2.
  - ∗ Hoff and BDA use different parameterization for π_2.
  - ∗ Do the example on Hoff pages 76-78.

† Likelihood Principle

- (Recall the definition of Likelihood) For the observed data, $X = x$, the function $\ell(\theta \mid x) = f(x \mid \theta)$, considered as a function of θ, is called the likelihood function.

  ⋆⋆ no guarantee that $\ell(\theta \mid x)$ as a function of θ is a pdf.

  ⋆⋆ *Intuitive reason for the name:* given x, the value of θ_1 is more likely to be the true parameter than θ_2 if $\ell(\theta_1 \mid x) > \ell(\theta_2 \mid x)$ (x would be more probable occurrence with θ_1).

  $f(x \mid \theta_1) > f(x \mid \theta_2)$

- **Likelihood Principle** The information brought by an observation x about θ is entirely contained in the likelihood $\ell(\theta \mid x)$. Moreover, if x_1 and x_2 are two observations depending on the same parameter θ, such that there exists a constant c satisfying

for $\theta \in \Theta$ prior data

$$\underline{\ell_1(\theta \mid x_1)} = c\,\underline{\ell_2(\theta \mid x_2)}$$

$\theta \neq c\theta$

prior distr $\ell(\theta \mid x)$

for every θ, they then bring the **same information about θ** and must lead to identical inference.

$\pi(\theta \mid x)$

$\ell_1(\theta \mid x_1) \propto \ell_2(\theta \mid x_2)$

- c does not depend on θ.

- **In other words,** In the inference about θ, after x is observed, all relevant experimental information is contained in the likelihood function for the observed x. Furthermore, the likelihood functions contain the **same information about θ** if they are proportional to each other.

† The Likelihood Principle is emphasized in Bayesian statistics!

- *Recall* The Bayesian approach is entirely based on the posterior distribution.

$$\underline{\pi(\theta \mid x)} = \frac{\cancel{c}\,\ell(\theta \mid x)\pi(\theta)}{\int_{\Theta} \cancel{c}\,\ell(\theta \mid x)\pi(\theta)d\theta}.$$

  That is, the posterior depends on data (x) only through $\underline{\ell(\theta \mid x)}$.

- Thus, the Likelihood Principle is automatically satisfied in a Bayesian setting.

Example 15 (JB, p28)– Testing fairness (1)

Suppose we are interested in testing θ, the unknown probability of heads for a possibly biased coin. Suppose we test $H0 : \theta = 1/2$ vs $Ha : \theta > 1/2$. An experiment is conducted by flipping the coin independently in a series of trials and 9 heads and 3 tails are observed. The information is not sufficient to fully specify the model, $f(x \mid \theta)$. Let's consider classical testing.

- **Scenario 1** The number of flips is pre-determined. $\boxed{N=12}$

X : # of heads

$X \mid \theta \sim \text{Binom}(12, \theta)$, $x \in \{0, 1, \ldots, 12\}$

$l_1(\theta \mid x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$

P-value : $p = \Pr(X \geq 9 \mid \theta = \frac{1}{2}) = \sum_{x=9}^{12} \Pr(X = x \mid \theta = \frac{1}{2})$

$= \boxed{0.075}$

$\alpha = 0.05 \Rightarrow$ fail to reject $H_0 : \theta = \frac{1}{2}$.

Example 15 (JB, p28)– Testing fairness (2)

- **Scenario 2** The number of tails is predetermined. $m = \#$ of tails $= \underline{3}$

X: # of heads , $\theta =$ Prob. of H

$X \mid \theta \sim$ Negative binomial $(3, 1-\theta)$

$$l_2(\theta \mid x) = \binom{x+3-1}{2} \theta^x (1-\theta)^3, \quad x = 0, 1, 2, \ldots\ldots$$

p-value $p = \Pr(X \geq 9 \mid \theta = \frac{1}{2}) = 1 - \Pr(X \leq 8 \mid \theta = \frac{1}{2})$

$= 0.0325$

$\Rightarrow$ using $\alpha = 0.05$, reject H_0

$$l_1(\theta \mid x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \propto l_2(\theta \mid x)$$

Example 15 (JB, p28)– Testing fairness (3)

- How about Bayesian inference?

①

$$\pi(\theta \mid x) \propto \underbrace{\binom{n}{x} \theta^{x} (1-\theta)^{n-x}} \pi(\theta)$$

$$\propto \theta^{9} (1-\theta)^{3} \pi(\theta)$$

②

$$\pi(\theta \mid x) \propto \binom{x+3-1}{2} \theta^{x} (1-\theta)^{3} \times \pi(\theta)$$

$$\propto \theta^{9} (1-\theta)^{3} \pi(\theta)$$

Example 15 (JB, p28)– Testing fairness (4)

* What does this imply?

- We did not really need to know anything about the "series of trials". That is, the rules governing when data collection stops are irrelevant to data interpretation.
- It is entirely appropriate to collect data until a point has been proven or disproven, or until the data collector runs out of time, money, or patience. — Edwards, Lindman, and Savage (1963, 193)

† Few Remarks!

- The correspondence of information from proportional likelihood functions applies *only* when the two likelihood functions are for the *same* parameter.
- The likelihood principle does not say all information about θ is contained in $\ell(\theta)$, just that all *experimental* information is.
- It is of fundamental importance to get the likelihood function right (the likelihood function should be a close approximation or representation of data).
- Also, see Example 1.3.5.
- Optional: read **Example 1.3.6 & Stopping Rule Principle** CR p17 & JB Chapter 7.7