Chapter 16. Introduction to Bayesian Methods for Inference

16.1. Under the assumption that X_1, ..., X_n are i.i.d. from $f(x;\theta)$, $\theta \in \Theta$, where θ is an unknown parameter, constant (not random),

we have learned 3 methods to estimate unknown parameter, say θ:

unbiased estimator,

MME,

MLE.

In this section, we study a new estimator: Bayes estimator, under

the Bayesian approach:

Conditional on θ, X_1, ..., X_n are i.i.d. from $f(x|\theta)$,

θ is a random variable with df $\pi(\theta)$,

$f(x|\theta)$ is a conditional df of $X|\theta$.

Bayes estimator of θ is $\hat{\theta} = E(\theta|\mathbf{X})$.

Recall the formula

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}. \tag{1}$$

Now

$f(\mathbf{x},\theta)$ is the joint df of $(\mathbf{X},\theta)$,

$f_{\mathbf{X}}(\mathbf{x})$ is the marginal df of $\mathbf{X}$,

$\pi(\theta)$ is the marginal df of θ, called **prior df** now,

$f(\mathbf{x}|\theta)$ is the conditional df of $\mathbf{X}|\theta$,

$\pi(\theta|\mathbf{x})$ is the conditional df of $\theta|\mathbf{X}$, called the **posterior df** now,

$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \int f(\mathbf{x},\theta)d\theta & \text{if } \theta \text{ is continuous} \\ \sum_\theta f(\mathbf{x},\theta) & \text{if } \theta \text{ is discrete.} \end{cases}$

$\pi(\theta) = \begin{cases} \int f(\mathbf{x},\theta)d\mathbf{x} & \text{if } X \text{ is continuous} \\ \sum_{\mathbf{x}} f(\mathbf{x},\theta) & \text{if } X \text{ is discrete.} \end{cases}$

$f(\mathbf{x}|\theta) = \frac{f(\mathbf{X},\theta)}{\pi(\theta)}$ by Eq. (1),

$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{X},\theta)}{f_{\mathbf{X}}(\mathbf{X})}$ by Eq. (1),

$E(\theta|\mathbf{X}=\mathbf{x}) = \begin{cases} \int \theta\pi(\theta|\mathbf{x})d\theta & \text{if } \theta \text{ is continuous} \\ \sum \theta\pi(\theta|\mathbf{x}) & \text{if } \theta \text{ is discrete.} \end{cases}$

Homework 16.1.1. Recall the Bayes set-up:

conditional on θ, X_1, ..., X_n are i.i.d. from $f(x|\theta)$.

Are X_i's i.i.d. from f_X ? Prove or disprove it through the assumption as follows.

$f(x|\theta)$ is the density of $bin(1,p)$, and $p \sim U(0,1)$.

Remark. Two ways to compute the Bayes estimator:

1. $E(\theta|\mathbf{X})$,
2. $E(\theta|T(\mathbf{X}))$ where T is a sufficient statistic.

They lead to the same estimator.

The second method is often simpler in derivation.

Example 1. Let X_1, ..., X_n be a random sample from $bin(k, \theta)$,
$\theta \sim beta(\alpha, \beta)$ with $\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$, $\theta \in [0, 1]$, where (k, α, β) is known.
Bayes estimator of θ ?
25. $X \sim beta(\alpha, \beta)$. $f(x) = \underline{\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}}$, if $x \underline{\in (0, 1)}$, $\mu = \underline{\frac{\alpha}{\alpha+\beta}}$, where $B(\alpha, \beta) = \underline{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}$
Sol. Recall $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic if θ is a parameter.
Two ways: (1) $E(\theta|\mathbf{X})$ (2) $E(\theta|T(\mathbf{X}))$.
Method 1. Based on $\mathbf{X}$.
$f(\mathbf{x}|\theta) = \prod_{i=1}^n \binom{k}{x_i}\theta^{x_i}(1-\theta)^{k-x_i}$
$= (\prod_{i=1}^n \binom{k}{x_i})\theta^{\sum_i x_i}(1-\theta)^{nk-\sum_i x_i}$.

$$\begin{aligned}
\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}, \theta)}{f_{\mathbf{X}}(\mathbf{x})} \\
&= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{f_{\mathbf{X}}(\mathbf{x})} \\
&\propto f(\mathbf{x}|\theta)\pi(\theta) \qquad\qquad \text{as } f_X \text{ does not depend on } \theta \\
&= \prod_{i=1}^n \binom{k}{x_i}\theta^{\sum_i x_i}(1-\theta)^{nk-\sum_i x_i}\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \\
&\propto \theta^{\sum_i x_i}(1-\theta)^{nk-\sum_i x_i}\theta^{\alpha-1}(1-\theta)^{\beta-1}\textbf{(main trick!!)} \\
&= \theta^{\sum_i x_i+\alpha-1}(1-\theta)^{kn-\sum_i x_i+\beta-1} \qquad (1)
\end{aligned}$$

$$\text{Thus } \theta|(\mathbf{X} = \mathbf{x}) \sim beta(\sum_i x_i + \alpha, nk - \sum_i x_i + \beta) \qquad (2)$$

$$\theta|(\mathbf{X} = \mathbf{x}) \sim beta(a, b).$$

Q: What is the meaning of Eq. (2) if $n = 0$?
The Bayes estimator is

$$\begin{aligned}
\hat{\theta} &= E(\theta|\mathbf{X}) \\
&= \frac{a}{a+b} \quad \textbf{why}? \\
&= \frac{\sum_i X_i + \alpha}{nk + \alpha + \beta} \\
&= \frac{1}{nk+\alpha+\beta}\frac{nk\sum_i X_i}{nk} + \frac{1}{nk+\alpha+\beta}(\alpha+\beta)\frac{\alpha}{\alpha+\beta} \\
&= \frac{nk}{nk+\alpha+\beta}\frac{\sum_i X_i}{nk} + \frac{\alpha+\beta}{nk+\alpha+\beta}\frac{\alpha}{\alpha+\beta} \\
&= r\frac{\sum_{i=1}^n X_i}{nk} + (1-r)\frac{\alpha}{\alpha+\beta} \qquad \approx \begin{cases} MLE & \text{if } r \approx 1 \text{ or } n \approx \infty \\ E(\theta) & \text{if } r \approx 0 \text{ or } n = 0, \end{cases} \qquad (3)
\end{aligned}$$

a weighted average of the MLE $\frac{\sum_{i=1}^n X_i}{nk}$ and the prior mean $\frac{\alpha}{\alpha+\beta}$.

Method 2. Based on the sufficient statistic $T = \sum_i X_i$.

$T|\theta \sim bin(nk, \theta)$? Yes, No, DNK

or $T \sim bin(nk, \theta)$? Yes, No, DNK

$$\begin{aligned}
f_{T|\theta}(t|\theta) &= \binom{nk}{t}\theta^t(1-\theta)^{nk-t}, \\
\pi(\theta|t) &= \frac{f_{T,\theta}(t,\theta)}{f_T(t)} \\
&= \frac{f_{T|\theta}(t|\theta)\pi(\theta)}{f_T(t)} \\
&= \frac{\binom{nk}{t}\theta^t(1-\theta)^{nk-t}\theta^{\alpha-1}(1-\theta)^{\beta-1}/B(\alpha,\beta)}{f_T(t)} \\
&= \theta^t(1-\theta)^{nk-t}\theta^{\alpha-1}(1-\theta)^{\beta-1}\frac{\binom{nk}{t}}{B(\alpha,\beta)f_T(t)} \\
&\propto \theta^{t+\alpha-1}(1-\theta)^{kn-t+\beta-1} \qquad \textbf{same as (1), why ?} \\
&\cdots
\end{aligned}$$

Additional HW:

448 [22] The Likelihood ratio test for H_o: $\theta \in \Theta_o$ v.s. H_a: $\theta \notin \Theta_o$ has a RR: $\{\lambda \_\_\_\_ k\}$, where

$\lambda =$ __________; $\hat\theta_o$ is the MLE under ____; $\hat\theta$ is the MLE under ____;

k satisfies $max\{P(RR): \theta \in \Theta_o\} =$ ____;

if n is large, then $-2\ln\lambda$ is approximated _______;

where $v =$ ___________; r and r_o = # of free parameters in Θ and in Θ_o, respectively.

key: $\underline{\leq}$, $\underline{\frac{L(\hat\theta_o)}{L(\hat\theta)}}$, $\underline{\Theta_o}$, $\underline{\Theta}$, $\underline{\alpha}$, $\underline{\chi^2(v)}$,

24(c).

$$\begin{aligned}
\lambda = \frac{L(\hat\theta_o)}{L(\hat\theta)} &= \frac{(\frac{\sum_i(x_i+y_i)}{2n})^{n\overline{x}+n\overline{y}}}{(\frac{\sum_i x_i}{n})^{n\overline{x}}(\frac{\sum_i y_i}{n})^{n\overline{y}}} \qquad \overline{x} = 20 \ \&\ \overline{y} = 22 \\
&= \frac{21^{100(20+22)}}{20^{100(20)}22^{100(22)}} \leq k = ? \qquad by\ 448[22] \\
-2\ln\lambda &\sim \chi^2(2-1) \\
-2\ln\lambda &= 9.53
\end{aligned}$$

H_o: $\theta_1 = \theta_2$ v.s.

H_1: $\theta_1 \neq \theta_2$

Test statistic: λ or $-2\ln\lambda$.

RR: $\lambda \leq k$? Yes, No, DNK

$-2\ln\lambda \leq \chi^2_{0.1,1}$? Yes, No, DNK

$-2\ln\lambda \geq \chi^2_{0.1,1}$? Yes, No, DNK