STAT 5430

Lecture 10, W, Feb 12

- Homework 3 is assigned in Canvas (due by Monday, Feb 17 by midnight)
  (practice on CRLB/UMVUE/Bayes)
- Solutions to Homeworks 1-2 posted.

- Exam 1 is scheduled for W, Feb 26 (two weeks)
  6-8 PM (location to be announced)
  - No regular class on W, Feb 26
  - See Canvas for study guide, practice exams
  - can bring 1 page formula sheet (front/back) with anything on it
  - see Canvas for a "canned" sheet
  - I'll provide table with STAT 5420 distributions (see Canvas)

Elements of Decision Theory

Posterior Distributions

Notation: For simplicity, write the random variables $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ and let $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ denote an observed value of $\underset{\sim}{X}$

prior (belief about θ before seeing data, $\underset{\sim}{X}$)

uncertainty in what value θ assumes

Bayes Set-up: Think of

(i) θ as a random variable on Θ with marginal pmf/pdf $\pi(\theta)$

(ii) $f(\underset{\sim}{x}|\theta) = f(x_1, x_2, \ldots, x_n|\theta)$ as the conditional pdf/pmf of $\underset{\sim}{X}$ *given* θ

(iii) $f(\underset{\sim}{x}, \theta) = f(\underset{\sim}{x}|\theta)\pi(\theta)$ as the joint pmf/pdf of $(\underset{\sim}{X}, \theta)$ together

← usual joint pdf/pmf of $\underset{\sim}{X}$

(iv) $m(\underset{\sim}{x}) = \int_\Theta f(\underset{\sim}{x}, \theta) d\theta$ is like a marginal pmf/pdf of $\underset{\sim}{X}$ with respect to the joint distribution of $(\underset{\sim}{X}, \theta)$ (given a $\underset{\sim}{x}$ value, integrate over θ)

Definition: The conditional pdf of θ (assumed continuous), *given* $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$,

$$f_{\theta|\underset{\sim}{x}}(\theta) = \frac{\overbrace{\boxed{f(\underset{\sim}{x}|\theta)}}^{\text{likelihood}}\overbrace{\boxed{\pi(\theta)}}^{\text{prior}}}{m(\underset{\sim}{x})}, \quad \theta \in \Theta \qquad \text{with } m(\underset{\sim}{x}) = \int_\Theta f(\underset{\sim}{x}|\theta)\pi(\theta)d\theta$$

$$\boxed{\int_{\Theta} f_{\theta|\underset{\sim}{x}}(\theta)d\theta = 1}$$

is called the **posterior pdf of θ** on Θ.

"updated distribution for θ" ← uncertainty/belief about θ after observing data

Key: $\underset{\text{posterior}}{f_{\theta|\underset{\sim}{x}}(\theta)}$ "$\propto$" (proportional to) "Likelihood" $f(\underset{\sim}{x}|\theta)$ × "prior" $\pi(\theta)$

Elements of Decision Theory

Finding Bayes Estimators

(use posterior dist)

More Notation: (only for clarity in motivating the next Theorem)

1. For any estimator/function $T = h(\underset{\sim}{X}) = h(X_1, X_2, \ldots, X_n)$ of $\underset{\sim}{X}$, the risk of T with respect to some loss function $L(t, \theta)$ is

risk
$$R_T(\theta) = \underbrace{\mathrm{E}_\theta L(T, \theta)}_{\text{earlier notation}} \equiv \underbrace{\mathrm{E}_{\underset{\sim}{X}|\theta} L\big(h(\underset{\sim}{X}), \theta\big)}_{\text{usual expectation of data given } \theta}$$

$$\mathrm{E}_{\underset{\sim}{X}|\theta} L\big(h(\underset{\sim}{X}), \theta\big) = \begin{cases} \sum_{(x_1, x_2, \ldots, x_n)} L\big(h(x_1, x_2, \ldots, x_n), \theta\big) f(x_1, x_2, \ldots, x_n|\theta) \\ \int L\big(h(x_1, x_2, \ldots, x_n), \theta\big) f(x_1, x_2, \ldots, x_n|\theta) dx_1 dx_2 \ldots dx_n \end{cases}$$

2. $\mathrm{E}_{(\theta)} R_T(\theta) = \int_\Theta R_T(\theta)\pi(\theta) d\theta$ (expectation with respect to $\pi(\cdot)$) Bayes risk

3. $\mathrm{E}_{(\underset{\sim}{X})} h(\underset{\sim}{X}) = \int h(\underset{\sim}{x}) m(\underset{\sim}{x}) dx_1 \ldots dx_n$ or $\mathrm{E}_{(\underset{\sim}{X})} h(\underset{\sim}{X}) = \sum_{\underset{\sim}{x}} h(\underset{\sim}{x}) m(\underset{\sim}{x})$

   $m(\underset{\sim}{x})$ is marginal pdf/pmf of $\underset{\sim}{X}$ in the joint distribution of $(\underset{\sim}{X}, \theta)$

Main idea: For an estimator $T = h(\underset{\sim}{X})$, the Bayes risk of T is

θ , $\underset{\sim}{X}$
parameter data

$$\begin{aligned}
\mathrm{BR}_T &= \mathrm{E}_{(\theta)} R_T(\theta) && \text{definition} \\
&= \mathrm{E}_{(\theta)}[\mathrm{E}_{\underset{\sim}{X}|\theta} L(T, \theta)] && [\text{given } \theta, \text{ expectation } \underset{\sim}{X}|\theta] \\
&= \mathrm{E}_{(\underset{\sim}{X}, \theta)} L(T, \theta) && \text{expectation with respect to } f(\underset{\sim}{x}, \theta) = f(\underset{\sim}{x}|\theta)\pi(\theta) \\
&= \mathrm{E}_{(\underset{\sim}{X})}[\mathrm{E}_{\theta|\underset{\sim}{x}} L(T, \theta)] && [\text{given } \underset{\sim}{X} = \underset{\sim}{x}, \text{ expectation } \theta|\underset{\sim}{x}]
\end{aligned}$$

for given $\underset{\sim}{x}$, pick $h(\underset{\sim}{x}) = T$ to minimize posterior risk $E_{\theta|\underset{\sim}{x}} L(h(x), \theta)$

To find an estimator $T = h(\underset{\sim}{X})$ to minimize the Bayes risk BR_T, it is enough, *at each fixed data* $\underset{\sim}{x}$ possibility of $\underset{\sim}{X}$, to pick the "$h(\underset{\sim}{x})$"-value that minimizes the so-called **posterior risk**

$$\mathrm{E}_{\theta|\underset{\sim}{x}} L(h(\underset{\sim}{x}), \theta) = \int_\Theta L(h(\underset{\sim}{x}), \theta) f_{\theta|\underset{\sim}{x}}(\theta) d\theta.$$

posterior pdf (given $\underset{\sim}{x}$)

pick $h(\underset{\sim}{x})$ to minimize this posterior risk for a given value of $\underset{\sim}{x}$

Elements of Decision Theory

Finding Bayes Estimators

← min maximax risk

Theorem: A Bayes estimator is an estimator that minimizes the "posterior risk" $\mathrm{E}_{\theta|\underset{\sim}{x}} L(h(\underset{\sim}{x}), \theta)$, over all estimators $T = h(\underset{\sim}{X})$, for fixed values $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ of $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$.

min posterior risk

Corollary: Let T_0 denote the Bayes estimator of $\gamma(\theta)$.

(1). If $L(t, \theta) = \big(t - \gamma(\theta)\big)^2$, then $T_0 = \mathrm{E}_{\theta|\underset{\sim}{x}} \gamma(\theta)$. *posterior mean of* $\gamma(\theta)$

(2). If $L(t, \theta) = \big|t - \gamma(\theta)\big|$, then $T_0 = median\big(\gamma(\theta)|\underset{\sim}{x}\big)$. *posterior median of* $\gamma(\theta)$

posterior risk $E_{\theta|\underset{\sim}{x}} L(h(\underset{\sim}{x}), \theta) = \underset{(\#)}{\int} (t - \gamma(\theta))^2 f_{\theta|\underset{\sim}{x}}(\theta)\, d\theta$, $t = h(\underset{\sim}{x})$ $\Rightarrow$ minimized $t = E_{\theta|\underset{\sim}{x}} \gamma(\theta) = \int \gamma(\theta) \cdot f_{\theta|\underset{\sim}{x}}(\theta)\, d\theta$ (#)

Example/continued: $X \sim$ Binomial(θ), $\theta \in (0, 1)$; uniform$(0, 1)$ prior for θ; $L(t, \theta) = (t - \theta)^2$. We found Bayes estimator $T_0 = \frac{X+1}{n+2}$ of $\gamma(\theta) = \theta$, but now try Corollary 1.

Solution: To find Bayes estimator of $\gamma(\theta) = \theta$

first find posterior pdf of θ:

$$\boxed{f_{\theta|\underset{\sim}{x}}(\theta) \propto f(x|\theta) \cdot \pi(\theta)}$$

$$\propto \cancel{\binom{n}{x}} \theta^x (1-\theta)^{n-x} \cdot 1, \quad 0 < \theta < 1$$

$$\propto \underbrace{\theta^x (1-\theta)^{n-x}}_{\text{Beta}(\underset{\alpha}{x+1},\ \underset{\beta}{n-x+1})}, \quad 0 < \theta < 1$$

$$\int_0^1 c\,\theta^x (1-\theta)^{n-x} dx \Rightarrow c = \frac{1}{B(x+1, n-x+1)}$$

$$f_{\theta|x}(\theta) = \frac{\theta^{x+1-1} (1-\theta)^{n-x+1-1}}{B(x+1, n-x+1)}, \quad 0 < \theta < 1$$

By corollary, $T_0 = E_{\theta|x}(\theta) = x+1/n+2$ //

Elements of Decision Theory

Finding Bayes Estimators

Definition. Let $\mathcal{F} = \{f(\underset{\sim}{x}|\theta) : \theta \in \Theta\}$ denote the class of joint pdf/pmf for $X_1, \ldots, X_n$. A class Π of priors is called a **conjugate family of priors for $\mathcal{F}$** if the posterior distribution is a member of Π for all $\pi \in \Pi$ and all $\underset{\sim}{x}$.

e.g. last example $\Theta = (0,1) \Rightarrow$ prior UNIF(0,1) $\equiv$ Beta(1,1) distribution

Found Beta(x+1, n-x+1) posterior dist.

check Beta(α,β) prior $\alpha,\beta>0$ gives Beta($\alpha+x$, $n-x+\beta$) posterior

Example. Let $X_1, \ldots, X_n$ be a random sample from UNIF$(0, \theta)$, $\theta > 0$. Find the Bayes estimator of θ with respect to an Exp(1) prior under the loss $L(t, \theta) = (t - \theta)^2/\theta^2$.

Solution: posterior pdf "$f_{\Theta|\underset{\sim}{x}}(\theta) \propto f(\underset{\sim}{x}|\theta) \cdot \underbrace{\pi(\theta)}_{e^{-\theta},\ \theta>0}$"

joint pdf of $X_1, \ldots, X_n$ given θ is

$$f(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} \{\theta^{-1} I(0<x_i<\theta)\}$$

indicator

$$= \begin{cases} \theta^{-n} & \text{if all } 0<x_i<\theta \Leftrightarrow 0<x_{(1)}<x_{(n)}<\theta \\ 0 & \text{o.w.} \end{cases}$$

min, max

So,

"proportional to"

$$f_{\Theta|\underset{\sim}{x}}(\theta) \propto \theta^{-n} I(0<x_{(1)}<x_{(n)}<\theta)\, e^{-\theta} I(\theta>0)$$

$$\propto \theta^{-n} e^{-\theta} I(x_{(n)}<\theta)$$

$$= c_0\, \theta^{-n} e^{-\theta} I(x_{(n)}<\theta)$$

(where $\int_{x_{(n)}}^{\infty} c_0\, \theta^{-n} e^{-\theta}\, d\theta = 1$ gives $c_0 = 1 \Big/ \int_{x_{(n)}}^{\infty} \theta^{-n} e^{-\theta}\, d\theta$)

Bayes estimator of θ minimizes posterior risk

$$E_{\theta|\underset{\sim}{x}} L(h(\underset{\sim}{x}),\theta) = \int_{x_{(n)}}^{\infty} \frac{(h(\underset{\sim}{x})-\theta)^2}{\theta^2} c_0 \theta^{-n} e^{-\theta} d\theta$$

minimize

$h(\underset{\sim}{x}) \equiv a$

$$g(a) \equiv \int_{x_{(n)}}^{\infty} \frac{(a-\theta)^2}{\theta^2} c_0 \theta^{-n} e^{-\theta} d\theta$$

given data fixed

$$= a^2 \int_{x_{(n)}}^{\infty} c_0 \theta^{-n-2} e^{-\theta} d\theta - 2a \int_{x_{(n)}}^{\infty} c_0 \theta^{-n-1} e^{-\theta} d\theta + \int_{x_{(n)}}^{\infty} c_0 \theta^{-n} e^{-\theta} d\theta$$

$2aC_1 - 2C_2 = 0$

$a_0 =$

is minimized

$$\text{at } a_0 = \int_{x_{(n)}}^{\infty} \theta^{-n-1} e^{-\theta} d\theta \Big/ \int_{x_{(n)}}^{\infty} \theta^{-n-2} e^{-\theta} d\theta$$

where $g'(a_0) = 0$.

Bayes Estimator of θ is

$$T_0 = \int_{x_{(n)}}^{\infty} \theta^{-n-1} e^{-\theta} d\theta \Big/ \int_{x_{(n)}}^{\infty} \theta^{-n-2} e^{-\theta} d\theta$$

Elements of Decision Theory

A Note on Bayes and Minimax Estimators

As we have seen, the minimax and Bayes principles are ways of determining estimators based on risk function considerations. There are also some interesting connections between the two types of estimators. As as example, the following result shows how to find a minimax estimator (which is hard) from a Bayes estimator (which can be done more easily).

Theorem: For some loss function $L(t,\theta)$, if T^* is a Bayes estimator with respect to *some* prior and the risk of T^* is constant (i.e., $R_{T^*}(\theta) = c$ for all $\theta \in \Theta$), then T^* is the minimax estimator under the same loss function.[1]

Example: For $X_1, \ldots, X_n$ iid Bernoulli(θ), $\theta \in \Theta = (0,1)$, find the minimax estimator of θ under the loss function $L(t,\theta) = (\theta - t)^2/\{\theta(1-\theta)\}$.

Solution: For this loss function, the Bayes estimator of θ with respect to the uniform(0,1) prior on Θ is given by $T_0 = \bar{X}_n$ (see Homework 3). Also note that, for any $\theta \in (0,1)$,

$$\begin{aligned}
R_{T_0}(\theta) \;=\; \mathrm{E}_\theta\left(\frac{(\theta-\bar{X}_n)^2}{\theta(1-\theta)}\right) &= \frac{1}{\theta(1-\theta)}\mathrm{MSE}_\theta(\bar{X}_n) \\
&= \frac{1}{\theta(1-\theta)}\mathrm{Var}_\theta(\bar{X}_n) \qquad \text{since } \bar{X}_n \text{ is unbiased} \\
&= \frac{1}{n} \qquad\qquad \text{since } \mathrm{Var}_\theta(X_1) = \theta(1-\theta) = n\mathrm{Var}_\theta(\bar{X}_n)
\end{aligned}$$

Hence, T_0 has constant risk so that, by the theorem, $T_0 = \bar{X}_n$ is also the minimax estimator of θ.

---

[1] *Proof*: Suppose that T^* is a Bayes estimator with respect to a prior pdf $\pi(\theta)$ on Θ. Then the Bayes risk of T^* is $BR_{T^*} = \int_\Theta R_{T^*}(\theta)\pi(\theta)d\theta = c\int_\Theta \pi(\theta)d\theta = c$, using that $R_{T^*}(\theta) = c$ is constant and $\int_\Theta \pi(\theta)d\theta = 1$. Since T^* is the Bayes estimator, it has minimal Bayes risk by definition so that, for any other estimator T, we have

$$c = BR_{T^*} \le BR_T = \int_\Theta R_T(\theta)\pi(\theta)d\theta \le \int_\Theta \left[\max_{\theta\in\Theta} R_T(\theta)\right]\pi(\theta)d\theta = \left[\max_{\theta\in\Theta} R_T(\theta)\right]\int_\Theta \pi(\theta)d\theta = \max_{\theta\in\Theta} R_T(\theta)$$

or $c \le \max_{\theta\in\Theta} R_T(\theta)$ for any estimator T. Because $\max_{\theta\in\Theta} R_{T^*}(\theta) = c$, it must be that T^* is minimax.