STAT 5430
Lecture 09, M, Feb 10

point estimators
- Homework 2 is assigned in Canvas (due by Monday, Feb 10, by midnight)

practice on CRLB/MSE decision theory (Bayes)
- Homework 3 is assigned in Canvas (due by Monday, Feb 17 by midnight)

Office hours Mine: FM, 12-1 PM & by appointment
TA (Min-Yi): WR 11-12 in Snedecor 2404

Elements of Decision Theory

Bayes Principle: Terminology

Definitions:

1. Let $\pi(\theta)$ be a pdf/pmf on Θ. Then, $\pi(\theta)$ is called a **prior**.

   ↙ distribution on parameter space Θ

2. Then, the **Bayes risk** of an estimator T with respect to $\pi(\theta)$ and loss function $L(t, \theta)$ is

   expectation of risk $R_T(\theta)$ w.r.t. prior $\pi(\theta)$

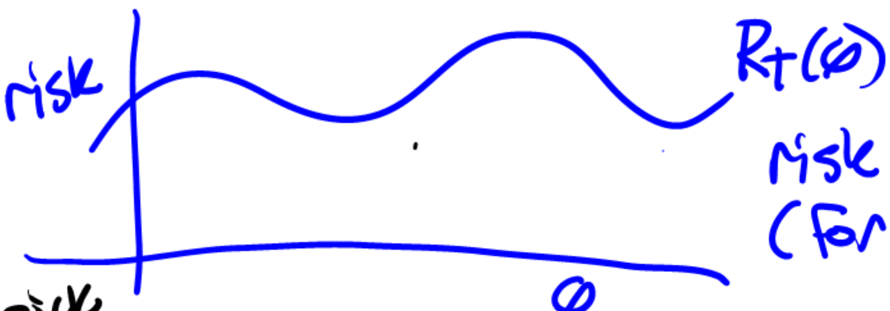
$$\mathrm{BR}_T = \begin{cases} \int_\Theta R_T(\theta)\pi(\theta)d\theta & \text{if } \pi(\cdot) \text{ is continuous} \\ \sum_{\theta\in\Theta} R_T(\theta)\pi(\theta) & \text{if } \pi(\cdot) \text{ is discrete} \end{cases}$$

   (average risk)

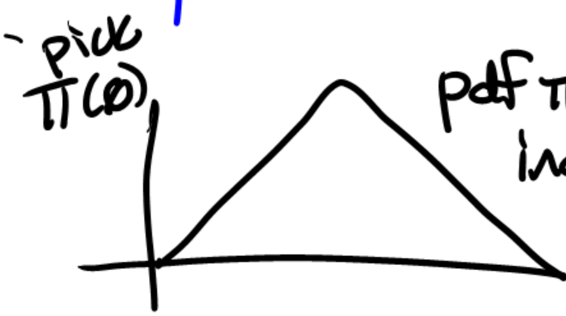
3. An estimator T_0 is called the **Bayes estimator** (with respect to the prior $\pi(\theta)$) if

$$\mathrm{BR}_{T_0} = \min_T \mathrm{BR}_T$$

   smallest Bayes risk



risk changes with θ (for given estimator T)



pdf $\pi(\cdot)$, prior indicates where you "think" θ might fall (uncertainty in what suspect for θ)

$\mathrm{BR}_T = \int_\Theta R_T(\theta)\pi(\theta)d\theta$ (average risk)

Elements of Decision Theory

Bayes Principle: Illustration

data/count (# of successes in "n" trials)

Example: Suppose X ~Binomial(n,θ), $\underbrace{0<\theta<1}$. Find the Bayes estimator of θ with respect to the Uniform$(0,1)$ prior & under the loss function $L(t,\theta)=(t-\theta)^2$.

Solution: The parameter space is $\Theta \equiv (0,1)$ & prior $\pi(\theta)=1$

If $T\equiv h(X)$ denotes an estimator of θ, then the risk of T is

$$\underbrace{R_T(\theta)}_{\text{risk}} \equiv \mathrm{E}_\theta\left[L(T,\theta)\right] = \mathrm{E}_\theta\left[(\overset{T=h(X)}{T}-\theta)^2\right] = \sum_{x=0}^{n}\binom{n}{x}\theta^x(1-\theta)^{n-x}(h(x)-\theta)^2$$

and the Bayes risk of T is

$$\mathrm{BR}_T = \int_0^1 R_T(\theta)\pi(\theta)d\theta = \sum_{x=0}^{n}\binom{n}{x}\underbrace{\int_0^1 \theta^x(1-\theta)^{n-x}(h(x)-\theta)^2 d\theta}_{\text{for each } x\text{, pick } h(x) \text{ to minimize this}}$$

To find a Bayes estimator $T_0\equiv h_0(X)$ (with minimal Bayes risk BR_{T_0}): for each possible outcome $x=0,\ldots,n$, we'd want to pick $h_0(x)$ to minimize

fixed

Beta function

$$\int_0^1 \theta^x(1-\theta)^{n-x}(h_0(x)-\theta)^2 d\theta \qquad \leftarrow \text{recall beta}(\alpha,\beta) \text{ pdf has form: } f(y)\equiv y^{\alpha-1}(1-y)^{\beta-1}/B(\alpha,\beta),\ 0<y<1$$

$y\equiv\theta$

$$\begin{aligned}
&= B(x+1,n-x+1)\int_0^1 \frac{y^{x+1-1}(1-y)^{n-x+1-1}}{B(x+1,n-x+1)}(h_0(x)-y)^2dy\\
&= B(x+1,n-x+1)\cdot \mathrm{E}[h_0(x)-Y_x]^2 \qquad \text{for } Y_x \sim\text{Beta}(\underset{\alpha}{x+1},\underset{\beta}{n-x+1})
\end{aligned}$$

which is minimized by

$$h_0(x) = \mathrm{E}Y_x = \frac{x+1}{x+1+n-x+1} = \frac{x+1}{n+2} \quad = \frac{\alpha}{\alpha+\beta}$$

[Fact: For a r.v. W with $\mathrm{E}W^2<\infty$, $g(a)\equiv \mathrm{E}[W-a]^2 = \mathrm{Var}(W)+[a-\mathrm{E}W]^2$ is minimized at $a=\mathrm{E}W$.]

So, BR_T is minimized at

$T_0 = \dfrac{X+1}{n+2}$ (T_0 is Bayes Estimator w.r.t. uniform prior)

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, $\underset{\sim}{X}$)

Bayes Set-up: Think of

(i) θ as a random variable on Θ with marginal pmf/pdf $\pi(\theta)$ ← uncertainty in what value θ assumes

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

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

(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* $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$,

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

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

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

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

Key: $\underbrace{f_{\theta|\underset{\sim}{x}}(\theta)}_{\text{posterior}}$ "$\propto$" (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 parameter, $\underset{\sim}{X}$ 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 $\mathrm{E}_{\theta|\underset{\sim}{x}} L(h(\underset{\sim}{x}), \theta)$

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 maximum risk

Theorem: A Bayes estimator is an estimator that minimizes the "posterior risk" $E_{\theta|\underset{\sim}{x}} L(h(\underset{\sim}{x}), \theta)$, 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 $T_0 = E_{\theta|\underset{\sim}{x}} \gamma(\theta)$. *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 $E_{\theta|\underset{\sim}{x}} L(h(\underset{\sim}{x}), \theta) = \int (t-\gamma(\theta))^2 f_{\theta|\underset{\sim}{x}}(\theta)\,d\theta \Rightarrow$ minimized $t = E_{\theta|\underset{\sim}{x}} \gamma(\theta) = \int \gamma(\theta) f_{\theta|\underset{\sim}{x}}(\theta)\,d\theta$

Θ $t = h(\underset{\sim}{x})$ Θ

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:

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

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

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

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

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