Statistical Decision Theory

STAT3023

Rachel Wang

School of Mathematics and Statistics, USyd

5 Oct, 2022

Full decision theory framework

In the full framework, we have

- A family of distributions $\mathcal{F} = \{f_\theta(\cdot) : \theta \in \Theta\}$ for a random vector $\mathbf{X}$ taking values in $\mathcal{X}$;
- A decision space $\mathcal{D}$, where each decision $d(\cdot)$ is a **function** mapping a possible value $\mathbf{x} \in \mathcal{X}$ into $\mathcal{D}$
- A non-negative-valued loss funciton such that when a decision d is made and the true distribution generating $\mathbf{X}$ is $f_\theta(\cdot)$, a loss of $L(d|\theta)$ is suffered.
- The risk funciton associated with decision function $d(\cdot)$ is:

$$R(\theta | d(\cdot)) = E_\theta \left(L(d(\underset{\sim}{X}) | \theta) \right)$$

$$\underset{\sim}{X} \sim f_\theta$$

Full decision theory framework

Example: Suppose we have 2 independent observations X_1, X_2 from an exponential distribution with mean θ. Take the loss as $L(d|\theta) = (d-\theta)^2$.

$f_\theta(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad x > 0$

- We already know that the CRLB for unbiased estimation of θ is $\theta^2/2$, attained by $\bar{X} = \frac{X_1+X_2}{2} = d_{\text{MVU}}(\mathbf{X})$.
- Consider the family of decisions $\{d_c(\cdot) : c > 0\}$ given by $d_c(\mathbf{X}) = c\bar{X}$.
- The risk of $d_c(\cdot)$ is:

$$R(\theta | d_c) = E_\theta\{(c\bar{X}-\theta)^2\}$$
$$= E_\theta(c^2\bar{X}^2 - 2c\theta\bar{X} + \theta^2)$$
$$= c^2 E_\theta(\bar{X}^2) - 2c\theta E_\theta(\bar{X}) + \theta^2$$

$$E_\theta(\bar{X}) = \theta$$
$$Var(\bar{X}) = \frac{1}{2} \cdot Var(X_1) = \frac{\theta^2}{2}$$

$$\mathbb{E}_\theta(\bar{X}^2) = \frac{\theta^2}{2} + \theta^2$$

$$= c^2\left(\frac{3}{2}\theta^2\right) - 2c\theta^2 + \theta^2$$

$$= \theta^2\left(\frac{3}{2}c^2 - 2c + 1\right)$$

$$\frac{\partial}{\partial c} R(\theta|d_c) = \theta^2(3c-2) = 0 \qquad \underline{c = \frac{2}{3}}$$

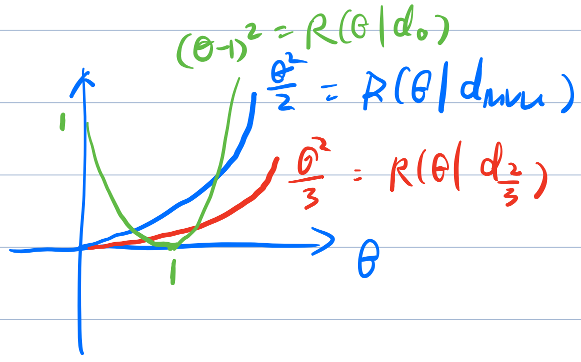
The best decision is $d_{\frac{2}{3}}(X) = \frac{2}{3}\bar{X}$

$$R(\theta|d_{\frac{2}{3}}) = \cdots = \frac{\theta^2}{3}$$

Although $\frac{2}{3}\bar{X}$ is biased, it has a smaller MSE (risk) than $\bar{X}$.
and this is true for all $\theta > 0$

$\frac{2}{3}\bar{X}$ is uniformly better than $\bar{X}$ (under the squared-error loss).

$d_0(X) = 1$, $R(\theta|d_0) = (\theta-1)^2 = \theta^2 - 2\theta + 1$



d_0 outperforms d_{MVU}, $d_{\frac{2}{3}}$ if true θ is around 1.

Full decision theory framework

- The above example shows neither d_0 nor $d_{2/3}$ is uniformly better than the other. Rather, there are ranges of θ for which each is better.
- It is not very useful to compare risk functions in a pointwise sense. In fact we need some "overall" measure of risk to encompass all θ values.

Overall risk measure

Bayes (or integrated) risk: For a given non-negative weight function $w(\cdot)$: ← weight function / prior

$$B_w(d) = \int_\Theta w(\theta) \cdot R(\theta | d)\, d\theta$$

If $\tilde{d}(\cdot)$ is s.t. $B_w(\tilde{d}) \le B_w(d)$ for any other decision function $d(\cdot)$, then $\hat{d}$ is said to be a Bayes procedure (or Bayes decision rule) w.r.t. weight / prior $w(\cdot)$.

Overall risk measure

Maximum risk: For a given subset $\Theta_0 \subseteq \Theta$, a decision rule $\hat{d}(\cdot)$ is said to be minimax (over Θ_0) if

$$\max_{\theta \in \Theta_0} R(\theta | \hat{d}) \leq \max_{\theta \in \Theta_0} R(\theta | d)$$

best decision in worst case scenario.

Finding Bayes procedures

- Bayes procedures can be found by reducing the problem to a simple prediction problem.
- Recall the **Bayes risk** of a decision rule $d(\cdot)$ (w.r.t. to a weight function/prior $w(\cdot)$) is

$$R(\theta|d) = \mathbb{E}[L(d(\underset{\sim}{x})|\theta)]$$

$$B_w(d) = \int_{\Theta} w(\theta) R(\theta|d)\, d\theta$$

$$= \int_{\Theta} w(\theta) \left(\int \cdots \int_{\mathcal{X}} L(d(\underset{\sim}{x})|\theta) f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x} \right) d\theta$$

$$= \int \cdots \int_{\mathcal{X}} \left[\int_{\Theta} L(d(\underset{\sim}{x})|\theta) \underline{w(\theta) f_\theta(\underset{\sim}{x})}\, d\theta \right] d\underset{\sim}{x}$$

assume $w(\theta) \cdot f_\theta(\underset{\sim}{x})$ is integrable over Θ.

$\mathbb{E}_{\theta|\underset{\sim}{x}}(L(d(\underset{\sim}{x})|\theta))$

conditional Bayes risk

define $m(\underset{\sim}{x}) = \int_\Theta w(\theta) f_\theta(\underset{\sim}{x})\, d\theta$

$$= \int_{\mathcal{X}} \cdots \int m(\underset{\sim}{x}) \underbrace{\left[\int_\Theta L(d(\underset{\sim}{x})|\theta) \cdot \frac{w(\theta) f_\theta(\underset{\sim}{x})}{m(\underset{\sim}{x})}\, d\theta \right]}_{} d\underset{\sim}{x}$$

$$p(\theta|\underset{\sim}{x}) = \frac{w(\theta) f_\theta(\underset{\sim}{x})}{m(\underset{\sim}{x})}, \quad \int_\Theta p(\theta|\underset{\sim}{x})\, d\theta = 1$$

$p(\theta|\underset{\sim}{x})$ can be viewed as a pdf of θ.

This is known as the posterior density of θ.

(conditional density of θ given $\underset{\sim}{X} = \underset{\sim}{x}$)

The inner integral is a simple prediction problem, based on a single draw of θ from $p(\theta|\underset{\sim}{x})$ with loss $L(d|\theta)$

If we know decision $\tilde{d}(\underset{\sim}{x})$ minimizes the risk this simple prediction problem,

$$\int_\Theta L(\tilde{d}(\underset{\sim}{x})|\theta)\, p(\theta|\underset{\sim}{x})\, d\theta$$

$$\le \int_\Theta L(d(\underset{\sim}{x})|\theta)\, p(\theta|\underset{\sim}{x})\, d\theta \quad \text{for any other decision } d.$$

then we also have $B_w(\tilde{d}) \le B_w(d)$.

Finding Bayes procedures

Example 1. Suppose $X_1, \dots, X_n$ are iid $N(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$ with deicision space $\mathcal{D}$ and loss $L(d|\theta)$.

(a) $\mathcal{D} = \mathbb{R}$, $L(d|\theta) = (d - \theta)^2$

(b) $\mathcal{D} = \mathbb{R}$, $L(d|\theta) = |d - \theta|$

(c) $\mathcal{D} = \mathbb{R}$, $L(d|\theta) = 1\{|d - \theta| > 1.96/\sqrt{n}\}$

(d) $\mathcal{D} = \{0, 1\}$,

$$L(d|\theta) = \begin{cases} L_0 & \text{if } d = 1, \theta \in \Theta_0 \\ L_1 & \text{if } d = 0, \theta \in \Theta_1 \\ 0 & \text{otherwise.} \end{cases}$$

choose d that maximises.

$L_d \cdot P_{\theta|\underset{\sim}{x}}(\theta \in \Theta_d)$

Find Bayes procedures of the above with $w(\theta) = 1$, the "flat prior".

Find the posterior of θ. (cond. density of θ given $\underset{\sim}{X}$)

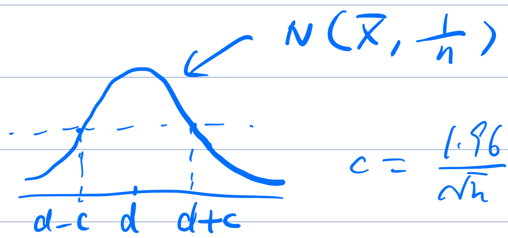
$p_{\theta|\underset{\sim}{X}} \sim N(\bar{x}, \frac{1}{n})$ see tute this week.

(a) $d(\underset{\sim}{X})$ is the mean of posterior, $d(\underset{\sim}{X}) = \bar{x}$

(Example on squared-error loss from Tuesday)

(b) $d(\underset{\sim}{X})$ is the median of posterior. $d(\underset{\sim}{X}) = \bar{X}$.

(c)



$c = \frac{1.96}{\sqrt{n}}$

$d(\underset{\sim}{X}) = \bar{X}$