Statistical Decision Theory

STAT3023

Rachel Wang

School of Mathematics and Statistics, USyd

4 Oct, 2022

Decision theory

- Many statistical procedures can be analysed through the general framework of statistical decision theory.
- We will begin with a special case - simple prediction problems.
- Suppose Y is a random variable from a known distribution. $\mathcal{D}$ is an arbitrary set which is called the decision space. For each possible value y of Y and a decision $d \in \mathcal{D}$, we measure the performance of d using a loss function $L(d|y) \geq 0$.
- Goal: choose d to minimise the expected loss:

$$R(d) = \mathbb{E}(L(d|Y))$$

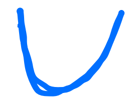
risk

Decision theory

Example: **Squared error loss.**
$\mathcal{D} = \mathbb{R}$, $L(d|y) = C(d-y)^2$ for some $C > 0$.

$$R(d) = \mathbb{E}[L(d|Y)] = C \cdot \mathbb{E}[(d-Y)^2]$$

$$= C\,(d^2 - 2\mathbb{E}Y \cdot d + \mathbb{E}Y^2)$$



$$R'(d) = C \cdot (2d - 2\mathbb{E}Y) = 0$$

$$d = \mathbb{E}Y \qquad R(d) = \mathrm{Var}(Y) \cdot C$$

best predictor of Y under the squared-error loss is $\mathbb{E}(Y)$.

Decision theory

Example: **Absolute error loss.**
$\mathcal{D} = \mathbb{R}$, $L(d|y) = C|d-y|$ for some $C > 0$. Y is continuous with cdf $F(\cdot)$ and density $f(\cdot)$.

$$R(d) = C \cdot E(|d-Y|)$$

$$= C \cdot \int_{-\infty}^{\infty} |d-y| f(y)\,dy$$

$$= C \cdot \left[\int_{-\infty}^{d} (d-y) f(y)\,dy + \int_{d}^{\infty} (y-d) \cdot f(y)\,dy \right]$$

$$= C \cdot \left[d \cdot F(d) - \underbrace{\int_{-\infty}^{d} y f(y)\,dy + \int_{d}^{\infty} y f(y)\,dy}_{-2\int_{-\infty}^{d} y f(y)\,dy + E(Y)} - d(1-F(d)) \right]$$

$$= C \cdot \Big[2d\,F(d) - d + \mathbb{E}(Y) - 2 \underbrace{\int_{-\infty}^{d} y f(y)\, dy}_{h(d)} \Big]$$

$h'(d) = d f(d)$ fundamental thm of calculus.

$$R'(d) = C \cdot \Big[2\big(F(d) + d f(d)\big) - 1 - 2 d f(d) \Big]$$

$$= C \cdot \big(2F(d) - 1 \big) = 0$$

$$\begin{cases} < 0 & \text{for } F(d) < \frac{1}{2} \quad R \text{ decreasing} \\ > 0 & \text{for } F(d) > \frac{1}{2} \quad R \text{ increasing.} \end{cases}$$

$R(d)$ is minimised at $F(d) = \frac{1}{2}$

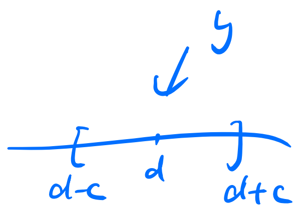
$d = F^{-1}\left(\frac{1}{2}\right)$ median of the distribution of Y.

Decision theory

Example: **0-1 (zero-one) loss.**
$\mathcal{D} = \mathbb{R}$, $L(d|y) = 1\{|d-y| > c\}$ for some $c > 0$. Y is continuous with density $f(\cdot)$, where $f(\cdot)$ is unimodal. That is, $f(y)$ is strictly increasing for $y < m$ and strictly decreasing for $y > m$ for some mode m. Further assume $f(y) > 0$ over an interval I with length at least $2c$.

$$L(d|y) = \begin{cases} 1 & |d-y| > c \\ 0 & |d-y| \le c \end{cases}$$

$$= \begin{cases} 0 & d-c \le y \le d+c \\ 1 & \text{otherwise} \end{cases}$$

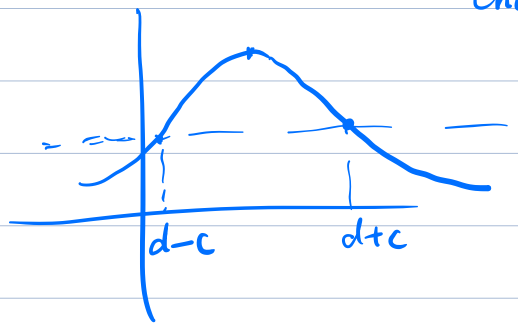


$$R(d) = E(L(d|Y)) = P(|d-Y| > c)$$

$$= P(Y < d-c) + P(Y > d+c)$$

$$= F(d-c) + 1 - F(d+c)$$

$$R'(d) = f(d-c) - f(d+c) = 0$$



chose d. s.t. the interval $d \pm c$ is a level set of $f(\cdot)$.

Decision theory

Example: **Discrete selection**.

Suppose $\mathbb{R}$ is partitioned into sets $S_1, S_2, \ldots, S_k$,
$\mathcal{D} = \{1, 2, \ldots, k\}$, and the loss is $L(d|y) = \sum_{j=1}^{k} L_{d,j} 1\{y \in S_j\}$,
where $L_{d,j}$ is a $k \times k$ loss matrix, such that

$$L_{d,d} = 0, \quad L_{d,j} = L_j \text{ for } d \neq j.$$

$$\begin{array}{l} d=1 \\ d=2 \\ \\ \\ \end{array} \begin{bmatrix} 0 & L_2 & L_3 & \cdots & L_k \\ L_1 & 0 & L_3 & \cdots & L_k \\ \vdots & L_2 & \ddots & & \vdots \\ L_1 & L_2 & \cdots & & 0 \end{bmatrix}$$

$$R(d) = E(L(d|Y))$$
$$= \sum_{j=1}^{k} L_{d,j} P(Y \in S_j)$$
$$= \sum_{\substack{j \neq d \\ j=1}}^{k} L_j P(Y \in S_j)$$
$$= \underbrace{\sum_{j=1}^{k} L_j P(Y \in S_j)}_{\text{indep of } d} - L_d P(Y \in S_d)$$

maximising over d. $L_d \cdot P(Y \in S_d)$.

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. $L(d(\underset{\sim}{x})|\theta)$
- The risk funciton associated with decision function $d(\cdot)$ is:

$$R(\theta|d(\cdot)) = E_\theta\left[L(d(\underset{\sim}{x})|\theta)\right] \qquad \underset{\sim}{x} \sim f_\theta$$