3. Decision Theoretic Approach

1. Introduction

Loss Function We define the *loss function* $\ell(t, \theta)$, which assigns disutility to each set of an estimate t and parameter value θ. Examples are

(a) $\ell(t, \theta) = (t - \theta)^2$, squared error loss

(b) $\ell(t, \theta) = |t - \theta|$, absolute error loss

(c) $\ell(t, \theta) = c\,\mathrm{I}\{|t - \theta| > \varepsilon\}$, fixed loss out of ε bound

(d) $\ell(t, \theta) = c(\theta)|t - \theta|^r$

Risk Function For an estimator $T = \tau(X)$, we define the *risk function* by

$$r(\tau, \theta) = \mathbf{E}_\theta \ell(T, \theta)$$

which can be thought of expected loss of an estimator for each value of θ. The risk functions corresponding to the loss functions in the above examples are

(a) $r(\tau, \theta) = \mathbf{E}_\theta(\tau(X) - \theta)^2$, MSE

(b) $r(\tau, \theta) = \mathbf{E}_\theta|\tau(X) - \theta|$, mean absolute error

(c) $r(\tau, \theta) = c\mathbf{P}_\theta\{|\tau(X) - \theta| > \varepsilon\}$

(d) $r(\tau, \theta) = c(\theta)\mathbf{E}_\theta|\tau(X) - \theta|^r$

Admissibility and Minimaxity Let τ_1 and τ_2 be two estimators for θ. We say that τ_1 is better than τ_2 if $r(\tau_1, \theta) \leq r(\tau_2, \theta)$ for all $\theta \in \Theta$ and $r(\tau_1, \theta) < r(\tau_2, \theta)$ for some $\theta \in \Theta$. In this case, we also say that τ_1 *dominates* (*beats*) τ_2.

Definition 1 *An estimator τ is said to be* admissible *if no other estimator beats τ.*

Example Let $X_1, \ldots, X_n$ be i.i.d. $\mathbf{N}(\mu, 1)$, and let $\tau_1(x) = x_1$ and $\tau_2(x) = \bar{x}$ be two estimators for μ. If the loss function is given by $\ell(t, \theta) = (t - \theta)^2$, then

$$\begin{aligned} r(\tau_1, \mu) &= \mathbf{E}_\mu(X_1 - \mu)^2 = 1 \\ r(\tau_2, \mu) &= \mathbf{E}_\mu(\overline{X} - \mu)^2 = \frac{1}{n} \end{aligned}$$

We can see clearly that τ_1 is dominated by τ_2.

For an estimator τ, $\sup_{\theta\in\Theta} r(\tau,\theta)$ measures its maximum risk. We define

Definition 2 *An estimator τ_* is called* minimax *if*

$$\sup_{\theta\in\Theta} r(\tau_*,\theta) \le \sup_{\theta\in\Theta} r(\tau,\theta)$$

for every other estimator τ.

2. Rao-Blackwell and Lehmann-Scheffe Theorem

Theorem 1 (Rao-Blackwell) *Suppose that $\ell(t,\theta)$ is convex in t, and that S is a sufficient statistic. Let $T = \tau(X)$ be an estimator for θ with finite mean and risk. Define $T_* = \mathbf{E}_\theta(T|S)$ and write $T_* = \tau_*(X)$. Then we have*

$$r(\tau_*,\theta) \le r(\tau,\theta)$$

Proof Since the loss function $\ell(t,\theta)$ is convex in t, we have from the Jensen's inequality that

$$\begin{aligned} \ell(T_*,\theta) &= \ell\left(\mathbf{E}_\theta(T|S),\theta\right) \\ &\le \mathbf{E}_\theta(\ell(T,\theta)|S) \end{aligned}$$

Taking expectations on both sides yields

$$\mathbf{E}_\theta \ell(T_*,\theta) \le \mathbf{E}_\theta \ell(T,\theta)$$

as we wanted to show. ■

Remarks

(a) If we write $f(S) = \mathbf{E}_\theta(T|S)$ and $S = \sigma(X)$, then $T_* = \tau_*(X)$ with $\tau_* = f\circ\sigma$. Note that $\mathbf{E}_\theta(T|S)$ is a function of only S, and not of θ, since S is sufficient.

(b) For a loss function $\ell(t,\theta)$ convex in t, an admissible estimator for θ must be a function of every sufficient statistic, and hence of minimal sufficient statistic. Otherwise, we may always improve it by taking expectation conditional on a sufficient statistic.

(c) If the loss function $\ell(t,\theta)$ is strictly convex in t for $\theta = \theta_0$, then $r(\tau_*, \theta_0) < r(\tau, \theta_0)$ unless $\tau_* = \tau$ a.s. P_{θ_0}.

Definition 3 *A statistic T is called* complete *if $\mathbf{E}_\theta(f(T)) = 0$ for all $\theta \in \Theta$ implies $f = 0$ a.s.* P_θ.

Remark A statistic T is complete if and only if there exists a unique function of T that is unbiased. To see this, let $f_1(T)$ and $f_2(T)$ be unbiased, i.e., $\mathbf{E}_\theta f_1(T) = \mathbf{E}_\theta f_2(T) = \theta$. Then $\mathbf{E}_\theta f(T) = 0$ with $f = f_1 - f_2$, and completeness of T implies that $f = 0$, that is $f_1 = f_2$ a.s. P_θ.

Theorem 2 (Lehmann-Scheffe) *If S is complete and sufficient and $T_* = f(S)$ is unbiased, then T_* is the UMVU estimator for the squared error loss function $\ell(t,\theta) = (t-\theta)^2$.*

Proof Obvious from Remark (b) of Rao-Blackwell. ■

Remark Given a complete and sufficient statistic S, it is now easy to obtain the UMVU estimator. We may indeed take any unbiased estimator U and let $T_* = \mathbf{E}_\theta(U|S)$. The resulting estimator T_* is the UMVU estimator, as one can easily see.

Examples:

(a) Let X_i, $i = 1, \ldots, n$, be i.i.d. $U(0,\theta)$. Recall that $T = \max_{1 \le i \le n} X_i$ is sufficient. The statistic T is indeed complete. To see this, note that

$$\begin{aligned}
\mathbf{P}_\theta\{T \le t\} &= \mathbf{P}_\theta\{X_1 \le t, \ldots, X_n \le t\} \\
&= (\mathbf{P}_\theta\{X_i \le t\})^n \\
&= \left(\frac{t}{\theta}\right)^n
\end{aligned}$$

and T has density

$$p_\theta(t) = \frac{nt^{n-1}}{\theta^n} \, \mathrm{I}\{0 \le t \le \theta\}$$

Now, $\mathbf{E}_\theta f(T) = 0$ for all θ implies

$$\int_0^\theta t^{n-1} f(t)\, dt = 0$$

for all θ, from which we deduce that $f = 0$ a.s.

(b) Let X_i, $i = 1, \ldots, n$, and T be given as above. Define an unbiased estimator $U = 2X_1$ of θ. Suppose $T = t$. Then X_1 can take t with probability $1/n$, since every X_i, $i = 1, \ldots, n$, is equally likely to have the maximum value. Moreover, when $X_1 \neq t$ with probability $(n-1)/n$, X_1 is uniformly distributed on $(0, t)$. Therefore,

$$\begin{aligned} \mathbf{E}_\theta(U|T = t) &= 2\mathbf{E}_\theta(X_1|T = t) \\ &= 2\left(\frac{1}{n}\, t + \frac{n-1}{n}\, \frac{t}{2}\right) \\ &= \frac{n+1}{n}\, t \end{aligned}$$

Thus

$$\tau_*(X) = \frac{n+1}{n} \max_{1 \le i \le n} X_i$$

is the UMVU estimator of θ.

3. The Bayesian Approach

For the Bayesian approach, the unknown parameter θ is regarded as a realization of the underlying random variable Θ. The *Bayes risk* for an estimator $T = \tau(X)$ is then defined by

$$r(\tau) = \mathbf{E}\, \ell(T, \Theta)$$

Remark Since the conventional risk function $r(\tau, \theta)$ can be viewed as

$$r(\tau, \theta) = \mathbf{E}(\ell(T, \Theta)|\Theta = \theta)$$

in the Bayesian approach, it follows that

$$r(\tau) = \mathbf{E}\, r(\tau, \Theta) = \int r(\tau, \theta) p(\theta)\, d\theta$$

where $p(\theta)$ is the prior density of Θ. The Bayes' risk may thus be regarded as an average of the conventional risk, which is taken with respect to our prior belief on the value of Θ.

Definition 4 *An estimator* $T_* = \tau_*(X)$ *is a Bayes estimator if it minimizes*

$$\mathbf{E}(\ell(T,\Theta)|X=x)$$

Remarks:

(a) The Bayes estimator minimizes the average loss after the observation is made.

(b) The Bayes estimator minimizes the Bayes risk, since

$$r(\tau) = \mathbf{E}\,\ell(T,\Theta) = \mathbf{E}(\mathbf{E}\,\ell(T,\Theta)|X)$$

for $T = \tau(X)$.

Proposition 3 *If* $\ell(t,\theta) = (t-\theta)^2$*, then the Bayes estimator* $\tau_*(x)$ *is the posterior mean of* Θ*, i.e.,*

$$\tau_*(x) = \mathbf{E}(\Theta|X=x)$$

Proof Note for $T = \tau(X)$

$$\begin{aligned}
&\mathbf{E}((T-\Theta)^2|X=x) \\
&= \mathbf{E}((T-\mathbf{E}(\Theta|X=x))^2|X=x) + \mathbf{E}((\mathbf{E}(\Theta|X=x)-\Theta)^2|X=x) \\
&= (\tau(x)-\tau_*(x))^2 + \mathbf{E}((\tau_*(x)-\Theta)^2|X=x)
\end{aligned}$$

which is minimized when $\tau = \tau_*$. ■

Example In the previous example, the conditional distribution of Θ given X was given by $Beta\left(\sum_i x_i + 1, n - \sum_i x_i + 1\right)$. We therefore have

$$\tau_*(x) = \frac{\sum_{i=1}^{n} x_i + 1}{n+2}$$

Note that the mean of $Beta(\alpha,\beta)$ distribution is $\alpha/(\alpha+\beta)$.

Theorem 4 *If τ_* is a Bayes estimator with constant conventional risk, then τ_* is minimax.*

Proof For any estimator τ, we have

$$\begin{aligned} \sup_{\theta\in\Theta} r(\tau_*,\theta) &= r(\tau_*) \\ &\le r(\tau) \\ &\le \sup_{\theta\in\Theta} r(\tau,\theta) \end{aligned}$$

as was to be shown. ■

4. Exercises

1. Let X_1 and X_2 be independent random variables with the underlying probability density

$$p(x,\theta) = \frac{1}{\theta}\,\mathrm{I}\{0 \le x \le \theta\}$$

(a) Find the MLE of θ.

(b) Let $T = \max\{X_1, X_2\}$ and $S = \min\{X_1, X_2\}$. Find the conditional distribution of S given $T = t$. Show that T is sufficient for θ.

(c) Using the squared-error loss function find a Bayes estimator of θ with respect to the uniform prior distribution on (0,1).

2. Let X be a single random observation from the density

$$p(x,\theta) = \frac{2x}{\theta^2}\,\mathrm{I}\{0 \le x \le \theta\}$$

where $\theta \in (0,1)$. The loss function is given by $\ell(t,\theta) = \theta^2(t-\theta)^2$. We define two estimators τ_1 and τ_2 of θ by

$$\tau_1(x) = \frac{3}{2}x \quad \text{and} \quad \tau_2(x) = \mathrm{I}[\frac{1}{2}, 1)$$

(a) Are τ_1 and τ_2 unbiased?

(b) Obtain the risk functions $r(\tau_1,\theta)$ and $r(\tau_2,\theta)$ of τ_1 and τ_2.

(c) Compute the Bayes risks $r(\tau_1)$ and $r(\tau_2)$, with respect to the uniform prior on $(0,1)$. Compare them.

3. Let X be a single random observation from a density given by

$$\frac{2x}{\theta^2}\,\mathrm{I}\{0 \le x \le \theta\}$$

where $\theta \in [0,1]$. In the classical approach, the density represents a member of a family of densities $p_\theta(x)$ parametrized by θ. On the other hand, the density is viewed for the Bayesian analysis as the conditional density $p(x|\theta)$ of X given $\Theta = \theta$, where Θ has the uniform distribution over $[0,1]$.

Now let the loss function be given by $\ell(t,\theta) = \theta^2(t-\theta)^2$, and define two estimators τ_1 and τ_2 of θ by

$$\tau_1(x) = \frac{3x}{2} \quad \text{and} \quad \tau_2(x) = \frac{x+1}{2}$$

Answer the following:

(a) Show that τ_1 is an unbiased estimator of θ.

(b) Show that τ_2 is the Bayes estimator of θ.

(c) Obtain the risk functions $r(\tau_1,\theta)$ and $r(\tau_2,\theta)$ of τ_1 and τ_2, respectively. Does one dominate the other? Compare the maximum risks of τ_1 and τ_2. Can you say anything about the minimaxity of τ_2?

(d) Compute the Bayes risks $r(\tau_1)$ and $r(\tau_2)$ for τ_1 and τ_2, and compare them.