Some basics of decision theory

1 Introduction

When a point estimate is desired, what is the Bayesian choice? Such questions can be addressed from a decision theory perspective. Suppose the data x follow a distribution with p.d.f. $f(x|\theta)$, $x \in \mathcal{X}$ and $\theta \in \Theta$. We call $\mathcal{X}$ the *sample space* and Θ *parameter space*. Any statistical inference, e.g. an estimate or rejection of a null hypothesis, can be viewed as an *action*. Suppose we consider allowable actions $a \in \mathcal{A}$ in the *action space* $\mathcal{A}$. Then a decision rule $d \in \mathcal{D}$ is a mapping $\mathcal{X} \to \mathcal{A}$. The general question is: *how to choose the decision rule*?

One can establish a systematic way of choosing decisions rules based on a loss function $L(\theta, a)$. For example, for parameter estimation, the most widely used loss function is the squared-error loss (SEL) function

$$L(\theta, a) = (\theta - a)^2.$$

Other alternatives include

- Weighted SEL: $L(\theta, a) = w(\theta)(\theta - a)^2$, where $w(\theta)$ is a weighting function;
- Absolute error loss: $L(\theta, a) = |\theta - a|$;
- 0-1 loss: $L(\theta, a) = 1(\theta \neq a)$.

Based on a loss function $L(\theta, a)$, one can then define the *risk* function of a decision rule $d(x)$. That is,

$$R(\theta, d) = E_{X|\theta}[L(\theta, d(x)] = \in L(\theta, d(x))f(x|\theta)dx.$$

Note that the risk is a function of θ for a given decision rule. Between two decision rules d_1 and d_2, if $R(\theta, d_1) \leq R(\theta, d_2)$ for all θ, we say d_2 is *inadmissible*. Obviously, it is still unclear to choose between two admissible decision rules. There are typically two solutions, *minimax* and *Bayes rule*.

The minimax approach is rather conservative by comparing the worst scenario of two decision rules. That is, the minimax rule d^* is given by

$$\sup_{\theta \in \Theta} R(\theta, d^*) = \inf_{d \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, d).$$

On the other hand, the Bayes rule is selected as the one with the lowest Bayes risk, which is an overall summary of the entire risk function. Suppose a prior distribution $\pi(\theta)$ is used. Then the Bayes risk is defined as

$$r(\pi, d) = E_\theta[R(\theta, d)].$$

The Bayes rule is then

$$d_\pi(x) = \arg\min_{d \in \mathcal{D}} r(\pi, d). \tag{1}$$

2 Bayes rule

Using the expression in (1) requires integration over θ consider all possible decision rules, hence it is computationally challenging. In many often-encountered problems, one can alleviate the complexity as follows. First, one can rewrite

$$r(\pi, d) = E_X[\rho(\pi, d(x))],$$

where $\rho(\pi, d(x)) = E_{\theta|x}[L(\theta, d(x))] = \int L(\theta, d(x)) f(\theta|x) d\theta$ is called the *Bayes posterior risk*. **If for any data x, there exists the same function $m_\pi(x)$ that minimizes $\rho(\pi, d(x))$, then $m_\pi(x)$ minimizes $r(\pi, d)$ and it is the Bayes rule.**

Example: Under the SEL, the Bayes estimate is $E(\theta|x)$.

There are two nice properties of Bayes rules.

1. A Bayes rule with a constant risk $R(\theta, d)$ is minimax.

2. A unique Bayes rule is admissible.

3 Interval estimation and hypothesis testing

The following is from Schervish (1995).

3.1 Credible interval

When considering bounded intervals as credible interval, the action space can be considered to be the set of ordered pairs (a_1, a_2) in which $a_1 \leq a_2$. Consider a loss function that penalizes excessive length above, below, and around Θ differently:

$$L(\theta, [a_1, a_2]) = a_2 - a_1 + \begin{cases} c_1(a_1 - \theta) & \theta < a_1 \\ 0 & a_1 \leq \theta \leq a_2 \\ c_2(\theta - a_2) & a_2 < \theta \end{cases} \tag{2}$$

Then the optimal interval is the interval between two quantiles of the posterior distribution.

Theorem Suppose that the posterior mean of Θ is finite and the loss is as in (2) with $c_1, c_2 > 1$. The formal Bayes rule is the interval between the $1/c_1$ and $1 - 1/c_2$ quantiles of the posterior distribution of Θ.

In the special case in which $c_1 = c_2 = 2/\alpha > 1$, the optimal interval runs from the $\alpha/2$ quantile of the posterior to the $1-\alpha/2$ quantile. This would be the usual **equal-tailed**, two sided posterior probability interval for Θ.

Theorem Suppose that $\Omega \in R$ and that the action space is the Borel σ-field of subsets of Ω. Suppose that the posterior distribution of Θ has a density $f_{\Theta|X}$ with respect to Lebesgue measure λ and that the loss is $L(\theta, B) = \lambda(B) + c(1 - I_B(\theta))$. Then, the formal Bayes rule is an **HPD region** of the form $B(x) = \{\theta : f_{\Theta|X}(\theta|x) \geq 1/c\}$

For the case in which the density is strongly unimodal (that is, $\{\theta : f_{\Theta|X}(\theta|x) > a\}$ is an interval for all a), the formal Bayes rule for loss function $L_l(\theta, [a_1, a_2]) = a_2 - a_1 + c(1 - I_{[a_1,a_2]}(\theta))$ is an HPD region. This loss function does not penalize differently for how short the interval is when it misses the parameter.

3.2 Hypothesis testing

Consider testing $H_0 : \theta \in \Theta_0$ versus $H_A : \theta \in \Theta - \Theta_0$. If we use the following 0-1 loss,

$$L(\theta, a) = \begin{cases} 1 & \text{if } a \neq 1_{\theta \in \Theta_0}, \\ 0 & \text{otherwise,} \end{cases}$$

the Bayes rule is

$$\delta^\pi(y) = \begin{cases} 1 & \text{if } P(\theta \in \Theta_0 \mid y) > P(\theta \notin \Theta_0 \mid y), \\ 0 & \text{otherwise,} \end{cases}$$

References

Schervish, M. J. (1995). *Theory of Statistics*, Springer.