A primer on Statistical Decision Theory: theoretical framework

Antonietta Mira
Swiss Finance Institute,
University of Lugano, Switzerland
jointly with
Filippo Macaluso

Statistical Decision Theory: Motivation

- Classical statistics uses sampling information to make inferences about unknown parameters
- Decision theory *combines* sampling information with the knowledge of the consequences of the taken decisions

Framework

- State of Nature: parametric space Θ
- Decision Space: $\mathcal{D}$. It is the space of all possible values of decisions/actions/rules
- Loss Function : $\ell(\theta, \delta(X))$. It is a function of $\theta \in \Theta$ and a decision $\delta \in \mathcal{D}$ based on the data $x \in X$
- Example
  $\delta(X) = \overline{X}$

Loss function properties and example

- A loss function $\ell(\theta, \delta(X))$ is a non negative random variable defined on $\Theta \times \mathcal{D}$
  It can be interpreted as a measure of how much we lose by choosing an action δ given θ (the real state of nature)

i) Absolute error loss $\mid \theta - \delta(x) \mid$

ii) Squared error loss $(\theta - \delta(x))^2$

- $\ell(\delta, \delta) = 0$
- $\ell(\delta, \theta)$ is non decreasing function of $|\delta - \theta|$

In order to compare which procedure is the best we can use two approaches:

Frequentist $\longrightarrow$ Frequentist Risk

Bayesian $\longrightarrow$ Posterior Risk

Frequentist Risk

Frequentist Risk is defined as

$$R(\theta, \delta(X)) = \mathbb{E}_\theta[\ell(\theta, \delta(X))]$$

Note that θ is fixed and the expectation is taken over X

Example: $\mathbb{E}_\theta[(\theta - \delta(X))^2] = V[\delta(X)] + bias(\delta(X))^2$

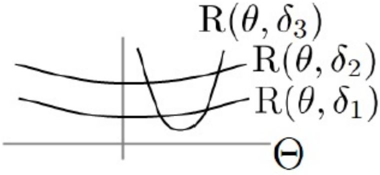

Figure 1: Frequentist Risk

Estimator Comparison

The Risk principle: a decision δ_1 is better than another decision δ_2 in the sense of risk if $R(\theta, \delta_1) \leq R(\theta, \delta_2)$ $\forall \theta \in \Theta$ then

Best Estimator(Uniformly minimum risk estimator)

$$\delta^* = arg \min_{\delta} R(\theta, \delta) \qquad \forall \theta \in \Theta$$

Usually it does not exist then we use the following criteria:

- Admissibility: A decision rule is admissible (with respect to the loss function) if and only if no other rule dominates it. Otherwise, we said it is inadmissible.
- Restricted class of estimator: i.e unbiased estimator

- Minimax. We look at $\sup_{\Theta} R(\theta, \delta)$. We consider the maximum value of the risk and we choose the *smallest maximum* risk

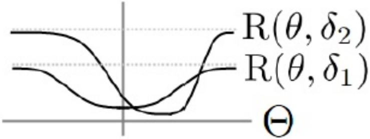

Figure 2: Minimax Frequentist Risk

Posterior Risk

$$\rho(\pi, \delta(x)) = \int_{\Theta} \ell(\theta, \delta(x)) p(\theta|x) d\theta$$

where $p(\theta|x) \propto \pi(\theta) p(x|\theta)$

The Expected Loss Principle: in comparing two actions δ_1 and δ_2 after data x had been observed, preferred action is the one for which the posterior expected loss is smaller. An action δ^* that minimizes the posterior expected loss is called *Bayes action*

$$\delta^*(x) = arg \min_{\delta} \rho(\pi, \delta(x))$$

Example(1)

Let $\ell(\theta, \delta(x)) = (\theta - \delta(x))^2$ then

$$\begin{aligned}
\rho &= \int (\theta - \delta(x))^2 p(\theta|x) d\theta \\
&= \delta(x)^2 - 2\delta(x) \int \theta p(\theta|x) d\theta + \int \theta^2 p(\theta|x) d\theta \\
\rightarrow & \quad \text{FOC} \quad \frac{\partial \rho}{\partial \delta(x)} = 2\delta(x) - 2 \int \theta p(\theta|x) d\theta = 0 \\
\rightarrow & \quad \delta^*(x) = \int \theta p(\theta|x) d\theta
\end{aligned}$$

then, δ^* is the posterior mean

Example(2)

Let $\ell(\theta, \delta(x)) = |\theta - \delta(x)|$ then

$$\begin{aligned} \rho &= \int |\theta - \delta(x)| p(\theta|x) d\theta \\ &= \int_{\theta \geq \delta} (\theta - \delta(x)) p(\theta|x) d\theta + \int_{\theta \leq \delta} (\delta(x) - \theta) p(\theta|x) d\theta \\ \rightarrow & \quad \text{FOC} \quad \frac{\partial \rho}{\partial \delta(x)} = -\int_{\theta \geq \delta} p(\theta|x) d\theta + \int_{\theta \leq \delta} p(\theta|x) d\theta = 0 \end{aligned}$$

then δ^* is the median of the posterior distribution