STATS 300A: Theory of Statistics I **Autumn 2016/17**

Lecture 7: October 18

Lecturer: Joseph Romano *Scribes: Kenneth Tay*

7.1 Minimum Risk Equivariant (MRE) Estimators

Say we are in a location model setting, i.e. the underlying family of distributions has the form $F(x-\theta)$, where F is a known distribution, $\theta \in \mathbb{R}$. Assume in addition that the loss function can be expressed as $L(\theta, d) = \rho(d-\theta)$ for some function ρ.

Recall that we have a characterization for all location equivariant estimators:

Theorem 7.1 *Let δ_0 be some location equivariant estimator.*

An estimator δ is location equivariant iff there exists a function v of $n-1$ arguments with

$$\delta(X) = \delta_0(X) - v(Y),$$

where $Y = (X_1 - X_n, \dots, X_{n-1}, X_n)$.

The following theorem allows us to find the "best" location equivariant estimator (i.e. the one with least risk):

Theorem 7.2 *Under the above set-up, suppose that there exists an equivariant δ_0 with finite risk. Assume that for each $y = (y_1, \dots, y_n)$, there exists a number $v(y) = v^*(y)$ which minimizes*

$$\mathbb{E}_0\{\rho[\delta_0(X) - v(y)] \mid Y = y\}.$$

Then the estimator $\delta^*(X) = \delta_0(X) - v^*(Y)$ is MRE.

Corollary 7.3 *Under the assumptions of Theorem 7.2,*

1. *If $\rho(d-\theta) = (d-\theta)^2$, then $v^*(Y) = \mathbb{E}_0[\delta_0(X) \mid Y]$.*
2. *If $\rho(d-\theta) = |d-\theta|$, then $v^*(Y) =$ any median of the conditional distribution $\delta_0(X) \mid Y$.*

7.1.1 MRE Estimator Examples

7.1.1.1 Normal distribution

Let X_i iid, $X_i \sim \mathcal{N}(\theta, 1)$ for all i.

Take $\delta_0 = \bar{X}$. It is location equivariant, and is complete sufficient as well. By Basu's Theorem, Y is independent of $\bar{X}$, and so

$$v^*(Y) = \mathbb{E}_0[\bar{X} \mid Y] = \mathbb{E}_0\bar{X} = 0.$$

Thus, $\bar{X}$ is an MRE estimator.

7.1.1.2 General distributions with density

Let $\mathcal{F} = \{$all univariate distributions F having density f with $\sigma^2(F) = 1\}$. Let $X_1, \ldots, X_n$ iid, $X_i \sim f(x - \theta)$ with $\theta = \mathbb{E}X_i$.

Let $r_n(F)$ be the risk of the MRE estimator of θ under squared error loss. Then $r_n(F) \leq r_n(\bar{X_n}) = \frac{1}{n}$.

7.1.1.3 Shifted exponential family

Let $E(\theta, b)$ represent the distribution with density

$$f_\theta(x) = \begin{cases} \frac{1}{b}e^{-(x-\theta)/b} & \text{for } x > \theta, \\ 0 & \text{otherwise.} \end{cases}$$

This has CDF $F_\theta(x) = 1 - \exp[-(x-\theta)/b]$, or survival function $\exp[-(x-\theta)/b]$.

Let the X_i iid come from $E(\theta, b)$ with θ unknown, b known. $X_{(1)}$ is a complete sufficient statistic for this model.

Take $\delta_0(X) = X_{(1)}$. By Basu's Theorem, $v(y)$ will not depend on Y. $v(y)$ is to be determined by minimizing

$$\mathbb{E}_{\theta=0}[\rho(X_{(1)} - v)].$$

- For squared-error loss, $v^* = \mathbb{E}_0 X_{(1)}$. Note that

$$\begin{aligned} P\{X_{(1)} > t\} &= \prod_{i=1}^n P\{X_i > t\} \\ &= \exp[-n(x-\theta)/b]. \end{aligned}$$

  By looking at the survival functions, we deduce that $X_{(1)} \sim E\left(\theta, \frac{b}{n}\right)$. Thus, $\mathbb{E}_0 X_{(1)} = \frac{b}{n}$ and the MRE estimator is $X_{(1)} - \frac{b}{n}$.

- For absolute error loss, $v^* =$ median of $X_{(1)}$, i.e.

$$\begin{aligned} 1 - \exp[-n(v^* - \theta)/b] &= \frac{1}{2}, \\ v^* &= \frac{b \log 2}{n}. \end{aligned}$$

  Thus, the MRE estimator is $X_{(1)} - \frac{b \log 2}{n}$.

7.1.2 Squared Error Loss and the Pitman Estimator

Assume that we are looking at squared error loss. From before, we know that the MRE is given by

$$\delta^* = \delta_0 - \mathbb{E}_0[\delta_0 | Y].$$

The following theorem gives us a "formula" to compute δ^*:

Theorem 7.4 *Under squared error loss, the MRE estimator is given by*

$$\delta^* = \frac{\int_{-\infty}^{\infty} u f(x_1 - u, \ldots, x_n - u) du}{\int_{-\infty}^{\infty} f(x_1 - u, \ldots, x_n - u) du}.$$

This is called the ***Pitman estimator****.*

Proof: Take $\delta_0 = X_n$. We wish to compute $\mathbb{E}_0[X_n \mid Y = y]$ for each y.

Let $y_n = x_n$. Then, for $i = 1, \ldots, n-1$, we have $x_i = y_i + x_n = y_i + y_n$. Calculate the Jacobian J:

$$J = \left(\frac{\partial y_i}{\partial x_j}\right)_{i,j} = \begin{pmatrix} 1 & 0 & \ldots & 0 & -1 \\ 0 & 1 & \ldots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -1 \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix}.$$

Since $\det A = 1$, the joint density of the Y's is given by

$$p_Y(y_1, \ldots, y_n) = f(y_1 + y_n, \ldots, y_{n-1} + y_n, y_n).$$

Hence, the conditional density of $Y_n \mid Y_1, \ldots, Y_{n-1}$ is

$$\frac{f(y_1 + y_n, \ldots, y_{n-1} + y_n, y_n)}{\int_{-\infty}^{\infty} f(y_1 + t, \ldots, y_{n-1} + t, t) dt},$$

so

$$\begin{aligned} \mathbb{E}_0[X_n \mid Y_1, \ldots, Y_n] &= \frac{\int_{-\infty}^{\infty} t f(y_1 + t, \ldots, y_{n-1} + t, t) dt}{\int_{-\infty}^{\infty} f(y_1 + t, \ldots, y_{n-1} + t, t) dt} \\ &= \frac{\int_{-\infty}^{\infty} t f(x_1 - x_n + t, \ldots, x_{n-1} - x_n + t, t) dt}{\int_{-\infty}^{\infty} f(x_1 - x_n + t, \ldots, x_{n-1} - x_n + t, t) dt} \\ &= \frac{\int_{-\infty}^{\infty} (x_n - u) f(x_1 - u, \ldots, x_n - u) du}{\int_{-\infty}^{\infty} f(x_1 - u, \ldots, x_n - u) du} \qquad \text{(change of variables with } u = x_n - t) \\ &= x_n - \frac{\int_{-\infty}^{\infty} u f(x_1 - u, \ldots, x_n - u) du}{\int_{-\infty}^{\infty} f(x_1 - u, \ldots, x_n - u) du}. \end{aligned}$$

Hence,

$$\delta^* = X_n - \mathbb{E}_0[X_n \mid Y] = \frac{\int_{-\infty}^{\infty} u f(x_1 - u, \ldots, x_n - u) du}{\int_{-\infty}^{\infty} f(x_1 - u, \ldots, x_n - u) du}.$$

■

Example: Consider the family of distributions Unif $\left(\theta - \frac{b}{2}, \theta + \frac{b}{2}\right)$, where b is known and θ is unknown. We have the density

$$f(x_1 - \theta, \ldots, x_n - \theta) = \begin{cases} \frac{1}{b^n} & \text{if } \theta - \frac{b}{2} \le x_{(1)} \le x_{(n)} \le \theta + \frac{b}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

Then the Pitman estimator (of θ) is given by

$$\frac{\int_{X_{(n)}-b/2}^{X_{(1)}+b/2} \frac{u}{b^n} du}{\int_{X_{(n)}-b/2}^{X_{(1)}+b/2} \frac{1}{b^n} du} = \frac{u^2/2\big|_{X_{(n)}-b/2}^{X_{(1)}+b/2}}{u\big|_{X_{(n)}-b/2}^{X_{(1)}+b/2}}$$
$$= \frac{(X_{(1)} + \frac{b}{2})^2 - (X_{(n)} - \frac{b}{2})^2}{2(X_{(1)} + b/2 - X_{(n)} + b/2)}$$
$$= \frac{X_{(1)} + X_{(n)}}{2}.$$

7.1.3 Comparing MRE and UMVU Estimators

When a UMVU estimator exists, it is typically the same for all convex loss functions. However, UMVU estimators may not exist for loss functions which are non-convex or bounded. UMVU estimators may also be inadmissible.

In contrast, MRE estimators generally exist for any loss function, but the solution usually depends on the choice of loss function. In addition, the Pitman estimator is generally admissible.

We have the following lemma for squared error loss:

Lemma 7.5 *For squared error loss,*

- *If δ is location equivariant with constant bias $b = \mathbb{E}_\theta \delta - \theta$, then $\delta - b$ is unbiased, location equivariant and has smaller risk than δ (unless $b = 0$).*
- *The unique MRE estimator is unbiased.*
- *If the UMVU estimator exists and is location equivariant, it must be MRE as well.*
- *If the UMVU estimator exists, then it is the Pitman estimator (except in exceptional settings).*

7.2 Bayes Estimators

With UMVU and MRE estimators, we restricted the class of estimators, then found the one with minimal risk among them. Instead of restricting the class of estimators, we could weaken the optimality criteria. One way to do this is through Bayes estimators.

Definition 7.6 *If we have a risk function $R(\theta, \delta)$, the **Bayes estimator** δ^* is the estimator which minimizes the quantity*

$$\int_{\theta \in \Omega} R(\theta, \delta) d\Lambda(\theta),$$

*where Λ is some fixed and specified measure (usually a probability measure on Ω). Λ is called the **prior distribution** for θ.*