STATS 300A: Theory of Statistics

Lecture 6 — October 8

Lecturer: Lester Mackey

Scribe: Evan Rosenman and Lin Fan

Warning: These notes may contain factual and/or typographic errors.

6.1 Optimal Equivariant Estimation

In this lecture, we will continue our discussion of optimal equivariant estimation under location models (the associated reading can be found in Keener 10.1-10.2). Recall from our last lecture that the goal in equivariant estimation is to enforce symmetries in our estimator, in accordance with symmetries that exist in the model and loss function. In the location family setting, we made this idea precise last lecture with the following definitions:

Definition 1 (Location-invariant decision problem). A decision problem is called locationinvariant if the family of densities $\mathcal{P} = \{f_{\theta} : \theta \in \Omega\}$ is location-invariant, i.e.,

$$f_{\theta+c}(x+c) = f_{\theta}(x),$$

and the loss function is location-invariant, i.e.,

$$L(\theta, d) = L(\theta + c, d + c) = \rho(d - \theta).$$

Definition 2 (Location equivariant estimator). An estimator is called location equivariant if

$$\delta(X_1 + c, \dots, X_n + c) = \delta(X_1, \dots, X_n) + c.$$

Examples of such δ include \overline{X}_n , the median of X_1, \ldots, X_n , and X_1 .

We also discovered a useful strategy for deriving minimum risk equivariant (MRE) estimators.

Theorem 1. If the decision problem is location invariant, δ_0 is location equivariant with finite risk, and if $v^*(y)$ minimizes $\mathbb{E}_0[\rho(\delta_0 - v(y))|Y = y]$ for each y, then an MRE estimator is $\delta_0(X) - v^*(Y)$, where $Y \triangleq (X_1 - X_n, \dots, X_{n-1} - X_n)$.

Proof. By Theorem 1.8 of TPE, in order to find an MRE, we must choose a function v^* to minimize

$$\mathbb{E}_{\theta=0}[\rho(\delta_0(X) - v(Y))] = \int \mathbb{E}_{\theta=0}[\rho(\delta_0(X) - v(Y))|Y = y]dP_0(y)$$
(6.1)

over v. In order to minimize the RHS, it is sufficient to minimize the integrand for all y. \Box

A corollary to this theorem, found in the supplemental text *Theory of Point Estimation*, is given below:

Fall 2015

Theorem 2. If ρ , the loss function, is convex and not monotone, then an MREE exists. If ρ is also *strictly* convex, then the MREE is unique.

We will now consider a number of examples.

Example 1. Suppose $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\theta, b)$ with an unknown location parameter θ and a known scale parameter b > 0. The density of each of the X_i has the form

$$p(x_i; \theta) = \frac{1}{b} \exp\left(-\frac{(x_i - \theta)}{b}\right) \mathbb{I}[x_i > \theta].$$

We begin by noting that $\min(X_1, \ldots, X_n)$ is a complete sufficient statistic for this model (see TPE Example 6.24, page 43). We need to choose a "base estimator" to start our procedure for finding an MREE. We see that $\delta_0(X) = \min(X_1, \ldots, X_n)$ is location equivariant as

$$\delta_0(X+c) = \min(X_1+c, \dots, X_n+c) = \delta_0(X) + c.$$

so we choose $\delta_0(X)$ as our base estimator. Since $Y = (X_1 - X_n, \dots, X_{n-1} - X_n)$ is ancillary (as it has no θ dependence), $\delta_0(X) \perp Y$ (by Basu's theorem), and hence

$$v^{\star}(y) = \operatorname*{argmin}_{v} \mathbb{E}_{0} \left[\rho(\delta_{0}(X) - v) | Y = y \right] = \operatorname*{argmin}_{v} \mathbb{E}_{0} \left[\rho(\delta_{0}(X) - v) \right] = v^{\star}$$

for all y.

The goal of the rest of this example is to find the constant v^* in the above expression, which will in turn lead to an MRE estimator by Theorem 1. We consider two specific forms of the loss function.

• Case 1: With squared error loss, i.e., $\rho(t) = t^2$, we have that:

$$\underset{v}{\operatorname{argmin}} \mathbb{E}_0\left[\rho(\delta_0(X) - v)\right] = \underset{v}{\operatorname{argmin}} \mathbb{E}_0\left[(\delta_0(X) - v)^2\right]$$

Taking the derivative of this quantity, we find that $v^* = \mathbb{E}_0 \delta_0(X)$. To evaluate this quantity, note that when $\theta = 0$,

$$\mathbb{P}\left(\delta_0(X) \ge t\right) = \mathbb{P}\left(X_i \ge t, \forall i\right) = \prod_{i=1}^n \mathbb{P}\left(X_i \ge t\right) = \begin{cases} \exp\left(-\frac{nt}{b}\right) & t \ge 0\\ 1 & t < 0 \end{cases},$$

which implies that $\delta_0(X) \sim \text{Exp}(0, b/n)$ with mean b/n. Thus an MRE estimator in this case is $\min(X_1, \ldots, X_n) - (b/n)$.

• Case 2: With absolute error loss, i.e., $\rho(t) = |t|$, it can be checked that v^* is the median of $\delta_0(X)$ with X_i distributed as Exp (0, b). That is, v^* satisfies the equation

$$\mathbb{P}_0(\delta_0(X) \leqslant v^*) = 1/2 = \mathbb{P}_0(\delta_0(X) \ge v^*),$$

which is equivalent to

$$1/2 = \mathbb{P}_0(\delta_0(X) \ge v^*) = \exp\left(-nv^*/b\right)$$

Solving the equation yields $v^* = (b/n) \ln 2$ and hence an MRE estimator $\min(X_1, \ldots, X_n) - (b/n) \ln 2$.

6.2 The Pitman Estimator of Location

Under squared error loss $\rho(t) = t^2$, we can derive an explicit form for the (unique) MRE estimator. First, we note that for any location equivariant δ_0 with finite risk, $v^*(Y) = \arg\min_v \mathbb{E}_0[(\delta_0(X) - v)^2|Y] = \mathbb{E}_0[\delta_0(X)|Y]$, so an MRE estimator is

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

We will show that in fact δ^* takes the more explicit form

$$\delta^*(X_1, ..., X_n) = \frac{\int_{-\infty}^{\infty} u \ f(X_1 - u, ..., X_n - u) du}{\int_{-\infty}^{\infty} f(X_1 - u, ..., X_n - u) du}.$$

In this form, δ^* is known as the *Pitman estimator of location*.

Proof. (Theorem 3.1.20 of TPE) Consider the simple location equivariant estimator $\delta_0(X) = X_n$. Our goal is to compute $\mathbb{E}_0[X_n|Y=y]$. To this end, define $Z \stackrel{\Delta}{=} (Y_1, \dots, Y_{n-1}, X_n)$ with joint density $p_{z,0}$ under location parameter 0 and note that

$$\mathbb{E}_0[X_n|Y] = \frac{\int_{-\infty}^{\infty} t \ p_{z,0}(Y_1, \dots, Y_{n-1}, t) dt}{\int_{-\infty}^{\infty} p_{z,0}(Y_1, \dots, Y_{n-1}, t) dt}.$$

To determine the form of $p_{z,0}$ we will perform a change of variables. Notice that

$$Z_i = Y_i = X_i - X_n, \forall i \le n - 1, \quad \text{and} \quad Z_n = X_n.$$

Furthermore,

$$p_{z,0}(z_1,...,z_n) = p_{x,0}(x_1,...,x_n) \left| \frac{\partial x}{\partial z} \right| = f(x_1,...,x_n) \left| \frac{\partial x}{\partial z} \right|$$

where the Jacobian matrix $\frac{\partial x}{\partial z}$ of partial derivatives has the form

$$\frac{\partial x}{\partial z} = \begin{pmatrix} \frac{\partial x_1}{\partial z_1} & \cdots & \frac{\partial x_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial z_1} & \cdots & \frac{\partial x_n}{\partial z_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Hence $\left|\frac{\partial x}{\partial z}\right| = 1$, and

$$p_{x,0}(x_1, \dots, x_n) = f(x_1, \dots, x_n) = f(y_1 + x_n, \dots, y_{n-1} + x_n, x_n).$$

Thus, we may write

$$\mathbb{E}_0[X_n|Y] = \frac{\int_{-\infty}^{\infty} t \ f(Y_1 + t, \dots, Y_{n-1} + t, t) dt}{\int_{-\infty}^{\infty} f(Y_1 + t, \dots, Y_{n-1} + t, t) dt}.$$

Changing variables once more to replace t with $X_n - u$, we obtain

$$\mathbb{E}_0[X_n|Y] = \frac{\int_{-\infty}^{\infty} (X_n - u) f(Y_1 + X_n - u, ..., Y_{n-1} + X_n - u, X_n - u) du}{\int_{-\infty}^{\infty} f(Y_1 + X_n - u, ..., Y_{n-1} + X_n - u, X_n - u) du},$$

that is,

$$\mathbb{E}_0[X_n|Y] = X_n - \frac{\int_{-\infty}^{\infty} u \ f(X_1 - u, ..., X_n - u) du}{\int_{-\infty}^{\infty} f(X_1 - u, ..., X_n - u) du}$$

Therefore, $\delta^*(X) = X_n - \mathbb{E}_0[X_n|Y]$, and equals the Pitman estimator as desired.

Let's compute the Pitman estimator of location in a specific example.

Example 2. Suppose $X_i \stackrel{\text{iid}}{\sim} U(\theta - \frac{b}{2}, \theta + \frac{b}{2})$ for b known. Then,

$$f(x_1 - \theta, ..., x_n - \theta) = \frac{1}{b^n} \mathbb{I}\left(\theta - \frac{b}{2} \le x_{(1)} \le x_{(n)} \le \theta + \frac{b}{2}\right).$$

The squared error MRE estimator is therefore

$$\delta^*(X) = \frac{\int u \frac{1}{b^n} \mathbb{I}[X_{(n)} - \frac{b}{2} \le u \le X_{(1)} + \frac{b}{2}] du}{\int \frac{1}{b^n} \mathbb{I}[X_{(n)} - \frac{b}{2} \le u \le X_{(1)} + \frac{b}{2}] du} = \frac{\frac{1}{2}((X_{(1)} + \frac{b}{2})^2 - (X_{(n)} - \frac{b}{2})^2)}{X_{(1)} + \frac{b}{2} - (X_{(n)} - \frac{b}{2})} = \frac{X_{(1)} + X_{(n)}}{2}$$

Note that the MRE estimator in the previous example was also unbiased. This is not a coincidence. Indeed, the following lemma shows us that, under squared error loss, the unique MRE estimator is always unbiased.

Lemma 1. (TPE 3.1.23) Under squared error loss:

- If $\delta(X)$ is location equivariant with constant bias b, then $\delta(X) b$ is unbiased and location equivariant with smaller risk than $\delta(X)$.
- The unique MRE estimator is unbiased.
- If UMVUE exists and is location equivariant, then it is also MRE.

Let us take a moment to compare several characteristic properties of MRE and UMRU estimators.

- 1(a) When an UMVUE exists, it is UMRU for any convex loss
- 1(b) No UMRUE exists for most bounded losses
- 1(c) UMVUEs are often inadmissible
- 2(a) An MREE exists for most losses, but the MRE estimator itself is often loss-dependent
- 2(b) The Pitman estimator is typically admissible

Moreover, while squared error MRE estimators are unbiased, MRE estimators under other losses satisfy a more general risk unbiasedness property.

6.3 Risk Unbiasedness

Definition 3. An estimator δ of $g(\theta)$ is called **risk unbiased** for a loss function $L(\theta, d)$ if

$$\mathbb{E}_{\theta}[L(\theta, \delta(X))] \leq \mathbb{E}_{\theta}[L(\theta', \delta(X))], \ \forall \theta', \ \theta.$$

In other words, on average, the true parameter penalizes the estimator δ no more than any false parameter.

Example 3. (Mean unbiasedness) If $L(\theta, d) = (d-\theta)^2$ (squared error loss), and for estimator δ we have $\mathbb{E}_{\theta}[\delta(X)^2] < \infty$, then for any θ'

$$\mathbb{E}_{\theta}[L(\theta',\delta(X))] = \mathbb{E}_{\theta}[\delta(X) - g(\theta')]^2 = \operatorname{Var}_{\theta}(\delta(X)) + (\mathbb{E}_{\theta}[\delta(X)] - g(\theta'))^2.$$

If $\mathbb{E}_{\theta}[\delta(X)]$ is always in range $(g) \triangleq g(\Omega) \triangleq \{g(\theta) : \theta \in \Omega\}$, then δ is **mean unbiased** if and only if $\mathbb{E}_{\theta}[\delta(X)] = g(\theta)$. Thus, risk unbiasedness under the squared error loss or mean unbiasedness is just our normal notion of unbiasedness.

Example 4. (Median unbiasedness) If $L(\theta, d) = |d - \theta|$ (absolute error loss), and for estimator δ we have $\mathbb{E}_{\theta}[|\delta(X)|] < \infty$, then risk unbiasedness is equivalent to

$$\mathbb{E}_{\theta}[|\delta(X) - g(\theta')|] \ge \mathbb{E}_{\theta}[|\delta(X) - g(\theta)|], \ \forall \ \theta', \ \theta.$$

In this case, if the median_{θ}($\delta(X)$) \in range(g), then δ is **median unbiased** if and only if $\text{median}_{\theta}(\delta(X)) = g(\theta)$.