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

Lecture 6: October 13

Lecturer: Joseph Romano — *Scribes: Kenneth Tay*

6.1 Adding UMVU Estimators

Suppose I have a model $X \sim P_\theta$ with squared-error loss. Suppose $T_1(X)$ is UMVU for $g_1(\theta)$ and $T_2(X)$ is UMVU for $g_2(\theta)$. Does it follow that $T_1(X) + T_2(X)$ is UMVU for $g_1(\theta) + g_2(\theta)$?

To answer this question, we will need the following theorem:

Theorem 6.1 *Let Δ be the set of all estimators with finite second moments, i.e.*

$$\Delta = \{\delta : \mathbb{E}_\theta \delta^2 < \infty \text{ for all } \theta\}.$$

Let $\mathcal{U}$ be the set of all unbiased estimators of zero.

Then, $\delta_0 \in \Delta$ is UMVU for $g(\theta) = \mathbb{E}_\theta \delta_0$ iff

$$Cov_\theta(\delta_0, U) = \mathbb{E}_\theta \delta_0 U = 0$$

for all θ, all $U \in \mathcal{U}$.

Proof: First, assume δ_0 is UMVU. Fix an unbiased estimator of zero U. For any $\lambda \in \mathbb{R}$, define

$$\delta_\lambda = \delta_0 + \lambda U.$$

Note that all the δ_λ's are unbiased. Since δ_0 is UMVU, we must have

$$\begin{aligned}
\text{Var}(\delta_\lambda) &= \text{Var}(\delta_0) + \lambda^2 \text{Var} U + 2\lambda \text{Cov}(\delta_0, U) \\
&\geq \text{Var}(\delta_0), \\
\lambda^2 \text{Var} U + 2\lambda \text{Cov}(\delta_0, U) &\geq 0
\end{aligned}$$

for all λ. Since the LHS is a quadratic function of λ with one root equal to 0, the only way this inequality can hold for all λ is if the other root is equal to 0 as well, i.e.

$$\begin{aligned}
-\frac{1}{2}\frac{\text{Cov}(\delta_0, U)}{\text{Var}(U)} &= 0, \\
\text{Cov}(\delta_0, U) &= 0.
\end{aligned}$$

In the other direction: assume that $\mathbb{E}_\theta \delta_0 U = 0$ for all θ and $U \in \mathcal{U}$, with δ_0 unbiased for $g(\theta)$.

Let δ be any other unbiased estimator of $g(\theta)$. Then $\delta - \delta_0$ is an unbiased estimator for zero, i.e.

$$\begin{aligned}
\mathbb{E}_\theta[\delta_0(\delta - \delta_0)] &= 0 && \text{for all } \theta, \\
\mathbb{E}_\theta \delta_0^2 &= \mathbb{E}_\theta \delta_0 \delta && \text{for all } \theta, \\
\mathbb{E}_\theta \delta_0^2 - g^2(\theta) &= \mathbb{E}_\theta \delta_0 \delta - g^2(\theta) && \text{for all } \theta, \\
\text{Var}(\delta_0) &= \text{Cov}(\delta_0, \delta) && \text{for all } \theta.
\end{aligned}$$

But the Cauchy-Schwarz Inequality tells us that

$$\mathrm{Cov}^2(\delta_0, \delta) \le \mathrm{Var}(\delta_0)\mathrm{Var}(\delta).$$

Hence, we get

$$\mathrm{Var}^2(\delta_0) \le \mathrm{Var}(\delta_0)\mathrm{Var}(\delta), \quad \Rightarrow \quad \mathrm{Var}(\delta_0) \le \mathrm{Var}(\delta),$$

as required. ■

We can use this theorem to answer our question. For any unbiased estimator of zero U, we have

$$\begin{aligned}\mathrm{Cov}(T_1 + T_2, U) &= \mathrm{Cov}(T_1, U) + \mathrm{Cov}(T_2, U) \\ &= 0 + 0 && \text{(by assumption)} \\ &= 0.\end{aligned}$$

Hence, $T_1 + T_2$ is indeed UMVU for $g_1(\theta) + g_2(\theta)$.

6.2 Location Families and Location Equivariance

Instead of using unbiasedness to restrict the class of estimators, we could use other restrictions, i.e. location equivariance.

Definition 6.2 *We say that $\delta(X_1, \dots, X_n)$ is* ***location equivariant*** *if $\delta(X_1 + c, \dots, X_n + c) = \delta(X_1, \dots, X_n) + c$ for all c.*

We say that $\delta(X_1, \dots, X_n)$ is ***location invariant*** *if $\delta(X_1 + c, \dots, X_n + c) = \delta(X_1, \dots, X_n)$ for all c.*

Definition 6.3 *A model is called a* ***location model*** *if the underlying family of distributions has the form $F(x - \theta)$, where F is a known distribution, $\theta \in \mathbb{R}$.*

In a location model, assume that the loss function is given by $L(\theta, d) = \rho(d - \theta)$ for some function ρ. This is a natural form for the loss function since shifting all values by a constant c should not affect the loss.

We seek to find the "best" location equivariant estimator, i.e. with minimum risk. Recall that in the general setting, we cannot talk about an estimator with uniformly minimum risk. However, in the location model setting, it turns out that the risk of a location equivariant estimator is constant (i.e. does not depend on θ), and so the idea of "best" makes sense. We prove this in the following proposition:

Proposition 6.4 *The bias, variance and risk functions of any location equivariant estimator are constant, i.e. do not depend on θ.*

Proof: If $X_1, \dots, X_n$ iid, $X_i \sim F(x - \theta)$, then $X_1 - \theta, \dots, X_n - \theta$ iid, $X_i - \theta \sim F(x)$.

For any θ,

$$\begin{aligned}\mathbb{E}_\theta \delta(X_1, \dots, X_n) &= \mathbb{E}_0[\delta(X_1 + \theta, \dots, X_n + \theta)] \\ &= \mathbb{E}_0[\delta(X_1, \dots, X_n)] + \theta,\end{aligned}$$

so

$$\mathrm{Bias}_\theta(\delta, \theta) = \mathbb{E}_\theta \delta - \theta = \mathbb{E}_0[\delta(X_1, \dots, X_n)]$$

does not depend on θ. Similarly,

$$\begin{aligned}\mathrm{Var}_\theta \delta(X_1,\ldots,X_n) &= \mathrm{Var}_0 \delta(X_1+\theta,\ldots,X_n+\theta)\\ &= \mathrm{Var}_0\left(\delta(X_1,\ldots,X_n)+\theta\right)\\ &= \mathrm{Var}_0\left(\delta(X_1,\ldots,X_n)\right),\end{aligned}$$

and

$$\begin{aligned}R(\delta,\theta) &= \mathbb{E}_\theta \rho\left(\delta(X_1,\ldots,X_n)-\theta\right)\\ &= \mathbb{E}_0 \rho\left(\delta(X_1+\theta,\ldots,X_n+\theta)-\theta\right)\\ &= \mathbb{E}_0 \rho\left(\delta(X_1,\ldots,X_n)\right)\end{aligned}$$

don't depend on θ. ■

6.2.1 Characterization of Location Equivariant Estimators

We have two (easy-to-prove) lemmas which will help us to characterize location equivariant estimators:

Lemma 6.5 *Let δ_0 be some location equivariant estimator.*

An estimator δ is location equivariant iff

$$\delta = \delta_0 + u,$$

where u is location invariant.

Lemma 6.6 *For $n > 1$, u is location invariant iff it is a function of the differences $Y_i = X_i - X_n$, $i = 1,\ldots,n-1$.*

Putting them together, we have the following theorem:

Theorem 6.7 *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 $Y = (X_1 - X_n,\ldots,X_{n-1}-X_n)$.

6.2.1.1 Case of $n = 1$

Say we only have $X_1 \sim F(x-\theta)$. Then the only location invariant functions are constant functions.

Since X_1 is a location equivariant estimator, Lemma 6.5 tells us that the only location equivariant estimators are $X_1 - v$ with v constant.

How can we choose v to minimize risk, i.e. $R(X_1 - v,\theta) = \mathbb{E}_\theta\rho(X_1 - v,\theta) = \mathbb{E}_0\rho(X_1 - v)$? Here, Theorem 1.7.15 from the book and its corollary help us:

Theorem 6.8 (1.7.15 in TPE) *Suppose ρ is convex on the real line, and X is some random variable. Assume that $\phi(v) = \mathbb{E}\rho(X - v)$ is finite for some v.*

If ρ is not monotone, then $\phi(v)$ has a minimum value, and the set of v^* which attain this minimum value form a closed interval.

If, in addition, ρ is strictly convex, then v^* is unique.

Corollary 6.9 *Under the assumptions of Theorem 6.8, suppose that ρ is even and X is symmetric about μ. Then $v^* = \mu$.*

Proof: If $\mu - \varepsilon$ is a minimizer of ϕ, then by symmetry $\mu + \varepsilon$ s a minimizer of ϕ too. Since $\mu \in [\mu - \varepsilon, \mu + \varepsilon]$, μ must be a minimizer too. ■

Thus, in order to minimize risk $\mathbb{E}_0 \rho(X_1 - v)$, if ρ is even, we must have $v = 0$.

6.2.1.2 The Normal Case

Let $X_1, \dots, X_n$ iid, $X_i \sim \mathcal{N}(\theta, 1)$. We have the following result:

Theorem 6.10 (Mini-Convolution Theorem) *Let δ be any location equivariant estimator of θ. Then*

$$\delta - \theta \overset{d}{=} \bar{X} - \theta + Y,$$

where Y is independent of $\bar{X}$. (Here, $\overset{d}{=}$ means "has the same distribution as".) In other words,

$$\mathcal{L}(\delta - \theta) = \mathcal{N}\left(0, \frac{1}{n}\right) * G,$$

where $Y \sim G$. (Here, $\mathcal{L}$ represents the law or distribution of a random variable.)

Proof: We write $\delta - \theta = \bar{X} - \theta + (\delta - \bar{X})$, and let $Y = \delta - \bar{X}$.

Note that Y is ancillary: Write $X_i = Z_i + \theta$, where Z_i's iid, $Z_i \sim \mathcal{N}(0,1)$. Then

$$\begin{aligned} \delta - \bar{X} &= \delta(Z_1 + \theta, \dots, Z_n + \theta) - \frac{\sum Z_i + \theta}{n} \\ &= \delta(Z_1, \dots, Z_n) - \frac{\sum Z_i}{n}, \end{aligned}$$

which does not depend on θ. By Basu's Theorem, Y must be independent of $\bar{X}$. ■

If we think of convolution as a process of "spreading out" the distribution, then for an optimal estimator, we would want as little spread as possible. This corresponds to the distribution $G =$ point mass at 0. Plugging this in, we get $\delta = \bar{X}$.

We can see the optimality of $\bar{X}$ in another way. From Theorem 6.7, we know that all location equivariant estimators for θ must be of the form $\delta = \bar{X} - v(Y_1, \dots, Y_{n-1})$. Assume that ρ is even (which is reasonable in a symmetric environment). We wish to minimize risk $\mathbb{E}_0 \rho(\bar{X} - v(Y_1, \dots, Y_{n-1}))$.

Note that by ancillarity and Basu's Theorem, $\bar{X}$ is independent of $v(Y_1, \dots, Y_{n-1})$. We can use the Law of Iterated Expectation to obtain

$$R = \mathbb{E}\left[\mathbb{E}\rho(\bar{X} - v) \mid Y = y\right].$$

For each value of y, minimizing the inner expectation is simply the case of $n = 1$, i.e. $v^* = 0$. Thus, $\bar{X}$ has minimal risk among all location equivariant estimators of θ.