STAT-510 **2017 Fall**

Lecture 6: Equivariance Principle

Instructor: Xiaohui Chen Scribe: Yutong Dai Last Modified: 2017-11-20

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

There are two types of equivariance,

- **Measurement Euqivariance**: It prescribes that the inference made should not depend on the measurement scale that is used.
- **Formal Equivariance**: If the two inference problems have the same formal structure in terms of the statistical model used, then the same inference procedure should be used in both problems. The elements of the model that must be the same are:
  - The parameter space Θ;
  - The family of distributions $\{f(x;\theta)|\theta \in \Theta\}$ for the sample;
  - The set of allowable inferences and consequences of wrong inferences.

These two types form the equivariance principle.

Definition 6.1 (Equivariance Principle) If $Y = g(X)$ is a change of measurement scale such that the model for Y has the same formal structure as the model for X, then an inference procedure should be both measurement equivariance and formal equivariance.

We use the following example to illustrate Equivariance Principle.

Example 6.2 Let $X \sim \text{binomial}(n,p)$ with n known and p unknown. Let $T(x)$ be the estimate of p that is used when $X = x$ is observed. Instead of using X to infer p, we can also use $Y = n - X$ to estimate p, where $Y \sim \text{binomial}(n,q)$ and $q = 1-p$. Let $T'(y)$ be the estimate of q that is used when $Y = y$ is observed. Then $1 - T'(y)$ is the estimate of p.

By the measurement equivariance, we require $T(x) = 1 - T'(y) = 1 - T'(n-x)$; Since both X, Y follow $\text{binomial}(n,p)$, i.e., the same statistical model, by the formal equivariance, we require $T(z) = T'(z), \forall z = 0, 1, .., n$.

Therefore, measurement equivariance and formal equivariance together require that

$$T(x) = 1 - T(n-x).$$

Equivariance Principle will greatly reduced and simplified the set of estimators we are willing to consider.

6.1 Shift-Equivariant Estimator

Proposition 6.3 Consider the location shift $y = x + a$ and assume $x \sim f(x-\theta)$. $T(x)$ is the equivariant estimator for $\theta \iff T(y) = T(x+a) = T(x) + a, \forall a \in R$.

Proof: $y \sim f(y-a-\theta) = f(y-\theta^*)$.

Suppose $T(x)$ is equivariant, then $T(y) - a = T(x)$, which implies $T(x+a) = a + T(x)$.

Suppose $T(x+a) = a + T(x)$, then just inverse the whole proof. ■

Theorem 6.4 If the underlying distribution $f(x;\theta)$ belongs to the location family, then the bias, variance and MSE of the shift-equivariant estimator of θ dose NOT depend on θ.

Proof: Denote $x = (x_1, ..., x_n)$ and assume $f(x_i) \overset{iid}{\sim} f(x-\theta)$. Let $z_i = x_i - \theta$, then $z_i \overset{iid}{\sim} f(z)$. Suppose $T(x)$ is the shift-equivariant estimator of θ. Then, by Proposition 6.3, $T(z) = T(x-\theta) = T(x) - \theta$.

As $\text{Bias}(T(X)) = E(T(x)) - \theta = E(T(Z))$, $\text{Bias}(T(X))$ doesn't depend on θ.

Since $\text{Var}(T(X)) = E[T(X) - E[T(X)]]^2 = E[T(Z) - E[T(Z)]]^2$, $\text{Var}(T(X))$ doesn't depend on θ. ■

6.1.1 Best Shift-Equivariant Estimator

Definition 6.5 The best shift-equivariant estimator is the one with minimal mean squared error.

This definition suggests that the best shift-equivariant estimator is unique, which can be easily proven by applying the Cauchy-Schwarz Inequality. Suppose, $T_1(X), T_2(X)$ are two different best shift-equivariant estimators. Denote $W(X) = \frac{1}{2}(T_1(X) + T_2(X))$ and $\text{Var}(T_1(X)) = \text{Var}(T_2(X)) = \sigma^2$.

$$\begin{aligned} \text{Var}(W(X)) &= \frac{1}{2}\sigma^2 + \frac{1}{2}\text{Cov}(T_1(X) + T_2(X)) \\ &\leq \frac{1}{2}\sigma^2 + \frac{1}{2}\sqrt{\text{Var}(T_1(X))\text{Var}(T_2(X))} = \sigma^2. \end{aligned} \tag{6.1}$$

Then equality holds $\iff T_1(X) = aT_2(X)$ (the best shift-equivariant estimators are unbiased, hence omitting the constant; the proof is given in the Proposition 6.6). Since $\text{Var}(T_1(X)) = \text{Var}(T_2(X))$, then $a = 1$. This incurs contradiction. Therefore, the best shift-equivariant estimator is unique.

Proposition 6.6

1. The best shift-equivariant estimator is unbiased.
2. If the $T(x)$ is the UMVUE and shift-equivariant, then it is the best shift-equivariant estimator.

Proof:

1. Suppose $T(x)$ is a shift-equivariant estimator. Denote $b = \text{Bias}(T(x)) = E(T(x)) - \theta$. Consider $T^*(X) = T(X) - b$, by Theorem 6.4, we know than $T^*(X)$ is an estimator. Furthermore, by Proposition 6.3, we know $T^*(X)$ is also shift-equivariant. Note that $\text{Bias}(T^*(X)) = 0$ and $MSE(T^*(X)) \leq MSE(T(X))$. All these mean that the best shift-equivariant estimator must fall into the unbiased estimator set.

2. By definition, it is true. ■

Theorem 6.7 (Existence;Pitman Estimator) The best shift-equivariant estimates given $X = x$ always exists for location family, it is given by the Pitman estimator,

$$T(x) = \frac{\int \theta \prod_{i=1}^n f(x_i;\theta)d\theta}{\int \prod_{i=1}^n f(x_i;\theta)d\theta}.$$

Proof: See [Book](http://stat.istics.net/MathStat/) P.337-338. ■

6.2 General equivariance

Definition 6.8 (Group) A set of elements $\{g : g \in G\}$ equipped with a binary operator $\circ$ is called a *group* if

1) (Closure) For every $g_1, g_2 \in G$, there exists $g \in G$ such that $g_1 \circ g_2 = g$.

2) (Associativity) For every $g_1, g_2, g_3 \in G$,

$$g_1 \circ g_2 \circ g_3 = g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3.$$

3) (Identity) $\exists e \in G$ such that for $\forall g \in G,\ e \circ g = g \circ e = g$.

4) (Inverse) $\forall g \in G, \exists\, g^{-1} \in G \text{s.t.} g^{-1} \circ g = g \circ g^{-1} = e$.

Definition 6.9 (Action) Suppose sample space is defined as $\mathcal{X} = \{(x_1, x_2, ..., x_n); x_i \in D_i\}$. Given a group G, its action onto $\mathcal{X}$ is defined by a one-to-one onto mapping (associate for each $g \in G$) $h_g : \mathcal{X}| \to \mathcal{X}$, which satisfies the following properties:

1) $\forall x \in \mathcal{X}, h_e(x) = x$;

2) $\forall x \in \mathcal{X}$ and $g_1, g_2 \in G$, $h_{g_1 \circ g_2}(x) = h_{g_1}(h_{g_2}(x))$.

Definition 6.10 Suppose the distribution of X is in the distribution family P and G is a group of transformation acts on $X \in \mathcal{X}$. The the statistical model $(X, \mathcal{X}, P)$ is invariant under group G if $\forall g \in G$, the distribution of $h_g(X)$ is also in the P.

Example 6.11 Suppose $X_i \overset{iid}{\sim} f(x - \mu)$, then the distribution of $X = (X_1, ..., X_n)$ falls in $\{\prod_i^n f(x_i - \mu) : \mu \in R\}$. Then the model $(X, \mathcal{X}, P)$ is invariant under shift group.

$\forall g \in G, h_g(x) = (x_1 + g, ..., x_n + g) = x^* \sim \prod_i^n f(x_i^* - \mu^*)$, where $\mu^* = g + \mu$.

Definition 6.12 (Induced Action) Given a Statistical model $(X, \mathcal{X}, P)$, where $P = \{p_\theta; \theta \in \Theta\}$. Suppose it is invariant under a group G. Consider the G's action on $\mathcal{X}$, then by the invariance, $h_g(X) = X^* \sim P_{\theta^*} \in P$. The induced action $\bar{h}_g$ onto Θ is defined as $h(\theta) = \theta^*$.

Example 6.13 (Example 6.11 continued) Since $\mu^* = g + \mu$, the induced action $\bar{h}_g(\mu) = \mu + g$.

Definition 6.14 (Equivariance Estimator) Given a Statistical model $(X, \mathcal{X}, P)$, where $P = \{p_\theta; \theta \in \Theta\}$. Suppose it is invariant under a group G. The estimator $W(X)$ for θ is called equivariant if $\forall g \in G$, $W(h_g(X)) = \bar{h}_g(W(X))$.

Example 6.15 The estimator $W(X) = \bar{X}_n$ is an equivariance estimator for the location-scale family $\{\prod_i \frac{1}{\sigma} f(\frac{x_i - \mu}{\sigma}); \mu \in R\}$.

The action is defined as $h_g(x) = (x_1 + g, ..., x_n + g)$,then $W(h_g(x)) = \bar{x}_n + g$. Since the $h_g(x) \sim \prod_i \frac{1}{\sigma} f(\frac{x_i - \mu^*}{\sigma})$, where $\mu^* = \mu + g$, we know the induced action is $\bar{h}_g(\mu) = \mu + g$, further $\bar{h}_g(W(x)) = \bar{x}_n + g$. So the $W(X)$ is an equivariant estimator for location-scale family.