Jimin Ding
Math 5062 Spring 2018 Scribe: Jiahui Lyu; Hiro Ono

6 Lecture 6: Invariant Estimation under Group Transformations

6.1 Invariant Problem under a Transformation

Let g be a bijection (1-1 and onto) of a sample space $\mathcal{X}$;
Let $\mathcal{P}$ be a family of distributions on $\mathcal{X}$, then

Definition 1: Invariance Model under a transformation. $\mathcal{P}$ is invariant to g if

$$X \sim P \in \mathcal{P}, \Rightarrow gX \sim P^* \in \mathcal{P}$$

and

$$\forall P^* \in \mathcal{P}, \exists P \text{ s.t. } X \sim P \to gX \sim P^*$$

The second condition ensures that g does not reduce the model.

Example: Normal with unknown variance under scale transformation. Let $g(x) = 2x$, and

$$\mathcal{P}_A = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$$
$$\mathcal{P}_B = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 1\}$$

Then $\mathcal{P}_A$ is invariant to g but $\mathcal{P}_B$ is not, as g reduces $\mathcal{P}_B$.

Induced Transformation: Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a model that is invariant under g. When $X \sim P_\theta, \theta \in \Theta$, then $\exists \theta^* \in \Theta s.t. gX \sim P_{\theta^*}$. Define $\bar{g}$ to be the transformation on the parameter space:

$$\bar{g} : \quad \Theta \to \Theta$$
$$\theta \mapsto \theta^*.$$

The function $\bar{g}$ is well-defined **iff** the model (parametrization) is identifiable. In other words, $\forall \theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$. In particular, the identifiability makes g a bijection:

Lemma 6.1. *If $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is invariant under a bijection (one-to-one and onto) g, and the model is identifiable, then $\bar{g}$ is a bijection.*

Proof.

1. onto: That g is onto follows from the second condition in the definition of model invariance under a transformation ($\forall P^* \in \mathcal{P}, \exists P$ s.t. $X \sim P \to gX \sim P^*$);

2. 1-to-1: To show that g is 1-to-1, if $\forall \theta^*$ s.t. $\bar{g}\theta_1 = \bar{g}\theta_2 = \theta^*$, we want to show that $\theta_1 = \theta_2$. By the identifiability, we only need to show that $P_{\theta_1} = P_{\theta_2}$. Then $\forall A$ (measurable sets) $\in \sigma(\mathcal{X})$

$$P_{\theta_1}(X \in A) = P_{\theta^*}(gX \in gA) = P_{\theta_2}(X \in A),$$

since $\{X \in A\} \equiv \{gX \in gA\}$. Thus, $P_{\theta_1} = P_{\theta_2}$, and by the identifiability of the model, $\Rightarrow \theta_1 = \theta_2$. □

Example: Recall Normal with unknown variance under scale transformation.
$\mathcal{P}_A = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$, $g(X) = a + bX$, $\bar{g}(\mu, \sigma^2) = (a + b\mu, b^2\sigma^2)$.
Here, $\Theta = \mathbb{R} \times \mathbb{R}^+, \bar{g}\Theta = \mathbb{R} \times \mathbb{R}^+$.

Definition 2: Invariant loss under a transformation. Let $\mathcal{P}$ be invariant under g (so $\bar{g}\Theta = \Theta$). A loss function $L(\theta, a) : \Theta \times \mathbb{A} \to \mathbb{R}^+$ is invariant if

$$\forall a \in \mathbb{A}, \exists!\ a^* \in \mathbb{A} \text{ s.t. } L(\theta, a) = L(\bar{g}\theta, a^*) \forall \theta \in \Theta$$

Induced Transformation: Denote $\tilde{g}a := a^*$, then $\tilde{g}$ is a bijection on $\mathbb{A}$:

$$\begin{aligned} \tilde{g} :\quad & \mathbb{A} \to \mathbb{A} \\ & a \mapsto a^*. \end{aligned}$$

$\tilde{g}$ is referred to as the induced transformation for decisions.

Examples: Normal under linear transformation.
$\mathcal{P}_A = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$
$g(X) = c + bX, \bar{g}(\mu, \sigma^2) = (c + b\mu, b^2\sigma^2)$
$\bar{g}(\mu, \sigma^2) = (c + b\mu, b^2\sigma^2)$
1. Squared error loss: $L_1(\theta, a) = (\mu - a)^2$.

$$\Rightarrow L_1(\bar{g}\theta, a^*) = (c + b\mu - a^*)^2.$$

If L_1 is invariant under g, then

$$L(\theta, a) = L(\bar{g}\theta, a^*), \quad \forall \mu \in \mathbb{R}.$$
$$\Leftrightarrow (\mu - a)^2 = (c + b\mu - a^*)^2, \quad \forall \mu \in \mathbb{R}$$

However, in this case, this is not possible unless $b = 1$ since the solution of above equation $a^* = a + c + (b - 1)\mu$ depends on μ. Hence squared error loss is not invariant under g.

2. Scaled squared error loss: $L_2(\theta, a) = \frac{(\mu-a)^2}{\sigma^2}$ (standardized by the variance).

$$\Rightarrow L_2(\bar{g}\theta, a^*) = \frac{(c + b\mu - a^*)^2}{b^2\sigma^2} = (\mu - \frac{a^* - c}{b})^2/\sigma^2.$$

Setting $L(\theta, a) = L(\bar{g}\theta, a^*)$, then

$$\frac{a^* - c}{b} = a \quad \Rightarrow \quad a^* = c + ba,$$

So $\tilde{g}a = c + ba$ is the induced transformation for decisions, and L_2 is invariant under g.

Definition 3: Invariant decision problem. A decision problem $(\Theta, \mathbb{A}, L)$ is invariant under g if

- invariant parameter space under the induced transformation $\bar{g}$ (defnition 1)
- invariant loss function under the induced transformation $\tilde{g}$ (definition 2)

6.2 Invariance Under a Group

Typically, if a problem in invariant under a transformation g, it is also invariant under a class of related transformations. We always take such a class to be a group.

Recall: a **Group** $\mathcal{G} = \{g : g \in \mathcal{G}\}$ is a class of transformations s.t.:

1. $\forall g_1, g_2 \in \mathcal{G}, g_1 \circ g_2 \in \mathcal{G}$
2. $\forall g \in \mathcal{G}, g^{-1} \in \mathcal{G}$ and $g \circ g^{-1} = g^{-1} \circ g = I$

Examples:

1. Location shift: Additive group.

2. Scale transformations: multiplicative group.

3. Linear transformations. Let $\mathcal{X} = \mathbb{R}$ then $\mathcal{G} = \{g : g(x) = c + bx, c \in \mathbb{R}, b \in \mathbb{R}\ \{0\}\}$ is a group on $\mathcal{X}$.

Recall normal model under linear transformation.
$\mathcal{P}_A = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$
$\mathcal{G} = \{g : g(x) = c + bx, c \in \mathbb{R}, b \in \mathbb{R}\ \{0\}\}$,
$L(\theta, a) = (\mu - a)^2/\sigma^2$.
For each single $g(x) = c + bx \in \mathcal{G}$, the induced transformations on Θ an $\mathbb{A}$ are $\bar{g}(\mu, \sigma^2) = (c + b\mu, b^2\sigma^2)$ and $\tilde{g}a = c + ba$ respectively.
We have known this decision problem $(\Theta, \mathbb{A}, L)$ is invariant under each $g \in \mathcal{G}$. Therefore, it is invariant under the group $\mathcal{G}$. Now consider the class of $\bar{g}$ and $\tilde{g}$:

$$\bar{\mathcal{G}} = \{\bar{g} : g \in \mathcal{G}\}, \tilde{\mathcal{G}} = \{\tilde{g} : g \in \mathcal{G}\},$$

are referred to as **Induced Groups**.

Lemma 6.2. *If $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is invariant under a group $\mathcal{G}$, then $\bar{\mathcal{G}} = \{\bar{g} : g \in \mathcal{G}\}$ and $\tilde{\mathcal{G}} = \{\tilde{g} : g \in \mathcal{G}\}$ are also groups.*

Examples:

1. Scale transformation
$\mathcal{P} = \{E(0, \theta) : \theta \in \Theta = \mathbb{R}\}, f(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$
$\mathcal{G} = \{g : g(x) = cx, c \in \mathbb{R}^+\}$
$L(\theta, a) = (1 - \frac{a}{\theta})^2 = \frac{(\theta - a)^2}{\theta^2}$. Then

- Invariant model under $\mathcal{G}$ (definition 1): for each $g \in \mathcal{G}$,
  - If $X \sim P_\theta, \theta \sim \Theta$, then $gX = X^* \sim P_{\theta^*}, \theta^* = c\theta$
  - If $P_{\theta^*} \in \mathcal{P}, \forall \theta^* \in \Theta, \exists \theta$ s.t. $\theta^* = c\theta$
- Induced group on Θ : $\bar{\mathcal{G}} = \{\bar{g} : \bar{g}\theta = c\theta, c \in \mathbb{R}^+\}$

- Invariant loss under $\mathcal{G}$ (definition 2): $\forall g \in \mathcal{G}, L(\theta, a) = L(\bar{g}\theta, a^*)$.
- Induced group on $\mathbb{A}$:

$$\begin{aligned} L(\bar{g}\theta, a^*) &= \frac{(c\theta - a^*)^2}{(c\theta)^2} = \frac{(c\theta - ca)^2}{(c\theta)^2} \\ &= \frac{(\theta - a)^2}{\theta^2} = L(\theta, a) \end{aligned}$$

$\Rightarrow \tilde{g}a = ca$, then we have the group $\tilde{\mathcal{G}} = \{\tilde{g} : \tilde{g}a = ca, c \in \mathbb{R}^+\}$.

- Invariant decision problem under $\mathcal{G}$ (definition 3).

Note that $\mathcal{G}, \bar{\mathcal{G}}$ and $\tilde{\mathcal{G}}$ are all the same group (all scale transformations). But this is not always the case, shown as follows.

2. Location and scale transformation
$\mathcal{P}_A = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$
$g(X) = c + bX, \tilde{g}a = c + ba, \bar{g}(\mu, \sigma^2) = (c + b\mu, b^2\sigma^2) \Rightarrow \mathcal{G} = \tilde{\mathcal{G}} \neq \bar{\mathcal{G}}$.

6.3 Invariant Decision Rule

Invariant principles: $X \sim P \in \mathcal{P}, X^* \sim P_{\theta^*} \in \mathcal{P}, L(\theta, a) = L(g\theta, \tilde{g}a)$; a good decision rule δ to estimate θ using X

- Formal invariance:

$$\hat{\theta^*} = \widehat{\bar{g}\theta} \quad \Leftrightarrow \quad \delta(X^*) = \delta(gX)$$

- Functional invariance

$$\hat{\theta^*} = \tilde{g}\hat{\theta} \quad \Leftrightarrow \quad \delta(X^*) = \tilde{g}\delta(X)$$

Combining the above two types of invariance we have an invariant decision rule.

Definition 4: Invariant Decision Rule. For a decision problem invariant under a group $\mathcal{G}$, an estimator is invariant if

$$\delta(gX) = \tilde{g}(\delta(X)), \forall g \in \mathcal{G}$$

Theorem 6.3. *The risk of an invariant estimation satisfies the following condition:*

$$R(\theta, \delta) = R(\bar{g}\theta, \delta), \forall \theta \in \Theta, \bar{g} \in \mathcal{G}.$$

Proof.

$$\begin{aligned} R(\theta, \delta) &= E_\theta[L(\theta, \delta(X))] \\ &= E_\theta[L(\bar{g}\theta, \tilde{g}\delta(X))] \text{ (invariant loss function)} \\ &= E_\theta[L(\bar{\theta}, \delta(gX))] \text{ (equivalence of } \delta : \delta(gX) = \tilde{g}\delta(X)) \\ &= E_{\theta^*}[L(\theta^*, \delta(X^*))] \\ &= R(\theta^*, \delta) \\ &= R(\bar{g}\theta, \delta) \end{aligned}$$

□

Definition. *Two points* $\theta_0, \theta_1 \in \Theta$ *are said to be equivalent if* $\exists \bar{g}$ *s.t.* $\bar{g}\theta_1 = \theta_2, \bar{g} \in \bar{\mathcal{G}}$. *The set of all such equivalent points is called an orbit:*

$$\Theta(\theta_0) = \{\bar{g}(\theta_0) : \bar{g} \in \bar{\mathcal{G}}\}$$

If all points in Θ *are equivalent (a single orbit defined by the group), then we say* $\bar{\mathcal{G}}$ *is* ***transitive***.

If $\bar{G}$ is transitive, for any θ, θ_0 we can choose $\bar{g} \in \bar{G}$ such that $\bar{g}(\theta) = \theta_0$. Combined this and the previous theorem, we obtain the next Corollary.

Corollary 6.4. *If* $\bar{G}$ *is transitive, then the risk of all invariant estimators is constant over parameter space.*

Theorem 6.5. *MRIE is risk-unbiased if* $\bar{G}$ *is transitive and* $\tilde{G}$ *is commutative.*

Examples: The followings are examples of commutative and non-commutative groups $\tilde{G}$.

1. Location transformation $\tilde{G}$ is commutative.
2. Scale transformation $\tilde{G}$ is commutative.
3. Scale and location transformation $\tilde{G}$ is not commutative.

Related Reading

1. Sh P256-267
2. LC chapter 3.1