Jimin Ding
Math 5062 Spring 2018 Scribe:Bei Wu

4 Lecture 4: Invariance/Equivariance

Motivating example: Let $X_1, \cdots, X_n \stackrel{\text{iid}}{\sim} Ber(p)$, $p \in (0,1)$. Let $\hat{p} = \delta(\boldsymbol{x}) = \delta(x_1, \cdots, x_n)$ be a decision rule. Consider $X_1^*, \cdots, X_n^* = 1 - X_1, \cdots, 1 - X_n \stackrel{\text{iid}}{\sim} Ber(p^*)$, where $p^* = 1 - p$. it is natural to use the same decision rule to estimate p^*, then

$$\hat{p^*} = \delta(\boldsymbol{x^*}) = \delta(x_1^*, \cdots, x_n^*) = \delta(1 - x_1, \cdots, 1 - x_n)$$

Also,we have already estimate $\hat{p}$, then, under the same transformation, we can estimate p^* as:

$$\hat{p^*} = 1 - \hat{p} = 1 - \delta(x_1, \cdots, x_n)$$

It is natural to request a decision rule δ such that $1 - \delta(x_1, \cdots, x_n) = \delta(1 - x_1, \cdots, 1 - x_n)$, which is a decision rule invariant under transformation. For example, $\delta(\boldsymbol{x}) = 1/n \sum_{i=1}^n x_i$ is invariant.

Recall: a **Group** $\mathcal{G} = \{g : g \in \mathcal{G}\}$ is a class of transformation 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 $g \circ g^{-1} = g^{-1} \circ g = I$

note: Location transformation is a group.

4.1 Location invariant

Location family: is invariant under location transformations

Definition 1. $\mathcal{P} = \{f(x, \theta); \theta \in \Theta\}$ is location invariant if $f(x^*; theta^*) = f(x; theta)$, where $x^* = x + c, \theta^* = \theta + c, \forall c \in \mathbb{R}$. c is location shift.
$(f(\boldsymbol{x}, \theta) = f(\boldsymbol{x} - \theta)$)

Examples:

1.$\mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2$ is known$\}$

2.$\mathcal{P} = \{E(\mu, \theta) : \mu \in \mathbb{R}, \theta$ is known$\}$

2. Loss function $L(\theta, a)$ is location invariant if $L(\theta^*, a^*) = L(\theta, a)$, where $\theta^* = \theta + c, a^* = a + c, \forall c \in \mathbb{R}$.
$(L(\theta, a) = L(\theta - a)$)

Examples:

1.$L(\theta, a) = (g(\theta) - a)^2$ when $g(\theta) = \theta$

2.(counter example) for motivating example, define $L(p,\delta) = \frac{(\delta-p)^2}{p(1-p)}$, then $L(p,\delta)$ is invariant under transformation $1-\delta, 1-p$, but not location invariant.

3. Estimation problem: $\hat{g}(\theta)$ is location invariant if the distribution family and loss functions are both location invariant.

4. An estimator δ is location invariant if $\delta(x^*) = \delta(x+c) = \delta(x) + c, \forall c \in \mathbb{R}$

MRIE minimum risk invariant estimator.

4.2 Properties of location invariant estimators

Theorem 4.1. *The bias, variance and risk of location invariance estimators are constant.(independent of θ)*

proof of bias is constant. The location invariant family has p.d.f. $f(\boldsymbol{x}, \theta) = f(\boldsymbol{x} - \theta) = f(x_1 - \theta, \cdots, x_n - \theta)$

$$\begin{aligned} bias = E(\delta(\boldsymbol{x})) - \theta &= \int \delta(\boldsymbol{x}) f(\boldsymbol{x} - \theta) dx - \theta \\ &= \int_{\mathbb{R}^n} \delta(x_1, \cdots, x_n) f(x_1 - \theta, \cdots, x_n - \theta) dx_1 \cdots dx_n - \theta \\ &= \int_{\mathbb{R}^n} \delta(s_1 + \theta, \cdots, s_n + \theta) f(s_1, \cdots, s_n) ds_1 \cdots ds_n - \theta \\ &= \int_{\mathbb{R}^n} [\delta(s_1, \cdots, s_n) + \theta] f(s_1, \cdots, s_n) ds_1 \cdots ds_n - \theta \\ &= \int_{\mathbb{R}^n} [\delta(s_1, \cdots, s_n)] f(s_1, \cdots, s_n) ds_1 \cdots ds_n \end{aligned} \tag{1}$$

Bias of δ is independent of θ, thus it is a constant. □

To find MRIE, we only need to compare the constant risks and find the δ^* which has the smallest constant risk.

Lemma 4.2. *Let δ_0 be a given location invariant estimator. Then any location invariant estimator δ satisfies*

$$\delta(x) = \delta_0(x) + u(x)$$

where $u(x+c) = u(x), \forall c \in \mathbb{R}, \forall x \in \mathbb{R}^n$

Proof. 1. $\delta_0(x) + u(x)$ is location invariant since $\delta_0(x) + u(x+c) = \delta_0(x) + c + u(x)$

2. Let $\delta(x)$ be any location invariant. Set $u(x) = \delta(x) - \delta_0(x)$, then

$$u(x+c) = \delta(x+c) - \delta_0(x+c) = \delta(x) + c - \delta_0(x) - c = u(x)$$

□

$u(x+c) = u(x), u(x_1+c,\cdots,x_n+c) = u(x_1,\cdots,x_n), \forall c \in \mathbb{R}$ set $c = -x_n$, then

$$u(x_1 - x_n, x_2 - x_n, \cdots, x_{n-1} - x_n, 0) = u(x_1, \cdots, x_n)$$

So u is a function in $\mathbb{R}^{n-1}$ and is a function of ancillary statistic $x_1-x_n, x_2-x_n, \cdots, x_{n-1}-x_n$

Theorem 4.3. *Let δ_0 be a location invariant estimator. Let $d_i = x_i - x_n, i = 1, \cdots, n-1$, and $d = (d_i, \cdots, d_n)$.Then a necessary and sufficient condition for δ is also location invariant is that $\exists$ a function u of n-1 arguments for which*

$$\delta(x) = \delta_0(x) + u(d), \forall x \in \mathbb{R}$$

Theorem 4.4. *Let $D = (x_1 - x_n, \cdots, x_{n-1} - x_n)^T$. Suppose $\exists$ a location invariant estimator δ_0 such that δ_0 has a finite risk. Assume $\forall y, \exists$ a $u^*(y)$ which minimizes $E_0[L(\delta_0(x) - u(y))|D = y]$. Then the MRIE exists and is $\delta^*(x) = \delta_0(x) - u^*(y)$.*

idea of proof. If a loss function is location invariant, it can be written as $L(\theta, \delta) = L(\theta - \delta)$. Then $R(\delta, \theta) = E(L(\theta, \delta)] = E(L(\theta - \delta)] = E_0[L(\delta)]$, since risk is independent of θ, we can set $\theta = 0$. And from theorem 4.3

$$E_0[L(\delta(x)] = E_0[L(\delta_0(x) - u(y))] = E[E_0[L(\delta_0(x) - u(y))|D = y]]$$

Thus if $u^*(y)$ minimize $E_0[L(\delta_0(x) - u(y))|D = y]$, $\delta^*(x) = \delta_0(x) - u^*(y)$ minimize the risk function $R(\delta, \theta)$ □

Corollary 4.5. *Suppose L is convex and not monotone, then MRIE exists. Furthermore, if L is strictly convex, then MRIE is unique.*

Examples:

1. If $L(\theta, a) = (\theta - a)^2$, then $u^*(y) = E_0[\delta_0(x)|D = y]$
2. If $L(\theta, a) = |\theta - a|$, then $u^*(y)$ is the conditional median of $\delta_0(x)$ given D.

4.3 Related Reading

1. Sh P251-255
2. LC Chapter 3.1

5 Lecture 5: More on Invariance/Equivariance

5.1 Properties of location invariant estimators

Examples

1. Let $X_1, X_2, \ldots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$ with an unknown $\mu \in \mathbb{R}$ and a known σ^2. Consider squared error loss function, $\delta_0(x) = \bar{X}$ is location invariant.
   By Basu's theorem, $D = (x_1 - x_n, \cdots, x_{n-1} - x_n)^T$ and $\bar{X}$ are independent, so
   $$u^*(d) = E_0(\bar{X}|x_1 - x_n, x_2 - x_n, \ldots, x_{n-1} - x_n) = E_0(\bar{X}) = 0.$$
   Furthermore, $\bar{X}$ is MRIE for all convex and even loss functions, since
   $$E_0\left[L\left(\delta_0(x) - u(d)\right)|D = d\right] = E_0\left[L\left(\delta_0(x) - u(d)\right)\right] = \int_{\mathbb{R}^n} L\left(\delta_0(x) - u(d)\right) f_0(x)dx$$
   is minimized if and only if $u(d) = 0$, since $f_0(x)$ is also even.

2. Let $X_1, X_2, \ldots, X_n$ be i.i.d. from the exponential distribution $E(\mu, 1)$ with an unknown $\mu \in \mathbb{R}$. $\delta_0(x) = X_{(1)}$ is location invariant. By Basu's theorem, $X_{(1)}$ is independent of D. We want to minimize $E_0\left[L\left(X_{(1)} - u(d)\right)|D = d\right] = E_0\left[L\left(X_{(1)} - u(d)\right)\right]$.
   If consider squared error loss function, then $u^* = E_0(X_{(1)}) = \frac{1}{n}$ since $X_{(1)} \sim E(0, n)$. And the MRIE is $\delta^*(x) = X_{(1)} - \frac{1}{n}$.
   If consider absolute loss function, then $u^* = median_0(X_{(1)}) = \frac{\log 2}{n}$ since
   $$\frac{1}{2} = F_1(x) = 1 - e^{-nx}, \quad F_1(x) \text{ is c.d.f. of } X_{(1)} \sim E(0, n).$$
   And the MRIE is $\delta^*(x) = X_{(1)} - \frac{\log 2}{n}$.

Theorem 5.1 (**Pitman Estimator**). *If we have location invariant estimation problem with squared error loss function $L(\theta - a) = (\theta - a)^2$, then*
$$\delta^*(x) = \frac{\int_{-\infty}^{\infty} uf(X_1 - u, X_2 - u, \ldots, X_n - u)du}{\int_{-\infty}^{\infty} f(X_1 - u, X_2 - u, \ldots, X_n - u)du}$$
is the MRIE of θ which is known as the Pitman estimator and is unbiased.

Proof. Consider $\delta_0(x) = X_n$, $u^*(d) = E_0(X_n|D = d)$.
Consider $1 - 1$ transformation: $X_1, \ldots, X_n \to Y_1, \ldots, Y_n$, where
$$Y_1 = X_1 - X_n, \ldots, Y_{n-1} = X_{n-1} - X_n, Y_n = X_n.$$

$$
\begin{aligned}
u^*(d) &= E_0(X_n|D=d) = E_0(Y_n|Y_1=d_1,\ldots,Y_{n-1}=d_{n-1}) \\
&= \int y_n f_{Y_n|Y_1,\ldots,Y_{n-1}} dy_n = \int y_n \frac{f_{Y_1,\ldots,Y_{n-1},Y_n}}{f_{Y_1,\ldots,Y_{n-1}}} dy_n \\
&= \int y_n \frac{f(y_1+y_n,\ldots,y_{n-1}+y_n,y_n)}{\int f(y_1+y_n,\ldots,y_{n-1}+y_n,y_n)dy_n} dy_n \\
&= \frac{\int y_n f(y_1+y_n,\ldots,y_{n-1}+y_n,y_n)dy_n}{\int f(y_1+y_n,\ldots,y_{n-1}+y_n,y_n)dy_n}, \qquad \text{let } y_n = x_n - u \\
&= \frac{\int (x_n-u) f(y_1+x_n-u,\ldots,y_{n-1}+x_n-u,x_n-u)d(-u)}{\int f(y_1+x_n-u,\ldots,y_{n-1}+x_n-u,x_n-u)d(-u)}, \qquad \text{let } y_i = x_i - x_n, y_n = x_n \\
&= \frac{\int (x_n-u) f(x_1-x_n+x_n-u,\ldots,x_{n-1}-x_n+x_n-u,x_n-u)d(-u)}{\int f(x_1-x_n+x_n-u,\ldots,x_{n-1}-x_n+x_n-u,x_n-u)d(-u)} \\
&= \frac{\int (x_n-u) f(x_1-u,\ldots,x_{n-1}-u,x_n-u)du}{\int f(x_1-u,\ldots,x_{n-1}-u,x_n-u)du} \\
&= \frac{x_n\int f(x_1-u,\ldots,x_{n-1}-u,x_n-u)du}{\int f(x_1-u,\ldots,x_{n-1}-u,x_n-u)du} - \frac{\int u f(x_1-u,\ldots,x_{n-1}-u,x_n-u)du}{\int f(x_1-u,\ldots,x_{n-1}-u,x_n-u)du} \\
&= x_n - \frac{\int u f(x_1-u,\ldots,x_{n-1}-u,x_n-u)du}{\int f(x_1-u,\ldots,x_{n-1}-u,x_n-u)du} \\
&= \delta_0(x) - \delta^*(x)
\end{aligned}
$$

Thus

$$
\delta^*(x) = \frac{\int u f(x_1-u,\ldots,x_{n-1}-u,x_n-u)du}{\int f(x_1-u,\ldots,x_{n-1}-u,x_n-u)du}.
$$

Let b be the constant bias of δ^*, then $\delta_1(x) = \delta^*(x) - b$ is a location invariant estimator of θ and

$$
R(\delta_1,\theta) = E\left[(\delta^*(x) - b - \theta)^2\right] = Var(\delta^*) \le Var(\delta^*) + b^2 = R(\delta^*,\theta).
$$

since δ^* is the MRIE, $b = 0$, then δ^* is unbiased. □

<u>Risk-unbiasness</u>: An estimator $\delta(x)$ for $g(\theta)$ is risk-unbiased if

$$
E_\theta\left[L\left(\theta,\delta(x)\right)\right] \le E_\theta\left[L\left(\theta',\delta(x)\right)\right], \quad \forall \theta' \ne \theta.
$$

Interpretation: $\delta(x)$ has the smallest risk at θ.

Theorem 5.2. *The MRIE of θ (location parameter) in a location invariant estimation problem (or decision problem) is risk-unbiased.*

5.2 Other non-convex loss functions

Theorem 5.3. *Suppose $0 \le L(t) \le M$ for all values of t and $L(t) \to M$ as $t \to +\infty$, and is risk-unbiased. The density of X is continuous a.e.. Then an MRIE of μ exists.*

Examples: MRIE under 0-1 loss
$L(\theta, a) = \mathbf{1}(|\theta - a| > k)$ with some known constant $k > 0$, u^* maximizes $P_0(|X - U| \leq k)$.
Suppose symmetric f, if it has unique mode, then $u^* = 0$.
For example, $N(\mu, \sigma^2)$ with $f(x; \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2(x-\mu)^2}\}$, $\delta_0 = \bar{X}$, $u^* = 0$, $\delta^*(x) = \bar{X}$.

Remarks: an MRIE (comparing with UMVUE)

1. When loss function is non-convex, MRIE typically still exists.
2. MRIE depends on the loss function even for convex loss function.
3. MRIE is often admissible (unlike UMVUE).
4. MRIE is often considered in location-scale families.
   (UMVUE is more for exponential families and UMVUE for location families usually does not exist.)

5.3 Related Reading

1. Sh P253-255
2. LC Chapter 3.1