STA732
Statistical Inference

Lecture 07: Equivariance

Yuansi Chen

Spring 2023

Duke University

https://www2.stat.duke.edu/courses/Spring23/sta732.01/

Recap from Lecture 06

1. Log-likelihood, score and Fisher information
2. Cramér-Rao lower bound and Hammersley-Chapman-Robbins ineq via **Cauchy-Schwarz**
3. UMVU has nice theory, but unbiased estimators are not always admissible

Goal of Lecture 07

1. Formulate equivariant estimation under location models
    - Location family
    - Location invariant loss
    - Location equivariant estimator
2. Formulate general equivariant estimation with group theory basics
3. Maximal invariant statistic

Chap. 10.1 in Keener or Chap. 3 in Lehmann and Casella

Where we are: to argue for the “the best” estimator in point estimation by restricting to a smaller set of estimators

- We have finished restricting to unbiased estimators
- We turn to restricting to estimators with ”symmetry”

Formulate equivariant estimation under location models

Location model

Def. location model

In a location model, the data $X = (X_1, \ldots, X_n)$ follows a joint probability density of the form

$$p_\theta(x) = f(x_1 - \theta, x_2 - \theta, \ldots, x_n - \theta)$$

where f is fixed and known, $\theta \in \mathbb{R}$ is the **location parameter**.

We denote $(X_1, \ldots, X_n) \sim \text{LocModel}(\theta)$. The familly $\mathscr{P} = \{\text{LocModel}(\theta), \theta \in \mathbb{R}\}$ is called a **location family**

Example: estimating mean from i.i.d. Gaussian samples

Property of location family

If we transform the data as follows

$$X_i' = X_i + a$$

Then the new data $X' = (X_1', \dots, X_n')$ has the joint density

$$p_\theta(x' - a) = f(x_1' - \theta - a, \dots, x_n' - \theta - a) = p_{\theta+a}(x')$$

We can also verify that $p_\theta(x) = p_{\theta+a}(x + a)$. Estimating $\theta + a$ from X' is the same problem as estimating θ from X!

We say that the family of distribution behaves naturally under location shift

Location invariant loss

Def. location invariant loss

A loss function L for the location parameter θ in a location family is called **invariant** if

$$L(\theta + a, d + a) = L(\theta, d), \forall a \in \mathbb{R}, \theta \in \mathbb{R}, d \in \mathbb{R}$$

In this case, we can write

$$L(\theta, d) = \rho(d - \theta)$$

if we define $\rho(x) = L(0, x)$.

Examples?

Location equivariant estimator

Motivation: since both the location family and the loss have invariance to location shifts, it is reasonable to restrict our estimator to respect this invariance

Def. location equivariant estimator

An estimator δ for the location parameter θ in a location family is **equivariant** of

$$\delta(X_1 + a, \dots, X_n + a) = \delta(X_1, \dots, X_n) + a$$

Examples of location equivariant estimators

Check the following estimators are equivariant for any location family

- $\frac{X_1+X_2+\ldots+X_n}{n}$
- median(X_i)
- $X_{(1)}$
- $\sum_{i=1}^n \alpha_i X_{(i)}$ for any fixed α_i, $\sum_{i=1}^n \alpha_i = 1$.

Equivariance makes risk constant

Thm. Page 197 in Keener, 3.1.4 in Lehmann and Casella

Given location family. Suppose δ is equivariant and L is invariant. Then the bias, risk and variance of δ has no θ dependence.

Equivariance makes risk constant

Thm. Page 197 in Keener, 3.1.4 in Lehmann and Casella

Given location family. Suppose δ is equivariant and L is invariant. Then the bias, risk and variance of δ has no θ dependence.

proof for the risk:

$$\begin{aligned} R(\theta, \delta) &= \mathbb{E}_\theta \rho(\delta(X) - \theta) \\ &= \mathbb{E}_\theta \rho(\delta(X - \theta + \theta) - \theta) \\ &\overset{(i)}{=} \mathbb{E}_\theta \rho(\delta(X - \theta)) \\ &\overset{(ii)}{=} \mathbb{E}_0 \rho(\delta(X')) \end{aligned}$$

(i) follows from equivariance, (ii) because of location family

Draw conceptual pictures for risk comparison (general, UMVU, equivariant estimation)

Minimum risk equivariant (MRE) estimator

Since the risk does not depend on θ, the graphs of risk functions for equivariant estimators cannot cross, we anticipate there will be a best equivariant estimator δ^*, called the minimum risk equivariant (MRE) estimator

If the set of equivariant estimators is not empty, the infimum exists. It is a question whether the infimum can be achieved.

Formulate general equivariant estimation with group theory basics

Group basics

A group is a set G together with a binary operation $* : G \times G \to G$, that satisfies

- **Associativity:** $(a * b) * c = a * (b * c), \quad \forall a, b, c \in G$
- **Identity:** $\exists e \in G, e * a = a, a * e = a, \quad \forall a \in G$
- **Inverse:** $\forall a \in G, \exists b \in G, a * b = e, b * a = e$

Group basics

A group is a set G together with a binary operation $* : G \times G \to G$, that satisfies

- **Associativity:** $(a * b) * c = a * (b * c), \quad \forall a, b, c \in G$
- **Identity:** $\exists e \in G, e * a = a, a * e = a, \quad \forall a \in G$
- **Inverse:** $\forall a \in G, \exists b \in G, a * b = e, b * a = e$

Examples

- $\mathbb{R}$ with addition, identity is 0
- $\mathbb{Z}$ with addition, identity is 0
- $\mathbb{R} \setminus \{0\}$ with multiplication, identitfy is 1
- $n \times n$ invertible matrices with matrix multiplication, identity $\mathbb{I}_n$

Group action on a set

Given a group G with identity e, and N is a set. a (left) group action of G on N is a function $\star : G \times N \to N$, that satisfies

- **Identity:** $e \star x = x, \quad \forall x \in N$
- **Compability:** $g \star (h \star x) = (gh) \star x, \quad \forall g, h \in G, x \in N$

Group action on a set

Given a group G with identity e, and N is a set. a (left) group action of G on N is a function $\star : G \times N \to N$, that satisfies

- **Identity:** $e \star x = x, \quad \forall x \in N$
- **Compability:** $g \star (h \star x) = (gh) \star x, \quad \forall g, h \in G, x \in N$

Examples

- G is $\mathbb{R}$ with addition, N is $\mathbb{R}$
- G is $n \times n$ invertible matrices, N is $\mathbb{R}^n$
- G is the rotation group, N is $\mathbb{R}^2$

Equivariance in group theory

Given a group G, two sets M and N endowed with group action $\star$. A map $F : M \to N$ is called equivariant with respect to the group action $\star$ if

$$F(g \star x) = g \star F(x)$$

for all $x \in M, g \in G$.

Conceptually, what is needed to formulate an equivariant estimation?

1. The family of distributions must behave naturally under group actions
2. The loss must be invariant
3. The estimator is restricted to be equivariant

1. Family behaves naturally

Given a family $\mathscr{P} = \{P_\theta, \theta \in \Omega\}$ and a group G. Suppose the group action $\star : G \times \mathbf{S} \to \mathbf{S}$ leaves the model invariant. If $g \star X$ has distribution $P_{\theta'}$, then $\theta' = \bar{g} \star \theta$, where $\bar{g}$ is a one-to-one mapping from Ω to Ω. So $\bar{g}$ form a group denoted by $\bar{G}$. We have for any event set B

$$P_{\bar{g}\star\theta}(X \in B) = P_\theta(g \star X \in B)$$

2. The loss must be invariant

A loss function L is invariant under the group action $\star$ if

$$L(\bar{g} \star \theta, \bar{g} \star d) = L(\theta, d)$$

for all $g \in G, \theta \in \Omega, d \in \Omega$.

3. The estimator is restricted to be equivariant

An estimator δ is equivariant with respect to the group action if

$$\delta(g \star x) = \bar{g} \star \delta(x).$$

Equivariance makes risk constant (general case)

Thm. 3.2.7 in Lehmann Casella

Given a family that behaves naturally under group action. If δ is equivariant and L is invariant, then the risk satisfies

$$R(\bar{g} \star \theta, \delta) = R(\theta, \delta)$$

for all $\theta \in \Omega$.

Remark: to ensure that the risk is constant, we need the group action to be transitive over Ω, that is for any $\theta \in \Omega$, there exists $\bar{g} \in G$ such that $\theta = \bar{g} \star \theta_0$.

proof:

Example 1: scale group

$X_1, \ldots X_n$ i.i.d. $\mathcal{N}(0, \sigma^2)$. Consider the estimation of σ^2. The group that acts on parameters is $\bar{G} = (0, \infty)$ with multiplication. The group that acts on sample space is $G = (0, \infty)^n$. Show that

- The family behaves naturally under group action
- The loss $L(\sigma, d) = \rho(d/\sigma)$ is invariant
- We can define MRE for scale estimation

Example 2: binomial transformation group

Let $X \sim \text{Binomial}(n, p), 0 < p < 1$. The groups are

$$g \star X = n - X$$
$$\bar{g} \star p = 1 - p$$

Show that the risks for estimating p and for estimating $1 - p$ are equal. And the group is not transitive.

Think about someone who interchanges the definition of head and tail when collecting binomial trials.

Maximal invariant statistic

What will be the form of MRE in equivariant location estimation?

We are back to the location estimation problem.

The general case is skipped

Location invariant function

A function h on $\mathbb{R}^n$ is called invariant if $h(x + a) = h(x)$ for all $x \in \mathbb{R}^n, a \in \mathbb{R}$.

A location invariant statistic

$$Y(X) = \begin{pmatrix} X_1 - X_n \\ \vdots \\ X_{n-1} - X_n \end{pmatrix}$$

is location invariant

In fact, any location invariant statistic is a function of Y

Suppose $h(X)$ is an arbitrary invariant function, take $a = -X_n$, we have

$$h(X) = h(X - X_n \mathbf{1}) = h(Y_1, \ldots, Y_{n-1}, 0).$$

For the above reason, Y is called maximal invariant

Y carries at least as much information about X an any other invariant statistics $h(X)$!

Optimal MRE estimators are constructed by conditioning on Y

Thm. 10.4 in Keener and 3.1.10 in Lehmann Casella

Consider equivariant estimation of a location parameter with an invariant loss ρ. Suppose δ_0 is an equivariant estimator with finite risk. Suppose for a.e. $y \in \mathbb{R}^{n-1}$, there is a value $v^* = v^*(y)$ that minimizes

$$\mathbb{E}_0 \left[\rho(\delta_0(X) - v) \mid Y = y\right]$$

over $v \in \mathbb{R}$. Then an MRE estimator is given by

$$\delta_0(X) - v^*(Y).$$

Will prove it in the next lecture

Summary

- Formulating **equivariant estimation** under location family requires three parts:
  - family behaves naturally
  - invariant loss
  - restricting to equivariant estimators
- Formulating general equivariant estimation is similar
- **Maximal invariant** statistic is useful

What is next?

- How to construct MRE estimators?

Thank you