Statistics

Week 6: Lehmann-Scheffé and consistency

Etienne Wijler

Econometrics and Data Science
Econometrics and Operations Research
Bachelor Program

VRIJE UNIVERSITEIT AMSTERDAM
SCHOOL OF BUSINESS AND ECONOMICS

Course overview

P4: Estimation

- Week 1 Probability Recap
- Week 2 Statistical Models
- Week 3 Data Reduction and MME
- Week 4 MLE and Evaluation
- Week 5 Estimator Optimality
- Week 6 Optimality and Consistency

P5: Inference

- Week 7 Hypothesis testing
- Week 8 Mean and Variance testing
- Week 9 Finding test statistics
- Week 10 Evaluating tests
- Week 11 Interval estimation
- Week 12 Asymptotic tests

When to stop improving?

Previously, you learned to improve estimators via Rao-Blackwellization.

Question: When to stop improving?

Answer: When we attain the Cramér-Rao lower bound.

Problem: the Cramér-Rao lower bound is not always attainable (or possible to derive)!

Idea: somehow measure the amount of "noise" left in our estimator.

Unbiased estimators of zero

First, let W be an unbiased estimator of $\tau(\theta)$ and U be an estimator such that $\mathbb{E}_\theta(U) = 0$ for all $\theta \in \Theta$.

Then, we define $\phi_a = W + aU$. Clearly, $\mathbb{E}(\phi_a) = \mathbb{E}(W) = \tau(\theta)$.

Additionally,

$$\mathbb{V}\mathrm{ar}_\theta \phi_a = \mathbb{V}\mathrm{ar}_\theta W + a^2 \mathbb{V}\mathrm{ar}_\theta U + 2a\mathbb{C}\mathrm{ov}_\theta(W, U).$$

Implication: If $\mathbb{C}\mathrm{ov}_\theta(W, U) \neq 0$, then we can choose a such that $a^2 \mathbb{V}\mathrm{ar}_\theta U + 2a\mathbb{C}\mathrm{ov}_\theta(W, U) < 0$. This would imply $\mathbb{V}\mathrm{ar}_\theta \phi_a < \mathbb{V}\mathrm{ar}_\theta W$!

Intuition: when an estimator is correlated with noise (U), it can be improved by removing this noise.

A new criterion for UMVUEs

Theorem (7.3.20)

An unbiased estimator W of $\tau(\theta_0)$ is UMVU if and only if $\mathbb{C}ov_\theta(W, U) = 0$ for all estimators U that satisfy $\mathbb{E}_\theta U = 0$.

Proof.

Only the "if" direction remains. Suppose W has $\mathbb{C}\mathrm{ov}_\theta(W, U) = 0$ for all estimators U that satisfy $\mathbb{E}_\theta U = 0$ and let W' be another unbiased estimator. Then

$$\begin{aligned} \mathbb{V}\mathrm{ar}_\theta W' &= \mathbb{V}\mathrm{ar}_\theta(W + W' - W) = \mathbb{V}\mathrm{ar}_\theta W + \mathbb{V}\mathrm{ar}_\theta(W' - W) + 2\mathbb{C}\mathrm{ov}_\theta(W, W' - W) \\ &= \mathbb{V}\mathrm{ar}_\theta W + \mathbb{V}\mathrm{ar}_\theta(W' - W) \geq \mathbb{V}\mathrm{ar}_\theta W, \end{aligned}$$

since $\mathbb{E}_\theta(W' - W) = 0$. $\square$

Complete sufficient statistics

Problem: it is hard/impossible to find all unbiased estimators of zero.

Solution: extend the idea of a sufficient statistic to imply uncorrelatedness with unbiased estimators of zero.

Definition (6.2.21)

Suppose we have a statistical model $\{f(x \mid \theta) \mid \theta \in \Theta\}$ for the random vector $\boldsymbol{X}$ and let $T(\boldsymbol{X})$ be a statistic. Then T is called *complete* if for all functions g such that $\mathbb{E}_\theta g(T) = 0$ for all $\theta \in \Theta$ we have that $g(T) = 0$ almost surely.

Interpretation: A complete sufficient statistic contains no irrelevant information about θ.

Implication: Let $g(T) = \mathbb{E}(U \mid T)$ with $\mathbb{E}_\theta g(T) = \mathbb{E}_\theta U = 0$. Then, $\mathbb{E}(U \mid T) = 0$, implying U and T are uncorrelated!

Lehmann - Scheffé

Important: It also holds that any estimator that is a function of a complete sufficient statistic is uncorrelated with unbiased estimators of zero!

Theorem (7.2.23, Lehmann-Scheffé)

Suppose T is a complete and sufficient statistic for θ_0 and let ϕ be a function. Then $\phi(T)$ is the best unbiased estimator for its expectation.

Proof.

Let U satisfy $\mathbb{E}_\theta U = 0$, then

$$\begin{aligned}\mathbb{C}\mathrm{ov}_\theta(\phi(T), U) &= \mathbb{E}_\theta(\phi(T)U) - \mathbb{E}_\theta(\phi(T))\mathbb{E}_\theta(U) = \mathbb{E}_\theta(\phi(T)U) = \mathbb{E}_\theta(\mathbb{E}(\phi(T)U \mid T)) \\ &= \mathbb{E}_\theta(\phi(T)\mathbb{E}(U \mid T)) = \mathbb{E}_\theta(\phi(T) \cdot 0) = 0.\end{aligned}$$

It follows by the previous Theorem that $\phi(T)$ must be the UMVUE of $\mathbb{E}_\theta\phi(T)$. $\square$

Implication of Lehmann-Scheffé

Any statistic which still contains all the information in the data about θ_0, but does not contain any stochastic noise, automatically is the UMVUE for its expectation.

Lehmann-Scheffé and the Bernoulli distribution

Example

Suppose we have the statistical model $\{\text{Bernoulli}(p) \mid p \in (0,1)\}$. We will show that $W(\boldsymbol{X}) = \overline{X}$ is the UMVUE for p_0 via Lehmann-Scheffé. Let $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$. We already know that T is sufficient. Hence, we must show that $T(\boldsymbol{X})$ is complete, after which we can conclude that $W(X) = \frac{1}{n}T(X)$ is the UMVUE of $\mathbb{E}W(\boldsymbol{X}) = p_0$.

Note: The above example is kind of unique. It is usually very hard to show that a sufficient statistic is complete.

Except....

Completeness and the exponential family

Lemma (6.2.25)

Let $X_1, \ldots, X_n$ be a random sample from a population in $\{g(x \mid \theta) \mid \theta \in \Theta\}$, where $\theta = (\theta_1, \ldots, \theta_k)$. If we can rewrite this statistical model as an exponential family of order k:

$$g(x \mid \theta) = h(x)c(\theta)e^{\sum_{j=1}^{k} w_j(\theta)t_j(x)}.$$

Then $T(\boldsymbol{X}) = (\sum_{i=1}^{n} t_1(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i))$ is complete if $\Theta \subseteq \mathbb{R}^k$ contains an open subset.

Note: The open subset condition is to prevent scenarios in which $\theta_2 = h(\theta_1)$.

Workflow to show that an estimator is UMVUE

Checklist:

1. Is $W(\boldsymbol{X})$ unbiased?
2. Is $W(\boldsymbol{X})$ a function of a sufficient statistic?
3. Is $g(x \mid \theta)$ a member of the exponential family?
4. Does $\Theta \subseteq \mathbb{R}^k$ contain an open subset?

Conclusion: If yes to all questions, then $W(\boldsymbol{X})$ is the UMVUE of $\mathbb{E}W(\boldsymbol{X})$. No need to derive a Cramér-Rao lower bound!

Lehmann-Scheffé examples

Example (Bernoulli revisited)

Consider again the statistical model $\{\text{Bernoulli}(p) \mid p \in (0,1)\}$. Show that $W(\boldsymbol{X}) = \overline{X}$ is the UMVUE for p_0.

Example (Normal(μ,σ^2))

We study the statistical model $\{\text{Normal}(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$. We have previously shown that an unbiased estimator of $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0^2)$ is given by $\hat{\boldsymbol{\theta}} = (\bar{X}, S^2)$, with

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n} X_i\right)^2.$$

Show that $\hat{\boldsymbol{\theta}}$ is the UMVUE of $\boldsymbol{\theta}_0$.

A final, challenging example

Example

We have the statistical model $\{\text{Poisson}(\lambda) \mid \lambda > 0\}$ and are interested in estimating $\tau(\lambda_0) = P(X_1 = 0) = e^{-\lambda_0}$. There is no immediate intuitive UMVUE candidate available. Our strategy will be as follows:

1. Find an easy but sub-optimal unbiased estimator.
2. Improve the estimator through Rao-Blackwellization.
3. Use Lehman-Scheffé to conclude optimality of the resulting estimator.

The initial estimator that we will consider is given by $W(\boldsymbol{X}) = \mathbb{1}_{\{X_1=0\}}$. We will additionally use the well-known result that the sum of n iid Poisson(λ) random variables is distributed Poisson($n\lambda$).

Course overview

P4: Estimation

- Week 1 Probability Recap
- Week 2 Statistical Models
- Week 3 Data Reduction and MME
- Week 4 MLE and Evaluation
- Week 5 Estimator Optimality
- Week 6 Optimality and Consistency

P5: Inference

- Week 7 Hypothesis testing
- Week 8 Mean and Variance testing
- Week 9 Finding test statistics
- Week 10 Evaluating tests
- Week 11 Interval estimation
- Week 12 Asymptotic tests

Asymptotic analysis

Previously: we looked at finite sample properties of estimators (n is a finite number).

Asymptotics: now we will investigate the properties of our estimators when n go to infinity.

Motivation: there are several reasons for such an "asymptotic analysis", including

1. calculations will simplify,
2. distribution may be misspecified,
3. independence may be relaxed.

Intuition: as n increases, we obtain more information about the true unknown density $g(x \mid \theta_0)$, and in the limit $n \to \infty$ we have all information.

Motivating examples (i)

Example (Bernoulli)

Suppose we have the $\{\text{Bernoulli}(p) \mid p \in [0,1]\}$ model, then we already know that $\frac{1}{n}\sum_{i=1}^n X_i \to p_0$ by the LLN, so in the limit we know the full distribution, as $P(X_1 = 1) = p_0$ and $P(X_1 = 0) = 1 - p_0$.

Example (Discrete density approximation)

Suppose we have a discrete statistical model $\{g(x \mid \theta) \mid \theta \in \Theta\}$ of densities that take value on $\mathbb{Z}$. Define the function

$$g(k) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i = k\}}, \qquad \forall k \in \mathbb{Z}.$$

Then, again by the LLN, we have $g(k) \to \mathbb{E}\mathbb{1}_{\{X_1 = k\}} = P(X_1 = k) = g(k \mid \theta_0)$ and thus we know the full distribution $g(x \mid \theta_0)$ as $n \to \infty$.

Motivating examples (ii)

Example

Suppose we have a continuous statistical model $\{g(x \mid \theta) \mid \theta \in \Theta\}$. Define the function

$$G(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}, \qquad \forall x \in \mathbb{R}.$$

Then, again by the LLN, we have $G(x) \to \mathbb{E}\mathbb{1}_{\{X_1 \leq x\}} = P(X_1 \leq x) = G(x \mid \theta_0)$ and thus we know the full distribution $g(x \mid \theta_0)$ as $n \to \infty$.

Interpretation: in the limit as $n \to \infty$, we can estimate any density with 100% accuracy.

Implication: when developing an estimator for a parameter, we should require that its estimation error vanishes as $n \to \infty$.

Consistency

Definition (10.1.1)

A sequence of estimators $W_n(X_1, \dots, X_n)$ is called consistent if for all $\epsilon > 0$ and $\theta \in \Theta$ we have

$$\lim_{n \to \infty} P_\theta(|W_n - \theta| > \epsilon) = 0.$$

Note: consistency states that the sequence of estimators converges in probability to θ for every $\theta \in \Theta$.

Implication: a consistent sequence of estimators satisfies $W_n \to \theta_0$ in probability. In that case, it is common terminology to say that "W_n is a consistent estimator of θ_0".

Tools: we can apply a lot of useful tools and limit theorems from probability theory to establish the consistency of (a sequence of) estimators.

Consistency example: LLN and CMT

Example

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf

$$g(x \mid \theta) = \frac{1}{2\theta} e^{-|x|/\theta}, \quad \sigma > 0.$$

Derive the method of moments estimator of θ_0 based on the lowest moment possible. Show that this MME is consistent.

Important: This example highlights a crucial strategy for deriving consistency.

1. identify that the estimator is a continuous function of an average,
2. apply the Law of Large Numbers (LLN) to the average,
3. Apply the Continuous Mapping Theorem (CMT) to the estimator.

An alternative route to consistency

Problem: The definition of convergence in probability is quite abstract and we cannot always fall back on general limit theorems such as in the previous example.

Solution: there exist more intuitive, sufficient conditions for consistency.

Definition

A sequence of estimators $(W_n)_{n\in\mathbb{N}}$ is called asymptotically unbiased if $\lim_{n\to\infty} \mathbb{B}\text{ias}_\theta(W_n) = 0$ for all $\theta \in \Theta$.

Theorem (10.1.3)

Suppose a sequence of estimators is asymptotically unbiased and that $\lim_{n\to\infty} \mathbb{V}ar_\theta(W_n) = 0$ for all $\theta \in \Theta$, then the sequence is consistent.

Consistency via asymptotic bias and variance

Example

Suppose we have the $\{\text{Normal}(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$ model. Then the sequence of estimators $W_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is consistent for μ_0, since $\overline{X}_n \sim \text{N}(\mu_0, \sigma_0^2/n)$ and thus for all $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ we have

$$\lim_{n\to\infty} \mathbb{B}\text{ias}_\mu(W_n) = \lim_{n\to\infty} \mathbb{E}_\mu(W_n) - \mu = \lim_{n\to\infty} \mu - \mu = 0,$$

$$\lim_{n\to\infty} \mathbb{V}\text{ar}_{\sigma^2}(W_n) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$$

MLE and consistency

MLE: the next result is the first out of two strong results for maximum likelihood estimators.

Theorem (10.1.6)

Suppose that some regularity conditions hold (they always do for members of the exponential family) and let τ be a continuous function. Then the sequence of maximum likelihood estimators $\tau(\hat{\theta}_{ML})$ is consistent for $\tau(\theta_0)$.

Example

Consider again the $\{\text{Normal}(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$ model. Then, the MLE of $\boldsymbol{\theta}_0$ is given by $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}^2)$ with $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Since the Normal($\mu$,$\sigma^2$) is a member of the exponential family, it holds that $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$.