ST5215 Advanced Statistical Theory, Lecture 19

HUANG Dongming

National University of Singapore

22 Oct 2020

Overview

Last time

- Consistency
- Asymptotic bias, variance, mean squared error

Today

- Asymptotic properties of MOM, UMVUE, sample quantiles

Recap: Asymptotic Criteria

Consistency is related to a sequence of $\{T_n; n = n_0, n_0+1, \ldots\}$ but say "T_n is consistent" for short

1. *consistent*: $T_n(X) \xrightarrow{P} \theta$ w.r.t. any $P \in \mathcal{P}$.
2. *a_n-consistent*: Suppose $\{a_n\} \to \infty$ and $a_n\{T_n(X) - \theta\} = O_P(1)$ w.r.t. any $P \in \mathcal{P}$.
3. *strongly consistent*: $T_n(X) \xrightarrow{a.s.} \theta$ w.r.t. any $P \in \mathcal{P}$
4. *L_r-consistent*: $T_n(X) \xrightarrow{L^r} \theta$ w.r.t. any $P \in \mathcal{P}$ for some fixed $r > 0$.

Suppose $a_n \to \infty$ or $a_n \to a > 0$ and $a_n\{T_n(X) - \theta\} \xrightarrow{\mathcal{D}} Y$.

1. asymptotic bias: EY/a_n, if $E|Y| < \infty$
2. asymptotic variance is $\mathrm{Var}(Y)/a_n^2$ and asymptotic mse is EY^2/a_n^2, if $EY^2 < \infty$

The *asymptotic relative efficiency of T_n' w.r.t. T_n* is defined to be

$$e_{T_n', T_n}(P) = \mathrm{amse}_{T_n}(P) / \mathrm{amse}_{T_n'}(P)$$

Recap: Method of Moments (Lecture 9)

Suppose X_i's are i.i.d. from P_θ and $E_\theta |X_1|^k < \infty$

- Let $\mu_j = E_\theta X_1^j$ be the jth moment of P_θ and suppose $\mu_j = h_j(\theta)$ for some functions h_j on $\mathcal{R}^k$ ($j = 1, \ldots, k$)
- $\hat{\mu}_{n,j} = \frac{1}{n}\sum_{i=1}^n X_i^j$ is the jth sample moment
- Any $\hat{\theta}$ that solves

$$\hat{\mu}_{n,j} = h_j(\hat{\theta}), \quad j = 1, \ldots, k$$

is a moment estimator of θ

Properties of MOM Estimators

Let $\mu = (\mu_1, \ldots, \mu_k)$, $\hat{\mu}_n = (\hat{\mu}_{n,1}, \ldots, \hat{\mu}_{n,k})$, and $h = (h_1, \ldots, h_k)$. Then

$$\mu = h(\theta) \text{ and } \hat{\mu}_n = h(\hat{\theta})$$

- If h^{-1} exists, the unique moment estimator of θ is $\hat{\theta}_n = h^{-1}(\hat{\mu}_n)$
- Furthermore, if h^{-1} continuous, then by SLLN and continuous mapping, $\hat{\theta}_n$ is strongly consistent
- If $g = h^{-1}$ is differentiable and $E|X_1|^{2k} < \infty$, by CLT and δ-method, we have

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathcal{D}} N\left(0, [\nabla g]^\top V_\mu \nabla g\right) \tag{1}$$

  where V_μ is a $k \times k$ matrix whose (i,j) th element is $\mu_{i+j} - \mu_i \mu_j$
  - In this case, the MOM estimator is $\sqrt{n}$-consistent
  - If $k = 1$, $\text{amse}_{\hat{\theta}_n}(\theta) = g'(\mu_1)^2 \sigma^2 / n$ where $\sigma^2 = \mu_2 - \mu_1^2$

Example 3.24

Let $X_1, ..., X_n$ be i.i.d. from P_θ indexed by $\theta = (\mu, \sigma^2)$, where $\mu = EX_1 \in \mathcal{R}$ and $\sigma^2 = \mathrm{Var}(X_1) \in (0, \infty)$. This includes cases like normal distributions and double exponential distributions with Lebesgue p.d.f.

$$\frac{1}{\sqrt{2}\sigma} e^{-\sqrt{2}|x-\mu|/\sigma}, \qquad \mu \in \mathcal{R}, \sigma > 0$$

- Since $EX_1 = \mu$ and $EX_1^2 = \mathrm{Var}(X_1) + (EX_1)^2 = \sigma^2 + \mu^2$, setting $\hat{\mu}_1 = \mu$ and $\hat{\mu}_2 = \sigma^2 + \mu^2$ we obtain the moment estimator

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

  Note that $\bar{X}$ is unbiased, but $\frac{n-1}{n} S^2$ is not.
- If P_θ is normal, then $\hat{\theta}$ is sufficient and nearly the same as the UMVUE
- If P_θ is double exponential, then $\hat{\theta}$ is not sufficient and can often be improved

Consider now the estimation of σ^2 when we know that $\mu = 0$.

- Obviously we cannot use the equation $\hat{\mu}_1 = \mu$ to solve the problem.
- Using $\mu_2 = \sigma^2$, we obtain the moment estimator

$$\hat{\sigma}^2 = \hat{\mu}_2 = n^{-1} \sum_{i=1}^{n} X_i^2. \tag{3}$$

- If P_θ is normal, this is a good estimator
- If P_θ is double exponential, Eq (3) is not sufficient for σ.
  - We should first make a transformation $Y_i = |X_i|$ (note that $EY_i = \sigma/\sqrt{2}$),
  - then obtain the moment estimator based on the transformed data:

$$2\bar{Y}^2 = 2\left(\frac{1}{n}\sum_{i=1}^{n} |X_i|\right)^2, \tag{4}$$

which is sufficient for σ^2.

Example 3.25

Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on (θ_1, θ_2), $-\infty < \theta_1 < \theta_2 < \infty$.

- Note that $EX_1 = (\theta_1 + \theta_2)/2$ and $EX_1^2 = (\theta_1^2 + \theta_2^2 + \theta_1\theta_2)/3$.
- Setting $\hat{\mu}_1 = EX_1$ and $\hat{\mu}_2 = EX_1^2$ and substituting θ_1 in the second equation by $2\hat{\mu}_1 - \theta_2$ (the first equation), we obtain that

$$(2\hat{\mu}_1 - \theta_2)^2 + \theta_2^2 + (2\hat{\mu}_1 - \theta_2)\theta_2 = 3\hat{\mu}_2, \tag{5}$$

which is the same as

$$(\theta_2 - \hat{\mu}_1)^2 = 3(\hat{\mu}_2 - \hat{\mu}_1^2). \tag{6}$$

- Since $\theta_2 > EX_1$, we obtain that

$$\hat{\theta}_2 = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \bar{X} + \sqrt{\tfrac{3(n-1)}{n} S^2} \tag{7}$$

$$\hat{\theta}_1 = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \bar{X} - \sqrt{\tfrac{3(n-1)}{n} S^2}. \tag{8}$$

- $\hat{\theta}_i$'s are not functions of $(X_{(1)}, X_{(n)})$ (complete and sufficient)

Asymptotic Properties of UMVUE

- UMVUE's are typically consistent (see Exercise 2 in the Tutorial)
- UMVUE's are exactly unbiased so that there is no need to discuss their asymptotic biases
- The amse can be used to assess the performance of a UMVUE if the exact form of mse are difficult to obtain
- In many cases, although the variance does not attain the Cramér-Rao lower bound, the ratio of the mse over the Cramér-Rao lower bound converges to 1

Example: Normal Families in Lecture 14

Model

$X_1, ..., X_n$ be i.i.d. from the $N(\mu, \sigma^2)$ distribution with an unknown $\mu \in \mathcal{R}$ and a known σ^2. $I(\mu) = n/\sigma^2$. We are interested in estimating μ^2

- Previously, we show that $h(\bar{X}) = \bar{X}^2 - \sigma^2/n$ is the UMVUE of μ^2 and the variance is
$$\mathrm{Var}(h(\bar{X})) = \frac{4\mu^2\sigma^2}{n} + \frac{2\sigma^4}{n^2}$$
while the Cramér-Rao lower bound is $4\mu^2\sigma^2/n$.
- If $\mu = 0$, the C-R lower bound $=0$ and is not informative
- If $\mu \neq 0$, the ratio of the two is
$$1 + \frac{2\sigma^4}{n^2} \times \frac{n}{4\mu^2\sigma^2} \to 1$$

Example: Poisson Families

Model

$X_1, ..., X_n$ be i.i.d. from the Poisson(λ) distribution with an unknown $\lambda > 0$. We are interested in estimating $\tau = P(X_1 = 0) = e^{-\lambda}$.

- We have showed that $T_n = \sum_{i=1}^n X_i$ is minimul sufficient and complete. Since $S_0 = I_{\{X_1=0\}}$ is unbiased, so by Lehmann-Scheffé Theorem
$$S_n = E(S_0 \mid T_n) = \left(1 - \frac{1}{n}\right)^{T_n}$$
is a UMVUE (Page 22 in Lecture 10)
- Note that $\log(S_n) = T_n/n\,[n\log(1-1/n)] = T_n/n\,[-1 + O(1/n)] = -T_n/n + O_p(1/n)$ since $T_n/n = O_p(1)$ by SLLN
- By CLT, $\sqrt{n}\,(T_n/n - \lambda) \xrightarrow{\mathcal{D}} N(0, \lambda)$
- By Slutsky's theorem and $O_p(1/\sqrt{n}) = o_p(1)$, we have $\sqrt{n}\,(\log S_n - (-\lambda)) \xrightarrow{\mathcal{D}} N(0, \lambda)$
- By δ-method, $\sqrt{n}\left(S_n - e^{-\lambda}\right) \xrightarrow{\mathcal{D}} N(0, \lambda e^{-2\lambda})$. So $amse_{S_n}(\lambda) = \lambda e^{-2\lambda}/n$=C-R lower bound

Asymptotic Property of Sample Quantiles

For any $\gamma \in (0,1)$, the $\lfloor \gamma n \rfloor$th order statistic is also called the γ-sample quantile. ($\lfloor a \rfloor$ denotes the largest integer that is no greater than a)

Theorem

1. *Let $X_1, X_2, \ldots$ be i.i.d. r.v.s with CDF F,*
2. *let $\gamma \in (0,1)$, and*
3. *let $\tilde{\theta}_n$ be the $\lfloor \gamma n \rfloor$th order statistic for $X_1, \ldots, X_n$*

Suppose $F(\theta) = \gamma$ and $F'(\theta)$ exists and is positive, then

$$\sqrt{n}\left(\tilde{\theta}_n - \theta\right) \Rightarrow N\left(0, \frac{\gamma(1-\gamma)}{[F'(\theta)]^2}\right)$$

- This shows that if the CDF is smooth and strictly increasing around θ, then the $F(\theta)$-sample quantile is asymptotically unbiased, and its amse is $\gamma(1-\gamma)/(nF'(\theta)^2)$

Proof

- For any fixed $a \in \mathcal{R}$, define

$$S_n(a) = \#\{i \leq n : X_i \leq a\}$$

- Then $\tilde{\theta}_n \leq a \Leftrightarrow S_n(a) \geq \lfloor \gamma n \rfloor$
- Therefore for any fixed $t \in \mathcal{R}$

$$P\left(\sqrt{n}(\tilde{\theta}_n - \theta) \leq t\right) = P\left(S_n(\theta + t/\sqrt{n}) \geq \lfloor \gamma n \rfloor\right)$$

- $S_n(a) \sim$Binom$(n, F(a))$, so it is tempting to use CLT. However, the probability parameter $F(a)$ changes along with n and CLT is not precise enough
- We need a uniform control on the convergence of CDF: *Berry-Esseen Theorem*

Theorem (Berry–Esseen Theorem)

There exists a positive universal constant C such that the following holds. Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random variables with $E\left(Y_1\right) = 0$, $E\left(Y_1^2\right) = \sigma^2 > 0$, and $E\left(\left|Y_1\right|^3\right) = \rho < \infty$. Let F_n be the CDF of

$$\frac{\sum_{i=1}^n Y_i}{\sigma\sqrt{n}}$$

and Φ be the CDF of the standard normal distribution, then

$$\sup_{y \in \mathcal{R}} \left|F_n(y) - \Phi(y)\right| \leq \frac{C\rho}{\sigma^3\sqrt{n}}$$

Proof (Cont.)

- Consider $Y_{n,i} = I_{X_i \le \theta + t/\sqrt{n}} - F(\theta + t/\sqrt{n})$
- Then $EY_{n,i} = 0$.
- For all n large enough, $EY_{n,i}^2 = F(\theta + t/\sqrt{n})(1 - F(\theta + t/\sqrt{n})) \ge \sigma_0^2 > 0$, $EY_{n,i}^3 \le 1$.
- By Berry–Esseen Theorem,

$$\begin{aligned} &\Big| P\left(S_n(\theta + t/\sqrt{n}) \ge \lfloor \gamma n \rfloor\right) \\ &\qquad - \Phi\left(- \frac{\lfloor \gamma n \rfloor - nF(\theta + t/\sqrt{n})}{\sqrt{nF(\theta + t/\sqrt{n})(1 - F(\theta + t/\sqrt{n}))}} \right) \Big| \\ &\le \frac{C}{\sigma_0^3 \sqrt{n}} \end{aligned}$$

- Note that

$$
\begin{aligned}
&\lim_n \frac{\lfloor \gamma n \rfloor - nF(\theta + t/\sqrt{n})}{\sqrt{nF(\theta + t/\sqrt{n})(1 - F(\theta + t/\sqrt{n}))}} \\
=&\frac{1}{\sqrt{F(\theta)(1-F(\theta)}} \lim_n \sqrt{n} \left(\lfloor \gamma n \rfloor / n - \gamma + \gamma - F(\theta + t/\sqrt{n}) \right) \\
=&\frac{1}{\sqrt{\gamma(1-\gamma)}} \left(-F'(\theta) t \right)
\end{aligned}
$$

- Since Φ is continuous, we conclude that for any $t \in \mathcal{R}$

$$
\lim_n \left| P\left(\sqrt{n}(\tilde{\theta}_n - \theta) \leq t \right) - \Phi(\frac{1}{\sqrt{\gamma(1-\gamma)}} F'(\theta) t) \right| = 0
$$

Tutorial

1 Let $X_1, X_2, \ldots$ be random variables. Show that $\{|X_n|\}$ is uniformly integrable if one of the following condition holds:

  (i) $\sup_n E\left|X_n\right|^{1+\delta} < \infty$ for a $\delta > 0$
  (ii) $P\left(|X_n| \geq c\right) \leq P(|X| \geq c)$ for all n and $c > 0$, where X is an integrable random variable.

2 Let $X_1, \ldots, X_n, \ldots$ be i.i.d. observations. Suppose that $T_n = T(X_{1:n})$ is an unbiased estimator of ϑ based on $X_1, \ldots, X_n$ such that for any n, $\mathrm{Var}\left(T_n\right) < \infty$, $\mathrm{Var}\left(T_n\right) \leq \mathrm{Var}\left(U_n\right)$ for any other unbiased estimator U_n of ϑ based on $X_1, \ldots, X_n$. Then T_n is consistent in mse.

3 Let $(X_1, \ldots, X_n)$ be a random sample of random variables from a population P with $EX_1^2 < \infty$ and $\bar{X}$ be the sample mean. Consider the estimation of $\mu = EX_1$.

(i) Let $T_n = \bar{X} + \xi_n$, where ξ_n is a random variable satisfying $\xi_n = 0$ with probability $1 - n^{-1}$ and $\xi_n = n$ with probability n^{-1}. Show that the bias of T_n is not the same as the asymptotic bias of T_n for any P.

(ii) Let $T_n = \bar{X} + \eta_n$, where η_n is a random variable that is independent of $X_1, \ldots, X_n$ and equals 0 with probability $1 - 2n^{-1}$ and ± 1 with probability n^{-1}. Show that the asymptotic mean squared error of T_n, the asymptotic mean squared error of $\bar{X}$, and the mean squared error of $\bar{X}$ are the same, but the mean squared error of T_n is larger than the mean squared error of $\bar{X}$ for any P.

Exercise 1

Let $X_1, X_2, \ldots$ be random variables. Show that $\{|X_n|\}$ is uniformly integrable if one of the following condition holds:

(i) $\sup_n E\,|X_n|^{1+\delta} < \infty$ for a $\delta > 0$

(ii) $P\,(|X_n| \geq c) \leq P(|X| \geq c)$ for all n and $c > 0$, where X is an integrable random variable.

Proof: Part (i)

- Since $I_{\{|X_n|>t\}} \leq t^{-\delta}|X_n|^{\delta}$, we have

$$E\left(|X_n|\, I_{\{|X_n|>t\}}\right) \leq E\,|X_n|^{1+\delta}\, t^{-\delta}$$

- Hence

$$\lim_{t\to\infty} \sup_n E\left(|X_n|\, I_{\{|X_n|>t\}}\right) \leq \sup_n E\,|X_n|^{1+\delta} \lim_{t\to\infty} t^{-\delta} = 0$$

Part (ii): Assume $P(|X_n| \geq c) \leq P(|X| \geq c)$ for all c

Use the identity that $E|Y| = \int_0^\infty P(|Y| > s)\, \mathrm{d}s$

$$
\begin{aligned}
\sup_n E\left(|X_n|\, I_{\{|X_n|>t\}}\right) &= \sup_n \int_0^\infty P\left(|X_n|\, I_{\{|X_n|>t\}} > s\right) ds \\
&= \sup_n \int_0^\infty P\left(|X_n| > s, |X_n| > t\right) ds \\
&= \sup_n \left(\int_0^t + \int_t^\infty\right) P\left(|X_n| > s, |X_n| > t\right) ds \\
&= \sup_n \left(tP\left(|X_n| > t\right) + \int_t^\infty P\left(|X_n| > s\right) ds\right) \\
&\leq tP(|X| > t) + \int_t^\infty P(|X| > s) ds \\
&\leq tE\left(\frac{|X|}{t} I_{\{|X|>t\}}\right) + \int_t^\infty P(|X| > s) ds \\
&\to 0
\end{aligned}
$$

as $t \to \infty$ when $E|X| < \infty$

Exercise 2

Let $X_1, \ldots, X_n, \ldots$ be i.i.d. observations. Suppose that $T_n = T(X_{1:n})$ is an unbiased estimator of ϑ based on $X_1, \ldots, X_n$ such that for any n, $\text{Var}(T_n) < \infty$, $\text{Var}(T_n) \leq \text{Var}(U_n)$ for any other unbiased estimator U_n of ϑ based on $X_1, \ldots, X_n$. Then T_n is consistent in mse.

Proof:

- Note that if $n > m$, then $\mathrm{Var}(T_n) \leq \mathrm{Var}(T_m)$ because T_m also depends on $X_1, \ldots, X_n$ and is unbiased
- For $2n$, consider an estimator $U_{2n} = \frac{T(X_{1:n}) + T(X_{(n+1):(2n)})}{2}$. It is unbiased and has variance

$$\mathrm{Var}(U_{2n}) = \frac{1}{2}\mathrm{Var}(T_n)$$

  since X_i's are i.i.d.
- So $\mathrm{Var}(T_{2n}) \leq 2^{-1}\mathrm{Var}(T_n)$
- Recursively, $\mathrm{Var}(T_{n2^k}) \leq 2^{-k}\mathrm{Var}(T_n) \to 0$ as $k \to \infty$
- Since $\mathrm{Var}(T_n)$ is non-increasing in n, we conclude that $\lim_n \mathrm{Var}(T_n) = 0$

Exercise 3: Part (i)

Let $(X_1, \ldots, X_n)$ be a random sample of random variables from a population P with $EX_1^2 < \infty$ and $\bar{X}$ be the sample mean. Consider the estimation of $\mu = EX_1$.

(i) Let $T_n = \bar{X} + \xi_n$, where ξ_n is a random variable satisfying $\xi_n = 0$ with probability $1 - n^{-1}$ and $\xi_n = n$ with probability n^{-1}. Show that the bias of T_n is not the same as the asymptotic bias of T_n for any P.

Proof:

- Since $E(\xi_n) = 1, E(T_n) = E(\bar{X}) + E(\xi_n) = \mu + 1$. This means that the bias of T_n is 1.
- Since $\xi_n \to_p 0$ and $\bar{X} \to_p \mu$ $T_n \to_p \mu$. Thus, the asymptotic bias of T_n is 0.

Exercise 3: Part (ii)

Let $(X_1, \ldots, X_n)$ be a random sample of random variables from a population P with $EX_1^2 < \infty$ and $\bar{X}$ be the sample mean. Consider the estimation of $\mu = EX_1$.

(ii) Let $T_n = \bar{X} + \eta_n$, where η_n is a random variable that is independent of $X_1, \ldots, X_n$ and equals 0 with probability $1 - 2n^{-1}$ and ± 1 with probability n^{-1}. Show that the asymptotic mean squared error of T_n, the asymptotic mean squared error of $\bar{X}$, and the mean squared error of $\bar{X}$ are the same, but the mean squared error of T_n is larger than the mean squared error of $\bar{X}$ for any P.

Proof:

- $\sqrt{n}(\bar{X} - \mu) \to_d N\left(0, \sigma^2\right)$, where $\sigma^2 = \operatorname{Var}(X_1)$.
- Since $\sqrt{n}\ \ \eta_n \to_p 0$, by Slutsky's theorem, $\sqrt{n}\left(T_n - \mu\right) = \sqrt{n}(\bar{X} - \mu) + \sqrt{n}\ \ \eta_n \to_d N\left(0, \sigma^2\right)$.
- Hence, the amse of T_n is the same as that of $\bar{X}$ and is equal to σ^2/n, which is the mse of $\bar{X}$.
- Since $E\left(\eta_n\right) = 0$, $E\left(T_n\right) = E(\bar{X}) = \mu$ and the mse of T_n is

$$\operatorname{Var}\left(T_n\right) = \operatorname{Var}(\bar{X}) + \operatorname{Var}\left(\eta_n\right) = \operatorname{Var}(\bar{X}) + \frac{2}{n} > \text{ mse of } \bar{X}$$