ST5215 Advanced Statistical Theory, Lecture 20

HUANG Dongming

National University of Singapore

27 Oct 2020

Overview

Last time

- Asymptotic properties of MOM, UMVUE, sample quantiles

Today

- A Uniform Strong Law of Large Numbers
- Consistency of MLEs, M-estimators
- Kullback-Leibler information, Shannon-Kolmogorov inequality

Recap: Asymptotic Properties of MOM, UMVU, and Quantile Estimators

1. Suppose the vector of first k moments is $\mu = (\mu_1, \ldots, \mu_k) = h(\theta)$ and the sample version is $\hat{\mu}_n = h(\hat{\theta}_n)$. If h^{-1} exists, the MOM estimator of θ is $\hat{\theta}_n = h^{-1}(\hat{\mu}_n)$
   - If h^{-1} continuous, then $\hat{\theta}_n$ is strongly consistent
   - If h^{-1} is differentiable and $E|X_1|^{2k} < \infty$, then we can apply CLT to $\hat{\mu}_n$ and use δ-method to obtain the asymptotic distribution of $\hat{\theta}_n$
2. A UMVUE with finite variance is consistent. In many cases, the ratio of its mse over the Cramér-Rao lower bound converges to 1
3. Suppose θ is the γ-quantile and $\tilde{\theta}_n$ is the sample γ-quantile, if $F'(\theta)$ exists and > 0, then

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

Wald's Consistency of MLE: Basic Idea

Let $X_1, \ldots, X_n$ be i.i.d. with density $f(x \mid \theta)$ w.r.t. a σ-finite measure ν, where $\theta \in \Theta$. θ_0 is the true value.

- The MLE is the value of θ that maximizes

$$\begin{aligned} \ell_n(\theta) - \ell_n(\theta_0) &= \log L_n(\theta) - \log L_n(\theta_0) \\ &= \sum_{j=1}^{n} \left(\log f(X_j \mid \theta) - \log f(X_j \mid \theta_0)\right) \end{aligned}$$

- From the Strong Law of Large Numbers,

$$\frac{1}{n} \log \frac{L_n(\theta)}{L_n(\theta_0)} = \frac{1}{n} \sum_{1}^{n} \log \frac{f(X_j \mid \theta)}{f(X_j \mid \theta_0)} \xrightarrow{\text{a.s.}} E_{\theta_0} \log \frac{f(X \mid \theta)}{f(X \mid \theta_0)}$$

- Suppose the model is identifiable
- If $\theta = \theta_0$, RHS$= 0$.
- If $\theta \neq \theta_0$, then RHS < 0.
- Moreover, if the model is *well-separated*, i.e., for any $\|\theta - \theta_0\| \geq \rho$, the RHS$< -\delta_\rho$, then eventually, $\frac{L_n(\theta)}{L_n(\theta_0)} < \exp(-n\delta_\rho)$, and the MLE must be lied in $B(\theta_0, \rho)$ (the ball center at θ_0 with radius ρ)

- A defect in the last argument: The SLLN only applies to a fixed θ, from which we cannot conclude that $\frac{L_n(\theta)}{L_n(\theta_0)}$ is uniformly small on $\|\theta - \theta_0\| \geq \rho$
- To make it precise, we need a Uniform SLLN

Theorem (USLLN, C)

Let $X_1, \ldots, X_n, \ldots,$ be i.i.d. sample from P and $U(x, \theta)$ be measurable on $\mathcal{X} \times \Theta$. Suppose

1. *$U(x, \theta)$ is a continuous in θ for any fixed x and for each θ, $\mu(\theta) = EU(X, \theta)$ is finite,*
2. *Θ is compact,*
3. *there exists a function $M(x)$ such that $EM(X) < \infty$ and $U(x, \theta) \leq M(x)$ for all x and θ.*

Then

$$P\left\{ \lim_{n\to\infty} \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{1}^{n} U\left(X_j, \theta\right) - \mu(\theta) \right| = 0 \right\} = 1$$

Proof

- By the continuity of $U(x, \cdot)$, its boundedness by $M(x)$, and DCT, $\mu(\theta)$ is continuous
- W.L.O.G., assume $\mu(\theta) \equiv 0$; otherwise consider $\tilde{U}(x, \theta) = U(x, \theta) - \mu(\theta)$
- Let
$$\varphi(x, \theta, \rho) = \sup_{\|\theta' - \theta\| < \rho} U\left(x, \theta'\right)$$
- φ is measurable, bounded by M and $\varphi(x, \theta, \rho) \to U(x, \theta)$ as $\rho \to 0$
- By DCT, $E\varphi(X, \theta, \rho) \to EU(X, \theta) = \mu(\theta) = 0$, as $\rho \to 0$
- Fixed $\varepsilon > 0$. For each θ, find ρ_θ so that $E\varphi\left(x, \theta, \rho_\theta\right) < \varepsilon$.
- Note that the collection of $B\left(\theta, \rho_\theta\right) = \{\theta' : |\theta - \theta'| < \rho_\theta\}$ for all θ covers Θ.
- By the compactness of Θ, there exists a finite sub-cover, say, $\Theta \subset \bigcup_{j=1}^m B\left(\theta_j, \rho_{\theta_j}\right)$

- For each $\theta \in \Theta$, there exists an index j such that $\theta \in B\left(\theta_j, \rho_{\theta_j}\right)$ and $U(x,\theta) \leq \varphi\left(x, \theta_j, \rho_{\theta_j}\right)$ for all x
- So

$$\sup_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} U\left(X_i,\theta\right) \leq \max_{1\leq j\leq m} \frac{1}{n}\sum_{i=1}^{n} \varphi\left(X_i,\theta_j,\rho_{\theta_j}\right)$$

- Apply SLLN to $\sum_{i=1}^{n} \varphi\left(X_i, \theta_j, \rho_{\theta_j}\right)$ for each $j = 1, \ldots, m$, and use the continuity of max, we have

$$\lim_n \max_{1\leq j\leq m} \frac{1}{n}\sum_{i=1}^{n} \varphi\left(X_i,\theta_j,\rho_{\theta_j}\right) = \max_{1\leq j\leq m} E\varphi\left(X,\theta_j,\rho_{\theta_j}\right) < \epsilon, \quad \text{a.s.}$$

- Therefore $\lim_n \sup_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} U\left(X_i,\theta\right) < \epsilon$, a.s. and thus,

$$\lim_n \sup_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} U\left(X_i,\theta\right) \leq 0, \quad \text{a.s.}$$

- Apply the same argument to $-U(x,\theta)$, we can conclude

$$\lim_n \sup_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} -U\left(X_i,\theta\right) \leq 0, \quad \text{a.s.}$$

- Note that $0 \leq \sup_\theta |g(\theta)| = \max\left\{\sup_\theta g(\theta), \sup_\theta\left[-g(\theta)\right]\right\}$

Kullback-Leibler Information

- In basic idea, the limit of the log likelihood ratio is $E_{\theta_0} \log \frac{f(X|\theta)}{f(X|\theta_0)}$
- Define the *Kullback-Leibler information number* as

$$K(f_0, f_1) = E_0 \log \frac{f_0(X)}{f_1(X)} = \int \log \frac{f_0(x)}{f_1(x)} f_0(x) d\nu(x)$$

- When f is indexed by θ, we may also write $K(\theta_0, \theta_1)$ for short.

Theorem (Shannon-Kolmogorov Information Inequality)

$K(f_0, f_1) \geq 0$ *with equality if and only if* $f_1(\omega) = f_0(\omega)$ ν*-a.e.*

Proof

- By Jensen's inequality with $-log(x)$, which is strictly convex, $K(f_0, f_1) = -E_0 \log \frac{f_1(X)}{f_0(X)} \geq -\log E_0 \frac{f_1(X)}{f_0(X)}$
- $E_0 \frac{f_1(X)}{f_0(X)} = \int_{f_0(x)>0} f_1(x) d\nu(x) \leq 1$
- The first equality holds if $f_0(X)/f_1(X)$ is a constant P_0-a.s.
- The second equality holds if $P_{f_1}(f_0(X) > 0) = 1$

Consistency of MLEs

Theorem (Continuous in θ)

Let $X_1, X_2, \ldots$ be i.i.d. from P, where P is in the family with density $f(x \mid \theta)$, $\theta \in \Theta$. Let θ_0 denote the true value of θ.
Suppose

1. Θ *is compact,*
2. $f(x \mid \theta)$ *is continuous in θ for all x,*
3. *there exists a function $M(x)$ such that $E_{\theta_0}|M(X)| < \infty$ and*

$$\log f(x \mid \theta) - \log f(x \mid \theta_0) \leq M(x), \qquad \text{for all } x \text{ and } \theta$$

4. *(identifiability) $f(x \mid \theta) = f(x \mid \theta_0)$ ν-a.e. $\Rightarrow \theta = \theta_0$.*

Then, for any sequence of maximum-likelihood estimates $\hat{\theta}_n$ of θ,

$$\hat{\theta}_n \xrightarrow{a.s.} \theta_0$$

Proof

- Let $U(x, \theta) = \log f(x \mid \theta) - \log f(x \mid \theta_0)$, and $\mu(\theta) = EU(X, \theta) = -K(\theta_0, \theta) \leq 0$
- Since $f(x \mid \theta)$ is continuous in θ for all x, so is $U(x, \theta)$.
- Following the proof of the Uniform SLLN, $\mu(\theta)$ is continuous in θ.
- By the identifiability and the Shannon-Kolmogorov Information Inequality, $\mu(\theta) < 0$ if $\theta \neq \theta_0$

- Fixed $\rho > 0$ and $W = \{\theta : \|\theta - \theta_0\| \geq \rho\}$. Then W is compact.
- Apply the Uniform SLLN, we have

$$P\left\{\lim_{n\to\infty} \sup_{\theta\in W} \left| \frac{1}{n} \sum_1^n U(X_j, \theta) - \mu(\theta) \right| = 0 \right\} = 1 \qquad (1)$$

- Let $\sup_{\theta\in W} \mu(\theta) = -\delta$. Since W is compact and μ is continuous, it achieves its maximal value in W and thus $\delta > 0$
- Denote by Ω_0 the event in Equation (1). When Ω_0 happens, we can choose N large enough such that for any $n \geq N$

$$\sup_{\theta\in W} \frac{1}{n} \sum_1^n U(X_j, \theta) < \sup_{\theta\in W} \mu(\theta) + \delta/2 = -\delta/2$$

- Note that $\frac{1}{n}\sum_1^n U\left(X_j, \hat{\theta}_n\right) \geq \frac{1}{n}\sum_1^n U(X_j, \theta_0) = 0$. So for $n \geq N$, $\hat{\theta}_n \notin W$, that is, $\|\hat{\theta}_n - \theta_0\| < \rho$.
- Since ρ is arbitrary, we have $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$

Remark 1 On the Continuity

- The assumption that $f(x \mid \theta)$ is continuous in θ for all x is too restricted. Actually, we only need to assume that $f(x \mid \theta)$ is *upper semicontinuous* (u.s.c.) in θ for all x, which means

$$\lim_{\rho \to 0} \left\{ \sup_{\|\theta' - \theta\| < \rho} f(x \mid \theta') \right\} = f(x \mid \theta)$$

Theorem (USLLN, u.s.c.)

Let $X_1, \ldots, X_n, \ldots,$ be i.i.d. sample from P and $U(x, \theta)$ be measurable on $\mathcal{X} \times \Theta$. Suppose

1. *Θ is compact,*
2. *$U(x, \theta)$ is upper semicontinuous in θ for all x,*
3. *$U(x, \theta) \leq M(x)$ for all x and θ, and $EM(X) < \infty$,*
4. *for all θ and for all sufficiently small $\rho > 0$, $\sup_{|\theta' - \theta| < \rho} U(x, \theta')$ is measurable in x*

Then $\limsup_{n \to \infty} \sup_{\theta \in \Theta} \frac{1}{n} \sum_1^n U(X_j, \theta) \leq \sup_{\theta \in \Theta} \mu(\theta)$ *a.s.*

Consistency of MLEs

Theorem (u.s.c. in θ)

Let $X_1, X_2, \ldots$ be i.i.d. from P, where P is in the family with density $f(x \mid \theta)$, $\theta \in \Theta$. Let θ_0 denote the true value of θ.
Suppose

1. *Θ is compact,*
2. *$f(x \mid \theta)$ is upper semicontinuous in θ for all x,*
3. *there exists a function $M(x)$ such that $E_{\theta_0}|M(X)| < \infty$ and*
$$\log f(x \mid \theta) - \log f(x \mid \theta_0) \leq M(x), \qquad \textit{for all } x \textit{ and } \theta$$
4. *for all $\theta \in \Theta$ and sufficiently small $\rho > 0, \sup_{|\theta' - \theta| < \rho} f(x \mid \theta')$ is measurable in x*
5. *(identifiability) $f(x \mid \theta) = f(x \mid \theta_0)$ ν-a.e. $\Rightarrow \theta = \theta_0$.*

Then any MLE of θ is strongly consistent

Remark 2 On the Compactness and Metric

When Θ is not compact, we have two remedies

1. Find a compact set and show that the estimators eventually lies in this compact set
2. Replace the space by its suitable compactification
   - In the theorem and proof, we can replace the norm of difference in θ, i.e., $\|\theta - \theta_0\|$ by any metric $d(\theta, \theta_0)$.
   - In this way, the theorems can be extended to a general metric space Θ (in contrary to merely a subset of Euclidean space), and hopefully Θ will be compact in the new topology induced by the metric
   - Accordingly, the strong consistency is defined as $d(\hat{\theta}_n, \theta_0) \xrightarrow{a.s.} 0$

Consistency of MLEs in General Metric Space

Theorem (u.s.c. in θ)

Let $X_1, X_2, \ldots$ be i.i.d. from P, where P is in the family with density $f(x \mid \theta)$, $\theta \in \Theta$. Let θ_0 denote the true value of θ.
Suppose

1. *Θ is a compact space with metric $d(\cdot, \cdot)$*
2. *$f(x \mid \theta)$ is upper semicontinuous in θ for all x,*
3. *there exists a function $M(x)$ such that $E_{\theta_0}|M(X)| < \infty$ and*
$$\log f(x \mid \theta) - \log f\left(x \mid \theta_0\right) \leq M(x), \qquad \textit{for all } x \textit{ and } \theta$$
4. *for all $\theta \in \Theta$ and sufficiently small $\rho > 0, \sup_{d(\theta', \theta) < \rho} f\left(x \mid \theta'\right)$ is measurable in x*
5. *(identifiability) $f(x \mid \theta) = f\left(x \mid \theta_0\right)$ ν-a.e. $\Rightarrow \theta = \theta_0$.*

Then any MLE $\hat{\theta}_n$ of θ is strongly consistent, $d(\hat{\theta}_n, \theta_0) \stackrel{a.s.}{\to} 0$

Example: Cauchy Distributions

Suppose X_i's are i.i.d. from the Cauchy distribution with location parameter θ, whose p.d.f. is

$$f_\theta(x) = \pi^{-1}\left(1 + (x-\theta)^2\right)^{-1}$$

- If $|\theta_0| \leq A$ is known for some $A > 0$
- Note that $f_\theta(x) \leq 1/\pi$ for all θ and x and $\int \frac{\log(1+x^2)}{1+x^2}\,\mathrm{d}x < \infty$, condition (3) holds
- Then the MLE restricted on $\Theta_A = [-A, A]$ is strongly consistent by the theorem

Example: Cauchy Distributions (Cont.)

Suppose X_i's are i.i.d. from the Cauchy distribution with location parameter θ, whose p.d.f. is

$$f_\theta(x) = \pi^{-1}\left(1 + (x-\theta)^2\right)^{-1}$$

Suppose the parameter space is $\mathcal{R}$. Since it is not compact, we can not use the theorem directly

- Let's consider an extended space $\bar{\Theta} = \mathcal{R} \cup \{\pm\infty\}$ equipped with metric $d(\theta_1, \theta_2) = |\arg\tan(\theta_1) - \arg\tan(\theta_2)|$ (note that $\arg\tan(\pm\infty) = \pm\pi/2$)
- Further define $U(x, -\infty) = -\infty$ and $U(x, \infty) = -\infty$ for all x
- Now the conditions in the theorem are satisfied, and we can conclude that $d(\hat{\theta}_n, \theta_0) \stackrel{a.s.}{\to} 0$
- Since $\theta_0 \neq \pm\infty$ and $\arg\tan$ is 1-1 and continuous, we can conclude that $\hat{\theta}_n \stackrel{a.s.}{\to} \theta_0$

Other Remarks

The theorems here rely on the continuity of p.d.f. but no differentiation is used

- In Theorem 4.17 in the textbook, conditions about the first and second order derivatives of the log likelihood are assumed, and the MLE is showed to be weakly consistent

The dominance of the log likelihood is crucial

- There exists an example where all other conditions are met and the MLE is not consistent
- In some cases, we can look for $M(x_1, \ldots, x_k)$ such that

$$\sup_{\theta \in \Theta} \sum_{i=1}^{k} \log \frac{f(x_i \mid \theta)}{f(x_i \mid \theta_0)} \leq M(x_1, \ldots, x_k)$$

  and $\mathbb{E}\left[M(X_1, \ldots, X_k)\right] < \infty$.
  Then we can divide the n observations into groups of k observations, and apply the uniform SLLN to the approximately n/k group sums

M-Estimators

A general method for finding an estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ is to maximize a criterion function of the type

$$\theta \mapsto S_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} s_\theta(X_i)$$

Here $s_\theta : \mathcal{X} \mapsto \overline{\mathbb{R}}$ are known functions.

- An estimator maximizing $S_n(\theta)$ over Θ is called an *M-estimator*
- Note that an MLE is is an M-estimator with $s_\theta(x) = \log f(x \mid \theta)$
- In the theorems of consistency of MLEs, we use $\log f(x \mid \theta)$ to obtain "$\mu(\theta) < 0$ for all $\theta \neq \theta_0$". This can be replaced by "$E_{\theta_0} s_\theta(X) < E_{\theta_0} s_{\theta_0}(X)$ for all $\theta \neq \theta_0$" for general M-estimators

Consistency of M-estimators

Theorem

Let S_n be random functions of θ and let S be a fixed function of θ such that

$$\sup_{\theta \in \Theta} |S_n(\theta) - S(\theta)| \xrightarrow{\mathcal{P}} 0, \tag{2}$$

and for every $\varepsilon > 0$,

$$\sup_{\theta : d(\theta, \theta_0) \geq \varepsilon} S(\theta) < S(\theta_0) \tag{3}$$

Then any sequence of estimators $\hat{\theta}_n$ with $S_n\left(\hat{\theta}_n\right) \geq S_n(\theta_0) - o_P(1)$ converges in probability to θ_0

- The proof is left for exercise. It is basically the same as the consistency theorem for MLEs
- Condition (2) is the uniform convergency
- Condition (3) is called *well-separation*
- The estimator $\hat{\theta}_n$ in this theorem *nearly maximize S_n*

Tutorial

1. Let $X_1, \ldots, X_n$ be i.i.d. from P with $EX_1^4 < \infty$ and unknown mean $\mu \in \mathcal{R}$ and variance $\sigma^2 > 0$. Consider the estimation of $\vartheta = \mu^2$ and the following three estimators: $T_{1n} = \bar{X}^2$, $T_{2n} = \bar{X}^2 - S^2/n$, $T_{3n} = \max\{0, T_{2n}\}$, where $\bar{X}$ and S^2 are the sample mean and variance.
   Show that the amse's of $T_{jn}, j = 1, 2, 3$, are the same when $\mu \neq 0$ but may be different when $\mu = 0$.
   Which estimator is the best in terms of the asymptotic relative efficiency when $\mu = 0$?

2. Let $X_1, \ldots, X_n$ be a random sample of random variables with $EX_i = \mu, \mathrm{Var}(X_i) = 1$, and $EX_i^4 < \infty$. Let $T_{1n} = n^{-1}\sum_{i=1}^n X_i^2 - 1$ and $T_{2n} = \bar{X}^2 - n^{-1}$ be estimators of μ^2, where $\bar{X}$ is the sample mean.
   Find the asymptotic relative efficiency of T_{1n} with respect to T_{2n}.

Exercise 1

Let $X_1, \ldots, X_n$ be i.i.d. from P with $EX_1^4 < \infty$ and unknown mean $\mu \in \mathcal{R}$ and variance $\sigma^2 > 0$. Consider the estimation of $\vartheta = \mu^2$ and the following three estimators: $T_{1n} = \bar{X}^2$, $T_{2n} = \bar{X}^2 - S^2/n$, $T_{3n} = \max\{0, T_{2n}\}$, where $\bar{X}$ and S^2 are the sample mean and variance.

Show that the amse's of $T_{jn}, j = 1, 2, 3,$ are the same when $\mu \neq 0$ but may be different when $\mu = 0$.

Proof: Suppose $\mu \neq 0$.

- By CLT and the δ-method, $\sqrt{n}\left(\bar{X}^2 - \mu^2\right) \xrightarrow{\mathcal{D}} N\left(0, 4\mu^2\sigma^2\right)$
- By SLLN, $S^2 \xrightarrow{\mathcal{P}} \sigma^2$ and, hence, $S^2/\sqrt{n} \xrightarrow{\mathcal{P}} 0$
- By Slutsky's theorem,

$$\sqrt{n}\left(T_{2n} - \mu^2\right) = \sqrt{n}\left(\bar{X}^2 - \mu^2\right) - S^2/\sqrt{n} \xrightarrow{\mathcal{D}} N\left(0, 4\mu^2\sigma^2\right)$$

- This shows that, amse of T_{2n} is the same as that of $T_{1n} = \bar{X}^2$
- Note that $\bar{X}^2 \xrightarrow{\mathcal{P}} \mu^2 > 0$. Hence

$$\lim_n P\left(T_{2n} \neq T_{3n}\right) = \lim_n P\left(T_{2n} < 0\right) = \lim_n P\left(\bar{X}^2 < S^2/n\right) = 0$$

- By Slutsky's theorem, $\sqrt{n}\left(T_{3n} - \mu^2\right) \xrightarrow{\mathcal{D}} N\left(0, 4\mu^2\sigma^2\right)$

Exercise 1: Suppose $\mu = 0$

- From $\sqrt{n}\bar{X} \xrightarrow{\mathcal{D}} N\left(0, \sigma^2\right)$, we conclude that $n\bar{X}^2 \xrightarrow{\mathcal{D}} \sigma^2 W$, where W has the chi-square distribution χ_1^2.
- Thus $n\left(T_{1n} - \mu^2\right) \xrightarrow{\mathcal{D}} \sigma^2 W$ and, hence, the amse of T_{1n} is $\sigma^4 EW^2/n^2 = 3\sigma^4/n^2$
- By Slutsky's theorem, $n\left(T_{2n} - \mu^2\right) = n\bar{X}^2 - S^2 \xrightarrow{\mathcal{D}} \sigma^2 W - \sigma^2$
- Hence, the amses of T_{2n} is $\sigma^4 E(W-1)^2/n^2 = \sigma^4 \operatorname{Var}(W)/n^2 = 2\sigma^4/n^2$
- Note that

  $$n\left(T_{3n} - \mu^2\right) = n \max\left\{0, T_{2n}\right\} = \max\left\{0, nT_{2n}\right\} \xrightarrow{\mathcal{D}} \max\left\{0, \sigma^2(W-1)\right\}$$

  since $\max\{0, t\}$ is continuous in t
- Then the amse of T_{3n} is $\sigma^4 E(\max\{0, W-1\})^2/n^2$
- Note that $E(\max\{0, W-1\})^2 = E\left[(W-1)^2 I_{\{W>1\}}\right] < E(W-1)^2$
- This shows that the amse of T_{3n} is the smallest when $\mu = 0$ and equal to those of T_{1n} and T_{2n} when $\mu \neq 0$.

Exercise 2

Let $X_1, \ldots, X_n$ be a random sample of random variables with $EX_i = \mu, \text{Var}(X_i) = 1$, and $EX_i^4 < \infty$. Let $T_{1n} = n^{-1}\sum_{i=1}^n X_i^2 - 1$ and $T_{2n} = \bar{X}^2 - n^{-1}$ be estimators of μ^2, where $\bar{X}$ is the sample mean.

Find the asymptotic relative efficiency of T_{1n} with respect to T_{2n}.

Proof:

- Note that $EX_1^2 = \text{Var}(X_1) + \mu^2 = 1 + \mu^2$.
- Apply the CLT to $\{X_i^2\}$ we obtain that

$$\sqrt{n}\left(T_{1n} - \mu^2\right) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n X_i^2 - \left(1 + \mu^2\right)\right] \xrightarrow{\mathcal{D}} N(0, \gamma)$$

  where $\gamma = \text{Var}\left(X_1^2\right)$.
- The amse of T_{1n} is γ/n

By the CLT, $\sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{D}} N(0,1)$

Suppose $\mu \neq 0$.

- By the δ-method (with $g(t) = t^2$ and $g'(t) = 2t$) and Slutsky's theorem,

$$\sqrt{n}\left(T_{2n} - \mu^2\right) = \sqrt{n}\left(\bar{X}^2 - \mu^2\right) - \frac{1}{\sqrt{n}} \xrightarrow{\mathcal{D}} N\left(0, 4\mu^2\right)$$

- The amse is $4\mu^2/n$

Suppose $\mu = 0$

- $n\left(T_{2n} - \mu^2\right) = n\bar{X}^2 - 1 = (\sqrt{n}\bar{X})^2 - 1 \xrightarrow{\mathcal{D}} W - 1$, where W has the chi-square distribution χ_1^2.
- Note that $E(W-1) = 0$ and $\mathrm{Var}(W-1) = 2$.
- The amse of T_{2n} is $2/n^2$

Therefore, the asymptotic relative efficiency of T_{1n} with respect to T_{2n} is

$$e_{T_1,T_2}(\mu) = \begin{cases} \frac{4\mu^2}{\mathrm{Var}\left(X_1^2\right)} = \frac{4\mu^2}{\left(EX_1^4 - (\mu^2+1)^2\right)} & \mu \neq 0 \\ \frac{2}{n\,\mathrm{Var}\left(X_1^2\right)} = \frac{2}{n\left(EX_1^4 - 1\right)} & \mu = 0 \end{cases}$$