Lecture 5. Asymptotic Theory [1]

We have seen,

1. In general, $E(\bar{X}_n) = \mu$, and $Var(\bar{X}_n) = \sigma^2/n$, where $\mu = E(X)$ and $\sigma^2 = Var(X)$.
2. Under Normality, $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

Can we say more about the distribution without the normality assumption?

- One approach is the asymptotic approach.
- In large samples (sample size n is large), averaging make all distributions look similar.

There are several core tools:

1. Weak Law of Large Numbers (WLLN)
2. Central Limit Theorem (CLT)
3. Continuous Mapping Theorem (CMT)
4. Delta Method

1 WLLN

We first explore the WLLN. For this, we need to understand convergence in a probablistic sense.

First, recall the convergence concept that you learned in calculus:

Definition (Calculus Convergence)**.** *A sequence $a_n \to a$ as $n \to \infty$, or in other words, the sequence converges to a, or $\lim_{n\to\infty} a_n = a$, if for all $\varepsilon > 0$ there is an $\bar{n} > \infty$ such that for all $n \geq \bar{n}$ we have $|a_n - a| \leq \varepsilon$.*

The calculus convergence is about a deterministic sequence. When we are considering a sequence of random variables (or in other words, a random sequence), a more appropriate convergence concept is convergence in probability:

Definition (Convergence in probability)**.** *A sequence of random variables $\{Z_n\}$ converges in probability to Z as $n \to \infty$, if for all $\varepsilon > 0$, we have*

$$\lim_{n\to\infty} P(|Z_n - Z| \geq \varepsilon) = 0.$$

We write this as $Z_n \to_p Z$ as $n \to \infty$, or $plim_{n\to\infty} Z_n = Z$, or $|Z_n - Z| = o_p(1)$.

[1] This lecture note is largely adapted from Professor Bruce Hansen's handwritten notes, however, all errors are mine.

In general, the limit Z can be a random variable, but in most cases where we use the concept of convergence in probability, Z is a constant.

Weak Law of Large Numbers (WLLN). If $\{X_1, X_2, \ldots, X_n, \ldots\}$ is an i.i.d. sequence from F and $E(|X_i|) < \infty$. Then we have $\bar{X}_n \to_p \mu$ as $n \to \infty$.

This means that $\bar{X}_n$ gets close to the parameter μ with high probabilitity when n gets large.

When an estimator converges in probability to the parameter that it is supposed to estimate, we say that the estimator is consistent:

Definition (Consistency)**.** *If an estimator $\hat{\theta}_n$ for θ satisfies $\hat{\theta}_n \to_p \theta$ as $n \to \infty$, then $\hat{\theta}_n$ is consistent for θ.*

Consistency is a nice property. It says that if n is large enough, $\hat{\theta}_n$ should be very close to θ with high probability, and thus can be interpreted as θ.

The WLLN immediately implies that $\bar{X}_n$ is a consistent estimator of μ.

We now prove the WLLN. The key ingredient of the proof is Chebyshev's inequality (or Markov's inequality), which we prove first.

Theorem (Markov's Inequality)**.** *For any random variable X with $E|X| < \infty$, and any $\lambda > 0$, we have*

$$P(|X| \geq \lambda) \leq \frac{E(|X|)}{\lambda}.$$

Proof. Note that

$$\begin{aligned}
\lambda P(|X| \geq \lambda) &= \int_{|x|\geq\lambda} \lambda f(x)dx \\
&\leq \int_{|x|\geq\lambda} |x| f(x)dx \\
&\leq \int_{-\infty}^{\infty} |x| f(x)dx \\
&= E(|X|).
\end{aligned}$$

Now divide both sides by λ and the conclusion of the theorem follows. □

Theorem (Chebychev's Inequality)**.** *For any random variable X with $\mu = E(X)$ and $Var(X) < \infty$, and any $\lambda > 0$, we have*

$$\Pr(|X - \mu| \geq \lambda) \leq \frac{Var(X)}{\lambda^2}.$$

Proof. Note that $P(|X - \mu| \geq \lambda) = P(|X - \mu|^2 \geq \lambda^2)$. Now apply Markov's inequality and we will get the Chebychev's inequality. □

Now we are ready to prove the WLLN stated above.

Proof of the WLLN. Take any $\varepsilon > 0$. Then since $E(\bar{X}) = \mu$ we have

$$P(|\bar{X}_n - \mu| \geq \varepsilon) = P(|\bar{X}_n - E\bar{X}_n| \geq \varepsilon) \leq \frac{Var(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}, \tag{1}$$

using Chebyshev's inequality and $Var(\bar{X}_n) = \sigma^2/n$. Since $\sigma^2/\varepsilon^2 < \infty$, we have $\sigma^2/(n\varepsilon^2) \to 0$ as $n \to \infty$. Then by definition of convergence in probability, we have

$$\bar{X}_n \to_p \mu.$$

□

Note that this proof assumes that $\sigma^2 < \infty$. This can be relaxed with more involved proof.

A nice property of convergence in probability is that it is preserved by continuous functions.

Theorem (Continuous Mapping Theorem (CMT))**.** *If $Z_n \to_p z$ as $n \to \infty$ and $g(\cdot)$ is continuous at z then*

$$g(Z_n) \to_p g(z) \text{ as } n \to \infty.$$

Proof. Since g is continuous at z, for all $\varepsilon > 0$, we can find $\delta > 0$ such that if $|Z_n - z| < \delta$ then $|g(Z_n) - g(z)| \leq \varepsilon$. That is, the event $|Z_n - z| < \delta$ (say we denote by A) implies the event $|g(Z_n) - g(z)| < \varepsilon$ (say we denote by B). We derived in the first couple of lectures that if $A \subseteq B$ then $P(A) \leq P(B)$. Therefore,

$$P(|Z_n - z| < \delta) \leq P(|g(Z_n) - g(z)| \leq \varepsilon).$$

The left-hand side converges to 1 as $n \to \infty$ due to $Z_n \to_p z$. Therefore, the right-hand side (which cannot be bigger than 1) also converges to 1 as $n \to \infty$. That shows that

$$P(|g(Z_n) - g(z)| > \varepsilon) \to 0 \text{ as } n \to \infty.$$

This shows $g(Z_n) \to_p g(z)$.

□

Now we give a few examples where the CMT can be used.

Example. *$\hat{\sigma}^2 \to_p \sigma^2$ as $n \to \infty$. To see why, consider that*

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - (\bar{X}_n - \mu)^2. \tag{2}$$

By WLLN, we have $\frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2 \to_p E[(X_i-\mu)^2] = \sigma^2$, *and* $\bar{X}_n - \mu \to_p 0$. *Then by CMT, we have*

$$\hat{\sigma}^2 \to_p \sigma^2 - 0 = \sigma^2.$$

Example. $s^2 \to_p \sigma^2$ *as* $n \to \infty$. *To see why, note that* $s^2 = \frac{n}{n-1}\hat{\sigma}^2$. *We have shown that* $\hat{\sigma}_n^2 \to_p \sigma^2$. *It is elementary that* $n/(n-1) \to 1$. *Therefore,*

$$s^2 \to_p 1 \times \sigma^2 = \sigma^2.$$

Example. $\hat{\sigma} \to \sigma$ *and* $s \to_p \sigma$. *For both, we can simply use the continuity of the function* $g(x) = \sqrt{x}$ *at all points* $x \geq 0$.

2 CLT

The WLLN does not provide a distributional approximation for $\bar{X}_n$. The CLT does. Before discussing the CLT, we first give a formal concept for distributional convergence:

Definition (Convergence in Distribution)**.** *A sequence of random variables/vectors* Z_n *converges in distribution to* Z *as* $n \to \infty$ *if*

$$F_n(x) = P(Z_n \leq x) \to F(x) = P(Z \leq x),$$

at all points of continuity of F. *We write this as* $Z_n \to_d Z$ *as* $n \to \infty$.

In most applications, F is a continuous distribution, and Z is a random variable/vector (r.v.).

The definition states that the distribution of Z_n gets close to the distribution of a r.v. Z.

In the special case that Z is actually a constant, $Z = c$, we have $Z_n \to_p c$ if and only if $Z_n \to_d c$.

What does it mean to say that a random variable is a constant? It means that the support of Z is a single point c and $\Pr(Z = c) = 1$.

Now we can state the CLT:

Theorem (CLT-Single Variate)**.** *If* $X_i : i = 1, 2, \ldots$ *are i.i.d. with* $E(X_i) = \mu$ *and* $Var(X_i) = \sigma^2 < \infty$ *and* $\sigma^2 > 0$ *then*

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to_d Z \sim N(0, 1).$$

Or equivalently,

$$\sqrt{n}(\bar{X}_n - \mu) \to_d N(0, \sigma^2).$$

<u>A Moment Based "Proof"</u>. Without loss of generality, assume that $\mu = 0$ and $\sigma^2 = 1$. According

to the calculations in previous lectures, we know

$$E(\bar{X}_n) = 0 \text{ and } Var(\sqrt{n}\bar{X}_n) = 1.$$

With more work, you could find that

$$E[(\sqrt{n}\bar{X}_n)^3] = \frac{1}{\sqrt{n}}E(X_i^3) \to 0,$$

and

$$E[(\sqrt{n}\bar{X}_n)^4] = \frac{E(X_i^4)}{n} + \frac{3(n-1)}{n} \to 3.$$

These moments match the same moments of $Z \sim N(0,1)$...

Formal Proof Based on MGF. Without loss of generality, assume that $\mu = 0$ and $\sigma^2 = 1$. Let $Z_n = n^{-1}\sum_{i=1}^n X_i$.

The MGF of Z_n is

$$\begin{aligned} M_n(t) &= E\exp\left(\frac{t}{\sqrt{n}}\sum_{i=1}^n X_i\right) \\ &= E\left[\Pi_{i=1}^n \exp\left(\frac{t}{\sqrt{n}}X_i\right)\right] \\ &= \Pi_{i=1}^n E\left[\exp\left(\frac{t}{\sqrt{n}}X_i\right)\right] \qquad \text{by i.i.d.} \\ &= M(t/\sqrt{n})^n, \end{aligned} \tag{3}$$

where $M(t)$ is the MGF of X_i. Take logarithm on both sides, and we have

$$\log M_n(t) = n \log M(t/\sqrt{n}). \tag{4}$$

Let $m(t) = \log M(t)$. We know that

$$\begin{aligned} M(0) &= 1 \\ M'(0) &= E(X_i) = 0 \\ M''(0) &= E(X_i^2) = 1. \end{aligned} \tag{5}$$

This implies that

$$\begin{aligned} m(0) &= \log(M(0)) = \log(1) = 0 \\ m'(0) &= \frac{M'(0)}{M(0)} = 0 \end{aligned}$$

$$m''(0) = \frac{M''(0)}{M(0)} - \frac{(M'(0))^2}{M(0)^2} = 1 \tag{6}$$

Then carry out a second order Taylor expansion and we get

$$\begin{aligned}\log(M_n(t)) &= nm(t/\sqrt{n}) \\ &\approx n\left[m(0) + m'(0)\frac{t}{\sqrt{n}} + \frac{1}{2}m''(0)\left(\frac{t}{\sqrt{n}}\right)^2\right] \\ &= n(0 + 0 + \frac{t^2}{2n}) \\ &= t^2/2. \end{aligned} \tag{7}$$

Hence, as $n \to \infty$,

$$\log M_n(t) \to t^2/2.$$

That implies that $M_n(t) \to \exp(t^2/2)$ as $n \to \infty$. The limit is the MGF of $N(0,1)$. We can then conclude that $Z_n \to_d N(0,1)$.

Sometimes, the single-variate CLT is not sufficient. We also need the multi-variate CLT given below without proof.

Theorem. *If $X_i : i = 1, 2, \ldots$ are i.i.d. random vectors, with $E(X_i) = \mu$ and $Var(X_i) = \Sigma$. Then*

$$\sqrt{n}(\bar{X}_n - \mu) \to_d N(0, \Sigma).$$

3 Continuous Mapping Theorem and Delta Method

With the concept of convergence in distribution, we can discuss the more general version of the continuous mapping theorem (CMT):

Theorem (CMT)**.** *If $Z_n \to_d Z$ and $g(\cdot)$ is continuous then*

$$g(Z_n) \to_d g(Z) \text{ as } n \to \infty.$$

Proof under the additional assumption that g is strictly increasing. Consider the derivation

$$P(g(Z_n) \leq x) = P(Z_n \leq g^{-1}(x)) \to P(Z \leq g^{-1}(x)) = P(g(Z) \leq x). \tag{8}$$

This holds for all $x \in R$. Thus, $g(Z_n) \to_d g(Z)$. □

Although our proof assumes monotonicity of g, the CMT holds without this additional assumption. The function g does not need to be $R \to R$. It can be $R^k \to R^p$ for two integers k and p. (A more general proof uses measure theoretical concepts of convergence in distribution.)

Example. *If $\sqrt{n}(\bar{X}_n - \mu) \to_d N(0, \sigma^2)$, then (with $g(x) = x^2/\sigma^2$):*

$$\frac{n(\bar{X}_n - \mu)^2}{\sigma^2} \to_d N(0,1)^2 = \chi^2_1.$$

When using the CMT with a vector Z_n that converges in distribution, it is in general not sufficient to just have the convergence in distribution of each elements of the random vector Z_n (marginal convergence). Thus it is useful to know when marginal convergence implies joint convergence:

Theorem. *Suppose that the random vector $Z_{1n} \to_d Z_1$ and the random vector $Z_{2n} \to_p z_2$ for a constant vector z_2. Then*

$$\begin{pmatrix} Z_{1n} \\ Z_{2n} \end{pmatrix} \to_d \begin{pmatrix} Z_1 \\ z_2 \end{pmatrix}.$$

The theorem states that if one component of a random vector converges in distribution and the other component of it converges in probability to a **constant**, then the two components converge in distribution jointly. A special case is where the other component is a deterministic sequence that converges in the calculus sense.

Example. *If $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \sigma^2)$, then $\hat{\theta}_n \to_p \theta$ as $n \to \infty$. This is because*

$$\hat{\theta}_n - \theta = \frac{1}{\sqrt{n}}(\sqrt{n}(\hat{\theta}_n - \theta) \to_d 0 \times N(0, \sigma^2) = 0.$$

Here the g function is $R^2 \to R$: $g(z_1, z_2) = z_1 z_2$. In this Example $Z_{1n} = 1/\sqrt{n}$ and $Z_{2n} = \sqrt{n}(\hat{\theta}_n - \theta)$.

Example (t-statistic convergence)**.** *For an i.i.d. sample $X_i : i = 1, \ldots, X_n$, recall that the t-statistic is defined to be*

$$t = \frac{\sqrt{n}(\bar{X}_n - \mu)}{s}.$$

The CLT implies that $\sqrt{n}(\bar{X}_n - \mu) \to_d N(0, \sigma^2)$ as $n \to \infty$ where $\sigma^2 = Var(X_i)$ and $\mu = E(X_i)$. For s, we have shown in the WLLN section that

$$s \to_p \sigma.$$

Thereofre,

$$(\sqrt{n}(\bar{X}_n - \mu), s)' \to_d (N(0, \sigma^2), \sigma).$$

Then by the CMT,

$$t \to_d \frac{N(0, \sigma^2)}{\sigma} = N(0, 1).$$

Another very useful technique is the Delta method. We introduce the single variate version first:

Theorem (Delta Method-Single Variate). *If $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \sigma^2)$, and $g(\cdot)$ is continuously differentiable in an open neighborhood of θ. Then*

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \to_d N(0, V),$$

where $V = (g'(\theta))^2\sigma^2$.

Proof. Take a Taylor expansion of $g(\hat{\theta}_n)$ around θ, which is valid since $\hat{\theta}_n \to_p \theta$ and g is continuously differentiable:

$$g(\hat{\theta}_n) \approx g(\theta) + g'(\theta)(\hat{\theta}_n - \theta).$$

So,

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \approx g'(\theta)\sqrt{n}(\hat{\theta}_n - \theta) \to_d g'(\theta)N(0, \sigma^2) = N(0, V). \tag{9}$$

More rigorously,

$$g(\hat{\theta}_n) = g(\theta) + g'(\theta_n^*)(\hat{\theta}_n - \theta),$$

where θ_n^* is a point on the line segment joining $\hat{\theta}_n$ and θ. Then $\theta_n^* \to_p \theta$, and $g'(\theta_n^*) \to_p g(\theta)$. □

The Delta method is very important in econometrics because most parameters of interests are functions of other parameters (that are often some moments of the population).

For example, $\beta = h(\theta)$, where $\theta = Eg(X)$ for some functions $h(\cdot)$ and $g(\cdot)$. Then we can use the following estimator for β:

$$\hat{\beta}_n = h\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right) = h(\hat{\theta}_n),$$

where $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$.

By CLT, we have

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \sigma^2),$$

where $\sigma^2 = Var(g(X_i))$.

Then by Delta Method,

$$\sqrt{n}(\hat{\beta}_n - \beta) \to_d N(0, V),$$

where $V = (g'(\theta))^2\sigma^2$. This gives a constructive distributional approximation to the estimate $\hat{\beta}_n$.

The multi-variate version of the Delta-Method can be useful as well. An example is given in the next section.

Theorem (Delta Method-Multivariate)**.** *If $h : R^k \to R^p$ is continuously differentiable at $\theta \in R^k$, and $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \Sigma)$. Then*

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta)) \to_d N(0, V),$$

where $V = H(\theta)\Sigma H(\theta)'$, and $H(\theta) = \frac{\partial}{\partial \theta'} h(\theta) = \begin{pmatrix} \frac{\partial}{\partial \theta_1} h_1(\theta) & \frac{\partial}{\partial \theta_2} h_1(\theta) & \dots & \frac{\partial}{\partial \theta_k} h_1(\theta) \\ \frac{\partial}{\partial \theta_1} h_2(\theta) & \frac{\partial}{\partial \theta_k} h_2(\theta) & \dots & \frac{\partial}{\partial \theta_k} h_2(\theta) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \theta_1} h_p(\theta) & \frac{\partial}{\partial \theta_2} h_p(\theta) & \dots & \frac{\partial}{\partial \theta_k} h_p(\theta) \end{pmatrix}$.

4 A Simple Moment-Based Estimation Technique

Suppose that $\theta = E(g(X_i))$, but our parameter of interest β is defined implicitly by the equation

$$h(\beta) = \theta,$$

for a function $h : R^k \to R^k$. Assume that $h(\cdot)$ is bijective, continuously differentiable, and $H(\beta) = \frac{\partial}{\partial \beta'} h(\beta)$ is invertible. Then, a moment-based estimator for β is the solution to the equation:

$$h(\hat{\beta}_n) = \hat{\theta}_n.$$

By CLT, we have $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \Sigma)$ where $\Sigma = Var(g(X_i))$. How do we obtain the asymptotic distribution of $\hat{\beta}_n$?

One way is to note that $\hat{\beta}_n = h^{-1}(\hat{\theta})$ and use the Multi-variate Delta Method to obtain

$$\sqrt{n}(\hat{\beta}_n - \beta) \to_d \frac{\partial}{\partial \theta'} h^{-1}(\theta) N(0, \Sigma) = \left(\frac{\partial}{\partial \beta'} h(\beta)\right)^{-1} N(0, \Sigma) = N(\mathbf{0}, H(\beta)^{-1}\Sigma(H(\beta)')^{-1}),$$

where the first equality holds by the implicit function theorem.

Alternatively, one could also conduct a Taylor expansion of $h(\hat{\beta}_n)$ around β:

$$h(\hat{\beta}_n) \approx h(\beta) + H(\beta)(\hat{\beta}_n - \beta).$$

Then

$$\sqrt{n}(\hat{\beta}_n - \beta) \approx -H(\beta)^{-1}(\hat{\theta}_n - \theta) \to_d N(\mathbf{0}, H(\beta)^{-1}\Sigma(H(\beta)')^{-1}).$$

Note that the $\approx$ can be made exact by using $H(\beta_n^*)$ instead of $H(\beta)$ for a value β_n^* on the line segment connecting $\hat{\beta}_n$ and β. Then one can use the continuous differentiability of $h(\cdot)$ and the invertibility of $H(\beta)$ to yield the conclusion.

The later derivation gets us close to the theory for the more general moment-based estimator: the generalized method of moment (GMM) estimator, which will be covered in Econ 710.

5 Problems

1. For the following sequences, show $a_n \to 0$ as $n \to \infty$:

   (a) $a_n = 1/n$

   (b) $a_n = \frac{1}{n}\sin\left(\frac{n\pi}{2}\right)$.

2. Consider a random variable X_n with the probability function

$$X_n = \begin{cases} -n & \text{with probability } 1/n \\ 0 & \text{with probability } 1-2/n \\ n & \text{with probability } 1/n \end{cases}$$

   (a) Does $X_n \to_p 0$ as $n \to \infty$?

   (b) Calculate $E(X_n)$.

   (c) Calculate $Var(X_n)$.

   (d) Now suppose the distribution is

$$X_n = \begin{cases} 0 & \text{with probability} 1-1/n \\ n & \text{with probability} 1/n \end{cases}$$

   Calculate $E(X_n)$.

   (e) Conclude that $X_n \to_p 0$ as $n \to \infty$ is not sufficient for $E(X_n) \to 0$.

3. A weighted sample mean takes the form $\bar{Y}^* = \frac{1}{n}\sum_{i=1}^n w_i Y_i$ for some non-negative constants w_i satisfying $\frac{1}{n}\sum_{i=1}^n w_i = 1$. Assume that $Y_i : i = 1, \ldots, n$ are i.i.d.

   (a) Show that $\bar{Y}^*$ is unbiased for $\mu = E(Y_i)$.

   (b) Calculate $Var(\bar{Y}^*)$.

   (c) Show that a sufficient condition for $\bar{Y}^* \to_p \mu$ is that $\frac{1}{n^2}\sum_{i=1}^n w_i^2 \to 0$. (Hint: use the Markov's or Chebyshev's Inequality).

   (d) Show that the sufficient condition for the condition in part (c) is $\max_{i\leq n} w_i/n \to 0$.

4. Take a random sample $\{X_1, \ldots, X_n\}$. Which statistic converges in probability by the weak law of large numbers and continuous mapping theorem, assuming the moment exists?

   (a) $\frac{1}{n}\sum_{i=1}^n X_i^2$.

   (b) $\frac{1}{n}\sum_{i=1}^n X_i^3$.

   (c) $\max_{i\leq n} X_i$.

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

(e) $\frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i}$, assuming $\mu = E(X_i) > 0$.

(f) $1\left(\frac{1}{n}\sum_{i=1}^n X_i > 0\right)$ where

$$1(a) = \begin{cases} 1 & \text{if } a \text{ is true} \\ 0 & \text{if } a \text{ is not true} \end{cases}$$

is called the indicator function of event a.

5. Take a random sample $\{X_1, \ldots, X_n\}$ where the support of X_i is a subset of $(0, \infty)$. Consider the sample geometric mean

$$\widehat{\mu} = \left(\Pi_{i=1}^n X_i\right)^{1/n}$$

and population geometric mean

$$\mu = \exp(E(\log(X)).$$

Assuming that μ is finite, show that $\widehat{\mu} \to_p \mu$ as $n \to \infty$.

For all the problems below, let $X_i : i = 1, \ldots, n$ be an i.i.d. sample.

6. Let $\mu_k = E(X^k)$ for some integer $k \geq 1$.

(a) Write down the natural moment estimator $\hat{\mu}_k$ of μ_k.

(b) Find the asymptotic distribution of $\sqrt{n}(\hat{\mu}_k - \mu_k)$ as $n \to \infty$, assuming that $E(X^{2k}) < \infty$.

7. Let $m_k = (E(X^k))^{1/k}$ for some integer $k \geq 1$.

(a) Write down the natural moment estimator $\hat{m}_k$ of m_k.

(b) Find the asymptotic distribution of $\sqrt{n}(\hat{m}_k - m_k)$ as $n \to \infty$, assuming that $E(X^{2k}) < \infty$.

8. Suppose $\sqrt{n}(\widehat{\mu} - \mu) \to_d N(0, v^2)$ and set $\beta = \mu^2$ and $\widehat{\beta} = \widehat{\mu}^2$.

(a) Use the Delta Method to obtain an asymptotic distribution for $\sqrt{n}(\widehat{\beta} - \beta)$.

(b) Now suppose $\mu = 0$. Describe what happens to the asymptotic distribution from the previous part.

(c) Improve on the previous answer. Under the assumption $\mu = 0$, find the asymptotic distribution $n\widehat{\beta}$.

(d) Comment on the differences between the answers in parts (a) and (c).

9. Let X be distributed Bernoulli $P(X = 1) = p$ and $P(X = 0) = 1 - p$ for some unknown $p \in (0, 1)$.

(a) Show that $p = E(X)$.

(b) Write down the natural moment estimator $\hat{p}_n$ of p.

(c) Find $Var(\hat{p}_n)$.

(d) Find the asymptotic distribution of $\sqrt{n}(\hat{p} - p)$ as $n \to \infty$.