

## Large Sample Theory

In statistics, we are interested in the properties of particular random variables (or “estimators”), which are functions of our data. In asymptotic analysis, we focus on describing the properties of estimators when the sample size becomes arbitrarily large. The idea is that given a reasonably large dataset, the properties of an estimator even when the sample size is finite are similar to the properties of an estimator when the sample size is arbitrarily large.

In these notes we focus on the large sample properties of *sample averages* formed from i.i.d. data. That is, assume that  $X_i \sim i.i.d.F$ , for  $i = 1, \dots, n, \dots$ . Assume  $EX_i = \mu$ , for all  $i$ . The sample average after  $n$  draws is  $\bar{X}_n \equiv \frac{1}{n} \sum_i X_i$ .

We focus on two important sets of large sample results:

- (1) **Law of large numbers:**  $\bar{X}_n \rightarrow EX$  as  $n \rightarrow \infty$ .
- (2) **Central limit theorem:**  $\sqrt{n}(\bar{X}_n - EX) \rightarrow N(0, \cdot)$ . That is,  $\sqrt{n}$  times a sample average looks like (in a precise sense to be defined later) a normal random variable as  $n$  gets large.

An important endeavor of asymptotic statistics is to show (1) and (2) under various assumptions on the data sampling process.



Consider a sequence of random variables  $Z_1, Z_2, \dots, Z_n$ .

### Convergence in probability:

$Z_n \xrightarrow{p} Z \iff$  for all  $\epsilon > 0$ ,  $\lim_{n \rightarrow \infty} Prob(|Z_n - Z| < \epsilon) = 1$ .

More formally: for all  $\epsilon > 0$  and  $\delta > 0$ , there exists  $n_0(\epsilon, \delta)$  such that for all  $n > n_0(\epsilon, \delta)$ :

$$Prob(|Z_n - Z| < \epsilon) > 1 - \delta.$$

Note that the limiting variable  $Z$  can also be a random variable. Furthermore, for convergence in probability, the random variables  $Z_1, Z_2, \dots$  and  $Z$  should be defined on the same probability space: if this common sample space is  $\Omega$  with elements  $\omega$ , then the statement of convergence in probability is that:

$$Prob(\omega \in \Omega : |Z_n(\omega) - Z(\omega)| < \epsilon) > 1 - \delta.$$

Corresponding to this convergence concept, we have the *Weak Law of Large Numbers* (WLLN), which is the result that  $\bar{X}_n \xrightarrow{p} \mu$ .

Earlier, we had used Chebyshev's inequality to prove a version of the WLLN, under the assumption that  $X_i$  are i.i.d. with mean  $\mu$  and variance  $\sigma^2$ .

Khinchine's law of large numbers only requires that the mean exists (i.e., is finite), but does not require existence of variances.

■■■

Recall Markov's inequality: for positive random variables  $Z$ , and  $p > 0$ ,  $\epsilon > 0$ , we have  $P(Z > \epsilon) \leq E(Z)^p / \epsilon^p$ . Take  $Z = |X_n - X^*|$ . Then if we have:

$$E[|X_n - X^*|^p] \rightarrow 0 :$$

that is,  $X_n$  converges in  $p$ -th mean to  $X^*$ , then by Markov's inequality, we also have  $P(|X_n - X^*| \geq \epsilon) \rightarrow 0$ . That is, convergence in  $p$ -th mean implies convergence in probability.

■■■

Some definitions and results:

- Define: the *probability limit* of a random sequence  $Z_n$ , denoted  $\text{plim}_{n \rightarrow \infty} Z_n$ , is a non-random quantity  $\alpha$  such that  $Z_n \xrightarrow{p} \alpha$ .
- Stochastic orders of magnitude:
  - “big-O.p.” (bounded in probability):  $Z_n = O_p(n^\lambda) \Leftrightarrow$  for every  $\delta$ , there exists a finite  $\Delta(\delta)$  and  $n^*(\delta)$  such that  $\text{Prob}(|\frac{Z_n}{n^\lambda}| > \Delta) < \delta$  for all  $n \geq n^*(\delta)$ .
  - “little-o.p.”:  $Z_n = o_p(n^\lambda) \Leftrightarrow \frac{Z_n}{n^\lambda} \xrightarrow{p} 0$ .
  - $Z_n = o_p(1) \Leftrightarrow Z_n \xrightarrow{p} 0$ .
- **Plim operator theorem:** Let  $Z_n$  be a  $k$ -dimensional random vector, and  $g(\cdot)$  be a function which is continuous at a constant  $k$ -vector point  $\alpha$ . Then

$$Z_n \xrightarrow{p} \alpha \implies g(Z_n) \xrightarrow{p} g(\alpha).$$

In other words:  $g(\text{plim } Z_n) = \text{plim } g(Z_n)$ .

*Proof:* See Serfling, Approximation Theorems of Mathematical Statistics, p. 24.

Don't confuse this: from the LLN, we have  $\text{plim } \frac{1}{n} \sum_i g(Z_i) = Eg(Z_i)$ . This is a distinct statement from  $\text{plim } g(\frac{1}{n} \sum_i Z_i) = g(EZ_i)$ , using plim operator result. Plim operator theorem allows you to move the “plim” inside the  $g(\cdot)$  function, but you cannot move the expectation operator inside  $g$  function.

- A useful intermediate result (“squeezing”): if  $Z_n \xrightarrow{p} \alpha$  (a constant) and  $Z_n^*$  lies between  $Z_n$  and  $\alpha$  with probability 1, then  $Z_n^* \xrightarrow{p} \alpha$ .

(Quick proof: for all  $\epsilon > 0$ ,  $|X_n - \alpha| < \epsilon \Rightarrow |X_n^* - \alpha| < \epsilon$ . Hence  $Prob(|X_n^* - \alpha| < \epsilon) \geq Prob(|X_n - \alpha| < \epsilon) \rightarrow 1$ .)



### Convergence almost surely

$Z_n \xrightarrow{as} Z \iff Prob(\omega : \lim_{n \rightarrow \infty} Z_n(\omega) = Z(\omega)) = 1$ . Let’s deep dive into this probability statement.

As with convergence in probability, almost sure convergence makes sense when the random variables  $Z_1, \dots, Z_n, \dots$  and  $Z$  are all defined on the same sample space, call it  $(\Omega, \mathcal{B}(\Omega), P)$ . Hence, for each element in the sample space  $\omega \in \Omega$ , we can associate the sequence  $Z_1(\omega), \dots, Z_n(\omega), \dots$  as well as the limit point  $Z(\omega)$ .

Define some set-theoretic concepts. Consider a sequence of sets  $S_1, \dots, S_n$ , all  $\in \mathcal{B}(\Omega)$ . Unless this sequence is monotonic, it’s difficult to talk about a “limit” of this sequence. So we define the “liminf” and “limsup” of this set sequence.

$$\begin{aligned} \liminf_{n \rightarrow \infty} S_n &\equiv \lim_{n \rightarrow \infty} \bigcap_{m=n}^{\infty} S_m = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} S_m \\ &= \{\omega \in \Omega : \omega \in S_n, \forall n \geq n_0(\omega)\}. \end{aligned}$$

Note that the sequence of sets  $\bigcap_{m=n}^{\infty} S_m$ , for  $n = 1, 2, 3, \dots$  is a non-decreasing sequence of sets. By taking the union, the liminf is this the limit of this monotone sequence of sets. That is, for all  $\omega \in \liminf S_n$ , there exists some number  $n_0(\omega)$  such that  $\omega \in S_n$  for all  $n \geq n_0(\omega)$ . Hence we can say that  $\liminf S_n$  is the set of outcomes which occur “eventually”.

$$\begin{aligned} \limsup_{n \rightarrow \infty} S_n &\equiv \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} S_m = \lim_{n \rightarrow \infty} \bigcup_{m=n}^{\infty} S_m \\ &= \{\omega \in \Omega : \text{for every } m, \omega \in S_{n_0(\omega, m)} \text{ for some } n_0(\omega, m) \geq m\}. \end{aligned}$$

Note that the sequence of sets  $\bigcup_{m=n}^{\infty} S_m$ , for  $n = 1, 2, 3, \dots$  is a non-increasing sequence of sets. Hence, limsup is limit of this monotone sequence of sets. The  $\limsup S_n$  is

the set of outcomes  $\omega$  which occurs within every tail of the set sequence  $S_n$ . Hence we say that an outcome  $\omega \in \limsup S_n$  occurs “infinitely often”.<sup>1</sup>

Note that

$$\liminf S_n \subset \limsup S_n.$$

Hence

$$P(\limsup S_n) \geq P(\liminf S_n)$$

and

$$\begin{aligned} P(\limsup S_n) = 0 &\rightarrow P(\liminf S_n) = 0 \\ P(\liminf S_n) = 1 &\rightarrow P(\limsup S_n) = 1. \end{aligned}$$

**Borel-Cantelli Lemma:** if  $\sum_{i=1}^{\infty} P(S_i) < \infty$  then  $P(\limsup_{n \rightarrow \infty} S_n) = 0$ .

Proof: for each  $m$ , we have

$$\begin{aligned} P(\limsup_{n \rightarrow \infty} S_n) &= P(\lim \left\{ \bigcup_{m=n}^{\infty} S_m \right\}) = \lim_{n \rightarrow \infty} P(\bigcup_{m=n}^{\infty} S_m) \\ &\leq \lim_{n \rightarrow \infty} \sum_{m \geq n} P(S_m) \end{aligned}$$

which equals zero (by the assumption that  $\sum_{i=1}^{\infty} P(S_i) < \infty$ ).

Since  $\{\bigcup_{m=n}^{\infty} S_m\}$  is a weakly increasing set sequence in  $n$ ,  $P(\bigcup_{m=n}^{\infty} S_m)$  is a weakly increasing sequence as well thus enabling the interchange of “limit” and “Prob” operations in the first equality. The  $\leq$  inequality comes from Boole’s inequality.

■■■

To apply this to almost-sure convergence, consider the sequence of sets

$$S_n \equiv \{\omega : |Z_n(\omega) - Z(\omega)| > \epsilon\},$$

for  $\epsilon > 0$ . Almost sure convergence is the statement that  $P(\limsup S_n) = 0$ :

- Then  $\bigcup_{m=n}^{\infty} S_m$  denotes the  $\omega$ ’s such that the sequence  $|Z_n(\omega) - Z(\omega)|$  exceeds  $\epsilon$  in the tail beyond  $n$ .

---

<sup>1</sup>However, note that all outcomes  $\omega \in \liminf S_n$  also occur an infinite number of times along the infinite sequence  $S_n$ .

- Then  $\lim_{n \rightarrow \infty} \bigcup_{m=n}^{\infty} S_m$  denotes all  $\omega$ 's such that the sequence  $|Z_n(\omega) - Z(\omega)|$  exceeds  $\epsilon$  in every tail: those  $\omega$  for which  $Z_n(\omega)$  “escapes” an  $\epsilon$ -ball around  $Z(\omega)$  infinitely often. For these  $\omega$ , either  $\lim Z_n(\omega)$  doesn't exist, or if it does,  $\lim Z_n(\omega) \neq Z(\omega)$ .
- Almost-sure convergence requires the probability of this set to equal zero, i.e.  $\Pr(\lim_{n \rightarrow \infty} \bigcup_{m=n}^{\infty} S_m) = 0$ .

■■■

**Example:**  $X_i \sim i.i.d. U[0, 1]$ . Show that  $X_{(1:n)} \equiv \min_{i=1, \dots, n} X_i \xrightarrow{as} 0$ .

Take  $S_n \equiv \{|X_{(1:n)}| > \epsilon\}$ . For all  $\epsilon > 0$ ,  $P(|X_{(1:n)}| > \epsilon) = P(X_{(1:n)} > \epsilon) = P(X_i > \epsilon, i = 1, \dots, n) = (1 - \epsilon)^n$ .

Hence,  $\sum_{n=1}^{\infty} (1 - \epsilon)^n = 1/\epsilon < \infty$ . So the conclusion follows by the BC Lemma.

■■■

**Theorem:**  $Z_n \xrightarrow{as} Z \implies Z_n \xrightarrow{p} Z$ .

**Proof:**

$$\begin{aligned}
0 &= P\left(\lim_{n \rightarrow \infty} \bigcup_{m \geq n}^{\infty} \{\omega : |Z_m(\omega) - Z(\omega)| > \epsilon\}\right) \\
&= \lim_{n \rightarrow \infty} P\left(\bigcup_{m \geq n}^{\infty} \{\omega : |Z_m(\omega) - Z(\omega)| > \epsilon\}\right) \\
&\geq \lim_{n \rightarrow \infty} P(\{\omega : |Z_n(\omega) - Z(\omega)| > \epsilon\}).
\end{aligned}$$

■■■

Corresponding to this convergence concept, we have the *Strong Law of Large Numbers* (SLLN), which is the result that  $\bar{X}_n \xrightarrow{as} \mu$ . Assume, as before, that  $X_1, X_2, X_3, \dots$  are iid distributed with mean  $\mu$  and variance  $\sigma^2$ . Consider that the sample averages are random variables  $\bar{X}_1(\omega), \bar{X}_2(\omega), \dots$  defined on the same probability space, say  $(\Omega, \mathcal{B}, P)$ , where each  $\omega$  indexes a sequence a sequence  $\bar{X}_1(\omega), \bar{X}_2(\omega), \bar{X}_3(\omega), \dots$ . Consider the set sequence

$$S_n = \{\omega : |\bar{X}_n(\omega) - \mu| > \epsilon\}.$$

SLLN is the statement that  $P(\limsup S_n) = 0$ .

**Proof (sketch; Davidson, pg. 296):** For convenience, consider  $\mu = 0$ . Given the above discussion, we set out to prove  $P(|\bar{X}_n| > \epsilon, i.o.) = 0$  for any  $\epsilon > 0$ . We intend

to verify that  $\sum_{n=1}^{\infty} P(|\bar{X}_n| > \epsilon) < \infty$  for all  $\epsilon > 0$ , and then apply the Borel-Cantelli Lemma.

As a starting point, let's use Chebyshev's inequality to bound  $P(|\bar{X}_n| > \epsilon)$  (as we did for the WLLN):

$$P(|\bar{X}_n| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}; \quad \text{but} \quad \sum_{n=1}^{\infty} \frac{1}{n\epsilon^2} = \infty.$$

Hence this is not good enough.

However, consider the subsequence with indices  $n^2$ :  $\{1, 4, 9, 16, \dots\}$ . Again using Chebyshev's inequality, we have<sup>2</sup>

$$P(|\bar{X}_{n^2}| > \epsilon) \leq \frac{\sigma^2}{n^2\epsilon^2}; \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{n^2\epsilon^2} = \frac{1}{\epsilon^2}(\pi^2/6) \approx \frac{1.64}{\epsilon^2} < \infty.$$

Hence, by BC lemma, we have that  $\bar{X}_{n^2} \xrightarrow{as} 0$ .

Next, examine the “omitted” terms from the sum  $\sum_{n=1}^{\infty} P(|\bar{X}_n| > \epsilon)$ . Let  $k$  be chosen such that  $n^2 \leq k \leq (n+1)^2$ . Note that:

$$\bar{X}_k - \bar{X}_{n^2} = \left(\frac{n^2}{k} - 1\right) \bar{X}_{n^2} + \frac{1}{k} \sum_{t=n^2+1}^k X_t$$

and, by independence of  $X_i$ , the two terms on RHS are uncorrelated. Hence

$$\begin{aligned} Var(\bar{X}_k - \bar{X}_{n^2}) &= \left(1 - \frac{n^2}{k}\right)^2 \frac{\sigma^2}{n^2} + \frac{k - n^2}{k^2} \sigma^2 \\ &= \sigma^2 \left(\frac{1}{n^2} - \frac{1}{k}\right) \leq \sigma^2 \left(\frac{1}{n^2} - \frac{1}{(n+1)^2}\right). \end{aligned}$$

Define  $D_{n^2} = \max_{k: n^2 \leq k < (n+1)^2} |\bar{X}_k - \bar{X}_{n^2}|$ . By Chebyshev's inequality, then, we have

$$P(D_{n^2} > \epsilon) \leq \frac{\sigma^2}{\epsilon} \left(\frac{1}{n^2} - \frac{1}{(n+1)^2}\right); \quad \text{so}$$

$$\sum_{n^2} P(D_{n^2} > \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \sum_{n^2} \left(\frac{1}{n^2} - \frac{1}{(n+1)^2}\right) < \frac{\sigma^2}{\epsilon^2} \sum_n \left(\frac{1}{n^2} - \frac{1}{(n+1)^2}\right) = \frac{\sigma^2}{\epsilon^2}$$

implying, using BC Lemma, that  $D_{n^2} \xrightarrow{as} 0$ .

---

<sup>2</sup>Another of Euler's greatest hits:  $\sum_{i=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$  (the Riemann zeta function).

Now, consider the general process  $\bar{X}_l$  for  $l = 1, 2, 3, 4, \dots$ . Consider any  $l$ , it must satisfy  $n^2 \leq l < (n+1)^2$  for some  $n$ . Note that, with probability 1,

$$\begin{aligned} |\bar{X}_l| &= |\bar{X}_l - \bar{X}_{n^2} + \bar{X}_{n^2}| \\ &\leq |\bar{X}_l - \bar{X}_{n^2}| + |\bar{X}_{n^2}| \\ &\leq D_{n^2} + |\bar{X}_{n^2}|. \end{aligned}$$

Hence  $P(|\bar{X}_l| > \epsilon, i.o.) \leq P(D_{n^2} + |\bar{X}_{n^2}| > \epsilon, i.o.) = 0$  since both  $D_{n^2}$  and  $X_{n^2}$  converge a.s. to 0. Thus,  $\bar{X}_l \xrightarrow{a.s.} 0$ . ■

■■■

**Convergence in Distribution:**  $Z_n \xrightarrow{d} Z$ .

A sequence of real-valued random variables  $Z_1, Z_2, Z_3, \dots$  converges in distribution to a random variable  $Z$  iff

$$\lim_{n \rightarrow \infty} F_{Z_n}(z) = F_Z(z) \quad \forall z \text{ s.t. } F_Z(z) \text{ is continuous.} \quad (1)$$

This is a statement about the CDFs of the random variables  $Z$ , and  $Z_1, Z_2, \dots$ . These random variables do not need to be defined on the same probability space.  $F_Z$  is also called the “limiting distribution” of the random sequence  $Z_n$ .

■■■

Alternative definitions of convergence in distribution (“Portmanteau theorem”):<sup>3</sup>

- Letting  $F([a, b]) \equiv F(b) - F(a)$  for  $a < b$ , convergence (1) is equivalent to

$$F_{Z_n}([a, b]) \rightarrow F_Z([a, b]) \quad \forall [a, b] \text{ s.t. } F_Z \text{ continuous at } a, b.$$

- For all bounded, continuous functions  $g(\cdot)$ ,

$$\int g(z) dF_{Z_n}(z) \rightarrow \int g(z) dF_Z(z) \quad \Leftrightarrow \quad Eg(Z_n) \rightarrow Eg(Z). \quad (2)$$

This definition of distributional convergence is more useful in advanced settings, because it is extendable to setting where  $Z_n$  and  $Z$  are general random elements taking values in metric space.

---

<sup>3</sup>For proofs of most results here, see Serfling, Approximation Theorems in Mathematical Statistics, ch. 1.

■■■

The following key theorem postulates a relation between convergence in distribution and pointwise convergence of characteristic functions. This treatment is taken from Durrett, *Probability: Theory and Examples*, pp. 80-100.

**Levy's continuity theorem:** Let  $X_n$  be a sequence of random variables with characteristic functions  $\phi_n(t)$ .

1. If  $X_n \xrightarrow{d} X_\infty$ , then  $\phi_n(t) \rightarrow \phi_\infty(t)$  for all  $t$ .
2. If  $\phi_n(t)$  converges pointwise in  $t$  to a limit  $\phi(t)$  which is continuous at 0, then the associated random variables  $X_n$  converges in distribution to a random variable  $X$  which has characteristic function  $\phi(t)$ .

The proof is advanced and involves results and concepts beyond this class.

**(Counter-)Example for part (2):** Let  $Z \sim N(0, 1)$  and define  $X_n = nZ$ , which is a normal distribution with mean 0 and variance  $n^2$ . In this case  $\phi_n(t) = \exp(-n^2 t^2 / 2)$  which converges to  $\mathbf{1}(t = 0)$  which is discontinuous at 0. The random sequence  $X_n$  converges to a random element  $X_\infty$  which equals  $+\infty$  or  $-\infty$  with probability 0.5 – which is not a random variable ( $X_\infty$  is not “tight”).

This theorem will be used to prove the CLT later. Recipe is the following: we will show (pointwise) convergence for a sequence of characteristic functions. The limit function is continuous at  $t = 0$ . Then we obtain convergence in distribution.

■■■

Distributional convergence, as defined above, is called “weak convergence”. This is because there are stronger notions of distributional convergence. One such notion is:

$$\sup_{A \in \mathcal{B}(\mathbb{R})} |P_{Z_n}(A) - P_Z(A)| \rightarrow 0$$

which is convergence in *total variation norm*.

Example: Multinomial distribution on  $[0, 1]$ .  $Z_n \in \{\frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \dots, \frac{n}{n} = 1\}$  each with probability  $\frac{1}{n}$ . We have  $Z_n \xrightarrow{d} U[0, 1]$ . Consider  $A = \mathbb{Q}$  (set of rational numbers in  $[0, 1]$ ).

■■■

Some definitions and results:



- **Slutsky Theorem:** If  $Z_n \xrightarrow{d} Z$ , and  $Y_n \xrightarrow{p} \alpha$  (a constant), then

(a)  $Y_n Z_n \xrightarrow{d} \alpha Z$

(b)  $Z_n + Y_n \xrightarrow{d} Z + \alpha$ .

- **Theorem \*:**

(a)  $Z_n \xrightarrow{p} Z \implies Z_n \xrightarrow{d} Z$

(b)  $Z_n \xrightarrow{d} Z \implies Z_n \xrightarrow{p} Z$  if  $Z$  is a constant.

Note: convergence in probability implies that (roughly speaking), the random variables  $Z_n$  (for  $n$  large enough) and  $Z$  frequently have the same numerical value. Convergence in distribution need not imply this, only that the CDF's of  $Z_n$  and  $Z$  are similar.



Note that the LLN tells us that  $\bar{X}_n \xrightarrow{p,as} \mu$ , which implies (trivially) that  $\bar{X}_n \xrightarrow{d} \mu$ , a *degenerate* limiting distribution.

For inference, we need to know how *far*  $\bar{X}_n$  is from  $\mu$ ; LLN does not tell us this.

The following result does.

**Central Limit Theorem:** (Lindeberg-Levy) Let  $X_i$  be i.i.d. with mean  $\mu$  and variance  $\sigma^2$ . Then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1). \quad (3)$$

$\sqrt{n}$  is also called the “rate of convergence” of the sequence  $\bar{X}_n$ . By definition, a rate of convergence is the lowest polynomial in  $n$  for which  $n^p(\bar{X}_n - \mu)$  converges in distribution to a nondegenerate distribution.

If you blow up  $\bar{X}_n$  by a constant (no matter how big), you still get a degenerate limiting distribution. If you blow up by  $n$ , then the sequence  $S_n \equiv n\bar{X}_n = \sum_{i=1}^n X_i$  will diverge.

Taking (3) as an approximation, we have that  $\bar{X}_n$  is “approximately” distributed  $N(\mu, \sigma^2/n)$ ; thus the rate of convergence of  $\sqrt{n}$  implies that the variance of  $\bar{X}_n$  approximately falls proportionally to  $n$  (doubling the sample size reduces the variance by half). This is expected as we have assumed that  $X_i$  are iid with mean  $\mu$  and variance  $\sigma^2$  which immediately implies that the variance of  $\bar{X}_n = \sigma^2/n$ .

**Proof via characteristic functions:** Consider  $X_i \sim i.i.d.$  with mean  $\mu$  and variance  $\sigma^2$ . Define:

$$Z_n \equiv \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \sum_i \frac{X_i - \mu}{\sqrt{n}\sigma} \equiv \frac{1}{\sqrt{n}} \sum_i W_i.$$

Note that,  $W_i \equiv (X_i - \mu)/\sigma$  are i.i.d. with mean  $\mathbb{E}W_i = 0$  and variance  $\mathbb{V}W_i = \mathbb{E}W_i^2 = 1$ .

Let  $\phi(t)$  denote characteristic function. Then

$$\phi_{Z_n}(t) = \left[ \phi_{\frac{W}{\sqrt{n}}}(t) \right]^n = [\phi_W(t/\sqrt{n})]^n.$$

Using the second-order Taylor expansion for  $\phi_W(t/\sqrt{n})$  around 0, we have:

$$\begin{aligned} \phi_W(t/\sqrt{n}) &= \phi_W(0) + \frac{t}{\sqrt{n}} i \mathbb{E}W + \frac{t^2}{2n} i^2 \mathbb{E}W^2 + o\left(\frac{t^2}{n}\right) \\ &= \phi_W(0) + 0 - \frac{t^2}{2n} * 1 + o\left(\frac{t^2}{n}\right) \end{aligned}$$

Now we have

$$\begin{aligned} \log \phi_{Z_n(t)}(t) &= n * \log \left\{ \phi_W(0) + 0 + \frac{(it)^2}{2n} * 1 + o\left(\frac{t^2}{n}\right) \right\} \\ &= n * \log \left\{ 1 + \frac{(it)^2}{2n} + o\left(\frac{t^2}{n}\right) \right\} \\ &\approx n * \left\{ \log 1 + \left[ \frac{(it)^2}{2n} + o\left(\frac{t^2}{n}\right) \right] * \log'(1) + o\left(\frac{t^2}{n}\right) \right\} \\ &= 0 - \frac{t^2}{2} + n * o\left(\frac{t^2}{n}\right) \\ &\xrightarrow{n \rightarrow \infty} \frac{-t^2}{2} \end{aligned}$$

The third line comes from Taylor-expanding  $\log \{\dots\}$  around 1. Hence

$$\phi_{Z_n(t)}(t) \xrightarrow{n \rightarrow \infty} \exp\left(\frac{-t^2}{2}\right) = \phi_{N(0,1)}(t).$$

Since  $\phi_{N(0,1)}(t)$  is continuous at  $t = 0$ , we have  $Z_n \xrightarrow{d} N(0, 1)$ . ■

■■■

**Asymptotic approximations for  $\bar{X}_n$**

CLT tells us that  $\sqrt{n}(\bar{X}_n - \mu)/\sigma$  has a limiting standard normal distribution. We can use this “exact” result to say something about the distribution of  $\bar{X}_n$ , even when  $n$  is finite. That is, we use the CLT to derive an *asymptotic approximation* for the finite-sample distribution of  $\bar{X}_n$ .

The approximation is as follows: starting with the result of the CLT:

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0, 1)$$

we “flip over” the result of the CLT:

$$\bar{X}_n \stackrel{a}{\sim} \frac{\sigma}{\sqrt{n}}N(0, 1) + \mu \Rightarrow \bar{X}_n \stackrel{a}{\sim} N(\mu, \frac{1}{n}\sigma^2).$$

The notation “ $\stackrel{a}{\sim}$ ” makes explicit that what is on the RHS is an approximation. Note that  $\bar{X}_n \xrightarrow{d} N(\mu, \frac{1}{n}\sigma^2)$  is definitely not true!

This approximation intuitively makes sense: under the assumptions of the LLCLT, we know that  $E\bar{X}_n = \mu$  and  $Var(\bar{X}_n) = \sigma^2/n$ . What the asymptotic approximation tells us is that the distribution of  $\bar{X}_n$  is approximately normal.<sup>4</sup>

### Asymptotic approximations for functions of $\bar{X}_n$

Oftentimes, we are not interested *per se* in approximating the finite-sample distribution of the sample mean  $\bar{X}_n$ , but rather functions of a sample mean. (Later, you will see that the asymptotic approximations for many statistics and estimators that you run across are derived by expressing them as sample averages.)

**Continuous mapping theorem:** Let  $g(\cdot)$  be a continuous function. Then

$$Z_n \xrightarrow{d} Z \implies g(Z_n) \xrightarrow{d} g(Z).$$

*Proof:* Serfling, Approximation Theorems of Mathematical Statistics, p. 25.

(Note: you still have to figure out what the limiting distribution of  $g(Z)$  is. But if you know  $F_X$ , then you can get  $F_{g(X)}$  by the change of variables formulas.)

Note that for any linear function  $g(\bar{X}_n) = a\bar{X}_n + b$ , deriving the limiting distribution of  $g(\bar{X}_n)$  is no problem (just use Slutsky’s Theorem to get  $a\bar{X}_n + b \stackrel{a}{\sim} N(a\mu + b, \frac{1}{n}a^2\sigma^2)$ ).

What about when  $g(\cdot)$  is nonlinear?

---

<sup>4</sup>Note that the approximation is exactly right if we assumed that  $X_i \sim N(\mu, \sigma^2)$ , for all  $i$ .

1. Use the **Mean Value Theorem**: Let  $g$  be a continuous function on  $[a, b]$  that is differentiable on  $(a, b)$ . Then, there exists (at least one)  $\lambda \in (a, b)$  such that

$$g'(\lambda) = \frac{g(b) - g(a)}{b - a} \Leftrightarrow g(b) - g(a) = g'(\lambda)(b - a).$$

Using the MVT, we can write

$$g(\bar{X}_n) - g(\mu) = g'(X_n^*)(\bar{X}_n - \mu) \quad (4)$$

where  $X_n^*$  is an RV strictly between  $\bar{X}_n$  and  $\mu$ .

2. On the RHS of Eq. (4):

- (a)  $g'(X_n^*) \xrightarrow{p} g'(\mu)$  by the “squeezing” result, and the plim operator theorem.
- (b) If we multiply by  $\sqrt{n}$  and divide by  $\sigma$ , we can apply the CLT to get  $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0, 1)$ .
- (c) Hence,

$$\sqrt{n} * \text{RHS} = \sqrt{n}[g(\bar{X}_n) - g(\mu)] \xrightarrow{d} N(0, g'(\mu)^2 \sigma^2), \quad (5)$$

using Slutsky’s theorem.

3. Now, in order to get the asymptotic approximation for the distribution of  $g(\bar{X}_n)$ , we “flip over” to get

$$g(\bar{X}_n) \overset{a}{\sim} N(g(\mu), \frac{1}{n}g'(\mu)^2 \sigma^2).$$

4. Check that  $g$  satisfies assumptions for these results to go through: continuity and differentiability for MVT, and continuous at  $\mu$  for plim operator theorem.

Examples:  $1/\bar{X}_n$ ,  $\exp(\bar{X}_n)$ ,  $(\bar{X}_n)^2$ , etc.

Eq. (5) is a general result, known as the **Delta method**. For the purposes of this class, I want you to derive the approximate distributions for  $g(\bar{X}_n)$  from first principles (as we did above).