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 $i = 1, \ldots, n, \ldots$. 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**: $\bar{X}_n \to EX$ as $n \to \infty$.

(2) **Central limit theorem**: $\sqrt{n}(\bar{X}_n - EX) \to N(0, \cdot)$. 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 $Z_1, Z_2, \ldots, Z_n$.

Convergence in probability:

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

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

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

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

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

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 μ and variance σ^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:

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

that is, X_n *converges in p-th mean* to X^*, then by Markov's inequality, we also have $P(|X_n - X^*| \geq \epsilon) \to 0$. 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 $\text{plim}_{n\to\infty} Z_n$, is a non-random quantity α 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 δ, there exists a finite $\Delta(\delta)$ and $n^*(\delta)$ such that $Prob(|\frac{Z_n}{n^\lambda}| > \Delta) < \delta$ 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 α. Then

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

  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: plim $\frac{1}{n}\sum_i g(Z_i) = Eg(Z_i)$, from the LLN, but this is distinct from plim $g(\frac{1}{n}\sum_i Z_i) = g(EZ_i)$, using plim operator result. The two are generally not the same; if $g(\cdot)$ is convex, then plim $\frac{1}{n}\sum_i g(Z_i) \geq$ plim $g(\frac{1}{n}\sum_i Z_i)$.

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

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

■■■

Convergence almost surely

$Z_n \overset{as}{\to} Z \iff Prob\,(\omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)) = 1.$

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

To understand the probability statement more clearly, assume that all these RVs are defined on the sample space $(\Omega, \mathcal{B}(\Omega), P)$.

Define some set-theoretic concepts. Consider a sequence of sets $S_1, \ldots, S_n$, 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.

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

Note that the sequence of sets $\bigcap_{m=n}^{\infty} S_m$, for $n = 1, 2, 3, \ldots$ 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 $\omega \in \lim\inf S_n$, 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 $\lim\inf S_n$ is the set of outcomes which occur "eventually".

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

Note that the sequence of sets $\bigcup_{m=n}^{\infty} S_n$, for $n = 1, 2, 3, \ldots$ 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 ω 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".[1]

Note that

$$\text{liminf } S_n \subset \text{limsup } S_n.$$

Hence

$$P(\text{limsup } S_n) \geq P(\text{liminf } S_n)$$

and

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

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 $\sum_{i=1}^{\infty} P(S_i) < \infty$).

The first equality (interchange of "limit" and "Prob" operations) is an application of the Monotone Convergence Theorem, which holds only because the sequence $\bigcup_{m=n}^{\infty} S_m$ is monotone.

■■■

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 $P(\text{limsup } S_n) = 0$:

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

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

- Then $\lim_{n\to\infty} \bigcup_{m=n}^{\infty} S_m$ denotes all ω's such that the sequence $|Z_n(\omega) - Z(\omega)|$ exceeds ϵ in every tail: those ω for which $Z_n(\omega)$ "escapes" the ϵ-ball around $Z(\omega)$ infinitely often. For these ω's, $\lim_{n\to\infty} Z_n(\omega) \neq Z(\omega)$.

- For almost-sure convergence, you require the probability of this set to equal zero, i.e. $\Pr(\lim_{n\to\infty} \bigcup_{m=n}^{\infty} S_m) = 0$.

■■■

Corresponding to this convergence concept, we have the *Strong Law of Large Numbers* (SLLN), which is the result that $\bar{X}_n \stackrel{as}{\to} \mu$. Consider that the sample averages are random variables $\bar{X}_1(\omega), \bar{X}_2(\omega), ...$ defined on the same probability space, say $(\Omega, \mathcal{B}, P)$, where each ω indexes a sequence a sequence $\bar{X}_1(\omega), \bar{X}_2(\omega), \bar{X}_3(\omega), \ldots$. Consider the set sequence

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

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

Proof (sketch; Davidson, pg. 296): For convenience, consider $\mu = 0$. From the above discussion, we see that the two statements $\bar{X}_n \stackrel{as}{\to} 0$ and $P(|\bar{X}_n| > \epsilon,\ i.o.) = 0$ for any $\epsilon > 0$ are equivalent. Therefore, we proceed in proving the SLLN by verifying 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, we see that Chebyshev's inequality is not good enough by itself; it tells us

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

However, consider the subsequence with indices n^2: $\{1, 4, 9, 16, \ldots\}$. Again using Chebyshev's inequality, we have[2]

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

Hence, by BC lemma, we have that $\bar{X}_{n^2} \stackrel{as}{\to} 0$. Next, examine the "omitted" terms from the sum $\sum_{n=1}^{\infty} P(|\bar{X}_n| > \epsilon)$. Define $D_{n^2} = \max_{n^2 \leq k < (n+1)^2} |\bar{X}_k - \bar{X}_{n^2}|$. Note that:

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

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

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

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

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 $D_{n^2} \stackrel{as}{\to} 0$.

Now, consider, for $n^2 \leq l < (n+1)^2$,

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

By the above discussion, the RHS converges a.s. to 0. Hence, for all integers l, we have that $|\bar{X}_l|$ is bounded by a random variable which converges a.s. to 0. Thus, $\bar{X}_l \stackrel{as}{\to} 0$. ■

■■■

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

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

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

■■■

Theorem: $Z_n \stackrel{as}{\to} Z \Longrightarrow Z_n \stackrel{p}{\to} Z$.

Proof:

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

■■■

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

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

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

This is a statement about the CDFs of the random variables Z, and $Z_1, Z_2, \ldots$. 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"):

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

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

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

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

  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.

- **Levy's continuity theorem**: A sequence $\{X_n\}$ of random variables converges in distribution to random variable X if and only if the sequence of characteristic function $\{\phi_{X_j}(t)\}$ converges pointwise (in t) to a function $\phi_X(t)$ which is continuous at the origin. Then ϕ_X is the characteristic function of X.

Hence, this ties together convergence of characteristic functions, and convergence in distribution. This theorem will be used to prove the CLT later.

- For proofs of most results here, see Serfling, Approximation Theorems in Mathematical Statistics, ch. 1.

■■■

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

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

which is convergence in *total variation norm*.

Example: Multinomial distribution on $[0, 1]$. $X_n \in \left\{\frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{n}{n} = 1\right\}$ each with probability $\frac{1}{n}$. Consider $A = \mathbb{Q}$ (set of rational numbers in $[0, 1]$).

■■■

Some definitions and results:

- **Slutsky Theorem:** If $Z_n \overset{d}{\to} Z$, and $Y_n \overset{p}{\to} \alpha$ (a constant), then
  (a) $Y_n Z_n \overset{d}{\to} \alpha Z$
  (b) $Z_n + Y_n \overset{d}{\to} Z + \alpha$.

- **Theorem *:**

$$(a)\ Z_n \overset{p}{\to} Z \Longrightarrow Z_n \overset{d}{\to} Z$$
$$(b)\ Z_n \overset{d}{\to} Z \Longrightarrow Z_n \overset{p}{\to} Z \text{ if } Z \text{ is a constant.}$$

  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.

- $Z_n \overset{d}{\to} Z \implies Z_n = O_p(1)$. Use $\Delta = \max(|F_Z^{-1}(1 - \epsilon)|, |F_Z^{-1}(\epsilon)|)$ in definition of $O_p(1)$.

■■■

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

This is not very useful for our purposes, because we are interested in knowing (say) how far $\bar{X}_n$ is from μ, which is unknown.

How do we "fix" $\bar{X}_n$, so that it has a non-degenerate limiting distribution?

Central Limit Theorem: (Lindeberg-Levy) Let X_i be i.i.d. with mean μ and variance σ^2. Then

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

$\sqrt{n}$ 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.

The rate of convergence $\sqrt{n}$ make sense: 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 $S_n \equiv n\bar{X}_n = \sum_{i=1}^n X_n$ will diverge.

Proof via characteristic functions: Consider $X_i \sim i.i.d.$ with mean μ and variance σ^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

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

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

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

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

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

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 "true" 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:

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

The notation "$\overset{a}{\sim}$" 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.[3]

[3]Note that the approximation is exactly right if we assumed that $X_i \sim N(\mu, \sigma^2)$, for all i.

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

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

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

The "problem" in deriving the distribution of $g(\bar{X}_n)$ arises when $g(\cdot)$ is nonlinear so that the $\bar{X}_n$ is "inside" the g function:

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

$$g(\bar{X}_n) - g(\mu) = g'(X_n^*)(\bar{X}_n - \mu) \tag{3}$$

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

2. On the RHS of Eq. (3):

   (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 σ, we can apply the CLT to get $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0, 1)$.

   (c) Hence,

$$\sqrt{n} * \text{RHS} = \sqrt{n}[g(\bar{X}_n) - g(\mu)] \xrightarrow{d} N(0, g'(\mu)^2\sigma^2), \tag{4}$$

   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 μ for plim operator theorem.

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

Eq. (4) 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).

1 Some extra topics (CAN SKIP)

1.1 Triangular array CLT

LLCLT assumes independence (across n) as well as identically distributed. We extend this to the "independent, non-identically distributed" setting.

Lindeberg-Feller CLT: For each n, let $Y_{n,1}, \ldots, Y_{n,n}$ be independent random variables with finite (possibly non-identical) means and variances $\sigma_{n,i}^2$ such that

$$(1/C_n)^2 \cdot \sum_{i=1}^{n} \mathbb{E}|Y_{n,i} - \mathbb{E}Y_{n,i}|^2 \mathbb{1}_{\{|Y_{n,i} - \mathbb{E}Y_{n,i}|/C_n > \epsilon\}} \to 0 \quad \text{every } \epsilon > 0$$

where $C_n \equiv [\sum_{i=1}^n \sigma_{n,i}^2]^{1/2}$. Then $\frac{[\sum_{i=1}^n (Y_{n,i} - \mathbb{E}Y_{n,i})]}{C_n} \xrightarrow{d} N(0, 1)$ ■

- The sampling framework is known as a "triangular array". Note that the "$(n-1)$-th observation in the n-th sample" ($Y_{n,n-1}$) need not coincide with $Y_{n-1,n-1}$, the "$(n-1)$-th observation in the $(n-1)$-th sample.
- C_n is just the standard deviation of $\sum_{i=1}^n Y_{n,i}$. Hence $\frac{[\sum_{i=1}^n (Y_{n,i} - \mathbb{E}Y_{n,i})]}{C_n}$ is a "standardized sum".

This is useful for showing asymptotic normality of Least-Squares regression. Let $y = \beta_0 X + \epsilon$ with ϵ i.i.d. with mean zero and variance σ^2, and x being an $n \times 1$ vector of covariates (here we have just 1 RHS variable).

The OLS estimator is $\hat{\beta} = (X'X)^{-1}X'Y$. To apply the LFCLT, we consider the normalized difference between the estimated and true β's:

$$((X'X)^{-1/2}/\sigma) \cdot (\hat{\beta} - \beta^0) = (1/\sigma) \cdot (X'X)^{-1/2} X'\epsilon \equiv (1/\sigma) \cdot \sum_{i=1}^{n} a_{n,i}\epsilon_i$$

where $a_{n,i}$ corresponds to the i-th component of the $1 \times n$ vector $(1/\sigma) \cdot (X'X)^{-1/2}X'$. So we just need to show that the Lindeberg condition applies to $Y_{n,i} = a_{n,i}\epsilon_i$. Note that $\sum_{i=1}^n V(a_{n,i}\epsilon_i) = V(\sum_{i=1}^n a_{n,i}\epsilon_i) = (1/\sigma^2) \cdot V((X'X)^{-1/2}X'\epsilon) = \sigma^2/\sigma^2 \cdot 1 = C_n^2$.

1.2 Convergence in distribution vs. a.s.

Combining two results above, we have

$$Z_n \overset{as}{\to} Z \Longrightarrow Z_n \overset{d}{\to} Z.$$

The converse is not generally true. (Indeed, $\{Z_1, Z_2, Z_3, \ldots\}$ need not be defined on the same probability space, in which case s.d. convergence makes no sense.)

However, consider

$$Z_n^* = F_{Z_n}^{-1}(U), \quad Z^* = F_Z^{-1}(U); \quad U \sim U[0,1].$$

Here $F_Z^{-1}(U)$ denotes the quantile function corresponding to the CDF $F(z)$:

$$F^{-1}(\tau) = \inf\{z : \ F(z) > \tau\} \quad \tau \in [0,1].$$

We have $F(F^{-1}(\tau)) = \tau$. (Note that quantile function is also right-continuous; discontinuity points of the quantile function arise where the CDF function is "flat".)

Then

$$P(Z_n^* \le z) = P(F_{Z_n}^{-1}(U) \le z) = P(U \le F_{Z_n}(z)) = F_{Z_n}(z)$$

so that $Z_n^* \stackrel{d}{=} Z_n$. (Even though their domains are different!) Similarly, $Z^* \stackrel{d}{=} Z$. The notation "$\stackrel{d}{=}$" means "identically distributed".

Moreover, it turns out (quite intuitive) that the convergence $F_{Z_n}^{-1}(U) \to F_Z^{-1}(U)$ fails only at points where $F_Z^{-1}(U)$ is discontinuous (corresponding to flat portions of $F_Z(z)$). Since these points of discontinuity are a countable set, their probability (under $U[0,1]$) is equal to zero, so that $F_{Z_n}^{-1}(U) \to F_Z^{-1}(U)$ for almost-all U.

So what we have here, is a result that for real-valued random variables $Z_n \stackrel{d}{\to} Z$, we can construct identically distributed variables such that both $Z_n^* \stackrel{d}{\to} Z^*$ and $Z_n^* \stackrel{as}{\to} Z^*$.

This is called the "Skorokhod construction".

Skorokhod representation: Let Z_n $(n = 1, 2, \ldots)$ be random elements defined on probability spaces $(\Omega_n, \mathbb{B}(\Omega_n), P_n)$ and $Z_n \stackrel{d}{\to} Z$, where X is defined on $(\Omega, \mathbb{B}(\Omega), P)$. Then there exist random variables Z_n^* $(n = 1, 2, \ldots)$ and Z^* defined on a common probability space $(\tilde{\Omega}, \mathbb{B}(\tilde{\Omega}), \tilde{P})$ such that

$$Z_n^* \stackrel{d}{=} Z_n \ (n = 1, 2, \ldots); \quad Z^* \stackrel{d}{=} Z; \quad Z_n^* \stackrel{as}{\to} Z^*.$$

Applications:

- Continuous mapping theorem: $Z_n \stackrel{d}{\to} Z \Rightarrow Z_n^* \stackrel{as}{\to} Z^*$; for $h(\cdot)$ continuous, we have $h(Z_n^*) \stackrel{as}{\to} h(Z^*) \Rightarrow h(Z_n^*) \stackrel{d}{\to} h(Z^*)$ which implies $h(Z_n) \stackrel{d}{\to} h(Z)$.

- Building on the above, if h is bounded, then we get

$$Eh(Z_n) = Eh(Z_n^*) \to Eh(Z^*) = Eh(Z)$$

under the bounded convergence theorem, which shows one direction of the "Portmanteau" theorem, Eq. (2).

1.3 Functional Central limit theorems

There are a set of distributional convergence results, which are known as functional CLT's (or Donsker theorems), because they deal with convergence of random *functions* (or, interchangeably, *processes*). These are indispensible tools in finance.

One of the simplest random functions is the **Wiener** process $\mathcal{W}(t)$, which is viewed as a random function on the unit interval $t \in [0, 1]$. This is also known as a Brownian motion process.

Features of the Wiener process:

1. $\mathcal{W}(0) = 0$
2. Gaussian marginals: $\mathcal{W}(t) \stackrel{d}{=} N(0, t)$; that is,

$$P(\mathcal{W}(t) \leq a) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{a} \exp\left(\frac{-u^2}{2t}\right) du.$$

3. Independent increments: Define $x_t \equiv \mathcal{W}(t)$. For any set $0 \leq t_0 \leq t_1 \leq \ldots \leq 1$, the differences

$$x_{t_1} - x_{t_0}, x_{t_2} - x_{t_1}, \ldots$$

are independent.

4. Given the two above features, we have that the increments are themselves normally distributed:

$$x_{t_i} - x_{t_{i-1}} \stackrel{d}{=} N(0, t_i - t_{i-1}).$$

Moreover, from

$$\begin{aligned} t_1 - t_0 = V(x_1 - x_0) &= E[(x_1 - x_0)^2] \\ &= Ex_1^2 - 2Ex_1x_0 + Ex_0^2 \\ &= t_1 + t_0 - 2Ex_1x_0 \end{aligned}$$

implying

$$Ex_1x_0 = Cov(x_1, x_0) = t_0.$$

5. Furthermore, we know that any finite collection $(x_{t_1}, x_{t_2}, \ldots)$ is jointly distributed multivariate normal, with mean 0 and variance matrix Σ as described above.

Take the conditions of the LLCLT: $X_1, X_2, \ldots$ are iid with mean zero and finite variance σ^2. For a given n, define the "partial sum" $S_k = X_1 + X_2 + \cdots + X_k$ for $k \leq n$. Now, for $t \in [0, 1]$, define the normalized "partial sum process"

$$\tilde{S}_n(t) = \frac{S_{[tn]}}{\sigma\sqrt{n}} + (tn - [tn]) * \frac{1}{\sigma\sqrt{n}} X_{[tn]+1}$$

where $[a]$ denotes the largest integer $\leq a$. $\tilde{S}_n(t)$ is a random function on $[0, 1]$. (Graph)

Then the functional CLT is the following result:

Functional CLT: $\tilde{S}_n \xrightarrow{d} \mathcal{W}$.

Note that $\tilde{S}_n$ is a random element taking values in $C[0, 1]$, the space of continuous functions on [0,1]. Proving this result is beyond our focus in this class.[4] ■

Note that the functional CLT implies more than the LLCLT. For example, take $t = 1$. Then $\tilde{S}_n(1) = \frac{\sum_{i=1}^n X_i}{\sigma\sqrt{n}} = \frac{\sqrt{n}\bar{X}_n}{\sigma}$ By the FCLT, we have

$$\begin{aligned} P(\tilde{S}_n(1) \leq a) &\rightarrow P(\mathcal{W}(1) \leq a) \\ &= P(N(0, 1) \leq a) \end{aligned}$$

which is the same result as the LLCLT.

Similarly, for $k_n < n$ but $\frac{k_n}{n} \rightarrow \tau \in [0, 1]$, we have

$$\tilde{S}_n\left(\frac{k}{n}\right) = \frac{\sum_{i=1}^k X_i}{\sigma\sqrt{n}} = \frac{k \cdot \bar{X}_k}{\sigma\sqrt{n}} \xrightarrow{d} N(0, \tau).$$

Also, it turns out, by a functional version of the continuous mapping theorem, we can get convergence results for functionals of the partial-sum process. For instance, $\sup_t \tilde{S}_n(t) \xrightarrow{d} \sup_t \mathcal{W}(t)$, and it turns out

$$P(\sup_t \mathcal{W}(t) \leq a) = \frac{2}{\sqrt{2\pi}} \int_0^a \exp(-u^2/2)\; du; \quad a \geq 0.$$

(Recall that $\mathcal{W}_0 = 0$, so $a < 0$ makes no sense.)

[4]See, for instance, Billingsley, *Convergence of Probability Measures*, ch. 2, section 10.