Lecture 3: Asymptotic Distributions

Remark 3.1 Recall deriving χ^2, t, and F distributions. The assumption in all of them was the random sample $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$. What happens when the random sample is **no longer** Gaussian?

Example 3.1 Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \text{Ber}(p)$. We can calculate $\bar{X}$ and denote it as the sample proportion. The distribution of $\bar{X}$ looks similar to a Gaussian distribution as the sample size n grows as the figure below. How can we formalize that as n grows, this distribution *converges* to some other distribution?

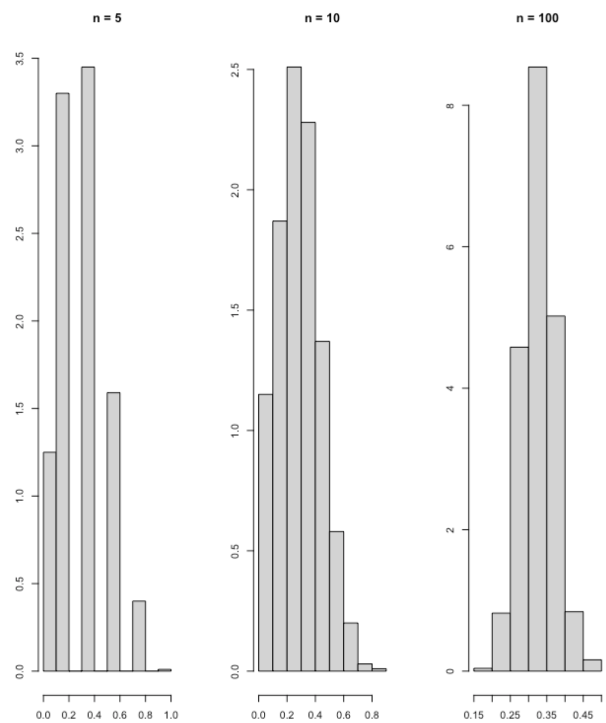

Figure 1: The distribution of $\bar{X}_n$ for $n = 5, 10, 100$, when $p = 1/3$ (repeated 100 times).

3.1 Convergence in probability

Definition 3.1 Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables and let X be a random variable defined on a sample space. We say that X_n **converges in probability** to X if, for all $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0,$$

or equivalently, $\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$. If so, we write $X_n \xrightarrow{P} X$.

$\boxed{\textbf{Example 3.2}}$ $X_n \sim \text{Unif}(1 - \frac{1}{n}, 1)$. Show that $X_n \xrightarrow{P} 1$ by definition.

Proof. Let $\epsilon > 0$, then

$$\begin{aligned}\lim_{n\to\infty} P(|X_n - 1| \geq \epsilon) &= \lim_{n\to\infty} P(1 - X_n \geq \epsilon) \quad \text{(we know } X_n \leq 1, \forall n\text{)} \\ &= \lim_{n\to\infty} P(X_n \leq 1 - \epsilon)\end{aligned}$$

We then are presented two cases: (1) $\epsilon \geq 1/n$, (2) $0 < \epsilon < 1/n$. In terms of case (2), we need not be concerned with it because are taking the limit as $n \to \infty$ and for all $n > \frac{1}{\epsilon}$, case (2) no longer holds. Hence, for sufficiently large n, we are in case (1) and

$$P\left(X_n \leq 1 - \frac{1}{n}\right) = 0 \implies P(X_n \leq 1 - \epsilon) = 0 \implies \lim_{n\to\infty} P(X_n \leq 1 - \epsilon) = 0$$

because $1 - \frac{1}{n}$ is the lower bound of support for X_n. □

$\boxed{\textbf{Theorem 3.1}}$ ***(Markov's Inequality)*** *If ϕ is a nondecreasing nonnegative function and $\phi(a) > 0$, then for a random variable, X,*

$$P(X \geq a) \leq \frac{\mathbb{E}(\phi(X))}{\phi(a)}.$$

Alternatively (a corollary), for a random variable Z, if $\mathbb{E}(|Z|^r) < \infty$ (where $r > 0$), then for any $\epsilon > 0$,

$$P(|Z| \geq \epsilon) \leq \frac{\mathbb{E}(|Z|^r)}{\epsilon^r}.$$

Proof. (Second part) Let $\epsilon > 0$. The key idea of this proof is expressing the probability as the expected value of an indicator function, which is given by

$$P(|Z| \geq \epsilon) = \mathbb{E}(\mathbf{1}\{|Z| \geq \epsilon\}).$$

The value of the indicator function is either 0 or 1, so

$$\mathbf{1}\{|Z| \geq \epsilon\} = 1 \cdot \mathbf{1}\{|Z|/\epsilon \geq 1\} \leq \left(\frac{|Z|}{\epsilon}\right)^r \cdot \mathbf{1}\{|Z|/\epsilon \geq 1\} \leq \frac{|Z|^r}{\epsilon^r}.$$

Taking the expectation on both sides, we obtain $P(|Z| \geq \epsilon) \leq \frac{\mathbb{E}(|Z|^r)}{\epsilon^r}$. □

$\boxed{\textbf{Theorem 3.2}}$ ***(Chebyshev's Inequality)*** *For a random variable X, if $\text{Var}(X) < \infty$, then for any $\epsilon > 0$,*

$$P(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}.$$

(In words, we can interpret this as "if the variance is small, then the random variable is centered around its mean.")

Proof. From Markov's inequality, by substituting $Z = X - \mathbb{E}(X)$ and $r = 2$, we can derive Chebyshev's inequality by

$$P(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}.$$

□

Theorem 3.3 *Suppose that a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ satisfies*

$$\lim_{n\to\infty} \text{Var}(X_n) = 0, \quad \lim_{n\to\infty} \mathbb{E}(X_n) = a.$$

Then $X_n \xrightarrow{P} a$.

Proof. By Markov's inequality, for any $\epsilon > 0$, we have

$$P(|X_n - a| \geq \epsilon) \leq \frac{\mathbb{E}[(X_n - a)^2]}{\epsilon^2}.$$

Recall $\text{Var}(X_n) = \mathbb{E}(X_n^2) - \mathbb{E}(X_n)^2$. From this, we know

$$\mathbb{E}[(X_n - a)^2] = \mathbb{E}(X_n^2) - 2a\mathbb{E}(X_n) + a^2 = \text{Var}(X_n) + (\mathbb{E}(X_n) - a)^2.$$

Thus, if $\lim_{n\to\infty} \text{Var}(X_n) = 0$ and $\lim_{n\to\infty} \mathbb{E}(X_n) = a$,

$$0 \leq \lim_{n\to\infty} P(|X_n - a| \geq \epsilon) \leq \lim_{n\to\infty} \frac{\mathbb{E}[(X_n - a)^2]}{\epsilon^2} = \lim_{n\to\infty} \frac{1}{\epsilon^2}\left[\text{Var}(X_n) + (\mathbb{E}(X_n) - a)^2\right] = 0.$$

[*Additional Exercise: Prove Example 3.2 using Theorem 3.3.*] □

Theorem 3.4 ***(Weak Law of Large Numbers)*** *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of i.i.d. random variables having common mean μ and variance $\sigma^2 < \infty$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then $\overline{X}_n \xrightarrow{P} \mu$.*

Proof. The mean and variance of $\overline{X}_n$ are μ and σ^2/n, respectively. By Chebyshev's, we have for every $\epsilon > 0$,

$$P(|\overline{X}_n - \mathbb{E}(\overline{X}_n)| \geq \epsilon) = P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Taking the limit, we know $\lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$ therefore $\overline{X}_n \xrightarrow{P} \mu$. Alternatively, we can apply Theorem 3.3, because we know $\mathbb{E}(\bar{X}_n) = \mu$ and $\lim_{n\to\infty} \text{Var}(\bar{X}_n) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$. □

Remark 3.2 One should **not** assume the existence of a mean (the first moment) as a given. There exist some random variables which do not have a first moment, and some random variables do not have a second moment (and variance). Take, for example, the Cauchy distribution. Let $X \sim \text{Cauchy}(0,1)$, then the probability density function is given by

$$f(x) = \frac{1}{\pi(x^2+1)} \cdot \mathbf{1}\{x \in (-\infty, \infty)\}$$

The mean

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x}{x^2+1} dx = \frac{1}{\pi}\left[\int_{-\infty}^{\infty} 1\, dx - \int_{-\infty}^{\infty} \frac{1}{1+x^2} dx\right] = \frac{1}{\pi}\left[\int_{-\infty}^{\infty} dx - \pi\right]$$

cannot be defined, and moreover, $E(|X|) = \infty$. Likewise, one can show that the variance cannot be defined (no second moment).

Theorem 3.5 ***(Continuous mapping theorem)***

(1) If $X_n \xrightarrow{P} X$ and f is a continuous function, then $f(X_n) \xrightarrow{P} f(X)$.
(2) If $X_n \xrightarrow{P} c$ where c is a constant and a function f is continuous at c, then $f(X_n) \xrightarrow{P} f(c)$.

Proof. See Theorem 5.1.4 in Hogg & Craig. □

Theorem 3.6 ***(Continuous mapping theorem multi-dimension)*** *For the sequence of random vectors $\boldsymbol{X_n} = (X_{n1}, \ldots, X_{nk})$ $(n = 1, 2, \ldots)$, and vector $\boldsymbol{c} = (c_1, \ldots, c_k)$, suppose we have $\boldsymbol{X_n} \xrightarrow{P} \boldsymbol{c}$. If a function g is continuous at $\boldsymbol{c}$, we have $g(\boldsymbol{X_n}) \xrightarrow{P} g(\boldsymbol{c})$.*

Proof. Since the function g is continuous at $\boldsymbol{c}$, the following holds:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } \|\boldsymbol{x} - \boldsymbol{c}\| < \delta \implies \|g(\boldsymbol{x}) - g(\boldsymbol{c})\| < \epsilon.$$

Therefore, for any $\epsilon > 0$, we can select such a δ so that we have the following relationship of probabilistic events

$$\{\|g(\boldsymbol{X_n}) - g(\boldsymbol{c})\| \geq \epsilon\} \subseteq \{\|\boldsymbol{X_n} - \boldsymbol{c}\| \geq \delta\}.$$

(*For some intuition, think about the graph $y = x^2$ and choose $c > 0$. Then imagine the set of y's that are more than ϵ deviated from c^2 compared to the set of x's that are more than δ deviated from c.*) Using the above set-relation, we know

$$P(\|g(\boldsymbol{X_n}) - g(\boldsymbol{c})\| \geq \epsilon) \leq P(\|\boldsymbol{X_n} - \boldsymbol{c}\| \geq \delta).$$

Since $\boldsymbol{X_n} \xrightarrow{P} \boldsymbol{c}$, we know $0 \leq \lim_{n\to\infty} P(\|g(\boldsymbol{X_n}) - g(\boldsymbol{c})\| \geq \epsilon) \leq \lim_{n\to\infty} P(\|\boldsymbol{X_n} - \boldsymbol{c}\| \geq \delta) = 0$. Thus, $g(\boldsymbol{X_n}) \xrightarrow{P} g(\boldsymbol{c})$. □

Corollary 3.1 *Suppose $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$ with a constant c. Then, the following holds.*

(1) $X_n + Y_n \xrightarrow{P} X + Y$.
(2) $cX_n \xrightarrow{P} cX$.
(3) $X_nY_n \xrightarrow{P} XY$.

Proof. Take (1) $f(x,y) = x+y$, (2) $f(x) = cx$, and (3) $f(x,y) = xy$ and apply Theorem 3.6. $\square$

Example 3.3 Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of i.i.d. random variables having mean μ and variance $\sigma^2 < \infty$. Suppose we have $\mathbb{E}(X_1^4) < \infty$ (ensures $\mathrm{Var}(S_n^2) < \infty$). Show $S_n^2 \xrightarrow{P} \sigma^2$.

Proof. Recall that through some algebra, we get the following identity for S_n^2.

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \frac{1}{n-1}\left(\left[\sum_{i=1}^{n} X_i^2\right] - n\overline{X}_n^2\right) = \frac{n}{n-1}\left(\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] - \overline{X}_n^2\right)$$

By the WLLN (Theorem 3.4), we know that $\bar{X}_n \xrightarrow{P} \mu$. Then, by the Continuous Mapping Theorem 3.5, we know $\bar{X}_n^2 \xrightarrow{P} \mu^2$. Next, we need to find what $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ converges to in probability. Notice that the WLLN applies for $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ because $\{X_n^2\}_{n=1}^{\infty}$ are i.i.d. because we are given that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. Therefore, $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \mathbb{E}(X_1^2)$. By Corollary 3.1,

$$\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] - \overline{X}_n^2 \xrightarrow{P} \mathbb{E}(X_1^2) - \mu^2 = \mathrm{Var}(X_1) = \sigma^2.$$

The proof is nearly done, but we need to apply the following lemma

> *Lemma:* If $X_n \xrightarrow{P} X$, and $a_n \xrightarrow{n\to\infty} a$ is a deterministic (non-random) sequence, then $a_nX_n \xrightarrow{P} aX$.

Using this lemma and noting that $\frac{n}{n-1} \to 1$ as $n \to \infty$, then we can conclude that

$$S_n^2 = \frac{n}{n-1}\left(\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] - \overline{X}_n^2\right) \xrightarrow{P} 1 \cdot \left(\mathbb{E}(X_1^2) - \mu^2\right) = \sigma^2.$$

$\square$

Remark 3.3 If we use the sample variance S_n^2 as the estimator of the population variance σ^2, we call S_n^2 a **consistent** estimator of σ^2 because $S_n^2 \xrightarrow{P} \sigma^2$. Consistency is considered an essential property of estimators because they get “closer” to the parameters when we have reasonably large sample size. We will discuss more on the **consistency** of an estimator later, but note that this concept is different from **unbiasedness** of an estimator (e.g., $\mathbb{E}(S_n^2) = \sigma^2$ for any n). This distinction is important because the estimators S_n^2 and S_n are referred to as the unbiased and biased estimators of σ^2 and σ, respectively, ($\mathbb{E}(S_n) \neq \sigma$ in general) whereas they are both consistent.

3.2 Convergence in distribution (weak convergence)

Remark 3.4 Convergence in probability indicates that as the sample size increases, the value of a statistic approaches the value of the parameter even without needing to know the distribution function of the statistic. Our next question is how close the statistic is to the distribution of the estimator? Moreover, can we obtain the margin of error of estimation? Convergence in distribution is a useful concept for these purposes. One representative example is the central limit theorem (CLT): if a random sample $X_1, \ldots, X_n$ with mean μ and variance σ^2, the distribution of $\sqrt{n}(\bar{X}_n - \mu)$ will be approximated by $\mathcal{N}(0, \sigma^2)$ given a sufficient large n. See Figure 1.

Definition 3.2 Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables and let X be a random variable. Let F_{X_n} and F_X be the cdfs of X_n and X, respectively. Let $C(F_X)$ denote the set of all points where F_X is continuous. We say that X_n *converges in distribution* to X (i.e., $X_n \xrightarrow{D} X$) if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \text{for all } x \in C(F_X).$$

Remark 3.5

(1) For statisticians and probabilists, convergence in probability and in distribution are in the field of *asymptotic theory*.

(2) For terminology, we say "the distribution of X is the **asymptotic distribution or the limiting distribution** of the sequence $\{X_n\}$" and refers to the CDF.

(3) Common notation: instead of saying $X_n \xrightarrow{D} X$, where X has a standard normal distribution, we may write $X_n \xrightarrow{D} \mathcal{N}(0,1)$ (this is an abuse of notation, but still recognized).

Example 3.4 Let a random variable X_n follow a binomial distribution Binom$(n, \lambda/n)$ (where $0 < \lambda/n < 1$) and a random variable $X \sim$ Poisson(λ). We can calculate

$$\begin{aligned}
P(X_n = k) &= \binom{n}{k}\left[\frac{\lambda}{n}\right]^k \left[1 - \frac{\lambda}{n}\right]^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{\lambda^k}{k!}\left[1 - \frac{\lambda}{n}\right]^{n-k} \\
&\xrightarrow{n\to\infty} \frac{\lambda^k}{k!} \cdot e^{-\lambda} \\
&= P(X = k), \quad k = 1, 2, \ldots.
\end{aligned}$$

Since these random variables take on integer values, we have the cumulative distribution function (cdf) at any point x as:

$$\text{cdf}_{X_n}(x) = \sum_{k:0\le k\le x} P(X_n = k) \xrightarrow{n\to\infty} \sum_{k:0\le k\le x} P(X = k) = \text{cdf}_X(x).$$

This is why we say that a Poisson distribution is the **limiting distribution of a binomial distribution**; Binom$(n, \lambda/n) \approx$ Poisson(λ), $n \to \infty$.

Theorem 3.7

(a) If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{D} X$. The converse is not always true.
(b) If $X_n \xrightarrow{D} b$, where b is some constant, then $X_n \xrightarrow{P} b$.
(c) If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} 0$, then $X_n + Y_n \xrightarrow{D} X$.

Proof. (b) Let $\epsilon > 0$. Using the alternative version of showing convergence in probability, we have

$$\begin{aligned}\lim_{n\to\infty} P(|X_n - b| \le \epsilon) &= \lim_{n\to\infty} P(-\epsilon \le X_n - b \le \epsilon) \\ &= \lim_{n\to\infty} P(b-\epsilon \le X_n \le b+\epsilon) \\ &= \lim_{n\to\infty} F_{X_n}(b+\epsilon) - F_{X_n}(b-\epsilon) \\ &= \lim_{n\to\infty} F_{X_n}(b+\epsilon) - \lim_{n\to\infty} F_{X_n}(b-\epsilon)\end{aligned}$$

Recall that $X_n \xrightarrow{D} b$ means

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) = \begin{cases} 0, & x < b \\ 1, & x \ge b \end{cases}.$$

Therefore,

$$\lim_{n\to\infty} P(|X_n - b| \le \epsilon) = \lim_{n\to\infty} F_{X_n}(b+\epsilon) - \lim_{n\to\infty} F_{X_n}(b-\epsilon) = 1 - 0 = 1$$

[Proofs of (a) and (c) can be found in Hogg & Craig Theorems 5.2.1 and 5.2.3] $\square$

Remark 3.6 As Theorem 3.7 shows, convergence in distribution is **weaker** than convergence in probability. Thus convergence in distribution is often called *weak convergence*.

Example 3.5 Convergence in probability is a way of saying that a sequence of random variables X_n is getting close to another random variable X. On the other hand, convergence in distribution is only concerned with the cdfs F_{X_n} and F_X. Here is a simple example of convergence in distribution but **not** in probability.

Let X be a continuous random variable with a pdf $f_X(x)$ that is symmetric about 0 (i.e., $f_X(-x) = f_X(x)$). Then the density of the random variable $-X$ is also $f_X(x)$. Thus, X and $-X$ have the same distributions. Define the sequence of random variables X_n as

$$X_n = X \cdot \mathbf{1}\{\text{n is odd}\} - X \cdot \mathbf{1}\{\text{n is even}\}.$$

Clearly, $F_{X_n}(x) = F_X(x)$ for all x in the support of X, so that $X_n \xrightarrow{D} X$. On the other hand, the sequence X_n does not get close to X. In particular, $X_n \not\to X$ in probability.

Theorem 3.8 *Let $\{X_n\}$ be a sequence of random variables with mgf $M_{X_n}(t)$ that exists for $-h < t < h$ for all n. Let X be a random variable with mgf $M(t)$, which exists for $|t| \le h_1 \le h$. If $\lim_{n\to\infty} M_{X_n}(t) = M(t)$ for $|t| \le h_1$, then $X_n \xrightarrow{D} X$.*

Proof. Omitted □

Lemma 3.1 *If $\lim_{n\to\infty} \psi(n) = 0$, then*

$$\lim_{n\to\infty} \left[1 + \frac{b}{n} + \frac{\psi(n)}{n}\right]^{cn} = \lim_{n\to\infty} \left[1 + \frac{b}{n}\right]^{cn} = e^{bc}.$$

Proof. Omitted (Note: use Theorem 3.8 and this Lemma to reprove Example 3.4). □

Theorem 3.9 ***(Central Limit Theorem)*** *Suppose a random sample $X_1, \dots, X_n$ are iid with mean and variance μ and σ^2 $(0 < \sigma < \infty)$, respectively, and let a random variable $Z \sim \mathcal{N}(0,1)$. Then, we have $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{D} \mathcal{N}(0,1)$, that is*

$$\lim_{n\to\infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le x\right) = P(Z \le x) \quad \textit{for all} \quad x : -\infty < x < \infty.$$

Proof. This theorem holds generally under the given conditions. However, we limit the proof to cases where the moment generating function of X_1 exists as an additional condition. The key idea for this proof is that the moment generating function of the standardized sample mean

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} = \frac{(X_1 + \cdots + X_n)/n - \mu}{\sigma/\sqrt{n}}$$

converges to the moment generating function of Z, where $\mathrm{mgf}_Z(t) = \exp\big(t^2/2\big)$. In other words, we need to apply Theorem 3.8 and show:

$$\lim_{n\to\infty} \mathrm{mgf}_{\sqrt{n}(\overline{X}_n - \mu)/\sigma}(t) = \mathrm{mgf}_Z(t) = \exp\big(t^2/2\big).$$

From the definition of the mgf and the independence of the random variables, we have

$$\begin{aligned}
\mathrm{mgf}_{\sqrt{n}(\overline{X}_n-\mu)/\sigma}(t) &= \mathbb{E}\left[\exp\left(t\sqrt{n}\frac{\overline{X}_n-\mu}{\sigma}\right)\right] \\
&= \mathbb{E}\left[\exp\left(\frac{t}{\sqrt{n}}\frac{X_1-\mu}{\sigma}+\cdots+\frac{t}{\sqrt{n}}\frac{X_n-\mu}{\sigma}\right)\right] \\
&= \left\{\mathbb{E}\left[\exp\left(\frac{t}{\sqrt{n}}\frac{X_1-\mu}{\sigma}\right)\right]\right\}^n \\
&= \left[\mathrm{mgf}_{(X_1-\mu)/\sigma}\left(\frac{t}{\sqrt{n}}\right)\right]^n
\end{aligned}$$

Let $m(t)=\mathrm{mgf}_{(X_1-\mu)/\sigma}(t)$. Then by the properties of the moment generating function and using the Taylor series expansion, we have

$$\begin{aligned}
m(t) &= m(0)+m'(0)t+\frac{m''(0)}{2!}t^2+R_t \\
&= 1+\mathbb{E}\left[\frac{X_1-\mu}{\sigma}\right]t+\mathbb{E}\left[\left(\frac{X_1-\mu}{\sigma}\right)^2\right]\frac{t^2}{2}+R_t \\
&= 1+\frac{t^2}{2}+R_t,
\end{aligned}$$

where, according to Taylor's theorem,

$$R_t=\frac{m''(\xi_t)-m''(0)}{2}\cdot t^2, \quad \xi_t\in(0,t).$$

With this, we can rewrite $m(t)$ as $m(t)=1+\frac{m''(\xi_t)}{2}t^2$. Then, replace t with $t/\sqrt{n}$, and add and subtract $t^2/2n$ (keep in mind Lemma 3.1)

$$\begin{aligned}
m\left(\frac{t}{\sqrt{n}}\right) &= 1+\frac{m''(\xi_t^*)t^2}{2n}, \quad \xi_t^*\in(0,t/\sqrt{n}) \\
&= 1+\frac{t^2}{2n}+\frac{(m''(\xi_t^*)-1)t^2}{2n}.
\end{aligned}$$

Putting everything together, we have

$$\begin{aligned}
\lim_{n\to\infty}\mathrm{mgf}_{\sqrt{n}(\overline{X}_n-\mu)/\sigma}(t) &= \lim_{n\to\infty}\left[\mathrm{mgf}_{(X_1-\mu)/\sigma}\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \lim_{n\to\infty}\left[1+\frac{t^2}{2n}+\frac{(m''(\xi_t^*)-1)t^2}{2n}\right]^n \\
&= \lim_{n\to\infty}\left[1+\frac{t^2}{2n}\right]^n \\
&= \exp\big(t^2/2\big).
\end{aligned}$$

Therefore, by Theorem 3.8, we have $X_n\xrightarrow{D}X$. (See Examples 5.3.3 and 5.3.4 in Hogg & Craig) □

$\boxed{\textbf{Theorem 3.10}}$ ***(Continuous mapping theorem)*** *Suppose X_n converges to X in distribution and g is a continuous function on the support of X. Then $g(X_n)$ converges to $g(X)$ in distribution.*

Proof. Omitted □

$\boxed{\textbf{Remark 3.7}}$ An often-used application of this theorem occurs when we have a sequence of random variables Z_n which converges in distribution to a standard normal random variable Z. Because the distribution of Z^2 is $\chi^2(1)$, it follows by Theorem 3.10 that Z_n^2 converges in distribution to a $\chi^2(1)$ distribution.

$\boxed{\textbf{Theorem 3.11}}$ ***(Slutsky's Theorem)*** *For random variables X_n, Y_n (for $n = 1, 2, \dots$), Z, and a real constant c, if $X_n \xrightarrow{D} Z$ and $Y_n \xrightarrow{P} c$, then the following hold:*

(a) $X_n + Y_n \xrightarrow{D} Z + c$,
(b) $X_n - Y_n \xrightarrow{D} Z - c$,
(c) $Y_n X_n \xrightarrow{D} cZ$,
(d) $X_n / Y_n \xrightarrow{D} Z/c \quad (c \neq 0)$.

Proof. This theorem holds in general, but here we will prove the case where the cumulative distribution function of Z is continuous.

(a) For any $\epsilon > 0$,

$$\begin{aligned} P(X_n + Y_n \le z) &= P(X_n + Y_n \le z, |Y_n - c| < \epsilon) + P(X_n + Y_n \le z, |Y_n - c| \ge \epsilon) \\ &\le P(X_n \le z - c + \epsilon) + P(|Y_n - c| \ge \epsilon). \end{aligned}$$

By the condition, $\lim_{n\to\infty} P(|Y_n - c| \ge \epsilon) = 0$, we have

$$\lim_{n\to\infty} P(X_n + Y_n \le z) \le P(Z \le z - c + \epsilon) \underset{\epsilon\downarrow 0}{\to} P(Z + c \le z),$$

as we tend $\epsilon \to 0$. Similarly, by symmetry, we also have

$$\begin{aligned} P(X_n + Y_n > z) &= P(X_n + Y_n > z, |Y_n - c| < \epsilon) + P(X_n + Y_n > z, |Y_n - c| \ge \epsilon) \\ &\le P(X_n > z - c - \epsilon) + P(|Y_n - c| \ge \epsilon). \end{aligned}$$

Again, by the condition, we have

$$\lim_{n\to\infty} P(X_n + Y_n > z) \le P(Z > z - c - \epsilon) \underset{\epsilon\downarrow 0}{\to} P(Z + c > z),$$

so $\lim_{n\to\infty} P(X_n + Y_n \le z) \ge P(Z + c \le z)$. Hence, under the assumption that the cumulative distribution function of Z is continuous,

$$\lim_{n\to\infty} P(X_n + Y_n \le z) = P(Z + c \le z), \quad \text{i.e., } X_n + Y_n \xrightarrow{d} Z + c$$

(b) If $Y_n \xrightarrow{P} c$ and $-Y_n \xrightarrow{P} -c$, immediate from (a).

(c) First, consider the case when $c = 0$. For any $\epsilon > 0$,

$$P(|Y_n X_n| \ge \epsilon) = P(|Y_n X_n| \ge \epsilon, |X_n| \ge k) + P(|Y_n X_n| \ge \epsilon, |X_n| < k)$$
$$\le P(|X_n| \ge k) + P(|Y_n| \ge \frac{\epsilon}{k}), \forall k > 0.$$

We can conclude, under the assumption of the continuity of the cumulative distribution function of Z,

$$0 \le \lim_{n\to\infty} P(|Y_n X_n| \ge \epsilon) \le P(Z > k) + P(Z \le -k) \to 0 \text{ as } k \to \infty.$$

Next, consider the general case of c. In this case, the given conditions are

$$Y_n - c \xrightarrow{P} 0, \quad X_n \xrightarrow{D} Z$$

Thus, if the conditions from the proof of the first case are satisfied, we have

$$(Y_n - c)X_n \xrightarrow{P} 0.$$

Also, under additional conditions on the cumulative distribution function of Z, it is clear that $cX_n \xrightarrow{D} cZ$. Therefore, from (a), we have

$$Y_n X_n = (Y_n - c)X_n + cX_n \xrightarrow{D} 0 + cZ = cZ.$$

(d) Using Theorem 3.6 with (c), (d) also holds.

□

Example 3.6 Let μ be the population mean and $0 < \sigma < \infty$ be the population standard deviation. Consider a random sample of size n from the population with sample mean $\overline{X}_n$ and sample standard deviation S_n. The *studentized sample mean* is defined as:

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$$

According to Theorem 3.9 (CLT) and Example 3.3, we have,

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z, \quad Z \sim N(0,1) \quad \text{and} \quad S_n \xrightarrow{P} \sigma.$$

Therefore, by Slutsky's theorem:

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{S_n} \xrightarrow{D} Z \cdot \frac{\sigma}{\sigma} = Z, \quad Z \sim N(0,1)$$

Specifically, if the population follows a normal distribution, the studentized sample mean follows a t-distribution with $n-1$ degrees of freedom. This implies that as n increases, the t-distribution approaches the standard normal distribution,

$$t(n-1) \xrightarrow{n\to\infty} N(0,1)$$

$\boxed{\textbf{Example 3.7}}$ (*Asymptotic distribution of sample variance*) Let μ be the population mean and $0 < \sigma < \infty$ be the population standard deviation. Consider a random sample of size n from the population with the sample variance S_n^2 and sample mean $\overline{X}_n$. The sample variance is defined as:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[(X_i - \mu) - (\overline{X}_n - \mu)\right]^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - (\overline{X}_n - \mu)^2\right].$$

Let $Y_i = (X_i - \mu)^2$ (for $i = 1, \ldots, n$) and assume these variables are independent and identically distributed. According to the CLT, when $\mathrm{Var}(Y_1) < \infty$, we have:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i - \mathbb{E}(Y_1)\right) \xrightarrow{D} W, \quad W \sim N(0, \mathrm{Var}(Y_1))$$

Given $\mathbb{E}(Y_1) = \sigma^2$ and $\mathrm{Var}(Y_1) = \mathbb{E}(Y_1^2) - (\mathbb{E}Y_1)^2 = \mathbb{E}[(X_1-\mu)^4] - \sigma^4$, we know that:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \sigma^2\right) \xrightarrow{D} W, \quad W \sim N(0, \mathbb{E}[(X_1-\mu)^4] - \sigma^4)$$

Also, by the CLT and WLLN:

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} Z, \quad Z \sim N(0, \sigma^2), \quad \text{and} \quad (\overline{X}_n - \mu) \xrightarrow{P} 0$$

Therefore, by Slutsky's theorem

$$\sqrt{n}(\overline{X}_n - \mu)(\overline{X}_n - \mu) \xrightarrow{D} 0 \times Z = 0,$$

which means $\sqrt{n}(\overline{X}_n - \mu)^2 \xrightarrow{P} 0$. Finally, when $\mathbb{E}[(X_1-\mu)^4] < \infty$,

$$
\begin{aligned}
\sqrt{n}\left(S_n^2-\sigma^2\right) &= \sqrt{n}\cdot\left\{\frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-(\overline{X}_n-\mu)^2\right]-\sigma^2\right\} \\
&= \sqrt{n}\left(\frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right]-\sigma^2\right)-\sqrt{n}\frac{n}{n-1}(\overline{X}_n-\mu)^2 \\
&= \sqrt{n}\left(\frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right]-\frac{n-1}{n-1}\sigma^2\right)-\sqrt{n}\frac{n}{n-1}(\overline{X}_n-\mu)^2 \\
&= \sqrt{n}\left(\frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-\sigma^2\right]+\frac{1}{n-1}\sigma^2\right)-\sqrt{n}\frac{n}{n-1}(\overline{X}_n-\mu)^2 \\
&= \frac{n}{n-1}\cdot\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-\sigma^2\right]+\frac{\sqrt{n}}{n-1}\sigma^2-\frac{n}{n-1}\cdot\sqrt{n}(\overline{X}_n-\mu)^2 \\
&\xrightarrow{D} W, \quad W\sim N(0,(\rho_4+2)\sigma^4),
\end{aligned}
$$

by Slutsky's theorem, where $\rho_4=\mathbb{E}\left[\left(\frac{X_1-\mu}{\sigma}\right)^4\right]-3$ (ρ_4 is the excess kurtosis).

Theorem 3.12 ***(Delta Method)***

- *(Univariate) Let $\{X_n\}$ be sequence of random variables where*

  $$\sqrt{n}(X_n-\theta)\xrightarrow{D}\mathcal{N}(0,\sigma^2).$$

  Suppose the function $g(x)$ is differentiable at some θ and $g'(\theta)\neq 0$. Then

  $$\sqrt{n}(g(X_n)-g(\theta))\xrightarrow{D}\mathcal{N}(0,\sigma^2\cdot(g'(\theta))^2).$$

- *(Multivariate) Let $\{\boldsymbol{X}_n\}$ be a k-dimensional sequence of random variables. Suppose*

  $$\sqrt{n}(\boldsymbol{X}_n-\boldsymbol{\theta})\xrightarrow{D}\mathcal{N}_k(\mathbf{0},\boldsymbol{\Sigma}) \quad \textit{as} \quad n\to\infty.$$

  Let g be some p-dimensional transformation such that $g(\boldsymbol{x})=(g_1(\boldsymbol{x}),\ldots,g_p(\boldsymbol{x}))^T$ where $1\leq p\leq k$. If the matrix of partial derivatives

  $$\boldsymbol{B}=\left[\frac{\partial g_i}{\partial x_j}\right], \quad i=1,\ldots,p; \quad j=1,\ldots,k$$

  are continuous and not 0 in a neighborhood of $\boldsymbol{\theta}$, then

  $$\sqrt{n}(g(\boldsymbol{X}_n)-g(\boldsymbol{\theta}))\xrightarrow{D}\mathcal{N}_p(\mathbf{0},\ \boldsymbol{B}(\boldsymbol{\theta})\cdot\boldsymbol{\Sigma}\cdot\boldsymbol{B}(\boldsymbol{\theta})^T).$$

Proof. See Theorem 5.2.9 of Hogg & Craig for univariate proof. Next, see Example 5.3.7. □