Lecture 19: Convergence

Asymptotic approach

- In statistical analysis or inference, a key to the success of finding a good procedure is being able to find some moments and/or distributions of various statistics.
- In many complicated problems we are not able to find exactly the moments or distributions of given statistics.
- When the sample size n is large, we may approximate the moments and distributions of statistics, using asymptotic tools, some of which are studied in this course.
- In an asymptotic analysis, we consider a sample $X = (X_1, ..., X_n)$ not for fixed n, but as a member of a sequence corresponding to $n = n_0, n_0 + 1, ...$, and obtain the limit of the distribution of an appropriately normalized statistic or variable $T_n(X)$ as $n \to \infty$.
- The limiting distribution and its moments are used as approximations to the distribution and moments of $T_n(X)$ in the situation with a large but actually finite n.

- This leads to some asymptotic statistical procedures and asymptotic criteria for assessing their performances.
- The asymptotic approach is not only applied to the situation where no exact method (the approach considering a fixed n) is available, but also used to provide a procedure simpler (e.g., in terms of computation) than that produced by the exact approach.
- In addition to providing more theoretical results and/or simpler procedures, the asymptotic approach requires less stringent mathematical assumptions than does the exact approach.

Definition 5.5.1 (convergence in probability)

A sequence of random variables Z_n, $i = 1, 2, \ldots$, converges in probability to a random variable Z iff for every $\varepsilon > 0$,

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

A sequence of random vectors Z_n converges in probability to a random vector Z iff each component of Z_n converges in probability to the corresponding component of Z.

Theorem 5.5.2 (Weak Law of Large Numbers (WLLN))

Let $X_1, \ldots, X_n$ be iid random variables with $E(X_i) = \mu$ and finite $\mathrm{Var}(X_i) = \sigma^2$. Then, the sample mean $\bar{X}$ converges in probability to μ.

Proof.

By Chebychev's inequality and Theorem 5.2.6,

$$P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

which converges to 0 as $n \to \infty$.

Remarks.

1. Although we write the sample mean as $\bar{X}$, it depends on n.
2. The WLLN states that the probability of the sample mean $\bar{X}$ being close to the population mean μ converges to 1.
3. The existence of a finite variance σ^2 is not needed; we only need the existence of $E(X_i)$, a proof will be given later.
4. The independence assumption is not necessary either: in the previous proof, we only need X_i's are uncorrelated.

Example.

Suppose that $X_1, ..., X_n$ are identically distributed with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$, and that

$$\text{Cov}(X_t, X_s) = \begin{cases} c & \text{if } |s-t| = 1 \\ 0 & \text{if } |s-t| > 1 \end{cases}$$

Then $\bar{X}$ converges in probability to μ, because

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2}\left(\sum_{i=1}^{n}\text{Var}(X_i) + \sum_{i\neq j}\text{Cov}(X_i, X_j)\right) \\ &= \frac{\sigma^2}{n} + \frac{(n-1)c}{n^2} \end{aligned}$$

and, Chebychev's inequality,

$$P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X})}{\varepsilon^2} = \frac{\sigma^2 + (1 - n^{-1})c}{n\varepsilon^2} \to 0$$

A proof of the WLLN using chf's

Let $X_1, ..., X_n$ be iid random variables with $E|X_1| < \infty$ and $E(X_i) = \mu$.
From the result for the chf (Theorem C1), the chf of X_1 is differentiable at 0 and

$$\phi_{X_1}(t) = 1 + \imath\mu t + o(|t|) \qquad \text{as } |t| \to 0.$$

Then, the chf of $\bar{X}$ is

$$\phi_{\bar{X}}(t) = \left[\phi_{X_1}\left(\frac{t}{n}\right)\right]^n = \left[1 + \frac{\imath\mu t}{n} + o\left(\frac{t}{n}\right)\right]^n \to e^{\imath\mu t}$$

for any $t \in \mathscr{R}$ as $n \to \infty$, because $(1 + c_n/n)^n \to e^c$ for any complex sequence $\{c_n\}$ satisfying $c_n \to c$.

The limiting function $e^{\imath\mu t}$ is the chf of the constant μ.

By Theorem C7, if $F_{\bar{X}}(x)$ is the cdf of $\bar{X}$, then

$$\lim_{n\to\infty} F_{\bar{X}}(x) = \begin{cases} 1 & x > \mu \\ 0 & x < \mu \end{cases}$$

This shows that $\bar{X}$ converges in probability to μ, because of Theorem 5.5.13 to be established later.

Theorem 5.5.4.

Let $Z_1, Z_2, \ldots$ be random vectors that converge in probability to a random vector Z and let h be a continuous function. Then $h(Z_1), h(Z_2), \ldots$ converges in probability to $h(Z)$.

Example 5.5.3.

Let $X_1, X_2, \ldots$ be iid random variable with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Consider the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{n(\bar{X}-\mu)^2}{n-1}$$

Define

$$Z_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2, \quad U_n = \bar{X} - \mu, \quad a_n = \frac{n}{n-1}$$

By the WLLN, (Z_n, U_n) converges in probability to $(\sigma^2, 0)$.
Note that $a_n \to 1$ and a_n is not random, but we can view that a_n converges in probability to 1.
Then, by Theorem 5.5.4, $S^2 = h(a_n, Z_n, U_n) = a_n(Z_n - U_n^2)$ converges in probability to $h(1, \sigma^2, 0) = \sigma^2$.

Example 5.5.5.

Consider $h(x) = \sqrt{x}$.
By Theorem 5.5.4, the sample standard deviation $S = h(S^2)$ converges in probability to the population standard deviation $\sigma = h(\sigma^2)$.

Note that convergence in probability is different from the convergence of a sequence of deterministic functions $g_n(x)$ to a function $g(x)$ for x in a set $A \subset \mathscr{R}^k$.
Similar to the convergence of deterministic functions (note that random variables are random functions), we have the following concept.

Definition 5.5.6 (convergence almost surely)

A sequence of random variables Z_n, $n = 1, 2, ...$, converges almost surely to a random variable Z iff

$$P\left(\lim_{n\to\infty} Z_n = Z\right) = 1.$$

A sequence of random vectors Z_n, $n = 1, 2, ...$, converges almost surely to a random vector Z iff each component of Z_n converges almost surely to the corresponding component of Z.

- The almost sure convergence of Z_n to Z means that there is an event N such that $P(N) = 0$ and for every element $\omega \in N^c$, $\lim_{n\to\infty} Z_n(\omega) = Z(\omega)$, which is almost the same as point-wise convergence for deterministic functions (Example 5.5.7).
- If a sequence of random vectors Z_n converges almost surely to a random vector Z, and h is a continuous function, then $h(Z_n)$ converges almost surely to $h(Z)$.
- If Z_n converges almost surely to Z, then Z_n converges in probability to Z.
- Convergence in probability, however, does not imply convergence almost surely (Example 5.5.8).
- If Z_n converges in probability fast enough, then it converges almost surely, i.e., if for every $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} P(|Z_n - Z| \geq \varepsilon) < \infty,$$

then Z_n converges almost surely to Z.

It is, however, not easy to construct an example of convergence in probability but not almost surely.

Similar to the WLLN in Theorem 5.5.2, we have the following result with almost sure convergence.

Theorem 5.5.9 (Strong Law of Large Numbers (SLLN))

Let $X_1,...,X_n$ be iid random variables with $E(X_i) = \mu$. Then, the sample mean $\bar{X}$ converges almost surely to μ.

Note that we still only require the existence of the mean, not the second order moment.

The proof is omitted, since it is out of the scope of the textbook.

Approximation to an integral

Suppose that $h(x)$ is a function of $x \in \mathscr{R}^k$.
In many applications we want to calculate an integral

$$\int_{\mathscr{R}^k} h(x)dx$$

If the integral is not easy to calculate, a numerical method is needed.

The following is the so called Monte Carlo approximation method, which is based on the SLLN.

Suppose that we can generate iid random vectors $X_1, X_2, \ldots$ from a pdf $p(x)$ on $\mathscr{R}^k$ satisfying that $p(x) > 0$ if $h(x) \neq 0$.

By the SLLN, with probability equal to 1 (almost surely),

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{h(X_i)}{p(X_i)} &= E\left(\frac{h(X_1)}{p(X_1)}\right) \\
&= \int_{\mathscr{R}^k} \frac{h(x)}{p(x)} p(x) dx \\
&= \int_{\mathscr{R}^k} h(x) dx
\end{aligned}
$$

Thus, we can approximate the integral by the average $\frac{1}{n}\sum_{i=1}^{n} \frac{h(X_i)}{p(X_i)}$ with a very large n.

We can actually find what is the large enough n to have a good approximation.

We often need to consider a convergence even weaker than convergence in probability.

Definition 5.5.10 (convergence in distribution)

A sequence of random variables Z_n, $n = 1, 2, ...$, converges in distribution to a random variable Z iff

$$\lim_{n\to\infty} F_{Z_n}(x) = F_Z(x), \qquad x \in \{y : F_Z(y) \text{ is continous } \}$$

where F_{Z_n} and F_Z are the cdf's of Z_n and Z, respectively.

- Note that we only need to consider the convergence at x that is a continuity point of F_Z.
- Note that cdf's, not pdf's or pmf's, are involved in this definition.
- In convergence in distribution, it is really the cdfs that converge, not the random variables; in fact, the random variables can be defined in different spaces, which is very different from the convergence in probability or almost surely.

Example 5.5.11.

Let $X_1, X_2, ...$ be iid from uniform on $(0,1)$ and $X_{(n)} = \max_{i=1,...,n} X_i$.
For every $\varepsilon > 0$,

$$\begin{aligned} P(|X_{(n)} - 1| \geq \varepsilon) &= P(X_{(n)} \leq 1 - \varepsilon) + P(X_{(n)} \geq 1 + \varepsilon) \\ &= P(X_{(n)} \leq 1 - \varepsilon) \\ &= P(X_i \leq 1 - \varepsilon, i = 1, ..., n) \\ &= (1 - \varepsilon)^n \end{aligned}$$

Hence, $X_{(n)}$ converges in probability to 1.
In fact, since $\sum_{n=1}^{\infty} (1-\varepsilon)^n < \infty$, $X_{(n)}$ converges almost surely to 1.
For any $t > 0$,

$$\begin{aligned} P(n(1 - X_{(n)}) \leq t) &= 1 - P(n(1 - X_{(n)}) \geq t) \\ &= 1 - P(X_{(n)} \leq 1 - t/n) \\ &= 1 - (1 - t/n)^n \\ &\to 1 - e^{-t} \end{aligned}$$

which is the cdf of the *exponential*$(0,1)$ distribution.

It is clear that $P(n(1-X_{(n)}) \le t) = 0$ if $t \le 0$.
Thus, $n(1-X_{(n)})$ converges in distribution to $Z \sim exponential(0,1)$.

The next theorem shows that convergence in distribution is weaker than convergence in probability and, hence, is also weaker than almost sure convergence.

Theorem 5.5.12.

If Z_n converges in probability to Z, then Z_n converges in distribution to Z.

Proof.

For any $x \in \mathscr{R}$ and $\varepsilon > 0$,

$$\begin{aligned} F_Z(x-\varepsilon) &= P\big(Z \le x-\varepsilon\big) \\ &\le P\big(Z_n \le x\big) + P\big(Z \le x-\varepsilon, Z_n > x\big) \\ &\le F_{Z_n}(x) + P\left(|Z_n - Z| > \varepsilon\right). \end{aligned}$$

Letting $n \to \infty$, we obtain that

$$F_Z(x-\varepsilon) \le \liminf_n F_{Z_n}(x).$$

Switching Z_n and Z in the previous argument, we can show that

$$F_Z(x+\varepsilon) \geq \limsup_n F_{Z_n}(x)$$

i.e.,

$$F_Z(x-\varepsilon) \leq \liminf_n F_{Z_n}(x) \leq \limsup_n F_{Z_n}(x) \leq F_Z(x+\varepsilon)$$

Since ε is arbitrary,

$$\lim_{\varepsilon\to 0} F_Z(x-\varepsilon) \leq \liminf_n F_{Z_n}(x) \leq \limsup_n F_{Z_n}(x) \leq \lim_{\varepsilon\to 0} F_Z(x+\varepsilon)$$

Now, if F_Z is continuous at x, then the limit on the far left hand side equals the limit on the far right hand side and both are equal to $F_Z(x)$, which shows that

$$F_Z(x) = \lim_{n\to\infty} F_{Z_n}(x).$$

Example.

The converse of Theorem 5.5.12 is not true in general.

Let $\theta_n = 1 + n^{-1}$ and X_n be a random variable having the *exponential*$(0,\theta_n)$ distribution, $n = 1, 2, ...$
Let X be a random variable $\sim$ *exponential*$(0,1)$.

For any $x > 0$, as $n \to \infty$,

$$F_{X_n}(x) = 1 - e^{-x/\theta_n} \to 1 - e^{-x} = F_X(x)$$

Since $F_{X_n}(x) \equiv 0 \equiv F_X(x)$ for $x \leq 0$, we have shown that X_n converges in distribution to X.

Does X_n converge in probability to X?

- Need further information about the random variables X and X_n.
- We consider two cases in which different answers can be obtained.

Case 1

Suppose that $X_n \equiv \theta_n X$ (then X_n has the given distribution).
$X_n - X = (\theta_n - 1)X = n^{-1}X$, which has the cdf $(1 - e^{-nx})I_{[0,\infty)}(x)$.
Then, X_n converges in probability to X, because, for any $\varepsilon > 0$,

$$P(|X_n - X| \geq \varepsilon) = e^{-n\varepsilon} \to 0$$

In fact, X_n converges almost surely to X, since

$$\sum_{n=1}^{\infty} e^{-n\varepsilon} < \infty$$

Case 2

Suppose that X_n and X are independent random variables.
Since the pdf's of X_n and $-X$ are $\theta_n^{-1}e^{-x/\theta_n}I_{(0,\infty)}(x)$ and $e^x I_{(-\infty,0)}(x)$, respectively, we have

$$P(|X_n - X| \le \varepsilon) = \int_{-\varepsilon}^{\varepsilon}\int \theta_n^{-1}e^{-x/\theta_n}e^{y-x}I_{(0,\infty)}(x)I_{(-\infty,x)}(y)dxdy,$$

which converges to (by the dominated convergence theorem)

$$\int_{-\varepsilon}^{\varepsilon}\int e^{-x}e^{y-x}I_{(0,\infty)}(x)I_{(-\infty,x)}(y)dxdy = 1 - e^{-\varepsilon}.$$

Thus,

$$P(|X_n - X| \ge \varepsilon) \to e^{-\varepsilon} > 0$$

for any $\varepsilon > 0$ and, therefore, X_n does not converge in probability to X.

In one situation, however, convergence in distribution is equivalent to convergence in probability, as the following result shows.

Theorem 5.5.13.

Z_n converges in probability to a constant c iff Z_n converges in distribution to c.

Proof.

The “only if" part is a special case of Theorem 5.5.12.
Hence, we only need to show the “if" part.
If Z_n converges in distribution to a constant c, then

$$\lim_{n\to\infty} P(Z_n \le x) = \begin{cases} 0 & x < c \\ 1 & x > c \end{cases}$$

which is the “cdf” of a constant c. (Note that the limit does not include the case of $x = 0$, which is a discontinuity point of the cdf of c.
For every $\varepsilon > 0$,

$$\begin{aligned} P(|Z_n - c| \ge \varepsilon) &= P(Z_n - c \ge \varepsilon) + P(Z_n - c \le -\varepsilon) \\ &= P(Z_n \ge c + \varepsilon) + P(Z_n \le c - \varepsilon) \\ &= 1 - P(Z_n < c + \varepsilon) + P(Z_n \le c - \varepsilon) \\ &\le 1 - P(Z_n < c + \varepsilon/2) + P(Z_n \le c - \varepsilon) \\ &\to 1 - 1 + 0 = 0 \end{aligned}$$

since $c - \varepsilon < c$ and $c + \varepsilon/2 > c$.
This proves that Z_n converges in probability to c.