1.2 Convergence of sequences of random variables

- If $\{X_n\}_{n=1}^{+\infty}$ is a sequence of random variables and X is a random variable, what meaning can we attribute to the statement "X_n converges to X",

$$X_n \longrightarrow X \quad \text{as } n \to +\infty \ ?$$

- Random variables are <u>functions</u>. There are various notions of convergence for sequences of functions.

- Consider the example $f_n : [0, 1] \to \mathbb{R}$

$$f_n(x) = \begin{cases} 0 \ , & 1/n \leq x \leq 1 \\ n - n^2 x \ , & 0 \leq x < 1/n \end{cases}$$

As $n \to +\infty$ to what is f_n converging to?

Notions of convergence: let $f_n, f : [0,1] \to \mathbb{R}$

- Pointwise convergence

$$f_n(x) \to f(x) \quad \text{for all } x \in [0,1]$$

- Uniform convergence

$$\sup_{x \in [0,1]} |f_n(x) - f(x)| \to 0$$

- Convergence in L^P

$$\int_0^1 |f_n(x) - f(x)|^p dx \to 0$$

- In measure

$$\mu(A_{n,\varepsilon}) \to 0 \quad \text{for all } \varepsilon > 0$$

$$\text{where } A_{n,\varepsilon} = \{x \in [0,1] : \ |f_n(x) - f(x)| > \varepsilon\}$$

For random variables, we have similar concepts but a slightly different language:

Definition 1.1 *Let $\{X_n\}_{n=1}^{+\infty}$ be a sequence of random variables and X a random variable, all defined in the same probability space $(\Omega, \mathcal{A}, P)$. Then, we say that*

- *X_n converges to X almost surely (or with probability 1, w.p. 1), $X_n \overset{as}{\to} X$, iff*

$$P[\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}] = 1$$

- *X_n converges to X in the rth mean ($r \geq 1$), $X_n \overset{r}{\to} X$, iff*

$$E[|X_n - X|^r] \to 0$$

- *X_n converges in probability to X, $X_n \overset{P}{\to} X$, iff for all $\varepsilon > 0$*

$$P(|X_n - X| > \varepsilon) \to 0$$

- *X_n converges to X in distribution, $X_n \overset{d}{\to} X$, iff*

$$F_n(x) \to F(x)$$

  for all x continuity point of F, where $F(x) = P(X \leq x)$ and $F_n(x) = P(X_n \leq x)$. ■

Remarks:

- The random variables are all defined in the same probability space $(\Omega, \mathcal{A}, P)$: Ω is the sample space, $\mathcal{A}$ is a σ-algebra of subsets of Ω (the events), X_n and X map Ω into $\mathbb{R}$. However, for the concept of convergence in distribution this restriction is not necessary as the definition only involves the cdf of the r.v.

- Convergence in distribution is really about the convergence of the sequence of probability distributions and not of the random variables themselves

- When defining convergence in the rth mean, it is assumed that the corresponding expected values exist: $E[|X_n|^r] < +\infty$ and $E[|X|^r] < +\infty$

- When $X_n \stackrel{1}{\to} X$, we say that X_n converges to X in mean; when $X_n \stackrel{2}{\to} X$, we say that X_n converges to X in quadratic mean

- The four criteria are obviously not equivalent:

Theorem 1.1 *In general, the following implications are valid:*

$$
\begin{tikzcd}
X_n \xrightarrow{as} X \arrow[dr, Rightarrow] & & \\
& X_n \xrightarrow{P} X \arrow[r, Rightarrow] & X_n \xrightarrow{d} X \\
X_n \xrightarrow{s} X \arrow[ur, Rightarrow] & & \\
X_n \xrightarrow{r} X \arrow[u, Rightarrow, "(r > s \geq 1)"'] & &
\end{tikzcd}
$$

No other implication is valid in general without imposing extra conditions.

Show that (i) convergence in mean implies convergence in probability, and (ii) convergence in probability implies convergence in distribution. In the homework problems, convergence in the rth mean implies convergence in the sth mean, $r > s \geq 1$.

Converses/useful results

- If $X_n \stackrel{P}{\to} X$, then there exists $\{n_k\}_{k=1}^{+\infty}$ such that $X_{n_k} \stackrel{as}{\to} X$ when $k \to +\infty$.
- If $|X_n|^r$ is uniformly integrable, then $(X_n \stackrel{P}{\to} X \Rightarrow X_n \stackrel{r}{\to} X$)
- *Skorokhod representation theorem*: If $X_n \stackrel{d}{\to} X$ then there exists a probability space $(\Omega', \mathcal{A}', P')$ and r.v. $\{Y_n\}$ and Y, defined in Ω', such that
  - $P'(Y_n \leq y) = P(X_n \leq y)$ and $P'(Y \leq y) = P(X \leq y)$ for all $y \in \mathbb{R}$ (X_n and Y_n are (marginally) equal in distribution, the same for X and Y)
  - $Y_n \stackrel{as}{\to} Y$
- $X_n \stackrel{P}{\to} c \Leftrightarrow X_n \stackrel{d}{\to} c$, where $c \in \mathbb{R}$ — See HW for a proof — very useful result
- Another very useful result: since $E[(X_n - \theta)^2] = \mathrm{Var}(X_n) + (E[X_n] - \theta)^2$ if $\mathrm{Var}(X_n) \to 0$ and $E[X_n] \to \theta$ we have convergence in mean square to θ, and hence convergence in probability to θ

Theorem 1.2 *(*Continuous mapping theorem*) Let* $h : \mathbb{R} \to \mathbb{R}$ *be a continuous function. Then:*

1. $X_n \stackrel{as}{\to} X \Rightarrow h(X_n) \stackrel{as}{\to} h(X)$
2. $X_n \stackrel{d}{\to} X \Rightarrow h(X_n) \stackrel{d}{\to} h(X)$
3. $X_n \stackrel{P}{\to} X \Rightarrow h(X_n) \stackrel{P}{\to} h(X)$

Show.

Theorem 1.3 *(Slutsky) Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables, X a random variable and c a real number. If $X_n \stackrel{d}{\to} X$ and $Y_n \stackrel{P}{\to} c$, then*

1. $X_n + Y_n \stackrel{d}{\to} X + c$
2. $Y_n \ X_n \stackrel{d}{\to} cX$
3. $X_n/Y_n \stackrel{d}{\to} X/c$ *as long as* $c \neq 0$.

Remarks:

- Suppose that $X_n \stackrel{d}{\to} X$ where $X \sim \mathrm{N}(0,1)$. Then, with $Z_n = -X_n$ we have $Z_n \stackrel{d}{\to} X$. However, $X_n + Z_n = 0$, hence $X_n + Z_n$ does not converge in distribution to $2X$ as one might (erroneously) expect.
- If $X_n \sim t(n)$, then $X_n \stackrel{d}{\to} \mathrm{N}(0,1)$. How do we show this?

1.3 Some important asymptotic results

- Law(s) of Large Numbers
- Central Limit Theorem
- Delta Method

Theorem 1.4 (Weak Law of Large Numbers) *Let $\{X_n\}_{n=1}^{+\infty}$ be a sequence of independent and identically distributed (iid) random variables, with $E[X_n] = \mu$ and $Var(X_n) = \sigma^2 < +\infty$. Let also $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, we have that*

$$\bar{X}_n \overset{P}{\to} \mu \ .$$

Proof: Immediate from Chebychev's inequality. Note that $E[\bar{X}_n] = \mu$ and $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$, hence, for any $\varepsilon > 0$, $P(|\bar{X}_n - \mu| > \varepsilon) \leq \mathrm{Var}(\bar{X}_n)/\varepsilon^2 = \sigma^2/(n\varepsilon^2)$ which goes to zero as $n \to +\infty$. (Or, perhaps simpler, see that $E[\bar{X}_n]) \to \mu$ and $\mathrm{Var}(\bar{X}_n) \to 0$, hence we have convergence in mean square, and hence in probability.)

The Strong Law of Large Numbers states that in fact we have almost sure convergence

Theorem 1.5 (Strong Law of Large Numbers) *In the same conditions as above, we have*

$$\bar{X}_n \stackrel{as}{\to} \mu \ .$$

Remark

- Actually, it is only necessary to assume that $E[|X_i|] < +\infty$ for both laws to hold, but the proofs become more difficult.
- Intuitively, the WLLN tell us that $\bar{X}_n$ becomes more and more concentrated around μ as n increases: $P(|\bar{X}_n - \mu| > \varepsilon) \to 0$ as $n \to +\infty$

Theorem 1.6 (Central Limit Theorem) *Let $\{X_n\}_{n=1}^{+\infty}$ be a sequence of iid random variables possessing finite variance. Let $\mu = E[X_n]$ and $\sigma^2 = Var(X_n)$. Let also $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and*

$$Z_n = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \ .$$

Then, we have that

$$Z_n \xrightarrow{d} Z \ ,$$

where $Z \sim N(0,1)$.

Remarks

- The CLT is valid as long as $\mathrm{Var}(X_n)$ is finite. In that case, the proof involves characteristic functions. We will be proving a weaker version of the theorem which assumes the existence of mgf. (Note that the existence of mgf implies the existence of all moments.)

Remarks (ctd)

- The CLT is often used to compute probabilities of the type $P(\bar{X}_n \leq x)$ approximating them by $\Phi(\sqrt{n}(x-\mu)/\sigma)$ for sufficiently large n.
- Intuitively, the CLT tells us that the <u>distribution</u> of $\bar{X}_n$ is well-approximated by a normal distribution for sufficiently large n (as long as the variance is finite). This is remarkable!
- Notation

$$\bar{X}_n \overset{a}{\sim} \mathrm{N}(\mu, \sigma^2/n)$$
$$\bar{X}_n - \mu \overset{a}{\sim} \mathrm{N}(0, \sigma^2/n)$$
$$\sqrt{n}(\bar{X}_n - \mu) \overset{a}{\sim} \mathrm{N}(0, \sigma^2)$$

- What does ‘sufficiently large n’ really mean?

Theorem 1.7 (Lévy's continuity theorem) *Suppose that* $\{X_n\}_{n=1}^{+\infty}$ *is a sequence of random variables and let* $M_n(s)$ *denote the mgf of* X_n, $n = 1, 2, \ldots$. *Additionally, assume that*

$$\lim_{n \to +\infty} M_n(s) = M(s)$$

for s *in a neighborhood of the origin, and that* $M(\cdot)$ *is the mgf of a random variable* X*. In these circumstances,*

$$X_n \xrightarrow{d} X \ .$$

Assuming the continuity theorem as proved, it's easy to prove the CLT:

- We have to show that for s in a neighborhood of the origin, the mgf of Z_n converges to $e^{s^2/2}$.
- Consider $Y_i = (X_i - \mu)/\sigma$. Then, $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$ and $M_{Z_n}(s) = [M_Y(s/\sqrt{n})]^n$, where M_Y is the mgf of Y_i.

- Expand M_Y as a power series around the origin retaining only the 2nd order terms, to conclude that

$$M_Y(s) = \sum_{k=0}^{2} M_Y^{(k)}(0)\frac{(s)^k}{k!} + r(s)$$

where $r(s)/s^2$ goes to zero as s approaches zero.

- Since $M^{(k)}(0) = E[Y^k]$, $E[Y_i] = 0$, and $\mathrm{Var}(Y_i) = 1$, it follows that

$$M_Y(s/\sqrt{n}) = 1 + \frac{(s/\sqrt{n})^2}{2!} + r(s/\sqrt{n})$$

where $nr(s/\sqrt{n})$ goes to zero as $n \to \infty$.

- Finally,

$$\begin{aligned}
\lim_{n\to+\infty} M_{Z_n}(s) &= \lim_{n\to+\infty}\left[1 + \frac{(s/\sqrt{n})^2}{2!} + r(s/\sqrt{n})\right]^n \\
&= \lim_{n\to+\infty}\left[1 + \frac{1}{n}(s^2/2 + nr(s/\sqrt{n}))\right]^n \\
&= e^{s^2/2}\ .
\end{aligned}$$

Remarks

1. The result stated above as Levi's continuity theorem is also known as Curtiss (1947) theorem.
2. In Curtiss (1942) there is an example of a sequence of cdf $\{F_n(x)\}$ which converges to a cdf $F(x)$, while the corresponding sequence of the mgf does not converge to the mgf of $F(x)$, although these exist for all t
3. Kozakiewicz (1947) gave conditions on the mgf's that ensure the reciprocal of Curtiss (1942) theorem
4. Recently, some results have been obtained which relax the existence of the mgf in a neighborhood of zero: see, eg, Mukherjea et al. (2006).

Applications

1. $\{X_n\}_{n=1}^{+\infty}$ iid $\mathrm{B}(1, \theta)$, where $\theta \in (0, 1)$. By the Central Limit Theorem,

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

On the other hand, the (Weak) Law of Large Numbers, ensures that $\bar{X}_n \xrightarrow{P} \theta$. By the Continuous Mapping Theorem,

$$\frac{\sqrt{\theta(1-\theta)}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{P} 1$$

and Slutsky's Theorem allows one to conclude that

$$\sqrt{n} \ \frac{\bar{X}_n - \theta}{\sqrt{\theta(1-\theta)}} \ \frac{\sqrt{\theta(1-\theta)}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} = \sqrt{n}\frac{\bar{X}_n - \theta}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{d} \mathrm{N}(0,1) \ ,$$

which in practice means that, for large n

$$P\left(\sqrt{n}\frac{\bar{X}_n - \theta}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \leq x\right) \approx \Phi(x) \ .$$

2. Compute $P(X \in A)$ using simple Monte Carlo

Start by noting that

$$P(X \in A) = E[Y] \quad \text{where } Y = I_A(X) \ .$$

Let $X_1, \ldots, X_M$ be iid random variables, with the same distribution as X, and $Y_i = I_A(X_i)$, $i = 1, \ldots, M$. Then, by the SLLN,

$$\bar{Y}_M = \frac{1}{M} \sum_{i=1}^{M} Y_i \overset{as}{\to} E[Y] = P(X \in A) \ .$$

In practice, this result can be utilized to compute an approximate value for $P(X \in A)$ in the following way

- generate in a computer M independent realizations of X, which we denote by $x_1, \ldots, x_M$
- for each x_i, compute $y_i = I_A(x_i)$
- for sufficiently large M

$$P(X \in A) \approx \frac{1}{M} \sum_{i=1}^{M} y_i = \frac{1}{M} \#\{i = 1, \ldots, M : \ x_i \in A\}$$

2. Simple Monte Carlo (ctd)

Simple Monte Carlo allows us to replace the analytical knowledge of a probability distribution by a sufficiently large sample of iid draws from that distribution — $x_1, \ldots, x_M$ — since (almost) all aspects of that probability distribution can be arbitrarily approximated using that sample.

Another example: continuous distribution with density f; how to compute $f(a)$ for some $a \in \mathbb{R}$?

$$
\begin{aligned}
f(a) &= \lim_{\delta \to 0} \frac{F(a+\delta) - F(a)}{\delta} \\
&= \frac{1}{\delta} \; \frac{1}{M} \# \{ i = 1, \ldots, M : a < x_i \leq a + \delta \}
\end{aligned}
$$

That is, the histogram of $x_1, \ldots, x_M$ is an approximation to the density of X

Theorem 1.8 (Delta Method) *Let $\{X_n\}_{n=1}^{+\infty}$ be a sequence of r.v. such that $\forall\theta \in \Theta$*

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

Let $\theta_0 \in \Theta$ and g be a differentiable function such that $g'(\theta_0) \neq 0$. Then,

$$\sqrt{n}\ (g(X_n) - g(\theta_0)) \xrightarrow{d} N(0, \sigma^2\ [g'(\theta_0)]^2)\ .$$

Proof: Expand g using a Taylor formula of order 1 around θ_0:

$$g(x) = g(\theta_0) + g'(\theta_0)(x - \theta_0) + r(x - \theta_0)$$

where r is such that $r(x-\theta_0)/(x-\theta_0) \to 0$ when $x \to \theta_0$. Hence,

$$\sqrt{n}\ (g(X_n) - g(\theta_0)) = \sqrt{n}\ g'(\theta_0)\ (X_n - \theta_0) + \sqrt{n}\ r(X_n - \theta_0)\ .$$

If we show that $\sqrt{n}\ r(X_n - \theta_0) \xrightarrow{P} 0$, then Slutsky's theorem guarantees the result. This is part of the homework problem sheet.

Application Suppose that $X_1, \ldots, X_n$ are iid $\mathrm{B}(1, \theta)$. Then, the CLT ensures that

$$\sqrt{n}\ \frac{\bar{X}_n - \theta}{\sqrt{\theta(1-\theta)}} \xrightarrow{d} \mathrm{N}(0,1)$$

which is equivalent to

$$\sqrt{n}\ (\bar{X}_n - \theta) \xrightarrow{d} \mathrm{N}(0, \theta(1-\theta))\ .$$

What is the asymptotic distribution of $Y_n = \ln \frac{\bar{X}_n}{1-\bar{X}_n}$? With $g(x) = \ln \frac{x}{1-x}$, it follows that $g'(x) = 1/[x(1-x)]$ and the Delta Method ensures that

$$\sqrt{n}\ \left(Y_n - \ln \frac{\theta}{1-\theta}\right) \xrightarrow{d} \mathrm{N}(0, [\theta(1-\theta)]^{-1})$$

which is often written as

$$Y_n \overset{a}{\sim} N\left(\ln \frac{\theta}{1-\theta}, \frac{[\theta(1-\theta)]^{-1}}{n}\right)$$

and we say that Y_n's approximate distribution is normal with mean value $\ln \frac{\theta}{1-\theta}$ (the asymptotic mean) and variance $\frac{[\theta(1-\theta)]^{-1}}{n}$ (the asymptotic variance).

Application Suppose $X_1, \ldots, X_n$ are $\mathrm{B}(1, \theta)$. The CLT ensures that

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

The asymptotic variance depends on the true value of θ.

Stabilize the variance using a transformation: consider g such that $g'(\theta) = 1/\sqrt{\theta(1-\theta)}$. Solution: $g(\theta) = 2 \arcsin \sqrt{\theta}$. Then:

$$\sqrt{n}\ (2 \arcsin \sqrt{\bar{X}_n} - 2 \arcsin \sqrt{\theta}) \xrightarrow{d} \mathrm{N}(0,1)\ .$$

In general, if $\sqrt{n}\ (Y_n - \mu) \xrightarrow{d} \mathrm{N}(0, h(\mu))$ then h must be such that $g'(\mu) = (\sqrt{h(\mu)})^{-1}$, hence

$$g(\mu) = \int_c^\mu \frac{1}{\sqrt{h(t)}}\ dt$$

where c is any constant such that the integral exists. In that case,

$$\sqrt{n}\ (g(Y_n) - g(\mu)) \xrightarrow{d} \mathrm{N}(0,1)$$