MODULE 5: TYPES OF CONVERGENCE

Ruben Zamar
Deapartment of Statistics
UBC

August 16, 2016

PART I

MODES OF CONVERGENCE

NOTATION

- $\{X_n\}$: sequence of random variables (or vectors)
- $X_n \sim F_n(x)$
- $X_0 \sim F_0$: random variable (or vector)

MOTIVATION

- Different ways in which the sequence $\{X_n\}$ gets close to X_0 as n increases

- **In applications**: The limit, X_0, is used to approximate X_n for large n

SOME TYPES OF CONVERGENCE

1. Almost sure convergence (Convergence with prob 1)
2. Convergence in probability
3. Convergence in distribution
4. Convergence in L_p

ALMOST SURE CONVERGENCE

- Suppose that X_n $(n = 0, 1, 2, 3, \ldots)$ are defined on the same probability space $(\Omega, \mathcal{F}, P)$
- DEFINITION: X_n **converges almost surely to** X_0 if

$$P\left(\lim_{n\to\infty} X_n = X_0\right) = 1.$$

Notation:

$$X_n \to X_0, \quad \text{a.s. } P$$

It can be shown that

$$P\left(\lim_{n\longrightarrow\infty} X_n = X\right) \; = \; 1 \Longleftrightarrow$$

$$\lim_{n\longrightarrow\infty} P\left(|X_m - X| < \epsilon, \vee \; m \geq n\right) \; = \; 1,$$

for all $\epsilon \; > \; 0.$

In other words:

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1 \Longleftrightarrow$$

for all $\epsilon > 0$.

$$P\left(\underline{\lim}_{n\to\infty}\left\{w : \left|X_n\left(w\right) - X\left(w\right)\right| < \epsilon\right\}\right) = 1.$$

- Other names for "almost sure convergence" are
  - **"convergence with probability 1"**
  - **"convergence almost everywhere"**

CONVERGENCE IN PROBABILITY

- Again, X_n and X_0 must be defined on the **same probability space** $(\Omega, \mathcal{F}, P)$

- So we can speak of the set of $w's$ where

$$|X_n(w) - X_0(w)| \qquad \text{is small.}$$

- DEFINITION: X_n **converges in probability to** X_0 if for all $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} P\left(|X_n - X_0| < \epsilon\right) = 1$$

- Notation

$$X_n \longrightarrow_p X_0.$$

- X_n **converges in distribution to** X_0 if

$$\lim_{n\longrightarrow\infty} F_n(x) = F_0(x),$$

for all point x **at which** $F(x)$ **is continuous.**

- The variables X_n may be defined on different probability spaces

RESTRICTION TO CONTINUITY POINTS

- The restriction to continuity points of $F(x)$ is a technicality needed to avoid counter intuitive situations

- A simple example. Let

$$X_n = 1/n \quad \text{with probability one}$$

- Of course

$$X_n = 1/n \longrightarrow 0$$

in all possible ways!

RESTRICTION TO CONTINUITY POINTS (continued)

However, for all $n \geq 1$

$$F_{X_n}(0) = 0$$

hence

$$F_{X_n}(0) \to 0 \neq F_0(0) = 1$$

Claim: the set D of continuity points of $F(x)$ is dense in R

Proof: the complement set

$$D^c = \left\{x : F(x) - F\left(x^-\right) > 0\right\}$$

is finite or countable.

RESTRICTION TO CONTINUITY POINTS (continued)

In fact

$$D^c = \cup_{n=1}^{\infty} \left\{ x : F(x) - F\left(x^-\right) > 1/n \right\}$$

and

$$\# \left\{ x : F(x) - F\left(x^-\right) > 1/n \right\} \leq n.$$

Therefore, for all point $x \in R$ there exist sequences $x_n \in D$ such that

$$x_n \uparrow x \qquad \text{and} \qquad x_n \downarrow x.$$

RELATION BETWEEN THESE TYPES OF CONVERGENCE

Type	Implies	a.s.	Probability	Distribution
a.s.	$\Rightarrow$	-	**Yes**	**Yes**
Probability	$\Rightarrow$	**No**	-	**Yes**
Distribution	$\Rightarrow$	**No**	**No**	-
Dist. to C	$\Rightarrow$	**No**	**Yes**	-

CONVERGENCE IS PRESERVED BY CONTINUOUS FUNCTIONS

Suppose g is a continuous function. Then

$$X_n \rightarrow X_0 \text{ a.s.} \Rightarrow g(X_n) \rightarrow g(X_0) \text{ a.s.} \quad \text{(easy)}$$

$$X_n \rightarrow_p X_0 \Rightarrow g(X_n) \rightarrow_p g(X_0) \quad \text{(mildly difficult)}$$

$$X_n \rightarrow_d X_0 \Rightarrow g(X_n) \rightarrow_d g(X_0) \quad \text{(fairly difficult)}$$

MARGINAL CONVERGENCE AND JOINT CONVERGENCE

Let

$$\mathbf{X}_n = \begin{pmatrix} X_{1,n} \\ \vdots \\ X_{m,n} \end{pmatrix} \quad \text{and} \quad \mathbf{X}_0 = \begin{pmatrix} X_{1,0} \\ \vdots \\ X_{m,0} \end{pmatrix}$$

$$X_{i,n} \quad \rightarrow \quad X_{i,0} \;\text{ a.s. } \; (i = 1, ..., p) \Rightarrow \mathbf{X}_n \rightarrow \mathbf{X}_0 \;\text{ a.s.}$$

$$X_{i,n} \quad \rightarrow_p \quad X_{i,0} \quad (i = 1, ..., p) \quad \Rightarrow \mathbf{X}_n \rightarrow_p \mathbf{X}_0$$

MARGINAL CONVERGENCE AND JOINT CONVERGENCE

However

$$X_{i,n} \;\; \to_d \;\; X_{i,0} \quad (i = 1, ..., p) \quad \text{does not imply}$$

$$\mathbf{X}_n \;\; \to_d \;\; \mathbf{X}_0$$

SLUTZKY'S THEOREM

(a) If $X_n \longrightarrow_d X_0$ and $Y_n \longrightarrow_p c$, then

$$X_n + Y_n \quad \longrightarrow_d X_0 + c.$$

(b) If $X_n \longrightarrow_d X_0$ and $Y_n \longrightarrow_p c$, then

$$X_n Y_n \quad \longrightarrow_d cX.$$

Suppose that $X_1, X_2, X_3, \ldots$ are iid with

$$E\{|X_1|\} < \infty \quad \text{and} \quad E(X_1) = \mu$$

Then

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \to_p \mu.$$

THE STRONG LOW OF LARGE NUMBERS (SLLN)

Suppose that $X_1, X_2, X_3, \ldots$ are iid with

$$E\{|X_1|\} < \infty \qquad \text{and} \qquad E(X_1) = \mu$$

Then

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \qquad \text{a.s.}$$

THE CENTRAL LIMIT THEOREM

Suppose that $X_1, ..., X_n, ...$ are iid with

$$E(X_1) = \mu \qquad \text{and} \qquad Var(X_1) = \sigma^2 < \infty$$

$$\sqrt{n}\left(\overline{X}_n - \mu\right)/\sigma \quad \rightarrow_d \quad N(0,1)$$

THE MULTIVARIATE CENTRAL LIMIT THEOREM

Suppose that $\mathbf{X}_1, ..., \mathbf{X}_n, ...$ are iid with mean

$$\boldsymbol{\mu} = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{pmatrix} \quad \text{and} \quad Var(X_j) = \sigma_j^2 < \infty, \ \ j = 1, 2, ..., p.$$

Then

$$\sqrt{n}\left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right) \to_d N\left(\mathbf{0}, \Sigma\right)$$

Equivlently

$$\sqrt{n}\Sigma^{-1/2}\left(\overline{\mathbf{X}}_n-\boldsymbol{\mu}\right)\rightarrow_d N\left(\mathbf{0},I\right)$$

where

$$\Sigma^{-1/2}=\sum_{j=1}^{p}\lambda_j^{-1/2}\mathbf{a}_j\mathbf{a}_j'$$

Here,

$$\lambda_1,...,\lambda_p \quad \text{are the eigenvalues of} \quad \Sigma$$

and

$$\mathbf{a}_1,...,\mathbf{a}_p \quad \text{are the corresponding eigenvectors.}$$

THE DELTA METHOD - UNIVARIATE

Suppose that

$$\sqrt{n}\left(X_n - \mu\right) \ \rightarrow_d \ N\left(0, \sigma^2\right)$$

Let $g\left(t\right)$ be a **continuously differentiable** function at μ.
Then

$$\sqrt{n}\left(g\left(X_n\right) - g\left(\mu\right)\right) \ \rightarrow_d \ N\left(0, \left[g'\left(\mu\right)\right]^2 \sigma^2\right)$$

THE DELTA METHOD - MULTIVARIATE

$$\sqrt{n}\left(\mathbf{X}_n - \boldsymbol{\mu}\right) \ \rightarrow_d \ N\left(\mathbf{0}, \Sigma\right)$$

$g\left(\mathbf{t}\right)$ is a **continuously differentiable** function at μ.
Then

$$\sqrt{n}\left(g\left(\mathbf{X}_n\right) - g\left(\boldsymbol{\mu}\right)\right) \ \rightarrow_d \ N\left(\mathbf{0}, \nabla_g\left(\boldsymbol{\mu}\right)' \Sigma \nabla_g\left(\boldsymbol{\mu}\right)\right)$$

$\nabla_g(\mathbf{t})$ is the gradient of g, that is,

$$\nabla_g(\mathbf{t}) = \left(\frac{\partial g(\mathbf{t})}{\partial t_j}\right) = \begin{pmatrix} \frac{\partial g(\mathbf{t})}{\partial t_1} \\ \frac{\partial g(\mathbf{t})}{\partial t_2} \\ \vdots \\ \frac{\partial g(\mathbf{t})}{\partial t_p} \end{pmatrix}.$$

Suppose that $X_1, X_2, ..., X_n$ are $i.i.d.$ Binomial$(1, p)$.

Then

$$\hat{p} \ \longrightarrow_p \ p \quad \text{by the WLLN}$$

and

$$\frac{\sqrt{n}\,(\hat{p} - p)}{\sqrt{p\,(1-p)}} \longrightarrow_d N\,(0,1)\,, \quad \text{by the CLT} \qquad (1)$$

EXAMPLE 1 (CONTINUED)

Hence

$$\frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}(1-\hat{p})}} \to_p 1 \quad \text{by continuity of } g(t) = \sqrt{t(1-t)} \tag{2}$$

By (1) and (2) and Slutzky's Theorem

$$\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{p(1-p)}} \frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}(1-\hat{p})}} \to_d N(0,1)$$

EXAMPLE 1 (CONTINUED)

That is

$$\frac{\sqrt{n}\left(\hat{p}-p\right)}{\sqrt{\hat{p}\left(1-\hat{p}\right)}} \rightarrow_d N\left(0,1\right)$$

EXAMPLE 2

The odds ratio is defined as

$$r = \frac{p}{1-p}$$

and an estimate is given by

$$\hat{r} = \frac{\hat{p}}{1-\hat{p}}$$

EXAMPLE 2 (CONTINUED)

Using (1) and the Delta Method with

$$
\begin{aligned}
g(t) &= \frac{t}{1-t} \\
g'(t) &= \frac{1-t+t}{(1-t)^2} = \frac{1}{(1-t)^2}
\end{aligned}
$$

EXAMPLE 2 (CONTINUED)

We have

$$\begin{aligned}
\sqrt{n}\left(\hat{p}-p\right) &\to_d N\left(0, p\left(1-p\right)\right) \\
\sqrt{n}\left(\hat{r}-r\right) &\to_d N\left(0, \left[\frac{1}{\left(1-p\right)^2}\right]^2 p\left(1-p\right)\right) \\
&= N\left(0, \frac{p}{\left(1-p\right)^3}\right)
\end{aligned}$$

Hence, for large n,

$$\hat{r} = \frac{\hat{p}}{1-\hat{p}} \sim N\left(\frac{p}{1-p}, \frac{p}{n\left(1-p\right)^{3}}\right) = N\left(r, \frac{r^{2}}{np\left(1-p\right)}\right)$$

EXAMPLE 3

Let $X_1, X_2, ..., X_n$ be *i.i.d.* with

$$\begin{aligned} E(X_1) &= \mu \\ Var(X_1) &= \sigma^2 \end{aligned}$$

$$Var\left((X_1 - \mu)^2\right) = \tau_4 < \infty$$

EXAMPLE 3

(Continued)
Consider the sample variance estimate,

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{n}.$$

Note that

$$\begin{aligned}
\hat{\sigma}_n^2 &= \frac{\sum_{i=1}^n \left[(X_i - \mu) - (\bar{X} - \mu)\right]^2}{n} \\
&= \frac{\sum_{i=1}^n \left[Y_i - \bar{Y}\right]^2}{n}, \quad \text{with } E\left(Y_i\right) = 0.
\end{aligned}$$

EXAMPLE 3 (CONTINUED)

Can assume w.l.g. that $\mu = 0$ and $\sigma^2 = E\left(Y^2\right)$

By the CLT

$$\sqrt{n}\ \left(\overline{Y^2} - \sigma^2\right)\ \rightarrow_d\ N\left(0, var\left(Y_1^2\right)\right) = N\left(0, \tau_4\right)$$

By the CLT and the WLLN

$$\sqrt{n}\ \overline{Y} \rightarrow_d N\left(0, \sigma^2\right) \tag{3}$$

and

$$\overline{Y} \rightarrow_p 0 \tag{4}$$

By (3) and (4) and Slutzky's Theorem

$$\sqrt{n}\ \overline{Y}^2 \rightarrow_p 0$$

Now

$$
\begin{aligned}
\sqrt{n}\left(\hat{\sigma}_n^2 - \sigma^2\right) &= \sqrt{n}\left[\overline{Y^2} - \overline{Y}^2 - \sigma^2\right] \\
&= \sqrt{n}\left[\overline{Y^2} - \sigma^2\right] - \sqrt{n}\ \overline{Y}^2 \\
&\rightarrow_d N\left(0, \tau_4\right) \qquad (5)
\end{aligned}
$$

EXAMPLE 3 (CONTINUED)

In the normal case, $\sqrt{n}\left(\hat{\sigma}_n^2 - \sigma^2\right) \to_d N\left(0, 2\sigma^4\right)$ because

$$
\begin{aligned}
\tau_4 &= var\left(Y_1^2\right) = var\left(\left(X_1 - \mu\right)^2\right) \\
&= \sigma^4 var\left(\left(\frac{X_1 - \mu}{\sigma}\right)^2\right) \\
&= 2\sigma^4 \quad \text{because} \quad \left(\frac{X_1 - \mu}{\sigma}\right)^2 \sim \chi^2_{(1)}.
\end{aligned}
$$

EXAMPLE 4

- Consider now the asymptotic distribution of $\hat{\sigma}$.

- That is, study the limiting behavior of

$$\sqrt{n}\left(\hat{\sigma}-\sigma\right)$$

- Recall that

$$\sqrt{n}\left(\hat{\sigma}_n^2-\sigma^2\right) \rightarrow_d N\left(0, \tau_4\right)$$

EXAMPLE 4 (CONTINUED)

Using the δ-method with

$$g(t) = \sqrt{t} \qquad g'(t) = \frac{1}{2\sqrt{t}}$$

$$g'\left(\sigma^2\right) = \frac{1}{2\sqrt{\sigma^2}} = \frac{1}{2\sigma}$$

We have

$$\sqrt{n}\left(\hat{\sigma} - \sigma\right) \to N\left(0, \frac{1}{4\sigma^2}\tau_4\right)$$

EXAMPLE 4 (CONTINUED)

In the normal case,

$$\sqrt{n}\left(\hat{\sigma}-\sigma\right) \rightarrow N\left(0, \frac{1}{4\sigma^2}2\sigma^4\right) = N\left(0, \frac{\sigma^2}{2}\right).$$