ST 501: Fundamentals of Statistical Inference

Convergence of Random Variables

Minh Tang

Fall 2024

Department of Statistics, North Carolina State University.

Convergence in probability

Definition
Let $\{X_n\}$ be a sequence of random variables. Then $\{X_n\}$ is said to converges in probability to $\alpha \in \mathbb{R}$ if and only if for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} P(|X_n - \alpha| > \epsilon) = 0.$$

There is a stronger notion of convergence almost surely, i.e, $\{X_n\}$ converges almost surely to $\alpha \in \mathbb{R}$ if for any $\epsilon > 0$ we have

$$P(\lim_{n\to\infty} |X_n - \alpha| > \epsilon) = 0.$$

Law of Large Numbers (LLN)

Theorem
Let $X_1, X_2, \ldots$ *be a sequence of iid random variables with finite mean* μ *and variance* σ^2*. Let* $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$*. Then* $\bar{X}_n$ *converges to* μ *almost surely.*

3

Proof of LLN (optional)

We only prove the convergence in probability of $\bar{X}_n$ to μ.

Recall that $\mathbb{E}[\bar{X}_n] = \mu$ and $\mathrm{Var}[\bar{X}_n] = \sigma^2/n$.

Then by Chebyshev's inequality we have for any $\epsilon > 0$ that

$$\mathbb{P}(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

Therefore, for any fixed $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| > \epsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$$

as desired.

Continuous mapping theorem

Theorem
Let $\{X_n\}$ converges in probability to α and g be a continuous function. Then $g(X_n)$ converges in probability to $g(\alpha)$.

See also problem 5.7 in your textbook.

Proof of continuous mapping theorem (optional)

The proof follows a standard δ and ϵ.

As g is continuous, for any $\epsilon > 0$ there exists a $\delta > 0$ depending on ϵ such that $|g(X_n) - g(\alpha)| < \epsilon$ for all $|X_n - \alpha| < \delta$.

We thus have, for any $\epsilon > 0$, that

$$\lim_{n\to\infty} P(|g(X_n) - g(\alpha)| < \epsilon) \geq \lim_{n\to\infty} P(|X_n - \alpha| < \delta) = 1$$

and thus $g(X_n)$ converges in probability to $g(\alpha)$.

Convergence in distribution

Definition
Let $\{X_n\}$ be a sequence of random variables. Let F_n be the cdf of X_n. Then $\{X_n\}$ is said to converge to a random variable X with cdf F if

$$\lim_{n\to\infty} F_n(z) = F(z)$$

at all point z for which F is continuous.

Central Limit Theorem

We now come to one of the most famous limit theorem in statistics. The version presented here is the easiest, albeit most restrictive. More general versions relax the assumption of identically distributed and existence of mgf.

Theorem

Let $X_1, X_2, \ldots$ *be a sequence of independent and identically distributed random variables. Suppose the mgf of* X_i *exists. Let* $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$. *Then*

$$\frac{\bar{X} - \mathbb{E}[\bar{X}]}{\sqrt{\mathrm{Var}[\bar{X}]}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$$

as $n \to \infty$. *Here* $\mathbb{E}[X_i] \equiv \mu$ *and* $\mathrm{Var}[X_i] \equiv \sigma^2$.

The convergence in the statement of the theorem means that, for any $a, b \in \mathbb{R}$, $a < b$,

$$\lim_{n\to\infty} P\Big(a \le \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le b\Big) = \Phi(b) - \Phi(a)$$

where Φ is the cdf of the standard normal.

In particular, we have, for n sufficiently large,

$$P\Big(|\bar{X}-\mu| \le \frac{k\sigma}{\sqrt{n}}\Big) \approx \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$$

Now

$$2\Phi(1) - 1 \approx 0.68, \quad 2\Phi(2) - 1 \approx 0.95,$$
$$2\Phi(3) - 1 \approx 0.997, \qquad 2\Phi(4) - 1 \approx 0.99994.$$

Comparing this with Chebyshev's inequality

$$P\Big(|\bar{X}-\mu| \le k\frac{\sigma}{\sqrt{n}}\Big) \ge 1 - \frac{1}{k^2},$$

the CLT gives a much tighter bound on $|\bar{X}-\mu|$.

9

Proof of the CLT (optional)

The proof is a straightforward use of mgf. Let $\tilde{X}_i = (X_i - \mu)/\sigma$. Then $\tilde{X}_i$ has mean 0 and variance 1. Furthermore

$$Z_n := \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{X}_i$$

Recall that, for any random variable Y, the mgf of $aY + b$ is

$$m_{aY+b}(t) = \mathbb{E}[\exp(taY + tb)] = e^{tb} m_Y(at)$$

As the $\tilde{X}_i$ are independent and identically distributed, we have

$$m_{Z_n}(t) = m_{\tilde{X}_1}(t/\sqrt{n}) \times m_{\tilde{X}_2}(t/\sqrt{n}) \times \cdots \times m_{\tilde{X}_n}(t/\sqrt{n}) = (m_{\tilde{X}}(t/\sqrt{n}))^n$$

where $m_{\tilde{X}}(t)$ is the mgf for any of the $\tilde{X}_i$.

10

We show that Z_n converges to standard normal by showing that $m_{Z_n}(t) \longrightarrow \exp(t^2/2)$, i.e., that

$$\lim_{n\to\infty} \log m_{Z_n}(t) = \lim_{n\to\infty} n \log m_{\tilde{X}}(t/\sqrt{n}) = \frac{t^2}{2}$$

Now as $n \to \infty$, $m_{\tilde{X}}(t/\sqrt{n}) \longrightarrow m_{\tilde{X}}(0) = 1$ and so

$$\lim_{n\to\infty} n \log m_{\tilde{X}}(t/\sqrt{n})$$

is of the form $\infty \times 0$ which is indeterminate.

We recall L'Hopital rule (for the indeterminate case)

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

11

Let $x = 1/\sqrt{n}$. We then have

$$\begin{aligned}
\lim_{n\to\infty} n \log m_{\tilde{X}}(t/\sqrt{n}) &= \lim_{x\to 0} \frac{\log m_{\tilde{X}}(tx)}{x^2} \\
&= \lim_{x\to 0} \frac{t m'_{\tilde{X}}(tx)/m_{\tilde{X}}(tx)}{2x} \\
&= \lim_{x\to 0} \frac{1}{2} \frac{t^2 m''_{\tilde{X}}(tx) m_{\tilde{X}}(tx) - t^2 (m'_{\tilde{X}}(tx))^2}{(m_{\tilde{X}}(tx))^2} = \frac{t^2}{2}
\end{aligned}$$

as $m'_{\tilde{X}}(0) = \mathbb{E}[\tilde{X}] = 0$ and $m''_{\tilde{X}}(0) = \mathbb{E}[\tilde{X}^2] = 1$.

12

EXAMPLE **F** *Normal Approximation to the Binomial Distribution*

Since a binomial random variable is the sum of independent Bernoulli random variables, its distribution can be approximated by a normal distribution. The approximation is best when the binomial distribution is symmetric—that is, when $p = \frac{1}{2}$. A frequently used rule of thumb is that the approximation is reasonable when $np > 5$ and $n(1 - p) > 5$. The approximation is especially useful for large values of n, for which tables are not readily available.

Suppose that a coin is tossed 100 times and lands heads up 60 times. Should we be surprised and doubt that the coin is fair?

13

CLT and delta method

The CLT can be combined with the delta method to yield more general CLTs.

Theorem
Suppose that $\sqrt{n}(X - \mu)$ *converges to a normal random variable with mean* 0 *and variance* σ^2. *Let* g *be a differentiable function with* $g'(\mu) > 0$. *Then*

$$\sqrt{n}(g(X) - g(\mu)) \xrightarrow{d} N(0, g'(\mu)^2 \times \sigma^2).$$

14

Example: Let $X_1, X_2, \ldots$ be iid Poisson rvs with rate λ. Let $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Find the limiting distribution for $(\bar{X}_n)^{1/2}$.

15

Example: Let $X_1, X_2, \ldots$ be iid Poisson rvs with rate λ. Let $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Find the limiting distribution for $(\bar{X}_n)^{1/2}$.

From the CLT we know

$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{d} N(0, \lambda)$$

Let $g(x) = x^{1/2}$. Then $g'(x) = \frac{1}{2x^{1/2}}$ and hence

$$\sqrt{n}(\bar{X}_n^{1/2} - \lambda^{1/2}) \xrightarrow{d} N(0, 1/4).$$

15

Example: Let $X_n \sim \mathrm{Bin}(n, p)$. Let $Y_n = X_n/n$. Find the limiting distribution for $\sqrt{Y_n(1 - Y_n)}$.

16

Example: Let $X_n \sim \mathrm{Bin}(n, p)$. Let $Y_n = X_n/n$. Find the limiting distribution for $\sqrt{Y_n(1 - Y_n)}$.

From the CLT we know

$$\sqrt{n}(Y_n - p) \xrightarrow{d} N(0, p(1 - p)).$$

Let $g(z) = \sqrt{z(1 - z)}$. Then $g'(z) = \frac{1-2z}{2\sqrt{z(1-z)}}$.

We therefore have

$$\sqrt{n}(g(Y_n) - \sqrt{p(1 - p)} \xrightarrow{d} N(0, \tfrac{1}{4}(1 - 2p)^2).$$

Q. What happens when $p = 1/2$?

16

Order statistics

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables. Let us consider a permutation of these X_i as

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

$X_{(k)}$ is the kth **order statistic**. These $\{X_{(k)}\}$ are **dependent**.

We are interested in the

1. marginal distribution of the minimum $X_{(1)}$
2. marginal distribution of the maximum $X_{(n)}$
3. marginal distribution of $X_{(k)}$ for a fixed k
4. joint distribution of $X_{(k)}$ and $X_{(\ell)}$ for fixed $k \neq \ell$.

Extremas

We start with the minimum.

Let x be given. Then $X_{(1)} \geq x$ if and only if $X_i \geq x$ for all i, i.e.,

$$P(X_{(1)} > x) = P(X_1 > x) \times \cdots \times P(X_n > x) = (1 - F(x))^n.$$

We therefore have

$$P(X_{(1)} \leq x) = 1 - (1 - F(x))^n$$

We can then derive the pmf for $X_{(1)}$ (assuming the X_i are discrete) or the pdf for $X_{(1)}$ (if the X_i are continuous).

When the X_i are continuous then $X_{(1)}$ has marginal pdf

$$f_{(1)}(x) = \frac{\mathrm{d}}{\mathrm{d}x} P(X_{(1)} \leq x) = n(1 - F(x))^{n-1} f(x)$$

The distribution of the maximum $X_{(n)}$ is argued similarly. For any given x, $X_{(n)} \leq x$ if and only if $X_i \leq x$ for all i, i.e.,

$$P(X_{(n)} \leq x) = P(X_1 \leq x) \times P(X_2 \leq x) \times \ldots P(X_n \leq x) = F(x)^n$$

19

The distribution of the maximum $X_{(n)}$ is argued similarly. For any given x, $X_{(n)} \leq x$ if and only if $X_i \leq x$ for all i, i.e.,

$$P(X_{(n)} \leq x) = P(X_1 \leq x) \times P(X_2 \leq x) \times \dots P(X_n \leq x) = F(x)^n$$

When the X_i are continuous then $X_{(n)}$ has marginal pdf

$$f_{(n)}(x) = nF(x)^{n-1} f(x).$$

19

Example: Extremas of uniform rvs

Let $X_1, \ldots, X_n$ be independent $\mathrm{Unif}(0,1)$. Then $X_{(n)}$ has cdf

$$P(X_{(n)} \leq x) = x^n$$

Let $x = 1 - \frac{\theta}{n}$ for $\theta > 0$. We then have

$$P\Big(X_{(n)} \leq 1 - \frac{\theta}{n}\Big) = \Big(1 - \frac{\theta}{n}\Big)^n \longrightarrow e^{-\theta}$$

as $n \to \infty$. Rearranging the above yield

$$P(n(1 - X_{(n)}) \geq \theta) \longrightarrow e^{-\theta}$$

Similarly

$$P(nX_{(1)} \geq \theta) \longrightarrow e^{-\theta}.$$

Conclusion The maximum and minimum of n independent $\mathrm{Unif}(0,1)$, properly transformed, converge to $\mathrm{Exp}(1)$.

21

Example: Extremas of exponential rvs

Let $X_1, \dots, X_n$ be independent $\mathrm{Exp}(1)$. Then $X_{(n)}$ has cdf

$$P(X_{(n)} \leq x) = (1 - e^{-x})^n$$

Let $x = \theta + \ln n$. Then, for $\theta > -\ln n$,

$$P(X_{(n)} \leq \theta + \ln n) = \big(1 - \exp(-(\theta + \ln n))\big) = \left(1 - \frac{e^{-\theta}}{n}\right)^n \longrightarrow e^{-e^{-\theta}}$$

Hence, properly transformed, $X_{(n)}$ converges to a Gumbel distribution with location $\mu = 0$ and scale $\beta = 1$.

22

In contrast, $X_{(1)}$ has cdf

$$P(X_{(1)} \geq x) = \left(e^{-x}\right)^n = e^{-nx}$$

which is an exponential with rate $1/n$. Taking $x = \frac{\theta}{n}$ we have

$$P(nX_{(1)} \geq \theta) = e^{-\theta}.$$

23

Example: Maximum of Gumbel rvs

Let $X_1, \ldots, X_n$ be independent Gumbel random variables with location $\mu = 0$ and scale $\beta = 1$. Then $X_{(n)}$ has cdf

$$P(X_{(n)} \leq x) = \exp(-e^{-x})^n = \exp(-ne^{-x})$$

Now taking $x = \theta + \ln n$ for $\theta \in \mathbb{R}$, we have

$$P(X_{(n)} \leq \theta + \ln n) = \exp(-ne^{-\theta - \ln n}) = e^{-e^{-\theta}}$$

which is again Gumbel distribution.

Note The Gumbel distribution turns out to be a quite special distribution. The maxima of iid normals is also Gumbel.

More interestingly, the Fisher-Tippett-Gnedenko theorem states that the maximum of n iid continuous rvs converge either to the Gumbel, Frechet, or Weibull distribution.

pdf and cdf of $X_{(k)}$

Given a fixed x, $X_{(k)} \leq x$ if and only if at least k of the X_i is less than or equal to x, i.e.,

$$P(X_{(k)} \leq x) = \sum_{i=k}^{n} \binom{n}{i} F(x)^i (1 - F(x))^{n-i}$$

When the X_i are continuous, $X_{(k)}$ has marginal pdf

$$\begin{aligned} f_{(k)}(x) &= \frac{\mathrm{d}}{\mathrm{d}x} P(X_{(k)} \leq x) \\ &= \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} (1 - F(x))^{n-k} f(x) \end{aligned}$$

after some tedious calculations.

25

Another (non-rigorous) argument for $f_{(k)}(x)$ goes as follows.

Consider a small interval $\mathcal{I} = (x - \epsilon, x + \epsilon]$.

Then $X_{(k)} \in \mathcal{I}$ if and only $(k-1)$ of the $X_i \leq x - \epsilon$ and $n - k$ of the $X_i \geq x + \epsilon$, i.e.,

$$P(X_{(k)} \in \mathcal{I}) = \frac{n! F(x-\epsilon)^{k-1}(F(x+\epsilon) - F(x-\epsilon))(1-F(x+\epsilon))^{n-k}}{(k-1)!(n-k)!}$$

We then have

$$\lim_{\epsilon \to 0} \frac{P(X_{(k)} \in \mathcal{I})}{2\epsilon} = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1}(1 - F(x))^{n-k} f(x).$$

26

We next derive the joint cdf for $X_{(k)}, X_{(\ell)}$ with $k < \ell$. Let $G(x, y) = P(X_{(k)} \leq x, X_{(\ell)} \leq y)$. If $x < y$ then

$$
\begin{aligned}
G(x,y) &= P(\text{at least } k \; X_i \leq x, \text{ at least } \ell \; X_i \leq y) \\
&= \sum_{s=\ell}^{n} \sum_{r=k}^{\ell} P(\text{exactly } r \; X_i \leq x, \text{ exactly } s \; X_i \leq y) \\
&= \sum_{s=\ell}^{n} \sum_{r=k}^{s} \tfrac{n!}{r!(s-r)!(n-s)!} F(x)^r (F(y) - F(x))^{s-r} (1 - F(y))^{n-s}
\end{aligned}
$$

Meanwhile, if $x \geq y$, then

$$P(X_{(k)} \leq x, X_{(\ell)} \leq y) = P(X_{(k)} \leq y, X_{(\ell)} \leq y) = P(X_{(\ell)} \leq y).$$

In the case when the X_i are continuous, taking partial derivatives (first wrt x then wrt y) of the above yield (after some quite tedious calculations)

$$f_{(k),(\ell)}(x,y) = \frac{n!F(x)^{k-1}(F(y)-F(x))^{\ell-1-k}(1-F(y))^{n-\ell}f(x)f(y)}{(k-1)!(\ell-1-k)!(n-\ell)!}$$

for $-\infty < x < y < \infty$.

28

Another (non-rigorous) argument goes as follows.

Consider an interval $\mathcal{I}_k = (x - \epsilon, x + \epsilon)$ and $\mathcal{I}_\ell = (y - \gamma, y + \gamma)$.

Then $X_{(k)} \in (x - \epsilon, x + \epsilon)$ and $X_{(\ell)} \in (y - \gamma, y + \gamma)$ if and only if

- $k - 1$ of the $X_i \leq x - \epsilon$
- $n - \ell$ of the $X_i \geq y + \gamma$
- the remaining $\ell - 1 - k$ of the $X_i \in (x + \epsilon, y - \gamma)$

29

That is to say

$$P(X_{(k)} \in \mathcal{I}_k, X_{(\ell)} \in \mathcal{I}_\ell)$$
$$= \frac{n! F(x^-)^{k-1}(F(x^+)-F(x^-))(F(y^-)-F(x^+))^{\ell-1-k}(F(y^+)-F(y^-))(1-F(y^+))^{n-k}}{(k-1)!(\ell-1-k)!(n-\ell)!}$$

and the joint pdf follows from evaluating

$$\lim_{\epsilon \to 0} \lim_{\gamma \to 0} \frac{P(X_{(k)} \in \mathcal{I}_k, X_{(\ell)} \in \mathcal{I}_\ell)}{2\epsilon \times 2\gamma}$$

30

Finally, using the same trick, we can obtain the joint pdf of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ whenever the X_i are continuous r.v.

As the X_i are continuous, $P(X_i = X_j) = 0$ for all $i \neq j$.

Let $x_1 < \cdots < x_n$ and consider intervals $\mathcal{I}_j = (x_j - \epsilon_j, x_j + \epsilon_j)$.

Then

$$P(X_{(1)} \in \mathcal{I}_1, \ldots, \mathcal{X}_{(n)} \in \mathcal{I}_{(n)}) = \frac{n!}{1!1!\ldots 1!} \prod_{i=1}^{n} (F(x_i + \epsilon_i) - F(x_i - \epsilon_i))$$

We then have

$$f_{X_{(1)}, X_{(2)}, \ldots, X_{(n)}}(x_1, x_2, \ldots, x_n) =$$
$$\lim_{\epsilon_1 \to 0, \epsilon_2 \to 0, \ldots, \epsilon_n \to 0} \frac{P(X_{(1)} \in \mathcal{I}_1, X_{(2)} \in \mathcal{I}_{(2)}, \ldots, X_{(n)} \in \mathcal{I}_{(n)})}{2^n \epsilon_1 \epsilon_2 \ldots \epsilon_n} =$$
$$n! f(x_1) f(x_2) \ldots f(x_n), \qquad x_1 < x_2 < \cdots < x_n.$$