Probability and Stochastic Processes I - Lecture 22

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/stac62/STAC622023.html

2023

- applications of probability theory are often concerned with approximations

- the underlying idea of "approximation" is the notion of a limit

- for example, for a sequence of real numbers $\{x_n : n \in \mathbb{N}\}$

Definition. The *limit* of $\{x_n : n \in \mathbb{N}\}$ exists if there is $x \in R^1$ such that for any $\varepsilon > 0$, there exists N_ε such that for all $n \geq N_\varepsilon$ then $|x_n - x| < \varepsilon$ and we write $\lim_{n\to\infty} x_n = x$.

then we approximate x by x_n for large n and try to say something about the error $|x_n - x|$ in this approximation

- if we have a sequence of r.v.'s $\{X_n : n \in \mathbb{N}\}$, then the *pointwise convergence* of X_n to r.v. X means $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for every $\omega \in \Omega$ but this is too strong and we weakened this to *convergence with probability 1*, namely, $X_n \overset{wp1}{\to} X$ if $P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$

- there are weaker forms of convergence that are useful

IV.1 Convergence in Distribution (Weak Convergence)

Definition IV.1.1 The sequence X_n of r.v.'s *converges in distribution* to r.v. X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for every continuity point x of the cdf F_X of X and we write $X_n \xrightarrow{d} X$. ■

- then $P_{X_n}((a,b]) = F_{X_n}(b) - F_{X_n}(a) \approx F_X(b) - F_X(a)$ for large n provided a, b are cty points of F_X

- so convergence in distribution is about approximating the distribution of a r.v. and not about approximating the value of the r.v.

Example IV.1.1 *Why restrict to convergence at continuity points of* F_X?

- suppose $P_{X_n}(\{-1/n\}) = P_{X_n}(\{1/n\}) = 1/2$ so

$$F_{X_n}(x) = \begin{cases} 0 & \text{if } x < -1/n \\ 1/2 & \text{if } -1/n \leq x < 1/n \\ 1 & \text{if } 1/n \leq x \end{cases}$$

- then as n gets bigger all the probability mass "piles up at 0" and let X be degenerate at 0 so

$$\begin{aligned} F_X(x) &= \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \leq x \end{cases} \\ \lim_{n\to\infty} F_{X_n}(x) &= \begin{cases} 0 & \text{if } x < 0 \\ 1/2 & \text{if } x = 0 \\ 1 & \text{if } 0 < x \end{cases} \end{aligned}$$

- so $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at every cty point of F_X but $\lim_{n\to\infty} F_{X_n}(0) \neq F_X(0)$ and 0 is not a cty point of F_X ■

Proposition IV.1.1 If $E(|X|^k) < \infty$, then $c_X(t) = \sum_{j=0}^{k} \frac{(it)^j}{j!}\mu_j + o(t^k)$ where the remainder $o(t^k)$ is a function of t satisfying $\lim_{t\to 0} o(t^k)/t^k = 0$.

Proof: We have, using integration by parts with $u = e^{is}, dv = (x-s)^n$, so $du = ie^{is}, v = -(x-s)^{n+1}/(n+1)$

$$\int_0^x (x-s)^n e^{is}\, ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1}\int_0^x (x-s)^{n+1} e^{is}\, ds \qquad (*)$$

and so

$$\begin{aligned} -i(e^{ix}-1) &= \int_0^x (x-s)^0 e^{is}\, ds \overset{\text{by } *}{=} x + i\int_0^x (x-s)^1 e^{is}\, ds \\ &\overset{\text{by } *}{=} x + \frac{ix^2}{2} + \cdots + \frac{i^{n-1}x^n}{n!} + \frac{i^n}{n!}\int_0^x (x-s)^n e^{is}\, ds \\ e^{ix} &= \sum_{j=0}^{n} \frac{(ix)^j}{j!} + \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}\, ds. \end{aligned}$$

Again, by *

$$\int_0^x (x-s)^{n-1} e^{is}\, ds = \frac{x^n}{n} + \frac{i}{n}\int_0^x (x-s)^n e^{is}\, ds \text{ which implies}$$

$$\int_0^x (x-s)^n e^{is}\, ds = \frac{n}{i}\left(\int_0^x (x-s)^{n-1} e^{is}\, ds - \frac{x^n}{n}\right) \text{ and}$$

$$e^{ix} = \sum_{j=0}^n \frac{(ix)^j}{j!} + \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is}-1)\, ds.$$

Therefore

$$\left| e^{ix} - \sum_{j=0}^n \frac{(ix)^j}{j!} \right| \le \min\left\{ \frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!} \right\} \text{ which implies}$$

$$\frac{1}{|t|^k}\left| c_X(t) - \sum_{j=0}^k \frac{(it)^j}{j!}\mu_j \right| = \frac{1}{|t|^k}\left| E\left(e^{itX} - \sum_{j=0}^k \frac{(itX)^j}{j!} \right) \right|$$

$$\le \frac{1}{|t|^k} E\left(\min\left\{ \frac{|tX|^{k+1}}{(k+1)!}, \frac{2|tX|^k}{k!} \right\} \right) = E\left(\min\left\{ \frac{|t||X|^{k+1}}{(k+1)!}, \frac{2|X|^k}{k!} \right\} \right)$$

and this upper bound is finite since $E(|X|^k) < \infty$ and goes to 0 as $t \to 0$ which proves the result. ■

Proposition IV.1.2 (*Continuity Theorem*) Suppose X_n is a sequence of r.v.'s. (i) If $X_n \xrightarrow{d} X$, then $c_{X_n}(t) \to c_X(t)$ for every t. (ii) If $c_{X_n}(t) \to c(t)$ for every t and c is continuous at 0, then c is the cf of a r.v. X such that $X_n \xrightarrow{d} X$.
Proof: Accept.

Proposition IV.1.3 (*Weak Law of Large Numbers*) If X_n is a sequence of i.i.d. r.v.'s with $E(X_i) = \mu \in R^1$, then

$$\frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{d} \mu \text{ (the r.v. with distribution generate at } \mu\text{)}.$$

Proof: Let X be degenerate at μ so $c_X(t) = \exp(it\mu)$ and note this is continuous at 0. Also,

$$\begin{aligned} c_{\frac{1}{n}S_n}(t) &= E\left(\exp\left(\frac{it}{n}\sum_{i=1}^{n} X_i\right)\right) \stackrel{i.i.d.}{=} c_{X_1}^n\left(\frac{t}{n}\right) \\ &= \left(1 + i\mu\frac{t}{n} + o\left(\frac{t}{n}\right)\right)^n \text{ (by Prop IV.1.1)} \\ &= \left(1 + i\mu\frac{t}{n}\right)^n \left(1 + \frac{o\left(\frac{t}{n}\right)}{1 + i\mu\frac{t}{n}}\right)^n \to \exp(it\mu) \end{aligned}$$

since, when $x_n \to 0$ and nx_n converges to a finite limit, then

$$\log(1 + x_n)^n = n\log(1 + x_n) = n(x_n - x_n^2/2 + x_n^3/3 - \cdots) \to \lim nx_n.$$

The result follows by the Continuity Theorem. ■

- the Strong Law of Large Numbers says

$$\frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{wp1}{\rightarrow} \mu$$

and we will prove that if $X_n \overset{wp1}{\rightarrow} X$, then $X_n \overset{d}{\rightarrow} X$ and so the the SLLN implies the WLLN

Proposition IV.1.4 (*The Central Limit Theorem*) If X_n is a sequence of i.i.d. r.v.'s with $E(X_i) = \mu \in R^1$, $Var(X_i) = \sigma^2$, then

$$Z_n = \frac{\frac{1}{n}S_n - \mu}{\sigma/\sqrt{n}} \overset{d}{\rightarrow} Z \sim N(0,1).$$

Proof: Note that

$$E\left(\frac{1}{n}S_n\right) = \mu, \qquad Var\left(\frac{1}{n}S_n\right) = \frac{\sigma^2}{n}$$

so Z_n has mean 0 and variance 1. Also $Y_i = (X_i - \mu)/\sigma$ has mean 0 and variance 1,

$$Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i$$

and $Y_1, \ldots, Y_n$ are i.i.d. Therefore

$$\begin{aligned} c_{Z_n}(t) &= c_{Y_1}^n\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(1 + \frac{it}{\sqrt{n}}E(Y_1) - \frac{t^2}{2n}E(Y_1^2) + o\left(\frac{t^2}{n}\right)\right)^n \text{ (by Prop IV.1.1)} \\ &= \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \to e^{-t^2/2} \end{aligned}$$

which is the cf of $Z \sim N(0,1)$ and the result follows by the Continuity Theorem. ■

Example IV.1.2 *Normal approximation to the binomial.*

- $X_1, X_2, \ldots$ i.i.d. Bernoulli(p), $E(X_i) = p$, $Var(X_i) = p(1-p)$ so $S_n \sim$ binomial(n, p)

- $\frac{1}{n}S_n =$ proportion of 1's in $X_1, X_2, \ldots, X_n$ then by CLT

$$\frac{\frac{1}{n}S_n - p}{\sqrt{p(1-p)/n}} \to N(0,1)$$

- so for large n with $Z \sim N(0,1)$

$$\begin{aligned}\Phi(b) - \Phi(a) &= P(a < Z \le b) \approx P\left(a < \frac{\frac{1}{n}S_n - p}{\sqrt{p(1-p)/n}} \le b\right) \\ &= P\left(np + a\sqrt{np(1-p)} < S_n \le np + b\sqrt{np(1-p)}\right)\end{aligned}$$

- note a, b reflect how long interval about mean is in terms of standard deviations ■

Example IV.1.3 *Poisson approximation to the binomial (rare events).*

- consider a situation where $X_1, X_2, \ldots, X_n$ i.i.d. Bernoulli(p_n) with $p_n = \lambda/n + o(1/n) \to 0$ with n (since $no(1/n) \to 0$, then $o(1/n) \to 0$)

- think of X_i as indicating whether or not, in n independent units, X_i is either on (1) or off (0) and the probability of being on is very small

- since $S_n \sim$ binomial$(n, \lambda/n + o(1/n))$, the expected number on is

$$np_n = \lambda + no(1/n) \to \lambda$$

- this permits working backwards from the expected number on to say $p_n = \lambda/n + o(1/n)$

- therefore,

$$
\begin{aligned}
P(S_n = k) &= \binom{n}{k}\left(\frac{\lambda}{n} + o(1/n)\right)^k \left(1 - \frac{\lambda}{n} - o(1/n)\right)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{\lambda^k}{k!}\left(1 + \frac{no(1/n)}{\lambda}\right)^k \left(1 - \frac{\lambda}{n}\right)^n \times \\
&\quad \left(1 - \frac{o(1/n)}{1 - \frac{\lambda}{n}}\right)^n \left(1 - \frac{\lambda}{n} - o(1/n)\right)^{-k} \\
&= \left[1\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{k}{n} + \frac{1}{n}\right)\right]\left(1 + \frac{no(1/n)}{\lambda}\right)^k \left(1 - \frac{o(1/n)}{1 - \frac{\lambda}{n}}\right)^n \times \\
&\quad \left(1 - \frac{\lambda}{n} - o(1/n)\right)^{-k} \frac{\lambda^k}{k!}\left(1 - \frac{\lambda}{n}\right)^n \to 1 \cdot 1 \cdot 1 \cdot 1 \cdot \frac{\lambda^k}{k!}e^{-\lambda} = \frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}
$$

using the expansion of $\log(1 + x_n)^n$ as in Prop.IV.1.3 for the limits

- so at any cty point of the Poisson(λ), say $y \in (k, k+1)$ where $k \in \mathbb{N}$

$$P(S_n \leq y) \rightarrow \sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda} = \text{ cdf of Poisson}(\lambda) \text{ at } y$$

which proves $S_n \xrightarrow{d}$ Poisson(λ)