STSCI 5080
Probability Models and Inference

Lecture 15: LLN and CLT

October 18, 2018

Convergence in probability

Definition

Random variables X_n converge in probability to another random variable X (that may be a constant) as $n \to \infty$, $X_n \xrightarrow{P} X$ in short, if

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

for any $\varepsilon > 0$.

Chebyshev's inequality

Theorem

Let X be a random variable with $E(X^2) < \infty$. Then

$$P(|X - E(X)| > x) \leq \frac{\mathrm{Var}(X)}{x^2}$$

for any $x > 0$.

LLN: setup

Suppose that we have a random sample from a cdf F:

$$X_1, \ldots, X_n \sim F \text{ i.i.d.}$$

and F has mean μ and variance $\sigma^2 > 0$.

Consider the sample mean

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

LLN

Theorem

We have $\overline{X}_n \xrightarrow{P} \mu$ *as* $n \to \infty$*, i.e.,*

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \varepsilon) = 0$$

for any $\varepsilon > 0$*.*

Proof

We know that $E(\overline{X}_n) = \mu$ and $\mathrm{Var}(\overline{X}_n) = \sigma^2/n$, so that by Chebyshev's inequality,

$$P(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

The right hand side $\to 0$ as $n \to \infty$. In addition, the probability is non-negative, so that by the sandwich rule, we have

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \varepsilon) = 0.$$

Example: Monte Carlo Integration

Suppose that we want to numerically evaluate

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x)f(x)dx \quad \text{for } X \sim f \text{ (pdf)}.$$

Assume that $E\{g(X)^2\} < \infty$.

- Generate $X_1, \ldots, X_n \sim f$ i.i.d.
- Random variables $g(X_1), \ldots, g(X_n)$ are i.i.d. as well.
- By LLN,

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{P} E\{g(X)\}.$$

  as $n \to \infty$.

Preliminary to CLT

CLT

Let $X_1, \ldots, X_n \sim F$ i.i.d. where F has mean μ and variance $\sigma^2 > 0$. If we normalize $\overline{X}_n$ as

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \qquad (*)$$

which has mean 0 and variance 1, is the cdf of (*) close to that of $N(0, 1)$?

Theorem

If $X_1, \ldots, X_n$ are independent and each X_i has $N(\mu_i, \sigma_i^2)$, then for any constants $\alpha_1, \ldots, \alpha_n$,

$$\sum_{i=1}^{n} \alpha_i X_i \sim N\left(\sum_{i=1}^{n} \alpha_i \mu_i, \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2\right).$$

Corollary

If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ i.i.d., then

$$\overline{X}_n \sim N(\mu, \sigma^2/n).$$

In particular,

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim N(0,1).$$

Proof of Theorem

The variable X_i has mgf

$$\psi_{X_i}(\theta) = e^{\mu_i\theta + \sigma_i^2\theta^2}.$$

By independence, the mgf of $Y = \sum_{i=1}^n \alpha_i X_i$ is

$$\begin{aligned}
\psi_Y(\theta) &= E(e^{\theta Y}) = E\left(e^{\theta \sum_{i=1}^n \alpha_i X_i}\right) = E\left(e^{\theta\alpha_1 X_1} \cdots e^{\theta\alpha_n X_n}\right) \\
&= E(e^{\theta\alpha_1 X_1}) \cdots E(e^{\theta\alpha_n X_n}) = \psi_{X_1}(\theta\alpha_1) \cdots \psi_{X_n}(\theta\alpha_n) \\
&= \exp\left(\theta \sum_{i=1}^n \alpha_i\mu_i + \theta^2 \sum_{i=1}^n \alpha_i^2\sigma_i^2/2\right).
\end{aligned}$$

This is the mgf of $N(\sum_{i=1}^n \alpha_i\mu_i, \sum_{i=1}^n \alpha_i^2\sigma_i^2)$.

Convergence in distribution

Definition

Let X_n and X be random variables with cdfs F_n and F, respectively. Then X_n converges in distribution to X, $X_n \xrightarrow{d} X$, if

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

at any continuity point of F. We also write $X_n \xrightarrow{d} F$, e.g., $X_n \xrightarrow{d} N(0,1)$.

Example 15.1

Example

Let X be a Bernoulli random variable with success probability p. The the cdf of X is

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} .$$

F is discontinuous at $x = 0$ and 1 (but continuous from right).

Example 15.2

Example

Let X be Bernoulli with success probability p, and $X_n = X + 1/n$. Does X_n converge in distribution to X?

Example 15.2

Example

Let X be Bernoulli with success probability p, and $X_n = X + 1/n$. Does X_n converge in distribution to X?

Answer: Yes. The cdf of X_n is

$$F_n(x) = P(X_n \leq x) = P(X \leq x - 1/n) = F(x - 1/n).$$

F is continuous at any $x \neq 0, 1$, and for any $x \neq 0, 1$, we have

$$\lim_{n\to\infty} F(x - 1/n) = F(x),$$

which implies that $X_n \xrightarrow{d} X$. However, at $x = 0, 1$,

$$\lim_{n\to\infty} F(x - 1/n) \neq F(x).$$

Example 15.3

Example

If $X_1, \ldots, X_n$, then show that $n(1 - X_{(n)}) \xrightarrow{d} Ex(1)$, where $X_{(n)} = \max_{1 \le i \le n} X_i$.

Example 15.3

Example

If $X_1, \ldots, X_n$, then show that $n(1 - X_{(n)}) \xrightarrow{d} Ex(1)$, where $X_{(n)} = \max_{1 \le i \le n} X_i$.

We know that the cdf of $X_{(n)}$ is

$$P(X_{(n)} \le x) = x^n \quad \text{for } 0 \le x \le 1.$$

Now, for $x \ge 0$,

$$\begin{aligned} P\{n(1 - X_{(n)}) \le x\} &= P(X_{(n)} \ge 1 - x/n) = 1 - P(X_{(n)} < 1 - x/n) \\ &= 1 - P(X_{(n)} \le 1 - x/n) = 1 - (1 - x/n)^n \to 1 - e^{-x}. \end{aligned}$$

If $x < 0$, then $P\{n(1 - X_{(n)}) \le x\} = 0$, and so $n(1 - X_{(n)}) \xrightarrow{d} Ex(1)$.

Theorem

If $X_n \xrightarrow{d} X$ and if X is continuous, then

$$\left.\begin{array}{l} P(a < X_n < b) \\ P(a \leq X_n < b) \\ P(a < X_n \leq b) \\ P(a \leq X_n \leq b) \end{array}\right. \rightarrow P(a < X < b)$$

for any $a < b$.

CLT

Denote by $\Phi(x)$ the cdf of $N(0,1)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

Theorem

Let $X_1, \ldots, X_n$ be a random sample from a cdf F, where F has mean μ and variance $\sigma^2 > 0$. Then

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1).$$

The CLT implies that

$$\left.\begin{array}{l} P(a < \sqrt{n}(\overline{X}_n - \mu)/\sigma < b) \\ P(a \leq \sqrt{n}(\overline{X}_n - \mu)/\sigma < b) \\ P(a < \sqrt{n}(\overline{X}_n - \mu)/\sigma \leq b) \\ P(a \leq \sqrt{n}(\overline{X}_n - \mu)/\sigma \leq b) \end{array}\right. \to \Phi(b) - \Phi(a)$$

for any $a < b$.
We note that

$$\Phi(-x) = 1 - \Phi(x)$$

for any $x > 0$.

Example 15.4

Example

Let $X_1, \ldots, X_{12} \sim U[0,1]$ i.i.d. Use CLT to approximate $P(|\overline{X}_{12} - 1/2| < 0.1)$.

Example 15.4

Example

Let $X_1, \ldots, X_{12} \sim U[0,1]$ i.i.d. Use CLT to approximate $P(|\overline{X}_{12} - 1/2| < 0.1)$.

We have

$$\mu = \frac{1}{2} \quad \text{and} \quad \sigma^2 = \frac{1}{12},$$

so that

$$P(12|\overline{X}_{12} - 1/2| < x) \approx \Phi(x) - \Phi(-x) = 2\Phi(x) - 1.$$

Hence,

$$\begin{aligned} P(|\overline{X}_{12} - 1/2| < 0.1) &= P(12|\overline{X}_{12} - 1/2| < 1.2) \\ &\approx 2\Phi(1.2) - 1 \\ &\approx 0.729. \end{aligned}$$

Theorem

if $Y_n \sim Bin(n, p)$, *then*

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

Proof

By definition, $Y_n = X_1, \ldots, X_n$ for independent Bernoulli trials $X_1, \ldots, X_n$ with success probability p. We note that $E(X_1) = p$ and $\mathrm{Var}(X_1) = p(1-p)$, and so

$$\frac{Y_n - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\overline{X}_n - p)}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

by CLT.

Example 15.5

Example

If $Y \sim Bin(100, 1/2)$, then use CLT to approximate $P(Y > 60)$.

Example 15.5

Example

If $Y \sim Bin(100, 1/2)$, then use CLT to approximate $P(Y > 60)$.

In this case, $np = 50$ and $np(1-p) = 25$. Hence,

$$P(Y > 60) = P\left(\frac{Y-50}{5} > \frac{60-50}{5}\right) \approx 1 - \Phi(2) \approx 0.023.$$

Proof of CLT

Theorem (Continuity theorem for mgfs)

Let X_n and X have mgfs ψ_n and ψ, respectively. If $\psi_n(\theta) \to \psi(\theta)$ for any θ in an open interval containing the origin, then $X_n \xrightarrow{d} X$.

Proof of CLT

Theorem (CLT)

Let $X_1, \ldots, X_n$ be a random sample from a cdf F, where F has mean μ and variance $\sigma^2 > 0$. Then

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1).$$

Proof of CLT using mgfs

Let

$$Y_i = \frac{X_i - \mu}{\sigma}, \; i = 1, \ldots, n.$$

We note that $Y_1, \ldots, Y_n$ are i.i.d. with mean zero and unit variance, and

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} = \sqrt{n}\overline{Y}_n.$$

Denote by $\psi(\theta)$ the mgf of Y_1: $\psi(\theta) = E(e^{\theta Y_1})$. The mgf of $\sqrt{n}\overline{Y}_n$ is

$$\begin{aligned}\psi_n(\theta) &= E(e^{\theta \sum_{i=1}^n Y_i/\sqrt{n}}) = E(e^{\theta Y_1/\sqrt{n}} \cdots e^{\theta Y_n/\sqrt{n}}) \\ &= E(e^{\theta Y_1/\sqrt{n}}) \cdots E(e^{\theta Y_n/\sqrt{n}}) = \{\psi(\theta/\sqrt{n})\}^n.\end{aligned}$$

Now, since $\psi'(0) = E(Y_1) = 0$ and $\psi''(0) = E(Y^2) = 1$, we can expand $\psi(\theta)$ as

$$\begin{aligned}\psi(\theta) &= \psi(0) + \psi'(0)\theta + \frac{\theta^2}{2}\psi''(0) + \theta^2 R(\theta) \\ &= 1 + \frac{\theta^2}{2} + \theta^2 R(\theta)\end{aligned}$$

by Taylor's theorem, where $\lim_{\theta\to 0} R(\theta) = 0$. Substituting this expansion, we have

$$\psi_n(\theta) = \{\psi(\theta/\sqrt{n})\}^n = \left(1 + \frac{\theta^2}{2n} + \frac{\theta^2}{n}R(\theta/\sqrt{n})\right)^n \to e^{\theta^2/2},$$

which is the mgf of $N(0,1)$. By the continuity theorem, we have $\sqrt{n}\overline{Y}_n \xrightarrow{d} N(0,1)$.