Mathematical Statistics and Data Analysis

Lecture 6: Review of Probability - Part V

Lyu Ni

DaSE@ECNU
(lni@dase.ecnu.edu.cn)

October 12, 2019

Outlines

1. Convergence
   - Convergence in Probability
   - Convergence in Distribution
2. Law of Large Numbers
   - Central Limit Theorem
3. m.g.f & ch.f

Reading Material

Textbook:

- Rice: Chapter 5;
- Mao: Chapter 4;

Motivative

We want to specifically examine each product and determine whether it is conforming or not.

- X_i: the performance of the ith product.
- If it is nonconforming, $X_i = 1$; otherwise, $X_i = 0$.
- $X_i \sim B(p), i = 1, 2, \cdots$, where p is the fraction defective in a population.

We wonder how many defective products among the n products. Let $S_n = \sum_{i=1}^{n} X_i$. We need to compute the probability

$$P(a \leq S_n \leq b) = P(a \leq X_1 + X_2 + \cdots + X_n \leq b)$$

It is known that $S_n \sim b(n, p)$. But it is still difficult to compute this probability when n is large (for example, $n = 1000$ or $n = 10000$).

Motivative (Con'd)

Suppose that Y is a random variable such that

$$P(a \leq S_n \leq b) \approx P(a \leq Y \leq b)$$

and it is relatively simple to compute the probability $P(a \leq Y \leq b)$. So, it is a good idea that we use the probability $P(a \leq Y \leq b)$ to approximate the probability $P(a \leq S_n \leq b)$.

Question

When can we say

$$P(a \leq S_n \leq b) \approx P(a \leq Y \leq b)?$$

Convergence in Probability

Definition

Let $X_1, X_2, \cdots$ be a sequence of random variables and let X be a random variable. X_n is said to **converge in probability** to X if, for any $\varepsilon > 0$,

$$P(|X_n - X| \geq \varepsilon) \to 0 \text{ as } n \to \infty$$

or

$$P(|X_n - X| < \varepsilon) \to 1 \text{ as } n \to \infty$$

- Notation: $X_n \xrightarrow{P} X$.
- Special Case: If $P(X = c) = 1$, then $X_n \xrightarrow{P} c$.

Convergence in Probability

Property

Suppose $\{X_n\}$ and $\{Y_n\}$ are two sequence of random variables. a and b are two constant. If

$$X_n \xrightarrow{P} a \text{ and } Y_n \xrightarrow{P} b,$$

then

- $X_n \pm Y_n \xrightarrow{P} a \pm b$;
- $X_n \times Y_n \xrightarrow{P} a \times b$;
- $X_n / Y_n \xrightarrow{P} a/b (b \neq 0)$.

Convergence in Distribution

Definition

Let $X_1, X_2, \cdots$ be a sequence of random variables with cumulative distribution functions $F_1(x), F_2(x), \cdots$ and let X be a random variable with distribution function $F(x)$. X_n is said to **converge in distribution** to X if

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

at every point at which F is continuous.

- Notation: $X_n \xrightarrow{L} X$ or $F_n(x) \xrightarrow{W} F(x)$.

Theorem

If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{L} X$; but not vice versa.

Convergence in Distribution

Example

Suppose the p.m.f. of a random variable X is

$$P(X=-1)=\frac{1}{2} \text{ and } P(X=1)=\frac{1}{2}$$

Let $X_n=-X$. X_n and X have the same distribution, that is, the c.d.f.s of X_n and X are the same. Thus, $X_n \xrightarrow{L} X$. But, for any $0<\varepsilon<2$, we have

$$P(|X_n-X|\geq\varepsilon)=P(2|X|\geq\varepsilon)=1\nrightarrow 0.$$

Thus, X_n does not converge in probability to X.

Theorem

Suppose that c is a constant. $X_n \xrightarrow{P} c$ if and only if $X_n \xrightarrow{L} c$.

Convergence in Distribution

Example

Suppose the p.m.f. of a random variable X is

$$P(X=-1)=\frac{1}{2} \text{ and } P(X=1)=\frac{1}{2}$$

Let $X_n = -X$. X_n and X have the same distribution, that is, the c.d.f.s of X_n and X are the same. Thus, $X_n \xrightarrow{L} X$. But, for any $0 < \varepsilon < 2$, we have

$$P(|X_n - X| \geq \varepsilon) = P(2|X| \geq \varepsilon) = 1 \nrightarrow 0.$$

Thus, X_n does not converge in probability to X.

Theorem

Suppose that c is a constant. $X_n \xrightarrow{P} c$ if and only if $X_n \xrightarrow{L} c$.

Law of Large Numbers

Let $X_1, X_2, \cdots$ be a sequence of random variables and let

$$p = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i).$$

Definition

The **Law of Large Number (LLN)** states that, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right) = 1,$$

- This is also called the **Weak Law of Large Numbers**.

Law of Large Numbers

Bernoulli Law of Large Numbers

Suppose that $X_1, X_2, \cdots$ is a sequence of independently and identically distributed Bernoulli random variables and the mean $p = E(X_1)$. Then, for any $\varepsilon > 0$, we have

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| < \varepsilon\right) = 1$$

Proof: Since $\sum_{i=1}^{n} X_i \sim b(n, p)$,

$$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = p \text{ and } Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{p(1-p)}{n}$$

Law of Large Numbers

Bernoulli Law of Large Numbers (Con'd)

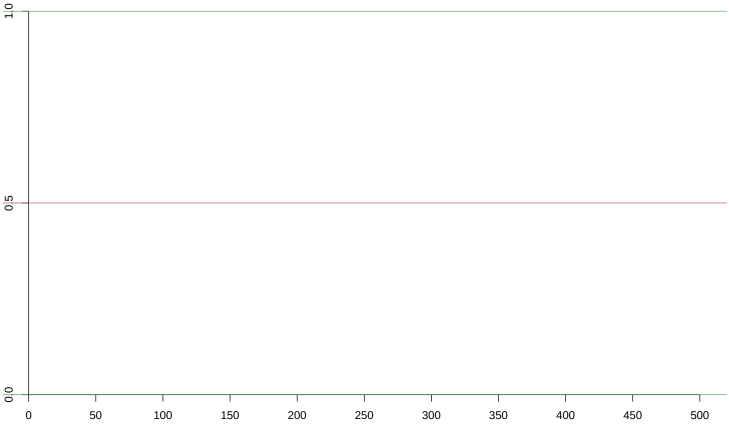
From the Chebyshev's inequality,

$$
\begin{aligned}
1 \geq P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| < \varepsilon\right) &\geq 1 - \frac{Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{\varepsilon^2} \\
&= 1 - \frac{p(1-p)}{n\varepsilon^2} \to 1,
\end{aligned}
$$

as $n \to \infty$. Therefore,

$$
\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| < \varepsilon\right) = 1
$$

Law of Large Numbers



Law of Large Numbers

Chebyshev Law of Large Numbers

Suppose that $\{X_n\}$ is a sequence of pairwise uncorrelated random variables. If the variance of X_i exists, and has an upper bound, that is, $Var(X_i) \leq c, i = 1, 2, \cdots$, then, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right) = 1.$$

Proof: Since $X_1, X_2, \cdots$ are pairwise uncorrelated,

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \leq \frac{c}{n}.$$

Law of Large Numbers

Chebyshev Law of Large Numbers (Con'd)

From the Chebyshev's inequality, for any $\varepsilon > 0$, we have

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| < \varepsilon\right) \geq 1 - \frac{Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{\varepsilon^2} \geq 1 - \frac{c}{n\varepsilon^2}$$

Thus,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| < \varepsilon\right) = 1$$

Law of Large Numbers

Markov Law of Large Numbers

If

$$\frac{1}{n} Var\left(\sum_{i=1}^{n} X_i\right) \to 0,$$

then, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right) = 1.$$

Proof: The result is easily obtained because of the Chebyshev's inequality.

Remark

$\frac{1}{n} Var(\sum_{i=1}^{n} X_i) \to 0$ is called the **Markov Condition**.

Law of Large Numbers

Example

Suppose that $\{X_n\}$ is a sequence of identically distributed random variables with a finite variance and X_n is correlated with X_{n-1} and X_{n+1} and uncorrelated with other X_i's. As we know,

$$\frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left\{\sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n-1} Cov(X_i, X_{i+1})\right\}$$

Let $Var(X_n) = \sigma^2$. Then $|Cov(X_i, X_j)| \leq \sigma^2$ and

$$\frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right) \leq \frac{1}{n^2}(n\sigma^2 + 2(n-1)\sigma^2) \to 0 \text{ as } n \to \infty$$

Thus, the Markov Condition is satisfied.

Law of Large Numbers

Wiener-Khinchin Law of Large Numbers

Let $\{X_n\}$ be a sequence of independently and identically distributed random variables. If $E(X_i)$ exists, then, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right) = 1.$$

Remark

Let $\{X_n\}$ be a sequence of independently and identically distributed random variables. If $E(|X_i|^k)$ exists where k is a positive integer, then, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i^k - \frac{1}{n}\sum_{i=1}^{n} E(X_i^k)\right| < \varepsilon\right) = 1.$$

Law of Large Numbers

Question:

Suppose that $0 \le f(x) \le 1$. Calculate the integral

$$J = \int_0^1 f(x)\mathrm{d}x$$

Solution - Method One Suppose that a random bivariate vector (X, Y) is distributed as a bivariate uniform distribution on $\{0 \le x \le 1, 0 \le y \le 1\}$. That is, $X \sim U(0,1)$, $Y \sim U(0,1)$ and X and Y are independent. Let

$$A = \{Y \le f(X)\}.$$

Then, the probability of A is

$$p = P(Y \le f(X)) = \int_0^1 \int_0^{f(x)} \mathrm{d}y\mathrm{d}x = \int_0^1 f(x)\mathrm{d}x = J$$

Law of Large Numbers

Method One (Con'd)

The algorithm is as follows:

- Generate $2n$ random numbers from $U(0,1)$;
- Randomly select two random numbers and obtain n pairs, that is, $(x_i, y_i), i = 1, 2, \cdots, n$;
- Let

$$S_n = \sum_{i=1}^{n} I\{y_i \le f(x_i)\}$$

- The integral J is estimated by S_n/n.

Law of Large Numbers

Question:

Suppose that $0 \leq f(x) \leq 1$. Calculate the integral

$$J = \int_0^1 f(x)\mathrm{d}x$$

Solution - Method Two Suppose that X is a uniform random variable on $(0, 1)$. The expectation of $Y = f(X)$ is

$$E(f(X)) = \int_0^1 f(x) \cdot 1\mathrm{d}x = J$$

Then, J is approximated by the estimate of the expectation of $f(X)$. By Wiener-Khinchin LLN, the mean of observations of $f(X)$ can be used as an estimate of the expectation of $f(X)$.

Law of Large Numbers

Method Two (Con'd)

The algorithm is as follows:

- Generate n random numbers from $U(0,1)$, $x_i, i = 1, 2, \cdots, n$;
- Calculate $f(x_i)$;
- Estimate J by

$$J \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

Central Limit Theorem

Motivate

If $X_1, X_2, \cdots$ is a sequence of independent random variables with mean μ and variance σ^2, and if

$$S_n = \sum_{i=1}^{n} X_i,$$

we know from WLLN that S_n/n converges to μ in probability. This followed from the fact that

$$Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2} Var(S_n) = \frac{\sigma^2}{n} \to 0$$

The central limit theorem is concerned not with the fact that the ratio S_n/n converges to μ but with how it fluctuates around μ.

Central Limit Theorem

Motivate (Con'd)

To analyze these fluctuations, we standardize

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

It can be verified that Z_n has mean 0 and variance 1. The central limit theorem states that the distribution of Z_n converges to the standard normal distribution.

Central Limit Theorem

De Moivre–Laplace Central Limit Theorem

Suppose that $X_1, X_2, \cdots$ is a sequence of independently and identically distributed Bernoulli random variables $B(p)$. Let

$$Z_n = \frac{S_n - np}{\sqrt{npq}}$$

where $q = 1 - p$. Then, for every real value z, we have

$$\lim_{n \to \infty} P(Z_n \leq z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} \mathrm{d}t$$

Central Limit Theorem

Question: How to use De Moivre–Laplace CLT?

Let $\beta = \Phi(z)$. From De Moivre–Laplace CLT,

$$P(Z_n \leq z) \approx \Phi(z) = \beta$$

- Given n, z, find the probability β;
- Given n, β, find the quantile z;
- Given z, β, find the sample size n;

Central Limit Theorem

Suppose that $X_1, X_2, \cdots$ is a sequence of independent random variables with finite expectations and finite variances, i.e.

$$E(X_i) = \mu_i, Var(X_i) = \sigma_i^2, i = 1, 2, \cdots$$

Let $S_n = \sum_{i=1}^n X_i$. Since

$$\begin{aligned} E(S_n) &= \mu_1 + \mu_2 + \cdots + \mu_n \\ \sigma(S_n) &= \sqrt{Var(S_n)} = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2} \end{aligned}$$

and let $\sigma(S_n) = B_n$, then S_n is standardized by

$$Z_n = \frac{S_n - (\mu_1 + \mu_2 + \cdots + \mu_n)}{B_n} = \sum_{i=1}^n \frac{X_i - \mu_i}{B_n}$$

Central Limit Theorem

Lindeberg Central Limit Theorem

Suppose that $\{X_n\}$ is a sequence of independent random variables satisfying, for any $\tau > 0$,

$$\lim_{n\to\infty} \frac{1}{\tau^2 B_n^2} \sum_{i=1}^{n} \int_{|x-\mu_i|>\tau B_n} (x-\mu_i)^2 f_i(x)\mathrm{d}x = 0.$$

Then, for every z, we have

$$\lim_{n\to\infty} P\left(\frac{1}{B_n}\sum_{i=1}^{n}(X_i-\mu_i) \le z\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-t^2/2}\mathrm{d}t.$$

Remark

This condition is also called **Lindeberg Condition**.

Central Limit Theorem

Consider that

$$Z_n = \sum_{i=1}^{n} \frac{X_i - \mu_i}{B_n}.$$

If $\frac{X_i - \mu_i}{B_n}$ is uniformly small, i.e. the probability of the event that A_{ni} occurs is small or approaches to zero where

$$A_{ni} = \left\{ \left| \frac{X_i - \mu_i}{B_n} \right| > \tau \right\} = \{ |X_i - \tau_i| > \tau B_n \}$$

for any $\tau > 0$. Thus, a sufficient condition is that

$$\lim_{n \to \infty} P \left(\max_{1 \le i \le n} |X_i - \mu_i| > \tau B_n \right) = 0$$

Central Limit Theorem

It is a fact that

$$
\begin{aligned}
P\left(\max_{1\le i\le n}|X_i-\mu_i|>\tau B_n\right) &= P\left(\cup_{i=1}^n(|X_i-\mu_i|>\tau B_n)\right) \\
&\le \sum_{i=1}^n P(|X_i-\mu_i|>\tau B_n)
\end{aligned}
$$

If X_i's are continuous r.v.s with a p.d.f. $f_i(x)$, then,

$$
\begin{aligned}
&\sum_{i=1}^n P(|X_i-\mu_i|>\tau B_n) \\
=\ &\sum_{i=1}^n \int_{|x-\mu_i|>\tau B_n} f_i(x)\mathrm{d}x \\
\le\ &\frac{1}{\tau^2 B_n^2}\sum_{i=1}^n \int_{|x-\mu_i|>\tau B_n}(x-\mu_i)^2 f_i(x)\mathrm{d}x
\end{aligned}
$$

Lyapunov Central Limit Theorem

Suppose that $\{X_n\}$ is a sequence of independent random variables. If there exists $\delta > 0$ such that

$$\lim_{n\to\infty} \frac{1}{B_n^{2+\delta}} \sum_{i=1}^{n} E(|X_i - \mu_i|^{2+\delta}) = 0,$$

then, for every z, we have

$$\lim_{n\to\infty} P\left(\frac{1}{B_n}\sum_{i=1}^{n}(X_i - \mu_i) \le z\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-t^2/2}\mathrm{d}t$$

where $\mu_i = E(X_i)$ and $B_n = \sqrt{\sum_{i=1}^{n} Var(X_i)}$

Moment-generating function

Definition

The **moment-generating function(m.g.f.)** of a random variable X is

$$M(t) = E(e^{tX})$$

if the expectation is defined.

- If X is discrete, $M(t) = \sum_x e^{tx} f(x)$;
- If X is continuous, $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) \mathrm{d}x$

Property

If the m.g.f. exists in an open interval containing zero, then

$$M^{(r)}(0) = E(X^r)$$

Characteristic function

Definition

The **characteristic function (ch.f.)** of a random variable X is

$$\phi(t) = E(e^{itX}) = E(\cos(tX) + iE(\sin(tX))$$

where $i = \sqrt{-1}$.

- If X is discrete, $\phi(t) = \sum_x e^{itx} f(x)$;
- If X is continuous, $\phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) \mathrm{d}x$

Remark

Since $|e^{itX}| \leq 1$, the ch.f. is thus defined for all distributions. But the m.g.f. may not exist for some distributions.