Chapter 4 Probability Limit Theorems

In the early days, the aim of probability theory is to reveal the inherent rule of random phenomena caused by a large number of random factors.

Bernoulli first recognized the importance to study an infinite sequence of random trials, and established the first limit theorem in probability theory—the law of large numbers.

4.1 Convergence in distribution and central limit theorems

de Moivre and Laplace presented that the observed error can be regard as the summation of a large number of independent and slight errors, and proved that the distribution of the observed error is approximated by a normal distribution—the central limit theorem.

Central limit theorems

Let S_n denote the number of successes in n Bernoulli trials, then $P(S_n = k) = b(k; n, p)$. In practice, people are usually interested to calculate

$$P(\alpha < S_n \le \beta) = \sum_{\alpha < k \le \beta} b(k; n, p).$$

The computation of the right hand side of the equality is generally very complex. However, it is found by de Moivre and Laplace that the binomial distribution can be well approximated by normal distribution when $n \longrightarrow \infty$.

Theorem

(de Moivre-Laplace) Let $\Phi(x)$ be the standard normal distribution function. We have for $-\infty < x < \infty$,

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{npq}} \le x\right) = \Phi(x).$$

When n is big enough, p is moderate, then

$$\begin{aligned} & P(\alpha \le S_n \le \beta) \\ = \ & P\Big(\frac{\alpha - np}{\sqrt{npq}} < \frac{S_n - np}{\sqrt{npq}} \le \frac{\beta - np}{\sqrt{npq}}\Big) \\ \approx \ & \Phi\Big(\frac{\beta - np}{\sqrt{npq}}\Big) - \Phi\Big(\frac{\alpha - np}{\sqrt{npq}}\Big). \end{aligned}$$

When n is big enough, p is moderate, then

$$\begin{aligned} & P(\alpha \le S_n \le \beta) \\ = \; & P\left(\frac{\alpha - np}{\sqrt{npq}} < \frac{S_n - np}{\sqrt{npq}} \le \frac{\beta - np}{\sqrt{npq}}\right) \\ \approx \; & \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right). \end{aligned}$$

如果把α和β替换成$\alpha - \frac{1}{2}$和$\beta + \frac{1}{2}$，则概率的近似结果更加精确，

$$P(\alpha \le S_n \le \beta) \approx \Phi\left(\frac{\beta + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - \frac{1}{2} - np}{\sqrt{npq}}\right).$$

Remark. Suppose $S_n \sim B(n,p)$.

- $np_n \to \lambda$, then
$$S_n \dot{\sim} P(np),$$
(in practical, if p is close to 0 (or 1), and np is not big (or not small), we use $P(\lambda)$ to approximate $B(n,p)$);
- fixed $0 < p < 1$, as $n \to \infty$,
$$S_n \dot{\sim} N(np, npq),$$
(in practical, if p is moderate, we use the normal distribution to approximate $B(n,p)$).

Definition

Let $\{\xi_n, n \geq 1\}$ be a sequence of random variables. If there exist two sequences of constants $B_n > 0$ and A_n such that

$$P\Big(\frac{1}{B_n}\sum_{k=1}^{n}\xi_k - A_n \leq x\Big) \to \Phi(x), \ \forall x,$$

then we say that $\{\xi_n\}$ obeys the central limit theorem.

Theorem

(Lindeberg-Lévy) Let $\{\xi_n, n \geq 1\}$ be a sequence of independent and identically distributed random variables. Let $S_n = \Sigma_{k=1}^n \xi_k$, $E\xi_1 = a$, $Var\xi_1 = \sigma^2$. Then the central limit theorem holds true, i.e., for any x,

$$P\left(\frac{S_n - na}{\sqrt{n}\sigma} \leq x\right) \to \Phi(x) \quad as \quad n \to \infty.$$

Jarl Waldemar Lindeberg

(August 1876– December 1932, Finnish)

Lemma

Suppose $F_n(x)$ and $F(x)$ have c.f.s $f_n(t)$ and $f(t)$, respectively. If $f_n(t) \to f(t)$ for each t, then

$$\text{for each continuous point } x \text{ of } F, \quad F_n(x) \to F(x).$$

Proof.

Proof. Let $f(t)$ and $f_n(t)$ be c.f.s of $\xi_1 - a$ and $\frac{S_n - na}{\sqrt{n}\sigma}$ respectively. Since $\xi_1, \xi_2, \cdots, \xi_n$ are i.i.d., we have $f_n(t) = (f(\frac{t}{\sqrt{n}\sigma}))^n$. And note that $E\xi_1 = a$, $Var\xi_1 = \sigma^2$, so the c.f. $f(t)$ has continuous derivative of 2-order, and $f'(0) = 0$, $f''(0) = -\sigma^2$. Using Taylor's expansion for f, we have

Proof. Let $f(t)$ and $f_n(t)$ be c.f.s of $\xi_1 - a$ and $\frac{S_n - na}{\sqrt{n}\sigma}$ respectively. Since $\xi_1, \xi_2, \cdots, \xi_n$ are i.i.d., we have $f_n(t) = (f(\frac{t}{\sqrt{n}\sigma}))^n$. And note that $E\xi_1 = a$, $Var\xi_1 = \sigma^2$, so the c.f. $f(t)$ has continuous derivative of 2-order, and $f'(0) = 0$, $f''(0) = -\sigma^2$. Using Taylor's expansion for f, we have

$$
\begin{aligned}
f(x) &= f(0) + f'(0)x + \frac{1}{2} f''(0)x^2 + o(x^2) \\
&= 1 - \frac{\sigma^2}{2}x^2 + o(x^2) \quad \text{as} \quad x \to 0.
\end{aligned}
$$

For given $t \in \mathbf{R}$,

For given $t \in \mathbf{R}$,

$$f\left(\frac{t}{\sqrt{n}\sigma}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \quad \text{as} \quad n \to \infty.$$

For given $t \in \mathbf{R}$,

$$f\left(\frac{t}{\sqrt{n}\sigma}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \quad \text{as} \quad n \to \infty.$$

Therefore

$$f_n(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{-\frac{t^2}{2}}.$$

The later is the c.f. of $N(0,1)$. Then Theorem 6 follows from the Lemma.

Example

When we do approximate calculation, the original data x_k rounds off to the m-th decimal place. In this way, the rounding error ξ_k can be regarded as a uniformly distributed random variable in $(-0.5 \cdot 10^{-m}, 0.5 \cdot 10^{-m}]$. If we obtain the sum $\sum_{k=1}^n x_k$ of n $x_k's$, how about the error according to the rounding principle?

Example

When we do approximate calculation, the original data x_k rounds off to the m-th decimal place. In this way, the rounding error ξ_k can be regarded as a uniformly distributed random variable in $(-0.5 \cdot 10^{-m}, 0.5 \cdot 10^{-m}]$. If we obtain the sum $\sum_{k=1}^{n} x_k$ of n $x_k's$, how about the error according to the rounding principle?

One may usually estimate the error of $\sum_{k=1}^{n} x_k$ by the sum of $\xi_k's$ upper bounds, that is $0.5 \cdot n \cdot 10^{-m}$. When n is very big, this number is also very big.

In fact, possibility that the error is so big is very small. Since $\{\xi_k\}$ are independent and identically distributed and $E\xi_k = 0$, $Var\xi_k = \sigma^2 = 10^{-2m}/12$, we have

In fact, possibility that the error is so big is very small. Since $\{\xi_k\}$ are independent and identically distributed and $E\xi_k = 0$, $Var\xi_k = \sigma^2 = 10^{-2m}/12$, we have

$$P(|\sum_{k=1}^{n}\xi_k| \le x\sqrt{n}\sigma) \approx 2\Phi(x) - 1$$

by Theorem 6. The above probability is 0.997 when $x = 3$. The probability that the error of the sum exceeds $3\sigma\sqrt{n} = 0.5 \cdot \sqrt{3} \cdot \sqrt{n} \cdot 10^{-m}$ is only 0.003. Obviously, for large n, the error bound is far smaller than $0.5 \cdot n \cdot 10^{-m}$.

Non i.i.d. case: Let $B_n^2 = \sum_{k=1}^n Var\xi_k$.

Theorem

(Lindeberg-Feller) Suppose that $\{\xi_k, k \geq 1\}$ is a sequence of indept. r.v.s. If the Lindeberg condition is satisfied:

$$\frac{1}{B_n^2}\sum_{k=1}^n \int_{|x-E\xi_k|\geq \varepsilon B_n} (x-E\xi_k)^2 dF_k(x) \to 0 \ \ \forall \epsilon > 0, \tag{1}$$

then for each x,

$$P\left(\frac{\sum_{k=1}^n (\xi_k - E\xi_k)}{B_n} \leq x\right) \to \Phi(x). \tag{2}$$

Note

$$
\begin{aligned}
Var(\xi_k) = & \int_{|x-E\xi_k|<\epsilon B_n} |x-E\xi_k|^2 dF_k(x) + \int_{|x-E\xi_k|\ge\epsilon B_n} |x-E\xi_k|^2 dF_k(x) \\
\le & \epsilon^2 B_n^2 + \sum_{k=1}^n \int_{|x-E\xi_k|\ge\epsilon B_n} |x-E\xi_k|^2 dF_k(x).
\end{aligned}
$$

Note

$$
\begin{aligned}
Var(\xi_k) =& \int_{|x-E\xi_k|<\epsilon B_n} |x-E\xi_k|^2 dF_k(x) + \int_{|x-E\xi_k|\geq\epsilon B_n} |x-E\xi_k|^2 dF_k(x) \\
\leq& \epsilon^2 B_n^2 + \sum_{k=1}^n \int_{|x-E\xi_k|\geq\epsilon B_n} |x-E\xi_k|^2 dF_k(x).
\end{aligned}
$$

So, the Lindeberg condition (1) also implies

$$
\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{Var\xi_k}{B_n^2} = 0 \quad \text{Feller's condition,} \tag{3}
$$

Note

$$
\begin{aligned}
Var(\xi_k) =& \int_{|x-E\xi_k|<\epsilon B_n} |x-E\xi_k|^2 dF_k(x) + \int_{|x-E\xi_k|\geq \epsilon B_n} |x-E\xi_k|^2 dF_k(x) \\
\leq& \epsilon^2 B_n^2 + \sum_{k=1}^n \int_{|x-E\xi_k|\geq \epsilon B_n} |x-E\xi_k|^2 dF_k(x).
\end{aligned}
$$

So, the Lindeberg condition (1) also implies

$$
\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{Var\xi_k}{B_n^2} = 0 \ \text{ Feller's condition,} \tag{3}
$$

Theorem

Conversely, if (2) and (3) hold, then (1) holds.

Theorem

(Lyapunov) Suppose that $\{\xi_k, k \geq 1\}$ *is a sequence of indep. r.v.s, which satisfy*

$$\frac{1}{\left(\sum_{k=1}^n Var\xi_k\right)^{1+\delta/2}} \sum_{k=1}^n E|\xi_k - E\xi_k|^{2+\delta} \to 0 \quad as\ n \to \infty,$$

then the central limit theorem holds true.

Proof.

Proof.

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-E\xi_k|\geq\varepsilon B_n}(x-E\xi_k)^2dF_k(x)$$

$$\leq$$

Proof.

$$
\begin{aligned}
&\frac{1}{B_n^2}\sum_{k=1}^n \int_{|x-E\xi_k|\ge \varepsilon B_n}(x-E\xi_k)^2 dF_k(x)\\
\le\ &\frac{1}{\epsilon^\delta}\frac{1}{B_n^{2+\delta}}\sum_{k=1}^n E|\xi_k-E\xi_k|^{2+\delta}\to 0.
\end{aligned}
$$

Aleksandr Mikhailovich Lyapunov

(June 1857 – November 1918)

Example

An insurance company issues two kinds of one-year-term life insurance with random claim amounts 10,000 yuan and 20,000 yuan respectively. The claim probability q_k and the number of insurant n_k are denoted Table below.

Type k	q_k	claim amounts b_k	n_k
1	0.02	1	500
2	0.02	2	500
3	0.10	1	300
4	0.10	2	500

The insurance company hopes that the probability that the sum of claims exceeds the total premium is only 0.05. Now the premium is priced according to the expectation value principle, that is, the premium of policy i is $\pi(X_i) = (1+\theta)EX_i$, it is required to estimate the value of θ.

Solution. $S = \sum_{i=1}^{1800} X_i$. θ is to satisfy $P(S \le \pi(S)) = 0.95$. While,

$$ES = \sum_{i=1}^{1800} EX_i = \sum_{k=1}^{4} n_k b_k q_k$$

$$= \quad 500 \cdot 1 \cdot 0.02 + 500 \cdot 2 \cdot 0.02 + 300 \cdot 1 \cdot 0.10 + 500 \cdot 2 \cdot 0.10$$

$$= \quad 160,$$

Solution. $S=\sum_{i=1}^{1800} X_i$. θ is to satisfy $P(S\le \pi(S))=0.95$. While,

$$
\begin{aligned}
ES &= \sum_{i=1}^{1800} EX_i = \sum_{k=1}^{4} n_k b_k q_k \\
&= 500\cdot 1\cdot 0.02 + 500\cdot 2\cdot 0.02 + 300\cdot 1\cdot 0.10 + 500\cdot 2\cdot 0.10 \\
&= 160,
\end{aligned}
$$

$$
\begin{aligned}
VarS &= \sum_{i=1}^{1800} VarX_i = \sum_{k=1}^{4} n_k b_k^2 q_k(1-q_k) \\
&= 500\cdot 1^2\cdot 0.02\cdot 0.98 + 500\cdot 2^2\cdot 0.02\cdot 0.98 \\
&\quad +300\cdot 1^2\cdot 0.10\cdot 0.90 + 500\cdot 2^2\cdot 0.10\cdot 0.90 \\
&= 256.
\end{aligned}
$$

From these we obtain the sum of premium

$$\pi(S) = (1+\theta)ES = 160(1+\theta).$$

From these we obtain the sum of premium

$$\pi(S) = (1+\theta)ES = 160(1+\theta).$$

According to the request, we have $P(S \le (1+\theta)ES) = 0.95$, that is

$$P\left(\frac{S-ES}{\sqrt{VarS}} \le \frac{\theta ES}{\sqrt{VarS}}\right) = P\left(\frac{S-ES}{\sqrt{VarS}} \le 10\theta\right) = 0.95.$$

Since,

$$\frac{1}{(Var(S))^{3/2}}\sum_{i=1}^{n} E[|X_i - E[X_i]|^3]$$

$$=\frac{\sum_{k=1}^{4} n_k\{b_k^3|1-q_k|^3 q_k + b_k^3|-q_k|^3(1-q_k)\}}{\{\sum_{k=1}^{4} n_k b_k^2 q_k(1-q_k)\}^{3/2}}$$

$$\le\frac{n\max_{k=1,\ldots,4} b_k^3}{n^{3/2}\{\min_k b_k^2 q_k(1-q_k)\}^{3/2}} \to 0.$$

Since,

$$\begin{aligned}
&\frac{1}{(Var(S))^{3/2}}\sum_{i=1}^{n} E[|X_i - E[X_i]|^3] \\
=&\frac{\sum_{k=1}^4 n_k\{b_k^3|1-q_k|^3 q_k + b_k^3|-q_k|^3(1-q_k)\}}{\{\sum_{k=1}^4 n_k b_k^2 q_k(1-q_k)\}^{3/2}} \\
\le&\frac{n \max_{k=1,\ldots,4} b_k^3}{n^{3/2}\{\min_k b_k^2 q_k(1-q_k)\}^{3/2}} \to 0.
\end{aligned}$$

By the Lyapunov CLT,

$$P\left(\frac{S-ES}{\sqrt{VarS}} \le x\right) \to \Phi(x).$$

One can approximately regard $\frac{S-ES}{\sqrt{VarS}}$ as a standard normal variable.
We have $10\theta = 1.645$, that is $\theta = 0.1645$.

4.1.1 Weak convergence of distribution functions

In the central limit theorem,

$$P\left(\frac{S_n - na}{\sqrt{n}\sigma} \le x\right) \to \Phi(x), \quad \text{for all } x,$$

means that the distribution of $\frac{S_n - na}{\sqrt{n}\sigma}$ converges to the standard normal distribution. This kind of convergence is called convergence in ditribution.

Definition

Let F be a cdf, $\{F_n, n \geq 1\}$ a sequence of cdfs. We say that F_n converges weakly to F, denoted by $F_n \xrightarrow{w} F$, if $F_n(x) \longrightarrow F(x)$ holds at every continuity point x of F as $n \longrightarrow \infty$.

Let ξ be a r.v., $\{\xi_n, n \geq 1\}$ a sequence of r.v.s, we say $\{\xi_n\}$ converges in distribution to ξ, denoted by $\xi_n \xrightarrow{d} \xi$, if the cdfs ξ_n's converges weakly to the cdf of ξ.

Remark 1. The limit function of a sequence of distribution functions is not necessarily a distribution function. For example, let

$$F_n(x) = \begin{cases} 0, & x < n, \\ 1, & x \geq n. \end{cases}$$

This distribution function converges pointwise to 0, but $F(x) \equiv 0$ is not a distribution function.

Remark 2. Since the set of discontinuous points of a distribution function F is at most countable, $F_n \xrightarrow{w} F$ means that F_n converges everywhere to F in a dense subset of $\mathbf{R}$.

Theorem

(Helly's first theorem) Let $\{F_n, n \geq 1\}$ *be a sequence of distribution functions. Then there exists a non-decreasing right-continuous function* F *(not necessarily a distribution function) with* $0 \leq F(x) \leq 1$, $x \in \mathbf{R}$, *and a subsequence* F_{n_k}, *such that* $F_{n_k}(x) \to F(x)$ *for every continuity point* x *of* F *as* $k \longrightarrow \infty$.

Main idea of the proof. For each given x, since $\{F_n(x)\}$ is a bounded sequence, there is a subsequence $\{F_{n_m}(x)\}$ and a number $F(x)$, such that

$$F_{n_m}(x) \to F(x).$$

Main idea of the proof. For each given x, since $\{F_n(x)\}$ is a bounded sequence, there is a subsequence $\{F_{n_m}(x)\}$ and a number $F(x)$, such that

$$F_{n_m}(x) \to F(x).$$

However, the subsequence $\{n_m\}$ may depend on the value of x, i.e., $n_m = n_m(x)$. What we want to do is to find an "uniform" sequence $\{n_m\}$ which does depend on x such the above convergence holds.

Proof. Let $r_1, r_2, \cdots,$ denote the set of rational numbers. That $0 \le F(x) \le 1$ means that $\{F_n(r_1)\}$ is a bounded sequence. So, there exists a convergent subsequence $\{F_{n_m^{(1)}}(r_1)\}$. Denote the limit by

$$G(r_1) = \lim_{m\to\infty} F_{n_m^{(1)}}(r_1).$$

Proof. Let $r_1, r_2, \cdots,$ denote the set of rational numbers. That $0 \le F(x) \le 1$ means that $\{F_n(r_1)\}$ is a bounded sequence. So, there exists a convergent subsequence $\{F_{n_m^{(1)}}(r_1)\}$. Denote the limit by

$$G(r_1) = \lim_{m\to\infty} F_{n_m^{(1)}}(r_1).$$

Then, consider the bounded sequence $\{F_{n_m^{(1)}}(r_2)\}$. There exists a further convergent subsequences $\{F_{n_m^{(2)}}(r_2)\}$. Denote the limit by

$$G(r_2) = \lim_{m\to\infty} F_{n_m^{(2)}}(r_2).$$

Repeating the procedure, we obtain

$$\{F_{n_m^{(k)}}\} \subset \{F_{n_m^{(k-1)}}\}, \ G(r_k) = \lim_{m\to\infty} F_{n_m^{(k)}}(r_k), \quad k \geq 2.$$

Now, consider the diagonal sequence $\{F_{n_m^{(m)}}\}$. Obviously,

$$\lim_{m\to\infty} F_{n_m^{(m)}}(r_k) = G(r_k), \quad \forall k \geq 1.$$

In addition, $F_n \nearrow \Longrightarrow G(r) \nearrow$ and also $0 \leq G(r) \leq 1$.

Let

$$F(x) = \lim_{r_j \downarrow x} G(r_j) = \inf_{r_j > x} G(r_j), \quad x \in \mathbf{R}.$$

Then $F(x) \nearrow$ and also $0 \le F(x) \le 1$, and $F(x)$ is right-continuous. Further, if $r < x < s$ and r, s are rational numbers, then

$$G(r) \le F(x) \le G(s).$$

Now for any continuous point x of F and $h > 0$, there are $r_i < r_j$ such that

$$x - h < r_i < x < r_j < x + h.$$

It follows that

$$\begin{array}{ccccc} F_{n_m^{(m)}}(r_i) & \leq F_{n_m^{(m)}}(x) \leq & F_{n_m^{(m)}}(r_j) \\ \downarrow & & \downarrow \\ F(x-h) \leq G(r_i) & & G(r_j) \leq F(x+h). \end{array}$$

Letting $m \to \infty$ yields

$$\begin{aligned} F(x-h) \le G(r_i) &= \lim_{m} F_{n_m^{(m)}}(r_i) \\ &\le \liminf_{m} F_{n_m^{(m)}}(x) \\ &\le \limsup_{m} F_{n_m^{(m)}}(x) \\ &\le \limsup_{m} F_{n_m^{(m)}}(r_j) \\ &= G(r_j) \le F(x+h). \end{aligned}$$

Letting $m \to \infty$ yields

$$
\begin{aligned}
F(x-h) \le G(r_i) &= \lim_m F_{n_m^{(m)}}(r_i) \\
&\le \liminf_m F_{n_m^{(m)}}(x) \\
&\le \limsup_m F_{n_m^{(m)}}(x) \\
&\le \limsup_m F_{n_m^{(m)}}(r_j) \\
&= G(r_j) \le F(x+h).
\end{aligned}
$$

Letting $h \to 0$ yields

$$
\liminf_m F_{n_m^{(m)}}(x) = \limsup_m F_{n_m^{(m)}}(x) = F(x).
$$

Theorem

(Helly's second theorem) Let F be a cdf, $\{F_n, n \geq 1\}$ a sequence of cdfs such that $F_n \xrightarrow{w} F$. If $g(x)$ is a bounded continuous function in $\mathbf{R}$, then

$$\int_{-\infty}^{\infty} g(x)dF_n(x) \longrightarrow \int_{-\infty}^{\infty} g(x)dF(x).$$

Main idea of Proof.

$$
\begin{aligned}
& \int_{-\infty}^{\infty} g(x) dF_n(x) \\
\approx\ & \sum g(x_{i-1}) \left[F_n(x_i) - F_n(x_{i-1})\right] \\
\to\ & \sum g(x_{i-1}) \left[F(x_i) - F(x_{i-1})\right] \\
& (\text{ if } x_i' s \text{ are continuous points of } F) \\
\approx\ & \int_{-\infty}^{\infty} g(x) dF(x).
\end{aligned}
$$

Proof. Since g is a bounded function, there must exist a constant $c > 0$ such that $|g(x)| < c$, $x \in \mathbf{R}$.

Proof. Since g is a bounded function, there must exist a constant $c > 0$ such that $|g(x)| < c$, $x \in \mathbf{R}$. For given $\delta > 0$ and $a > 0$ with $\pm a$ being continuous points of F, select
$-a = x_0 < x_1 < \cdots < x_m = a$ such that x_is are continuous points of F and $|\Delta x| =: \max_i |x_i - x_{i-1}| < \delta$.

Proof. Since g is a bounded function, there must exist a constant $c > 0$ such that $|g(x)| < c$, $x \in \mathbf{R}$. For given $\delta > 0$ and $a > 0$ with $\pm a$ being continuous points of F, select $-a = x_0 < x_1 < \cdots < x_m = a$ such that x_is are continuous points of F and $|\Delta x| =: \max_i |x_i - x_{i-1}| < \delta$. Let

$$g_m(x) = \begin{cases} g(x_{i-1}), & \text{if } x_{i-1} < x \le x_i, \\ 0, & \text{if } x \le -a \text{ or } x > a. \end{cases}$$

It is easily seen that

$$\int_{-\infty}^{\infty} g_m(x)dF(x) = \sum_{i=1}^{m} g(x_{i-1})(F(x_i) - F(x_{i-1})),$$

It is easily seen that

$$\begin{aligned}\int_{-\infty}^{\infty} g_m(x)dF(x) &= \sum_{i=1}^{m} g(x_{i-1})(F(x_i) - F(x_{i-1})), \\ &= \sum_{i=1}^{m} \int_{x_{i-1}}^{x_i} g(x_{i-1})dF(x),\end{aligned}$$

It is easily seen that

$$\begin{aligned}
\int_{-\infty}^{\infty} g_m(x)dF(x) &= \sum_{i=1}^{m} g(x_{i-1})(F(x_i)-F(x_{i-1})), \\
&= \sum_{i=1}^{m} \int_{x_{i-1}}^{x_i} g(x_{i-1})dF(x), \\
\int_{-\infty}^{\infty} g(x)dF(x) &= \int_{-\infty}^{-a} g(x)dF(x) \\
&\quad + \sum_{i=1}^{m} \int_{x_{i-1}}^{x_i} g(x)dF(x) + \int_{a}^{\infty} g(x)dF(x).
\end{aligned}$$

It follows that

$$
\begin{aligned}
& \left|\int_{-\infty}^{\infty} g(x)dF(x) - \int_{-\infty}^{\infty} g_m(x)dF(x)\right| \\
\leq\ & \left|\int_{-\infty}^{-a} g(x)dF(x)\right| + \left|\int_{a}^{\infty} g(x)dF(x)\right| \\
& + \sum_{i=1}^{m} \int_{x_{i-1}}^{x_i} |g(x) - g(x_{i-1})| dF(x) \\
\leq\ & c\left[F(-a) + 1 - F(a)\right] + \max_i \max_{x_{i-1} \leq x \leq x_i} |g(x) - g(x_{i-1})|.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& \left|\int_{-\infty}^{\infty} g(x)dF(x) - \int_{-\infty}^{\infty} g_m(x)dF(x)\right| \\
\leq\ & \left|\int_{-\infty}^{-a} g(x)dF(x)\right| + \left|\int_{a}^{\infty} g(x)dF(x)\right| \\
& + \sum_{i=1}^{m} \int_{x_{i-1}}^{x_i} |g(x) - g(x_{i-1})| dF(x) \\
\leq\ & c\left[F(-a) + 1 - F(a)\right] + \max_i \max_{x_{i-1}\leq x\leq x_i} |g(x) - g(x_{i-1})|.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& \left|\int_{-\infty}^{\infty} g(x)dF_n(x) - \int_{-\infty}^{\infty} g_m(x)dF_n(x)\right| \\
\leq\ & c\left[F_n(-a) + 1 - F_n(a)\right] + \max_i \max_{x_{i-1}\leq x\leq x_i} |g(x) - g(x_{i-1})|.
\end{aligned}
$$

While

$$
\begin{aligned}
&\left|\int_{-\infty}^{\infty} g_m(x)dF_n(x) - \int_{-\infty}^{\infty} g_m(x)dF(x)\right| \\
\leq\ & \sum_{i=0}^{m} |g(x_{i-1})| \left[|F_n(x_i) - F(x_i)| + |F_n(x_{i-1}) - F(x_{i-1})|\right] \\
\leq\ & 2cm \max_i |F_n(x_i) - F(x_i)|.
\end{aligned}
$$

It follows that

$$\left|\int_{-\infty}^{\infty} g(x)dF_n(x) - \int_{-\infty}^{\infty} g(x)dF(x)\right|$$

$$\leq$$

It follows that

$$
\begin{aligned}
& \left|\int_{-\infty}^{\infty} g(x)dF_n(x) - \int_{-\infty}^{\infty} g(x)dF(x)\right| \\
\leq\ & c\left[F_n(-a) + 1 - F_n(a)\right] + c\left[F(-a) + 1 - F(a)\right] \\
& +2\max_i \max_{x_{i-1}\leq x\leq x_i} |g(x) - g(x_{i-1})| \\
& +2cm\max_i |F_n(x_i) - F(x_i)| \\
\leq\ &
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& \left|\int_{-\infty}^{\infty} g(x)dF_n(x) - \int_{-\infty}^{\infty} g(x)dF(x)\right| \\
\leq\ & c\left[F_n(-a)+1-F_n(a)\right] + c\left[F(-a)+1-F(a)\right] \\
& +2\max_i \max_{x_{i-1}\leq x\leq x_i} |g(x)-g(x_{i-1})| \\
& +2cm\max_i |F_n(x_i)-F(x_i)| \\
\leq\ & 2c\left[F(-a)+1-F(a)\right] + 2\max_{\substack{|x-y|<\delta \\ |x|,|y|\leq a}} |g(x)-g(y)| \\
& +4cm\max_{i=0,\cdots,m} |F_n(x_i)-F(x_i)|.
\end{aligned}
$$

Now, for given $\epsilon > 0$, we fist choose $a = a(\epsilon) > 0$ ($\pm a$ be continuous points of F) such that

$$F(-a) + 1 - F(a) < \epsilon/(6c).$$

Now, for given $\epsilon > 0$, we fist choose $a = a(\epsilon) > 0$ ($\pm a$ be continuous points of F) such that

$$F(-a) + 1 - F(a) < \epsilon/(6c).$$

Secondly, choose $\delta > 0$ such that

$$\max_{|x-y|<\delta;\ |x|,|y|\le a} |g(x) - g(y)| < \epsilon/6.$$

Now, for given $\epsilon > 0$, we fist choose $a = a(\epsilon) > 0$ ($\pm a$ be continuous points of F) such that

$$F(-a) + 1 - F(a) < \epsilon/(6c).$$

Secondly, choose $\delta > 0$ such that

$$\max_{|x-y|<\delta;\ |x|,|y|\le a} |g(x) - g(y)| < \epsilon/6.$$

Thirdly, choose m and x_is such that x_is are continuous points of F and $|x_i - x_{i-1}| < \delta$.

Now, for given $\epsilon > 0$, we fist choose $a = a(\epsilon) > 0$ ($\pm a$ be continuous points of F) such that

$$F(-a) + 1 - F(a) < \epsilon/(6c).$$

Secondly, choose $\delta > 0$ such that

$$\max_{|x-y|<\delta;\ |x|,|y|\le a} |g(x) - g(y)| < \epsilon/6.$$

Thirdly, choose m and x_is such that x_is are continuous points of F and $|x_i - x_{i-1}| < \delta$. Finally, choose $N = N(\epsilon)$ such that

$$\max_{i=0,\cdots,m} |F_n(x_i) - F(x_i)| < \epsilon/(12cm) \quad n \ge N.$$

The proof is now completed.

The second proof. Let $U \sim U(0,1)$. Then $\xi_n = F_n^{-1}(U) \sim F_n$, $\xi = F^{-1}(U) \sim F$. Then

$$\int_{-\infty}^{\infty} g(x)dF_n(x) = Eg(\xi_n) = \int_0^1 g\big(F_n^{-1}(y)\big)dy,$$

$$\int_{\infty}^{\infty} g(x)dF(x) = Eg(\xi) = \int_0^1 g\big(F^{-1}(y)\big)dy.$$

The second proof. Let $U \sim U(0,1)$. Then $\xi_n = F_n^{-1}(U) \sim F_n$, $\xi = F^{-1}(U) \sim F$. Then

$$\int_{-\infty}^{\infty} g(x)dF_n(x) = Eg(\xi_n) = \int_0^1 g\big(F_n^{-1}(y)\big)dy,$$
$$\int_{\infty}^{\infty} g(x)dF(x) = Eg(\xi) = \int_0^1 g\big(F^{-1}(y)\big)dy.$$

It can be shown that

$$F_n \xrightarrow{w} F \Longleftrightarrow F_n^{-1}(y) \to F^{-1}(y) \;\; \forall y \in C(F^{-1}),$$

where $C(F^{-1})$ is the set of continuity points of F^{-1}.

Hence

$$F_n^{-1}(y) \to F^{-1}(y) \ \ a.e.\ L.$$

It follows that

$$g\big(F_n^{-1}(y)\big) \to g\big(F^{-1}(y)\big) \ \ a.e.\ L.$$

Hence

$$F_n^{-1}(y) \to F^{-1}(y) \;\; a.e. \; L.$$

It follows that

$$g\big(F_n^{-1}(y)\big) \to g\big(F^{-1}(y)\big) \;\; a.e. \; L.$$

So

$$\int_{-\infty}^{\infty} g(x)dF_n(x) \to \int_{\infty}^{\infty} g(x)dF(x).$$

Remark.

- If $F_n(x), F(x) \nearrow$ are right continuous, and for any continuous point x of F, $F_n(x) \to F(x)$, then for continuous points $A < B$ of F, and continuous $g(\cdot)$,

$$\int_A^B g(x)dF_n(x) \longrightarrow \int_A^B g(x)dF(x).$$

Proof. For F, we redefine it as

$$F^*(x) = \begin{cases} 1, & x \geq B, \\ \frac{F(x)-F(A)}{F(B)-F(A)}, & A \leq x \leq B, \\ 0, & x \leq A; \end{cases}$$

and for a function $g(x)$, we redefine it as $g^*(x) = g(B)$, $x \geq B$; $g^*(x) = g(x)$, $A \leq x \leq B$; $g^*(x) = g(A)$, $x \leq A$. Then

$$F_n^* \xrightarrow{w} F^*,$$

$$\int_{-\infty}^{\infty} g^*(x) dF_n^*(x) \to \int_{-\infty}^{\infty} g^*(x) dF^*(x).$$

Remark.

- If $g(x)$ is continuous, and $\{g_t(x)\}$ satisfy $|g_t(x)| \leq c$ and $|g_t(x) - g_t(y)| \leq |g(x) - g(y)|$, then uniformly in t,

$$\int_{-\infty}^{\infty} g_t(x) dF_n(x) \longrightarrow \int_{-\infty}^{\infty} g_t(x) dF(x).$$

Proof

$$\left|\int_{-\infty}^{\infty} g_t(x)dF_n(x) - \int_{-\infty}^{\infty} g_t(x)dF(x)\right|$$

$$=$$

Proof

$$
\begin{aligned}
& \left| \int_{-\infty}^{\infty} g_t(x) dF_n(x) - \int_{-\infty}^{\infty} g_t(x) dF(x) \right| \\
= & \left| \int_0^1 g_t(F_n^{-1}(y)) dy - \int_0^1 g_t(F^{-1}(y)) dy \right| \\
\leq & \int_0^1 \left| g_t(F_n^{-1}(y)) - g_t(F^{-1}(y)) \right| dy \\
\leq &
\end{aligned}
$$

Proof

$$
\begin{aligned}
& \left| \int_{-\infty}^{\infty} g_t(x) dF_n(x) - \int_{-\infty}^{\infty} g_t(x) dF(x) \right| \\
= & \left| \int_0^1 g_t(F_n^{-1}(y)) dy - \int_0^1 g_t(F^{-1}(y)) dy \right| \\
\le & \int_0^1 \left| g_t(F_n^{-1}(y)) - g_t(F^{-1}(y)) \right| dy \\
\le & \int_0^1 \min\left\{ 2c, \left| g(F_n^{-1}(y)) - g(F^{-1}(y)) \right| \right\} dy
\end{aligned}
$$

Proof

$$
\begin{aligned}
& \left| \int_{-\infty}^{\infty} g_t(x) dF_n(x) - \int_{-\infty}^{\infty} g_t(x) dF(x) \right| \\
= & \left| \int_0^1 g_t(F_n^{-1}(y)) dy - \int_0^1 g_t(F^{-1}(y)) dy \right| \\
\le & \int_0^1 \left| g_t(F_n^{-1}(y)) - g_t(F^{-1}(y)) \right| dy \\
\le & \int_0^1 \min\left\{ 2c, \left| g(F_n^{-1}(y)) - g(F^{-1}(y)) \right| \right\} dy \\
& \to 0.
\end{aligned}
$$

Theorem

(Lévy's continuity theorem) Let F *be a cdf,* $\{F_n, n \geq 1\}$ *a sequence of cdfs. If* $F_n \xrightarrow{w} F$, *then the corresponding sequence of characteristic functions* $\{f_n(t)\}$ *converges to the characteristic function* $f(t)$ *of* F *uniformly in* t *on any finite interval.*

Theorem

(Lévy's continuity theorem) Let F *be a cdf,* $\{F_n, n \geq 1\}$ *a sequence of cdfs. If* $F_n \xrightarrow{w} F$*, then the corresponding sequence of characteristic functions* $\{f_n(t)\}$ *converges to the characteristic function* $f(t)$ *of* F *uniformly in* t *on any finite interval.*

Proof. Let $g_t(x) = e^{itx}$, then

Theorem

(Lévy's continuity theorem) Let F be a cdf, $\{F_n, n \geq 1\}$ a sequence of cdfs. If $F_n \xrightarrow{w} F$, then the corresponding sequence of characteristic functions $\{f_n(t)\}$ converges to the characteristic function $f(t)$ of F uniformly in t on any finite interval.

Proof. Let $g_t(x) = e^{itx}$, then $|g_t(x)| = 1$ and $\sup_{|t| \leq b} |g_t(x) - g_t(y)| \leq b|x - y|$.

Theorem

(The converse limit theorem) Let $f_n(t)$ be characteristic function of distribution function $F_n(x)$, if for every t, $f_n(t) \longrightarrow f(t)$, and $f(t)$ is continuous on $t = 0$, then $f(t)$ must be a characteristic function of some distribution function F, and $F_n \xrightarrow{w} F$.

Before the proof, we need a lemma.

Lemma

If $f(t)$ is the characteristic function of a distribution function F, then

$$\int_{|x|\geq 2/u} dF(x) \leq \frac{1}{u}\int_{-u}^{u}(1-f(t))dt.$$

Proof. For any $u > 0$,

$$\frac{1}{2u}\int_{-u}^{u}(1-f(t))dt = \frac{1}{2u}\int_{-u}^{u}\int_{-\infty}^{\infty}(1-e^{itx})dF(x)dt$$

Proof. For any $u > 0$,

$$\frac{1}{2u}\int_{-u}^{u}(1-f(t))dt = \frac{1}{2u}\int_{-u}^{u}\int_{-\infty}^{\infty}(1-e^{itx})dF(x)dt$$
$$= \int_{-\infty}^{\infty}\int_{-u}^{u}\frac{1}{2u}(1-e^{itx})dtdF(x) = \int_{-\infty}^{\infty}\left(1-\frac{\sin ux}{ux}\right)dF(x)$$

Proof. For any $u > 0$,

$$
\begin{aligned}
&\frac{1}{2u}\int_{-u}^{u}(1-f(t))dt = \frac{1}{2u}\int_{-u}^{u}\int_{-\infty}^{\infty}(1-e^{itx})dF(x)dt \\
= & \int_{-\infty}^{\infty}\int_{-u}^{u}\frac{1}{2u}(1-e^{itx})dtdF(x) = \int_{-\infty}^{\infty}\left(1-\frac{\sin ux}{ux}\right)dF(x) \\
\geq & \int_{|x|\geq 2/u}\left(1-\frac{\sin ux}{ux}\right)dF(x) \geq \frac{1}{2}\int_{|x|\geq 2/u}dF(x).
\end{aligned}
$$

Proof.

We want to prove that there exists a cdf F such that for any subsequence $\{n'\}$ there is a further subsequence $\{n''\}$ for which

$$F_{n''} \xrightarrow{w} F.$$

It suffices to prove that for any subsequence $\{n'\}$ there is a further subsequence $\{n''\}$ and a cdf $F(x)$ (which may depend on the subsequence) such that

$$F_{n''} \xrightarrow{w} F. \qquad (*)$$

In fact, if (*) holds then

$$f_{n''}(t) \to f_F(t),$$

due to Lévy's continuity theorem, here $f_F(t)$ is the c.f. of F.

In fact, if (*) holds then

$$f_{n''}(t) \to f_F(t),$$

due to Lévy's continuity theorem, here $f_F(t)$ is the c.f. of F. By the assumption of the theorem, we must have $f_F \equiv f$.

In fact, if (*) holds then

$$f_{n''}(t) \to f_F(t),$$

due to Lévy's continuity theorem, here $f_F(t)$ is the c.f. of F. By the assumption of the theorem, we must have $f_F \equiv f$. So, $f(t)$ is a c.f. and $F(x)$ is uniquely determined by $f(t)$(and hence does not depend on the subsequence). And further, (*) means that $F_n \xrightarrow{w} F$.

Now, by Helly's first theorem, there exists a non-decreasing right-continuous function F (not necessarily a distribution function) with $0 \le F(x) \le 1$, $x \in \mathbf{R}$, and a subsequence $\{n''\} \subset \{n'\}$ such that

$$F_{n''}(x) \to F(x), \forall \text{ continuous point } x \text{ of } F.$$

Now, by Helly's first theorem, there exists a non-decreasing right-continuous function F (not necessarily a distribution function) with $0 \le F(x) \le 1$, $x \in \mathbf{R}$, and a subsequence $\{n''\} \subset \{n'\}$ such that

$$F_{n''}(x) \to F(x), \forall \text{ continuous point } x \text{ of } F.$$

Next, it suffices to show that $F(x)$ is a cdf, i.e., $F(+\infty) - F(-\infty) = 1$.

Notice that if $a > 0$ and $\pm a$ are continuous points of F, then

$$F(a) - F(-a) = \lim_{n''} \left[F_{n''}(a) - F_{n''}(-a) \right].$$

Notice that if $a > 0$ and $\pm a$ are continuous points of F, then

$$F(a) - F(-a) = \lim_{n''} \left[F_{n''}(a) - F_{n''}(-a) \right].$$

We need only to show that for any given $\epsilon > 0$, if a is sufficiently large, then

$$\limsup_{n} \int_{|x| \geq a} dF_n(x) \leq \epsilon. \qquad (**)$$

So,

$$\begin{aligned}
&\limsup_n \int_{|x|\geq 2/u} dF_n(x) \\
\leq\ & \limsup_n \frac{1}{u}\int_{-u}^{u} |1-f_n(t)|dt \\
\leq\ & \frac{1}{u}\int_{-u}^{u} |1-f(t)|dt
\end{aligned}$$

So,

$$
\begin{aligned}
& \limsup_n \int_{|x|\geq 2/u} dF_n(x) \\
\leq \quad & \limsup_n \frac{1}{u}\int_{-u}^{u} |1-f_n(t)|dt \\
\leq \quad & \frac{1}{u}\int_{-u}^{u} |1-f(t)|dt
\end{aligned}
$$

Since $f(t)$ is continuous at $t=0$, we can choose $u>0$ small enough such that $|1-f(t)|<\epsilon/2$ whenever $|t|\leq u$. And then (**) is proved.

Summary: The following are equivalent:

1. $F_n \xrightarrow{w} F$ ($\xi_n \xrightarrow{d} \xi$);
2. $$\int g(x)dF_n(x) \to \int g(x)dF(x) \qquad (\#)$$
   $$(E[g(\xi_n)] \to E[g(\xi)])$$
   for every bounded, continuous function g;
3. (#) holds for every bounded, uniformly continuous function;
4. (#) holds for every bounded, continuous function g having bounded, continuous derivatives of each order;
5. $f_n(t) \to f(t)$ for all t.

Multi-dimensional Case:

Definition

Let $\{F, F_n; n \geq 1\}$ be a sequence of CDF on $\mathbb{R}^d$,

$$C_F = \{\boldsymbol{x} = (x_1, \ldots, x_d) : F_j(x_j) = F_j(x_j + 0), j = 1, \ldots, d\}.$$

If

$$F_n(\boldsymbol{x}) \to F(\boldsymbol{x}), \ \ \forall \boldsymbol{x} \in C_F,$$

then call F_n is weakly convergent to F, written as $F_n \xrightarrow{d} F$.
If the CDF of $\boldsymbol{X}_n$ is weakly convergent to the CDF of $\boldsymbol{X}$, we call $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$.

多维依分布收敛有一维依分布收敛同样的性质.

Theorem

The following are equivalent:

1. $\boldsymbol{X}_n = (X_{n1}, \ldots, X_{nd}) \xrightarrow{d} \boldsymbol{X} = (X_1, \ldots, X_d)$;
2. *For every bounded, continuous function* g
$$E[g(\boldsymbol{X}_n)] \to E[g(\boldsymbol{X})]; \qquad (\#)$$
3. *(#) holds for every bounded, continuous function* g *having bounded, continuous derivatives of each order;*
4. $f_n(\boldsymbol{t}) \to f(\boldsymbol{t})$ *for all* $\boldsymbol{t}$.

Theorem

(Cramér-Wold device) Let $\boldsymbol{X}_n = (X_{n1}, \ldots, X_{nd})$ *and* $\boldsymbol{X} = (X_1, \ldots, X_d)$ *be random vectors. Then* $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$ *if and only if*

$$a_1 X_{n1} + \cdots + a_d X_{nd} \xrightarrow{d} a_1 X_1 + \cdots + a_d X_d, \quad \forall a_1, \ldots, a_d \in \mathbb{R}.$$

Proof. Let $f_n(\boldsymbol{t})$ and $f(\boldsymbol{t})$ be the c.f.s of $\boldsymbol{X}_n$ and $\boldsymbol{X}$. Then the c.f.s of $a_1X_{n1}+\cdots+a_dX_{nd}$ and $a_1X_1+\cdots+a_dX_d$ are $f_n(ta_1,\cdots,ta_d)$ and $f(ta_1,\cdots,ta_d)$. So

$$\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$$

Proof. Let $f_n(\boldsymbol{t})$ and $f(\boldsymbol{t})$ be the c.f.s of $\boldsymbol{X}_n$ and $\boldsymbol{X}$. Then the c.f.s of $a_1X_{n1}+\cdots+a_dX_{nd}$ and $a_1X_1+\cdots+a_dX_d$ are $f_n(ta_1,\cdots,ta_d)$ and $f(ta_1,\cdots,ta_d)$. So

$$\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X} \Longleftrightarrow f_n(\boldsymbol{t}) \to f(\boldsymbol{t}) \ \forall \boldsymbol{t}$$

Proof. Let $f_n(\boldsymbol{t})$ and $f(\boldsymbol{t})$ be the c.f.s of $\boldsymbol{X}_n$ and $\boldsymbol{X}$. Then the c.f.s of $a_1X_{n1} + \cdots + a_dX_{nd}$ and $a_1X_1 + \cdots + a_dX_d$ are $f_n(ta_1, \cdots, ta_d)$ and $f(ta_1, \cdots, ta_d)$. So

$$\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X} \Longleftrightarrow f_n(\boldsymbol{t}) \to f(\boldsymbol{t}) \ \forall \boldsymbol{t}$$

$$\Longleftrightarrow f_n(ta_1, \cdots, ta_d) \to f(ta_1, \cdots, ta_d) \ \forall t \text{ and } \boldsymbol{a}$$

Proof. Let $f_n(\boldsymbol{t})$ and $f(\boldsymbol{t})$ be the c.f.s of $\boldsymbol{X}_n$ and $\boldsymbol{X}$. Then the c.f.s of $a_1X_{n1}+\cdots+a_dX_{nd}$ and $a_1X_1+\cdots+a_dX_d$ are $f_n(ta_1,\cdots,ta_d)$ and $f(ta_1,\cdots,ta_d)$. So

$$\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X} \Longleftrightarrow f_n(\boldsymbol{t}) \to f(\boldsymbol{t})\ \forall \boldsymbol{t}$$

$$\Longleftrightarrow f_n(ta_1,\cdots,ta_d) \to f(ta_1,\cdots,ta_d)\ \forall t \text{ and } \boldsymbol{a}$$

$$\Longleftrightarrow a_1X_{n1}+\cdots+a_dX_{nd} \xrightarrow{d} a_1X_1+\cdots+a_dX_d,\quad \forall \boldsymbol{a}.$$

Example

Prove the Poisson approximation of binomial distributions by the method of characteristic function.

Proof.

Example

Prove the Poisson approximation of binomial distributions by the method of characteristic function.

Proof. Let $\xi_n \sim B(n, p_n)$, and $\lim_{n\to\infty} np_n = \lambda$. Then its c.f. is $f_n(t) = (p_n e^{it} + q_n)^n$.

Example

Prove the Poisson approximation of binomial distributions by the method of characteristic function.

Proof. Let $\xi_n \sim B(n, p_n)$, and $\lim_{n\to\infty} np_n = \lambda$. Then its c.f. is $f_n(t) = (p_n e^{it} + q_n)^n$. So

$$f_n(t) = (1 + \frac{n \cdot p_n(e^{it} - 1)}{n})^n$$

Example

Prove the Poisson approximation of binomial distributions by the method of characteristic function.

Proof. Let $\xi_n \sim B(n, p_n)$, and $\lim_{n\to\infty} np_n = \lambda$. Then its c.f. is $f_n(t) = (p_n e^{it} + q_n)^n$. So

$$f_n(t) = (1 + \frac{n \cdot p_n(e^{it} - 1)}{n})^n \to e^{\lambda(e^{it}-1)}.$$

This is just the c.f. of $P(\lambda)$.

Example

Prove the Poisson approximation of binomial distributions by the method of characteristic function.

Proof. Let $\xi_n \sim B(n, p_n)$, and $\lim_{n\to\infty} np_n = \lambda$. Then its c.f. is $f_n(t) = (p_n e^{it} + q_n)^n$. So

$$f_n(t) = (1 + \frac{n \cdot p_n(e^{it} - 1)}{n})^n \to e^{\lambda(e^{it}-1)}.$$

This is just the c.f. of $P(\lambda)$. It follows that from the converse limit theorem, the binomial distribution $B(n, p_n)$ converges in distribution to the Poisson distribution $P(\lambda)$.

4.1.2 Properties

4.1.2 Properties

1. Let $\{F_n, n \geq 1\}$ be a sequence of distribution functions. If $F_n \xrightarrow{d} F$, and F is a continuous distribution function, then

$$\sup_x |F_n(x) - F(x)| \to 0.$$

(Proof as exercise).

2. Let ξ be a random variable, $\{\xi_n, n \geq 1\}$ a sequence of random variables, $g(x)$ a continuous function on $\mathbf{R}$. If $\xi_n \xrightarrow{d} \xi$, then $g(\xi_n) \xrightarrow{d} g(\xi)$.

2. Let ξ be a random variable, $\{\xi_n, n \geq 1\}$ a sequence of random variables, $g(x)$ a continuous function on $\mathbf{R}$. If $\xi_n \xrightarrow{d} \xi$, then $g(\xi_n) \xrightarrow{d} g(\xi)$.

Proof. Let $h(y)$ be a bounded continuous function. Then, $h \circ g(x)$ is a bounded continuous function. So

2. Let ξ be a random variable, $\{\xi_n, n \geq 1\}$ a sequence of random variables, $g(x)$ a continuous function on $\mathbf{R}$. If $\xi_n \xrightarrow{d} \xi$, then $g(\xi_n) \xrightarrow{d} g(\xi)$.

Proof. Let $h(y)$ be a bounded continuous function. Then, $h \circ g(x)$ is a bounded continuous function. So

$$E\left[h\left(g(\xi_n)\right)\right] = Eh \circ g(\xi_n) \to Eh \circ g(\xi) = E\left[h\left(g(\xi)\right)\right].$$

附录

Proof of the Lindeberg-Feller CLT. Without loss of generality, we assume $E\xi_k = 0$. Write $\sigma_k^2 = Var(\xi_k)$. Let $g_k(t)$ be the c.f. of ξ_k. Then the c.f. of $\frac{\sum_{k=1}^n \xi_k}{B_n}$ is

$$f_n(t) = \prod_{k=1}^n g_k\left(\frac{t}{B_n}\right).$$

附录

Proof of the Lindeberg-Feller CLT. Without loss of generality, we assume $E\xi_k = 0$. Write $\sigma_k^2 = Var(\xi_k)$. Let $g_k(t)$ be the c.f. of ξ_k. Then the c.f. of $\frac{\sum_{k=1}^n \xi_k}{B_n}$ is

$$f_n(t) = \prod_{k=1}^n g_k\left(\frac{t}{B_n}\right).$$

Write

$$g_k\left(\frac{t}{B_n}\right) = 1 - \frac{t^2\sigma_k^2}{2B_n^2} + a_{nk},$$

$$a_{nk} = \int_{-\infty}^{\infty} \left(e^{\frac{itx}{B_n}} - 1 - \frac{itx}{B_n} - \frac{1}{2}\left(\frac{itx}{B_n}\right)^2\right) dF_k(x).$$

Note

$$\left|e^{ix}-1-ix\right|\le \frac{x^2}{2},$$
$$\left|e^{ix}-1-ix-\tfrac{1}{2}(ix)^2\right|\le \frac{|x|^3}{6},$$

$$|a_{nk}|\le \int_{-\infty}^{\infty}\frac{(tx)^2}{B_n^2}dF_k(x)=\frac{t^2\sigma_k^2}{B_n^2},$$

$$|a_{nk}|\le \int_{|tx|<\epsilon B_n}\frac{|tx|^3}{6B_n^3}dF_k(x)+\int_{|tx|\ge\epsilon B_n}\frac{(tx)^2}{B_n^2}dF_k(x)$$
$$\le \epsilon t^2\frac{\sigma_k^2}{B_n^2}+t^2\frac{1}{B_n^2}\int_{|tx|\ge\epsilon B_n}x^2dF_k(x).$$

Under the Lindeberg condition (1),

$$\sum_{k=1}^{n}|a_{nk}| \le \epsilon + t^2\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|tx|\ge \epsilon B_n} x^2 dF_k(x) \to 0,$$

as $n \to \infty$ and then $\epsilon \to 0$.

Under the Lindeberg condition (1),

$$\sum_{k=1}^{n} |a_{nk}| \le \epsilon + t^2 \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|tx| \ge \epsilon B_n} x^2 dF_k(x) \to 0,$$

as $n \to \infty$ and then $\epsilon \to 0$.

Write

$$\ln(1+z) = z + \theta(z) z^2, |z| \le \frac{1}{2},$$

where $|\theta(z)| \le 1$ for $|z| \le 1/2$.

Hence

$$\ln g_k\left(\frac{t}{B_n}\right) = -\frac{t^2\sigma_k^2}{2B_n^2} + a_{nk} + \theta_k\left(-\frac{t^2\sigma_k^2}{2B_n^2} + a_{nk}\right)^2,$$

where $|\theta_k| \le 1$.

Hence

$$\ln g_k\left(\frac{t}{B_n}\right) = -\frac{t^2\sigma_k^2}{2B_n^2} + a_{nk} + \theta_k\left(-\frac{t^2\sigma_k^2}{2B_n^2} + a_{nk}\right)^2,$$

where $|\theta_k| \le 1$. It follows that

$$\sum_{k=1}^{n} \ln g_k\left(\frac{t}{B_n}\right) \to -\frac{t^2}{2}.$$

Hence

$$f_n(t) \to e^{-\frac{t^2}{2}}.$$

Now, suppose (3) and (2) are satisfied. Then

$$\begin{aligned}\sum_{k=1}^n\left(-\frac{t^2\sigma_k^2}{2B_n^2}+a_{nk}\right)^2 \le & 4t^2\sum_{k=1}^n\left(\frac{\sigma_k^2}{B_n^2}\right)^2\\ \le & 4t^2\frac{\max_k\sigma_k^2}{B_n}\to 0.\end{aligned}$$

By the CLT (2),

$$\sum_{k=1}^{n} \ln g_k\left(\frac{t}{B_n}\right) \to -\frac{t^2}{2}.$$

So

$$\sum_{k=1}^{n} a_{nk} \to 0.$$

Hence

$$\sum_{k=1}^{n} \int_{-\infty}^{\infty} \left(\cos\left(\frac{tx}{B_n}\right) - 1 - \frac{1}{2}\left(\frac{tx}{B_n}\right)^2\right) dF_k(x) \to 0.$$

Note

$$\cos x - 1 + \frac{1}{2}x^2 = -2\sin^2\frac{x}{2} + \frac{1}{2}x^2 = \frac{1}{2}x^2\left(1 - \left(\frac{\sin\frac{x}{2}}{\frac{x}{2}}\right)^2\right),$$

$$\begin{aligned}
&\int_{-\infty}^{\infty}\left(\cos\left(\frac{tx}{B_n}\right) - 1 - \frac{1}{2}\left(\frac{tx}{B_n}\right)^2\right)dF_k(x) \\
\geq &\int_{|tx|/B_n\geq\epsilon}\left(\cos\left(\frac{tx}{B_n}\right) - 1 - \frac{1}{2}\left(\frac{tx}{B_n}\right)^2\right)dF_k(x) \\
\geq &\frac{1}{2}\left(1 - \left(\frac{\sin\epsilon/2}{\epsilon/2}\right)^2\right)\frac{t^2}{B_n^2}\int_{|tx|/B_n\geq\epsilon} x^2 dF_k(x).
\end{aligned}$$

Hence,

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|tx|/B_n\geq\epsilon} x^2 dF_k(x) \to 0.$$

(1) is satisfied.

Theorem

Let $\{\xi_n; n \geq 1\}$ *be a sequence of independent and identically distributed random variables. Let* $S_n = \Sigma_{k=1}^n \xi_k$. *Suppose*

$$\frac{S_n - na}{\sqrt{n}} \xrightarrow{d} Y \quad as \quad n \to \infty,$$

Then $Var\xi_1 = \sigma^2$ *is finite,* $E\xi_1 = a$ *and* $Y \sim N(0, \sigma^2)$.

Proof. Symmetrization: $\{\xi_n'; n \geq 1\}$ be a sequence of i.i.d. random variables such that $\xi_1' \overset{d}{=} \xi_1$, and the two sequences $\{\xi_n; n \geq 1\}$ and $\{\xi_n', n \geq 1\}$ are independent. Then $\{\eta_n =: \xi_n - \xi_n'; n \geq 1\}$ is a sequence of i.i.d symmetric random variables,

$$\frac{\sum_{k=1}^n \eta_k}{\sqrt{n}} = \frac{\sum_{k=1}^n \xi_k - na}{\sqrt{n}} - \frac{\sum_{k=1}^n \xi_k' - na}{\sqrt{n}} \xrightarrow{d} Y - Y',$$

whereY, Y' are i.i.d. Denote $Z = Y - Y'$.

Truncation: for any $c > 0$, by symmetry,

$$\begin{aligned} \eta_k =& \eta_k I\{|\eta_k| \le c\} + \eta_k I\{|\eta_k| > c\} \\ \text{and } \widetilde{\eta}_k =:& \eta_k I\{|\eta_k| \le c\} - \eta_k I\{|\eta_k| > c\} \end{aligned}$$

have the same distribution. It follows that

$$
\begin{aligned}
&P\left(\left|\frac{\sum_{k=1}^n \eta_k I\{|\eta_k| \le c\}}{\sqrt{n}}\right| \ge x\right)\\
=&P\left(\left|\frac{\sum_{k=1}^n (\eta_k + \widetilde{\eta}_k)}{\sqrt{n}}\right| \ge 2x\right)\\
\le&P\left(\left|\frac{\sum_{k=1}^n \eta_k}{\sqrt{n}}\right| \ge x\right) + P\left(\left|\frac{\sum_{k=1}^n \widetilde{\eta}_k}{\sqrt{n}}\right| \ge x\right)\\
=&2P\left(\left|\frac{\sum_{k=1}^n \eta_k}{\sqrt{n}}\right| \ge x\right).
\end{aligned}
$$

Note $E[\eta_1 I\{|\eta_1|\le c\}]=0$, $Var(\eta_1 I\{|\eta_1|\le c\})=E[\eta_1^2 I\{|\eta_1|\le c\}]=:\sigma_c^2<\infty$.
By CLT,

$$\frac{\sum_{k=1}^n \eta_k I\{|\eta_k|\le c\}}{\sqrt{n}\sigma_c}\xrightarrow{d} N(0,1).$$

Note $E[\eta_1 I\{|\eta_1|\le c\}]=0$, $Var(\eta_1 I\{|\eta_1|\le c\})=E[\eta_1^2 I\{|\eta_1|\le c\}]=:\sigma_c^2<\infty$.
By CLT,

$$\frac{\sum_{k=1}^n \eta_k I\{|\eta_k|\le c\}}{\sqrt{n}\sigma_c}\xrightarrow{d} N(0,1).$$

On the other hand,

$$\frac{\sum_{k=1}^n \eta_k}{\sqrt{n}}\xrightarrow{d} Z.$$

Note $E[\eta_1 I\{|\eta_1| \le c\}] = 0$, $Var(\eta_1 I\{|\eta_1| \le c\}) = E[\eta_1^2 I\{|\eta_1| \le c\}] =: \sigma_c^2 < \infty$.
By CLT,

$$\frac{\sum_{k=1}^n \eta_k I\{|\eta_k| \le c\}}{\sqrt{n}\sigma_c} \xrightarrow{d} N(0,1).$$

On the other hand,

$$\frac{\sum_{k=1}^n \eta_k}{\sqrt{n}} \xrightarrow{d} Z.$$

Taking the limit yields

$$P\left(|N(0,1)| \ge \frac{x}{\sigma_c}\right) \le 2P\left(|Z| \ge x\right), \quad \pm x \in C(F_Z).$$

Choose a $x = x_0$ such that $2P(|Z| \geq x) < P(|N(0,1)| \geq 1)$. Then

$$E[\eta_1^2 I\{|\eta_1| \leq c\}] = \sigma_c^2 \leq x_0^2, \quad \forall c.$$

Letting $c \to \infty$ yields

$$\iint (x-y)^2 dF_\xi(x) dF_\xi(y) = E(\xi_1 - \xi_1')^2 = E\eta_1^2 < \infty.$$

So, exists y such that

$$\int (x-y)^2 dF_\xi(x) < \infty.$$

Hence

$$E\xi_1^2 = \int x^2 dF_\xi(x) \leq 2\int (x-y)^2 dF_\xi(x) + 2y^2 < \infty.$$

By CLT again,

$$\frac{\sum_{k=1}^{n} \xi_k - nE\xi_1}{\sqrt{n}} \xrightarrow{d} N(0, Var(\xi_1)).$$

Hence

$$a = E\xi_1, \quad Y \sim N(0, Var(\xi_1)).$$