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.

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.

4.1 Convergence in distribution and central limit theorems

4.1.1 Weak convergence of distribution functions

Definition 1 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. The condition that $F_n(x) \longrightarrow F(x)$ for every continuity point x of F is not a strong condition in Definition 1. For example, let

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

Then $F_n(x)$ converges pointwise to $F(x)$ except at point $x = 0$, while $x = 0$ is a unique discontinuous point of $F(x)$. Thus it follows from Definition 1 that $F_n \xrightarrow{w} F$.

Remark 3. 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 1 (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

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

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,

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

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

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

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 2 (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

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

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.

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

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

Remark.

1. If $g(x)$ is continuous, and $\{g_t(x)\}$ satisfy $|g_t(x)| \le c$ and $|g_t(x) - g_t(y)| \le |g(x) - g(y)|$, then uniformly in t,
$$\int_{-\infty}^{\infty} g_t(x)dF_n(x) \longrightarrow \int_{-\infty}^{\infty} g_t(x)dF(x).$$

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

Theorem 3 (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 3 (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 3 (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 4 (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 $F_n \xrightarrow{w} F$.

Theorem 4 (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 $F_n \xrightarrow{w} F$.

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

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

Proof. 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

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

Now, for any $u > 0$,

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

Now, for any $u > 0$,

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

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

Now, for any $u > 0$,

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

So,

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

So,

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

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. $\int g(x)dF_n(x) \to \int g(x)dF(x)$ for every bounded, continuous function g;
3. $\int g(x)dF_n(x) \to \int g(x)dF(x)$ for every bounded, uniformly continuous function;
4. $\int g(x)dF_n(x) \to \int g(x)dF(x)$ for every bounded, continuous function g having bounded, continuous derivatives of each order;
5. $f_n(t) \to f(t)$ for all t.

Example 1. Prove the Poisson approximation of binomial distributions by the method of characteristic function.

Proof.

Example 1. 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 1. 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 1. 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 1. 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).

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 $g(\xi_n) \xrightarrow{d} g(\xi)$. (Proof.)

3. Let $\{a_n, n \geq 1\}$ and $\{b_n, n \geq 1\}$ be two sequences of constants, F a distribution function, $\{F_n, n \geq 1\}$ a sequence of distribution functions. If $a_n \to a$, $b_n \to b$, $F_n \xrightarrow{w} F$, then

$$F_n(a_n x + b_n) \to F(ax + b),$$

where x is such that $ax + b$ is a continuity point of F.

Proof. Let x be s.t. $ax+b$ is a continuity point of F, and let $\varepsilon>0$ be s.t. F is continuous on the point $ax+b\pm\varepsilon$.

Proof. Let x be s.t. $ax+b$ is a continuity point of F, and let $\varepsilon > 0$ be s.t. F is continuous on the point $ax+b\pm\varepsilon$. Obviously, $a_n x + b_n \to ax+b$. So, for n large enough,

$$ax+b-\varepsilon \le a_n x + b_n \le ax+b+\varepsilon,$$

Proof. Let x be s.t. $ax+b$ is a continuity point of F, and let $\varepsilon>0$ be s.t. F is continuous on the point $ax+b\pm\varepsilon$. Obviously, $a_nx+b_n\to ax+b$. So, for n large enough,

$$ax+b-\varepsilon\le a_nx+b_n\le ax+b+\varepsilon,$$

which implies

$$F_n(ax+b-\varepsilon)\le F_n(a_nx+b_n)\le F_n(ax+b+\varepsilon).$$

Proof. Let x be s.t. $ax+b$ is a continuity point of F, and let $\varepsilon>0$ be s.t. F is continuous on the point $ax+b\pm\varepsilon$. Obviously, $a_nx+b_n\to ax+b$. So, for n large enough,

$$ax+b-\varepsilon\le a_nx+b_n\le ax+b+\varepsilon,$$

which implies

$$F_n(ax+b-\varepsilon)\le F_n(a_nx+b_n)\le F_n(ax+b+\varepsilon).$$

Since $F_n\xrightarrow{w}F$, letting $n\to\infty$ yields

Proof. Let x be s.t. $ax+b$ is a continuity point of F, and let $\varepsilon > 0$ be s.t. F is continuous on the point $ax+b\pm\varepsilon$. Obviously, $a_nx+b_n \to ax+b$. So, for n large enough,

$$ax+b-\varepsilon \le a_nx+b_n \le ax+b+\varepsilon,$$

which implies

$$F_n(ax+b-\varepsilon) \le F_n(a_nx+b_n) \le F_n(ax+b+\varepsilon).$$

Since $F_n \xrightarrow{w} F$, letting $n\to\infty$ yields

$$\begin{aligned} F(ax+b-\varepsilon) &\le \liminf_{n\to\infty} F(a_nx+b_n) \\ \le &\limsup_{n\to\infty} F_n(a_nx+b_n) \le F(ax+b+\varepsilon). \end{aligned}$$

Proof. Let x be s.t. $ax+b$ is a continuity point of F, and let $\varepsilon>0$ be s.t. F is continuous on the point $ax+b\pm\varepsilon$. Obviously, $a_nx+b_n\to ax+b$. So, for n large enough,

$$ax+b-\varepsilon\le a_nx+b_n\le ax+b+\varepsilon,$$

which implies

$$F_n(ax+b-\varepsilon)\le F_n(a_nx+b_n)\le F_n(ax+b+\varepsilon).$$

Since $F_n\xrightarrow{w}F$, letting $n\to\infty$ yields

$$\begin{aligned}F(ax+b-\varepsilon)&\le \liminf_{n\to\infty}F(a_nx+b_n)\\ \le\ &\limsup_{n\to\infty}F_n(a_nx+b_n)\le F(ax+b+\varepsilon).\end{aligned}$$

Letting $\varepsilon\to 0$, the proof is complete.

Corollary. If $\xi_n \xrightarrow{d} \xi$, $a_n \to a$, $b_n \to b$, then $a_n\xi_n + b_n \xrightarrow{d} a\xi + b$, $(a_n, a \neq 0)$.

Proof.

Corollary. If $\xi_n \xrightarrow{d} \xi$, $a_n \to a$, $b_n \to b$, then $a_n\xi_n + b_n \xrightarrow{d} a\xi + b$, $(a_n, a \neq 0)$.

Proof. It suffices to observe that the distribution functions of $a_n\xi_n + b_n$ and $a\xi + b$ are $F_n(\frac{x-b_n}{a_n})$ and $F(\frac{x-b}{a})$ respectively, when $a_n > 0$ and $a > 0$;

Corollary. If $\xi_n \xrightarrow{d} \xi$, $a_n \to a$, $b_n \to b$, then $a_n\xi_n + b_n \xrightarrow{d} a\xi + b$, $(a_n, a \neq 0)$.

Proof. It suffices to observe that the distribution functions of $a_n\xi_n + b_n$ and $a\xi + b$ are $F_n(\frac{x-b_n}{a_n})$ and $F(\frac{x-b}{a})$ respectively, when $a_n > 0$ and $a > 0$; and are $1 - F_n(\frac{x-b_n}{a_n} - 0)$ and $1 - F(\frac{x-b}{a} - 0)$ respectively, when $a_n < 0$ and $a < 0$.

4.1.3 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

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

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

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

i.e.,

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

When n is big enough, p is moderate, then

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

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 $B(n,p)$).

Example 3. Rolling a fair coin, how many times need one roll to ensure the probability the proportion of heads is between 0.4 and 0.6 is not smaller than 90%.

Solution.

Example 3. Rolling a fair coin, how many times need one roll to ensure the probability the proportion of heads is between 0.4 and 0.6 is not smaller than 90%.

Solution. Let n be the number of times of rolling the coin, S_n the number of times of appearing head, then $S_n \sim B(n, 1/2)$. Note that n is to satisfy

$$P(0.4 < \frac{S_n}{n} \leq 0.6) \geq 0.9.$$

From Theorem 5,

$$
\begin{aligned}
& P(0.4 < \frac{S_n}{n} \le 0.6) \\
= \quad & P(\frac{0.4n - n/2}{\sqrt{n/4}} < \frac{S_n - n/2}{\sqrt{n/4}} \le \frac{0.6n - n/2}{\sqrt{n/4}})
\end{aligned}
$$

From Theorem 5,

$$
\begin{aligned}
& P(0.4 < \frac{S_n}{n} \leq 0.6) \\
= \;& P(\frac{0.4n - n/2}{\sqrt{n/4}} < \frac{S_n - n/2}{\sqrt{n/4}} \leq \frac{0.6n - n/2}{\sqrt{n/4}}) \\
\approx \;& \Phi(0.2\sqrt{n}) - \Phi(-0.2\sqrt{n}) \\
= \;& 2\Phi(0.2\sqrt{n}) - 1.
\end{aligned}
$$

From Theorem 5,

$$
\begin{aligned}
& P(0.4 < \frac{S_n}{n} \leq 0.6) \\
= \ & P(\frac{0.4n - n/2}{\sqrt{n/4}} < \frac{S_n - n/2}{\sqrt{n/4}} \leq \frac{0.6n - n/2}{\sqrt{n/4}}) \\
\approx \ & \Phi(0.2\sqrt{n}) - \Phi(-0.2\sqrt{n}) \\
= \ & 2\Phi(0.2\sqrt{n}) - 1.
\end{aligned}
$$

Take $n \geq 69$ such that the above equality ≥ 0.9.

Definition 2 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

$$\frac{1}{B_n}\sum_{k=1}^{n}\xi_k - A_n \xrightarrow{d} N(0,1),$$

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

Theorem 6 (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.,

$$\frac{S_n - na}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1) \quad as \quad n \to \infty.$$

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 $t \in \mathbf{R}$,

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

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

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

Therefore

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

The later is the c.f. of $N(0,1)$. Then Theorem 6 follows from Theorem 4.

Example 4. 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 4. 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 6 (Lindeberg-Feller) Suppose that $\{\xi_k, k \ge 1\}$ is a sequence of indept. r.v.s. If the Lindeberg condition is satisfied:

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

then

$$\frac{\sum_{k=1}^n(\xi_k - E\xi_k)}{B_n} \xrightarrow{d} \Phi(x). \tag{2}$$

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

Theorem 6 (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

$$\frac{\sum_{k=1}^n (\xi_k - E\xi_k)}{B_n} \xrightarrow{d} \Phi(x). \tag{2}$$

Conversely, if (2) and

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

then (1) holds.

Theorem 7 (Lyapunov) Suppose that $\{\xi_k, k \geq 1\}$ is a sequence of indep. r.v.s, which satisfy

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

then the central limit theorem holds true.

Proof.

Theorem 7 (Lyapunov) Suppose that $\{\xi_k, k \geq 1\}$ is a sequence of indep. r.v.s, which satisfy

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

then the central limit theorem holds true.

Proof.

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

$$\leq$$

Theorem 7 (Lyapunov) Suppose that $\{\xi_k, k \geq 1\}$ is a sequence of indep. r.v.s, which satisfy

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

then the central limit theorem holds true.

Proof.

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

$$\leq \frac{1}{\epsilon^\delta} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^n E|\xi_k - E\xi_k|^{2+\delta} \to 0.$$

Example 7. 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. $S = \Sigma_{i=1}^{1800} X_i$. θ is to satisfy $P(S \le \pi(S)) = 0.95$. While,

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

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

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

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

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

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

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