MH2500 Probability and Introduction to Statistics

Handout 9 - Limit Theorems

Synopsis

We discuss two famous theorems in probability, the law of large numbers, and the central limit theorem.
The law of large numbers, essentially says that when an experiment is repeated many times, the proportion of experiments with successful outcomes is approximately the probability of success. The central limit theorem quantifies how close the proportion is to the probability.

- The Law of Large Numbers
- Convergence in Distribution and the Central Limit Theorem

The Law of Large Numbers

1. If we toss a fair coin many times, we expect about half of them are heads.

2. While detained as a prisoner during World War II, South African mathematician John Kerrich tossed a coin 10,000 times. He obtained 5067 heads.

3. Mathematically, the law of large numbers tells us that happens.

4. The coin tosses are random independent events. Let the random variable X_i denote whether we obtain a head on the i-th toss, i.e., 1 if it is a head and 0 if it is a tail. Then the X_i's are independent.

5. The proportion of heads is then given by

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

6. The law of large numbers tells us that $\overline{X}_n$ approaches $\frac{1}{2}$.

Law of large numbers

In the case of a fair coin toss, the X_i's are Bernoulli random variables with $p = 1/2$,

$$E(X_i) = \frac{1}{2} \qquad Var(X_i) = \frac{1}{4},$$

and for $n = 10000$,

$$\mathrm{Var}(\overline{X}_{10,000}) = \qquad .$$

Hence the standard deviation is .

Kerrich's observation was 0.5067. This is 0.0067 away from the mean, which is slightly more than one standard deviation away.

This is consistent with Chebyshev's inequality in the form

$$P(|\overline{X}_n - \mu| > k\sigma) \leq 1/k^2.$$

Law of large numbers

Theorem A – Law of Large Numbers

Let $X_1, X_2, \ldots, X_i, \ldots$ be a sequence of independent random variables with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Let $\overline{X_n} = n^{-1} \sum_{i=1}^{n} X_i$. Then, for any $\epsilon > 0$,

$$P(|\overline{X}_n - \mu| > \epsilon) \to 0 \qquad \text{as } n \to \infty.$$

Proof of Theorem A

Proof: We first find $E(\overline{X}_n)$ and $\mathrm{Var}(\overline{X}_n)$:

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

Since the X_i are independent,

$$\mathrm{Var}(X_n) = \qquad\qquad = \quad .$$

The desired result now follows immediately from Chebyshev's inequality, which states that

$$P(|\overline{X_n} - \mu| > \epsilon) \leq$$

$$= \qquad \to 0 \qquad \text{as } n \to \infty$$

Convergence

If a sequence of random variables, $\{Z_n\}$, is such that

for any $\epsilon > 0$ and for some scalar α,

$$P(|Z_n - \alpha| > \epsilon) \to 0 \qquad \text{as } n \to \infty$$

then Z_n is said to **converge in probability** to α.

There is a stronger version of convergence. Z_n is said to **converge almost surely to** α if

$$P(|Z_n \to \alpha|) = 1.$$

Theorem A asserts that $\overline{X}_n$ converges in probability to μ and is also known as the *weak* law of large numbers. There is a *strong* law of large numbers which asserts that $\overline{X}_n$ converges almost surely to μ but we will not discuss this.

Example

Ex.5.4.1

Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma_i^2$. Show that if $n^{-2}\sum_{i=1}^{n} \sigma_i^2 \to 0$, then $\overline{X}_n \to \mu$ in probability.

. . . .

First, note that

$$\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 \to 0 \qquad \text{as } n \to \infty.$$

By Chebyshev's inequality, for any $\epsilon > 0$,

$$P(|\overline{X}_n - \mu| > \epsilon) \leq$$

Hence $\overline{X}_n$ converges to μ in probability.

Convergence in Distribution

In this section, we begin with introductory terminology and theory, to prepare us for the most famous limit theorem in probability theory, the central limit theorem, which describes a way to approximate probabilities of a sequence of independent random variables.

Definition

Let $X_1, X_2, \ldots$ be a sequence of random variables with cdf $F_1, F_2, \ldots$ and let X be a random variable with cdf F.

We say that X_n **converges in distribution** to X if

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

at every point x where F is continuous.

Continuity Theorem

Recall that a distribution function is uniquely determined by its mgf M_n. The following theorem (proof omitted) states that this unqiue determination holds for limits as well.

Theorem A (Continuity Theorem)

Let F_n be a sequence of cdf with the corresponding moment-generating function M_n. Let F be a cdf with the moment-generating function M.

If $M_n(t) \to M(t)$ for all t in an open interval containing zero, then $F_n(x) \to F(x)$ at all continuity points of F.

Example A – Approximating Poisson

Show that the Poisson distribution can be approximated by the normal distribution for large values of λ.

Let $\lambda_1, \lambda_2, \ldots$ be an increasing sequence with $\lambda_n \to \infty$ and let $\{X_n\}$ be a sequence of Poisson random variables with the corresponding parameters.

. . . .

Since the X_n's are Poisson, $E(X_n) = Var(X_n) = \lambda_n$.

To approximate the Poisson distribution by a normal distribution function, the normal must have the same mean and variance as the Poisson does.

However $\lambda_n \to \infty$ then creates a problem.

Example A con't

This is overcomed by standardizing the random variables, that is, by letting

$$Z_n = \qquad\qquad = \qquad\qquad .$$

Then $E(Z_n) = 0$, and $\text{Var}(Z_n) = 1$.

Next, we show that the mgf of Z_n converge to the mgf of the standard normal.

The mgf of X_n is
$$M_{X_n}(t) = e^{\lambda_n(e^t - 1)}.$$

By property C of §4.5, the mgf of Z_n is

$$\begin{aligned} M_{Z_n}(t) &= e^{-t\sqrt{\lambda_n}} M_{X_n}\left(\frac{1}{\sqrt{\lambda_n}}\right) \\ &= e^{-t\sqrt{\lambda_n}} e^{\lambda_n(e^{t/\sqrt{\lambda_n}} - 1)}. \end{aligned}$$

Example A con't

It is easier to work with the expression

$$\log M_{Z_n}(t) = -t\sqrt{\lambda_n} + \lambda_n(e^{t/\sqrt{\lambda_n}} - 1).$$

Using the power series expansion $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, we see that

$$\lim_{n\to\infty} \log M_{Z_n}(t) = \frac{t^2}{2}$$

or

$$\lim_{n\to\infty} M_{Z_n}(t) = e^{t^2/2}.$$

This last expression is the mgf of the standard normal distribution.

Thus we have shown that a standardized Possion random variable converges in distribution to a standard normal variable as $\lambda \to \infty$.

Example B

A certain type of particle is emitted at a rate of 900 per hour. What is the probability that more than 950 particles will be emitted in a given hour if the counts form a Poisson process?

Let X be a Poisson random variable with mean 900.

$$P(X > 950) =$$

$$\approx$$

$$= \qquad ,$$

where Φ is the standard normal cdf. For comparison, the exact proability is 0.04712.

Suppose $X_1, X_2, \ldots$ is a sequence of independent random variables with mean μ and variance σ^2, and suppose

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

We know from the law of large numbers that S_n/n converges to μ in probability. This followed from the fact that

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

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.

Standardizing the S_n's, set

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

so that Z_n has mean 0 and standard deviation 1. Then then the central limit theorem states that Z_n converges to the standard normal.

Theorem B – Central Limit Theorem

Central Limit Theorem

Let $X_1, X_2, \ldots$ be a sequence of independent random variables having mean 0 and variance σ^2 and the common distribution function F and moment-generating function M defined in a neighbourhood of zero. Let

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

Then

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) = \Phi(x), \qquad -\infty < x < \infty.$$

Proof of CTL

Proof: Let $Z_n = S_n/(\sigma\sqrt{n})$. We show that the mgf of Z_n tends to the mgf of the standard normal distributions.

Since S_n is a sum of independent random variables,

$$M_{S_n}(t) = [M(t)]^n$$
$$M_{Z_n}(t) = \left[M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

$M(s)$ has a Taylor series expansion about zero:

$$M(s) = M(0) + sM'(0) + \frac{1}{2}M''(0) + \epsilon_s,$$

where $\epsilon_s/s^2 \to 0$ as $s \to 0$. Since $E(X) = 0$, $M'(0) = 0$, and $M''(0) = \sigma^2$.

Proof of CLT con't

As $n \to \infty$, we see that $t/(\sigma\sqrt{n}) \to 0$, and so

$$M\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \epsilon_n,$$

where $\dfrac{\epsilon_n}{t^2/(n\sigma^2)} \to 0$ as $n \to \infty$.
Thus we have

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + \epsilon_n\right)^n.$$

It can be shown that if $a_n \to a$, then

$$\lim_{n\to\infty}\left(1 + \frac{a_n}{n}\right)^n = e^a.$$

From this result, it follows

$$M_{Z_n}(t) = e^{t^2/2} \qquad \text{as } n \to \infty,$$

where $\exp(t^2/2)$ is the mgf of the standard normal distribution.

Remarks

1. We have given one of the simplest version of the central limit theorem. (There are many other versions.)
2. We assumed the mgf exists, which is a rather strong assumption.
3. For practical purposes, e.g. for statistics, the limiting result is not of primary interest.
4. Statisticians are more interested in its use as an approximation with finite values of n.

Example C

Let $U_1, U_2, \ldots, U_{12}$ be independent uniform random variables on [0,1]. Let

$$Z = \left(\sum_{n=1}^{12} U_n \right) - 6$$

and $S_n = \sum_{i=1}^{n} Z_i$. Explain why this distribution of S_n is close to normal.

. . . .

Clearly $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$.

Thus, by the central limit theorem, as n gets large, S_n is approximately normal.

Also, as the distribution is symmetric about zero, the approximation is even for small n.

Example D

Recall the sum of n independent exponential ranom variables with parameter $\lambda = 1$ follows a gamma distribution with $\lambda = 1$ and $\alpha = n$. Would the standard normal to be a good approximation for the standardized gamma for small values of n? Why?

. . . .

As the exponential density is skewed, we do not expect to get a good approximation for small values of n. The approximation imporves as n increases.

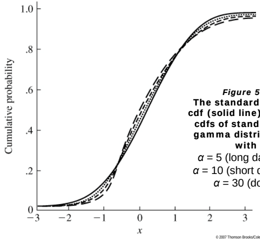

Figure 5.2
The standard normal cdf (solid line) and the cdfs of standardized gamma distributions with
$\alpha = 5$ (long dashes),
$\alpha = 10$ (short dashes),
$\alpha = 30$ (dots).

Normal approximation to the Binomial Distribution

Normal approximation to Binomail

Binomial distributions may be approximated by a normal distribution.

The approximation is best when the binomial distribution is symmetric, i.e., when $p = 1/2$.

The approximation is reasonable when $np > 5$ and $np(1-p) > 5$.

The approximation is useful for large values of n, for which tables are not readily available.

Example F

Suppose that a coin is tossed 100 times and lands up heads 60 times. Should we be surprised that the coin is fair?

Suppose the coin is fair. Then the probability of obtaining a head is $1/2$. Let X denote the number of heads obtained in tossing the coin 100 times. Then X has a binomial distribution with $n = 100$, $p = \frac{1}{2}$.
Then

$$E(X) = np = 50 \qquad \text{and} \qquad \mathrm{Var}(X) = np(1-p) = 25.$$

$P(X = 50)$ or $P(X = 60)$ are all small numbers and would not help to answer the question. Rather, we calculate the probability of "a deviation as extreme as or more extreme than 60", i.e., $P(X \geq 60)$.

We approximate this probability using the normal distribution:

$$\begin{aligned} P(X \geq 60) &= \\ &\approx \\ &= \end{aligned}$$

The probability is rather small, so the fairness of the coin is questionable.