Probability & Statistics (1)

Limit Theorems (I)

Asst. Prof. Chan, Chun-Hsiang

Master program in Intelligent Computing and Big Data, Chung Yuan Christian University, Taoyuan, Taiwan
Undergraduate program in Intelligent Computing and Big Data, Chung Yuan Christian University, Taoyuan, Taiwan
Undergraduate program in Applied Artificial Intelligence, Chung Yuan Christian University, Taoyuan, Taiwan

Outlines

Introduction

- 在機率論裡面，**極限定理(limit theorem)**扮演舉足輕重的地位。
  其中，有兩個定理特別重要:
  - **大數法則(laws of large numbers)**
  - **中央極限定理(central limit theorems)**

Chebyshev's Inequality and the Weak Law of Large Numbers

- 我們先從一個比較知名的不等式開始 – Markov's Inequality
- **Proposition 1 Markov's Inequality**

If X is a random variable that takes only nonnegative values, then, for any value $a > 0$,

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Proof:

For $a > 0$, let

$$I = \begin{cases} 1 & if\ X \geq a \\ 0 & otherwise \end{cases}, since\ X \geq 0 \Rightarrow I \leq \frac{X}{a}$$

Chebyshev’s Inequality and the Weak Law of Large Numbers

Take expectations of the preceding inequality yields

$$E[I] \leq \frac{E[X]}{a}$$

$$which, because\ E[I] = P\{X \geq a\}, proves\ the\ result$$

As a corollary, we obtain **Proposition 2**.

Chebyshev's Inequality and the Weak Law of Large Numbers

- **Proposition 2 Chebyshev's Inequality**

If X is a random variable with finite mean μ and variance σ^2, then, for any value $k > 0$,

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Proof:

Since $(X-\mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P\{(X-\mu)^2 \geq k^2\} <= \frac{E[(X-\mu)^2]}{k^2}$$

Chebyshev's Inequality and the Weak Law of Large Numbers

$$P\{(X-\mu)^2 \geq k^2\} <= \frac{E[(X-\mu)^2]}{k^2}$$

Since $(X-\mu)^2 \geq k^2$ if and only if $|X-\mu| \geq k$, so we obtain

$$P\{|X-\mu| \geq k\} \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

And the proof is complete.

Chebyshev's Inequality and the Weak Law of Large Numbers

- **範例一**

假設今天某一半導體工廠生廠A晶圓，一周可以生產的片數為隨機變數，目前已知平均一周為50片。

試問: (a) 生產超過75片的機率為何? (b) 若已知生產片數的變異數為25，那麼生產數量介於40至60片的機率為何?

Solution:

Let X be the number of items that will be produced in a week.

(a) By Markov's inequality [生產超過75片的機率為何?]

$$P\{X > 75\} \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

Chebyshev’s Inequality and the Weak Law of Large Numbers

(a) By Chebyshev’s inequality [若已知生產片數的變異數為25，那麼生產數量介於40至60片的機率為何?]

$$P\{|X-50| \geq 10\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

Hence,

$$P\{|X-50| < 10\} \geq 1 - \frac{1}{4} = \frac{3}{4}$$

So the probability that this week’s production will be between 40 and 60 is at least 75%.

Chebyshev’s Inequality and the Weak Law of Large Numbers

- **範例二**

如果X為uniform distributed over (0,10)，已知$E[X]=5$與$Var(X)=\frac{25}{3}$，且符合Chebyshev's inequality

$$P\{|X-5|>4\}\leq\frac{25}{3(16)}\approx 0.52$$

Whereas the exact result is

$$P\{|X-5|>4\}=0.2$$

Chebyshev's Inequality and the Weak Law of Large Numbers

- **Proposition 3**

If $Var(X) = 0$, then

$$P\{X = E[X]\} = 1$$

In other words, the only random variables having variances equal to 0 are those which are constant with probability 1.

Proof:

By Chebyshev's inequality, we have, for any $n \geq 1$,

$$P\left\{|X - \mu| > \frac{1}{n}\right\} = 0$$

Letting $n \to \infty$ and using the continuity property of probability yields

$$0 = \lim_{n\to\infty} P\left\{|X - \mu| > \frac{1}{n}\right\} = P\left\{\lim_{n\to\infty}\left\{|X - \mu| > \frac{1}{n}\right\}\right\} = P\{X \neq \mu\}$$

Chebyshev's Inequality and the Weak Law of Large Numbers

- **Theorem 1 The Weak Law of Large Numbers**

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$. Then, for any $\varepsilon > 0$,

$$P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0 \; as \; n \to \infty$$

Chebyshev's Inequality and the Weak Law of Large Numbers

Proof:

We shall prove the theorem only under the additional assumption that the random variables have a finite variance σ^2. Now, since

$$E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = \mu \text{ and } Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

It follows from Chebyshev's inequality that

$$P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

And the result is proven.

The Central Limit Theorem

- 中央極限定理(central limit theorem, CLT)在機率論裡面最重要的定理之一。簡單來說，這個定理就在說當你今天有大量相互獨立隨機變數的均值，其分布會收斂於常態分布(normal distribution)。
- **Theorem 2 The Central Limit Theorem**
- Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

Tends to be standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$

The Central Limit Theorem

$$P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}}\, dx \text{ as } n \to \infty$$

The key to the proof of the central limit theorem is the following lemma, which we state without proof.

Lemma 1

Let $Z_1, Z_2, \ldots$ be a sequence of random variables having distribution functions F_Z and moment generating functions $M_{Z_n}, n \geq 1$; and let Z be a random variable having distribution function F_Z and moment generating function M_Z. If $M_{Z_n}(t) \to M_Z(t)$ for all t, then $F_{Z_n}(t) \to F_Z(t)$ for all t at which $F_Z(t)$ is continuous.

The Central Limit Theorem

- **Proof of the Central Limit Theorem**

Let us assume at first that $\mu = 0$ and $\sigma^2 = 1$. We shall prove the theorem under the assumption that the moment generating function of the X_i, $M(t)$, exists and is finite. Now, the moment generating function of $\frac{X_i}{\sqrt{n}}$ is given by

$$E\left[\exp\left\{\frac{tX_i}{\sqrt{(n)}}\right\}\right] = M\left(\frac{t}{\sqrt{(n)}}\right)$$

Thus, the moment generating function of $\sum_{i=1}^{n}\frac{X_i}{\sqrt{n}}$ is given by $\left[M\left(\frac{t}{\sqrt{(n)}}\right)\right]^n$.

Let $L(t) = logM(t)$

The Central Limit Theorem

- And note that

$$L(0) = 0$$

$$L'(0) = \frac{M'(0)}{M(0)} = \mu = 0$$

$$L''(0) = \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} = E[X^2] = 1$$

The Central Limit Theorem

- Now, to prove the theorem, we must show that $\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n \to e^{\frac{t^2}{2}}$ as $n \to \infty$, or, equivalently, that $nL\left(\frac{t}{\sqrt{(n)}}\right) \to \frac{t^2}{2}$ as $n \to \infty$. To show this, note that

$$\lim_{n\to\infty} \frac{L\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}} = \lim_{n\to\infty} \frac{-L'\left(\frac{t}{\sqrt{n}}\right) n^{-\frac{3}{2}} t}{-2n^{-2}}$$

$$= \lim_{n\to\infty} \left[\frac{L'\left(\frac{t}{\sqrt{n}}\right) t}{2n^{-\frac{1}{2}}}\right] = \lim_{n\to\infty} \left[\frac{L''\left(\frac{t}{\sqrt{n}}\right) n^{-\frac{3}{2}} t^2}{-2n^{-\frac{3}{2}}}\right] = \lim_{n\to\infty} [L''\left(\frac{t}{\sqrt{n}}\right)\frac{t^2}{2}] = \frac{t^2}{2}$$

The Central Limit Theorem

- 範例三

如果投擲10顆公平的骰子，試問點數合介於30至40點之間的機率為何?

Solution:

Let X_i denote the value of the ith die, $i = 1,2,\ldots,10$. Since

$$E(X_i) = \frac{7}{2}, Var(X_i) = E\left[X_i^2\right] - (E[X])^2 = \frac{35}{12}$$

The central limit theorem yields

$$P\{29.5 \le X \le 40.5\} = P\left\{\frac{29.5-35}{\sqrt{\frac{350}{12}}} \le \frac{X-35}{\sqrt{\frac{350}{12}}} \le \frac{40.5-35}{\sqrt{\frac{350}{12}}}\right\}$$

$$\approx 2\Phi(1.0184) - 1 \approx 0.692$$

The Central Limit Theorem

- **範例四**

令$X_i, i = 1,2,\ldots,10$，為獨立隨機變數，uniformly distributed over (0,1)。試計算$P\{\sum_{i=1}^{10} X_i > 6\}$的近似值。

Solution:

Since $E[X_i] = \frac{1}{2}$ $and\ Var(X_i) = \frac{1}{12}$, we have, by the central limit theorem,

$$P\left\{\sum_{i=1}^{10} X_i > 6\right\} = P\left\{\frac{\sum_1^{10} X_i - 5}{\sqrt{10\left(\frac{1}{12}\right)}} > \frac{6-5}{\sqrt{10\left(\frac{1}{12}\right)}}\right\} \approx 1 - \Phi\left(\sqrt{1.2}\right) \approx 0.1367$$

The Central Limit Theorem

- **範例五**

期末考完，某位老師需要批改50份考卷。假設批改每一份考卷都是獨立，平均需要花20分鐘，標準差為4分鐘。試問: 該老師在最一開始的450分鐘內至少批改完25份考卷的機率為何?

Solution:

If we let X_i be the time that it takes to grade exam i, then

$$X = \sum_{i=1}^{25} X_i$$

is the time it takes to grade the first 25 exams.

The Central Limit Theorem

$$E[X] = \sum_{i=1}^{25} E[X_i] = 25(20) = 500$$

$$Var(X) = \sum_{i=1}^{25} Var(X_i) = 25(16) = 400$$

Consequently, with Z being a standard normal random variable, we have

$$P\{X \le 450\} = P\left\{\frac{X - 500}{\sqrt{(400)}} \le \frac{450 - 500}{\sqrt{(400)}}\right\} \approx P\{Z \le -2.5\}$$

$$= P\{Z \ge 2.5\} = 1 - \Phi(2.5) = 0.006$$

The Strong Law of Large Numbers

- **Theorem 3 Central Limit Theorem for Independent Random Variables**

Let $X_1, X_2, \dots,$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \; as \; n \to \infty$$

As an application of the strong law of large numbers, suppose that a sequence of independent trials of some experiment is performed.

The Strong Law of Large Numbers

Let E be a fixed event of the experiment, and denote by $P(E)$ the probability that E occurs on any particular trial.

Letting

$$X_i = \begin{cases} 1 & if\ E\ occurs\ on\ the\ ith\ trial \\ 0 & if\ E\ dose\ not\ occur\ on\ theith\ trial \end{cases}$$

We have, by the strong law of large numbers, that with probability 1.

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow E[X] = P(E)$$

Since $X_1 + X_2 + \cdots + X_n$ represents the number of times that the event E occurs in the first n trials, we may interpret abovementioned equation as starting that, with probability 1, the limiting proportion of time that the event E occurs is just $P(E)$.

The Strong Law of Large Numbers

- Although the theorem can be proven without this assumption, our proof of the strong law of large numbers will assume that the random variables X_i have a finite fourth moment. That is, we will suppose that $E[X_i^4] = P(E)$.

The Strong Law of Large Numbers

- **Proof:**

Assume that μ, the mean of the X_i, is equal to 0. Let $S_n = \sum_{i=1}^{n} X_i$ and consider

$$E[S_n^4]$$
$$= E[(X_1 + \cdots + X_n)(X_1 + \cdots + X_n) \times (X_1 + \cdots + X_n)\,(X_1 + \cdots + X_n)]$$

Expanding the right side of the preceding equation results in terms of the form

$$X_i^4, X_i^3 X_j, X_i^2 X_j^2, X_i^2 X_j X_k, and\ X_i X_j X_k X_l$$

where $i, j, k, and\ l$ are all different. Because all the X_i have mean 0, it follows by independence that

$$E[X_i^3 X_j] = E[X_i^3]E[X_j] = 0$$

The Strong Law of Large Numbers

$$E[X_i^2 X_j X_k] = E[X_i^2]E[X_j]E[X_k] = 0$$
$$E[X_i X_j X_k X_l] = E[X_i]E[X_j]E[X_k]E[X_l] = 0$$

Now, for a given pair i and j, there will be $\binom{4}{2} = 6$ terms in the expansion that will equal $X_i^2 X_j^2$. Hence, upon expanding the preceding product and taking expectations term by term, it follows that

$$E[S_n^4] = nE\left[X_i^4\right] + 6\binom{n}{2}E\left[X_i^2 X_j^2\right] = nK + 3n(n-1)E\left[X_i^2\right]E[X_j^2]$$

Where we have once again made use of the independence assumption.

The Strong Law of Large Numbers

- Now since

$0 \leq Var(X_i^2) = E[X_i^4] - (E[X_i^2])^2$

We have $(E[X_i^2])^2 \leq E[X_i^4] = K$

Therefore, from the preceding, we obtain

$E[S_i^4] \leq nK + 3n(n-1)K$

which implies that

$E\left[\frac{S_n^4}{n^4}\right] \leq \frac{K}{n^3} + \frac{3K}{n^2}$

The Strong Law of Large Numbers

Therefore,

$$E\left[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4}\right] = \sum_{n=1}^{\infty} E\left[\frac{S_n^4}{n^4}\right] < \infty$$

But the preceding implies that, with probability 1, $\sum_{n=1}^{\infty} \frac{S_n^4}{n^4} < \infty$. (For if there is a positive probability that the sum is infinite, then its expected value is infinite.) But the convergence of a series implies that its nth term goes to 0; so we can conclude that, with probability 1,

$$\lim_{n \to \infty} \frac{S_n^4}{n^4} = 0$$

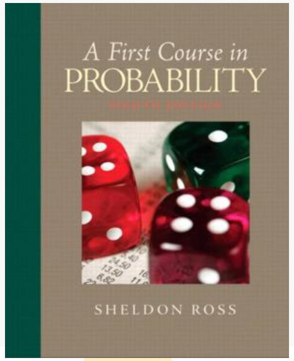
The Strong Law of Large Numbers

- But if $\frac{S_n^4}{n^4} = \left(\frac{S_n}{n}\right)^4$ goes to 0, then so must $\frac{S_n}{n}$; hence, we have proven that, with probability 1,
- $\frac{S_n}{n} \to 0 \; as \; n \to \infty$
- When μ, the mean of the X_i, is not equal to 0, we can apply the preceding argument to the random variables $X_i - \mu$ to obtain that with probability 1,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{(X_i - \mu)}{n} = 0; or\ equivalently, \lim_{n\to\infty} \sum_{i=1}^{n} \frac{X_i}{n} = \mu$$

which proves the result.

[#14] Assignment



- Selected Problems from Sheldon Ross Textbook [1].

8.1. Suppose that X is a random variable with mean and variance both equal to 20. What can be said about $P\{0 < X < 40\}$?

8.2. From past experience, a professor knows that the test score of a student taking her final examination is a random variable with mean 75.

(a) Give an upper bound for the probability that a student's test score will exceed 85. Suppose, in addition, that the professor knows that the variance of a student's test score is equal to 25.

(b) What can be said about the probability that a student will score between 65 and 85?

(c) How many students would have to take the examination to ensure, with probability at least .9, that the class average would be within 5 of 75? Do not use the central limit theorem.

8.4. Let $X_1, \ldots, X_{20}$ be independent Poisson random variables with mean 1.

(a) Use the Markov inequality to obtain a bound on

$$P\left\{\sum_1^{20} X_i > 15\right\}$$

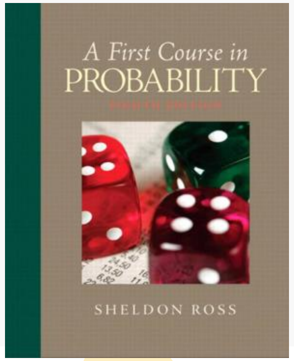
(b) Use the central limit theorem to approximate

$$P\left\{\sum_1^{20} X_i > 15\right\}.$$

8.5. Fifty numbers are rounded off to the nearest integer and then summed. If the individual round-off errors are uniformly distributed over $(-.5, .5)$, approximate the probability that the resultant sum differs from the exact sum by more than 3.

[1] Sheldon Ross. A First of Course in Probability. 8th edition.

[#14] Assignment



8.6. A die is continually rolled until the total sum of all rolls exceeds 300. Approximate the probability that at least 80 rolls are necessary.

8.7. A person has 100 light bulbs whose lifetimes are independent exponentials with mean 5 hours. If the bulbs are used one at a time, with a failed bulb being replaced immediately by a new one, approximate the probability that there is still a working bulb after 525 hours.

[1] Sheldon Ross. A First of Course in Probability. 8th edition.

Reference

Ross, S. (2010). *A first course in probability*. Pearson.

The End

If you have any questions, please do not hesitate to ask me.

Thank you for your attention))