Large Random Samples

Chapter 6 of Probability and Statistics

Yiping Cheng
ypcheng@bjtu.edu.cn
10-Jul-2017

School of Electronic and Information Engineering
Beijing Jiaotong University

Chapter Contents

The Chebyshev Inequality

The Chebyshev Inequality – Version 1

Theorem [Chebyshev Inequality]: Let X be a random variable such that $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$, and $0 < \sigma^2 < \infty$. Then for every $\kappa > 0$,

$$P(|\frac{X-\mu}{\sigma}| \geq \kappa) \leq \frac{1}{\kappa^2}.$$

Proof.

Let $Z = \frac{X-\mu}{\sigma}$ and $W = \begin{cases} 0 & \text{for } |Z| < \kappa, \\ \kappa^2 & \text{for } |Z| \geq \kappa. \end{cases}$ We have $W \leq Z^2$. Then

$$1 = E(Z^2) \geq E(W) = \kappa^2 \cdot P(W = \kappa^2) = \kappa^2 \cdot P(|Z| \geq \kappa).$$

Thus finally, $P(|Z| \geq \kappa) \leq \frac{1}{\kappa^2}$. □

The Chebyshev Inequality – Version 2

Theorem [Chebyshev Inequality]: Let X be a random variable such that $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$, and $\sigma^2 < \infty$. Then for every $t > 0$,

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Proof.

Let $Y = X - \mu$ and $W = \begin{cases} 0 & \text{for } |Y| < t, \\ t^2 & \text{for } |Y| \geq t. \end{cases}$ We have $W \leq Y^2$. Then

$$\sigma^2 = E(Y^2) \geq E(W) = t^2 \cdot P(W = t^2) = t^2 \cdot P(|Y| \geq t).$$

Thus finally, $P(|Y| \geq t) \leq \frac{\sigma^2}{t^2}$. □

Properties of the Sample Mean

Theorem [Mean and Variance of Sample Mean]: Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Let $\overline{X}_n$ be the sample mean. Then

$$E(\overline{X}_n) = \mu \text{ and } \mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Proof.

It follows from Theorem 4.2.1 and 4.2.4 that

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

Furthermore, since $X_1, \ldots, X_n$ are independent, Theorem 4.3.4 and 4.3.5 say that

$$\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

□

Example – Determining the Required Number of Observations

Suppose that a random sample is to be taken from a distribution for which the value of the mean μ is not known, but for which it is known that the standard deviation σ is 2 units or less. We shall determine how large the sample size must be in order to make the probability at least 0.99 that $|\overline{X}_n - \mu|$ will be less than 1 unit.

Example – Determining the Required Number of Observations

Suppose that a random sample is to be taken from a distribution for which the value of the mean μ is not known, but for which it is known that the standard deviation σ is 2 units or less. We shall determine how large the sample size must be in order to make the probability at least 0.99 that $|\overline{X}_n - \mu|$ will be less than 1 unit.

Solution: Since $\sigma^2 \leq 2^2 = 4$, it follows from the Chebyshev inequality that

$$P(|\overline{X}_n - \mu| \geq 1) \leq \frac{\frac{\sigma^2}{n}}{1} = \frac{4}{n}.$$

Since n must be chosen so that $P(|\overline{X}_n - \mu| < 1) \geq 0.99$, it follows that n must be chosen so that $4/n \leq 0.01$. Hence, it is required that $n \geq 400$.

Convergence of Random Variables

Convergence in Probability

Definition [Convergence in Probability]: A sequence $Z_1, Z_2, \ldots$ of random variables is said to *converge to b in probability* if for every number $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|Z_n - b| < \varepsilon) = 1.$$

This property is denoted by

$$Z_n \xrightarrow{P} b.$$

In other words, Z_n converges to b in probability if the probability that Z_n lies in each given interval around b, no matter how small this interval may be, approaches 1 as $n \to \infty$.

Convergence in Probability Is Invariant under Continuous Functions

Theorem: If $Z_n \xrightarrow{P} b$, and if $g(z)$ is a function continuous at $z = b$, then $g(Z_n) \xrightarrow{P} g(b)$.

Proof.

For every $\varepsilon > 0$, there is δ such that $|z - b| < \delta \Rightarrow |g(z) - g(b)| < \varepsilon$, and hence for every n,

$$P(|Z_n - b| < \delta) \leq P(|g(Z_n) - g(b)| < \varepsilon).$$

Since $Z_n \xrightarrow{P} b$, we have

$$\lim_{n\to\infty} P(|Z_n - b| < \delta) = 1,$$

and hence

$$\lim_{n\to\infty} P(|g(Z_n) - g(b)| < \varepsilon) = 1.$$

$\square$

In fact, this theorem can be extended to any finite number of variables. If $Z_n^{(1)} \xrightarrow{P} b^{(1)}, \ldots, Z_n^{(k)} \xrightarrow{P} b^{(k)}$, and if $g(z^{(1)}, \ldots, z^{(k)})$ is a function continuous at $(b^{(1)}, \ldots, b^{(k)})$, then $g(Z_n^{(1)}, \ldots, Z_n^{(k)}) \xrightarrow{P} g(b^{(1)}, \ldots, b^{(k)})$.

Almost Sure (or Probability 1) Convergence

Definition [Almost Sure Convergence]: A sequence $Z_1, Z_2, \ldots$ of random variables is said to *converge almost surely (or with probability 1) to b* if

$$P(\lim_{n\to\infty} Z_n = b) = 1.$$

This property is denoted by

$$Z_n \xrightarrow{a.s.} b.$$

Note: Again, this convergence property is invariant under continuous functions.

Convergence in Distribution

Definition [Convergence in Distribution]: A sequence $Z_1, Z_2, \ldots$ of random variables is said to *converge in distribution* to a c.d.f. F if for every $x \in \mathbb{R}$ at which F is continuous,

$$\lim_{n\to\infty} P(Z_n \le x) = F(x).$$

This property is denoted by

$$Z_n \xrightarrow{D} F.$$

Relationship between Types of Convergence

These statements are given without proof:

1. Almost sure convergence implies convergence in probability, but the converse is not true.

2. Convergence in probability to b is equivalent to convergence in distribution to the c.d.f.

$$F_{\text{single\_point}(b)}(x) = \begin{cases} 1 & \text{for } x \geq b \\ 0 & \text{otherwise.} \end{cases}$$

The Law of Large Numbers

The Weak Law

Theorem [Weak Law of Large Numbers]: Suppose that $X_1, X_2, \ldots$ form an infinite-size random sample from a distribution for which the mean is μ. Let $\overline{X}_n$ denote the sample mean of $X_1, \ldots, X_n$. Then

$$\overline{X}_n \xrightarrow{P} \mu.$$

Partial proof for the finite variance case

Let the variance of the distribution be σ^2. It then follows from the Chebyshev inequality that for every number $\varepsilon > 0$,

$$P(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Hence,

$$P(|\overline{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

Therefore

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1,$$

which means $\overline{X}_n \xrightarrow{P} \mu$. $\square$

The Strong Law

Theorem [Strong Law of Large Numbers]: Suppose that $X_1, X_2, \ldots$ form an infinite-size random sample from a distribution for which the mean is μ. Let $\overline{X}_n$ denote the sample mean of $X_1, \ldots, X_n$. Then

$$\overline{X}_n \xrightarrow{a.s.} \mu.$$

Remark

- This strong law implies the weak law, since almost sure convergence implies convergence in probability.
- Proof of this theorem involves characteristic functions, so it is omitted.
- It still does not require finite variance.

The Central Limit Theorem

The Central Limit Theorem

Theorem [Central Limit Theorem]: If the random variables $X_1, X_2, \ldots$ form an infinite-size random sample from a distribution with mean μ and variance σ^2 $(0 < \sigma^2 < \infty)$, then $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ converges in distribution to the standard normal distribution, i.e., for each fixed number x,

$$\lim_{n\to\infty} P\left[\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le x\right] = \Phi(x)$$

Remark

If a large random sample is taken from any distribution with mean μ and variance σ^2, regardless of whether this distribution is discrete or continuous, then the distribution of the random variable $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ will be approximately the standard normal distribution. Therefore, the distribution of $\overline{X}_n$ will be approximately the normal distribution with mean μ and variance σ^2/n, or, equivalently, the distribution of the $\sum_{i=1}^{n} X_i$ will be approximately the normal distribution with mean $n\mu$ and variance $n\sigma^2$.

Approximating sample distribution using LLN or CLT

RV	Approx. Distr. by LLN	Approx. Distr. by CLT
$\sum_{i=1}^{n} X_i$	single_point$(n\mu)$	normal$(n\mu, n\sigma^2)$
$\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}$	single_point$(\sqrt{n}\mu)$	normal$(\sqrt{n}\mu, \sigma^2)$
$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$	single_point(μ)	normal$(\mu, \frac{\sigma^2}{n})$

From the above table it can be seen: In terms of approximating the distribution of a random variable based on a random sample, it is more accurate to use the central limit theorem than to use the law of large numbers, provided that the variance of the distribution is finite and known.

Approximation Using Central Limit Theorem – Example

Sampling from a Uniform Distribution. Suppose that a random sample of size $n = 12$ is taken from the uniform distribution on the interval $[0, 1]$. Approximate the value of $P(|\overline{X} - \frac{1}{2}| \leq 0.1)$.

Approximation Using Central Limit Theorem – Example

Sampling from a Uniform Distribution. Suppose that a random sample of size $n = 12$ is taken from the uniform distribution on the interval $[0, 1]$. Approximate the value of $P(|\overline{X} - \frac{1}{2}| \leq 0.1)$.

Solution: $E(X_i) = \frac{1}{2}$ and $\mathrm{Var}(X_i) = \frac{1}{12}$, hence $E(\overline{X}) = \frac{1}{2}$ and $\mathrm{Var}(\overline{X}) = \frac{1}{144}$. According to the central limit theorem, the RV $\overline{X}$ approximately $\sim$ normal$(\frac{1}{2}, \frac{1}{144})$. Hence,

$$P(|\overline{X} - \frac{1}{2}| \leq 0.1) = P(\frac{|\overline{X} - \frac{1}{2}|}{1/12} \leq 1.2)$$

$$\approx 2\Phi(1.2) - 1 = 0.7698.$$

Correction for Continuity in Using CLT

Some applications of the central limit theorem allow us to approximate the probability that an integer-valued discrete random variable X lies in an integer-ended interval $[a, b]$ by the probability that a normal random variable lies in that interval.

Correction for Continuity:

$$P(a \leq X \leq b) = P(a - 0.5 < X < b + 0.5).$$

Correction for Continuity in Using CLT – Example

Coin Tossing. A fair coin is tossed 20 times and all tosses are independent. What is the probability of obtaining exactly 10 heads?

Solution Method 1: Let X be the number of heads. Then $X \sim \text{binomial}(n = 20, p = \frac{1}{2})$.

$$P(X = 10) = \binom{20}{10}(\frac{1}{2})^{10}(1 - \frac{1}{2})^{20-10} = 0.1762.$$

Correction for Continuity in Using CLT – Example

Coin Tossing. A fair coin is tossed 20 times and all tosses are independent. What is the probability of obtaining exactly 10 heads?

Solution Method 1: Let X be the number of heads. Then $X \sim \text{binomial}(n = 20, p = \frac{1}{2})$.

$$P(X = 10) = \binom{20}{10} (\frac{1}{2})^{10} (1 - \frac{1}{2})^{20-10} = 0.1762.$$

Solution Method 2: $E(X) = np = 10$, $\text{Var}(X) = np(1-p) = 5$. Thus the distribution of X can be approximated by normal$(10, 5)$.

$$\begin{aligned} P(X = 10) &= P(9.5 < X < 10.5) \\ &= P(\frac{9.5 - 10}{\sqrt{5}} < \frac{X - 10}{\sqrt{5}} < \frac{10.5 - 10}{\sqrt{5}}) \\ &\approx \Phi(0.2236) - \Phi(-0.2236) = 0.177. \end{aligned}$$

Homework for Chapter 6

- P359: 6
- P370: 4, 10
- P374: 2
- P375: 1