R403: Probabilistic and Statistical Computations with R

Topic 5: Limiting Distributions. Stochastic Convergence

Kaloyan Ganev

2022/2023

Lecture Contents

Introduction

Introduction

- A quick and rather informal introduction offered
- Some applications in R will also be shown

Some Important Inequalities

Some Important Inequalities

- Assume that the distribution of the random variable X is unknown but its mean (and possibly its variance) is known
- The following two inequalities allow to determine bounds to probabilities in such cases[1]
- **Markov's inequality:** If $X \geq 0$, then for any $a > 0$

$$P(X \geq a) \leq \frac{\mathsf{E}(X)}{a}$$

Interpretation: The probability of observing a value of X that is far away from the mean is getting smaller with a increasing

[1] In case the distribution is known, the probabilities can be computed exactly.

Some Important Inequalities (2)

- **Chebyshev's inequality:** If X has mean μ and variance σ^2, then for any $k > 0$

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

  Interpretation: The probability of observing an absolute deviation of a random variable from its mean greater than or equal to k standard deviations from the mean is bounded by $1/k^2$

- Implication: The probability of being within k standard deviations is greater than or equal to $1 - \frac{1}{k^2}$, i.e.

$$P(|X - \mu| < k\sigma) \geq \frac{1}{k^2}$$

- Note: there is no requirement that $X \geq 0$

Modes of Stochastic Convergence

The Weak Law of Large Numbers

- Main idea: the sample mean of a large number of i.i.d. random variables is very close to the theoretical mean with high probability
- Let $X_1, X_2, \ldots$ be a sequence of $i.i.d.(\mu, \sigma^2)$ random variables
- Take a sample of size n; denote the sample mean by $\overline{X}$, i.e.

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

- Take expectations of both sides:

$$\mathsf{E}(\overline{X}) = \frac{\mathsf{E}(X_1) + \mathsf{E}(X_2) + \ldots + \mathsf{E}(X_n)}{n} = \frac{\mu + \mu + \ldots + \mu}{n} = \mu$$

The Weak Law of Large Numbers (2)

- Calculate the variance of both sides:

$$\mathsf{Var}(\overline{X}) = \frac{\mathsf{Var}(X_1) + \mathsf{Var}(X_2) + \ldots + \mathsf{Var}(X_n)}{n^2} = \frac{\sigma^2 + \sigma^2 + \ldots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

- Use the Chebyshev inequality (which can also be written as[2] $P(|X - \mu| \geq k) \leq \dfrac{\sigma^2}{k^2}$):

$$P(|\,\mathsf{E}(X) - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

- This is valid for each $\varepsilon > 0$
- The RHS goes to 0 as $n \to \infty$

[2] Note that σ^2 here denotes the variance of X, and not that of $\overline{X}$

The Weak Law of Large Numbers (3)

- This leads to the following theorem:[3]

Theorem 1 (Weak Law of Large Numbers)

If $X_1, X_2, \ldots,$ are i.i.d. random variables with mean μ,

$$\lim_{n\to\infty} P(|\,\mathsf{E}(X) - \mu| \geq \varepsilon) = 0, \quad \forall \varepsilon > 0$$

- We will illustrate that result in R
- We draw 100000 random numbers from a normal distribution[4] with $\mu = 5$ and $\sigma^2 = 9$
- Look at the script named `law_of_large_numbers.R`

[3] Known also as the Khinchin Theorem.

[4] Distribution and parameters are picked arbitrarily; play with the code by changing them.

Convergence in Probability

- The type of convergence that is seen in the Law of Large Numbers is not however the same as convergence of deterministic sequences
- The difference is that the sequence is one of random variables here
- To make things precise, we first recall what deterministic convergence is

Definition 1 (Convergence of a deterministic sequence)

The sequence of real numbers $a_1, a_2, \ldots$ converges to the real number a if $\forall \varepsilon > 0\ \exists n_0 : |a_n - a| \leq \varepsilon,\ \forall n \geq n_0$. We denote this by

$$\lim_{n\to\infty} a_n = a$$

Convergence in Probability (2)

- Given the preceding definition, we can state the definition of convergence in probability

Definition 2 (Convergence in probability)

If $X_1, X_2, \ldots$ is a sequence of random variables and a is a real number, the sequence X_n converges in probability to a if

$$\lim_{n\to\infty} P(|X_n - a| \geq \varepsilon) = 0, \quad \forall \varepsilon > 0$$

- Note that we do not require that the random variables are independent
- The Weak Law of Large Numbers uses in fact the notion of convergence in probability

Convergence in Probability (3)

- If the sequence of random variables $X_1, X_2, \dots$ converges in probability to a, then when n becomes large, almost all of the probability mass/density will be within ε from a
- This allows to rewrite the definition of convergence in probability as follows: $\forall \varepsilon > 0,\ \forall \delta > 0$ there exists some n_0 such that

$$P(|X_n - a| \geq \varepsilon) \leq \delta, \quad \forall n \geq n_0$$

- ε here is known as *accuracy level*, while δ is known as *confidence level*

Convergence in Distribution

Definition 3 (Convergence in Distribution)

Let $X_1, X_2, \ldots$ be a sequence of random variables with cdfs $F_{X_1}, F_{X_2}, \ldots$, and also let X be a random variable with cdf F_X. Denote by $C(F_X)$ the set of values of X for which F_X is continuous. Then the sequence X_n converges in distribution to X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x), \quad \forall x \in C(F_X)$$

- Notation:

$$X_n \xrightarrow{d} X$$

- The distribution of X is called the *asymptotic* or *limiting* distribution of the sequence $\{X_n\}$

Central Limit Theorem

- Start from the Khinchin Theorem
- It states that with the increase of n, the distribution of $\overline{X}$ gets more and more concentrated around the true mean μ
- That also means its variance (σ^2/n) converges to 0
- Take now the sum

$$S_n = X_1 + \ldots + X_n = n\,\mathsf{E}(X) = n\overline{X}$$

- The variance of this sum is

$$\mathsf{Var}(S_n) = \mathsf{Var}(n\overline{X}) = n^2\,\mathsf{Var}(\overline{X}) = n\sigma^2$$

- When $n \to \infty$, $n\sigma^2 \to \infty$

Central Limit Theorem (2)

- Consider a case between those two extremes: the deviation of S_n from its mean,[5] i.e. $S_n - n\mu$
- Let $X_1, X_2, \ldots$ be a sequence of $i.i.d.(\mu, \sigma^2)$ random variables
- Define

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

- The mathematical expectation of Z_n equals

$$\mathsf{E}(Z_n) = \mathsf{E}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) = 0,$$

and its variance correspondingly is

$$\mathsf{Var}(Z_n) = \mathsf{Var}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) = 1$$

[5] The mean is $\mathsf{E}(S_n) = \mathsf{E}(X_1) + \ldots + \mathsf{E}(X_n) = \mu + \ldots + \mu = n\mu$.

Central Limit Theorem (3)

Theorem 2 (Central Limit Theorem)

Let $X_1, X_2, \ldots$ *be a sequence of* $i.i.d.(\mu, \sigma^2)$ *random variables. Define*

$$Z_n = \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

Then the cdf of Z_n *converges to the standard normal cdf, i.e.*

$$\lim_{n\to\infty} P(Z_n \leq z) = \mathbf{\Phi}(z), \quad \forall z$$

where

$$\mathbf{\Phi}(z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} e^{-\frac{x^2}{2}}\, dx$$

The Strong Law of Large Numbers

- Also relates to the convergence of the sample mean to the true population one
- However, the type of convergence that it uses is different

Definition 4 (Strong Law of Large Numbers)

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. If $\mathrm{E}(X_i) = \mu$, then the sequence of sample means $M_n = \dfrac{X_1 + \ldots + X_n}{n}$ converges to μ with probability 1, i.e.

$$P\left(\lim_{n\to\infty} \frac{X_1 + \ldots + X_n}{n} = \mu\right) = 1$$

The Strong Law of Large Numbers (2)

- To get insight into the definition, consider the sample space $\mathcal{S}$ consisting of infinite sequences of real numbers $(x_1, x_2, \ldots)$
- In other words, the sequence of random variables $X_1, X_2, \ldots$ will produce one realization, i.e. one such sequence of numbers
- Consider $A \subset \mathcal{S}$ consisting of only the sequences $(x_1, x_2, \ldots)$ which have an average μ when $n \to \infty$

$$(x_1, x_2, \ldots) \in A \Leftrightarrow \lim_{n \to \infty} \frac{x_1 + \ldots + x_n}{n} = \mu$$

- The Strong Law states that all of the probability mass is in A which is not guaranteed for the Weak Law

Almost Sure Convergence

Definition 5

If $X_1, X_2, \ldots$ is a sequence of random variables and c is a real number, X_n converges almost surely (i.e. with probability 1) given that

$$P\left(\lim_{n\to\infty} X_n = c\right) = 1$$

- Note: There is no independence requirement concerning random variables
- The Strong Law uses the notion of almost sure convergence
- Notation

$$X_n \overset{a.s.}{\to} c$$

- Important: Almost sure convergence $\Rightarrow$ Convergence in probability $\Rightarrow$ Convergence in distribution

References

- Bertsekas, D., and J. Tsitsiklis (2008): *Introduction to Probability*, Athena Scientific, 2nd ed.
- Hogg, R., J. McKean, and A. Craig (2018): *Introduction to Mathematical Statistics*, Pearson, 8th ed.
- https://www.analyticsvidhya.com/blog/2019/05/statistics-101-introduction-central-limit-theorem/
- https://statistical-engineering.com/clt-summary/