SDS7102: Linear Models and Extensions

Simple Asymptotics

Qiang Sun, Ph.D. <qiang.sun@mbzuai.ac.ae>

These slides are due to Eric Moulines.

August 19, 2025

MBZUAI

Introduction

- It is often necessary to consider the distribution of a random variable that is itself a function of several random variables, for example, $Y = g(X_1, \cdots, X_n)$; a simple example is the sample mean of random variables $X_1, \cdots, X_n$.
- Unfortunately, finding the distribution exactly is often very difficult or very time-consuming even if the joint distribution of the random variables is known exactly. In other cases, we may have only partial information about the joint distribution of $X_1, \cdots, X_n$ in which case it is impossible to determine the distribution of Y.
- However, when n is large, it may be possible to obtain approximations to the distribution of Y even when only partial information about $X_1, \cdots, X_n$ is available; in many cases, these approximations can be remarkably accurate.

Introduction

- Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with mean μ and variance σ^2 and define

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

  to be their sample mean; we would like to look at the behaviour of the distribution of $\bar{X}_n$ when n is large.

- First of all, it seems reasonable that $\bar{X}_n$ will be close to μ if n is sufficiently large; that is, the random variable $\bar{X}_n - \mu$ should have a distribution that, for large n, is concentrated around 0 or, more precisely,

$$P\left[\left|\bar{X}_n - \mu\right| \leq \epsilon\right] \approx 1,$$

  when ϵ is small. (Note that $\mathrm{Var}\left(\bar{X}_n\right) = \sigma^2/n \to 0$ as $n \to \infty$.)

Chebyshev's inequality

Theorem

Suppose that X is a random variable with $E\left(X^2\right) < \infty$. Then for any $\epsilon > 0$,

$$P[|X| > \epsilon] \leq \frac{E\left(X^2\right)}{\epsilon^2}.$$

Introduction

- It is also possible to look at the difference between $\bar{X}_n$ and μ on a "magnified" scale; we do this by multiplying the difference $\bar{X}_n - \mu$ by $\sqrt{n}$ so that the mean and variance are constant.
- Thus define
$$Z_n = \sqrt{n}\left(\bar{X}_n - \mu\right)$$
and note that $E\left(Z_n\right) = 0$ and $\mathrm{Var}\left(Z_n\right) = \sigma^2$.
- We can now consider the behaviour of the distribution function of Z_n as n increases. If this sequence of distribution functions has a limit (in some sense) then we can use the limiting distribution function to approximate the distribution function of Z_n (and hence of $\bar{X}_n$).

Introduction

For example, if we have

$$P\left(Z_n \leq x\right) = P\left(\sqrt{n}\left(\bar{X}_n - \mu\right) \leq x\right) \approx F_0(x)$$

then

$$\begin{aligned} P\left(\bar{X}_n \leq y\right) &= P\left(\sqrt{n}\left(\bar{X}_n - \mu\right) \leq \sqrt{n}(y - \mu)\right) \\ &\approx F_0(\sqrt{n}(y - \mu)) \end{aligned}$$

provided that n is sufficiently large to make the approximation valid.

Convergence in probability

Definition

Let $\{X_n\}, X$ be random variables. Then $\{X_n\}$ converges in probability to X as $n \to \infty$ $(X_n \to_p X)$ if for each $\epsilon > 0$,

$$\lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0.$$

Convergence in distribution

Definition

Let $\{X_n\}, X$ be random variables. Then $\{X_n\}$ converges in distribution to X as $n \to \infty \, (X_n \to_d X)$ if

$$\lim_{n\to\infty} P\left(X_n \leq x\right) = P(X \leq x) = F(x).$$

for each continuity point of the cumulative distribution function F.

Note that the number of discontinuity points of the function F is at most countable.

Convergence in distribution

- It is important to remember that $X_n \to_d X$ implies convergence of distribution functions and not of the random variables themselves.
- For this reason, it is often convenient to replace $X_n \to_d X$ by $X_n \to_d F$ where F is the distribution function of X, that is, the limiting distribution; for example, $X_n \to {}_d N\left(0, \sigma^2\right)$ means that $\{X_n\}$ converges in distribution to a random variable that has a Normal distribution (with mean 0 and variance σ^2).

Convergence in distribution

- If $X_n \to_d X$ then for sufficiently large n we can approximate the distribution function of X_n by that of X; thus, convergence in distribution is potentially useful for approximating the distribution function of a random variable.
- However, the statement $X_n \to_d X$ does not say how large n must be in order for the approximation to be practically useful. To answer this question, we typically need a further result dealing explicitly with the approximation error as a function of n.

Maximum of uniform random variables

Suppose that $X_1, \cdots, X_n$ are i.i.d. Uniform random variables on the interval $[0, 1]$ and define

$$M_n = \max(X_1, \cdots, X_n)$$

- Show that $M_n \to_p 1$.
- Find the limiting distribution of $n(1 - M_n)$.

Decimal representation

Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with

$$P(X_i = j) = \frac{1}{10} \quad \text{for } j = 0, 1, 2, \cdots, 9$$

and define

$$U_n = \sum_{k=1}^{n} \frac{X_k}{10^k}$$

- Find the limiting distribution of U_n.

Links between convergence in probability and in distribution

Theorem

Let $\{X_n\}, X$ *be random variables.*

1. *If* $X_n \to_p X$ *then* $X_n \to_d X$*.*
2. *If* $X_n \to_d \theta$ *(a constant) then* $X_n \to_p \theta$*.*

Continuous Mapping Theorem

Theorem

Suppose that $g(x)$ is a continuous real-valued function.

1. *If $X_n \rightarrow_p X$ then $g\left(X_n\right) \rightarrow_p g(X)$.*
2. *If $X_n \rightarrow_d X$ then $g\left(X_n\right) \rightarrow_d g(X)$.*

The assumption of continuity can also be relaxed somewhat. For example, Theorem 3.2 will hold if g has a finite or countable number of discontinuities provided that these discontinuity points are continuity points of the distribution function of X. For example, if $X_n \rightarrow_d \theta$ (a constant) and $g(x)$ is continuous at $x=\theta$ then $g\left(X_n\right) \rightarrow_d g(\theta)$.

Slutsky's Theorem

Theorem

Suppose that $X_n \to_d X$ *and* $Y_n \to_p \theta$ *(a constant). Then*

1. $X_n + Y_n \to_d X + \theta$.
2. $X_n Y_n \to_d \theta X$.

Delta Method

Theorem

Suppose that

$$a_n (X_n - \theta) \to_d Z$$

where θ *is a constant and* $\{a_n\}$ *is a sequence of constants with* $a_n \uparrow \infty$*. If* $g(x)$ *is a function with derivative* $g'(\theta)$ *at* $x = \theta$ *then*

$$a_n (g(X_n) - g(\theta)) \to_d g'(\theta) Z.$$

Convergence of moments

- If $X_n \rightarrow {}_d X$ (or $X_n \rightarrow {}_p X$), it is tempting to say that $E\left(X_n\right) \rightarrow E(X)$; however, this statement is not true in general.
- For example, suppose that $P\left(X_n=0\right)=1-n^{-1}$ and $P\left(X_n=n\right)=n^{-1}$. Then $X_n \rightarrow_p 0$ but $E\left(X_n\right)=1$ for all n (and so converges to 1).
- To ensure convergence of moments, additional conditions are needed; these conditions effectively bound the amount of probability mass in the distribution of X_n concentrated near $\pm\infty$ for large n.

Convergence of moments

Theorem

If $X_n \to_d X$ *and* $|X_n| \leq M$ *(finite) then* $E(X)$ *exists and* $E(X_n) \to E(X)$.

Weak Law of Large Numbers

Theorem

Suppose that $X_1, X_2, \cdots$ *are i.i.d. random variables with* $E(X_i) = \mu$ *where* $E(|X_i|) < \infty$). *Then*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to_p \mu$$

as $n \to \infty$.

Convergence of the sample median

- Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with a distribution function $F(x)$. Assume that the X_i 's have a unique median $\mu(F(\mu) = 1/2)$; in particular, this implies that for any $\epsilon > 0, F(\mu + \epsilon) > 1/2$ and $F(\mu - \epsilon) < 1/2$.
- Let $X_{(1)}, \cdots, X_{(n)}$ be the order statistics of the X_i's and define $Z_n = X_{(m_n)}$ where $\{m_n\}$ is a sequence of positive integers with $m_n/n \to 1/2$ as $n \to \infty$. For example, we could take $m_n = n/2$ if n is even and $m_n = (n+1)/2$ if n is odd; in this case, Z_n is essentially the sample median of the X_i's.
- Show that $Z_n \to_p \mu$ as $n \to \infty$.