SDS7102: Linear Models and Extensions

Confidence Intervals and More on Convergence

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

September 22, 2025

MBZUAI

Outline

In this lecture, we will cover

- Confidence Intervals;
- More on Convergence of Random Variables;
- and o_p and O_p notations.

Confidence Intervals

Confidence Intervals

- Suppose $X_1, \ldots, X_n$ are random variables with some joint distribution depending on a parameter θ that may be real- or vector-valued.
- An example is the same average $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ for estimating the mean μ, where $X_1, \ldots, X_n$ are i.i.d. random variables with mean μ and variance σ^2.
- It is often of interest to find interval estimators for θ that are likely to contain the true value of θ.

Confidence Intervals

Definition (Confidence Intervals)

A p confidence interval for a parameter θ is an interval $(L(X_1, \ldots, X_n), U(X_1, \ldots, X_n))$ such that

$$P_\theta(L(X_1, \ldots, X_n) \le \theta \le U(X_1, \ldots, X_n)) = p$$

for all θ in the parameter space.

The number p is called the confidence level of the interval or coverage probability (or simply coverage).

Lower and upper confidence bounds

- Similarly, we can define lower and upper confidence bounds.
- Suppose that
$$P_\theta[\theta \geq L(X)] = p$$
for some statistic $L(X)$ and for all θ; then $L(X)$ is called a p lower confidence bound for θ.
- Likewise, if
$$P_\theta[\theta \leq U(X)] = p$$
for some statistic $U(X)$ and for all θ; then $U(X)$ is called a p upper confidence bound for θ.

Some misconceptions about confidence intervals

- The interpretation ofconfidence intervals is frequently misunderstood.
- Much of the confusion stems from the fact that confidence intervals are defined in terms ofthe distribution of $\boldsymbol{X} = (X_1, \ldots, X_n)$ but, in practice, are stated in terms ofthe observed values ofthese random variables leaving the impression that a probability statement is being made about θ rather than about the random interval.
- However, given data $\boldsymbol{X} = \boldsymbol{x}$, the interval $[L(x), U(x)]$ will either contain the true value of θ or not contain the true value of θ; under repeated sampling, p ofthese intervals will contain the true value of θ.

Exact vs approximate confidence intervals

- An exact confidence interval is derived from the exact distribution of the data.
- In many problems, it is difficult or impossible to find an exact confidence interval.
- In such cases, we often resort to approximate confidence intervals that are valid in large samples:

$$P_\theta(L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})) \approx p.$$

An example of exact and approxiamte confidence intervals

- Suppose that $X_1, \ldots, X_n$ are i.i.d. Normal random variables with mean μ and variance 1. Then $\sqrt{n}(\bar{X} - \mu) \sim N(0,1)$ and so

  $$P_\mu[-1.96 \leq \sqrt{n}(\bar{X} - \mu) \leq 1.96] = 0.95,$$

  which is the same as

  $$P_\mu[\bar{X} - 1.96/\sqrt{n} \leq \mu \leq \bar{X} + 1.96/\sqrt{n}] = 0.95.$$

- If we assume only that $X_1, \ldots, X_n$ are i.i.d. with mean μ and variance 1 (not necessarily normally distributed), we have (by the CLT),

  $$P_\mu[-1.96 \leq \sqrt{n}(\bar{X} - \mu) \leq 1.96] \approx 0.95,$$

  if n is sufficiently large. Thus, using the same argument as above, the interval with endpoints $\bar{X} \pm 1.96/\sqrt{n}$ is an approximate 95% confidence interval for μ.

More on Convergence of Random Variables

Definitions

Definition (Convergence in Distribution)

Let X_n and X be random variables. We say that X_n converges in distribution to X, denoted $X_n \xrightarrow{d} X$, if

$$F_{X_n}(x) \to F_X(x)$$

for all points x at which F_X is continuous, where F_{X_n} and F_X are the cumulative distribution functions of X_n and X, respectively.

Definition (Convergence in Probability)

Let X_n and X be random variables. We say that X_n converges in probability to X, denoted $X_n \xrightarrow{p} X$, if for every $\epsilon > 0$,

$$P(\|X_n - X\| > \epsilon) \to 0.$$

Definitions

Definition (Convergence almost surely)

Let X_n and X be random variables. We say that X_n converges almost surely to X, denoted $X_n \stackrel{a.s.}{\to} X$, if

$$P(\lim_{n\to\infty} X_n = X) = 1.$$

The three notions of convergence correspond to central limit theroem, weak law of large numbers, and strong law of large numbers, respectively.

The strong law of large numbers appears to be of less interest in statistics. Usually the weak law of large numbers, according to which $\bar{X} := \sum_{i=1}^{n} X_i \stackrel{p}{\to} \mathbb{E}X_1$, suffices.

Portmanteau Lemma

Lemma (Portmanteau Lemma)

For any random vectors X_n and X the following statements are equivalent.

1. $P(X_n \leq x) \to P(X \leq x)$ *for all continuity points of* $x \mapsto P(X \leq x)$;
2. $E[f(X_n)] \to E[f(X)]$ *for all bounded, continuous functions* f;
3. $E[f(X_n)] \to E[f(X)]$ *for all bounded, Lipschitz functions* f;
4. $\liminf E[f(X_n)] \geq E[f(X)]$ *for all nonnegative, continuous functions* f;
5. $\liminf P(X_n \in G) \geq P(X \in G)$ *for every open set* G;
6. $\limsup P(X_n \in F) \leq P(X \in F)$ *for every closed set* F;
7. $P(X_n \in B) \to P(X \in B)$ *for all Borel sets* B *with* $P(X \in \partial B) = 0$, *where* $\partial B = \bar{B} - \mathring{B}$ *is the boundary of* B.

Proof of Portmanteau Lemma

See Lemma 2.2 by van der Vaart (2000).

Continuous Mapping Theorem

Theorem (Continuous Mapping Theorem)

Let $g : \mathbb{R}^k \to \mathbb{R}^m$ *be continuous at every point of a set* $C \subseteq \mathbb{R}^k$ *such that* $\mathbb{P}(X \in C) = 1$*. Then:*

1. *If* $X_n \xrightarrow{d} X$*, then* $g(X_n) \xrightarrow{d} g(X)$*.*
2. *If* $X_n \xrightarrow{p} X$*, then* $g(X_n) \xrightarrow{p} g(X)$*.*
3. *If* $X_n \xrightarrow{as} X$*, then* $g(X_n) \xrightarrow{p} g(X)$*.*

Proof of Continuous Mapping Theorem

See Theorem 2.3 by van der Vaart (2000).

Some results regarding the three notions of convergence

Theorem

Let X_n, X and Y_n be random vectors. Then

1. $X_n \overset{as}{\to} X$ *implies* $X_n \overset{p}{\to} X$*;*
2. $X_n \overset{p}{\to} X$ *implies* $X_n \overset{d}{\to} X$*;*
3. $X_n \overset{d}{\to} c$ *for a constant* c *if and only if* $X_n \overset{p}{\to} c$*;*
4. *If* $X_n \overset{d}{\to} X$ *and* $\|X_n - Y_n\| \overset{p}{\to} 0$*, then* $Y_n \overset{d}{\to} X$*;*
5. *If* $X_n \overset{d}{\to} X$ *and* $Y_n \overset{p}{\to} c$ *for a constant* c*, then* $(X_n, Y_n) \overset{d}{\to} (X, c)$*;*
6. *If* $X_n \overset{p}{\to} X$ *and* $Y_n \overset{p}{\to} Y$*, then* $(X_n, Y_n) \overset{p}{\to} (X, Y)$*.*

Poof

See Theorem 2.7 by van der Vaart (2000).

Slustky's Lemma

Lemma (Slustky's Lemma)

Let X_n and Y_n be random vectors. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ for a constant c, then

- $X_n + Y_n \xrightarrow{d} X + c$;
- $Y_n X_n \xrightarrow{d} cX$;
- $Y_n^{-1} X_n \xrightarrow{d} ac^{-1} X$.

Proof

See Lemma 2.8 by van der Vaart (2000).

Example 1: t-statistics

Example

Let $Y_1, Y_2, \ldots$ be independent, identically distributed random variables with $\mathsf{E}Y_1 = 0$ and $\mathsf{E}Y_1^2 < \infty$. Then the *t*-statistic $\sqrt{n}\bar{Y}_n/S_n$, where $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2$ is the sample variance, is asymptotically standard normal.

To see this, first note that by two applications of the weak law of large numbers and the continuous-mapping theorem for convergence in probability

$$S_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n Y_i^2 - \bar{Y}_n^2\right) \xrightarrow{p} \mathsf{E}Y_1^2 - (\mathsf{E}Y_1)^2 = \operatorname{var} Y_1.$$

Again by the continuous-mapping theorem, S_n converges in probability to $\operatorname{sd} Y_1$. By the central limit theorem $\sqrt{n}\bar{Y}_n$ converges in law to the $N(0, \operatorname{var} Y_1)$ distribution. Finally, Slutsky's lemma gives that the sequence of *t*-statistics converges in distribution to $N(0, \operatorname{var} Y_1)/\operatorname{sd} Y_1 = N(0,1)$. □

Example 2: Confidence intervals

Example

Let T_n and S_n be sequences of estimators satisfying

$$\sqrt{n}(T_n - \theta) \rightsquigarrow N(0, \sigma^2), \qquad S_n \xrightarrow{p} \sigma^2,$$

for certain parameters θ and σ^2 depending on the underlying distribution, for every distribution in the model. Then $\theta = T_n \pm S_n/\sqrt{n} z_\alpha$ is a confidence interval for θ of asymptotic level $1 - 2\alpha$. More precisely, we have that the probability that θ is contained in $[T_n - S_n/\sqrt{n} z_\alpha, T_n + S_n/\sqrt{n} z_\alpha]$ converges to $1 - 2\alpha$. This is a consequence of the fact that the sequence $\sqrt{n}(T_n - \theta)/S_n$ is asymptotically standard normally distributed. $\square$

Example 3: Applications of Weak Law of Large Numbers

Example

- 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 $n \to \infty$.

The o_p and O_p Notations

Bounded in probability

Definition (Tightness)

A random variable X is called tight if for every $\epsilon > 0$ there exists a constant M such that $P(\|X\| > M) < \epsilon$.

Definition (Bounded in probability)

A sequence of random vectors $\{X_n : n \geq 1\}$ is called bounded in probability or uniformly tight if M can be chosen the same for every X_n: For every $\epsilon > 0$ there exists a constant M such that $\sup P(\|X_n\| > M) < \epsilon$.

The o_p and O_p Notations

- The notation $o_p(1)$ (”small oh-P-one”) is short for a sequence of random vectors that converges to zero in probability.
- The notation $O_p(1)$ (”big oh-P-one”) is short for a sequence of random vectors that is bounded in probability.
- More generally, for a given sequence of random variables R_n,
  - $R_n = o_p(R_n)$ means $X_n = Y_n R_n$ and $Y_n \xrightarrow{p} 0$;
  - $R_n = O_p(R_n)$ means $X_n = Y_n R_n$ and $Y_n = O_p(1)$.

Rules for Calculation

There are many rules of calculus with o and O symbols, which we apply without comment. For instance,

$$\begin{aligned}
o_P(1) + o_P(1) &= o_P(1) \\
o_P(1) + O_P(1) &= O_P(1) \\
O_P(1)o_P(1) &= o_P(1) \\
(1 + o_P(1))^{-1} &= O_P(1) \\
o_P(R_n) &= R_n o_P(1) \\
O_P(R_n) &= R_n O_P(1) \\
o_P(O_P(1)) &= o_P(1).
\end{aligned}$$

Two more rules

Lemma

Let R be a function defined on domain in $\mathbb{R}^k$ such that $R(0) = 0$. Let X_n be a sequence of random vectors with values in the domain of $\mathbb{R}$ that converges in probability to zero. Then, for every $p > 0$,

1. *if $R(h) = o(\|h\|^p)$ as $h \to 0$, then $R(X_n) = o_P(\|X_n\|^p)$;*
2. *if $R(h) = O(\|h\|^p)$ as $h \to 0$, then $R(X_n) = O_P(\|X_n\|^p)$.*

See Lemma 2.12 by van der Vaart (2000).