Week 9

Statistics 251

Zhongjian Wang

Department of Statistics
The University of Chicago

Where are we?

Central Limit Theorem

Proof of CLT

Examples

Strong Law of Large Numbers

Proof of (s)LLN

Examples

Revision

Recall the weak law of large numbers,
Let $X_1, X_2, \ldots$ be a sequence of thdependent and Identically distributed random vanables, each having finite mean $E[X_i] = \mu$. Then, for any $\varepsilon > 0$,

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

In some sense means, $\left|\frac{X_1+\cdots+X_n-n\mu}{n}\right| \to 0$. What about

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

or in other words, how fast does it converge?

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

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

Where are we?

Central Limit Theorem

Proof of CLT

Examples

Strong Law of Large Numbers

Proof of (s)LLN

Examples

Proof:

We first assume mean $\mu = 0$, variance $\sigma^2 = 1$. (Why we can do this?)
Let $M(t)$ be the moment generating function of X_i. Then the moment generating function of $\sum_{i=1}^{n} X_i/\sqrt{n}$ is given by $\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$.
Let $L(t) = \log M(t)$, then

$$\log\left(\left[M\left(\frac{t}{\sqrt{n}}\right)\right)\right]^n = nL(t/\sqrt{n})$$

Proof conti

What is the limit of $nL(t/\sqrt{n})$ when $n \to \infty$?

$$\begin{aligned}
\lim_{n\to\infty} \frac{L(t/\sqrt{n})}{n^{-1}} &= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} \text{ (by L'Hôpital's rule)} \\
&= \lim_{n\to\infty} \left[\frac{L'(t/\sqrt{n})t}{2n^{-1/2}}\right] \\
&= \lim_{n\to\infty} \left[\frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}}\right] \text{ (by L'Hôpital's rule)} \\
&= \lim_{n\to\infty} \left[L''\left(\frac{t}{\sqrt{n}}\right)\frac{t^2}{2}\right] \\
&= \frac{t^2}{2} \text{ (as } L''(0) = \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} = 1\text{)}
\end{aligned}$$

Proof conti

So as $n \to \infty$, $[M(t/\sqrt{n})]^n \to e^{t^2/2}$.

What distribution has generating function $e^{t^2/2}$?

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 the standard normal as $n \to \infty$.

Where are we?

Notes when applying CLT

- When can CLT be applied?
  CLT only applies when $N \to \infty$. Empirically, only when $N \geq 30$, you can have confidence to apply it. If $N < 10$, unless otherwise noted, we should use Markov inequalities.
- Z score
  When the random variable converges to normal distribution, in some case we need to know the exact value of integral $\phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}$.

a	-2.58	-1.96	-1.645	0	1.645	1.96	2.58
$\phi(a)$	0.005	0.025	0.05	0.5	0.95	0.975	0.995

Example

Let $X_i, i = 1, \ldots, 100$ be i.i.d. uniform random variables on $(0, 1)$. Approximate

$$\mathbb{P}\left(\sum_{i=1}^{100} X_i > 57.5\right).$$

Notice that $\mathbb{E}[X_i] = \frac{1}{2}$ and $\mathrm{Var}(X_i) = \frac{1}{12}$. This means

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{100} X_i > 57.5\right) &= \mathbb{P}\left(\frac{X_1 + \cdots + X_{100} - 50}{\sqrt{100 \times \frac{1}{12}}} > \frac{7.5}{\sqrt{100 \times \frac{1}{12}}}\right) \\
&= \mathbb{P}\left(\frac{X_1 + \cdots + X_{100} - 50}{\sqrt{100 \times \frac{1}{12}}} > 2.598\right) \\
&\approx 1 - \Phi(2.58) \approx 0.005.
\end{aligned}$$

Where are we?

Central Limit Theorem

Proof of CLT

Examples

Strong Law of Large Numbers

Proof of (s)LLN

Examples

Comparison with weak law of large numbers

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

weak for any $\epsilon > 0$,

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

strong with probability 1 ,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \to 0 \quad \text{as} \quad n \to \infty$$

The strong implies the weak.

Convergence of random variables

A random variable sequence $\{X_n\}$ converges to $X \in R$

- almost surely if
$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$
e.g. Strong law of large numbers
- in probabilties if
$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) = 0 \quad \text{for any } \epsilon > 0$$
e.g. Weak law of large numbers
- in distribution if
$$\lim_{n\to\infty} \mathbb{P}\left(X_n \in A\right) = \mathbb{P}(X \in A) \quad \text{for cointinuous set } A \text{ in } R$$
e.g. Central limit theorem

Where are we?

Proof

Although the theorem can be proven without this assumption, we will suppose that $\mathbb{E}\left[X_i^4\right] = K < \infty$.
Let $S_n = \sum_{i=1}^n X_i$ and consider

$$\begin{aligned}\mathbb{E}\left[S_n^4\right] =& \mathbb{E}\left[(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)\right. \\ &\left.\times (X_1 + \cdots + X_n)(X_1 + \cdots + X_n)\right]\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[S_n^4\right] &= n\mathbb{E}\left[X_i^4\right] + 6\begin{pmatrix} n \\ 2 \end{pmatrix}\mathbb{E}\left[X_i^2 X_j^2\right] \\ &= nK + 3n(n-1)\mathbb{E}\left[X_i^2\right]\mathbb{E}\left[X_j^2\right]\end{aligned}$$

Proof conti

We know

$$0 \le \text{Var}\left(X_i^2\right) \mathbb{E}\left[X_i^4\right] - \left(\mathbb{E}\left[X_i^2\right]\right)^2$$

So,

$$\left(\mathbb{E}\left[X_i^2\right]\right)^2 \le \mathbb{E}\left[X_i^4\right] = K.$$

Now,

$$\mathbb{E}\left[S_n^4\right] \le nK + 3n(n-1)K$$

Hence

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

What if with some positive probability,

$$\sum_{n=1}^{\infty} S_n^4/n^4 \quad \text{diverges?}$$

Where are we?

Alternative (classic) definition of probability of an event

We suppose that an experiment, whose sample space is S, is repeatedly performed under exactly the same conditions. For each event E of the sample space S, we define $n(E)$ to be the number of times in the first n repetitions of the experiment that the event E occurs. Then $P(E)$, the probability of the event E, is defined as

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

Now let

$$X_i = \begin{cases} 1 & \text{if } E \text{ occurs on the } i \text{ th trial} \\ 0 & \text{if } E \text{ does not occur on the } i \text{ th trial} \end{cases}$$

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

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

Bernstein Polynomials

Let $f(x)$ be a continuous function defined for $0 \leq x \leq 1$. Consider the functions

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}$$

and prove that

$$\lim_{n\to\infty} B_n(x) = f(x)$$

Hint: Let $X_1, X_2, \ldots$ be independent Bernoulli random variables with mean x. Show that

$$B_n(x) = E\left[f\left(\frac{X_1 + \cdots + X_n}{n}\right)\right]$$