Section 14 Law of Large Number

TA: Yasi Zhang

May 19, 2022

About CLT: When you read the problem and want to use CLT, first take notice that it is the case I (Normal) or the case II (other distributions) (sample size must be large).

1 Parameter Estimation

Parameters θ are given in probability while they are not given in statistics (or in real life)!

e.g. You never know the exact $P(X = head)$ when tossing a coin. You can only toss many many times to estimate the probability.

True parameter θ is not a random variable! It could be regarded as an unknown constant while $\hat{\theta}$ is considered as a random variable.

How do we judge an estimate is good or not?

- Bias: $Bias(\hat{\theta}) = E\hat{\theta} - \theta$
- Consistency: $\hat{\theta} \xrightarrow{p} \theta$?
- MSE: $E(\hat{\theta} - \theta)^2 = Bias^2(\hat{\theta}) + Var(\hat{\theta})$

1.1 Practice

4. Suppose $\{X_1, X_2, ..., X_2\}$ is an i.i.d. random sample from some population with unknown mean μ and variance σ^2. Define parameter $\theta = (\mu - 2)^2$.

(a) Suppose $\hat{\theta} = (\bar{X}_n - 2)^2$ is an estimator for θ, where $\bar{X}_n$ is the sample mean. Show that $\hat{\theta}$ is not unbiased for θ. [*Hint:* $\bar{X}_n - 2 = \bar{X}_n - \mu + \mu - 2$.]

(b) Find an unbiased estimator for θ. Hint: Consider an estimator which corrects the bias of $\hat{\theta} = (\bar{X}_n - 2)^2$.

Suppose $\{X_i\}$ are i.i.d. and $E(X_i) = \mu, Var(X_i) = \sigma^2$. Our interest is to estimate μ, using the following class of estimator:

$$\hat{\mu} = \sum_{i=1}^{n} c_i X_i$$

a. Show that $\hat{\mu}$ is unbiased for μ if and only if $\sum_{i=1}^{n} c_i = 1$.
b. Find the best unbiased estimator $\hat{\mu}^*$. Hint: Minimize $Var(\hat{\mu})$. Use Cauchy-Schwarz Inequality.

Remark 1 *We may pick X_1 as our estimate for μ and it is an unbiased estimater, but it is not as good as the estimate given in (b).*

2 Asymptotic Theory

Lemma 19 (7.2). **[Markov's Inequality]:** Suppose X is a random variable and $g(X)$ is a nonnegative function. Then for any $\epsilon > 0$, and any $k > 0$, we have

$$P[g(X) \geq \epsilon] \leq \frac{E[g(X)^k]}{\epsilon^k}.$$

Definition 74 (7.6). **[Convergence in Probability]:** A sequence of random variables $\{Z_n, n = 1, 2, \cdots\}$ converges in probability to a random variable Z if for every small constant $\epsilon > 0$,

$$P[|Z_n - Z| > \epsilon] \to 0 \text{ as } n \to \infty.$$

When Z_n converges in probability to Z, we write $\lim_{n\to\infty} P(|Z_n - Z| > \epsilon) = 0$ for every $\epsilon > 0$, or $p\lim_{n\to\infty} Z_n = Z$, or $Z_n \xrightarrow{p} Z$, or $Z_n - Z = o_P(1)$, or $Z_n - Z \xrightarrow{p} 0$.

Definition 73 (7.5). **[L_p-convergence]:** Let $0 < p < \infty$, and let $\{Z_n, n = 1, 2, \cdots\}$ be a sequence of random variables with $E|Z_n|^p < \infty$, and let Z be a random variable with $E|Z|^p < \infty$. Then Z_n converges in L_p to Z if

$$\lim_{n\to\infty} E|Z_n - Z|^p = 0.$$

Theorem 84 (7.4). **[Weak Law of Large Numbers (WLLN)]:** Let $\mathbf{X}^n = (X_1, \cdots, X_n)$ be an IID random sample with $E(X_i) = \mu$ and $\text{var}(X_i) = \sigma^2 < \infty$. Define $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Then for any given constant $\epsilon > 0$ and as $n \to \infty$,

$$\begin{aligned} P[|\bar{X}_n - \mu| &\leq \epsilon] \to 1, \\ \text{or } \bar{X}_n - \mu \xrightarrow{p} 0 \quad &\text{or} \quad \bar{X}_n - \mu = o_P(1). \end{aligned}$$

Lemma 22 (7.6). **[Continuity]:** Suppose $g(\cdot)$ is a continuous function, and Z_n converges in probability to Z. Then $g(Z_n)$ also converges in probability to $g(Z)$. That is, if $g(\cdot)$ is continuous, then $Z_n \xrightarrow{p} Z$ as $n \to \infty$ implies

$$g(Z_n) \xrightarrow{p} g(Z) \text{ as } n \to \infty$$

If $a_n \xrightarrow{p} a$, $b_n \xrightarrow{p} b$,then

$$\begin{aligned} a_n + b_n &\xrightarrow{p} a + b \\ a_n b_n &\xrightarrow{p} ab \\ Ca_n &\xrightarrow{p} Ca \\ g(a_n) &\xrightarrow{p} g(a) \end{aligned}$$

where $g(\cdot)$ is a continuous function and C is a constant.

Remark 2 *Since $S_n^2 \xrightarrow{p} \sigma^2$ (You can think about how to prove it), then $\sqrt{S_n^2} \xrightarrow{p} \sigma$, although $E\sqrt{S_n^2} \neq \sigma$.*

Since $\bar{X}_n \xrightarrow{p} \mu$, $\bar{X}_n^k \xrightarrow{p} \mu^k$.

3 Practice

Example 175 (7.11). Suppose $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID random sample from a $U[0, \theta]$ distribution, where $\theta > 0$ is an unknown parameter. Define a statistic $Z_n = \max_{1 \le i \le n}(X_i)$. Is Z_n consistent for θ?

Hint: $P(\max_{i=1}^n (X_i) < a) = P(X_1 < a, X_2 < a, \ldots, X_n < a)$
$P(\min_{i=1}^n (X_i) > a) = P(X_1 > a, X_2 > a, \ldots, X_n > a)$

4. Let $X_1, X_2, \ldots$ be a sequence of random variables such that

$$\mathbb{P}\left(X_n = \frac{1}{n}\right) = 1 - \frac{1}{n^2} \quad \text{and} \quad \mathbb{P}(X_n = n) = \frac{1}{n^2}.$$

Does X_n converge in probability? Does X_n converge in quadratic mean?

1. Suppose $\{X_i\}$ are i.i.d and follow $U(0, 1)$. Let

$$Z_n = (\Pi_{i=1}^n X_i)^{1/n}$$

Show that $Z_n \xrightarrow{p} C$, where C is a constant and find the value of C.
Hint: If $a_n \xrightarrow{p} a$, then $g(a_n) \xrightarrow{p} g(a)$ where $g(\cdot)$ is a continuous function.

4 Take-Home Practice

Show that $S_n^2 \xrightarrow{p} \sigma^2$ when X_i are iid normal.
Then, show that $S_n^2 \xrightarrow{p} \sigma^2$ when X_i are only iid.
Hint: Given that $a_n \xrightarrow{p} a$, we have $\frac{n}{n-1} a_n \xrightarrow{p} a$.

Example 175 (7.11). Suppose $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID random sample from a $U[0, \theta]$ distribution, where $\theta > 0$ is an unknown parameter. Define a statistic $Z_n = \max_{1 \le i \le n}(X_i)$. Is Z_n consistent for θ?

Solution: Given $\{|Z_n - Z| > \epsilon\} = \{Z_n - Z > \epsilon\} \cup \{Z_n - Z < -\epsilon\}$, we have

$$
\begin{aligned}
P(|Z_n - \theta| > \epsilon) &= P(Z_n > \theta + \epsilon) + P(Z_n < \theta - \epsilon) \\
&= P(Z_n < \theta - \epsilon) \\
&= P\left[\max_{1 \le i \le n}(X_i) < \theta - \epsilon\right] \\
&= P(X_1 < \theta - \epsilon, X_2 < \theta - \epsilon, \cdots, X_n < \theta - \epsilon) \\
&= \prod_{i=1}^{n} P(X_i < \theta - \epsilon) \text{ by independence} \\
&= \left(\frac{\theta - \epsilon}{\theta}\right)^n \\
&= \left(1 - \frac{\epsilon}{\theta}\right)^n \\
&\to 0 \text{ as } n \to \infty \text{ for any given } \epsilon > 0.
\end{aligned}
$$

It follows that Z_n is consistent for θ. The statistic $Z_n = \max_{1 \le i \le n} |X_i|$ is called an order statistic which involves some sort of ranking for the n random variables in the random sample $\mathbf{X}^n$.

Take-Home Practice

iid normal case: show the convergence in L2, i.e. $E(S_n^2 - \sigma^2)^2 \to 0$
iid case:

$$S_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2\right)$$
$$= \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}_n^2$$

where $c_n = d_n = \frac{n}{n-1} \to 1$.

Applying the law of large numbers to X_i^2:

$$n^{-1}\sum_{i=1}^{n} X_i^2 \xrightarrow{\text{P}} \mathbb{E}(X_i^2) = \sigma^2 + \mu^2$$

$$\overline{X}_n \xrightarrow{\text{P}} \mathbb{E}(X_i) = \mu \Rightarrow \overline{X}_n^2 \xrightarrow{\text{P}} \mu^2$$

Therefore, from theorem 6.5.e, $S_n^2 = c_n n^{-1}\sum_{i=1}^{n} X_i^2 - d_n\overline{X}_n^2 \xrightarrow{\text{P}} \sigma^2 + \mu^2 - \mu^2 = \sigma^2$.

4.

(a) $E\hat{\theta} = E(\overline{X}_n - 2)^2 \overset{?}{=} (\mu-2)^2$

$= E(\overline{X}_n - \mu + \mu - 2)^2$

$= \underbrace{E(\overline{X}_n - \mu)^2}_{\text{Var}(\overline{X}_n)} + 2E\underbrace{(\overline{X}_n - \mu)}_{0}(\mu-2) + (\mu-2)^2$

$= \frac{\sigma^2}{n} + (\mu-2)^2$

$\text{Bias}(\hat{\theta}) = E\hat{\theta} - \theta = \frac{\sigma^2}{n} \neq 0.$

Biased!

(b)

① $\tilde{\theta} = (\overline{X}_n - 2)^2 - (\overline{X}_n - \mu)^2$ (X)

② $\tilde{\theta} = (\overline{X}_n - 2)^2 - \frac{S_n^2}{n}$ (✓)

6.

(a) $E\hat{\mu} = E\sum_{i=1}^{n} c_i X_i$

$= \sum_{i=1}^{n} c_i E X_i$

$= \left(\sum_{i=1}^{n} c_i\right) \mu$

if and only if

① $E\hat{\mu} = \mu \Rightarrow \sum_{i=1}^{n} c_i = 1$

② $\sum_{i=1}^{n} c_i = 1 \Rightarrow E\hat{\mu} = \mu$

(b) $Var(X_i) = \sigma^2$

$\min \sum_{i=1}^{n} c_i^2$

$s.t. \sum_{i=1}^{n} c_i = 1$

By Cauchy-Schwarz Inequality,

$$\left(\sum_{i=1}^{n} (c_i \cdot 1)\right)^2 \leq \left(\sum_{i=1}^{n} c_i^2\right)\left(\sum_{i=1}^{n} 1\right)$$

$$(E(S \cdot 1))^2 \leq ES^2 E1$$

$\Rightarrow \sum_{i=1}^{n} c_i^2 \geq \frac{1}{n}$

The equality holds if and only if (linear dependent)

$c_i = c \cdot 1, \forall i$ where c is a constant

Thus, $c_1 = c_2 = \cdots = c_n$

Since $\sum c_i = 1$

Hence, $c_1 = c_2 = \cdots = c_n = \frac{1}{n}$, $\hat{\mu}^* = \frac{X_1 + \cdots + X_n}{n}$

Remark: Given $\{X_i\}_{i=1}^{n}$, we have multiple choices for $\hat{\mu}$

when $Var(X_i) = \sigma^2$, $\overline{X}_n$ is the best

- $\hat{\mu} = \overline{X}_n$
- $\hat{\mu} = X_1$ or $\frac{X_1 + X_2}{2}$ unbiased. but its variance too large

4.

$$EX_n = \frac{1}{n}\left(1-\frac{1}{n^2}\right) + n\frac{1}{n^2}$$
$$= \frac{2}{n} - \frac{1}{n^3}$$
$$= \frac{1}{n}\left(2-\frac{1}{n^2}\right)$$

$$P(|X_n - 0| > \varepsilon)$$
$$= P(X_n > \varepsilon)$$
$$< \frac{EX_n}{\varepsilon} = \left(\frac{2}{n} - \frac{1}{n^3}\right)\frac{1}{\varepsilon} \to 0$$

Thus $X_n \xrightarrow{P} 0$

$$E(X_n - 0)^2$$
$$= EX_n^2$$
$$= \frac{1}{n^2}\left(1-\frac{1}{n^2}\right) + n^2\frac{1}{n^2}$$
$$= \frac{1}{n^2} - \frac{1}{n^4} + 1 \not\to 0$$

Thus, $X_n \overset{q.m.}{\not\to} 0$

More strictly,

$$E(X_n - X)^2$$
$$= EX_n^2 - 2EX_nX + X^2$$
$$= \frac{1}{n^2} - \frac{1}{n^4} + 1 + \frac{2X}{n}\left(2-\frac{1}{n^2}\right) + X^2$$
$$\to 1 + X^2 \geq 1, \ \forall X$$

$$\log Z_n = \frac{1}{n} \log \prod X_i$$

$$= \frac{1}{n} \sum \log X_i \longrightarrow E \log X_i$$

$$E \log X_i = \int_0^1 \log x \, dx$$

$$= (y \log y - y)\Big|_0^1$$

$$= -1$$

Thus, $\log Z_n \xrightarrow{P} -1$

$$Z_n \xrightarrow{P} e^{-1}$$