Solutions Asymptotic Statistics Course

Data Analytics Lab

January 2020

Exercise 1

Prove Theorem 1.5

> **Theorem 1.5 Weak Law of Large Number**
> Suppose X_i are i.i.d. random variables with finite mean μ (i.e. $\mathbb{E}[X_i] = \mu$) and finite variance.
> Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, then $\bar{X}_n \xrightarrow{\mathcal{P}} \mu$

If we consider $x_i \quad$ i.i.d., $\quad \mathbb{E}[X_i] = \mu, \quad \mathrm{Var}(x_i) = \sigma^2 < \infty$

$$\bar{X}_n = \frac{1}{n}\sum x_i \quad \text{show that} \quad \bar{X}_n \longrightarrow \mu$$

That is, $\forall \varepsilon > 0, \forall \delta > 0, \exists n^* > 0 \quad$ s.t. $\quad \forall n \geqslant n^*$

$$\Pr\left(\left\|\bar{X}_n - \mu\right\| \geqslant \varepsilon\right) \leqslant \delta$$

Proof:

By linearity of the expectation ($\mathbb{E}(aX) = a\mathbb{E}(X)$) we have $\mathbb{E}\left[\bar{X}_n\right] = \mu$.
Indeed, $\mathbb{E}\left[\bar{X}\right] = \mathbb{E}\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n}\sum \mathbb{E}[X_i] = \frac{1}{n}\sum \mu = \mu$
Similarly, since we in presence of i.i.d. data, $\sum_{i=1}^{n} \mathrm{Var}(X_i) = n\,\mathrm{Var}(X_i) = n\sigma^2$.
Hence, $\mathrm{Var}\left(\bar{X}_n\right) = \frac{1}{n^2}\sum \mathrm{Var}(X_i) = \frac{1}{n^2}\sum \sigma^2 = \frac{\sigma^2}{n}$

> **Inequality of Chebyshev**
>
> $$\forall k > 0: \quad \Pr\left(\|\bar{X}_n - \mu\| \geq k\frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}$$

We then define by variable change $\tilde{\varepsilon}$ and defining therefore $\boxed{\tilde{\varepsilon} = k\frac{\sigma}{\sqrt{n}},\, k^2 = \frac{\tilde{\varepsilon}^2 n}{\sigma^2} \text{ and } \frac{1}{k^2} = \frac{\sigma^2}{n\tilde{\varepsilon}^2}}$

$$\Rightarrow \forall \tilde{\varepsilon} > 0: \quad \Pr\left(\|\bar{X}_n - \mu\| \geq \tilde{\varepsilon}\right) \leq \frac{\sigma^2}{n\tilde{\varepsilon}^2} \tag{1}$$

Then, let's take $\varepsilon > 0, \quad \delta > 0$ arbitrarily
Let's take $\tilde{\varepsilon} := \varepsilon$ and take $n^* > 0$ such that $\frac{\sigma^2}{n^*\varepsilon^2} < \delta$.
Therefore

$$\frac{\sigma^2}{n\tilde{\varepsilon}^2} \leq \frac{\sigma^2}{n^*\tilde{\varepsilon}^2} \leq \delta \tag{2}$$

Hence we have by (1) and (2), $\forall n \geq n^*$,

$$\Pr\left(\|\bar{X}_n - \mu\| \geq \tilde{\varepsilon})\right) \leq \frac{\sigma^2}{n\tilde{\varepsilon}^2} \leq \frac{\sigma^2}{n^*\tilde{\varepsilon}^2} \leq \delta \tag{3}$$

$$\Pr\left(\|\bar{X}_n - \mu\| \geq \tilde{\varepsilon})\right) \leq \delta \tag{4}$$

Exercise 2

Prove Corollary 1.4 for $k = \infty$

Corollary 1.4 Consistency equivalence

If $p < \infty$, then for all $k \in \mathbb{N}^* \cup \{\infty\}$

$$\lim_{n\to\infty} \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geq \varepsilon\right) = 0$$

if and only if

$$\lim_{n\to\infty} \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_k \geq \varepsilon\right) = 0$$

That is:

$$\lim_{n\to\infty} \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geq \varepsilon\right) = 0 \Leftrightarrow \lim_{n\to\infty} \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_k \geq \varepsilon\right) = 0$$

Definition 1.2 Consistency

The estimator $\hat{\boldsymbol{\theta}}$ is said to be consistent if it converges in probability to $\boldsymbol{\theta}_0$, i.e.

for all $\varepsilon > 0$

$$\lim_{n\to\infty} \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geq \varepsilon\right) = 0$$

More precisely, by the definition of the limit, this means that the estimator $\hat{\boldsymbol{\theta}}$ is consistent if $\forall \varepsilon > 0, \forall \delta > 0, \exists n^* \geq 0$ such that $\forall n \geq n^*$ we have

$$\Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geq \varepsilon\right) \leq \delta$$

Norm of a vector definitions & Inequalities

Consider the vector $x \in \mathbb{R}^p$

Definition of the norm $\|x\|_k$:

$\|x\|_k := \left(\sum_{i=1}^p |x_i|^k\right)^{1/k} = \left(|x_1|^k + |x_2|^k + |x_3|^k + \ldots + |x_p|^k\right)^{1/k}$

Definition of the infinite norm $\|\mathbf{x}\|_\infty$:

$\|x\|_\infty := \max\limits_{i\in\{1,\ldots,p\}} |x_i| \in \mathbb{R}_+$

L_2 and L_∞ inequation :

We can show that

$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{p}\|x\|_\infty$

- For the left hand side: $\|x\|_\infty = \max\limits_{i\in\{1,\ldots,p\}} |x_i| \leq \sqrt{\sum_{i=1}^p x_i^2} = \|x\|_2$
  Consider $j \in \{1, \ldots, p\}$ and $\max\limits_{i\in\{1,\ldots,p\}} |x_i| = x_j$. Then, $\|x\|_\infty = \|x\|_2$ i. f. f. $\forall i \neq j, x_i = 0$
- For the right hand side: $\|x\|_2 = \sqrt{\sum_{i=1}^p |x_i|^2} \leq \sqrt{\sum_{i=1}^p \max\limits_{i\in\{1,\ldots,p\}} |x_i|^2} = \sqrt{p \max\limits_{i\in\{1,\ldots,p\}} |x_i|^2} = \sqrt{p} \max\limits_{i\in\{1,\ldots,p\}} |x_i| = \sqrt{p}\,\|x\|_\infty$
  Therefore,
  $\|x\|_2 \leq \sqrt{p}\,\|x\|_\infty$

Proof:

When proving equivalence statement, we want to prove that $A \Leftrightarrow B$, which implies that $A \Rightarrow B$ and $B \Rightarrow A$. Therefore, we generally structure the proof as such:

- Prove $A \Rightarrow B$
- Prove $B \Rightarrow B$

Here, the structure of the proof is a bit different. We will first prove $B \Rightarrow A$ and then prove $A \Rightarrow B$. Also,
$A = \lim_{n\to\infty} \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geq \varepsilon\right) = 0$
and
$B = \lim_{n\to\infty} \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_\infty \geq \varepsilon\right) = 0$

Prove $B \Rightarrow A$

Hypothesis: Let us consider that $\forall \varepsilon > 0, \forall \delta > 0, \quad \exists n^* > 0 \quad$ such that $\forall n \geqslant n^*$,

$$\Pr\left(\left\|\hat{\theta} - \theta_0\right\|_\infty \geqslant \varepsilon\right) \leqslant \delta$$

Let's take $\varepsilon > 0, \quad \delta > 0 \quad$ arbitrarily and let's find $n^* > 0$ such that $\forall n \geqslant n^*$,

$$\Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geqslant \varepsilon\right) \leqslant \delta$$

We define $\tilde{\varepsilon} := \frac{\varepsilon}{\sqrt{p}}$ where p is the dimension of the vector θ_0. By hypothesis, for $\tilde{\varepsilon} := \frac{\varepsilon}{\sqrt{p}} \geq 0$ and $\delta \geq 0$, we know that $\exists n_0^* > 0$ such that $\forall n \geqslant n_0^*$

$$\Pr\left(\left\|\hat{\theta} - \theta_0\right\|_\infty \geqslant \tilde{\varepsilon}\right) \leqslant \delta \tag{5}$$

Therefore, rewriting (5) we obtain

$$\Pr(\left\|\hat{\theta} - \theta_0\right\|_\infty \geq \frac{\varepsilon}{\sqrt{p}}) \leq \delta$$
$$\Pr(\left\|\hat{\theta} - \theta_0\right\|_\infty \sqrt{p} \geq \varepsilon) \leq \delta$$

Since $\left\|\hat{\theta} - \theta_0\right\|_2 \leq \sqrt{p}\left\|\hat{\theta} - \theta_0\right\|_\infty$ and setting $n^* = n_0^*$, we have $\forall n \geq n^*$,

$$\Pr\left(\left\|\hat{\theta} - \theta_0\right\|_\infty \geqslant \varepsilon\right) \leqslant \delta \Rightarrow \Pr(\left\|\hat{\theta} - \theta_0\right\|_2 \geq \varepsilon) \leq \delta$$

Which therefore implies that $\|\cdot\|_\infty$-consistant $\Rightarrow$ $\|\cdot\|_2$-consistant

Prove $A \Rightarrow B$

Hypothesis: Let us consider that $\forall \varepsilon > 0, \forall \delta > 0, \exists n^* > 0$ such that $\forall n \geqslant n^*$,

$$\Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geqslant \varepsilon\right) \leqslant \delta$$

Let's take $\varepsilon > 0, \quad \delta > 0$ arbitrarily and let's find $n^* > 0$ such that $\forall n \geqslant n^*$,

$$\Pr\left(\left\|\hat{\theta} - \theta_0\right\|_\infty \geqslant \varepsilon\right) \leqslant \delta$$

By hypothesis, $\exists n^* > 0$ such that $\forall n \geqslant n^* \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geqslant \varepsilon\right) \leqslant \delta$.
Since $\left\|\hat{\theta} - \theta_0\right\|_\infty \leq \left\|\hat{\theta} - \theta_0\right\|_2$
$\forall n \geqslant n^* \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geqslant \varepsilon\right) \leqslant \delta \Rightarrow \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_\infty \geqslant \varepsilon\right) \leqslant \delta$
Which therefore implies that $\|\cdot\|_2$-consistant $\Rightarrow$ $\|\cdot\|_\infty$-consistant

□

Exercise 3

In $\mathbb{R}^\infty$, are the $\|\cdot\|_\infty$ -consistency and the $\|\cdot\|_2$ -consistency equivalent?

<u>Answer:</u>

No! for example because we can have $\|X_n\|_\infty \to 0$ while $\|X_n\|_2 \to \infty$

$\square$

Exercise 4

Prove Theorem 1.6 using Definition 1.2

Theorem 1.6 Continuous Mapping Theorem

Suppose $Y_n \xrightarrow{p} \mu$ then $g(Y_n) \xrightarrow{p} g(\mu)$ if $g(\cdot)$ is a *continuous function*.

Reminder

Let $f : \mathbb{R} \to \mathbb{R}$, and $x_o \in \mathbb{R}$

f is *continuous* in x_0 if $\forall \varepsilon > 0, \exists \delta_{\varepsilon,x_0} > 0$ such that $\forall x \in \mathbb{R}$

$$\|x - x_0\| \leq \delta_{\varepsilon,x_0} \implies \|f(x) - f(x_0)\| \leq \varepsilon$$

Definition 1.2 Consistency

The estimator $\hat{\boldsymbol{\theta}}$ is said to be consistent if it converges in probability to $\boldsymbol{\theta}_0$, i.e.

for all $\varepsilon > 0$

$$\lim_{n\to\infty} \Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geq \varepsilon\right) = 0$$

More precisely, by the definition of the limit, this means that the estimator $\hat{\boldsymbol{\theta}}$ is consistent if $\forall \varepsilon > 0, \forall \delta > 0, \exists n^* \geq 0$ such that $\forall n \geq n^*$ we have

$$\Pr\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geq \varepsilon\right) \leq \delta$$

Proof:

Here we want to prove $A \Rightarrow B$ where

$A = Y_n \xrightarrow{p} \mu$

and

$B = g(Y_n) \xrightarrow{p} g(\mu)$ where g is a continuous function ($\mathbb{R} \to \mathbb{R}$)

Hypothesis:

We first pose our starting hypothesis, the convergence in probability of y_n to μ.

$\forall \varepsilon > 0, \forall \delta > 0, \exists n^*_{\varepsilon,\delta}$ such that $\forall n \geq n^*_{\varepsilon,\delta}$

$$\Pr(\|y_n - \mu\| \geq \varepsilon) \leq \delta$$

Let's take $\varepsilon > 0$, $\delta > 0$ arbitrary and find $n* > 0$ such that $\forall n \geqslant n^*$

$$\Pr\left(\|g\left(Y_n\right) - g(\mu)\| \geqslant \varepsilon\right) \leqslant \delta$$

We now define subsets of $\mathbb{R}$ where y_n satisfies two conditions.

$\forall \delta' > 0$, define $B_{\delta'} := \{y \in \mathbb{R} | \exists x, \|x - y\| < \delta'$ and $\|g(x) - g(y)\| \geq \varepsilon\}$

Let's now consider that one of these conditions is fullfilled for some n.

Indeed, suppose that $\|g\left(y_n\right) - g(\mu)\| \geq \varepsilon$ for some n,

Then, either $\mu \in B_{\delta'}$ or $\mu \notin B_{\delta'}$, and if so, then $\|x - y\| \geq \delta'$

$\Rightarrow \{\mu \in B_{\delta'}\}$ or $\{\|y_n - \mu\| \geqslant \delta'\}$

Recall that if A and B are event and if $A \subset B$ then $\Pr(A) \leq \Pr(B)$

Therefore, $\Pr\left(\|g\left(y_n\right) - g(\mu)\| \geqslant \varepsilon\right) \leq \Pr\left(\{\mu \in B_{\delta'}\} \sqcup \{\|y_n - \mu\| \geqslant \delta'\}\right) = \Pr\left(\{\mu \in B_{\delta'}\}\right) + \Pr\left(\{\|y_n - \mu\| \geqslant \delta'\}\right)$

That is $\Pr\left(\|g\left(y_n\right) - g(\mu)\| \geqslant \varepsilon\right) \leqslant \Pr\left(\mu \in B_{\delta'}\right) + \Pr\left(\|y_n - \mu\| \geqslant \delta'\right\}$

Now, since the function g is continuous in $\mu, \exists \delta_{\varepsilon,\mu} > 0$ such that $\left[\|y - \mu\| < \delta_{\varepsilon,\mu} \Rightarrow \|g(y) - g(\mu)\| < \varepsilon\right]$

Therefore since one of the condition of the set $B_{\delta'}$ is $\|g(x) - g(y)\| \geq \varepsilon$,

$\Rightarrow \Pr\left(\mu \in B_{\delta_{\varepsilon,\mu}}\right) = 0$

and

$\Rightarrow \Pr\left(\|g\left(y_n\right) - g(\mu)\| \geqslant \varepsilon\right) \leqslant 0 + \Pr\left(\|y_n - \mu\| \geqslant \delta_{\varepsilon,\mu}\right)$

Therefore, since we have by hypothesis that $\exists n^*_{\delta_{\varepsilon,\mu},\delta} \geq 0$ such that $\forall n \geq n^*_{\delta_{\varepsilon,\mu},\delta}$

$$\Pr\left(\|y_n - \mu\| \geqslant \delta_{\varepsilon,\mu}\right) \leqslant \delta$$

We have that for $n^*_{\delta_{\varepsilon,\mu},\delta} \geqslant 0$ and $\forall n \geqslant n^*_{\delta_{\varepsilon,\mu},\delta}$

$$\Pr\left(\|g\left(y_n\right) - g(\mu)\| \geqslant \varepsilon\right) \leq \Pr\left(\|y_n - \mu\| \geqslant \delta_{\varepsilon,\mu}\right) \leqslant \delta$$

□

Note that this proof can be generalized to any function $g : \mathbb{X} \to \mathbb{Y}$

Exercise 5

Prove Theorem 1.8 using Theorem 1.7

Theorem 1.8 Consistency of Z-estimators

Let $\hat{\boldsymbol{\theta}} := \text{argzero}_{\boldsymbol{\theta}\in\Theta}\, \hat{\mathbf{g}}_n(\boldsymbol{\theta})$. If there exists a function $\mathbf{g}_0(\boldsymbol{\theta})$ such that:

- (C.1) $g_0(\theta)$ has a unique root in Θ at θ_0[a]
- (C.2) Θ is compact
- (C.3) $g_0(\theta)$ is continuous in θ
- (C.4) $\hat{\mathbf{g}}_n(\theta)$ converges uniformly in probability to $\mathbf{g}_0(\boldsymbol{\theta})$

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\|\hat{\mathbf{g}}_n(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})\right\| \xrightarrow{\mathcal{P}} 0$$

then we have $\hat{\boldsymbol{\theta}} \xrightarrow{\mathcal{P}} \boldsymbol{\theta}$

[a] this means that $\mathbf{g}_0(\boldsymbol{\theta}) = 0$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

Proof:

Set $\hat{Q}_n(\theta) = -\,\|\hat{g}_n(\theta)\|$ and $Q_0(\theta) := -\,\|g_0(\theta)\|$

We only need to check C4 of Theorem 1,7

Theorem 1.7 Consistency of Extremum estimators

If there is a function $Q_0(\theta)$ such that:

- (C.1) $Q_0(\theta)$ is uniquely maximized in θ_0
- (C.2) Θ is compact [a]
- (C.3) $Q_0(\theta)$ is continuous in θ
- (C.4) $\widehat{Q}_n(\theta)$ converges uniformly in probability to $Q_0(\theta)$[b]

then we have $\hat{\boldsymbol{\theta}} \xrightarrow{\mathcal{P}} \boldsymbol{\theta}$

[a] Compact, in finite dimension, means that Θ is both closed (i.e. containing all its limit points) and bounded (i.e. all its points are contained in a ball). [b] $\widehat{Q}_n(\theta)$ is said to converges uniformly in probability to $Q_0(\theta)$ if $\sup_{\boldsymbol{\theta}\in\Theta} |Q_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \xrightarrow{\mathcal{P}} 0$

$\left|\widehat{Q}_n(\theta) - Q_0(\theta)\right| = |-\,\|\widehat{g}_n(\theta)\| + \|g_0(\theta)\|\,|$

Δ ineq

$$\leqslant \|g_0(\theta) - \hat{g}_n(\theta)\|$$

Done by C,4 of Theroem 1.8

□

Exercise 6

Consider the following estimator:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} - ||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}(\boldsymbol{\theta})||^2_{\mathbf{W}}$$

where $||\mathbf{x}||^2_{\mathbf{A}} = \mathbf{x}^{\mathbf{T}}\mathbf{A}\mathbf{x}$, and where $\hat{\mu} \in \mathbb{R}^p, \mu(\theta) \in \mathbb{R}^p, W \in \mathbb{R}^{p\times p}$ and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. Note that the $\mathbf{W}$ may be random and/or depend on $\boldsymbol{\theta}$ but we assume that W > 0.

Is it possible to show that

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argzero}} \quad \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}(\boldsymbol{\theta})?$$

If so, do we need any additional conditions?

Simplify the conditions of Theorem 1.7 (or 1.8) for this estimator.

<u>Proof:</u>

Yes if $\frac{\partial}{\partial\theta^{\top}}||\hat{\mu} - \mu(\theta)||^2_{\mathbf{W}}$ and if $\hat{\theta} \in \operatorname{Int}(\mathbb{B})$

Assume $\hat{\mu} := \hat{\mu}(\theta_0, n)$

$\mu(\theta) := \lim_{n\to\infty} \hat{\mu}(\theta, n)$

Def $\hat{g}_n(\theta) := \hat{\mu}(\theta_0, n) - \mu(\theta)$

$g_0(\theta) = \mu(\theta_0) - \mu(\theta)$

Then $\hat{g}_n(\theta) - g(\theta) = \hat{\mu}(\theta_0, n) - \mu(\theta_0) \underset{n\to\infty}{\longrightarrow} 0$

So C4 is not necessary but I don't know if we assumed too much

□