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

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

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

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

Proof:
$\mathbb{E}\left[\bar{X}_n\right] = \mu$, by linearity of the expectation, $\operatorname{Var}\left(\bar{X}_n\right) = \frac{1}{n^2}\sum \operatorname{var}(x_i) = \frac{\sigma^2}{n}$

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

Taking $\tilde{\varepsilon}$ as an auxiliary for ε and defining therefore $\boxed{\tilde{\varepsilon} = k\frac{\sigma}{\sqrt{n}} \text{and similarly, } k^2 = \frac{\tilde{\varepsilon}^2 n}{\sigma^2}}$

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

Then, let's take $\varepsilon > 0, \quad \delta > 0$ arbitrarly

Let's take $\tilde{\varepsilon} := \varepsilon$ and take $n^\star > 0$ such that $\frac{\sigma^2}{n^\star \varepsilon^2} < \delta$.

Hence, $\forall n \geq n^\star$, we have

$$\Pr(\|\bar{x}_n - \mu\| \geq \varepsilon) \leq \delta$$

$\square$

Exercise 2

Prove Corollary 1.4

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

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

Where $\|\mathbf{x}\|_k := \left(\sum_{i=1}^{p} |x_i|^k\right)^{1/k}$

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:

Hypothesis: $\forall \varepsilon > 0, \forall \delta > 0, \quad \exists n^\star > 0 \quad$ s.t. $\quad \forall n \geqslant n^\star$

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

Since $\left\|\hat{\theta} - \theta_0\right\|_2 \leq p^{1/2}\|\hat{\theta} - \theta_0\|_\infty$, $\quad \boxed{p\text{, the dimension of } \theta_0}$

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

$\forall n \geqslant n^\star \quad P_r\left(\left\|\hat{\theta} - \theta_0\right\|_2 \geqslant \varepsilon\right) \leqslant \delta$

We know that $\exists n_0^\star > 0$ such that

$\Pr(\left\|\hat{\theta} - \theta_0\right\|_\infty \geq \frac{\varepsilon}{p^{\frac{1}{2}}}) \leq \delta$

$\Pr(\left\|\hat{\theta} - \theta_0\right\|_\infty p^{\frac{1}{2}} \geq \varepsilon) \leq \delta$

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

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

Since $\left\|\hat{\theta} - \theta_0\right\|_\infty \leq \left\|\hat{\theta} - \theta_0\right\|_2$

□

Exercise 3

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

Answer:

No! for example because of the fact that $\left\|\hat{\theta} - \theta_0\right\|_\infty \leq \left\|\hat{\theta} - \theta_0\right\|_2$.

Also, we can have $\|X_n\|_\infty \longrightarrow 0$ while $\|X_n\|_2 \longrightarrow \infty$

□

Exercise 4

Prove Theorem 1.6 using Definition 1.2

Theorem 1.6 Continuous Mapping Theorem

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

Reminder

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

$$\|x - x_0\| \leq \delta \implies \|f(x) - f(x_0)\| \leq \varepsilon$$

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

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

<u>**Proof:**</u>

Hypothesis

$\forall \varepsilon > 0, \forall \delta > 0, \exists n^*_{\varepsilon,\delta}$ 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 $\forall n \geqslant n^\star$

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

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

Suppose that $\|g\left(y_n\right) - g(\mu)\| > \varepsilon$ for some n,

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

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

Since $g()$ is continuous in $\mu, \exists \delta_{\varepsilon,\mu} > 0$ such that

$\left[\|y - \mu\| < \delta_{\varepsilon,\mu} \Rightarrow \|g(y) - g(\mu)\| < \varepsilon\right]$

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

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

Therefore,

$\Rightarrow$ for $n^\star_{\delta_{\varepsilon,\mu'},\delta} \geqslant 0$ and $\forall n \geqslant n^\star_{\delta_{\varepsilon,\mu'},\delta}$ we have $\Pr\left(\|y_n - \mu\| \geqslant \delta_{\varepsilon,\mu}\right) \leqslant \delta$

$\square$

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:

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

where $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^{\mathbf{T}}\mathbf{A}\mathbf{x}$, 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 $\frac{\partial}{\partial\theta^{\top}}\|\hat{\mu} - \mu(\theta)\|_{\mathbf{W}}^2$ and if $\hat{\theta} \in \operatorname{Int}(\mathbb{B})$

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

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

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

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

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

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

□