Question: What happens when $n \to \infty$

Observations: $X_1, X_2, \dots, X_n$ to be the first n terms of an infinite sequense of i.i.d. r.v.

$X_1, \dots, X_n, \dots$

Consider a family $\{\mathbb{P}_\vartheta\}_{\vartheta \in \Theta}$ of probability measures on $(\mathcal{X}, \mathcal{G})$ and the statistical model

$$(\mathbb{X}, \mathcal{F}, \{\mathbb{P}_\vartheta\}_{\vartheta \in \Theta}) = (\mathcal{X}^{\mathbb{N}}, \mathcal{G}^{\otimes \mathbb{N}}, \{\mathbb{P}_\vartheta\}_{\vartheta \in \Theta})$$

llowing us to define the infinite sequence:

$$\{\underline{X}_n\}_{n=1}^{\infty} = \{(X_1, \dots, X_n)\}_{n=1}^{\infty}$$

of samples observations.

The parameter of interest is.

$$g : \Theta \longrightarrow \mathbb{R}^p,$$

We let $\Gamma := \{ g(\vartheta) : \vartheta \in \Theta \}$ be the parameter space.

$\{T_n\}_{n=1}^{\infty}$. for $T_n: \mathcal{X}^{\otimes N} \to \Gamma$ a sequence

of estimators of $g(\vartheta)$.

Types of convergence:

(in Probability) Definition: Let $\{Z_n\}_{n=1}^{\infty}$ and Z be $\mathbb{R}^p$-valued r.v. defined on the same probability space. We say that $Z_n \xrightarrow{\mathbb{P}} Z$.

if $\forall \varepsilon > 0$: $\lim_{n\to\infty} \mathbb{P}[\|Z_n - Z\| > \varepsilon] = 0$

(in Distribution) Definition: Let $\{Z_n\}_{n=1}^{\infty}$ and Z be $\mathbb{R}^p$-valued r.v. We say that $Z_n \xrightarrow{d} Z$, if for all continuous and bounded functions f,

$$\lim_{n\to\infty} \mathbb{E}[f(Z_n)] = \mathbb{E} f(Z).$$

Remark: Convergence in ~~distribution~~ probability implies convergence in distribution

Example: (CLT) Let $X_1, \ldots, X_n, \ldots$ be i.i.d real-valued r.v. with mean μ and variance σ^2.

Let $\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ be the average of the first n terms. Then by CLT.

$$\underbrace{\sqrt{n}\,(\bar{X}_n - \mu)}_{Z_n} \xrightarrow{D} \underbrace{\mathcal{N}(0, \sigma^2)}_{Z}.$$

Remark: $\bar{X}_n - \mu$ by an order $\frac{1}{\sqrt{n}}$ in probability

$$\bar{X}_n - \mu \cong \frac{1}{\sqrt{n}} O_P(1) \qquad \text{if } \sigma^2 < \infty$$

Theorem: (Slutsky's thm) Let $(\{Z_n, A_n\}, Z)$ be a collection of $\mathbb{R}^p$-valued r.v. and $a \in \mathbb{R}^p$ be a vector of constants. Assume that $Z_n \xrightarrow{d} Z$ and $A_n \xrightarrow{P} a$. Then.

$$A_n^T Z_n \xrightarrow{d} a^T Z$$

Consistency and asymptotic Normality

Definition: A sequence of estimators $\{T_n\}$ of $g(\vartheta)$ is called consistent if

$$T_n \xrightarrow{\mathbb{P}_\vartheta} g(\vartheta)$$

* **Definition:** A sequence of estimators $\{T_n\}$ of $g(\vartheta)$ is called asymptotically normal if with asymptotic covariance matrix V_ϑ, if

$$\sqrt{n}\,(T_n - g(\vartheta)) \xrightarrow{d_\vartheta} \mathcal{N}(0, V_\vartheta).$$

* **Example:** sample mean is a consistent estimator for estimating the population mean.

Suppose $X_1, \dots, X_n$ i.i.d. r.v. $\mathbb{E}[X_i] = \mu$

and $\mathrm{Var}(X_i) = \sigma^2 < \infty$.

The sample mean $T_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

By L.L.N. $T_n \xrightarrow{\mathbb{P}} \mu \quad n \to \infty$

Remark: By increasing the sample size, we improve accuracy of our estimator of the population parameter.

Consistency M-estimators :

T : parameter space $\quad \Gamma := \{ g(\vartheta) : \vartheta \in \Theta \}$

$g : \Theta \to \mathbb{R}^p$

$\gamma = g(\vartheta)$

<u>Definition</u> theoretical risk: for $c \in \Gamma$

$$\mathcal{R}(c) := \mathbb{E}_\vartheta \, \rho_c(x) = \rho(x, c)$$

where $\quad \rho_c : \mathcal{X} \to \mathbb{R}$ is a loss function

$\rho \quad \mathcal{X} \times \mathbb{R}^p \to \mathbb{R}.$

and

$$\gamma := \arg\min_{c \in \Gamma} \underbrace{\mathbb{E}_\vartheta \, \rho_c(x)}_{=: \mathcal{R}(c)}$$

<u>Definition</u> : Empirical risk

$$\hat{\mathcal{R}}_n(c) := \frac{1}{n} \sum_{i=1}^{n} \rho_c(X_i), \quad c \in \Gamma$$

$\rho(x_i, c)$

<u>Definition</u> : The M-estimator $\hat{\gamma}_n$ of γ is defined as

$$\hat{\gamma}_n := \arg\min_{c \in \Gamma} \frac{1}{n} \underbrace{\sum_{i=1}^{n} \rho(X_i, c)}_{=: \hat{\mathcal{R}}_n(c)}.$$

Theorem: Suppose that

(A1) Γ : compact space.

(A2) $c \longmapsto \rho_c(x)$ is continuous for all x.

(A3) $\mathbb{P}_\vartheta [\sup_{c \in \Gamma} |\rho_c|] < \infty$

Then we have the uniform convergence.

$$\sup_{c \in \Gamma} \underbrace{\left| \frac{1}{n} \sum_{i=1}^{n} \rho_c(X_i) - \mathbb{E}_\vartheta \rho_c(X) \right|}_{|\hat{R}_n(c) - R(c)|} \longrightarrow 0 \quad \mathbb{P}_\vartheta\text{-a.s.}$$

By L.L.N. if $\mathbb{E}_\vartheta |\rho_c| < \infty \Longrightarrow$

$$\left| \underbrace{\frac{1}{n} \sum_{i=1}^{N} \rho_c(X_i)}_{\hat{R}_n(c)} - \underbrace{\mathbb{E} \rho_c(X)}_{R(c)} \right| \longrightarrow 0 \quad \mathbb{P}_\vartheta\text{-a.s}$$

$$|\hat{R}_n(c) - R(c)| \longrightarrow 0 \quad \mathbb{P}\text{-a.s.}$$

$\digamma$ $\hat{\gamma}_n \xrightarrow{\mathbb{P}_\vartheta\text{-a.s}} \gamma$ $\qquad R(\hat{\gamma}_n) \xrightarrow{\mathbb{P}_\vartheta\text{-a.s}} R(\gamma)$

$$\Longrightarrow \hat{\gamma}_n \xrightarrow{\mathbb{P}_\vartheta\text{-a.s.}} \gamma \quad \perp$$

Then minimizer γ of $R(c)$ is if γ well-seperated
called well seperated
if $\forall \varepsilon > 0 \quad \inf\{R(c) : c \in \Gamma, \|c - \gamma\| > \varepsilon\} > R(\gamma)$

<u>Theorem</u>: Suppose that

$$\sup_{c \in \Gamma} |\hat{R}_n(c) - R(c)| \longrightarrow 0 \quad \mathbb{P}_\vartheta\text{-a.s.}$$

then $\hat{R}_n(\hat{\gamma}_n) \longrightarrow R(\gamma)$ $\mathbb{P}_\vartheta$-a.s.

<u>Proof</u>: By uniform convergence

$$0 \le \mathbb{E}_\vartheta\left(\rho_{\hat{\gamma}_n}(x) - \rho_\gamma(x)\right)$$

$$= \mathbb{E}_\vartheta\left(\rho_{\hat{\gamma}_n}(x) - \rho_\gamma(x)\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\rho_{\hat{\gamma}_n}(x_i) - \rho(x_i)\right)$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\left(\rho_{\hat{\gamma}_n}(x_i) - \rho(x_i)\right)$$

$$= \mathbb{E}_\vartheta\left(\rho_{\hat{\gamma}_n}(x)\right) - \mathbb{E}_\vartheta\left(\rho_\gamma(x)\right) - \frac{1}{n}\sum_{i=1}^{n}\rho_{\hat{\gamma}_n}(x_i) + \frac{1}{n}\sum_{i=1}^{n}\rho_\gamma(x_i)$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\rho_{\hat{\gamma}_n}(x_i) - \frac{1}{n}\sum_{i=1}^{n}\rho_\gamma(x_i)$$

$$\le \mathbb{E}_\vartheta\, \rho_{\hat{\gamma}_n}(x) - \frac{1}{n}\sum_{i=1}^{n}\rho_{\hat{\gamma}_n}(x_i) + \frac{1}{n}\sum_{i=1}^{n}\rho_\gamma(x_i) - \mathbb{E}_\vartheta\left(\rho_\gamma(x)\right)$$

$$= -\left(\frac{1}{n}\sum_{i=1}^{n}\rho_{\hat{\gamma}_n}(x_i) - \mathbb{E}_\vartheta\left(\rho_{\hat{\gamma}_n}(x)\right)\right) + \frac{1}{n}\sum_{i=1}^{n}\rho_\gamma(x_i) - \mathbb{E}_\vartheta\left(\rho_\gamma(x)\right)$$

$$\le \left|\frac{1}{n}\sum_{i=1}^{n}\rho_{\hat{\gamma}_n}(x_i) - \mathbb{E}_\vartheta\left(\rho_{\hat{\gamma}_n}(x)\right)\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\rho_\gamma(x_i) - \mathbb{E}_\vartheta\left(\rho_\gamma(x)\right)\right|$$

$$\le 2 \sup_{c \in \Gamma}\left|\frac{1}{n}\sum_{i=1}^{n}\rho_c(x_i) - \mathbb{E}_\vartheta\left(\rho_c(x)\right)\right|$$

$$\longrightarrow 0 \quad \mathbb{P}_\vartheta\text{-a.s.}$$