Statistical Inference (II)

- $X_1, \dots, X_n \sim F$ with $F_{X_i}(x) = \mathbb{P}\{X_i \leq x\}$

$\uparrow$

Approach 1:

$F \in \mathcal{F} = \{ f(x;\theta) \mid \theta \in \Theta \}$

ex: i) Gaussian:

$$\mathcal{F} = \left\{ f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma^2}\right)^2} \;\middle|\; \begin{array}{l} \mu \in \mathbb{R} \\ \sigma \in \mathbb{R}^+ \end{array} \right\}$$

Remark: F is the cumulated distribution function (cdf)

f is the probability density function (pdf)

$$F(x) = \int_{-\infty}^{x} f(u)\,du$$

ii) Bernoulli distrib. (for binary data)

$$\mathcal{F} = \{ p(x, \theta) = \theta^x (1-\theta)^{1-x} \mid \theta \in [0,1] \}$$

with $p(x, \cdot)$ probability mass function (pmf)

Approach II (Non parametric)

$$\mathcal{F} = \{ F \mid F \text{ probability distribution} \}$$

Recall: $F(\cdot) \in [0,1]$ and F is "increasing"

$\Rightarrow F(x_2) \geq F(x_1) \; \forall \, x_1 \leq x_2$

•• Regression / Classification

observations: $(x_1, y_1), \dots, (x_n, y_n)$

x_i: regressor, feature, independent variable

y_i: response variable, dependent variable

Problem: estimate $\underbrace{\mathbb{E}(Y \mid X = x)}_{r(x)}$

Equivalent to say

$$Y = r(X) + \varepsilon \qquad \text{with } \mathbb{E}(\varepsilon) = 0$$

Ex $\quad r(X) = \alpha + \beta X \qquad$ (linear reg.)

Point estimation

Estimator : $X_1, \dots, X_N$ (i.i.d.) $\sim F_\theta$

θ true value

$$\hat{\theta}_N = g(X_1, \dots, X_N)$$

$$\text{bias}(\hat{\theta}_N) := \mathbb{E}_\theta(\hat{\theta}_N) - \theta$$

Ex: $X_1, \dots, X_N \overset{i.i.d.}{\sim} \mathcal{B}(p)$

$$\hat{p}_N = \frac{1}{N} \sum_{i=1}^N X_i, \qquad \mathbb{E}_p(\hat{p}_N) = \frac{1}{N} \sum_{i=1}^N \underbrace{\mathbb{E}_p(X_i)}_{p}$$

$$= \frac{1}{\cancel{N}} \cdot \cancel{N} p = p$$

Unbiased

Def 3c: $\hat{\theta}_N$ is a consistent estimator of θ if $\hat{\theta}_N \xrightarrow{\mathbb{P}_\theta} \theta$

Convergence in probability:

$$\lim_{N \to +\infty} \mathbb{P}_\theta\{ |\hat{\theta}_N - \theta| \geq \varepsilon \} = 0$$

$$se(\hat{\theta}_N) = \left(\mathbb{V}ar_\theta(\hat{\theta}_N)\right)^{1/2}$$

Ex: $X_1, \dots, X_N \overset{\text{i.i.d.}}{\sim} \mathcal{B}(p)$

$$\hat{p}_N = \frac{1}{N}\sum_{i=1}^N x_i \quad \Rightarrow \quad \mathbb{E}_p(\hat{p}_N) = p$$

$$\mathbb{V}ar_p(\hat{p}_N) = \mathbb{V}ar_p\left(\frac{1}{N}\sum_{i=1}^N x_i\right) =$$

$$= \frac{1}{N^2}\mathbb{V}ar_p\left(\sum_{i=1}^N x_i\right)$$

$$= \frac{1}{N^2}\sum_{i=1}^N \underbrace{\mathbb{V}ar(x_i)}_{p(1-p)} = \frac{1}{N^{\cancel{2}}}\cancel{N}\, p(1-p)$$

$$= \frac{p(1-p)}{N}$$

$$se(\hat{p}_N) \rightarrow \hat{se}(\hat{p}_N)$$

$$\frac{p(1-p)}{N} \rightarrow \frac{\hat{p}_N(1-\hat{p}_N)}{N}$$

We can use the bias of $\hat{\theta}_N$ and $\mathbb{V}ar_\theta(\hat{\theta}_N)$ order to compute the MEAN SQUARED ERROR $MSE_\theta(\hat{\theta}_N)$

$$MSE(\hat{\theta}_N) := \mathbb{E}_\theta[\hat{\theta}_N - \theta]^2$$

$$\overset{\text{Th.}}{=} \mathbb{V}ar_\theta(\hat{\theta}_N) + bias^2(\hat{\theta}_N)$$

proof: $\mathbb{E}_\theta[\hat{\theta}_N - \bar{\theta}_N + \bar{\theta}_N - \theta]^2 =$ with $\bar{\theta}_N = \mathbb{E}_\theta[\hat{\theta}_N]$

$$= \mathbb{E}_\theta[\hat{\theta}_N - \bar{\theta}_N]^2 + \cancel{\mathbb{E}}[\bar{\theta}_N - \theta]^2 + 2\,\mathbb{E}(\hat{\theta}_N - \bar{\theta}_N)(\bar{\theta}_N - \theta)$$

$$= \mathbb{V}ar_\theta(\hat{\theta}_N) + bias^2_\theta(\hat{\theta}_N)$$

$$2(\bar{\theta}_N - \theta)\underbrace{\mathbb{E}(\hat{\theta}_N - \bar{\theta}_N)}$$

$$0 = \mathbb{E}_\theta(\hat{\theta}_N) - \mathbb{E}_\theta(\hat{\theta}_N)$$

$\square$

i) Consistent estimator is such that

$$\hat{\theta}_N \xrightarrow{\mathbb{P}_\theta} \theta$$

ii) $$MSE_\theta(\hat{\theta}_N) = bias^2(\hat{\theta}_N) + \mathbb{V}ar_\theta(\hat{\theta}_N)$$

Theorem: if $\hat{\theta}_N$ is unbiased and

$$\lim_{N \to \infty} \mathbb{V}ar_\theta(\hat{\theta}_N) = 0$$

$\Rightarrow$ $\hat{\theta}_N$ is consistent

Proof: $$\mathbb{P}_\theta\{|\hat{\theta}_N - \theta| > \varepsilon\} \xrightarrow[N \to +\infty]{?} 0 \qquad \varepsilon > 0$$

this is what we have to show

$$\mathbb{P}_\theta\{|\theta_N - \theta| > \varepsilon\} = \mathbb{P}_\theta\{|\theta_N - \theta|^2 > \varepsilon^2\} \leq$$

$$\leq \frac{\mathbb{E}_\theta[(\hat{\theta}_N - \theta)^2]}{\varepsilon^2} =$$

by Markov inequality

$$= \frac{MSE(\hat{\theta}_N)}{\varepsilon^2} = \frac{\cancelto{0}{bias^2(\hat{\theta}_N)} + \mathbb{V}ar(\hat{\theta}_N)}{\varepsilon^2} =$$

$$= \cancel{\frac{\mathbb{V}ar(\hat{\theta})}{\varepsilon^2 N}} \xrightarrow[N \to +\infty]{} 0 \qquad \square$$

- Exercise :

$x_1, \dots, x_N \overset{i.i.d.}{\sim} \mathcal{P}(\lambda)$, $\hat{\lambda}_N = \frac{1}{N} \sum_{i=1}^{N} x_i$

Compute the bias, se and MSE of $\hat{\lambda}_N$.

i) $\mathbb{E}_\lambda(\hat{\lambda}_N) = \mathbb{E}_\lambda\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\mathbb{E}(x_i)}_{=\lambda}$

$$= \frac{\cancel{N}}{\cancel{N}}\lambda \qquad \text{UNBIASED}$$

$$\Rightarrow \quad \text{bias}(\hat{\lambda}_N) = \underbrace{\mathbb{E}(\hat{\lambda}_N)}_{=\lambda} - \lambda = 0 .$$

ii) $\mathbb{V}ar_\lambda(\hat{\lambda}_N) = \mathbb{V}ar_\lambda\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) = \frac{1}{N^2}\sum_{i=1}^{N} \underbrace{\mathbb{V}ar(x_i)}_{\lambda}$

$$= \frac{1}{N^{\cancel{2}}} \cdot \cancel{N}\lambda = \frac{\lambda}{N} \quad \text{INFINITESIMAL} \left(\xrightarrow[N \to +\infty]{} 0\right)$$

iii) $\text{MSE}(\hat{\lambda}_N) = \underbrace{\text{bias}^2(\hat{\lambda}_N)}_{=0} + \mathbb{V}ar_\lambda(\hat{\lambda}_N) = \frac{\lambda}{N}$

if $\hat{\theta}_N$ is unbiased, it is asymptotically normally distributed if

$$\frac{\hat{\theta}_N - \theta}{se(\hat{\theta}_N)} \xrightarrow{d} N(0,1)$$