Model

$X_1, X_2, \cdots, X_n$ i.i.d. from unknown distribution F.

Parametric modeling

" Finite-dimensional " Estimation

e.g. Exponential Distribution

$\theta = \lambda$ (" rate " parameter)

$f(\theta, x) = \lambda e^{-\lambda x}$

e.g. Normal Distribution

$\theta = (\mu, \sigma^2)$

$$f(\theta, x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(x-\mu)^2/\sigma^2)$$

How should we estimate the unknown "true" value θ_0 ?

e.g. Exponential Distribution

$X \overset{D}{=} Exp(\lambda_0)$

$$\begin{cases} EX = 1/\lambda_0 \\ var\ X = 1/\lambda_0^2 \\ P(X \le x) = 1 - e^{-\lambda_0 x} \end{cases}$$

① $\bar{X}_n \to EX = 1/\lambda_0$

$\hat{\lambda} = 1/\bar{X}_n$

② $\bar{S}_n = 1/\lambda_0^2$

$\hat{\lambda} = 1/S_n$

③ $\frac{1}{n}\sum_{i=1}^{n} I(X \le x) \to 1 - e^{-\lambda_0 x}$

$e^{-\hat{\lambda} x} = \frac{1}{n}\sum_{i=1}^{n} I(X_i > x)$

$\hat{\lambda} = \cdots$

Which estimator do we choose ?

$$n^{\frac{1}{2}} (\hat{\lambda} - \lambda_0) \Rightarrow \eta N(0, 1)$$

Choose the estimator that minimizes the variance !

Principle of Maximum Likelihood

e.g. X_1, X_2, X_3 i.i.d. $Ber(p_0)$

$0, 1, 1$

$$L(p) = (1-p) p^2$$

$$\hat{p} = \underset{p \in [0,1]}{\arg\max} L(p)$$

$$\hat{p} = \frac{2}{3}$$

e.g. X_1, X_2, X_3 i.i.d. $N(\mu_0, \sigma_0^2)$

$$L_n(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)$$

$$l_n(\mu, \sigma) = \sum_{i=1}^{n} \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(X_i-\mu)^2}{2\sigma^2} \right]$$

$$\frac{\partial l}{\partial \mu} = - \sum_{i=1}^{n} \frac{(X_i-\hat{\mu})^2}{\hat{\sigma}^2} = 0$$

$$\Rightarrow \hat{\mu} = \bar{X}_n$$

$$\frac{\partial l}{\partial \sigma^2} = - \frac{2}{2} \frac{1}{\hat{\sigma}^2} + \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{2\hat{\sigma}^4} = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

$$= \left(\frac{n-1}{n}\right) S_n^2$$

S_n^2 is preferred

$\hookrightarrow \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

- when $n = 1$, S_n is undefined
- $ES_n^2 = \sigma_0^2$; $E\hat{\sigma}^2 = \left(\frac{n-1}{n}\right) \sigma_0^2$

↑ unbiased

$X_1, X_2, \cdots, X_n$ i.i.d. $f(\theta_0, \cdot)$

$$L_n(\theta) = \prod_{i=1}^{n} f(\theta, X_i)$$

$$l_n(\theta) = \sum_{i=1}^{n} \log f(\theta, X_i)$$

$$\frac{1}{n} l_n(\theta) \to E_{\theta_0} \log f(\theta, X_1) \quad \text{a.s.}$$

$$\frac{1}{n} l_n(\theta) - \frac{1}{n} l_n(\theta_0) \xrightarrow{a.s.} E_{\theta_0}[\log f(\theta, X_1) - \log f(\theta_0, X_1)]$$

$$= E_{\theta_0} \log \frac{f(\theta, X_1)}{f(\theta_0, X_1)}$$

$$\overset{\text{unless } X \text{ is deterministic}}{<} \underbrace{\log E_{\theta_0} \frac{f(\theta, X)}{f(\theta_0, X_1)}}_{\int_{\mathbb{R}} \frac{f(\theta, x)}{f(\theta_0, x)} \cdot f(\theta_0, x)\,dx = 1} = 0$$

"consistency" $\hat{\theta}_n \to \theta_0$ a.s.

---

non-identifiability:

$$f(\theta_1, \cdot) = f(\theta_2, \cdot)$$

identifiability:

MLE consistent

$X_1, X_2, \cdots, X_n$ i.i.d. $\text{Gamma}(\overset{\text{rate}}{\lambda}, \overset{\text{shape}}{\alpha})$

$$f(\lambda, \alpha, x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$$

$\alpha = 1 \Rightarrow$ Exponential Dist.

$$L(\lambda, \alpha) = \prod_{i=1}^{n} \lambda(\lambda x_i)^{\alpha-1} e^{-\lambda x_i} / \Gamma(\alpha)$$

$$l(\lambda, \alpha) = \sum_{i=1}^{n} [\alpha \log \lambda + (\alpha - 1) \log x_i - \lambda x_i - \underline{\log \Gamma(\alpha)}]$$

No closed form for $\hat{\lambda}$ and $\hat{\alpha}$

- Just optimize numerically $\ell_n(\cdot)$
  - can be expensive

We might want to avoid the computational issue.

$$\operatorname{Var} X = \frac{\alpha}{\lambda^2}$$

$$E X = \frac{\alpha}{\lambda}$$

$$\Rightarrow \quad S_n^2 = \frac{\hat{\alpha}}{\hat{\lambda}^2}$$

$$\bar{X}_n = \frac{\hat{\alpha}}{\lambda}$$

- "Methods of Moment"
  ↳ can be used as an "initial guess" for MLE

MLE: - statistically efficient
  ↳ minimizing variance in CLT
- computationally expensive

MoM: - less efficient statistically
- computationally faster often time

Special cases of "Estimating Equations"

$$E_{\theta_0} g(\theta_1, X) = 0 \quad \text{iff} \quad \theta_1 = \theta_0$$

$$\left(\begin{array}{l} \theta_0 \text{ is the soln of } E_{\theta_0} g(\theta, X) = 0 \\ \bar{\theta}_n \text{ is the soln of } \frac{1}{n}\sum_{i=1}^{n} g(\theta, X_i) = 0 \\ \bar{\theta}_n \xrightarrow{a.s.} \theta_0 , \quad \sqrt{n}(\bar{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, \frac{\operatorname{var}_{\theta_0} g(\theta_0, X)}{(E_{\theta_0} g'(\theta_0, X))^2}\right) \end{array} \right)$$

$$\frac{1}{n}\sum_{i=1}^{n} s(\theta, x_i) \xrightarrow{LLN} E_{\theta_0} s(\theta, x_1) \quad a.s.$$

$$\frac{1}{n}\sum_{i=1}^{n} s(\hat{\theta}, x_i) = 0$$

MLE:

$$s(\theta, x) = \nabla_\theta \frac{f(\theta, x)}{f(\theta, x)}$$

$$\left[\frac{1}{n}\sum_{i=1}^{n} \frac{\nabla_\theta f(\hat{\theta}, x_i)}{f(\hat{\theta}, x_i)} = 0 \right]$$

gradient of log f

MoM:

moment

$$s(\theta, x) = k(x) - E_\theta k(x_1)$$

$$EX = k_1(x)$$

$$EX^2 = k_2(x)$$

$$E_{\theta_0} s(\theta_0, x) = 0$$

MLE:

$$\frac{1}{n}\sum_{i=1}^{n} \overset{0}{\overset{\parallel}{s(\hat{\theta}, x_i)}} - \frac{1}{n}\sum_{i=0}^{n} s(\theta_0, x_i)$$

$$= -\frac{1}{n}\sum_{i=0}^{n} s(\theta_0, x_i)$$

If $\hat{\theta}$ is constitent for θ_0, tylor expand:

$$\frac{1}{n}\sum_{i=1}^{n} s'(\xi, x_i) \cdot n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$$

$$= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[s(\theta_0, x_i) - \overset{0}{\overset{\parallel}{E_{\theta_0} s(\theta_0, x_i)}}\right]$$

$$\downarrow P$$

$$E_{\theta_0} s'(\theta_0, x_1)$$

$$\Downarrow$$

$$N(0, Var_{\theta_0} s(\theta_0, x_1))$$

So,

$$n^{\frac{1}{2}}(\hat{\theta} - \theta_0) \Rightarrow N\left(0, \frac{Var_{\theta_0} s(\theta_0, x_1)}{E_{\theta_0} s'(\theta_0, x_1)^2}\right)$$