Nov 6, 2023

Maximum Likelihood Estimator

(Gold standard for estimation)
Asymptotic efficiency property: variance of MLE is the smallest from the class of such estimators.

* KL - divergence

The measure that is used to evaluate proximity b/w 2 distn functions, typically 2 density fns

Suppose that f and g are density functions and $X \sim f(\cdot)$ (random variable).

Then KL divergence.

$$KL(f \| g) = E_x\left[\log \frac{f(X)}{g(X)}\right]$$

θ_0 - parameter of true data generating model (DGP)
G

$X \sim f(\cdot, \theta_0)$

Family of models $\{f(\cdot, \theta), \theta \in \Theta\}$ Parameter space.

How are we going to find a model from the set of possible models? → Look at KL divergence.

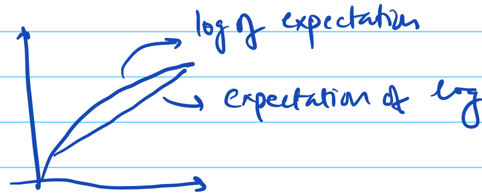
expectation wrt θ_0 (use pdf of θ_0)

$$E_{\theta_0}\left[\log \frac{f(X, \theta_0)}{f(X, \theta)}\right]$$

divergence b/w actual & sample

It is the smallest when $\theta = \theta_0$

$$E_{\theta_0}\left[\log \frac{f(X,\theta_0)}{f(X,\theta)}\right] = -E_{\theta_0}\left[\log \frac{f(X,\theta)}{f(X,\theta_0)}\right]$$



Using Jensen's Inequality :-

$$-E_{\theta_0}\left[\log \frac{f(X,\theta)}{f(X,\theta_0)}\right] \geq -\log\left(E_{\theta_0}\left[\frac{f(X,\theta)}{f(X,\theta_0)}\right]\right)$$

$$E_{\theta_0}\left[\frac{f(X,\theta)}{f(X,\theta_0)}\right] = \int_{-\infty}^{\infty} \frac{f(x,\theta)}{f(x,\theta_0)}\, f(x,\theta_0)\, dx$$

Density of true generating f^n

$$= 1.$$

$$\Rightarrow -\log\left(E_{\theta_0}\left[\frac{f(X,\theta)}{f(X,\theta_0)}\right]\right) = 0$$

$$KL(f\|g) = E_{X\sim f}\left[\log \frac{f(X)}{g(X)}\right] \geq 0$$

This is going to min when $g(X) = f(X)$

Or in the above model (the parametric model);

$$E_{\theta_0}\left[\log \frac{f(X,\theta_0)}{f(X,\theta)}\right] \quad \text{is min when } \theta = \theta_0.$$

(*) Rewriting;

$$E_{\theta_0}[\log f(X,\theta_0)] - E_{\theta_0}[\log f(X,\theta)]$$

$$\theta_0 = \underset{\theta \in \Theta}{\operatorname{argmin}} -E_{\theta_0}[\log f(X,\theta)]$$

$$= \underset{\theta \in \Theta}{\operatorname{argmax}} \underbrace{E_{\theta_0}[\log f(X,\theta)]}_{\text{log-likelihood function}}$$

* Take fixed values of a set :–

$\{x_i\}_{i=1}^{n}$

$$L(\theta) = E_{\theta_0}[\log f(X,\theta)]$$

$$\hat{L}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log f(x_i,\theta) \qquad \text{(sample log-likelihood fn)}$$

$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}}\ \hat{L}(\theta)$ is the MLE.

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

(Max. across these f^n would work as a R.V.)

Example :—

Assume $X \sim N(\mu_0, 1)$

Potential set of models $\{N(\mu, 1),\ \mu \in \mathbb{R}\}$

$$f(x, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

log likelihood f^n :

$$\log f(x, \mu) = -\frac{1}{2}\log(2\pi) - \frac{(x-\mu)^2}{2}$$

$$\hat{l}(\mu) = -\frac{1}{2}\log(2\pi) - \frac{1}{2n}\sum_{i=1}^{n}(x_i-\mu)^2$$

Estimator :

$$\hat{\mu}_{MLE} = \underset{\mu}{\operatorname{argmax}}\left(-\frac{1}{2}\log(2\pi) - \frac{1}{2n}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

Taking derivation to find the max;

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_{MLE}) = 0$$

?

$$\Rightarrow \hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{P} \mu_0 \quad (\text{LLN})$$

(assuming iid : in this course we only did LLN for iid.)

Next class: –

1. Go back to risk f^n & use taylor expansion for linearilization. (like in delta function).