Estimators

Monday, 28 March 2022 11:50

Unbiased Estimators:

An estimator of a parameter θ is a function $T = T(x)$ which we use to estimate θ from an observation of X. T is said to be unbiased if:

$$\mathbb{E}(T) = T$$

Ex: Suppose we have $X_1, X_2 \dots X_n$ are IID $B(1,p)$. p is unknown

Consider an estimator for p: $\hat{p}(x) = \sum x_i / n$.

$$\mathbb{E}\hat{p}(x) = \mathbb{E}\left[\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right] = \frac{1}{n}\left(\underbrace{\mathbb{E}X_1 + \dots + \mathbb{E}(X_n)}\right)$$

$$= \frac{1}{n}(np) = p.$$ — unbiased estimator!

$$\tilde{p} = \frac{1}{3}(X_1 + 2X_2)$$

$$\mathbb{E}\tilde{p}(x) = \frac{1}{3}\mathbb{E}(X_1 + 2X_2) = \frac{1}{3}(\mathbb{E}X_1 + 2\mathbb{E}X_2)$$

$$= \frac{1}{3}(p + 2p) = p.$$ $\rightarrow$ unbiased estimator!

Sufficient Statistics:

Our goal is to infer 'θ' from the data.

Data : X

Statistics is the function of the data.

$$T(X)$$

The MLE, if it exists, is always a function of a Sufficient Statistics.

$T = T(x_1, x_2, \dots, x_n) \Rightarrow$ it Summarises all information in $\{x_1 \dots x_n\}$ which is relevant to inference about θ.

Theorem:

The Statistics T is Sufficient for θ if and only if $f(x|\theta)$ Can be expressed as

$$lik(\theta) = f(x|\theta) = g(T(x), \theta) \cdot h(x) \leftarrow$$

factorization criterion

Ex: $X_1 \dots X_n \sim \text{Poisson}(\lambda)$

Need to estimate λ.

$$lik(\lambda) = f(x|\lambda) = \prod_{i=1}^{n} \left\{ \frac{\lambda^{x_i} e^{-\lambda}}{x!} \right\}$$

$$= \frac{\lambda^{\sum_i x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

$$= g(T(x), \lambda)\, h(x).$$

if $g(T(x),\lambda) = \lambda^{\sum x_i} e^{-n\lambda}$

$h(x) = \frac{1}{\prod_{i=1}^{n} x_i}$

Sufficient Statistics here is $\boxed{t = \sum_i x_i}$

If $T(X)$ is a sufficient statistics then so are statistic like $T(x)/n$ and $\log T(x)$.

Mean Squared Error:

(estimator) $\hat{\theta}$ θ (true)

If $\hat{\theta}$ is an unbiased estimator $(\mathbb{E}\hat{\theta} = \theta)$ then $\mathbb{E}(\hat{\theta} - \theta)^2$ is the variance of $\hat{\theta}$.

If $\hat{\theta}$ is a biased estimator $(\mathbb{E}\hat{\theta} \neq \theta)$ then $\mathbb{E}(\hat{\theta} - \theta)^2$ is not the variance of $\hat{\theta}$.

$\mathbb{E}(\hat{\theta} - \theta)^2$ is still a useful quantity to measure the mean squared error (MSE) of $\hat{\theta}$.

the Mean-squared error (MSE) of θ

Ex: Estimator A: $\hat{p}_1 = \frac{1}{n}(X_1 + \cdots + X_n)$

Estimator B: $\hat{p}_2 = \frac{1}{3}(X_1 + 2X_2)$.

$$Var(\hat{p}_1) = Var\left[\frac{1}{n}(X_1 + \cdots + X_n)\right] = \frac{Var(X_1) + \cdots + Var(X_n)}{n^2}$$

$$= \frac{n\,p(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

$$Var(\hat{p}_2) = var\left[\frac{1}{3}(X_1 + 2X_2)\right] = \frac{1}{9}\left(Var(X_1) + 4Var(X_2)\right)$$

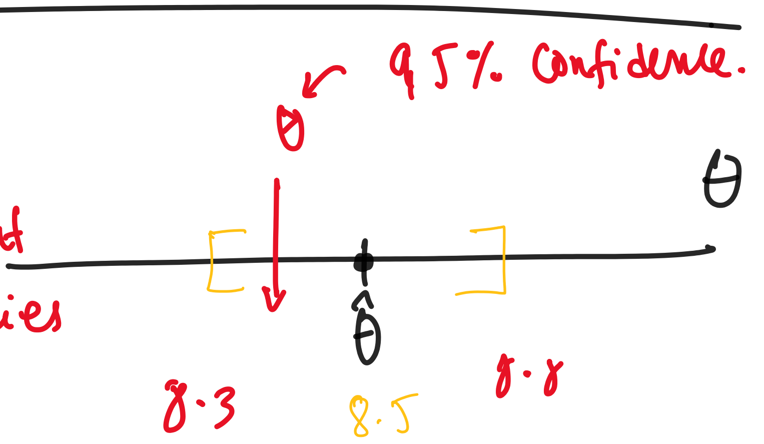
$$= 1/9 \cdot 5p(1-p).$$

$$Var(\hat{p}_1) < var(\hat{p}_2)$$

$$Var(\hat{p}_1)/var(\hat{p}_2) \to 0 \quad \text{as } n \to \infty$$

Confidence Interval:

100α% Confident that the true value lies within this interval.

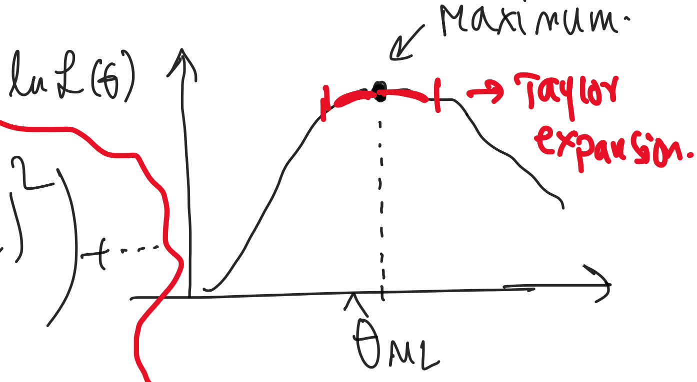


Consider a general likelihood function $\mathcal{L}(\theta)$; let's do a Taylor expansion of the log-likelihood function around its Maximum, $\hat{\theta}_{ML}$

vanishes.

$$\ln \mathcal{L}(\theta) = \ln \mathcal{L}(\hat{\theta}_{ML}) + \underbrace{\frac{\partial \ln \mathcal{L}(\theta)}{\partial \theta}\bigg|_{\theta=\hat{\theta}_{ML}} (\theta - \hat{\theta}_{ML})}_{\text{vanishes}}$$

$$+ \frac{1}{2} \frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial \theta^2}\bigg|_{\theta=\hat{\theta}_{ML}} (\theta - \hat{\theta}_{ML})^2 + \dots$$



$$\boxed{\begin{aligned} \mathcal{L}(\theta) &\approx \mathcal{L}(\hat{\theta}_{ML}) \exp\left(-\frac{1}{2} \frac{(\theta - \theta_{ML})^2}{\Sigma_\theta}\right) + \dots \\ \frac{1}{\Sigma_\theta} &= - \frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial \theta^2}\bigg|_{\theta=\hat{\theta}_{ML}} . \end{aligned}}$$

error on $\hat{\theta}_{ML}$ → confidence interval.

Second derivative of the log-likelihood function :

Fisher Matrix.

Example: Gaussian Case: (μ, σ^2)

$$\hat{\mu}_{ML} = \bar{X} \quad \rightarrow \quad \Sigma_{\hat{\mu}_{ML}} = \sigma^2/N$$

This means that the uncertainty on our ML estimate for μ is proportional to $1/\sqrt{N}$ with N being the Number of measurements.

100 α % confidence interval.

$\alpha = 0.95$

$\alpha = 0.99$

Confidence interval: $[\mu_{min}, \mu_{max}]$ is a 100 α %

$$P(\mu_{min} < \mu < \mu_{max}) = \alpha$$

→ $[\hat{\mu}_{ML} - \Sigma_{\hat{\mu}_{ML}} < \mu < \hat{\mu}_{ML} + \Sigma_{\hat{\mu}_{ML}}]$ is 68% confidence interval. ("1σ interval").

→ $[\hat{\mu}_{ML} - 2\Sigma_{\hat{\mu}_{ML}} < \mu < \hat{\mu}_{ML} + 2\Sigma_{\hat{\mu}_{ML}}]$ is a 95.4% confidence interval ("2σ interval")