

Stochastic Processes

Estimation Theory



Hamid R. Rabiee



Overview

- Reading Assignment
 - Chapter 7 of C.B. book.

- Further Resources
 - MIT Open Course Ware



Methods of finding estimations:

- Method of Moments
- Maximum Likelihood
- Bayes Estimators



Method of Moments

- Oldest, simplest method
- Dates back to Karl Pearson in 1880
- Let x_1, x_2, \dots, x_n be a sample from $f(x | \theta_1, \dots, \theta_k)$
- Assume first k sample moments are equal to corresponding k population moments.

$$\begin{aligned} m_1 &= \frac{1}{n} \sum_{i=1}^n x_i & \mu_1 &= E[X] & \mu_1(\theta_1, \dots, \theta_k) &= m_1 \\ m_2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 & \mu_2 &= E[X^2] & \mu_2(\theta_1, \dots, \theta_k) &= m_2 \\ & & \vdots & & & \\ m_k &= \frac{1}{n} \sum_{i=1}^n x_i^k & \mu_k &= E[X^k] & \mu_k(\theta_1, \dots, \theta_k) &= m_k \end{aligned}$$

- Solve resulting system of equations.



Method of Moments

➤ Example 1:

➤ Let x_1, x_2, \dots, x_n be iid samples from binomial(k, p),
i.e.:

$$p(x_i = x | k, p) = \binom{k}{x} p^x (1 - p)^{k-x}$$

Find an estimation of k, p .



Method of Moments

- Example 1:
 - Let x_1, x_2, \dots, x_n be iid samples from Binomial(k, p), i.e.:

$$p(x_i = x | k, p) = \binom{k}{x} p^x (1-p)^{k-x}$$

$$\begin{cases} \frac{1}{n} \sum x_i = \bar{X} = E(X) = kp \\ \frac{1}{n} \sum x_i^2 = E(X^2) = kp(1-p) + k^2 p^2 \end{cases} \Rightarrow$$

$$\tilde{k} = \frac{\bar{X}^2}{\bar{X} - \left(\frac{1}{n}\right) \sum (x_i - \bar{X})^2}$$

$$\tilde{p} = \frac{\bar{X}}{\tilde{k}}$$



Maximum Likelihood Estimation (MLE)

- Likelihood function:

$$L(\theta|x) = L(\theta_1, \dots, \theta_k|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta_1, \dots, \theta_k)$$

- For each sample point x , let $\hat{\theta}(x)$ be the parameter value at which $L(\theta|x)$ attains its maximum as a function of θ .
- The MLE estimator of θ based on a sample x is $\hat{\theta}(x)$.
- The MLE is the parameter point for which the observed sample is more likely.



Maximum Likelihood Estimation, Cont'd

$$\hat{\theta} = \operatorname{argmax}_{\theta} L(X|\theta)$$

- If the likelihood function is differentiable (in θ_i), possible conditions for the MLE are the values $(\theta_1, \dots, \theta_k)$ that solve:

$$\frac{\partial}{\partial \theta_i} L(\theta|x) = 0, i = 1, \dots, k$$

- Note that the solutions are possible candidates. To find exact MLE we should check

$$\frac{d^2}{d\theta^2} L(\theta|x) \Big|_{\theta=\bar{x}} < 0$$



Maximum Likelihood Estimation

- Example 2:
 - Let x_1, x_2, \dots, x_n be iid samples from $N(\theta, 1)$.
Find and MLE of θ .



Maximum Likelihood Estimation

➤ Example 2:

$$L(\theta|X) = \prod_{i=1}^n \frac{1}{(2\pi)^{.5}} e^{-\left(\frac{1}{2}\right)(x_i-\theta)^2} =$$
$$\frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\left(\frac{1}{2}\right) \sum_{i=1}^n (x_i-\theta)^2}$$
$$\frac{d}{d\theta} L(\theta|X) = 0 \Rightarrow \sum_{i=1}^n (x_i - \theta) = 0$$
$$\Rightarrow \tilde{\theta} = \bar{X}$$



Maximum Likelihood Estimation

- Sometimes it's more convenient to use log likelihood.
- Let x_1, x_2, \dots, x_n be iid samples from Bernouli(p), then the likelihood function is:

$$L(p|X) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$
$$\log L(p|X) = \sum x_i \log p + (n - \sum x_i) \log(1-p)$$

$$\frac{d}{dp} \log L(p|X) = 0 \implies \tilde{p} = \frac{\sum x_i}{n}$$



Maximum Likelihood Estimation, Cont'd

➤ Invariance property of MLE:

If $\hat{\theta}$ is the MLE, then for any function $\tau(\theta)$ the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.



Maximum Likelihood Estimation, Cont'd

➤ To use two-variate calculus to verify that a function $H(\theta_1, \theta_2)$ has a maximum at $(\hat{\theta}_1, \hat{\theta}_2)$, it must be shown that the following three conditions hold:

a) First order partial derivatives are zero:

$$\frac{\partial}{\partial \theta_1} H(\hat{\theta}_1, \hat{\theta}_2) = \frac{\partial}{\partial \theta_2} H(\hat{\theta}_1, \hat{\theta}_2) = 0$$

b) At least one second order partial derivative is negative:

$$\frac{\partial^2}{\partial \theta_1^2} H(\hat{\theta}_1, \hat{\theta}_2) < 0 \quad \text{or} \quad \frac{\partial^2}{\partial \theta_2^2} H(\hat{\theta}_1, \hat{\theta}_2) < 0$$

c) The Jacobian of second order derivatives is positive:

$$\begin{vmatrix} \frac{\partial^2}{\partial \theta_1^2} H(\hat{\theta}_1, \hat{\theta}_2) & \frac{\partial^2}{\partial \theta_1 \partial \theta_2} H(\hat{\theta}_1, \hat{\theta}_2) \\ \frac{\partial^2}{\partial \theta_2 \partial \theta_1} H(\hat{\theta}_1, \hat{\theta}_2) & \frac{\partial^2}{\partial \theta_2^2} H(\hat{\theta}_1, \hat{\theta}_2) \end{vmatrix} > 0$$



Bayes Estimators

- Suppose that we have a prior distribution for θ : $\pi(\theta)$
- Let $f(x|\theta)$ be the sampling distribution, then conditional distribution of θ given the sample x is:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}$$

where $m(x)$ is the marginal distribution of x :

$$m(x) = \int f(x|\theta) \pi(\theta) d\theta$$

