IE605: Engineering Statistics

Lecture 12

Manjesh K. Hanawal

Previous Lecture:

- Likelihood functions

This Lecture:

- Method of moments
- Bayes method
- Evaluating Estimators
- Cramer's Rao bound

Other Point Estimators:

- Maximum Likelihood estimator
- Method of moments
- Bayes method
- Expectation Maximization (EM) method

Method of Moments (MM)

- Method of Moments(MM) is one the oldest method for finding point estimator (since 1800!)
- MM estimators are found by equating the first k sample moments to the corresponding population moments
- $\boldsymbol{X}$ be a sample from pmf/pdf $f(\boldsymbol{x}|(\theta_1, \theta_2, \ldots, \theta_k))$

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i^1, \qquad \mu_1' = \mathbb{E}(X^1)$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \qquad \mu_2' = \mathbb{E}(X^2)$$

$$\vdots$$

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \qquad \mu_k' = \mathbb{E}(X^k)$$

Usually, μ_j'S will be function of $(\theta_1, \theta_2, \ldots, \theta_k)$, say $\mu_j'(\theta_1, \theta_2, \ldots, \theta_k)$

Solving MMs

- Obtain k equations by equating

$$
\begin{aligned}
m_1 &= \mu_1'(\theta_1, \theta_2, \ldots, \theta_k) \\
m_2 &= \mu_2'(\theta_1, \theta_2, \ldots, \theta_k) \\
&\vdots \\
m_k &= \mu_k'(\theta_1, \theta_2, \ldots, \theta_k)
\end{aligned}
$$

Normal Method of Moments: $X = (X_1, X_2, \ldots, X_n)$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$. $\boldsymbol{\theta} = (\theta_1, \theta_2) = (\mu, \sigma^2)$.

- $m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ (first moment), $m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ (second moment)
- $\mu_1'(\theta_1, \theta_2) = \mu$ (mean), $\mu_2'(\theta_1, \theta_2) = \mu^2 + \sigma^2$ (second moment).
- $$\frac{\sum_{i=1}^{n} X_i}{n} = \theta_1 \quad \text{and} \quad \frac{\sum_{i=1}^{n} X_i^2}{n} = \theta_1^2 + \theta_2.$$

Example: Binomial method of moments

$\boldsymbol{X} = (X_1, X_2 \ldots, X_n)$ are iid, $X_i \sim Bin(k, p)$. k & p are unknown.

Bayes Estimators

- In classical approach θ is unknown but fixed
- Based on the observed sample $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ drawn from population with parameter θ, we obtain knowledge of θ
- In **Bayesian approach** θ is assumed to be random quantity drawn from a distribution known as **prior distribution**
- Prior distribution is a subjective belief formulated before data are seen (hence prior distribution)
- When sample is observed from population, the prior distribution is updated. The udate is called **Posterior distribution**

Baye's method

- Denote prior distributions as $P(\theta)$
- Conditional pmf of samples under θ as $P(\boldsymbol{x}|\theta)$
- Joint pmf of samples as $P(\boldsymbol{x})$
- Conditional distibution of θ given $\boldsymbol{x}$ as $P(\theta|\boldsymbol{x})$

By Baye's formula:

$$P(\theta|\boldsymbol{x})p(\boldsymbol{x}) = P(\boldsymbol{x}|\theta)P(\theta)$$

$$\implies P(\theta|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|\theta)p(\theta)}{P(\boldsymbol{x})}$$

$$\implies P(\theta|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|\theta)P(\theta)}{\int P(\boldsymbol{x}|\theta)P(\theta)d\theta}$$

For continuous case replace pmf $p(\boldsymbol{x}|\theta)$ by pdf $f(\boldsymbol{x}|\theta)$.

Example: Binomial Bayes Estimator

$Y \sim Bin(n, p)$. Prior distribution of $p \sim Beta(\alpha, \beta)$. $Y = y$ is observed. Find posterior, i.e., $P(p|y) = \frac{P(y|p)P(p)}{P(y)}$

$$\begin{aligned} P(y, p) =& P(y|p)P(p) \\ =& \left[\binom{n}{y} p^y (1-p)^{n-y}\right] \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\right] \\ =& \left[\binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1}\right] \end{aligned}$$

$$\begin{aligned} P(y) =& \int_0^1 P(y, p) dp \\ =& \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)} \end{aligned}$$

Binomial Bayes Estimator

$$
\begin{aligned}
P(p|y) &= \frac{P(y|p)P(p)}{P(y)} \\
&= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1} \\
&\sim \; Beta(y+\alpha, n-y+\beta).
\end{aligned}
$$

- Natural estimator is mean of posterior distribution

$$\hat{p} = \frac{y+\alpha}{\alpha+\beta+n}$$

- Estimator from prior distribution is $\frac{\alpha}{(\alpha+\beta)}$. From samples is $\frac{y}{n}$.

$$\hat{p} = \left(\frac{n}{\alpha+\beta+n}\right)\left(\frac{y}{n}\right) + \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\left(\frac{\alpha}{\alpha+\beta}\right)$$

Conjugate family

Let $\mathcal{F}$ denote a familly of pmf/pdf. A class Π of prior distribution is **conjugate family** for $\mathcal{F}$ if any prior distribution for a pdf/pmf $f \in \mathcal{F}$ from Π also results in a posterior in Π.

- Beta family is a conjugate for Binomial family
- Normal family is a conjugate for Normal family with unknown mean and known variance.

Evaluating Estimators

- Different method can give different estimator
- Basic criteria for evaluation of estimators

> Mean Squared Error (MSE) of an estimator W of a parameter θ is a function of θ defined by $\mathbb{E}_\theta(W-\theta)^2$

- Any increasing function of $|W-\theta|$ would serve as a measure of goodness of an estimator
- MSE has two advantages: It is tractable and it has good interpretation

$$\begin{aligned}\mathbb{E}_\theta(W-\theta)^2 &= Var_\theta W + (\mathbb{E}_\theta W-\theta)^2\\ &= Var_\theta W + (Bias_\theta W)^2\end{aligned}$$

Unbiased estimator

An estimator W for parameter θ is called unbiased estimator if $\mathbb{E}_\theta W = \theta$ for all θ, i.e., $Bias_\theta W = 0$.

Cramer-Rao's Bound

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a sample with pdf $f(\boldsymbol{x}|\theta)$, and let $W(\boldsymbol{X})$ be any estimator such that $Var_\theta W(\boldsymbol{X}) < \infty$. Then

$$Var_\theta W(\boldsymbol{X}) \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta W(\boldsymbol{X})\right)^2}{\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta)\right)^2\right)}$$