Lecture 6

Casella and Berger

Sections 7.2.3, 7.2.4, 7.3.1

Example 6.1 (MLE in linear regression models): Suppose that random variables $Y_1, ..., Y_n$ satisfy

$$Y_i = \boldsymbol{\beta}' \mathbf{X}_i + \epsilon_i, i = 1, ..., n$$

where $\mathbf{X}_1, ..., \mathbf{X}_n$ are p-dimensional covariates, and $\epsilon_1, ..., \epsilon_n$ are i.i.d. $\sim N(0, \sigma^2)$, σ^2 is unknown. Find the MLE of $\boldsymbol{\beta}$ and σ^2, and the distribution of $\widehat{\boldsymbol{\beta}}$.

Example 6.2: Suppose that $X_1, ..., X_n$ are i.i.d. following the distribution

$$P(X_i \le x|\alpha,\beta) = \begin{cases} 0 & \text{if} x < 0 \\ (x/\beta)^\alpha & \text{if} 0 \le \text{x} \le \beta \\ 1 & \text{if} x > \beta \end{cases}$$

where α and β are positive.

(a) Find a two-dimensional sufficient statistic for (α, β).

(b) Fine the MLEs of α and β.

Bayes Estimators

Fundamental difference between the Bayesian approach and the frequentist approach: parameters θ are also random and follow a prior distribution $\pi(\theta)$ in the Bayesian approach.

Sampling distribution: $f(\mathbf{x}|\theta)$

Prior distribution: $\pi(\theta)$

Posterior distribution (conditional distribution of θ given the sample $\mathbf{x}$):

$$\pi(\theta|\mathbf{x}) = f(\mathbf{x}|\boldsymbol{\theta})\pi(\theta)/m(\mathbf{x}),$$

where $m(\mathbf{x})$ is the marginal distribution of $\mathbf{X}$, that is,

$$m(\mathbf{x}) = \int f(\mathbf{x}|\theta)\pi(\theta)d\theta.$$

The posterior distribution is used to make statements about θ, which is still considered a random quantity. The mean of posterior distribution can be used as a point estimate of θ.

Example 6.3 (Binomial Bayes estimation): Let $X_1, ..., X_n$ be i.i.d. Bernoulli(p). We assume the prior distribution on p is beta(α, β). Find a Bayes estimator of p.

Definition 6.1: Let $\mathcal{F}$ denote the class of pdfs or pmfs $f(x|\theta)$ (indexed by θ). A class Π of prior distributions is a *conjugate family* for $\mathcal{F}$ if the posterior distribution is in the class Π for all $f \in \mathcal{F}$, all priors in Π, and all $x \in \mathcal{X}$.

Example 6.4 (Normal Bayes estimators): Let $X \sim N(\theta, \sigma^2)$, and suppose that the prior distribution on θ is $N(\mu, \tau^2)$. Here we assume that σ^2, μ, and τ^2 are all known. Find a Bayes estimator of θ.

Expectation-Maximization (EM) Algorithm

The idea of EM-algorithm is to replace one difficult likelihood maximization with a sequence of easier maximizations whose limit is the answer to the original problem. It is particularly suited to "missing data" problems.

Complete data: $(\mathbf{Y}, \mathbf{X}) \sim f(\mathbf{y}, \mathbf{x}|\theta)$

Observed data or incomplete data: $\mathbf{Y} \sim f(\mathbf{y}|\theta)$

Objective: estimate θ based on the observable data

Approaches:

a). Direct maximization of $L(\theta|\mathbf{y}) = f(\mathbf{y}|\theta)$

b) EM-algorithm: maximize $E[\log L(\theta|\mathbf{y}, \mathbf{x})|\widehat{\theta}^{(k)}, \mathbf{y}]$ based on the current estimate of θ, and obtain $\widehat{\theta}^{(k+1)}$. We can show that $L(\widehat{\theta}^{(k+1)}|\mathbf{y}) \geq L(\widehat{\theta}^{(k)}|\mathbf{y})$.

Example 6.5 (ABO blood type): In a population of n individuals, the number of individuals with blood types A, B, AB, and O are n_A, n_B, n_{AB}, and n_O, respectively. Assume random mating in this population, i.e., $p_{AB} = p_A p_B, p_{AO} = p_A p_O, ..., p_{OO} = p_O p_O$. Estimate the frequencies of alleles A, B, and O in this population.

Example 6.6 (Multiple Poisson rates): We observe $X_1, ..., X_n$ and $Y_1, ..., Y_n$, all mutually independent, where $Y_i \sim Poisson(\beta\tau_i)$ and $X_i \sim Poisson(\tau_i)$. This would model, for instance, the incidence of a disease, Y_i, where the underlying rate is a function of an overall effect β and an additional factor τ_i. We do not see τ_i but get information on it through X_i. Suppose that the value of x_1 was missing, describe the EM-algorithm for estimating β and $\tau_i, i = 1, ..., n$.

Methods of Evaluating Estimators

Definition 6.2 (Mean Squared Error) The *mean squared error* (MSE) of an estimator W of a parameter θ is the function of θ defined by $E_\theta(W-\theta)^2$.

Definition 6.3 The *bias* of a point estimator W of a parameter θ is the difference between the expected value of W and θ; that is, $\text{Bias}_\theta W = E_\theta W - \theta$. An estimator is called unbiased if $E_\theta W = \theta$ for all θ.

Example 6.7 (Normal MSE) Let $X_1, ..., X_n$ be iid $N(\mu, \sigma^2)$.
(1).Calculate the MSEs for $\bar{X}$ and S^2 as estimators of μ and σ^2.
(2). Calculate the MSE of the MLE of σ^2.