More on MLE; Bayes Estimators

Tiandong Wang

Department of Statistics
Texas A&M University

Restricted MLE

Special attention is needed to make sure $\widehat{\theta} \in \Theta$

- Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\theta, 1)$, where $\theta \geq 0$.
- Example: $X_1, \cdots, X_n \overset{iid}{\sim} Bin(1, p)$, where $0 \leq p \leq 1$.
- Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where $\mu \geq 0$.

Other cases: monotone likelihood

- Example: $X_1, \cdots, X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$, where $\theta > 0$.
- Example: $X_1, \cdots, X_n \overset{iid}{\sim}$ exponential location family with pdf

$$f(x) = \exp^{-(x-\theta)}, \quad \text{if} \quad x \geq \theta$$

- Example: $X_1, \cdots, X_n \overset{iid}{\sim} \text{Unif}(\theta - 1/2, \theta + 1/2)$
- Example: Let X by a single observation taking values from $\{0, 1, 2\}$ according to P_θ, where $\theta = \theta_0$ or θ_1. The probability of X is summarized

	$x = 0$	$x = 1$	$x = 2$
$\theta = \theta_0$	.8	.1	.1
$\theta = \theta_1$	.2	.3	.5

Practical issues with MLE

- Multimodality:
  - Likelihood function can be **multimodal**, often have to use numerical techniques to try to maximize (no closed-form max).
  - Can get stuck in local modes.
  - **Solutions**:
    1. Choose models such that L is convex in θ.
    2. Heuristic search, multiple starting points.
    3. Satisfied with a local maximum.
- Flatness and sensitivity:
  - $L(\theta; \mathbf{x})$ can be pretty flat near the max.
  - So a slightly different sample $\mathbf{x}$ may give a very different MLE.

Remarks on the MLE:

- The MLE $\widehat{\theta}(\mathbf{x})$ is the value for which the observed sample $\mathbf{x}$ is most likely; possess some optimal properties (Chapter 10)
- The MLE can be numerically sensitive to the variation in the data. Example: $Bin(k, p)$.
- If T is (minimal) sufficient for θ, then the MLE $\widehat{\theta}$ must be a function of T. By factorization theorem, we have

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}),$$

  and the MLE $\widehat{\theta}$ should maximize $g(T(\mathbf{X})|\theta)$. Therefore, the MLE is a function of the (minimal) sufficient statistic.

Invariance Property of MLE

Theorem

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

Examples:

- Example: $X_1, \cdots, X_n \overset{iid}{\sim} Bin(1, p)$. Find the MLE of $\sqrt{p(1-p)}$
- Example: $X_1, \cdots, X_n \overset{iid}{\sim} Poi(\lambda)$. Find the MLE of $P(X \leq 1)$
- Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Find the MLE of μ/σ
- Example: Find the MLE of the population median
- Find the MLE for c such that $P(\bar{X} > c) = 0.025$. (the 97.5% percentile of the distribution of $\bar{X}$.

Bayes Estimators

- Different from classical approaches, in the Bayesian approach θ is considered as a random quantity whose variation can be described by a probability distribution (called the prior distribution).
- A sample is then taken from a population indexed by θ and the prior distribution is updated with this sample information. The updated prior is called the posterior distribution.

Important Concepts in Bayes Estimation

- prior distribution of θ: $\theta \sim \pi(\theta)$
- sampling distribution of $\mathbf{y}$ given θ: $\mathbf{y}|\theta \sim f(\mathbf{y}|\theta)$
- posterior distribution of θ: $\pi(\theta|\mathbf{y}) = f(\mathbf{y}|\theta)\pi(\theta)/m(\mathbf{x})$
- marginal distribution of $\mathbf{y}$: $m(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta)d\theta$
- posterior mean of θ: $E(\theta|\mathbf{y}) = \int \theta\pi(\theta|\mathbf{y})d\theta$ (Bayes estimator of θ)

Bayes Estimation: Examples

- Example: Assume $X_1, \cdots, X_n$ iid $Bin(1, p)$. Assume the prior distribution on p is $Beta(\alpha, \beta)$ with known parameters (α, β). Find the posterior distribution of p and the Bayes estimator of p
  Special case: $\pi(p) \sim Unif(0, 1)$
- Remark: If $T(\mathbf{x})$ is a sufficient statistic, then the posterior density of θ is $\pi(\theta|\mathbf{x}) = \pi(\theta|T(\mathbf{x}))$

Bayes Estimation: Conjugate family

Let $\mathcal{F}$ denote the class of pdfs or pmfs $f(x|\theta)$. 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 $x \in \mathcal{X}$

Examples:

- Example:(Beta-Binomial Conjugate)
- Example:(Gamma-Poisson Conjugate)
- Example: (Normal-Normal Conjugate)
  Let $X_1, \cdots, X_n$ be iid $\sim N(\theta, \sigma^2)$, with θ unknown and σ^2 known. Suppose that the prior distribution of θ is $N(\mu, \tau^2)$. Here we assume both μ and τ^2 are given. Find the posterior distribution of θ.

Bayesians vs.Frequentists