Review for Midterm II - STAT 411

Chap 6. Maximum Likelihood Methods

§ 6.4 Multiparameter Case: Estimation

- Parameters: Let $X_1, \dots, X_n$ be iid $\sim f(x; \boldsymbol{\theta})$, where $\theta = (\theta_1, \dots, \theta_p)'$.
  Likelihood function: $L(\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i; \boldsymbol{\theta})$;
  Log likelihood function: $l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(x_i; \boldsymbol{\theta})$.
  MLE: The value of θ which maximizes $L(\boldsymbol{\theta})$ or $l(\boldsymbol{\theta})$ is called the maximum likelihood estimator (mle) of θ and denoted by $\hat{\boldsymbol{\theta}}$.
- Under regularity conditions, the gradient is

$$\nabla \log f(x, \theta) = \left(\frac{\partial}{\partial \theta_1} \log f(X; \boldsymbol{\theta}), \cdots, \frac{\partial}{\partial \theta p} \log f(X; \boldsymbol{\theta}) \right)$$

$$E\left[\nabla \log f(x, \theta)\right] = 0, E\left[\frac{\partial}{\partial \theta_j} \log f(X; \boldsymbol{\theta})\right] = 0, \quad \text{for } j = 1, \dots, p\,;$$

$$\text{Cov}\left(\frac{\partial}{\partial \theta_j} \log f(X; \boldsymbol{\theta}),\ \frac{\partial}{\partial \theta_k} \log f(X; \boldsymbol{\theta})\right) = -E\left[\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log f(X; \boldsymbol{\theta})\right]$$

- Fisher information matrix:

$$I(\boldsymbol{\theta}) = Cov\left(\nabla \log f(x, \theta)\right) = (I_{jk})_{p \times p},$$

  where

$$I_{jk} = -E\left[\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log f(X; \boldsymbol{\theta})\right]$$

- Theorem: Under regularity conditions, if $Y = u(X_1, \dots, X_n)$ is an unbiased estimate of θ_j, then

$$Var(Y) \geq \frac{1}{n} [I^{-1}(\boldsymbol{\theta})]_{jj}$$

  .
- Theorem 6.4.1: Let $X_1, \dots, X_n$ be iid $\sim f(x; \boldsymbol{\theta})$. Under regularity conditions, any consistent solution sequence $\hat{\theta}_n$ of the likelihood equation $\frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}) = 0$

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{D} N_p(\mathbf{0},\ \mathbf{I}^{-1}(\boldsymbol{\theta}))$$

- Corollary. Let g be a transformation $g(\theta) = (g_1(\theta), \cdots, g_k(\theta))^T$, such that $1 \leq k \leq p$. and the $k \times p$ matrix of partial derivatives:

$$\mathbf{B} = \left[\frac{\partial g_i}{\partial \theta_j}\right], i = 1, \cdots, k; j = 1, \cdots, p$$

has continuous elements and does not vanish in a neighborhood of θ. Let $\hat{\eta} = g\left(\hat{\theta}\right)$, then $\hat{\eta}$ is the mle of $\eta = g\left(\theta\right)$, and

$$\sqrt{n}(\hat{\eta} - \eta) \xrightarrow{D} N\left(\mathbf{0}, \mathbf{B}\mathbf{I}^{-1}\left(\theta\right)\ \mathbf{B}^{T}\right)$$

and the information matrix for η is $I\left(\eta\right) = \left(\mathbf{B}\mathbf{I}^{-1}\left(\theta\right)\ \mathbf{B}^{T}\right)^{-1}$

Chap 7. Sufficiency

§ 7.1 Measures of Quality of Estimators

- **MVUE**: For fixed n and a given sample $X_1, \ldots, X_n$, the statistic $Y = u(X_1, \ldots, X_n)$ is called a *minimum variance unbiased estimator* (MVUE) of θ, if $E(Y) = \theta$ and if $\mathrm{Var}(Y) \leq \mathrm{Var}(Z)$ for every other unbiased estimator Z of θ.

§ 7.2 A Sufficient Statistic for a Parameter

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. from a distribution that has pdf or pmf $f(x;\theta)$, $\theta \in \Omega$. Let $Y_1 = u(X_1, \ldots, X_n)$ be a statistic that has pdf or pmf $f_{Y_1}(y;\theta)$.
- **Sufficient Statistic**: Y_1 is a *sufficient statistic* for θ if and only if the joint conditional distribution of $X_1, \ldots, X_n$ given Y_1 does not depend on θ . In other words, $\left[\prod_{i=1}^{n} f(x_i;\theta)\right] / f_{Y_1}(u(x_1, \ldots, x_n);\theta) = H(x_1, \ldots, x_n)$, which does not depend on θ .
- **Factorization Theorem** (Neyman): Y_1 is a sufficient statistic for θ if and only if $\prod_{i=1}^{n} f(x_i;\theta) = k_1[u(x_1, \ldots, x_n);\theta] \cdot k_2(x_1, \ldots, x_n)$ for some nonnegative functions k_1 and k_2 .

§ 7.3 Properties of a Sufficient Statistic

- Let $X_1, X_2, \ldots, X_n$ be i.i.d.$\sim f(x;\theta)$, $\theta \in \Omega$.
- **Rao-Blackwell Theorem**: Let $Y_1 = u_1(X_1, \ldots, X_n)$ be a sufficient statistic for θ, and let $Y_2 = u_2(X_1, \ldots, X_n)$ be an unbiased estimator of θ . Then $E(Y_2|Y_1) = \varphi(Y_1)$ is another unbiased estimator of θ whose variance is less than that of Y_2 .
  Corollary: Any MVUE of θ must be a function of the sufficient statistic.
- Theorem: If $Y_1 = u_1(X_1, \ldots, X_n)$ is sufficient for θ and the mle $\hat{\theta}$ of θ exists uniquely, then $\hat{\theta}$ must be a function of Y_1 .

§ 7.4 Completeness and Uniqueness

- **Completeness**: Let $Y_1 \sim f(y;\theta)$, $\theta \in \Omega$. Suppose the condition $E[u(Y_1)] = 0$ for all θ always implies that $u(y) \equiv 0$ except on a zero-probability set. Then $\{f(y;\theta) : \theta \in \Omega\}$ is called a *complete family* of probability functions and Y_1 is said to be *complete* for $\theta \in \Omega$.
  Note: The completeness of Y_1 is to guarantee the uniqueness of the unbiased estimator of θ among the functions of Y_1 .

- **Theorem (Lehmann-Scheffé)**: Let $X_1, X_2, \ldots, X_n$ be i.i.d.$\sim f(x;\theta)$, $\theta \in \Omega$. Let $Y_1 = u(X_1, \ldots, X_n)$ be a complete sufficient statistic for θ . If a function $\varphi(Y_1)$ of Y_1 is an unbiased estimator of θ, then $\varphi(Y_1)$ must be the unique MVUE of θ .

§ 7.5 The Exponential Class of Distributions

- **Exponential class**: A family $\{f(x;\theta) : \theta \in (\gamma, \delta) \subset \mathbb{R}\}$ of pdfs or pmfs of the form $f(x;\theta) = \exp\{p(\theta)K(x) + S(x) + q(\theta)\}$, $x \in \mathcal{S}$.

- **Regular exponential class**: A member of exponential class satisfies
  (1) $\mathcal{S}$ does not depend on θ;
  (2) $p(\theta)$ is nontrivial and continuous;
  (3.1) if X is continuous then $K'(x)$ and $S(x)$ are continuous, where $K'(x)$ is not always 0; (3.2) if X is discrete then $K(x)$ is nontrivial.

- Examples of regular exponential class: beta, gamma (exponential, chi-square), normal, binomial (Bernoulli), geometric, negative binomial, Poisson

- Theorem: Let $X_1, \ldots, X_n$ be i.i.d.$\sim f(x;\theta)$, $\theta \in \Omega$, which belongs to the regular exponential class. Let $Y_1 = \sum_{i=1}^n K(X_i)$. Then
  (1) $Y_1 \sim f_{Y_1}(y;\theta) = R(y)\exp[p(\theta)y + nq(\theta)]$, for $y \in \mathcal{S}_{Y_1}$ and some positive function $R(y)$. Neither $\mathcal{S}_{Y_1}$ nor $R(y)$ depends on θ .
  (2) $E(Y_1) = -nq'(\theta)/p'(\theta)$.
  (3) $Var(Y_1) = n\left[p''(\theta)q'(\theta) - q''(\theta)p'(\theta)\right]/(p')^3$.
  (4) Y_1 is a complete sufficient statistic for θ .

- Theorem: Let Y be a complete sufficient statistic for θ and $g(Y)$ be a one-to-one function of Y, then $g(Y)$ is also a complete sufficient statistic for θ.

§ 7.6 Functions of a Parameter

- Suppose Y is complete sufficient for θ. Let $\eta = g(\theta)$ is the parameter of interest and $T = T(Y)$ is an unbiased estimator of η. Then T is the MVUE of η.

- If Y is an mle, then $T(Y)$ can be constructed on Y by the functional invariance of mle.
- Statistic $T(Y)$ also can be obtained by the conditional expectation of an unbiased estimator of $g(\theta)$ given the sufficient statistic Y (Rao-Blackwell Thm and Lehmann and Scheffé Thm).

Practice Problems

- Chapter 6:
  Exercise § 6.4 - 3, 11; Example § 6. 4 - 3, 4
- Chapter 7:
  Exercise § 7.2 - 9; § 7.3 - 4; § 7.4 - 4, 6, 7; § 7.5 - 6, 10; § 7.6 - 9 .