Review for Midterm-II

Chapter 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 $\text{Var}(Y) \leq \text{Var}(Z)$ for every other unbiased estimator Z of θ.

- **Loss function**: Let $\delta(Y)$ be a point estimator of θ ($\delta(y)$ is called a *decision function* or a *decision rule*). A nonnegative function $\mathcal{L}[\theta, \delta(y)]$ that reflects the seriousness of the difference between $\delta(y)$ and θ is called the *loss function.* For examples, $\mathcal{L}(\theta, \delta) = (\theta - \delta)^2$ (squared-error loss function), $\mathcal{L}(\theta, \delta) = |\theta - \delta|$ (absolute-error loss function).

- **Risk function**: $R(\theta, \delta) = E(\mathcal{L}[\theta, \delta(Y)])$

- **Minimax principle**: $\delta_0(y)$ is called a *minimax decision function* if $\max_\theta R(\theta, \delta_0) \leq \max_\theta R(\theta, \delta)$ for every other decision function $\delta(y)$.

- **Likelihood principle**: Suppose two different sets of data from possible two different random experiments lead to respective likelihood functions, $L_1(\theta)$ and $L_2(\theta)$. If $L_1(\theta)$ and $L_2(\theta)$ are proportional to each other, then a statistician should obtain the same estimate of θ from either.

§ 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 .

- For practice: Example 7.2.1, Example 7.2.5, Example 7.2.6

§ 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 .
- For practice: Example 7.3.1

§ 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 and 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 .
- For practice: Example 7.4.1

§ 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 $X_1, \dots, X_n$ 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) $Var(Y_1) = n\left[p''(\theta)q'(\theta) - q''(\theta)p'(\theta)\right]/[p'(\theta)^3]$.
(4) Y_1 is a complete sufficient statistic for θ .

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

- For practice: Example 7.5.1, Example 7.5.2

§ 7.6 Functions of a Parameter

- Suppose $Y = \hat{\theta}$ is complete sufficient for θ. Let $\delta = g(\theta)$ is the parameter of interest and $T = T(Y)$ is an unbiased estimator of δ. Then T is teh MVUE of δ.

- 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).

§ 7.7 The Case of Several Parameters

- Let $X_1, \dots, X_n$ be i.i.d.$\sim f(x;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Omega \subset R^p$, $x \in \mathcal{S}$.
Let $\mathbf{Y} = (Y_1, \dots, Y_m)' \sim f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})$,
where $Y_i = u_i(X_1, \dots, X_n)$, $i = 1, \dots, m$.

- **Joint sufficiency**: $\mathbf{Y}$ is said to be *jointly sufficient* for $\boldsymbol{\theta}$ if and only if $\left[\prod_{i=1}^n f(x_i;\boldsymbol{\theta})\right]/f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) = H(x_1, \dots, x_n)$ does not depend on $\boldsymbol{\theta}$.

- **Extended factorization theorem**: $\mathbf{Y}$ is jointly sufficient for $\boldsymbol{\theta}$ if and only if $\prod_{i=1}^n f(x_i;\boldsymbol{\theta}) = k_1(\mathbf{y};\boldsymbol{\theta}) \cdot k_2(x_1, \dots, x_n)$ for some nonnegative functions k_1 and k_2 .

- **Completeness** (case of several parameters): Suppose the condition $E[u(Y_1, \dots, Y_m)] = 0$ for all $\boldsymbol{\theta} \in \Omega$ always implies that $u(y_1, \dots, y_m) \equiv 0$ except on a zero-probability set. Then $\mathbf{Y} = (Y_1, \dots, Y_m)'$ is said to be complete for $\boldsymbol{\theta}$.

- **Extended theorem (Lehmann and Scheffé)**: Suppose $\mathbf{Y}$ is jointly complete and sufficient for $\boldsymbol{\theta}$. Let $\delta = g(\boldsymbol{\theta})$ is the parameter of interest and $T = T(\mathbf{Y})$ is an unbiased estimator of δ. Then T is the unique MVUE of δ .

- **Regular exponential class (case of several parameters)**: Let $X \sim f(x;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Omega \subset R^p$. Suppose $f(x;\boldsymbol{\theta}) = \exp\left\{\sum_{j=1}^{m} p_j(\boldsymbol{\theta})K_j(x) + S(x) + q(\boldsymbol{\theta})\right\}$, $x \in \mathcal{S}$. We say that it is a member of the *regular exponential class* if
(1) $p = m$, and $\mathcal{S}$ does not depend on $\boldsymbol{\theta}$;
(2) Ω contains a nonempty, m-dimensional open rectangle;
(3) $p_j(\boldsymbol{\theta})$, $j = 1, \ldots, m$ are nontrivial, functionally independent, continuous functions of $\boldsymbol{\theta}$;
(4.1) If X is continuous, then $K_j'(x)$'s are continuous and no one is a linear homogeneous function of the others, and $S(x)$ is continuous;
(4.2) If X is discrete, then $K_j(x)$'s are nontrivial and no one is a linear homogeneous function of the others.

- **Theorem (regular exponential class)**: Let $X_1, \ldots, X_n$ be i.i.d.$\sim f(x;\boldsymbol{\theta})$, which belongs to the regular exponential class. Let $\mathbf{Y} = (Y_1, \ldots, Y_m)'$, where $Y_j = \sum_{i=1}^{n} K_j(X_i)$, $j = 1, \ldots, m$. Then
(1) $\mathbf{Y} \sim f(\mathbf{y};\boldsymbol{\theta}) = R(\mathbf{y}) \exp\left\{\sum_{j=1}^{m} p_j(\boldsymbol{\theta})y_j + nq(\boldsymbol{\theta})\right\}$. Neither the support of $\mathbf{Y}$ nor $R(\mathbf{y})$ depends on $\boldsymbol{\theta}$.
(2) $Y_1, \ldots, Y_m$ are joint complete sufficient statistics for $\boldsymbol{\theta}$, if $n > m$.

- Theorem: Let $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ be joint complete sufficient statistics for $\boldsymbol{\theta}$ and $\mathbf{g}(\mathbf{Y}) = (g_1(\mathbf{Y}), \ldots, g_m(\mathbf{Y}))'$ is a one-to-one mapping of $\mathbf{Y}$. Then $(g_1(\mathbf{Y}), \ldots, g_m(\mathbf{Y}))$ are also joint complete sufficient statistics for $\boldsymbol{\theta}$.

- **Regular exponential class (k-dimensional random vector)**: Let $\mathbf{X}$ be a k-dimensional random vector with pdf or pmf $f(\mathbf{x};\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Omega \subset R^p$. Suppose $f(\mathbf{x};\boldsymbol{\theta}) = \exp\left\{\sum_{j=1}^{m} p_j(\boldsymbol{\theta})K_j(\mathbf{x}) + S(\mathbf{x}) + q(\boldsymbol{\theta})\right\}$, $\mathbf{x} \in \mathcal{S} \subset \mathbb{R}^{\urcorner}$.
We say that $f(\mathbf{x};\boldsymbol{\theta})$ is a member of the *regular exponential class* if
(1) $p = m$; (2) $\mathcal{S}$ does not depend on $\boldsymbol{\theta}$; and (3) the regularity conditions similar to those of one-dimensional case hold.

- **Theorem (k-dimensional regular exponential class)**: Suppose $\mathbf{X}$ is a k-dimensional random vector with pdf or pmf $f(\mathbf{x};\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^{\gtrdot}$, which belongs to the regular exponential class. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from $\mathbf{X}$ and let $\mathbf{Y} = (Y_1, \ldots, Y_m)'$, where $Y_j = \sum_{i=1}^{n} K_j(\mathbf{X}_i)$, $j = 1, \ldots, m$. Then
(1) $(Y_1, \ldots, Y_m)$ are joint complete sufficient statistics for $\boldsymbol{\theta} \in \Omega$.
(2) Let $\delta = g(\boldsymbol{\theta})$ be the parameter of interest and $T = h(\mathbf{Y})$ is an unbiased estimator of δ. Then T is the unique MVUE of δ .

- For practice: Example 7.7.2, Example 7.7.3, Example 7.7.5

§ 7.8 Minimal Sufficiency and Ancillary Statistics

- Let $X_1, \ldots, X_n$ be i.i.d.$\sim f(x;\theta)$, $x \in \mathcal{S}$, $\theta \in \Omega$.

- **Minimal sufficient statistic**: A sufficient statistic Y is called a *minimal sufficient statistic* for θ if, for any other sufficient statistic T of θ, Y is a function of T .
  Note: If both Y_1 and Y_2 are minimal sufficient statistics for θ, then $Y_1 = g(Y_2)$ for some one-to-one function g .

- **Theorem (minimal sufficiency)**: Let $T = T(X_1, \ldots, X_n)$ be a statistic. Suppose $\prod_{i=1}^n [f(x_i;\theta)/f(z_i;\theta)]$ does not depend on θ if and only if $T(x_1, \ldots, x_n) = T(z_1, \ldots, z_n)$. Then T is a minimal sufficient statistic for θ .

- Theorems: (1) Suppose the mle $\hat{\theta}$ of θ is also sufficient for θ. Then $\hat{\theta}$ must be a minimal sufficient statistic for θ .
  (2) Suppose Y is a minimal sufficient statistic for θ and $g(Y)$ is a one-to-one function of Y. Then $g(Y)$ is also minimal sufficient for θ .

- **Theorem (Lehmann and Scheffé)**: If a complete sufficient statistic exists, it must be minimal sufficient.

- **Ancillary statistic**: A statistic whose distribution does not depend on the parameter θ is called an *ancillary statistic*.

- **Location model and location invariant statistics**: Let $W_1, \ldots, W_n$ be i.i.d. random variables with pdf $f(w)$ which does not depend on θ . Let $X_i = \theta + W_i$, $-\infty < \theta < \infty$, $i = 1, \ldots, n$, known as a *location model.* The common pdf of X_i is $f(x-\theta)$. Then $\{f(x-\theta) : -\infty < \theta < \infty\}$ is called a *location family.*
  Let $Z = u(X_1, \ldots, X_n)$ be a statistic such that $u(x_1 + d, \ldots, x_n + d) = u(x_1, \ldots, x_n)$ for all $d \in \mathbb{R}$. Then Z is a *location-invariant statistic* whose distribution does not depend on θ .
  Examples of location invariant statistics: sample variance S^2, sample range $\max_i\{X_i\} - \min_i\{X_i\}$.

- **Scale model and scale invariant statistics**: Let $W_1, \ldots, W_n$ be i.i.d. random variables with pdf $f(w)$ which does not depend on θ . Let $X_i = \theta W_i$, $\theta > 0$, $i = 1, \ldots, n$, known as a *scale model.* The common pdf of X_i is $f(x/\theta)/\theta$. Then $\{f(x/\theta)/\theta : \theta > 0\}$ is called a *scale family.*
  Let $Z = u(X_1, \ldots, X_n)$ be a statistic such that $u(cx_1, \ldots, cx_n) = u(x_1, \ldots, x_n)$ for all $c > 0$. Then Z is a *scale-invariant statistic* whose distribution does not depend on θ .
  Examples of scale invariant statistics: $\min_i\{X_i\}/\max_i\{x_i\}$, $X_1^2/\sum_{i=1}^n X_i^2$.

- **Location and scale invariant statistics**: Let $W_1, \ldots, W_n$ be i.i.d. random variables with pdf $f(w)$ which does not depend on θ . Let $X_i = \theta_1 + \theta_2 W_i$, $i = 1, \ldots, n$, known as a *location and scale model.* The common pdf of X_i is $f((x-\theta_1)/\theta_2)/\theta_2$. Then $\{f((x-\theta_1)/\theta_2)/\theta_2 : -\infty < \theta_1 < \infty, \theta_2 > 0\}$ is called a *location and scale family.*

Let $Z = u(X_1, \ldots, X_n)$ be a statistic such that $u(cx_1 + d, \ldots, cx_n + d) = u(x_1, \ldots, x_n)$ for all $c > 0, d \in \mathbb{R}$. Then Z is a *location and scale invariant statistic* whose distribution does not depend on θ .
Examples of location and scale invariant statistics:
$[\max_i\{X_i\} - \min_i\{X_i\}]/S$, $(X_1 - \bar{X})/S$.

- For practice: Example 7.8.1, Example 7.8.4

§ 7.9 Sufficiency, Completeness and Independence

- Let $X_1, \ldots, X_n$ be i.i.d.$\sim f(x;\theta)$, $\theta \in \Omega$.

- Theorem: Let Y_1 be a sufficient statistic for θ and let Z be another statistic which is independent of Y_1 . Then Z is an ancillary statistic.

- **Theorem (Basu's)**: Suppose Y_1 is complete and sufficient for $\theta \in \Omega$. Then Y_1 is independent of every ancillary statistic.
  Note: Basu's theorem allows us to deduce the independence of two statistics without even finding the joint distribution of the two statistics.

- For practice: Example 7.9.1, Example 7.9.5