Final Exam Review - Part III

Chapter 7. Sufficiency

§ 7.7 The Case of Several Parameters

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

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

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

- **Completeness** (case of several parameters): Suppose the condition $E[u(Y_1, \ldots, Y_m)] = 0$ for all $\boldsymbol{\theta} \in \Omega$ always implies that $u(y_1, \ldots, y_m) \equiv 0$ except on a zero-probability set. Then $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ 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 $\eta = g(\boldsymbol{\theta})$ is the parameter of interest and $T = T(\mathbf{Y})$ is an unbiased estimator of η. Then T is the unique MVUE of η .

- **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 $f(\mathbf{x};\boldsymbol{\theta}) = \exp\left\{\sum_{j=1}^m p_j(\boldsymbol{\theta})K_j(\mathbf{x}) + S(\mathbf{x}) + q(\boldsymbol{\theta})\right\}$, $\mathbf{x} \in \mathcal{S} \subset R^k$.
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 $f(\mathbf{x};\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Omega \subset R^m$, 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 $\eta = g(\boldsymbol{\theta})$ be the parameter of interest and $T = h(\mathbf{Y})$ is an unbiased estimator of η. Then T is the unique MVUE of η .

- Practice Problem: Example 7.7-1, Exercise 7.7-13

§ 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 .

- **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 $d \in R$. Then Z is a *location-invariant statistic* whose distribution does not depend on θ . Examples: 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: $X_{(1)}/X_{(n)}$, $X_1^2/\sum_{i=1}^n X_i^2$.

- **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 $c > 0, d \in R$. Then Z is a *location and scale invariant statistic* whose distribution does not depend on θ . Examples: $(X_1 - \bar{X})/S$, $[\max_i\{X_i\} - \min_i\{X_i\}]/S$.

  Practice Problems: Exercise § 7.8 - 1, 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.

Practice Problems: Exercise § 7.9 - 5, 7.

Chapter 8. Optimal Tests of Hypothesis

§8.1 Most Powerful Tests

- Hypothesis testing (general setup): Let $X_1, \ldots, X_n$ be i.i.d.$\sim f(x;\theta)$, $\theta \in \Theta = \Theta_0 \bigcup \Theta_1$, where $\Theta_0 \bigcap \Theta_1 = \emptyset$. Let S be the support of $X = (X_1, \ldots, X_n)'$. We want to test the null hypothesis $H_0 : \theta \in \Theta_0$ versus the alternative hypothesis $H_1 : \theta \in \Theta_1$.
  (1) Critical region (rejection region): C$\subset S$ such that,
  we reject H_0 if and only if $x = (x_1, \ldots, x_n)' \in$C .
  (2) Size of the test (significance level, Type I error):
  $\alpha = \max_{\theta \in \Theta_0} P_\theta(X \in C)$.
  (3) Power function: $\gamma_C(\theta) = P_\theta(X \in C)$, $\theta \in \Theta_1$.

- **Best critical region (Best Test)**: To test $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, let C be the critical region, which is a subset of $S \subset R^n$. We say that C is a best critical region of size α, $0 < \alpha < 1$ if
  (1) $P_{\theta_0}(X \in C) = \alpha$;
  (2) For any other critical region $A \subset S$ of the same size α, we must have
  $P_{\theta_1}(X \in C) \geq P_{\theta_1}(X \in A)$.
  In other words, C is the most powerful critical region of size α . The test based on C is called the most powerful test of size α .

- **Theorem (Neyman-Pearson)**: Let $X_1, \ldots, X_n$ be i.i.d.$\sim f(x;\theta)$, $\theta \in \{\theta_0, \theta_1\}$. The likelihood function $L(\theta; x) = \prod_{i=1}^n f(x_i;\theta)$, for $x = (x_1, \ldots, x_n)' \in S$. Let C be a subset of S and let k be a positive number such that
  (a) $L(\theta_0; x)/L(\theta_1; x) \leq k$ for each $x \in C$;
  (b) $L(\theta_0; x)/L(\theta_1; x) \geq k$ for each $x \notin C$;
  (c) $\alpha = P_{H_0}(X \in C)$.
  Then C is a best critical region of size α for testing the simple hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$.
  Note: (1) The conditions (a), (b), and (c) are also necessary for region C to be a best critical region of size α .
  (2) In the case of continuous distributions, the best critical region C of size α is unique in the probability sense.

- **Theorem (power of test)**: Let C be the best critical region of size α for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. Let $\gamma_C(\theta_1) = P_{\theta_1}(X \in C)$ denote the power of the test based on C . Then $\gamma_C(\theta_1) \geq \alpha$.
  In other words, a lower bound of the power of the most powerful test of size α is α .

- Theorem (nonparametric case): Let $X_1, \ldots, X_n$ be an arbitrary sample. It is desired to test the simple hypothesis "H_0 : the joint pdf (or pmf) is $g(x_1, \ldots, x_n)$" versus "H_1 : the joint pdf (or pmf) is $h(x_1, \ldots, x_n)$" . Then $C \subset R^n$ is a best

critical region of size α if, for $k > 0$,
(1) $g(x_1, \ldots, x_n)/h(x_1, \ldots, x_n) \le k$ for $(x_1, \ldots, x_n)' \in C$;
(2) $g(x_1, \ldots, x_n)/h(x_1, \ldots, x_n) \ge k$ for $(x_1, \ldots, x_n)' \notin C$
(3) $\alpha = P_{H_0}\left[(X_1, \ldots, X_n)' \in \mathcal{C}\right]$.

- For practice: Example 8.1.2

§ 8.2 Uniformly Most Powerful Tests

- **UMP critical region**: A critical region C is called a uniformly most powerful (UMP) critical region of size α for testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ if, for each $\theta_1 \in \Theta_1$, C is a best critical region of size α for testing H_0 against $H_1' : \theta = \theta_1$.
  The test based on the UMP critical region C is called a UMP test.

- **Monotone likelihood ratio**: The likelihood function $L(\theta; x)$, $x = (x_1, \ldots, x_n)'$, is said to have monotone likelihood ratio (mlr) in the statistic $Y = u(X_1, \ldots, X_n)$ if $L(\theta_1; x)/L(\theta_2; x)$ is a monotone function of $y = u(x_1, \ldots, x_n)$ as long as $\theta_1 < \theta_2$.

- Two-step standard procedure for finding a UMP test of size α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \in \Theta_1$, where Θ_1 might be $\theta > \theta_0$, $\theta < \theta_0$, or $\theta \ne \theta_0$:
  Step 1: For each fixed $\theta_1 \in \Theta_1$, find a best critical region C of size α for testing H_0 against $H_1' : \theta = \theta_1$ based on the Neyman-Pearson theorem.
  Step 2: Check if C depends on θ_1 . If it does not, then C is a UMP critical region of size α for testing H_0 against H_1; otherwise there is no UMP test for this case.

- Theorem: If $L(\theta; x)$ has mlr in the statistic $Y = u(X)$, then a UMP test for $H_0 : \theta \le \theta_0$ against $H_1 : \theta > \theta_0$ exists. Furthermore,
  (1) if it is monotone increasing, the UMP critical region takes the form of $\{(x_1, \ldots, x_n) : u(x_1, \ldots, x_n) \le C\}$;
  (2) if it is monotone decreasing, the UMP critical region takes the form of $\{(x_1, \ldots, x_n) : u(x_1, \ldots, x_n) \ge C\}$.
  Note: The case of $H_0 : \theta \ge \theta_0$ against $H_1 : \theta < \theta_0$ is similar.

- Theorem: Let $X_1, \ldots, X_n$ be i.i.d.$\sim f(x; \theta)$, where

$$f(x; \theta) = \exp\{p(\theta)K(x) + S(x) + q(\theta)\}$$

  belongs to the regular exponential class. If $p(\theta)$ is monotone, then the likelihood function $L(\theta; x)$ has mlr in $Y = \sum_{i=1}^{n} K(X_i)$.
  For example, if $p(\theta)$ is monotone increasing, then $L(\theta; x)$ has monotone decreasing likelihood ratio in Y .

- For practice: Example 8.2.1, Example 8.2.2, Example 8.2.5

§ 8.3 Likelihood Ratio Tests

- **Unbiased Test**: A test for $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ is said to be unbiased, if its power never falls below the significance level. In other words, if $\alpha = \max_{\theta \in \Theta_0} P_\theta[\text{ reject } H_0]$, then $P_\theta[\text{ reject } H_0] \geq \alpha$ for each $\theta \in \Theta_1$.

- **Likelihood ratio test**: For testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, the likelihood ratio test statistic is

$$\Lambda = \frac{\max_{\theta \in \Theta_0} L(\theta; \mathbf{x})}{\max_{\theta \in \Theta} L(\theta; \mathbf{x})} ,$$

  where $\Theta = \Theta_0 \cup \Theta_1$.
  Note that $0 < \Lambda \leq 1$. If H_0 is true, Λ should be close to 1; if H_1 is true, Λ should be smaller.

- **Likelihood ratio principle**: Reject H_0 if and only if $\Lambda \leq \lambda_0$, where $\lambda_0 < 1$ is a constant determined by the significance level α such that $P_{\theta_0}(\Lambda \leq \lambda_0) = \alpha$, where θ_0 is the boundary point of Θ_0 and Θ_1.

- ***p*-value**: The so-called p-value is the probability that the test statistic under H_0 is at least as extreme as the particular observed value.
  A small enough p-value indicates the rejection of H_0 .

- **Wilks's Theorem**: As the sample size n approaches ∞, the test statistic $-2\log(\Lambda)$ will be asymptotically χ^2-distributed with degrees of freedom equal to the difference in dimensionality of Θ and Θ_0.

- For practice: Example 8.3.1, Example 8.3.3, Exercise 8.3.12.