THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

STAT3602 Statistical Inference

(2020-2021 First Semester)

Example Class 7 [Sufficiency and Completeness]

Date: 4/11/2020

1 Review: Chapter 4

1.1 Sufficiency and Likelihood

1. **Likelihood and loglikelihood functions of θ.**

$$\ell_{\boldsymbol{X}}(\theta) \propto f(\boldsymbol{X} \mid \theta)$$

$$S_{\boldsymbol{X}}(\theta) = \ln \ell_{\boldsymbol{X}}(\theta)$$

2. **Sufficiency.**
   Statistic $T = T(X)$ is sufficient for θ if the conditional distribution of X given T is free of θ.

3. **Theorem 4.2.6.** The following three statements are equivalent:
   - Statistic $T = T(\boldsymbol{X})$ is sufficient for θ
   - For any samples X, X', $\quad T(X) = T(X') \Rightarrow \ell_X(\theta) \propto \ell_{X'}(\theta)$
   - **Factorization Criterion.** There exists a function $g(\cdot)$ such that for any sample $\boldsymbol{X}, \ell_{\boldsymbol{X}}(\theta) \propto g(T(\boldsymbol{X}), \theta)$

4. **Minimal Sufficiency.**
   (a) **Definition:** $T(X)$ is minimal sufficient for θ if it is sufficient and is a function of every other sufficient statistic.
   (b) **Theorem 4.3.2**
   $T(\boldsymbol{X})$ is minimal sufficient for θ if and only if

$$\text{for any samples } \boldsymbol{X}, \boldsymbol{X}', \quad T(\boldsymbol{X}) = T\left(\boldsymbol{X}'\right) \Leftrightarrow \ell_{\boldsymbol{X}}(\theta) \propto \ell_{\boldsymbol{X}'}(\theta)$$

5. **Sufficiency for Exponential Family $\boldsymbol{X} \sim f(\boldsymbol{x} \mid \boldsymbol{\pi}) \propto h(\boldsymbol{x}) \exp\left\{\sum_{j=1}^{k} \pi_j t_j(\boldsymbol{x})\right\}$.**
   Assume that the natural parameter space Π is not contained in an affine hyperplane of the form $\left\{\boldsymbol{\pi} : \sum_{i=1}^{k} c_i \pi_i = b\right\} \subset \mathbb{R}^k$ for c_i 's not all equal to 0. Then $T(\boldsymbol{X}) = (t_1(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X}))$ is minimal sufficient for $\boldsymbol{\pi}$.

1.2 Completeness

1. **Definition.**
   A sufficient statistic $T = T(\boldsymbol{X})$ is complete for θ if and only if, for any real function $g(\cdot)$ not depending on θ,

$$\mathbb{E}_\theta[g(T(\boldsymbol{X}))] = 0 \quad \forall \theta \quad \Rightarrow \quad \mathbb{P}_\theta\{g(T(\boldsymbol{X})) = 0\} = 1 \quad \forall \theta.$$

   The latter condition says that $g(\cdot)$ is a zero function almost surely under any θ.

2. **Lehmann-Scheffé Theorem.**
   If $T(\boldsymbol{X})$ is complete sufficient for θ, then $T(\boldsymbol{X})$ is minimal sufficient.

3. **Theorem for (Exponential family)**
   If the natural parameter space Π contains an open rectangle, i.e. a nonempty set of the form $(a_1, b_1) \times \cdots \times (a_k, b_k) \subset \mathbb{R}^k$, then the natural statistic $T(\boldsymbol{X}) = (t_1(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X}))$ is complete for the natural parameter $\pi = (\pi_1, \ldots, \pi_k)$.

2 Problems

1. Let $X_1, \ldots, X_n$ be independent Poisson random variables with X_j having parameter $j\lambda$, where $\lambda > 0$ is an unknown parameter.

   (a) Find a minimal sufficient statistic for λ.

   (b) What is its distribution?

   (c) Is it a complete sufficient statistic?

2. Let $X_1, \ldots, X_n$ be i.i.d. with a common density function

$$f(x) = \begin{cases} \dfrac{2}{3\theta}, & 0 \leq x \leq \theta/2 \\ \dfrac{4}{3\theta}, & \theta/2 < x \leq \theta \end{cases}$$

   Let Y_n be the maximum of $X_1, \ldots, X_n$, and N be the number of observations less than $Y_n/2$. Rewrite the X_i 's which are greater than or equal to $Y_n/2$ as $Y_{N+1} \leq Y_{N+2} \leq \cdots \leq Y_n$.

   (a) Show that the likelihood is

$$\ell_{\mathcal{X}}(\theta) = \left(\frac{4}{3\theta}\right)^n \left(\frac{1}{2}\right)^{N+\#\{i: Y_{N+i} \leq \theta/2, i=1,\ldots,n-N\}} 1\{Y_n \leq \theta\}$$

   (b) Prove that $(N, Y_{N+1}, \ldots, Y_n)$ is a sufficient statistic for θ.

3. A married man who frequently talks on his mobile is well known to have conversations the lengths of which are independent, identically distributed random variables, distributed as exponential with mean $1/\lambda$. His wife has long been irritated by his behavior and knows, from infinitely many observations, the exact value of λ. In an argument with her husband, the woman produces $t_1, \ldots, t_n$, the times of n telephone conversations, to prove how excessive her husband is. He suspects that she has randomly chosen the observations, conditional on their all being longer than the expected length of conversation. Assuming he is right in his suspicion, the husband wants to use the data he has been given to infer the value of λ. What is the minimal sufficient statistic he should use? Find the maximum likelihood estimator for λ

4. Suppose we observe $(X_{1,\ldots,} X_n, Y_1, \ldots, Y_m)$ where $X_1, \ldots, X_n, Y_1 \ldots, Y_m$ are independent random variables, with $X_i \sim N\left(\mu, \sigma_1^2\right)$ and $Y_j \sim N\left(\mu, \sigma_2^2\right)$ for $0 \leq i \leq n$ and $0 \leq j \leq m$. Assuming $\mu \in \mathbb{R}$ and $\sigma_1^2, \sigma_2^2 \in \mathbb{R}^+$, show that $T = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2, \sum_{j=1}^m Y_i, \sum_{j=1}^m Y_i^2\right)$ is a sufficient statistic. Show that T is not a complete sufficient statistic.