Mathematical Statistics2 Tutoring5

Seung Bong Jung

Seoul National University

November 1, 2021

Comparison of Estimators

Let $X_1, \dots, X_n$ be a random sample from a population with pdf $f(\cdot;\theta)$ for $\theta \in \Theta \subset \mathbb{R}^d$. Below, we write $X = (X_1, \dots, X_n)$. How to compare different estimators of a parameter of interest, say $g(\theta) \in \mathbb{R}$.

- Loss function: $L(\cdot,\cdot) : \Theta \times \mathcal{A} \to \mathbb{R}_+$, where $\mathcal{A} \subset \mathbb{R}$ is an action space for the estimation of $g(\theta)$ such that $L(\theta, g(\theta)) = 0$. For example,

$$L(\theta, a) = (a - g(\theta))^2, \quad L(\theta, a) = |a - g(\theta)|.$$

- Risk function: For an estimator $\delta(X)$ of $g(\theta)$,

$$R(\theta, \delta) = E_\theta L(\theta, \delta(X)).$$

In the case of the squared error loss, the risk $R(\theta, \delta)$ is the mean squared error of $\delta(X)$.

Difficulty with Uniform Comparison

- One would prefer δ_1 to δ_2 if $R(\theta, \delta_1) \leq R(\theta, \delta_2)$ for all $\theta \in \Theta$, and $R(\theta, \delta_1) < R(\theta, \delta_2)$ for some $\theta \in \Theta$.
- The difficulty is that there exists no estimator that is best in this sense. As if it exists, then $r(\theta, \delta_0) \overset{\theta \in \Theta}{\equiv} 0$, which leads to a contradiction.

Optimal Estimation

- Restricted class of estimators: One may find an estimator δ_0 in the class of unbiased estimators of $g(\theta)$ such that

$$E_\theta(\delta_0(X) - g(\theta))^2 \leq E_\theta(\delta(X) - g(\theta))^2 \text{ for all } \theta \in \Theta$$

  for any unbiased estimator $\delta(X)$. If exists, such an estimator is called UMVUE (Uniformly Minimum Variance Unbiased Estimator).

- Global measures of performance: The estimator that minimizes the maximum risk $r(\delta) = \max_{\theta\in\Theta} R(\theta, \delta)$ is called a minimax estimator. The estimator that minimizes the average risk

$$r(\delta; \pi) = \int_\Theta R(\theta, \delta)\pi(\theta)d\theta$$

  for a weight function π is called a Bayes estimator.

The Idea of Sufficiency

Suppose that we observe X_1 and X_2 that are i.i.d. Bernoulli(θ) random variables, where $0 < \theta < 1$.

- The distribution of $Y = X_1 + X_2$;

$$P_\theta(Y = 0) = (1 - \theta)^2,\ P_\theta(Y = 1) = 2\theta(1 - \theta),\ P_\theta(Y = 2) = \theta^2.$$

- When we observe Y, we may produce (X_1^*, X_2^*), without knowledge of the true θ, that has the same distribution of the original (X_1, X_2), as follows:
  (1) Put $(X_1^*, X_2^*) = (0, 0)$ when $Y = 0$; (2) put $(X_1^*, X_2^*) = (1, 1)$ when $Y = 2$; (3) conduct a randomized experiment and put $(X_1^*, X_2^*) = (1, 0)$ and $(X_1^*, X_2^*) = (0, 1)$, each with probability $1/2$ when $Y = 1$.
- Thus, for any estimator $\delta(X_1, X_2)$ of $g(\theta)$ one may find an estimator that depends only on Y rather than (X_1, X_2) but has the same risk as $\delta(X_1, X_2)$.

Sufficient Statistic

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf $f(\cdot;\theta)$ for $\theta \in \Theta \subset \mathbb{R}^d$. We write $X = (X_1, \ldots, X_n)$ below.

- Sufficient statistic for $\theta \in \Theta$: A statistic $Y = u(X)$ is called a sufficient statistic if the conditional distribution of X given Y does not depend on $\theta \in \Theta$, i.e., if

$$P_{\theta_1}(X \in A | Y = y) \overset{\theta_1, \theta_2 \in \Theta}{\equiv} P_{\theta_2}(X \in A | Y = y)$$

  for all A and for all y.

Factorization Theorem

A statistic $Y = u(X)$ is a sufficient statistic for $\theta \in \Theta$ if and only if there exist functions f_1 and f_2 such that

$$\prod_{i=1}^{n} f(x_i; \theta) = f_1(u(x), \theta) \cdot f_2(x) \text{ for all } x \text{ and for all } \theta \in \Theta.$$

Sufficient Statistic: Examples

- $U(\theta_1 - \theta_2, \theta_1 + \theta_2)$, $\theta = (\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R}_+$ is a sufficient statistic: Assume $n \geq 2$.

$$\prod_{i=1}^{n} f(x_i; \theta) = (2\theta_2)^{-n} I_{(\theta_1 - \theta_2, \infty)}(x_{(1)}) I_{(-\infty, \theta_1 + \theta_2)}(x_{(n)}),$$

so that $Y = (X_{(1)}, X_{(n)})$ is a sufficient statistic for $\theta \in \mathbb{R} \times \mathbb{R}_+$.

Sufficient Statistic: Examples

- Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are random samples from the bivariate normal distribution with $E(X_1) = E(Y_1) = 0$, $Var(X_1) = Var(Y_1) = 1$, and $Cov(X_1, Y_1) = \theta$ for $\theta \in (-1, 1)$. Then, the sufficient statistic for θ is given by $(\sum_{i=1}^n (X_i^2 + Y_i^2), \sum_{i=1}^n X_i Y_i)$.
- This is because

$$\prod_{i=1}^n f(x_i; \theta) \propto (1 - \theta^2)^{-n/2}$$
$$\exp\left(-\frac{1}{2\pi(1-\theta^2)}\left(\sum_{i=1}^n X_i^2 - 2\theta \sum_{i=1}^n X_i Y_i + \sum_{i=1}^n Y_i^2\right)\right)$$

Minimal Sufficient Statistic

- There are many sufficient statistics for a given model. As an extreme example, $X = (X_1, \ldots, X_n)$ itself is also a sufficient statistic. As another example, consider the case where X_1, X_2, X_3 are i.i.d. Bernoulli(θ) random variables and $\theta \in (0, 1)$. For latter model, sufficient statistics include (X_1, X_2, X_3), $(X_1 + X_2, X_3)$, $(X_1 + X_2, X_2)$, $(X_1, X_2 + X_3)$, and $X_1 + X_2 + X_3$.
- Minimal sufficient statistic: A sufficient statistic is called a minimal sufficient statistic (MSS) if it is a function of any sufficient statistic.

Properties of Sufficient Statistic

- Any statistic that is a $1-1$ function of a sufficient statistic for $\theta \in \Theta$ is also a sufficient statistic for $\theta \in \Theta$.
- Any statistic that is a $1-1$ function of MSS is also an MSS.
- MLE and SS: The unique MLE of θ, when it exists, is a function of any sufficient statistic for $\theta \in \Theta$.
- Existence of MSS: If the MLE of $\theta \in \Theta$ is unique and it is a sufficient statistic for $\theta \in \Theta$, then it is a minimal sufficient statistic for $\theta \in \Theta$.
- Suppose that there exists $\theta_0 \in \Theta$ such that supp$(f(\cdot;\theta)) \subset$ supp$(f(\cdot;\theta_0))$ for all $\theta \in \Theta$. Then, the statistic $T(X)$, as a function defined on Θ in such a way that $T(X)(\theta) = \prod_{i=1}^{n}(f(X_i;\theta)/f(X_i;\theta_0))$, is an MSS.

Characterization of Minimal Sufficiency

- Suppose that $f_\theta(\cdot)$ is pdf with $\theta \in \Theta \subset \mathbb{R}^d$. Let $X = (X_1, \ldots, X_n)$ be random samples from the distribution with pdf $f_\theta(\cdot)$ and assume that $T(X)$ and $T(Y)$ a sufficient statistic for θ.
- Then, $T(X)$ is a minimal sufficient statistic for θ if and only if

$$\frac{\prod_{i=1}^n f_\theta(x_i)}{\prod_{i=1}^n f_\theta(y_i)} \text{ is independent of } \theta \Leftrightarrow T(X) = T(Y)$$

  for another random samples $Y = (Y_1, \ldots, Y_n)$ from f_θ.
- The proof follows from factorization theorem.

MSS: Examples

Let $X_1, \ldots, X_n$ $(n \geq 2)$ be a random sample from $U(\theta_1 - \theta_2, \theta_1 + \theta_2)$, $\theta = (\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R}_+$. For this model, $Y = (X_{(1)}, X_{(n)})$ is an MSS.
Proof: The sufficiency was established before. Let the model is reparametrized by $\eta = (\eta_1, \eta_2)$, where $\eta_1 = \theta_1 - \theta_2$ and $\eta_2 = \theta_1 + \theta_2$. Then,

$$\prod_{i=1}^{n} f(x_i; \eta) = (\eta_2 - \eta_1)^{-n} I_{(-\infty, x_{(1)}]}(\eta_1) I_{[x_{(n)}, \infty)}(\eta_2).$$

Clearly, $(\hat{\eta}_1, \hat{\eta}_2) = (X_{(1)}, X_{(n)})$ is the unique MLE, so that $(\hat{\theta}_1, \hat{\theta}_2)$ defined by

$$\hat{\theta}_1 = (X_{(1)} + X_{(n)})/2, \quad \hat{\theta}_2 = (X_{(n)} - X_{(1)})/2$$

is the unique MLE of $(\theta_1, \theta_2) = ((\eta_1 + \eta_2)/2, (\eta_2 - \eta_1)/2)$. Since $(\hat{\theta}_1, \hat{\theta}_2)$ is a $1-1$ function of the SS $(X_{(1)}, X_{(n)})$, it is also an SS and thus an MSS. This establishes that $(X_{(1)}, X_{(n)})$ is an MSS since it is a $1-1$ function of $(\hat{\theta}_1, \hat{\theta}_2)$.

MSS: Examples

- Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are random samples from the bivariate normal distribution with $E(X_1) = E(Y_1) = 0$, $Var(X_1) = Var(Y_1) = 1$, and $Cov(X_1, Y_1) = \theta$ for $\theta \in (-1, 1)$. Then, the minimal sufficient statistic for θ is given by $(\sum_{i=1}^n (X_i^2 + Y_i^2), \sum_{i=1}^n X_i Y_i)$.
- The sufficiency was established before. Let $\eta_1(\theta) = -\frac{1}{2}(1 - \theta^2)$ and $\eta_2(\theta) = \theta/(1 - \theta^2)$. Define a map $\eta : \Theta \to \mathbb{R}^2$ by $\eta(\theta) = (\eta_1(\theta), \eta_2(\theta))$, where $\Theta = (-1, 1)$. By characterization of minimal sufficiency, it suffices to show that the set of all differences $\eta(\theta_0) - \eta(\theta_1)$ for $\theta_0, \theta_1 \in \Theta$, denoted by $\eta(\Theta) \ominus \eta(\Theta)$, spans $\mathbb{R}^2$.