Sufficiency

Motivation. We hope to separate the information contained in the data into the information relevant for making inference about θ and the information irrelevant for these inferences. In other words, we would like to compress the data to, e.g. $T(X)$, without loss of information. (Actually, it often turns out that some part of the data carries no information about the unknown distribution that produces the data)

Benefits:

1. increasing computational efficiency and decreasing storage requirements
2. involving irrelevant information may increase an estimator's risk (see Rao-Blackwell Theorem)
3. Improving the scientific interpretability of our data

Definition of Sufficient Statistic

Suppose X has a distribution from $\mathcal{P} = \{ P_\theta : \theta \in \Omega \}$. A statistic T is **sufficient** for θ if, for every t in the range $\mathcal{T}$ of T, the conditional distribution $P_\theta(X \mid T(X) = t)$ is independent of θ.

Example: Let $X_i \sim Ber(\theta)$ i.i.d., $i = 1, ..., n$. Show that $T = \sum_{i=1}^n X_i$ is sufficient for θ.

Neyman Fractorization Theorem. Suppose a family $\{ P_\theta : \theta \in \Omega \}$ of distributions have joint mass functions or densities $\{ p(x; \theta) : \theta \in \Omega \}$. Then a statistic T is sufficient for θ if and only if there are functions h and g such that the density/mass function can be written

$$p(x; \theta) = h(x) \cdot g(T(x), \theta).$$

Proof: To be presented in class (for the discrete case).

Examples:

1. Let $Y_i \sim Uniform(0, \theta)$ i.i.d., $i = 1, ..., n$. Show that $T = Y_{(n)}$ is sufficient for θ.

2. Let $X_i \sim N(\theta, 1)$ i.i.d., $i = 1, ..., n$. Show that $T = \sum_{i=1}^{n} X_i$ is sufficient for θ.

The Rao-Blackwell Theorem

Suppose X is distributed according to $P_\theta(x) \in \{P_\theta : \theta \in \Omega\}$ and a statistic $T(X)$ is sufficient for θ. Given any estimator $\delta(X)$ of θ, define $\eta(T) = E_\theta[\delta(X)|T(X)]$. If the loss function $\mathcal{L}(\theta, \delta(X))$ is convex and the risk function $R(\theta, \delta(X)) = E[\mathcal{L}(\theta, \delta(X))] < \infty$, then $R(\theta, \eta) \leq R(\theta, \delta)$. If $\mathcal{L}$ is strictly convex, then the inequality is strict unless $\delta = \eta$.

Note that the loss function reflects the degree of wrongness of an estimate. The commonly used quadratic loss function is defined as $\mathcal{L}(\theta, \delta) = (\theta - \delta(X))^2$.

Proof of Rao-Blackwell: by Jensen's inequality and iterated expectation.

Jensen's Inequality

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to R$ be an integrable random variable, i.e. $E[|X|] < \infty$.

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function such that $\varphi(X)$ is integrable. Then

$$\varphi\big(\mathbb{E}[X]\big) \;\leq\; \mathbb{E}\big[\varphi(X)\big].$$

Moreover, if φ is strictly convex, then equality holds if and only if X is almost surely constant.

For example, $(E(X))^2 \leq E(X^2)$.

Some notes

- Why we need T to be sufficient?
  - When T is sufficient, $\eta(T) = E_\theta[\delta(X)|T(X)]$ will be independent of θ, and so $\eta(T)$ is a statistic.
  - Let us consider $\mathcal{L}(\theta, \eta(T)) = (\theta - \eta(T))^2 = (\theta - E_\theta[\delta(X)|T])^2 = (E_\theta[(\theta - \delta(X))|T])^2$. Note that $E_\theta[\theta|T] = \int \theta f(x|T;\theta)\mathrm{d}x = \theta \int f(x|T;\theta)\mathrm{d}x$. We have the last equation because $f(x|T)$ is independent of θ and so we do not have to worry that the support could be different

Minimal Sufficiency

Definition. Suppose $T(X)$ is sufficient for $P = \{P_\theta : \theta \in \Omega\}$. For any other sufficient statistic $S(X)$, if we can always find a function f such that $T = f(S)$, then T is minimally sufficient.

($T = f(S)$ means (i) the knowledge of S implies the knowledge of T, and (ii) T provides a greater reduction of data unless f is one-to-one.)

A d-parameter exponential family has pdf in the following form

$$p(x, \theta) = h(x) \exp[\sum_{i=1}^{d} \eta_i(\theta) T_i(x) - A(\theta)],$$

which is of full rank if $\eta(\Theta) = \{\eta_1(\theta), ..., \eta_d(\theta)\}$ has non-empty interior in $\Re^d$ and $T_1(x), ..., T_d(x)$ are linearly independent. In a full rank exponential family, the natural sufficient statistic $T = (T_1, ..., T_d)$ is minimally sufficient.

Examples

- Let $X_1, ..., X_n$ be iid and follow a normal distribution $N(\mu, \sigma^2)$. Find the minimal sufficient statistic for μ and σ^2.

Relevant Readings on Sufficient Statistics

Chapters 2-5 of the Robert Keener book.