Lecture 6 Principle of Data Reduction

Ma Xuejun

School of Mathematical Sciences

Soochow University

https://xuejunma.github.io

Outline

Introduction

From these on, we start doing statistics.

- **Probability** Suppose $X_1, \cdots, X_n$ are i.i.d. $f_\theta(\cdot)$, θ fixed. what is the probability of observing

$$\mathbf{X} = (X_1, \cdots, X_n) \;\; \mathbf{x} = (x_1, \cdots, x_n)$$

The answer is $f_\theta(x_1) f_\theta(x_2) \cdots f_\theta(x_n)$.

- **Statistics** We know $X_1, \cdots, X_n$ are i.i.d. $f_\theta(\cdot)$, but we don't know the exact value of θ. Suppose that $\mathbf{X} = \mathbf{x}$ is observed, what is the most likely θ?

From midterm, we know that if θ' is the true parameter, then

$$E\left[\frac{f_{\theta'}(X_1) f_{\theta'}(X_2) \cdots f_{\theta'}(X_n)}{f_{\theta''}(X_1) f_{\theta''}(X_2) \cdots f_{\theta''}(X_n)}\right] \geq 1$$

or

$$\frac{E\left[\sum_{i=1}^n \log f_{\theta'}(X_i)\right]}{E\left[\sum_{i=1}^n \log f_{\theta''}(X_i)\right]} \geq 1 \text{ for } \theta'' \neq \theta'.$$

So we have the likelihood principle:

An estimate of the true parameter is $\hat{\theta}$, which maximize $\sum_{i=1}^n \log f_\theta(X_i)$

The Sufficiency Principle

A sufficient statistic for a parameter θ is a statistic that captures all the information about θ contained in the sample.

- **Sufficiency Principle** If $T(\mathbf{X})$ is a sufficient statistic for θ, then any information about θ should depend on the sample $\mathbf{X}$ only through the value $T(\mathbf{X})$. That is, if $\mathbf{x}$ and $\mathbf{y}$ are two sample points such that $T(\mathbf{x}) = T(\mathbf{y})$, then the inference about θ should be the same whether $\mathbf{X} = \mathbf{x}$ or $\mathbf{X} = \mathbf{y}$ is observed.

Sufficient Statistics

Definition 6.2.1

A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.

The above definition is not easy to check whether a statistic $T(\mathbf{X})$ is a sufficient statistic.

Theorem 6.2.2

If $p(\mathbf{x}|\theta)$ is the pdf or pmf of $\mathbf{X}$, and $q(\mathbf{t}|\theta)$ is the pdf or pmf of $T(\mathbf{X})$, then $T(\mathbf{X})$ is a sufficient statistic for θ if, for every $\mathbf{x}$,

$$\frac{p(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)} \equiv \text{constant in } \quad \theta$$

.

Example 6.2.3

Let $X_1, X_2, \cdots, X_n$ be i.i.d. Bernoulli(p). Let

$$T(\mathbf{X}) = X_1 + X_2 + \cdots + X_n.$$

Then

$$\begin{aligned}
p(\mathbf{x}|p) &= p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)} \\
q(t|p) &= \binom{n}{t} p^t (1-p)^{n-t} \\
\frac{p(\mathbf{X}|p)}{q(T(\mathbf{x})|p)} &= \frac{p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)}}{\binom{n}{x_1+\cdots+x_n} p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)}} \\
&= \frac{1}{\binom{n}{T(\mathbf{x})}} \quad \text{does not depend on } \theta
\end{aligned}$$

Example 6.2.4

Let $X_1, X_2, \cdots, X_n$ be i.i.d.$N(\mu, \sigma^2)$, where σ is unknown.

$$T(\mathbf{X}) = (X_1 + X_2 + \cdots + X_n)/n$$

is sufficient for μ.

$$\begin{aligned} f(\mathbf{x}|\mu) &= \prod_{i=1}^{n} (2\pi)^{-1/2} \sigma^{-1} \exp\left(-(x_i - \mu)^2/(2\sigma^2)\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left[-\sum_{i=1}^{n} (x_i - \mu)^2/(2\sigma^2)\right] \\ &= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 - \frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right] \end{aligned}$$

$\bar{\mathbf{X}} \sim N(\mu, \sigma^2/n)$

$$f_{\bar{\mathbf{X}}}(t|\mu) = (2\pi\sigma^2/n)^{-n/2} \exp\left[-\frac{n}{2\sigma^2}(t - \mu)^2\right]$$

So

$$\frac{f(\mathbf{x}|\mu)}{f_{\bar{\mathbf{X}}}(t|\mu)} = \frac{(2\pi)^{-n/2}}{(2\pi n^{-1})^{-1/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2\right]$$

does not depend on μ.

Example 6.2.5 (Sufficient Order Statistic)

Suppose X_1, X_2, $\cdots$, X_n are i.i.d. $f(x)$. Then

$$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$$

is sufficient for $f(\cdot)$. $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ is the order statistic of $X_1, X_2, \ldots, X_n$.

Theorem 6.2.6, Factorial Theorem

Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of a sample $\mathbf{X}$. A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exist functions $g(t|\theta)$ and $h(\mathbf{x})$ such that, for all sample point $\mathbf{x}$ and all parameter points θ,

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x}|\theta))h(\mathbf{x})$$

This theory is most useful in finding out sufficient statistic

- **Example 6.2.7** $X_1, X_2, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$, σ known, we have

$$f(\mathbf{x}|\mu) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right]$$

since $\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$ does not involve μ,
$\bar{\mathbf{X}} = \frac{1}{n}(X_1 + \ldots + X_n)$ is a sufficient statistic for μ.

Example 6.2.8: Uniform Sufficient Statistic

Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations from discrete Uniform distribution on 1, 2, $\cdots, \theta$.

$$f(x|\theta) = \begin{cases} 1/\theta, & x = 1, 2, \cdots, \theta \\ 0, & \text{otherwise} \end{cases}$$

Thus the joint pmf of $X_1, \ldots, X_n$ is

$$f(\mathbf{x}|\theta) = \begin{cases} \theta^{-n}, & x_i \in \{1, 2, \cdots, \theta\} \text{ for } i = 1, 2, \cdots, n \\ 0, & \text{otherwise} \end{cases}$$

Let

$$\begin{aligned} f(\mathbf{x}|\theta) &= \theta^{-n} I(x)_{\{1,2,\ldots,\theta\}} = \theta^{-n} I(\max\{x_i\})_{\{\max\{x_i\} \leq \theta\}} \\ g(t|\theta) &= \theta^{-n}, t \leq \theta \\ &= \theta^{-n} \cdot 1[t \leq \theta] \end{aligned}$$

Then

$$f(\mathbf{x}|\theta) = g\left(\max_{1 \leq i \leq n} \{x_i\} | \theta\right) \cdot h(\mathbf{x})$$

$\Longrightarrow T(\mathbf{X}) = \max_{1 \leq i \leq n} \{X_i\}$ is a sufficient statistic for θ.

Example 6.2.9

$X_1, \cdots, X_n \sim N(\mu, \sigma^2)$.

$$\begin{aligned} f(\mathbf{x}|\mu,\sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 - \frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\} \\ &= h(\mathbf{x})g(T_1(\mathbf{x}), T_2(\mathbf{x})|\mu,\sigma^2) \end{aligned}$$

Here,

$$h(\mathbf{x}) \equiv 1$$

$$g(t_1, t_2|\mu,\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{n-1}{2\sigma^2}\cdot t_2 - -\frac{n-1}{2\sigma^2}(t_1-\mu)^2\right\}$$

Hence, $T_1(\mathbf{x}) = \bar{X}$, $T_2(\mathbf{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ are sufficient statistics.

Theorem 6.2.10

Let $X_1, X_2, \cdots, X_n$ be i.i.d. observations from a pdf or pmf $f(x|\boldsymbol{\theta})$ that belongs to an exponential family given by

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)$, $d \leq k$. Then

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \sum_{j=1}^{n} t_2(X_j), \cdots, \sum_{j=1}^{n} t_k(X_j)\right)$$

is a sufficient statistic for $\boldsymbol{\theta}$

- **Example** Let $X_1, X_2, \cdots, X_n$ be i.i.d. Gamma(α, β), then $T(\mathbf{X}) = \left(\sum_{j=1}^{n} \log X_j, \sum_{j=1}^{n} X_j\right)$ are sufficient for (α, β).
- **Example** Let $X_1, X_2, \cdots, X_n$ be i.i.d. Uniform(α, β), $\alpha < \beta$, then $(\min_{1\leq i\leq n} X_i, \max_{1\leq i\leq n} X_i)$ is sufficient for (α, β).