Principles of Data Reduction

* In statistics we use a sample $X_1, \ldots, X_n$ to make inference about a parameter θ through the use of a statistic $T(\boldsymbol{X})$.

* The aim of data reduction is to keep all of the relevant information in the sample through a smaller number of statistics.

* Any statistic $T(\boldsymbol{X})$ defines a partition of the sample space into sets

$$A_t = \{\boldsymbol{x} : T(\boldsymbol{X} = t)\}$$

* Within such a partition we treat two samples, $\boldsymbol{x}$ and $\boldsymbol{y}$ as equal if $T(\boldsymbol{x}) = T(\boldsymbol{y})$.

The Sufficiency Principle

* The sufficiency principle relies on the concept of a sufficient statistic.

* Such statistics are functions of the data which contain all of the information about the parameter of interest.

Definition 9.1
A statistic $T(\boldsymbol{x})$ is a sufficient statistic for a parameter θ if the conditional distribution of the sample $\boldsymbol{X}$ given the value of $T(\boldsymbol{X})$ does not depend on θ.

The Sufficiency Principle

If $T(\boldsymbol{X})$ is a sufficient statistic for θ then any inference about θ should depend on the sample $\boldsymbol{X}$ only through the value of $T(\boldsymbol{X})$. If two sample points $\boldsymbol{x}$ and $\boldsymbol{y}$ have $T(\boldsymbol{x}) = T(\boldsymbol{y})$ then inference about θ should be the same whether $\boldsymbol{X} = \boldsymbol{x}$ or $\boldsymbol{X} = \boldsymbol{y}$.

Characterizing Sufficient Statistics

Theorem 9.1

Let $\boldsymbol{X}$ be a random vector with pdf or pmf $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$ and let $T(\boldsymbol{X})$ be a statistic with pdf or pmf $f_T(t \mid \theta)$. Then $T(\boldsymbol{X})$ is a sufficient statistic if, for every $\boldsymbol{x}$ with $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta > 0)$, the ratio

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_T(T(\boldsymbol{x}) \mid \theta)}$$

is constant as a function of θ.

Theorem 9.2 (Factorization Criterion)

Let $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$ be the joint pdf (pmf) of a sample $\boldsymbol{X}$. A statistic $T(\boldsymbol{X})$ is a sufficient statistic for θ if, and only if, there exist two non-negative functions $g(t, \theta)$ and $h(\boldsymbol{x})$ such that $h(\boldsymbol{x})$ is free of θ and for all sample points $\boldsymbol{x}$

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) = g(T(\boldsymbol{x}), \theta)h(\boldsymbol{x}).$$

Sufficient Statistics in the Exponential Family

Theorem 9.3

If $X_1, \ldots, X_n$ is a random sample from a distribution having an exponential family form

$$f(x; \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left\{\sum_{j=1}^{k} w_j(\boldsymbol{\theta})t_j(x)\right\}$$

Then a sufficient statistic for $\boldsymbol{\theta}$ is the vector

$$T(X_1, \ldots, X_n) = \left\{\sum_i t_1(x_i), \ldots, \sum_i t_k(x_i)\right\}$$

Multiple Sufficient Statistics

* In general, there are many sufficient statistics available for any model.

* Clearly any function of a sufficient statistic is also a sufficient statistic.

* The sample is a sufficient statistic (although it results in no data reduction).

* For *iid* data, the ordered sample is also a sufficient statistic.

Minimal Sufficient Statistics

Definition 9.2

A sufficient statistic $T(\boldsymbol{X})$ *is minimal sufficient if, and only if, for every other sufficient statistic* $T'(\boldsymbol{X})$, $T(\boldsymbol{X})$ *is a function of* $T'(\boldsymbol{X})$.

* By saying that $T(\boldsymbol{X})$ is a function of $T'(\boldsymbol{X})$ we simply mean that if $\boldsymbol{x}$ and $\boldsymbol{y}$ are sample points such that $T'(\boldsymbol{x}) = T'(\boldsymbol{y})$ then $T(\boldsymbol{x}) = T(\boldsymbol{y})$.

Theorem 9.4

Let $f(\boldsymbol{x} \mid \theta)$ *be the pdf of a sample* $\boldsymbol{X}$ *and suppose that there exists a function* $T(\boldsymbol{x})$ *such that, for any two points* $\boldsymbol{x}$ *and* $\boldsymbol{y}$, *the ratio* $f(\boldsymbol{x} \mid \theta)/f(\boldsymbol{y} \mid \theta)$ *is constant in* θ *if, and only if,* $T(\boldsymbol{x}) = T(\boldsymbol{y})$. *Then* $T(\boldsymbol{X})$ *is a minimal sufficient statistic for* θ.

Complete Families of Distributions

Definition 9.3

Suppose $f(t \mid \theta)$ is the family of distributions for a statistic $T(\boldsymbol{X})$ indexed by the parameter θ. This family of distributions is called complete if

$$\mathsf{E}_\theta(g(T)) = 0 \implies \mathsf{P}_\theta(g(T) = 0) = 1 \text{ for every } \theta.$$

* A statistic $T(\boldsymbol{X})$ whose sampling distribution is from a complete family is often called a complete statistic.

* Unfortunately, it is often quite difficult to prove completeness.

Theorem 9.5 (Bahadur's Theorem)

If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic.

Complete Statistics in the Exponential Family

Theorem 9.6

Suppose that $X_1, \ldots, X_n$ is a random sample from an exponential family with pdf or pmf given by

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

and further suppose that the set $\{w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})\}$ contains an open subset in $\mathbb{R}^k$.

Then the statistic

$$T(X_1, \ldots, X_n) = \left\{\sum_i t_1(x_i), \ldots, \sum_i t_k(x_i)\right\}$$

is complete and minimal sufficient.

Ancillary Statistics

Definition 9.4

Suppose that $\boldsymbol{X}$ is a random sample from a distribution depending on a parameter θ. A statistic $S(\boldsymbol{X})$ is called an ancillary statistic if the sampling distribution of $S(\boldsymbol{X})$ does not depend on θ.

Theorem 9.7 (Basu's Theorem)

If $T(\boldsymbol{X})$ is a complete and minimal sufficient statistic then $T(\boldsymbol{X})$ is independent of any ancillary statistic.

Likelihood Principle

Definition 9.5

Let $\boldsymbol{X}$ be a random sample with joint pdf (or pmf) $f(\boldsymbol{x} \mid \theta)$. If $\boldsymbol{x}$ is the observed value of $\boldsymbol{X}$ then the likelihood function is the function of θ given by

$$L(\theta \mid \boldsymbol{x}) = f(\boldsymbol{x} \mid \theta)$$

* We can use the likelihood to compare the probability of observing $\boldsymbol{X} = \boldsymbol{x}$ under different parameter values.

* If $L(\theta_1 \mid \boldsymbol{x}) > L(\theta_2 \mid \boldsymbol{x})$ then we may think of θ_1 as being more plausible than θ_2 given that we did actually observe $\boldsymbol{X} = \boldsymbol{x}$.

The Likelihood Principle

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are two sample points such that

$$L(\theta \mid \boldsymbol{x}) = C(\boldsymbol{x}, \boldsymbol{y}) L(\theta \mid \boldsymbol{y})$$

then the inference drawn from $\boldsymbol{x}$ should be identical to those drawn from $\boldsymbol{y}$.

* Note that if $T(\boldsymbol{X})$ is a sufficient statistic and $\boldsymbol{x}$ and $\boldsymbol{y}$ are such that $T(\boldsymbol{x}) = T(\boldsymbol{y})$ then the sufficiency principle says that inference should be the same and so does the likelihood principle.