STAT 713 sp 2023 Lec 02 slides

Data reduction part 2: Minimality, ancillarity

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Setup and notation

- Let $\mathbf{X} = (X_1, \ldots, X_n)$ represent the set of rvs from an experiment.
- Use $\mathbf{x} = (x_1, \ldots, x_n)$ to represent a specific set of values for the rvs in $\mathbf{X}$.
- Let $f(\mathbf{x}; \theta)$ or $p(\mathbf{x}; \theta)$ denote the joint pdf or pmf of the rvs in $\mathbf{X}$, resp.
- The distribution of $\mathbf{X}$ depends on a parameter (or some parameters) $\theta \in \Theta$.
- A function $T(\mathbf{X})$ of the rvs in $\mathbf{X}$ is called a *statistic*.

Goal: Learn about θ from a realization of $\mathbf{X}$ via the value of a statistic $T(\mathbf{X})$.

Key concepts of data reduction in the rough:

1. A *sufficient statistic* carries all the information about θ from $\mathbf{X}$.
2. A *minimal sufficient statistic* carries the above and no more than this.
3. An *ancillary statistic* carries no information about θ.
4. A *complete statistic* retains info about θ under any non-deg. transformation.

We think of computing a statistic $T(\mathbf{X})$ as "reducing" or summarizing the data.

Example: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Exponential}(\lambda)$ and consider the statistics

$$T_1(\mathbf{X}) = X_{(1)}, \qquad T_2(\mathbf{X}) = \bar{X}_n, \qquad T_3(\mathbf{X}) = S_n^2,$$

$$T_4(\mathbf{X}) = (\textstyle\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2), \quad T_5(\mathbf{X}) = X_{(1)}/X_{(n)}, \quad T_6(\mathbf{X}) = (X_{(1)}, \dots, X_{(n)}).$$

Sufficiency, *minimality*, *ancillarity*, and *completeness* address (resp) the questions

1. Which reductions of the data do not discard any information about λ?
2. Which ones keep all relevant information about λ, but discard all else?
3. Which ones discard all information about λ?
4. Which ones retain info about λ under any non-deg. transformation?

Table of Contents

Exercise: Let $X_1, X_2, X_3 \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$, $p \in (0, 1)$, and consider the statistics

$$T(\mathbf{X}) = X_1 + X_2 + X_3 \quad \text{and} \quad T'(\mathbf{X}) = (X_1, X_1 + X_2 + X_3),$$

both of which are sufficient statistics for p. Now:

1. List the elements in the support $\mathcal{X}$ of $\mathbf{X} = (X_1, X_2, X_3)$.
2. Give the supports $\mathcal{T}$ and $\mathcal{T}'$ of $T(\mathbf{X})$ and $T'(\mathbf{X})$.
3. Identify the partition of the sample space $\mathcal{X}$ induced by each statistic.
4. Which suff. statistic corresponds to a coarser partition of the sample space?

We want maximal data reduction without losing information about θ.

Minimal sufficient statistic

A sufficient statistic $T(\mathbf{X})$ is called a *minimal sufficient statistic* if it is a function of any other sufficient statistic.

Hard to find minimal sufficient statistics using the definition. Use following result:

Theorem (Find a minimal suff. statistic, cf. Thm 6.2.13 in CB)

Let $\mathbf{X}$ have joint pdf/pmf $f(\mathbf{x};\theta)$ with support not depending on θ. Let $T(\mathbf{x})$ be a function such that for any $\mathbf{x},\mathbf{y} \in \mathcal{X}$

$$f(\mathbf{x};\theta)/f(\mathbf{y};\theta) \text{ is constant as a function of } \theta \iff T(\mathbf{x}) = T(\mathbf{y}).$$

Then $T(\mathbf{X})$ is a minimal sufficient statistic for θ.

Exercise: Prove the result.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$, where $p \in (0, 1)$.

Check whether $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a minimal sufficient statistic for p.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$, $\sigma > 0$.

Check if $T(\mathbf{X}) = (\bar{X}_n, S_n^2)$ is a minimal sufficient statistic for (μ, σ^2).

If the support of $\mathbf{X}$ depends on θ, we check minimal sufficiency more carefully:

Theorem (Minimal suff. if support depends on the parameter)

Let $\mathbf{X}$ *have joint pdf/pmf* $f(\mathbf{x};\theta)$ *and let* $T(\mathbf{x})$ *be a function st for any* $\mathbf{x},\mathbf{y}$,

(i) $\{\theta : f(\mathbf{x};\theta) > 0\} = \{\theta : f(\mathbf{y};\theta) > 0\}$ *and*

(ii) $f(\mathbf{x};\theta)/f(\mathbf{y};\theta)$ *is constant as a function of* θ *on* $\{\theta : f(\mathbf{y};\theta) > 0\}$

$\iff$ $T(\mathbf{x}) = T(\mathbf{y})$. *Then* $T(\mathbf{X})$ *is a minimal sufficient statistic.*

Exercise: Let $X_1,\dots,X_n \overset{\text{ind}}{\sim} \text{Uniform}(a,b)$, where $-\infty < a < b < \infty$.

Find a minimal sufficient statistic for (a,b).

Theorem (Non-uniqueness of minimal sufficient statistics)

Any one-to-one function of a minimal sufficient statistic is also a minimal sufficient statistic.

Exercise: Prove the result.

Example: For $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$, $\sigma > 0$, both

$$T(\mathbf{X}) = (\bar{X}_n, S_n^2) \quad \text{and} \quad T'(\mathbf{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$$

are minimal sufficient statistics for (μ, σ^2).

Minimality is not about the dimension of the statistic.

Exercise: Let $X_1, \dots, X_n$ be iid rvs equal to j with probability p_j for $j = 1, \dots, k$, where $p_1 + \cdots + p_k = 1$. Find two minimal sufficient statistics for $(p_1, \dots, p_k)$.

Table of Contents

Some statistics discard all information about θ.

Ancillary statistic

A statistic $S(\mathbf{X})$ is an *ancillary statistic* if its distribution does not depend on θ.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\theta, 1)$. Show that $S(\mathbf{X}) = S_n^2$ is ancillary to θ.

Ancillary statistics are easy to find for location-scale families.

Recall location-scale families

The family of pdfs $\{f(x;\mu,\sigma) : \mu \in \mathbb{R}, \sigma > 0\}$ is a location-scale family if

$$f(x;\mu,\sigma) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right)$$

for some "standard" pdf f_Z for all $\mu \in \mathbb{R}$ and $\sigma > 0$.

Exercise: Find the location and/or scale parameters for the following families:

1. Exponential(λ), $\lambda > 0$.
2. Uniform($\mu - \theta, \mu + \theta$), $\mu \in \mathbb{R}$, $\theta > 0$.
3. Gamma(α_0, β), where $\alpha_0 > 0$ is known and $\beta > 0$.
4. Weibull(ν_0, β), where $\nu_0 > 0$ is known and $\beta > 0$.

A function $S(x_1, \ldots, x_n)$ is *location-scale-invariant* if for any $a > 0$ and $b \in \mathbb{R}$

$$S(x_1, \ldots, x_n) = S(ax_1 + b, \ldots, ax_n + b)$$

for all $(x_1, \ldots, x_n)$.

Theorem (Ancillary statistics for location-scale families)

If $\mathbf{X}$ a rs from a L-S family, then $S(\mathbf{X})$ is ancillary if $S(\mathbf{x})$ is L-S-invariant.

Exercise: Prove the above.

Exercise: Check for ancillarity of the statistics:

1. For $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Uniform}(0, \theta)$, check
$$S_1(\mathbf{X}) = \frac{X_{(1)}}{X_{(n)}} \quad \text{and} \quad S_2(\mathbf{X}) = \frac{X_1 - X_2}{\bar{X}_n}.$$
2. For $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, check
$$S_1(\mathbf{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right) \quad \text{and} \quad S_2(\mathbf{X}) = \frac{X_2 - X_1}{X_3 - X_2}.$$
3. For $X_1, \dots, X_n \overset{\text{ind}}{\sim} f(x; \theta) = e^{-(x-\theta)} e^{-e^{-(x-\theta)}}$, check
$$S_1(\mathbf{X}) = \frac{X_1}{X_1 + \cdots + X_n} \quad \text{and} \quad S_2(\mathbf{X}) = X_{(n)} - X_{(1)}.$$

Questions:

1. Are ancillary statistics worthless? No; see discussion in Sec 6.2.3 in CB.
2. Are ancillary statistics independent of minimal sufficient statistics?

This leads us to consider the property of *completeness*.