Principles of Data Reduction[1]

[1] Based on Casella and Berger's (2002) *Statistical Inference* and on Migon and Gamerman's (1999) *Statistical Inference: An Integrated Approach*.

Outline

Statistic

Statistics: functions of the sample.

$\Rightarrow$ Any statistic, $T(\boldsymbol{X})$, defines a form of data reduction or data summary.

$\Rightarrow$ Data reduction as a partition of the sample space $\mathcal{X}$.

Let $\mathcal{T} = \{t : t = T(\boldsymbol{x}) \text{ for some } \boldsymbol{x} \in \mathcal{X}\}$ be the image of $\mathcal{X}$ under $T(\boldsymbol{x})$. Then $T(\boldsymbol{x})$ partitions the sample space into sets $A_t, t \in \mathcal{T}$, defined by $A_t = \{\boldsymbol{x} : T(\boldsymbol{x})\}$.

A statistic summarizes the data in that, rather than reporting the entire sample $\boldsymbol{x}$, it reports only that $T(\boldsymbol{x}) = t$ or, equivalently, $\boldsymbol{x} \in A_t$.

Sufficiency, likelihood and invariance principles

Sufficiency Principle

Does not discard information about θ while achieving some summarization of the data.

Likelihood Principle

Describes a function of the parameter, determined by the observed sample, that contains all the information about θ tha is available from the sample.

Invariance Principle

Preserves some important features of the model.

Sufficiency Principle

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

Sufficient Statistics: A statistic $T(\boldsymbol{X})$ is a *sufficient statistic for* $\boldsymbol{\theta}$ if the conditional distibution of the sample $\boldsymbol{X}$ given the value of $T(\boldsymbol{X})$ does not depend on $\boldsymbol{\theta}$.

Likelihood Principle

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are two sample points such that $L(\boldsymbol{\theta}|\boldsymbol{x})$ is proportional to $L(\boldsymbol{\theta}|\boldsymbol{y})$, that is, there exists a constant $C(\boldsymbol{x}, \boldsymbol{y})$ such that

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = C(\boldsymbol{x}, \boldsymbol{y})L(\boldsymbol{\theta}|\boldsymbol{y}) \qquad \text{for all } \boldsymbol{\theta},$$

then the conclusions drawn from $\boldsymbol{x}$ and $\boldsymbol{y}$ should be identical.

Formal Sufficiency Principle

Experiment, E: triple $(\boldsymbol{X}, \boldsymbol{\theta}, \{f(\boldsymbol{x}|\boldsymbol{\theta})\})$.

$Ev(E, \boldsymbol{x})$: evidence about $\boldsymbol{\theta}$ arising from experiment E and $\boldsymbol{x}$.

Formal Sufficiency Principle: Consider

- Experiment $E = (\boldsymbol{X}, \boldsymbol{\theta}, \{f(\boldsymbol{x}|\boldsymbol{\theta})\})$, and
- $T(\boldsymbol{X})$ sufficient statistic for $\boldsymbol{\theta}$.

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are sample points satisfying $T(\boldsymbol{x}) = T(\boldsymbol{y})$, then

$$Ev(E, \boldsymbol{x}) = Ev(E, \boldsymbol{y})$$

Conditionality Principle

Let $E_i = (\boldsymbol{X}_i, \boldsymbol{\theta}, \{f_i(\boldsymbol{x}_i|\boldsymbol{\theta})\})$ for $i = 1, 2$ be two experiments, where only the unknown parameter $\boldsymbol{\theta}$ need be common between the two experiments.

Consider the mixed experiment in which the random variable J is observed, where $P(J = 1) = P(J = 2) = 1/2$ (independent of $\boldsymbol{\theta}$, $\boldsymbol{X}_1$, or $\boldsymbol{X}_2$), and the experiment E_J is performed. Formally, the performed experiment is

$$E^* = (\boldsymbol{X}^*, \boldsymbol{\theta}, \{f^*(\boldsymbol{x}^*|\boldsymbol{\theta})\}),$$

where $\boldsymbol{X}^* = (j, \boldsymbol{X}_j)$ and

$$f^*(\boldsymbol{x}^*|\boldsymbol{\theta}) = f^*((j, x_j)|\boldsymbol{\theta}) = \frac{1}{2} f_j(x_j|\boldsymbol{\theta}).$$

Then,

$$Ev(E^*, (j, x_j)) = Ev(E_j, x_j)$$

Conditionality Principle

The Conditionality Principle simply says that if one of two experiments is randomly chosen and the chosen experiment is done, yielding data $\boldsymbol{x}$, the information about $\boldsymbol{\theta}$ *depends only on the experiment performed*.

$\Rightarrow$ The Likelihood Principle can be derived from the Formal Sufficiency Principle and the Conditionality Principle

Formal Likelihood Principle

Consider the experiments $E_i = (\boldsymbol{X}_i, \boldsymbol{\theta}, \{f_i(\boldsymbol{x}_i|\boldsymbol{\theta})\})$ for $i = 1, 2$, where the unknown parameter $\boldsymbol{\theta}$ is the same in both experiments.

Suppose $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ are sample points from E_1 and E_2, respectively, such that

$$L(\boldsymbol{\theta}|\boldsymbol{x}_2^*) = CL(\boldsymbol{\theta}|\boldsymbol{x}_1^*)$$

for all $\boldsymbol{\theta}$ and for some constant C that may depend on $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ but not $\boldsymbol{\theta}$.

Then

$$Ev(E_1, \boldsymbol{x}_1^*) = Ev(E_1, \boldsymbol{x}_2^*)$$

Likelihood Principle Corolary

If

$$E = (\boldsymbol{X}, \theta, \{f(\boldsymbol{x}|\boldsymbol{\theta})\})$$

is an experiment, then $Ev(E, \boldsymbol{x})$ should depend on E and $\boldsymbol{x}$ only through $L(\boldsymbol{\theta}|\boldsymbol{x})$.

Birnbaum's Theorem: The Formal Likelihood Principle follows from the Formal Sufficiency Principle and the Conditionality Principle. The converse is also true.

Classical definition of sufficiency

Let $\boldsymbol{X}$ be a random quantity with probability density function (pdf) $p(\boldsymbol{x}|\boldsymbol{\theta})$.

Then, the statistic $\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{X})$ is sufficient for the parameter $\boldsymbol{\theta}$ if

$$p(\boldsymbol{x}|\boldsymbol{t}, \boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{t}).$$

Bayesian sufficiency

If $\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{X})$ is a sufficient statistic for $\boldsymbol{\theta}$, then

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = p(\boldsymbol{\theta}|\boldsymbol{t}), \text{ for all priors } p(\boldsymbol{\theta}).$$

Proof: $p(\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{x}, \boldsymbol{t}|\boldsymbol{\theta})$ if $\boldsymbol{t} = \boldsymbol{T}(\boldsymbol{x})$ and 0, if $\boldsymbol{t} \neq \boldsymbol{T}(\boldsymbol{X})$. So,

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{\theta}) &= p(\boldsymbol{x}|\boldsymbol{t}, \boldsymbol{\theta})p(\boldsymbol{t}|\boldsymbol{\theta}) \\ &= p(\boldsymbol{x}|\boldsymbol{t})p(\boldsymbol{t}|\boldsymbol{\theta}) \end{aligned}$$

But, by Bayes theorem,

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{x}) &\propto p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{t})p(\boldsymbol{t}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &\propto p(\boldsymbol{t}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|\boldsymbol{t}). \end{aligned}$$

□

Neyman's factorization

Definition (Bayesian): The statistics $\boldsymbol{T}(\boldsymbol{X})$ is sufficient for $\boldsymbol{\theta}$ if there is a function f such that

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{\theta}, \boldsymbol{t}).$$

Neyman's factorization: The statistic $\boldsymbol{T}$ is sufficient for $\boldsymbol{\theta}$ if and only if

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = f(\boldsymbol{t}, \boldsymbol{\theta})g(\boldsymbol{x})$$

where f and g are non-negative functions.

Proof: $(\Rightarrow)$ We have already seen that $p(\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{t})p(\boldsymbol{t}|\boldsymbol{\theta})$. Then it is enough to define

$$\begin{aligned} g(\boldsymbol{x}) &= p(\boldsymbol{x}|\boldsymbol{t}) = p(\boldsymbol{x}|\boldsymbol{T}(\boldsymbol{x})) \\ f(\boldsymbol{t}, \boldsymbol{\theta}) &= p(\boldsymbol{t}|\boldsymbol{\theta}) \end{aligned}$$

completing this part of the proof.

Neyman's factorization

($\Leftarrow$) Conversely, we have that $p(\boldsymbol{x}|\boldsymbol{\theta}) = f(\boldsymbol{t}, \boldsymbol{\theta})g(\boldsymbol{x})$. Defining $A_{\boldsymbol{t}} = \{\boldsymbol{x} : \boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{t}\}$, the pdf of $\boldsymbol{T}|\boldsymbol{\theta}$ is

$$p(\boldsymbol{t}|\boldsymbol{\theta}) \;=\; \int_{A_{\boldsymbol{t}}} p(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x} = f(\boldsymbol{t}, \boldsymbol{\theta}) \int_{A_{\boldsymbol{t}}} g(\boldsymbol{x})d\boldsymbol{x} = f(\boldsymbol{t}, \boldsymbol{\theta})G(\boldsymbol{x}),$$

and so, $f(\boldsymbol{t}, \boldsymbol{\theta}) = p(\boldsymbol{t}|\boldsymbol{\theta})/G(\boldsymbol{x})$. Also, from the theorem,

$$f(\boldsymbol{t}, \boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{\theta})/g(\boldsymbol{x}).$$

Equating the two forms for $f(\boldsymbol{t}, \boldsymbol{\theta})$ leads to

$$\frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{t}|\boldsymbol{\theta})} = \frac{g(\boldsymbol{x})}{G(\boldsymbol{x})}$$

Since $p(\boldsymbol{x}|\boldsymbol{t}, \boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{\theta})/p(\boldsymbol{t}|\boldsymbol{\theta})$, then

$$p(\boldsymbol{x}|\boldsymbol{t}, \boldsymbol{\theta}) = \frac{g(\boldsymbol{x})}{G(\boldsymbol{x})} = p(\boldsymbol{x}|\boldsymbol{t}).$$

Thus, $\boldsymbol{T}$ is sufficient for $\boldsymbol{\theta}$. $\square$