Principles of Data Reduction

\footnote{Based on Casella and Berger's (2002) \textit{Statistical Inference} and on Migon and Gamerman's (1999) \textit{Statistical Inference: An Integrated Approach}.}
Outline

Statistics

Sufficiency Principle

Likelihood Principle

Formal Sufficiency Principle

Conditionality Principle

Formal Likelihood Principle

Birnbaum’s Theorem

Bayesian sufficiency

Neyman’s factorization
Statistics: functions of the sample.

⇒ Any statistic, $T(X)$, defines a form of data reduction or data summary.

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

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

A statistic summarizes the data in that, rather than reporting the entire sample $x$, it reports only that $T(x) = t$ or, equivalently, $x \in A_t$. 
Sufficiency, likelihood and invariance principles

**Sufficiency Principle**
Does not discard information about $\theta$ 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 $\theta$ that is available from the sample.

**Invariance Principle**
Preserves some important features of the model.
Sufficiency Principle

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

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

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

\[
L(\theta|x) = C(x, y)L(\theta|y)
\]

for all \( \theta \),

then the conclusions drawn from \( x \) and \( y \) should be identical.
Formal Sufficiency Principle

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

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

Formal Sufficiency Principle: Consider

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

If $x$ and $y$ are sample points satisfying $T(x) = T(y)$, then

$$Ev(E, x) = Ev(E, y)$$
Conditionality Principle

Let \( E_i = (X_i, \theta, \{f_i(x_i|\theta)\}) \) for \( i = 1, 2 \) be two experiments, where only the unknown parameter \( \theta \) 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 \( \theta, X_1, \) or \( X_2 \)), and the experiment \( E_J \) is performed. Formally, the performed experiment is

\[
E^* = (X^*, \theta, \{f^*(x^*|\theta)\}),
\]

where \( X^* = (j, X_j) \) and

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

Then,

\[
Ev(E^*, (j, x_j)) = Ev(E_j, x_j)
\]
The Conditionality Principle simply says that if one of two experiments is randomly chosen and the chosen experiment is done, yielding data $x$, the information about $\theta$ 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 \( E_i = (X_i, \theta, \{f_i(x_i|\theta)\}) \) for \( i = 1, 2 \), where the unknown parameter \( \theta \) is the same in both experiments.

Suppose \( x_1^* \) and \( x_2^* \) are sample points from \( E_1 \) and \( E_2 \), respectively, such that

\[
L(\theta|x_2^*) = CL(\theta|x_1^*)
\]

for all \( \theta \) and for some constant \( C \) that may depend on \( x_1^* \) and \( x_2^* \) but not \( \theta \).

Then

\[
Ev(E_1, x_1^*) = Ev(E_1, x_2^*)
\]
Likelihood Principle Corollary

If

\[ E = (X, \theta, \{f(x|\theta)\}) \]

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

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

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

$$p(x|t, \theta) = p(x|t).$$
Bayesian sufficiency

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

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

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

$$p(x|\theta) = p(x|t, \theta)p(t|\theta)$$

$$= p(x|t)p(t|\theta)$$

But, by Bayes theorem,

$$p(\theta|x) \propto p(x|\theta)p(\theta) = p(x|t)p(t|\theta)p(\theta)$$

$$\propto p(t|\theta)p(\theta) \propto p(\theta|t).$$
Neyman’s factorization

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

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

\[ \text{Neyman’s factorization:} \quad \text{The statistic } T \text{ is sufficient for } \theta \text{ if and only if} \]

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

where $f$ and $g$ are non-negative functions.

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

$$g(x) = p(x|t) = p(x|T(x))$$

$$f(t, \theta) = p(t|\theta)$$

completing this part of the proof.
Neyman’s factorization

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

$$p(t|\theta) = \int_{A_t} p(x|\theta)dx = f(t, \theta) \int_{A_t} g(x)dx = f(t, \theta)G(x),$$

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

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

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

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

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

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

Thus, $T$ is sufficient for $\theta$. \qed