Note on the principle of data reduction

stma

June 11, 2024

1 Introduction

A set of principles, with respect to the data reduction, prescribe the use of statistical instruments to reduce the data prior to the inference of an unknown parameter θ, more or less without loss of nice properties such as accuracy, precision, robustness and so on. Although the concept sounds appealing and still remain the theoretical importance as its own right, the interest have been lessened in the descriptive statistics.

The concept dated back in early 1920's, at least when they became prominent. Fisher had contributed wide range of foundations in the context, especially first introduced the *Sufficiency Principle*, Birnbaum is known for *Likelihood Principle*, and Pitman for *Equivariance Principle*.

Since the statements can be formulated with a set of experiments on which a specific common parameter is estimated, we first introduce the definition of experiment.

Formally, a statistical experiment is composed of a triple $E = (\mathbf{X}, \theta, \{f(\mathbf{x} \mid \theta) \mid \theta \in \Theta\})$ such that

- $\mathbf{X}$ is a random vector together with the underlying sample space,
- θ is the (possibly vector valued) parameter to which an inference is made,
- $\{f(\mathbf{x} \mid \theta) \mid \theta \in \Theta\}$ is the θ-parameterized family of probability distributions in which an experimenter will expect to find a "true" distribution that $\mathbf{X}$ should follow.

Since much of our concern involve the identification of "inferential processes", we further assume that a statistical experiment includes a pair $\langle T, S \rangle$ of transformation T and estimator S, which gives rise to what the evidence of inference should be.

Before move on to the full exposition of so called principle of data reduction, one need to identify the experimental procedure on which a data reduction method can apply, as Figure 1 shows.

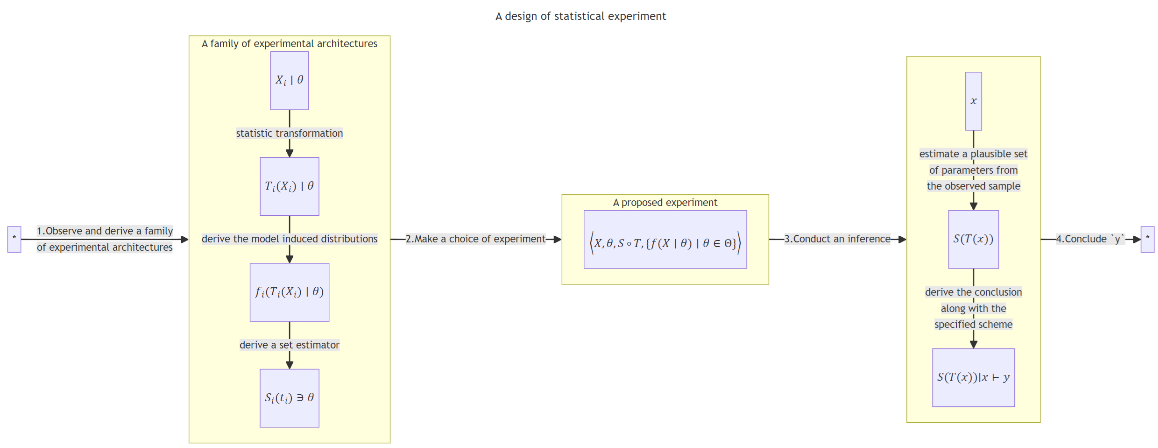

Figure 1: A design of statistical experiment

A scheme of experiment, generally being specified to constitute the experimental design, is implicit here because it complicates the overall story of data reduction. The scheme, of course, is significantly important that defines a set of rules and criterion that attest the sample validity and the credibility of estimated output; the acceptable degree of errors is one of such.

The Table 1 shows the three types of principles of data reduction, together with the statistical instruments employed to reduce the data.

Name	**Definition**	**Formal definition**	**Statistical instruments**
Sufficiency Principle	If $T(X)$ is a sufficient statistic for θ, then any inference about θ should depend 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 θ should be the same whether $X = x$ or $X = y$ is observed.	Consider experiment $E = (X, \theta, S \circ T, \{f(x \mid \theta)\})$ and suppose $T(X)$ is a sufficient statistic for θ. If x and y are sample points satisfying $T(x) = T(y)$, then $\mathrm{Ev}_S(E, x) = \mathrm{Ev}_S(E, y)$.	• (Minimal) Sufficient statistic • Complete statistic • Ancillary statistic
Likelihood Principle	If x and y are two sample points such that $L(\theta \mid x)$ is proportional to $L(\theta \mid y)$, that is, there exists a constant $C(x, y)$ such that $L(\theta \mid x) = C(x, y)L(\theta \mid y)\,\forall\theta$, then the conclusions drawn from x and y should be identical.	Suppose that we have two experiments, $E_1 = (X_1, \theta, \{f_1(x_1 \mid \theta)\})$ and $E_2 = (X_2, \theta, \{f_2(x_2 \mid \theta)\})$, where the unknown parameter θ is the same in both experiments. Suppose x_1^* and x_2^* are sample points from E_1 and E_2, respectively, such that $L(\theta \mid x_2^*) = CL(\theta \mid x_1^*)$ for all θ and some constant C that may depend on x_1^* and x_2^* but not θ. Then $\mathrm{Ev}(E_1, x_1^*) = \mathrm{Ev}(E_2, x_2^*)$.	• Likelihood function
Equivariance Principle	If $Y = g(X)$ is a change of measurement scale such that the model for Y has the same formal structure as the model for X, then an inference procedure should be both measurement equivariant and formally equivariant.	N/A	• Group of transformations

Table 1: Principles of data reduction

The *Formal Likelihood Principle* can be expressed in terms of *Formal Sufficiency Principle* together with the additional condition with respect to the experimental design, called *Conditionality Principle*, and vice verca, due to the following Birnbaum's theorem.

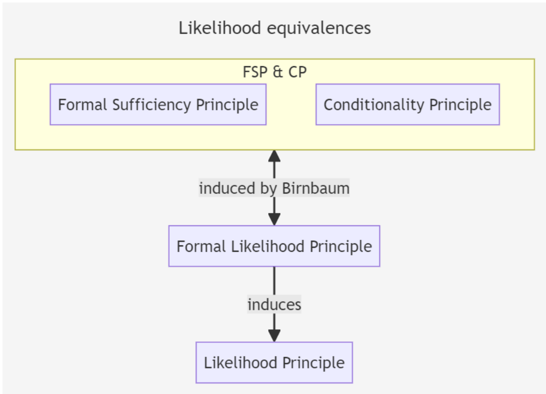

Figure 2: Schematic figure of Birnhaum's Theorem

Theorem 1. The Formal Likelihood Principle follows from the Formal Sufficiency Principle and the Conditionality Principle. The converse is also true. [1]

2 Caveat on the use of terms

We should be alerted that the definition of (statistical) experiment may vary in the contexts, as well as the evidence function, denoted Ev.

The evidence function, either not well defined or not to be formulated, needs the specified measure of "evidence" to be mathematically well-defined, while the interpretation of Ev is clear, i.e. the most likely, plausible, probable or reasonable conclusion about the concerned parameter inferred from the given sample against the experimental model.

3 Examples

3.1 Sufficiency Principle

For a series of discrete uniform random variables $X_1, \dots, X_n \sim u(1, \theta)$, $T(\mathbf{X}) = \max_i X_i$ is a sufficient statistic for θ. For the unbiased continuous estimator, we need $T(\mathbf{X}) = \frac{n+1}{n} \max_i X_i$.

To make a conclusion about the population from the given sample, if the median $u_{1/2}$ is larger than certain value for instance, we only need to check the maximum of the sample.

3.2 Likelihood Principle

Suppose $X \sim B(12, \theta)$, $Y \sim NB(3, \theta)$. Then $L(\theta \mid X = 3) = 220\theta^3(1-\theta^9)$ and $L(\theta \mid Y = 12) = 55\theta^3(1-\theta^9)$.

We may then want to prescribe the experiment of Y – number of Bernoulli trials to make 3 successes – to be relatively insignificant at the value of $Y = 12$, hence choose to do for X at the value of $X = 3$ instead.

3.3 Equivariance Principle

The set of binomial pmfs $\{f(X, n \mid \theta) \mid \theta \in \Theta\}$ is invariant under the group $\mathcal{G} = \{(id : X \mapsto X), (g : X \mapsto n - X)\}$, since the transformed random variable $Y = n - X$ follows $B(n, \theta)$, the same formal structure as before.

For a fixed estimator $S(X)$ of the success probability θ, we may prescribe that the equation $S(X) = 1 - S(n - X)$ holds, where we only need to investigate the estimates of $S(0), S(1), \dots, S([n/2])$.

In general, we can also describe the Equivariance Principle in terms of the following commutative diagram, where χ is a family of distributions parameterized by θ (cf. equivalently, a set of random variables that follow the distributions), $\mathcal{E}$ a set of estimators for $\theta' \in \Theta$, and ρ assigns an estimator to each distribution from which the estimator has been derived.

$$\begin{tikzcd} \chi \arrow[r, "g"] \arrow[d, "\rho"] & \chi \arrow[d, "\rho"] \\ \mathcal{E} \arrow[r, "g^*"] & \mathcal{E} \end{tikzcd} \qquad \begin{tikzcd} X \arrow[r, mapsto, "g"] \arrow[d, mapsto, "\rho"] & n - X \arrow[d, mapsto, "\rho"] \\ S(X) \arrow[r, mapsto, "g^*"] & 1 - S(X) = S(n - X) \end{tikzcd}$$

Note that the prescription of *Equivariance Principle* requires the invariance of the model under the group of transformations defined as follows.

Definition. Let $\mathcal{F} = \{f(\mathbf{x} \mid \theta) \mid \theta \in \Theta\}$ be a statistical model, i.e. a set of pdfs or pmfs for $\mathbf{X}$ in the specified form, and let $\mathcal{G}$ be a group of transformations of the sample space χ. The model $\mathcal{F}$ is *invariant under the group* $\mathcal{G}$ if $\forall\theta \in \Theta, \forall g \in \mathcal{G}, \exists! \theta' \in \Theta$ such that $\mathbf{Y} = g(\mathbf{X}) \sim f(\mathbf{y} \mid \theta')$ if $\mathbf{X} \sim f(\mathbf{x} \mid \theta)$.

Hence the diagram on the right-hand side makes sense since the transformation of a random variable determines the unique parameter, for which the corresponding estimator is uniquely derived by the transformation.

References

[1] George Casella and Roger Berger. *Statistical Inference*. Duxbury Resource Center, June 2001. ISBN: 0534243126.