Data reduction in statistical inference

TECHNISCHE UNIVERSITEIT EINDHOVEN
Faculteit Wiskunde en Informatica

Memorandum COSOR 92-03

Data reduction in statistical inference

E.E.M. van Berkum
H.N. Linssen
D.A. Overdijk

Eindhoven, February 1992
The Netherlands

Data reduction in statistical inference

E.E.M. van Berkum
H.N. Linssen
D.A. Overdijk

Department of Mathematics and Computing Science
University of Technology
P.O. Box 513, 5600 MB Eindhoven
The Netherlands

Abstract: Generalized definitions of the invariance, (partial) sufficiency and ancillarity principles are given in a measure-theoretic context. Data reduction in statistical inference is described in terms of these principles.

0. Introduction

The concepts of sufficiency and ancillarity in statistical inference were introduced by Fisher. Many authors contributed to the generalization of these concepts (see Bhapkar (1989, 1991) for references).

In this paper we treat the problem of data reduction in a measure-theoretic context, using generalized definitions of the sufficiency, ancillarity and invariance principles. The paper claims that every statistical inference problem can be reduced to an inference problem in a reference model, as described in section 9.

In section 1 we give the mathematical prerequisites that are used in the other sections. Section 2 describes in general terms the probability structure of the observational evidence. Section 3 and 4 describe its sufficient reduction and ancillary conditioning.

Statistical inference may have a given form, e.g. there may be a parameter of interest. This interest specification together with sufficiently reduced and conditioned probability structure constitute an inference model as described in section 5. The interest specification makes invariant reduction possible as defined in section 6. Invariant reduction as defined by Barnard (1963) is a special case. In section 7 and 8 the concepts of partially sufficient reduction and partially ancillary conditioning are introduced. These definitions generalize other definitions of partial sufficiency and ancillarity (see Bhapkar). In section 9 we propose a sequence in which the transformations described in the previous sections must be applied. Through this sequence of transformations every probability structure is changed into one or more reference models. Section 10 gives a number of examples.

1. Preliminaries

In this section we collect preliminaries for reference in subsequent sections. Therefore the reading of the paper can start with section two and the relevant parts of this section can be read when they are referenced.

1.1. Let λ be a map from a set X into a set Y. We refer to X as the domain and to Y as the codomain of λ. The powerset of X, i.e. the collection of subsets of X, is written as $P(X)$. The maps

$$\lambda_\rightarrow : P(X) \rightarrow P(Y) \,,$$

$$\lambda^\leftarrow : P(Y) \rightarrow P(X) \,,$$

are defined as

(1.1.1)
$$\lambda_\rightarrow(A) := \{\lambda(a) \in Y | a \in A\}, \quad A \subset X \,,$$

$$\lambda^\leftarrow(B) := \{b \in X | \lambda(b) \in B\}, \quad B \subset Y \,.$$

Let $\mathcal{F}$ be a σ-field of subsets of the codomain Y, i.e. $(Y, \mathcal{F})$ is a measurable space. We refer to

1

(1.1.2) $\sigma(\lambda) := \{\lambda^{\leftarrow}(B) \subset X | B \in \mathcal{F}\}$

as the σ-field of subsets of the domain X generated by the map λ from X into the measurable space $(Y, \mathcal{F})$.

1.2.   Let X be a set and let $\mathcal{A} \subset P(X)$ be a collection of subsets of X. The intersection of the σ-fields of subsets of X containing $\mathcal{A}$ is said to be the σ-field $\sigma(\mathcal{A})$ generated by $\mathcal{A}$.

1.3.   Let $(X, \mathcal{F})$ be a measurable space. For $A \subset X$ we write

(1.3.1) $\mathcal{F}|A := \{B \cap A | B \in \mathcal{F}\}$,

and we refer to the σ-field $\mathcal{F}|A$ on A as the trace of $\mathcal{F}$ on A.

1.4.   We now formulate a theorem for later use.

Theorem.   Let $\lambda : X \rightarrow (X_1, \mathcal{F}_1)$ be a surjection from a set X into a measurable space $(X_1, \mathcal{F}_1)$, and let $\mathcal{F}_0 \subset \sigma(\lambda)$ be a σ-field. We have

i $\lambda^{\leftarrow}\lambda_{\rightarrow}(A) = A$ for all $A \in \sigma(\lambda)$.

ii $\{\lambda_{\rightarrow}(A) | A \in \mathcal{F}_0\} \subset \mathcal{F}_1$ is a σ-field.

Proof.   Let $A \in \sigma(\lambda)$. There exists $B \in \mathcal{F}_1$ such that $A = \lambda^{\leftarrow}(B)$. Since the map λ is a surjection, it follows that

$$\lambda_{\rightarrow}\lambda^{\leftarrow}(B) = B .$$

and we can then conclude that

$$\lambda^{\leftarrow}\lambda_{\rightarrow}(A) = \lambda^{\leftarrow}\lambda_{\rightarrow}\lambda^{\leftarrow}(B) = \lambda^{\leftarrow}(B) = A ,$$

which proves i.

It is easily verified that

$$\{\lambda_{\rightarrow}(A) | A \in \mathcal{F}_0\} = \{B \in \mathcal{F}_1 | \lambda^{\leftarrow}(B) \in \mathcal{F}_0\} .$$

The collection on the right-hand side is a σ-field, which completes the proof of the theorem.

1.5.   Let X be a set and let I be an index set such that to every $i \in I$ there correspond a measurable space $(X_i, \mathcal{F}_i)$ and a map $\lambda_i : X \rightarrow X_i$. The σ-field on X generated by the collection

$$\Lambda := \{\lambda_i | i \in I\}$$

2

of maps on X is written as $\sigma(\Lambda)$ and

$$(1.5.1) \qquad \sigma(\Lambda) := \sigma\left(\bigcup_{i \in I} \sigma(\lambda_i)\right) ;$$

see 1.2 and (1.1.2).

1.6. Let $(X_1, \mathcal{F}_1)$ and $(X_2, \mathcal{F}_2)$ be two measurable spaces. A bijection $\lambda : X_1 \to X_2$ is said to be a measurable isomorphism between $(X_1, \mathcal{F}_1)$ and $(X_2, \mathcal{F}_2)$, if both λ and its inverse λ^{-1} are measurable. If $(X_1, \mathcal{F}_1) = (X_2, \mathcal{F}_2)$, then we refer to λ as a measurable automorphism on $(X_1, \mathcal{F}_1)$.
Let λ be a measurable isomorphism between $(X_1, \mathcal{F}_1)$ and $(X_2, \mathcal{F}_2)$. Note that the bijection $\lambda_\to : \mathcal{F}_1 \to \mathcal{F}_2$ satisfies

$$(1.6.1) \qquad \lambda_\to(A \cap B) = \lambda_\to(A) \cap \lambda_\to(B)$$

for all $A, B \in \mathcal{F}_1$.

1.7. Let δ be a measurable automorphism on the measurable space $(X, \mathcal{F})$; see 1.6. The σ-field

$$(1.7.1) \qquad J(\delta) := \{A \in \mathcal{F} | \delta_\to(A) = A\}$$

is called the σ-field of invariant sets under δ.

1.8. Let $(X_1, \mathcal{F}_1)$ and $(X_2, \mathcal{F}_2)$ be two measurable spaces. A bijection $\lambda : \mathcal{F}_1 \to \mathcal{F}_2$ is said to be an isomorphism between the σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$, if

$$(1.8.1) \qquad \lambda(A \cap B) = \lambda(A) \cap \lambda(B)$$

for all $A, B \in \mathcal{F}_1$. If $(X_1, \mathcal{F}_1) = (X_2, \mathcal{F}_2)$, then we refer to λ as an automorphism on the σ-field $\mathcal{F}_1$. It follows from (1.6.1) that a measurable isomorphism between $(X_1, \mathcal{F}_1)$ and $(X_2, \mathcal{F}_2)$ induces an isomorphism between the σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$. The converse, however, is not true in general.

1.9. Let γ be an automorphism on the σ-field $\mathcal{F}$ of a measurable space $(X, \mathcal{F})$; see 1.8. The σ-field

$$(1.9.1) \qquad J(\gamma) := \{A \in \mathcal{F} | \gamma(A) = A\}$$

is called the σ-field of invariant sets under γ.

1.10. Let $(X, \mathcal{F})$ be a measurable space. The set of probability measures on $\mathcal{F}$ is denoted by $\widehat{\mathcal{F}}$. For $A \in \mathcal{F}$ consider the map

$$(1.10.1) \quad \lambda_A(p) := p(A) \in [0,1], \quad p \in \widehat{\mathcal{F}}$$

from $\widehat{\mathcal{F}}$ into the unit interval. The σ-field of Borel subsets of the unit interval is written as $\mathcal{B}[0,1]$. The σ-field generated by the collection

$$\{\lambda_A | A \in \mathcal{F}\}$$

of maps from $\widehat{\mathcal{F}}$ into the measurable space $([0,1], \mathcal{B}[0,1])$ is denoted by $\overline{\mathcal{F}}$; see (1.5.1). We refer to $(\widehat{\mathcal{F}}, \overline{\mathcal{F}})$ as the space of probability measures on the σ-field $\mathcal{F}$. Let $\mathcal{P} \subset \widehat{\mathcal{F}}$ be a collection of probability measures on $\mathcal{F}$. We refer to $(\mathcal{P}, \overline{\mathcal{F}}|\mathcal{P})$ as a space of probability measures on $\mathcal{F}$; see (1.3.1).

1.11. Let $(X, \mathcal{F})$ be a measurable space and let $\mathcal{F}_0 \subset \mathcal{F}$ be a σ-field. Furthermore, let $(\mathcal{P}, \overline{\mathcal{F}}|\mathcal{P})$ be a space of probability measures on $\mathcal{F}$; see 1.10. The marginal probability measure on $\mathcal{F}_0$ corresponding to $p \in \mathcal{P}$ is denoted by $\varphi(p)$. We refer to the map

$$(1.11.1) \quad \varphi : (\mathcal{P}, \overline{\mathcal{F}}|\mathcal{P}) \to (\widehat{\mathcal{F}}_0, \overline{\mathcal{F}}_0)$$

as the marginalization map on $\mathcal{P}$ corresponding to $\mathcal{F}_0$. Here $(\widehat{\mathcal{F}}_0, \overline{\mathcal{F}}_0)$ is the space of probability measures in $\mathcal{F}_0$; see 1.10. The σ-field $\sigma(\varphi)$ generated by the map φ from $\mathcal{P}$ into $(\widehat{\mathcal{F}}_0, \overline{\mathcal{F}}_0)$ is the smallest σ-field on $\mathcal{P}$ such that the map

$$\lambda_A(p) := p(A), \quad p \in \mathcal{P} ,$$

from $\mathcal{P}$ into $([0,1], \mathcal{B}[0,1])$ is measurable for all $A \in \mathcal{F}_0$.
Hence,

$$(1.11.2) \quad \sigma(\varphi) \subset \overline{\mathcal{F}}|\mathcal{P} .$$

Let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}$ be two σ-fields. The marginalization maps on $\mathcal{P}$ corresponding to $\mathcal{F}_0$ and $\mathcal{F}_1$ are written as φ_0 and φ_1 respectively. We have

$$(1.11.3) \quad \sigma(\varphi_0) \subset \sigma(\varphi_1) .$$

2. Probability structure

Consider an experiment. The set of possible outcomes of the experiment is denoted by Ω. The σ-field of events on Ω is written as Σ. The measurable space (Ω, Σ) is said to be the sample space of the experiment. Let $(\widehat{\Sigma}, \overline{\Sigma})$ be the space of probability measures on Σ; see 1.10. The probability distribution on Σ corresponding to the outcome of the experiment is not known. However, a subset $\mathcal{P} \subset \widehat{\Sigma}$ is given such that the probability distribution of the outcome of the experiment is in the set $\mathcal{P}$. The space $(\mathcal{P}, \overline{\Sigma}|\mathcal{P})$ of probability measures on Σ is referred to as the probability model for the outcome of the experiment; see 1.10.

Consider the σ-ring

4

$$(2.1) \quad \mathcal{N} := \{A \in \Sigma | p(A) = 0 \text{ for all } p \in \mathcal{P}\}$$

of the so-called negligible subsets of the sample space. All equalities and inclusions between sets in Σ have to be interpreted modulo $\mathcal{N}$, i.e. for all $A, B \in \Sigma$

$$(2.2) \quad A \subset B :\Leftrightarrow A \setminus B \in \mathcal{N} .$$

The two measurable spaces

> sample space: $\quad (\Omega, \Sigma)$,
>
> probability model: $\quad (\mathcal{P}, \overline{\Sigma} | \mathcal{P})$

are said to constitute the probability structure of the experiment.

3. Sufficiency

Let (Ω, Σ) be the sample space and $(\mathcal{P}, \overline{\Sigma} | \mathcal{P})$ the probability model of a probability structure. Consider a statistic S on the sample space, i.e. a measurable map from (Ω, Σ) into a measurable space (Ω_1, Σ_1). The σ-field on Ω generated by S is denoted by $\sigma(S) \subset \Sigma$; see (1.1.2). The marginalization map from $\mathcal{P}$ into the space $(\widehat{\sigma(S)}, \overline{\sigma(S)})$ of probability measures on $\sigma(S)$ is written as φ; see 1.11. The conditional probability measure on Σ given $S = \omega_1 \in \Omega_1$ corresponding to $p \in \mathcal{P}$ is denoted by $\psi(\omega_1, p)$ and for $\omega_1 \in \Omega_1$ we write

$$(3.1) \quad \mathcal{P}(\omega_1) := \{p \in \mathcal{P} | \psi(\omega_1, p) \text{ exists}\} .$$

For fixed $\omega_1 \in \Omega_1$ we consider the map

$$(3.2) \quad \psi(\omega_1, \cdot) : (\mathcal{P}(\omega_1), \overline{\Sigma} | \mathcal{P}(\omega_1)) \rightarrow (\widehat{\Sigma}, \overline{\Sigma}) ,$$

where $(\widehat{\Sigma}, \overline{\Sigma})$ is the space of probability measures on Σ. The σ-field on $\mathcal{P}(\omega_1)$ generated by $\psi(\omega_1, \cdot)$ is denoted by $\sigma(\psi(\omega_1, \cdot))$.

The statistic S is said to be sufficient, if the conditional probability measure $\psi(\omega_1, p)$ is independent of $p \in \mathcal{P}(\omega_1)$ for all $\omega_1 \in \Omega_1$, i.e.

$$(3.3) \quad \sigma(\psi(\omega_1, \cdot)) = \{\emptyset, \mathcal{P}(\omega_1)\}$$

for all $\omega_1 \in \Omega_1$. The sufficient statistic S is said to be minimal sufficient, if for all sufficient statistics S' on (Ω, Σ) we have

$$(3.4) \quad \sigma(S') \subset \sigma(S) \Rightarrow \sigma(S') = \sigma(S) .$$

In general there exists a unique minimal sufficient statistic S. Let S be the unique minimal sufficient statistic. So for every sufficient statistic S' we have

5

(3.5) $\quad \sigma(S) \subset \sigma(S')$.

We now transform the given probability structure and we refer to this transformation as the sufficient reduction of the probability structure under consideration.
The sample space of the new probability structure is

(3.6) $\quad (\Omega, \sigma(S))$.

The probability model of the new probability structure can be written as

(3.7) $\quad (\mathcal{P}_1, \overline{\sigma(S)}|\mathcal{P}_1)$,

where

$$\mathcal{P}_1 := \{\varphi(p) \in \widehat{\sigma(S)} | p \in \mathcal{P}\} .$$

The conditional probability measure $\psi(\omega_1, p)$ is independent of $p \in \mathcal{P}$ and therefore the map

(3.8) $\quad \varphi : (\mathcal{P}, \overline{\Sigma}|\mathcal{P}) \to (\mathcal{P}_1, \overline{\sigma(S)}|\mathcal{P}_1)$

is a measurable isomorphism; see 1.6.
If

(3.9) $\quad \sigma(S) = \Sigma$,

then the sufficient reduction of the probability structure is said to be trivial.
For an illustration of the concepts in this section we refer to example 10.1.

4. Conditioning

Let (Ω, Σ) be the sample space and $(\mathcal{P}, \overline{\Sigma}|\mathcal{P})$ the probability model of a probability structure. Consider a statistic C from the sample space into a measurable space (Ω_1, Σ_1). The σ-field on Ω generated by C is denoted by $\sigma(C) \subset \Sigma$; see (1.1.2). The marginalization map from $\mathcal{P}$ into the space $(\widehat{\sigma(C)}, \overline{\sigma(C)})$ of probability measures on $\sigma(C)$ is written as φ; see 1.11. The σ-field on $\mathcal{P}$ generated by φ is written as $\sigma(\varphi) \subset \overline{\Sigma}|\mathcal{P}$. The statistic C is said to be an ancillary statistic, if the probability distribution of C is independent of $p \in \mathcal{P}$, i.e.

(4.1) $\quad \sigma(\varphi) = \{\emptyset, \mathcal{P}\}$.

The ancillary statistic C is said to be a maximal ancillary statistic, if for all ancillary statistics C' on (Ω, Σ) we have

(4.2) $\quad \sigma(C) \subset \sigma(C') \Rightarrow \sigma(C) = \sigma(C')$.

In general there does not exist a unique maximal ancillary statistic.

Let C be a maximal ancillary statistic from (Ω, Σ) into (Ω_1, Σ_1). The conditional probability measure on Σ given $C = \omega_1 \in \Omega_1$ corresponding to $p \in \mathcal{P}$ is denoted by $\psi(\omega_1, p)$ and for $\omega_1 \in \Omega_1$ we write

$$(4.3) \quad \mathcal{P}(\omega_1) := \{p \in \mathcal{P} | \psi(\omega_1, p) \text{ exists}\} \ .$$

For fixed $\omega_1 \in \Omega_1$ we consider the map

$$(4.4) \quad \psi(\omega_1, \cdot) : (\mathcal{P}(\omega_1), \overline{\Sigma} | \mathcal{P}(\omega_1)) \to (\widehat{\Sigma}, \overline{\Sigma}) \ ,$$

where $(\widehat{\Sigma}, \overline{\Sigma})$ is the space of probability measures on Σ.

The probability distribution of the maximal ancillary statistic C is independent of $p \in \mathcal{P}$ and therefore we can consider the experiment as a mixture of experiments corresponding to the possible values of C in Ω_1. So we transform, for every $\omega_1 \in \Omega_1$, the given probability structure in the following way.
Fix $\omega_1 \in \Omega_1$. The sample space of the new probability structure is

$$(4.5) \quad (\Omega, \Sigma) \ .$$

The probability model of the new probability structure can be written as

$$(4.6) \quad (\mathcal{P}^{\omega_1}, \overline{\Sigma} | \mathcal{P}^{\omega_1}) \ ,$$

where

$$\mathcal{P}^{\omega_1} := \{\psi(\omega_1, p) \in \widehat{\Sigma} | p \in \mathcal{P}(\omega_1)\} \ .$$

If

$$(4.7) \quad \sigma(C) = \{\emptyset, \Omega\} \ ,$$

then the maximal ancillary statistic C is said to be trivial. For an illustration of the concepts in this section we refer to the examples 10.2 and 10.3.

5. Inference model

Let (Ω, Σ) be the sample space and let $(\mathcal{P}, \overline{\Sigma} | \mathcal{P})$ be the probability model of a probability structure such that the sufficient reduction is trivial, i.e. for every sufficient statistic S on (Ω, Σ) we have

$$(5.1) \quad \sigma(S) = \Sigma \ ,$$

and every ancillary statistic is trivial, i.e. for every ancillary statistic C on (Ω, Σ) we have

(5.2) $\quad \sigma(C) = \{\emptyset, \Omega\}$.

Let p_0 be the probability distribution of the outcome of the experiment. It may be that one is interested only in a specific aspect of p_0, i.e. a σ-field $\mathcal{R} \subset \overline{\Sigma}|\mathcal{P}$ is specified such that every inferential statement can be written as $p_0 \in A$ with $A \in \mathcal{R}$. We refer to $\mathcal{R}$ as the σ-field of interest.

The map (4.4) is not a measurable isomorphism in general, and therefore the initial specification of the σ-field $\mathcal{R}$ of interest must take place in a probability structure satisfying (5.2).
The triple

$$\begin{aligned}
&\text{sample space:} \quad (\Omega, \Sigma) , \\
&\text{probability model: } (\mathcal{P}, \overline{\Sigma}|\mathcal{P}) , \\
&\sigma\text{-field of interest: } \mathcal{R} \subset \overline{\Sigma}|\mathcal{P} ,
\end{aligned}$$

is said to constitute an inference model for the experiment.

6. Invariance

Let (Ω, Σ) be the sample space, $(\mathcal{P}, \overline{\Sigma}|\mathcal{P})$ the probability model and $\mathcal{R} \subset \overline{\Sigma}|\mathcal{P}$ the σ-field of interest of an inference model. The set of automorphisms of the σ-field Σ is denoted by Γ; see 1.8, and the set of measurable automorphisms of $(\mathcal{P}, \overline{\Sigma}|\mathcal{P})$ by Δ; see 1.6. Introduce the set

(6.1) $\quad V := \{(\gamma, \delta) \in \Gamma \times \Delta | \delta(p)(\gamma(A)) = p(A) \text{ for all } p \in \mathcal{P}, A \in \Sigma\}$.

For $(\gamma, \delta) \in V$ let $J(\gamma) \subset \Sigma$ be the σ-field of invariant sets under γ, and let $J(\delta) \subset \overline{\Sigma}|\mathcal{P}$ be the σ-field of invariant sets under δ; see (1.9.1) and (1.7.1). The marginalization map on $\mathcal{P}$ corresponding to $J(\gamma)$ is written as φ_γ, i.e. for $(\gamma, \delta) \in V$

(6.2) $\quad \varphi_\gamma : (\mathcal{P}, \overline{\Sigma}|\mathcal{P}) \to (\widehat{J(\gamma)}, \overline{J(\gamma)})$,

where $(\widehat{J(\gamma)}, \overline{J(\gamma)})$ is the space of probability measures on $J(\gamma)$; see 1.11. The σ-field on $\mathcal{P}$ generated by φ_γ is denoted by $\sigma(\varphi_\gamma) \subset \overline{\Sigma}|\mathcal{P}$.

Consider a statistic I on the sample space. The σ-field on Ω generated by I is denoted by $\sigma(I) \subset \Sigma$; see (1.1.2). The marginalization map from $\mathcal{P}$ into the space $(\widehat{\sigma(I)}, \overline{\sigma(I)})$ of probability measures on $\sigma(I)$ is written as φ; see 1.11. The σ-field on $\mathcal{P}$ generated by φ is denoted by $\sigma(\varphi) \subset \overline{\Sigma}|\mathcal{P}$. The statistic I is said to be invariant, if there exists a set $B \subset V$ such that for all $(\gamma, \delta) \in B$

(6.3) $\quad \sigma(I) \subset J(\gamma)$ and $\mathcal{R} \subset \sigma(\varphi)$.

8

Let I be an invariant statistic. We show that

$$(6.4) \qquad \mathcal{R} \subset \sigma(\varphi) \subset \sigma(\varphi_\gamma) \subset J(\delta) \subset \overline{\Sigma}|\mathcal{P}$$

for all $(\gamma, \delta) \in B$. We prove (6.4).
For $(\gamma, \delta) \in V, A \in J(\gamma)$ and $0 \le x \le 1$ consider

$$W := \{ p \in \mathcal{P} | p(A) \le x \} \in \sigma(\varphi_\gamma) \ .$$

For $p \in W$ we get

$$\delta(p)(A) = \delta(p)(\gamma(A)) = p(A) \le x \ ,$$

$$\delta^{-1}(p)(A) = \delta^{-1}(p)(\gamma^{-1}(A)) = p(A) \le x \ .$$

We conclude that

$$W \in J(\delta) \ ,$$

and therefore

$$(6.5) \qquad \sigma(\varphi_\gamma) \subset J(\delta) \text{ for all } (\gamma, \delta) \in V \ .$$

The statement (6.4) now follows from (6.3), (6.5) and (1.11.3).

The invariant statistic I is said to be minimal invariant, if for all invariant statistics I' we have

$$(6.6) \qquad \sigma(I') \subset \sigma(I) \Rightarrow \sigma(I') = \sigma(I) \ .$$

We now discuss the existence of a unique invariant statistic. Introduce the set

$$(6.7) \qquad Q := \{ (\gamma, \delta) \in V | \mathcal{R} \subset \sigma(\varphi_\gamma) \} \ ,$$

and the σ-field

$$(6.8) \qquad \Sigma_m := \bigcap_{(\gamma, \delta) \in Q} J(\gamma) \subset \Sigma \ .$$

Let I_m be a statistic such that

$$(6.9) \qquad \sigma(I_m) = \Sigma_m \ ,$$

and let φ_m be the marginalization map from $\mathcal{P}$ into the space $(\widehat{\Sigma}_m, \overline{\Sigma}_m)$ of probability measures on Σ_m. The σ-field on $\mathcal{P}$ generated by φ_m is denoted by $\sigma(\varphi_m) \subset \overline{\Sigma}|\mathcal{P}$. We have

9

$$(6.10) \quad \sigma(\varphi_m) \subset \bigcap_{(\gamma,\delta)\in Q} \sigma(\varphi_\gamma) \, .$$

We now conclude the following.
If I_m is invariant, i.e.

$$(6.11) \quad \mathcal{R} \subset \sigma(\varphi_m) \, ,$$

then I_m is the unique minimal invariant statistic.

We transform the given inference model and refer to this transformation as the invariant reduction of the inference model under consideration.
Let I be a minimal invariant statistic on the sample space. The sample space of the new inference model is

$$(6.12) \quad (\Omega, \sigma(I)) \, .$$

The probability model of the new inference model can be written as

$$(6.13) \quad (\mathcal{P}_1, \overline{\sigma(I)}|\mathcal{P}_1) \, ,$$

where

$$\mathcal{P}_1 := \{\varphi(p) \in \widehat{\sigma(I)} | p \in \mathcal{P}\} \, .$$

For $A \in \mathcal{R} \subset \sigma(\varphi)$ we have by use of theorem 1.4

$$\varphi^{\leftarrow}\varphi_{\rightarrow}(A) = A \, ,$$

and

$$\mathcal{R}_1 := \{\varphi_{\rightarrow}(A) | A \in \mathcal{R}\} \text{ is a } \sigma-\text{field on } \mathcal{P}_1 \, .$$

Therefore the σ-field $\mathcal{R}_1 \subset \overline{\sigma(I)}|\mathcal{P}_1$ is the σ-field of interest in the new inference model. If

$$(6.14) \quad \sigma(I) = \Sigma \, ,$$

then the invariant reduction of the inference model is said to be trivial.
For an illustration of the concepts in this section we refer to the examples 10.4 and 10.5.

10

7. Partial sufficiency

Let (Ω, Σ) be the sample space, $(\mathcal{P}, \overline{\Sigma}|\mathcal{P})$ the probability model and $\mathcal{R} \subset \overline{\Sigma}|\mathcal{P}$ the σ-field of interest of an inference model. Consider a statistic R from (Ω, Σ) into a measurable space (Ω_1, Σ_1). The σ-field on Ω generated by R is denoted by $\sigma(R) \subset \Sigma$; see (1.1.2). The marginalization map from $\mathcal{P}$ into the space $(\widehat{\sigma(R)}, \overline{\sigma(R)})$ of probability measures on $\sigma(R)$ is written as φ; see 1.11. The σ-field on $\mathcal{P}$ generated by φ is written as $\sigma(\varphi) \subset \overline{\Sigma}|\mathcal{P}$. The conditional probability measure on Σ given $R = \omega_1 \in \Omega_1$ corresponding to $p \in \mathcal{P}$ is denoted by $\psi(\omega_1, p)$ and for $\omega_1 \in \Omega_1$ we write

$$(7.1) \qquad \mathcal{P}(\omega_1) := \{p \in \mathcal{P}|\psi(\omega_1, p) \text{ exists}\} \ .$$

For fixed $\omega_1 \in \Omega_1$ we consider the map

$$(7.2) \qquad \psi(\omega_1, \cdot) : (\mathcal{P}(\omega_1), \overline{\Sigma}|\mathcal{P}(\omega_1)) \to (\widehat{\Sigma}, \overline{\Sigma}) \ ,$$

where $(\widehat{\Sigma}, \overline{\Sigma})$ is the space of probability measures on Σ. The σ-field on $\mathcal{P}(\omega_1)$ generated by $\psi(\omega_1, \cdot)$ is denoted by $\sigma(\psi(\omega_1, \cdot))$.

The statistic R is said to be partially sufficient if the interesting aspect of $p \in \mathcal{P}$ is a function of the marginal probability distribution $\varphi(p)$, i.e.

$$(7.3) \qquad \mathcal{R} \subset \sigma(\varphi) \ ,$$

and the conditional probability distribution $\psi(\omega_1, p)$ is independent of the marginal probability distribution $\varphi(p)$ for all $\omega_1 \in \Omega_1$, i.e.

$$(7.4) \qquad \sigma(\psi(\omega_1, \cdot)) \cap \sigma(\varphi)|\mathcal{P}(\omega_1) = \{\emptyset, \mathcal{P}(\omega_1)\}$$

for all $\omega_1 \in \Omega_1$.
The partially sufficient statistic R is said to be minimal partially sufficient, if for all partially sufficient statistics R' on (Ω, Σ) we have

$$(7.5) \qquad \sigma(R') \subset \sigma(R) \Rightarrow \sigma(R') = \sigma(R) \ .$$

We conjecture that there exists a unique minimal partially sufficient statistic in general.
Let R be a minimal partially sufficient statistic. We now transform the given inference model and refer to this transformation as the partially sufficient reduction of the inference model under consideration.
The sample space of the new inference model is

$$(7.6) \qquad (\Omega, \sigma(R)) \ .$$

The probability model of the new inference model can be written as

11

(7.7) $\quad (\mathcal{P}_1, \overline{\sigma(R)}|\mathcal{P}_1)$,

where

$$\mathcal{P}_1 := \{\varphi(p) \in \widehat{\sigma(R)} | p \in \mathcal{P}\} \ .$$

For $A \in \mathcal{R} \subset \sigma(\varphi)$ we have by use of theorem 1.4

$$\varphi^{\leftarrow}\varphi_{\rightarrow}(A) = A \ ,$$

and

$$\mathcal{R}_1 := \{\varphi_{\rightarrow}(A) | A \in \mathcal{R}\} \text{ is a } \sigma\text{-field on } \mathcal{P}_1 \ .$$

Therefore the σ-field $\mathcal{R}_1 \subset \overline{\sigma(R)}|\mathcal{P}_1$ is the σ-field of interest in the new inference model. If

(7.8) $\quad \sigma(R) = \Sigma$,

then the partially sufficient reduction of the inference model is said to be trivial. For an illustration of the concepts in this section we refer to the examples 10.6 and 10.7.

8. Partial conditioning

Let (Ω, Σ) be the sample space, $(\mathcal{P}, \overline{\Sigma}|\mathcal{P})$ the probability model and $\mathcal{R} \subset \overline{\Sigma}|\mathcal{P}$ the σ-field of interest of an inference model. Consider a statistic A from (Ω, Σ) into a measurable space (Ω_1, Σ_1). The σ-field on Ω generated by A is denoted by $\sigma(A) \subset \Sigma$; see (1.1.2). The marginalization map from $\mathcal{P}$ into the space $(\widehat{\sigma(A)}, \overline{\sigma(A)})$ of probability measures on $\sigma(A)$ is written as φ; see 1.11. The σ-field on $\mathcal{P}$ generated by φ is written as $\sigma(\varphi) \subset \overline{\Sigma}|\mathcal{P}$. The conditional probability measure on Σ given $A = \omega_1 \in \Omega_1$ corresponding to $p \in \mathcal{P}$ is denoted by $\psi(\omega_1, p)$ and for $\omega_1 \in \Omega_1$ we write

(8.1) $\quad \mathcal{P}(\omega_1) := \{p \in \mathcal{P} | \psi(\omega_1, p) \text{ exists}\}$.

For fixed $\omega_1 \in \Omega_1$ we consider the map

(8.2) $\quad \psi(\omega_1, \cdot) : (\mathcal{P}(\omega_1), \overline{\Sigma}|\mathcal{P}(\omega_1)) \to (\widehat{\Sigma}, \overline{\Sigma})$,

where $(\widehat{\Sigma}, \overline{\Sigma})$ is the space of probability measures on Σ. The σ-field on $\mathcal{P}(\omega_1)$ generated by $\psi(\omega_1, \cdot)$ is denoted by $\sigma(\psi(\omega_1, \cdot))$.

The statistic A is said to be a partially ancillary statistic, if the marginal probability distribution $\varphi(p)$ is independent of the interesting aspect of $p \in \mathcal{P}$, i.e.

(8.3) $\quad \mathcal{R} \cap \sigma(\varphi) = \{\emptyset, \mathcal{P}\}$,

and the interesting aspect of $p \in \mathcal{P}$ is a function of the conditional probability distribution $\psi(\omega_1, p)$ for all $\omega_1 \in \Omega_1$, i.e.

$$(8.4) \qquad \mathcal{R}|\mathcal{P}(\omega_1) \subset \sigma(\psi(\omega_1, \cdot))$$

for all $\omega_1 \in \Omega_1$.

The partially ancillary statistic A is said to be a maximal partially ancillary statistic, if for all partially ancillary statistics A' on (Ω, Σ) we have

$$(8.5) \qquad \sigma(A) \subset \sigma(A') \Rightarrow \sigma(A) = \sigma(A') .$$

In general there does not exist a unique maximal partially ancillary statistic.

Let A be a maximal partially ancillary statistic from (Ω, Σ) into (Ω_1, Σ_1). For every $\omega_1 \in \Omega_1$ we transform the inference model in the following way.

Fix $\omega_1 \in \Omega_1$. The sample space of the new inference model is

$$(8.6) \qquad (\Omega, \Sigma) .$$

The probability model of the new inference model can be written as

$$(8.7) \qquad (\mathcal{P}^{\omega_1}, \overline{\Sigma}|\mathcal{P}^{\omega_1}) ,$$

where

$$\mathcal{P}^{\omega_1} := \{\psi(\omega_1, p) \in \widehat{\Sigma} | p \in \mathcal{P}(\omega_1)\} .$$

For $A \in \mathcal{R}|\mathcal{P}(\omega_1) \subset \sigma(\psi(\omega_1, \cdot))$ we have by use of theorem 1.4

$$\psi^{\leftarrow}(\omega_1, \cdot)\psi_{\rightarrow}(\omega_1, \cdot)(A) = A ,$$

and

$$\mathcal{R}_1^{\omega_1} := \{\psi_{\rightarrow}(\omega_1, \cdot)(A) | A \in \mathcal{R}|\mathcal{P}(\omega_1)\}$$

is a σ-field on $\mathcal{P}^{\omega_1}$. Therefore the σ-field $\mathcal{R}_1^{\omega_1} \subset \overline{\Sigma}|\mathcal{P}^{\omega_1}$ is the σ-field of interest in the new inference model.

If

$$(8.8) \qquad \sigma(A) = \{\emptyset, \Omega\} ,$$

then the partially ancillary statistic A is said to be trivial.

For an illustration of the concepts in this section we refer to the examples 10.8 and 10.9.

13

9. Reference model

In this section we propose a sequence in which the transformations described in the previous sections must be performed. We refer to this scheme of data reduction as the SCIRA data reduction, where

<div style="text-align:center">

S: sufficiency,
C: conditioning,
I: invariance,
R: partial sufficiency,
A: partial conditioning.

</div>

Let (Ω, Σ) be the sample space and $(\mathcal{P}, \overline{\Sigma}|\mathcal{P})$ the probability model of a probability structure. By use of the unique minimal sufficient statistic S we transform the probability structure $[(\Omega, \Sigma), (\mathcal{P}, \overline{\Sigma}|\mathcal{P})]$ as described in section 3.

Let (Ω_S, Σ_S) be the sample space and $(\mathcal{P}_S, \overline{\Sigma}_S|\mathcal{P}_S)$ the probability model of the new probability structure. Hence, for every sufficient statistic S on (Ω_S, Σ_S) we have

$$(9.1) \qquad \sigma(S) = \Sigma_S \; ;$$

see (3.9). By use of a maximal ancillary statistic C we transform the probability structure $[(\Omega_S, \Sigma_S), (\mathcal{P}_S, \overline{\Sigma}_S|\mathcal{P}_S)]$ as described in section 4.

Let (Ω_C, Σ_C) be the sample space and $(\mathcal{P}_C, \overline{\Sigma}_C|\mathcal{P}_C)$ the probability model of the new probability structure. Hence, for every ancillary statistic C on (Ω_C, Σ_C) we have

$$(9.2) \qquad \sigma(C) = \{\emptyset, \Omega_C\} \; ;$$

see (4.7) and for every sufficient statistic S on (Ω_C, Σ_C) we have

$$(9.3) \qquad \sigma(S) = \Sigma_C \; ; \qquad .$$

see (5.2) and (5.1). We now specify the σ-field $\mathcal{R} \subset \overline{\Sigma}_C|\mathcal{P}_C$ of interest. The triple $[(\Omega_C, \Sigma_C), (\mathcal{P}_C, \overline{\Sigma}_C|\mathcal{P}_C), \mathcal{R}]$ constitutes the inference model under consideration; see section 5. By use of a minimal invariant statistic I on (Ω_C, Σ_C) we transform the inference model under consideration as described in section 6.

Let (Ω_I, Σ_I) be the sample space, $(\mathcal{P}_I, \overline{\Sigma}_I|\mathcal{P}_I)$ the probability model and $\mathcal{R}_I \subset \overline{\Sigma}_I|\mathcal{P}_I$ the σ-field of interest of the new inference model. As illustrated by example 10.4, there may exist a nontrivial invariant statistic on (Ω_I, Σ_I). In that case the invariant reduction of the inference model $[(\Omega_I, \Sigma_I), (\mathcal{P}_I, \overline{\Sigma}_I|\mathcal{P}_I), \mathcal{R}_I]$ as described in section 6 must be applied. Now suppose that every invariant statistic on (Ω_I, Σ_I) is trivial. By use of a minimal partially sufficient statistic R on (Ω_I, Σ_I) we transform the inference model $[(\Omega_I, \Sigma_I), (\mathcal{P}_I, \overline{\Sigma}_I|\mathcal{P}_I), \mathcal{R}_I]$ as described in section 7.

Let (Ω_R, Σ_R) be the sample space, $(\mathcal{P}_R, \overline{\Sigma}_R|\mathcal{P}_R)$ the probability model and $\mathcal{R}_R \subset$

$\overline{\Sigma}_R|\mathcal{P}_R$ the σ-field of interest of the new inference model. Suppose that every invariant or partially sufficient statistic on (Ω_R, Σ_R) is trivial. If this is not true, then we start with the invariant reduction again. By use of a maximal partially ancillary statistic A we tranform the inference model $[(\Omega_R, \Sigma_R), (\mathcal{P}_R, \overline{\Sigma}_R|\mathcal{P}_R), \mathcal{R}_R]$ as described in section 8.

Let (Ω_A, Σ_A) be the sample space, $(\mathcal{P}_A, \overline{\Sigma}_A|\mathcal{P}_A)$ the probability model and $\mathcal{R}_A \subset \overline{\Sigma}_A|\mathcal{P}_A$ the σ-field of interest of the new inference model. If every invariant or partially sufficient or partially ancillary statistic on (Ω_A, Σ_A) is trivial, then the inference model $[(\Omega_A, \Sigma_A), (\mathcal{P}_A, \overline{\Sigma}_A|\mathcal{P}_A), \mathcal{R}_A]$ is said to be a reference model. If the new inference model is not a reference model, then we start with the invariant reduction again. Hence, the SCIRA data reduction ultimately yields a reference model.

The SCIRA data reduction scheme is depicted in the figure below. Double arrrows should be followed only if the corresponding reduction is nontrivial.
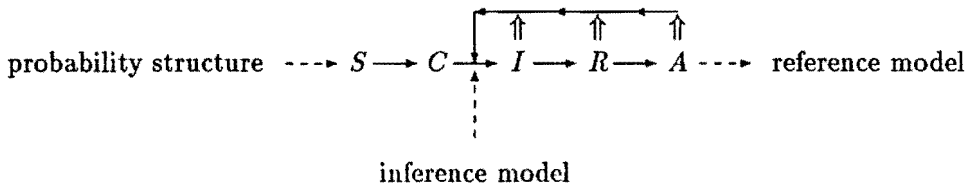
The SCIRA data reduction is illustrated in the examples 10.10 and 10.11.

The reduction and conditioning transformations described in the previous sections seem to be self-evident. However, the specific sequence described above may be open to debate. The authors considered alternative sequences but all of these seem to be inadequate in the sense that for every sequence there is a crucial example where it leads to unsatisfactory results.
So, in statistical inference, every probability structure should be transformed according to SCIRA and every statistical inference problem is equivalent to the construction of inference rules in a specific reference model. This means that the theory presented here may be used to separate statistical inference from the art of decision making.
This paper does not solve any particular statistical inference problem. However, in a forthcoming paper the authors will describe a general method to construct verifiable and precise inference rules. Again, like the SCIRA data reduction, this method is an elaboration and extension of Fisher's work, especially of his idea of sampling the reference set (Fisher (1961)).

10.   Examples

In this section we give some examples to illustrate the concepts given in sections 2-9. In some of these examples both the set Ω of the sample space (Ω, Σ) and the set $\mathcal{P}$ of the probability model $(\mathcal{P}, \overline{\Sigma}|\mathcal{P})$ are finite. First we present the notation that we use in such examples.
The probability structure is represented by a matrix M as shown below.

15

$$
\begin{array}{cc|c}
\mathcal{P} & A & \\
\downarrow & B & \\
& C & M \\
& \vdots & \\
\hline
& & 1 \quad 2 \quad 3 \quad \cdots \\
& & \Omega \quad \rightarrow
\end{array}
$$

Here the possible outcomes are denoted by $1, 2, 3, \ldots$ and the probability measures on Σ by $A, B, C, \ldots$. The σ-field of events is equal to the powerset $P(\Omega)$. By use of the matrix M we can compute probabilities as follows. The probability of the event $\{j\}$, corresponding to a probability measure $p \in \mathcal{P}$ on Σ, is equal to

$$(10.1) \quad p(\{j\}) = m_{ij} / \left(\sum_{j \in \Omega} m_{ij} \right) ,$$

where i is the number of the row that corresponds to the measure p.

Example 10.1.

The probability structure is given by

$$
\begin{array}{c|ccc}
A & 0 & 0 & 5 \\
B & 2 & 2 & 1 \\
C & 1 & 1 & 3 \\
\hline
& 1 & 2 & 3
\end{array} \quad .
$$

Consider the statistic S defined by $S(1) = S(2) = 1$, and $S(3) = 3$ $(\Omega_1 = \{1, 3\})$. We have $\mathcal{P}(1) = \{B, C\}$, and $\mathcal{P}(3) = \mathcal{P}$; see (3.1). The statistic S is sufficient, since

$$\sigma(\psi(1, \cdot)) = \{\emptyset, \mathcal{P}(1)\} ,$$

and

$$\sigma(\psi(3, \cdot)) = \{\emptyset, \mathcal{P}(3)\} .$$

After sufficient reduction the probability structure is represented by

$$
\begin{array}{c|cc}
A & 0 & 5 \\
B & 4 & 1 \\
C & 2 & 3 \\
\hline
& 1,2 & 3
\end{array} \quad .
$$

Example 10.2.

A well-known example in the literature concerns the case where there are two measuring instruments available with different measurement standard deviations. A fair coin is tossed to decide which to use. The statistic C defined by $C = i$ if the i-th instrument is used $(i = 1, 2)$ is an ancillary statistic. Conditioning leads to two probability structures. See Dawid (1982) for more details.

Example 10.3.

The probability structure is given by

16

$$
\begin{array}{c|cccc}
A & 1 & 2 & 3 & 4 \\
\hline
B & 0 & 3 & 4 & 3 \\
\hline
& 1 & 2 & 3 & 4
\end{array} \quad .
$$

In this example there exist two maximal ancillary statistics. Consider the statistic C defined by $C(1) = C(2) = 1, C(3) = C(4) = 2$ $(\Omega_1 = \{1,2\})$. The statistic C is ancillary since

$$\sigma(\varphi) = \{\emptyset, \mathcal{P}\} \, ,$$

and is maximal. By conditioning on the value of C we obtain two probability structures. They are represented by

$$
\begin{array}{c}
C = 1 \\
\begin{array}{c|cccc}
A & 1 & 2 & 0 & 0 \\
\hline
B & 0 & 3 & 0 & 0 \\
\hline
& 1 & 2 & 3 & 4
\end{array}
\end{array}
\qquad
\begin{array}{c}
C = 2 \\
\begin{array}{c|cccc}
A & 0 & 0 & 3 & 4 \\
\hline
B & 0 & 0 & 4 & 3 \\
\hline
& 1 & 2 & 3 & 4
\end{array}
\end{array} \quad .
$$

The ancillary statistic C' defined by $C'(1) = C'(3) = 1, C'(2) = C'(4) = 2$ is also maximal.

In the next examples we have to deal with a σ-field of interest $\mathcal{R}$. In the examples with a finite set $\mathcal{P}$ and a finite set Ω, we denote the σ-field of interest as follows. The probability measures are written as $A_1, A_2, ..., A_{n_A}, B_1, B_2, ..., B_{n_B}, ...$. This notation means that the σ-field of interest $\mathcal{R}$ is equal to

$$(10.2) \quad \mathcal{R} = \sigma(\{A_1, ..., A_{n_A}\}, \{B_1, ..., B_{n_B}\}, ...) \, .$$

Example 10.4.
The inference model is given by

$$
\begin{array}{c|cccc}
A_1 & 1 & 0 & 2 & 1 \\
A_2 & 0 & 1 & 2 & 1 \\
A_3 & 1 & 1 & 1 & 1 \\
B_1 & 1 & 1 & 2 & 0 \\
\hline
& 1 & 2 & 3 & 4
\end{array} \quad .
$$

The only nontrivial element of the set V of (6.1) is the pair (γ, δ), where

$$\gamma(\{1\}) = \{2\}, \gamma(\{2\}) = \{1\}, \gamma(\{3\}) = \{3\}, \gamma(\{4\}) = \{4\} \, ,$$
and
$$\delta(A_1) = A_2, \delta(A_2) = A_1, \delta(A_3) = A_3, \delta(B_1) = B_1 \, .$$

Using (6.7) (Choose $Q = V$), and (6.8)-(6.11), we conclude that there exists a unique minimal invariant statistic I_m with $\sigma(I_m) = \sigma(\{1,2\}, \{3\}, \{4\})$. After invariant reduction the following inference model results

17

$A_{1,2}$	1	2	1
A_3	2	1	1
B_1	2	2	0
	1,2	3	4

.

In this inference model there again exists a nontrivial invariant statistic. Invariant reduction leads to the inference model

$A_{1,2,3}$	3	1
B_1	4	0
	1,2,3	4

.

Remark.

Barnard (1963) demands that the group of transformations on $\mathcal{P}$ operates transitively on $\{A_1, A_2, A_3\}$. So, according to this definition invariant reduction is not allowed in this case.

Example 10.5.

Let X and S^2 be independent random variables with $X \sim N(\mu, \sigma^2)$ and $S^2/\sigma^2 \sim \chi^2$. So, $\Omega = I\!R \times I\!R^+$ and every probability measure in $\mathcal{P}$ can be represented by (μ, σ^2). The interesting aspect of $p \in \mathcal{P}$ is σ^2. The groups of translations on the first component of (μ, σ^2) and the first component of (X, S^2) respectively, lead to the unique minimal invariant statistic defined by $I(X, S^2) = S^2$. In the transformed inference model the set Ω of the sample space can be represented by $\{s^2 | s^2 \in I\!R^+\}$ and the set $\mathcal{P}_1$ of the probability model by $\{\sigma^2 | \sigma^2 \in I\!R^+\}$.

Example 10.6.

The inference model is given by

A_1	0	4	2	2	2
A_2	2	2	2	2	2
A_3	1	1	1	3	4
B_1	1	1	4	1	3
	1	2	3	4	5

.

The statistic R is defined by $R(1) = R(2) = 1, R(3) = 3, R(4) = 4, R(5) = 5$ ($\Omega_1 = \{1, 3, 4, 5\}$). We have

$$\sigma(\varphi) = \sigma(\{A_1, A_2\}, \{A_3\}, \{B_1\}),$$

and

$$\sigma(\psi(1, \cdot)) = \sigma(\{A_1\}, \{A_2, A_3, B_1\}).$$

Conditions (7.3) and (7.4) are satisfied, so R is a partially sufficient statistic. R is minimal and unique. After partially sufficient reduction the following inference model results

$A_{1,2}$	4	2	2	2
A_3	2	1	3	4
B_1	2	4	1	3
	1,2	3	4	5

.

18

Example 10.7.

Let X_1, X_2 and Y be independent random variables such that $X_1, X_2 \sim N(\mu, \sigma_1^2)$, and $Y \sim N(0, \sigma_2^2)$, with $\mu \in I\!R, \sigma_1^2 \in I\!R^+$ and $\sigma_2^2 < 1$. The minimal sufficient statistic is $(X_{(1)}, X_{(2)}, Y)$, where $X_{(1)}$ and $X_{(2)}$ are order statistics. The set $\mathcal{P}$ of the probability model is represented by $\{(\mu, \sigma_1^2, \sigma_2^2) | \mu \in I\!R, \sigma_1^2 \in I\!R^+, \sigma_2^2 < 1\}$. The interesting aspect of $p \in \mathcal{P}$ is μ. The statistic R defined by $R(X_{(1)}, X_{(2)}, Y) = (X_{(1)}, X_{(2)})$ is the unique minimal partially sufficient statistic. In the transformed inference model the set $\mathcal{P}_1$ is represented by $\{(\mu, \sigma_1^2) | \mu \in I\!R, \sigma_1^2 \in I\!R^+\}$.

Remark.

The statistic $(X_{(1)}, X_{(2)})$ is not weakly p-sufficient in the sense defined by Bhapkar (1991), and it is not p-sufficient in the complete sense, because the σ-field of interest is a proper subset of $\sigma(\varphi)$; see (7.3) and Definition 3.2. of Bhapkar (1991).

Example 10.8.

The inference model is given by the transformed model of example 10.6.

$$
\begin{array}{c|cccc}
A_1 & 4 & 2 & 2 & 2 \\
A_2 & 2 & 1 & 3 & 4 \\
B_1 & 2 & 4 & 1 & 3 \\
\hline
 & 1 & 2 & 3 & 4
\end{array}
$$

The statistic A is defined by $A(1) = A(2) = 1, A(3) = A(4) = 2$ ($\Omega_1 = \{1, 2\}$). We have

$$\sigma(\psi(1, \cdot)) = \sigma(\{A_1, A_2\}, \{B_1\}),$$

$$\sigma(\psi(2, \cdot)) = \sigma(\{A_1\}, \{A_2\}, \{B_1\}),$$

and

$$\sigma(\varphi) = \sigma(\{A_1, B_1\}, \{A_2\}).$$

Conditions (8.3) and (8.4) are satisfied, so A is a partially ancillary statistic. It is also maximal and unique. After conditioning on the value of A the following inference models result

$$
\begin{array}{c|cc}
 & \multicolumn{2}{c}{A = 1} \\
A_{1,2} & 2 & 1 \\
B_1 & 1 & 2 \\
\hline
 & 1 & 2
\end{array}
\qquad
\begin{array}{c|cc}
 & \multicolumn{2}{c}{A = 2} \\
A_1 & 14 & 14 \\
A_2 & 12 & 16 \\
B_1 & 7 & 21 \\
\hline
 & 3 & 4
\end{array}
$$

where columns with only zeroes are deleted; see (2.1) and (2.2).

Example 10.9.

Consider the probability structure defined in example 10.5. However, the interesting aspect of $p \in \mathcal{P}$ is μ. The statistic A defined by $A(X, S^2) = S^2$ is the unique maximal partially ancillary statistic. For every s^2 we transform the inference model as follows.

19

The set Ω of the sample space can be represented by $\{x|x \in I\!\!R\}$ and the set $\mathcal{P}^{s^2}$ of the probability model by $\{(\mu, \sigma^2)|\mu \in I\!\!R, \sigma^2 \in I\!\!R^+\}$.

Remark 1.
The statistic A is not weakly p-ancillary in the sense defined by Bhapkar (1989), and it is not p-ancillary in a complete sense, because the σ-field of interest is a proper subset of $\sigma(\psi(s^2, \cdot))$; see (8.4) and Definition (2.3) of Bhapkar (1989).

Remark 2.
The statistics X and S^2 are independent. So, the conditional distribution of X, given $S^2 = s^2$, is independent of s^2. On the other hand, the distribution of X depends on σ^2. Fisher (1961) indicated a method to construct inference rules in this and similar cases.

Example 10.10.
In this example all the transformations of the SCIRA data reduction occur once. The initial probability structure is given by

A	0	4	2	2	1	1	2	4
B	2	2	2	2	1	1	2	4
C	1	1	1	3	2	2	2	4
D	1	1	4	1	1	2	2	4
E	1	1	4	1	2	1	2	4
	1	2	3	4	5	6	7	8

.

Sufficient reduction leads to the probability structure

A	0	4	2	2	1	1	6
B	2	2	2	2	1	1	6
C	1	1	1	3	2	2	6
D	1	1	4	1	1	2	6
E	1	1	4	1	2	1	6
	1	2	3	4	5	6	7,8

.

There exists a unique maximal ancillary statistic. Conditioning on its value leads to the two following probability structures

1.

A,B,C,D,E	1
	7,8

,

2.

A	0	4	2	2	1	1
B	2	2	2	2	1	1
C	1	1	1	3	2	2
D	1	1	4	1	1	2
E	1	1	4	1	2	1
	1	2	3	4	5	6

.

In probability structure 1 only a trivial σ-field of interest can be specified; see section 5.

We specify the σ-field of interest $\mathcal{R}$ in probability structure 2 as follows

$$\mathcal{R} = \sigma(\{A, B, C\}, \{D, E\}) \,.$$

Define accordingly $A_1 := A, A_2 := B, A_3 := C, B_1 := D$ and $B_2 := E$. There exists a unique minimal invariant statistic defined by $I(5) = I(6) = 5$ and $I(i) = i$ for $i = 1, 2, 3, 4$. The transformed inference model is given by the model in example 10.6, where B_1 is replaced by $B_{1,2}$ and the possible outcomes are adjusted accordingly. Now the invariant statistics are trivial. The partially sufficient reduction leads to a transformed inference model given in example 10.8, where A_1 is replaced by $A_{1,2}, A_2$ by A_3 and B_1 by $B_{1,2}$. In this model the invariant statistics and the partially sufficient statistics are trivial. There exists a unique maximal partially ancillary statistic; see also example 10.8. So SCIRA data reduction leads to three reference models

$A_{1,2,3}$	1	2
$B_{1,2}$	2	1
	1,2	3

$A_{1,2}$	14	14
A_3	12	16
$B_{1,2}$	7	21
	4	5,6

$A_{1,2,3}B_{1,2}$	1
	7,8

.

Example 10.11.
The inference model is given by

A_1	1	3	1	5	7	7
A_2	6	8	6	0	2	2
B_1	4	3	2	1	0	14
	1	2	3	4	5	6

.

This is an example where one has to apply a partially ancillary reduction twice. The invariant and partially sufficient statistics are trivial. The statistic A defined by $A(1) = A(2) = A(3) = A(4) = 1, A(5) = A(6) = 2$ ($\Omega_1 = \{1,2\}$) is the unique maximal partially ancillary statistic. After conditioning on the value of A two inference models result

1. $A = 2$

$A_{1,2}$	1	1
B_1	0	2
	5	6

2. $A = 1$

A_1	1	3	1	5
A_2	3	4	3	0
B_1	4	3	2	1
	1	2	3	4

.

Model 1 is a reference model, since the invariant, partially sufficient and partially ancillary statistics are trivial. In model 2 the invariant and partially sufficient statistics are trivial; however, there exists a nontrivial partially ancillary statistic A' defined by $A'(1) = A'(2) = 1, A'(3) = A'(4) = 2$. After conditioning on the value of A' two reference models result

$A' = 1$

A_1	7	21
A_2	12	16
B_1	16	12
	1	2

$A' = 2$

A_1	1	5
A_2	6	0
B_1	4	2
	3	4

.

References

Barnard, G.A. (1963). Some Logical Aspects of the Fiducial Argument. *J. Roy. Stat. Soc.* B25, 111-114.

Bhapkar, V.P. (1989). Conditioning on ancillary statistics and loss of information in the presence of nuisance parameters. *J. Statist. Plann. Inference* 21, 139-160.

Bhapkar, V.P. (1991). Loss of information in the presence of nuisance parameters and partial sufficiency. *J. Statist. Plann. Inference* 28, 185-203.

Dawid, A.P. (1982). Statistical Inference I. *Encyclopedia of statistical sciences.* Wiley, New York.

Fisher, R.A. (1961). Sampling the reference set. *Sankhya A* 23, 3-7.