ANNALES DE L'I. H. P., SECTION B

WERNER FLACKE
NORBERT THERSTAPPEN

Bayesian sufficient statistics and invariance

Annales de l'I. H. P., section B, tome 15, nº 4 (1979), p. 303-314
<http://www.numdam.org/item?id=AIHPB_1979__15_4_303_0>

Ann. Inst. Henri Poincaré, Section B :
Vol. XV, n° 4, 1979, p. 303-314. *Calcul des Probabilités et Statistique.*

Bayesian sufficient statistics and invariance

by

Werner FLACKE and **Norbert THERSTAPPEN** (*)
Lehrstuhl für Mathematik, Fachrichtung Operations Research,
RWTH Aachen, D-5100 Aachen, Templergraben 57

SUMMARY. — The relations between Bayesian sufficient statistics and sufficient statistics are examined and Bayesian sufficiency is composed with the invariance reduction scheme.

An essential sufficient statistic is defined and conditions are given under which it is equivalent to a Bayesian sufficient statistic.

Various counterexamples show that the statement: « Bayesian sufficiency implies sufficiency » is not true in general.

RÉSUMÉ. — La connexion entre une statistique exhaustive et une statistique bayésienne exhaustive est étudiée. Le modèle statistique bayésien est composé avec la notion de réduction invariante.

Après la définition d'une statistique exhaustive essentielle conditions sont faites que cette statistique est équivalente d'une statistique bayésienne exhaustive. Nous montrons sur des exemples que l'exhaustivité bayésienne n'implique pas l'exhaustivité.

---

1. PRELIMINARIES

Let $(X, \mathfrak{A}, P)$ be a basic probability space associated with an observed random variable X, where $P = \{ P_\theta \mid \theta \in \Theta \}$ is an identifiable family of probability measures on $(X, \mathfrak{A})$; that means: $\theta \neq \theta' \Rightarrow P_\theta \neq P_{\theta'}$. Θ can be

(*) Mailing Address, N. Therstappen, Cyprianusweg 6, D-5100 Aachen.

assumed as a metric space. If there is no given metric, take $\rho(\theta_1, \theta_2) = \sup \{ | P_{\theta_1}(A) - P_{\theta_2}(A) | A \in \mathfrak{A} \}$. Let be $\mathfrak{T}$ the σ-algebra of Borel sets of Θ and H a family of probability distributions on $(\Theta, \mathfrak{T})$. Demand the families P and H be dominated by σ-finite measures μ, ζ and denote the belonging versions of the Radon-Nikodym densities by $f(x, \theta)$ and $h(\theta)$ respectively. According to Bayes theorem, the posterior density of θ as a version of the conditional density of θ, given $X = x$, is:

$$h(\theta \mid X = x) = \frac{f(x, \theta)h(\theta)}{\int_\Theta f(x, \tau)h(\tau)d\zeta(\tau)} \chi_{\overline{N}_H \times \Theta}(x, \theta),$$

with

$$N_H = \left\{ x \in X \Big/ \int_\Theta f(x, \tau)h(\tau)d\zeta(\tau) = 0 \right\} . \overline{N}$$

denotes the complement and χ_N the indicator function of a set N.

The following notation is used in the sequel.

P_H a. e. means P_θ a. e. for H almost all $\theta \in \Theta$. $P_H \times K$ a. e. means $P_\theta \times K$ a. e. for H almost all $\theta \in \Theta$; $H, K \in H \cup \{ \zeta \}$. In the same way P_H and $P_H \times K$ null sets may be defined. Every $\mu \times \zeta$ null set is a $P_H \times \zeta$ null set for $H \in H \cup \{ \zeta \}$.

2. BAYESIAN SUFFICIENCY

Let Y be a statistic on $(X, \mathfrak{A})$ into $(Y, \mathfrak{B})$. $Y(P_\theta)$, $\theta \in \Theta$, and $Y(\mu)$ are the image measures under the mapping Y and denote by $f_Y(y, \theta)$ a version of the density of $Y(P_\theta)$ relative to $Y(\mu)$. Thus the posterior density of θ, given $Y = y$, becomes:

$$h_Y(\theta \mid Y = y) = \frac{f_Y(y, \theta)h(\theta)}{\int_\Theta f_Y(y, \tau)h(\tau)d\zeta(\tau)} \chi_{\overline{N}_H^Y \times \Theta}(y, \theta),$$

with

$$N_H^Y = \left\{ y \in Y \Big/ \int_\Theta f_Y(y, \tau)h(\tau)d\zeta(\tau) = 0 \right\} . Y(P_\theta(N_H^Y)) = 0$$

for H almost all $\theta \in \Theta$.

Now a Bayesian definition of a sufficient statistic can be given.

DEFINITION 0. — Y on $(X, \mathfrak{A}, P)$ to a measurable space $(Y, \mathfrak{B})$ is called Bayesian sufficient for H, if for all $H \in H$:

$$h_Y(\theta \mid Y = y) \circ Y(x) = h(\theta \mid X = x) P_H \times \zeta \text{ a. e.}$$

($h_Y \circ Y$ denotes the composition of h_Y and Y). ┘

The usual theorem about the relation between the Bayesian and non-Bayesian definitions of sufficiency states: a statistic is Bayesian sufficient for H if and only if it is sufficient for P [5] [6].

The following example shows that this theorem is not valid for arbitrary H.

EXAMPLE 1. — Let be $\theta_0 \in \Theta$, $H = \{ \varepsilon_{\theta_0} \}$ and $\zeta = \varepsilon_{\theta_0}$, where ε_{θ_0} is the Dirac measure. Under these conditions every statistic on $(X, \mathfrak{A})$ is Bayesian sufficient for $H . h(\theta)$ is ζ a. e. determined by $h(\theta_0) = 1$. So

$$h(\theta \mid \mathrm{X} = x) = h_{\mathrm{Y}}(\theta \mid \mathrm{Y} = y) = 1, \quad \text{for} \quad \theta = \theta_0$$

and 0 otherwise $\mu \times \zeta$ a. e. ⌋

As is easily shown the statement ≫ sufficiency implies Bayesian sufficiency ≪ is true in general.

Because of the following principal difficulty the converse statement can't be shown. Fundamental for the Bayesian statistical model is the existence of prior measures, therefore the equality of the densities in definition 0 can only be required for ζ almost all θ. Sufficiency omits prior measures and postulates a version of the conditional probability independent of all θ.

The following case demonstrates the conceptional difference. θ real, $(\mathrm{X}_1, \mathrm{X}_2)$ independent normal $N(\theta, 1)$ if θ is irrational, $N(\theta, \theta^2)$ if θ is rational. With $\zeta = N(0, 1)$, $\mathrm{S} = (\mathrm{X}_1 + \mathrm{X}_2)/2$ is Bayesian sufficient but not sufficient.

This leads to the following definition.

DEFINITION 2. — Given $(X, \mathfrak{A}, P)$; $P = \{ \mathrm{P}_\theta / \theta \in \Theta \}$ and $(\Theta, \mathfrak{T}, \zeta)$. A statistic $\mathrm{Y} : (X, \mathfrak{A})$ to $(Y, \mathfrak{B})$ is called essentially sufficient (ess. sufficient) with respect to ζ, if for all sets $\mathrm{A} \in \mathfrak{A}$, $\mathrm{B} \in \mathfrak{B}$

$$\mathrm{P}_\theta(\mathrm{A} \cap \mathrm{Y}^{-1}(\mathrm{B})) = \int_{\mathrm{B}} \mathrm{P}(\mathrm{A} \mid \mathrm{Y} = y) d\mathrm{Y}(\mathrm{P}_\theta(y))$$

for ζ almost all θ, where $\mathrm{P}(\mathrm{A} \mid \mathrm{Y} = y)$ is independent of the class of probability measures P. Definition 2 is valid for sufficient statistics. If Θ is countable and $\zeta(\theta) > 0 \; \forall \theta \in \Theta$, ess. sufficiency implies sufficiency. To prove that a Bayesian sufficient statistic is ess. sufficient three conditions are imposed.

CONDITION I.

For all $\Theta' \subset \Theta$ with $\zeta(\Theta') = 1$ let be $P' = \{ \mathrm{P}_\theta \mid \theta \in \Theta' \} \equiv P \equiv \mu$. (this condition insures that $\mathrm{N_H}$ is a μ-null set; $P \equiv P'$ signifies $P \ll P'$ and $P' \ll P$).

Choose the ζ-densities of the probability measures $\mathrm{H} \in H$ so that there exists a countable covering $\{\Theta_H \mid H \in H^*,\ H^* \subset H\}$ of Θ, where

$$\Theta_H : = \left\{ \theta \mid \frac{dH}{d\zeta}(\theta) > 0 \right\}$$

and H^* is countable.

Now distinguish two cases. II : one Θ_H forms the covering; III: the covering is formed by several Θ_H.

CONDITION II.

There exists a $H \in H$ so that a version of the density of H is positive $(h(\theta) > 0)\ \forall \theta \in \Theta$.

CONDITION III.

There is a countable subset $H^* \subset H$ to which corresponds a set $\mathbb{H}$ consisting of the versions of the ζ-density of H, $H \in H^*$, with

i) $\exists(\theta_0 \in \Theta)\ \forall(h \in \mathbb{H})(f(x, \theta_0) > 0 \mu \text{ a. e. } \wedge h(\theta_0) > 0)$.

ii) $\forall(\theta \in \Theta)\ \exists(h \in \mathbb{H})(h(\theta) > 0)$.

Now we can prove

THEOREM 3. — *a)* Let H be a family of probability distributions on $(\Theta, \mathfrak{T})$, then the measurable mapping Y: $(X, \mathfrak{A}) \to (Y, \mathfrak{B})$ is Bayesian sufficient for H if it is ess. sufficient for P.

b) If either conditions I and II or conditions I and III hold then Y is ess. sufficient for P if it is Bayesian sufficient for H.

Proof. — *a)* (analogously to Zacks [*6*]).

b) (*i*) Assume I and II. Let $H \in H$ be the measure with ζ-density $h > 0$.

$$f(x, \theta) = \frac{f(x, \theta)}{\int_\Theta f(x, \tau)h(\tau)d\zeta(\tau)} \int_\Theta f(x, \tau)h(\tau)d\zeta(\tau) P_\zeta \text{ a. e. } \forall \theta \in \Theta.$$

(i. e. for all $(x, \theta) \in \overline{N}_H \times \Theta$). If follows that

$$f(x, \theta) = [h_Y(\theta \mid Y = y).Y(x)] \frac{1}{h(\theta)} \int_\Theta f(x, \tau)h(\tau)d\zeta(\tau) P_\zeta \times \zeta \text{ a.e.}$$

With

$$K(x) = \int_\Theta f(x, \tau)h(\tau)d\zeta(\tau) \qquad \text{and} \qquad W(y, \theta) = h_Y(\theta \mid Y = y] \frac{1}{h(\theta)}$$

$$f(x, \theta) = K(x)[W(y, \theta) \circ Y(x)] P_\zeta \times \zeta \text{ a. e.}$$

which means $\forall(\theta \in \Theta')$ with $\zeta(\Theta') = 1$ we have

$$f(x, \theta) = K(x)[W(y, \theta) \circ Y(x)] P_\theta \text{ a. e.}$$

for ζ almost all $\theta \in \Theta'$. Hence, because of I follows

$$f(x, \theta) = \mathrm{K}(x)[\mathrm{W}(y, \theta) \circ \mathrm{Y}(x)]\mu \text{ a. e.}$$

for ζ almost all $\theta \in \Theta$.

ii) Assume I and III. III (*ii*) states that the sets $\Theta_h = \{ \theta \mid h(\theta) > 0, h \in \mathbb{H} \}$ form a countable partition of Θ.

Hence

$$\sum_{h \in \mathbb{H}} \frac{h(\theta \mid \mathrm{X} = x)h(\theta_0)}{h(\theta_0 \mid \mathrm{X} = x)h(\theta)} \chi_{\Theta_h} = \sum_{h \in \mathbb{H}} \frac{[h_{\mathrm{Y}}(\theta \mid \mathrm{Y} = y) \circ \mathrm{Y}(x)]h(\theta_0)}{[h_{\mathrm{Y}}(\theta_0 \mid \mathrm{Y} = y) \circ \mathrm{Y}(x)]h(\theta)} \chi_{\Theta_h}$$

$:= \mathrm{W}(y, \theta) \circ \mathrm{Y}(x)$. But

$$\sum_{h \in \mathbb{H}} \chi_{\Theta_h} \frac{f(x, \theta) \int_\Theta f(x, \tau)h(\tau)d\zeta(\tau)}{\int_\Theta f(x, \tau)h(\tau)d\zeta(\tau) f(x, \theta_0)} f(x, \theta_0) = f(x, \theta)\mathrm{P}_\zeta \times \zeta \text{ a. e.}$$

It follows as in the proof of part *b* (*i*) of this theorem that

$$f(x, \theta) = \mathrm{K}(x)[\mathrm{W}(y, \theta) \circ \mathrm{Y}(x)]\mu \text{ a. e.}$$

for ζ almost all θ with $\mathrm{K}(x) = f(x, \theta_0)$. ⌟

Example 2 showed that theorem 3*b* isn't valid without assumptions. In particular condition I can't be dispensed with without substitute. The next two examples shows the same is true for conditions II, III *ii*) and III *i*).

EXAMPLE 4. — Let Θ be a countable set, say $\Theta = \{ \theta_i \mid i \in \mathbb{N} \}$. Let

$$\zeta = \sum_{i \in \mathbb{N}} c_i \varepsilon_{\theta_i}, \; c_i \in \mathbb{R}_+, \; \sum c_i = 1 \qquad \text{and} \qquad H = \{ \varepsilon_{\theta_1} \}$$

with

$$h(\theta_j) = \begin{cases} 1/c_i & j = 1 \\ 0 & \text{otherwise} \end{cases}$$

These assumptions fulfil the conditions of I but the assertions of example 2 are preserved (the calculation is parallel to that of example 2).

Define $H = \{ \mathrm{H}_i \mid \mathrm{H}_i = \varepsilon_{\theta_i} \}$ so that $h_i(\theta_j) = \begin{cases} 1/c_i & i = j \\ 0 & i \neq j \end{cases}$

then I and III *ii*) but not III *i*) are fulfilled and again every statistic on $(X, \mathfrak{A})$ is Bayesian sufficient for H. ⌟

In example 4 the requirement that there exists a $\theta_0 \in \Theta$ so that for all

$h \in \mathbb{H} h(\theta_0) > 0$ was not correct while the condition $f(x, \theta_0) > 0 \mu$ a. e. can be thoroughly fulfilled. The next example shows that even if

$$h(\theta_0) > 0 \ \forall (h \in \mathbb{H})$$

is correct and conditions I, III *i*) hold $f(x, \theta_0) > 0$ can't be dispensed with without substitute.

EXAMPLE 5. — Let Θ, ζ be as in example 4, $(X, \mathfrak{A}) = (\mathbb{R}^1, \mathfrak{B}^1)$ the Lebesgue-Borel measurable space and λ^1 the Lebesgue measure on $(\mathbb{R}^1, \mathfrak{B}^1)$. Define

$$H = \left\{ \mathrm{H}_i \mid \mathrm{H}_i = \frac{1}{2}\theta_1 + \frac{1}{2}\theta_i \right\}$$

and

$$h_i(\theta_j) = \begin{cases} 1/2c_1 & j = 1 \\ 1/2c_i & j = i \\ 0 & \text{otherwise} \end{cases}$$

$$h_i(\theta_j \mid \mathrm{X} = x) = \begin{cases} \dfrac{(1/2c_1) f(x, \theta_1)}{(1/2c_1) f(x, \theta_1) + (1/2c_i) f(x, \theta_i)} & j = 1 \\ \dfrac{(1/2c_i) f(x, \theta_i)}{(1/2c_1) f(x, \theta_1) + (1/2c_i) f(x, \theta_i)} & j = i \\ 0 & \text{otherwise} \end{cases}$$

Let $\mathrm{M} \subset X$ with $\mu(\mathrm{M}) > 0$ and let be $f(x, \theta_1) \equiv 0$ on M. Define

$\mathrm{Y_M}(x) = \begin{cases} x & x \in \overline{\mathrm{M}} \\ x_0 & x \in \mathrm{M} \end{cases}$ $x_0 \in \mathrm{M}$. Then $h_i(\theta_j \mid \mathrm{X} = x) = (h_i)_{\mathrm{Y_M}}(\theta_j \mid \mathrm{Y} = y) \circ \mathrm{Y}(x)$

$\mu \times \zeta$ a. e. which means, $\mathrm{Y_M}$ is Bayesian sufficient for H. Here I, III *ii*) and $h(\theta_0) > 0 \ \forall (h \in \mathbb{H})$ (choose $\theta_0 = \theta_1$) are fulfilled. This doesn't include sufficiency. P can still be chosen arbitrarily. Let $\mathrm{P}_{\theta_i} = \mathrm{P}_i$ be a measure with λ^1-density $f(x, i) = \begin{cases} 1/i & x \leqslant i \\ 0 & x > i \vee x < 0. \end{cases}$ Define $\mathrm{M} = [2, 4]$, $x_0 = 3$, $f(x', 1) = 0 \ \forall x' \in \mathrm{M}$. $\mathrm{Y_M}$ generates the σ-subalgebra

$$\mathfrak{A}(\mathrm{Y_M}) = \{ \mathrm{M} \} \cup (\overline{\mathrm{M}} \cap \mathfrak{B}^1)$$

where $\overline{\mathrm{M}} \cap \mathfrak{B}^1$ is the trace of $\mathfrak{B}^1$ in $\overline{\mathrm{M}}$. Suppose $\mathrm{Y_M}$ is a sufficient statistic for P. For each $\mathrm{B} \in \mathfrak{B}^1$ denote by $\mathrm{P}(\mathrm{B} \mid \mathrm{Y_M})$ a version of the conditional expectation of χ_{B} given $\mathrm{Y_M}$ which is independent of P_i, $i \in \mathbb{N}$. Because of the $\mathfrak{A}(\mathrm{Y_M})$ measurability $\mathrm{P}(\mathrm{B} \mid \mathrm{Y_M})$ must be constant on M. Hence

$$\int_{\mathrm{M}} \mathrm{P}(\mathrm{B} \mid \mathrm{Y_M}) d\mathrm{P}_3 = \mathrm{P}_3(\mathrm{B} \cap \mathrm{M}) \qquad \int_{\mathrm{M}} \mathrm{P}(\mathrm{B} \mid \mathrm{Y_M}) d\mathrm{P}_4 = \mathrm{P}_4(\mathrm{B} \cap \mathrm{M}).$$

Let B = [2, 3] and $P(B \mid Y_M) = b$ on M. Then

$$\left.\begin{array}{l} bP_3(M) = P_3(B \cap M) = P_3(B) \\ bP_4(M) = P_4(B) \end{array}\right\} \Rightarrow \frac{P_3(B)}{P_3(M)} = \frac{P_4(B)}{P_4(M)} \Rightarrow 1 = \frac{1}{2}$$

That's a contradiction to sufficiency and likewise to ess. sufficiency, for Θ is countable, $\zeta(\theta) > 0 \ \forall \theta \in \Theta$.

3. BAYESIAN INVARIANTLY SUFFICIENT STATISTICS

Now invariance properties are needed. Let G be a group of one to one measurable transformations of X onto X and let $P = \{ P_\theta \mid \theta \in \Theta \}$ be invariant under G. The invariance of P means that $\forall (g \in G)$ and $\forall (\theta \in \Theta)$ there exists a unique $\theta' \in \Theta$ (P is identifiable so that the uniqueness of θ' will be satisfied) such that the distribution of $g(X)$ is given by $P_{\theta'}$ whenever the distribution of X is given by P_θ. The parameter θ', uniquely determined by g and θ, is denoted by $\bar{g}(\theta)$. Then $\overline{G} = \{ \bar{g}/g \in G \}$ is a group of measurable transformations of Θ onto itself.

The group of transformations induces a partition of X to orbits, where an orbit of x_0, $x_0 \in X$, relative to G is the set $G(x_0) = \{ x/g(x_0) = x, g \in G \}$. A statistic Y on X is invariant if it is constant on orbits i. e.

$$f(x) = f(gx) \forall (x \in X) \forall (g \in G).$$

An invariant function Y on X is called maximal invariant if $f(x) = f(x')$ implies the existence of a $g \in G$ for which $x = gx'$. Maximal invariants always exist and all invariants are functions of a maximal invariant; if Y is invariant under G then its distribution depends only on a maximal invariant function α on Θ under $\overline{G}$ with range $\Gamma = \alpha(\Theta)$.

Let U: $(X, \mathfrak{A}) \to (U, \mathfrak{C})$ be a maximal invariant statistic on X under G and let $f_U(u, \theta)$ be the $U(\mu)$ density of $U(P_\theta)$. Define $U(P_\theta) = P_\gamma$, $\gamma = \alpha(\theta)$ and $f_U(u, \theta) = f_U(u, \gamma)$. Hence $E(f(x, \theta) \mid \mathfrak{A}(U)) = f_U(u, \gamma) \circ U(x) P_\theta$ a. e. where $E(f(x, \theta) \mid \mathfrak{A}(U))$ de notes a version of the conditional expectation of $f(x, \theta)$ given $\mathfrak{A}(U)$. Following from the factorization of conditional expectations [6] $E(h(\theta) \mid \mathfrak{A}(\alpha))$ only depends on γ.

DEFINITION 6.

$$h_U(\gamma \mid U = u) = \frac{f_U(u, \gamma)\tilde{h}(\gamma)}{\int_\Theta f_U(u, \tau) h(\tau) d\zeta(\tau)} \chi_{\overline{N}_H^U \times \Gamma}(u, \gamma)$$

with

$$N_H^U = \left\{ u \Big/ \int_\Theta f_U(u, \tau) h(\tau) d\zeta(\tau) = 0 \right\} \quad \text{and} \quad \tilde{h}(\gamma) \circ \alpha(\theta) = E(h(\theta) \mid \mathfrak{A}(\alpha)). \ \lrcorner$$

Applying the definition of the posterior density we would receive the denominator $\int_\Gamma f_U(u,\gamma)\tilde{h}(\gamma)d\alpha(\zeta)(\gamma)$. Nevertheless the following identity holds.

$$\begin{aligned}\int_\Gamma f_U(u,\gamma)^{h}(\gamma)d\alpha(\zeta)(\gamma) &= \int_\Theta \Big[f_U(u,\gamma)\circ\alpha(\tau)\Big][\tilde{h}(\gamma)\circ\alpha(\tau)]d\zeta(\tau)\\ &= \int_\Theta f_U(u,\tau)\mathrm{E}(h(\tau)\mid\mathfrak{A}(\alpha))d\zeta(\tau) = \int_\Theta \mathrm{E}(f_U(u,\tau)h(\tau)\mid\mathfrak{A}(\alpha))d\zeta(\tau)\\ &= \int_\Theta f_U(u,\tau)h(\tau)d\zeta(\tau),\ \Theta\in\mathfrak{A}(\alpha).\end{aligned}$$

LEMMA 7. — $\mathrm{E}(h_U(\theta \mid U = u) \mid \mathfrak{A}(\alpha)) = h_U(\gamma \mid U = u)\circ\alpha(\theta)P_{\alpha(H)} \times \zeta$ a. e· that means $P_\gamma \times \zeta$ a. e. for $\alpha(H)$ almost all $\gamma\in\Gamma$.

Proof. —

$$\begin{aligned}\mathrm{E}(h_U(\theta\mid U=u)\mid\mathfrak{A}(\alpha)) &= \mathrm{E}\left(\left(\frac{f_U(u,\theta)h(\theta)}{\int_\Theta f_U(u,\tau)h(\tau)d\zeta(\tau)}\right)\Big|\ \mathfrak{A}(\alpha)\right)\\ &= \frac{1}{\int_\Theta f_U(u,\tau)h(\tau)d\zeta(\tau)}\mathrm{E}(f_U(u,\theta)h(\theta)\mid\mathfrak{A}(\alpha))\\ &= \frac{1}{\int_\Theta f_U(u,\tau)h(\tau)d\zeta(\tau)}f_U(u,\theta)\mathrm{E}(h(\theta)\mid\mathfrak{A}(\alpha))\\ &= h_U(\gamma\mid U=u)\circ\alpha(\theta)\zeta \text{ a. e. for fixed } \theta\in\Theta\ \lrcorner\end{aligned}$$

Let $Y : (X, \mathfrak{A}) \to (Y, \mathfrak{B})$ be a Bayesian sufficient statistic for which $Y(gx) = Y(gx')$ whenever $Y(x) = Y(x')$ then G induces a group G_Y of transformations g_Y on Y. Here g_Y is defined by $g_Y y = Y(gx)\ \forall(y\in Y)$ and all x satisfying $Y(x) = y$. If Z is invariant on Y under G_Y then $Z\circ Y$ is invariant on X under G. With the following definition the possible reduction of data is given.

DEFINITION 8. — A function $V : (X, \mathfrak{A}) \to (V, \mathfrak{D})$ is invariantly Bayesian sufficient for H under G, if

i) V is invariant under G ($V = W\circ U$).

ii) $\forall(H\in H)\ h_U(\gamma\mid V = v)\circ W(u) = h_U(\gamma\mid U = u)P_{\alpha(H)}\times\alpha(\zeta)$ a. e. with $h_V(\gamma\mid V = v) = (h_U)_W(\gamma\mid W = v)$. $\lrcorner$

The following scheme represents the routes of reduction.

$$\begin{array}{cccc}\Theta & (X,\mathfrak{A}) & \xrightarrow[\ Y\ ]{} & (Y,\mathfrak{B})\\ \downarrow{\scriptstyle\alpha} & \downarrow{\scriptstyle U} & & \downarrow{\scriptstyle Z}\\ \Gamma & (U,\mathfrak{C}) & \xrightarrow[\ W\ ]{V} & (V,\mathfrak{D})\end{array}$$

Definition 8 states that V is Bayesian sufficient for

$$H_\alpha = \alpha(H) = \{ \alpha(\mathrm{H})/\mathrm{H} \in H \}$$

where $\alpha : \Theta \to \Gamma$ is a maximal invariant function under $\overline{\mathrm{G}}$. Regard Γ as a quotient space formed by the $\overline{\mathrm{G}}$ orbits, equipped with the quotient topology which is the finest topology keeping α continous. The purpose is to transfer condition I from Θ to Γ but the quotient topology is too coarse in general.

LEMMA 9. — If condition I is valid for ζ, it is valid too for $\alpha(\zeta)$.

Proof. — (I) Assume $\Gamma' \subset \Gamma$ with $\alpha(\zeta)(\Gamma') = 1$ then $\{ \mathrm{P}_\gamma / \gamma \in \Gamma' \}$ and U(P) are equivalent. Let be $\mathrm{N}^{\mathrm{U}} \in \mathfrak{C}$ with $\mathrm{P}_\gamma(\mathrm{N}^{\mathrm{U}}) = 0$, $\gamma \in \Gamma'$ then $\mathrm{P}_\theta(\mathrm{U}^{-1}(\mathrm{N}^{\mathrm{U}})) = 0$ for all $\theta \in \alpha^{-1}(\Gamma')$. Thus $1 = \alpha(\zeta)(\Gamma')$. Hence for all $\theta \in \Theta \mathrm{P}_\theta(\mathrm{U}^{-1}(\mathrm{N}^{\mathrm{U}})) = 0$ and finally $\mathrm{P}_\gamma(\mathrm{N}^{\mathrm{U}}) = 0$ for all $\gamma \in \Gamma$. ⌟

LEMMA 10. — If condition II holds, it also holds for H_α and $\alpha(\zeta)$.

Proof. — Let $\mathrm{H} \in H$ fulfil condition II. $h_\alpha(\gamma) \circ \alpha(\theta) = \mathrm{E}(h(\theta) \mid \mathfrak{A}(\alpha))$ implies $h_\alpha(\gamma) \circ \alpha(\theta) \geqslant 0 \zeta$ a. e. Then

$$h'_\alpha(\gamma) = \begin{cases} h_\alpha(\gamma) & h_\alpha(\gamma) > 0 \\ 1 & h_\alpha(\gamma) = 0 \end{cases}$$

is also a density of H_α relative to $\alpha(\zeta)$. ⌟

EXAMPLE 11. — It is impossible to transfer condition III in the same way. Be $\Theta = \mathbb{R}^1_+$, $\Gamma = \{ [n-1, n[/ n \in \mathbb{N} \}$ and let be ζ the λ^1-continuous probability measure with density c, $c = \sum_{i \in \mathbb{N}} c_i \chi_{[i-1,i[}$ with $\sum_{i \in \mathbb{N}} c_i = 1$ and $c_i > 0$ for all $i \in \mathbb{N}$. Let H_i be the measure with ζ-density

$$h_i = \frac{1}{c_i} \chi_{[i-1,i[} + \chi_{\{0\}}, \quad i \in \mathbb{N}. H = \{ \mathrm{H}_i / i \in \mathbb{N} \}.$$

Then condition III *ii*) is fulfilled for H while the item $\chi_{\{0\}}$ synthesizes III *i*) (given $f(x, 0) > 0$). Transfering the conditions to the image space we get III *ii*) with $h_{i\alpha} = 1/c_i \varepsilon_{[i-1,i[}$ but III *i*) isn't valid any longer because of

$$c_\alpha = \sum_{i \in \mathbb{N}} c_i \varepsilon_{[i-1,i[} \quad ⌟$$

THEOREM 12. — Let $\mathrm{V} : (X, \mathfrak{A}) \to (V, \mathfrak{D})$ be an invariantly ess. sufficient statistic for P under G, then V is invariantly Bayesian sufficient for H under G. If conditions I and II are fulfilled, then invariantly Bayesian sufficiency implies invariantly ess. sufficiency.

Proof. — V is invariantly Bayesian sufficient for H under G if and only if W is Bayesian sufficient for H_α. Let V be invariantly ess. sufficient for P under G. Then W is ess. sufficient for $U(P)$. Hence W is Bayesian sufficient for H_α. Now let W be Bayesian sufficient for H_α. Then H_α fulfils I and II. Hence W is ess. sufficient for $U(P)$.

The theorem of Stein holds under several conditions (see [*4*]). One of these conditions states.

CONDITION IV.

There is a set $A_0 \in \mathfrak{A}(V)$ of P measure 1 and an invariant conditional probability distribution Q on $\mathfrak{A} \times X$ with $Q(A, x) = 0 \; \forall(A \in \mathfrak{A}) \; \forall(x \notin A_0)$.

The invariance of the conditional probability distribution means:

$$\forall(x \in X) \; \forall(A \in \mathfrak{A}) \; \forall(g \in G) \; Q(gA, gx) = Q(a, x).$$

Now we get a Bayesian version for the theorem of Stein.

THEOREM 13. — Let conditions I, II, IV or I, III, IV be valid then (*), condition IV can be substituted by any of the other conditions of [*4*].

Proof. — Y is ess. sufficient for P. It follows that $Z \circ Y$ is invariantly ess. sufficient for P under G [*4*]. Hence $Z \circ Y$ is invariantly Bayesian sufficient.

EXAMPLE 14. — Let be $(X, \mathfrak{A}) = (\mathbb{R}^n, \mathfrak{B}^n)$; $X = (X_1, \dots, X_n)$; $X_1, \dots, X_n$ independent identical distributed random variables with distribution $N(\theta_1, \theta_2^2)$; we define $\Theta = \mathbb{R}^1 \times \mathbb{R}^1_+$ and

$$G = \{ g_c / g_c : \mathbb{R}^n \to \mathbb{R}^n \; ; \; g_c(x_1, \dots, x_n) = (x_1 + c, \dots, x_n + c) \} \; ; \; c \in \mathbb{R}$$

$$\bar{G} = \{ \bar{g}_c / \bar{g}_c : \Theta \to \Theta \; ; \; \bar{g}_c(\theta_1, \theta_2) = (\theta_1 + c, \theta_2) \}$$

$U : \mathbb{R}^n \to \mathbb{R}^{n-1}$ defined by $U(x_1, \dots, x_n) = (x_1 - x_n, \dots, x_{n-1} - x_n) =: (u_1, \dots, u_{n-1})$

$$Y : \mathbb{R}^n \to \mathbb{R}^1 \times \mathbb{R}^1_+ \text{ defined by } Y(x_1, \dots, x_n) = \left(\frac{1}{n} \sum_1^n x_i, \sum_1^n (x_i - \bar{x})^2 \right)$$

$$=: (Y_1(x_1, \dots, x_n), Y_2(x_1, \dots, x_n)) =: (Y_1, Y_2)$$

$$g_{c_y}(y_1, y_2) := (y_1 + c, y_2) = Y(x_1 + c, \dots, x_n + c) = \gamma(g_c(x_1, \dots, x_n)).$$

Hence $Z(y) = y_2$. Let be $\alpha : \Theta \to \Gamma$, $\Gamma = \mathbb{R}^1_+$ be defined by $\alpha(\theta_1, \theta_2) = \theta_2$, then α is continuous (and IV is fulfilled); ζ must be a probability measure on

(*) $h_Y(\theta \mid Y = y) \circ Y(x) = h(\theta \mid X = x)$ implies
$h_{Z \circ Y}(\gamma \mid Z \circ Y = v) \circ W(u) = h_U(\gamma \mid U = u)$ with $W \circ U = Z \circ Y$)

$(\Theta, \Theta \cap \mathfrak{B}^2)$ with a positive $\lambda^2_{|\Theta}$-density $\xi(\theta_1, \theta_2)$ (II fulfilled). Let be $h' = h\xi$ and $H = \{ \text{H/H probability measure on } (\Theta, \Theta \cap \mathfrak{B}^2) \text{ with } \zeta\text{-density } h(\theta_1, \theta_2) > 0 \}$. With some calculations follows

$$\mathrm{Z} \circ \mathrm{Y}(x_1, \ldots, x_n) = \sum_1^n \Big(x_i - \frac{1}{n}, \sum_1^n x_j \Big)^2 = \mathrm{W}(u_1, \ldots, u_{n-1})$$
$$= \frac{1}{n} \Big[(n-1) \sum_1^{n-1} u_i^2 - \sum_{\substack{i,j=1 \\ i \neq j}}^{n-1} u_i u_j \Big]$$

Further

$$f(x_1, \ldots, x_n\,; \theta_1, \theta_2) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\theta_2^n} \cdot \exp \Big\{ -\frac{1}{2\theta_2^2} \sum_1^n (x_i - \bar{x})^2 - \frac{n}{2\theta_2^2} (\bar{x} - \theta_1)^2 \Big\}$$
$$h_\alpha(\theta_2) = h_\alpha(\gamma) = \int_{-\infty}^{\infty} h(\theta_1, \theta_2) \xi(\theta_1, \theta_2) d\lambda^1(\theta_1)$$
$$f_{\mathrm{U}}(u_1, \ldots, u_{n-1}\,; \theta_2) = \frac{1}{(2\pi)^{(n-1)/2} \theta_2^{n-1} \mid \mathrm{M} \mid^{1/2}} \exp \Big\{ -\frac{1}{2\theta_2^2} u \mathrm{M}^{-1} u^{\mathrm{T}} \Big\}$$

Here $\mathrm{M} = \begin{pmatrix} 2 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \ldots & 2 \end{pmatrix}$ is a $(n-1) \times (n-1)$ matrix. U has the distribution $N(0, \theta_2^2 \mathrm{M})$. Hence

$$h_{\mathrm{U}}(\theta_2 \mid \mathrm{U} = u) = \frac{\int_{-\infty}^{\infty} h'(\theta_1, \theta_2) d\lambda^1(\theta_1) \theta_2^{1-n} \exp\left(-(1/2\theta_2^2) u \mathrm{M}^{-1} u^{\mathrm{T}}\right)}{\int_{-\infty}^{\infty} d\lambda^1(\theta_1) \int_0^{\infty} d\lambda^1(\theta_2) h'(\theta_1, \theta_2) \theta_2^{1-n} \exp\left(-(1/2\theta_2^2) u \mathrm{M}^{-1} u^{\mathrm{T}}\right)}$$

Here Y_2/θ_2^2 is χ^2_{n-1}-distributed. Y_2 has the λ^1-density $f_{\mathrm{Y}_2}(y_2\,;\,\theta_2)$ $= (\mathrm{Z} \circ \mathrm{Y} = \mathrm{Y}_2)$

$$= \frac{1}{\theta_2^{n-1}} \frac{1}{2^{(n-1)/2} \Gamma(n - 1/2)} y_2^{(n-3)/2} \exp\left(-\frac{1}{2\theta_2^2} y_2 \right).$$

So we get

$$h_{\mathrm{Y}_2}(\theta_2 \mid \mathrm{Y}_2 = y_2) = \frac{\int_{-\infty}^{\infty} h'(\theta_1, \theta_2) d\lambda^1(\theta_1) \theta_2^{1-n} \exp\left((-1/2\theta_2^2) y_2\right)}{\int_{-\infty}^{\infty} d\lambda^1(\theta_1) \int_0^{\infty} d\lambda^1(\theta_2) h'(\theta_1, \theta_2) \theta_2^{1-n} \exp\left((-1/2\theta_2^2) y_2\right)}.$$

It holds $y_2 = u \mathrm{M}^{-1} u^{\mathrm{T}}$ with $\mathrm{M}^{-1} = -\frac{1}{n} \begin{pmatrix} 1-n & \ldots & 1 \\ \cdot & \ddots & \vdots \\ 1 & \ldots & 1-n \end{pmatrix}$ and finally

$h_{\mathrm{U}}(\theta_2 \mid \mathrm{U} = u) = h_{\mathrm{Y}_2}(\theta_2 \mid \mathrm{Y}_2 = y_2)$ ⌟

REFERENCES

[1] D. Basu, *On Sufficiency and Invariance Essays in Probability and Statistics.* Chapel Hill, N. C. University of North Carolina Press, 1970, p. 61-84.

[2] H. Bauer, *Probability Theory and Elements of Measure Theory.* New York, Holt, Rinehart and Winston, 1972.

[3] N. Bourbaki, *Éléments de Mathématique.* Intégration Chap. 1-4, Paris, Hermann, 1965.

[4] W. J. Hall, R. A. Wijsman, J. K. Ghosh, The Relationship between Sufficiency and Invariance with Applications in Sequential Analysis. *Ann. Math. Statist.*, t. **36**, 1965, p. 575-614.

[5] H. Raiffa, R. Schlaifer, *Applied Statistical Decision Theory*, Cambridge, Mass. The M. I. T. Press, 1961.

[6] S. Zacks, *The Theory of Statistical Inference*, New York, J. Wiley, 1971.

(Manuscrit reçu le 3 mai 1978)