Ann. Inst. Statist. Math.
Vol. 40, No. 4, 715–726 (1988)

SUFFICIENCY AND JENSEN'S INEQUALITY FOR CONDITIONAL EXPECTATIONS

DIETER MUSSMANN

Institut für Mathematische Statistik der Universität Münster, Einsteinstraße 62, D-4400 Münster, West Germany

(Received May 12, 1986; revised July 22, 1987)

Abstract. For finite sets of probability measures, sufficiency is characterized by means of certain positively homogeneous convex functions. The essential tool is a discussion of equality in Jensen's inequality for conditional expectations. In particular, it is shown that characterizations of sufficiency by Csiszár's f-divergence (1963, *Publ. Math. Inst. Hung. Acad. Sci. Ser. A*, **8**, 85–107) and by optimal solutions of a Bayesian decision problem used by Morse and Sacksteder (1966, *Ann. Math. Statist.*, **37**, 203–214) can be proved by the same method.

Key words and phrases: Sufficiency, Jensen's inequality, f-divergence, Bayesian decision problem.

1. Introduction

Let $P_1,\dots, P_k$ be probability measures on a measurable space $(\Omega, \mathscr{A})$. Put $\mu = (1/k)(P_1 + \cdots + P_k)$ and $p_i = dP_i/d\mu$, $1 \le i \le k$. Let b_i be the probability that P_i is the "true" distribution. Then

$$\int \sup \{b_1 p_1,\dots, b_k p_k\} d\mu$$
$$= \sup \left\{ \sum_{i=1}^{k} b_i \int f_i p_i d\mu \colon \sum_{i=1}^{k} f_i = 1_\Omega, f_i \text{ test for each } 1 \le i \le k \right\},$$

is the maximum probability of guessing the "true" P_i (cf. Morse and Sacksteder (1966), Torgersen ((1976), p. 194) and Györfi and Nemetz (1977)). Morse and Sacksteder used these quantities in order to characterize sufficiency. For the same purpose and $k = 2$, Csiszár (1963, 1967) employed f-divergences $\int g \circ (p_1, p_2) d\mu$, where $g(x_1, x_2) = f(x_1/x_2)x_2$ for $x_1, x_2 > 0$ and f is a convex function on $(0, \infty)$ (see Lemma 1.3 and Example 1(c)).

We shall characterize sufficiency by means of integrals $\int g \circ (b_1 p_1, \ldots, b_k p_k) d\mu$ where $b_i > 0$ and g is a positively homogeneous convex map with values in $(-\infty, \infty]$ which equals the sup-function or a k-dimensional generalization of the above function g used by Csiszár. In particular, we shall see that the results of Csiszár (1963, 1967) and Morse and Sacksteder (1966) can be proved by the same method. The essential tool is a discussion of equality in Jensen's inequality for conditional expectations and certain positively homogeneous convex maps which have similar properties as the sup-function (see Lemma 1.1) and the map g introduced in Lemma 1.3. Integrals of the above type with a general real valued g have been studied by Györfi and Nemetz (1977) as a measure of dissimilarity of distributions. For later reference we state the following obvious result.

LEMMA 1.1. *For $x = (x_1, \ldots, x_k) \in [0, \infty)^k$ put $g(x) = \sup\{x_1, \ldots, x_k\}$ and $C_i = \{x \in [0, \infty)^k : x_i = g(x)\}$. Then g is a positively homogeneous convex function; and if $x, y \in [0, \infty)^k$, $0 < t < 1$, and $g(tx + (1-t)y) = tg(x) + (1-t)g(y)$, then $x, y \in C_i$ for some $1 \leq i \leq k$.*

The definition of a positively homogeneous or convex real valued function defined on a convex cone or subset of a linear space makes sense for an extended real valued function the range of which is contained in $(-\infty, \infty]$ if one assumes $0\infty = 0$, $b\infty = \infty$ for $b > 0$, and $\infty + b = \infty$ for $-\infty < b \leq \infty$. Concerning continuity we consider the usual topology on $(-\infty, \infty]$. Using these conventions, functions like $1/x$ or $-\log x$ can be considered as continuous convex maps from $[0, \infty)$ to $(-\infty, \infty]$ which carry 0 into ∞. In Lemma 1.3 we shall see that positively homogeneous convex maps from $[0, \infty)^k \backslash \{0\}$ to $(-\infty, \infty]$, $k \geq 2$, which are defined by means of a continuous convex map from $[0, \infty)^{k-1}$ to $(-\infty, \infty]$ have similar properties as g in Lemma 1.1. If $x = (x_1, \ldots, x_k) \in [0, \infty)^k$, put $x' = (x_1, \ldots, x_{k-1})$. For $h \in (0, \infty)^{k-1}$ define $L_{h1} = \{u \in [0, \infty)^{k-1} : \langle h, u \rangle \leq 1\}$, $L_{h2} = \{u \in [0, \infty)^{k-1} : \langle h, u \rangle \geq 1\}$, $C_{h1} = \{x \in [0, \infty)^k : \langle h, x' \rangle \leq x_k\}$, and $C_{h2} = \{x \in [0, \infty)^k : \langle h, x' \rangle \geq x_k\}$ ($\langle\ ,\ \rangle$ denotes the inner product). For $0 < s < \infty$ put $x/s = (x_1/s, \ldots, x_k/s)$.

LEMMA 1.2. *Let f be a continuous convex map from $[0, \infty)^{k-1}$ to $(-\infty, \infty]$. Then the following assertions hold.*

(a) *There exists $\lim_{s \downarrow 0} sf(y/s) \in (-\infty, \infty]$ for all $y \in [0, \infty)^{k-1}$.*

(b) *Suppose there is an $h \in (0, \infty)^{k-1}$ such that if $u, v \in [0, \infty)^{k-1}$, $0 < t < 1$, and $f(tu + (1-t)v) = tf(u) + (1-t)f(v)$, then $u, v \in L_{hi}$ for some $i \in \{1, 2\}$. Then $f(y) < \infty$ for all $y \in [0, \infty)^{k-1} \backslash \{0\}$.*

PROOF. (a) For fixed $y \in [0, \infty)^{k-1}$ let F denote the map from $[0, \infty)$ to $(-\infty, \infty]$ which carries t into $f(ty)$. F is convex. This implies that there is a t_0 such that $F(t) < \infty$ for all $t \geq t_0$, or $F(t) = \infty$ for all $t \geq t_0$. In the first case the existence of $\lim_{t\uparrow\infty}(1/t)F(t) \in (-\infty, \infty]$ is well known; in the second case $(1/t)F(t) = \infty$ for all $t \geq t_0$. Now

$$\lim_{t\uparrow\infty}(1/t)F(t) = \lim_{s\downarrow 0} sF(1/s) = \lim_{s\downarrow 0} sf(y/s) .$$

(b) If $y \in [0,\infty)^{k-1}$ and $y \neq 0$, then there are $b < 1 < c$ such that $by \in L_{h1}$, $cy \in L_{h2}\backslash L_{h1}$. Putting $t = (c-1)/(c-b)$, $u = by$, and $v = cy$, we get $y = tu + (1-t)v$ and $f(y) < tf(u) + (1-t)f(v) \leq \infty$.

LEMMA 1.3. *Let f be given as in Lemma 1.2 and let g denote the map from $[0,\infty)^k\backslash\{0\}$ to $(-\infty,\infty]$ defined as*

$$g(x) = x_k f(x'/x_k) \quad \textit{if} \quad x_k > 0$$

and

$$g(x) = \lim_{s\downarrow 0} sf(x'/s) \quad \textit{if} \quad x_k = 0 .$$

Then the following assertions hold.

(a) *g is measurable, positively homogeneous, and convex.*

(b) *If f fulfills the premise of Lemma 1.2(b), then $g(x) < \infty$ for all $x \in [0,\infty)^k\backslash\{0\}$ with $x' \neq 0$ and $x_k \neq 0$.*

(c) *Suppose that f fulfills the premise of Lemma 1.2(b) and that $x, y \in [0,\infty)^k\backslash\{0\}$, $y_k > 0$, $0 < t < 1$, and $g(tx + (1-t)y) = tg(x) + (1-t)g(y)$. Then $x, y \in C_{hi}$ for some $i \in \{1,2\}$.*

PROOF. Lemma 1.3(b) follows immediately from Lemma 1.2(b).

(a) From the definition of g its positive homogeneity is clear. Since f is continuous, we have

$$g(x) = \lim_n (x_k + (1/n))f(x'/(x_k + (1/n))) ,$$

for all $x \in [0,\infty)^k\backslash\{0\}$. Hence g is measurable.

The convexity of g follows because, if $x, y \in [0,\infty)^k\backslash\{0\}$, $0 < t < 1$, and $s_n = t(x_k + 1/n) + (1-t)(y_k + 1/n)$, then

$$\begin{aligned} g(tx + (1-t)y) &= \lim_n s_n f((tx' + (1-t)y')/s_n) \\ &= \lim_n s_n f((t(x_k + 1/n)/s_n)x'/(x_k + 1/n) \\ &\qquad + ((1-t)(y_k + 1/n)/s_n)y'/(y_k + 1/n)) \\ &\leq \lim_n s_n((t(x_k + 1/n)/s_n)f(x'/(x_k + 1/n)) \end{aligned}$$

$$+((1-t)(y_k+1/n)/s_n)f(y'/(y_k+1/n)))$$
$$=tg(x)+(1-t)g(y).$$

(c) If $t<c<1$, $z=cx+(1-c)y$, and $b=t/c$, then $tx+(1-t)y=bz+(1-b)y$ and $bg(z)+(1-b)g(y)\le tg(x)+(1-t)g(y)$. Hence $g(bz+(1-b)y)=bg(z)+(1-b)g(y)$. Replacing x in Proof of (a) by z, t by b, and s_n by $tz_k+(1-t)y_k$, we obtain $z'/z_k, y'/y_k \in L_{hi}$ for some $i\in\{1,2\}$. This implies $z,y\in C_{hi}$ for some $i\in\{1,2\}$, and therefore $x,y\in C_{hi}$ for some $i\in\{1,2\}$.

Examples 1. (a) If f is a strictly convex continuous map from $[0,\infty)^{k-1}$ to $(-\infty,\infty]$, the assumption in Lemma 1.2(b) is fulfilled for each $h\in(0,\infty)^{k-1}$.

(b) Suppose $h\in(0,\infty)^{k-1}$ and $f_h(u)=|\langle h,u\rangle-1|$. Then $g_h(x)=|\langle h,x'\rangle-x_k|$ corresponds to f_h.

(c) For $k=2$, $\int g\circ(p_1,p_2)d\mu$ is Csiszár's (1963, 1967) f-divergence. Examples are given in Csiszár ((1963), pp. 86–87, and (1967), p. 301) and Györfi and Nemetz (1977).

(d) Suppose $k=2$. Define strictly convex functions f_t by $f_t(u)=-u^t$ if $0<t<1$ and $f_t(u)=u^t$ if $1<t<\infty$. Then $g_t(x_1,x_2)=-x_1^tx_2^{1-t}$ if $0<t<1$ and $g_t(x_1,x_2)=x_1^tx_2^{1-t}$ if $1<t<\infty$ correspond to f_t.

(e) If $f(u)=u\log u$, then $g(x)=\infty$ iff $x_2=0$ and $x_1\neq 0$. Thus, even if f is real valued, g is in general not.

2. Sufficiency and Jensen's inequality

First we study equality in Jensen's inequality for conditional expectations where the convex functions have some special properties like g in Lemma 1.1 or 1.3. For arbitrary convex functions this problem has been investigated in full generality by Kozek and Suchanecki (1980) and earlier by Pfanzagl (1974*b*). It seems that their results do not (at least not in a simple way) imply Proposition 2.1 below.

In the following let $\mathscr{A}$ be a σ-algebra over a set Ω. Suppose $\mathscr{S}\subset\mathscr{A}$ is a sub-σ-algebra and P is a probability measure on $\mathscr{A}$. For A, $B\in\mathscr{A}$ we shall write $A\subset B[P]$ and $A=B[P]$ if $P(A\cap B^c)=0$ and $P(A\cap B^c)+P(A^c\cap B)=0$, respectively. Recall that for each measurable map Y from Ω to $[0,\infty]$ a conditional expectation $E(Y|\mathscr{S})$ with values in $[0,\infty]$ can be defined (cf. Bauer (1968), p. 244). If Y is a measurable map from Ω to $(-\infty,\infty]$ with $E(Y^-|\mathscr{S})<\infty$ P-a.e., one puts $E(Y|\mathscr{S})=E(Y^+|\mathscr{S})-E(Y^-|\mathscr{S})$. Furthermore, we set $P(A|\mathscr{S})=E(1_A|\mathscr{S})$ for $A\in\mathscr{A}$, and $E(Z|\mathscr{S})=(E(Z_1|\mathscr{S}),\dots,E(Z_k|\mathscr{S}))$ if $Z_1,\dots,Z_k$ are measurable maps from Ω to $(-\infty,\infty]$ such that each $E(Z_i^-|\mathscr{S})<\infty$ P-a.e. The next lemma is implied by results in Rubin and Wesler (1958) and Pfanzagl (1974*b*). R

denotes the set of real numbers.

LEMMA 2.1. *Let $C \subset R^k$ be a convex Borel set and g a measurable convex function from C to $(-\infty, \infty]$ such that $C' = \{x \in C: g(x) < \infty\}$ is not empty. For $i = 1,\ldots,k$ let X_i be a real valued measurable function on Ω with $E|X_i| < \infty$. Suppose $X(\omega) \in C$ for all $\omega \in \Omega$ where $X = (X_1,\ldots,X_k)$. Then the following assertions hold.*

(a) $E((g \circ X)^- | \mathscr{S}) < \infty$ *P-a.e.*

(b) $g \circ E(X|\mathscr{S}) \leq E(g \circ X | \mathscr{S})$ *P-a.e.*

(c) *If $E(g \circ X|\mathscr{S}) < \infty$ P-a.e., then there is a Markov kernel ψ from $(\Omega, \mathscr{A})$ to $(R^k, \mathscr{B}^k)$ such that $\psi(\cdot, B) = P(\{X \in B\}|\mathscr{S})$ P-a.e. for all $B \in \mathscr{B}^k$, $\psi(\omega, C') = 1$, and $\int y\psi(\omega, dy) \in C'$ for all $\omega \in \Omega$.*

(d) *If $g \circ E(X|\mathscr{S}) = E(g \circ X|\mathscr{S}) < \infty$ P-a.e., then ψ can be chosen such that $g\left(\int y\psi(\omega, dy)\right) = \int g(y)\psi(\omega, dy)$ for all $\omega \in \Omega$.*

The premise of the following proposition holds for the situation in Lemma 1.1 or 1.3(b) (see Corollaries 2.1 to 2.3).

PROPOSITION 2.1. *Assume the situation of Lemma 2.1. Moreover, let $\tilde{C}, C_1,\ldots, C_n$ be convex subsets of C with the following properties: If $x \in \tilde{C}$, $y \in C$, $0 < t < 1$, and $g(tx + (1-t)y) = tg(x) + (1-t)g(y)$, then $x, y \in C_i$ for some $i = 1,\ldots,n$; $\tilde{C} \neq \varnothing$, $C = C_1 \cup \cdots \cup C_n$, $C \setminus \tilde{C}$ is convex, and $C_i \cap \bigcap_{j \in J} C_j^c$ is convex for all $i = 1,\ldots,n$ and $\varnothing \neq J \subset \{1,\ldots,n\}$. Suppose*

$$g \circ E(X|\mathscr{S}) = E(g \circ X|\mathscr{S}) < \infty \qquad P\text{-a.e.}$$

Then

$$\left\{ E(X|\mathscr{S}) \in C_i \cap \bigcap_{j \neq i} C_j^c \cap \tilde{C} \right\} \subset \{X \in C_i\}[P],$$

for all $i = 1,\ldots,n$.

PROOF. We only need to prove the assertion for $i = 1$. Put $M_1 = C_1$, $M_i = C_i \cap \bigcap_{j<i} C_j^c$ for $2 \leq i \leq n$, $H_i = \{X \in M_i\}$, and $I = \{i: P(H_i) > 0\}$. For $i \in I$ define $P_i = P(\cdot \cap H_i)/P(H_i)$. Let $E_i(\cdot|\mathscr{S})$ denote the conditional expectation with respect to $\mathscr{S}$ and P_i, $i \in I$. Let ψ be given as in Lemma 2.1. For $i \in I$ there are Markov kernels ψ_i such that $\psi_i(\cdot, B) = P_i(\{X \in B\}|\mathscr{S})$ P_i-a.e., $\psi_i(\cdot, M_i) = 1$, and $\int y\psi_i(\omega, dy) \in M_i$ for all $\omega \in \Omega$ and $B \in \mathscr{B}^k$. We have

$$E(h \circ X|\mathscr{S}) = \sum_{i \in I} P(H_i|\mathscr{S}) E_i(h \circ X|\mathscr{S}) \qquad P\text{-a.e.},$$

if h is a measurable function with $E|h \circ X| < \infty$ (see Pfanzagl (1974*b*), p. 492). Hence

$$\int y\psi(\cdot\,, dy) = \sum_{i \in I} \psi(\cdot\,, M_i) \int y\psi_i(\cdot\,, dy) \qquad P\text{-a.e.}\,,$$

and

$$\int g(y)\psi(\cdot\,, dy) = \sum_{i \in I} \psi(\cdot\,, M_i) \int g(y)\psi_i(\cdot\,, dy) \qquad P\text{-a.e.}$$

By Jensen's inequality,

$$g\left(\int y\psi_i(\omega, dy)\right) \leq \int g(y)\psi_i(\omega, dy)\,,$$

for all $\omega \in \Omega$ and $i \in I$.

We conclude that there is a P-null set N such that

$$\int y\psi(\omega, dy) = \sum_{i \in I} \psi(\omega, M_i) \int y\psi_i(\omega, dy)\,,$$

$$g\left(\int y\psi(\omega, dy)\right) = \int g(y)\psi(\omega, dy)\,,$$

and

$$\begin{aligned} g\left(\int y\psi(\omega, dy)\right) &\leq \sum_{i \in I} \psi(\omega, M_i) g\left(\int y\psi_i(\omega, dy)\right) \\ &\leq \sum_{i \in I} \psi(\omega, M_i) \int g(y)\psi_i(\omega, dy) \\ &= \int g(y)\psi(\omega, dy) \qquad \text{for all} \qquad \omega \in \Omega \backslash N\,. \end{aligned}$$

Hence, if $\omega \in \Omega\backslash N$, $\int y\psi(\omega, dy) \in \tilde{C}$, and $0 < \psi(\omega, M_i) < 1$ for some $i \in I$, then

$$\int y\psi(\omega, dy) = \psi(\omega, M_i) \int y\psi_i(\omega, dy) + (1 - \psi(\omega, M_i))x$$

and

$$g\left(\int y\psi(\omega, dy)\right) = \psi(\omega, M_i) g\left(\int y\psi_i(\omega, dy)\right) + (1 - \psi(\omega, M_i))g(x)\,,$$

for some $x \in C$. Since $C\backslash\tilde{C}$ is convex, $\int y\psi_i(\omega, dy)$ or x is in $\tilde{C}$. Thus

$\int y\psi(\omega, dy)$, x, $\int y\psi_i(\omega, dy) \in C_j$ for some $1 \leq j \leq n$. Using $\int y\psi_i(\omega, dy) \in M_i$, we see that $\int y\psi(\omega, dy) \in C_1 \cap \bigcap_{j\geq 2} C_j^c \cap \tilde{C}$ implies $\psi(\omega, M_i) = 0$ for $i = 2,\ldots, n$. We conclude

$$\int 1_{\{X \in M_i\}} 1_{\{E(X|\mathscr{P}) \in C_1 \cap \bigcap_{j\geq 2} C_j^c \cap \tilde{C}\}} dP = \int \psi(\cdot, M_i) 1_{\{E(X|\mathscr{P}) \in C_1 \cap \bigcap_{j\geq 2} C_j^c \cap \tilde{C}\}} dP = 0 ,$$

for all $i = 2,\ldots, n$. Since $\sum_{j\geq 2} M_j = C_1^c$, the proposition follows.

From Proposition 2.1 we get the following special case of Theorem 2 in Pfanzagl (1974*b*).

COROLLARY 2.1. *Assume the situation of Lemma* 2.1. *If g is strictly convex and $g \circ E(X|\mathscr{P}) = E(g \circ X|\mathscr{P}) < \infty$ P-a.e., then $X = E(X|\mathscr{P})$ P-a.e.*

PROOF. The premise of Proposition 2.1 holds for each $h \in R^k$ if we put $n = 2$, $C_1 = \{x \in C: \langle h, x\rangle \leq 1\}$, $C_2 = \{x \in C: \langle h, x\rangle \geq 1\}$, and $\tilde{C} = C$. Let Q denote the set of rational numbers. Then

$$\begin{aligned}\{X \neq E(X|\mathscr{P})\} &= \bigcup_{h\in Q^k} \{\langle h, X\rangle > 1 > \langle h, E(X|\mathscr{P})\rangle\} \\ &= \bigcup_{h\in Q^k} (\{E(X|\mathscr{P}) \in C_1 \backslash C_2\} \cap \{X \in C_2 \backslash C_1\}) ,\end{aligned}$$

is a P-null set.

COROLLARY 2.2. *Assume the situation of Lemma* 2.1 *with* $C = [0, \infty)^k \backslash \{0\}$. *Let g be given as in Lemma* 1.3 *and $h \in (0, \infty)^k$ as in Lemma* 1.2(b). *If $g \circ E(X|\mathscr{P}) = E(g \circ X|\mathscr{P}) < \infty$ P-a.e., then the following assertions hold.*

(a) $\{E(X|\mathscr{P}) \in C_{hi} \backslash C_{hj}\} \cap \{E(X_k|\mathscr{P}) > 0\} \subset \{X \in C_{hi}\}[P]$ *for* $i \neq j$.

(b) $\{E(X_k|\mathscr{P}) > 0\} \subset \bigcup_{i=1}^{2} (\{E(X|\mathscr{P}) \in C_{hi}\} \cap \{X \in C_{hi}\})[P]$.

PROOF. (a) The premise of Proposition 2.1 holds with $\tilde{C} = \{x \in C: x_k > 0\}$, $n = 2$, $C_1 = C_{h1}$ and $C_2 = C_{h2}$.

(b) We have

$$\begin{aligned}&\{E(X_k|\mathscr{P}) > 0\} \Big\backslash \left(\bigcup_{i=1}^{2} (\{E(X|\mathscr{P}) \in C_{hi}\} \cap \{X \in C_{hi}\}) \right) \\ &\quad = \{E(X_k|\mathscr{P}) > 0\} \cap ((\{E(X|\mathscr{P}) \notin C_{h1}\} \cap \{X \notin C_{h2}\}) \\ &\qquad \cup (\{E(X|\mathscr{P}) \notin C_{h2}\} \cap \{X \notin C_{h1}\})) .\end{aligned}$$

By (a), the right hand side is a P-null set.

Some notation: Let $M \subset [0,\infty)^k$ have the property that for all $x \in [0,\infty)^k$ there is a $t \geq 0$ with $tx \in M$; for example $M = \left\{ x \in [0,\infty)^k \colon \sum_{i \in J} x_i = 1 \right\}$ where $\emptyset \neq J \subset \{1,\dots,k\}$. Suppose that $D \subset M \cap (0,\infty)^k$ is dense in M. For each $x, a \in R^k$ we put $x^{(a)} = (a_1 x_1, \dots, a_k x_k)$. The random vector $X^{(a)}$ is defined analogously.

COROLLARY 2.3. *Assume the situation of Lemma* 2.1 *with* $C = [0,\infty)^k$. *Let* g *and* $C_1, \dots, C_k$ *be given as in Lemma* 1.1. *Then the following assertions hold.*

(a) *If* $g \circ E(X \mid \mathscr{S}) = E(g \circ X \mid \mathscr{S})$ *P-a.e., then*

$$\left\{ E(X \mid \mathscr{S}) \in C_i \cap \bigcap_{j \neq i} C_j^c \right\} \subset \{X \in C_i\}[P],$$

for $i = 1, \dots, k$.

(b) *If* $g \circ E(X^{(b)} \mid \mathscr{S}) = E(g \circ X^{(b)} \mid \mathscr{S})$ *P-a.e. for all* $b \in D$, *then*

$$\Omega = \bigcup_{i=1}^{k} (\{E(X \mid \mathscr{S}) \in C_i\} \cap \{X \in C_i\})[P].$$

PROOF. (a) The premise of Proposition 2.1 holds with $\tilde{C} = C$ and $n = k$.

(b) We have

$$\bigcap_{i=1}^{k} (\{E(X \mid \mathscr{S}) \notin C_i\} \cup \{X \notin C_i\})$$
$$\subset \bigcup_{b \in D} \bigcup_{i=1}^{k} \left(\{E(X^{(b)} \mid \mathscr{S}) \in C_i \cap \bigcap_{j \neq i} C_j^c\} \right.$$
$$\left. \cap \{X^{(b)} \notin C_i\} \right) [P].$$

Replacing X in (a) by $X^{(b)}$, shows that the right hand side is a P-null set.

We shall modify Corollaries 2.2 and 2.3 in order to obtain criteria for sufficiency for finite sets of probability measures (see Corollaries 2.4 to 2.8). The essential result is Theorem 2.1. First we need another lemma. Suppose D and M are defined as above. Let $D_1 \subset (0,\infty)^{k-1}$ and $D_2 \subset (0,\infty)^k$ be given such that $\{(h_1 a_1, \dots, h_{k-1} a_{k-1}, a_k) \colon h \in D_1 \text{ and } a \in D_2\}$ is a dense subset of M.

LEMMA 2.2. *For all* $x, y \in [0,\infty)^k \setminus \{0\}$ *with* $y_k > 0$, *the following assertions are equivalent.*

(a) $x_k > 0$ *and* $x/x_k = y/y_k$.

(b) *For all* $h \in D_1$ *and* $a \in D_2$ *there is an* $i \in \{1,2\}$ *with* $x^{(a)}, y^{(a)} \in C_{hi}$, *where* C_{hi} *has the same meaning as in Lemma* 1.3.

(c) *For all* $b \in D$ *there is an* $i \in \{1,\dots,k\}$ *with* $x^{(b)}, y^{(b)} \in C_i$, *where* C_i *is defined as in Lemma* 1.1.

PROOF. Obviously, (b) and (c) follow from (a) since in these assertions the convex sets are even cones.

(b) implies (a): Since $y_k > 0$, there are $h \in D_1$ and $a \in D_2$ such that $y^{(a)} \in C_{h1} \setminus C_{h2}$. Hence $x^{(a)} \in C_{h1}$. This implies $x_k > 0$ since $x \neq 0$. We conclude that for all $h \in D_1$ and $a \in D_2$, there is an $i \in \{1,2\}$ such that $x^{(a)}/x_k, y^{(a)}/y_k \in C_{hi}$. Assume $x/x_k \neq y/y_k$. Then, there are $h \in D_1$ and $a \in D_2$ such that $x^{(a)}/x_k \in C_{h1} \setminus C_{h2}$ and $y^{(a)}/y_k \in C_{h2} \setminus C_{h1}$. This is a contradiction.

(c) implies (a): Since $y_k > 0$, there is a $b \in D$ with $y^{(b)} \in C_k$. Hence $x_k > 0$ since $x \neq 0$. Now we use induction on k. The case $k = 1$ is clear. Suppose the assertion holds for $k - 1 \geq 1$. If $x_i = 0$ for all $i < k$, we get $y_i = 0$ for all $i < k$. Assume that $x_i > 0$ for some $i < k$. Without loss of generality let $x_1 > 0$ and $x_k = y_k = 1$. Hence $y_1 > 0$. Put

$$\tilde{M} = \{(z_1,\dots,z_{k-1}) : (z_1,\dots,z_k) \in M \text{ for some } z_k < \sup\{z_i : i < k\}\},$$

$$\tilde{D} = \{(b_1,\dots,b_{k-1}) : (b_1,\dots,b_k) \in D \text{ for some } b_k < \sup\{b_i : i < k\}\}.$$

We get $x_i/x_1 = y_i/y_1$ for all $i < k$. Suppose $x_1 \neq y_1$. Then $x/x_1 \neq y/y_1$. Hence, there is a $b \in D$ with $x^{(b)}/x_1 \in C_j$ for some $j < k$ and $y^{(b)}/y_1 \in C_k$. This is a contradiction.

THEOREM 2.1. *Suppose that in Corollary* 2.3 *we have* $g \circ E(X^{(b)}|\mathscr{P}) = E(g \circ X^{(b)}|\mathscr{P})$ *P-a.e. for all* $b \in D$, *or suppose that in Corollary* 2.2 *we have* $g \circ E(X^{(a)}|\mathscr{P}) = E(g \circ X^{(a)}|\mathscr{P}) < \infty$ *P-a.e. for all* $a \in D_2$ *and f fulfills the assumption of Lemma* 1.2(b) *for all* $h \in D_1$. *Then the following assertions hold.*

(a) $\{E(X_k|\mathscr{P}) > 0\} = \{X_k > 0\}[P]$.

(b) $X/X_k 1_{\{X_k>0\}} = E(X|\mathscr{P})/E(X_k|\mathscr{P}) 1_{\{E(X_k|\mathscr{P})>0\}}$ *P-a.e.*

(c) *If* $\sum_{i=1}^{k} X_i = 1$ *P-a.e., then*

$$X 1_{\{X_k>0\}} = E(X|\mathscr{P}) 1_{\{E(X_k|\mathscr{P})>0\}} \qquad P\text{-a.e.}$$

PROOF. It is known that $\{X_k > 0\} \subset \{E(X_k|\mathscr{P}) > 0\}[P]$. Therefore (a) and (b) follow from Corollaries 2.2, 2.3 and Lemma 2.2.

(c): Because of (a) and (b) we have the following equalities

$$\begin{aligned} X1_{\{X_k>0\}} &= X1_{\{X_k>0\}}\Big/\Big(\sum_j X_j\Big) \\ &= E(X\,|\mathscr{S})\,1_{\{E(X_k|\mathscr{S})>0\}}\Big/\Big(\Big(1+\sum_{i=1}^{k-1} E(X_i|\mathscr{S})/E(X_k|\mathscr{S})\Big)E(X_k|\mathscr{S})\Big) \\ &= E(X\,|\mathscr{S})\,1_{\{E(X_k|\mathscr{S})>0\}}\Big/\Big(\sum_j E(X_j|\mathscr{S})\Big) \\ &= E(X\,|\mathscr{S})\,1_{\{E(X_k|\mathscr{S})>0\}} \qquad P\text{-a.e.} \end{aligned}$$

If in Theorem 2.1 the function f is equal to f_t defined in Example 1(d) $(0<t<1)$, we can put $D_2=\{(1,1)\}$ and the corresponding premise in Theorem 2.1 means that equality holds in Hölder's inequality for conditional expectations ($t=1/p$).

Now we turn to sufficiency. Let μ and $p_1,\dots,p_k$ be given as in Section 1. Define $p_i'=dP_i|\mathscr{S}/d\mu|\mathscr{S}$. We have $p_i'=E(p_i|\mathscr{S})$ μ-a.e. (conditional expectation with respect to μ), and $\mathscr{S}$ is sufficient for $\{P_1,\dots,P_k\}$ iff $p_i=p_i'$ for all $i=1,\dots,k$. The following corollaries are implied by Jensen's inequality and Theorem 2.1 where we put $P=\mu$ and $X_i=(1/k)p_i$ for all $i=1,\dots,k$.

COROLLARY 2.4. *$\mathscr{S}$ is sufficient for $\{P_1,\dots,P_k\}$ iff $\int\sup\{b_1p_1,\dots,b_kp_k\}\,d\mu=\int\sup\{b_1p_1',\dots,b_kp_k'\}\,d\mu$ for all $b\in D$.*

PROOF. Using Theorem 2.1(c) and the permutation invariance of the sup-function, we get

$$(p_1,\dots,p_k)\,1_{\{p_i>0\}}=(p_1',\dots,p_k')\,1_{\{p_i'>0\}} \qquad \mu\text{-a.e.}$$

for $1\le i\le k$.

By other methods Corollary 2.4 has been proved by Morse and Sacksteder ((1966), Theorem 2) for $D=M=[0,\infty)^k$. In the following let $\phi_2,\dots,\phi_k$ be permutations of $\{2,\dots,k\}$ with $\phi_i(k)=i$.

COROLLARY 2.5. *Assume the situation of Corollary 2.2. Suppose the premise of Lemma 1.2(b) is fulfilled for all $h\in D_1$. Then $\mathscr{S}$ is sufficient for $\{P_1,\dots,P_k\}$ if $\int g\circ(a_1p_1',a_2p_{\phi(2)}',\dots,a_kp_{\phi(k)}')\,d\mu=\int g\circ(a_1p_1,a_2p_{\phi(2)},\dots,a_kp_{\phi(k)})\,d\mu<\infty$ for all $a\in D_2$ and $\phi=\phi_2,\dots,\phi_k$.*

PROOF. From Theorem 2.1(c) we get

$$(p_1,p_{\phi(2)},\dots,p_{\phi(k)})\,1_{\{p_{\phi(k)}>0\}}=(p_1',p_{\phi(2)}',\dots,p_{\phi(k)}')\,1_{\{p_{\phi(k)}'>0\}} \qquad \mu\text{-a.e.}\,,$$

for $\phi = \phi_2, \ldots, \phi_k$. Hence $p_i = p_i'$ μ-a.e. for $i = 2, \ldots, k$; and this implies $p_1 = p_1'$ μ-a.e.

For $k = 2$, Corollary 2.5 has been proved in Mussmann (1979). Of course, in this case ϕ_2 is a trivial mapping.

COROLLARY 2.6. *Assume the situation of Corollary 2.2 with strictly convex f. Then $\mathscr{S}$ is sufficient for $\{P_1, \ldots, P_k\}$ if*

$$\int g \circ (p_1', p_{\phi(2)}', \ldots, p_{\phi(k)}') \, d\mu = \int g \circ (p_1, p_{\phi(2)}, \ldots, p_{\phi(k)}) \, d\mu < \infty ,$$

for $\phi = \phi_2, \ldots, \phi_k$.

PROOF. Since f is strictly convex, in Corollary 2.5 we can take $D_1 = (0, \infty)^{k-1}$ and $D_2 = \{(1, 1, \ldots, 1)\}$.

For $k = 2$, Corollary 2.6 has been proved by Csiszár ((1963), Satz 1, and (1967), p. 310).

COROLLARY 2.7. *Suppose $h \in (0, \infty)^{k-1}$ is fixed and $D_1 = \{h\}$. Then $\mathscr{S}$ is sufficient for $\{P_1, \ldots, P_k\}$ iff*

$$\int |a_1 h_1 p_1 + a_2 h_2 p_{\phi(2)} + \cdots + a_{k-1} h_{k-1} p_{\phi(k-1)} - a_k p_{\phi(k)}| \, d\mu$$

$$= \int |a_1 h_1 p_1' + a_2 h_2 p_{\phi(2)}' + \cdots + a_{k-1} h_{k-1} p_{\phi(k-1)}' - a_k p_{\phi(k)}'| \, d\mu ,$$

for all $a \in D_2$ and $\phi = \phi_2, \ldots, \phi_k$.

PROOF. Replace g in Corollary 2.5 by g_h from Example 1(b).

COROLLARY 2.8. Let $\tilde{D}$ be a dense subset of the set $\tilde{M} = \{(a_1, \ldots, a_{k-1}) \in (0, \infty)^{k-1} : a_1 + \cdots + a_{k-1} = 1\}$. Then $\mathscr{S}$ is sufficient for $\{P_1, \ldots, P_k\}$ iff for each $A \in \mathscr{A}$, $a \in \tilde{D}$, and $\phi = \phi_2, \ldots, \phi_k$ there is an $\mathscr{S}$-measurable test υ_A with

$$\int 1_A \, dP_{\phi(k)} \leq \int \upsilon_A \, dP_{\phi(k)}$$

and

$$\int 1_A \, d(a_1 P_1 + a_2 P_{\phi(2)} + \cdots + a_{k-1} P_{\phi(k-1)})$$

$$\geq \int \upsilon_A \, d(a_1 P_1 + a_2 P_{\phi(2)} + \cdots + a_{k-1} P_{\phi(k-1)}) .$$

PROOF. A short computation shows that the inequalities of Corollary 2.8 imply that the equalities in Corollary 2.7 hold with $h=(1,1,\ldots,1)$, $M=\tilde{M}\times\{1\}$, and $D_2=\tilde{D}\times\{1\}$.

For $k=2$, Corollary 2.8 is Pfanzagl's (1974*a*) characterization of sufficiency. The proof of Corollary 2.5 shows that in Corollaries 2.5 to 2.8 the permutations $\phi_2,\ldots,\phi_k$ are only needed if $k\geq 3$ and if $P_2,\ldots,P_{k-1}$ are not absolutely continuous with respect to P_k.

REFERENCES

Bauer, H. (1968). *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, de Gruyter, Berlin-New York.

Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten, *Publ. Math. Inst. Hung. Acad. Sci. Ser. A*, **8**, 85–107.

Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.*, **2**, 299–318.

Györfi, L. and Nemetz, T. (1977). *f*-dissimilarity: A general class of separation measures of several probability measures, *Colloq. Math. Soc. János Bolyai*, **16**, 309–321.

Kozek, A. and Suchanecki, Z. (1980). Multifunctions of faces for conditional expectations of selectors and Jensen's inequality, *J. Multivariate Anal.*, **10**, 579–598.

Morse, N. and Sacksteder, R. (1966). Statistical isomorphism, *Ann. Math. Statist.*, **37**, 203–214.

Mussmann, D. (1979). Sufficiency and *f*-divergences, *Studia Sci. Math. Hungar.*, **14**, 37–41.

Pfanzagl, J. (1974*a*). A characterization of sufficiency by power functions, *Metrika*, **21**, 197–199.

Pfanzagl, J. (1974*b*). Convexity and conditional expectations, *Ann. Probab.*, **2**, 490–494.

Rubin, H. and Wesler, O. (1958). A note on convexity in Euclidean *n*-space, *Proc. Amer. Math. Soc.*, **9**, 522–523.

Torgersen, E. N. (1976). Comparison of statistical experiments, *Scand. J. Statist.*, **3**, 186–208.