Studia Scientiarum Mathematicarum Hungarica 14 (1979), 37—41

SUFFICIENCY AND f-DIVERGENCES

by
D. MUSSMANN

0. Introduction and summary

We shall characterize sufficiency of a sub-σ-field for two probability measures by f-divergences, where f is a convex function which is not affine. This extends a theorem of CSISZÁR. As a corollary we obtain a criterion for sufficiency in terms of total variations. This criterion may be applied to prove PFANZAGL's characterization of sufficiency by power functions. An essential tool is a result on the attainment of equality in Jensen's inequality for conditional expectations for convex functions which are not necessarily strictly convex.

1. Preliminaries

The first part of this section is concerned with strict inequality in some well-known inequalities for convex functions. In the second part we state Neyman's criterion for sufficiency in a form suitable for our purposes.

Let f be a convex (continuous) function defined on $I:=]a, b[$, where $-\infty \leqq a < b \leqq +\infty$. Furthermore we shall suppose f to be not affine, i.e. not of type $\alpha x+\beta$. We denote the right derivative of f at $x\in I$ by $D^+(f; x)$ and define $[xy](f):=(f(x)-f(y))/(x-y)$, where $x, y\in I$ such that $x\neq y$. If a, resp. b, is finite, we define $f(a):=\lim_{r\downarrow a} f(r)$, resp. $f(b):=\lim_{r\uparrow b} f(r)$; $f(a)$, $f(b)$ may be $+\infty$.

The assumption that f is not affine is equivalent to each of the following two conditions:

(1.1) There exists an $x_0\in I$ such that

$$f(x_0) < \frac{y-x_0}{y-x} f(x) + \frac{x_0-x}{y-x} f(y) \quad \text{whenever} \quad x, y\in I \quad \text{and} \quad x < x_0 < y.$$

(1.2) There exists an $x_0\in I$ such that

$$f\big(tx+(1-t)\,y\big) < tf(x)+(1-t)f(y)$$

whenever

$$x, y\in I, \quad x < x_0 < y, \quad \text{and} \quad t\in]0, 1[.$$

Otherwise every $x_0\in I$ would be contained in an open interval on which f is affine (see [1], p. 232, Satz 2), hence f would be affine on I.

By the same reasoning as in [1], p. 197, 8.3.4.3 Satz, we get from (1.2):

(1.3) $[rx](f)<[yx](f)<[yr](f)$ whenever $x, y, r\in I$, $x<x_0<y$, and $x<r<y$ (same x_0 as in (1.2)).

Again by the same reasoning as in [1], p. 198, Satz (3a), we obtain from (1.3): (1.4) $f(y) > D^+(f; x)(y-x)+f(x)$ whenever $x, y \in I$ and $x < x_0 < y$ or $y < x_0 < x$. The strict inequality in (1.4) is even valid if a, resp. b, is finite and $y=a$, resp. $y=b$, because $[yx](f)$ is isotone in y.

If $b=\infty$, we have $\lim_{y\to\infty} [ry](f) = \lim_{y\to\infty} \frac{f(y)}{y}$. Combining this fact with (1.3), we have:

(1.5) $f(r) < f(x)+(r-x) \lim_{y\to\infty} \frac{f(y)}{y}$ whenever $r, x \in I$, $x < x_0$, and $x < r$.

We close this section with a variant of Neyman's criterion. Let P and Q be probability measures defined on a measurable space $(\Omega, \mathscr{A})$ and let $P=\mu_1+\mu_2$ be the Lebesgue decomposition of P with respect to Q, $\mu_1 \ll Q$ and $\mu_2 \perp Q$.

A sub-σ-field $\mathscr{S}$ of $\mathscr{A}$ is sufficient for P, Q iff

(1.6) there is an $\mathscr{S}$-measurable $\frac{d\mu_1}{dQ}$ and a $T \in \mathscr{S}$ such that $\mu_2(T)=0$ and $Q(T)=1$.

This assertion seems to be known. Concerning the "if" part, observe that the density with respect to μ_2+Q of P, resp. Q, is $(1-1_T)+1_T \cdot \frac{d\mu_1}{dQ}$, resp. 1_T.

2. Equality in Jensen's inequality

We employ the same notation as in 1; the point x_0 is of special importance. In the following, expectations are understood with respect to the probability measure Q. Let X be a Q-integrable function with values in $\bar{I}$ (closure of I in $]-\infty, +\infty[$).

(2.1) THEOREM. *If there is equality in Jensen's inequality* $f(\mathbf{E}(X|\mathscr{S})) \le \mathbf{E}(f\cdot X|\mathscr{S})[Q]$, *then* $Q\{X < x_0 < \mathbf{E}(X|\mathscr{S})\}=0$ *and* $Q\{\mathbf{E}(X|\mathscr{S}) < x_0 < X\}=0$.

PROOF. It is known that there exists a Markov kernel φ from $(\Omega, \mathscr{A})$ to $(\bar{I}, \mathscr{B}_1 \cap \bar{I})$ such that $\mathbf{E}(g\circ X|\mathscr{S}) = \int g(r)\varphi(\cdot, dr)[Q]$ whenever $\int g\circ X dQ$ exists.

Define $h(\omega) := \int r\varphi(\omega, dr)$, we may suppose $-\infty < h(\omega) < \infty$ for all $\omega \in \Omega$. Our premise can be rewritten in the form:

(2.2) There exists an $N \in \mathscr{A}$ of Q-measure zero such that $f(h(\omega)) = \int f(r)\varphi(\omega, dr)$ unless $\omega \in N$.

Now suppose ω to be not in N. We shall show

(2.3) $\varphi(\omega,]x_0, b])=0$ whenever $a \le h(\omega) < x_0$, and $\varphi(\omega, [a, x_0[)=0$ whenever $x_0 < h(\omega) \le b$.

First we assume $h(\omega) \notin I$, i.e. $-\infty < a = h(\omega)$ or $h(\omega) = b < +\infty$, then $\varphi(\omega, \cdot)$ equals Dirac measure in a, resp. in b, hence (2.3) is valid.

Suppose next that $h(\omega) \in I$. Replacing x with $h(\omega)$ in (1.4) we obtain $f(y) > D^+(f; h(\omega))\,(y-h(\omega))+f(h(\omega))$ whenever $y \in \bar{I}$ and $h(\omega) < x_0 < y$ or $y < x_0 < h(\omega)$ ($\ge$ holds for every $y \in \bar{I}$). Integrating with respect to $\varphi(\omega, \cdot)$ the assertion (2.3) follows.

To conclude the proof observe that for $A \in \mathscr{A}$ we have $\mathbf{E}(1_A|\mathscr{S})=\varphi(\cdot, A)[Q]$ and then using (2.3) we infer

$$Q\{X < x_0 < \mathbf{E}(X|\mathscr{S})\} = \int 1_{\{h > x_0\}} \cdot 1_{\{X < x_0\}}\, dQ = \int 1_{\{h > x_0\}} \varphi(\cdot, [a, x_0[)\, dQ = 0.$$

In an analogous manner we get $Q\{\mathbf{E}(X|\mathscr{S}) < x_0 < X\} = 0$.

REMARK. Pfanzagl ([6], p. 493, Theorem 2) proved that for strictly convex functions equality in Jensen's inequality implies $\mathbf{E}(X|\mathscr{S})=X[Q]$. This is easily obtained from (2.1). Indeed, for a strictly convex f the $x_0 \in I$ is arbitrary, hence, if M denotes the rational numbers of I, then

$$Q\{X \neq \mathbf{E}(X|\mathscr{S})\} \leqq \sum_{r \in M} Q(\{X < r < \mathbf{E}(X|\mathscr{S})\} \cup \{\mathbf{E}(X|\mathscr{S}) < r < X\}) = 0.$$

(2.4) COROLLARY. *Suppose* $a=0$ *and* $b=+\infty$. *Let* D *denote a countable dense subset of* I.

If $f(\mathbf{E}(\beta X|\mathscr{S}))=\mathbf{E}(f\circ(\beta X)|\mathscr{S})[Q]$ *for all* $\beta \in D$, *then* $\mathbf{E}(X|\mathscr{S})=X[Q]$.

PROOF. If D is dense in I, so is $\left\{\frac{x_0}{\beta}\colon \beta \in D\right\}$. Replacing X with βX in (2.1) we get

$$Q\{\mathbf{E}(X|\mathscr{S}) \neq X\} \leqq \sum_{\beta \in D} Q\left(\left\{\mathbf{E}(X|\mathscr{S}) < \frac{x_0}{\beta} < X\right\} \cup \left\{X < \frac{x_0}{\beta} < \mathbf{E}(X|\mathscr{S})\right\}\right) = 0.$$

REMARK. If X is bounded, $X < M < \infty$, then in (2.4) we only need $\left\{\frac{x_0}{\beta}\colon \beta \in D\right\}$ to be dense in $]0, M[$.

Now we shall investigate equality in another inequality. Suppose $a=0$ and $b=+\infty$ and let Y and Z be measurable functions on Ω with values in $[0, \infty[$. Replacing x with $Y(\omega)$ and r with $Y(\omega)+Z(\omega)$ in (1.5) we obtain:

(2.5) *If* $f(Y+Z)=f(Y)+Z \lim_{s\to\infty} \frac{f(s)}{s}[Q]$, *then* $Q\{Y < x_0,\ Z > 0\}=0$.

Furthermore, let D denote a countable subset of I with cluster point zero.

(2.6) *If* $f(\beta Y+\beta Z)=f(\beta Y)+\beta Z \lim_{s\to\infty} \frac{f(s)}{s}$ *for all* $\beta \in D$, *then* $Q\{Z>0\}=0$.

This holds because of (2.5) and $\Omega = \bigcup_{\beta \in D} \{\beta Y < x_0\}$.

REMARK. If f is strictly convex, x_0 is arbitrary, and therefore $Q\{Z>0\}=0$ follows even from the premise of (2.5). If Y is bounded, we need only one β in (2.6).

3. The main result and applications

In this section we present a condition in terms of f-divergences which is equivalent to sufficiency. This is an extension of a result obtained by Csiszár for a strictly convex f. Our condition is easily seen to be fulfilled if a certain condition for total variations or Pfanzagl's condition for power functions hold.

We employ the same notation as in 1. We suppose $a=0$ and $b=\infty$. For $\beta>0$ the Lebesgue decomposition of βP with respect to Q is $\beta P=\beta\mu_1+\beta\mu_2$.

(3.1) DEFINITION. $I_f(\beta P, Q):=\int f\circ\left(\beta\frac{d\mu_1}{dQ}\right)dQ+\beta\mu_2(\Omega)\lim_{s\to\infty}\frac{f(s)}{s}$ is called the *f-divergence* of βP with respect to Q (see [2], [3], [4]).

Set $P':=P|\mathscr{S}$ and in the same manner Q', μ_1', and μ_2'.

(3.2) THEOREM. *Let D denote a countable dense subset of I. Suppose $I_f(\beta P', Q')<<\infty$ for every $\beta\in D$. $\mathscr{S}$ is sufficient for P, Q iff $I_f(\beta P', Q')=I_f(\beta P, Q)$ for every $\beta\in D$.*

PROOF. The "only if" part is known (and follows at once from (1.6)), and the "if" part is known for strictly convex f with $\beta=1$ ([2], p. 90, Satz 1; [3], p. 310; [4], p. 141, Satz 17.2).

By the same reasoning as in [4], p. 145, one shows:

(3.3) $I_f(\beta P', Q')=I_f(\beta P, Q)$ iff $f\circ\mathbf{E}\left(\beta\frac{d\mu_1}{dQ}\middle|\mathscr{S}\right)=\mathbf{E}\left(f\circ\left(\beta\frac{d\mu_1}{dQ}\right)\middle|\mathscr{S}\right)$ $[Q]$ and

$$f\left(\mathbf{E}\left(\beta\frac{d\mu_1}{dQ}\middle|\mathscr{S}\right)+\beta\frac{dm_1}{dQ'}\right)=f\left(\mathbf{E}\left(\beta\frac{d\mu_1}{dQ}\middle|\mathscr{S}\right)\right)+\beta\frac{dm_1}{dQ'}\lim_{s\to\infty}\frac{f(s)}{s}\ [Q]$$

where m_1 denotes with respect to Q' absolutely continuous part of μ_2'.

Now, replacing X with $\frac{d\mu_1}{dQ}$ in (2.4) we get $\mathbf{E}\left(\frac{d\mu_1}{dQ}\middle|\mathscr{S}\right)=\frac{d\mu_1}{dQ}$ $[Q]$, hence there exists an $\mathscr{S}$-measurable $\frac{d\mu_1}{dQ}$.

Replacing Y with $\mathbf{E}\left(\frac{d\mu_1}{dQ}\middle|\mathscr{S}\right)$ and Z with $\frac{dm_1}{dQ'}$ in (2.6) we obtain $\frac{dm_1}{dQ'}=0$ $[Q]$, whence $Q'\perp\mu_2'$, i.e. there exists a $T\in\mathscr{S}$ such that $Q(T)=Q'(T)=1$ and $\mu_2(T)==\mu_2'(T)=0$. Then the sufficiency of $\mathscr{S}$ follows by (1.6).

(3.4) COROLLARY. *$\mathscr{S}$ is sufficient for P, Q iff $\|\beta P-Q\|=\|\beta P'-Q'\|$ for all $\beta\in D$ (D as in (3.2); $\|\ \|$ denotes the total variation).*

PROOF. Take $f(x)=|x-1|$. Let p and q be densities of P, resp. Q, with respect to $\nu:=P+Q$. It is straightforward to show that $I_f(\beta P, Q)=\int|\beta p-q|d\nu$ and

$$\|\beta P-Q\|=\max\left\{\frac{1}{2}\int|\beta p-q|\,d\nu+\frac{1}{2}(\beta-1),\ \frac{1}{2}\int|\beta p-q|\,d\nu-\frac{1}{2}(\beta-1)\right\}.$$

Hence $\|\beta P-Q\|=\|\beta P'-Q'\|$ iff

$$I_f(\beta P, Q)=I_f(\beta P', Q').$$

In order to characterize sufficiency, Pfanzagl ([5], p. 197) presents the following condition:

(3.5) For every $A\in\mathscr{A}$ there exists an $\mathscr{S}$-measurable test φ such that $P(A)\geqq\geqq\int\varphi\,dP$ and $Q(A)\leqq\int\varphi\,dQ$.

If (3.5) holds, then obviously for every $A\in\mathcal{A}$ there exists an $\mathcal{S}$-measurable test ψ such that $P(A)\leqq\int\psi\,dP$ and $Q(A)\geqq\int\psi\,dQ$.

The inequalities may be multiplied by $\beta>0$. It follows that for every $A\in\mathcal{A}$ there exists an $\mathcal{S}$-measurable test χ such that

$$|\beta P(A)-Q(A)|\leqq\left|\beta\int\chi\,dP-\int\chi\,dQ\right|\leqq\|\beta P'-Q'\|,$$

hence $\|\beta P-Q\|=\|\beta P'-Q'\|$ and (3.4) can be applied.

The following example shows that (3.4) is not valid with $D=\{1\}$.

EXAMPLE. $\Omega=\{1,2,3\}$, $\mathcal{A}=\mathcal{P}(\Omega)$. Let P, resp. Q, be given by $(p_1,p_2,p_3)=\left(\frac{1}{8},\frac{1}{4},\frac{5}{8}\right)$, resp. $(q_1,q_2,q_3)=\left(\frac{5}{8},\frac{1}{4},\frac{1}{8}\right)$. Let $\mathcal{S}$ denote the σ-field generated by $\{3\}$. $\frac{dP}{dQ}$ takes the values $\frac{1}{5}$, 1, 5, whence $\frac{dP}{dQ}$ is not $\mathcal{S}$-measurable and $\mathcal{S}$ not sufficient for P, Q. But $\|P-Q\|=\frac{4}{8}=\|P'-Q'\|$ and furthermore $\|\lambda-\mu\|=\|\lambda'-\mu'\|$ when λ and μ are in the convex hull of $\{P,Q\}$.

I wish to thank D. PLACHKY who suggested Corollary 3.4 and its application to Pfanzagl's characterization of sufficiency by power functions.

REFERENCES

[1] AUMANN, G. und HAUPT, D.: *Einführung in die reelle Analysis I*, De Gruyter, Berlin—New York, (1974).

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

[3] CSISZÁR, I.: Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hung.* **2** (1967), 299—318.

[4] HEYER, H.: *Mathematische Theorie statistischer Experimente*, Springer, Berlin—Heidelberg—New York, (1973).

[5] PFANZAGL, J.: A characterization of sufficiency by power function, *Metrika* **21** (1974), 197—199.

[6] PFANZAGL, J.: Convexity and conditional expectations, *Ann. Prob.* **2** (1974), 490—494.

(Eingegangen am 28. Juni 1977)

INSTITUT FÜR MATHEMATISCHE STATISTIK DER
WESTFÄLISCHEN WILHELMS-UNIVERSITÄT MÜNSTER
ROXELER STRASSE 64
D—4400 MÜNSTER
WEST GERMANY