SUFFICIENCY AND THE SEPARATION
OF STRONGLY CONVEX SETS OF PROBABILITY MEASURES

R. V. RAMAMOORTHI

(Communicated by William D. Sudderth)

ABSTRACT. We show that the equivalence of two notions of sufficiency is related to the separation of two orthogonal, strongly convex sets of probability measures by a universally measurable set.

Consider a measurable space $(\mathscr{X}, \mathscr{A})$, and let $M(\mathscr{X})$ be the set of all probability measures on $(\mathscr{X}, \mathscr{A})$. Equip $M(\mathscr{X})$ with the canonical σ-algebra $\mathscr{M}$ generated by the functions $m \to m(A)$, m in $M(\mathscr{X})$, and $A \in \mathscr{A}$. A subset M of $M(\mathscr{X})$ is said to be strongly convex if, for every probability measure μ on M, the barycenter $\int_M P\mu(dP)$ is in M. Two strongly convex sets M and N are orthogonal if for any $P \in M$ and $Q \in N$ there is a set $A_{P,Q}$ in A such that $P(A_{P,Q}) = 1$ and $Q(A_{P,Q}) = 0$. M and N are said to be uniformly orthogonal if there is a set A in $\mathscr{A}$ such that $P(A) = 1$ for all P in M and $Q(A) = 0$ for all Q in N; in such a case, A is said to separate M and N. A natural question in this context is, "if two strongly convex sets M and N are orthogonal, then are they uniformly orthogonal?" This question has received much attention in recent times, for instance see [8] and the references therein. We show in this article that the above question is equivalent to a problem arising in the study of "sufficiency" in mathematical statistics.

Throughout this paper, we assume that $(\mathscr{X}, \mathscr{A})$ is a standard Borel space and that $\mathscr{B} \subset \mathscr{A}$ is a countably generated σ-algebra. If M is an analytic set of probability measures on $(\mathscr{X}, \mathscr{A})$, then $(\mathscr{X}, \mathscr{A}, M)$ is called a standard Borel experiment.

Let P be a probability measure on $(\mathscr{X}, \mathscr{A})$. For any bounded $\mathscr{A}$-measurable function f, the conditional expectation of f given $\mathscr{B}$ under P is a function g such that (i) g is $\mathscr{B}$-measurable and (ii) $\int_B g dP = \int_B f dP$ for all B in $\mathscr{B}$. We shall denote by $E_P(f|\mathscr{B})$ any version of the conditional expectation of f given $\mathscr{B}$. Doob [4] defined a somewhat weaker notion of conditional expectation. According to Doob, $\overline{g}$ is a conditional expectation of f given $\mathscr{B}$ if (i$'$) $\overline{g}$ is measurable with respect to the P-completion of $\mathscr{B}$ and (ii$'$) $\int_B \overline{g} dP = \int_B f dP$ for all B in $\mathscr{B}$.

Let M be a family of probability measures on $(\mathscr{X}, \mathscr{A})$ and, as before, let $\mathscr{B}$ be a sub-σ-algebra of $\mathscr{A}$. We now define various notions of "sufficiency" of $\mathscr{B}$ for $(\mathscr{X}, \mathscr{A}, M)$. The following definition, essentially due to Fisher, is the one prevalent in statistical theory.

Definition 1. $\mathscr{B}$ is said to be *sufficient* for $(\mathscr{X}, \mathscr{A}, M)$ if, for any bounded $\mathscr{A}$-measurable function f, there is a $\mathscr{B}$-measurable function g such that for all P in M, $E_P(f|\mathscr{B}) = g[P]$.

From a different standpoint, Kolmogorov [5] proposed another notion of sufficiency which we shall call Bayes sufficiency. According to Kolmogorov, $\mathscr{B}$ is Bayes sufficient if for every "prior distribution" ξ on M, the "posterior distribution" on M given $\mathscr{A}$ is the same as that given $\mathscr{B}$. Formally let any ξ, a probability on M, define a measure λ_ξ on $M \times \mathscr{X}$ by $\lambda_\xi(E \times A) = \int_E P(A)\xi(dP)$.

Definition 2. $\mathscr{B}$ is *Bayes sufficient* for M if, for all ξ and E in $\mathscr{E}$,

$$E_{\lambda_\xi}(I_{E \times \mathscr{X}}|\overline{\mathscr{A}}) = E_{\lambda_\xi}(I_{E \times \mathscr{X}}|\overline{\mathscr{B}})[\lambda_\xi],$$

where $\overline{\mathscr{A}} = \{M \times A : A \in \mathscr{A}\}$ and $\overline{\mathscr{B}} = \{M \times B : B \in \mathscr{B}\}$.

Motivated by Doob's definition of conditional expectation, we define yet another concept of sufficiency. This concept, while essentially a technical variation of Definition 1, serves to elucidate the relationship between Bayes sufficiency and sufficiency. Towards this end, we define the universal completion $\hat{\mathscr{B}}$ of $\mathscr{B}$ as

$$\hat{\mathscr{B}} = \cap\{\mathscr{B}^P : P \in M_{\mathscr{B}}(\mathscr{X})\},$$

where $M_{\mathscr{B}}(\mathscr{X})$ is the set of all probability measures on $(\mathscr{X}, \mathscr{B})$ and $\mathscr{B}^P$ is the usual completion of $\mathscr{B}$ with respect to P. We next state a lemma which is used often in this paper. The proof of the lemma can be found in [11, Lemma 2.2, Corollary].

Lemma 1. *Any probability measure P on $(\mathscr{X}, \mathscr{B})$ admits an extension to a probability measure P^* on $(\mathscr{X}, \mathscr{A})$. Equivalently, if $T : \mathscr{X} \to [0, 1]$ is measurable and P is a probability measure on the Borel σ-algebra of $[0, 1]$, then there is a probability measure P^* on $(\mathscr{X}, \mathscr{A})$ such that $P = P^* T^{-1}$.*

The following proposition gives alternative descriptions of $\hat{\mathscr{B}}$. Lemma 1 yields (i), and (ii) can be proved by using the von Neumann selection theorem and the fact that the class of universally measurable functions is closed under composition.

Proposition 1. (i) $\hat{\mathscr{B}} = \cap\{\mathscr{B}^{P|\mathscr{B}} : P \in M(\mathscr{X})\}$. (ii) *If B is a universally measurable subset of $\mathscr{X}$ and is a union of $\mathscr{B}$ atoms, then $B \in \hat{\mathscr{B}}$.*

Definition 3. $\mathscr{B}$ is *Doob sufficient* for $(\mathscr{X}, \mathscr{A}, M)$ if, for any bounded $\mathscr{A}$-measurable function f, there is a $\hat{\mathscr{B}}$-measurable function g such that for all P in M $g = E_P(f|\mathscr{B})[P]$.

It is easy to see that if $\mathscr{B}$ is sufficient, then it is Doob sufficient. We shall next show that Doob sufficiency implies Bayes sufficiency.

Proposition 2. *If M is dominated by a σ-finite measure λ and $\mathscr{B}$ is Doob sufficient for M, then $\mathscr{B}$ is sufficient for M.*

Proof. Let f be a bounded $\mathscr{A}$-measurable function. Then there is a bounded $\hat{\mathscr{B}}$-measurable function g such that

$$\int_B g\, dP = \int_B f\, dP \quad \text{for all } B \in \mathscr{B} \text{ and } P \in M.$$

Now get a $\mathscr{B}$-measurable g such that $\overline{g} = g[\lambda]$. Since $P \ll \lambda$ for every P in M, $\overline{g} = g[P]$, and hence

$$\int_B \overline{g}\, dP = \int_B f\, dP$$

for every B in $\mathscr{B}$ and $P \in M$. This establishes the sufficiency of $\mathscr{B}$.

Corollary. *If $\mathscr{B}$ is Doob sufficient for M, then $\mathscr{B}$ is pairwise sufficient for M; i.e., $\mathscr{B}$ is sufficient for every pair $\{P, P'\} \subset M$.*

The next easy result relates Bayes sufficiency to sufficiency and is proved in [10].

Proposition 3. *$\mathscr{B}$ is Bayes sufficient for M iff, for each probability measure ξ on M, there is a set E_ξ in $\mathscr{M}$ such that $\xi(E_\xi) = 1$ and $\mathscr{B}$ is sufficient for E_ξ.*

The next proposition is an immediate consequence of Proposition 25.3A of [7].

Proposition 4. *$\mathscr{B}$ is Bayes sufficient iff, for every prior ξ on M and every bounded $\mathscr{A}$-measurable function $f: \mathscr{X} \to \mathbb{R}$, there is a bounded, $\mathscr{B}$-measurable function $g: \mathscr{X} \to \mathbb{R}$ such that*

$$\int_E \int_B f(x)\, dP(x)\, d\xi(P) = \int_E \int_B g(x)\, dP(x)\, d\xi(P)$$

for all E-measurable subset of M and $B \in \mathscr{B}$.

If M is a subset of $M(\mathscr{X})$, then denote by $\hat{M}$ the strongly convex set generated by M. That is, $\hat{M} = \{P_\xi: \xi \text{ a probability measure on } M\}$, where P_ξ is the measure $P_\xi(A) = \int P(A)\, d\xi(P)$.

Theorem 1. *$\mathscr{B}$ is Bayes sufficient for M iff $\mathscr{B}$ is pairwise sufficient for $\hat{M}$.*

Proof. Suppose that $\mathscr{B}$ is Bayes sufficient for M. If P_{ξ_1} and P_{ξ_2} are in $\hat{M}$, then with $\xi = (\xi_1 + \xi_2)/2$ we have by Proposition 3, a set $E_\xi \subset M$ such that $\mathscr{B}$ is sufficient for E_ξ. Therefore, given that f is $\mathscr{A}$-measurable, there is a $\mathscr{B}$-measurable function g_ξ such that $g_\xi = E_P(f|\mathscr{B})[P]$, $P \in E_\xi$. Since $\xi_1(E_\xi) = 1$ and $\xi_2(E_\xi) = 1$, it is easy to see that

$$g_\xi = E_{P_{\xi_1}}(f|\mathscr{B}) \quad \text{and} \quad g_\xi = E_{P_{\xi_2}}(f|\mathscr{B}),$$

so that $\mathscr{B}$ is sufficient for $(\mathscr{X}, \mathscr{A}, \{P_{\xi_1}, P_{\xi_2}\})$.

For the converse, let $\mathscr{M}_n$ be an increasing sequence of finite algebras generating the canonical σ-algebra $\mathscr{M}$ on M. Denote the atoms of $\mathscr{M}_n$ by $E_n^1, \ldots, E_n^{k(n)}$. For any probability measure ξ on M, define ξ_i^n by

$$\xi_i^n(E) = \frac{\xi(E \cap E_n^i)}{\xi(E_n^i)} \quad \text{for } E \in \mathscr{M}$$
$$= \xi(E) \quad \text{if } \xi(E_n^i) = 0.$$

Since pairwise sufficiency implies sufficiency for every finite collection, we have that $\mathscr{B}$ is sufficient for $\{P_{\xi_1^n}, \ldots, P_{\xi_{k(n)}^n}\}$. Therefore, for any bounded $\mathscr{A}$-measurable function f, we have a $\mathscr{B}$-measurable function g_n such that

$$g_n = E_{P_{\xi_i^n}}(f | \mathscr{B}) \quad \text{for } i = 1, \ldots, k(n).$$

Hence g_n is a version of the conditional expectation of f given $\mathscr{M}_n \times \mathscr{B}$ under the probability measure λ_ξ. Set

$$g(x) = \begin{cases} \lim_n g_n(x) & \text{if } \lim_n g_n(x) \text{ exists} \\ 0 & \text{otherwise} \end{cases}.$$

Since $\mathscr{M}_n \times \mathscr{B} \uparrow \mathscr{M} \times \mathscr{B}$, the martingale convergence theorem ensures that g is a version of the conditional expectation of f given $M \times \mathscr{B}$ under the probability measure λ_ξ. So, by Proposition 4, $\mathscr{B}$ is Bayes sufficient.

Theorem 2. *If $\mathscr{B}$ is Doob sufficient, then $\mathscr{B}$ is Bayes sufficient.*

Proof. If $\mathscr{B}$ is Doob sufficient for M, then it is easy to see that $\mathscr{B}$ is Doob sufficient for the strongly convex set $\hat{M}$ generated by M. By Proposition 1, $\mathscr{B}$ is pairwise sufficient for $\hat{M}$, and by Theorem 1 it follows that $\mathscr{B}$ is Bayes sufficient for M. □

We have thus shown that sufficiency implies Doob sufficiency, which in turn implies Bayes sufficiency. Even under the standard Borel assumptions that we have made, however, it is known that Bayes sufficiency does not imply sufficiency. An example to this effect was given in [2]. This example uses an earlier example from [1] of two strongly convex Borel sets M and N of probability measures which are orthogonal but not uniformly so. The next theorem exhibits the connection between the sufficiency problem and the separation of strongly convex sets of probability measures.

Theorem 3. *In ZFC the following statements are equivalent.*
   (i) *Any two orthogonal, strongly convex, analytic sets of probability measures on the unit interval can be separated by a universally measurable set.*
   (ii) *In any standard Borel experiment, a countably generated sub-σ-algebra is Doob sufficient iff it is Bayes sufficient.*

The following lemma is useful in proving the theorem.

Lemma 2. *Suppose* (i) *of the theorem is known to be true in a particular case. Then, if M and N are strongly convex, analytic sets of probability measures on $(\mathscr{X}, \mathscr{A})$ such that, for every $P \in M$ and $Q \in M$, there is a B in $\mathscr{B}$ such that $P(B) = 1$ and $Q(B) = 0$, then M and N can be separated by a set in $\hat{\mathscr{B}}$.*

Proof of lemma. Since $\mathscr{B}$ is countably generated, there is a function ψ $\mathscr{X} \rightarrow [0, 1]$ such that $(\mathscr{X}, \mathscr{B})$ and $\psi(\mathscr{X})$ equipped with its Borel σ-algebra are isomorphic. Further, the sets $\{P\psi^{-1} : P \in M\}$ and $\{Q\psi^{-1} : Q \in N\}$, thought of as measures on $[0, 1]$, are analytic and, by Lemma 1, strongly convex. By (i) there is a universally measurable subset F of $[0, 1]$ which separates the two sets of probability measures. $\psi^{-1}(F)$ separates M and N and, by Proposition 1, is in $\hat{\mathscr{B}}$. □

Proof of theorem. (i) $\Rightarrow$ (ii). Suppose that the standard Borel experiment consists of the analytic set M of probability measures on the standard Borel space $(\mathscr{X}, \mathscr{A})$ and that $\mathscr{B}$ is a countably generated sub-σ-algebra of $\mathscr{A}$ which is Bayes sufficient for M.

Let f be a bounded, $\mathscr{A}$-measurable function. Choose a version $g(x, P)$ of $E_P(f|\mathscr{B})$ which is $\mathscr{B} \times \mathscr{M}$-measurable. Assume without loss of generality that $0 \leq f \leq 1$ and $0 \leq g(x, P) \leq 1$. For each n and i such that $0 \leq i \leq n$, define

$$B_{n,i}^{P} = \left\{ x : \frac{i-1}{n} < g(x, P) \leq \frac{i}{n} \right\}.$$

Let $M_{n,i} = \{P \in M : P(B_{n,i}^{P}) > 0\}$. For each P in $M_{n,i}$, define the restriction $\overline{P}_{n,i}$ by $\overline{P}_{n,i}(A) = \frac{P(A \cap B_{n,i}^{P})}{P(B_{n,i}^{P})}$. Set $\overline{M}_{n,i} = \{\overline{P} : P \in M_{n,i}\}$ and $\hat{M}_{n,i}$ to be the strongly convex set generated by $\overline{M}_{n,i}$. Note that $M_{n,i}$ is Borel in M and the function $T(P) = \overline{P}_{n,i}$ from $M_{n,i}$ onto $\overline{M}_{n,i}$ is measurable. This shows that $\overline{M}_{n,i}$ is analytic and, consequently [3, p. 196], $\hat{M}_{n,i}$ is also analytic.

We next show that $\hat{M}_{n,i}$ are pairwise orthogonal. In fact we shall show that if $P_1 \in \hat{M}_{n,i}$ and $P_2 \in \hat{M}_{n,j}$, $i \neq j$, then there is a set B in $\mathscr{B}$ such that $P_1(B) = 1$ and $P_2(B) = 0$. To see this, let ξ be a probability measure on $\overline{M}_{n,i}$ and η be a probability measure on $\overline{M}_{n,j}$ such that $P_\xi = P_1$ and $P_\eta = P_2$. Let ξ^* and η^* be liftings of the measures ξ and η to $M_{n,i}$ and $M_{n,j}$, as provided by Lemma 1. Consider the measure $\mu = \xi^*/2 + \eta^*/2$. Then there exists, by Proposition 3, a measurable set $E \subset M$ and a $\mathscr{B}$-measurable function g such that $\mu(E) = 1$ and $g(x) = g(x, P)[P]$ for all P in E. Since $B_{n,i}^{P} \in \mathscr{B}$, if $P \in M_{n,i} \cap E$, we have $g(x) = g(x, P)[\overline{P}_{n,i}]$. Therefore $B = \{x : (i-1)/n < g(x) \leq i/n\}$ satisfies $\overline{P}_{n,i}(B) = 1$ for $P \in \overline{M}_{n,i} \cap E$, and since $\xi^*(M_{n,i} \cap E) = 1$, we have $P_1(B) = 1$. Similarly $P_2(B) = 0$.

Using the lemma and an easy argument, we get disjoint sets $U_{n,1}$, $U_{n,2}$, $\ldots$, $U_{n,n}$ in $\hat{\mathscr{B}}$ such that $P(U_{n,i}) = 1$ for P in $\hat{M}_{n,i}$ and $P(U_{n,i}) = 0$ for

P in $\hat{M}_{n,j}$. For each n, define $g_n(x) = \sum_{i=1}^{n} \frac{i}{n} I_{U_{n,i}}$. It is easily seen that

$$g(x) = \begin{cases} \lim_n g_n(x) & \text{if it exists} \\ 0 & \text{otherwise} \end{cases}$$

satisfies $g(x) = g(x, P)[P]$ for all P in M. Since each $U_{n,i}$ is in $\hat{\mathscr{B}}$, $g(x)$ is $\hat{\mathscr{B}}$-measurable. This establishes the Doob sufficiency of $\mathscr{B}$.

(ii) $\Rightarrow$ (i). Let M_0 and M_1 be two orthogonal, strongly convex, analytic sets of probability measures on $[0, 1]$. Denote by $\mathscr{Y}_0 = \{0\} \times [0, 1]$ and $\mathscr{Y}_1 = \{1\} \times [0, 1]$.

Set $\mathscr{X} = \mathscr{Y}_0 \cup \mathscr{Y}_1$, $\mathscr{A}$ to be the Borel σ-algebra on $\mathscr{X}$, and π to be the function $\pi(i, y) = y$ for $i = 0, 1$. For P in M_0, define the probability measure $\overline{P}$ on $\mathscr{X}$ by $\overline{P}(A) = P(\pi(A \cap \mathscr{Y}_0))$ and $\overline{M}_0$ to be $\{\overline{P}: P \in M_0\}$. Similarly, for Q in M_1, define $\overline{Q}$ by $\overline{Q}(A) = Q(\pi(A \cap \mathscr{Y}_1))$ and $\overline{M}_1 = \{\overline{Q}: Q \in M_1\}$. Note that $[0, 1]$, $\mathscr{Y}_0$, $\mathscr{Y}_1$, are all isomorphic, and so are $M_0, \overline{M}_0$ and $M_1, \overline{M}_1$. We shall next show that $\mathscr{B}$, the σ-algebra generated by π, is Bayes sufficient for $\overline{M} = (\overline{M}_0 \cup \overline{M}_1)$.

Towards this end, note that $\mathscr{B}$ is pairwise sufficient for $\overline{M}$. This is so because, if $P \in M_0$ and $Q \in M_1$, then $I_{\pi^{-1}(B)}$, where B is a separating set for P, Q, is a version of $d\overline{P}/(d\overline{P} + \overline{Q})$. Also, if $P, Q \in M_0$, then $(dP/(dP + Q)) \circ \pi$ is a version of $d\overline{P}/(d\overline{P} + \overline{Q})$. A similar argument works when $P, Q \in M_1$.

In order to establish Bayes sufficiency, since $\overline{M}_0$ and $\overline{M}_1$ are strongly convex, the strongly convex set generated by $\overline{M}$ is the set of all measures of the form $\alpha \overline{P}_1 + (1-\alpha)\overline{Q}$, where $\overline{P} \in \overline{M}_0$, $\overline{Q} \in \overline{M}_1$, and $0 \le \alpha \le 1$. It is known [6] that if $\mathscr{B}$ is pairwise sufficient for $\overline{M}$, it is also pairwise sufficient for the convex set (note: not strongly convex) generated by $\overline{M}$. By Theorem 1, $\mathscr{B}$ is Bayes sufficient for $\overline{M}$.

(ii) now implies that $\mathscr{B}$ is Doob sufficient for $\overline{M}$. Let f^* be a $\hat{\mathscr{B}}$-measurable function such that

$$f^* = E_P(I_{\mathscr{Y}_0}|\mathscr{B})[p] \quad \text{for all } P \in \overline{M}.$$

Then it is easily verified that $\pi\{x: f^*(x) = 1\}$ is universally measurable and separates M_0 and M_1.

Under Martin's Axiom, Mokobodzki [8, Theorem 3.7; 3, Chapter XI] showed that (i) of Theorem 3 holds. In view of Theorem 3, we can now state

Theorem 4. *Assume $ZFC + MA$. In any standard Borel space Bayes sufficiency is equivalent to Doob sufficiency.*

The only two cases when Bayes sufficiency is known to imply sufficiency are those in which the family P is dominated by a σ-finite measure and those in which the measures in P are all discrete [9]. In both these situations, pairwise sufficiency itself implies sufficiency, and consequently the additional assumption of Bayes sufficiency plays no essential part. In view of some positive results known about the problem of separating orthogonal, strongly convex sets

of probability measures [8], it is likely that there would be cases other than the dominated and discrete when Bayes sufficiency would imply sufficiency. We, however, do not know any results in this direction.

Acknowledgment

References

1. D. Blackwell, *There are no Borel SPLIFs*, Ann. Probab. **8** (1980), 1189–1190.

2. D. Blackwell and R. V. Ramamoorthi, *A Bayes but not classically sufficient statistic*, Ann. Statist. **10** (1982), 1025–1026.

3. C. Dellacherie and P. A. Meyer, *Probabilities and potential C*, North-Holland, 1988.

4. J. L. Doob, *Stochastic processes*, Wiley, New York, 1953.

5. A. N. Kolmogorov, *Definition of center of dispersion and measure of accuracy from a finite number of observations*, Izv. Akad. Nauk Er. Mat. **6** (1942), 3–32. (in Russian)

6. E. L. Lehmann, *Testing statistical hypotheses*, 2nd ed., Wiley, New York, 1986.

7. M. Loeve, *Probability theory*, Van Nostrand, 1955.

8. R. D. Mauldin, D. Preiss, and H. V. Weizsacker, *Orthogonal transition kernels*, Ann. Probab. **11** (1983), 970–988.

9. R. V. Ramamoorthi, *On pairwise sufficiency and sufficiency in standard Borel spaces*, Sankhya Ser. A **42** (1980), 139–145.

10. K. K. Roy and R. V. Ramamoorthi, *Relationship between Bayes, classical and decision theoretic sufficiency*, Sankhya Ser. A **41** (1979), 48–58.

11. V. S. Varadarajan, *Groups of automorphisms of Borel spaces*, Trans. Amer. Math. Soc. **109** (1963), 191–220.

DEPARTMENT OF STATISTICS AND PROBABILITY, MICHIGAN STATE UNIVERSITY, EAST LANSING, MICHIGAN 48824