



Marginal sufficiency

B V RAO

Chennai Mathematical Institute, SIPCOT IT Park, Kelambakkam 603 103, India
E-mail: bvrao@cmi.ac.in

MS received 30 June 2020; accepted 30 September 2020

Abstract. A proof that marginal sufficiency implies sufficiency is presented.

Keywords. Marginal sufficiency; conditional expectation; dominated family.

2010 Mathematics Subject Classification. 62B05, 62B10, 60A10.

1. Introduction

This note contains no new result. In 1968, J K Ghosh wrote an article ‘Marginal Sufficiency Implies Sufficiency’. It was Tech Report No. 54 of Research and Training School, Indian Statistical Institute. Here is the abstract as appeared in [1]:

“Suppose that X_1, \dots, X_n are independently distributed random variables and $T(X_1, \dots, X_n)$ is a statistic sufficient for each X_i . Is T sufficient for (X_1, \dots, X_n) ? In other words, does marginal sufficiency imply sufficiency if X_1, \dots, X_n are independent? The rather surprising answer to this question (given in Theorem 1) is yes, at least in the dominated case. (Of course, the answer is no if the X ’s are dependent.)

The problem arose in an unpublished paper of Dr. V S Huzurbazar who conjectured the correct answer.”

It was a very difficult paper to read. Few years later, R R Bahadur presented, in a lecture at ISI, Kolkata, an understandable proof in a special case. Soon after (around 1974), the present author worked out, in full generality, a proof of this theorem for private circulation. None of these was ever published.

In the meanwhile, the Russian statistician V N Sudakov [5] published his version. This is what he says:

“Several years ago, the Indian statistician V S Huzurbazar advanced the hypothesis that in the case of a repeated sample, the marginal sufficiency of a statistic implies its sufficiency. The proof of this hypothesis was announced in a preprint published by Ghosh in 1968, but my colleagues and I, all specialists in mathematical statistics, have run into considerable difficulties in attempting to reconstruct the complete proof from Ghosh’s preprint. We have not even been able to extricate a reasonably clear idea that might be of help in arriving at the required proof. Below we set forth the proof of a somewhat more general proposition.”

Many years later, the Japanese statistician Hirokichi Kudo [3] also published his version. He extends this concept and relates to his work with Edward Barankin and Kusama.

These published proofs at various generalities, which I saw recently, lack Bahadurian elegance and beauty. It appeared worthwhile to put on record the complete proof of the circulated notes mentioned above.

2. The result

Consider R^n with its Borel sigma-field \mathcal{B} . Let \mathcal{P} be a family of probabilities on \mathcal{B} . Let $\mathcal{A} \subset \mathcal{B}$ be a sub σ -field. Say that \mathcal{A} is marginally sufficient for \mathcal{P} if every real valued bounded Borel function on R^n depending only on one coordinate has a common version of conditional expectation given \mathcal{A} . More precisely, if $f : R \rightarrow R$ is a bounded Borel function and $1 \leq i \leq n$ and $f_i(x_1, \dots, x_n) = f(x_i)$; then there is an \mathcal{A} measurable h_i such that

$$E_P(f_i|\mathcal{A}) = h_i \quad \text{a.e. } [P] \quad \text{for all } P \in \mathcal{P}.$$

Recall that \mathcal{A} is sufficient for \mathcal{P} if given a bounded Borel function f on R^n , there is a bounded \mathcal{A} -measurable h such that

$$E_P(f|\mathcal{A}) = h \quad \text{a.e. } [P] \quad \text{for all } P \in \mathcal{P}.$$

The family \mathcal{P} is dominated if there is a probability μ on \mathcal{B} such that $P \ll \mu$ for all $P \in \mathcal{P}$. Say that \mathcal{P} is a family of product probabilities if each $P \in \mathcal{P}$ is a product probability, that is, of the form $P = \nu_1 \times \nu_2 \times \dots \times \nu_n$ for some probabilities ν_i on R . Here then is the result.

Theorem 1. *If \mathcal{P} is a dominated family of product probabilities on R^n and if $\mathcal{A} \subset \mathcal{B}$ is marginally sufficient, then \mathcal{A} is sufficient.*

Statistical intuition goes as follows. A statistic (that is, sigma-field) is sufficient for a parameter (that is, a family of probabilities) if it contains all the information of the sample (X_1, \dots, X_n) . If the sample is independent (that is, if the probabilities are product probabilities), then this is same as sum of the informations in each X_i and thus enough to contain all the information about each of the random variables.

The concept of marginal sufficiency is due to V S Huzurbazar (in an unpublished paper) and he conjectured the theorem above. It was proved by J K Ghosh. The proof we present below follows ideas of R R Bahadur.

3. Proof

We need to deal with expressions having zero in the denominator. To avoid interruption in what follows, we make temporary conventions: a/b is taken as zero if $a = 0$. $a/0$ is taken as $+\infty$ if $a > 0$. $\log 0 = -\infty$ and $\log \infty = \infty$. These conventions are only to help us write formulae neatly without casifying. All Radon–Nikodym derivatives are non-negative everywhere (may of course take value zero).

Suppose $(\Omega, \mathcal{B}, \mu)$ is a probability space; P, Q two probabilities on \mathcal{B} with $P \ll \mu$ and $Q \ll \mu$; $dP = f d\mu$ and $dQ = g d\mu$.

Here is a quick review of three well known facts:

(1) $\int \log \frac{f}{g} dQ$ exists, may be $-\infty$ but not $+\infty$.

Proof. Set

$$A = (f > 0; g = 0); \quad B = (f > 0; g > 0); \quad C = (f = 0, g > 0).$$

The facts $\int_B \frac{f}{g} dQ = \int_B f d\mu = P(B) < \infty$ and $\log \frac{f}{g} \leq \frac{f}{g}$ imply that the integral $\int_B \log \frac{f}{g} dQ$ exists and may be $-\infty$ but not $+\infty$. Further $\int_C \log \frac{f}{g} dQ$ is zero or $-\infty$ according as $Q(C) = 0$ or $Q(C) > 0$. Since $Q(B \cup C) = 1$, these two observations prove (1). □

(2) $\int \log \frac{f}{g} dP$ exists, may be $+\infty$ but not $-\infty$.

Proof. Similar argument as above applies or you can deduce from above. □

(3) $\int \log \frac{f}{g} dP = \int \log \frac{f}{g} dQ$ iff P is same as Q .

Proof. We only need to show that equality of integrals implies $P = Q$. Note that $Q(C) > 0$ implies $\int \log \frac{f}{g} dQ = -\infty$ and whereas $\int \log \frac{f}{g} dP \neq -\infty$. Thus hypothesis implies $Q(C) = 0$. Clearly, $Q(A) = 0$. Thus $Q(B) = 1$. Similarly we conclude that $P(A) = 0$ and since $P(C) = 0$, we have $P(B) = 1$. Thus P and Q are mutually absolutely continuous. Further the integrals are finite in view of (1) and (2). Since now $dP = \frac{f}{g} dQ$ we can rewrite the hypothesis as $\int [1 - \frac{f}{g}] \log \frac{f}{g} dQ = 0$. Since the integrand is always non-positive, it must be zero a.e. $[Q]$. □

Let us now further assume that $\mathcal{A} \subset \mathcal{B}$ is a sub sigmafield and $\mu^*(\omega, B)$ is regular conditional probability on \mathcal{B} given \mathcal{A} under μ .

Thus (i) for each fixed ω , the map $B \mapsto \mu^*(\omega, B)$ is a probability on \mathcal{B} and (ii) for each fixed $B \in \mathcal{B}$, the map $\omega \mapsto \mu^*(\omega, B)$ is a version of the conditional expectation $E_\mu(I_B | \mathcal{A})$. In particular, if $B \in \mathcal{B}$ and $A \in \mathcal{A}$, then $\mu(A \cap B) = E_\mu[A : \mu^*(\cdot, B)]$. Here we have used the notation $E_\mu(A : \varphi)$ for $\int_A \varphi d\mu$. We use $\mu^*(\omega, h)$ for $\int h(\eta) \mu^*(\omega, d\eta)$. Thus $\mu^*(\omega, I_B) = \mu^*(\omega, B)$, where as usual I_B is indicator function of the set B . Set

$$F = \{\omega : 0 < \mu^*(\omega, f) < \infty\}; \quad G = \{\omega : 0 < \mu^*(\omega, g) < \infty\}.$$

Then we have the following:

(4) $F, G \in \mathcal{A}$; $P(F) = 1$ and $Q(G) = 1$.

Proof. Since f is non-negative $\mu^*(\omega, f)$ makes sense and is \mathcal{A} measurable. Thus $F \in \mathcal{A}$. Since $\int f d\mu = 1$, we see $f(\eta)$ is $\mu^*(\omega, d\eta)$ integrable for a.e. ω $[\mu]$ and $\mu^*(\cdot, f)$ is in fact a version of $E_\mu(f | \mathcal{A})$. In particular, $P\{\omega : \mu^*(\omega, f) = \infty\} = 0$. Further, if $A = \{\omega : \mu^*(\omega, f) = 0\}$ then $A \in \mathcal{A}$ so that

$$P(A) = \int_A f d\mu = \int_A E_\mu(f | \mathcal{A}) d\mu = \int_A \mu^*(\omega, f) d\mu = 0.$$

Thus $P(F) = 1$. Similar argument holds for G . □

(5) Define

$$P^*(\omega, B) = \begin{cases} \frac{\mu^*(\omega, fI_B)}{\mu^*(\omega, f)} & \text{for } \omega \in F \\ \frac{\mu^*(\omega, gI_B)}{\mu^*(\omega, g)} & \text{for } \omega \in G - F \\ \mu^*(\omega, B) & \text{for } \omega \in (F \cup G)^c \end{cases}$$

Then P^* is regular conditional probability on \mathcal{B} given \mathcal{A} under P .

Proof. In view of (4), $0 < \mu^*(\omega, f) < \infty$ for $\omega \in F$ and $0 < \mu^*(\omega, g) < \infty$ for $\omega \in G$. So the definition of P^* makes sense. It is easy to see that for each ω , the map $B \mapsto P^*(\omega, B)$ is a probability. Fix $B \in \mathcal{B}$. The map $\omega \mapsto P^*(\omega, B)$ is \mathcal{A} measurable as observed in the proof of (4). Need to verify the integral equation that defines conditional expectation. Take any $A \in \mathcal{A}$.

$$\begin{aligned} \int_A P^*(\omega, B) dP &= \int I_{A \cap F} P^*(\omega, B) f d\mu && [P(F) = 1, dP = f d\mu] \\ &= E_\mu\{I_{A \cap F} P^*(\cdot, B) f\} && [\int \varphi d\mu = E_\mu\{\varphi\}, \text{notation}] \\ &= E_\mu\{I_{A \cap F} P^*(\cdot, B) E_\mu(f | \mathcal{A})\} && [A \cap F; P^* : \mathcal{A} \text{ measurable}] \\ &= E_\mu\{I_{A \cap F} P^*(\cdot, B) \mu^*(\cdot, f)\} && [E_\mu(f | \mathcal{A}) = \mu^*(\cdot, f)] \\ &= E_\mu\{I_{A \cap F} \mu^*(\cdot, f I_B)\} && [\text{definition of } P^* \text{ on } F] \\ &= E_\mu\{I_{A \cap F} E_\mu(I_B f | \mathcal{A})\} && [\mu^*(\cdot, \varphi) = E_\mu(\varphi | \mathcal{A})] \\ &= E_\mu\{E_\mu(I_{A \cap F} I_B f | \mathcal{A})\} && [A \cap F \in \mathcal{A}] \\ &= \int I_{A \cap F} I_B f d\mu && [\text{smoothing property}] \\ &= \int_{A \cap F} I_B dP && [dP = f d\mu] \\ &= \int_A I_B dP && [P(F) = 1]. \end{aligned}$$

Thus the map $\omega \mapsto P^*(\omega, B)$ is a version of $E_P(I_B | \mathcal{A})$. □

(6) Define

$$Q^*(\omega, B) = \begin{cases} \frac{\mu^*(\omega, gI_B)}{\mu^*(\omega, g)} & \text{for } \omega \in G \\ \frac{\mu^*(\omega, fI_B)}{\mu^*(\omega, f)} & \text{for } \omega \in F - G \\ \mu^*(\omega, B) & \text{for } \omega \in (F \cup G)^c. \end{cases}$$

Then Q^* is regular conditional probability on \mathcal{B} given \mathcal{A} under Q .

Proof. This is similar to above. □

(7) For any $A \in \mathcal{A}$, if $A \subset F \cap G$, then

$$P(A) = 0 \Leftrightarrow \mu(A) = 0 \Leftrightarrow Q(A) = 0.$$

Proof. Of course, $\mu(A) = 0$ implies $P(A) = 0 = Q(A)$. Let $A \subset F \cap G$; $A \in \mathcal{A}$; and $P(A) = 0$. Then $P(A) = \int_A f d\mu = \int_A E_\mu(f | \mathcal{A}) d\mu = 0$ but $E_\mu(f | \mathcal{A}) > 0$ on $A \subset F$ so that $\mu(A) = 0$; which in turn implies $Q(A) = 0$ too. Similar argument applies if we start with $Q(A) = 0$. □

(8) Let h be a \mathcal{B} -measurable function integrable under P, Q and μ . Suppose that there is a \mathcal{A} -measurable h^* such that $E_P(h | \mathcal{A}) = h^*$ a.e.[P] and $E_Q(h | \mathcal{A}) = h^*$ a.e.[Q]. Then there exists a set H such that the following hold:

- (a) $H \in \mathcal{A}$ and $P(H) = Q(H) = 1$.
 (b) for $\omega \in H$; h is integrable w.r.t. both $P^*(\omega, \cdot)$ and $Q^*(\omega, \cdot)$.
 (c) for $\omega \in H$, $P^*(\omega, h) = Q^*(\omega, h)$.

Proof. Set

$$I = \{\omega : P^*(\omega, |h|) = \infty\}; \quad J = \{\omega : Q^*(\omega, |h|) = \infty\} \quad N = I \cup J.$$

First observe that the above sets are in \mathcal{A} . Since h is P integrable and P^* is regular conditional probability given \mathcal{A} under P , we conclude that $P(I) = 0$. Using (7), we see $P(I \cap F \cap G) = 0 = Q(I \cap F \cap G)$. Similarly h being Q integrable and Q^* being regular conditional probability under Q , we have $Q(J) = 0$. Again by (7), $Q(J \cap F \cap G) = 0 = P(J \cap F \cap G)$. Thus $P(N \cap F \cap G) = 0 = Q(N \cap F \cap G)$.

Further on $F^c \cup G^c$, both P^* and Q^* coincide. Hence for $\omega \in F^c \cup G^c$ we have $\omega \in I$ iff $\omega \in J$. In particular, $N \cap (F^c \cup G^c) \subset I \cap J$. Now using the fact that $P(I) = 0 = Q(J)$, we conclude that P as well as Q measures of $N \cap (F^c \cup G^c)$ are zero.

Combining the two conclusions above, we have $P(N) = Q(N) = 0$. Set

$$K = \{\omega : h^*(\omega) \neq P^*(\omega, h)\} \cap F \cap G, \quad L = \{\omega : h^*(\omega) \neq Q^*(\omega, h)\} \cap F \cap G.$$

By hypothesis, $E_P(h|\mathcal{A}) = h^*$ a.e.[P] so that $P(K) = 0$ and so from (7), $Q(K) = 0$ as well. Similarly $Q(L) = P(L) = 0$. Thus $P(K \cup L) = Q(K \cup L) = 0$.

Let now $H = \Omega - (N \cup K \cup L)$. Clearly $P(H) = Q(H) = 1$. Also $H \in \mathcal{A}$. Thus H satisfies (a).

Take now $\omega \in H$. Since $\omega \notin N$, we conclude that h is integrable w.r.t. $P^*(\omega, \cdot)$ as well as w.r.t. $Q^*(\omega, \cdot)$. Thus H satisfies (b).

Let $\omega \in H$. In case $\omega \in F \cap G$, then $\omega \notin K \cup L$ so that $h^*(\omega) = P^*(\omega, h) = Q^*(\omega, h)$. On the other hand, if $\omega \in F^c \cup G^c$ then $P^*(\omega, \cdot)$ and $Q^*(\omega, \cdot)$ are same and hence so are $P^*(\omega, h)$ and $Q^*(\omega, h)$. Thus H satisfies the condition (c) also. \square

Let us now further assume that $\Omega = R^2$ and \mathcal{B} is its Borel sigma-field.

Then the following is well-known.

(9) Regular conditional probability $\mu^*(\omega, B)$ on \mathcal{B} given \mathcal{A} under μ exists.

Thus we have at our disposal Propositions 4–8 above.

Let us now further assume that $P = P_1 \times P_2$ and $Q = Q_1 \times Q_2$ and $\mu = \mu_1 \times \mu_2$ are product probabilities where P_i, Q_i, μ_i are probabilities on R . Assume $dP_i = f_i d\mu_i$ and $dQ_i = g_i d\mu_i$ for $i = 1, 2$.

Observe that

$$\begin{aligned} dP &= f d\mu; & f(\omega_1, \omega_2) &= f_1(\omega_1) f_2(\omega_2). \\ dQ &= g d\mu; & g(\omega_1, \omega_2) &= g_1(\omega_1) g_2(\omega_2). \end{aligned}$$

Then we have the following.

(10) If \mathcal{A} is marginally sufficient for $\{P, Q\}$ then it is sufficient for $\{P, Q\}$.

Proof. Set $h_i = \log(f_i/g_i)$ for $i = 1, 2$. Even though h_i is a function of only one variable, we regard it as function on the product space not depending on the other variable. Further, we will also use the fact that any function of one variable which is integrable on that coordinate space, when treated as a function of two variables, is integrable on the product space with product probability and value of the integral remains the same.

Use (8) and get H such that $P(H) = Q(H) = 1$ and

$$\omega \in H \Rightarrow P^*(\omega, h_i^\pm) = Q^*(\omega, h_i^\pm) \text{ for } i = 1, 2.$$

Of course (8) only says for integrable functions. In our case, using marginal sufficiency and considering $h_i^\pm \wedge n$ we can get a set H_n as in (8), and then take $H = \cap H_n$ to see (8(a)) holds. Further, (8(c)) too holds by an application of monotone convergence theorem. Only point is that nothing about finiteness of the common value in (8(c)) can be asserted at this stage — which was the content of (8(b)) when we assumed h there was integrable.

From (1), we see h_i^+ is Q_i -integrable and h_i^- is P_i -integrable. Thus

$$Q^*(\omega, h_i^+) < \infty \text{ a.e. } [Q], \quad P^*(\omega, h_i^-) < \infty \text{ a.e. } [P].$$

Set

$$I = \{\omega : Q^*(\omega, h_i^+) = \infty \text{ for some } i\} \cap F \cap G,$$

$$J = \{\omega : P^*(\omega, h_i^-) = \infty \text{ for some } i\} \cap F \cap G.$$

Then $Q(I) = 0 = P(J)$. But both are subsets of $F \cap G$ and are in \mathcal{A} , so using (7) we get $P(N) = 0 = Q(N)$, where $N = I \cup J$. Also note $N \in \mathcal{A}$.

Now take $\Omega^* = H - N$. Since H as well as N are in \mathcal{A} , we note that $\Omega^* \in \mathcal{A}$. Since $P(H) = 1 = Q(H)$, we conclude that $P(\Omega^*) = Q(\Omega^*) = 1$. We now claim that if $\omega \in \Omega^*$ then the two probabilities $P^*(\omega, \cdot)$ and $Q^*(\omega, \cdot)$ are same. By definition of P^* and Q^* , this is true for points in $F^c \cup G^c$. So we only need to consider the case $\omega \in \Omega^* \cap F \cap G$.

Since $\omega \in H$, $\int h_i^\pm(\eta)P^*(\omega, d\eta) = \int h_i^\pm(\eta)Q^*(\omega, d\eta)$. Since $\omega \in F \cap G$,

$$dP^*(\omega, \cdot) = f d\mu^*(\omega, \cdot)/a; \quad a = \mu^*(\omega, f),$$

$$dQ^*(\omega, \cdot) = g d\mu^*(\omega, \cdot)/b; \quad b = \mu^*(\omega, g).$$

Now

$$\int \log \frac{f/a}{g/b}(\eta)P^*(\omega, d\eta) = \int \log \frac{f_1}{g_1} dP^*(\omega, \cdot)$$

$$+ \int \log \frac{f_2}{g_2} dP^*(\omega, \cdot) - [\log a - \log b]$$

(because of finiteness of the integrals involved)

$$= \int \log \frac{f_1}{g_1} dQ^*(\omega, \cdot) + \int \log \frac{f_2}{g_2} dQ^*(\omega, \cdot) - [\log a - \log b]$$

(because $\omega \in \Omega^*$)

$$= \int \log \frac{f/a}{g/b}(\eta) dQ^*(\omega, d\eta).$$

Now (3) completes the proof. □

We shall now prove the theorem.

Proof. We shall consider R^2 , the same proof works for any n . Since the family of probabilities is dominated, by Halmos–Savage theorem [2], it suffices to show that \mathcal{A} is pairwise sufficient. Given two product probabilities $P_1 \times P_2$ and $Q_1 \times Q_2$, we take $\mu_i = \frac{P_i + Q_i}{2}$ and $\mu = \mu_1 \times \mu_2$. Since we have product family, (10) above completes the proof. \square

4. Remarks

Remark 1. The special case of Bahadur alluded to earlier is when all probabilities are equivalent so that null sets or infinities do not create nuisance.

Remark 2. That the family is product family is important. Consider R^2 and the two probabilities: P is uniform distribution (Lebesgue measure) on the unit square $[0, 1] \times [0, 1]$ and Q is the uniform distribution (one-dimensional Lebesgue measure) on the diagonal $\{(x, x) : 0 \leq x \leq 1\}$. Let \mathcal{A} be the trivial sigma-field $\{\emptyset, R^2\}$.

Then \mathcal{A} is marginally sufficient for $\{P, Q\}$ but not sufficient.

Remark 3. That the family is dominated is also important. Though we do not have an example in the Euclidean set up, it is possible to construct an example by taking point masses on a large set $\Omega \times \Omega$ with \mathcal{B} as the power set of $\Omega \times \Omega$ and \mathcal{A} as the product sigma field with power set on each coordinate.

Remark 4. Though it is a theorem about sufficiency, it is purely probabilistic and information theoretic in nature.

Remark 5. It is interesting to note that several years ago Maitra [4] used the fact that invariant sigma field is sufficient for the family of invariant probabilities to express them as mixtures of ergodic measures. We do not know if the present theorem is relevant in that context.

Remark 6. There is an obvious tensor product interpretation of the result which makes it amenable for formulation in operator theory or free probability theory. We do not know if the above result is true and interesting in that context.

References

- [1] Ghosh J K, Abstract No. 59, *Sankhya A* **31**(1) (1969) 91
- [2] Halmos P R and Savage L J, Application of the Radon–Nikodym theorem to the theory of sufficient statistics, *Ann. Math. Statist.* **20**(2) (1949) 225–241
- [3] Kudo H, On marginal sufficiency *Statistics and Decisions* **4** (1986) 301–320
- [4] Maitra A, Integral representations of invariant measures, *Trans. Amer. Math. Soc.* **229** (1977) 209–225
- [5] Sudakov V N, The marginal sufficiency of statistics, *Zap. Nauchn. Sem. LOMI* **29** (1972) 92–101, Engl. Transl.: *J. Soviet Math.* **3** (1975) 792–800