Chapter 3
Sufficiency

Anirban DasGupta

It was the Fall of 1978. I had just finished my masters in statistics and started out as a PhD student in the stat-math division at the ISI in Calcutta. Teachers of the calibre of B.V. Rao and Ashok Maitra had taught me an enormous amount of mathematics and probability theory. But deep inside me I was curious to learn much more of statistical theory. Unfortunately, Basu had already left and moved to the US, and C.R. Rao was rarely seen in the Calcutta center. I considered following Basu to Tallahassee, but my friend Rao Chaganty warned me that the weather was so outlandishly good that I would probably never graduate. My other favorite teacher T. Krishnan was primarily interested in applied statistics, and J.K. Ghosh had only just returned from his visit to Pittsburgh. I remember being given a problem on admissibility; but, alas, that too turned out to be a modest extension of Karlin [30].

ISI allowed its students an unlimited amount of laziness and vagrancy, and I exploited this executive nonchalance gratuitously. I was not doing anything that I wanted to admit. Stat-Math was then located in an unpretentious, dark old building across from the central pond in the main campus. One day I was intrigued to see a new face; a visitor from Australia, someone whispered. In a week or so, the office sent out an announcement of a course on sufficiency by our visitor; the name was Terence P. Speed. That is how I first met Terry 34 years ago, and became one of his early students. Much later, I came to know that he was professionally and personally close to Basu, who had an enduring influence on my life. Together, Terry and Basu prepared a comprehensive bibliography of sufficiency [8]. They had intended to write a book, but communication at great distances was not such a breeze 40 years ago, and the book never came into being. Most recently, Terry and I worked together on summarizing Basu's work for the Selected Works series of Springer. I am deeply honored and touched to be asked to write this commentary on Terry's contributions to statistics, and particularly to sufficiency. Terry has worked on such an incredible variety of areas and problems that I will limit myself to just a few of his contributions that have directly influenced my own work and education. Sufficiency is certainly

A. DasGupta
Department of Statistics, Purdue University and Indian Statistical Institute
e-mail: dasgupta@stat.purdue.edu

S. Dudoit (ed.), *Selected Works of Terry Speed*, Selected Works in Probability and Statistics, 71
DOI 10.1007/978-1-4614-1347-9_3, © Springer Science+Business Media, LLC 2012

one of them. My perspective and emphasis will be rather different from other survey articles on it, such as Yamada and Morimoto [51].

For someone who does not believe in a probability model, sufficiency is of no use. It is also of only limited use in the robustness doctrine. I think, however, that the importance of sufficiency in inference must be evaluated in the context of the time. The idea of data summarization in the form of a low dimensional statistic without losing information must have been intrinsically attractive and also immensely useful when Fisher first formulated it [23]. In addition, we now know the various critical links of sufficiency to both the foundations of statistics, and to the elegant and structured theory of optimal procedures in inference.

For example, the links to the (weak and the strong) likelihood principle and conditionality principle are variously summarized in the engaging presentations in Barnard [3], Basu [6], Berger and Wolpert [10], Birnbaum [14], Fraser [26], and Savage [42]. And we are also all aware of such pillars of the mathematical theory of optimality, the Rao-Blackwell and the Lehmann-Scheffé theorem [12, 35], which are inseparably connected to sufficient statistics. At the least, sufficiency has acted as a nucleus around which an enormous amount of later development of ideas, techniques, and results have occurred. Some immediate examples are the theory of ancillarity, monotone likelihood ratio, exponential families, invariance, and asymptotic equivalence [5, 17, 18, 22, 33, 36, 38]. Interesting work relating sparse order statistics (e.g., a small fraction of the largest ones) to approximate sufficiency is done in Reiss [40], and approximate sufficiency and approximate ancillarity are given a direct definition, with consequences, in DasGupta [20]. We also have the coincidence that exact and nonasymptotic distributional and optimality calculations can be done precisely in those cases where a nontrivial sufficient statistic exists. The fundamental nature of the idea of sufficiency thus cannot be minimized; not yet.

Collectively, Kolmogorov, Neyman, Bahadur, Dynkin, Halmos, and Savage, among many other key architects, put sufficiency on the rigorous mathematical pedal. If $\{P, P \in \mathscr{P}\}$ is a family of probability measures on a measurable space $(\Omega, \mathscr{A})$, a sub σ-field $\mathscr{B}$ of $\mathscr{A}$ is *sufficient* if for each measurable set $A \in \mathscr{A}$, there is a (single) $\mathscr{B}$ measurable function g_A such that $g_A = E_P(I_A|\mathscr{B})$, a.e. $(P) \forall P \in \mathscr{P}$. This is rephrased in terms of a *sufficient statistic* by saying that if $T : (\Omega, \mathscr{A}) \longrightarrow (\Omega', \mathscr{A}')$ is a mapping from the original (measurable) space to another space, then T is a sufficient statistic if $\mathscr{B} = \mathscr{B}_T = T^{-1}(\mathscr{A}')$ is a sufficient sub σ-field of $\mathscr{A}$. In a classroom situation, the family $\mathscr{P}$ is often parametrized by a finite dimensional parameter θ, and we describe sufficiency as the conditional distribution of any other statistic given the sufficient statistic being independent of the underlying parameter θ. Existence of a fixed dimensional sufficient statistic for all sample sizes is a rare phenomenon for regular families of distributions, and is limited to the multiparameter exponential family (Barankin and Maitra [2], Brown [16]; it is also mentioned in Lehmann [34]). Existence of a fixed dimensional sufficient statistic in location-scale families has some charming (and perhaps unexpected) connections to the Cauchy-Deny functional equation [29, 32, 39].

Sufficiency corresponds to summarization without loss of information, and so the maximum such possible summarization is of obvious interest. A specific sub

σ-field $\mathscr{B}^*$ is a *minimal sufficient* sub σ-field if for any other sufficient sub σ-field $\mathscr{B}$, we have the inclusion that $\mathscr{B}^* \vee \mathscr{N}_{\mathscr{P}} \subseteq \mathscr{B} \vee \mathscr{N}_{\mathscr{P}}$, where $\mathscr{N}_{\mathscr{P}}$ is the family of all $\mathscr{P}$-null members of $\mathscr{A}$. In terms of statistics, a specific sufficient statistic T^* is minimal sufficient if given any other sufficient statistic T, we can write T^* as $T^* = h \circ T$ a.e. $\mathscr{P}$, i.e., a minimal sufficient statistic is a function of every sufficient statistic. A sufficient statistic that is also *boundedly complete* is minimal sufficient.

This fact does place completeness as a natural player on the scene rather than as a mere analytical necessity; of course, another well known case is Basu's theorem [4]. The converse is not necessarily true; that is, a minimal sufficient statistic need not be boundedly complete. The location parameter t densities provide a counterexample, where the vector of order statistics is minimal sufficient, but clearly not boundedly complete. It is true, however, that in somewhat larger families of densities, the vector of order statistics is complete, and hence boundedly complete [9]. If we think of a statistic as a partition of the sample space, then the partitions corresponding to a minimal sufficient statistic T^* can be constructed by the rule that $T^*(x) = T^*(y)$ if and only if the likelihood ratio $\frac{f_\theta(x)}{f_\theta(y)}$ is independent of θ. Note that this rule applies only to the dominated case, with $f_\theta(x)$ being the density (Radon-Nikodym derivative) of P_θ with respect to the relevant dominating measure.

Halmos and Savage [28] gave the *factorization theorem* for characterizing a sufficient sub σ-field, which says that if each $P \in \mathscr{P}$ is assumed to be absolutely continuous with respect to some P_0 (which we may pick to be in the convex hull of $\mathscr{P}$), then a given sub σ-field $\mathscr{B}$ is sufficient if and only if for each $P \in \mathscr{P}$, we can find a $\mathscr{B}$ measurable function g_P such that the identity $dP = g_P dP_0$ holds. Note that we insist on g_P being $\mathscr{B}$ measurable, rather than being simply $\mathscr{A}$ measurable (which would be no restriction, and would not serve the purpose of data summarization). Once again, in a classroom situation, we often describe this as T being sufficient if and only if we can write the joint density $f_\theta(x)$ as $f_\theta(x) = g_\theta(T(x))p_0(x)$ for some g and p_0. The factorization theorem took the guessing game out of the picture in the dominated case, and is justifiably regarded as a landmark advance. I will shortly come to Terry Speed's contribution on the factorization theorem.

Sufficiency comes in many colors, which turn out to be equivalent under special sets of conditions (e.g. Roy and Ramamoorthi [41]). I will loosely describe a few of these notions. We have *Blackwell sufficiency* [15] which corresponds to sufficiency of an experiment as defined via comparison of experiments [48, 50], *Bayes sufficiency* which corresponds to the posterior measure under any given prior depending on the data x only through $T(x)$, and *prediction sufficiency* (also sometimes called *adequacy*) which legislates that to predict an unobserved Y defined on some space $(\Omega'', \mathscr{A}'')$ on the basis of an observed X defined on $(\Omega, \mathscr{A})$, it should be enough to only consider predictors based on $T(X)$. See, for example, Takeuchi and Akahira [49], and also the earlier articles Bahadur [1] and Skibinsky [44]. I would warn the reader that the exact meaning of prediction sufficiency is linked to the exact assumptions on the prediction loss function. Likewise, Bayes sufficiency need not be equivalent to ordinary sufficiency unless $(\Omega, \mathscr{A})$ is a standard Borel space, i.e., unless $\mathscr{A}$ coincides with the Borel σ-field corresponding to some compact metrizable topology on Ω.

Consider now the enlarged class of probability distributions defined as $P_C(A) = P(X \in A\,|\,Y \in C), P \in \mathscr{P}, C \in \mathscr{A}''$. Bahadur leads us to the conclusion that prediction sufficiency is equivalent to sufficiency in this enlarged family of probability measures. A major result due to Terry Speed is the derivation of a factorization theorem for characterizing a prediction sufficient statistic in the dominated case [45]. A simply stated but illuminating example in Section 6 of Speed's article shows why the particular version of the factorization theorem he gives can be important in applications. As far as I know, a theory of partial adequacy, akin to partial sufficiency [7, 25, 27], has never been worked out. However, I am not sure how welcome it will now be, considering the diminishing importance of probability and models in prevalent applied statistics.

Two other deep and delightful papers of Terry that I am familiar with are his splendidly original paper on spike train deconvolution [37], and his paper on Gaussian distributions over finite simple graphs [47]. These two papers are precursors to what we nowadays call independent component analysis and graphical models. Particularly, the spike train deconvolution paper leads us to good problems in need of solution. However, I will refrain from making additional comments on it in order to spend some time on a most recent writing of Terry that directly influenced me.

In his editorial column in the *IMS Bulletin* [46], Terry describes the troublesome scenario of irreconcilable quantitative values obtained in bioassays conducted under different physical conditions at different laboratories (actually, he describes, specifically, the example of reporting the expression level of the HER2 protein in breast cancer patients). He cites an earlier classic paper of Youden [52], which I was not previously familiar with. Youden informally showed the tendency of a point estimate derived from one experiment to fall outside of the error bounds reported by another experiment. In Youden's cases, this was usually caused by an unmodelled latent bias, and once the bias was taken care of, the conundrum mostly disappeared.

Inspired by Terry's column, I did some work on reconcilability of confidence intervals found from different experiments, even if there are no unmodelled biases. What I found rather surprised me. Theoretical calculations led to the conclusion that in as few as 10 experiments, it could be quite likely that the confidence intervals would be nonoverlapping. In meta-analytic studies, particularly in clinical trial contexts, the number of experiments combined is frequently 20, 25, or more. This leads to the apparently important question: how does one combine independent confidence intervals when they are incompatible? We have had some of our best minds think about related problems; for example, Fisher [24], Birnbaum [13], Koziol and Perlman [31], Berk and Cohen [11], Cohen et al. [19], and Singh et al. [43]. Holger Dette and I recently collaborated on this problem and derived some exact results and some asymptotic theory involving extremes [21]. It was an exciting question for us, caused by a direct influence of Terry.

Human life is a grand collage of countless events and emotions, triumphs and defeats, love and hurt, joy and sadness, the extraordinary and the mundane. I have seen life from both sides now, tears and fears and feeling proud, dreams and schemes and circus crowds. But it is still my life's illusion of those wonderful years in the seventies that I recall fondly in my life's journey. Terry symbolizes that fantasy and uncomplicated part of my life. I am grateful to have had this opportunity to write a few lines about Terry; *prendre soin*, Terry, my teacher and my friend.

References

[1] R. R. Bahadur. Sufficiency and statistical decision functions. *Ann. Math. Statist.*, 25:423–462, 1954.

[2] E. W. Barankin and A. P. Maitra. Generalization of the Fisher-Darmois-Koopman-Pitman theorem on sufficient statistics. *Sankhyā Ser. A*, 25:217–244, 1963.

[3] G. A. Barnard. Comments on Stein's "A remark on the likelihood principle". *J. Roy. Stat. Soc. A*, 125:569–573, 1962.

[4] D. Basu. On statistics independent of a complete sufficient statistic. *Sankhyā*, 15:377–380, 1955.

[5] D. Basu. The family of ancillary statistics. *Sankhyā*, 21:247–256, 1959.

[6] D. Basu. Statistical information and likelihood. *Sankhyā Ser. A*, 37(1):1–71, 1975. Discussion and correspondence between Barnard and Basu.

[7] D. Basu. On partial sufficiency: A review. *J. Statist. Plann. Inference*, 2(1): 1–13, 1978.

[8] D. Basu and T. P. Speed. Bibliography of sufficiency. Mimeographed Technical report, Manchester, 1975.

[9] C. B. Bell, D. Blackwell, and L. Breiman. On the completeness of order statistics. *Ann. Math. Statist.*, 31:794–797, 1960.

[10] J. O. Berger and R. L. Wolpert. *The Likelihood Principle*, volume 6 of *Lecture Notes – Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 2nd edition, 1988.

[11] R. H. Berk and A. Cohen. Asymptotically optimal methods of combining tests. *J. Am. Stat. Assoc.*, 74(368):812–814, 1979.

[12] P. J. Bickel and K. A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*, volume I of *Holden-Day Series in Probability and Statistics*. Holden-Day, Inc., San Francisco, CA, 1977.

[13] A. Birnbaum. Combining independent tests of significance. *J. Am. Stat. Assoc.*, 49:559–574, 1954.

[14] A. Birnbaum. On the foundations of statistical inference. *J. Am. Stat. Assoc.*, 57:269–326, 1962.

[15] D. Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 93–102, Berkeley and Los Angeles, 1951. University of California Press.

[16] L. Brown. Sufficient statistics in the case of independent random variables. *Ann. Math. Statist.*, 35:1456–1474, 1964.

[17] L. D. Brown. *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, volume 9 of *Lecture Notes – Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1986.

[18] L. D. Brown and M. G. Low. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Stat.*, 24(6):2384–2398, 1996.
[19] A. Cohen, J. I. Marden, and K. Singh. Second order asymptotic and nonasymptotic optimality properties of combined tests. *J. Statist. Plann. Inference*, 6(3): 253–276, 1982.
[20] A. DasGupta. Extensions to Basu's theorem, factorizations, and infinite divisibility. *J. Statist. Plann. Inference*, 137:945–952, 2007.
[21] A. DasGupta and H. Dette. On the reconcilability and combination of independent confidence intervals. Preprint, 2011.
[22] P. Dawid. Basu on ancillarity. In *Selected Works of Debabrata Basu*, Selected Works in Probability and Statistics, pages 5–8. Springer, 2011.
[23] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. Lond. A*, 222:309–368, 1922.
[24] R. A. Fisher. *Statistical Methods for Research Workers*. Oliver & Boyd, 14th, revised edition, 1970.
[25] D. A. S. Fraser. Sufficient statistics with nuisance parameters. *Ann. Math. Statist.*, 27:838–842, 1956.
[26] D. A. S. Fraser. *The Structure of Inference*. John Wiley & Sons Inc., New York, 1968.
[27] J. Hájek. On basic concepts of statistics. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics*, pages 139–162, Berkeley, CA, 1967. University of California Press.
[28] P. R. Halmos and L. J. Savage. Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.*, 20:225–241, 1949.
[29] V. S. Huzurbazar. *Sufficient Statistics: Selected Contributions*. Marcel Dekker, New York, NY, 1976.
[30] S. Karlin. Admissibility for estimation with quadratic loss. *Ann. Math. Statist.*, 29:406–436, 1958.
[31] J. A. Koziol and M. D. Perlman. Combining independent chi-squared tests. *J. Am. Stat. Assoc.*, 73(364):753–763, 1978.
[32] K.-S. Lau and C. R. Rao. Integrated Cauchy functional equation and characterizations of the exponential law. *Sankhyā Ser. A*, 44(1):72–90, 1982.
[33] L. Le Cam. Sufficiency and approximate sufficiency. *Ann. Math. Statist.*, 35: 1419–1455, 1964.
[34] E. L. Lehmann. *Testing Statistical Hypotheses*. John Wiley & Sons Inc., New York, 1959.
[35] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998.
[36] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, New York, third edition, 2005.
[37] L. Li and T. P. Speed. Parametric deconvolution of positive spike trains. *Ann. Stat.*, 28(5):1279–1301, 2000.
[38] M. Nussbaum. Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Stat.*, 24(6):2399–2430, 1996.

[39] C. R. Rao and D. N. Shanbhag. Recent results on characterization of probability distributions: A unified approach through extensions of Deny's theorem. *Adv. Appl. Prob.*, 18(3):660–678, 1986.
[40] R. D. Reiss. A new proof of the approximate sufficiency of sparse order statistics. *Stat. Prob. Lett.*, 4:233–235, 1986.
[41] K. K. Roy and R. V. Ramamoorthi. Relationship between Bayes, classical and decision theoretic sufficiency. *Sankhyā Ser. A*, 41(1-2):48–58, 1979.
[42] L. J. Savage. The foundations of statistics reconsidered. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Vol. I*, pages 575–586, Berkeley, CA, 1961. University of California Press.
[43] K. Singh, M. Xie, and W. E. Strawderman. Combining information from independent sources through confidence distributions. *Ann. Stat.*, 33(1):159–183, 2005.
[44] M. Skibinsky. Adequate subfields and sufficiency. *Ann. Math. Statist.*, 38: 155–161, 1967.
[45] T. P. Speed. A factorisation theorem for adequate statistics. *Aust. J. Stat.*, 20 (3):240–249, 1978.
[46] T. P. Speed. Enduring values. *IMS Bulletin*, 39(10):10, 2010.
[47] T. P. Speed and H. T. Kiiveri. Gaussian Markov distributions over finite graphs. *Ann. Stat.*, 14(1):138–150, 1986.
[48] C. Stein. Notes on the comparison of experiments. Technical report, University of Chicago, 1951.
[49] K. Takeuchi and M. Akahira. Characterizations of prediction sufficiency (adequacy) in terms of risk functions. *Ann. Stat.*, 3(4):1018–1024, 1975.
[50] E. Torgersen. *Comparison of Statistical Experiments*, volume 36 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1991.
[51] S. Yamada and H. Morimoto. Sufficiency. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, volume 17 of *Lecture Notes – Monograph Series*, pages 86–98. Institute of Mathematical Statistics, Hayward, CA, 1992.
[52] W. J. Youden. Enduring values. *Technometrics*, 14(1):1–11, 1972.

Sankhyā : The Indian Journal of Statistics
1976, Volume **38**, Series A, Pt. 2, pp. 194–196.

A NOTE ON PAIRWISE SUFFICIENCY AND COMPLETIONS

By T. P. SPEED
University of Western Australia, Nedlands

SUMMARY. The result of Halmos and Savage that pairwise sufficiency implies sufficiency for dominated families of measures is obtained here by a simple technique. We also use this technique to obtain an easy proof of the equality between the two notions of completion with respect to a dominated family of probability measures for sub-σ-fields in a measure space. The proofs have natural generalisations to the so called campact case.

1. Introduction

In their proof that pairwise sufficiency implies sufficiency for dominated families of measures Halmos and Savage (1949) use four technical Lemmas. One of the purposes of the present note is to describe a simple method of obtaining this result directly. We find that the technique used also provides a simpler method of deriving a result of Le Bihan, Littaye-Petit and Petit (1970) asserting the equality between two notions of completion with respect to a family of probability measures for sub-σ-fields in a measure space. We close the note with an indication of how simple generalisations of these two results to the so-called compact case can be obtained.

2. Pairwise sufficiency implies sufficiency in the dominated case

Let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$ in a measure space $(\mathcal{X}, \mathcal{A})$ and suppose that $\mathcal{B}$ is pairwise sufficient for a countable family $\mathcal{P}_0$ of probability measures on $\mathcal{A}$; that is, for any bounded $\mathcal{A}$ measurable function f on $\mathcal{X}$ and a pair $\{P, Q\} \subseteq \mathcal{P}_0$, there exists a $\mathcal{B}$-measurable function $f_{P,Q}$ on $\mathcal{X}$ such that

$$f_{P,Q} = E_P^{\mathcal{B}} f \qquad \text{a.s. } P$$

and

$$f_{P,Q} = E_Q^{\mathcal{B}} f \qquad \text{a.s. } Q.$$

Put $f^* = \bigvee_P \bigwedge_Q f_{P,Q}$ the sup and inf extending over $\mathcal{P}_0$, and observe that for every $P \epsilon \mathcal{P}_0$ we have

$$\bigwedge_Q f_{P,Q} \leqslant f^* \leqslant \bigvee_Q f_{Q,P}.$$

We can now see that for every $P \epsilon \mathcal{P}_0$

$$f^* = E_P^{\mathcal{B}} f \qquad \text{a.s. } P \qquad \dots \text{ (1)}$$

and the proof that $\mathcal{B}$ is sufficient for $\mathcal{P}_0$ on $\mathcal{A}$ is completed. For a dominated family $\mathcal{P}$ of probability measures on $\mathcal{A}$ take a countable equivalent subset $\mathcal{P}_0$ of $\mathcal{P}$, see

Halmos and Savage (1949) for this fact, and argue as above obtaining one f^* such that (1) holds. Now take $Q \in \mathscr{P}\setminus\mathscr{P}_0$ and consider $\mathscr{P}_0 \cup \{Q\}$. We can produce an f^{**} such that (1) holds with f^* replaced by f^{**}, and $f^{**} = E_Q^{\mathscr{B}} f$ a.s. Q. But then we see that $f^* = f^{**}$ a.s. P for all $P \in \mathscr{P}_0$, and hence $f^* = f^{**}$ a.s. Q; i.e. $f^* = E_Q^{\mathscr{B}} f$ a.s. Q.

3. Pairwise and strong completions of a sub-σ-field

In the notation used above let us define the *pairwise completion* of $\mathscr{B}$ with respect to $\mathscr{P}$ on $\mathscr{A}$, in symbols $\hat{\mathscr{B}}^{[\mathscr{P},\mathscr{A}]}$ or just $\hat{\mathscr{B}}$ where no confusion can result, to be all elements $A \in \mathscr{A}$ with the following property : for each pair $\{P, Q\} \subseteq \mathscr{P}$ there exists $B = B_{P,Q} \in \mathscr{B}$ such that $P(A\Delta B) = Q(A\Delta B) = 0$. We can also define the *strong completion* $\bar{\mathscr{B}} = \bar{\mathscr{B}}^{[\mathscr{P},\mathscr{A}]}$ of $\mathscr{B}$ with respect to $\mathscr{P}$ on $\mathscr{A}$ to be all elements $A \in \mathscr{A}$ for which there exists $B \in \mathscr{B}$ such that $P(A\Delta B) = 0$ for every $P \in \mathscr{P}$. Using analytic sets it was proved in Marie-Francoise Le Bihan *et al* (1970) that for a dominated family $\mathscr{P}$ we have $\bar{\mathscr{B}} = \hat{\mathscr{B}}$. Let us see that this result follows easily using the argument of Section 2. If $A \in \hat{\mathscr{B}}$, then for each pair $\{P, Q\} \subseteq \mathscr{P}_0$, a countable equivalent subfamily of $\mathscr{P}$, there exists $B_{P,Q} \in \mathscr{B}$ such that $P(A\Delta B_{P,Q}) = Q(A\Delta B_{P,Q}) = 0$. Put

$$B = \bigcup_P \bigcap_Q B_{P,Q},$$

the unions and intersections being taken over $\mathscr{P}_0$, and we see that $P(A\Delta B) = 0$ for all $P \in \mathscr{P}_0$, and hence for all $P \in \mathscr{P}$. This proves that $\hat{\mathscr{B}} \subseteq \bar{\mathscr{B}}$ and the reverse conclusion is immediate.

4. Compactness

Both of the above results generalise to the case of a compact family of probability measures $\mathscr{P}$ on $(\mathscr{X}, \mathscr{A})$. A family $\{f_P : P \in \mathscr{P}\}$ of bounded $\mathscr{B}$-measurable functions on $\mathscr{X}$ will be called *pairwise* [respectively finitely, countably, completely] *compatible* if for every subfamily $Q \subseteq \mathscr{P}$ consisting of two [respectively finitely, countably, arbitrarily many] measures, there exists a $\mathscr{B}$-measurable function f_Q such that for all $P \in Q$ we have $f_Q = f_P$ a.s. P. Pitcher (1965) defined the notion of compactness for families $\mathscr{P}$ of probabilities and an equivalent form of it is the following : $\mathscr{P}$ is compact on the sub-σ-field $\mathscr{B}$ of $\mathscr{A}$ if every countably compatible family $\{f_P, P \in \mathscr{P}\}$ of bounded $\mathscr{B}$-measurable functions is completely compatible.

Let us observe that a dominated family $\mathscr{P}$ of probabilities on $(\mathscr{X}, \mathscr{A})$ is compact on every sub-σ-field $\mathscr{B}$ of $\mathscr{A}$. As before we first take a countable equivalent subset $\mathscr{P}_0 \subseteq \mathscr{P}$. Any pairwise compatible family $\{f_P : P \in \mathscr{P}_0\}$ of $\mathscr{B}$-measurable functions is clearly completely compatible here : simply put $f^* = \bigvee_P \bigwedge_Q f_{P,Q}$ where $f_{P,Q}$ is the function defining the compatibility of f_P and f_Q, and the sup and inf are taken over $\mathscr{P}_0$. Now take $Q \in \mathscr{P}\setminus\mathscr{P}_0$. By considering $\mathscr{P}_0 \cup \{Q\}$ we find a

$\mathcal{B}$-measurable f^{**} such that $f^{**} = f_P$ a.s. P for all $P \in \mathcal{P}_0$, and $f^{**} = f_Q =$ a.s. Q. This first condition implies that $f^{**} = f^*$ a.s. P for all $P \in \mathcal{P}_0$ and so $f^{**} = f^* = f_Q$ a.s. Q and the proof is complete.

These remarks prove that the following result is a generalisation of those outlined in Sections 2, 3 and provides an alternative approach to them, although the details are not dissimilar.

Theorem : *Let $\mathcal{P}$ be a family of probabilities over $(\mathcal{X}, \mathcal{A})$ and $\mathcal{B}$ a sub-σ-field of $\mathcal{A}$ such that $\mathcal{P}$ is compact on $\mathcal{B}$. Then*

(i) $\overline{\mathcal{B}} = \hat{\mathcal{B}}$, *and further,*

(ii) *$\mathcal{B}$ is sufficient for $\mathcal{P}$ on $\mathcal{A}$ whenever it is pairwise sufficient for $\mathcal{P}$ on $\mathcal{A}$.*

Proof : (i) For $A \in \hat{\mathcal{B}}$ we take f_P to be the indicator of any $B \in \mathcal{B}$ for which $P(A\Delta B) = 0$; such exist and it follows from the assumption on A that $\{f_P : P \in \mathcal{P}\}$ is a pairwise compatible family of $\mathcal{B}$-measurable functions.

This easily extends to countable compatibility by using the sup-inf trick and so the compactness assumption on $\mathcal{B}$ ensures the existence of a $\mathcal{B}$-measurable f with $f = f_P$ a.s. P for each $P \in \mathcal{P}$. Clearly $B = \{f = 1\} \in \mathcal{B}$ satisfies $P(A\Delta B) = 0$, $P \in \mathcal{P}$ and the proof of (i) is complete.

(ii) For a bounded $\mathcal{A}$-measurable function f we put $f_P = E_P^{\mathcal{B}} f$. The system $\{f_P : P \in \mathcal{P}\}$ is again a pairwise compatible family of $\mathcal{B}$-measurable functions by the pairwise sufficiency assumption and again this lifts to countable compatibility using the sup-inf trick. The proof is completed by invoking the compactness of $\mathcal{P}$ on $\mathcal{B}$.

References

PAUL R. HALMOS and SAVAGE, L. J. (1949) Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.*; **20**, 225-241.

TOKITAKE KUSAMA and SAKATURO YAMADA (1972) : On compactness of the statistical structure and sufficiency. *Osaka J. Math.*, **9**, 11-18.

MARIE-FRANÇOISE LE BIHAN, MONIQUE LITTAYE-PETIT and JEAN-LUC PETIT (1970) : Exhaustivité par paire. *C. R. Acad. Sci. Paris Ser. A.*, **270**, 1753-1756.

PITCHER, T. S. (1965) : A more general property than domination for sets of probability measures. *Pacific J. Math.*, **15**, 597-611.

Paper received : July, 1975.

Austral. J. Statist., **20** (**3**), 1978, 240–249.

A FACTORISATION THEOREM FOR ADEQUATE STATISTICS[1]

T. P. Speed

Department of Mathematics, University of Western Australia

1. Introduction

In his interesting paper [3] Goro Ishii convincingly demonstrated the usefulness of the notion of adequacy in a number of situations by using this concept, together with completeness, to derive optimality properties of some known predictors. These proofs were all based upon analogues of the well known Rao–Blackwell and Lehmann–Scheffé theorems and to help recognise the adequacy of the statistics under discussion, Ishii cited an extension of the Fisher–Neyman factorisation criterion due to Sugiura and Morimoto [7]. I have not been able to consult this work.

The purpose of this paper is to state and prove a factorisation theorem of the type mentioned above. Our result is slightly more general than the one cited in [3]; we do not suppose a product structure for the underlying measure space and so in this respect our situation is more like that in the original paper by Skibinsky [5], and further, we do not suppose the dominating measure to be a probability measure. We begin by reviewing some simple facts regarding conditional expectations and then prove a factorisation theorem characterising conditional independence. A section is devoted to organising the known factorisation theorem characterising sufficiency in the dominated case, and may be of some independent interest. Finally a combination of these results gives the theorem of our title.

2. Conditional expectations

Our basic setting is a measurable space $(\mathscr{X}, \mathscr{A})$ equipped with a σ-finite measure μ; these will remain fixed throughout the paper. In this and the next section we will be considering a probability measure P on $\mathscr{A}$ which is absolutely continuous with respect to μ; let its Radon–Nikodym derivative be p and let us write this relationship as $P = p\,.\,\mu$.

For any sub-σ-field $\mathscr{C} \subseteq \mathscr{A}$ on which the restriction $\mu_{\mathscr{C}}$ of μ remains σ-finite we can define a conditional expectation operator $E_\mu^{\mathscr{C}}$

[1] Manuscript received January 6, 1977; revised November 22, 1977.

in the usual way: for any measurable non-negative or μ-integrable $f:\mathscr{X}\to\mathbb{R}$ there exists a $\mathscr{C}$-measurable function $E_\mu^{\mathscr{C}}f$ satisfying

$$\int_C E_\mu^{\mathscr{C}} f\, d\mu_{\mathscr{C}} = \int_C f\, d\mu \quad (C\in\mathscr{C}). \tag{2.1}$$

If we suppose that $\mathscr{C}$ is $[\mu, \mathscr{A}]$-complete i.e. contains all the elements of $\mathscr{A}$ of zero μ-measure, (2.1) defines (see Neveu [6] p. 1) a unique μ-equivalence class of $\mathscr{C}$-measurable functions and we will adopt the usual procedure of denoting this class or a representative of this class by $E_\mu^{\mathscr{C}}f$. The relation between $E_\mu^{\mathscr{C}}$ and the usual operator $E_P^{\mathscr{C}}$ is as follows: for any measurable non-negative f

$$E_P^{\mathscr{C}}f = E_\mu^{\mathscr{C}}fp/E_\mu^{\mathscr{C}}p \quad \text{a.s. } P \tag{2.2}$$

where $p = dP/d\mu$ satisfies $P(\{E_\mu^{\mathscr{C}}p=0\})=0$. This is not hard to prove from the definitions and can be found in Loève [4] pp. 344–345; see also Neveu [6] pp. 16–17 for a brief discussion. Another relation which we need below is the following minor modification of the result just noted: if $\mathscr{B}\supseteq\mathscr{C}$ is another sub-σ-field of $\mathscr{A}$ and f *is also* $\mathscr{B}$-measureable, then

$$E_P^{\mathscr{C}}f = E_\mu^{\mathscr{C}}fp_1/E_\mu^{\mathscr{C}}p_1 \quad \text{a.s. } P \tag{2.3}$$

where

$$p_1 = dP_{\mathscr{B}}/d\mu_{\mathscr{B}} = E_\mu^{\mathscr{B}}p.$$

3. Conditional independence

Let $\mathscr{B}_1$, $\mathscr{B}_2$ and $\mathscr{C}$ be sub-σ-fields of $\mathscr{A}$. We say that $\mathscr{B}_1$ *and* $\mathscr{B}_2$ *are conditionally P-independent given* $\mathscr{C}$ if for all non-negative $\mathscr{B}_i$-measurable functions f_i $(i=1,2)$:

$$E_P^{\mathscr{C}}f_1f_2 = E_P^{\mathscr{C}}f_1E_P^{\mathscr{C}}f_2 \quad \text{a.s. } P. \tag{3.1}$$

It is well-known that this is equivalent to: for all non-negative $\mathscr{B}_2\vee\mathscr{C}$-measurable functions f_2:

$$E_P^{\mathscr{B}_1\vee\mathscr{C}}f_2 \quad \text{is } \mathscr{C}\text{-measurable, and so} = E_P^{\mathscr{C}}f_2 \tag{3.2}$$

where we suppose that $\mathscr{C}$ is $[P, \mathscr{A}]$-complete. For a proof of essentially this assertion see Loève [4] pp. 563–564. It is also shown in Loève that if $\mathscr{B}_1$ and $\mathscr{B}_2$ are conditionally P-independent given $\mathscr{C}$ then so also are $\mathscr{B}_1\vee\mathscr{C}$ and $\mathscr{B}_2\vee\mathscr{C}$.

When $\mu_{\mathscr{C}}$ is σ-finite the operator $E_\mu^{\mathscr{C}}$ is well-defined and we can give an analogous definition of *conditional μ-independence*. It will then follow that (3.1) and (3.2) with μ instead of P are still equivalent.

Conditional P-independence of σ-fields has a formulation in terms of the factorisation of a density just as independence of σ-fields does, but I am unable to locate the following result in the literature. It is surely well known in its special case as the corollary. Let $(\mathscr{X}, \mathscr{A})$, μ, P and p be as in §2 and suppose that $\mathscr{B}_1$, $\mathscr{B}_2$ and $\mathscr{C}$ are sub-σ-fields of $\mathscr{A}$; for simplicity we will assume that $\mathscr{B}_1 \vee \mathscr{C} \vee \mathscr{B}_2 = \mathscr{A}$, and that $\mathscr{C}$ is $[\mu, \mathscr{A}]$-complete.

Proposition 1. *Let $\mu_{\mathscr{C}}$ be σ-finite and $\mathscr{B}_1$ and $\mathscr{B}_2$ be conditionally μ-independent given $\mathscr{C}$. If $\mathscr{B}_1$ and $\mathscr{B}_2$ are conditionally P-independent given $\mathscr{C}$ then we have the factorisation*

$$p = p_1 p_2 q^{-1} I_{\{q>0\}} \quad \text{a.s. } P \tag{3.3}$$

where $p_i = dP_{\mathscr{B}_i \vee \mathscr{C}}/d\mu_{\mathscr{B}_i \vee \mathscr{C}}$ $(i=1,2)$ and $q = dP_{\mathscr{C}}/d\mu_{\mathscr{C}}$. Conversely, if p can be factorised

$$p = f_1 f_2 \quad \text{a.s. } P \tag{3.4}$$

where f_i is $\mathscr{B}_i \vee \mathscr{C}$-measurable $(i=1,2)$, then $\mathscr{B}_1$ and $\mathscr{B}_2$ are conditionally P-independent given $\mathscr{C}$.

Loosely speaking the result asserts that when $\mathscr{B}_1$ and $\mathscr{B}_2$ are conditionally μ-independent given $\mathscr{C}$ and $P \ll \mu$, a necessary and sufficient condition for $\mathscr{B}_1$ and $\mathscr{B}_2$ to be conditionally P-independent given $\mathscr{C}$ is that the density p of P with respect to μ factorises into the product of a non-negative $\mathscr{B}_1 \vee \mathscr{C}$-measurable function and a non-negative $\mathscr{B}_2 \vee \mathscr{C}$-measurable function.

Proof. Suppose that $\mathscr{B}_1$ and $\mathscr{B}_2$ are conditionally P-independent given $\mathscr{C}$. We will prove (3.3) and it is here that we use the fact that $\mathscr{B}_1 \vee \mathscr{C} \vee \mathscr{B}_2 = \mathscr{A}$; trivial modifications would allow us to drop this assumption. Take $B_1 \in \mathscr{B}_1$, $C \in \mathscr{C}$ and $B_2 \in \mathscr{B}_2$. Then

$$\begin{aligned}
&\int_{B_1 C B_2} p_1 p_2 q^{-1} I_{\{q>0\}}\, d\mu \\
&= \int_{C \cap \{q>0\}} p_1 I_{B_1} p_2 I_{B_2} q^{-1}\, d\mu \\
&= \int_{C \cap \{q>0\}} E_\mu^{\mathscr{C}}(p_1 I_{B_1} p_2 I_{B_2}) q^{-1}\, d\mu_{\mathscr{C}} && \text{as } q \text{ is } \mathscr{C}\text{-measurable,} \\
&= \int_{C \cap \{q>0\}} E_\mu^{\mathscr{C}}(p_1 I_{B_1}) E_\mu^{\mathscr{C}}(p_2 I_{B_2}) q^{-1}\, d\mu_{\mathscr{C}} && \text{using cond. } \mu\text{-independence} \\
&= \int_{C \cap \{q>0\}} \frac{E_\mu^{\mathscr{C}}(p_1 I_{B_1})}{q} \frac{E_\mu^{\mathscr{C}}(p_2 I_{B_2})}{q}\, q\, d\mu_{\mathscr{C}} \\
&= \int_{C \cap \{q>0\}} E_P^{\mathscr{C}}(I_{B_1}) E_P^{\mathscr{C}}(I_{B_2})\, dP_{\mathscr{C}} \\
&\qquad\qquad \text{using 2.3 and the fact that } q = dP_{\mathscr{C}}/d\mu_{\mathscr{C}} = E_\mu^{\mathscr{C}} p, \\
&= \int_{C \cap \{q>0\}} E_P^{\mathscr{C}}(I_{B_1 B_2})\, dP_{\mathscr{C}} && \text{using cond. } P\text{-independence}
\end{aligned}$$

$$= \int_{C\cap\{q>0\}} I_{B_1B_2}\, dP$$
$$= P(C\cap\{q>0\}B_1B_2)$$
$$= P(CB_1B_2) \quad \text{since} \quad P(\{q>0\}) = 1$$
$$= \int_{B_1CB_2} p\, d\mu \quad \text{as} \quad P = p\,.\,\mu.$$

This implies the relation (3.3) as sets of the form B_1CB_2 form a π-system generating $\mathcal{B}_1 \vee \mathcal{C} \vee \mathcal{B}_2 = \mathcal{A}$.

For the converse suppose that p can be factorised as in (3.4). It is easy to check that $P(\{f_1=0\}) = P(\{f_2=0\}) = P(\{E_\mu^{\mathcal{B}_1\vee\mathcal{C}} f_1f_2 = 0\})$ follows from the fact that $P = f_1f_2\,.\,\mu$. Let f be a non-negative $\mathcal{B}_2\vee\mathcal{C}$-measurable function; then

$$E_P^{\mathcal{B}_1\vee\mathcal{C}} f = E_\mu^{\mathcal{B}_1\vee\mathcal{C}} fp / E_\mu^{\mathcal{B}_1\vee\mathcal{C}} p$$
$$= E_\mu^{\mathcal{B}_1\vee\mathcal{C}} ff_1f_2 / E_\mu^{\mathcal{B}_1\vee\mathcal{C}} f_1f_2$$
$$= f_1 E_\mu^{\mathcal{B}_1\vee\mathcal{C}} ff_2 / f_1 E_\mu^{\mathcal{B}_1\vee\mathcal{C}} f_2$$
$$= E_\mu^{\mathcal{B}_1\vee\mathcal{C}} ff_2 / E_\mu^{\mathcal{B}_1\vee\mathcal{C}} f_2$$
$$= E_\mu^{\mathcal{C}} ff_2 / E_\mu^{\mathcal{C}} f_2$$
$$= \text{a } \mathcal{C}\text{-measurable function,}$$

provided we suppose that $\mathcal{C}$ is $[\mu, \mathcal{A}]$-complete. This completes the proof.

Corollary 1. *Suppose that $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$, $\mu = \mu_1 \otimes \mu_2$ where the μ_i are σ-finite on $\mathcal{A}_i$ $(i=1,2)$, and that $\mathcal{B}_1$, $\mathcal{B}_2$ are the σ-algebras generated by the coordinate projections. Let $\mathcal{C} \subseteq \mathcal{A}_1$ be a $[\mu_1, \mathcal{A}_1]$-complete sub-σ-field of $\mathcal{A}_1$ such that μ_1 remains σ-finite when restricted to $\mathcal{C}$, and let us also denote the sub-σ-field of $\mathcal{A}$ isomorphic to $\mathcal{C}$ by $\mathcal{C}$. Then for a probability measure $P \ll \mu$ we have $\mathcal{B}_1$ and $\mathcal{B}_2$ conditionally P-independent given $\mathcal{C}$ if and only if the density $p = dP/d\mu$ can be factorised*

$$p = f_1 f_2 \quad \text{a.s. } \mu$$

where f_1 is $\mathcal{B}_1$-measurable and f_2 is $\mathcal{C}\vee\mathcal{B}_2$-measurable.

Proof. This is an immediate consequence of the proposition, for the hypotheses of the corollary imply that $\mathcal{B}_1$ and $\mathcal{B}_2$ are actually μ-independent and hence conditionally μ-independent given $\mathcal{C} \subseteq \mathcal{B}_1$.

4. Sufficiency

Let $\mathcal{P}$ be a family of probability measures on $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{C}$ be sub-σ-field of $\mathcal{A}$. We say that $\mathcal{C}$ *is sufficient for $\mathcal{P}$ on $\mathcal{B}$* if for any $B \in \mathcal{B}$

there exists a $\mathscr{C}$-measurable function φ_B such that for every $P \in \mathscr{P}$

(4.1) $$\varphi_B = E^{\mathscr{C}}_{\mathscr{P}} I_B \quad \text{a.s. } P$$

We will always suppose that $\mathscr{C}$ is $[\mathscr{P}, \mathscr{A}]$-complete i.e. contains all elements of $\mathscr{A}$ having zero P-measure for every $P \in \mathscr{P}$. In this section we may take $\mathscr{B} = \mathscr{A}$ without loss of generality, but in section 4 we will revert to the more general situation.

To understand the factorisation theorem of Fisher and Neyman let us consider the case $\mathscr{P} = \{P, Q\}$ with $P \ll Q$. We will denote dP/dQ by g and suppose $\mathscr{C} \subseteq \mathscr{A}$ to be $[Q, \mathscr{A}]$-complete.

Proposition 2. *$\mathscr{C}$ is sufficient for $\{P, Q\}$ on $\mathscr{A}$ if and only if g is $\mathscr{C}$-measurable.*

Proof. We choose an arbitrary $A \in \mathscr{A}$ and hope to prove that there exists a φ_A satisfying (4.1) if and only if g is $\mathscr{C}$-measurable. But (4.1) in this case implies that such a φ_A, if it exists, must be unique a.s. Q, and so must actually be $E^{\mathscr{C}}_Q I_A$. Thus we are really trying to prove: the necessary and sufficient condition that for all $A \in \mathscr{A} E^{\mathscr{C}}_Q I_A = E^{\mathscr{C}}_P I_A$ a.s. P, is $g = dP/dQ$ be $\mathscr{C}$-measurable.

Let us integrate $E^{\mathscr{C}}_Q I_A$ and $E^{\mathscr{C}}_P I_A$ over $C \in \mathscr{C}$ with respect to $P = g \,.\, Q$; we obtain

$$\int_C E^{\mathscr{C}}_Q I_A \, dP = \int_C E^{\mathscr{C}}_Q I_A g \, dQ = \int_C I_A E^{\mathscr{C}}_Q g \, dQ = \int_A I_C E^{\mathscr{C}}_Q g \, dQ;$$

$$\int_C E^{\mathscr{C}}_P I_A \, dP = \int_C I_A \, dP = \int_C I_A g \, dQ = \int_A I_C g \, dQ.$$

Now if $\mathscr{C}$ is sufficient for $\{P, Q\}$ on $\mathscr{A}$, equivalently, if for all $A \in \mathscr{A}$, $E^{\mathscr{C}}_Q I_A = E^{\mathscr{C}}_P I_A$ a.s. P, then the first terms in the two above equations coincide and hence so must the last. Taking $C = \mathscr{X}$ and varying $A \in \mathscr{A}$ we have proved that $E^{\mathscr{C}}_Q g = g$ a.s. Q; but $\mathscr{C}$ is $[Q, \mathscr{A}]$-complete and so g is $\mathscr{C}$-measurable.

On the other hand, if g is $\mathscr{C}$-measurable then the last terms in the above two equations, and hence the first, must coincide. By the remarks beginning the proof, this means that $\mathscr{C}$ is sufficient for $\{P, Q\}$ on $\mathscr{A}$. This completes the proof.

Now let us suppose that $\mathscr{P}$ is dominated by our σ-finite measure μ on $(\mathscr{X}, \mathscr{A})$. It is well known (see Halmos and Savage [2]) that this implies the existence of a countable subset $\mathscr{P}_0 \subseteq \mathscr{P}$ equivalent to $\mathscr{P}$—simply take (notation as in [6] p. 121) $\mathscr{P}_0$ for which

$$\sup_{P \in \mathscr{P}_0} I\{dP/d\mu > 0\} = \operatorname*{ess\,sup}_{P \in \mathscr{P}} I\{dP/d\mu > 0\}$$

—*and so a probability measure* Q in the countable convex hull of $\mathscr{P}$ equivalent to $\mathscr{P}$. Furthermore, if $\mathscr{C}$ is sufficient for $\mathscr{P}$ on $\mathscr{B}$, $\mathscr{C}$ is also sufficient for $\mathscr{P} \cup \{Q\}$ on $\mathscr{B}$: the same φ_A works in (4.1) for Q since

$Q=\sum a_nP_n(a_n\geq 0, \sum a_n=1,\ P_n\in\mathscr{P})$. Thus we have the following result assuming that $\mathscr{C}$ is $[\mathscr{P},\mathscr{A}]$-complete, and thus also $[Q,\mathscr{A}]$-complete:

Corollary 1. *$\mathscr{C}$ is sufficient for $\mathscr{P}$ on $\mathscr{A}$ if and only if for every $P\in\mathscr{P}$, $g_P=dP/dQ$ is $\mathscr{C}$-measurable.*

The most frequently used form of the factorisation theorem is that of Bahadur [1] given in the next corollary.

Corollary 2. *Let $\mathscr{P}$ be dominated by μ on $(\mathscr{X},\mathscr{A})$. Then for a $[\mathscr{P},\mathscr{A}]$-complete sub-$\sigma$-field $\mathscr{C}\subseteq\mathscr{A}$ to be sufficient for $\mathscr{P}$ on $\mathscr{A}$ it is necessary and sufficient that there exists a non-negative function h, and for each $P\in\mathscr{P}$, non-negative functions g_P defined on $\mathscr{X}$ such that*
(i) h is $\mathscr{A}$-measurables;
(ii) for all $P\in\mathscr{P}$, g_P is $\mathscr{C}$-measurable;
(iii) for all $P\in\mathscr{P}$, $P=g_Ph\,.\,\mu$ on $\mathscr{A}$.

Proof. This is an immediate consequence of the previous corollary; simply take Q as above, and $h=dQ/d\mu$.

5. Adequacy

Let $\mathscr{P}$ be a family of probability measures on $(\mathscr{X},\mathscr{A})$ and $\mathscr{B}_1$, $\mathscr{B}_2$, $\mathscr{C}$ sub-σ-fields of $\mathscr{A}$. Slightly paraphrasing Skibinsky [5] we say that *$\mathscr{C}$ is adequate for $\mathscr{B}_1$ with respect to $\mathscr{B}_2$ and $\mathscr{P}$* if (i) $\mathscr{C}$ is sufficient for $\mathscr{P}$ on $\mathscr{B}_1$, and (ii) $\mathscr{B}_1$ and $\mathscr{B}_2$ are, for all $P\in\mathscr{P}$, conditionally P-independent given $\mathscr{C}$. In Skibinsky [5] $\mathscr{C}\subseteq\mathscr{B}_1$, but although we could replace $\mathscr{B}_1$ by $\mathscr{B}_1\vee\mathscr{C}$ to achieve this, and no generality would be lost in doing so, we prefer the more symmetric situation natural to a formulation involving conditional independence. We will also suppose the $\mathscr{C}$ is $[\mathscr{P},\mathscr{A}]$-complete, and that $\mathscr{B}_1\vee\mathscr{C}\vee\mathscr{B}_2=\mathscr{A}$.

As we did with sufficiency, it is instructive to formulate the factorisation theorem in the case $\mathscr{P}=\{P,Q\}$ where $P=g\,.\,Q$ on $\mathscr{A}$.

Proposition 3a. *If $\mathscr{C}$ is adequate for $\mathscr{B}_1$ with respect to $\mathscr{B}_2$ and $\{P,Q\}$ then g is $\mathscr{C}\vee\mathscr{B}_2$-measurable.*

Proof. The conditional independence of $\mathscr{B}_1$ and $\mathscr{B}_2$ given $\mathscr{C}$ with respect to P and Q gives, using Proposition 1 with Q replacing μ

$$g=g_1g_2q^{-1}I_{\{q>0\}}\quad \text{a.s. } Q \tag{5.1}$$

where

$$g_i=dP_{\mathscr{B}_i\vee\mathscr{C}}/dQ_{\mathscr{B}_i\vee\mathscr{C}}=E_Q^{\mathscr{B}_i\vee\mathscr{C}}g\quad (i=1,2)$$

and

$$q=dP_{\mathscr{C}}/dQ_{\mathscr{C}}=E_Q^{\mathscr{C}}g.$$

But $\mathscr{C}$ is sufficient for $\{P, Q\}$ on $\mathscr{B}_1$, and hence $\mathscr{B}_1 \vee \mathscr{C}$, and so by Proposition 2

(5.2) $$g_1 = dP_{\mathscr{B}_1 \vee \mathscr{C}}/dQ_{\mathscr{B}_1 \vee \mathscr{C}} \quad \text{is } \mathscr{C}\text{-measurable.}$$

This fact combined with (5.1) implies that g is $\mathscr{C} \vee \mathscr{B}_2$-measurable.

To see that we cannot get a complete analogue of Proposition 2 by proving that the adequacy of $\mathscr{C}$ follows from the measurability of g with respect to $\mathscr{C} \vee \mathscr{B}_2$ we need only consider the case $P = Q$. For in this case $g = 1$ is certainly measurable $\mathscr{C} \vee \mathscr{B}_2$, the sufficiency part of the adequacy of $\mathscr{C}$ trivial but the conditional independence part false in general. In fact the converse requires a conditional independence assumption.

Proposition 3b. *If $\mathscr{B}_1$ and $\mathscr{B}_2$ are conditionally Q-independent given $\mathscr{C}$ and g is $\mathscr{C} \vee \mathscr{B}_2$-measurable, then $\mathscr{C}$ is adequate for $\mathscr{B}_1$ with respect to $\mathscr{B}_2$ and $\{P, Q\}$.*

Proof. The fact that $\mathscr{B}_1$ and $\mathscr{B}_2$ are conditionally P-independent given $\mathscr{C}$ is an immediate consequence of Proposition 1 now that we have assumed their conditional Q-independence given $\mathscr{C}$. To see that $\mathscr{C}$ is sufficient for $\{P, Q\}$ on $\mathscr{B}_1$ or, what is the same $\mathscr{B}_1 \vee \mathscr{C}$, we prove that $g_1 = dP_{\mathscr{B}_1 \vee \mathscr{C}}/dQ_{\mathscr{B}_1 \vee \mathscr{C}}$ is $\mathscr{C}$-measurable, and then use Proposition 2. But this is also a consequence of the assumed conditional Q-independence for, as g is $\mathscr{C} \vee \mathscr{B}_2$-measurable by hypothesis, $g_1 = E_Q^{\mathscr{B}_1 \vee \mathscr{C}} g$ is $\mathscr{C}$-measurable by (3.2).

Having considered the factorisation theorem for adequate sub-σ-fields in this very special case, the way is now clear to formulate an analogue of Bahadur's result (given as Corollary 2 to Proposition 2 above).

Theorem 1. *Let $(\mathscr{X}, \mathscr{A}, \mu)$ be a σ-finite measure space, $\mathscr{P}$ a family of probability measures on $\mathscr{A}$ dominated by μ, $\mathscr{B}_1$, $\mathscr{B}_2$, $\mathscr{C}$ sub-σ-fields of $\mathscr{A}$, such that $\mathscr{B}_1 \vee \mathscr{C} \vee \mathscr{B}_2 = \mathscr{A}$, $\mu_{\mathscr{C}}$ is σ-finite and $\mathscr{C}$ is $[\mathscr{P}, \mathscr{A}]$-complete. Suppose further that $\mathscr{B}_1$ and $\mathscr{B}_2$ are conditionally μ-independent given $\mathscr{C}$. Then a necessary and sufficient condition for $\mathscr{C}$ to be adequate for $\mathscr{B}_1$ with respect to $\mathscr{B}_2$ and $\mathscr{P}$ is that there exist a non-negative h, and for each $P \in \mathscr{P}$ non-negative functions g_P, on $\mathscr{X}$ such that*

(*i*) *h is $\mathscr{B}_1 \vee \mathscr{C}$-measurable,*

(*ii*) *for all $P \in \mathscr{P}$, g_P is $\mathscr{C} \vee \mathscr{B}_2$-measurable,*

(*iii*) *for all $P \in \mathscr{P}$, $P = g_P h \,.\, \mu$ on $\mathscr{A}$.*

Proof. Suppose that $\mathscr{C}$ is adequate for $\mathscr{B}_1$ with respect to $\mathscr{B}_2$ and $\mathscr{P}$. Then by Proposition 1 for each $P \in \mathscr{P}$ we can write

$$\frac{dP}{d\mu} = g_{P,1} g_{P,2} q_P^{-1} I_{\{q_P > 0\}} \quad \text{a.s. } \mu$$

where $g_{P,i}$ is $\mathscr{B}_i \vee \mathscr{C}$-measurable $(i = 1, 2)$ and q_P is $\mathscr{C}$-measurable.

A FACTORISATION THEOREM FOR ADEQUATE STATISTICS

Furthermore $g_{P,1} = dP_{\mathscr{B}_1 \vee \mathscr{C}}/d\mu_{\mathscr{B}_1 \vee \mathscr{C}}$. By Corollary 2 to Proposition 2 these can all be factorised

$$g_{P,1} = g'_P \,.\, h \quad \text{a.s.} \quad [P, \mathscr{B}_1 \vee \mathscr{C}]$$

where g'_P is $\mathscr{C}$-measurable and h is $\mathscr{B}_1 \vee \mathscr{C}$-measurable. Putting

$$g_P = g'_P g_{P,2} q_P^{-1} I_{\{q_P > 0\}},$$

a $\mathscr{C} \vee \mathscr{B}_2$-measurable function, and keeping the common h completes the proof of the necessity of this factorisation.

Conversely, suppose that functions h and g_P can be found satisfying (i), (ii) and (iii) of the theorem. Then by Proposition 1 $\mathscr{B}_1$ and $\mathscr{B}_2$ are, for all P, conditionally P-independent given $\mathscr{C}$. The sufficiency of $\mathscr{C}$ for $\mathscr{P}$ on $\mathscr{B}_1 \vee \mathscr{C}$ follows more or less as it did in the proof of Proposition 3b. We note that $g_{P,1} = d\mu_{\mathscr{B}_1 \vee \mathscr{C}}/d_{\mathscr{B}_1 \vee \mathscr{C}}$ can be obtained as $E_\mu^{\mathscr{B}_1 \vee \mathscr{C}}(dP/d\mu)$ and so

$$g_{P,1} = E_\mu^{\mathscr{B}_1 \vee \mathscr{C}}(g_P h) = (E_\mu^{\mathscr{B}_1 \vee \mathscr{C}} g)h = (E_\mu \, g)h$$

since h is $\mathscr{B}_1 \vee \mathscr{C}$ measurable, g is $\mathscr{C} \vee \mathscr{B}_2$-measurable, and (3.2) applies with μ instead of P. The proof is completed by invoking Corollary 2 to Proposition 2. This completes the proof of our main result.

Corollary 1. *Suppose that $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2$, $\mathscr{A} = \mathscr{A}_1 \otimes \mathscr{A}_2$, $\mu = \mu_1 \otimes \mu_2$ where the $\mathscr{B}_i$ are σ-finite on $\mathscr{A}_1$ $(i = 1, 2)$ and that $\mathscr{B}_1$, $\mathscr{B}_2$ are the σ-fields generated by the coordinate projections. Let $\mathscr{C} \subseteq \mathscr{A}_1$ be a sub-σ-field of $\mathscr{A}_1$ such that μ_1 remains σ-finite when restricted to $\mathscr{C}$ and let us also denote the sub-σ-field of $\mathscr{A}$ isomorphic to $\mathscr{C}$ by $\mathscr{C}$. Then a necessary and sufficient condition for $\mathscr{C}$ to be adequate for $\mathscr{B}_1$ with respect to $\mathscr{B}_2$ and a family $\mathscr{P}$ of probability measures dominated by μ is that there exist a non-negative function h, and for each $P \in \mathscr{P}$ non-negative functions g_P defined on $\mathscr{X}$ such that*

(i) h is $\mathscr{B}_1$-measurable,

(ii) for all $P \in \mathscr{P}$, g_P is $\mathscr{C} \vee \mathscr{B}_2$-measurable,

(iii) for all $P \in \mathscr{P}$, $P = g_P h \,.\, \mu$ on $\mathscr{A}$.

Proof. Once we observe that $\mathscr{B}_1$ and $\mathscr{B}_2$ are μ-independent, and hence conditionally μ-independent given $\mathscr{C} \subseteq \mathscr{B}_1$, there is nothing left to prove. This corollary includes the result cited as Theorem 2.2 in Ishii [3].

6. An Illustration

The examples in Ishii's paper [3] all concern cases in which the σ-fields $\mathscr{B}_1$ and $\mathscr{B}_2$ are actually P-independent for each $P \in \mathscr{P}$. Apart from a problem noted in the next section, the theorem of §5 above applies to Ishii's examples, and we now consider a simple situation in which a more general type of behaviour takes place.

Let $(X_1, X_2, \ldots, X_n, X_{n+1})$ be jointly normally distributed random variables such that for each i, $1 \le i \le n+1$, the distribution of X_1 given $X_{i-1}, X_{i-2}, \ldots, X_0$ is $N(\rho X_{i-1}, 1)$, where X_0 is given (i.e. held constant) and $|\rho|<1$. Put $\mathscr{X}_1 = \mathbb{R}^n$, $\mathscr{X}_2 = \mathbb{R}$, and

$$\mu_1(dx_1 \ldots . dx_n) = (2\pi)^{-n/2} \exp\left(-\tfrac{1}{2}\sum_1^n x_i^2\right) dx_1 \ldots dx_n,$$

$$\mu_2(dx_{n+1}) = (2\pi)^{-1/2} \exp\left(-\tfrac{1}{2}x_{n+1}^2\right) dx_{n+1}.$$

The conditions on $(X_1, \ldots, X_{n+1})$ are then seen to be equivalent to the requirement that their joint distribution P has density with respect to $\mu = \mu_1 \otimes \mu_2$ given by

$$\frac{dP}{d\mu} = \exp\left[\rho \sum_1^{n+1} x_{i-1}x_i - \tfrac{1}{2}\rho^2 \sum_1^{n+1} x_i^2\right].$$

Defining the sub-σ-field $\mathscr{C} = \sigma(\sum_1^n X_{i-1}X_i, \sum_1^n X_i^2, X_n)$ of $\mathscr{B}_1 = \sigma(X_1, \ldots, X_n)$, and putting $\mathscr{B}_2 = \sigma(X_{n+1})$, we see that the conditions of Corollary 1 in §5 above are satisfied, viewing $\mathscr{P}$ as the class of probabilities indexed by ρ, $|\rho|<1$. Thus the triple $(\sum_1^n X_{i-1}X_i, \sum_1^n X_i^2, X_n)$ is an *adequate* reduction of $(X_1, \ldots, X_n)$ as far as *estimation* of ρ and *prediction* of X_{n+1} is concerned. It is not hard to see that X_n could not be omitted from the triple, and this is, of course, intuitively obvious.

The reason why we chose to use the particular probability measure $\mu = \mu_1 \otimes \mu_2$ as the dominating measure, rather than the more natural dominating measure λ^{n+1}, $(n+1)$-dimensional Lebesgue measure, is explained below.

7. Limitations

In all of the results involving μ as a dominating measure we have had to suppose $\mu_{\mathscr{C}}$ σ-finite in order to be able to formulate conditional μ-independence given $\mathscr{C}$. This is a strong assumption and is violated in many simple examples.

Example. Let $\mathscr{X} = \mathbb{R}^3$, $\mathscr{A} = \mathscr{R}^3$ and $\mu = \lambda^3$ (Lebesgue measure) corresponding to three real random variables X_1, X_2, X_3 defined as the coordinate projections. One frequently considers a system such as

$$\mathscr{B}_1 = \sigma(X_1, X_2),\ \mathscr{B}_2 = \sigma(X_3) \text{ and } \mathscr{C} = \sigma(X_1 + X_2) \subseteq \mathscr{B}_1$$

and would certainly hope that since $\mathscr{B}_1$ and $\mathscr{B}_2$ are, in a sense, μ-independent, one would also have $\mathscr{B}_1$ and $\mathscr{B}_2$ conditionally μ-independent given $\mathscr{C}$. But μ is not σ-finite when restricted to $\mathscr{C}$, every set in $\mathscr{C}$ having zero or infinite μ-measure.

In hardly needs stating that in many examples of interest the μ will be exactly of the kind in the example, and a similar remark applies to the sub-σ-field $\mathscr{C}$, since many sufficient σ-fields are based upon sums.

Thus, although it seems very reasonable to assert that the theorem in §5 is in a sense a natural analogue of the Halmos–Savage formulation of the Fisher–Neyman theorem, it simply does not cover many examples of interest to statisticians. This also applies to the result of Sugiura and Morimoto cited in Ishii [3].

When the sample $(\mathscr{X}, \mathscr{A}, \mu)$ has a Euclidean structure it appears possible to prove a theorem which includes the examples mentioned but what seems difficult at the moment is the formulation of a result *generalising Theorem* 1 to cover all such cases. The problem is this: a satisfactory theory of conditional expectations, and hence of conditional independence, given sub-σ-fields on which the basic measure is not σ-finite, has yet to be developed. One might approach the problem via the notion of *set of σ-finiteness* Neveu [6] p. 16–17 but so far this has not been worked out. On the other hand there may be no alternative to formulating Euclidean problems in an Euclidean measure-theoretic framework cf. Tjur [8].

8. Acknowledgments

I would like to thank Professor D. Basu for his constant interest in this and related work, Professor J. F. C. Kingman for remarks which clarified some confusion in my mind, and Dr. M. L. Thornett, whose interest also contributed to this note. I am also indebted to Professor H. Morimoto for encouraging me to publish this extension of his result with M. Sugiura.

References

[1] Bahadur R. R. (1954) "Sufficiency and statistical decision functions." *Ann. Math. Statist.* 25, 423–462.

[2] Halmos, Paul R. and Savage, L. J. (1949) "Application of the Radon-Nikodym theorem to the theory of sufficient statistics." *Ann. Math. Statist.* 20, 225–241.

[3] Ishii, Goro (1969) "Optimality of unbiased predictors." *Ann. Inst. Statist. Math.* 21, 471–488.

[4] Loève, Michel (1963) *Probability Theory* Third Edition. Van Nostrand, Princeton N.J.

[5] Skibinsky, Morris (1967) "Adequate subfields and sufficiency." *Ann. Math. Statist.* 38, 155–161.

[6] Neveu, Jacques (1972) *Martingales a temps discret.* Masson et Cie, Paris.

[7] Sugiura, M. and Morimoto, H. (1969) "Factorisation theorem for adequate σ-field." (in Japanese). *Sūgaku* 21.

[8] Tjur, Tue (1974). *Conditional Probability Distributions.* Institute of Mathematical Statistics, University of Copenhagen.