Commentary on D. Basu's Papers on Sufficiency and Related Topics

T.P. Speed

It is an honour and a pleasure to be able to offer this commentary on some of D. Basu's papers on sufficiency and related topics. In the early 1970s, we thought that we might write a book together on sufficiency. We had many discussions and exchanged a fair amount of material. In particular, we prepared a bibliography on sufficiency which was reasonably comprehensive at the time (Basu and Speed (1975)). But the planned book never came to fruition. I have not worked on this topic since the late 1970s. So the experience of writing this commentary has been a pleasant walk down the memory lane. However, it also means that I may be unaware of some relevant later developments. Accordingly, I begin with an apology in advance to the readers for any such oversight or errors. *Caveat lector!*

General background

I think it is worth setting the scene for Basu's work on sufficiency. A new refugee from the then East Pakistan, he began working towards a PhD in 1950, at the Indian Statistical Institute, under the direction of C. R. Rao. We can assume that during this period, he gained a thorough grounding in the theory and philosophy of statistical inference, in particular, on the work of Fisher. Basu's doctoral dissertation was on estimation and testing in a decision theoretic framework, and to a smaller extent on some characterisation problems for normal distributions. It was not very Fisherian in style, but more mathematical. Undoubtedly, it was influenced by the work of Neyman, Pearson, Wald, notably Rao, and perhaps others. After submitting his PhD thesis (to the Calcutta University) in 1953, he went to the University of California at Berkeley as a Fulbright scholar. It was a long trip by ship. By the end of his time there, if not even before, I believe that he would have thoroughly absorbed the modern version of the Neyman-Pearson-Wald approach to inference, being defined and taught by Erich Lehmann, Charles Stein, Henry Scheffé, Lucien Le Cam, and Jerzy Neyman himself. In fact, some major papers of these statisticians were published in *Sankhyā*, then edited by P. C. Mahalanobis.

Sufficiency background

Fisher introduced sufficiency in his famous 1922 paper (Fisher (1922)) on the mathematical foundations of theoretical statistics. Stigler (1973) is a good source for more background. In this paper, Fisher decreed that if θ is the parameter of concern, and a statistic T contains the whole of the

T.P. Speed (✉)
Department of Statistics, University of California, Berkeley, CA 94720
e-mail: terry@stat.berkeley.edu

A. DasGupta (ed.), *Selected Works of Debabrata Basu*, Selected Works in Probability and Statistics,
DOI 10.1007/978-1-4419-5825-9_9,

information that the full sample supplies as to the value of θ, then for any other statistic T_2, and for any θ, the conditional distribution of T_2 given T should be the same under all θ. There and in later papers, Fisher presented several properties of sufficient statistics. In the 1930s, Neyman gave a rigorous proof of the factorisation criterion (Neyman (1935)), while Pitman, Koopman, Darmois, and Brown (Darmois (1935), Koopman (1936), Pitman (1936), Brown (1964)) independently discovered that under suitable regularity conditions, Exponential families were the only classes of distributions to possess nontrivial sufficient statistics at all sample sizes, and minimal sufficient statistics that have the same affine dimension as the parameter space to which θ belongs. Kolmogorov (1942) introduced *Bayes sufficiency*, though Basu was not to be seriously interested in it till the early 1970s. An important development on the practical implications of sufficiency that also came in the 1940s was the independent discovery by C. R. Rao and David Blackwell (Rao (1945), Blackwell (1947)) of the theorem that bears their names. Perhaps more significantly for Basu's work on sufficiency, a 1949 paper by P. R. Halmos and L. J. Savage (Halmos and Savage (1949)) elegantly placed sufficiency within the framework of measure theory, and replaced Fisher's parametric families by a more or less arbitrary family of probabilities, through the consideration of *sufficient σ-fields*. Halmos and Savage obtained their best results under the assumption of a *dominated family of probabilities*, that is, that each member of the family of probabilities possessed a density (*Radon-Nikodym derivative*) with respect to a common σ-finite measure. Their measure-theoretic approach was adopted in nearly all of Basu's papers on sufficiency.

The main themes in Basu's sufficiency papers

The Trinity. Kolmogorov introduced the famous triple $(\Omega, \mathcal{A}, P)$ (Kolmogorov (1933)). In his writings on sufficiency, Basu used the expanded triple $(\Omega, \mathcal{A}, \mathcal{P})$, where $\mathcal{P}$ is a general family of probability measures on $\mathcal{A}$. He used X or $\mathcal{X}$ instead of Ω, and called $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ *the trinity*. The precise nature of parametrization played little or no role in most of his research on sufficiency, the principal exceptions being his discussions of invariance and partial sufficiency, and some of his famous counterexamples.

Null sets. Much of our intuition in statistics is developed for the dominated case, where each of our probability measures in $\mathcal{P}$ has a density with respect to a common σ-finite dominating measure. With just a single probability measure P, we only need to take care with P-null sets. When we work with sufficiency, we need to pay attention to null sets more generally. With a family of probability measures $\mathcal{P}$, the relationships between the P-null sets for different members P of $\mathcal{P}$, and the sets which are P-null for all P in $\mathcal{P}$ (called $\mathcal{P}$-null sets), and the different completions of the underlying measure space all play a critical role. Much of Basu's work on sufficiency is marked by very careful treatment of considerations of null sets.

Completeness and other joint aspects of a statistic and the family of probabilities. In 1950, Lehmann and Scheffé published a landmark paper which highlighted the importance of the notions of *completeness* and *bounded completeness* of a family of distributions (Lehmann and Scheffé (1950)). Later, *weak completeness* came into the picture. Completeness was an essential ingredient of one of Basu's most well known contributions, namely *Basu's theorem*.

Ordering, maximality, and minimality of σ-fields. Another fundamental contribution in the 1950 paper of Lehmann and Scheffé was the introduction of the notion of *minimal sufficiency*. In the measure theoretic framework, this amounted to identification of a minimal sufficient σ-field. On several occasions, Basu studied the ordering of sub-σ-fields with various specific properties, and the questions of existence of maximal or minimal elements among them. As a rule, this was not a simple matter.

The papers (As in Author Bibliography)

Basu's Theorem (paper 15)

Basu's theorem (Basu (1955)) says that a boundedly complete sufficient statistic and any ancillary statistics are independently distributed under all θ. This is Theorem 2 in the paper. This was Basu's first paper on sufficiency, and arguably the most well known. Numerous applications of Basu's theorem, including many in probability theory, are detailed in the commentary of Anirban DasGupta and of Malay Ghosh in this volume; for earlier references on applications of Basu's theorem, see the beautiful exposition in Boos and Hughes-Oliver (1998), and also see DasGupta (2007). Theorem 1 in the paper was a converse, but not correct as stated. In a later paper, Basu gave a correct converse, which describes conditions under which a statistic which is independent of a sufficient statistic under all θ must be ancillary. This has been used in the literature on higher order asymptotics to establish approximate ancillarity of certain P-values; for example, see Lauritzen (2008). It is also worth pointing out that although we regard Basu's theorem purely as a result in statistical inference, it is also a tremendously effective tool in probabilistic calculations. Students of probability would be better equipped if they were trained in applying Basu's theorem to greatly simplify many distributional calculations.

Sufficiency and Finite Population Sampling (papers 29, 27, 26)

In some sense, we can see Basu at his best in his papers on finite population sampling. These papers have several goals, all of which he achieves neatly and eloquently. The first goal involves setting the statistical notion of sampling from a finite universe within the same mathematical framework of all other statistical models, by defining a suitable trinity $(\mathcal{X}, \mathcal{A}, \mathcal{P})$. Basu argues that it is natural to take $\mathcal{A}$ as the set of all subsets of the sample space $\mathcal{X}$, and $\mathcal{P}$ as an undominated family of discrete probabilities on $\mathcal{A}$. He then shows that a maximal sufficiency reduction is always at hand. These are also probably the papers in which he shows his Bayesian transition for the first time. One piece of evidence of this is his theorem that once the survey data from the finite universe has been obtained, inference should no longer depend on the sampling design that was actually used. Paper 29 (Basu (1970)) contains the now famous and colorful example of *Basu's elephants*. This example has led many statisticians of subsequent generations to think about the exact role and relevance of sample space based optimality criteria, such as admissibility. The example of Basu's elephants was the subject of an entire book on survey sampling (Brewer (2002)). Paper 27 also had the goal of showing that the counterexamples given by Pitcher (1957) and Burkholder (1961) concerning sufficiency in the undominated case need not discourage statisticians. Indeed, paper 27 shows what a difference an $\mathcal{A}$ makes.

Sufficiency and Invariance (papers 25, 18)

Calling upon invariance (under transformations preserving a statistical model) to select one from competing decision procedures originated in the late 1940s, though undoubtedly there were earlier instances. The approach was widely used in Erich Lehmann's classic text *Testing Statistical Hypotheses*, first published in 1959 (Lehmann (1959)). Invariance is a tool for data reduction, and so is sufficiency. Charles Stein (in some unpublished work), Burkholder (1960) (see Hall, Wijsman, and Ghosh (1965)), Hall, Wijsman, and Ghosh (1965), Berk and Bickel (1968), and Berk (1972) explored the relationship between sufficiency and invariance reductions of the sample data. These papers form the background for Basu's research in this area, which includes his characteristic search for clear and simple proofs, compelling motivation, a desire to deal very carefully with null sets, and is also of very significant expository value. Let me summarize it briefly. Suppose that $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ is a statistical trinity, and $\mathcal{G}$ a group of one-to-one bimeasurable transformations of $(\mathcal{X}, \mathcal{A})$ onto itself which are measure-preserving, i.e., $Pg^{-1} = P$ for all $P \in \mathcal{P}$ and all $g \in \mathcal{G}$. It is not hard to prove that for any $g \in \mathcal{G}$, the sub-σ-field $\mathcal{A}_g = \{A \in \mathcal{A} : g^{-1}A = A\}$ is sufficient for the triple $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ and so interest naturally turns to $\mathcal{A}(\mathcal{G}) = \cap_g \mathcal{A}_g$. The sub-$\sigma$-field $\mathcal{A}(\mathcal{G})$ is the σ-field of *$\mathcal{G}$-invariant sets*. Closely related are the sub-σ-fields of *essentially* and *almost $\mathcal{G}$-invariant* sets (see Basu (1970) for the exact definitions). With these preliminaries, Basu explores conditions under which a minimal sufficient sub-field $\mathcal{T}$ is contained in or coincides with the sub-field of almost $\mathcal{G}$-invariant sets. In the second half of the paper, Basu turns to parameter-preserving transformations, foreshadowing his work on partial sufficiency. But his focus here is firmly on normal models.

In paper 25 (Basu and Ghosh (1969)), Basu and Ghosh introduced the concept of *nontrivial weak completeness*. Nontrivial weak completeness means that there are no sets A, not $\mathcal{P}$- equivalent to the empty set or the entire space $\mathcal{X}$, such that $P(A)$ is constant in $P \in \mathcal{P}$. The principal aim of this paper was to explore families $\mathcal{P}$ which are not weakly complete. This was almost entirely restricted to translation parameter families on the real line, circle, or some other compact or locally compact group. Several interesting connections with theorems from harmonic analysis, and as was customary with Basu, a number of interesting examples were described. But no neat general results really came into light.

Partial Sufficiency (paper 39)

When we read Basu's work, it appears that Basu embraced the Bayesian approach to statistical inference because of the failure of the other approaches to deal adequately with inference concerning what he termed as sub-parameters, that is, functions of the global parameter. I think he found sufficiency compelling when inference concerning the entire parameter was the goal, despite some of the problems and paradoxes involving ancillary statistics. In this case, reduction to the likelihood function is the maximal possible reduction, which he probably found appropriate. However, when he turned to ways of carrying out inference for sub-parameters of interest, eliminating nuisance parameters not of interest, and the procession of forms of partial sufficiency, he did not find any solution that stood up to his creative scrutiny (Basu (1978)). Although there has been a lot more work on this topic since Basu's 1978 paper, I don't think that any general satisfactory solution to the problem has emerged. Today, non-Bayesians deal with nuisance parameters on a case-by-case basis, at times aided by special results, such as Barndorff-Nielsen's formula (Barndorff-Nielsen (1983)), or special tools, such as conditional, partial or profile likelihoods (Cox (1975), Severini (1994)). Bayesians would often integrate out all the nuisance parameters, perhaps with some type of a *noninformative prior* (Bernardo and Smith (1994)). Satisfactory general approaches concerning sub-parameters and elimination of nuisance parameters seem as far away today as they did when Basu wrote his probing article in 1978.

Sufficiency and Coherence (paper 46)

In the late 1960s and 1970s, several authors sought to broaden the domain of the nice results due to Halmos and Savage (1949) concerning general, pairwise, and minimal sufficiency, which they proved undeer the assumption that the family of probabilities was a dominated family. Pitcher (1965) defined the notion of *compactness* of a family of probability measures, Mussman (1972) introduced *weak domination*, Hasegawa and Perlman (1974) gave us *coherence*, while Le Cam (1964) explored related ideas within the framework of vector lattices. Siebert (1979) connects the first three, showing that they are essentially equivalent. While Siebert's publication predates Basu and Cheng (1981), it came after S. C. Cheng's 1978 Florida State PhD thesis written under Basu's direction, from which paper 46 most likely derived. This paper had an expository flavor. But it also gives a useful addendum to the converse part of Basu's theorem. Basu's original converse (Basu (1958)) was in terms of the family of probabilities being *connected*. Koehn and Thomas (1975) strengthened this theorem of Basu to lay down a necessary and sufficient condition for the converse to Basu's theorem to hold. This result of Koehn and Thomas said that non-ancillary statistics independent of a sufficient statistic (under all θ) exist if and only if the family of probabilities admit a *splitting set*. In paper 46, Basu and Cheng show that under the condition of coherence, this necessary and sufficient condition of Koehn and Thomas is exactly the same as Basu's original connectedness condition. Thus, under coherence, the two theorems of Basu precisely characterize the relationship between ancillarity and sufficiency through their independence, a very clean conclusion.

References

Barndorff-Nielsen, O. E. (1983). On a formula for the conditional distribution of the maximum likelihood estimator, Biometrika, 70, 343–365.

Basu, D. (1955). On statistics independent of a complete sufficient statistic, Sankhy$\overline{a}$, 15, 4, 377–380.

Basu, D. (1958). On statistics independent of a sufficient statistic, Sankhy$\overline{a}$, 20, 223–226.

Basu, D. and Ghosh, J. K. (1969). Invariant sets for translation parameter families of measures, Ann. Math. Statist., 40, 162–174.

Basu, D. (1970). An essay on the logical foundations of survey sampling, Foundations of Statistical Inference, 203-242, Holt, Rinehart and Winston, Toronto, Canada.

Basu, D. (1970). On sufficiency and invariance, Essays in Probability and Statistics, 61–84, UNC Press, Chapel Hill.

Basu, D. and Speed, T. P. (1975). Bibliography of sufficiency, Mimeographed Technical Report, Manchester.

Basu, D. (1978). On partial sufficiency: A review, Jour. Statist. Planning Inf., 1, 1–13.

Basu, D. and Cheng, S. C. (1981). A note on sufficiency in coherent models, Internat. J. Math. Sci., 3, 571–582.

Berk, R. and Bickel, P. J. (1968). On invariance and almost invariance, Ann. Math. Statist., 39, 1573–1576.

Berk, R. (1972). A note on sufficiency and invariance, Ann. Math. Statist., 43, 647–650.

Bernardo, J. M. and Smith, A. F. M. (1994). Bayesian Theory, Wiley, New York.

Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation, Ann. Math. Statist., 18, 105–110.

Boos, D. D. and Hughes-Oliver, J. M. (1998). Applications of Basu's theorem, Amer. Statist., 52, 218–221.

Brewer, K. (2002). Combined Survey Sampling Inference: Weighing of Basu's Elephants, Arnold, London,

Brown, L. D. (1964). Sufficient statistics in the case of independent random variables, Ann. Math. Statist., 35, 1456–1475.

Burkholder, D. (1960). The relation between sufficiency and invariance, I: theory. Invited address at the Central Regional Meeting of the Institute of Mathematical Statistics, Lafayette, Indiana.

Burkholder, D. (1961). Sufficiency in the undominated case, Ann. Math. Statist., 32, 1191–1200.

Cox, D. R. (1975). Partial likelihood, Biometrika, 62, 269–276.

Darmois, G. (1935). Sur les lois de probabilités a estimation exhaustive, C. R. Acad. Sc. Paris, 200, 1265–1266.

DasGupta, A. (2007). Extensions to Basu's theorem, factorizations, and infinite divisibility, Jour. Statist. Planning Inf., 945–952.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics, Phil. Trans. Royal Soc. (London), A222, 309–368.

Hall, W. J., Wijsman, R. J., and Ghosh, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis, Ann. Math. Statist., 36, 575–614.

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

Hasegawa, M. and Perlman, M. D. (1974). On the existence of a minimal sufficient sub-field, Ann. Statist., 2, 1049–1055.

Koehn, U. and Thomas, D. L. (1975). On statistics independent of a sufficient statistic: Basu's lemma, Amer. Statist., 29, 1, 40–42.

Kolmogorov, A. N. (1933). Foundations of the Theory of Probability Springer, Berlin.

Kolmogorov, A. N. (1942). Definition of center of dispersion and measure of accuracy from a finite number of observations, Izv. Akad. Nauk Er. Mat., 6, 3–32. (In Russian)
Koopman, B. O. (1936). On distributions admitting a sufficient statistic, Trans. Amer. Math. Soc., 39, 399–409.
Lauritzen, S. (2008). Ancillarity and Conditional Inference, Lecture Notes, Oxford University, UK.
Le Cam, L. (1964). Sufficiency and approximate sufficiency, Ann. Math. Statist., 35, 1419–1455.
Lehmann, E. L. and Scheffé, H. (1950). Completeness, similar regions, and unbiased estimation, Sankhy$\overline{a}$, 10, 305–340.
Lehmann, E. L. (1959). Testing Statistical Hypotheses, Wiley, New York.
Mussman, O. (1972). Vergleich von Experimenten im schwach dominierten Fall, Z. Wahrsch. Verw. Gebiete, 24, 295–308.
Neyman, J. (1935). Su un teorema concernente le cosiddette statistiche sufficienti, Inst. Ital. Atti Giorn, 6, 320–334.
Pitcher, T. S. (1957). Sets of measures not admitting necessary and sufficient statistics or subfields, Ann. Math. Statist., 28, 267–268.
Pitman, E. J. G. (1936). Sufficient statistics and intrinsic accuracy, Proc. Camb. Philos. Soc., 32, 567–579.
Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc., 37, 81–91.
Severini, T. A. (1994). On the approximate elimination of nuisance parameters by conditioning, Biometrika, 81, 649–661.
Siebert, E. (1979). Statistical experiments and their conical measures, Z. Wahrsch. Verw. Gebiete, 45, 247–258.
Stigler, S. (1973). Studies in the history of probability and statistics, Laplace, Fisher, and the discovery of the concept of sufficiency, Biometrika, 60, 439–445.