

Selected Works in Probability and Statistics

Selected Works of Debabrata Basu

Selected Works in Probability and Statistics

For other titles published in this series, go to
www.springer.com/series/8556

Anirban DasGupta
Editor

Selected Works of Debabrata Basu

 Springer

Editor

Anirban DasGupta
Department of Statistics and Mathematics
Purdue University
150 N. University Street
West Lafayette, IN 47907, USA
dasgupta@stat.purdue.edu

ISBN 978-1-4419-5824-2 e-ISBN 978-1-4419-5825-9
DOI 10.1007/978-1-4419-5825-9
Springer New York Dordrecht Heidelberg London

© Springer Science+Business Media, LLC 2011

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface to the Series

Springer's Selected Works in Probability and Statistics series offers scientists and scholars the opportunity of assembling and commenting upon major classical works and honors the work of distinguished scholars in probability and statistics. Each volume contains the original papers, original commentary by experts on the subject's papers, and relevant biographies and bibliographies.

Springer is committed to maintaining the volumes in the series with free access on SpringerLink, as well as to the distribution of print volumes. The full text of the volumes is available on SpringerLink with the exception of a small number of articles for which links to their original publisher is included instead. These publishers have graciously agreed to make the articles freely available on their websites. The goal is maximum dissemination of this material.

The subjects of the volumes have been selected by an editorial board consisting of Anirban DasGupta, Peter Hall, Jim Pitman, Michael Sørensen, and Jon Wellner.



Preface and Introduction

When Springer approached me with a proposal for editing a collection of Dev Basu's writings and articles, with original commentaries from experts in the field, I accepted this invitation with a sense of pride, joy, and anxiety. I was a direct student of Dev Basu at the Indian Statistical Institute, and accepted this task with a sense of apprehension. I was initially attracted to Basu because of his clarity of exposition and a dignified and inspiring presence in the classrooms at the ISI. He gave us courses on combinatorics, distribution theory, central limit theorems, and the random walk. To this date, I regard those courses to be the best I have ever taken on any subject in my life. He never brought any notes, never opened the book, and explained and derived all of the material in class with an effortlessness that I have never again experienced in my life.

Then, I got to read some of his work, on sufficiency and ancillarity, survey sampling, the likelihood principle, the meaning of the elusive word *information*, the role of randomization in design and in inference, eliminating nuisance parameters, his mystical and enigmatic counterexamples, and also some of his highly technical work using abstract algebra, techniques of complex and Fourier analysis, and on putting statistical concepts in a rigorous measure theoretic framework. I realized that Basu was also a formidable mathematician. The three signature and abiding qualities of nearly all of Basu's work were clarity of exposition, simplicity, and an unusual originality in thinking and in presenting his arguments. Anyone who reads his paper on randomization analysis (Basu (1980)) will ponder about the use of permutation tests and the role of a statistical model. Anyone who reads his papers on survey data, the likelihood principle, information, ancillarity, and sufficiency will be forced to think about the foundations of statistical practice. The work was fascinating and original, and influenced generations of statisticians across the world. Dennis Lindley has called Basu's writings on foundations "among the most significant contributions of our time to the foundations of statistical inference."

Foundations can be frustrating, and disputes on foundations can indeed be never-ending. Although the problems that we, as statisticians, are solving today are different, the fundamentals of the subject of statistics have not greatly changed. Depending on which particular foundational principle we more believe in, it is still the fundamentals laid out by Fisher, Pearson, Neyman, Wald, and the like, that drive statistical inference. Despite decades and volumes devoted to debates over foundational issues, these issues still remain important. In his commentary on Basu's work on survey sampling in this volume, Alan Welsh says "... and this is characteristic of Basu and one of the reasons (that) his papers are still so valuable; it does challenge the usual way ... Statistics has benefitted enormously that Basu made that journey." I could not say it any better. It is with this daunting background that I undertook the work of editing this volume.

This book contains 23 of Basu's most significant articles and writings. These are reprints of the original articles, presented in a chronological order. It also contains eleven commentaries written by some of our most distinguished scholars in the areas of foundations and statistical inference. Each commentary gives an original and contemporary critique of a particular aspect or some particular contribution in Basu's work, and places it in perspective. The commentaries are by George Casella and V. Gopal, Phil Dawid, Tom DiCiccio and Alastair Young, Malay Ghosh, Jay Kadane, Glen Meeden,

Robert Serfling, Jayaram Sethuraman, Terry Speed, and Alan Welsh. I am extremely grateful to each of these discussants for the incredible effort and energy that they have put into writing these commentaries. This book is a much better statistical treasure because of these commentaries.

Terry Speed has eloquently summarized a large portion of Basu's research in his commentary. My comments here serve to complement his. Basu was born on July 5, 1924 in the then undivided Bengal and had a modest upbringing. He obtained an M.A. in pure mathematics from Dacca University in the late forties. His initial interest in mathematics no doubt came from his father, Dr. N. M. Basu, an applied mathematician who worked on the mathematical theory of elasticity under the supervision of Nobel laureate C. V. Raman. Basu told us that shortly after the partition of India and Pakistan, he became a refugee, and crossed the borders to come over to Calcutta. He found a job as an actuary. The work failed to give him any intellectual satisfaction at all. He quit, and at considerable risk, went back to East Pakistan. The adventure did not pay off. He became a refugee for a second time and returned to Calcutta. Here, he came to know of the ISI, and joined the ISI in 1950 as a PhD student under the supervision of C. R. Rao. Basu's PhD dissertation "Contributions to the Theory of Statistical Inference" was nearly exclusively on pure decision theory, minimaxity and admissibility, and risks in testing problems under various loss functions (Basu (1951, 1952a, 1952b)). In Basu (1951), a neat counterexample shows that even the most powerful test in a simple vs. simple testing problem can be inconsistent, if the iid assumption for the sequence of sample observations is violated. In Basu (1952a), an example is given to show that if the ordinary squared error loss is just slightly altered, then a best unbiased estimate with respect to the altered loss function would no longer exist, even in the normal mean problem. Basu (1952b) deals with admissible estimation of a variance for permutation invariant joint distributions and for stationary Markov chains with general convex loss functions. It would seem that C. R. Rao was already thinking of characterization problems in probability, and Basu was most probably influenced by C. R. Rao. Basu wrote some early articles on characterizing normal distributions by properties of linear functions, a topic in which Linnik and his students were greatly interested at that time. This was a passing phase.

In 1953, he came to Berkeley by ship as a Fulbright scholar. He met Neyman, and listened to many of his lectures. Basu spoke effusively of his memories of Neyman, Wald, and David Blackwell at many of his lectures. It was during this time that he learned frequentist decision theory and the classical theory of statistics, extensively and deeply. At the end of the Fulbright scholarship, he went back to India, and joined the ISI as a faculty member. He later became the Dean of Studies, a distinguished scientist, and officiating Director of the Research and Training School at the ISI. He pioneered scholastic aptitude tests in India that encourage mathematics students to understand a topic and stop focusing on cramming. The widely popular aptitude test book Basu, Chatterji, and Mukherjee (1972) is an institution in Indian mathematics all by itself. He also had an ingrained love for beautiful things. One special thing that he personally developed was a beautiful flower garden around the central lake facing the entrance of the main residential campus. It used to be known as *Basu's garden*. He also worked for some years as the chief superintendent of the residential student hostels, and when I joined there, I heard stories about his coming by in the wee hours of the morning to check on students, always with his dog. I am glad that he wasn't the superintendent when I joined, because my mischievous friends and I were stealing coconuts from the campus trees around 4:00 am every morning for an early morning carb boost. I can imagine his angst and disbelief and the angry outrage of his puzzled dog at some of his dedicated students hanging from coconut trees at four in the morning, and a few others hiding in the bushes on guard.

The subject of statistics was growing rapidly in the years around the second world war. The most fundamental questions were being raised, and answered. In quick succession, we had the *Neyman-Pearson lemma*, *Cramér-Rao inequality*, *Rao-Blackwell theorem*, *the Lehmann-Scheffé theorem*, *Neyman's proof of the factorization theorem*, *Mahalanobis's D^2 -statistic*, *the Wald test*, *the score test of Rao*, and *Wald and Le Cam's masterly and all encompassing formulation and development of decision theory as a unified basis for inference*. This was also a golden age for the ISI. Fisher was

spending time there, and so were Kolmogorov, and Haldane. Basu joined the ISI as a PhD student in that fertile and golden era of statistics and the ISI. He was clearly influenced, and deeply so, by Kolmogorov's rigorous measure theoretic development of probability theory, and simultaneously by Fisher's prodigious writings on the core of statistics, maximum likelihood, sufficiency, ancillarity, fiducial distributions, randomization tests, and so on. This influence of Kolmogorov and Fisher is repeatedly seen in much of Basu's later work. The work on foundations is obviously influenced by Fisher's work, and the technical work on sufficiency, ancillarity, and invariance (Basu (1967, 1969)) is very clearly influenced by Kolmogorov. It is not too much of a stretch to call some of Basu's work on sufficiency, ancillarity, and invariance research on abstract measure theory.

Against this backdrop, we see Basu's first transition into raising and answering questions that have something fundamental and original about them. Among the most well known is what everyone knows simply as *Basu's theorem* (Basu (1955a)). It is the only result in statistics and probability that is listed in Wikipedia's *list of Indian inventions and discoveries*, significant scientific inventions and discoveries originating in India in all of recorded history. A few other entries in this list are the hookah, muslin, cotton, pajamas, private toilets, swimming pools, hospitals, plastic surgery, diabetes, jute, diamonds, the number zero, differentials, the Ramanujan theta function, the AKS primality test, Bose-Einstein statistics, and the Raman effect.

The direct part of Basu's theorem says that if X_1, \dots, X_n have a joint distribution $P_{n,\theta}$, $\theta \in \Theta$, $T(X_1, \dots, X_n)$ is boundedly complete and sufficient, and $S(X_1, \dots, X_n)$ is ancillary, then T and S are independent under each $\theta \in \Theta$. The theorem has very pretty applications, and I will mention a few. But, first, I would like to talk a little more of the context of this theorem. He was led to Basu's theorem when he was asked the following question. Consider iid $N(\mu, 1)$ observations X_1, \dots, X_n . Then, every location invariant statistic is ancillary; is the converse true? The converse is not true, and so Basu wanted to characterize the class of all ancillary statistics in this situation. The reverse part of Basu's theorem answers this question; in general, suppose $T(X_1, \dots, X_n)$ is boundedly complete and sufficient. Then, $S(X_1, \dots, X_n)$ is ancillary only if it is independent of T under each θ . Typically, in applications, one would take T to be the minimal sufficient statistic, which has the best chance of being also complete. Without completeness, an ancillary statistic need not be independent of T .

Returning to the more well known part of Basu's theorem, namely the direct part, there is an element of sheer beauty about the result. A sufficient statistic is supposed to capture all the information about the parameter that the full data could supply, and an ancillary statistic has none to offer at all. We can think of a rope, with T and S at the two ends of the rope, and θ placed in the middle. T has everything to do with θ , and S has nothing to do with θ whatsoever. They must be independent! The theorem brings together sufficiency, information, ancillarity, completeness, and conditional independence. Terry Speed (Speed (2010), IMS Bulletin), calls it *a beautiful theorem*, which indeed it is. Basu later worked on various other aspects of ancillarity and selection of reference sets; all of these are comprehensively discussed in Phil Dawid's commentary in this volume.

Basu's theorem isn't only pretty. It has also been used by many researchers in diverse areas of statistics and probability. To name a few, the theorem has been used extensively in distribution theory, in deriving Barndorff-Nielsen's magic formula (Barndorff-Nielsen (1983), Small (2010)), in proving theorems on infinitely divisible distributions, in goodness of fit testing (and in particular for finding the mean and the variance of the *WE* statistic for testing exponentiality), and of late in small area estimation. Hogg and Craig (1956), Lehmann (1981), Boos and Hughes-Oliver (1998), and Ghosh (2002) have previously described some of these applications. Larry Brown has provided some very powerful applications of a more general (but less statistically intuitive) version of Basu's theorem in Brown (1986), and Malay Ghosh has indicated applications to empirical Bayes problems in his commentary in this volume. Here are a few of my personal favorite applications of Basu's theorem to probability theory. The attractive part of these examples is that you save on clumsy or boring calculations by using Basu's theorem in a clever way. The final results can be obtained without

using Basu's theorem, but in a pedestrian sort of way. In contrast, by applying Basu's theorem, you do it elegantly.

Example 1 (Independence of Mean and Variance for a Normal Sample) Suppose X_1, X_2, \dots, X_n are iid $N(\eta, \tau^2)$ for some η, τ . Suppose \bar{X} is the mean and s^2 the variance of the sample values X_1, X_2, \dots, X_n . The goal is to prove that \bar{X} and s^2 are independent, whatever be η and τ . First note the useful reduction that if the result holds for $\eta = 0, \tau = 1$, then it holds for all η, τ . Indeed, fix any η, τ , and write $X_i = \eta + \tau Z_i, 1 \leq i \leq n$, where Z_1, \dots, Z_n are iid $N(0, 1)$. Now,

$$\left(\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2 \right) \stackrel{\mathcal{L}}{=} \left(\eta + \tau \bar{Z}, \tau^2 \sum_{i=1}^n (Z_i - \bar{Z})^2 \right).$$

Therefore, \bar{X} and $\sum_{i=1}^n (X_i - \bar{X})^2$ are independently distributed under (η, τ) if and only if \bar{Z} and $\sum_{i=1}^n (Z_i - \bar{Z})^2$ are independently distributed. This is a step in getting rid of the parameters η, τ from consideration. But, now, we will import a parameter! Embed the $N(0, 1)$ distribution into a larger family of $\{N(\mu, 1), \mu \in \mathcal{R}\}$ distributions. Consider now a fictitious sample Y_1, Y_2, \dots, Y_n from $P_\mu = N(\mu, 1)$. The joint density of $Y = (Y_1, Y_2, \dots, Y_n)$ is a one parameter Exponential family density with the natural sufficient statistic $T(Y) = \sum_{i=1}^n Y_i$. And, of course, $\sum_{i=1}^n (Y_i - \bar{Y})^2$ is ancillary. Since this is an Exponential family, and the parameter space for μ obviously has a nonempty interior, all the conditions of Basu's theorem are satisfied, and therefore, under each $\mu, \sum_{i=1}^n Y_i$ and $\sum_{i=1}^n (Y_i - \bar{Y})^2$ are independently distributed. In particular, they are independently distributed under $\mu = 0$, i.e., when the samples are iid $N(0, 1)$, which is what we needed to prove. This is surely a very pretty proof of that classic fact in distribution theory.

Example 2 (An Exponential Distribution Result) Suppose X_1, X_2, \dots, X_n are iid Exponential random variables with mean λ . Then, by transforming (X_1, X_2, \dots, X_n) to $\left(\frac{X_1}{X_1 + \dots + X_n}, \dots, \frac{X_{n-1}}{X_1 + \dots + X_n}, X_1 + \dots + X_n \right)$, one can show by carrying out the necessary Jacobian calculation that $\left(\frac{X_1}{X_1 + \dots + X_n}, \dots, \frac{X_{n-1}}{X_1 + \dots + X_n} \right)$ is independent of $X_1 + \dots + X_n$. We can show this without doing any calculations by using Basu's theorem.

For this, once again, by writing $X_i = \lambda Z_i, 1 \leq i \leq n$, where the Z_i are iid standard Exponentials, first observe that $\left(\frac{X_1}{X_1 + \dots + X_n}, \dots, \frac{X_{n-1}}{X_1 + \dots + X_n} \right)$ is a (vector) ancillary statistic. Next observe that the joint density of $X = (X_1, X_2, \dots, X_n)$ is a one parameter Exponential family, with the natural sufficient statistic $T(X) = X_1 + \dots + X_n$. Since the parameter space $(0, \infty)$ obviously contains a nonempty interior, by Basu's theorem, under each $\lambda, \left(\frac{X_1}{X_1 + \dots + X_n}, \dots, \frac{X_{n-1}}{X_1 + \dots + X_n} \right)$ and $X_1 + \dots + X_n$ are independently distributed.

Example 3 (A Weak Convergence Result Using Basu's Theorem) Suppose X_1, X_2, \dots are iid random vectors with a uniform distribution in the d -dimensional unit ball. For $n \geq 1$, let $d_n = \min_{1 \leq i \leq n} \|X_i\|$, and $D_n = \max_{1 \leq i \leq n} \|X_i\|$. Thus, d_n measures the distance to the closest data point from the center of the ball, and D_n measures the distance to the farthest data point. We find the limiting distribution of $\rho_n = \frac{d_n}{D_n}$. Although this can be done by using Slutsky's theorem, the Borel-Cantelli lemma, and some direct algebra, we will do so by an application of Basu's theorem.

Toward this, note that for $0 \leq u \leq 1$,

$$P(d_n > u) = (1 - u^d)^n; \quad P(D_n > u) = 1 - u^{nd}.$$

As a consequence, for any $k \geq 1$,

$$E[D_n]^k = \int_0^1 ku^{k-1}(1-u^{nd})du = \frac{nd}{nd+k},$$

and,

$$E[d_n]^k = \int_0^1 ku^{k-1}(1-u^d)^n du = \frac{n!\Gamma\left(\frac{k}{d}+1\right)}{\Gamma\left(n+\frac{k}{d}+1\right)}.$$

Now, embed the uniform distribution in the unit ball into the family of uniform distributions in balls of radius θ and centered at the origin. Then, D_n is complete and sufficient, and ρ_n is ancillary. Therefore, once again, by Basu's theorem, D_n and ρ_n are independently distributed under each $\theta > 0$, and so, in particular under $\theta = 1$. Thus, for any $k \geq 1$,

$$\begin{aligned} E[d_n]^k &= E[D_n \rho_n]^k = E[D_n]^k E[\rho_n]^k \\ \Rightarrow E[\rho_n]^k &= \frac{E[d_n]^k}{E[D_n]^k} = \frac{n!\Gamma\left(\frac{k}{d}+1\right)}{\Gamma\left(n+\frac{k}{d}+1\right)} \frac{nd+k}{nd} \\ &\sim \frac{\Gamma\left(\frac{k}{d}+1\right) e^{-n} n^{n+1/2}}{e^{-n-k/d} \left(n+\frac{k}{d}\right)^{n+\frac{k}{d}+1/2}} \end{aligned}$$

(by using Stirling's approximation)

$$\sim \frac{\Gamma\left(\frac{k}{d}+1\right)}{\frac{k}{n^{\frac{1}{d}}}}.$$

Thus, for each $k \geq 1$,

$$E\left[n^{1/d} \rho_n\right]^k \rightarrow \Gamma\left(\frac{k}{d}+1\right) = E[V]^{k/d} = E[V^{1/d}]^k,$$

where V is a standard Exponential random variable. This implies, because $V^{1/d}$ is uniquely determined by its moment sequence, that

$$n^{1/d} \rho_n \xrightarrow{\mathcal{L}} V^{1/d},$$

as $n \rightarrow \infty$.

Example 4 (An Application of Basu’s Theorem to Quality Control) Herman Rubin kindly suggested that I give this example. He used Basu’s theorem to answer a question on statistical quality control while consulting some engineers in the fifties. Here is the problem, which is simple to state.

In a production process, the measurement X of a certain product has to fall between *conformity limits* a, b . For design, as well as for quality monitoring, the management wants to estimate what percentage of items are currently meeting the conformity specifications. That is, we wish to estimate $\theta = P(a < X < b)$. Suppose that for estimation purpose, we have obtained data values X_1, X_2, \dots, X_n , which we assume are iid $N(\mu, \sigma^2)$ for some μ, σ . Then, $\theta = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$. Standard plug-in estimates are asymptotically efficient. But quality control engineers have an inherent preference for the UMVUE. We derive the UMVUE below in closed form by using Basu’s theorem and the Lehmann-Scheffé theorem.

By the Lehmann-Scheffé theorem, the UMVUE is the conditional expectation

$$\begin{aligned} E(I_{a \leq X_1 \leq b} | \bar{X}, s) &= P(a \leq X_1 \leq b | \bar{X}, s) \\ &= P\left(\frac{a - \bar{X}}{s} \leq \frac{X_1 - \bar{X}}{s} \leq \frac{b - \bar{X}}{s} | \bar{X}, s\right) \end{aligned}$$

Now, (\bar{X}, s) are jointly sufficient and complete, and $\frac{X_1 - \bar{X}}{s}$ is evidently ancillary. Therefore, by Basu’s theorem, we get the critical simplification that the conditional distribution of $\frac{X_1 - \bar{X}}{s}$ given (\bar{X}, s) is the same as the unconditional distribution (whatever it is) of this ancillary statistic $\frac{X_1 - \bar{X}}{s}$. Hence, the UMVUE of θ is

$$= P\left(\frac{a - \bar{X}}{s} \leq \frac{X_1 - \bar{X}}{s} \leq \frac{b - \bar{X}}{s}\right) = F_n\left(\frac{b - \bar{X}}{s}\right) - F_n\left(\frac{a - \bar{X}}{s}\right),$$

where $F_n(t)$ denotes the CDF of the unconditional distribution of $\frac{X_1 - \bar{X}}{s}$.

We can, perhaps a little fortunately, compute this in closed form. It is completely obvious that the mean of F_n is zero and that the second moment is $\frac{1}{n}$. With a little calculation, which we will omit, F_n can be shown to be a Beta distribution on $[-1, 1]$ (*in fact, even this fact, which I am not proving here, can be proved by using Basu’s theorem*). In other words, F_n has the density

$$f_n(x) = \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\sqrt{\pi}\Gamma(\alpha)}(1 - x^2)^{\alpha-1}, \quad -1 \leq x \leq 1.$$

The value of α must be $\frac{n-1}{2}$, by virtue of the second moment being $\frac{1}{n}$. Hence, the UMVUE of θ is

$$F_n\left(\frac{b - \bar{X}}{s}\right) - F_n\left(\frac{a - \bar{X}}{s}\right)$$

where

$$F_n(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)} \int_{-1}^t (1 - x^2)^{\frac{n-3}{2}} dx$$

$$= \frac{1}{2} + \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)} t {}_2F_1\left(\frac{1}{2}, \frac{3-n}{2}; \frac{3}{2}; t^2\right),$$

where ${}_2F_1$ denotes the usual *Gauss Hypergeometric function*. This describes the UMVUE of θ in closed form. Herman Rubin or I have not personally seen this very closed form derivation involving the Hypergeometric function anywhere, but it is another instance where Basu’s theorem makes the problem solvable without breaking our back.

Example 5 (A Covariance Calculation) Suppose X_1, \dots, X_n are iid $N(0, 1)$, and let \bar{X} and M_n denote the mean and the median of the sample set X_1, \dots, X_n . By using our old trick of importing a mean parameter μ , we first observe that the difference statistic $\bar{X} - M_n$ is ancillary. By Basu’s theorem, $X_1 + \dots + X_n$ and $\bar{X} - M_n$ are independent under each μ , which implies

$$\begin{aligned} \text{Cov}(X_1 + \dots + X_n, \bar{X} - M_n) &= 0 \Rightarrow \text{Cov}(\bar{X}, \bar{X} - M_n) = 0 \\ &\Rightarrow \text{Cov}(\bar{X}, M_n) = \text{Cov}(\bar{X}, \bar{X}) = \text{Var}(\bar{X}) = \frac{1}{n}. \end{aligned}$$

We have achieved this result without doing any calculations at all.

Example 6 (Application to Infinite Divisibility) Infinitely divisible distributions are important in both the theoretical aspects of weak convergence of partial sums of triangular arrays, and in real applications. Here is one illustration of the use of Basu’s theorem in producing unconventional infinitely divisible distributions. The example is based on the following general theorem (DasGupta (2006)), whose proof uses *both* Basu’s theorem and the Goldie-Steutel law (Goldie (1967)).

Theorem Let f be any homogeneous function of two variables, i.e., suppose $f(cx, cy) = c^2 f(x, y) \forall x, y$ and $\forall c > 0$. Let Z_1, Z_2 be iid $N(0, 1)$ random variables and Z_3, Z_4, \dots, Z_m any other random variables such that (Z_3, Z_4, \dots, Z_m) is independent of (Z_1, Z_2) . Then for any positive integer k , and an arbitrary measurable function g , $f^k(Z_1, Z_2)g(Z_3, Z_4, \dots, Z_m)$ is infinitely divisible.

A large number of explicit densities can be proved to be densities of infinitely divisible distributions by using this theorem. Here are a few, each supported on the entire real line.

(i) $f(x) = \frac{1}{\pi} K_0(|x|)$, where K_0 denotes the Bessel K_0 function;

(ii) $f(x) = \frac{2 \log(|x|)}{\pi^2(x^2 - 1)}$;

(iii) $f(x) = \frac{1}{\sqrt{2\pi}} \left(1 - \sqrt{2\pi}|x|e^{x^2/2}\Phi(-|x|)\right)$;

(iv) $f(x) = \frac{1}{2} \frac{1}{(1 + |x|)^2}$.

Note that the density in (iv) is the so called *GT density*. Of course, we can introduce location and scale parameters into all of these, and make families of infinitely divisible distributions.

Continuing on with some other significant contributions of Basu, his mathematical work as well as his foundational writings on survey sampling helped put sampling theory on a common mathematical footing with the rest of rigorous statistical theory, and even more, in raising and clarifying really fundamental issues. Alan Welsh has discussed the famous example of *Basu's elephants* (Basu (1971)) in his commentary in this volume. This is the solitary example that I know of where a single example has led to the writing of a book with the example in its title (Brewer (2002)). The elephants example must be understood in the context of the theme and also the time. It was written when the Horvitz-Thompson estimator for a finite population total was gaining theoretical popularity, and many were accepting the estimator as an automatic choice. Basu's example reveals a fundamental flaw in the estimator in particular, and in the wisdom of sample space based optimality, more generally. The example is so striking and entertaining, that I cannot but reproduce it here.

Example 7 (Basu's Elephants) “The circus owner is planning to ship his 50 adult elephants and so he needs a rough estimate of the total weight of the elephants. As weighing an elephant is a cumbersome process, the owner wants to estimate the total weight by weighing just one elephant. Which elephant should he weigh? So the owner looks back at his records and discovers a list of the elephants' weights taken 3 years ago. He finds that 3 years ago Sambo the middle-sized elephant was the average (in weight) elephant in his herd. He checks with the elephant trainer who reassures him (the owner) that Sambo may still be considered to be the average elephant in his herd. Therefore, the owner plans to weigh Sambo and take $50y$ (where y is the present weight of Sambo) as an estimate of the total weight $Y = Y_1 + \dots + Y_{50}$ of the 50 elephants. But the circus statistician is horrified when he learns of the owner's purposive sampling plan. “How can you get an unbiased estimate of Y this way?” protests the statistician. So together they work out a compromise sampling plan. With the help of a table of random numbers, they devise a plan that allots a selection probability of $99/100$ to Sambo and equal selection probabilities of $1/4900$ to each of the other 49 elephants. Naturally, Sambo is selected, and the owner is happy. “How are you going to estimate Y ?”, asks the statistician. “Why? The estimate ought to be $50y$ of course,” says the owner. “Oh! No! That cannot possibly be right,” says the statistician. “I recently read an article in the *Annals of Mathematical Statistics* where it is proved that the Horvitz-Thompson estimator is the unique hyperadmissible estimator in the class of all generalized polynomial unbiased estimators.” “What is the Horvitz-Thompson estimate in this case?”, asks the owner, duly impressed. “Since the selection probability for Sambo in our plan was $99/100$,” says the statistician, “the proper estimate of Y is $\frac{100y}{99}$ and not $50y$.” “And how would you have estimated Y ,” enquires the incredulous owner, “if our sampling plan made us select, say, the big elephant Jumbo?” “According to what I understand of the Horvitz-Thompson estimation method,” says the unhappy statistician, “the proper estimate of Y would then have been $4900y$, where y is Jumbo's weight.” That is how the statistician lost his circus job (and perhaps became a teacher of statistics!).”

Some of my other personal favorites are a number of his counterexamples. The examples always used to have something dramatic or penetrating about them. He would take a definition, or an idea, or an accepted notion, and chase it to its core. Then, he would give a remarkable example to reveal a fundamental flaw in the idea and it would be very very difficult to refute it. One example of this is his well known *ticket example* (Basu (1975)). The point of this example was to argue that blind use of the maximum likelihood estimate, even if there is just one parameter, is risky. In the ticket example, Basu shows that the MLE overestimates the parameter by a huge factor with a probability nearly equal to one. The example was constructed to make the likelihood function have a global peak at the wrong place. Basu drives home the point that one must look at the entire likelihood function, and not just where it peaks. Very reasonable, especially these days, when so many of us just throw the data into a computer and get the MLE and feel happy about it. Jay Kadane has discussed Basu's epic paper on likelihood (Basu (1975)) for this volume.

In this very volume, Robert Serfling discusses Basu's definition of asymptotic efficiency through concentration of measures (Basu (1956)) and the counterexample which puts his definition at the extreme opposite pole of the traditional definition of asymptotic efficiency. An important aspect of Basu's definition of asymptotic relative efficiency is that it isn't wedded to asymptotic normality, or \sqrt{n} -consistency. You could use it, for example, to compare the mean and the midrange in the normal case, which you cannot do according to the traditional definition of asymptotic relative efficiency.

A third example, but of less conceptual gravity, is his example of an inconsistent MLE (Basu (1955b)). The most famous example in that domain is certainly the Neyman-Scott example (Neyman and Scott (1948)). In the Neyman-Scott example, the inconsistency is caused by a nonvanishing bias, and once the bias is corrected, consistency is retrieved. Basu's example is pathological statistically, but like all his examples, this too makes the point in the most extreme conceivable way. The inconsistent MLE isn't fixable in his example.

One important point about Basu's writings is that it is never clear that he does not like the procedures that classical criteria produce. In numerous writings, he uses a time tested classical procedure. *But he only questions the rationale behind choosing them.* This distinction is important. I feel that in these matters, he is closer to Dennis Lindley, who too, reportedly holds the view that classical statistics generally produces fine procedures, but using the wrong reasons. This seems to be very far from a dogmatic view that all classical procedures deserve to be discarded because of where they came from. But, even when Basu questioned the criteria for selecting a procedure in his writings, and in his seminars, it was always in the best of spirits (George Casella comments in this volume that "the banter between Basu and Kempthorne is fit for a comedy.")

Shortly after he taught us at the ISI, Basu left India and moved to the USA. He joined the faculty of the Florida State University, causing, according to his daughter Monimala, a family rebellion. They would happily settle in Sydney, or Denmark, or Ottawa, or Sheffield, where Basu used to visit frequently, even Timbuktu, but not in a small town in Florida. After he moved to the US, one can see a distinct change of perspective and emphasis in Basu's work. He now started working on more *practical things*; concrete elimination of nuisance parameters, modelling Bayesian bioassays with missing data (Basu and Pereira (1982a)), randomization tests (1980), and Bayesian nonparametrics. His involvement in Bayesian nonparametrics resulted in a beautifully written paper on Dirichlet processes starting from absolute scratch (Basu and Tiwari (1982b)). Jayaram Sethuraman superbly discusses this paper in this volume. In a way, Basu's involvement in Bayesian nonparametrics was perhaps a little surprising. This is not because he could not deal with abstract measure theory; he was an expert on it! But, because, Basu repeatedly expressed his deep rooted skepticism about putting priors on *more than three or four parameters*. He never said what he would do in problems with many parameters. But he would not accept improper priors, or even empirical Bayes. He simply said that he does not know what to do if one has many parameters, because you then just can't write your elicited information into honest priors. In some ways, Basu was a half hearted Bayesian. But, he was very forthcoming.

Basu returned permanently to India 1986. He still taught and lectured at the ISI. I last saw Basu in 1995 at East Lansing. He, his wife Kalyani, and daughter Monimala were all visiting his son Shantanu, who was then a postdoctoral scientist at Michigan State University. I spent a few hours with him, and I told him what I was working on. I asked him if he would like to give a seminar at Purdue. He said that the time came some years ago that he should now only listen, *to young people like you*, and not talk. He said that his time has passed, and he only wants to learn now, and not profess. He often questioned himself. In a nearly spiritual mood, he once wrote (Ghosh (1988)): "What have I got to offer? I am afraid nothing but a set of negative propositions. But with all humility, let me draw the attention of the would be reader to the ancient Vedic (Hindu scripture) saying- "neti, neti, . . . , iti," which means- "not this, not this, . . . , this!"

Basu went back to Calcutta from Lansing, and I received letters from him periodically. In the March of 2001, when I was visiting Larry Brown at the Wharton school, one morning an e-mail came from

B.V. Rao at Calcutta. The e-mail said that he is duty bound to give me the saddest news, that Dr. Basu passed away the night before. I went to Larry Brown's office and gave him the news. Larry looked at me, as if he did not believe what I said, and I saw his eyes glistening up, and he said - "that's just too bad. He was such a good guy." However, the idealism and the inspiration live on, as strongly as in 1973, when he walked into that classroom at the ISI to meet twenty 16 year olds, and fifty minutes later, we were all in love with probability. I know that I speak for numerous people who got to know Basu as closely as we did, that he was a personification of purity, in scholarship and in character. There was an undefinable poetic and ethereal element in the man, his personality, his presence, his writings, and his angelic disposition, that is very very hard to find. He is missed dearly, but he lives in our memory. The legacy remains ever so strong. I do tell my students in my PhD theory course; *read Basu*.

West Lafayette, Indiana, USA

Anirban DasGupta

Preface Bibliography

- Barndorff-Nielsen, O. (1983). On a formula for the conditional distribution of the maximum likelihood estimator, *Biometrika*, 70, 343–365.
- Basu, D. (1951). A note on the power of the best critical region for increasing sample sizes, *Sankhyā*, 11, 187–190.
- Basu, D. (1952a). An example of non-existence of a minimum variance estimator, *Sankhyā*, 12, 43–44.
- Basu, D. (1952b). On symmetric estimators in point estimation with convex weight functions, *Sankhyā*, 12, 45–52.
- Basu, D. (1955a). On statistics independent of a complete sufficient statistic, *Sankhyā*, 15, 377–380.
- Basu, D. (1955b). An inconsistency of the method of maximum likelihood, *Ann. Math. Statist.*, 26, 144–145.
- Basu, D. (1956). The concept of asymptotic efficiency, *Sankhyā*, 17, 193–196.
- Basu, D. (1967). Problems related to the existence of minimal and maximal elements in some families of subfields, *Proc. Fifth Berkeley Symp. Math. Statist. and Prob.*, I, 41–50, Univ. California Press, Berkeley.
- Basu, D. (1969). On sufficiency and invariance, in *Essays on Probability and Statistics*, R.C. Bose et al. eds., University of North Carolina Press, Chapel Hill.
- Basu, D. (1971). An essay on the logical foundations of survey sampling, with discussions, in *Foundations of Statistical Inference*, V. P. Godambe and D. A. Sprott eds., Holt, Rinehart, and Winston of Canada, Toronto.
- Basu, D. (1975). Statistical information and likelihood, with discussion and correspondence between Barnard and Basu, *Sankhyā*, Ser. A, 37, 1–71.
- Basu, D. (1980). Randomization analysis of experimental data: The Fisher randomization test, with discussions, *Jour. Amer. Statist. Assoc.*, 75, 575–595.
- Basu, D., Chatterji, S., and Mukherjee, M. (1972). *Aptitude Tests for Mathematics and Statistics*, Statistical Publishing Society, Calcutta.
- Basu, D. and Pereira, C. (1982a). On the Bayesian analysis of categorical data: The problem of nonresponse, *Jour. Stat. Planning Inf.*, 6, 345–362.
- Basu, D. and Tiwari, R. (1982b). A note on the Dirichlet process, in *Essays in Honor of C.R. Rao*, 89–103, G. Kallianpur, P.R. Krishnaiah, and J.K. Ghosh eds., North-Holland, Amsterdam.
- Boos, D. and Hughes-Oliver, J. (1998). Applications of Basu's theorem, *Amer. Statist.*, 52, 218–221.
- Brewer, K. (2002). *Combined Survey Sampling Inference: Weighing Basu's Elephants*, Arnold, London.
- Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families*, Institute of Mathematical Statistics, Hayward, California.
- DasGupta, A. (2006). Extensions to Basu's theorem, infinite divisibility, and factorizations, *Jour. Stat. Planning Inf.*, 137, 945–952.
- Ghosh, J. K. (1988). *A Collection of Critical Essays by Dr. D. Basu*, Springer-Verlag, New York.
- Ghosh, M. (2002). Basu's theorem with applications: A personalistic review, *Sankhyā*, Ser. A, 64, 509–531.
- Goldie, C. (1967). A class of infinitely divisible random variables, *Proc. Cambridge Philos. Soc.*, 63, 1141–1143.
- Hogg, R. and Craig, A. T. (1956). Sufficient statistics in elementary distribution theory, *Sankhyā*, 17, 209.
- Lehmann, E. L. (1981). An interpretation of completeness and Basu's theorem, *Jour. Amer. Statist. Assoc.*, 76, 335–340.
- Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations, *Econometrica*, 16, 1–32.
- Small, C. G. (2010). *Expansions and asymptotics for statistics*, Chapman and Hall, Boca Raton.
- Speed, T. (2010). *IMS Bulletin*, 39, 1, 9–10.

Acknowledgment I am thankful to Peter Hall for his critical help in acquiring the reprint rights of some of Basu's most important articles. Alan Welsh rescued me far too many times from an impending disaster. B.V. Rao and Bill Strawderman went over this preface very very carefully and made numerous suggestions for improving on an earlier draft. My longtime and dear friend Srinivas Bhogle graciously helped me with editing the preface. Amarjit Budhiraja, Shih-Chuan Cheng, Doug Crabill, Ritabrata Dutta, Ingram Olkin, Sastry Pantula, Bhamidi Shankar and Ram Tiwari helped me with acquisition of some of the articles. Shantanu and Monimala Basu gave me important biographical information. Springer's scientific editors Patrick Carr and Marc Strauss were tremendously helpful. Integra Software Services at Pondicherry did a fabulous job of producing this volume. And, John Kimmel was just his usual self, the best supporter of every good cause.

Acknowledgements

This series of selected works is possible only because of the efforts and cooperation of many people, societies, and publishers. The series editors originated the series and directed its development. The volume editors spent a great deal of time compiling the previously published material and the contributors provided comments on the significance of the papers. The societies and publishers who own the copyright to the original material made the volumes possible and affordable by their generous cooperation:

American Institute of Physics
American Mathematical Society
American Statistical Association
Applied Probability Trust
Bernoulli Society
Cambridge University Press
Canadian Mathematical Society
Elsevier
Foundation of the Scandinavian Journal of Statistics
Indian Statistical Institute
Institute of Mathematical Statistics
International Chinese Statistical Association
International Statistical Institute
John Wiley and Sons
l'Institut Fourier
London Mathematical Society
New Zealand Statistical Association
Oxford University Press
Polish Academy of Sciences
Princeton University and the Institute for Advanced Studies
Springer
Statistical Society of Australia
University of California Press
University of Illinois, Department of Mathematics
University of Michigan, Department of Mathematics
University of North Carolina Press

Contents

Basu's Work on Randomization and Data Analysis	1
George Casella and Vikneswaran Gopal	
Basu on Ancillarity	5
Philip Dawid	
Conditional Inference by Estimation of a Marginal Distribution	9
Thomas J. DiCiccio and G. Alastair Young	
Basu's Theorem	15
Malay Ghosh	
Basu's Work on Likelihood and Information	23
Joseph B. Kadane	
Basu on Survey Sampling	25
Glen Meeden	
Commentary on Basu (1956)	27
Robert J. Serfling	
Commentary on <i>A Note on the Dirichlet Process</i>	31
Jayaram Sethuraman	
Commentary on D. Basu's Papers on Sufficiency and Related Topics	35
T.P. Speed	
Basu on Randomization Tests	41
A.H. Welsh	
Basu on Survey Sampling	45
A.H. Welsh	
On Symmetric Estimators in Point Estimation with Convex Weight Functions	51
D. Basu	
An Inconsistency of the Method of Maximum Likelihood	59
D. Basu	
On Statistics Independent of a Complete Sufficient Statistic	61
D. Basu	
The Concept of Asymptotic Efficiency	65
D. Basu	

On Statistics Independent of Sufficient Statistics	69
D. Basu	
On Sampling With and Without Replacement	73
D. Basu	
The Family of Ancillary Statistics	81
D. Basu	
Recovery of Ancillary Information	91
D. Basu	
Problems Relating to the Existence of Maximal and Minimal Elements in Some Families of Statistics (Subfields)	105
D. Basu	
Invariant sets for translation-parameter Families of Measures	115
D. Basu and J. K. Ghosh	
Role of the Sufficiency and Likelihood Principles in Sample Survey Theory	129
D. Basu	
On Sufficiency and Invariance	143
D. Basu	
An Essay on the Logical Foundations of Survey Sampling, Part One	167
D. Basu	
Statistical Information and Likelihood	207
D. Basu	
On the Elimination of Nuisance Parameters	279
Debabrata Basu	
On Partial Sufficiency: A Review	291
D. Basu	
Randomization Analysis of Experimental Data: The Fisher Randomization Test	305
D. Basu	
Ancillary Statistics, Pivotal Quantities and Confidence Statements	327
A Note on Sufficiency in Coherent Models	343
D. Basu and S.C. Cheng	
A Note on the Dirichlet Process	355
D. Basu and R. C. Tiwari	
Conditional Independence in Statistics	371
D. Basu and Carlos A. B. Pereira	
A Note on Blackwell Sufficiency and a Skibinsky Characterization of Distributions	385
D. Basu and Carlos A. B. Pereira	
Learning Statistics from Counter Examples: Ancillary Statistics	391
D. Basu	

Contributors

George Casella Distinguished Professor of Statistics, University of Florida, Gainesville, FL 32611, casella@stat.ufl.edu

Philip Dawid Statistical Laboratory, Centre for Mathematical Sciences, CB3 0WB, UK, apd@statslab.cam.ac.uk

Thomas J. DiCiccio ILR School, Cornell University, Ithaca, NY 14853, tjd9@cornell.edu

Malay Ghosh Distinguished Professor of Statistics, University of Florida, Gainesville, FL 32611, ghoshm@stat.ufl.edu

Vikneswaran Gopal Department of Statistics, University of Florida, Gainesville, FL 32611, viknesh@stat.ufl.edu

Joseph B. Kadane, Emeritus Leonard J. Savage University Professor of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, kadane@stat.cmu.edu

Glen Meeden Chairman and Head, School of Theoretical Statistics, University of Minnesota, Minneapolis, MN 55455, glen@stat.umn.edu

Robert J. Serfling Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75083, serfling@utdallas.edu

Jayaram Sethuraman, Emeritus Robert O. Lawton Distinguished Professor of Statistics, Florida State University, Tallahassee, FL 32306, sethu@ani.stat.fsu.edu

T.P. Speed Department of Statistics, University of California, Berkeley, CA 94720, terry@stat.berkeley.edu

A.H. Welsh Centre for Mathematics and its Applications, The Australian National University, Canberra, ACT 0200, Australia, alan.welsh@anu.edu.au

G. Alastair Young Department of Mathematics, Imperial College, London, UK, alastair.young@imperial.ac.uk

Author Bibliography

1. D. Basu. Contributions to the Theory of Statistical Inference, PhD Dissertation, Submitted to the University of Calcutta, 1953.
2. D. Basu. A note on the power of the best critical region for increasing sample sizes, *Sankhyā*, 11, 187–190, 1951.
3. D. Basu. On the limit points of relative frequencies, *Sankhyā*, 11, 379–382, 1951.
4. D. Basu. On the independence of linear functions of independence chance variables, *Bull. Inst. Internat. Statist.*, 23, II, 83–96, 1951.
5. D. Basu. On the minimax approach to the problem of estimation. *Proc. Nat. Inst. Sci. India*, 18, 287–299, 1952.
6. D. Basu. An example of non-existence of a minimum variance estimator, *Sankhyā*, 12, 43–44, 1952.
7. D. Basu. On symmetric estimators in point estimation with convex weight functions, *Sankhyā*, 12, 45–52, 1952.
8. D. Basu. On a class of admissible estimators of the normal variance, *Sankhyā*, 12, 57–62, 1952.
9. D. Basu. Choosing between two simple hypotheses and the criterion of consistency, *Proc. Nat. Inst. Sci. India*, 19, 841–849, 1953.
10. D. Basu and R. G. Laha. On some characterizations of the normal distribution, *Sankhyā*, 13, 359–362, 1954.
11. D. Basu. On the optimum character of some estimators in multistage sampling problems, *Sankhyā*, 13, 363–368, 1954.
12. D. Basu. A note on mappings of probability spaces, *Vestnik Leningrad Univ.*, 10, 5, 33–35, 1955.
13. D. Basu. An inconsistency of the method of maximum likelihood, *Ann. Math. Statist.*, 26, 144–145, 1955.
14. D. Basu. A note on the structure of a stochastic model considered by V. M. Dandekar, *Sankhyā*, 15, 251–253, 1955.
15. D. Basu. On statistics independent of a complete sufficient statistic, *Sankhyā*, 15, 377–380, 1955.
16. D. Basu. A note on the theory of unbiased estimation, *Ann. Math. Statist.*, 26, 345–348.
17. D. Basu. The concept of asymptotic efficiency, *Sankhyā*, 17, 193–196, 1956.
18. D. Basu. A note on the multivariate extension of some theorems related to the univariate normal distribution, *Sankhyā*, 17, 221–224.
19. D. Basu. On statistics independent of sufficient statistics, *Sankhyā*, 20, 223–226, 1958.
20. D. Basu. On sampling with and without replacement, *Sankhyā*, 20, 287–294, 1958.
21. D. Basu. The family of ancillary statistics, *Sankhyā*, 21, 247–256, 1959.
22. D. Basu. Recovery of ancillary information, *Sankhyā*, 26, 3–16, 1964.
23. D. Basu. Sufficiency and model preserving transformations, Technical Report, Dept. of Statistics, University of North Carolina, Chapel Hill, 1965.

24. D. Basu. Problems related to the existence of minimal and maximal elements in some families of subfields, Proc. Fifth Berkeley Symp. Math. Statist. and Prob., I, 41–50, Univ. California Press, Berkeley, 1967.
25. D. Basu and J. K. Ghosh. Invariant sets for translation parameter families of distributions, Ann. Math. Statist., 40, 162–174, 1969.
26. D. Basu. Role of sufficiency and likelihood principle in sample survey theory, Sankhyā, 31, 441–454, 1969.
27. D. Basu and J. K. Ghosh. Sufficient statistics in sampling from a finite universe, Bull. Inst. Internat. Statist., 42, 850–858, 1969.
28. D. Basu. On sufficiency and invariance, in *Essays on Probability and Statistics*, 61–84, R.C. Bose et al. eds., University of North Carolina Press, 1970.
29. D. Basu. An essay on the logical foundations of survey sampling, with discussions, in *Foundations of Statistical Inference*, 203–242, V. P. Godambe and D. A. Sprott eds., Holt, Rinehart, and Winston of Canada, Toronto, 1971.
30. D. Basu. Discussion of “A note on Basu’s examples of anomalous ancillary statistics” by G. A. Barnard and D. A. Sprott, in *Foundations of Statistical Inference*, 163–176, V. P. Godambe and D. A. Sprott eds., Holt, Rinehart, and Winston of Canada, Toronto, 1971.
31. D. Basu, S. Chatterji, and M. Mukherjee. Aptitude Tests for Mathematics and Statistics, Statistical Publishing Society, Calcutta, 1972.
32. D. Basu. Discussion of “The use of the structural model in combining data bearing on the same characteristics” by J. B. Whitney in *Foundations of Statistical Inference*, 393–408, V. P. Godambe and D. A. Sprott eds., Holt, Rinehart, and Winston of Canada, Toronto, 1972.
33. D. Basu. Discussion of “Exact tests, confidence regions, and estimates” by P. Martin-Löf, in *Foundational Questions in Statistical Inference*, 121–138, Memoirs, No. 1, Dept. Theoretical Statist., Inst. Math., Univ. Aarhus, Aarhus, 1974.
34. D. Basu. Discussion of “Conditionality, pivotals, and robust estimation” by G. A. Barnard, in *Foundational Questions in Statistical Inference*, 61–80, Memoirs, No. 1, Dept. Theoretical Statist., Inst. Math., Univ. Aarhus, Aarhus, 1974.
35. D. Basu. Discussion of “Sufficiency, Prediction, and Extreme models,” by S. Lauritzen, in *Foundational Questions in Statistical Inference*, 249–269, Memoirs, No. 1, Dept. Theoretical Statist., Inst. Math., Univ. Aarhus, Aarhus, 1974.
36. Discussion of “Redundancy and its use as a quantitative measure of the deviation between a statistical hypothesis and a set of observational data” by P. Martin-Löf, in *Foundational Questions in Statistical Inference*, 1–42, Memoirs, No. 1, Dept. Theoretical Statist., Inst. Math., Univ. Aarhus, Aarhus, 1974.
37. D. Basu. Statistical information and likelihood, with discussion and correspondence between Barnard and Basu, Sankhyā, Ser. A, 37, 1–71, 1975.
38. D. Basu. On the elimination of nuisance parameters, Jour. Amer. Statist. Assoc., 72, 355–366, 1977.
39. D. Basu. On partial sufficiency: A review, Jour. Stat. Planning Inf., 2, 1–13, 1978.
40. D. Basu. On the relevance of randomization in data analysis, with discussions, *Survey Sampling and Measurement*, 267–339, N. Namboodiri ed., Academic Press, New York, 1978.
41. D. Basu. Discussion of “In dispraise of the exact tes” by Joseph Berkson, Jour. Stat. Planning Inf., 3, 189–192, 1979.
42. D. Basu. Survey Theory, in *Incomplete Data on Sample Surveys*, 3, 407–410, Academic Press, New York, 1979.
43. D. Basu. Randomization analysis of experimental data: The Fisher randomization test, with discussions, Jour. Amer. Statist. Assoc., 75, 575–595, 1980.
44. D. Basu. On ancillary statistics, pivotal quantities, and confidence statements, in *Topics in Applied Statistics*, 1–29, V. P. Chaubey and T. D. Dwivedi eds., Concordia University Publication, 1981.

Author Bibliography

45. D. Basu. Basu Theorems, *Encyclopedia of Statistical Sciences*, 1, 193–196, S. Kotz and N. L. Johnson eds., Wiley, New York, 1982.
46. D. Basu and S. C. Cheng. A note on sufficiency in coherent models, *Internat. Jour. Math. Math. Sci.*, 3, 571–582, 1981.
47. D. Basu. Likelihood, *Proceedings of National Symposium on Probability and Statistics*, Univ. Sao Paulo, Brazil, 1981.
48. D. Basu and C. B. Pereira. On the Bayesian analysis of categorical data: The problem of nonresponse, *Jour. Stat. Planning Inf.*, 6, 345–362, 1982.
49. D. Basu and R. Tiwari. A note on the Dirichlet process, in *Essays in Honor of C.R. Rao*, 89–103, G. Kallianpur, P.R. Krishnaiah, and J.K. Ghosh eds., North-Holland, Amsterdam, 1982.
50. D. Basu and C. B. Pereira. Conditional independence in statistics, *Sankhyā*, Ser. A, 45, 324–337, 1983.
51. D. Basu and C. B. Pereira. A note on Blackwell sufficiency and a Skibinsky characterization of distributions, *Sankhyā*, Ser. A, 45, 99–104, 1983.
52. D. Basu. *Statistical Information and Likelihood: A Collection of Critical Essays*, J. K. Ghosh ed., *Lecture Notes in Statistics*, 45, Springer-Verlag, New York, 1988.
53. D. Basu and C. B. Pereira. Blackwell sufficiency and Bernoulli experiments, *Rebrape*, 4, 137–145, 1990.
54. D. Basu. Learning statistics from counterexamples, *Bayesian Analysis in Statistics and Econometrics*, *Lecture Notes in Statistics*, 75, Springer-Verlag, New York, 1992.

Basu's Work on Randomization and Data Analysis

George Casella and Vikneswaran Gopal

1 Introduction

Sir R. A. Fisher put forward the idea that randomization is a necessary component of any designed experiment. It is accepted without question by most practitioners of statistics. Yet in the two papers

1. Basu, D. (1978) Relevance of randomization in data analysis. *Survey sampling and measurement* 267-292.
2. Basu, D. (1980) Randomization analysis of experimental data: the Fisher randomization test. *Journal of the American Statistical Association* **75** (371) 575-582.

Basu wonders out loud if randomization is really that important. He argues his case in the context of survey sampling, and when analyzing data using a randomization test.

In [1] Basu covers the survey sampling situation, and the randomization test is the topic of [2]. Although he acknowledges that there is a place for randomization in surveys (see Section 4 of [1]), his belief is the opposite for the randomization test. It is important to note the difference between the randomizations discussed in the two papers. In [1], Basu focuses on *prerandomization* - how to pick a sample from a sampling frame, and how it affects the subsequent analysis of data. In [2], the focus is on the *randomization test*, which was first introduced by Fisher. The two types of randomization are intricately linked, as the first provides a basis for the second. In essence, Basu argues that the absence of prerandomization does not make a dataset worthless, however, because of the total dependence of the randomization test on prerandomization, a randomization test is never valid.

In this commentary, we provide a short summary of Basu's ideas on randomization. That he did not write a great deal more on this topic is, in his own words, "a measure of my diffidence on the important question of the relevance of randomization at the data analysis stage".

G. Casella (✉)

Distinguished Professor, Department of Statistics, University of Florida, Gainesville, FL 32611. Supported by National Science Foundation Grants DMS-0631632 and SES-0631588
e-mail: casella@stat.ufl.edu

V. Gopal (✉)

PhD Candidate, Department of Statistics, University Florida, 102 Griffin-Floyd Hall, Gainesville, FL 32611
e-mail: viknesh@stat.ufl.edu

A. DasGupta (ed.), *Selected Works of Debabrata Basu*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4419-5825-9_1, © Springer Science+Business Media, LLC 2011

2 Survey Sampling

The main question posed in [1] is about how to analyze the data generated by a survey or experiment. With a series of examples, Basu demonstrates the disadvantages of a frequentist approach, which is closely tied to the exact sampling plan used.

We highlight one of his more striking examples here. Suppose we have a well-defined finite population \mathcal{P} , consisting of individually identifiable objects called units. We can perceive of \mathcal{P} as the set $\{1, 2, \dots, N\}$. Corresponding to each $j \in \mathcal{P}$, there exists an unknown quantity Y_j . The goal of sampling is typically to estimate some function of (Y_1, Y_2, \dots, Y_N) . The method of achieving this is through a sampling plan \mathcal{S} , by which we mean a set of rules, following which we can arrive at a subset s of \mathcal{P} .

Suppose also that we have a machine that produces $N = 100$ units in a day. However, it is possible for the machine to malfunction at some point, after which *it only produces defective products*. Using the definitions of the previous paragraph, Y_i take on values 1 or 0, depending on whether they are defective or functioning. The aim is to estimate $\theta = \sum Y_i$, the total number of defective products manufactured in a day, by drawing a sample from the N units.

Randomization is injected into the experiment through the choice of the sampling plan. Should we draw a simple random sample? Maybe a stratified sample? Whatever \mathcal{S} we chose, the result of drawing a sample of size 4 would be recorded as, say,

$$Y_{17} = 0, Y_{24} = 0, Y_{40} = 1, Y_{73} = 1.$$

What then would a non-Bayesian statistician do with this data? To apply a randomization analysis, the probability of this sample with respect to the sampling scheme would have to be computed. A complicated enough scheme might even preclude this. A Bayesian, on the other hand, would observe that regardless of the sampling scheme applied, we know that $61 \leq \theta = \sum Y_i \leq 76$, since the first defective occurred in the set $\{25, 26, \dots, 40\}$. Moreover, the likelihood function would be constant over the set $\{61, 62, \dots, 76\}$ and we simply base all inference on this. Thus, Basu is invoking the Conditionality and Likelihood Principles to conclude that at the data analysis stage, the exact nature of the sampling plan is not important. He also points out that it in this case a sequential purposive sampling plan would serve our need better.

Notice that the example has been carefully set-up so that the non-Bayesian would be somewhat confused. For example, θ as it is presented here, would not be viewed as a parameter in classical statistics. But Basu, being a Bayesian, does not make a distinction between a random variable and a parameter. The way Basu presents the problem, a Bayesian analysis offers itself as the most natural thing to do. Such an approach avoids the need for obsessive randomization, and extracts information from the sample obtained rather than basing inference on samples that were not drawn.

3 The Test of Randomization

With [2], Basu places the randomization test under his microscope. At the end of his analyses, he concludes that he is unable to justify the use of the randomization test.

In the initial segments of the paper, Basu presents a version of the Fisher randomization test as a precursor to nonparametric tests such as the sign test and the Wilcoxon signed-rank test. Following that, he speculates that Fisher lost interest and belief in the randomization test. The final section of the paper is the most entertaining one. It contains a fictitious conversation between a scientist, a statistician and Basu himself. The three individuals discuss the randomization test introduced by Fisher in Chapter III (Section 21) of [3].

The scientist wishes to test whether a new diet is an improvement over the standard one. 30 animals are divided into 15 homogeneous pairs and from each pair, the scientist selects one subject for the treatment and the other one for the control. The response is the amount of weight gained in a subject after, say, 6 weeks. The data for each pair are recorded as (t_i, c_i) . Suppose that for this particular experiment, the scientist records that $t_i - c_i > 0$ for all i , and that $T = \sum_i (t_i - c_i)$ is a large positive number.

H_0 states that the new diet makes no difference to the response. If this null hypothesis were true, it would mean that any difference in response for the i -th pair must have been caused by “nuisance” factors such as subject differences. Under H_0 then, the significance level of the observed statistic would be $\Pr(T' \geq T | H_0) = (1/2)^{15}$, assuming that all treatment assignments were equally likely. Basu takes the position that the randomization test should be applicable even if the randomization were not so. Specifically, he asks why the randomization test yields a different significance level if a biased coin were used to assign treatments within each pair. This apparent breakdown of the methodology is one of the reasons that leads Basu to recommend that the test not be used.

In introducing the article [2] in an earlier collection [4], Basu poses similar questions with regard to the famous tea-tasting experiment, which was also introduced by Fisher in [3].

A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or tea infusion was first added to the cup. . . . Our experiment consists in mixing eight cups of tea, four in one way and four in the other, and presenting them to the subject for judgement in a random order.

The subject knows that there are 4 cups of each kind, and her task is to pick out the two groups of cups. Fisher argued that under the (null) hypothesis that the lady does not have the ability to distinguish, if we use the number of matches between the true grouping and the lady's grouping as a statistic, the significance level of a perfect grouping by the lady is given by

$$\Pr(T \geq 8 | H_0) = \frac{1}{70}$$

Basu asks a series of questions of this approach:

Why randomize? Was it because we wanted to keep the Lady in the dark about the actual layout? But then, why did we have to tell the Lady that there were exactly four cups of each kind in the layout and that all the 70 choices were equally likely? Why couldn't we choose just any haphazard looking layout and keep the lady uninformed about the choice? But then, how could we compute the significance level? Instead of randomizing over the full 70 point set, couldn't we randomize over a smaller, say, 10 point set of haphazard arrangements? How can we explain that in that case the same data (x, y) with $T = 8$ will be associated with a significance level of $1/10$? Why are we holding the Lady's response y as fixed and playing this probability game with the ancillary statistic x ?

Fisher went some way to explaining some of these questions when he described the purpose of randomization, in Chapter II (Section 10) of [3].

The element in the experimental procedure which contains the essential safeguard, is that the two modifications of the test beverage are to be prepared “in random order”. . . . The phrase “random order” itself, however, must be regarded as an incomplete instruction, standing as a kind of shorthand symbol for the full procedure of randomization, by which the validity of the test of significance may be guaranteed against corruption by the causes of disturbance which have not been eliminated.

Fisher says that randomization is what solves the problem of not being able to hold every single factor other than the treatment condition constant. The only solution is to ensure that every treatment allocation has an equal chance of occurring. Any other probability distribution on the treatment assignments could introduce a confounding factor.

For example, suppose that in the diet experiment, a coin that yields a treatment assignment of (t_i, c_i) with probability $1/4$ rather than $1/2$ is used. Then this is clearly against the requirement spelt out by Fisher, because for example, a treatment allocation with 15 (c_i, t_i) 's is more likely than one

with 15 (t_i, c_i) 's. If the animals were kept in a pen divided into 30 cells in a 15×2 arrangement, it is possible that the cells on the left obtained more sunlight and hence caused the animals to gain more weight. This would make the control treatment look good, since more animals on the left would be assigned that treatment.

The validity of the randomization test depends on the prerandomization being carried out properly, which requires that *all* treatment assignments be equally likely. Granted, Fisher never explicitly stated that when he said randomize, he meant for us to impose a uniform distribution on the treatment allocations. However, even if he had made his intentions explicit, would Basu have let him so lightly? We think not. Unless Fisher gave a sound mathematical argument as to why all treatment allocations should be equally likely, Basu's points would still be relevant and fair.

4 Final Thoughts

It is a tremendous joy to read Basu's papers. He presents his view in such a convincing manner that one almost feels ashamed at believing anything to the contrary. However, it is clear from the final sections of [1] that he does not suffer terribly from tunnel vision; he dissects his own arguments and tries to come up with explanations for possible criticisms of his points. It is also evident that he welcomes a good debate. The discussions at the end of [2] provide ample evidence for this. The banter between Basu and Kempthorne in particular, is fit for a comedy (be sure not to miss it!).

(Re-)Reading Basu's papers, which combine an inimitable style of writing with impactful examples, is an educating, enlightening and entertaining experience. At best, we question our assumptions and beliefs, which leads us to gain new insights into classical statistical concepts. At "worst", we embark on a journey to becoming Bayesian.

References

- [1] Basu, D. (1978) Relevance of randomization in data analysis. *Survey sampling and measurement* 267–292.
- [2] Basu, D. (1980) Randomization analysis of experimental data: the Fisher randomization test. *Journal of the American Statistical Association* **75** (371) 575–582.
- [3] Fisher, R.A. (1966) The design of experiments.
- [4] Ghosh, J.K. (1988) Statistical Information and Likelihood, Lecture Notes in Statistics **45**

Basu on Ancillarity

Philip Dawid

1 The origins of ancillarity

The term “ancillary statistic” was introduced by R. A. Fisher (Fisher 1925) in the context of maximum likelihood estimation. Fisher regarded the likelihood function as embodying all the information that the data had to supply about the unknown parameter. At a purely abstract level, this might be regarded as simply an application of the sufficiency principle (SP), since as a function of the data the whole likelihood function (*modulo* a positive constant factor — a gloss we shall henceforth take as read) is minimal sufficient; but that principle says nothing about what we should do with the likelihood function when we have it. Fisher went beyond this stark interpretation, regarding the actual form of the likelihood function as itself somehow embodying the appropriate inference. In some cases, such as full exponential families, the maximum likelihood estimator (MLE) is itself sufficient, fully determining the whole course of the likelihood function; but more generally it is only in many-one correspondence with the likelihood function, so that two different sets of data can have associated likelihood functions whose maxima are in same place, but nevertheless differ in shape. Initially, for Fisher, an *ancillary* statistic (from the Latin “ancilla”, meaning handmaiden) denoted a quantity calculated from the data which “lent support” to the MLE, by providing additional information about the shape of the likelihood function, over and above the position of its maximum — for example, higher derivatives of the log-likelihood at the MLE. If we regard the spikiness of the likelihood function as telling us something about the (data-dependent) precision of the MLE, we might select a suitable ancillary statistic to quantify this precision: this appears to have been Fisher’s original motivation. According to this understanding of an ancillary statistic as describing the shape of the likelihood function, it is necessarily a function of the minimal sufficient statistic. Ideally, the MLE together with its handmaiden would fully determine the likelihood function, the pair then constituting a minimal sufficient statistic.

Fisher (1934) then considered the working out of these general concepts in the special case of a location model, where the MLE fully determines the location of the likelihood function, but is entirely uninformative as to its shape; while the configuration statistic, *i.e.*, the set of pairwise differences between the observations, constitutes an ancillary statistic, fully determining the shape of the likelihood function, but uninformative about its location. For this model (though, for Fisher’s original definition, not necessarily more generally) it is also true that the distribution of this ancillary statistic is entirely independent of the value of the unknown location parameter; furthermore, the conditional distribution of the maximum likelihood estimator, given the configuration, has a density

P. Dawid (✉)
Statistical Laboratory, Centre for Mathematical Sciences, CB3 0WB, UK
e-mail: apd@statslab.cam.ac.uk

that has the same shape as the likelihood function. At a certain point, Fisher decided that it was such properties, rather than his original handmaiden conception, that were of crucial general importance, and from that point on the word “ancillary” was used to mean “having a distribution independent of the parameter”. Associated with this was the somewhat vague idea of a “conditionality principle” (CP), whereby it is the conditional distribution of the data, given the ancillary statistic, that is regarded as supplying the appropriate “frame of reference” for determining the precision of our estimate. As a simple example lending support to this principle, suppose we first toss a fair coin, and then take 10 observations if it lands heads up, or 100 if it lands tails up. The coin toss does not depend upon the parameter (it is ancillary in the revised sense, although not necessarily in the original sense), and so cannot, of itself, be directly informative about it; but it does determine the precision of the experiment subsequently performed, and it does seem eminently sensible to condition on the number of observations actually taken to obtain a realistic measure of realised precision.

At an abstract level, CP can be phrased as requiring that any inference should be (or behave as if it were) conducted in the frame of reference that conditions on the realised value of an ancillary statistic. One can attempt to draw analogies between this CP and the sufficiency principle, SP, which tells us that our inference should always be (or behave as if it were) based on a sufficient statistic. But is important to note that in either case there may be a choice of statistics of the relevant kind, and we would like to be able to apply the principle simultaneously for all such. Considering first the case of sufficiency, suppose T_1 is sufficient and, in accordance with SP, we are basing our inference on T_1 alone. If now T_1 is a function of T_2 , then T_2 is also sufficient: but the property that our inference should be based on T_2 alone is automatically inherited from this property holding for the “smaller” statistic T_1 , so we do not need to take any explicit steps to ensure this. In particular, if we can find a *smallest* sufficient statistic T_0 , a function of any other sufficient statistic, then basing our inference on T_0 will automatically satisfy SP with respect to *any* choice of sufficient statistic. It is well known that, subject only to mild regularity conditions, such a smallest (“minimal”) sufficient statistic can generally be found. Hence it is pretty straightforward to satisfy SP simultaneously with respect to every sufficient statistic: simply base inference on the minimal sufficient statistic.

The case of ancillarity appears very similar, though with the functional ordering reversed. Suppose S_1 is ancillary, and, in accordance with CP, we are basing inference on the conditional distribution of the data, given S_1 . If now S_2 is a function of S_1 , then S_2 is also ancillary; and the property that inference is conditioned on S_2 is automatically inherited from this property holding for the “larger” statistic S_1 . This analysis suggests that — in close analogy with the case of the minimal sufficient statistic — we should aim to identify a *largest* ancillary statistic S_0 , of which every ancillary statistic would be a function. Then conditioning on S_0 would automatically satisfy CP, simultaneously with respect to every choice of ancillary statistic.

2 Enter Basu

The above analysis appears unproblematic, and might be thought to make a compelling case for always conditioning on the largest ancillary statistic — an apparently straightforward enterprise. But then along comes Basu, and suddenly things are not so clear!

Basu presented theory and counter-examples to show that in general there is *no* unique largest ancillary statistic, conditioning on which would allow us to apply the conditionality principle unambiguously. Typically there will be a multiplicity of ancillary statistics that are maximal, *i.e.*, cannot be expressed as a non-trivial function of any other ancillary; and in this case no single largest ancillary can exist. Even in what would seem to be the simplest special case, of two independent observations from the normal location model, having $X_i \sim \mathcal{N}(\theta, 1)$ ($i = 1, 2$), there is no largest ancillary: for any $c \in [-\infty, +\infty]$, the statistic S_c defined as $X_1 - X_2$ if $X_1 + X_2 > c$, $X_2 - X_1$ otherwise, is