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 "$\cdots$, and this is characteristic of Basu and one of the reasons (that) his papers are still so valuable; it does challenge the usual way $\cdots$. 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 $X_1, \cdots, X_n$ have a joint distribution $P_{n,\theta}$, $\theta \in \Theta$, $T(X_1, \cdots, X_n)$ is boundedly complete and sufficient, and $S(X_1, \cdots, X_n)$ 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 $X_1, \cdots, X_n$. 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 $T(X_1, \cdots, X_n)$ is boundedly complete and sufficient. Then, $S(X_1, \cdots, X_n)$ 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 $X_1, X_2, \cdots, X_n$ are iid $N(\eta, \tau^2)$ for some η, τ. Suppose $\overline{X}$ is the mean and s^2 the variance of the sample values $X_1, X_2, \cdots, X_n$. The goal is to prove that $\overline{X}$ 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 $X_i = \eta + \tau Z_i, 1 \le i \le n$, where $Z_1, \cdots, Z_n$ are iid $N(0, 1)$. Now,

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

Therefore, $\overline{X}$ and $\sum_{i=1}^n (X_i - \overline{X})^2$ are independently distributed under (η, τ) if and only if $\overline{Z}$ and $\sum_{i=1}^n (Z_i - \overline{Z})^2$ 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 $Y_1, Y_2, \cdots, Y_n$ from $P_\mu = N(\mu, 1)$. The joint density of $Y = (Y_1, Y_2, \cdots, Y_n)$ is a one parameter Exponential family density with the natural sufficient statistic $T(Y) = \sum_{i=1}^n Y_i$. And, of course, $\sum_{i=1}^n (Y_i - \overline{Y})^2$ 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* μ, $\sum_{i=1}^n Y_i$ and $\sum_{i=1}^n (Y_i - \overline{Y})^2$ 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 $X_1, X_2, \cdots, X_n$ are iid Exponential random variables with mean λ. Then, by transforming $(X_1, X_2, \cdots, X_n)$ to $\left(\frac{X_1}{X_1 + \cdots + X_n}, \cdots, \frac{X_{n-1}}{X_1 + \cdots + X_n}, X_1 + \cdots + X_n \right)$, one can show by carrying out the necessary Jacobian calculation that $\left(\frac{X_1}{X_1 + \cdots + X_n}, \cdots, \frac{X_{n-1}}{X_1 + \cdots + X_n} \right)$ is independent of $X_1 + \cdots + X_n$. We can show this without doing any calculations by using Basu's theorem.

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

Example 3 (A Weak Convergence Result Using Basu's Theorem) Suppose $X_1, X_2, \cdots$ are iid random vectors with a uniform distribution in the d-dimensional unit ball. For $n \ge 1$, let $d_n = \min_{1 \le i \le n} ||X_i||$, and $D_n = \max_{1 \le i \le n} ||X_i||$. 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 $0 \le u \le 1$,

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

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,

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

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 $k \geq 1$,

$$E[d_n]^k = E[D_n \rho_n]^k = E[D_n]^k E[\rho_n]^k$$

$$\Rightarrow E[\rho_n]^k = \frac{E[d_n]^k}{E[D_n]^k} = \frac{n!\Gamma\left(\dfrac{k}{d} + 1\right)}{\Gamma\left(n + \dfrac{k}{d} + 1\right)}\frac{nd + k}{nd}$$

$$\sim \frac{\Gamma\left(\dfrac{k}{d} + 1\right)e^{-n}n^{n+1/2}}{e^{-n-k/d}\left(n + \dfrac{k}{d}\right)^{n+\frac{k}{d}+1/2}}$$

(by using Stirling's approximation)

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

Thus, for each $k \geq 1$,

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

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

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

as $n \to \infty$.

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 $X_1, X_2, \cdots, X_n$, 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

$$E\left(I_{a \leq X_1 \leq b} \mid \overline{X}, s\right) = P\left(a \leq X_1 \leq b \mid \overline{X}, s\right)$$

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

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

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

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

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

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

where

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

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

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

$$\mathrm{Cov}(X_1 + \cdots + X_n, \overline{X} - M_n) = 0 \Rightarrow \mathrm{Cov}(\overline{X}, \overline{X} - M_n) = 0$$

$$\Rightarrow \mathrm{Cov}(\overline{X}, M_n) = \mathrm{Cov}(\overline{X}, \overline{X}) = \mathrm{Var}(\overline{X}) = \frac{1}{n}.$$

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 $Z_3, Z_4, \cdots, Z_m$ any other random variables such that $(Z_3, Z_4, \ldots, Z_m)$ is independent of (Z_1, Z_2). Then for any positive integer k, and an arbitrary measurable function g, $f^k(Z_1, Z_2)g(Z_3, Z_4, \cdots, Z_m)$ 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) $f(x) = \dfrac{1}{\pi} K_0(|x|)$, where K_0 denotes the Bessel K_0 function;

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

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

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

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 $Y = Y_1 + \ldots + Y_{50}$ 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, $\cdots$, iti," which means- "not this, not this, $\cdots$, 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.