

# 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,  $\theta$  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  $15 \times 2$  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**