ROLE OF RANDOMIZATION IN BAYESIAN ANALYSIS – AN EXPOSITORY OVERVIEW

by

Jayanta K. Ghosh
Purdue University and I.S.I.

Technical Report #05-04

Department of Statistics
Purdue University
West Lafayette, IN USA

May 2005

ROLE OF RANDOMIZATION IN BAYESIAN ANALYSIS – AN EXPOSITORY OVERVIEW

by

Jayanta K. Ghosh

Abstract

The evolving role of randomization in Bayesian Analysis as well as arguments for and against randomization is discussed. We note that Bayesian Analysis is moving towards a substantially reduced role for randomization, but it is a sample surveys area where more future work of this kind is needed.

1. Introduction

We provide a relatively non-technical overview of how the Bayesian approach to randomization has changed over time. The overview depends primarily on the cited papers of Basu (1988), Kadane and Seidenfeld (1999), Rubin (1978) and some extracts from Savage, quoted in Kadane and Seidenfeld (1999). These references are supplemented by more recent papers of Berry (2004) and Spiegelhalter (2004).

Most of the paper deals with the role of randomization in Bayesian analysis of clinical trials (section 5), but there is some discussion of random samples (section 4) and randomization tests (section 3). Section 6 contains concluding remarks.

2. Bayesian Views on Randomization – An Imaginary Conversation

We present an imaginary conversation by suitably arranging written views of the distinguished Bayesians mentioned in the Introduction. For example, the first two paragraphs below are from Savage (1961,1962), quoted at the beginning of the cited article of Kadane and Seidenfeld(1999).

Savage: "Applying the theory (of personal probability) naively one quickly comes to the conclusion that randomization is without value for statistics. This conclusion does not sound right...

The need for randomization presumably lies in the imperfection of actual people and,

perhaps, in the fact that more than one person is ordinarily concerned with an investigation."

Basu: "The mere artifact of randomization cannot generate any information that is not there already. However, in survey practice situations will arise where it will be necessary to insist upon a random sample. But this will be only to safeguard against some unknown biases.

The inner consistency of the Bayesian point of view is granted...(But) who can be a true Bayesian and live with thousands of parameters. Survey statistics is more an art than a science."

Kadane and Seidenfeld: "Many have criticized randomization analysis for failing the likelihood principle: see Basu (1981). This reply to Fisher, Kempthorne and others, who defend randomization analysis, is we think, what Savage means by the "naive" Bayesian rejection of randomization.... When only one decision maker is relevant we would not randomize."

After considering Bayesian alternatives to randomization, Kadane and Seidenfeld go on to observe: "Thus even a sophisticated (rather than naive) Bayesian defense of randomization....fails to establish it as a sine qua non of sound experimental methods."

Rubin: "Classical randomized designs stand out as especially appealing mechanisms designed to make inference for causal effects straightforward by limiting the sensitivity of a valid Bayesian Analysis."

Kadane and Seidenfeld: "Randomization is one way to accomplish this but it is not unique in having the virtue. We join with Savage and many others, however, in his respect for randomization as a statistical tool for enhancing interpersonal communication."

At first reading there seems to be major differences. A second or third reading of the original papers shows a basic unity along with important differences. As Kadane and Seidenfeld (1999) point out, there are two kinds of experiments - experiments to learn and experiments to prove. The second kind of experiment involves at least two decision makers, of which one is the Bayesian who designed the experiment and analyzed the data and the other represents other Bayesians who have to be convinced. There, i.e, in the second set

of experiments, randomization helps ensure trust in the analysis. Most Bayesians would agree with the sophisticated view expressed by Kadane and Seidenfeld (1999) in the last two extracts from their paper.

This paper is about learning - my learning. It's not a paper to prove. Hopefully, it will help other Bayesians who haven't made up their mind about randomization but would like to do so.

We explore these basic issues in some more detail below.

3. Randomization and Permutation Tests

Randomization appears in classical inference and design of experiment – either to reduce the effect of unknown biases by some form of averaging over randomizations or to justify permutation tests by introducing exchangeability. Randomization also appears in classical inference via devices like randomized tests.

While the idea of using randomization to cope with unknown biases seems reasonable, each of the other two applications is contrary to principles of Bayesian Analysis with a single decision maker. The single decision maker is supposed to maximize his posterior expectation of utility. He may, but need not, randomize if two or more actions are optimal.

If randomization is used at the design stage, it violates the likelihood principle since the likelihood is the same no matter what randomization is used at the design stage to draw a sample from a population or allocate different treatments to different sampling units. In particular, randomization will not have any effect on the posterior. Therefore any method of analysis, in which randomization plays an important role, can't be based solely on likelihood or posterior. We examine below the logical difficulties associated with such tests. Logical difficulties arise from the fact that certain observed or observable quantities are ignored even though they may contain relevant information.

We consider Example B of Kadane and Seidenfeld (1999) on the permutation t-test for the difference of means of two normal populations with possibly unequal variances. Let (x_i,y_i), $i = 1, 2 \ldots, n$, be n observations. Use a random permutation of y's to generate data with a new pairing, (x_i, y_{j_i}), $i = 1, 2 \ldots, n$. The differences $\{z_i = (x_i - y_{j_i}), i = 1, 2 \ldots, n\}$

are equally likely over the permutations. Let

$$t = \sqrt{(n-1)}\bar{z}/S_z,$$
$$S_z^2 = \sum_1^n (z_i - \bar{z})^2/(n-1).$$

For large n, the permutation distribution of t is approximately N(0,1) under the null hypothesis of same mean of X and Y. When we use this distribution obtained by averaging over $n!$ permutations, we ignore the random pairing and hence, S_z, which are not ancillary and contain information.

Kadane and Seidenfeld show how odd the argument is by examining the case of $n = 2$. There are only two values of the t-statistic since $n! = 2$. The values are

$$[(x_1 + x_2) - (y_1 + y_2)]/[|x_1 - x_2| - |y_1 - y_2|]$$

and

$$[(x_1 + x_2) - (y_1 + y_2)]/[|x_1 - x_2| + |y_1 - y_2|].$$

It is clear that the first value of $|t|$ is larger, i.e, more significant because of the smaller value of the ignored S_z in the denominator of t. On the other hand, the conditional distribution of t given S_z is degenerate. Thus the permutation t distribution has an unclear logical status. Similar objections apply to permutation t or F tests for randomized block experiments. None of these are acceptable to a Bayesian.

A second way in which randomization may enter analysis is via a randomized decision function which, given data, puts a probability distribution over the space of actions and chooses an action at random. This would be acceptable to a Bayesian only if this distribution sits on actions that minimize the posterior risk.

4. Randomization and Sample Surveys

Random sampling from a finite population with or without replacement, is supposed to ensure exchangeability of x_i's that are sampled (and hence seen) and remaining x_i's that are not sampled (and hence not seen). Exchangeability leads to a natural estimate of population total, based on the assumption that the mean of the unseen x_i's can be

estimated by the mean of seen x_i's,

$$\sum_{i\in \text{ sample}} x_i + \left(\frac{1}{n} \sum_{i\in \text{ sample}} x_i\right)(N-n) = N\bar{x}$$

Here $N - n$ = number of unseen units.

The catch here is that labels of the units are ignored. If labels are not ignored exchangeability doesn't hold.

Even more difficult to justify are pps (probability proportional to size) sampling designs and the corresponding Horvitz-Thomson (HT) estimate, which are routinely used but remain repugnant to a Bayesian. The HT estimate for the population total is

$$\sum_{i=1}^{n} x_i/p_i$$

where p_i = probability of choosing the ith unit. In case $p_i = \frac{1}{n}$ this reduces to the estimate for random samples with equal probability. Basu (1988) presents a hilarious example of how absurd the estimate can be if the p_i's are poorly chosen. Basu's Circus Example: A circus owner wants a rough estimate of the total weight of fifty adult elephants. He chooses Sambo as typical (based on past measurements), measures Sambo's current weight and multiplies by fifty. That is his estimate.

The circus statistician is "horrified" by this ad hoc estimate. He chooses a sampling plan "that allots a selection probability 99/100 to Sambo and 1/4900 to the rest." The object is to choose Sambo with a high probability and yet have a valid estimate of variance. Naturally, Sambo is selected and the statistician produces the HT estimate, namely, (Sambo's weight) which is approximately about (1/50)th of the owner's rough estimate. (You should read Basu's own description of the problem.) This is a clever example of an unbiased estimate with a very large variance. So even a classical statistician would be shocked. Basu seems to be saying that this is what would often happen if one uses a sampling design with unequal probabilities. If a random sample with equal probabilities were used, the result would still be bad compared with the owner's common sense estimate but not as absurd as above.

The common sense estimate makes sense to a subjectivist Bayesian. He too may want to choose Sambo as typical, based on the prior information consisting of the past

measurements. But what would be a simple, generally applicable Bayesian procedure for drawing a sample based on available information? What would be a Bayes estimate for the population total and an estimate of its variance? The parametric super population based approaches have not survived competition with the design based approaches. It's possible that one needs to be a nonparametric Bayesian, even in these relatively simple examples, i.e, sampling problems with small sample size.

We consider another simple example to indicate that we seem to lack simple Bayesian alternatives to randomization.

A Presidential candidate in the U.S. wants an estimate of the proportion of voters who will vote for him. This is an experiment to learn, the usual sample size is small (1000 to 1200). Can a Bayesian do better than a simple random sample?

If the sample size is quite large, one ought to be able to come up with good Bayesian nonparametric answers. But I haven't seen any.

5. Bayesian Approach to Clinical Trials with or without Randomization

Most of this section is based on Rubin (1978) and Kadane and Seidenfeld (1999). Rubin assumes there is a target population and the data arise by random sampling from the target population and random experiment of treatments (as in a randomized block assignment).

Rubin notes that "intuitively the causal effect of one treatment relative to another for a particular experimental unit is the difference between the result, if instead, the unit had been exposed to a second treatment." Clearly, both observed and unobserved random variables are relevant.

Clinical trials are experiments to prove, with various groups, namely, the statisticians, doctors and FDA, who are to be convinced by the study undertaken by the pharmaceutical company. In addition, there are ethical concerns relating to patients.

Data Table

	Covariate (X)	Which Treatment (W)	Post-Treatment Value (Y)
1.	(X_{11}, X_{01})	W_1	$Y_{11}, \ldots, Y_{T1}$
2.	(X_{02}, X_{02})	W_2	$Y_{12}, \ldots, Y_{T2}$
$\vdots$			
N.	...		

Here,

X_{11} = observed part of x for first unit.

X_{01} = unobserved part of x for first unit.

Y_{11} = value of Y if unit 1 gets the first treatment.

Y_{T1} = value of Y if unit 1 gets the T th treatment.

N = # units in target population.

$W_i = 0$ indicates ith unit was not selected and so not exposed to any treatment.

W_i = t indicates ith unit was selected and exposed to treatment t.

For the ith unit with $W_i = t$, $(Y_{1i}, \ldots, Y_{t-1,i}, Y_{t+1,i}, \ldots)$ = unobserved $\left(Y_{(0)}\right)$, Y_{ti} = observed $\left(Y_{(1)}\right)$.

Model for data:

$\left(X_{(1)}, Y_{(1)}, W\right)$ are observed random vectors. $\left(X_{(0)}, Y_{(0)}\right)$ are unobserved random vectors. Joint density is:

$$f(X, Y|\Pi)k(W|X, Y, \Pi)$$

where Π is what Rubin calls parameters in the model, the factor k models assignment and the first factor is:

$$f(X, Y|\Pi) = \prod_{i=1}^{n} f(X_i, Y_i|\Pi)$$

The object is to make inference on Π.

Definition: The assignment mechanism $k(W|X,Y,\Pi)$ is ignorable if k depends on (X,Y,Π) only through the observed part $\left(X_{(1)},Y_{(1)}\right)$.

Rubin remarks as follows, "The more involved the assignment mechanism, the more complex must be the recording mechanism if the assignment mechanism is to be ignorable".

Ignorability has an important implication. In general, i.e, without ignorability,

$$P\{Y_{(0)}|X_{(1)},Y_{(1)},W\}=\frac{\int\int k(W|X,Y,\Pi)f(X,Y|\Pi)p(\Pi)d\Pi dX_{(0)}}{\int\int\int k(W|X,Y,\Pi)f(X,Y|\Pi)f(X,Y|\Pi)p(\Pi)d\Pi dX_{(0)}dY_{(0)}}.$$

Under ignorability the factor "k" comes out of the integrals in the numerator and denominator and so gets canceled. This implies $k(W|X,Y,\Pi)$ need not be modeled.

We can now list the advantages of randomization as suggested by Rubin.

1. It leads to ignorable k. Thus k need not be modeled, nor would one have to model unobserved values in terms of observed values. This makes the Bayesian analysis robust.

2. It "yields data having more than one treatment condition for any distinct value of covariate." This also implies robustness. Rubin points out if two units with identical $X_{(1)}$ represent two treatments, then randomization must have been used. This allows balancing of covariates for different treatments and hence achieves robustness in modeling effect of covariates.

Various ignorable alternatives to randomization have been suggested in recent years. One of these, due to Kadane and Sedransk, has been discussed by doctors, lawyers, and philosophers and implemented at Johns Hopkins, vide Kadane (1996) and Kadane and Seidenfeld (1999). The main features are summarized below.

1. Appoint a small number of experts on the disease and treatments.

2. The group chooses a single indicator of outcome "of most reasonably of concern to a patient."

3. The group agrees on a few diagnostic variables.

4. Each expert's opinion is elected as a dynamic probabilistic function A(3) about the

outcome indicator. This may involve modeling.

5. Based on (4), each expert has a preferred treatment as a function of the diagnostic variables. Given the values of these variables, determine the preferred treatments and choose one at random.

This is ignorable and more ethical than randomization. Also it would not be easy to deviate from this protocol. As observed by Kadane and Seidenfeld (1999), "The desire that the analysis of an experiment not involve a judgement of the motivation of the experimenter seems natural for science. ...This consideration independent of the prior and the utility function of the experimenter can be conducted under any design that puts the experiment on "automatic pilot" once it is started ... Randomization is one way to accomplish this but it is not unique in having this virtue."

Can such protocols yield data which are as robust as those arising from experiments with randomization? Berry (2004) argues with considerable theoretical, and numerical evidence that we learn just as much. Of course, it would be good to subject real data from new studies to critical evaluation. Berry also points out that ethical considerations for patients may increase the number of volunteers and thus lead to more reliable studies. Apparently, FDA has accepted some Bayesian recommendations. Other studies of the same kind include Spiegelhalter(2004), who explores many other aspects of health care and Christen et al. (2004), who provide the technical details for an implementable Bayesian protocol.

6. Concluding Remarks

Our review focuses on three applications of randomization – permutation tests, clinical trials and sample surveys, – all of which disturb Bayesians. The first of these, namely, permutation tests are simple, but the simplicity comes with a price, one has to ignore part of the information. The second application, namely, clinical trials seem to be moving away from randomization to well planned, ethical, but more complex alternatives. In the case of the third application of randomization, namely, design based sample surveys, Basu(1988) had pointed out various logical difficulties. However, except for the use of post-sampling stratification, small area estimation and ingenious similar ideas, the basic philosophical structure of survey sampling has not apparently changed much, see, for example, Rao(1999)

and Ghosh (1999). This remains a promising area neglected by Bayesians.

References

Basu, D. (1988) Statistical Information and Likelihood, – A Collection of Critical Essays by D. Basu, editor J.K. Ghosh, Springer, New York.

Berry, D. A. (2004) Bayesian Statistics and the Efficiency and Ethics of Clinical Trials, *Statistical Science*, 19, 175-187.

Christen, J.A. Muller, P., Wathen, K. and Wolf, J. (2004) Bayesian Randomized Clinical Trials: A Decision Theoretic Sequential Design, *The Canadian Journal of Statistics,* 32, No.4, 1-16.

Ghosh, J.K. (1999) Discussion of "Some Current Trends in Sample Survey Theory and Methods", *Sankhyā* Ser B, 61, 36-40.

Kadane, J.B. (1996 Ed) Bayesian Methods and Ethics in a Clinical Trial Design, J. Wiley and Sons, New York.

Kadane, J.B. and Seidenfeld, T.(1999) Randomization in a Bayesian Perspective, in Rethinking the Foundations of Statistics, J.B. Kadane, M.J. Schervish, T. Seidenfeld, Eds, Cambridge University Press, Cambridge, U.K.

Rao, J.N.K. (1999) Some Current Trends in sample Survey Theory and Methods, *Sankhyā* Ser B, 61, 1-25.

Rubin, D. (1978) Bayesian Inference for Causal Effects: The Role of Randomization, *Ann Statist.*, 6, 34-58.

Savage, L.J. (1961) The Foundations of Statistics Reconsidered. Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I, 575–586, J. Neyman, Ed., Univ. California Press, Berkeley, Calif.

Savage, L.J. (1962) Subjective Probability and Statistical Practice, in The Foundations of Statistical Inference. M.S. Bartlett, Ed., Methuen, London.

Spiegelhalter, D.J. (2004) Incorporating Bayesian Ideas into Health-Care Evaluation, *Statistical Science*, 19, 156-174.