A Bayesian Guide to the Unknown

by

I. Richard Savage

FSU Statistics Report M 249
OSR Technical Report No. 72

January 1973
The Florida State University
Department of Statistics
Tallahassee, Florida

A Bayesian Guide to the Unknown

by

I. Richard Savage
The Florida State University
Tallahassee, Florida 32306

0. Introduction

Statistics: A Guide to the Unknown is a collection of 44 essays describing applications of statistics. The intended audience was the non-specialist in probability and statistics, such as, all people associated with secondary education. Even the editors have noted that these essays are useful in university education. I believe the professional statistician will find the Guide interesting and challenging.

In this review I first give a rough guide to the Guide (Section 1). This is followed by a partial list of the challenges found in the essays (Section 2). Finally, I look as a Bayesian at the essays (Section 3).

1. Something for everyone

The essays include many topics of varying importance, such as, can famous people postpone their deaths until a birthday (by David P. Philips), sampling versus auditing of waybills to increase profit (by John Neter), or the projection of future populations (by Nathan Keyfitz).

The influence of the editors is sensed in several ways. They worked very hard to invite outstanding authorities for each essay. The essays are between five and thirteen pages in length with an average length of about nine pages. In most essays there is a statement of the substantive problem, techniques for finding solutions, data sources, conclusions, and key references.

The Guide is not meant for systematic pursuit. The reader is encouraged to move from essay to essay without the necessity of reading connective editorial comment. (In a few places there are cross references between essays, for example, page 157, page 191, page 259, page 323 and page 359 all introduce "correlation" and all but the first refers to C. A. Whitney's essay on astrostastics for technical detail.)

The most apparent editorial influence is in the scholarly apparatus. The table of contents is arranged by subject. It includes authors, essay titles, and short abstracts. The basic divisions are man in his biologic (political, social, physical) world. The essays are classified by data sources: samples, available data, surveys and questionnaires, experiments, and quasi experiments. The essays are also classified by tools:

- Estimation
- Hypothesis testing
- Bayesian analysis
- Data analysis
- Tables, graphs and maps
- Percents and rates, standardization and adjustment
- Time series and index construction
- Probability and models
Sampling and randomization
Correlation and regression
Data reduction and classification
Tests and measurements
Forecasting and prediction
Decision making.

I fear that I have not given a vivid picture of the richness of the variety in the Guide. As a final effort I paraphrase a segment of the index:

Factor analysis
Failure
Fair-share coefficient
False negative and positive
Fame
Federalist Papers, The
Federal Reserve Board
Federal statistics
Felonies
Fertility
Field experiment
Field First Sergeant
Field trial
File organization
Filler words
Fillmore, Millard
Financial policy
Finnegan's Wake
Fin whales
Firms
Fisher, R. A.
Fixed-market-basket index
Flueck, J. A.
Food and Agricultural Organization.
2. The challenges

As a personal challenge, I often attempted to fill in technical details. One could go to the library for that detail. But if statistics has coherence one should be able to develop theory from a few good suggestions. Besides, if your students are reading these essays they will often wish instant guidance not involving your visiting the library.

These remarks are well illustrated by D. G. Chapman's "The plight of the whales". In his seven pages he tells us very much about the natural history of whales, the whaling industry problem, methods for measuring the whale populations, and policy consequences of knowledge of whale population dynamics. He gives vivid descriptions of capture-recapture, catch-per-day, and age analysis methods for measuring population sizes.

In fact, he gives several variations of capture-recapture. From the data Chapman says we can "infer" the population size. His presentation does not include sampling errors so that "infer" means the solution for an unknown in a deterministic proportion with three terms given. I can recall some of the relevant literature which allows me to formulate the objective sampling theory (hypergeometric) and objective modes of inference, such as, maximum likelihood estimation. Without visiting the library, I am not sure how much Bayesian analysis exists. Preparing
a report on "The Bayesian analysis of the capture-recapture method"
is a reasonable challenge.

In speaking of the catch-per-day method Chapman says "the results
obtained in this way must be combined with estimates obtained in other
ways". The formal combining of statistical data (estimates) is a major
problem that is seldom carefully developed. This might be an area for
library research rather than second guessing. It might be a call for
research!

Because the Guide is of high quality one is challenged to find the
few misprints. Page 346 of Hayward R. Alker, Jr.'s discussion of the
Lorenz curve has several difficulties: Instead of just "Figure 1" there
should have been labels for "Figure 1(a)" and "Figure 1(b)". The
sentence beginning "Let us draw..." refers to two lines which I thought
the author would have put into Figure 1(a) but apparently he really
wants me to do the drawing. He incorrectly says the diagonal in Figure
1(a) makes a 45° angle with the vertical. There are places where I
might not be able to say that an author is wrong although I strongly
disagree. Hanns Schwarz in "The use of subjective probability methods
in estimating demand" turned me off with his hypothetical atomic cabin
cruisers. (See Section 3.4 for a further critical remark on Schwarz.)

The major challenge of these essays is to read and interpret them
in a consistent manner from your viewpoint of the foundations of stat-
istics. The many authors and topics will tax any specific viewpoint.
If you easily move from objective to subjective or from inferences to decisions, then your job is relatively easy. Few of the authors make explicit foundational statements so that even the eclectic is challenged to state the intended foundations of most of the papers.

3. A Bayesian viewpoint

As I read the essays I tried to react to them from my viewpoint of the foundations of statistics: All probability statements are summaries of an (corporate) individual's beliefs about acceptable betting odds. Complete problems are in an economic setting and are solved by maximizing expected utility.

My viewpoint is not prominent in the editorial work. "Utility" and "subjective probability" do not appear in the index. (What happened to the title of H. Schwarz's essay?). The only Bayesian references in the classification scheme and index are to the essay "Deciding authorship" by Frederick Mosteller and David L. Wallace. The classification of methods cites seven essays under "decision making". There is a most appropriate index entry under "decision theory" which refers to B. W. Brown Jr.'s, "Statistics, scientific method, and smoking."

My lack of energy prohibits me from systematically covering all of the essays. So I will skip about. This will prevent me from showing you that I can be consistent in a Bayesian viewpoint but it will illustrate
the kinds of considerations I have gone through.

3.1 A good place to begin is the essay by Mosteller and Wallace on use of textual analysis for deciding authorship (between Hamilton and Madison) of the twelve *Federalist* essays which had a single but unknown author. The economic aspect of the problem is nearly absent; professors who draw conclusions which are later counter indicated might obtain adverse (monetary) rewards.

"A chief motivation" of the authors was "comparing several different statistical approaches". Hence, we have a meta-experiment which I will not attempt to analyze.

The detailed Bayesian work of Mosteller and Wallace was involved but roughly they did the following: Known writings of Hamilton and Madison were examined to detect differences in frequencies in word usages. By trial and error, differences were located, such as, Hamilton used "while" and Madison used "whilst". Substantial data were used to express these differences as sampling distributions. That is, Mosteller and Wallace acted as if they had no uncertainty about the (linguistic) behavior of Madison and Hamilton. In view of the large amount of data this seems appropriate; the Bayesian analysis would be robust. For the disputed essays the likelihood ratios were astronomically in favor of Madison.

How or what prior probabilities were used in the analysis is not hinted at. A plausible prior was

\[
\text{Pr(Hamilton wrote a specific set of } k \text{ out of 12 essays)} = p^k q^{12-k} \text{ where } 0 < p = 1 - q < 1 \text{ and } k = 0, 1, \ldots, 12.
\]
And reasonable choices of \( p \) would be \( 0.1 \leq p \leq 0.9 \). For these priors and other plausible models the likelihood ratios dominate the analysis and hence they conclude the probability is exceedingly near one that Madison wrote all of the papers. I wish that Mosteller and Wallace gave some of their thought on the choice of prior.

One can think of situations where the above analysis would be upset. What if it were physically impossible for one of Hamilton or Madison to do all of that writing? But those fellows were very industrious.

3.2 Robert G. Miller's "The probability of rain" starts with an example of the economic use of probability forecasts. He continues with the puzzling

"Probability distributions may be arrived at in various ways. The most common method is to use the human judgement of an experienced weather expert. He considers all the evidence and on the basis of his experience chooses a number that he thinks express the chance of rain. Another way of generating a probability distribution is to apply statistical methods to weather data stored in government archives. This essay describes a method for arriving at such a probability distribution based on the statistical evidence of past years."

Miller's procedure will be described below. But I give my interpretation now. He is a Bayesian with an economic leaning. Because there is much available data his judgement places much weight on past (relative) frequencies of events being near his probabilities. Most Bayesian analysts develop independently their sampling distributions and prior distributions.
Miller uses the technical mode of assessing his joint distribution of predictions and prediction variables.

The problem is to predict (dry, little rain, little snow, rain, snow) at Hartford, Connecticut at 7:00 AM during the next six hours with the use of 7:00 AM observations. (Miller's not including the season or current precipitation states makes the problem strange.) For some reason, the judgement of an experienced forecaster, he works with seven particular weather variables from twenty-five particular weather stations. He uses data for three years to develop his prediction device (joint distribution) and the data of a fourth year to test his device. He has a massive data pile $7 \times 25 \times 3 \times 365$ (plus the precipitation states), about 175,000 observations. Even so, there is not much historical data about snow for it does not snow often. Data reduction, selection, and transformation yielded two linear combinations (discriminant functions) involving 16 of the $7 \times 25$ weather variables from the stations. I gather it was assumed that the $7 \times 25$ variables have a multivariate normal distribution for each precipitation state; the historical data yielded estimates for the parameters; and the main reduction was done in terms of likelihood ratios. Finally, for each combination of discriminant function values, he found the relative frequencies of the precipitation states which are then used to generate prior probabilities.

When the above procedure is used at full scale the historical data would be nearly overwhelming in its relative frequency form. Of course
many subjective judgements are made in selecting variables to measure. It is possible that a good predictor could improve upon these predictions. For special situations such as shade tobacco harvesting time, one might find additional economic ways of making predictions. On the other hand, Miller's procedure is available at tolerable costs. (There is little evidence that public statistical problems have received sufficient analysis and discussion to encourage good solutions.)

3.3 It is difficult to assess probabilities which are very near 0. In my case ESP remains very unlikely in spite of much data. The essay "Deathday and birthday: an unexpected connection" by David P. Phillips contains evidence that famous people tend to postpone death until after a birthday. A precisely formulated hypothesis of this sort would be given small probability by me. So small that the data of this essay still leaves me very doubtful; not so small as my probability for ESP.

My reactions to the essay are: Get more data. Look for plausible alternative explanations. The first suggestion might cause hard feelings. If the new data is collected in the same manner from people with similar cultures it might contain the same traps which I suspect are in the current data, see the essay by B. W. Brown, Jr. Alternative explanations should always be sought. In this case we have: Famous people tend to be elderly and hence they die at ages where they should have a higher death rate a month after a birthday than a month before it. This effect is small but
it is in the correct direction. There is the possibility that famous people celebrate their birthdays too well and hence bring on excessive deaths. Mr. Phillips suggested the opposite, famous people hold death off to obtain their birthday celebrations. It would be hard to get data to distinguish between these hypotheses. I must say my hypothesis sounds more likely. Famous people are often required to travel and do formal extensive entertaining for a birthday. We may be killing our distinguished elders with kindness.

Phillips did get more data after he made his first observations. The additional data were supportive but not as strong as the original. He also found a dose response relationship: the more famous people had lower death rates prior to their birthdays than had the less famous.

3.4 Schwarz speaks of "subjective probability". He is interested in the probability of a particular action - the purchase of an atomic cabin cruiser. He would have done better to discuss "subject's probability". On page 214, he attempts to distinguish between "objective" and "subjective" probability but his only interest appears to be "objective". His essay does cover two important methodology topics: How to assess behavior from attitudes and the measurement of behavior by probabilities of occurrence.

3.5 I close with some miscellaneous comments.
3.5.1. You should do statistics in areas where you are an expert. I was surprised to learn that post operative death rates in legitimate hospitals could vary by a factor of 20, see page 20.

3.5.2. The "Drug screening" essay by Charles W. Dunnett is very Bayesian and decision theoretic.

3.5.3. Bayesian language appears in the writing of people who may not have thought about foundations: "thus [smoking] is likely to be its [bronchitis] cause" page 82, "After all, educated opinion is always the weighing of probabilities." page 94.

3.5.4. S. James Press in "Police manpower versus crime" talks like a Bayesian: "intervals of credibility", "degree of belief", "chance was very great (95%) that the true decrease...", and "highly unlikely in the future." I wonder how "frequentists" read Press?

3.5.5. Donald T. Campbell speaks of "quasi experiments" where social or other constraints prevent the use of randomization and experimental versus control groups. These are situations where expert opinion should be placed in a Bayesian framework and apparently Campbell does that. It should be noted that a pure Bayesian would not pay the price of randomization.
3.5.6. Edward R. Tufte in "Registration and voting" discusses a correlation analysis of proportions of registrants and voters in 104 large United States cities of 1968. The Bayesian attempts to give significance to this kind of study by applying the results to other situations. From the frequency viewpoint there is no sampling and hence no larger population to generalize about.

3.5.7. The primary purposes of many of the essays are to describe models, explain processes, or to present methods of measurement. Thus there is little Bayesian reaction. "Calibrating college board scores" by William H. Angoff is typical of this class. Of course his institution, Education Testing Service, has many Bayesian problems.

3.5.8. Robert Hooke considers decision problems with non monetary utility, for example, runs in baseball. Such examples show the catholic nature of decision theory.

3.5.9. Phillip J. McCarthy in "The consumer price index" points out the probability sample of communities is to be used over an extended time. This should puzzle the frequentist more than the Bayesian.

* * * * * * * * * * * * * * * *

So read, enjoy, and think.

* * * * * * * * * * * * * * *
14. Key Words

<table>
<thead>
<tr>
<th>Applied Statistics</th>
<th>Link A</th>
<th>Link B</th>
<th>Link C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Role</td>
<td>Role</td>
<td>Role</td>
</tr>
<tr>
<td></td>
<td>Wt.</td>
<td>Wt.</td>
<td>Wt.</td>
</tr>
<tr>
<td>Bayesian Statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Theory</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>