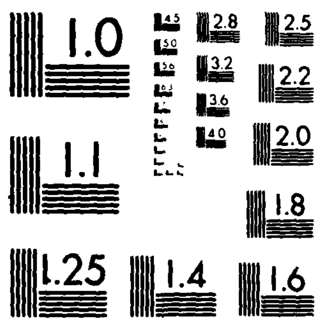
AD-A111 691 CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH CENTER F/G 12/1
ASSESSMENT OF SUBJECTIVE PROBABILITY.(U)
DEC 81 R E BARLOW AFOSR-81-0122
UNCLASSIFIED ORC-81-23 NL

END
DATE
FILMED
4-82
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

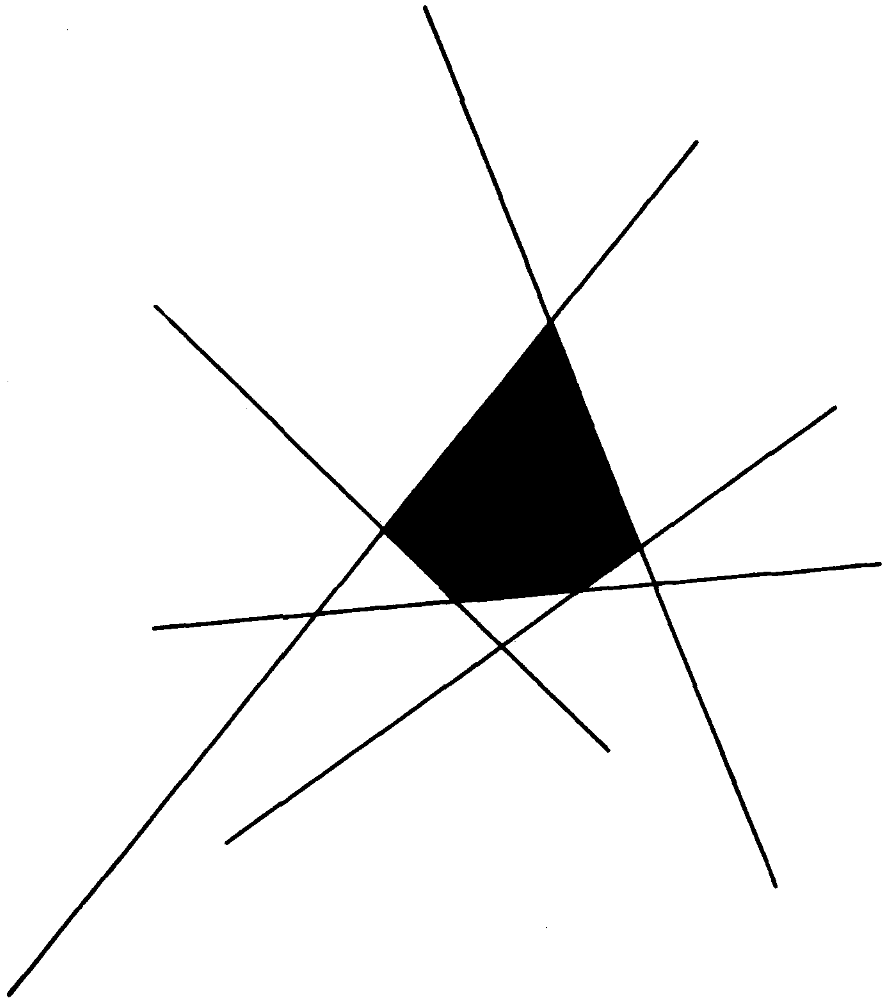
ORC 81-23
DECEMBER 1981

(12)

ASSESSMENT OF SUBJECTIVE PROBABILITY

by
RICHARD E. BARLOW

ADA111691



DTIC FILE COPY

OPERATIONS RESEARCH CENTER

82 03 05 020

DTIC ELECTE MAR 5 1982 D H

UNIVERSITY OF CALIFORNIA · BERKELEY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

ASSESSMENT OF SUBJECTIVE PROBABILITY†

by

Richard E. Barlow
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

DECEMBER 1981 ORC 81-23

†This research has been partially supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-81-0122 and the Mathematics and Statistics Division, Lawrence Livermore National Laboratory with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

This was a keynote address presented at the 1981 DOE Statistical Symposium held at Brookhaven, Long Island, New York, November 4, 1981. It will appear in the Proceedings of that Symposium.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC 81-23	2. GOVT ACCESSION NO. AD-A111691	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ASSESSMENT OF SUBJECTIVE PROBABILITY		5. TYPE OF REPORT & PERIOD COVERED Research Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Richard E. Barlow		8. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0122
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS United States Air Force Air Force Office of Scientific Research Bolling Air Force Base, D.C. 20332		12. REPORT DATE December 1981
		13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Probability Assessment Aids
Scoring Rules
Bayesian Paradigm

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ASSESSMENT OF SUBJECTIVE PROBABILITY*

Richard E. Barlow

Operations Research Center
University of California
Berkeley, California 94720

Abstract

The assessment of subjective probability is of great interest in risk analysis. Some aids for assessing subjective probability are surveyed. The connection with statistical inference and recent papers on statistical foundations are discussed.

I. Introduction

Recent DOE and NRC sponsored projects have involved the assessment of subjective probability. For example, the Seismic Safety Margins Research Program at Lawrence Livermore National Laboratory concerning nuclear power plant safety has involved assessing expert opinion relative to the strength of critical components (George and Mensing[1]). Combining expert opinion was necessary due to the lack of an experimental data base.

EPRI, Bechtel and others have been interested in coal fired electrical power plant availability. Data on rare failure events are again difficult to obtain. Especially in designing a new power plant, subjective failure probability assessments are required.

A third area of experimental research where subjective assessments are required is that of extrapolation procedures for late effects of radiation at low doses (P. G. Groer and R. E. Barlow[2]). The human health effects of low doses of ionizing radiation have been the subject of much controversy. A recognition of the inevitable subjective assessments involved might have mitigated some of this controversy.

Problems of the above kind are not about to go away. Instead, they will in all likelihood become even more important and pressing in future years.

How are these probability assessments to be made? It is easy to criticize the current literature concerning probability assessment methodology and I disagree with many of the methods used. However, some groups of people have been assessing probability for a long time. Examples are

- Odds Makers: "Jimmie the Greek"
- Weather Forecasters
- Actuaries
- Seismologists.

All of us respect the odds assessments of "Jimmie the Greek." We may have less respect for weather forecasters. A school of mathematical actuaries is concerned with so-called credibility formulas. These are formulas for determining insurance premiums based on experience as well as subjective judgement, e.g. Bühlmann,[3] Jewell.[4] This school, however, seems divided between subjective Bayesians and empirical Bayesians. A prominent seismologist has recently assessed the probability of a great earthquake in the state of California within the next ten years to be one-half. In many cases, though perhaps not in this instance, a probability of one-half is used when the forecaster has little knowledge concerning the event in question.

II. Assessment Aids

Bruno de Finetti was one of the earliest exponents of subjective probability. His ideas (cf. de Finetti,[5] Chapter 5) are the basis for much of the current research on probability assessment aids. Some of the most promising are:

1. *Linear Programming* as an aid in checking for probabilistic consistency as well as for exploring the probabilistic consequences of assessed probabilities;

*Work supported by the Air Force Office of Scientific Research under Grant AFOSR-81-0122 and the Mathematics and Statistics Division, Lawrence Livermore National Laboratory.

DTIC
COPY
INSPECTED
2

2. *Scoring Rules* as aids in encouraging serious probability assessment as well as evaluation of assessments in terms of actual event outcomes;
3. *Methods* for *Combining Expert Opinion*. Lindley et al.,[6] Lindley[7] and Lindley[8] are some of the best references concerned with combining expert opinion.

III. Linear Programming

For the last two years, I have asked my reliability class to assess their probabilities of the U.C. Bears winning certain football games. For example, I asked their probabilities for the (conditional) event that the Bears beat

Washington	Arizona $\mid F_2 = E_1$	UCLA $\mid F_3 = E_2$
E_1	$E_2 \mid F_2$	$E_3 \mid F_3$
p_1	p_2	p_3

where $p_i = P[E_i \mid F_i]$ and F_1 is the sure event. $E_1 = 1$ if the Bears beat Washington and $E_1 = 0$ otherwise. $E_2F_2 = 1$ if the Bears beat Arizona *and* Washington, *so* that p_2 is the conditional probability they beat Arizona *given* they beat Washington. Of course the Bears usually lose and we quickly lose interest. However, before the games are played, the students construct the following table of possible outcomes:

Table I

Implicit Probabilities	p_1 (E_1,F_1)	p_2 (E_2,F_2)	p_3 (E_3,F_3)
w_1	(E_{11},F_{11})	(E_{21},F_{21})	(E_{31},F_{31})
w_2	(E_{12},F_{12})	(E_{21},F_{22})	(E_{32},F_{32})
$\vdots$	$\vdots$	$\vdots$	$\vdots$
w_8	(E_{18},F_{18})	(E_{28},F_{28})	(E_{38},F_{38})

The j^{th} possible outcome corresponds to the pairs: (E_{1j},F_{1j}) , (E_{2j},F_{2j}) , (E_{3j},F_{3j}) . $E_{ij} = 1$ if E_i also occurs in the j^{th} possible outcome and $E_{ij} = 0$ otherwise. For example, the first row of the table could be

$$(1,1) \quad (1,1) \quad (1,1)$$

signifying the Bears won all these games and the conditioning events of course were also true in this case. The students then use a linear program devised by Bob Nau[9] to determine whether or not their assessed probabilities are consistent. If there exist $w_j^* > 0$, $j = 1,2, \ldots, 8$ such that

$$p_i = \sum_{j=1}^{8} E_{ij}F_{ij}w_j^* \Big/ \sum_{j=1}^{8} F_{ij}w_j^* \qquad (1)$$

then the student's probabilities are consistent (we say he is strictly coherent). If $w_j^* \geq 0$, and some $w_j^* = 0$, then the student is only coherent and he has implicitly assigned probability zero to a logically possible outcome. If no such w_i^*'s exist, then he is inconsistent (we say he is incoherent) and should revise his opinion.

In his Ph.D. thesis, Bob Nau[9] formulated the problem of determining (strict) coherence as a linear programming problem. This was done in the context of a two person zero sum game involving a Bookie (the probability assessor) and a Bettor (his adversary). The Bookie sets fair prices p_i $i = 1,2, \ldots, n$ for one dollar payoffs on event pairs

$$(E_1,F_1), \ldots, (E_n,F_n) .$$

The bettor is allowed to make z_i bets on (or against) the event pairs (E_i,F_i) . The Bookie charges the Bettor p_iz_i dollars for the privilege. If $E_iF_i = 1$, the Bookie pays z_i dollars to the Bettor. If $F_i = 0$, the Bookie refunds the Bettor his money and the bet is called off. If $E_i = 0$ and $F_i = 1$, the Bettor receives nothing. The Bookie's net gain is then

$$\sum_{i=1}^{n} (p_i - E_i)z_iF_i ,$$

which can of course be negative. If the j^{th} outcome occurs, let the Bettor receive $y_o + y_j$ and hence

$$y_o + y_j + \sum_{i=1}^{n} (p_i - E_{ij})F_{ij}z_i = 0$$

since it is a zero sum game.

Given $p_1, p_2, \ldots, p_n$, consider the problem

$$\begin{aligned} &\text{Maximize} \quad y_o \\ &\text{subject to} \quad y_o + y_j + \sum_{i=1}^{n} (p_i - E_{ij}) z_i F_{ij} = 0 \qquad j = 1,2, \ldots, m, \\ &\sum_{j=1}^{m} \beta_j y_j = 1 \\ &y_j \geq 0 \qquad j = 1,2, \ldots, m. \end{aligned}$$

We must also specify $\beta_j > 0$, $j = 1,2, \ldots, m$ (the Bettor's probabilities on outcomes). Usually, we choose $\beta_1 = \beta_2 = \ldots = \beta_m = 1/m$. Let y_o^* be the solution to our problem. Then the Bettor is guaranteed to make at least y_o^*. It turns out that if $y_o^* > 0$, the Bookie is inconsistent and he should revise his prices ("probabilities"). If $y_o^* = 0$, the Bookie is coherent and if $y_o^* < 0$ he is strictly coherent.

By using this linear programming approach, the probability assessor can quickly determine if his probabilities are consistent. He can also explore the consequences of his probability assessments by performing a sensitivity analysis with the linear program and its dual.

IV. Scoring Rules

How can the probability assessor be encouraged to think carefully about his probability assessments? One answer may be the scoring rule. For example, suppose a weather forecaster states his probabilities r_1, r_2, r_3 for the events

Rain	Snow	Sun
E_1	E_2	E_3
r_1	r_2	r_3

Suppose further that the events are carefully defined so that they are mutually exclusive and exhaustive. The so-called Brier scoring rule is

$$\sum_{i=1}^{n} (r_i - E_i)^2 \qquad (2)$$

where $E_i = 1$ if E_i occurs and $E_i = 0$ otherwise. Clearly, scores close to zero are best. This rule has been advocated as a method for scoring weather forecasters.

To obtain a rule reflecting the user's (say a farmer's) concerns, let u_{dk} be the user's utility if he makes decision d and outcome k (say snow) occurs. Given forecast $r = (r_1, r_2, r_3)$, the user may then calculate

$$G(r) = \underset{d}{\text{maximum}} \sum_{k=1}^{n} u_{dk} r_k$$

to determine his best decision. Note that $G(r)$ is convex in r but not strictly convex. Let

$$G^*(r) = (G_1^*(r), \ldots, G_n^*(r))$$

be a subgradient to G at r; i.e., for all forecasts p

$$G(p) \geq G(r) + \langle p - r, G^*(r) \rangle$$

where $\langle \cdot, \cdot \rangle$ denotes inner product.

Now suppose outcome k occurs and p is the perfect forecast for k; i.e., $p_k = 1$ while $p_i = 0$ for $i \neq k$. Let

$$S_k(r) = G(p) - G(r) - \langle p - r, G^*(r) \rangle$$

be our scoring rule. Then in this particular case

$$G(p) = \max_d u_{dk}$$

and

$$S_k(r) = \max_d u_{dk} - u_{ik} \qquad (3)$$

when $\sum_{k=1}^{n} u_{ik} r_k \geq \sum_{k=1}^{n} u_{jk} r_k$ for all $j \neq i$. Clearly, (3) is the value to the user of perfect information minus the value of information contained in r when k occurs. This scoring rule has a property which Savage[10] called *proper*. We say that any scoring rule $S_k(r)$ is (strictly) proper iff

$$\sum_{k=1}^{n} r_k S_k(r) \underset{(<)}{\leq} \sum_{k=1}^{n} r_k S_k(q)$$

for all forecasts r and q. Hence, if you believe r but hedge and say q, *you* will expect to receive a worse score than if you are honest and say r. Scoring rule (3) is not strictly proper. However, (2) is strictly proper. L. J. Savage[10] explored in depth the use and properties of scoring rules as aids in probability assessment.

He (somewhat loosely) characterized proper and strictly proper scoring rules. Eduardo Haim[11] in his Ph.D. thesis clarified and extended Savage's work in several directions.

V. The Bayesian Paradigm

I spent the 1975-76 academic year at Florida State University in Tallahassee. My purpose was to complete a book on *Statistical Reliability Theory* with Frank Proschan. At the time, I was working on total time on test processes. I saw this as a way of unifying life test procedures. At the same time, I started attending lectures by Dev Basu on statistical inference. It was Lehmann's hypothesis testing course and Lehmann's book was the text. However, I noticed something strange - Basu never opened the book. He was obviously not following it. Instead, he was giving a very elegant, measure theoretic treatment of the concepts of sufficiency, ancillarity, and invariance. He was interested in the concept of information - what it meant - how it fitted in with contemporary statistics. As he looked at the fundamental ideas, the logic behind their use seemed to evaporate. I was shocked. I didn't like priors. I didn't like Bayesian statistics. But after the smoke had cleared, that was all that was left.

Berger[12] at Purdue University has recently written a new graduate level test on decision theory. In his preface, he says that he had intended to adopt a neutral position vis-a-vis the various statistical approaches. However, in the course of writing the book, he turned into a rabid Bayesian. He says "There was no single cause for this conversion; just a gradual realization that things seemed to ultimately make sense only when looked at from the Bayesian viewpoint."

Basu loves counterexamples. He is like an art critic in the field of statistical inference. He would find a counterexample to the Bayesian approach if he could. So far, he has failed in this respect.

Recently, Basu[13] wrote the following: "It is about 12 years now that I finally came to the sad conclusion that most of the statistical methods that I had learned from pioneers like Karl Pearson, Ronald Fisher and Jerzy Neyman and survey practitioners like Morris Hanson, P. C. Mahalanobis and Frank Yates are *logically untenable*." I believe he is right. Read Basu.[14]

At the beginning of this talk, I was concerned with the problem of assessing probabilities for seemingly very specialized problems of current interest. However, from the Bayesian point of view, this seems to be pretty much the problem of modern statistics. After modeling, the problem of statistical inference is the assessment and calculation of probabilities. However, this is not a trivial problem!

As I understand it, some of the main points of the Bayesian paradigm are:

1. All probabilities are subjective. However, the calculus of probability is essentially the same as for frequentists;
2. All probabilities are conditional on current information available to the analyst;
3. Probability is a measure of the analyst's uncertainty about *unknown* quantities;
4. *Information* is anything which changes the analyst's probability distributions about unknown quantities of interest.

References

1. L. L. George and R. W. Mensing, "Using Subjective Percentiles and Test Data for Estimating Fragility Functions," Proceedings of the 1980 DOE Statistical Symposium, pp. 27-35, October 1980.
2. P. G. Groer and R. E. Barlow, "Some Concepts, Terminology, and Methodology for Radiation Risk Assessment," Radiation Research, to appear, 1982.
3. H. Bühlmann, Mathematical Methods in Risk Theory, Springer-Verlag, New York, 1970.
4. W. S. Jewell, "Models in Insurance: Paradigms, Puzzles, Communications and Revolutions," presented at the 21st International Congress of Actuaries, Zurich and Lausanne, June 19-26, 1980.

5. B. de Finetti, Theory of Probability, Vol. I, Wiley, New York, 1974.

6. D. V. Lindley, A. Tversky and R. V. Brown, "On the Reconciliation of Probability Assessments," J. Roy. Stat. Soc. A, 142, Part 2, pp. 146-180, 1979.

7. D. V. Lindley, "Reconciling Continuous Probability Assessments," Decision Science Consortium, 1981; submitted to Operations Research.

8. D. V. Lindley, "The Improvement of Probability Judgements," J. Roy. Stat. Soc. A, 1982, to appear.

9. R. Nau, "Coherent Assessment of Subjective Probability," Ph.D. Dissertation, University of California, Berkeley, 1981.

10. L. J. Savage, "Elicitation of Personal Probabilities and Expectations," JASA, Vol. 66, No. 336, pp. 783-801, 1971.

11. E. Haim, "The Characterization of Strictly Proper Scoring Rules in Decision-Making," Ph.D. Dissertation, University of California, Bekeley, 1981.

12. J. O. Berger, Statistical Decision Theory, Springer-Verlag, New York, 1980.

13. D. Basu, "A Discussion on Survey Theory," Florida State University Statistics Technical Report, 1979.

14. D. Basu, "Statistical Information and Likelihood," Sankhyā, Vol. 37, pp. 1-71, 1975.

8

DTIC