A program for evaluating, fitting, and debiasing subjective probability distributions

HERBERT MOSKOWITZ

and

WILLIAM I. BULLERS

*Krannert Graduate School of Management
Purdue University
West Lafayette, Indiana 47907*

Decision makers must often make or obtain subjective probability distribution (SPD) assessments associated with some uncertain quantity (u.q.). The u.q. may concern a social decision problem such as predicting the safety of a nuclear power plant. Or it may concern a business decision problem such as forecasting the expected return from an investment or the time to complete an activity in a large-scale project. Such assessments may exhibit certain biases, some of which are expressible in terms of the moments of a probability distribution. The programmed algorithm analyzes and evaluates actually assessed SPDs in terms of these and other biases. It can also fit beta probability density functions (pdfs) to the actual assessments, statistically test the goodness of fit of the betas to the assessed distributions, and evaluate their normative goodness. The program provides several manipulative capabilities for "debiasing" SPDs. The algorithm has been designed for either direct fractile assessments (Moskowitz & Bullers, in press; Alpert & Raiffa, Note 1; Selvidge, Note 2) or PERT- (Program Evaluation and Review Technique) type assessments, as, for example, employed in the assessment of activity times in large-scale projects (Weist & Levy, 1969).

Description. The programmed algorithm evaluates either directly assessed or beta fitted SPDs using performance measures of (1) external validation, (2) prediction accuracy, (3) skewness, and (4) modality. External validation measures the second moment bias, and is concerned with whether assessments are properly calibrated, that is, whether an assessed distribution reflects an assessor's true state of knowledge or is "too tight" or "too loose." A "too tight" (loose) distribution implies that an assessment expresses more (less) knowledge than it should. External validation measures include the interquartile (IQ) range—the percent of time the true value lies between the .25 and .75 fractiles ($x_{.25}$ and $x_{.75}$), which should be 50%—and the "surprise index" (SI)—the percent of time the true value lies either below or above the .01 and .99 fractiles ($x_{.01}$ and $x_{.99}$), which should be 2% (Alpert & Raiffa, Note 1). Prediction accuracy measures the first moment bias and precision. First moment bias denotes how far the median (.50 fractile) or mode deviates from the true value. Precision measures the dispersion of an SPD, which reflects the amount of

uncertainty of (or confidence in) an assessment. Skewness is also measureable directly from the parameters of the beta fits to the actually assessed SPDs. The modality of an SPD is determined from the precise and approximate measures described in Alpert and Raiffa (Note 1).

Beta pdfs are fitted to both the PERT and direct fractile assessments as follows. For the PERT assessments, the beta parameters r and n are obtained from (Freund, 1971, p. 172).

$$r = \mu[(1 - \mu)/(\sigma^2) - 1] \quad (1)$$

and

$$n = (1 - \mu)[\mu(1 - \mu)/(\sigma^2) - 1], \quad (2)$$

where μ and σ denote the mean and standard deviation, respectively. These parameters are estimated by (MacCrimmon & Ryavec, 1966)

$$\mu = (a + 4m + b)/6 \quad (3)$$

and

$$\sigma = (b - a)/6, \quad (4)$$

where a , m , and b are the lower bound, most likely, and upper bound estimates. Carter's approximation formula (Pearson & Hartley, 1954) is used to generate the resulting fractiles directly from the parameters of the beta.

For the direct fractile assessments, beta pdfs are fitted in either one of two ways: (1) A beta is fitted to the quartiles and median (viz., .25, .75, and .50 fractiles), which is then used to generate the extreme tail fractiles (viz., the .01, and .99 fractiles) from the fitted betas; (2) a beta is fitted to the tail fractiles and median (viz., .01, .99, and .50 fractiles), which is then used to generate the quartiles (viz., the .25 and .75 fractiles). The precise fitting procedures are outlined in Moskowitz and Bullers (in press) and follow Alpert and Raiffa (Note 1). The goodness of the beta fits to the PERT-type and fractile assessments is tested using a Kolmogorov-Smirnov statistic.

The beta pdfs are then manipulated, if desired, using scaling constants on the parameters to "debias" the assessed distributions. One approach, which follows Alpert and Raiffa (Note 1), involves finding an optimal scaling factor ρ^* ($0 \leq \rho \leq 1$), which reduces the beta parameters r and n proportionally. This, in effect, spreads the distribution to reduce the second moment bias inherent in human assessments. A second method manipulates separate scaling constants for r and n . This method finds $\vec{\rho} = (\rho_1^*, \rho_2^*)$, which reduces both first and second moment biases (Moskowitz & Bullers, in press).

Input. The input consists of five cards plus data. Card 1 reads in the type of data input (PERT or fractile assessments), m , the number of data observations (SPDs assessed), and a data report heading. Card 2 specifies the format of the m data cards to follow. Data cards 3 through $m+2$, inclusive, input the m PERT or fractile assessments, one per card. Card $m+3$ specifies the type of analysis to be performed on the preceding data group and an analysis report heading. Card $m+4$ specifies the format of the analysis parameter card to follow. Card $m+5$ inputs the analysis parameters, including (a) the true value of the uncertain quantity, (b) a divisor for converting input data to 0 – 1 proportions, (c) and (d) parameters to specify whether the beta pdf r and n parameters are to be manipulated (rescaled), and (e), (f), and (g) ρ beginning, ending, and incremental values, respectively, for adjusting beta parameters r and n . Analysis control cards ($m+4$, $m+5$, and $m+6$) may be repeated with different analysis types and/or analysis parameters to accommodate multiple analysis runs for each data group input. Similarly, the entire data group and analysis control inputs may be repeated in a program run for different data.

Output. The printed output includes: (1) a listing of the data group control cards, (2) the input data, (3) the analysis control cards, (4) individual observation analysis reports, and (5) summary analysis reports. Outputs 4 and 5 depend on whether PERT or fractile assessments are being analyzed. For analysis of the PERT assessments, the individual observation output (4) includes: (4a) input PERT a , m , b values, (4b) the estimated range ($b - a$), (4c) the range ratio $[(b - a)/m]$, (4d) computations of the mean, standard deviation, mode, first moment bias, and beta distribution parameters α and β where $\alpha = r$ and $\beta = n - r$. For the beta distribution fit to the PERT assessments, the output includes: (4e) calculated fractiles $x_{.01}$, $x_{.25}$, $x_{.50}$, $x_{.75}$, and $x_{.99}$; (4f) ranges ($x_{.99} - x_{.01}$) and ($x_{.75} - x_{.25}$); (4g) range ratios $(x_{.99} - x_{.01})/(x_{.75} - x_{.25})$ and $(x_{.99} - x_{.01})/x_{.50}$; (4h) first moment bias; and (4i) mean absolute deviations of the fitted $x_{.01}$, $x_{.50}$, and $x_{.99}$ fractiles from the PERT a , m , and b inputs. PERT summary statistics (5) include: (5a) means and variances of the individual observation parameters of output 4; and (5b) true value category counts (Alpert & Raiffa, Note 1).

For analysis of the fractile assessments, the individual observation output includes: (4a) input $x_{.01}$, $x_{.25}$, $x_{.50}$, $x_{.75}$, and $x_{.99}$ fractiles; (4b) ranges ($x_{.99} - x_{.01}$) and ($x_{.75} - x_{.25}$); (4c) range ratios $(x_{.99} - x_{.01})/(x_{.75} - x_{.25})$ and $(x_{.99} - x_{.01})/x_{.50}$; (4d) first moment bias; and (4e) a multimodality flag. For the beta distributions fit to the quartile or tail fractiles, the individual observation output includes: (4f) α and β parameters; (4g) fractiles, ranges, range ratios, and first moment biases similar to those obtained with the

original fractile assessments; and (4h) mean absolute deviations of the fitted fractiles from the original fractile assessments. Fractile summary statistics (5) for the original fractile assessments, beta quartile-fit fractiles, and beta tail-fit fractiles include: (5a) means and variances of the parameters of output reports (4), (5b) true value category counts, and (5c) bimodality and approximate multimodality counts. The fractile outputs (4 and 5) are also available for each value of the ρ parameter when varying the beta parameters r and n . Analysis options exist for reducing the amount of printed output in 4 and 5 when varying ρ for the fractile assessments.

Computer and Language. The program is written in FORTRAN IV for the CDC 6500 computer. Conversion to IBM and other computer systems is straightforward.

Restrictions. Input data arrays are currently dimensioned to handle a maximum of 100 assessed SPDs per data group. As such, the program requires about 32K (octal) to load and run. This restriction on sample size can easily be altered by changing a single dimension statement and a variable initialization value.

As currently implemented, the algorithm requires that all input data be convertible to 0 – 1 proportions to enable fitting beta distributions to the original assessments. Thus, an upper limit for the value of assessments to the unknown quantity must be input to the program for converting "almanac data" to proportion data. The capability for fitting other specific density functions, such as the gamma or normal, to the assessments can be easily accommodated.

Limitations of problem size are primarily a function of core memory capacity and CPU time requirements. Varying r and n parameters of beta distributions over a wide range significantly increases CPU time. As an illustration, CPU time for producing all outputs for 18 assessed SPDs without varying r and n parameters was about 1 sec. When manipulating r and n over a ρ range of .25 to 2.00 by increments of .25, the same 18 observations had a CPU time of about 13 sec for detailed individual output and 5 sec for summary output.

Availability. A description of the computer program input and complete program listing may be obtained by writing Herbert Moskowitz, Krannert Graduate School of Management, Purdue University, West Lafayette, Indiana 47907.

REFERENCE NOTES

1. Alpert, M., & Raiffa, H. *A progress report on the training of probability assessors*. Unpublished manuscript, Harvard University, Cambridge, Massachusetts, 1969.
2. Selvidge, J. *Experimental comparison of different methods for assessing the extremes of probability distributions by the fractile method*. Report No. 75-13. Business Research Division, Graduate School of Business Administration, University of Colorado, Boulder, Colorado, March 1975.

fractile method. Report No. 75-13, Business Research Division, Graduate School of Business Administration, University of Colorado, Boulder, Colorado, March 1975.

REFERENCES

- FREUND, J. *Mathematical statistics*. Englewood Cliffs: N.J.: Prentice-Hall, 1971.
- MACCRIMMON, K. R., & RYAVEC, C. A. An analytical study of the PERT assumptions. In E. Buffa (Ed.), *Readings in production and operations management*. New York: Wiley, 1966.
- MOSKOWITZ, H., & BULLERS, W. I. Modified PERT versus fractile assessment of subjective probability distributions. *Organizational Behavior and Human Performance*, in press.
- PERSON, E., & HARTLEY, H. (Eds.), *Biometrika tables for statisticians* (Vol. 1). Cambridge: Cambridge University Press, 1954.
- WEIST, J., & LEVY, F. K. *Management guide to PERT/CPM*. Englewood Cliffs, N.J.: Prentice-Hall, 1969.

(Accepted for publication September 13, 1978.)