RR-73-11

THE ELICITATION OF CONTINUOUS PROBABILITY DISTRIBUTIONS

James E. Matheson

Robert L. Winkler

November 1973

The Elicitation of Continuous Probability Distributions*

James E. Matheson** and Robert L. Winkler***

1. Introduction

Various methods have been developed to aid an individual in assessing (encoding) personal probabilities to be used in inferential and decision-making situations (e.g. see Winkler [12], and Spetzler and Stael von Holstein [9]). Included among these elicitation procedures are scoring rules, which encourage an assessor to reveal his opinions and to make his stated probabilities correspond with his judgments. Scoring rules, which involve the computation of a score based on the assessor's stated probabilities and on the event that actually occurs, are useful in the evaluation of probability assessors as well as in the elicitation process itself. For general discussions of scoring rules, see Winkler [13], Murphy and Winkler [7], Stael von Holstein [10], and Savage [8].

The development of scoring rules has, in general, been restricted to the elicitation of individual probabilities or

*This research was supported in part by the U.S. National Science Foundation under Grants GA-31735 and GA-41232.

**Decision Analysis Group, Stanford Research Institute, U.S.A.

***Graduate School of Business, Indiana University, U.S.A.; research scholar at the International Institute for Applied Systems Analysis, Laxenburg, Austria.

discrete probability distributions. In Winkler [13], it is pointed out that scoring rules developed for discrete situations can be used to elicit continuous probability distributions through the use of randomly generated partitions that are not known to the assessor at the time of elicitation. The purpose of this paper is to develop classes of scoring rules based on the entire density function (or equivalently, the distribution function) rather than just on a set of probabilities determined from the density function via a partition. We generate an extremely rich set of scoring rules that includes previously developed rules (including discrete rules) as special cases and provides the experimenter with a great deal of flexibility in choosing a rule that is particularly appropriate for a given situation. The families of rules generated in Section 2 are based on binary scoring rules, and the families of rules generated in Section 3 are based on another type of payoff function, or scoring function.

2. The Generation of Scoring Rules for Continuous Distributions from Scoring Rules for Binary Situations

Consider the assessment of the probability of a single event E. We assume a subject assigns probability p to the occurrence of the event, but when asked to reveal his probability assignment states a probability r which might not be equal to p. A scoring rule S(r) gives the subject a payoff

$S(r) = S_1(r)$ is the event occurs and $S(r) = S_2(r)$ if it does not. The subject's expected payoff for this binary situation is accordingly

$$E(S(r)) = pS_1(r) + (1 - p)\, S_2(r) \quad , \tag{1}$$

and the scoring rule is defined as strictly proper if

$$E(S(p)) > E(S(r)) \quad , \quad \text{for } r \neq p \ . \tag{2}$$

The notion of scoring rules can be generalized quite easily to the assessment of any discrete probability distribution. Let E_i represent the ith event (or ith value of a random variable), where $i \varepsilon I$ and I is finite or countably infinite. Moreover, let p_i and r_i correspond to p and r in the binary situation, and suppose that the scoring rule $S(r_1, r_2, \ldots)$ gives the subject a payoff $S_j(r_1, r_2, \ldots)$ if E_j occurs. Then

$$E(S(r_1, r_2, \ldots)) = \sum_{j \varepsilon I} p_j S_j(r_1, r_2, \ldots) \quad , \tag{3}$$

and S is strictly proper if

$$E(S(p_1, p_2, \ldots)) > E(S(r_1, r_2, \ldots))$$

when

$$r_i \neq p_i \quad , \quad \text{for any } i \varepsilon I \ . \tag{4}$$

The literature regarding such rules is fairly extensive; several forms of strictly proper scoring rules have been developed (e.g. see the references given in Section 1). Three frequently-encountered examples are the quadratic, logarithmic, and spherical scoring rules, which are, respectively,

$$S_j(r_1,r_2,\ldots) = 2r_j - \sum_{i\varepsilon I} r_i^2 \quad , \tag{5}$$

$$S_j(r_1,r_2,\ldots) = \log r_j \quad , \tag{6}$$

and

$$S_j(r_1,r_2,\ldots) = r_j/(\sum_{i\varepsilon I} r_i^2)^{\frac{1}{2}} \quad . \tag{7}$$

Scoring rules have been extended to the continuous case by limiting arguments. If x is the revealed value of the variable of interest and $r(\cdot)$ represents the density function assigned by the subject, continuous analogs of the quadratic, logarithmic, and spherical scoring rules are, respectively,

$$S(r(\cdot)) = 2r(x) - \int_{-\infty}^{\infty} r^2(x)\ dx \quad , \tag{8}$$

$$S(r(\cdot)) = \log r(x) \quad , \tag{9}$$

and

$$S(r(\cdot)) = r(x)/(\int_{-\infty}^{\infty} r^2(x)\ dx)^{\frac{1}{2}} \quad . \tag{10}$$

Rules such as these are strictly proper scoring rules for the continuous case. It might be argued, however, that such rules are somewhat deficient. For example, they are sensitive to the probability density function at the precise point of the revealed value of the variable, but not to the amount of probability mass nearby. The following development generates continuous rules from binary rules to produce new continuous rules that are sensitive to the entire density function, not just to the density at a single value. In this sense, the rules generated here can be thought of as sensitive to distance (e.g. see Stael von Holstein [10]).

Consider the assessment of a probability distribution for a variable defined on the real line. We assume the subject assigns probability distribution function $F(\cdot)$ to the variable, but when asked to reveal his probability assignment states $R(\cdot)$. Let x be the revealed value of the variable and let u be an arbitrary real number we shall use to divide the variable into two intervals (see Figure 1), $I_1 = (-\infty,u]$ and $I_2 = (u,\infty)$. Let E be the event that x falls in I_1. Applying the previous scoring rule with the identification $p = F(u)$ and $r = R(u)$, we have

$$S(R(u)) = \begin{cases} S_1(R(u)) & \text{if } x \varepsilon I_1 , \\ S_2(R(u)) & \text{if } x \varepsilon I_2 , \end{cases} \tag{11}$$

and

$$E(S(R(u))) = F(u)\ S_1(R(u)) + [1 - F(u)]\ S_2(R(u)) \quad . \tag{12}$$

If S is strictly proper, then

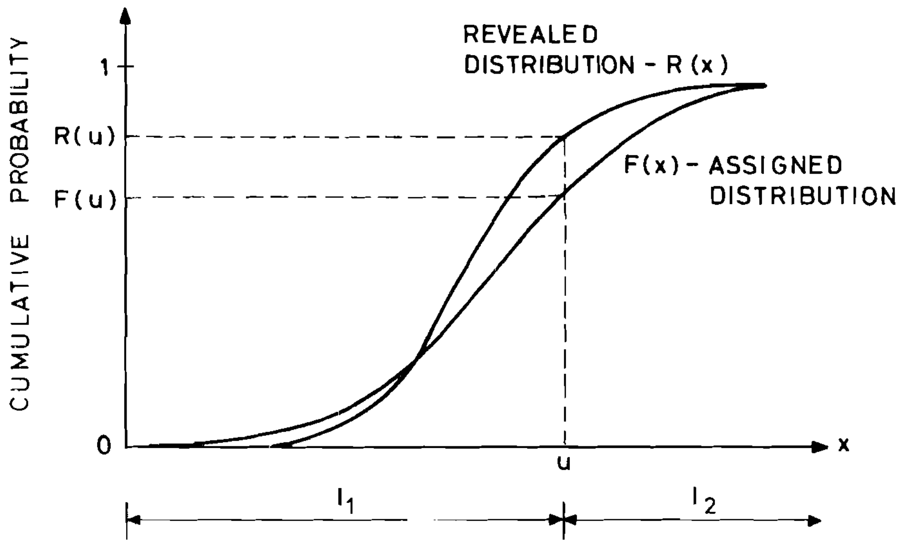

FIGURE 1. GENERATION OF PROBABILITY-ORIENTED SCORING RULES

$$E(S(F(u))) > E(S(R(u))) \quad , \quad \text{if } F(u) \neq R(u) . \tag{13}$$

Thus, the subject will maximize the expected payoff by setting $R(u) = F(u)$. If the subject does not know the value of u, he clearly should set $R(\cdot) = F(\cdot)$; however, his payoff depends strongly on the arbitrarily selected value of u. To eliminate this dependence, we can simply integrate $S(R(u))$ over all u and pay the subject this amount, which is

$$S^*(R(\cdot)) = \int_{-\infty}^{x} S_2(R(u)\ du + \int_{x}^{\infty} S_1(R(u))\ du \quad . \tag{14}$$

The corresponding expected score is

$$E(S^*(R(\cdot))) = \int_{-\infty}^{\infty} E(S(R(u)))\ du \quad . \tag{15}$$

Equations (14) and (15) are in direct analogy with Equations (11) and (12). Equation (15) can be derived as the expectation of Equation (14) with an interchange of order of integration. If S is strictly proper, then S* is strictly proper, and the subject maximizes his expected payoff by setting $R(u) = F(u)$ for each u. The above approach is applicable to some scoring rules such as the quadratic scoring rule, which will be considered later. However, the required integral may not exist for many other important rules, such as the logarithmic, so a more general method is needed.

To increase the generality and usefulness of the above result, we assume that the experimenter selects a probability distribution function $G(\cdot)$ for u. After a value of x has been revealed, he pays the subject the expected score using this distribution. The expected score given the revealed value x is

$$S^{**}(R(\cdot)) = E_{u|x}\,(S(R(\cdot)))$$

$$= \int_{-\infty}^{x} S_2(R(u))\, dG(u) + \int_{x}^{\infty} S_1(R(u))\, dG(u) \quad , \tag{16}$$

and before x is revealed the subject's expected score is

$$E(S^{**}(R(\cdot))) = \int_{-\infty}^{\infty} E(S(R(u)))\, dG(u) \quad . \tag{17}$$

Since S is strictly proper, S** is also strictly proper, and the subject maximizes his expected payoff by setting $R(u) = F(u)$ for each u. Incidentally, note that the experimenter could simply generate a single value from $G(\cdot)$ and use that value to reward the subject via Equation (11). However, although the mathematical results are identical, it seems preferable to pay the expected score given by Equation (16) instead of the score obtained from a single value generated from $G(\cdot)$.

If we write Equation (17) in density form,

$$E(S^{**}(R(\cdot))) = \int_{-\infty}^{+\infty} E(S(R(u)))\, g(u)\, du \quad , \tag{18}$$

we see that $g(\cdot)$ serves as a weighting function which should encourage the subject to pay more attention to his assessments where $g(u)$ is highest. Thus, if certain regions of values of the variable are of particular interest, the experimenter might make $g(\cdot)$ higher in these regions than it is elsewnere. Of course, $g(\cdot)$ could be a general weighting function (i.e. it is not necessary for $G(\cdot)$ to be a probability distribution function), but this does not increase the generality of our results. Technically, $G(\cdot)$ must be selected so that the integral of Equation (18), which depends on both the scoring rule and the probability distributions, will exist. If the interval of definition is finite or the integrand is well-behaved, $g(u)$ can be selected as uniform or "diffuse" to yield the earlier results of Equations (14) and (15).

This process generates continuous scoring rules from each binary scoring rule. For instance, consider the quadratic scoring rule defined by

$$S_1(r) = -(1 - r)^2$$

and (19)

$$S_2(r) = -r^2 \quad ,$$

with

$$E(S(r)) = -p(1 - r)^2 - (1 - p)\, r^2$$

$$= -(p - r)^2 - p(1 - p) \quad . \tag{20}$$

The generated continuous quadratic case defined by Equations (16) and (17) is a payoff of

$$S^{**}(R(\cdot)) = -\int_{-\infty}^{x} R^2(u)\ dG(u) - \int_{x}^{\infty} [1 - R(u)]^2\ dG(u) \qquad (21)$$

and an expected score of

$$E(S^{**}(R(\cdot))) = -\int_{-\infty}^{+\infty} [F(u) - R(u)]^2\ dG(u)$$

$$- \int_{-\infty}^{+\infty} F(u)[1 - F(u)]\ dG(u) \quad . \qquad (22)$$

If $G(\cdot)$ is "diffuse," then $dG(u)$ is replaced by du and the above equations have interesting graphical interpretations; it is left to the reader to sketch them.[1] If the subject sets $R(u) = F(u)$, then his expected score is

$$E(S^{**}(F(\cdot))) = -\int_{-\infty}^{+\infty} F(u)[1 - F(u)]\ dG(u) \quad , \qquad (23)$$

which is a measure of the dispersion in his true probability assignment. Thus, Equation (22) is the sum of two terms, the first rewarding honesty and the second rewarding expertise or sharpness. Although partitioning of the quadratic scoring rule and the resulting "attributes" measured by elements of various partitions have been studied (e.g. Murphy and Epstein [6], Murphy [4, 5], it appears that partitioning of the function representing the expected score has not been considered.

---

[1]During the final preparation of this paper, we learned that Brown (personal communication) has used a different approach to generate a rule that is apparently equivalent to the rule given by Equation (21) with $dG(u)$ replaced by du.

Although this work was motivated by the desire for better continuous scoring rules, the results are applicable for any probability distribution function $F(\cdot)$. Thus, they are applicable to the discrete case. Moreover, the continuous case can be discretized by choosing $G(\cdot)$ as a step function. For example, suppose that $G(\cdot)$ is a step function with positive steps $g_1, g_2, \ldots, g_n$ at $u_1 < u_2 < \cdots < u_n$ and that $R(u_i) = R_i$ and $F(u_i) = F_i$ for $i = 1,2,\ldots,n$. Then the quadratic binary scoring rule generates

$$S^{**}(R(\cdot)) = -\sum_{i=1}^{j-1} R_i^2 g_i - \sum_{i=j}^{n-1} (1-R_i)^2 g_i , \quad \text{if } x = u_j \qquad (24)$$

and

$$E(S^{**}(R(\cdot))) = -\sum_{i=1}^{n-1} (F_i - R_i)^2 g_i - \sum_{i=1}^{n-1} F_i(1-F_i) g_i . \qquad (25)$$

If $g_1 = g_2 = \cdots = g_n$, this is not the usual quadratic rule, but it is equivalent to the ranked probability score (e.g. Epstein [1], Murphy [3], Stael von Holstein [10]), which has quite different properties. In particular, the ranked probability score is sensitive to distance, and the procedures discussed in this section can be used to generate classes of scoring rules that are sensitive to distance for the continuous case as well as the discrete case.

3. The Generation of Scoring Rules for Continuous Distribution from Payoff Functions other than Binary Scoring Rules

The rules generated in Section 2 are based on binary scoring rules. Other rules for continuous distributions can be generated from different types of payoff functions. As in Section 2, we assume that the subject assigns probability distribution function $F(\cdot)$ to the variable of interest but states $R(\cdot)$ when asked to reveal his probability assignment. In order to treat the cases of discrete points and zero-probability intervals we shall define the inverse functions

$$F^{-1}(z) = \min_u \{u | F(u) \geq z\} \tag{26}$$

and

$$R^{-1}(z) = \min_u \{u | R(u) \geq z\} \tag{27}$$

for all $z \varepsilon (0,1)$. The typical case is illustrated in Figure 2.

For any arbitrary $z \varepsilon [0,1]$, let the subject receive a payoff according to the rule $T(R^{-1}(z))$. If T is strictly proper, then

$$E(T(F^{-1}(z))) > E(T(R^{-1}(z))) \quad , \quad \text{if } R^{-1}(z) \neq F^{-1}(z) \quad , \tag{28}$$

where

$$E(T(R^{-1}(z))) = \int_{-\infty}^{\infty} T(R^{-1}(z)) \; dF(x) \quad . \tag{29}$$

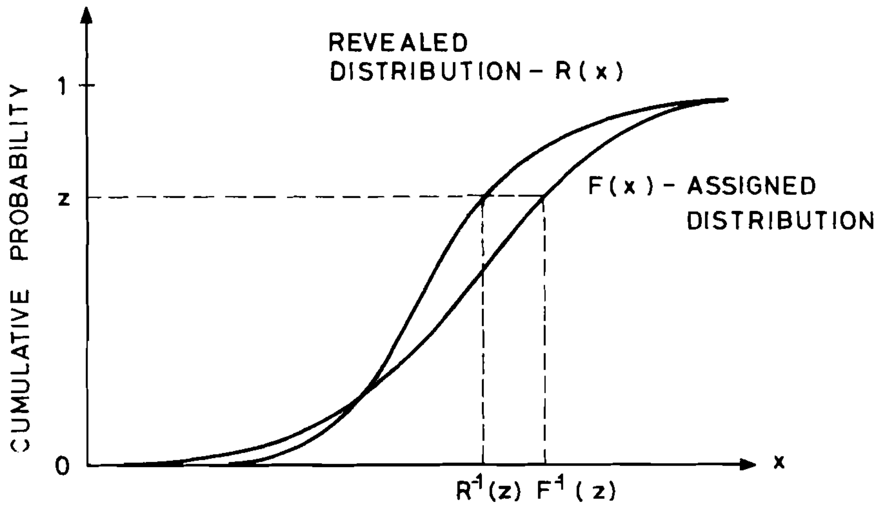

FIGURE 2. GENERATION OF VALUE-ORIENTED SCORING RULES

For example, let T represent a payoff function for a Bayesian point estimation problem under linear loss (e.g. see Winkler [14], pp. 397-405). For this situation, which is often called the "newsboy problem," the payoff function can be represented as follows:

$$T(R^{-1}(z)) = \begin{cases} \pi(x) - (1-z)[R^{-1}(z) - x] & \text{if } x \leq R^{-1}(z) , \\ \pi(x) - z[x - R^{-1}(z)] & \text{if } x > R^{-1}(z) , \end{cases} \tag{30}$$

where $\pi(x)$ is a function of x that represents the payoff if $x = R^{-1}(z)$ (i.e. if the newsboy orders exactly the right number of papers). Assuming that

$$\int_{-\infty}^{\infty} \pi(x) \; dF(x)$$

converges,

$$E(T(R^{-1}(z))) = \int_{-\infty}^{\infty} \pi(x) \; dF(x) - \int_{-\infty}^{R^{-1}(z)} (1-z)[R^{-1}(z) - x] \; dF(x) - \int_{R^{-1}(z)}^{\infty} z[x - R^{-1}(z)] \; dF(x) \tag{31}$$

is maximized only for $R^{-1}(z) = F^{-1}(z)$.

If the subject does not know the value of z, he should set $R(\cdot) = F(\cdot)$; however, the actual payoff depends strongly on the arbitrarily selected value of z. To eliminate this dependence, we integrate over all z and pay the subject

$$T^*(R^{-1}(\cdot)) = \int_0^1 T(R^{-1}(z))\ dz \quad . \tag{32}$$

The expected score is then

$$E(T^*(R^{-1}(\cdot))) = \int_{-\infty}^{\infty} \int_0^1 T(R^{-1}(z))\ dzdF(x)$$

$$= \int_0^1 E(T(R^{-1}(z)))\ dz \quad . \tag{33}$$

The integration over z is analogous to the integration over u in Section 2. If T is strictly proper then T* is also strictly proper, and the subject maximizes his expected payoff by setting $R(\cdot) = F(\cdot)$.

We can now generalize the above result in a manner analogous to the generalization represented by Equation (16) in Section 2. Assume that the experimenter selects a probability distribution function $H(\cdot)$ for z. After a value of x has been revealed, the subject is paid the expected score using H:

$$T^{**}(R^{-1}(\cdot)) = E_{z|x}\ (T(R^{-1}(\cdot))) = \int_0^1 T(R^{-1}(z))\ dH(z) \quad . \tag{34}$$

Before x is revealed, the subject's expected score is

$$E(T^{**}(R^{-1}(\cdot))) = \int_{-\infty}^{\infty} \int_{0}^{1} T(R^{-1}(z))\ dH(z)\ dF(x)$$

$$= \int_{0}^{1} E(T(R^{-1}(z)))\ dH(z) \quad . \qquad (35)$$

For example, the payoff generated by the scoring rule of Equation (30) is

$$T^{**}(R^{-1}(\cdot)) = \pi(x) - \int_{0}^{R(x)} z[x - R^{-1}(z)]\ dH(z)$$

$$- \int_{R(x)}^{1} (1 - z)[R^{-1}(z) - x]\ dH(z) \quad . \qquad (36)$$

If T is strictly proper, the subject maximizes his expected payoff from Equation (35) by setting R(·) = F(·). H is similar to G in that dH(·) serves as a weighting function which should encourage the subject to pay more attention to his assessments where dH(·) is highest. For example, if the experimenter is particularly concerned about the extreme tails of the distribution, he might select a U-shaped dH(·): if the middle of the distribution is of interest, dH(·) might be taken to be symmetric and unimodal with mode at z = 0.5. Of course, dH(·) can simply be uniform, in which case Equations (34) and (35) reduce to Equations (32) and (33). If only certain fractiles are of interest, H(·) can be chosen as a step function with positive steps $h_1, h_2, \ldots, h_m$ at $z_1 < z_2 < \cdots < z_m$.

In this section we have generated a family of scoring rules, with each member of the family corresponding to a particular choice of T(•) and H(•). This is similar to the situation covered in Section 2, where each member of the family of scoring rules that is generated corresponds to a particular choice of S(•) and G(•). The two families are completely different, however. The scoring rule S(•) is defined on the probability space (the unit interval), whereas T(•) is defined on the space of values of the variable of interest (the real line). In practice, the choice of a particular rule might be based primarily on convenience and on psychological considerations relating to the elicitation procedure. For instance, experimental results suggest that different elicitation techniques may yield quite different results (e.g. see Tversky and Kahneman [11] and Kahneman and Tversky [2]). Clearly such factors need to be investigated further.

References

[1] Epstein, E.S. "A Scoring System for Probability Forecasts of Ranked Categories," Journal of Applied Meteorology, 8 (1969), 985-987.

[2] Kahneman, D. and Tversky, A. "Subjective Probability: A Judgment of Representativeness," Cognitive Psychology, 3 (1972), 430-454.

[3] Murphy, A.H. "On the 'Ranked Probability Score'," Journal of Applied Meteorology, 8 (1969), 988-989.

[4] Murphy, A.H. "Scalar and Vector Partitions of the Probability Score: Part I. Two-State Situation," Journal of Applied Meteorology, 11 (1972), 273-282.

[5] Murphy, A.H. "Scalar and Vector Partitions of the Probability Score: Part II. N-State Situation," Boulder, Colorado, National Center for Atmospheric Research, unpublished manuscript, 1972.

[6] Murphy, A.H. and Epstein, E.S. "Verification of Probabilistic Predictions: A Brief Review," Journal of Applied Meteorology, 6 (1967), 748-755.

[7] Murphy, A.H. and Winkler, R.L. "Scoring Rules in Probability Assessment and Evaluation," Acta Psychologica, 34 (1970), 273-286.

[8] Savage, L.J. "Elicitation of Personal Probabilities and Expectations," Journal of the American Statistical Association, 66 (1971), 783-801.

[9] Spetzler, C.S. and Stael von Holstein, C.-A.S. "Probability Encoding in Decision Analysis," Management Science, in press, 1974.

[10] Stael von Holstein, C.-A.S. Assessment and Evaluation of Subjective Probability Distributions. Stockholm: Economic Research Institute, Stockholm School of Economics, 1970.

[11] Tversky, A. and Kahneman, D. "Availability: A Heuristic for Judging Frequency and Probability," Oregon Research Institute Research Bulletin, 11, No. 6 (1971).

[12] Winkler, R.L. "The Assessment of Prior Distributions in Bayesian Analysis," Journal of the American Statistical Association, 62 (1967), 776-800.

[13] Winkler, R.L. "The Quantification of Judgment: Some Methodological Suggestions," Journal of the American Statistical Association, 62 (1967), 1105-1120.

[14] Winkler, R.L. An Introduction to Bayesian Inference and Decision. New York: Holt, Rinehart and Winston, 1972.