DOCUMENT RESUME

ED 135 799 TM 005 666

AUTHOR Brown, Thomas A.
TITLE Admissible Scoring Systems for Continuous Distributions.
INSTITUTION Rand Corp., Santa Monica, Calif.
PUB DATE Aug 74
NOTE 27p.
AVAILABLE FROM The Rand Corporation, 1700 Main Street, Santa Monica, California 90406 (P-5235, $1.50)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Mathematical Models; *Prediction; *Probability; *Scoring; *Statistics
IDENTIFIERS *Admissible Probability Testing

ABSTRACT

The defining property of an admissible scoring system is that any individual perceives himself as maximizing his expected score by reporting his true subjective distribution. The use of admissible scoring systems as a measure of probabilistic forecasts is becoming increasingly well-known in those cases where the forecast is a discrete distribution over a finite number of alternatives. Most serious forecasts which are made in the real world seem to be forecasts of quantities rather than choices between a finite number of alternatives. In such cases as this, it seems much more natural to ask the forecaster to specify a continuous probability distribution which represents his expectations rather than trying to re-cast a basically continuous process into a discrete one. To construct an admissible scoring system for a continuous distribution, a collection of possible bets can be postulated on a continuous variable, and an admissible scoring system can be constructed as the net pay-off to a forecaster who takes all bets (and only those bets) which appear favorable on the basis of his reported distribution. Mathematical models for this and alternative systems are presented. (Author/BW)

ADMISSIBLE SCORING SYSTEMS FOR CONTINUOUS DISTRIBUTIONS

Thomas A. Brown

August 1974

P-5235

The Rand Paper Series

Papers are issued by The Rand Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests; Papers are not reports prepared in fulfillment of Rand's contracts or grants. Views expressed in a Paper are the author's own, and are not necessarily shared by Rand or its research sponsors.

The Rand Corporation
Santa Monica, California 90406

ADMISSIBLE SCORING SYSTEMS FOR CONTINUOUS DISTRIBUTIONS

Thomas A. Brown

I. INTRODUCTION

The use of admissible scoring systems as a measure of probabilistic forecasts is becoming increasingly well-known in those cases where the forecast is a discrete distribution over a finite number of alternatives (e.g.: Will it rain or not? Will Dewey, Truman, Wallace, or Thurmond be elected? Will the Rams or the Vikings win the game?). The defining property of an admissible scoring system is that any individual perceives himself as maximizing his expected score by reporting his true subjective distribution. That is to say, if you want to beat the system the best way to do it is to be honest.

Most serious forecasts which are made in the real world seem to be forecasts of *quantities* (e.g.: What will be the total U.S. wheat production during 1974? How many tanks will there be in the Egyptian Army on July 1, 1975? What will be the Dow-Jones average on January 2, 1976?) rather than choices between a finite number of alternatives. In such cases as this, it seems much more natural to ask the forecaster to specify a continuous probability distribution which represents his expectations rather than trying to re-cast a basically continuous process into a discrete one. But how can we construct an admissible scoring system for a continuous distribution? There are three basic approaches which seem to work:

(1) We can regard the continuous distribution as the limit of a discrete one, and derive a continuous

admissible scoring system as the limit of a sequence or discrete ones.

(2) We can create continuous admissible scoring systems by exploiting the Schwartz inequality, or by using other well-known inequalities of mathematical analysis.

(3) We can postulate a collection of possible bets on a continuous variable, and construct an admissible scoring system as the net pay-off to a forecaster who takes all bets (and only those bets) which appear favorable on the basis of his reported distribution. This is an exact analogue to the "gambling house" construction method which may be used to discover discrete admissible scoring systems.

Of the three techniques, I tend to prefer the third because it gives greater insight into what actually lies behind an admissible scoring system, and suggests ways to tailor the scoring system to accomplish one's goals more effectively in a given situation. But let us discuss each of the methods in turn.

II. LIMITS OF DISCRETE SCORING SYSTEMS

Suppose that the domain of possible answers to a forecasting question is an interval D. For example, if we are forecasting the temperature at noon on May 1 in Santa Monica we might take D to be all temperatures between 32° and 130°. Suppose a forecaster specifies a probability density function $f(x)dx$ over D which he asserts is his best subjective estimate of what the temperature will be. How should we reward him when the true temperature becomes known?

We could convert the problem into a discrete one by the following device: divide D into n small intervals, each of length Δx. Choose a set of n points $\{x_i\}$ in such a way that x_i is in the i^{th} interval. If $r(x)$ is continuous and Δx is small enough, then the forecaster is asserting (approximately) that there is a probability $r(x_i)\Delta x$ that the true temperature is in the i^{th} interval. This is a probabilistic forecast over a finite number (n) of alternatives, so we could use any discrete admissible scoring system on it. For the sake of definiteness, let us apply the quadratic admissible scoring system. This means that if the true answer is in the i^{th} interval, the forecaster would be rewarded as follows:

$$f_n(i) = 2r(x_i)\Delta x - \sum_{j=1}^{n}[r(x_j)\Delta x]^2$$

Note that the pay-off becomes small if n is large (since Δx is then small). Recall that if you multiply an admissible scoring system by a constant, you get another admissible scoring system. Therefore we have another admissible scoring system if we "renormalize" the one above by dividing out Δx:

$$f_n^*(i) = 2r(x_i) - \sum_{j=1}^{n} [r(x_j)]^2 \Delta x$$

If r(x) is continuous, then if we let $n \to \infty$ the sequence of scores $f_n^*(i)$ clearly goes to a limit F(t), where

$$f(t) = 2r(t) - \int_D [r(x)]^2 dx$$

In this expression, t stands for the "true answer." Because of the way in which it was constructed, we would expect the expression above to be an admissible scoring system on continuous distributions: any deviation from the true subjective r(x) would be reflected in a less than optimal score on the f_n^* for n sufficiently large, and therefore (one would think) in a less than optimal score on f. But to make this argument rigorous is somewhat cumbersome. A much more efficient way to provide a rigorous proof that f(t) is admissible is to invoke the Schwartz inequality, which we shall do in the next section.

Other discrete admissible scoring systems may be introduced as the quadratic was above, although the details of the renormalization process differ from case to case. If you use the logarithmic discrete scoring system the continuous system derived is simply

$$g(t) = \log r(t)$$

If you use the "spherical" scoring system of Masanao Toda, then the corresponding continuous system is

$$h(t) = \frac{r(t)}{\sqrt{\int r(x)^2 dx}}$$

This limiting process is a good way to discover continuous scoring systems, but it is a poor way to *prove* that a scoring system has the "admissible property."

III. EXPLOITING SOME WELL-KNOWN INEQUALITIES

Once a continuous admissible scoring system has been discovered (by means of a limiting process applied to a discrete admissible scoring system, or otherwise), it is usually not very difficult to provide a rigorous proof that the scoring system is, in fact, admissible by applying more or less well-known theorems and techniques from the field of integral inequalities. By way of illustration, let us begin by attacking the quadratic scoring system f(t), mentioned above. If we let s(t)dt denote the "true" probability density function, and r(t)dt denote the probability density function specified by the respondent, then we must prove that the respondent will maximize his expected score by making $r(t) = s(t)$. Putting this into symbols, we must prove that

$$\int 2s(t)r(t)dt - \int s(t)dt \int r(x)^2 dx \le \int 2s(t)^2 dt - \int s(t)dt \int s(x)^2 dx$$

Since s(t)dt is a probability density function, we have

$$\int s(t)dt = 1$$

Therefore we seek to show

$$\int 2s(t)r(t)dt - \int r(x)^2 dx \le \int s(x)^2 dx$$

which is the same as showing

$$0 \le \int [s(t) - r(t)]^2 dt$$

Since the above inequality obviously holds, and is a strict inequality unless $s(t) = r(t)$, except on a set of measure zero, it follows that the "quadratic" continuous scoring system does indeed have the admissible property.

Now let us turn our attention to the "logarithmic" continuous scoring system, g(t). As before, let s(t)dt denote the "true" probability density function, and let r(t)dt denote the one specified by the respondent. Our task is to prove that

$$\int s(t)\ \log\ s(t)dt \geq \int s(t)\ \log\ r(t)dt$$

and that equality holds if and only if s(t) = r(t). But this is the same as proving

$$0 \geq \int s(t)\ \log\ \left[\frac{r(t)}{s(t)}\right]dt$$

We know that

$$\log x \leq x - 1$$

with equality holding if and only if x = 1 (indeed, the right-hand side is a tangent line to the left-hand side of the above inequality). Therefore

$$\int s(t)\ \log\ \left[\frac{r(t)}{s(t)}\right]dt \leq \int s(t)\left\{\left[\frac{r(t)}{s(t)}\right] - 1\right\}dt = \int r(t)dt - \int s(t)dt$$
$$= 1 - 1 = 0$$

The inequality is strict unless r(t)/s(t) = 1 almost everywhere, and thus our desired result is established. Note, by the way, that we made use of the fact that r(t) integrates to one over the whole space, while we did not have to use this fact in establishing that the quadratic continuous scoring system is admissible. Indeed, the logarithmic scoring system as defined here can be "beaten" if you allow the respondent to specify improper distributions (ones which integrate to more than one), while this is not the case with the quadratic scoring system. As a practical matter, this means that any real-world implementation of the logarithmic scoring system must include a test (perhaps followed by a renormalization) to ensure that he is specifying a proper distribution. This test is unnecessary in the case of the quadratic scoring system for the respondent is only "hurting himself" if he specifies an improper distribution.

Finally, let us look at the "spherical" continuous scoring system. The inequality which we must establish in order to prove that this is an admissible scoring system is the following:

$$\frac{\int s(t)r(t)dt}{\sqrt{\int r(x)^2 dx}} \leq \frac{\int s(t)^2 dt}{\sqrt{\int s(x)^2 dx}}$$

This is readily transformed into

$$\int s(t)r(t)dt \leq \sqrt{\int s(t)^2 dt \int r(t)^2 dt}$$

which is the very well-known Schwartz inequality, with equality holding if and only if $s(t) = r(t)$ almost everywhere.

All three of the scoring systems which we have so far discussed have a common handicap in that they do not seem to take adequate account of the *topology* of the real line. This is to say, if forecaster A asserted that the true answer was certain to appear between 10 and 11, and the true answer was 12, common sense indicates that B had done a better job than A: he had put his distribution closer to the answer than A had. But none of the schemes we have discussed would give him any credit for that. If A and B both used rectangular distributions, they would both get identical scores, as follows:

Scoring System	A's Score	B's Score
Quadratic	-1	-1
Logarithmic	$-\infty$	$-\infty$
Spherical	0	0

We will now turn to a construction technique which can readily produce admissible scoring systems on continuous distributions which do not have this handicap.

IV. A SEQUENCE OF BETS CONSTRUCTION METHOD

A decisionmaker is interested in forecasts which help him make more intelligent bets about the future. If we take this aphorism seriously, it suggests that we construct a scoring system for probabilistic forecasters on the basis of how well a gambler would do who made all wagers (and only those wagers) which offered a positive expected payoff according to the distribution specified by the forecaster. This approach has been very successful as a technique for deriving the best-known discrete admissible scoring systems and gaining new insights into their properties. Let us try to apply it to derive continuous admissible scoring systems.

First of all, what is a typical bet in a continuous context? Let y be a real random variable, let x be some fixed real number, and let r be a fixed real number between zero and one. A typical bet would be for me to agree to pay you an amount r if y turns out to be greater than x, on condition that you pay me an amount 1 - r if y turns out to be less than or equal to x. This bet will look favorable to me if I believe the probability that y will be less than or equal to x is greater than r. My pay-off may be written symbolically as follows:

$$\ell_x(r, y) = \begin{cases} 1 - r & \text{if } y \le x \\ -r & \text{if } y > x \end{cases}$$

Similarly we could make bets where I receive the positive pay-off if $y > x$. Symbolically, such a bet could be written

$$u_x(r, y) = \begin{cases} -r & \text{if } y \le x \\ 1 - r & \text{if } y > x \end{cases}$$

This bet will look favorable to me if I believe the probability that $y \le x$ is less than $1 - 4$; otherwise, it will look unfavorable to me. In order to avoid minor technical problems with infinity, let us suppose [illegible] is certain (and known to everyone) that -L [illegible] ere L is some fixed real number). Let us suppos[illegible] for each x, we have a continuous spectrum of infinitesimal bets (proportional to the bets above) with r ranging from zero to one. Let $R(x)$ denote our "subjective probability" that $y \le x$. Then we would take up all the "lower" bets (of the type denoted by the function ℓ_x above) for which $r < R(x)$, and we would take up all the "upper" bets (of the type denoted by the function u_x above) for which $r < 1 - R(x)$. If t denotes the "true value" which the random variable y assumes, then our net pay-off for lower and upper bets taken at this value of x will be

$$\int_0^{R(x)} \ell_x(r, t)\,dr \quad \text{and} \quad \int_0^{1-R(x)} u_x(r, t)\,dr$$

respectively. We can calculate these integrals explicitly as follows:

$$\int_0^{R(x)} \ell_x(r, t)\,dr = \begin{cases} R(x) - \dfrac{R(x)^2}{2} & \text{if } t \le x \\[2ex] \dfrac{-R(x)^2}{2} & \text{if } t > x \end{cases}$$

$$\int_0^{1-R(x)} u_x(r, t)\,dr = \begin{cases} -\dfrac{(1-R(x))^2}{2} & \text{if } t \le x \\[2ex] 1 - R(x) - \dfrac{(1-R(x))^2}{2} & \text{if } t > x \end{cases}$$

These net pay-offs themselves constitute bets. If we imagine these bets as being distributed continuously over the interval $-L < x < L$, then we see that our grand total pay-off function will be as follows:

$$F(t) = \int_{-L}^{+L} dx \int_{0}^{R(x)} \ell_x(r, t)dr + \int_{-L}^{+L} dx \int_{0}^{1-R(x)} \quad , t)dr$$

It should be very clear that this pay-off scheme constitutes an admissible scoring system, because we would perceive any deviation between the R(x) we *reported* and the R(x) we *actually believed* as being equivalent either to rejecting some bets which were favorable to us or accepting some which were unfavorable.

Let's take a closer look at the scoring system we have just derived. Carrying out some obvious transformations shows that it can be expressed in the following form:

$$F(t) = \int_{-L}^{t} [1 - R(x)]dx + \int_{t}^{L} R(x)dx - \tfrac{1}{2} \int_{-L}^{+L} [R^2(x) + \big(1 - R(x)\big)^2]dx$$

Suppose a respondent does not make use of the freedom he has to specify a distribution, but simply makes a "point estimate" d. That is to say, he reports the following cumulative function:

$$R(x) = \begin{cases} 0 \text{ if } x < d \\ 1 \text{ if } x \geq d \end{cases}$$

Then a simple calculation shows

$$F(t) = L = |t - d|$$

This is quite a satisfying result: the penalty he suffers is exactly the amount by which he missed the true answer. What could be more natural?

What score does an individual expect to make if he believes in and reports a cumulative distribution R(x)? This is easily calculated as follows:

$$\text{Respondent's Expected Score} = \int_{-L}^{+L} R'(t)F(t)dt$$

$$= \int_{-L}^{+L} R'(t)dt \int_{-L}^{t} 1 - R(x)dx + \int_{-L}^{+L} R'(t)dt$$

$$\int_{t}^{L} R(x)dx - \tfrac{1}{2}\int_{-L}^{+L} R^2(x) + \left(1 - R(x)\right)^2 dx$$

$$= \int_{-L}^{+L}\int_{x}^{L} R'(t)\left(1 - r(x)\right)dtdx$$

$$+\int_{-L}^{+L}\int_{0}^{x} R'(t)R(x)dtdx$$

$$- \tfrac{1}{2}\int_{-L}^{+L} R^2(x) + \left(1 - R(x)\right)^2 dx$$

$$= \tfrac{1}{2}\int_{-L}^{+L} R^2(x) + \left(1 - R(x)\right)^2 dx$$

Note the interchange of order of integration in the calculation above. It is easy to see that a respondent's maximum expectation occurs when he is completely certain of the right answers (he then expects to score L) and the minimum expectation occurs when he feels there is a 50%

chance y = -L and a 50% chance that y = +L (in which case his expectation is L/4). If he feels that y is equally likely to assume any value between -L and +L, then his expectation is 2L/3. The reader may feel that the expected pay-off is not sufficiently sensitive to the precision of the respondent's estimate; but recall that any admissible scoring system may be multiplied by a positive constant or have any constant added to it. Thus we could "renormalize" to secure any degree of [illegible]tivity desired.

To verify directly [illegible] F(t) is an admissible scoring system, let us introduce S(x) to stand for the true cumulative function of the random variable y. Then the absolute expected pay-off to a response R(x) is as follows:

$$\text{Absolute Expected Score} = \int_{-L}^{+L} S'(t)F(t)\,dt$$

$$= \int_{-L}^{+L} S'(t)\,dt \int_{-L}^{t} \left(1 - R(x)\right)dx$$

$$+ \int_{-L}^{+L} S'(t)\,dt \int_{t}^{+L} R(x)\,dx - \tfrac{1}{2}\int_{-L}^{+L} R^2(x) + \left(1 - R(x)\right)^2 dx$$

$$= \int_{-L}^{+L} \Big\{\left(1 - S(x)\right)\left(1 - R(x)\right) + S(x)R(x)$$

$$- \tfrac{1}{2}\left(R^2(x)\right) - \tfrac{1}{2}\left(1 - R(x)\right)^2 \Big\}\,dx$$

$$= \tfrac{1}{2}\int_{-L}^{+L} \left\{S^2(x) + \left(1 - S(x)\right)^2\right\}dx - \int_{-L}^{+L} \left(S(x) - R(x)\right)^2 dx$$

The first integral in the right-hand side above depends only on the true distribution S(x); the second integral is obviously minimized if and only if the respondent's distribution equals the true distribution almost everywhere. This confirms that $F(t)$ is, indeed, an admissible scoring system.

The first integral on the right-hand side is of interest in itself, for it gives a general expression for the maximum score which an individual can expect against a given distribution. It can be interpreted as a constant minus the integral of the variance of the two-alternative distribution "greater than x or less than x" across all x. The reader may suspect that there is some deep relationship between this quantity and the variance of the distribution represented by $S(t)$ itself. Note, however, that if we multiply the random variable y by a positive constant K, then this quantity is multiplied by K while the variance of y is multiplied by K^2.

V. GENERALIZATIONS

It is quite clear from the nature of the "sequence of bets" construction method that we can generate other admissible scoring systems by varying the functions $\ell_x(r, t)$ and $u_x(r, t)$ in the expression

$$\int_{-L}^{+L} dx \int_{0}^{R(x)} \ell_x(r, t)dr + \int_{-L}^{+L} dx \int_{0}^{1-R(x)} u_x(r, t)dr$$

In fact, if $\varphi(x, r)$ and $\Psi(x, r)$ are any positive functions of x and r whatsoever, then we generate an admissible scoring system on continuous distributions by taking

$$\ell_x(r, y) = \begin{array}{rl} (1 - r)\varphi(x, r) & \text{if } y \leq x \\ -r\ \varphi(x, r) & \text{if } y > x \end{array}$$

$$u_x(r, y) = \begin{array}{rl} -r\ \Psi(x, r) & \text{if } y \leq x \\ (1 - 4)\Psi(x, r) & \text{if } y > x \end{array}$$

in the expression above. Whatever functions we use, however, we will always come out with a scoring system in terms of the cumulative probability function rather than the probability density function. This is because of the *form* of the bets we are permitting: they are all bets that y will fall in a given half-time (essentially). One could think of admissible scoring systems based on the probability density function (like the three discussed in Sections II and III above) as being generated by sequences of bets placed on whether or not y falls in a sequence of smaller and smaller intervals. Which class of scoring systems is more appropriate depends on the details of the particular application you have in mind.

VI. ASSESSING REALISM

Repeated experiments have indicated that it is a common human characteristic to overstate high probabilities and understate low ones. Or, putting it another way, to overestimate the degree of one's knowledge. Or, putting it still another way, to report subjective probability distributions which are too "tight." Some individuals exhibit the opposite form of behavior, however, and tend to hedge too much rather than too little. In any information system involving subjective probability estimates we would like to detect when either of these behavior patterns are present and provide appropriate feedback in order to help the estimators improve their behavior.

The way this has been done in Rand CAAPT implementations in the past is that the individual's external validity graph has been estimated (generally using a one-parameter linear least squares technique) and feedback has been based on this estimate. It is hard to see what would be the continuous analogue of an external validity graph, however. Therefore I believe that we should provide feedback to respondents (urging them to hedge more or hedge less) strictly on the basis of whether they score worse or better (over a number of questions) than they expected.

Let me give a heuristic justification for this approach. If an individual gets a significantly worse score than he expected, what does that indicate? It indicates that events he thought were relatively likely did not occur as frequently as he expected, while events he thought were relatively unlikely occurred more frequently than he expected. He would have done better if he had hedged his bets more. On the other hand, if his score is better than he expected, that indicates just the reverse, and he would have made an even better score if he had not hedged his bets so much.

Of course, an individual will rarely make *exactly* his expected score. In order to determine whether the difference between his expected score and his achieved score is great enough to justify corrective feedback, it seems reasonable to calculate the variance of his expected score and provide feedback only if his actual score falls more than two standard deviations (say) away from his expected score.

To make these ideas definite, let us work out specific formulas for the case of the quadratic scoring system $\left(f(t)\text{ derived in Section II}\right)$. Suppose an individual answers a set of n questions, giving a response $r_i(x)$ on the i^{th} question. His expected score (m_i) on the i^{th} question is then given by

$$m_i = \int [r_i(x)]^2 dx.$$

An easy calculation shows that the variance on the i^{th} question (σ_i^2) is given by

$$\sigma_i^2 = \left\{4 \int [r_i(x)]^3 dx - m_i^2\right\}$$

If the questions are independent of one another, then the variance of the total score will be simply the sum of the variances of the scores on the individual questions. In actual practice, the questions will probably not be completely independent of one another, and this will mean that the variance in the total score will be somewhat greater. This does not undermine our basic concept of using the variance calculated as though the questions are independent as a standard for providing feedback: it only means we must not try to be too precise in making statements about what the individual viewed his chance of making such a large or such a small score as being. The expected score on the whole test will be the sum of the expected scores on the individual questions, whether the questions are independent or not. So let us take

$$m = \sum_{i=1}^{n} m_i$$

$$\sigma^2 = \sum_{i=1}^{n} \sigma_i^{\,2}$$

$$T = 2\sum_{i=1}^{n} r_i(t) - \sum_{i=1}^{n} m_i$$

We might then provide feedback in the following form:

If $T < m - 4\sigma$: "Your score is significantly worse than you expected. You would have done much better if you had put more spread into your responses and not claimed to be so certain about the correct values."

If $m - 4\sigma \le T < m - 2\sigma$: "Your score is somewhat worse than you expected. You would have done better if you had hedged more in your responses."

If $m - 2\sigma \le T \le m + 2\sigma$: No corrective feedback.

If $m + 2\sigma < T \le m + 4\sigma$: "Your score is somewhat better than you expected. If would be even better if you had not hedged your responses quite so much."

If $m + 4\sigma < T$: "Your score is significantly better than you expected. You would have made an even better score if you had had more confidence and had not put so much spread into your responses.

VII. EXPERIMENTA~~ ~~ ~~ ~~NTATION

We seem to have some sort of answer to every question I can think of about CAAPT for continuous distributions. Therefore we should proceed at once to some sort of demonstration program so that we can garner some informal empirical experience before investing in a program designed for serious use and systematic test. JOSS seems like a good vehicle for such a demonstration program for the following reasons:

1. It's the only system in which I am proficient.
2. It lends itself to outside demonstrations.
3. Its limitation to alpha-numeric IO is a limitation shared by many systems which might be good vehicles for follow-on programs (e.g., Jerry Shure's system).

There are some serious difficulties with the JOSS system, however:

1. Its IO speed is very slow for this application.
2. Lack of graphic output makes some natural feedback schemes infeasible.

Even if the above difficulties mean that the JOSS program is ineffective, however, the work we put into it will not be entirely wasted: much of the work required in preparing the JOSS program will be required for any continuous CAAPT routine. So let us charge ahead with the outline of a plan for such an experimental JOSS implementation.

The scoring system used will be the quadratic scoring system introduced in Section II above. Feedback (at the end of a set of questions) will be based on the individual's expected score and variance (as discussed in Section VI above). The respondent will have two options:

1. Feedback after every question and at the end of the question set.
2. Feedback at the end of test only.

Questions will be in sets of up to nine. The i^{th} question in a set will be "part 90 + i." The true answer to the i^{th} question will be T(i); the number of questions in the set will be T(o); the normalization number (to be multiplied by the raw score on the question) will be U(i).

The responses will be demanded in the following format:

There is a .01 chance the true answer is less than:
There is a .20 chance the true answer is less than:
There is a .40 chance the true answer is less than:
There is a .60 chance the true answer is less than:
There is a .80 chance the true answer is less than:
There is a .99 chance the true answer is less than:

The above format will be used on the first question in a set only. Thereafter the following format will be used:

... .01...less than:
... .20...less than:
... .40...less than:
... .60...less than:
... .80...less than:
... .99...less than:

The six percentile breaks elicited on the i^{th} question would be stored as R(i, 1), R(i, 2)...R(i, 6). In calculating the reward function, we hypothesize that the probability density function is constant between the specified percentile breaks. This hypothesis enables us to make easy computations and provide feed-back well suited to alpha-numeric output. Specifically, if we ignore the density less than .01 and greater than .99, we can calculate the integral in the scoring function as follows:

$$Q = \frac{.0361}{R(i,2) - R(i,1)} + \frac{.0361}{R(i,6) - R(i,5)} +$$

$$\sum_{j=2}^{4} \frac{.04}{R(i,j+1) - R(i,j)}$$

We will also temporarily store the probability densities as follows:

$$D(1) = \frac{.19}{R(i,2) - R(i,1)}$$

$$D(2) = \frac{.20}{R(i,3) - R(i,2)}$$

$$D(3) = \frac{.20}{R(i,4) - R(i,3)}$$

$$D(4) = \frac{.20}{R(i,5) - R(i,4)}$$

$$D(5) = \frac{.19}{R(i,6) - R(i,5)}$$

After all six of the R's have been elicited from a respondent, then we will provide feed-back of the following form:

IF THE TRUE ANSWER IS:	YOU WILL SCORE:
less than R(i,1)	-U(i)·Q
greater than or equal to R(i,1) but less than R(i,2)	U(i)·[D(1) - Q]
greater than or equal to R(i,2) but less than R(i,3)	U(i)·[D(2) - Q]
greater than or equal to R(i,3) but less than R(i,4)	U(i)·[D(3) - Q]
greater than or equal to R(i,4) but less than R(i,5)	U(i)·[D(4) - Q]
greater than or equal to R(i,5) but less than R(i,6)	U(i)·[D(5) - Q]
greater than or equal to R(i,6)	-U(i)·Q

The underlined quantities above will, of course, be expressed in numerical terms. The above format will be one of two alternative feed-back formats. The more condensed version, used after the first question is over with, will be the following:

IF THE TRUE ANSWER IS:	YOU WILL SCORE:
less than R(i,1)	-U(i)·Q
between R(i,2)	U(i)·[D(1) - Q]
between R(i,3)	U(i)·[D(2) - Q]
between R(i,4)	U(i)·[D(3) - Q]
between R(i,5)	U(i)·[D(4) - Q]
between R(i,6)	U(i)·[D(5) - Q]
greater than or equal to	-U(i)·Q

The respondent will then be given an opportunity to revise his answer (if he wishes). If he is satisfied that his response is the best he can make, he will signal that fact to the machine. The machine will then report the true answer to the individual along with the score he achieved.

In order to form the basis for analytical feed-back at the end of the question set, we will keep a tally of true total score, expected total score, and expected variance in total score as quantities L, M, and N respectively. Therefore before we proceed to the next question we must carry out the following operations:

$$S(i) = \text{True score on question } i$$

$$E(i) = \text{Expected score on question } i$$

$$= .19 \cdot U(i) \cdot [D(1) + D(5)] + .2 \cdot U(i) \cdot \sum_{j=2}^{4} D(j) - U(i) \cdot Q$$

$$\text{Set } L = L + S(i)$$

$$\text{Set } M = M + E(i)$$

$$\text{Set } N = N + .02 \cdot [U(i) \cdot Q]^2 + .19\ U(i)^2 \cdot [(D(1) - Q)^2 + (D(5) - Q)^2] + .2 \cdot U(i)^2 \sum_{j=2}^{4} [D(j) - Q]^2 - E(i)^2$$

After all questions in the set have been answered the machine will provide one of the five feed-back messages specified in Section VI above, depending on whether $L < M - 4\sqrt{N}$, $M - 4\sqrt{N} \leq L < M - 2\sqrt{N}$, $M + 2\sqrt{N} < L \leq M + 4\sqrt{N}$, or $M + 4\sqrt{N} < L$