VOL. 106, NO. 7 MONTHLY WEATHER REVIEW JULY 1978

The Family of Quadratic Scoring Rules

CARL-AXEL S. STAËL VON HOLSTEIN[1]
Stanford Research Institute-Scandinavia, Stockholm, Sweden

ALLAN H. MURPHY
National Center for Atmospheric Research,[2] Boulder, CO 80307

(Manuscript received 1 December 1977, in final form 24 March 1978)

ABSTRACT

A family of quadratic scoring rules (QSR's) is defined. Some properties of these scoring rules are described, and it is demonstrated that QSR's are strictly proper. The probability (or Brier) score and the ranked probability score are shown to be special cases of the general QSR.

A geometrical framework for the representation of QSR's is presented. This framework facilitates formulation of QSR's and provides insight into the properties of these scoring rules, including the sensitive-to-distance property. The relationships between QSR's and measures of the value of (probability) forecasts are briefly discussed.

The richness of the family of QSR's provides the evaluator with considerable flexibility in choosing a scoring rule that is particularly suited to the situation at hand.

1. Introduction

The Probability, or Brier, score (PS) (Brier, 1950) is frequently used by meteorologists (and others) to evaluate probability forecasts. The PS is a strictly proper scoring rule (Murphy and Epstein, 1967) and possesses several other desirable properties (e.g., see Murphy, 1970). However, the widespread acceptance and use of the PS among meteorologists has not deterred evaluators (i.e., researchers and practitioners in the area of forecast evaluation) from their efforts to find even more suitable evaluation measures.

Epstein (1969) formulated a scoring rule that he called the ranked probability score (RPS), and the RPS has been shown to possess some properties that appear to make it more attractive than the PS in certain situations. Murphy (1969a) proved that the RPS is a strictly proper scoring rule and Staël von Holstein (1970) showed that it is sensitive to distance. This latter property, which relates to the weights associated with the events of concern, is *not* shared by the PS. These two scoring rules and their properties have been compared by Murphy (1970).

It is of interest to note that both the PS and the RPS have a quadratic form (e.g., see Murphy, 1971). In this regard, Murphy and Staël von Holstein (1975) showed that the RPS can be obtained from the PS by a linear transformation. The purpose of this paper is to define and study a general family of quadratic scoring rules (QSR's). The existence of this family of scoring rules provides the evaluator with considerable flexibility, in that he can choose a set of weights—specifically, a member of the family of QRS's—that is particularly suited to the situation at hand.

The family of QSR's is defined in Section 2. We also show that the PS and RPS are special cases of the general QSR in this section. Some properties of the QSR are considered in Section 3. In particular, we prove that the QSR is a strictly proper scoring rule. A geometrical framework for the representation of QSR's is described in Section 4, and the sensitive-to-distance concept is investigated within this framework. The use of QSR's and QSR's as value measures are discussed in Section 5, including some considerations related to the selection of a QSR for a particular situation and the existence of relationships between QSR's and measures of the (*ex post*) value of probability forecasts. Section 6 consists of a summary and conclusion.

2. Quadratic scoring rules (QSR's): Definitions

a. Forecasts and observations

In this paper, we assume that the range of values of the variable of concern has been divided into a set of n mutually exclusive and collectively exhaustive events. A probability forecast is then a (row) vector

[1] Current affiliation: Scandinavian Airlines System, Bromma, Sweden.

[2] The National Center for Atmospheric Research is sponsored by the National Science Foundation.

0027-0644/78/0917-0924$05.00

$\mathbf{r}=(r_1, \ldots, r_n)$, where r_i is the probability assigned to event i $(i=1, \ldots, n)$. An observation is a (row) vector $\mathbf{d}_k=(d_{k1}, \ldots, d_{kn})$, where $d_{ki}=1$ if $i=k$ and $d_{ki}=0$ if $i\neq k$. A categorical forecast $\mathbf{r}_k$ denotes a forecast that assigns probability one to the kth event (i.e., $\mathbf{r}_k=\mathbf{d}_k$).

b. General definitions

Let $\mathbf{C}=\{c_{ij}\}$ denote an $n\times n$ symmetric and positive definite matrix. Further, let $\mathrm{QSR}_k(\mathbf{r},\mathbf{C})$ denote the general QSR for the forecast $\mathbf{r}$ and an observation $\mathbf{d}_k$. Then,

$$\mathrm{QSR}_k(\mathbf{r},\mathbf{C})=(\mathbf{r}-\mathbf{d}_k)\mathbf{C}(\mathbf{r}-\mathbf{d}_k)', \quad (1)$$

or

$$\mathrm{QSR}_k(\mathbf{r},\mathbf{C})=\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(r_i-d_{ki})(r_j-d_{kj}). \quad (2)$$

It should be noted that the condition that $\mathbf{C}$ be symmetric is not a restriction. If $\mathbf{C}$ is not symmetric, then we can define $\mathbf{B}=(\mathbf{C}+\mathbf{C}')/2$, where $\mathbf{C}'$ is the transpose of $\mathbf{C}$ (i.e., $c_{ij}'=c_{ji}$). $\mathbf{B}$ is then a symmetric matrix, and it is easily seen from Eq. (2) that

$$\mathrm{QSR}_k(\mathbf{r},\mathbf{C}')=\mathrm{QSR}_k(\mathbf{r},\mathbf{C})$$

and that

$$\mathrm{QSR}_k(\mathbf{r},\mathbf{B})=(\tfrac{1}{2})[\mathrm{QSR}_k(\mathbf{r},\mathbf{C})+\mathrm{QSR}_k(\mathbf{r},\mathbf{C}')] =\mathrm{QSR}_k(\mathbf{r},\mathbf{C}).$$

The matrix $\mathbf{C}$ represents a matrix of weights. That is, c_{ij} is the weight assigned to the difference, or error, $(r_i-d_{ki})(r_j-d_{kj})$ $(i, j=1, \ldots, n)$. Differences among these weights can then be considered to reflect the relative importance of errors associated with different events or different combinations of events.

The family of QSR's can be written in a different form using the following result from linear algebra: A symmetric matrix $\mathbf{C}$ is positive definite if and only if there exists a nonsingular matrix $\mathbf{A}$ such that $\mathbf{C}=\mathbf{A}\mathbf{A}'$ (Graybill, 1969, pp. 317–318).[3] Thus, we can write

$$\mathrm{QSR}_k(\mathbf{r})=(\mathbf{r}-\mathbf{d}_k)\mathbf{A}\mathbf{A}'(\mathbf{r}-\mathbf{d}_k)', \quad (3)$$

where $\mathbf{A}$ is a nonsingular matrix.[4] Let

$$\mathbf{R}=\mathbf{r}\mathbf{A} \quad (4)$$

and

$$\mathbf{D}_k=\mathbf{d}_k\mathbf{A}. \quad (5)$$

Then,

$$\mathrm{QSR}_k(\mathbf{r})=(\mathbf{R}-\mathbf{D}_k)(\mathbf{R}-\mathbf{D}_k)'=\sum_{i=1}^{n}(R_i-D_{ki})^2. \quad (6)$$

It is of interest to note that $\mathbf{R}$ and $\mathbf{D}_k$ in Eqs. (4) and (5) are linear transformations, using the matrix $\mathbf{A}$, of $\mathbf{r}$ and $\mathbf{d}_k$, respectively. Moreover, the transformed observation vector $\mathbf{D}_k$ $(k=1, \ldots, n)$ is equal to the kth row of the transformation matrix $\mathbf{A}=\{a_{ij}\}$. That is, $D_{ij}=a_{ij}$ $(i, j=1, \ldots, n)$.

c. Special cases: the PS and the RPS

The PS can be defined as

$$\mathrm{PS}_k(\mathbf{r})=(\mathbf{r}-\mathbf{d}_k)(\mathbf{r}-\mathbf{d}_k)' \quad (7)$$

(Winkler and Murphy, 1968). Thus, the PS is the special case of the QSR in which both $\mathbf{C}$ and $\mathbf{A}$ are identity matrices. It should be mentioned that the term "quadratic scoring rule" has often been used as a synonym for the PS (e.g., Winkler and Murphy, 1968). In this paper, we use this term to denote the family of scoring rules defined by Eqs. (1) and (2), since the PS is not the only scoring rule of quadratic form.

The RPS can be defined as

$$\mathrm{RPS}_k(\mathbf{r})=(\mathbf{R}-\mathbf{D}_k)(\mathbf{R}-\mathbf{D}_k)', \quad (8)$$

where

$$R_i=\sum_{j=1}^{i} r_j \quad (i=1, \ldots, n) \quad (9)$$

and

$$D_{ki}=\sum_{j=1}^{i} d_{kj} \quad (i=1, \ldots, n) \quad (10)$$

(see Murphy, 1971). The matrix $\mathbf{A}$ that defines the linear transformations from $\mathbf{r}$ and $\mathbf{d}_k$ to $\mathbf{R}$ and $\mathbf{D}_k$, respectively, in this case is an upper triangular matrix with all of the elements on and above the principal diagonal equal to one. Murphy (1971) has referred to $\mathbf{R}$ as a cumulative (probability) forecast and $\mathbf{D}_k$ as a cumulative observation if event k occurs ($D_{ki}=0$ for $i<k$ and $D_{ki}=1$ for $i\geqslant k$).

3. Some properties of QSR's

a. Range of scores

Since the matrix $\mathbf{C}$ is assumed to be positive definite, it follows from Eq. (1) that

$$\mathrm{QSR}_k(\mathbf{r},\mathbf{C})\geqslant 0,$$

with equality if and only if $\mathbf{r}=\mathbf{d}_k$ (i.e., if and only if $\mathbf{r}=\mathbf{r}_k$). Thus, $\mathrm{QSR}_k(\mathbf{r},\mathbf{C})$ is equal to zero only when the forecast is perfect. Since a lower score is better, we say that QSR has a negative orientation (Winkler and Murphy, 1968).

The maximum value of $\mathrm{QSR}_k(\mathbf{r},\mathbf{C})$ occurs for some categorical forecast $\mathbf{r}_i$, where $i\neq k$. This value is equal to

$$\max_i\,(c_{ii}+c_{kk}-2c_{ik}).$$

For the PS, $c_{ij}=1$ for $i=j$ and $c_{ij}=0$ for $i\neq j$. Thus, the maximum value of the PS is two. In the case of the RPS, $c_{ij}>0$ for all i and j and the maximum value is equal to

$$\max_k\,(k-1, n-k).$$

[3] The matrix $\mathbf{A}$ need not be $n\times n$ but it must be of rank n. The use of an enlarged $n\times N$ $(N>n)$ matrix $\mathbf{A}$ may be instructive in certain situations. In this regard, the authors acknowledge the valuable comments of J. Hilden, University of Copenhagen.

[4] We will sometimes write $\mathrm{QSR}_k(\mathbf{r})$ instead of $\mathrm{QSR}_k(\mathbf{r},\mathbf{C})$ when it is clearly understood which matrix $\mathbf{C}$ is used.

b. Scores for categorical and uniform forecasts

For a categorical forecast $\mathbf{r}=\mathbf{r}_i$, the value of $\mathrm{QSR}_k(\mathbf{r},\mathbf{C})$ is equal to $c_{ii}+c_{kk}-2c_{ik}$. In the case of the PS, this value is zero if $i=k$ and two if $i\neq k$. For the RPS, the score for categorical forecasts is equal to $|i-k|$.

For a uniform forecast, $r_i=1/n$ for all i and we have

$$\mathrm{QSR}_k(\mathbf{r},\mathbf{C})=\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}[(1/n)-d_{ki}][(1/n)-d_{kj}]$$

or

$$\mathrm{QSR}_k(\mathbf{r},\mathbf{C})=(1/n^2)[\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}-2n\sum_{i=1}^{n} c_{ik}+n^2c_{kk}].$$

In the case of the PS ($c_{ij}=1$ for $i=j$ and $c_{ij}=0$ for $i\neq j$), the score for uniform forecasts is then $(n-1)/n$. For the RPS, the score for uniform forecasts is $(1/6n)\times[(n-1)(2n-1)-6(k-1)(n-k)]$.

c. Strictly proper QSR's

Let the (row) vector $\mathbf{p}=(p_1,\ldots,p_n)$ denote the forecaster's true belief (judgment). Then, $p_i(\geqslant 0)$ represents the forecaster's judgment that event i will occur $(i=1,\ldots,n)$. The stated forecast $\mathbf{r}$ may not be equal to the judgment $\mathbf{p}$. However, we (as evaluators) would like to use scoring rules that encourage the forecaster to make his forecast correspond exactly to his true belief (i.e., to make $\mathbf{r}=\mathbf{p}$). In this regard, let $S_k(\mathbf{r})$ denote the score assigned by the scoring rule S when event k occurs, and let $S(\mathbf{r},\mathbf{p})$ be the forecaster's (subjective) expected score, where

$$S(\mathbf{r},\mathbf{p})=\sum_{k=1}^{n} p_kS_k(\mathbf{r}). \tag{11}$$

Then, the scoring rule S is said to be strictly proper (Murphy and Epstein, 1967),[5] if

$$S(\mathbf{r},\mathbf{p})>S(\mathbf{p},\mathbf{p}) \quad \text{for all } \mathbf{r}\neq\mathbf{p}.^{6} \tag{12}$$

That is, S is strictly proper if and only if the forecaster's expected score is minimized when he sets $\mathbf{r}=\mathbf{p}$. We will now show that the QSR is strictly proper.

Let $\mathrm{QSR}(\mathbf{r},\mathbf{p})$ denote the forecaster's expected quadratic score. Then, from Eq. (1),

$$\mathrm{QSR}(\mathbf{r},\mathbf{p})=\sum_{k=1}^{n} p_k(\mathbf{r}-\mathbf{d}_k)\mathbf{C}(\mathbf{r}-\mathbf{d}_k)',$$

or

$$\mathrm{QSR}(\mathbf{r},\mathbf{p})=\sum_{k=1}^{n} p_k(\mathbf{rCr}'-\mathbf{rCd}_k'-\mathbf{d}_k\mathbf{Cr}'+\mathbf{d}_k\mathbf{Cd}_k'),$$

or

$$\mathrm{QSR}(\mathbf{r},\mathbf{p})=\mathbf{rCr}'-\mathbf{rCp}'-\mathbf{pCr}'+\sum_{k=1}^{n} p_kc_{kk},$$

or

$$\mathrm{QSR}(\mathbf{r},\mathbf{p})=(\mathbf{r}-\mathbf{p})\mathbf{C}(\mathbf{r}-\mathbf{p})'-\mathbf{pCp}'+\sum_{k=1}^{n} p_kc_{kk}. \tag{13}$$

Since $\mathbf{C}$ is positive definite, the first term in Eq. (13)

$$(\mathbf{r}-\mathbf{p})\mathbf{C}(\mathbf{r}-\mathbf{p})'\geqslant 0 \quad \text{for all } \mathbf{r},$$

with equality if and only if $\mathbf{r}=\mathbf{p}$, while the last two terms in Eq. (13) are independent of $\mathbf{r}$. Therefore,

$$\mathrm{QSR}(\mathbf{r},\mathbf{p})\geqslant\mathrm{QSR}(\mathbf{p},\mathbf{p}) \quad \text{for all } \mathbf{r},$$

with equality if and only if $\mathbf{r}=\mathbf{p}$. Thus, the general QSR is strictly proper.[7]

An alternative derivation of the family of strictly proper QSR's is presented in the appendix.

4. Geometrical representation of QSR's

a. The geometrical framework

Let π^n denote the set of all possible forecasts, where

$$\pi^n=\{\mathbf{r}\,|\,r_i\geqslant 0, \sum_i r_i=1; i=1,\ldots,n\}.$$

The set π^n represents a regular $(n-1)$-dimensional simplex. A regular simplex is a unit line segment when $n=2$, an equilateral triangle when $n=3$, and a regular tetrahedron when $n=4$. The n vertices of the simplex represent the n categorical forecasts $\mathbf{r}_k$ as well as the n possible observations $\mathbf{d}_k$ $(k=1,\ldots,n)$.

We shall denote the set of all possible transformed forecasts $\mathbf{R}$ $(=\mathbf{rA})$ by Π^n. Since $\mathbf{A}$ is a nonsingular matrix, it follows that Π^n is a (generally irregular) $(n-1)$-dimensional simplex. It also follows that a one-to-one correspondence exists between a point $\mathbf{r}$ in π^n and a point $\mathbf{R}$ in Π^n. The n vertices of the simplex Π^n represent the n transformed categorical forecasts $\mathbf{R}_k$ (where $\mathbf{R}_k=\mathbf{r}_k\mathbf{A}$) as well as the n transformed observations $\mathbf{D}_k$ $(=\mathbf{d}_k\mathbf{A})$. Moreover, since Π^n is a simplex, it follows that the maximum value of the general QSR occurs for a categorical forecast. Specifically, the maximum value occurs for the categorical forecast that corresponds to the vertex that is furthest away from the vertex that corresponds to the transformed observation (see Section 3.a).

Since the matrix $\mathbf{A}$ has $\mathbf{D}_k$ as its kth row (see Section 2.b), we can write Eq. (4) as

$$\mathbf{R}=\sum_{k=1}^{n} r_k\mathbf{D}_k. \tag{14}$$

[5] They used the term "proper." The present terminology was introduced by Murphy (1969b) and Staël von Holstein (1970).

[6] This definition assumes that S has a negative orientation. The inequality would be reversed for a scoring rule with a positive orientation.

[7] A reviewer has indicated that this result is a special case of a result obtained by DeGroot (1970, p. 234) within the context of "Bayes estimates."

That is, the transformed forecast $\mathbf{R}$ is a convex linear combination (i.e., a linear combination with non-negative weights) of the vectors corresponding to the vertices of Π^n. Thus, the simplex Π^n is a convex set.

Let $\Delta(\mathbf{R},\mathbf{D}_k)$ denote the (Euclidean) distance between $\mathbf{R}$ and $\mathbf{D}_k$ in this framework, where

$$\Delta(\mathbf{R},\mathbf{D}_k)=[\sum_{i=1}^{n}(R_i-D_{ki})^2]^{\frac{1}{2}}. \tag{15}$$

Thus, from Eqs. (6) and (15),

$$\mathrm{QSR}_k(\mathbf{r})=[\Delta(\mathbf{R},\mathbf{D}_k)]^2. \tag{16}$$

That is, the QSR is the square of the distance between $\mathbf{R}$ and $\mathbf{D}_k$.

De Finetti (1962) showed that the PS could be given a geometrical interpretation (see also Epstein and Murphy, 1965). The geometrical framework for the RPS was discovered by Murphy and Staël von Holstein (1975).

b. An example

As an example of the geometrical representation of QSR's, we consider a case with three events ($n=3$). Let

$$\mathbf{A}=\begin{matrix}1 & 1 & 0.5\\ 0 & 0.8 & 1\\ 0 & 0 & 1\end{matrix}$$

The simplex Π^3—the set of all possible transformed forecasts $\mathbf{R}$—for this case is depicted in a Cartesian coordinate system in Fig. 1 as a triangle bounded by solid lines. The equilateral triangle π^3, which represents the set of all forecasts $\mathbf{r}$, is bounded by dashed lines.

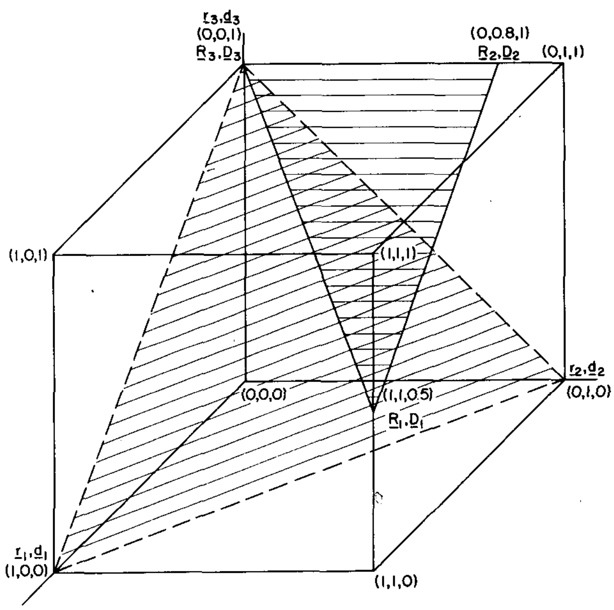

FIG. 1. The geometrical framework for the representation of forecasts and observations π^3 (bounded by dashed lines), and of transformed forecasts and observations Π^3 (bounded by solid lines). The latter corresponds to a particular transformation matrix $\mathbf{A}$ (see text).

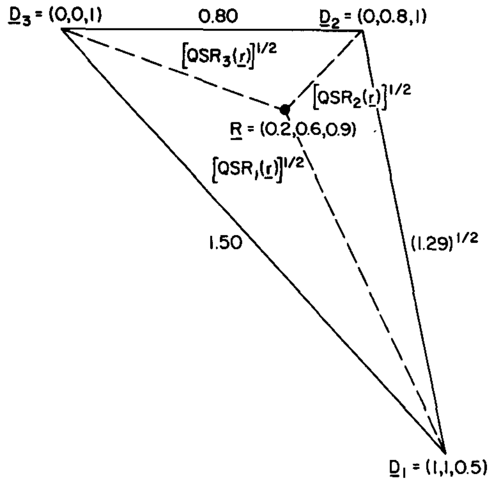

FIG. 2. The transformed forecast $\mathbf{R}=(0.2, 0.6, 0.9)$ depicted within the framework of the triangle Π^3 (see text).

Let $\mathbf{r}=(0.2, 0.5, 0.3)$. Then, the transformed forecast is $\mathbf{R}=(0.2, 0.6, 0.9)$, and this forecast is shown within the framework of the triangle Π^3 in Fig. 2. The dashed lines in Fig. 2 represent the distances between the transformed forecast $\mathbf{R}$ and the transformed observations $\mathbf{D}_1=(1, 1, 0.5)$, $\mathbf{D}_2=(0, 0.8, 1)$, and $\mathbf{D}_3=(0, 0, 1)$. The squares of these distances are $\mathrm{QSR}_1(\mathbf{r})$, $\mathrm{QSR}_2(\mathbf{r})$, and $\mathrm{QSR}_3(\mathbf{r})$, respectively.

c. Sensitive-to-distance QSR's

An evaluator may also be interested in whether or not the scoring rule of concern is sensitive to distance. To illustrate the sensitive-to-distance concept, consider two forecasts $\mathbf{r}$ and $\mathbf{r}^*$ in the case $n=4$, where $\mathbf{r}=(0.1, 0.5, 0.3, 0.1)$ and $\mathbf{r}^*=(0.3, 0.3, 0.3, 0.1)$. Further, suppose that the third event is observed on this occasion [i.e., $\mathbf{d}=\mathbf{d}_3=(0, 0, 1, 0)$]. Note that $\mathbf{r}$ assigns more probability than $\mathbf{r}^*$ to events "close" to the observed event, and a scoring rule that includes considerations of distance (in this sense) should give $\mathbf{r}$ a better score than $\mathbf{r}^*$. This property appears to be particularly important when the events represent an ordered variable (e.g., temperature, precipitation amount). The discussion that follows is based upon definitions formulated by Staël von Holstein (1970).

A forecast $\mathbf{r}^*$ is said to be more distant from the event that is observed than another forecast $\mathbf{r}$ if $\mathbf{r}^*\neq\mathbf{r}$ and if

$$\sum_{j=1}^{i} r_j^* \geqslant \sum_{j=1}^{i} r_j \quad (i=1, \ldots, k-1) \tag{17}$$

and

$$\sum_{j=i+1}^{n} r_j^* \geqslant \sum_{j=i+1}^{n} r_j \quad (i=k, \ldots, n), \tag{18}$$

where the event k is the observed event. Then, a scoring rule S is said to be sensitive to distance if

$$S_k(\mathbf{r}^*)>S_k(\mathbf{r}), \tag{19}$$

whenever $\mathbf{r}^*$ is more distant than $\mathbf{r}$ from the event that occurs. For the family of QSR's, Eq. (19) implies that the distance between $\mathbf{R}^*$ and $\mathbf{D}_k$ is greater than the distance between $\mathbf{R}$ and $\mathbf{D}_k$. Then, it follows that a QSR is sensitive to distance if $\mathbf{R}^*$ is further away than $\mathbf{R}$ from $\mathbf{D}_k$ (if event k occurs) whenever $\mathbf{r}^*$ is more distant than $\mathbf{r}$ from event k.

As indicated in Section 1, the PS is not and the RPS is sensitive to distance. The latter was proved by Staël von Holstein (1970). A simpler proof is possible, which follows directly from the definition of the RPS used in this paper. From Eq. (8), we have

$$\mathrm{RPS}_k(\mathbf{r}^*)=\sum_{i=1}^{n}(R_i^*-D_{ki})^2,$$

or

$$\mathrm{RPS}_k(\mathbf{r}^*)=\sum_{i=1}^{k-1}(R_i^*)^2+\sum_{i=k}^{n}(1-R_i^*)^2,$$

and, since $R_i^*\geqslant R_i$ for all $i=1,\ldots,k-1$ and $R_i^*\leqslant R_i$ for all $i=k,\ldots,n$ [see Eqs. (17) and (18)],

$$\mathrm{RPS}_k(\mathbf{r}^*)>\sum_{i=1}^{k-1}(R_i)^2+\sum_{i=k}^{n}(1-R_i)^2$$

or

$$\mathrm{RPS}_k(\mathbf{r}^*)>\mathrm{RPS}_k(\mathbf{r}).$$

We will illustrate the sensitive-to-distance concept for the case $n=3$, using the QSR defined in Fig. 2 as an example. Fig. 3 shows the set of all transformed forecasts $\mathbf{R}^*$ that correspond to forecasts $\mathbf{r}^*$ that are more distant (as well as the set of all $\mathbf{R}^*$ that correspond to forecasts that are less distant) than $\mathbf{r}=(0.2, 0.5, 0.3)$ —which corresponds to $\mathbf{R}=(0.2, 0.6, 0.9)$— for each of the three cases that the first, second and third event occurs. When $n=3$, the inequalities (17) and (18) define two lines through $\mathbf{r}$ that are parallel to two sides of the triangle π^3. The linear transformation $\mathbf{A}$ (from $\mathbf{r}$ to $\mathbf{R}$) maintains this property. Thus, the conditions that determine the corresponding boundaries in the transformed triangle Π^3 are once again two lines (through $\mathbf{R}$) that are parallel to two sides of this triangle.

An examination of the geometrical representation of different QSR's for the case $n=3$ indicates that a QSR will be sensitive to distance if the angle of the triangle Π^3 at $\mathbf{D}_2$ is greater than or equal to 90°. In this regard, the angle at $\mathbf{D}_2$ is equal to 90° in the case of the RPS (see Murphy and Staël von Holstein, 1975). It may be possible to extend this result to the general case in which Π^n is an $(n-1)$-dimensional simplex. As yet, we have been unable to make this extension.

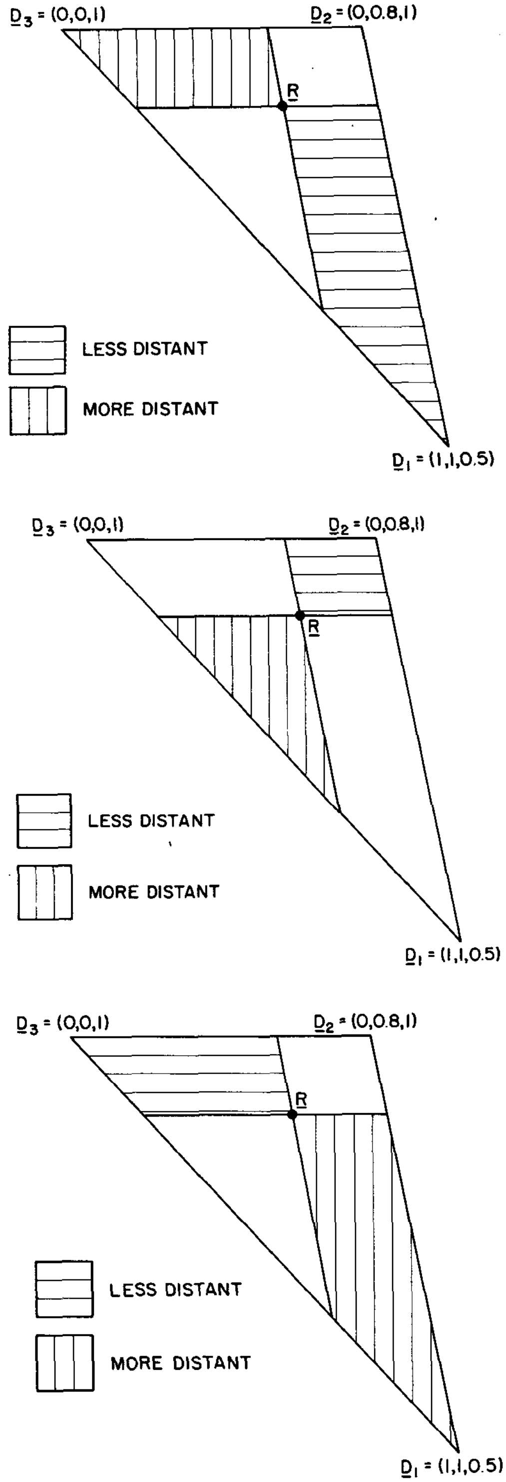

FIG. 3. Representation, within Π^3, of the sets of transformed forecasts that correspond to forecasts that are more and less distant than the forecast corresponding to $\mathbf{R}$ when (a) event 1 occurs, (b) event 2 occurs and (c) event 3 occurs.

Finally, the fact that the PS is not sensitive to distance can be readily demonstrated within this geometrical framework. Fig. 4 depicts the situation for the PS for the same forecast $\mathbf{r}=(0.2, 0.5, 0.3)$ when the first event occurs. It is clear from this figure that

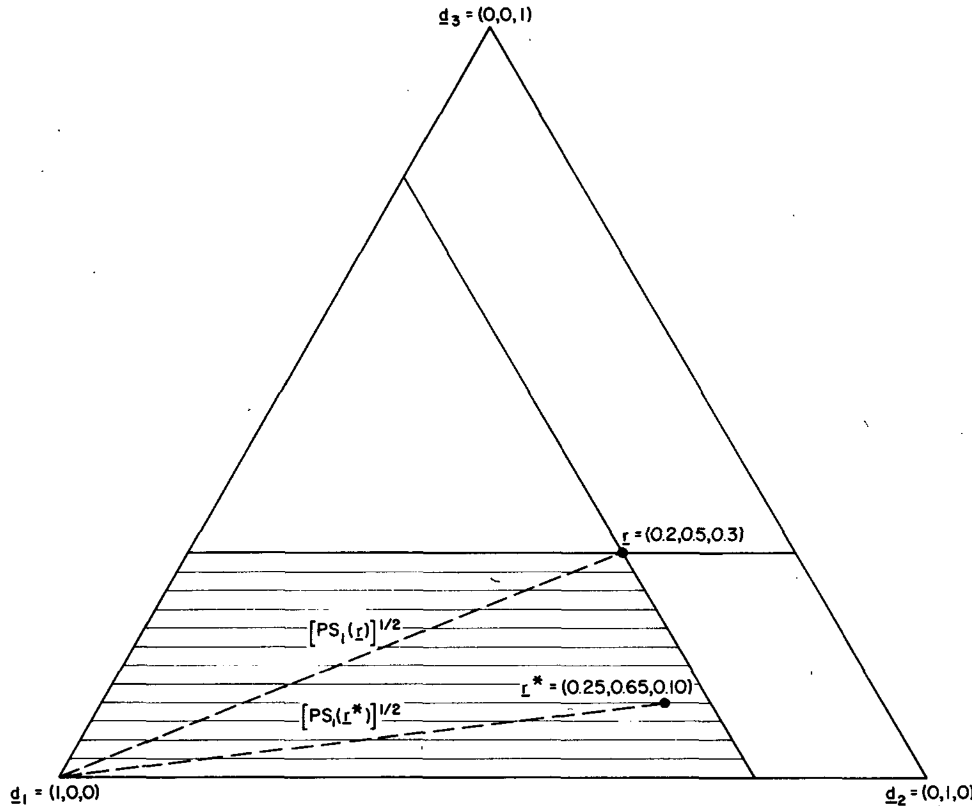

FIG. 4. Representation, within π^3, of forecasts $\mathbf{r}=(0.2,0.5,0.3)$ and $\mathbf{r}^*=(0.25,0.65,0.10)$ and the set of forecasts that is less distant than $\mathbf{r}$ when the first event occurs.

forecasts can be found that are less distant than **r** from **d**, but that receive a higher score. For example, the forecast $\mathbf{r}^*=(0.25, 0.65, 0.10)$ receives a score of 0.995, whereas **r** receives a score of 0.980.

5. Use of QSR's and QSR's as value measures

a. Selection of a QSR

In this section we briefly discuss some considerations related to choosing a QSR for use in a particular situation. These considerations involve the shape of the simplex Π^n, and it is convenient to explore these issues within the context of a specific example. In Fig. 5 we depict a particular simplex Π^n for the case $n=3$. This triangle has one side—the side opposite the point, or vertex, $\mathbf{D}_3$—that is much shorter than the other two sides. This situation implies that relatively large weights will be assigned to errors involving the third event, while much smaller weights will be associated with errors involving the first and second events. For example, a categorical forecast of the third event ($\mathbf{r}_3$) is penalized much more heavily than a categorical forecast of the first event ($\mathbf{r}_1$) when the second event occurs. Specifically, in this case, $\mathrm{QSR}_2(\mathbf{r}_1)=0.25$ and $\mathrm{QSR}_2(\mathbf{r}_3)=1.01$. Also, loosely speaking, the distribution of probability between the first and second events is not very important since the points $\mathbf{D}_1$ and $\mathbf{D}_2$ are quite close to one another.

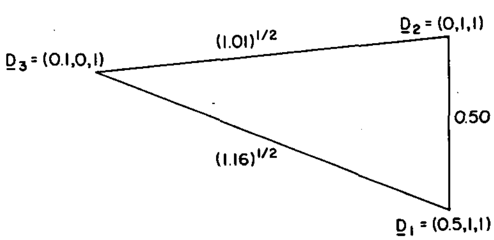

FIG. 5. The geometrical framework Π^3 for a QSR for which the third event is more important than the other two events.

Thus, by a judicious choice of the relative sizes of the sides of the triangle Π^3, the evaluator can reflect the perceived importance of forecasting different events. It is clear that this argument generalizes to situations involving more than three events. In this regard, it should be noted that the transformation matrix $\mathbf{A}=\{a_{ij}\}$ has $\mathbf{D}_k$—the transformed observation—as its kth row [i.e., $a_{ij}=D_{ij}$ (see Section 2b)]. Therefore, the evaluator can initiate the process of selecting an appropriate quadratic scoring rule by defining the shape of the simplex Π^n, and then he can work backwards to find the matrices **A** and **C** and the corresponding QSR.

b. QSR's *as measures of value*

It is of some interest to note that, under certain conditions, QSR's are equivalent (i.e., linearly related) to measures of the *ex post* value of probability forecasts. The measures of value of concern here are the so-called expected-utility measures formulated within the context of parameterized decision-making situations, in which the parameters are assumed to possess certain probability distributions. For example, Murphy (1966) showed that the PS is equivalent to the expected-utility measure in the familiar cost-loss ratio situation—a two-action, two-state decision-making situation involving a single parameter (the cost-loss ratio)—when the cost-loss ratio is assumed to be uniformly distributed. Epstein (1969) derived the RPS as an expected-utility measure within the framework of an M-action, M-state generalization of the cost-loss ratio situation. This situation also involved only one parameter (namely, the cost-loss ratio), and it was assumed to have a uniform distribution. More recently, Pearl (1975) has shown that the PS is equivalent to an expected-utility measure in a two-action, two-state situation involving two parameters. In this case, the parameters were assumed to be independent and to be exponentially distributed. For a more detailed discussion of these (and other) relationships between familiar scoring rules and certain expected-utility measures, refer to Murphy (1975).

In a more general vein, Savage (1971, p. 799) states that ". . . every proper scoring rule can be viewed as a share in a business and that every such share leads to at least a weakly proper scoring rule . . . ."[8] Thus, every QSR is related to some expected-utility measure. Unfortunately, it is extremely difficult to work backwards from a particular QSR to the "underlying" measure of value. Nevertheless, the existence of such relationships in certain situations provides an incentive to search for similar relationships in other situations, and in some cases it may provide an additional justification for the use of particular QSR's.

6. Summary and conclusion

In this paper, we have formulated a family of quadratic scoring rules (QSR's) and proved that QSR's are strictly proper. The probability score (PS) and the ranked probability score (RPS) were shown to be special cases of the general QSR. The sensitive-to-distance nature of QSR's was also investigated (earlier studies have demonstrated that the PS is not and the RPS is sensitive to distance), and some possible conditions defining the class of sensitive-to-distance QSR's were proposed.

A geometrical framework for the representation of forecasts, observations, and (QSR) scores was described. Within this framework, which consists of an irregular $(n-1)$-dimensional simplex, the forecasts and observations are represented by a suitably chosen linear transformation of the original forecasts and observations, and the score assigned by the QSR is represented by the square of the (Euclidean) distance between the point in the simplex that corresponds to the forecast of concern and the vertex of the simplex that corresponds to the relevant observation. The framework also provides useful insights into the properties of QSR's, including the sensitive-to-distance property. This framework represents a generalization of geometrical frameworks for the PS and the RPS that were described in previous papers.

A process by which an evaluator can formulate QSR's designed for particular purposes was described. An example of the use of this process was also presented. Finally, the relationships between certain QSR's and particular measures of the (*ex post*) value of probability forecasts were briefly examined.

The family of QSR's is quite rich and will allow the evaluator to define a scoring rule that assigns weights to the events of concern that reflect their relative importance.

[8] In our terminology, Savage's terms "proper" and "weakly proper" are equivalent to strictly proper and proper, respectively.

APPENDIX

An Alternative Derivation of Strictly Proper QSR's

The family of strictly proper QSR's can also be derived from a result obtained by Savage (1971). He states that a scoring rule is strictly proper if and only if

$$S_k(\mathbf{r}) = J(\mathbf{r}) - \sum_{i=1}^{n} r_i J_i'(\mathbf{r}) + J_k'(\mathbf{r}) + a_k, \qquad \text{(A1)}$$

where $J(\mathbf{r})$ is a strictly convex function, $J_i'(\mathbf{r})$ is the partial derivative of $J(\mathbf{r})$ with respect to r_i $(i=1, \ldots, n)$, and a_k is a constant that may depend on k but not on $\mathbf{r}$. The most general quadratic function of n variables is

$$J(\mathbf{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} u_{ij} r_i r_j + \sum_{i=1}^{n} v_i r_i + w.$$

This function is strictly convex if and only if the matrix $\mathbf{U} = \{u_{ij}\}$ is positive definite. The derivative of $J(\mathbf{r})$ is then

$$J_i'(\mathbf{r}) = \sum_{j=1}^{n} (u_{ij} + u_{ji}) r_j + v_i,$$

and Eq. (A1) becomes

$$S_k(\mathbf{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} u_{ij} r_i r_j + \sum_{i=1}^{n} v_i r_i + w - \sum_{i=1}^{n} r_i \Big[\sum_{j=1}^{n} (u_{ij} + u_{ji}) r_j + v_i\Big] + \sum_{j=1}^{n} (u_{kj} + u_{jk}) r_j + v_k + a_k,$$

or

$$S_k(\mathbf{r}) = w + v_k - \sum_{i=1}^{n}\sum_{j=1}^{n} u_{ij} r_i r_j + \sum_{j=1}^{n} (u_{kj} + u_{jk}) r_j + a_k. \quad \text{(A2)}$$

On the other hand, the general QSR is defined as

$$\mathrm{QSR}_k(\mathbf{r}) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(r_i - d_{ki})(r_j - d_{kj})$$

[see Eq. (2)] or

$$\mathrm{QSR}_k(\mathbf{r}) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} r_i r_j - \sum_{j=1}^{n} (c_{jk} + c_{kj}) r_j + c_{kk}. \quad \text{(A3)}$$

Thus, a comparison of Eqs. (A2) and (A3) indicates that, since the matrix $\mathbf{C} = \{c_{ij}\}$ is positive definite,

$$S_k(\mathbf{r}) = -\mathrm{QSR}_k(\mathbf{r}) + w + v_k - c_{kk} + a_k$$

or

$$S_k(\mathbf{r}) = -\mathrm{QSR}_k(\mathbf{r}) + z_k,$$

where $z_k = w + v_k - c_{kk} + a_k$. The term z_k can be set equal to zero without any loss of generality. Thus, we find that the two scoring rules are equivalent. More precisely, any scoring rule that is at the same time quadratic in $\mathbf{r}$ and strictly proper is of the form proposed in Eq. (1) except perhaps for a constant that does not depend upon the forecast $\mathbf{r}$. The scoring rule $S_k(\mathbf{r})$ derived from Savage's result has a positive orientation (i.e., a larger score is better), and it will yield scores that are less than or equal to zero.

REFERENCES

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.

DeGroot, M. H., 1970: *Optimal Statistical Decisions*. McGraw-Hill, 489 pp.

de Finetti, B., 1962: Does it make sense to speak of "good probability appraisers"? *The Scientist Speculates: An Anthology of Partly-Baked Ideas*, I. J. Good, General Ed., Basic Books, Inc., 357–364.

Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8**, 985–987.

——, and A. H. Murphy, 1965: A note on the attributes of probabilistic predictions and the probability score. *J. Appl. Meteor.*, **4**, 297–299.

Graybill, F. A., 1969: *Introduction to Matrices with Applications in Statistics*. Wadsworth Publishing Co., 372 pp.

Murphy, A. H., 1966: A note on the utility of probabilistic predictions and the probability score in the cost-loss ratio decision situation. *J. Appl. Meteor.*, **5**, 534–537.

——, 1969a: On the ranked probability score. *J. Appl. Meteor.*, **8**, 988–989.

——, 1969b: A note on proper and strictly proper scoring rules. Ann Arbor, University of Michigan, Department of Meteorology and Oceanography, unpublished manuscript, 10 pp.

——, 1970: The ranked probability score and the probability score: A comparison. *Mon. Wea. Rev.*, **98**, 917–924.

——, 1971: A note on the ranked probability score. *J. Appl. Meteor.*, **10**, 155–156.

——, 1975: Relationships between measures of the accuracy and value of probability forecasts: some recent results. Boulder, National Center for Atmospheric Research, unpublished manuscript, 35 pp.

——, and E. S. Epstein, 1967: A note on probability forecasts and "hedging." *J. Appl. Meteor.*, **6**, 1002–1004.

——, and C.-A. S. Staël von Holstein, 1975: A geometrical framework for the ranked probability score. *Mon. Wea. Rev.*, **103**, 16–20.

Pearl, J., 1975: An economic basis for certain methods of evaluating probabilistic forecasts. Los Angeles, University of California, School of Engineering and Applied Science, Rep. UCLA-ENG-7561, 17 pp.

Savage, L. J., 1971: Elicitation of personal probabilities and expectations. *J. Amer. Statis. Assoc.*, **66**, 783–801.

Staël von Holstein, C.-A. S., 1970: A family of strictly proper scoring rules which are sensitive to distance. *J. Appl. Meteor.*, **9**, 360–364.

Winkler, R. L., and A. H. Murphy, 1968: "Good" probability assessors. *J. Appl. Meteor.*, **7**, 751–758.