NOTES AND CORRESPONDENCE

A Note on Gandin and Murphy's Equitable Skill Score

JOSEPH P. GERRITY, JR.

National Meteorological Center, Washington, D.C.

21 August 1991 and 2 March 1992

ABSTRACT

Gandin and Murphy introduced an "equitable skill score" for use in evaluating categorical forecasts. For forecasts involving more than two categories, the elements of the scoring matrix are not defined uniquely. In this note, a specific formula for the general multiple-category scoring matrix is presented and proven to satisfy the necessary conditions for "equitability." It is shown that, while it is not the only possible scoring matrix satisfying these necessary conditions, it is compatible with a logical condensation of the general K-category problem into a set of $K - 1$ two-category problems. Each of the two-category problems is associated with one of the $K - 1$ partitions defining the categories of the original problem.

1. Introduction

In their recent paper, Gandin and Murphy (1992) defined an "equitable" scoring matrix **S** for use in the verification of categorical forecasts. The rank of the scoring matrix equals the number of classes used in the categorization of the event being forecast. The elements of **S** are solely dependent on the marginal probabilities of occurrence of each category, or class. When attention is focused exclusively on the accuracy of the forecasts, rather than their utility for some specific application, the scoring matrix is symmetric.

Define a matrix **E** (contingency table) such that its element $e(i, j)$ is the relative frequency with which an observation falls in class i when the forecast predicted class j. In terms of **E** and **S**, Gandin and Murphy's "equitable" skill score ESS is defined as the trace of the matrix product of **S** transpose and **E**.

The number of criteria for the "equitability" of the score ESS depends upon the rank of **E** (i.e., the number of classes into which the event is categorized). For a K-class problem there are $K + 1$ criteria; namely, ESS must vanish for each of the K forecasting strategies that always predicts just one of the classes to occur, and ESS must equal unity for a perfect forecast. This set of criteria also implies that any forecast produced randomly will receive an equitable skill score of zero.

Using the $K + 1$ criteria, $K + 1$ of the $K(K + 1)/2$ distinct elements of the scoring matrix **S** may be determined. This leaves open the specification of the remaining $(K + 1)(K - 2)/2$ elements. Gandin and Murphy present an analysis of the constraints that restrict the choice of these remaining elements of the scoring matrix.

It is the purpose of this note to draw attention to a closed formula for the elements of **S**, which depends solely on the marginal probabilities giving the frequency of the occurrence of each of the K categories. The formula provides a scoring matrix that satisfies the "equitability" criteria and is free of ambiguity. A second objective is to demonstrate that the "equitable" score provided by the cited formula may also be computed as the arithmetic mean of $K - 1$ "equitable" scores, each of which applies to a two-class problem formulated in terms of the $K - 1$ thresholds partitioning the K classes of the original problem.

2. Closed formula for K-class equitable score

Let $P(r)$ be the relative frequency with which class r of the event is observed in a (large) sample of forecasts. Based on $P(r)$ define the following:

$$D(n) \equiv \frac{1 - \sum_{r=1}^{n} P(r)}{\sum_{r=1}^{n} P(r)} \qquad (1a)$$

$$R(n) = \frac{1}{D(n)}. \qquad (1b)$$

Observe that $D(n)$ is the ratio of the probability that an observation falls into a class with index greater than

Corresponding author address: Dr. Joseph P. Gerrity, Jr., NOAA/NMC, WWB, 5200 Auth Rd., Camp Springs, MD 20746.

n to the probability that it falls into a class with index less than or equal to n; $R(n)$ is the reciprocal of this ratio of probabilities. These parameters have an obvious meaning whenever the classes are ordinally related.

In terms of D and R, the elements of a K-class equitable scoring matrix may be written

$$s_{n,n} = \kappa\Big[\sum_{r=1}^{n-1} R(r) + \sum_{r=n}^{K-1} D(r)\Big]; \quad n = (1, 2, \cdots, K) \tag{2a}$$

$$s_{m,n} = \kappa\Big[\sum_{r=1}^{m-1} R(r) + \sum_{r=m}^{n-1} (-1) + \sum_{r=n}^{K-1} D(r)\Big]; \quad 1 \leqslant m < K, \quad m < n \leqslant K \tag{2b}$$

$$s_{n,m} = s_{m,n}, \quad 2 \leqslant n \leqslant K, \quad 1 \leqslant m < n \tag{2c}$$

$$\kappa \equiv \frac{1}{K-1}.$$

Equation (2a) gives the elements on the diagonal of $\mathbf{S}$. The remaining elements of the upper triangle are given by Eq. (2b). The lower triangle elements Eq. (2c) follow from the symmetry of $\mathbf{S}$.

When $K = 2$, $P(2) = 1 - P(1)$, and the equitable scoring matrix $\mathbf{S}$ has the elements:

$$s(1, 1) = P(2)/P(1) \quad s(1, 2) = -1$$
$$s(2, 1) = -1 \quad s(2, 2) = P(1)/P(2),$$

which is precisely the result given by Gandin and Murphy. They noted that in the two-class problem, the "equitable" skill score ESS is identical to the Hanssen and Kuipers discriminant (skill score) (Woodcock 1976). The ESS based on the scoring matrix defined in Eq. (2) may serve as a generalization of the Hanssen and Kuipers discriminant to the K-class problem.

By examination of Eq. (2b), it becomes clear that the greater the difference between the indices m and n, the greater the disparity between the forecast and observed categories, and the more often "minus 1" is added to the scoring weight. This results in the desirable property that the greater the categorical error, the heavier the penalty.

3. Origin of the closed form for the K-class scoring matrix

In the context of the verification of quantitative precipitation forecasts (QPFs), it is customary to calculate scores (e.g., threat score) for several thresholds and to display the scores graphically as a function of the thresholds. This practice leaves open the question of how to discriminate fairly (or equitably) between two methods of prediction that differ in performance across the thresholds. For example, the nested-grid model often achieves better threat scores for low thresholds (e.g., 0.01″) than for higher thresholds (e.g., 1″) when compared to the ETA model (Mesinger et al. 1990).

It was attractive, therefore, to apply the ESS to the QPF verification problem, in order to obtain a single score that equitably weights performance at the various thresholds. The general method outlined by Gandin and Murphy (1991) for the calculation of equitable scoring matrices for multiple-class problems (one threshold separating each class) seemed overly complicated.

Fortunately, it was discovered that when the ESS for a three-class categorization of the QPF was set to the arithmetic average of the ESS for the two two-class problems, generated by partitioning the contingency table at the thresholds separating its three classes, the resulting score satisfied the criteria for equitability. In turn, the ESS for a four-class problem could be set to the arithmetic average of the three ESS's applicable to the three-class problems generated by partitioning the four-class table at its three thresholds. Again it was possible to demonstrate algebraically that the resulting score satisfied the criteria for equitability.

Thanks to an observation made by Murphy on the form of the main diagonal of the three- and four-class scoring matrices, it was possible to find the pattern revealed in the form of the elements of the K-class equitable scoring matrix. The validity of the assertion that the scoring matrix defined by Eq. (2) is equitable was confirmed computationally by application to random contingency tables and perfect forecast contingency tables up to rank 6. An algebraic proof of the assertion is presented in the next section.

4. Algebraic proof of equitability

The first K criteria for equitability of the scoring matrix $\mathbf{S}$ are the following:

$$\mathrm{ESS} \equiv \mathrm{Tr}(\mathbf{S}^{\mathrm{T}}\mathbf{E}^{n}) = 0, \tag{3}$$

where $\mathbf{E}^n$ is any one of the K matrices containing the probabilities of occurrence $P(r)$ as its nth column with zeros elsewhere; for example, using the notation $P(r) \equiv P_r$ one has for $n = 2$:

$$\begin{pmatrix} 0 & P_1 & 0 & 0\cdots0 \\ 0 & P_2 & 0 & 0\cdots0 \\ 0 & P_3 & 0 & 0\cdots0 \\ \cdot & \cdot & \cdot & \cdots\cdot \\ 0 & P_K & 0 & 0\cdots0 \end{pmatrix}.$$

This corresponds to the condition that the expected value of ESS vanishes for forecasting strategies that always predict just one of the classes to occur; in the example, class 2 is always forecast.

The remaining criterion is that for a perfect forecast the ESS must equal unity, which may be expressed by

$$\mathrm{ESS} \equiv \mathrm{Tr}(\mathbf{S}^{\mathrm{T}}\mathbf{E}^{p}) = 1, \tag{4}$$

where $\mathbf{E}^p$ is given by:

$$\begin{pmatrix} P_1 & 0 & 0 \cdots 0 \\ 0 & P_2 & 0 \cdots 0 \\ 0 & 0 & P_3 \cdots 0 \\ \cdot & \cdot & \cdots\cdots \\ 0 & 0 & 0 \cdots P_K \end{pmatrix}.$$

Recalling that the matrix $\mathbf{S}$ is symmetric, Eqs. (3) and (4) may be written as Eqs. (5) and (6), respectively,

$$\mathrm{ESS} = \sum_{n=1}^{K} P(n) s_{m,n} = 0, \quad (m = 1, \cdots, K) \qquad (5)$$

and

$$\mathrm{ESS} = \sum_{n=1}^{K} P(n) s_{n,n} = 1. \qquad (6)$$

To demonstrate that Eq. (5) is satisfied, the general term $m = M$ may be selected and it must be shown that

$$\sum_{n=1}^{K} P(n) s_{M,n} = 0. \qquad (7)$$

Using the expression in Eq. (2) gives the following for the elements s in Eq. (7):

$$s_{M,1} = \frac{\sum_{r=1}^{M-1} (-1) + \sum_{r=M}^{K-1} D(r)}{K-1}$$

$$s_{M,2} = \frac{R(1) + \sum_{r=2}^{M-1} (-1) + \sum_{r=M}^{K-1} D(r)}{K-1}$$

$$s_{M,3} = \frac{\sum_{r=1}^{2} R(r) + \sum_{r=3}^{M-1} (-1) + \sum_{r=M}^{K-1} D(r)}{K-1}$$

$\cdots$

$$s_{M,M} = \frac{\sum_{r=1}^{M-1} R(r) + \sum_{r=M}^{K-1} D(r)}{K-1}$$

$\cdots$

$$s_{M,K-1} = \frac{\sum_{r=1}^{M-1} R(r) + \sum_{r=M}^{K-2} (-1) + D(K-1)}{K-1}$$

$$s_{M,K} = \frac{\sum_{r=1}^{M-1} R(r) + \sum_{r=M}^{K-1} (-1)}{K-1}. \qquad (8)$$

When Eqs. (8) are introduced into Eq. (7), the result may be written out as follows (the irrelevant factor $K - 1$ has been suppressed):

$$\mathrm{ESS} = P(1)[1 - M + \sum_{s=M}^{K-1} D(s)]$$
$$+ \sum_{r=2}^{M} P(r)[r - M + \sum_{s=1}^{r-1} R(s) + \sum_{s=M}^{K-1} D(s)]$$
$$+ \sum_{r=M+1}^{K} P(r)[M - r + \sum_{s=1}^{M-1} R(s) + \sum_{s=r}^{K-1} D(s)]. \qquad (9)$$

Reorganization of the summations gives

$$\mathrm{ESS} = \sum_{r=1}^{M-1} \{R(r)[1 - \sum_{s=1}^{r} P(s)] - \sum_{s=1}^{r} P(s)\}$$
$$+ \sum_{r=M}^{K-1} \{D(r)[\sum_{s=1}^{r} P(s)] - \sum_{s=r+1}^{K} P(s)\}. \qquad (10)$$

When Eqs. (1) for D and R are used in Eq. (10), it is readily seen that the expression for ESS vanishes for each value of the index r. This proves that the proposed K-class scoring matrix satisfies the first K criteria for equitability. The satisfaction of the remaining criterion may be demonstrated by introducing Eq. (2a) into Eq. (6) to obtain

$$\mathrm{ESS} = P(1)\frac{[D(1) + D(2) + \cdots + D(K-1)]}{K-1}$$
$$+ P(2)\frac{[R(1) + D(2) + \cdots + D(K-1)]}{K-1}$$
$$+ \cdots$$
$$+ P(K)\frac{[R(1) + R(2) + \cdots + R(K-1)]}{K-1}. \qquad (11)$$

Summing by columns and using the definitions of D and R [Eqs. (1)], it is easy to obtain the result ESS = 1 as required for equitability.

5. Computation of *K*-class ESS from two-class ESS's

As noted in section 3, it was found that an equitable skill score for the three- and four-class contingency tables could be constructed from the arithmetic average of two and three two-class partitions, respectively, of the original contingency tables. It is the goal of this section to demonstrate that this result is valid for the general case of an original K-class contingency table.

Let the K-class contingency table be represented by the matrix $\mathbf{E}$ as defined in the Introduction. Consider the set of $(K - 1)$ $2 \times K$ matrices $\mathbf{C}_n$ that have the elements $c_n(i, j)$ defined by:

$$c_n(1, j) = 1, \quad j \leqslant n < K \tag{12a}$$

$$c_n(1, j) = 0, \quad K \geqslant j > n \tag{12b}$$

$$c_n(2, j) = 0, \quad j \leqslant n < K \tag{12c}$$

$$c_n(2, j) = 1, \quad K \geqslant j > n. \tag{12d}$$

The partitioning of the matrix $\mathbf{E}$ into a set of $(K - 1)$ two-class matrices may be expressed by

$$\mathbf{E}^{(n)} = \mathbf{C}_n \mathbf{E} \mathbf{C}_n^{\mathrm{T}}, \quad 1 \leqslant n \leqslant K - 1 \tag{13}$$

or using the definition of $\mathbf{C}_n$ and $\mathbf{E}$,

$$\mathbf{E}^{(n)} = \begin{bmatrix} \sum_{k=1}^{n} \sum_{l=1}^{n} e(k, l) & \sum_{k=1}^{n} \sum_{l=n+1}^{K} e(k, l) \\ \sum_{k=n+1}^{K} \sum_{l=1}^{n} e(k, l) & \sum_{k=n+1}^{K} \sum_{l=n+1}^{K} e(k, l) \end{bmatrix}. \tag{13a}$$

The equitable scoring matrix for each 2×2 matrix is

$$S^{(n)} = \begin{bmatrix} D(n) & -1 \\ -1 & R(n) \end{bmatrix}, \tag{14}$$

where $D(n)$ and $R(n)$ are defined in Eq. (1). The equitable skill score for each of the $K - 1$ two-class partitions may be denoted by ESS(n),

$$\mathrm{ESS}(n) = \mathrm{Tr}[\mathbf{S}^{(n)} \mathbf{E}^{(n)}]. \tag{15}$$

It is asserted that an equitable skill score ESS for the K-class problem may be defined by the arithmetic mean,

$$\mathrm{ESS} = (K - 1)^{-1} \sum_{n=1}^{K-1} \mathrm{ESS}(n) \tag{16}$$

or

$$\mathrm{ESS} = (K - 1)^{-1} \sum_{n=1}^{K-1} \mathrm{Tr}[\mathbf{S}^{(n)} \mathbf{C}_n \mathbf{E} \mathbf{C}_n^{\mathrm{T}}]. \tag{17}$$

It is straightforward but tedious to demonstrate that the score provided by Eq. (17) is identical to that provided by applying the K-class equitable scoring matrix, defined by Eqs. (2), to the K-class matrix $\mathbf{E}$.

The first step in the demonstration involves the insertion of Eqs. (13a) and (14) into Eq. (17). With some little effort one may express ESS as follows,

$$\begin{aligned} \mathrm{ESS} = \frac{1}{K - 1} \sum_{n=1}^{K-1} \Big[D(n) \sum_{k=1}^{n} \sum_{l=1}^{n} e(k, l) \\ - \sum_{k=n+1}^{K} \sum_{l=1}^{n} e(k, l) - \sum_{k=1}^{n} \sum_{l=n+1}^{K} e(k, l) \\ + R(n) \sum_{k=n+1}^{K} \sum_{l=n+1}^{K} e(k, l) \Big]. \end{aligned} \tag{18}$$

Next one must expand Eq. (18) and collect the coefficients of each element $e(kl)$. The result is that the coefficients are identical to the elements of the scoring matrix $\mathbf{S}$ [defined in Eq. (2)] which multiply the element $e(kl)$ in the expansion of the trace of the matrix product of $\mathbf{S}$ transpose and $\mathbf{E}$. To save space, the detailed exposition is left for the interested reader.

6. Summary

An explicit formula has been presented for the calculation of Gandin and Murphy's (1992) equitable skill score for a general, K-class contingency table. An algebraic proof of the formula's validity was given. Finally, it was demonstrated that the calculation of the score may be carried out using the arithmetic average of the $K - 1$ "equitable" skill scores computed for the two-class problems generated by partitioning the original contingency table at its $K - 1$ thresholds.

It should be noted that although the equitable scoring matrix presented here has an attractive simplicity, it is by no means the only scoring matrix that can satisfy the equitability criteria for a K-class problem. The theoretical reasons for this have been adequately covered by Gandin and Murphy. As related to the scoring matrix presented in this note, it may be pointed out that an equitable score will result from any normalized linear combination of the $K - 1$ two-class equitable scores, each associated with one of the thresholds of the original K-class problem. In the absence of sufficient reason for an alternative weighting, the arithmetic mean seemed to be appropriate.

Acknowledgments. The insight and encouragement of Lev Gandin and Allan Murphy were essential to the conduct of this work. Thanks are also due to Thomas Schlatter, who carefully reviewed the manuscript and suggested points for clarification.

REFERENCES

Gandin, L., and A. H. Murphy, 1992: Equitable skill scores for categorical forecasts. *Mon. Wea. Rev.*, **120**, 361–370.

Mesinger, F., T. L. Black, D. W. Plummer, and J. H. Ward, 1990: ETA model precipitation forecasts for a period including Tropical Storm Allison. *Wea. Forecasting,* **5,** 483–493.

Woodcock, F., 1976: The evaluation of yes/no forecasts for scientific and administrative purposes. *Mon. Wea. Rev.*, **104**, 1209–1214.