

A Note on the Ranked Probability Score^{1,2}

ALLAN H. MURPHY

Dept. of Meteorology and Oceanography, The University of Michigan, Ann Arbor

4 August 1970 and 29 September 1970

The ranked probability score (RPS) (Epstein, 1969) is both *strictly proper* (Murphy, 1969)³ and *sensitive to distance* (Staël von Holstein, 1970),⁴ and, as a result, the RPS appears to be a particularly appropriate scoring rule for the evaluation of probability forecasts of *ordered* variables (Murphy, 1970). Recently, F. Sanders, in a personal communication,⁵ has described a scoring rule T (say), where

$$T(\mathbf{r}, \mathbf{d}) = (1/K) \sum_{i=1}^K (\sum_{k=1}^i r_k - \sum_{k=1}^i d_k)^2,$$

¹Supported in part by the National Science Foundation (Atmospheric Sciences Section) under Grant GA-1707.

²Contribution No. 187 from the Department of Meteorology and Oceanography, The University of Michigan.

³A scoring rule is strictly proper if a forecaster can maximize (or minimize, whichever is appropriate) his expected score if and only if he makes his forecasts correspond (exactly) to his judgments.

⁴A scoring rule is sensitive to distance if the scoring rule assigns a higher (i.e., "better") score to a forecast which is "less distant" from the event, or class, which occurs. Whether or not a particular scoring rule is sensitive to distance depends upon both the scoring rule and the definition of distance. For example, the RPS is sensitive to distance according to one, but not the other, of two (known) definitions of distance (Staël von Holstein, 1970; Murphy, 1970).

⁵According to Sanders, the scoring rule T was formulated about 1967 by R. J. Thompson.

which he believes also possesses these properties.⁶ The purposes of this note are to demonstrate: 1) that T is *equivalent*, i.e., linearly related, to the RPS, which means that T is both strictly proper and sensitive to distance; and 2) that, as a result, the RPS, as well as the probability score (PS) (Brier, 1950), are related to the difference between the forecast and the observed *probability* distributions, which provides an additional measure of support for the use of the RPS as an evaluation measure.

When class j occurs, $T(\mathbf{r}, \mathbf{d})$ becomes $T_j(\mathbf{r})$, where

$$T_j(\mathbf{r}) = (1/K) \left[\sum_{i=1}^{j-1} (\sum_{k=1}^i r_k)^2 + \sum_{i=j}^K (\sum_{k=1}^i r_k - 1)^2 \right],$$

OR

$$T_j(\mathbf{r}) = (1/K) \left[\sum_{i=1}^{j-1} (\sum_{k=1}^i r_k)^2 + \sum_{i=j}^{K-1} (\sum_{k=i+1}^K r_k)^2 \right].$$

Now,

$$\text{RPS}_j(\mathbf{r}) = 1 - [1/(K-1)] \left[\sum_{i=1}^{j-1} (\sum_{n=1}^i r_n)^2 + \sum_{i=j}^{K-1} (\sum_{n=i+1}^K r_n)^2 \right]$$

⁶The vector $\mathbf{r} = (r_1, \dots, r_K)$, where $r_k \geq 0$ and $\sum_k r_k = 1$, represents the forecast and the vector $\mathbf{d} = (d_1, \dots, d_K)$, where d_k equals one if class k occurs and zero otherwise, represents the observation.

(Murphy, 1970, p. 924). Thus,

$$T_j(\mathbf{r}) = [(K-1)/K][1 - \text{RPS}_j(\mathbf{r})].$$

Since a scoring rule which is equivalent to a scoring rule which is strictly proper and sensitive to distance is itself strictly proper (Winkler and Murphy, 1968) and sensitive to distance (Staël von Holstein, 1970), T is both strictly proper and sensitive to distance.

The standard expression for the PS is

$$\text{PS}(\mathbf{r}, \mathbf{d}) = \sum_{i=1}^K (r_i - d_i)^2.$$

Let $r(x_i)$ and $d(x_i)$ denote the (probability) mass functions [the discrete counterparts of (probability) density functions] for the forecast and the observation, respectively, where x_i represents the i th class of the variable of concern x . Then,

$$\text{PS}(\mathbf{r}, \mathbf{d}) = \sum_{i=1}^K [r(x_i) - d(x_i)]^2.$$

Thus, the PS represents the sum of the squares of the differences between the forecast and the observed (probability) mass functions. Now, since T is equivalent to the RPS, the RPS can be expressed simply as

$$\text{RPS}(\mathbf{r}, \mathbf{d}) = \sum_{i=1}^K \left(\sum_{k=1}^i r_k - \sum_{k=1}^i d_k \right)^2,$$

or, in terms of (probability) mass functions, as

$$\text{RPS}(\mathbf{r}, \mathbf{d}) = \sum_{i=1}^K \left[\sum_{k=1}^i r(x_k) - \sum_{k=1}^i d(x_k) \right]^2.$$

Note that the

$$\sum_{k=1}^i r(x_k)$$

and the

$$\sum_{k=1}^i d(x_k),$$

$i=1, \dots, K$, describe the forecast and the observed cumulative (probability) mass functions, respectively [the discrete counterparts of (probability) distribution functions]. Let $R(x_i)$, where

$$R(x_i) = \sum_{k=1}^i r(x_k),$$

and $D(x_i)$, where

$$D(x_i) = \sum_{k=1}^i d(x_k),$$

denote these functions. Then,

$$\text{RPS}(\mathbf{r}, \mathbf{d}) = \sum_{i=1}^K [R(x_i) - D(x_i)]^2.$$

Thus, the RPS represents the sum of the squares of the differences between the forecast and the observed cumulative (probability) mass functions. Therefore, the RPS, as well as the PS, are concerned with the (square of the) differences between the forecast and the observed probability distributions. The relationship between the RPS and these probability distributions provides 1) a convenient means of comparing the RPS and the PS, and 2) an additional measure of support for the use of the RPS as an evaluation measure for forecasts of ordered variables.

REFERENCES

- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1-3.
- Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8**, 985-987.
- Murphy, A. H., 1969: On the "ranked probability score." *J. Appl. Meteor.*, **8**, 988-989.
- , 1970: The ranked probability score and the probability score: A comparison. *Mon. Wea. Rev.*, **98**, 917-924.
- 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.