A New Decomposition of the Brier Score: Formulation and Interpretation

ALLAN H. MURPHY

Department of Atmospheric Sciences, Oregon State University, Corvallis, OR 97331

24 March 1986 and 3 May 1986

ABSTRACT

A new decomposition of the Brier score is described. This decomposition is based on conditional distributions of forecast probabilities given observed events, and, as a result, it differs in a fundamental way from most previous partitions of quadratic verification measures. The new decomposition consists of 1) a term involving the variances of the conditional distributions and 2) a term related to the mean errors in the forecasts, which involves the squared differences between the means of the conditional distributions and the respective mean observations (the latter are necessarily either zero or one). Decreases in these variances and/or mean errors generally lead to improvements in the Brier score. The decomposition may be useful in verification studies, since it appears to provide additional insight into the quality of probabilistic forecasts.

1. Introduction

The Brier score (Brier, 1950) is simply the mean-square error of probabilistic forecasts, and, as such, it is a measure of the accuracy of the forecasts. This measure possesses several desirable properties, and it is widely used—both in its original form and in the form of a skill score—to evaluate the results of probability forecasting programs (see Murphy and Daan, 1985). One such property relates to the fact that it is possible to partition or decompose the Brier score into measures of other attributes of the forecasts (and/or observations). For example, decompositions developed by Sanders (1963) and Murphy (1973) yield measures of reliability and resolution, two important attributes of probabilistic forecasts. Thus, these partitions provide valuable information regarding the quality of such forecasts beyond that embodied in the overall Brier score itself.

Partitions of the Brier score formulated to date have generally involved the conditional distributions of the observations (i.e., observed events) given the forecasts. For example, the measure of reliability in the partitions formulated by Sanders (1963) and Murphy (1973) is the weighted average of the squared differences between the conditional means of the observations (i.e., conditional observed relative frequencies) given the forecast probabilities and the probabilities themselves, where the weights are the relative frequencies with which the probability values are used. It is also possible to approach the verification problem by considering the conditional distributions of the forecast probabilities given the observed events, together with the sample relative frequencies of these events, and the primary purpose of this note is to describe a decomposition of the Brier score based on this alternative approach to forecast verification.

The new decomposition of the Brier score is formulated in section 2. In section 3 we provide interpretations for the terms in this decomposition and briefly discuss the relationships between these terms and other measures of forecast quality. Section 4 consists of a short summary and some concluding remarks.

2. Formulation of new decomposition

Consider a two-event—say, precipitation/no precipitation—situation involving a sample of n forecasts and the corresponding observations. Let r_i denote the probability of precipitation on the ith occasion and let d_i denote the observation of precipitation on the ith occasion, where $d_i = 1$ if precipitation occurs and $d_i = 0$ otherwise ($i = 1, \ldots, n$). In this situation, the average Brier score (BS) (Brier, 1950) can be expressed as follows:

$$\mathrm{BS} = (2/n) \sum_{i=1}^{n} (r_i - d_i)^2. \tag{1}$$

(The 2 in this expression arises because the scores associated with the occurrence and nonoccurrence of the events are equal.)

The sample of n forecasts (and observations) can be divided into two subsamples based on the occurrence and nonoccurrence of precipitation. Specifically, let n_1 denote the number of occasions on which $d_i = 1$ and let n_0 denote the number of occasions on which $d_i = 0$ ($n_1 + n_0 = n$). Moreover, let r_{1j} denote the forecast probability of precipitation on the jth of the n_1 occasions ($j = 1, \cdots, n_1$) and let r_{0k} denote the forecast probability of precipitation on the kth of the n_0 occasions ($k = 1, \cdots, n_0$). Then BS in (1) can be rewritten as follows:

$$BS = (2/n)\left[\sum_{j=1}^{n_1}(r_{1j}-1)^2 + \sum_{k=1}^{n_0}(r_{0k}-0)^2\right], \quad (2)$$

or

$$BS = (2/n)\left(\sum_{j=1}^{n_1} r_{1j}^2 + \sum_{k=1}^{n_0} r_{0k}^2 - 2\sum_{j=1}^{n_1} r_{1j} + n_1\right), \quad (3)$$

or, since

$$\bar{r}_1 \equiv (1/n_1)\sum_{j=1}^{n_1} r_{1j},$$

$$BS = (2/n)\left(\sum_{j=1}^{n_1} r_{1j}^2 + \sum_{k=1}^{n_0} r_{0k}^2 - 2n_1\bar{r}_1 + n_1\right). \quad (4)$$

Adding and subtracting $(2/n)(n_1\bar{r}_1^2 + n_0\bar{r}_0^2)$ on the right-hand side (RHS) of (4) $[\bar{r}_0 \equiv (1/n_0)\sum_{k=1}^{n_0} r_{0k}]$, this equation becomes

$$BS = (2/n)\left[\left(\sum_{j=1}^{n_1} r_{1j}^2 - n_1\bar{r}_1^2\right) + \left(\sum_{k=1}^{n_0} r_{0k}^2 - n_0\bar{r}_0^2\right)\right] + (2/n)(n_1\bar{r}_1^2 - 2n_1\bar{r}_1 + n_1 + n_0\bar{r}_0^2), \quad (5)$$

or, since

$$\mathrm{Var}(r_1) \equiv (1/n_1)\left(\sum_{j=1}^{n_1} r_{1j}^2 - n_1\bar{r}_1^2\right)$$

and

$$\mathrm{Var}(r_0) \equiv (1/n_0)\left(\sum_{k=1}^{n_0} r_{0k}^2 - n_0\bar{r}_0^2\right),$$

$$BS = (2/n)[n_1\,\mathrm{Var}(r_1) + n_0\,\mathrm{Var}(r_0)] + (2/n)[n_1(\bar{r}_1-1)^2 + n_0(\bar{r}_0-0)^2]. \quad (6)$$

Finally, let $\bar{d}_1 = n_1/n$ and $\bar{d}_0 = n_0/n$ ($\bar{d}_1 + \bar{d}_0 = 1$). Then BS in (6) can be rewritten as follows:

$$BS = 2[\bar{d}_1\,\mathrm{Var}(r_1) + \bar{d}_0\,\mathrm{Var}(r_0)] + 2[\bar{d}_1(\bar{r}_1-1)^2 + \bar{d}_0(\bar{r}_0-0)^2]. \quad (7)$$

The expression for BS in (7) represents the basic form of the new decomposition of the Brier score. However, it will be convenient for some purposes to have a symbolic (and simpler) expression for this decomposition. Thus, if we let $\mathrm{Var}(r) \equiv \bar{d}_1\,\mathrm{Var}(r_1) + \bar{d}_0\,\mathrm{Var}(r_0)$ and $\bar{E}^2(r) \equiv \bar{d}_1(\bar{r}_1 - 1)^2 + \bar{d}_0(\bar{r}_0 - 0)^2$, then

$$BS = 2\,\mathrm{Var}(r) + 2\bar{E}^2(r). \quad (8)$$

It should be noted that the decomposition in (8) represents a special case of a general and well-known result, according to which the mean-square error is the sum of the variance of the errors and the square of the mean error.

3. Interpretation and discussion

In interpreting the terms on the RHS of (7), it should be recognized that the subsamples of forecasts and observations on which the decomposition is based constitute two conditional distributions: (i) the distribution of forecast probabilities given the occurrence of precipitation and (ii) the distribution of forecast probabilities given the occurrence of no precipitation. From this perspective, it is evident that the two terms in the new decomposition represent simple functions of summary measures of these distributions. Specifically, the first term on the RHS of (7) is twice the weighted average of the variances of the two distributions, where the weights are the corresponding sample relative frequencies of precipitation and no precipitation, respectively. This term ranges from zero when $r_{1j} = \bar{r}_1$ for all j and $r_{0k} = \bar{r}_0$ for all k to one-half when $r_{1j} = 1$ on $n_1/2$ occasions, $r_{1j} = 0$ on $n_1/2$ occasions, $r_{0k} = 1$ on $n_0/2$ occasions and $r_{0k} = 0$ on $n_0/2$ occasions. Thus, a decrease in the variance of the forecasts in either subsample will lead to a decrease in the Brier score (recall that smaller Brier scores are better).

The second term on the RHS of (7) is twice the weighted average of the squared differences between the average subsample forecast probability and the corresponding average subsample observation (in this decomposition, the latter is necessarily either one or zero). Once again, the weights are the sample relative frequencies of the two events. This mean-error term ranges from zero when $r_{1j} = 1$ for all j and $r_{0k} = 0$ for all k to two when $r_{1j} = 0$ for all j and $r_{0k} = 1$ for all k. As expected, a decrease in the mean error associated with either subsample will lead to a decrease in the Brier score.

As an example of the use of this decomposition, we consider matched samples of objective and subjective precipitation probability forecasts formulated for approximately 17 locations in the Western Region of the National Weather Service during the 1984–85 cool season (October–March). These forecasts are valid for the period 12–24 hours after the 0000 GMT cycle time. Each sample consists of 2959 forecasts, and the overall sample relative frequency of measurable precipitation ($\bar{d}_1$) is 0.187. The average objective forecast probabilities for periods with and without measurable precipitation are 0.446 ($=\bar{r}_1$) and 0.107 ($=\bar{r}_0$), respectively. Corresponding values for the subjective forecasts are 0.532 and 0.103. With regard to the terms in the decomposition of BS, $\bar{E}^2(r) = 0.067$ and $\mathrm{Var}(r) = 0.030$ for the objective forecasts and $\bar{E}^2(r) = 0.050$ and $\mathrm{Var}(r) = 0.031$ for the subjective forecasts. Thus, the subjective forecasts exhibit a smaller mean error and a slightly larger variance than the objective forecasts. This result is consistent with the tendency for the extreme probabilities (say, 0%, 80%, 90%, 100%) to be used more frequently by the forecasters than by the objective forecasting system (e.g., see Murphy, 1985). The overall

Brier scores for the objective and subjective forecasts are 0.194 and 0.162, respectively, indicating that the accuracy of the latter is greater than that of the former for these samples.

Although the expression for the decomposition in (7) possesses an attractive and symmetric form, it should be noted that this decomposition can also be expressed as follows:

$$\mathrm{BS} = (2/n)\sum_{i=1}^{n} r_i^2 + 2\bar{d}_1(1 - 2\bar{r}_1) \qquad (9)$$

[see (4)]. The first term on the RHS of (9) is simply twice the mean square of *all* the forecast probabilities (i.e., the r_i), and the second term represents a measure of mean error involving the subsample of forecasts for which $d_i = 1$. The first term is always nonnegative and attains a minimum value when $r_i = \bar{r}$ for all

$$i\,[\bar{r} \equiv (1/n)\sum_{i=1}^{n} r_i].$$

On the other hand, the second term is negative (desirable) when $\bar{r}_1 > \frac{1}{2}$ and positive (undesirable) when $\bar{r}_1 < \frac{1}{2}$.

In formulating the new decomposition of the Brier score, we have taken a different approach than that employed in developing previous decompositions of this quadratic verification measure. Thus, it is not surprising that the terms in the new partition possess quite different interpretations than the terms in most earlier decompositions. However, the new decomposition is similar in some respects to the so-called covariance decomposition of the Brier score recently described by Yates and Curley (1985). The latter involves five terms, one of which corresponds to the variance term in the partition described here; the other four terms arise from a further decomposition of the mean-error term in this new partition.

It is also of interest to note that some similarities exist between this approach and the signal detection theory (SDT) approach to forecast verification described by Mason (1982). Specifically, Mason defines two measures of forecast quality of special significance in the SDT approach, and these measures involve the same basic quantities—the means and variances of the distributions of forecasts in the two subsamples—as the decomposition described here. On the basis of a preliminary study, however, no simple relationships appear to exist between the SDT measures and the terms in the new decomposition.

4. Conclusion

A new decomposition of the Brier Score has been described in this paper. Unlike most previous partitions, this decomposition is based on dividing the sample of forecasts and observations into subsamples on the basis of the occurrence of the respective events. Here, each subsample constitutes a conditional distribution of forecast probabilities given that a particular event has occurred. The two terms in the decomposition represent weighted averages of summary measures of these distributions. Specifically, one term involves the variances of the conditional distributions, with smaller variances generally leading to better Brier scores. The other term relates to the conditional means of the forecast probabilities and represents a measure of the mean error in the forecasts, with smaller errors yielding better Brier scores.

We also briefly discussed the relationships between the terms in this decomposition and measures of forecast quality based on other decompositions of the Brier score or on alternative approaches to forecast verification. These relationships warrant more detailed investigation. Although the terms in the new decomposition appear to provide additional insight into the quality of probabilistic forecasts, it will be necessary to gain some experience with this partition before it is possible to judge its ultimate usefulness in the context of forecast verification.

The decomposition of the Brier score presented here was formulated within the context of a two-event situation. However, it is relatively straightforward to extend this decomposition to cover multiple-event (two-or-more-event) situations. Moreover, an analogous decomposition can be developed for other probabilistic quadratic verification measures such as the ranked probability score (Epstein, 1969; Murphy, 1971).

Acknowledgments. I. B. Mason provided helpful comments on an earlier version of this paper. The data involving the objective and subjective precipitation probability forecasts were kindly provided by G. M. Carter. This work was supported in part by the National Science Foundation under Grant ATM-8507495.

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 probabilities of ranked categories. *J. Appl. Meteor.*, **8**, 985–987.

Mason, I. B., 1982: A model for the assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.

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

——, 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600.

——, 1985: Probabilistic weather forecasting. *Probability, Statistics, and Decision Making in the Atmospheric Sciences.* A. H. Murphy and R. W. Katz, Eds., Westview Press, 337–377.

——, and H. Daan, 1985: Forecast evaluation. *Probability, Statistics, and Decision Making in the Atmospheric Sciences.* A. H. Murphy and R. W. Katz, Eds., Westview Press, 379–437.

Sanders, F., 1963: On subjective probability forecasting. *J. Appl. Meteor.*, **2**, 191–201.

Yates, J. F., and S. P. Curley, 1985: Conditional distribution analyses of probabilistic forecasts. *J. Forecasting*, **4**, 61–73.