Scoring Rules and the Brier Score

A *scoring rule* is a mathematical tool used to evaluate probabilistic forecasts. A substantial field of study regarding scoring rules exists, as the simple question “What makes a scoring rule a good scoring rule?’ leads to many rich and complex questions in probability and mathematical modeling. For now, we will set aside such questions and focus on a particular scoring rule: the Brier Score.

The Brier Score is essentially a measure of the error in a probabilistic forecast. As such, a lower Brier Score indicates a better forecast. Here’s how it works.

Suppose you are making a forecast about tomorrow’s weather (Brier Scores have their origins in measuring the accuracy of weather forecasts) and you forecast a 70% chance of rain.

The key observation with Brier Scoring is that this is really two forecasts: the first forecast is a 70% chance of rain; the second is a 30% chance of no rain. The Brier Score computes the squared error in both forecasts and adds them.

For example, suppose it rains tomorrow. The probability of rain is now taken to be 100% (it rained!) and the probability of no rain is 0% (it rained!). Thus, the Brier Score is the sum of the squares of the errors of these two forecasts:

$$(1-0.7)^2+(0-0.3)^2=0.3^2+0.3^2=0.18$$

Notice that if you had forecasted an 80% chance of rain, your Brier Score would have been:

$$(1-0.8)^2+(0-0.2)^2=0.2^2+0.2^2=0.08$$

Here we see that a better forecast results in a lower Brier Score.

On the other hand, if you forecast a 70% chance of rain tomorrow and it does not rain, then the Brier Score for your forecast is

$$(0-0.7)^2+(1-0.3)^2=0.7^2+0.7^2=0.98$$

Here the probability of rain is now taken to be 0% (it did not rain) and the probability of no rain is 100%.

Here are a few simple questions to consider:

- What is the best possible Brier Score?
- What is the worst possible Brier Score?
- How might you implement Brier Scoring in a forecast of an event with more than two possible outcomes?

Patrick Honner
www.PatrickHonner.com

Now, consider the following scenarios in the context of Brier Scoring.

Scenario 1

Suppose you and your team are forecasting the outcome of repeated tosses of a fair coin.

Teammate A suggests forecasting tails with probability 50%, arguing that since it's equally likely that the coin is either heads or tails, this is the most realistic forecast.

Teammate B suggests forecasting tails with probability 100% in order to achieve a better Brier Score on the successful flips.

Evaluate these two arguments. What do you think your team should do?

Scenario 2

Suppose you and your team are forecasting the outcome of repeated tosses of a coin, which you have reason to believe is weighted to show tails 75% of the time.

Teammate A suggests forecasting tails with probability 75%, again arguing that this is the most realistic forecast.

Teammate B suggests forecasting tails with probability 100%, again arguing that you should aim to optimize your Brier Score on the successes.

Teammate C suggests forecasting tails with probability 50%, to hedge against the possibility of flipping heads.

Teammate D suggests forecasting tails with probability 90%, as a compromise between Teammate A's strategy and Teammate B's strategy.

What do you think of these arguments? What should your team do?

Follow-Up to Scenario 2

After 6 tosses of the coin from Part 2, the results are 3 Heads and 3 Tails. How might this influence your decision making?

Afterwards, check out this [interactive Scratch program]() that demonstrates how the Brier Score is a *proper* scoring rule. That is, to optimize your score, it is best to bet your true belief! There is no edge to be gained by betting anything else.