A Note on Probability Aggregation and Conditional Accuracy Measures
Brian Weatherson - April 25, 2011

Assume that we care about a set of n propositions, $\{p_1, ..., p_n\}$. Let w be a world, and for any world w and proposition p, define $T(p, w)$ as p is p is true in w, and $\neg p$ if p is false in w. And say $V(p, w)$ is 1 if p is true in w, and 0 otherwise. Define the conditional inaccuracy of a probability function Pr in world w as follows:

$$A(\text{Pr}, w) = \sum_{i=1}^{n} (V(p_i, w) - \text{Pr}(p_i | T(p_1, w) \wedge \cdots \wedge T(p_{i-1}, w) \wedge T(p_{i+1}, w) \ldots T(p_n, w)))^2$$

We will say $A(\text{Pr})$ in a world w is $A(\text{Pr}, w)$.

This is basically a Brier score measure for conditional probabilities. For any world, and any proposition, we look at the difference between the truth value of that proposition and the conditional probability of that proposition. The condition for the conditional probability is the long conjunction that sets the truth value of each of the other propositions to their truth value in that world.

Apart from using conditional probabilities, I've made this different from a regular Brier score in a couple of ways. First, I've just focussed on the 'positive' propositions, $p_1, ..., p_n$, not the 'negative' ones, $\neg p_1, ..., \neg p_n$. But adding them in would just double the value of A, since $(V(p, w) - \text{Pr}(p, w))^2 = ((1 - V(p, w)) - \text{Pr}(\neg p, w))^2$. And I haven't used a sum here not an average; i.e., I haven't divided the whole thing by n. But since we're only going to care about comparative accuracy over a fixed set of propositions, that won't matter a lot.

Let's consider a very simple case for now, just because the algebra is easy. We'll consider just the case where $n = 2$. And instead of calling them p_1 and p_2, we'll call the propositions p and q. Imagine that we have two probability functions Pr_1 and Pr_2. And according to both Pr_1 and Pr_2, p and q are probabilistically independent. We'll write x_i and y_i for $\text{Pr}_i(p)$ and $\text{Pr}_i(q)$ respectively, for $i \in \{1, 2\}$.

Now assume that we want to 'blend' those two functions into a single function Pr. For any given function Pr we can work out its expected conditional inaccuracy according to both Pr_1 and Pr_2. Our aim is to minimize the average of these two expected inaccuracies.

If we were measuring inaccuracy by looking at just the probabilities of p and q, and not their conditional probabilities, it is fairly easy to show that we would do this by letting Pr be a simple average of Pr_1 and Pr_2. But I would have guessed this isn't true when we are measuring conditional inaccuracy, because we might to better by, say, taking the averages of the conditional probabilities, and working out the unconditional probabilities from those. The point of this note is to investigate these claims.

To simplify matters, let's introduce some shorthand. We'll say $z_1 = \text{Pr}(p|q), z_2 = \text{Pr}(p|\neg q), z_3 = \text{Pr}(q|p)$ and $z_4 = \text{Pr}(q|\neg p)$. And we'll identify the four salient worlds, (really world classes) as follows: $w_1 = p \wedge q, w_2 = p \wedge \neg q, w_3 = \neg p \wedge q$ and $w_4 = \neg p \wedge \neg q$. Now we can work out some values for $A(\text{Pr}, w)$.

$$\begin{aligned}
A(\text{Pr}, w_1) &= (\text{Pr}(p|q) - 1)^2 + (\text{Pr}(q|p) - 1)^2 \\
&= (z_1 - 1)^2 + (z_3 - 1)^2 \\
A(\text{Pr}, w_2) &= (\text{Pr}(p|\neg q) - 1)^2 + \text{Pr}(q|p)^2 \\
&= (z_2 - 1)^2 + z_3^2 \\
A(\text{Pr}, w_3) &= \text{Pr}(p|q)^2 + (\text{Pr}(q|\neg p) - 1)^2 \\
&= z_1^2 + (z_4 - 1)^2 \\
A(\text{Pr}, w_4) &= \text{Pr}(p|\neg q)^2 + \text{Pr}(q|\neg p)^2 \\
&= z_2^2 + z_4^2
\end{aligned}$$

For any random variable X, we'll say $E_i(X)$ is the expected value of X according to Pr_i for $i \in \{1, 2\}$, and $E_3(X) = {}^{E_1(X) + E_2(X)}/2$. Then $E_1(A(\text{Pr}))$ is

$$
\begin{aligned}
&\sum_{i=1}^{4} \Pr{}_i(w_i)A(\Pr, w_i)\\
&=x_1y_1((z_1-1)^2+(z_3-1)^2)+x_1(1-y_1)((z_2-1)^2+z_3^2)+(1-x_1)y_1(z_1^2+(z_4-1)^2)+(1-x_1)(1-y_1)(z_2^2+z_4^2)\\
&=z_1(x_1y_1z_1-2x_1y_1+(1-x_1)y_1z_1)+\\
&\quad z_2(x_1(1-y_1)z_2-2x_1(1-y_1)+(1-x_1)(1-y_1)z_2)+\\
&\quad z_3(x_1y_1z_3-2x_1y_1+x_1(1-y_1)z_3)+\\
&\quad z_4((1-x_1)y_1z_4-2(1-x_1)y_1+(1-x_1)(1-y_1)z_4)+\\
&\quad 2x_1y_1+x_1(1-y_1)+(1-x_1)y_1\\
&=z_1(y_1z_1-2x_1y_1)+z_2((1-y_1)z_2-2x_1(1-y_1))+z_3(x_1z_3-2x_1y_1)+z_4((1-x_1)z_4-2(1-x_1)y_1)+x_1+y_1
\end{aligned}
$$

A similar calculation gives us this value for $E_2(A(\Pr))$:

$$z_1(y_2z_1-2x_2y_2)+z_2((1-y_2)z_2-2x_2(1-y_2))+z_3(x_2z_3-2x_2y_2)+z_4((1-x_2)z_4-2(1-x_2)y_2)+x_2+y_2$$

Now we want to minimize the average of these two sums, subject to the constraint that each z_1 must in in $[0,1]$. (Note: I think that's enough to ensure Pr is a probability function, but I need to double check this.) Since no terms in the summation include more than one z_i, we can approach the four minimsation tasks indepdently. Rather than try to minimise $E_3(A(\Pr))$, which is what we're officially trying to do, we'll minimise $2E_3(A(\Pr))$; this just saves having to deal with too many fractions!

Let's start with z_1. The relevant term in the sum we're trying to minimise is:

$$z_1(y_1z_1-2x_1y_1)+z_1(y_2z_1-2x_2y_2)$$

If we differentiate with respect to z_1, the resulting derivative is:

$$2y_1z_1-2x_1y_1+2y_2z_1-2x_2y_2=z_1(2y_1+2y_2)-2x_1y_1-2x_2y_2$$

And that is 0 when

$$z_1=\frac{x_1y_1+x_2y_2}{y_1+y_2}$$

And note that the second derivative is just $2y_1+2y_2$, which is non-negative, so we've got a genuine minimum here, not merely a turning point.

Now we'll do z_2. The relevant term we're trying to minimise is:

$$z_2((1-y_1)z_2-2x_1(1-y_1))+z_2((1-y_2)z_2-2x_2(1-y_2))$$

If we differentiate with respect to z_2, the resulting derivative is:

$$2(1-y_1)z_2-2x_1(1-y_1)+2(1-y_2)z_2-2x_2(1-y_2)=z_2(2(1-y_1)+2(1-y_2))-2x_1(1-y_1)-2x_2(1-y_2)$$

And that is 0 when

$$z_2=\frac{x_1(1-y_1)+x_2(1-y_2)}{(1-y_1)+(1-y_2)}$$

And note that the second derivative is just $2(1-y_1)+2(1-y_2)$, which is non-negative, so we've got a genuine minimum here, not merely a turning point.

I won't go through the other two cases in detail, but similar arguments show that the inaccuracy is minimised when z_3 and z_4 are:

$$
\begin{aligned}
z_3&=\frac{x_1y_1+x_2y_2}{x_1+x_2}\\
z_4&=\frac{(1-x_1)y_1+(1-x_2)y_2}{(1-x_1)+(1-x_2)}
\end{aligned}
$$

And the only way to have z_1 to z_4 to take those values is for Pr to be the linear average of $\Pr_1$ and $\Pr_2$. That is, $\Pr(p \wedge q) = (\Pr_1(p \wedge q) + \Pr_2(p \wedge q))/2$, and so on for the other Boolean combinations of p and q. That this is sufficient to get z_1 to z_4 to work out that way is a simple algebra exercise. That this is necessary follows from the fact that once you set the values of the four conditional probabilities, the whole probability function is determined. For instance, the following equation must hold:

$$\Pr(p) = \frac{z_2 + z_1 z_4 - z_2 z_4}{1 - z_1 z_3 + z_1 z_4 + z_2 z_3 - z_2 z_4}$$

Working out the rest of Pr given z_1 through z_4 is left as an exercise, but it isn't especially challenging.

So we've proven the following result:

> Let $\Pr_1$ and $\Pr_2$ be probability functions such that p and q are probabilistically independent according to each. Say the inaccuracy of a probability function is measured by the conditional Brier score measure described on page 1. Then the probability function Pr that minimises the average expected inaccuracy according to $\Pr_1$ and $\Pr_2$ is the linear average of $\Pr_1$ and $\Pr_2$.

This is not what I would have expected. We are looking for some way to 'mix' various conditional probabilities. Normally, mixing fractions by averaging their numerators and denominators does not produce particularly sensible or interesting outcomes. But that seems to be what has happened here.

There are three big questions for further research that I haven't started to explore.

- Does the result hold up if we have more than two probability functions?
- Does the result hold up if we have more than two propositions?
- Does the result hold up if we replace the Brier score approach with some other (credence-eliciting) accuracy measure?

I don't know the answers to these questions, but even the simple result proved here seemed surprising enough to post.