

Playing the score – exploring beyond the hedge

Ian Jolliffe

University of Reading

Outline

- Hedging, propriety, equitability, consistency
 - What are they?
 - When are they relevant?
- Starting point/inspiration
 - ‘What does “hedging” a forecast mean ...’ by Laurie Wilson & Beth Ebert in the FAQ section of http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html

Metaverification

- The term ‘metaverification’ was used by Murphy (1996) to describe ‘determining whether or not verification measures [“scores” for brevity] satisfy specific criteria and/or possess particular properties’
- Such criteria include propriety, consistency and equitability, all of which relate to the avoidance of hedging



Hedging



- Hedging has a variety of definitions, but is commonly taken in everyday use to mean ‘placing bets on the opposite side in order to cut losses or guarantee a minimum amount of winnings’. In other words a forecast allows more than one (conflicting) possibility.
- The term is fairly well-known in meteorology, though not very often used in print. When it is, it is taken to mean that it occurs (Murphy, 1978) ‘whenever a forecaster’s judgement and forecast differ’.

To hedge or not to hedge

- ‘A meteorologist who prepares probability forecasts should not “hedge,” i.e. the meteorologist’s probabilities should express his true beliefs’
- ‘A meteorologist whose forecasts are evaluated with a particular scoring system can, and should, be expected to “hedge” to obtain the best possible score’
- Both quotations express plausible positions. Both are from Murphy & Epstein (1967), the latter deriving from a panel discussion reported in BAMS (1952)

Hedging and scores

- Hedging is used to make some sort of gain over what can be achieved without hedging
- In everyday usage the gain is financial
- In meteorology the gain is a better value or expected value of some score used to assess/verify forecasts. Hence hedging is ‘playing the score’.
- To make both quotations compatible we can restrict ourselves to using scores for which hedging is impossible – we need proper, or consistent, or perhaps equitable, scores

Propriety



- Next they should be taught about equitability and consistency.

Propriety II

- For probability forecasts, a (strictly) proper scoring system is one for which the forecaster obtains the best possible expected score by forecasting his/her true beliefs (and only by doing so) – Murphy & Epstein, 1967
- The Brier score is the best known proper score – there are others (logarithmic, spherical –Winkler & Murphy, 1968; Winkler, 1996), also plenty of theory and discussion (e.g Gneiting & Raftery, 2007, Winkler + discussants 1996). Equally, many scores (e.g. linear) are not proper.

Equitability



- ‘All’ unskilled forecasts should have the same expected score
- Not so obviously related to hedging as propriety
- But if a score is not equitable, it can be hedged in the sense that a forecaster who knows (s)he has little skill may do better using an unskilled forecast with a better expected score

Propriety and equitability

- No scoring system for probability forecasts can be both proper and equitable – noted at the last of these workshops (Montreal), and at last written up
- Given the choice, which would you prefer?
 - Propriety or equitability?

Equitability and probability forecasts

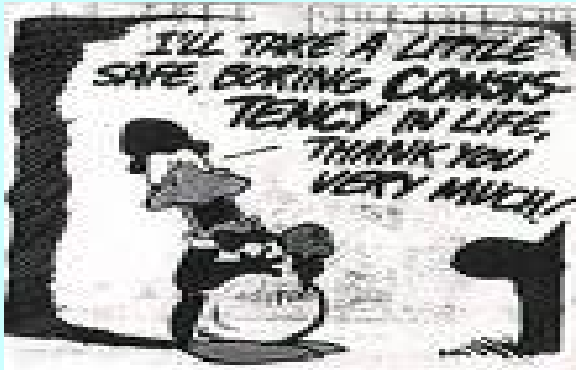
- Not only is equitability incompatible with propriety, but it is rather difficult to achieve equitability at all for probability forecasts
- Any score is a function $S(d)$ of the difference $d=f-o$, where f is forecast probability and o the corresponding observation, which is always 0 or 1
- If $S(d)$ is required to be symmetric [$S(d)=S(-d)$], then equitability is impossible unless the event being forecast has a base rate/climatology θ equal to 0.5
- In this case if the unskilled expected score is zero any score with $S(d) = -S(1-d)$ will be equitable, but the limitation to $\theta=0.5$ is somewhat restrictive

Equitability and probability forecasts II

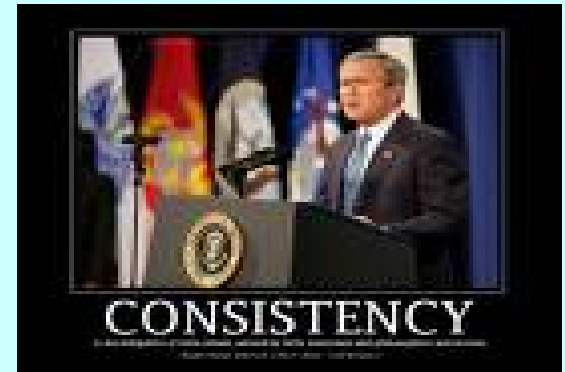
- What about allowing non-symmetry of $S(d)$?
- It is then relatively easy to get an equitable score
- Suppose again that the unskilled expected score is zero, and (arbitrarily) set $S(0)=-1$. Also let θ be the base rate and $r = \theta/(1-\theta)$.
- Then for $d>0$, the score
 - $S(d)=[d/(1-\theta)]-1$; $S(-d)=(d/\theta)-1$ is equitable
- But such scores rapidly become non-symmetric as θ and d move away from zero e.g for $\theta=0.6$, $S(d)=(5d/3)-1$ and $S(-d)=(5d/2)-1$. At the end of the range $S(1)=r$, $S(-1)=1/r$, equating to $3/2$ and $2/3$ in the example, and becoming 9 and $1/9$ for $\theta=0.9$.

Equitability and probability forecasts III

- Is such asymmetry ever desirable?
 - I can envisage the possibility of asymmetry based on cost/loss considerations, but for equitability, the nature of the asymmetry is tied to the base rate



Consistency



- For deterministic forecasts ‘consistency’ takes the place of ‘propriety’
- Like ‘hedging’ the meaning is slightly different from everyday usage
- There are ‘consistent’ forecasts – those that correspond with the forecaster’s judgments i.e. the forecaster does not hedge (Murphy, 1993)
- There are also ‘consistent’ performance measures (scores) ...

Consistent scores

- For the definition of consistency given by Murphy & Daan (1985) we need to assume that any forecaster really has a probability distribution for the variable to be forecast and a rule or directive determines the deterministic forecast to be made, given the forecaster's probability distribution
- Then a score is consistent with the directive if that score is minimised by forecasting using that directive. For example if the directive is 'forecast the mean' for a continuous variable, then mean square error is a consistent score as it is minimised by forecasting the mean.

Consistent scores - remarks

- The assumption of how the forecaster behaves (forecast deterministically when his/her beliefs are probabilistic) implies that the forecaster always hedges in the sense that true beliefs are not forecast
- However, it is the opposite of hedging in its everyday usage (make forecasts less definite), and is not done to improve a score
- As a slight digression, Murphy (1978) argues that ‘the desire to eliminate hedging should encourage forecasters to express ... forecasts in probabilistic terms’
- It seems to me that the definition of consistency could be turned around to say that a directive is consistent with a score, rather than a score consistent with a directive. Are others uneasy with the definition? What do we mean by hedging for deterministic forecasts?

Hedging for deterministic forecasts

- First what do we mean by hedging?
 - For probability forecasts it implies improving ‘expected score’. But expectation is with respect to the forecaster’s true beliefs.
- For deterministic forecasts either the forecaster’s true beliefs are
 - deterministic (and clearly wrong) or
 - Probabilistic and unknown
- So does hedging now imply improving actual score? Or is there another definition?

Back to equitability

- Although apparently not very useful for probability forecasts, it is often made a requirement for deterministic categorical forecasts
- Does equitability rule out hedging when hedging implies improvement of actual score?
- Does non-equitability necessarily imply that a score can be hedged?

Equitability – a conjecture

- Equitability ensures that hedging is impossible for deterministic categorical forecasts
- It works for the Pierce skill score, an equitable score for binary forecasts
 - Transferring a proportion of forecasts of an event to forecasts of no event or, conversely, transferring a proportion of no-event forecasts to ‘event’, reduces the Pierce skill score
 - But is this the only way that a forecaster can diverge from his/her true beliefs?

Non-equitable scores

- Non-equitable scores may or may not be hedged, depending on details of the data
- Consider ‘Proportion Correct’ in a (2x2) table, a non-equitable score, $(a+d)/n$
- If $a > b$ and $d > c$ the score cannot be improved by transferring a proportion of forecasts to non-forecasts, or vice versa; otherwise it can

	Observe event	Observe no event
Forecast event	a	b
Forecast no event	c	d

Questions? Or answers!

