The effect of ensemble size on verification measures for binary-event forecasts

Chris Ferro

Walker Institute
Department of Meteorology
University of Reading

- Brier score
- Effect of ensemble size
- RPS, Logarithmic and ROC scores

Brier score definition

Times	$t = 1, \ldots, n$
Binary observations	$I_t = 0$ or 1
Probabilistic forecasts	P_t

Brier score

$$B = \frac{1}{n} \sum_{t=1}^{n} (P_t - I_t)^2$$

Brier (1950, MWR)

Brier score definition

Times	$t = 1, \ldots, n$
Binary observations	$I_t = 0$ or 1
Probabilistic forecasts	$P_t =$ proportion of ensemble members that forecast the event

Brier score

$$B(m) = \frac{1}{n} \sum_{t=1}^{n} (P_t - I_t)^2$$

for ensemble size m.

Brier (1950, MWR)

Ensemble size effect

Assume
1. data are stationary
2. ensemble members are exchangeable

Expected Brier score

$$E[B(m)] = E[B(\infty)] + \frac{A}{m}$$

Richardson (2001, QJRMS)

A measures sharpness

$$A \propto \frac{1}{4} - E\left[\left(P - \frac{1}{2}\right)^2\right]$$

Ensemble size effect

Assume 1. data are stationary
2. ensemble members are exchangeable

Expected Brier score

$$E[B(m)] = E[B(\infty)] + \frac{A}{m}$$

Richardson (2001, QJRMS)

A measures sharpness

$$A \propto \frac{1}{4} - E\left[\left(P - \frac{1}{2}\right)^2\right]$$

Ensemble size effect

Assume
1. data are stationary
2. ensemble members are exchangeable

Expected Brier score

$$E[B(m)] = E[B(\infty)] + \frac{A}{m}$$

Richardson (2001, QJRMS)

A measures sharpness

$$A \propto \frac{1}{4} - E\left[\left(P - \frac{1}{2}\right)^2\right]$$

Estimate Brier score for other ensemble sizes

Given ensemble size m, estimate

$$
\begin{aligned}
E[B(M)] &= E[B(\infty)] + \frac{A}{M} \\
&= E[B(m)] - \frac{A}{m} + \frac{A}{M}
\end{aligned}
$$

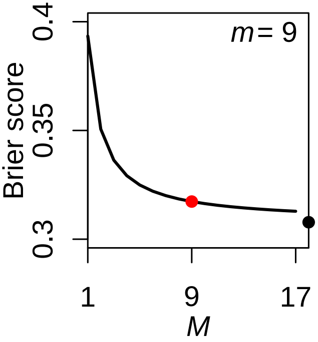
Unbiased estimator for $E[B(M)]$ is

$$B(m) - \frac{M-m}{M(m-1)n} \sum_{t=1}^{n} P_t(1-P_t)$$

Estimate Brier score for other ensemble sizes

Given ensemble size m, estimate

$$E[B(M)] = E[B(\infty)] + \frac{A}{M} = E[B(m)] - \frac{A}{m} + \frac{A}{M}$$



Unbiased estimator for $E[B(M)]$ is

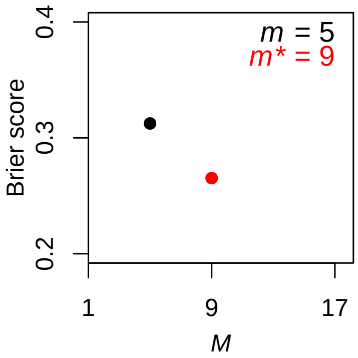
$$B(m) - \frac{M-m}{M(m-1)n} \sum_{t=1}^{n} P_t(1-P_t)$$

Compare Brier scores

Compare two sets of forecasts:

P_t with ensemble size m
P_t^* with ensemble size m^*

Identical verifying observations

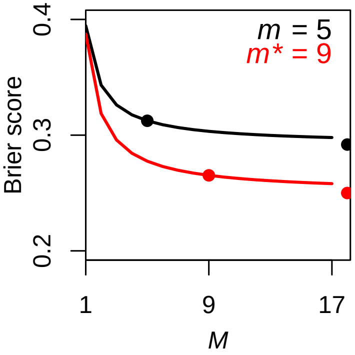


Compare Brier scores

Compare two sets of forecasts:

P_t with ensemble size m
P_t^* with ensemble size m^*

Identical verifying observations



Compare Brier scores

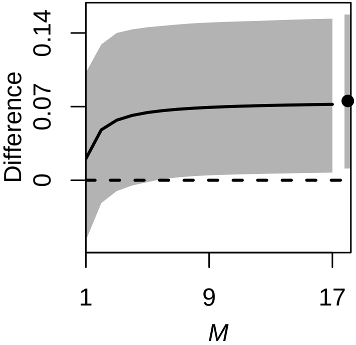
Compare two sets of forecasts:

P_t with ensemble size m
P_t^* with ensemble size m^*

Identical verifying observations

Bootstrap confidence interval for

$$E[B(M)] - E[B^*(M)]$$



Other scores

- Multi-category Brier / Ranked probability scores

$$\frac{1}{K}\sum_{k=1}^{K} B_k(m)$$

cf. Müller et al. (2005, JClim)

- Logarithmic score

$$S(m) = -\sum_{t=1}^{n}[I_t \log P_t + (1-I_t)\log(1-P_t)]$$

- Area $R(m)$ under ROC curve

- Unbiased estimators for $E[S(M)]$ and $E[R(M)]$ iff $M \leq m$

Other scores

- Multi-category Brier / Ranked probability scores

$$\frac{1}{K}\sum_{k=1}^{K} B_k(m)$$

cf. Müller et al. (2005, JClim)

- Logarithmic score

$$S(m) = -\sum_{t=1}^{n}[I_t \log P_t + (1-I_t)\log(1-P_t)]$$

- Area $R(m)$ under ROC curve

- Unbiased estimators for $E[S(M)]$ and $E[R(M)]$ iff $M \leq m$

Conclusion

- Easy to estimate effect of ensemble size on Brier scores
- Can use confidence intervals to compare scores
- Also possible for RPS, Logarithmic and ROC scores

Paper and R code available at

www.met.rdg.ac.uk/∼sws02caf

c.a.t.ferro@reading.ac.uk