

Discussion Paper No. 1182

**A PROOF OF CALIBRATION VIA
BLACKWELL'S APPROACHABILITY THEOREM**

by

Dean P. Foster*

Department of Statistics

Wharton School, The University of Pennsylvania

February 27, 1997

* Foster is also a visiting professor in the Center for Mathematical Studies in Economics and Management Science at Northwestern University, Winter 1997.

A proof of Calibration via Blackwell's Approachability Theorem

Dean P. Foster*

February 27, 1997

Abstract

Over the past few years many proofs of calibration have been presented (Foster and Vohra (1991, 1997), Hart (1995), Fudenberg and Levine (1995), Hart and Mas-Colell (1996)). Does the literature really need one more? Probably not, but this algorithm for being calibrated is particularly simple and doesn't require a matrix inversion. Further the proof follows directly from Blackwell's approachability theorem. For these reasons it might be useful in the class room.

*This work was done while I was visiting the Center for Mathematical Studies in Economics and Management Science, Northwestern University. Permanent affiliation: Dept. of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104. Email: foster@bellsark.wharton.upenn.edu.

Thanks to Sergiu Hart who provided the proof of the only result in the paper. The algorithm is a modification of the original algorithm in Foster and Vohra (1991).

Suppose at time t a forecast, f_t , is made which takes on the value of the midpoint of each of the intervals $[0, 1/m]$, $[1/m, 2/m]$, \dots , $[(m-1)/m, 1]$, namely, $\frac{2i-1}{2m}$ for i equals 1 to m . Let A_t^i the vector of indicators as to which forecast is actually made:

$$A_t^i = \begin{cases} 1 & \text{if } f_t = \frac{2i-1}{2m} \\ 0 & \text{otherwise} \end{cases}$$

Let X_t be the outcome at time t . We can now define the empirical frequency ρ_t^i as:

$$\rho_T^i = \frac{\sum_{t=1}^T X_t A_t^i}{\sum_{t=1}^T A_t^i}$$

Hopefully, ρ_T^i lies in the interval $[\frac{i-1}{m}, \frac{i}{m}]$. If so, the forecast is approximately calibrated. If not, I will measure how far outside the interval it is by two distances: \bar{d}_T^i and \bar{e}_T^i (for deficit and excess) which are defined as:

$$\begin{aligned} \bar{d}_T^i &= \frac{1}{T} \sum_{t=1}^T (\frac{i-1}{m} - X_t) A_t^i = [\frac{i-1}{m} - \rho_T^i] \bar{A}_T^i \\ \bar{e}_T^i &= \frac{1}{T} \sum_{t=1}^T (X_t - \frac{i}{m}) A_t^i = [\rho_T^i - \frac{i}{m}] \bar{A}_T^i \end{aligned}$$

where $\bar{A}_T^i = \sum A_t^i / T$. I will show that the following forecasting rule will drive both of these distances to zero:

1. If there exist an i^* such that $\bar{e}_T^{i^*} \leq 0$ and $\bar{d}_T^{i^*} \leq 0$, then forecast $\frac{2i^*-1}{2m}$.
2. Otherwise, find an i^* such that $\bar{d}_T^{i^*} > 0$ and $\bar{e}_T^{i^*-1} < 0$ then randomly forecast either $\frac{2i^*-1}{2m}$ or $\frac{2i^*+1}{2m}$ with probabilities:

$$P\left(f_{T+1} = \frac{2i^*-1}{2m}\right) = 1 - P\left(f_{T+1} = \frac{2i^*+1}{2m}\right) = \frac{\bar{d}_T^{i^*}}{\bar{d}_T^{i^*} + \bar{e}_T^{i^*-1}}$$

It is clear that an i^* can be found in step 2, since $i = 1$ always under forecasts and $i = m$ always over-forecasts.

The L-1 calibration score:

$$C_{1,T} \equiv \sum_{i=0}^m |\rho_t^i - \frac{2i+1}{m}| \bar{A}_T^i = \frac{1}{2m} + \sum_{i=1}^m \max(\bar{d}_T^i, \bar{c}_T^i)$$

so showing that all the \bar{c}_T^i and \bar{d}_T^i converge to zero, implies that $C_{1,T}$ converges to $\frac{1}{2m}$.

Theorem 1 (Foster and Vohra) *For all $\epsilon > 0$, there exists a forecasting method which is calibrated in the sense that $C_{1,T} < \epsilon$ if T is sufficiently large. In particular the above algorithm will achieve this goal if $m \geq \frac{1}{\epsilon}$.*

Consider this as a game between a statistician and nature. The statistician picks the forecast f_t and nature picks the data sequence X_t . The statistician's goal is to force all of the \bar{c}^i and \bar{d}^i to be negative (or at least approach this in the limit). Nature's goal is to keep the statistician from doing this. This set up is a game of "approachability" which was studied by Blackwell. He found a necessary and sufficient condition for a set to be approachable.

Theorem 2 (Blackwell 1956) *Let L_{ij} be a vector valued payoff taking values in R^n , where the statistician picks an i from \mathcal{I} at round i and nature picks a strategy j from \mathcal{J} at time t . Let G be a convex subset of R^n . Let $a \in R^n$ and let $c \in G$ be the closest point in G to the point a . Then G is approachable by the statistician if for all such a , there exist a weight vector w_i such that for all $j \in \mathcal{J}$,*

$$\left(\sum_{i \in \mathcal{I}} w_i L_{ij} - c \right)' (a - c) \leq 0. \quad (1)$$

To prove Theorem 1, we need to translate the calibration game into a Blackwell approachability game. The set of strategies for the statistician, \mathcal{I} ,

is the set of the m different forecasts. The set of strategies for nature, \mathcal{J} , is the set $\{0, 1\}$. Define

$$\begin{aligned} c_X^i &= (X - \frac{i}{m})A^i \\ d_X^i &= (\frac{i-1}{m} - X)A^i \end{aligned}$$

The vector loss is the vector of all the (d^i, c^i) 's, in otherwords, it is a point in R^{2m} . The goal set $G \subset R^{2m}$ is $G = \{x \in R^{2m} | (\forall k)x_k \leq 0\}$. Let $\bar{c}_T^i = \frac{1}{T} \sum_{t=1}^T c_{X_t}^i$ and $\bar{d}_T^i = \frac{1}{T} \sum_{t=1}^T d_{X_t}^i$. The $(d_X^i, c_X^i)_i$ will be our L_{ij} in the Blackwell game, and $(\bar{d}^i, \bar{c}^i)_i$ will be the point c . The closest point in G to the current average $a = (\bar{d}^i, \bar{c}^i)_{i \in \mathcal{I}}$ is

$$c = \left((\bar{d}^i)^-, (\bar{c}^i)^- \right)_{i \in \mathcal{I}}.$$

where we have defined the positive and negative parts as $x^+ = \max(0, x)$ and $x^- = \min(0, x)$. The weight vector w is the vector of probability of forecasting i/k .

Proof: (Hart 1996) Now to check equation (1). Writing it in terms of d^i 's and c^i 's equation (1) is:

$$\sum_{i=1}^m \left((w^i d^i - (\bar{d}^i)^-)(\bar{d}^i - (\bar{d}^i)^-) + (w^i c^i - (\bar{c}^i)^-)(\bar{c}^i - (\bar{c}^i)^-) \right) \leq 0$$

from $x - x^- = x^+$ equation (1) is equivalent to

$$\sum_{i=1}^m \left((w^i d^i - (\bar{d}^i)^-)(\bar{d}^i)^+ + (w^i c^i - (\bar{c}^i)^-)(\bar{c}^i)^+ \right) \leq 0$$

Since, $(x^-)(x^+) = 0$, it is sufficient to show:

$$\sum_{i=1}^m w^i (c^i (\bar{c}^i)^+ + d^i (\bar{d}^i)^+) \leq 0.$$

If step 1. of the algorithm is used the weight vector is just $w^{i^*} = 1$ if i^* is the forecast chosen and zero otherwise. So $w^i \neq 0$ only when both $(\bar{d}^i)^+$ and $(\bar{c}^i)^+$ are zero, so the entire sum is zero.

If step 2. is used, the non-zero terms are w^{i^*} and w^{i^*-1} . But, $(\bar{c}^{i^*})^+$ is zero and $(\bar{d}^{i^*-1})^+$ is zero. So, it is sufficient to show:

$$w^{i^*} d^{i^*} (\bar{d}^{i^*})^+ + w^{i^*-1} c^{i^*-1} (\bar{c}^{i^*-1})^+ \leq 0$$

But, $d^{i^*} = -c^{i^*-1}$, so it is sufficient to show:

$$w^{i^*} (\bar{d}^{i^*})^+ - w^{i^*-1} (\bar{c}^{i^*-1})^+ \leq 0$$

But, this follows (with equality) from the definition of our probabilities. \square

References

- Blackwell, D. (1956). "An analog of the minimax theorem for vector pay-offs." *Pacific Journal of Mathematics*, **6**, 1-8.
- Foster, D. and R. Vohra. (1991) "Asymptotic Calibration." technical report, Graduate School of Business, University of Chicago, Chicago IL.
- Foster, D. and R. Vohra. (1997) "Asymptotic Calibration." manuscript.
- Fudenberg, D. and D. Levine. (1997) "An easier way to calibrate." manuscript, 1995.
- Hart, S., personal communication, 1995.
- Hart, S. and A. Mas-Colell. (1997) "A simple adaptive procedure leading to correlated equilibrium."