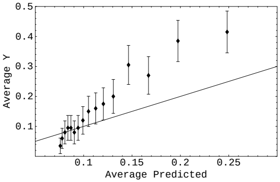


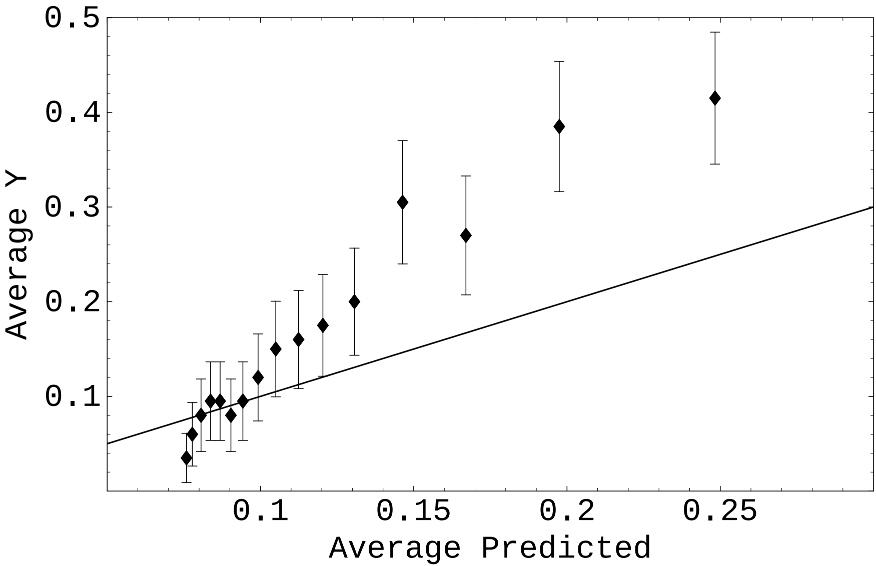
Calibration

Dean P. Foster

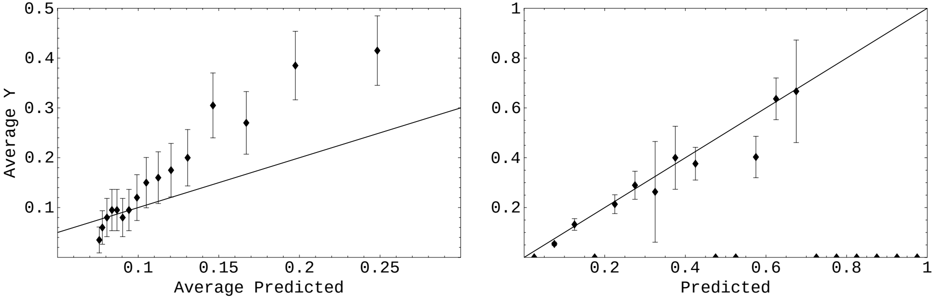
Outline

- Calibration for humans
- Calibration for big data
- Theory of calibrated
- Game theory:
  - Convergence to correlated equilibria
  - Convergence to NE

What is calibration?



Corrected by Pool Adjacent Violators



'Then you should say what you mean,' the March Hare.

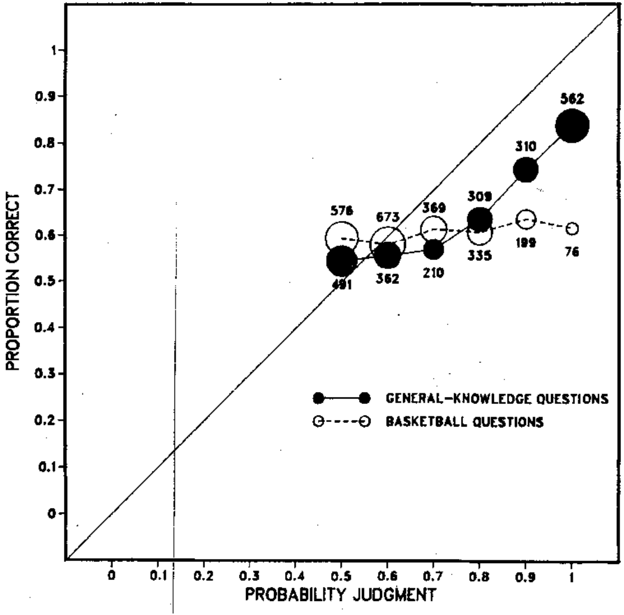
Calibration is unbiasedness

- Want $E(Y - \hat{Y}) \approx 0$.
- Actually we want more:

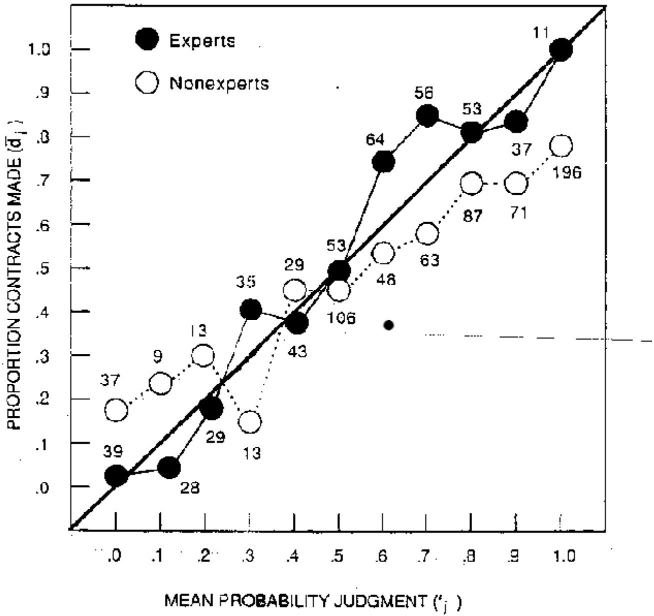
$$E(Y - \hat{Y} | \hat{Y} \approx c) \approx 0$$

  for all c.

Human behavior: without incentives



Human behavior: With incentives!



Calibration theory

"Suppose in a long (conceptually infinite) sequence of weather forecasts, we look at all those days for which the forecast probability of precipitation was, say, close to some given value p and then determine the long run proportion f of such days on which the forecast event (rain) in fact occurred. If $f = p$ the forecaster may be termed well calibrated."

Phillip Dawid

Calibration theory: example

Calibration is a minimal condition for performance

- On sequence: 0 1 0 1 0 1 0 ...
- A constant forecast of .5 is calibrated
- A constant forecast of .6 is not calibrated

Calibration theory: example

Calibration is a minimal condition for performance

- On sequence: 0 1 0 1 0 1 0 ...
- A constant forecast of .5 is calibrated
- A constant forecast of .6 is not calibrated
- Isn't a forecast of .1 .9 .1 .9 .1 .9 ... better?

Calibration theory: example

Calibration is a minimal condition for performance

- On sequence: 0 1 0 1 0 1 0 ...
- A constant forecast of .5 is calibrated
- A constant forecast of .6 is not calibrated
- Isn't a forecast of .1 .9 .1 .9 .1 .9 ... better?
  - Yes, it has higher "resolution."
  - But, it isn't calibrated.
  - Science calls it accuracy vs precision (or "trueness" as VIM says we should call it since 2008)

Calibration is easy to achieve

proof:
apply minimax theorem.

Calibration is easy to achieve

proof:
apply minimax theorem.

- Game: between the statistician and Nature.
- Natures strategy is a stochastic process.
- If the statistician knew the process she could easily “win.”
- By the minimax theorem she can always win.

Calibration is easy to achieve

proof:
apply minimax theorem.

- Game: between the statistician and Nature.
- Natures strategy is a stochastic process.
- If the statistician knew the process she could easily "win."
- By the minimax theorem she can always win.

Theorem (with Johnson 2013)

An exponential smooth close to calibrated.

Calibration in regression

Warm-up Goal: $E(Y - \hat{Y} | X = c) = 0$

- This can be guaranteed by doing a polynomial regression on X.

Calibration in regression

Warm-up Goal: $E(Y - \hat{Y}|X = c) = 0$

- This can be guaranteed by doing a polynomial regression on X.

But, what if $X = \hat{Y}$?

Calibration in regression

Real Goal: $E(Y - \hat{Y} | \hat{Y} = c) = 0$

- This can be guaranteed by doing a polynomial regression on $\hat{Y}$

Calibration in regression

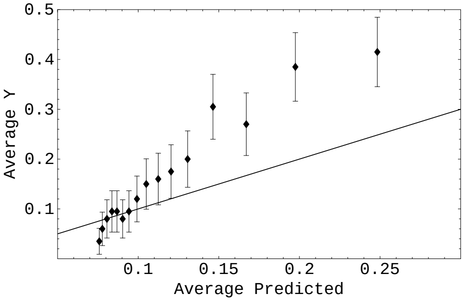
Real Goal: $E(Y - \hat{Y} | \hat{Y} = c) = 0$

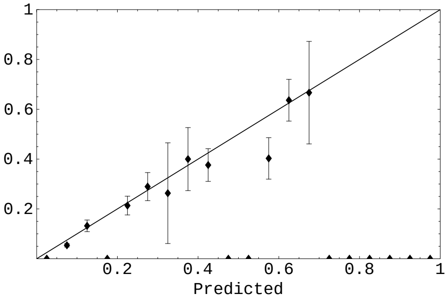
- This can be guaranteed by doing a polynomial regression on $\hat{Y}$

Computing $\hat{Y}$ now entails finding a fixed point.

Applied calibration

- First compute $Y \sim X$ to generate $\hat{Y}$
- Now fit a regression of Y on a polynomial of $\hat{Y}$
- Work really well!





Calibration theory: Paranoia

Tricking a forecasting method:

- If you predict $p > .5$, nature picks no rain
- If you predict $p \leq .5$ nature picks rain

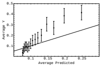
Calibration theory: Paranoia

Tricking a forecasting method:

- If you predict $p > .5$, nature picks no rain
- If you predict $p \leq .5$ nature picks rain
- But, if we treat .4999 and .5000 as about the same forecasts, then this attack fails.
  - Leads to different definitions
  - Leads to different algorithms

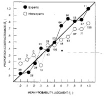
Summary so far: Handout

Talk on Calibration by Dean Foster



- Works well for big data since only costs a few more degrees of freedom.
- "Variable selection in data mining: Building a predictive model for bankruptcy," Foster and Stine, *JASA*, 2004.

- "Precision and Accuracy of Judgmental Estimation," Foster and Yaniv, *Journal of Behavioral Decision Making* (1997).
- "Graininess of Judgment Under Uncertainty: An Accuracy - informativeness Tradeoff," Foster and Yaniv *Journal of Experimental Psychology: General*, 1995.
- We looked at confidence intervals.
- Humans actually are responding to the social utility function.



"Suppose in a long (conceptually infinite) sequence of weather forecasts, we look at all those days for which the forecast probability of precipitation was, say, close to some given value p and then determine the long run proportion f of such days on which the forecast event (rain) in fact occurred. If $f = p$ the forecaster may be termed well calibrated."

Phillip Dawid

- "Asymptotic Calibration," Foster and Vohra, *Biometrika*, 1998.
- "A proof of Calibration via Blackwell's Approachability Theorem," Foster *GEB* 1999.
- "Regret in the On-line Decision Problem," Foster and Vohra, *GEB* 1999. (See also *AI-STATS* 2012 and *MOR* 2014.)
- "Deterministic Calibration and Nash Equilibrium" Foster and Kakade, *COLT*, 2004.

Convergence to Correlated Equilibrium

- "Calibrated Learning and Correlated Equilibrium," Foster and Vohra *Games and Economic Behavior*, 1997.
  - Playing calibrated forecasts will lead to correlated equilibria
  - Playing no-interal regret actions will converge to correlated equilibria
- Extended in "A general class of adaptive strategies, " by Hart and Mas-Colell 2001.

"If there is intelligent life on other planets, in a majority of them, they would have discovered correlated equilibrium before Nash equilibrium."

Roger Myerson

Convergence to Nash Equilibrium

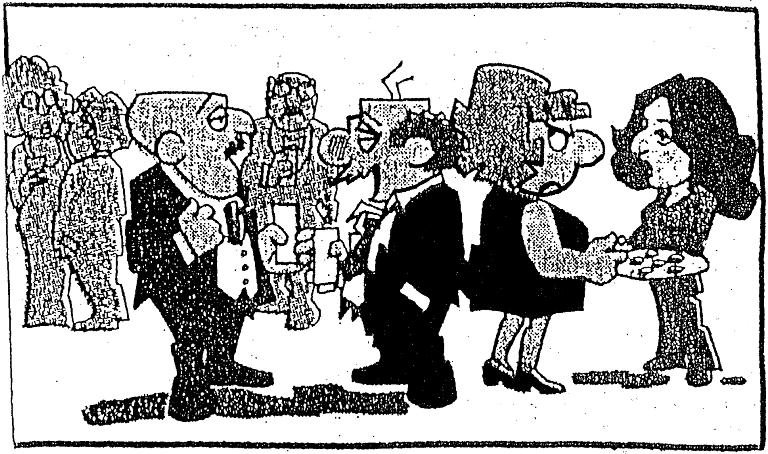
- Yes: You can learn NE from a grain of truth. (Kalai and Lehrer, 1993).
- No: Not exactly. (Nachbar 1997, Foster and Young 2001)
- Yes: Via exhaustive search–i.e. very slowly. (Foster and Young, 2006)
- No: Hart and Mas-Colell 2011.
- Yes: Via public, deterministic calibration which is very slow (Foster and Kakade, 2008, Foster and Hart, 2016)
- For all but the smallest games, it is basically no.

Recommendations

- Use isotonic link functions to calibrate regressions
- Use fixed point based calibration for time series
- Use no-internal regret for game theory
- Let go of Nash equilibrium

So, when is paranoia justifiable? Game theory

What is an equilibrium?



"LORETTA'S DRIVING BECAUSE I'M DRINKING, AND I'M DRINKING BECAUSE SHE'S DRIVING."

Fictitious play model

- The first player predicts the second player
- The second player predicts the first player
- Each plays a best reply to their predictions
- Called fictitious play

Convergences for fictitious play

- For zero sum games: it is easy (basically an interior point method for LP)
- For general games, calibration leads to correlated equilibrium
- Roger Meyerson: “2 out of 3 intelligent species discover Correlated equilibrium before Nash equilibrium.”

Convergences for fictitious play

- For zero sum games: it is easy (basically an interior point method for LP)
- For general games, calibration leads to correlated equilibrium
- Roger Meyerson: “2 out of 3 intelligent species discover Correlated equilibrium before Nash equilibrium.”
- Calibration is stronger than you need–it gets all forecasts right.

No internal regret

When asked if he had any regrets, Winston Churchill said, “I wish I’d bet on black every time I bet red and vice versa.”

No internal regret

When asked if he had any regrets, Winston Churchill said, “I wish I’d bet on black every time I bet red and vice versa.”

- $R^{i \to j}$ measures how much better off one would have been if all i’s were switched to j
- Find a stationary distribution of this flow (easy LP)
- It will end up having no-regrets in the long run
- It is better in many ways than using calibration

Recommendations

- Use calibration to clean up regressions
- Use fixed point based calibration to clean up time series predictions
- Use no-internal regret forecasts for game theory

Recommendations

- Use calibration to clean up regressions
- Use fixed point based calibration to clean up time series predictions
- Use no-internal regret forecasts for game theory

Thanks!