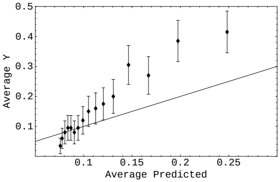


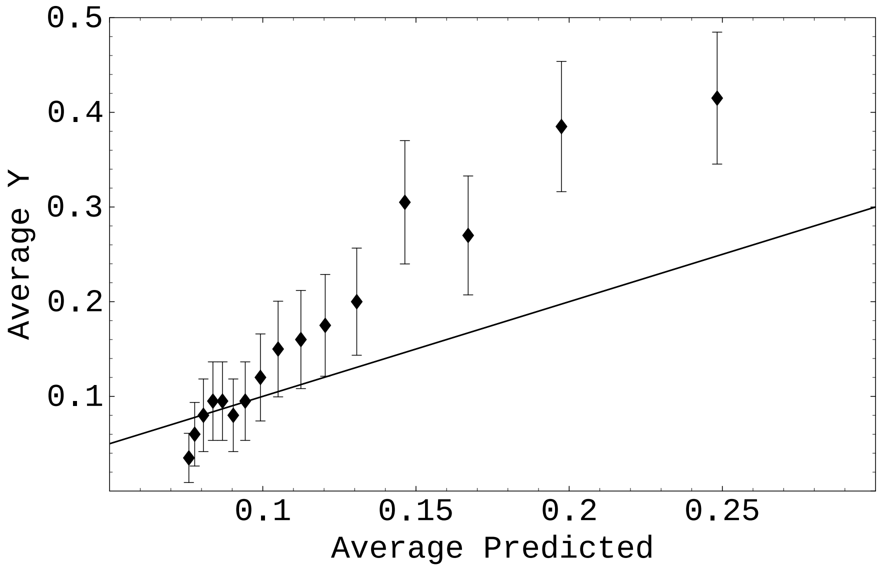
Calibration

Dean P. Foster
Amazon

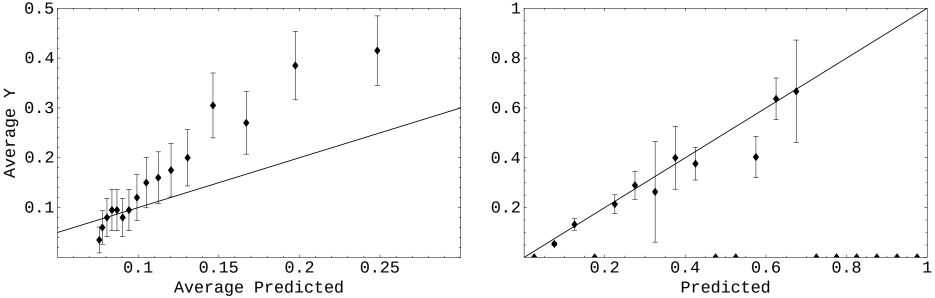
Outline

- 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.

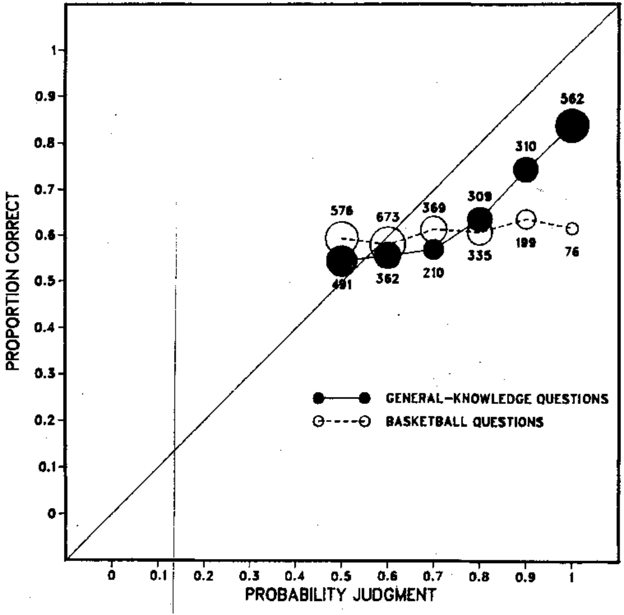
Calibration is unbiasedness

- 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



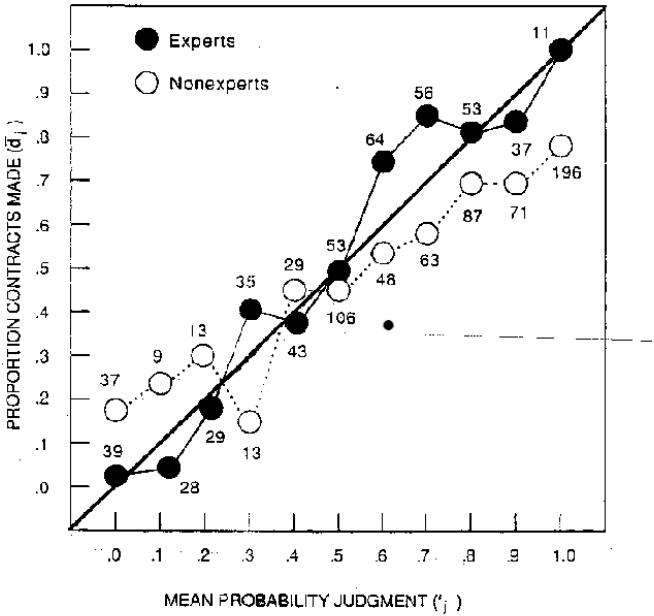
Human respond to incentives

- Classic over confidence can be elicited by: “Give me a 95% CI for the number of floors in the empire state building.”

Human respond to incentives

- Classic over confidence can be elicited by: “Give me a 95% CI for the number of floors in the empire state building.”
- How many covered 104? (without using your phone)
- But people are responding to real incentives: What will impress their friends!
- So need to figure out what impresses people.
- Turns out 50% coverage intervals impress more friends than 95% coverage intervals do. (Foster and Yaniv 1995)

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.

Calibration in regression

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

- Define $R = Y - \hat{Y}$
- saying: $E(R|X = c) = 0$ for all c is equivalent to
  - $E(R) = 0$
  - $E(RX) = 0$
  - $E(RX^2) = 0$
  - $E(RX^3) = 0$
  - $\cdots$
- This can be guaranteed by simply putting a polynomial of X into our regression.

Calibration in regression

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

- Define $R = Y - \hat{Y}$
- saying: $E(R|X = c) = 0$ for all c is equivalent to
  - $E(R) = 0$
  - $E(RX) = 0$
  - $E(RX^2) = 0$
  - $E(RX^3) = 0$
  - $\cdots$
- This can be guaranteed by simply putting a polynomial of X into our regression.

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

Calibration in regression

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

- Define $R = Y - \hat{Y}$
- Saying: $E(R|X = c) = 0$ for all c is equivalent to
  - $E(R) = 0$
  - $E(R\hat{Y}) = 0$
  - $E(R\hat{Y}^2) = 0$
  - $E(R\hat{Y}^3) = 0$
  - $\cdots$
- This can be guaranteed by “simply” putting a polynomial of $\hat{Y}$ into our regression.

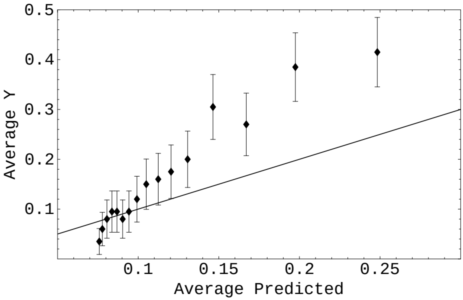
Calibration in regression

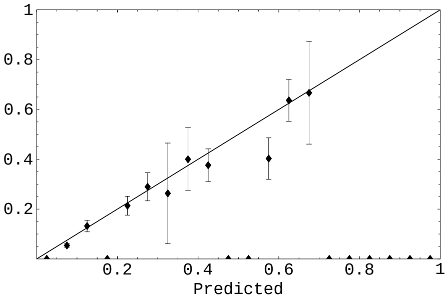
Real Goal: $E(Y - \hat{Y} | \hat{Y} = c) = 0$

- Define $R = Y - \hat{Y}$
- Saying: $E(R|X = c) = 0$ for all c is equivalent to
  - $E(R) = 0$
  - $E(R\hat{Y}) = 0$
  - $E(R\hat{Y}^2) = 0$
  - $E(R\hat{Y}^3) = 0$
  - $\cdots$
- This can be guaranteed by "simply" putting a polynomial of $\hat{Y}$ into our regression.
- Uff da: Computing $\hat{Y}$ now entails finding a fixed point.

Calibration in regression

- First compute $Y \sim X$ to generate $\hat{Y}$
- Now calibrate Y vs $\hat{Y}$
  - Isotonic regression
  - Empirically estimated link function
  - Pool adjacent violators (PAV)
  - Practical method: $Y \sim Poly(\hat{Y})$ for say a 5th degree polynomial





Calibration in time series

- We can do the same trick for predicting the next event
- It is easy to do the regression: We have previous $\hat{y}$'s and so can fit a polynomial to them.
- The hard part is finding the fixed point for the next round.

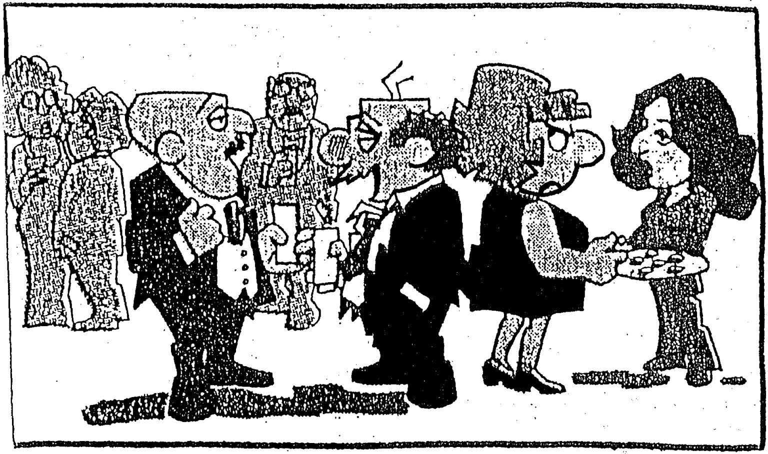
Calibration theory: Paranoia

Using this forecasting method can be fooled:

- 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
- Theorem: Using polynomials in $\hat{y}$ will lead to a "weakly calibrated" forecast.

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."

Definition of equilibrium?

A correlated equilibrium satisfies:

- Player 1 is conditionally rational:

$$E(U_1(A_1, A_2)|\mathcal{F}_1) = \max_a E(U_1(a, A_2)|\mathcal{F}_1)$$

- Player 2 is conditionally rational:

$$E(U_2(A_1, A_2)|\mathcal{F}_2) = \max_a E(U_1(A_1, a)|\mathcal{F}_2)$$

Definition of equilibrium?

A correlated equilibrium satisfies:

- Player 1 is conditionally rational:

$$E(U_1(A_1, A_2)|\mathcal{F}_1) = \max_a E(U_1(a, A_2)|\mathcal{F}_1)$$

- Player 2 is conditionally rational:

$$E(U_2(A_1, A_2)|\mathcal{F}_2) = \max_a E(U_1(A_1, a)|\mathcal{F}_2)$$

If $\mathcal{F}_1 \perp \mathcal{F}_2$ then this is a Nash equilibrium

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

- forecast by average
  - zero sum converges
  - Shapely game cycles converges
  - All proven in the 1950's
- Calibrated forecasts
  - any game converges to correlated equilibrium
- Weakly calibrated forecasts
  - If players use a continuous ϵ-best reply, then these will converge to a correlated equilibrium
  - Can be tweaked to make it converge to a Nash equilibrium

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
- It will end up having no-regrets in the long run
- It is cleaner than using calibration

Random conclusions

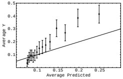
- Use calibration to clean up regressions / time series
  - Isotonic regressions (Sham and others)
  - Time series (Sham and me, Sergiu and me)
  - Applied regression (Bob Stine and me)
- Fixed point are deeply connected to calibration (ask Sham)
- Learning in Game theory
  - no-internal regret is computational easy for fairly large games and converges to CE
  - In fact, you only need to have no-internal regret against similar strategic choices (Sasha and me)
  - Calibration is over-kill for CE but can be used for NE

Random conclusions

- Use calibration to clean up regressions / time series
  - Isotonic regressions (Sham and others)
  - Time series (Sham and me, Sergiu and me)
  - Applied regression (Bob Stine and me)
- Fixed point are deeply connected to calibration (ask Sham)
- Learning in Game theory
  - no-internal regret is computational easy for fairly large games and converges to CE
  - In fact, you only need to have no-internal regret against similar strategic choices (Sasha and me)
  - Calibration is over-kill for CE but can be used for NE

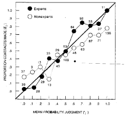
Thanks!

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