Comparative Testing of Experts
The Truth is Relative

Nabil Al-Najjar & Jonathan Weinstein

Northwestern University
MEDS

October 18, 2007

▸ Skip slide

"O False and treacherous Probability,

Enemy of truth, and friend of wickednesse;

With whose bleare eyes Opinion learnes to see,

Truth's feeble party here, and barrennesse."

–Keynes

The problem

Skip slide

- Unknown probability distribution P generating observations
  - e.g. weather conditions, stock prices, or GDP levels,
- The true distribution P is unrestricted (need not be independent or identical across periods)
- In each period, an expert makes a probabilistic forecast that he claims is based on his knowledge of P
- *Can this claim be tested?*

Literature

Skip slide

Calibration Literature: **Foster & Vohra, 1998**

Fudenberg & Levine, 1999; Sandroni, Smorodinsky & Vohra, 2003; Kalai, Lehrer, & Smorodinsky, 1999; Lehrer, 2001.

General Tests: **Sandroni, 2003**

Dekel & Feinberg, 2006; Olszewski & Sandroni, 2006 and 2007.

Multiple Experts:

Feinberg & Stewart, 2006.

The problem formally: *Strategies*

Skip slide

- A is a fixed finite set of outcomes
- $\Delta(\cdot)$ is probability measures on a space
- Each period $t = 1, \ldots, n$ an expert submits a forecast

$$\alpha(t) \in \Delta(A)$$

- A *forecasting strategy* is $f = \{f^t\}_{t=1}^n$ where

$$f^t : H^{t-1} \rightarrow \Delta(A)$$

- F^n is the set of all forecasting strategies
- $\Delta(F^n)$ is the set of expert's mixed strategies

The problem formally: *Tests*

Skip slide

A single expert test is any function

$$T^n : A^n \times \Delta(A^n) \to \{0, 1\}$$

A strategic expert's payoff is the expected probability of passing the test:

$$z(P, \varphi) = \int_{A^n} \int_{\Delta(A^n)} T^n(a^n, Q)\, d\varphi(Q)\, dP(a^n).$$

The test T^n *passes the truth with probability* $1 - \epsilon$ if:

$$z(P, P) \equiv P\{T^n(a^n, P) = 1\} > 1 - \epsilon.$$

Impossibility theorem

Skip slide

Theorem (Sandroni, 03, IJGT):

For any test T that passes the truth with probability $1 - \epsilon$ there is an expert's mixed strategy φ that passes the test with probability $1 - \epsilon$ regardless of the true distribution.

- ... in particular, on any sequence of outcomes
- Any test that is is "generous enough" is "too generous"

Is there at least one informed expert?

Skip slide

Is there a way to determine if among two experts at least one is informed?

Consider the function:

$$\tau : H^n \to \{0, 1\}$$

Interpret $\tau(a^n, f_0, f_1) = 1$ to mean that at least one expert is informed

Impossibility theorem

Skip slide

Theorem

Suppose that τ *is such that for every* P, f_0 *and* f_1

$$P\{a^n : \tau(a^n, f_0, f_1) = 1\} > 1 - \epsilon \quad \text{if either } f_0 = f_P \text{ or } f_1 = f_P.$$

Then for every mixed strategy φ_0 *of Expert 0 there is a mixed strategy* φ_1 *of Expert 1 such that for every* a^n

$$\varphi_0 \times \varphi_1\{(f_0, f_1) : \tau(a^n, f_0, f_1) = 1\} > 1 - \epsilon.$$

Multiple experts: *Definitions*

Skip slide

Now assume that there are two experts: 0 and 1.

An *n-period comparative test* is any (measurable) function

$$T^n : A^n \times F^n \times F^n \to \{0,\ 0.5,\ 1\}$$

such that for every $f, f' \in F^n$ and a^n,

$$T^n(a^n, f, f') = 1 - T^n(a^n, f', f).$$

$i = T^n(h^n)$ means that the test picks expert i after observing the history of forecasts and Nature's realizations

Comparative test

Skip slide

Set $L_0(h^0) = 1$

For $t > 0$

$$L_t(h^t) = \frac{f_1^t(h^{t-1})(a(t))}{f_0^t(h^{t-1})(a(t))} L_{t-1}(h^{t-1})$$

Expert 1 is chosen if $L_n(h^n) > 1$

Expert 0 is chosen if $L_n(h^n) < 1$

The test returns “0.5” if $L_n(h^n) = 1$

Main possibility theorem

Skip slide

Theorem

For every $\epsilon > 0$, there is an integer K such that for all integers n, distributions P, and mixed forecasting strategies φ_0, φ_1 with at least one informed expert, there is P-probability at least $1 - \epsilon$ that either

(a) *T^n picks an informed expert; or*

(b) *The two experts' forecasts are ϵ-close in all but K periods.*

K depends on ϵ but not n, P or the strategies

Intuition for the test

Skip slide

- Identical to: assigning 50-50 prior to each expert announcing true probabilities, Bayesian updating
- So, mathematically equivalent to updating concerning a player's type in a reputation game
- Realized states are like signals generated by actions of a player with two types, the probabilities announced are like the signal distribution conditional on each type

Sketch of the proof 1

Skip slide

Assume Expert 0 is informed and that he reports the truth

The stochastic process $\{L_t\}$ is a supermartingale under the distribution induced by the strategy of Expert 0

As in Fudenberg & Levine (1992), define $\{\tilde{L}_t\}$ to be the faster process obtained from $\{L_t\}$ through a sequence of stopping times that contains all finite histories at which

$$\left| f_0^t(h^{t-1})(a(t)) - f_1^t(h^{t-1})(a(t)) \right| > \epsilon$$

Sketch of the proof 2

Skip slide

Fudenberg and Levine show that $\{\tilde{L}_t\}$ is an active supermartingale with activity ϵ.

Their Theorem A.1 implies that for any $\epsilon > 0$ there is an integer K such that for any active supermartingale $\{\tilde{L}_t\}$

$$P\left[\sup_{k>K} \tilde{L}_k < 1\right] > 1 - \epsilon.$$

K depends only on ϵ and not on the true stochastic process P or the forecasting strategy f_1.

How big a loophole is left?

Skip slide

The test potentially gives the false expert a loophole–if he can match the predictions of the true expert.

To win the reputation game the imposter must win the "matching game"–how likely is this?

Scope of strategic manipulations

Skip slide

Family of incomplete-information constant-sum games between Expert 0 and Expert 1, parametrized by $n = 1, 2 \ldots$ and $\mu \in \Delta(\Delta(A^n))$:

- Nature chooses $P \in \Delta(A^n)$ according to μ;
- Expert 0 is informed of P; Expert 1 only knows μ;
- The two players simultaneously choose forecasts;
- Nature then chooses a^n according to P;
- The payoff of Expert 1 is

  $$T^n\left(a^n, f_0, f_1\right),$$

  where T^n is the test constructed in Theorem 2.
- The payoff of Expert 0 is $1 - T^n\left(a^n, f_0, f_1\right)$.

REPUTATION GAME SLIDES USED HERE;

THEY APPEAR AT THE END OF THIS DOCUMENT

Preliminary

Skip slide

Identify $\mu \in \Delta(\Delta(A^n))$ with its one-step-ahead conditionals,

$$\mu^t(\cdot|\alpha^{t-1}) \in \Delta(\Delta(A))$$

$\mathcal{M}(\epsilon, \delta, L) \subset \Delta(\Delta(A^n))$ consists of all μ such that there are at least L periods $1 \le t \le n$ such that for μ-a.e. h^n

$$\max_{p \in \Delta(A)} \mu^t(B_\epsilon(p)|\alpha^{t-1}) < 1 - \delta.$$

- In words, in each of at least L periods μ does not concentrate its mass in some small ball
- Condition becomes less restrictive as n becomes large.

Possibility theorem: *Bayesian game*

Skip slide

Theorem

For every ϵ and $\delta > 0$ there is an integer L such that for every $\mu \in \mathcal{M}(\epsilon, \delta, L)$ the value of the game to Expert 1 is less than ϵ.

If μ is sufficiently diffuse, player 0 is much better informed than player 1, and the value of player 1 is small.

Comparative test: *Infinite horizon*

Skip slide

L^t is defined exactly as before

Test picks Expert 0 if $\lim_{n\to\infty} L_n(h^n) < 1$

Test picks Expert 1 if $\lim_{n\to\infty} L_n(h^n) > 1$

Test picks at random if either $\lim_{n\to\infty} L_n(h^n) = 1$ or this sequence fails to converge.

Possibility theorem: *Infinite horizon*

▸ Skip slide

Theorem

For any distribution P and mixed forecasting strategies φ_0, φ_1 *with at least one informed expert, with P-probability 1 either*

(a) *T picks an informed expert; or*

(b) $\lim_{t\to\infty} |f_0^t(h^{t-1}) - f_1^t(h^{t-1})| = 0.$

Possibility theorem: *Infinite horizon*

Skip slide

$\mathcal{M}(\epsilon, \delta) \subset \Delta(\Delta(A^\infty))$ consists of all μ such that for μ-a.e. infinite history h^∞, for infinitely many periods,

$$\max_{p \in \Delta(A)} \mu^t(B_\epsilon(p) | \alpha^{t-1}) < 1 - \delta.$$

Theorem

For every $\epsilon,\ \delta > 0$ and $\mu \in \mathcal{M}(\epsilon, \delta)$ the value of the game to Expert 1 is zero.

What drives the single-expert result?

Skip slide

Proof of impossibility in the single-expert case:

The minimax theorem

$$\max_{\varphi\in\Delta(\Delta(A^n))} \min_{P\in\Delta(A^n)} z(P,\varphi) = \min_{P\in\Delta(A^n)} \max_{\varphi\in\Delta(\Delta(A^n))} z(P,\varphi).$$

But..

$$\min_{P\in\Delta(A^n)} \max_{\varphi\in\Delta(\Delta(A^n))} z(P,\varphi) \geq \min_{P\in\Delta(A^n)} z(P,P) > 1-\epsilon.$$

Therefore

$$\max_{\varphi\in\Delta(\Delta(A^n))} \min_{P\in\Delta(A^n)} z(P,\varphi) \geq 1-\epsilon.$$

This looks like magic!

Skip slide

This a “hide-and-seek” game with a continuum of locations.
How can the seeker (the expert) beat the hider (Nature)?

Resolution: Nature has no meaningful randomizations

Minimax Theorem really says:

$$\max_{\varphi \in \Delta(\Delta(A^n))} \min_{\mu \in \Delta(\Delta(A^n))} z(\mu, \varphi) = \min_{\mu \in \Delta(\Delta(A^n))} \max_{\varphi \in \Delta(\Delta(A^n))} z(\mu, \varphi).$$

But any of Nature’s mixed strategy μ is payoff-equivalent to pure strategy P_μ obtained by reduction of compound lotteries:

$$z(\mu, \varphi) = z(P_\mu, \varphi), \quad \forall \mu, \varphi \in \Delta(\Delta(A^n)).$$

True probabilities and multiple experts

Skip slide

- In the single-expert case, Nature's compound lotteries are equivalent to their reduction
- With multiple experts, the presence of the informed expert's forecasts breaks the observational equivalence between μ and P_μ
- ... now the game is much harder.

Let's Play "The Reputation Game"

- I take 5 cards at random from a deck with 27 red and 27 green cards
- I draw three cards from the 5, with replacement (so A={red, green},n=3)
- One player knows only the information on this slide, μ_1, the other knows the composition of the 5-card subset and thus P

Let's Play Again

- This time I draw (with replacement) a new 5-card subset each period, then draw one card from each
- Again, one player only knows the process on this slide, μ_2, the other knows the composition of each subset and thus P

One more game

- Same process as last time
- One player knows the composition of the three 5-card subsets, μ, the other knows the actual draw, P (a point-mass distribution)
- The player who "knew the truth" last time knows exactly the same amount this time, but suddenly that player is "uninformed" of the "true" distribution