Calibration and Approachability

Luciano Pomatto

Caltech

Forecasting

- Every day a forecaster announces a probability p of rain.
- We want to understand if and how these predictions can be refuted empirically.

Probabilistic predictions:

- weather and climate (Gneiting and Raftery, 2005)
- aggregate output and inflation (Bank of England, Diebold, Tay and Wallis, 1997), epidemics (Alkema, Raftery and Clark, 2007)
- seismic hazard (Jordan et al., 2011)
- financial risk (Timmermann, 2000)
- demographic variables (Raftery et al., 2012)
- elections (Tetlock, 2005), etc.

Forecasting

A sample of surveys on forecasting:

- Murphy, Winkler,'84: Probability Forecasting in Metereology, *JASA*.
- Diebold et al., '97: Evaluating Density Forecasts, *IER*.
- Gneiting, Katzfuss, '14: Probability Forecasting, *Ann.Rev.Econ*.

A recurrent question: what is a good criterion for judging the quality of probabilistic forecasts?

A common theme: good forecasts should display "*statistical compatibility between the probabilistic forecasts and the realizations; essentially, the observations should be indistinguishable from random draws from the predictive distributions. [Murphy and Winkler]*"

Calibration Test

Journal of the American Statistical Association, September 1984

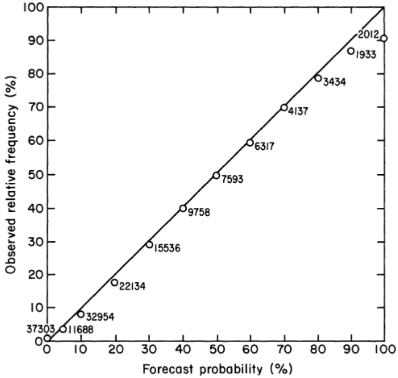

Figure 2. Reliability Diagram for Subjective Precipitation Probability Forecasts Formulated at 87 National Weather Service Offices During the Period April 1977–March 1979 (Charba and Klein 1980). The number adjacent to each point indicates the number of forecasts with that forecast probability.

Murphy and Winkler, *JASA*, 1984

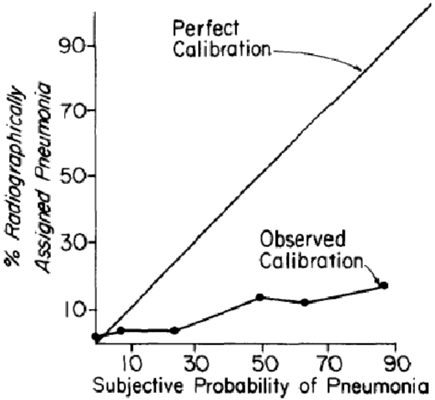

Figure 1. Relationship between physicians' subjective probability of pneumonia and the actual probability of pneumonia.

Christensen et al. , Jour. Exper. Psych, 1981

Many variations across fields: density forecasts, Value-at-Risk, etc.

Formal Definitions

- In each period an outcome, 0 or 1, is publicly observed.
- $\Omega = \{0,1\}^\infty$: set of all **paths**.
- $\Delta(\Omega)$: set of all Borel probability measures on Ω.
- ω_t : realization at time t along path ω.
- $(f_1, f_2, \ldots)$: **forecasting rule**. Each f_t takes values in $[0,1]$ and is measurable w.r.t. the information available at time $t-1$.

Calibration

- Partition $[0,1]$ into $[0,1/m],\ldots,[(m-1)/m,1]$. Let $M_i = \frac{2i-i}{2m}$.

- Given ω consider the frequency

$$\rho^T_{M_i}(\omega) = \frac{\sum_{t=1}^{T} \omega_t 1\{f_t = M_i\}}{\sum_{t=1}^{T} 1\{f_t = M_i\}}$$

(let $\rho^T_{M_i} = M_i$ if the denominator is 0)

- $\rho^T_{M_i}$: frequency of 1's conditional on the forecast being M_i.

Calibration

- Partition $[0,1]$ into $[0,1/m],\ldots,[(m-1)/m,1]$. Let $M_i = \frac{2i-i}{2m}$.

- Given ω consider the frequency

$$\rho^T_{M_i}(\omega) = \frac{\sum_{t=1}^T \omega_t 1\{f_t = M_i\}}{\sum_{t=1}^T 1\{f_t = M_i\}}$$

(let $\rho^T_{M_i} = M_i$ if the denominator is 0)

- $\rho^T_{M_i}$: frequency of 1's conditional on the forecast being M_i.

Calibration

Definition

A forecasting rule $(f_1, f_2, \ldots)$ is $1/m$-calibrated along path ω if

$$\limsup_T |\rho^T_{M_i}(\omega) - M_i| \leq \frac{1}{2m}$$

for every i such that $\limsup_T \sum_{t=1}^T 1\{f_t = M_i\}(\omega) > 0$.

- Along a given path ω many different forecasting rules can be calibrated. The test can be made harder by coupling it with a collection of **checking rules** (e.g. Sandroni, Smorodinsky, Vohra, 1999).

Fundamental property

- Let $P \in \Delta(\Omega)$ be the true law governing the data.
- P is unrestricted and unknown.

Dawid (1982)
A forecaster who predicts according to P passes the calibration test P-a.s.

Hence:

- **Type-I error free**: No risk of rejecting the correct predictions of an expert who knows the true law.
- The tester is not required to have any preconceived theory about the problem at hand. The forecaster can be evaluated on purely empirical ground.

Tests and incentive problems

Two main approaches:

1. Contract theory: forecasters as agents advising a principal about the best course of action.

2. Statistical tests: alternative to standard contracts. Used when:

    - Forecasts lack an easily identifiable user.
      (e.g. National Weather Service, Macroeconomics)

    - Contracts are impractical.

    - The decision problem is not well defined.
      (e.g. testing of scientific theories)

Key issue:
Forecasters may be concerned about their reputation.

Adverse selection

Consider:

- An expert informed about the true probabilistic law governing the data.
- A forecaster who is ignorant about the data generating process but is interested in passing the test.

The calibration test cannot discriminate between the two.

Adverse selection

Theorem (Foster and Vohra, 1998)

For every m, there exists a probability measure ζ over forecasting rules such that *for every* ω, the realized sequence of forecasts

$$(f_1, f_2, \ldots)$$

generated by ζ is almost surely $1/m$-calibrated along the path ω.

Adverse selection

Theorem (Foster and Vohra, 1998)
For every m, there exists a probability measure ζ over forecasting rules such that *for every* ω, the realized sequence of forecasts

$$(f_1, f_2, \ldots)$$

generated by ζ is almost surely $1/m$-calibrated along the path ω.

- The calibration test can be passed *without* any knowledge of the actual data generating process.

Algorithm (Binary Case)

Two measures of miscalibration, *deficit* and *excess* :

$$d_T = \left(\frac{1}{2} - \rho^T_{3/4}\right) \frac{1}{T} \sum_{t=1}^{T} 1\{f_t = 3/4\}$$

$$e_T = \left(\rho^T_{1/4} - 1/2\right) \frac{1}{T} \sum_{t=1}^{T} 1\{f_t = 1/4\}$$

Calibration requires $d_T \leq 0$ and $e_T \leq 0$ as $T \to \infty$

Algorithm (Binary Case)

1. If $d_T \leq 0$ let $f_{T+1} = 3/4$;
2. If $e_T \leq 0$ let $f_{T+1} = 1/4$; otherwise
3. Choose $f_{T+1} = 1/4$ with prob.

$$\frac{d_T}{d_T + e_T}$$

and $f_{T+1} = 3/4$ with prob.

$$\frac{e_T}{d_T + e_T}$$

i.e. randomize with weights that are inversely proportional to the degrees of miscalibration.

Algorithm (Binary Case)

1. If $d_T \leq 0$ let $f_{T+1} = 3/4$;
2. If $e_T \leq 0$ let $f_{T+1} = 1/4$; otherwise
3. Choose $f_{T+1} = 1/4$ with prob.

$$\frac{d_T}{d_T + e_T}$$

and $f_{T+1} = 3/4$ with prob.

$$\frac{e_T}{d_T + e_T}$$

i.e. randomize with weights that are inversely proportional to the degrees of miscalibration. Immediate to extend to the non-binary case.

Approachability

- While the algorithm is simple looking, a proof is highly nontrivial.
- I will cover a proof due to Foster and Hart-Mas-Colell.
- It relies on the theory of **repeated zero-sum vector games**, initiated by Blackwell (1956).

Approachability

- While the algorithm is simple looking, a proof is highly nontrivial.
- I will cover a proof due to Foster and Hart-Mas-Colell.
- It relies on the theory of **repeated zero-sum vector games**, initiated by Blackwell (1956).
- A (finite) **zero-sum vector game** is a game $\langle u, A, B \rangle$ where

$$u : A \times B \to \mathbb{R}^m$$

- We are going to consider the corresponding repeated game.

Blackwell (1956)

Definition

A set $C \subseteq \mathbb{R}^m$ is **approachable** if there exists a strategy $(\sigma^t)_{t=1}^{\infty}$ of Player 1 such that for any strategy of Player 2

$$dist\left(\frac{1}{T}\sum_{t=1}^{T} u\left(a^t, b^t\right), C\right) \to 0 \text{ a.s.}$$

Definition

An halfspace $\{x \in \mathbb{R}^m : \langle x, \lambda \rangle \geq \beta\}$ is **enforceable** if there exists a randomization $\pi \in \Delta(A)$ such that

$$\left\langle \sum_{a \in A} \pi(a)\, u(a, b), \lambda \right\rangle \geq \beta \quad \text{for all } b \in B.$$

Blackwell (1956)

Theorem (Blackwell, 1956)
A closed convex set C is approachable if and only if every halfspace containing C is enforceable.

- Original motivation: an analog of von Neumann's minmax theorem.
- Applications across several fields: repeated games (Aumann and Maschler, 1995), learning in games (Hart and Mas-Colell, 2000), contract theory (Chassang, 2013), algorithmic game theory, and machine learning (Cesa-Bianchi and Lugosi, 03).
- Key to the applicability of the result is its constructive proof.

Main Ideas:

- Look at the accumulated average payoff vector

$$w_T = \frac{1}{T} \sum_{t=1}^{T} u\left(a^t, b^t\right)$$

- Let $c \in C$ be the point of minimum distance from w_T.
- Then

$$C \subseteq \left\{x \in \mathbb{R}^m : \langle x - c, w_T - c \rangle \leq 0\right\}.$$

Main Ideas:

- **Blackwell Strategy**: play according to a randomization π such that

$$\left\langle \underbrace{\sum \pi\left(a^{T+1}\right) u\left(a^{T+1}, b\right)}_{\text{expected vector payoff}} - c, w_T - c \right\rangle \leq 0 \quad \text{for all } b.$$

- A geometric argument suggests that $E[\delta_{T+1}]$, the expected distance from C, shrinks over time.

Main Ideas:

Let $\delta_{T+1} = \min_{x \in C} \left\| \frac{1}{T+1} \sum_{t=1}^{T+1} u\left(a^t, b^t\right) - x \right\|^2$

$$\delta_{T+1} \leq \left\| \frac{T}{T+1}\left(w_T - c\right) + \frac{1}{T+1}\left(u\left(a^{T+1}, b^{T+1}\right) - c\right) \right\|^2$$

Main Ideas:

Let $\delta_{T+1} = \min_{x \in C} \left\| \frac{1}{T+1} \sum_{t=1}^{T+1} u\left(a^t, b^t\right) - x \right\|^2$

$$\delta_{T+1} \leq \left\| \frac{T}{T+1}\left(w_T - c\right) + \frac{1}{T+1}\left(u\left(a^{T+1}, b^{T+1}\right) - c\right) \right\|^2$$

$$\delta_{T+1} \leq \frac{T^2}{(1+T)^2}\delta_T + \frac{2T}{(1+T)^2}\left\langle u\left(a^{T+1}, b^{T+1}\right) - c, w_T - c \right\rangle + \frac{M}{(1+T)^2}$$

Main Ideas:

Let $\delta_{T+1} = \min_{x \in C} \left\| \frac{1}{T+1} \sum_{t=1}^{T+1} u\left(a^t, b^t\right) - x \right\|^2$

$$\delta_{T+1} \leq \left\| \frac{T}{T+1}\left(w_T - c\right) + \frac{1}{T+1}\left(u\left(a^{T+1}, b^{T+1}\right) - c\right) \right\|^2$$

$$\delta_{T+1} \leq \frac{T^2}{(1+T)^2}\delta_T + \frac{2T}{(1+T)^2}\left\langle u\left(a^{T+1}, b^{T+1}\right) - c, w_T - c\right\rangle + \frac{M}{(1+T)^2}$$

Taking expectations:

$$E[\delta_{T+1}] \leq \frac{T^2}{(1+T)^2} E[\delta_T] + \frac{M}{(1+T)^2}$$

$\implies E[\delta_T] \leq \frac{M}{T}$. Hence $E[\delta_{T+1}] \to 0$ as $T \to \infty$.

Back to Calibration:

Two measures of miscalibration:

$$d_T = \left(\frac{1}{2} - \rho^T_{3/4}\right) \frac{1}{T} \sum_{t=1}^{T} 1\{f_t = 3/4\}$$

$$e_T = \left(\rho^T_{1/4} - \frac{1}{2}\right) \frac{1}{T} \sum_{t=1}^{T} 1\{f_t = 1/4\}$$

We want to find a strategy such that (e_T, d_T) converges to the negative orthant as $T \to \infty$, no matter how data unfolds.

Back to Calibration:

Two measures of miscalibration:

$$d_T = \left(\frac{1}{2} - \rho^T_{3/4}\right) \frac{1}{T} \sum_{t=1}^{T} 1\{f_t = 3/4\}$$

$$e_T = \left(\rho^T_{1/4} - \frac{1}{2}\right) \frac{1}{T} \sum_{t=1}^{T} 1\{f_t = 1/4\}$$

We want to find a strategy such that (e_T, d_T) converges to the negative orthant as $T \to \infty$, no matter how data unfolds.

Plan of action: establish that (e_T, d_T) is the average accumulated vector payoff in an appropriate repeated zero-sum game against Nature. Then show that the negative orthant is approachable.

Blackwell's strategy will lead to Foster and Vohra's algorithm.

Proof of the Calibration Theorem

Consider a repeated zero-sum game where in each period t,

- Nature chooses a realization $\omega_t \in \{0, 1\}$
- The forecaster chooses a prediction $f_t \in \{1/4, 3/4\}$

Payoffs:

$$u(\omega_t, 1/4) = \left(\omega_t - \frac{1}{2}, 0\right)$$

$$u(\omega_t, 3/4) = \left(0, \frac{1}{2} - \omega_t\right)$$

Proof of the Calibration Theorem

Consider a repeated zero-sum game where in each period t,

- Nature chooses a realization $\omega_t \in \{0, 1\}$
- The forecaster chooses a prediction $f_t \in \{1/4, 3/4\}$

Payoffs:

$$u(\omega_t, 1/4) = \left(\omega_t - \frac{1}{2}, 0\right)$$

$$u(\omega_t, 3/4) = \left(0, \frac{1}{2} - \omega_t\right)$$

Accumulated average payoff = vector of miscalibration scores:

$$w_T = \frac{1}{T} \sum_{t=1}^{T} u(\omega_t, f_t) = (e_T, d_T)$$

Proof of the Calibration Theorem

We must find probabilities $\pi(1/4)$ and $\pi(3/4)$ such that

$$\left\langle \pi(1/4)\left(\omega_t - \frac{1}{2}, 0\right) + \pi(3/4)\left(0, \frac{1}{2} - \omega_t\right), \left(e_T^+, d_T^+\right) \right\rangle \leq 0 \text{ for all } \omega_t$$

A solution is obtained by the following randomization:

- If $e_T^+ = 0$ set $\pi(1/4) = 1$
- If $d_T^+ = 0$ set $\pi(3/4) = 1$
- Otherwise set

$$\frac{\pi(1/4)}{\pi(3/4)} = \frac{d_T}{e_T}$$

Testing Strategic Forecasters

Luciano Pomatto

Caltech

Adverse selection

Foster and Vohra (1998)
There exists a randomized forecasting algorithm that requires no knowledge about the data generating process and makes the forecaster calibrated with high probability, no matter what data is realized.

Fundamental question: What tests cannot be manipulated?

Model: Basic Ingredients

- In each period an outcome from a finite set $\mathcal{X}$ is publicly observed.
- $\Omega = \mathcal{X}^\infty$: set of all **paths**.
- $\Delta(\Omega)$: set of all Borel probability measures on Ω.

Model: Empirical Tests

- A forecaster claims to know the law $P \in \Delta(\Omega)$ governing the data.
- A tester is interested in evaluating this claim.

Timing:

1. The tester designs a test

$$T : \Omega \times \Delta(\Omega) \to \{\mathrm{pass}, \mathrm{reject}\}$$

2. The forecaster observes T and reports a prediction P.

3. Nature produces a path $\omega \in \Omega$.

4. $T(\omega, P)$ determines acceptance or rejection.

Example · Likelihood-Ratio Test

1. The forecaster announces P.

2. The test produces a competing theory P^* such that

$$P(E) = 1 \text{ and } P^*(E) = 0 \text{ for some } E \subseteq \Omega$$

3. $T(\omega, P) = \text{pass}$ if and only if

$$\lim_{n\to\infty} \frac{P(\omega^n)}{P^*(\omega^n)} = \infty$$

By a martingale convergence argument, $P\{\omega : T(\omega, P) = \text{pass}\} = 1$ for every P.

Adverse Selection

The forecaster is either:

- A **true expert** who knows the true law P governing the data, and reports it truthfully.
- A **strategic forecaster** uninformed but interested in passing the test.

A (mixed) strategy is a randomization $\zeta \in \Delta(\Delta(\Omega))$ with finite support.

Examples:

1. Fix a benchmark measure P^* with full support and a time n.
2. The forecaster announces $P \in \Delta(\Omega)$.
3. $T(\omega, P) = \text{pass}$ if and only if

$$\frac{P(\omega^n)}{P^*(\omega^n)} > 1$$

Examples:

1. Fix a benchmark measure P^* with full support and a time n.
2. The forecaster announces $P \in \Delta(\Omega)$.
3. $T(\omega, P) = \text{pass}$ if and only if

$$\frac{P(\omega^n)}{P^*(\omega^n)} > 1$$

There exists a strategy $\zeta \in \Delta(\Delta(\Omega))$ such that

$$\zeta\{P : T(\omega, P) = \text{pass}\} \geq 1 - \frac{1}{2^n}$$

for every $\omega \in \Omega$.

Desiderata

- I : The test does not reject a true expert.

Definition

A test T passes the truth with probability $1 - \varepsilon$ if for all $P \in \Delta(\Omega)$

$$P\{\omega : T(\omega, P) = \text{pass}\} \geq 1 - \varepsilon$$

Desiderata

- II : The test decides in finite time.

Definition

A test T is finite if for every $P \in \Delta(\Omega)$ there exists a time N_P such that $T(\cdot, P)$ is measurable with respect to $\mathcal{F}_{N_P}$.

Sandroni(2003)

Theorem

Consider a test T that is finite and passes the truth with probability $1-\epsilon$. Then, for every $\delta > 0$ there exists a strategy ζ such that

$$\zeta\{\omega : T(\omega, P) = \mathrm{pass}\} \geq 1 - \epsilon - \delta$$

for every path ω.

- The test can be ignorantly passed without any knowledge about the data generating process.

Proof

Theorem (Fan)

Let X and Y be convex subsets of two topological vector spaces and consider a function

$$V : X \times Y \to [0,1]$$

If

- X is compact
- $V(\cdot, y)$ is convex and lsc
- $V(x, \cdot)$ is concave

then

$$\min_x \sup_y V = \sup_y \min_x V$$

- See Fan (1953) or Mertens (1986), *"The minmax theorem for USC-LSC payoff functions," IJGT.*

Generalizations

- Shmaya (2008): prequential tests.
- Olszewski and Sandroni (2008): future-independent tests.

Nonmanipulable Tests

- Dekel and Feinberg (2006) and Olszewski and Sandroni (2009) overturned the negative results in the literature by constructing nonmanipulable tests.
- **Their approach:** to allow for tests that may not return, for some realization, a pass/fail decision in finite time.
- But first, a detour.

Verification and Falsification

Let

- 1: a "white swan"
- 0: a "black swan"

consider the claim **"all swans are white"**

$$\omega = (1, 1, 1, \dots).$$

- No **finite** evidence can verify this claim.
- A single black swan is enough to falsify it.

Verification and Falsification

Let

- 1: a “white swan”
- 0: a “black swan”

consider the claim **“all swans are white”**

$$\omega = (1, 1, 1, \dots).$$

- No **finite** evidence can verify this claim.
- A single black swan is enough to falsify it.

Popper’s essential claim: falsification is a better maxim than verification for empirical research.

Verification and Falsification

More generally, consider a set $A \subseteq \Omega$. Then

- A is open $\iff \forall \omega \in A$, there exists n s.t. $\omega^n \subseteq A$

i.e. if A is true then it can be verified using finite evidence.

- A is closed $\iff \forall \omega \notin A$, there exists n s.t. $\omega^n \subseteq A^c$

i.e. if A is false then it can be falsified using finite evidence.

Verification and Falsification

Definition

T is a verification test if

$$\text{for all } P \text{ the set } \{\omega : T(\omega, P) = \mathrm{pass}\} \text{ is open} \tag{1}$$

T is a falsification test if

$$\text{for all } P \text{ the set } \{\omega : T(\omega, P) = \mathrm{pass}\} \text{ is closed} \tag{2}$$

(1) : any theory P can be verified by the test.

(2) : any theory P can be falsified by the test.

Falsifiability and Nonmanipulable Tests

Theorem

Consider a verification test T that passes the truth with probability $1-\epsilon$. The test can be ingnorantly passed with probabililty $1-\epsilon-\delta$.

Theorem (DF-OS)

For every ϵ there exists a falsification test T that passes the truth with probability $1-\epsilon$ and is nonmanipulable: for every strategy ζ the set

$$\{\omega : \zeta(\{Q : T(\omega, Q) = \mathrm{pass}\}) = 0\}$$

is nonempty and open.

Proof

- Fix $\overline{\omega} = (1, 1, 1, \ldots)$.
- Let $R^n = \overline{\omega}^n - \{\overline{\omega}\}$
- R^n : set of paths where the first n consecutive swans are white but a black swan is eventually observed.

Proof

- Fix $\overline{\omega} = (1, 1, 1, \ldots)$.
- Let $R^n = \overline{\omega}^n - \{\overline{\omega}\}$
- R^n : set of paths where the first n consecutive swans are white but a black swan is eventually observed.
- $R^n \supseteq R^{n+1}$ and $\bigcap_n R^n = \emptyset$

Proof

- Fix $\overline{\omega} = (1, 1, 1, \ldots)$.
- Let $R^n = \overline{\omega}^n - \{\overline{\omega}\}$
- R^n : set of paths where the first n consecutive swans are white but a black swan is eventually observed.
- $R^n \supseteq R^{n+1}$ and $\bigcap_n R^n = \emptyset$
- For every $P \in \Delta(\Omega)$, $P(R^n) \downarrow 0$ as $n \to \infty$.
- For every P fix $n_P \in \mathbb{N}$ large enough s.t. $P(R^{n_P}) < \epsilon$

Proof

- Fix $\overline{\omega} = (1, 1, 1, \ldots)$.
- Let $R^n = \overline{\omega}^n - \{\overline{\omega}\}$
- R^n : set of paths where the first n consecutive swans are white but a black swan is eventually observed.
- $R^n \supseteq R^{n+1}$ and $\bigcap_n R^n = \emptyset$
- For every $P \in \Delta(\Omega)$, $P(R^n) \downarrow 0$ as $n \to \infty$.
- For every P fix $n_P \in \mathbb{N}$ large enough s.t. $P(R^{n_P}) < \epsilon$
- Define T as $T(\omega, P) = \text{reject} \iff \omega \in R^{n_P}$

Proof

- Fix $\overline{\omega} = (1, 1, 1, \dots)$.
- Let $R^n = \overline{\omega}^n - \{\overline{\omega}\}$
- R^n : set of paths where the first n consecutive swans are white but a black swan is eventually observed.
- $R^n \supseteq R^{n+1}$ and $\bigcap_n R^n = \emptyset$
- For every $P \in \Delta(\Omega)$, $P(R^n) \downarrow 0$ as $n \to \infty$.
- For every P fix $n_P \in \mathbb{N}$ large enough s.t. $P(R^{n_P}) < \epsilon$
- Define T as $T(\omega, P) = \text{reject} \iff \omega \in R^{n_P}$
- Key Property : the sets R^{n_P} are nested.

Category Tests

Ideally, we would like to establish that for every strategy ζ the set of paths

$$\{\omega : \zeta(\{Q : T(\omega, Q) = \mathrm{pass}\}) = 0\}$$

where a strategic forecaster is rejected is "large" or "typical."

Following Dekel and Feinberg (2006), the literature has taken as a desideratum tests where the above set is topologically large.

Category Tests

Theorem (DF-OS)

For every ϵ there exists a falsification test T that passes the truth with probability $1 - \epsilon$ and such that for every strategy ζ the set

$$\{\omega : \zeta(\{Q : T(\omega, Q) = \mathrm{pass}\}) = 0\}$$

is open and dense.

A Decision Theoretic Perspective

Payoffs:

- 0 outside option
- $w > 0$ if $T = \mathrm{pass}$
- $d < 0$ if $T = \mathrm{reject}$
- Uninformed forecasters choose according to Wald's maxmin criterion.

Maxmin expected payoff:

$$v = \sup_{\zeta} \inf_{P} E_{P \otimes \zeta} \left[w 1\{T = \mathrm{pass}\} + d 1\{T = \mathrm{reject}\} \right]$$

$v < 0$ whenever T is nonmanipulable.

A Decision Theoretic Perspective

Payoffs:

- 0 outside option
- $w > 0$ if $T = \text{pass}$
- $d < 0$ if $T = \text{reject}$
- Uninformed forecasters choose according to Wald's maxmin criterion.

Maxmin expected payoff:

$$v = \sup_{\zeta} \inf_{P} E_{P \otimes \zeta} \left[w1\{T = \text{pass}\} + d1\{T = \text{reject}\} \right]$$

$v < 0$ whenever T is nonmanipulable. So, only informed experts will participate to the test. Hence, T can screen between informed and uninformed forecasters.

Another Look at Nonmanipulable Tests

- Fix $\overline{\omega} = (1, 1, 1, \ldots)$.
- Let $R^n = \overline{\omega}^n - \{\overline{\omega}\}$
- For every $P \in \Delta(\Omega)$, $P(R^n) \downarrow 0$ as $n \to \infty$.

Another Look at Nonmanipulable Tests

- Fix $\overline{\omega} = (1, 1, 1, \ldots)$.
- Let $R^n = \overline{\omega}^n - \{\overline{\omega}\}$
- For every $P \in \Delta(\Omega)$, $P(R^n) \downarrow 0$ as $n \to \infty$. Why?

Another Look at Nonmanipulable Tests

- Fix $\overline{\omega} = (1, 1, 1, \ldots)$.
- Let $R^n = \overline{\omega}^n - \{\overline{\omega}\}$
- For every $P \in \Delta(\Omega)$, $P(R^n) \downarrow 0$ as $n \to \infty$. Why?
- Not a mere technicality.

The Axioms of Probability

- Consider an agent, Bob, who thinks the data will unfold according to P.
- Say $P(\overline{\omega}) > 0$. Then $P(R^n) \downarrow 0$ is equivalent to

$$P(\overline{\omega} | \underbrace{1, \ldots, 1}_{\text{n}}) \to 1 \text{ as } n \to \infty$$

- After sufficiently many white swans are observed the statement "all swans are white" becomes a virtual certainty. Bob is willing to conclude, from **finite evidence**, that a **universal law** of nature is true.
- Is Bob justified in making this inference?

The Axioms of Probability

- Consider an agent, Bob, who thinks the data will unfold according to P.
- Say $P(\overline{\omega}) > 0$. Then $P(R^n) \downarrow 0$ is equivalent to

$$P(\overline{\omega} | \underbrace{1, \ldots, 1}_{\text{n}}) \to 1 \text{ as } n \to \infty$$

- After sufficiently many white swans are observed the statement "all swans are white" becomes a virtual certainty. Bob is willing to conclude, from **finite evidence**, that a **universal law** of nature is true.
- Is Bob justified in making this inference? Hume: in the absence of a logical justification for it, induction must either be accepted on faith or must be rejected.

The Axioms of Probability

- Consider an agent, Bob, who thinks the data will unfold according to P.
- Say $P(\overline{\omega}) > 0$. Then $P(R^n) \downarrow 0$ is equivalent to

$$P(\overline{\omega} | \underbrace{1, \ldots, 1}_{\text{n}}) \to 1 \text{ as } n \to \infty$$

- An ongoing debate on "Bayesian orgulity" in philosophy of science revisits these questions in the language of modern probability.
- Intrinsic in Kolmogorov's axioms of probability (σ-additivity) is a form of faith in induction.
- In two papers with Nabil Al-Najjar and Alvaro Sandroni, we discuss these ideas in the context of testing forecasters.

Other approaches:

- Assumptions about the data-generating process.
- Testing multiple forecasters (Al-Najjar and Weinstein, 2008, Feinberg and Stewart, 2008, Olszewski and Sandroni, 2008).
- Impose a computational bound on strategies (Fortnow and Vohra, 2008, Hu and Shmaya, 2013).
- Focus on a specific decision problem (Olszewski and Peski, 2011, and Gradwohl and Salant, 2011).
- Evaluate the forecaster's performance against a default artificial predictor, by means of a scoring rule (Wilks, 2011, *Statistical Methods in the Atmospheric Sciences*, and Lambert et al. 2011).