Truth-Incentive Surveys: Choice Matching and Posterior Probabilities

Based on papers with Dražen Prelec, Benjamin Tereick, Blake Riley, Sonja Radas and Hrvoje Šikić

The problem

- An old problem in statistics (Savage 1971, Bernardo 1979): pay a single respondent so that he reveals his true belief about a probability distribution.
- An extension to a game of many respondents: to design **survey questionnaires/scoring rules** that are **truth inducing** and/or help identify individuals who have **expertise**. For example, we want responses that help
  - predicting political or other events
  - providing useful information in market research and focus groups ....
- **The issues:**
  - What questions to ask?
  - How to reward the responses?

Outline and preview of results

- Two parts:
  (i) A new truth-incentive algorithm called "Choice-matching".

  (ii) A uniqueness result: under a "locality" condition, the only algorithm that ranks respondents according to the posterior probabilities of the true state of nature is Bayesian Truth Serum, BTS, (D. Prelec., "Science" 2004).

Example

- A company wants to evaluate a new product on a sample of customers:
  they are asked to evaluate it on a scale of 1 to 5.
- Choice-matching, in a nutshell:
  Each respondent picks a number from 1 to 5 and is also asked to predict what percentage of other respondents pick each of the numbers.
  The respondents are paid according to the accuracy of their own prediction and the accuracy of those who pick the same number.

Choice-matching

- No knowledge of prior by designer.
- unverifiable truth
- Transparent and comprehensible payment rule.
- Rests on mild assumptions.
- Extends scope of truth-serums:
  the additional question need not be about the predictions of others; not all respondents need to respond to it.

Basic Framework and Notation

- There is a multiple-choice question with answer options $A = \{1, ..., M\}$.
- Each respondent $r \in \{1, ..., n\}$ submits a report (x^r, y^r): An **answer** x^r, for example $x^r = (1, 0, 0, 0, 0)$, and a **prediction** y^r, for example $y^r = (10\%, 40\%, 5\%, 30\%, 15\%)$.
- The average answer, that is, the vector of actual percentages is denoted $\bar{x}$. The answers of respondents other than r are denoted x^{-r}, the average of those is $\bar{x}^{-r}$.
- The truthful answer of r is his **x-type** $t^r \in A$ with $t = \left(t^1, ..., t^n\right)$.

Assumptions in Basic Framework

Let $\mathcal{E}^r$ denote the event that each $i \in A$ is the honest answer of at least one respondent other than r.

1. **Common Prior.** The prior distribution of x-types is common and known to all respondents.
2. **Non-Degeneracy.** For any respondent r and any x-type t: $Pr\left(\mathcal{E}^r \mid \, t\right) > 0$.
3. **Stochastic Relevance.** For any two respondents r, s:

$$E\left[\bar{t}^{-r} \mid \, \mathcal{E}^r, \, t^r\right] \neq E\left[\bar{t}^{-s} \mid \, \mathcal{E}^s, \, t^s\right] \text{ if } t^r \neq t^s$$

4. **Universal Updating.** For any two respondents r, s:

$$E\left[\bar{t}^{-r} \mid \, \mathcal{E}^r, t^r\right] = E\left[\bar{t}^{-s} \mid \, \mathcal{E}^s, t^s\right] \text{ if } t^r = t^s$$

Strictly proper scoring rules

Given an integer $M > 1$, we say that functions $f(p; j)$, $j = 1, ..., M$, form a **strictly proper scoring rule** (SPSR) if, for any probability vectors $p = (p_1, \ldots, p_M)$, $q = (q_1, \ldots, q_M)$, $q \neq p$, we have:

$$\sum_{j=1}^{M} p_j f(p; j) > \sum_{j=1}^{M} p_j f(q; j) \tag{1}$$

Define

$$S(p) = S(p, \bar{x}) := \sum_{j=1}^{M} \bar{x}_j \, f(p; j) \tag{2}$$

Choice-matching

- Notation: $\rho(y^r) \equiv S(y^r, \bar{x}^{-r})$, with S a strictly proper scoring rule: r's **prediction score**.
- $\bar{\rho}^{-r}(k)$: Average prediction score of all the respondents other than r who reported answer k.

- Choice-matching Payment Rule:
If there is **no** $k \in A$ such that $\bar{x}_k^{-r} = 0$, then, for $0 < \lambda < 1$,

$$R(x^r, y^r) = \lambda \rho(y^r) + (1 - \lambda) \bar{\rho}^{-r}(x^r)$$

and otherwise $R(x^r, y^r) = 0$.

Choice-matching – Bayesian Nash Equilibrium

Proposition.

Assume the respondents aim to maximize the expected reward. Under the baseline assumptions, truth-telling is a Bayesian Nash equilibrium.

- The first term is maximized if r declares y^r truthfully.
- If every respondent other than r tells the truth, then the truthful y^r is, if r's type is k,

$$y^r = y^{r,k} := E\left[\bar{t}^{-r} \mid \mathcal{E}^r,\, t^r\right]$$

- The corresponding difference between non-deviation and deviating from t^r to some other response x^r with $x_i^r = 1$ is:

$$Pr(\mathcal{E}^r \mid t^r) \times (1 - \lambda) E\left[\bar{S}^{-r}(t_x^r) - \bar{S}^{-r}(x^r) \mid t^r, \mathcal{E}^r\right]$$

$$= Pr(\mathcal{E}^r \mid t^r) \times (1 - \lambda) E\left[S(y^{r,k}) - S(y^{r,i}) \mid t^r, \mathcal{E}^r\right] > 0$$

Relaxing Assumptions

- Most controversial assumptions: Common Prior and Universal updating.
- Relaxation: Result holds when posteriors of individuals with same x-type are more similar to each other than those of individuals with different x-types, in the appropriate distance,

$$d\left(y^{r,k}, y^{s,k}\right) < d\left(y^{r,k}, y^{s',k'}\right)$$

  and

$$d\left(p^1,\, q\right) \leq d\left(p^2,\, q\right) \Longleftrightarrow \sum_i q_i\, f(p^1; i) \geq \sum_i q_i\, f(p^2; i)$$

- For example, d is entropy and $f(x) = \log(x)$.

Extending the general principle behind choice-matching

- The second question does not need to be a prediction.
- In the example, each respondent could choose from a list of other, existing products. Respondent r receives the product he chooses with probability λ and otherwise receives the product chosen by a respondent randomly selected among those giving the same star rating.
- Example: MCQ asks respondent to dis/agree with "the fiscal stimulus applied in 2009 accelerated the recovery of the US economy". The auxiliary question could be a prediction about GDP, or interest rates or the unemployment rate.
- Reducing the burden: respondents could be asked to predict the percentages of ratings higher and lower than 3 stars.
- Not necessary that all respondents submit the prediction response.

Other mechanisms

- Bayesian Truth Serum, Prelec (2004): infinite number of respondents; complex; it identifies experts.
- Peer prediction:
  - Miller, N., Resnick, P. and Zeckhauser, R. (2005): requires knowing the common prior.
  - Zhang and Chen (2014) modified the method without the knowledge of common prior; relatively complex.
- No prediction question, but requires distributional assumptions on prior or estimating it from a large amount of data: Radanovic and Faltings (2015); Radanovic, Faltings and Jurca (2016); Shnayder et al. (2016); Agarwal, Mandal, Parkes and Shah (2017); Liu and Chen (2017).
- Witkowski and Parkes (2012): binary choice; complex
- Baillon (2017): trading, not responses: binary choice
- Radanovic and Faltings (2013): assumption $y_k^{r,k} > y_k^{r,\ell}$.
- Radanovic and Faltings (2014): score for type-declaration depends on prediction – all must submit prediction.

Budget balancing: sum of rewards = zero

- To prevent collusion

$$R^0(x^r, y^r) = R(x^r, y^r) - \frac{1}{n}\sum_{s=1}^{n} R(x^s, y^s)$$

- Alternatively, each respondent receives the score from each of the $n-1$ subsamples of which he is part, minus the total score of the subsample of which he is not part:

$$R^1(x^r, y^r, x^{-r}, y^{-r})$$

$$= \frac{1}{n-1}\sum_{s \neq r}\left[R(x^r, y^r, x^{-r,s}, y^{-r,s}) - R(x^s, y^s, x^{-r,s}, y^{-r,s})\right]$$

where $x^{-r,s}$ excludes the answers r and s.

Choice-matching – Summary

- Easy to explain: plenty of experimental experience in explaining scoring rules. Choice-matching adds a minimal amount of additional difficulty.
- Weak set of assumptions: Although conditioning on $\mathcal{E}^r$ has disadvantages, not important if n is moderately large.
- Can be extended to cover a large variety of settings.

Mechanisms monotone in posteriors

- The aim: to characterize such mechanisms.
- Assumption: equilibrium scores depend only on local posterior probabilities
- Result 1: such scores rank the respondents in terms of their posterior probabilities, in any type-separating equilibrium.
- Result 2: under smoothness conditions, the only such equilibrium payoffs are logarithmic, up to a linear transformation.

Recalling Bayesian Truth Serum (BTS)

$X_i^r \in \{0,1\}$: equal to one for the chosen response $i \in \{1,\ldots,M\}$.
$Y_i^r \in [0,1]$: respondent's r response on what percentage will choose i as the correct answer. State of nature:

$$\bar{X} = \lim_n \frac{1}{n}\sum_{r=1}^{n} X^r$$

Geometric mean $\bar{y}_j$:

$$\log \bar{y}_j := \lim_{n\to\infty} \frac{1}{n}\sum_{r=1}^{n} \log y_j^r$$

The Bayesian Truth Serum (BTS) score function:

$$BTS^r = \sum_{j=1}^{M} x_j^r \log \frac{\bar{x}_j}{\bar{y}_j} + \sum_{j=1}^{M} \bar{x}_j \log \frac{y_j^r}{\bar{x}_j}$$

Theorem. (Prelec 2004, CPRS 2017)
Assume the types are iid conditionally on the state of nature Ω. BTS scoring is equivalent to the budget-balanced logarithmic payoffs. More precisely, we have

$$BTS^r = \log Pr(\bar{X} = \bar{x} | X^r = x^r) - \lim_{n \to \infty} \frac{1}{n} \sum_{s=1}^{n} \log Pr(\bar{X} = \bar{x} | X^s = x^s)$$

or, in a different notation,

$$BTS^r = \log(Pr(\Omega = i \mid T^r = k))$$

$$- \sum_{j=1}^{M} Pr(T^r = j \mid \Omega = i) \log(Pr(\Omega = i \mid T^r = j))$$

The model

- Infinitely many players of $M > 1$ different types.
-$T^r \in \{1, \ldots, M\}$: player r's type. Exchangeable random variables, independent conditional on the true state of nature $\Omega \in \{1, ..., N\}$, $N > 1$.

$$Q = [q_{ki}] = [Pr(T^r = k, \Omega = i)]$$

The common prior matrix Q is known to the players, but not to the planner, who knows only M.

Introduce **type probabilities**

$$s_k = Pr(T^r = k)$$

and **posteriors**

$$Z = [z_k^i] = [Pr(\Omega = i \mid T^r = k)]$$

- a^r: response vector from a set of at least M possible responses;
- $f(a^r, a^{-r})$: scoring function (mechanism);

Locality Condition

Assumption. Fix $M > 1$.
- (i) Scoring function $f(a^r, a^{-r})$ is symmetric in the elements of a^{-r}, and for every prior matrix Q, f allows a strictly separating NE.
- (ii) **Locality condition.** In that equilibrium, the payoffs are functions $F_i : (0,1)^{2M} \to \mathbb{R}$, of the form $F_i(z_k^i, z_{-k}^i; s_k, s_{-k})$. Moreover, F does not change with permutations of z_{-k}, s_{-k}, and

$$\sum_{i=1}^{N} z_k^i F_i(z_k^i, z_{-k}^i; s_k, s_{-k}) > \sum_{i=1}^{N} z_k^i F_i(z_j^i, z_{-j}^i; s_j, s_{-j})$$

We call the family $\{F_i\}$ a **Posterior-Local Equilibrium Payoff System (PLEPS)**.

Examples

Logarithmic PLEPS:

$$F_i(z^i_k, z^i_{(-k)}) = \log(z^i_k)$$

A PLEPS **different from logarithmic**, $M = 3$ types:

$$p^i = z^i_k \ , \quad (q^i, r^i) = z^i_{-k}$$

$$F(p, q, r) = K \cdot log(p) + p^4 - 2p^3(q + r) - 6p(qr^2 + q^2 r)$$

- However, **up to first order it has to be log**:
  First, the solution to the problem

$$\min_{q^i} \left\{ \sum_i p^i [F_i(p^i, q^i, r^i) - F_i(q^i, p^i, r^i)] + \lambda \sum_i q^i \right\}$$

  is $q^i = p^i$, where λ is a Lagrange multiplier for the constraint $\sum_i q^i = 1$. The first order condition is

$$\partial_q F_i(p^i, p^i, r^i) - \partial_p F_i(p^i, p^i, r^i) = -\frac{\lambda}{p^i}$$

- Thus, for $q \approx p$,

$$F(p, q, r) - F(q, p, r) \approx [\partial_q F(p, p, r) - \partial_p F(p, p, r)](q - p)$$

$$= \lambda(1 - \frac{q}{p}) \approx \lambda(\log(p) - \log(q))$$

Score rankings

Theorem. Functions F_i comprising a PLEPS satisfy:

$$\text{If } z_k^i > z_j^i, \text{ then } F_i(z_k^i, z_{-k}^i) > F_i(z_j^i, z_{-j}^i) \tag{3}$$

Thus, if a scoring rule results in an SSNE realized via a PLEPS, the players will be ranked according to their posteriors in that SSNE.

When are the SSNE payoffs logarithmic?

We assume here $N \geq 3$.
Assumption A. For all i, and all type probabilities s_p, s_q, s_r, the second mixed derivative (assumed to exist)

$$\partial_{pq}\left[F_i(p^i, q^i, r^i; s_p, s_q, s_r) - F_i(q^i, p^i, r^i; s_q, s_p, s_r)\right]$$

of the difference in scores of two types with posteriors p^i and q^i respectively, **does not depend on other type's posteriors r^i.**

Proposition.

Consider a PLEPS system $\{F_i\}$ such that Assumption A holds. Then, if, for some $p^0 \in (0,1)$ and for any fixed type probabilities s_p, s_q, s_r the function $F_i(p^i, q^i, r^i; s_p, s_q, s_r)$ can be expanded as an infinite Taylor series around the point $(p^i, q^i, r^i) = (p^0, \ldots, p^0) \in (0,1)^M$, then, necessarily, the following *Additive Representation* (AR) holds:

$$F_i(p^i, q^i, r^i; s_p, s_q, s_r) = G_i(p^i; s_p, s_q, s_r) + H_i(p^i, q^i, r^i; s_p, s_q, s_r) \tag{4}$$

where H_i is a function that is symmetric in all the pairs $(p^i, s_p), (q^i, s_q), (r^i, s_r)$, $i = 1, \ldots, N$.

Theorem.
Consider a PLEPS consisting of functions $F_i(p^i, q^i, r^i; s_p, s_q, s_r)$, $i = 1, 2, \ldots, N$, that satisfy the assumptions of the proposition. Assume also that F_i is such that G_i is symmetric in all s_k variables, for every fixed p^i, $i = 1, \ldots, N$. Then, we have, for some functions λ and B of type probabilities $S = (s_p, s_q, s_r)$,

$$G_i(p^i, s_p, s_q, s_r) = \lambda(S) \log p^i + B_i(S)$$

In particular, if the corresponding PLEPS is budget-balanced, the EP with posterior p_i is given by

$$F_i(p^i, q^i, r^i; s_p, s_q, s_r) = \lambda(S) \log p^i - \lambda(S) \sum_{t=p,q,r} s_t^i \log t^i \qquad (5)$$

where s_t^i is the conditional probability of the type with posterior t in state i.

Proofs

Proof of the theorem.
From the FOC for truth-telling for function G we obtain

$$-\lambda \frac{1}{p} = \partial_q [G(p;S) - G(q;S)]\Big|_{q=p} = -G'(p;S)$$

This implies, for some $B = B_i(S_i)$,

$$G(p;S) = \lambda \log(p) + B$$

Ranking by posteriors.

Lemma.

Let $0 < a \leq 1$, $p, q \in (0, a)$, and $p > q$. If A, B are such that

$$pA + (a-p)B > 0 \ , \quad q(-A) + (a-q)(-B) > 0$$

then $A > 0$ and $B < 0$.

Proof. If $A = 0$ then $(a-p)B > 0$ and $(a-q)(-B) > 0$, a contradiction. Then, $B \neq 0$, and $sign(A) = -sign(B)$. Suppose $A < 0$. Then $B > 0$. From $(a-p)B > -pA$ we get $-B < \frac{p}{a-p}A$. From the second inequality we get a contradiction:

$$0 < q(-A) + (a-q)(-B) < q(-A) + \frac{a-q}{a-p}pA = Aa\frac{p-q}{a-p} < 0$$

Assume $M = 2, N = 2$. Denote

$$p := z_1^1, q := z_2^1, \text{ so that } Z = \begin{bmatrix} p & 1-p \\ q & 1-q \end{bmatrix}$$

Suppose $p > q$. The truth-inducing property implies

$$pF_1(p,q)+(1-p)F_2(1-p,1-q) > pF_1(q,p)+(1-p)F_2(1-q,1-p)$$

$$qF_1(q,p)+(1-q)F_2(1-q,1-p) > qF_1(p,q)+(1-q)F_2(1-p,1-q)$$

This leads to

$$p[F_1(p,q)-F_1(q,p)]+(1-p)[F_2(1-p,1-q)-F_2(1-q,1-p)] > 0$$

$$q[F_1(q,p)-F_1(p,q)]+(1-q)[F_2(1-q,1-p)-F_2(1-p,1-q)] > 0$$

We set $a = 1, A = F_1(p,q) - F_1(q,p)$ and $B = F_2(1-p,1-q) - F_2(1-q,1-p)$, and apply the lemma to obtain $F_1(p,q) > F_1(q,p)$ and $F_2(1-p,1-q) < F_2(1-q,1-p)$.

BTS implementing log-posterior payoffs.

Denote $p_{ij} = Pr(X_i^r = 1, X_j^s = 1)$, so that

$$Pr(X_i^r = x^r | X_j^s = x^s) = \frac{p_{ij}}{\sum_{k=1}^{M} p_{kj}}$$

- Property I: $y_j^r = \sum_{i=1}^{M} x_i^r \frac{p_{ij}}{\sum_{k=1}^{M} p_{ki}}$
- Property II: $\log Pr(X^s = x^s \mid X^r = x^r) = \sum_{j=1}^{M} x_j^s \log y_j^r$,
- Property III:

$\log Pr(X^r = x^r \mid \bar{X} = \bar{x}) = \log \sum_{k=1}^{M} x_k^r \bar{x}_k = \sum_{k=1}^{M} x_k^r \log \bar{x}_k$.

Let x^s be any values such that

$$\bar{x}_k = \lim_n \frac{1}{n} \sum_s x_k^s$$

From Property II we have

$$\sum_{k=1}^{M} \bar{x}_k \log y_k^r = \lim_n \frac{1}{n} \sum_s \log Pr(X^s = x^s | X^r = x^r)$$

and

$$\sum_{k=1}^{M} x_k^r \log \bar{y}_k = \lim_n \frac{1}{n} \sum_s \log Pr(X^r = x^r | X^s = x^s)$$

$$= \log\left(Pr(X^r = x^r|\bar{X} = \bar{x}) \lim_n \Pi_{s=1}^n \frac{Pr^{1/n}(X^s = x^s|X^r = x^r)}{Pr^{1/n}(X^r = x^r|X^s = x^s)} \right)$$

$$= \log\left(Pr(X^r = x^r|\bar{X} = \bar{x}) \frac{\lim_n \Pi_{s=1}^n Pr^{1/n}(X^s = x^s)}{Pr(X^r = x^r)} \right)$$

$$= \log Pr(\bar{X} = \bar{x}|X^r = x^r) - \log Pr(\bar{X} = \bar{x}) + \lim_n \frac{1}{n} \sum_s \log Pr(X^s = x^s)$$

$$= \log Pr(\bar{X} = \bar{x}|X^r = x^r) - \lim_{n\to\infty} \frac{1}{n} \sum_{s=1}^n \log Pr(\bar{X} = \bar{x}|X^s = x^s) \ ,$$

since the last two terms do not depend on r, and $\sum_r BTS^r = 0$.

Logarithm as the only payoff of a PBEPS.

Proof of the proposition:
Set $M = 3$. Denote

$$\bar{p} = p - p_0, \bar{q} = q - p_0, \bar{r} = r - p_0$$

By the smoothness assumption, we can write

$$F(p,q,r) = \sum_{n=0}^{\infty} a_n \bar{p}^n + \sum_{n=1}^{\infty} b_n(\bar{q}^n + \bar{r}^n)$$

$$+ \sum_{m,n=1}^{\infty} c_{m,n}\bar{p}^m(\bar{q}^n + \bar{r}^n) + \sum_{m,n=1}^{\infty} d_{m,n}\bar{q}^m\bar{r}^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n}\bar{p}^l\bar{q}^m\bar{r}^n$$

Sufficient to show:

$$c_{m,n}(s_p, s_q, s_r) = d_{m,n}(s_q, s_p, s_r)\ , \quad e_{l,m,n}(s_p, s_q, s_r) = e_{n,l,m}(s_q, s_p, s_r)$$

because then we can write

$$F(p,q,r) = \sum_{n=0}^{\infty}[a_n(s_p, s_q, s_r) - b_n^q(s_q, s_p, s_r)]\bar{p}^n + H(p,q,r)$$

We get that from

$$-\lambda = \partial_q F(p,p,r,s_p,s_q,s_r) - \partial_p F(p,p,r,s_q,s_p,s_r)$$

and

$$0 = \partial_{pqr}[F(p,q,r) - F(q,p,r)]$$

Conclusions

- (i) Any strictly separating equilibrium corresponding to PLEPS necessarily ranks the respondents according to the relative size of their posterior value;
(ii) Under additional assumptions on the sensitivity of score differences, the budget balanced strictly separating equilibria necessarily result in logarithmic payoffs.