UNIVERSITÀ DEGLI STUDI ROMA TRE

COLLANA DEL
DIPARTIMENTO DI ECONOMIA

PROBABILITY FORECASTS AND PREDICTION MARKETS

A. Philip Dawid - Julia Mortera

ISSN 2279-6916 Working papers

(Dipartimento di Economia Università degli studi Roma Tre) (online)

Working Paper n° 250, 2019

I Working Papers del Dipartimento di Economia svolgono la funzione di divulgare tempestivamente, in forma definitiva o provvisoria, i risultati di ricerche scientifiche originali. La loro pubblicazione è soggetta all'approvazione del Comitato Scientifico.

Per ciascuna pubblicazione vengono soddisfatti gli obblighi previsti dall'art. l del D.L.L. 31.8.1945, n. 660 e successive modifiche.

Copie della presente pubblicazione possono essere richieste alla Redazione.

REDAZIONE:
Dipartimento di Economia
Università degli Studi Roma Tre
Via Silvio D'Amico, 77 - 00145 Roma
Tel. 0039-06-57335655   fax 0039-06-57335771
E-mail: dip_eco@uniroma3.it
http://dipeco.uniroma3.it

DIPARTIMENTO DI ECONOMIA

PROBABILITY FORECASTS AND PREDICTION MARKETS

A. Philip Dawid - Julia Mortera

Probability Forecasts and Prediction Markets

Julia Mortera* A. Philip Dawid†

Abstract

We give an overview of various topics tied to the expression of uncertainty about a variable or event by means of a probability distribution. We first consider methods used to evaluate a single probability forecaster and then revisit methods for combining several experts' distributions. We describe the implications of imposing coherence constraints, based on a specific understanding of "expertise". In the final part we revisit from a statistical standpoint some results in the economics literature on prediction markets, where individuals sequentially place bets on the outcome of a future event. In particular we consider the case of two individuals who start with the same probability distribution but have different private information, and take turns in updating their probabilities. We note convergence of the announced probabilities to a limiting value, which may or may not be the same as that based on pooling their private information.

Key Words: consensus; expert opinion; information aggregation; probability forecast; sequential prediction
JEL No. C530, C180, D700, D810, G140, E370

1 Introduction

In this Chapter we give an overview of various topics tied to probability forecasting, *i.e.*, the use of probability distributions to express uncertainty about future events—a problem area to which Steve Fienberg made seminal contributions. In particular, we consider methods for assisting and assessing a single forecaster; methods for combining the probability forecasts of several forecasters; and prediction markets, where forecasters take turns to announce their current probabilities, taking into account previous announcements.

We start, in § 2, by reviewing methods for motivating and assessing a single forecaster. Important tools here include *proper scoring rules*, which motivate the forecaster

*Università Roma Tre, Italy and University of Bristol, UK.
†University of Cambridge, UK.

to give honest predictions; *calibration*, which compares average forecasts to observed frequencies, and *resolution* and *refinement*, which reflect expertise in the subject area.

In § 3 we describe methods for *opinion pooling*, where a decision maker consults a number of experts who give their opinions as probability statements, and needs to combine these somehow. We consider model-based and axiomatic approaches, and the application of coherence constraints under a specific definition of what constitutes expertise as seen by You, the decision-maker. These approaches are illustrated with the linear and logarithmic opinion pools.

Section 4 discusses *prediction markets*, which are venues where individuals trade predictions on uncertain future events, and allow participants to stake bets on future events. The individuals take it in turns to update their probabilities for a future event, taking into account the previously announced probabilities of the other individuals, which may be based on unannounced private information. We show that there will always be convergence to a limiting value, but this may or may not be the same as the value they could achieve if they were able to pool all their private information.

Finally § 5 reviews the main contributions of this Chapter.

2 Evaluating a single probability forecaster

Following Dawid (1986), consider a forecaster F who is required to describe his uncertainty about some unknown event A (coded $A = 1$ if A happens, $A = 0$ if not) by quoting a value $q \in [0, 1]$, intended to be interpreted as his personal probability for the event A. So long as $q \neq 0$ or 1, one might consider that neither outcome of A could discredit F's quote. Nevertheless a higher value is clearly better when $A = 1$, and a lower value when $A = 0$. Here we consider ways of motivating and evaluating F, both for single and for multiple events.

2.1 Scoring rules

To induce F to give an honest prediction, we might penalise him with a loss $S(a, q)$, depending on his quoted probability forecast q and the eventual outcome a (= 0 or 1) of A. Such a loss function S is termed a *scoring rule*. We assume that the forecaster F wishes to minimise his expected loss. Let $p = \Pr(A = 1)$ be F's true subjective probability of A. Then when he quotes probability value q his expected loss is $S(p, q) := pS(1, q) + (1 - p)S(0, q)$. The forecaster should thus choose q to minimise $S(p, q)$. The scoring rule S is called *[strictly] proper* if, for any true probability p, the expected loss $S(p, q)$ is minimised if [and only if] $q = p$. Under such a scoring rule, honesty is the best policy.

There is a wide variety of proper scoring rules, which can be tailored to emphasise different parts of the probability range. Important examples are the following.

a) The Brier score or quadratic loss function (Brier, 1950; de Finetti, 1954):

$$\begin{aligned} S(1,q) &= (1-q)^2 \\ S(0,q) &= q^2. \end{aligned}$$

b) The logarithmic scoring rule (Good, 1952):

$$\begin{aligned} S(1,q) &= -\log q \\ S(0,q) &= -\log(1-q). \end{aligned}$$

As well as motivating honesty before the event, a proper scoring rule can be used after the event, to quantify the quality of the forecaster's performance, in the light of the observed outcome a, by means of the realised score $S(a,q)$ (a lower value being better). Different forecasters, with their differing q's, can thus be compared.

When a forecaster makes a sequence of probability forecasts, for multiple events, additional evaluation criteria become available. In particular, we can assess the *calibration* and *resolution* of the forecasts issued.

2.2 Calibration

Suppose that, over a long sequence, F has issued probability forecast p_i for event A_i. Now choose $\pi \in [0,1]$, and consider all those occasions i for which $p_i = \pi$ (to a good enough approximation). Supposing there are many such occasions, let $\rho(\pi)$ denote the relative frequency of success ($A_i = 1$) on these occasions. A plot of $\rho(\pi)$ against π is the forecaster's *calibration curve*, and the forecaster is said to be *well-calibrated*, or *probability calibrated*, when he is "getting the relative frequencies right", *i.e.*, $\rho(\pi) \approx \pi$ for all values of π used. In meteorology, calibration is called also termed validity or reliability, and a well-calibrated forecaster is called perfectly reliable. It is shown in Dawid (1982) that, when events arise and are predicted in sequence, probability calibration is a necessary (though not sufficient) requirement of a good forecaster.

2.3 Resolution

Probability calibration is a fairly weak constraint on a forecaster. It will hold for the "”naïve forecaster", who quotes the same value q for every A_i, so long as $q = \pi_0$, the overall relative requence of success; as well as for the ideal "perfect forecaster", who has a crystal ball and so can always give probability 1 to the outcome a_i of A_i that actually occurs (so $q_i = a_i$). Although both are well-calibrated, the latter is doing a much more useful forecasting job than the former.

More generally, a good forecaster should be able to issue many forecasts close to the extreme values 0 or 1, with few intermediate values, while remaining well-calibrated. The same criterion can be applied to an uncalibrated forecaster, if we first replace each issued probability q_i by its recalibrated version $r_i = \rho(q_i)$. The term "resolution" refers to the extent to which a forecaster's (possibly recalibrated) forecast probabilities are widely dispersed on the unit interval. Thus a weather forecaster's resolution is a reflection of his knowledge of, and skill in, forecasting the weather, whereas his calibration addresses his entirely different ability to quantify his uncertainty appropriately.

Resolution can be quantified in various ways, for example by the variance of the recalibrated forecasts. More generally, let S be an arbitrary proper scoring rule. We might assess a forecaster's overall performance, over n events, by his total achieved penalty score, $S_+ := \sum_{i=1}^n S(a_i, q_i)$, where q_i is his quoted probability for A_i, and $a_i = 0$ or 1 is the outcome of A_i. This total score can be used to compare different forecasters. Now introduce the *entropy function* associated with S, $H(p) := S(p,p)$, which is a concave function of p; and the associated *discrepancy function*, $D(p,q) := S(p,q) - H(p)$, which is non-negative and vanishes for $q = p$. (For the Brier score, $H(p) = p(1-p)$, and $D(p,q) = (p-q)^2$.) Let $r_i = \rho(q_i)$ be the recalibrated version of q_i. Then (DeGroot and Fienberg, 1983) we can decompose $S_+ = S_1 + S_2$, where

$$\begin{aligned} S_1 &= \sum_{i=1}^n D(r_i, q_i) \\ S_2 &= \sum_{i=1}^n H(r_i). \end{aligned}$$

We see that $S_1 \geq 0$, with equality if $r_i = q_i$ for all i: S_1 thus penalises poor calibration. As for S_2, since H is concave it is smaller when the recalibrated forecasts (r_i) are clustered near 0 and 1, and thus S_2 penalises poor resolution. We can use these components of the overall score to compare forecasters in terms of their calibration and/or their resolution.

We can further decompose $S_2 = nH(\pi_0) - S_3$, where $S_3 = \sum_{i=1}^n D(r_i.\pi_0)$, and $\pi_0 = n^{-1} \sum_{i=1}^n a_i$ is the overall relative frequency of success. Since the first term is fixed, a larger S_3 indicates better resolution. For the Brier score this delivers the variance criterion.

2.4 Refinement

DeGroot and Fienberg (1983) describe a partial ordering between forecasters which is related to resolution. This is based on the theory of sufficiency in the comparison of statistical experiments (Blackwell, 1951).

Consider two forecasters, F and F', who issue respective forecasts (q_i) and (q_i') for the same sequence of events (A_i), with outcomes (a_i). We can suppose both forecasters are well-calibrated; if not, we work with their recalibrated forecasts, rather than the raw values. Then we say that F is *more refined* than F' if there exist a specification of a conditional distribution, $p(q' \mid q)$, of q' given q, such that, both for $\alpha = 0$ and for $\alpha = 1$, we can generate the (empirical) distribution of q', for those events having $a = \alpha$, by first generating q from its distribution given $a = \alpha$, and then generating q' from $p(q' \mid q)$. In this case we can consider q' as a noisy version of q, with the noise unrelated to the true outcome—which suggests that F' has poorer performance than F. Indeed, it can be shown that, when this property holds, the resolution score S_2 for F will not exceed that for F', for any proper scoring rule S.

3 Combining several opinions

It is sometimes necessary to construct a single opinion by combining a number of individual opinions. A decision maker might consult a number of experts (financial, meteorological, medical, *etc.*) before reaching a final decision.

We can distinguish three types of problem:

(a). When opinions are expressed as probability distributions. For reviews of opinion pooling in this setting, see for example Genest and Zidek (1986); DeGroot and Mortera (1991); Dawid et al. (1995); Clemen and Winkler (1999); Ranjan and Gneiting (2010). Marschak and Radner (1972) developed team theory from an economic perspective.

(b). Group decision making when opinions are expressed as preferences among alternatives: see for example Arrow (1951); Luce and Raiffa (1958); Laffont (1979).

(c). Meta-analysis where different quantitative methods are used to combine the results of different studies on the same topic.

Here we will be concerned only with problems of type (a). We shall suppose that the experts' opinions are expressed as probability distributions, over a fixed set of events and quantities of interest, but the data underlying those opinions remain undisclosed.

The ideal approach to merging several experts' views would be for each of them to report all the data and background knowledge on which his or her opinions are based, and for You, the decision maker, to combine all this information with Your own prior opinions, and any additional data You may have, using Bayes's theorem; but in the absence of access to the underlying data, You can only work with the experts' opinions, expressed as probability distributions. Their distributions will most probably differ. Your task is to combine these differing opinions, somehow, into a single distribution to use as Your own. Let the k experts $E_1, E_2, \ldots E_k$, give their probability predictions $\Pi_1, \Pi_2, \ldots \Pi_k$, for an uncertain quantity, perhaps an event A or the parameter θ of a distribution. You must pool the experts' distributions to form Your resulting aggregate or pooled distribution, Π.

3.1 Model-based approach

In this approach the experts' opinions are modelled as data (for You) and, on combining the data with Your own prior opinions, using Bayes's theorem, You can construct Your own Posterior distribution. This approach is taken by, among others, Winkler (1981); French (1986); Lindley (1983); Berger and Mortera (1991). For the case of a single event, Winkler (1981) assumes that the various log-odds have a multivariate normal distribution.

However, the process by which probability assessments are generated is not in general very easy to formalise—unlike the mechanisms by which experimental data are typically generated. The model needs to take into account the decision maker's opinion, the dependence between that and the experts' opinions, the interdependencies among the experts' opinions, and the dependence between all of these and the quantity of interest.

Group with complete interaction In the case where all experts exchange information, the problem of consensus of opinions expressed as probability distributions is an example of complete interaction. DeGroot (1974) considers a group of individuals who must act together as a team or committee, each individual in the group having his/her own subjective probability distribution for the unknown quantity. After these are all announced (round 1), each expert updates his distribution to a linear combination of all the distributions. This procedure is repeated over many rounds, the weights varying between experts but being fixed over time. DeGroot (1974) presents a condition under which the group eventually reach agreement on a common probability distribution. The model can also be applied to problems of reaching a consensus when the opinion of each member of the group is represented as a point estimate, rather than as a probability distribution. Aumann (1976) studied the dynamics of reaching a consensus through Bayesian dialogue, where conflicting opinions in a group are due solely to the fact that the members have different information sets.

Group with partial interaction The theory above bears a resemblance to the Delphi technique (Pill, 1971), used to reach agreement among a panel of experts. The Delphi technique is a purely empirical procedure and is not based on any underlying mathematical model. Again, it is applied iteratively in a sequence of rounds. At each round, the individuals are informed of the opinions of the others in the group and allowed to reassess their own opinion before proceeding to the next round. Because of the empirical nature of the Delphi technique it provides no conditions under which the experts can be expected to reach agreement or terminate the iterative process.

3.2 Axiomatic approach

In this approach a series of axioms are laid down which an opinion pooling method should satisfy. For example, if all the experts agree on a certain property—*e.g.*, that certain events are independent—then one might require that this should be preserved in the aggregated distribution. Another such property is invariance with respect to marginalisation, *i.e.* You would attain the same aggregated opinion if you first aggregate overall and then marginalise, or if each expert gives his marginal distribution and You then aggregate those. Depending on the properties assumed, a variety of aggregation methods can be derived. Among these we will discuss the *linear opinion pool* and the *logarithmic opinion pool*.

Linear Opinion Pool Stone (1961) considered the linear opinion pool:

$$\Pi = \sum_{i=1}^{m} w_i \Pi_i, \tag{1}$$

where $w_i \geq 0$ and $\sum_{i=1}^{m} w_i = 1$. He suggested that the opinion pool is democratic if You use equal weights $w_1 = w_2 = \cdots = w_m = 1/m$.

The linear opinion pool has both advantages and disadvantages. McConway (1981) proved that if you require the marginalization property then the rule for aggregation must be linear (if at least three non trivial events exist). The weights w_i can be interpreted as reflecting the previous performance of the experts. DeGroot and Mortera (1991) derived the optimal weights according to a criterion based on the Brier score.

Ranjan and Gneiting (2010) show that the linear opinion pool is uncalibrated, even when the individual probability forecasts are calibrated.

Logarithmic Opinion Pool The logarithmic opinion pool is given by

$$\log \pi = w_0 + \sum_{i=1}^{m} w_i \log \pi_i,$$

where π [resp., π_i] is the density function of Π [resp., Π_i], and w_0, a function of $(w_1, \ldots, w_m)$, is chosen to ensure that Π is a probability distribution. This was derived by Weerahandi and Zidek (1981). An important property of the logarithmic opinion pool is its consistency under aggregating and updating *i.e.* if You first aggregate opinions and then update the pooled opinion when new information is available, or if the experts first update their opinions with the new information and You then aggregate these. However, the weights do not have a simple interpretation, and if a probability given by any expert for an event is zero then the pooled probability is zero, whatever weight he/she has, and whatever the other experts' opinions are.

3.3 Coherent combination

Dawid et al. (1995) investigate coherent methods for combining experts' opinions, when these are expressed as probabilities for some fixed event A. Neither axiomatic nor modelling assumptions are made. Instead a restricted definition is used of what constitutes *expertise*, as seen by You, the decision-maker: an expert is considered to be someone who "shares Your world-view", *i.e.* if You both had identical information, You would both have identical opinions. However, the expert may know more than You do. It is assumed that the probabilities the expert provides are correctly and coherently computed.

Suppose You have access to k different experts. If You were to obtain a probability for an event A from just one of these, You should adopt it as Your own; but the different experts' probabilities will typically differ, since they will be based on differing information. You require a *combination formula* to apply to the full collection $(\Pi_1, \cdots, \Pi_k)$ of expert probabilites, to compute Your own probability Π.

Before You consult the experts, their various reports $(\Pi_1, \cdots, \Pi_k)$ will be, for You, uncertain random quantities, jointly distributed together with the uncertain event A of interest. Let P^* denote Your overall joint distribution on the random quantities $(\Pi_1, \cdots, \Pi_k, A)$ and let P denote the implied distribution for the (Π_i), marginalizing over A.

The laws of coherence require that, on learning all the experts' probabilities, You should assign probability $P^*(A \mid \Pi_1, \cdots, \Pi_k)$ to A. This yields the combination formula

$$\Pi = \Phi(\Pi_1, \cdots, \Pi_k) \equiv P^*(A \mid \Pi_1, \cdots, \Pi_k). \tag{2}$$

Note that, if expert i bases Her probability on observation of X_i, then $\Pi_i \equiv P^*(A \mid X_i)$ where here P^* is extended to encompass the (X_i); but in general the value of X_i will not be fully recoverable from that of Π_i, so that the right hand side of (2) will not usually be the same as $P^*(A \mid X_1, \cdots, X_k)$.

The question addressed is: When will a given combination formula Φ be coherently compatible with some joint distribution P for the experts' reported opinions? *i.e.*, When will there be some overall joint distribution P^* under which $\Pi_i \equiv P^*(A \mid \Pi_i)$, the implied distribution for $(\Pi_1, \cdots, \Pi_k)$ is P, and (2) holds?

Compatibility Consider the case of $k = 2$ experts and an event A, where expert E_i observes X_i and reports $\Pi_i \equiv P^*(A \mid X_i)$, for $i = 1, 2$. Then, from the definition of an expert, $\Pi_i \equiv P^*(A \mid \Pi_i)$, and defining $\Phi(\Pi_1, \Pi_2) \equiv P^*(A \mid \Pi_1, \Pi_2)$, Φ must satisfy:

$$0 \le \Phi(\Pi_1, \Pi_2) \le 1 \tag{3}$$

$$\mathrm{E}_P(\Phi \mid \Pi_i) = \Pi_i, \; i = 1, 2. \tag{4}$$

Then $\mathrm{E}_P(\Phi) = \mathrm{E}_P(\Pi_1) = \mathrm{E}_P(\Pi_2) = \pi_0$, say. Thus, by Bayes's theorem

$$p^*(\pi_1, \pi_2 \mid A) \equiv \pi_0^{-1}\Phi(\pi_1, \pi_2)p(\pi_1, \pi_2),$$

where $\pi_0 = P^*(A)$.

The pair (P, Φ) determine a unique distribution P^* for (Π_1, Π_2, A) with $P^*(A \mid \Pi_i) \equiv \Pi_i$, $P^*(A) = \pi_0$. Conditions (3) and (4) are necessary and sufficient conditions for logical consistency and the pair (P, Φ) are then termed *compatible*.

Characterizations In general, the problems of characterizing all Φ's compatible with a given P, and *vice versa*, are difficult. This set, defined by (3) and (4), is convex, but not generally a simplex. It might be empty, or contain just one member, or many.

Let P denote a joint distribution for (Π_1, Π_2) having $\mathrm{E}(\Pi_1) = \mathrm{E}(\Pi_2) = \pi_0$; and let Φ be a combination formula. Define a finite measure Q by $dQ(\pi_1, \pi_2) := \Phi(\pi_1, \pi_2)p(\pi_1, \pi_2)$, and let P_i and Q_i be the marginals for Π_i under P and Q respectively. Dawid et al. (1995) show that Φ and P are a compatible pair if and only if $dQ_i(\pi_i) \equiv \pi_i dP_i(\pi_i) = dP_i^*(\pi_i)$ for $i = 1, 2$. Given P, this shows that the problem of finding a compatible Φ reduces to that of characterizing measures Q having specified marginals and with $dQ/dP \leq 1$.

As a corollary of the above, for any absolutely continuous coherent joint distribution P for (Π_1, Π_2), there exists a compatible combination formula Φ that takes values 0 and 1 only. This implies that it is logically consistent that the combination of opinions could deliver absolute subjective certainty as to whether the event A holds or not.

Most of the literature on combining opinions uses axiomatic properties or modelling assumptions to derive particular pooling recipes. Compared with these, the assumptions for coherent pooling are less restrictive.

This analysis also offers guidance for assessing pooling formulae that have been suggested from other approaches. Thus consider combination rules that can be expresses in the generalized linear form

$$g(\Phi) \equiv \alpha_1 g(\Pi_1) + \alpha_2 g(\Pi_2) + c, \tag{5}$$

for some monotonic continuous function g. These include those considered in § 3.2

Linear opinion pool Dawid et al. (1995) show that, with (3) and (4), a generalised linear opinion pool

$$\Phi \equiv \alpha_1 \Pi_1 + \alpha_2 \Pi_2 + c \tag{6}$$

has $\alpha_0 + \alpha_1 + \alpha_2 = 1$ where $c = \alpha_0 \pi_0$, with $\pi_0 := P^*(A)$. In particular, if $c \neq 0$ every distribution P^* compatible with Φ must assign the same prior probability $\pi_0 = c/(1 - \alpha_1 - \alpha_2)$ to A. When both experts agree, You will adopt their common forecast if and only if it is the exactly same as Your prior probability for A.

When α_1, α_2 and $c = \alpha_0 \pi_0$ are all non-zero, (4) implies that

$$\mathrm{E}_P(\Pi_2 \mid \Pi_1) \equiv \lambda \Pi_1 + (1 - \lambda)\pi_0 \tag{7}$$

$$\mathrm{E}_P(\Pi_1 \mid \Pi_2) \equiv \mu \Pi_2 + (1 - \mu)\pi_0 \tag{8}$$

where $\lambda := (1-\alpha_1)/\alpha_2$, $\mu := (1-\alpha_2)/\alpha_1$, *i.e.* each Π_i has a linear regression on the other.

Conversely, any joint distribution P on $[0,1]^2$ that satisfies (7) and (8) is compatible with a Φ of form (6), for

$$\begin{aligned} \alpha_1 &= \frac{1-\lambda}{1-\lambda\mu} \\ \alpha_2 &= \frac{1-\mu}{1-\lambda\mu} \\ c &= \pi_0(1-\alpha_1-\alpha_2) \end{aligned}$$

as long as it gives probability 1 to the event $0 \le \alpha_1\Pi_1 + \alpha_2\Pi_2 + c \le 1$, thus satisfying (3). This characterizes all distributions P compatible with Φ in ((6)).

Note that not all choices of the α coefficients are coherent. Since, from (7) and (8), $\lambda\mu$ is the squared correlation ρ^2 between Π_1 and Π_2, we must have $0 \le (1-\alpha_1)(1-\alpha_2)/\alpha_1\alpha_2 < 1$. In contrast with an assumption commonly made for linear opinion pools, it is not coherent for α_1, α_2 and c all to be strictly positive, since then both regression coefficients in (7) and (8) would exceed 1.

Logarithmic opinion pool A generalised logarithmic opinion pool can be expressed as

$$\operatorname{logit}\Phi \equiv \alpha_1 \operatorname{logit}\Pi_1 + \alpha_2 \operatorname{logit}\Pi_2 + \alpha_0 \operatorname{logit}\pi_0, \tag{9}$$

where $\operatorname{logit} x = \log\{x/(1-x)\}$.

This combination formula can arise as follows. Given A or $\bar{A}$, let (X_1, X_2) be bivariate normal with $\operatorname{var}(X_i \mid A) = \operatorname{var}(X_i \mid \bar{A}) = 1$, and $\operatorname{cov}(X_1, X_2 \mid A) = \operatorname{cov}(X_1, X_2 \mid \bar{A}) = \rho$, with $\rho^2 \neq 1$ and $\mathrm{E}(X_i \mid A) = \delta_i/2$ and $\mathrm{E}(X_i \mid \bar{A}) = -\delta_i/2$ for $i = 1, 2$. Let $\Pi_i := P^*(A \mid X_i)$ and $\Phi := P^*(A \mid X_1, X_2)$. Bayes's theorem yields

$$\operatorname{logit}\Pi_i = \operatorname{logit}\pi_0 + \delta_i X_i,$$

and

$$\operatorname{logit} P^*(A \mid X_1, X_2) = \operatorname{logit}\pi_0 + (1-\rho^2)^{-1}\left\{(\delta_1 - \rho\delta_2)X_1 + (\delta_2 - \rho\delta_1)X_2\right\}.$$

Hence (9) holds, with $\alpha_1 = (1-\rho\eta)/(1-\rho^2)$, and $\alpha_2 = (1-\rho\eta^{-1})/(1-\rho^2)$, where $\eta = \delta_2/\delta_1$, and $\alpha_0 = 1 - \alpha_1 - \alpha_2$.

Again we have $\alpha_0 + \alpha_1 + \alpha_2 = 1$, and if $\alpha_0 \neq 0$, (9) determines π_0. Since $\alpha_0 = -(1-\rho^2)\alpha_1\alpha_2$, again α_0, α_1 and α_2 cannot all be strictly positive.

Conditional independence If the two experts' opinions are conditionally independent given both A and $\bar{A}$, $\Pi_1 \perp\!\!\!\perp \Pi_2 \mid (A, \bar{A})$, then

$$\operatorname{logit}\Phi \equiv \operatorname{logit}\Pi_1 + \operatorname{logit}\Pi_2 + c, \tag{10}$$

where $c = -\operatorname{logit}\pi_0$.

Dawid et al. (1995) prove the following theorem which characterizes all joint distributions compatible with (10).

Theorem 1 *A necessary and sufficient condition for a joint density* $f(\pi_1, \pi_2)$ *to be compatible with* Φ *in (10) is that*

$$f(\pi_1, \pi_2) \equiv \{[1-\pi_0)\pi_1\pi_2 + \pi_0(1-\pi_1)(1-\pi_2)]/(\pi_0(1-\pi_0))\}g(\pi_1, \pi_2) \qquad (11)$$

where $\pi_0 = (1+e^c)^{-1}$ *and* g *is a density such that* $\mathrm{E}_g(\Pi_1 \mid \Pi_2) \equiv \mathrm{E}_g(\Pi_2 \mid \Pi_1) \equiv \pi_0$. *In this case,* π_0 *is the common expectation of* Π_1 *and* Π_2 *under* f *and, thus, the prior probability of* A.

4 Prediction Markets

This part of the paper revisits some results appearing in the economics literature from a statistical point of view.

A prediction market—also known as a predictive market, an information market, a decision market, or a virtual market—-is a venue where experts can trade predictions on uncertain future events, and can stake bets on various events occurring. Such events might be, for example, an election result, a terrorist attack, a natural disaster, commodity prices, quarterly sales or sporting outcomes. Prediction markets also offer trade in possible future outcomes on securities markets, in which case participants who use it are buying something like a futures contract. The Iowa Electronic Markets (`http://tippie.uiowa.edu/iem/`) of the University of Iowa Henry B. Tippie College of Business is one of the main prediction markets in operation. Also companies like Google have their own internal prediction markets. Prediction markets sometimes operate as an open market like the stock market, or as a closed market akin to a betting pool. A prediction market translates the wisdom of crowds into predictive probabilities.

For example, suppose that in a prediction market one can bet whether A occurs (before time t), and individuals can trade contracts among each other. Consider a contract that pays 1 if event A occurs, and 0 otherwise. Say the current market price for the contract is 0.58. Offers to buy and sell are fixed at 0.57 and at 0.59, respectively. Now you can either pay 0.59 instantly, or post an offer to pay 0.58 and see if any one is willing to sell at that price. If so the new market price, 0.58, becomes the consensus probability.

Prediction markets have been discussed by Aldous (2013); Arrow et al. (2008); Hanson (2003); Chen et al. (2010); Hanson et al. (2006); Wolfers and Zitzewitz (2008); Strähl and Ziegel (2015), among others.

4.1 Basic setup

We shall focus on the opinions of a specific individual, "You", possibly but not necessarily a participant in the market, and how these opinions change in the light of accumulating experience. We suppose that Your opinions are expressed as a joint probability distribution, Pr, over all relevant variables. Other individuals may have their own probabilities for various events, but for You these are data. In the sequel, all probabilities are computed under Your distribution Pr.

We shall again interpret the term "expert" in the sense of DeGroot (1988); Dawid et al. (1995). That is, an individual E is an expert (for You) if E started with exactly the same joint probability distribution Pr over all relevant variables as You, and has observed everything that You have observed, and possibly more. Then when You learn (just) the probability Π that E assigns to some event A, Your updated probability for A will be Π. That is, You will agree with the expert.

In the context of a prediction market, suppose that experts $E_1, E_2, \ldots$, sequentially announce their probability predictions $\Pi_1, \Pi_2, \ldots$, for a future event A. Thus E_i is the expert that makes the forecast at time i, and we allow that the same expert could make forecasts at different times. At time i expert E_i has access to all previous forecasts $\Pi_1, \ldots, \Pi_{i-1}$, and possibly additional private information H_i; but E_i will typically not have access to the private information sets $H_1, \ldots, H_{i-1}$ that the previous experts used in formulating their forecasts. However, in some markets there is an option for forecasters to leave comments, which could give partial information K_i (which might be empty) about H_i. We assume that each forecaster is aware of all such past comments. Thus $\Pi_i = \Pr(A \mid T_i)$, where $T_i := (K_1, \Pi_1, \ldots, K_{i-1}, \Pi_{i-1}, H_i)$ is the total information available to E_i.

The full public information available just after time i is $S_i := (K_1, \Pi_1, K_2, \Pi_2, \ldots, K_i, \Pi_i)$. Note that S_i and T_i both contain all the information made public up to time $i-1$. They differ however in the information they contain for time i: T_i incorporates the totality, H_i, of expert E_i's information, both public, K_i, and private, whereas S_i incorporates only E_i's public information, K_i, and her announced probability forecast, Π_i, for A at time i. The information sets (T_i) are not in general increasing with i, since H_i is included in T_i but need not be in T_{i+1}. The information sets (S_i) are however increasing. The following Lemma and Corollary show that, for You, for the purposes of predicting A both information sets T_i and S_i are equivalent, and Your associated prediction is just the most recently announced probability forecast.

Lemma 1 $\Pr(A \mid S_i) = \Pr(A \mid T_i) = \Pi_i$.

Proof. Since $T_i \supseteq S_i \ni \Pi_i$,

$$\begin{aligned} \Pr(A \mid S_i) &= \mathrm{E}\{\Pr(A \mid T_i) \mid S_i\} \\ &= \mathrm{E}(\Pi_i \mid S_i) \\ &= \Pi_i \\ &= \Pr(A \mid T_i). \end{aligned}$$

□

Corollary 2 *If You observe the full public information S_i, and have no further private information, Your conditional probability for A is just the last announced forecast Π_i.*

4.2 Convergence

From Lemma 1 and the fact that the information sequence (S_i) is increasing, we have:

Corollary 3 *The sequence (Π_i) is a martingale with respect to (S_i).*

Then by Corollary 3 and the martingale convergence theorem, we have:

Corollary 4 *As $i \to \infty$, Π_i tends to a limiting value Π_∞.*

The variable Π_∞ is random in the sense that it depends on the initially unknown (to You) information sequence $S_\infty := \lim S_i$ that will materialise, but will be a fixed value for any such sequence.

A perhaps surprising implication of Corollary 4 is that, eventually, introduction of new experts will not appreciably change the probability You assign to A—whatever new private information they may bring will be asymptotically negligible compared with the accumulated public information. We term Π_∞ the *consensus probability* of A, and the information S_∞ on which it is based the *consensus information set*. The information S_∞ is *common knowledge* for all experts in the sense of Aumann (1976): see Geanakoplos (1992a,b); Nielsen (1984); McKelvey and Page (1986).

It might be considered that the limiting value Π_∞ has succeeded in integrating all the private knowledge of the infinite sequence of experts. As we shall see below this is sometimes, but not always, the case.

4.3 Two experts

As a special case, suppose we have a finite set of experts, $E_1, \ldots, E_N$, and we take $E_{N+1} = E_1$ (so $H_{N+1} = H_1$), $E_{N+2} = E_2$, *etc.* Thus we repeatedly cycle through the experts. Continuing for many such cycles, eventually we will get convergence to some Π_∞, at which point each expert will not be changing her opinion based on the total sequence of publicly announced forecasts, even though she may still have access to additional private information.

At convergence, it will thus make no difference to expert E_i to incorporate (again) her private information H_i. Consequently we have:

Proposition 5 *For each i, $A \perp\!\!\!\perp H_i \mid S_\infty$.*

In the sequel we consider in detail the case $N = 2$ of two experts, who alternate $E_1, E_2, E_1, E_2, \ldots$ in updating and announcing their forecasts. Geanakoplos and Polemarchakis (1982) have studied this in the case that there is no side-information, and each expert E_i's set of possible private information has finite cardinality, k_i say. They show that exact consensus is reached in at most $k_1 + k_2$ rounds.

Dutta and Polemarchakis (2014) give a simple example, with two experts, that shows that the order in which the experts play can matter. In their example they show that when one of the experts starts they reach complete consensus (*i.e.*, equivalent to pooling their private information), whereas on changing the starting order they only reach a limited consensus. Dutta and Polemarchakis (2014) further show that if an expert has additional information this can produce a weaker consensus. They call this "obfuscation".

4.4 Vacuous consensus

We start with some examples where the experts learn nothing from each other's forecasts—although they would learn more if they were able to communicate and pool their private data.

Example 1 Parity check

This example is essentially the same as that described by Geanakoplos and Polemarchakis (1982, p. 198).

Let X_1, X_2 be independent fair coin tosses. Expert E_i observes only X_i $(i = 1, 2)$. Let A be the event $X_1 = X_2$. This has prior probability 0.5.

On observing his private information X_1, whatever value it may take, E_1's probability of A is unchanged, at 0.5. His announcement of that value is therefore totally uninformative about the value of X_1. Consequently E_2 can only condition on her private information about X_2—which similarly has no effect. The sequence of forecasts will thus be 0.5, 0.5, 0.5, ... Convergence is immediate, but to a vacuous state.

However, if the experts could pool their data, they would learn the value of A with certainty. □

Example 2 Bivariate normal

With this example, we generalise from predicting an uncertain event to predicting an uncertain quantity.

Suppose that E_1 observes X_1, and E_2 observes X_2, where (X_1, X_2) have a bivariate normal distribution with means $\mathrm{E}(X_i) = 0$, variances $\mathrm{var}(X_i) = 1$, and unknown correlation coefficient ρ—which is what they have to forecast. Let ρ have a prior distribution Π_0. Since X_1 is totally uninformative about ρ, E_1's first forecast is again Π_0, and so is itself uninformative. Again, E_2 has learned nothing relevant to ρ, and so outputs forecast Π_0; and so on, leading to immediate convergence to a vacuous state. However the pooled data (X_1, X_2) is informative about ρ (though does not determine ρ with certainty). □

In the above examples, each expert's private information was marginally independent of the event or variable, generically Y say, being forecast, with the immediate result that the consensus forecast was vacuous, the same as the prior forecast. Conversely, suppose the consensus is vacuous. That is to say,

$$Y \perp\!\!\!\perp S_\infty. \tag{12}$$

From Proposition 5 (trivially generalised) we have

$$Y \perp\!\!\!\perp H_i \mid S_\infty. \tag{13}$$

Combining (12) and (13), we obtain $Y \perp\!\!\!\perp (H_i, S_\infty)$ whence, in particular,

$$Y \perp\!\!\!\perp H_i.$$

Hence the consensus will be vacuous if and only if each expert's private information is, marginally, totally uninformative. The argument extends trivially to any finite number of experts.

4.5 Complete consensus

We use the term *complete consensus* to refer to the case that the consensus forecast will be the same as the forecast based on the totality of the private information available to all the individual forecasters. A simple situation where this will occur is when Π_i is a one-to-one function of H_i, so that, by announcing Π_i, expert E_i fully reveals her private information.

Example 3 Overlapping Bernoulli trials
Let θ be a random variable with a distribution over $[0, 1]$ having full support. Given θ, let $Y_0 \sim B(n_0, \theta)$, $Y_1 \sim B(n_1, \theta)$, $Y_2 \sim B(n_2, \theta)$, and $A \sim B(1, \theta)$, all independently.

Suppose E_1 observes $X_1 = Y_0 + Y_1$, and E_2 observes $X_2 = Y_0 + Y_2$. At the first stage, E_1 computes and announces $\Pi_1 = \Pr(A \mid X_1)$—which is a one-to-one function of X_1. For example, under a uniform prior distribution for θ, $\Pi_1 = (X_1 + 1)/(n_0 + n_1 + 2)$. Then at stage 2, E_2 will have learned X_1, and also has private information X_2. Thus

$\Pi_2 = \Pr(A \mid X_1, X_2)$, the correct forecast given the pooled private information of E_1 and E_2 (though different from that based on full knowledge of (Y_1, Y_2, Y_2)). Further cycles will not change this probability, which will be the consensus. □

Example 4 Linear prediction
Consider variables $\mathbf{X} = (X_1, \ldots, X_k)$, $\mathbf{Z} = (Z_1, \ldots, Z_h)$ and (scalar) Y, all being jointly normally distributed with non-singular dispersion matrix. Expert 1 observes $H_1 = \mathbf{X}$, Expert 2 observes $H_2 = \mathbf{Z}$, and they have to forecast Y. Each time an expert announces her predictive distribution for Y, she is making known the value of her predictive mean of Y, which will be some linear combination of the predictor variables $(\mathbf{X}, \mathbf{Z})$. So generically we would expect convergence of the forecasts, after at most $\min\{k, h\}$ rounds, to the full forecast based on the pooled information $(\mathbf{X}, \mathbf{Z})$. This has been shown by Dutta and Polemarchakis (2014). However they did not give a numerical illustration, which we now supply.

We have made use of the 93CARS dataset (Lock, 1993), containing information on new cars for the 1993 model year. There are $n = 82$ complete cases with information on 26 variables, including price, mpg ratings, engine size, body size, and other features. We took $\mathbf{X} = (X_1, \ldots, X_{11})$ to be the variables 7 to 17, $\mathbf{Z} = (Z_1, \ldots, Z_9)$ to be the variables 18 to 26, and Y to be variable 5 (Midrange Price).

Let S denote the uncorrected sum-of-squares-and-products matrix based on the data for these variables. The fictitious model we shall consider for the prediction game has $(\mathbf{X}, \mathbf{Z}, Y)$ multivariate normal, with mean $\mathbf{0}$ and dispersion matrix $\Sigma = S/n$. The predictive distribution of Y, based on any collection of linear transforms of the X's and Z's, will then be normal, with a mean-formula that can be computed by running the zero-intercept sample linear regression of Y on those variables, and variance that will not depend on the values of the predictors. Note that, although our calculations are based on the sample data, the values computed are not estimates, but are the correct values for our fictitious model.

Let U_1 be the variable so obtained from the sample regression Y on $\mathbf{X} \equiv (X_1, \ldots, X_{11})$. Recall that both experts are supposed to know the model, hence Σ, and know which variables each is observing. Consequently both know the form of U_1, but initially only E_1, who knows the values of $(X_1, \ldots, X_{11})$, can compute its value, u_1 say. Since his round-1 forecast for Y is normal with mean u_1, while its variance is already computable by both experts, the effect of E_1 issuing his forecast is to make the value u_1 of U_1 public knowledge.

It is now E_2's turn to play. At this point she knows the values of U_1 and $(Z_1, \ldots, Z_9)$, and her forecast is thus obtained from the sample regression of Y on these variables. Let this regression function (computable by both experts) be V_1; then at this round E_2 effectively makes the value v_1 of V_1 public.

Now at round 2, E_1 regresses Y on $(X_1, \ldots, X_{11}, V_1)$ (U_1, which is a linear function of his privately known X's, being redundant), and announces the value u_2 of the computed regression function U_2. And so on.

The relevant computations are easy to conduct using the statistical software package `R` (R Development Core Team, 2011). At each stage, we compute the 82 fitted values based on the regression just performed. These can then be used as values for the new predictor variable to be included in the next regression. Moreover, convergence of the forecast sequence will be reflected in convergence of these fitted values.

As a numerical illustration, suppose E_1 has observed

$$\mathbf{X} = \mathbf{x} = (16, 25, 2, 1, 8, 4.6, 295, 6000, 1985, 0, 20.0),$$

and E_2 has observed

$$\mathbf{Z} = \mathbf{z} = (5, 204, 111, 74, 44, 31.0, 14, 3935, 1).$$

Before entering the prediction market, E_1's point forecast for Y, based on his data $\mathbf{X} = \mathbf{x}$, is $u_1 = 40.6163$, and E_2's point forecast for Y, based on her data $\mathbf{Z} = \mathbf{z}$, is $v_0 = 30.6316$. If they could combine their data, the forecast, based on $(\mathbf{X}, \mathbf{Z}) = (\mathbf{x}, \mathbf{z})$, would be 39.73925.

On entering the market, the sequence of their predictions is as given in Table 1. We

i:	1	2	3	4	5	6	7	8	9	10	...
u_i:	40.62	39.49	39.34	39.51	39.55	39.54	39.66	39.75	39.73917	39.73925	...
v_i:	38.28	39.40	39.46	39.54	39.56	39.63	39.67	39.74	39.73924	39.73925	...

Table 1: Sequence of market predictions for Y

observe convergence, both for the fitted values and the predicted standard deviations, from round 10 onwards. As soon as E_1 has access to the values of $U_1, \ldots, U_9$, he effectively knows $Z_1, \ldots, Z_9$, and his forecast becomes the same as that based on the pooled data. And as soon as E_1 makes that public, E_2 can make the same forecast. The predictions of both experts will remain the same thereafter.

As a second illustration, suppose E_1 has observed

$$\mathbf{X} = \mathbf{x} = (22, 30, 1, 0, 4, 3.5, 208, 5700, 2545, 1, 21.1),$$

and E_2 has observed

$$\mathbf{Z} = \mathbf{z} = (4, 186, 109, 69, 39, 27.0, 13, 3640, 0).$$

Before entering the prediction market, E_1's point forecast for Y is 27.80968, and E_2's point forecast is 36.593865. Their market forecasts converge at round 10 to 31.22983, the forecast based on all the data.

These two examples illustrate within-sequence convergence, to a data-dependent limit. □

4.6 Limited consensus

In all the above examples, convergence was either to a vacuous state, or to a complete consensus based on the totality of the pooled private information. As the following example shows, it is also possible to converge to an intermediate state.

Example 5 Suppose θ and X_1 have independent $N(0,1)$ distributions, while, given (θ, X_1), $X_2 \sim N(\theta X_1, 1)$. Expert E_1 observes $H_1 = X_1$, while E_2 observes $H_2 = X_2$. The interest is in predicting θ. A sufficient statistic for θ, based on the combined data (X_1, X_2), is $(X_1X_2, |X_1|) = (S_1, S_2)$, say. The posterior distribution is

$$\theta \mid (S_1, S_2) = (s_1, s_2) \sim N\left(\frac{s_1}{1+s_2^2}, \frac{1}{1+s_2^2}\right).$$

Straightforward computations deliver the joint density of (X_1, X_2), marginalising over θ:

$$f(x_1, x_2) = (2\pi)^{-1}(1+x_1^2)^{-\frac{1}{2}} \exp -\frac{1}{2}\left(x_1^2 + \frac{x_2^2}{1+x_1^2}\right). \tag{14}$$

Because (14) is unchanged if we change the sign of either or both of x_1 and x_2, we deduce (what may be obvious from the symmetry of the whole set-up):

Proposition 6 *Conditionally on* $|X_1|$ *and* $|X_2|$, $\text{sign}(X_1)$ *and* $\text{sign}(X_2)$ *behave as independent fair coin-flips.*

At the first round, E_1 declares his posterior for θ, based on X_1—but, since $X_1 \perp\!\!\!\perp \theta$ this supplies no information at all about θ. (So we would get the same answer if E_2 were to go first—the order in which they announce their opinions does not matter.)

Now E_2 goes. Since $X_2 \mid \theta \sim N(0, 1+\theta^2)$, with sufficient statistic $|X_2|$, E_2 is effectively putting $|X_2|$ into the public pot.

At the start of round 2, E_1 knows X_1 and $|X_2|$. By Proposition 6, $\text{sign}(X_2)$ is still equally likely to be 1 or -1. So E_1 knows S_2, but only knows $|S_1|$—for him, S_1 is either $|S_1|$ or $-|S_1|$, each being equally likely. His posterior is thus a 50–50 mixture of the associated posteriors

$$N\left(\frac{|S_1|}{1+S_2^2}, \frac{1}{1+S_2^2}\right)$$

and

$$N\left(\frac{-|S_1|}{1+s_2^2}, \frac{1}{1+S_2^2}\right).$$

On E_1's now announcing this mixture posterior, he is effectively communicating $(|S_1|, S_2) \equiv (|X_1| \times |X_2|, |X_1|)$. The total information in the public pot is thus now equivalent to $(|X_1|, |X_2|)$.

It is now E_2's turn again. At this point she knows $(|X_1|, X_2)$, so $(|S_1|, S_2)$—but still does not know $\text{sign}(S_1)$, which again behaves as a coin-flip. Her forecast distribution is thus exactly the same as E_1's. So we get convergence to the above mixture posterior at round 2. But this limiting forecast is not the same as that based on the pooled data, which would be the relevant single component of the mixture.

Note that, at convergence, the pool of public knowledge is $(|X_1|, |X_2|)$. Since θ has the identical mixture posterior whether conditioned on $(|X_1|, |X_2|)$, on $(X_1, |X_2|)$, or on $(|X_1|, X_2)$, we have both $\theta \perp\!\!\!\perp X_1 \mid (|X_1|, |X_2|)$ and $\theta \perp\!\!\!\perp X_2 \mid (|X_1|, |X_2|)$, in accordance with Proposition 5. □

It might appear that the above behaviour is highly dependent on the symmetry of the problem, but this is not so. As the following analysis shows, the same limited consensus behaviour arises on breaking the symmetry.

Example 6 Consider the same problem as in Example 5 above, with the sole modification that the prior distribution of θ is now $N(\mu, 1)$, where μ is non-zero. The posterior distribution of θ, based on the full data (X_1, X_2) or its sufficient statistic (S_1, S_2), is now

$$\theta \mid (S_1, S_2) = (s_1, s_2) \sim \Pi(s_1, s_2) := N\left(\frac{\mu + s_1}{1 + s_2^2}, \frac{1}{1 + s_2^2}\right).$$

The following result is immediate.

Proposition 7 *Given only* $|S_1| = m_1, S_2 = m_2$*, the posterior distribution is a mixture:*

$$\theta \sim M(m_1, m_2) = \pi(1)\Pi(m_1, m_2) + \pi(-1)\Pi(-m_1, m_2) \tag{15}$$

where

$$\pi(j) = \Pr(\text{sign}(S_1) = j \mid |S_1| = m_1, S_2 = m_2) \quad (j = \pm 1). \tag{16}$$

Proposition 8 *Conditionally on* $|X_1|$ *and* $|X_2|$*:*

(i). $\text{sign}(X_1) \perp\!\!\!\perp \text{sign}(X_1 X_2)$

(ii). $\text{sign}(X_2) \perp\!\!\!\perp \text{sign}(X_1 X_2)$

Proof. (i) The joint density of (X_1, X_2), marginalising over θ, is

$$f(x_1, x_2) = (2\pi)^{-1}(1 + x_1^2)^{-\frac{1}{2}} \exp -\frac{1}{2}\left(x_1^2 + \frac{(x_2 - \mu x_1)^2}{1 + x_1^2}\right).$$

This is unchanged if we change the signs of both x_1 and x_2. Consequently, given $|X_1| = m_1, |X_2| = m_2$, $\Pr(X_1 = m_1, X_2 = m_2) = \Pr(X_1 = -m_1, X_2 = -m_2)$, while $\Pr(X_1 = m_1, X_2 = -m_2) = \Pr(X_1 = -m_1, X_2 = m_2)$. But this is equivalent to

$$\begin{aligned} \Pr(\text{sign}(X_1) = 1, \text{sign}(X_1 X_2) = 1) &= \Pr(\text{sign}(X_1) = -1, \text{sign}(X_1 X_2) = 1) \\ \Pr(\text{sign}(X_1) = 1, \text{sign}(X_1 X_2) = -1) &= \Pr(\text{sign}(X_1) = -1, \text{sign}(X_1 X_2) = -1). \end{aligned}$$

Thus $\Pr(\text{sign}(X_1) = 1 \mid \text{sign}(X_1 X_2) = 1) = \Pr(\text{sign}(X_1) = 1 \mid \text{sign}(X_1 X_2) = -1) = \frac{1}{2}$, which in particular implies $\text{sign}(X_1) \perp\!\!\!\perp \text{sign}(X_1 X_2)$.

(ii) We have

$$\begin{aligned}
\Pr(\text{sign}(X_2) = 1 \mid \text{sign}(X_1X_2) = 1) &= \Pr(\text{sign}(X_1) = 1 \mid \text{sign}(X_1X_2) = 1) \\
\Pr(\text{sign}(X_2) = 1 \mid \text{sign}(X_1X_2) = -1) &= \Pr(\text{sign}(X_1) = -1, \text{sign}(X_1X_2) = -1)
\end{aligned}$$

So from (i), conditional on $|X_1| = m_1, |X_2| = m_2$, $\Pr(\text{sign}(X_2) = 1 \mid \text{sign}(X_1X_2) = 1) = \Pr(\text{sign}(X_2) = 1 \mid \text{sign}(X_1X_2) = -1) = \frac{1}{2}$ so that, in particular, $\text{sign}(X_2) \perp\!\!\!\perp \text{sign}(X_1X_2)$.

□

In the first round, E_1 and E_2 behave exactly as before, and again, at the start of round 2, the public pot contains $|X_2|$. So now E_1 knows X_1 and $|X_2|$. In terms of the sufficient statistic he knows $(|S_1|, S_2)$, but does not know $\text{sign}(S_1)$. Moreover, by Proposition 8(i), his additional knowledge of $\text{sign}(X_1)$ contains no relevant further information about $\text{sign}(S_1)$. Consequently, he will compute and announce the mixture posterior $M(|S_1|, S_2)$. From this it is possible to deduce the values of $|S_1|$ and S_2. Hence at this point the public pot contains $(|S_1|, S_2)$.

Now E_2 knows $(|S_1|, S_2)$, but is still ignorant of $\text{sign}(S_1)$. And again, although she has the additional knowledge of $\text{sign}(X_2)$, by Proposition 8(ii) this contains no relevant further information about $\text{sign}(S_1)$. Consequently, E_2 will have the same posterior distribution $M(|S_1|, S_2)$, which will be the final (but limited) consensus.

(Note that an essentially identical analysis will hold with any prior distribution for θ.)

□

5 Discussion

Probability forecasts take explicit account of the uncertainty concerning an unknown quantity or event. We have described three important tools for motivating and assessing the performance of a single forecaster. A proper scoring rule induces the forecaster to give honest predictions, and can also be used to evaluate performance after the event. For forecasts made for a sequence of events, calibration measures success in quantifying uncertainty. Resolution measures how close calibrated forecasts come to actual outcomes, and thus reflects expertise in the subject area. Refinement is a relation between the resolutions of different forecasters, which is useful for comparing them.

When there are multiple expert forecasters, You require a method for combining their forecasts. There is strong empirical evidence that probability forecasts suitably combining all the experts' opinions generally result in better predictive performance—this is similar to the case of Bayesian model averaging, a coherent mechanism for accounting for model uncertainty which improves predictive performance. In a Bayesian approach to combining experts' opinions, the decision maker models the experts' opinions and combines them with His/Her own prior opinion, and any additional data He/She may have, using Bayes's theorem. An alternative axiomatic approach imposes constraints that a combination formula is required to satisfy. We have described in detail an approach, based on a specific understanding of "expertise", which imposes only coherence constraints. For the linear opinion pool, the most popular method for combining probability forecasts, coherence requires that not all weights are strictly positive.

We have given a detailed account of prediction markets, with special attention to the case where two experts take it in turns to update their probability of a future event, conditioning only on the revealed probabilities of the other. We have displayed a variety of behaviours for such a process. There will always be convergence to a limiting value, but this may or may not be the same as what could be achieved if the experts were able to pool all their private information.

We have supposed that, although each expert may be unaware of the private information held by the other, he does at least know which variables the other expert knows—though not their values. When even this cannot be assumed there will be much greater freedom for an expert to update his own probability on the basis of the revealed probabilities of the other. Nevertheless this freedom is restricted. The theory of Dawid et al. (1995) relates to combining the announced probabilities of a number of experts without necessarily knowing the private variables on which these are based. It would be challenging, but valuable, to extend this to the sequential case.

Acknowledgements

The authors would like to thank Mario Tirelli.

References

Aldous, D. J. (2013). Using prediction market data to illustrate undergraduate probability. *American Mathematical Monthly*, 120(7):583–593.

Arrow, K. J. (1951). *Social Choice and Individual Values*. J. Wiley, New York.

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P., Nelson, F. D., et al. (2008). The promise of prediction markets. *Science*, 320(5878):877–878.

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics*, 4(6):1236–1239.

Berger, J. O. and Mortera, J. (1991). Bayesian analysis with limited communication. *Journal of Statistical Planning and Inference*, 28(1):1–24.

Blackwell, D. (1951). Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102, Berkeley, Calif. University of California Press.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3.

Chen, Y., Dimitrov, S., Sami, R., Reeves, D., Pennock, D., Hanson, R., Fortnow, L., and Gonen, R. (2010). Gaming prediction markets: Equilibrium strategies with a market maker. *Algorithmica*, 58(4):930–969.

Clemen, R. T. and Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):187–203.

Dawid, A. P. (1982). The well-calibrated Bayesian. *Journal of the American Statistical Association*, 77(379):605–610.

Dawid, A. P. (1986). Probability forecasting. In Kotz, S., Johnson, N. L., and Read, C. B., editors, *Encyclopedia of Statistical Sciences*, volume 7, pages 210–218. Wiley-Interscience.

Dawid, A. P., DeGroot, M. H., and Mortera, J. (1995). Coherent combination of experts' opinions (with Discussion). *TEST*, 4(2):263–314.

de Finetti, B. (1954). Media di decisioni e media di opinioni. *Bulletin of the International Statistical Institute*, 24(2):144–157.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121.

DeGroot, M. H. (1988). A Bayesian view of assessing uncertainty and comparing expert opinion. *Journal of Statistical Planning and Inference*, 20(3):295–306.

DeGroot, M. H. and Fienberg, S. E. (1983). The comparison and evaluation of forecasters. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 32(1-2):12–22.

DeGroot, M. H. and Mortera, J. (1991). Optimal linear opinion pools. *Management Science*, 37(5):546–558.

Dutta, J. and Polemarchakis, H. (2014). Convergence to agreement and the role of public information. *Mathematics and Financial Economics*, 8(4):399–404.

French, S. (1986). Calibration and the expert problem. *Management Science*, 32(3):315–321.

Geanakoplos, J. (1992a). Common knowledge. In *Proceedings of the 4th Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 254–315, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Geanakoplos, J. (1992b). Common knowledge. *Journal of Economic Perspectives*, 6(1):53–82.

Geanakoplos, J. D. and Polemarchakis, H. M. (1982). We can't disagree forever. *Journal of Economic Theory*, 32(1):192–200.

Genest, C. and Zidek, J. V. (1986). Combining probability distributions: A critique and an annotated bibliography (with Discussion). *Statistical Science*, (1):114–148.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Series B*, 14:107–114.

Hanson, R. (2003). Combinatorial information market design. *Information Systems Frontiers*, 5(1):105–119.

Hanson, R., Oprea, R., and Porter, D. (2006). Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior and Organization*, 60(4):449–459.

Laffont, J.-J. (1979). *Aggregation and Revelation of Preferences*. Elsevier North-Holland, New York.

Lindley, D. V. (1983). Reconciliation of probability distributions. *Operations Research*, 31(5):866–880.

Lock, R. H. (1993). The 1993 new car data. *Journal of Statistics Education*, 1. `http://www.amstat.org/publications/jse/v1n1/datasets.lock.html`.

Luce, R. D. and Raiffa, H. (1958). *Games and Decisions: Introduction and Critical Survey*. Wiley, New York.

Marschak, J. and Radner, R. (1972). *Economic Theory of Teams*. Yale University Press.

McConway, K. J. (1981). Marginalization and linear opinion pools. *Journal of the American Statistical Association*, 76(374):410–414.

McKelvey, R. D. and Page, T. (1986). Common knowledge, consensus, and aggregate information. *Econometrica*, 54:109–127.

Nielsen, L. T. (1984). Common knowledge, communication, and convergence of beliefs. *Mathematical Social Sciences*, 8(1):1–14.

Pill, J. (1971). The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio-Economic Planning Sciences*, 5(1):57–71.

R Development Core Team (2011). R: A Language and Environment for Statistical Computing. ISBN 3-900051-07-0.

Ranjan, R. and Gneiting, T. (2010). Combining probability forecasts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):71–91.

Stone, M. (1961). The opinion pool. *The Annals of Mathematical Statistics*, 32(4):1339–1342.

Strähl, C. and Ziegel, J. F. (2015). Cross-calibration of probabilistic forecasts. arXiv:1505.05314.

Weerahandi, S. and Zidek, J. V. (1981). Multi-Bayesian statistical decision theory. *Journal of the Royal Statistical Society: Series A (General)*, 144(1):85–93.

Winkler, R. L. (1981). Combining probability distributions from dependent information sources. *Management Science*, 27(4):479–488.

Wolfers, J. and Zitzewitz, E. (2008). Prediction markets in theory and practice. In Blume, L. and Durlauf, S., editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, London, 2nd edition.