Forecast Aggregation

Itai Arieli, Yakov Babichenko, and Rann Smorodinsky

Faculty of Industrial Engineering, Technion—Israel Institute of Technology.

February 12, 2017

Abstract

Bayesian experts with a common prior that are exposed to different evidence possibly make contradicting probabilistic forecasts. A policy maker who receives the forecasts must aggregate them in the best way possible. This is a challenge whenever the policy maker is not familiar with the prior nor the model and evidence available to the experts. We propose a model of non-Bayesian forecast aggregation and adapt the notion of *regret* as a means for evaluating the policy maker's performance. Whenever experts are Blackwell ordered taking a weighted average of the two forecasts, the weight of which is proportional to its precision (the reciprocal of the variance), is *optimal*. The resulting regret is equal $\frac{1}{8}(5\sqrt{5}-11) \approx 0.0225425$, which is 3 to 4 times better than naive approaches such as choosing one expert at random or taking the non-weighted average.

1 Introduction

Just the other day we were planning our weekend activities and looked at the forecast for the weather in Tel-Aviv on Friday, January 27. In particular what

Research supported by ISF grant 2016301, Technion VPR grants, the joint Microsoft-Technion e-Commerce Lab, the Bernard M. Gordon Center for Systems Engineering at the Technion, and the TASP Center at the Technion.

interested us was the probability for rain (precipitation). Accuweather's precipitation forecast was 77% while Yahoo! had a forecast of 60% and the Weather Channel was at 90% (all three screenshots are provided in the appendix). It was unclear to us how to aggregate these conflicting forecasts although we knew all three were reputable sources and were using sound weather models and reliable data.

Our dilemma was not unique. In fact many of us face such conflicting sources of advice from experts on a daily basis. Forecasts from reliable pollsters on the outcome of the presidential elections, medical prognosis from trusted physicians, investment advice from experienced financial pundits and more.

This challenge is in fact the crux of the working of many governing bodies. In the political arena we often see ministers and legislators that are chosen electives and must decide on critical issues and policies with any subject matter expertise. Such public electives dictate health care policies, decide on military development and deployment, financial regulation and so on without prior medical / military / financial background. To do so they reach out to expert advice. Whether through ad-hoc committees, civil servants with years of experience, lobbyists and more. Similar to chosen electives, board members of commercial companies are often season business people with managerial experience but are often not familiar with industry domain of the company on which board they reside. Such board members essentially need to aggregate input from various experts to make a decision. In fact, the judicial system is a classical setting where person subject-matter ignorant individual must make a verdict based on aggregation of conflicting expert advice.

We consider a model with three agents. There are two experts who provide a forecast about the probability over some given future event. The experts agree on the prior but have access to different evidence and so form different posterior forecasts. A third agent, the policy maker (PM), completely lacks knowledge and is uneducated both with respect to the subject matter as well as the way the experts derive their advice. The PM's task is to optimally aggregate the two, possibly conflicting, forecasts. How should we evaluate the PM and what is his best course of action. These are the questions we are interested in.

To study this we should elucidate three aspects of the model:

- How to model an expert? Assume the PM would like to know the state of nature which is chosen from some prior probability distribution. A forecaster provides a probability distribution over these states. A forecaster is considered an expert if he knows the prior distribution, receives a private signal and uses Bayes rule to produces a forecast.

- How to model an ignorant policy maker? A policy maker is ignorant if he knows that the experts are indeed Bayesian but does not know the prior nor the signal structure. The PM observes the forecasts and uses these to produce his own subjective probability.

- How to measure the accuracy of a forecast? The PM, being exposed to public scrutiny must make the forecast as accurate as possible. A natural and prevalent family of measures of forecast accuracy, and the one we adopt here, is that of proper scoring rules (see Brier [4]) and in particular the square loss function. Proper scoring rules are a family of utility function that motivate a Bayesian forecaster to be truthful about his forecast. As the PM is not a Bayesian then he seeks to minimize his regret, computed as the difference between his score and that of the better expert. Unfortunately for the PM he does not know which is the better expert.

An expert PM, in contrast with our ignorant PM, is one who knows the prior and the models used by the experts. In our Bayesian framework the optimal scheme to aggregate the experts' advice and form a prediction is to apply Bayes rule, conditioning on the available information, namely the forecasts made by the experts. Unfortunately, this will not work for our ignorant PM who is not familiar with the prior nor with the model (signal structure) of both experts. Given the inadequacy of a Bayesian framework we turn to a more stringent criterion to evaluate the performance of the PM. We say that the PM can guarantee a regret of α if his expected score is no more than α higher than that of the better informed expert, for any prior and information structure for the experts. With this in mind we ask what is the lowest regret the PM can guarantee.

What is left ambiguous is what is meant by a 'better informed expert'. There are a two natural ways to interpret this. First, the ex-post approach which determines which forecast was better given the actual realization. In our forecasting setting it is clear that if the the forecasted event is realized then the expert with higher probability is the ex-post better whereas in the opposite case it is the other expert. Alternatively, in the ex-ante approach the more informed expert is determined at the outset from the Bayesian framework. Without any limitations on the framework there is no universally agreed approach to determine which expert is better ex-ante. The determination of what constitutes a better expert does make sense when the agents can be ordered according to the celebrated notion of Blackwell ordering [3]. According to this approach the two experts use the same model for computing forecasts, based on observable data. The better informed expert is the one who has more data. In other words, the better informed expert has access to the data available to the less informed expert but has collected even more data. As the data collected by each (which we deem as the signal) is non verifiable the PM cannot know which expert is superior.

In this paper we provide a closed form formula for aggregating forecasts that minimizes regret with respect to the ex-ante more informed expert (whose identity is anonymous). Our main result makes use of the notion of precision in statistics. The precision of a random variable is the reciprocal of its variance. For a Bernoulli random variable with parameter x let us denote the precision by $\phi(x) = \frac{1}{x(1-x)}$. We characterize the PM's optimal aggregation policy which is the following function we deem as *precision work*:

$$f(x_1, x_2) = \begin{cases} \frac{\sqrt{\phi(x_1)}x_1 + \sqrt{\phi(x_2)}x_2}{\sqrt{\phi(x_1)} + \sqrt{\phi(x_2)}} & \text{if } |x_1 - x_2| \geq 0.4 \\ \frac{\phi(x_1)x_1 + \phi(x_2)x_2}{\phi(x_1) + \phi(x_2} & \text{if } |x_1 - x_2| < 0.4. \end{cases} \tag{1}$$

Surprisingly, this prediction function guarantees an accuracy of at least $\frac{1}{8}(5\sqrt{5} - 11) \approx 0.0225425$. Moreover, $\frac{1}{8}(5\sqrt{5} - 11)$ is the optimal value agent can guarantee. Namely, for every aggregation scheme there exist an information structure where one agent is more informed than the other such that the difference between the more informed expert utility and agents utility is at least $\frac{1}{8}(5\sqrt{5}-11)$.

To appreciate how well this scheme performs let compare it with a few 'naive' approaches to the problem:

- The standard average, with equal weights, yields a regret that is as high as $\frac{1}{16} = 0.0625$, almost a three fold higher than the precision work.
- Choosing randomly one of the two experts with equal probability results in a regret as high as $\frac{1}{8}$
- If the PM Chooses to adopt the forecast with minimal uncertainty, as measured by entropy (in other words the more extreme forecast), and ignores the other, then his regret is as high as 4 times the regret bounds from the optimal precision work.

1.1 Related Literature

A closely related research topic is that of expert testing and in particular multiple expert testing. Al-Najjar and Weinstein [1] and Feinberg and Stewart [10] ask how a policy maker should identify which of the experts is the better informed one. In their model each expert has his subjective prior and the experts report their posterior probabilities after observing some data. This is in contrast with our approach where experts agree on the prior but observe some private signal. A 'test' takes as input a sequence of pairs composed of the realized data and the posteriors, and outputs an order over the experts. Thus, this literature is mute whenever the forecasts are made just once. One natural test for ranking experts, in the context of prediction in financial markets, is that of portfolio returns. Sandroni [11] shows that indeed the better informed expert outperforms the less informed one in the long run.

How should a policy maker integrate advice of multiple experts is essentially the crux of a strand of the machine learning literature knows as regret minimization. In that model an ignorant PM (the 'machine' in their case) repeatedly receives input from multiple experts and takes an action based on the vector of experts advice. At each stage, the PM is paid according to some function that depends on her action and the temporal state of nature. The goal of the PM is then to choose a sequence of actions that would perform as well as the best expert. The notion of a best expert is taken in hindsight. Namely we compare the PM's aggregate payoff to the payoff he would have received had he consistently chosen to follow the advice of a single agent. The goal is then to minimize this with respect to the best expert (this is the PM's 'regret'). The literature provides a variety of settings and schemes for choosing actions such that the average per-stage regret goes to zero. The reader is referred to Cesa-Bianchi and Lugosi [5] for a review on this topic.

Regret minimization is closely related to game theory as it turns out that if players in a repeated game use regret minimization schemes then the resulting play converges to an equilibrium of the game. This may happen even when players are not fully familiar with the structure of the game. This was initially observed by Hannan [8] and more recently by Foster and Vohra [7]. The reader is referred to Hart and Mas-Collel [9] for a review. The major distinction between the above mentioned work on machine learning and the current paper is that we do not consider a repeated setting but a one-shot model, for which the regret minimization is mute.

2 Model and main result

A PM must set a policy without knowing the state of nature, $\omega \in \Omega = \{0, 1\}$. This could represent whether global warming occurs, whether there is danger of some epidemic or whether a new technology that is being developed is viable. The PM consults two Bayesian experts who share a common prior but receive different signals. Let S_i be the signal space available to expert i and let $\mathbf{P} \in \Delta(\Omega \times S_1 \times S_2)$ be the common prior. We refer to the tuple (S_1, S_2, p) as an *information structure* and denote by $\mathcal{I}$ the set of all information structures.

Our work focuses on the case where one expert is more informed than the other. In other words, the more informed expert learns the signal of the less informed one and some additional signal. Formally:

Definition 1. An information structure (S_1, S_2, p) is *Blackwell ordered*, if there exists some set, S_2', such that $S_2 = (S_1 \times S_2')$ or, symmetrically, there exists some set, S_1', such that $S_1 = (S_1' \times S_2)$. Let $\mathcal{BI}$ denote the set of all Blackwell ordered information structures.

In words, the better informed expert has access to the signal available to the less informed expert and he receives an additional private signal. This notion is equivalent to the notion of Blackwell domination and Blackwell ordering, see [3].

The two experts report their posterior marginal probabilities over Ω, which the PM uses to provide his own aggregated forecast. In particular, the PM is not familiar with the information structure but only knows it is an element of $\mathcal{BI}$. In this context we refer to a function $g : [0,1]^2 \to [0,1]$ as an *aggregation scheme* and denote by $\mathcal{G}$ the set of all such aggregation schemes.

To ensure all agents report their forecasts truthfully their are incentivized by a *proper scoring rule*. A proper scoring rule is a function of a forecaster's report and the realization with the property that it is ex-ante maximized when the forecaster is truthful, see Brier [4]. In particular, the agents are endowed with the following proper scoring rule as their utility function:

$$l(a, \omega) = -(a - \omega)^2. \tag{2}$$

Given an information structure $\mathbf{P} = (S_1, S_2, p)$ the expected utility of the PM who predicts a is $l(a) = \mathbb{E}_{\mathbf{P}}[l(a, \omega)]$. The PM does not know the information structure and hence cannot maximize this. Instead his objective is to minimize his *r*egret. Recall the PM knows that one expert is better informed than the other but does not which is the better informed. His regret compares his utility with the counterfactual option where he would have known the identity of the better informed expert and would have simply adopted his forecast. Formally, given an information structure $\mathbf{P} = (S_1, S_2, p)$ (general one or Blackwell ordered one) and an aggregation scheme f the let $R(\mathbf{P}, g)$ be defined as follows:

$$R(\mathbf{P}, g) = \max_{i=1,2} \mathbb{E}_{\mathbf{P}}[l(x_i)] - \mathbb{E}_{\mathbf{P}}[l(g(x_1, x_2))]. \tag{3}$$

where x_i is expert i's posterior belief of $\omega = 1$ given his signal (that is drawn according to $\mathbf{P}$). Recall the PM does not know the information structure $\mathbf{P}$ and so cannot minimize $R(\mathbf{P}, g)$. Instead he seeks to minimize the *r*egret of his aggregation scheme, g, defined as

$$\sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, g).$$

Our challenge is to construct an aggregation scheme g that minimizes the regret whenever the experts are Blackwell ordered. The value of the minimal regret, v, is therefore:

$$v = \inf_{g} \sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, g) \tag{4}$$

where the infimum is taken across all aggregation schemes g.

To state our main result we borrow the notion of a precision of a random variable from the statistics literature:

Definition 2. The *p*recision of a random variable is the reciprocal of its. For a Bernoulli random variable with parameter x it is $\phi(x) = \frac{1}{x(1-x)}$. The *precision scheme* is an aggregation scheme of the form:

$$f(x_1, x_2) = \begin{cases} \frac{\sqrt{\phi(x_1)}}{\sqrt{\phi(x_1)}+\sqrt{\phi(x_2)}} x_1 + \frac{\sqrt{\phi(x_2)}}{\sqrt{\phi(x_1)}+\sqrt{\phi(x_2)}} x_2 & \text{if } |x_1 - x_2| \geq 0.4 \ \wedge \ x_1, x_2 \notin \{0, 1\} \\ \frac{\phi(x_1)}{\phi(x_1)+\phi(x_2)} x_1 + \frac{\phi(x_2)}{\phi(x_1)+\phi(x_2)} x_2 & \text{if } |x_1 - x_2| < 0.4 \wedge \ x_1, x_2 \notin \{0, 1\} \end{cases}$$

In addition, for $x_1, x_2 < 1$ set $f(0, x_2) = f(x_1, 0) = 0$ and for $x_1, x_2 > 0$ set $f(1, x_2) = f(x_1, 1) = 1$.

Note that f is not defined for the pair of forecasts $(0, 1)$ and $(1, 0)$ which are inconsistent with our Bayesian model and will never occur with positive probability.

We are now ready to state our main theorem:

Theorem 1. *Using a precision scheme the PM can guarantee a regret of* $\frac{1}{8}(5\sqrt{5} - 11) \approx 0.0225425$. *Moreover, no aggregation scheme guarantees a lower regret.*

Remark 2.1. Given an information structure $\mathbf{P}$ we note that $R(\mathbf{P}, g)$and $l(g(x_1, x_2))$ differ by an additive constant, independent of g. Hence choosing g to minimize the former is equivalent to maximizing the latter. This suggests that the above formulation also captures the case where the PM wants to maximize his utility.

3 Naive aggregation schemes

To appreciate the regret obtained by the precision scheme and as a warm-up, before moving to the proof of the main theorem, we study the regret of two natural aggregation schemes.

3.1 The simple average

Consider the naive aggregation scheme $f(x_1, x_2) = \frac{1}{2}x_1 + \frac{1}{2}x_2$, which is exactly the naive Degroot [6] opinion formation function. This function does not perform well in the following information structure.

Assume that the prior is $Pr(\omega = 1) = \frac{1}{2}$. Expert 1 receives no information, whereas expert 2 receives the complete information (i.e., expert 2 knows the realized state ω). In this case the pair of forecasts will be $(\frac{1}{2}, 0)$ and $(\frac{1}{2}, 1)$ with probabilities $\frac{1}{2}, \frac{1}{2}$. PM's forecast will be either $\frac{1}{4}$ or $\frac{3}{4}$ with probabilities $\frac{1}{2}, \frac{1}{2}$. PM's forecast always is $\frac{1}{4}$-far from the best expert forecast. Therefore the regret in the square-loss utilities is at least (in fact also exactly) $\frac{1}{16}$.

3.2 The minimal entropy scheme

In a Bayesian framework, whenever one of the experts forecast is extreme ($x_i \in \{0, 1\}$) while the other is not then the corresponding event surely happened and the PM must adopt this forecast independently of the information structure. A naive generalization of this is to follow the forecaster which forecasts is more informational, in terms of entropy. This implies adopting the more extreme forecaster. Formally,

$$f(x_1, x_2) = \begin{cases} x_1 \text{ if } |x_1 - \frac{1}{2}| > |x_2 - \frac{1}{2}| \\ x_2 \text{ otherwise.} \end{cases}$$

As it turns out, this aggregation scheme does not always perform well.

To see this we use the identification between Blackwell ordered information structures and martingales of size 2 (see Section 4.1 below). Consider the posterior belief martingale (X_1, X_2) with expectation $\frac{1}{2}$ where $X_1 = 0.2, 0.8$ with equal

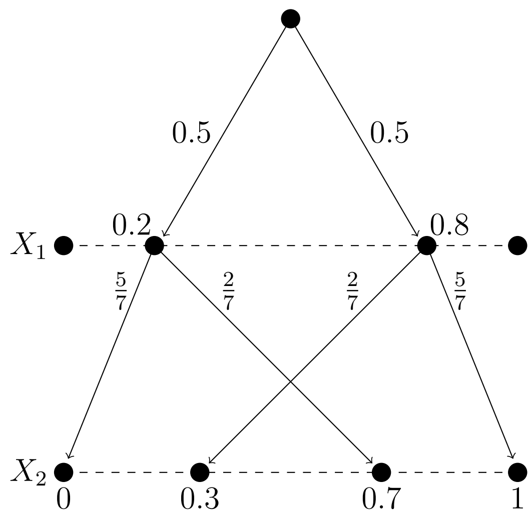


Figure 1: The martingale X_1, X_2.

probabilities. The conditional probabilities for X_2 are: $P(X_2 = 0|X_1 = 0.2) = \frac{5}{7}$, $P(X_2 = 0.7|X_1 = 0.2) = \frac{2}{7}$ and symmetrically $P(X_2 = 1|X_1 = 0.8) = \frac{5}{7}$, $P(X_2 = 0.3|X_1 = 0.8) = \frac{2}{7}$. Figure 1 visualizes this martingale.

In this information structure with probability $\frac{1}{7}$ the PM observes the pair of forecasts $(0.2, 0.7)$. Not knowing which is the better informed expert he predicts 0.2 whereas the better informed expert's forecast is 0.7. Symmetrically, with probability $\frac{1}{7}$ the PM observes the pair $(0.8, 0.3)$ and predicts 0.8 which, once again, is 0.5 away far from the forecast of the better informed expert. Thus, the induced regret is at least $\frac{2}{7}\frac{1}{4}$ which is even worse than that of the average aggregation scheme.

The analysis of the two naive forecast aggregation schemes and the corresponding information structures suggest that a regret minimizing aggregation scheme should assign weights to the forecasts that do depend on their distance from $\frac{1}{2}$ (higher distance– more weight) but not too radically. Note that this is exactly what happens in the precision scheme.

Before we turn to the proof let us introduce an alternative representation of Blackwell ordered information structures using martingales.

4 Proof of main result

In this section we reformulate forecast aggregation problem as follows: We first replace the information structure which is quite general with the induced posteriors which form a martingale. We argue that the martingale approach is without loss of generality. We then proceed to consider aggregation schemes that treat the experts in an anonymous way. Intuitively this follows from the fact that the PM has no knowledge which of the two experts is the better informed. Finally we recast the forecast aggregation problem as a problem of computing the value of a zero-sum game played between the PM who is restricted to anonymous aggregation functions and an adversary who chooses a martingale.

4.1 Martingales

Given an information scheme, $\mathbf{P}$, let X_0, X_1, X_2 denote the marginal distribution over Ω corresponding to the prior and the posteriors of expert 1 and expert 2. We observe that whenever $\mathbf{P} \in \mathcal{BI}$ is a Blackwell ordered information structure where expert 2 is more informed than expert 1 then the sequence X_0, X_1, X_2, forms a martingale. Symmetrically, the sequence X_0, X_2, X_1 forms a martingale whenever expert 1 is more informed than expert 1. The opposite also holds, namely for any martingale Z_0, Z_1, Z_2 with realization in the unit interval there exists some $\mathbf{P} \in \mathcal{BI}$ for which the corresponding posteriors, X_0, X_1, X_2, equal Z_0, Z_1, Z_2. This follows from The second statement follows from standard spiting arguments as in Aumann and Maschler [2].

Using the martingale representation of the problem, the regret $R(\mathbf{P}, f)$ has the following simple form:

$$R(\mathbf{P}, f) = \mathbb{E}_{\mathbf{P}}[(f(x_1, x_2) - x_2)^2],$$

whenever expert 2 is more informed. This follows from the following observation.

Lemma 1. *Let $\mathbf{P} \in \mathcal{BI}$ where expert 2 Blackwell dominates expert 1 and let g be any aggregation scheme. Then*

$$\mathbb{E}_{\mathbf{P}}[l(x_2, \omega) - l(g(x_1, x_2), \omega)] = \mathbb{E}_{\mathbf{P}}[(g(x_1, x_2) - x_2)^2].$$

This Lemma uses standard properties of the square loss utility, and its proof is relegated to Appendix A.

4.2 Anonymous aggregation schemes

We call an aggregation scheme g *anonymous* if $g(x_1, x_2) = g(x_2, x_1)$ for every $x_1, x_2 \in [0, 1]$. Our first step is to observe that we can restrict attention to *anonymous* aggregation schemes.

Proposition 1. *For every aggregation scheme g there exists an anonymous aggregation scheme $\tilde{g}$ such that*

$$\sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, g) \geq \sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, \tilde{g}).$$

The proof is done by showing that the anonymous aggregation scheme

$$\tilde{g} = \frac{g(x_1, x_2) + g(x_2, x_1)}{2}$$

is weakly better (in the worst case) than the aggregation scheme g. The formal proof is relegated to Appendix A.

Proposition 1 shows that the regret the PM can guarantee with anonymous aggregation scheme is the same regret as in the general case. Therefore we can restrict the PM to choose an anonymous aggregation schemes. As a result we can restrict the set on Blackwell information structures to those where Expert 2 is necessarily more informed one.

4.3 A zero sum game

Using the martingale formulation of Blackwell information structures and Proposition 1 imply the following reformulation the problem of computing the value v (defined in Equation (4)) as computing the value of the following two-player zero-sum game Γ:

- An adversary chooses a martingale X_0, X_1, X_2 of length three with values in $[0, 1]$.

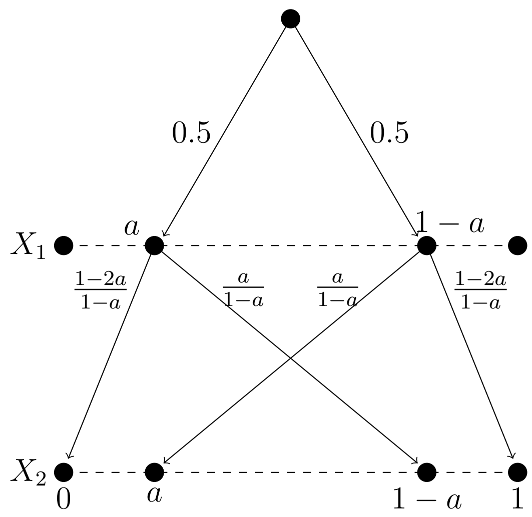


Figure 2: The optimal martingale is obtained for $a = \frac{1}{4}(3 - \sqrt{5})$.

- The PM simultaneously chooses an *anonymous* aggregation scheme g.
- The adversary's payoff is the expectation of the random variable $(g(X_1, X_2) - X_2)^2$.

Sion's theorem [12] implies that the game Γ has a value and for every ϵ both the adversary and the PM have an ϵ-optimal pure strategies. We show more than that and derive explicitly an optimal pure strategies both for the adversary and the PM.

Proposition 2. *For the game* Γ *we have*

- $Val(\Gamma) = \frac{1}{8}(5\sqrt{5} - 11)$.
- *An optimal strategy for the PM is the precision scheme (see Definition 2).*
- *An optimal strategy for the adversary is the martingale depicted in Figure 2.*

Given the above formulation of the strategies of Γ it is quite straightforward to prove these indeed form an equilibrium and that they induce the claimed value. We relegate this proof to Appendix A.

Moving back from our auxiliary game Γ to the original problem formulation, the PM should use the exact same symmetric precision scheme. The regret is

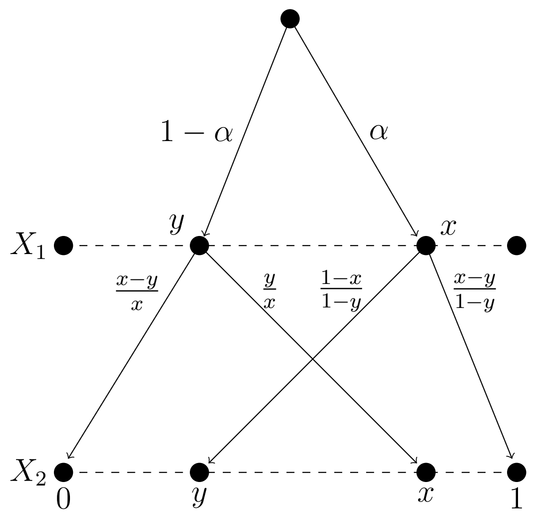


Figure 3: The martingale $\mathbf{P}_{x,y,\alpha}$.

then maximized when the information structure used has the following martingale formulation: choose between the two experts which is more informed with equal probabilities. Then proceed with the corresponding martingale.

The curious reader might ask himself how does one come up with the specific form of the optimal aggregation scheme. We provide partial intuition in the following section (Section 5).

5 Some intuition for the precision scheme

The existence of pure approximate optimal strategies in the game Γ, as previously argued, is a corollary of Sion's Min-Max Theorem [12]. Once this was verified we focused on studying the adversary's Min-Max pure strategies. Note the simplest pure strategies available to the adversary that achieve positive regret against a Bayesian PM, are martingales of the form demonstrated in Figure 3. Namely, martingales for which X_1 takes on two values x, y with probabilities $\alpha, 1-\alpha$ and X_2 can take only two pairs of values, depending on the outcome of X_1. These pairs can either be $(x, 1)$ or $(0, y)$. The corresponding probabilities are uniquely determined by the martingale property. Let us denote this family of martingales by $\mathcal{P}$, where each $\mathbf{P} \in \mathcal{P}$ is a function of three parameters, $\mathbf{P} = \mathbf{P}_{x,y,\alpha}$.

The importance of the family $\mathcal{P}$ is due to the following observation:

Proposition 3. *The adversary's MinMax strategy is in $\mathcal{P}$.*

In hindsight, Proposition 3 obviously follows from Theorem 1, which proof, in turn does not rely this proposition. However, our original proof of Proposition 3 was independent and did not use the specific structure of the optimal aggregation scheme. We omit the proof which follows tedious computations.

If the PM knows the martingale $\mathbf{P}_{x,y,\alpha}$, then his anonymous best reply is (we leave the computational details out):

$$g_{\mathbf{P}_{x,y,\alpha}}(x_1,x_2) = \begin{cases} \frac{(1-\alpha)\frac{(1-y)}{(1-x)}\cdot x+\alpha\frac{x}{y}\cdot y}{(1-\alpha)\frac{1-y}{1-x}+\alpha\frac{x}{y}} & \text{if } \{x_1,x_2\}=\{x,y\} \\ 0 & \text{if } x_1=0 \text{ or } x_2=0 \\ 1 & \text{if } x_1=1 \text{ or } x_2=1. \end{cases} \tag{5}$$

The resulting regret for the PM is therefore:

$$R(\mathbf{P}_{x,y,\alpha}, g_{\mathbf{P}_a}) = 2(x-y)^2 \frac{(1-\alpha)\frac{(1-y)}{(1-x)}\cdot \alpha\frac{x}{y}}{(1-\alpha)\frac{(1-y)}{(1-x)}+\alpha\frac{x}{y}}. \tag{6}$$

Given a pair of values, (x,y), the corresponding α that maximizes the regret, $R(\mathbf{P}_{x,y,\alpha}, g_{\mathbf{P}_a})$, is denoted by $\alpha^*(x,y)$. From first order conditions we derive a closed form solution for $\alpha^*(x,y)$:

$$\alpha^*(x,y) = \frac{\sqrt{y(1-y)}}{\sqrt{y(1-y)}+\sqrt{x(1-x)}}.$$

A PM that does not know α but knows x,y (because these are the two forecasts of the experts), may assume that adversary chooses $\alpha=\alpha^*$. In such a case PM's forecast is obtained by replacing α with the formula for $\alpha^*(x,y)$ in the first case of Equation (5):

$$\begin{aligned}\mathbf{P}_{x,y,\alpha^*}(\omega=1|\{x_1,x_2\}=\{x,y\}) &= \frac{\left(\sqrt{y(1-y)}\right)x+\left(\sqrt{x(1-x)}\right)y}{\sqrt{y(1-y)}+\sqrt{x(1-x)}} \\ &= \frac{\sqrt{\phi(x)}}{\sqrt{\phi(x)}+\sqrt{\phi(y)}}\cdot x + \frac{\sqrt{\phi(y)}}{\sqrt{\phi(x)}+\sqrt{\phi(y)}}\cdot y \end{aligned} \tag{7}$$

where ϕ is the precision function. Note that this coincides with the precision scheme whenever the two forecasts are sufficiently distinct.

Recall that we derive equation (7) under the assumption that the adversary chooses the value $\alpha = \alpha^*(x, y)$, which is optimal against a *Bayesian* PM. Once the PM fixes the scheme provided by equation (7), we can reconsider the optimal martingale for the adversary. It turns out that in some cases the adversary can choose some value $\alpha \neq \alpha^*(x, y)$ and increase the PM's regret. This happens only for martingales $\mathbf{P}_{x,y,\alpha}$ where $x - y < 0.4$. The interim conclusion was that the scheme provided in equation (7) guarantees the regret $\frac{1}{8}(5\sqrt{5} - 11)$ for all martingales $\mathbf{P}_{x,y,\alpha}$ where $x - y \geq 0.4$. By adjusting the weights to $(\frac{\phi(x)}{\phi(x)+\phi(y)}, \frac{\phi(y)}{\phi(x)+\phi(y)})$, whenever for $x - y < 0.4$, we derive the optimal scheme.[1]

6 Concluding remarks

We have shown that a completely ignorant PM can still make good use of contradicting forecasts suggested by two Bayesian experts who share a common prior but do not share their evidence. We borrow the notion of 'regret' to evaluate the PM and construct an optimal aggregation scheme for the PM for the square loss proper scoring rule. The resulting regret is shown to be $3 - 4$ better than those obtained by naive approaches.

Natural extension for the problem are to study the case where there are more than 2 experts and to consider the case where experts re not necessarily Blackwell ordered. For the latter one could propose an alternative notion of regret, which compares the PM's forecasts with what would be obtained by a single forecaster exposed to the evidence of both experts (which is the same as that obtained by a Bayesian PM or alternatively the forecast obtained by allowing the experts access to each other's forecasts (we refer the reader to Geanakoplos and Polemarchakis [?] for the relevant analysis). Note that when experts are Blackwell order the latter notion of regret is exactly the same as the one used throughout the paper.

[1] The adjustment was inspired by simulations that demonstrated the reason for the failure of aggregation scheme (7) for close forecasts.

Turning back to our notion of regret, $R(\mathbf{P}, g) = \max\{\mathbb{E}_{\mathbf{P}}[l(x_1)], \mathbb{E}_{\mathbf{P}}[l(x_2)]\} - \mathbb{E}_{\mathbf{P}}[l(g(x_1, x_2))]$, we have an example where the adversary can guarantee a regret that exceeds $\frac{1}{8}(5\sqrt{5} - 11)$. The details of the example see provided in Appendix B.

References

[1] Al-Najjar, N. I. and Weinstein, J., (2008). "Comparative Testing of Experts." *Econometrica*, 76: 541–559.

[2] R.J. Aumann, M. Maschler, (1995). "Repeated Games with Incomplete Information," *MIT Press,* Cambridge.

[3] Blackwell, D., (1953). "Equivalent comparison of experiments," *Annals of Mathematical Statistics,* 24, 265–272.

[4] Brier, G.W., (1950). "Verification of forecasts expressed in terms of probability," *M*onthly weather review 78, 1–3.

[5] Cesa-Bianchi, N. and Lugosi, G., (2006). "Prediction, Learning, and Games." Cambridge University Press.

[6] DeGroot, M. H., (1974), "Reaching a Consensus," *Journal of the American Statistical Association* 69(345): 118–121.

[7] Foster, D. P., and R. V. Vohra, (1998). "Asymptotic Calibration," *Biometrika* 85, 379–390.

[8] Hannan, J. (1957). "Approximation to Bayes risk in repeated play." In M. Dresher, A. W. Tucker, and P. Wolfe (Eds.), Contributions to the Theory of Games, Vol. III, Ann. Math. Stud. 39, pp. 97–139. Princeton University Press.

[9] Hart, S. and Mas-Colell, A., (2013). "Simple Adaptive Strategies - From Regert-Matching to Uncoupled Dynamics." World Scientific Series in Economic Theory - Vol.4

[10] Feinberg, Y. and Stewart, C. (2008). "Testing Multiple Forecasters." *Econometrica*, 76: 561–582.

[11] Sandroni, A., (2000). "Do Markets Favor Agents Able to Make Accurate Predictions?" *Econometrica*, 68(6): 1303–1341.

[12] Sion, M., (1958). "On general minimax theorems," *P*acific Journal of Mathematics, 8, (1): 171–176.

A Proofs

Proof of Lemma 1. Assume for w.l.o.g. that $i = 2$ we shall show first that for any realization x_1, x_2

$$\mathbb{E}_{\mathbf{P}}[l(x_2, \omega) - l(g(x_1, x_2), \omega)|x_1, x_2] = (f(x_1, x_2) - x_2)^2.$$

To see this note that since expert 2 Blackwell dominates expert 1 we have by definition that $\mathbf{P}(\omega = 1|x_1, x_2) = \mathbf{P}(\omega = 1|x_2) = x_2$. Therefore,

$$\begin{aligned}
&\mathbb{E}_{\mathbf{P}}[l(x_2, \omega) - l(f(x_1, x_2), \omega)|x_1, x_2] = \\
&\mathbf{P}(\omega = 1|x_2)[(f(x_1, x_2) - 1)^2 - (x_2 - 1)^2] + \mathbf{P}(\omega = 0|x_2)[(f(x_1, x_2))^2 - (x_2)^2] = \\
&x_2[(f(x_1, x_2) - 1)^2 - (x_1 - 1)^2] + (1 - x_2)[(f(x_1, x_2))^2 - (x_2)^2] = \\
&[x_2(f(x_1, x_2)^2 - 2x_2 f(x_1, x_2) + x_2 - (x_2)^3 + 2(x_2)^2 - x_2] + \\
&[(f(x_1, x_2))^2 - (x_2)^2 - x_2(f(x_1, x_2)^2 + (x_2)^3] = \\
&[(f(x_1, x_2)^2 - 2x_2 f(x_1, x_2) + (x_2)^2] = (f(x_1, x_2) - x_2)^2.
\end{aligned}$$

Hence by the law of total probability

$$\begin{aligned}
&\mathbb{E}_{\mathbf{P}}[l(x_2, \omega) - l(f(x_1, x_2), \omega)] = \mathbb{E}_{\mathbf{P}}\big(\mathbb{E}_{\mathbf{P}}[l(x_2, \omega) - l(f(x_1, x_2), \omega)|x_1, x_2]\big) = \\
&\mathbb{E}_{\mathbf{P}}[(f(x_1, x_2) - x_2)^2]
\end{aligned}$$

□

Proof of Proposition 1. Let g' be the aggregation scheme that is defined as follows:

$$g'(x_1, x_2) = g(x_2, x_1).$$

That is g' switch the roles of the two experts and treats expert 1 as expert 2 and vice versa. It is easy to see that $\sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, g) = \sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, g')$.

Indeed, let $\mathbf{P} \in \Delta(\Omega \times S_1 \times S_2) \in \mathcal{BI}$ be any information structure with expert i as the dominate expert. Let $\mathbf{P}'$ be the probability that is obtained by switching the roles of expert 1 and expert 2. That is $\mathbf{P}'$ is obtained from $\mathbf{P}$ by letting expert 1 observe the signal of expert 2 in $\mathbf{P}$ and letting expert 2 observe the signal of expert 1. It readily follows that

$$R(\mathbf{P}, g) = R(\mathbf{P}, g')$$

for $j = 3 - i$. Therefore,

$$\sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, g) = \sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}', g').$$

Consider the following anonymous aggregation scheme $\tilde{g}$

$$\tilde{g}(x_1, x_2) = \frac{g(x_1, x_2) + g'(x_1, x_2)}{2} = \frac{g(x_1, x_2) + g(x_2, x_1)}{2}.$$

We claim that $\sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, \tilde{g}) \leq \sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, g)$. To see this note that for every $x_i \in [0, 1]$ it follows from convexity that

$$(\tilde{g}(x_1, x_2) - x_i)^2 = (\frac{g(x_1, x_2) + g'(x_1, x_2)}{2} - x_i)^2 \leq$$
$$\frac{1}{2}(g(x_1, x_2) - x_i)^2 + \frac{1}{2}(g'(x_1, x_2) - x_i)^2$$

Therefore, in particular for every $\mathbf{P} \in \mathcal{BI}$

$$R(\mathbf{P}, \tilde{g}) \leq \frac{1}{2} R(\mathbf{P}, g) + \frac{1}{2} R(\mathbf{P}, g').$$

Hence

$$R(\mathbf{P}, \tilde{g}) \leq \max\{R(\mathbf{P}, g), R(\mathbf{P}, g')\}.$$

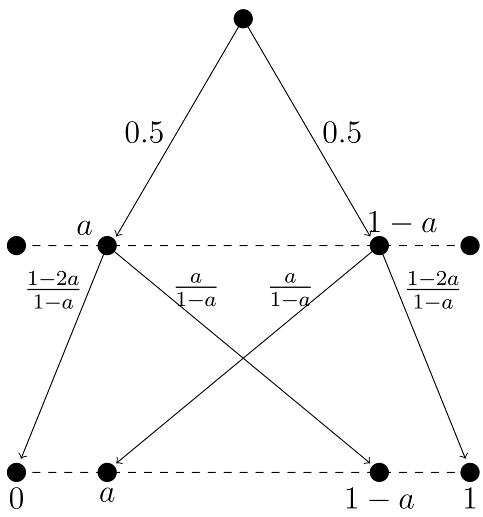
As a result we can deduce that

$$\sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, \tilde{g}) \leq \max\{\sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, g), \sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, g')\} = \sup_{\mathbf{P} \in \mathcal{BI}} R(\mathbf{P}, g).$$

□

Proof of Theorem 1 and Proposition 2. The first step is to show that the adversary can guarantee the value $v = \frac{1}{8}(5\sqrt{5} - 11)$. That is to show that $Val(\Gamma) \geq \frac{1}{8}(5\sqrt{5} - 11)$.

To see this consider the following class of belief martingales:



This is a class of distributions over posterior belief $F_a \in \Delta([0,1]^2)$ depending on a parameter $a \in [0, \frac{1}{2}]$. F_a has the following interpretation. The less informative expert 1 has two possible posterior probabilities $x_1 = a$ and $x_1 = 1 - a$ each of which obtains with probability $\frac{1}{2}$. Conditional on $x_1 = a$ expert 2's posterior is $x_2 = 0$ with probability $\frac{1-2a}{1-a}$ and $x_2 = 1 - a$ with probability $\frac{a}{1-a}$. Similarly, conditional on $x_1 = 1 - a$ expert 2's posterior is $x_2 = a$ with probability $\frac{a}{1-a}$ and $x_2 = 1$ with probability $\frac{1-2a}{1-a}$. Since both $a = 0 \cdot \frac{1-2a}{1-a} + (1-a) \cdot \frac{a}{1-a}$ and $1 - a = a \cdot \frac{a}{1-a} + 1 \cdot \frac{1-2a}{1-a}$ we have that $\mathbb{E}_{F_a}(x_2|x_1) = x_1$.

There exists an information structure $\mathbf{P}_a \in \mathcal{F}$ that induces F_a as a distribution of the posterior beliefs for the experts.

Given a martingale structure F and the corresponding information structure $\mathbf{P} \in \mathcal{BI}$, we ask what is the best anonymous aggregation scheme for the PM given $\mathbf{P}$. The answer is clearly obtained using Bayesian updating. Restricting attention to anonymous forecasts is equivalent to conceal the information of who says what from the PM. Namely, for any forecast vector (x_1, x_2) the agent only knows that one of the experts predicted x_1 and the other predicted x_2. That is, he only knows that the unordered set of forecasts $\{x_1, x_2\}$. Had he known the information structure $\mathbf{P}$ his best forecast is $g_{\mathbf{P}}(x_1, x_2) = \mathbf{P}(\omega = 1|\{x_1, x_2\})$.

We next calculate $g_{\mathbf{P}_a}$ where $\mathbf{P}_a \in \mathcal{BI}$ is the information structure corresponding to F_a. We claim that $g_{\mathbf{P}_a}$ has the following form

$$g_{\mathbf{P}_a}(x_1,x_2) = \begin{cases} \frac{1}{2} & \text{if } \{x_1,x_2\}=\{a,1-a\} \\ 0 & \text{if } x_2 = 0 \\ 1 & \text{if } x_2 = 1 \end{cases}$$

To see this note first that if $x_2 \in \{0,1\}$ which happens with probability $\frac{1-2a}{1-a}$ then the Bayesian PM who knows $\mathbf{P}_a$ infers x_2 and thus, as expert 2 knows the realised state ω, so does the PM. In this case $\mathbf{P}_a(\omega|\{x_2,x_1\}) = x_2$. The only other case is where $\{x_1,x_2\}=\{a,1-a\}$. Since $\mathbf{P}_a((x_1,x_2)=(a,1-a)) = \mathbf{P}_a((x_1,x_2) = (1-a,a)) = \frac{a}{2(1-a)}$ we have that

$$g_{\mathbf{P}_a}(a,1-a) = g_{\mathbf{P}_a}(1-a,a) = \frac{1}{2}.$$

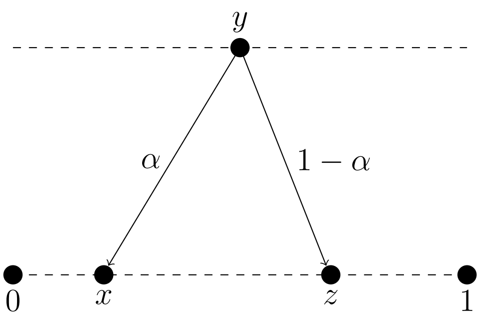
Therefore, by Lemma 1

$$R(\mathbf{P}_a, g_{\mathbf{P}_a}) =$$
$$\mathbb{E}_{\mathbf{P}_a}[(x_2 - g_{\mathbf{P}_a}(x_1,x_2))^2] = 0 \cdot \frac{1-2a}{1-a} + \frac{a}{(1-a)}(\frac{1}{2}-a)^2.$$

Maximizing the regret $\frac{a}{(1-a)}(\frac{1}{2}-a)^2$ across all values $a \in [0,\frac{1}{2}]$ yields a regret of exactly $v = \frac{1}{8}(5\sqrt{5}-11)$ for the information structure $\mathbf{P}_a$ that corresponds to $a = \frac{1}{4}(3-\sqrt{5}) \approx 0.190983$.

Our next goal is to show that $Val(\Gamma) \leq v$. We do it by showing that the above aggregation scheme f guarantees a regret at most v to the PM.

Consider the following martingale distributions $F_{x,y,z} \in \Delta([0,1]^2)$ that depend upon three parameters $x,y,z \in [0,1]$ such that $x \leq y \leq z$.



The prior probability of $\omega = 1$ in the corresponding information structure is y and expert 1 receives no further information. Expert 2's posterior probability of $\omega = 1$

is x with probability α and z with probability $1-\alpha$, where α is chosen such that $\alpha x + (1-\alpha)z = y$. That is,

$$\alpha = \frac{z-y}{z-x}.$$

Hence, by Lemma **??** there exists an information structure $\mathbf{P}_{x,y,z} \in \mathcal{F}$ with $F_{x,y,z}$ as the corresponding distribution over the posterior probabilities of the experts.

Lemma 2. *For any anonymous aggregation scheme $g \in \mathcal{G}$ the following equality holds:*

$$\sup_{\mathbf{P}\in\mathcal{BI}} R(\mathbf{P}, g) = \sup_{(x,y,z)\in[0,1]^3 : x\le y\le z} R(\mathbf{P}_{x,y,z}, g).$$

The lemma states that for the fixed anonymous aggregation scheme g in order to calculate the best response value for the adversary $\sup_{\mathbf{P}\in\mathcal{BI}} R(\mathbf{P}, g)$ one can restrict attention to the class $\mathbf{P}_{x,y,z}$ above.

Proof. Clearly, it is sufficient to show that $\sup_{\mathbf{P}\in\mathcal{BI}} R(\mathbf{P}, g) \le \sup_{x,y,z} R(\mathbf{P}_{x,y,z}, g)$. To see this let $\mathbf{P} \in \mathcal{BI}$ be any distribution. Since g is anonymous we can assume with no lose of generality that expert 2 dominates expert 1. We shall show that there exists $\mathbf{P}_{x,y,z}$ such that

$$R(\mathbf{P}, g) \ge R(\mathbf{P}_{x,y,z}, g).$$

We first let $F_y \in \Delta([0,1]^2)$ be the class of distribution (which contains the class $F_{x,y,z}$), with a corresponding information structure $\mathbf{P}_x$, for which the posterior distribution of expert 1 is fixed and equal y with probability 1 and the distribution of expert 2's posterior probability x_2 has the martingale property, namely, $\mathbb{E}_{F_y}(x_2) = y$. The first step is to show that there exists an appropriate $\mathbf{P}_y$ for some value y such that:

$$R(\mathbf{P}, g) \ge R(\mathbf{P}_y, g).$$

To see this let $Y \subset [0,1]$ the set of posterior distribution for expert 1 that obtained with positive probability under $\mathbf{P}$. We can write the expected regret as follows:

$$R(\mathbf{P}, g) = \sum_{y\in Y} \mathbb{E}_{\mathbf{P}}[l(x_2,\omega) - l(g(x_1,x_2),\omega) | x_1 = y].$$

Therefore there exists a value $y \in Y$ such that $R(\mathbf{P}, g) \geq \mathbb{E}_{\mathbf{P}}[l(x_2, \omega) - l(f(x_1, x_2), \omega)|x_1 = y]$. Hence we can simply take $F_y = \mathbf{P}(x_2|x_1 = y)$.

Consider now the set class of distribution $F_y \in \Delta([0,1]^2)$ for $y \in [0,1]$. A simple inductive argument (over the size of the support) shows that every such distribution can be expressed as a convex combination of distribution of the form $F_{x,y,z}$. Namely, for every F_y there exist positive numbers $\{\alpha_k\}_{k=1}^n$ that sum up to 1 and corresponding distributions $\{F_{x_k,y,z_k}\}_{k=1}^n$ that satisfy

$$F_y = \sum_{k=1}^{n} \alpha_k F_{x_k,y,z_k}.$$

Therefore we can write:

$$\mathbb{E}_{\mathbf{P}}[l(x_2, \omega) - l(f(x_1, x_2), \omega)|x_1 = y] = \sum_{k=1}^{n} \mathbb{E}_{\mathbf{P}_{x_k,y,z_k}}[l(x_2, \omega) - l(f(x_1, x_2), \omega)].$$

Hence there exists $1 \leq m \leq n$ for which

$$\mathbb{E}_{\mathbf{P}}[l(x_2, \omega) - l(f(x_1, x_2), \omega)] \leq \mathbb{E}_{\mathbf{P}_y}[l(x_2, \omega) - l(f(x_1, x_2), \omega)] \leq$$
$$\mathbb{E}_{\mathbf{P}_{x_m,y,z_m}}[l(x_2, \omega) - l(f(x_1, x_2), \omega)].$$

This concludes the proof of the lemma. □

We now explain how to prove the fact that f guarantees a payoff at most v for the PM. By Lemma 2

$$R(\mathbf{P}, f) = \sup_{(x,y,z) \in [0,1]^3 : x \leq y \leq z} R(\mathbf{P}_{x,y,z}, g).$$

Consider the following four compact ranges in $[0,1]^3$.

$$K_1 = \{(x, y, z)|x \leq y \leq z,\ |x - y| \geq 0.4,\ |z - y| \geq 0.4\}$$
$$K_2 = \{(x, y, z)|x \leq y \leq z,\ |x - y| \leq 0.4,\ |z - y| \geq 0.4\}$$
$$K_3 = \{(x, y, z)|x \leq y \leq z,\ |x - y| \geq 0.4,\ |z - y| \leq 0.4\}$$
$$K_4 = \{(x, y, z)|x \leq y \leq z,\ |x - y| \leq 0.4,\ |z - y| \leq 0.4\}$$

We note that every triplet (x, y, z) such that $x \leq y \leq z$ there exists $1 \leq m \leq 4$ such that $(x, y, z) \in K_m$ and the regret of f on each of the K_m is determined by a fixed loss function. For example for (x,y,z) in K_1 we have

$$\begin{aligned}&\mathbb{E}_{\mathbf{P}_{x,y,z}}[l(x_2,\omega)-l(f(x_1,x_2),\omega)] =\\ &\frac{z-y}{z-x}\Big(\frac{\big(\sqrt{x(1-x)}\big)y+\big(\sqrt{y(1-y)}\big)x}{\sqrt{x(1-x)}+\sqrt{y(1-y)}}-x\Big)^2\\ &+\frac{y-x}{z-x}\Big(\frac{(\sqrt{y(1-y)})z+(\sqrt{z(1-z)})y}{\sqrt{y(1-y)}+\sqrt{z(1-z)}}-z\Big)^2\end{aligned}$$

Similarly, on range K_2 we have

$$\begin{aligned}&\mathbb{E}_{\mathbf{P}_{x,y,z}}[l(x_2,\omega)-l(f(x_1,x_2),\omega)] =\\ &\frac{z-y}{z-x}\Big(\frac{(x(1-x))y+(y(1-y))x}{x(1-x)+y(1-y)}-x\Big)^2\\ &+\frac{y-x}{z-x}\Big(\frac{(\sqrt{y(1-y)})z+(\sqrt{z(1-z)})y}{\sqrt{y(1-y)}+\sqrt{z(1-z)}}-z\Big)^2.\end{aligned}$$

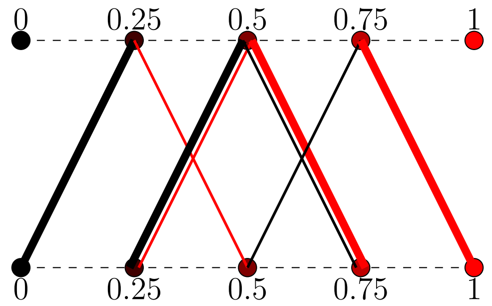
Similar expression obtained for ranges K_4 and K_3.

Hence essentially in order to show that f guarantees a regret at most v to the PM one can restrict attention to solve four simple optimization problem at the four compact domains $\{K_i\}_{i=1,2,3,4}$. One can easily show that in each range the respective function is differentiable with bounded derivative. So essentially all is left to do is to solve four maximization problems with only three variables for each. This is done via standard application of Matlab, which show that the global maximum of $R(x,y,z)$ is obtained at two points $(0,\frac{1}{4}(3-\sqrt{5}),1-\frac{1}{4}(3-\sqrt{5}))$ and $(\frac{1}{4}(3-\sqrt{5}),1-\frac{1}{4}(3-\sqrt{5}),1)$. The global maximum is equal $\frac{1}{8}(5\sqrt{5}-11)$. □

B Counter-example for general domains

.

Consider the following information structure.



There are two states $\{0, 1\}$. Condition on state 0 a black edge is drawn: the two thick black edges are drawn with probability $\frac{3}{8}$ and the two thin black edges are drawn with probability $\frac{1}{8}$. Similarly condition on state 1 a red edge is drawn: the thick red edges are drawn with probability $\frac{3}{8}$ and thin red edges are drawn with probability $\frac{1}{8}$. The prior probability of every state ω is $\frac{1}{2}$. Expert 1 is told the upper node of the chosen edge and expert 2 is told the lower node of the chosen edge. The experts communicate their forecasts for $\omega = 1$ after observing their chosen node. For example, if the first black thick edge joining the upper value 0.25 to the lower value 0 is drawn, then expert 2 observes the lower left hand node of this edge, and hence knows that the probability of $\omega = 1$ is zero. Expert 1, however, observes the upper left hand node and using Bayesian updating he assigns probability 0.25 that the state is $\omega = 1$. Similarly if the red edge, joining the upper value 0.25 to the lower value 0.5, is drawn, then expert 1 assign probability 0.25 that the state is $\omega = 1$ and expert 2 assigns $\omega = 1$ probability $\frac{1}{2}$. The PM, which is also Bayesian in this case, only observes the forecast of the two experts without the identity of who made which forecast. So if he hears the recommendations $\{0.25, 0.5\}$, for example, he knows that the chosen edge is either the thin red edge joining the upper 0.25 to the lower 0.5, the thick black edge joining the upper 0.5 to the lower 0.25, or the thin red edge joining the upper 0.5 to the lower 0.25. Hence his forecast would be 0.4. Using a simple calculation one can show that the expected regret of the PM from expert 2 is $0.025 > v$.

C Contradicting weather forecasts

Figure 4: Forecast in Yahoo! website

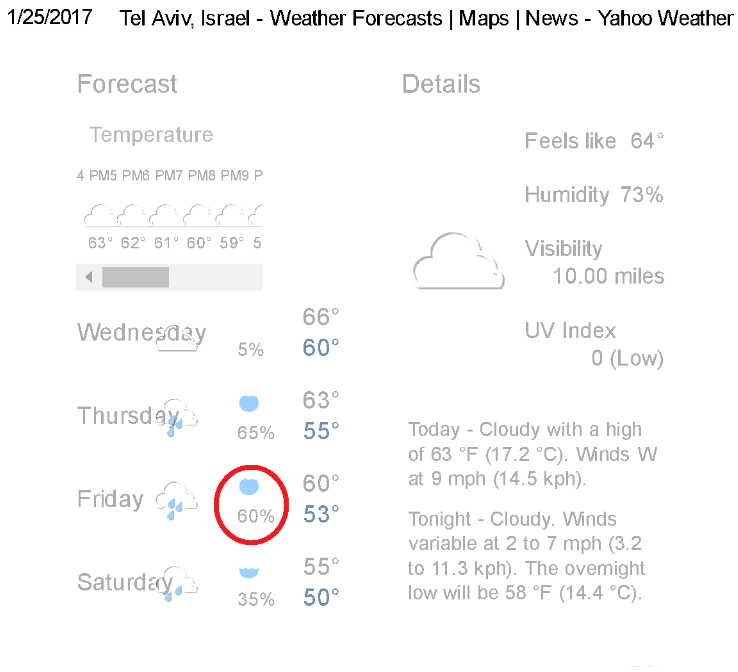


Figure 5: Forecast in Accuweather website

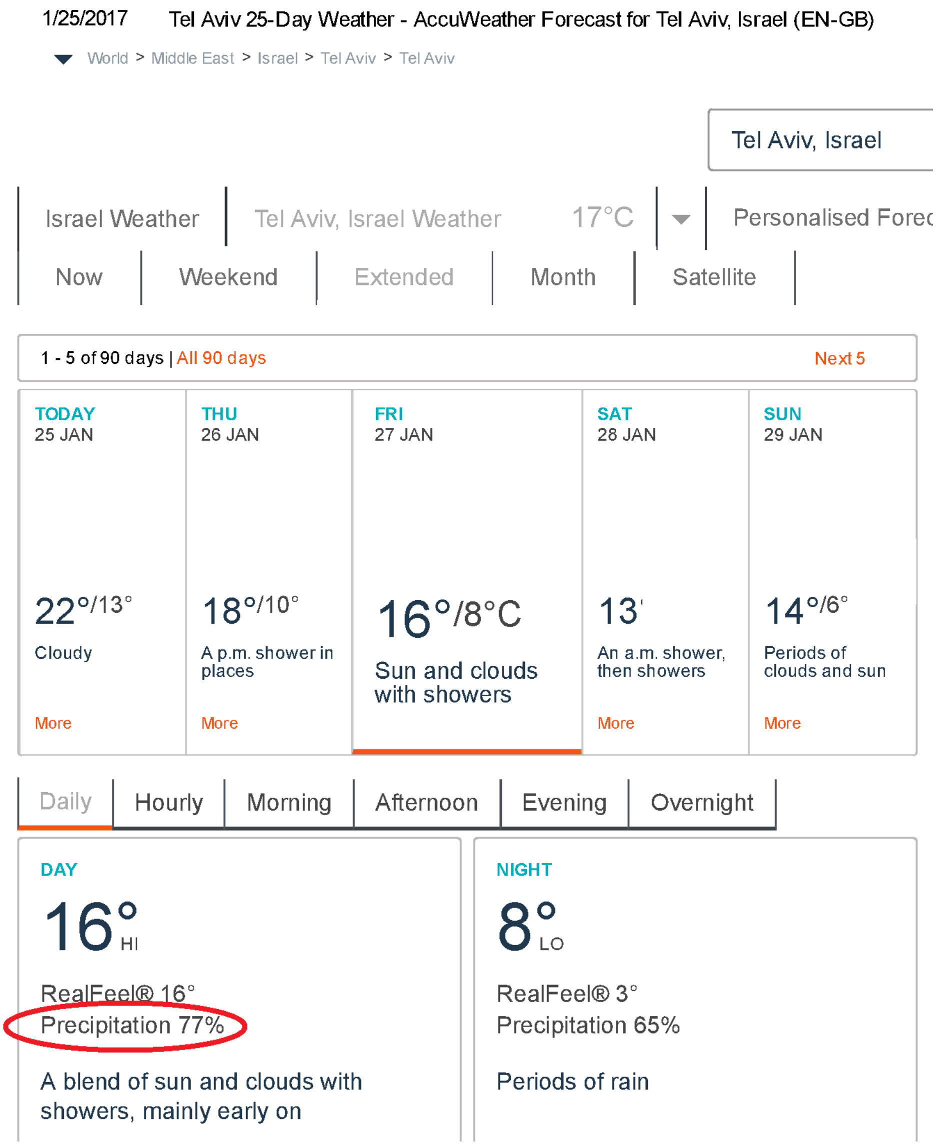
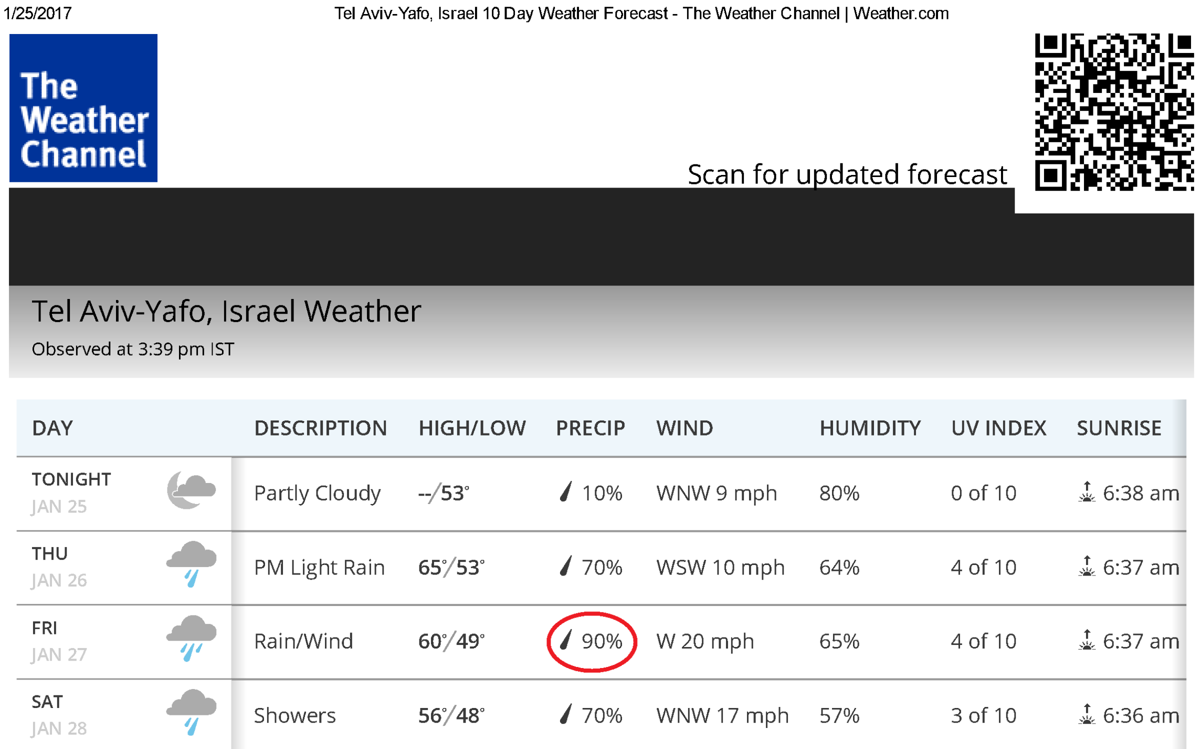


Figure 6: Forecast in Weather-Channel website



1/25/2017 Tel Aviv-Yafo, Israel 10 Day Weather Forecast - The Weather Channel | Weather.com

The Weather Channel

Scan for updated forecast

Tel Aviv-Yafo, Israel Weather

Observed at 3:39 pm IST

DAY	DESCRIPTION	HIGH/LOW	PRECIP	WIND	HUMIDITY	UV INDEX	SUNRISE
TONIGHT JAN 25	Partly Cloudy	--/53°	10%	WNW 9 mph	80%	0 of 10	6:38 am
THU JAN 26	PM Light Rain	65°/53°	70%	WSW 10 mph	64%	4 of 10	6:37 am
FRI JAN 27	Rain/Wind	60°/49°	90%	W 20 mph	65%	4 of 10	6:37 am
SAT JAN 28	Showers	56°/48°	70%	WNW 17 mph	57%	3 of 10	6:36 am