Elicitation for sets of probabilities and distributions

Mark J. Schervish[1] **Teddy Seidenfeld**[2] **Ruobin Gong**[3] **Joseph B. Kadane**[1] **Rafael B. Stern**[4]

[1]Statistics, Carnegie Mellon University, USA
[2]Philosophy and Statistics, Carnegie Mellon University, USA
[3]Statistics, Rutgers University, USA
[4]Statistics, University of Sao Paulo, Brazil

Abstract

We investigate techniques for applying strictly proper scoring rules to elicit arbitrary sets of probabilities for an event and for eliciting sets of countably additive (finite dimensional) joint distributions. We contrast E-admissibility, Maximality, and Γ-Maximin as three IP decision rules for these elicitations. The techniques we investigate apply with sets of probabilities that need not be convex or even connected, and with distributions that may lack moments. We address some challenges to applying these techniques for eliciting merely finitely additive probability distributions.

Keywords. elicitation, proper scoring rules, finite additivity, agglutinated probability

1. Introduction

In de Finetti's [2] seminal presentation on the foundations of subjective probability, he formalizes two coherence criteria: coherence$_1$ and coherence$_2$. These criteria are associated, respectively, with two different decision problems.

Coherence$_1$: governing fair prices for buying and selling bounded random variables.

Coherence$_2$: governing forecasting bounded random variables subject to Brier score, i.e. squared error loss.

In each setting, coherence requires decision making that respects (uniform) dominance with respect to a partition of possible states. Coherence$_1$ requires that the sum of each (finite) set of contracts for buying and selling random variables at a Bookie's fair prices, when those contracts are chosen by an opposing Gambler, are immune to a uniform sure loss to the Bookie. That is, the Gambler cannot form a finite set of contracts at the Bookie's fair prices where the Bookie's net outcome is uniformly dominated in each possible state by the status quo, represented by the constant 0. Coherence$_2$ requires that, for each finite set of forecasts, the sum of their Brier scores is immune to a uniform improvement relative to a rival set of forecasts for the same set of variables, with respect to the same set of possible states.

De Finetti shows that these two criteria are equivalent [2, section 3.3.7] in the sense that a set of fair prices are coherent$_1$ if and only if they serve as a coherent$_2$ set of forecasts. Let P be a finitely additive probability function and let $E_P(\cdot)$ be its corresponding finitely additive expectation function, defined over bounded random variables.[1] De Finetti establishes that the coherent$_1$ and coherent$_2$ sets of fair prices/forecasts are exactly those generated by the class of finitely additive expectation functions.

Nonetheless, de Finetti notes [2, section 3.6.3] the two coherence criteria are not equivalent as operational idealizations for eliciting an agent's subjective probabilities. Specifically, he points out that the first context – setting fair prices for random variables – introduces incentives for strategic play by the Bookie some of which are avoided in the second context – forecasting. That is, though a Bookie should offer only coherent$_1$ fair prices, upon knowing who is the Gambler in the game creates incentives for the Bookie to alter her/his announced fair prices. Then the announced fair prices need not report the Bookie's "honest" personal probabilities.

Example 1.1 (Fair prices for Betting against an expert)**.** Suppose the Bookie is called on to set a fair price, $\Pr(F)$ for buying/selling the indicator function $F(\omega)$ for an event F. That is, in state ω, $F(\omega) = 1$ if $\omega \in F$ and $F(\omega) = 0$ if $\omega \notin F$.

Specifically, the Bookie is required to move first and announce a fair price $\Pr(F)$ for the event F. Then the Gambler plays second and chooses a real-valued quantity β to form a contract that, depending upon which state ω obtains, results in a payoff to the Bookie of

$$\beta(F(\omega) - \Pr(F)).$$

[1]Hereafter, we assume that finitely additive probabilities are defined on a measurable space $\langle \Omega, \mathcal{B} \rangle$; where, Ω is a partition of pairwise disjoint and collectively exhaustive possible *states*, and $\mathcal{B}$ is an algebra of events. Random variables are assumed to be measurable with respect to the measure space $\langle \Omega, \mathcal{B}, P \rangle$.

The game is 0-sum, so that the payoff to the Gambler is

$$-\beta(F(\omega) - \Pr(F)).$$

If the Gambler chooses $\beta > 0$ (respectively, $\beta < 0$), the contract calls for the Bookie to buy from (respectively, to sell to) the Gambler β-many units of the variable F, at the price of $\Pr(F)$ per unit.

Coherence$_1$ requires that $0 \le \Pr(F) \le 1$, with each value in the unit interval a coherent$_1$ fair price. However, suppose the Gambler is an expert about the event F and knows the value of $F(\omega)$. Then if $0 < \Pr(F) < 1$, the Gambler can choose the sign of β, positive or negative, to put the Bookie on the losing side of the contract, where the Bookie's payoff is negative. And by choosing $|\beta|$ as large as is permitted, the Gambler wins (and the Bookie loses) as large a fortune as the rules permit.

In summary, the expert Gambler has no strategy β that results in the outcome 0 (uniformly) dominating the Bookie's outcome, $\beta(F(\omega) - \Pr(F))$, for each possible state ω. Nonetheless, the Gambler is aware of a strategy β that yields a negative outcome for the Bookie in the actual state ω that is realized. Though the Bookie's offer of a fair price, $0 < \Pr(F) < 1$, does not result in a "sure-loss" in the sense of a Book, if played against the expert Gambler it does result in the Bookie losing, and losing for sure!

If the Bookie believes that the Gambler is such an expert then only extreme fair odds, $\Pr(F) \in \{0, 1\}$, offer the Bookie a prospect of not losing. And the Bookie, who plays first in this sequential, 2-person, 0-sum game, maximizes subjective expected utility by choosing the extreme price $\Pr(F) = 1$ (respectively, $\Pr(F) = 0$) if her/his personal probability P satisfies $P(F) > .5$ (respectively, $P(F) < .5$), with indifference between the two extreme prices just in case $P(F) = .5$. In this evidently unfair game, bold play is optimal for the Bookie! Also evident, strategic pricing for the random variable F elicits merely the information whether the Bookie's personal probability satisfies $P(F) \le .5$ or whether $P(F) \ge .5$.

Example 1.2 (An IP version of Example 1.1)**.** Note that a similar phenomenon occurs if, in the IP spirit of Smith's theory of *Pignic Odds* [16], the Bookie is permitted to offer a pair of one-side prices, $\underline{\Pr}(F) \le \overline{\Pr}(F)$, where the lower price is for buying F when the Gambler chooses $\beta > 0$, and the upper price is for selling F when the Gambler chooses $\beta < 0$. Then if the Bookie offers one sided odds with $0 < \underline{\Pr}(F)$ and $\overline{\Pr}(F) < 1$, the Gambler who knows whether or not F obtains, can make the Bookie accept a losing contract. However, better for the Bookie than when obliged to give fair (two-sided) prices, the game with one-sided pricing affords the Bookie a safe strategy: Set $0 = \underline{\Pr}(F)$ and $\overline{\Pr}(F) = 1$. Then, against the expert Gambler, the Bookie is assured of neither a loss nor a gain. But then the IP elicitation is even less informative than in de Finetti's game, as the Bookie's Pignic Odds form the vacuous interval, $[0, 1]$, of all probabilities for the event F.

2. Eliciting a set of probabilities for an event using a strictly proper scoring rule

In response to such operational defects in elicitation of personal probability that arises with his 2-person fair pricing game, de Finetti [2, chapter 3] uses coherence$_2$ with the 1-person forecasting game, where forecasts are evaluated by Brier score loss.[2]

Fix a scoring rule S, used to assess forecasts. Assume S is a loss function.

Let $f(F)$ be a forecast for the event F.

Let P be a (finitely additive) probability.

Denote by $\mathbf{E}_{P,S}(f(F))$ the P-expected S-score for $f(F)$.

A scoring rule S is proper provided that, with respect to a finitely additive probability P, the forecaster maximizes expected S-utility (i.e., minimizes expected S-loss,) by setting $f(F) = P(F)$.

The scoring rule S is strictly proper provided this solution is unique.

The scoring rule S is continuous if for each sequence $\{P_i(F) = p_i\}$ with $\lim_i p_i = p = P(F)$,

$$\lim_i \mathbf{E}_{P_i,S}(p_i) = \mathbf{E}_{P,S}(p).$$

Schervish [9, Theorem 4.2] characterizes all (strictly) proper (continuous) scoring rules for events, as follows.

The forecast $0 \le x \le 1$ for event F is scored using the pair of functions $g_i(x)$ for $i \in \{0, 1\}$, where the subscript 'i' for g_i, is the value of the indicator function $F(\omega)$ for the event F. The forecast x for event F is given score $g_1(x)$ if F obtains, and score $g_0(x)$ if F^c obtains.

Theorem 2.1 (Theorem 4.2 in [9])**.** *Let g_i, $i \in \{0, 1\}$ be a pair of left-continuous functions satisfying*

$$g_i(x) = \lim_{t \to x} g_i(t) \text{ for } x \in \{0, 1\},$$
$$\text{with } g_i(t_i) \text{ finite for some } 0 \le t_i \le 1.$$

The scoring rule S defined by the pair $\{g_1, g_0\}$ is proper if and only if there exists a measure $\lambda(dq)$ on the half-open interval $[0, 1)$ such that, for each $0 \le x < 1$,

$$g_1(x) = g_1(t_1) + \int_{[x,t_1)} (1 - q)\lambda(dq)$$

[2] Coherence$_2$ mitigates some, but not all of the challenges to elicitation that arise from strategic forecasting. See Schervish et al. [11], section 6.3 for an illustration where Brier score invites strategic forecasting for an event F. The agent has a strict preference between forecasts of F based on which of two currencies is used to pay the Brier score. But the same agent has no incentive for using one currency rather than the other when giving fair odds for pricing bets on the event F. This problem is created because the two currencies have different exchange rates in different possible states.

and

$$g_0(x) = g_0(t_0) + \int_{[t_0,x)} q\lambda(dq).$$

The scoring rule S is strictly proper if and only if, in addition, λ gives positive measure to every non-degenerate interval. S is continuous provided that $\lambda(\{x\}) = 0$ for each point $0 \le x \le 1$.

In our applications of this result, without loss of generality, we assume that $t_0 = 0$, $t_1 = 1$, and that $g_0(0) = g_1(1) = 0$.

Next, we prove a Lemma that is central to our investigation of IP-elicitation. Assume that S is a strictly proper scoring rule, which then admits the representation in Theorem 2.1, above. Let P be a probability function with $P(F) = p$. So, p is uniquely the P-optimal forecast $f(F)$ for event F under the score S:

$$\text{argmin}_{f(F)}\{\mathbf{E}_{P,S}(f(F))\} = p.$$

Lemma 2.1. *Consider forecasts for event F, $f(F) \in \{a, b, c, d\}$, where $a < b < p < c < d$.*
Then $\mathbf{E}_{P,S}(p) < \mathbf{E}_{P,S}(b) < \mathbf{E}_{P,S}(a)$,
and $\mathbf{E}_{P,S}(p) < \mathbf{E}_{P,S}(c) < \mathbf{E}_{P,S}(d)$.
In other words, the P-expected utility of the S-score for a forecast $f(F)$ of event F, decreases monotonically as the difference from the P-optimal forecast increases.

Proof. Immediate from the result (A4) in [9]. Equation (A4) asserts that,
for $b < p$, $\mathbf{E}_{P,S}(b) - \mathbf{E}_{P,S}(p) = \int_{[b,p)}(p-q)\lambda(dq) > 0$
and
for $p < c$, $\mathbf{E}_{P,S}(c) - \mathbf{E}_{P,S}(p) = \int_{[p,c)}(q-p)\lambda(dq) > 0$.
Then, for $a < b < p < c < d$,
$\mathbf{E}_{P,S}(a) - \mathbf{E}_{P,S}(b) = \int_{[a,b)}(p-q)\lambda(dq) > 0$, and
$\mathbf{E}_{P,S}(d) - \mathbf{E}_{P,S}(c) = \int_{[c,d)}(q-p)\lambda(dq) > 0$. □

Next, we investigate eliciting a set of imprecise probabilities for an event, F, by identifying the set of admissible forecasts for F subject to a strictly proper scoring rule S. We consider three IP decision rules adapted to arbitrary sets of probabilities[3]

- E-admissibility, see Levi [7]
- Maximality or M-admissibility, see Sen [15] and Walley [18]
- Γ-Maximin. See Gilboa and Schmeidler [3] for an axiomatization in the setting of Anscombe-Aumann horse-lotteries.[4]

[3]Typically, these three decision rules are formulated with respect to a convex set of probability distributions. See, for instance, [6], chapter 8. Here, we relax that assumption about the IP set of probabilities. We review motivations for this generalization in Example 3.1.

[4]In the context of Wald [17] statistical decisions, where the statistician minimizes maximal risk, i.e., Minimax, the theory of Γ-Minimax is developed in papers by, e.g., Hodges Jr and Lehmann [5], Robbins et al. [8], and Good [4].

Definition 2.1.
- $X \in \mathcal{X}$ is E-admissible if, for some $P \in \mathcal{P}$, X maximizes P-expected utility with respect to the option set $\mathcal{X}$.
- $X \in \mathcal{X}$ is M-admissible if, for each $X' \in \mathcal{X}$, there exists $P \in \mathcal{P}$, where the P-expected utility of X is at least as great as the P-expected utility of X'. That is, no alternative X' exceeds X in expected utility with respect to all the elements of the IP set $\mathcal{P}$.
- $X \in \mathcal{X}$ is Γ-Maximin admissible if, for each $X' \in \mathcal{X}$,

  $$\inf_{P\in\mathcal{P}}[P\text{-expectation}(X')] \le \inf_{P\in\mathcal{P}}[P\text{-expectation}(X)].$$

  That is, X maximizes over $\mathcal{X}$ the minimum P-expected utility with respect to $\mathcal{P}$ in $\mathcal{P}$.

When $\mathcal{P}$ is the "vacuous" set of all probability distributions, we refer to the Γ-Maximin solution as Maximin, in accord with the older statistical literature on Minimax theory. In order to assure existence of an "equalizer" Maximin solution in the elicitation problems investigated here, for our results about Γ-Maximin we add the assumption that the scoring rule is continuous.

When $\mathcal{P}$ is a singleton set, i.e., in the "precise" case, each of these three decision rules reduces to maximizing expected utility. Evidently, if X is E-admissible then it is M-admissible. The converse is true in restricted circumstances. See [13] for details.

Let $\mathcal{P}_F$ be an arbitrary set of probabilities for event F.

Let $\mathcal{H}(\mathcal{P}_F)$ be the convex hull of $\mathcal{P}_F$ and $\overline{\mathcal{H}}(\mathcal{P}_F)$ the closed convex hull of $\mathcal{P}_F$.

Fix S as a strictly proper scoring rule.

Let $\mathcal{F}^E_{\mathcal{P}_F,S}(f(F))$ be the set of E-admissible forecasts for event F with respect to the IP set $\mathcal{P}_F$ and scoring rule S. Similarly, let $\mathcal{F}^M_{\mathcal{P}_F,S}(f(F))$ be the set of M-admissible forecasts, and $\mathcal{F}^\Gamma_{\mathcal{P}_F,S}(f(F))$ the set of Γ-Maximin admissible forecasts.

Proposition 2.1. *(i) $\mathcal{F}^E_{\mathcal{P}_F,S}(f(F)) = \mathcal{P}_F$.*

(ii) $\mathcal{F}^M_{\mathcal{P}_F,S}(f(F)) = \overline{\mathcal{H}}(\mathcal{P}_F)$.

(iii) Let S be a continuous scoring rule. Let p^*_S be the Γ-Maximin S-forecast of F with respect to the "vacuous" IP interval $p_F \in [0,1]$, i.e., p^*_S is the Maximin solution.

Assume $0 < p^*_S < 1$; so that $g_i(p^*_S) < \infty$ for $i = 0, 1$. Let $p_{\max} = \max\{\overline{\mathcal{H}}(\mathcal{P}_F)\}$ and $p_{\min} = \min\{\overline{\mathcal{H}}(\mathcal{P}_F)\}$.

(a) $\mathcal{F}^\Gamma_{\mathcal{P}_F,S}(f(F)) = p^*_S$, if $p^*_S \in \mathcal{H}(\mathcal{P}_F)$.

(b) $\mathcal{F}^\Gamma_{\mathcal{P}_F,S}(f(F)) = p_{\max}$, if $p_{\max} \le p^*_S$.

(c) $\mathcal{F}^\Gamma_{\mathcal{P}_F,S}(f(F)) = p_{\min}$, if $p_{\min} \ge p^*_S$.

In summary, for eliciting the IP set $\mathcal{P}_F$ of probabilities for a specific event F:

(i) E-admissibility elicits the target set $\mathcal{P}_F$.

(ii) Maximality elicits the IP set that is the closed convex hull of the target set $\mathcal{P}_F$.

(iii) Assume S is a continuous scoring rule. Γ-Maximin elicits the precise Maximin forecast p^*_S if that belongs to $\overline{\mathcal{H}}(\mathcal{P}_F)$; otherwise, Γ-Maximin elicits the point in $\overline{\mathcal{H}}(\mathcal{P}_F)$ that is nearest to p^*_S.

Proof. We use scores formulated as losses; hence, the Bayesian decision maker with a precise probability seeks to minimize expected score, etc.

Because S is strictly proper, each (countably additive) non-trivial mixed "randomized" strategy forecast σ for event F fails to minimize expected S-loss, regardless the forecaster's precise probability $P(F)$.[5] By Theorem 2 of [10, page 210] then, with respect to the binary partition $\{F, F^c\}$, σ is dominated in S-score by some other (possibly mixed strategy) forecast.[6] So σ is inadmissible for each of the three IP decision rules. Thus, for identifying the E-admissible, M-admissible and Γ-Maximin admissible sets of forecasts for an event F, we may restrict attention to sets of "pure," i.e., non-randomized forecasts of F.

(i) Because the scoring rule S is strictly proper, for each $P \in \mathcal{P}_F$ the forecast $f(F) = P(F)$ uniquely minimizes P-expected loss. Hence the set of E-admissible forecasts for the event F, satisfies $\mathcal{F}^E_{\mathcal{P}_F,S}(f(F)) = \mathcal{P}_F$.

(ii) Assume $P^*(F) \in \mathcal{H}(\mathcal{P}_F)$. We use Lemma 2.1 to show that $P^*(F) \in \mathcal{F}^M_{\mathcal{P}_F,S}(f(F))$. That is, we show that there does not exist a rival forecast $f(F)$ where, for each $P \in \mathcal{P}_F$, $\mathbf{E}_{P,S}(f(F)) < \mathbf{E}_{P,S}(P^*(F))$.

Since, for each IP decision problem, an option is M-admissible if it is E-admissible, by (i) $\mathcal{P}_F \subseteq \mathcal{F}^M_{\mathcal{P}_F,S}(f(F))$. So, suppose that $P^*(F) \in \mathcal{H}(\mathcal{P}_F)$ but $P^*(F) \notin \mathcal{P}_F$. Then, without loss of generality, there exist a pair, $P_1(F), P_2(F) \in \mathcal{P}_F$, $P_1(F) < P_2(F)$, and for some $0 < x < 1$, $P^*(F) = xP_1(F) + (1-x)P_2(F)$. Then $P_1(F) < P^*(F) < P_2(F)$.

We show that there does not exist a forecast $f(F)$ where, for each $P \in \mathcal{P}_F$, $\mathbf{E}_{P,S}(f(F)) < \mathbf{E}_{P,S}(P^*(F))$.

Suppose $P^*(F) < f(F)$. Then, by Lemma 2.1

$$\mathbf{E}_{P_1,S}(f(F)) > \mathbf{E}_{P_1,S}(P^*(F)).$$

Or else suppose $P^*(F) > f(F)$. Then, by Lemma 2.1

$$\mathbf{E}_{P_2,S}(f(F)) > \mathbf{E}_{P_2,S}(P^*(F)).$$

So, $P^*(F)$ is an M-admissible forecast for event F with respect to the IP set $\mathcal{P}_F$.

Next, we show that if $f(F) \notin \overline{\mathcal{H}}(\mathcal{P}_F)$ then $f(F)$ is M-inadmissible. Recall, $p_{\max} = \max\{\overline{\mathcal{H}}(\mathcal{P}_F)\}$ and $p_{\min} = \min\{\overline{\mathcal{H}}(\mathcal{P}_F)\}$. Assume that $f(F) \notin \overline{\mathcal{H}}(\mathcal{P}_F)$. So, either

$$p_{\max} < f(F) \text{ or } f(F) < p_{\min}.$$

By Lemma 2.1, then (respectively), either for every $P \in \mathcal{P}_F$, $\mathbf{E}_{P,S}(p_{\max}) < \mathbf{E}_{P,S}(f(F))$, or for every $P \in \mathcal{P}_F$, $\mathbf{E}_{P,S}(p_{\min}) < \mathbf{E}_{P,S}(f(F))$.

Therefore, $f(F)$ is an M-inadmissible forecast. This establishes that $\mathcal{F}^M_{\mathcal{P}_F,S}(f(F)) \subseteq \overline{\mathcal{H}}(\mathcal{P}_F)$.

To complete (ii), arguing indirectly, we show that each of $p_{\max}$ and $p_{\min}$ is M-admissible.

Suppose $p_{\max}$ is an M-inadmissible forecast. So then $P_{\max}(F) \notin \mathcal{P}_F$ and there exists a rival forecast p^* where, for each $P \in \mathcal{P}_F$, $\mathbf{E}_{P,S}(p^*) < \mathbf{E}_{P,S}(p_{\max})$. Let $\{P_i \in \mathcal{P}_F : i = 1, \dots\}$ be a sequence of probabilities where $\lim_i\{p_i)\} = p_{\max}$ (or $= p_{\min}$, respectively). We write out the details for $p_{\max}$. By hypothesis,

$$\mathbf{E}_{P_i,S}(p^*) = p_i g_1(p^*) + (1-p_i)g_0(p^*) <$$
$$\mathbf{E}_{P_i,S}(p_{\max}) = p_i g_1(p_{\max}) + (1-p_i)g_0(p_{\max}).$$

But, $\lim_{i\to\infty} \mathbf{E}_{P_i,S}(p_{\max}) = \mathbf{E}_{P_{\max},S}(p_{\max})$ and $\lim_{i\to\infty} \mathbf{E}_{P_i,S}(p^*) = \mathbf{E}_{P_{\max},S}(p^*)$ Then, $\mathbf{E}_{P_{\max},S}(p^*) \leq \mathbf{E}_{P_{\max},S}(p_{\max})$ However, since S is strictly proper,

$$\mathbf{E}_{P_{\max},S}(p_{\max}) < \mathbf{E}_{P_{\max},S}(p^*).$$

The same reasoning applies to show that $p_{\min}$ is M-admissible.

(iii) Assume S is a continuous scoring rule. Regarding admissible Γ-Maximin S-forecasts, the "equalizer" strategy, $f(F) = p^*_S$ (where $g_0(p^*_S) = g_1(p^*_S) = c$), is uniquely the Maximin forecast. It is uniquely Maximin because $\mathbf{E}_{P,S}(p^*_S) = c$ for each probability P, and since S is strictly proper, p^*_S uniquely minimizes the P-expected S-loss for the probability $P^*(F) = p^*_S$.

(a) Let $P_1(F), P_2(F) \in \mathcal{P}_F$, with $p_1 < p^*_S < p_2$. Let $x < p^*_S < y$ be two forecasts. Then by Lemma 2.1, $c = \mathbf{E}_{P_1,S}(p^*_S) < \mathbf{E}_{P_1,S}(y)$ and $c = \mathbf{E}_{P_2,S}(p^*_S) < \mathbf{E}_{P_2,S}(x)$. Hence, if $p^*_S \in$ interior $\mathcal{H}(\mathcal{P}_F)$, then for each rival forecast $z \neq p^*_S$, there exists $P \in \mathcal{P}_F$ where $c = \mathbf{E}_{P,S}(p^*_S) < \mathbf{E}_{P,S}(z)$. Hence, $\mathcal{F}^\Gamma_{\mathcal{P}_F,S}(f(F)) = p^*_S$.

For demonstrating cases (iii-b) and (iii-c), we introduce the following Lemma.

[5] Because of agglutinated masses with merely finitely additive probabilities, a f.a. mixed strategy forecast for event F that agglutinates mass at the value $P(F) = p$ also may maximize P-expected utility for a strictly proper scoring rule, in addition to the pure forecast p. See Schervish et al [14], Example 1 for an illustration involving a discontinuous scoring rule.

[6] In terms of the representation of scoring rule S given in Theorem 2.1, say that forecast x dominates forecast y provided that $g_i(x) < g_i(y)$ for $i = 0, 1$. When S is a continuous, strictly proper scoring rule, then if forecast y is dominated, then it is dominated by some coherent$_2$ forecast. However, with a discontinuous, strictly proper scoring rule, though an incoherent$_2$ forecast is dominated, it may be dominated only by other incoherent$_2$ (countably additive) forecasts. See Schervish et al. [10] Examples 3 and 8.

Lemma 2.2. *Let $p_1 < p_2 \le p_S^* \le p_3 < p_4$. Then $\mathbf{E}_{P_1,S}(p_2) < \mathbf{E}_{P_2,S}(p_1)$ and $\mathbf{E}_{P_4,S}(p_3) < \mathbf{E}_{P_3,S}(p_4)$.*

Proof. See the appendix. □

(b) Assume $p_{\max} \le p_S^*$.
For ease of notation let $P(F) = p$.
Then by Lemma 2.2, for each $P < P_{\max}$,

$$\mathbf{E}_{P,S}(p_{\max}) < \mathbf{E}_{P_{\max},S}(p).$$

Therefore, $\mathcal{F}^{\Gamma}_{\mathcal{P}_F,S}(f(F)) = p_{\max}$.

(c) Follows the same reasoning as for (iii-b).

□

Remark. We emphasize that the results for E-admissibility and M-admissibility elicitation of an IP set of probabilities, $\mathcal{P}_F$, for an event F, are invariant over the strictly proper scoring rule, S, that is used in the IP elicitation. However, the Γ-Maximin IP elicitation does depend upon S through the S-Minimax solution p_S^*. Next, we illustrate this in Example 2.1.

Example 2.1 (Two strictly proper scoring rules with different equalizer forecasts)**.** Let S_1 be the (strictly proper) Brier score for forecasting event F. That is, $g_1(f(F)) = (1 - f(F))^2$ and $g_0(f(F)) = f(F)^2$. Then $f(F) = \frac{1}{2}$ is the Brier score equalizer forecast, where $g_1(f(F)) = g_0(f(F))$, with constant score $\frac{1}{4}$.

For contrast, define the strictly proper score S_2 where:

$$g_1(f(F)) = 1-3f(F)^2+2f(F)^3 \text{ and } g_0(f(F)) = 2f(F)^3.$$

Then $f(F) = \sqrt{\frac{1}{3}} \approx 0.577$ is the S_2 equalizer forecast, where $g_1(f(F)) = g_0(f(F))$, with constant score $2(1/3)^{3/2} \approx 0.385$.

3. Eliciting a set of c.d.f.'s

In this section, we review basic results for eliciting an IP set of joint probability distributions for a finite dimensional random variable using a technique for combining countably many strictly proper forecasts. Because of the complications for our technique in the presence of agglutinated probability masses that may arise with merely finitely additive probabilities, first we investigate results when each distribution in the set is countably additive – subsection 3.1. In subsection 3.2, we address agglutinations and other challenges associated with eliciting a merely finitely additive probability distribution.

3.1. A scoring rule for a set of countably additive joint distributions. We begin by adapting the central ideas in Schervish et al. [12] for combining strictly proper scoring rules for a denumerable set of events to form a strictly proper scoring rule for a finite dimensional, countably additive joint distribution. Let $\mathcal{B} = \{B_n\}, n = 1, \dots$, be a sequence of events and let $\mathbf{S} = \{S_n\}, n = 1, 2, \dots$, be a sequence of bounded, strictly proper scoring rules, where the scoring rule S_k scores forecasts $0 \le p_k \le 1$ for the indicator B_k of the event B_k. That is, for each n, there is a finite b_n such that $S_n : \{0,1\} \times [0,1] \to [0, b_n)$. Since multiplying a strictly proper scoring rule by a positive constant gives another strictly proper scoring rule, without loss of generality assume $b_n = 2^{-n}$.

Lemma 3.1.

$$\mathbf{S}(\mathcal{B}, \{p_n\}, n = 1, \dots) = \sum_{n=1}^{\infty} S_n(B_n, p_n) < \infty \qquad (3.1)$$

is a strictly proper scoring rule for eliciting the set of coherent precise countably additive probabilities $\mathbf{p} = \{P(B_n) : n = 1, \dots\}$.

That is, consider the decision problem where a forecaster is subject to the loss (3.1) for the sequence of forecasts $\mathbf{p} = \{p_n\}$ for the sequence of events $\{B_n\}$. Let $\mathbf{E}_P[\sum_{n=1}^{\infty} S_n(B_n, p_n)]$ be the forecaster's expected loss of the sequence of forecasts $\mathbf{p} = \{p_n\}$, calculated with respect to her/his joint, countably additive probability $\mathbf{P}$ for the events $\{B_n\}$. Lemma 3.1 asserts that straightforward forecasting, setting $p_n = \mathbf{P}(B_n)$, uniquely minimizes the forecaster's expected loss in (3.1).

Proof. Since each $S_n > 0$, and the joint distribution P is countably additive, the monotone convergence theorem implies

$$\mathbf{E}_P\left[\sum_{n=1}^{\infty} S_n(B_n, p_n)\right] = \sum_{n=1}^{\infty} \mathbf{E}_P[S_n(B_n, p_n)]. \qquad (3.2)$$

The forecaster minimizes the r.h.s. of (3.2) by minimizing each term separately. Since each scoring rule is strictly proper, this minimization occurs (uniquely) for the sequence of forecasts $\{p_n\}$ where, for each B_n, $\mathbf{P}(B_n) = p_n$. □

Lemma 3.1 helps us create a strictly proper scoring rule to elicit a joint distribution for a finite set of random variables, as follows.[7]

Let $\mathbf{X} = \{X_1, \dots, X_m\}$ be a finite set of random variables and let $\mathbf{P}$ be a countably additive joint probability over $\mathbf{X}$. Then $\mathbf{P}$ is determined by its joint *c.d.f.*

$$F_{\mathbf{X}}(x_1, \dots, x_m) = \mathbf{P}(\{X_j \le x_j\} : j = 1, \dots, m),$$

where $(x_1, \dots, x_m)$ ranges over all elements of $\mathbb{R}^m$.

Axiom of Continuity: Let A be an event and $\{A_n : n = 1, \dots\}$ a sequence of (downward) nested events,

[7]The technique we use here extends to creating a strictly proper scoring rule for a countably additive "nonparametric" joint distribution over denumerably many random variables, provided that joint distribution is determined by the class of all finite dimensional joint distributions.

$A_{n+1} \subseteq A_n$, with $A = \cap_n A_n$. Then $\lim_{n\to\infty} P(A_n) = P(A)$.

By Kolmogorov's *Axiom of Continuity*, equivalent to countable additivity, the *c.d.f.* $F_{\mathbf{X}}$ is determined by its values at rational coordinates of $(x_1, \dots, x_m)$ – sequences in $\mathbb{Q}^m$. There are denumerably many such sequences in the measurable space for the c.d.f. of $\mathbf{X}$. Then, using the technique from Lemma 3.1 for creating a strictly proper scoring rule for eliciting the joint probability for countably many events, we construct a strictly proper scoring rule for eliciting $F_{\mathbf{X}}$. Here are some details.

Let $\mathbf{y} = \{y_n = (y_{n,1}, \dots, y_{n,m}) : n = 1, \dots\}$ be a sequence of elements from $\mathbb{Q}^m$. Define the denumerably many events

$$B_n = \{X_j \leq y_{n,j} : j = 1, \dots, m\} \text{ for } n = 1, \dots .$$

where, for each $j = 1, \dots, m$, the sequence $\{y_{n,j}\}$ converges downward to the value x_j. Then the strictly proper scoring rule (3.1) applied to the sequence of events $\{B_n\}$ elicits the value of the cumulative distribution $F_{\mathbf{X}}(x_1, \dots, x_m)$. So, eliciting the countable set of values, $F_{\mathbf{X}}(x_1, \dots, x_m)$ at the countable set $\{(x_1, \dots, x_m)\}$ of sequences in $\mathbb{Q}^m$, identifies the joint distribution $\mathbf{P}(\mathbf{X})$.

Next, let $\mathcal{P}_{\mathbf{X}}$ be a set of joint, countably additive probabilities over a finite set $\mathbf{X} = \{X_1, \dots, X_m\}$ of random variables. Let $\mathbf{S}$ be a scoring rule of the form in (3.1) for eliciting the set $\mathcal{P}_{\mathbf{X}}$ through the set of admissible forecasts for the set $F_{\mathbf{X}}$.

Paralleling the notation from Section 2 we define:

$\mathcal{F}^E_{\mathcal{P}_{\mathbf{X}},\mathbf{S}}(\mathbf{p})$ to be the set of E-admissible forecasts for the joint distribution $\mathbf{P}$ with respect to the set $\mathcal{P}_{\mathbf{X}}$ and scoring rule $\mathbf{S}$.

Similarly, let $\mathcal{F}^M_{\mathcal{P}_{\mathbf{X}},\mathbf{S}}(\mathbf{p})$ be the set of M-admissible forecasts, and let $\mathcal{F}^\Gamma_{\mathcal{P}_{\mathbf{X}},\mathbf{S}}(\mathbf{p})$ the set of Γ-Maximin admissible forecasts.

As in Proposition 2.1, we require a continuous, strictly proper scoring rule for our analysis of Γ-Maximin forecasting. In order to assure that the strictly proper scoring rule of the form (3.1) is continuous, we introduce the following conditions, namely that the summands in (3.1) all be the same continuous scoring rule multiplied by summable constants.

Using the representation from Theorem 2.1, let $s = [(g_0, g_1), \lambda]$ be a bounded (by 1) continuous scoring rule for the probability of an event. For $n = 1, \dots$, let $s_n = b_n s$, where each $b_n > 0$ and $\sum_n b_n = 1$. We show next that $\mathbf{S} = \sum_n s_n$ is a continuous, strictly proper scoring rule with respect to the sequence of probability forecasts in the following metric topologies:

For the argument space, let $x = \{x_n\}$ and $y = \{y_n\}$ be two possible sequences of elicited probabilities for orthants.

Let $d_a(x, y) = \sup_n |x_n - y_n|$. For the triangle inequality, note that

$$d_a(x, z) = \sup_n |x_n - z_n| = \sup_n |x_n - y_n + y_n - z_n| \leq$$
$$\sup_n |x_n - y_n| + \sup_n |y_n - z_n| = d_a(x, y) + d_a(y, z).$$

For the scoring rule space, the total score is a real number, so we use the absolute value of the difference as a metric.

That is, for each $\epsilon > 0$, let $\delta_\epsilon > 0$ be such that whenever $|t-h| < \delta_\epsilon$ with $t, h \in [0,1]$ then $|g_j(t) - g_j(h)| < \epsilon$ for $j = 0, 1$. If $d_a(x, y) < \delta_\epsilon$, then

$$\Big|\sum_n s_n(B_n, x_n) - \sum_n s_n(B_n, y_n)\Big| \leq$$
$$\sum_n |s_n(B_n, x_n) - s_n(B_n, y_n)| =$$
$$\sum_n b_n |s(B_n, x_n) - s(B_n, y_n)| \leq \epsilon.$$

Thus, $\mathbf{S}$ is a continuous scoring rule of the form (3.1).

Proposition 3.1. *(i)* $\mathcal{F}^E_{\mathcal{P}_{\mathbf{X}},\mathbf{S}}(\mathbf{p}) = \mathcal{P}_{\mathbf{X}}$.

(ii) $\mathcal{F}^M_{\mathcal{P}_{\mathbf{X}},\mathbf{S}}(\mathbf{p}) = \overline{\mathcal{H}}(\mathcal{P}_{\mathbf{X}})$.

(iii) Let $\mathbf{P}^*_S$ *be the Minimax forecast of the joint distribution* $\mathbf{P}(\mathbf{X})$ *using a continuous* $\mathbf{S}$*-score in (3.2).*

(a) $\mathcal{F}^\Gamma_{\mathcal{P}_{\mathbf{X}},\mathbf{S}}(\mathbf{p}) = \mathbf{P}^*_S$*, if* $\mathbf{P}^*_S \in \mathcal{H}(\mathcal{P}_F)$*, otherwise;*

(b) $\mathcal{F}^\Gamma_{\mathcal{P}_F,\mathbf{S}}(\mathbf{p}) = \mathbf{P}^\dagger_S$ *where* $\mathbf{P}^\dagger_S = argmin_{P \in \overline{\mathcal{H}}(\mathcal{P}_{\mathbf{X}})} \mathbf{E}_{\mathbf{P}^*_S}[\mathbf{S}(P)]$.

Proof. Parts (i) and (ii) follow using the identical reasoning from Proposition 2.1, applied to the elicitation of the set of joint distributions, $\mathcal{P}_{\mathbf{X}}$, separately at each event B_n in the r.h.s. of (3.2). Part (iii-a) follows by the same reasoning as used in part (iii-a) of Proposition 2.1. Part (iii-b) follows by applying Lemma 2.2 to identify that point $\mathbf{P}^\dagger_S$ in $\overline{\mathcal{H}}(\mathcal{P}_{\mathbf{X}})$ "closest" in $\mathbf{S}$-score to $\mathbf{P}^*_S$. Evidently, $\mathbf{P}^\dagger_S$ is a point on the boundary of $\overline{\mathcal{H}}(\mathcal{P}_{\mathbf{X}})$. □

Parallel with the situation in Proposition 2.1 (iii) we extend Example 2.1 to illustrate the sensitivity of the Maximin solution in Proposition 3.1 (iii) to the selection of the continuous strictly proper scoring rule $\mathbf{S}$.

Example 3.1 (Example 2.1, continued). Let $\Omega = \{\omega_1, \dots, \omega_n\}$ and consider the random variable $X(\omega_k) = k$. Let $B_k = \{\omega_1, \dots, \omega_k\}$ for $k = 1, \dots, n$. So the values $\{P(B_k) = P(X \leq k) : k = 1, \dots, n\}$ identify the *cdf* for the distribution $P(X)$. Let $y_k = f(B_k)$ be a forecast for the event B_k. Then, in Lemma 3.1, with $\mathbf{S} = \mathbf{S}_1 =$ Brier score, the Maximin forecasts satisfy $y_k = \frac{1}{2}$ for $k = 1, \dots, n-1$, and (of course) $y_n = 1$. This *cdf* corresponds to an extreme probability density function on Ω: $P(\{\omega_1\}) = P(\{\omega_n\}) = \frac{1}{2}$ and $P(\{\omega_k\}) = 0$ for $1 < k < n$.

By similar reasoning, with $\mathbf{S} = \mathbf{S}_2$ in Lemma 3.1, the Maximin forecasts satisfy $y_k = \sqrt{\frac{1}{3}}$ for $k = 1, \ldots, n-1$, and $y_n = 1$. This *cdf* corresponds to a different extreme probability density function on Ω: $P(\{\omega_1\}) = \sqrt{\frac{1}{3}}$, $P(\{\omega_n\}) = 1 - \sqrt{\frac{1}{3}}$ and $P(\{\omega_k\}) = 0$ for $1 < k < n$.

Evidently, for each continuous, strictly proper scoring rule $\mathbf{S}$, the Maximin solution $\mathbf{P}^*_S$ for the *cdf* $P(X)$ selects a distribution that is a point on the boundary of the simplex of distributions over Ω, with support the two states $\{\omega_1, \omega_n\}$. This is in sharp contrast with the Maximin solution for forecasting the probability density function for the n-many atoms of the space, $\{z_k = f(\{\omega_k\}) : k = 1, \ldots, n\}$. For example, using Brier score, the Maximin forecasts are the uniform distribution, $z_k = 1/n$, the center of mass of the simplex.

Remark. Note that the distributions elicited through Proposition 3.1 need not admit moments.

Next, we give an example that motivates the generality of the IP sets permitted in Proposition 3.1, where $\mathcal{P}_\mathbf{X}$ may fail to be convex.

Example 3.2. Let $\mathbf{X} = \{X_1, \ldots, X_m\}$ be quantities of interest. Consider an IP set $\mathcal{P}_\mathbf{X}$ of k-many joint distributions $\{P_1(\mathbf{X}), \ldots, P_k(\mathbf{X})\}$ used to represent a set of k-many expert opinions about these quantities. Suppose that, included in $\mathbf{X}$ are the indicator functions for two events, A and B.

It may be that the experts are unanimous in their judgements that A and B are probabilistically independent, for instance, when they agree that these two are causally independent events. Then, in general, the set $\overline{\mathcal{H}}(\mathcal{P}_\mathbf{X})$, whose finitely many extreme points are a subset of $\mathcal{P}_\mathbf{X}$, includes probability distributions where A and B are not independent.

For instance, suppose that $P_1(A, B)$ has A and B independent events with $P_1(A) = 0.4$ and $P_1(B) = 0.2$. so that $P_1(AB) = 0.08$, etc. Similarly, suppose that $P_2(A, B)$ has A and B independent events with $P_2(A) = 0.6$ and $P_2(B) = 0.8$. so that $P_2(AB) = 0.48$, etc. Then $P_3(A, B) = 0.5P_1(A, B) + 0.5P_2(A, B)$ belongs to $\overline{\mathcal{H}}(\mathcal{P}_\mathbf{X})$. But $0.5 = P_3(A) < P_3(A|B) = \frac{0.28}{0.5} = 0.56$. Hence, with P_3, B is positively relevant to A, contrary to the probability judgements, P_1 and P_2, of the two experts.

Thus, we find it is important to consider IP sets that are not convex. And for the reasons illustrated in Example 2.1, we are not satisfied using M-admissibility as the decision rule for eliciting such IP sets with the techniques developed here, based on strictly proper scoring rules.

3.2. Eliciting a merely finitely additive probability distribution. In [1, section 7.3], de Finetti notes the possibility of agglutinated masses with merely finitely additive probabilities. That is, in violation of the Axiom of Continuity for countably additive probabilities, it may be that $\{A_n : n = 1, \ldots\}$ is a sequence of (downward) nested events, $A_n \supseteq A_{n+1}$, with $A = \cap_n A_n \neq \emptyset$, but where for each A_n, $P(A_n) = a > b = P(A)$. Then probability mass $c = (a - b) > 0$ is agglutinated to the nonempty set A by the sequence $\{A_n\}$. Here is an illustration of the challenge posed by agglutinations as it affects the technique used in section 3.1.

Example 3.3. Let X be a bounded but non-simple random variable, e.g. with values $0 \leq x < 1$. Let $P([0, x]) = x$ for each $0 \leq x < 0.5$. Suppose $P[0, 0.5] = 0.6$. Then the *cdf* F_X for $P(X)$ fails to distinguish between the countably additive case where there is a discrete point mass, $P(\{x = 0.5\}) = 0.1$, and infinitely many different merely finitely additive probability distributions where a mass $0 < c \leq 0.1$ is agglutinated (by a sequence of intervals converging from the left) to the point $x = 0.5$, and the remaining mass is assigned to the point $x = 0.5$, i.e., where

$$P(\{x = 0.5\}) = 0.1 - c \text{ and}$$
$$P((x, 0.5]) = 0.6 - x, \text{ for each } 0 \leq x < 0.5.$$

Thus, the successful elicitation of the *cdf* F_X by the method used with Lemma 3.1 is inadequate for identifying the possibly finitely additive distribution $P(X)$ without checking each point of discontinuity in the *cdf* for possible agglutinations (from the left).

One response to this challenge is to introduce a second round of forecasting, designed specifically for eliciting agglutinated masses associated with a finitely additive *cdf* that is only approximated by the countably additive *cdf* $F(X)$ identified in the first round of elicitation based on the method of Lemma 3.1. That is, suppose that the procedure for eliciting a precise countably additive *cdf* $F(X)$ using the method of Lemma 3.1 yields (at most) denumerably many discontinuities at the real values x_n ($n = 1, \ldots$), where these x_n need not be rational values. Let this first round of elicitation yields the point masses $P(X = x_n) = p_n > 0$, for $n = 1, \ldots$. Then, in order to identify whether the forecaster's probability distribution is merely finitely additive because some of this point mass p_n is agglutinated (from below) at $X = x_n$, in a second round of elicitation, use the strictly proper scoring rule S to elicit probabilities q_n for events $C_n = \{X < x_n : n = 1, \ldots\}$. Note that the events C_n differ from the events B_n used in the first round elicitation of $F(X)$ with Lemma 3.1 in two ways:

1. The cutoff points, x_n, need not be rational values.
2. The C_n are defined by strict inequalities.

Let $r_n = F(X \leq x_n) - q_n$. Then $p_n - r_n$ is the finitely additive probability agglutinated (from below) at $(X = x_n)$.[8] This is a 2-round procedure since identifying the

[8] For assessing agglutination from above, apply these methods to the cdf of $Y = -X$.

locations of the point masses, x_n, from $F(X)$, may require first completing an infinite set of elicitations using the events B_n.

In addition to, and separate from this challenge to elicitation posed by agglutination, also there is also the more familiar issue that, since P may be merely finitely additive, the probabilities for an infinite sequence of disjoint sets provides only a lower bound on their union. So, the probabilities provided by a *cdf* for an infinite number of disjoint intervals for the random variable X, in general, is insufficient for determining the probability that X takes its value in their union, independent of the matter of agglutinated probability masses at points within these intervals.

4. Summary, and some open questions

In this paper we have investigated basic results for using strictly proper scoring rules to elicit sets of probabilities and sets of joint distributions in an IP setting. We consider admissible sets of forecasts for eliciting a set of probabilities of a given event. Using a strictly proper scoring rule for forecasting a sequence of events, we consider admissible sets of forecasts for finite dimensional, countably additive joint probabilities, as identified by their cumulative probability distribution functions. We present our findings for three common IP decision rules: E-admissibility, Maximality, and Γ-Maximin. In general, E-admissibility permits identifying an exact target set of probabilities for an event, or target set of countably additive joint distributions, regardless the structure of those IP sets. Maximality recovers the closed convex hull of the target set. And Γ-Maximin generates a precise probability or a precise distribution from the target set that is (or is "closest to") the Maximin solution to the forecasting problem. Since the Maximin solution depends upon which strictly proper scoring rule is used to assess forecasts, only the first two methods give elicitations that are invariant over which strictly proper scoring rule is applied.

In our work on eliciting IP sets of joint (finite dimensional) countably additive distributions, since our method identifies the *cdf* for a countably additive joint distribution, the distributions so elicited need not have moments. But, for the same reason, our methods face the challenge of agglutinated masses when the distributions are merely finitely additive. We present a 2-stage elicitation procedure for addressing finitely additive agglutinations.

In this paper we have not discussed applying our methods for eliciting sets of conditional probabilities, or sets of conditional probability distributions. We are interested in those questions. The following example is offered in the hope of provoking the reader's interest too.

Example 4.1 ((Example 3.1, continued) Γ-maximin for eliciting a *cdf*, and a conditional *cdf*)**.** Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and consider the random variable $X(\omega_k) = k$ for $k = 1, 2, 3$. Let $B_k = \{\omega_1, \dots, \omega_k\}$ for $k = 1, 2, 3$. So the values $\{P(B_k) = P(X \le k) : k = 1, 2, 3\}$ identify the *cdf* for the distribution $P(X)$. Let $y_k = f(B_k)$ be a forecast for the event B_k. Use Brier score to assess each of the three forecasts and use the sum of the three scores as the strictly proper scoring rule. Since, $B_3 = \Omega$, for each coherent forecast, $y_3 = 1$, when $(B_3 - y_3)^2 = 0$, and the sum of the three Brier scores is equal to $(B_1 - y_1)^2 + (B_2 - y_2)^2 + 0$. As noted in Example 3.1, the Maximin (equalizer) forecasts are $y_1 = y_2 = \frac{1}{2}$, with total score $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ regardless which of the three states obtain. However, this forecast corresponds to an extreme probability distribution $P(\omega_1) = P(\omega_3) = \frac{1}{2}$ and $P(\omega_2) = 0$. By Maximin standards, the rival forecasts $z_1 = 1/3$, $z_2 = 2/3$, and $z_3 = 1$ (corresponding to the uniform distribution $Q(\{\omega_1\}) = Q(\{\omega_2\}) = Q(\{\omega_3\}) = 1/3$) is inferior because, if ω_3 obtains, the rival z-forecasts have a worse Brier score than the y-forecasts: $1/9 + 4/9 = 5/9 > \frac{1}{2}$. Now consider the situation where the forecaster learns that state ω_1 does not obtain. Then the conditional y-forecasts, y^*, given $\{\omega_2, \omega_3\}$, are $y_1^* = y_2^* = 0$, and $y_3^* = 1$, corresponding to $P(\{\omega_3\}|\{\omega_2, \omega_3\}) = 1$, whereas the rival conditional z-forecasts, z^*, given $\{\omega_2, \omega_3\}$, are $z_1^* = 0$, $z_2^* = 1/2$, and $z_3^* = 1$, corresponding to $Q(\{\omega_3\}|\{\omega_2, \omega_3\}) = \frac{1}{2}$. Here the conditional z-forecasts are Maximin for the conditional elicitation problem, not the conditional y-forecasts!

A. Proof of Lemma 2.2

Lemma 2.2 regulates admissible Γ-Maximin S-forecasts for an event F, by identifying the greater of $\mathbf{E}_{P_1,S}(p_2)$ and $\mathbf{E}_{P_2,S}(p_1)$ whenever p_1 and p_2 lie on the same side of the Minimax forecast p_S^*.

Let p_S^* be the "equalizer" Minimax strategy, where $g_0(p_S^*) = g_1(p_S^*) = c$.

Proof of Lemma 2.2. Since S is fixed in this proof, we simplify the notation. That is, we demonstrate $\mathbf{E}_{P_1}(p_2) < \mathbf{E}_{P_2}(p_1)$. The proof that $\mathbf{E}_{P_4}(p_3) < \mathbf{E}_{P_3}(p_4)$ follows similarly.

By the representation of expected scores given via Thrm. 4.2, we have:

$$\mathbf{E}_{P_1}(p_2) = \int_{p_2}^{1} p_1(1-q)\lambda(dq) + \int_{0}^{p_2} (1-p_1)q\lambda(dq).$$

$$\mathbf{E}_{P_2}(p_1) = \int_{p_1}^{1} p_2(1-q)\lambda(dq) + \int_{0}^{p_1} (1-p_2)q\lambda(dq).$$

By assumption that p^* is the equalizer Minimax forecast:

$$c = \int_{p^*}^{1} (1-q)\lambda(dq) = \int_{0}^{p^*} q\lambda(dq).$$

Write,

$$\int_{p_2}^{1}(1-q)\lambda(dq) = c + \int_{p_2}^{p^*}(1-q)\lambda(dq),$$

and

$$\int_{0}^{p_1} q\lambda(dq) = c - \int_{p_1}^{p^*} q\lambda(dq).$$

Then, $\mathbf{E}_{P_2}(p_1) - \mathbf{E}_{P_1}(p_2)$

$$\begin{aligned}
&= \int_0^{p_1} [(1-p_2)q - (1-p_1)q]\,\lambda(dq) \\
&\quad + \int_{p_1}^{p_2} [p_2(1-q) - (1-p_1)q]\,\lambda(dq) \\
&\quad + \int_{p_2}^{1} [p_2(1-q) - p_1(1-q)]\,\lambda(dq) \\
&= -(p_2-p_1)\int_0^{p_1} q\lambda(dq) \\
&\quad + \int_{p_1}^{p_2} [p_2(1-q) - (1-p_1)q]\,\lambda(dq) \\
&\quad + (p_2-p_1)\int_{p_2}^{1}(1-q)\lambda(dq) \\
&= -(p_2-p_1)\left[c - \int_{p_1}^{p^*} q\lambda(dq)\right] + p_2\lambda((p_1,p_2]) \\
&\quad + (p_1-p_2-1)\int_{p_1}^{p_2} q\lambda(dq) \\
&\quad + (p_2-p_1)\left[c + \int_{p_2}^{p^*}(1-q)\lambda(dq)\right] \\
&= (p_2-p_1)\left[\int_{p_1}^{p^*} q\lambda(dq) + \int_{p_2}^{p^*}(1-q)\lambda(dq)\right] \\
&\quad + p_2\lambda((p_1,p_2]) + (p_1-p_2-1)\int_{p_1}^{p_2} q\lambda(dq) \\
&= (p_2-p_1)\left[\lambda((p_2,p^*]) + \int_{p_1}^{p_2} q\lambda(dq)\right] \\
&\quad + p_2\lambda((p_1,p_2]) - (p_2-p_1+1)\int_{p_1}^{p_2} q\lambda(dq) \\
&= (p_2-p_1)\lambda((p_2,p^*]) + p_2\lambda((p_1,p_2]) - \int_{p_1}^{p_2} q\lambda(dq) \\
&\geq (p_2-p_1)\lambda((p_2,p^*]) + p_2\lambda((p_1,p_2]) - p_1\lambda((p_1,p_2]) \\
&= (p_2-p_1)\lambda((p_1,p^*]) \\
&\geq 0.
\end{aligned}$$

The inequalities are strict for a strictly proper scoring rule. □

References

[1] Bruno de Finetti. *Probability, Induction and Statistics*. New York: Wiley, 1972.

[2] Bruno de Finetti. *Theory of Probability*. Vol. 1. New York: Wiley, 1974.

[3] Itzhak Gilboa and David Schmeidler. “Maximin Expected Utility with Non-unique Prior”. In: *Journal of Mathematical Economics* 18 (1989), pp. 141–153. DOI: 10.1016/0304-4068(89)90018-9.

[4] Irving John Good. “Rational decisions”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 14.1 (1952), pp. 107–114. DOI: 10.1111/j.2517-6161.1952.tb00104.x.

[5] JL Hodges Jr and Eric L Lehmann. “Some problems in minimax point estimation”. In: *Selected Works of EL Lehmann*. Springer, 2011, pp. 15–30. DOI: 10.1214/aoms/1177729838.

[6] Nathan Huntley, Robert Hable, and Matthias Troffaes. “Decision Making”. In: *Introduction to Imprecise Probabilities*. Ed. by Thomas Augustin, Frank P.A. Coolen, Gert de Cooman, and Matthias C.M. Troffaes. Chichester, UK: Wiley, 2014. Chap. 8, pp. 190–206.

[7] Isaac Levi. *The Enterprise of Knowledge*. Cambridge: MIT Press, 1980.

[8] Herbert Robbins et al. “Asymptotically subminimax solutions of compound statistical decision problems”. In: *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. Vol. 1. University of California Press Berkeley. 1951, pp. 131–148. DOI: 10.1525/9780520411586-011.

[9] Mark J. Schervish. “A General Method for Comparing Probability Assessors”. In: *The Annals of Statistics* 17.4 (1989), pp. 1856–1879. DOI: 10.1214/aos/1176347398.

[10] Mark J. Schervish, Teddy Seidenfeld, and Joseph B. Kadane. “Proper Scoring Rules, Dominated Forecasts, and Coherence”. In: *Decision Analysis* 6 (2009), pp. 202–221. DOI: 10.1287/deca.1090.0153.

[11] Mark J. Schervish, Teddy Seidenfeld, and Joseph B. Kadane. “The Effects of Exchange Rates on Statistical Decisions”. In: *Philosophy of Science* 80 (2013), pp. 504–532. DOI: 10.1086/673267.

[12] Mark J. Schervish, Teddy Seidenfeld, and Joseph B. Kadane. “Dominating Countably Many Forecasts”. In: *Annals of Statistics* 42 (2014), pp. 728–756. DOI: 10.1214/14-AOS1203.

[13] Mark J. Schervish, Teddy Seidenfeld, Joseph B. Kadane, and Isaac Levi. "Extensions of Expected Utility Theory and Some Limitations of Pairwise Comparisons". In: *ISIPTA-03*. 2003. URL: https://isipta03.sipta.org/proc/papers/037.pdf.

[14] Mark J. Schervish, Teddy Seidenfeld, Rafael B. Stern, and Joseph B. Kadane. "What finite-additivity can add to decision theory". In: *Statistical Methods & Applications* 29 (2020), pp. 237–263. DOI: 10.1007/s10260-019-00486-6.

[15] Amartya K. Sen. "Social choice theory: a re-examination". In: *Econometrica* 45 (1977), pp. 53–89. DOI: 10.2307/1913287.

[16] Cedric A.B. Smith. "Consistency in statistical inference and decision". In: *Journal of the Royal Statistical Society, Series B* 23 (1961), pp. 1–37. DOI: 10.1111/j.2517-6161.1961.tb00388.x.

[17] Abraham Wald. "Statistical decision functions". In: *The Annals of Mathematical Statistics* (1949), pp. 165–205.

[18] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Vol. 42. Monographs on Statistics and Applied Probability. London, UK: Chapman and Hall, 1991.