JMLR: Workshop and Conference Proceedings vol 40:1–18, 2015

Vector-Valued Property Elicitation

Rafael Frongillo RAF@CS.BERKELEY.EDU
Harvard University

Ian A. Kash IANKASH@MICROSOFT.COM
Microsoft Research

Abstract

The elicitation of a statistic, or *property* of a distribution, is the task of devising proper scoring rules, equivalently proper losses, which incentivize an agent or algorithm to truthfully estimate the desired property of the underlying probability distribution or data set. Leveraging connections between elicitation and convex analysis, we address the vector-valued property case, which has received little attention in the literature despite its applications to both machine learning and statistics.

We first provide a very general characterization of linear and ratio-of-linear properties, the first of which resolves an open problem by unifying and strengthening several previous characterizations in machine learning and statistics. We then ask which vectors of properties admit nonseparable scores, which cannot be expressed as a sum of scores for each coordinate separately, a natural desideratum for machine learning. We show that linear and ratio-of-linear do admit nonseparable scores, and provide evidence for a conjecture that these are the only such properties (up to link functions). Finally, we give a general method for producing identification functions and address an open problem by showing that convex maximal level sets are insufficient for elicitability in general.

Keywords: Elicitation, property, proper loss, scoring rule, identification function, expectile.

1. Introduction

Scoring rules, equivalently proper losses, seek to elicit distributions from an agent or algorithm, and are fundamental objects of study in machine learning, statistics, and economics. Recently, there has been much work on a natural extention of scoring rules, to elicit statistics, or *properties* of distributions rather than the entire distribution itself. As Steinwart et al. (2014) point out, these more general scoring rules in a sense capture the power of empirical risk minimization algorithms. Yet despite the fact that most machine learning applications, and many in statistics and economics, are concerned with extracting multi-dimensional information from data, most work in the scoring rules literature has focused on the real-valued property case. In this paper, we address this gap by providing several new results and foundations for a more general study of property elicitation.

The study of properties in scoring rules can be traced to Savage, who considered the problem of eliciting *linear* properties, which are simply the expected value of a random variable (Savage, 1971). Since then, this case has been studied in both the machine learning and scoring rules literatures (Osband and Reichelstein, 1985; Saerens, 2000; Banerjee et al., 2005; Patton, 2011; Gneiting, 2011; Abernethy and Frongillo, 2012), where the conclusion has been that all linear properties are elicitable, and the scores eliciting them can be characterized in terms of *Bregman divergences*. In all cases, however, authors make various simplifying assumptions. Some authors, e.g. (Banerjee et al., 2005; Gneiting, 2011), assume that the report space coincides with the set of possible outcomes. Moreover, all previous results require some sort of differentiablity of the score; Gneiting (2011) points this out, stating that "a challenging, nontrivial problem is to unify and strengthen these results, both in univariate and multivariate settings." In Section 3, we achieve this unification, removing all differentiability assumptions.

Beyond linear properties, many authors have addressed other common statistics, such as quantiles, ratios of expectations, and expectiles (Osband and Reichelstein, 1985; Gneiting and Raftery, 2007; Gneiting, 2011; Grant and Gneiting, 2013). It was perhaps Osband (1985) and Lambert, Pennock, and Shoham (2008), however, who first considered the following general problem: given an outcome space Ω and an *arbitrary* map $\Gamma : \Delta(\Omega) \to \mathcal{R}$, under what circumstances can we construct a proper scoring rule $\mathsf{S} : \mathcal{R} \times \Omega \to \mathbb{R}$ for Γ, i.e. where $\Gamma(p) \in \operatorname{argmax}_{r\in\mathcal{R}} \mathbb{E}_{\omega\sim p}[\mathsf{S}(r, \omega)]$ for every $p \in \Delta(\Omega)$? Moreover, what is the full classification of functions Γ which can be elicited in this way? Several authors have made significant contributions toward answering this general question for the case $\mathcal{R} = \mathbb{R}$ (Lambert et al., 2008; Lambert, 2011; Gneiting, 2011; Gneiting and Katzfuss, 2014; Steinwart et al., 2014), but much less has been done for $\mathcal{R} = \mathbb{R}^k$.

The natural first step from general scalar properties to general vector-valued properties is to consider properties $\Gamma : \mathcal{P} \to \mathbb{R}^k$ composed entirely of elicitable scalar properties $\gamma_i : \mathcal{P} \to \mathbb{R}$. We dub this the *vector of properties* case. Here the question of elicitability is trivial: the score which simply sums proper scores S_i for each coordinate property γ_i will elicit Γ; we call such scores *separable*. Hence it remains only to determine the full class of scores which elicit vectors of properties. In particular, when can we devise a *nonseparable* score for a given Γ?

The question of score separability is important in many machine learning applications. For example, in regression problems it is common to use a notion of error called the *Mahalanobis distance*, which corresponds to the geometry of the data, namely $(x-y)^\top M(x-y)$ for some matrix M which is estimated as the inverse covariance of a sample (Mahalanobis, 1936; De Maesschalck et al., 2000). When M is (highly) non-diagonal, practitioners would be highly dissatisfied with separable losses, as they would suggest that, e.g., certain points are outliers when in fact they are not, or vice versa (Rousseeuw and Leroy, 2005).

Surprisingly, Osband (1985) showed that there do not exist separable scores for pairs of quantiles; see also (Fissler and Ziegel, 2015). Lambert et al. (2008) show that if a scoring rule for a vector of properties is accuracy-rewarding, in the sense that making any coordinate-wise improvement of the report yields a higher score, then the score must be separable; this aligns with our intuition from above, and indeed we are precisely interested in cases where this notion of accuracy may not apply. In Section 4, we show the first nonseparable scores for a vector of nonlinear properties, by fully characterizing the scores for ratios of expectations with the same denominator, and provide evidence for a conjecture that these are the *only* properties which admit nonseparable scores. We then go on to extend Osband's result to show that expectiles do not admit nonseparable scores.

Finally, we conclude by exploring the general case, and ask when a general vector-valued property Γ is elicitable. Building on the idea of *identification*, introduced by Osband (1985), we provide a new construction of identifications functions (which intuitively are derivatives of proper scores) using the derivative of Γ. Prior techniques for constructing an identification function required either manual analysis of the Γ of interest or an existing scoring rule for Γ to differentiate. As our construction is purely mechanical, it provides a general way to construct scoring rules for properties we do not already know how to elicit, as well as a technique to show that some properties are not elicitable. We leverage this to give the first known example of a property with convex maximal level sets that is not elicitable (at least not by any twice differentiable score), thus showing that a natural sufficient condition that Lambert et al. (2008) give for the scalar case does not generalize.

In summary, our contributions improve the understanding of several aspects of vector-valued property elicitation. A number of our results are only part way towards a full understanding, but even these allow us to demonstrate interesting new facts. Through the course of the paper we show:

1. A unified characterization of linear properties (Theorem 11);
2. The first nonlinear property with a nonseparable score (Theorem 13);
3. That expectiles have no nonseparable scores (Corollary 17);
4. That convex maximal level sets, a sufficient condition for elicitability in the scalar case, are not sufficient in the vector case (Example 1).

We begin with notation and a literature review.

2. Background

2.1. Classical scoring rules

A classical scoring rule takes a reported distribution and an observed outcome and determines the payoff to the agent, who is assumed to maximize his or her expected score.

Definition 1 *Given outcome space Ω and convex*[1] *set of probability measures $\mathcal{P} \subseteq \Delta(\Omega)$, a* scoring rule *is a function $\mathsf{S} : \mathcal{P} \times \Omega \to \mathbb{R}$ such that $\mathsf{S}(p', \cdot)$ is integrable with respect to each $p \in \mathcal{P}$.*[2] *We write $\mathsf{S}(p', p) \doteq \mathbb{E}_{\omega \sim p}[\mathsf{S}(p', \omega)]$. We say S is* proper *if for all $p, q \in \mathcal{P}$,*

$$\mathsf{S}(q, p) \leq \mathsf{S}(p, p). \tag{1}$$

If the inequality in (1) is strict for all $q \neq p$, then S is strictly proper.

Scoring rule characterizations make use of the concept of a subgradient from convex analysis.

Definition 2 *Let $\mathcal{T}$ be a subset of a real vector space $\mathcal{V}$. Given function $G : \mathcal{T} \to \mathbb{R}$, a linear function $dG_t : \mathcal{V} \to \mathbb{R}$ is a* subgradient *to G at t if for all $t' \in \mathcal{T}$,*[3]

$$G(t') \geq G(t) + dG_t(t' - t). \tag{2}$$

When $\mathcal{T} = \mathcal{P}$ we will additionally require $dG_{p'}(p) = \mathbb{E}_p[f]$ for some $\mathcal{P}$-integrable $f : \Omega \to \mathbb{R}$. We denote by ∂G_t or $\partial G(t)$ the set of subgradients to G at t, and $\partial G = \cup_{t \in \mathcal{T}} \partial G_t$.

While we focus on the scoring rules setting, many of our results hold for a more general elicitation setting due to Frongillo and Kash (2014), which builds on a large body of work on scoring rules and mechanism design (a field of economics) for situations where the information is directly revealed. Here the agent has a private type $t \in \mathcal{T}$, where $\mathcal{T}$ is a subset of a topological vector space,

1. Many results can be extended to the non-convex case, see (Frongillo and Kash, 2014).
2. To permit scoring rules such as the log score, it is common to relax the requirement that $\mathsf{S}(p, \omega) \in \mathbb{R}$, and relax integrability to quasi-integrability (Gneiting and Raftery, 2007). We ignore these concerns in this paper, but many of our results can be adapted accordingly.
3. While subgradients are often required to be continuous, several of our results do not require this; cf. (Lonard, 2006).

and one only requires that the score to the agent following a report r be *affine* in t. We make use of this greater generality in the proof of Theorem 11, so we state such a more general theorem here. The special case of this theorem applied to scoring rules with convex sets of probability measures is the standard characterization due to Gneiting and Raftery (2007).

Theorem 3 ((Frongillo and Kash, 2014)) *Let $\mathsf{S} : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ be given such that $\mathsf{S}(t', \cdot) : \mathcal{T} \to \mathbb{R}$ is affine for all $t' \in \mathcal{T}$. Then S is proper if and only if there exists some convex $G : \mathsf{Conv}(\mathcal{T}) \to \mathbb{R}$, and some selection of subgradients $\{dG_t\}_{t \in \mathcal{T}} \in \partial G$, such that*

$$\mathsf{S}(t', t) = G(t') + dG_{t'}(t - t'). \tag{3}$$

2.2. Extending to properties

We wish to generalize the notion of a scoring rule, from eliciting a full distribution from $\mathcal{P}$ to accepting reports from a space $\mathcal{R}$ which is different from $\mathcal{P}$. To even discuss what a proper scoring rule is in this setting, we need a notion of a correct report r for a given distribution p. Building on notation introduced by Lambert et al. (2008), and following (Frongillo and Kash, 2014), we encapsulate this notion by a function[4] Γ which specifies the "true" r for a given p.

Definition 4 *Let $\mathcal{P} \subseteq \Delta(\Omega)$ be a convex set of distributions, and $\mathcal{R} \subseteq \mathbb{R}^k$ be some given set of valid reports.*[5] *A* property *is a surjective function $\Gamma : \mathcal{P} \to \mathcal{R}$ which associates a correct report value to each distribution; in particular, $\mathcal{R} = \Gamma(\mathcal{P})$. We let $\Gamma_r \doteq \{p \in \mathcal{P} \,|\, r = \Gamma(p)\}$ denote a* level set *of Γ, the set of distributions p corresponding to report value r.*

We extend the notion of a scoring rule to this setting, where the report space is $\mathcal{R}$ instead of $\mathcal{P}$ itself. The definition is straightforward: the correct report should maximize the expected score.

Definition 5 *A scoring rule $\mathsf{S} : \mathcal{R} \times \Omega \to \mathbb{R}$* elicits *a property $\Gamma : \mathcal{P} \to \mathcal{R}$ if for all $p \in \mathcal{P}$,*

$$\Gamma(p) = \operatorname*{argsup}_{r \in \mathcal{R}} \mathsf{S}(r, p). \tag{4}$$

If we merely have $\Gamma(p) \in \operatorname{argsup}_{r \in \mathcal{R}} \mathsf{S}(r, p)$, we say S weakly elicits *Γ. A property is* elicitable *if some scoring rule elicits it.*

Note that it is certainly possible to write down S such the argsup in (4) is not well defined. This corresponds to some distributions having no optimal report, which we view as violating a minimal requirement for a sensible scoring rule, and hence are not concerned with such cases.

Frongillo and Kash (2014) show how to generalize Theorem 3 to this setting. We state the relevant special case here, which we use in the proof of Lemma 9.

Theorem 6 ((Frongillo and Kash, 2014)) *Let property $\Gamma : \mathcal{P} \to \mathcal{R}$ and scoring rule S be given. Then S elicits Γ if and only if there exists some convex $G : \mathcal{P} \to \mathbb{R}$, some $\mathcal{D} \subseteq \partial G$, and some bijection $\varphi : \mathcal{R} \to \mathcal{D}$ with $\Gamma(p) = \varphi^{-1}(\mathcal{D} \cap \partial G_p)$, such that for all $r \in \mathcal{R}$ and $p \in \mathcal{P}$,*

$$\mathsf{S}(r, p) \;=\; G(p_r) + \varphi(r)(p - p_r), \tag{5}$$

where $\{p_r\}_{r \in \mathcal{R}} \subseteq \mathcal{P}$ satisfies $r' \in \Gamma(p_{r'})$ for all r'.

4. Prior work has allowed Γ to more generally be a multivalued map rather than a function, e.g. for quantiles.
5. Infinite-dimensional report spaces are also of interest, and we briefly discuss how some of our results extend in Sec. 3.

2.3. Integral representations and identification

Savage (1971) observed that to obtain a truthful report π of an agent's value for a commodity, one can simply offer to sell one unit of the commodity at each of a sequence of prices, in even intervals between 0 and the reported price π'. Savage notes that "one unit" could easily be replaced by "any positive amount" and the elicitation would still be valid; making this continuous and selecting some positive amount $\lambda(\pi)$ yields the representation,

$$\mathsf{S}(\pi', \pi) = \int_0^{\pi'} \lambda(\alpha)(\pi - \alpha)d\alpha, \tag{6}$$

since the agent gains value $\lambda(\alpha)\pi$ for each transaction but pays $\lambda(\alpha)\alpha$.

This integral representation of scoring rules has come up repeatedly in the literature; that of Schervish (1989) is perhaps the most famous, but it also appears in Osband (1985; 1989), Lambert et al. (2008; 2011), Gneiting (2011), and Steinwart et al. (2014). An equivalent formulation uses what is called an *identification function* $V(r, \omega)$, a notion introduced by Osband (1985). Identification functions specify the level sets of Γ by having zero expected value precisely when ω is drawn from some $p \in \Gamma_r$. For example, one could define $V(\alpha, \pi) = (\pi - \alpha)$ to recover eq. (6).

Definition 7 *Let property $\Gamma : \mathcal{P} \to \mathcal{R}$ be given, where $\mathcal{R} \subseteq \mathbb{R}^k$. A function $V : \mathcal{R} \times \Omega \to \mathbb{R}^k$ is an* identification function *for Γ, or* identifies Γ, *if for almost every $r \in \mathcal{R}$ it holds that $p \in \Gamma_r \iff V(r, p) = 0 \in \mathbb{R}^k$, where as with $S(r, p)$ above we write $V(r, p) \doteq \mathbb{E}_p[V(r, \omega)]$.*[6]

Lambert et al. (2008) show that a scoring rule is proper for $\Gamma : \mathcal{P} \to \mathbb{R}$ if and only if one can write $\mathsf{S}(r, \omega) = \int_0^r \lambda(r')V(r', \omega)dr'$, where λ is a nonnegative function and V is an identification function uniquely determined by Γ. This is a special case of an earlier result due to Osband (1985), which we restate below. Note that one direction of this result restricts to the relative interior $\text{relint}(\mathcal{R})$ because there is additional flexibility at other points. See further discussion and an example in Section 3. We make extensive use of this theorem in Sections 4 and 5.

Theorem 8 (Osband (1985)) *Let V be a differentiable identification function for $\Gamma : \mathcal{P} \to \mathcal{R}$ for $\mathcal{R} \subseteq \mathbb{R}^k$. Then if S is a twice-differentiable*[7] *strictly proper scoring rule eliciting Γ, for all $r \in \text{relint}(\mathcal{R})$ we can write $\nabla_r \mathsf{S}(r, \omega) = (\Lambda(r)V(r, \omega))^\top$ for $\Lambda : \mathcal{R} \to \mathbb{R}^{k \times k}$ such that $\Lambda(r)\nabla_r V(r, p)$ is negative semi-definite for all $p \in \Gamma_r$. Conversely, if Λ and V are as above and $\Lambda(r)\nabla_r V(r, p)$ is negative definite for all $p \in \Gamma_r$, $\mathsf{S}(r, \omega) = \int_0^r V(r', \omega)^\top \Lambda(r')dr'$ elicits Γ.*

3. Linear properties

This section focuses on the case that Γ is a *linear* function, in the sense that for all $\alpha \in \mathbb{R}$ and all $p_1, p_2 \in \mathcal{P}$ we have $\Gamma(\alpha p_1 + p_2) = \alpha\Gamma(p_1) + \Gamma(p_2)$. In particular, we will consider Γ to be the mean of a random variable, $\Gamma(p) = \mathbb{E}_p[\phi] = \int_\Omega \phi(\omega)dp(\omega)$. As discussed in the introduction, this

6. We have made two relaxations relative to other definitions of identification functions in the literature: (1) V need only identify Γ almost everywhere, and (2) we do not require any order-preservation of V, in part because it is unclear how to define in higher dimensions; we suggest a definition in Sec. 5.

7. Here and throughout, whenever we discuss differentiable S or V, we require differentiation and expectations to commute, i.e. $\nabla_r V(r, p) = \mathbb{E}_p[\nabla_r V(r, \omega)]$. As V is continuously differentiable and thus locally Lipschitz with some constant $L(r, \omega)$, it would suffice for example if L were $\mathcal{P}$-integrable (which e.g. holds for ratio of expectations with $L = b$), allowing the dominated convergence theorem to apply.

case is well-studied in statistics and machine learning, yet previous characterizations make various different assumptions and in particular all require some sort of differentiability. The goal of this section is to unify these characterizations and remove the differentiability assumptions.

That every linear Γ is elicitable can be seen by taking the following score:

$$\mathsf{S}(r, p) = G(r) + dG_r(\Gamma(p) - r), \tag{7}$$

where $G : \mathcal{R} \to \mathbb{R}$ is convex. To see why, consider $\hat{r} = \Gamma(p)$ to be the type, and invoke Theorem 3. By linearity of Γ, the resulting score is still affine in p, which suffices by the affine score framework of (Frongillo and Kash, 2014). (One can also directly verify with one application of the subgradient inequality.) But what forms of S other than (7) weakly elicit Γ?

Note that (7) depends p only though Γ, which is why we can effectively replace $\mathcal{P}$ by $\mathcal{R} = \Gamma(\mathcal{P})$. We now state the crucial lemma of this section, which says that we can always "collapse" the type space to $\mathcal{R}$ in this way, modulo a linear term which does not depend on r. The proof follows from a careful but simple path integration argument.[8]

Lemma 9 *Let $\mathsf{S} : \mathcal{R} \times \Omega \to \mathbb{R}$ be given which weakly elicits a linear property Γ, where $\mathcal{R}$ and $\mathcal{P}$ are convex. Then there exists $\mathcal{P}$-integrable $f : \Omega \to \mathbb{R}$ such that for all $r' \in \mathrm{relint}(\mathcal{R})$ and all $p_1, p_2 \in \mathcal{P}$ such that $\Gamma(p_1) = \Gamma(p_2)$,*

$$\mathsf{S}(r', p_1) - \mathsf{S}(r', p_2) = \mathbb{E}_{p_1}[f] - \mathbb{E}_{p_2}[f]. \tag{8}$$

From Lemma 9, we now conclude that any scoring rule eliciting a linear property Γ must be well-behaved on the affine hull of Γ_r for all $r \in \mathrm{relint}(\mathcal{R})$.

Lemma 10 *Let $\mathsf{S} : \mathcal{R} \times \Omega \to \mathbb{R}$ be given which weakly elicits a linear property $\Gamma : \mathcal{P} \rightrightarrows \mathcal{R}$. Then S weakly elicits Γ on expanded type space $\hat{\mathcal{P}} = \{\alpha p + (1-\alpha)p' : \alpha \in \mathbb{R},\ \exists r \in \mathrm{relint}(\mathcal{R})\ p, p' \in \Gamma_r\}$.*

Proof Let $p_1, p_2 \in \Gamma_r$ and $p = p_1 + \alpha(p_2 - p_1)$ for $\alpha \in \mathbb{R}$. One easily verifies that with the f from Lemma 9 we have $\mathsf{S}(r', p) = \mathsf{S}(r', p_1) + \alpha(\mathbb{E}_{p_2}[f] - \mathbb{E}_{p_1}[f])$, for any $r' \in \mathrm{relint}(\mathcal{R})$. As S weakly elicits Γ and the f term is independent of the report, the result now follows. ■

The main result of this section now follows. Note that, as previous work has done (Lambert et al., 2008), we must be careful to restrict to the relative interior of $\mathcal{R}$ (see Example 2 in Appendix A). To this end, we say that S weakly elicits Γ *on* $\mathcal{P}'$ if $\Gamma(p') \in \mathrm{argsup}_r\, \mathsf{S}(r, p')$ for all $p' \in \mathcal{P}'$. Similarly, we say S weakly elicits Γ on $\mathcal{R}'$ if it weakly elicits Γ on $\Gamma^{-1}(\mathcal{R}')$.

Theorem 11 *Let $\Gamma(p) = \mathbb{E}_{\omega\sim p}[\phi(\omega)]$ where $\phi : \Omega \to \mathcal{R}$ and $\mathcal{R} = \Gamma(\mathcal{P}) \subseteq \mathbb{R}^k$. Scoring rule $\mathsf{S} : \mathcal{R} \times \Omega \to \mathbb{R}$ weakly elicits Γ on $\mathrm{relint}(\mathcal{R})$ if and only if there exists some convex $G : \mathcal{R} \to \mathbb{R}$ with subgradients $\{dG_r\}_{r\in\mathcal{R}}$, and some $\mathcal{P}$-integrable $f : \Omega \to \mathbb{R}$, such that for all $r \in \mathrm{relint}(\mathcal{R})$ and $p \in \mathcal{P}$,*

$$\mathsf{S}(r, \omega) \;=\; G(r) + dG_r(\phi(\omega) - r) + f(\omega) + h(r, \omega)\ , \tag{9}$$

where $\forall r' \in \mathcal{R}, p' \in \mathcal{P}\ \mathbb{E}_{p'}[h(r', \omega)] = 0$.

8. Omitted proofs can be found in Appendix A.

Theorem 11 generalizes (Abernethy and Frongillo, 2012) by removing the need for a differentiability assumption, resolving the problem proposed by Gneiting (2011). Intuitively, Theorem 11 is using the fact from Theorem 6 that $\mathsf{S}(\Gamma(p), p)$ has to be flat along the parallel level sets of Γ, which imposes a very rigid structure on the subgradients at interior points. However, Example 2 in Appendix A shows that all bets are off for boundary points. Finally, note that Theorem 11 applies to some nonlinear properties as well: for any invertible $\psi : \mathcal{R}' \to \mathcal{R}$, replacing r by $\psi(r)$ in the right-hand side of eq. (9) gives a characterization of scores $\mathsf{S} : \mathcal{R}' \times \Omega \to \mathbb{R}$ eliciting $\Gamma'(p) \doteq \psi^{-1}(\Gamma(p))$. Incorporating link functions in this way allows our results to capture, e.g., generalized linear models.

We conclude by noting several possible extensions. While we have assumed $\mathcal{R} \subseteq \mathbb{R}^k$ throughout, many of the results in this section may apply if $\mathcal{R}$ is a subset of a more general topological vector space. Such an extension would allow us to capture examples such as the identity property (for distributions with infinite support), projections from a distribution over a multidimensional outcome to the marginal distribution for a single dimension, and the (countably infinite) property that consists of reporting all the moments of the distribution. A crucial place the finiteness of $\mathcal{R}$ is used is in the very final step of the proof, where we need to pass from the $\mathsf{S}(r, p)$ we have constructed to a $\mathsf{S}(r, \omega)$. This construction works more generally as long as for all linear functions $d : \mathcal{R} \to \mathbb{R}$ and all distributions $p \in \mathcal{P}$, we have $\mathbb{E}_p[d(\phi)] = d(\mathbb{E}_p[\phi]) = d(\Gamma(p))$, i.e., that Fubini's theorem holds. This also requires some care to ensure $\mathbb{E}_p[\phi]$ is in fact well defined for more complicated choices of $\mathcal{R}$. Additionally, while we only considered real-valued scoring rules $\mathsf{S} : \mathcal{R} \times \Omega \to \mathbb{R}$, this can also be relaxed to the extended reals via the same type of arguments and conditions as seen in (Gneiting and Raftery, 2007; Frongillo and Kash, 2014).

4. Vectors of elicitable properties

Given some property $\Gamma : \mathcal{P} \to \mathbb{R}^k$, we will say Γ is a *vector of elicitable properties* if one can write $\Gamma(p) = [\gamma_1(p), \ldots, \gamma_k(p)]$ where each $\gamma_i : \mathcal{P} \to \mathbb{R}$ is an elicitable scalar property. Clearly, Γ is elicitable, as from the elicitability of each γ_i we have scores S_i eliciting γ_i, and thus $\mathsf{S}(r, \omega) = \sum_{i=1}^k \mathsf{S}_i(r_i, \omega)$ elicits Γ. It remains therefore to characterize the scores which elicit Γ. In particular, it is natural to ask when there exist *nonseparable* scores, where no such summable separation across coordinates of r exists.

In fact, we have already seen nonseparable scores for vectors of properties. Taking Γ to be a vector of linear properties and $G(r) = \|r\|^2 + (\sum_i r_i)^2$, Theorem 11 gives a totally nonseparable score. As we saw in Section 3, the linear case is essentially as flexible as the classical scoring rule setting, in which any strictly convex function generates a scoring rule. Thus we may ask, do all vectors of properties admit nonseparable scores? Surprisingly, the answer is no: Osband (1985) shows that any vector of properties containing an α-quantile must be at least partially separable; see also Fissler and Ziegel (2015). We now formalize these notions of separability.

Definition 12 *A scoring rule* S *for* Γ *is* partially separable *if one can write* $\mathsf{S}(r, \omega) = \mathsf{S}_1(\{r_i : i \in \mathcal{I}_1\}, \omega) + \mathsf{S}_2(\{r_i : i \in \mathcal{I}_2\}, \omega)$ *for some partition* $\mathcal{I}_1, \mathcal{I}_2$ *of* $[k]$*. Likewise,* S *is* totally separable *if* $\mathsf{S}(r, \omega) = \sum_{i=1}^k \mathsf{S}_i(r_i, \omega)$*. We say that* S *is* nonseparable *if it is not partially separable, and* totally nonseparable *if for all open sets* $R \subseteq \mathcal{R} = \Gamma(\mathcal{P})$*,* $\mathsf{S}|_R$ *is a nonseparable strictly proper scoring rule for* $\Gamma|_{\Gamma^{-1}(R)}$.

A vector with a quantile and two linear properties admits a partially separable score, where the linear part is nonseparable, but not a nonseparable score. We introduce *totally* nonseparable scores because it is possible to construct scores which are "separable" for all but one part of $\mathcal{R}$ (see Example 3 in the appendix).

As linear scores can be nonseparable, but quantile scores must be separated, the question then becomes, for which vectors of properties do nonseparable scores exist? In particular, are there any *nonlinear* properties which admit nonseparable scores? We find our candidate in the ratio of expectations, $\Gamma(p) = \mathbb{E}_p[a]/\mathbb{E}_p[b]$; as we show, taking a vector of such ratios with the same denominator has as much flexibility as in the linear case. Let $\Gamma(p) = \mathbb{E}_p[\phi]/\mathbb{E}_p[b]$ for $\phi : \Omega \to \mathbb{R}^k$ and $b : \Omega \to \mathbb{R}$ with $\mathbb{E}_p[b] > 0$ for all $p \in \mathcal{P}$. Appealing to Theorem 8, we use the identification function $V(r,\omega) = \phi(\omega) - r\,b(\omega)$. Now choosing strictly convex $G : \mathbb{R}^k \to \mathbb{R}$ and letting

$$\mathsf{S}(r,\omega) = G(r)\,b(\omega) + \nabla_r G \cdot (\phi(\omega) - r\,b(\omega)), \tag{10}$$

we compute $\nabla_r \mathsf{S}(r,p) = \nabla_r^2 G\,(\mathbb{E}_p[\phi] - r\,\mathbb{E}_p[b])$. This is proper from Theorem 8 with $\Lambda(r) = \nabla_r^2 G$, as $\nabla_r^2 \mathsf{S}(r,p) = -\mathbb{E}_p[b]\,\nabla_r^2 G(r)$ is negative definite for all $p \in \Gamma_r$. Moreover, if we choose G as above for the linear case, S is again totally nonseparable.

The form (10) is very similar to eq. (22) of Gneiting (2011), which characterizes the scoring rules for a real-valued ratio of expectations, which restated in our notation is $\mathsf{S}(r,\omega) = g'(r)(a(\omega) - rb(\omega)) - b(\omega)(g(\omega) - g(r)) - g'(\omega)(a(\omega) - \omega b(\omega))$ for some convex function g. Ignoring expressions which depend solely on the outcome, and hence do not affect the elicitation, gives $\mathsf{S}(r,\omega) = g(r)b(\omega) + g'(r)(a(\omega) - rb(\omega))$; our expression is therefore the natural generalization to higher dimensions. This suggests that perhaps (10) is a characterization here as well. We indeed show this to be the case, and by leveraging results from Section 3, we can also generalize our result to allow nondifferentiable scores.

Theorem 13 *Let $\phi : \Omega \to \mathcal{R}$ and $b : \Omega \to \mathbb{R}$. Then scoring rule $\mathsf{S} : \mathcal{R} \times \Omega \to \mathbb{R}$ weakly elicits the ratio of expectations $\Gamma(p) = \mathbb{E}_p[\phi]/\mathbb{E}_p[b]$ on* relint$(\mathcal{R})$ *if and only if there exists some convex $G : \mathcal{R} \to \mathbb{R}$ with subgradients $\{dG_r\}_{r\in\mathcal{R}}$, and some $\mathcal{P}$-integrable $f : \Omega \to \mathbb{R}$, such that for all $r \in$* relint$(\mathcal{R})$ *and ω,*

$$\mathsf{S}(r,\omega) \;=\; G(r)\,b(\omega) + dG_r(\phi(\omega) - r\,b(\omega)) + f(\omega) + h(r,\omega)\;, \tag{11}$$

where $\forall r' \in \mathcal{R}, p' \in \mathcal{P}\ \mathbb{E}_{p'}[h(r',\omega)] = 0$.

We have now seen that linear properties and ratios of expectations (with the same denominator) allow totally nonseparable scores. In fact, we conjecture that these are the only such cases; all scores for other vectors of properties must be separable. This would mean that all such scores satisfy a very restrictive condition known as *accuracy rewarding* (see Theorem 5 of (Lambert et al., 2008)).

Conjecture 14 *Let $\Gamma(p) = [\gamma_1(p), \ldots, \gamma_k(p)]$ where each $\gamma_i : \mathcal{P} \to \mathbb{R}$ is elicitable. There exist totally nonseparable differentable scores eliciting Γ if and only if each γ_i is a link of a ratio of expectations with the same denominator; that is, there exists $b : \Omega \to \mathbb{R}$ such that for each i there exist $\phi_i : \Omega \to \mathbb{R}$ and invertible $\psi_i : \mathbb{R} \to \mathbb{R}$ such that $\gamma_i(p) = \psi_i^{-1}(\mathbb{E}_p[\phi_i]/\mathbb{E}_p[b])$.*

In the remainder of this section we provide some partial results toward this conjecture that help motivate it. We begin by showing that, if S is totally nonseparable, the Λ from Theorem 8 has considerable structure.

Proposition 15 *Let $V(r,\omega) = [v_1(r_1,\omega),\ldots,v_k(r_k,\omega)]$ be a continuously differentiable identification function for Γ, where each v_i identifies each γ_i. If S is a twice differentiable totally nonseparable score with $\nabla_r \mathsf{S}(r,\omega) = \Lambda(r)V(r,\omega)$ for some Λ, then there does not exist an open path-connected $R \subset \mathcal{R}$ such that $\Lambda(r)$ is reducible for almost every $r \in R$.*[9]

Building on this result, we further show that whenever $\Lambda(r)$ is irreducible, $\nabla_r V(r,p)$ is a scalar multiple of the same diagonal matrix for all $p \in \Gamma_r$.

Lemma 16 *Let $V = [v_1,\ldots,v_k]$ and Γ as in Proposition 15. If S a twice differentiable score, then there exist functions $\beta_i(r)$ such that $\beta_i(r)v_i'(r_i,p) = \beta_j(r)v_j'(r_j,p)$ for all r where $\Lambda(r)$ is irreducible and all $p \in \Gamma_r$, where $v_i'(\cdot,p) = \partial/\partial_{r_i} v_i(\cdot,p)$.*

For intuition for the conjecture, suppose that the constraints from Lemma 16 held for *all* p, rather than just $p \in \Gamma_r$. Then that would imply that $v_i'(r_i,\omega) = g_i(r_i)b(\omega)$ for some $b : \Omega \to \mathbb{R}$ and some functions g_i. Integrating, we have $v_i(r_i,\omega) = b(\omega)f_i(r_i) - \phi_i(\omega)$ for some function ϕ_i and where f_i is an antiderivative of g_i. But by definition of v_i we have $\gamma_i(p) = r_i \iff v_i(r_i,p) = 0$, giving $\gamma_i(p) = f_i^{-1}(\mathbb{E}_p[\phi_i]/\mathbb{E}_p[b])$, which would prove the conjecture.

While these partial results are insufficient to prove the conjecture, they suffice to show that some natural properties have no (twice differentiable[10]) totally nonseparable score. In particular, we describe how to recover Osband's result that this holds for α-quantiles, and give a new result that it holds for τ-expectiles. For α-quantiles, taking $\mathcal{P}$ to be distributions on the reals with continuous densities with full support, we take $v_i(r_i,\omega) = \alpha_i - \mathbb{1}_{r_i \geq \omega}$. Evaluating at $p \in \Gamma_r$ and differentiating gives $v_i'(r_i,p) = -p(r_i)$. As these can vary essentially arbitrarily for $p \in \Gamma_r$, the conditions of Lemma 16 will not be satisfied for any fixed choice of the $\beta_i(r)$.

Finally, consider the τ-expectile, which is a type of generalized quantile introduced by Newey and Powell (1987). The τ-expectile is defined by the solution $x = \mu_\tau$ to the equation

$$\mathbb{E}_p\left[|\mathbb{1}_{x \geq \omega} - \tau|(x-\omega)\right] = 0\ .$$

One can check that $v_i(r_i,\omega) = |\tau_i - \mathbb{1}_{r-\omega}|(\omega - r)$ identifies the τ_i-expectile, and as the derivative again contains a $p(r_i)$ term, we have the following.

Corollary 17 *Expectiles have no twice differentiable totally nonseparable score.*

5. General vector-valued properties

5.1. Generating identification functions

In this subsection, we provide a natural way of generating identification functions without first having a proper score to differentiate. We restrict to the case of finite Ω, with $n = |\Omega|$ and $\mathcal{R} \subseteq \mathbb{R}^k$. In addition, we assume that Γ satisfies the following condition, which generalizes a condition that is sufficient for elicitability in the scalar case (Lambert et al., 2008).

$$\text{Condition:} \quad \begin{array}{l}\Gamma \text{ is twice differentiable and has convex maximal level sets} \\ \text{of dimension } \dim(\mathcal{P}) - \dim(\mathcal{R}) = n-k-1\end{array} \tag{12}$$

9. A matrix A is irreducible if the directed graph with edges for nonzero entries of A is strongly connected.
10. This result can likely be extended to the non-twice-differentiable case, due to Alexandrov's theorem that convex functions are twice differentiable almost everywhere (Rockafellar, 1997).

We say a set S is *convex maximal* (in $\mathcal{P}$) if S is the intersection of an affine subspace and $\mathcal{P}$. Equivalently, S contains all points in its affine extension which lie in $\mathcal{P}$.

Proposition 18 *Let $\Gamma : \mathcal{P} \to \mathcal{R}$ satisfy condition (12), and let $\hat{p} : \mathcal{R} \to \mathcal{P}$ be any continuously differentiable function such that $\Gamma \circ \hat{p} = \mathrm{id}_{\mathcal{R}}$ and $\mathrm{rank}(\nabla_{\hat{p}(r)}\Gamma) = n - k$ for almost every r. Then the function*

$$V(r, \omega) = \left(\nabla_{\hat{p}(r)}\Gamma\right)(\delta_\omega - \hat{p}(r)) \tag{13}$$

identifies Γ, where δ_ω is the point distribution on ω.

This proposition benefits from our weakening of the definition of identification to only require it to hold almost everywhere. In particular, if Γ has inflection points $\nabla_p\Gamma$ would be 0 at such points. Thus, this weakening allows our results to hold for examples such as $\Gamma(p) = p_1^3$.

We illustrate the power of Proposition 18 below, by showing that identifications of standard properties can be recovered. Note that such a differentiable right inverse $\hat{p}$ of Γ may fail to exist. In practice, however, the particular choice of $\hat{p}$ is often irrelevant, as in most cases below, and even when the domain of Γ is infinite-dimensional, such as for the quantile, this process often still produces a valid identification function. For these reasons we view this result as a theoretical justification for a construction of identification functions which is quite useful in practice.

Linear and Ratio of expectations. For $\Gamma(p) = Ap/b \cdot p$, $A \in \mathbb{R}^{k\times n}$, we obtain $\nabla_p\Gamma = (b \cdot p)^{-2}((b\cdot p)A - Apb^\top) = (b\cdot p)^{-1}(A - \Gamma(p)b^\top)$. Noting that $\nabla_p\Gamma p = \Gamma(p) - \Gamma(p) = 0$, this gives $V(r)\,p = \nabla_{\hat{p}(r)}\Gamma p = (b\cdot\hat{p}(r))^{-1}(Ap - rb^\top)$. Scaling by $b \cdot \hat{p}(r) \neq 0$, which does not change the identification property, gives the usual V from Theorem 13, as well as the linear case when $b = \mathbb{1}$.

Quantiles. While we have not shown Proposition 18 for the infinite outcome setting, we can nonetheless motivate the traditional α-quantile identification function $V(r, \omega) = \alpha - \mathbb{1}_{r\geq\omega}$. In particular, $\nabla_p\Gamma = \alpha - \mathbb{1}_{\Gamma(p)\geq\omega}$. This gives $V(r, \omega) = (\alpha - \mathbb{1}_{\Gamma(\hat{p}(r))\geq\omega})(\delta_\omega - \hat{p}(r)) = (\alpha - \mathbb{1}_{r\geq\omega}) - (\alpha - \alpha)$.

Finally, we conjecture that this derivation provides a simple test for elicitability. For intuition, suppose that Γ is elicitable and $\nabla_r V(r, p)$ is symmetric for $p \in \Gamma_r$. Then we have $\nabla_r^2 S(r, p) = (\Lambda(r)(\nabla_r V(r, p)))^\top$, where all 3 matrices are symmetric, $\nabla_r^2 S(r, p)$ is negative semidefinite, and $\Lambda(r)$ is positive semidefinite. If both were definite, we could conclude that $\nabla_r V(r, p)$ was negative definite. This intuition, while insufficient for the conjecture, does provide a natural candidate scoring rule for cases where $\nabla_r V(r, p)$ is symmetric.

Conjecture 19 *Let Γ satisfy condition (12). Then Γ is elicitable iff it is elicited by $S(r', \omega) = \int_0^{r'} V(r, \omega)^\top dr$, where V is given by (13).*

5.2. Ordered identification and convexity

Several authors have defined identification functions for real-valued properties to have an additional *order-sensitive* quality, requiring $v(r, p) > 0 \iff r > \Gamma(p)$ (Lambert et al., 2008; Lambert, 2011; Steinwart et al., 2014). Notably, the corresponding characterizations then force the weights $\lambda(r)$ to be nonnegative. At first glance, there does not appear to be a natural notion of order-sensitivity for vector-valued properties. Here we propose such a notion, using Γ to determine the order. The key idea follows from Proposition 18.

Corollary 20 *Let $\Gamma : \mathcal{P} \to \mathcal{R}$ satisfy condition (12). Then for V and $\hat{p}$ as in Proposition 18, any twice differentiable scoring rule $\mathsf{S} : \mathcal{R} \times \Omega \to \mathbb{R}$ eliciting Γ satisfies $\nabla_r \mathsf{S}(r,\omega) = (\Lambda(r)V(r,\omega))^\top$, for $\Lambda(r) \in \mathbb{R}^{k\times k}$ positive semi-definite such that*

$$\Lambda(r)\left(\left(\nabla^2_{\hat{p}(r)}\Gamma\right)\left((p-\hat{p}(r))\otimes\nabla_r\hat{p}\right) - I\right) \tag{14}$$

is negative semi-definite for all $p \in \Gamma_r$.

Just as $\lambda \geq 0$ in the scalar case, we now have a class of identification functions which force $\Lambda \geq 0$ (positive semi-definite). Hence, we may define a Γ-order-preserving identification function as one for which any score eliciting Γ with $\nabla_r \mathsf{S}(r,\omega) = (\Lambda(r)V(r,\omega))^\top$ has $\Lambda(r) \geq 0$ almost everywhere.

This notion of identification has strong ties to convexity. Many explicit characterizations of scoring rules for scalar properties have forms that resemble Bregman divergences; see e.g. (Gneiting, 2011, eqs. 18, 22, 28). Using the fact that the Hessian of a convex function is positive semi-definite, we can provide some intuition for this. Suppose S is as above with $\Lambda(r) = \nabla^2_r G(r)$ for some convex function G. Then integrating by parts gives

$$\begin{aligned}
\mathsf{S}(r,p) &= \nabla_r G\, V(r,p) - \nabla_0 G\, V(0,p) - \textstyle\int_0^r \nabla_{r'} G \nabla_{r'} V(r',p)\, dr' \\
&= \nabla_r G \nabla_{\hat{p}(r)}\Gamma(p-\hat{p}(r)) - \ell\, p - \textstyle\int_0^r \nabla_{r'} G\left(\left(\nabla^2_{\hat{p}(r')}\Gamma\right)\left((p-\hat{p}(r'))\otimes\nabla_{r'}\hat{p}\right) - I\right) dr' \\
&= G(r) + \nabla_r G \nabla_{\hat{p}(r)}\Gamma(p-\hat{p}(r)) - \ell\, p - \textstyle\int_0^r \nabla_{r'} G\left(\nabla^2_{\hat{p}(r')}\Gamma\right)\left((p-\hat{p}(r'))\otimes\nabla_{r'}\hat{p}\right) dr'.
\end{aligned}$$

We now see that the first two terms resemble a nonlinearly transformed linear approximation of G as in the linear case, e.g. eq. (9). The third term is irrelevant, and the final term may provide clues as to which properties Γ are elicitable; see Section 5.3.

5.3. Toward a general characterization

Thus far, we have not seen examples of elicitable properties which are not simply vectors of elicitable properties. This leaves the question, are there elicitable properties which cannot be decomposed coordinate-wise? One simple way to generate such properties is with a *link function*, an invertible map $\psi : \mathcal{R} \to \hat{\mathcal{R}}$. Given some property $\Gamma : \mathcal{P} \to \mathcal{R}$, clearly S elicits Γ if and only if $\hat{\mathsf{S}}(r,\omega) = \mathsf{S}(\psi^{-1}(r),\omega)$ elicits $\hat{\Gamma} = \psi \circ \Gamma$. Thus, while the variance is not elicitable (Osband, 1985; Lambert et al., 2008), the vector-valued property (mean, variance) is via the link $\psi : \{x,y\} \mapsto \{x, y-x^2\}$ from the first and second moment.

While link functions are important in machine learning, and affect the optimization properties of the score or loss, they are merely relabeling the report values, leaving the underlying structure of Γ the same. In particular, the level set structure $L(\Gamma) = \{\Gamma_r : r \in \mathcal{R}\}$ remains fixed under a link operation. It is therefore natural to ask, which level set structures L are elicitable, in the sense that there is some elicitable property Γ such that $L = L(\Gamma)$. So far, we have only seen elicitable L corresponding to vector-valued properties which are composed of elicitable scalar properties, leaving the following question open.

Question 1 *For what partitions L of $\mathcal{P}$ does $L = L(\Gamma)$ for some elicitable Γ?* [11]

11. Given Theorem 6, this is essentially a question about subgradient maps obtainable from convex functions.

Lambert et al. (2008) ask whether having each element of L be convex maximal is sufficient in general. Surprisingly, the following example shows that this is not the case, even when one further requires all level sets to be flats of dimension $n-k-1$. In fact, from our above discussion of level set structures, Example 1 says that no property with level sets $L(\Gamma)=\{\{[a-f(b)c,c,b]:c\in\mathbb{R}\}:(a,b)\in\mathbb{R}^2\}$ is elicitable.[12]

Example 1 *Let $\mathcal{T}=\mathbb{R}^3$, $\mathcal{R}=\mathbb{R}^2$, and define $\Gamma(t)=[t_1+f(t_3)t_2,\ t_3]$ for some differentiable f. One easily checks that $V(r,t)=[t_1+f(r_2)t_2-r_1,t_3-r_2]$ identifies Γ, and thus by Theorem 8, any score S eliciting Γ must have $\nabla_r^2 S(r,t)=\Lambda(r)\nabla_r V(r,t)$ for all $t\in\Gamma_r$, for some $\Lambda:\mathbb{R}^2\to\mathbb{R}^{2\times 2}$. We easily calculate*

$$\nabla_r V(\,\cdot\,,t)=\begin{bmatrix}-1 & t_2 f'(r_2)\\ 0 & -1\end{bmatrix}\ .$$

If $f'(r_2)=0$ for all r_2 then Γ is a linear property and hence elicitable by Theorem 11. If $f'(r_2)\neq 0$, then $\nabla_r V(\,\cdot\,,t)$ depends on t_2, and one can check that t_2 is not fixed within a level set Γ_r. Hence, regardless of Λ, the product $\Lambda(r)\nabla_r V(r,t)$ will fail to be symmetric for some $t\in\Gamma_r$, and we conclude that Γ is not elicitable (at least not by a twice differentiable score). Note that we can easily apply this reasoning to show that e.g. the property $\Gamma(p)=[\mathbb{E}_p[\phi_1]+f(\mathbb{E}_p[\phi_3])\mathbb{E}_p[\phi_2],\ \mathbb{E}_p[\phi_3]]$ is not elicitable, for some $\mathcal{P}$-integrable $\phi:\Omega\to\mathbb{R}^3$.

As with Lemma 16 for the vector of properties case, this example shows how our understanding of the structure of elicitable properties can be used to rule out the elicitability of properties that lack the required structure. Extending our techniques to deepen this understanding seems a promising approach to attack Question 1.

ACKNOWLEDGMENTS

We would like to thank Jake Abernethy for providing some insights for the construction of V using $\nabla\Gamma$, Peter Key for help with the intuition for Conjecture 19, and Panos Toulis and Matus Telgarsky for answering numerous technical questions. We also thank Tilmann Gneiting, Tobias Fissler, and other participants of the 2014 Workshop on Propriety and Elicitability in Heidelberg.

12. Equivalently, recalling Theorem 6, no (smooth) convex function can have a subgradient map matching $L(\Gamma)$.

References

J. Abernethy and R. Frongillo. A characterization of scoring rules for linear properties. In *Proceedings of the 25th Conference on Learning Theory*, 2012.

A. Archer and R. Kleinberg. Truthful germs are contagious: a local to global characterization of truthfulness. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 21–30, 2008.

A. Banerjee, X. Guo, and H. Wang. On the optimality of conditional expectation as a Bregman predictor. *IEEE Transactions on Information Theory*, 51(7):2664–2669, July 2005. ISSN 0018-9448. doi: 10.1109/TIT.2005.850145.

R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart. The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50(1):1–18, January 2000. ISSN 0169-7439. doi: 10.1016/S0169-7439(99)00047-7.

Tobias Fissler and Johanna F. Ziegel. Higher order elicitability and Osband's principle. *arXiv:1503.08123 [math, q-fin, stat]*, March 2015. arXiv: 1503.08123.

Rafael Frongillo and Ian Kash. General truthfulness characterizations via convex analysis. In *Web and Internet Economics*, pages 354–370. Springer, 2014.

T. Gneiting. Making and Evaluating Point Forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

T. Gneiting and A.E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Tilmann Gneiting and Matthias Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151, 2014.

Kyrill Grant and Tilmann Gneiting. Consistent scoring functions for quantiles. In *From Probability to Statistics and Back: High-Dimensional Models and ProcessesA Festschrift in Honor of Jon A. Wellner*, pages 163–173. Institute of Mathematical Statistics, 2013.

N.S. Lambert. Elicitation and Evaluation of Statistical Forecasts. *Preprint*, 2011.

N.S. Lambert, D.M. Pennock, and Y. Shoham. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138, 2008.

Christian Lonard. A Set of Lecture Notes on Convex Optimization with some Applications to Probability Theory. *Technical Report*, 2006.

P. C. Mahalanobis. On the generalised distance in statistics. *Proc. Nat. Inst. Sci. India, A, 2, 49*, 55, 1936.

Whitney K. Newey and James L. Powell. Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pages 819–847, 1987.

Kent Osband. Optimal Forecasting Incentives. *Journal of Political Economy*, 97(5):1091–1112, October 1989. ISSN 0022-3808. doi: 10.2307/1831887.

Kent Osband and Stefan Reichelstein. Information-eliciting compensation schemes. *Journal of Public Economics*, 27(1):107–115, June 1985. ISSN 0047-2727. doi: 10.1016/0047-2727(85)90031-3.

Kent Harold Osband. *Providing Incentives for Better Cost Forecasting*. University of California, Berkeley, 1985.

Andrew J. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256, 2011.

R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*. Springer, October 2011. ISBN 9783540627722.

R.T. Rockafellar. *Convex analysis*, volume 28 of *Princeton Mathematics Series*. Princeton University Press, 1997.

Peter J. Rousseeuw and Annick M. Leroy. *Robust regression and outlier detection*, volume 589. John Wiley & Sons, 2005.

Marco Saerens. Building cost functions minimizing to some summary statistics. *Neural Networks, IEEE Transactions on*, 11(6):1263–1271, 2000.

L.J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, pages 783–801, 1971.

M. J. Schervish. A general method for comparing probability assessors. *The Annals of Statistics*, 17(4): 1856–1879, 1989.

Ingo Steinwart, Robert Williamson, Siyu Zhang, and others. Elicitation and Identification of Properties. In *Proceedings of The 27th Conference on Learning Theory*, pages 482–526, 2014.

C. Zlinescu. *Convex Analysis in General Vector Spaces*. World Scientific Publishing Company, Incorporated, January 2002. ISBN 9789812380678.

Appendix A. Omitted material

Proof of Lemma 9

Proof By Theorem 6, we have some $\varphi : \mathcal{R} \to \mathcal{D}$ such that $dG_p \doteq \varphi(\Gamma(p))$ is a selection of subgradients of convex function $G(p) \doteq \mathsf{S}(\Gamma(p), p)$, and moreover that

$$\mathsf{S}(r', p_1) - \mathsf{S}(r', p_2) = \varphi(r')(p_1 - p_2) \; . \tag{15}$$

Note that as $\mathcal{P}$ is convex and $\mathcal{R} = \Gamma(\mathcal{P})$, we have $\mathrm{relint}(\mathcal{R}) = \Gamma(\mathrm{relint}(\mathcal{P}))$ (cf. (Rockafellar and Wets, 2011, Prop. 2.44)); hence as $r' \in \mathrm{relint}(\mathcal{R})$, we have some $p' \in \Gamma_{r'} \cap \mathrm{relint}(\mathcal{P})$. Now by definition of relint (see (Zlinescu, 2002, pp. 2-3)) we have some $\alpha > 0$ and $v \doteq \alpha(p_1 - p_2)$ such that $p' + v \in \mathcal{P}$.

Let $L_{x,y}$ denote the line integral of the vector field $\{dG_p\}_{p\in\mathcal{P}}$ between points $x, y \in \mathcal{P}$:

$$L_{x,y} \doteq \int_0^1 dG_{x+\lambda(y-x)}(y - x)d\lambda \; . \tag{16}$$

As $\{dG_p\}_{p\in\mathcal{P}}$ is a selection of subgradients, it satisfies two conditions which we will need: (i) *weak monotonicity*, which asserts that $dG_p(p' - p) \leq dG_{p'}(p' - p)$ for all $p, p' \in \mathcal{P}$; (ii) *path independence*, which says that the value of a path integral depends only on the endpoints of the path, and in our notation implies $L_{x,y} + L_{y,z} = L_{x,z}$ for all $x, y, z \in \mathcal{P}$. Weak monotonicity is sufficient for the integrand in eq. (16) to be monotone in λ and thus Reimann integrable; cf. (Archer and Kleinberg, 2008, Thm 3.1). Furthermore, path independence ensures that

$$L_{p_1,\, p_1+v} + L_{p_1+v,\, p'+v} = L_{p_1,\, p'} + L_{p',\, p'+v} \; . \tag{17}$$

Using linearity of Γ, note that $\Gamma(v) = \alpha(\Gamma(p_1) - \Gamma(p_2)) = 0$, and hence $\Gamma(p + v) = \Gamma(p)$ for all $p \in \mathcal{P}$. Applying this fact and $dG = \varphi \circ \Gamma$ to eq. (17), we now have

$$\begin{aligned} L_{p_1+v,\, p'+v} &= \int_0^1 \varphi\Big(\Gamma(p_1 + v + \lambda(p' - p_1))\Big)(p' - p_1)d\lambda \\ &= \int_0^1 \varphi\Big(\Gamma(p_1 + \lambda(p' - p_1))\Big)(p' - p_1)d\lambda = L_{p_1,\, p'}, \end{aligned}$$

whence we conclude $L_{p_1,\, p_1+v} = L_{p',\, p'+v}$. Using similar reasoning, we have $L_{p_1,\, p_1+v} = \int_0^1 \varphi(r)(v)d\lambda = \varphi(r)(v)$, and likewise $L_{p',\, p'+v} = \varphi(r')(v)$. Thus, dividing both of these terms by $\alpha > 0$, we obtain $\varphi(r')(p_1 - p_2) = \varphi(r)(p_1 - p_2)$. Now fix any $r_0 \in \mathrm{relint}(\mathcal{R})$, and recall from Definition 2 that as $\varphi(r_0) \in \partial G$ we have some $\mathcal{P}$-integrable $f : \Omega \to \mathbb{R}$ such that $\varphi(r_0)(p) = \mathbb{E}_p[f]$ for all $p \in \mathcal{P}$. Applying the above with r_0 in place of r' now gives us $\mathbb{E}_{p_1}[f] - \mathbb{E}_{p_2}[f] = \varphi(r_0)(p_1 - p_2) = \varphi(r)(p_1 - p_2) = \varphi(r')(p_1 - p_2)$. ■

Example 2

Example 2 *Let $\mathcal{T} = \{(t_1, t_2) \in [0, 1]^2 : t_2 \leq t_2\}$, and $\Gamma(t) = t_1$. Then the following affine score elicits Γ:*

$$\mathsf{A}(r)(t) = \begin{cases} 2rt_1 - r^2 & \text{if } r > 0 \\ -t_2 & \text{if } r = 0 \end{cases} . \tag{18}$$

To see this, note that $\mathsf{A}(0)(t) \leq 0$ *for all* t*, so when* $t_1 > 0$ *reporting* $r = t_1$ *is optimal. For* $t_1 = 0$*, the only type in* $\mathcal{T}$ *also has* $t_2 = 0$*, and hence* $r = 0$ *strictly dominates. Thus, we have strict elicitation. But* A *cannot be written in the form* (9) *as* A *depends on* t_2 *in a way that is not constant in* r*.*

Proof of Theorem 11

Proof It is trivial that the given form is proper: f plays no part in the elicitation, and as discussed after eq. (7), the remainder is proper by Theorem 3. For the converse, let S weakly eliciting Γ on relint$(\mathcal{R})$ be given, at let f be the functional from Lemma 9. Now let $\hat{p} : \mathcal{R} \to \mathcal{V}$ be a linear right inverse of Γ , satisfying $\Gamma \circ \hat{p} = \mathrm{id}_{\mathcal{R}}$, and define $\mathsf{S}^{\mathcal{R}}(r', r) \doteq \mathsf{S}(r', \hat{p}(r)) - \mathbb{E}_{\hat{p}(r)}[f]$, which can be done by Lemma 10. Then for $p \in \Gamma_r$,

$$\mathsf{S}(r', p) - \mathbb{E}_p[f] - \mathsf{S}^{\mathcal{R}}(r', \Gamma(p)) = \mathsf{S}(r', p) - \mathsf{S}(r', \hat{p}(r)) + \mathbb{E}_{\hat{p}(r)}[f] - \mathbb{E}_p[f] = 0, \tag{19}$$

where the last equality uses Lemma 9. By linearity of $\hat{p}$, $\mathsf{S}^{\mathcal{R}}$ satisfies the conditions of Theorem 3 for type space relint$(\mathcal{R})$, so that $\mathsf{S}^{\mathcal{R}}(r', \Gamma(p)) = G(r') + dG_{r'}(\Gamma(p)) - dG_{r'}(r')$. Now by the fact that $\mathcal{R} \subseteq \mathbb{R}^k$, we can write $dG_{r'}(\Gamma(p)) = \mathbb{E}_p[dG_{r'}(\phi)]$. Letting $Z(\mathcal{P}) = \{h : \Omega \to \mathbb{R} : \mathbb{E}_p[h] = 0 \; \forall p \in \mathcal{P}\}$, we have shown $S(r, p) - \mathbb{E}_p[f] - G(r) - dG_r \cdot (\mathbb{E}_p[\phi] - r) = 0$, which implies

$$\forall r \in \mathrm{relint}(\mathcal{R}), \; \big(h(r, \cdot) : \omega \mapsto S(r, \omega) - f(\omega) - G(r) - dG_r \cdot (\phi(\omega) - r)\big) \in Z(\mathcal{P}) \,,$$

from which the form (9) follows. ■

Example 3 *Let* $\gamma : \mathcal{P} \to \mathbb{R}$ *be an arbitrary elicitable property, and fix some* $\hat{r}$*. By characterizations for the scalar case (Lambert, 2011; Steinwart et al., 2014), we can describe the level set for* $\hat{r}$ *as a hyperplane; specifically, we can write* $\gamma_{\hat{r}} = \{p \in \mathcal{P} : p \cdot x = c\}$ *where* x *is orthogonal to the simplex and* $c \in \mathbb{R}$*. We can simply "slice"* γ *and insert a linear property in the gap; define* $\hat{\gamma}$ *by*

$$\hat{\gamma}(p) = \begin{cases} \gamma(p) & p \cdot x < c \\ \hat{r} + (p \cdot x - c) & c \leq p \cdot x \leq c + \epsilon \\ \gamma(p - \epsilon x) + \epsilon & p \cdot x > c \,, \end{cases} \tag{20}$$

as long as γ *can be extended to* $\mathcal{P} \cup \{p - \epsilon x : p \in \mathcal{P}, p \cdot x > c\}$*. This* $\hat{\gamma}$ *is still elicitable by the same characterizations (all level sets are still convex maximal). Now given any vector of properties* $\Gamma = [\gamma_1, \ldots, \gamma_k]$*, we can simply perform this operation on each* γ_i*, and in principle we can then devise scores which are nonseparable in the region* $[\hat{r}_1, \hat{r}_1 + \epsilon] \times \cdots \times [\hat{r}_k, \hat{r}_k + \epsilon]$*.*

As a concrete instance of this, consider $\gamma(p) = q_\alpha(p)$ *to be the* α*-quantile of* p*. Then we have* $x = \mathbb{1}_{\hat{r} < \omega}$ *and* $c = 1 - \alpha$*, since* $q_\alpha(p) = \hat{r}$ *exactly when* $x \cdot p = \mathbb{E}_p[x] = 1 - \alpha$*. Writing* $p(r, \infty) = \mathbb{E}_p[\mathbb{1}_{r < \omega}]$*, this gives us*

$$\hat{\gamma}(p) = \begin{cases} q_\alpha(p) & p(\hat{r}, \infty) < 1 - \alpha \\ \hat{r} + p(\hat{r}, \infty) + \alpha - 1 & 1 - \alpha \leq p(\hat{r}, \infty) \leq 1 - \alpha + \epsilon \\ q_{\alpha - \epsilon}(p) + \epsilon & p(\hat{r}, \infty) > 1 - \alpha + \epsilon \,, \end{cases} \tag{21}$$

which is well-defined for all $p \in \mathcal{P}$. The identification function in this case is then

$$\hat{v}(r,\omega) = \begin{cases} \mathbb{1}_{r \le \omega} - \alpha & r < \hat{r} \\ \mathbb{1}_{\hat{r} \le \omega} - (\alpha + r - \hat{r}) & \hat{r} \le r \le \hat{r} + \epsilon \\ \mathbb{1}_{r \le \omega + \epsilon} - (\alpha + \epsilon) & r > \hat{r} + \epsilon \; . \end{cases} \tag{22}$$

Note how r is outside the indicator function in the interval $[\hat{r}, \hat{r} + \epsilon]$, giving us a clear linear relationship.

Proof of Theorem 13

Proof We first must establish that the set $\mathrm{dom}(G) = \{\mathbb{E}_p[\phi]/\mathbb{E}_p[b] : p \in \mathcal{P}\}$ is convex. Consider $p, p' \in \mathcal{P}$, $p_\lambda \doteq \lambda p + (1-\lambda)p'$ and $f(\lambda) = \mathbb{E}_{p_\lambda}[\phi]/\mathbb{E}_{p_\lambda}[b]$. Then by linearity of expectation we may write $f(\lambda) = (\lambda v + (1-\lambda)v')/(\lambda x + (1-\lambda)x')$, where $v = \mathbb{E}_p[\phi]$, $v' = \mathbb{E}_{p'}[\phi]$, $x = \mathbb{E}_p[b]$, $x' = \mathbb{E}_{p'}[b]$. Now we have

$$\frac{d}{d\lambda} f(\lambda) = \frac{(v - v')(x' + \lambda(x - x')) - (x - x')(v' + \lambda(v - v'))}{(x' + \lambda(x - x'))^2} = \frac{vx' - v'x}{(x' + \lambda(x - x'))^2} \; ,$$

which shows that $f(\lambda)$ parametrizes a line segment between v/x and v'/x'. As $\mathcal{P}$ is convex, $\mathrm{dom}(G)$ must contain this entire line segment, and hence is convex.

To show the form (11) elicits Γ, we first assume without loss of generality that $f = h = 0$, as they play no role in the optimization. Now simply factor out the b term and apply the subgradient inequality:

$$\begin{aligned} \frac{1}{\mathbb{E}_p[b]} \mathsf{S}(r,p) &= G(r) + dG_r \left(\frac{\mathbb{E}_p[\phi]}{\mathbb{E}_p[b]} - r \right) \\ &= G(r) + dG_r(\Gamma(p) - r) \le G(\Gamma(p)) = \frac{1}{\mathbb{E}_p[b]} \mathsf{S}(\Gamma(p), p). \end{aligned}$$

For the converse, let S be given and define $\psi : p \mapsto \mathbb{E}_p[b]^{-1} p$ and $\mathcal{P}^1 = \psi(\mathcal{P})$, where $\mathcal{P}$ is the domain of Γ. Then for all $p \in \mathcal{P}^1$ we have $\Gamma(p) = \mathbb{E}_p[\phi]$. Furthermore, as $\mathsf{S}(r, \alpha p) = \alpha \mathsf{S}(r,p)$ is well-defined by linearity for all constants $\alpha \in \mathbb{R}$, it must be the case that S elicits Γ on $\mathcal{P}^1$. Applying Theorem 11, we have

$$\mathsf{S}(r, \psi(p)) = G(r) + dG_r \left(\mathbb{E}_{\psi(p)}[\phi] - r \right) + \mathbb{E}_{\psi(p)}[f].$$

Now applying $\mathsf{S}(r,p) = \mathbb{E}_p[b] \mathsf{S}(r, \mathbb{E}_p[b]^{-1} p) = \mathbb{E}_p[b] \mathsf{S}(r, \psi(p))$ and $\mathbb{E}_p[b] \, \mathbb{E}_{\psi(p)}[f] = \mathbb{E}_p[f]$, the form (11) follows. ■

For simplicity the following example is stated for affine scores; the proof of Proposition 18 easily extends to the affine score case with e.g. $V(r,t) = \nabla_{\hat{t}(r)} \Gamma(t - \hat{t}(r))$.

Proof of Proposition 15

Proof Let r be a report value at which $\Lambda(r)$ is reducible. As $\nabla_r V(r,\omega)$ is diagonal, by Theorem 8, $\Lambda(r)$ must be symmetric. Hence, there exists a partition $\mathcal{I}_1, \mathcal{I}_2$ of the indices such that $\lambda_{ij}(r) =$

$\lambda_{ji}(r) = 0$ for all $i \in \mathcal{I}_1$, $j \in \mathcal{I}_2$; without loss of generality, we can write $\Lambda(r) = \begin{bmatrix} \Lambda_1(r) & 0 \\ 0 & \Lambda_2(r) \end{bmatrix}$. Furthermore, $\Lambda_i(r)$ must be a function solely of r_j for $j \in \mathcal{I}_i$: if not, then without loss of generality we have $i_1, j_1 \in \mathcal{I}_1$ and $j_2 \in \mathcal{I}_2$ such that $\lambda_{i_1,j_1}(r)$ depends on r_{j_2}. This implies for all $p \in \Gamma_r$ that $(\nabla_r^2 \mathsf{S}(r,p))_{i_1,j_1} = (\Lambda(r)\nabla_r V(r,p))_{i_1,j_1}$ depends on r_{j_2}, as $(\nabla_r V(r,p))_{ii} = v_i'(r_i)$ depends only on r_i for all i so the dependence on r_{j_2} cannot be canceled. As $\Lambda(r)\nabla_r V(r)\, p$ is a Hessian, this means that its (i_1, j_2) entry must be nonzero, and as ∇V is diagonal we must have $\lambda_{i_1,j_2}(r) \neq 0$, which is a contradiction.

Now suppose for a contradiction that we have an open region $R \subset \mathcal{R}$ such that $\Lambda(r)$ is reducible for all $r \in R$. Fixing $r_* \in R$, for all $r \in R$ we can write $\mathsf{S}(r,\omega) = \mathsf{S}(r^*,\omega) + \int_{r^*}^{r} (\Lambda(r')V(r',\omega))^\top dr' = \mathsf{S}(r^*,\omega) + \int_{r^*_{[1]}}^{r_{[1]}} (\Lambda(r')V(r',\omega))^\top dr' + \int_{r^*_{[2]}}^{r_{[2]}} (\Lambda(r')V(r',\omega))^\top dr' = \mathsf{S}_1(r_{[1]}), p) + \mathsf{S}_2(r_{[2]}, p)$, where $r^* = r^*_{[1]} \oplus r^*_{[2]}$ and $r = r_{[1]} \oplus r_{[2]}$. Hence, S cannot be totally nonseparable. ■

Proof of Lemma 16

Proof By Theorem 8, we have some Λ such that $(\nabla_r \mathsf{S})^\top = \Lambda(r)V(r)$ and for all r, $\Lambda(r)\nabla_r V(r,p)$ is (symmetric and) negative definite for all $p \in \Gamma_r$. By the form of V, we have $\nabla_r V(r,p)$ is a diagonal matrix with diagonal $[v_1'(r_1,p), \ldots, v_k'(r_k,p)]$.

Assume that $\Lambda(r)$ is irreducible. Let i,j such that $\lambda_{ji}(r) \neq 0$; symmetry of $\Lambda(r)\nabla_r V(r,p)$ gives us $\lambda_{ij}(r)v_j'(r_j,p) = \lambda_{ji}(r)v_i'(r_i,p)$, which implies that for all $p \in \Gamma_r$

$$v_i'(r_i,p)/v_j'(r_j,p) = \lambda_{ij}(r)/\lambda_{ji}(r), \tag{23}$$

By irreducibility of $\Lambda(r)$, we have a path of such nonzero entries between any i and j, and multiplying the corresponding ratios gives eq. (23) for all i,j. ■

Proof of Proposition 18

Proof By convex maximality of Γ_r and definition of $\hat{p}$, we must have $\nabla_{\hat{p}(r)}\Gamma(q - \hat{p}(r)) = 0$ for all $p \in \Gamma_r$. By linearity then, we must have $\operatorname{span}(\{p - \hat{p}(r) : p \in \Gamma_r\}) \subseteq \ker \nabla_{\hat{p}(r)}\Gamma$. As $\nabla_{\hat{p}(r)}\Gamma$ is of rank $n-k$ almost everywhere, we must have $\operatorname{span}(\{p - \hat{p}(r) : p \in \Gamma_r\}) = \ker \nabla_{\hat{p}(r)}\Gamma$ for almost every r. We conclude that for almost every r, $V(r,p) = 0 \iff p \in \Gamma_r$. ■

Proof of Corollary 20

Proof Apply Theorem 8 with V, which identifies Γ by Proposition 18. Fixing $p \in \Gamma_r$, we have

$$\begin{aligned}\nabla_r V(r,p) &= \left(\nabla^2_{\hat{p}(r)}\Gamma\right)\left((p - \hat{p}(r)) \otimes \nabla_r \hat{p}\right) - \nabla_{\hat{p}(r)}\Gamma\ \nabla_r \hat{p} \\ &= \left(\nabla^2_{\hat{p}(r)}\Gamma\right)\left((p - \hat{p}(r)) \otimes \nabla_r \hat{p}\right) - I,\end{aligned}$$

as $\Gamma \circ \hat{p} = \mathrm{id}_{\mathcal{R}}$. The form (14) follows. Taking $p = \hat{p}(r)$ forces $\Lambda(r)$ to be positive semi-definite. ■