Journal of Mathematical Economics 46 (2010) 1200–1210

Contents lists available at ScienceDirect

Journal of Mathematical Economics

journal homepage: www.elsevier.com/locate/jmateco

Proper scoring rules with arbitrary value functions

Fang Fang[a], Maxwell B. Stinchcombe[b,*], Andrew B. Whinston[c]

[a] Department of Information Systems and Operations Management, CSU San Marcos, United States
[b] Department of Economics, UT Austin, United States
[c] Department of Information, Risk & Operations Management, UT Austin, United States

ARTICLE INFO

Article history:
Received 3 December 2009
Received in revised form 9 September 2010
Accepted 9 September 2010
Available online 25 September 2010

Keywords:
Expert opinions
Elicitation
Proper scoring rules
Value functions
Convex extensions of functions
Bauer simplexes
Choquet's theorem

ABSTRACT

A scoring rule is proper if it elicits an expert's true beliefs as a probabilistic forecast, and it is strictly proper if it uniquely elicits an expert's true beliefs. The value function associated with a (strictly) proper scoring rule is (strictly) convex on any convex set of beliefs. This paper gives conditions on compact sets of possible beliefs Θ that guarantee that every continuous value function on Θ is the value function associated with some strictly proper scoring rule. Compact subsets of many parametrized sets of distributions on $\mathbb{R}^k$ satisfy these conditions.

1. Introduction

An expert believes that $\mathbf{p}$ is the distribution of a random variable X. $\mathbf{p}$ belongs to Θ, a subset of $\Delta = \Delta(M)$, where M is the range of X, usually $\mathbb{R}^k$, $k \geq 1$, and Δ is the convex of possible probability distributions on M. A **scoring rule** is a $\mathbb{R}$-valued function $(\mathbf{r}, x) \mapsto h(\mathbf{r}, x)$, depending on the report $\mathbf{r}$ and the realization, x, of the random variable X. Knowing the scoring rule, the expert submits a response or report, $\mathbf{r}$, belonging to Θ. The expert's utility to reporting $\mathbf{r}$ is $E^{\mathbf{p}}u(h(\mathbf{r}, X)) = \int u(h(\mathbf{r}, x))\, d\mathbf{p}(x)$ where $u : \mathbb{R} \to \mathbb{R}$ is their von Neumann–Morgenstern utility function.

Interest focuses on scoring rules that motivate the expect to report their true $\mathbf{p}$, the so-called "proper" scoring rules. Originally introduced by Brier (1950) to provide meteorologists with a reward system for accurate forecasts that they would not want to "game," proper scoring rules are now known to be deeply connected to many areas in statistics.[1]

1.1. (Strictly) proper scoring rules

A scoring rule is Θ-**proper** if for all $\mathbf{p} \in \Theta$,

$$\mathbf{p} \in \mathrm{argmax}_{\mathbf{r} \in \Theta} E^{\mathbf{p}} u(h(\mathbf{r}, X)), \tag{1}$$

* Corresponding author. Tel.: +1 512 475 8515; fax: +1 512 471 3510.
E-mail addresses: fangfang@csusm.edu (F. Fang), maxwell@eco.utexas.edu (M.B. Stinchcombe), abw@uts.cc.utexas.edu (A.B. Whinston).

[1] Gneiting and Raftery (2007) give the linkages of scoring to information measures, entropy functions, and Bregman divergences, develop new scoring rules, and give extensive references to the uses that have been made. Dawid (2007) ties proper scoring rules to the geometry of Bayesian decision theory.

0304-4068/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jmateco.2010.09.002

and **proper** if it is Δ-proper. It is **strictly** Θ-**proper** if for all $\mathbf{p} \in \Theta$,

$$\{\mathbf{p}\} = \mathrm{argmax}_{\mathbf{r} \in \Theta} E^{\mathbf{p}} u(h(\mathbf{r}, X)), \tag{2}$$

and **strictly proper** if it is strictly Δ-proper.[2]

There is almost no loss in studying (strictly) proper scoring rules for risk neutral experts. If $h(\cdot, \cdot)$ is a (strictly) proper scoring rule for an expert who is risk neutral, then $u^{-1} \circ h$ is (strictly) proper for an expert with von Neumann–Morgenstern utility function u, although some limitations arise if u is not invertible. We study only the case of a risk neutral expert, i.e. $u(r) \equiv r$. Theoretically, extensions to the case of an invertible nonlinear u are immediate.[3]

1.2. Value functions

Associated with every scoring rule h is the value function

$$V_h(\mathbf{p}) = \max_{\mathbf{r} \in \Delta} E^{\mathbf{p}} h(\mathbf{r}, X). \tag{3}$$

McCarthy (1956) and Savage (1971) characterized scoring rules and their value functions when M, the range of X, is finite.

1.2.1. McCarthy–Savage for finite outcomes

(1) *The value function V_h associated with a (strictly) proper scoring rule h is (strictly) convex on Δ,*
(2) *any (strictly) convex Lipschitz function V on Δ is the value function for some (strictly) proper scoring rule h, and*
(3) *any (strictly) convex function V on* relint (Δ) *is the value function for some (strictly)* relint (Δ)*-proper scoring rule h.*

Part (3) allows for unbounded scoring rules such as the logarithmic scoring rule given in Example 2. A short sketch of the proof is informative. The (strict) convexity of V_h comes from the same basic convex analysis arguments that deliver e.g. the convexity of the profit function in neoclassical economics (e.g. [Mas-Colell et al., 1995, Proposition 5.C.1]). Given a strictly convex V, let $DV(\mathbf{r})$ denote (any element of) its subgradient at $\mathbf{r}$ so that convexity delivers

$$V(\mathbf{p}) \geq V(\mathbf{r}) + (\mathbf{p} - \mathbf{r})DV(\mathbf{r}) \tag{4}$$

with strict inequality when $\mathbf{p} \neq \mathbf{r}$ and V is strictly convex. Let δ_x denote point mass on x and set $h(\mathbf{r}, x) = V(\mathbf{r}) + (\delta_x - \mathbf{r})DV(\mathbf{r})$ so that $E^{\mathbf{p}} h(\mathbf{r}, X) = V(\mathbf{r}) + \sum_x \mathbf{p}(x)(\delta_x - \mathbf{r})DV(\mathbf{r}) = V(\mathbf{r}) + (\mathbf{p} - \mathbf{r})DV(\mathbf{r})$ where $\mathbf{p}(x)$ denotes the probability that $X = x$ under the probability $\mathbf{p}$. Now (4) reads $V(\mathbf{p}) \geq E^{\mathbf{p}} h(\mathbf{r}, X)$ with equality when $\mathbf{p} = \mathbf{r}$ so that $V(\mathbf{p}) = E^{\mathbf{p}} h(\mathbf{p}, X)$ and V is the value function associated with h.

There are two complications to the general study of scoring rules when X takes on infinitely many values. First, it may be desirable to restrict attention to subsets $\mathcal{D}$ of the set of all probabilities. Second, some scoring rules take on the value $-\infty$ and/or integrate to $-\infty$. To handle these possibilities in a unified fashion requires a bit of care (see Gneiting and Raftery's (2007) extension of Hendrickson and Buehler (1971)), but the essential result is as in the finite case — a scoring rule is (strictly) proper if and only if its value function is (strictly) convex.

1.3. What this paper does

This paper gives conditions on sets $\Theta \subset \Delta$ that guarantee that every continuous function, or every smooth function, on Θ can be extended to some V_h on Δ where h is a strictly proper scoring rule. Since $V_h(\cdot)$ is convex, the conditions cannot allow Θ to contain any non-trivial convex sets: for extension of smooth functions in the case that X takes on only finitely many values, Theorem A requires that Θ be a compact subset of the extreme points of a compact convex set K where the boundary of K satisfies curvature conditions; for the extension of continuous functions in the case that X takes on infinitely many values, Theorem B requires that Θ be a compact subset of the extreme points of a Bauer simplex.[4]

The subsets of Δ of the most interest to us are of the form $\{\mathbf{p}_\theta : \theta \in \Theta\}$ where Θ is a compact subset of $\mathbb{R}^\ell$. Theorem C shows that, generically among the continuous parametrizations, $\theta \mapsto \mathbf{p}_\theta$, $\{\mathbf{p}_\theta : \theta \in \Theta\}$ is the compact set of the extreme points of a Bauer simplex if $k \geq \ell$, i.e. if the dimension of the range of X is not smaller than the dimension of the set of parameters. Theorem D shows that many of the classic sets of parametrized distributions satisfy the Bauer condition. Combined with Theorem B, these results means that, both generically and in a large number of well-known cases, arbitrary continuous functions of $\theta \in \mathbb{R}^\ell$ are the value functions for strictly proper scoring rules when θ parametrizes a probability $\mathbf{p}_\theta$.

---

[2] Within the class of strictly proper scoring rules are the **effective scoring rules**. Introduced and studied by Friedman (1983), an effective scoring rule is one where the experts' payoff is larger iff the report is metrically closer to their true beliefs.

[3] See e.g. Allen (1987), Kadane and Winkler (1988) for the theoretical considerations, and Bickel (2007) for discussions of the performance of different scoring rules in the presence of non-linearities.

[4] The original reference is Bauer (1963), but see also Phelps (2001) or Choquet [Choquet, 1969, Section 28].

1.4. Why we care

If experts with higher precision posteriors are the ones with more knowledge and competence, then their opinions are more valuable to the principal. Suppose that $v(\mathbf{p})$ is the principal's value of the report of expert $\mathbf{p}$. Offering a scoring rule h with $V_h = v$ on Θ motivates experts to produce a report only if the value of the report is higher than the expert's cost of producing the report. This observation motivated our original interest in controlling the value functions by choice of scoring rule.

More generally, in many mechanism design models, the agent's private information belongs to a parametrized class of distributions and part of the problem is to motivate the agent to reveal this information. The results here tell us that, both in general and in many specific examples, there are strictly proper scoring rules that solve the revelation problem while simultaneously controlling the payoffs that the agent receives.

1.5. Outline

Section 2 covers the case that X takes values in a finite set, Section 3 covers the same territory when X takes values in an infinite set. Section 4 covers the genericity result and the verifications that several of the classic parametrizations of probabilities on $\mathbb{R}^k$ satisfy the Bauer property. Finally, Section 5 sketches several applications of these results.

2. Finite outcomes

In this section, X takes on n distinct values, the probability distribution to be elicited is a point $\mathbf{p} \in \Theta \subset \Delta \subset \mathbb{R}^n$, $\Delta = \{\mathbf{p} \in \mathbb{R}^n_+ : \sum_x \mathbf{p}(x) = 1\}$ where $\mathbf{p}(x)$ denotes the probability that $X = x$ under the probability $\mathbf{p}$.

2.1. Scoring rule examples

We begin with examples of proper scoring rules that are not strictly proper, and then turn to strictly proper examples.

Example 1 *(Three proper scoring rules).* For these three rules, any $\mathbf{r}$ with the same mean, respectively median, respectively mode as $\mathbf{p}$ is an optimal response.

Name	$h(\mathbf{r}, x)$	V_h
Quadratic loss	$-(\Sigma_i x_i \mathbf{r}(x_i) - x)^2$	$V_{qloss}(\mathbf{p}) = -Var^{\mathbf{p}}(X)$
Absolute deviation	$-\lvert m(\mathbf{r}) - x\rvert$	$V_{ad}(\mathbf{p}) = -E^{\mathbf{p}}\lvert m(\mathbf{p}) - X\rvert$
Zero-one	$\frac{1}{\#M(\mathbf{r})} 1_{M(\mathbf{r})}(x)$	$V_{mode}(\mathbf{p}) = \max_x \mathbf{p}(x)$

where $m(\mathbf{r}) := \max\ \{x : \mathbf{r}((-\infty, x]) \leq (1/2)\}$, denotes the median of $\mathbf{r}$, $M(\mathbf{r}) := \{x' : \mathbf{r}(x') = \max{}_x \mathbf{r}(x)\}$, denotes the mode or set of modal points of $\mathbf{r}$.

In the first case, $\mathbf{r} \mapsto E^{\mathbf{p}} h_{qloss}(\mathbf{r}, X)$ is smooth and concave, but not strictly concave as any $\mathbf{r}$ having the same mean as $\mathbf{p}$ is optimal. In the second and third cases, $E^{\mathbf{p}} h$ is not concave in $\mathbf{r}$. In the second case, any $\mathbf{r}$ having the same median as $\mathbf{p}$ is optimal, in the third, any $\mathbf{r}$ having the same or a smaller set of modal points, $M(\mathbf{r}) \subset M(\mathbf{p})$, is optimal. In all three cases, the associated value functions are convex, but not strictly convex.

Example 2 *(Three strictly proper scoring rules).* From Brier (1950), Good (1952), and Roby (1965), we have the quadratic, logarithmic, and spherical scoring rules respectively.

Name	$h(\mathbf{r}, x)$	Associated value
Quadratic	$2\mathbf{r}(x) - \langle \mathbf{r}, \mathbf{r} \rangle$	$V_{quad}(\mathbf{p}) = \langle \mathbf{p}, \mathbf{p} \rangle$
Logarithmic	$\log(\mathbf{r}(x))$	$V_{\log}(\mathbf{p}) = \Sigma_i p_i \log(p_i)$
Spherical	$\mathbf{r}(x)/\lVert\mathbf{r}\rVert$	$V_{sph}(\mathbf{p}) = \lVert\mathbf{p}\rVert$

where $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes the inner product of vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lVert\mathbf{x}\rVert := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ is the length of the vector $\mathbf{x}$.

In each case, $\mathbf{r} \mapsto E^{\mathbf{p}} h(\mathbf{r}, X)$ is smooth and strictly concave on Δ (or on $\text{relint}(\Delta)$), the directional derivative is equal to 0 for all directions in Δ iff $\mathbf{r} = \mathbf{p}$, and the associated value functions are strictly convex on Δ (or on $\text{relint}(\Delta)$).

2.2. Proper scoring rules and cheap talk

For Blackwell (1953, 1951), an *experiment* gives rise to a random signal, and each possible signal leads to an updated, posterior belief $\mathbf{p} \in \Delta$. Blackwell then reduces the experiment to the induced distribution over possible $\mathbf{p}$'s. Given a $\mathbf{p}$, i.e. given a signal, the decision maker solves the problem

$$U^*(\mathbf{p}) = \max_{a \in A} \int U(a, x)\, d\mathbf{p}(x) \tag{5}$$

where A is a space of actions and $U(\cdot,\cdot)$ is the utility function depending on the outcome x and the action a. Interpreting Θ as a special case of A, this is the same as the problem we have been assuming that the expert solves. Blackwell ranks experiments by the expectation of the convex value function $U^*:\Delta\to\mathbb{R}$ under the induced distribution over possible $\mathbf{p}$'s.

Let $a^*(\mathbf{p})$ denote the solution in (5). If $a^*(\cdot)$ is invertible on Δ, then observing a^* reveals the decision maker's $\mathbf{p}$ and U^* is strictly convex on Δ. If $a^*(\mathbf{p})=a^*(\mathbf{q})$ for $\mathbf{p}\neq\mathbf{q}$, then observing an $a\in a^*(\mathbf{p})$ does not reveal $\mathbf{p}$ as $U^*(\alpha\mathbf{p}+(1-\alpha)\mathbf{q})=\alpha U^*(\mathbf{p})+(1-\alpha)U^*(\mathbf{q})$ for all $\alpha\in[0,1]$. However, just as in the (proof of the) revelation principle, there is a cheap talk interpretation: expand A to $B:=A\times\Delta$ in (5); extend U to $\mathbb{U}$ by setting $\mathbb{U}((a,\mathbf{r}),x)=U(a,x)$; define $b^*(\mathbf{p})$ as the associated argmax set; for all $\mathbf{p}$, $b^*(\mathbf{p})$ includes $(a^*(\mathbf{p}),\mathbf{p})$, and observing *this* element of $b^*(\mathbf{p})$ reveals $\mathbf{p}$. Since $\mathbf{p}$ does not affect utility, we can understand it as "cheap talk."

2.3. Value functions on subsets of Δ

We now give conditions on $\Theta\subset\Delta$ guaranteeing that rich classes of continuous value functions on Θ are associated with strictly proper scoring rules. We use the following notation: Con denotes the set of convex functions on Δ, Con° the set of strictly convex functions; the corresponding sets of restrictions to $\Theta\subset\Delta$ are $\mathrm{Con}_{|\Theta}$ and $\mathrm{Con}^\circ_{|\Theta}$; $C(\Theta)$ denotes the set of continuous functions on Θ; $C^2(\Theta)$ the sup norm dense set of functions on Θ that have twice continuously differentiable extensions to some open neighborhood of Θ; and $\mathrm{extr}(K)$ denotes the set of extreme points of a convex set K.

Theorem A. *Suppose that K is a compact convex subset of Δ, and that Θ is a compact subset of* $\mathrm{extr}(K)$. *Then,*

(1) *if K is a convex polytope, then $C(\Theta)=\mathrm{Con}^\circ_{|\Theta}$,*
(2) *if K is strictly convex with a C^2 boundary in its affine hull, and has bounded and everywhere strictly positive radius of curvature, $C^2(\Theta)\subset\mathrm{Con}_{|\Theta}$, and for each $v\in C^2(\Theta)$ there exists a strictly Θ-proper scoring rule h such that, the restriction of V_h to Θ is equal to v.*

Proof. For (1), suppose that $\Theta\subset\mathrm{extr}(K)$ where K is a convex polytope. Adding extreme points to K if necessary, there is no loss in assuming that the interior of $\mathrm{co}(\Theta)$ is non-empty relative to Δ. Any $v:\Theta\to\mathbb{R}$ can be represented as finitely many points, $((\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n))$ where $y_i=v(\mathbf{x}_i)$ and $\Theta=\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$. We wish to show the existence of a strictly convex $V:\Delta\to\mathbb{R}$ with $V(\mathbf{x}_i)=y_i$. We first show the existence of a convex V_0, then modify it to a strictly convex V.

The existence of a convex V_0: Let $\mathbf{x}_0=(1/n)\sum_i\mathbf{x}_i$. For $y_0<\underline{y}:=\min_i y_i$, consider the lines in $\Delta\times\mathbb{R}$,

$$L_i(y_0)=\{r[(\mathbf{x}_i,y_i)-(\mathbf{x}_0,y_0)]:r\geq 0\},\tag{6}$$

and the cone $C(y_0)\subset\Delta\times\mathbb{R}$ with vertex $(\mathbf{x}_0,y_0)$ defined by $C(y_0)=(\mathbf{x}_0,y_0)+\{\sum_i\lambda_iL_i(y_0):\lambda_i\geq 0\}$. The slice through $C(y_0)$ at $\underline{y}$ is $S(y_0):=C(y_0)\cap\{(\mathbf{x},\underline{y}):\mathbf{x}\in\Delta\}$. $S(y_0)$ has at most n extreme points, and as $y_0\downarrow-\infty$, the finitely many extreme points of $S(y_0)$ converge to the n-point set Θ. Since each $\mathbf{x}_i$ is a positive distance from the convex hull of the $\mathbf{x}_j$, $j\neq i$, for sufficiently negative y_0, $C(y_0)$ is a convex cone with the lines L_i as the extreme rays. Being a convex subset of $\Delta\times\mathbb{R}$ with non-empty interior, $C(y_0)$ is the graph of a convex function, $V_0(\cdot|y_1,\ldots,y_n):\Delta\to\mathbb{R}$. Since the points $(\mathbf{x}_i,y_i)$ belong to the extreme rays, $V_0(\mathbf{x}_i\,|\,y_1,\ldots,y_n)=y_i$ for each i.

The existence of a strictly convex V: The domain of V_0 can, without loss, be extended to A', the affine hull of Δ. Let $A=A'-\mathbf{p}$ for any $\mathbf{p}\in\Delta$ be the translation of A' to be a vector subspace of $\mathbb{R}^n$. Let Z be a mean 0 Gaussian distribution with support equal to A. For each $\epsilon>0$, Jensen's inequality and the support assumption on Z guarantee that the function $V_\epsilon(\mathbf{x}\,|\,y_1,\ldots,y_n)=E\,V_0(\mathbf{x}+\epsilon Z\,|\,y_1,\ldots,y_n)$ is strictly convex.

At each $\mathbf{x}_i$, $V_\epsilon(\mathbf{x}_i\,|\,y_1,\ldots,y_n)>y_i$. However, because the subgradients of V_0 are bounded, as $\epsilon\downarrow 0$, $V_\epsilon(\cdot\,|\,y_1,\ldots,y_n)$ converges uniformly to $V_0(\cdot\,|\,y_1,\ldots,y_n)$. We now show that, for sufficiently small ϵ, it is possible to change the $y_1,\ldots,y_n$ to $y'_1,\ldots,y'_n$ so that $V_\epsilon(\mathbf{x}_i|y'_1,\ldots,y'_n)=y_i$.

It is sufficient to show that for small $\epsilon>0$, the following mapping is invertible,

$$M(y'_1,\ldots,y'_n):=V_\epsilon(\mathbf{x}_1|(y'_1,\ldots,y'_n)),\ldots,V_\epsilon(\mathbf{x}_n|(y'_1,\ldots,y'_n)).\tag{7}$$

For $\epsilon\simeq 0$, $\partial M_i/\partial y'_i\simeq 1$ and $\partial M_i/\partial y'_j\simeq 0$ for $i\neq j$, both evaluated at $(y_1,\ldots,y_n)$. Therefore, Hessian has a dominant diagonal, so that M is invertible.

For (2), we suppose first that K has non-empty relative interior. Since K has a C^2 boundary with bounded and everywhere strictly positive radius of curvature, K is strictly convex and $\mathrm{extr}(K)=\partial K$ [Schneider, 1993, pp. 106–111].

Pick arbitrary $v\in C^2(\Theta)$. Since Θ is a compact subset of ∂K, v can be extended to a C^2 function on (an open neighborhood of) ∂K, also denoted v. We show the existence of a convex $V:\Delta\to\mathbb{R}$ such that $V_{|K}=v$.

Without loss, we first shift and rescale v so that $0<v(\mathbf{q})<1$ for all $\mathbf{q}\in\partial K$. Let $\mathbf{x}_0$ be a point in the relative interior of K. For each $\mathbf{q}\neq\mathbf{x}_0$ in the affine hull of Δ, define $r(q)\in\mathbb{R}$ to be the unique number r such that $\mathbf{x}_0+r(\mathbf{q}-\mathbf{x}_0)\in\partial K$, and define $\mathbf{p}(\mathbf{q})=\mathbf{x}_0+r(\mathbf{q})(\mathbf{q}-\mathbf{x}_0)\in\partial K$. From [Schneider, 1993, op. cit. and Theorem 2.2.9 and seq. (p. 78)], $r(\cdot)$ and $\mathbf{p}(\cdot)$ are C^2. For each $\epsilon\geq 0$, define the C^2 function $V_\epsilon(\mathbf{q})=r(\mathbf{q})[1+\epsilon v(\mathbf{p}(\mathbf{q}))]$ for $\mathbf{q}\neq\mathbf{x}_0$, and $V_\epsilon(\mathbf{x}_0)=0$. We will show that for small enough $\epsilon>0$, V_ϵ is convex. For any such sufficiently small ϵ, the function $V:=(1/\epsilon)[V_\epsilon-1]$ is the convex function with $V_{|K}=v$.

As $\epsilon\downarrow 0$, $V_\epsilon^{-1}(1)$ clearly converges to ∂K in the sense of Hausdorff distance. We show that it converges sufficiently smoothly that the principal radii of curvature (i.e. the eigenvalues of the reverse Weingarten map [Schneider, 1993, p. 105]) of $V_\epsilon^{-1}(1)$

converges to those of ∂K. Consider the mapping from points of the form $(\mathbf{p}, y) \in \partial K \times (0, 1)$, to $M(\mathbf{p}, y) = (\mathbf{r}, 1)$ such that $\mathbf{r}$ is on the line jointing $\mathbf{x}_0$ and $\mathbf{p}$ and $V_\epsilon(\mathbf{r}) = 1$. This is a diffeomorphism, so that $\mathbf{p} \mapsto M(\mathbf{q}, \epsilon v(\mathbf{p}))$ is C^2. Since the eigenvalue map is smooth away from 0, as $\epsilon \downarrow 0$, the principal radii of curvature of $V_\epsilon^{-1}(1)$ converge uniformly to those of ∂K, so that $V_\epsilon^{-1}(1)$ has everywhere strictly positive radius of curvature, and $V_\epsilon(\cdot)$ is a convex function.

Consider the scoring rule defined by $h(\mathbf{r}, \mathbf{p}) = (\mathbf{r} - \mathbf{p})DV(\mathbf{p}) - V(\mathbf{r}) + 2V(\mathbf{p})$ where for each $\mathbf{p}$, $DV(\mathbf{p})$ belongs to the subgradient of V. This is a concave function, and for each $\mathbf{p}$, its supgradient is 0 if $\mathbf{r} = \mathbf{p}$. Further, $h(\mathbf{p}, \mathbf{p}) = V(\mathbf{p})$, so this scoring rule is associated with V. Finally, since ∂K has strict curvature, for any $\mathbf{r} \neq \mathbf{p}$, $(\mathbf{r} - \mathbf{p})DV(\mathbf{p}) < 0$, so it is strictly ∂K-proper, hence strictly Θ-proper.

Finally, if K has empty relative interior, the previous work delivers a convex function V on A, the affine hull of K. Every point $\mathbf{q}$ in Δ has a unique expression as a point $\mathbf{p} \in A$ and a vector $\mathbf{r}$ with $\mathbf{r} - \mathbf{p}$ orthogonal to A. Extend V to $\widehat{V}(\mathbf{q}) = V(\mathbf{p}) + \|\mathbf{r} - \mathbf{p}\|^2$. □

We conjecture that being Lipschitz rather than C^2 is sufficient for Theorem A(2), but non-Lipschitz functions on extr(K) need not extend to convex function an open neighborhood of K.

Example 3. Let $K = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| \le 1\}$ so that extr$(K) = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1\}$ is a set with constant curvature. For $\mathbf{x} = (x_1, x_2) \in$ extr(K), $x_2(x_1) = \pm\sqrt{1 - x_1^2}$. For all such $\mathbf{x}$, we define $f(x_1, x_2) = +\sqrt{|x_2|}$ so that f is uniformly continuous, but not Lipschitz. Fix any $0 < h < 1$, we ask what value a continuous convex extension, $\hat{f}$, of f must take at the point $\mathbf{y} = (-1, h) \in \mathbb{R}^2$.

For any ϵ small enough that $+\sqrt{1 - (-1+\epsilon)^2} < h$,

$$f(-1+\epsilon, +\sqrt{1-(-1+\epsilon)^2}) = +\sqrt[4]{\epsilon(2-\epsilon)}.$$

From this, we see that the slope of the line joining the two points in the graph of f, $((-1, 0), 0)$ and $((-1+\epsilon, x_2(-1+\epsilon)), +\sqrt[4]{\epsilon(2-\epsilon)})$ is larger than any n for sufficiently small ϵ. This and the convexity of $\hat{f}$ in turn imply that $\hat{f}(-1, h) \ge nh$ for each n. Since h was arbitrary, there is no continuous extension of f to any open neighborhood of K.

3. Infinite outcomes

In this section, X takes on values in a complete separable metric space (M, d), and the probability to be elicited is a point $\mathbf{p} \in \Theta \subset \Delta(M)$ where $\Delta(M)$ is the set of countably additive probabilities on $\mathcal{B}_M$, the Borel σ-field on M. For our applications, $M = \mathbb{R}^k$, $k \ge 1$, and we will often restrict our attention to convex subsets $\mathcal{D}(M)$ of $\Delta(M)$, e.g. taking $\mathcal{D}(M)$ to be the set of probability distributions on $M = \mathbb{R}$ having finite first and second moments.[5]

Closure and continuity in $\Delta(M)$ is taken with respect to the weak* topology — $\mathbf{p}_n \to \mathbf{p}$ if and only if $\int f\,d\mathbf{p}_n \to \int f\,d\mathbf{p}$ for every continuous bounded $f : M \to \mathbb{R}$. Probabilities on $\Delta(M)$ are probabilities on the associated Borel σ-field, which can also be characterized as the smallest σ-field containing the sets $\{\nu \in \Delta(M) : \nu(A) \le r\}$ where $A \subset M$ is measurable and $r \in [0, 1]$. As needed, we will consider $\mathcal{D}(M)$ and $\Delta(M)$ as subset of the topological vector space ca(M), the set of countably additive, finite, signed measures, ξ, on $\mathcal{B}_M$ with the weak* topology, that is, the weakest topology making the mappings $\xi \mapsto \int f\,d\xi$ continuous, f a continuous bounded function on M.

3.1. Scoring rule examples

Restricted to the appropriate sets of probability distributions, the scoring rules given in Examples 1 and 2 have fairly immediate extensions to the infinite case. We give two of the extensions to indicate what is involved, and then give two separate extensions of the quadratic loss ideas.

Example 4 *(Extensions of finite examples).* Let $\Delta^2(\mathbb{R})$ denote the set of all probability distributions on $\mathbb{R}$ having finite second moments. The quadratic loss scoring rule is $h_{qloss}(\mathbf{r}, x) = -\left(\int y\,d\mathbf{r}(y) - x\right)^2$, and any $\mathbf{r} \in \Delta^2(\mathbb{R})$ with $E^{\mathbf{r}}X = E^{\mathbf{p}}X$ is a solution to $\max \int h_{qloss}(\mathbf{r}, x)\,d\mathbf{p}(x)$, so that h_{qloss} is a $\Delta^2(\mathbb{R})$-proper, but not strictly proper, scoring rule.

Let $\Delta^2(\lambda)$ denote the set of all probabilities $\mathbf{q}$ having a square integrable density, also denoted $\mathbf{q}$, with respect to Lebesgue measure λ. The spherical scoring rule, given by $h_{sph}(\mathbf{r}, x) = \mathbf{r}(x)/\|\mathbf{r}\|$ where $\|\mathbf{r}\|$ is the L^2 norm of the density $\mathbf{r}$, is a strictly $\Delta^2(\lambda)$-proper scoring rule.

Example 5 *(Integrated quadratic loss rules).* Let $\Delta^1(\mathbb{R})$ denote the set of all probability distributions on $\mathbb{R}$ having finite first moments. Matheson and Winkler (1976) give the **continuous ranked probability score**, $h_{crps}(\mathbf{r}, x) = -\int_{\mathbb{R}} (F_{\mathbf{r}}(y) - 1_{(-\infty, y]}(x))^2\,dy$ where $F_{\mathbf{r}}(y) := \mathbf{r}((-\infty, y])$ is the cdf associated with $\mathbf{r}$. This is the integral of Brier's quadratic scoring rule for all of the binary events $X \le y$ and $X > y$.[6]

The logic of the proof of the McCarthy–Savage theorem yields the following, which is essentially a summation of well-chosen quadratic loss rules.

[5] This allows the use of unbounded scoring rules, e.g. quadratics, without needing to consider integrals taking the value $-\infty$.
[6] Gneiting and Raftery (2007) [Section 4.2] give several useful reformulations of this rule and relatively easy methods to calculate the associated value function.

Example 6 *(Extensions to complete separable metric spaces).* It is well-known (e.g. [Corbae et al., 2009, Ch. 9]) that there exists a countable set $\{f_i : i \in \mathbb{N}\}$ of continuous functions on M taking values in $[-1, +1]$ with the property that $\mathbf{p}_n$ weak* converges to $\mathbf{p}$ iff $\langle f_i, \mathbf{p}_n\rangle \to \langle f_i, \mathbf{p}\rangle$ for all for all $i \in \mathbb{N}$ (where $\langle g, \mathbf{q}\rangle := \int g\, d\mathbf{q}$). The function

$$V(\mathbf{p}) = \frac{1}{2}\sum_i \frac{(\langle f_i, \mathbf{p}\rangle)^2}{2^i} \tag{8}$$

is strictly convex, and at any $\mathbf{r} \in \Delta$, its derivative is the continuous function

$$\mathrm{D}V(\mathbf{r}) = \sum_i f_i \cdot \frac{\langle f_i, \mathbf{r}\rangle}{2^i}. \tag{9}$$

Since V is strictly convex, for all $\mathbf{r} \neq \mathbf{p}$, $V(\mathbf{p}) > V(\mathbf{r}) + (\mathbf{p} - \mathbf{r})\mathrm{D}V(\mathbf{r})$, so that setting $h(\mathbf{r}, x)$ equal to $V(\mathbf{r}) + (\delta_x - \mathbf{r})\mathrm{D}V(\mathbf{r})$ yields

$$h(\mathbf{r}, x) = V(\mathbf{r}) + \sum_i \frac{(f_i(x) - \langle f_i, \mathbf{r}\rangle)\langle f_i, \mathbf{r}\rangle}{2^i}. \tag{10}$$

Therefore, $E^{\mathbf{p}} h(\mathbf{r}, X) = V(\mathbf{r}) + (\mathbf{p} - \mathbf{r})\mathrm{D}V(\mathbf{r})$, and for all $\mathbf{r} \neq \mathbf{p}$, $E^{\mathbf{p}} h(\mathbf{p}, X) > E^{\mathbf{p}} h(\mathbf{r}, X)$ so that we have a strictly proper scoring rule. If one took the sum over a finite subset of the f_i, one would have a proper but not strictly proper scoring rule.

3.2. Simplexes

A compact convex S in a topological vector space $\mathfrak{X}$ is a **simplex** if every point in S has a unique representation as a convex combination of the extreme points of K. If S is a subset of the vector space $\mathbb{R}^n$, a simplex is the set of convex combinations of at most $n+1$ affinely independent points. If S is a subset of the vector space $\mathrm{ca}(M)$, we must also take integrals as convex combinations, and the convex combinations are known as **resultants**

Definition 1. For K a compact convex subset of $\Delta(M)$ and η a probability distribution on K, the **resultant of** η, denoted $\mathbf{x}_\eta$, is the unique element $\mathbf{x} \in K$ with such that for every continuous linear $L : K \to \mathbb{R}$, $L(\mathbf{x}) = \int L(\mathbf{y})\, d\eta(\mathbf{y})$.

More than finite weighted combinations are needed: in $\Delta(\mathbb{R})$, $S := \{\mathbf{p} \in \Delta(\mathbb{R}) : \mathbf{p}([0, 1]) = 1\}$ is a compact and convex set; $\mathrm{extr}\,(S)$ is the compact set of point masses on $[0, 1]$; $\mathrm{co}\,(\mathrm{extr}\,(S))$, the convex hull of $\mathrm{extr}\,(S)$, contains only the finitely supported probabilities while $\overline{\mathrm{co}}(\mathrm{extr}\,(S)) = S$; and every $\mathbf{p} \in S$ has a unique expression as the resultant of a probability distribution over $\mathrm{extr}\,(S)$.

Another way to understand the resultant is as the "average" probability distribution when a distribution is drawn according to η. For any measurable $A \subset M$, picking $\mathbf{y} \in \Delta(M)$ according to η gives rise to the random variable $\mathbf{y}(A)$. By the usual approximation arguments,[7] for every measurable $A \subset M$ and every probability distribution η on a set of probabilities, the resultant of η satisfies

$$\mathbf{x}_\eta(A) = \int \mathbf{y}(A)\, d\eta(\mathbf{y}). \tag{11}$$

We are interested in simplexes having the property that the set of extreme points is a closed set. Already in $\mathbb{R}^3$, the set of extreme points of a compact and convex set need not be a closed set.

Example 7. If $A \subset \mathbb{R}^3$ contains the points $(0, 0, +1)$, $(0, 0, -1)$ and $\{\mathbf{x} \in \mathbb{R}^3 : x_3 = 0, ||\mathbf{x} - (1, 0, 0)|| = 1\}$, and $K = \overline{\mathrm{co}}(A)$ is its closed convex hull, then $\mathrm{extr}\,(K)$ is the non-closed set $A \setminus \{(0, 0, 0)\}$ because $(0, 0, 0) = \frac{1}{2}(0, 0, +1) + \frac{1}{2}(0, 0, -1)$.

Definition 2. A compact convex $S \subset \mathfrak{X}$ is a **simplex** if for every element $\mathbf{x} \in S$, there is a unique probability $\eta_{\mathbf{x}}$ for which $\mathbf{x}$ is the resultant, and $\eta_{\mathbf{x}}(A) = 1$ for every analytic set A containing $\mathrm{extr}\,(S)$. It is a **Bauer simplex** if $\mathrm{extr}\,(S)$ is a closed (hence compact) set, in which case $\eta_{\mathbf{x}}(\mathrm{extr}\,(S)) = 1$.

In $\mathbb{R}^n$, every simplex S is a Bauer simplex. As noted above, $S = \{\mathbf{p} \in \Delta(\mathbb{R}) : \mathbf{p}([0, 1]) = 1\}$ is a Bauer simplex. In general, if S is a simplex and Θ is a closed subset of $\mathrm{extr}\,(S)$, then $\overline{\mathrm{co}}(\Theta)$ is a Bauer simplex.

For a compact convex K, let $C_{aff}(K)$ denote the set of continuous affine functions on K and $C(\mathrm{extr}\,(K))$ the set of continuous functions on $\mathrm{extr}\,(K)$. For us, the following is the crucial characterization result for Bauer simplexes.[8]

Bauer's Theorem. *If S is a simplex, then $\mathrm{extr}\,(S)$ is closed if and only if the restriction mapping, $a \mapsto a_{|\mathrm{extr}\,(S)}$, defines an isometric isomorphism between $C_{aff}(S)$ and $C(\mathrm{extr}\,(S))$.*

---

[7] Indicators of measurable subsets are nearly indicators of closed subsets which are nearly continuous functions.

[8] Bauer (1963) contains this and several other characterizations of the class of simplexes having closed sets of extreme point. Expositions of this and related results can be found in Choquet (1969), Phelps [Phelps, 2001, Proposition 11.1], or Goodearl [Goodearl, 1986, Ch. 7].

For us, extr (S) will be of the form $\{\mathbf{p}_\theta : \theta \in \Theta\}$ where $\theta \leftrightarrow \mathbf{p}_\theta$ is a homeomorphism. Pick any continuous function, $\mathbf{p}_\theta \mapsto v(\theta)$, on Θ, and any strictly convex V on $\mathcal{D}$. Restricted to extr (S), the difference between the two is a continuous function. Adding the associated affine function to V gives a strictly convex value function that is equal to v on $\{\mathbf{p}_\theta : \theta \in \Theta\}$.

Bauer's result has very clear geometric intuition in the case that Δ is 2-dimensional: for S to be a simplex in this case, it must be the convex hull of 3 (or fewer) affinely independent probabilities, $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$; any function f : extr $(S) \to \mathbb{R}$ is continuous; f determines and is determined by the three points $(\mathbf{p}_1, f(\mathbf{p}_1))$, $(\mathbf{p}_2, f(\mathbf{p}_2))$, and $(\mathbf{p}_3, f(\mathbf{p}_3))$ in $\Delta \times \mathbb{R}$; these three points determine a plane in $\Delta \times \mathbb{R}$; and this plane is the graph of an affine function on Δ. For the isometric isomorphism part of the result, note that, restricted to S, the sup norm distance between any two affine functions is maximized at the extreme points.

3.3. Value functions on subsets of $\Delta(M)$

We shall see that the most useful condition for applications is the first one.

Theorem B. *Suppose that Θ is a compact subset of a convex $\mathcal{D} \subset \Delta(M)$.*

(1) *If $\overline{\text{co}}(\Theta)$ is a Bauer simplex and Θ is the set of extreme points of $\overline{\text{co}}(\Theta)$, then every $v \in C(\Theta)$ is the restriction to Θ of a continuous strictly convex function V on $\mathcal{D}$.*
(2) *If Θ is a compact subset of* extr (K), *$K \subset \mathcal{D}$ compact and convex, the set of restrictions of the continuous strictly convex functions on K is dense in $C(\Theta)$ in the topology of pointwise convergence.*

If the conditions of Theorem B(1) are satisfied, then any continuous value function on Θ is associated with some $\mathcal{D}$-strictly proper scoring rule. What is striking about the proof of this part of the result is that any continuous value function, v, on Θ, is a continuous affine function away from the value function for *any* strictly proper scoring rule. This means that, starting from one's favorite strictly proper scoring rule, $h(\mathbf{r}, x)$, and needs only consider scoring rules of the form $h(\mathbf{r}, x) + g(x)$, g continous, to achieve the desired value function.

In Theorem B(2), neither Θ nor K need be a simplex, but it would be useful to be able to improve on the topology of pointwise convergence.

Proof. Pick arbitrary $v \in C(\Theta)$ and an arbitrary strictly convex and continuous $V : \mathcal{D} \to \mathbb{R}$. Define $f \in C(\Theta)$ as $f = V_{|\Theta} - v$ so that $V_{|\Theta} - f = v$. Since V and v are continuous, f is also continuous. Since $\overline{\text{co}}(\Theta)$ is a Bauer simplex, f determines a unique continuous affine function, a, on $\overline{\text{co}}(\Theta)$. By the Hahn–Banach theorem, this has a further extension, $\hat{a}$, to $\mathcal{D}$. The function $V - \hat{a}$ is strictly convex on $\mathcal{D}$, and restricted to Θ, it is equal to v.

Now suppose that K is compact and convex (not necessarily a simplex), and that Θ is a compact subset of extr (K). Pick $v \in C(\Theta)$, a finite set $\mathbf{p}_1, \ldots, \mathbf{p}_n$ in Θ, and $\epsilon > 0$. We must show that there exists a strictly convex function, V, on K with $|V(\mathbf{p}_i) - v(\mathbf{p}_i)| < \epsilon$ for $i = 1, \ldots, n$. Let S denote the affine hull of the $\mathbf{p}_i$. From Theorem A(1), we know that there exists a strictly convex $V^\circ : S \to \mathbb{R}$ with $|V^\circ(\mathbf{p}_i) - v(\mathbf{p}_i)| = 0$ for each i. There is no loss in assuming that the topology on K is generated by a Hilbert space norm.[9] Therefore we can express every $\mathbf{q} \in \Delta$ uniquely as the sum of $\mathbf{p}(\mathbf{q})$, its orthogonal projection onto S, and $\mathbf{p} - \mathbf{p}(\mathbf{q})$. Extend V° to a strictly convex function on Δ by setting $V(\mathbf{q}) = V^\circ(\mathbf{p}(\mathbf{q})) + \delta\,||\,\mathbf{p} - \mathbf{p}(\mathbf{q})\,||^2$, $\delta > 0$. □

4. Verifications of the Bauer property

The first result in this section shows that generic continuous parametrizations result in Bauer simplexes. The second part of this section verifies that the same is true for several classical parametrized sets of distributions.

4.1. Prevalence of the Bauer property

We begin with

Definition 3. A continuous parametrization, $f : \Theta \to \Delta(M)$, Θ compact, has the **Bauer property** if $\overline{\text{co}} f(\Theta)$ is a Bauer simplex and $f(\Theta)$ is its compact set of extreme points.

We will show that a prevalent[10] set of parametrizations has the Bauer property when $k \geq \ell$ where $M = \mathbb{R}^k$ and $\Theta \subset \mathbb{R}^\ell$. As comforting as this is for the general applicability of the theory, there is no guarantee that any particular parametrization, f, has the Bauer property.

Suppose that X takes values in $\mathbb{R}^k$, that the principal wants the expert to report on the joint distribution of k random variables. In many cases, one can easily show that the requirement $k \geq \ell$ cannot be avoided.

---

[9] The weak* distance between $\mathbf{p}$ and $\mathbf{q}$ can be metrized using the ℓ_2-norm $\left(\sum_i \frac{1}{2^i} |\int f_i\, d\mathbf{p} - \int f_i\, d\mathbf{q}|^2\right)^{1/2}$ where $\{f_i : i \in \mathbb{N}\}$ is any uniformly bounded set of continuous functions determining weak* convergence.

[10] Prevalence is an infinite dimensional analogue of full Lebesgue sets, and shyness an analogue of Lebesgue null sets. Its extension to convex infinite dimensional subsets of topological vector spaces is due to Anderson and Zame (2001).

Let $\mathfrak{X} = C(\Theta;\text{ca}(\mathbb{R}^k))$ denote the set of continuous functions from Θ to $\text{ca}(\mathbb{R}^k)$ with the uniform topology. This is the weakest topology containing sets of the form $\{g:(\forall\theta)[g(\theta)\in h(\theta)+U]\}$, $h\in C(\Theta;\text{ca}(\mathbb{R}^k))$ and U an open subset of $\text{ca}(\mathbb{R}^k)$. This makes $C(\Theta;\text{ca}(\mathbb{R}^k))$ into a topological vector space. Relativized to $\Delta(\mathbb{R}^k)\subset\text{ca}(\mathbb{R}^k)$, the weak* topology can be metrized by e.g. the Prokhorov metric, ρ, making it a complete separable metric space. The convex set $C = C(\Theta;\Delta(\mathbb{R}^k))\subset\mathfrak{X}$ is the set of continuous parametrizations of probability distributions on $\mathbb{R}^k$. C inherits the uniform topology from $\mathfrak{X}$, and this can be metrized by $d(f,g) = \max_\theta \rho(f(\theta), g(\theta))$. With this metric, C is also a complete separable metric space, which is required for prevalence to have intuitive properties.

Theorem C. *If Θ is homeomorphic to $[0,1]^\ell$, X takes values in $\mathbb{R}^k$, and $k\geq\ell$, then a prevalent subset of C has the Bauer property.*

Proof. We replace Θ with its homeomorphic image, $[0,1]^\ell$, and divide the argument into two parts, $k=\ell$ and $k>\ell$.

Suppose first that $k=\ell$. Let E denote the set of $f\in C(\Theta;\Delta(\mathbb{R}^k))$ that do not have the Bauer property. An easy projection argument shows that E is universally measurable. We show that E is relatively shy at z where $z(\theta)\equiv\delta_0$, i.e. $z(\theta)$ is identically equal to point mass at the origin in $\mathbb{R}^k$. Pick an arbitrary $\delta\in(0,1]$ and open neighborhood of 0, $W\subset C(\Theta;\text{ca}(\mathbb{R}^k))$. Since every Borel probability on the complete separable metric space $C(\Theta;\Delta(\mathbb{R}^k))$ is tight, it is sufficient to show the existence of a Borel probability μ such that $\mu(E+h)=0$ for all $h\in\mathfrak{X}$ and $\mu((\delta C+(1-\delta)z)\cap(W+z))=1$.

For $r>0$, we first construct a probability η_r on $C(\Theta;\Delta(\mathbb{R}^k))$ with the property that for all $\theta\in\Theta$, $\eta_r(\theta)(B_r(0))=1$ where $B_r(0)$ is the ball with radius r around 0. We will set $\mu=\delta\eta_r+(1-\delta)z$ and verify that $\mu((\delta C+(1-\delta)z)\cap(W+z))=1$ and $\mu(E+h)=0$ for all $h\in\mathfrak{X}$.

1. *Defining* η_r: Let $\mathbf{1}$ denote the vector of 1's in $\mathbb{R}^k$. For $x\in\mathbb{R}^k$ and $s>0$, define the box $[x, x+s\mathbf{1}]$ to be the set $\{y\in\mathbb{R}^k : x\leq y\leq x+s\mathbf{1}\}$. Let $\{G_k : k\in\mathbb{N}\}$ enumerate a countably infinite, disjoint collection of box subsets of $\{y\in\mathbb{R}^k : 0<||y||<r\}$, $r>0$. For each box $G_k=[x_k, x_k+s_k\mathbf{1}]$, define $g_k^0(\theta)=x_k+(s_k/3)\theta$ and $g_k^1(\theta)=(x_k+(2s_k/3)\mathbf{1})+(s_k/3)\theta$. This guarantees that $g_k^0(\Theta)$ and $g_k^1(\Theta)$ are disjoint box subsets of G_k. Let Y_k be an iid collection of Bernoulli random variables with $P(Y_k=0)=P(Y_k=1)=1/2$. Define a random point masses $p_k(\theta)$ as being point mass on $g_k^{Y_k}(\theta)$, and define η_r as the distribution of the random element of $C(\Theta;\Delta(\mathbb{R}^k))$ defined by $\sum_k(1/2^k)p_k(\theta)$.
   We define $\mu=\delta\eta_r+(1-\delta)z$.
2. *Verifying that* $\mu((\delta C+(1-\delta)z)\cap(W+z))=1$: As $r\downarrow 0$, $d(\eta_r, z)\downarrow 0$. Since $(W+z)\cap C$ is open, for sufficiently small r, $\mu((\delta C+(1-\delta)z)\cap(W+z))=1$.
3. *Verifying* $\mu(E+h)=0$ *for all* $h\in\mathfrak{X}$: Let $\nu_1'\neq\nu_2'$ be two probabilities on Θ. For any sequence $y:=(y_1,y_2,y_3,\ldots)$ be a sequence of realizations of the iid sequence $(Y_1,Y_2,Y_3,\ldots)$ and let $f_y(\theta)=\sum_k(1/2^k)\delta_{g_k^{y_k}(\theta)}$ denote the corresponding element of C. We will show that for a probability one set of y's, the resultants of ν_1' and ν_2' under $f_y(\theta)+h(\theta)$ are different.
   Re-scaling the two strictly positive measures $(\nu_1'-\nu_2')^+$ and $(\nu_1'-\nu_2')^-$ into probabilities, we have two probabilities ν_1 and ν_2. For any $g\in C$, the resultants of ν_1' and ν_2' are different iff the resultants of ν_1 and ν_2 are different. Further, there exists a measurable $A\subset\Theta$ such that $\nu_1(A)=1$ and $\nu_2(A)=0$.
   Depending on whether $y_k=0$ or $y_k=1$, for each $\theta\in\Theta$, $f_y(\theta)$ puts mass $1/2^k$ on one of the two disjoint box subsets of G_k, $g_k^0(\Theta)$ or $g_k^1(\Theta)$. This means that the probability that the resultants of ν_1 and ν_2 under $f_y(\theta)+h(\theta)$ agree on the box G_k is (at most) 1/2 because ν_1 and ν_2 are supported on disjoint sets. Since the Y_k are iid Bernoulli's, the probability that the resultants agree on K boxes is $1/2^K$. Thus, the probability that the resultants agree is less than or equal to $1/2^K$ for arbitrary K.

Suppose now that $k>\ell$. In the previous proof, redefine $g_k^0(\theta)$ as $x_k+(s_k/3)(\theta,\mathbf{0})$ and $g_k^1(\theta)$ as $(x_k+(2s_k/3))+(s_k/3)(\theta,\mathbf{0})$ where $(\theta,\mathbf{0})\in\mathbb{R}^k$ is vector having θ in its first ℓ components and 0 in the last $(k-\ell)$ components. The rest of the argument proceeds without change. □

The technique of proof for Theorem C used a distribution over smooth mappings from Θ to $\Delta(\mathbb{R}^k)$. This means that the genericity result also holds within the class of smooth parametrizations.

There are some reasons to believe that the condition $k\geq\ell$ cannot be avoided. One can show that if there is a compact $K\subset\mathbb{R}^k$ such that $f(\theta)(K)=1$ for all $\theta\in\Theta$ and $f(\Theta)$ determines continuous linear functions on $\Delta(K)$, then K must be homeomorphic to Θ. This means that the condition that $k\geq\ell$ must be tight in some cases. Combined with the following example, this leads us to conjecture that the set of continuous parametrizations having the Bauer property is prevalent if and only if $k\geq\ell$.

Example 8. Suppose that Θ is a compact subset of $\mathbb{R}^2$ with non-empty interior and consider the set of uniform distributions $\{U[a,b]:(a,b)\in\Theta\}$. Since the interior of Θ is non-empty, there exists $a<b<c<d$ such that $(a,c),(a,d),(b,c),(b,d)\in\Theta$. For appropriate choice of $\alpha\in(0,1)$, the resultants of $(1/2)U[a,d]+(1/2)U[b,c]$ and $\alpha U[a,c]+(1-\alpha)U[b,d]$ are equal and $(a,b)\mapsto U[a,b]$ does not have the Bauer property.

4.2. Specific verifications

We now verify that many well-known classes of parametrized distributions have the Bauer property. The following elementary Lemma will be useful.

Lemma 1. *If $K \subset \Delta(M)$ is compact and the mapping from $\eta \in \Delta(K)$ to the resultant $\mathbf{x}_\eta \in \overline{\text{co}}(K)$ is 1-to-1, then $\overline{\text{co}}(K)$ is a Bauer simplex and K its compact set of extreme points.*

Proof. The set of resultants, $\{\mathbf{x}_\eta : \eta \in \Delta(K)\}$ is exactly $\overline{\text{co}}(K)$. We need to show that $K = \text{extr}(\overline{\text{co}}(K))$. From this and the compactness of K, the 1-to-1 property is the definition of $\overline{\text{co}}(K)$ being a Bauer simplex. Pick $\mathbf{x} \in K$. Since the resultant mapping is 1-to-1 and the resultant of $\delta_\mathbf{x}$ is $\mathbf{x}$, the unique $\eta \in \Delta(K)$ having resultant $\mathbf{x}$ is $\delta_\mathbf{x}$. By [Phelps, 2001, Lemma 1.4, p. 6], this means that $\mathbf{x}$ is an extreme point of $\overline{\text{co}}(K)$. □

A simple example[11] contains the intuition for how the 1-to-1 condition for the resultant mapping works. Suppose that $\Theta = \{1, 2, 3\}$, and that $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$ are distinct, as would be required by $\theta \leftrightarrow \mathbf{p}_\theta$ being a homeomorphism. Problems would arise if $\mathbf{p}_2 = \alpha\mathbf{p}_1 + (1-\alpha)\mathbf{p}_3$. However, this could only happen if $\mathbf{p}_2$ were the resultant of both $\delta_{\mathbf{p}_2}$ and of $\alpha\delta_{\mathbf{p}_1} + (1-\alpha)\delta_{\mathbf{p}_3}$.

We will use the following two pieces of notation: associated with a $\mathbb{R}$-valued random variable X is a **location-scale** class of distributions, $\{\mathbf{p}_{r,\sigma} : r \in \mathbb{R}, \sigma \in \mathbb{R}_+\}$, where $\mathbf{p}_{r,\sigma}$ is the distribution of $r + \sigma X$; and given $\mathbf{p} \in \Delta(\mathbb{R})$ and $k \geq 1$, $\perp^k(\mathbf{p}) \in \Delta(\mathbb{R}^k)$ is the joint distribution of the random vector $(X_1, \ldots, X_k)$ where the $X_i \in \mathbb{R}$ are i.i.d. with distribution $\mathbf{p}$.

Theorem D. *The following classes $\{\mathbf{p}_\theta : \theta \in \Theta\}$ have the Bauer property:*

(1) $\{\mathbf{p}_{r,\sigma} : \sigma \in S\}$, *S a compact subset of $\mathbb{R}_+$ where X is non-degenerate and has finite moments of all orders, or where X is a standard Cauchy distribution*;
(2) $\{\mathbf{p}_{r,\sigma} : r \in R\}$, *R a compact subset of $\mathbb{R}$ where X is non-degenerate and has finite moments of all orders, or where X is a standard Cauchy distribution*;
(3) $\{\perp^2(\mathbf{p}_{\mu,\sigma}) : (\mu, \sigma) \in \Theta\}$, *$\Theta$ a compact subset of $\mathbb{R} \times \mathbb{R}_{++}$ where $\mathbf{p}_{\mu,\sigma}$ is the Gaussian distribution on $\mathbb{R}$ with mean μ and variance σ^2*;
(4) $\{\mathbf{g}_q : q \in \Theta\}$ *where Θ is a compact subset of $(0, 1)$ where $\mathbf{g}_q$ is the geometric distribution with parameter q*, i.e. $\mathbf{g}_q(\{k\}) = (1-q)^k q$ *for $k = 0, 1, \ldots$; and*
(5) $\{\mathbf{P}_\lambda : \lambda \in \Theta\}$ *for Θ a compact subset of $(0, \infty)$ where $\mathbf{P}_\lambda$ is the Poisson distribution with parameter λ*, i.e. $\mathbf{P}_\lambda(\{k\}) = e^{-\lambda}(\lambda^k/k!)$ *for $k = 0, 1, \ldots$.*

Several comments are in order.

i. By appropriate choice of the distribution of X, Theorem D(1) covers compact classes of uniform distributions that have one end fixed, e.g. $\{U[0, \theta] : \theta \in S\}$. It also covers compact classes of exponential distributions, as well as compact classes of Erlang, χ^2, Gamma and Weibull distributions with fixed shape parameters.
ii. The presence of the Cauchy distribution in Theorem D(1) and (2) indicates that it is not the moment conditions that are crucial for the Bauer property.
iii. If K is a compact subset of $\{\perp^2(\mathbf{p}) : \mathbf{p} \in \Delta(\mathbb{R})\}$, then every $\mathbf{q} \in K$ is an extreme point of $\overline{\text{co}}(K)$. This is the first part of proving the Bauer property.
iv. If $\varphi : M \leftrightarrow M$ is a homeomorphism, then $\{\mathbf{p}_\theta : \theta \in \Theta\}$ has the Bauer property if and only if $\{\mathbf{p}_\theta\varphi^{-1} : \theta \in \Theta\}$ does.

Proof. In each of the cases, $\theta \leftrightarrow \mathbf{p}_\theta$ is a homeomorphism, and the arguments for the 1-to-1 property of Lemma 1 have the same outline. First, suppose that η is a distribution on $\{\mathbf{p}_\theta : \theta \in \Theta\} \subset \Delta(M)$ and $\mathbf{x}_\eta \in \Delta(M)$ is its resultant. We have

$$\mathbf{x}_\eta(E) = \int_\Theta \mathbf{p}_\theta(E)\, d\eta(\theta) = \int_\Theta \left[\int_M 1_E(m)\, d\mathbf{p}_\theta(m) \right] d\eta(\theta), \tag{12}$$

which means that for any measurable $f : M \to \mathbb{R}$,

$$\int_M f(m)\, d\mathbf{x}_\eta(m) = \int_\Theta \left[\int_M f(m)\, d\mathbf{p}_\theta(m) \right] d\eta(\theta). \tag{13}$$

Denote the middle integral, $[\int_M f(m)\, d\mathbf{p}_\theta(m)]$, by $M_f(\theta)$. Let $\eta_1 \neq \eta_2$ be two distributions on Θ, equivalently, on Θ's homeomorphic image, $\{\mathbf{p}_\theta : \theta \in \Theta\}$. If they have the same resultant, $\mathbf{x}_{\eta_1} = \mathbf{x}_{\eta_2}$, then for all f,

$$\int_\Theta M_f(\theta)\, d\eta_1(\theta) = \int_\Theta M_f(\theta)\, d\eta_2(\theta). \tag{14}$$

The arguments will depend on showing that the set of M_f is sufficiently rich that (14) can be satisfied for all f iff $\eta_1 = \eta_2$. We do this in more detail for (1) than for the subsequent claims.

(1) We discuss the finite moments case before the Cauchy case.

[11] Suggested by a referee.

(a) Suppose that X has moments of all orders and that $\eta_1, \eta_2 \in \Delta(S)$ are two probabilities having the same resultant, $\mathbf{x}$. We will show that $\eta_1 = \eta_2$. Consider the moments of $\mathbf{x}$, i.e. $\int f(m)\, d\mathbf{x}(m)$ where $f(m) = m^k$, $k = 0, 1, 2, \ldots$. These are finite because X has finite moments of all orders and S is compact, hence bounded. Since $\mathbf{x} = \mathbf{x}_{\eta_1} = \mathbf{x}_{\eta_2}$,

$$\int_S \left[\int m^{2k}\, d\mathbf{p}_{0,\sigma}(m) \right] d(\eta_1 - \eta_2)(\sigma) = 0 \tag{15}$$

for $k = 0, 1, 2, \ldots$. For $k = 0$, this yields $\int_S 1\, d(\eta_1 - \eta_2)(\sigma) = 0$. For $k \geq 1$, this yields

$$EX^{2k} \cdot \int_S \sigma^{2k}\, d(\eta_1 - \eta_2)(\sigma) = 0 \tag{16}$$

where $EX^{2k} > 0$ for $k = 0, 1, \ldots$ because X is non-degenerate. The span of the functions $\sigma \mapsto \sigma^{2k}$ contains all polynomials in σ^2. Since $S \subset \mathbb{R}_+$, this is an algebra in $C(S)$ containing the constants and separating points. By the Stone–Weierstrass theorem, it is dense in $C(S)$. By continuity, this implies that for all $f \in C(S)$, $\int_S f(\sigma)\, d(\eta_1 - \eta_2)(\sigma) = 0$, i.e. $\eta_1 = \eta_2$.

(b) Suppose now that X is a standard Cauchy distribution. As above, it is sufficient to prove the result with $r = 0$. The characteristic function of $\mathbf{p}_{0,\sigma}$ is $e^{-|\sigma t|}$. The span of the functions $\sigma \mapsto e^{-|\sigma t|}$ contains an algebra containing the constants and separating points in $S \subset \mathbb{R}_+$. As above, this implies that if two probabilities on $\{\mathbf{p}_{0,\sigma} : \sigma \in S\}$ have the same resultant, then they are equal.

(2) An inductive variant of previous argument in the finite moments case shows that $\int_R r^k\, d(\eta_1 - \eta_2)(r) = 0$, $k = 0, 1, \ldots$, which means that $\eta_1 = \eta_2$ because R is compact. In the Cauchy case, as a function of r, the span of the characteristic functions contains an algebra containing the constants and separating points in $R \subset \mathbb{R}$.

(3) This follows from analytic continuation arguments for characteristic functions given in [Barndorff-Nielsen, 1965, Corollary 4].

(4) The probability generating function[12] for a geometric distribution $\mathbf{g}_q$ is $G_q(s) = q/s(1-q)$. For Θ a compact subset of $(0, 1)$, the closure of the span of the functions $q \mapsto G_q(s)$ in $C(\Theta)$ contains its own k th derivatives with respect to q for all k. Evaluating these at $s = 1$, we find that the closed span contains $1/(1-q)^{2k}$ for $k = 0, 1, \ldots$. Therefore the closed span contains an algebra containing the constants and separating points in $C(\Theta)$.

(5) The probability generating function for a Poisson distribution $\mathbf{P}_\lambda$ is $G_\lambda(s) = e^{\lambda(s-1)}$. For Θ a compact subset of $(0, \infty)$, the closure of the span of the functions $\lambda \mapsto G_\lambda(s)$ in $C(\Theta)$ contains its own k th derivatives with respect λ for all k. Evaluating these at $s = 0$, we find that the closed span contains λ^k for $k = 0, 1, \ldots$, hence is dense in $C(\Theta)$. □

5. Sketches of applications

We believe that the results here are likely to be useful in a variety of mechanism design problems. However, the agent/expert is often interested in the action that the principal takes. As Kadane and Winkler (1988) emphasize, for proper scoring rules to work, the expert should have no stake in the principal's action. Our first Proposition shows that the ability to control the value function sometimes allows us to circumvent this problem.

Our second Proposition is a corollary of our location-scale result, Theorem D(1). We show that it is possible to strictly elicit both the location and the scale of a location scale class while implementing any continuous function of the scale as the value to the expert.

5.1. Circumventing the no stakes condition

Suppose that the principal wishes to learn the expert's beliefs, $\mathbf{p} \in \Theta$, $\Theta \subset \mathcal{D}$, because they will solve a problem of the form

$$\max_{a \in A} \int g(a, x)\, d\mathbf{p}(x), \tag{17}$$

where A is compact, g is bounded, and jointly continuous in a and x. Let $a^*(\mathbf{p})$ denote the solution to the problem in (17). We suppose that the expert's utility is of the form $u(h, a)$ where h is the value of the scoring rule and a is the principal's choice of action. Given $\mathbf{p} \mapsto a^*(\mathbf{p})$, the expert solves the following variant of (1),

$$\max_{\mathbf{r} \in \Theta} E^{\mathbf{p}} u(h(\mathbf{r}, X), a^*(\mathbf{r})), \tag{18}$$

and if $\{\mathbf{p}\} = \operatorname{argmax}_{\mathbf{r} \in \Theta} E^{\mathbf{p}} u(h(\mathbf{r}, X), a^*(\mathbf{r}))$, we say that the scoring rule is **strictly Θ-proper with stakes**.

---

[12] A probability distribution on the integers can be regarded as a point in ℓ_1. For all $s \in [0, 1]$, the point $\mathbf{y}(s) := (s^0, s^1, s^2, s^3, \ldots)$ belongs to ℓ_∞. The probability generating function for $\mathbf{p}$ is the function $G_{\mathbf{p}}(s) = \langle \mathbf{p}, \mathbf{y}(s) \rangle = \sum_k \mathbf{p}(\{k\}) s^k$. The mapping $\mathbf{p} \leftrightarrow G_{\mathbf{p}}$ is continuous, linear, and invertible.

Definition 4. The expert satisfies the **no stakes condition** if for all values h of the scoring rule and for all $a, a' \in A$, the expert's utility satisfies $u(h, a) = u(h, a')$. The expert's utility is **continuous and quasi-linear** if it is of the form $u(h, a) = h + w(a)$ for some continuous $w : A \to \mathbb{R}$.

The only quasi-linear preferences satisfying the no stakes condition have $w(a) = w(a')$ for all $a, a' \in A$. Our interest focuses on continuous and quasi-linear violations of the no stakes condition.

Proposition 1. *If the expert's utility is continuous and quasi-linear, $\mathbf{p} \mapsto a^*(\cdot)$ is singleton-valued on Θ, and Θ satisfies the Bauer simplex condition, then for any continuous $v : \Theta \to \mathbb{R}$, there exists a scoring rule that is strictly Θ-proper with stakes for which the value function is v.*

Proof. By the usual properties of uhc correspondences (e.g. [Corbae et al., 2009, Section 6.1]), if $a^*(\cdot)$ is singleton valued, then $\mathbf{p} \mapsto w(a^*(\mathbf{p}))$ is continuous. By Theorem B(1), there exists a strictly convex $V : \mathcal{D} \to \mathbb{R}$ for which the value function is $v(\mathbf{p}) - w(a^*(\mathbf{p}))$. □

Note that this design problem requires that the principal know the expert's tradeoffs between h and a, and that the utility difference between h and h' should not depend on a.

5.2. Eliciting both location and scale

Suppose that the random variable X to be forecast belongs to a non-degenerate location-scale class having moments of all orders. A quadratic loss scoring rule elicits the location, and one can interpret the associated value, a multiple of σ^2, as a 'penalty.' The next result shows that one can cancel the penalty.

Proposition 2. *Suppose that X is a non-degenerate, mean 0 random variable having finite moments of all orders, and for all $(\mu, \sigma) \in R \times S \subset \mathbb{R} \times \mathbb{R}_+$, let $\mathbf{p}_{\mu,\sigma}$ denote the distribution of $\mu + \sigma X$. If $R \times S$ is compact, then for all continuous $v : S \to \mathbb{R}$, there exists a strictly Θ-proper scoring rule for $\Theta = \{\mathbf{p}_{\mu,\sigma} : (\mu, \sigma) \in R \times S\}$ with for which the value function is v.*

Proof. Let $\mathcal{D}$ denote the set of distributions on $\mathbb{R}$ having finite moments of all orders. For each $\mathbf{r} \in \Theta$, let $m(\mathbf{r})$ denote the mean of any random variable having distribution $\mathbf{r}$, and let $\sigma(\mathbf{r})$ denote the scale. Let $h_1(\mathbf{r}, x) = -(m(\mathbf{r}) - x)^2$. By Theorems B(1) and D(1), there exists a scoring rule $h_2(\mathbf{r}, x)$ for which reporting $\mathbf{r}$ is strictly Θ-proper and for which the associated value function is $v(\sigma) + \sigma^2$. Consider the value function $h(\mathbf{r}, x) = h_1(\mathbf{r}, x) + h_2(\mathbf{r}, x)$. □

In Fang et al. (2009), we show how to use this result to recruit sets of experts whose forecasts should optimally be weighted by their truthfully reported scale/precision, and we explicitly give scoring rules in terms of the value function to be implemented.

References

Allen, F., 1987. Discovering personal probabilities when utility functions are unknown. Manage. Sci. 33 (4), 542–544.
Anderson, R.M., Zame, W.R., 2001. Genericity with infinitely many parameters. Adv. Theor. Econ. 1, Art. 1, 64 pp. (electronic).
Barndorff-Nielsen, O., 1965. Identifiability of mixtures of exponential families. J. Math. Anal. Appl. 12, 115–121.
Bauer, H., 1963. Kennzeichnung kompakter Simplexe mit abgeschlossener Extremalpunktmenge. Arch. Math. 14, 415–421.
Bickel, J.E., 2007. Some comparisons among quadratic, spherical, and logarithmic scoring rules. Decis. Anal. 4 (2), 49–65.
Blackwell, D., 1951. Comparison of experiments. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, Berkeley and Los Angeles, 1951. University of California Press, pp. 93–102.
Blackwell, D., 1953. Equivalent comparisons of experiments. Ann. Math. Stat. 24, 265–272.
Brier, G., 1950. Verification of forecasts expressed in terms of probabilities. Monthly Weather Rev. 78 (1), 1–3.
Choquet, G., 1969. In: Marsden, J., Lance, T., Gelbart, S. (Eds.), Lectures on Analysis. Vol. II: Representation Theory. W.A. Benjamin, Inc., New York/Amsterdam.
Corbae, P.D., Stinchcombe, M.B., Zeeman, J., 2009. An Introduction to Mathematical Analysis for Economic Theory and Econometrics. Princeton University Press, Princeton, NJ.
Dawid, A.P., 2007. The geometry of proper scoring rules. Ann. Inst. Stat. Math. 59 (1), 77–93.
Fang, F., Stinchcombe, M.B., Whinston, A.B., 2009. Acquiring and aggregating costly information from sources of differing quality. Working paper, Dept. Econ., U.T. Austin.
Friedman, D., 1983. Effective scoring rules for probabilistic forecasts. Manage. Sci. 29, 447–454.
Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102 (477), 359–378.
Good, I.J., 1952. Rational decisions. J. Roy. Stat. Soc. Ser. B 14, 107–114.
Goodearl, K.R., 1986. Partially Ordered Abelian Groups with Interpolation, vol. 20 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI.
Hendrickson, A.D., Buehler, R.J., 1971. Proper scores for probability forecasters. Ann. Math. Stat. 42, 1916–1921.
Kadane, J.B., Winkler, R.L., 1988. Separating probability elicitation from utilities. J. Am. Stat. Assoc. 83 (402), 357–363.
Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic Theory. Oxford University Press, New York.
Matheson, J.E., Winkler, R.L., 1976. Scoring rules for continuous probability distributions. Manage. Sci. 22, 1087–1096.
McCarthy, J., 1956. Measures of the value of information. Proc. Natl. Acad. Sci. U.S.A. 42 (9), 654–655.
Phelps, R.R., 2001. Lectures on Choquet's Theorem, vol. 1757 of Lecture Notes in Mathematics, 2nd ed. Springer-Verlag, Berlin.
Roby, T.B., 1965. Belief states and the use of evidence. Behav. Sci. 10 (3), 255–270.
Savage, L.J., 1971. Elicitation of personal probabilities and expectations. J. Am. Stat. Assoc. 66, 783–801.
Schneider, R., 1993. Convex Bodies: The Brunn–Minkowski Theory, vol. 44 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge.