*The Annals of Mathematical Statistics*
1971, Vol. 42, No. 6, 1916 1921

# PROPER SCORES FOR PROBABILITY FORECASTERS[1]

BY ARLO D. HENDRICKSON AND ROBERT J. BUEHLER

*Virginia Polytechnic Institute and University of Minnesota*

A probability forecaster is asked to give a density $p$ of a random variable $\omega$. In return he gets a reward (or score) depending on $p$ and on a subsequently observed value of $\omega$. A scoring rule is called proper if the expected score is maximized when the true density is chosen. The present paper uses convex analysis to generalize McCarthy's characterization of proper scoring rules.

**1. Introduction and summary.** Let $(\Omega, \mathscr{A}, \mu)$ be a measure space and let $\mathscr{P}$ be a convex class of probability densities with respect to the measure $\mu$. A scoring rule $f$ is a mapping from $\mathscr{P}$ into the class $\mathscr{L}$ of random variables on $\Omega$. Assume a forecaster has knowledge of a probability density $p \in \mathscr{P}$, and is to receive the score (or actual payment) $f(p)$ for his disclosure of $p$. Since $f(p)$ is a random variable, the score depends on the outcome of the experiment $\omega \in \Omega$. The score $f$ has been called *proper* if

$$E_p(f(p)) \geqq E_p(f(q)) \qquad \text{for all } p, q \in \mathscr{P} \tag{1}$$

where $E_p(\cdot)$ is the mathematical expectation with respect to the density $p$. If (1) holds, then the forecaster will maximize his expected score with respect to $p$ by disclosing this density $p$. To avoid difficulties in (1) we will assume $E_p(f(q))$ exists and is finite.

The first suggested use of a scoring rule was apparently by Brier (1950) in connection with weather forecasting. The independent work of Good (1952) explicitly considered condition (1). For more recent work, see for example de Finetti (1962), Winkler (1969), Savage (1970), and Staël von Holstein (1970). The latter notes that scoring rules have also been called payoff or reward or incentive functions, and gives an excellent bibliography.

While the above context in which $p$ is known to the forecaster is adequate for our purposes, other points of view are possible. For example the cumulative score of any forecaster may be used as a measure of his forecasting ability; or the stated $p$ may be regarded as defining a subjective probability. But the present paper is concerned only with the purely mathematical problems connected with the characterization of functions $f$ satisfying (1) (or strictly satisfying (1)). Our main result is Theorem 3.1, which modifies and generalizes a theorem of McCarthy (1956), using a generalization of Rockafellar's (1970) definition of subgradient.

Theorems 4.2, 4.3, and 4.4 give additional conditions to ensure that there exists an $f$ satisfying the requirements of Theorem 3.1. The necessary preliminary definitions and theorems are given in Section 2.

Received January 29, 1971.

[1] Research supported by NSF Grant GP–9556.

**2. Some concepts of convex analysis.** The space $\mathscr{L}$ of random variables on $(\Omega, \mathscr{A})$ is a vector space with an inner product defined whenever it exists by

$$p \cdot q = \int p(\omega)q(\omega)d\mu(\omega). \tag{2}$$

Let $\mathscr{L}_1 = \mathscr{L}_1(\mathscr{P})$ be the set of all $q \in \mathscr{L}$ such that $p \cdot q$ is defined for all $p \in \mathscr{P}$. The range of a scoring rule $f$ defined on $\mathscr{P}$ is assumed to be contained in $\mathscr{L}_1(\mathscr{P})$. The following relation is crucial for applying convex analysis to studying (1):

$$E_p(q) = p \cdot q \qquad \text{if } p \in \mathscr{P}. \tag{3}$$

For given $f$, the expected score $H$ is defined on $\mathscr{P}$ by

$$H(p) = p \cdot f(p), \tag{4}$$

and condition (1) is equivalent to

$$H(p) \geqq p \cdot f(q) \qquad \text{for all } p, q \in \mathscr{P}. \tag{5}$$

The condition that $f$ be *strictly* proper is

$$H(p) > p \cdot f(q) \qquad \text{if } p \neq q. \tag{6}$$

The following is a generalization of Rockafellar's (1970) definition of subgradient to the infinite-dimensional case:

DEFINITION 2.1. If $H$ is defined on a convex set $D \subset \mathscr{L}$ and if there exists $q \in D$ and $q^* \in \mathscr{L}_1(D)$ such that

$$H(p) \geqq (p-q) \cdot q^* + H(q) \qquad \text{for all } p \in D, \tag{7}$$

then $q^*$ is a *subgradient* of $H$ at $q$ (relative to $D$).

It can be shown in the Euclidean case that a subgradient of a convex function $H$ is unique and equal to the gradient at every point where $H$ is differentiable (Rockafellar (1970), Theorem 25.1).

THEOREM 2.1. *If $H$ has a subgradient $q^*$ at each point $q$ in a convex set $D$, then $H$ is convex on $D$.*

PROOF. For any $p, q \in D$, let $p_1{}^*$ be a subgradient of $H$ at $p_1 = (1-\lambda)p + \lambda q$. Then $H(p) \geqq (p-p_1) \cdot p_1{}^* + H(p_1)$ and $H(q) \geqq (q-p_1) \cdot p_1{}^* + H(p_1)$. It follows that $(1-\lambda)H(p) + \lambda H(q) \geqq H(p_1)$.

The following is a variant of Euler's theorem.

THEOREM 2.2. *If $H$ is homogeneous of degree $r$ on a convex cone $D$ and has a subgradient $q^* \in \mathscr{L}_1(D)$ for some $q \in D$, then $rH(q) = q \cdot q^*$.*

PROOF. The inequality $\lambda^r H(q) = H(\lambda q) \geqq (\lambda q - q) \cdot q^* + H(q)$ implies

$$\frac{\lambda^r - 1}{\lambda - 1} H(q) \geqq q \cdot q^* \qquad \text{if } \lambda > 1,$$

with the reverse inequality if $0 < \lambda < 1$. Thus $rH(q) = q \cdot q^*$.

It can be shown further that if $q^*$ is a subgradient at $q$ of $H$ in Theorem 2.2, then $(\lambda q)^* = \lambda^{r-1}q^*$ is a subgradient of $H$ at $\lambda q$ for all $\lambda > 0$.

In the sequel we will use the term "homogeneous" to mean "homogeneous of the first degree." If $H$ is homogeneous on a convex set $C$, then by letting $H(\lambda p) = \lambda H(p)$ we can always extend the domain of $H$ to the convex cone $D = \{\lambda p : p \in C, \lambda > 0\}$.

**3. McCarthy's Theorem.** McCarthy (1956) stated without proof a characterization of proper scoring rules for the case when $\mathscr{P}$ is the class of discrete distributions on a finite set $\Omega$. Our Theorem 3.1 applies to more general $\mathscr{P}$ and distinguishes between strict and non-strict inequalities.

THEOREM 3.1. *A scoring rule $f$ mapping $\mathscr{P}$ into $\mathscr{L}_1$ satisfies* (1) [*strictly*] iff *there exists a function $H$ defined on $D = \{\lambda p : p \in \mathscr{P}, \lambda > 0\}$ which is* (a) *homogeneous,* (b) *convex* [*strictly convex on $\mathscr{P}$*], *and* (c) *such that $f(p)$ is a subgradient of $H$ relative to $D$ at $p$ for all $p \in \mathscr{P}$. The function $H$ satisfies $H(\lambda p) = \lambda p \cdot f(p)$.*

PROOF. Assuming (1) holds, define $H(\lambda p) = \lambda p \cdot f(p)$. Using (1),

$$H(\lambda p) \geqq \lambda p \cdot f(q) = (\lambda p - q) \cdot f(q) + H(q)$$

for all $p, q \in \mathscr{P}$, $\lambda > 0$, which establishes (c). Finally (b) follows from Theorem 2.1.

Conversely, (a), (b), (c) imply by Theorem 2.2 that $H(\lambda p) = \lambda p \cdot f(p)$, and substituting this into the subgradient inequality gives the desired result (1).

Strict inequality in (1) is equivalent to no subgradient of $H$ at $p$ being a subgradient of $H$ at $q$, if $p, q \in \mathscr{P}$, $p \neq q$. This is equivalent to $H$ being strictly convex on $\mathscr{P}$.

EXAMPLE 3.1. A familiar example is the logarithmic score suggested by Good (1952) for the binomial case. In the general case we put

$$f(p) = \log p. \tag{8}$$

A well-known inequality shows that $f$ is strictly proper. If $\mu$ is finite, then $H(\lambda p) = \lambda p \cdot f(p)$ is finite for $\lambda p \in \mathscr{L}_2{}^+$ where $\mathscr{L}_2{}^+ = \{q : q(\omega) \geqq 0 \text{ for all } \omega \in \Omega, \int q^2 d\mu < \infty\}$ and $H$ is continuous with respect to the $\mathscr{L}_2$ norm $\|\cdot\|$ defined by the inner product (2).

In the finite discrete case let the density $p(\omega)$ be replaced by a vector $p$ of probabilities $p_j$. For this $p \in \mathscr{P}$, $H(p) = \sum p_j \log p_j$, but for $q = \lambda p \in D$, $H(q) = \lambda \sum p_j \log p_j = \sum q_j \log (q_j / \sum q_k)$. Marschak (1960), page 97, attempted to show that the logarithmic score gave a counterexample to McCarthy's theorem, erroneously considering the gradient of $\sum p_j \log p_j$ rather than of $\sum q_j \log (q_j / \sum q_k)$. A proper understanding of the theorem requires a clear distinction between $\mathscr{P}$ and $D$ not explicitly stated in McCarthy's paper.

If we wish to define $f$ on $D$ as well as on $\mathscr{P}$, then a natural choice is $f(\lambda p) = f(p)$. In particular for the logarithmic case $f(q) = \log (q / \int q d\mu)$. Unlike $\log q$, this $f(q)$ is a subgradient of $H$ for all $q \in D$.

EXAMPLE 2.3. For $\mathscr{P} \subset \mathscr{L}_2(\mu)$, the "quadratic" score $f(p) = 2p - \|p\|^2$ (Brier (1950), de Finetti (1962)) is strictly proper. $H(\lambda p) = \lambda \|p\|^2$.

EXAMPLE 3.3. For $\mathscr{P} \subset \mathscr{L}_2(\mu)$, the "spherical" score $f(p) = p/\|p\|$ is strictly proper. $H(\lambda p) = \lambda \|p\|$.

**4. Expected score functions.** It might be asked what class of homogeneous and convex functions on $D$ satisfy the additional requirement of Theorem 3.1 of having subgradients relative to $D$ at each point in $\mathscr{P}$. The following is an example of a function which has no subgradients and yet is homogeneous and convex.

EXAMPLE 4.1. Let $\mathscr{P}$ be the class of continuous, bounded densities ($\sup p(\omega) < \infty$) on $(R, \mathscr{B}, \mu)$ where $\mu$ is Lebesgue measure and $\mathscr{B}$ consists of the Borel sets. Define $H(p) = \sup_\omega p(\omega)$. Then $H$ is clearly convex on $\mathscr{P}$. However, $H$ is neither continuous at any $p \in \mathscr{P}$ (with respect to $\|p\|$} nor does $H$ have a subgradient for any $p \in \mathscr{P}$.

Let $\mathscr{H} \subset \mathscr{L}$ be a Hilbert space, $R$ the real numbers, and let $\mathscr{H} \times R$ have the usual product topology and inner product. $\mathscr{H}$ can be taken to be the smallest closed subspace of $\mathscr{L}_2(\mu)$ containing $\mathscr{P}$, where $\mathscr{P} \subset \mathscr{L}_2(\mu)$. If $D \subset \mathscr{H}$ is the convex domain of a real-valued function $H$, then the epigraph of $H$, epi $(H) \subset \mathscr{H} \times R$, is the set $\{(p, \alpha) : \alpha \geqq H(p), p \in D\}$. $H$ is a convex function iff epi $(H)$ is a convex set.

The following is a partial converse to Theorem 2.1.

THEOREM 4.1. *Let $D$ be a convex set in $\mathscr{H}$ whose interior is nonempty. Let $H$ be a convex function on $D$ which is continuous at a point $p \in$ int $(D)$. Then $H$ has a subgradient $q^* \in \mathscr{H}$ at each point $q \in$ int $(D)$.*

PROOF. The assumptions imply epi $(H)$ is a convex subset of the Hilbert space $\mathscr{H} \times R$ whose interior is nonempty. If this is satisfied then epi $(H)$ has a closed hyperplane of support through each of its boundary points. (See for example Valentine (1964), Theorems 2.15 and 4.1.) The supporting hyperplane at the boundary point $(q, H(q))$ is seen to give one of the following inequalities for some $q^* \in \mathscr{H}$:

$$H(p) \geqq (p-q) \cdot q^* + H(q) \qquad \text{for all } p \in D, \tag{9}$$

or

$$q \cdot q^* \geqq p \cdot q^* \qquad \text{for all } p \in D. \tag{10}$$

Clearly, (10) is satisfied only if $q \in$ bdry $(D)$. Hence (9) is satisfied if $q \in$ int $(D)$.

THEOREM 4.2. *If the set of densities $D \subset \mathscr{H}$ is a convex set and if $H$ is convex and homogeneous on $D$ and continuous at a point $p$ in the interior of $D$, then there exists $f$ such that conditions* (4) *and* (5) *hold on the interior of $D$. The range of $f$ may be taken in $\mathscr{H}$.*

PROOF. Whenever $p \in \text{int}(D)$, apply Theorem 4.1, and let $f(p)$ be a subgradient of $H$ at $p$. The proof follows from Theorem 2.2.

The following theorems give equivalent conditions on $f$ for continuity conditions on $H$. We assume the range of $f$ is in $\mathscr{H}$.

THEOREM 4.3. *If $D$ is a convex cone in $\mathscr{H}$ whose interior is nonempty and if $H$ and $f$ satisfy* (4) *and* (5) *on $D$, then $H$ is continuous at $p \in \text{int}(D)$* iff *there exists a neighborhood of $p$ on which $\|f(\cdot)\|$ is bounded.*

PROOF. Let $p$, $p_n \in D$, $\|p_n - p\| \to 0$ as $n \to \infty$. Let $q_n = f(p_n)/\|f(p_n)\|^2$. Then $H(p_n+q_n) \geqq (p_n+q_n)\cdot f(p_n) = H(p_n)+1$. Thus, if $H$ is continuous at $p$, we cannot have $\|q_n\| \to 0$. Hence $\|f(\cdot)\|$ is bounded on a neighborhood of $p$.

Conversely, if $\|f(\cdot)\|$ is bounded on a neighborhood of $p$ then by the Cauchy-Schwarz inequality $(p_n - p)\cdot f(p_n) \to 0$ if $\|p_n - p\| \to 0$. This implies $\limsup H(p_n) = \limsup p\cdot f(p_n) \leqq H(p)$. Also $\liminf H(p_n) \geqq \lim p_n \cdot f(p) = H(p)$. Hence, if $\|p_n - p\| \to 0$ then $H(p_n) \to H(p)$.

We will now assume that $f$ is defined on the convex cone $D$ such that

$$f(\lambda p) = f(p) \qquad \text{if } p \in D,\ \lambda > 0. \tag{11}$$

This condition, although natural, is not necessary because the homogeneous function $H$ may have several subgradients at any $p \in D$.

COROLLARY 4.1. *If $f$ and $H$ satisfy* (4), (5), *and* (11) *on a convex cone $D \subset \mathscr{H}$, then $H$ is continuous on the interior of $D$* iff *$\|f(\cdot)\|$ is bounded on every closed set contained in $D$.*

PROOF. Since $f(\lambda p) = f(p)$ if $\lambda > 0$, $\|f(\cdot)\|$ is bounded on every closed set contained in $D$ is equivalent to $\|f(\cdot)\|$ bounded on every compact set in $D$, which is equivalent to the requirement of Theorem 4.3 that $\|f(\cdot)\|$ be locally bounded at each point $p \in \text{int}(D)$.

THEOREM 4.4. *If $H$ and $f$ satisfy conditions* (4), (5), *and* (11) *for all points in a Hilbert space $\mathscr{H}$, then the following are equivalent*:

(i) *$H$ is continuous*;

(ii) *$H$ is bounded on the sphere $\{p \in \mathscr{H} : \|p\| = 1\}$*;

(iii) *$\|f\|$ is bounded.*

PROOF. We need only show (ii) implies (iii). This follows from $H(f(q)/\|f(q)\|) \geqq (f(p)/\|f(q)\|)\cdot f(q) = \|f(q)\|$.

## REFERENCES

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

DE FINETTI, B. (1962). Does it make sense to speak of "good probability appraisers"? In Good (1962) 357–364.

GOOD, I. J. (1952). Rational decisions. *J. Royal Statist. Soc. Ser. B* **14** 357–364.

GOOD, I. J. ed. (1962). *The Scientist Speculates*. Basic Books, New York.

MCCARTHY, J. (1956). Measures of the value of information. *Proc. Nat. Acad. Sci.* **42** 654–655.

MARSCHAK, J. (1960). Remarks on the economics of information. *Contributions to Scientific Research in Management*. Univ. of California Press.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.

SAVAGE, L. J. (1970). The elicitation of personal probabilities and expectations. Unpublished.

STAËL VON HOLSTEIN, C.-A. S. (1970). *Assessment and Evaluation of Subjective Probability Distributions*. The Economic Research Institute at the Stockholm School of Economics, Stockholm.

VALENTINE, F. A. (1964). *Convex Sets*. McGraw-Hill, New York.

WINKLER, R. L. (1969). Scoring rules and the evaluation of probability assessors. *J. Amer. Statist. Assoc.* **64** 1073–1077.