Experimental Economics, 1:43–62 (1998)

Axiomatic Characterization of the Quadratic Scoring Rule

REINHARD SELTEN
University of Bonn, Laboratorium für experimentelle Wirtschaftsforschung, Adenauerallee 24-42, D-53113 Bonn
email: selten@lab.econ1.uni-bonn.de

Abstract

In the evaluation of experiments often the problem arises of how to compare the predictive success of competing probabilistic theories. The quadratic scoring rule can be used for this purpose. Originally, this rule was proposed as an incentive compatible elicitation method for probabilistic expert judgments. It is shown that up to a positive linear transformation, the quadratic scoring rule is characterized by four desirable properties.

Keywords: scoring rules, descriptive statistics

JEL Classification: C4

1. Introduction

Probabilistic theories for experimental decision situations need to be compared with respect to their predictive success. We may, for example, think of a learning experiment, in which in each of T periods $1, \ldots, T$ a subject has to choose an action. The set of available actions may not be the same one in every period; it may depend on previous history, and even the number of actions may be history dependent. Suppose we want to compare two learning theories that make probabilistic predictions for all choice situations that may arise in the course of an experiment. If we look at the actions of a specific subject in periods $1, \ldots, T$, we may ask which of both theories is more successful in the prediction of the subject's behavior.

If the theories to be compared were deterministic, one could just count the numbers of correct predictions. An easy generalization to the case of probabilistic predictions suggests itself: one forms the sum of all probabilities specified for observed choices and takes it as a measure of predictive success. This method has been referred to as the *linear scoring rule* (Stael von Holstein, 1970). The linear scoring rule has bad properties. Therefore several alternatives have been proposed in the literature (Brier, 1950; Roby, 1965; Toda, 1963; Winkler, 1969; Murphy and Winkler, 1970; Stael von Holstein, 1970; Matheson and Winkler, 1976; Friedman, 1983). One of these alternatives is the quadratic scoring rule. As far as the author knows, Brier (1950) was the first one who described this rule. He discussed it in the context of weather forecasting.

A scoring rule measures the predictive success of a theory for every period separately. For every period a score is computed that depends on the predicted probabilities and the actually observed action. Suppose that the true probabilities are known, and assume that at least two of them are positive and that one of the actions—say, action j—has a higher probability than all the other alternatives. As we show in Section 2, in this situation the linear scoring rule has the undesirable property that the highest expected score is not obtained by the correct probabilistic theory but by the wrong deterministic theory that predicts action j with certainty. This has been pointed out already by Brier (1950).

Incentive compatibilty—in the sense that the correct theory always is the only one that obtains the highest expected score—is a minimal requirement for a scoring rule. A scoring rule that is accepted as an instrument for the evaluation of competing probabilistic theories should not drive theoretical research in a wrong direction. It should provide incentives to search for a theory that comes as near to reality as possible.

The quadratic scoring rule is based on the idea that the score should reflect nearness of the predicted probability distribution to the observed outcome. In order to be able to measure a distance, the observation is interpreted as a frequency distribution: The relative frequency is 1 for the observed action and zero for every other alternative. The score is equal to 1 minus the squared distance between the predicted probability distribution and this relative frequency distribution. As we show in Section 2, the quadratic scoring rule is incentive compatible. This, too, has been pointed out by Brier (1950).

The quadratic scoring rule is not the only incentive-compatible one. Another example is the logarithmic scoring rule of Toda (1963). The logarithmic score is the logarithm of the predicted probability of the observed action or, in other words, the likelihood of the observation. The logarithmic scoring rule has a close connection to the maximum likelihood principle. However, in spite of this theoretical advantage, the logarithmic scoring rule is not really recommendable. On the one hand, it is too sensitive with respect to differences between very small probabilities and, on the other hand, it is sometimes not sensitive enough, in the sense that in some situations it does not matter whether the truth is near to the prediction or far from it.

In Sections 2.5 and 2.6 the undesirable properties of the logarithmic scoring rule are described in detail. In Section 2.7 we describe a whole family of incentive compatible scoring rules containing the quadratic scoring rule as a special case. Moreover, an additional rule first proposed by Roby (1965), the spherical scoring rule is described in Section 2.8.

In Section 4 it is shown that up to a positive linear transformation, the quadratic scoring rule is characterized by four axioms. These axioms are introduced and discussed in Section 3. Axiom 1 requires symmetry with respect to a renumbering of the actions. The second axiom concerns the consequences of adding a new action whose predicted probability is zero. It is required that this operation does not change the scores of old alternatives. However, nothing is said about the score for the new alternative.

The remaining two axioms are expressed in terms of the expected score loss of a predicted probability distribution $p = (p_1, \ldots, p_n)$ compared with predicting the true probability distribution $r = (r_1, \ldots, r_n)$. Incentive compatibility, required by axiom 3, can be expressed by saying that the expected score loss of p at r should always be positive for $p \neq r$.

Axiom 4 postulates that the expected score loss of p at q is equal to the expected score loss of q at p. This means that in a comparison between two theories p and q the mistake of predicting p if q is right is not judged to be more or less severe than the opposite mistake of predicting q if p is right. It seems to be natural to require that a reasonable scoring rule exhibits this kind of neutrality.

At the end of Section 4 the role of the axioms in the proof of the main result is discussed and the question is answered which of the scoring rules described in Section 2 satisfies which of the axioms. The paper ends with some concluding remarks in Section 5.

2. Preliminaries

In this section we first introduce some definitions and notational conventions. Then we look at specific scoring rules and examine whether they are incentive compatible or not.

2.1. Definitions and notational conventions

For $n = 1, 2, \ldots$ the symbol Δ_n denotes the set of all probability distributions $p = (p_1, \ldots, p_n)$ over the integers $1, \ldots, n$, where p_i stands for the probability of i. For mathematical reasons it is convenient not to exclude the trivial case $n = 1$. The distribution $p = (1)$ is the only element of Δ_1. We make use of the Kronecker symbol defined as follows for pairs of integers i and j:

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

The *ith unit n-vector* $\delta(i, n)$ is defined as follows:

$$\delta(i, n) = (\delta_{i1}, \ldots, \delta_{in}).$$

This vector has 1 at its ith place and zeros everywhere else. As has been explained in the introduction, $\delta(i, n)$ can be interpreted as the observed relative frequency distribution over the available actions $1, \ldots, n$, if action i has been taken.

For any two distributions $p = (p_1, \ldots, p_n)$ and $r = (r_1, \ldots, r_n)$ a *convex linear combination* of p and r is a distribution $q = (q_1, \ldots, q_n)$ with

$$q_i = (1 - \alpha)p_i + \alpha r_i \quad \text{for } i = 1, \ldots, n$$

where α is a real number with $0 \leq \alpha \leq 1$. We also express this relationship between r, p, and q by

$$q = (1 - \alpha)p + \alpha r.$$

2.1.1. Scoring rules. For $i = 1, 2, \ldots$ the union of all Δ_n with $n = i, i + 1, \ldots$ is denoted by Λ_i. A *scoring rule* is a sequence $S = S_1, S_2, \ldots$ of *scoring functions* S_i such that S_i assigns a score $S_i(p)$ to every $p \in \Lambda_i$. Here $S_i(p)$ is either a real number or $-\infty$. A scoring rule is *real valued*, if $S_i(p)$ is always a real number. (The general definition permits the

score $-\infty$ in order to cover the case of the logarithmic scoring rule.) Formally a scoring rule $S = S_1, S_2, \ldots$ also assigns a score. $S_1(p)$ to the only element $p = (1)$ of Δ_1, even if this is without practical significance.

2.1.2. Expected scores and expected score losses. The concepts of expected score and expected score loss will be first defined for real valued scoring rules. The definition will be extended to other scoring rules later. Let $S = S_1, S_2, \ldots$ be real valued. Let $r = (r_1, \ldots, r_n)$ and $p = (p_1, \ldots, p_n)$ be two probability distributions over $1, \ldots, n$. We interpret r as the true distribution and p as the predicted one. The *expected score* of p at r for a real valued scoring rule $S = S_1, S_2, \ldots$ is defined as follows:

$$V(p \mid r) = \sum_{i=1}^{n} r_i S_i(p).$$

For the sake of brevity the notation $V(p \mid r)$ does not express the dependence on the scoring rule. The *expected score loss* $L(p \mid r)$ of p at r is the difference

$$L(p \mid r) = V(r \mid r) - V(p \mid r).$$

2.1.3. Extension to general scoring rules. We now extend the definition of *expected score* and *expected score loss* to scoring rules which are not real valued. This extension will be based on the same formulas as in the case of the real valued scoring rules, complemented by conventions about the evaluation of terms involving $-\infty$. The following convention applies to the evaluation of the right-hand side of the equation for $V(p \mid r)$:

$$r_i S_i(p) = \begin{cases} -\infty & \text{for } r_i > 0 \text{ and } S_i(p) = -\infty \\ 0 & \text{for } r_i = 0 \text{ and } S_i(p) = -\infty \end{cases}$$

$V(p \mid r)$ has the value $-\infty$ if at least one of the terms $r_i S_i(p)$ is evaluated as $-\infty$. If all terms are real the expected score loss is the sum of all these terms. The convention for the evaluation of the right hand-side of the equation for $L(p \mid r)$ is as follows:

$$L(p \mid r) = \begin{cases} +\infty & \text{if } V(r \mid r) > -\infty \text{ and } V(p \mid r) = -\infty \\ -\infty & \text{if } V(r \mid r) = -\infty \text{ and } V(p \mid r) > -\infty \\ 0 & \text{if } V(r \mid r) = V(p \mid r) = -\infty \end{cases}$$

The formula for $L(p \mid r)$ is directly applied if $V(r \mid r)$ and $V(p \mid r)$ are both real.

2.1.4. Comment. The convention applied to $V(p \mid r)$ seems to be reasonable in view of

$$\lim_{s_i \to -\infty} r_i S_i = \begin{cases} -\infty & \text{for } r_i > 0 \\ 0 & \text{for } r_i = 0 \end{cases}$$

The first and the second line of the convention for $L(p \mid r)$ have an analogous straightforward interpretation. The third line is based on the idea that no expected score loss results by the transition from r to p or vice versa, if the expected score is $-\infty$ for both of them.

2.1.5. Incentive compatibility. A scoring rule $S_1, S_2, \ldots$ is called *incentive compatible* if for all $p, r \in \Delta_n$ with $p \neq r$ the expected score loss $L(p \mid r)$ for S is positive ($+\infty$ counts as positive).

Note that incentive compatibility does not exclude the possibility of two distributions $p, r \in \Delta_n$ with $V(r \mid r) > -\infty$ and $V(p \mid r) = -\infty$. In this case $L(p \mid r)$ is evaluated as $+\infty$ by the first line of the convention for $L(p \mid r)$. However, it is clear that the conditions of the second and the third lines are excluded by incentive compatibility.

2.2. The linear scoring rule

The scoring rule $S = S_1, S_2, \ldots$ with

$$S_i(p) = p_i$$

for all $p = (p_1, \ldots, p_n) \in \Lambda_i$ with $i = 1, 2, \ldots$ is called the *linear scoring rule*. In the following $S = S_1, S_2, \ldots$ denote this rule. We now show that the linear scoring rule is not incentive compatible.

Assume that in $r = (r_1, \ldots, r_n) \in \Lambda_i$ one of the probabilities r_i—say, r_j— is greater than each of the other ones but smaller than 1. Consider the distribution $\delta(j, n)$, which concentrates all the probability on j. We have

$$L(\delta(j, n) \mid r) = -r_j + \sum_{i=1}^{n} r_i^2 < -r_j + \sum_i r_i r_j = r_j \left[-1 + \sum_i r_i \right] = 0.$$

Therefore, the linear scoring rule fails to be incentive compatible. Under mild assumptions on the true distribution r, the deterministic theory that predicts the most probable action j with certainty achieves a higher expected score than the correct stochastic theory. It can also be seen without difficulty that this wrong deterministic theory maximizes the expected score.

2.3. The quadratic scoring rule

The scoring rule $S = S_1, S_2, \ldots$ with

$$S_i(p) = 1 - \sum_{j=1}^{n} (\delta_{ij} - p_j)^2$$

for every $p = (p_1, \ldots p_n) \in \Lambda_i$ and $i = 1, 2, \ldots$ is called the *quadratic* scoring rule. In the following $S = S_1, S_2, \ldots$ will denote the quadratic scoring rule. The definition determines the score of p as 1 minus the squared Euclidean distance of p and the ith unit n vector $\delta(i, n)$. It can be seen easily that the formula for $S_i(p)$ can be rewritten as follows:

$$S_i(p) = 2p_i - \sum_{j=1}^{n} p_j^2.$$

We now show that the quadratic scoring rule is incentive compatible. The second formula for $S_i(p)$ yields

$$V(p \mid r) = \sum_{i=1}^{n} 2r_i p_i - \sum_{j=1}^{n} p_j^2$$

$$V(p \mid r) = \sum_{i=1}^{n} r_i^2 - \sum_{i=1}^{n} (r_i - p_i)^2$$

$$L(p \mid r) = \sum_{i=1}^{n} (r_i - p_i)^2.$$

This shows that the quadratic scoring rule is incentive compatible. Moreover, the expected score loss is the squared Euclidean distance between r and p.

2.4. The logarithmic scoring rule

The scoring rule $S = S_1, S_2, \ldots$ with

$$S_i(p) = \ln p_i$$

for every $p = (p_1, \ldots, p_n) \in \Delta_i$ and $i = 1, 2, \ldots$ is called the *logarithmic* scoring rule. For $p_i = 0$, the right-hand side is interpreted as $-\infty$. In the following, $S = S_1, S_2, \ldots$ denotes the logarithmic scoring rule.

We now show that the logarithmic scoring rule is incentive compatible. For this purpose we prove that $L(p \mid r)$ has its maximum with respect to p at $p = r$ and nowhere else. Let $P(r)$ be the set of all $p \in \Delta_n$ with $p_i = 0$ if and only if $r_i = 0$. It will be shown that for every $p \in \Delta_n$ not in $P(r)$ we can find a $q \in \Delta_n$ such that the expected score of q is greater than that of p. In this way the task of maximizing the expected score $V(p \mid r)$ over Δ_n is reduced to that of maximizing it over $P(r)$.

Assume that $p \in \Delta_n$ does not belong to $P(r)$. First, consider the case that $p_i = 0$ holds for some i with $r_i > 0$. In this case the expected score is $-\infty$. The score becomes a real number if all p_i with $p_i = 0$ and $r_i > 0$ are slightly increased and some positive components of p are slightly decreased in a way which does not reduce them to zero and keeps the distribution within Δ_n. If the resulting distribution $q = (q_1, \ldots, q_n)$ does not belong to $P(r)$, it must have the property that for at least one i we have $q_i > 0$ and $r_i = 0$. In this case, we can increase the expected score by changing all these components of q to zero and by increasing all q_j with $q_j < r_j$ to an extent that keeps them below r_j. It is clear that this can be done in a way which results in a distribution belonging to $P(r)$. We have reduced the task of maximizing the expected score to the task of maximizing it over $P(r)$.

It can be seen immediately that within $P(r)$ the expected score $V(p \mid r)$ is a continuously differentiable and strictly concave function of all p_i with $r_i > 0$. Therefore, a $p \in P(r)$ at which the first-order conditions for an extremum of the expected score are satisfied with respect to these p_i must be a maximizer of $V(p \mid r)$ over $P(r)$. Moreover, there can be only one such maximizer. An easy Lagrange argument yields the following first-order conditions:

$$\frac{r_i}{p_i} = \lambda \quad \text{for all } p_i \text{ with } r_i > 0.$$

These conditions are satisfied at $p = r$ with $\lambda = 1$. It follows that the expected score $V(p \mid r)$ is maximized over $P(r)$ and therefore over Δ_n at $r = p$ and nowhere else. We can conclude that the logarithmic scoring rule is incentive compatible.

2.5. *An insensitivity property*

Consider a predicted distribution p and a true distribution r. Suppose that r is a convex linear combination of p and a distribution q different from p:

$$r = (1 - \alpha)q + \alpha p \quad \text{with } 0 \leq \alpha < 1$$

Obviously r is the nearer to p in any reasonable sense, the greater α is. Therefore, it seems to be a desirable property of a scoring rule that the expected score loss $L(p \mid r)$ is decreased by an increase of α. However, the logarithmic scoring rule is insensitive for some p, q, in the sense that $L(p \mid r)$ is not changed by an increase of α. We now formally state this insensitivity property.

Insensitivity property. Let $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_n)$ be two distributions in Δ_n where n is an integer greater than 1. Assume that for at least one j we have $q_j > 0$ and $p_j = 0$. Then we have

$$L(p \mid (1 - \alpha)\, q + \alpha p) = +\infty \quad \text{for } 0 \leq \alpha < 1.$$

We shall show that every incentive compatible scoring rule $S = S_1, S_2, \ldots$ with

$$S_j(p) = -\infty \quad \text{for } p_j = 0$$

has the insensitivity property. Let $S = S_1, S_2, \ldots$ be a scoring rule satisfying these conditions. Obviously the logarithmic scoring rule is a special case of such a scoring rule. Incentive compatibility means that we must always have

$$V(r \mid r) > V(p \mid r) \quad \text{for } p \neq r.$$

Obviously this cannot be true in the case $V(r \mid r) = -\infty$. Therefore the expected score with respect to S must satisfy the condition

$$-V(r \mid r) > -\infty \quad \text{for every } r.$$

Let p and q be as described in the insensitivity property and let j be an integer with $q_j > 0$ and $p_j = 0$. The jth component of $(1 - \alpha)q + \alpha p$ is $(1 - \alpha)q_j$ and therefore positive. In view of $S_j(p) = -\infty$ it follows that we have

$$V(p \mid (1 - \alpha)q + \alpha p)) = -\infty \quad \text{for } 0 \leq \alpha < 1.$$

This together with $V(r \mid r) > -\infty$ yields the insensitivity property.

Result. Let $S = S_1, S_2, \ldots$ be an incentive compatible scoring rule with $S_j(p) = -\infty$ in the case that $p_j = 0$ holds for the jth component of p. Then S has the insensitivity property.

Remark. The result implies that the logarithmic scoring rule has the insensitivity property.

2.6. *Hypersensitivity of the logarithmic scoring rule*

In this section, too, $S = S_1, S_2, \ldots$ stands for the logarithmic scoring rule. In the following we discuss a property of the logarithmic scoring rule called *hypersensitivity*. Roughly speaking, hypersensitivity means that the expected score reacts very strongly to small differences of small probabilities.

We use the notation $|r - p|$ for the Euclidic distance between two distributions r, $p \in \Delta_n$ with $n = 1, 2, \ldots$. The hypersensitivity property can be formally expressed as follows.

Hypersensitivity. *For* $n = 2, 3, \ldots$ *the following two assertions* (a) *and* (b) *hold*: (a) *let* $r, p \in \Delta_n$ *be two distributions with* $r_j > 0$ *and* $p_j = 0$ *for at least one* j; *then* $V(p \mid r) = -\infty$; (b) *for every* $\varepsilon > 0$ *and every* $M > 0$, *we can find* $r, p \in \Delta_n$ *with* $r_i > 0$ *and* $p_i > 0$ *for* $i = 1, \ldots, n$ *such that* $|r - p| < \varepsilon$ *and* $L(p \mid r) > M$.

We now show that the logarithmic scoring rule has this property. Assertion a is an immediate consequence of $S_i(p) = -\infty$ for $p_i = 0$. We turn our attention to b. Let $r = (r_1, \ldots, r_n)$ and let $p = (p_1, \ldots, p_n)$ be as follows:

$$r_1 = \varepsilon_1, r_i > 0 \quad \text{for } i = 2, \ldots, n$$
$$p_1 = \varepsilon_2,\ p_2 = r_2 + \varepsilon_1 - \varepsilon_2,\ p_i = r_i \quad \text{for } i = 3, \ldots, n$$

with

$$0 < \varepsilon_2 < \varepsilon_1 < \frac{\varepsilon}{2}.$$

It can be seen easily that $|r - p| < \varepsilon$ holds. We have

$$L(r \mid p) = \varepsilon_1 \ln \frac{\varepsilon_1}{\varepsilon_2} + r_2 \ln \frac{r_2}{r_2 + \varepsilon_1 - \varepsilon_2}.$$

For $\varepsilon_2 \to 0$, the expression on the right-hand side approaches the limit $+\infty$. This shows that b holds.

Comment. Part a of the hypersensitivity property shows that any theory that wrongly excludes an action as impossible has no chance to be judged to be better than any other theory that predicts positive probabilities for all actions, if sufficiently many data are collected. In this respect it makes no difference how improbable the excluded action is.

A theorist who is guided by the logarithmic scoring rule is well advised not to specify zero probabilities for very improbable actions. In this way one can protect oneself against the consequences of part a. However, in general, it will be very difficult to judge how small a very small probability should be. Usually there will be no good theoretical reasons to specify a probability as 10^{-5} rather than 10^{-10}. However, in view of part b of the hypersensitivity property, such differences can be of crucial importance for the comparison of the two theories. The example used for the proof of b illustrates this point.

The use of the logarithmic scoring rule implies the value judgment that small differences between small probabilities should be taken very seriously and that wrongly describing something extremely improbable as having zero probability is an unforgivable sin. The author thinks that this value judgment is unacceptable. Therefore, he looks at the hypersensitivity property as a very undesirable one.

2.7. *The power rule family*

Let α be a real number with $\alpha > 1$. Consider the scoring rule $S = S_1, S_2, \ldots$ with

$$S_i(p) = \alpha\, p_i^{\alpha-1} - (\alpha - 1) \sum_{j=1}^{n} p_j^{\alpha}$$

for every $p = (p_1, \ldots, p_n) \in \Lambda_i$ and for $i = 1, 2, \ldots$ We call S the α-power scoring rule. Obviously, one obtains the quadratic scoring rule as a special case for $\alpha = 2$.

We now show that every α-power scoring rule with $\alpha > 1$ is incentive compatible. Let $S = S_1, S_2, \ldots$ be one of these scoring rules. S is real valued. In order to prove incentive compatibility, it is sufficient to show that for $r \in \Lambda_n$ with $n = 1, 2, \ldots$ the expected score $V(p \mid r)$ is maximized over Λ_n with respect to p at $p = r$ and nowhere else. The expected score $V(p \mid r)$ for r, $p \in \Lambda_n$ can be written as a sum of n functions $f_i(p_i)$:

$$V(p \mid r) = \sum_{i=1}^{n} f_i(p)$$

with

$$f_i(p_i) = r_i \alpha\, p_i^{\alpha-1} - (\alpha - 1) p_i^{\alpha} \quad \text{for } i = 1, \ldots, n.$$

The derivatives of these functions are as follows:

$$f_i'(p_i) = \alpha(\alpha - 1)[r_i - p_i] p_i^{\alpha-2} \quad \text{for } i = 1, \ldots, n.$$

Obviously, $f_i'/(p_i)$ is positive for $0 < p_i < r_i$ and negative for $p_i > r_i$. This shows that $f_i(p_i)$ attains its maximum at $r_i = p_i$ and nowhere else. It follows that $V(p \mid r)$, too, attains its maximum at $r = p$ and nowhere else. Therefore every α-power scoring rule with $\alpha > 1$ is incentive compatible.

2.8. The spherical scoring rule

For $n = 1, 2, \ldots$ and every $p = (p_1, \ldots, p_n) \in \Lambda_n$, let

$$|p| = \sqrt{\sum_{i=1}^{n} p_i^2}$$

the *norm* of p. The scoring $S = S_1, S_2, \ldots$ with

$$S_i(p) = \frac{p^i}{|p|}$$

for $i = 1, \ldots, n$ and every $p \in \Lambda_i$ is called the *spherical scoring rule*. The name is due to the fact that the mapping from p to the score vector $(S_1(p), \ldots, S_n(p))$ transforms p to a point on the unit sphere by multiplying p by the factor $1/|p|$. The greater this factor is the more sensitive the spherical scores $S_i(p)$ are with respect to small changes of $|p|$. At the corners of Λ_n this factor $1/|p|$ is equal to 1. It attains its maximum within this simplex at the midpoint $(1/n, \ldots, 1/n)$ of Λ_n. There $1/|p|$ has the value $\sqrt{n}$. Unlike the logarithmic scores the spherical ones do not exhibit any hypersensitivity with respect to changes of very small probabilities. On the contrary, for fixed n the spherical scores are most sensitive near the center of Λ_n.

It will now be shown that the spherical scoring rule is incentive compatible. For this rule we have

$$L(p \mid r) = \frac{1}{|p|}\left[|r|\,|p| - \sum_{i=1}^{n} p_i r_i\right].$$

It is a well-known fact, called *Cauchy's inequality* (see, e.g., Ostrowski, 1965: 217) that

$$\left[\sum_{i=1}^{n} r_i p_i\right]^2 < \sum_{i=1}^{n} r_i^2 \sum_{i=1}^{n} p_i^2$$

holds for vectors $r = (r_1, \ldots, r_n)$ and $p = (p_1, \ldots, p_n)$ with nonnegative components unless they are proportional in the sense that for some $\lambda \neq 0$ we have $p = \lambda r$. Since in our case both r and p are probability distributions over $1, \ldots, n$ this means that the inequality holds for $p \neq r$. Consequently, the expected spherical score loss $L(p \mid r)$ at r is positive for every p with $p \neq r$. The spherical scoring rule is incentive compatible.

2.9. Summary of the results on specific scoring rules

It has been shown that the linear scoring rule is not incentive compatible. The quadratic scoring rule, the logarithmic scoring rule, all α-power scoring rules with $\alpha > 1$ and the

spherical scoring rule are incentive compatible. The quadratic scoring rule is the α-power rule with $\alpha = 2$.

Every incentive compatible scoring rule with $S_i(p) = -\infty$, if $p_i = 0$ holds for the ith component of p, has the insensitivity property of Section 2.5 in particular the logarithmic scoring rule has this property. Roughly speaking this property means that in some cases the expected score loss does not adequately respond to how far the true distribution is from the predicted one.

The logarithmic scoring rule also has the hypersensitivity property discussed in Section 2.6. This property results in too much weight given to small differences between small probabilities. The logarithmic scoring rule lacks sensitivity in some situations and is hypersensitive in others. For this reason the author thinks that the logarithmic scoring rule is not recommendable.

The power family shows that there are infinitely many incentive compatible scoring rules.

3. The axioms

In the following we first introduce some convenient notation. The formal statement of the four axioms follows. Finally, the intuitive justification of the axioms is discussed.

3.1. Notation

In the axioms $S = S_1, S_2, \ldots$ denotes an arbitrary scoring rule as defined in Section 2.1. Consider a permutation π of the numbers $1, \ldots, n$ or in other words a one-to-one function of $\{1, \ldots, n\}$ onto itself, which maps i to $\pi(i)$. For every permutation π of $1, \ldots, n$ we define a one-to-one mapping of Δ_n onto itself, for which we also use the symbol π. For every distribution $p = (p_1, \ldots, p_n)$ the π-*image* $q = (q_1, \ldots, q_2)$ is defined as follows:

$$q_{\pi(i)} = p_i \quad \text{for } i = 1, \ldots, n.$$

The π-image of p is denoted by $\pi(p)$. We now define an *elongation function* θ that maps a distribution $p = (p_1, \ldots, p_n) \in \Delta_n$ to a distribution $\theta(p) \in \Delta_{n+1}$ by adding zero as the $(n+1)$th component and changing nothing else:

$$\theta(p) = (p_1, \ldots, p_n, 0).$$

The symbols $V(p \mid r)$ and $L(p \mid r)$ will be used as before to denote expected score and expected score loss, respectively, for the scoring rule under consideration.

3.2. Axioms

Axiom 1 (symmetry) For every $n = 1, 2, \ldots$ and every permutation π of the numbers $1, \ldots, n$ and every $p \in \Delta_n$ we have

$$S_{\pi(i)}(\pi(p)) = S_i(p) \quad \text{for } i = 1, \ldots, n.$$

Axiom 2 (elongation invariance) For every $n = 1, 2, \ldots$ and every $p \in \Delta_n$ we have

$$S_i(\theta(p)) = S_i(p) \quad \text{for } i = 1, \ldots, n.$$

Axiom 3 (incentive compatibility) S is incentive compatible.

Axiom 4 (neutrality) For every $n = 1, 2, \ldots$ and any two $p \in \Delta_n$ and $q \in \Delta_n$ the expected score loss of p at q equals the expected score loss of q at p:

$$L(p \mid q) = L(q \mid p).$$

3.3. Discussion of the axioms

Scores should not depend on the numbering of the alternatives. Not more than this is expressed by axiom 1. Score function symmetry in this sense seems to be an indispensable requirement for a reasonable scoring rule.

Axiom 2 requires that scores for the alternatives $1, \ldots, n$ should not be influenced by a changed description of the same situation which differs from the original one only by the inclusion of an impossible $(n + 1)$th alternative. Blowing up p by an $(n + 1)$th zero component should not make any difference as far as the scores $S_1(p), \ldots, S_n(p)$ are concerned. This seems to be a reasonable requirement. Note that axiom 3 does not say anything about $S_{n+1}(\theta(p))$. If alternative $n + 1$ was actually observed, then the description of the situation as involving only n alternatives were wrong and the elongation were not just an irrelevant change of a correct description of the situation but a necessary correction of a wrong one. Nevertheless the case of a spurious change of the description by the addition of an impossible alternative should be treated correctly.

Incentive compatibility, required by axiom 3 already has been extensively discussed in the introduction. It is clear that incentive compatibility is an indispensable property of a scoring rule.

The interpretation of axiom 4 becomes clear if one looks at the hypothetical case that one and only one of two theories p and q is right, but it is not known which one. The expected score loss of the wrong theory is a measure of how far it is from the truth. It is only fair to require that this measure is "neutral" in the sense that it treats both theories equally. If p is wrong and q is right, then p should be considered to be as far from the truth as q in the opposite case that q is wrong and p is right.

A scoring rule should not be prejudiced in favor of one of both theories in the contest between p and q. The severity of the deviation between them should not be judged differently depending on which of them is true or false.

A scoring rule which is not neutral is discriminating on the basis of the location of the theories in the space of all probability distributions over the alternatives. Theories in some parts of this space are treated more favorably than those in some other parts without any justification. Therefore, the neutrality axiom 4 is a natural requirement to be imposed on a reasonable scoring rule.

4. Characterization of the quadratic scoring rule

We first show that the quadratic scoring rule satisfies axioms 1 to 4. Then we state our main result in Section 4.2. However, the proof of this theorem will be given only at the end of Section 4. Before this can be done it is necessary to derive some intermediate results.

4.1. The axioms are satisfied

We prove the following result.

Lemma 1. *The quadratic scoring rule satisfies axioms* 1 *to* 4.

Proof: It is clear by the definition of the quadratic scoring rule in Section 2.3 that axioms 1 and 2 are satisfied. There we also have seen that the quadratic scoring rule is incentive compatible. Axiom 3 is satisfied. As we have seen in Section 2.3 the expected score loss $L(p \mid r)$ with respect to the quadratic scoring rule is the squared Euclidean distance between p and r. Therefore axiom 4 is satisfied. □

4.2. Statement of the main result

Let $S = S_1, S_2, \ldots$ and $R = R_1, R_2, \ldots$ be two scoring rules. We say that R is a *positive linear transformation* of S if a positive number α and a real number β exists such that

$$R_i(p) = \alpha S_i(p) + \beta$$

holds for $i = 1, 2, \ldots$ and for every $p \in \Lambda_i$. Instead of $S_i(p)$ with $p = (p_1, \ldots, p_n)$ and $p \in \Lambda_i$. we also write $S_i(p_1, \ldots, p_n)$.

It is clear that R is a positive linear transformation of S if and only if S is a positive linear transformation of R. A *normed* scoring rule $S = S_1, S_2, \ldots$ is a scoring rule with the following properties:

$$S_1(0, 1) = -1$$
$$S_1(1, 0) = +1.$$

We now state our main result.

Theorem. *There is one and only one normed scoring rule which satisfies axioms* 1 *to* 4. *This is the quadratic scoring rule. A scoring rule satisfies axioms* 1 *to* 4 *if and only if it is a positive linear transformation of the quadratic scoring rule.*

Proof: The proof will be given in Section 4.5, after the derivation of some intermediate results. □

4.3. Reduction to the case of a normed scoring rule

In the following it will be our aim to show that a scoring rule satisfies axioms 1 to 4 if and only if it is a positive linear transformation of a normed scoring rule satisfying axioms 1 to 4. In this way the task of proving the theorem is reduced to the task of showing that there is one and only one normed scoring rule satisfying axioms 1 to 4, namely the quadratic scoring rule.

Lemma 2. *Let R be a scoring rule and let S be a positive linear transformation of R. Then S satisfies axioms* 1 *to* 4 *if and only if R satisfies axioms* 1 *to* 4.

Proof: It has been pointed out above that a scoring rule S is a positive linear transformation of a scoring rule R if and only if R is a positive linear transformation of S. It can be seen without difficulty for each of the axioms 1 to 4 that it is invariant with respect to positive linear transformations, in the sense that any positive linear transformation of a scoring rule satisfying it, also satisfies it. Therefore, the assertion holds. □

Lemma 3. *A scoring rule R is a positive linear transformation of a normed scoring rule if and only if the following condition is satisfied*:

$$R_1(1,0) > R_1(0,1) > -\infty.$$

Proof: Suppose that the condition is satisfied. It can be seen immediately that in this case a R is a positive linear transformation of a normed scoring rule S. Now suppose that we have:

$$R_1(1,0) \leq R_1(0,1).$$

Obviously in this case it is not possible to find a normed scoring rule S, such that R is a positive linear transformation of S. The same is true for $R_1(0,1) = -\infty$, since under a positive linear transformation the image of a real score is real. It follows that the assertion holds. □

Lemma 4. *Let R be a scoring rule satisfying axioms* 1, 3 *and* 4. *Then the condition*

$$R_1(1,0) > R_1(0,1) > -\infty$$

is satisfied.

Proof: Since R is incentive compatible by axiom 3 the expected score with respect to R has the property

$$V(r \mid r) > V(p \mid r) \quad \text{for } p \neq r.$$

Therefore, we have

$$V(r \mid r) > -\infty \quad \text{for every } r.$$

Incentive compatibility together with the definition of the expected score also permits the conclusion:

$$R_1(1, 0) = V((1, 0) \mid (1, 0)) > V((0, 1) \mid (1, 0)) = R_1(0, 1).$$

In order to prove the lemma it remains to show

$$R_1(0, 1) > -\infty.$$

We now indirectly prove that this inequality holds. Assume $R_1(0, 1) = -\infty$ and let w be the distribution.

$$w = (.5, .5).$$

We have

$$L((0, 1) \mid w) = V(w \mid w) - V((0, 1) \mid w)$$

and

$$V((0, 1) \mid w) = .5R_1(0, 1) + .5R_2(0, 1) = -\infty.$$

In view of the second inequality of this proof, this yields

$$L((0, 1) \mid w) = +\infty.$$

The neutrality axiom 4 yields

$$L(w \mid (0, 1)) = L((0, 1) \mid w) = +\infty.$$

We have

$$L(w \mid (0, 1)) = V((0, 1) \mid (0, 1)) - V(w \mid (0, 1))$$

or equivalently

$$L(w \mid (0, 1)) = V((0, 1) \mid (0, 1)) - .5R_2(0, 1).$$

The symmetry axiom 1 yields

$$R_2(0, 1) = R_1(1, 0).$$

We already know that $R_1(1, 0)$ is greater than $R_i(0, 1)$ and therefore greater than $-\infty$. This together with the second inequality of this proof yields the conclusion that the right hand

side of the last inequality for $L(w \mid (0, 1))$ is a real number and therefore smaller than $+\infty$, contrary to what has been shown with the help of axiom 4. □

Lemma 5. *Let R be a scoring rule satisfying axioms* 1 *to* 4. *Then R is a positive linear transformation of a normed scoring rule S satisfying axioms* 1 *to* 4.

Proof: Lemma 4 shows that the condition required by Lemma 3 for R is satisfied. Therefore R is a positive linear transformation of a normed scoring rule S. In view of Lemma 2 this scoring rule S satisfies axioms 1 to 4. □

Remark. Lemma 5 has an important consequence. If there is one and only one normed-scoring rule S which satisfies axioms 1 to 4 then every scoring rule R satisfying axioms 1 to 4 is a positive linear transformation of this normed scoring rule S; and in view of Lemma 2 every positive linear transformation of S satisfies axioms 1 to 4. Therefore, in order to prove the theorem, it is sufficient to show that the quadratic scoring rule is the only one satisfying axioms 1 to 4. We know already by Lemma 1 that it satisfies these axioms.

4.4. No other normed scoring rule satisfies the axioms

We first derive a result on the unit vector $\delta(1, n) = (1, 0, \ldots, 0)$

Lemma 6. *Let* $S = S_1, S_2, \ldots$ *be a normed scoring rule satisfying axiom* 2. *Then the following equations hold*:

$$S_1(\delta(1, n)) = 1 \quad \text{for } n = 1, 2, \ldots$$
$$S_2(\delta(1, n)) = -1 \quad \text{for } n = 2, 3, \ldots.$$

Proof: Since S is normed, it follows by the elongation invariance axiom 2 that we have

$$S_1(1) = S_1(1, 0) = 1.$$

Therefore, the first equation of the lemma holds for $n = 1$. Since S is normed both equations hold for $n = 2$. Suppose that both equations are valid for some n. It is an immediate consequence of axiom 2 that the equations also hold for $n + 1$ instead of n. The assertion of the lemma follows by induction. □

Lemma 7. *Let* $S = S_1, S_2, \ldots$ *be a normed scoring rule satisfying axioms* 1 *and* 2. *Then we have*

$$S_i(\delta(j, n)) = \begin{cases} 1 & \text{for } i = j \\ -1 & \text{for } i \neq j \end{cases}$$

for $i, j = 1, \ldots, n$ *and* $n = 1, 2, \ldots$.

Proof: We first prove the assertion for the special case $j = 1$. Lemma 6 shows that in this case the assertion holds for $i = 1$ and $i = 2$. We have to show that it also holds for $i = 3, \ldots, n$. Let i be one of these numbers and consider the permutation π_{2i} of $1, \ldots, n$ which exchanges 2 and i and leaves everything else unchanged. π_{2i} maps $\delta(1, n)$ to itself. Therefore, it follows by the symmetry axiom 1 that we have:

$$S_i(\delta(1, n)) = \begin{cases} 1 & \text{for } i = j \\ -1 & \text{for } i \neq j \end{cases}$$

for $i = 1, \ldots, n$. Now consider the permutation π_{1j} of $1, \ldots, n$ which exchanges 1 and j and leaves everything else unchanged. π_{1j} maps $\delta(1, n)$ to $\delta(j, n)$. In view of the symmetry axiom 1, this together with the equation for $S_i(\delta(1, n))$ yields the assertion of the lemma. □

Lemma 8. *The quadratic scoring rule is the only normed scoring rule satisfying axioms* 1, 2 *and* 4.

Proof: For every $p \in \Delta_n$ the neutrality axiom yields

$$L(\delta(i, n) \mid p) = L(p \mid \delta(i, n))$$

for $i = 1, \ldots, n$ and $n = 1, 2, \ldots$. This is equivalent to

$$V(p \mid p) - V(\delta(i, n) \mid p) = V(\delta(i, n) \mid \delta(i, n)) - V(p \mid \delta(i, n)).$$

With the help of Lemma 7 we obtain:

$$\begin{aligned} V(p \mid p) - p_i(+1) - (1 - p_i)(-1) &= 1 - S_i(p) \\ V(p \mid p) - 2p_i + 1 &= 1 - S_i(p) \end{aligned}$$

This yields the following *preliminary formula* for $S_i(p)$:

$$S_i(p) = 2p_i - V(p, p).$$

If the right-hand side is inserted for $S_i(p)$ in the definition of $V(p, p)$ we obtain:

$$V(p \mid p) = \sum_{i=1}^{n} p_i S_i(p) = \sum_{j=1}^{n} p_j (2p_j - V(p \mid p))$$

$$V(p \mid p) = -V(p \mid p) + 2 \sum_{i=1}^{n} p_j^2$$

Therefore, we have:

$$V(p \mid p) = \sum_{j=1}^{n} p_j^2$$

The right-hand side can be inserted for $V(p \mid p)$ in the preliminary formula for $S_i(p)$. This yields:

$$S_i(p) = 2p_i - \sum_{j=1}^{n} p_j^2$$

for $i = 1, \ldots, n$ and $n = 1, 2, \ldots$ The formula is nothing else than the quadratic scoring rule. □

4.5. Proof of the theorem

The theorem stated in Section 4.2 is proved in the following. As has been pointed out in the remark at the end of 4.3 the intermediate results obtained earlier have shown that it is sufficient for the proof of the theorem to show that the quadratic scoring rule is the only normed scoring rule satisfying axioms 1 to 4. Lemma 8 permits this conclusion. Therefore, the theorem holds.

4.6. Remark on the role of the axioms in the proof

In the following we shall look at the results which had to be proved in order to show that the quadratic scoring rule is the only one satisfying axioms 1 to 4. Not every scoring rule is a positive linear transformation of a normed scoring rule. For example, the logarithmic scoring rule cannot be obtained in this way, since it has the property $S_1(0, 1) = -\infty$. The condition of Lemma 3 is necessary and sufficient for the property of a scoring rule to be obtainable as a positive linear transformation of a normed one. Lemma 4 shows that axioms 1, 3, and 4 imply this conditions. Axiom 2 does not enter the picture, since the condition concerns only the first scoring function applied to distributions over two alternatives. Axiom 4 is of crucial importance here, as the example of the logarithmic scoring rule shows. The logarithmic scoring rule satisfies axioms 1 and 3 but not 4.

After the reduction to the case of a normed scoring rule it had to be shown that the quadratic scoring rule is the only normed scoring rule satisfying axioms 1 to 4. The elongation axiom 2 comes into play with Lemma 6. It is used to determine the scores assigned by the first two scoring functions to unit vectors $(1, \ldots, 0)$ with 1 in the first component. Later axiom 2 is not anymore directly applied. Only symmetry considerations based on axiom 1 are used in the proof of Lemma 7 in order to determine all scores assigned by scoring functions to arbitrary unit vectors. It is remarkable that these scores are fully determined by axioms 1 and 2 together with the property of being normed.

After the scores for the unit vectors have been obtained, the neutrality axiom 4 is the only one directly applied. It is used only at the beginning of the proof of Lemma 8 in order to establish the equality of the expected score losses of $\delta(i, n)$ at p and of p at $\delta(i, n)$ for arbitrary $p \in \Delta_n$. The remainder of the proof is entirely based on elementary algebraic manipulations.

Interestingly the incentive compatibility axiom 3 is not anymore needed after the reduction to the case of a normed scoring rule. However, in order to achieve this reduction it is of

crucial importance for deriving the conclusion of Lemma 4 that $R_1(1, 0)$ is greater than $R_1(0, 1)$. An axiom asserting this relationship could replace incentive compatibility in our characterization of the quadratic scoring rule.

4.7. Remark on the other scoring rules described in Section 2

It can be seen easily that all scoring rules described in Section 2 satisfy the symmetry axiom 1 and the elongation axiom 2. As has been shown in Section 2 the linear scoring rule is not incentive compatible, but all other scoring rules described there satisfy axiom 3.

Our characterization of the quadratic scoring rule permits the conclusion that other scoring rules satisfying axioms 1, 2 and 3 do not satisfy axiom 4. The linear scoring rule does not satisfy axiom 4, since for this rule we have

$$L(p \mid q) = \sum_{i=1}^{n} q_i^2 - q_i p_i$$

and

$$L(p \mid q) = \sum_{i=1}^{n} p_i^2 - p_i q_i$$

for $p, q \in \Delta_n$. Obviously both expressions are in general not equal to each other. Among all the scoring rules described in Section 2, the quadratic scoring rule is the only one satisfying axiom 4.

5. Concluding remarks

As has been shown, four plausible axioms characterize the quadratic scoring rule up to positive linear transformations. This lends support to the idea that the quadratic scoring rule should be used for the comparison of competing probabilistic theories wherever this is possible.

Of course, it is a methodological decision to use the quadratic scoring rule as the criterion of predictive success. However, such decisions should not be looked on as arbitrary. It is possible to put them on an axiomatic basis. If this is done, the discussion of the merits of specific methods becomes more transparent.

The quadratic scoring rule can be applied to fully specified probabilistic theories that predict probability distributions over available alternatives. Sometimes such theories may contain some unknown parameters. The quadratic scoring rule cannot be applied in such cases unless the parameters are specified. One can estimate them in a way that maximizes the quadratic score sum. It seems to be appropriate to compare two theories with unknown parameters in this way, provided the number of parameters to be estimated from the data is the same in both cases.

Different measures of predictive success must be used for different types of theories. Elsewhere the author has axiomatized a measure of predictive success for area theories

(Selten, 1991). Area theories do not predict probability distributions but areas in which outcomes should tend to lie. Here the difficulty arises that one cannot just count hits. The size of the area must be taken into account. The smaller the area the better is the prediction. A balance between precision (smallness of the area) and accuracy (relative number of hits) must be struck.

In the experimental literature sometimes theories are considered that predict neither outcome areas nor probability distributions over outcomes but rather areas of probability distributions over outcomes. Thus Hey and Orme (1994) look at the explanation of lottery choices by theories involving utility maximization subject to errors. In such cases, a subset of a space of probability distributions is predicted but only realizations are observed. If the probability distributions are known up to a few undetermined parameters, one can estimate them from the data in a way that maximizes quadratic score sums. If, however, the subspace of predicted probability distributions is not a parametric family but is delineated by qualitative properties like unimodality and so on, a different approach must be taken. It is an interesting open problem to find a reasonable measure of predictive success for such situations.

Acknowledgments

I am greatly indebted to Daniel Friedman, who provided valuable advise and made me aware of relevant literature. My thanks also go to Fang Fang Tang who found the paper by Brier and to Rika Fülling for typing the manuscript. Support by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303, is gratefully acknowledged.

References

Brier, Glenn W. (1950). "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review*. 78(1), 1–3.

Friedman, Daniel (1983). "Effective Scoring Rules for Probabilistic Forecasts." *Management Science*. 29, 447–454.

Hey, I.D. and Orme, Ch. (1994). "Investigating Generalisations of Expected Utility Theory Using Experimental Data." *Econometrica*. 62, 1261–1326.

Matheson, J.E. and Winkler, R.L. (1976). "Scoring Rules for Continuous Probability Distributions." *Management Science*. 22, 1078–1096.

Murphy, A.H. and Winkler, R.L. (1970). "Scoring Rules in Probability Assessment and Evaluation." *Acta Psych*. 34, 273–286.

Ostrowski, A. (1965). *Vorlesungen über Differential und Integralrechnung. 2. Auflage*. Basel-Stuttgart: Birkhäuser-Verlag.

Roby, T.B. (1965). *Belief States: A Preliminary Empirical Study*. Decision Science Laboratory, L.G. Hascom Field.

Selten, R. (1991). "Properties of a Measure of Predictive Success." *Mathematical Social Sciences*. 21, 153–167.

Stael von Holstein, C.-A.S. (1970). "Measurement of Subjective Probability." *Acta Psych*. 34, 146–159.

Toda, Masanao (1963). "Measurement of Subjective Probability Distribution." Report 3, State College, Pennsylvania, Institute for Research, Division of Mathematical Psychology.

Winkler, R.L. (1969). "Scoring Rules and the Evaluation of Probability Assessors." *Journal of the American Statistical Association*. 64, 1073–1078.