The Annals of Statistics
2012, Vol. 40, No. 1, 593–608
DOI: 10.1214/12-AOS972

PROPER LOCAL SCORING RULES ON DISCRETE SAMPLE SPACES

BY A. PHILIP DAWID, STEFFEN LAURITZEN AND MATTHEW PARRY[1]

University of Cambridge, University of Oxford and University of Otago

A *scoring rule* is a loss function measuring the quality of a quoted probability distribution Q for a random variable X, in the light of the realized outcome x of X; it is *proper* if the expected score, under any distribution P for X, is minimized by quoting $Q = P$. Using the fact that any differentiable proper scoring rule on a finite sample space $\mathcal{X}$ is the gradient of a concave homogeneous function, we consider when such a rule can be *local* in the sense of depending only on the probabilities quoted for points in a nominated neighborhood of x. Under mild conditions, we characterize such a proper local scoring rule in terms of a collection of homogeneous functions on the cliques of an undirected graph on the space $\mathcal{X}$. A useful property of such rules is that the quoted distribution Q need only be known up to a scale factor. Examples of the use of such scoring rules include Besag's pseudo-likelihood and Hyvärinen's method of ratio matching.

1. Introduction. Let $\mathcal{X}$ be a finite set, let $\mathcal{A}$ be the set of real vectors $\boldsymbol{\alpha} = (\alpha_x : x \in \mathcal{X})$ with each $\alpha_x > 0$, and let $\mathcal{P} = \{\mathbf{p} \in \mathcal{A} : \sum_x p_x = 1\}$ be the set of such vectors corresponding to strictly positive probability distributions on $\mathcal{X}$. We will use P for the distribution determined by $\mathbf{p}$ (similarly Q for $\mathbf{q}$), and generally do not distinguish between them. For $\boldsymbol{\alpha} \in \mathcal{A}$, $C \subseteq \mathcal{X}$ we write $\boldsymbol{\alpha}_C := (\alpha_x : x \in C)$, and similarly $\mathbf{p}_C$.

Consider a game between Forecaster and Nature, where Forecaster quotes a distribution $Q \in \mathcal{P}$ as representing his uncertainty about a quantity X taking values in $\mathcal{X}$, and Nature then reveals $X = x$. A *scoring rule* [see, e.g., Dawid (1986)] is a function $S : \mathcal{X} \times \mathcal{P} \to \mathbb{R}$. The interpretation is that $S(x, Q)$ measures the loss suffered by Forecaster for the above outcome of the game.

For $P \in \mathcal{P}$ we define $S(P, Q) := \sum_x p_x S(x, Q)$, the expected score when Forecaster quotes Q, and Nature generates X from P. The scoring rule S is *proper* if always $S(P, Q) \geq S(P, P)$, so that it is always optimal to quote a distribution Q matching the real uncertainty P; S is *strictly proper* if furthermore $S(P, Q) > S(P, P)$ when $Q \neq P$.

Received April 2011; revised January 2012.

[1]Supported by EPSRC Statistics Mobility Fellowship EP/E009670.

MSC2010 subject classifications. Primary 62C99; secondary 62A99.

Key words and phrases. Concavity, entropy, Euler's theorem, supergradient, homogeneous function.

The generalized *entropy function*, or *uncertainty function*, $H : \mathcal{P} \to \mathbb{R}$, associated with a proper scoring rule S is given by $H(P) := S(P, P)$. Then H is a concave function on $\mathcal{P}$. We also introduce the associated *divergence* or *discrepancy* function $d : \mathcal{P} \times \mathcal{P} \to \mathbb{R}$, where $d(P, Q) := S(P, Q) - H(P)$. Then $d(P, Q) \geq 0$, with equality if $Q = P$ (and only in this case if S is strictly proper).

As well as being of intrinsic interest, proper scoring rules have a range of applications. For example, if $\mathcal{Q} = \{Q_\theta : \theta \in \Theta\}$ is a smooth parametric statistical model, we might estimate θ, based on a random sample $(x_1, \ldots, x_n)$, by minimizing the *empirical discrepancy*, $d(\hat{P}_n, Q_\theta)$, where $\hat{P}_n$ is the empirical distribution of the data. This is equivalent to minimizing $\sum_{i=1}^n S(x_i, Q_\theta)$. Implementing this by setting the derivative of this criterion to 0 will yield an unbiased estimating equation [Dawid and Lauritzen (2005), Dawid (2007)], from which we, under suitable smoothness assumptions, can deduce statistical properties of the associated estimator such as consistency and asymptotic normality. For the well-known *logarithmic score* $S(x, Q) = -\log q_x$ this procedure leads to the maximum likelihood estimator, but it is of interest to use other scoring rules and estimators, for example, because they can lead to greatly simplified calculations. We illustrate this in Section 4 for capture–recapture experiments and pseudo-likelihood estimation for image analysis; see also Czado, Gneiting and Held (2009) for a range of other applications of proper scoring rules to discrete data.

Parry, Dawid and Lauritzen (2012) investigated when a proper scoring rule with $\mathcal{X}$ an interval on the real line can be *local* in the sense that $S(x, Q)$ depends on Q only through the value $f(x)$ of the density f of Q and the values $f^{(k)}(x)$ of a finite number of the derivatives of f at the realized outcome x and Ehm and Gneiting (2012) studied rules with $k = 2$ in further detail. It was shown in Parry, Dawid and Lauritzen (2012) that any proper local scoring rule is a linear combination of the logarithmic score and what was termed a *key local scoring rule*; and that any such key local scoring rule is 0-*homogeneous* in the sense that $S(x, Q)$ can be evaluated when f is only known up to proportionality. The results in this article for discrete sample spaces parallel these. However, in the case of discrete $\mathcal{X}$ we have to redefine locality using a neighborhood structure on the space $\mathcal{X}$, and use somewhat different techniques of proof.

The organization of the paper is as follows. In Section 2 we review results from Hendrickson and Buehler (1971) characterizing proper scoring rules as supergradients of concave functions.

In Section 3 we formally define what it means for a scoring rule to be local with respect to a neighborhood system and show that if the homogeneous extension of the scoring rule is local, the neighborhood system must be determined by an undirected graph. We also describe a general additive construction of local scoring rules. Section 4 gives examples of local scoring rules and their use. In Section 5 we proceed in parallel to Parry, Dawid and Lauritzen (2012) by a variational argument to characterize local scoring rules as solutions to a key differential equation and, under an additional condition on the neighborhood system, we show that such local scoring rules can be expanded in additive terms indexed by complete subsets of an undirected graph.

2. Homogeneous proper scoring rules. Further analysis is facilitated by recasting the problem in terms of homogeneous functions and using the fundamental characterization of proper scoring rules given by McCarthy (1956) and Hendrickson and Buehler (1971).

2.1. *Homogeneous functions.* A function $f : \mathcal{A} \to \mathbb{R}$ is called (*positive*) *homogeneous of order* h, or h-*homogeneous*, if

$$f(\lambda\boldsymbol{\alpha}) \equiv \lambda^h f(\boldsymbol{\alpha}) \qquad \text{for all } \lambda > 0. \tag{1}$$

In this paper we shall only need homogeneity of orders 0 and 1. If f is differentiable, (1) will hold if and only if f satisfies *Euler's equation*:

$$\sum_x \alpha_x \frac{\partial f}{\partial \alpha_x} = hf. \tag{2}$$

Even when f is not differentiable, in some circumstances we can reinterpret (2) so as to continue to apply.

DEFINITION 2.1. A vector $\nabla f(\boldsymbol{\alpha}) \in \mathcal{A}$ is a *supergradient* to f at $\boldsymbol{\alpha}$ if, for all $\boldsymbol{\beta} \in \mathcal{A}$,

$$f(\boldsymbol{\alpha}) + (\boldsymbol{\beta} - \boldsymbol{\alpha})^{\mathrm{T}} \nabla f(\boldsymbol{\alpha}) \geq f(\boldsymbol{\beta}).$$

When f is differentiable at $\boldsymbol{\alpha}$ and has a supergradient $\nabla f(\boldsymbol{\alpha})$ there, it must coincide with the gradient vector $(\partial f/\partial \alpha_x : x \in \mathcal{X})$. Lemma 2.1 below and Corollary 2.2 extend Euler's equation (2) to homogeneous functions with a supergradient and are equivalent to Theorem 2.1 of Hendrickson and Buehler (1971) and subsequent remarks, so we omit the proofs here.

LEMMA 2.1. *Suppose f is h-homogeneous, and has a supergradient $\nabla f(\boldsymbol{\alpha})$ at $\boldsymbol{\alpha}$. Then*

$$\boldsymbol{\alpha}^{\mathrm{T}} \nabla f(\boldsymbol{\alpha}) = hf(\boldsymbol{\alpha}). \tag{3}$$

COROLLARY 2.2. *Suppose f is 1-homogeneous. Then $\mathbf{S}$ is a supergradient of f at $\boldsymbol{\alpha}$ if and only if*

$$\boldsymbol{\beta}^{\mathrm{T}} \mathbf{S} \geq f(\boldsymbol{\beta})$$

for all $\boldsymbol{\beta} \in \mathcal{A}$, with equality when $\boldsymbol{\beta} = \boldsymbol{\alpha}$.

By the supporting hyperplane theorem, a function f is concave on $\mathcal{A}$ if and only if it has a supergradient at each $\boldsymbol{\alpha} \in \mathcal{A}$ (not necessarily unique if f is not differentiable at $\boldsymbol{\alpha}$). A *supergradient function* ∇f associates a specific choice of supergradient $\nabla f(\boldsymbol{\alpha})$ with each point $\boldsymbol{\alpha} \in \mathcal{A}$. If f is h-homogeneous, (3) holds at each $\boldsymbol{\alpha} \in \mathcal{A}$ for any choice of supergradient function ∇f.

2.2. *Homogeneous scoring rules.* Clearly, any scoring rule $S(x, P)$ can readily be extended to $\mathcal{A}$ by defining $S(x, \boldsymbol{\alpha}) := S(x, \boldsymbol{\alpha}/\alpha_+)$, where $\alpha_+ := \sum_{y\in\mathcal{X}} \alpha_y$. So extended, $S(x, \boldsymbol{\alpha})$ is a 0-homogeneous function of $\boldsymbol{\alpha}$ for every x and we say that $S(x, \boldsymbol{\alpha})$ is a 0-*homogeneous scoring rule*.

McCarthy (1956) states that a 0-homogeneous scoring rule S is proper if and only if it can be expressed as the supergradient of a concave 1-homogeneous function $H : \mathcal{A} \to \mathbb{R}$. This is formally proved in Hendrickson and Buehler (1971) and stated below in Theorems 2.3 and 2.4.

THEOREM 2.3. *Suppose $H : \mathcal{A} \to \mathbb{R}$ is concave and* 1-*homogeneous. Let ∇H be a supergradient of H, and, for $x \in \mathcal{X}$, $\mathbf{p} \in \mathcal{P}$, define $S(x, \mathbf{p})$ to be the x-component of the vector $\mathbf{S}(\mathbf{p}) := \nabla H(\mathbf{p})$. Then S is a proper scoring rule, and the associated entropy at $\mathbf{p}$ is $H(\mathbf{p})$.*

We note that the definition $\mathbf{S}(\boldsymbol{\alpha}) := \nabla H(\boldsymbol{\alpha})$ can be used to extend the domain of S from $\mathcal{X} \times \mathcal{P}$ to $\mathcal{X} \times \mathcal{A}$. The supergradient function ∇H can be taken to be 0-homogeneous and then $S(x, \boldsymbol{\alpha})$ is a 0-homogeneous function of $\boldsymbol{\alpha}$.

For the converse direction, starting with a scoring rule S defined on $\mathcal{X} \times \mathcal{P}$, we let $S(x, \boldsymbol{\alpha})$ denote its 0-homogeneous extension as described above and let $\mathbf{S}(\boldsymbol{\alpha})$ be the vector with x-component $S(x, \boldsymbol{\alpha})$.

THEOREM 2.4. *Suppose that $S(x, \boldsymbol{\alpha})$ is a* 0-*homogeneous proper scoring rule. Define $H(\boldsymbol{\alpha}) := \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{S}(\boldsymbol{\alpha})$. Then H is* 1-*homogeneous and concave, and $\mathbf{S}(\boldsymbol{\alpha})$ is a supergradient of H at $\boldsymbol{\alpha}$.*

As a consequence we obtain the following symmetry relation for the partial derivatives of any 0-homogeneous proper scoring rule.

COROLLARY 2.5. *If S is a* 0-*homogeneous proper scoring rule, and $S(x, \boldsymbol{\alpha})$ is continuously differentiable on $\mathcal{A}$ for each $x \in \mathcal{X}$, then*

$$\frac{\partial S(x, \boldsymbol{\alpha})}{\partial \alpha_y} = \frac{\partial S(y, \boldsymbol{\alpha})}{\partial \alpha_x}. \tag{4}$$

PROOF. In this case $H(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{S}$ is differentiable on $\mathcal{A}$, so $\mathbf{S}(\boldsymbol{\alpha})$ is its gradient. It immediately follows that H is twice continuously differentiable. Then (4) follows from $\partial^2 H/\partial\alpha_y\,\partial\alpha_x = \partial^2 H/\partial\alpha_x\,\partial\alpha_y$. $\square$

EXAMPLE 2.1. Examples of proper scoring rules are the *Brier score* $S(x, \mathbf{p}) = \|\mathbf{p}\|^2 - 2p_x$ [Brier (1950)], where $\|\mathbf{p}\|^2 = \sum_x p_x^2$, with 1-homogeneous entropy function $H(\boldsymbol{\alpha}) = -\|\boldsymbol{\alpha}\|^2/\alpha_+$ and the *spherical score* $S(x, \mathbf{p}) = -p_x/\|\mathbf{p}\|$ with 1-homogeneous entropy function $H(\boldsymbol{\alpha}) = -\|\boldsymbol{\alpha}\|$ [Good (1971), Dawid (2007)].

We shall say that the entropy function H is *regular* if it is continuous on $\mathcal{A}$ and its closure $\operatorname{cl} H$ as a concave function [Rockafellar (1970), page 52] is finite on the closed cone $\bar{\mathcal{A}} = \{\boldsymbol{\alpha} : \alpha_x \geq 0, x \in \mathcal{X}\}$. In other words, H is regular if it can be extended by continuity to have finite values for all $\boldsymbol{\alpha} \in \bar{\mathcal{A}}$.

Clearly, since $H(\boldsymbol{\alpha}) = \sum_x \alpha_x S(x, \boldsymbol{\alpha})$, $H(\boldsymbol{\alpha})$ is certainly regular if $S(x, P)$ is bounded in P for each x. Both the Brier score and the spherical score satisfy this requirement and have regular entropy functions, but in general boundedness is not necessary for regularity.

3. Local scoring rules. In general, as for the Brier and spherical score, $S(x, P)$ will depend on every element of P. We are interested in cases where this is not so.

3.1. *Locality.* Suppose we specify, for each $x \in \mathcal{X}$, a set $N_x \subseteq \mathcal{X}$ (the *neighborhood* of x), containing x, and require that the proper scoring rule $S(x, P)$ be expressible as a function of x and the restriction $\mathbf{p}_{N_x}$ of $\mathbf{p}$ to N_x:

$$S(x, P) = s(x, \mathbf{p}_{N_x}).$$

We say that such a scoring rule is $\mathcal{N}$*-local*, where $\mathcal{N} = \{N_x : x \in \mathcal{X}\}$ is the neighborhood system. Similarly, its 0-homogeneous extension is said to be $\mathcal{N}$*-local* if $S(x, \boldsymbol{\alpha}) = s(x, \boldsymbol{\alpha}_{N_x})$. Note this property is strictly stronger; see Section 3.2 below.

Suppose that the 0-homogeneous extension of S is continuously differentiable and $\mathcal{N}$-local. We then obtain from (4) that, if $x \notin N_y$,

$$\frac{\partial S(x, \boldsymbol{\alpha})}{\partial \alpha_y} = \frac{\partial S(y, \boldsymbol{\alpha})}{\partial \alpha_x} = 0,$$

so that without loss of generality we can also require $y \notin N_x$. Hence, for scoring rules with $\mathcal{N}$-local 0-homogeneous extensions we can assume that the neighborhood relation is symmetric and so determined by an undirected graph $\mathcal{G}$ so that $y \in N_x$ if and only if $x = y$ or x–y, that is, x and y are neighbors in $\mathcal{G}$. We then also say that the scoring rule and its extension are $\mathcal{G}$*-local*.

We note that *a scoring rule with a* $\mathcal{G}$*-local* 0*-homogeneous extension only depends on* P *through its conditional distribution* $\mathbf{p}_{|N_x}$ *of* X *given* $X \in N_x$, that is, it satisfies

$$S(x, P) = s(x, \mathbf{p}_{|N_x}). \tag{5}$$

In particular, only knowledge of $\mathbf{p}$ up to a constant factor is necessary to calculate $S(x, P)$. Conversely, *the* 0*-homogeneous extension of any scoring rule satisfying* (5) *is* $\mathcal{G}$*-local.*

3.2. *Logarithmic score.* The simplest case of a local scoring rule is where $S(x, P)$ is a function only of x and p_x, and is thus $\mathcal{G}_0$-local for the totally disconnected graph $\mathcal{G}_0$. It is well known [Bernardo (1979)] that (for $\#\mathcal{X} > 2$) a scoring rule with this property is proper (and is then strictly proper) if and only if it has the form

$$S(x, P) = a(x) - \lambda \ln p_x \tag{6}$$

with $\lambda > 0$. For $a(x) = 0$ this is known as the *log-score*.

As described previously, any scoring rule has a 0-homogeneous extension which in this case is $a(x) - \lambda \ln \alpha_x + \lambda \ln \alpha_+$; however, the extension depends, not just on α_x, but on α_y for all $y \in \mathcal{X}$. Hence, although the scoring rule itself is local, its 0-homogenous extension is not, reflected in the fact that knowledge of $\mathbf{p}$ up to a constant factor is not sufficient for calculating the log-score. In fact, there is no nontrivial proper scoring rule with a $\mathcal{G}_0$-local 0-homogeneous extension.

Note that the (Shannon) entropy function $H(\boldsymbol{\alpha}) = -\lambda \sum_x \alpha_x \log(\alpha_x/\alpha_+)$ for the log-score is regular although the log-score itself is unbounded.

3.3. *Additive scoring rules.* Here we describe a simple way of constructing a 0-homogeneous local scoring rule. Let $\mathcal{B}$ be a collection of subsets of $\mathcal{X}$, define $\mathcal{A}_B := \{\boldsymbol{\alpha}_B : \boldsymbol{\alpha} \in \mathcal{B}\}$, and let $H_B : \mathcal{A}_B \to \mathbb{R}$ be a concave and 1-homogeneous function of $\boldsymbol{\alpha}_B$—and thus also, by extension of its domain, of $\boldsymbol{\alpha}$. Let $\nabla H_B \in \mathcal{A}_B$ be a 0-homogeneous supergradient of H_B on $\mathcal{A}_B$; this is also a 0-homogeneous supergradient on the extended domain $\mathcal{A}$, if we define its components for $x \notin B$ as 0. By the results of Section 2, this determines a proper scoring rule $S_B(x, \boldsymbol{\alpha})$. Moreover, S_B vanishes if $x \notin B$, and otherwise depends on $\boldsymbol{\alpha}$ only through $\boldsymbol{\alpha}_B$. We now let

$$S(x, \boldsymbol{\alpha}) = \sum_{B \in \mathcal{B}} S_B(x, \boldsymbol{\alpha}_B), \qquad H(\boldsymbol{\alpha}) = \sum_{B \in \mathcal{B}} H_B(\boldsymbol{\alpha}_B) \tag{7}$$

and these define a proper and 0-homogeneous scoring rule and its associated 1-homogeneous entropy function. We shall say that a scoring rule and entropy function satisfying (7) are *$\mathcal{B}$-additive*. When each H_B is a differentiable function of $\boldsymbol{\alpha}_B$, the gradient of H will be the unique associated scoring rule S of form (7).

We note that if we define an undirected graph $\mathcal{G}$ by x–y if and only if $x, y \in B$ for some $B \in \mathcal{B}$, we have that the (0-homogenous extension of) *any $\mathcal{B}$-additive scoring rule is $\mathcal{G}$-local*. If $\mathcal{C}$ denotes the collection of all cliques of $\mathcal{G}$, that is, all maximal complete subsets of $\mathcal{X}$, we can collect terms appropriately and rewrite the expansions in (7) above as

$$S(x, \boldsymbol{\alpha}) = \sum_{C \in \mathcal{C}} s_C(x, \boldsymbol{\alpha}_C) \tag{8}$$

and

$$H(\boldsymbol{\alpha}) = \sum_{C \in \mathcal{C}} h_C(\boldsymbol{\alpha}_C). \tag{9}$$

We say that a scoring rule S and entropy function H having the forms of (8) and (9) are $\mathcal{G}$*-additive.*

We remark that the above constructions can also be applied straightforwardly to the case of a countably infinite sample space $\mathcal{X}$, so long as every set $B \in \mathcal{B}$ is finite.

We shall in Section 5 give conditions for the converse to hold, that is, conditions for a $\mathcal{G}$-local scoring rule to be $\mathcal{G}$-additive as above, without necessarily demanding each term of the decomposition (9) to be concave or 1-homogeneous.

4. Examples. This section gives some examples of $\mathcal{G}$-additive and $\mathcal{G}$-local scoring rules.

4.1. *Local scoring rules for integer-valued outcomes.* We first consider cases where the outcomes are nonnegative integers.

EXAMPLE 4.1 (Pair scoring rule). Suppose $\mathcal{X} = \{0, 1, 2, \ldots\}$, and let the graph $\mathcal{G}$ have edges between successive integers. The cliques are just the pairs, $C_x := \{x, x+1\}$ $(x = 0, 1, \ldots)$, and a concave, 1-homogeneous local entropy function on C_x has the form $H_x(\alpha_x, \alpha_{x+1}) = \alpha_x G_x(\alpha_{x+1}/\alpha_x)$ with G_x concave. The associated additive scoring rule is

$$S(x, P) = G'_{x-1}\left(\frac{p_x}{p_{x-1}}\right) + G_x\left(\frac{p_{x+1}}{p_x}\right) - \frac{p_{x+1}}{p_x} G'_x\left(\frac{p_{x+1}}{p_x}\right) \qquad (x = 0, 1, \ldots)$$

with the first term absent if $x = 0$. The total score based on a sample $(x_1, \ldots, x_n)$ in which the frequency of y is f_y $(y = 0, 1, \ldots)$ is thus

$$\sum_{y=0}^{\infty} f_y G_y(v_y) + (f_{y+1} - f_y v_y) G'_y(v_y)$$

with $v_y := p_{y+1}/p_y$. If, for example, we wished to fit the Poisson model $p_x \propto \theta^x/x!$, we could estimate θ by minimizing the total empirical score

$$\sum_{y=0}^{\infty} f_y G_y\left(\frac{\theta}{y+1}\right) + \left(f_{y+1} - \frac{f_y}{y+1}\theta\right) G'_y\left(\frac{\theta}{y+1}\right).$$

Taking $G_x(v) = -(x+1)^a v^m / m(m-1)$ for $m \neq 0, 1$, we obtain the unbiased estimating equation

$$\theta \sum_{y \geq 0} \frac{f_y}{(y+1)^{m-a}} - \sum_{y \geq 0} \frac{f_{y+1}}{(y+1)^{m-a-1}} = 0 \tag{10}$$

yielding a simple explicit formula for the estimate. When $m = a$, we recover the maximum likelihood estimate $\hat{\theta} = \bar{x}$.

EXAMPLE 4.2 (Capture–recapture). Consider the following experiment performed to estimate the number N of fish in a lake. On c consecutive occasions we catch a fish, at random, and then replace it. When a fish is first caught it is given a unique tag, so that it can be recognized on recapture. Each fish $i = 1, \dots, N$ in the lake has an associated random variable X_i, the number of times it is caught. For large c and N we can approximate the distribution of X_i by the Poisson distribution with mean $\theta = c/N$. We will know f_x, the number of fish caught x times, for $x > 0$, but not f_0, the number of fish never caught. The observed data thus arise from a truncated Poisson distribution for X, conditioned on $X > 0$. If we can estimate θ we can estimate $N = c/\theta$. However, because of the need to work with the normalization constant of the truncated Poisson distribution, the maximum likelihood estimate of θ cannot be expressed in explicit form and must be determined numerically.

Homogeneous local scoring rules can be used to avoid the normalization constant problem and obtain an explicit estimate. We simply modify the above analysis of the full Poisson model by removing the edge 0–1 from the neighborhood graph $\mathcal{G}$, together with its associated local entropy function. Equivalently, we redefine $G_0 \equiv 0$. With the other explicit choices made above, the resulting estimating equation is given by (10) but with the sums now over $y \geq 1$. For $m = a$ we obtain

$$\tilde{\theta} = \frac{c - f_1}{n}, \tag{11}$$

where $n = \sum_{x \geq 1} f_x$ is the number of different fish caught. Note that $c - f_1$ is the number of times a catch yields a fish which is already marked. In comparison the maximum likelihood estimate $\hat{\theta}$ satisfies

$$\hat{\theta} = \frac{c - ne^{-\hat{\theta}}}{n},$$

so f_1 in (11) is replaced by the estimate of its expectation $\mathrm{E}_{\hat{\theta}}(F_1) = ne^{-\hat{\theta}}$.

In the interests of robustness we might also omit other data, and again this is easily done. For example, let the only edge in $\mathcal{G}$ be 1–2; equivalently, we take $G_x = 0$ for $x \neq 1$. Then, as well as f_0, the counts $f_3, f_4, \dots$ are also excluded, only the terms for $y = 1$ remain in (10), and we obtain the robust Zelterman estimate [Zelterman (1988)]:

$$\check{\theta} = \frac{2f_2}{f_1}.$$

4.2. *Local scoring rules for product spaces.* Suppose our discrete sample space is itself a product space, $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k$. A point x of $\mathcal{X}$ has the form $x = (x_1, \dots, x_k)$. We can define a useful symmetric neighborhood relation on $\mathcal{X}$ by x–y if, for some i, $x^{\setminus i} = y^{\setminus i}$, where $x^{\setminus i} := (x_j : j \neq i)$. A maximal clique of the associated graph $\mathcal{G}$ is then defined by an index $i \in \{1, \dots, k\}$ and a vector $\xi^{\setminus i} \in \mathcal{X}^{\setminus i} := \times_{j \neq i} \mathcal{X}_j$, and has the form $C_{i,\xi^{\setminus i}} := \{x : x^{\setminus i} = \xi^{\setminus i}\}$. Within such a clique, only the value of x_i can vary, over the space $\mathcal{X}_i$.

We can introduce, for such a clique $C = C_{i,\xi^{\setminus i}}$, a 1-homogeneous concave function H_C of $\boldsymbol{\alpha}_C$. Its gradient will determine a 0-homogeneous proper scoring rule $S_C(x, \boldsymbol{\alpha})$, vanishing unless $x \in C$, that is, $x^{\setminus i} = \xi^{\setminus i}$, and in this case depending on $\boldsymbol{\alpha}$ only through $\boldsymbol{\alpha}_C$. In particular, $S_C(x, P)$ depends on P only through the implied conditional distribution $P(\cdot \mid X^{\setminus i} = \xi^{\setminus i})$, for X_i, given $X^{\setminus i} = \xi^{\setminus i}$.

Conversely, any proper scoring rule defined for outcomes in $\mathcal{X}_i$ and distributions over $\mathcal{X}_i$ can be applied to the observed value x_i of X_i and the conditional distribution $P(\cdot \mid X^{\setminus i} = \xi^{\setminus i})$ (and taken as 0 if $x^{\setminus i} \neq \xi^{\setminus i}$): when denormalized, this will be of the above form. The general $\mathcal{G}$-additive scoring rule can then be formed by aggregating a collection of such single-clique component scoring rules:

$$S(x, P) = \sum_C S_C\{x_i, P(\cdot \mid X^{\setminus i} = \xi^{\setminus i})\} 1(x^{\setminus i} = \xi^{\setminus i}). \tag{12}$$

4.2.1. *Specialization*. Although it is allowable that the form of the component scoring rule S_C in (12) might vary with the conditioning values $\xi^{\setminus i}$ that, together with the index i, determine the clique C, this level of generality will rarely be needed, and we might thus restrict attention to proper scoring rules of the form

$$S(x, P) = \sum_{i=1}^{k} S_i\{x_i, P(\cdot \mid X^{\setminus i} = x^{\setminus i})\}, \tag{13}$$

where S_i is a proper scoring rule for variables in, and distributions over, $\mathcal{X}_i$. The associated discrepancy function is

$$d(P, Q) = \sum_i \mathrm{E}_{X \sim P} d_i\{P(\cdot \mid X^{\setminus i}), Q(\cdot \mid X^{\setminus i})\},$$

where d_i is the discrepancy function associated with S_i.

Recall that we are assuming that P, Q are everywhere positive distributions. If now each S_i is strictly proper, then $d(P, Q) = 0$ if and only if, for all i and $x^{\setminus i}$, $Q(\cdot \mid X^{\setminus i} = x^{\setminus i}) = P(\cdot \mid X^{\setminus i} = x^{\setminus i})$. But (with strict positivity) this can only occur if $Q = P$, so S is strictly proper.

4.2.2. *Markov models*. Suppose now that we have an undirected graph $\mathcal{K}$ with vertices $\{1, \dots, k\}$, and we restrict attention to distributions P that are Markov with respect to $\mathcal{K}$. Any component score S_C in (12), or S_i in (13), will then depend only on the value x_i of X_i, and the conditional distribution of X_i *given the neighbors* $X^{\mathrm{ne}(i)}$ of i in $\mathcal{K}$. It is possible to calculate this conditional distribution without having access to the normalizing constant of the overall distribution P, which is often hard to compute. In particular, in the estimation context described in the Introduction, this can greatly ease construction and solution of the unbiased estimating equation associated with this scoring rule.

A prominent example of a scoring rule of this kind is the *pseudo-likelihood* function introduced by Besag (1975):

EXAMPLE 4.3 (Pseudo-likelihood). When every component scoring rule S_i in (13) is the log score, the overall rule will be just the negative logarithm, $S(x, P) = -\log \mathrm{PL}(P, x)$, of the pseudo-likelihood function, defined, for a joint distribution P and outcome vector x, as

$$\mathrm{PL}(P, x) := \prod_i P(X_i = x_i \mid X^{\setminus i} = x^{\setminus i})$$

(where in the context of a Markov model the conditioning variables $X^{\setminus i}$ can be reduced to $X^{\mathrm{ne}(i)}$). Hence general properties of proper scoring rules can be applied to pseudo-likelihood. In particular, a maximum pseudo-likelihood estimator will typically be consistent under independent and identically distributed repetitions (though this argument does not address consistency under increasing dimension k, which is more relevant in many applications of pseudo-likelihood).

Replacing the log score with the Brier score leads to the method of *ratio matching* [Hyvärinen (2007)].

EXAMPLE 4.4 (Ratio matching). For the case $\mathcal{X}_i = \{0, 1\}$, take every component score S_i in (13) to be the Brier score, leading to the overall scoring rule

$$S(x, P) = \sum_i \{x_i - P(X_i = 1 \mid X^{\setminus i} = x^{\setminus i})\}^2. \tag{14}$$

For a parametric model $\mathcal{Q} = \{Q_\theta : \theta \in \Theta\}$, we could estimate θ by minimizing $\sum_i S(x_i, Q_\theta)$. This would equivalently minimize the empirical discrepancy $d(\hat{P}_n, Q_\theta)$, where

$$d(P, Q) = \sum_{i=1}^{k} \sum_{\xi^{\setminus i} \in \{0,1\}^{k-1}} P(X^{\setminus i} = \xi^{\setminus i})\{P^{\xi^{\setminus i}}(X_i = 1) - Q^{\xi^{\setminus i}}(X_i = 1)\}^2$$

with $P^{\xi^{\setminus i}}(X_i = 1) = P(X_i = 1 | X^{\setminus i} = \xi^{\setminus i})$, etc. This can be shown to agree with the more complex formula (13) of Hyvärinen (2007).[2]

5. Characterizing local scoring rules. Any positive linear combination of the log-score $-\lambda \ln p_x$ and a $\mathcal{G}$-additive score of form (8) will be $\mathcal{G}$-local. We now develop a converse to this result, assuming henceforth that $S(x, P)$ is continuously differentiable on $\mathcal{P}$. Under additional conditions on the neighborhood relation $\mathcal{N}$, we show that any proper such local scoring rule must be $\mathcal{G}$-local for a suitably

[2]The further analysis in that paper does not agree with our (14), and appears to contain some errors.

defined graph $\mathcal{G}$ and equal to a positive linear combination of the log-score and a 0-homogeneous $\mathcal{G}$-additive score.

We say x is *related* to y and write $x \sim y$, if $x, y \in N_z$ for some neighborhood $N_z \in \mathcal{N}$. Let $\rho(x) := \{y : y \sim x\}$ denote the set of *relatives* of x. Consider now the following condition on the neighborhood system $\mathcal{N}$:

CONDITION 5.1. There exist $y_1, y_2 \in \mathcal{X}$ such that, with $\rho_i := \rho(y_i)$:

$$\rho_1 \cap \rho_2 = \varnothing, \tag{15}$$

$$\rho_1 \cup \rho_2 \neq \mathcal{X}. \tag{16}$$

Note that in the special case of the trivial neighborhood system $\mathcal{N}_0$, that is, $N_x = \{x\}$ for all x, this condition is equivalent to the condition $\#\mathcal{X} > 2$, as required for the log score to be the only proper $\mathcal{N}_0$-local scoring rule; see Section 3.2.

Assume now that S is a proper scoring rule. For fixed $P \in \mathcal{P}$, $S(P, Q)$ is then minimized in Q, subject to $Q \in \mathcal{P}$, at $Q = P$. Introducing, for each P, a Lagrange multiplier $\lambda(P)$ for the constraint $\sum_x q_x = 1$, we must thus have

$$\sum_x p_x \frac{\partial}{\partial p_y} S(x, P) + \lambda(P) = 0 \qquad \text{for all } y \in \mathcal{X}. \tag{17}$$

In the case of a 0-homogeneous proper local scoring rule, we could without loss of generality assume that the neighborhood system $\mathcal{N} = \{N_x, x \in \mathcal{X}\}$ was determined by an undirected graph $\mathcal{G}$. In general this is not necessarily the case, as the following example shows.

EXAMPLE 5.1. A simple example that does not satisfy the condition is the neighborhood system determined by the undirected graph 1–2 3, where $\rho(1) = \rho(2) = \{1, 2\}$, $\rho(3) = \{3\}$. For this graph we can define a scoring rule as follows:

$$S(1, P) = S(2, P) = (1 - p_1 - p_2)^2,$$
$$S(3, P) = (1 - p_3)^2.$$

Then S is $\mathcal{G}$-local, and can easily be shown to be proper (it is an affine transformation of the Brier score for the event $X = 3$). However, its 0-homogeneous extension is $S(1, \boldsymbol{\alpha}) = S(2, \boldsymbol{\alpha}) = (\alpha_3/\alpha_+)^2$, $S(3, \boldsymbol{\alpha}) = \{(\alpha_1 + \alpha_2)/\alpha_+\}^2$, where $\alpha_+ := \alpha_1 + \alpha_2 + \alpha_3$. Thus the 0-homogeneous extension of S is not $\mathcal{G}$-local, and in particular not $\mathcal{G}$-additive.

For neighborhood systems $\mathcal{N}$ which satisfy Condition 5.1 we have the following lemma.

LEMMA 5.1. *Suppose S is proper and $\mathcal{N}$-local. If Condition 5.1 holds, then $\lambda(P)$ satisfying (17) is constant on $\mathcal{P}$.*

PROOF. For $\mathcal{N}$-local S, condition (17) gives for any $y \in \mathcal{X}$:

$$-\lambda(P) = \sum_{\{x : y \in N_x\}} p_x \frac{\partial S(x, P)}{\partial p_y} \tag{18}$$

as $\partial S(x, P)/\partial p_y = 0$ unless $y \in N_x$. For any term in the sum in (18), $S(x, P)$, and thus $\partial S(x, P)/\partial p_y$, depends only on $\mathbf{p}_{N_x}$, hence, since $y \in N_x$, only on $\mathbf{p}_{\rho(y)} = \{p_z : z \in \rho(y)\}$. Taking $y = y_1$, this implies that $\lambda(P)$ depends only on $\mathbf{p}_1 := \{p_z : z \in \rho_1\}$; similarly, $\lambda(P)$ depends only on $\mathbf{p}_2 := \{p_z : z \in \rho_2\}$.

By (16) we can take $w \in \mathcal{X} \setminus (\rho_1 \cup \rho_2)$. Starting at $\mathbf{p}$, consider a change $\delta\mathbf{p}_1$ to $\mathbf{p}_1$, such that $p_x + \delta p_x \in (0, 1)$ $(x \in \rho_1)$, and $\delta p_1^+ := \sum_{x \in \rho_1} \delta p_x \in (p_w - 1, p_w)$. Extend the variation $\delta\mathbf{p}_1$ to the whole of $\mathbf{p}$ by $\delta p_w = -\delta p_1^+$, $\delta p_x = 0$ (all other x). Then $\mathbf{p} + \delta\mathbf{p} \in \mathcal{P}$. Since $\lambda(\mathbf{p})$ depends only on $\mathbf{p}_2$, which has not changed, $\lambda(\mathbf{p} + \delta\mathbf{p}) = \lambda(\mathbf{p})$. But we can also express $\lambda(P)$ as $\lambda^*(\mathbf{p}_1)$, whence λ^* must be constant in an open neighborhood of $\mathbf{p}_1$. It follows that $\lambda(P)$ is constant on $\mathcal{P}$. $\square$

We now have that, under Condition 5.1, any $\mathcal{N}$-local proper scoring rule must satisfy:

For all $P \in \mathcal{P}$ and all $y \in \mathcal{X}$,

$$\sum_x p_x \frac{\partial}{\partial p_y} S(x, P) = -\lambda \tag{19}$$

for some scalar $\lambda \in \mathbb{R}$.

We note that a particular $\mathcal{N}$-local solution of (19) is given by the log-score, $S(x, P) = -\lambda \ln p_x$. Because (19) is linear, the general solution is thus $S = -\lambda \ln p_x + S_0$, where, for all $P \in \mathcal{P}$ and all $y \in \mathcal{X}$, S_0 satisfies the *key equation*:

$$\sum_x p_x \partial/\partial p_y S(x, P) = 0. \tag{20}$$

We thus can, and henceforth shall, restrict attention to such *key local* scoring rules. We next show that the 0-homogeneous extension of a key local scoring rule is $\mathcal{G}$-additive for a suitable undirected graph $\mathcal{G}$.

Let $H(P) := \sum_x p_x S(x, P)$ be the associated entropy function. Then (20) implies $S(y, P) = \partial H(P)/\partial p_y$. It follows that

$$\frac{\partial S(x, P)}{\partial p_y} = \frac{\partial S(y, P)}{\partial p_x}. \tag{21}$$

Hence if $y \in N_x$ but $x \notin N_y$, $\partial S(x, P)/\partial p_y$ must nevertheless vanish. Let $\mathcal{G}$ be the undirected graph in which x and y are neighbors if *both* $x \in N_y$ and $y \in N_x$. We call $\mathcal{G}$ the *symmetric core* of $\mathcal{N}$. Then any key $\mathcal{N}$-local proper scoring rule must in fact be $\mathcal{G}$-local. So we henceforth confine attention to $\mathcal{G}$-locality for an undirected graph $\mathcal{G}$ and assume that the neighborhoods are determined by $\mathcal{G}$ as $N_x = \{x\} \cup \mathrm{bd}(x)$.

LEMMA 5.2. *Under Condition 5.1, if S is a key $\mathcal{G}$-local scoring rule, its 0-homogeneous extension is $\mathcal{G}$-local.*

PROOF. With $P = \boldsymbol{\alpha}/\alpha_+$ we obtain by differentiation that, for any y,

$$\frac{\partial S(x,\boldsymbol{\alpha})}{\partial \alpha_y} = \frac{1}{\alpha_+}\left\{\frac{\partial S(x,P)}{\partial p_y} - \sum_x p_x \frac{\partial}{\partial p_x} S(y,P)\right\} = \frac{1}{\alpha_+}\frac{\partial S(x,P)}{\partial p_y}$$

as (20) and (21) imply that the second term within braces vanishes. Hence the result follows. □

Before we proceed to show that under Condition 5.1, any key $\mathcal{G}$-local scoring rule is $\mathcal{G}$-additive, the following result is useful.

LEMMA 5.3. *Suppose $x \neq y$, Condition 5.1 holds, and S is key $\mathcal{G}$-local. Then its entropy function H satisfies*

$$\frac{\partial^2 H(\boldsymbol{\alpha})}{\partial \alpha_x\, \partial \alpha_y} = 0.$$

PROOF. In this case, by Theorem 2.4, $\partial^2 H(\boldsymbol{\alpha})/\partial\alpha_x\,\partial\alpha_y = \partial S(x,\boldsymbol{\alpha})/\partial\alpha_y = 0$ by Lemma 5.2. □

The following lemma is straightforward:

LEMMA 5.4. *Let $f(\boldsymbol{\alpha})$ be a twice continuously differentiable function. Then $\partial^2 f(\boldsymbol{\alpha})/\partial\alpha_x\,\partial\alpha_y = 0$ if and only if*

$$f(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}_W, \alpha_x, \alpha_y^*) + f(\boldsymbol{\alpha}_W, \alpha_x^*, \alpha_y) - f(\boldsymbol{\alpha}_W, \alpha_x^*, \alpha_y^*) \tag{22}$$

for some (and then any) values α_x^*, α_y^*, where $W := \mathcal{X} \setminus \{x, y\}$.

We now proceed to establish $\mathcal{G}$-additivity for any key $\mathcal{G}$-local scoring rule:

THEOREM 5.5. *Let $\mathcal{C}$ be the set of maximal cliques of the graph $\mathcal{G}$ satisfying Condition 5.1, and let S be a key $\mathcal{G}$-local scoring rule. Then we can express the scoring rule and associated entropy function as*

$$S(\boldsymbol{\alpha}) = \sum_{C\in\mathcal{C}} s_C(\boldsymbol{\alpha}_C), \qquad H(\boldsymbol{\alpha}) = \sum_{C\in\mathcal{C}} h_C(\boldsymbol{\alpha}_C). \tag{23}$$

Further, if H is regular, each term h_C in the expansion can be chosen to be 1-homogeneous.

PROOF. The proof parallels that of the Hammersley–Clifford theorem as given in Grimmett (1973); see also Lauritzen (1996), page 36. For a fixed $\boldsymbol{\alpha}^*$ and all subsets $A \subseteq \mathcal{X}$ we define

$$\eta_A(\boldsymbol{\alpha}_A) = H(\boldsymbol{\alpha}_A, \boldsymbol{\alpha}^*_{\mathcal{X}\setminus A}) \tag{24}$$

and note that then $H(\boldsymbol{\alpha}) = \eta_{\mathcal{X}}(\boldsymbol{\alpha})$. Next, we define for all $B \subseteq \mathcal{X}$

$$h_B(\boldsymbol{\alpha}_B) = \sum_{A\,:\,A\subseteq B} (-1)^{|B\setminus A|} \eta_A(\boldsymbol{\alpha}_A). \tag{25}$$

The Möbius inversion formula [see, e.g., Lauritzen (1996), page 239] then yields: for all $A \subseteq \mathcal{X}$,

$$\eta_A(\boldsymbol{\alpha}_A) = \sum_{B\,:\,B\subseteq A} h_B(\boldsymbol{\alpha}_B).$$

Thus, taking $A = \mathcal{X}$, we have established

$$H(\boldsymbol{\alpha}) = \sum_{B\,:\,B\subseteq \mathcal{X}} h_B(\boldsymbol{\alpha}_B). \tag{26}$$

We next show that all terms h_B in (26) vanish unless B is a complete set and (23) then follows by collecting appropriate terms. So suppose there exist $x, y \in B$ with $x \not\sim y$. We then let $D = B \setminus \{x, y\}$ and write the expression in (25) as

$$h_B(\boldsymbol{\alpha}_B) = \sum_{A\,:\,A\subseteq D} (-1)^{|D\setminus A|} \bigl(\eta_A - \eta_{A\cup\{x\}} - \eta_{A\cup\{y\}} + \eta_{A\cup\{x,y\}}\bigr),$$

where we have abbreviated $\eta_U := \eta_U(\boldsymbol{\alpha}_U) = H(\boldsymbol{\alpha}_U, \alpha^*_{\mathcal{X}\setminus U})$, etc. But each of the terms in this expansion vanishes by Lemma 5.4 and hence h_B vanishes as required.

If H is regular, we can choose $\alpha^*_x = 0$ for all $x \in \mathcal{X}$ so each function η_A in (24) and hence each h_B in (25) will be 1-homogeneous.

This establishes the desired expansion of the entropy function. The expansion for the scoring rule is obtained by forming gradients. $\square$

It does not seem to be true in general that each term in (23) can also be chosen to be concave. If this is indeed the case, S and H are built up additively from proper scoring rules and entropy functions defined on cliques.

6. Summary and discussion. We have defined a proper scoring rule $S(x, P) = s(x, \mathbf{p})$ for discrete sample space $\mathcal{X}$ to be local relative to a neighborhood system $\mathcal{N} = \{N_x\}_{x\in\mathcal{X}}$ if each $s(x, \mathbf{p})$ only depends on $\mathbf{p}$ through its restriction $\mathbf{p}_{N_x}$ to N_x, and shown how to construct such scoring rules from additive components. Conversely, we have shown that under appropriate regularity conditions any proper local scoring rule has this structure, although the additive components may not in general each correspond to proper scoring rules.

A definition of homogeneous local scoring rule for a general well-behaved topological outcome space $\mathcal{X}$ that would unify the discrete and continuous case would be to say that a scoring rule is homogeneous and local if it satisfies

$$S(x, P) = S\{x, P(\cdot\,|N_x)\} \tag{27}$$

for every open neighborhood N_x of x and every $x \in \mathcal{X}$.

Clearly, the homogeneous local scoring rules investigated in Parry, Dawid and Lauritzen (2012) satisfy this requirement. It would be interesting to obtain a complete characterization of proper scoring rules satisfying (27).

Acknowledgment. We are grateful to Valentina Mameli for comments on an earlier version of this article.

REFERENCES

BERNARDO, J.-M. (1979). Expected information as expected utility. *Ann. Statist.* **7** 686–690. MR0527503

BESAG, J. (1975). Statistical analysis of non-lattice data. *J. Roy. Statist. Soc. Ser. D* (*The Statistician*) **24** 179–195.

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

CZADO, C., GNEITING, T. and HELD, L. (2009). Predictive model assessment for count data. *Biometrics* **65** 1254–1261. MR2756513

DAWID, A. P. (1986). Probability forecasting. In *Encyclopedia of Statistical Sciences* (S. Kotz, N. L. Johnson and C. B. Read, eds.) **7** 210–218. Wiley, New York.

DAWID, A. P. (2007). The geometry of proper scoring rules. *Ann. Inst. Statist. Math.* **59** 77–93. MR2396033

DAWID, A. P. and LAURITZEN, S. L. (2005). The geometry of decision theory. In *Proceedings of the Second International Symposium on Information Geometry and Its Applications* 22–28. Univ. Tokyo, Tokyo, Japan.

EHM, W. and GNEITING, T. (2012). Local proper scoring rules of order two. *Ann. Statist.* **40** 609–637.

GOOD, I. J. (1971). Comment on "Measuring information and uncertainty," by Robert J. Buehler. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) 337–339. Holt, Rinehart and Winston, Toronto.

GRIMMETT, G. R. (1973). A theorem about random fields. *Bull. Lond. Math. Soc.* **5** 81–84. MR0329039

HENDRICKSON, A. D. and BUEHLER, R. J. (1971). Proper scores for probability forecasters. *Ann. Math. Statist.* **42** 1916–1921. MR0314430

HYVÄRINEN, A. (2007). Some extensions of score matching. *Comput. Statist. Data Anal.* **51** 2499–2512. MR2338984

LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. Clarendon Press, Oxford, UK. MR1419991

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

PARRY, M. F., DAWID, A. P. and LAURITZEN, S. L. (2012). Proper local scoring rules. *Ann. Statist.* **40** 561–592.

ROCKAFELLAR, R. T. (1970). *Convex Analysis. Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton, NJ. MR0274683

ZELTERMAN, D. (1988). Robust estimation in truncated discrete distributions with application to capture–recapture experiments. *J. Statist. Plann. Inference* **18** 225–237. MR0922210

A. P. DAWID
STATISTICAL LABORATORY
CENTRE FOR MATHEMATICAL SCIENCES
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD
CAMBRIDGE CB3 0WB
UNITED KINGDOM
E-MAIL: A.P.Dawid@statslab.cam.ac.uk

S. LAURITZEN
DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
SOUTH PARKS ROAD
OXFORD OX1 3TG
UNITED KINGDOM
E-MAIL: steffen@stats.ox.ac.uk

M. PARRY
DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF OTAGO
P.O. BOX 56
DUNEDIN 9054
NEW ZEALAND
E-MAIL: mparry@maths.otago.ac.nz