## Martin Jacobsen

## Discrete Exponential Families: <br> Deciding when the Maximum Likelihood Estimator Exists and is Unique



## Martin Jacobsen

DISCRETE EXPONENTIAL FAMILIES: DECIDING WHEN THE MAXIMUM LIKELIHOOD ESTIMATOR EXISTS AND IS UNIQUE

Preprint 1988 No. 1

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

February 1988

## Summary

A necessary and sufficient condition is given, for the maximum likelihood estimator in some discrete exponential families to exist and be unique. The condition does not rely on a sufficient reduction of the observations. Several examples are given including the Cox regression model and the dose-response, Bradley-Terry and Rasch models.

AMS Subject Classification 62G05, 62AlO.

Key words: Cox regression model, dose-response model, Bradley-Terry model, Rasch model, exponential family, maximum likelihood estimator.

## 1. Introduction

Let $X$ be a random variable with values in a measurable space ( $E, E$ ), and suppose that the distribution of $X$ belongs to a full exponential family of distributions, i.e. the distribution of $X$ has a density $f_{\theta}$ with respect to a given $\sigma$-finite measure on ( $E, E$ ), which has the form

$$
f_{\theta}(x)=\frac{1}{\varphi(\theta)} \exp (\langle\theta, T(x)\rangle+b(x)), \quad(x \in E)
$$

where $T: E \rightarrow \mathbb{R}^{p}, b: E \rightarrow \mathbb{R}$ and $\theta \in \theta \subset \mathbb{R}^{p}$ with

$$
\theta=\left\{\theta \in \mathbb{R}^{\mathrm{p}}: \varphi(\theta)=\int \exp (\langle\theta, \mathrm{T}(\mathrm{x})\rangle+\mathrm{b}(\mathrm{x})) \mu(\mathrm{dx})\langle\infty\}\right.
$$

Here 〈•,•〉 denotes the inner product on $\mathbb{R}^{p},\langle y, z\rangle=\sum_{k=1}^{p} y_{k} z_{k}$ for $y=\left(y_{1}, \cdots, y_{p}\right), z=\left(z_{1}, \cdots, z_{p}\right)$.

Suppose one observation $x$ of $X$ is availabe. Barndorff-Nielsen's [2, Theorem 9.13] main result on estimating $\theta$ on the basis of $x$ states that the maximum-likelihood estimator $\hat{\theta}=\hat{\theta}(x)$ exists and is unique if and only if the observed value $T(x)$ of the sufficient statistic belongs to the interior of the convex hull of the support of the measure $\mu T^{-1}$ on $\mathbb{R}^{p}$.

Especially for discrete exponential families, it can be difficult to verify whether Barndorff-Nielsen's condition is satisfied. The purpose of this paper is to present an alternative condition, and to show by a num-
ber of examples, how the condition works in practice. It should be stressed that our condition does not rely on first transforming the observation by the sufficient statistic $T$. The first example involves the Cox regression model, where a necessary and sufficient condition is given for Cox's partial likelihood to attain its maximal value at a unique point. Amazingly enough, this condition does not seem to have been noticed before.

## 2. ML-estimation in some discrete exponential families

Suppose $X$ is a discrete random variable with values in a finite or countable set $E$. Assume that the distribution of $X$ belongs to a full exponential family, i.e.

$$
\begin{equation*}
\mathrm{P}_{\theta}(\mathrm{X}=\mathrm{x})=\frac{1}{\varphi(\theta)} \exp (\langle\theta, \mathrm{T}(\mathrm{x})\rangle+\mathrm{b}(\mathrm{x})) \tag{2.1}
\end{equation*}
$$

with $T: E \rightarrow \mathbb{R}^{p}, b: E \rightarrow \mathbb{R}$. To avoid technicalities we shall assume the parameter space to be all of $\mathbb{R}^{\mathbb{P}}$, so that

$$
\varphi(\theta)=\sum_{x \in E} \exp (\langle\theta, T(x)\rangle+b(x))\langle\infty
$$

for all $\theta \in \mathbb{R}^{p}$.

Based on an observation $x$ of $X$, we wish to estimate $\theta$. Barndorff-Nielsen's Theorem [2] shows that the ML-estimator $\hat{\theta}=\hat{\theta}(x)$ of $\theta$ exists and is unique if and only if the observed value $T(x)$ of the sufficient statistic belongs to the interior of the convex hull in $\mathbb{R}^{p}$
of the points $T(y), y \in E$.

Inspired by the first example below, the Cox regression model, we shall present an alternative condition. The idea is to write the likelihood function (2.1) in the form
(2.2) $\quad P_{\theta}(X=x)=\left[1+\sum_{y \neq x} \exp (\langle\theta, T(y)-T(x)\rangle+b(y)-b(x))\right]^{-1}$
and then exploit some simple analytic properties of functions of this form.

If $X=\left(X_{1}, \cdots, X_{K}\right)$ is a vector of independent components where the distribution of each $X_{i}$ belongs to an exponential family, it will prove useful to write the total likelihood as a product of functions of the form appearing on the right of (2.2).

Thus, consider a function $f: \mathbb{R}^{p} \rightarrow(0, \infty)$ of the form

$$
\begin{equation*}
f(\theta)=\prod_{i=1}^{K}\left[1+\sum_{j \in A_{i}} a_{i j} \exp \left\langle\theta, v_{i j}\right\rangle\right]^{-1} \tag{2.3}
\end{equation*}
$$

for $K$ a positive integer, $A_{i}$ for each $i$ finite or countably infinite index set, all $a_{i j}>0$ and $v_{i j} \in \mathbb{R}^{p}$. The vectors $v_{i j}$ will be called the structure vectors for $f$.

As it stands $f$ is defined for all $\theta \in \mathbb{R}^{p}$, but to ensure that
f $>0$, we make the following

Assumption For all $i$, the series

$$
\sigma_{i}(\theta)=\sum_{j \in A_{i}} a_{i j} \exp \left\langle\theta, v_{i j}\right\rangle
$$

converges for all $\theta \in \mathbb{R}^{p}$.

We note the following facts, easily understood for instance, as properties of Laplace transforms:

The series $\sigma_{i}(\theta)$ converges for all $\theta \in \mathbb{R}^{p}$ iff the series

$$
\sum_{j \in A_{i}} a_{i j} e^{B\left\|v_{i j}\right\|}
$$

converges for all $B \geq 0$. In that case $\sigma_{i}$ is twice differentiable with gradient

$$
\begin{equation*}
D \sigma_{i}(\theta)=\sum_{j \in A_{i}} a_{i j} e^{\left\langle\theta, v_{i j}\right\rangle} v_{i j} \tag{2.4}
\end{equation*}
$$

and Hessian

$$
\begin{equation*}
D^{2} \sigma_{i}(\theta)=\sum_{j \in A_{i}} a_{i j} e^{\left\langle\theta, v_{i j}\right\rangle} v_{i j} \otimes 2 \tag{2.5}
\end{equation*}
$$

Notation: for $v=\left(v_{1}, \cdots, v_{p}\right) \in \mathbb{R}^{p},\|v\|=\langle v, v\rangle^{1 / 2}$ is the Euclidean
norm and $\mathrm{v}^{\otimes 2}$ is the $\mathrm{p} \times \mathrm{p}$ matrix $\left(\mathrm{v}_{\mathrm{k}} \mathrm{v}_{\mathrm{l}}\right), 1 \leq \mathrm{k}, \mathrm{l} \leq \mathrm{p}$.

Returning now to (2.3) it is clear that $f>0$ is twice differentiable. Introducing $l=\log f$ the properties of $f$ that we shall use may be summarized as follows, where given a collection of vectors $v \in \mathbb{R}^{p}$, span\{v\} (conv\{v\}) denotes the linear subspace (convex hull) spanned by the collection, and where for $A$ a $p \times p$ symmetric matrix, $A>0$ means that $A$ is positive definite.

Theorem (a) The function $l=\log f$ is concave. It is strictly concave if and only if either of the following three conditions is satisfied:
(a i)

$$
\begin{equation*}
\operatorname{span}\left\{v_{i j}: i=1, \cdots, K, j \in A_{i}\right\}=\mathbb{R}^{p} \tag{aii}
\end{equation*}
$$

there exists $\theta \in \mathbb{R}^{p}$ such that $-D^{2} l(\theta)>0$,
(a iii) for all $\theta \in \mathbb{R}^{p}, \quad-D^{2} l(\theta)>0$,
(b) The function $f$ attains its maximal value at a unique point $\hat{\theta} \in \mathbb{R}^{p}$ if and only if either of the following two conditions is satisfied:
(b i)
(b ii)
$0 \in \operatorname{int} \operatorname{conv}\left\{\mathrm{v}_{\mathrm{i} j}: i=1, \cdots, K, j \in A_{i}\right\}$, there does not exist $\alpha \in \mathbb{R}^{p}$ with $\alpha \neq 0$ such that for all $\mathrm{i}, \mathrm{j},\left\langle\alpha, \mathrm{v}_{\mathrm{ij}}\right\rangle \geq 0$.

In particular, if either condition is satisfied, $l$ is strictly concave.

A proof is sketched in the appendix. The reader should consult Albert and

Anderson [1] for results very similar in appearance to this theorem. They consider likelihood functions for independent observations from different multinomial distributions and are led to study the behaviour of functions given by an expression similar to (2.2), but where differences between parameter vectors rather than differences between $T$-values appear in the exponents. Nevertheless, their results on existence and uniqueness of maximum likelihood estimators certainly follow from ours. It should be stressed however that the statistical emphasis in [1] and this paper, concerns two very different matters.

The reader is reminded, that in the examples to follow, functions of the form (2.3) appear as products of exponential family likelihoods with each factor looking like the right hand side of (2.2). Of course that expression itself is (2.3) with $K=1, A_{1}=E \backslash\{x\}$, $a_{1 j}=$ $\exp (b(j)-b(x)), v_{1 j}=T(j)-T(x)$.

## 3. Examples

### 3.1. The Cox regression model

In its simplest form, the Cox model (Cox [4]), models N independent strictly positive random variables $Y_{1}, \cdots, Y_{N}$, such that

$$
\mathrm{P}\left(\mathrm{Y}_{v}>\mathrm{t}\right)=\exp \left[-\int_{0}^{\mathrm{t}} \lambda(\mathrm{~s}) \mathrm{e}^{\left\langle\theta, \mathrm{z}_{v}(\mathrm{~s})\right\rangle} \mathrm{ds}\right]
$$

with $\lambda$ an unknown baseline hazard, $\theta \in \mathbb{R}^{\mathbf{p}}$ an unknown vector of regression parameters, and, for each $v$ and $s, z_{v}(s)$ a p-vector of
observed, time dependent covariates.

The more sophisticated versions may allow for censorings or truncations, and the covariates may be random. Always however, inference about $\theta$ is made on the basis of Cox's partial likelihood $C(\theta)$ (Cox $[4,5]$ ) which has the following form: on a possibly random interval of observation, $K \leq N$ of the values of the $Y_{v}$ are observed, $0<t_{1}<\cdots<t_{K}$ say, and then

$$
C(\theta)=\prod_{i=1}^{K} \frac{\exp \left\langle\theta, z_{v_{i}}\left(\mathrm{t}_{\mathrm{i}}\right)\right\rangle}{\sum_{v \in R_{i}} \exp \left\langle\theta, z_{v}\left(\mathrm{t}_{\mathrm{i}}\right)\right\rangle}
$$

with $v_{i}$ the $v$ for which $Y_{v}=t_{i}$, and $R_{i}$ the risk set of $v$ under observation just before time $t_{i}$, in particular $v_{i} \in R_{i}$ always.

Rewriting $C$ in the form (2.3) gives

$$
\mathrm{C}(\theta)=\prod_{\mathrm{i}=1}^{\mathrm{K}}\left[1+\sum_{v \in \mathrm{R}_{\mathrm{i}} \backslash v_{i}} \exp \left\langle\theta, z_{v}\left(\mathrm{t}_{\mathrm{i}}\right)-{z_{v_{i}}}\left(\mathrm{t}_{\mathrm{i}}\right)\right\rangle\right]^{-1}
$$

so the structure vectors are the contrast covariate vectors $z_{v}\left(\mathrm{t}_{\mathrm{i}}\right)-\mathrm{z}_{v_{\mathrm{i}}}\left(\mathrm{t}_{\mathrm{i}}\right)$ for $\mathrm{i}=1, \cdots, \mathrm{~K}, v \in \mathrm{R}_{\mathrm{i}} \backslash v_{\mathrm{i}}$. The Theorem now shows that $\log C$ is concave always and strictly concave iff

$$
\operatorname{span}\left\{z_{v}\left(\mathrm{t}_{\mathrm{i}}\right)-\mathrm{z}_{v_{i}}\left(\mathrm{t}_{\mathrm{i}}\right): \mathrm{i}=1, \cdots, K, v \in \mathrm{R}_{\mathrm{i}} \backslash v_{\mathrm{i}}\right\}=\mathbb{R}^{p}
$$

Further, the Cox estimator that maximizes $C$, exists and is unique iff

$$
0 \in \operatorname{int} \operatorname{conv}\left\{z_{v}\left(t_{i}\right)-z_{v_{i}}\left(t_{i}\right): i=1, \cdots, K, v \in R_{i} \backslash v_{i}\right\}
$$

or equivalently, there is no $\alpha \in \mathbb{R}^{p}, \alpha \neq 0$ such that for all $i$, $v \in \mathrm{R}_{\mathrm{i}} \backslash v_{\mathrm{i}}$

$$
\left\langle\alpha, z_{v}\left(\mathrm{t}_{\mathrm{i}}\right)-\mathrm{z}_{\mathrm{v}_{\mathrm{i}}}\left(\mathrm{t}_{\mathrm{i}}\right)\right\rangle \geq 0 .
$$

Thus, for estimating $\theta$, the behaviour of the contrast covariate vectors $z_{v}\left(t_{i}\right)-z_{v_{i}}\left(t_{i}\right)$ for $i=1, \cdots, K, v \in R_{i} \backslash v_{i}$ is crucial, and it is seen that $\theta$ may be estimated uniquely by maximizing the partial likelihood, except in the cases where there is some linear combination of the $p$ covariates such that at each $t_{i}$, the value of the linear combination for $v_{i}$ exceeds or equals the value for all other $v \in R_{i} \backslash v_{i}$.

It is known of course, that in a survival study with, say, age as one covariate, $\theta_{\text {age }}$ cannot be estimated if it is always the oldest individual in the risk group that is observed to die. We now see that as a consequence of the mathematical structure of the partial likelihood, with e.g. blood pressure a second covariate, $\theta$ cannot be estimated if always the individual in the risk set, with, say, the smallest value of $3 \times$ age minus $7 \times$ blood pressure, is observed to die!

An example, where the Cox-analysis does not yield a unique estimator for $\theta$ is recorded in Bryson and Johnson [3]. व

All the following examples involve genuine exponential models, where we discuss which observations are extreme, i.e. are such that the ML-estimator either does not exist, or if it does, is not unique.
3.2 The dose-response model. Let $d_{1}<\cdots<d_{k}$ be real numbers and for $\mathrm{i}=1, \cdots, k$ consider $n_{i} \geq 1$ i.i.d. $0-1$ valued random variables $X_{i j}, j=1, \cdots, n_{i}, \quad$ such that

$$
\begin{equation*}
P\left(X_{i j}=1\right)=1-P\left(X_{i j}=0\right)=\frac{\exp \left(\theta_{1}+\theta_{2} d_{i}\right)}{1+\exp \left(\theta_{1}+\theta_{2} d_{i}\right)} \tag{3.1}
\end{equation*}
$$

with all $X_{i j}$ mutually independent.

From Barndorff-Nielsen [2] or Larsen [10] it is known that an observation vector $\left(x_{i j}\right)$ is extreme iff there is a critical $i_{0}$ such that either

$$
x_{i j}=0 \text { for } i<i_{0}, \text { all } j, \quad x_{i j}=1 \text { for } i>i_{0} \text {, all } j
$$

or

$$
x_{i j}=1 \text { for } i<i_{0}, \text { all } j, \quad x_{i j}=1 \text { for } i>i_{0} \text {, all } j
$$

with arbitrary values allowed for $x_{i_{0}}, j=1, \cdots, n_{i_{0}}$.

The proof is based on Barndorff-Nielsen's theorem quoted above and involves drawing a convex polygon with $2 k$ sides in $\mathbb{R}^{2}$ and deciding
whether $T\left(\left(x_{i j}\right)\right)$ is in the interior of this polygon, with $T\left(\left(x_{i j}\right)\right)=$ $\left(\underset{i, j}{ } \mathrm{x}_{\mathrm{ij}},{ }_{\mathrm{i}, \mathrm{j}} \mathrm{d}_{\mathrm{i}} \mathrm{x}_{\mathrm{ij}}\right)$ the observed value of the minimal sufficient statistic.

We shall now see how our Theorem may be used to give an easy proof. The likelihood is written in the form (2.3) with each $\mathrm{x}_{\mathrm{ij}}$ contributing one factor which, cf. (3.1), is

$$
\begin{array}{ll}
{\left[1+\exp \left(-\theta_{1}-\theta_{2} d_{i}\right)\right]^{-1}} & \text { if } \\
x_{i j}=1 \\
{\left[1+\exp \left(\theta_{1}+\theta_{2} d_{i}\right)\right]^{-1}} & \text { if } \\
x_{i j}=0
\end{array}
$$

Thus, referring to (2.3), index $i$ there is the double index $i j, K=\Sigma n_{i}$ and each $A_{i j}$ contains one element with structure vector $\mathrm{v}_{\mathrm{ij}}$,

$$
\begin{aligned}
v_{i j} & =(-1, \\
v_{i j} & =\left(d_{i}\right) \\
1, & \text { if }
\end{aligned} x_{i j}=1 .
$$

Immediately, using part (a) of the Theorem, it is seen that if $k \geq 2$ the log-likelihood is always strictly concave. Further, since all vectors $v_{i j}$ sit on the two parallel lines $(-1, t)_{t \in \mathbb{R}}$ and $(1, t)_{t \in \mathbb{R}}$ in $\mathbb{R}^{2}$, appealing to Figure 1 and using (b i) from the Theorem, it is seen that there is a unique ML-estimator iff

$$
\mathrm{d}_{1+}+\mathrm{d}_{0+}>0, \quad \mathrm{~d}_{1-}+\mathrm{d}_{0-}<0
$$

with $d_{1+}\left(d_{1-}\right)$ the largest (smallest) value of $-d_{i}$ for $i$ such that
$\mathbf{x}_{\mathbf{i j}}=1$ for some j , and similarly $\mathrm{d}_{0+}\left(\mathrm{d}_{0_{-}}\right)$the largest (smallest) $d_{i}$ for $i$ such that $x_{i j}=0$ for some $j$.


Fig. 1 The structure vectors for the dose-response model.

It is not difficult to see that this criterion for existence and uniqueness of the ML-estimator is equivalent to that given by Barndorff-Nielsen and Larsen. Rather than present the argument, we only mention that the equivalence is established even more easily using (bii) of the Theorem, as the following more elaborate example will show. Our point in proceeding via (b i) has been to show the difference between our approach and that of Barndorff-Nielsen (deciding whether a given point belongs to the interior of a trapezoid determined by the observations as opposed to deciding whether a point determined from the observations belongs to a given convex polygon with 2 k sides).
3.3. Logistic regression with 2 independent variables. Let $d_{1}<\cdots<d_{k}$ as before and let $e_{1} \prec \cdots<e_{m}$. For each $i=1, \cdots, k, j=1, \cdots, m$ let $X_{i j \mu}$ for $\mu=1, \cdots, n_{i j}$ be i.i.d. $0-1$ valued random variables with

$$
P\left(X_{i j \mu}=1\right)=1-P\left(X_{i j \mu}=0\right)=\frac{\exp \left(\theta_{1}+\theta_{2} d_{i}+\theta_{3} e_{j}\right)}{1+\exp \left(\theta_{1}+\theta_{2} d_{i}+\theta_{3} e_{j}\right)}
$$

and all $X_{i j \mu}$ independent.

There need not be observations corresponding to all km combinations $\left(d_{i}, e_{j}\right)$, so we allow for $n_{i j}=0$. However, to ensure that observations are present for each $\mathrm{d}_{\mathrm{i}}$ and each $\mathrm{e}_{\mathrm{j}}$, we assume $\underset{\mathrm{j}}{\mathrm{in}} \mathrm{i}_{\mathrm{j}}>0$ for all i, $\sum_{i} \mathrm{n}_{\mathrm{ij}}>0$ for all j.

Let ( $\mathrm{x}_{\mathrm{i} j \mu}$ ) denote the observation vector. Proceeding as in the previous example, the likelihood is written in the form (2.3), each $x_{i j \mu}$ contributing one factor with the structure vectors $v_{i j \mu} \in \mathbb{R}^{3}$ for $i=$ $1, \cdots, \mathrm{k}, \mathrm{j}=1, \cdots, \mathrm{~m}, \mu=1, \cdots, \mathrm{n}_{\mathrm{ij}}$ given by

$$
\mathrm{v}_{\mathrm{i} j \mu}=\left[\begin{array}{lll}
(-1, & \mathrm{d}_{\mathrm{i}},- & \left.\mathbf{e}_{\mathrm{j}}\right) \\
(1, & \text { if } \mathrm{x}_{\mathrm{i} j \mu}=1 \\
( & \left.\mathbf{e}_{\mathrm{j}}\right) & \text { if } \mathrm{x}_{\mathrm{i} j \mu}=0
\end{array}\right.
$$

Suppose ( $\mathrm{x}_{\mathrm{i} j \mu}$ ) is an extreme observation, so that there is not a unique ML-estimator. By (b ii) from the Theorem, we can find $\alpha=\left(\alpha_{1}, \alpha_{2}, \alpha_{3}\right) \neq 0$ such that $\left\langle\alpha, \mathrm{v}_{\mathbf{i j} \mu}\right\rangle \geq 0$ for all $\mathrm{i}, \mathrm{j}, \mu$, i.e. for all $\mathrm{i}, \mathrm{j}$ with $\mathrm{n}_{\mathrm{i} j} \geq$ 1 ,

$$
\alpha_{1}+\alpha_{2} \mathrm{~d}_{\mathrm{i}}+\alpha_{3} \mathrm{e}_{\mathrm{j}}\left[\begin{array}{ll}
\leq 0 & \text { if } \mathrm{x}_{\mathrm{i} j \mu}=1 \text { for all } \mu  \tag{3.2}\\
\geq 0 & \text { if } \mathrm{x}_{\mathrm{i} j \mu}=0 \text { for all } \mu \\
=0 & \text { if there is } \mu, \mu^{\prime} \text { such that } \\
& \mathrm{x}_{\mathrm{i} j \mu}=1, \mathrm{x}_{\mathrm{i} j \mu^{\prime}}=0
\end{array} .\right.
$$

Of course, if all $\mathrm{x}_{\mathrm{i} j \mu}=1$ (or = 0), the observation is extreme, so assume at least one $x_{i j \mu}=1$ and one $x_{i j \mu}=0$. Then, because of
(3.2), $\alpha_{2}=\alpha_{3}=0$ would force $\alpha_{1}=0$ which is impossible, hence $\left(\alpha_{2}, \alpha_{3}\right) \neq(0,0)$ and the equation $\alpha_{1}+\alpha_{2} d+\alpha_{3} e=0$ defines a straight line in the (d,e)-plane. Since $\alpha_{1}+\alpha_{2} d+\alpha_{3} e>0$ for all points (d,e) on one side and $<0$ for all (d,e) on the other side of the line, it emerges that an extreme observation ( $\mathrm{x}_{\mathrm{i} j \mu}$ ) has the structure shown in Figure 2 , where each $\mathrm{x}_{\mathrm{ij} \mu}$ is attached to the point $\left(\mathrm{d}_{\mathrm{i}}, \mathrm{e}_{\mathrm{j}}\right)$.


Fig. 2 The extreme observations for logistic regression.

If conversely, the observation pattern is as shown in Fig. 2, trivially $\alpha \neq 0$ can be found such that (3.2) holds. Thus we have shown that an observation ( $\mathrm{x}_{\mathrm{i} j \mu}$ ) is extreme iff either for all $\mathrm{i}, \mathrm{j}, \mu \mathrm{x}_{\mathrm{i} j \mu}=$ 1 (or $=0$ ) or else there is a straight line in the ( $\mathrm{d}, \mathrm{e}$ )-plane separating $\mathrm{x}_{\mathrm{i} j \mu}=1$ from $\mathrm{x}_{\mathrm{i} j \mu}=0$ with arbitrary $0-1$ configurations allowed on the line itself.

In particular, in order for the ML-estimator to be unique, it is (of course) necessary that the points $\left(d_{i}, e_{j}\right)$ with $n_{i j} \geq 1$ do not lie on a straight line.

The generalization to more than 2 independent variables is immediate.

### 3.4 The Bradley-Terry model (See Zermelo [12] for the model and an

 early derivation of the criterion obtained below for an observation to be extreme). In a tournament with $k$ players, when $i$ plays against $j$, $i$ wins with probability$$
\begin{equation*}
\frac{e^{\theta_{i}}}{e^{\theta_{i}}+e^{\theta^{\prime}}} \tag{3.3}
\end{equation*}
$$

and loses (j wins) with the complementary probability.

We assume that $i$ plays against $j n_{i j}$ times (so $n_{i j}=n_{j i}$ ) with the outcomes of all games mutually independent. We allow for some $n_{i j}$ to be 0 but assume the tournament to be closed: for all $i \neq j$ there exists $i=i_{0} \neq i_{1} \neq \cdots \neq i_{m-1} \neq i_{m}=j$, such that $n_{i_{\mu-1}} i_{\mu} \geq 1$ for $\mu=1, \cdots, \mathrm{~m}$.

The probabilities (3.3) are unchanged if $\theta=\left(\theta_{1}, \cdots, \theta_{k}\right)$ is replaced by $\left(\theta_{1}+c, \cdots, \theta_{k}+c\right)$ for an arbitrary $c$. We therefore assume from now on that $\theta_{k}=0$ and note that since the tournament is closed the parametrization of the model is unique and it makes sense to ask whether for a given observation, there is a unique ML-estimator of $\left(\theta_{1}, \cdots, \theta_{k-1}\right)$.

The likelihood is of the form (2.3) with factors of the form

$$
\left(1+e^{\theta} j^{-\theta} i\right)^{-1} \text { if } i \text { has played against } j \text { and won, }
$$

this factor corresponding to the structure vector $v \in \mathbb{R}^{k-1}$

$$
\begin{gathered}
\mathbf{i} \quad \mathbf{j} \\
(\cdots,-1, \cdots, 1, \cdots)
\end{gathered}
$$

with all entries not shown equal to 0 , and where, if $i=k$, respectively $j=k$, the -1 , respectively 1 , component is ignored.

Consider now an extreme observation, so that by (bii) of the Theorem there exists $\alpha=\left(\alpha_{1}, \cdots, \alpha_{k-1}\right) \neq 0$ with all $\langle\alpha, \mathrm{v}\rangle \geq 0$. Introducing $\alpha_{k}=0$ it is seen that this is equivalent to demanding that for all pairs (i,j) of players with $i<j$,
$\alpha_{i} \leq \alpha_{j}$ if $i$ has won all games against $j$
$\alpha_{i} \geq \alpha_{j}$ if $i$ has lost all games against $j$
$\alpha_{i}=\alpha_{j}$ if $i$ has both won and lost a game against $j$.

Ordering the players in a sequence $\left(i_{1}, \cdots, i_{k}\right)$ such that $\alpha_{i_{1}} \leq \cdots \leq \alpha_{i_{k}}$ (with the 0 for $\alpha_{k}$ included), since $\left(\alpha_{1}, \cdots, \alpha_{k-1}\right) \neq 0$ there must be a sharp inequality somewhere. With $\alpha_{i_{l-1}}<\alpha_{i_{l}}$, define $A=\left\{i_{1}, \cdots, i_{l-1}\right\}, \quad B=\left\{i_{l}, \cdots, i_{k}\right\}$, thus partitioning the set of players into two non-empty subsets. Clearly from (3.4), any game between $i \in A, j \in B$ is won by $i \in A$, and the results of the tournament must have the form displayed in Figure 3, where
$a+(-)$ in the $i^{\prime}$ th row, $j^{\prime}$ th column signifies that all games played between $i$ and $j$ have been won (lost) by $i$, while the symbol "any" allows for arbitrary results in the games between the players concerned.


Fig. 3 The extreme observations for the Bradley-Terry model.

If conversely the observation is of the form in Fig.3, if e.g. player $k \in B$, let $\beta<0$ and define $\left(\alpha_{1}, \cdots, \alpha_{k-1}\right)$ by $\alpha_{i}=\beta$ if ifA, $\alpha_{i}=0$ if $i \in B$. Then, with $v$ the structure vector for a game played between $i$ and $j,\langle\alpha, v\rangle=0$ if both $i, j \in A$ or both $i, j \in B$, while $\langle\alpha, v\rangle=-\beta\rangle 0$ if $i \in A, j \in B$, hence the observation is extreme.

Thus, for a closed tournament, an observation is extreme iff it is possible to divide the set of players into two non-empty subsets A and B such that all games between $i \in A, j \in B$ are always won by $i$.

Special cases of extreme observations: one player wins (respectively loses) all his games.

Also note that for a non-closed tournament, any observation fits with the pattern in Fig. 3: since the tournament is not closed, it is possible to find $A \neq \varnothing$ with $A^{c} \neq \varnothing$ such that $i \in A$ never plays against $j \in A^{c}$.
3.5 The Rasch model. Rasch [11] proposed the following model for evaluating questionnaires where $k$ persons answer the same $m$ questions: for $i=1, \cdots, k, j=1, \cdots, m$ the probability that person $i$ answers question $j$ correctly (wrongly) is

$$
\begin{equation*}
\frac{e^{\theta_{i}+\xi_{j}}}{1+e^{\theta_{i}+\xi_{j}}} \quad\left[\frac{1}{1+e^{\theta_{i}+\xi_{j}}}\right] \tag{3.5}
\end{equation*}
$$

all responses being mutually independent. Thus, there is a binary random variable attached to each pair (i,j) and here we shall allow for $n_{i j} \geq 0$ i.i.d. copies of that binary response to be observed, generalizing the case $n_{i j}=1$ for all $i, j$ usually subsumed for the questionnaire interpretation of the model.

We shall assume that $\sum_{i} n_{i j} \geq 1, \sum_{j} n_{i j} \geq 1$ for all $j$ and $i \quad$ respectively, and further that the system is closed in the sense that for all $i \neq i^{\prime}$ there exists a sequence ( $i_{0}, \cdots, i_{l}$ ) of persons with $i=i_{0}, i^{\prime}=i_{l}$ and a sequence $\left(j_{0}, \cdots, j_{l-1}\right)$ of questions such that


$$
\begin{aligned}
& \text { Replacing }(\theta, \xi)=\left(\theta_{1}, \cdots, \theta_{k} ; \xi_{1}, \cdots, \xi_{m}\right) \text { by }\left(\theta^{\prime}, \xi^{\prime}\right)= \\
& \left(\theta_{1}+c, \cdots, \theta_{k}+c ; \xi_{1}-c, \cdots, \xi_{m}-c\right) \text { for an arbitrary } c \text { does not }
\end{aligned}
$$ change the probabilities (3.5), but since the system is closed, if we adopt the convention $\xi_{\mathrm{m}}=0$, as we do from now on, the parametrization is unique.

The problem is then to discuss when the ML-estimator for $\left(\theta_{1}, \cdots, \theta_{k}, \xi_{1}, \cdots, \xi_{m-1}\right)$ exists and is unique. The likelihood is brought on the form (2.3) by allowing a factor for each of the $n_{i j}$ responses of i to $\mathrm{j}, \mathrm{i}=1, \cdots, \mathrm{k}, \mathrm{j}=1, \cdots, \mathrm{~m}$, and this representation yields structure vectors $v \in \mathbb{R}^{k+m-1}$ of the form

where we show only the non-zero coordinates and use the convention that if $j=m$, the component -1 or 1 in the position marked $j$ is ignored.

Suppose we are given an extreme observation. By (b ii) of the Theorem, we can find $\gamma=\left(\alpha_{1}, \cdots, \alpha_{k}, \beta_{1}, \cdots, \beta_{m-1}\right) \neq 0$ such that $\langle\gamma, v\rangle \geq 0$ for all $v$ of the form (3.6) specified by the observation. Equivalently, putting $\beta_{m}=0$ (and reminding the reader that $n_{i j} \geq 2$ is allowed, so that several i.i.d. responses to the same question by the same person may occur), for all i,j

$$
\begin{array}{ll}
\alpha_{i}+\beta_{j} \leq 0 & \text { if } i \text { always answers } j \text { correctly } \\
\alpha_{i}+\beta_{j} \geq 0 & \text { if } i \text { always answers } j \text { wrongly }  \tag{3.7}\\
\alpha_{i}+\beta_{j}=0 & \text { if } i \text { answers } j \text { both correctly and wrongly. }
\end{array}
$$

Reordering persons and questions we can obtain

$$
\alpha_{i_{1}} \leq \cdots \leq \alpha_{i_{k}}, \quad \beta_{\mathbf{j}_{1}} \leq \cdots \leq \beta_{\mathbf{j}_{\mathrm{m}}}
$$

with a sharp inequality somewhere, and hence after a little reflection also the following table showing the sign of $\alpha_{i}+\beta_{j}$ :

the horizontal line separating $i_{1} \leq \cdots \leq i_{l-1}$ from $i_{l} \leq \cdots \leq i_{k}$ with $i_{l-1}<i_{l}$, and similarly the vertical line separating
$j_{1} \leq \cdots \leq j_{p-1}$ from $j_{p} \leq \cdots \leq j_{m}$ with $j_{p-1}<j_{p}$. Let $A=\left\{i_{1}, \cdots, i_{l-1}\right\}, B=\left\{j_{1}, \cdots, j_{p-1}\right\}$. (Since by assumption we only know that $\left(\alpha_{1}, \cdots, \alpha_{k}, \beta_{1}, \cdots, \beta_{m-1}\right) \neq 0$, it is possible that e.g. all $\alpha_{i}$ are equal so that $A$ or $A^{c}$ is empty. These degenerate cases must therefore be allowed as special cases of the sign configuration above).

Referring to (3.7) it is now clear that an extreme observation has the following structure:


Fig. 4 The extreme observations for the Rasch model.

For instance all responses by $i \in A$ to $j \in B$ must be correct, while for $j \in B^{c}$ an arbitrary answer is allowed.

If conversely the observation fits into the pattern from Fig. 4, then if e.g. $m \in B^{c}$ we have $\langle\gamma, v\rangle \geq 0$ for all structure vectors $v$ given by the observation provided $\gamma=\left(\alpha_{1}, \cdots, \alpha_{k} ; \beta_{1}, \cdots, \beta_{m-1}\right) \neq 0$ satisfies

$$
\alpha_{i}=\left\{\begin{array}{ll}
0 & i \in A \\
\alpha & i \in A^{c}
\end{array} \quad \beta_{j}=\left\{\begin{array}{rl}
-\alpha & j \in B \\
0 & j \in B^{c}
\end{array}\right.\right.
$$

where $\alpha>0$.

Thus an observation is extreme iff there is a subset $A \subset\{1, \cdots, k\}$ and a subset $B \subset\{1, \cdots, m\}$ such that (i) all questions $j \in B$ are alway sanswered correctly by any $i \in A$, and (ii) all questions $j \in B^{C}$ are always answered wrongly by any $i \in A^{c}$.

Special cases of extreme observations: one person always answers questions correctly (or wrongly), or one question is always answered correctly (or wrongly) by all persons.

The criterion above for an observation to be extreme is known, see Fischer [8,9].
3.6 A multiplicative Poisson model. The following model is proposed for a tournament where $k$ teams are paired of $f$ in a given number of games, and where the result of any game between teams $i$ and $j$ consists of
two scores by $i$ against $j$, and by $j$ against $i: i$ plays $j$ $n_{i j} \geq 0$ times (so $n_{i j}=n_{j i}$ ), the results of all games are independent and the two scores in any given game are also independent, and if i plays $j$, the score by $i$ against $j$ ( $j$ against i) follows a Poisson distribution with expectation

$$
\mathrm{e}^{\theta_{\mathrm{i}}+\xi_{\mathrm{j}}}\left[\mathrm{e}^{\left.\theta_{\mathrm{j}}+\xi_{\mathrm{i}}\right]}\right.
$$

(A minimal sufficient statistic is given by the vector comprising for each team $i$, the total score obtained by $i$ together with the total score against i. If a high score by a team means that the team plays well one would rank $i$ as better than $j$ if $\theta_{i}+\xi_{j}>\theta_{j}+\xi_{i}$, i.e. that team is best for which $\theta_{i}-\xi_{i}$ is largest. The $\theta_{i}$ may be thought of as attack and the $\xi_{j}$ as defense parameters).

We shall assume that for all $i, \sum_{j \neq i} n_{i j} \geq 1$ and also that the tournament is closed in the sense that for any $i \neq j$ there exists $i=i_{0} \neq i_{1} \neq \cdots \neq i_{m-1} \neq i_{m}=j$ such that all $n_{i_{l-1} i_{l}} \geq 1$ (cf. Example 3.4).

Replacing $\theta=\left(\theta_{1}, \cdots, \theta_{k} ; \xi_{1}, \cdots, \xi_{k}\right)$ by $\theta^{\prime}=\left(\theta_{1}+c, \cdots, \theta_{k}+c\right.$; $\xi_{1}-\mathrm{c}, \cdots, \xi_{\mathrm{k}}-\mathrm{c}$ ) does not change the Poisson parameters above, but by requiring, as we do from now on, that $\xi_{k}=0$, we obtain a unique parametrization.

Any game played between i and j contributes two factors to the
likelihood when writing it in the form (2.3), namely, if the score is $x$ by $i$ against $j$ and $y$ by $j$ against $i$,

$$
\begin{aligned}
& \frac{1}{x!} e^{\left(\theta_{i}+\xi_{j}\right) x} \exp \left[-e^{\left.\theta_{i}+\xi_{j}\right]} \frac{1}{y!} e^{\left(\theta_{j}+\xi_{i}\right) y} \exp \left[-e^{\left.\theta_{j}+\xi_{i}\right]}\right.\right. \\
& =\left[1+\sum_{\mathrm{z}=0}^{\infty} \frac{\mathrm{x}!}{\mathrm{z}!} \exp \left(\left(\theta_{\mathrm{i}}+\xi_{\mathrm{j}}\right)(\mathrm{z}-\mathrm{x})\right)\right]^{-1} \\
& \text { z } \neq \mathrm{x} \\
& \times\left[1+\sum_{\substack{\mathrm{z}=0 \\
\mathrm{z} \neq \mathrm{y}}}^{\infty} \frac{\mathrm{y}!}{\mathrm{z}!} \exp \left(\left(\theta_{\mathrm{j}}+\xi_{\mathrm{i}}\right)(\mathrm{z}-\mathrm{y})\right)\right]^{-1} .
\end{aligned}
$$

Thus any one score in any game generates infinitely many structure vectors, that however are simply related, viz. the vectors for the score $x$ by $i$ against $j$ are $(z-x) \tilde{v} \in \mathbb{R}^{2 k-1}$ for $z=0,1, \cdots, z \neq x$ with

$$
\tilde{v}=(\underbrace{\cdots, 1, \cdots}_{\theta} ; \underbrace{\cdots, 1, \cdots}_{\xi})
$$

where if $j=k$ the $j$ component 1 is ignored, and only non-zero entries are shown.

Consider now an extreme observation so that we can find $\boldsymbol{\gamma}=$ $\left(\alpha_{1}, \cdots, \alpha_{k} ; \beta_{1}, \cdots, \beta_{k-1}\right) \in \mathbb{R}^{2 k-1}$ with $\langle\gamma, \mathrm{v}\rangle \geq 0$ for all structure vectors $v$ determined by the observation. Introducing $\beta_{k}=0$, it is clear that this amounts to, for any $i \neq j$ with $n_{i j} \geq 1$,
(3.8) $\alpha_{i}+\beta_{j} \begin{cases}\geq 0 & \text { if } i \text { always has score } 0 \text { against } j \\ =0 & \text { if } i \text { at least once has score } \geq 1 \text { against } j \text {. }\end{cases}$

Rearranging the teams in two different ways so that $\alpha_{i_{1}} \leq \cdots \leq \alpha_{i_{k}}$, $\beta_{j_{1}} \leq \cdots \leq \beta_{j_{k}}$, just as in the previous example we obtain the following table for the sign of $\alpha_{i}+\beta_{j}$

corresponding to a division of the teams into two groups in two different ways, $A, A^{c}$ and $B, B^{c}$.

Since $r \neq 0$, and by (3.8), $\alpha_{i}+\beta_{j} \geq 0$ whenever $n_{i j} \geq 1$, it follows because the tournament is closed that $\alpha_{i}+\beta_{j}>0$ for at least one pair (i,j) with $n_{i j} \geq 1$, in particular $A^{c} \neq \varnothing, B^{c} \neq \varnothing$. The only way to account for the minus signs in the table is that for $i \in A$, $j \in B, n_{i j}=0$ (in particular $A$ or $B$ may be empty). Thus the following observation pattern emerges for scores by any $i$ against any $j$.


Fig. 5 The extreme observations for the multiplicative Poisson model.

Conversely, if the observation agrees with Fig. 5, and e.g. k $\in \mathrm{B}^{\mathrm{c}}$, if $\quad \gamma=\left(\alpha_{1}, \cdots, \alpha_{k} ; \beta_{1}, \cdots, \beta_{k-1}\right) \neq 0$ is given by

$$
\alpha_{i}=\left\{\begin{array}{ll}
0 & i \in A, \\
\alpha & i \in A^{c}
\end{array}, \quad \beta_{j}=\left\{\begin{array}{cl}
-\alpha & j \in B \\
0 & j \in B^{c}
\end{array}\right.\right.
$$

with $\alpha\rangle 0$, one checks that $\langle\gamma, \mathrm{v}\rangle \geq 0$ for all structure vectors $v$.

Thus, for a closed tournament an observation is extreme iff there exists $C \subset\{1, \cdots, k\}, D \subset\{1, \cdots, k\}$, both non-empty and such that $\underline{n}_{\mathrm{i} j} \underline{\geq 1}$ for at least one $\mathrm{i} \in \mathrm{C}, \mathrm{j} \in \mathrm{D}$ and furthermore it holds that (i) always $i \in C$ has score 0 against any $j \in D$, (ii) i $\in C^{c}$ never plays any $j \in D^{\mathbf{c}}$.

Special cases of extreme observations: one team always has score 0 against any team it plays, or one team always has 0 scored against it. $\square$
3.6 A Bradley-Terry model allowing draws. As in Example 3.4 we consider a closed tournament with $k$ players, but do not restrict the outcome of a game to a win for one of the two players, but allow also for a draw. Thus a game between $i$ and $j$ results in a win for $i$, $a$ draw, $a$ win for $j$ with probabilities

$$
\begin{equation*}
\frac{e^{\theta_{i}}}{e^{\theta_{i}} \xi_{i} \xi_{i} \xi_{j_{+e}} \theta_{j}} \tag{3.9}
\end{equation*}
$$

$$
\frac{e^{\xi_{i}+\xi_{j}}}{e^{\theta_{i}}+e^{\xi_{i}+\xi_{j}}+e_{j}^{\theta}}, \quad \frac{e^{\theta_{j}}}{e^{\theta_{i}} \xi^{\xi_{i}+\xi_{j}}{ }^{\theta} e_{j}}
$$

respectively. It is of course assumed that the results of different games are independent.

Similar models have been proposed by Davidson [6], Davidson and Beaver [7]. A minimal sufficient statistic is obtained by counting for each player the total number of wins and the total number of draws.

We assume now that $\xi_{\mathrm{k}}=0$. Since the tournament is closed, the parametrization is then unique.

We write the likelihood in the form (2.3) with each game played contributing one factor with two structural vectors $\in \mathbb{R}^{2 \mathrm{k}-1}$. For any game played between $i$ and $j$, (3.9) shows these two vectors to be

if i wins and

$$
\begin{align*}
& (\cdots, 1, \cdots, 0, \cdots, \cdots,-1, \cdots,-1, \cdots)  \tag{3.11}\\
& (\cdots, 0, \cdots, 1, \cdots ; \cdots,-1, \cdots,-1, \cdots)
\end{align*}
$$

if $i$ and $j$ draw. As usual, if e.g. $j=k$, the components on the rightmost coordinate marked $j$ are to be deleted, and all entries not shown are zero.

Consider an extreme observation and find $\gamma=\left(\alpha_{1}, \cdots, \alpha_{k} ; \beta_{1}, \cdots, \beta_{k-1}\right) \neq 0$ such that $\langle\gamma, \mathrm{v}\rangle \geq 0$ for all structure vectors v determined by the observation. Writing $\beta_{k}=0$ and referring
to (3.10) and (3.11) it is seen that this amounts to the following conditions for any pair (i,j) of players that have played each other at least once, and where the five categories refer to the games between $i$ and $j$ only:
(i) if one player, i say, always wins,

$$
\alpha_{i} \leq \beta_{i}+\beta_{j}, \quad \alpha_{i} \leq \alpha_{j}
$$

(ii) if all games are drawn,

$$
\alpha_{i} \geq \beta_{i}+\beta_{j}, \quad \alpha_{j} \geq \beta_{i}+\beta_{j}
$$

(iii) if one player, i say, wins at least once, draws at least once and never loses,

$$
\alpha_{i}=\beta_{i}+\beta_{j}, \quad \alpha_{i} \leq \alpha_{j}
$$

(iv) if both $i$ and $j$ win at least once and never draw,

$$
\alpha_{i} \leq \beta_{i}+\beta_{j}, \quad \alpha_{i}=\alpha_{j}
$$

(v) if both $i$ and $j$ win at least once and at least one game is drawn

$$
\alpha_{i}=\beta_{i}+\beta_{\mathrm{j}}, \quad \alpha_{\mathrm{i}}=\alpha_{\mathrm{j}}
$$

Working towards a characterization of the extreme observations, it proves useful to introduce $\delta_{i}=\beta_{i}-\alpha_{i}$ and study the signs for $\delta_{i}+\beta_{j}$, reordering the players in two different ways according to increasing values of $\delta_{i}$ and $\beta_{j}$ respectively. It quickly turns out that the rather crude sign analysis that worked successfully in the two previous examples, in the present case only yields a class of observations, that al though it certainly contains all the extreme ones, also comprises a host of non-extreme observations.

By gradually refining the sign analysis, we have been able to arrive at the class of extreme observations exhibited in Figure 6, but do not believe that it comprises all extremes. Fig. 6 serves to show that extreme observations in exponential families may have a very complicated structure.

In Fig. 6 the set of players is partitioned into three subsets in two different ways, together yielding a partitioning into nine subsets.

Any observation fitting with the pattern in Fig. 6 is extreme. The figure shows which outcomes are allowed in the games played between any two players $i, j$, categorized according to which of the nine subsets they belong to. In each cell of the resulting $9 \times 9$ table, seven classes of outcomes are allowed, corresponding to (i) - (v) above plus the two extras obtained by interchanging $i$ and $j$ in (i) and (iii). The following notation is used for the seven classes:
(i) i always wins: >, j always wins: く
(ii) all games drawn: ~
(iii) i never loses: $\underset{\sim}{ }$, j never loses: ふ
(iv) games never drawn: not ~
(v) any outcome allowed: any

The tabel is of course symmetric in $i$ and $j$, so all information is contained on and e.g. below the diagonal.


Fig. 6 A class of extreme observations for the Bradley-Terry model with draws.

We claim that if the observation fits with Fig. 6, then $\langle\gamma, \mathrm{v}\rangle \geq 0$ for all the structure vectors, provided $\gamma=\left(\alpha_{1}, \cdots, \alpha_{k}, \beta_{1}, \cdots, \beta_{k-1}\right) \neq 0$ is chosen in the following fashion: the $\beta_{j}$ (including $\beta_{k}=0$ ) are
constant on each of the sets $R, S, T$, the $\alpha_{i}$ are constant on each of the nine subsets of players and determined from

$$
\begin{equation*}
\beta_{\mathrm{R}}>\beta_{\mathrm{T}}, \quad \beta_{\mathrm{S}}=\frac{1}{2}\left(\beta_{\mathrm{R}}+\beta_{\mathrm{T}}\right) \tag{3.12}
\end{equation*}
$$

as follows:

$$
\begin{array}{ll}
\alpha_{\mathrm{KR}}=\alpha_{\mathrm{LS}}=\alpha_{\mathrm{MT}} & =2 \beta_{\mathrm{S}}, \\
\alpha_{\mathrm{LR}}=\alpha_{\mathrm{MS}} & =\beta_{\mathrm{R}}+\beta_{\mathrm{S}}, \\
\alpha_{\mathrm{KS}}=\alpha_{\mathrm{LT}} & =\beta_{\mathrm{S}}+\beta_{\mathrm{T}},  \tag{3.13}\\
\alpha_{\mathrm{KT}} & =2 \beta_{\mathrm{T}}, \\
\alpha_{\mathrm{MR}} & \\
& =2 \beta_{\mathrm{R}} .
\end{array}
$$

For the proof one must verify 45 groups of inequalities!

As mentioned above Fig. 6 does not seem to contain all extreme observations. It may be shown however that by replacing the ">" for $i \in K \cap S, j \in L \cap S$ by "not $\sim$ " and the "く" for $i \in L \cap R, j \in L \cap S$ by "not $\sim$ ", and symmetrizing, one obtains a table that comprise all extreme observations. Unfortunately, not all observations in this slightly more general table are extreme.

Simpler examples of extreme observations may be obtained from by Fig. 6 allowing one or more of the nine subsets to be empty, the only requirement being that $\gamma$ given by (3.12) and (3.13) is $\neq 0$. Special cases: one player wins all games (or draws all games, or loses all
games). Also, the corner configuration with only $K \cap R$ and $M \cap T$ non-empty yields extreme observations.

## Appendix

Proof of the Theorem
With f given by (2.3) and the assumption below (2.3) in force, $l=\log \mathrm{f}$ is twice differentiable with the gradient and the Hessian found easily using (2.4), (2.5).

For $\beta \in \mathbb{R}^{p}$, a simple application of Jensen's inequality shows that $\left\langle-\mathrm{D}^{2} l(\theta) \beta, \beta\right\rangle \geq 0$ with equality iff for all $\mathrm{i}, \mathrm{j} \in \mathrm{A}_{\mathrm{i}},\left\langle\mathrm{v}_{\mathrm{ij}}, \beta\right\rangle=0$. Thus $l$ is concave and if $\operatorname{span}\left\{\mathrm{v}_{\mathrm{ij}}\right\}=\mathbb{R}^{\mathrm{p}},-\mathrm{D}^{2} l(\theta)$ is positive definite for all $\theta$. If on the other hand $\operatorname{span}\left\{v_{i j}\right\} \neq \mathbb{R}^{p}$, choosing $\beta \perp \operatorname{span}\left\{\mathrm{v}_{\mathrm{ij}}\right\}$ above shows that $-\mathrm{D}^{2} \imath(\theta)$, although positive semidefinite, is nowhere positive definite. Thus (a i) - (a iii) are equivalent.

The equivalence between (b i) - (b ii) is standard. We show that $f$ has a unique maximum iff (b ii) holds.

Suppose $l(\hat{\theta})=\sup l$ and that $\hat{\theta}$ is unique. Then necessarily $\operatorname{span}\left\{\mathrm{v}_{\mathrm{ij}}\right\}=\mathbb{R}^{\mathrm{p}}$ since otherwise $l(\hat{\theta})=l(\hat{\theta}+\beta)$ for any $\beta \perp \operatorname{span}\left\{\mathrm{v}_{\mathrm{ij}}\right\}$. Thus, by part (a), $l$ is strictly concave and by standard convexity theory, for all directions $\theta_{0} \in \mathbb{R}^{p}$ where $\left\|\theta_{0}\right\|=1$,

$$
\begin{equation*}
\lim _{B \rightarrow \infty} l\left(B \theta_{O}\right)=-\infty \tag{A1}
\end{equation*}
$$

or, equivalently
(A2)

$$
\lim _{B \rightarrow \infty} f\left(B \theta_{0}\right)=0
$$

Obviously, for each i

$$
\lim _{B \rightarrow \infty} \sum_{j \in A_{i}} a_{i j} e^{\left\langle B \theta_{0}, v_{i j}\right\rangle}=\infty
$$

iff $\left.\left\langle\theta_{0}, v_{i j}\right\rangle\right\rangle 0$ for some $j \in A_{i}$, with the limit a finite constant otherwise. Thus (A2) holds iff for every direction $\left.\theta_{0},\left\langle\theta_{0}, v_{i j}\right\rangle\right\rangle 0$ for some $i, j$. Taking an arbitrary $\alpha \in \mathbb{R}^{p}$ with $\alpha \neq 0$ and using this on $\theta_{0}= \pm \alpha /\|\alpha\|$, it is seen that (bii) holds.

Conversely, if (b ii) and, afortiori, (b i) hold, (A2) and hence (A1) is true for all directions $\theta_{0}$. By (bi), $\operatorname{span}\left\{\mathrm{v}_{\mathrm{ij}}\right\}=\mathbb{R}^{p}$ so $l$ is strictly concave and therefore has at most one maximum. By (A1), if C is close enough to $-\infty$, the set $M=\{\theta: l(\theta) \geq C\}$ is non-empty and compact, and since $l$ is continuous, it has a maximum on $M$. $\quad$

## Acknowledgement

I would like to thank Søren Tolver Jensen for helpful discussions.

## REFERENCES

[1] Albert, A. and Anderson, J.A. On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. Biometrika $\underline{71}$ (1984), 1-10.
[2] Barndorff-Nielsen, O. Information and Exponential Families in Statistical Theory. Wiley, Chichester, 1978.
[3] Bryson, M.C. and Johnson, M.E. The Incidence of Monotone Likelihood in the Cox Model. Technometrics 23 (1981), 381-383.
[4] Cox, D.R. Regression Models and Life Tables (with discussion). J.Roy.Statist.Soc.Ser. B 34 (1972), 187-220.
[5] Cox, D.R. Partial Likelihood. Biometrika 62 (1975), 269-276.
[6] Davidson, R.R. On extending the Bradley-Terry Model to accomodate Ties in Paired Comparison Experiments. J.Amer.Statist.Assoc. 65 (1970), 317-328.
[7] Davidson, R.R. and Beaver, R.J. On Extending the Bradley-Terry Model to incorporate within-pair Order Effects. Biometrics 33 (1977), 693-702.

Fischer, G.H. Einfuihrung in die Theorie psychologischer Tests. Huber, Bern, 1974.
[9] Fischer, G.H. On the Existence and Uniqueness of Maximum-Likelihood Estimates in the Rasch Model. Psychometrika 46 (1981), 59-77.
[10] Larsen, J. Estimation in Exponential Families. University of Copenhagen, Institute of Mathematical Statistics, Preprint 3 (1971).
[11] Rasch, G. Probabilistic Models for some Intelligence and Attainment Tests. Pædagogisk Institut, Copenhagen (1960).
[12] Zermelo, E. Die Berechnung der Turnier-Ergebnisse als ein Maximum-Problem der Wahrscheinlichkeitsrechnung. Math.Z. 29 (1929), 436-460.

GOPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS，UNIVERSITETSPARKEN 5， 2100 COPENHAGEN $\emptyset, ~ D E N M A R K$, TELEPHONE＋ 451353133 ．

No． 1 Jensen，S申ren Tolver and Johansen，S申ren：Estimation of Proportional Covariances．

No． 2 Rootzén，Holger：Extremes，Loads，and Strengths．

No． 3 Bertelsen，Aksel：On the Problem of Testing Reality of a Complex Multivariate Normal Distribution．

No． 4 Gill，Richard D．and Johansen，S申ren：Product－Integrals and Counting Processes．

No． 5 Leadbetter，M．R．and Rootzén，Holger：Extremal Theory for Stochastic Processes．

No． 6 Tjur，Tue：Block Designs and Electrical Networks．
No． 7 Johansen，S申ren：Statistical Analysis of Cointegration Vectors．
No． 8 Bertelsen，Aksel：On the Problem of Testing Reality of a Complex Multivariate Normal Distribution，II．

No． 9 Andersson，S．A．and Perlman，M．D．：Group－Invariant Analogues of Hadamard＇s Inequality．

No． 10 Hald，Anders：Two Generalizations of the Problem of Points by Bernoulli，de Moivre and Montmort．

No． 11 Andersson，Steen Arne：The Lattice Structure of Orthogonal Linear Models and Orthogonal Variance Component Models．

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN $\emptyset$, DENMARK, TELEPHONE + 451353133 .

No. 1 Jacobsen, Martin: Discrete Exponential Families: Deciding when the Maximum Likelihood Estimator Exists and is Unique.

