Remarks on optimal designs in a paired comparison experiment using the Bradley-Terry model
van Berkum, E.E.M.

Published: 01/01/1982

Document Version
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Remarks on optimal designs in a paired comparison experiment using the Bradley-Terry model

by

Emiel E.M. van Berkum

Eindhoven, January 1982

The Netherlands
SUMMARY

A number of methods for estimation of the parameters in a paired comparison experiment are considered. This is done by comparing the (asymptotic) variance-covariance matrices of the estimators. The methods considered are: Springall (1973), Beaver (1977a, 1977b) and El-Helbawy/Bradley (1978). It is shown that these three methods yield the same variance-covariance matrix. When the treatment parameters are equal, the variance-covariance matrix is proportional to the variance-covariance matrix, which is obtained when an ordinary least-squares method is used. In constructing optimal designs it is possible to restrict oneself to this variance-covariance matrix. Some errors by Springall (1973) are corrected.

KEYWORDS: Paired comparison; Bradley-Terry model; Optimal designs, Weighted least-squares; Factorials.
1. INTRODUCTION

The paired comparison experiment has \( T \) treatments or items, \( T_1, \ldots, T_t \), with \( n_{ij} \) judgements or comparisons of \( T_i \) and \( T_j \), \( n_{ij} \geq 0 \), \( n_{ii} = 0 \), \( n_{ji} = n_{ij} \), \( i, j = 1, \ldots, t \).

We define \( n_{ij} \) as the number of times treatment \( T_i \) is preferred to \( T_j \), when \( i \) and \( j \) were compared, \( n_{ij} = n_{ji} \), \( n_{ij} + n_{ji} = n_{ij} \). Bradley and Terry (1952) provided a paired-comparison model. The model postulates the existence of treatment parameters, \( \pi_i \) for \( T_i \), \( \pi_i > 0 \), \( i = 1, \ldots, t \), such that the probability \( \pi_{ij} \) of selecting \( T_i \) when compared with \( T_j \) is

\[
\pi_{ij} = \frac{\pi_i}{(\pi_i + \pi_j)} , \quad i \neq j .
\]

Here \( \ln \pi \) is the parameter which has to be estimated. Since (1.1) is scale-independent, a convenient, scale-determining constraint,

\[
\sum_{i} \mu_i = 0
\]

where

\[
\mu_i = \ln \pi_i , \quad i = 1, \ldots, t ,
\]

is used.

In "classical" experiments in which no comparisons are made, but only direct observations, many results are achieved in constructing optimal designs. Several criteria, for example D-optimality, G-optimality, have been developed. Very few results are available in paired comparison experiments.

However, the developed criteria are also applicable in paired comparison experiments. Many of these criteria are a function of the variance-covariance matrix of the estimators.
The results achieved are:

Quenouille and John (1971) presented $2^n$ factorial paired comparison designs, which can be constructed so as to reduce the number of pairs required by neglecting or losing information on higher order interactions. Their analysis assumes that the variances of the observations are equal. Springall (1973) considers the model

$$\ln \pi_i = \sum_{k=1}^{s} x_{ik} \beta_k, \ i = 1, \ldots, t,$$

and gives so-called analogue designs. These are designs in which the elements of the paired comparison variance-covariance matrix are proportional to the elements of the "classical" variance-covariance matrix with the same design points.

El-Helbawy and Bradley (1978) give the variance-covariance matrix in factorial paired comparison experiments and present optimal designs for some examples. Beaver (1977a, 1977b) considers the analysis of paired comparison experiments using a weighted least squares approach, but does not discuss optimal designs.

2. COMPARISON OF THE VARIANCE-COVARIANCE MATRICES

i) Springall (1973) assumes that the $\mu_i$, $i = 1, \ldots, t$, are functions of continuous independent variables, $x_1, \ldots, x_s$ (1.4). The results are - in a slightly different notation - $(\hat{\beta}_1 - \beta_1), \ldots, (\hat{\beta}_s - \beta_s)$ have the asymptotic $s$-variate normal distribution with zero means and variance-covariance matrix

$$(\lambda_{rq})^{-1}, \ r, q = 1, \ldots, t,$$

where

$$\lambda_{rq} = \sum_{i<j} \sum\sum \pi_{ij} \phi_{ij} (x_{ir} - x_{jr})(x_{iq} - x_{jq}),$$

where

$$\phi_{ij} = \pi_{ij}/(\pi_i + \pi_j)^2 = \pi_{ij}^2 / \pi_i \pi_j.$$
and \( \hat{\beta}_i \) is the maximum likelihood estimator of \( \beta_i \).

We may formulate (2.1) in a different fashion:

Let \( X \) be the matrix of dimension \( t \times s \), the elements of which are the \( x_{ik} \) from (1.4). This matrix may be regarded as the design matrix in the classical experiment.

Define

\[
G = \begin{pmatrix}
1 & -1 \\
1 & -1 \\
\vdots & \vdots \\
1 & -1 \\
1 & \ldots & -1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & \ldots & -1 & \ldots & -1 \\
1 & \ldots & -1 \\
1 & -1 \\
\end{pmatrix}
\]

a matrix of dimension \( \binom{t}{2} \times t \), having one + 1, one - 1, and \( t-2 \) zeroes in each row, such that

\[
(G'G)_{ij} = \begin{cases} 
-1, & \text{if } i \neq j, \\
t-1, & \text{if } i = j.
\end{cases}
\]

The matrix \( G \) is related to a design in which all items are compared with each other just once \((n_{ij} = 1, i,j = 1,\ldots,t, i \neq j)\).
Define

\[ D = GX, \]
\[ \phi(\pi) = \text{diag}(n_{12}^2, n_{13}^3, \ldots, n_{t-1}^t), \]
\[ \pi = (x_{11}, \ldots, x_{nt})', \]

\[ \phi(\pi) = \begin{pmatrix} n_{12} & n_{13} & \cdots & n_{1t} \\ n_{23} & n_{2t} & \cdots & n_{t-1t} \\ n_{t-12} & n_{t-13} & \cdots & n_{tt} \end{pmatrix}, \]

a matrix of dimension \( (t^2) \times (t^2) \).

The following result may be readily obtained:

\[ \lambda_{rq} = -\sum_{i<j} x_{ir}^2 x_{ij} x_{jl} - \sum_{i>j} x_{ir} x_{ij} x_{i'l} + \sum_{i<j} x_{ir} x_{ij} x_{i'l} + \sum_{i>j} x_{ir} x_{ij} x_{i'l} = \]
\[ = \sum_{i<j} x_{ir} (G'\phi(\pi)G)_{ij} x_{ij} x_{i'l}. \]

Hence

\[ (\lambda_{rq})^{-1} = (D'\phi(\pi)D)^{-1}. \]

Remarks

- Springall does not derive the variance-covariance matrix \( (\lambda_{rq}) \), but the variance-covariance matrix \( (\nu_{rq})^{-1} \) of the estimators \( \hat{\xi}_i \) of \( \xi_i \), where

\[ \hat{\xi}_i = \exp(\beta_i), i = 1, \ldots, t. \]

The result is

\[ \nu_{rq} = \frac{1}{\xi_r \xi_q} \sum_{i<j} n_{ij} \phi_{ij} (x_{ir} - x_{jr}) (x_{iq} - x_{jq}). \]

Then Springall uses \( \lambda_{rq} = \nu_{rq} / (\xi_r \xi_q) \). This is not correct, it should be

\[ \lambda_{rq} = \xi_r \xi_q \nu_{rq}. \]

- Springall uses a generalization of the Bradley-Terry model, the model of Rao and Kupper (1967), including a threshold parameter \( \eta \). This parameter can be interpreted as the threshold of sensory perception for the judge.
If we let $\theta = \exp(n)$, the preference probabilities can be written as

$$\pi_{i,j} = \pi_i / (\pi_i + \theta \pi_j)$$

$$\pi_{j,i} = \pi_j / (\pi_j + \theta \pi_i)$$

$$\pi_{0,j} = \pi_i \pi_j (\theta^2 - 1) / ((\pi_i + \theta \pi_j)(\pi_j + \theta \pi_i))$$

There are $t + 1$ unknown parameters and the variance-covariance matrix has dimension $(t+1) \times (t+1)$. Springall gives the following results

$$\nu_{00} = 2n_0 \frac{\theta^2 + 1}{(\theta^2 - 1)^2} - \sum_{i<j} n_{ij} \phi_{ij}^{-1}$$

$$\nu_{0r} = \frac{1}{\xi_r} \sum_{i<j} n_{ij} \phi_{ij}(x_{ir} - x_{ir}) \quad r = 1, \ldots, s$$

$$\nu_{r} = \frac{1}{\xi_q} \sum_{i<j} n_{ij} \phi_{ij}(x_{ir} - x_{ir})(x_{iq} - x_{iq}) \quad r, q = 1, \ldots, s$$

where

$$\phi_{ij} = \frac{\theta^2 \pi_i \pi_j (\theta \pi_i^2 + \pi_i^2) + 2\pi_i \pi_j}{(\pi_i + \theta \pi_j)^2 (\pi_j + \theta \pi_i)^2},$$

$n_0$ is the total number of times no preference was declared in the experiment.

However, these statements cannot be correct, since $n_0$ is a random variable.

They should be

$$\nu_{00} = \sum_{i<j} n_{ij} \pi_i \pi_j (\theta^2 + (\theta \pi_i^2 + \theta \pi_j^2) + 2\pi_i \pi_j + 4\pi_i \pi_j (\theta^2 - 1))$$

$$\nu_{0r} = \frac{1}{\xi_r} \sum_{i<j} n_{ij} \pi_i \pi_j \frac{\theta^2 (\theta \pi_i^2 - \pi_i^2)(x_{ir} - x_{ir})}{(\pi_i + \theta \pi_j)^2 (\pi_j + \theta \pi_i)^2}$$

$\nu_{rq}$ as above.
ii) El-Helbawy and Bradley (1978) analyse factorials. Let \( n \) be the number of factors; the \( i \)-th factor has \( b_i \) levels, so that

\[
  t = \prod_{i=1}^{n} b_i.
\]

The general problem is to estimate the parameters \( \mu_i, i = 1,\ldots,t \), under the conditions

\[
(2.4) \quad \begin{bmatrix} 1' \\ \mathbf{B}_m \end{bmatrix} \mu = 0,
\]

where

\[
\begin{align*}
\mu &= (\mu_1, \ldots, \mu_t)', \\
1'_t &= (1, \ldots, 1)', \\
1'_t \mu &= 0 \text{ is the usual condition,}
\end{align*}
\]

\( \mathbf{B}_m \mu = 0 \) is the assumption that \( m \) specified orthonormal treatment contrasts are zero.

The problem is solved by estimating the other \( t-m-1 \) orthonormal contrasts; these can written as linear combinations of the \( \mu_i \):

\[
(2.5) \quad \theta_1 = \mathbf{B}_m^* \mu,
\]

where \( \mathbf{B}_m^* \) is a matrix of dimension \((t-m-1) \times t\), and

\[
\begin{bmatrix} \mathbf{1}'_t / \sqrt{t} \\ [\mathbf{1}'_t / \sqrt{t} \mathbf{B}_m \mathbf{B}_m^*] \end{bmatrix} = \mathbf{I}.
\]

It follows that

\[
(2.6) \quad \mu = \mathbf{B}_m^* \theta.
\]
The results are

\[(2.7)(\hat{\theta}_1 - \theta_1)\] has the asymptotic \((t-m-1)\) variate normal distribution with zero means and variance-covariance matrix \((B_m^* \Lambda B_m^*)^{-1}\), where

\[
A(\pi)_{ij} = -n_{ij}^2 \text{ if } i \neq j
\]

\[
= \sum_{k \neq i} n_{ik} \phi_{ik} \text{ if } i = j.
\]

Now it is easy to show that the following holds

\[
B_m^* = \frac{1}{\sqrt{t}} X',
\]

where \(X\) can be regarded as the design matrix in the classical experiment.

Hence \((2.5)\) is equivalent to

\[\mu = X\beta,\]

and the estimator of \(\beta\) is \(\hat{\beta} = \hat{\theta}_1 / \sqrt{t} \).

Now

\[\text{var}(\frac{1}{\sqrt{t}} \hat{\theta}_1) = (tB_m^* \Lambda(\pi)B_m^*)^{-1} = (X' \Lambda(\pi)X)^{-1} = (X'G'\phi(\pi)GX)^{-1}\]

hence \((2.7)\) may be rewritten as:

\[(2.9) \quad \text{var} \hat{\beta} = (D' \phi(\pi)D)^{-1}.\]


Define

\[\pi = \left(\pi_{12}, \pi_{21}, \pi_{13}, \pi_{31}, \ldots, \pi_{1t}, \pi_{t1}, \pi_{23}, \pi_{32}, \ldots, \pi_{t-1t}, \pi_{tt-1}\right)',\]

\[p_{ij} = \frac{n_{i,j}}{n_{i,j}}, \text{ an estimate of } \pi_{ij},\]

\[p = \left(p_{12}, p_{21}, p_{13}, p_{31}, \ldots, p_{1t}, p_{t1}, p_{23}, p_{32}, \ldots, p_{t-1t}, p_{tt-1}\right)',\]

\[f_{ij}(\pi) = \ln(\pi_{ij}/\pi_{ij}) = \ln \pi_i - \ln \pi_j,\]
$F._{\pi} = (f_{12}, f_{13}, \ldots, f_{1t}, f_{23}, \ldots, f_{t-1t})'$,

$H = \left\{ \frac{\partial F_{i}}{\partial \pi_{j}} | \pi_{i} = p_{i} \right\}$, a matrix of dimension $(t \times 2(t))$ ,

$V$ is a block diagonal matrix of dimension $2(t \times 2(t))$, having as blocks $V(p_{ij})$ ,

$V(p_{ij}) = \frac{1}{n_{ij}} \begin{pmatrix} p_{ij}p_{ji} & -p_{ij}p_{ji} \\ -p_{ij}p_{ji} & p_{ij}p_{ji} \end{pmatrix}$,

$S = HVH'$.

Under certain rather weak conditions, the following holds. If $F_.(\pi) = Z\beta$, where $Z$ is a known matrix of dimension $(t \times s)$ and $\beta$ an unknown vector of parameters, then the best asymptotic normal estimator of $\beta$ is given by

(2.10) $\hat{\beta} = (Z'S^{-1}Z)^{-1}Z'S^{-1}F_.(p.)$,

for which

(2.11) $\text{var } \hat{\beta} = (Z'S^{-1}Z)^{-1}$.

Beaver uses the model $\pi_{i} = \sum_{k=1}^{s} x_{ij} \beta_{k}$, so that $F_.(\pi)$ may be written as a linear function of $s$ independent parameters $\beta_{i}$.

The results can be rewritten as follows; it is easy to verify that $S^{-1} = \phi(p)$; $\phi(p)$ is the matrix $\phi(\pi)$, in which the $\pi_{ij}$ have been replaced by the estimates $p_{ij}$. It can be shown that

$F_.(\pi) = G \ln \pi = GX\beta = D\beta$ ;

hence

$\text{var } \hat{\beta} = (D'\phi(p)D)^{-1}$.
When the parameters are known, we have

\[ \text{var } \beta = (D'\phi(\pi)D)^{-1}. \]  

3. APPLICATION IN OPTIMAL DESIGNING; ORDINARY LEAST-SQUARES

As is shown by (2.3), (2.9) and (2.12), the three methods considered yield the same variance-covariance matrix. Many criteria in optimal designing are functions of this matrix. However, the variance-covariance matrix \( \phi(\pi) \) is a function of the unknown parameters \( \pi_i \), and no estimates of the parameters are available, when the design is chosen.

The variance-covariance matrix can be simplified, if the assumption of no treatment differences is made:

\[ \forall i : \pi_i = 1. \]

Actually, this is the choice Springall and El-Helbawy/Bradley make, when they give applications of their results. Under this condition the variance-covariance matrix may be written as follows

\[ (D'\phi(1_\pi)D)^{-1}, \]

where

\[ \phi(1_\pi) = \frac{1}{2} \text{ diag}(n_{12}, n_{13}, \ldots, n_{1T}, n_{23}, \ldots, n_{T-1T}). \]

As will be shown, the matrix \((D'\phi(1_\pi)D)^{-1}\) is proportional to the variance-covariance matrix, obtained by the method of ordinary least squares.

Let

\[ y = D^* \beta + e, \]
where

\[ y \in \mathbb{R}^N, \]  
the vector of observed differences or preferences,

\[ N = \sum_{i<j} n_{ij}, \]  
\[ D^*, \]  
the design matrix of dimension \( N \times s \),

\[ \beta = (\beta_1, \ldots, \beta_s)', \]  
e is the disturbance vector with \( Ee = 0, \) \( \text{var } e = \sigma^2 I. \)

In general the assumption \( \text{var } e = \sigma^2 I \) does not hold when paired comparisons are applied.

The matrix \( D^* \) may be written as follows

\[ (3.4) \quad D^* = G^* X, \]  
where \( X \) is the design matrix in a classical experiment, \( G^* \) is a matrix analogous to \( G. \) It has in each row one \(+1\), one \(-1\) and \( t - 2 \) zeroes; a row is repeated \( n_{ij} \) times, when the items \( T_i \) and \( T_j \), which it is related to, are compared \( n_{ij} \) times.

The least squares estimator is

\[ \hat{\beta} = (D^* 'D^*)^{-1}D^* 'y, \]  
and

\[ (3.5) \quad \text{var } \hat{\beta} = (D^* 'D^*)^{-1}\sigma^2 . \]

This may be rewritten as:

\[ D^* 'D^* = X'G^* G^* X = 4(X'G'(1_\tau)GX) = 4D'\Phi(1_\tau)D ; \]

hence

\[ (3.6) \quad \text{var } \hat{\beta} = \frac{\sigma^2}{4} (D'\Phi(1_\tau)D)^{-1} . \]
This matrix in (3.6) is proportional to the matrix in (3.2). When choosing optimal designs, we may restrict ourselves to this ordinary least squares variance-covariance matrix. Actually, the designs given by Springall and El-Helbawy/Bradley for $2^n$ factorials, may be found by using the method developed by Quenouille and John.

These results can be extended to other models, for example the Rao Kupper model, with threshold parameter $\eta$. Further research will be centred on developing optimal designs by investigating the variance-covariance matrix

$$(D'\Phi(I_{\ell}D)^{-1})$$.
References


