On a connection between the Bradley–Terry model and the Cox proportional hazards model

Yuhua Su*, Mai Zhou

Department of Statistics, University of Kentucky, Lexington, KY 40506-0027, USA

Received 25 August 2004; received in revised form 29 August 2005
Available online 4 November 2005

Abstract

This article addresses a connection between the Bradley and Terry [1952a. The rank analysis of incomplete block designs. I. The method of paired comparisons. Biometrika 39, 324–345] model and Cox [1972. Regression models and life-tables. J. Roy. Statist. Soc. B 24, 187–220 (with discussion)] proportional hazards model. We show that the partial likelihood of random variables that satisfy the stratified proportional hazards assumption of Cox [1972. Regression models and life-tables. J. Roy. Statist. Soc. B 24, 187–220 (with discussion)] coincides with the likelihood given by the Bradley–Terry (BT) model for rank order events. Such a connection not only allows many available software for the Cox model to be used for regression analysis based on the BT model, but also enables the new developments to fit certain rank order data by using ideas that stem from the Cox model and its extensions available in many recent literatures from survival analysis.

© 2005 Elsevier B.V. All rights reserved.

MSC: primary 00A69; secondary 60G25

Keywords: Rank order data; Partial likelihood

1. Introduction—rank order data

In games such as horse racing and chess, ordered name lists are often of interest and recorded after the comparison of the performance of contestants. Suppose there are \( N \) independent (unobservable) random variables \( T_1, \ldots, T_N \) representing the performance of \( N \) contestants.

If we compare two random variables \( T_i \) and \( T_j \), where \( 1 \leq i, j \leq N \), then the event \( \{T_i > T_j\} \) is called a rank-order event for a paired comparison experiment. In general, if we compare \( K \) \( (2 \leq K \leq N) \) random variables, \( T_{i_1}, \ldots, T_{i_K} \), where \( i_j \) are distinct numbers from \( \{1, \ldots, N\} \), then the event \( \{T_{i_1} \geq \cdots \geq T_{i_K}\} \), is called a rank-order event for a multiple comparison experiment. The rank order data is a collection of outcomes of rank order events.

*Corresponding author.

E-mail address: yuhua.su@yale.edu (Y. Su).
Example 1 ($N = 7$, $K = 2$). Consider the example of paired comparisons in Agresti (2002, p. 438). Table 1 gives the outcomes of the games within the Eastern Division of the American league in the 1987 baseball season.

In this example, New York and Baltimore had 13 matches (comparisons) and New York wins 10 of them. From those observations, we may want to estimate the probabilities of rank order events, for example, the event that New York would defeat Baltimore.

2. Model definitions

Bradley–Terry model is a popular model used to estimate the probabilities of rank order events based on data as described in Section 1.

2.1. The Bradley–Terry (BT) models

Bradley and Terry (1952a) do not assume specific distributions for $T_i$ in the model description but postulate that every subject $i$ can have a parameter $\pi_i > 0$, and assume that the event $\{T_i > T_j\}$ has a probability given by $\pi_j/\left(\pi_i + \pi_j\right)$.

Definition 1 (The BT model ($K = 2$)). If $\pi_i (> 0)$ is a parameter associated with $T_i$, $i = 1, \ldots, N$, then the probability of the event $\{T_i > T_j\}$ is given by

$$P_B(T_i > T_j) = \frac{\pi_j}{\pi_i + \pi_j}, \quad i \neq j, \quad i, j = 1, \ldots, N.$$  (1)

Bradley and Terry (1952b) propose a model for triple comparisons where the events $\{T_i > T_j > T_k\}$ are assumed to have probabilities given by $\pi_k/\left(\pi_i + \pi_j + \pi_k\right) \times \pi_j/\left(\pi_i + \pi_j\right)$. From here we can easily generalize the BT model to multiple comparisons.

Definition 2 (The BT model ($K > 2$)). The BT model of $K$-subject matches assigns the probabilities by the following rule,

$$P_B(T_{i_1} > \cdots > T_{i_k}) = \prod_{j=2}^{K} \frac{\pi_{j}}{\sum_{j=1}^{K} \pi_{j}},$$  (2)

where $i_j$’s are distinct numbers in $\{1, \ldots, N\}$; $\pi_i (> 0)$’s are parameters associated with $T_i$’s, respectively.

Estimation of the parameters $\pi_i$ can then be obtained by the maximum likelihood method.
2.2. The Cox proportional hazards model

In the Cox proportional hazards model, one assumption is the existence of the hazard functions \( \lambda_i(t) \) for random variables \( T_i \), \( 1 \leq i \leq N \). This model, introduced by Cox (1972) presumes that a parameter \( \pi_i(>0) \) of \( T_i \) affects the hazard function in a multiplicative manner according to

\[
\lambda_i(t) = \lambda_0(t) \pi_i, \quad i = 1, \ldots, N
\]

(3)

where \( \lambda_0(t) \) is an unspecified baseline hazard function. In the survival analysis, usually \( \pi_i = \exp(Z_i^T \beta) \), where \( Z_i \) is a \( 1 \times q \) vector of covariates for subject \( i \) and \( \beta \) is a \( 1 \times q \) vector of unknown parameters.

There are many extensions to the proportional hazards model (see Therneau and Grambsch, 2000). The one that is relevant to our study here is the stratified proportional hazards model. Basically we shall consider each match/comparison among \( K \) contestants to be a distinct stratum.

**Definition 3** \((2 \leq K \leq N)\). In a stratified Cox proportional hazards model, a random variable, \( T_i \), in a stratum/comparison \( s \) has the following hazard function:

\[
\lambda_i(t) = \lambda_s(t) \pi_i, \quad j = 1, \ldots, K,
\]

(4)

where \( \pi_i \)'s are distinct numbers in \([1, \ldots, N]\), and \( \lambda_s(t) \) is the common (but un-specified) baseline hazard shared by random variables \( T_{i1}, \ldots, T_{ik} \), in this particular comparison.

This model allows the baseline hazard \( \lambda_i(t) \) to change from strata to strata. In the analysis of the stratified proportional hazards model, the statistical inference is typically based on the so called partial likelihood. For details please see Therneau and Grambsch (2000, Chapter 3).

**Example 1 (Continued).** Note that every team plays every other team exactly 13 times. There are \( \binom{7}{2} = 21 \) different combinations of pairings between two teams. The total number of pairwise comparisons is \( 21 \times 13 = 273 \). Therefore, the stratified Cox model have 21 different baselines here: \( s = 1, 2, \ldots, 21 \).

Let \((i,j) (1 \leq i < j \leq 7)\) span over all the combinations of possible pairings. Denote the performance of team \( q \) in the comparison \((i,j)\) by random variable \( T_{q(i,j)} \), \( q \in \{i,j\}, 1 \leq i < j \leq 7 \). The stratified Cox proportional hazards model for this paired comparison assumes that the hazards of those random variables satisfy

\[
\hat{\lambda}_{q(i,j)}(t) = \hat{\lambda}_{(i,j)}(t) \pi_q \quad \text{where} \quad q = i \text{ or } j, \quad 1 \leq i < j \leq 7.
\]

(5)

\( \hat{\lambda}_{(1,2)}(t), \ldots, \hat{\lambda}_{(6,7)}(t) \) are unspecified baseline hazards. They may be different from each other.

3. Main results

Now we are ready to describe the main results of this paper.

**Theorem 1.** Suppose that independent random variables \( T_i \) satisfy the stratified Cox proportional hazards assumption (4) and \( T_i \) and \( T_j \) belong to one stratum, then we have (i) \( P(T_i > T_j) = \pi_i / (\pi_i + \pi_j) \); (ii) The contribution to the partial likelihood from the random variables \( T_i \) and \( T_j \) with \( \{T_i > T_j\} \) is also \( \pi_j / (\pi_i + \pi_j) \).

The proof of the theorem is straightforward and will be omitted.

Recall the BT model Definition 1, we immediately conclude that the likelihood of BT model and the partial likelihood of the stratified Cox proportional hazards model are identical. This is an important finding with far reaching consequences: (i) as illustrated in the next section, software developed for stratified Cox proportional hazards model can be used to obtain MLE in BT model since the two estimators are identical; (ii) MLE in BT model can share large sample properties obtained by powerful counting process martingale theory for partial likelihood estimators in Cox model; (iii) likelihood ratio tests are going to be identical in two models, as well as Fisher information matrices; (iv) methods to treat ties proposed in the two different models can borrow strength from each other and give us more options; and (v) more importantly, many extensions to the Cox model, namely handling of right censored observations and use of time dependent covariates, can be introduced to the BT model in analyzing rank order data and give us interesting results. For details please see Su (2003).
In multiple comparisons, we have $K$ subjects in a comparison. Let \( i_1, \ldots, i_K \) be the set of indices of the subjects engaged in a comparison. As in the paired comparison case, we have the following theorem.

**Theorem 2.** Let \( T_{i_1}, \ldots, T_{i_K} \) be independent random variables with distributions \( F_{i_1}(t, \pi_{i_1}), \ldots, F_{i_K}(t, \pi_{i_K}) \), respectively. Then,

\[
P(T_{i_1} > \cdots > T_{i_K}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dF_{i_1}(t_1, \pi_{i_1}) \cdots dF_{i_{K-1}}(t_{K-1}, \pi_{i_{K-1}}) dF_{i_K}(t_K, \pi_{i_K}).
\]

If furthermore, \( T_{i_1}, \ldots, T_{i_K} \) satisfy the stratified Cox proportional hazards assumption (4) and are in one stratum, then the BT model will assign the probability correctly to the event \( \{T_{i_1} > \cdots > T_{i_K}\} \), i.e.

\[
P(T_{i_1} > \cdots > T_{i_K}) = \frac{\prod_{j=2}^{K} \pi_{i_j}}{\sum_{j=1}^{K} \pi_{i_j}}.
\]

The contribution to the partial likelihood from observations \( T_{i_1}, \ldots, T_{i_K} \) with \( \{T_{i_1} > \cdots > T_{i_K}\} \) is the same as the corresponding likelihood given by the BT model in (2).

### 4. Application

Now we illustrate fitting the BT model using the \texttt{coxph} procedure in R. All the observations for a game should be in one stratum. In Example 1, each game yields two observations for two subjects. Both observations should be in a stratum. After entering the data in counting process style (cf. SAS Institute, Inc., 2000), the following output can be obtained. (For complete R and SAS code please see web page \texttt{www.ms.uky.edu/~mai/research/ExampleBTcox}.)

Call:
\[
\text{coxph(formula = Surv(ta, tp, ev) ~ z2 + z3 + z4 + z5 + z6 + z7 + strata(games), weights = sc, method = "efron")}
\]

<table>
<thead>
<tr>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>z2</td>
<td>-0.145</td>
<td>0.865</td>
<td>0.311</td>
<td>-0.466</td>
</tr>
<tr>
<td>z3</td>
<td>-0.287</td>
<td>0.751</td>
<td>0.310</td>
<td>-0.925</td>
</tr>
<tr>
<td>z4</td>
<td>-0.334</td>
<td>0.716</td>
<td>0.310</td>
<td>-1.076</td>
</tr>
<tr>
<td>z5</td>
<td>-0.474</td>
<td>0.623</td>
<td>0.311</td>
<td>-1.525</td>
</tr>
<tr>
<td>z6</td>
<td>-0.898</td>
<td>0.408</td>
<td>0.317</td>
<td>-2.835</td>
</tr>
<tr>
<td>z7</td>
<td>-1.581</td>
<td>0.206</td>
<td>0.343</td>
<td>-4.607</td>
</tr>
</tbody>
</table>

Likelihood ratio test = 34.0 on 6 df, \( p = 6.84e-06 \) n = 84

The coef for \( z1 \) is fixed at zero and thus \( \exp(\text{coef}) = 1 \). The estimates of \( \pi \)'s given by the \texttt{coxph} procedure are 1.0865, 0.751, 0.716, 0.623, 0.408 and 0.206. They are for Milwaukee, Detroit, Toronto, New York, Boston, Cleveland and Baltimore respectively. Using (1), the predicted probability of the event that New York would defeat Baltimore is 0.716/(0.716 + 0.206) \( \approx 0.777 \). The predicted probability of the event that New York would defeat Boston is 0.716/(0.716 + 0.623) \( \approx 0.535 \), etc. The results agree with Agresti (2002).

### 5. Discussion—ties

To cover the possibility of an appreciable amount of ties, Cox (1972) suggests the conditional logistic likelihood. If the continuous time assumption can be made, Therneau and Grambsch (2000) address different methods to handle tied data and their computation issues. Those methods can be easily applied to the rank-order data analysis. Davidson (1970) discusses an extension of the BT model for paired comparisons to situations which allow an expression of no preference, and compared its performance with a model proposed by Rao and Kupper (1967). Beaver and Rao (1973) have developed a model to account for ties in triple...
comparisons. Unlike the approach of Cox, these extensions of the BT model introduce an additional parameter.

6. Acknowledgements

Professors Ali (1977) and Zelterman (2002) kindly offered invaluable suggestions during the preparation of this paper.

References