AN ALTERNATIVE TO THE PROPORTIONAL HAZARDS MODEL

BY

TERENCE J. O'NEILL

TECHNICAL REPORT NO. 93
JULY 1984

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 5 R01 GM21215-10

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
AN ALTERNATIVE TO THE PROPORTIONAL HAZARDS MODEL

by

Terence J. O'Neill

TECHNICAL REPORT NO. 93
July 1984

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 5 R01 GM21215-10

Also prepared under the Auspices of National Science Foundation Grant MCS80-24649 and issued as Technical Report No. 218, Stanford University, Department of Statistics.

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
AN ALTERNATIVE TO THE PROPORTIONAL HAZARDS MODEL

by

Terence J. O'Neill
Australian National University

ABSTRACT

The proportional hazards model (PHM) is a very successful model for censored data because of

(i) easy estimation,
(ii) no model for the underlying hazard is required,
(iii) highly efficient estimation is possible and,
(iv) easy incorporation of covariate information.

An alternative model called the inverse proportional hazards model (IPHM) is proposed in this paper and it is shown that it retains most of the attractive features of the PHM. It is only marginally more difficult to estimate. Although the IPHM is closely related to a PHM, it is quite different in terms of the relationships it implies across the covariates. It should prove to be a useful alternative to the PHM.

Keywords: Proportional Hazards; Hazard; Cox Regression; Inverse Proportional Hazards.
1. Introduction

The proportional hazards model (PHM) which was introduced by Cox (1972) has become the standard "non-parametric" or "distribution free" method for the analysis of censored failure time data with covariates. The usual definition of the hazard is that if $S(t)$ is a continuous survival function then the associated hazard function is

$$h(t) = -\frac{S'(t)}{S(t)}$$

with an equivalent definition for discrete distributions. The PHM says that the effect of the covariates is multiplicative on the hazard and that the hazard at $t$ for an individual $\lambda$ with covariate value $z_\lambda$ is

$$h_\lambda(t) = h_0(t)\gamma_\lambda$$

where $\gamma_\lambda$ is some function of $z_\lambda$ and some unknown parameters $\beta$, usually taken to be for mathematical convenience

$$\gamma_\lambda = \exp(\beta'z_\lambda)$$

Along with the PHM, Cox (1972) proposed an estimation procedure for $\beta$ which avoided any consideration or estimation of $h_0(t)$ and hence $S_0(t)$ by using the partial likelihood:
If \( \mathcal{R}(t) \) is the set of individuals known to be alive at \( t \), then the probability that individual \( (i) \) fails at \( t(i) \), where, assuming there is only one death at each time point, the ordered times of death are

\[ t(1) < \ldots < t(k), \]

given \( \mathcal{R}(t(i)) \) and the fact that a death occurs at \( t(i) \) is

\[ \frac{\gamma(i)}{\sum_{\lambda \in \mathcal{R}(t(i))} \gamma(\lambda)} = \frac{\exp(\beta'z(i))}{\sum_{\lambda \in \mathcal{R}(t(i))} \exp(\beta'z(\lambda))}. \]

The partial likelihood is defined as

\[ PL = \prod_{i=1}^{k} \frac{\gamma(i)}{\sum_{\lambda \in \mathcal{R}(t(i))} \gamma(\lambda)} \]

and this is maximized to obtain an estimate of \( \hat{\beta} \), \( \tilde{\beta} \), avoiding any consideration of the underlying hazard function \( h_0(t) \).

The PHM and PL approach has been shown to be a reasonable procedure with high efficiency by several people (Cox 1975, Efron 1977, Oakes 1977, Kay 1977). Thus its virtues are

(i) easy estimation,
(ii) no basic hazard required,
(iv) quite efficient,

(iii) easy incorporation of covariate information.

However it should be remembered that the PHM is a fairly strong assumption in itself and so we should not always expect the model to fit. For example, in terms of survival distributions, the PHM is stated as

\[ S_\lambda(t) = S_0(t) \gamma_\lambda \]

or

\[ \log(-\log(S_\lambda(t))) = \log \gamma_\lambda + \log(-\log(S_0(t))) \quad (1.1) \]

Thus the survival distribution at different levels of the covariate can never cross (unless the covariate is time dependent). One check of the model if there are strata with appreciable numbers is to replace \( S_\lambda \) in (1.1) with the Kaplan Meier estimate for the stratum and see if the parallel relationship (1.1) holds across the strata.

If this or one of the other possible checks indicates that the model does not fit well then it would be nice to have another model to try which at least retains some of the desirable features of the PHM.

This paper describes one possible alternative model called the Inverse Proportional Hazards Model (IPHM) which is closely related to the PHM. The IPHM retains almost all of the attractive features of the PHM and yet is very distinct in terms of the model fitted.
Section 2 gives a reliability interpretation to motivate both the PHM and IPHM which can also be used to embed the models in a larger class. Section 3 discusses appropriate likelihoods for both models and Section 4 considers estimation of the parameters and of the covariance matrix of the estimators. In Section 5, all the models are applied to an example. In Section 6 some remarks on choices between the models are made and a discussion is given in Section 7.
2. A Reliability Interpretation

We have seen that one form of the PHM is

\[ S_\lambda(t) = (S_0(t))^{\gamma_\lambda} . \]

If we think of this relationship in a reliability context and let \( \gamma_\lambda \) be a number of components and \( S_0(t) \) the survival distribution of each component, then \( S_\lambda(t) \) is the survival distribution of a system where we wire the components up in series and assume the components act independently.

Figure 1. Pictorial Representation of the IPHM and PHM

Hence under the PHM model, the system fails when the first component fails. Clearly there are other ways that you can wire up a system. One obvious way is a parallel system where the system lasts until the last component fails.
If $F$ is used to denote a distribution function then the distribution function of the survival time of the system if only one component is required to keep it operating is

$$F_t(t) = F(t)^{\gamma_t}.$$  \hspace{1cm} (2.1)

This model will be termed the Inverse Proportional Hazards Model (IPHM) for the following reason. Suppose $T$ has distribution function $F_t$ and let $V = 1/T$. Then $V$ has survival function

$$\Pr(V \geq v) = \Pr(T \leq 1/v)$$
$$= F_t(1/v)$$
$$= F_0(1/v)^{\gamma_t}$$
$$= S_0(v)^{\gamma_t}$$

where

$$S_0(v) = F_0(1/v).$$

Thus if a set of survival times satisfies an IPHM then the inverse survival times satisfy a PHM. The relevance of this fact to the estimation process will be explored in the next section.
A more general model which contains both the PHM and the IPHM is a parallel system where failure of the system occurs when either a certain number or a certain proportion of the components have failed. This model will be explored further later.
3. The PHM and IPHM Likelihoods

As noted in the previous section, inverse transformation of failure times converts an IPHM to a PHM. However that does not necessarily mean that the PL is appropriate. If there are no censorings then the PL method is applicable and there are no complications. However a censored time on the reciprocal scale means that the individual died before, rather than after, the time of censoring and so there is no way that the PL can be constructed in the usual manner.

Let us reconsider the problem on the original time scale and see what happens when we try to form the usual partial likelihood. The hazard function for (2.1) is

$$h_\ell(t) = \gamma_\ell f_0(t) F_0(t)^{\gamma_\ell - 1}\left[1 - F_0(t)^{\gamma_\ell}\right]$$

and when we take the ratio of hazard functions,

$$\frac{h_{\ell_1}(t)}{h_{\ell_2}(t)} = \frac{\gamma_{\ell_1}}{\gamma_{\ell_2}} \cdot \frac{F_0(t)^{\gamma_{\ell_1}}}{1 - F_0(t)^{\gamma_{\ell_1}}} \cdot \frac{1 - F_0(t)^{\gamma_{\ell_2}}}{F_0(t)^{\gamma_{\ell_2}}}$$

So $F_0(t)$ would appear in a partial likelihood in a particularly awkward form.
There is another facet of the PHM that has not been mentioned yet. Even with the PL approach, it is usual to also attempt to estimate the underlying survival distribution and this is usually done by parameterizing $S_0(t)$. A discrete model is taken with hazard $1 - \alpha_i$ at $t_{(i)}$. A similar parameterization is often used to show that the Kaplan Meier estimate is the "non-parametric maximum likelihood estimator" when there are no covariates. Assuming for the moment that $\beta$ is known and that there is a single death at each time point, the likelihood for $\alpha_i$ is then

$$
\left(1 - \alpha_i^{-Y(i)}\right)^{\gamma(i)} \prod_{\ell \in R(t_{(i)})} \alpha_i^{\gamma_{\ell}}
$$

(3.1)

and maximizing this with respect to $\alpha_i$, we obtain

$$
\hat{\alpha}_i = \left(1 - \gamma(i) \prod_{\ell \in R(t_{(i)})} \gamma_{\ell}\right)^{1/\gamma(i)}
$$

(3.2)

Since $\beta$ is not usually known, the common practice is to replace $\beta$ with the PL estimate $\tilde{\beta}$ in (3.1) and (3.2).

An alternative approach mentioned by Kalbfleisch and Prentice (1980) and attributed by them to Meier which was further discussed by Bailey (1979, 1983, 1984) is to treat both $\alpha_i$ and $\beta$ as unknown in (3.1) and estimate them both by maximum likelihood to obtain $\hat{\alpha}$ and $\hat{\beta}$. Thus we are led to the proportional hazards likelihood
\[
\text{PHL} = \prod_{i=1}^{k-1} \left[ \prod_{\lambda \in D_i} \left( 1 - \alpha_i \right)^{\gamma_i} \right] \prod_{\lambda \in R(t_{(i)})} \gamma_i \alpha_i
\]  

(3.3)

where \( D_i \) is the set of individuals who die at \( t_{(i)} \). Maximization of this likelihood is discussed in the next section.

Since the partial likelihood approach to the IPHM does not work, we would like to construct a similar likelihood to (3.3) for the IPHM. Under the PHM for the reciprocal times, the parameterization for \( V_0 \) which has \( \gamma_0 = 1 \) was, using \( 1 - \delta_i \) for the hazard at \( v_{(i)} \),

\[
\Pr(V_0 \geq v) = \prod_{v_{(i)} \leq v} \delta_i
\]

and hence

\[
P(T_0 \leq t) = P(V_0 \geq 1/t)
\]

\[
= \prod_{v_{(i)} \leq 1/t} \delta_i
\]

\[
= \prod_{t_{(i)} \geq t} \alpha_i
\]

say, and thus take

\[
F_0(t) = \prod_{t_{(i)} \geq t} \alpha_i
\]
where $\alpha_i = 0$ since we require $F_0(t) = 0$, $t < t_{(1)}$. This yields an
IPHM likelihood of

$$\text{IPHL} = \prod_{i=2}^{k} \left\{ \prod_{\ell \in D_i} \left( 1 - \alpha_{i} \right) \prod_{\ell \in C_i} \left( 1 - A_{i} \gamma_{\ell} \right) \prod_{\ell \in E_i} \alpha_{i} \gamma_{\ell} \right\} \quad (3.4)$$

where

$$A_{i} = \prod_{j \geq i} \alpha_{j},$$

$C_i = \{\text{individuals censored in } (t_{(i-1)}, t_{(i)}]\}$ and $E_i = \{\text{individuals who died before } t_{(i)}\}, i = 1, \ldots, k$. Note that as usual we are assuming that the largest observation is uncensored (or it is treated as such if it is censored).

This is a more difficult likelihood than the PHL because of the terms $1 - A_i \gamma_{\ell}$. Methods of maximization are considered in the next section.
4. Estimation

This section describes the general method of maximizing the likelihoods (3.3) and (3.4). Complete details of the method are given in Appendix A. For clarity in this section we will assume that only one death occurs at each time point. Appendix A describes the more general case of multiple deaths. Since the PHL is simpler we will consider it first for insight. The PHL has been discussed in detail by Bailey (1979, 1983, 1984) who calls it the general ML method for joint estimation of \( \hat{\beta} \) and \( S_0 \).

Differentiating the logarithm of PHL with respect to \( \alpha_i \) and equating it to zero, we obtain

\[
\hat{\gamma}(i) \left[ 1 - \hat{\alpha}(i) \right]^{-1} = \sum_{\ell \in R(t_i)} \hat{\gamma}_\ell
\]

or

\[
\hat{\alpha}_i = \left( 1 - \frac{\hat{\gamma}(i)}{\sum_{\ell \in R(t_i)} \hat{\gamma}_\ell} \right)^{1/\hat{\gamma}(i)}.
\]

Note the similarity of this equation to (3.2). Thus we only need solve

\[
\frac{\partial \hat{G}}{\partial \hat{\beta}} = 0
\]

(4.1)
where we regard the $\hat{a}$'s as functions of $\hat{\beta}$ given implicitly by the equations

$$\frac{\partial l}{\partial \alpha_i} = 0.$$ 

The equation (4.1) can be solved by a Newton-Rhapson procedure but it should be remembered that the $\hat{a}$'s are to be regarded as functions of $\hat{\beta}$ and so the appropriate matrix to use in the Newton-Rhapson procedure is

$$\left( \frac{\partial^2 l}{\partial \beta \partial \beta'} + k-1 \sum_{i=1}^{k-1} \frac{\partial^2 l}{\partial \beta \partial \alpha_i} \frac{\partial \alpha_i}{\partial \beta'} \right)^{-1}$$

where since $\frac{\partial l}{\partial \alpha_i} = 0$,

$$\frac{\partial \alpha_i}{\partial \beta'} = -\frac{\partial^2 l}{\partial \alpha_i \partial \beta'} / \frac{\partial^2 l}{\partial \alpha_i^2}$$

and so the matrix for the Newton-Rhapson procedure is

$$\left\{ \frac{\partial^2 l}{\partial \beta \partial \beta'} - k-1 \sum_{i=1}^{k-1} \left( \frac{\partial^2 l}{\partial \alpha_i \partial \beta'} \right)^{-1} \left( \frac{\partial^2 l}{\partial \alpha_i \partial \beta'} \right) \right\}^{-1} = -V_{22}$$
say. By Appendix B, it follows that this also estimates the asymptotic variance matrix of \( \hat{\beta} \) and

\[
\frac{\partial \alpha}{\partial \beta'} \begin{pmatrix}
V_{22} &= V_{12} \\
\end{pmatrix} = V_{21}
\]

estimates the covariance of \( \hat{\alpha} \) and \( \hat{\beta} \).

It remains to estimate the variance of \( \hat{\alpha} \). Let

\[
\frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \beta'} = C = \begin{bmatrix}
C_{11} & C_{12} \\
C_{21} & C_{22}
\end{bmatrix}
\]

where \( C_{22} \) is \( p \times p \) where \( p \) is the dimension of \( \beta \). Let

\[
-C^{-1} = V = \begin{pmatrix}
V_{11} & V_{12} \\
V_{21} & V_{22}
\end{pmatrix}
\]
Then we already have $V_{22}$ and $V_{12}$ and

$$V_{11} = - C_{11}^{-1} + C_{11}^{-1} C_{12} V_{21}$$

and hence since $C_{11}$ is diagonal, the covariance matrix for $\hat{\gamma}$, $\hat{\beta}$ is easily estimated. Thus the computation of $\hat{\gamma}$ and $\hat{\beta}$ and the estimation of their covariance matrix for the PHL is fairly easy and only requires the inversion of a $p \times p$ matrix.

The procedure for the IPHL is similar. Note however that we cannot solve for $\hat{\alpha}_i$ in terms of $\hat{\beta}$ only. Also, $\frac{\partial^2 \ell}{\partial \alpha \partial \beta}$ has no zero entries and so initially we are faced with a maximization of a likelihood which has $k - 1 + p$ unknown parameters. If $k$ is at all large then this will not be computationally feasible. Instead of attempting this maximization we convert to an equivalent system of equations which is computationally more manageable. Let

$$f_i = \alpha_i \frac{\partial \ell}{\partial \alpha_i} - \alpha_{i+1} \frac{\partial \ell}{\partial \alpha_{i+1}}, \ i = 2, \ldots, k - 1$$

$$f_k = \frac{\partial \ell}{\partial \alpha_k},$$

and

$$f_{k+1} = \frac{\partial \ell}{\partial \beta}.$$
Then

\[
\tilde{f} = \begin{pmatrix}
\tilde{f}_1 \\
\cdots \\
\tilde{f}_k \\
\tilde{f}_{k+1}
\end{pmatrix} = D \frac{\partial \tilde{L}}{\partial \tilde{z}}
\]

where

\[
D = \begin{pmatrix}
\alpha_2 & -\alpha_3 & 0 \\
\alpha_3 & -\alpha_4 & \alpha_{k-1} & -\alpha_k \\
0 & 1 & 0 \\
0 & 1 & 1
\end{pmatrix}
\]

Since the \( \alpha \)'s are constrained to lie between 0 and 1, the solutions of \( \tilde{f} = 0 \) and \( \frac{\partial \tilde{L}}{\partial (\tilde{z}', \tilde{\beta}')} = 0' \) are identical. Now \( f_i = 0 \) is equivalent to

\[
\hat{\alpha}_i = \left[ 1 - \hat{\gamma}_i \left( 1 - \hat{\alpha}_{i+1} \right)^{-1} \hat{\gamma}_{i+1} \right]^{-1} \hat{\gamma}_{i+1} + \sum_{\ell \in C_{i+1}} \left( \hat{\gamma}_{\ell} \left( 1 - \hat{\alpha}_{i+1} \right)^{-1} \hat{\gamma}_{\ell} \right)^{-1} \hat{\gamma}_{\ell} \right]^{1/\hat{\gamma}_i}
\]

(4.2)
and note that the right hand side of (4.2) only involves \( \hat{\alpha}_{i+1}, \ldots, \hat{\alpha}_k, \hat{\beta} \). Hence we can sequentially express \( \hat{\alpha}_k, \hat{\alpha}_{k-1}, \ldots, \hat{\alpha}_2 \) in terms of \( \hat{\alpha}_k \) and \( \hat{\beta} \). Thus we wish to solve \( f_k = 0, f_{k+1} = 0 \) for \( \hat{\alpha}_k, \hat{\beta} \) regarding \( \hat{\alpha}_{k-1}, \ldots, \hat{\alpha}_2 \) as implicit function of \( \hat{\alpha}_k \) and \( \hat{\beta} \). For a Newton-Raphson procedure we require

\[
\begin{pmatrix}
\frac{\partial f_k}{\partial \hat{\alpha}_k} \\
\frac{\partial f_{k+1}}{\partial \hat{\beta}}
\end{pmatrix} = A^{-1}
\begin{pmatrix}
\frac{\partial f_k}{\partial (\hat{\alpha}_k, \hat{\beta})} \\
\frac{\partial f_{k+1}}{\partial (\hat{\alpha}_k, \hat{\beta})}
\end{pmatrix}
\]

say, and thus we require \( \frac{\partial \alpha_i}{\partial \alpha_k}, \frac{\partial \alpha_i}{\partial \hat{\beta}}, i = 2, \ldots, k-1 \). Now

\[
\frac{\partial f_i}{\partial \alpha_i} \frac{\partial \alpha_i}{\partial \hat{\beta}} + \frac{\partial f_i}{\partial \alpha_{i+1}} \frac{\partial \alpha_{i+1}}{\partial \hat{\beta}} + \cdots + \frac{\partial f_i}{\partial \alpha_{k-1}} \frac{\partial \alpha_{k-1}}{\partial \hat{\beta}} + \frac{\partial f_i}{\partial \hat{\beta}} = 0
\]

or

\[
\frac{\partial \alpha_i}{\partial \hat{\beta}} = -\left( \frac{\partial f_i}{\partial \hat{\beta}} + \frac{\partial f_i}{\partial \alpha_i} \frac{\partial \alpha_i}{\partial \hat{\beta}} + \cdots + \frac{\partial f_i}{\partial \alpha_{k-1}} \frac{\partial \alpha_{k-1}}{\partial \hat{\beta}} \right) / \frac{\partial f_i}{\partial \alpha_i}
\] (4.3)
with a similar relationship holding for $\alpha_k$ with $\alpha_k$ replacing $\beta$ in (4.3). Hence $\frac{\partial \alpha_1}{\partial \alpha_k}$ and $\frac{\partial \alpha_1}{\partial \beta}$ can be found in the same sequential manner as $\alpha_{k-1}, \ldots, \alpha_2$.

Note that

$$
\frac{\partial}{\partial (\alpha_k, \beta')}
\begin{bmatrix}
\alpha_2 \\
\vdots \\
\alpha_{k-1}
\end{bmatrix}
= 
\begin{bmatrix}
f_2 \\
f_{k-1}
\end{bmatrix}
^{-1}
\begin{bmatrix}
f_2 \\
f_{k-1}
\end{bmatrix}
\frac{\partial}{\partial (\alpha_2, \ldots, \alpha_{k-1})}
\frac{\partial}{\partial (\alpha_k, \beta')}
$$

but using (4.3) is computationally much quicker. So we have reduced the problem to a Newton-Rhapson procedure on $p+1$ variables instead of $k-1+p$.

Once again it remains to estimate the covariance matrix of $\hat{\alpha}$ and $\hat{\beta}$ avoiding the inversion of a $(k-1+p) \times (k-1+p)$ matrix if possible.

Recall that

$$
D \frac{\partial \ell}{\partial (\alpha, \beta)} = f
$$

Thus since

$$
\frac{\partial \ell}{\partial (\alpha, \beta)} = 0
$$
at $\hat{\alpha}, \hat{\beta}$ it follows that

$$D \frac{\partial^2 \lambda}{\partial \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \partial \left( \begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} \right)} = \begin{pmatrix} \partial f \\ \partial \alpha' \partial \beta' \end{pmatrix}$$

or

$$V = \left( - \frac{\partial^2 \lambda}{\partial \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \partial \left( \begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} \right)} \right)^{-1} D^{-1} = \begin{pmatrix} \partial f \\ \partial \alpha' \partial \beta' \end{pmatrix} \begin{pmatrix} \partial \alpha' \partial \beta' \\ \partial \alpha' \partial \beta' \end{pmatrix}^{-1}$$

at $\hat{\alpha}, \hat{\beta}$ where $V$ is the estimated covariance matrix of $(\hat{\alpha}', \hat{\beta}')$.

Hence if

$$C = \begin{pmatrix} \partial f \\ \partial \hat{\alpha}' \partial \hat{\beta}' \end{pmatrix}$$

then the estimated covariance matrix of $\hat{\alpha}, \hat{\beta}$ is $\sim C^{-1} \sim$. Thus we would like to invert $C$. Now let

$$C^{-1} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

where $C_{22}$ is $(p+1) \times (p+1)$. 
From Appendix B

\[ \tilde{C}^{22} = A \]

and

\[
\begin{bmatrix}
\alpha_2 \\
\vdots \\
\alpha_{k-1}
\end{bmatrix}
\tilde{A}^{(\alpha_k \beta')} = \tilde{C}^{12} \]

It remains to find \( \tilde{C}^{11} \) and \( \tilde{C}^{21} \). Note that

\[
\tilde{C} = \begin{bmatrix}
\tilde{C}_{11} & \tilde{C}_{12} \\
\tilde{C}_{21} & \tilde{C}_{22}
\end{bmatrix}
\]

where \( \tilde{C}_{11} \) is \((k-1) \times (k-1)\) and upper triangular and hence is easy to invert. Thus obtain \( \tilde{C}^{11} \) and \( \tilde{C}^{21} \) using the relationships

\[
\tilde{C}^{11} = \tilde{C}_{11}^{-1} - \tilde{C}^{12} \tilde{C}_{21} \tilde{C}_{11}^{-1}
\]

and

\[
\tilde{C}^{21} = -\tilde{C}_{22}^{-1} \tilde{C}_{21} \tilde{C}^{11}
\]

Thus \( \tilde{C}^{-1} \) and hence \( \tilde{V} \) is quite easy numerically to find.
Thus the solution of the IPHL is only slightly more difficult than that of the PHL. The asymptotic properties of the PHL estimates have been established by Bailey (1979, 1983, 1984). The same properties can be established in the same manner for the IPHL using the inverse relationship. This will be carried out by the author in another paper. In the next section both methods are applied to an example.
5. An Example

Consider the data of Table 1.1, page 2, Kalbfleisch and Prentice (1980) which was originally from Pike (1966).

Table 1

<table>
<thead>
<tr>
<th>Days to Vaginal Cancer Mortality in Rats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Group 2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Censored.

A preliminary plot was made in Figure 2 to test the appropriateness of both the PHM and IPHM for this set of data. The complementary log log plots of the Kaplan Meier estimates seem to indicate that both models should fit the data with perhaps the IPHM being more suitable.

Both models were filled to the data. For PHM the model is:

Group 1: Survival Distribution $S_0(t)$

Group 2: Survival Distribution $S_0(t)^{exp(\beta)}$
Using Kaplan Meier estimates for each group complement log-log plots for carcinogenesis data.

Figure 2
The PHL yielded

\[ \hat{\beta} = -0.595, \quad \hat{\sigma}_\beta = 0.348 \]

and

\[ \text{deviance} = -2 \log L(\hat{\alpha}, \hat{\beta}) = 268.5. \]

By contrast the PL yielded

\[ \tilde{\beta} = -0.596, \quad \tilde{\sigma}_\beta = 0.289. \]

The discrepancy between \( \hat{\sigma}_\beta \) and \( \tilde{\sigma}_\beta \) may warrant further investigation.

The IPHM is:

Group 1: Distribution Function \( F_0(t) \)

Group 2: Distribution Function \( F_0(t)^{\exp(\beta)} \)

The IPHL gave

\[ \hat{\beta} = 0.248, \quad \hat{\sigma}_\beta = 0.328 \]

and

\[ \text{deviance} = -2 \log L(\hat{\alpha}, \hat{\beta}) = 224.7. \]

Note that the \( \hat{\beta} \)'s for the two models have different signs as expected. The difference between the groups is much less compelling in the IPHM.
The deviance for the IPHL is much lower than the deviance for PHL, indicating a possible superiority of the IPHM. This point will be discussed further in the next section.

Finally, the estimates of the distributions for the models are plotted for each group with the Kaplan Meier estimates in Figures 3 and 4. There is very little to choose between the estimates on these plots.

The issue of choosing between the models is discussed in more detail in the next section.
Estimated Distribution Functions for Carcinogenesis Data

FIGURE 3
Estimated Distribution Functions for Carcinogenesis Data

Group 2

Distribution Functions

- Dots - Kaplan-Meier
- Dashes - Proportional Hazards
- Solid - Inverse Proportional Hazards

Lines are

Time
6. Choice of Models

In this section, we wish to consider how you can choose between two non nested models in a "non-parametric" censored data with covariates context. Two possible methods, stratum plots and embedding in a larger model, which assist in the choice of a model have already been discussed. Two other possible methods are residual analysis and using some overall measure of fit.

A method of residual analysis for the PHM was described by Lawless (1982). In the PHM, we have that for an uncensored survival time \( t \),

\[-\log S_\lambda(t) \sim \text{standard exponential}\]

and so a residual is defined by

"residual" = \( -\hat{\gamma}_1 \log \hat{S}_0(t_i) \).

The residuals should be approximately a random sample from a standard exponential with censored residuals being treated as censored exponentials. Thus if we plot \(-\log(\text{Kaplan Meier Estimate})\) it should result in a straight line with intercept 0 and slope 1. Censored times are inconvenient when we wish to plot residuals against other factors and so it is common to correct the censored residuals using the fact that if \( Y \) is standard exponential then

\[ \mathbb{E}[Y | Y > y] = 1 + y. \]
Hence for a censored \( t_i \) use

\[
1 - \log(\hat{s}_0(t_i) \hat{\gamma}_L)
\]

as the residual. The residual analysis for the IPHM is identical with \( F \) replacing \( S \).

In order to choose adequately between the two models, we need some overall measure of fit. If we were in a log linear parametric format, then an approach of the type of Efron (1983b) would be appropriate. In the "non-parametric" maximum likelihood context, two possibilities for some overall measure are the deviance and some measure of prediction error.

Define the deviance as

\[
E[-2 \log L(t_1, \hat{\alpha}, \hat{\beta})]
\]

where \( t_1 \) has the same experimental set up as the data at hand. The sample deviance for the data \( t \) is

\[
-2 \log L(t, \hat{\alpha}, \hat{\beta})
\]

Recall that the model is chosen to give the lowest possible sample deviance and so we can expect it to be a downward biased estimator of the deviance. Hence we cannot be sure if a difference in sample deviances for non nested models is solely due to these biases. We require
some more unbiased estimate. This problem is analogous to estimating misclassification error for classification rules and several recommendations for estimation have been recently made by Efron (1983a). Estimates of a similar type (cross-validated, bootstrap) can be expected to work equally well in this situation. Since $L$ in these models is not really a likelihood, it would be nice to have another quantity other than deviance to assist in the choice of a model.

Since we are interested in estimating distribution functions or survival distributions and their relationship across covariates, it may be more natural to consider a prediction error criterion which directly involves the distribution function. For an uncensored observation $t$ with estimated distribution function $\hat{F}$ and true distribution function $F$, an appropriate error might be a function of $F(t) - \hat{F}(t)$. Since we do not know $F(t)$, replace it by its expectation, $1/2$, and so a suitable error might be a function of

$$|1/2 - \hat{F}(t)|.$$

One nice such function is

$$-2 \log [2 \min(\hat{F}(t), 1 - \hat{F}(t))].$$

Note that if $F$ were the true distribution then this would have a $\chi^2_2$ distribution. Then the prediction error is defined as
where \( t_{11} \) are new observations obtained under the same experimental conditions as \( t \). For censored observations, for the same reasons as in the residuals, we use \( \hat{F}(t)/e \) instead of \( \hat{F}(t) \).

Note that we can use \( F \) and \( S \) interchangeably in (6.1). Although the sample value of \( P \) is not likely to be as strongly biased as the value of the deviance, it would still be wise to use an estimation procedure of the type suggested by Efron (1983a) such as cross-validation or bootstrap.

The use of these criteria is being further investigated by the author.

7. Discussion

A viable alternative to the Proportional Hazards Model, the Inverse Proportional Hazards Model has been introduced and discussed. It is closely related to the PHM, but is very distinct in terms of the implications about the relationships across covariates.

Thus it should prove to be a useful class of models to compare with the PHM. It has been shown to be only slightly more difficult computationally than the PHM.
Appendix A: Estimation for the IPHM and PHM

I. IPHM

Recall that the likelihood is given by (3.4). Then

\[ \frac{\partial^2 L}{\partial \alpha_i \partial \alpha_i} = \{-u_{1i} - \sum_{j=2}^{i} v_{ij} + t_i - t_0\} \alpha_i^{-1}, \]

\[ \frac{\partial^2 L}{\partial \alpha_i \partial \alpha_j} = \begin{cases} i \sum_{m=2}^{i} v_{2m} (\alpha_i \alpha_j)^{-1}, & i < j \\ \sum_{m=2}^{k} \log A_j - (\log \alpha_j) w_{ij} - (\log A_j) x_{ij} \end{cases} \]

\[ \frac{\partial^2 L}{\partial \beta \partial \beta'} = \begin{cases} \sum_{i=1}^{k} \log A_j - \sum_{j=2}^{k} \log \alpha_j + w_{3j} (\log \alpha_j)^2 + x_{2j} \log A_j \end{cases} \]

and

\[ \frac{\partial^2 L}{\partial \beta \partial \alpha_i} = \{r_i - \sum_{i=0}^{w_{4i}} \log \alpha_i - \sum_{j=2}^{i} (x_{1j} + x_{4j} \log A_j)\} \alpha_i^{-1} \]

where
\[ u_{1i} = \sum_{\lambda \in D_i} \gamma_{\lambda} (1 - \alpha_{\lambda})^{-1}, \]
\[ u_{2i} = \sum_{\lambda \in D_i} \gamma_{\lambda}^2 \alpha_{\lambda} (1 - \alpha_{\lambda})^{-2}, \]
\[ v_{1i} = \sum_{\lambda \in C_i} \gamma_{\lambda} (1 - A_{\lambda})^{-1}, \]
\[ v_{2i} = \sum_{\lambda \in C_i} \gamma_{\lambda}^2 A_{\lambda} (1 - A_{\lambda})^{-2}, \]
\[ t_0 = \sum_{\lambda \in C_1} \gamma_{\lambda}, \]
\[ t_{1i} = \sum_{\lambda \in D_i \cup C_i} \gamma_{\lambda}, \]
\[ t_i = \sum_{j=1}^{i} t_{1j}, \]
\[ s_{1i} = \sum_{\lambda \in D_i} \frac{\partial \gamma_{\lambda}}{\partial \beta}, \]
\[ s_{2i} = \sum_{\lambda \in D_i} \frac{\partial^2 \gamma_{\lambda}}{\partial \beta \partial \beta'}, \]
\[ r_{1i} = s_{1i} + \sum_{\lambda \in C_i} \frac{\partial \gamma_{\lambda}}{\partial \beta}, \]
\[ r_{2i} = s_{2i} + \sum_{\lambda \in C_i} \frac{\partial^2 \gamma_{\lambda}}{\partial \beta \partial \beta'}, \]
\[ r_i = \sum_{j=1}^{i} r_{1j}, \]
\[ X_0 = X_{11} - X_{11} , \]

\[ w_{1i} = \sum \frac{\gamma \alpha}{D_{1}} \left( 1 - \alpha \right)^{-1} \frac{\partial^2 \gamma}{\partial \beta \partial \beta'} , \]

\[ w_{2i} = \sum \frac{\gamma \alpha}{D_{1}} \left( 1 - \alpha \right)^{-1} \frac{\partial \gamma}{\partial \beta} \frac{\partial \gamma}{\partial \beta'} , \]

\[ w_{3i} = \sum \frac{\gamma \alpha}{D_{1}} \left( 1 - \alpha \right)^{-2} \frac{\partial \gamma}{\partial \beta} \frac{\partial \gamma}{\partial \beta'} , \]

\[ w_{4i} = \sum \frac{\gamma \alpha}{D_{1}} \left( 1 - \alpha \right)^{-2} \frac{\partial \gamma}{\partial \beta} \frac{\partial \gamma}{\partial \beta'} , \]

\[ x_{1i} = \sum \frac{\gamma \alpha}{C_{1}} \left( 1 - \alpha \right)^{-1} \frac{\partial \gamma}{\partial \beta} , \]

\[ x_{2i} = \sum \frac{\gamma \alpha}{C_{1}} \left( 1 - \alpha \right)^{-1} \frac{\partial \gamma}{\partial \beta} \frac{\partial \gamma}{\partial \beta'} , \]

\[ x_{3i} = \sum \frac{\gamma \alpha}{C_{1}} \left( 1 - \alpha \right)^{-2} \frac{\partial \gamma}{\partial \beta} \frac{\partial \gamma}{\partial \beta'} , \]

\[ x_{4i} = \sum \frac{\gamma \alpha}{C_{1}} \left( 1 - \alpha \right)^{-2} \frac{\partial \gamma}{\partial \beta} \frac{\partial \gamma}{\partial \beta'} . \]

and

\[ x_{4i} = \sum \frac{\gamma \alpha}{C_{1}} \left( 1 - \alpha \right)^{-2} \frac{\partial \gamma}{\partial \beta} \frac{\partial \gamma}{\partial \beta'} . \]

Then

\[ f = u_{1,i+1} + v_{1,i+1} - t_{1,i+1} - u_{1,i} , i = 2, \ldots, k-1 . \]
Also \( \hat{\alpha}_i \) can be expressed in terms of \( \hat{\alpha}_{i+1}, \ldots, \hat{\alpha}_k, \hat{\beta} \) by solving \( f_i = 0 \) or

\[
\hat{u}_{i+1} - u_{i+1} + v_{i+1} - t_{i+1} = 0, \quad i = k-1, k-2, \ldots, 2.
\]

Note that \( \alpha_k \) must be chosen large enough so that all the \( \alpha_i, \quad i = 2, \ldots, k-1 \) determined by the implicit relationship are between 0 and 1. Otherwise \( \alpha_k \) is not an allowable value. Now

\[
\frac{\partial f_i}{\partial \alpha_i} = -\alpha_i^{-1} u_{2i},
\]

\[
\frac{\partial f_i}{\partial \alpha_{i+1}} = (u_{2, i+1} + v_{2, i+1}) \alpha_{i+1}^{-1},
\]

\[
\frac{\partial f_i}{\partial \alpha_j} = v_{2, i+1} \alpha_j^{-1}, \quad j = i+2, \ldots, k,
\]

and

\[
\frac{\partial f_i}{\partial \hat{\beta}} = (\log \alpha_i^{i+1}) w_{4, i+1} + (\log A_{i+1}) x_{4, i+1} - \pi_{1, i+1} - (\log \alpha_i) w_{4, i} + w_{1, i+1} - w_{1, i} + x_{1, i+1}
\]

and hence \( \frac{\partial \alpha_i}{\partial \alpha_k} \) and \( \frac{\partial \alpha_i}{\partial \hat{\beta}} \) can be obtained using (4.3).

Now regarding \( \alpha_2, \ldots, \alpha_{k-1} \) as implicit functions of \( \alpha_k \) and \( \hat{\beta} \), we have
\[ \frac{\partial f_k}{\partial \alpha_k} = -\alpha_k^{-1} f_k - \alpha_k^{-1} \left\{ \alpha_k^{-1} u_{2k} + \sum_{j=2}^{k} z_{1j} v_{2j} \right\}, \]

\[ \frac{\partial f_k}{\partial \beta} = -\alpha_k^{-1} \left[ w_{ik} + (\log \alpha_k) w_{4k} + \sum_{j=2}^{k} \left\{ x_{1j} + (\log A_j) x_{4j} \right\} \right. \]

\[ \left. - \frac{r}{z_0} + \sum_{j=2}^{k-1} \frac{z_{2j}}{z_{1j}} \right], \]

\[ \frac{\partial f_{k+1}}{\partial \alpha_k} = \frac{s_{11} z_{12}}{z_2} + \sum_{j=2}^{k} \left( \frac{z_{1j}}{z_{2j}} \right) - \alpha_j^{-1} \left( \frac{\partial \alpha_j}{\partial \alpha_k} \right) w_{1j} - \alpha_j^{-1} (\log \alpha_j) \frac{\partial \alpha_j}{\partial \alpha_k} \cdot \left( \frac{w_{4j}}{z_{4j}} - \frac{z_{1j}}{z_{2j}} \cdot (\log A_j) \right), \]

and

\[ \frac{\partial f_{k+1}}{\partial \beta} = \frac{s_{21} \log A_2}{z_2} + \sum_{j=2}^{k} \left\{ \frac{r_{2j}}{z_{2j}} (\log A_j) - \frac{w_{2j}}{z_{2j}} \log \alpha_j - \frac{x_{2j}}{z_{2j}} \log A_j \right\} + \frac{s_{11} z_{22}}{z_2} + \sum_{j=2}^{k-1} \left( \frac{z_{1j} \cdot \frac{\partial \alpha_j}{\partial \beta}}{z_{2j}} \right) - \alpha_j^{-1} (\log \alpha_j) \frac{w_{4j}}{z_{4j}} - (\log A_j) \frac{x_{4j}}{z_{2j}} - \alpha_j^{-1} \frac{w_{1j}}{z_{1j}} \frac{\partial \alpha_j}{\partial \beta} - \frac{x_{1j} \cdot \frac{\partial \alpha_j}{\partial \beta}}{z_{2j}}, \]

where

\[ z_{1i} = \sum_{m=i}^{k} \alpha_m^{-1} \frac{\partial \alpha_m}{\partial \alpha_k}, \]
and

\[ z_{2i} = \sum_{m=1}^{k-1} \alpha_m^{-1} \frac{\partial \alpha_m}{\partial \beta}. \]

II. PHM

Recall that the likelihood is given by (3.3). Then

\[ \frac{\partial \ell}{\partial \alpha_i} = (b_{1i} - u_{1i}) \alpha_i^{-1}, \]

\[ \frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{k-1} \left( b_{2i} - w_{1i} \right) \log \alpha_i, \]

\[ \frac{\partial^2 \ell}{\partial \alpha_i \partial \alpha_j} = 0, \quad i \neq j, \]

\[ \frac{\partial^2 \ell}{\partial \alpha_i^2} = (u_{1i} - u_{2i} - b_{1i}) \alpha_i^{-2}, \]

\[ \frac{\partial^2 \ell}{\partial \alpha_i \partial \beta} = \left\{ b_{2i} - w_{1i} - (\log \alpha_i) w_{4i} \right\} \alpha_i^{-1}, \]

and

\[ \frac{\partial^2 \ell}{\partial \beta \partial \beta} = \sum_{i=1}^{k-1} \left\{ b_{3i} - (\log \alpha_i) w_{3i} - w_{2i} \right\} \log \alpha_i \]
where

\[ b_{1i} = \sum_{t(i)} \gamma_{i} \],

\[ b_{2i} = \sum_{t(i)} \frac{\partial \gamma_{i}}{\partial \beta} \],

and

\[ b_{3i} = \sum_{t(i)} \frac{\partial^{2} \gamma_{i}}{\partial \beta \partial \beta^*}. \]

Now, regarding \( \hat{\alpha}_{i} \) as an implicit function of \( \tilde{\beta} \) determined by \( \frac{\partial \tilde{\beta}}{\partial \alpha_{i}} = 0 \), we obtain

\[
\frac{\partial \alpha_{i}}{\partial \tilde{\beta}} = -\frac{\partial^{2} \gamma_{i}}{\partial \alpha_{i}} / \frac{\partial^{2} \gamma_{i}}{\partial \alpha_{i}^{2}}
\]

\[ = \alpha_{i}^{-1} \left[ b_{2i} - w_{1i} - (\log \alpha_{1}) w_{4i} \right] / (u_{2i} - u_{1i}), \]

and regarding \( f = \frac{\partial \tilde{\beta}}{\partial \tilde{\beta}} \) as a function of \( \tilde{\beta} \) only,

\[
\frac{\partial f}{\partial \tilde{\beta}} = k^{-1} \sum_{i=1}^{k-1} \left[ \left\{ b_{3i} - (\log \alpha_{1}) w_{3i} \right\} \log \alpha_{i}ight.

- \left\{ b_{2i} - w_{1i} - (\log \alpha_{1}) w_{4i} \right\} \left\{ b_{2i} - w_{1i} - (\log \alpha_{1}) w_{4i} \right\} ^{'}

\left. / (u_{2i} - u_{1i}) \right] .
\]
Suppose we have a set of equations

\[ g \left( x_1, x_2 \right) = 0, \quad (B.1) \]

\[ h \left( x_1, x_2 \right) = 0. \quad (B.2) \]

Consider using (B.1) to implicitly define \( \sim x_1 \) in terms of \( x_2 \) and then regard (B.2) as a function of \( x_2 \) only. From (B.1)

\[ \frac{\partial g}{\partial x_1} \frac{\partial x_1}{\partial x_2} + \frac{\partial g}{\partial x_1} = 0 \]

or

\[ \frac{\partial x_1}{\partial x_2} = -\left( \frac{\partial g}{\partial x_1} \right)^{-1} \frac{\partial g}{\partial x_1}. \]

From (B.2), regarding it as a function of \( x_2 \) only, we obtain the matrix necessary for a Newton-Raphson procedure as

\[ \left( \begin{array}{cc} \frac{\partial h}{\partial x_2} + \frac{\partial h}{\partial x_1} \frac{\partial x_1}{\partial x_2} \\ \frac{\partial h}{\partial x_1} \end{array} \right) = \left( \begin{array}{cc} \frac{\partial h}{\partial x_2} & \frac{\partial h}{\partial x_1} \left( \frac{\partial g}{\partial x_1} \right)^{-1} \frac{\partial g}{\partial x_1} \end{array} \right)^{-1} \approx A. \]
say. Hence if

\[ C = \mathcal{A} \left( \begin{array}{c} g' \\ h' \end{array} \right) / \mathcal{A} \left( x_1', x_2' \right) \]

\[ = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \]

and

\[ C^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \]

then

\[ C_{22}^{22} = A \]

and

\[ C_{12}^{12} = \frac{\partial x_1}{\partial x_2} A. \]

Also note that

\[ \frac{\partial x_1}{\partial x_2} A \frac{\partial x_1'}{\partial x_2'} = C_{11} - C_{11}^{-1} \]
and further that $\frac{\partial x_1}{\partial x_2}$ and hence $A, \frac{\partial x_1}{\partial x_2}^\sim A$ and $(\frac{\partial x_1}{\partial x_2})^\sim A(\frac{\partial x_1}{\partial x_2})^\sim$ are invariant with respect to the form of $g$, that is any $g$ which gives the same implicit solution for $x_1$ will yield the same expression for these quantities. However note that $C_{11}$ is not invariant.
References


