Inference for the Proportional Mean Residual Life Model

By David Oakes and Tamraparni Dasu

University of Rochester and AT & T Labs Research

We review the proportional mean residual life model for the analysis of reliability and survival data. In the single sample case with known baseline distribution the Fisher information matrix for the proportionality parameter is derived. A class of weighted ratio estimators is defined and it is shown that the right choice of weight function yields an asymptotically efficient estimator. It is conjectured that the methodology will extend to the two-sample case, i.e. with unknown baseline distribution.

1. Introduction Oakes and Dasu [15] introduced the *proportional mean residual life (PMRL)* model for the analysis of reliability and survival data . As its name suggests, in the two sample case this model implies that the mean residual life (MRL) functions for the two samples $e_j(x) = E(X_j - x|X_j > x)$, $(j = 1, 2)$ are in a constant (i.e. x-free) ratio θ,

$$e_2(x) = e(x;\theta) = \theta e_1(x). \tag{1}$$

We assume that the corresponding survivor functions $S_j(x) = P(X_j > x)$,$(j = 1, 2)$ are absolutely continuous.

In many ways the MRL function provides a more natural basis for the modeling of such data than the hazard function - the basis for Cox's proportional hazards model ([4]). The former summarizes the entire residual life distribution, whereas the latter relates only to the risk of immediate failure. In industrial reliability studies the MRL function may therefore be more important than the hazard function in the planning of strategies for maintenance and replacement. Demographers have used the *life expectancy* or *expectation of life* function $e(x) + x$ for centuries in studies of human populations. Hall and Wellner [10, 11] gave a detailed discussion of the properties of the MRL function. They characterized the class of distributions with linear MRL, $e(x) = ax + b$, and showed that the only continuous distributions with this property are the Pareto, exponential and a certain class of rescaled beta distributions.

The MRL function does have one serious disadvantage for statistical work. It is highly dependent on the tail behavior of the survivor function, and is therefore hard to estimate with precision, especially when no parametric form can be assumed. The model (1), if it is appropriate, would be expected to lead to substantial gains in efficiency in the estimation of each $e_j(x)$ for large x.

In this paper we study parametric and nonparametric methods for the analysis of data from (1). We concentrate mainly on the one-sample problem, where the baseline survivor function $S_1(x)$ is assumed known, but we also consider the the two-sample problem where both functions are unknown. In Section 2 we review some known results concerning the MRL and its sample estimate, and present conditions for the existence of a PMRL family. Section 3 considers maximum likelihood estimation of θ in the one-sample case, derives a simple expression for Fisher's information and compares the asymptotic variance of the maximum likelihood estimator (MLE) with that of the simplest nonparametric estimator, the ratio of ordinary sample means. Still in the one-sample problem, Section 4 proposes a class of non-parametric estimators for θ, based on the ratio of a linear combination of the estimated MRL functions for the sample at a series of prespecified *support points*, and show that as these support points become dense, the asymptotic variance approaches the reciprocal of Fisher's information. The corresponding vector of coefficients converges to a discrete necessarily positive delta function spike at the origin $x = 0$ and a "density", which may take positive and/or negative values, over $x > 0$. Section 5 considers the corresponding class of estimators for the two-sample problem. Explicit results are harder to obtain in this case. The optimal vector of coefficients can be determined under the null hypothesis $S_2(x) = S_1(x)$ i.e. $\theta = 1$ of equality of the two survivor functions, and also for general values of θ when the distributions are from the Pareto family. A computational recipe is given to determine the optimal vector of coefficients in the general case.

The PMRL has attracted some interest among researchers, for example Asadi [2] and Ma [12, 13] focus on multivariate extensions of the PMRL property. However the only published work on inference for the PMRL model, by Maguluri and Zhang [14], exploits the fact that the hazard function for the "forward recurrence time density" $f_R(x) \propto S(x)$ is the reciprocal of $e(x)$, so that the PMRL model for $S(x, \theta)$ translates to a proportional hazards model for $f_R(x, \theta)$. Maguluri and Zhang [14] use this relationship to derive a consistent estimator for the regression coefficients β in an extended model with the constant θ replaced by a function $\exp(\beta^{\mathrm{T}}\mathbf{z})$ of a covariate vector $\mathbf{z}$. However they do not address the efficiency properties of their estimator.

Some of the results of the present paper are contained in [6], which also includes more detailed examination of certain special cases of model (1). We acknowledge many useful discussions with Jack Hall.

2. Mean Residual Life and the PMRL Model The mean residual life function $e(x)$ for a survivor function $S(x)$ exists for all x if and only if the ordinary expectation $\mu = e(0) = E(X)$ is finite. It is well known, see [8] or [3, p. 128], that for any continuous distribution $S(x)$ is determined by its MRL function

$$e(x) = E(X - x | X > x) = \frac{\int_x^\infty S(u)du}{S(x)}$$

by the inversion formula

$$S(x) = \frac{e(0)}{e(x)} \exp\left\{ - \int_0^x \frac{du}{e(u)} \right\}. \tag{2}$$

Since $e(x)S(x) \to 0$ as $x \to \infty$ it follows that $\int_0^x \{e(u)\}^{-1}du \to \infty$ as $x \to \infty$. Also, it is easily seen that the expectation of life function $e(x) + x$ must be non-decreasing in x, so that $e'(x) \geq -1$ for all x. Integration by parts gives a simple formula for the variance $\sigma(x) = \text{var}(X - x|X > x)$ of the residual life at age x,

$$\sigma(x) = \frac{\int_x^\infty e^2(u)f(u)du}{S(x)}. \tag{3}$$

This formula, attributed by Hall and Wellner [11] to Pyke [16], becomes more intuitive when viewed from a martingale perspective. It shows that $\sigma(x)$ is finite for all x if and only if $\sigma(0) = \text{var}(X)$ is finite.

The exponential distribution has the simplest form of the MRL function, $e(x) = b$, a constant. When the MRL is linear in x, i.e. $e(x) = ax + b$, (2) gives

$$S(x) = \left(\frac{b}{ax+b}\right)_+^{(1/a)+1}, \tag{4}$$

where the subscript + means that we take the positive part of the expression in parentheses. This function is a valid survivor function provided $b > 0$ and $a > -1$. We refer to (4) as the Hall-Wellner family. For $a > 0$, $a = 0$ and $-1 < a < 0$ we obtain respectively a Pareto distribution, an exponential distribution and a beta distribution $Beta(1, -1/a - 1)$, rescaled to have support $[0, -b/a]$. The natural estimator of the mean residual life function $e(x)$ at age x, based on a random sample $X_1, \ldots, X_n$ of size n from the survivor function $S(x)$, is the sample mean of the residual lifetimes of the observations that exceed x, that is

$$\hat{e}(x) = \frac{\sum(X_i - x)I(X_i - x)}{\sum I(X_i - x)},$$

where the summations run from $i = 1$ to n, and $I(y) = 1$ or $I(y) = 0$ according as $y > 0$ or $y \leq 0$, so that the denominator $\sum I(X_i - x) = N(x)$ is the total number of observations that exceed x. If $N(x) = 0$ then we arbitrarily set $\hat{e}(x) = 0$. Properties of this estimator were studied by Yang [17] and by Csörgő, Csörgő and Horváth [5]. These authors showed that the process

$$Z_n(x) = \sqrt{n}\{\hat{e}(x) - e(x)\}$$

converges in law to a Gaussian process $Z(x)$ with zero mean and covariance function

$$C\{Z(x), Z(y)\} = \frac{\sigma(x \vee y)}{S(x \wedge y)}. \tag{5}$$

Here $x \vee y$ and $x \wedge y$ denote $\max(x, y)$ and $\min(x, y)$ respectively, and the convergence is over any finite interval $[0, L]$ contained in the support of $S(x)$.

Since the form of the covariance function $C(x, y)$ is important in our work we give a brief heuristic derivation here. Suppose that $0 < x < y < T$, where T lies in the support of $S(x)$, and consider $\text{cov}\{\hat{e}(x), \hat{e}(y)\}$. We condition on the vectors $\{\mathbf{I}(x), \mathbf{I}(y)\}$, where $\mathbf{I}(z) = \{I(X_i - z)$, for $z = x, y$ and $i = 1, \ldots, n\}$. The possibility that $N(x) = 0$ may be ignored, since the probability of this event converges to zero geometrically fast in n. The iterated expectation formula gives

$$\begin{aligned}\text{cov}\{\hat{e}(x), \hat{e}(y)\} = {} & E[\text{cov}\{\hat{e}(x), \hat{e}(y)|\}\mathbf{I}(x), \mathbf{I}(y)] \\ & + \text{cov}[E\{\hat{e}(x)|\mathbf{I}(x), \mathbf{I}(y)\}, E\{\hat{e}(y)|\mathbf{I}(x), \mathbf{I}(y)\}].\end{aligned}$$

The second term converges to zero geometrically fast in n, since $E\{\hat{e}(y)|\mathbf{I}(y)\} = e(y)$ unless $N(y) = 0$. The first term times n is

$$nE\left\{\frac{1}{N(x)}\frac{1}{N(y)}\sum_{i:X_i>y}\operatorname{cov}(X_i - x, X_i - y|X_i > y)\right\}$$

$$= \quad nE\left\{\frac{1}{N(x)}\frac{1}{N(y)}N(y)\operatorname{cov}(X - x, X - y|X > y)\right\}$$

$$= \quad nE\left\{\frac{1}{N(x)}\operatorname{var}(X - y|X > y)\right\} \to \frac{\sigma(y)}{S(x)} \quad \text{as } n \to \infty,$$

as required.

Oakes and Dasu [15] used (2) to show that the PMRL model (1) implies the relationship

$$S_2(x) = S(x;\theta) = S_1(x)\left\{\int_x^\infty \frac{S_1(u)du}{e_1(0)}\right\}^{1/\theta - 1} \tag{6}$$

between the corresponding survivor functions. If $S_1(x)$ is a survivor function and $0 < \theta \leq 1$ then $S(x;\theta)$ as defined by (6) is always a survivor function, but if $\theta > 1$ this is not always true. To see this, consider the corresponding density $f(x;\theta) = -S'(x;\theta)$, which may be written in the form

$$f(x;\theta) = -\frac{\mu_1}{e_1^2(x)}\exp\left\{-\int_0^x \frac{du}{\theta e_1(u)}\right\}\left\{e_1'(x) + \frac{1}{\theta}\right\}, \tag{7}$$

where $\mu_1 = e_1(0)$. If $e_1'(x) \geq 0$ for all x this will be non-negative for all $\theta > 0$, but if $1/\theta_0 = \max[0, -\inf\{e_1(x)\}] > 0$, then we must have $\theta \leq \theta_0$ for (4) to be a probability density function. In the sequel we shall assume that θ is an interior point of the parameter space, which implies that $e'(x,\theta) + 1 > c$ for all x and some $c > 0$.

We define the PMRL family generated by a baseline survivor function $S_1(x)$ to be the parametric family $\{S(x;\theta), 0 < \theta < \theta_0\}$, where $S(x;\theta)$ is given by (6) and we set $\theta_0 = \infty$ if $e_1'(x) > 0$ for all x. The simplest PMRL family is the exponential, with $S_1(x) = \exp(-\rho x)$, $(\rho > 0)$ and $S(x;\theta) = \exp(-\rho x/\theta)$, $(\theta > 0)$. The Pareto baseline survivor function $S_1(x) = (1 + \rho x)^{-\alpha}$ $(\alpha > 0, \rho > 0)$, a reparameterization of (4), also defines a PMRL family, with

$$S(x;\theta) = \left(\frac{1}{1 + \rho x}\right)^\gamma, \tag{8}$$

where $\gamma = 1 + \frac{\alpha - 1}{\theta}$.

One further example of a PMRL family is worth noting. Suppose that $e(x;\theta) = \theta\exp(-\alpha x)$, where $0 < \alpha < 1$ and $\theta < 1$. The inversion formula gives

$$S(x;\theta) = \exp\left\{\alpha x - \frac{(e^{\alpha x} - 1)}{\alpha\theta}\right\}, \tag{9}$$

a form of the Gompertz distribution.

3. Maximum Likelihood Inference - Fisher Information In this section we consider maximum likelihood inference for the parameter θ of the proportional mean residual life model generated by an arbitrary baseline survivor function $S_1(x)$. Equation (7) gives

$$\log f(x;\theta) = \log \mu_1 - 2\log e_1(x) - \frac{1}{\theta}\int_0^x \frac{du}{e_1(u)} + \log\left\{\frac{1}{\theta} + e_1'(x)\right\}.$$

The score function (derivative of the log-likelihood) is

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = \frac{1}{\theta^2}\int_0^x \frac{du}{e_1(u)} - \frac{1}{\theta + \theta^2 e_1'(x)}.$$

The second derivative of the log-likelihood may be written in terms of $e(x;\theta) = \theta e_1(x)$ and its derivative $e'(x;\theta)$ in x as

$$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} = -\frac{2}{\theta^2}\int_0^x \frac{du}{e(u;\theta)} + \frac{1}{\theta^2}\left[1 - \frac{e'(x;\theta)^2}{\{1+e'(x;\theta)\}^2}\right].$$

The Fisher information $I(\theta)$ per observation is

$$E\left\{-\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}\right\} = \frac{1}{\theta^2}\left[2\int_0^\infty \int_0^x \frac{du}{e(u;\theta)} f(x;\theta)dx - 1 + \int_0^\infty \frac{e'(x;\theta)^2}{\{1+e'(x;\theta)\}^2} f(x;\theta)dx\right].$$

With the aid of the identity $e' = -1 + ef/S$ the last term may be written as

$$1 + \int \frac{S(x;\theta)^2}{e(x;\theta)^2 f(x;\theta)}dx - 2\int \frac{S(x;\theta)}{e(x;\theta)}dx.$$

Since θ is an interior point of the parameter space we have $e' + 1 > c > 0$ so that $S/(ef) < c^{-1}$, so that both integrals above are finite. After a little further algebra and use of Fubini's theorem, we obtain the expression

$$I(\theta) = \frac{1}{\theta^2}\int \frac{S(x;\theta)^2}{e(x;\theta)^2 f(x;\theta)}dx. \tag{10}$$

When the baseline survivor function $S_1(x)$ is known, standard large sample theory shows that the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta)$, where $\hat{\theta}$ is the maximum likelihood estimator of θ, is normal with mean zero and variance $I(\theta)^{-1}$.

For the exponential distribution with $S(x;\theta) = \exp(-x/\theta)$ we find that $I(\theta) = \theta^{-2}$ as expected. For the Pareto distribution (8) (with α known) we find that

$$I(\theta) = \left(\frac{\alpha - 1}{\alpha - 1 + \theta}\right)^2 = \left(\frac{\gamma - 1}{\gamma}\right)^2.$$

For the Gompertz family (9) we obtain, after some manipulation,

$$I(\theta) = 1 + 2\theta + \theta e^{\phi} E_1(\phi),$$

where $\phi = 1/\theta - 1$ and $E_1(z) = \int_z^\infty (e^{-u}/u)du$ is the exponential integral [1]

Formula (10) yields an interesting though unsurprising inequality for $I(\theta)$. The Cauchy-Schwartz inequality gives

$$\left\{\int S(x;\theta)dx\right\}^2 \le \left\{\int \frac{S^2(x;\theta)dx}{e^2(x;\theta)f(x;\theta)}\right\}\left\{\int e^2(x;\theta)f(x;\theta)dx\right\}.$$

Using Pyke's formula for the variance, (3), we obtain

$$I(\theta) \ge \frac{\{E_\theta(X)\}^2}{\mathrm{var}_\theta(X)}. \tag{11}$$

This inequality states that the asymptotic variance of the parametric estimator is always less than or equal to $(1/n)\times$ the coefficient of variation, which equals the variance of the simple ratio estimator $\tilde{\theta} = \bar{X}/\mu_1$. Equality in (11) is obtained if and only if

$$\frac{S(x;\theta)}{e(x;\theta)f(x;\theta)} \propto e(x;\theta),$$

where the constant of proportionality may depend on θ but not on x. Rearrangement and integration yield

$$\frac{1}{\int_x^\infty S(u;\theta)du} = \frac{A}{S(x;\theta)} + B,$$

for some constants $A(\theta)$ and $B(\theta)$. This can be integrated again to give the the quantile function $S^{-1}(1-\cdot)$, but with some awkward constants of integration. It seems likely that there exist families other than the exponential for which $\tilde{\theta}$ is asymptotically efficient at a particular value of θ, but there can be no such family for which $\tilde{\theta}$ is asymptotically efficient for all θ.

4. Ratio Estimators - Single Sample Problem We have seen that the ratio $\hat{\theta}_1 = \bar{X}/\mu_1$ of the sample mean to the population baseline mean gives a consistent estimator of θ that is asymptotically normally distributed with

$$\sqrt{n}(\hat{\theta}_1 - \theta) \sim N(0, (\sigma/\mu_1)^2).$$

For any prespecified point x i.e. in the support of X, the ratio $\hat{e}(x)/e_1(x)$ also gives a consistent asymptotically normally distributed estimator of θ. For any integer k and vectors $\mathbf{x} = \{x_1, \dots x_k\}^{\mathrm{T}}$ and $\mathbf{w} = \{w_1, \dots w_k\}^{\mathrm{T}}$ we may consider the ratio of linear combinations

$$\tilde{\theta}_k(\mathbf{x}, \mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\hat{\mathbf{e}}(\mathbf{x})}{\mathbf{w}^{\mathrm{T}}\mathbf{e}_1(\mathbf{x})},$$

where $\hat{\mathbf{e}}(\mathbf{x}) = \{\hat{e}(x_1), \dots, \hat{e}(x_k)\}^{\mathrm{T}}$ and $\mathbf{e}_1(\mathbf{x}) = \{e_1(x_1), \dots, \hat{e}_1(x_k)\}^{\mathrm{T}}$. It is understood that some of the coefficients w_i may be negative. The variance of $\tilde{\theta}(\mathbf{x}, \mathbf{w})$ can be simply calculated in terms of the matrix $\mathbf{\Sigma}$ with entries

$$\sigma_{ij} = \frac{\sigma(x_i \vee x_j)}{S(x_i \wedge x_j)}.$$

In fact

$$\text{var}\{\theta(\mathbf{x},\mathbf{w})\} = \frac{\mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w}}{(\mathbf{w}^{\mathrm{T}}\mathbf{e}_1)^2}.$$

It follows by simple matrix algebra that, for given x, the variance of $\tilde{\theta}(\mathbf{x},\mathbf{w})$ is minimized for

$$\mathbf{w}^* = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{e}_1}{\mathbf{e}_1^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{e}_1},$$

and that the achievable minimum variance is

$$V^*(\mathbf{x}) = \frac{1}{\mathbf{e}_1^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{e}_1}.$$

Although the optimal vector $\mathbf{w}^*$ depends on the unknown parameter θ, a standard argument shows that replacing θ by a preliminary $\sqrt{n}$-consistent estimate, for example the ratio $\bar{X}/\mu_1$, leaves the asymptotic distribution of the estimators unaffected. Of course in the single sample problem considered here the vector $e_1(\mathbf{x})$ is known, since it is a function of the known baseline survivor function $S_1(x)$.

The special structure of the matrix $\boldsymbol{\Sigma}$ allows it to be inverted explicitly. For $\boldsymbol{\Sigma} = \mathbf{DBD}$, where $\mathbf{D}$ is the diagonal matrix with elements $\sigma_i = \sigma(x_i)$, $(i = 1, ..., k)$ and $\mathbf{B}$ has elements $b_{ij} = b_l$ where $l = min(i,j)$ and $b_l = 1/(\sigma S)_l$. Here and in the sequel we write $(S)_i = S_1(x_i)$ and $(\sigma S)_i = \sigma(x_i)S_1(x_i)$ for notational economy. The inverse of $\mathbf{B}$ is a tridiagonal matrix with diagonal elements $b_{i,i}^{(-1)} = (b_i - b_{i-1})^{-1} + (b_{i+1} - b_i)^{-1}$, $(i = 1, \ldots, k-1)$, $b_{k,k}^{(-1)} = (b_k - b_{k-1})^{-1}$ and off-diagonal elements $b_{i,i+1}^{(-1)} = b_{i+1,i}^{(-1)} = -(b_{i+1} - b_i)^{-1}$, $(i = 1, \ldots, k-1)$. After a little further algebra we find that $\boldsymbol{\Sigma}^{-1}$ has elements

$$\Sigma_{1,1}^{(-1)} = \frac{-(S)_1^2}{(\sigma S)_2 - (\sigma S)_1},$$

$$\Sigma_{i,i}^{(-1)} = -(S)_i^2\left\{\frac{1}{(\sigma S)_i - (\sigma S)_{i-1}} + \frac{1}{(\sigma S)_{i+1} - (\sigma S)_i}\right\}, \quad (i = 2, \ldots, k-1),$$

$$\Sigma_{kk}^{(-1)} = -\frac{S_k}{\sigma_k}\frac{(\sigma S)_{k-1}}{(\sigma S)_k - (\sigma S)_{k-1}},$$

$$\Sigma_{i,i+1}^{(-1)} = \Sigma_{i+1,i}^{(-1)} = \frac{(S)_{i+1}(S)_i}{(\sigma S)_{i+1} - (\sigma S)_i}, \quad (i = 1, \ldots, k-1).$$

$$\Sigma_{i,j}^{(-1)} = 0, \quad (|i-j| > 1).$$

The reciprocal of the optimal variance is

$$V^*(\mathbf{x}) = \mathbf{e}_1^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{e}_1 = -\sum_{i=1}^{k-1}\frac{\{(eS)_{i+1} - (eS)_i\}^2}{\{(\sigma S)_{i+1} - (\sigma S)_i\}} + \frac{(Se)_k^2}{(S\sigma)_k}.$$

Once the optimal vector w^* has been chosen for a given x it is theoretically possible to perform a second optimization, over x for the given k, to find the optimal choice of support points $x_1, \ldots, x_k$. Dasu [6] gives some explicit evaluations of x_2 for the case $k = 2$ with $x_1 = 0$ in the Hall-Wellner family. The coefficient w_2^* turns out to be positive for the rescaled beta distribution but negative for the Pareto. For the latter distribution the contribution of the second support point is to downweight the influence of the extreme observations on the estimated ratio of means.

Finally we consider the limiting case as the set of points x_i become dense in the support of $S_1(x)$. It is easily seen that $\mathbf{w}^*$ has as its limiting form a delta function spike at $x = 0$ of magnitude

$$dW^*(0) = \frac{1}{\mu^2 f(0)},$$

and for $x > 0$ a continuous component

$$dW^*(x) = S(x)\frac{d}{dx}\left\{\frac{S(x)}{f(x)e(x)^2}\right\}dx.$$

Here and in the remainder of this section we drop the argument θ and write $\mu = e(0)$.

We can show directly from (5) that the variance of the estimator

$$\tilde{\theta}^* = \frac{\int \hat{e}(x)dW^*(x)}{\int e_1(x)dW^*(x)}$$

is equal to the reciprocal of the Fisher information $I(\theta)$ given in (10), showing that, in the single sample case, the ratio estimator achieves full asymptotic efficiency with appropriate choice of $W(t)$. We do require the additional assumption that $\sigma(x)$ is finite as well as $e(x)$.

We first show that

$$\int e(x)dW^*(x) = \theta^2 I(\theta).$$

We have

$$\begin{aligned}
\int e(x)dW^*(x) &= e(0)dW^*(0) + \int_0^\infty e(x)dW^*(x) \\
&= \frac{\mu}{\mu^2 f(0)} + \int_0^\infty e(x)S(x)\frac{d}{dx}\left\{\frac{S(x)}{f(x)e(x)^2}\right\}dx \\
&= \frac{1}{\mu f(0)} + \int_{x=0}^\infty \int_{y=x}^\infty S(y)\frac{d}{dx}\left\{\frac{S(x)}{f(x)e(x)^2}\right\}dxdy \\
&= \frac{1}{\mu f(0)} + \int_{y=0}^\infty S(y)\int_{x=0}^y \frac{d}{dx}\left\{\frac{S(x)}{f(x)e(x)^2}\right\}dxdy \\
&= \frac{1}{\mu f(0)} + \int_{y=0}^\infty S(y)\left\{\frac{S(y)}{f(y)e(y)^2} - \frac{1}{f(0)\mu^2}\right\}dy \\
&= \int_{y=0}^\infty \frac{S(y)^2}{f(y)e(y)^2}dy = \theta^2 I(\theta).
\end{aligned}$$

We now examine $\text{var}\{\int \hat{e}(x)dW^*(x)\}$. This requires a little care, because of the atom $dW^*(0)$. We have

$$
\begin{aligned}
&n\left[\text{var}\left\{\int_0^\infty \hat{e}(x)dW^*(x)\right\} + dW^*(0)^2\text{var}\{\hat{e}(0)\}\right] \\
&= 2n\int_{x=0}^\infty \int_{y=x}^\infty \text{cov}\{\hat{e}(x),\hat{e}(y)\}dW^*(x)dW^*(y) \\
&\to 2\int_{x=0}^\infty \int_{y=x}^\infty \frac{\sigma(y)}{S(x)}dW^*(x)dW^*(y) \\
&= 2\int_{y=0}^\infty \sigma(y)dW^*(y)\left[\frac{1}{\mu^2 f(0)} + \int_{x=0}^y \frac{d}{dx}\left\{\frac{S(x)}{f(x)e(x)^2}\right\}dx\right] \\
&= 2\int_{y=0}^\infty \sigma(y)\frac{S(y)}{f(y)e(y)^2}dW^*(y).
\end{aligned}
$$

Since $\sigma(y)S(y) = \int_y^\infty e(z)^2 f(z)dz$ by (3) this expression equals

$$
\begin{aligned}
&2\int_{y=0}^\infty \int_{z=y}^\infty e(z)^2\, f(z)\frac{dW^*(y)}{f(y)e(y)^2}dz \\
&= \int_{z=0}^\infty e(z)^2 f(z)\left[\frac{1}{\mu^4 f(0)^2} + \int_{y=0}^z \frac{S(y)}{f(y)e(y)^2}\frac{d}{dy}\left\{\frac{S(y)}{f(y)e(y)^2}\right\}dy\right]dz \\
&= \int_{z=0}^\infty e(z)^2 f(z)\left[\left\{\frac{S(z)}{f(z)e(z)^2}\right\}^2 - \frac{1}{\mu^4 f(0)^2}\right] \\
&= \int_{z=0}^\infty \frac{S(z)^2}{f(z)e(z)^2}dz - \frac{\text{var}\{\hat{e}(0)\}}{\mu^4 f(0)^2}.
\end{aligned}
$$

The second terms cancel so that

$$n\text{var}\left\{\int_0^\infty \hat{e}(x)dW^*(x)\right\} \to \theta^2 I(\theta).$$

Putting these results together and using $e(x) = \theta e_1(x)$ we obtain

$$n\text{var}\left\{\frac{\int \hat{e}(x)dW^*(x)}{\int e_1(x)dW^*(x)}\right\} \to \{I(\theta)\}^{-1},$$

as claimed.

5. Ratio Estimators - Two-Sample Problem In the previous work we have assumed that the baseline survivor function $S_1(x)$ was known. We now relax this assumption and consider the case of two-samples $X_1 = (X_{11},\ldots,X_{1m})$, $X_2 = (X_{21},\ldots,X_{2n})$ from $S_1(x)$ and $S_2(x;\theta)$ respectively. The first subscript denotes the sample, the second (if present) the observation.

Likelihood estimation for the two-sample problem would involve estimation of the unknown function $S_1(x)$, and its derivatives, which is beyond the scope of this paper. However, as in the one sample case, the simple ratio of means, now

$$\tilde{\theta} = \bar{X}_2/\bar{X}_1,$$

in an obvious notation, will still be consistent and asymptotically normally distributed for θ as will the ratio of linear combinations

$$\hat{\theta}(\mathbf{x}, \mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\hat{\mathbf{e}}_2(\mathbf{x})}{\mathbf{w}^{\mathrm{T}}\hat{\mathbf{e}}_1(\mathbf{x})},$$

where the notation follows that of the previous section. For given k, and support points $0 = x_1, \ldots, x_k$ the optimal weights and the variance of the resulting estimate can be estimated empirically from the sample.

The limiting variance of the estimator is given by

$$n\mathrm{var}\{\tilde{\theta}(\mathbf{x}, \mathbf{w})\} \to \frac{\mathbf{w}^{\mathrm{T}}(\mathbf{\Sigma}_2 + \theta^2\mathbf{\Sigma}_1)\mathbf{w}}{(\mathbf{w}^{\mathrm{T}}\mathbf{e_1})^2}.$$

Choice of an optimal vector of coefficients now requires inversion of the matrix $\mathbf{\Sigma}_2 + \theta^2\mathbf{\Sigma}_1$. There is no immediately obvious explicit formula for this inverse except under the "null hypothesis" $\theta = 1$ when also $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1$, when the optimal weights are the same as those derived in the previous section. The formulas of Section 4 extend easily when $\Sigma_2 \propto \Sigma_1$ also. However this can be true for all θ only in the Hall-Wellner family for which a simple parametric maximum likelihood estimator of θ is available.

In the general case, since both $\mathbf{\Sigma}_2^{-1}$ and $\mathbf{\Sigma}_1^{-1}$ are band-diagonal it follows that $\mathbf{\Sigma}_2^{-1} + \theta^{-2}\mathbf{\Sigma}_1^{-1}$ is also band-diagonal. Graybill [7, Section 8.3] gives a necessary and sufficient condition for the inverse of a matrix to be a band-diagonal matrix. It must be of expressible in the form

$$\begin{pmatrix} b_1 & b_2 & b_3 & b_4 & \ldots & b_k \\ b_2 & a_2b_2 & a_2b_3 & a_2b_4 & \ldots & a_2b_k \\ b_3 & a_2b_3 & a_3b_3 & a_3b_4 & \ldots & a_3b_k \\ b_4 & a_2b_4 & a_3b_4 & a_4b_4 & \ldots & a_4b_k \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_k & a_2b_k & a_3b_k & a_4b_k & \ldots & a_kb_k \end{pmatrix}.$$

This provides a recipe for computing

$$(\mathbf{\Sigma}_2^{-1} + \theta^{-2}\mathbf{\Sigma}_1^{-1})^{-1},$$

although not an explicit formula.

Finally Hall [9] gave the general matrix result

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1}(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{B}^{-1},$$

which shows that the inverse of the matrix $(\mathbf{A} + \mathbf{B})$ can be expressed in terms of $\mathbf{A}^{-1}$, $\mathbf{B}^{-1}$ and $(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}$. These formulas provide a computational recipe for inverting $\mathbf{\Sigma}_2 + \theta^2\mathbf{\Sigma}_1$ and hence for calculating the optimal vector of coefficients $\mathbf{w}(\mathbf{x})$. However the results are not transparent and the limiting behavior as the number of support points become dense is not clear.

6. Discussion The PMRL provides a simple example of a semiparametric estimation problem which is not rank invariant, in contrast to the proportional hazards model and the proportional odds model which have received much recent attention.

There remain many unanswered questions with regard to this model. First, it would be of interest to derive an explicit or implicit formula for the optimal function $w^*(x)$ in the two-sample case. Secondly, we may conjecture that use of this optimal weight function would lead to an asymptotically efficient estimator. In proving this an explicit formula for the semiparametric Fisher information in the two-sample case would be useful. Finally, since the optimal weight function depends on the unknown survivor function $S_1(x)$ and its derivatives as well as the unknown parameter θ, construction of a feasible efficient estimator is likely to remain a hard problem, even if the other questions can be answered satisfactorily.

Acknowledgments This work was supported in part by grant R01 CA52572 from the National Cancer Institute.

REFERENCES

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions.* Dover, New York, 1965.

[2] M. Asadi. Some general characterizations of the bivariate gumbel distribution and bivariate lomax distribution based on truncated expectations. *Journal of Multivariate Analysis*, 67:190–202, 1998.

[3] D.R. Cox. *Renewal Theory.* Methuen, London, 1962.

[4] D.R. Cox. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B*, 34:187–220, 1972.

[5] M. Csörgő, S. Csörgő, , and L. Horváth. *An Asymptotic Theory for Empirical Reliability and Concentration Processes.* Springer, New York, 1986.

[6] T. Dasu. *The Proportional Mean Residual Life Model.* PhD thesis, University of Rochester, 1990.

[7] F.A. Graybill. *Introduction to Matrices with Applications in Statistics.* Wadsworth, Belmont, 1969.

[8] E.I. Gumbel. Eine Darstellung der Sterbetafel. *Biometrika*, 16:283–96, 1924.

[9] W.J. Hall. Efficiency of weighted averages. Technical report, University of Rochester Department of Biostatistics, 1992.

[10] W.J. Hall and J.A. Wellner. Estimation of mean residual life. 1979.

[11] W.J. Hall and J.A. Wellner. Mean residual life. In M. Csörgő, D.A. Dawson, J.N.K. Rao, and A.K.Md.E. Saleh, editors, *Statistics and Related Topics*, pages 169–184, New York, 1981. North Holland.

[12] C.S. Ma. Multivariate survival functions characterized by constant product of mean remaining lives and hazard rates. *Metrika*, 44:71–83, 1996.

[13] C.S. Ma. Characteristic properties of multivariate survival functions in terms of residual life distributions. *Metrika*, 47:227–40, 1998.

[14] G. Maguluri and C.H. Zhang. Estimation in the mean residual life regression-model. *Journal of the Royal Statistical Society Series B*, 56:477–89, 1994.

[15] D. Oakes and T. Dasu. A note on residual life. *Biometrika*, 77:409–10, 1990.

[16] R. Pyke. Spacings (with discussion). *Journal of the Royal Statistical Society Series B*, 27:395–449, 1965.

[17] G. Yang. Estimation of a biometric function. *Annals of Statistics*, 6:112–16, 1978.

Deparment of Biostatistics
School of Medicine and Dentistry
Box 630
Rochester, NY 14642 USA
oakes@bst.rochester.edu

Rm E247
AT&T Shannon Lab
180 Park Avenue
Florham Park, NJ 07932 USA
tamr@research.att.com