Rank estimates in a class of semiparametric two-sample models.*

By

Dorota M. Dabrowska, Kjell A. Doksum and Ryozo Miura

University of California, Los Angeles, University of California Berkeley
and Osaka City University

Technical Report No. 83
November 1986
(revised March 1988)

*Research partially supported by National Science Foundation Grant DMS-85-41787 and National Institute of General Medical Sciences Grant SS-Y1RO1GM35416-01.

Department of Statistics
University of California
Berkeley, California

Rank estimates in a class of semiparametric two-sample models.*

Dorota M. Dabrowska, Kjell A. Doksum and Ryozo Miura

University of California, Los Angeles, University of California, Berkeley and Osaka City University

Summary

We consider a two-sample semiparametric model involving a real parameter θ and a nuisance parameter F which is a distribution function. This model includes the proportional hazard, proportional odds, linear transformation and Harrington-Fleming (1982) models. We propose two types of estimates based on ranks. The first is a rank approximation to Huber's (1981) M-estimates and the second is a Hodges-Lehmann (1963) type rank inversion estimate. We obtain asymptotic normality and efficiency results. The estimates are consistent and asymptotically normal generally but fully efficient only for special cases.

Key words: semiparametric transformation models, M-estimates based on ranks, Hodges-Lehmann estimates.

* Research partially supported by National Science Foundation Grant DMS-85-41787 and National Institute of General Medical Sciences Grant SSS-Y1RO1GM35416-01.

1. Introduction. We consider the two sample problem where $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ are independent random samples from populations with continuous distribution functions F and G, respectively. Many of the models in which rank (partial, marginal) likelihood methods are useful can be put in the form

$$(1.1) \qquad F(t) = D(H(t), \theta_1); \; G(t) = D(H(t), \theta_2)$$

where H(t) is an unknown continuous distribution function, $D(u,\theta)$ is a known continuous distribution function on (0,1), and θ_1 and θ_2 are in some parameter set Θ.

For inference based on rank likelihood, the above model is equivalent to the model obtained by using the distributions of $U_i = F(X_i) = D(H(X_i), \theta_1)$ and $V_j = F(Y_j) = D(H(Y_j), \theta_1)$. These distributions are

$$\tilde{F}(u) = u, u \in (0,1); \; \tilde{G}(v) = D(D^{-1}(v,\theta_1), \theta_2), v \in (0,1)$$

In the case where $\{D(u,\theta) : \theta \in \Theta\}$ is a group under composition satisfying $D(u,1) = u$, $D(D^{-1}(u,\theta_1), \theta_2) = D(u,\theta)$, $\theta = \theta_2/\theta_1$, we can write

$$(1.2) \qquad \tilde{F}(u) = u, u \in (0,1); \; \tilde{G}(v) = D(v, \theta), v \in (0,1)$$

From this point on we assume that (1.2) is satisfied. The distribution function F is treated as a nuisance parameter and we consider the problem of estimating θ. This model goes back to Lehmann (1953), and includes the following models that have important applications in survival analysis, reliability, and other areas.

Example 1.1. *Proportional Hazard Model.* If F and G have proportional hazards, then $D(v,\theta) = 1 - [1 - v]^{\frac{1}{\theta}}$, $\theta > 0$. Lehmann (1953) and Savage (1956) considered testing in this model. Cox (1972, 1975) developed estimation procedures in a much more general regression problem with censored data.

Example 1.2. *Proportional Odds Model.* For any continuous distribution H the odds rate is defined by $r_H = H/(1 - H)$. If F and G have proportional odds rates, in the sense that $r_G(t) = \theta^{-1} r_F(t)$, then $D(v,\theta) = v\,[\,(1 - v)\theta + v\,]^{-1}$. This model has been considered by Ferguson (1967) and Bickel (1986) in the two-sample case and in more general regression models by Bennett (1983) and Pettitt (1984), among others.

Example 1.3. *Proportional γ-Odds Model.*

For any continuous distribution function H, the γ-odds rate is defined by

$$
\begin{aligned}
r_{H,\gamma}(t) &= \{\,[\,1 - H(t)\,]^{-\gamma} - 1\}/\gamma, \gamma > 0 \\
&= -\log[\,1 - H(t)\,], \qquad \gamma = 0
\end{aligned}
$$

If F and G have proportional γ-odds, then

$$
\begin{aligned}
D(v,\theta) &= 1 - \left\{\frac{\theta(1 - v)^{\gamma}}{1 - (1 - v)^{\gamma} + \theta(1 - v)^{\gamma}}\right\}^{\frac{1}{\gamma}}, \gamma > 0 \\
&= 1 - (1 - v)^{\frac{1}{\theta}}, \qquad \gamma = 0
\end{aligned}
$$

This model, which has been considered by Harrington and Flemming (1982), Clayton & Cuzick (1986), Bickel (1986), and Dabrowska and Doksum (1986), reduces to Example 1.1 when $\gamma = 0$, and to Example 1.2 when $\gamma = 1$.

Example 1.4. *Transformation Shift Model.* Let Q be any known continuous distribution function which is strictly increasing on the whole real line. If F and G satisfy the transformation shift model where for some increasing transformation h, X_i and Y_j can be written

$$
\begin{aligned}
h(X_i) &= \mu_1 + \varepsilon_i, i = 1,\dots,m \\
h(Y_j) &= \mu_2 + \varepsilon_j, j = 1, \cdots, n
\end{aligned}
\tag{1.3}
$$

with $\mu_2 - \mu_1 = \log\theta$, $N = m + n$ and $\varepsilon_1, \cdots, \varepsilon_N$ independent with distribution function

Q, then $D(v,\theta) = Q(Q^{-1}(v) - \log\theta)$. To see this, set $h(t) = Q^{-1}(F(t)) + \mu_1$. Then (1.3) is equivalent to (1.2) with $D(v,\theta) = Q(Q^{-1}(v) - \log\theta)$. This is an extension of the power transformation model where h(t) is of the form $\mathrm{sign}(t)|t|^{\lambda}$ or $[\mathrm{sign}(t)|t|^{\lambda} - 1]/\lambda$, and where Q is the standard normal distribution function. See Anscombe and Tukey (1952), Tukey (1957), Box and Cox (1964), Bickel and Doksum (1981) and Doksum (1987). This transformation shift model reduces to the proportional odds model if we take Q to be the logistic distribution function $L(x) = 1/[1 + e^{-x}]$.

Theory and methods for dealing with semiparametric models and partial likelihood have been developed by Begun (1981), Begun, Hall, Huang and Wellner (1983), Begun and Wellner (1983), Wellner (1986) and Wong (1986) among others. However, these methods do not lead to tractable efficient scores or tractable efficient estimates for any of the above models except the proportional hazard model. For arbitrary θ_0 and under certain regularity conditions, Bickel (1986) obtained the asymptotically optimal rank test for testing $H_0: \theta = \theta_0$, vs $H_1: \theta > \theta_0$. His regularity conditions are satisfied by the γ-odds model with $\gamma \geq 1$. However, the optimal test statistic is a nonlinear rank statistic whose value can be obtained only after solving certain functional equations numerically on the computer. Attempts to extend these methods to obtain estimates that are asymptotically efficient in the semiparametric sense have not yet succeeded. Doksum (1987) and Dabrowska and Doksum (1988) propose a resampling scheme for computing maximum partial (rank) likelihood estimators in general semiparametric transformation regression models with censored data. These estimates perform well in Monte Carlo simulation studies, but the theoretical properties are difficult to establish. Clayton and Cuzick (1986) proposed estimates for the proportional γ-odds model that also apply to regression and censored data, but the theoretical properties of these estimates are also not well understood.

Methods based on local (near $\theta = 1$) approximations to the likelihood have been developed by Pettitt (1984), Doksum (1987), and Dabrowska and Doksum (1986). These estimates are asymptotically normal for θ in neighborhoods of $\theta = 1$, however they are not consistent for fixed $\theta \neq 1$.

In this paper we consider the uncensored case and introduce two classes of estimates. The advantage of these estimates is that they are relatively simple to implement in practice and that their properties can readily be derived and understood. Moreover they are based on intuitive estimation equations that are immediate extensions of the familiar M and R estimate equations. Although the new estimates are not fully efficient for all values if θ, such fully efficient estimates are not available for $\gamma \neq 0$.

In Section 2 we introduce estimates of θ that can be regarded as rank approximations to Huber's (1981) M-estimates based on score function ψ. We show asymptotic normality of these RAM (rank approximate M) estimates. We compare these estimates with the asymptotically optimal estimates for F known, and find that for a certain range of parameter values, not much efficiency is lost. In fact, the score function ψ can be chosen so that the estimate is fully efficient at $\theta = 1$.

In Section 3 we introduce estimates of θ based on the Hodges-Lehmann (1963) rank inversion idea and obtain asymptotic normality of these estimates. For appropriate choices of the score functions, these estimates have the same asymptotic distribution as the RAM estimates.

2. M estimates based on ranks. In this section we introduce estimates that, in an approximate sense, are M estimates based on ranks. We start by assuming that the distribution F of the X's is known and introduce M estimates of θ that depend on F. Let

$$U_i = F(X_i),\ V_j = F(Y_j),\ i = 1, \cdots, m,\ j = 1, \cdots, n.$$

The joint distribution of $U_1, \cdots, U_m$, $V_1, \cdots, V_n$ is $\prod_{i=1}^{m} u_i \prod_{j=1}^{n} D(v_j, \theta)$, so that $(V_1, \cdots, V_n)$ is sufficient for θ. Let $\psi(v, \theta)$ be a function which is monotone decreasing in θ, and satisfies $E_{\theta_0}\psi(V, \theta_0) = 0$ where V is distributed according to $D(v, \theta_0)$ and θ_0 is the true parameter value. An M-estimate (see Huber (1981)) of θ is defined as a solution to the equation $\Sigma\psi(V_j, \theta) = 0$.

Let $N = m + n$. When F is unknown, we define $\hat{F}(u) = mF_m(u)/(m+1)$ for $u \in [X_{(1)}, X_{(m)}]$, $\hat{F}(u) = 1/(N+1)$ for $u < X_{(1)}$ and $\hat{F}(u) = N/(N+1)$ for $u > X_{(m)}$. Here $X_{(1)}$ and $X_{(m)}$ denote the first and the last order statistics of the X_i's and $F_m(u) = m^{-1}\#\{i : X_i \le u\}$.

We set $\hat{V}_j = \hat{F}(Y_j)$, and we let $\hat{\theta}$ be any "solution" to

$$\sum_{j=1}^{n} \psi(\hat{V}_j, \theta) = 0.$$

More precisely, $\hat{\theta}$ is any value in the interval $[\theta^*, \theta^{**}]$ where $\theta^* = \sup\{\theta : \Sigma\psi(\hat{V}_j, \theta) > 0\}$ and $\theta^{**} = \inf\{\theta : \Sigma\psi(\hat{V}_j, \theta) < 0\}$.

In terms of the rank likelihood, $\hat{V}_1, \cdots, \hat{V}_n$ are sufficient for θ. To see this, let $Y_{(1)} < \cdots < Y_{(n)}$ be the ordered statistics and let $K_{(j)} = m\hat{F}(Y_{(j)})$. Then it is easy to check that $K_{(1)}, \cdots, K_{(n)}$ are equivalent to the ranks. Thus we call $\hat{\theta}$ a *RAM (Rank*

Approximate M) estimate.

Here is an example where we get an explicit formulae for $\hat{\theta}$.

Example 2.1. Assuming F is known, the M-estimate based on $\psi(v,\theta) = -\theta - \log(1-v)$ is the MLE (Maximum Likelihood Estimate) for the proportional hazard model of Example 1.1. This estimate is $n^{-1}\Sigma - \log(1 - F(Y_j))$. The corresponding RAM estimate of θ which applies when F is unknown is
$\hat{\theta} = n^{-1}\sum_{j=1}^{n} -\log(1 - \hat{F}(Y_j))$

We return to the general case and show asymptotic normality of the RAM estimates. We assume throughout that the limits $\pi_0 = \lim_{N\to\infty}(m/N)$ and $\pi_1 = \lim_{N\to\infty}(n/N)$ exist and are strictly between 0 and 1. Further, we assume that ψ is continuously differentiable in u and we set $\psi'(u,\theta) = \frac{\partial}{\partial u}\psi(u,\theta)$ and $d(u,\theta) = \frac{\partial}{\partial u}D(u,\theta)$. Define

$$\lambda(\theta) = \int_0^1 \psi(u,\theta)d(u,\theta_0)du, \quad \lambda_1(\theta) = \int_0^1 \psi_1(u,\theta)du,$$

$$\psi_1(u,\theta) = \int_0^u \psi'(v,\theta)d(v,\theta_0)dv,$$

$$\sigma^2(\theta) = \pi_0^{-1}\{\int_0^1 \psi_1^2(u,\theta)du - \lambda_1^2(\theta)\} + \pi_1^{-1}\{\int_0^1 \psi^2(u,\theta)d(u,\theta_0)du - \lambda^2(\theta)\}.$$

Assume that for θ in a neighbourhood of the true parameter value θ_0, the following conditions hold:

A.1 $|\psi(\cdot,\theta)| = O(r^a)$ and $|\psi'(\cdot,\theta)| = O(r^{a+1})$

where $r(u) = [u(1-u)]^{-1}$ and $a = 1/2 - \tau$ for some $0 < \tau < 1/2$.

A.2 $\int r(u)^{1-\eta}d(u,\theta)du < \infty$ uniformly in θ for some $0 < \eta < \tau$

A.3 $m\int u^m d(u,\theta)du = O(1)$ and $m\int(1-u)^m d(u,\theta)du = O(1)$

uniformly in θ and m.

The assumption A.3 is satisfied whenever the density $d(u,\theta)$ is bounded uniformly in θ in a neighbourhood of θ_0. In particular it holds for the proportional hazard model with $\theta_0 \le 1$ and γ-rate models of Examples 1.2 and 1.3. Condition A.2 is satisfied in all three examples.

Theorem 2.1. Suppose that $\psi'(u,\theta)$ and $\lambda(\theta)$ are continuous in θ for θ in a neighborhood of the true parameter value θ_0 and suppose that θ_0 is the unique point with $\lambda(\theta_0) = 0$. Assume that $\sigma^2(\theta)$ is finite, nonzero and continuous in a neighborhood of θ_0 and that A.1, A.2 and A.3 hold. Then $\sqrt{N}\lambda(\hat{\theta})$ is asymptotically normally distributed with mean zero and variance $\sigma^2(\theta_0)$. If in addition, if $\lambda'(\theta_0)$ exists and $\lambda'(\theta_0) < 0$, then $\sqrt{N}(\hat{\theta} - \theta_0)$ is asymptotically normal with mean zero and variance $\sigma^2(\theta_0)/[\lambda'(\theta_0)]^2$.

Proof. First we note that asymptotically all $\tilde{\theta} \in [\theta^*, \theta^{**}]$ are equivalent so it is enough to consider θ^*.

By the assumed monotonicity of $\psi(u,\theta)$ in θ, the function $\lambda(\theta)$ is monotone decreasing. Let y be fixed. Since $\lambda(\theta_0) = 0$ and λ is continuous, for N sufficiently large, there is θ_N such that $y = -\sqrt{N}\lambda(\theta_N)$. In fact, $\theta_N = -\lambda^{-1}(-y/\sqrt{N})$. Let G_n denote the empirical distribution function of the Y's. Then

$$P(-\sqrt{N}\lambda(\theta^*) < y) \;=\; P(\theta^* < \theta_N) \;=$$

$$P\Big[\frac{\sqrt{N}[\int\psi(\hat{F},\theta_N)dG_n - \lambda(\theta_N)]}{\sigma(\theta_N)} \le \frac{y}{\sigma(\theta_N)}\Big] \to \Phi\left[\frac{y}{\sigma(\theta_0)}\right]$$

To see this we note that

$$(2.1)\quad \sqrt{N}\,[\,\int\psi(\hat{F},\theta_N)dG_n - \lambda(\theta_N)\,]$$
$$= \sqrt{N}\,[\,\int(F_m - F)\psi'(F,\theta_N)dG + \int\psi(F,\theta_N)d(G_n - G)\,] + r_N,$$

where r_N is a remainder term. We have

$$\int(F_m - F)\psi'(F,\theta_N)dG$$
$$= m^{-1}\sum_{i=1}^{m}\int[\,I(X_i < x) - F(x)\,]\psi'(F(x),\theta_N)d(F(x),\theta_0)dF(x) = \sum_{i=1}^{m}A_{im}$$

which is a sum of independent identically distributed (iid) random variables. By assumption A.1, $|A_{im}| = O(1)r^{1/2-\tau+\eta}(F(X_i))\int r^{1-\eta}(u)d(u,\theta_0)du$. By assumption A.2, in a neighbourhood of θ_0 the deterministic part of this upper bound is uniformly bounded from above. Further, for some $\tau_1 > 0$ the random part of this bound has an absolute moment of order $2 + \tau_1$, which is uniformly bounded above.

Further, $\int\psi(F,\theta_N)d(G_n - G)$ is a sum iid random variables. By assumptions A.1 and A.2, they have a finite absolute moment of order $2 + \tau_2$ for some $\tau_2 > 0$. Berry-Esseen's theorem completes the proof of the asymptotic normality of the first two terms. The remainder term r_N is considered in Section 4.

Note that when $\theta_0 = 1$, the asymptotic variance of $\hat{\theta}$ reduces to

$$(\pi_0^{-1} + \pi_1^{-1})\,E\psi^2(U,1)\,/\,[\,E\frac{\partial}{\partial\theta}\psi(U,\theta)\,|_{\theta=1}\,]^2$$

where U is uniform on (0,1). This is exactly the same as the asymptotic variance of the M estimate for the model (1.1) with H known. Thus if we choose

$$\psi(v,\theta) = \frac{\partial}{\partial\theta}\log d(v,\theta)$$

then $\hat{\theta}$, in addition to being consistent and asymptotically normal for general θ_0, is asymptotically efficient when $\theta_0 = 1$.

If we consider the RAM estimate $\hat{\theta} = n^{-1}\sum_{j=1}^{n}-\log(1 - \hat{F}(Y_j))$, we find that in the proportional hazard model the asymptotic variance is $\pi_0^{-1}\theta(2-\theta)^{-1}+\pi_1^{-1}\theta^2$ for $\theta \le 1$. We can compare this estimate with the Cox partial likelihood estimate. The usual parametrization in the Cox proportional hazard model is in terms of $\beta = \ln\theta$. If β^* is the Cox partial likelihood estimate then from Efron (1977) we find that its asymptotic variance is given by

$$[\int_0^1 \{\pi_1^{-1} + \pi_0^{-1}\theta u^{(\theta-1)/\theta}\}^{-1} du\,]^{-1}.$$

The asymptotic variance of $\hat{\beta}_0 = \ln\hat{\theta}$ where $\hat{\theta}$ is the RAM estimate is given by

$$\pi_0^{-1}\theta^{-1}(2-\theta)^{-1} + \pi_1^{-1}$$

for $\theta \le 1$. For $\pi_0 = \pi_1 = 1/2$ the asymptotic relative efficiency of $\hat{\beta}_0$ with respect to β^* is equal to 1, 0.951, 0.863, 0.757 and 0.647 for $\theta = 1, 1/2, 1/4, 1/8$ and 1/16.

Example 2.2. *Proportional γ-odds model.* Our main application is to the proportional γ rate model, $\gamma > 0$. The ψ function corresponding to the MLE for the F known case is

$$\psi(u,\theta) \;=\; \gamma^{-1} - (1+\gamma^{-1})\frac{(1-u)^{\gamma}\theta}{1-(1-u)^{\gamma}+\theta(1-u)^{\gamma}}. \tag{2.2}$$

Using Theorem 2.1, we find that the asymptotic variance of the RAM estimate simplifies when γ is of the form $\gamma = 1/k$, where $k \ge 1$ is an integer. In this case the asymptotic variance is

$$\frac{\theta^2}{\pi_1}\frac{2+k}{k} + \frac{\theta^2}{(1-\theta)^2}\frac{1}{\pi_0}\left\{\frac{(1+k)^2}{k}\theta^{k+2}\sum_{i=0}^{3+k}\binom{3+k}{i}\frac{1}{i+k}\left[\frac{1-\theta}{\theta}\right]^i - \left[\frac{k+\theta}{k}\right]^2\right\}$$

For $\gamma = k = 1$, the proportional odds model, the asymptotic variance is $3\theta^2\pi_1^{-1} + (0.2)\theta\pi_0^{-1}(4+7\theta+4\theta^2)$ while for $k = 2$, it equals $2\pi_1^{-1}\theta^2 + \theta\pi_0^{-1}\{18+23\theta+12\theta^2+3\theta^3\}/28$.

To judge the performance of the RAM estimate based on (2.1), we compute the efficiency of this estimate with respect to the MLE for the γ-odds model with F known and the X's and Y's distributed as $D(H(x), \theta_1)$ and $D(H(y), \theta_2)$, respectively, where D is given in Example 1.3, and $\theta = \theta_2/\theta_1$. For $m = n$, this efficiency is given by $e(\hat{\theta}_{RAM}, \hat{\theta}_{MLE}) = 60\theta(8 + 44\theta + 8\theta^2)^{-1}$ for $k = 1$ and $112\theta(18 + 79\theta + 12\theta^2 + 3\theta^3)^{-1}$ for $k = 2$.

Here is a brief table of these efficiencies.

Table 1 about here.

We see from Table 2.1 that when $\gamma = 1$ and 1/2, $\hat{\theta}$ is quite efficient for θ in the range (0.5, 2). The efficiency increases as γ increases. In fact it is easy to show that as $\gamma \to \infty$ the efficiency tends to one for all θ. The efficiency given is a lower bound in the efficiency of $\hat{\theta}$ with respect to the asymptotically optimal estimate based on the ranks.

3. Rank inversion estimates. In this section, we introduce rank-inversion estimates based on the ideas of Hodges-Lehmann (1963). Again, we start by assuming that F is known and let U_i, V_j and $D(u,\theta)$ be as in Section 2. In particular, we assume that $D(u, \theta)$ is monotone decreasing in θ. Note that $U_1, \cdots, U_m$, $D(V_1,\theta), \cdots, D(V_n,\theta)$ all have the same distribution when $\theta = \theta_0$, where θ_0 is the true value of the parameter. Let $R_i(\theta)$ denote the rank of U_i among $U_1, \cdots, U_m$, $D(V_1,\theta), \cdots, D(V_n,\theta)$, and let

$$T_N(\theta) = m^{-1}\sum_{i=1}^{m} J_N\left[\frac{R_i(\theta)}{N+1}\right]$$

denote a linear rank function with monotone increasing score function J_N. The F-known Hodges-Lehmann estimate of θ is obtained by solving $T_N(\theta) = \int_0^1 J(u)du$ for θ, where $J(u)$ is the limit of $J_N(u)$. Without loss of generality, we assume $\int_0^1 J(u)du = 0$.

Suppose now that F is unknown. Let $X_{(1)} < \cdots < X_{(m)}$ be the vector of order statistics of X_i's Let $\tilde{F}$ be defined by

$$\tilde{F}(u) = \frac{u + iX_{(i+1)} - (i+1)X_{(i)}}{(m+1)(X_{(i+1)} - X_{(i)})}$$

for $X_{(i)} \le u \le X_{(i+1)}$, $i = 1, \cdots, m-1$. Thus on the interval $[X_{(1)}, X_{(m)}]$, $\tilde{F}$ is a linearized version of the right-continuous distribution function $mF_m/(m+1)$, where $F_m(u) = m^{-1}\#\{i : X_i \le u\}$. Further, let $Y_{(1)}$ and $Y_{(n)}$ be the first and the last order statistics of the Y_j's. If $Y_{(1)} < X_{(1)}$ or $Y_{(n)} > X_{(m)}$, then we extend $\tilde{F}$ to the interval $[\min(X_{(1)}, Y_{(1)}), \max(X_{(m)}, Y_{(n)})]$ linearly with $\tilde{F}(Y_{(1)}) = 1/(N+1)$ if $Y_{(1)} < X_{(1)}$ and $\tilde{F}(Y_{(n)}) = N/(N+1)$ if $Y_{(n)} > X_{(m)}$.

Further, let $\tilde{R}_i(\theta)$ be the the rank of $\tilde{F}(X_i)$ among $\tilde{F}(X_1), \cdots, \tilde{F}(X_m)$, $D(\tilde{F}(Y_1), \theta), \cdots, D(\tilde{F}(Y_n), \theta)$. Let $\tilde{\theta}_R$ be any "solution" to

$$\tilde{T}_N(\theta) = m^{-1}\sum_{i=1}^{m} J_N\left(\frac{\tilde{R}_i(\theta)}{N+1}\right) = 0$$

More precisely, let $\tilde{\theta}_R$ be any point in $[\theta_R^*, \theta_R^{**}]$ where $\theta_R^* = \sup\{\theta : \tilde{T}_N(\theta) < 0\}$ and $\theta_R^{**} = \inf\{\theta : \tilde{T}_N(\theta) > 0\}$. Similar estimates have also been considered by Doksum and Nabeya (1984) and Miura (1985).

Example 3.1. Assuming F is known, the Hodges-Lehmann type rank estimate based on $J(u) = 2u - 1$ is asymptotically optimal for the proportional odds model. Let $L(x) = 1/[1 + e^{-x}]$ be the logistic distribution function and note that if we set

$W_i = L^{-1}(F(X_i))$ and $Z_j = L^{-1}(F(Y_j))$, then W_i and Z_j follow a logistic shift model ($W_i \sim L(w)$, $Z_j \sim L(z - \log\theta)$) with parameter $\log\theta$. Since the ranks are invariant under the increasing transformation L^{-1}, it follows that the Hodges-Lehmann estimate of θ is

$$\tilde{\theta}_{HL} = \exp\{\underset{i,j}{\text{median}}(Z_j - W_i)\}.$$

The corresponding $\tilde{\theta}_R$, which is appropriate when F is unknown, is

$$\tilde{\theta}_R = \exp\{\underset{i,j}{\text{median}}(L^{-1}(\tilde{F}(Y_j) - L^{-1}(\tilde{F}(X_i))\}.$$

We return to the general case and show the asymptotic normality of $\sqrt{N}(\tilde{\theta}_R - \theta_0)$. From (1.2), we have

$$(3.1) \qquad D(D^{-1}(u,\theta),\delta) = D(u,\delta/\theta), \quad D(u,1) = u.$$

Let

$$\dot{D}(u,\theta) = \frac{\partial}{\partial\theta}D(u,\theta) \quad \eta = \int_0^1 J'(u)\dot{D}(u,1)du$$

$$\tau^2(\theta) = \pi_1^{-1}\int_0^1 J^2(u)du + \pi_0^{-1}[\int_0^1 \alpha^2(u)du - (\int_0^1 \alpha(u)du)^2]$$

where α is defined by

$$\frac{d\alpha(u)}{du} = J'(D(u,\theta))[d(u,\theta)]^2$$

Further, let $G_n^\theta(u) = n^{-1}\#[j : D(\tilde{F}(Y_j),\theta) \le u] = G_n\tilde{F}^{-1}D(u,\theta^{-1})$, $H_N^\theta(u) = \{mF_m\tilde{F}^{-1}(u) + nG_n^\theta(u)\}/(N+1)$, $G^\theta(u) = D(D^{-1}(u,\theta),\theta_0) = D(u,\theta_0/\theta)$ (by (3.1)) and $H^\theta(u) = \pi_0 u + \pi_1 G^\theta(u)$. In terms of these functions, we have

$$\tilde{T}_N(\theta) = m^{-1}\sum_{i=1}^m J_N\left(\frac{\tilde{R}_i(\theta)}{N+1}\right) = \int J_N(H_N^\theta(\tilde{F}))dF_m = \int J_N(H_N^\theta)dF_m\tilde{F}^{-1}.$$

Assume

B.1 $r_{1N} = \sqrt{N}\int\{J_N(H_N^\theta) - J(H_N^\theta)\}dF_m\tilde{F}^{-1} \to_P 0$ as $N \to \infty$ uniformly for θ in a neighbourhood of θ_0.

Moreover,

B.2 J is a differentiable function with bounded continuous derivative J', and $0 < \int J^2(u)du < \infty$.

Finally, we assume that the limits $\pi_0 = \lim_{N\to\infty}(m/N)$ and $\pi_1 = \lim_{N\to\infty}(n/N)$ exist and are strictly between 0 and 1.

Theorem 3.1. If $D(u,\theta)$ is decreasing in θ, and if the preceeding conditions hold then $\sqrt{N}(\tilde{\theta}_R - \theta_0)$ has asymptotically a normal distribution with mean zero and variance $\theta_0^2\, \tau^2(\theta_0)/\eta^2$.

Proof. As in the case of Hodges-Lehmann (1963), θ_R^*, θ_R^{**} and any point between them, such as $\tilde{\theta}_R$, will have the same asymptotic distribution. Further, $P(\sqrt{N}\left[\theta_R^*/\theta_0 - 1\right] \le t) = P(\sqrt{N}\tilde{T}_N(\delta) > 0)$ where $\delta = \theta_0(1 + t/\sqrt{N})$. We have

$$\begin{aligned}\sqrt{N}\tilde{T}_N(\delta) &= \sqrt{N}\int J_N(H_N^\delta(u))dF_m\tilde{F}^{-1}(u)\\ &= \sqrt{N}\int J(H_N^\delta(u))d\,[\,F_m\tilde{F}^{-1}(u) - u\,] + \sqrt{N}\int[\,J(H_N^\delta(u) - J(H^\delta(u))\,]\,du\\ &\quad + \sqrt{N}\int[\,J(H^\delta(u) - J(u)\,]du + r_{1N} = I_1 + I_2 + I_3 + r_{1N}.\end{aligned}$$

Note that $I_1 = -\sqrt{N}\int[\,F_m\tilde{F}^{-1} - u\,]dJ(H_N^\delta(u))$. This term is bounded in absolute value by $\sup\sqrt{N}\,|\,F_m\tilde{F}^{-1}(u) - u\,|\int|\,J'(u)\,|\,du$, which tends to zero since $\sup|\,F_m\tilde{F}^{-1}(u) - u\,| \le 2/(m+1)$ and $\int|\,J'(u)\,|\,du < \infty$ by B.2.

The second term can be written as

$$I_2 = \sqrt{N}\int J'(H^\delta(u))\,[\,H_N^\delta(u) - H^\delta(u)\,]\,du + r_{2N}$$

where

$$r_{2N} = \sqrt{N}\int(H_N^\delta - H^\delta)\{J'(\tilde{H}^\delta) - J'(H^\delta)\}du$$

and $\tilde{H}^\delta$ assumes values between H_N^δ and H^δ. We have

$$\begin{aligned}(3.2)\qquad &\sqrt{N}\int J'(H^\delta(u))\,[\,H_N^\delta(u) - H^\delta(u)\,]\,du\\ &= \sqrt{N}\int J'(H^\delta(u))\,[\,m/(N+1)F_m\tilde{F}^{-1}(u) - \pi_0 u\,]\,du\\ &+ \sqrt{N}\int J'(H^\delta(u))(n/(N+1))\,[\,G_n\tilde{F}^{-1}D^{-1}(u,\delta) - G\tilde{F}^{-1}D^{-1}(u,\delta)\,]\,du\\ &+ \sqrt{N}\int J'(H^\delta(u))\,[\,n/(N+1)G\tilde{F}^{-1}D^{-1}(u,\delta) - \pi_1 G F^{-1}D^{-1}(u,\delta)\,]\,du\end{aligned}$$

The first term converges in probability to 0 by an argument similar to I_1. The next two terms of this expansion converge weakly to

$$\pi_1\{\pi_1^{-1/2}\int J'(u)\,B_2(u)du - \pi_0^{-1/2}\int J'(D(u,\theta_0))d^2(u,\theta_0)B_1(u)du\}$$

where B_1 and B_2 are independent Brownian bridges. This follows since $\sqrt{n}\,(G_n - G)$ and $\sqrt{m}\,(\tilde{F}^{-1;} - F^{-1})$ converge weakly to $B_2(G)$ and $-f\,(F^{-1})\,B_1(u)$, respectively and $\sup|\tilde{F}^{-1} - F^{-1}| = o_P(1)$. The standard Skorokhod construction yields the desired result. Further, $r_{2N} \to_P 0$ by B.2, $\sup|H_N^\delta - H^\delta| \to_P 0$ and $\sup\sqrt{N}(H_N^\delta - H^\delta) = O_P(1)$.

Finally,

$$I_3 = \sqrt{N}\int J'(u)\pi_1\{D(u,\theta_0/\delta) - u\}du + r_{4N} \to -\pi_1 t\int_0^1 J'(u)\dot{D}(u,1)du$$

and

$$r_{4N} = \sqrt{N}\int[\,J'(\tilde{H}^\delta) - J'(u)\,]\,\pi_1\,[\,D(u,\theta_0\,/\,\delta) - u\,]\,du$$

where $\tilde{H}^\delta$ assumes values between H^δ and $H \equiv u$. This term converges to 0 by assumption B.2 and Taylor expansion of $D(u,\theta_0\,/\,\delta)$.

If we let $J(u) = [\,1 + \gamma^{-1}\,]\,(1-u)^\gamma - \gamma^{-1}$, $\gamma \geq 1$, then in the proportional γ-odds model, $\hat{\theta}_R$ given in Example 3.1 will have the same asymptotic variance as the RAM estimate based on (2.2). See example 2.2.

4. Proof of Theorem 2.1: remainder terms. The remainder term r_N in (2.1) is given by

$$r_N = \sqrt{N}\int\{\psi(\hat{F},\theta_N) - \psi(F,\theta_N)\}dG_n - \sqrt{N}\int(F_m - F)\psi'(F,\theta_N)dG.$$

For small $\gamma \in (0,1)$ define $S_\gamma = [F^{-1}(\gamma), F^{-1}(1-\gamma)]$. Further let $\Delta_N = [Y_{(1)}, Y_{(n)}]$ and $E_N = [X_{(1)}, X_{(m)}]$ where $X_{(1)}$, $X_{(m)}$, $Y_{(1)}$ and $Y_{(n)}$ are the first and last order statistics among X_i's and Y_j's. Then, after some algebra, $r_N = \sum_{i=1}^{7} r_{iN}$ where

$$r_{1N} = \sqrt{N}\int_{\Delta_N\cap S_\gamma}\{\psi'(F^*,\theta_N) - \psi'(F,\theta_N)\}(\hat{F} - F)dG_n$$

$$r_{2N} = \sqrt{N}\int_{\Delta_N\cap S_\gamma}\psi'(F,\theta_N)(\hat{F} - F_m)\,dG_n$$

$$r_{3N} = \sqrt{N}\int_{\Delta_N\cap S_\gamma}\psi'(F,\theta_N)(F_m - F)d(G_n - G)$$

$$r_{4N} = -\sqrt{N}\int_{\Delta_N^c\cup S_\gamma^c}\psi'(F,\theta_N)(F_m - F)dG$$

$$r_{5N} = \sqrt{N}\int_{\Delta_N\cap S_\gamma^c\cap E_N}\psi'(F^*,\theta_N)(\hat{F} - F)dG_n$$

$$r_{6N} = \sqrt{N}\int_{\Delta_N\cap S_\gamma^c\cap E_N^c}\psi(\hat{F},\theta_N)dG_n$$

$$r_{7N} = -\sqrt{N}\int_{\Delta_N\cap S_\gamma^c\cap E_N^c}\psi(F,\theta_N)dG_n$$

Here F^* is a random function assuming values between F and $\hat{F}$. We shall show that for any fixed γ, r_{1N}, r_{2N}, r_{3N} and r_{6N} converge in probability to 0 and r_{4N}, r_{5N} and r_{7N} converge in probability to 0 as $\gamma \to 0$ and $N \to \infty$.

Lemma 4.1. For fixed γ, $r_{1N} \to_P 0$ as $N \to \infty$.

Proof. Given $\gamma \in (0,1)$ let $\Omega_{\gamma N} = \{\omega: \sup_{S_\gamma}|\hat{F} - F| < \gamma/2\}$. Then $r_{1N} = I(\Omega_{\gamma N})r_{1N} + I(\Omega_{\gamma N}^c)r_{1N}$. The second term converges in probability to 0, by the Glivenko - Cantelli theorem. Further, we have $\sup_{S_\gamma}\sqrt{m}|\hat{F} - F| = O_p(1)$. The function $\psi'(u,\theta_N)$ is uniformly continuous on $[\gamma/2, 1-\gamma/2]$ and $|F^* - F| \le |\hat{F} - F|$. Therefore, by the Glivenko-Cantelli theorem,

$$I(\Omega_{\gamma N}|r_{1N}| \le O_p(1)\,\sup_{S_\gamma}|\psi'(F^*,\theta_N) - \psi'(F,\theta_N)| \underset{P}{\to} 0.$$

Lemma 4.2. For fixed γ, $r_{2N} \underset{P}{\to} 0$ as $N \to \infty$.

Proof. This follows since $|\hat{F} - F_m| \le 1/(m+1)$ and $\int_{S_\gamma} \psi'(F)dG_n \underset{P}{\to} \int_{S_\gamma} \psi'(F)dG$.

Lemma 4.3. For fixed γ, $r_{3N} \to_P 0$ as $N \to \infty$.

Proof. For each positive integer k define a function χ_k on $[0,1]$ by

$$\chi_k(0) = 0$$
$$\chi_k(s) = (i-1)/k \text{ for } (i-1)/k < s \le i/k \quad i = 1,\dots,k.$$

Then, by Lemma 4.3 of Ruymgaart et al. (1972)

$$\sup\sqrt{m}\,|F_m F^{-1}\chi_k(F) - \chi_k(F) - F_m + F| \to_P 0$$

as $k, m \to \infty$. Furthermore $|r_{3N}| \le \sum_{i=1}^{3} r_{3iN}$ where

$$r_{31N} = \sqrt{N} \int_{\Delta_N \cap S_\gamma} |(F_m - F)\psi'(F,\theta_N) - (F_m F^{-1}\chi_k(F) - \chi_k(F))\psi'(\chi_k(F),\theta_N)|\,dG_n$$
$$r_{32N} = \sqrt{N} \int_{\Delta_N \cap S_\gamma} |(F_m - F)\psi'(F,\theta_N) - (F_m F^{-1}\chi_k(F) - \chi_k(F))\psi'(\chi_k(F),\theta_N)|\,dG$$
$$r_{33N} = |\sqrt{N} \int_{\Delta_N \cap S_\gamma} (F_m F^{-1}\chi_k(F) - \chi_k(F))\psi'(\chi_k(F),\theta_N)d(G_n - G)|.$$

The proof is similar to Corollary 5.5 in Ruymgaart et. al (1972). Given $\varepsilon > 0$ there exist constants M and $\eta_{kN} \to 0$ as $k,N \to \infty$ such that the sets

$$\Omega_N = \{\sup\sqrt{m}\,|F_m - F| \le M\}$$
$$\Omega_{kN} = \sup\{\sqrt{m}\,|F_m - F - F_m F^{-1}\chi_k(F) + \chi_k(F)| \le \eta_{kN}\}$$

have probability at least $1 - \varepsilon$. Further, the function $\psi'(u,\theta_N)$ is bounded by M_γ and uniformly continuous on $[0, 1-\gamma]$. Finally

$$\xi_{k\gamma N} = \sup_{S\gamma} |\psi'(F,\theta_N) - \psi'(\chi_k(F),\theta_N)| \to 0.$$

Therefore, for i=1,2

$$I(\Omega_{kN} \cap \Omega_N) r_{3iN} \le (\eta_{kN} M_\gamma + M\xi_{k\gamma N})\sqrt{N/m} \to 0.$$

Finally, for each $\omega \in \Omega_N$, the integrand of r_{33N} is a step function assuming value a_{ikN}, $|a_{ikN}| \le M(M_\gamma + \xi_{k\gamma N})\sqrt{N/m}$ on the interval $(F^{-1}((i-1)/k), F^{-1}(i/k)] = R_{ikN}$ and

$$I(\Omega_N) r_{33N} = |\sum_{i=1}^{k} a_{ikN} \int_{R_{ikN}} d(G_n - G)| \le 2kM(M_\gamma + \xi_{k\gamma N})\sqrt{N/m} \sup|G_n - G| \to_P 0$$

as $N \to \infty$.

Lemma 4.4. As $\gamma \downarrow 0$ and $N \to \infty$, $r_{4N} \to_P 0$.

Proof. Let $\varepsilon > 0$ be fixed. By Theorem A1 of Shorack (1972), there exists a constant M such that the set

$$\Omega_N = \{\sqrt{m}(F_m - F) \le M r^{-1/2+\tau-\eta}(F)\}$$

has probability at least $1 - \varepsilon$. By assumption A.1

$$(4.1) \qquad I(\Omega_N)|r_{4N}| \le O(1)M\sqrt{N/m} \int_{\Delta_N^c \cup S_\gamma^c} r(F)^{1-\eta} dG.$$

Let us consider the integral

$$(4.2) \qquad \int_{S_\gamma^c} r(u)^{1-\eta} d(u, \theta) du.$$

By assumption A.2, we can find a value $\tilde{\gamma}$ of γ such that (4.2) is less than ε provided $\gamma \le \tilde{\gamma}$. For this $\tilde{\gamma}$ there exists $\tilde{N}$ such that $P(\Delta_N \supset S_{\tilde{\gamma}}) > 1 - \varepsilon$ provided $N > \tilde{N}$. It follows that the integral on the right of (4.1) is less than ε with probability larger than $1 - \varepsilon$ for $\gamma \le \tilde{\gamma}$ and $N \ge \tilde{N}$. Thus $r_{4N} \underset{P}{\to} 0$ as $\gamma \downarrow 0$ and $N \to \infty$.

Lemma 4.5. As $\gamma \downarrow 0$ and $N \to \infty$, $r_{5N} \underset{P}{\to} 0$.

Proof. Let $\varepsilon > 0$ be fixed. By Lemmas 6.1 and 6.2 in Ruymgaart et al. (1972), there exist constants M_1 and M_2 such that the sets $\Omega_{1N} = \{\sqrt{m}\,|\hat{F} - F| \le M_1 r^{-1/2+\tau-\eta}(F) \text{ on } E_N\}$ and $\Omega_{2N} = \{r^{a+1}(F^*) \le M_2 r^{a+1}(F) \text{ on } E_N\}$ have probability at least $1 - \varepsilon$. Then

$$EI(\Omega_{1N} \cap \Omega_{2N})\,|r_{5N}| \le O(1)M_1M_2\sqrt{N/m}\int_{S_\gamma^c} r^{1-\eta}(F)dG.$$

The right-hand side converges to 0 as $\gamma \downarrow 0$.

Lemma 4.6. For fixed γ, $r_{6N} \xrightarrow[P]{} 0$ as $N \to \infty$.

Proof. Let $\varepsilon > 0$ be fixed. Given $\gamma \in (0,1)$, we can find $\tilde{N}$ such that the set $\Omega_N = \{S_\gamma \subset E_N\}$ has probability at least $1 - \varepsilon$ for $N > \tilde{N}$. By assumption A.1, $EI(\Omega_N)\,|r_{6N}| \le O((N/m)^{1-\tau})m^{1-\tau}P(Y \in E_N^c)$. But

$$P(Y \in E_N^c) = \int u^m d(u,\theta_N)du + \int (1-u)^m d(u,\theta_N)du.$$

By assumption A.3, $EI(\Omega_N)|r_{6N}| \to 0$.

Lemma 4.7. As $\gamma \downarrow 0$ and $N \to \infty$, $r_{7N} \xrightarrow[P]{} 0$.

Proof. By assumption A.1,

$$|r_{7N}| \le \sqrt{N} \int_{\Delta_N \cap S_\gamma^c \cap E_N^c} r(F)^a dG_n$$

Hölders inequality yields

$$E|r_{7N}| \le \sqrt{N/m}\{mP(Y_i \in E_N^c)\}^{1/2}\{\int_{S_\gamma^c} r(F)^{2a}dG\}^{1/2}.$$

By assumption A.2 and A.3, $mP(Y_i \in E_N^c) = O(1)$ and the integral on the right-hand side converges to 0 as $\gamma \downarrow 0$.

References

ANSCOMBE, F. J. and TUKEY, J. W. (1954). The criticism of transformation. Unpublished manuscript.

BEGUN, J. M. (1981). A class of rank estimators of relative risk. Institute of Statistics Mimeo Series No.1381. Univ. of North Carolina at Chapel Hill.

BEGUN,J.M. & WELLNER, J.A. (1983). Asymptotic efficiency of relative risk estimates. In *Contributions to Statistics: Essays in Honor of Norman L. Johnson*, ed by P.K. Sen.

BEGUN, J.M., HALL, W.J., HUANG, W.M. & WELLNER, J.A. (1983). Information and asymptotic efficiency in parametric — nonparametric models. *Ann. Statist.* **11**, 432-452.

BENNETT, S. (1983). Log-logistic regression models for survival data. *Appl. Statist.*, **32**, 165-171.

BICKEL, P. J. (1986). Efficient testing in a class of transformation models. Tech. Report, University of California, Berkeley.

BICKEL, P. J. and DOKSUM, K. A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.*, **76**, 293-311.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformation. *J. Roy. Statist. Soc. B*, **26**, 211-252.

CLAYTON, D. and CUZICK, J. (1986). The semiparametric Pareto model for regression analysis of survival times. Proc. ISI, Amsterdam.

COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. B*, **34**, 187-202.

COX, D. R. (1975). Partial likelihood. *Biometrika*, **62**, 269-276.

DABROWSKA, D. M. and DOKSUM, K. A. (1988). Partial likelihood in transformation models with censored data. To appear in *Scand. J. Statist.*

DOKSUM, K. A. (1987). An extension of partial likelihood methods for proportional hazard model to general transformation models. *Ann. Statist.*, **15**, (March).

DOKSUM, K. A. and NABEYA, S. (1984). Estimation in proportional hazard and log-linear models. *J. Statist. Plan. Infer.* **9**, 297-303.

FERGUSON, T. S. (1967). Mathematical Statistics (page 257). Academic Press, New York.

HARRINGTON, D. P. and FLEMING, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika*, **69**, 533-546.

HODGES, J. L. JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.*, **34**, 598-611.

HUBER, P. J. (1981). Robust Statistics. Wiley, New York.

LEHMANN, E. L. (1953). The power of rank tests. *Ann. Math. Statist.*, **24**, 23-43.

MIURA, R. (1985). Hodges-Lehmann type estimates and Lehmann's alternative models. A special lecture presented at the annual meeting of Japanese Mathematical Society, the Division of Statistical Mathematics. April, 1985. (In Japanese).

PETTITT, A. N. (1984). Proportional odds model for survival data and estimates using ranks. *J. Roy. Statist. Soc. C*, **33**, 169-175.

RUYMGAART, F. H., SHORACK, G. R. and VAN ZWET, W. R. (1972). Asymptotic normality of nonparametric tests for independence. *Ann. Math. Statist.*, **43**, 1122-1135.

SAVAGE, I. R. (1956). Contributions to the theory of rank order statistics: the two sample case. *Ann. Math. Statist.*, **31**, 415-426.

TUKEY, J. W. (1957). On the comparative anatomy of transformations. *Ann. Math. Statist.*, **28**, 602-632.

WELLNER, J. A. (1986). Semiparametric models; progress and problems. *Proc. ISI*, Amsterdam.

WONG, W. H. (1986). Theory of partial likelihood. *Ann. Statist.*, **14**, 88-123.