

ON THE PERFORMANCE OF ESTIMATES IN PROPORTIONAL  
HAZARD AND LOG-LINEAR MODELS

Kjell A. Doksum

Department of Statistics, University of California, Berkeley

1. Introduction

Let  $T_1, T_2, \dots, T_n$  be independent survival times with  $T_i$  having distribution function (d.f.)  $F_i$ , density  $f_i$  and hazard rate  $\lambda_i(t) = f_i(t)/[1-F_i(t)]$ .

One model often used in the analysis of survival experiments is the proportional hazard model where

$$(1) \quad \lambda_i(t) = \Delta_i \lambda(t) \quad , \quad t \geq 0$$

for some constant  $\Delta_i > 0$ . Here  $\lambda(t) = f(t)/[1-F(t)]$  for d.f.  $F$  with density  $f$ .

In a different context, this model was considered by Lehmann (1953) and Savage (1956) in the equivalent form  $F_i(t) = 1 - [1-F(t)]^{\Delta_i}$ , some d.f.  $F$ . It was used by Cox (1972) in situations where the distribution of  $T_i$  depends on  $p$  covariates  $x_{i1}, \dots, x_{ip}$ . Cox modeled this dependence by assuming

$$(2) \quad \lambda_i(t) = \Delta_i \lambda(t) \quad , \quad \Delta_i = \exp\left(\sum_{j=1}^p x_{ij} \beta_j\right) \quad ,$$

where  $\tilde{\beta} = (\beta_1, \dots, \beta_p)^T$  is a vector of regression coefficients.

Another model often used with survival distributions is the scale model where

$$(3) \quad F_i(t) = G(t/\tau_i) \quad , \quad \text{some } \tau_i > 0 \quad , \quad \text{some d.f. } G \quad .$$

When  $F_i$  depends on covariates, one way to model this dependence is by writing

$$(4) \quad \log T_i = \sum_{j=1}^p x_{ij} \theta_j + e_i \quad .$$

Here  $\underline{x} = (x_{ij})$  are the same covariates as before and  $\underline{\theta} = (\theta_1, \dots, \theta_p)^T$  is a vector of regression coefficients. Note that (4) is a special case of (3) with  $\tau_i = \exp(\sum_{j=1}^p x_{ij} \theta_j)$  and  $G(t) = H(\log t)$  where  $H$  is the d.f. of  $e_i$ .

In certain studies, there will be censoring variables  $C_1, \dots, C_n$ , and one observes  $T'_i = \min(T_i, C_i)$ , and  $\delta_i = I[T_i \leq C_i]$ , rather than  $T_i$ , where  $I$  is the indicator function.

Cox (1972) has introduced partial likelihood estimates for the model (2); and Miller (1976), Buckley and James (1979), and Koul, Susarla and Van Ryzin (1981) have considered least squares type estimators for the model (4).

In the next sections, we will first show that the models (1) and (3) coincide only for the Weibull model and then make asymptotic comparisons between the Cox estimates, least squares type estimates and rank estimates. In the Weibull model, the rank estimates are asymptotically optimal. Efficiency results are obtained in very special cases.

## 2. The Equivalence of the Proportional Hazard and Log-linear Models

The result that the proportional hazard and log-linear models coincide only when  $T_i$  has a Weibull distribution has appeared in Doksum (1975), Kalbfleisch and Prentice (1980, p.34) and Louis (1981). Only the second reference gives a proof and in this proof the covariates  $\underline{x}$  are allowed to vary and in fact are allowed to be functions of the regression coefficients.

We give a different proof which requires that (i) the proportional hazard model (1) and scale model (3) coincide when  $\tau_i$  and  $\Delta_i$  are unity, and (ii) that they coincide for at least one value of  $\tau_i$  different from unity. We also need the regularity condition: (iii) For some  $a > -1$ ,  $\lim_{t \rightarrow 0^+} [\lambda(t)/t^a]$  exists and is positive.

The proof proceeds as follows: From (i) we conclude that  $G$  in (3) equals the  $F$  in  $\lambda(t) = f(t)/[1-F(t)]$  of model (1). Now (3) and (ii) implies that  $\lambda_i(t) = \lambda(t/\tau_i)/\tau_i$  for some  $\tau_i \neq 1$ . When this is combined with (1), we obtain

$$(5) \quad \lambda(t/\tau) = \tau\Delta\lambda(t) \quad , \quad \text{all } t \quad , \quad \text{some } \tau \neq 1 \quad ,$$

where we dropped the subscripts on  $\tau_i$  and  $\Delta_i$ . We will show that (5) implies that  $\lambda(t)$  must be the failure rate of a Weibull distribution, i.e., that  $\lambda(t)$  is of the form  $\lambda(t) = ct^{\gamma-1}$ , some  $\gamma > 0$ .

First suppose that  $0 < \lambda(0) < \infty$ , then  $\lambda(0) = \Delta\tau\lambda(0)$  implies  $\Delta\tau = 1$ , i.e.,  $\lambda(t/\tau) = \lambda(t)$  for all  $t \geq 0$ . Now when  $\tau \neq 1$  and (iii) holds, this implies  $\lambda(t) = \text{constant}$ , i.e., the model is exponential.

Next suppose that  $\lambda(0) = 0$ . Let  $h(t) = \lambda(t)/t^a$  where  $a$  is given in (iii). Using (5), we find  $h(t/\tau) = \Delta\tau^{a+1}h(t)$ . Now since  $0 < h(0^+) < \infty$ ,  $h(0^+) = \Delta\tau^{a+1}h(0^+)$  implies  $\Delta\tau^{a+1} = 1$ , thus  $h(t/\tau) = h(t)$  for all  $t > 0$ . Since  $\tau \neq 1$ , this implies  $h(t) = \text{constant}$ , i.e.,  $\lambda(t) = ct^a$ , and the model is Weibull.

Equations that include equation (5) as a special case can be found in Kuczma (1968, p.47) and Nabeya (1974), but the present solution is not given there.

In the Weibull model we use the notation where  $T_i$  had d.f.

$$(6) \quad F_i(t) = 1 - \exp\{-(t/\tau_i)^\gamma\} \quad .$$

Here  $\log \tau_i = \sum_{j=1}^p x_{ij} \theta_j$  as in (4). In the Cox model (2), the model (6) corresponds to  $\lambda(t) = t^{\gamma-1}$ ,  $\Delta_i = \tau_i^{-\gamma}$ . Thus  $\log \tau_i = -\gamma^{-1} \sum_{j=1}^p x_{ij} \beta_j$ , and the correspondence between  $\tilde{\theta}$  and  $\tilde{\beta}$  in the Weibull model is  $\tilde{\theta} = -\gamma^{-1} \tilde{\beta}$ .

### 3. The Estimates

#### 3.1 Least Squares (L.S.) Type Estimates

We consider only the uncensored case. The asymptotic variance of L.S. type estimates has been obtained for certain types of censoring by Miller (1976) and Koul, Susarla and Van Ryzin (1981). We consider the model (4) with  $\underline{x}$  of full rank and  $e_1, \dots, e_n$  i.i.d. The variance of the L.S. estimate  $\hat{\underline{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$  is then  $\sigma^2 (\underline{x}^T \underline{x})^{-1}$ , where  $\sigma^2 = \text{Var}(e_1)$ . If we specialize to the Weibull model (6) we find that  $e_i$  has d.f. H given by

$$H(t) = 1 - \exp[-\exp(\gamma t)] \quad ,$$

and variance  $\text{Var}(e_i) = \text{Var}(\log T_i) = \pi^2/6\gamma^2$ .

Note that  $E(e_i)$  is not equal to zero; in fact  $E(e_i) = -\bar{E}/\gamma$  where  $\bar{E}$  = Euler's constant  $\approx .5772$ . It follows that the L.S. type estimates are not necessarily consistent for the Weibull model. Thus, if  $p=2$ , and  $\log T_i = \theta_1 + \theta_2 x_i + e_i$ , then  $\hat{\theta}_1$  converges in probability to  $\theta_1 - E/\gamma$  while  $\hat{\theta}_2$  is consistent. This can be "fixed" by reparametrizing: Set  $\log T_i = \theta'_1 + \theta_2 x_i + e_i$ , where  $e'_i = e_i - E(e_i)$  and  $\theta'_1 = \theta_1 + E(e_i)$ . Note that  $E(e_i)$  is unknown if  $\gamma$  is, but we can think of the L.S. estimate as an estimate of the intercept after the errors have been adjusted to have mean zero.

#### 3.2 Cox Estimates

Relevant asymptotic results can be found in papers by Efron (1977), Oakes (1977), Aalen (1978, 1980), Bailey (1979), Tsiatis (1981), and the book by Kalbfleisch and Prentice (1980). In the model (2) with no censoring, let  $t_{(1)} < \dots < t_{(n)}$  be the ordered observed survival times and let  $\underline{x}_{(i)} = (x_{(i)1}, \dots, x_{(i)p})$  be the covariates corresponding to  $t_{(i)}$ . Then the Cox estimate  $\hat{\underline{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$  is the value that maximizes the Cox partial likelihood

$$L_c = \prod_{i=1}^n \left\{ \frac{\exp \underline{x}_{(i)} \underline{\beta}}{\sum_{s=1}^n \exp \underline{x}_{(s)} \underline{\beta}} \right\} \quad .$$

The asymptotic covariance matrix of  $(\hat{\beta}_1, \dots, \hat{\beta}_p)$  is the inverse of the expected value of the observed Cox partial information matrix defined by

$$I_{kl}^c = \sum_{i=1}^n \left\{ \frac{\sum_{j=i}^n x_{(j)k} x_{(j)\ell} \exp\{x_{(j)} \beta\}}{\sum_{j=i}^n \exp\{x_{(j)} \beta\}} - \frac{\left( \sum_{j=i}^n x_{(j)k} \exp\{x_{(j)} \beta\} \right) \left( \sum_{j=i}^n x_{(j)\ell} \exp\{x_{(j)} \beta\} \right)}{\left( \sum_{j=i}^n \exp\{x_{(j)} \beta\} \right)^2} \right\}.$$

Note that the only quantity that is random in this expression is the index (j) in  $x_{(j)}$ ,  $x_{(j)k}$  and  $x_{(j)\ell}$ .

### 3.3 Rank Estimates

We consider the log-linear model (4). Properties of estimates based on ranks were developed for the two-sample problem by Hodges and Lehmann (1963), considered for Type II censoring by Doksum (1967), extended to simple linear regression by Adichie (1967), and to multiple linear regression by Jureckova (1971). Let  $R_i$  be the rank of  $T_i$  among  $T_1, \dots, T_n$ . Since  $\text{Rank}(T_i) = \text{Rank}(\log T_i)$ , the rank approach to log-linear models reduces the estimation problem to the problem of estimating the parameters in a linear model for  $Y_i = \log T_i$ . The idea in the above references is to use the estimates  $\hat{\theta}_1, \dots, \hat{\theta}_p$  obtained by "inverting" linear rank statistics of the type

$$S_j = \frac{1}{n} \sum_{i=1}^n (x_{ij} - x_{.j}) J_n \left( \frac{R_i}{n+1} \right); \quad j=1, \dots, p$$

where  $J_n \left( \frac{1}{n+1} \right), \dots, J_n \left( \frac{n}{n+1} \right)$  are given scores (constants) and  $x_{.j} = \frac{1}{n} \sum_{i=1}^n x_{ij}$ . When  $H_0: \theta = \underline{0}$  holds, the distribution of  $\underline{S} = (S_1, \dots, S_p)$  tends to be concentrated near its mean  $E_{H_0}(\underline{S}) = \underline{0}$ . When  $\theta \neq \underline{0}$ , let  $R_i^\theta$  denote the rank of  $Y_i - x_{i\cdot} \theta$ , where  $x_{i\cdot} = (x_{i1}, \dots, x_{ip})$ , and let  $S_j(Y - x\theta) = \frac{1}{n} \sum_{i=1}^n (x_{ij} - x_{.j}) J_n \left( \frac{R_i^\theta}{n+1} \right)$ . When  $\theta$  is the true value of the parameter, the distribution of  $S_j(Y - x\theta)$  will be concentrated near zero, thus the idea is to use the estimate  $\hat{\theta}$  which "solves"  $S_j(Y - x\hat{\theta}) = 0$ ,  $j=1, \dots, p$ , for  $\hat{\theta}$ . Exact definitions and conditions are in the above references.

Note that since the ranks are invariant under additions of constants, i.e.,  $\text{Rank}(\log T_i + a) = \text{Rank}(\log T_i)$ , this approach can not be used to estimate  $\alpha$  in the model  $\log T_i = \alpha + \beta x_i + e_i$ . Adichie (1967) and Jureckova (1971) introduce rank estimates for  $\alpha$ . We do not treat those here.

$\hat{\theta}$  is consistent and  $(\hat{\theta} - \theta)$  (standardized) is asymptotically normal with mean zero.  $\hat{\theta} - \theta$  has approximate covariance matrix

$$A^2 B^{-2} (\mathbf{x}^T \mathbf{x})^{-1}$$

where  $A^2 = \int_0^1 J^2(u) du - [\int_0^1 J(u) du]^2$ ,  $B = \int_{-\infty}^{\infty} [\frac{d}{dx} J(H(x))] dH(x)$ , and  $J$  is the limiting score function,  $J(u) = \lim_{n \rightarrow \infty} J_n(\frac{[nu]+1}{n+1})$ ,  $0 < u < 1$ . Here  $[ ]$  is the greatest integer function.

Let  $\sigma^2(\hat{\theta}; J, G)$  denote the asymptotic variance vector of  $[(\hat{\theta}_1 - \theta_1)/b_1, \dots, (\hat{\theta}_p - \theta_p)/b_p]$  where  $b_j = [\sum_{i=1}^n (x_{i1} - x_{.j})^2]^{-1/2}$ . If the distribution  $G$  is known, and thus  $H$  is also known,  $\sigma^2(\hat{\theta}; J, G)$  is minimized by choosing  $J_n(\frac{i}{n+1}) = E[\phi(U^{(i)}, H)]$ , where  $\phi(u, H) = -h'(H^{-1}(u))/h(H^{-1}(u))$ ,  $h$  is the density of  $H$ , and  $U^{(i)}$  is the  $i^{\text{th}}$  uniform order statistic in a sample of size  $n$ . Another optimal choice is the simpler form  $J_n(\frac{i}{n+1}) = \phi(\frac{i}{n+1}, H)$ . These results follow from Hájek and Šidák (1967) and the above references.

In particular, if  $T_i$  has the Weibull distribution (6), then the optimal  $J_n$  is  $J_n(\frac{i}{n+1}) = E\{-\log(1-U^{(i)})\} = \sum_{j=n+1-i}^n (1/j)$ , the exponential or Savage scores. The asymptotically equivalent simpler version is  $J_n(\frac{i}{n+1}) = -\log\{1-[i/(n+1)]\}$ . Note that these functions do not depend on the shape parameter  $\gamma$  of the Weibull model, thus the exponential scores estimate  $\hat{\theta}_{\sim R}$  obtained by setting  $J_n(\frac{i}{n+1}) = \sum_{j=n+1-i}^n (1/j)$  or  $J_n(\frac{i}{n+1}) = -\log\{1-[i/(n+1)]\}$  minimizes the asymptotic variance uniformly in  $\gamma$ . This optimality does not hold only in the class of rank estimates, but over the class of all "regular" estimates including least squares and Cox estimates.

The exponential scores estimate has another strong optimality property. It is asymptotically minimax over the class of increasing failure rate average

(IFRA) distributions. More precisely, let  $\sigma^2(\hat{\theta}; J) = \sup_G \sigma^2(\hat{\theta}, J, G)$ , where the sup is over all  $G$  continuous and IFRA. The estimate which minimizes  $\sigma^2(\hat{\theta}; J)$  is the exponential scores estimate; moreover for this estimate, the maximum approximate variance (i.e., the maximum of  $A^2 B^{-2}$ ) is attained at the exponential distribution. In fact the approximate covariance matrix  $\Sigma(G)$  of  $\hat{\theta}_R$  is such that for the exponential distribution, it reduces to the familiar matrix  $(\tilde{x}^T \tilde{x})^{-1}$ . Thus we can think of  $(\tilde{x}^T \tilde{x})^{-1}$  as a lower bound for the covariance matrix of  $\hat{\theta}_R$  for IFRA distributions. This result leads immediately to simple bounds on the standard error of  $\hat{\theta}_R$  and confidence intervals for  $\theta$ . These results are extensions of Doksum (1967).

Rank estimates for Type II censoring was considered in the two sample case by Doksum (1967). Rank test statistics for censored samples have been considered by Rao-Savage and Sobel (1960), Gastwirth (1965), Gehan (1965), Mantel (1966), Efron (1967), Basu (1968), Doksum (1969), Johnson and Mehrotra (1972), Peto and Peto (1972), Cox (1972), Crowley (1974), Prentice (1978) and Aalen (1978), among others.

The following is an extension of the exponential scores statistic: The survival times are ordered. The first survival time is given score =  $\frac{1}{n_1}$ , the  $(k+1)^{st}$  is given the score of the  $k^{th}$  plus the reciprocal of the number  $n_{k+1}$  of subjects at risk right before the  $(k+1)^{st}$  death. The censoring time  $C_i$  is given the score of the largest survival time  $T$  to the left of  $C_i$  plus one. "One" can be interpreted as the average of the possible scores to the right of (and including) the score of  $T$ . If there is no survival time  $T$  to the left of  $C_i$ ,  $C_i$  gets score one. If this scheme is used, the asymptotic normality and optimality of the exponential scores estimate carries over to Type II censored samples in the two-sample case (Doksum (1967)).

#### 4. Comparisons

From the considerations in Section 3, we know that for the Weibull model without any censoring, the rank exponential scores estimate  $\hat{\theta}_R$  is asymptotically

optimal. The asymptotic efficiency of the least squares type estimate  $\hat{\theta}_{\sim LS}$  is  $e(\hat{\theta}_{\sim LS}, \hat{\theta}_{\sim R}) = (6/\pi^2) = .61$  for all values of the Weibull parameter. To study the efficiency of the Cox estimate, we need to consider the two-sample problem. We let the parameter of interest be the ratio  $\delta$  of the means of the survival distributions. When  $\gamma = 1$ , we find

$\delta$	1	2	4	8	16
$e(\hat{\delta}_{\text{COX}}, \hat{\delta}_{\text{LS}})$	1.6	1.5	1.2	.83	.55
$e(\hat{\delta}_{\text{COX}}, \hat{\delta}_{\text{R}})$	1	.90	.71	.50	.33

From the results of Section 3, a qualitatively similar result should hold for  $\gamma \neq 1$ , Type II censoring and more general designs, but we do not have exact figures.

## 5. DISCUSSION

The asymptotics for the Weibull and increasing failure rate average models clearly favor the rank exponential scores estimate. However, this estimate is hard to compute and its finite sample size properties are not well known. Moreover, if we consider a different model such as the log normal model for the distribution of  $T_i$ , then the LS type estimate will be best in the case of no censoring. In this model, the optimal rank estimate would be the rank normal scores estimate.

## ACKNOWLEDGEMENT

This research was partially supported by NSF grant MCS-81-92349.

## REFERENCES

- Aalen, O. (1978). Nonparametric inference for a family of counting processes. Annals of Statistics 6, 701-726.
- Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. Proceedings, Sixth International Conference, Mathematical Statistics and Probability Theory, Wisla, Poland, Springer Lecture Notes in Statistics.
- Adichie, J. (1967). Estimation of regression parameters based on rank tests. Annals of Mathematical Statistics 38, 894-904.
- Basu, A.P. (1967). On a generalized Savage statistic with applications to life testing. Annals of Mathematical Statistics 39, 1591-1604.
- Bailey, K.R. (1979). The general maximum likelihood approach to the Cox regression model. PhD dissertation, University of Chicago, Chicago, Ill.
- Buckley, J. and James, I. (1979). Linear regression with censored data. Biometrika 66, 429-436.
- Cox, D.R. (1972). Regression models and life tables (with discussion). Journal of the Royal Statistical Society B 34, 187-220.
- Crowley, J. (1974). Asymptotic normality of a new nonparametric statistic for use in organ transplant studies. Journal of the American Statistical Association 69, 1006-1011.
- Doksum, K.A. (1967). Asymptotically optimal statistics in some models with increasing failure rate averages. Annals of Mathematical Statistics 38, 1731-1739.
- Doksum, K.A. (1969). Minimax results for IFRA scale alternatives. Annals of Mathematical Statistics 40, 1778-1783.
- Doksum, K.A. (1975). Measures of difference in reliability. Proceedings of the Conference on Reliability and Fault Tree Analysis, Barlow, Fussell and Singpurwalla, eds., SIAM, 427-499.
- Efron, B. (1967). The two-sample problem with censored data. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. IV. University of California Press, Berkeley, California, 831-853.

- Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association 72, 557-565.
- Gastwirth, J.L. (1965). Asymptotically most powerful rank tests for the two-sample problem with censored data. Annals of Mathematical Statistics 36, 1243-1247.
- Gehan, E.A. (1965). A generalized Wilcoxon test for comparing arbitrary singly-censored samples. Biometrika 52, 203-223.
- Hájek, J. and Šidák, Z. (1967). Theory of Rank Tests. Academic Press, New York.
- Hodges, J. and Lehmann, E. (1963). Estimates of location based on rank tests. Annals of Mathematical Statistics 34, 598-611.
- Johnson, R.A. and Mehrotra, K.G. (1972). Locally most powerful rank tests for the two-sample problem with censored data. Annals of Mathematical Statistics 43, 823-831.
- Jureckova, J. (1971). Nonparametric estimate of regression coefficients. Annals of Mathematical Statistics 42, 1328-1338.
- Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.
- Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right censored data. Annals of Statistics 9, 1276-1288.
- Kuczma, M. (1968). Functional Equations in a Single Variable, Polish Scientific Publishers, Warszawa.
- Lehmann, E.L. (1953). The power of rank tests. Annals of Mathematical Statistics 24, 23-43.
- Louis, T.A. (1981). Nonparametric analysis of an accelerated failure time model. Biometrika 68, 381-390.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemotherapy Reports 50, 163-170.
- Miller, R.G. (1976). Least squares regression with censored data. Biometrika 63, 449-464.

Nabeya, Seiji (1974). On the functional equation  $f(p+qx+rf(x)) = a+bx+rf(x)$ .

Aequationes Mathematicae 11, 199-211.

Oakes, D. (1977). The asymptotic information in censored survival data.

Biometrika 64, 441-448.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society A 135, 185-206.

Prentice, R.L. (1978). Linear rank tests with right censored data. Biometrika 65, 167-179.

Rao, V.V.R., Savage, I.R. and Sobel, M. (1960). Contributions to the theory of rank order statistics: The two-sample censored case. Annals of Mathematical Statistics 31, 415-426.

Savage, I.R. (1956). Contributions to the theory of rank order statistics - the two-sample case. Annals of Mathematical Statistics 27, 590-615.

Tsiatis, A. (1981). A large sample study of Cox's regression model. Annals of Statistics 9, 93-108.