A
Numerical Techniques for Maximization

Many of the procedures discussed in this volume require maximizing the log likelihood or partial log likelihood function. For many models, it is impossible to perform this maximization analytically, so, numerical methods must be employed. In this appendix, we shall summarize some techniques which can be used in both univariate and multivariate cases. The reader is referred to a text on statistical computing, such as Thisted (1988), for a more detailed discussion of these techniques.

A.1 Univariate Methods

Suppose we wish to find the value x which maximizes a function $f(\,)$ of a single variable. Under some mild regularity conditions, x maximizes f if the score equation $f'(x)$ equals 0 and $f''(x) < 0$. We present three numerical methods which attempt to find the maximum of $f(\,)$ by solving the score equation. Some care must be taken when using these routines because they do not ensure that the second derivative of f is negative at the solution we find.

The first technique is the bisection method. Here, the algorithm starts with two initial values, x_L and x_U, which bracket the root of $f'(x) = 0$, that is, $f'(x_L) \cdot f'(x_U) < 0$. A new guess at the root is taken to be the midpoint of the interval (x_L, x_U), namely, $x_N = (x_L + x_U)/2$. If $f'(x_L)$ and $f'(x_N)$ have the same sign, x_L is replaced by x_N, otherwise x_U is replaced by x_N. In either case, the algorithm continues with the new values of x_L and x_U until the desired accuracy is achieved. At each step, the length of the interval $(x_U - x_L)$ is a measure of the largest possible difference between our updated guess at the root of $f'()$ and the actual value of the root.

A second method, of use when one has good initial starting values and a complicated second derivative of $f()$, is the secant method or *regula falsi*. Again, we start with two initial guesses at the root, x_0 and x_1. These guesses need not bracket the root. After i steps of the algorithm, the new guess at the root of $f'(x)$ is given by

$$x_{i+1} = x_i - f'(x_i)(x_i - x_{i-1})/[f'(x_i) - f'(x_{i-1})]. \qquad (A.1)$$

Iterations continue until convergence. Typical stopping criteria are

$$|x_{i+1} - x_i| < \gamma, \; |f'(x_{i+1})| < \gamma$$

or

$$|(x_{i+1} - x_i)/x_i| < \gamma,$$

where γ is some small number.

The third method is the Newton–Raphson technique. Here, a single initial guess, x_o, of the root is made. After i steps of the algorithm, the updated guess is given by

$$x_{i+1} = x_i - f'(x_i)/f''(x_i). \qquad (A.2)$$

Again, the iterative procedure continues until the desired level of accuracy is met. Compared to the secant method, this technique has the advantage of requiring a single starting value, and convergence is quicker than the secant method when the starting values are good. Both the secant and Newton–Raphson techniques may fail to converge when the starting values are not close to the maximum.

EXAMPLE A.1 Suppose we have the following 10 uncensored observations from a Weibull model with scale parameter $\lambda = 1$ and shape parameter α, that is, $h(t) = \alpha t^{\alpha-1} e^{-t^\alpha}$.

Data: 2.57, 0.58, 0.82, 1.02, 0.78, 0.46, 1.04, 0.43, 0.69, 1.37

To find the maximum likelihood estimator of α, we need to maximize the log likelihood $f(\alpha) = \ln L(\alpha) = n\ln(\alpha) + (\alpha - 1)\sum \ln(t_j) - \sum t_j^\alpha$.

Here, $f'(\alpha) = n/\alpha + \sum \ln(t_j) - \sum t_j^\alpha \ln(t_j)$, and $f''(\alpha) = -n/\alpha^2 - \sum t_j^\alpha [\ln(t_j)]^2$.

Applying the bisection method with $\alpha_L = 1.5$ and $\alpha_U = 2$ and stopping the algorithm when $|f'(\alpha)| < 0.01$, we have the following values:

Step	α_L	α_U	α_N	$f'(\alpha_L)$	$f'(\alpha_U)$	$f'(\alpha_N)$
1	1.5	2	1.75	1.798	−2.589	−0.387
2	1.5	1.75	1.625	1.798	−0.387	0.697
3	1.625	1.75	1.6875	0.697	−0.387	0.154
4	1.6875	1.75	1.71875	0.154	−0.387	−0.116
5	1.6875	1.71875	1.70313	0.154	−0.116	0.019
6	1.70313	1.71875	1.71094	0.019	−0.116	−0.049
7	1.70313	1.71094	1.70704	0.019	−0.049	−0.015
8	1.70313	1.70704	1.70509	0.019	−0.015	0.002

So, after eight steps the algorithm stops with $\hat{\alpha} = 1.705$.

For the secant method, we shall start the algorithm with $\alpha_o = 1$ and $\alpha_1 = 1.5$. The results are in the following table:

Step	α_{i-1}	α_i	$f'(\alpha_{i-1})$	$f'(\alpha_i)$	α_{i+1}	$f'(\alpha_{i+1})$
1	1	1.5	7.065	1.798	1.671	0.300
2	1.5	1.671	1.798	0.300	1.705	0.004

Here, using the same stopping rule $|f'(\alpha)| < 0.01$, the algorithm stops after two steps with $\hat{\alpha} = 1.705$.

For the Newton–Raphson procedure, we use an initial value of $\alpha_o = 1.5$. The results of the algorithm are in the following table.

i	α_{i-1}	$f'(\alpha_{i-1})$	$f''(\alpha_{i-1})$	α_i	$f'(\alpha_i)$
1	1.5	1.798	−8.947	1.701	0.038
2	1.701	0.038	−8.655	1.705	2×10^{-6}

Again, using the same stopping rule $|f'(\alpha)| < 0.01$, the algorithm stops after two steps with $\hat{\alpha} = 1.705$. Notice the first step of the Newton–Raphson algorithm moves closer to the root than the secant method.

A.2 Multivariate Methods

We present three methods to maximize a function of more than one variable. The first is the method of steepest ascent which requires only the vector of first derivatives of the function. This method is robust to the starting values used in the iterative scheme, but may require a large number of steps to converge to the maximum. The second is the multivariate extension of the Newton–Raphson method. This method, which requires both the first and second derivatives of the function, converges quite rapidly when the starting values are close to the root, but may not converge when the starting values are poorly chosen. The third, called Marquardt's (1963) method, is a compromise between these two methods. It uses a blending constant which controls how closely the algorithm resembles either the method of steepest ascent or the Newton–Raphson method.

Some notation is needed before presenting the three methods. Let $f(\mathbf{x})$ be a function of the p-dimensional vector $\mathbf{x} = (x_1, \ldots, x_p)^t$. Let $\mathbf{u}(\mathbf{x})$ be the p-vector of first order partial derivatives of $f(\mathbf{x})$, that is,

$$\mathbf{u}(\mathbf{x}) = [u_1(\mathbf{x}), \ldots, u_p(\mathbf{x})]^t, \tag{A.3}$$

where

$$u_j(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_j}, j = 1, \ldots, p.$$

Let $\mathbf{H}(\mathbf{x})$ be the $p \times p$ Hessian matrix of mixed second partial derivatives of $f(\mathbf{x})$, defined by

$$\mathbf{H}(\mathbf{x}) = (H_{ij}(\mathbf{x})),\ i, j = 1, \ldots, p \text{ where } H_{ij}(\mathbf{x}) = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}. \tag{A.4}$$

The method of steepest ascent starts with an initial guess, $\mathbf{x}_o$, of the point which maximizes $f(\mathbf{x})$. At any point, the gradient vector $\mathbf{u}(\mathbf{x})$ points the direction of steepest ascent of the function $f(\mathbf{x})$. The algorithm moves along this direction by an amount d to a new estimate of the maximum from the current estimate. The step size d is chosen to maximize the function in this direction, that is, we pick d to maximize $f[\mathbf{x}_k + d\mathbf{u}(\mathbf{x}_k)]$. This requires maximizing a function of a single variable, so that any of the techniques discussed earlier can be employed.

The updated guess at the point which maximizes $f(\mathbf{x})$ is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + d\mathbf{u}(\mathbf{x}_k). \tag{A.5}$$

The second method is the Newton–Raphson method which, like the method of steepest ascent, starts with an initial guess at the point which maximizes $f(\mathbf{x})$. After k steps of the algorithm, the updated estimate of the point which maximizes $f(\mathbf{x})$ is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}(\mathbf{x}_k)^{-1}\mathbf{u}(\mathbf{x}_k). \tag{A.6}$$

The Newton–Raphson algorithm converges quite rapidly when the initial guess is not too far from the maximum. When the initial guess is poor, the algorithm may move in the wrong direction or may take a step in the correct direction, but overshoot the root. The value of the function should be computed at each step to ensure that the algorithm is moving in the correct direction. If $f(\mathbf{x}_k)$ is smaller than $f(\mathbf{x}_{k+1})$, one option is to cut the step size in half and try $\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}(\mathbf{x}_k)^{-1}\mathbf{u}(\mathbf{x}_k)/2$. This procedure is used in SAS and BMDP in the Cox regression procedure.

The third method is Marquardt's (1963) compromise between the method of steepest ascent and the Newton–Raphson method. This

method uses a constant, γ, which blends the two methods together. When γ is zero, the method reduces to the Newton–Raphson method, and, as $\gamma \to \infty$, the method approaches the method of steepest ascent. Again, the method starts with an initial guess, $\mathbf{x}_o$. Let $\mathbf{S}_k$ be the $p \times p$ diagonal scaling matrix with diagonal element $(|\mathbf{H}_{ii}(\mathbf{x}_k)|^{-1/2})$. The updated estimate of the maximum is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{S}_k(\mathbf{S}_k\mathbf{H}(\mathbf{x}_k)\mathbf{S}_k + \gamma\mathbf{I})^{-1}\mathbf{S}_k\mathbf{u}(\mathbf{x}_k),$$

where $\mathbf{I}$ is the identity matrix. Typically, the algorithm is implemented with a small value of γ for the first iteration. If $f(\mathbf{x}_1) < f(\mathbf{x}_o)$, then, we are having difficulty approaching the maximum and the value of γ is increased until $f(\mathbf{x}_1) > f(\mathbf{x}_o)$. This procedure is iterated until convergence is attained. For the final step of the algorithm, a "Newton–Raphson" step with $\gamma = 0$ is taken to ensure convergence.

In the multivariate maximization problem, there are several suggestions for declaring convergence of these algorithms. These include stopping when $f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) < \epsilon$ (or $|[f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)]/f(\mathbf{x}_k)| < \epsilon$); when $\sum \mathbf{u}_j(\mathbf{x}_{k+1})^2 < \epsilon$ (or $\max[|\mathbf{u}_1(\mathbf{x}_{k+1})|, \ldots, |\mathbf{u}_p(\mathbf{x}_{k+1})|] < \epsilon$) or when $\sum(x_{k+1,j} - x_{k,j})^2 < \epsilon$ (or $\max[|x_{k+1,1} - x_{k,1}|, \ldots, |x_{k+1,p} - x_{k,p}| < \epsilon$).

EXAMPLE A.2 We shall fit a two-parameter Weibull model with survival function $S(t) = \exp(-\lambda t^\alpha)$ to the ten observations in Example A.2. Here the log likelihood function is given by

$$L(\lambda, \alpha) = n \ln \lambda + n \ln \alpha + (\alpha - 1)\sum \ln t_i - \lambda \sum t_i^\alpha.$$

The score vector $\mathbf{u}(\lambda, \alpha)$ is expressed by

$$u_\lambda(\lambda, \alpha) = \frac{\partial L(\lambda, \alpha)}{\partial \lambda} = \frac{n}{\lambda} - \sum t_i^\alpha$$

$$u_\alpha(\lambda, \alpha) = \frac{\partial L(\lambda, \alpha)}{\partial \alpha} = \frac{n}{\alpha} + \sum \ln t_i - \lambda \sum t_i^\alpha \ln t_i$$

and the Hessian matrix is

$$\mathbf{H}(\lambda, \alpha) = \begin{pmatrix} -\frac{n}{\lambda^2} & -\sum t_i^\alpha \ln t_i \\ -\sum t_i^\alpha \ln t_i & -\frac{n}{\alpha^2} - \lambda \sum t_i^\alpha (\ln t_i^2) \end{pmatrix}$$

To apply the method of steepest ascent, we must find the value of d_k which maximizes $L[(\lambda_k + d_k u_\lambda[\lambda_k, \alpha_k]), (\alpha_k + d_k u_\alpha[\lambda_k, \alpha_k])]$. This needs to be done numerically and this example uses a Newton–Raphson algorithm. Convergence of the algorithm is declared when the maximum of $|u_\lambda|$ and $|u_\alpha|$ is less than 0.1. Starting with an initial guess of $\alpha = 1$ and $\lambda = 10/\sum t_i = 1.024$, which leads to a log likelihood of -9.757, we have the following results:

Step k	λ_k	α_k	$L(\lambda, \alpha)$	u_λ	u_α	d_k
0	1.024	1.000	−9.757	0.001	7.035	0.098
1	1.025	1.693	−7.491	0.001	−1.80	0.089
2	0.865	1.694	−7.339	0.661	0.001	0.126
3	0.865	1.777	−7.311	0.000	−0.363	0.073
4	0.839	1.777	−7.307	0.121	0.000	0.128
5	0.839	1.792	−7.306	0.000	−0.072	0.007

Thus the method of steepest ascent yields maximum likelihood estimates of $\hat{\lambda} = 0.839$ and $\hat{\alpha} = 1.792$ after 5 iterations of the algorithm.

Applying the Newton–Raphson algorithm with the same starting values and convergence criterion yields

Step k	λ_k	α_k	u_λ	u_α	$H_{\lambda\lambda}$	$H_{\alpha\alpha}$	$H_{\alpha\lambda}$
0	1.024	1.000	0.001	7.035	−9.537	−13.449	−1.270
1	0.954	1.530	−0.471	1.684	−10.987	−8.657	−3.34
2	0.838	1.769	0.035	0.181	−14.223	−7.783	−1.220
3	0.832	1.796	−0.001	0.001	−14.431	−7.750	−4.539

This method yields maximum likelihood estimates of $\hat{\lambda} = 0.832$ and $\hat{\alpha} = 1.796$ after three iterations.

Using $\gamma = 0.5$ in Marquardt's method yields

Step k	λ_k	α_k	u_λ	u_α	$H_{\lambda\lambda}$	$H_{\alpha\alpha}$	$H_{\alpha\lambda}$
0	1.024	1.000	0.001	7.035	−9.537	−13.449	−1.270
1	0.993	1.351	−0.357	3.189	−10.136	−9.534	−2.565
2	0.930	1.585	−0.394	1.295	−11.557	−8.424	−3.599
3	0.883	1.701	−0.275	0.523	−12.813	−8.049	−4.176
4	0.858	1.753	−0.162	0.218	−13.591	−7.891	−4.453
5	0.845	1.777	−0.087	0.094	−14.013	−7.817	−4.581

Here, the algorithm converges in five steps to estimates of $\hat{\lambda} = 0.845$ and $\hat{\alpha} = 1.777$.

B Large-Sample Tests Based on Likelihood Theory

Many of the test procedures used in survival analysis are based on the asymptotic properties of the likelihood or the partial likelihood. These test procedures are based on either the maximized likelihood itself (likelihood ratio tests), on the estimators standardized by use of the information matrix (Wald tests), or on the first derivatives of the log likelihood (score tests). In this appendix, we will review how these tests are constructed. See Chapter 9 of Cox and Hinkley (1974) for a more detailed reference.

Let $\mathbf{Y}$ denote the data and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ be the parameter vector. Let $L(\boldsymbol{\theta} : \mathbf{Y})$ denote either the likelihood or partial likelihood function. The maximum likelihood estimator of $\boldsymbol{\theta}$ is the function of the data which maximizes the likelihood, that is, $\hat{\boldsymbol{\theta}}(\mathbf{Y}) = \hat{\boldsymbol{\theta}}$ is the value of $\boldsymbol{\theta}$ which maximizes $L(\boldsymbol{\theta} : \mathbf{Y})$ or, equivalently, maximizes $\log L(\boldsymbol{\theta} : \mathbf{Y})$.

Associated with the likelihood function is the efficient score vector $\mathbf{U}(\boldsymbol{\theta}) = [U_1(\boldsymbol{\theta}), \ldots, U_p(\boldsymbol{\theta})]$ defined by

$$U_j(\boldsymbol{\theta}) = \frac{\delta}{\delta\theta_j} \ln L(\boldsymbol{\theta} : \mathbf{Y}). \tag{B.1}$$

In most regular cases, the maximum likelihood estimator is the solution to the equation $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$. The efficient score vector has the property that its expected value is zero when the expectation is taken with respect to the true value of $\boldsymbol{\theta}$.

A second key quantity in large-sample likelihood theory is the Fisher information matrix defined by

$$\mathbf{i}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[\mathbf{U}(\boldsymbol{\theta})^t\mathbf{U}(\boldsymbol{\theta})] = E_{\boldsymbol{\theta}}\left[-\frac{\delta}{\delta\boldsymbol{\theta}}\mathbf{U}(\boldsymbol{\theta})\right]$$

$$= \left\{-E_{\boldsymbol{\theta}}\left[\frac{\delta^2}{\delta\theta_j\delta\theta_k}\ln L(\boldsymbol{\theta} : \mathbf{Y})\right]\right\}, \quad j = 1,\ldots,p, \; k = 1,\ldots,p. \tag{B.2}$$

Computation of the expectation in (B.2) is very difficult in most applications of likelihood theory, so a consistent estimator of $\mathbf{i}$ is used. This estimator is the observed information, $\mathbf{I}(\boldsymbol{\theta})$, whose (j, k)th element is given by

$$I_{j,k}(\boldsymbol{\theta}) = -\frac{\delta^2 \ln L(\boldsymbol{\theta} : \mathbf{Y})}{\delta\theta_j\delta\theta_k}, \quad j, k = 1,\ldots,p. \tag{B.3}$$

The first set of tests based on the likelihood are for the simple null hypothesis, $H_o : \boldsymbol{\theta} = \boldsymbol{\theta}_o$. The first test is the *likelihood ratio test* based on the statistic

$$\chi^2_{LR} = -2[\ln L(\boldsymbol{\theta}_o : \mathbf{Y}) - \ln L(\hat{\boldsymbol{\theta}} : \mathbf{Y})] \tag{B.4}$$

This statistic has an asymptotic chi-squared distribution with p degrees of freedom under the null hypothesis.

A second test, called the *Wald test*, is based on the large-sample distribution of the maximum likelihood estimator. For large samples, $\hat{\boldsymbol{\theta}}$ has a multivariate normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\mathbf{i}^{-1}(\boldsymbol{\theta})$ so the quadratic form $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)\mathbf{i}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)^t$ has a chi-squared distribution with p degrees of freedom for large samples. Using the observed information as an estimator of the Fisher information, the Wald statistic is expressed as

$$\chi^2_W = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)\mathbf{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)^t \tag{B.5}$$

which has a chi-squared distribution with p degrees of freedom for large samples when H_o is true.

The third test, called the *score or Rao test*, is based on the efficient score statistics. When $\boldsymbol{\theta} = \boldsymbol{\theta}_o$, the score vector $\mathbf{U}(\boldsymbol{\theta}_o)$ has a large-sample multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{i}(\boldsymbol{\theta})$. This leads to a test statistic given by

$$\chi^2_S = \mathbf{U}(\boldsymbol{\theta}_o)\mathbf{i}^{-1}(\boldsymbol{\theta}_o)\mathbf{U}^t(\boldsymbol{\theta}_o).$$

As for the Wald test, the Fisher information is replaced in most applications by the observed information, so the test statistic is given by

$$\chi^2_S = \mathbf{U}(\boldsymbol{\theta}_o)\mathbf{I}^{-1}(\boldsymbol{\theta}_o)\mathbf{U}^t(\boldsymbol{\theta}_o). \tag{B.6}$$

Again, this statistic has an asymptotic chi-squared distribution with p degrees of freedom when H_o is true. The score test has an advantage in

many applications in that the maximum likelihood estimates need not be calculated.

EXAMPLE B.1 Suppose we have a censored sample of size n from an exponential population with hazard rate λ. We wish to test the hypothesis that $\lambda = 1$. Let (T_i, δ_i), $i = 1,\ldots,n$, so that the likelihood, $L(\lambda; (T_i, \delta_i), i = 1,\ldots,n)$, is given by $\prod_{i=1}^n \lambda^{\delta_i} e^{-\lambda T_i} = \lambda^D e^{-\lambda S}$ where $D = \sum_{i=1}^n \delta_i$ is the observed number of deaths and $S = \sum_{i=1}^n T_i$ is the total time on test (see Section 3.5). Thus,

$$\ln L(\lambda) = D\ln\lambda - \lambda S, \tag{B.7}$$

$$U(\lambda) = \frac{d}{d\lambda}\ln L(\lambda) = \frac{D}{\lambda} - S, \tag{B.8}$$

and

$$I(\lambda) = -\frac{d^2}{d\lambda^2}\ln L(\lambda) = \frac{D}{\lambda^2}. \tag{B.9}$$

Solving B.8 for λ gives us the maximum likelihood estimator, $\hat{\lambda} = D/S$. Using these statistics,

$$\chi^2_S = \left(\frac{D}{1} - S\right)^2 \cdot \left(\frac{1^2}{D}\right) = \frac{(D-S)^2}{D},$$

$$\chi^2_W = \left(\frac{D}{S} - 1\right)^2 \cdot \frac{D}{(D/S)^2} = \frac{(D-S)^2}{D}$$

$$\chi^2_{LR} = -2\{(D\ln 1 - 1\cdot S) - [D\ln(D/S) - (D/S)\cdot S]\}$$

$$= 2[S - D + D\ln(D/S)]$$

In this case, note that the Wald and Rao tests are identical. All three of these statistics have asymptotic chi-squared distributions with one degree of freedom.

All three test statistics can be used to test composite hypotheses. Suppose the parameter vector $\boldsymbol{\theta}$ is divided into two vectors $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ of lengths p_1, and p_2, respectively. We would like to test the hypothesis $H_o : \boldsymbol{\psi} = \boldsymbol{\psi}_o$. Here $\boldsymbol{\phi}$ is a nuisance parameter. Let $\hat{\boldsymbol{\phi}}(\boldsymbol{\psi}_o)$ be the maximum likelihood estimates of $\boldsymbol{\phi}$ obtained by maximizing the likelihood with respect to $\boldsymbol{\phi}$, with $\boldsymbol{\psi}$ fixed at $\boldsymbol{\psi}_o$. That is, $\hat{\boldsymbol{\phi}}(\boldsymbol{\psi}_o)$ maximizes $\ln L[(\boldsymbol{\psi}_o, \boldsymbol{\phi}) : \mathbf{Y}]$ with respect to $\boldsymbol{\phi}$. We also partition the information matrix $\mathbf{I}$ into

$$\mathbf{I} = \begin{pmatrix} \mathbf{I}_{\psi\psi} & \mathbf{I}_{\psi\phi} \\ \mathbf{I}_{\phi\psi} & \mathbf{I}_{\phi\phi} \end{pmatrix}, \tag{B.10}$$

where $\mathbf{I}_{\psi\psi}$ is of dimension $p_1 \times p_1$, $\mathbf{I}_{\phi\phi}$ is of dimension $p_2 \times p_2$, $\mathbf{I}_{\psi\phi}$ is $p_1 \times p_2$, and $\mathbf{I}^t_{\phi\psi} = \mathbf{I}_{\psi\phi}$. Notice that a partitioned information matrix has an inverse which is also a partitioned matrix with

$$\mathbf{I}^{-1} = \begin{pmatrix} \mathbf{I}^{\psi\psi} & \mathbf{I}^{\psi\phi} \\ \mathbf{I}^{\phi\psi} & \mathbf{I}^{\phi\phi} \end{pmatrix}, \tag{B.11}$$

With these refinements, the three statistics for testing $H_o : \psi = \psi_o$ are given by

Likelihood ratio test:

$$\chi^2_{LR} = -2\{\ln L[(\psi_o, \hat{\phi}(\psi_o) : \mathbf{Y})] - \ln L(\hat{\theta} : \mathbf{Y})\}, \tag{B.12}$$

Wald test:

$$\chi^2_W = (\hat{\psi} - \psi_o)[\mathbf{I}^{\psi\psi}(\hat{\psi}, \hat{\phi})]^{-1}(\hat{\psi} - \psi_o)^t, \tag{B.13}$$

and score test:

$$\chi^2_S = \mathbf{U}_\psi[\psi_o, \hat{\phi}(\psi_o)][\mathbf{I}^{\psi\psi}(\psi_o, \hat{\phi}(\psi_o))]\mathbf{U}^t[\psi_o, \hat{\phi}(\psi_o)]. \tag{B.14}$$

All three statistics have an asymptotic chi-squared distribution with p_1 degrees of freedom when the null hypothesis is true.

EXAMPLE B.2 Consider the problem of comparing two treatments, where the time to event in each group has an exponential distribution. For population one, we assume that the hazard rate is λ whereas for population two, we assume that the hazard rate is $\lambda\beta$. We shall test $H_o : \beta = 1$ treating λ as a nuisance parameter. The likelihood function is given by

$$L[(\lambda, \beta) : D_1, D_2, S_1, S_2] = \lambda^{D_1+D_2}\beta^{D_2}\exp(-\lambda S_1 - \lambda\beta S_2) \tag{B.15}$$

where D_i is the number of events and S_i is the total time on test in the ith sample, $i = 1, 2$. From (B.15),

$$\ln L[(\beta, \lambda)] = (D_1 + D_2)\ln\lambda + D_2\ln\beta - \lambda S_1 - \lambda\beta S_2, \tag{B.16}$$

$$U_\beta(\beta, \lambda) = \frac{\delta}{\delta\beta}\ln L(\beta, \lambda) = \frac{D_2}{\beta} - \lambda S_2, \tag{B.17}$$

$$U_\lambda(\beta, \lambda) = \frac{\delta}{\delta\lambda}\ln L(\beta, \lambda) = \frac{D_1 + D_2}{\lambda} - S_1 - \beta S_2, \tag{B.18}$$

$$I_{\beta\beta}(\beta, \lambda) = -\frac{\delta^2\ln L(\beta, \lambda)}{\delta\beta^2} = \frac{D_2}{\beta^2}, \tag{B.19}$$

$$I_{\lambda\lambda}(\beta, \lambda) = -\frac{\delta^2\ln L(\beta, \lambda)}{\delta\lambda^2} = \frac{D_1 + D_2}{\lambda^2}, \tag{B.20}$$

and

$$I_{\beta\lambda} = -\frac{\delta^2\ln L(\beta, \lambda)}{\delta\lambda\delta\beta} = S_2. \tag{B.21}$$

Solving the system of equations $U_\beta(\beta, \lambda) = 0$, $U_\lambda(\beta, \lambda) = 0$ yields the global maximum likelihood estimators $\hat{\beta} = S_1D_2/(S_2D_1)$ and $\hat{\lambda} = D_1/S_1$.

Solving $U_\lambda(\beta, \lambda) = 0$, for β fixed at its value under H_o, yields $\hat{\lambda}(\beta = 1)$, denoted by $\hat{\lambda}(1)$, $= (D_1 + D_2)/(S_1 + S_2)$. Thus, we have from B.12 a likelihood ratio test statistic of

$$\begin{aligned}\chi^2_{LR} &= -2\{[(D_1 + D_2)\ln[\hat{\lambda}(1)] - \hat{\lambda}(1)(S_1 + S_2)] \\ &\quad - [(D_1 + D_2)\ln[\hat{\lambda}] + D_2\ln[\hat{\beta}] - \hat{\lambda}S_1 - \hat{\lambda}\hat{\beta}S_2]\}. \\ &= 2D_1\ln\left[\frac{D_1(S_1 + S_2)}{S_1(D_1 + D_2)}\right] + 2D_2\ln\left[\frac{D_2(S_1 + S_2)}{S_2(D_1 + D_2)}\right].\end{aligned}$$

From B.19–B.21,

$$I^{\beta\beta}(\beta, \lambda) = \frac{\beta^2(D_1 + D_2)}{[D_2(D_1 + D_2) - (\lambda\beta S_2)^2]}$$

so, the Wald test is given by

$$\begin{aligned}\chi^2_W &= (\hat{\beta} - 1)^2\left\{\frac{\hat{\beta}^2(D_1 + D_2)}{[D_2(D_1 + D_2) - (\hat{\lambda}\hat{\beta}S_2)^2]}\right\}^{-1} \\ &= \frac{D_1^2(S_1D_2 - S_2D_1)^2}{D_2S_1^2(D_1 + D_2)}\end{aligned}$$

The score test is given by

$$\begin{aligned}\chi^2_S &= (D_2 - \hat{\lambda}(1)S_2)^2\frac{(D_1 + D_2)}{[D_2(D_1 + D_2) - (\hat{\lambda}(1)S_2)^2]} \\ &= \frac{[D_2(S_1 + S_2) - (D_1 + D_2)S_2]^2}{D_2(S_1 + S_2)^2 - (D_1 + D_2)S_2^2}\end{aligned}$$

If, for example, $D_1 = 10$, $D_2 = 12$, $S_1 = 25$, and $S_2 = 27$, then $\chi^2_{LR} = 0.0607$, $\chi^2_W = 0.0545$ and $\chi^2_S = 0.0448$, all nonsignificant when compared to a chi-square with one degree of freedom.

Solutions to Chapter 13

13.1 $T = 14.8$, $V = 107.5$, $Z = 1.4$, $p = 0.1539$. No evidence of random effect.

13.3 (a) Standard Cox model $b = -1.035$, SE $(b) = 0.44$, $p = 0.0187$.
(b) Gamma Frailty Model $b = -1.305$, SE $(b) = 0.528$, $p = 0.0133$.
Estimate of $\theta = 0.713$, SE $= 0.622$, Wald p-value of test of $\theta = 0 = 0.2517$, likelihood ratio p-value $= 0.1286$.

13.5 (a) See 13.1.
(b) Adjusted SE $= 0.3852$, test statistic $= -1.035/0.3852$, $p = .0072$.

Bibliography

Aalen, O. O. Statistical Inference for a Family of Counting Processes. Ph.D. Dissertation, University of California, Berkeley, 1975.

Aalen, O. O. Nonparametric Estimation of Partial Transition Probabilities in Multiple Decrement Models. *Annals of Statistics* 6 (1978a): 534–545.

Aalen, O. O. Nonparametric Inference for a Family of Counting Processes. *Annals of Statistics* 6 (1978b): 701–726.

Aalen, O. O. A Model for Non-Parametric Regression Analysis of Counting Processes. In *Lecture Notes on Mathematical Statistics and Probability*, 2, W. Klonecki, A. Kozek, and J. Rosiski, eds. New York: Springer-Verlag, 1980, pp. 1–25.

Aalen, O. O. Heterogeneity in Survival Analysis. *Statistics in Medicine* (1988): 1121–1137.

Aalen, O. O. A Linear Regression Model for the Analysis of Lifetimes. *Statistics in Medicine* 8 (1989): 907–925.

Aalen, O. O. Modeling Heterogeneity in Survival Analysis by the Compound Poisson Distribution. *Annals of Applied Probability* 2 (1992): 951–972.

Aalen, O. O. Further Results on the Nonparametric Linear Regression Model in Survival Analysis. *Statistics in Medicine* 12 (1993): 1569–1588.

Aalen, O. O. and Johansen, S. An Empirical Transition Matrix for Nonhomogeneous Markov Chains Based on Censored Observations. *Scandinavian Journal of Statistics* 5 (1978): 141–150.

Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. In *2nd International Symposium of Information Theory and Control*, E. B. N. Petrov and F. Csaki, eds. Akademia Kiado, Budapest, pp. 267–281.

American Joint Committee for Staging and End-Result Reporting. *Manual for Staging of Cancer*, 1972.

Andersen, P. K. Testing Goodness of Fit of Cox's Regression and Life Model. *Biometrics* 38 (1982): 67–77. Correction: 40 (1984): 1217.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. Linear Nonparametric Tests for Comparison of Counting Processes, with Application to Censored Survival Data (with Discussion). *International Statistical Review*, 50 (1982): 219–258. Amendment: 52 (1984): 225.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. Censoring, Truncation and Filtering in Statistical Models Based on Counting Processes. *Contemporary Mathematics* 80 (1988): 19–60.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. *Statistical Models Based on Counting Processes*. New York: Springer-Verlag, 1993.

Andersen, P. K., Bentzon, M. W., and Klein, J. P. Estimating the Survival Function in the Proportional Hazards Regression Model: A Study of the Small Sample Size Properties. *Scandinavian Journal of Statistics* 23 (1996): 1–12.

Andersen, P. K., Klein, J. P., Knudsen, K. M., and Tabanera-Palacios, R. Estimation of the Variance in a Cox's Regression Model with Shared Frailties. *Biometrics* 53 (1997): 1475–1484.

Andersen, P. K. and Væth, M. Simple Parametric and Nonparametric Models for Excess and Relative Mortality. *Biometrics* 45 (1989): 523–535.

Arjas, E. A Graphical Method for Assessing Goodness of Fit in Cox's Proportional Hazards Model. *Journal of the American Statistical Association* 83 (1988) 204–212.

Arjas, E. A. and Haara, P. A Note on the Exponentiality of Total Hazards Before Failure. *Journal of Multivariate Analysis* 26 (1988): 207–218.

Avalos, B. R., Klein, J. L., Kapoor, N., Tutschka, P. J., Klein, J. P., and Copelan, E. A. Preparation for Marrow Transplantation in Hodgkin's and Non-Hodgkin's Lymphoma Using Bu/Cy. *Bone Marrow Transplantation* 13 (1993): 133–138.

Barlow, R. E. and Campo, R. Total Time on Test Processes and Application to Failure Data Analysis. In *Reliability and Fault Tree Analysis*, R. E. Barlow, J. Fussell, and N. D. Singpurwalla, eds. SIAM, Philadelphia, (1975) pp. 451–481.

Batchelor, J. R. and Hackett, M. HLA Matching in the Treatment of Burned Patients with Skin Allographs. *Lancet* 2 (1970): 581–583.

Beadle, G. F., Come, S., Henderson, C., Silver, B., and Hellman, S. A. H. The Effect of Adjuvant Chemotherapy on the Cosmetic Results after Primary Radiation Treatment for Early Stage Breast Cancer. *International Journal of Radiation Oncology, Biology and Physics* 10 (1984a): 2131–2137.

Beadle, G. F., Harris, J. R., Silver, B., Botnick, L., and Hellman, S. A. H. Cosmetic Results Following Primary Radiation Therapy for Early Breast Cancer. *Cancer* 54 (1984b): 2911–2918.

Berman, S. M. Note on Extreme Values, Competing Risks and Semi-Markov Processes. *The Annals of Mathematical Statistics* 34 (1963): 1104–1106.

Berretoni, J. N. Practical Applications of the Weibull Distribution. *Industrial Quality Control* 21 (1964): 71–79.

Beyer, W. H. *CRC Handbook of Tables for Probability and Statistics*. Boca Raton, Florida: CRC Press, 1968.

Bie, O., Borgan, Ø., and Liestøl, K. Confidence Intervals and Confidence Bands for the Cumulative Hazard Rate Function and Their Small Sample Properties. *Scandinavian Journal of Statistics* 14 (1987): 221–233.

Billingsley, P. *Convergence of Probability Measures*. New York: John Wiley and Sons, 1968.

Borgan, Ø. and Liestøl, K. A Note on Confidence Intervals and Bands for the Survival Curve Based on Transformations. *Scandinavian Journal of Statistics* 17 (1990): 35–41.

Breslow, N. E. A Generalized Kruskal–Wallis Test for Comparing K Samples Subject to Unequal Patterns of Censorship. *Biometrika* 57 (1970): 579–594.

Breslow, N. E. Covariance Analysis of Censored Survival Data. *Biometrics* 30 (1974): 89–99.

Breslow, N. E. Analysis of Survival Data under the Proportional Hazards Model. *International Statistics Review* 43 (1975): 45–58.

Brookmeyer, R. and Crowley, J. J. A Confidence Interval for the Median Survival Time. *Biometrics* 38 (1982a): 29–41.

Brookmeyer, R. and Crowley, J. J. A K-Sample Median Test for Censored Data. *Journal of the American Statistical Association* 77 (1982b): 433–440.

Brown, J. B. W., Hollander, M., and Korwar, R. M. Nonparametric Tests of Independence for Censored Data, with Applications to Heart Transplant Studies. In *Reliability and Biometry: Statistical Analysis of Lifelength*, F. Proschan and R. J. Serfling, eds. Philadelphia: SIAM, 1974, pp. 327–354.

Buckley, J. D. Additive and Multiplicative Models for Relative Survival Rates. *Biometrics* 40 (1984): 51–62.

Chiang, C. L. *Introduction to Stochastic Processes in Biostatistics*. New York: John Wiley and Sons, 1968.

Chiang, C. L. *The Lifetable and its Applications*. Malabar, Florida: Krieger, 1984.

Christensen, E., Schlichting, P., Andersen, P. K., Fauerholdt, L., Schou, G., Pedersen, B. V., Juhl, E., Poulsen, H., and Tygstrup, N. Updating Prognosis and Therapeutic Effect Evaluation in Cirrhosis Using Cox's Multiple Regression Model for Time-Dependent Variables. *Scandinavian Journal of Gastroenterology* 21 (1986): 163–174.

Chung, C. F. Formulae for Probabilities Associated with Wiener and Brownian Bridge Processes. Technical Report 79, Laboratory for Research in Statistics and Probability, Ottawa, Canada: Carleton University, 1986.

Clausius, R. Ueber Die Mittlere Lange Der Wege. *Ann. Phys. Lpzg* 105 (1858): 239–58.

Clayton, D. G. A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika* 65 (1978): 141–151.

Clayton, D. G. and Cuzick, J. Multivariate Generalizations of the Proportional Hazards Model (with Discussion). *Journal of the Royal Statistical Society A* 148 (1985): 82–117.

Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatter Plots. *Journal of the American Statistical Association*, 74 (1979): 829–836.

Collett, D. *Modeling Survival Data in Medical Research*. New York: Chapman and Hall, 1994.

Commenges, D. and Andersen, P. K. Score Test of Homogeneity for Survival Data. *Lifetime Data Analysis* 1 (1995): 145–160.

Contal, C. and O'Quigley, J. An Application of Change Point Methods in Studying the Effect of Age on Survival in Breast Cancer. *Computational Statistics and Data Analysis* 30 (1999): 253–270.

Copelan, E. A., Biggs, J. C., Thompson, J. M., Crilley, P., Szer, J., Klein, J. P., Kapoor, N., Avalos, B. R., Cunningham, I., Atkinson, K., Downs, K., Harmon, G. S., Daly, M. B., Brodsky, I., Bulova, S. I., and Tutschka, P. J. Treatment for Acute Myelocytic Leukemia with Allogeneic Bone Marrow Transplantation Following Preparation with Bu/Cy. *Blood* 78 (1991): 838–843.

Cornfield, J. A. and Detre, K. Bayesian Life Table Analysis. *Journal of the Royal Statistical Society Series B* 39 (1977): 86–94.

Costigan, T. M. and Klein. J. P. Multivariate Survival Analysis Based On Frailty Models, A. P. Basu, ed. *Advances In Reliability*, New York: North Holland, pp. 43–58.

Cox, D. R. The Analysis of Exponentially Distributed Lifetimes with Two Types of Failure. *The Journal of the Royal Statistical Society* B21 (1959): 411–421.

Cox, D. R. *Renewal Theory*. London: Methune, 1962.

Cox, D. R. Regression Models and Life Tables (with Discussion). *Journal of the Royal Statistical Society B* 34 (1972): 187–220.

Cox, D. R. and Hinkley, D. V. *Theoretical Statistics*. New York: Chapman and Hall, 1974.

Cox, D. R. and Oakes, D. *Analysis of Survival Data.* New York: Chapman and Hall, 1984.
Cox, D. R. and Snell, E. J. A General Definition of Residuals (with Discussion). *Journal of the Royal Statistical Society B* 30 (1968): 248–275.
Crowley, J. J. and Storer, B. E. Comment On 'A Reanalysis of the Stanford Heart Transplant Data', by M. Aitkin, N. Laird, and B. Francis. *Journal of the American Statistical Association,* 78 (1983): 277–281.
Cutler, S. J. and Ederer, F. Maximum Utilization of the Life Table Method in Analyzing Survival. *Journal of Chronic Diseases* 8 (1958): 699–712.
Dabrowska, D. M., Doksum, K. A., and Song, J. K. Graphical Comparison of Cumulative Hazards for Two Populations. *Biometrika* 76 (1989): 763–773.
David, H. A. *Order Statistics.* New York: John Wiley and Sons, 1981.
David, H. A., and Moeschberger, M. L. *The Theory of Competing Risks.* London: Charles Griffin, 1978.
Davis, D. J. An Analysis of Some Failure Data. *Journal of the American Statistical Association* 47 (1952): 113–150.
Doll, R. The Age Distribution of Cancer: Implications for Models of Carcinogens. *Journal of the Royal Statistical Society, Series A* 134 (1971): 133–66.
Efron, B. The Two Sample Problem with Censored Data. In *Proceedings of the Fifth Berkeley Symposium On Mathematical Statistics and Probability.* New York: Prentice-Hall, (1967): 4, 831-853.
Efron, B. The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association* 72 (1977): 557–565.
Elandt–Johnson, R. C. and Johnson, N. L. *Survival Models and Data Analysis.* New York: John Wiley and Sons, 1980.
Epstein, B. The Exponential Distribution and Its Role in Life Testing. *Industrial Quality Control* 15 (1958): 2–7.
Epstein, B. and Sobel, M. Some Theorems Relevant to Life Testing from an Exponential Distribution. *Annals of Mathematical Statistics* 25 (1954): 373–81.
Escobar, L. A. and Meeker, W. Q. Assessing Influence in Regression Analysis with Censored Data. *Biometrics* 48 (1992): 507–528.
Feigl, P. and Zelen, M. Estimation of Exponential Survival Probabilities with Concomitant Information. *Biometrics* 21 (1965): 826–838.
Feinleib, M. A Method of Analyzing Log Normally Distributed Survival Data with Incomplete Follow-Up. *Journal of the American Statistical Association,* 55 (1960): 534–545.
Ferguson, T. S. A Bayesian Analysis of Some Nonparametric Problems. *Annals of Statistics* 1 (1973): 209–230.
Ferguson, T. S. and Phadia, E. G. Bayesian Nonparametric Estimation Based on Censored Data. *Annals of Statistics* 7 (1979): 163–186.
Finkelstein, D. M. A Proportional Hazards Model for Interval-Censored Failure Time Data. *Biometrics* 42 (1986): 845–854.
Finkelstein, D. M. and Wolfe, R. A. A Semiparametric Model for Regression Analysis of Interval-Censored Failure Time Data. *Biometrics* 41 (1985): 933–945.
Fleming, T. R. and Harrington, D. P. A Class of Hypothesis Tests for One and Two Samples of Censored Survival Data. *Communications In Statistics* 10 (1981): 763–794.
Fleming, T. R. and Harrington, D. P. *Counting Processes and Survival Analysis.* New York: John Wiley and Sons, 1991.
Fleming, T. R., Harrington, D. P., and O'Sullivan, M. Supremum Versions of the Log-Rank and Generalized Wilcoxon Statistics. *Journal of the American Statistical Association* 82 (1987): 312–320.

Fleming, T. R., O'Fallon, J. R., O'Brien, P. C., and Harrington, D. P. Modified Kolmogorov-Smirnov Test Procedures with Application to Arbitrarily Right Censored Data. *Biometrics* 36 (1980): 607–626.
Freireich, E. J., Gehan, E., Frei, E., Schroeder, L. R., Wolman, I. J., Anbari, R., Burgert, E. O., Mills, S. D., Pinkel, D., Selawry, O. S., Moon, J. H., Gendel, B. R., Spurr, C. L., Storrs, R., Haurani, F., Hoogstraten, B., and Lee, S. The Effect of 6-Mercaptopurine on the Duration of Steroid-Induced Remissions in Acute Leukemia: A Model for Evaluation of Other Potentially Useful Therapy. *Blood* 21 (1963): 699–716.
Galambos, J. Exponential Distribution. In *Encyclopedia of Statistical Science,* N. L. Johnson and S. Kotz, eds. New York: John Wiley and Sons, Vol. 2, pp. 582–587.
Galambos, J. and Kotz, S. *Characterizations of Probability Distributions. Lecture Notes in Mathematics* 675. Heidelberg: Springer-Verlag, 1978.
Gasser, T. and Müller, H. G. Kernel Estimation of Regression Functions. In *Smoothing Techniques for Curve Estimation, Lecture Notes in Mathematics* 757. Berlin: Springer-Verlag, 1979, pp. 23–68.
Gastrointestinal Tumor Study Group. A Comparison of Combination and Combined Modality Therapy for Locally Advanced Gastric Carcinoma. *Cancer* 49 (1982): 1771–1777.
Gatsonis, C., Hsieh, H. K., and Korway, R. Simple Nonparametric Tests for a Known Standard Survival Based on Censored Data. *Communications in Statistics—Theory and Methods* 14 (1985): 2137–2162.
Gehan, E. A. A Generalized Wilcoxon Test for Comparing Arbitrarily Singly Censored Samples. *Biometrika* 52 (1965): 203–223.
Gehan, E. A. and Siddiqui, M. M. Simple Regression Methods for Survival Time Studies. *Journal of the American Statistical Association* 68 (1973): 848–856.
Gelfand, A. E. and Smith, A. F. M. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* 85 (1990): 398–409.
Gill, R. D. Censoring and Stochastic Integrals. *Mathematical Centre Tracts.* Amsterdam: Mathematisch Centrum, 1980, 124.
Gill, R. D. Discussion of the Paper by D. Clayton and J. Cuzick. *Journal of the Royal Statistical Society A* 148 (1985): 108–109.
Gill, R. D. and Schumacher, M. A Simple Test of the Proportional Hazards Assumption. *Biometrika* 74 (1987): 289–300.
Gomez, G., Julia, O., and Utzet, F. Survival Analysis for Left Censored Data. In *Survival Analysis: State of the Art,* J. P. Klein and P. Goel, eds. Boston: Kluwer Academic Publishers, 1992, pp. 269–288.
Gompertz, B. On the Nature of the Function Expressive of the Law of Human Mortality and on the New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London* 115 (1825): 513–585.
Gooley, T. A., Leisenring, W., Crowley, J., and Storer, B. Estimation of Failure Probabilities in the Presence of Competing Risks: New Representations of Old Estimators, *Statistics in Medicine* 18 (1999): 695–706.
Greenwood, M. The Natural Duration of Cancer. In *Reports On Public Health and Medical Subjects* 33. London: His Majesty's Stationery Office, 1926, pp. 1–26.
Gross, S. and Huber–Carol, C. Regression Analysis for Discrete and Continuous Truncated and Eventually Censored Data. In *Survival Analysis: State of the Art,* J. P. Klein and P. Goel, eds. Boston: Kluwer Academic Publishers, pp. 289–308.
Guerts, J. H. L. On the Small Sample Performance of Efron's and Gill's Version of the Product Limit Estimator Under Proportional Hazards. *Biometrics* 43 (1987): 683–692.
Gumbel, E. J. *Statistics of Extremes.* New York: Columbia University Press, 1958.

Hall, W. J. and Wellner, J. A. Confidence Bands for a Survival Curve from Censored Data. *Biometrika* 67 (1980): 133–143.

Hamburg, B. A., Kraemer, H. C., and Jahnke, W. A Hierarchy of Drug Use in Adolescence Behavioral and Attitudinal Correlates of Substantial Drug Use. *American Journal of Psychiatry* 132 (1975): 1155–1163.

Harrington, D. P. and Fleming, T. R. A Class of Rank Test Procedures for Censored Survival Data. *Biometrika* 69 (1982): 133–143.

Heckman, J. J. and Honore, B. E. The Identifiability of the Competing Risks Model, *Biometrika* 76 (1989): 325–330.

Hjort, N. L. Nonparametric Bayes Estimators Based on Beta Processes in Models for Life History Data. *Annals of Statistics* 18 (1990): 1259–1294.

Hjort, N. L. Semiparametric Estimation of Parametric Hazard Rates. In *Survival Analysis: State of the Art*, J. P. Klein and P. Goel, eds. Boston: Kluwer Academic Publishers, 1992, pp. 211–236.

Hoel, D. G. and Walburg, H. E. Survival Analysis of Survival Experiments, *Journal of the National Cancer Institute* 49 (1972): 361–372.

Horner, R. D. Age at Onset of Alzheimer's Disease: Clue to the Relative Importance of Etiologic Factors? *American Journal of Epidemiology*, 126 (1987): 409–414.

Hougaard, P. A Class of Multivariate Failure Time Distributions. *Biometrika* 73 (1986a): 671–678.

Hougaard, P. Survival Models for Heterogeneous Populations Derived from Stable Distributions. *Biometrika* 73 (1986b): 387–396.

Howell, A. *A SAS Macro for the Additive Regression Hazards Model.* Master's Thesis, Medical College of Wisconsin, Milwaukee, Wisconsin, 1996.

Huffer, F. W. and McKeague, I. W. Weighted Least Squares Estimation for Aalen's Additive Risk Model. *Journal of the American Statistical Association* 86 (1991): 114–129.

Hyde, J. Testing Survival under Right Censoring and Left Truncation. *Biometrika* 64 (1977): 225–230.

Hyde, J. Survival Analysis with Incomplete Observations. In *Biostatistics Casebook*, R. G. Miller, B. Efron, B. W. Brown, and L. E. Moses, eds. New York: John Wiley and Sons, 1980, pp. 31–46.

Ichida, J. M., Wassell, J. T., Keller, M. D., and Ayers, L. W. Evaluation of Protocol Change in Burn-Care Management Using the Cox Proportional Hazards Model with Time-Dependent Covariates. *Statistics in Medicine* 12 (1993): 301–310.

Izenman, A. J. Recent Developments in Nonparametric Density Estimation. *Journal of the American Statistical Association* 86 (1991): 205–224.

Jesperson, N. C. B. *Discretizing a Continuous Covariate in the Cox Regression Model*, Research Report 86/2, Statistical Research Unit, University of Copenhagen, 1986.

Johansen, S. An Extension of Cox's Regression Model. *International Statistical Review* 51 (1983): 258–262.

Johnson, N. L. and Kotz, S. *Distributions in Statistics: Continuous Multivariate Distributions.* New York: John Wiley and Sons, 1970.

Johnson, W. and Christensen, R. Bayesian Nonparametric Survival Analysis for Grouped Data. *Canadian Journal of Statistics* 14, (1986): 307–314.

Kalbfleisch, J. D. and Prentice, R. L. Marginal Likelihoods Based on Cox's Regression and Life Model. *Biometrika* 60 (1973): 267–278.

Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data.* New York: John Wiley and Sons, 1980.

Kao, J. H. K. A Graphical Estimation of Mixed Weibull Parameters in Life-Testing Electron Tubes. *Technometrics* 1 (1959): 389–407.

Kaplan, E. L. and Meier, P. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* 53 (1958): 457–481.

Kardaun, O. Statistical Analysis of Male Larynx-Cancer Patients—A Case Study. *Statistica Nederlandica* 37 (1983): 103–126.

Kay, R. Goodness-of-Fit Methods for the Proportional Hazards Model: A Review. *Reviews of Epidemiology Santé Publications* 32 (1984): 185–198.

Keiding, N. Statistical Inference in the Lexis Diagram. *Philosophical Transactions of the Royal Society of London A* 332 (1990): 487–509.

Keiding, N. Independent Delayed Entry. In *Survival Analysis: State of the Art*, J. P. Klein and P. Goel, eds. Boston: Kluwer Academic Publishers, 1992, pp. 309–326.

Keiding, N. and Gill, R. D. Random Truncation Models and Markov Processes. *Annals of Statistics* 18 (1990): 582–602.

Kellerer, A. M. and Chmelevsky, D. Small-Sample Properties of Censored-Data Rank Tests. *Biometrics* 39 (1983): 675–682.

Klein, J. P. Small-Sample Moments of Some Estimators of the Variance of the Kaplan–Meier and Nelson–Aalen Estimators. *Scandinavian Journal of Statistics* 18 (1991): 333–340.

Klein, J. P. Semiparametric Estimation of Random Effects Using the Cox Model Based on the EM Algorithm. *Biometrics* 48 (1992): 795–806.

Klein, J. P. and Moeschberger, M. L. The Asymptotic Bias of the Product-Limit Estimator Under Dependent Competing Risks. *Indian Journal of Productivity, Reliability and Quality Control* 9 (1984): 1–7.

Klein, J. P. and Moeschberger, M. L. Bounds on Net Survival Probabilities for Dependent Competing Risks. *Biometrics* 44 (1988): 529–538.

Klein, J. P., Keiding, N., and Copelan, E. A. Plotting Summary Predictions in Multistate Survival Models: Probability of Relapse and Death in Remission for Bone Marrow Transplant Patients. *Statistics in Medicine* 12 (1994): 2315–2332.

Klein, J. P., Keiding, N., and Kreiner, S. Graphical Models for Panel Studies, Illustrated on Data from the Framingham Heart Study. *Statistics in Medicine* 14 (1995): 1265–1290.

Klein, J. P., Moeschberger, M. L., Li, Y. H. and Wang, S.T. Estimating Random Effects in the Framingham Heart Study. In *Survival Analysis: State of the Art*, J. P. Klein and P. Goel, eds. Boston: Kluwer Academic Publishers, 1992, pp. 99–120.

Koziol, J. A. A Two-Sample Cramér–Von Mises Test for Randomly Censored Data. *Biometrical Journal* 20 (1978): 603–608.

Kuo, L. and Smith, A. F. M. Bayesian Computations in Survival Models via the Gibbs Sampler. In *Survival Analysis: State of the Art*, J. P. Klein and P. Goel, eds. Boston: Kluwer Academic Publishers, 1992, pp. 11–24.

Lagakos, S. W. The Graphical Evaluation of Explanatory Variables in Proportional Hazards Regression Models. *Biometrika* 68 (1981): 93–98.

Lagakos, S. W., Barraj, L. M., and Degruttola, V. Nonparametric Analysis of Truncated Survival Data, with Application to AIDS. *Biometrika* 75 (1988): 515–523.

Lai, T. L. and Ying, Z. Estimating a Distribution Function with Truncated and Censored Data. *Annals of Statistics* 19 (1991): 417–442.

Latta, R. B. A Monte Carlo Study of Some Two-Sample Rank Tests with Censored Data. *Journal of the American Statistical Association* 76 (1981): 713–719.

Lee, E. W., Wei, L. J., and Amato, D. A. Cox-Type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations. In *Survival Analysis: State of the Art*, J. P. Klein and P. Goel, eds. Boston: Kluwer Academic Publishers, 1992, pp. 237–248.

Lee, L. and Thompson, W. A., Jr. Results on Failure Time and Pattern for the Series System. In *Reliability and Biometry: Statistical Analysis of Lifelength*, F. Proshan and R. J. Serfling, eds. Philadelphia: SIAM, 1974, pp. 291-302.

Lee, P. N. and O'Neill, J. A. The Effect Both of Time and Dose Applied on Tumor Incidence Rate in Benzopyrene Skin Painting Experiments. *British Journal of Cancer* 25 (1971): 759–70.

Lee, S. and Klein, J. P. Bivariate Models with a Random Environmental Factor. *Indian Journal of Productivity, Reliability and Quality Control* 13 (1988): 1–18.

Li, Y., Klein, J. P., and Moeschberger, M. L. Effects of Model Misspecification in Estimating Covariate Effects in Survival Analysis for a Small Sample Size. *Computational Statistics and Data Analysis* 22 (1996): 177–192.

Liang, K. -Y., Self, S. G., and Chang, Y. -C. Modeling Marginal Hazards in Multivariate Failure-Time Data. *Journal of Royal Statistical Society* B 55; 441–463, 1993.

Lieblein, J. and Zelen, M. Statistical Investigation of the Fatigue Life of Deep-Groove Ball Bearings. *Journal of Research, National Bureau of Standards* 57 (1956): 273–316.

Lin, D. Y. MULCOX2: A General Program for the Cox Regression Analysis of Multiple Failure Time Data. *Computers in Biomedicine* 40 (1993): 279–293.

Lin, D. Y. and Wei, L. J. Robust Inference for the Cox Proportional Hazards Model. *Journal of the American Statistical Association* 84 (1989): 1074–1078.

Lin, D. Y. and Ying, Z. Semiparametric Analysis of the Additive Risk Model. *Biometrika* 81 (1994): 61–71.

Lin, D. Y. and Ying, Z. Semiparametric Analysis of General Additive-Multiplicative Hazard Models for Counting Processes. *Annals of Statistics* 23 (1995): 1712–1734.

Lin, D. Y. and Ying, Z. Additive Regression Models for Survival Data. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, D. Y. Lin and T. R. Fleming, eds. New York: Springer, 1997, pp. 185–198.

Lindley, D. V. and Singpurwalla N. A. Multivariate Distributions for the Reliability of a System of Components Sharing a Common Environment. *Journal of Applied Probability* 23 (1986): 418–431.

Makeham, W.M. On the Law of Mortality and the Construction of Annuity Tables. *Journal of the Institute of Actuaries*, 8 (1860): 301–310.

Mantel, N., Bohidar, N. R., and Ciminera, J. L. Mantel–Haenszel Analysis of Litter-Matched Time-To-Response Data, with Modifications for Recovery of Interlitter Information. *Cancer Research* 37 (1977): 3863–3868.

Marquardt, D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal of Applied Mathematics* 11 (1963): 431–441.

Matthews, D. E. and Farewell, V. T. On Testing for a Constant Hazard Against a Change-Point Alternative (Corr. v41, 1103). *Biometrics* 38 (1982): 463–468.

McCarthy, D. J., Harman, J. G., Grassanovich, J. L., Qian, C., and Klein, J. P. Combination Drug Therapy of Seropositive Rheumatoid Arthritis. *The Journal of Rheumatology* 22 (1995): 1636–1645.

McCullagh, P. and Nelder, J. A. *Generalized Linear Models*, 2nd Ed. London: Chapman and Hall, 1989.

McGilchrist, C. A. and Aisbett, C. W. Regression with Frailty in Survival Analysis. *Biometrics* 47 (1991): 461–466.

McKeague, I. W. Asymptotic Theory for Weighted Least-Squares Estimators in Aalen's Additive Risk Model. *Contemporary Mathematics* 80 (1988): 139–152.

Miller, R. G. and Siegmund, D. Maximally Selected Chi-Square Statistics. *Biometrics* 38 (1982): 1011–1016.

Moeschberger, M. L and Klein, J. P. A Comparison of Several Methods of Estimating the Survival Function When There is Extreme Right Censoring. *Biometrics* 41 (1985): 253–259.

Morsing, T. *Competing Risks in Cross-Over Designs*. Technical Report, Department of Mathematics, Chalmers University of Technology, Goteborg, 1994.

Nahman, N. S., Middendorf, D. F., Bay, W. H., McElligott, R., Powell, S., and Anderson, J. Modification of the Percutaneous Approach to Peritoneal Dialysis Catheter Placement Under Peritoneoscopic Visualization: Clinical Results in 78 Patients. *Journal of The American Society of Nephrology* 3 (1992): 103–107.

Nair, V. N. Confidence Bands for Survival Functions with Censored Data: A Comparative Study. *Technometrics* 14 (1984): 265–275.

National Longitudinal Survey of Youth. *NLS Handbook*. Center for Human Resource Research. The Ohio State University, Columbus, Ohio, 1995.

Nelson, W. Theory and Applications of Hazard Plotting for Censored Failure Data. *Technometrics* 14 (1972): 945–965.

Nelson, W. *Applied Life Data Analysis*. New York: John Wiley and Sons, 1982.

Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sørensen, T. I. A. A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models. *Scandinavian Journal of Statistics* 19 (1992): 25–43.

Odell, P. M., Anderson, K. M., and D'Agostino, R. B. Maximum Likelihood Estimation for Interval-Censored Data Using a Weibull-Based Accelerated Failure Time Model. *Biometrics* 48 (1992): 951–959.

Peace, K. E. and Flora, R. E. Size and Power Assessment of Tests of Hypotheses on Survival Parameters. *Journal of the American Statistical Association* 73 (1978): 129–132.

Pepe, M. S. Inference for Events with Dependent Risks in Multiple Endpoint Studies. *Journal of the American Statistical Association* 86 (1991): 770–778.

Pepe, M. S. and Fleming, T. R. Weighted Kaplan–Meier Statistics: A Class of Distance Tests for Censored Survival Data. *Biometrics* 45 (1989): 497–507.

Pepe, M. S. and Fleming, T. R. Weighted Kaplan–Meier Statistics: Large Sample and Optimality Considerations. *Journal of the Royal Statistical Society B* 53 (1991): 341–352.

Pepe, M. S. and Mori, M. Kaplan-Meier, Marginal or Conditional Probability Curves in Summarizing Competing Risks Failure Time Data? *Statistics in Medicine* 12 (1993): 737–751.

Pepe, M. S., Longton, G. Pettinger, Mori, M., Fisher, L. D., and Storb, R. Summarizing Data on Survival, Relapse, and Chronic Graft-Versus-Host Disease After Bone Marrow Transplantation: Motivation for and Description of New Methods. *British Journal of Haematology* 83 (1993): 602–607.

Peterson, A. V., Jr. Bounds for a Joint Distribution Function with Fixed Sub-Distribution Functions: Application to Competing Risks. *Proceedings of the National Academy of Sciences* 73 (1976): 11–13.

Peto, R. and Lee, P. N. Weibull Distributions for Continuous-Carcinogenesis Experiments. *Biometrics*, 29 (1973): 457–470.

Peto, R. and Peto, J. Asymptotically Efficient Rank Invariant Test Procedures (with Discussion). *Journal of the Royal Statistical Society A* 135 (1972): 185–206.

Peto, R. and Pike, M. C. Conservatism of the Approximation $\Sigma(O - E)^2/E$ in the Log Rank Test for Survival Data or Tumor Incidence Data. *Biometrics* 29 (1973): 579–584.

Pike, M. C. A Method of Analysis of a Certain Class of Experiments in Carcinogenesis. *Biometrics* 22 (1966): 142–161.

Prentice, R. L. and Marek, P. A Qualitative Discrepancy Between Censored Data Rank Tests. *Biometrics* 35 (1979): 861–867.

Qian, C. Time-Dependent Covariates in a General Survival Model with Any Finite Number of Intermediate and Final Events. Unpublished Doctoral Dissertation, The Ohio State University, Columbus, Ohio, 1995.

Ramlau–Hansen, H. The Choice of a Kernel Function in the Graduation of Counting Process Intensities. *Scandinavian Actuarial Journal* (1983a): 165–182.

Ramlau–Hansen, H. Smoothing Counting Process Intensities by Means of Kernel Functions. *Annals of Statistics* 11 (1983b): 453–466.

Rosen, P. and Rammler, B. The Laws Governing the Fineness of Powdered Coal. *Journal of Inst. Fuels* 6 (1933): 29–36.

Sacher, G. A. On the Statistical Nature of Mortality with Special References to Chronic Radiation Mortality. *Radiation* 67 (1956): 250–257.

Schoenfeld, D. Partial Residuals for the Proportional Hazards Regression Model. *Biometrika* 69 (1982): 239–241.

Schumacher, M. Two-Sample Tests of Cramér–Von Mises and Kolmogorov–Smirnov Type for Randomly Censored Data. *International Statistical Review* 52 (1984): 263–281.

Sedmak, D. D., Meineke, T. A., Knechtges, D. S., and Anderson, J. Prognostic Significance of Cytokeratin-Positive Breast Cancer Metastases. *Modern Pathology* 2 (1989): 516–520.

Sheps, M. C. Characteristics of a Ratio Used to Estimate Failure Rates: Occurrences Per Person Year of Exposure. *Biometrics* 22 (1966): 310–321.

Sickle-Santanello, B. J., Farrar, W. B., Keyhani-Rofagha, S., Klein, J. P., Pearl, D., Laufman, H., Dobson, J., and O'Toole, R. V. A Reproducible System of Flow Cytometric DNA Analysis of Paraffin Embedded Solid Tumors: Technical Improvements and Statistical Analysis. *Cytometry* 9 (1988): 594–599.

Slud, E. V. Nonparametric Identifiability of Marginal Survival Distributions in the Presence of Dependent Competing Risks and a Prognostic Covariate. In *Survival Analysis: State of the Art*, J. P. Klein and P. Goel, eds. Boston: Kluwer Academic Publishers, 1992, pp. 355–368.

Slud, E. V. and Rubinstein, L. V. Dependent Competing Risks and Summary Survival Curves. *Biometrika* 70 (1983): 643–649.

Smith, R. M. and Bain, L. J. An Exponential Power Life-Testing Distribution. *Communications in Statistics-Theory and Methods* 4 (1975): 469–481.

Stablein, D. M. and Koutrouvelis, I. A. A Two-Sample Test Sensitive to Crossing Hazards in Uncensored and Singly Censored Data. *Biometrics* 41 (1985): 643–652.

Storer, B. E. and Crowley, J. J. A Diagnostic for Cox Regression and General Conditional Likelihoods. *Journal of the American Statistical Association* 80 (1985): 139–147.

Susarla, V. and Van Ryzin, J. Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations. *Journal of the American Statistical Association* 61 (1976): 897–902.

Tarone, R. E. and Ware, J. H. On Distribution-Free Tests for Equality for Survival Distributions. *Biometrika* 64 (1977): 156–160.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. Martingale-Based Residuals for Survival Models. *Biometrika* 77 (1990): 147–160.

Thisted, R. A. *Elements of Statistical Computing*. New York: Chapman and Hall, 1988.

Thompson, W. A., Jr. On the Treatment of Grouped Observations in Life Studies. *Biometrics* 33 (1977): 463–470.

Thomsen, B. L. A Note on the Modeling of Continuous Covariates in Cox's Regression Model. Research Report 88/5, Statistical Research Unit, University of Copenhagen, 1988.

Tsai, W. -Y. Testing The Assumption of Independence of Truncation Time and Failure Time. *Biometrika* 77 (1990): 169–177.

Tsiatis, A. A Nonidentifiability Aspect of the Problem of Competing Risks. *Proceedings of the National Academy of Sciences* 72 (1975): 20–22.

Tsuang, M. T. and Woolson, R. F. Mortality in Patients with Schizophrenia, Mania and Depression. *British Journal of Psychiatry*, 130 (1977): 162–166.

Turnbull, B. W. Nonparametric Estimation of a Survivorship Function with Doubly Censored Data. *Journal of the American Statistical Association* 69 (1974): 169–173.

Turnbull, B. W. The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *Journal of the Royal Statistical Society B* 38 (1976): 290–295.

Turnbull, B. W. and Weiss, L. A Likelihood Ratio Statistic for Testing Goodness of Fit with Randomly Censored Data. *Biometrics* 34 (1978): 367–375.

U.S. Department of Health and Human Services. Vital Statistics of the United States, 1959.

U.S. Department of Health and Human Services. Vital Statistics of the United States, 1990.

Wagner, S. S. and Altmann, S. A. What Time Do the Baboons Come Down from the Trees? (An Estimation Problem). *Biometrics* 29 (1973): 623–635.

Wang, S. T., Klein, J. P., and Moeschberger, M. L. Semiparametric Estimation of Covariate Effects Using the Positive Stable Frailty Model. *Applied Stochastic Models and Data Analysis* 11 (1995): 121–133.

Ware, J. H. and DeMets, D. L. Reanalysis of Some Baboon Descent Data. *Biometrics* 32 (1976): 459–463.

Wei, L. J., Lin, D. Y., and Weissfeld, L. Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association* 84 (1989): 1065–1073.

Weibull, W. A Statistical Theory of the Strength of Materials. *Ingeniors Vetenskaps Akakemien Handlingar* 151 (1939): 293–297.

Weibull, W. A Statistical Distribution of Wide Applicability. *Journal of Applied Mechanics* 18 (1951): 293–297.

Weissfeld, L. A. and Schneider, H. Influence Diagnostics for the Weibull Model to Fit to Censored Data. *Statistics and Probability Letters* 9 (1990): 67–73.

Wellner, J. A. A Heavy Censoring Limit Theorem for the Product Limit Estimator. *Annals of Statistics* 13 (1985): 150–162.

Woolson, R. F. Rank Tests and a One-Sample Log Rank Test for Comparing Observed Survival Data to a Standard Population. *Biometrics* 37 (1981): 687–696.

Wu, J.-T. Statistical Methods for Discretizing a Continuous Covariate in a Censored Data Regression Model. Ph.D. Dissertation, The Medical College of Wisconsin, 2001.

Zheng, M. and Klein, J. P. A Self-Consistent Estimator of Marginal Survival Functions Based on Dependent Competing Risk Data and an Assumed Copula. *Communications in Statistics-Theory and Methods* A23 (1994): 2299–2311.

Zheng, M. and Klein, J. P. Estimates of Marginal Survival for Dependent Competing Risks Based on an Assumed Copula. *Biometrika* 82 (1995): 127–138.

Author Index

Subject Index