Likelihood-based Interval Estimation of Functionals

by

Ronnie Teng Chee Lee

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 1997

0-612-30619-4

Abstract

In this thesis, we are concerned with interval estimation for functions of parameters (or functionals). In particular, we explore topics involving profile likelihood-based interval estimation of functionals in parametric models, as well as models with "missing" data.

Our study is motivated by the inadequacy of intervals based on the large sample normal approximation of test statistics when the normality assumption is not warranted. For example, in parametric settings, interval estimates for functionals based on the delta method have been known to perform poorly in applications. Although many univariate problems admit simple transformations that improve the large sample approximation, analogous approaches do not necessarily carry over to multi-parameter settings in a straightforward manner. For missing data problems. use of the observed information matrix in conjunction with the EM algorithm does not always yield satisfactory interval estimates for essentially the same reasons. While profile likelihood-based approaches to interval estimation are familiar in parametric statistical inference, its use in missing data and semi-parametric settings is not as well-known.

Chapter 1 of the thesis introduces the basic elements of likelihood-based interval estimation, with emphasis on using the profile likelihood to construct interval estimates. We describe the extension of the approach to handle functionals, via Madansky (1965). A few examples serve to round out the discussion. For parametric models, it is well-known that a simple correction factor applied to the likelihood ratio statistic (LRS) improves the quality of the approximation to the reference χ^2 distribution. This factor is known as the Bartlett correction and has routinely been applied to tests of hypotheses concerning a parameter vector or a sub-vector of it. In chapter 2, we derive a Bartlett correction to the LRS for testing a parameter function. Unlike the standard approach, our method is based on the basic assumptions and framework of the Lagrange multiplier technique. In many practical situations, we show that this approach can yield a simpler implementation. The improved performance of the corrected LRS is illustrated with examples and evaluation of coverage probabilities.

In chapter 3, we utilize the Lagrange multiplier technique, in conjunction with the EM algorithm, to obtain profile likelihood-based CIs for functionals in "missing" data settings. This yields an alternative interval estimate to those based on the observed information matrix (such as the approach of Louis, 1982). The resulting procedures are computationally intensive, but there is a potential gain in other aspects. To reduce the computational workload, we adapt the EM1 algorithm described by Rai and Matthews (1993) to the parameter function setting.

Chapter 4 is devoted to likelihood-based interval estimation of functionals in failure time models. We first consider Aitkin and Clayton's (1980) "formulation" of the parametric Cox proportional hazards model as a generalized linear model. We adapt their approach in order to derive likelihood-based interval estimates for common failure model functionals, such as quantiles and survival probabilities. Next we consider location-scale failure time models. Therneau (1992) recently fit this class of models via the method of iteratively reweighted least squares (IRLS). We demonstrate that IRLS can be conveniently adapted to constrained maximization. We also show that a Lagrange multiplier argument can be applied to this setting to provide interval estimates of some other useful functionals which are not available via the regular profile likelihood approach.

A concluding chapter provides suggestions for future research.

Acknowledgements

I would like to acknowledge and thank various persons who have made this thesis possible. Firstly, I am grateful to Prof. David Matthews for his patient and kind supervision in the writing of the thesis, and for suggesting the ideas which led to the present work. I also thank Professors Richard Cook and Mary Thompson for serving on my committee and providing many constructive comments that improved the quality of the thesis. The criticism and advice of Professor James Lindsey is also gratefully acknowledged. This research would be impossible without excellent computing assistance. Generous help from Rebekah Lim and Wayne Liu prevented a major disaster. I am indebted to my wife, Rebekah, for her support through times of discouragement in writing the thesis. I am also grateful to Chris Perumala for his concern and help many times throughout our stay in Canada. Finally. I thank the National University of Singapore for their generous financial support through the Overseas Graduate Scholarship.

Contents

List of Figures

Chapter 1

Introduction

1.1 Outline of Thesis

In this thesis, we are concerned with interval estimation for functions of parameters (or functionals). In particular, we explore topics involving profile likelihood-based interval estimation of functionals in parametric models, as well as models with "missing" data.

Our study is motivated by the inadequacy of intervals based on the large sample normal approximation of test statistics when the normality assumption is not warranted. For example, in parametric settings, interval estimates for functionals based on the delta method have been known to perform poorly in applications. Although many univariate problems admit simple transformations that improve the large sample approximation, analogous approaches do not necessarily carry over to multivariate settings in a straightforward manner. For missing data problems, use of the observed information matrix in conjunction with the EM algorithm does not always yield satisfactory interval estimates for essentially the same reasons. While profile likelihood-based approaches to interval estimation are familiar in parametric statistical inference, its use in missing data and semi-parametric settings is not as well-known.

For parametric models, it is well-known that a simple correction factor applied to the likelihood ratio statistic (LRS) improves the quality of the approximation to the reference χ^2 distribution (for example, Lawley, 1956). This factor is known as the Bartlett factor and has been applied routinely to tests of hypotheses concerning a parameter vector or a sub-vector of it. In chapter 2, we derive a Bartlett correction to the LRS for testing a parameter function. Unlike the standard approach, our method is based on the basic assumptions and framework of the Lagrange multiplier technique. In many practical situations, we show that this approach can yield a simpler implementation. The improved performance of the corrected LRS is illustrated with examples and evaluation of coverage probabilities.

In chapter 3, we utilize the Lagrange multiplier technique, in conjunction with the EM algorithm, to obtain profile likelihood-based CIs for functionals in "missing" data settings. This yields an alternative interval estimate to those based on the observed information matrix (such as the approach of Louis, 1982). The resulting procedures are computationally intensive, but there is a potential gain in other aspects. To reduce the computational workload, we adapt the EM1 algorithm proposed by Rai and Matthews (1993) to the parameter function setting.

Chapter 4 is devoted to likelihood-based interval estimation of functionals in failure time models. We first consider Aitkin and Clayton's (1980) "formulation" of the parametric Cox proportional hazards model as a generalized linear model. We adapt their approach to supply likelihood-based interval estimates for common failure model functionals, such as quantiles and survival probabilities. Next we consider location-scale failure time models. Therneau (1995) recently fit this class of models via the method of iteratively reweighted least squares. This is a convenient method for adaptation to constrained maximization. We also show that a Lagrange multiplier argument can be applied to this setting to provide interval estimates of some other useful functionals which are not available via the regular profile likelihood approach.

The remainder of this chapter introduces the basic elements of likelihood-based interval estimation, with emphasis on profile likelihood-based interval construction. We describe the extension of the approach to handle functionals, via Madansky (1965). A few examples serve to round out the discussion.

1.2 Likelihood-based Interval Estimation

In this section, we briefly review some common approaches for likelihood-based interval estimation. These are based on the score statistic, the likelihood ratio statistic and the maximum likelihood estimator (MLE). Let $Y = (Y_1, \ldots, Y_n)$ be an observed random vector with cumulative distribution function F that depends on an unknown parameter $\theta = (\theta^1, \ldots, \theta^{p+q})$. Denote the log likelihood function for θ based on Y by $\ell(\theta) = \ell$, where we suppress the dependence on θ for notational convenience. Confidence regions (CR's) for a sub-vector, say $\theta_1 = (\theta^1, \ldots, \theta^p)$, are often of interest.

The ML approach makes direct use of the MLE $\hat{\theta}_1 = (\hat{\theta}^1, \ldots, \hat{\theta}^p)$. We require the observed information matrix, which is the matrix of minus the second derivatives of ℓ with respect to θ, evaluated at $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$. Write $\nu_{\theta_1\theta_1}(\hat{\theta})$ for the leading submatrix of the inverse of the observed information matrix. Then under suitable regularity conditions (see Cox and Hinkley, 1974), the Wald statistic

$$\mathcal{W} = (\hat{\theta}_1 - \theta_1)^T \, \nu^{-1}_{\theta_1\theta_1}(\hat{\theta}) \, (\hat{\theta}_1 - \theta_1) \, ,$$

has an asymptotic χ_p^2 distribution under the null hypothesis that the true value of the parameter is θ_1. A $100(1-\alpha)\%$ CR for θ_1 is given by the set

$$\{\theta_1 | \mathcal{W}(\theta_1) \leq \chi^2_{p,1-\alpha}\},$$

where $\chi^2_{p,1-\alpha}$ is the associated $(1-\alpha)-$quantile of the chi-squared distribution with p degrees of freedom. This procedure gives rise to an elliptical CR for θ_1, centred at $\hat{\theta}_1$.

Another approach is based on the score statistic,

$$\mathcal{U} = S_{\theta_1}^T \ \nu_{\theta_1\theta_1}(\theta_1, \tilde{\theta}_2(\theta_1)) \ S_{\theta_1},$$

where

$$S_{\theta_1} = \left[\frac{\partial \ell}{\partial \theta_1}\right]_{\theta_2 = \tilde{\theta}_2(\theta_1)},$$

and $\tilde{\theta}_2(\theta_1)$ is the MLE of $\theta_2 = (\theta^{p+1}, \ldots, \theta^{p+q})$ corresponding to a fixed value of θ_1. Since S_{θ_1} is asymptotically normally distributed with zero mean and covariance matrix $\nu^{-1}_{\theta_1\theta_1}(\theta_1, \tilde{\theta}_2(\theta_1))$ under the null hypothesis, it follows that $\mathcal{U}$ is approximately χ_p^2 distributed. The score statistic offers computational savings over the previous approach as the full MLE does not have to be obtained.

A third approach is based on the LRS. Due to its central role in this proposal, we discuss this method of interval estimation separately in the following section.

1.3 Profile Likelihood-based Inference

In this section, we outline profile likelihood-based interval estimation for vector parameters, and its extension to include functional parameters. The latter development is a relatively recent innovation that originated with Madansky (1965), and was utilized more recently by Cox and Oakes (1984) in a survival analysis setting. Matthews (1988a) applied the methodology in the context of independent binomial random variables to obtain approximate CIs for complicated functionals. Some theoretical properties of the method were discussed by Critchley *et al.* (1988). illustrating its scope and generality.

Following the presentation in the last section, approximate CRs for θ_1 can be based on the profile log likelihood function $\ell(\theta_1, \tilde{\theta}_2(\theta_1))$. The asymptotic χ_p^2 distribution of the LRS, $2\{\ell(\hat{\theta}) - \ell(\theta_1, \tilde{\theta}_2(\theta_1))\}$. is well-known. and can be employed to obtain approximate confidence regions for θ_1. It is also well-known that the LRS and the statistics discussed in the previous section are equivalent to first order. so that higher-order comparisons are generally necessary to assess their relative merits.

In certain models, the dimension of θ_2 is "large" or increases with the sample size (as in the Neyman-Scott problem). The profile likelihood method is known to be potentially misleading in such cases, since insufficient information is available in the sample for precise estimation of the effects of interest. after eliminating the nuisance parameters. Numerous methods have been proposed to correct the profile likelihood in such situations (via conditional likelihood or other adjustments; see for example. Cox and Reid (1987) and McCullagh and Tibshirani (1990)). The methodology developed in subsequent chapters assumes that such problematic considerations do not arise. For most practical problems, this is a reasonable assumption and CRs based on the profile likelihood are satisfactory.

Madansky (1965) extended the profile likelihood method to obtain CIs for functionals as follows. Suppose we require an approximate CI for some functional, $f(\theta) \in \Re$, where f possesses continuous first partial derivatives with respect to

θ. A profile likelihood for $f(\theta)$ can be constructed by setting $f(\theta) = \omega$, $\omega \in \Omega$, where Ω is the set of values of f for which the constrained MLE $\tilde{\theta}$ exists. Using the method of Lagrange multipliers, we find $\sup_{\theta: f(\theta)=\omega} \ell(\theta) = \ell(\tilde{\theta}(\omega))$. From Critchley *et. al* (1988), given a fixed ω and provided $\tilde{\theta}(\omega)$ exists, there is a unique value of the Lagrange multiplier ξ corresponding to this value of ω. Therefore, we can also write $\tilde{\theta}(\omega) = \tilde{\theta}(\xi)$. This suggests an alternative way to locate the constrained MLEs (which we illustrate in a subsequent section). Under the regularity conditions mentioned in the previous section, the asymptotic distribution of the LRS, $2\{\ell(\hat{\theta}) - \ell(\tilde{\theta})\}$, is generally closely approximated by the χ^2_1; this fact can be employed to yield the required CI for $f(\theta)$. The profile likelihood-based method of interval estimation for functionals can also be extended to cover the case of multiple functionals, as indicated by Silvey (1959). Generally, intervals obtained by this method better reflect the information in the data, compared to the usual Wald-type confidence intervals based on the delta method. In the following section, we illustrate this method of interval estimation.

1.4 Examples

Example 1.1 Estimating the Number Needed to Treat in Independent Binomial Samples

Consider the randomized trial reported by Oxner *et al.* (1992) in which the endoscopic injection of adrenaline and ethanolamine was investigated for its efficacy in adults with bleeding or non-bleeding peptic ulcers. Eligible patients were all adults admitted to a British district general hospital for suspected gastrointestinal hemorrhage and who had an endoscopy within 16 hours of admission. The data are summarized in the following 2 × 2 table. In this setting, we assume that

Table 1.1: Endoscopic injection data for patients with bleeding or non-bleeding peptic ulcers, from Oxner *et al.* (1992).

Therapy	Number of Events	Number Randomized
Control	21	45
Injection	8	48
Total	29	93

$X_i \sim B(n_i, p_i)$, for $i = 1, 2$. The log likelihood function is

$$\ell(p_1, p_2) = \sum_{i=1}^{2} \{x_i \log p_i + (n_i - x_i) \log(1 - p_i)\}.$$

Suppose we are interested in interval estimation of the functional $\theta = (p_1 - p_2)^{-1}$. In medical terms, θ is the number needed to treat (see Laupacis *et al.*, 1988) or simply. NNT. This parameter is especially useful to clinicians due to its interpretation as the average number of patients that need to receive treatment 2 rather than treatment 1 in order to prevent one adverse event. In epidemiological terms. the NNT is the reciprocal of the absolute risk reduction (ARR). the difference in the event rates for the treatment and control arms of the trial.

The procedure for obtaining interval estimates for the NNT is analytically simpler if we first obtain CIs for $\theta^{-1} = p_1 - p_2$. The end-points of the interval estimate for θ^{-1} can then be inverted to yield the required CI for θ. We note that, due to the nature of the function θ^{-1} over the domain $\theta \in [-1, 1]$, disjoint interval estimates for θ might result. We form the augmented log likelihood function

$$\ell_\xi(p_1, p_2) = \ell(p_1, p_2) + \xi\{\theta^{-1} - (p_1 - p_2)\}.$$

This is particularly straightforward to maximize with respect to the p_i's. In conjunction with the range restriction for probabilities, we obtain

$$\tilde{p}_1 = \frac{n_1 + \xi - \sqrt{(n_1 + \xi)^2 - 4\xi x_1}}{2\xi}, \qquad \tilde{p}_2 = \frac{\xi - n_2 + \sqrt{(\xi - n_2)^2 + 4\xi x_2}}{2\xi}$$

as the constrained MLEs for p_1 and p_2. The LRS, based on the profile log likelihood for θ^{-1}, is given by

$$W(\xi) = 2\{\ell(\hat{p}_1, \hat{p}_2) - \ell(\tilde{p}_1, \tilde{p}_2)\},$$

where the unconstrained MLEs $\hat{p}_i = \frac{x_i}{n_i}$ $(i = 1.2)$. The $100(1-\alpha)\%$ CI for θ^{-1} is given by the set $\{\xi | W(\xi) \le \chi^2_{1.1-\alpha}\}$. A repeated bisection routine can be used to locate the required end-points. For the data in Table 1.1, the approximate 95% CI for $p_1 - p_2$ is (0.116, 0.474); the corresponding interval for $(p_1 - p_2)^{-1}$ is therefore (2.11. 8.63).

It is helpful to consider briefly the geometrical details of the above procedure. Figure 1.1 shows a contour plot of the log likelihood surface for (p_1, p_2). The surface attains a unique maximum at the point $(\hat{p}_1, \hat{p}_2)$ (indicated by "+"). The curve (or line, in this case) $\theta = p_1 - p_2$ is drawn for a range of values of $\theta \in [-1.1]$ (dashed lines). Given an arbitrary fixed $\theta^{-1} = \theta_0^{-1}$, the maximum of the slice of the log likelihood surface, corresponding to the set $\{(p_1, p_2) : p_1 - p_2 = \theta_0\}$, is located. Provided the constrained maximum exists, there is a unique value of the Lagrange multiplier, say ξ_0, corresponding to this maximum point. In the notation established above, the maximum of the log likelihood surface for the constrained set of (p_1, p_2) values is therefore $(\tilde{p}_1(\xi_0), \tilde{p}_2(\xi_0))$. The set $\{(\tilde{p}_1(\xi), \tilde{p}_2(\xi)\}$, or equivalently, $\{(\tilde{p}_1, \tilde{p}_2) | \theta^{-1} \in [-1, 1]\}$, is indicated by the dashed curve in the figure. By recalibrating the ξ scale in terms of θ, we obtain the plot of the LRS based on the profile log likelihood for the NNT (Figure 1.2).

Figure 1.1: Likelihood and constraint contours for the NNT based on the data of Oxner *et al.* (1992). The contour levels displayed are 0.1, 0.3, 0.5, 0.7 and 0.9. The superimposed dashed lines (- - -) are level curves of $\theta = 1/(p_1 - p_2)$, corresponding to NNT values of 12 (top left), 6, 3, and 2 (bottom right). The curve (– – –) indicates the locus of constrained maxima defined by the profile log likelihood of the NNT.

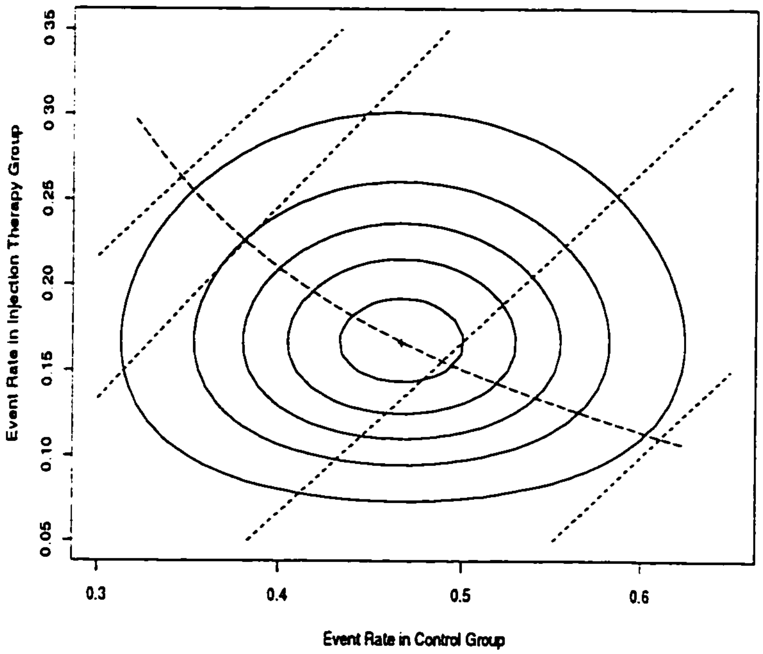
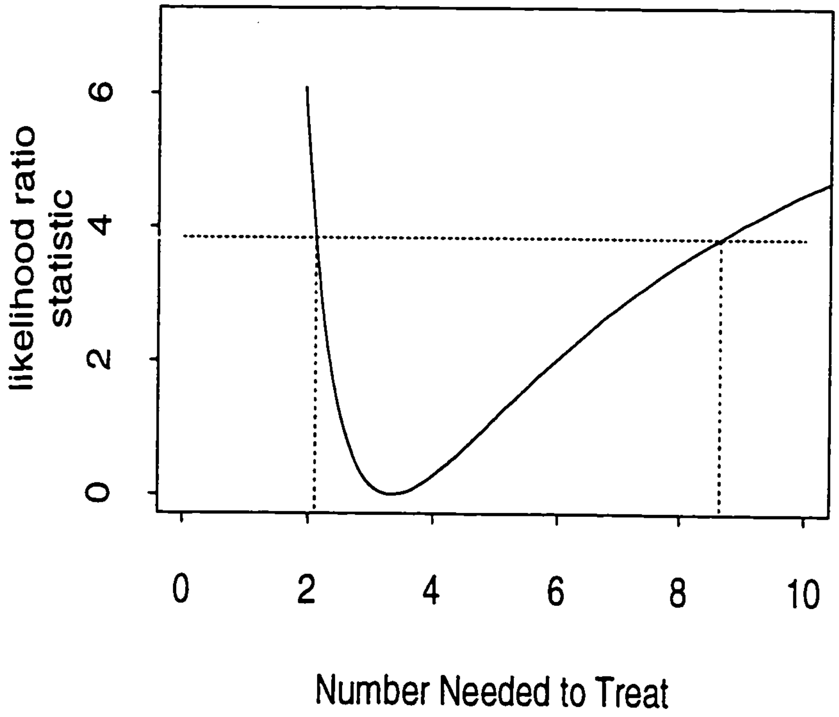


Figure 1.2: Likelihood ratio statistic for the NNT based on the data in Table 1.1. The height of the horizontal dotted line is 3.841 units. The vertical dashed lines mark the endpoints of an approximate 95% CI for the NNT.



Example 1.2 Estimating the Number Needed to Treat in a Failure Time Setting

A major difficulty arises in comparing the NNT estimates from different studies if the duration of follow-up differs from trial to trial. In this case, the reported event rates and NNT estimates are no longer comparable. One solution to this problem is to adopt suitable parametric assumptions, such as constant hazards for events under both treatment regimes, and then rescale the information originally reported as 2 × 2 tables of response status by treatment group to a common time-frame. However, as Cook *et al.* (1995) point out in an unpublished manuscript, when the observed response is the time until an event of interest is recorded and a failure time analysis is possible, such parametric assumptions are unnecessary.

The example in this section is based on the North American Symptomatic Carotid Endarterectomy Trial (NASCET); see NASCET Collaborators (1991). This was a large randomized controlled trial that has reported the relative merits of best medical therapy versus best medical therapy plus carotid endarterectomy in preventing subsequent strokes in patients with high grade ($\geq$ 70%) symptomatic carotid stenosis. Patients were randomized to one of the two treatments and followed for an average of approximately two years. The response measurement for each subject consisted of days since randomization to treatment, with severe stroke or death as the clinically important outcomes. The status of each subject at the conclusion of the trial was also recorded.

Cook *et al.* (1995) also analyzed this data set, providing, among other graphical output, a Kaplan-Meier plot of the occurrence of events over time for the two treatment groups and the Kaplan-Meier estimates of the NNT at various points in time. In the following, we obtain pointwise, nonparametric interval estimates of the NNT, using the profile likelihood-based method.

Without loss of generality, let $t_{11} < \ldots < t_{1n_1}$ and $t_{21} < \ldots < t_{2n_2}$ denote the distinct event times from the two arms of the trial, with event multiplicities $\{d_{ij}, \ j = 1, \ldots, n_i, \ i = 1, 2\}$. Let r_{ij} represent the risk set at $t_{ij}-$, and c_{ij} the number censored at $t_{ij}+$. In this example, we consider deriving a CI for the associated NNT parameter, $\theta_t = \{\mathcal{F}_1(t) - \mathcal{F}_2(t)\}^{-1}$, where $\mathcal{F}_i(t)$ is the survivor function for treatment group i. Let $f_i(t)$ denote the corresponding density functions. We assume that censoring is uninformative in the sense that an observed censoring time t conveys only the information that the latent survival time exceeds t. The likelihood function, based on the data, is given by

$$L = \prod_{j \in \mathcal{D}_1} \{f_1(t_{1j})\}^{d_{1j}} \{\mathcal{F}_1(t_{1i})\}^{c_{1j}} \prod_{j \in \mathcal{D}_2} \{f_2(t_{2j})\}^{d_{2j}} \{\mathcal{F}_2(t_{2j})\}^{c_{2j}} ,$$

where $\{\mathcal{D}_j\}$, $j = 1, 2$, represent the distinct event times for the two groups. Without imposing distributional assumptions on the failure times, the argument of Kaplan and Meier (1958) can be used to show that L is maximized over the set of discrete distribution functions with point masses located at the times in $\mathcal{D}_1 \cup \mathcal{D}_2$.

Accordingly, we define $h_{ij} = h_i(t_{ij}) = f_i(t_{ij})/\mathcal{F}_i(t_{ij}-)$, $j = 1, \ldots, n_i$, $i = 1, 2$, and regard the h_{ij} as binomial proportions (i.e., conditioning on the risk sets at each failure time). More specifically, h_{ij} is the probability of failure at t_{ij}, conditional on surviving up to $t_{ij}-$. From the preceding argument, the log likelihood for the data from each arm of the trial can be written as

$$\ell_i(h) = \sum_{j=1}^{n_i} \{ \ d_{ij} \log(h_{ij}) + (r_{ij} - d_{ij}) \log(1 - h_{ij}) \ \} \ , \quad i = 1, 2.$$

This follows from considering the discrete form of an empirical survivor function, i.e., $\mathcal{F}_i(t) = \prod^{(t)}\{1 - h_i(t_{ij})\}$, where $\prod^{(t)}$ denotes product over j, $t_{ij} < t$. The form of the log likelihood function is exactly that for independent binomial samples.

The method of interval construction we use is based on the sampling distribution of the nonparametric LRS. Thomas and Grunkmeier (1975) first used it in the context of a homogeneous failure time sample subject to right-censorship to derive confidence intervals for the survivor function. They showed informally that the LRS in that setting has asymptotically a chi-squared distribution with one degree of freedom. Cox and Oakes (1984) recently used a Lagrange multiplier argument in conjunction with the nonparametric LR approach to derive approximate CIs for the survivor function. The asymptotic sampling distribution of the LRS used in the preceding work was rigorously derived by Li (1995). In particular, he established that the LRS used in Thomas and Grunkmeier (1975) is the nonparametric analogue of the LRS in parametric settings. It is possible to adapt Li's approach to establish rigorously the asymptotic distribution of the LRS for our problem.

As in the previous example, it is more convenient to obtain a CI for θ_t^{-1} first, and then invert the end-points of the interval to obtain the required interval estimate for θ_t. Clearly, the unconstrained MLEs are $\hat{h}_{ij} = d_{ij}/r_{ij}$. To obtain the constrained MLEs corresponding to a fixed $\theta_t^{-1} = \{\mathcal{F}_1(t) - \mathcal{F}_2(t)\}$, we maximize the log likelihood

$$\ell_\xi(h) = \ell_1(h) + \ell_2(h) + \xi\{\theta_t^{-1} - \mathcal{F}_1(t) + \mathcal{F}_2(t)\}.$$

Since $\mathcal{F}_i(t) = \prod_{j|t_{ij}<t}(1-h_{ij}) = \prod^{(t)}(1-h_{ij})$, we obtain the score equations

$$\frac{d_{1j}}{h_{1j}} - \frac{r_{1j}-d_{1j}}{1-h_{1j}} + \xi\frac{\prod^{(t)}(1-h_{1j})}{1-h_{1j}} = 0 \quad (t_{1j} < t) \tag{1.1}$$

$$\frac{d_{2j}}{h_{2j}} - \frac{r_{2j}-d_{2j}}{1-h_{2j}} - \xi\frac{\prod^{(t)}(1-h_{2j})}{1-h_{2j}} = 0 \quad (t_{2j} < t) \tag{1.2}$$

$$\frac{d_{ij}}{h_{ij}} - \frac{r_{ij} - d_{ij}}{1 - h_{ij}} = 0 \quad (t_{ij} \geq t) \tag{1.3}$$

For $t_{ij} \geq t$, the constrained MLEs are equal to the unconstrained estimates, that is, $\tilde{h}_i(t_{ij}) = \hat{h}_i(t_{ij}) = \frac{d_{ij}}{r_{ij}}$. For $t_{ij} < t$, we need to solve equations (1.1) and (1.2) simultaneously to obtain the constrained ML estimates $\{\tilde{h}_{ij}(\xi)\}$. To facilitate the numerical work, we reexpress equations (1.1), (1.2) as

$$h_{1j} = \frac{d_{1j}}{r_{1j} - \xi \mathcal{F}_1(t)} \tag{1.4}$$

$$h_{2j} = \frac{d_{2j}}{r_{2j} + \xi \mathcal{F}_2(t)} \tag{1.5}$$

for $t_{ij} < t$. Functional iteration (Conte and De Boor, 1980) can be applied to equations (1.4) and (1.5) to obtain the constrained MLEs of $\{h_{ij}\}$. Once the constrained MLEs $\{\tilde{h}_{ij}(\xi)\}$ are known, the value of the LRS

$$W(\xi) = 2\sum_{i=1}^{2}\sum^{(t)}\left\{d_{ij}\log\left(\frac{\hat{h}_{ij}}{\tilde{h}_{ij}}\right) + (r_{ij} - d_{ij})\log\left(\frac{1-\hat{h}_{ij}}{1-\tilde{h}_{ij}}\right)\right\}$$

can be determined. A numerical method, such as repeated bisection, can be used to locate $\xi_+ > 0$ and $\xi_- < 0$ such that $W(\xi_+) = W(\xi_-) = \chi^2_{1,1-\alpha}$. These values of ξ, i.e., ξ_+ and ξ_-, and the associated constrained MLEs, $\tilde{h}_{ij}(\xi_+)$ and $\tilde{h}_{ij}(\xi_-)$, yield the end-points, $\{\tilde{\mathcal{F}}_1(\xi_-) - \tilde{\mathcal{F}}_2(\xi_-)\}^{-1}$ and $\{\tilde{\mathcal{F}}_1(\xi_+) - \tilde{\mathcal{F}}_2(\xi_+)\}^{-1}$, of the required $100(1-\alpha)\%$ CI for θ_t.

The Kaplan-Meier plot for the NASCET data (Figure 1.3) reveals that the estimated survival curves corresponding to the two arms of the trial intersect at approximately 100 days post-randomization. Prior to this time, the medical therapy arm (Group 1) appears to fare better; thereafter, the therapeutic benefit of surgical treatment for high grade symptomatic carotid stenosis (Group 2) becomes evident.

Although it is possible to derive symmetric variance bounds for the nonparametric survivor function heuristically (cf. Kalbfleisch and Prentice (1980)) to obtain interval estimates for the NNT, we have not done so in this case. As seen below, the normality assumption is inadequate in the present case. Nonparametric, likelihood-based, pointwise 95% CIs for θ_t at roughly 50-day intervals are displayed in Figure 1.4. The pointwise maximum likelihood estimates of the NNT are denoted by the symbol "X".

Since the Kaplan-Meier estimates of the survivor functions for the two arms of the trial are quite close during most of the first year of follow-up, the resulting interval estimates during this period are the union of two disjoint intervals of the form $(-\infty, \theta_1)$ and (θ_2, ∞), where $\theta_1 < 0$ and $\theta_2 > 0$. The point estimate for θ_t belongs to one of these disjoint parts of the pointwise 95% CI. This unusual interval estimate can be interpreted as indicating that the NASCET data provide no evidence that the survival functions for the two arms of the study differ during the first year following treatment. Subsequently, the 95% pointwise CIs for the NNT are all continuous intervals of finite length lying below the line $\theta_t = 0$, thereby indicating the long-term advantage enjoyed by patients who received the combined medical and surgical treatment and who survived the initial period of treatment indifference. The additional information supplied by the NNT interval estimates indicates that, in the long-run, combined medical and surgical treatment for high grade symptomatic carotid stenosis should prevent one adverse event for at least 21 months in approximately 9 patients.

Figure 1.3: Kaplan-Meier estimates of the survival functions, by treatment group, for the NASCET data.

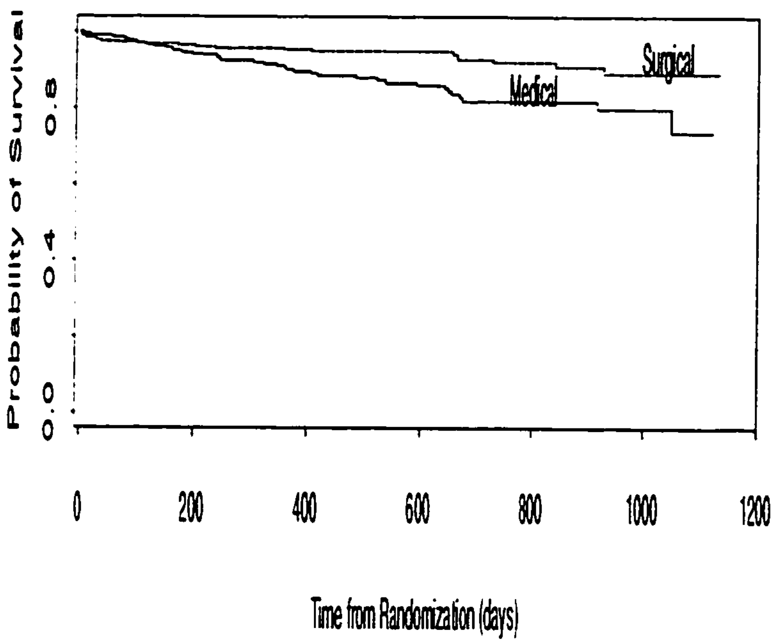
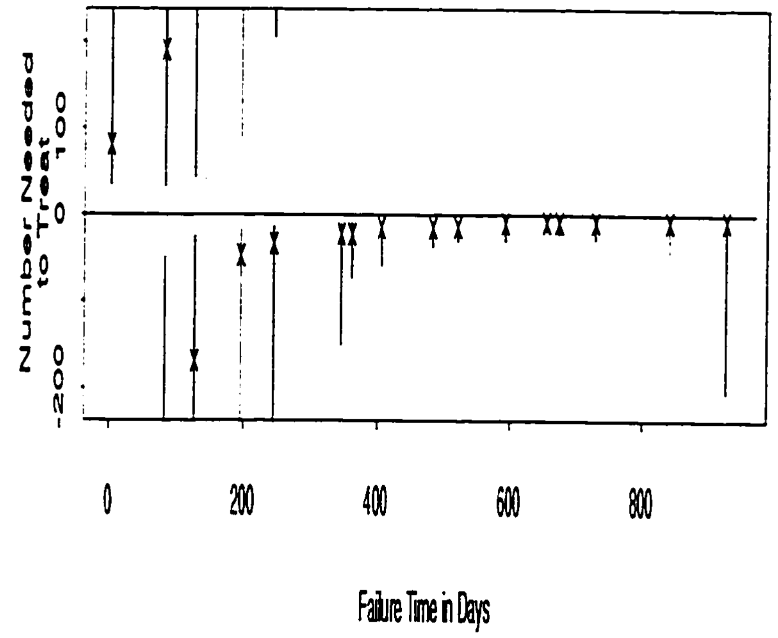


Figure 1.4: Nonparametric, pointwise 95% confidence intervals for the time-dependent NNT approximately every 50 days, based on the NASCET data. Point estimates are indicated by X. Each confidence interval can be a single, continuous interval of NNT values of finite length ($t < 8$ or $t > 304$) or the union of two disjoint intervals, each of infinite length, that exclude the value 0, $8 \leq t \leq 304$.



Chapter 2

Improved Likelihood-based CIs for Functionals

2.1 Introduction

For parametric statistical inference, a conceptually simple way to improve the asymptotic χ^2 approximation of the LRS is by means of a multiplicative factor called the Bartlett correction factor. The basic idea is to scale the LRS by the inverse of its asymptotic mean (the Bartlett factor). The remarkable effect of this adjustment is that the resulting pivotal statistic possesses cumulants which differ from those of χ^2 by terms which are $O(n^{-2})$, where n refers to the sample size. This approximation was first given by Bartlett (1937), and a general method for deriving the correction was provided by Lawley (1956). Barndorff-Nielsen and Cox (1984) give an account of parametric Bartlett correction, and in particular obtained the adjustment via the saddlepoint method. It is well-known that the Bartlett correction unambiguously improves the error rate in the case of continuous random variables. There is, however, some empirical evidence showing that corresponding improvements are not guaranteed in the case of discrete random variables (Frydenberg and Jensen, 1989). Even in this latter case, Kolassa (1994) asserts that Bartlett correction does improve accuracy in many cases.

The applicability of Bartlett correction extends to empirical likelihood as well. DiCiccio *et al.* (1991) showed that empirical likelihood is Bartlett-correctable: in particular, a Bartlett factor is available for a wide range of parameters such as means, variances, covariances, etc. Hall and La Scala (1990) provide a survey of these results. However, recent work by Lazar and Mykland (1995) demonstrate that Bartlett correction does not work in the empirical likelihood setting when forming a confidence region for a subset of the parameters of interest. In addition, apart from empirical likelihood, no Bartlett corrections are as yet available for the nonparametric setting.

In this chapter we derive an alternative, and in some cases, simpler way to obtain the Bartlett factor in the parameter function setting. The motivation for our approach stems from the Lagrange multiplier technique introduced by Madansky (1965) to provide likelihood-based interval estimates for functions of parameters. Madansky's method is a simple alternative to the usual approach of reparametrizing the model in terms of the function of interest and other "nuisance" parameters. The proposed approach is particularly useful when the Lagrange multiplier argument yields closed-form solutions for estimates of constrained parameters. In this case, the Bartlett factor can be obtained using the assumptions and framework of the Lagrange multiplier technique. In section 2.2, we briefly review some basic results on Bartlett adjustment of the LRS for tests of hypotheses concerning parameter vectors. Following Lawley (1956), and based on the framework and assumptions of the Lagrange multiplier technique, we obtain an appropriate Bartlett factor for likelihood-based confidence intervals for parameter functions in section 3. We illustrate our solution in the general case with some examples, and conclude the chapter by investigating the coverage probabilities of the profile LRS in selected cases.

2.2 Bartlett Factor for Parameter Vectors : A Review

Since the derivation of Bartlett factors for parameter functions is based on that for parameter vectors, a brief introduction to the technique and notation of the latter case is included in this section.

We assume a statistical model with full parameter vector $\theta \in \Re^{p+q}$, and data Y generated by the model. We also assume sufficient regularity exists and that the second partial derivatives ℓ_{rr} (defined below) are of order n, where n is related to the number of observations. Following Lawley (1956), standard conventions for denoting arrays and summation are used. According to these conventions, the indices $r, s, t, \ldots$ range over $1, \ldots, p+q$. Differentiation is indicated by subscripts, so $\ell_r = \frac{\partial \ell}{\partial \theta^r}$, $\ell_{rs} = \frac{\partial^2 \ell}{\partial \theta^r \partial \theta^s}$, etc. Let $\lambda_{rs} = E\{\ell_{rs}\}$, $\lambda_{rst} = E\{\ell_{rst}\}$, etc., and define $\mathcal{L}_r = \ell_r$, $\mathcal{L}_{rs} = \ell_{rs} - \lambda_{rs}$, $\mathcal{L}_{rst} = \ell_{rst} - \lambda_{rst}$, etc. We further denote $(\lambda_{rs})_t = \frac{\partial \lambda_{rs}}{\partial \theta^t}$, $(\lambda_{rst})_u = \frac{\partial \lambda_{rst}}{\partial \theta^u}$ and $(\lambda_{rs})_{tu} = \frac{\partial^2 \lambda_{rs}}{\partial \theta^t \partial \theta^u}$. The quantities λ_{rs}, λ_{rst}, etc. are generally of order $O(n)$. The variables ℓ_r, $\mathcal{L}_{rs}$, $\mathcal{L}_{rst}$, etc. are typically of order $O_p(n^{1/2})$ (this holds when we have an i.i.d. sample, for example). Denote the matrix inverse of (λ_{rs}) by (λ^{rs}).

Suppose we are interested in the subset $\theta_2 = (\theta^{p+1}, \ldots, \theta^{p+q})$, with θ_1 defined accordingly as its complement. Let $\tilde{\theta}_1 = \tilde{\theta}_1(\theta_2)$ denote the MLE of θ_1, given a fixed value of θ_2. Lawley (1956) showed that $E[2\{\ell(\hat{\theta}_1, \theta_2) - \ell(\theta_1, \theta_2)\}] = p + \epsilon_p + O(n^{-2})$, where

$$
\begin{aligned}
\epsilon_p &= \lambda^{rs}\lambda^{tu}\{\frac{1}{4}\lambda_{rstu} - (\lambda_{rst})_u + (\lambda_{rt})_{su}\} - \lambda^{rs}\lambda^{tu}\lambda^{vw}\{\frac{1}{6}\lambda_{rtv}\lambda_{suw} + \frac{1}{4}\lambda_{rtu}\lambda_{svw} \\
&- \lambda_{rtv}(\lambda_{sw})_u - \lambda_{rtu}(\lambda_{sw})_v + (\lambda_{rt})_v(\lambda_{sw})_u + (\lambda_{rt})_u(\lambda_{sw})_v\}\,.
\end{aligned}
$$

Considerable simplification results when $\lambda_{rs} = 0 = \lambda^{rs}$ for $r \neq s$. Such a case occurs when, for example, the components of θ are orthogonal. As indicated by Lawley. the formula for ϵ_p in this case is simply

$$
\begin{aligned}
\epsilon_p &= \sum_r\sum_s\{\frac{1}{4}\lambda_{rrss} - (\lambda_{rrs})_s + (\lambda_{rs})_{rs}\}/(\lambda_{rr}\lambda_{ss}) + \sum_r\sum_s\sum_t\{\frac{1}{6}\lambda^2_{rst} + \frac{1}{4}\lambda_{rss}\lambda_{rtt} \\
&- \lambda_{rst}(\lambda_{rs})_t - \lambda_{rss}(\lambda_{rt})_t + (\lambda_{rs})_t(\lambda_{rt})_s + (\lambda_{rs})_s(\lambda_{rt})_t\}/(-\lambda_{rr}\lambda_{ss}\lambda_{tt})\,,
\end{aligned}
$$

where the summation convention has been dropped in this special instance. Based on the general formula for ϵ_p, it follows that the expectation of the LRS. $2\{\ell(\hat{\theta}_1.\hat{\theta}_2) - \ell(\theta_1.\theta_2)\}$, is given by $p + q + \epsilon_{p+q} + O(n^{-2})$. The term ϵ_{p+q} is determined according to the formula for ϵ_p, with subscripts and superscripts now running from 1 to $p+q$. Combining these results, the expectation of the LRS

$$
\begin{aligned}
W &= 2\{\ell(\hat{\theta}_1,\hat{\theta}_2) - \ell(\tilde{\theta}_1,\theta_2)\} \\
&= 2\{\ell(\hat{\theta}_1,\hat{\theta}_2) - \ell(\theta_1,\theta_2) - \ell(\tilde{\theta}_1,\theta_2) + \ell(\theta_1,\theta_2)\} \\
&= 2\{\ell(\hat{\theta}_1,\hat{\theta}_2) - \ell(\theta_1,\theta_2)\} - 2\{\ell(\tilde{\theta}_1,\theta_2) - \ell(\theta_1,\theta_2)\}
\end{aligned}
$$

is thus

$$(p + q + \epsilon_{p+q}) - (p + \epsilon_p) + O(n^{-2}) = q + \epsilon_{p+q} - \epsilon_p + O(n^{-2})\,.$$

The adjusted LRS, $W' = (1 + \frac{\epsilon_{p+q}-\epsilon_p}{q})^{-1}W$, possesses cumulants closer to those of the χ^2_q distribution than the original unscaled statistic. Specifically, the cumulants of W' and the χ^2_q distribution are equivalent to order n^{-1}. Additional arguments are, however, required to make the desired conclusion that $W' \sim \chi^2_q + O(n^{-2})$; for example, see McCullagh (1987).

2.3 Bartlett Factor for Parameter Functions

2.3.1 Derivation

In this section, we consider the problem of deriving the Bartlett factor for the parameter function case. We continue to assume a fully parametric statistical model indexed by θ, as defined previously. Before proceeding with the development proper, we note the possibility of applying the proposed technique to nonparametric settings for which the nonparametric likelihood function assumes the form of a "parametric" likelihood. For example, Cox and Oakes (1984) show that the likelihood function based on a random sample of failure times subject to right-censorship has the form of a likelihood from independent binomials. However, the connection between this "parametric" likelihood and the nonparametric likelihood is not explored here.

It is helpful to think of confidence interval construction for $f(\theta)$ as a problem in hypothesis testing, as follows. Consider a fixed value of $f(\theta)$, say $f(\theta) = \omega$; let $h(\theta) = f(\theta) - \omega$. Approximate confidence intervals for f can be constructed by finding the set of ω that are not rejected by the hypothesis $H_0 : h(\theta) = 0$. A natural way of handling this problem is to reparametrize the model in terms of $f(\theta)$ and some other "nuisance" parameters, where the latter are chosen to make the mapping one-to-one. Lawley's method may then be applied directly to obtain the Bartlett factor. While this approach is straightforward in principle, at least in the case of single parameter functions, it can complicate the processes of computing the constrained MLEs and the Bartlett factor. For example, for orthogonal parameters. $\lambda_{rs} = \lambda^{rs} = 0$, $r \neq s$, which greatly simplifies the computation of ϵ_{p+q} in the Bartlett factor. This condition does not necessarily apply under the reparametrized model. For the case of multiple functions of interest, it may not be straightforward to find a suitable one-to-one reparametrization of the parameters. This problem is avoided by our approach; as long as the multiple constraints on the model parameters are not dependent, our method proceeds along the lines for the single functional case. In the following. we derive another solution to the problem.

The following derivation utilizes the implicit function theorem (cf. Protter and Morrey. 1991, chap.14). This theorem provides conditions under which the equation $h(\theta) = 0$ can be solved explicitly for one parameter in terms of the remaining parameters. The Lagrange multiplier method is valid provided we can perform this operation (even if only in principle). Even for certain implicitly-defined functions. it is clear that the equation $h(\theta) = 0$ can be routinely solved for one of its variables. In general. we will rely on results for implicit functions to indicate whether such solutions exist. For our present purposes, we suppose that $h(\theta)$ and its first partial derivatives are continuous on an open set in $\Re^{p+q}$ containing $\tilde{\theta} = \tilde{\theta}(\omega)$. for $\omega \in \Omega$. Without loss of generality, we further assume that

$$h(\tilde{\theta}^1, \ldots, \tilde{\theta}^{p+q-1}, \tilde{\theta}^{p+q}) = 0$$

and

$$\frac{\partial h(\tilde{\theta}^1, \ldots, \tilde{\theta}^{p+q-1}, \tilde{\theta}^{p+q})}{\partial \theta^{p+q}} \neq 0 .$$

where the LHS of the second term above denotes the first partial derivative of h with respect to θ^{p+q}, evaluated at $\tilde{\theta}$. If $\tilde{\theta}$ is the constrained MLE of θ (obtained via the Lagrange multiplier technique), then the assumption $h(\tilde{\theta}) = 0$ is automatically satisfied. Under these conditions, we can express θ^{p+q} as a function of $(\theta^1, ..., \theta^{p+q-1})$, i.e., $\theta^{p+q} = F(\theta^1, ..., \theta^{p+q-1})$ (Protter and Morrey, 1991, Theorem 14.2).

Following the preceding development, we only need to consider cases where $h(\theta) = 0$ permits a solution for θ^{p+q} in terms of the other parameters. Then $H_0 : h(\theta) = 0$ and $H_0 : \theta = g(\beta)$ are equivalent, for some vector-valued function $g : \Re^{p+q-1} \rightarrow \Re^{p+q}$, where $\beta = (\beta^1, ..., \beta^{p+q-1})^T$ and $\text{rank}[\frac{\partial\theta}{\partial\beta}] = p+q-1$ (Rao, 1973, p. 418). For example, for $\theta = (\theta^1, \theta^2, \theta^3)^T$, $h(\theta) = \frac{\theta^1 - \theta^2}{\theta^3}$, we have $\beta = (\beta^1, \beta^2)^T$ and $g(\beta) = (\beta^1, \beta^1, \beta^2)^T$. For a more complicated case, consider a functional in a failure time setting, $h(\theta) = \sum_{t_j<t} \log(1 - h_j) - \log(1 - \omega)$, where t is some arbitrary fixed index t denoting a predetermined censoring time, for instance. The h_j's are the conditional probabilities of failure at t_j, given survival up to t_j-. Suppose $\text{card}\{j : t_j < t\} = s$. We can write $g(\beta) = (\beta^1, ..., \beta^{s-1}, 1 - \frac{1-\omega}{\prod_{i=1}^{s-1} \beta^i})^T$, by letting $(h_1, ..., h_{s-1}) \equiv (\beta^1, ..., \beta^{s-1})$. Functionals that afford explicit solution for the constraint $h(\theta) = 0$ in terms of one of the parameters give rise to straightforward computation of the Bartlett factor, as demonstrated below. For the general case of functionals that afford explicit solution only in principle, we also indicate how the Bartlett factor can be calculated.

The LRS, $W(\xi)$, based on the profile log likelihood function is defined by

$$\begin{aligned} \frac{1}{2}W(\xi) &= \sup_{\theta\in\Theta} \ell(\theta) - \sup_{\theta:h(\theta)=0} \ell(\theta) \\ &= \ell(\hat{\theta}) - \ell(\tilde{\theta}) \end{aligned}$$

where $\tilde{\theta}$ is the constrained MLE of θ under H_0 and Θ denotes the unrestricted parameter space. From the equivalence of $H_0 : h(\theta) = 0$ and $H_0 : \theta = g(\beta)$, we have that $\ell(\theta) = \ell(\beta)$ under H_0. It is also known that $\sup_{\beta\in\Re^{p+q-1}} \ell(\beta) = \sup_{\theta:h(\theta)=0} \ell(\theta)$.

Therefore, under the null hypothesis, it follows that

$$\frac{1}{2}W(\xi) = \ell(\hat{\theta}) - \ell(\hat{\beta}) = \{\ell(\hat{\theta}) - \ell(\theta)\} - \{\ell(\hat{\beta}) - \ell(\beta)\} \tag{2.1}$$

and we know that, in general, $W(\xi)$ is asymptotically distributed as χ^2_1.

The Bartlett factor is obtained by taking the expectation of (2.1). From Lawley (1956), $E[\ 2\{\ell(\hat{\theta}) - \ell(\theta)\}\ ]$ and $E[\ 2\{\ell(\hat{\beta}) - \ell(\beta)\}\ ]$ are $p + q + \epsilon_{p+q} + O(n^{-2})$ and $p + q - 1 + \epsilon_{p+q-1} + O(n^{-2})$, respectively. The asymptotic mean of $W(\xi)$ is $1 + \epsilon_{p+q} - \epsilon_{p+q-1}$ with error rate $O(n^{-2})$, and therefore the Bartlett-adjusted LRS is $(1 + \epsilon_{p+q} - \epsilon_{p+q-1})^{-1}W$. The Bartlett factor derived here is in fact described by Barndorff-Nielsen and Blaesild (1986) as the adjustment associated with testing a submodel of a full model. The preceding derivation highlights their point. As opposed to the usual approach in which suitable one-to-one reparametrizations of the parameters must be found, our method is relatively simple. The method proceeds along the same lines when multiple functionals are of interest, as long as the multiple constraints on the model parameters are not dependent.

The terms ϵ_{p+q} and ϵ_{p+q-1} are generally functions of the unknown parameters, i.e., $\epsilon_{p+q} = \epsilon_{p+q}(\theta)$ and $\epsilon_{p+q-1} = \epsilon_{p+q-1}(\beta)$. In general, unknown parameters in the Bartlett factor may be replaced by consistent estimators without affecting the order of approximation. The technique of Lagrange multipliers can be used to obtain $\tilde{\theta}$, which is also known to be consistent under fairly general conditions (Silvey, 1959). For the parameter function setting, we can regard a subset of the original parameter vector as a nuisance vector. Under H_0, $\sup_{\beta \in R^{p+q-1}} \ell(\beta) = \sup_{\theta : h(\theta)=0} \ell(\theta)$, so that we can regard β as the nuisance vector. Hence, provided the conditions for ML estimation of θ hold, the usual asymptotic properties also continue to apply to the parameter function setting.

For the case of a single functional, standard results can be used to establish the efficacy of the Bartlett correction. This follows since the null parameter space can be parametrized in terms of the functional (as a new parameter) and $p + q - 1$ other parameters (not necessarily the same $p + q - 1$ parameters from the unconstrained parameter space). Suitable one-to-one parametrizations can often be found in this case. For multiple functionals, the preceding approach may not be so conveniently adapted. We therefore offer the following verification for the single functional setting; its extension to multiple functionals is straightforward.

We use the result that the rth cumulant of the LRS, $2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_1)\}$, is given by

$$\kappa_r = 2^{r-1}(r-1)!q\{1 + q^{-1}(\epsilon_{p+q} - \epsilon_p)\}^r + O(n^{-2}).$$

Let $\omega = f(\theta) \in \Re$. As we have previously shown, under H_0, the profile LRS for $f(\theta)$ is

$$W(\xi) = 2\{\ell(\hat{\theta}) - \ell(\tilde{\theta})\} = 2\{\ell(\hat{\theta}) - \ell(\hat{\beta})\},$$

where $\theta \in \Re^{p+q}$ and $\beta \in \Re^{p+q-1}$. Alternatively, we observe that $\ell(\hat{\theta}) = \ell(\hat{\omega})$. To see this geometrically, recall the plot of level curves (Figure 1.1) in the example of Section 1.4.1. Mathematically, by the invariance principle, $f(\hat{\theta})$ is the MLE of $f(\theta)$. Also, $\ell(\hat{\beta}) = \ell(\omega, \tilde{\theta}_1, ..., \tilde{\theta}_{p+q-1})$ via $h(\tilde{\theta}) = 0$. Therefore W is equivalent to $2\{\ell(\hat{\omega}) - \ell(\omega)\}$. The rth cumulant of W is therefore $2^{r-1}(r-1)!\{1 + (\epsilon_{p+q} - \epsilon_{p+q-1})\}^r + O(n^{-2})$ (cf. McCullagh, 1987). By neglecting terms of order n^{-2} or less. the cumulant generating function (c.g.f.) of W is given by

$$\begin{aligned} K_W(t) &= t\{1 + (\epsilon_{p+q} - \epsilon_{p+q-1})\} + \frac{2t^2}{2!}\{1 + (\epsilon_{p+q} - \epsilon_{p+q-1})\}^2 \\ &+ \frac{8t^3}{3!}\{1 + (\epsilon_{p+q} - \epsilon_{p+q-1})\}^3 + \ldots \end{aligned}$$

$$= \sum_{r=1}^{\infty} \frac{t^r}{r} 2^{r-1} \{1 + (\epsilon_{p+q} - \epsilon_{p+q-1})\}^r$$

and hence the c.g.f. of W' is

$$K_{W'}(t) = \sum_{r=1}^{\infty} \frac{t^r}{r} 2^{r-1},$$

the c.g.f. of a χ_1^2 random variable.

In general, for testing $H_0 : h(\theta) = 0$, where $h(\theta) = (h_1(\theta), ..., h_r(\theta))^T$ $(r < p+q)$, the Bartlett adjustment is given by $r + \epsilon_{p+q} - \epsilon_{p+q-r}$. We essentially utilize the implicit function theorem for systems of equations. In this case, it is also known that $W(\xi)$ is distributed approximately as χ_r^2 (Silvey, 1959). The above argument can be easily modified to verify the efficacy of the correction in this general case.

Now suppose we want to compute the Bartlett factor for testing $H_0 : h(\theta) = 0$, where h has ith component $h_i(\theta)$, $i = 1, ..., r$. Assuming that the conditions of the implicit function theorem are satisfied, we have

$$\theta^{p+q+1-i} = F^i(\theta^1, ..., \theta^{p+q-r})$$

for $i = 1, ..., r$. Protter and Morrey (1991) show that, by applying the chain rule, we get

$$\frac{\partial h_i}{\partial \theta^v} + \sum_{k=p+q+1-r}^{p+q} \left\{ \frac{\partial h_i}{\partial \theta^k} \frac{\partial F^k}{\partial \theta^v} \right\} = 0,$$

for $i = 1, ..., r$, $v = 1, ..., p+q-r$. This can be treated as a system of $p+q-r$ equations in $p+q-r$ unknowns, $\{\frac{\partial F^k}{\partial \theta^p}\}$. Since the determinant of the coefficients $\{\frac{\partial h_i}{\partial \theta^k}\}$, $k = p+q+1-r, ..., p+q$, is also assumed not to vanish at the constrained MLE, this system of equations can be solved uniquely. For the case of one functional of interest, we can write $\theta^{p+q} = F(\theta^1, ..., \theta^{p+q-1})$. To calculate the terms in ϵ_{p+q}

and ϵ_{p+q-1}, we apply the preceding result to obtain

$$\frac{\partial F}{\partial \theta^i} = -\frac{\partial h/\partial \theta^i}{\partial h/\partial \theta^{p+q}}$$

for $i = 1, ..., p+q-1$. Computation of the Bartlett factor is certainly more laborious when θ^{p+q} cannot be expressed explicitly as a function of the remaining parameters, but should still be straightforward.

2.3.2 Computational Note

The modified algorithm for computing Bartlett-adjusted confidence intervals for $f(\theta) \in \Re$ is:

1. Set the tolerance level TOL
2. Set the value of the Lagrange multiplier ξ (or equivalently, ω)
3. Calculate $\tilde{\theta}(\xi)$
4. Compute the Bartlett factor $1 + b = 1 + b(\tilde{\theta})$
5. Compute $W' = W(\xi)/(1 + b)$
6. Compare W' with the $(1-\alpha)$−quantile, $\chi^2_{1,1-\alpha}$. If $|W' - \chi^2_{1,1-\alpha}| < TOL$, stop the iteration and accept the current ξ value. Otherwise, compute the updated value of ξ and return to step 3.

From the algorithm we obtain (ξ^-, ξ^+), from which we can compute the endpoints, $f(\tilde{\theta}(\xi^-))$ and $f(\tilde{\theta}(\xi^+))$ of the Bartlett-adjusted, profile likelihood-based CI for $f(\theta)$. Obvious changes to the algorithm are required for the case $f(\theta) \in \Re^r$, $1 < r < p + q$.

Table 2.1: Occurrence of acute mononucleosis-like syndrome and seroconversion (from Tindall *et. al*, 1988).

Occurrences of acute clinical illness	Cases	Controls	Total
≥ 1	36	10	46
0	3	15	18
Total	39	25	64

2.4 Examples

2.4.1 Log-odds Ratio in Binomial Sampling

The data summarized in Table 2.1 were collected by Tindall *et al.* (1988) as part of a case-control study concerning the possible association between the occurrence of an acute mononucleosis-like syndrome and the event of seroconversion in individuals at high risk of infection with the human immunodeficiency virus. Type 1 (HIV-1). The researchers used retrospective interview techniques to determine that. during the study period, of 39 male homosexuals whose blood serum changed from HIV- (seronegative) to HIV+ (seropositive), a total of 36 experienced at least one occurrence of acute clinical illness. Among the 25 controls, HIV- male homosexuals who did not undergo seroconversion during the study period. only 10 individuals suffered one or more episodes of acute clinical illness. In this study, the relative risk. ψ. of seroconversion among individuals at risk of HIV infection who experienced one or more episodes of acute clinical illness. might be of interest.

The observed data consist of independent binomial observations, $X_i \sim B(n_i, p_i)$. for $i = 1, 2$. The functional of interest, ψ, is the odds ratio, $\left\{\frac{p_1/(1-p_1)}{p_2/(1-p_2)}\right\}$, or equivalently, the log-odds ratio, $\omega = \log\psi$. It is numerically easier to work with ω for this example; hence, a profile likelihood-based CI for ψ will be obtained by inverting the end-points of the interval estimate for ω (the details are similiar to Example 1.1 of section 1.4). The constrained MLEs for p_i, $i = 1, 2$, in this case are

$$\tilde{p}_1 = \frac{x_1 - \xi}{n_1}, \qquad \tilde{p}_2 = \frac{x_2 + \xi}{n_2}.$$

corresponding to a fixed value, ξ, of the Lagrange multiplier. The Bartlett factor for this problem may also be found by adopting a suitable 1-1 reparametrization of the parameters. For example, we can reparametrize the model in terms of ω and $\delta = \frac{p_2}{1-p_2}$, treating the latter term as the nuisance parameter and working with the profile likelihood for ω. The LRS for testing a fixed value of ω is given by $2\{\ell(\hat{\omega}.\hat{\delta}) - \ell(\omega.\tilde{\delta})\}$. However. this approach does not yield a simple (closed-form) solution for the constrained MLE of δ. A univariate root-finding routine can be used in this case. but much greater numerical effort may be required for multi-parameter problems. We may also note that. for this parametrization, it is not difficult to find the value of the Lagrange multiplier corresponding to a fixed ω. This indirect approach may be used to avoid the numerical method of obtaining $\tilde{\delta}$. However. it is not always easy to implement this approach for other functionals of interest (such as the NNT in the next example). In contrast. the Lagrange multiplier argument can lead to simpler solutions for some multi-parameter problems.

Note that ϵ_2 may be computed easily using the simplified form for ϵ_p (see section 2.2). For ϵ_1, we let $p_1 = \frac{\exp(\beta)}{1+\exp(\beta)}$, which implies, via the constraint, that $p_2 = \frac{\exp(\beta-\omega)}{1+\exp(\beta-\omega)}$. In this case, further simplification results since

$$\begin{aligned}(\lambda_{111})_1 &= -n_1 A(\beta)[1 - A(\beta)]A_1 - n_2 A(\beta - \omega)[1 - A(\beta - \omega)]A_2 \\ &= -\nu_1(1 - 6\ p_1 + 6\ p_1^2) - \nu_2(1 - 6\ p_2 + 6\ p_2^2)\end{aligned}$$

$$\begin{aligned}
&= \lambda_{1111}, \\
(\lambda_{11})_1 &= -n_1 A(\beta)[1 - A(\beta)]\{1 - 2A(\beta)\} - n_2 A(\beta - \omega)[1 - A(\beta - \omega)]\{1 - 2A(\beta - \omega)\} \\
&= -\nu_1(1 - 2\,p_1) - \nu_2(1 - 2\,p_2) \\
&= \lambda_{111}, \\
(\lambda_{11})_{11} &= -n_1 A(\beta)[1 - A(\beta)]A_1 - n_2 A(\beta - \omega)[1 - A(\beta - \omega)]A_2 \\
&= \lambda_{1111},
\end{aligned}$$

where $\nu_i = n_i p_i(1 - p_i)$, $A(x) = \frac{\exp(x)}{1+\exp(x)}$, and

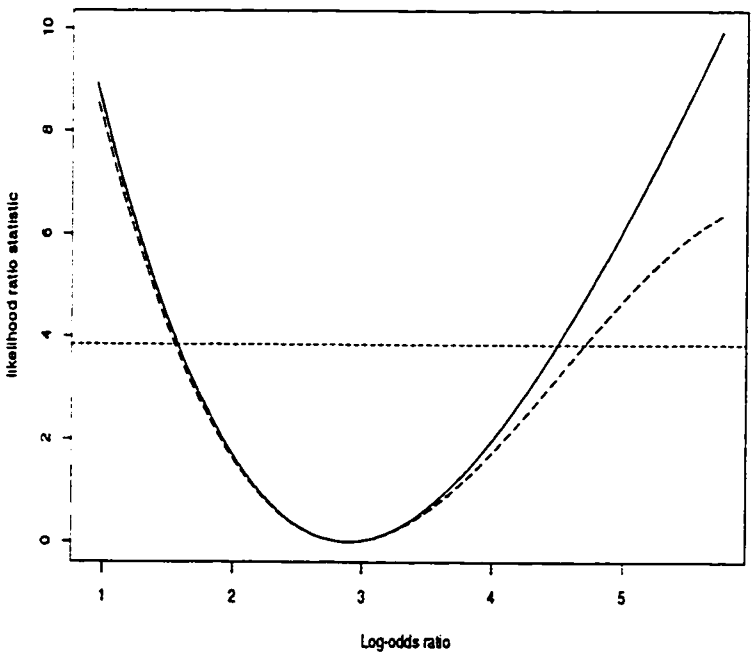
$$A_1 = 1 - 6n_1 A(\beta) + 6n_1 A^2(\beta),$$
$$A_2 = 1 - 6n_2 A(\beta - \omega) + 6n_2 A^2(\beta - \omega).$$

Hence we obtain

$$\begin{aligned}
\epsilon_2 &= \frac{1}{6\nu_1} + \frac{1}{6\nu_2} - \frac{1}{6n_1} - \frac{1}{6n_2}, \\
\epsilon_1 &= \frac{1}{4}\frac{\lambda_{1111}}{\lambda_{11}^2} - \frac{5}{12}\frac{\lambda_{111}^2}{\lambda_{11}^3} \\
&= -\frac{1}{4(\nu_1 + \nu_2)} + \frac{3}{2(\nu_1 + \nu_2)^2}\left(\frac{\nu_1^2}{n_1} + \frac{\nu_2^2}{n_2}\right) \\
&+ \frac{5}{12(\nu_1 + \nu_2)^3}\left\{\nu_1(1 - 2p_1) + \nu_2(1 - 2p_2)\right\}^2.
\end{aligned}$$

The constrained MLEs, $\tilde{p}_i$, $i = 1, 2$, can be substituted for p_i to obtain a consistent estimate of the Bartlett factor. Applying these results to the data, we obtain the adjusted 95% CI for the odds ratio as (4.84, 110.5), while the unadjusted interval is (4.87, 90.5). Figure 2.1 shows the LRS and Bartlett-adjusted LRS for the log-odds ratio. For our data, the adjustment is most pronounced at the upper endpoint of the interval. In contrast, the standard interval based on the maximum likelihood estimate for ψ is (4.33, 74.76), while the exact interval is (3.80, 109.5) (Matthews, 1988b). The method for obtaining the exact interval is based on a conditional likelihood argument and is discussed, for example, in Chapter 4 of Cox (1970). While the LRS appears to be reasonably approximated by a quadratic curve, the ML interval is too narrow; in the long run, the ML method presumably results in a coverage that is below nominal level.

Figure 2.1: Likelihood ratio statistic for the odds ratio, based on the data in Table 2.1. The solid curve is the uncorrected LRS and the dashed curve is the corresponding Bartlett-corrected LRS. The height of the horizontal dashed line is 3.841 units.



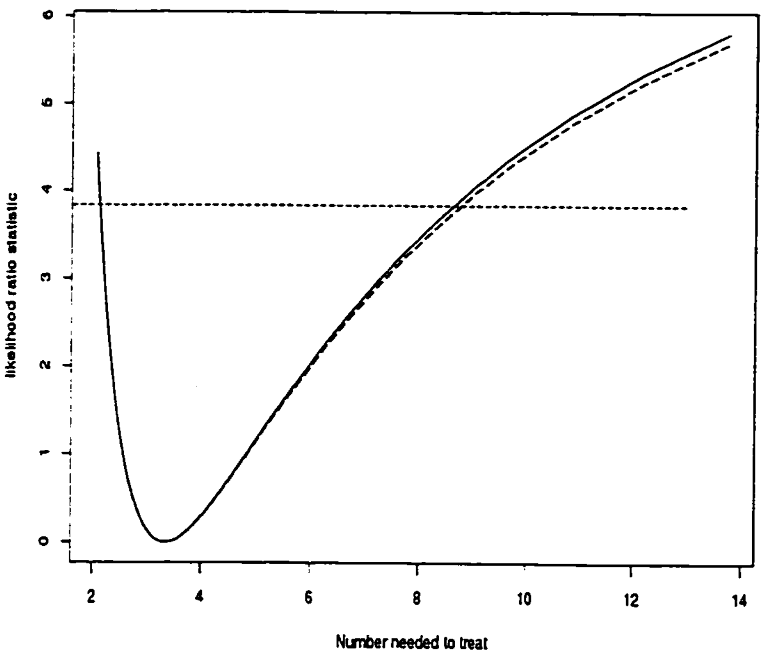
2.4.2 Number Needed to Treat

A Bartlett-adjusted interval estimate for the NNT parameter in the endoscopic injection trial (section 1.4) can similiarly be obtained. We find that

$$\begin{aligned}
\epsilon_2 &= \frac{1}{6\nu_1} + \frac{1}{6\nu_2} - \frac{1}{6n_1} - \frac{1}{6n_2}, \\
\epsilon_1 &= \left(\frac{n_1^2}{\nu_1} + \frac{n_2^2}{\nu_2}\right)^{-2} \left\{ \frac{n_1^4(1 - 3p_1 + 3p_1^2)}{2\nu_1^3} + \frac{n_2^4(1 - 3p_2 + 3p_2^2)}{2\nu_2^3} \right\} \\
&- \frac{1}{3}\left(\frac{n_1^2}{\nu_1} + \frac{n_2^2}{\nu_2}\right)^{-3} \left\{ \frac{n_1^3(1 - 2p_1)}{\nu_1^2} + \frac{n_2^3(1 - 2p_2)}{\nu_2^2} \right\}^2,
\end{aligned}$$

where the ν_i are as previously defined. The constrained MLEs for p_i, $i = 1, 2$ (cf. Chapter 1) can be substituted into the preceding formulae to obtain a consistent estimate of the Bartlett factor. Using the data from section 1.4, the unadjusted and Bartlett-corrected 95% confidence intervals are, respectively, (2.11, 8.63) and (2.10, 8.77). Thus, the correction results in a slight elongation of the interval estimate; Figure 2.2 displays the plot of the relevant statistics. By way of comparison, the usual interval based on the normal approximation is (2.08, 8.33).

Figure 2.2: Likelihood ratio statistic for the NNT, based on the data in Table 1.1. The solid curve is the uncorrected LRS and the dashed curve is the corresponding Bartlett-corrected statistic. The height of the horizontal dashed line is 3.841 units.



2.4.3 Linear Predictor in a GLM

Cordeiro (1983) derived a Bartlett factor for the regression coefficients in the context of generalized linear models (GLM). Based on his results, and using the technique developed above, a Bartlett factor for testing hypotheses of the form $H_0 : \beta^T \mathbf{x}_0 = \omega_0$ can be obtained, where $\mathbf{x}_0$ and ω_0 are known. Some notation is first introduced in the following.

The probability density function (p.d.f.) of a response variable Y_i is of the form

$$\pi(y;\theta_i,\phi_i) = \exp\{\phi_i[y\theta_i - b(\theta_i) + c(y)] + d(\phi_i, y)\}$$

where $b(\cdot)$, $c(\cdot)$ and $d(\cdot)$ are known functions. A more common parametrization of π in the GLM literature is that of McCullagh and Nelder (1989), viz.

$$\pi(y;\theta_i,\phi_i) = \exp\{(y\theta_i - b(\theta_i))/a(\phi_i) + c(\phi_i, y)\}$$

for some known functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$. However, this slightly different parametrization does not require major revisions in the following derivation of the Bartlett factor. As will become clear below, we only need to substitute $a^{-1}(\phi_i)$ for ϕ_i in the formulae for λ_{rs}, λ_{rst}, etc., and the matrix $diag\{\phi_1, \ldots, \phi_n\}$ becomes $diag\{a^{-1}(\phi_1), \ldots, a^{-1}(\phi_n)\}$. For GLMs, the linear predictor, $\eta_i = \beta^T \mathbf{x}_i$, $\beta = (\beta_1, \ldots, \beta_p)^T$, is related to the mean of Y_i by a suitable 1-1 function, i.e., $E(Y_i) = \mu_i = g(\eta_i)$. By setting the score equation to zero and solving, we obtain $\mu_i = \frac{\partial b(\theta_i)}{\partial \theta_i}$. It can be shown that, for any GLM,

$$\begin{aligned}
\lambda_{rs} &= -\sum_{i=1}^{n} \phi_i w_i x_{ir} x_{is}, \\
\lambda_{rst} &= -\sum_{i=1}^{n} \phi_i \left[3\frac{d\mu}{d\eta}\frac{d^2\theta}{d\eta^2} + \left(\frac{d\theta}{d\eta}\right)^3 \frac{d^2\mu}{d\theta^2}\right]_i x_{ir} x_{is} x_{it}, \\
\lambda_{rstu} &= -\sum_{i=1}^{n} \phi_i \left[\frac{d^2 w}{d\eta^2} + \frac{d^2\theta}{d\eta^2}\frac{d^2\mu}{d\eta^2} + 2\frac{d^3\theta}{d\eta^3}\frac{d\mu}{d\eta}\right]_i x_{ir} x_{is} x_{it} x_{iu},
\end{aligned}$$

where $w = (d\mu/d\eta)^2/V$, and $V = d\mu/d\theta$ is known as the variance function. The remaining terms in the Bartlett factor are easily obtained from these terms. Let $K = \{-\lambda_{rs}\}$, $W = diag\{w_1, \ldots, w_n\}$ and $\Phi = diag\{\phi_1, \ldots, \phi_n\}$. Applying Lawley's (1956) results to the GLM setting, Cordeiro showed that

$$\epsilon_p = \frac{1}{4} tr(\Phi H Z_d^2) - \frac{1}{3} 1^T \Phi G Z^{(3)} (F + G) \Phi 1 + \frac{1}{12} 1^T \Phi F (2Z^{(3)} + 3 Z_d Z Z_d) F \Phi 1, \quad (2.2)$$

where $Z = X(X^TW\Phi X)^{-1}X^T = \{z_{ij}\}$, $Z^{(3)} = \{z_{ij}^3\}$ and

$$\begin{aligned}
H &= diag\left\{ \frac{1}{V}\frac{d^2\mu}{d\eta^2}\left[\frac{d^2\mu}{d\eta^2} - 4w\frac{dV}{d\mu}\right] + w^2\left[\frac{2}{V}\left(\frac{dV}{d\mu}\right)^2 - \frac{d^2V}{d\mu^2}\right] \right\}. \\
F &= diag\left\{ \frac{1}{V}\frac{d\mu}{d\eta}\frac{d^2\mu}{d\eta^2} \right\}, \\
G &= diag\left\{ \frac{1}{V}\frac{d\mu}{d\eta}\frac{d^2\mu}{d\eta^2} - \frac{1}{V^2}\frac{dV}{d\mu}\left(\frac{d\mu}{d\eta}\right)^3 \right\}. \\
Z_d &= diag\{z_{11}, \dots, z_{nn}\}.
\end{aligned}$$

Now consider the null hypothesis $H_0 : \beta^T\mathbf{x}_0 = \omega$, for some fixed value ω and given $\mathbf{x}_0$. Without loss of generality, assume that $x_{0p} \neq 0$. Then we may write

$$\beta_p = \frac{\omega - \sum_{i=1}^{p-1}\beta_i x_{0i}}{x_{0p}}.$$

Hence

$$\begin{aligned}
\beta^T\mathbf{x}_i &= \beta_1(x_{i1} - \frac{x_{01}x_{ip}}{x_{0p}}) + \ldots + \beta_{p-1}(x_{i,p-1} - \frac{x_{0,p-1}x_{ip}}{x_{0p}}) + \frac{\omega x_{ip}}{x_{0p}} \\
&= \gamma'\mathbf{v}_i + \omega v_{ip},
\end{aligned}$$

where $v_{ip} = \frac{x_{ip}}{x_{0p}}$, $v_{ij} = (x_{ij} - x_{0j}\frac{x_{ip}}{x_{0p}})$ for $j = 1, \dots, p-1$, $\mathbf{x}_i = (x_{i1}, \dots, x_{i,p-1})^T$, and $\gamma = (\beta_1, \dots, \beta_{p-1})^T$. Let $X_0 = \{v_{ij}\}$ denote the design matrix under H_0; Z_0, $Z_0^{(3)}$ and Z_{0d} are accordingly defined. Under the null hypothesis, the general form of λ_{rs}, λ_{rst}, λ_{rstu} etc., is unchanged, except for having v_{ij} in place of x_{ij}, $j \neq p$. This follows because these terms are made up of two types of partial derivatives. The first type (which includes $\frac{d\mu}{d\eta}$, $\frac{d\theta}{d\eta}$, $\frac{dV}{d\mu}$, for example) depends only on the choice of link function, or the relation between the mean response and the canonical parameter. Their general form is therefore unaffected under H_0. The remaining (non-zero) partial derivatives are $\{\frac{\partial\eta}{\partial\beta_r}\}$, which under H_0 are given by $\{v_r = (v_{1r}, \dots, v_{nr})^T\}$.

Therefore, the structure of H, F and G are unchanged under the null.

Except for substituting the constrained versions of the design matrices for the unrestricted ones, the derivation of ϵ_{p-1} follows the steps taken for ϵ_p exactly. Using these facts, it is not difficult to verify that ϵ_{p-1} is computed by substituting Z_0. $Z_0^{(3)}$ and Z_{0d} into equation (2.2). Therefore

$$\begin{aligned}\epsilon_p - \epsilon_{p-1} &= \frac{1}{4}tr\{\Phi H(Z_d^2 - Z_{0d}^2)\} - \frac{1}{3}1^T\Phi G(Z^{(3)} - Z_0^{(3)})(F+G)\Phi 1 \\ &+ \frac{1}{12}1^T\Phi F\{2(Z^{(3)} - Z_0^{(3)}) + 3(Z_d Z Z_d - Z_{0d}Z_0 Z_{0d})\}F\Phi 1.\end{aligned}$$

In this case. we note also that the reparametrization $(\beta_1, \ldots, \beta_p) \rightarrow (\beta_1, \ldots, \beta_{p-1}, \delta)$. where $\delta = \beta^T\mathbf{x}_0$, is 1-1 and leads to ϵ_{p-1} in a straightforward way as well. However. it can be more difficult to find suitable 1-1 reparametrizations if one is interested in a simultaneous confidence region for multiple linear predictors. This problem is briefly discussed below. [*Note.* For given values of the multiple linear predictors. it is helpful to link confidence region construction to a test of hypothesis. Under the null. one is therefore testing a subset of the transformed parameters. i.e.. the linear predictors of interest.]

Since the constrained MLE $\tilde{\beta}$ is a consistent estimator for β under regularity conditions. we can substitute it into the expression for the Bartlett factor. The S-Plus glm() function can be used with an offset. We note briefly that the approach of Nyquist (1991) or a Lagrange multiplier argument can also be used to obtain an iterative procedure for computing $\tilde{\beta}$. The latter follows from the same approach used in Kim and Taylor (1995). Let $\hat{\beta} = (X^TW\Phi X)^{-1}X^TW\Phi\mathbf{z}$ denote the unrestricted MLE of β, where $\mathbf{z}$ is the adjusted dependent variate. The iterative scheme is given by

$$\tilde{\beta}^{(m+1)} = \hat{\beta} + (X^T W \Phi X)^{-1} \mathbf{x_0}^T \{\mathbf{x_0} (X^T W \Phi X)^{-1} \mathbf{x_0}^T\}^{-1} \; \{\omega - \hat{\beta}^T \mathbf{x_0}\} .$$

where the elements on the right side of the equation are evaluated at $\tilde{\beta}^{(m)}$.

Bartlett factors for linear combinations of linear predictors can be derived in the same way. An important example of a functional in this class is the difference in risk score between two subjects with covariate values $\mathbf{x_1}$ and $\mathbf{x_2}$. In the GLM context. this can represent the difference in means. rates or log-odds ratio between two individuals with distinct covariate values. The null hypothesis is $H_0 : \beta^T(\mathbf{x_1} - \mathbf{x_2}) = \omega \in \Re$. Since we can express β_p as

$$\beta_p = (x_{1p} - x_{2p})^{-1} \{\omega - \sum_{j=1}^{p-1} \beta_j (x_{1j} - x_{2j})\} .$$

the form of ϵ_{p-1} here is identical to the earlier expression. with the appropriate modification to the design matrix.

In industrial quality improvement experiments. quite often one is interested in the optimum combination of factor levels that maintains the mean response at some desired level. while minimizing process variability. In this context. it might be of interest to compare the variability of the product (or process) at two prescribed settings of the factors involved. In the following example. we assume constant dispersion. although the joint modelling of both mean and dispersion as functions of covariates might be a better approach in the industrial setting; see. for example. McCullagh and Nelder (1989. p. 357). When the variance of the responses is some suitable function of the mean (e.g. Gamma. inverse Gaussian distributions). we can usually obtain the required Bartlett-adjusted interval estimates. Consider the case of Gamma random variables with different means but common shape parameter.

The Gamma(μ_i, ν) density is given by

$$\pi(y_i; \mu, \nu) = \Gamma^{-1}(\nu)\left(\frac{\nu}{\mu_i}\right)^{\nu} y^{\nu-1} \exp\left(-\frac{\nu}{\mu_i} y_i\right), \qquad y_i \geq 0,$$

and therefore $\mathrm{var}(y_i) = \mu_i^2/\nu$. We assume the usual GLM set-up, with $\mu_i = g(\eta_i)$, $\eta_i = \beta^T \mathbf{x_i}$. For analytically simpler results, we consider the null hypothesis H_0 : $\log \frac{var(y_1)}{var(y_2)} = \omega$, i.e., H_0 : $\log \frac{\mu_1^2}{\mu_2^2} = \omega$. For the reciprocal link function i.e., the canonical link, H_0 is equivalent to specifying that $2 \log \frac{\eta_2}{\eta_1} = \omega$, which implies

$$\beta_p = (x_{2p} - \gamma x_{1p})^{-1} \sum_{j=1}^{p-1} \beta_j (\gamma x_{1j} - x_{2j}),$$

where $\gamma = \exp(\omega/2)$. Hence, under H_0,

$$\eta_i \equiv \beta^T \mathbf{x_i} = \sum_{j=1}^{p-1} \beta_j \{x_{ij} + (\frac{x_{ip}}{x_{2p} - \gamma x_{1p}})(\gamma x_{1j} - x_{2j})\}.$$

By defining the elements of the design matrix, v_{ij}, as the expression in the brace brackets, the computation of ϵ_{p-1} proceeds in the usual way. Another common choice for the link function is the log link, $\log \mu = \eta$, in which case $H_0 : \eta_1 = \eta_2 + \omega/2$. The details are similiar to the previous case and are omitted.

A Bartlett-adjusted confidence region for a set of linear predictors may also be obtained. This can be useful, for example, in a logistic regression context in which one is interested in making simultaneous inference on a finite number J of survival probabilities, p_j, $j = 1, \ldots, J$, where $p_j = \frac{\eta_j}{1+\eta_j}$, $\eta_j = \beta^T \mathbf{x_j}$ and $\mathbf{x_j}$ are fixed and known covariates. Consider then $q < p$ linear predictors of interest, $\eta_i = \beta^T \mathbf{x_i}$, $i = 1, \ldots, q$. We let X denote the $q \times p$ matrix with ith row given by $\mathbf{x_i}^T$. Given $\{\omega_i\}$ and $\{x_{ij}\}$, $i = 1, \ldots, q$, $j = 1, \ldots, p$, we have the set of q constraints

$\{\omega_i = \beta^T x_i\}$. We rewrite the constraints as

$$X_q\beta^{(q)} = \Omega - X_{p-q}\beta^{(p-q)}$$

where $X_q = \{x_{ij}\}$, $i, j = 1, \ldots, q$, $X_{p-q} = \{x_{ij}\}$, $i, j = q+1, \ldots, p$, $\Omega = \{\omega_i\}$ and $\beta^{(r)}$ are the partitions of the coefficient vector conforming to each X matrix. For q small, this can be solved with some effort for $\beta^{(q)}$ in terms of $\beta^{(p-q)}$, and the method proceeds in a straightforward way.

Remark. Barndorff-Nielsen and Blaesild (1986) derived a method of obtaining the Bartlett factor in a distinct problem setting, although their applications also included a GLM. Basically, their correction factor applies for testing the null hypothesis that the original p model parameters can be expressed as linear combinations of $q < p$ parameters. Suppose the original model is indexed by $\theta \in \Re^p$. The null model is reparametrized as $\theta = \beta \mathbf{y}$, where $\beta \in \Re^q$, $q < p$ and $\mathbf{y}$ is a known vector. As shown in their paper, the constrained portion of the Bartlett factor may be obtained from the quantities computed in the unrestricted portion, using a series of translation formulae. Hence their method and ours address different problems. For example, their method does not handle the case of obtaining a Bartlett-adjusted confidence region for p linear predictors (or a subset of them), given the fixed covariate values.

The following examples illustrate some of the procedures discussed in this subsection.

Example 2.4.3a Mean Lifetime in Exponential Survival

For uncensored exponential lifetime data, Lawless (1982, example 6.3.2) obtained interval estimates for the mean survival time based on an exact method and a large-sample normal approximation. However, due to the moderate size of the sample, the normal approximation did not perform very well. Using the same data, we illustrate below the effect of a Bartlett correction on interval estimation for the mean lifetime of a given subject.

For the exponential regression model, it is assumed that, given $\mathbf{x}$, the p.d.f. of T is

$$f(t|\mathbf{x}) = \mu^{-1}\exp\left(-\frac{t}{\mu}\right), \qquad t > 0 .$$

Following Lawless (1982), we also assume a logarithmic link function, $\log\mu = \beta^T\mathbf{x}$. It is straightforward to verify for this model that, except for Z (and hence $Z^{(3)}$ and Z_d), the other matrices in equation (2.2) are given by the identity matrix. Let $\omega_0 = \exp(\beta_0 + \beta_1 x_0)$ denote the mean lifetime for a subject with covariate value x_0. We have $Z = X(X^TX)^{-1}X^T$, and using (2.2), we obtain

$$\begin{aligned}\epsilon_2 &= \frac{1}{6}1^T Z^{(3)}1 - \frac{1}{4}\{tr(Z_d^2) - 1^T Z_d Z Z_d 1\} \\ &= \frac{1}{6}\sum_{i,j} z_{ij}^3 - \frac{1}{4}1^T Z_d(I - Z)Z_d 1.\end{aligned}$$

where I denotes the identity matrix. Under the constrained model, the linear predictor for the ith subject is $\eta_i = \log\omega_0 + \beta_1(x_i - x_0)$. The design matrix for the constrained model, X_0, has elements $\{x_i - x_0\}$. Similiarly, $Z_0 = X_0(X_0^TX_0)^{-1}X_0^T$. Therefore

$$\epsilon_1 = \frac{1}{6}\sum_{i,j} z_{0ij}^3 - \frac{1}{4}1^T Z_{d0}(I - Z_0)Z_{d0}1.$$

For this example, both portions of the Bartlett factor do not depend on β (ϵ_1 does not even depend on ω). The term ω affects only the constrained MLEs for the regression coefficients, i.e., it enters in the LRS $2\{\ell(\hat{\beta}) - \ell(\tilde{\beta})\}$.

Table 2.2: Failure time, t_i, in weeks and white blood count, from Feigl and Zelen (1965).

t_i	WBC	x_i	t_i	WBC	x_i
65	2300	3.36	143	7000	3.85
156	750	2.88	56	9400	3.97
100	4300	3.63	26	32000	4.51
134	2600	3.41	22	35000	4.54
16	6000	3.78	1	100000	5.00
108	10000	4.02	1	100000	5.00
121	10000	4.00	5	52000	4.72
4	17000	4.23	65	100000	5.00
39	5400	3.73			

Lawless (1982) considered the above data from Feigl and Zelen (1965), consisting of the failure time in weeks and white blood count (WBC) of 17 leukaemia patients. In the table, $x = \log_{10} \text{WBC}$. Using these data, we consider estimating the mean life, μ_0, for a subject with $x = 4$. The approximate 95% CI based on the profile likelihood for μ_0 is (36.4, 96.1). The Bartlett correction yields the interval (35.9, 98.2), while the exact confidence interval was found by Lawless to be (37.9, 102.0). At the lower end-point, both likelihood-based intervals yield very similiar estimates. The Bartlett-adjusted interval is however more accurate at the upper limit, and its length is also closer to that of the exact interval.

Example 2.4.3b ED Level in Logistic Regression

Duncan *et al.* (1984) report the results of a study initiated to investigate the effect of premedication on the dose requirement in children of the anaesthetic thiopentone. The study involved observations on 490 children aged 1-12 years. These patients were divided into 4 groups, three of which received different types of premedication.

Table 2.3: Dose-response data for patients treated with TDP.

Dose (mg/kg)	Number of Responses	Total
2.0	7	15
2.5	14	21
3.0	15	20
3.5	13	14
4.0	11	12
4.5	5	8
5.0	10	11
5.5	22	22
6.0	13	14

No premedication was administered to the fourth group of patients. All the children subsequently received an injection of 2.0-8.5 mg/kg of thiopentone in steps of 0.5 mg/kg. The anaesthetic was administered to each patient over a 10-second interval. Twenty seconds after the injection, the eyelash reflex was tested; if the eyelash reflex was disabled, the patient was considered to have responded to the anaesthetic.

In the following, we apply our method to only a subset of the data, viz., to a group of 137 patients who were premedicated orally with TDP (trimeprazine, droperiodol and physeptone) and atropine; see Table 2.3. A binary logistic regression model is often appropriate for the analysis of dose-response data. We assume the observed data consist of independent $Y_i \sim \mathrm{B}(n_i, p_i)$, where

$$p_i = \frac{\exp\{\beta(d_i - \gamma)\}}{1 + \exp\{\beta(d_i - \gamma)\}},$$

for $i = 1, \ldots, k$. Equivalently, $\lambda_i = \operatorname{logit}(p_i) = \beta(d_i - \gamma)$. A better parametrization is given by $\operatorname{logit}(p_i) = \alpha + \beta(d_i - \bar{d})$, where $\bar{d}$ is the average dose level: however, we stick with the present model for purposes of illustration. Define the parameter of interest, the ED100p, to be the dose ω_p satisfying

$$\text{logit}(p) = \lambda_p = \beta(\omega_p - \gamma),$$

so that $\omega_p = \gamma + \frac{\lambda_p}{\beta}$. The log likelihood function for γ and β is

$$\begin{aligned} l(\gamma, \beta) &= \sum_{i=1}^{k} \{y_i \log p_i + (n_i - y_i) \log(1 - p_i)\} \\ &= \sum_{i=1}^{k} y_i \beta(d_i - \gamma) - \sum_{i=1}^{k} n_i \log[\, 1 + \exp\{\beta(d_i - \gamma)\}\,]. \end{aligned}$$

For this model, the iterative weights $w_i = n_i p_i(1 - p_i)$, and $\Phi = diag\{n_i\}$. The joint unrestricted MLEs can be obtained using commonly available software, e.g. glm() in S-Plus. For a fixed value of ω_p, we substitute $\gamma = \omega_p - \frac{\lambda_p}{\beta}$ into the linear predictor, and the constrained estimate of β can routinely be obtained using the glm() function by specifying a regression model with offset and no intercept term. To obtain the Bartlett factor for a fixed value of ω, we note that

$$\begin{aligned} H &= diag\{-p(1-p)(6p^2 - 10p + 1)\}, \\ F &= diag\{p(1-p)(1-2p)\}, \\ G &= 0, \end{aligned}$$

where $p = \frac{\exp(\eta)}{1+\exp(\eta)}$. Thus

$$\begin{aligned} \epsilon_2 &= \frac{1}{4} tr(\Phi H Z_d^2) + \frac{1}{12} 1^T \Phi F (2Z^{(3)} + 3Z_d Z Z_d) F \Phi 1, \\ &= -\frac{1}{4} \sum_i n_i h_{ii} z_{ii}^2 + \frac{1}{6} \sum_{i,j} z_{ij}^3 (n_i f_{ii})(n_j f_{jj}) + \frac{1}{4} \sum_{i,j} z_{ij} (n_i f_{ii} z_{ii})(n_j f_{jj} z_{jj}), \end{aligned}$$

where h_{ii} and f_{ii} are the ith diagonal elements of H and F, respectively. The constrained version of the design matrix is $X_0 = \{d_i - \omega_p\}$. The term ϵ_1 is straightforward to compute, and p_i can be estimated consistently by $\tilde{p}_i = \frac{\exp\{\tilde{\beta}(d_i - \tilde{\gamma})\}}{1+\exp\{\tilde{\beta}(d_i - \tilde{\gamma})\}}$.

Based on the data in Table 2.2, we obtain the following. The unadjusted and Bartlett-adjusted 95% CIs for the ED50 are (1.26, 2.42) and (1.16. 2.43). respectively. The corresponding interval based on the normal approximation is (1.54. 2.59).

Figure 2.3: Likelihood ratio statistic (solid curve) and corresponding Bartlett-corrected statistic (dashed curve) for the ED50. based on the data in Table 2.3. The height of the horizontal dashed line is 3.841 units.

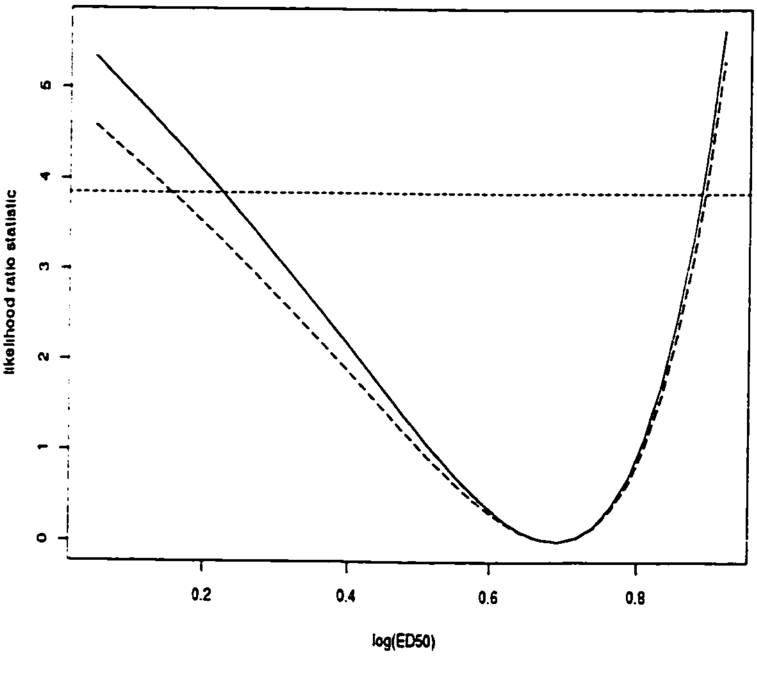


Figure 2.3 shows the LRS and its Bartlett-adjusted version. It also indicates clearly the inadequacy of the normality assumption in this case. The unadjusted 95%

Table 2.4: Inhalation test data from CIBA-GEIGY, from Racine *et al.* (1986).

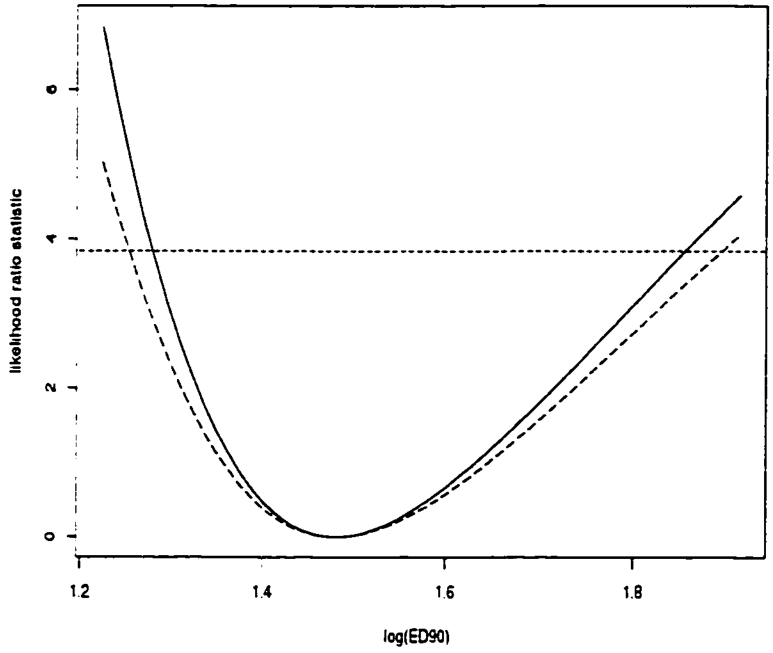
Dose (mg/ml)	Number of Animals	Number of Deaths
422	5	0
744	5	1
948	5	3
2069	5	5

likelihood-based interval for the ED90 is (3.59. 6.40), compared to the adjusted interval of (3.5. 6.66). The interval estimate based on the normal approximation is (3.43. 5.64). Figure 2.4 shows the relevant curves. Compared to the ED50. the Bartlett factor has a larger effect on the interval estimate of the ED90.

To provide an indication of the potential difference that can result from a Bartlett correction in an application, we consider a small sample below. The data summarized in Table 2.4 were obtained from an inhalation acute toxicity test carried out at CIBA-GEIGY and reported in Racine *et al.* (1986). Using the model adopted in Example 2.4.3b. we obtain the unrestricted MLEs. 7.930 and 6.797. for β and γ. respectively. For illustration. we consider interval estimation of the ED50 and ED90. The MLE of the ED50 is 895. and its approximate 95% profile likelihood-based CI is (726.4, 1225.2). The Bartlett-adjusted interval is (670.1, 1451.8). There is a substantial (57%) increase in length of the likelihood-based interval as a result of the Bartlett correction. Even on the log dose scale, the increase is about 48% over the unadjusted interval estimate. This contrasts with the situation in Example 2.4.3b. The normal approximation yields the symmetric interval estimate (880.0. 910.8). which clearly does not reflect the asymmetry in the given likelihood. For the LD90, we obtain an MLE of 1181.1, and a profile likelihood-based 95% CI of (932.9, 2779.1). The Bartlett correction yields the interval estimate

(812.8. 4705.5). These are in sharp contrast to the normal approximation which gives (1072.7, 1300.5).

Figure 2.4: Likelihood ratio statistic (solid curve) and corresponding Bartlett-corrected statistic (dashed curve) for the ED90, based on the data in Table 2.3. The height of the horizontal dashed line is 3.841 units.



For small samples, the Bartlett correction can result in a very conservative interval estimate for the ED50 and ED90, relative to the interval estimates based on the uncorrected profile likelihood and the normal approximation. An indication of this conservative effect is the substantial overlap between the Bartlett-adjusted intervals for the ED50 and ED90 based on the inhalation data. In a subsequent section, we conduct a more systematic evaluation of the Bartlett adjustment in terms of coverage probabilities of the likelihood-based interval estimate.

Example 2.4.3c Quantiles for Normal Samples

Suppose we have a random sample, $Y_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$. We are interested in obtaining an interval estimate for the pth quantile $y_p = \mu + z_p\sigma$, where z_p is the pth quantile of the standard normal distribution. For uncensored data, Lawless (1982) describes an exact method based on the noncentral t distribution. We compare below intervals based on the exact method, the profile likelihood and the Bartlett-adjusted profile likelihood.

Since the normal linear model is a special case of the GLM framework, we can analyze this case using the preceding techniques. However, we also can obtain simple closed-form expressions for the constrained MLEs using standard methods, which we indicate briefly as follows. The score equations based on the augmented log likelihood function are

$$\frac{\partial \ell_\xi}{\partial \mu} = \sigma^{-2} \sum_i (y_i - \mu) - \xi = 0, \tag{2.3}$$

$$\frac{\partial \ell_\xi}{\partial \sigma} = -n\sigma^{-1} + \sigma^{-3} \sum_i (y_i - \mu)^2 - \xi z_p = 0, \tag{2.4}$$

where ξ denotes a fixed value of the Lagrange multiplier. Equation (2.3) implies $\mu = \bar{y} - n^{-1}\xi\sigma^2$, which is substituted into equation (2.4) to yield a quartic equation in σ,

$$n^{-1}\xi^2\sigma^4 - \xi y_p \sigma^3 - n\sigma^2 + (n-1)s^2 = 0\,,$$

where $s^2 = (n-1)^{-1}\sum_i (y_i - \bar{y})^2$. This equation can be solved routinely by standard numerical methods but is more tedious to implement. Alternatively, let $\omega_p = y_p$,

Table 2.5: Interval estimates for selected data from Snook and Ciriello (1991).

	Exact Interval	Profile-based	Bartlett-adjusted
Male	(46.649, 56.025)	(46.374, 55.458)	(46.346, 55.450)
Female	(23.882, 27.461)	(23.756, 27.626)	(23.735, 27.620)

for some fixed ω_p. We can substitute $\mu = \omega_p - u_p\sigma$ into the log likelihood function and maximize it with respect to σ. The resulting score equation

$$\frac{\partial \ell(\sigma)}{\partial \sigma} = -\frac{n}{\sigma} - \frac{1}{2\sigma^2}\{(n-1)s^2 - n(\omega_p - u_p\sigma)^2\} = 0$$

is quadratic and hence admits a closed-form solution for the constrained estimator $\tilde{\sigma}$. For numerical illustration, we consider just two cases from the data of Snook and Ciriello (1991), who conducted an ergonomic trial to study a variety of manual handling tasks consistent with worker capabilities and limitations. The sample size, mean and standard deviation for males performing carrying tasks are, respectively, 38, 32.45 and 10.98. The corresponding data for females are 27, 19.03 and 3.93. Based on the normality assumption and assuming a 95% level of confidence, we obtain the exact, unadjusted and Bartlett-adjusted profile likelihood-based CIs for $y_{0.10}$, shown in Table 2.5. A smaller data set involving 23 'lifetimes' of deep groove ball bearings was also analyzed; the data can be found in Example 5.2.2 of Lawless (1982). For this data set, the exact interval for $y_{0.10}$ is (23.379, 39.592). The corresponding unadjusted and Bartlett-adjusted likelihood-based CIs are (24.369, 40.210) and (24.239, 40.174), respectively. For these moderate size samples, profile likelihood-based intervals compare quite well with the exact intervals. The Bartlett correction does not always improve accuracy in a given case; however, in the examples the length of the interval is increased slightly by the adjustment. From a computational viewpoint, both exact and approximate methods require iterative schemes. In the exact method, two quantiles for the non-central t distribution have to be located numerically; for our examples, this task seems to be more time-consuming than locating the end-points of the LRS. In the method which substitutes for the parameter μ into the log likelihood, the numerical burden is relatively lower since each iteration requires only about 1-2 steps to locate the constrained MLEs corresponding to a fixed value of ω. In contrast, the exact method can require up to 100 iterations to locate both quantile points. Overall, the approximate method is a feasible alternative for computing interval estimates in moderate-size samples.

2.5 Coverage Probabilities

In this section, we investigate the coverage probabilities of the Bartlett-adjusted LRS, based on some of the preceding examples.

2.5.1 Odds Ratio in Binomial Sampling

Exact coverage probabilities were calculated for different combinations of (p_1, p_2) and n_1, n_2. Two sample configurations, $(0, 0)$ and (n_1, n_2), provide no information on ψ. Hence, the coverage probabilities should be conditioned on the non-occurrence of these two events. The coverage probabilities in Table 2.6 were calculated on this basis. The first row of each combination of probabilities gives the exact coverage probability of a nominal 95% confidence interval generated by the Bartlett-adjusted LRS. The second row is the corresponding quantity for the unadjusted LRS.

From the tables, we find that the Bartlett-adjusted LRS possesses coverage probabilities at least equal to those of the unadjusted LRS, over the parameter combinations considered. The Bartlett factor appears to yield best effects, roughly, for those (p_1, p_2) pairs on the "perimeter" of the tables. This is especially evident in Table 2.6c. However, the correction can at times result in a conservative interval estimate. As noted by Williams (1986), the uncorrected intervals, in contrast, tend to be anti-conservative. The results also appear to indicate "convergence" to the nominal coverage level as sample size increases.

Table 2.6 Exact conditional coverage probabilities for nominal 95% CIs based on the Bartlett-adjusted (upper row) or unadjusted (lower row) LRS of the odds ratio, $\psi = \{p_1/(1-p_1)\}/\{p_2/(1-p_2)\}$, in independent binomial sampling with sample sizes n_1 and n_2.

a. $n_1 = n_2 = 20$

			p_2		
p_1	0.10	0.25	0.50	0.75	0.90
0.10	0.976	0.931	0.929	0.949	0.975
	0.929	0.924	0.927	0.925	0.906
0.25	0.931	0.949	0.950	0.951	0.949
	0.924	0.948	0.941	0.930	0.925
0.50	0.929	0.950	0.957	0.950	0.929
	0.927	0.941	0.957	0.941	0.927
0.75	0.949	0.951	0.950	0.949	0.931
	0.925	0.930	0.941	0.948	0.924
0.90	0.975	0.949	0.929	0.931	0.976
	0.906	0.925	0.927	0.924	0.929

b. $n_1 = n_2 = 30$

			p_2		
p_1	0.10	0.25	0.50	0.75	0.90
0.10	0.947	0.943	0.941	0.934	0.936
	0.926	0.932	0.936	0.931	0.913
0.25	0.943	0.951	0.950	0.959	0.934
	0.932	0.940	0.946	0.943	0.931
0.50	0.941	0.950	0.948	0.950	0.941
	0.936	0.946	0.948	0.946	0.936
0.75	0.934	0.959	0.950	0.951	0.943
	0.931	0.943	0.946	0.940	0.932
0.90	0.936	0.934	0.941	0.943	0.947
	0.913	0.931	0.936	0.932	0.926

c. $2n_1 = n_2 = 40$

			p_2		
p_1	0.10	0.25	0.50	0.75	0.90
0.10	0.953	0.942	0.945	0.947	0.950
	0.928	0.927	0.915	0.924	0.922
0.25	0.948	0.948	0.951	0.949	0.947
	0.941	0.942	0.946	0.941	0.931
0.50	0.944	0.953	0.945	0.953	0.944
	0.938	0.942	0.945	0.942	0.938
0.75	0.947	0.949	0.951	0.948	0.948
	0.931	0.941	0.946	0.942	0.941
0.90	0.950	0.947	0.945	0.942	0.953
	0.922	0.924	0.915	0.927	0.928

d. $n_1 = n_2 = 40$

			p_2		
p_1	0.10	0.25	0.50	0.75	0.90
0.10	0.941	0.950	0.948	0.951	0.940
	0.935	0.940	0.947	0.937	0.928
0.25	0.950	0.950	0.950	0.956	0.951
	0.940	0.946	0.949	0.956	0.937
0.50	0.948	0.950	0.943	0.950	0.948
	0.947	0.949	0.943	0.949	0.947
0.75	0.951	0.956	0.950	0.950	0.950
	0.937	0.956	0.949	0.946	0.940
0.90	0.940	0.951	0.948	0.950	0.941
	0.928	0.937	0.947	0.940	0.935

2.5.2 Number Needed to Treat

Table 2.7 gives the exact coverage probabilities for the Bartlett-adjusted and unadjusted LRS for the NNT parameter considered in Example 1.1. The nominal level of confidence is 95%.

The results here also show that the coverage of the Bartlett-corrected LRS uniformly equals or exceeds that of the unadjusted LRS for the (p_1, p_2) pairs considered. However, the improvement in coverage is not as marked as in the odds ratio setting.

Table 2.7 Exact coverage probabilities for nominal 95% CIs based on the Bartlett-adjusted (upper row) or unadjusted (lower row) LRS of the NNT, $(p_1 - p_2)^{-1}$, in independent binomial sampling with sample sizes n_1 and n_2.

a. $n_1 = n_2 = 20$

			p_2		
p_1	0.10	0.25	0.50	0.75	0.90
0.10	0.961	0.942	0.952	0.942	0.970
	0.915	0.934	0.952	0.936	0.969
0.25	0.942	0.949	0.948	0.948	0.942
	0.934	0.948	0.943	0.930	0.936
0.50	0.952	0.948	0.957	0.948	0.952
	0.952	0.943	0.957	0.943	0.952
0.75	0.942	0.948	0.948	0.949	0.942
	0.936	0.930	0.943	0.948	0.934
0.90	0.970	0.942	0.952	0.942	0.961
	0.969	0.936	0.952	0.934	0.915

b. $n_1 = n_2 = 30$

	p_2				
p_1	0.10	0.25	0.50	0.75	0.90
0.10	0.946	0.947	0.947	0.943	0.951
	0.925	0.937	0.944	0.943	0.922
0.25	0.950	0.951	0.948	0.948	0.943
	0.940	0.940	0.946	0.948	0.943
0.50	0.947	0.948	0.948	0.948	0.947
	0.944	0.946	0.948	0.946	0.944
0.75	0.943	0.948	0.948	0.951	0.947
	0.943	0.948	0.946	0.940	0.937
0.90	0.951	0.943	0.947	0.947	0.946
	0.922	0.943	0.944	0.937	0.925

c. $2n_1 = n_2 = 40$

	p_2				
p_1	0.10	0.25	0.50	0.75	0.90
0.10	0.951	0.939	0.946	0.936	0.927
	0.926	0.933	0.945	0.936	0.927
0.25	0.951	0.948	0.949	0.946	0.948
	0.950	0.942	0.947	0.939	0.937
0.50	0.953	0.951	0.945	0.951	0.953
	0.948	0.947	0.945	0.947	0.948
0.75	0.948	0.946	0.949	0.948	0.951
	0.937	0.939	0.947	0.942	0.950
0.90	0.927	0.936	0.946	0.939	0.951
	0.927	0.936	0.945	0.933	0.926

2.5.3 ED100p in Dose-Response Study

Since the sampling distribution for a table of dose-response data is the product of independent binomials, exact coverage probabilities can also be calculated for this example. For our purposes, we looked at 13 different dose-response models shown in Table 2.8.

The model involved a three-point assay, with dose levels d_1, d_2, d_3, and equal sample sizes $n_i = 20$, $i = 1, 2, 3$. Let $\mathbf{p} = \mathbf{p}(d_1, d_2, d_3)$ denote the vector of response probabilities, given dose levels (d_1, d_2, d_3). The nominal level of confidence is 95%, and the coverage probabilities were calculated for $\omega_{0.1}$, $\omega_{0.5}$ and $\omega_{0.9}$. Coverage probabilities for intervals based on the unadjusted and Bartlett-adjusted LRS were obtained for the dose-response models in Table 2.8.

The results again show the Bartlett-adjusted LRS to be more conservative relative to the unadjusted LRS. Roughly speaking, the correction appears to be most efficacious for those combinations of $\mathbf{p} < 0.5$. The Bartlett correction results in coverage probabilities that appear to tend to 1 as $\mathbf{p}$ nears 1. In contrast, over the same range the uncorrected LRS yields coverage probabilities slightly below the nominal level, in most cases.

Table 2.8 Exact coverage probabilities for nominal 95% CIs based on the Bartlett-adjusted (upper row) or unadjusted (lower row) LRS of the ED100p, ω_p, in independent binomial sampling with sample sizes $n_i = 20$, $i = 1, 2, 3$.

p	ED10	ED50	ED90
0.057, 0.069, 0.083	0.955	0.947	0.942
	0.918	0.942	0.938
0.114, 0.190, 0.299	0.954	0.950	0.957
	0.943	0.945	0.950

Table 2.8 (Cont.)

p	ED10	ED50	ED90
0.310, 0.400, 0.500	0.949	0.946	0.943
	0.933	0.935	0.937
0.359, 0.404, 0.451	0.948	0.948	0.952
	0.939	0.935	0.932
0.268, 0.404, 0.557	0.946	0.948	0.947
	0.936	0.935	0.937
0.350, 0.380, 0.400	0.956	0.955	0.957
	0.942	0.943	0.935
0.404, 0.450, 0.500	0.958	0.959	0.958
	0.949	0.944	0.948
0.301, 0.352, 0.407	0.956	0.956	0.959
	0.942	0.945	0.937
0.500, 0.590, 0.680	0.957	0.957	0.963
	0.948	0.946	0.945
0.670, 0.720, 0.760	0.970	0.963	0.969
	0.944	0.951	0.944
0.856, 0.858, 0.860	0.980	0.980	0.979
	0.946	0.940	0.941
0.845, 0.855, 0.865	0.981	0.978	0.976
	0.942	0.941	0.944
0.920, 0.936, 0.948	0.994	0.990	0.987
	0.940	0.939	0.915

Chapter 3

Confidence Intervals via Profile Likelihood-based Methods with the EM Algorithm

3.1 Introduction

The EM algorithm has commonly been employed to obtain estimates of scalar or vector parameters in incomplete data settings. Corresponding measures of precision for these estimates are usually based on the observed information and obtained by the method of Louis (1982). These measures of precision are based on the assumption of asymptotic normality of the estimates derived by the EM algorithm. It would be useful to consider obtaining more appropriate estimates of precision in situations where the normality assumption is not warranted.

For this reason, a profile likelihood-based approach for obtaining interval estimates is considered in this chapter. While this approach is familiar in parametric statistical inference, its application in 'incomplete' data settings is not widespread nor well documented, particularly where the object of interest is a functional (as opposed to a parameter vector or subset of it). The idea of using a profile likelihood approach in the missing data context originated from Turnbull and Mitchell (1984) who obtained approximate interval estimates for the quantiles of a survival distribution. Although the basic idea of Turnbull and Mitchell (1984) is in principle straightforward, its implementation via the EM algorithm may not be a trivial task in general. For example, for the statistical analysis of carcinogenicity trials, Dewanji and Kalbfleisch (1986) indicated that the profile likelihood approach may be used to obtain interval estimates for complicated functions of the model parameters. However, the potentially intensive computations involved led them to develop an alternative approach based on the normal approximation. Usually, for constrained models, there is no guarantee that the M-step continues to yield closed-form expressions for updating the estimates. A major advantage of the EM algorithm is thereby lost. To reduce the computational burden, we adapt the approach of Rai and Matthews (1993) to the parameter function case considered in this chapter, and illustrate the procedure through some examples.

Kim and Taylor (1995) have recently described an EM algorithm for estimating model parameters subject to linear restrictions. By utilizing the linearity of the functional and a Lagrange multiplier argument, they showed that profile likelihood-based CIs for a linear functional can be obtained directly from the algorithm. We extend and apply the Lagrange multiplier technique to various missing data scenarios where the functional of interest may be nonlinear in the parameters. The objective is to obtain profile likelihood-based CIs for more general functionals. When closed-form solutions are not readily available in the M-step for the Lagrange multiplier approach, we adapt the approach of Rai and Matthews (1993) as a possible solution.

General notation and the basic EM algorithm are first presented in section 3.2. We extend the algorithm to deal with the interval estimation of functionals, first via a direct profile likelihood approach and then via Lagrange multipliers. As iterative solutions are usually required, we also adapt the EM1 algorithm of Rai and Matthews (1993) to the parameter function setting and verify its self-consistency. In section 3.3 we illustrate the methods by example. A final problem we explore is the possible use of the novel method described in Gu (1996) to derive interval estimates for functionals, as well as its relation to the usual approach (based on profile likelihood).

3.2 The EM Algorithm and the Lagrange Multiplier Method

The Basic EM Algorithm

In this chapter, we will assume a parametric statistical model indexed by parameter $\theta \in \Theta \subset \Re^p$. We consider only cases where the missing data mechanism is uninformative in the sense of Little and Rubin (1987). Let T and T_{obs} denote the complete and observed data, respectively. For example, in a failure time setting. T_{obs} can be the tuple (x, v), where $x = \min\{T, c\}$, v is a function that indicates whether a subject is censored and c is the observed censoring time. The complete data, T, in this case is the failure time that would have been observed for each subject in the absence of censoring. We represent the complete data log likelihood by

$$\ell_0(\theta) = \ell_0(\theta; T)$$

and the observed data log likelihood by

$$\ell(\theta) = \ell(\theta; T_{obs}) .$$

Since $\ell(\theta; T_{obs})$ can be a complicated function with no obvious maximum and a complicated form for the information matrix, direct maximization may be tedious or even intractable. The EM algorithm attempts to circumvent or reduce the difficulty of this problem by utilizing an indirect maximization argument. Let $\ell_1(\theta' : T|T_{obs})$ denote the log likelihood arising from the conditional density of T given T_{obs}. The EM algorithm utilizes the simple identity,

$$\ell(\theta') = Q(\theta', \theta) - R(\theta', \theta) \tag{3.1}$$

where $Q(\theta', \theta) = E[\ell_0(\theta'; T)|T_{obs}; \theta]$ and $R(\theta', \theta) = E[\ell_1(\theta'; T)|T_{obs}, \theta]$. It is important to note that θ' is an argument of the complete data log likelihood, while θ is the parameter of the conditional distribution of T given T_{obs} which is used to compute the conditional expectation. Since $R(\theta', \theta) \leq R(\theta, \theta)$ for any θ', θ, by Jensen's inequality, and

$$\ell(\theta') - \ell(\theta) = [Q(\theta', \theta) - Q(\theta, \theta)] - [R(\theta', \theta) - R(\theta, \theta)], \tag{3.2}$$

it follows that $\ell(\theta') \geq \ell(\theta)$, provided θ' is chosen to maximize $Q(\theta', \theta)$. Generalized EM (GEM) algorithms were also defined in Dempster *et al.* (1977), where θ' is chosen at each M-step to give a non-decreasing Q.

The steps of the EM algorithm can be summarized as follows. In the E-step of the algorithm, given a current estimate $\theta^{(m)}$ of θ, we compute $Q(\theta', \theta^{(m)})$ as a function of the argument θ'. In the M-step we maximize $Q(\theta', \theta^{(m)})$ with respect to θ' to obtain the updated estimate, $\theta^{(m+1)}$. An attractive feature of using the EM algorithm is that by appropriately "redefining" the problem, relatively simpler complete data solutions may be utilized. In particular, closed-form solutions for the MLEs might be available in the M-step. Under fairly general conditions, the sequence $\{\theta^{(m)}\}$ will converge to the MLE $\hat{\theta}$ which maximizes the log likelihood $\ell(\theta; T_{obs})$. Theoretical properties of the EM algorithm are discussed in Wu (1983).

Interval estimates for θ will often be of interest as well. However, additional work will be required since the EM algorithm does not automatically generate measures of precision for the estimates. When sample sizes are "sufficiently large," CIs for θ can be based on the method of Louis (1982). His formula for computing the observed information matrix works by taking expectations, conditional on the incomplete data, of complete data quantities. One shortcoming of the method is that the required algebra can be difficult for some problems, especially when the complete data information matrix has to be derived. It is also known that, since the observed data do not generally constitute an i.i.d. sample, the large sample normality of the likelihood function is not assured. Asymptotic standard errors based on the information matrix are therefore more questionable in the incomplete data case. In the subsequent development, we assume sufficient regularity exists so that likelihood-based inferences remain valid.

The EM1 Algorithm

There may be instances where the M-step of the EM algorithm does not admit a closed-form solution, so that an important advantage of the algorithm is lost. A common strategy in this case is to utilize a numerical tool such as the Newton-Raphson algorithm to implement the M-step. This will increase the numerical burden, but more importantly, it is not difficult to encounter problems with convergence.

A possible solution to these problems was proposed by Rai and Matthews (1993), who employed a slight modification of the EM algorithm at the M-step, giving the so-called EM1 algorithm. Briefly, their recommendation is to replace the usual M-step with a one-step maximization, as follows. Let

$$I(\theta^{(m)}) = -\frac{\partial^2 Q}{\partial\theta\partial\theta^T}|_{\theta=\theta^{(m)}}$$

so that, for example,

$$I(\theta^*) = -\frac{\partial^2 Q}{\partial\theta\partial\theta^T}|_{\theta=\theta^*},$$

where $\theta^* = h\theta^{(m)} + (1-h)\theta^{(m+1)}$, $0 < h < 1$. We assume that $I(\theta^{(m)})$ and $I(\theta^*)$ are positive definite (we may not need the positive definiteness of $I(\theta^*)$). Given the current value of θ, say $\theta^{(m)}$, the EM1 algorithm chooses an $s^{(m)}$, $0 < s^{(m)} < 1$, to update the value of θ according to

$$\theta^{(m+1)} = \theta^{(m)} + s^{(m)} I^{-1}(\theta^{(m)}) S(\theta^{(m)}) \qquad (3.3)$$

where $S(\theta^{(m)}) = \frac{\partial Q(\theta,\theta')}{\partial\theta}|_{\theta=\theta^{(m)}}$ and θ' is a current estimate of θ. This is equivalent to performing one cycle of the Newton-Raphson algorithm, with step length $s^{(m)}$ at the mth cycle. Rai and Matthews (1993) verified the self-consistency of this version of the EM algorithm in the exponential family setting and also demonstrated its efficiency. For completeness, we give a more general proof, as follows. By a Taylor series expansion of $Q(\theta^{(m+1)}, \theta^{(m)})$ about $\theta^{(m)}$ (partial derivatives are with respect to the argument $\theta^{(m+1)}$), we obtain

$$Q(\theta^{(m+1)}, \theta^{(m)}) - Q(\theta^{(m)}, \theta^{(m)}) = (\theta^{(m+1)} - \theta^{(m)})^T S(\theta^{(m)}) - \frac{1}{2}(\theta^{(m+1)} - \theta^{(m)})^T \times$$

$$\begin{aligned} & I(\theta^{*})(\theta^{(m+1)} - \theta^{(m)}) \\ = \; & s^{(m)} S(\theta^{(m)})^{T} I^{-1}(\theta^{(m)})[I_p - \frac{1}{2} s^{(m)} \times \\ & I(\theta^{*}) I^{-1}(\theta^{(m)})] S(\theta^{(m)}) \end{aligned}$$

by equation (3.3). Hence, given $0 < s^{(m)} < 1$, we only need to show that $I^{-1}(\theta^{(m)}) \, [\, I_p - \frac{1}{2} s^{(m)} I(\theta^{*}) I^{-1}(\theta^{(m)}) \,] \geq 0$. We observe that, for a non-terminal iterate, $S(\theta^{(m)}) \neq 0$. Further, $I^{-1}(\theta^{(m)}) > 0$. So for sufficiently small $s^{(m)}$, $I^{-1}(\theta^{(m)}) \, [\, I_p - \frac{1}{2} s^{(m)} I(\theta^{*}) I^{-1}(\theta^{(m)}) \,]$ can be made nonnegative definite. A similar proof can be found in McLachlan and Krishnan (1997). The EM1 algorithm will prove useful for the interval estimation of functionals in subsequent sections.

Interval Estimates for Functionals

Now suppose we are interested in obtaining an interval estimate for a functional $f(\theta) \in \Re$: we assume that $\frac{\partial f}{\partial \theta}$ exists for $\theta \in \Theta$. The conventional approach uses Louis' (1982) method to derive the observed information matrix first. The delta method is then applied to obtain the variance of the estimate for the functional. However. we shall be concerned with a profile likelihood-based approach to the problem.

Consider the hypothesis $H_0 : f(\theta) = \omega$. Suppose that H_0 can be explicitly expressed as $H_0 : \theta = g(\beta)$ for some vector-valued function g and parameter vector $\beta \in \Re^{p+q-1}$ (cf. chapter 2). Then under H_0 the complete data log likelihood, $\ell_0(\beta{:}\,T)$, can be maximized in the usual way via the EM algorithm. In general, since the estimating equations for β do not always yield closed-form solutions for the parameter estimate, a numerical method such as the Newton-Raphson algorithm is required. Let $G(\beta) = \frac{\partial\theta}{\partial\beta}$. The Newton-Raphson algorithm for β is given by

$$[G(\beta)]^T \left[\frac{\partial^2 Q}{\partial\theta\partial\theta^T}\right] G(\beta)\ (\beta^{(m+1)} - \beta^{(m)}) = [G(\beta)]^T \frac{\partial Q}{\partial\theta} \qquad (3.4)$$

where $\frac{\partial Q}{\partial\theta}$ and $\frac{\partial^2 Q}{\partial\theta\partial\theta^T}$ are evaluated at $\theta^{(m)}$. It is well-known that the numerical scheme given by (3.4), commonly employed in the M-step of the EM algorithm, does not always perform satisfactorily.

From the previous subsection, we see that insertion of an appropriate step size into equation (3.4) leads to a corresponding EM1 algorithm for the parameter function case, via (3.3). The self-consistency of this algorithm also follows from the proof above, by observing the following. Since $G(\beta)$ is assumed to have rank $p+q-1$, we can apply a standard result (Rao and Toutenburg (1995)) to deduce that $\{G(\beta)^T I(\theta^{(m)})G(\beta)\}$, and hence $\{G(\beta)^T I(\theta^{(m)})G(\beta)\}^{-1}$, are positive definite.

Before proceeding further with the profile likelihood method in the missing data context, it is appropriate at this point to consider, at least briefly, any theoretical justification for using the profile likelihood to obtain interval estimates for functionals. It is well-known in parametric statistical inference that, under mild regularity conditions, the maximized relative log likelihood has an approximate χ^2_p distribution, where p is the number of restrictions on the parameter vector under the null hypothesis. An analogous result was obtained by Owen (1988) for the nonparametric setting. Subsequent papers (e.g., Li, 1995; Murphy, 1995) rigorously established the asymptotic distribution of the nonparametric LRS for interval estimation of some functionals in censored survival data models. At present, there are no general arguments that similiarly apply to the use of the profile likelihood in missing data contexts. Since the EM algorithm is essentially a numerical tool for maximizing the observed data log likelihood, we have to individually establish the asymptotic distribution of the LRS for each missing data problem. The other alternative is to use the observed information matrix, which is also questionable in view of the data not necessarily being i.i.d. (see Little and Rubin, 1987, pp. 88). Specifically, the asymptotic normality of the MLE does not hold in general for correlated data.

A related point concerns the important assumption of data which are 'missing at random' (MAR), which we make throughout this chapter. Its importance is underscored by the fact that when data are not MAR, maximum likelihood inference based on the observed data log likelihood can lead to inconsistent MLEs (Little and Rubin, 1987). This implies that interval estimation of functionals under these conditions will also be inconsistent (since for each ω, H_0 defines a distinct estimation problem as in the unconstrained estimation of the model parameters).

A Lagrange Multiplier Formulation

We may also consider a Lagrange multiplier approach to the problem of obtaining profile likelihood-based interval estimates for functionals. In the "missing data" context, this technique has not been fully utilized nor studied, except for Kim and Taylor (1995) who used the EM algorithm to test linear hypotheses $H_0 : A\theta = a$, where A is a known $q \times p$ matrix with $\text{rank}(A) = q < p$, and a is a known $q \times 1$ vector. They developed a restricted version of the EM algorithm based on Newton-Raphson iteration and a simple Lagrange multiplier argument. To provide the motivation for our adaptation of the Lagrange multiplier technique to general functionals, we briefly outline the key steps leading to the algorithm of Kim and Taylor (1995).

Let $\ell(\theta|y)$ represent the log likelihood for θ based on a completely observed random sample y. Let S_U and I_U represent the score function and observed information matrix, respectively. The restricted log likelihood function is $\ell_\xi(\theta|y) = \ell(\theta|y) - \xi^T(a - A\theta)$, where ξ is a vector of Lagrange multipliers. The score function and observed information matrix for the restricted log likelihood are clearly $S_R = S_U + A^T\xi$ and $I_R = I_U$. Now let $\theta_R^{(m)}$ denote the estimate of θ at the mth cycle of the Newton-Raphson scheme, i.e.,

$$\begin{aligned} \theta_R^{(m+1)} &= \theta_R^{(m)} + I_R^{-1} S_R \\ &= \theta_R^{(m)} + I_U^{-1} S_U + I_U^{-1} A^T \xi \ ; \end{aligned}$$

this follows from the relations between the restricted and unrestricted score and information matrices.

An important step in their derivation is to recognize that $\theta_R^{(m)} + I_U^{-1} S_U = \theta_U^{(m+1)}[\theta_R^{(m)}]$, where $\theta_U^{(m+1)}[\theta_R^{(m)}]$ denotes the estimate of θ in the $(m + 1)$th cycle of the unrestricted Newton-Raphson scheme, based on the mth cycle estimate, $\theta_R^{(m)}$. The preceding step of their method therefore depends crucially on the linearity of the restrictions. The resulting expression for $\theta_R^{(m+1)}$ is substituted into the constraint,

$$a - A(\theta_U^{(m+1)}[\theta_R^{(m)}] + I_U^{-1} A^T \xi) = 0 \, .$$

This is solved for ξ, which is finally substituted into the expression for θ_R to yield

$$\theta_R^{(m+1)} = \theta_U^{(m+1)}[\theta_R^{(m)}] + I_U^{-1} A^T (A I_U^{-1} A^T)^{-1} (a - A\theta_U^{(m+1)}[\theta_R^{(m)}]) \, .$$

This algorithm generates the "alternating" sequences $\{\theta_U^{(m)}\}$, $\{\theta_R^{(m)}\}$, $m = 0, 1, 2, \ldots$ of estimates, and terminates when $\theta_R^{(m)}$ has converged. For application to missing data problems, the score function and information matrix are replaced by $\frac{\partial Q}{\partial \theta}$ and $\frac{\partial^2 Q}{\partial \theta \partial \theta^T}$, respectively.

To conclude this brief summary of Kim and Taylor's method, note that an important assumption of their derivation is the requirement that the sequence of iterates $\{\theta_R^{(m)}\}$ satisfy the constraint for all m. Results in numerical analysis (Gill *et al.* 1992, pp. 157-61) show that the sequence of iterates are feasible provided $\theta_R^{(0)}$ is a feasible point. For problems with a small number of constraints, suitable feasible initial points can usually be found without difficulty. Numerical methods may be necessary as the number of constraints increases. To summarize, Kim and Taylor's version of the EM algorithm can be useful in the kinds of situations considered in their paper. In particular, when closed-form solutions exist for the unrestricted estimates, their algorithm is most useful since it provides a simple way of computing the constrained estimates; no analytical expressions for these quantities are required.

In the following, we expand the "scope" of the Lagrange multiplier technique to handle nonlinear functionals. When $H_0 : h(\theta) = 0$ can be expressed as $H_0 : \theta = g(\beta)$ implicitly, we may proceed as follows. The obvious modification of the EM algorithm at the E-step consists of deriving the conditional expectation of a constrained log likelihood function, in place of the usual complete data log likelihood $\ell_0(\theta; T)$. Therefore, at the E-step we calculate

$$\begin{aligned} Q_\xi(\theta', \theta) &= E\{\ell_0(\theta'; T) + \xi h(\theta') | T_{obs}; \theta\} \\ &= Q(\theta', \theta) + \xi h(\theta'), \end{aligned}$$

where $h(\theta') = \omega - f(\theta')$, for some arbitrary fixed value of ω and Lagrange multiplier ξ. In principle, the function Q_ξ is easy to maximize at the M step of the algorithm. If a closed-form exists for the constrained estimator, the method proceeds as in unconstrained problems. In this case, an important advantage of the profile likelihood approach is that the complete data information matrix does not need to be derived. Where no closed-form solution is readily available, an EM1 algorithm based on earlier arguments can also be implemented, given a fixed value of the Lagrange multiplier. That is, we can apply the algorithm

$$\left[\frac{\partial^2 Q}{\partial\theta\partial\theta^T}+\xi\frac{\partial^2 h}{\partial\theta\partial\theta^T}\right](\theta^{(m+1)}-\theta^{(m)})=s^{(m)}\left[\frac{\partial Q}{\partial\theta}+\xi\frac{\partial h}{\partial\theta}\right], \tag{3.5}$$

where $0 < s < 1$. The positive definiteness of the first term on the LHS of equation (3.5) is guaranteed provided $\xi\frac{\partial^2 h}{\partial\theta\partial\theta^T} \geq 0$. For linear functionals, the term $\xi\frac{\partial^2 h}{\partial\theta\partial\theta^T}$ disappears, so the usual assumptions suffice.

The argument of Cox and Oakes (1984, p.171) described at the beginning of this section can be adapted to establish the self-consistency of this constrained version of the EM algorithm. For the constrained problem, we seek to maximize the Lagrangian function, $\ell(\theta')+\xi h(\theta')$. From equation (3.2), it follows immediately that θ' should be chosen to maximize $Q_\xi(\theta',\theta)$.

A Computational Note

Given an arbitrary value of $f(\theta) = \omega$, it might be difficult to identify a suitable initial value for β in general. The problems caused by a poor starting value are well-known in the case of the Newton-Raphson algorithm. However, in applications the EM algorithm tends to be less sensitive to the choice of starting value. In our numerical examples, we find this to be the case when the M-step admits closed-form solutions. Choice of initial values can be more crucial when the EM1 algorithm is employed. An ad hoc procedure to obtain the upper bound of the CI for f is as follows (the same technique applies for the lower bound). First identify $f(\hat{\theta}) = \hat{\omega}$, and then let $\hat{\theta}$ be the initial value for the next iteration, for some $\omega > \hat{\omega}$. This approach is adopted in our numerical examples. However, we note that in our examples the above procedure does not always work well for the Lagrange multiplier form of the EM1 algorithm.

3.3 Examples

3.3.1 Profile Likelihood Approach

Example 3.1

By considering a homogeneous random sample of failure times as a special case of grouped multinomial data. Cox and Oakes (1984. pp. 175-77) derived the product-limit estimator of the corresponding survivor function in an alternative way. Their approach. while being more "indirect," does offer some interesting insights on the versatility of the EM algorithm and. furthermore, covers more general censoring schemes than right-censoring.

In their derivation. the basic statistical model takes the form of a multinomial trial with p categories $\{a_1, \ldots, a_p\}$, with associated outcome probabilities $\{\pi_1, \ldots, \pi_p\}$. For a random sample of n subjects. the complete data consist of the multinomial frequencies K_j which record the number of failures at a_j. $j = 1, \ldots, p$. Each subject i observed to fail at a_j can be regarded as yielding the complete information $S_i = \{a_j\}$. If subject i is censored at a_j then $S_i = \{a_{j+1}, a_{j+2}, \ldots, a_p\}$, where a_p is the terminal failure time. The complete data log likelihood function

$$l_0(\pi', K) = \sum_{j=1}^{p-1} K_j \log \pi_j' + K_p \log(1 - \pi_1' - \ldots - \pi_{p-1}')$$

is linear in the complete data sufficient statistics $\{K_j\}$, $j = 1, \ldots, p-1$. Hence

$$Q(\pi', \pi) = \sum_{j=1}^{p-1} E(K_j | S, \pi) \log \pi_j' + E(K_p | S, \pi) \log(1 - \pi_1' - \ldots - \pi_{p-1}'),$$

where $E(K_j|S,\pi) = \sum_{i=1}^{n} \frac{g_{ij}\pi_j}{|S_i|}$, $g_{ij} = 1$ if $a_j \in S_i$, 0 otherwise, and $|S_i| = \sum_{j \in S_i} \pi_j$.

Suppose the functional of interest is the survivor function, $\mathcal{F}(t) = 1 - \sum^{(t)} \pi_j'$. Let $1 - \sum^{(t)} \pi_i' = \omega$. Solving for π_1', we get $\pi_1' = 1 - \omega - \sum_{2 \le j < t} \pi_j'$. Therefore, under the constraint, the complete data log likelihood function ℓ_0 becomes

$$\sum_{j=2}^{p-1} K_j \log \pi_j' + K_p \log(\omega - \sum_{t \le j < p} \pi_j') + K_1 \log(1 - \omega - \sum_{2 \le j < t} \pi_j'),$$

The score equations are

$$\frac{\partial \ell_0}{\partial \pi_j'} = \frac{K_j}{\pi_j'} - \frac{K_1}{1 - \omega - \sum_{2 \le i < t} \pi_i'} = 0, \quad 2 \le j < t \tag{3.6}$$

$$\frac{\partial \ell_0}{\partial \pi_j'} = \frac{K_j}{\pi_j'} - \frac{K_p}{\omega - \sum_{t \le i < p} \pi_i'} = 0, \quad t \le j < p \tag{3.7}$$

Equation (3.6) implies that

$$\pi_j' = (1 - \omega) \frac{K_j}{\sum_{i<t} K_i}.$$

while (3.7) implies

$$\pi_j' = \omega \frac{K_j}{\sum_{t \le i} K_i}.$$

Combining the E and M-steps, we get

$$\tilde{\pi}_j = \begin{cases} (1 - \omega) \frac{E(K_j|S,\pi)}{\sum^{(t)} E(K_i|S,\pi)}, & \text{if } a_j < t \\ \omega \frac{E(K_j|S,\pi)}{n - \sum^{(t)} E(K_i|S,\pi)}, & \text{otherwise} \end{cases}$$

as the constrained EM estimates for $\{\pi_j\}$ corresponding to a fixed value of ω. We compute $2\{\ell(\hat{\pi}) - \ell(\tilde{\pi}(\omega)\}$ and compare it with the relevant χ_1^2 quantile point. These steps are iterated until the end-points of the interval are obtained. [For this problem, Li (1995) and Murphy (1995) have established the asymptotic distribution of the LRS.]

The preceding estimate is analogous to what Turnbull and Mitchell (1984) obtained in the context of carcinogenicity trials. Their interval estimate for the median survival time for subjects with tumours was obtained by renormalizing the unrestricted EM estimates of p_l ($\equiv \pi_j$ in our case) to satisfy the constraint introduced. This simple device of renormalization may be very inconvenient to implement when multiple functionals are of simultaneous interest.

To obtain an approximate interval estimate for the median survival time $t_{0.5}$ using our approach, we proceed along the following lines. First set $1-\sum^{(t)} \pi_i' = 0.5$. where t is an arbitrary, fixed value for the median survival. The constrained MLEs for π_i' corresponding to each fixed choice for t are found using the same method discussed above, yielding a nonparametric profile likelihood for the median survival time. To obtain an approximate 95% CI for $t_{0.5}$, we locate the abscissa for this profile likelihood corresponding to a drop of two log units from the unconstrained maximum. As noted by Turnbull and Mitchell, it is difficult to ascribe an exact level of confidence for intervals derived in this manner, even though they may usefully be regarded as approximate $100(1-\alpha)\%$ CIs.

For a simple numerical illustration, consider the data in Table 3.1 showing the remission times of leukaemia patients (Gehan, 1965). Censored observations are indicated by an asterisk. The treatment group received the drug 6-mercaptopurine (6-MP) while the other subjects served as controls. Treatment allocation was randomized. It has been pointed out that the study design was not a routine randomized controlled trial and that the censoring scheme was in fact sequential in nature (cf. Cox and Oakes (1984); Venables and Ripley (1994)). These aspects of the trial should be taken into account in a detailed analysis of the data. For our present

Table 3.1: Remission time in weeks, Gehan (1965).

6-MP	6*	6*	6*	6	7	9*	10*	10	11*	13	16	17*
	19*	20*	22	23	25*	32*	32*	34*	35*			
control	1	1	2	2	3	4	4	5	5	8	8	8
	8	11	11	12	12	15	17	22	23			

Table 3.2: Interval estimates for the median survival. $\alpha = 0.10$.

Group	Profile Likelihood	Kaplan-Meier	MLE
6-MP	(16. NA)	(16. NA)	23
Control	(3. 11)	(5. 12)	8

purposes. we shall ignore these caveats. and implement our method of obtaining an interval estimate for the median survival time in both treatment and control groups. The results are summarized in Tables 3.2 and 3.3.

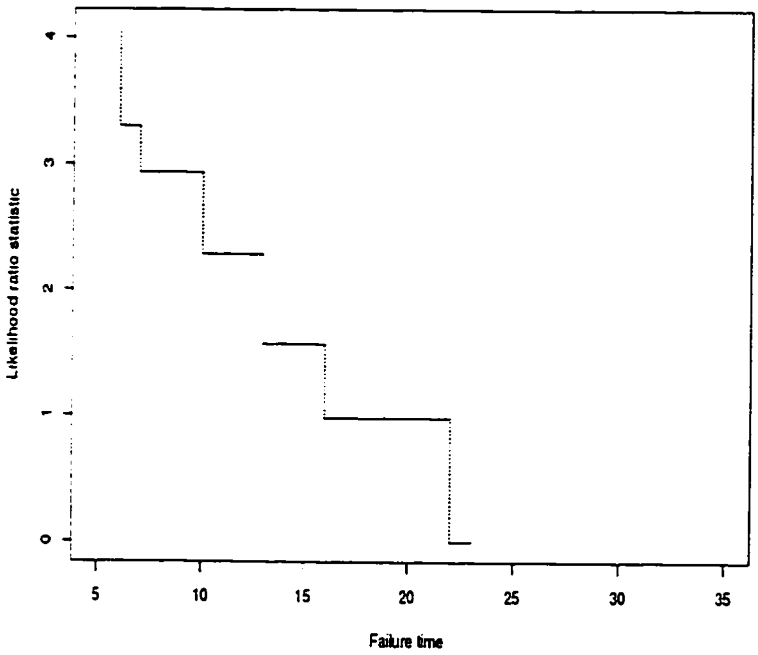
The Kaplan-Meier interval estimates were obtained by fitting the survival curves and finding the time t for which $S(t) \leq 0.5$. i.e.. the lower and upper limits are the intersection of a horizontal line at 0.5 with the lower and upper confidence bands for $S(t)$. If the upper confidence band for $S(t)$ does not reach 0.5. the upper limit is unknown, hence the NA symbol. We note that. although this is a standard

Table 3.3: Interval estimates for the median survival. $\alpha = 0.05$.

Group	Profile Likelihood	Kaplan-Meier	MLE
6-MP	(13, NA)	(16, NA)	23
Control	(3, 12)	(4, 12)	8

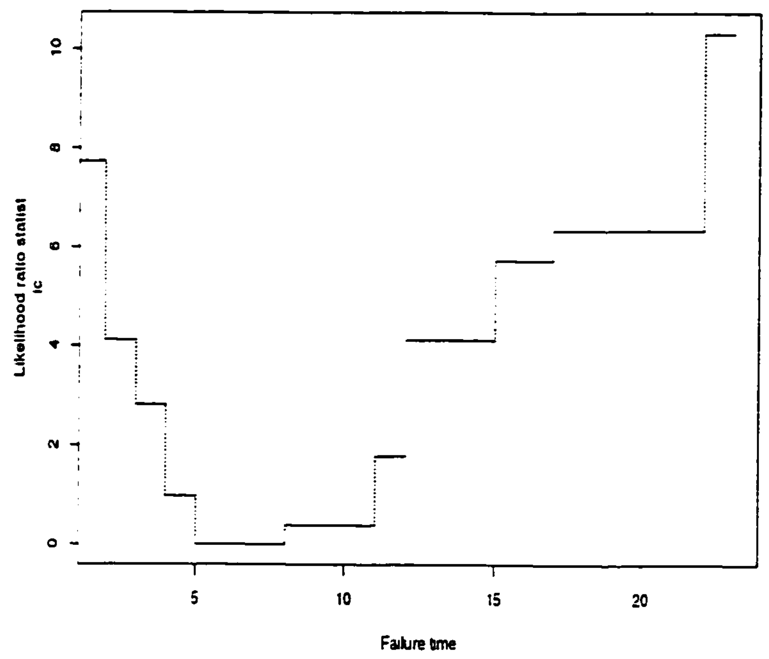
method of obtaining interval estimates for quantiles, the resulting interval estimates do not possess the ascribed confidence level. This follows since the confidence intervals based on the Kaplan-Meier estimate are point-wise confidence intervals for the survivor function, evaluated at fixed time points. Hence, the likelihood-based approach ought to be the preferred solution to this estimation problem.

Figure 3.1: Nonparametric likelihood ratio statistic for the median survival in the 6-MP group of leukaemia patients, based on the data in Table 3.1.



Note that in the case of likelihood-based interval estimates, a value of NA is obtained for the 6-MP group because the MLE occurs at the final death point. For the data in this example, the profile likelihood-based intervals are longer in three out of four cases. For the control group, they also tend to be more asymmetrical. Figures 3.1 and 3.2 display the nonparametric likelihood ratio statistics for the two groups of subjects.

Figure 3.2: Nonparametric likelihood ratio statistic for the median survival time in the control group of leukaemia patients, based on the data in Table 3.1.



Example 3.2

Consider a variance components problem discussed in Little and Rubin (1987, p. 149) and Kim and Taylor (1995). The model suggested is

$$y_{ij} = \alpha_i + e_{ij}, \tag{3.8}$$

where the parameters α_i describe the primary effects of interest and the residuals e_{ij} denote the secondary effects, $i = 1, ..., J$, $j = 1, ..., n_i$. The modelling assumptions are

$$\alpha_i \sim N(\mu, \sigma_\alpha^2) \quad \text{i.i.d.},$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \text{i.i.d.}.$$

If we treat the α_i and y_{ij} as missing and observed data, respectively, we can obtain an EM algorithm for estimating $\theta = (\mu, \sigma_\alpha^2, \sigma_e^2)^T$ (Little and Rubin, 1987, p. 151). The complete data log likelihood is given by

$$-\frac{IJ}{2}\log\sigma_e^2 - \frac{1}{2}\sum_i\sum_j\frac{(y_{ij}-\alpha_i)^2}{\sigma_e^2} - \frac{I}{2}\log\sigma_\alpha^2 - \frac{1}{2}\sum_i\frac{(\alpha_i-\mu)^2}{\sigma_\alpha^2}. \tag{3.9}$$

Maximizing (3.9) with respect to θ yields the complete data MLEs,

$$\begin{aligned}\hat{\mu} &= N^{-1}\sum_i \alpha_i,\\ \hat{\sigma}_\alpha^2 &= N^{-1}\sum_i \alpha_i^2 - \hat{\mu}^2,\\ \hat{\sigma}_e^2 &= (\sum_i n_i)^{-1}\left\{\sum_i\sum_j(y_{ij}-\bar{y}_{i\cdot})^2 + \sum_i n_i(\bar{y}_{i\cdot}-\alpha_i)^2\right\}.\end{aligned}$$

where $\bar{y}_{i\cdot} = {n_i}^{-1}\sum_{j=1}^{n_i} y_{ij}$.

Little and Rubin utilize Bayes theorem to show that the conditional distributions of the primary effects, α_i, given the value of the data $\{y_{ij}\}$ and θ, are independent and normally distributed with mean $w_i\mu + (1-w_i)\bar{y}_{i\cdot}$ and variance ν_i, $i = 1, ..., n_i$, i.e.

$$[\alpha_i|\{y_{ij}\}, \theta] \sim N(w_i\mu + (1-w_i)\bar{y}_{i\cdot\cdot}, \nu_i)$$

independently, where $w_i = \sigma_\alpha^{-2}\nu_i$ and $\nu_i = (\sigma_\alpha^{-2} + n_i\,\sigma_e^{-2})^{-1}$. Applying the E-step of the algorithm, it is straightforward to show that the updated MLE of θ at the conclusion of cycle $m+1$ of the algorithm has components

$$\begin{aligned} \mu^{(m+1)} &= N^{-1}\{ \sum_i \left[w_i^{(m)}\mu^{(m)} + (1-w_i^{(m)})\bar{y}_{i\cdot}\right] \}. \\ (\sigma_\alpha^2)^{(m+1)} &= N^{-1}\{ \sum_i \left[w_i^{(m)}\mu^{(m)} + (1-w_i^{(m)})\bar{y}_{i\cdot}\right]^2 + \sum_i \nu_i^{(m)} \} - (\mu^{(m)})^2. \end{aligned}$$

and

$$(\sigma_e^2)^{(m+1)} = (\sum_i n_i)^{-1}\{ \sum_i \sum_j (y_{ij} - \bar{y}_{i\cdot})^2 + \sum_i n_i[w_i^{(m)2}(\mu^{(m)} - \bar{y}_{i\cdot})^2 + \nu_i^{(m)}] .$$

where $w_i^{(m)}$ denotes the value of w_i based on the mth cycle estimates. Kim and Taylor (1995) used their restricted EM algorithm to test the hypothesis $H_0 : \sigma_\alpha^2 = \Omega\sigma_e^2$. for some fixed value Ω. by rewriting H_0 as a linear hypothesis. Note that in their example, $N = I$ and $n_i = J$ $(i = 1, \dots, I)$. Alternatively. we may use the following procedure. For a fixed value of Ω. substituting $\Omega\sigma_e^2$ for σ_α^2 yields the complete data log likelihood under H_0 as

$$-\frac{1}{2}\frac{\sum_i\sum_j(y_{ij}-\alpha_i)^2}{\sigma_e^2} - \frac{1}{2}\frac{\sum_i(\alpha_i-\mu)^2}{\Omega\sigma_e^2} - \frac{IJ}{2}\log(\sigma_e^2) - \frac{I}{2}\log(\Omega\sigma_e^2) \qquad (3.10)$$

Let $(\tilde{\mu}, \tilde{\sigma}_e^2)^T$ denote the complete data constrained MLEs of $(\mu, \sigma_e^2)^T$. Clearly $\tilde{\mu} = \hat{\mu}$. while

$$\begin{aligned} \tilde{\sigma}_e^2 &= \frac{\sum\sum(y_{ij}-\alpha_i)^2 + \Omega^{-1}\sum_i(\alpha_i - \tilde{\mu})^2}{I(J+1)} \\ &= \frac{\sum\sum(y_{ij}-\bar{y}_{i\cdot})^2 + J\sum_i(\bar{y}_{i\cdot} - \alpha_i)^2 + \Omega^{-1}[\sum_i \alpha_i^2 - 2\tilde{\mu}\sum_i \alpha_i + I\tilde{\mu}^2]}{I(J+1)} \end{aligned}$$

The E-step is straightforward to implement since we already obtained $E(\alpha_i|y, \mu^{(m)}, (\sigma_e^2)^{(m)})$

Table 3.4: Variance components data cited in Kim and Taylor (1995).

Unit	y_{ij}			
1	76	64	85	75
2	58	75	81	66
3	49	63	62	46
4	74	71	85	90
5	66	74	81	79

and $E(\alpha_i^2|y, \mu^{(m)}, (\sigma_e^2)^{(m)})$. We obtain

$$\tilde{\mu}^{(m+1)} = (1 + J\Omega)^{-1}\{\tilde{\mu}^{(m)} + J\Omega\bar{y}_{..}\}.$$

where $\bar{y}_{..} = (IJ)^{-1}\sum_i \sum_j y_{ij}$, and

$$\begin{aligned}(\tilde{\sigma}_e^2)^{(m+1)} &= (J+1)^{-1}(\tilde{\sigma}_e^2)^{(m)} + \frac{1}{I(J+1)}\sum\sum y_{ij}^2 \\ &\quad - \frac{J}{I(J+1)(1+J\Omega)}\{\, 2\tilde{\mu}^{(m)}\sum \bar{y}_{i\cdot} - I(\tilde{\mu}^{(m)})^2 + J\Omega\sum \bar{y}_{i\cdot}^2 \,\}\end{aligned}$$

as the constrained EM estimates for $(\mu, \sigma_e^2)^T$. We illustrate this alternative approach with the following data. Table 3.4 contains the data on Apex Enterprises from Neter *et al.* (1985). Each row represents the evaluation ratings given by a personnel officer to four candidates selected at random. Five personnel officers were randomly selected to give ratings. Kim and Taylor applied their algorithm to these data, obtaining the profile likelihood-based 90% CI (0.131, 2.466) for $\sigma_\alpha^2/\sigma_e^2$. In contrast, we obtained (0.070, 3.599).

The same procedure can be used on a variety of linear model designs. In particular, Laird (1982) provides a general framework on which the preceding method might be applied.

Example 3.2 (Cont.)

The variance components example considered previously admitted a convenient closed-form solution for the constrained MLEs at the M-step. In general. even for linear functions of parameters, such simple solutions are not available. The EM1 algorithm may be useful for these cases, and can be an attractive alternative to Kim and Taylor's (1995) algorithm especially when there is no closed-form solution for the unconstrained estimates and the number of linear constraints is small.

For illustration, we consider the simple case of one linear function of interest. with $H_0 : \theta^T \mathbf{x}_0 = \omega \in \Re$. By expressing θ_p as a function of the other parameters. we obtain the matrix $G = \{\frac{\partial \theta}{\partial \beta}\}$. with the identity matrix I_{p-1} as its first $p-1$ rows. and

$$\left(-\frac{x_{01}}{x_{0p}}, \ldots, -\frac{x_{0,p-1}}{x_{0p}} \right),$$

as its pth row. As usual. x_{0j} denotes the jth element of $\mathbf{x}_0$. Using this approach on the variance components data in Table 3.4. we obtained the same interval estimate. i.e. (0.070. 3.599). For our data, we found that the EM1 algorithm was quite stable with respect to the choice of a starting value (although we might also want to verify that different choices lead to different rates of convergence. in future work). Convergence was also quite rapid for a selection of different starting values. In addition, as in the case of Rai and Matthews (1993), setting $s^{(p)} = 1, \forall p$, was sufficient to ensure convergence.

Example 3.3

In this example, we consider using the EM1 algorithm to obtain approximate CIs for some important functionals in the context of carcinogenicity studies incorporating numerous interim sacrifices. The data from such studies are incomplete in the sense that the time to occurrence of tumour in live animals is not clinically observable: the presence or absence of tumour is determined by autopsy following the death or sacrifice of the animal. If tumour onset was observable, standard nonparametric survival analysis methods could be used to analyze important quantities such as tumour incidence. Animal survival and sacrifice experiments are commonly used to provide important information for identifying carcinogens and estimating carcinogenic effects.

Hoel and Walburg (1972) first investigated data from carcinogenicity studies involving occult tumours in live animals. They made a useful distinction between rapidly lethal tumours and incidental tumours. In the former, time to death following tumour onset is short, and therefore it is a good proxy for the time to tumour onset. Hence an analysis may be based on time to death with tumour. On the other hand, incidental tumours have no effect on the death rate and the proportion of deaths with tumour provides an estimate of the tumour prevalence at that time. Other papers (e.g. Kodell and Nelson (1980), Dinse and Lagakos (1982) and Turnbull and Mitchell (1984)) utilize cause-of-death information, while McKnight and Crowley (1984) and Dewanji and Kalbfleisch (1986) provide extensive surveys of nonparametric methods of estimation in occult tumour studies. In both the latter papers, numerous interim sacrifices are required for analyses.

For the present example, we shall not be concerned with the relative merits of the preceding approaches to analyzing occult tumour data but will instead illustrate the use of the EM1 algorithm in deriving interval estimates for some functionals.

We refer to the basic framework described by Dewanji and Kalbfleisch (1986), who addressed the problem of providing a nonparametric estimate of the tumour incidence rate. We suppose that n animals are subjected to insults with a carcinogen and followed over time. Let X_1 represent the time until occurrence of tumour ($J = 1$) or until death without tumour ($J = 2$). Let X_2 represent the time (from the start of study) until death with tumour. We also assume a discrete time setting for X_i, $i = 1, 2$.

Let

$$\lambda_j(u) = P(X_1 = u, J = j | X_1 \geq u), \quad j = 1, 2\,,$$

$$\lambda_{11}(t|u) = P(X_2 = t | J = 1, X_2 \geq t, X_1 = u), \quad t = u, u+1, \dots, M\,,$$

where $u = 1, 2, \dots, M$. We further assume an independent scheme for determining the time of sacrifice Y_i for the ith animal. Specifically, we have $P(Y_i = j) = q_j$, $j = 1, \dots, M$ and $\sum q_j = 1$. We also define the survivor functions

$$S_1(u) = P(X_1 \geq u) = \prod_{v<u} [1 - \lambda_1(v) - \lambda_2(v)]$$

$$S_2(t|u) = P(X_2 \geq t | X_1 = u, J_1 = 1) = \prod_{u \leq v < t} [1 - \lambda_{11}(v|u)], \quad t > u\,.$$

where $u = 1, 2, \dots, M$.

The parameter vector for this model is $\lambda = (\lambda_1(1), \dots, \lambda_1(M), \lambda_2(1), \dots, \lambda_2(M), \lambda_{11}(1|1), \lambda_{11}(2|1), \dots, \lambda_{11}(M|M))^T$, since only transition probabilities up to time M can be estimated. Dewanji and Kalbfleisch (1986) used the EM algorithm to obtain closed-form estimates for λ. Since tumour onset is unobservable, they specified a complete data problem in place of the original one. In this complete data formulation, we suppose that the data consist of the processes

$$\begin{aligned}
N_j(u) &= \#\{i : X_{1i} = u, X_{1i} \leq Y_i, J_i = j\}, \quad j = 1, 2, \\
N_{11}(t|u) &= \#\{i : X_{1i} = u, X_{21} = t, J_i = 1, Y_i \geq t\}, \\
Y_0(t) &= \#\{i : Y_i = t < X_{1i}\}, \\
Y_1(t|u) &= \#\{i : X_{1i} = u \leq Y_i = t < X_{2i}, J_i = 1\}
\end{aligned}$$

where X_{1i}, X_{2i} and J_i denote the values of X_1, X_2 and J respectively, for the ith animal. The process $N_1(u)$ represents the number of tumour onsets occurring at time u and $N_2(u)$ the number of deaths without tumour at time u. Likewise, $N_{11}(t|u)$ is the number of deaths at time t of animals with onset time u. The process $Y_0(t)$ records the number of tumour-free animals sacrificed at t, while $Y_1(t|u)$ is the number of animals with tumour onset at time u and sacrificed at time t. We also let $R_1(u)$ denote the number of animals at risk of death without disease or of tumour onset at time $u-$, and $R_{11}(t|u)$ the number at risk of death at time $t-$ with disease onset at $u \leq t$. The complete data likelihood function is given by

$$\begin{aligned}
L_0 = \prod_{u=1}^{M} (\; & \lambda_1(u)^{N_1(u)} \lambda_2(u)^{N_2(u)} [1 - \lambda_1(u) - \lambda_2(u)]^{R_1(u) - N_1(u) - N_2(u)} \times \\
& \prod_{t=u}^{M} \{\lambda_{11}(t|u)^{N_{11}(t|u)} [1 - \lambda_{11}(t|u)]^{R_{11}(t|u) - N_{11}(t|u)}\} \;);
\end{aligned}$$

this is easily maximized to give the complete data MLEs

$$\begin{aligned}
\hat{\lambda}_j(u) &= N_j(u)/R_1(u), \quad j = 1, 2, \\
\hat{\lambda}_{11}(t|u) &= N_{11}(t|u)/R_{11}(t|u), \quad t \geq u, \; u = 1, 2, ..., M,
\end{aligned}$$

for the Markov version of the problem. Let $N_{11}(t)$ and $Y_1(t)$ denote respectively the numbers of deaths and sacrifices with tumour present at time t. In the E-step. Dewanji and Kalbfleisch (1986) obtain

$$\begin{aligned} E\{N_1(u)|N_{11}(\cdot), Y_1(\cdot), \lambda^{(m)}\} &= \sum_{t\geq u}[N_{11}(t)P^{(m)}(u,t) + Y_1(t)R^{(m)}(u,t)], \\ E\{N_{11}(t|u)|N_{11}(\cdot), Y_1(\cdot), \lambda^{(m)}\} &= N_{11}(t)P^{(m)}(u,t), \\ E\{Y_1(t|u)|N_{11}(\cdot), Y_1(\cdot), \lambda^{(m)}\} &= Y_1(t)R^{(m)}(u,t), \end{aligned}$$

where

$$P(u,t) = P(X_1 = u, J = 1|X_2 = t) = \frac{\lambda_1(u)S_1(u)\lambda_{11}(t|u)S_2(t|u)}{\sum_{v\leq t}\lambda_1(v)S_1(v)\lambda_{11}(t|v)S_2(t|v)},$$

$$R(u,t) = P(X_1 = u|X_1 \leq t < X_2, J = 1) = \frac{\lambda_1(u)S_1(u)S_2(t+1|u)}{\sum_{v\leq t}\lambda_1(v)S_1(v)S_2(t+1|v)}.$$

These follow by noting that, conditionally on the set $\{N_{11}(\cdot), Y_1(\cdot), \lambda^{(m)}\}$, $N_1(u)$, $N_{11}(t|u)$ and $Y_1(t|u)$ are binomial variates. Suppose we are interested in interval estimates for functions of λ. For example, the cumulative hazard for tumour onset at time t,

$$\Lambda_1(t) = \sum_{u=1}^{t}\lambda_1(u),$$

the subdistribution function,

$$F_1(t) = P(X_1 \leq t, J = 1) = \sum_{u=1}^{t}\lambda_1(u)S_1(u),$$

or the prevalence of disease among surviving animals at time t,

$$P(t) = \frac{\sum_{u\leq t}\lambda_1(u)S_1(u)S_2(t|u)}{S_1(t) + \sum_{u\leq t}\lambda_1(u)S_1(u)S_2(t|u)}$$

may be of interest.

Let us consider $F_1(t)$. We begin by fixing $F_1(t) = \omega \in [0, 1]$. Since there is no apparent closed-form solution for the constrained estimator in this case, we apply the EM1 algorithm. We first solve for $\lambda_1(1)$ as

$$\lambda_1(1) = \frac{\omega - \{1 - \lambda_2(1)\}\gamma(t)}{1 - \gamma(t)}$$

where $\gamma(t) = \sum_{u=2}^{t} \lambda_1(u) \prod_{1<v<u}\{\lambda_3(v)\}$ and $\lambda_3(u) = 1 - \lambda_1(v) - \lambda_2(v)$. The components of the matrix $G = \frac{\partial \theta}{\partial \beta}$ are not difficult to obtain in this case. Moreover, the analytical effort required for the delta method is comparable to that used here: in the delta method we need to compute $\frac{\partial f}{\partial \theta}$, for a functional f. Most of the work involves finding

$$\frac{\partial \gamma(t)}{\partial \lambda_1(j)} = \begin{cases} 1 - \frac{\gamma(t) - \lambda_1(2)}{\lambda_3(2)}, & j = 2 \\ \prod_{1<v<j}\{\lambda_3(v)\} - \frac{\gamma(t) - \sum_{i=2}^{j}\{\lambda_1(i) \prod_{1<v<i} \lambda_3(v)\}}{\lambda_3(j)}, & 2 < j < t - 1 \\ \prod_{1<v<t} \lambda_3(v), & j = t \\ 0, & j > t \end{cases}$$

and

$$\frac{\partial \gamma(t)}{\partial \lambda_2(j)} = \begin{cases} -\frac{\gamma(t) - \sum_{i=2}^{j}\{\lambda_1(i) \prod_{1<v<i} \lambda_3(v)\}}{\lambda_3(j)}, & 2 \le j \le t - 1 \\ 0, & j \ge t \end{cases}$$

The unconstrained EM estimates for λ admit closed forms (and hence do not require computation of quantities such as $\frac{\partial^2 Q}{\partial \lambda \partial \lambda^T}$). Second derivative matrices are, however, required in the EM1 algorithm for obtaining the constrained estimates. In general, this might involve considerable additional computation, but in many practical instances (including this one), $\frac{\partial^2 Q}{\partial \lambda \partial \lambda^T}$ is readily obtained.

For the prevalence function $P(t)$, we proceed in an indirect route. We shall first obtain an approximate CI for $V(t) = \frac{S_1(t)}{\sum_{u\le t}\lambda_1(u)S_1(u)S_2(t|u)}$. By using the fact that $V(t)$ is a monotonic function in t and the $1-1$ relation $P(t) = \{1+V(t)\}^{-1}$ between $V(t)$ and $P(t)$, we can obtain a corresponding interval estimate for $P(t)$. Let $V(t) = \omega$, for some $\omega \in [0,1]$. Note that

$$\sum_{u\le t}\lambda_1(u)S_1(u)S_2(t|u) = \lambda_1(1)\{\rho(t) - \delta(t)\} + (1-\lambda_2(1))\delta(t),$$

where

$$\rho(t) = \prod_{1\le v<t}(1-\lambda_{11}(v|1)).$$

$$\delta(t) = \sum_{j=2}^{t}\{\lambda_1(j)\prod_{2\le u<j}\lambda_3(u)\prod_{j\le v<t}(1-\lambda_{11}(v|j))\},$$

and we define $\prod_{2\le u<2}\lambda_3(u) = 1$. We can also re-express $S_1(t)$ as

$$(1-\lambda_2(1))\{\prod_{1<v<t}\lambda_3(v)\} - \lambda_1(1)\{\prod_{1<v<t}\lambda_3(v)\}.$$

Substituting these into the constraint, we can solve for $\lambda_1(1)$ as

$$\begin{aligned}\lambda_1(1) &= \frac{(1-\lambda_2(1))\prod_{1<v<t}\lambda_3(v) - \omega(1-\lambda_2(1))\delta(t)}{\prod_{1<v<t}\lambda_3(v) + \omega\{\rho(t)-\delta(t)\}}\\ &= (1-\lambda_2(1))\left\{1 - \frac{\omega\rho(t)}{\prod_{1<v<t}\lambda_3(v) + \omega\{\rho(t)-\delta(t)\}}\right\}.\end{aligned}$$

Now we need to work out the matrix G for this case. This involves computing the quantities

$$\frac{\partial\delta(t)}{\partial\lambda_1(j)} = \begin{cases} \lambda_1(j)\prod_{2\le v<j}\lambda_3(v)\prod_{j\le v<t}[1-\lambda_{11}(v|j)] & \\ -\frac{\delta(t)-\sum_{i=2}^{j}\{\lambda_1(i)\prod_{2<v<i}\lambda_3(v)\prod_{i<v<t}[1-\lambda_{11}(v|i)]\}}{\lambda_3(j)}, & 2\le j\le t\\ 0, & j>t\end{cases}$$

$$\frac{\partial\delta(t)}{\partial\lambda_2(j)} = \begin{cases} 0, & j = 1 \\ -\frac{\delta(t)-\sum_{i=2}^{j}\{\lambda_1(i)\prod_{2<v<i}\lambda_3(v)\prod_{i<v<t}[1-\lambda_{11}(v|i)]\}}{\lambda_3(j)}, & 2 \leq j < t \\ 0, & j \geq t \end{cases}$$

$$\frac{\partial\delta(t)}{\partial\lambda_{11}(j|i)} = \begin{cases} 0, & i = 1, \forall j \\ -\frac{\lambda_1(i)\prod_{2<v<i}\lambda_3(v)\prod_{i<v<t}[1-\lambda_{11}(v|i)]}{1-\lambda_{11}(j|i)}, & 2 \leq i < t,\ i \leq j < t \\ 0, & i, j \geq t \end{cases}$$

Given these partial derivatives and corresponding ones for $\rho(t)$ (which are straightforward), it is relatively easy to obtain the partial derivatives of $\lambda_1(1)$ with respect to the other parameters; see the appendix.

Berlin *et al.* (1979) report the results of a comparative assay with serial sacrifice data. For our purposes, we consider the data summarized in Tables 3.5 and 3.6 below: Dewanji and Kalbfleisch (1986) provide a complete listing of the data categorized according to the disease of interest. Table 3.7 displays some interval estimates for the prevalence function. The level of confidence is 0.95. To obtain the left and right endpoints of each interval, we used the unrestricted MLE as a starting point for the iteration. For the glomerulosclerosis data, this choice proved to be satisfactory. Figures 3.3 and 3.4 display the likelihood ratio statistics for $V(t)$ for both irradiated and control groups at $t = 201 - 300$. In both cases, the normal approximation appears to be adequate.

Table 3.5: Summary of the data for glomerulosclerosis as the disease of interest. Control group

t	0-	101-	201-	301-	401-	501-	601-	≥ 701
	100	200	300	400	500	600	700	
N_2	1	9	8	5	6	4	2	1
N_{11}	1	16	55	71	98	127	177	138
Y_0	58	41	19	9	2	1	1	0
Y_1	14	24	44	31	37	32	32	15

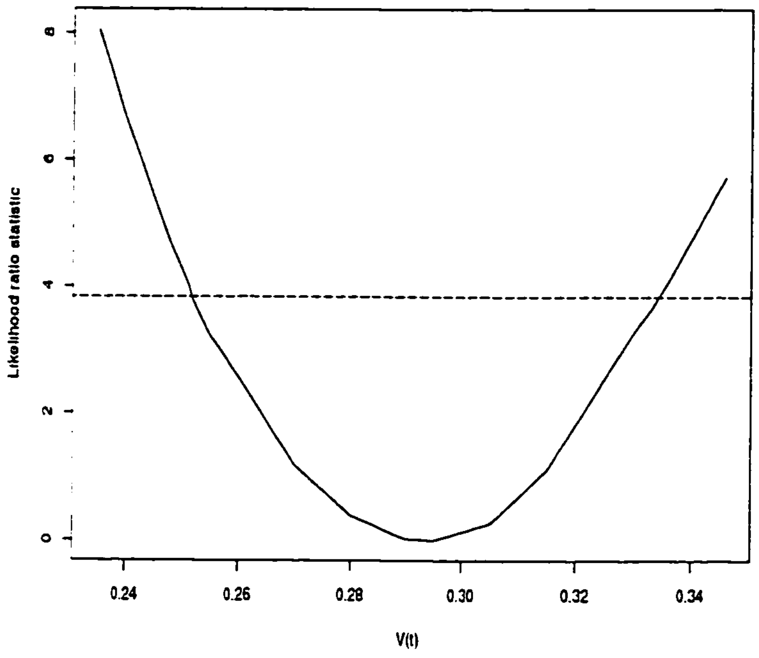
Table 3.6: Irradiated group

t	0-	101-	201-	301-	401-	501-	601-	≥ 701
	100	200	300	400	500	600	700-	
N_2	4	61	65	24	12	5	6	2
N_{11}	4	144	257	204	150	99	65	9
Y_0	56	43	16	2	1	1	1	0
Y_1	11	25	50	41	38	30	26	2

Table 3.7: Interval estimates of prevalence.

	Control group		Irradiated group	
t	$\hat{P}_t$	CI	$\hat{P}_t$	CI
201 – 300	0.7092	(0.6937, 0.7208)	0.7722	(0.7587, 0.7911)
301 – 400	0.7750	(0.7658, 0.8066)	0.9389	(0.9372, 0.9400)

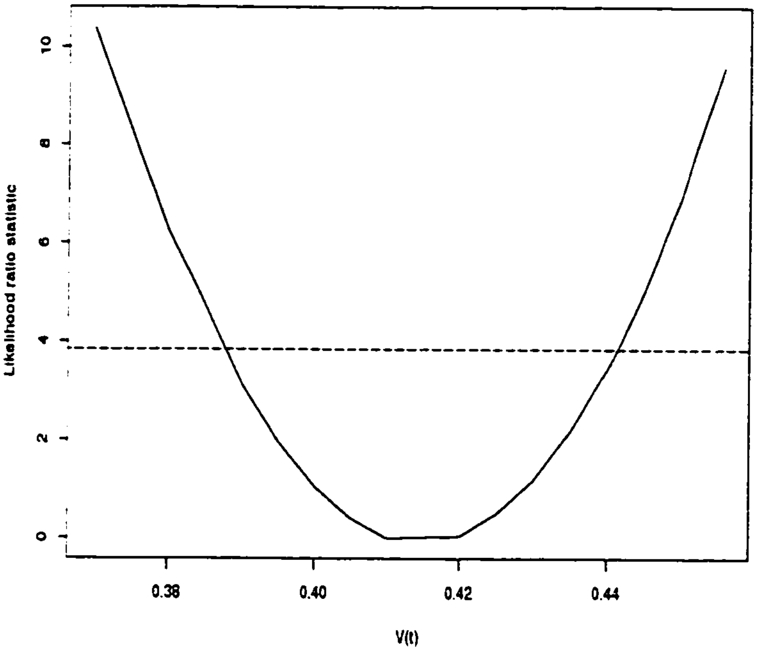
Figure 3.3: Likelihood ratio statistic for $V(t) = S_1(t)/\{\sum_{u\leq t} \lambda_1(u)S_1(u)S_2(t|u)\}$ during the interval $201 \leq t \leq 300$ days, based on the data for irradiated animals summarized in Table 3.5. The height of the horizontal dashed line is 3.841 units.



3.3.2 Lagrange Multiplier Approach

The following examples illustrate the Lagrange multiplier approach. In particular, they show that a Lagrange multiplier approach can be useful in "missing" data contexts by simplifying the analytical work in some cases. For certain models and functionals, numerical computations can potentially be relieved by using a Lagrange multiplier argument.

Figure 3.4: Likelihood ratio statistic for $V(t) = S_1(t)/\{\sum_{u<t}\lambda_1(u)S_1(u)S_2(t|u)\}$ during the interval $201 \leq t \leq 300$ days, based on the data for control animals summarized in Table 3.5. The height of the horizontal dashed line is 3.841 units.



Example 3.4

Recall that the survivor function was the functional of interest in Example 3.1. A Lagrange multiplier approach can also be implemented, as follows. Define

$$Q_\xi(\pi', \pi) = Q(\pi', \pi) + \xi\{\omega - \mathcal{F}(t)\}.$$

At the M-step, the score equations for $\{\pi_j'\}$ are

$$\frac{\partial Q_\xi}{\partial \pi_j'} = \frac{E(K_j|S,\pi)}{\pi_j'} - \frac{E(K_p|S,\pi)}{\pi_p'} + \xi = 0$$

for $\{j : a_j < t\}$, and

$$\frac{\partial Q_\xi}{\partial \pi_j'} = \frac{E(K_j|S,\pi)}{\pi_j'} - \frac{E(K_p|S,\pi)}{\pi_p'} = 0$$

for $\{j : a_j \geq t\}$. Let $\Lambda(\xi, S, \pi) = \frac{(n+\xi)\pm\sqrt{(n+\xi)^2 - 4\xi\sum_{j\geq t} E(K_j|S,\pi)}}{2\xi\sum_{j\geq t} E(K_j|S,\pi)}$. The score equations can be solved to yield

$$\pi_j' = \begin{cases} E(K_j|S,\pi)\frac{\Lambda(\xi,S,\pi)}{1-\xi\Lambda(\xi,S,\pi)}, & (j : a_j < t) \\ E(K_j|S,\pi)\Lambda(\xi,S,\pi), & (j : a_j \geq t) \end{cases}$$

as the EM estimates for $\{\pi\}$ corresponding to a fixed value of ξ.

Combining the E and M-steps, we obtain an iterative procedure for computing the constrained MLEs $\{\tilde{\pi}\}$. We can compare the observed LRS, $2\{l(\hat{\pi}) - l(\tilde{\pi})\}$, with the relevant χ_1^2 quantile point, and obtain the limits of the Lagrange multiplier, $[\xi_-, \xi_+]$. These can be used to yield the required CI for $\mathcal{F}(t)$. The procedure employed here may be contrasted with the approach of Thomas and Grunkmeier (1975), who obtained a simpler expression for the constrained MLEs. However, the approach used here can be extended easily to handle more general patterns of missing data, as indicated in Cox and Oakes (1984).

A simple extension of the previous example is to consider two independent samples of failure times. As in Example 1.2, we can obtain an interval estimate of the NNT parameter via this approach. Following the notation of the preceding example, let $S_1 = (S_{11}, ..., S_{1n_1})$ and $S_2 = (S_{21}, ..., S_{2n_2})$ represent the observed data from the two samples, where n_i is the number of observations in sample i, $i = 1, 2$. The atoms of distinct failure times for each sample are denoted by $\{a_{ij}\}$.

$i = 1, 2$; $j = 1, ..., n_i$. The NNT parameter in this continuous-time setting is

$$\{\mathcal{F}_1(t) - \mathcal{F}_2(t)\}^{-1} = \{\sum^{(t)} \pi_{2j} - \sum^{(t)} \pi_{1j}\}^{-1},$$

where $\sum^{(t)}$ denotes summation over subjects with failure times less than t, for the respective samples. As in Chapter 1, it is simpler to first obtain an interval estimate for $\mathcal{F}_1(t) - \mathcal{F}_2(t)$, and then invert the result to obtain the corresponding interval estimate for the NNT. By using arguments similiar to those employed in Example 3.1, we obtain the constrained estimates

$$\tilde{\pi}_{1j} = \begin{cases} \frac{\lambda_1}{1-\xi\lambda_1} E(K_{1j}|S_1, \pi_1), & (j : a_{1j} < t) \\ \lambda_1 E(K_{1j}|S_1, \pi_1), & (j : a_{1j} \geq t) \end{cases}$$

$$\tilde{\pi}_{2j} = \begin{cases} \frac{\lambda_2}{1+\xi\lambda_2} E(K_{2j}|S_2, \pi_2), & (j : a_{2j} < t) \\ \lambda_2 E(K_{2j}|S_2, \pi_2), & (j : a_{2j} \geq t) \end{cases}$$

where

$$\lambda_1 = \lambda_1(S_1, \pi_1, \xi) = \frac{\xi + n_1 \pm \sqrt{(\xi + n_1)^2 - 4\xi \sum_{a_{1j} \geq t} E(K_{1j}|S_1, \pi_1)}}{2\xi \sum_{a_{1j} \geq t} E(K_{1j}|S_1, \pi_1)},$$

$$\lambda_2 = \lambda_2(S_2, \pi_2, \xi) = \frac{\xi - n_2 \pm \sqrt{(\xi - n_2)^2 + 4\xi \sum_{a_{2j} \geq t} E(K_{2j}|S_2, \pi_2)}}{2\xi \sum_{a_{2j} \geq t} E(K_{2j}|S_2, \pi_2)},$$

and $\pi_1 = (\pi_{11}, ..., \pi_{1n_1})^T$, $\pi_2 = (\pi_{21}, ..., \pi_{2n_2})^T$ denote the current estimates of the parameters.

Example 3.5

In Example 3.3, interest centered on interval estimation of $\Lambda(t)$, $F_1(t)$ and $P(t)$. However, we did not discuss the interval estimation of $\Lambda(t)$ then as a Lagrange multiplier approach is simpler. Let $\Lambda(t) = \omega$, for some $\omega > 0$. By differentiating the Lagrangian, $\ell_0 + \xi\{\omega - \Lambda(t)\}$, with respect to the parameters, we can easily obtain the score vector. As in Example 3.1, only the score equations for $\lambda_1(u), u \leq t$ are affected:

$$\frac{N_1(u)}{\lambda_1(u)} - \frac{R_1(u) - N_1(u) - N_2(u)}{\lambda_3(u)} - \xi = 0,$$

where $\lambda_3(u) = 1 - \lambda_1(u) - \lambda_2(u)$. These equations can be solved explicitly to yield

$$\bar{\lambda}_1(u) = \begin{cases} \frac{N_1(u)\bar{\lambda}_3(u)}{R_1(u) - N_1(u) - N_2(u) + \xi\bar{\lambda}_3(u)}, & u \leq t \\ \frac{N_1(u)}{R_1(u)}, & u > t \end{cases}$$

$$\bar{\lambda}_2(u) = \begin{cases} \frac{N_2(u)\bar{\lambda}_3(u)}{R_1(u) - N_1(u) - N_2(u)}, & u \leq t \\ \frac{N_2(u)}{R_1(u)}, & u > t \end{cases}$$

where

$$\bar{\lambda}_3(u) = \frac{\xi - R_1(u) \pm \sqrt{(R_1(u) - \xi)^2 - 4\xi(R_1(u) - N_1(u))}}{2\xi(R_1(u) - N_1(u))}\{R_1(u) - N_1(u) - N_2(u)\}.$$

In this case, the EM1 algorithm is not required. However, we need to numerically evaluate which root of the solution is feasible. Since the constraint does not affect the E-step, we can substitute the conditional expectations of $\{N_1(u)\}$ and $\{R_1(u)\}$ into the preceding formulae.

Next, we show that when a Lagrange multiplier argument is applied to the case of $F_1(t)$, relatively simple numerical techniques can be used at the M-step of the EM algorithm; in particular, fixed-point methods can be used in place of

Newton-type iterative schemes (as used in the EM1 algorithm). Consider fixing $F_1(t) = \omega$. We form the Lagrangean, $\ell_0 + \xi\{\omega - F_1(t)\}$, from which the partial derivatives with respect to the parameters are straightforward to obtain. Clearly, the estimating equations for $\{\lambda_{11}(k|i)\}$, $k \geq i$, $\forall i$, $\{\lambda_1(j)\}$ and $\{\lambda_2(j)\}$, $j > t$, are the same as those for the unrestricted model; these can be solved explicitly in terms of the observed data and the previous cycle updates. However, closed-form estimates are not available for the remaining parameters. To tackle the problem, the obvious strategy is to use the EM1 algorithm to update these parameters, given the matrix of second partial derivatives for the parameters concerned. We shall not implement this approach here. A numerical illustration of the Lagrange multiplier form of the EM1 algorithm will be considered in Example 3.7. Instead, we point out the following alternative numerical approach which can potentially relieve the numerical burden (cf. Example 1.2).

It is straightforward to show that the estimating equations for the parameters in $F_1(t)$ can be written as

$$\lambda_1(j) = \frac{N_1(j)\ [\ 1 - \lambda_2(j)\ ]}{R_1(j) - N_2(j) + \xi\ [\ 1 - \sum_{i=1}^{j} \lambda_2(i) S_1(i) - F_1(t)\ ]}\ ,$$

$$\lambda_2(j) = \frac{N_2(j)\ [\ 1 - \lambda_1(j)\ ]}{R_1(j) - N_1(j) - \xi\ [\ F_1(t) - \sum_{i=1}^{j} \lambda_1(i) S_1(i)\ ]}\ ,$$

for $j \leq t$. Given the previous estimates and the observed data, this is a fixed-point system of equations which can potentially be solved by the method of functional iteration. A fixed-point system of equations can similarly be obtained for the case of $P(t)$. Finally, we note that when the usual profile likelihood approach is used, the score equations cannot be similarly written as a fixed-point system of equations.

Example 3.6

When the complete data belong to the regular exponential family, Dempster *et. al.* (1977) show that the EM algorithm yields an intuitively appealing result. Suppose the density of T_i is a regular member of the exponential family in its natural parametrization, i.e.,

$$f_i(t;\theta) = \exp[\theta^T S_i(t) + A_i(t) + B_i(\theta)],$$

where $S_i(t)$ is a $p \times 1$ vector of linearly independent functions of t, $A_i(t)$ is a scalar function of t and $B_i(\theta)$ is a scalar function of the natural parameter θ. It can be shown that the E and M-steps in this case combine to give

$$E(\mathbf{S}|\theta') = E(\mathbf{S}|T_{obs}:\theta)$$

as the updating scheme, where $\mathbf{S} = (S_1(t), \ldots, S_p(t))^T$. This resembles the estimating equation for the complete data setting which equates each element of the score vector to its expected value. An attractive result of this is that computer programmes for the complete data problem can be utilized in the corresponding missing data case.

Consider the interval estimation of the functional $f(\theta)$; let us assume a fixed value for $f(\theta)$, say $f(\theta) = \omega$. For constrained maximization, the E-step is unchanged, while the M-step is obtained by a simple modification to that for the unrestricted problem. The E-step consists of computing $E(\mathbf{S}|T_{obs};\theta)$, as before. In the M-step, we solve for θ' in

$$E(\mathbf{S}|\theta') + \xi\frac{\partial h(\theta')}{\partial \theta'} = E(\mathbf{S}|T_{obs};\theta).$$

where $h(\theta') = \omega - f(\theta')$. This follows by setting $\frac{\partial Q_\xi}{\partial \theta'} = 0$ and noting that $\frac{\partial B_i(\theta)}{\partial \theta} = E(\mathbf{S}|T_{obs};\theta)$ (from the unconstrained problem). In general, numerical methods are required to solve the resulting system of equations. Unfortunately. complete data computer code cannot be applied as routinely as was possible for the unconstrained maximization problem. Specially written programmes appear unavoidable. However, in chapter 5 we indicate a possible numerical method to reduce the computational load.

To illustrate the constrained estimation procedure. consider the case of a partially observed normal random sample (Little and Rubin. 1987, p. 93). Assume that we have a random sample $X_i \sim N(\mu,\sigma)$, $i = 1,\ldots,n$, of which l are observed. $Y_{obs} = (Y_1,\ldots,Y_l)^T$. and $n - l$ are missing at random. Here $\theta = (\theta_1,\theta_2)^T$. where $\theta_1 = \frac{\mu}{\sigma^2}$. and $\theta_2 = -\frac{1}{2\sigma^2}$. and $S = (\sum_{i=1}^n Y_i,\ \sum_{i=1}^n Y_i^2)^T$. Suppose we are interested in obtaining an approximate 100 $(1-\alpha)$% CI for the functional $f(\theta) = -\frac{2\theta_2}{\theta_1^2} = (\frac{\sigma}{\mu})^2$. the square of the coefficient of variation. In the E-step, we obtain

$$E(\sum_i Y_i \mid Y_{obs},\theta) = \sum_{i=1}^{l} y_i - (n-l)\frac{\theta_1}{2\theta_2}$$

$$E(\sum_i Y_i^2 \mid Y_{obs},\theta) = \sum_{i=1}^{l} y_i^2 + (n-l)(\frac{\theta_1^2}{4\theta_2^2} - \frac{1}{2\theta_2}).$$

For the M-step. we also need

$$E(\sum_i Y_i \mid \theta') = -n\theta_1'(2\theta_2')^{-1}$$

$$E(\sum_i Y_i^2 \mid \theta') = n\left[\frac{(\theta_1')^2}{4(\theta_2')^2} - (2\theta_2')^{-1}\right].$$

The E and M-steps together yield the updating equations for $(\mu, \sigma)^T$ as

$$n\mu' + 2\xi\frac{(\sigma')^4}{(\mu')^3} = \sum_{i=1}^{l} y_i + (n-l)\mu ,$$

$$n[(\mu')^2 + (\sigma')^2] - 2\xi\frac{(\sigma')^4}{(\mu')^2} = \sum_{i=1}^{l} y_i^2 + (n-l)(\mu^2 + \sigma^2) .$$

[note for unconstrained maximization, i.e., $\xi = 0$, this system is easily solved]. These may be re-expressed in the form

$$\mu^{(m+1)} = n^{-1}\sum_{i=1}^{l} y_i + (1 - \frac{l}{n})\mu^{(m)} - 2\xi\frac{(\sigma^{(m+1)})^4}{(\mu^{(m+1)})^3} ,$$

$$(\sigma^{(m+1)})^2 = n^{-1}\sum_{i=1}^{l} y_i^2 + (1 - \frac{l}{n})[(\mu^{(m)})^2 + (\sigma^{(m)})^2] - (\mu^{(m+1)})^2 + 2\xi\frac{(\sigma^{(m+1)})^4}{n(\mu^{(m+1)})^2}$$

where $(\mu^{(m)}, \sigma^{(m)})^T$ denotes the estimate of $(\mu, \sigma)^T$ at the mth cycle of the EM algorithm. Given a fixed ξ and current estimates $(\mu^{(m)}, \sigma^{(m)})^T$, a numerical method (e.g. fixed-point iteration) is required to obtain the updated estimates. The LRS can be computed and compared with the relevant χ_1^2 quantile.

In some problems, by using a suitable monotone transformation of the functional of interest, one might obtain closed-form expressions for the M-step. For our example, instead of $f(\theta) = -\frac{2\theta_2}{\theta_1^2} = (\frac{\sigma}{\mu})^2$, consider $1/f = (\frac{\mu}{\sigma})^2$. Then the vector of derivatives with respect to θ is $(\frac{-\theta_1}{\theta_2}, \frac{\theta_1^2}{2\theta_2^2})$. The E-step remains unchanged, while the M-step becomes

$$-n\frac{\theta_1'}{2\theta_2'} + \xi\frac{\theta_1'}{\theta_2'} = \sum_{i=1}^{l} y_i - (n-l)\frac{\theta_1}{2\theta_2} , \qquad (3.11)$$

$$n\left[\frac{(\theta_1')^2}{4(\theta_2')^2} - \frac{1}{2\theta_2'}\right] - \xi\frac{(\theta_1')^2}{2(\theta_2')^2} = \sum_{i=1}^{l} y_i^2 + (n-l)\left(\frac{\theta_1^2}{4\theta_2^2} - \frac{1}{2\theta_2}\right) . \qquad (3.12)$$

From equation (3.11), we obtain

$$\theta_1' = A_1 \theta_2' ,$$

where $A_1 = \frac{\sum_{i=1}^{l} y_i - (n-l)\frac{\theta_1}{2\theta_2}}{\xi - \frac{n}{2}}$. This can be substituted into equation (3.12) to give

$$\theta_2'(nA_1^2 - 2\xi A_1^2 - 2A_2) = 2n ,$$

where $A_2 = \sum_{i=1}^{l} y_i^2 + (n-l)(\frac{\theta_1^2}{4\theta_2^2} - \frac{1}{2\theta_2})$. Hence we choose

$$\theta_2' = \frac{2n}{nA_1^2 - 2\xi A_1^2 - 2A_2}$$

as the current update, given the data, previous estimates and a fixed value of the Lagrange multiplier.

Example 3.2 (Cont.)

Closed-form expressions can similarly be obtained at the M-step for Example 3.2, if we consider $\log \frac{\sigma_\alpha^2}{\sigma_e^2}$ instead of $\frac{\sigma_\alpha^2}{\sigma_e^2}$. Given a fixed value of the Lagrange multiplier, ξ, the M-step gives

$$\frac{\partial \ell_\xi}{\partial \sigma_\alpha^2} = -\frac{I}{2\sigma_\alpha^2} + \frac{1}{2}\sum_i \frac{(\alpha_i - \mu)^2}{\sigma_\alpha^4} - \frac{\xi}{\sigma_\alpha^2} = 0,$$

$$\frac{\partial \ell_\xi}{\partial \sigma_e^2} = -\frac{IJ}{2\sigma_e^2} + \frac{1}{2}\sum_i \sum_j \frac{(y_{ij} - \alpha_i)^2}{\sigma_e^4} + \frac{\xi}{\sigma_e^2} = 0,$$

and the equation for μ is unchanged. Hence,

$$\tilde{\sigma}_\alpha^2 = \sum_i \frac{(\alpha_i - \tilde{\mu})^2}{I + 2\xi},$$

Table 3.8: Results of Apex data analysis.

$(\mu^{(0)}, (\sigma_e^{(0)})^2, (\sigma_\alpha^{(0)})^2)$	ξ	$(\tilde{\mu}, \tilde{\sigma}_e^2, \tilde{\sigma}_\alpha^2)$	$\frac{\tilde{\sigma}_\alpha^2}{\tilde{\sigma}_e^2}$
(71, 20, 200)	1.294	(71, 120.026, 8.536)	0.071
(71, 8, 114)	-1.607	(71, 61.506, 221.383)	3.599

$$\tilde{\sigma}_e^2 = \sum_i \sum_j \frac{(y_{ij} - \alpha)^2}{IJ - 2\xi},$$

and $\tilde{\mu} = I^{-1} \sum_i \alpha_i$ as before. For numerical illustration, we consider the Apex data of Example 3.2. For this data set, we find that the EM algorithm is sensitive to the choice of starting values. Various starting values were used and the final log likelihood values were compared to obtain the constrained MLEs. We also found that, in contrast to the profile likelihood approach, the unrestricted MLE was not a suitable starting point for both end-points of the interval estimate: the algorithm could not increase the log likelihood from this initial position. Table 3.8 summarizes the details of the procedure; the level of confidence is fixed at 0.90. In the table, $(\mu^{(0)}, (\sigma_e^{(0)})^2, (\sigma_\alpha^{(0)})^2)$ denotes the starting value of our iteration, while $(\tilde{\mu}, \tilde{\sigma}_e^2, \tilde{\sigma}_\alpha^2)$ represents the constrained MLE. For negative ξ values, it took 8-9 iterations for convergence while only 1-2 iterations were required for each positive ξ value. We also used the Lagrange multiplier form of the EM1 algorithm to obtain an interval estimate for $\frac{\sigma_\alpha^2}{\sigma_e^2}$ directly. Let $\ell_\xi = \ell_0 + \xi\{\omega - \frac{\sigma_\alpha^2}{\sigma_e^2}\}$ and $\theta = (\mu, \sigma_e^2, \sigma_\alpha^2)^T$. We obtain

$$\frac{\partial \ell_\xi}{\partial \theta} = \left(\sum_i \frac{(\alpha_i - \mu)}{\sigma_\alpha^2}, -\frac{IJ}{2\sigma_e^2} + \frac{1}{2} \sum_i \sum_j \frac{(y_{ij} - \alpha_i)^2}{\sigma_e^4} + \frac{\xi \sigma_\alpha^2}{\sigma_e^4}, -\frac{I}{2\sigma_\alpha^2} + \frac{1}{2} \sum_i \frac{(\alpha_i - \mu)^2}{\sigma_\alpha^4} - \frac{\xi}{\sigma_e^2} \right)^T.$$

Table 3.9: Results of Apex data analysis using the EM1 algorithm with Lagrange multiplier.

$(\mu^{(0)}, (\sigma_e^{(0)})^2, (\sigma_\alpha^{(0)})^2)$	ξ	$(\bar{\mu}, \bar{\sigma}_e^2, \bar{\sigma}_\alpha^2)$	$\frac{\bar{\sigma}_\alpha^2}{\bar{\sigma}_e^2}$
(71, 7, 114)	-47.65	(71, 100.326, 7.932)	0.079
(71, 116, 41)	6.06	(71, 58.795, 210.227)	3.575

The M-step clearly does not yield a closed-form solution for θ. To implement the M-step numerically, we also need $\frac{\partial \ell_\xi^2}{\partial\theta\partial\theta^T}$, i.e.,

$$\begin{pmatrix} -\frac{I}{\sigma_\alpha^2} & 0 & -\sum_i \frac{(\alpha_i-\mu)}{\sigma_\alpha^4} \\ 0 & \frac{IJ}{2\sigma_e^4} - \sum_i \sum_j \frac{(y_{ij}-\alpha_i)^2}{\sigma_e^6} - \frac{2\xi\sigma_\alpha^2}{\sigma_e^6} & \frac{\xi}{\sigma_e^4} \\ -\sum_i \frac{(\alpha_i-\mu)}{\sigma_\alpha^4} & \frac{\xi}{\sigma_e^4} & \frac{I}{2\sigma_\alpha^4} - \sum_i \frac{(\alpha_i-\mu)^2}{\sigma_\alpha^6} \end{pmatrix}$$

This algorithm was also sensitive to the choice of starting value, so different starting values were used and their respective log likelihood values compared to determine the constrained MLE. The results are summarized in Table 3.9. The number of iterations for each fixed ξ value is comparable to those required for the calculations summarized in Table 3.8. However, the total number of iterations required to locate the end-points of the interval estimate is greater in this case since $|\xi|$ is larger.

3.4 EM Algorithm and Decomposable Likelihoods

Recently, Gu (1996) developed versions of the EM algorithm by suitably decomposing the likelihood function of the data (typically called the incomplete data likelihood). A novel feature of his method is that, given the mathematical form of the likelihood function, the Q function can be written down without specifying the probability structure associated with the corresponding complete data likelihood.

Using this approach, closed-form parameter estimates for certain models were derived and shown to be consistent with those based on the usual EM algorithm. In the following, we describe the basic approach derived by Gu, and clarify some points concerning the method. We then utilize the Lagrange multiplier argument to derive interval estimates for functionals in some models.

3.4.1 Introduction

Gu's approach deals with likelihoods of the form

$$\log L(\phi) = \sum_{i \in \mathbf{I}} a_i \log p_i(\phi) \, .$$

where $a_i \geq 0$, $p_i(\phi) \geq 0$, $\mathbf{I}$ is an arbitrary index set, and $\phi \in \Omega$ is the unknown parameter (vector). An example of such a likelihood was encountered in Example 3.1. We further suppose that $p_i(\phi)$ admits a simple decomposition, viz., $p_i(\phi) = \sum_{j \in \mathbf{I}_i} f_j(\phi)$, where $f_j(\phi) \geq 0$ for any $j \in \mathbf{I}_i$, and $\mathbf{I}_i$ represents a subset in a partition of the index set $\mathbf{I}$. Based on the form of the likelihood function, Gu defined the Q function

$$Q(\phi'|\phi) = \sum_i a_i \sum_{j(i)} \frac{f_j(\phi)}{\sum_{k(i)} f_k(\phi)} \log f_j(\phi') \, .$$

where $\sum_i$ and $\sum_{j(i)}$ denote $\sum_{i \in \mathbf{I}}$ and $\sum_{j \in \mathbf{I}_i}$, respectively. Gu's generalization of the EM algorithm consists of the following steps:

E-step. Compute $Q(\phi'|\phi)$.

M-step. Maximize $Q(\phi'|\phi)$ with respect to ϕ'.

The algorithm can be shown to satisfy the same basic properties as the usual EM algorithm. In fact, Gu states that the generalized version of the EM algorithm yields the usual version under certain conditions (see below).

We can make the following observations concerning the generalized algorithm. The first pertains to the generality of the algorithm. Specifically, the algorithm is applicable when the complete data model can be expressed in the form of a multinomial distribution. Another comment concerns the rate of convergence of the modified algorithm. In Example 1 of Gu (1996), it was noted that different decompositions of the likelihood led to different rates of convergence to the unique parameter estimates. The example presented data in which 197 animals are distributed multinomially into four categories, so that the observed data consist of

$$\mathbf{y} = (y_1, y_2, y_3, y_4) = (125, 18, 20, 34).$$

A genetic model for the population specifies cell probabilities

$$\left(\frac{1}{2} + \frac{1}{4}\pi, \frac{1}{4}(1-\pi), \frac{1}{4}(1-\pi), \frac{1}{4}\pi\right),$$

with $0 \le \pi \le 1$. The observed data log likelihood is

$$\ell(\pi) = y_1 \log(\frac{1}{2} + \frac{1}{4}\pi) + (y_2 + y_3)\log(1-\pi) + y_4 \log \pi.$$

For this log likelihood, Gu obtained two decompositions, resulting in the following Q functions,

$$\begin{aligned} Q_1(\pi'|\pi) &= y_1 \frac{1/2}{1/2+\pi/4} \log 1/2 + y_1 \frac{\pi/4}{1/2+\pi/4} \log \pi' + (y_2+y_3)\log(1-\pi') + y_4 \log \pi'. \\ Q_2(\pi'|\pi) &= \frac{3y_1\pi/4}{3\pi/4+(1-\pi)/2} \log \pi' + \frac{y_1(1-\pi)/2}{3\pi/4+(1-\pi)/2} \log(1-\pi') \\ &+ (y_2+y_3)\log(1-\pi') + y_4 \log \pi'. \end{aligned}$$

The values of π that maximizes these Q functions are, respectively, $\frac{A_1(\pi)}{A_1(\pi)+38}$ and

Table 3.10: A comparison of the convergence rates of the two versions of the EM algorithm in Example 1 (Gu, 1996).

	$\pi^{(k+1)} = \frac{A_1(\pi^{(k)})}{A_1(\pi^{(k)})+38}$		$\pi^{(k+1)} = \frac{A_2(\pi^{(k)})}{A_2(\pi^{(k)})+B(\pi^{(k)})}$	
k	$\pi^{(k)}$	$\lvert\pi^{(k)} - \pi^*\rvert$	$\pi^{(k)}$	$\lvert\pi^{(k)} - \pi^*\rvert$
0	0.2000000	0.4268215	0.2000000	0.4268215
1	0.5441658	0.0826574	0.3456391	0.2811824
2	0.6153352	0.0116863	0.4530849	0.1737366
5	0.6267936	0.0000276	0.5935573	0.0332642
10	0.6268215	0.0000000	0.6250739	0.0017476

$\frac{A_2(\pi)}{A_2(\pi)+B(\pi)}$, where

$$A_1(\pi) = 125\pi/(2+\pi) + 34.$$

$$A_2(\pi) = 375\pi/(2+\pi) + 34$$

and

$$B(\pi) = 250(1-\pi)/(2+\pi) + 38.$$

Using these decompositions, Gu performed a comparison of their convergence rates: a portion of the numerical results are reproduced in Table 3.10. Gu noted that the first decomposition was uniformly superior and therefore remarked that careful consideration is required for selecting a suitable decomposition of a given likelihood. For this particular example, we find that the difference in convergence rates can be explained by the proportion of 'missingness' induced by the decomposition. This follows from the relation

$$I(\phi|Y_{obs}) = -\frac{\partial^2 Q}{\partial\phi\partial\phi^T} + \frac{\partial^2 R}{\partial\phi\partial\phi^T},$$

which has the well-known interpretation:

Observed information = complete information - missing information. Dempster *et al.* (1977) established that the rate of convergence of the EM algorithm is related to the preceding quantities. They showed that the greater the proportion of missing data, the slower the rate of convergence. Specifically, if the EM iterates $\phi^{(m)}$ converge to ϕ^*, then for $\phi^{(m)}$ near ϕ^*,

$$|\phi^{(m+1)} - \phi^*| = \lambda|\phi^{(m)} - \phi^*|\,.$$

where λ is the ratio of the missing to the complete information for scalar ϕ or the largest eigenvalue of the corresponding matrix for vector ϕ. Hence for the first decomposition in Gu's example, $\lambda = 0.1328$, while for the second decomposition, $\lambda = 0.5517$. Further research into these aspects of Gu's version of the EM algorithm in the multi-parameter and constrained parameter settings would be desirable.

3.4.2 Application to Constrained Maximization

In this subsection, we shall explore the use of Gu's simplified EM algorithm to yield interval estimates for functionals in some problems. As in the case of the usual EM algorithm, standard arguments show that Gu's version of the EM algorithm continues to be self-consistent for the constrained parameter setting. Instead of delving into more theoretical properties of the algorithm, we shall illustrate the use of the algorithm for some simple constrained problems.

Example 3.7

This example is taken from Gu (1996). We consider a failure time experiment involving a control group, labelled 0, and J dose groups. Let $\lambda_j(t)$ be the hazard function for group j at time t. A restricted proportional hazards model is defined by

$$\lambda_j(t) = \lambda_0(t)\theta_j$$

where $0 \leq \theta_j \leq 1$, $1 \leq j \leq J$, and $\theta_0 = 1$. Let $t_1 < \cdots < t_m$ denote the m distinct failure times from the sample; for convenience we denote $\lambda_j(t_k)$ by $\lambda_j(k)$. For these data, the log likelihood function can be written as

$$\ell(\lambda, \theta) = \sum_{j=0}^{J}\sum_{k=1}^{m} I_{jk}\log\lambda_j(k) + \sum_{j=0}^{J}\sum_{k=1}^{m}\Delta_{jk}\log\{1 - \lambda_j(k)\}$$

where I_{jk} is the multiplicity of failures from group j at t_k, and Δ_{jk} is the number of subjects in group j surviving beyond t_k. By using the decomposition $1 - \lambda_j(k) = \{1 - \lambda_0(k)\} + \{\lambda_0(k)(1 - \theta_j)\}$, Gu obtained

$$\begin{aligned} Q(\lambda', \theta' | \lambda, \theta) &= \sum_{j=0}^{J}\sum_{k=1}^{m} I_{jk}\log\lambda_j'(k) \\ &+ \sum_{j=0}^{J}\sum_{k=1}^{m}(1 - f_{jk})\Delta_{jk}\log\{\lambda_0'(k)(1 - \theta_j')\} \\ &+ \sum_{j=0}^{J}\sum_{k=1}^{m} f_{jk}\Delta_{jk}\log\{1 - \lambda_0'(k)\}, \end{aligned}$$

where $f_{jk} = \frac{1-\lambda_0(k)}{1-\theta_j\lambda_0(k)}$. Closed-form solutions for the parameter estimates can be easily derived by maximizing Q.

For this problem, the cumulative hazard of failure for dose group r, $\sum_{u=1}^{t}\lambda_r(u)$, might be of interest. (We focus on just one cumulative hazard as we can extend the method easily to include more cumulative hazards.) We find that it is simpler to first obtain an interval estimate for $\log\sum_{u=1}^{t}\lambda_r(u)$, and then to exponentiate the end-points of the interval to get the desired interval estimate. The augmented Q

function is given by

$$Q_\xi(\lambda',\theta'|\lambda,\theta) = Q(\lambda',\theta'|\lambda,\theta) + \xi[\omega - \log\{\theta'_r \sum_{u=1}^{t} \lambda'_0(u)\}] .$$

Maximizing Q_ξ with respect to θ is straightforward. For $j \neq r$, the constrained estimate for θ_j is equal to the unrestricted estimate, i.e.,

$$\tilde{\theta}_j = \frac{\sum_k I_{jk}}{\sum_k\{I_{jk} + (1 - f_{jk})\Delta_{jk}\}} .$$

while

$$\tilde{\theta}_r = \frac{\sum_k I_{rk} - \xi}{\sum_k I_{rk} - \xi + \sum_k(1 - f_{rk})\Delta_{rk}} .$$

For $k > t$, the constrained estimate of $\lambda_0(k)$ is the same as the unrestricted one.

$$\tilde{\lambda}_0(k) = \frac{\sum_{j=0}^{J} I_{jk}}{\sum_{j=0}^{J}\{(1 - w_{jk})\Delta_{jk}\}} .$$

where $w_{jk} = \frac{1-\theta_j}{1-\theta_j\lambda_0(k)}$. For $k \leq t$, we do not obtain closed-form estimates. However, a simple numerical procedure such as functional iteration can be applied to

$$\tilde{\lambda}_0(k) = \frac{\delta \pm \sqrt{\delta^2 + 4\xi\{\sum_k I_{jk} + \sum_k(1 - f_{jk})\Delta_{jk}\}}}{2\xi} .$$

where

$$\delta = \xi - \sum_k I_{jk} - \sum_k (1 - f_{jk})\Delta_{jk} - \sum_k f_{jk}\Delta_{jk} \sum_{u=1}^{t} \tilde{\lambda}_0(u) .$$

For a fixed ξ, we compare $2\{\ell(\hat{\lambda},\hat{\theta}) - \ell(\tilde{\lambda},\tilde{\theta})\}$ with the relevant $(1-\alpha)$ quantile point of the chi-squared distribution with 1 df. If these are discrepant, we adjust the value of ξ accordingly and iterate. Because of the nonparametric setting in which the likelihood function arose, it is difficult to assign an exact confidence level to the resulting interval estimate. These intervals are nevertheless usefully regarded as approximate $100(1-\alpha)\%$ confidence intervals.

Example 3.8

Frydman (1992) considers a three-state Markov process with irreversible transitions and interval-censored data. She develops a nonparametric maximum likelihood procedure for the estimation of the parameters of this model. In particular, joint MLEs of F. the distribution function of time in the first state. and Λ_2, the cumulative intensity of transitions from state 2 to state 3. were obtained. Gu (1996) obtained equivalent estimates for this model, using his novel approach. In both cases however. no interval estimates for the functionals were provided. In Frydman's solution. the complicated form of the likelihood function mitigates against a likelihood-based approach. Gu's method opens up the possibility of a tractable likelihood-based approach. We illustrate this with a few functionals mentioned in Frydman (1992). after first outlining some notation.

Consider the log likelihood function derived by Frydman:

$$\begin{aligned}\ell(z,\lambda) &= \sum_{j=1}^{J} \log \sum_{i=1}^{I} \delta_{ij} z_i + \sum_{n=1}^{N} d_n \log \lambda_n + \sum_{m=1}^{M} \sum_{G_m} \log(1-\lambda_n) \\ &+ \sum_{m=1}^{M} \log\{\sum_{i=1}^{I} \prod_{(r_i,R_m]} (1-\lambda_n)\alpha_{im} z_i\},\end{aligned}$$

where δ_{ij} and α_{im} are indicator variables. d_n is a positive integer. G_m is an interval such that

$$\sum_{G_m} \log(1-\lambda_n) = \sum_{n=1}^{N} \log\{1-\lambda_n I(t_n^* \in G_m)\}.$$

t_n^*, r_i, R_m are positive real numbers, and

$$\prod_{(r_i,R_m]} (1-\lambda_n)\alpha_{im}z_i = \prod_{n=1}^{N}\{1-\lambda_n I(t_n^* \in (r_i, R_m])\}\alpha_{im}z_i .$$

Frydman parametrized the nonparametric likelihood function in terms of (F, Λ_2), and showed that maximum likelihood estimation of these "parameters" is equivalent to maximizing $\ell(z, \lambda)$ with respect to (z, λ), subject to

$$\sum_{i=1}^{I} z_i = 1, \ z_i \geq 0 \quad (1 \leq i \leq I).$$

and

$$0 < \lambda_n \leq 1 \quad (1 \leq n \leq N).$$

Gu (1996) shows that the Q function is given by

$$\begin{aligned} Q(z', \lambda' | z, \lambda) &= \sum_{j=1}^{J}\sum_{i=1}^{I} \frac{\delta_{ij}z_i}{\sum_{p=1}^{I}\delta_{pj}z_p}\log z_i' \\ &+ \sum_{n=1}^{N} d_n \log \lambda_n' + \sum_{m=1}^{M}\sum_{G_m}\log(1-\lambda_n') \\ &+ \sum_{m=1}^{M}\sum_{i=1}^{I} \frac{\alpha_{im}z_i\prod_{(r_i,R_m]}(1-\lambda_n)}{\sum_{p=1}^{I}\alpha_{pm}z_p\prod_{(r_p,R_m]}(1-\lambda_n)}\{\log \prod_{(r_i,R_m]}(1-\lambda_n') + \log z_i'\}, \end{aligned}$$

which is easily maximized with respect to (z', λ'). For a fixed time point s, consider interval estimation for the distribution function $F(s)$.

$$F(s) = \begin{cases} 0, & s < l_1 \\ \sum_{p=1}^{i} z_p, & r_i \leq s < l_{i+1} \quad (1 \leq i \leq I-1) \\ 1, & s \geq r_I \end{cases}$$

where the l_i's and r_i's are real and ordered, $l_1 \leq r_1 < l_2 \leq r_2 < \cdots < l_I \leq r_I$. Let

$$\mu_{im}(z, \lambda) = \sum_j \frac{\delta_{ij} z_i}{\sum_p \delta_{pj} z_p} + \sum_m \frac{\alpha_{im} z_i \prod_{(r_i, R_m]}(1 - \lambda_n)}{\sum_p \alpha_{pm} z_p \prod_{(r_p, R_m]}(1 - \lambda_n)}$$

This maximization problem is very similiar to the one considered in Example 3.1. Using the Lagrangian with multiplier ξ, we easily obtain the constrained MLEs for z as

$$\bar{z}_i = \begin{cases} \frac{\Lambda}{1+\xi\Lambda}\mu_{im}(z, \lambda), & \{i : r_i \leq s\} \\ \Lambda\mu_{im}(z, \lambda), & \{i : s < r_i\} \end{cases}$$

where

$$\Lambda = \frac{\xi - \sum_{i=1}^N \mu_{im}(z, \lambda) \pm \sqrt{(\xi - \sum_{i=1}^N \mu_{im}(z, \lambda))^2 + 4\xi \sum_{i>s} \mu_{im}(z, \lambda)}}{2\xi \sum_{i>s} \mu_{im}(z, \lambda)}.$$

The form of the constrained MLE for λ is the same as that of the unrestricted estimate. Analogous processes for the cumulative intensity,

$$\Lambda_2((0, t]) = \begin{cases} \sum_{n=1}^N \lambda_n I(t_n^* \leq t), & t \leq t_{max} \\ \text{undefined}, & t > t_{max} \end{cases}$$

lead to the constrained MLEs

$$\bar{\lambda}_n = \frac{\zeta_{mi}(z, \lambda) + \gamma_{mi}(z, \lambda) \pm \sqrt{(\zeta + \gamma_{mi}(z, \lambda))^2 - 4\xi I(t_n^* \leq t) d_n}}{2\xi I(t_n^* \leq t)}$$

for $t \leq t_{max}$, where

$$\begin{aligned} \zeta_{mi}(z, \lambda) &= d_n + \xi I(t_n^* \leq t) + \gamma_{mi}(z, \lambda), \\ \gamma_{mi}(z, \lambda) &= \frac{\alpha_{im} z_i \prod_{(r_i, R_m]}(1 - \lambda_n)}{\sum_{p=1}^I \alpha_{pm} z_p \prod_{(r_p, R_m]}(1 - \lambda_n)}. \end{aligned}$$

The constrained MLEs for z are unchanged in form. Even simpler closed-form estimates for the parameters exist, when the functional of interest is the conditional survivor function in state 2, $G_s(v) = \prod_{n=1}^{N}\{1 - \lambda_n I(s < t_n^* \leq v + s)\}$, s, v being fixed. In this case, it is easier to obtain an interval estimate for $\log G_s(v)$, and transform accordingly to obtain the interval for $G_s(v)$. The constrained MLEs for λ are easily found to be

$$\bar{\lambda}_n = \frac{d_n}{d_n + \sum_m I(t_n^* \in G_m) + \sum_i \sum_m \gamma_{mi}(z, \lambda) I(t_n^* \in (r_i, R_m]) + \xi I(s < t_n^* \leq v + s)};$$

the constrained MLEs for z are unchanged in form.

Let us consider the cumulative intensity for transitions from states 1 to 2,

$$\Lambda_1((0, s]) = \begin{cases} 0, & s < l_1 \\ \sum_{p=1}^{i} z_p (1 - \sum_{k=1}^{p-1} z_k)^{-1}, & r_i \leq s < l_{i+1}, \quad 1 \leq i \leq I \end{cases}$$

and undefined, otherwise. Given the form of this function, no closed-form estimates for z are available. We shall use the EM1 algorithm to obtain the constrained MLEs numerically. For a fixed s, $r_i \leq s < l_{i+1}$, we set $\Lambda_1((0, s]) = \omega$, for an arbitrary ω. This equation is solved explicitly for z_i,

$$z_i = \{\omega - \sum_{p=1}^{i-1} z_p (1 - \sum_{k=1}^{p-1} z_k)^{-1}\}\{1 - \sum_{k=1}^{i-1} z_k\}.$$

It is relatively straightforward to obtain $\frac{\partial z_i}{\partial z_j}$, $j = 1, ..., i-1$, and $\frac{\partial^2 Q}{\partial z \partial \lambda}$, so that the EM1 algorithm is not difficult in principle to implement. We can also consider the following procedure for easing the numerical burden. This involves alternatively estimating z, given λ, and vice versa. Thus, in the M-step of the algorithm for z, we only need $\frac{\partial z_i}{\partial z_j}$ and $\frac{\partial^2 Q}{\partial z \partial z^T}$.

Chapter 4

Profile Likelihood-based Interval Estimation in Failure Time Models

4.1 Introduction

In this chapter. we consider alternative ways of deriving profile likelihood-based confidence intervals for functionals in various failure time models. Specifically. we look at the parametric Cox proportional hazards (PH) model first. The standard way to derive likelihood-based interval estimates in this setting is to maximize the full likelihood subject to the constraint imposed by the functional. This method however requires specially written programmes for each application. Aitkin and Clayton (1980) showed that the parametric Cox model can be fit by a clever manipulation of the model into a GLM. We shall adapt their approach to yield likelihood-based interval estimates for some common failure time functionals. This approach makes use of available software and is easy to implement. The proposed method is extended to the semi-parametric Cox model with piecewise-constant baseline hazards. as well as to failure models with nonlinear predictors.

Another formulation of parametric survival models was discussed in Therneau (1995). leading to the accelerated failure model as a special case. Maximum likelihood estimation in this case is carried out by iteratively reweighted least squares (IRLS). Again. likelihood-based interval estimation of functionals requires specially written programmes since the usual survival packages supply standard errors based on the normal approximation. Using the IRLS formulation, we show how profile likelihood-based interval estimates for some common functionals can be easily obtained using available software. Nevertheless. there are some other useful functionals which are not handled by the regular profile likelihood approach. For these functionals, a Lagrange multiplier argument facilitates interval estimation.

4.2 The GLM Approach for Parametric Models

4.2.1 Introduction

While Cox's (1972) semi-parametric regression model is commonly used in biomedical work, parametric Cox failure time models can be useful in certain applications. In this section, we focus on the GLM formulation of the parametric Cox model described by Aitkin and Clayton (1980). This novel approach facilitates the use of GLIM (or other generalized linear model software) to fit proportional hazard regression models to right-censored survival data. Based on this method, common parametric failure time distributions such as the exponential, Weibull or extreme value distributions may be fit by expressing the likelihood in each case as a "Poisson" likelihood with a log-linear model for the "Poisson" mean.

Our interest lies in the interval estimation of functions of the model parameters. Usually, interval estimates for the regression coefficients can be derived by assuming the MLEs are asymptotically normal. The delta method can then be used to derive the asymptotic standard errors for the MLE of the functional. Since the normality assumption may not always be warranted, especially in the presence of moderate to heavier levels of censoring, a likelihood-based approach is desirable. The standard approach of obtaining such intervals would be via regular profile likelihood, i.e., to maximize the likelihood with respect to the parameters, subject to the constraint imposed by the functional. This involves a considerable amount of programming to suit each specific model and functional. Our proposed method makes use of the available software and is easy to implement. In the following, we briefly describe the GLM approach of Aitkin and Clayton (1980). We then adapt their approach to obtain likelihood-based interval estimates for some important functionals in common failure time models.

Let t_i, $i = 1, \ldots, n$ represent the event times of n individuals. Let w_i denote the corresponding indicator variables, taking the value 1 for uncensored, and 0 for censored events. The density and survivor functions for the failure times are denoted by $f(t)$ and $S(t)$, respectively. The hazard rate for the ith individual with covariate value $\mathbf{x_i}$ is assumed to be given by

$$h(t_i; \mathbf{x_i}) = \lambda(t_i; \gamma) \exp(\beta^T \mathbf{x_i}).$$

where $\lambda(t; \gamma)$ represents the parametrized baseline hazard function, indexed by the vector γ. Under the assumption of independent censoring, the likelihood function is given by

$$
\begin{aligned}
L &= \prod_{i=1}^{n} \{\ [f(t_i)]^{w_i}[S(t_i)]^{1-w_i}\ \} \\
&= \prod_{i} \{\ [\lambda(t_i;\gamma)\exp(\beta^T \mathbf{x}_i)]^{w_i}\exp(-\Lambda(t_i;\gamma))e^{\beta^T \mathbf{x}_i}\ \} \\
&= \prod_{i} \{\ [\mu_i^{w_i} e^{-\mu_i}]\ [\lambda(t_i;\gamma)/\Lambda(t_i;\gamma)]^{w_i}\ \},
\end{aligned}
$$

where $\mu_i = \Lambda(t_i;\gamma)\exp(\beta^T \mathbf{x}_i)$ and $\Lambda(t;\gamma) = \int_0^t \lambda(u;\gamma)du$. Aitkin and Clayton make the important observation that the first term in L has the form of the likelihood function for n independent "Poisson" variates w_i with means μ_i. The log-linear model for the hazard implies a log-linear model for the "Poisson" mean μ_i. i.e.. $\log\mu_i = \log\Lambda(t_i;\gamma) + \beta^T \mathbf{x}_i$. Since the second term of L does not depend on β. a simple iterative procedure to estimate $(\beta.\gamma)^T$ is to begin with an initial estimate of γ. Using GLIM (or any software that can handle GLMs). we fit the Poisson log-linear model with fixed offset. $\log\Lambda(t;\gamma)$. The estimates of β are then substituted into the likelihood equations for γ. and solved to yield an updated estimate for γ. These steps are alternated until convergence criteria are met. Compared to a direct maximization of L through simultaneous solution of the likelihood equations. the present approach is more convenient to implement since it utilizes GLM software to estimate the regression coefficients. In some common models. $\hat{\gamma}$ is also available in closed-form. given $\hat{\beta}$.

The matrix $\{\frac{\partial^2 \ell}{\partial\beta\partial\beta^T}\}$ is easily obtained from the GLM fitting procedure, while $\{\frac{\partial^2 \ell}{\partial\beta\partial\gamma}\}$ and $\frac{\partial^2 \ell}{\partial\gamma^2}$ have to be calculated separately. These may be combined to yield the observed information. In general, the variances of $\hat{\beta}$ will be underestimated by the output of any GLM routine since γ was assumed known.

4.2.2 Likelihood-based Interval Estimation

In this section, we explore a new approach to obtain profile likelihood-based interval estimates for functionals in common parametric failure time models. We shall focus on the case where, under the constraint imposed by the functional, the model for the log "Poisson" means retains its log-linear form. The estimation procedure for these models then follows directly from Aitkin and Clayton's approach. A few useful functionals fall under this category, e.g., $\beta^T(\mathbf{x}_1 - \mathbf{x}_2)$, $\Lambda(t_0; \mathbf{x}_1)/\Lambda(t_0; \mathbf{x}_2)$, where t_0 denotes a fixed (known) time point, and $\mathbf{x}_1$, $\mathbf{x}_2$ are two distinct (known) covariate values. When the constrained model is nonlinear in the log "Poisson" means, we also indicate briefly a possible solution to the problem via the simple technique of linearization.

We consider firstly the case where the log-linear model applies under the constraint. For illustration, suppose the functional of interest is $S(t_0; \mathbf{x}_0) = \exp\{-\Lambda(t_0; \gamma)\, e^{\beta^T \mathbf{x}_0}\}$, where $\mathbf{x}_0$ denotes a fixed and known covariate vector, and t_0 is a fixed time. Compared to a full likelihood approach, our method is fairly easy to implement for this functional and the survival distributions in question, particularly since the constrained estimator for γ continues to admit a closed-form solution and the constrained estimates for β are also obtained by fitting a log-linear model for the "Poisson" means. Assume a fixed value of $S(t_0; \mathbf{x}_0)$, say $S(t_0; \mathbf{x}_0) = \omega$, $\omega \in [\, 0, 1\,]$. Provided $x_{0p} \neq 0$, we can solve for β_p as

$$\beta_p = \frac{1}{x_{0p}} \,[\, \log\{-\log(\omega)\} - \log \Lambda(t_0; \gamma) - \sum_{j=1}^{p-1} \beta_j x_{0j} \,]\, .$$

Hence, under $H_0 : S(t_0; \mathbf{x_0}) = \omega$, the likelihood function is

$$L = \prod_i [\ (\mu_i')^{w_i} e^{-\mu_i'}\ ][\ \lambda(t_i;\gamma)/\Lambda(t_i;\gamma)\ ]^{w_i},$$

where

$$\mu_i' = \Lambda(t_i;\gamma) \exp\left[\frac{x_{ip}}{x_{0p}} \log\left\{ \frac{-\log\omega}{\Lambda(t_0;\gamma)} \right\} + \sum_{j=1}^{p-1} \beta_j z_{ij} \right].$$

$$z_{ij} = x_{ij} - x_{ip}\frac{x_{0j}}{x_{0p}},\quad j \neq p.$$

Given ω and an initial estimate, $\gamma^{(0)}$, of γ, we fit the log-linear model,

$$\log\mu_i' = \log\Lambda(t_i;\gamma^{(0)}) + \frac{x_{ip}}{x_{0p}} \log\left\{ \frac{-\log(\omega)}{\Lambda(t_0;\gamma^{(0)})} \right\} + \sum_{j=1}^{p-1} \beta_j z_{ij}.$$

The updated value, $\mu^{(1)}$, of μ' is used in the score equations for γ to obtain $\gamma^{(1)}$. These steps are iterated till convergence, and the LRS is compared to the relevant quantile point of the chi-square distribution with one degree of freedom. The value of ω is then increased or decreased, accordingly, with the next cycle. The interval estimation of quantiles may also be of interest. A simple but computationally intensive solution for interval estimation of Q_p, the pth quantile of the failure time distribution, consists of finding the set of t_0 such that the $100(1-\alpha)\%$ CI for $S(t_0, \mathbf{x_0})$ covers $(1-p)$.

Consider the following simple cases discussed in Aitkin and Clayton (1980).

Exponential Distribution

We assume $\Lambda(t;\gamma) = \Lambda(t) = t$, so that $f(t) = \exp(\beta^T\mathbf{x} - te^{\beta^T\mathbf{x}})$. Since $\lambda(t)/\Lambda(t) = t^{-1}$ is independent of unknown parameters, the iterative scheme just involves the first part of L. We only need to fit the log-linear model

$$\log \mu_i' = \log t_i + \frac{x_{ip}}{x_{0p}} \log \left\{ \frac{-\log(\omega)}{t_0} \right\} + \sum_{j=1}^{p-1} \beta_j z_{ij} \, .$$

Weibull and Extreme Value Distributions

We assume $\Lambda(t;\gamma) = t^\gamma$, which implies that $\lambda(t;\gamma)/\Lambda(t;\gamma) = \gamma/t$. Given an initial estimate, $\gamma^{(0)}$, we fit the model

$$\log \mu_i' = \gamma^{(0)} \log t_i + \frac{x_{ip}}{x_{0p}} \{\log(-\log \omega) - \gamma^{(0)} \log t_0\} + \sum_{j=1}^{p-1} \beta_j z_{ij} \, .$$

The log likelihood function is

$$\ell = \log \gamma (\sum_i w_i) + \sum_i (w_i \log \mu_i' - \mu_i') \, ,$$

ignoring constant terms. The score equation for γ,

$$\frac{\partial \ell}{\partial \gamma} = \frac{\sum_i w_i}{\gamma} + \sum_i (w_i - \mu_i') \left(\log t_i - \frac{x_{ip}}{x_{0p}} \log t_0 \right) = 0$$

can be solved to yield an updated estimate for γ as

$$\tilde{\gamma} = \frac{\sum_i w_i}{\sum_i (\tilde{\mu}_i' - w_i)(\log t_i - \frac{x_{ip}}{x_{0p}} \log t_0)} \, .$$

As noted by Aitkin and Clayton (1980), by transforming t to $\exp(t)$, we obtain the Weibull density. Hence the preceding iterative procedure can be used to obtain an interval estimate for $S(t_0; \mathbf{x}_0)$ in the case of the extreme value distribution. We only need to substitute t for $\log t$.

Generalized Extreme Value Distribution

The Weibull and extreme value distributions are members of a more general family of distributions, viz., the generalized extreme value distribution with density function

$$f(t) = \alpha\delta t^{\delta-1}e^{\alpha t^{\delta}}\exp(-e^{\alpha t^{\delta}}), \qquad -\infty < t < \infty, \alpha > 0\,.$$

For this distribution, $\Lambda(t;\gamma) = \exp(\alpha t^{\delta})$ and $\lambda(t;\gamma)/\Lambda(t;\gamma) = \alpha\delta t^{\delta-1}$. Omitting constant terms, the log likelihood function is

$$\ell = d\log\alpha + d\log\delta + (\delta-1)\sum_i w_i\log t_i + \sum_i(w_i\log\mu_i^{'} - \mu_i^{'})\,,$$

where $d = \sum_i w_i$. Given an initial estimate for $\gamma = (\alpha,\delta)^T$, we fit the model

$$\log\mu_i^{'} = \alpha t_i^{\delta} + \frac{x_{ip}}{x_{0p}}\{\log(-\log\omega) - \alpha {t_0}^{\delta}\} + \sum_{j=1}^{p-1}\beta_j z_{ij}\,.$$

The likelihood equations for $(\alpha,\delta)^T$ in this case are

$$\begin{aligned}\frac{\partial\ell}{\partial\alpha} &= \frac{d}{\alpha} + \sum_i(w_i - \mu_i^{'})\left(t_i^{\delta} - \frac{x_{ip}}{x_{0p}}{t_0}^{\delta}\right) = 0\,.\\ \frac{\partial\ell}{\partial\delta} &= \frac{d}{\delta} + \sum_i w_i\log t_i + \alpha\sum_i(w_i - \mu_i^{'})\left(t_i^{\delta}\log t_i - \frac{x_{ip}}{x_{0p}}{t_0}^{\delta}\log t_0\right) = 0\,.\end{aligned}$$

where the constrained MLEs of α and δ satisfy

$$\begin{aligned}\tilde{\alpha} &= d\left[\sum_i(\tilde{\mu}_i^{'} - w_i)\left(t_i^{\tilde{\delta}} - \frac{x_{ip}}{x_{0p}}{t_0}^{\tilde{\delta}}\right)\right]^{-1},\\ \tilde{\delta} &= d\left[\tilde{\alpha}\sum_i(\tilde{\mu}_i^{'} - w_i)\left(t_i^{\tilde{\delta}}\log t_i - \frac{x_{ip}}{x_{0p}}{t_0}^{\tilde{\delta}}\log t_0\right) - \sum_i w_i\log t_i\right]^{-1}.\end{aligned}$$

Given starting estimates for $(\alpha,\delta)^T$, the likelihood equations can be used in the same way as in Aitkin and Clayton (1980) to update the estimates.

Extension to the Piecewise Exponential Cox Model

In this subsection, we consider extending the previous methods to a version of the Cox PH model which is popular in epidemiology. This model is given by

$$\lambda_i(t) = \lambda_j \exp(\beta^T \mathbf{x_i}), \qquad t \in (a_{j-1}, a_j], \quad j = 1, ..., M,$$

where λ_j are (unknown) constants and we partition the observation interval into M subintervals. This implies a constant hazard within each subinterval.

Aitkin *et al.* (1989) also considered this model and showed that two simple approaches could be employed to fit this model. The first method they describe follows that of Aitkin and Clayton (1980) and is practicable provided the number of observations is not large, i.e., less than several hundred. The other method outlined by Aitkin *et al.* is similar and is based on recognizing that the likelihood equation for β is identical to that for a related Poisson distribution. We shall focus on the first of these two methods and show that likelihood-based interval estimates for $S(t_0; \mathbf{x_0})$ can be derived using the approach proposed earlier in this section. This will be useful since, for the methods described in Aitkin *et. al* (1989), standard errors for $\lambda^T = (\lambda_1, ..., \lambda_M)$ are not available from their fitting procedure. Hence other methods are required to obtain the standard errors of both β and λ. As seen in the previous chapters, the profile likelihood-based method has the advantage that explicit standard errors for both β and λ are not required to compute interval estimates for functions of these parameters.

First we describe the setup leading to the likelihood function. The *i*th subject experiences a sequence of "censorings" at $a_1, a_2, ...$, and either genuine censoring or death at t_i, where $a_{N_i-1} < t_i < a_{N_i}$. Let h_{ij} denote the hazard function for the *i*th subject in the interval $I_j = (a_{j-1}, a_j]$, with

$$h_{ij} = \lambda_j \exp(\beta^T \mathbf{x}_i).$$

Let w_{ij} and e_{ij} denote the censoring indicator and exposure time for subject i in I_j, respectively. Then

$$S_{ij}(t) = \exp(-h_{ij} t)$$

is the conditional survivor function for subject i in I_j, and the contribution of this subject to the likelihood is

$$L_i = \prod_{j=1}^{N_i} h_{ij}^{w_{ij}} S_{ij}(e_{ij}).$$

The complete likelihood function is

$$\begin{aligned} L &= \prod_{i=1}^{n} L_i \\ &= \prod_i \prod_j \theta_{ij}^{w_{ij}} \exp(-\theta_{ij}) / \prod_i \prod_j e_{ij}^{w_{ij}} \end{aligned}$$

where $\theta_{ij} = e_{ij} \lambda_j \exp(\beta^T \mathbf{x}_i)$. This gives us the familiar "Poisson" representation with a log-linear model for the mean,

$$\log \theta_{ij} = \log e_{ij} + \log \lambda_j + \beta^T \mathbf{x}_i.$$

It is not difficult to show that, given $\hat{\beta}$, $\hat{\lambda}_j$ satisfies

$$\hat{\lambda}_j = \frac{\sum_{i \in R_j} w_{ij}}{\sum_{i \in R_j} e_{ij} \exp(\hat{\beta}^T \mathbf{x}_i)}, \quad j = 1, \ldots, M,$$

where R_j is the risk set in I_j.

Now we may consider the interval estimation of $S(t_0; \mathbf{x_0})$ using the first approach, i.e., adapting Aitkin and Clayton's (1980) method to the Cox model. Let $t_0 \in (a_{k-1}, a_k]$ be a fixed time point. We also define $e_{0j} = a_j - a_{j-1}$ for $j = 1, \ldots, k-1$, and $e_{0k} = t_0 - a_{k-1}$. Then, $\log S(t_0; \mathbf{x_0})$ can be written as $-(\sum_j \lambda_j e_{0j} + \lambda_k e_{0k}) \exp(\beta^T \mathbf{x_0})$ and hence it would be easier to work with this functional instead of $S(t_0; \mathbf{x_0})$. Setting $\log S(t_0; \mathbf{x_0}) = \omega \in [0, 1]$, we solve for β_p as

$$\beta_p = x_{0p}^{-1} \left\{ \log \left(\frac{-\omega}{\sum_j \lambda_j e_{0j} + \lambda_k e_{0k}} \right) - \sum_{j=1}^{p-1} \beta_j x_{0j} \right\}.$$

This implies a log-linear model for θ_{ij},

$$\log \theta'_{ij} = \log e_{ij} + \log \lambda_j + \frac{x_{ip}}{x_{0p}} \left\{ \log \left(\frac{-\omega}{\sum_j \lambda_j e_{0j} + \lambda_k e_{0k}} \right) \right\} + \sum_{j=1}^{p-1} \beta_j v_{ij},$$

where $v_{ij} = x_{ij} - x_{ip} \frac{x_{0j}}{x_{0p}}$, $j = 1, \ldots, p-1$. Given initial values for λ_j, this model can be fit in the usual manner. To update the λ_j, we use the following likelihood equations. For $\{j : a_j > t_0\}$, the likelihood equations are the same as those for the unrestricted model, yielding the constrained MLEs

$$\tilde{\lambda}_j = \frac{\sum_{i \in R_j} w_{ij}}{\sum_{i \in R_j} e_{ij} \exp(\tilde{\beta}^T \mathbf{x}_i)}.$$

For $\{j : a_j \leq t_0\}$,

$$\frac{\partial \ell}{\partial \lambda_j} = \sum_{i \in R_j} (1 - r_{jp})(w_{ij} - \theta'_{ij}),$$

where

$$r_{jp} = \frac{x_{ip} e_{0j} / x_{0p}}{\sum_j \lambda_j e_{0j} + \lambda_k e_{0k}}.$$

By setting $\frac{\partial \ell}{\partial \lambda_j} = 0$, we obtain after some rearrangement,

$$\bar{\lambda}_j = \frac{\sum_i w_{ij}}{\sum_i e_{ij}\bar{\theta}'_{ij} + r_{jp}\{\sum_i x_{ip}w_{ij} - \bar{\lambda}_j \sum_i x_{ip}e_{ij}\bar{\theta}'_{ij}\}},$$

where the summations above are over R_j. This set of equations is in the form of a fixed-point system and the usual numerical methods may be applied. However, convergence to the constrained MLE is expected to depend on suitable choice of starting values for λ_j. As noted previously, this approach is impractical when the number of observations is large. This method is most likely to be useful in small to moderate sample size situations, since the standard assumption of normality may not be appropriate then.

Nonlinear Models

In this brief remark, we indicate how Aitkin and Clayton's approach can be adapted when the constrained model for the "Poisson" means is nonlinear in the regression coefficients. This is especially useful for obtaining profile likelihood-based interval estimates for other functionals of interest in the failure time context that cannot be obtained using the methods described previously. Although there are many ways to handle nonlinearity, we shall only consider the simple technique of linearization here. Aside from the simplicity of this technique, it also fits nicely into the previous GLM framework.

Under the constraint of interest, the model for the "Poisson" means can be written as

$$\log \mu_i' = g(\mathbf{x}; \theta) + \sum_{j=1}^{p-1} \beta_j x_{ij},$$

where g is a nonlinear function of $\theta = (\beta_1, ..., \beta_{p-1})^T$. For a suitable choice of initial

Table 4.1: Interval estimates for survival probabilities.

t_0	6-MP	MLE (6-MP)	Control	MLE (Control)
10	(0.553, 0.869)	0.75	(0.142, 0.422)	0.265
16	(0.481, 0.846)	0.70	(0.072, 0.345)	0.180
23	(0.417, 0.824)	0.64	(0.031, 0.283)	0.120

values. θ_0, the approximation

$$g(\mathbf{x};\theta) \approx g(\mathbf{x};\theta_0) + (\theta-\theta_0)^T \frac{\partial g}{\partial \theta}|_{\theta=\theta_0}$$

is appropriate. Using this approximation, the constrained model becomes

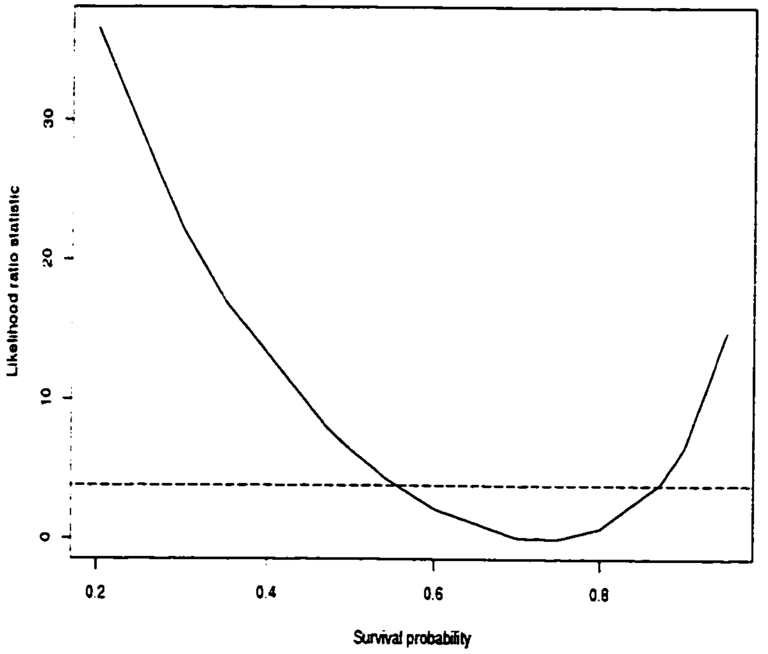
$$\log \mu_i' = g(\mathbf{x};\theta_0) - \theta_0^T \frac{\partial g}{\partial \theta}|_{\theta=\theta_0} + \sum_{j=1}^{p-1} \beta_j(x_{ij} + \{\frac{\partial g}{\partial \theta}|_{\theta=\theta_0}\}_j),$$

where $\{\frac{\partial g}{\partial \theta}|_{\theta=\theta_0}\}_j$ denotes the partial derivative of g with respect to the j component of θ, evaluated at θ_0. Given a suitable θ_0, this model can be fit in the usual way. The multiple functional case can be similiarly handled.

A Numerical Example

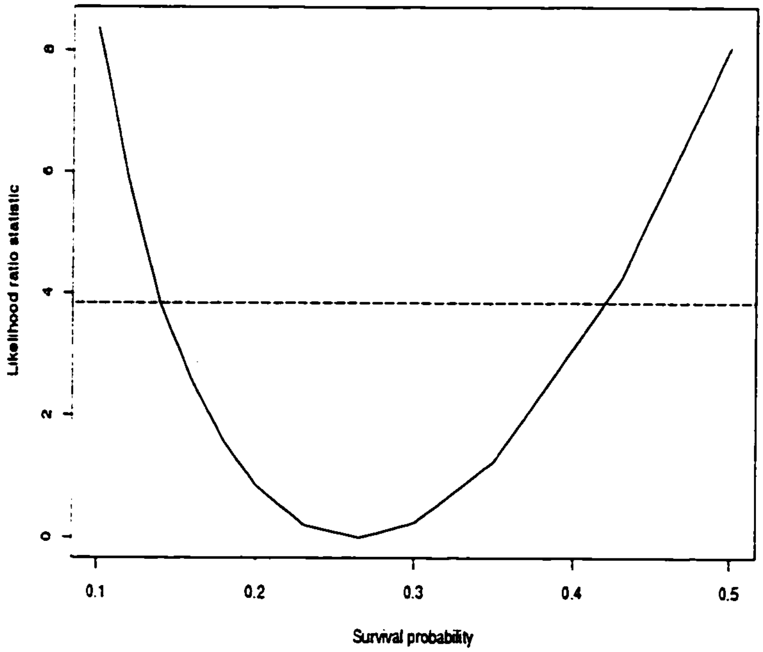
For a simple application of the methods discussed so far, consider the two-sample data on remission times of leukaemia patients (Gehan, 1965; see Table 3.1). Approximate 95% CIs for $S(t_0; x_0)$ are displayed in Table 4.1, for a selection of t_0 values. The calculations assume that remission times in each group of patients follow a Weibull distribution. The proposed procedure shows up the asymmetry in the interval estimate. Figures 4.1 and 4.2 show the likelihood ratio statistics for the respective survivor functions for the two groups at 10 weeks. If we assume an extreme value distribution for the remission times. we obtain for the 6-MP group at 10 weeks an interval estimate of (0.622, 0.906); the corresponding estimate for the control group is (0.195, 0.489). These estimates are quite similiar to the ones obtained from the Weibull model.

Figure 4.1: Likelihood ratio statistic for interval estimation of the survival probability at 10 weeks in the 6-MP group of leukaemia patients (see Table 3.1). The height of the horizontal dashed line is 3.841 units.



Next we consider a simple application of the linearization technique for the Weibull model. Suppose one is also interested in the ratio, $\frac{S(t_0;\mathbf{x}_1)}{S(t_0;\mathbf{x}_2)}$, where t_0 is a

Figure 4.2: Likelihood ratio statistic for interval estimation of the survival probability at 10 weeks in the control group of leukaemia patients (see Table 3.1). The height of the horizontal dashed line is 3.841 units.



fixed time. $\mathbf{x}_1 = (1, 1)^T$ and $\mathbf{x}_2 = (1, 0)^T$. This is a measure of the relative rates of survival for the treatment and control groups. For analytical simplicity, let us consider obtaining an interval estimate for $\log\left\{\frac{S(t_0;\mathbf{x}_1)}{S(t_0;\mathbf{x}_2)}\right\}$ first, and then invert the end-points of the interval estimate to obtain the corresponding values for $\frac{S(t_0;\mathbf{x}_1)}{S(t_0;\mathbf{x}_2)}$. By setting $\log\left\{\frac{S(t_0;\mathbf{x}_1)}{S(t_0;\mathbf{x}_2)}\right\} = \omega$, $\omega \geq 0$, we obtain

$$\beta_0 = \log\left\{\frac{\omega}{\Lambda_0(t_0;\gamma)(1 - e^{\beta_1})}\right\}.$$

Table 4.2: Interval estimates for the ratio of survival probabilities.

t_0	95% CI	MLE
10	(1.811, 5.795)	3.034
16	(1.786, 9.865)	3.781
23	(2.241, 22.176)	5.755

This implies the following log-linear model for the "Poisson" mean.

$$\log \mu_i = \gamma \log\left(\frac{t_i}{t_0}\right) + \log \omega - \log[1 - \exp(\beta_1)] + \beta_1 x_{1i}.$$

The linearization technique gives us the approximation.

$$\begin{aligned}\log \mu_i &= \gamma \log\left(\frac{t_i}{t_0}\right) + \log \omega - \log[1 - \exp(\beta_1^{(0)})] \\ &+ \beta_1^{(0)} \frac{\exp(\beta_1^{(0)})}{\exp(\beta_1^{(0)}) - 1} + \beta_1 \left\{ x_{1i} - \frac{\exp(\beta_1^{(0)})}{\exp(\beta_1^{(0)}) - 1} \right\}.\end{aligned}$$

where $\beta_1^{(0)}$ is an initial estimate for β_1. Table 4.2 displays the approximate 95% CIs for $\frac{S(t_0:\mathbf{x}_1)}{S(t_0:\mathbf{x}_2)}$ at a sample of t_0 values. This result provides evidence of a higher rate of survival for the treatment group at the given time points. Figure 4.3 shows the likelihood ratio statistic for the log survival probability ratio at 10 weeks. Numerically, the implementation for this example was fairly straightforward in that the routine was stable with respect to the choice of starting value for β_1. For interest, we also obtained interval estimates for the survival ratio from an extreme value model; these are summarized in Table 4.3. The results are similiar to those of Table 4.2.

Table 4.3: Interval estimates for the ratio of survival probabilities (extreme value model).

t_0	95% CI	MLE
10	(1.791, 4.495)	2.678
16	(1.701, 8.440)	3.579
23	(2.298, 34.261)	6.753

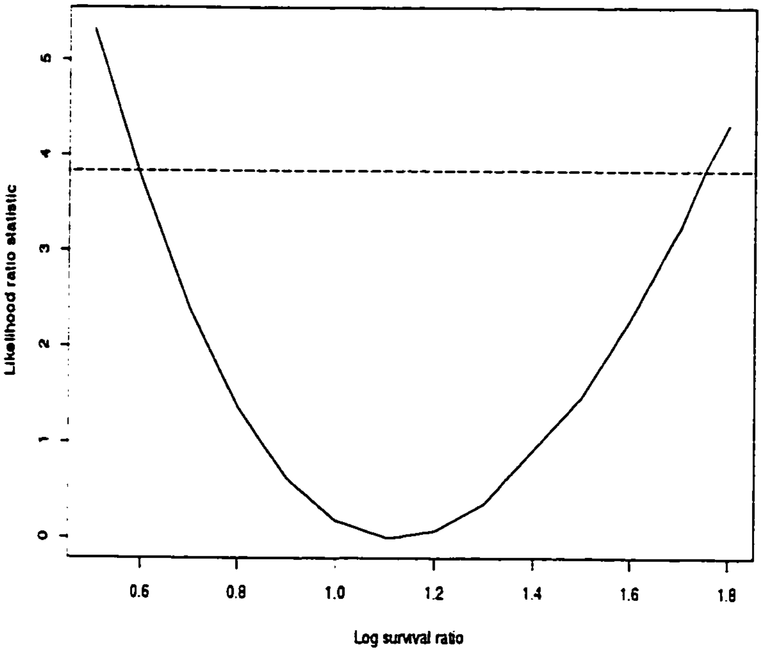
Extensions to more difficult contexts, such as multivariate GLMs, is an area for future research.

4.3 An IRLS Approach for Parametric Models

4.3.1 Introduction

An alternative approach for analyzing parametric survival models is based on the class of univariate location-scale models. Lawless (1982) provides a comprehensive summary of statistical methods for such models, based on data that are right-censored. Many standard statistical software routines implement maximum likelihood estimation for the parameters in location-scale models. The asymptotic covariance matrix of the MLEs is routinely used to generate standard errors and confidence intervals for parameters. Recently, Therneau (1995) proposed a novel way to estimate the parameters of the survival model, using the iteratively reweighted least squares (IRLS) method. His approach is implemented via the survreg() function in S-Plus, and provides a convenient way to analyze the effect of covariates on response time in common parametric survival distributions, while taking into

Figure 4.3: Likelihood ratio statistic for interval estimation of the log survival probability ratio, 6-MP versus control at 10 weeks, based on the remission times in Table 3.1. The height of the horizontal dashed line is 3.841 units.



account the effects of various types of censoring. A detailed and far-ranging discussion of the scope of IRLS, covering applications to generalized linear models. quasi-likelihood. robust regression. etc., was presented by Green (1984).

In this section, we show how the IRLS approach can be adapted to supply likelihood-based intervals for important functionals in location-scale lifetime models when responses are subject to right-, left- and interval censoring. For right-censored failure times, Lawless (1982) obtained likelihood-based interval estimates for functionals such as quantiles, using primarily the Newton-Raphson algorithm when the ML equations do not admit closed-form solutions. In general, Green

(1984) indicated that the IRLS method tends to be more stable numerically. While the survreg() function also routinely supplies standard errors of estimates based on the normality assumption, these might not always be appropriate, especially when samples are small to moderate size or when censoring is heavy. The techniques we propose are simple to implement since they utilize existing routines in S-Plus; however, other packages can potentially be used as well. We begin by briefly outlining Therneau's adaptation of the IRLS approach to the parametric failure time setting. Subsequently, we indicate how the IRLS approach can be modified to yield profile likelihood-based CIs for several important functionals in the failure time context.

Let the data be represented by $(t_i^l, t_i^u, \mathbf{x_i})$, $i = 1, \ldots, n$, where $t_i^l \leq t_i^u$ are times and $\mathbf{x_i}$ is the covariate vector for the ith subject. As usual, define $\eta_i = \beta^T \mathbf{x_i}$. For a failure time, $t_i^l = t_i^u = t_i$. For right-censored subjects, $t_i^l = t_i$ and $t_i^u = \infty$. For left-censored subjects, $t_i^l = 0$ and $t_i^u = t_i$. All other choices of t_i^l and t_i^u correspond to interval-censored observations. We assume that

$$z_i \equiv \frac{a(t_i) - \beta^T \mathbf{x_i}}{\sigma} \sim f$$

for some distribution f and differentiable function a. Define $z_i^l = [a(t_i^l) - \beta^T \mathbf{x_i}]\sigma^{-1}$ and $z_i^u = [a(t_i^u) - \beta^T \mathbf{x_i}]\sigma^{-1}$. The likelihood function arising from the observed data is

$$L = \prod_{\mathcal{D}} f(z_i)/\sigma \prod_{\mathcal{R}} \int_{z_i^l}^{\infty} f(s)ds \prod_{\mathcal{L}} \int_{-\infty}^{z_i^u} f(s)ds \prod_{\mathcal{I}} \int_{z_i^l}^{z_i^u} f(s)ds\,,$$

where the sets $\mathcal{D}$, $\mathcal{R}$, $\mathcal{L}$ and $\mathcal{I}$ represent uncensored, right-censored, left-censored and interval-censored observations, respectively. The log likelihood is

$$\ell = \sum_{\mathcal{D}} [g_1(z_i) - \log \sigma] + \sum_{\mathcal{R}} g_2(z_i^l) + \sum_{\mathcal{L}} g_3(z_i^u) + \sum_{\mathcal{I}} g_4(z_i^l, z_i^u)\,.$$

where the g's are the logarithms of the individual terms (except for σ^{-1}) in L.

It is easy to verify that

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \left\{ x_{ij} \frac{\partial g}{\partial \eta_i} \right\} = (X^T U)_j \, , \tag{4.1}$$

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^{n} \left\{ x_{ij} x_{ik} \frac{\partial^2 g}{\partial \eta_i^2} \right\} = -(X^T D X)_{jk} \, , \tag{4.2}$$

where X denotes the $n \times p$ matrix of regressors. $U = \{\frac{\partial g}{\partial \eta_i}\}$, and $D = \{-\frac{\partial^2 g}{\partial \eta_i^2}\}$. (Note that the observations are independent, so $\frac{\partial^2 g}{\partial \eta_i \partial \eta_j} = 0$ for $i \neq j$: this means D should be diagonal). For a fixed σ, numerically maximizing ℓ with respect to β, via the Newton-Raphson algorithm, is equivalent to an IRLS regression with weights D and adjusted dependent variate $\eta + D^{-1}U$, where $\eta = X\beta$. This was shown by Therneau as follows. The Newton-Raphson iterative scheme is given by

$$\beta^{(m+1)} = \beta^{(m)} + (X^T D^{(m)} X)^{-1} X^T U^{(m)} \, ,$$

or equivalently,

$$(X^T D^{(m)} X) \Delta^{(m+1)} = X^T U^{(m)}, \tag{4.3}$$

where $D^{(m)}$ and $U^{(m)}$ are the matrices D and U evaluated at the mth cycle of the algorithm, and $\Delta^{(m)} = \beta^{(m+1)} - \beta^{(m)}$ is the mth cycle update. Adding $X^T D^{(m)} X \beta^{(m)}$ to both sides of equation (4.3), we get

$$(X^T D^{(m)} X) \beta^{(m+1)} = (X^T D^{(m)})(\eta^{(m)} + (D^{(m)})^{-1} U^{(m)})$$

where $\eta^{(m)} = X\beta^{(m)}$. The IRLS scheme iterates between the estimation of β and σ; this is unlike the case of exponential family models in GLMs, where σ is estimated after $\hat{\beta}$ is obtained. The latter follows from the fact that σ drops out of the score equations for β, and can therefore be estimated separately from β.

4.3.2 Likelihood-based Interval Estimation

The standard approach to obtaining likelihood-based interval estimates is to directly maximize the log likelihood function with respect to the parameters under the constraint imposed by the functional. This is usually implemented numerically by the Newton-Raphson algorithm. However, it is not possible to use some standard packages (e.g. SAS) to implement the constrained maximization entailed, so that special programs have to be written for each failure time model and functional of interest.

The methods we describe in this section offer a simple way to obtain these interval estimates using any software that implements the IRLS approach. For some common functionals of interest, a regular profile likelihood approach, coupled with some manipulation of existing computer code, is usually sufficient to yield the desired likelihood-based CIs. However, this may not work for other functionals. In these cases, the Lagrange multiplier technique introduced in chapter 1 may prove useful. We employ the same argument for this parametric failure time setting. For the functionals considered, this technique is attractive in the sense that the IRLS regression model continues to apply under the constraint.

In the following, we consider the regular profile likelihood approach and show how to adapt the survreg() function to the parameter function setting. First, for $p \in [0, 1]$, we define $Q_p = \inf\{t : S(t; \mathbf{x}) \leq 1 - p\}$ to be the pth quantile point for the distribution of T. The corresponding quantile point for f is denoted by q_p. For fixed t_0, let $S(t_0; \mathbf{x_0})$ represent the survivor function for a subject with covariate value $\mathbf{x_0}$. For illustration, we will obtain a likelihood-based interval estimate for

$S(t_0; \mathbf{x_0})$. We set $S(t_0; \mathbf{x_0}) = \omega$, where $\omega \in [0, 1]$. For many practical cases, we can also assume that $a(\cdot)$ is a monotonic function. In this case, since $P(T > t_0; \mathbf{x_0}) = P(Z > \frac{a(t_0) - \beta^T \mathbf{x_0}}{\sigma})$, we obtain

$$\frac{a(t_0) - \beta^T \mathbf{x_0}}{\sigma} = q_{1-\omega} \, .$$

Solving for β_p, we obtain

$$\beta_p = \frac{a(t_0) - \sigma q_{1-\omega}}{x_{0p}} - \frac{1}{x_{0p}} \sum_{j=1}^{p-1} \beta_j x_{0j} \, ,$$

provided $x_{0p} \neq 0$. For a fixed value of σ and $q_{1-\omega}$, this sets up an IRLS regression with modified weights and adjusted dependent variates. To see this, we substitute the expression for β_p into the log likelihood, obtaining the constrained log likelihood

$$\ell_c = \sum_{\mathcal{D}} g_1(z_{*i}) - \log \sigma + \sum_{\mathcal{R}} g_2(z^l_{*i}) + \sum_{\mathcal{C}} g_3(z^u_{*i}) + \sum_{\mathcal{I}} g_4(z^l_{*i}, z^u_{*i}) \, .$$

where

$$z_{*i} = \frac{a(t_i) - \eta_{*i}}{\sigma} \, .$$

$$\eta_{*i} = \frac{x_{ip}}{x_{0p}} \{a(t_0) - \sigma q_{1-\omega}\} + \sum_{j=1}^{p-1} \beta_j v_{ij} \, ,$$

$$v_{ij} = x_{ij} - \frac{x_{ip}}{x_{0p}} x_{0j} \, .$$

In the same way, (z^l_{*i}, z^u_{*i}) now represent the end-points of the censoring interval for the ith "incomplete" observation. For notational simplicity, we will suppress the subscript i in most of the subsequent development. It follows that

$$\begin{aligned} \frac{\partial \ell_c}{\partial \beta_j} &= (V^T U_*)_j \\ \frac{\partial^2 \ell_c}{\partial \beta_j \partial \beta_k} &= -(V^T D_* V)_{jk} \end{aligned}$$

where $V = \{v_{ij}\}$, $j = 1, \ldots, p-1$, $U_* = \{\frac{\partial g_*}{\partial \eta_{*i}}\}$, $D_* = \{-\frac{\partial^2 g_*}{\partial \eta_{*i} \partial \eta_{*j}}\}$, and g_{*j}, $j = 1, \ldots 4$, are the respective g_j's with z_* as arguments. Hence, maximizing ℓ_c with respect to β is equivalent to an IRLS regression with weights D_* and adjusted dependent variate, $\eta_* + D_*^{-1} U_*$.

The elements of U_* and D_* can be obtained routinely by applying the chain rule. However, it is not difficult to see that they are equivalent to the elements of U and D, except for a change of argument from z to z_*, by noting the following. The partial derivatives of the g_j's with respect to z_* and z are essentially identical. The partial derivatives of z with respect to η are identical to those of z_* with respect to η_*. For the constrained model, we therefore have

$$\begin{aligned} \frac{\partial g_{*1}}{\partial \eta_*} &= \frac{\partial z_*}{\partial \eta_*} \frac{\partial g_{*1}}{\partial z_*} \\ &= -\sigma^{-1} \frac{f'(z_*)}{f(z_*)} \\ \frac{\partial g_{*4}}{\partial \eta_*} &= \frac{\partial z_*^u}{\partial \eta_*} \frac{\partial g_{*4}}{\partial z_*^u} + \frac{\partial z_*^l}{\partial \eta_*} \frac{\partial g_{*4}}{\partial z_*^l} \\ &= -\sigma^{-1} \left[\frac{f(z_*^u) - f(z_*^l)}{F(z_*^u) - F(z_*^l)} \right] \\ \frac{\partial^2 g_{*1}}{\partial \eta_*^2} &= \left(\frac{\partial z_*}{\partial \eta_*} \right)^2 \frac{\partial^2 g_{*1}}{\partial z_*^2} + \frac{\partial g_{*1}}{\partial z_*} \frac{\partial^2 z_*}{\partial \eta_*^2} \\ &= \sigma^{-2} \frac{f''(z_*)}{f(z_*)} - \left(\frac{\partial g_{*1}}{\partial \eta_*} \right)^2 \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 g_{*4}}{\partial \eta_*^2} &= \left(\frac{\partial z_*^u}{\partial \eta_*}\right)^2 \frac{\partial^2 g_{*4}}{\partial (z_*^u)^2} + \frac{\partial g_{*4}}{\partial z_*^u}\frac{\partial^2 z_*^u}{\partial \eta_*^2} + \left(\frac{\partial z_*^l}{\partial \eta_*}\right)^2 \frac{\partial^2 g_{*4}}{\partial (z_*^l)^2} + \frac{\partial g_{*4}}{\partial z_*^l}\frac{\partial^2 z_*^l}{\partial \eta_*^2} \\
&= \sigma^{-2}\left[\frac{f'(z_*^u) - f'(z_*^l)}{F(z_*^u) - F(z_*^l)}\right] - \left(\frac{\partial g_{*4}}{\partial \eta_*}\right)^2 .
\end{aligned}$$

Given these formulae and a fixed initial value for σ, an IRLS regression can be performed to supply the estimates of β. Since β and σ have to be estimated jointly (unlike the solution involving GLMs; see section 4.2), we also need to keep track of the changes to the derivatives involving σ as follows.

We can relieve the burden of the computations by taking note of the following. For any given term, the chain of partial derivatives employed correspond to their unrestricted counterparts, e.g., in place of $\frac{\partial z}{\partial \sigma}$ we have $\frac{\partial z_*}{\partial \sigma}$. The partial derivatives of the g_j's with respect to z_* and z are identical, except for a change in the argument. The partial derivatives of z with respect to η are identical to those of z_* with respect to η_*. Wherever it is clear to do so, we make use of these facts, together with the form of the partial derivatives under the unrestricted model, to obtain the corresponding terms under the constrained model. Let $\delta_i = z_{*i} - q_{1-\omega}\frac{x_{ip}}{x_{0p}}$, $\delta_i^u = z_{*i}^u - q_{1-\omega}\frac{x_{ip}}{x_{0p}}$ and $\delta_i^l = z_{*i}^l - q_{1-\omega}\frac{x_{ip}}{x_{0p}}$. In the following, we suppress the subscript i for notational simplicity. We obtain

$$\begin{aligned}
\frac{\partial g_{*1}}{\partial \log \sigma} &= \sigma \frac{\partial g_{*1}}{\partial \sigma} \\
&= \sigma \frac{\partial z_*}{\partial \sigma}\frac{\partial g_{*1}}{\partial z_*} \\
&= -\delta \frac{f'(z_*)}{f(z_*)} \\
\frac{\partial g_{*4}}{\partial \log \sigma} &= \sigma \frac{\partial g_{*4}}{\partial \sigma} \\
&= \sigma \left[\frac{\partial z_*^u}{\partial \sigma}\frac{\partial g_{*4}}{\partial z_*^u} + \frac{\partial z_*^l}{\partial \sigma}\frac{\partial g_{*4}}{\partial z_*^l}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sigma\left[\, -\sigma^{-1}\delta^{u}\frac{\partial g_{*4}}{\partial z_{*}^{u}} - \sigma^{-1}\delta^{l}\frac{\partial g_{*4}}{\partial z_{*}^{l}} \,\right] \\
&= -\frac{\delta^{u} f(z_{*}^{u}) - \delta^{l} f(z_{*}^{l})}{F(z_{*}^{u}) - F(z_{*}^{l})} \\
\frac{\partial^2 g_{*1}}{\partial(\log\sigma)^2} &= \sigma^2\frac{\partial^2 g_{*1}}{\partial\sigma^2} + \frac{\partial g_{*1}}{\partial\log\sigma} \\
&= \sigma^2\left[\, \left(\frac{\partial z_{*}}{\partial\sigma}\right)^2 \frac{\partial^2 g_{*1}}{\partial z_{*}^2} + \frac{\partial g_{*1}}{\partial z_{*}}\frac{\partial^2 z_{*}}{\partial\sigma^2} \,\right] + \frac{\partial g_{*1}}{\partial\log\sigma} \\
&= \sigma^2\left[\, \sigma^{-2}\delta^2\left\{\frac{f''(z_{*})}{f(z_{*})} - \left(\frac{f'(z_{*})}{f(z_{*})}\right)^2\right\} + 2\delta\sigma^{-2}\frac{f'(z_{*})}{f(z_{*})} \,\right] + \frac{\partial g_{*1}}{\partial\log\sigma} \\
&= \frac{\delta^2 f''(z_{*}) + \delta f'(z_{*})}{f(z_{*})} - \left(\frac{\partial g_{*1}}{\partial\log\sigma}\right)^2 \\
\frac{\partial^2 g_{*4}}{\partial(\log\sigma)^2} &= \sigma^2\frac{\partial^2 g_{*4}}{\partial\sigma^2} + \frac{\partial g_{*4}}{\partial\log\sigma} \\
&= \sigma^2\left[\, \left(\frac{\partial z_{*}^{u}}{\partial\sigma}\right)^2 \frac{\partial^2 g_{*4}}{\partial(z_{*}^{u})^2} + \frac{\partial z_{*}^{u}}{\partial\sigma}\frac{\partial z_{*}^{l}}{\partial\sigma}\frac{\partial^2 g_{*4}}{\partial z_{*}^{u}\partial z_{*}^{l}} + \frac{\partial g_{*4}}{\partial z_{*}^{u}}\frac{\partial^2 z_{*}^{u}}{\partial\sigma^2} + \cdots \,\right] + \frac{\partial g_{*4}}{\partial\log\sigma} \\
&= \sigma^2[\ \sigma^{-2}(\delta^{u})^2\left\{\frac{f'(z_{*}^{u})}{F(z_{*}^{u}) - F(z_{*}^{l})} - \left(\frac{f(z_{*}^{u})}{F(z_{*}^{u}) - F(z_{*}^{l})}\right)^2\right\} \\
&+ \sigma^{-2}\delta^{u}\delta^{l}\left\{\frac{f(z_{*}^{u})f(z_{*}^{l})}{F(z_{*}^{u}) - F(z_{*}^{l})}\right\}^2 + 2\delta^{u}\sigma^{-2}\frac{f(z_{*}^{u})}{F(z_{*}^{u}) - F(z_{*}^{l})} + \cdots] + \frac{\partial g_{*4}}{\partial\log\sigma} \\
&= \frac{(\delta^{u})^2 f'(z_{*}^{u}) - (\delta^{l})^2 f'(z_{*}^{l}) + \delta^{u} f(z_{*}^{u}) - \delta^{l} f(z_{*}^{l})}{F(z_{*}^{u}) - F(z_{*}^{l})} - \left(\frac{\partial g_{*4}}{\partial\log\sigma}\right)^2 \\
\frac{\partial^2 g_{*1}}{\partial\log\sigma\,\partial\eta_{*}} &= \sigma\frac{\partial^2 g_{*1}}{\partial\sigma\,\partial\eta_{*}} \\
&= \sigma\left[\, \frac{\partial z_{*}}{\partial\sigma}\frac{\partial z_{*}}{\partial\eta_{*}}\frac{\partial^2 g_{*1}}{\partial(z_{*})^2} + \frac{\partial g_{*1}}{\partial z_{*}}\frac{\partial^2 z_{*}}{\partial\sigma\,\partial\eta_{*}} \,\right] \\
&= \sigma\left[\, -\sigma^{-1}\delta(-\sigma^{-1})\frac{\partial^2 g_{*1}}{\partial(z_{*})^2} + \frac{\partial g_{*1}}{\partial z_{*}}(\sigma^{-2}) \,\right] \\
&= \delta\sigma^{-1}\frac{f''(z_{*})}{f(z_{*})} - \frac{\partial g_{*1}}{\partial\eta_{*}}\left(1 + \frac{\partial g_{*1}}{\partial\log\sigma}\right) \\
\frac{\partial^2 g_{*4}}{\partial\eta_{*}\partial\log\sigma} &= \sigma\frac{\partial^2 g_{*4}}{\partial\eta_{*}\partial\sigma}
\end{aligned}
$$

$$
\begin{aligned}
&= \sigma\frac{\partial z^u_*}{\partial \eta_*}\left[\frac{\partial z^u_*}{\partial\sigma}\frac{\partial^2 g_{*4}}{\partial (z^u_*)^2} + \frac{\partial z^l_*}{\partial\sigma}\frac{\partial^2 g_{*4}}{\partial z^u_*\partial z^l_*}\right] + \sigma\frac{\partial g_{*4}}{\partial z^u_*}\frac{\partial^2 z^u_*}{\partial\eta_*\partial\sigma} + \cdots \\
&= \sigma^{-1}\frac{\delta^u f'(z^{*u}) - \delta^l f'(z^{*l})}{F(z^{*u}) - F(z^{*l})} - \frac{\partial g_{*4}}{\partial \eta^*}\left(1 + \frac{\partial g_{*4}}{\partial \log\sigma}\right).
\end{aligned}
$$

In the derivations, "$\cdots$" indicates similiar computations for terms involving z^l_*. The computation of the partial derivatives of g_2 and g_3 make use of the formulae for g_4, by setting the appropriate value for the integral limits. That is, $z^u = \infty$ for g_2, and $z^l = -\infty$ for g_3.

The implementation of the constrained maximization step is not as straightforward, however. Under the constrained model, we saw that

$$
\eta_{*i} = \frac{x_{ip}}{x_{0p}}\{a(t_0) - \sigma q_{1-\omega}\} + \sum_{j=1}^{p-1}\beta_j v_{ij},
$$

is the form of the linear predictor for subject i. If we use this form directly to construct the model equation in survreg(), the routine application of survreg() will not work, since the model equation now explicitly involves σ. In theory, we need to plug in the re-estimated σ in the model equation after each complete cycle of the IRLS method, i.e., after having estimated β given an initial value for σ and β. There does not seem to be an easy way of implementing these steps, apart from a substantial modification of the internal code of survreg().

To overcome the above problems, an indirect approach requiring only minor modification of the internal code of survreg() can be utilized, as follows. First, we observe that z_{*i} can be written as

$$
\begin{aligned}
z_{*i} &= \frac{a(t_i) - \{\frac{x_{ip}}{x_{0p}}a(t_0) + \sum_{j=1}^{p-1}\beta_j v_{ij}\}}{\sigma} + \frac{x_{ip}}{x_{0p}}q_{1-\omega} \\
&= z'_{*i} + \frac{x_{ip}}{x_{0p}}q_{1-\omega},
\end{aligned}
$$

where $z'_{*i} = \frac{a(t_i)-\eta'_{*i}}{\sigma}$ and $\eta'_{*i} = \frac{x_{ip}}{x_{0p}}a(t_0) + \sum_{j=1}^{p-1}\beta_j v_{ij}$. By specifying a model with offset $\frac{\mathbf{x_p}}{x_{0p}}a(t_0)$, where $\mathbf{x_p}^T = (x_{1p}, \ldots, x_{np})$, survreg() routinely computes z'_{*i}. The required terms, z_{*i}, can then be obtained by a relatively simple change in the code, namely adding $\frac{\mathbf{x_p}}{x_{0p}}q_{1-\omega}$ to z'_{*}. This takes care of the partial derivatives with respect to η_{*}. For the partial derivatives involving σ, there is the additional change from z to δ. Since δ is simply z'_{*}, this change is also quite easily effected in the code. As in the case of unconstrained maximization of the log likelihood, the numerical properties of the IRLS approach in the constrained parameter problem are not easy to establish. While the sequence of estimates is expected to converge in many practical instances, it is difficult to establish sufficient conditions for convergence: see Green (1984).

Another important functional in the life data context is the quantile point for subjects with given covariate value. Suppose we are interested in the pth quantile lifetime, Q_p, for subjects with given covariates x_0, i.e., $Q_p = \inf\{t : S(t, x_0) \leq 1-p\}$. An approximate $100(1-\alpha)\%$ CI for Q_p is given by

$$\{\omega : \ell(\hat{\beta}, \hat{\sigma}) - \ell(\tilde{\beta}, \tilde{\sigma}) \leq \chi^2_{1.\alpha}/2\}\,,$$

where $(\hat{\beta}, \hat{\sigma})^T$ are the unrestricted MLEs and $(\tilde{\beta}, \tilde{\sigma})^T$ are the constrained MLEs under $H_0 : Q_p = \omega$, for some fixed $\omega > 0$. Analogous to other problems discussed in this section, we have $P(T > \omega; \mathbf{x_0}) = P(Z > \frac{a(\omega)-\beta^T\mathbf{x_0}}{\sigma}) = 1-p$, which implies that

$$\frac{a(\omega) - \beta^T\mathbf{x_0}}{\sigma} = q_{1-p}\,.$$

The remaining steps follow very closely those taken for the survivor function, and are omitted. The method discussed in this section appears to be easiest to implement for functions of β not involving σ, provided η maintains its linear form under

H_0. For linear functions of β, under the null hypothesis we have

$$z_{*i} = \frac{a(t_i) - \eta_{*i}}{\sigma},$$

where η_{*i} is a linear function of $(\beta_1, ..., \beta_{p-1})^T$. Thus, to fit the constrained model we simply specify a model with fixed offset and no intercept term, in the model equation for survreg(). No modification of the internal code of survreg() is needed in this case.

A Numerical Example

Table 4.4 summarizes lung cancer data from Glasser (1965). The response variable, Y, is the logarithm of survival time in days for patients with primary lung tumours. The starred values denote censored observations. Age (in years) and performance status are two covariates available for modelling the survival experience of this group of patients. The performance ratings 1-3 indicate complete hospitalization, 4-6 indicate partially confined to hospital and 7-10 indicate ability to care for oneself. For purposes of illustrating the methods discussed previously, let us suppose the model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \quad i = 1, ..., n,$$

is adequate, where $\epsilon_i \sim N(0, \sigma^2)$ and independent.

Table 4.4: Lung cancer data from Glasser (1965).

Log survival time (y)	Age (x_1)	Performance status (x_2)
1.94	42	4
2.23	67	6
1.94	62	4
1.98	52	6
2.23	57	5
1.59	58	6
2.13	55	6
1.80	63	7
2.32	44	5
1.92	62	7
2.15*	51	7
2.05*	64	10
2.48*	54	8
2.42*	64	3
2.56*	54	9
2.56*	57	9

By using our proposed method, we obtain the following approximate 90% CIs for $S(t_0; \mathbf{x_0})$, where $\mathbf{x_0} = (x_{01}, x_{02})^T$; see Table 4.5. We arbitrarily set $t_0 = 2.015$ and $x_{01} = 56.625$ years. Figures 4.4 and 4.5 show the LRS for $S(t_0; \mathbf{x_0})$, for the last two groups of patients in Table 4.5, i.e., values of 5 and 7 for performance status. The numerical implementation for this example was fairly straightforward. However, at times the survreg() routine fails to converge in the default number of steps specified in the internal controls. We simply increased the default until the likelihood values converged.

Table 4.5: Interval estimates for $S(t_0; \mathbf{x}_0)$ for patients aged 56.6 years.

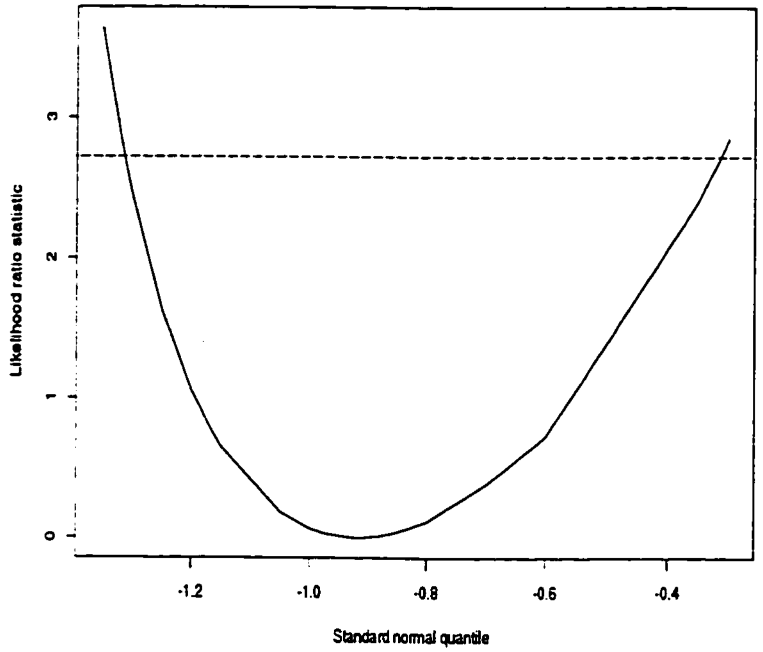
x_{02}	90% CI	MLE
3	(0.322, 0.883)	0.633
5	(0.624, 0.905)	0.821
7	(0.782, 0.968)	0.932

A Lagrange Multiplier Approach

The method used in the previous section may not work for some functionals. An example is $S(t_0, \mathbf{x}_1)/S(t_0, \mathbf{x}_2)$, which can be regarded as a measure of the relative magnitude of the probability of survival at t_0 for two distinct subjects with covariate vectors, $\mathbf{x}_1$, $\mathbf{x}_2$. If we set $S(t_0, \mathbf{x}_1)/S(t_0, \mathbf{x}_2) = \omega$, where ω is some fixed value of the functional, in general it is not possible to solve explicitly for β_p (or some other parameter) in terms of the remaining parameters. Another useful functional is $S(t_1, \mathbf{x}_0)/S(t_2, \mathbf{x}_0)$, where $\mathbf{x}_0$ is a fixed covariate value, and t_1, t_2 denote distinct times. This functional can be used to assess the change of probabilities from one point to another. Other functionals which are not easily handled by the method of the previous section are $P(t_1 \leq T \leq t_2; \mathbf{x}_1)$ and $P(t_1 \leq T \leq t_2; \mathbf{x}_1)/P(t_1 \leq T \leq t_2; \mathbf{x}_2)$, where $\mathbf{x}_1$, $\mathbf{x}_2$ are known, distinct covariate values. In some failure time contexts, the latter functional may usefully be regarded as the relative risk of an event in the interval $[t_1,\ t_2]$.

A Lagrange multiplier argument can be applied to obtain likelihood-based CIs for functionals that can be expressed as linear combinations of g_i, $i = 1, \ldots, 4$. Some useful functionals in the failure time context, including the preceding ones, can be expressed in this manner. An advantage of this approach over the usual profile likelihood approach is that an IRLS regression also applies under the constrained model. Therefore, with some modification of the computer code, we can continue

Figure 4.4: Likelihood ratio statistic for interval estimation of the survival probability for lung cancer patients aged 56.6 years who were partially confined to hospital, based on the data in Table 4.4. The height of the horizontal dashed line is 3.841 units.

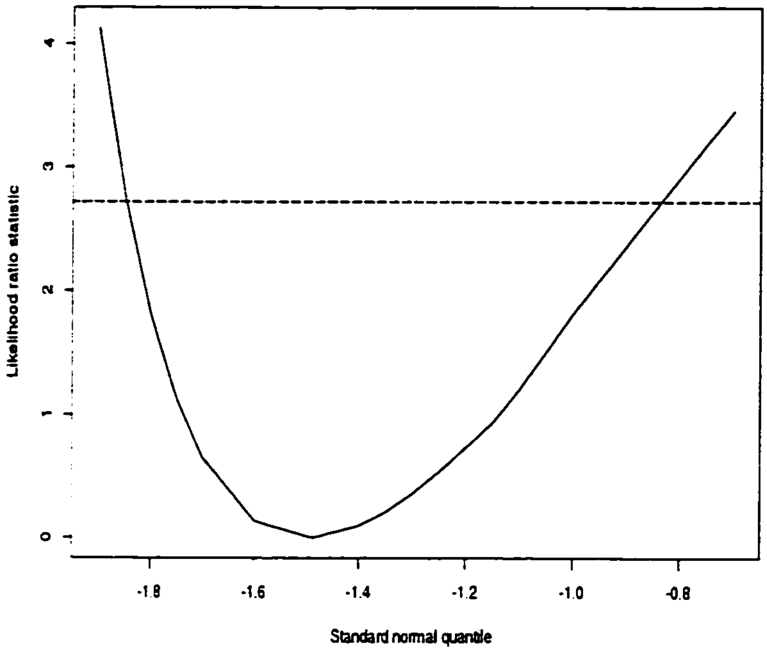


to use the survreg() function in S-Plus.

We introduce the method by considering a simple example. Consider the interval estimation of $S(t_0, \mathbf{x}_0)$, which was also performed earlier. It is easier to first obtain an interval estimate for $\log S(t_0, \mathbf{x}_0)$, and then invert to obtain the corresponding estimate for $S(t_0, \mathbf{x}_0)$. Since $S(t_0, \mathbf{x}_0) = P(Z > z_0)$, where $z_0 = \frac{a(t_0)-\beta^T \mathbf{x}_0}{\sigma}$, we maximize the augmented log likelihood

$$\ell_\xi = \ell - \xi g_2(z_0)$$

Figure 4.5: Likelihood ratio statistic for interval estimation of the survival probability for lung cancer patients aged 56.6 years who lived at home, based on the data in Table 4.4. The height of the horizontal dashed line is 3.841 units.



with respect to β to obtain the constrained MLEs corresponding to a fixed value of $\log S(t_0, x_0)$. Given a fixed value of ξ, we obtain

$$\frac{\partial \ell_\xi}{\partial \beta_j} = \sum_{i=1}^{n} x_{ij} \frac{\partial g}{\partial \eta_i} - \xi x_{0j} \frac{\partial g_2}{\partial \eta_0}. \qquad (4.4)$$

$$\frac{\partial^2 \ell_\xi}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^{n} x_{ij} x_{ik} \frac{\partial^2 g}{\partial \eta_i^2} - \xi x_{0j} x_{0k} \frac{\partial^2 g_2}{\partial \eta_0^2}, \qquad (4.5)$$

where $\eta_0 = \beta^T \mathbf{x_0}$. The implementation of these steps proceeds as follows. For easy exposition, we split the implementation into two simple parts, each corresponding to whether $\xi > 0$ or $\xi < 0$. For $\xi < 0$, we can write the preceding equations as

$$\frac{\partial \ell_\xi}{\partial \beta_j} = \sum_{i=0}^{n} x_{ij} \frac{\partial g}{\partial \eta_i} = (X^T U_x)_j,$$

$$\frac{\partial^2 \ell_\xi}{\partial \beta_j \partial \beta_k} = \sum_{i=0}^{n} x_{ij} x_{ik} d_i = -(X^T D_x X)_{jk},$$

where

$$d_i = \begin{cases} -\frac{\partial^2 g}{\partial \eta_i^2}, & i = 1, \dots, n \\ -|\xi| \frac{\partial^2 g_2}{\partial \eta_i^2}, & i = 0 \end{cases}$$

$$U_x = \left(\frac{\partial g}{\partial \eta_1}, \dots, \frac{\partial g}{\partial \eta_n}, \xi \frac{\partial g_2}{\partial \eta_0} \right)^T,$$

and

$$D_x = diag\{d_i\}.$$

This leads directly to another IRLS regression which can be carried out by survreg(), provided we account for the additional term in the score vector and information matrix.

For $\xi > 0$, equations (4.4) and (4.5) can be viewed as the original unrestricted model terms less an "adjustment" term,

$$\frac{\partial \ell_\xi}{\partial \beta_j} = (X^T U)_j - x_{0j} (U_x)_{n+1},$$

$$\frac{\partial^2 \ell_\xi}{\partial \beta_j \partial \beta_k} = -\{(X^T D X)_{jk} - x_{0j} x_{0k} d_0\}.$$

Simple modifications to the code will enable us to implement the constrained maximization step for $\xi > 0$. Similar adjustments are entailed for the partial derivatives involving σ.

It is not difficult to extend the procedure to linear functionals of g_i, namely. $\sum_{i=1}^{k} a_i g_i$, where a_i are real numbers (usually, $a_i = \pm 1$). For linear functions of β, it appears that the direct substitution method is easier to use, even though we can still employ the Lagrange multiplier technique. To see the latter, we note that for linear functionals, the information matrix is unchanged. Therefore, we only need to modify the score function terms accordingly (see above). The convergence properties of the proposed procedure are not known at present. It is possible that the procedure might break down either by not converging or by converging to points other than the constrained MLE.

For a numerical illustration, consider the data of Glasser (1965) again. Let us suppose interest centers on a comparison of the relative rate of survival between patients who are able to care for themselves (Group 1), and those who are either confined to hospital or partially confined to hospital (Group 2). That is, we consider the model

$$Y_i = \alpha_0 + \alpha_1 v_{1i} + \alpha_2 v_{2i} + \epsilon_i, \quad i = 1, \ldots, n,$$

where the ϵ_i's satisfy the independent, normal assumptions, v_1 is age and

$$v_{2i} = \begin{cases} 1, & i \in \text{Group 1} \\ 0, & i \in \text{Group 2} \end{cases}$$

A one-tailed test of $\alpha_2 = 0$ is significant at the 5% level, so the model will be used to demonstrate the Lagrange multiplier approach. We define $\mathbf{x}_1 = (1,\ 56.625,\ 0)^T$ and $\mathbf{x}_2 = (1,\ 56.625,\ 1)^T$. Using our approach, we obtain an approximate 95% CI for $\frac{S(2.015;\mathbf{x}_1)}{S(2.015;\mathbf{x}_2)}$; we use the same value $t_0 = 2.015$ for illustration. This is given by (0.717, 0.862); the MLE is 0.766. This result provides evidence of superior survival experience for Group 1 patients at the given time point and age profile.

Numerically, the convergence criterion in survreg() could not always be met in the default number of iterations, so that we adopted the same remedy as in the previous example.

Remarks. For linear functions of β, an alternative Lagrange multiplier argument may be used to obtain constrained MLEs for β, along the lines of Kim and Taylor (1995). To illustrate, consider interval estimation of $\eta_0 = \beta^T \mathbf{x_0}$. Let S_ξ and I_ξ represent the score and information matrix based on the augmented log likelihood ℓ_ξ; the Lagrange multiplier is ξ. These are related to the corresponding unrestricted quantities via $S_\xi = S + \xi \mathbf{x_0}$ and $I_\xi = I$. Let $\beta^{(m)}$ and $\beta_\xi^{(m)}$ denote the unrestricted and constrained estimates, respectively, at the mth cycle of a Newton-Raphson iteration. Then

$$\begin{aligned}\beta_\xi^{(m+1)} &= \beta_\xi^{(m)} + I_\xi^{-1} S_\xi \\ &= \beta_\xi^{(m)} + I^{-1}S + \xi I^{-1}\mathbf{x_0} \\ &= \beta^{(m+1)}(\beta_\xi^{(m)}) + \xi I^{-1}\mathbf{x_0}.\end{aligned}$$

where $\beta^{(m+1)}(\beta_\xi^{(m)}) = \beta_\xi^{(m)} + I^{-1}S$. Substituting this expression into the constraint $\eta_0 = \omega$, and solving for ξ, we get

$$\xi = (\mathbf{x_0}^T I^{-1} \mathbf{x_0})^{-1} \left\{ \omega - (\beta^{(m+1)}(\beta_\xi^{(m)}))^T \mathbf{x_0} \right\} .$$

This is substituted into the Newton-Raphson iteration to obtain

$$\beta_\xi^{(m+1)} = \beta^{(m+1)}(\beta_\xi^{(m)}) + (\mathbf{x_0}^T I^{-1} \mathbf{x_0})^{-1} \left\{ \omega - (\beta^{(m+1)}(\beta_\xi^{(m)}))^T \mathbf{x_0} \right\} I^{-1}\mathbf{x_0} .$$

We obtain a similiar solution when there is more than one linear constraint; for additional details, see Kim and Taylor (1995). □

Chapter 5

Further Work

This thesis has explored the profile likelihood approach to interval estimation of functionals in various settings. The proposed procedures are computationally intensive compared to the standard approach based on the normal approximation. However, with modern computing facilities, the likelihood-based approach will become increasingly easier to implement.

In chapter 2, we proposed a simple method of obtaining a Bartlett correction for functionals in parametric inference. It would be of interest to also derive a Bartlett factor in this setting more directly, i.e., by explicitly accounting for the Lagrange multiplier in the computation of the Bartlett factor. Cordeiro (1993) provided useful matrix formulae for computing the Bartlett factor in two distinct situations, viz., in tests of null hypotheses which specify a parameter vector in the presence of nuisance parameters, and for testing a scalar parameter which is orthogonal to the remaining parameters. These ideas may be utilized in future work to provide more easily accessible formulae for the Bartlett factor of Chapter 2. Currently, there is also some interest in the derivation of appropriate Bartlett factors for the nonparametric and semi-parametric settings. For the latter, some results were obtained by Gu and Zheng (1993) for the semi-parametric Cox proportional hazards model with a single covariate. Although no Bartlett correction is known for the nonparametric setting apart from empirical likelihood, Mykland (1995) provides arguments for the existence of such correction factors. His arguments can be summarized as follows. In parametric inference, it is well-known that, in practical problems, the LRS is preferrable to the studentized score statistic for interval estimation (cf. chapter 1). This idea is exploited in Mykland (1995) who proposed a "likelihood" construction (called dual likelihood) for martingale-based inference. The concept of dual likelihood can be used in particular to construct a dual likelihood ratio statistic which is useful for hypothesis testing and interval estimation. It is then argued that dual likelihood can essentially be regarded as a likelihood in the usual sense. Further, for independent data, the dual LRS coincides with the nonparametric LRSs derived from nonparametric considerations. For example, in the case of right-censored survival data from a homogeneous population, the dual LRS for testing a fixed value of the cumulative hazard function is the same as the nonparametric LRS based on the point process likelihood (assuming continuous cumulative hazard). It may therefore be argued that the existence of a Bartlett factor for empirical and point process LRSs follows as a corollary to the existence of such a correction for the parametric setting. We hope to use these ideas in future work.

In Example 3.6 of chapter 3, we proposed a Lagrange multiplier solution to the constrained maximization problem in the important context of exponential family models. Since numerical solutions are unavoidable in general, it may be worthwhile to find algorithms which are superior in terms of ease of implementation. In this short note, we point out a useful technique which can be applied to that setting. This technique arises from the fixed-point theorem in analysis, but its connection to the EM algorithm has not been fully exploited thus far. Our adaptation of this technique to the "missing" data context is motivated by a recent paper of Navidi (1997) which focuses on the one-parameter exponential family model. He shows that $E(S|\theta)$ and $E(S|T_{obs};\theta)$ are both increasing functions of θ in a neighbourhood of the MLE $\hat{\theta}$, with $E(S|\theta)$ increasing more rapidly. Further, $E(S|T_{obs};\theta)-E(S|\theta) = \ell'(\theta)$. Hence,

$$E(S|\hat{\theta}) = E(S|T_{obs};\hat{\theta}),$$

which is identical to the algorithm derived in Cox and Oakes (1984).

By a graphical illustration, Navidi then showed that the EM estimates converge monotonically to $\hat{\theta}$. In fact, these findings are closely related to the fixed point theory in analysis. Briefly, for some differentiable function $f : \Re \rightarrow \Re$, the theory sets sufficient conditions for the existence of a fixed point x_0 of f, i.e., $f(x_0) = x_0$. The same conditions establish that the iteration defined by $x_{n+1} = f(x_n)$, $n = 1, 2, \ldots$, converges to x_0 under the assumptions of the theorem. The connection to Navidi's work is seen most directly for the normal distribution with mean μ and known variance σ^2. Consider a single random variable from this distribution. The canonical parameter θ is μ in this case, and $E(S|\theta') = \theta' = \mu'$. Hence,

$$\mu' = E(S|T_{obs};\mu)$$

defines the updating scheme according to the EM algorithm, and we can also deduce by the fixed point theory that this algorithm converges to the MLE, $\hat{\mu}$. This suggests an alternative implementation for the EM algorithm.

For brevity, let $F(\theta)$ and $G(\theta)$ represent $E(S|\theta)$ and $E(S|T_{obs};\theta)$, respectively, and let $H(\theta) = G(\theta) - F(\theta)$. Given that $F'(\theta) > G'(\theta)$, the problem is to find the value of θ satisfying $H(\theta) = 0$. This problem is thus equivalent to finding the zeroes of H, and is discussed, for example, in Protter and Morrey (1991). Its solution makes use of the fixed point theorem. Additional conditions are required, which we outline in the following. Suppose then that we desire to locate the zeroes of $f : \Re \rightarrow \Re$. We assume that $f(x)$ is differentiable on $I = [a, b]$, with $f(a)$ and $-f(b)$ contained in $[0, b-a]$, and there exists some k, k' such that $-1 < k' \leq f'(x) \leq k < 0$ for all $x \in I$. Then, there exists an $x_0 \in [a, b]$ such that $f(x_0) = 0$. Furthermore, for any $x_1 \in [a, b]$, the iteration $x_{n+1} = f(x_n)$, $n = 1, 2, \ldots$ converges to x_0. For our problem, the conditions need to be checked in each application, even though it is known that for the function of interest H, $H'(\hat{\theta}) = \ell''(\hat{\theta}) < 0$. This result thus yields the iteration

$$\theta^{(m+1)} = E(S|T_{obs}; \theta^{(m)}) - E(S|\theta^{(m)})$$

as a feasible alternative to the usual updating scheme. Given a suitable choice of $\theta^{(m)}$, this yields a convenient way to implement the EM algorithm for scalar θ. This numerical scheme will be considered in subsequent work. It would also be very useful to see whether a multivariate version of the above works as well. This algorithm could then be applied to the constrained maximization problems encountered in Example 3.6.

As a general theoretical basis for profile likelihood-based intervals remains elusive, numerical investigation of their coverage properties and comparisons with results based on the usual normal approximation will be helpful. This research will undoubtedly involve a substantial amount of computational effort.

In chapter 4, we described some techniques for obtaining likelihood-based interval estimates for failure time functionals. It would be desirable to evaluate the coverage probabilities as well as other properties of the proposed methods. For the IRLS approach, the convergence properties of the technique in the constrained maximization setting are not known at present, and are an area for future work.

In many biomedical applications, interest focuses on the effect of covariates on response measurements. To date, the most popular model in biomedical work is the Cox (1972) proportional hazards regression model. It would be useful to see if the techniques discussed in Section 4.2 could be adapted to yield likelihood-based interval estimates for functionals in this semi-parametric Cox model. In the following, we summarize some preliminary work on this problem, as well as indicate future research directions for this area.

Suppose the lifetimes of n subjects are distributed independently with hazard function

$$\lambda(t, z_i(t)) = \lambda_0(t)\exp\{\beta^T z_i(t)\}\,,$$

where $\lambda_0(t)$ is an unknown baseline hazard function, $\beta \in \Re^p$ is a p-dimensional vector of regression coefficients, and $z_i(t) \in \Re^p$ represents a known p-dimensional covariate for subject i at time t. Then the full likelihood function based on the observed data is

$$L = \prod_i \left[\lambda_0(t_i)\exp\{\beta^T z_i(t_i)\}\exp\left\{-\int_0^{t_i}\lambda_0(u)\exp\{\beta^T z_i(u)\}du\right\}\right]\,.$$

By defining $Y_i^\beta(u) = \exp\{\beta' z_i(u)\}I\{u \leq t_i\}$ and $Y^\beta(u) = \sum_i Y_i^\beta(u)$, we can write

$$L = \prod_i \frac{Y_i^\beta(t_i)}{Y^\beta(t_i)}\{\prod_i Y^\beta(t_i)\lambda_0(t_i)\ \}\exp\left\{-\int_0^\infty Y^\beta(u)\lambda_0(u)du\right\}\,.$$

Let $\mathcal{R}_i = \{j : t_j \geq t_i\}$ represent the risk set at t_i. The first term of the preceding expression,

$$L_P(\beta) = \prod_i \frac{Y_i^\beta(t_i)}{Y^\beta(t_i)} = \prod_i \frac{\exp\{\beta' z_i(t_i)\}}{\sum_{j\in\mathcal{R}_i}\exp\{\beta' z_j(t_i)\}}\,,$$

is the Cox partial likelihood, and may be used to estimate β. By assuming λ_0 to be piecewise constant on intervals of length ϵ, Johansen (1983) showed that $\hat{\beta} = \sup L_P(\beta)$ and

$$\hat{\Lambda}(t) = \sum^{(t)} \frac{1}{Y^{\hat{\beta}}(t_i)} \tag{5.1}$$

jointly maximize the full log likelihood. In equation (5.1), the arguments $\{t_i\}$ are the distinct event times less than t. The estimators $\hat{\Lambda}(t)$ and $\hat{\beta}$ can be interpreted as nonparametric (or generalized) maximum likelihood estimators. Alternatively, one may assume that λ_0 is piecewise constant between distinct failure times. In this case, the Breslow estimator is obtained (Breslow, 1974). It is known that equation (5.1) coincides with the Breslow estimate of Λ on intervals of the form $(t_i, t_{i+1}]$.

The asymptotic distribution theory of standard test statistics has been covered in detail by various authors; see, for example, Fleming and Harrington (1991). Some key findings are summarized as follows. Analogous to standard likelihood theory, the usual proofs in the censored data context focus on establishing the approximate normal distribution of the score vector, since this result leads directly to the asymptotic distribution for $\hat{\beta}$. Confidence intervals for β or subsets of it are commonly based on the approximate normal distribution of $\hat{\beta}$. While the partial LRS is also used in inference about β (or subsets of it), we have not seen its use in the case of parameter functions. The Breslow estimator also has an asymptotic normal distribution; the joint asymptotic distribution of $\hat{\beta}$ and the Breslow estimator can in turn be used to yield an approximate normal distribution for $S(t, \mathbf{x_0})$, the survivor function for subjects with given covariates $\mathbf{x_0}$.

In some applications, a confidence interval for the linear predictor may be of interest. For example, in a binary logistic regression model, one may wish to obtain confidence bounds on the probability, $p(Y|\mathbf{x})$, of a response ($Y = 1$) conditional on a given vector of covariates $\mathbf{x}$. The usual approach is to find a confidence bound for the linear predictor $\eta = \beta^T\mathbf{x}$ first, and then use the one-to-one relationship between η and $p(Y|\mathbf{x})$ to obtain the required interval for $p(Y|\mathbf{x})$.

Alho (1992) used a profile likelihood-based method to construct confidence intervals for the linear predictor in generalized linear models. His approach can also be adapted to the Cox partial likelihood to obtain a corresponding interval estimate for the linear predictor in the PH model. We assume, for the time being, that all the covariates are fixed, i.e., not time-dependent. To test the hypothesis $H_0 : \beta^T\mathbf{z_0} = \omega$, for some arbitrary fixed $\mathbf{z_0}$ and ω, the S-Plus routine coxph() can be used with the appropriate modification of the covariate matrix to yield the constrained MLEs for the regression parameter β. The partial likelihood ratio statistic (PLRS) for testing H_0 is given by

$$W_P = 2\{\ell_P(\hat{\beta}) - \ell_P(\tilde{\beta})\},$$

where $\ell_P(\beta) = \log L_P(\beta)$, $\hat{\beta}$ is the maximum partial likelihood estimator (MPLE), and $\tilde{\beta}$ is the constrained partial likelihood estimator when $\beta^T\mathbf{z_0} = \omega$. Analogous to a parametric setup, the PLRS may be used to obtain likelihood-based intervals for the linear predictor in the usual way. Assuming $z_{0p} \neq 0$, under H_0 we may write

$$\beta_p = \frac{\omega - \sum_{i=1}^{p-1} \beta_i z_{0i}}{z_{0p}},$$

which leads to the constrained form of the linear predictor, $\beta^T\mathbf{z_i} = \gamma^T\mathbf{y_i} + \omega y_{ip}$, where $y_{ip} = \frac{z_{ip}}{z_{0p}}$, $y_{ij} = (z_{ij} - z_{0j}\frac{z_{ip}}{z_{0p}})$ for $j = 1, \dots, p-1$, $\mathbf{z_i} = (z_{i1}, \dots, z_{i,p-1})^T$, and $\gamma = (\beta_1, \dots, \beta_{p-1})^T$. Under H_0, the likelihood function is, up to a proportionality constant,

$$L = \prod_i \frac{Y_i^\gamma(t_i)}{Y^\gamma(t_i)} \{ \prod_i Y^\gamma(t_i)\lambda_0(t_i) \} \exp\left\{ - \int_0^\infty Y^\gamma(u)\lambda_0(u)du \right\} ,$$

where $Y_i^\gamma(u) = \exp\{\gamma^T \mathbf{y}_\mathbf{i} + \omega y_{ip}\} I\{u \le t_i\}$ and $Y^\gamma(u) = \sum_i Y_i^\gamma(u)$. Given a fixed value for ω, the first term of this expression is the partial likelihood for γ. Therefore, we can obtain the constrained MPLE, $\tilde{\gamma}$, by appropriately modifying the design matrix of the regression problem. The constrained MPLE can be obtained by specifying a fixed offset ωy_{ip} in the S-Plus routine coxph() and substituting the adjusted covariates y_{ij} for x_{ij}. Subject to sufficient regularity conditions, $W_P \sim \chi_1^2$ under H_0. An approximate $100(1-\alpha)\%$ CI for the linear predictor thus consists of the set $\{\omega : W_P(\omega) \le \chi^2_{1.1-\alpha}\}$. We note that the above steps carry over without modification when the covariates are time-dependent.

It may also be of interest to derive the joint MLEs of β and $\Lambda_0(t)$ under H_0. By utilizing the piecewise constant assumption for λ_0, and following the same argument outlined in Johansen (1983), it can be seen that $\tilde{\beta}$ and

$$\tilde{\Lambda}_0(t) = \sum^{(t)} \frac{1}{Y^{\tilde{\beta}}(t_i)}$$

jointly maximize the constrained likelihood function. This follows since the score equations for λ_0 are unchanged in form, under H_0.

Some heuristic ideas can be utilized to obtain profile-based CIs for $S(t, \mathbf{z})$, the survivor function for an individual with covariate value $\mathbf{z}$. We continue to assume that the baseline hazard rates are piecewise constant. In some studies, e.g. population or demographic studies, reliable information on baseline parameters may be available, perhaps based on historic data from comparable studies, censuses, etc.

For the continuous-time PH model, assuming that the survivor function $S_0(t)$ is known from such sources of information, the methods of the last section can be adapted to yield likelihood-based interval estimates for $S(t, \mathbf{z}) = S_0(t)^{\exp(\beta^T \mathbf{z})}$.

Usually no information is available on $S_0(t)$ apart from the assumption of its functional relation to $S(t, \mathbf{z})$, i.e., the proportional hazards assumption. In this case, we can utilize Johansen's approach to derive an iterative procedure to obtain an interval estimate for $S(t, \mathbf{z})$. Consider a fixed value of $S(t, \mathbf{z})$, say $S(t, \mathbf{z}) = \omega$. For convenience, we take t to be a time point just after a death time, t_i, i.e. set $t = t_i+$. Solving for β_p, we get

$$\beta_p = \frac{1}{z_p} \; [\; \log \left\{ \frac{-\log(\omega)}{\Lambda_0(t)} \right\} - \sum_{i=1}^{p-1} \beta_i z_i \;].$$

assuming $z_p \neq 0$. Substituting β_p into the likelihood function, we obtain

$$L = \prod_i \left\{ \frac{e^{\gamma^T \mathbf{y}_i}}{\sum_j e^{\gamma^T \mathbf{y}_j}} \right\} \prod_i \exp\{\frac{z_{ip}}{z_p} \log \frac{-\log \omega}{\Lambda_0(t)}\} \prod_i \left\{ \lambda_0(t_i) \sum_j e^{\gamma^T \mathbf{y}_j} \right\} \exp\{- \int_0^\infty \lambda_0(u) \sum_j Y_j^{\gamma}(u) du\}$$

as the likelihood function under the null hypothesis. The first term of L is the partial likelihood for γ, and can be maximized in the standard way to give the MPLE, $\hat{\gamma}$. Given $\hat{\gamma}$, the next step is to maximize with respect to λ_0,

$$\begin{aligned} \log L(\hat{\gamma}. \Lambda_0) &= \sum_i (\frac{z_{ip}}{z_p})\{\log(-\log \omega) - \log \Lambda_0(t)\} + \sum_i \log \lambda_0(t_i) - \int_0^\infty \lambda_0(u) \sum_j Y_j^{\hat{\gamma}}(u) du \\ &= \sum_i (\frac{z_{ip}}{z_p})\{\log(-\log \omega) - \log[\sum_i \lambda_0(t_i) \int_{I_i \cap [0,\, t]} du]\} + \sum_i \log \lambda_0(t_i) \\ &- \sum_i \lambda_0(t_i) \int_{I_i} \sum_j \exp[\; \frac{z_{jp}}{z_p} \log\{\frac{-\log \omega}{\Lambda_0(t)}\} + \beta^T \mathbf{x}_j] I\{u \leq t_j\} du. \end{aligned}$$

In the above equation, we expressed $\Lambda_0(t)$ as $\sum_i \lambda_0(t_i) \int_{I_i \cap [0,\, t]} du$. Note that, for

$t_i > t$,

$$\frac{\partial \ell(\hat{\gamma}, \Lambda_0)}{\partial \lambda_0(t_i)} = \frac{1}{\lambda_0(t_i)} - \int_{I_i} \sum_j e^{\frac{z_{jp}}{z_p} \log\{\frac{-\log \omega}{\Lambda_0(t)}\} + \beta^T \mathbf{x}_j} I\{u \leq t_j\} du,$$

while

$$\begin{aligned} \frac{\partial \log L(\hat{\gamma}, \Lambda_0)}{\partial \lambda_0(t_i)} &= \frac{1}{\lambda_0(t_i)} - \int_{I_i} \sum_j e^{\frac{z_{jp}}{z_p} \log\{\frac{-\log \omega}{\Lambda_0(t)}\} + \beta^T \mathbf{x}_j} I\{u \leq t_j\} du + \frac{\sum_i \frac{z_{ip}}{z_p} \int_{I_i \cap [0,\, t]} du}{\sum_j \lambda_0(t_j) \int_{I_j \cap [0,\, t]} du} \\ &+ \frac{\int_{I_i \cap [0,\, t]} du}{\sum_i \lambda_0(t_i) \int_{I_i \cap [0,\, t]} du} \sum_j \lambda_0(t_j) \int_{I_j} \sum_k (\frac{z_{kp}}{z_p}) e^{\frac{z_{kp}}{z_p} \log\{\frac{-\log \omega}{\Lambda_0(t)}\} + \beta^T \mathbf{x}_k} I\{u \leq t_k\} du \end{aligned}$$

for $t_j \leq t$. By assuming, as in Johansen (1983), that λ_0 is piecewise constant on intervals of length ϵ, the likelihood equations for $t_j \leq t$ turn out to be

$$\lambda_0(t_i) = \left[\int_{I_i} Y^{\hat{\gamma}}(u) du - \frac{\sum_i \frac{z_{ip}}{z_p} + \sum_j \lambda_0(t_j) \int_{I_j} \sum_k (\frac{z_{kp}}{z_p}) Y_k^{\hat{\gamma}}(u) du}{\sum^{(t)} \lambda_0(t_j)}\right]^{-1},$$

where $Y_k^{\hat{\gamma}}(u) = \exp\{\frac{z_{kp}}{z_p} \log[\frac{-\log \omega}{\Lambda_0(t)}] + \beta^T \mathbf{x}_k\} I\{u \leq t_k\}$. The likelihood equations for $t_j > t$ can be solved to give

$$\lambda_0(t_i) = \left[\int_{I_i} Y^{\hat{\gamma}}(u) du\right]^{-1}.$$

Given the constrained estimates, we compare the observed value of the LRS corresponding to the current value of ω with the relevant χ_1^2 quantile. Based on this comparison, we either accept the current value of ω as an endpoint of the interval estimate for $S(t, \mathbf{z})$, or increase (decrease) the current value appropriately. The procedure halts when both end-points are located.

For future research, it would be desirable to see if the methods described above could be implemented, as well as to look into interval estimates for other functionals. Numerical evaluation of the coverage properties of these interval estimates, at least in the simplest cases, would be desirable.

Appendix A

Some Partial Derivatives for Example 3.3

Some partial derivatives used in the EM1 algorithm to obtain an interval estimate for the prevalence function, $P(t)$. Let

$$\delta(t) = \lambda_1(2) \prod_{2 \leq v < t} [1 - \lambda_{11}(v|2)] + \sum_{j=3}^{t} \{ \lambda_1(j) \prod_{1 < v < j} \lambda_3(v) \prod_{j \leq v < t} [1 - \lambda_{11}(v|j)] \}.$$

$$\gamma(t) = \prod_{1 \leq v < t} [1 - \lambda_{11}(v|1)],$$

$$\Gamma(t) = \{ \prod_{1 < v < t} \lambda_3(v) + \omega[\gamma(t) - \delta(t)]\}^2,$$

and

$$\Lambda(t) = \prod_{1 < v < t} \lambda_3(v) + \omega[\gamma(t) - \delta(t)].$$

$$\frac{\partial\lambda_1(1)}{\partial\lambda_1(j)} = \begin{cases} -\frac{1-\lambda_2(1)}{\Gamma(t)}\left\{\omega\gamma(t)\left[\frac{\prod_{1<v<t}}{\lambda_3(j)} + \omega\frac{\partial\delta(t)}{\partial\lambda_1(j)}\right]\right\}, & 2 \le j < t \\ -\frac{1-\lambda_2(1)}{\Gamma(t)}\left\{\omega\gamma(t)\left[\omega\frac{\partial\delta(t)}{\partial\lambda_1(t)}\right]\right\}, & j = t \\ 0, & j > t \end{cases}$$

$$\frac{\partial\lambda_1(1)}{\partial\lambda_2(j)} = \begin{cases} -\left\{1 - \omega\frac{\gamma(t)}{\prod_{1<v<t}\lambda_3(v)+\omega[\gamma(t)-\delta(t)]}\right\} - \frac{1-\lambda_2(1)}{\Gamma(t)}\left\{\omega\gamma(t)\left[\omega\frac{\partial\delta(t)}{\partial\lambda_2(j)}\right]\right\}, & j = 1 \\ -\frac{1-\lambda_2(1)}{\Gamma(t)}\left\{\omega\gamma(t)\left[\frac{\prod_{1<v<t}\lambda_3(v)}{\lambda_3(j)} + \omega\frac{\partial\delta(t)}{\partial\lambda_1(t)}\right]\right\}, & 2 \le j < t \\ 0, & j \ge t \end{cases}$$

$$\frac{\partial\lambda_1(1)}{\partial\lambda_{11}(j|1)} = \begin{cases} -\frac{1-\lambda_2(1)}{\Gamma(t)}\omega\frac{\gamma(t)}{1-\lambda_{11}(j|1)}\{\Lambda(t) - \omega\gamma(t)\}, & 1 \le j < t \\ 0, & j \ge t \end{cases}$$

For $2 \le k < t$,

$$\frac{\partial\lambda_1(1)}{\partial\lambda_{11}(j|k)} = \begin{cases} -\frac{1-\lambda_2(1)}{\Gamma(t)}\omega^2\gamma(t)\frac{\partial\delta(t)}{\partial\lambda_{11}(j|k)}, & k \le j < t \\ 0, & j \ge t \end{cases}$$

$$\frac{\partial\lambda_1(1)}{\partial\lambda_{11}(j|k)} = 0,$$

for $k > t$.

Bibliography

Aitkin, M. and Clayton, D. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Appl. Statist.*, **29**. 156–63.

Aitkin. M.. Anderson, D., Francis. B., and Hinde. J. (1989). *Statistical Modelling in GLIM.* Clarendon Press. Oxford.

Alho. J. M. (1992). On the computation of likelihood ratio and score test based confidence intervals in generalized linear models. *Statist. Med.*, **11**, 923–30.

Barndorff-Nielsen, O. E. and Blaesild, P. (1986). A note on the calculation of Bartlett adjustments. *J. R. Statist. Soc. B*, **48**, 353–58.

Barndorff-Nielsen, O. E. and Cox, D. R. (1984). Bartlett adjustment to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. R. Statist. Soc. B*, **46**, 483–98.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. A*, **160**, 268.

Berlin, B., Brodsky, J., and Clifford, P. (1979). Testing disease dependence in survival experiments with serial sacrifice. *J. Am. Statist. Assoc.*, **74**, 5–14.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics*. **30**, 89–99.

Conte, S. D. and De Boor, C. (1980). *Elementary Numerical Analysis*. McGraw-Hill, New York.

Cook, R. D., Cook, D. J., and Sackett, D. L. (1995). Modelling the clinical impact of treatment from published randomized trials. Unpublished manuscript.

Cordeiro, G. M. (1983). Improved likelihood ratio statistics for generalized linear models. *J. R. Statist. Soc. B*, **45**, 404–13.

Cordeiro, G. M. (1993). General matrix formulae for computing Bartlett corrections. *Statist. and Prob. Letters*, **16**, 11–18.

Cox, D. R. (1970). *Analysis of Binary Data*. Chapman and Hall. London.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. Royal Statist. Soc. B*, **34**, 187–220.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Methuen. London.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall. London.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B*, **49**, 1–39.

Critchley, F., Ford, I., and Rihal, O. (1988). Interval estimation based on the profile likelihood: Strong Lagrangian theory with applications to discrimination. *Biometrika*, **75**, 21–28.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Royal Statist. Soc. B*, **39**, 1–38.

Dewanji, A. and Kalbfleisch, J. D. (1986). Nonparametric methods for survival sacrifice experiments. *Biometrics*, **42**, 325–41.

DiCiccio, T. J., Hall, P., and Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.*, **19**, 1053–61.

Dinse, G. E. and Lagakos, S. W. (1982). Nonparametric estimation of lifetime and disease onset distributions from incomplete observations. *Biometrics*, **38**, 921–32.

Duncan, B. B. A., Zaimi, F., Newman, G. B., Jenkins, J. G., and Aveling, W. (1984). Effect of premedication on the induction dose of thiopentine in children. *Anaesthesia*, **39**, 426–28.

Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, **21**, 826–38.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Frydenberg, M. and Jensen, J. L. (1989). Is the 'Improved Likelihood Ratio Statistic' really improved in the discrete case? *Biometrika*, **76**, 655–661.

Frydman, H. (1992). A nonparametric estimation procedure for a periodically observed three-state Markov process, with application to AIDS. *J. Royal Statist. Soc. B*, **54**, 853–66.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, **52**, 203–23.

Gill, P., Murray, W., and Wright, M. H. (1992). *Practical Optimization*. Academic Press, London.

Glasser, M. (1965). Regression analysis with dependent variable censored. *Biometrics*, **21**, 300–307.

Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives (with discussion). *J. Royal Statist. Soc. B*, **46**, 149–92.

Gu, M. G. and Zheng, Z. K. (1993). On the Bartlett adjustment for the partial likelihood ratio test in the Cox regression model. *Statistica Sinica*, **3**, 543–555.

Gu, X. L. (1996). *Statistical Analysis of Incomplete Data Arising in Biomedical Studies*. Unpublished Ph. D. Thesis, University of Waterloo.

Hall, P. and La Scala (1990). Methodology and algorithms of empirical likelihood. *I.S.I. Review*, **58**, 109–27.

Hoel, D. G. and Walburg, H. E. (1972). Statistical analysis of survival experiments. *J. Nat. Cancer Institute*, **49**, 361–72.

Johansen, S. (1983). An extension of Cox's regression model. *Inter. Statist. Rev.*, **51**, 165–174.

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

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observation. *J. Am. Statist. Assoc.*, **53**, 457–81.

Kim. D. and Taylor, J. M. G. (1995). The restricted EM algorithm for maximum likelihood estimation under linear restrictions on the parameters. *J. Am. Statist. Assoc.*, **90**, 708–16.

Kodell, R. L. and Nelson, C. J. (1980). An illness-death model for the study of the carcinogenic process using survival/sacrifice data. *Biometrics*, **36**, 267–77.

Kolassa. J. E. (1994). *Series Approximation Methods in Statistics.* Springer. New York.

Laird. N. (1982). Computation of Variance Components Using the EM Algorithm. *J. Statist. Comp. Simul.*, **14**, 295–303.

Laupacis, A., Sackett. D. L., and Roberts, R. S. (1988). As Assessment of Clinically Useful Measures of the Consequences of Treatment. *The New England Journal of Medicine*, **318**, 1728–33.

Lawless. J. F. (1982). *Statistical Models and Methods for Lifetime Data.* John Wiley, New York.

Lawley. D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**, 295–303.

Lazar, N. and Mykland, P. A. (1995). Empirical likelihood in the presence of nuisance parameters. Technical report no. 400. Dept. of Statistics, University of Chicago.

Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statistics & Probability Letters*, **25**, 95–104.

Little, R. J. and Rubin, D. B. (1987). *Statistical Analysis of Missing Data.* Wiley, New York.

Louis, T. A. (1982). Finding the observed information matrix using the EM Algorithm. *J. R. Statist. Soc. B*, **44**, 226–33.

Madansky, A. (1965). Approximate confidence limits for the reliability of series and parallel systems. *Technometrics*, **7**, 495–503.

Matthews, D. E. (1988a). Likelihood-based confidence intervals for functions of many parameters. *Biometrika*, **75**, 139–44.

Matthews, D. E. (1988b). Profile Likelihood-based Interval Estimation of the Odds Ratio. Unpublished manuscript.

McCullagh, P. (1987). *Tensor Methods in Statistics.* Chapman and Hall, London.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Chapman and Hall, London.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *J. R. Statist. Soc. B*, **52**, 325–44.

McKnight, B. and Crowley, J. (1984). Tests for differences in tumour incidence based on animal carcinogenesis experiments. *J. Am. Statist. Assoc.*, **79**, 639–48.

McLachlan, G. and Krishnan, T. (1997). *The EM Algorithm and Extensions.* Wiley, New York.

Murphy, S. A. (1995). Likelihood ratio-based confidence intervals in survival analysis. *J. Am. Statist. Assoc.*, **90**, 1399–1405.

Mykland, P. A. (1995). Dual likelihood. *Ann. Statist.*, **23**, 396–421.

NASCET Collaborators (1991). Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *New England Journal of Medicine*, **325**, 445–453.

Navidi, W. (1997). A graphical illustration of the EM algorithm. *The American Statistician*, **51**, 29–31.

Neter, J., Wasserman, W., and Kutner, M. H. (1985). *Applied Linear Statistical Models (2nd ed.)*. Richard Irwin, Homewood.

Nyquist, H. (1991). Restricted estimation of generalized linear models. *Appl. Statist.*, **40**, 133–41.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for single functional. *Biometrika*, **75**, 237–49.

Oxner, R. B., Simmonds, N. J., Gertner, D. J., Nightingale, J. M., and Burnham, W. R. (1992). Controlled trial of endoscopic injection treatment for bleeding from peptic ulcers with visible vessels. *Lancet*, **339**, 966–968.

Protter, M. H. and Morrey, C. B. (1991). *A First Course in Real Analysis*. Springer, New York.

Racine, A., Grieve, A. P., Fluhler, H., and Smith, A. F. M. (1986). Bayesian methods in practice: experiences in the pharmaceutical industry. *Appl. Statist.*, **35**, 93–150.

Rai, S. N. and Matthews, D. E. (1993). Improving the EM algorithm. *Biometrics*, **49**, 587–91.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* Wiley, New York.

Rao, C. R. and Toutenburg, H. (1995). *Linear Models: Least Squares and Alternatives.* Springer, New York.

Silvey, S. D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.*, **30**, 389–407.

Snook, S. T. and Ciriello, V. M. (1991). The design of manual handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, **34**, 1197–1213.

Therneau, T. (1995). *S-Plus Guide to Statistical and Mathematical Analysis.* MathSoft, Seattle.

Thomas, D. R. and Grunkmeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Am. Statist. Assoc.*, **70**, 865–71.

Tindall, B., Barker, S., Donovan, B., Barnes, T., Roberts, J., Kronenberg, C., Gold, J., Penny, R., Cooper, D., *et al.* (1988). Characterisation of the acute clinical illness associated with human immunodeficiency virus infection. *Archives of Internal Medicine*, **148**, 945–949.

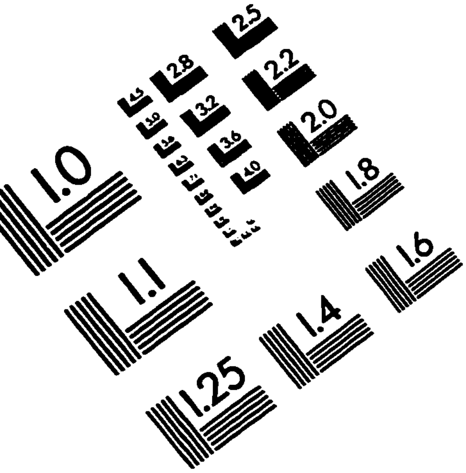
Turnbull, B. and Mitchell, T. (1984). Nonparametric estimation of the distribution of time to onset and time to death for specific disease in survival/sacrifice experiments. *Biometrics*, **40**, 41–50.

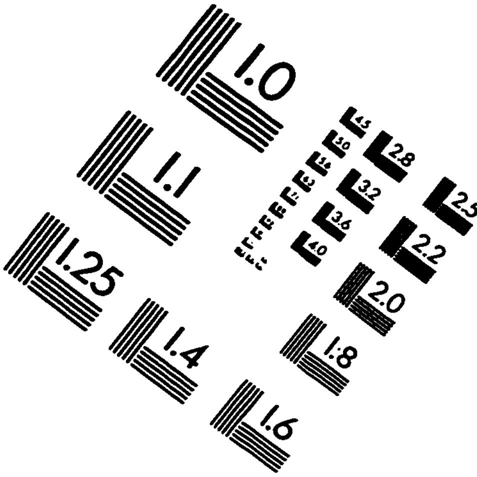
Venables, W. N. and Ripley, B. D. (1994). *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.

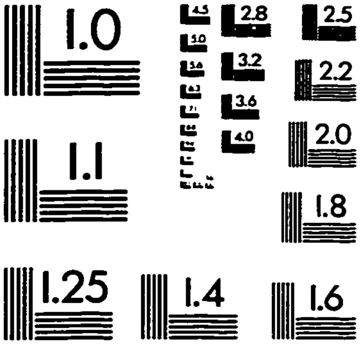
Williams, D. A. (1986). Interval estimation of the median lethal dose. *Biometrics*, **42**, 641–46.

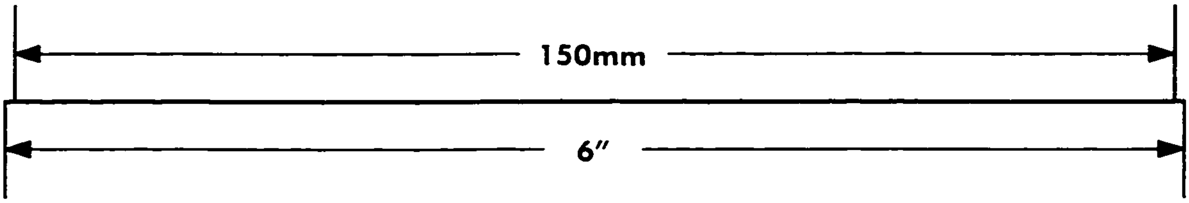
Wu. C. F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.*, **11**, 95–103.

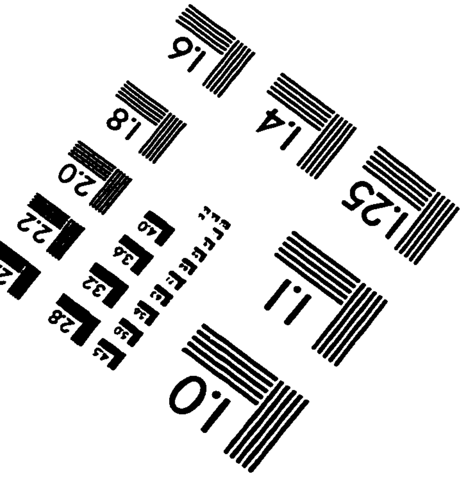
IMAGE EVALUATION TEST TARGET (QA-3)











APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989