MIXED MODEL ANALYSES OF CENSORED NORMAL DISTRIBUTIONS VIA THE EM ALGORITHM

by

Fraser B. Smith and Ronald W. Helms

Department of Biostatistics, University of
North Carolina at Chapel Hill, NC.

Institute of Statistics Mimeo Series No. 1898T

April 1992

MIXED MODEL ANALYSES OF CENSORED NORMAL DISTRIBUTIONS VIA THE EM ALGORITHM

by

Fraser B. Smith

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements of the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill

1992

Approved by:

________________ Advisor

________________ Reader

________________ Reader

________________ Reader

________________ Reader

ABSTRACT

FRASER B. SMITH. Mixed Model Analyses of Censored Normal Distributions via the EM Algorithm. (Under the direction of Ronald W. Helms.)

The analysis of censored data from repeated measures and crossover studies is a frequently occurring problem. The purpose of this work is to develop a method to estimate parameters in general linear mixed models with fixed censoring, noninformative random censoring, or informative random censoring. The proposed method is an extension of maximum likelihood estimation and is applicable to normal data from longitudinal studies where the effects of serial correlation are negligible. Current methods dealing with such topics are limited in that no methods are available to address parameter estimation in general linear mixed models with nonterminal informative censoring and the methods developed for fixed and noninformatively censored data are computationally infeasible.

General convergence properties of the EM algorithm described in Cox and Oakes (1984) for general linear univariate models are discussed. Cox and Oakes restricted their discussion to fixed censoring where censoring values were considered to be predetermined constants. These results are extended to the case of general linear univariate models with noninformative and informative random censoring.

Subsequently this approach is extended so that it can be used for parameter estimation in general linear mixed models. Unlike previous approaches, this method has the advantages of not requiring computations of high-dimensional integrals, not requiring the inversion of large matrices, and is not restricted to random intercept models or studies with noninformative or fixed censoring. This method is applied to data from a placebo-controlled, double-blind crossover, dose-ranging study to assess the short-term efficacy of an antianginal drug in patients with chronic stable angina. Censoring was informative and nonterminal, i.e., was not due to death or withdrawal from the study.

ACKNOWLEDGEMENTS

I gratefully acknowledge my dissertation advisor, Dr. Ron Helms, for his encouragement, guidance, and support, and for the numerous hours he contributed to this research. I also thank my other committee members, Dr. Gerardo Heiss, Dr. Jim Hosking, Dr. Larry Kupper, and Dr. Paul Stewart, and acknowledge Dr. Bahjat Qaqish for his suggestions and comments.

I express my gratitude also to Dr. Bernard Chaitman of the St. Louis University Medical Center for allowing me to use his data and to Ms. Margery Cruise and Dr. David Frankel, who introduced me to the Chaitman data while I was working at Miles Canada in 1986. Finally, I express my deep appreciation to my family for their encouragement and financial support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

The analysis of incomplete data from repeated measures and crossover studies is a frequently occurring problem. Data can be incomplete due to:

(a) missing observations where no information is available about the response of interest, and

(b) truncation or censoring of the response of interest.

Data can be missing or censored for reasons that are either related or unrelated to the outcome of interest and for reasons that are either planned or accidental. Unavailable data are said to be missing if the experimental protocol specifies that the data are to be collected but, for reasons beyond the control of the investigator, the data could not be obtained. Censored data are a special form of incomplete data in which there is some information about the missing value, namely, that the data if available, would have been outside a specified bound. For example, if a device for measuring blood pressure could not measure blood pressure lower than 30 mm Hg or higher than 200 mm Hg, a blood pressure value outside the interval [30, 200] would result in an (unavailable) censored data value.

Censored data problems are common in follow-up studies. For example, in a clinical trial investigating the efficacy of a new treatment for lung cancer the event of interest could be the survival time of lung cancer patients. Right-censoring occurs when patients withdraw from the study, die from another cause, or when they are alive at the end of the study. Note that this is an example of terminal censoring where censoring is due to death or withdrawal from the study and patients are no longer available for subsequent observation.

Several categories of censoring are important in the context of follow-up studies: fixed censoring, noninformative random censoring, and informative random censoring. Fixed censoring occurs when the termination of follow-up for each individual is predetermined in advance. Consequently the censoring values (times of termination of follow-up) are predetermined constants and censoring occurs if the event of interest does not occur prior to the scheduled end of follow-up. Note that censored data values are typically survival times but in some cases the "schedule" is based on a variable other than time. For example, in a dose-response study of the dose required to obtain a 20% reduction in pulmonary function, censoring typically occurs after a patient reaches a predetermined maximum dose.

Random censoring occurs when the termination of follow-up for each individual is not predetermined and observation is terminated by a randomly occurring event prior to the occurrence of the event of interest. The censoring value is the time to censoring. For example, in a follow-up study of cancer patients, the censoring time could be the observation time recorded until observation is terminated because the patient withdraws from the study, as for example, due to death from a cause other than cancer.

Random censoring is noninformative when the failure values and censoring values are stochastically independent and their distributions depend on two sets of functionally independent parameters. If the failure values and censoring values are correlated and/or depend on the same parameters, then the censoring is informative: the value of the censoring variable carries information about the distribution of the time-to-failure distribution. In the previous example, censoring due to withdrawal from the study is informative if patients withdrew because they were not doing well because of an effect of the treatment or if the probability of dying from competing causes depends upon whether the subject has cancer.

Data that are missing but not censored can be missing at random, missing completely at random, or not missing at random. A missing data value is missing at random if the probability of response is independent of the outcome variable of interest that would have been observed if the value were not missing (Little and Rubin 1987, p. 13). A missing data value is missing completely at random if it is missing at random and if the probability of response is independent of any predictor variables of the outcome variable of interest.

For the purpose of making likelihood-based inferences, the missing-data mechanism is ignorable (Little and Rubin 1987, p. 15) if the missing data values are missing at random. Some examples of ignorably missing data might be: data that was missing due to laboratory errors, to unrelated illness, or because the subject moved out of town. Fixed left censored data are clearly not ignorable since the data values are missing when they fall below a known threshold. Therefore analysis of a reduced sample excluding the censored data is subject to bias.

To illustrate the difference between missing and censored data, consider an example in Wei, Lin, and Weissfeld (1989) of a randomized clinical trial to evaluate the effectiveness of ribavirin, a drug used to treat AIDS patients. Patients were randomized into one of three treatment groups: placebo, low-dose ribavirin and high-dose ribavirin. Blood samples for each patient were collected after four, eight, and twelve weeks of treatment. Measurements of p24 antigen levels, important markers of HIV-1 infection were repeatedly taken for a period of four weeks. The response of interest was the "viral load" in each blood sample, which was the number of days until virus positivity (p24>100 picograms/ml) was detected. Ideally each patient in the study would have had three such event times. However some observations were missing because patients did not make the scheduled number of visits or because serum specimens were inadequate for laboratory analysis. Censoring occurred when the culture required a longer period of time to register as virus positive than was achievable in the laboratory or when the serum sample was contaminated during the assay procedure before virus positivity was detected. The authors assumed that censoring was noninformative and that missing data were missing at random. Censoring was <u>nonterminal</u> as it was possible to obtain further blood samples after censoring occurred.

A classic approach used to analyze incomplete longitudinal data is to delete the entire observation vector (row of $\underset{\sim}{Y}$) for any subject with missing or censored data and to use multivariate techniques. This process is known as "casewise" deletion (Timm 1970). When the proportion of subjects with incomplete data is high a great deal of information is lost. When the data are not missing at random this can also be a source of bias. Casewise deletion is still practiced by popular software for multivariate model analysis as, for example, SAS PROC GLM.

Another commonly used approach is to delete observations with censored data and to analyze the remaining data using mixed model techniques which accommodate uncensored missing and mistimed data and covariates that change over time. Using this approach it is no longer necessary to delete the entire observation vector whenever the data are incomplete. However the deletion of censored observations still results in a loss of information and can lead to biased results if censoring is informative.

An alternate approach is to impute missing or censored data values. For example, Lam, Chaitman, Crean, Blum and Waters (1985) conducted a placebo-controlled, double-blind crossover, dose-ranging study to assess the duration and extent of antianginal effects of Nisoldipine in patients with stable angina pectoris. Efficacy was determined by assessing the results of treadmill exercise tolerance tests, in which the time to onset of angina was the primary response of interest. In some cases, right censoring occurred when patients became exhausted and had to stop running on the treadmill before they got angina. Instead of excluding the censored data, peak exercise duration was used to calculate mean exercise time to angina. However, if censoring occurred peak exercise time underestimated the time to angina. Therefore, if the treatment was effective, this approach was conservative because it underestimated the duration of the drug's antianginal effects. However because peak exercise duration time and time to onset of angina were highly correlated, it is plausible that individuals were censored when they were at unusually high risk of failure (i.e., censoring was informative). Therefore it may have been preferable to impute censored data values rather than delete these observations entirely.

Note that in this example the censoring was nonterminal: censoring was not due to death or withdrawal from the study and censored patients returned for subsequent treatments. If patients died because of unrelated illnesses or withdrew from the study, data for subsequent visits were assumed to be missing at random.

Previous work by Wu and Carroll (1988), Wu and Bailey (1988, 1989), and

Schluchter (1991) considered a specific case of terminal informative censoring where there was interest in comparing rates of change of a series of measurements of a single continuous response variable (e.g., one-second forced expiratory volume, tumor growth, decline in renal function) between two treatment groups in a longitudinal study. Each individual received only one treatment. Right censoring caused by death or withdrawal made any subsequent measurements impossible. For example, when steeper slopes were correlated with longer periods of observation this was symptomatic of informative censoring.

These techniques are not applicable to the crossover example because:

(i) censoring was not due to death or withdrawal,

(ii) censoring did not affect subsequent measurements, and

(iii) each patient received multiple treatments.

In addition, Lam et al. (1985) were interested in the main effects of treatment, period and sequence at one or more time intervals rather than comparing rates of change over time between treatments. Problems of this type involving nonterminal censoring will be considered in this dissertation.

Maximum likelihood estimates for general linear models with incomplete data frequently cannot be obtained analytically. Instead it is usually necessary to use iterative procedures. Literature pertaining to the use of these procedures will be summarized chronologically in Section 1.2. Likelihood functions for general linear univariate models with noninformative right censoring are derived in Section 1.2.1. The behavior of the EM algorithm and theory applicable to general linear univariate models with fixed right censoring will be reviewed in Section 1.2.2, followed by a review of the literature pertaining to mixed models with noninformative censoring in Section 1.2.3. Two papers will be discussed in detail. Pettitt's (1985) paper used a frequentist approach in conjunction with the EM

Algorithm to obtain parameter estimates in mixed models with noninformative right censoring while Carriquiry, Gianola, and Fernando (1987) used a Bayesian approach in conjunction with the Newton-Raphson algorithm to obtain parameter estimates in random intercept models with fixed left censoring. The discussion will highlight computational problems associated with these approaches. In many problems computations using these approaches are difficult or intractable involving high-dimensional integration or the inversion of large matrices. Finally after reviewing the existing literature, the objectives of this research will be outlined in Section 1.3.

In order to remain clearly focussed, this dissertation will deal specifically with the use of the general linear mixed model to obtain parameter estimates for normal or lognormal data containing censored observations. (Normal distribution theory can be used for both distributions if logarithms of the response are used instead of actual data values when the dependent variable is lognormally distributed.) The purpose of this work is to simplify existing computational approaches used to estimate parameters in mixed models with fixed or noninformative random censoring and to extend these techniques to parameter estimation in mixed models with informative censoring (e.g., data from the crossover study by Lam et al. 1985). It will be assumed that correlations between measurements within an individual are not dependent on time between measurements. This is a reasonable first approach for longitudinal studies where the effects of serial correlations are negligible.

Cox and Oakes' (1984) application of the EM algorithm to data from regular exponential families with fixed censoring and discussion of general convergence properties is reviewed in Section 1.2.2 and, as part of this research, is extended to random noninformative and informative censoring in Chapter 2 with emphasis on the general linear univariate model. Applications of the EM algorithm for parameter estimation in mixed models with noninformative and informative right censoring are discussed in detail in Chapter 3.

1.2 Literature Review

Several papers have been written outlining parametric methods for the analysis of univariate normal or lognormal failure time data. Sampford and Taylor (1959) developed an iterative procedure to obtain parameter estimates for right censored data from randomized block experiments. When censoring occurred, the conditional expected value of the dependent variable was substituted for the unknown value in the usual maximum likelihood formulae for complete data. Wolynetz (1974) examined the problem of making statistical inferences from normally distributed Type I right censored data. Sampford and Taylor's (1959) method was found to be an efficient procedure for finding maximum likelihood estimates.

After reading Dempster, Laird, and Rubin's (1977) paper, Wolynetz (1979a, b) wrote a FORTRAN program using the EM algorithm to compute maximum likelihood estimates in linear models with censored normal data and normal data confined between two finite limits. Wolynetz (1979a, b) also used the EM algorithm to obtain maximum likelihood estimation techniques for grouped normal data, i.e., where for $i=1, \ldots, m$, Y_i is known but for $i=m+1, \ldots, n$, Y_i is only known to lie between two constants, a_i and b_i. Similarly, Swan (1969a, b, 1977) obtained maximum likelihood estimates for grouped normal data using the Newton-Raphson algorithm.

Subsequently Wolynetz and Binns (1983) reanalyzed dairy cattle survival data using Weibull and lognormal distributions after an inconsistency in published results was attributed to the authors' incorrect assumption that an exponential distribution fit the data. The choice of an exponential distribution was probably made because, as Breslow (1974) noted, researchers often prefer to use other parametric distributions such as the exponential, Weibull, and Gompertz to fit survival data because they are perceived to be mathematically more tractable and conceptually and computationally simpler than the normal or log normal distribution.

Schmee and Hahn (1979) and Chatterjee and McLeish (1986) proposed iterative least squares procedures similar to the method proposed by Sampford and Taylor (1959) whereby censored observations were replaced by their conditional expectations given current parameter estimates. Following Schmee and Hahn's suggestion, Aitkin (1981) outlined a computational procedure used for maximum likelihood estimation using the EM algorithm and compared variance estimators obtained by both methods.

Other parametric distributions (e.g., the exponential distribution), semi-parametric distributions (e.g., the Cox proportional hazards model) and nonparametric procedures have been proposed in the literature for analyzing univariate survival data and are too numerous to review here. [See, for example, Elandt-Johnson and Johnson (1980).] Attempts have also been made to analyze correlated failure time data with noninformative censoring using these techniques. For example, Wei, Lin, and Weissfeld (1989) proposed a semiparametric method for the analysis of incomplete failure time data that used the Cox proportional hazards model to formulate marginal distributions of failure times and estimate regression parameters in the Cox models by maximizing failure-specific partial likelihoods. No specific structure of dependence among the distinct failure times for each subject was imposed.

Bissette, Carr, Koch, Adams, and Sheps (1986) used weighted least squares methods to analyze incidence density rates from two-period crossover studies. The incidence density rates were defined as

$$\hat{\lambda}=\frac{\text{number of people experiencing the event}}{\text{total time at risk}}$$

where time to event is the maximum likelihood estimator for the hazard (scale) parameter when time to event data have an exponential distribution.

Only a few papers have been written about the use of mixed model techniques to analyze correlated survival data, perhaps due to the perception that other methods were mathematically and computationally more tractable. These include Pettitt's (1985) paper that used the EM algorithm to analyze data from mixed models with noninformative right censoring and work by Carriquiry (1985) and Carriquiry, Gianola, and Fernando (1987) that used a Bayesian approach to estimate fixed effects and variance components for random intercept models with fixed left censoring.

Papers that are relevant to this dissertation will be discussed in subsequent sections. Cox and Oakes' (1984) application of the EM algorithm to data from regular exponential families with fixed censoring and discussion of general convergence properties is reviewed in Section 1.2.2 and extended to random noninformative and informative censoring in Chapter 2. Section 1.2.3 reviews Pettitt's (1985) paper and work by Carriquiry (1985) and Carriquiry, Gianola, and Fernando (1987).

1.2.1 Likelihood Functions: General Linear Univariate Models with Noninformative Right Censoring

Consider a random sample of K individuals with 1 observation per subject from a normal population with common parameters $\underset{\sim}{\beta}$ and σ_e^2. The General Linear Univariate Model is

$\underset{\sim}{Y}^* = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e}$,

where

$\underset{\sim}{Y}^*$ is a K $\times$ 1 vector of failure values which may or may not be observed,

$\underset{\sim}{X}$ is a K $\times$ p known constant matrix of rank $r \leq p$,

$\underset{\sim}{\beta}$ is a p $\times$ 1 vector of unknown constant 'fixed' population parameters,

$\underset{\sim}{e}$ is a K x 1 vector of unobservable random errors,

$\underset{\sim}{e} \sim N(\underset{\sim}{0}, \underset{\sim}{I}_K \sigma_e^2)$,

and σ_e^2 is an unknown within-subject variance component.

Therefore $\underset{\sim}{Y}^* \sim N(\underset{\sim}{X}\underset{\sim}{\beta}, \underset{\sim}{I}\,\sigma_e^2)$

with density

$$f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\beta}, \sigma_e^2) = \left[\frac{1}{2\pi\sigma_e^2}\right]^{\frac{K}{2}} \times \exp\left[-\frac{1}{2}(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{I}\,\sigma_e^2)^{-1}(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta})\right]$$

and log likelihood of the complete data $\underset{\sim}{Y}^*$

$$l_o(\underset{\sim}{\beta}, \sigma_e^2 \mid \underset{\sim}{Y}^*) = \log f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\beta}, \sigma_e^2)$$

$$= -\frac{K}{2}\left[\log(2\pi) + \log(\sigma_e^2)\right] - \frac{1}{2}\left[(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{I}\,\sigma_e^2)^{-1}(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta})\right].$$

Note that $\underset{\sim}{Y}^*$ denotes the complete data vector in the absence of censoring.

Similarly let $\underset{\sim}{C}^* = \underset{\sim}{H}\underset{\sim}{\alpha} + \underset{\sim}{\epsilon}$

where

$\underset{\sim}{C}^*$ is a K $\times$ 1 vector of censoring values which may or may not be observed,

$\underset{\sim}{H}$ is a K $\times$ p_h known constant matrix of rank $r_h \leq p_h$,

$\underset{\sim}{\alpha}$ is a p_h $\times$ 1 vector of unknown constant 'fixed' population parameters,

$\underset{\sim}{\epsilon}$ is a K $\times$ 1 vector of unobservable random errors,

$\underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \underset{\sim}{I}_K \sigma_\epsilon^2)$,

and σ_ϵ^2 is an unknown within-subject variance component.

Therefore $\underset{\sim}{C}^* \sim N(\underset{\sim}{H}\underset{\sim}{\alpha}, \underset{\sim}{I}\,\sigma_\epsilon^2)$

with density

$$g_o(\underset{\sim}{C}^* \mid \underset{\sim}{\alpha}, \sigma_\epsilon^2) = \left[\frac{1}{2\pi\sigma_\epsilon^2}\right]^{\frac{K}{2}} \times \exp\left[-\frac{1}{2}(\underset{\sim}{C}^* - \underset{\sim}{H}\underset{\sim}{\alpha})'(\underset{\sim}{I}\,\sigma_\epsilon^2)^{-1}(\underset{\sim}{C}^* - \underset{\sim}{H}\underset{\sim}{\alpha})\right]$$

and log likelihood

$$l_o(\underset{\sim}{\alpha}, \sigma_\epsilon^2 \mid \underset{\sim}{C}^*) = \log g_o(\underset{\sim}{C}^* \mid \underset{\sim}{\alpha}, \sigma_\epsilon^2)$$

$$= -\frac{K}{2}[\log(2\pi) + \log(\sigma_\epsilon^2)] - \frac{1}{2}[(\underset{\sim}{C}^* - \underset{\sim}{H}\underset{\sim}{\alpha})'(\underset{\sim}{I}\,\sigma_\epsilon^2)^{-1}(\underset{\sim}{C}^* - \underset{\sim}{H}\underset{\sim}{\alpha})].$$

Define: $Y_i = \min(Y_i^*, C_i^*)$

$$\delta_i = \mathfrak{B}(Y_i = Y_i^*) = \begin{cases} 1 \text{ if } Y_i = Y_i^* \\ 0 \text{ otherwise} \end{cases}$$

where $\mathfrak{B}$ is the Boolean function (Helms 1988). Observations Y_i for which $\delta_i=0$ are called censored values and observations for which $\delta_i=1$ are called uncensored values or failures (i.e., $Y_i=Y_i^*$ when $\delta_i=1$).

Let Φ denote the cumulative distribution function of $Z_i^* = \dfrac{Y_i^* - \underset{\sim}{X}_i\underset{\sim}{\beta}}{\sigma_e}$

and G denote the survival distribution function of C_i^*.

Theorem 1.2.1: Assuming that

1. $\underset{\sim}{Y}^*$ and $\underset{\sim}{C}^*$ and independent and
2. Parameters of the distribution of $\underset{\sim}{Y}^*$ are functionally independent of the parameters of the distribution of $\underset{\sim}{C}^*$

then the likelihood used to obtain maximum likelihood estimates of $\underset{\sim}{\beta}$ and σ_e^2 is

$$L(\underset{\sim}{\beta}, \sigma_e^2 \mid \underset{\sim}{Y}) \propto \prod_{i=1}^{K} [\, f_{Y^*}(y_i \mid \underset{\sim}{\beta}, \sigma_e^2)]^{\delta_i} \left[1 - \Phi\left(\frac{y_i - \underset{\sim}{X}_i \underset{\sim}{\beta}}{\sigma_e} \right)\right]^{1-\delta_i}$$

Proof: [This is a greatly expanded version of a proof by Lawless (1982, pp. 37-38). This likelihood function is also derived in Kalbfleisch and Prentice (1980).]

The mixed p.d.f. of (Y, δ) is

$$f_{Y,\,\delta}(y, \delta{=}1) = \lim_{\Delta y \to 0^+} \frac{P(y \le Y \le y+\Delta y,\ \delta{=}1)}{\Delta y}$$

$$= \lim_{\Delta y \to 0^+} \frac{P(y \le Y^* \le y+\Delta y,\ C^*>Y^*)}{\Delta y}$$

$$= \lim_{\Delta y \to 0^+} \frac{P[\ (y \le Y^* \le y+\Delta y) \cap \{\ (C^*>y+\Delta y) \cup (Y^*<C^* \le y+\Delta y)\ \}\ ]}{\Delta y}$$

$$= \lim_{\Delta y \to 0^+} \frac{1}{\Delta y} P[\ \{\ (y \le Y^* \le y+\Delta y) \cap (C^*>y+\Delta y)\ \}$$

$$\cup \{\ (y \le Y^* \le y+\Delta y) \cap (Y^*<C^* \le y+\Delta y)\ \}\ ]$$

$$= \lim_{\Delta y \to 0^+} \frac{1}{\Delta y} P[\ (y \le Y^* \le y+\Delta y) \cap (C^*>y+\Delta y)\ ]$$

$$+ \lim_{\Delta y \to 0^+} \frac{1}{\Delta y} P\ [\ (y \le Y^* \le y+\Delta y) \cap (Y^*<C^* \le y+\Delta y)\ ]\ .$$

$$- \lim_{\Delta y \to 0^+} \frac{1}{\Delta y} P\left[(y \le Y^* \le y+\Delta y) \cap (C^*>y+\Delta y) \cap (Y^*<C^* \le y+\Delta y) \right].$$

However $P[(y \le Y^* \le y+\Delta y) \cap (Y^*<C^* \le y+\Delta y)]$

$$\le P[(y \le Y^* \le y+\Delta y) \cap (y \le C^* \le y+\Delta y)]$$

and $P[(y \le Y^* \le y+\Delta y) \cap (C^*>y+\Delta y) \cap (Y^*<C^* \le y+\Delta y)]$

$$= P[(y \le Y^* \le y+\Delta y) \cap (y+\Delta y<C^* \le y+\Delta y)]=0.$$

Since Y_i^* is independent of C_i^*

$$\lim_{\Delta y \to 0^+} \frac{1}{\Delta y} P[(y \le Y^* \le y+\Delta y) \cap (C^*>y+\Delta y)]$$

$$= \lim_{\Delta y \to 0^+} \frac{P(y \le Y^* \le y+\Delta y)}{\Delta y} P(C^*>y)$$

$$= f_{Y^*}(y) P(C_i^*>y) = f_{Y^*}(y) G_{C_i^*}(y)$$

and $\lim_{\Delta y \to 0^+} \frac{1}{\Delta y} P[(y \le Y^* \le y+\Delta y) \cap (Y^*<C^* \le y+\Delta y)]$

$$\le \lim_{\Delta y \to 0^+} \frac{1}{\Delta y} P[y \le Y^* \le y+\Delta y] P[y \le C^* \le y+\Delta y]$$

$$= f_{Y^*}(y) \times 0 = 0.$$

Therefore $f_{Y,\delta}(y, \delta=1) = f_{Y^*}(y) G_{C_i^*}(y)$.

Similarly

$$f_{Y,\delta}(y, \delta=0) = g_{C^*}(y) P(Y^*>y)$$

$=g_{C^*}(y)\ [\ 1\text{-}\Phi(z)\ ]$

where $z_i = \frac{y_i - \underset{\sim}{X}_i \underset{\sim}{\beta}}{\sigma_e}$.

Therefore the joint density of Y and δ is

$$f_{Y,\delta}(y, \delta) = [\ f_{Y^*}(y)\ \ G_{C^*}(y)\ ]^{\delta}\ \ [g_{C^*}(y)\ \{1 - \Phi(z)\ ]^{1-\delta},$$
$$\delta \in \{0, 1\},\ -\infty \le y \le \infty$$

and the joint density of the sampling distribution of (Y_i, δ_i) is

$$\prod_{i=1}^{K} [\ f_{Y^*}(y_i)\ \ G_{C^*}(y_i)\ ]^{\delta_i}\ \ [g_{C^*}(y_i)\ \{1 - \Phi(\ z_i\}\ ]^{1-\delta_i}.$$

If $G(y_i)$ does not involve $\underset{\sim}{\beta}$ or σ_e^2 censoring is noninformative, terms involving g and G can be neglected, and the likelihood is

$$L(\underset{\sim}{\beta}, \sigma_e^2 |\ \underset{\sim}{Y}) \propto \prod_{i=1}^{K} [\ f_{Y^*}(y_i \mid \underset{\sim}{\beta}, \sigma_e^2)]^{\delta_i} \left[1 - \Phi\!\left(\frac{y_i - \underset{\sim}{X}_i \underset{\sim}{\beta}}{\sigma_e}\right)\right]^{1-\delta_i}.$$

Q.E.D.

Type I censoring, i.e., when $C_i^* = y_i = c_i$, a predetermined fixed constant can be considered to be a special case in which each C_i^* has a different degenerate distribution with probability mass at the fixed point $y_i = c_i$ (Lawless 1982, pg. 38). This is because censoring is noninformative and therefore terms involving g and G, whether they are fixed or random, do not involve the parameters of interest.

1.2.2 EM Algorithm

The theory behind the EM algorithm for regular exponential families is given in Dempster et al. (1977). This will be discussed in Section 1.2.2.1. Subsequently in Section 1.2.2.2 general convergence properties of the EM algorithm described in Cox and Oakes (1984) will be discussed. Finally Sections 1.2.2.3 and 1.2.2.4 will focus specifically on applications of the EM algorithm to right censored survival data.

1.2.2.1 EM Algorithm for Regular Exponential Families

The distribution of the complete-data vector, $\underset{\sim}{Y}^* \sim N(\underset{\sim}{X}\underset{\sim}{\beta}, \underset{\sim}{I}\sigma_e^2)$ with density

$$f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\beta}, \sigma_e^2) = \left[\frac{1}{2\pi\sigma_e^2}\right]^{\frac{K}{2}} \times \exp\left[-\frac{1}{2}(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{I}\sigma_e^2)^{-1}(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta})\right]$$

and log likelihood

$$l_o(\underset{\sim}{\beta}, \sigma_e^2 \mid \underset{\sim}{Y}^*) = \log f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\beta}, \sigma_e^2)$$

$$= -\frac{K}{2}[\log(2\pi) + \log(\sigma_e^2)] - \frac{1}{2}[(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{I}\sigma_e^2)^{-1}(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta})]$$

is a member of the exponential class of distributions. The density has the regular exponential-family form

$$f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\theta}) = b(\underset{\sim}{Y}^*) \exp[\underset{\sim}{\theta}' t(\underset{\sim}{Y}^*)] / a(\underset{\sim}{\theta}) \qquad (1.2.1)$$

where $\underset{\sim}{\theta} = \{\underset{\sim}{\beta}, \sigma_e^2\}$ denotes the parameter vector that is restricted to a (p+1)-dimensional convex set $\underset{\sim}{\Psi}$ such that (1.2.1) defines a density for all $\underset{\sim}{\theta}$ in $\underset{\sim}{\Psi}$ and

$$a(\underset{\sim}{\theta}) = \int_{\mathcal{Y}^*} b(\underset{\sim}{Y}^*) \exp[\underset{\sim}{\theta}' t(\underset{\sim}{Y}^*)]\, d\underset{\sim}{Y}^*$$

where $\mathcal{Y}^*$ denotes the set of every possible value of the random variable Y^*; i.e. $\mathcal{Y}^*$ is the sample space of $\underset{\sim}{Y}^*$ (Dempster et al. 1977, p. 1).

For a given $\underset{\sim}{Y}^*$, maximizing

$$l_o(\underset{\sim}{\theta}\,|\underset{\sim}{Y}^*)=\log f_o(\underset{\sim}{Y}^*|\,\underset{\sim}{\theta}\,)=-\log a(\underset{\sim}{\theta}\,)+\log b(\underset{\sim}{Y}^*)+\underset{\sim}{\theta}'\,t(\underset{\sim}{Y}^*)$$

is equivalent to maximizing $-\log a(\underset{\sim}{\theta}\,)+\underset{\sim}{\theta}'\,t(\underset{\sim}{Y}^*)$.

The log likelihood of the incomplete (i.e., observed) data can be obtained in the form

$$l(\underset{\sim}{\theta}\,|\underset{\sim}{Y}\,)=\log f(\underset{\sim}{Y}\,|\,\underset{\sim}{\theta}\,)$$

where Dempster et al. (1977) define the marginal density of the observed data as

$$f(\underset{\sim}{Y}\,|\,\underset{\sim}{\theta}\,)=\int_{\mathcal{Y}^*(\underset{\sim}{Y})} f_o(\underset{\sim}{Y}^*\,|\,\underset{\sim}{\theta}\,)\,d\underset{\sim}{Y}^*$$

where the $\underset{\sim}{Y}$ are a subset of the sample space $\mathcal{Y}^*$ and the corresponding $\underset{\sim}{Y}^*$ in $\mathcal{Y}^*$ are not observed directly but only indirectly through $\underset{\sim}{Y}$. Alternatively, Carriquiry(1985) partitions $\underset{\sim}{Y}^*$ into an observed data vector ($\underset{\sim}{Y}$) and a missing data vector ($\underset{\sim}{M}$) and integrates out the missing data. The marginal density of the observed data is defined as

$$f(\underset{\sim}{Y}\,|\,\underset{\sim}{\theta}\,)=\int_{\underset{\sim}{M}} f_o(\underset{\sim}{Y}^*\,|\,\underset{\sim}{\theta}\,)\,d\underset{\sim}{M}.$$

It is interesting to note that

$$f(\underset{\sim}{Y}|\,\underset{\sim}{\theta}\,)=\frac{1}{a(\underset{\sim}{\theta}\,)}\int_{\mathcal{Y}^*(\underset{\sim}{Y})} b(\underset{\sim}{Y}^*)\exp[\,\underset{\sim}{\theta}'\,t(\underset{\sim}{Y}^*)]\,d\underset{\sim}{Y}^*$$

$$=\frac{a(\underset{\sim}{\theta} \mid \underset{\sim}{Y})}{a(\underset{\sim}{\theta})}$$

and where $a(\underset{\sim}{\theta} \mid \underset{\sim}{Y})= \int_{\mathcal{Y}^*(\underset{\sim}{Y})} b(\underset{\sim}{Y}^*) \exp[\underset{\sim}{\theta}' t(\underset{\sim}{Y}^*)]\, d\underset{\sim}{Y}^*$.

Dempster et al. (1977, equation 2.7) and Carriquiry (1985, equation 4.18) define the conditional density of $\underset{\sim}{Y}^*$ given $\underset{\sim}{Y}$ and $\underset{\sim}{\theta}$ to be

$$m(\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\theta})=\frac{f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\theta})}{f(\underset{\sim}{Y} \mid \underset{\sim}{\theta})}=\frac{b(\underset{\sim}{Y}^*) \exp[\underset{\sim}{\theta}' t(\underset{\sim}{Y}^*)]}{a(\underset{\sim}{\theta} \mid \underset{\sim}{Y})}.$$

Both $f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\theta})$ and $m(\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\theta})$ are from exponential families with the same natural parameters $\underset{\sim}{\theta}$ and the same vector of complete-data sufficient statistics $t(\underset{\sim}{Y}^*)$ but with different sample spaces, where the sample space of $\underset{\sim}{Y}$ is a subspace of $\underset{\sim}{Y}^*$.

Therefore the log likelihood of the observed data

$$\begin{aligned} l(\underset{\sim}{\theta} \mid \underset{\sim}{Y}) &= \log f(\underset{\sim}{Y} \mid \underset{\sim}{\theta}) \\ &= \log f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\theta}) - \log m(\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\theta}) \\ &= -\log a(\underset{\sim}{\theta}) + \log a(\underset{\sim}{\theta} \mid \underset{\sim}{Y}). \end{aligned}$$

Differentiating with respect to $\underset{\sim}{\theta}$,

$$\frac{\partial}{\partial \underset{\sim}{\theta}} l(\underset{\sim}{\theta} \mid \underset{\sim}{Y}) = -\frac{\partial}{\partial \underset{\sim}{\theta}} \log a(\underset{\sim}{\theta}) + \frac{\partial}{\partial \underset{\sim}{\theta}} \log a(\underset{\sim}{\theta} \mid \underset{\sim}{Y})$$

$$= -\frac{1}{a(\underset{\sim}{\theta})} \frac{\partial}{\partial \underset{\sim}{\theta}} a(\underset{\sim}{\theta}) + \frac{1}{a(\underset{\sim}{\theta} \mid \underset{\sim}{Y})} \frac{\partial}{\partial \underset{\sim}{\theta}} a(\underset{\sim}{\theta} \mid \underset{\sim}{Y})$$

$$= -\frac{1}{a(\underset{\sim}{\theta})} \int_{\mathcal{Y}^*} b(\underset{\sim}{Y}^*) \frac{\partial}{\partial \underset{\sim}{\theta}} \exp[\underset{\sim}{\theta}' t(\underset{\sim}{Y}^*)]\, d\underset{\sim}{Y}^*$$

$$+\frac{1}{a(\underset{\sim}{\theta} \mid \underset{\sim}{Y})} \int_{\mathcal{Y}^*(\underset{\sim}{Y})} b(\underset{\sim}{Y}^*) \frac{\partial}{\partial \underset{\sim}{\theta}} \exp[\, \underset{\sim}{\theta}' t(\underset{\sim}{Y}^*)]\, d\underset{\sim}{Y}^*$$

$$= -\frac{1}{a(\underset{\sim}{\theta})} \int_{\mathcal{Y}^*} b(\underset{\sim}{Y}^*) \; \exp[\, \underset{\sim}{\theta}' t(\underset{\sim}{Y}^*)] \; t(\underset{\sim}{Y}^*)\, d\underset{\sim}{Y}^*$$

$$+\frac{1}{a(\underset{\sim}{\theta} \mid \underset{\sim}{Y})} \int_{\mathcal{Y}^*(\underset{\sim}{Y})} b(\underset{\sim}{Y}^*) \; \exp[\, \underset{\sim}{\theta}' t(\underset{\sim}{Y}^*)] \; t(\underset{\sim}{Y}^*)\, d\underset{\sim}{Y}^*$$

$$= -E[\, t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}] + E[\, t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\theta}],$$

where $E[\, t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}]$ is an integral over the whole domain of $\underset{\sim}{Y}^*$ and $E[\, t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\theta}]$ is an integral over the whole domain of the unobserved data. For details see Dempster et al. (1977, pp. 1-5).

As a result the EM algorithm can be expressed in two steps:

E Step: Compute $t^{(r)}(\underset{\sim}{Y}^*) = E[\, t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\theta}^{(r-1)}]$.

M Step: Obtain $\underset{\sim}{\theta}^{(r)}$ as a solution to $E[\, t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}^{(r)}] = t^{(r)}(\underset{\sim}{Y}^*)$

Convergence of $\underset{\sim}{\theta}_1$ to $\hat{\underset{\sim}{\theta}}$, the maximum likelihood estimator of $\underset{\sim}{\theta}$ therefore implies that

$$E[\, t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}; \hat{\underset{\sim}{\theta}}] = E[\, t(\underset{\sim}{Y}^*) \mid \hat{\underset{\sim}{\theta}}]$$

since

$$\frac{\partial}{\partial \underset{\sim}{\theta}} l(\underset{\sim}{\theta})\Big|_{\underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}} = \underset{\sim}{0}.$$

1.2.2.2 Behavior of the EM Algorithm

Cox and Oakes (1984) show how the log likelihood $l_o(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}, \underset{\sim}{\delta})$ never decreases at any iteration of the EM algorithm. Although they specifically use the EM algorithm to obtain maximum likelihood estimates for right censored distributions that are members of regular exponential families, this proof is more general in that it applies to other distributions as well as those from regular exponential families. Their proof uses the function

$Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})=E[l_o(\underset{\sim}{\theta}_1, \underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}]$,

the conditional expectation of the log likelihood of $\underset{\sim}{Y}^*$ given the observed data $(\underset{\sim}{Y}, \underset{\sim}{\delta})$, where $\delta_i=\mathfrak{B}(Y_i=Y_i^*)$ and $\underset{\sim}{Y}$ and $\underset{\sim}{\delta}$ are assumed to be fixed and known. This function was defined in Section 3 of Dempster et al. (1977). Q has two arguments; $\underset{\sim}{\theta}_1$ is an argument of the full likelihood L_o while $\underset{\sim}{\theta}$ is the parameter of the conditional distribution of $\underset{\sim}{Y}^*$ given $(\underset{\sim}{Y}, \underset{\sim}{\delta})$ which is used in computations involving the conditional expectation. The EM algorithm obtains a value $\underset{\sim}{\theta}^*$ that maximizes $f(\underset{\sim}{Y}^*; \underset{\sim}{\theta})$.

Recall from Section 1.2.2.1 that $l(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}, \underset{\sim}{\delta})=\log f(\underset{\sim}{Y}, \underset{\sim}{\delta} \mid \underset{\sim}{\theta}_1)$

and $m(\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}_1)=\dfrac{f_o(\underset{\sim}{Y}^* | \underset{\sim}{\theta}_1)}{f(\underset{\sim}{Y}, \underset{\sim}{\delta} \mid \underset{\sim}{\theta}_1)}$.

Therefore $l(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}, \underset{\sim}{\delta})=\log f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\theta}_1) - \log m(\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}_1)$

$$=l_o(\underset{\sim}{\theta}_1, \underset{\sim}{Y}^*) - l_1(\underset{\sim}{\theta}_1, \underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta})$$

where

$l_1(\underset{\sim}{\theta}_1, \underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta})=\log m(\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}_1)$.

Therefore $l(\underset{\sim}{\theta}_1)=E[\, l_o(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\,] - E[\, l_1(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\,]$

$=Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) - R(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})$,

where $R(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})=E[\, l_1(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\,]$.

Consequently, $l(\underset{\sim}{\theta}_1) - l(\underset{\sim}{\theta})=[Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) - Q(\underset{\sim}{\theta}, \underset{\sim}{\theta})] - [R(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) - R(\underset{\sim}{\theta}, \underset{\sim}{\theta})]$.

Using Lemma 1 from Dempster et al. (1977), for any pair $(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})$ in $\underset{\sim}{\Psi}$ X $\underset{\sim}{\Psi}$, $R(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) \leq R(\underset{\sim}{\theta}, \underset{\sim}{\theta})$. This follows because the expected value of a concave density function is maximized at the true value of the parameter. Because $\underset{\sim}{\theta}_1$ is chosen in the M step to maximize $Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})$ for a previously given value of $\underset{\sim}{\theta}$, $Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) \geq Q(\underset{\sim}{\theta}, \underset{\sim}{\theta})$. Therefore $l(\underset{\sim}{\theta}_1) - l(\underset{\sim}{\theta}) \geq 0$, so each iteration of the EM algorithm cannot decrease the log likelihood function. Note that when the maximum likelihood estimate of $\underset{\sim}{\theta}$ is obtained, the maximum likelihood estimate $\hat{\underset{\sim}{\theta}}$ must satisfy the self-consistency condition $Q(\underset{\sim}{\theta}_1, \hat{\underset{\sim}{\theta}}) \leq Q(\hat{\underset{\sim}{\theta}}, \hat{\underset{\sim}{\theta}})$, and it is therefore impossible to increase the value of the log likelihood function at subsequent iterations. Cox and Oakes (1984, p.165) point out that concavity of l_o with respect to the parameters of interest ($\underset{\sim}{\beta}$ and σ_e^2) does not necessarily imply concavity of l.

1.2.2.3 Derivation of the EM Algorithm; General Linear Univariate Models with Fixed Right Censoring

Cox and Oakes (1984) describe how the EM algorithm can be used to compute maximum likelihood estimates of $\underset{\sim}{\beta}$ and σ_e^2 for right censored data from univariate exponential families. The censoring mechanism they consider is fixed, i.e. $C_i^*=c_i$, a predetermined constant. They make the point that the EM algorithm is very useful when the likelihood of the complete data, $\underset{\sim}{Y}^*$ has a much simpler form than the likelihood of the observed data ($\underset{\sim}{Y}$, $\underset{\sim}{\delta}$), which is the case with the right censored normal data described above.

When $\underset{\sim}{Y}^*$ is a member of the exponential class of distributions,

$f_o(\underset{\sim}{Y}^* | \underset{\sim}{\theta}) = b(\underset{\sim}{Y}^*) \exp [\, \underset{\sim}{\theta}' \, t(\underset{\sim}{Y}^*) \,] \,/\, a(\underset{\sim}{\theta})$ and

$$\frac{\partial l_o(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} = t(\underset{\sim}{Y}^*) - \frac{\partial}{\partial \underset{\sim}{\theta}} \log a(\underset{\sim}{\theta})$$

$$= t(\underset{\sim}{Y}^*) - E[\, t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta} \,].$$

For a right censored sample ($\underset{\sim}{Y}$, $\underset{\sim}{\delta}$) the function

$$\begin{aligned} Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) &= E[l_o(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] \\ &= E[\, \underset{\sim}{\theta}_1' t(\underset{\sim}{Y}^*) - \log a(\underset{\sim}{\theta}_1) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta} \,] + \log b(\underset{\sim}{Y}^*) \\ &= E[\, \underset{\sim}{\theta}_1' t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta} \,] - \log a(\underset{\sim}{\theta}_1) + \log b(\underset{\sim}{Y}^*). \end{aligned}$$

For a concave function l_o, Q is maximized with respect to $\underset{\sim}{\theta}_1$ when

$$0 = \frac{\partial Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}_1} = E[\, t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta} \,] - \frac{\partial \log a(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1}$$

$$=E[\ t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\ ] - E[\ t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}_1]$$

or $$E[\ t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\ ]=E[\ t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}_1]. \tag{1.2.2}$$

The corresponding E and M steps are:

E-step: Compute $\tilde{t}\,(\underset{\sim}{Y}^*)=E[\ t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}\,; \underset{\sim}{\theta}\ ]$

M-step: Obtain $\underset{\sim}{\theta}_1$ as the solution to $E[\ t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}_1]=\tilde{t}\,(\underset{\sim}{Y}^*)$. (1.2.3)

In the r-th iteration these correspond to:

E-step: Compute $t^{(r)}(\underset{\sim}{Y}^*)=E[\ t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}^{(r-1)}]$

M-step: Obtain $\underset{\sim}{\theta}^{(r)}$ as the solution to $E[\ t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}^{(r)}]=t^{(r)}(\underset{\sim}{Y}^*)$.

In the expectation step,

$$E[\ t(Y_i^*) \mid Y_i=y_i,\ \delta_i,\ \underset{\sim}{\theta}\ ]=$$

$$\delta_i E[\ t(Y_i^*) \mid Y_i=y_i,\ \delta_i=1,\ \underset{\sim}{\theta}\ ]+(1-\delta_i)E[\ t(Y_i^*) \mid Y_i=c_i,\ \delta_i=0,\ \underset{\sim}{\theta}\ ]$$

$$=\delta_i E[\ t(Y_i^*) \mid Y_i^*=y_i,\ \underset{\sim}{\theta}\ ]+(1-\delta_i)E[\ t(Y_i^*) \mid Y_i^*>c_i,\ \underset{\sim}{\theta}\ ]$$

$$=\delta_i t(y_i)+(1-\delta_i)E[\ t(Y_i^*) \mid Y_i^*>c_i;\ \underset{\sim}{\theta}\ ]$$

with corresponding density function

$$f_0(Y^* \mid Y^*>c)=\frac{f_o(Y^*)}{\int_c^\infty f_o(Y^*)\ dY^*}=\frac{f_o(Y^*)}{1-\Phi(z)}$$

where $z_i=\frac{c_i-\underset{\sim}{X}_i\underset{\sim}{\beta}}{\sigma_e}$.

Therefore

$E[\, t(Y_i^*) \mid Y_i^* > c_i, \underset{\sim}{\theta} \,]$

$$= \frac{1}{1-\Phi\left(\frac{c_i - \underset{\sim}{X}_i \underset{\sim}{\beta}}{\sigma_e^2}\right)} \int_{c_i}^{\infty} t(Y_i^*)\, f_o(Y_i^* \mid \underset{\sim}{\theta})\, dY_i^*.$$

Recall that the likelihood function for the observed data $(\underset{\sim}{Y}, \underset{\sim}{\delta})$ can be expressed as

$l(\underset{\sim}{\theta}_1 \mid \underset{\sim}{Y}, \underset{\sim}{\delta}) = l_o(\underset{\sim}{\theta}_1, \underset{\sim}{Y}^*) - l_1(\underset{\sim}{\theta}_1, \underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta})$

where l_o is the log likelihood for the complete data $\underset{\sim}{Y}^*$ that would be observed if there was no censoring and l_1 is the log likelihood of the conditional distribution of $\underset{\sim}{Y}^*$ given $(\underset{\sim}{Y}, \underset{\sim}{\delta})$.

When $\underset{\sim}{Y}^* \sim N(\underset{\sim}{X}\underset{\sim}{\beta}, \underset{\sim}{I}\, \sigma_e^2)$,

$$L(\underset{\sim}{\beta}, \sigma_e^2 \mid \underset{\sim}{Y}, \underset{\sim}{\delta}) \propto \prod_{i=1}^{K} [\, f_o(Y_i^* \mid \underset{\sim}{\beta}, \sigma_e^2)]^{\delta_i} \left[1-\Phi\left(\frac{c_i - \underset{\sim}{X}_i \underset{\sim}{\beta}}{\sigma_e}\right)\right]^{1-\delta_i}.$$

Therefore $l(\underset{\sim}{\beta}, \sigma_e^2 \mid \underset{\sim}{Y}, \underset{\sim}{\delta})$

$$\propto \sum_{i=1}^{K} \left\{ \delta_i \log f_o(Y_i^* \mid \underset{\sim}{\beta}, \sigma_e^2) + (1-\delta_i) \log \left[1-\Phi\left(\frac{c_i - \underset{\sim}{X}_i \underset{\sim}{\beta}}{\sigma_e}\right)\right]\right\}$$

$$= \sum_{i=1}^{K} \log f_o(Y_i^* \mid \underset{\sim}{\beta}, \sigma_e^2) - \sum_{i=1}^{K} (1-\delta_i) \log \frac{f_o(Y_i^* \mid \underset{\sim}{\beta}, \sigma_e^2)}{\left[1-\Phi\left(\frac{c_i - \underset{\sim}{X}_i \underset{\sim}{\beta}}{\sigma_e}\right)\right]}$$

where $\dfrac{f_o(Y_i^* \mid \underset{\sim}{\beta}, \sigma_e^2)}{\left[1-\Phi\left(\frac{c_i - \underset{\sim}{X}_i \underset{\sim}{\beta}}{\sigma_e}\right)\right]}$

is the conditional density of Y_i^* given $Y_i^* > c_i$ with the same form as the unconditional density of Y_i^* except that the range of the density is restricted to

$Y_i^* > c_i$.

As before

$$\frac{\partial}{\partial \underset{\sim}{\theta}} \log \frac{f_o(Y_i^* \mid \underset{\sim}{\theta})}{[1 - F(c_i \mid \underset{\sim}{\theta})]} = t(Y_i^*) - E[\, t(Y_i^*) \mid Y_i^* > c_i, \underset{\sim}{\theta}\,].$$

Therefore $\dfrac{\partial l(\underset{\sim}{\theta}\,;\, \underset{\sim}{Y}, \underset{\sim}{\delta})}{\partial \underset{\sim}{\theta}}$

$$= \sum_{i=1}^{K} \{\, t(Y_i^*) - E[\, t(Y_i^*) \mid \underset{\sim}{\theta}\,]\,\}$$

$$- \sum_{i=1}^{K} (1-\delta_i)\{\, t(Y_i^*) - E[\, t(Y_i^*) \mid Y_i^* > c_i, \underset{\sim}{\theta}\,]\,\}$$

$$= \sum_{i=1}^{K} \{\, E[\, t(Y_i^*) \mid \underset{\sim}{Y} = \underset{\sim}{y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\,] - E[\, t(Y_i^*) \mid \underset{\sim}{\theta}\,]\,\}$$

the difference between the conditional and unconditional expectations of the complete-data sufficient statistics and when $\underset{\sim}{\theta}_1 = \underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}$ in (1.2.2) then $\hat{\underset{\sim}{\theta}}$ is also a solution of the

likelihood equation $\dfrac{\partial l(\underset{\sim}{\theta}\,;\, \underset{\sim}{Y}, \underset{\sim}{\delta})}{\partial \underset{\sim}{\theta}} = \underset{\sim}{0}$.

1.2.2.4 EM Computations; General Linear Univariate Models with Fixed Right Censoring

EM Computations are summarized in this section. In the E step, the conditional expected values of the 'complete-data' sufficient statistic are computed from the observed data and current estimates of the parameters, while in the M step, new estimates of the unknown parameters are computed using the conditional expected values of the 'complete-data' sufficient statistics in the maximum likelihood equations.

E Step:

In the r-th iteration, the estimation step computes the conditional expected values of the complete-data sufficient statistics given the observed data $\underset{\sim}{Y}$ and the estimated values of the parameters from the $(r-1)$-st iteration. Not all of the $\underset{\sim}{Y}^*$ are observed so the E step will estimate the complete-data sufficient statistics that involve $\underset{\sim}{Y}^*$.

A set of complete-data sufficient statistics for this problem is

$\underset{\sim}{X}'\underset{\sim}{Y}^*$ and $\underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^*$.

The expected values of the complete-data sufficient statistics may be denoted,

$\underset{\sim}{t}_1^{(r)}=\underset{\sim}{X}'\, E[\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$ and

$t_2^{(r)}=E[\underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$.

When $\delta_i=0$ expectations involving the i-th element of the complete-data sufficient statistics can be computed as:

$E[Y_i^* \mid Y_i=c_i, \delta_i=0, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$

$$=E[Y_i^* \mid Y_i^* > c_i, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=\int_{c_i}^{\infty} Y_i^* \frac{f_o(Y_i^* \mid \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2})}{1-\Phi(z_i^{(r-1)})} dY_i^*$$

$$=\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}+\sigma_e^{(r-1)} \frac{\phi(z_i^{(r-1)})}{1-\Phi(z_i^{(r-1)})}$$

where $z_i^{(r-1)}=\dfrac{c_i-\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}}{\sigma_e^{(r-1)}}$

and

$$E[Y_i^{*2} \mid Y_i=y_i, \delta_i=0, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=E[Y_i^{*2} \mid Y_i^* > c_i, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=\int_{c_i}^{\infty} Y_i^{*2} \frac{f_o(Y_i^* \mid \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2})}{1-\Phi(z_i^{(r-1)})} dY_i^*$$

$$=[\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}]^2+\sigma_e^{(r-1)^2} +\sigma_e^{(r-1)}(c_i+\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}) \frac{\phi(z_i^{(r-1)})}{1-\Phi(z_i^{(r-1)})} .$$

Both of these expectations appear in Aitkin (1981).

M Step:

The r-th iteration of the M step obtains $\underset{\sim}{\beta}^{(r)}$ and $\sigma_e^{(r)^2}$ as solution to equation 1.2.3. For the complete-data problem the maximum likelihood estimates are:

$$\hat{\underset{\sim}{\beta}}=(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}'\underset{\sim}{Y}^*$$

and

$$\hat{\sigma}_e^2 = \frac{1}{K}(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta}).$$

$$= \frac{1}{K}(\underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^* - 2\underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim}{Y}^* + \underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta}).$$

Initial values of maximum likelihood estimates are obtained by treating the censored data as if they were uncensored. Convenient initial estimates are:

$$\underset{\sim}{\beta}^{(0)} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}$$

and

$$\sigma_e^{(0)^2} = \frac{1}{K}(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta}^{(0)})'(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta}^{(0)}).$$

In the r-th iteration maximum likelihood estimates which maximize the likelihood function using the expected values of the complete-data sufficient statistics obtained from the previous iteration of the E step are:

$$\underset{\sim}{\beta}^{(r)} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{t}_1^{(r)}$$

and

$$\sigma_e^{(r)^2} = \frac{1}{K}[t_2^{(r)} - 2\underset{\sim}{\beta}^{(r-1)'}\underset{\sim}{t}_1^{(r)} + \underset{\sim}{\beta}^{(r-1)'}\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)}].$$

1.2.3 Mixed Models with Noninformative Right Censoring or Fixed Left Censoring

Consider a random sample of K individuals with n_i observations for the i-th subject such that $n=\sum_{i=1}^{K} n_i$. The n observations are assumed to be a sample from a normal population with common parameters β, D, and σ_e^2. The General Linear Mixed Model is

$$Y_i^* = X_i\beta + B_i d_i + e_i, \tag{1.2.4}$$

where

Y_i^* is an $n_i \times 1$ vector of failure values which may or may not be observed,

X_i is an $n_i \times p$ known constant matrix of rank $r \le p$,

β is a p x 1 vector of unknown constant 'fixed' population parameters,

B_i is an $n_i \times q$ known matrix corresponding to the random effects,

d_i is a q x 1 vector of unknown individual parameters,

e_i is an n_i x 1 vector of unobservable random errors ,

$d_i \sim N(0, D)$ independent of $e_i \sim N(0, I_{n_i}\sigma_e^2)$,

D is a positive-definite symmetric $q \times q$ covariance matrix of random effects, d_i,

and σ_e^2 is an unknown within-subject variance component.

Therefore $Y_i^* \sim N(X_i\beta, \Sigma_i)$, each with marginal density

$$f_o(Y_i^* \mid \beta, \Sigma_i) = \left[\frac{1}{2\pi}\right]^{\frac{n_i}{2}} |\Sigma_i|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}(Y_i^* - X_i\beta)'(\Sigma_i)^{-1}(Y_i^* - X_i\beta)\right]$$

and log likelihood of the complete data subvector Y_i^*

$l_o(\beta, \Sigma_i \mid Y_i^*) = \log f_o(Y_i^* \mid \beta, \Sigma_i)$

$$= -\frac{n_i}{2}\log(2\pi) - \frac{1}{2}\log|\underset{\sim}{\Sigma}_i| - \frac{1}{2}\,[(\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i\underset{\sim}{\beta})'\,(\underset{\sim}{\Sigma}_i)^{-1}(\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i\underset{\sim}{\beta})]$$

where $\underset{\sim}{\Sigma}_i = \underset{\sim}{B}_i\underset{\sim}{D}\underset{\sim}{B}_i' + \underset{\sim}{I}_{n_i}\sigma_\epsilon^2$

is the positive-definite symmetric covariance matrix of $\underset{\sim}{Y}_i^*$.

Note that $\underset{\sim}{Y}^*$ denotes the complete data vector in the absence of censoring.

Similarly let $\underset{\sim}{C}_i^* = \underset{\sim}{H}_i\underset{\sim}{\alpha} + \underset{\sim}{J}_i\underset{\sim}{v}_i + \underset{\sim}{\epsilon}_i$,

where

$\underset{\sim}{C}_i^*$ is an $n_i \times 1$ vector of censoring values which may or may not be observed,

$\underset{\sim}{H}_i$ is an $n_i \times p_h$ known constant matrix of rank $r_h \leq p_h$,

$\underset{\sim}{\alpha}$ is a $p_h \times 1$ vector of unknown constant 'fixed' population parameters,

$\underset{\sim}{J}_i$ is an $n_i \times q_h$ known matrix corresponding to the random effects,

$\underset{\sim}{v}_i$ is a q_h x 1 vector of unknown individual parameters,

$\underset{\sim}{\epsilon}_i$ is an $n_i \times 1$ vector of unobservable random errors,

$\underset{\sim}{v}_i \sim N(\underset{\sim}{0}, \underset{\sim}{V})$ independent of $\underset{\sim}{\epsilon}_i \sim N(\underset{\sim}{0}, \underset{\sim}{I}_{n_i}\sigma_\epsilon^2)$,

$\underset{\sim}{V}$ is a positive-definite symmetric $q_h \times q_h$ covariance matrix of random effects, $\underset{\sim}{v}_i$,

and σ_ϵ^2 is an unknown within-subject variance component.

Therefore $\underset{\sim}{C}_i^* \sim N(\underset{\sim}{H}_i\underset{\sim}{\alpha}, \underset{\sim}{\Gamma}_i)$

with density

$$g_o(\underset{\sim}{C}_i^* \mid \underset{\sim}{\alpha}, \underset{\sim}{\Gamma}_i) = \left[\frac{1}{2\pi}\right]^{\frac{n_i}{2}} |\underset{\sim}{\Gamma}_i|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}(\underset{\sim}{C}_i^* - \underset{\sim}{H}_i\underset{\sim}{\alpha})'(\underset{\sim}{\Gamma}_i)^{-1}(\underset{\sim}{C}_i^* - \underset{\sim}{H}_i\underset{\sim}{\alpha})\right]$$

and log likelihood

$$l_o(\underset{\sim}{\alpha}, \underset{\sim}{\Gamma}_i \mid \underset{\sim}{C}_i^*) = \log g_o(\underset{\sim}{C}_i^* \mid \underset{\sim}{\alpha}, \underset{\sim}{\Gamma}_i)$$

$$= -\frac{n_i}{2}\log(2\pi) - \frac{1}{2}\log|\underset{\sim}{\Gamma}_i| - \frac{1}{2}\,[(\underset{\sim}{C}_i^* - \underset{\sim}{H}_i\underset{\sim}{\alpha})'\,(\underset{\sim}{\Gamma}_i)^{-1}(\underset{\sim}{C}_i^* - \underset{\sim}{H}_i\underset{\sim}{\alpha})]$$

where $\underset{\sim}{\Gamma}_i = \underset{\sim}{J}_i \underset{\sim}{V} \underset{\sim}{J}_i' + \underset{\sim}{I}_{n_i} \sigma_\epsilon^2$

is the positive-definite symmetric covariance matrix of $\underset{\sim}{C}_i^*$.

Define: $Y_{ij} = \min(Y_{ij}^*, C_{ij}^*)$

$\delta_{ij} = \mathfrak{B}(Y_{ij} = Y_{ij}^*)$

where $\mathfrak{B}$ is the Boolean function (Helms 1988) and let $\underset{\sim}{\delta}_i$ denote the vector of indicator variables for non-right censoring in the i-th individual, where right censoring occurs when the log censoring time C_{ij}^* is less than the log survival time Y_{ij}^*.

Pettitt (1985) assumes that censoring is noninformative, i.e. that

1. $\underset{\sim}{Y}_i^*$ and $\underset{\sim}{C}_i^*$ and independent and
2. Parameters of the distribution of $\underset{\sim}{Y}_i^*$ are functionally independent of the parameters of the distribution of $\underset{\sim}{C}_i^*$

and defines the likelihood used to obtain maximum likelihood estimates of $\underset{\sim}{\beta}$, $\underset{\sim}{D}$, and σ_e^2 to be

$$L(\underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2 \mid \underset{\sim}{Y}) = \prod_{i=1} L_i(\underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2 \mid \underset{\sim}{Y}_i)$$

where

$$L_i(\underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2 \mid \underset{\sim}{Y}_i) \propto \prod_{l \in \mathfrak{C}} \int_{Y_{il}}^{\infty} f_o(\underset{\sim}{Y}_i^* \mid \underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2) \prod_{l \in \mathfrak{C}} dY_{il}^*$$

and $\mathfrak{C}$ denotes the set of missing or right censored observations in $\underset{\sim}{Y}_i^*$. Alternatively, after partitioning $\underset{\sim}{Y}_i^*$ into an observed data vector ($\underset{\sim}{Y}_i$) and a missing data vector ($\underset{\sim}{M}_i$) (Carriquiry 1985), the likelihood function for the i-th individual can be written in more compact notation as

$$L_i(\underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2 \mid \underset{\sim}{Y}_i) \propto \int_{\underset{\sim}{M}_i} f_o(\underset{\sim}{Y}_i^* \mid \underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2) \; d\underset{\sim}{M}_i \, .$$

Pettitt (1985) solved this problem using the EM Algorithm. Using his approach it is necessary to estimate the random effects ($\underset{\sim}{d}_i$) in the E step. As a result the complete-data sufficient statistics involve functions of $\underset{\sim}{Y}_i^*$ and $\underset{\sim}{d}_i$ and expected values of the complete-data sufficient statistics are computed given $\underset{\sim}{\beta}$, $\underset{\sim}{D}$, and σ_e^2 but not $\underset{\sim}{d}_i$.

Recall that $\underset{\sim}{Y}_i^* \sim N(\underset{\sim}{X}_i\underset{\sim}{\beta}, \underset{\sim}{\Sigma}_i)$ where

$$\underset{\sim}{\Sigma}_i = \underset{\sim}{B}_i \underset{\sim}{D} \underset{\sim}{B}_i' + \underset{\sim}{I}_{n_i}\sigma_e^2$$

and consequently the $\underset{\sim}{Y}_i^*$ are not conditionally independent.

Some of the complete-data sufficient statistics involve functions of $\underset{\sim}{Y}_i^*$. Given $\underset{\sim}{\beta}$, $\underset{\sim}{D}$, σ_e^2 the expectations of Y_{ij}^* and $Y_{ij}^* Y_{ik}^*$ are

$$E[Y_{ij}^* \mid Y_{ij}, \delta_{ij}; \underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2] = \frac{\int_{\underset{\sim}{M}_i} Y_{ij}^* \, f(\underset{\sim}{Y}_i^*; \underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2) \, d\underset{\sim}{M}_i}{\int_{\underset{\sim}{M}_i} f(\underset{\sim}{Y}_i^*; \underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2) \, d\underset{\sim}{M}_i} \tag{1.2.5}$$

and

$$E[Y_{ij}^* Y_{ik}^* \mid Y_{ij}, Y_{ik}, \delta_{ij}, \delta_{ik}; \underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2] = \frac{\int_{\underset{\sim}{M}_i} Y_{ij}^* Y_{ik}^* \, f(\underset{\sim}{Y}_i^*; \underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2) \, d\underset{\sim}{M}_i}{\int_{\underset{\sim}{M}_i} f(\underset{\sim}{Y}_i^*; \underset{\sim}{\beta}, \underset{\sim}{D}, \sigma_e^2) \, d\underset{\sim}{M}_i}$$

where the denominator of (1.2.5) is the likelihood of the observed data.

Consequently, using Pettitt's approach, in order to obtain the expected values of the complete-data sufficient statistics which are functions of $\underset{\sim}{Y}_i^*$ it is often necessary to carry out high-dimensional integrations. This is computationally infeasible for many real-data problems.

A mixed model procedure for the analysis of multivariate normal survival data using a Bayesian approach was developed by Carriquiry (1985) and Carriquiry et al. (1987). They described how to estimate fixed effects and variance components for a random intercept model when records were left censored at time c, a predetermined fixed constant. Unlike Pettitt (1985) they estimated the random effects as well as fixed effects. Given a random sample of K individuals with n_i observations for the i'th individual such that $n=\sum_{i=1}^{K} n_i$, the n observations are assumed to be a sample from a normal distribution with common parameters $\underset{\sim}{\beta}$, $\underset{\sim}{D}$, and σ_e^2.

Stacking the $\underset{\sim}{Y}_i^*$ matrices defined in (1.2.4), let

$\underset{\sim}{Y}^* = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{B}\underset{\sim}{d} + \underset{\sim}{e}$,

where

$\underset{\sim}{Y}^*$ is an n x 1 vector of failure values which may or may not be observed,

$\underset{\sim}{X}$ is an n x p known constant matrix of rank $r \leq p$,

$\underset{\sim}{\beta}$ is a p x 1 vector of unknown constant 'fixed' population parameters,

$\underset{\sim}{B}$ is an n x Kq known block-diagonal design matrix corresponding to the random effects,

$\underset{\sim}{d}$ is a Kq x 1 vector of unknown individual parameters,

$\underset{\sim}{e}$ is an n x 1 vector of unobservable random errors,

$\underset{\sim}{d} \sim N(\underset{\sim}{0}, \underset{\sim}{A}\sigma_d^2)$ independent of $\underset{\sim}{e} \sim N(\underset{\sim}{0}, \underset{\sim}{I}\,\sigma_e^2)$,

$\underset{\sim}{A}$ is a positive-definite symmetric Kq x Kq covariance matrix,

and σ_d^2 and σ_e^2 are the unknown between- and within-subject variance components. Therefore

$$\underset{\sim}{Y} \sim N(X\underset{\sim}{\beta}, \underset{\sim}{\Sigma})$$

where

$$\underset{\sim}{\Sigma} = \underset{\sim}{B}\ \underset{\sim}{A}\underset{\sim}{B}'\sigma_d^2 + \underset{\sim}{I}_n\sigma_e^2$$

is the positive-definite symmetric covariance matrix of $\underset{\sim}{Y}$.

For a given individual

$$\underset{\sim}{Y}_i^* = \underset{\sim}{X}_i\underset{\sim}{\beta} + \underset{\sim}{B}_i\ \underset{\sim}{d}_i + \underset{\sim}{\varepsilon}_i, \quad i=1, \ldots, K,$$

Carriquiry et al. (1987) suggest conditioning on both the fixed and random effects so that

$$(\underset{\sim}{Y}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}) \sim N(\underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{B}\underset{\sim}{d}, \underset{\sim}{I}_n\sigma_e^2)$$

and the elements of $(\underset{\sim}{Y}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d})$ are conditionally independent, eliminating the need to compute multi-dimensional integrals.

The likelihood function can be simplified by partitioning $\underset{\sim}{Y}$ into a vector of n_1 uncensored observations and n_2 censored observations,

$$\underset{\sim}{Y}^* = \begin{bmatrix} \underset{\sim}{Y}_1^* \\ \underset{\sim}{Y}_2^* \end{bmatrix}.$$

The likelihood of the conditional distribution of $(\underset{\sim}{Y}_1 \mid \underset{\sim}{\beta}, \underset{\sim}{d})$ is:

$$L[\underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_e^2 \mid (\underset{\sim}{Y}_1 \mid \underset{\sim}{\beta}, \underset{\sim}{d})] = \left[\frac{1}{2\pi\sigma_e^2}\right]^{\frac{n_1}{2}} \exp\left[-\frac{1}{2}\sum_{l=1}^{n_1}\frac{(Y_l - \underset{\sim}{X}_l\underset{\sim}{\beta} - \underset{\sim}{B}_l\ \underset{\sim}{d}_l)^2}{\sigma_e^2}\right] \qquad (1.2.6)$$

while, for censored observations,

$$\Pr(Y_{2l} < c \mid \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_e^2) = \Phi(z_l), \quad l = n_1+1, \ldots, n$$

where

$$z_l = \frac{c - \underset{\sim}{X}_l\underset{\sim}{\beta} - \underset{\sim}{B}_l\ \underset{\sim}{d}_l}{\sigma_e}$$

and

$\Phi(z_l)$ is the probability of the l-th observation being left censored at c.

Therefore

$$\Pr(Y_{2,n_1+1}<c, \ldots, Y_{2,n}<c \mid \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_e^2) = \prod_{l=n_1+1}^{n} \Phi(z_l) \tag{1.2.7}$$

which is the joint probability of all observations in Y_2 being left censored. The product of (1.2.6) and (1.2.7) gives the likelihood function for the whole sample (Carriquiry 1985, Carriquiry et al. 1987):

$$L[\underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_e^2 \mid (\underset{\sim}{Y} \mid \underset{\sim}{\beta}, \underset{\sim}{d}), \underset{\sim}{\delta}] \propto f[\underset{\sim}{Y}, \underset{\sim}{\delta} \mid \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_e^2]$$

$$= \left[\frac{1}{2\pi\sigma_e^2}\right]^{\frac{n_1}{2}} \exp\left[-\frac{1}{2}\sum_{l=1}^{n_1}\frac{(Y_{1l} - \underset{\sim}{X}_l\underset{\sim}{\beta} - \underset{\sim}{B}_l\underset{\sim}{d}_l)^2}{\sigma_e^2}\right] \times \prod_{l=n_1+1}^{n} \Phi(z_l) \tag{1.2.8}$$

<u>Prior</u> <u>Distributions</u>:

Unlike the mixed model approach where $\underset{\sim}{\beta}$ is assumed fixed, the Bayesian approach assumes that all parameters are random variables. Harville (1974, 1976, 1977) shows that restricted maximum likelihood estimation of variance components is equivalent to Bayesian procedures with flat priors for $\underset{\sim}{\beta}$ and the components of σ_d^2 using all the data. Suitable prior distributions for this problem are:

$\Pi_1(\underset{\sim}{\beta}) \propto$ constant,

$\Pi_2(\underset{\sim}{d} \mid \sigma_d^2) = N_{Kq}(\underset{\sim}{0}, \underset{\sim}{A}\sigma_d^2)$, and

$\Pi_3(\sigma_d^2, \sigma_e^2) \propto$ constant.

This leads to the joint prior distribution

$$\Pi(\underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2) \propto \Pi_1(\underset{\sim}{\beta}) \cdot \Pi_2(\underset{\sim}{d} \mid \sigma_d^2) \cdot \Pi_3(\sigma_d^2, \sigma_e^2) \tag{1.2.9}$$

<u>Posterior</u> <u>Distribution</u>:

The joint distribution of $\underset{\sim}{Y}$, $\underset{\sim}{\delta}$, $\underset{\sim}{\beta}$, $\underset{\sim}{d}$, σ_d^2, and σ_e^2 is equal to the product of the conditional likelihood in (1.2.8) and the joint prior distribution in (1.2.9), i.e.

$$p(\underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2) = f(\underset{\sim}{Y}, \underset{\sim}{\delta} \mid \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2) \times \Pi(\underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2) .$$

Using Bayes Theorem

$$p(\underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2 \mid \underset{\sim}{Y}, \underset{\sim}{\delta}) = \frac{p(\underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2)}{\int_{\underset{\sim}{\beta}}\int_{\underset{\sim}{d}}\int_{\sigma_d^2}\int_{\sigma_e^2} p(\underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2)\, d\sigma_e^2\, d\sigma_d^2\, d\underset{\sim}{d}\, d\underset{\sim}{\beta}}$$

$$= \frac{p(\underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2)}{p(\underset{\sim}{Y}, \underset{\sim}{\delta})}.$$

Since $p(\underset{\sim}{Y}, \underset{\sim}{\delta})$ does not depend on any of the parameters, maximizing $p(\underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2 \mid \underset{\sim}{Y}, \underset{\sim}{\delta})$ is equivalent to maximizing $p(\underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2)$ with respect to $\underset{\sim}{\beta}$, $\underset{\sim}{d}$, σ_d^2, or σ_e^2.

Therefore the posterior distribution of $\underset{\sim}{\beta}$, $\underset{\sim}{d}$, σ_d^2, and σ_e^2 is

$$p(\underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2 \mid \underset{\sim}{Y}, \underset{\sim}{\delta})$$

$$= \frac{p(\underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2)}{p(\underset{\sim}{Y}, \underset{\sim}{\delta})}$$

$$\propto p(\underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2)$$

$$= (\sigma_e^2)^{-\frac{n_1}{2}} \left|\sigma_d^2\right|^{-\frac{Kq}{2}}$$

$$\times \exp\left\{-\frac{1}{2}\left[\sum_{l=1}^{n_1} \frac{(y_l - \underset{\sim}{X}_l\underset{\sim}{\beta} - \underset{\sim}{B}_l\underset{\sim}{d}_l)^2}{\sigma_e^2} + \sum_{i=1}^{K} \underset{\sim}{d}_i'(\underset{\sim}{A}\sigma_d^2)^{-1}\underset{\sim}{d}_i\right]\right\}$$

$$\times \prod_{l=n_1+1}^{n} \Phi(z_l). \tag{1.2.10}$$

It is usually not practical to perform the integrations needed to obtain parameter estimates from (1.2.10) directly. Instead Carriquiry (1985) and Carriquiry et al. (1987) developed an iterative approach using the Newton-Raphson algorithm to obtain estimates of fixed and random effects assuming variances were known and then used these to estimate variance components. This involved estimating the posterior modes or maximum a posteriori values of $\underset{\sim}{\beta}$ and $\underset{\sim}{d}$ of the joint posterior distribution in equation 1.2.10 using a procedure that is referred to as "maximum a posteriori" estimation (Beck and Arnold 1977) which can be viewed as a Bayesian extension of maximum likelihood estimation. Posterior mode estimators are equivalent to maximum likelihood estimates for parameters with flat priors (Laird and Ware 1982), but in this case we do not have flat priors for the random effects.

Newton-Raphson algorithm:

Let $L(\underset{\sim}{\beta}, \underset{\sim}{d})=\log p(\underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_d^2, \sigma_e^2 \mid \underset{\sim}{Y}, \underset{\sim}{\delta})$. The Newton-Raphson algorithm is used to obtain estimates of $\underset{\sim}{\beta}$ and $\underset{\sim}{d}$ where

$$\begin{bmatrix} \underset{\sim}{\beta}^{(r+1)} \\ \underset{\sim}{d}^{(r+1)} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\beta}^{(r)} \\ \underset{\sim}{d}^{(r)} \end{bmatrix}$$

$$+ \begin{bmatrix} -\dfrac{\partial L^2(\underset{\sim}{\beta}, \underset{\sim}{d})}{\partial \underset{\sim}{\beta}\, \partial \underset{\sim}{\beta}'} & -\dfrac{\partial L^2(\underset{\sim}{\beta}, \underset{\sim}{d})}{\partial \underset{\sim}{\beta}\, \partial \underset{\sim}{d}'} \\ -\dfrac{\partial L^2(\underset{\sim}{\beta}, \underset{\sim}{d})}{\partial \underset{\sim}{d}\, \partial \underset{\sim}{\beta}'} & -\dfrac{\partial L^2(\underset{\sim}{\beta}, \underset{\sim}{d})}{\partial \underset{\sim}{d}\, \partial \underset{\sim}{d}'} \end{bmatrix}^{-}_{(\underset{\sim}{\beta}=\underset{\sim}{\beta}^{(r)},\ \underset{\sim}{d}=\underset{\sim}{d}^{(r)})}$$

$$\times \begin{bmatrix} \dfrac{\partial L(\underset{\sim}{\beta}, \underset{\sim}{d})}{\partial \underset{\sim}{\beta}} \\ \dfrac{\partial L(\underset{\sim}{\beta}, \underset{\sim}{d})}{\partial \underset{\sim}{d}} \end{bmatrix}_{(\underset{\sim}{\beta} = \underset{\sim}{\beta}^{(r)},\ \underset{\sim}{d} = \underset{\sim}{d}^{(r)})}$$

For large samples, this requires the inversion of very large matrices of the order of the number of subjects in order to obtain estimates of fixed and random effects.

1.3 Statement of the problem and outline

The mixed model procedures described in Section 1.2.3 are summarized in Table 1.3.1. The use of the EM algorithm is often preferable to gradient methods (e.g., Newton-Raphson algorithm, Method of Scoring) when maximum likelihood solutions are more easily obtained for the likelihood of the complete data conditional on the observed data than for the likelihood of the observed data alone. The likelihood of the complete data for the General Linear Mixed Model has a much simpler form than the likelihood corresponding to the General Linear Mixed Model with censored data. Using the EM algorithm, computations involved in obtaining parameter estimates in the M step are straightforward. However, using Pettitt's (1985) approach, computations in the E-step involving estimation of the random effects and censored data can involve high-dimensional integrations when censoring occurs.

This was not a problem for Carriquiry et al. (1987) because they used an extension of maximum likelihood estimation known as maximum a posteriori estimation. However, using this approach in conjunction with the Newton-Raphson algorithm other complications arose because parameter estimates had to be obtained using the likelihood of the observed data. Although they eliminated the difficulty of having to perform high-dimensional integrations this was offset by the fact that their method required the inversion of very large matrices, of the order of the number of subjects, in order to obtain estimates of fixed and random effects. For example, it would be necessary to invert a 41×41 matrix to estimate the fixed and random effects in the example from Wei, Lin, and Weissfeld (1989) described in Section 1.1. In this example, there were only 36 patients. If there had been ten times as many patients, the dimension of the matrix would be 365×365.

Similarly, for 1000 patients, it would be necessary to invert a 1005×1005 matrix. However, in the M-step of the EM algorithm the dimensions of the matrices required to estimate fixed and random effects would always be 5×5 and 1×1, respectively, for each subject regardless how many subjects there were.

Carriquiry, Gianola, and Fernando's (1987) method was also restricted to random intercept models. In fact, if they had attempted to fit a model with a random intercept and a random slope, the dimension of the matrices they would have to invert would be almost double the dimensions for the random intercept model.

The approach taken in this dissertation is to use maximum a posteriori estimation instead of maximum likelihood estimation in order to avoid the problem of having to compute high-dimensional integrals and to use the EM algorithm instead of the Newton-Raphson algorithm. This approach takes full advantage of the simple computational form of the likelihood for the complete data and avoids having to invert large matrices.

Unlike Pettitt's (1985) approach, the EM algorithm for noninformative right censoring proposed in Section 3.1 uses maximum a posteriori estimation to obtain parameter estimates of the random effects in the M-step, eliminating the need to compute multi-dimensional integrations in the E-step. This is because expectations involving the responses for the i-th individual are conditionally independent given estimates of both fixed and random effects.

This approach is also not restricted to study designs with fixed or noninformative random censoring. The approach developed in Section 3.1 for noninformative random censoring mechanisms will be extended to informative censoring in Section 3.2.

TABLE 1.3.1
SUMMARY OF MIXED MODEL PROCEDURES

PARAMETER ESTIMATION	PETTITT (1985)	CARRIQUIRY et al. (1987)	PROPOSED METHOD
Maximum Likelihood Estimation	X		
Maximum a Posteriori Estimation		X	X
COMPUTATIONAL ALGORITHM			
EM Algorithm	X		X
Newton-Raphson Algorithm		X	
TYPE OF CENSORING			
Fixed Censoring		X	X
Noninformative Random Censoring	X		X
Informative Censoring			X
LIMITATIONS			
High-Dimensional Integrations	X		
Inversion of Large Matrices		X	
Restricted to Random Intercept Models		X	

TABLE 1.3.2

DIMENSIONS OF MATRICES THAT MUST BE INVERTED IN ORDER TO ESTIMATE FIXED AND RANDOM EFFECTS

Fixed Effects (#)	Random Effects (#)	Carriquiry et al.[1] (1987)	Pettitt (1985) and Proposed Method Fixed Effects	Pettitt (1985) and Proposed Method Random Effects
5	36	41 × 41	5 × 5	1 × 1
5	360	365 × 365	5 × 5	1 × 1
5	1000	1005 × 1005	5 × 5	1 × 1

1. Carriquiry et al. (1987) estimate fixed and random effects simultaneously.

II. GENERAL LINEAR UNIVARIATE MODELS WITH RANDOM CENSORING

The Cox and Oakes (1984) method assumes that the censoring values are predetermined constants (c_i) that are known to the investigator in advance. In practice the c_i usually are not known in advance and cannot be treated as fixed constants. In this chapter the Cox and Oakes (1984) method is extended to random noninformative right censoring in Section 2.1 and further extended to informative censoring in Section 2.2.

2.1 General Linear Univariate Models with Noninformative Right Censoring

In Section 2.1.1 the EM algorithm is derived for the general linear univariate model with noninformative right censoring. Corresponding EM Computations are described in Section 2.1.2.

2.1.1 General Linear Univariate Models with Noninformative Right Censoring; Derivation of the EM Algorithm

The method of Cox and Oakes (1984) described in Section 1.2.2.3 can easily be extended to noninformative right censoring. In the expectation step,

$$E[\ t(Y_i^*) \mid Y_i{=}y_i,\ \delta_i,\ \underset{\sim}{\theta}\,]=$$

$$\delta_i E[\ t(Y_i^*) \mid Y_i{=}y_i,\ \delta_i{=}1,\ \underset{\sim}{\theta}\,]+(1-\delta_i)E[\ t(Y_i^*) \mid Y_i{=}y_i,\ \delta_i{=}0,\ \underset{\sim}{\theta}\,]$$

$$=\delta_i E[\ t(Y_i^*) \mid Y_i^*{=}y_i,\ \underset{\sim}{\theta}\,]+(1-\delta_i)E[\ t(Y_i^*) \mid Y_i^*{>}y_i,\ \underset{\sim}{\theta}\,]$$

$$=\delta_i t(y_i)+(1-\delta_i)E[\ t(Y_i^*)\ |\ Y_i^*>y_i;\ \underset{\sim}{\theta}\ ]$$

with corresponding density function

$$f_0(Y^*\ |\ Y^*>y)=f_0(Y^*\ |\ Y=y,\ \delta=0)$$

$$=\frac{f_0(Y^*=y^*,\ Y=y,\ \delta=0)}{f_0(\ Y=y,\ \delta=0)}=\frac{f_0(Y^*=y^*,\ C^*=y)}{\int_y^\infty f_o(Y^*,\ C^*)\ dY^*}\ .$$

The last step follows from an argument analogous to the one in Section 1.1.

If Y^* and C^* are independent,

$$f_0(Y^*\ |\ Y^*>y)=\frac{g_{C^*}(y)\ f_o(Y^*)}{g_{C^*}(y)\ \int_y^\infty f_o(Y^*)\ dY^*}=\frac{f_o(Y^*)}{1-\Phi(z)}$$

where $z_i=\frac{y_i-\underset{\sim}{X}_i\underset{\sim}{\beta}}{\sigma_e}$.

Therefore

$$E[\ t(Y_i^*)\ |\ Y_i^*>y_i,\ \underset{\sim}{\theta}\ ]$$

$$=\frac{1}{1-\Phi\left(\frac{y_i-\underset{\sim}{X}_i\underset{\sim}{\beta}}{\sigma_e^2}\right)}\int_{y_i}^\infty t(Y_i^*)\ f_o(Y_i^*\ |\ \underset{\sim}{\theta}\ )\ dY_i^*.$$

Recall that the likelihood function for the observed data $(\underset{\sim}{Y},\ \underset{\sim}{\delta}\ )$ can be expressed as

$$l(\underset{\sim}{\theta}_1|\ \underset{\sim}{Y},\ \underset{\sim}{\delta}\ )=l_o(\underset{\sim}{\theta}_1,\ \underset{\sim}{Y}^*)-l_1(\underset{\sim}{\theta}_1,\ \underset{\sim}{Y}^*\ |\ \underset{\sim}{Y},\ \underset{\sim}{\delta}\ )$$

where l_o is the log likelihood for the complete data $\underset{\sim}{Y}^*$ that would be observed if there was no censoring and l_1 is the log likelihood of the conditional distribution of $\underset{\sim}{Y}^*$ given $(\underset{\sim}{Y}, \underset{\sim}{\delta})$. When $\underset{\sim}{Y}^* \sim N(\underset{\sim}{X}\underset{\sim}{\beta}, \underset{\sim}{I}\,\sigma_e^2)$,

$$L(\underset{\sim}{\beta}, \sigma_e^2 \mid \underset{\sim}{Y}, \underset{\sim}{\delta}) \propto \prod_{i=1}^{K} [\, f_o(Y_i^* \mid \underset{\sim}{\beta}, \sigma_e^2)]^{\delta_i} \left[1 - \Phi\!\left(\frac{y_i - \underset{\sim}{X}_i\underset{\sim}{\beta}}{\sigma_e}\right)\right]^{1-\delta_i}.$$

Therefore $l(\underset{\sim}{\beta}, \sigma_e^2 \mid \underset{\sim}{Y}, \underset{\sim}{\delta})$

$$\propto \sum_{i=1}^{K} \left\{ \delta_i \log f_o(Y_i^* \mid \underset{\sim}{\beta}, \sigma_e^2) + (1-\delta_i)\log\left[1 - \Phi\!\left(\frac{y_i - \underset{\sim}{X}_i\underset{\sim}{\beta}}{\sigma_e}\right)\right]\right\}$$

$$= \sum_{i=1}^{K} \log f_o(Y_i^* \mid \underset{\sim}{\beta}, \sigma_e^2) - \sum_{i=1}^{K} (1-\delta_i)\log \frac{f_o(Y_i^* \mid \underset{\sim}{\beta}, \sigma_e^2)}{\left[1 - \Phi\!\left(\frac{y_i - \underset{\sim}{X}_i\underset{\sim}{\beta}}{\sigma_e}\right)\right]}$$

where $\dfrac{f_o(Y_i^* \mid \underset{\sim}{\beta}, \sigma_e^2)}{\left[1 - \Phi\!\left(\frac{y_i - \underset{\sim}{X}_i\underset{\sim}{\beta}}{\sigma_e}\right)\right]}$

is the conditional density of Y_i^* given $Y_i^* > y_i$ with the same form as the unconditional density of Y_i^* except that the range of the density is restricted to $Y_i^* > y_i$.

As before

$$\frac{\partial}{\partial \underset{\sim}{\theta}} \log \frac{f_o(Y_i^* \mid \underset{\sim}{\theta})}{[1 - F(y_i \mid \underset{\sim}{\theta})]} = t(Y_i^*) - E[\, t(Y_i^*) \mid Y_i^* > y_i, \underset{\sim}{\theta}].$$

Therefore $\dfrac{\partial l(\underset{\sim}{\theta}; \underset{\sim}{Y}, \underset{\sim}{\delta})}{\partial \underset{\sim}{\theta}}$

$$= \sum_{i=1}^{K} \{\, t(Y_i^*) - E[\, t(Y_i^*) \mid \underset{\sim}{\theta}] \,\}$$

$$-\sum_{i=1}^{K}(1-\delta_i)\{\, t(Y_i^*) - E[\, t(Y_i^*) \mid Y_i^*>y_i, \underset{\sim}{\theta}\,]\,\}$$

$$=\sum_{i=1}^{K}\{\, E[\, t(Y_i^*) \mid \underset{\sim}{Y}=\underset{\sim}{y}\,, \underset{\sim}{\delta}\,, \underset{\sim}{\theta}\,] - E[\, t(Y_i^*) | \, \underset{\sim}{\theta}\,]\,\}$$

the difference between the conditional and unconditional expectations of the complete-data sufficient statistics and when $\underset{\sim}{\theta}_1=\underset{\sim}{\theta}=\hat{\underset{\sim}{\theta}}$ in (1.2.2) then $\hat{\underset{\sim}{\theta}}$ is also a solution of the

likelihood equation $\frac{\partial l(\underset{\sim}{\theta}\,;\, \underset{\sim}{Y}\,,\, \underset{\sim}{\delta}\,)}{\partial \underset{\sim}{\theta}}=\underset{\sim}{0}$.

2.1.2 General Linear Univariate Models with Noninformative Right Censoring; EM Computations

EM Computations are summarized in this section. In the E step, the conditional expected values of the 'complete-data' sufficient statistic are computed from the observed data and current estimates of the parameters, while in the M step, new estimates of the unknown parameters are computed using the conditional expected values of the 'complete-data' sufficient statistics in the maximum likelihood equations.

E Step:

In the r-th iteration, the estimation step computes the conditional expected values of the complete-data sufficient statistics given the observed data $\underset{\sim}{Y}$ and the estimated values of the parameters from the $(r-1)$-st iteration. Not all of the $\underset{\sim}{Y}^*$ are observed so the E step will estimate the complete-data sufficient statistics that involve $\underset{\sim}{Y}^*$.

A set of complete-data sufficient statistics for this problem is

$\underset{\sim}{X}'\underset{\sim}{Y}^*$ and $\underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^*$.

The expected values of the complete-data sufficient statistics may be denoted,

$\underset{\sim}{t}_1^{(r)}=\underset{\sim}{X}'\,E[\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$ and

$t_2^{(r)}=E[\underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$.

When $\delta_i=0$ expectations involving the i-th element of the complete-data sufficient statistics can be computed as:

$E[Y_i^* \mid Y_i=y_i, \delta_i=0, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$

$$=E[Y_i^* \mid Y_i^* > y_i, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=\int_{y_i}^{\infty} Y_i^* \frac{f_o(Y_i^* \mid \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2})}{1-\Phi(z_i^{(r-1)})} dY_i^*$$

$$=\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}+\sigma_e^{(r-1)} \frac{\phi(z_i^{(r-1)})}{1-\Phi(z_i^{(r-1)})}$$

where $z_i^{(r-1)}=\frac{y_i-\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}}{\sigma_e^{(r-1)}}$

and

$$E[Y_i^{*2} \mid Y_i=y_i, \delta_i=0, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=E[Y_i^{*2} \mid Y_i^* > y_i, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=\int_{y_i}^{\infty} Y_i^{*2} \frac{f_o(Y_i^* \mid \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2})}{1-\Phi(z_i^{(r-1)})} dY_i^*$$

$$=[\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}]^2+\sigma_e^{(r-1)^2}+\sigma_e^{(r-1)}(y_i+\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}) \frac{\phi(z_i^{(r-1)})}{1-\Phi(z_i^{(r-1)})}.$$

<u>M Step</u>:

The r-th iteration of the M step obtains $\underset{\sim}{\beta}^{(r)}$ and $\sigma_e^{(r)^2}$ as the solution to equation 1.2.3. For the complete-data problem the maximum likelihood estimates are:

$$\hat{\underset{\sim}{\beta}}=(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}^*$$

and

$$\hat{\sigma}_e^2=\frac{1}{K}(\underset{\sim}{Y}^*-\underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{Y}^*-\underset{\sim}{X}\underset{\sim}{\beta}).$$

$$=\frac{1}{K}(\underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^*-2\underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim}{Y}^*+\underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta}).$$

Initial values of maximum likelihood estimates are obtained by treating the censored data as if they were uncensored. Convenient initial estimates are:

$$\underset{\sim}{\beta}^{(0)}=(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}$$

and

$$\sigma_e^{(0)^2}=\frac{1}{K}(\underset{\sim}{Y}-\underset{\sim}{X}\underset{\sim}{\beta}^{(0)})'(\underset{\sim}{Y}-\underset{\sim}{X}\underset{\sim}{\beta}^{(0)})$$

In the r-th iteration maximum likelihood estimates which maximize the likelihood function using the expected values of the complete-data sufficient statistics obtained from the previous iteration of the E step are:

$$\underset{\sim}{\beta}^{(r)}=(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\ \underset{\sim}{t}_1^{(r)}$$

and

$$\sigma_e^{(r)^2}=\frac{1}{K}[t_2^{(r)}-2\underset{\sim}{\beta}^{(r-1)'}\ \underset{\sim}{t}_1^{(r)}+\underset{\sim}{\beta}^{(r-1)'}\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)}].$$

2.2 General Linear Univariate Models with Informative Right Censoring

In Section 2.2.1 the EM algorithm is derived for the general linear univariate model with informative right censoring. Corresponding EM Computations are described in Section 2.2.2.

2.2.1 General Linear Univariate Models with Informative Right Censoring; Derivation of the EM Algorithm

Consider a random sample of K individuals with one observation per subject from a normal population with common parameters $\underset{\sim}{\gamma}$, $\underset{\sim}{G}$, and $\underset{\sim}{\Omega}$. The General Linear Univariate Model is

$$\underset{\sim}{W}^*=\begin{bmatrix}\underset{\sim}{Y}^*\\ \underset{\sim}{C}^*\end{bmatrix}=\begin{bmatrix}\underset{\sim}{X} & \underset{\sim}{0}\\ \underset{\sim}{0} & \underset{\sim}{H}\end{bmatrix}\begin{bmatrix}\underset{\sim}{\beta}\\ \underset{\sim}{\alpha}\end{bmatrix}+\begin{bmatrix}\underset{\sim}{e}\\ \underset{\sim}{\epsilon}\end{bmatrix}$$

$$=\underset{\sim}{A}\underset{\sim}{\gamma}+\underset{\sim}{\nu}$$

where

$\underset{\sim}{W}^*$ is a $2n \times 1$ vector of failure and censoring values which may or may not be observed,

$\underset{\sim}{A}=\begin{bmatrix}\underset{\sim}{X} & \underset{\sim}{0}\\ \underset{\sim}{0} & \underset{\sim}{H}\end{bmatrix}$ is a $2n \times 2p$ known constant matrix of rank $2r \leq 2p$,

$\underset{\sim}{\gamma}=\begin{bmatrix}\underset{\sim}{\beta}\\ \underset{\sim}{\alpha}\end{bmatrix}$ is a $2p \times 1$ vector of unknown constant 'fixed' population parameters,

$\underset{\sim}{\nu}=\begin{bmatrix}\underset{\sim}{e}\\ \underset{\sim}{\epsilon}\end{bmatrix}$ is a $2n \times 1$ vector of unobservable random errors,

$\underset{\sim}{\nu} \sim N(\underset{\sim}{0}, \underset{\sim}{\Omega} \otimes \underset{\sim}{I}_n)$, and

$$\underset{\sim}{\Omega} = \begin{bmatrix} \sigma_e^2 & \rho\sigma_e\sigma_\epsilon \\ \rho\sigma_e\sigma_\epsilon & \sigma_\epsilon^2 \end{bmatrix}$$ consists of the unknown within-subject variance components, where

$$\underset{\sim}{\Omega} \otimes \underset{\sim}{I}_n = \begin{bmatrix} \underset{\sim}{I}_n\sigma_e^2 & \underset{\sim}{I}_n\rho\sigma_e\sigma_\epsilon \\ \underset{\sim}{I}_n\rho\sigma_e\sigma_\epsilon & \underset{\sim}{I}_n\sigma_\epsilon^2 \end{bmatrix}.$$

Define: $\delta_{1i} = \mathfrak{B}(Y_i = Y_i^*)$, $\delta_{2i} = \mathfrak{B}(C_i = C_i^*)$,

and $\underset{\sim}{W}$ to be the observed values of $\underset{\sim}{W}^*$,

where $\mathfrak{B}$ is the Boolean function (Helms 1988).

The joint bivariate normal distribution of the complete-data vector,

$\underset{\sim}{W}^* \sim N(\underset{\sim}{A}\underset{\sim}{\gamma}, \underset{\sim}{\Omega} \otimes \underset{\sim}{I})$

with density

$$f_o(\underset{\sim}{W}^* \mid \underset{\sim}{\gamma}, \Omega) = \left[\frac{1}{2\pi}\right]^K \mid \underset{\sim}{\Omega} \mid^{-\frac{K}{2}} \times \exp\left[-\frac{1}{2}(\underset{\sim}{W}^* - \underset{\sim}{A}\underset{\sim}{\gamma})'(\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I})(\underset{\sim}{W}^* - \underset{\sim}{A}\underset{\sim}{\gamma})\right]$$

and log likelihood

$$l_o(\underset{\sim}{\gamma}, \underset{\sim}{\Omega} \mid \underset{\sim}{W}^*) = \log f_o(\underset{\sim}{W}^* \mid \underset{\sim}{\gamma}, \underset{\sim}{\Omega})$$

$$= -K\log(2\pi) - \frac{K}{2}\log\mid \underset{\sim}{\Omega} \mid - \frac{1}{2}[(\underset{\sim}{W}^* - \underset{\sim}{A}\underset{\sim}{\gamma})'(\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I})(\underset{\sim}{W}^* - \underset{\sim}{A}\underset{\sim}{\gamma})]$$

is a member of the exponential class of distributions. The density has the regular exponential-family form

$f_o(\underset{\sim}{W}^* \mid \underset{\sim}{\theta}) = b(\underset{\sim}{W}^*) \exp[\underset{\sim}{\theta}' t(\underset{\sim}{W}^*)] / a(\underset{\sim}{\theta})$ and

$$\frac{\partial l_o(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} = t(\underset{\sim}{W}^*) - \frac{\partial}{\partial \underset{\sim}{\theta}} \log a(\underset{\sim}{\theta})$$

$$= t(\underset{\sim}{W}^*) - E[\, t(\underset{\sim}{W}^*) \mid \underset{\sim}{\theta}\,].$$

where $\underset{\sim}{\theta} = \{\underset{\sim}{\beta},\ \sigma_e^2,\ \underset{\sim}{\alpha},\ \sigma_\epsilon^2,\ \rho\}$ denotes the parameter vector that is restricted to a (p+1)-dimensional convex set $\underset{\sim}{\Psi}$ such that $f_o(\underset{\sim}{W}^* \mid \underset{\sim}{\theta})$ defines a density for all $\underset{\sim}{\theta}$ in $\underset{\sim}{\Psi}$ and

$$a(\underset{\sim}{\theta}) = \int_{\mathcal{W}^*} b(\underset{\sim}{W}^*) \exp[\, \underset{\sim}{\theta}' \, t(\underset{\sim}{W}^*)\,] \; d\underset{\sim}{W}^*$$

where $\mathcal{W}^*$ is the sample space of $\underset{\sim}{W}^*$.

For a right censored sample $(\underset{\sim}{W}, \underset{\sim}{\delta})$ the function

$$\begin{aligned} Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) &= E[l_o(\underset{\sim}{\theta}_1 \mid \underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] \\ &= E[\, \underset{\sim}{\theta}_1' t(\underset{\sim}{W}^*) - \log a(\underset{\sim}{\theta}_1) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] + \log b(\underset{\sim}{W}^*) \\ &= E[\, \underset{\sim}{\theta}_1' t(\underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] - \log a(\underset{\sim}{\theta}_1) + \log b(\underset{\sim}{W}^*). \end{aligned}$$

For a concave function l_o, Q is maximized with respect to $\underset{\sim}{\theta}_1$ when

$$0 = \frac{\partial Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}_1} = E[\, t(\underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] - \frac{\partial \log a(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1}$$

$$= E[\, t(\underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] - E[\, t(\underset{\sim}{W}^*) \mid \underset{\sim}{\theta}_1]$$

or $$E[\, t(\underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] = E[\, t(\underset{\sim}{W}^*) \mid \underset{\sim}{\theta}_1].$$

The corresponding E and M steps are:

E-step: Compute $t^{(r)}(\underset{\sim}{W}^*) = E[\, t(\underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}^{(r-1)}]$

M-step: Obtain $\underset{\sim}{\theta}^{(r)}$ as the solution to $E[\, t(\underset{\sim}{W}^*) \mid \underset{\sim}{\theta}^{(r)}] = t^{(r)}(\underset{\sim}{W}^*)$. (2.2.1)

With bivariate right censoring, there are three cases to be considered:

(1) $\delta_{1i}=\delta_{2i}=1$, i.e., both Y_i^* and C_i^* are known,

(2) $\delta_{1i}=1$ and $\delta_{2i}=0$ which implies that $C_i^*>Y_i^*=Y_i$

and (3) $\delta_{1i}=0$ and $\delta_{2i}=1$ which implies that $Y_i^*>C_i^*=C_i$.

It is assumed that at least one of the Y_i or C_i is observed. If both are missing (i.e., $\delta_{1i}=\delta_{2i}=0$), this observation is assumed to be ignorably missing.

In the expectation step,

$$E[\ t(W_i^*)\mid W_i=w_i,\ \underset{\sim}{\delta}_i,\ \underset{\sim}{\theta}\ ]=$$

$$\delta_{1i}\delta_{2i}E[\ t(W_i^*)\mid W_i=w_i,\ \delta_{1i}=1,\ \delta_{2i}=1,\ \underset{\sim}{\theta}\ ]$$
$$+(1-\delta_{1i})\delta_{2i}E[\ t(W_i^*)\mid W_i=w_i,\ \delta_{1i}=0,\ \delta_{2i}=1,\ \underset{\sim}{\theta}\ ]$$
$$+\delta_{1i}(1-\delta_{2i})E[\ t(W_i^*)\mid W_i=w_i,\ \delta_{1i}=1,\ \delta_{2i}=0,\ \underset{\sim}{\theta}\ ]$$

$$=\delta_{1i}\delta_{2i}E[\ t(W_i^*)\mid W_i^*=w_i,\ \underset{\sim}{\theta}\ ]$$
$$+(1-\delta_{1i})\delta_{2i}E[\ t(W_i^*)\mid Y_i^*>y_i,\ C_i^*=y_i,\ \underset{\sim}{\theta}\ ]$$
$$+\delta_{1i}(1-\delta_{2i})E[\ t(W_i^*)\mid Y_i^*=y_i,\ C_i^*>y_i,\ \underset{\sim}{\theta}\ ]$$

$$=\delta_{1i}\delta_{2i}t(w_i)$$
$$+(1-\delta_{1i})\delta_{2i}E[\ t(W_i^*)\mid Y_i^*>y_i,\ C_i^*=y_i,\ \underset{\sim}{\theta}\ ]$$
$$+\delta_{1i}(1-\delta_{2i})E[\ t(W_i^*)\mid Y_i^*=y_i,\ C_i^*>y_i,\ \underset{\sim}{\theta}\ ]$$

with corresponding density functions

$$f_0(Y^*\mid Y^*>y,\ C^*=y)=f_0(Y^*\mid Y=y,\ \delta_1=0,\ \delta_2=1)$$

$$=\frac{f_0(Y^*=y^*,\ Y=y,\ \delta_1=0,\ \delta_2=1)}{f_0(\ Y=y,\ \delta_1=0,\ \delta_2=1)}=\frac{f_0(Y^*=y^*,\ C^*=y)}{\int_y^\infty f_o(Y^*,\ C^*)\ dY^*}$$

$$=\frac{g_{C^*}(y)\ f_o(Y^*|\ C^*=y)}{g_{C^*}(y)\int_y^\infty f_o(Y^*|\ C^*=y)\ dY^*}$$

$$=\frac{f_o(Y^*|\ C^*=y)}{\int_y^\infty f_o(Y^*|\ C^*=y)\ dY^*}=\frac{f_o(Y^*|\ C^*=y)}{1-\Phi(z_{Y^*|C^*})}$$

where $z_{Y^*|C^*,\ i}=\dfrac{y_i-\left[\underset{\sim}{X}_i\underset{\sim}{\beta}+\rho\frac{\sigma_{Y^*}}{\sigma_{C^*}}(y_i-\underset{\sim}{H}_i\underset{\sim}{\alpha})\right]}{\sigma_{Y^*}(1-\rho^2)^{\frac{1}{2}}}$.

Similarly,

$f_0(C^*\ |\ Y^*=y,\ C^*>y)=f_0(C^*\ |\ Y=y,\ \delta_1=1,\ \delta_2=0)$

$$=\frac{f_0(C^*=c^*,\ Y=y,\ \delta_1=1,\ \delta_2=0)}{f_0(\ Y=y,\ \delta_1=1,\ \delta_2=0)}=\frac{f_0(Y^*=y,\ C^*=c^*)}{\int_y^\infty f_o(Y^*,\ C^*)\ dC^*}$$

$$=\frac{f_{Y^*}(y)\ g_o(C^*|\ Y^*=y)}{f_{Y^*}(y)\int_y^\infty g_o(C^*|\ Y^*=y)\ dC^*}$$

$$=\frac{g_o(C^*|\ Y^*=y)}{\int_y^\infty g_o(C^*|\ Y^*=y)\ dC^*}=\frac{g_o(C^*|\ Y^*=y)}{1-\Phi(z_{C^*|Y^*})}$$

where $z_{C^*|Y^*,\ i}=\dfrac{y_i-\left[\underset{\sim}{H}_i\underset{\sim}{\alpha}+\rho\frac{\sigma_{C^*}}{\sigma_{Y^*}}(y_i-\underset{\sim}{X}_i\underset{\sim}{\beta})\right]}{\sigma_{C^*}(1-\rho^2)^{\frac{1}{2}}}$.

Therefore

$$E[\ t(W_i^*) \mid Y_i^*>y_i,\ C_i^*=y_i,\ \underset{\sim}{\theta}\ ]$$

$$=\frac{1}{1-\Phi(z_{Y^*\mid C^*,\ i})}\int_{y_i}^{\infty} t(W_i^*)\ f_o(Y_i^* \mid C_i^*=y_i,\ \underset{\sim}{\theta}\ )\ dY_i^*$$

and

$$E[\ t(W_i^*) \mid Y_i^*=y_i,\ C_i^*>y_i,\ \underset{\sim}{\theta}\ ]$$

$$=\frac{1}{1-\Phi(z_{C^*\mid Y^*,\ i})}\int_{y_i}^{\infty} t(W_i^*)\ g_o(C_i^* \mid Y_i^*=y_i,\ \underset{\sim}{\theta}\ )\ dC_i^*.$$

2.2.2 General Linear Univariate Models with Informative Right Censoring; EM Computations

EM Computations are summarized in this section. In the E step, the conditional expected values of the 'complete-data' sufficient statistic are computed from the observed data and current estimates of the parameters, while in the M step, new estimates of the unknown parameters are computed using the conditional expected values of the 'complete-data' sufficient statistics in the maximum likelihood equations.

E Step:

In the r-th iteration, the estimation step computes the conditional expected values of the complete-data sufficient statistics given the observed data $\underset{\sim}{W}$ and the estimated values of the parameters from the $(r-1)$-st iteration. Not all of the $\underset{\sim}{W}^*$ are observed so the E step will estimate the complete-data sufficient statistics that involve $\underset{\sim}{W}^*$.

A set of complete-data sufficient statistics for this problem is

$$\underset{\sim}{A}'(\underset{\sim}{\Omega}^{-1}\otimes\underset{\sim}{I})\,\underset{\sim}{W}^*, \qquad (\underset{\sim}{Y}^*-\underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{Y}^*-\underset{\sim}{X}\underset{\sim}{\beta}),$$

$$(\underset{\sim}{C}^*-\underset{\sim}{H}\underset{\sim}{\alpha})'(\underset{\sim}{C}^*-\underset{\sim}{H}\underset{\sim}{\alpha}), \qquad \text{and} \qquad (\underset{\sim}{Y}^*-\underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{C}^*-\underset{\sim}{H}\underset{\sim}{\alpha}).$$

The r-th iteration of the E-step consists of evaluating the expectations:

$$\underset{\sim}{t}_1^{(r)} = \underset{\sim}{A}'(\underset{\sim}{\Omega}^{(r-1)^{-1}}\otimes\underset{\sim}{I})\,E[\,\underset{\sim}{W}^* \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\gamma}^{(r-1)}, \underset{\sim}{\Omega}^{(r-1)}\,]$$

$$\underset{\sim}{t}_2^{(r)} = E[\,(\underset{\sim}{Y}^*-\underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{Y}^*-\underset{\sim}{X}\underset{\sim}{\beta}) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=E[\, \underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^{*} \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}] - 2\underset{\sim}{\beta}^{(r-1)\prime}\underset{\sim}{X}'\, E[\, \underset{\sim}{Y}^{*} \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$+\underset{\sim}{\beta}^{(r-1)\prime}\underset{\sim}{X}'\,\underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)}$$

$$\underset{\sim}{t}_3^{(r)} = E[\,(\underset{\sim}{C}^{*}-\underset{\sim}{H}\underset{\sim}{\alpha})'(\underset{\sim}{C}^{*}-\underset{\sim}{H}\underset{\sim}{\alpha}) \mid \underset{\sim}{C}, \underset{\sim}{\delta}, \underset{\sim}{\alpha}^{(r-1)}, \sigma_\epsilon^{(r-1)^2}]$$

$$=E[\, \underset{\sim}{C}^{*\prime}\underset{\sim}{C}^{*} \mid \underset{\sim}{C}, \underset{\sim}{\delta}, \underset{\sim}{\alpha}^{(r-1)}, \sigma_\epsilon^{(r-1)^2}] - 2\underset{\sim}{\alpha}^{(r)\prime}\underset{\sim}{H}'\, E[\, \underset{\sim}{C}^{*} \mid \underset{\sim}{C}, \underset{\sim}{\delta}, \underset{\sim}{\alpha}^{(r-1)}, \sigma_\epsilon^{(r-1)^2}]$$

$$+\underset{\sim}{\alpha}^{(r-1)\prime}\underset{\sim}{H}'\,\underset{\sim}{H}\,\underset{\sim}{\alpha}^{(r-1)}$$

and

$$\underset{\sim}{t}_4^{(r)} = E[\,(\underset{\sim}{Y}^{*}-\underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{C}^{*}-\underset{\sim}{H}\underset{\sim}{\alpha}) \mid \underset{\sim}{Y}, \underset{\sim}{C}, \underset{\sim}{\delta}, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}, \underset{\sim}{\alpha}^{(r-1)}, \sigma_\epsilon^{(r-1)^2}]$$

$$=E[\, \underset{\sim}{Y}^{*\prime}\underset{\sim}{C}^{*} \mid \underset{\sim}{Y}, \underset{\sim}{C}, \underset{\sim}{\delta}, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}, \underset{\sim}{\alpha}^{(r-1)}, \sigma_\epsilon^{(r-1)^2}]$$

$$-\underset{\sim}{\beta}^{(r-1)\prime}\underset{\sim}{X}'\, E[\, \underset{\sim}{C}^{*} \mid \underset{\sim}{C}, \underset{\sim}{\delta}, \underset{\sim}{\alpha}^{(r-1)}, \sigma_\epsilon^{(r-1)^2}]$$

$$-\underset{\sim}{\alpha}^{(r-1)\prime}\underset{\sim}{H}'\, E[\, \underset{\sim}{Y}^{*} \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\beta}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$+\underset{\sim}{\beta}^{(r-1)\prime}\underset{\sim}{X}'\,\underset{\sim}{H}\underset{\sim}{\alpha}^{(r-1)}.$$

When δ_{1i}=0 and δ_{2i}=1 expectations involving the i-th element of the complete-data sufficient statistics can be computed as:

$$E[Y_i^* \mid Y_i^*>y_i,\ C_i^*=y_i,\ \underset{\sim}{\theta}^{(r-1)}]$$

$$=\int_{y_i}^{\infty} Y_i^* \frac{f_o(Y_i^* | C_i^*=y_i,\ \underset{\sim}{\theta}^{(r-1)})}{1-\Phi(z_{Y^*|C^*,\ i}^{(r-1)})}\, dY_i^*$$

$$=\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}+\rho^{(r-1)}\frac{\sigma_e^{(r-1)}}{\sigma_\epsilon^{(r-1)}}\,(y_i-\underset{\sim}{H}_i\underset{\sim}{\alpha}^{(r-1)})+\sigma_e^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}\frac{\phi(z_{Y^*|C^*,\ i}^{(r-1)})}{1-\Phi(z_{Y^*|C^*,\ i}^{(r-1)})}$$

where

$$z_{Y^*|C^*,\ i}^{(r-1)}=\frac{y_i-\left[\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}+\rho^{(r-1)}\frac{\sigma_e^{(r-1)}}{\sigma_\epsilon^{(r-1)}}(y_i-\underset{\sim}{H}_i\underset{\sim}{\alpha}^{(r-1)})\right]}{\sigma_e^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}},$$

and

$$E[Y_i^{*2} \mid Y_i^*>y_i,\ C_i^*=y_i,\ \underset{\sim}{\theta}^{(r-1)}]$$

$$=\int_{y_i}^{\infty} Y_i^{*2} \frac{f_o(Y_i^* | C_i^*=y_i,\ \underset{\sim}{\theta}^{(r-1)})}{1-\Phi(z_{Y^*|C^*,\ i}^{(r-1)})}\, dY_i^*$$

$$=[\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}+\rho^{(r-1)}\frac{\sigma_e^{(r-1)}}{\sigma_\epsilon^{(r-1)}}\,(y_i-\underset{\sim}{H}_i\underset{\sim}{\alpha}^{(r-1)})]^2+\sigma_e^{(r-1)^2}(1-\rho^{(r-1)^2})$$

$$+\sigma_e^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}} \times\ [y_i+\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)}+\rho^{(r-1)}\frac{\sigma_e^{(r-1)}}{\sigma_\epsilon^{(r-1)}}\,(y_i-\underset{\sim}{H}_i\underset{\sim}{\alpha}^{(r-1)})\ ]$$

$$\times\ \frac{\phi(z_{Y^*|C^*,\ i}^{(r-1)})}{1-\Phi(z_{Y^*|C^*,\ i}^{(r-1)})}\,.$$

Similarly, when $\delta_{1i}=1$ and $\delta_{2i}=0$ expectations involving the i-th element of the complete-data sufficient statistics can be computed as:

$$E[C_i^* \mid Y_i^*=y_i,\ C_i^*>y_i,\ \underset{\sim}{\theta}^{(r-1)}]$$

$$=\int_{y_i}^{\infty} C_i^* \frac{g_o(C_i^* \mid Y_i^*=y_i,\ \underset{\sim}{\theta}^{(r-1)})}{1-\Phi(z^{(r-1)}_{C^* \mid Y^*,\ i})}\, dC_i^*$$

$$=\underset{\sim}{H}_i\underset{\sim}{\alpha}^{(r-1)}+\rho^{(r-1)}\frac{\sigma_\epsilon^{(r-1)}}{\sigma_e^{(r-1)}}\,(y_i-\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)})+\sigma_\epsilon^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}\ \frac{\phi(z^{(r-1)}_{C^* \mid Y^*,\ i})}{1-\Phi(z^{(r-1)}_{C^* \mid Y^*,\ i})}$$

where

$$z^{(r-1)}_{C^* \mid Y^*,\ i}=\frac{y_i-\left[\underset{\sim}{H}_i\underset{\sim}{\alpha}^{(r-1)}+\rho^{(r-1)}\frac{\sigma_\epsilon^{(r-1)}}{\sigma_e^{(r-1)}}(y_i-\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)})\right]}{\sigma_\epsilon^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}},$$

and

$$E[C_i^{*2} \mid Y_i^*=y_i,\ C_i^*>y_i,\ \underset{\sim}{\theta}^{(r-1)}]$$

$$=\int_{y_i}^{\infty} C_i^{*2} \frac{g_o(C_i^* \mid Y_i^*=y_i,\ \underset{\sim}{\theta}^{(r-1)})}{1-\Phi(z^{(r-1)}_{C^* \mid Y^*,\ i})}\, dC_i^*$$

$$=[\underset{\sim}{H}_i\underset{\sim}{\alpha}^{(r-1)}+\rho^{(r-1)}\frac{\sigma_\epsilon^{(r-1)}}{\sigma_e^{(r-1)}}\,(y_i-\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)})]^2+\sigma_\epsilon^{(r-1)^2}(1-\rho^{(r-1)^2})$$

$$+\sigma_\epsilon^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}\ \times\ [y_i+\underset{\sim}{H}_i\underset{\sim}{\alpha}^{(r-1)}+\rho^{(r-1)}\frac{\sigma_\epsilon^{(r-1)}}{\sigma_e^{(r-1)}}\,(y_i-\underset{\sim}{X}_i\underset{\sim}{\beta}^{(r-1)})\ ]$$

$$\times\ \frac{\phi(z^{(r-1)}_{C^* \mid Y^*,\ i})}{1-\Phi(z^{(r-1)}_{C^* \mid Y^*,\ i})}.$$

M Step:

The r-th iteration of the M step obtains $\underset{\sim}{\beta}^{(r)}$, $\underset{\sim}{\alpha}^{(r)}$, $\sigma_e^{(r)^2}$, $\sigma_\epsilon^{(r)^2}$, and $\rho^{(r)}$ as the solution to equation 2.2.1. For the complete-data problem the maximum likelihood estimates are:

$$\hat{\underset{\sim}{\gamma}} = \begin{bmatrix} \hat{\underset{\sim}{\beta}} \\ \hat{\underset{\sim}{\alpha}} \end{bmatrix} = [\, \underset{\sim}{A}' \, (\hat{\underset{\sim}{\Omega}}^{-1} \otimes \underset{\sim}{I}) \, \underset{\sim}{A} \,]^{-1} \, \underset{\sim}{A}' \, (\hat{\underset{\sim}{\Omega}}^{-1} \otimes \underset{\sim}{I}) \, \underset{\sim}{W}^*,$$

$$\hat{\sigma}_e^2 = \frac{1}{K}(\underset{\sim}{Y}^* - \underset{\sim}{X}\hat{\underset{\sim}{\beta}})'(\underset{\sim}{Y}^* - \underset{\sim}{X}\hat{\underset{\sim}{\beta}}),$$

$$\hat{\sigma}_\epsilon^2 = \frac{1}{K}(\underset{\sim}{C}^* - \underset{\sim}{H}\hat{\underset{\sim}{\alpha}})'(\underset{\sim}{C}^* - \underset{\sim}{H}\hat{\underset{\sim}{\alpha}}) \, ,$$

and

$$\hat{\rho} = \frac{1}{K\hat{\sigma}_e\hat{\sigma}_\epsilon}(\underset{\sim}{Y}^* - \underset{\sim}{X}\hat{\underset{\sim}{\beta}})'(\underset{\sim}{C}^* - \underset{\sim}{H}\hat{\underset{\sim}{\alpha}}).$$

Initial values of maximum likelihood estimates are obtained by treating the censored data as if they were uncensored. Convenient initial estimates are:

$$\underset{\sim}{\gamma}^{(0)} = (\underset{\sim}{A}'\underset{\sim}{A})^{-1} \, \underset{\sim}{A}'\underset{\sim}{W},$$

$$\sigma_e^{(0)^2} = \frac{1}{K}(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta}^{(0)})'(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta}^{(0)}),$$

$$\sigma_\epsilon^{(0)^2} = \frac{1}{K}(\underset{\sim}{C} - \underset{\sim}{H}\underset{\sim}{\alpha}^{(0)})'(\underset{\sim}{C} - \underset{\sim}{H}\underset{\sim}{\alpha}^{(0)}),$$

and

$$\rho^{(0)} = \frac{1}{K\sigma_e^{(0)}\sigma_\epsilon^{(0)}}(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta}^{(0)})'(\underset{\sim}{C} - \underset{\sim}{H}\underset{\sim}{\alpha}^{(0)}).$$

In the r-th iteration maximum likelihood estimates which maximize the likelihood function using the expected values of the complete-data sufficient statistics obtained from the previous iteration of the E step are:

$$\underset{\sim}{\gamma}^{(r)}=[\ \underset{\sim}{A}'\ (\underset{\sim}{\Omega}^{(r-1)^{-1}} \otimes \underset{\sim}{I}\ )\ \underset{\sim}{A}\ ]^{-1}\ \times\ \underset{\sim}{t}_1^{(r)},$$

$$\sigma_e^{(r)^2}=\frac{1}{K}\ t_2^{(r)},$$

$$\sigma_\epsilon^{(r)^2}=\frac{1}{K}\ t_3^{(r)},$$

and

$$\rho^{(r)}=\frac{1}{K\sigma_e^{(r-1)}\sigma_\epsilon^{(r-1)}}\ t_4^{(r)}\ .$$

III. MIXED MODELS WITH RANDOM CENSORING

In this chapter, the methods discussed in Chapter 2 for general linear univariate models with random censoring are extended to mixed models with noninformative censoring in discussed in Section 3.1 and extended further to informative censoring in Section 3.2.

3.1 Mixed Models with Noninformative Right Censoring.

Likelihood functions for complete data are derived in Section 3.1.1, and theory and applications of the EM algorithm to mixed models with random noninformative right censoring are discussed in Sections 3.1.2 and 3.1.3.

3.1.1 Mixed Models with Noninformative Right Censoring; Likelihood Functions

If there is no censoring, the probability density function (pdf) of $\underset{\sim}{Y}^*$ given $\underset{\sim}{\beta}$ and $\underset{\sim}{d}$ is:

$$f(\underset{\sim}{Y}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}, \sigma_e^2)$$

$$= \prod_{i=1}^{K} f(\underset{\sim}{Y}_i^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i, \sigma_e^2)$$

$$= \prod_{i=1}^{K} \left[\frac{1}{2\pi\sigma_e^2}\right]^{\frac{n_i}{2}} \times \exp\left[-\tfrac{1}{2}\,[(\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i\underset{\sim}{\beta} - \underset{\sim}{B}_i\underset{\sim}{d}_i)'\,(\sigma_e^2\,\underset{\sim}{I}\,n_i)^{-1}(\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i\underset{\sim}{\beta} - \underset{\sim}{B}_i\underset{\sim}{d}_i)]\;\right]$$

$$= \left[\frac{1}{2\pi\sigma_e^2}\right]^{\frac{n}{2}}$$

$$\times \exp\; -\tfrac{1}{2}\left[\sum_{i=1}^{K} [(\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i\underset{\sim}{\beta} - \underset{\sim}{B}_i\underset{\sim}{d}_i)'\,(\sigma_e^2\,\underset{\sim}{I}\,n_i)^{-1}(\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i\underset{\sim}{\beta} - \underset{\sim}{B}_i\,\underset{\sim}{d}_i)]\right] \qquad (3.1.1)$$

Suitable flat prior distributions for this problem are:

$\Pi_1(\underset{\sim}{\beta}) \propto$ constant, and

$\Pi_2(\underset{\sim}{D}, \sigma_e^2) \propto$ constant,

and a convenient prior for $\underset{\sim}{d}_i$ is

$\Pi_3(\underset{\sim}{d}_i \mid \underset{\sim}{D}) = N_q(\underset{\sim}{0}, \underset{\sim}{D})$.

This leads to the joint prior distribution

$$\Pi(\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2) \propto \Pi_1(\underset{\sim}{\beta}) \cdot \prod_{i=1}^{K} \Pi_3(\underset{\sim}{d}_i \mid \underset{\sim}{D}) \cdot \Pi_2(\underset{\sim}{D}, \sigma_e^2). \qquad (3.1.2)$$

The joint p.d.f. of the distribution of $\underset{\sim}{Y}^*$, $\underset{\sim}{\beta}$, $\underset{\sim}{d}$, $\underset{\sim}{D}$, and σ_e^2 is equal to the product of the conditional density in (3.1.1) and the joint prior distribution in (3.1.2), i.e.

$$p(\underset{\sim}{Y}^*, \underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2) = f(\underset{\sim}{Y}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2) \times \Pi(\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2).$$

Using Bayes Theorem

$$p(\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2 \mid \underset{\sim}{Y}^*) = \frac{p(\underset{\sim}{Y}^*, \underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2)}{\int_{\underset{\sim}{\beta}}\int_{\underset{\sim}{d}}\int_{\underset{\sim}{D}}\int_{\sigma_e^2} p(\underset{\sim}{Y}^*, \underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2)\, d\sigma_e^2\, d\underset{\sim}{D}\, d\underset{\sim}{d}\, d\underset{\sim}{\beta}}$$

$$= \frac{p(\underset{\sim}{Y}^*, \underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2)}{p(\underset{\sim}{Y}^*)}.$$

Since $p(\underset{\sim}{Y}^*)$ does not depend on any of the parameters, maximizing $p(\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2 \mid \underset{\sim}{Y}^*)$ is equivalent to maximizing $p(\underset{\sim}{Y}^*, \underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2)$ with respect to $\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}$, or σ_e^2.

Therefore $p(\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2 \mid \underset{\sim}{Y}^*) = \dfrac{p(\underset{\sim}{Y}^*, \underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2)}{p(\underset{\sim}{Y}^*)}$

$$\propto p(\underset{\sim}{Y}^*, \underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2)$$

$$= (\sigma_e^2)^{-\frac{n}{2}} \, |\underset{\sim}{D}|^{-\frac{K}{2}}$$

$$\times \exp\left[-\frac{1}{2} \sum_{i=1}^{K} [(\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i \underset{\sim}{\beta} - \underset{\sim}{B}_i \underset{\sim}{d}_i)' (\underset{\sim}{I}_{n_i} \sigma_e^2)^{-1} (\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i \underset{\sim}{\beta} - \underset{\sim}{B}_i \underset{\sim}{d}_i) + \underset{\sim}{d}_i' \underset{\sim}{D}^{-1} \underset{\sim}{d}_i] \right].$$

Therefore the logarithm of the posterior distribution for the parameters $\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}$, and σ_e^2 is:

$$p(\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2 \mid \underset{\sim}{Y}^*) \propto -\frac{1}{2} \Big\{ n \log(\sigma_e^2) + K \log |\underset{\sim}{D}|$$

$$+ \sum_{i=1}^{K} (\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i \underset{\sim}{\beta} - \underset{\sim}{B}_i \underset{\sim}{d}_i)' (\underset{\sim}{I}_{n_i} \sigma_e^2)^{-1} (\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i \underset{\sim}{\beta} - \underset{\sim}{B}_i \underset{\sim}{d}_i) + \sum_{i=1}^{K} \underset{\sim}{d}_i' \underset{\sim}{D}^{-1} \underset{\sim}{d}_i \Big\}.$$

The <u>maximum</u> <u>a</u> <u>posterior</u> <u>estimators</u> (Beck and Arnold 1977) of $\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}$, and σ_e^2 are the coordinates of the mode of the posterior distribution, treated as an analog of a likelihood function (i.e., the parameters are variables and the data are constants). Posterior mode estimators are equivalent to maximum likelihood estimates for parameters with flat priors (Laird and Ware 1982), but in this case we do not have flat priors for the random effects.

An equivalent derivation using a frequentist approach is given in Fairclough and Helms (1984) for an artificial General Mixed Model which treats the unobserved $\underset{\sim}{d}_i$'s as missing data, thus facilitating the use of the EM algorithm. The model is defined as:

$$\begin{bmatrix} \underset{\sim}{Y}_i^* \\ \underset{\sim}{d}_i \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}_i \underset{\sim}{\beta} \\ \underset{\sim}{0} \end{bmatrix} + \begin{bmatrix} \underset{\sim}{B}_i \\ \underset{\sim}{I}_q \end{bmatrix} \underset{\sim}{d}_i + \begin{bmatrix} \underset{\sim}{e}_i \\ \underset{\sim}{0} \end{bmatrix}.$$

The log of the likelihood function can then be written as:

$$l_o(\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2) = -\tfrac{1}{2}\Bigg\{ (n+Kq)\log(2\pi) + \sum_{i=1}^{K} \log \begin{vmatrix} \underset{\sim}{\Sigma}_i & \underset{\sim}{B}_i \underset{\sim}{D} \\ \underset{\sim}{D}\underset{\sim}{B}_i' & \underset{\sim}{D} \end{vmatrix}$$

$$+ \sum_{i=1}^{K} \begin{bmatrix} \underset{\sim}{Y}_i^* - \underset{\sim}{X}_i \underset{\sim}{\beta} \\ \underset{\sim}{d}_i - \underset{\sim}{0} \end{bmatrix}' \begin{bmatrix} \underset{\sim}{\Sigma}_i & \underset{\sim}{B}_i \underset{\sim}{D} \\ \underset{\sim}{D}\underset{\sim}{B}_i' & \underset{\sim}{D} \end{bmatrix}^{-1} \begin{bmatrix} \underset{\sim}{Y}_i^* - \underset{\sim}{X}_i \underset{\sim}{\beta} \\ \underset{\sim}{d}_i - \underset{\sim}{0} \end{bmatrix} \Bigg\}$$

$$= -\tfrac{1}{2}\Big\{ (n+Kq)\log(2\pi) + n\log(\sigma_e^2) + K\log|\underset{\sim}{D}|$$

$$+ \sum_{i=1}^{K} (\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i \underset{\sim}{\beta} - \underset{\sim}{B}_i \underset{\sim}{d}_i)'(\underset{\sim}{I}_{n_i}\, \sigma_e^2)^{-1}(\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i \underset{\sim}{\beta} - \underset{\sim}{B}_i \underset{\sim}{d}_i) + \sum_{i=1}^{K} \underset{\sim}{d}_i' \underset{\sim}{D}^{-1} \underset{\sim}{d}_i \Big\}.$$

Therefore the posterior distribution of the parameters $\underset{\sim}{\beta}$, $\underset{\sim}{d}$, $\underset{\sim}{D}$, and σ_e^2 given $\underset{\sim}{Y}^*$, assuming flat priors for fixed effects and variance components and normal priors for random effects is proportional to the likelihood of $\underset{\sim}{\beta}$, $\underset{\sim}{D}$, and σ_e^2 given $\underset{\sim}{Y}^*$ and $\underset{\sim}{d}$.

Fairclough and Helms (1984) showed that the maximum likelihood estimates of the parameters $\underset{\sim}{\beta}$, $\underset{\sim}{d}_i$, $\underset{\sim}{D}$, and σ_e^2 are:

$$\hat{\beta} = \left[\sum_{i=1}^{K} X_i' X_i\right]^{-1} \sum_{i=1}^{K} X_i' (Y_i - B_i \hat{d}_i),$$

$$\hat{d}_i = [\hat{\sigma}_e^2 \hat{D}^{-1} + B_i' B_i]^{-1} B_i' (Y_i - X_i \hat{\beta}),$$

$$\hat{D} = \frac{1}{K} \sum_{i=1}^{K} \hat{d}_i \hat{d}_i',$$

$$\hat{\sigma}_e^2 = \frac{1}{n} \left[\sum_{i=1}^{K} (Y_i^* - X_i \hat{\beta} - B_i \hat{d}_i)' (Y_i^* - X_i \hat{\beta} - B_i \hat{d}_i)\right],$$

and if $D = \sum_{g=1}^{m} \tau_g G_g$ then

$$\hat{\tau} = \frac{1}{K} \left[< \text{trace } (\hat{D}^{-1} G_g \hat{D}^{-1} G_h) >_{gh} \right]^{-1} \left[< \sum_{i=1}^{K} \hat{d}_i' \hat{D}^{-1} G_g \hat{D}^{-1} \hat{d}_i >_g \right].$$

3.1.2 Mixed Models with Noninformative Right Censoring; Derivation of the EM Algorithm

The EM Algorithm (Dempster et al. 1977) can be used to compute maximum a posteriori estimates of $\underset{\sim}{\beta}$, $\underset{\sim}{d}$, $\underset{\sim}{D}$, and σ_e^2. The distribution of the complete-data vector,

$$(\underset{\sim}{Y}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}) \sim N(\underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{B}\underset{\sim}{d},\ \underset{\sim}{I}_n\sigma_e^2)$$

with density

$f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{I}_n\sigma_e^2)$

$$= \left[\frac{1}{2\pi\sigma_e^2}\right]^{\frac{n}{2}} \times \exp\left[-\frac{1}{2}\sum_{i=1}^{K}(\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i\underset{\sim}{\beta} - \underset{\sim}{B}_i\underset{\sim}{d}_i)'(\underset{\sim}{I}_n\sigma_e^2)^{-1}(\underset{\sim}{Y}_i^* - \underset{\sim}{X}_i\underset{\sim}{\beta} - \underset{\sim}{B}_i\underset{\sim}{d}_i)\right]$$

is a member of the exponential class of distributions. The density has the regular exponential-family form

$f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\theta}) = b(\underset{\sim}{Y}^*) \exp[\,\underset{\sim}{\theta}'\, t(\underset{\sim}{Y}^*)\,] / a(\underset{\sim}{\theta})$

where $\underset{\sim}{\theta} = \{\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2\}$ denotes the parameter vector that is restricted to a (p+1)-dimensional convex set $\underset{\sim}{\Psi}$. Applications of the EM Algorithm to densities from Regular Exponential Families were discussed in Section 1.2.2.1. The behavior of the EM Algorithm was previously discussed in Section 1.2.2.2 for the General Univariate Model. The proof is similar for General Linear Mixed Models except that $\underset{\sim}{\theta} = \{\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2\}$ instead of $\{\underset{\sim}{\beta}, \sigma_e^2\}$ and $\delta_{ij} = \mathfrak{B}(Y_{ij} = Y_{ij}^*)$ instead of $\mathfrak{B}(Y_i = Y_i^*)$. In addition, the posterior distribution function contains prior information about $\underset{\sim}{d}$. For mixed models, the function

$Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) = E[l_o(\underset{\sim}{\theta}_1, \underset{\sim}{Y}^*) + \log \pi(\underset{\sim}{d}_1 \mid \underset{\sim}{D}_1) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}]$,

the conditional expectation of the logarithm of the posterior distribution function of

$\underset{\sim}{Y}^*$ given the observed data ($\underset{\sim}{Y}$, $\underset{\sim}{\delta}$), where $\delta_{ij}=\mathfrak{B}(Y_{ij}=Y^*_{ij})$ and $\underset{\sim}{Y}$ and $\underset{\sim}{\delta}$ are assumed to be fixed and known. Q has two arguments; $\underset{\sim}{\theta}_1$ is an argument of the full likelihood L_o while $\underset{\sim}{\theta}$ is the parameter of the conditional distribution of $\underset{\sim}{Y}^*$ given ($\underset{\sim}{Y}$, $\underset{\sim}{\delta}$) which is used in computations involving the conditional expectation. The EM algorithm obtains a value $\underset{\sim}{\theta}^*$ that maximizes $f(\underset{\sim}{Y}^*; \underset{\sim}{\theta})$.

For this problem the logarithm of the posterior distribution function of $\underset{\sim}{\theta}_1$ is

$$p(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}, \underset{\sim}{\delta}) = \log f(\underset{\sim}{Y}, \underset{\sim}{\delta} \mid \underset{\sim}{\theta}_1) + \log \pi(\underset{\sim}{d}_1 | \underset{\sim}{D}_1)$$

and $m(\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}_1) = \dfrac{f_o(\underset{\sim}{Y}^* | \underset{\sim}{\theta}_1)}{f(\underset{\sim}{Y}, \underset{\sim}{\delta} \mid \underset{\sim}{\theta}_1)}$.

Therefore $p(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}, \underset{\sim}{\delta}) = \log f_o(\underset{\sim}{Y}^* \mid \underset{\sim}{\theta}_1) - \log m(\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}_1) + \log \pi(\underset{\sim}{d}_1 \mid \underset{\sim}{D}_1)$

$$= l_o(\underset{\sim}{\theta}_1, \underset{\sim}{Y}^*) - l_1(\underset{\sim}{\theta}_1, \underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}) + \log \pi(\underset{\sim}{d}_1 \mid \underset{\sim}{D}_1)$$

where

$$l_1(\underset{\sim}{\theta}_1, \underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}) = \log m(\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}_1).$$

Therefore $p(\underset{\sim}{\theta}_1) = E[\, l_o(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}^*) + \log \pi(\underset{\sim}{d}_1 | \underset{\sim}{D}_1) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\,] - E[\, l_1(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\,]$

$$= Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) - R(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})$$

where $R(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) = E[\, l_1(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\,]$.

Consequently,

$$p(\underset{\sim}{\theta}_1) - p(\underset{\sim}{\theta}) = [Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) - Q(\underset{\sim}{\theta}, \underset{\sim}{\theta})] - [R(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) - R(\underset{\sim}{\theta}, \underset{\sim}{\theta})].$$

Using Lemma 1 from Dempster et al. (1977), for any pair ($\underset{\sim}{\theta}_1$, $\underset{\sim}{\theta}$) in $\underset{\sim}{\Psi}$ X $\underset{\sim}{\Psi}$, $R(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) \leq R(\underset{\sim}{\theta}, \underset{\sim}{\theta})$. This follows because the expected value of a concave density function is maximized at the true value of the parameter. Because $\underset{\sim}{\theta}_1$ is chosen in the M step to maximize $Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})$ for a previously given value of $\underset{\sim}{\theta}$, $Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) \geq Q(\underset{\sim}{\theta}, \underset{\sim}{\theta})$. Therefore $p(\underset{\sim}{\theta}_1) - p(\underset{\sim}{\theta}) \geq 0$, so each iteration of the EM algorithm cannot decrease the posterior distribution function. Note that when the maximum a posteriori estimator of $\underset{\sim}{\theta}$ is obtained, the maximum a posteriori estimator, $\hat{\underset{\sim}{\theta}}$ must satisfy the self-consistency condition $Q(\underset{\sim}{\theta}_1, \hat{\underset{\sim}{\theta}}) \leq Q(\hat{\underset{\sim}{\theta}}, \hat{\underset{\sim}{\theta}})$, and it is therefore impossible to increase the value of the posterior distribution function at subsequent iterations. Note that concavity of p_o with respect to the parameters of interest ($\underset{\sim}{\beta}$, $\underset{\sim}{d}$, $\underset{\sim}{D}$, and σ_e^2) does not necessarily imply concavity of p.

The method of Cox and Oakes (1984) can easily be extended to noninformative right censoring in mixed models. When $\underset{\sim}{Y}^*$ is a member of the exponential class of distributions,

$$f_o(\underset{\sim}{Y}^* | \underset{\sim}{\theta}) = b(\underset{\sim}{Y}^*) \exp [\underset{\sim}{\theta}' t(\underset{\sim}{Y}^*)] / a(\underset{\sim}{\theta}).$$

For mixed models, $p_o(\underset{\sim}{\theta} | \underset{\sim}{Y}, \underset{\sim}{\delta}) = \log f_o(\underset{\sim}{Y}, \underset{\sim}{\delta} | \underset{\sim}{\theta}) + \log \pi(\underset{\sim}{d} | \underset{\sim}{D})$ and

$$\frac{\partial p_o(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} = t(\underset{\sim}{Y}^*) - \frac{\partial}{\partial \underset{\sim}{\theta}} \log a(\underset{\sim}{\theta}) + \frac{\partial}{\partial \underset{\sim}{\theta}} \log \pi(\underset{\sim}{d} | \underset{\sim}{D})$$

$$= t(\underset{\sim}{Y}^*) - E[\, t(\underset{\sim}{Y}^*) | \underset{\sim}{\theta}] + \frac{\partial}{\partial \underset{\sim}{\theta}} \log \pi(\underset{\sim}{d} | \underset{\sim}{D}).$$

For a right censored sample ($\underset{\sim}{Y}$, $\underset{\sim}{\delta}$) the function

$$\begin{aligned} Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) &= E[l_o(\underset{\sim}{\theta}_1 | \underset{\sim}{Y}^*) + \log \pi(\underset{\sim}{d}_1 | \underset{\sim}{D}_1) | \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] \\ &= E[\, \underset{\sim}{\theta}_1' t(\underset{\sim}{Y}^*) - \log a(\underset{\sim}{\theta}_1) | \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] + \log b(\underset{\sim}{Y}^*) + \log \pi(\underset{\sim}{d}_1 | \underset{\sim}{D}_1) \\ &= E[\, \underset{\sim}{\theta}_1' t(\underset{\sim}{Y}^*) | \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] - \log a(\underset{\sim}{\theta}_1) + \log b(\underset{\sim}{Y}^*) + \log \pi(\underset{\sim}{d}_1 | \underset{\sim}{D}_1). \end{aligned}$$

For a concave function p_o, Q is maximized with respect to $\underset{\sim}{\theta}_1$ when

$$0 = \frac{\partial Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}_1} = E[\, t(\underset{\sim}{Y}^*) | \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] - \frac{\partial \log a(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1} + \frac{\partial}{\partial \underset{\sim}{\theta}_1} \log \pi(\underset{\sim}{d}_1 | \underset{\sim}{D}_1)$$

$$=E[\,t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\,] - E[\,t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}_1] + \frac{\partial}{\partial \underset{\sim}{\theta}_1} \log \pi(\underset{\sim}{d}_1 \mid \underset{\sim}{D}_1)$$

or
$$E[\,t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}\,] = E[\,t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}_1] - \frac{\partial}{\partial \underset{\sim}{\theta}_1} \log \pi(\underset{\sim}{d}_1 \mid \underset{\sim}{D}_1). \qquad (3.1.3)$$

The corresponding E and M steps are:

E-step: Compute $\tilde{t}(\underset{\sim}{Y}^*) = E[\,t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}; \underset{\sim}{\theta}\,]$

M-step: Obtain $\underset{\sim}{\theta}_1$ as the solution to $E[\,t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}_1] - \frac{\partial}{\partial \underset{\sim}{\theta}_1} \log \pi(\underset{\sim}{d}_1 \mid \underset{\sim}{D}_1) = \tilde{t}(\underset{\sim}{Y}^*)$.

In the r-th iteration these correspond to:

E-step: Compute $t^{(r)}(\underset{\sim}{Y}^*) = E[\,t(\underset{\sim}{Y}^*) \mid \underset{\sim}{Y}, \underset{\sim}{\delta}, \underset{\sim}{\theta}^{(r-1)}]$

M-step: Obtain $\underset{\sim}{\theta}^{(r)}$ as the solution to

$$E[\,t(\underset{\sim}{Y}^*) \mid \underset{\sim}{\theta}^{(r)}] - \frac{\partial}{\partial \underset{\sim}{\theta}^{(r)}} \log \pi(\underset{\sim}{d}^{(r)} \mid \underset{\sim}{D}^{(r)}) = t^{(r)}(\underset{\sim}{Y}^*). \qquad (3.1.4)$$

By conditioning on both the fixed and random effects,

$(\underset{\sim}{Y}_i^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i) \sim N(\underset{\sim}{X}_i \underset{\sim}{\beta} + \underset{\sim}{B}_i \underset{\sim}{d}_i, \underset{\sim}{I}_{n_i} \sigma_e^2)$.

In the expectation step,

$E[\,t(Y_{ij}^*) \mid Y_{ij} = y_{ij}, \delta_{ij}, \underset{\sim}{\theta}\,] =$

$\delta_{ij} E[\,t(Y_{ij}^*) \mid Y_{ij} = y_{ij}, \delta_{ij} = 1, \underset{\sim}{\theta}\,] + (1 - \delta_{ij}) E[\,t(Y_{ij}^*) \mid Y_{ij} = y_{ij}, \delta_{ij} = 0, \underset{\sim}{\theta}\,]$

$= \delta_{ij} E[\,t(Y_{ij}^*) \mid Y_{ij}^* = y_{ij}, \underset{\sim}{\theta}\,] + (1 - \delta_{ij}) E[\,t(Y_{ij}^*) \mid Y_{ij}^* > y_{ij}, \underset{\sim}{\theta}\,]$

$= \delta_{ij} t(y_{ij}) + (1 - \delta_{ij}) E[\,t(Y_{ij}^*) \mid Y_{ij}^* > y_{ij}; \underset{\sim}{\theta}\,]$

with corresponding density function

$$f_0(Y^* \mid Y^*>y)=f_0(Y^* \mid Y=y,\ \delta=0)$$

$$=\frac{f_0(Y^*=y^*,\ Y=y,\ \delta=0)}{f_0(\ Y=y,\ \delta=0)}=\frac{f_0(Y^*=y^*,\ C^*=y)}{\int_y^{\infty} f_o(Y^*,\ C^*)\ dY^*}\ .$$

The last step follows from an argument analogous to the one in Section 1.1.

If Y^* and C^* are independent,

$$f_0(Y^* \mid Y^*>y)=\frac{g_{C^*}(y)\ f_o(Y^*)}{g_{C^*}(y)\ \int_y^{\infty} f_o(Y^*)\ dY^*}=\frac{f_o(Y^*)}{1-\Phi(z)}$$

where $z_{ij}=\dfrac{y_{ij}-\underset{\sim}{X}_{ij}\underset{\sim}{\beta}-\underset{\sim}{B}_{ij}\ \underset{\sim}{d}_i}{\sigma_e}$.

Therefore

$E[\ t(Y^*_{ij}) \mid Y^*_{ij}>y_{ij},\ \underset{\sim}{\theta}\,]$

$$=\frac{1}{1-\Phi\left(\dfrac{y_{ij}-\underset{\sim}{X}_{ij}\underset{\sim}{\beta}-\underset{\sim}{B}_{ij}\ \underset{\sim}{d}_i}{\sigma_e}\right)}\int_{y_{ij}}^{\infty} t(Y^*_{ij})\ f_o(Y^*_{ij} \mid \underset{\sim}{\theta}\,)\ dY^*_{ij}.$$

Recall that the likelihood function for the observed data $(\underset{\sim}{Y},\underset{\sim}{\delta}\,)$ can be expressed as

$l(\underset{\sim}{\theta}_1 \mid \underset{\sim}{Y},\underset{\sim}{\delta}\,)=l_o(\underset{\sim}{\theta}_1,\ \underset{\sim}{Y}^*)-l_1(\underset{\sim}{\theta}_1,\ \underset{\sim}{Y}^* \mid \underset{\sim}{Y},\underset{\sim}{\delta}\,)$

where l_o is the log likelihood for the complete data $\underset{\sim}{Y}^*$ that would be observed if there was no censoring and l_1 is the log likelihood of the conditional distribution of $\underset{\sim}{Y}^*$ given $(\underset{\sim}{Y},\underset{\sim}{\delta}\,)$.

When $(\underset{\sim}{Y}^*_i \mid \underset{\sim}{\beta},\ \underset{\sim}{d}_i) \sim N(\underset{\sim}{X}_i\underset{\sim}{\beta}+\underset{\sim}{B}_i\underset{\sim}{d}_i,\ \underset{\sim}{I}\,\sigma_e^2)$,

$$P(\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2 \mid \underset{\sim}{Y}, \underset{\sim}{\delta})$$

$$\propto \prod_{i=1}^{K} \left\{ \prod_{j=1}^{n_i} [\, f_o(Y_{ij}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i, \underset{\sim}{D}, \sigma_e^2)]^{\delta_{ij}} \times \left[1-\Phi\left(\frac{y_{ij} - \underset{\sim}{X}_{ij}\underset{\sim}{\beta} - \underset{\sim}{B}_{ij}\,\underset{\sim}{d}_i}{\sigma_e}\right)\right]^{1-\delta_{ij}} \right\}$$

$$\times \; \pi(\underset{\sim}{d}_i \mid \underset{\sim}{D}) \; .$$

Therefore $p(\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2 \mid \underset{\sim}{Y}, \underset{\sim}{\delta})$

$$\propto \sum_{i=1}^{K} \left\{ \sum_{j=1}^{n_i} \delta_{ij} \log f_o(Y_{ij}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i, \sigma_e^2) \right.$$

$$\left. + \sum_{j=1}^{n_i} (1-\delta_{ij}) \log\left[1-\Phi\left(\frac{y_{ij} - \underset{\sim}{X}_{ij}\underset{\sim}{\beta} - \underset{\sim}{B}_{ij}\,\underset{\sim}{d}_i}{\sigma_e}\right)\right] + \log \pi(\underset{\sim}{d}_i \mid \underset{\sim}{D}) \right\}$$

$$= \sum_{i=1}^{K} \sum_{j=1}^{n_i} \log f_o(Y_{ij}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i, \sigma_e^2) - \sum_{i=1}^{K} \sum_{j=1}^{n_i} (1-\delta_{ij}) \log \frac{f_o(Y_{ij}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i, \sigma_e^2)}{\left[1-\Phi\left(\frac{y_{ij} - \underset{\sim}{X}_{ij}\underset{\sim}{\beta} - \underset{\sim}{B}_{ij}\,\underset{\sim}{d}_i}{\sigma_e}\right)\right]}$$

$$+ \sum_{i=1}^{K} \log \pi(\underset{\sim}{d}_i \mid \underset{\sim}{D})$$

where $\dfrac{f_o(Y_{ij}^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i, \sigma_e^2)}{\left[1-\Phi\left(\frac{y_{ij} - \underset{\sim}{X}_{ij}\underset{\sim}{\beta} - \underset{\sim}{B}_{ij}\,\underset{\sim}{d}_i}{\sigma_e}\right)\right]}$

is the conditional density of Y_{ij}^* given $Y_{ij}^* > y_{ij}$ with the same form as the unconditional density of Y_{ij}^* except that the range of the density is restricted to $Y_{ij}^* > y_{ij}$.

As before

$$\frac{\partial}{\partial \underset{\sim}{\theta}_{ij}} \log \frac{f_o(Y^*_{ij}|\ \underset{\sim}{\theta}_i)}{[1-F(y_{ij}|\ \underset{\sim}{\theta}_i)\ ]} = t(Y^*_{ij}) - E[\ t(Y^*_{ij})\ |\ Y^*_{ij}>y_{ij}, \underset{\sim}{\theta}_i].$$

Therefore $\frac{\partial p(\underset{\sim}{\theta}\,;\, \underset{\sim}{Y}, \underset{\sim}{\delta}\,)}{\partial \underset{\sim}{\theta}}$

$$=\sum_{i=1}^{K} \sum_{j=1}^{n_i} \{\ t(Y^*_{ij}) - E[\ t(Y^*_{ij})|\ \underset{\sim}{\theta}\,] + \frac{\partial}{\partial \underset{\sim}{\theta}_i}\log \pi(\underset{\sim}{d}_i|\ \underset{\sim}{D})\ \}$$

$$-\sum_{i=1}^{K} \sum_{j=1}^{n_i} (1-\delta_{ij})\{\ t(Y^*_{ij}) - E[\ t(Y^*_{ij})\ |\ Y^*_{ij}>y_{ij}, \underset{\sim}{\theta}\,]+ \frac{\partial}{\partial \underset{\sim}{\theta}_i}\log \pi(\underset{\sim}{d}_i|\ \underset{\sim}{D})\ \}$$

$$=\sum_{i=1}^{K} \sum_{j=1}^{n_i} \{\ E[\ t(Y^*_{ij})\ |\ \underset{\sim}{Y}=\underset{\sim}{y}\,, \underset{\sim}{\delta}\,, \underset{\sim}{\theta}\,] - E[\ t(Y^*_{ij})|\ \underset{\sim}{\theta}\,]+ \frac{\partial}{\partial \underset{\sim}{\theta}_i}\log \pi(\underset{\sim}{d}_i|\ \underset{\sim}{D})\ \}$$

the difference between the conditional and unconditional expectations of the complete-data sufficient statistics plus the partial derivative of the log prior distribution for the random effects ($\underset{\sim}{d}$). When $\underset{\sim}{\theta}_1=\underset{\sim}{\theta}=\hat{\underset{\sim}{\theta}}$ in (3.1.3) then $\hat{\underset{\sim}{\theta}}$ is also a solution of the

equation $\frac{\partial p(\underset{\sim}{\theta}\,;\, \underset{\sim}{Y}, \underset{\sim}{\delta}\,)}{\partial \underset{\sim}{\theta}}=\underset{\sim}{0}$.

3.1.3 Mixed Models with Noninformative Right Censoring; EM Computations

EM Computations are summarized in this section. In the E step, the conditional expected values of the 'complete-data' sufficient statistic are computed from the observed data and current estimates of the parameters, while in the M step, new estimates of the unknown parameters are computed using the conditional expected values of the 'complete-data' sufficient statistics in the maximum a posteriori estimating equations.

E Step:

In the r-th iteration, the estimation step computes the conditional expected values of the complete-data sufficient statistics given the observed data $\underset{\sim}{Y}$ and the estimated values of the parameters from the (r − 1)-st iteration. Not all of the $\underset{\sim}{Y}^*$ are observed so the E step will estimate the complete-data sufficient statistics that involve $\underset{\sim}{Y}^*$. Note that the random individual parameters $\underset{\sim}{d}_i$ are estimated in the M step, not the E step. This is in contrast to the usual application of the EM algorithm for situations where all of the $\underset{\sim}{Y}_i^*$ are known and complete-data sufficient statistics involving $\underset{\sim}{d}_i$ are estimated in the E step. By conditioning on both the fixed and random effects

$$(\underset{\sim}{Y}_i^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i) \sim N(\underset{\sim}{X}_i\underset{\sim}{\beta} + \underset{\sim}{B}_i\underset{\sim}{d}_i, \underset{\sim}{I}_{n_i}\sigma_e^2)$$

and the $(\underset{\sim}{Y}_i^* \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i)$, i=1, 2, ..., N, are conditionally independent, eliminating the need to compute multi-dimensional integrals.

A set of complete-data sufficient statistics for this problem is

$\{\underset{\sim}{X}_i'\underset{\sim}{Y}_i^*\}_{i=1}^K$, $\{\underset{\sim}{B}_i'\underset{\sim}{Y}_i^*\}_{i=1}^K$, and $\{\underset{\sim}{Y}_i^{*\prime}\underset{\sim}{Y}_i^*\}_{i=1}^K$.

The expected values of the complete-data sufficient statistics may be denoted,

$\underset{\sim}{t}_{1i}^{(r)}=\underset{\sim}{X}_i' \, E[\, \underset{\sim}{Y}_i^* \mid \underset{\sim}{Y}_i, \underset{\sim}{\delta}_i, \underset{\sim}{\beta}^{(r-1)}, \underset{\sim}{d}_i^{(r-1)}, \underset{\sim}{D}^{(r-1)}, \sigma_e^{(r-1)^2}]$

$\underset{\sim}{t}_{2i}^{(r)}=\underset{\sim}{B}_i' \, E[\, \underset{\sim}{Y}_i^* \mid \underset{\sim}{Y}_i, \underset{\sim}{\delta}_i, \underset{\sim}{\beta}^{(r-1)}, \underset{\sim}{d}_i^{(r-1)}, \underset{\sim}{D}^{(r-1)}, \sigma_e^{(r-1)^2}]$ and

$t_{3i}^{(r)}=E[\, \underset{\sim}{Y}_i^{*\prime}\underset{\sim}{Y}_i^* \mid \underset{\sim}{Y}_i, \underset{\sim}{\delta}_i, \underset{\sim}{\beta}^{(r-1)}, \underset{\sim}{d}_i^{(r-1)}, \underset{\sim}{D}^{(r-1)}, \sigma_e^{(r-1)^2}]$.

When $\delta_{ij}=0$, $Y_{ij}^*>Y_{ij}=C_{ij}^*$ and expectations involving the (i,j)-th element of the complete-data sufficient statistics can be computed as:

$$E[\, Y_{ij}^* \mid Y_{ij}=y_{ij}, \delta_{ij}=0, \underset{\sim}{\beta}^{(r-1)}, \underset{\sim}{d}_i^{(r-1)}, \underset{\sim}{D}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=E[\, Y_{ij}^* \mid Y_{ij}^*>y_{ij}, \underset{\sim}{\beta}^{(r-1)}, \underset{\sim}{d}_i^{(r-1)}, \underset{\sim}{D}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=\int_{y_{ij}}^{\infty} Y_{ij}^* \frac{f_o(\, Y_{ij}^* \mid \underset{\sim}{\beta}^{(r-1)}, \underset{\sim}{d}_i^{(r-1)}, \underset{\sim}{D}^{(r-1)}, \sigma_e^{(r-1)^2})}{1-\Phi(z_{ij}^{(r-1)})} \, dY_{ij}^*$$

$$=\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}_{ij}\,\underset{\sim}{d}_i^{(r-1)}+\sigma_e^{(r-1)} \frac{\phi(z_{ij}^{(r-1)})}{1-\Phi(z_{ij}^{(r-1)})}$$

where $z_{ij}^{(r-1)}=\dfrac{y_{ij}-\left[\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}_{ij}\,\underset{\sim}{d}_i^{(r-1)}\right]}{\sigma_e^{(r-1)}}$, and

$$E[\, Y_{ij}^{*2} \mid Y_{ij}=y_{ij}, \delta_{ij}=0, \underset{\sim}{\beta}^{(r-1)}, \underset{\sim}{d}_i^{(r-1)}, \underset{\sim}{D}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=E[\, Y_{ij}^{*2} \mid Y_{ij}^*>y_{ij}, \underset{\sim}{\beta}^{(r-1)}, \underset{\sim}{d}_i^{(r-1)}, \underset{\sim}{D}^{(r-1)}, \sigma_e^{(r-1)^2}]$$

$$=\int_{y_{ij}}^{\infty} Y_{ij}^{*2} \frac{f_o(Y_{ij}^{*} \mid \underset{\sim}{\beta}^{(r-1)}, \underset{\sim}{d}_i^{(r-1)}, \underset{\sim}{D}^{(r-1)}, \sigma_e^{(r-1)^2})}{1-\Phi(z_{ij}^{(r-1)})} dY_{ij}^{*}$$

$$=[\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}_{ij}\ \underset{\sim}{d}_i^{(r-1)}]^2+\sigma_e^{(r-1)^2}$$

$$+\sigma_e^{(r-1)}(y_{ij}+\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}_{ij}\ \underset{\sim}{d}_i^{(r-1)})\ \frac{\phi(z_{ij}^{(r-1)})}{1-\Phi(z_{ij}^{(r-1)})}\ .$$

<u>M Step</u>:

The r-th iteration of the M step obtains $\underset{\sim}{\beta}^{(r)}$, $\underset{\sim}{d}^{(r)}$, $\underset{\sim}{D}^{(r)}$, and $\sigma_e^{(r)^2}$ as the solution to equation 3.1.4. Initial values of maximum a posteriori estimates are obtained using unweighted regression and by treating censored data as if they were uncensored. Convenient initial estimates are:

$$\underset{\sim}{\beta}^{(0)}=\left[\sum_{i=1}^{K}\underset{\sim}{X}_i'\underset{\sim}{X}_i\right]^{-1}\sum_{i=1}^{K}\underset{\sim}{X}_i'\underset{\sim}{Y}_i,$$

$$\underset{\sim}{d}_i^{(0)}=[\underset{\sim}{B}_i'\underset{\sim}{B}_i]^{-1}\underset{\sim}{B}_i'(\underset{\sim}{Y}_i-\underset{\sim}{X}_i\underset{\sim}{\beta}^{(0)}),$$

$$\underset{\sim}{D}^{(0)}=\frac{1}{K}\sum_{i=1}^{K}\underset{\sim}{d}_i^{(0)}\underset{\sim}{d}_i^{(0)\prime},$$

$$\sigma_e^{(0)^2}=\frac{1}{n}\left[\sum_{i=1}^{K}(\underset{\sim}{Y}_i^{*}-\underset{\sim}{X}_i\underset{\sim}{\beta}^{(0)})'\ (\underset{\sim}{Y}_i^{*}-\underset{\sim}{X}_i\underset{\sim}{\beta}^{(0)})\right],$$

and if $\underset{\sim}{D}=\sum_{s=1}^{m}\sigma_s\underset{\sim}{S}_s$ then

$$\underset{\sim}{\sigma}^{(0)}=\underset{\sim}{1}\ .$$

In the r-th iteration maximum a posteriori estimates which maximize the posterior distribution function using the expected values of the complete-data sufficient statistics obtained from the previous iteration of the E step are computed:

$$\underset{\sim}{\beta}^{(r)} = \left[\sum_{i=1}^{K} \underset{\sim}{X}_i' \underset{\sim}{X}_i\right]^{-1} \sum_{i=1}^{K} (\underset{\sim}{t}_{1i}^{(r)} - \underset{\sim}{X}_i' \underset{\sim}{B}_i \underset{\sim}{d}_i^{(r-1)}),$$

$$\underset{\sim}{d}_i^{(r)} = [\sigma_e^{(r-1)^2} \underset{\sim}{D}^{(r-1)^{-1}} + \underset{\sim}{B}_i' \underset{\sim}{B}_i]^{-1} (\underset{\sim}{t}_{2i}^{(r)} - \underset{\sim}{B}_i' \underset{\sim}{X}_i \underset{\sim}{\beta}^{(r-1)}),$$

$$\underset{\sim}{D}^{(r)} = \frac{1}{K} \sum_{i=1}^{K} \underset{\sim}{d}_i^{(r)} \underset{\sim}{d}_i^{(r)'},$$

and

$$\sigma_e^{(r)^2} = \frac{1}{n} \sum_{i=1}^{K} \left[t_{3i}^{(r)} - 2(\underset{\sim}{\beta}^{(r-1)'} \underset{\sim}{t}_{1i}^{(r)} + \underset{\sim}{d}_i^{(r-1)'} \underset{\sim}{t}_{2i}^{(r)}) + \underset{\sim}{d}_i^{(r-1)'} \underset{\sim}{B}_i' \underset{\sim}{B}_i \underset{\sim}{d}_i^{(r-1)}\right].$$

If $\underset{\sim}{D} = \sum_{s=1}^{m} \tau_s \underset{\sim}{S}_s$ then

$$\underset{\sim}{\tau}^{(r)} = \frac{1}{K}\left[< \text{trace } (\underset{\sim}{D}^{(r-1)^{-1}} \underset{\sim}{S}_s \underset{\sim}{D}^{(r-1)^{-1}} \underset{\sim}{S}_h) >_{sh}\right]^{-1}$$

$$\times \left[< \sum_{i=1}^{K} \underset{\sim}{d}_i^{(r-1)'} \underset{\sim}{D}^{(r-1)^{-1}} \underset{\sim}{S}_s \underset{\sim}{D}^{(r-1)^{-1}} \underset{\sim}{d}_i^{(r-1)} >_s\right]$$

(For details, see Fairclough and Helms 1984).

3.2 Mixed Models with Informative Right Censoring

Informative censoring is discussed in Section 3.2. Likelihood functions for complete data are derived in Section 3.2.1, and theory and applications of the EM algorithm to mixed models with random informative right censoring are discussed in Sections 3.2.2 and 3.2.3.

3.2.1 Mixed Models with Informative Right Censoring; Likelihood Functions

Consider a random sample of K individuals with n_i observations for the i-th subject such that $n=\sum_{i=1}^{K} n_i$. The n observations are assumed to be a sample from a normal population for which the corresponding the General Linear Mixed Model is

$$\underset{\sim}{W}^*=\begin{bmatrix} \underset{\sim}{Y}^* \\ \underset{\sim}{C}^* \end{bmatrix}=\begin{bmatrix} \underset{\sim}{X} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{H} \end{bmatrix}\begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\alpha} \end{bmatrix}+\begin{bmatrix} \underset{\sim}{B} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{J} \end{bmatrix}\begin{bmatrix} \underset{\sim}{d} \\ \underset{\sim}{v} \end{bmatrix}+\begin{bmatrix} \underset{\sim}{e} \\ \underset{\sim}{\varepsilon} \end{bmatrix}$$

$$=\underset{\sim}{A}\underset{\sim}{\gamma}+\underset{\sim}{M}\underset{\sim}{g}+\underset{\sim}{\nu},$$

where

$\underset{\sim}{W}^*$ is a 2n × 1 vector of failure values and censoring values which may or may not

be observed,

$\underset{\sim}{A}=\begin{bmatrix} \underset{\sim}{X} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{H} \end{bmatrix}$ is a 2n × 2p known constant matrix of rank $2r \leq 2p$,

$\underset{\sim}{\gamma}=\begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\alpha} \end{bmatrix}$ is a 2p × 1 vector of unknown constant 'fixed' population parameters,

$\underset{\sim}{M}=\begin{bmatrix} \underset{\sim}{B} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{J} \end{bmatrix}$ is a 2n $\times$ 2q known matrix corresponding to the random effects,

$\underset{\sim}{g}=\begin{bmatrix} \underset{\sim}{d} \\ \underset{\sim}{v} \end{bmatrix}$ is a 2Kq $\times$ 1 vector of unobservable individual parameters,

$\underset{\sim}{\nu}=\begin{bmatrix} \underset{\sim}{e} \\ \underset{\sim}{\epsilon} \end{bmatrix}$ is a 2n $\times$ 1 vector of unobservable random errors,

$\underset{\sim}{g}_i \sim N(\underset{\sim}{0}, \underset{\sim}{G})$ independent of $\underset{\sim}{\nu}_i \sim N(\underset{\sim}{0}, \underset{\sim}{\Omega} \otimes \underset{\sim}{I}_{n_i})$,

$\underset{\sim}{G}$ is a positive-definite symmetric 2q $\times$ 2q covariance matrix of random effects, $\underset{\sim}{g}_i$, and

$\underset{\sim}{\Omega}=\begin{bmatrix} \sigma_e^2 & \rho\sigma_e\sigma_\epsilon \\ \rho\sigma_e\sigma_\epsilon & \sigma_\epsilon^2 \end{bmatrix}$ consists of the unknown within-subject variance components,

where

$$\underset{\sim}{\Omega} \otimes \underset{\sim}{I}_{n_i}=\begin{bmatrix} \underset{\sim}{I}_{n_i}\sigma_e^2 & \underset{\sim}{I}_{n_i}\rho\sigma_e\sigma_\epsilon \\ \underset{\sim}{I}_{n_i}\rho\sigma_e\sigma_\epsilon & \underset{\sim}{I}_{n_i}\sigma_\epsilon^2 \end{bmatrix}.$$

Define: $\delta_{1ij}=\mathfrak{B}(Y_{ij}=Y^*_{ij})$, $\delta_{2ij}=\mathfrak{B}(Y_{ij}=C^*_{ij})$, and $\underset{\sim}{W}$ to be the observed values of $\underset{\sim}{W}^*$, where $\mathfrak{B}$ is the Boolean function (Helms 1988).

If there is no censoring, the joint bivariate pdf normal of $\underset{\sim}{W}^*$ given $\underset{\sim}{\gamma}$ and $\underset{\sim}{g}$ is:

$$f(\underset{\sim}{W}^* \mid \underset{\sim}{\gamma}, \underset{\sim}{g}, \underset{\sim}{\Omega})$$

$$= \prod_{i=1}^{K} f(\underset{\sim}{W}_i^* \mid \underset{\sim}{\gamma}, \underset{\sim}{g}_i, \underset{\sim}{\Omega})$$

$$= \prod_{i=1}^{K} \left[\frac{1}{2\pi}\right]^{n_i} \mid \underset{\sim}{\Omega} \mid^{-\frac{n}{2}} \times \exp\left[-\frac{1}{2}(\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)'(\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i})(\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)\right]$$

$$= \left[\frac{1}{2\pi}\right]^{n} \mid \underset{\sim}{\Omega} \mid^{-\frac{n}{2}}$$

$$\times \exp\left[-\frac{1}{2}\sum_{i=1}^{K} (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)'(\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i})(\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)\right]. \quad (3.2.1)$$

Suitable flat prior distributions for this problem are:

$\Pi_1(\underset{\sim}{\gamma}) \propto$ constant,

$\Pi_2(\underset{\sim}{G}) \propto$ constant,

$\Pi_3(\sigma_e^2) \propto$ constant,

$\Pi_4(\sigma_\epsilon^2) \propto$ constant,

$\Pi_5(\rho) \propto$ constant,

and a convenient prior for $\underset{\sim}{g}_i$ is

$\Pi_6(\underset{\sim}{g}_i \mid \underset{\sim}{G}) = N_q(\underset{\sim}{0}, \underset{\sim}{G})$.

This leads to the joint prior distribution

$$\Pi(\underset{\sim}{\gamma}, \underset{\sim}{g}, \underset{\sim}{G}, \sigma_e^2, \sigma_\epsilon^2, \rho)$$

$$\propto \Pi_1(\underset{\sim}{\gamma}) \cdot \Pi_3(\sigma_e^2) \cdot \Pi_4(\sigma_\epsilon^2) \cdot \Pi_5(\rho) \cdot \prod_{i=1}^{K} \Pi_6(\underset{\sim}{g}_i \mid \underset{\sim}{G}) \cdot \Pi_2(G) \,. \quad (3.2.2)$$

Let $\underset{\sim}{\theta} = \{\underset{\sim}{\gamma}, \underset{\sim}{g}, \underset{\sim}{G}, \sigma_e^2, \sigma_\epsilon^2, \rho\}$ denote the parameter vector that is restricted to a (p+1)-dimensional convex set Ψ. The joint p.d.f. of the distribution of W^* and θ is equal to the product of the conditional density in (3.2.1) and the joint prior distribution in (3.2.2), i.e.

$p(W^*, \theta) = f(W^* \mid \theta) \times \Pi(\theta)$.

Using Bayes Theorem

$$p(\theta \mid W^*) = \frac{p(W^*, \theta)}{\int_{\theta} p(W^*, \theta)\, d\theta} = \frac{p(W^*, \theta)}{p(W^*)}.$$

Since $p(W^*)$ does not depend on any of the parameters, maximizing $p(\theta \mid W^*)$ is equivalent to maximizing $p(W^*, \theta)$ with respect to θ.

Therefore $p(\theta \mid W^*) = \dfrac{p(W^*, \theta)}{p(W^*)} \propto p(W^*, \theta)$

$$= |\Omega|^{-\frac{n}{2}} \, |G|^{-\frac{K}{2}}$$

$$\times \exp\left\{-\frac{1}{2} \sum_{i=1}^{K} \left[(W_i^* - A_i\gamma - M_i g_i)' (\Omega^{-1} \otimes I_{n_i}) (W_i^* - A_i\gamma - M_i g_i) + g_i' G^{-1} g_i \right] \right\}.$$

Therefore the logarithm of the posterior distribution for the parameters γ, g, G, σ_e^2, σ_ϵ^2, and ρ is:

$$p(\gamma, g, G, \sigma_e^2, \sigma_\epsilon^2, \rho \mid W^*)$$

$$\propto -\frac{1}{2} \left\{ n \log |\Omega| + K \log |G| \right.$$

$$+\sum_{i=1}^{K}\left[\ (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)' (\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i) + \underset{\sim}{g}_i'\underset{\sim}{G}^{-1}\underset{\sim}{g}_i\ \right] \Big\}.$$

The <u>maximum</u> <u>a</u> <u>posterior</u> <u>estimators</u> (Beck and Arnold 1977) of $\underset{\sim}{\gamma}$, $\underset{\sim}{g}$, $\underset{\sim}{G}$, σ_e^2, σ_ϵ^2, and ρ are the coordinates of the mode of the posterior distribution, treated as an analog of a likelihood function (i.e., the parameters are variables and the data are constants). Posterior mode estimators are equivalent to maximum likelihood estimates for parameters with flat priors (Laird and Ware 1982), but in this case we do not have flat priors for the random effects.

As in Section 3.1 the log likelihood function can also be derived using a frequentist approach for an artificial General Mixed Model which treats the unobserved $\underset{\sim}{g}_i$'s as missing data, thus facilitating the use of the EM algorithm. The model is defined as:

$$\begin{bmatrix} \underset{\sim}{W}_i^* \\ \underset{\sim}{g}_i \end{bmatrix} = \begin{bmatrix} \underset{\sim}{A}_i\underset{\sim}{\gamma} \\ \underset{\sim}{0} \end{bmatrix} + \begin{bmatrix} \underset{\sim}{M}_i \\ \underset{\sim}{I}_q \end{bmatrix} \underset{\sim}{g}_i + \begin{bmatrix} \underset{\sim}{\nu}_i \\ \underset{\sim}{0} \end{bmatrix}.$$

Let $\underset{\sim}{\Sigma}_{\underset{\sim}{W}_i^*} = \underset{\sim}{M}_i'\underset{\sim}{G}\underset{\sim}{M}_i + \underset{\sim}{\Omega} \otimes \underset{\sim}{I}_{n_i}$.

The log of the likelihood function can then be written as:

$$l_o(\underset{\sim}{\gamma}, \underset{\sim}{g}, \underset{\sim}{G}, \sigma_e^2, \sigma_\epsilon^2, \rho) = -\frac{1}{2}\Big\{ (2n+2Kq)\log(2\pi) + \sum_{i=1}^{K} \log \begin{vmatrix} \underset{\sim}{\Sigma}_{\underset{\sim}{W}_i^*} & \underset{\sim}{M}_i\underset{\sim}{G} \\ \underset{\sim}{G}\underset{\sim}{M}_i' & \underset{\sim}{G} \end{vmatrix}$$

$$+\sum_{i=1}^{K} \begin{bmatrix} \underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} \\ \underset{\sim}{g}_i - \underset{\sim}{0} \end{bmatrix}' \begin{bmatrix} \underset{\sim}{\Sigma}_{\underset{\sim}{W}_i^*} & \underset{\sim}{M}_i\underset{\sim}{G} \\ \underset{\sim}{G}\underset{\sim}{M}_i' & \underset{\sim}{G} \end{bmatrix}^{-1} \begin{bmatrix} \underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} \\ \underset{\sim}{g}_i - \underset{\sim}{0} \end{bmatrix} \Big\}$$

$$= -\frac{1}{2}\Big\{ (2n+2Kq)\log(2\pi) + n\log|\Omega| + K\log|G|$$

$$+\sum_{i=1}^{K}\begin{bmatrix} W_i^* - A_i\gamma \\ g_i - 0 \end{bmatrix}'\begin{bmatrix} \Sigma_{W_i^*} & M_i G \\ G M_i' & G \end{bmatrix}^{-1}\begin{bmatrix} W_i^* - A_i\gamma \\ g_i - 0 \end{bmatrix}\Big\}.$$

$$\begin{bmatrix} \Sigma_{W_i^*} & M_i G \\ G M_i' & G \end{bmatrix}^{-1} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}^{-1} = H^{-1}$$

$$= \begin{bmatrix} J^{-1} & E_{12} \\ E_{21} & E_{22} \end{bmatrix},$$

where $J = H_{11} - H_{12}\, H_{22}^{-1}\, H_{21} = \Sigma_{W_i^*} - M_i G G^{-1} G M_i'$

$$= \Sigma_{W_i^*} - M_i G M_i' = \Omega \otimes I_{n_i},$$

$$E_{12} = -J^{-1}\, H_{12}\, H_{22}^{-1} = [\Sigma_{W_i^*} - M_i G M_i']^{-1} M_i G G^{-1}$$
$$= [\,\Omega^{-1} \otimes I_{n_i}]\, M_i,$$

$$E_{21} = -H_{22}^{-1}\, H_{21}\, J^{-1} = M_i'[\Sigma_{W_i^*} - M_i G M_i']^{-1},$$
$$= M_i'\,[\,\Omega^{-1} \otimes I_{n_i}\,], \text{ and}$$

$$E_{22} = H_{22}^{-1} + H_{22}^{-1} H_{21}\, J^{-1} H_{12}\, H_{22}^{-1}$$
$$= G^{-1} + G^{-1}(G M_i')\,[\Sigma_{W_i^*} - M_i G M_i']^{-1}\, M_i G G^{-1}$$
$$= G^{-1} + M_i'\,[\,\Omega^{-1} \otimes I_{n_i}\,]\, M_i.$$

Therefore

$$\begin{bmatrix} \underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} \\ \underset{\sim}{g}_i - \underset{\sim}{0} \end{bmatrix}' \begin{bmatrix} \Sigma_{\underset{\sim}{W}_i^*} & \underset{\sim}{M}_i\underset{\sim}{G} \\ \underset{\sim}{G}\underset{\sim}{M}_i' & \underset{\sim}{G} \end{bmatrix}^{-1} \begin{bmatrix} \underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} \\ \underset{\sim}{g}_i - \underset{\sim}{0} \end{bmatrix}$$

$$= \begin{bmatrix} \underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} \\ \underset{\sim}{g}_i \end{bmatrix}' \begin{bmatrix} [\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i}] & -[\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i}]\underset{\sim}{M}_i \\ -\underset{\sim}{M}_i'[\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i}] & \underset{\sim}{G}^{-1} + \underset{\sim}{M}_i'[\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i}]\underset{\sim}{M}_i \end{bmatrix} \begin{bmatrix} \underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} \\ \underset{\sim}{g}_i \end{bmatrix}$$

$$= (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)' (\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i) + \underset{\sim}{g}_i'\underset{\sim}{G}^{-1}\underset{\sim}{g}_i \,.$$

Therefore

$$l_o(\underset{\sim}{\gamma}, \underset{\sim}{g}, \underset{\sim}{G}, \sigma_e^2, \sigma_\epsilon^2, \rho) = -\tfrac{1}{2}\Big\{ (2n+2Kq)\log(2\pi) + n\log|\underset{\sim}{\Omega}| + K\log|\underset{\sim}{G}|$$

$$+ \sum_{i=1}^{K}\Big[(\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)' (\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i) + \sum_{i=1}^{K}\underset{\sim}{g}_i'\underset{\sim}{G}^{-1}\underset{\sim}{g}_i \Big] \Big\}.$$

Therefore the posterior distribution of the parameters $\underset{\sim}{\gamma}$, $\underset{\sim}{g}$, $\underset{\sim}{G}$, σ_e^2, σ_ϵ^2, and ρ given $\underset{\sim}{W}^*$, assuming flat priors for fixed effects and variance components and normal priors for random effects is proportional to the likelihood of $\underset{\sim}{\gamma}$, $\underset{\sim}{G}$, σ_e^2, σ_ϵ^2, and ρ given $\underset{\sim}{W}^*$ and $\underset{\sim}{g}$.

<u>Theorem</u> <u>3.2.1</u>: Maximum a posteriori estimates of the parameters $\underset{\sim}{\gamma}$, $\underset{\sim}{g}_i$, $\underset{\sim}{G}$, σ_e^2, σ_ϵ^2, and ρ satisfy the following equations:

$$\hat{\underset{\sim}{\gamma}} = \begin{bmatrix} \hat{\underset{\sim}{\beta}} \\ \hat{\underset{\sim}{\alpha}} \end{bmatrix} = \Big[\sum_{i=1}^{K} \underset{\sim}{A}_i' (\hat{\underset{\sim}{\Omega}}^{-1} \otimes \underset{\sim}{I}_{n_i}) \underset{\sim}{A}_i \Big]^{-1} \sum_{i=1}^{K} \underset{\sim}{A}_i' (\hat{\underset{\sim}{\Omega}}^{-1} \otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{M}_i\hat{\underset{\sim}{g}}_i),$$

$$\hat{\underset{\sim}{g}}_i = [\underset{\sim}{M}_i' (\hat{\underset{\sim}{\Omega}}^{-1} \otimes \underset{\sim}{I}_{n_i}) \underset{\sim}{M}_i + \hat{\underset{\sim}{G}}^{-1}]^{-1} \underset{\sim}{M}_i' (\hat{\underset{\sim}{\Omega}}^{-1} \otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\hat{\underset{\sim}{\gamma}}),$$

$$\hat{\sigma}_e^2=\frac{1}{n}(\underset{\sim}{Y}^* - \underset{\sim}{X}\hat{\underset{\sim}{\beta}} - \underset{\sim}{B}\hat{\underset{\sim}{d}})'(\underset{\sim}{Y}^* - \underset{\sim}{X}\hat{\underset{\sim}{\beta}} - \underset{\sim}{B}\hat{\underset{\sim}{d}}),$$

$$\hat{\sigma}_\epsilon^2=\frac{1}{n}(\underset{\sim}{C}^* - \underset{\sim}{H}\hat{\underset{\sim}{\alpha}} - \underset{\sim}{J}\hat{\underset{\sim}{v}})'(\underset{\sim}{C}^* - \underset{\sim}{H}\hat{\underset{\sim}{\alpha}} - \underset{\sim}{J}\hat{\underset{\sim}{v}}),$$

and

$$\hat{\rho}=\frac{1}{n\hat{\sigma}_e\hat{\sigma}_\epsilon}(\underset{\sim}{Y}^* - \underset{\sim}{X}\hat{\underset{\sim}{\beta}} - \underset{\sim}{B}\hat{\underset{\sim}{d}})'(\underset{\sim}{C}^* - \underset{\sim}{H}\hat{\underset{\sim}{\alpha}} - \underset{\sim}{J}\hat{\underset{\sim}{v}}) .$$

<u>Proof</u>:

$$\frac{\partial p(\underset{\sim}{\gamma}, \underset{\sim}{g}, \underset{\sim}{G}, \sigma_e^2, \sigma_\epsilon^2, \rho \mid \underset{\sim}{W}^*)}{\partial \underset{\sim}{\gamma}}$$

$$= -\frac{1}{2}\frac{\partial}{\partial \underset{\sim}{\gamma}}\left[\sum_{i=1}^{K} (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)' (\underset{\sim}{\Omega}^{-1}\otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)\right]$$

$$= -\frac{1}{2}\left[-2\sum_{i=1}^{K} \underset{\sim}{A}_i' (\underset{\sim}{\Omega}^{-1}\otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)\right].$$

Therefore $\frac{\partial p(\underset{\sim}{\gamma}, \underset{\sim}{g}, \underset{\sim}{G}, \sigma_e^2, \sigma_\epsilon^2, \rho \mid \underset{\sim}{W}^*)}{\partial \underset{\sim}{\gamma}}=0$ implies that

$$\hat{\underset{\sim}{\gamma}} = \left[\sum_{i=1}^{K} \underset{\sim}{A}_i' (\hat{\underset{\sim}{\Omega}}^{-1}\otimes \underset{\sim}{I}_{n_i}) \underset{\sim}{A}_i\right]^{-1} \sum_{i=1}^{K} \underset{\sim}{A}_i' (\hat{\underset{\sim}{\Omega}}^{-1}\otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{M}_i\hat{\underset{\sim}{g}}_i) .$$

$$\frac{\partial p(\underset{\sim}{\gamma}, \underset{\sim}{g}, \underset{\sim}{G}, \sigma_e^2, \sigma_\epsilon^2, \rho \mid \underset{\sim}{W}^*)}{\partial \underset{\sim}{g}_i}$$

$$= -\frac{1}{2}\frac{\partial}{\partial \underset{\sim}{g}_i}\sum_{i=1}^{K}\left[(\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i)' (\underset{\sim}{\Omega}^{-1}\otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i\underset{\sim}{\gamma} - \underset{\sim}{M}_i\underset{\sim}{g}_i) + \underset{\sim}{g}_i'\underset{\sim}{G}^{-1}\underset{\sim}{g}_i' \right]$$

$$= -\frac{1}{2}\left[-2\,\underset{\sim}{M}_i'\,(\underset{\sim}{\Omega}^{-1}\otimes\underset{\sim}{I}_{n_i})\,(\underset{\sim}{W}_1^*-\underset{\sim}{A}_i\underset{\sim}{\gamma}-\underset{\sim}{M}_i\underset{\sim}{g}_i)+2\underset{\sim}{G}^{-1}\underset{\sim}{g}_i'\right]$$

$$=\underset{\sim}{M}_i'\,(\underset{\sim}{\Omega}^{-1}\otimes\underset{\sim}{I}_{n_i})\,(\underset{\sim}{W}_1^*-\underset{\sim}{A}_i\underset{\sim}{\gamma}-\underset{\sim}{M}_i\underset{\sim}{g}_i)-\underset{\sim}{G}^{-1}\underset{\sim}{g}_i'$$

$$=\underset{\sim}{M}_i'\,(\underset{\sim}{\Omega}^{-1}\otimes\underset{\sim}{I}_{n_i})\,(\underset{\sim}{W}_1^*-\underset{\sim}{A}_i\underset{\sim}{\gamma})-[\,\underset{\sim}{M}_i'\,(\underset{\sim}{\Omega}^{-1}\otimes\underset{\sim}{I}_{n_i})\,\underset{\sim}{M}_i+\underset{\sim}{G}^{-1}]\,\underset{\sim}{g}_i'.$$

Therefore $\dfrac{\partial p(\underset{\sim}{\gamma},\underset{\sim}{g},\underset{\sim}{G},\sigma_e^2,\sigma_\epsilon^2,\rho\mid\underset{\sim}{W}^*)}{\partial\underset{\sim}{g}_i}=0$ implies that

$$\hat{\underset{\sim}{g}}_i=[\,\underset{\sim}{M}_i'\,(\hat{\underset{\sim}{\Omega}}^{-1}\otimes\underset{\sim}{I}_{n_i})\,\underset{\sim}{M}_i+\hat{\underset{\sim}{G}}^{-1}]^{-1}\,\underset{\sim}{M}_i'\,(\hat{\underset{\sim}{\Omega}}^{-1}\otimes\underset{\sim}{I}_{n_i})\,(\,\underset{\sim}{W}_1^*-\underset{\sim}{A}_i\hat{\underset{\sim}{\gamma}}).$$

$$\frac{\partial p(\underset{\sim}{\gamma},\underset{\sim}{g},\underset{\sim}{G},\sigma_e^2,\sigma_\epsilon^2,\rho\mid\underset{\sim}{W}^*)}{\partial\underset{\sim}{\Omega}}$$

$$=-\frac{1}{2}\left\{\frac{\partial}{\partial\underset{\sim}{\Omega}}\,n\log|\underset{\sim}{\Omega}|\right.$$

$$\left.+\frac{\partial}{\partial\underset{\sim}{\Omega}}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\,\underset{\sim}{g}_i)'\,\underset{\sim}{\Omega}^{-1}\,(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\,\underset{\sim}{g}_i)\right\}$$

$$=-\frac{1}{2}\left\{2n\,\underset{\sim}{\Omega}^{-1}-n\,\mathrm{diag}(\underset{\sim}{\Omega}^{-1})\right.$$

$$-2\underset{\sim}{\Omega}^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\,\underset{\sim}{g}_i)\,(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\,\underset{\sim}{g}_i)'\,\underset{\sim}{\Omega}^{-1}$$

$$\left.+\mathrm{diag}\left[-\underset{\sim}{\Omega}^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\,\underset{\sim}{g}_i)\,(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\,\underset{\sim}{g}_i)'\,\underset{\sim}{\Omega}^{-1}\right]\right\}$$

Setting $\dfrac{\partial p(\underset{\sim}{\gamma}, \underset{\sim}{g}, \underset{\sim}{G}, \sigma_e^2, \sigma_\epsilon^2, \rho \mid \underset{\sim}{W}^*)}{\partial \underset{\sim}{\Omega}}=0$ implies that

$$2\underset{\sim}{\Omega}^{-1}\left[n-\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\underset{\sim}{g}_i)(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\underset{\sim}{g}_i)'\underset{\sim}{\Omega}^{-1}\right]$$

$$-\operatorname{diag}\left\{\underset{\sim}{\Omega}^{-1}\left[n-\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\underset{\sim}{g}_i)(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\underset{\sim}{g}_i)'\underset{\sim}{\Omega}^{-1}\right]\right\}$$

$$=0$$

$$\rightarrow \hat{\underset{\sim}{\Omega}}=\begin{bmatrix} \hat{\sigma}_e^2 & \hat{\rho}\hat{\sigma}_e\hat{\sigma}_\epsilon \\ \hat{\rho}\hat{\sigma}_e\hat{\sigma}_\epsilon & \hat{\sigma}_\epsilon^2 \end{bmatrix}$$

$$=\frac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\underset{\sim}{g}_i)(\underset{\sim}{W}_{ij}^*-\underset{\sim}{A}_{ij}\underset{\sim}{\gamma}-\underset{\sim}{M}_{ij}\underset{\sim}{g}_i)'$$

$$=\frac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\begin{bmatrix}(Y_{ij}^*-\underset{\sim}{X}_{ij}\underset{\sim}{\beta}-\underset{\sim}{B}_{ij}\underset{\sim}{d}_i) \\ (C_{ij}^*-\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}-\underset{\sim}{J}_{ij}\underset{\sim}{v}_i)\end{bmatrix}\begin{bmatrix}(Y_{ij}^*-\underset{\sim}{X}_{ij}\underset{\sim}{\beta}-\underset{\sim}{B}_{ij}\underset{\sim}{d}_i) \\ (C_{ij}^*-\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}-\underset{\sim}{J}_{ij}\underset{\sim}{v}_i)\end{bmatrix}'$$

$$\rightarrow \hat{\sigma}_e^2=\frac{1}{n}(\underset{\sim}{Y}^*-\underset{\sim}{X}\hat{\underset{\sim}{\beta}}-\underset{\sim}{B}\hat{\underset{\sim}{d}})'(\underset{\sim}{Y}^*-\underset{\sim}{X}\hat{\underset{\sim}{\beta}}-\underset{\sim}{B}\hat{\underset{\sim}{d}})$$

$$\hat{\sigma}_\epsilon^2=\frac{1}{n}(\underset{\sim}{C}^*-\underset{\sim}{H}\hat{\underset{\sim}{\alpha}}-\underset{\sim}{J}\hat{\underset{\sim}{v}})'(\underset{\sim}{C}^*-\underset{\sim}{H}\hat{\underset{\sim}{\alpha}}-\underset{\sim}{J}\hat{\underset{\sim}{v}})$$

and

$$\hat{\rho}=\frac{1}{n\hat{\sigma}_e\hat{\sigma}_\epsilon}(\underset{\sim}{Y}^*-\underset{\sim}{X}\hat{\underset{\sim}{\beta}}-\underset{\sim}{B}\hat{\underset{\sim}{d}})'(\underset{\sim}{C}^*-\underset{\sim}{H}\hat{\underset{\sim}{\alpha}}-\underset{\sim}{J}\hat{\underset{\sim}{v}}).$$

Q.E.D.

3.2.2 Mixed Models with Informative Right Censoring; Derivation of the EM Algorithm

The EM Algorithm (Dempster et al. 1977) can be used to compute maximum a posteriori estimates of $\underset{\sim}{\gamma}$, $\underset{\sim}{g}$, $\underset{\sim}{G}$, σ_e^2, σ_ϵ^2, and ρ. The complete-data vector,

$(\underset{\sim}{W}^* | \underset{\sim}{\gamma}, \underset{\sim}{g}) \sim N(\underset{\sim}{A}\underset{\sim}{\gamma} + \underset{\sim}{M}\underset{\sim}{g}, \underset{\sim}{\Omega})$

has a bivariate normal density function

$f_o(\underset{\sim}{W}^* | \underset{\sim}{\gamma}, \underset{\sim}{g}, \Omega)$

$$= \left[\frac{1}{2\pi}\right]^n |\underset{\sim}{\Omega}|^{-\frac{1}{2}} \times \exp\left[-\frac{1}{2}(\underset{\sim}{W}^* - \underset{\sim}{A}\underset{\sim}{\gamma} - \underset{\sim}{M}\underset{\sim}{g})'(\underset{\sim}{\Omega})^{-1}(\underset{\sim}{W}^* - \underset{\sim}{A}\underset{\sim}{\gamma} - \underset{\sim}{M}\underset{\sim}{g})\right].$$

This distribution is a member of the exponential class of distributions. The density has the regular exponential-family form

$f_o(\underset{\sim}{W}^* | \underset{\sim}{\theta}) = b(\underset{\sim}{W}^*) \exp[\underset{\sim}{\theta}' t(\underset{\sim}{W}^*)] / a(\underset{\sim}{\theta})$ and

$$\frac{\partial l_o(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} = t(\underset{\sim}{W}^*) - \frac{\partial}{\partial \underset{\sim}{\theta}} \log a(\underset{\sim}{\theta})$$

$$= t(\underset{\sim}{W}^*) - E[t(\underset{\sim}{W}^*) | \underset{\sim}{\theta}].$$

where $\underset{\sim}{\theta} = \{\underset{\sim}{\gamma}, \underset{\sim}{g}, \sigma_e^2, \sigma_\epsilon^2, \rho\}$ denotes the parameter vector that is restricted to a (p+1)-dimensional convex set $\underset{\sim}{\Psi}$ such that $f_o(\underset{\sim}{W}^* | \underset{\sim}{\theta})$ defines a density for all $\underset{\sim}{\theta}$ in $\underset{\sim}{\Psi}$ and

$$a(\underset{\sim}{\theta}) = \int_{W^*} b(\underset{\sim}{W}^*) \exp[\underset{\sim}{\theta}' t(\underset{\sim}{W}^*)] \, d\underset{\sim}{W}^*$$

where $\mathcal{W}^*$ is the sample space of $\underset{\sim}{W}^*$.

For a right censored sample $(\underset{\sim}{W}, \underset{\sim}{\delta})$ the function

$$Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}) = E[l_o(\underset{\sim}{\theta}_1 \mid \underset{\sim}{W}^*) + \log \pi(\underset{\sim}{g}_1 \mid \underset{\sim}{G}_1) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}]$$
$$= E[\underset{\sim}{\theta}_1' t(\underset{\sim}{W}^*) - \log a(\underset{\sim}{\theta}_1) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] + \log b(\underset{\sim}{W}^*) + \log \pi(\underset{\sim}{g}_1 \mid \underset{\sim}{G}_1)$$
$$= E[\underset{\sim}{\theta}_1' t(\underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] - \log a(\underset{\sim}{\theta}_1) + \log b(\underset{\sim}{W}^*) + \log \pi(\underset{\sim}{g}_1 \mid \underset{\sim}{G}_1).$$

For a concave function p_o, Q is maximized with respect to $\underset{\sim}{\theta}_1$ when

$$0 = \frac{\partial Q(\underset{\sim}{\theta}_1, \underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}_1} = E[t(\underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] - \frac{\partial \log a(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1} + \frac{\partial}{\partial \underset{\sim}{\theta}_1} \log \pi(\underset{\sim}{g}_1 \mid \underset{\sim}{G}_1)$$

$$= E[t(\underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] - E[t(\underset{\sim}{W}^*) \mid \underset{\sim}{\theta}_1] + \frac{\partial}{\partial \underset{\sim}{\theta}_1} \log \pi(\underset{\sim}{g}_1 \mid \underset{\sim}{G}_1)$$

or
$$E[t(\underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}] = E[t(\underset{\sim}{W}^*) \mid \underset{\sim}{\theta}_1] - \frac{\partial}{\partial \underset{\sim}{\theta}_1} \log \pi(\underset{\sim}{g}_1 \mid \underset{\sim}{G}_1).$$

The corresponding E and M steps are:

E-step: Compute $t^{(r)}(\underset{\sim}{W}^*) = E[t(\underset{\sim}{W}^*) \mid \underset{\sim}{W}, \underset{\sim}{\delta}, \underset{\sim}{\theta}^{(r-1)}]$

M-step: Obtain $\underset{\sim}{\theta}^{(r)}$ as the solution to

$$E[t(\underset{\sim}{W}^*) \mid \underset{\sim}{\theta}^{(r)}] - \frac{\partial}{\partial \underset{\sim}{\theta}^{(r)}} \log \pi(\underset{\sim}{g}^{(r)} \mid \underset{\sim}{G}^{(r)}) = t^{(r)}(\underset{\sim}{W}^*). \qquad (3.2.3)$$

With bivariate right censoring, there are three cases to be considered:

(1) $\delta_{1ij} = \delta_{2ij} = 1$, i.e., both Y_{ij}^* and C_{ij}^* are known,

(2) $\delta_{1ij} = 1$ and $\delta_{2ij} = 0$ which implies that $C_{ij}^* > Y_{ij}^* = Y_{ij}$

and (3) $\delta_{1ij} = 0$ and $\delta_{2ij} = 1$ which implies that $Y_{ij}^* > C_{ij}^* = C_{ij}$.

It is assumed that at least one of the Y_{ij} or C_{ij} is observed. If both are missing (i.e., $\delta_{1ij} = \delta_{2ij} = 0$), this observation is assumed to be ignorably missing.

In the expectation step,

$$E[\, t(W^*_{ij}) \mid W_{ij}=w_{ij}, \underset{\sim}{\delta}_{ij}, \underset{\sim}{\theta}\,]=$$

$$\delta_{1ij}\delta_{2ij}E[\, t(W^*_{ij}) \mid W_{ij}=w_{ij}, \delta_{1ij}=1, \delta_{2ij}=1, \underset{\sim}{\theta}\,]$$
$$+(1-\delta_{1ij})\delta_{2ij}E[\, t(W^*_{ij}) \mid W_{ij}=w_{ij}, \delta_{1ij}=0, \delta_{2ij}=1, \underset{\sim}{\theta}\,]$$
$$+\delta_{1ij}(1-\delta_{2ij})E[\, t(W^*_{ij}) \mid W_{ij}=w_{ij}, \delta_{1ij}=1, \delta_{2ij}=0, \underset{\sim}{\theta}\,]$$

$$=\delta_{1ij}\delta_{2ij}E[\, t(W^*_{ij}) \mid W^*_{ij}=w_{ij}, \underset{\sim}{\theta}\,]$$
$$+(1-\delta_{1ij})\delta_{2ij}E[\, t(W^*_{ij}) \mid Y^*_{ij}>y_{ij}, C^*_{ij}=y_{ij}, \underset{\sim}{\theta}\,]$$
$$+\delta_{1ij}(1-\delta_{2ij})E[\, t(W^*_{ij}) \mid Y^*_{ij}=y_{ij}, C^*_{ij}>y_{ij}, \underset{\sim}{\theta}\,]$$

$$=\delta_{1ij}\delta_{2ij}t(w_{ij})$$
$$+(1-\delta_{1ij})\delta_{2ij}E[\, t(W^*_{ij}) \mid Y^*_{ij}>y_{ij}, C^*_{ij}=y_{ij}, \underset{\sim}{\theta}\,]$$
$$+\delta_{1ij}(1-\delta_{2ij})E[\, t(W^*_{ij}) \mid Y^*_{ij}=y_{ij}, C^*_{ij}>y_{ij}, \underset{\sim}{\theta}\,]$$

with corresponding density functions

$$f_0(Y^* \mid Y^*>y, C^*=y)=f_0(Y^* \mid Y=y, \delta_1=0, \delta_2=1)$$

$$=\frac{f_0(Y^*=y^*, Y=y, \delta_1=0, \delta_2=1)}{f_0(\, Y=y, \delta_1=0, \delta_2=1)}=\frac{f_0(Y^*=y^*, C^*=y)}{\int_y^\infty f_o(Y^*, C^*)\, dY^*}$$

$$=\frac{g_{C^*}(y)\, f_o(Y^* \mid C^*=y)}{g_{C^*}(y)\int_y^\infty f_o(Y^* \mid C^*=y)\, dY^*}$$

$$=\frac{f_o(Y^* \mid C^*=y)}{\int_y^\infty f_o(Y^* \mid C^*=y)\, dY^*}=\frac{f_o(Y^* \mid C^*=y)}{1-\Phi(z_{Y^* \mid C^*})}$$

where $z_{Y^*\mid C^*,\ ij}=\dfrac{y_{ij}-\left[\underset{\sim}{X}_{ij}\underset{\sim}{\beta}+\underset{\sim}{B}_{ij}\,\underset{\sim}{d}_i+\rho\frac{\sigma_{Y^*}}{\sigma_{C^*}}[y_{ij}-(\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}+\underset{\sim}{J}_{ij}\,\underset{\sim}{v}_i)]\right]}{\sigma_{Y^*}(1-\rho^2)^{\frac{1}{2}}}$.

and

$$f_0(C^* \mid Y^*=y,\ C^*>y)=f_0(C^* \mid Y=y,\ \delta_1=1,\ \delta_2=0)$$

$$=\frac{f_0(C^*=c^*,\ Y=y,\ \delta_1=1,\ \delta_2=0)}{f_0(\ Y=y,\ \delta_1=1,\ \delta_2=0)}=\frac{f_0(Y^*=y,\ C^*=c^*)}{\int_y^\infty f_o(Y^*,\ C^*)\ dC^*}$$

$$=\frac{f_{Y^*}(y)\ g_o(C^*\mid Y^*=y)}{f_{Y^*}(y)\int_y^\infty g_o(C^*\mid Y^*=y)\ dC^*}$$

$$=\frac{g_o(C^*\mid Y^*=y)}{\int_y^\infty g_o(C^*\mid Y^*=y)\ dC^*}=\frac{g_o(C^*\mid Y^*=y)}{1-\Phi(z_{C^*\mid Y^*})}$$

where $z_{C^*\mid Y^*,\ ij}=\dfrac{y_{ij}-\left[\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}+\underset{\sim}{J}_{ij}\,\underset{\sim}{v}_i+\rho\frac{\sigma_{C^*}}{\sigma_{Y^*}}[y_{ij}-(\underset{\sim}{X}_{ij}\underset{\sim}{\beta}+\underset{\sim}{B}_{ij}\,\underset{\sim}{d}_i)]\right]}{\sigma_{C^*}(1-\rho^2)^{\frac{1}{2}}}$.

Therefore

$$E[\ t(W^*_{ij}) \mid Y^*_{ij}>y_{ij},\ C^*_{ij}=y_{ij},\ \underset{\sim}{\theta}\,]$$

$$=\frac{1}{1-\Phi(z_{Y^*\mid C^*,\ ij})}\int_{y_{ij}}^\infty t(W^*_{ij})\ f_o(Y^*_{ij} \mid C^*_{ij}=y_{ij},\ \underset{\sim}{\theta}\,)\ dY^*_{ij}$$

and

$$E[\ t(W^*_{ij})\ |\ Y^*_{ij}=y_{ij},\ C^*_{ij}>y_{ij},\ \underset{\sim}{\theta}\ ]$$

$$=\frac{1}{1-\Phi\left(z_{C^*\,|\,Y^*,\ ij}\right)}\int_{y_{ij}}^{\infty} t(W^*_{ij})\ g_o(C^*_{ij}\ |\ Y^*_{ij}=y_{ij},\ \underset{\sim}{\theta}\ )\ dC^*_{ij}.$$

3.2.3 Mixed Models with Informative Right Censoring; EM Computations

EM Computations are summarized in this section. In the E step, the conditional expected values of the 'complete-data' sufficient statistic are computed from the observed data and current estimates of the parameters, while in the M step, new estimates of the unknown parameters are computed using the conditional expected values of the 'complete-data' sufficient statistics in the maximum a posteriori estimating equations.

E Step:

In the r-th iteration, the estimation step computes the conditional expected values of the complete-data sufficient statistics given the observed data $\underset{\sim}{W}$ and the estimated values of the parameters from the $(r-1)$-st iteration. Not all of the $\underset{\sim}{W}^*$ are observed so the E step will estimate the complete-data sufficient statistics that involve $\underset{\sim}{W}^*$.

A set of complete-data sufficient statistics for this problem is

$$\sum_{i=1}^{K} \underset{\sim}{A}_i' (\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{M}_i \underset{\sim}{g}_i),$$

$$\{ \underset{\sim}{M}_i' (\underset{\sim}{\Omega}^{-1} \otimes \underset{\sim}{I}_{n_i}) (\underset{\sim}{W}_i^* - \underset{\sim}{A}_i \underset{\sim}{\gamma})\}_{i=1}^{K},$$

$$(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{B}\underset{\sim}{d})'(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{B}\underset{\sim}{d}),$$

$$(\underset{\sim}{C}^* - \underset{\sim}{H}\underset{\sim}{\alpha} - \underset{\sim}{J}\underset{\sim}{v})'(\underset{\sim}{C}^* - \underset{\sim}{H}\underset{\sim}{\alpha} - \underset{\sim}{J}\underset{\sim}{v}), \text{ and}$$

$$(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{B}\underset{\sim}{d})'(\underset{\sim}{C}^* - \underset{\sim}{H}\underset{\sim}{\alpha} - \underset{\sim}{J}\underset{\sim}{v}) .$$

The expected values of the complete-data sufficient statistics may be denoted,

$$\underset{\sim}{t}_1^{(r)} = \sum_{i=1}^{K} \underset{\sim}{A}_i' \, (\underset{\sim}{\Omega}^{(r-1)^{-1}} \otimes \underset{\sim}{I}_{n_i}) \times \left\{ \tilde{\underset{\sim}{W}}_i^* - \underset{\sim}{M}_i \underset{\sim}{g}_i^{(r-1)} \right\}$$

where

$$\tilde{\underset{\sim}{W}}_i^* = E[\, \underset{\sim}{W}_i^* \mid \underset{\sim}{W}_i, \ \underset{\sim}{\delta}_i, \ \underset{\sim}{\gamma}^{(r-1)}, \ \underset{\sim}{g}_i^{(r-1)}, \ \underset{\sim}{G}^{(r-1)}, \ \sigma_e^{(r-1)^2}, \ \sigma_\epsilon^{(r-1)^2}, \ \rho^{(r-1)}] \, ,$$

$$\underset{\sim}{t}_{2i}^{(r)} = \underset{\sim}{M}_i' \, (\underset{\sim}{\Omega}^{(r-1)^{-1}} \otimes \underset{\sim}{I}_{n_i}) \, (\tilde{\underset{\sim}{W}}_1^* - \underset{\sim}{A}_i \underset{\sim}{\gamma}^{(r-1)})$$

$$t_3^{(r)} = E[\, (\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{B}\underset{\sim}{d})'(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{B}\underset{\sim}{d})$$

$$\mid \underset{\sim}{Y}, \ \underset{\sim}{\delta}, \ \underset{\sim}{\beta}^{(r-1)}, \ \underset{\sim}{d}^{(r-1)}, \ \underset{\sim}{D}^{(r-1)}, \ \sigma_e^{(r-1)^2}]$$

$$= E[\, (\underset{\sim}{Y}^{*\prime}\underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \ \underset{\sim}{\delta}, \ \underset{\sim}{\beta}^{(r-1)}, \ \underset{\sim}{d}^{(r-1)}, \ \underset{\sim}{D}^{(r-1)}, \ \sigma_e^{(r-1)^2}]$$

$$-2 \, (\underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)} + \underset{\sim}{B}\underset{\sim}{d}^{(r-1)})' \, \tilde{\underset{\sim}{Y}}^*$$

$$+ \, (\underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)} + \underset{\sim}{B}\underset{\sim}{d}^{(r-1)})' \, (\underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)} + \underset{\sim}{B}\underset{\sim}{d}^{(r-1)})$$

where $\tilde{\underset{\sim}{Y}}^* = E[\, \underset{\sim}{Y}^* \mid \underset{\sim}{Y}, \ \underset{\sim}{\delta}, \ \underset{\sim}{\beta}^{(r-1)}, \ \underset{\sim}{d}^{(r-1)}, \ \underset{\sim}{D}^{(r-1)}, \ \sigma_e^{(r-1)^2}]$,

$$t_4^{(r)}=E[\,(\underset{\sim}{C}^* - \underset{\sim}{H}\underset{\sim}{\alpha} - \underset{\sim}{J}\underset{\sim}{v})'(\underset{\sim}{C}^* - \underset{\sim}{H}\underset{\sim}{\alpha} - \underset{\sim}{J}\underset{\sim}{v})$$

$$\mid \underset{\sim}{Y},\ \underset{\sim}{\delta},\ \underset{\sim}{\alpha}^{(r-1)},\ \underset{\sim}{v}^{(r-1)},\ \underset{\sim}{V}^{(r-1)},\ \sigma_\epsilon^{(r-1)^2}]$$

$$=E[\,(\underset{\sim}{C}^{*\prime}\underset{\sim}{C}^* \mid \underset{\sim}{Y},\ \underset{\sim}{\delta},\ \underset{\sim}{\alpha}^{(r-1)},\ \underset{\sim}{v}^{(r-1)},\ \underset{\sim}{V}^{(r-1)},\ \sigma_\epsilon^{(r-1)^2}]$$

$$-2\,(\underset{\sim}{H}\underset{\sim}{\alpha}^{(r-1)}+\underset{\sim}{J}\underset{\sim}{v}^{(r-1)})'\,\tilde{\underset{\sim}{C}}^*$$

$$+\,(\underset{\sim}{H}\underset{\sim}{\alpha}^{(r-1)}+\underset{\sim}{J}\underset{\sim}{v}^{(r-1)})'\,(\underset{\sim}{H}\underset{\sim}{\alpha}^{(r-1)}+\underset{\sim}{J}\underset{\sim}{v}^{(r-1)})$$

where $\tilde{\underset{\sim}{C}}^*=E[\,(\underset{\sim}{C}^* \mid \underset{\sim}{Y},\ \underset{\sim}{\delta},\ \underset{\sim}{\alpha}^{(r-1)},\ \underset{\sim}{v}^{(r-1)},\ \underset{\sim}{V}^{(r-1)},\ \sigma_\epsilon^{(r-1)^2}]$

and

$$t_5^{(r)}=E[\,(\underset{\sim}{Y}^* - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{B}\underset{\sim}{d})'(\underset{\sim}{C}^* - \underset{\sim}{H}\underset{\sim}{\alpha} - \underset{\sim}{J}\underset{\sim}{v})$$

$$\mid \underset{\sim}{W}_i,\ \underset{\sim}{\delta}_i,\ \underset{\sim}{\gamma}^{(r-1)},\ \underset{\sim}{g}_i^{(r-1)},\ \underset{\sim}{G}^{(r-1)},\ \sigma_e^{(r-1)^2},\ \sigma_\epsilon^{(r-1)^2},\ \rho^{(r-1)}]$$

$$=E[\,(\underset{\sim}{Y}^{*\prime}\underset{\sim}{C}^* \mid \underset{\sim}{W}_i,\ \underset{\sim}{\delta}_i,\ \underset{\sim}{\gamma}^{(r-1)},\ \underset{\sim}{g}_i^{(r-1)},\ \underset{\sim}{G}^{(r-1)},\ \sigma_e^{(r-1)^2},\ \sigma_\epsilon^{(r-1)^2},\ \rho^{(r-1)}]$$

$$-\,\tilde{\underset{\sim}{Y}}^{*\prime}\,(\underset{\sim}{H}\underset{\sim}{\alpha}^{(r-1)}+\underset{\sim}{J}\underset{\sim}{v}^{(r-1)}) - (\underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}\underset{\sim}{d}^{(r-1)})'\,\tilde{\underset{\sim}{C}}^*$$

$$+\,(\underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}\underset{\sim}{d}^{(r-1)})'\,(\underset{\sim}{H}\underset{\sim}{\alpha}^{(r-1)}+\underset{\sim}{J}\underset{\sim}{v}^{(r-1)}).$$

When $\delta_{1ij}=0$ and $\delta_{2ij}=1$ expectations involving the i-th element of the complete-data sufficient statistics can be computed as:

$$E[Y^*_{ij} \mid Y^*_{ij}>y_{ij},\ C^*_{ij}=y_{ij},\ \underset{\sim}{\theta}^{(r-1)}]$$

$$=\int_{y_{ij}}^{\infty} Y^*_{ij}\ \frac{f_o(Y^*_{ij}\mid C^*_{ij}=y_{ij},\ \underset{\sim}{\theta}^{(r-1)})}{1-\Phi(z^{(r-1)}_{Y^*\mid C^*,\ ij})}\ dY^*_{ij}$$

$$=\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}_{ij}\underset{\sim}{d}_i^{(r-1)}$$

$$+\rho^{(r-1)}\frac{\sigma_e^{(r-1)}}{\sigma_\epsilon^{(r-1)}}\left(y_{ij}-\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}^{(r-1)}-\underset{\sim}{J}_{ij}\underset{\sim}{v}_i^{(r-1)}\right)+\sigma_e^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}\frac{\phi(z^{(r-1)}_{Y^*\mid C^*,\ ij})}{1-\Phi(z^{(r-1)}_{Y^*\mid C^*,\ ij})}$$

where

$$z^{(r-1)}_{Y^*\mid C^*,\ ij}=\frac{y_{ij}-\left[\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}_{ij}\underset{\sim}{d}_i^{(r-1)}+\rho^{(r-1)}\frac{\sigma_e^{(r-1)}}{\sigma_\epsilon^{(r-1)}}[y_{ij}-(\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}^{(r-1)}+\underset{\sim}{J}_{ij}\underset{\sim}{v}_i^{(r-1)})]\right]}{\sigma_e^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}},$$

and

$$E[Y^{*2}_{ij} \mid Y^*_{ij}>y_{ij},\ C^*_{ij}=y_{ij},\ \underset{\sim}{\theta}^{(r-1)}]$$

$$=\int_{y_{ij}}^{\infty} Y^{*2}_{ij}\ \frac{f_o(Y^*_{ij}\mid C^*_{ij}=y_{ij},\ \underset{\sim}{\theta}^{(r-1)})}{1-\Phi(z^{(r-1)}_{Y^*\mid C^*,\ ij})}\ dY^*_{ij}$$

$$=[\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}_{ij}\underset{\sim}{d}_i^{(r-1)}$$

$$+\rho^{(r-1)}\frac{\sigma_e^{(r-1)}}{\sigma_\epsilon^{(r-1)}}\left(y_{ij}-\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}^{(r-1)}-\underset{\sim}{J}_{ij}\underset{\sim}{v}_i^{(r-1)}\right)]^2+\sigma_e^{(r-1)^2}(1-\rho^{(r-1)^2})$$

$$+\sigma_e^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}$$

$$\times\ [y_{ij}+\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}_{ij}\underset{\sim}{d}_i^{(r-1)}$$

$$+\rho^{(r-1)}\frac{\sigma_e^{(r-1)}}{\sigma_\epsilon^{(r-1)}}(y_{ij}-\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}^{(r-1)}-\underset{\sim}{J}_{ij}\underset{\sim}{v}_i^{(r-1)})\ ]$$

$$\times\ \frac{\phi(z_{Y^*\,|\,C^*,\ ij}^{(r-1)})}{1-\Phi(z_{Y^*\,|\,C^*,\ ij}^{(r-1)})}\ .$$

Similarly, when δ_{1ij}=1 and δ_{2ij}=0 expectations involving the i-th element of the complete-data sufficient statistics can be computed as:

$$E[C_{ij}^*\ |\ Y_{ij}^*=y_{ij},\ C_{ij}^*>y_{ij},\ \underset{\sim}{\theta}^{(r-1)}]$$

$$=\int_{y_{ij}}^{\infty} C_{ij}^*\ \frac{g_o(C_{ij}^*|\ Y_{ij}^*=y_{ij},\ \underset{\sim}{\theta}^{(r-1)})}{1-\Phi(z_{C^*\,|\,Y^*,\ ij}^{(r-1)})}\ dC_{ij}^*$$

$$=\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}^{(r-1)}+\underset{\sim}{J}_{ij}\underset{\sim}{v}_i^{(r-1)}+\rho^{(r-1)}\frac{\sigma_\epsilon^{(r-1)}}{\sigma_e^{(r-1)}}(y_{ij}-\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}-\underset{\sim}{B}_{ij}\underset{\sim}{d}_i^{(r-1)})$$

$$+\sigma_\epsilon^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}\ \frac{\phi(z_{C^*\,|\,Y^*,\ ij}^{(r-1)})}{1-\Phi(z_{C^*\,|\,Y^*,\ ij}^{(r-1)})}$$

where

$$z_{C^*\,|\,Y^*,\ ij}^{(r-1)}=\frac{y_{ij}-\left[\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}^{(r-1)}+\underset{\sim}{J}_{ij}\underset{\sim}{v}_i^{(r-1)}+\rho^{(r-1)}\frac{\sigma_\epsilon^{(r-1)}}{\sigma_e^{(r-1)}}[y_{ij}-(\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}+\underset{\sim}{B}_{ij}\underset{\sim}{d}_i^{(r-1)})]\right]}{\sigma_\epsilon^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}},$$

and

$$E[C_{ij}^{*2}\ |\ Y_{ij}^*=y_{ij},\ C_{ij}^*>y_{ij},\ \underset{\sim}{\theta}^{(r-1)}]$$

$$=\int_{y_{ij}}^{\infty} C_{ij}^{*2} \frac{g_o(C_{ij}^*|\, Y_{ij}^*=y_{ij},\ \underset{\sim}{\theta}^{(r-1)})}{1-\Phi(z_{C^*|Y^*,\, ij}^{(r-1)})}\, dC_{ij}^*$$

$$=[\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}^{(r-1)}+\underset{\sim}{J}_{ij}\underset{\sim}{v}_i^{(r-1)}+\rho^{(r-1)}\frac{\sigma_\epsilon^{(r-1)}}{\sigma_e^{(r-1)}}(y_{ij}-\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}-\underset{\sim}{B}_{ij}\underset{\sim}{d}_i^{(r-1)})]^2$$

$$+\sigma_\epsilon^{(r-1)^2}(1-\rho^{(r-1)^2})$$

$$+\sigma_\epsilon^{(r-1)}\left(1-\rho^{(r-1)^2}\right)^{\frac{1}{2}}$$

$$\times\ [y_{ij}+\underset{\sim}{H}_{ij}\underset{\sim}{\alpha}^{(r-1)}+\underset{\sim}{J}_{ij}\underset{\sim}{v}_i^{(r-1)}$$

$$+\rho^{(r-1)}\frac{\sigma_\epsilon^{(r-1)}}{\sigma_e^{(r-1)}}(y_{ij}-\underset{\sim}{X}_{ij}\underset{\sim}{\beta}^{(r-1)}-\underset{\sim}{B}_{ij}\underset{\sim}{d}_i^{(r-1)})\ ]$$

$$\times\ \frac{\phi(z_{C^*|Y^*,\, ij}^{(r-1)})}{1-\Phi(z_{C^*|Y^*,\, ij}^{(r-1)})}\ .$$

<u>M</u> <u>Step</u>:

The r-th iteration of the M step obtains $\underset{\sim}{\gamma}^{(r)}$, $\underset{\sim}{g}^{(r)}$, $\underset{\sim}{G}^{(r)}$, $\sigma_e^{(r)^2}$, $\sigma_\epsilon^{(r)^2}$, and $\rho^{(r)}$ as the solution to equation 3.2.3. Initial values of maximum a posteriori estimates are obtained using unweighted regression and by treating censored data as if they were uncensored. Convenient initial estimates are:

$$\underset{\sim}{\gamma}^{(0)}=\left[\sum_{i=1}^{K}\underset{\sim}{A}_i'\underset{\sim}{A}_i\right]^{-1}\ \sum_{i=1}^{K}\underset{\sim}{A}_i'\underset{\sim}{W}_i,$$

$$\underset{\sim}{g}_i^{(0)}=[\underset{\sim}{M}_i'\underset{\sim}{M}_i]^{-1}\underset{\sim}{M}_i'(\underset{\sim}{W}_i-\underset{\sim}{A}_i\underset{\sim}{\gamma}^{(0)}),$$

$$\underset{\sim}{G}^{(0)}=\frac{1}{K}\sum_{i=1}^{K}\underset{\sim}{g}_i^{(0)}\underset{\sim}{g}_i^{(0)\prime},$$

$$\sigma_e^{(0)^2}=\frac{1}{n}\,(\underset{\sim}{Y}^*-\underset{\sim}{X}\underset{\sim}{\beta}^{(0)}-\underset{\sim}{B}\underset{\sim}{d}^{(0)})'\,(\underset{\sim}{Y}^*-\underset{\sim}{X}\underset{\sim}{\beta}^{(0)}-\underset{\sim}{B}\underset{\sim}{d}^{(0)}),$$

$$\sigma_\epsilon^{(0)^2}=\frac{1}{n}\,(\underset{\sim}{C}^*-\underset{\sim}{H}\underset{\sim}{\alpha}^{(0)}-\underset{\sim}{J}\underset{\sim}{v}^{(0)})'\,(\underset{\sim}{C}^*-\underset{\sim}{H}\underset{\sim}{\alpha}^{(0)}-\underset{\sim}{J}\underset{\sim}{v}^{(0)}),$$

$$\rho^{(0)}=\frac{1}{n\,\sigma_e^{(0)}\sigma_\epsilon^{(0)}}\,(\underset{\sim}{Y}^*-\underset{\sim}{X}\underset{\sim}{\beta}^{(0)}-\underset{\sim}{B}\underset{\sim}{d}^{(0)})'\,(\underset{\sim}{C}^*-\underset{\sim}{H}\underset{\sim}{\alpha}^{(0)}-\underset{\sim}{J}\underset{\sim}{v}^{(0)}),$$

and if $\underset{\sim}{G}=\sum_{s=1}^{m}\sigma_s\,\underset{\sim}{S}_s$ then

$$\underset{\sim}{\sigma}^{(0)}=\underset{\sim}{1}\,.$$

In the r-th iteration maximum a posteriori estimates which maximize the posterior distribution function using the expected values of the complete-data sufficient statistics obtained from the previous iteration of the E step are computed:

$$\underset{\sim}{\gamma}^{(r)}=\left[\sum_{i=1}^{K}\underset{\sim}{A}_i'\,(\underset{\sim}{\Omega}^{(r-1)^{-1}}\otimes\underset{\sim}{I}_{n_i})\,\underset{\sim}{A}_i\right]^{-1}\underset{\sim}{t}_1^{(r)},$$

$$\underset{\sim}{g}_i^{(r)}=[\,\underset{\sim}{M}_i'\,(\underset{\sim}{\Omega}^{(r-1)^{-1}}\otimes\underset{\sim}{I}_{n_i})\,\underset{\sim}{M}_i+\underset{\sim}{G}^{(r-1)^{-1}}]^{-1}\;\underset{\sim}{t}_{2i}^{(r)},$$

$$\underset{\sim}{G}^{(r)}=\frac{1}{K}\sum_{i=1}^{K}\underset{\sim}{g}_i^{(r)}\underset{\sim}{g}_i^{(r)\prime},$$

$$\sigma_e^{(r)^2}=\frac{1}{n}\,t_3^{(r)},$$

$$\sigma_\epsilon^{(r)^2}=\frac{1}{n}\,t_4^{(r)},$$

$$\rho^{(r)} = \frac{1}{n\sigma_e^{(r-1)}\sigma_\epsilon^{(r-1)}}\, t_5^{(r)},$$

and if $\underset{\sim}{G} = \sum_{s=1}^{m} \tau_s \underset{\approx}{S}_s$ then

$$\underset{\sim}{\tau}^{(r)} = \frac{1}{K}\left[< \text{trace}\,(\underset{\sim}{G}^{(r-1)^{-1}} \underset{\approx}{S}_s \underset{\sim}{G}^{(r-1)^{-1}} \underset{\approx}{S}_h) >_{sh} \right]^{-1}$$

$$\times \left[< \sum_{i=1}^{K} \underset{\approx}{g}_i^{(r-1)'} \underset{\sim}{G}^{(r-1)^{-1}} \underset{\approx}{S}_s\, \underset{\sim}{G}^{(r-1)^{-1}} \underset{\approx}{g}_i^{(r-1)} >_s \right].$$

IV. EXERCISE TOLERANCE TESTS OF PATIENTS WITH CHRONIC STABLE ANGINA

4.1 Introduction

Calcium channel blocking drugs have been found to be useful in the treatment of patients with chronic stable angina. Nisoldipine, a dihydropyridine slow channel calcium blocker is a potent coronary vasodilator and could be beneficial for patients with angina. Lam et al. (1985) conducted a placebo-controlled, double-blind crossover, dose-ranging study to assess the short-term efficacy of nisoldipine as an antianginal drug in humans. Maximal treadmill exercise tests at 1, 3 and 8 hours after ingestion of a single dose of a placebo or a 5, 10, or 20 mg. oral dose of nisoldipine were used to assess the duration of its effects and the effective oral dose.

The efficacy of nisoldipine was assessed in terms of time to onset of angina (denoted by Y^*), the primary response of interest, and other response variables. Because the greatest effects occurred 3 hours after oral ingestion, we shall examine the 3-hour data.

The dataset described in Lam et al. (1985), reproduced in Appendix A, was used to obtain parameter estimates assuming that no censoring occurred. Those parameter estimates were used as parameter values in a program that generated artificial data for 80 subjects. (The details are described in Section 4.3.) Fixed and informative right censoring were induced artificially and resulting parameter estimates are compared to those obtained without censoring in Section 4.4. Fixed censoring occurs when time to onset of angina exceeds a predetermined value. Informative right-censoring occurs when time to onset of angina exceeds maximal exercise time (denoted by C^*), the censoring variable. In this situation, patients become exhausted and have to stop running on the treadmill before they get angina. However because peak exercise duration time and time to onset of angina are highly correlated, it is plausible that individuals are censored when they were at unusually high risk of failure.

4.2 Description of the Experiment and Data

The study population consisted of 12 male or female patients with a mean age of 58 years (range 46 to 66) who had a history of angina pectoris which had been stable for at least three months. Prior to acceptance into the study, exercise-induced angina occurred during upright treadmill exercise tests in all of these patients, with specified associated ECG changes. Three patients had a previous myocardial infarction and one patient had prior coronary bypass surgery. Patients were excluded from the study if they had a history of predominant rest angina or ST segment elevation during an episode of chest pain, severe hypertension, valvular heart disease, congestive heart failure, intraventricular conduction disturbances or severe ventricular arrhythmias on the electrocardiogram at rest.

Patients were randomized to one of four treatment sequences and received a different dose each visit. Prior to entering the study, each patient had had at least two exercise tests and were familiar with the test environment. All cardiovascular medications were stopped at least 48 hours before the study with the exception of nitroglycerin, which was stopped at least eight hours before the exercise test. Patients also refrained from smoking and drinking coffee or tea at least eight hours before the exercise test. A single oral dose consisting of four tablets was given at 8:00 AM. The placebo dose consisted of four placebo tablets, the 5 mg. dose consisted of one 5 mg. tablet of nisoldipine and 3 placebo tablets, the 10 mg. dose consisted of two 5 mg. tablets of nisoldipine and two placebo tablets, and the 20 mg. dose consisted of four 5 mg. tablets of nisoldipine. Modified Naughton treadmill exercise tests were performed at 9:00 AM, 11:00 AM and 4:00 PM. Exercise tests were terminated due to severe angina or extreme fatigue. The series of tests involved four study days per patient and were completed within 2 weeks for each patient. There was a 48 to 72 hour washout period after each visit. Previous data had indicated that a 48-hour washout period was sufficient to eliminate carryover effects.

4.3 Computational Issues

Recall that the mixed model for $\underset{\sim}{W}_i^*$ is given by

$$\underset{\sim}{W}_i^* = \underset{\sim}{A}_i^* \underset{\sim}{\gamma} + \underset{\sim}{M}_i^* \underset{\sim}{g}_i + \underset{\sim}{\nu}_i$$

$$= \begin{bmatrix} \underset{\sim}{Y}_i^* \\ \underset{\sim}{C}_i^* \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}_i & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{H}_i \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\alpha} \end{bmatrix} + \begin{bmatrix} \underset{\sim}{B}_i & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{J}_i \end{bmatrix} \begin{bmatrix} \underset{\sim}{d}_i \\ \underset{\sim}{v}_i \end{bmatrix} + \begin{bmatrix} \underset{\sim}{e}_i \\ \underset{\sim}{\epsilon}_i \end{bmatrix}$$

where

$\underset{\sim}{W}_i^*$ is a $2n_i \times 1$ vector of failure and censoring values which may or may not be observed,

$\underset{\sim}{A}_i = \begin{bmatrix} \underset{\sim}{X}_i & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{H}_i \end{bmatrix}$ is a $2n_i \times 2p$ known constant matrix of rank $2r \leq 2p$,

$\underset{\sim}{\gamma} = \begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\alpha} \end{bmatrix}$ is a $2p \times 1$ vector of unknown constant ‘fixed’ population parameters,

$\underset{\sim}{M}_i = \begin{bmatrix} \underset{\sim}{B}_i & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{J}_i \end{bmatrix}$ is a $2n_i \times 2q$ known matrix corresponding to the random effects,

$\underset{\sim}{g}_i = \begin{bmatrix} \underset{\sim}{d}_i \\ \underset{\sim}{v}_i \end{bmatrix}$ is a $2q \times 1$ vector of unobservable individual parameters,

$\underset{\sim}{\nu}_i = \begin{bmatrix} \underset{\sim}{e}_i \\ \underset{\sim}{\epsilon}_i \end{bmatrix}$ is a $2n_i \times 1$ vector of unobservable random errors,

$\underset{\sim}{g}_i \sim N(\underset{\sim}{0}, \underset{\sim}{G})$ independent of $\underset{\sim}{\nu}_i \sim N(\underset{\sim}{0}, \underset{\sim}{\Omega} \otimes \underset{\sim}{I}_{n_i})$,

$\underset{\sim}{G}$ is a positive-definite symmetric 2q × 2q covariance matrix

of random effects, $\underset{\sim}{g}_i$, and

$$\underset{\sim}{\Omega} = \begin{bmatrix} \sigma_e^2 & \rho\sigma_e\sigma_\epsilon \\ \rho\sigma_e\sigma_\epsilon & \sigma_\epsilon^2 \end{bmatrix}$$ consists of the unknown within-subject variance components.

The formulae given in Sections 3.1.3 and 3.2.3 were used to obtain parameter estimates, with the exception of initial starting values. These were estimated assuming noninformative censoring and using the EM algorithm to compute maximum likelihood estimates of $\underset{\sim}{\beta}$, $\underset{\sim}{D}$, and σ_e^2.. The same initial estimates of $\underset{\sim}{\beta}$, $\underset{\sim}{D}$, and σ_e^2 were used in the informative censoring program. It was assumed that $\underset{\sim}{\alpha}^{(0)}=\underset{\sim}{\beta}^{(0)}$, $\underset{\sim}{V}^{(0)}=\underset{\sim}{D}^{(0)}$, $\sigma_\epsilon^{(0)^2}=\sigma_e^{(0)^2}$, cov($\underset{\sim}{V}^{(0)}$, $\underset{\sim}{D}^{(0)}$)=0.5, and ρ=0.5. These estimates were then used to obtain initial estimates of the random effects.

Convergence was assumed to have occurred if the maximum relative change (over all estimators) in one iteration was less than 0.001 for all of the structural parameters. Using the values of the parameters at convergence, the observed log posterior distribution functions assuming fixed or noninformative censoring and informative censoring were:

$$p(\underset{\sim}{\beta}, \underset{\sim}{d}, \underset{\sim}{D}, \sigma_e^2 | \underset{\sim}{Y}, \underset{\sim}{\delta})$$

$$\propto \sum_{i=1}^{K} \left\{ \sum_{j=1}^{n_i} \delta_{ij} \log f_o(Y_{ij} \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i, \sigma_e^2) \right.$$

$$\left. + \sum_{j=1}^{n_i} (1-\delta_{ij}) \log\left[1 - \Phi\left(\frac{y_{ij} - \underset{\sim}{X}_{ij}\underset{\sim}{\beta} - \underset{\sim}{B}_{ij}\underset{\sim}{d}_i}{\sigma_e}\right)\right] + \log \pi(\underset{\sim}{d}_i | \underset{\sim}{D}) \right\}$$

and

$$p(\underset{\sim}{\gamma}, \underset{\sim}{g}, \underset{\sim}{G}, \sigma_e^2, \sigma_\epsilon^2, \rho \mid \underset{\sim}{W}, \underset{\sim}{\delta})$$

$$\propto \sum_{i=1}^{K} \Bigg\{ \sum_{j=1}^{n_i} \delta_{1ij}\delta_{2ij} \log f_o(W_{ij} \mid \underset{\sim}{\gamma}, \underset{\sim}{g}_i, \sigma_e^2, \sigma_\epsilon^2, \rho)$$

$$+\sum_{j=1}^{n_i} (1-\delta_{1ij})\delta_{2ij} \log\left[1-\Phi(z_{Y^*\mid C^*,\ ij})\ g(C_{ij} \mid \underset{\sim}{\alpha}, \underset{\sim}{v}_i, \sigma_\epsilon^2)\right]$$

$$+\sum_{j=1}^{n_i} \delta_{1ij}(1-\delta_{2ij}) \log\left[1-\Phi(z_{C^*\mid Y^*,\ ij})\ f(Y_{ij} \mid \underset{\sim}{\beta}, \underset{\sim}{d}_i, \sigma_e^2)\right]$$

$$+ \log \pi(\underset{\sim}{g}_i \mid \underset{\sim}{G}) \Bigg\}, \text{ respectively,}$$

where $$z_{Y^*\mid C^*,\ ij} = \frac{y_{ij} - \left[\underset{\sim}{X}_{ij}\underset{\sim}{\beta} + \underset{\sim}{B}_{ij}\ \underset{\sim}{d}_i + \rho\frac{\sigma_{Y^*}}{\sigma_{C^*}}[y_{ij} - (\underset{\sim}{H}_{ij}\underset{\sim}{\alpha} + \underset{\sim}{J}_{ij}\ \underset{\sim}{v}_i)]\right]}{\sigma_{Y^*}(1-\rho^2)^{\frac{1}{2}}}$$

and $$z_{C^*\mid Y^*,\ ij} = \frac{y_{ij} - \left[\underset{\sim}{H}_{ij}\underset{\sim}{\alpha} + \underset{\sim}{J}_{ij}\ \underset{\sim}{v}_i + \rho\frac{\sigma_{C^*}}{\sigma_{Y^*}}[y_{ij} - (\underset{\sim}{X}_{ij}\underset{\sim}{\beta} + \underset{\sim}{B}_{ij}\ \underset{\sim}{d}_i)\right]}{\sigma_{C^*}(1-\rho^2)^{\frac{1}{2}}}.$$

Estimates of the within-subject variance components, σ_e^2 and σ_ϵ^2, are biased, in part due to the large number of fixed and random effects that had to be estimated. After convergence, "reduced-biased" estimates were computed by multiplying both components by $\frac{n}{df_e}$ and $\frac{n}{df_\epsilon}$, where the error degrees of freedom for σ_e^2, $df_e = n - \text{rank}(\underset{\sim}{X}, \underset{\sim}{B})$, while the error degrees of freedom for σ_ϵ^2, $df_\epsilon = n - \text{rank}(\underset{\sim}{H}, \underset{\sim}{J})$. These "reduced-biased" estimates were used to compute $\underset{\sim}{\Omega}_i$, $\underset{\sim}{\Sigma}_{\underset{\sim}{W}_i^*}$, and estimates of variance of fixed effects, $V(\underset{\sim}{\gamma}) = \sum_{i=1}^{K} \underset{\sim}{A}_{1i}' \underset{\sim}{\Sigma}_{\underset{\sim}{W}_i^*}^{-1} \underset{\sim}{A}_{1i}$, where $\underset{\sim}{A}_{1i}$ consists of the rows in $\underset{\sim}{A}_i$ that correspond to uncensored data in $\underset{\sim}{W}$. Consequently, standard errors of fixed effects are usually lower when censoring occurs. When comparing approximate standard errors of fixed effects with and without censored data, it may be preferable to use all the rows of $\underset{\sim}{A}_i$ instead of using $\underset{\sim}{A}_{1i}$.

4.4 Generation of Data

The values of the dependent variables $\underset{\sim}{Y}_i^*$ and $\underset{\sim}{C}_i^*$ were generated using a mixed model with linear covariance structure as described in Section 3.2 using parameter estimates obtained from the dataset from Lam et al. (1985).

In this example, there were two random intercept parameters for each subject, one for $\underset{\sim}{Y}_i^*$ and one for $\underset{\sim}{C}_i^*$ (i.e., q=1). In addition, the design matrices for both $\underset{\sim}{Y}_i^*$ and $\underset{\sim}{C}_i^*$ were equal, i.e., $\underset{\sim}{X}_i=\underset{\sim}{H}_i$ and $\underset{\sim}{B}_i=\underset{\sim}{J}_i$. The estimated parameter vector of fixed effects for $\underset{\sim}{Y}^*$ consisted of the following:

$$\hat{\underset{\sim}{\beta}}'=[\hat{\beta}_{\text{int}},\ \hat{\beta}_{5\text{ mg.}},\ \hat{\beta}_{10\text{ mg.}},\ \hat{\beta}_{20\text{ mg.}},\ \hat{\beta}_2,\ \hat{\beta}_3,\ \hat{\beta}_4]$$
$$=[376.6,\ 74.6,\ 93.3,\ 78.8,\ -10.0,\ -43.3,\ 1.67]$$

where β_{int} denotes the intercept for Y*, $\beta_{5\text{ mg.}}$, $\beta_{10\text{ mg.}}$, and $\beta_{20\text{ mg.}}$ denote incremental effects of 5, 10, and 20 mg. doses of nisoldipine, respectively, on Y*, and β_2, β_3, and β_4 denote incremental effects of the second, third, and fourth periods on Y*.

Similarly, the estimated fixed effects parameter vector for $\underset{\sim}{C}^*$ was estimated to be

$$\hat{\underset{\sim}{\alpha}}'=[\hat{\alpha}_{\text{int}},\ \hat{\alpha}_{5\text{ mg.}},\ \hat{\alpha}_{10\text{ mg.}},\ \hat{\alpha}_{20\text{ mg.}},\ \hat{\alpha}_2,\ \hat{\alpha}_3,\ \hat{\alpha}_4]$$
$$=[429.2,\ 64.2,\ 72.9,\ 92.9,\ -5.8,\ -30.4,\ -7.1].$$

The vector of between-subject variance components $\underset{\sim}{\tau}$ consisted of the following:

$$\underset{\sim}{\tau}=\begin{cases}\tau_1 & =\text{Var(d)} \quad =27750,\\ \tau_2 & =\text{Var(v)} \quad =26729,\\ \tau_3 & =\text{Cov(d,v)}=26538.\end{cases}$$

The "reduced-biased" estimates of the within-subject variance components for Y* and C* and correlation coefficient were

σ_e^2=3395, σ_e^2=2653, and ρ=0.771.

Using the SAS function RANNOR with fixed seeds chosen from a table of random digits, $\underset{\sim}{g}_i$ and $\underset{\sim}{\nu}_i$ were generated for each subject. Because $\underset{\sim}{g}_i \sim \text{NID}(\underset{\sim}{0}, \underset{\sim}{G})$, data can be generated by setting

$$\underset{\sim}{g}_i = \underset{\sim}{G}^{\frac{1}{2}} \underset{\sim}{Z}_{g_i} \text{ and } \underset{\sim}{W}_{ij} = \underset{\sim}{A}_{ij} \underset{\sim}{\gamma} + \underset{\sim}{M}_{ij} \underset{\sim}{g}_i + \underset{\sim}{\Omega}^{\frac{1}{2}} \underset{\sim}{Z}_{ij},$$

where $\underset{\sim}{Z}_{g_i}$ and $\underset{\sim}{Z}_{ij}$ are 2×1 NID(0, 1) random variables and $\underset{\sim}{G}^{\frac{1}{2}}$ and $\underset{\sim}{\Omega}^{\frac{1}{2}}$ are the Cholesky square roots of $\underset{\sim}{G}$ and $\underset{\sim}{\Omega}$. Data were generated for 80 subjects with up to four visits per subject (Appendix B). One of four treatment sequences was randomly assigned to each subject with equal probability using the SAS function RANUNI. To ensure that some data were missing at random, observations were deleted with a 15% probability. This would reflect a situation in which, on average, subjects failed to show up for their scheduled appointments 15% of the time.

This data were analyzed with and without censoring. Right censoring was induced by censoring either the time to onset of angina or the maximal exercise time, whichever was greater if they exceeded the seventy-fifth percentiles. In this dataset, the seventy-fifth percentiles were 540 seconds for time to onset of angina and 556 seconds for MET. Ten percent (28) of the 273 observed angina times and 19% (53) of the 273 realized values of maximal exercise time were censored. Note that computations that assume censoring is informative will make adjustments for 30% of the observations that have either censored angina or maximal exercise times. If censoring is assumed to be fixed, maximal exercise time is irrelevant and adjustments only need to be made for 10% of the censored angina times.

4.5 Results

As a first step, the dataset described in Lam et al. (1985) was used to obtain parameter estimates for Y^*, assuming no censoring occurred. Parameter estimates were obtained (1) using maximum likelihood estimation whereby random effects were estimated in the E-step of the EM Algorithm, and (2) using maximum a posteriori estimation and estimating the random effects in the M-step. Parameter estimates are given in Table 4.5.1. Because random effects are computed differently in the two methods, different estimates for fixed intercept and random intercepts were computed. Fixed effects for the 5, 10, and 20 mg. doses of nisoldipine and for the second, third, and fourth periods were identical, but had slightly different estimates of approximate standard errors. The estimates of $\underset{\sim}{D}$ were similar while the "reduced-biased" estimates of within-subject variances were 3272 (error df=41) and 3370 (error df=30).

Subsequent computations involved the randomly generated dataset consisting of 80 fictitious subjects with up to 4 visits per subject. Structural parameter estimates (i.e., parameter estimates excluding the incidental parameters, $\underset{\sim}{g}$) obtained from the model are given in Table 4.5.2 along with corresponding estimates obtained with fixed and informatively right-censored data. Parameter estimates obtained using the complete data were compared with those obtained using the censored data assuming censoring was informative and fixed.

Parameter estimates are given in Table 4.5.2. Unlike the previous example, this example was an incomplete design in which approximately 15% of the visits were missing at random (noninformatively). Parameter estimates for fixed effects for the uncensored and informatively censored data were very similar, but not identical. When censoring was assumed to be fixed there were more discrepancies between parameter estimates obtained for uncensored data and censored data. This is not surprising because the correlation between Y^* and C^* was ignored.

The estimated variance components that were computed assuming fixed censoring were somewhat smaller than those that were computed using uncensored or informatively censored data. This resulted in somewhat larger estimates of approximate asymptotic standard errors for fixed effects. Approximate asymptotic standard errors of fixed effects parameters obtained using informatively censored data were smaller than those obtained using the uncensored data.

Results in Tables 4.5.1 and 4.5.2 indicate that time to onset of angina and peak exercise time were both prolonged by each dose of nisoldipine. In these data, the 10 mg. dose was optimal for prolonging time to onset of angina while the 20 mg. dose was optimal for prolonging maximal exercise time. Period effects were minimal in the original 12-patient dataset but appeared to be present in the larger 80-patient dataset, most likely due in part to the larger sample size and corresponding smaller standard errors.

Maximum a posteriori informative censoring estimates and maximum likelihood estimates and approximate asymptotic standard errors from Tables 4.5.2 and 4.5.3 are plotted in Figure 4.5.1, along with the population estimates (i.e, the original parameter values, denoted by $*$, that were used to generate the larger sample). The predicted response, Time to Angina (seconds), is plotted against estimates of fixed effects parameters for incremental effects of the 5, 10, and 20 mg. doses of nisoldipine and incremental effects of the second, third, and fourth test days. Approximate 95% confidence intervals for incremental effects of the 5, 10, and 20 mg. doses do not include zero. The largest incremental effect corresponded to the 10 mg. dose where the predicted increase in time to onset of angina was 95.3

seconds using all the data, 83.5 seconds after deleting censored observations, and 91.0 seconds using the casewise deletion method. The overlap of the approximate confidence intervals suggest that the treatment effects are not significantly different from one another but do seem to be significantly different from the placebo. Incremental effects due to the second and fourth test days do not appear to be statistically significant at the α=0.05 level. However, patients appeared to perform poorly on the third test day with predicted decreases in time to angina of about 60 seconds.

Maximum likelihood estimates in Table 4.5.3 were computed using mixed model techniques for complete data after (1) deleting 10% of the 273 observations with censored values of time to onset of angina and (2) "Casewise Deletion", where patients were excluded from the analysis if any of their angina times were either censored or missing. Casewise deletion resulted in the greatest loss of information - 46 of the 80 patients (58%) were excluded from the analysis. As a result, one would expect these confidence intervals to be the most inaccurate and to have the largest standard errors.

The discrepancies between the confidence intervals appear to be particularly evident for the 20 mg. dose. Using the casewise deletion method, the predicted increase in time to onset of angina was only 31.2 seconds compared to 63.0 seconds using the informatively censored angina times and 53.3 seconds using the mixed model approach assuming that censored angina times were missing at random. Note also that the confidence intervals for the casewise deletion method are wider and would lead one to believe that the incremental effect of the 20 mg. was the least effective dose.

TABLE 4.5.1
PARAMETER ESTIMATES USING DATA FROM LAM ET AL. (K=12 PATIENTS) ASSUMING THAT NO CENSORING OCCURRED

Dependent Variable=Time to Onset of Angina

Parameters	Maximum Likelihood Estimation (80 iterations)	Maximum a Posteriori Estimation (91 iterations)
Fixed Effects:	Estimate (a.s.e.)	Estimate (a.s.e.)
β_{int}	386.0 (52.8)	381.6 (52.8)
$\beta_{5\ mg.}$	74.6 (23.4)	74.6 (23.7)
$\beta_{10\ mg.}$	93.3 (23.4)	93.3 (23.7)
$\beta_{20\ mg.}$	78.8 (23.4)	78.9 (23.7)
β_2	−10.0 (23.4)	−10.0 (23.7)
β_3	−43.3 (23.4)	−43.3 (23.7)
β_4	1.7 (23.4)	1.7 (23.7)
Random Effects:	Estimate	Estimate
d_1	91.3	95.4
d_2	13.6	17.8
d_3	−118.8	−114.4
d_4	239.4	243.4
d_5	101.0	105.1
d_6	−305.7	−301.2
d_7	0.2	4.5
d_8	−53.2	−48.9
d_9	−26.5	−22.2
d_{10}	−143.0	−138.7

Parameters	Maximum Likelihood Estimation (80 iterations)	Maximum a Posteriori Estimation (91 iterations)
d_{11}	43.9	48.1
d_{12}	335.4	339.2
Variance Components:		
D	27754	27545
σ_e^2	3272	3370

a.s.e. = approximate asymptotic standard error

TABLE 4.5.2
MAXIMUM A POSTERIORI PARAMETER ESTIMATES USING SIMULATED DATASET (K=80 PATIENTS)

Dependent Variables=Time to Onset of Angina,
Maximal Exercise Time.

Parameters	Complete Data (83 iterations)	Informative Censoring (80 iterations)	Fixed Censoring* (52 iterations)
Fixed Effects:	Estimate (a.s.e.)	Estimate (a.s.e.)	Estimate (a.s.e.)
β_{int}	382.2 (17.3)	383.5 (10.0)	386.8 (15.0)
$\beta_{\text{5 mg.}}$	60.5 (11.5)	61.8 (10.5)	57.2 (12.0)
$\beta_{\text{10 mg.}}$	94.3 (11.1)	95.3 (10.4)	96.5 (12.2)
$\beta_{\text{20 mg.}}$	62.5 (10.7)	63.0 (9.3)	58.1 (11.3)
β_2	−13.8 (11.2)	−15.2 (10.0)	−16.7 (12.1)
β_3	−57.9 (11.3)	−59.0 (10.2)	−59.4 (12.1)
β_4	−4.6 (10.9)	−4.6 (9.5)	−9.0 (11.6)
α_{int}	438.8 (15.6)	438.9 (8.9)	
$\alpha_{\text{5 mg.}}$	53.4 (9.8)	48.0 (9.2)	
$\alpha_{\text{10 mg.}}$	59.2 (9.4)	55.9 (8.7)	
$\alpha_{\text{20 mg.}}$	74.7 (9.1)	70.5 (8.6)	
α_2	−13.6 (9.5)	−14.7 (8.7)	
α_3	−38.6 (9.6)	−38.4 (8.9)	
α_4	−23.4 (9.2)	−24.7 (8.3)	

Parameters	Complete Data (83 iterations)	Informative Censoring (80 iterations)	Fixed Censoring* (52 iterations)
Variance Components:	Estimates	Estimates	Estimates
D	16759	17072	16348
V	14360	13264	
cov(D, V)	14798	14478	
σ_e^2	3850	3909	3835
σ_ϵ^2	2781	2754	
ρ	0.790	0.791	

a.s.e.=approximate asymptotic standard error

* Note: Censoring was informative but parameter estimates in the third column were obtained assuming fixed censoring.

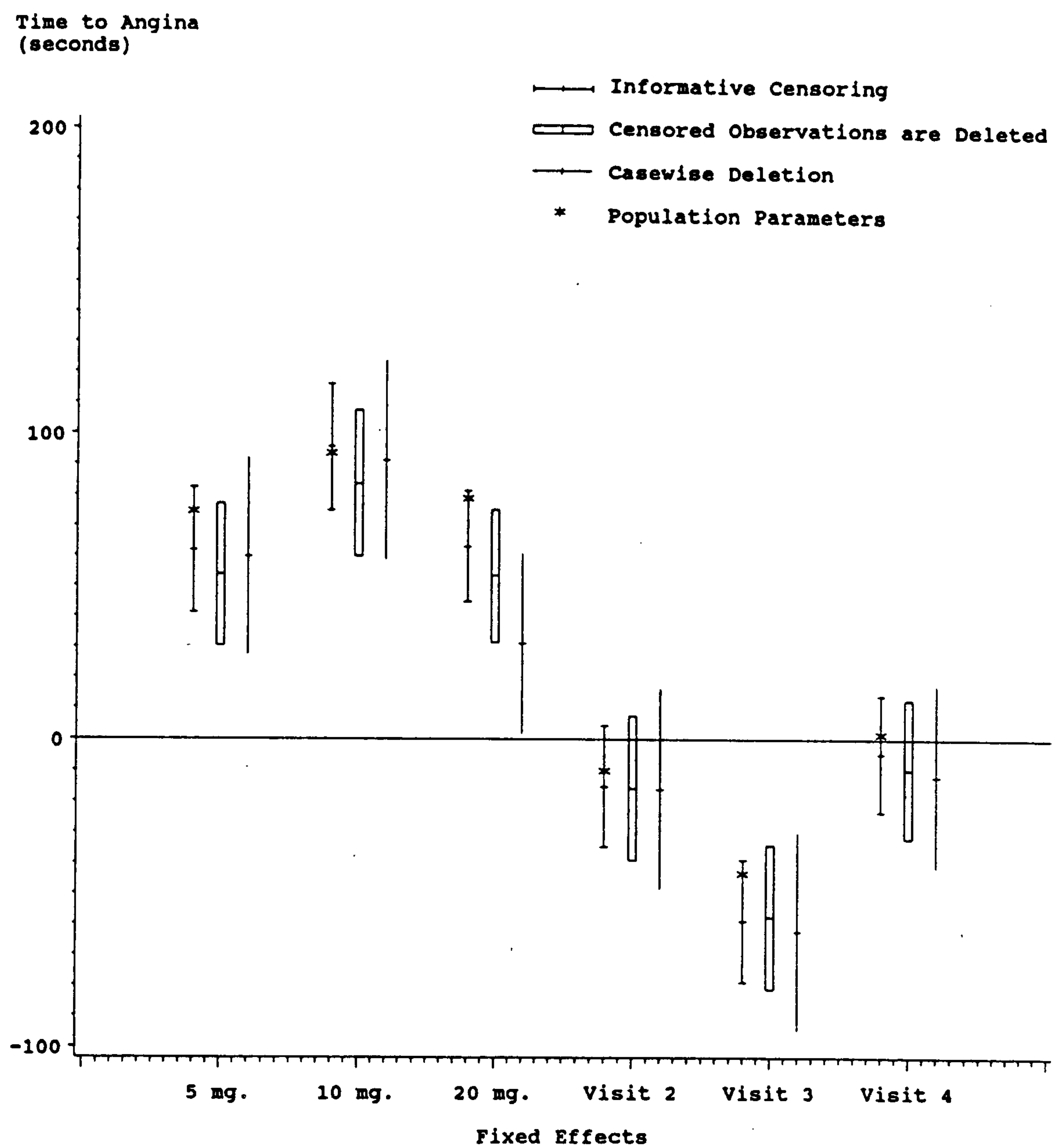
TABLE 4.5.3
MAXIMUM LIKELIHOOD ESTIMATES USING SIMULATED DATASET

Dependent Variable=Time to Onset of Angina

Parameters	Censored Observations Deleted (K=78) (38 iterations)	Casewise Deletion (K=34) (91 iterations)
Fixed Effects:	Estimate (a.s.e.)	Estimate (a.s.e.)
β_{int}	383.2 (17.2)	368.0 (26.8)
$\beta_{5\ mg.}$	53.8 (11.9)	59.6 (16.5)
$\beta_{10\ mg.}$	83.5 (12.1)	91.0 (16.5)
$\beta_{20\ mg.}$	53.3 (11.1)	31.2 (15.0)
β_2	−15.8 (11.9)	−16.1 (16.5)
β_3	−57.7 (12.0)	−62.4 (16.5)
β_4	−9.7 (11.4)	−12.0 (15.0)
Variance Components:		
D	15549	17551
σ_e^2	3650	3540

a.s.e. = approximate asymptotic standard error

Figure 4.5.1
Approximate 95% Confidence Intervals for Incremental Effects of Dose and Visit



4.6 Summary

Results from a demonstration of this method using the treadmill exercise data from the cross-over study by Lam et al. (1985) and the simulated dataset generated using this data were summarized in this chapter. In each example, the algorithm was monotonic: the logarithm of the posterior distribution function increased monotonically until the parameter estimates converged within the specified tolerance. Estimates of fixed effects and variance components obtained after censoring approximately 10 to 30% of the data were reasonably close to the original complete-data estimates and were well within two standard errors of one another.

Parameter estimates for incremental effects of dose were somewhat higher when right-censoring was taken into account than when censored observations were deleted. This would always be the case if censored data had been analyzed as if they were uncensored. The least accurate method was casewise deletion, where 58% of the patients were excluded from the analysis.

V. SUMMARY AND RECOMMENDATIONS FOR FUTURE RESEARCH

5.1 Summary

The purpose of this work was to simplify existing computational approaches used to estimate parameters in general linear mixed models with fixed or noninformative random censoring and to extend the use of these techniques to include parameter estimation in mixed models with informative censoring. This method is applicable to normal data from longitudinal studies where the effects of serial correlation are negligible.

For this type of problem, the EM algorithm was preferable to gradient methods (e.g., Newton-Raphson algorithm, Method of Scoring) because the likelihood of the complete data for the General Linear Mixed Model has a much simpler form than the likelihood corresponding to the General Linear Mixed Model with censored data. The EM algorithm also does not require the inversion of large matrices. Using the Newton-Raphson algorithm, this problem can become overwhelming when the number of subjects is large and gets worse when there are multiple random coefficients for each subject.

Using an extension of maximum likelihood estimation known as maximum *a posteriori* estimation, computations involved in obtaining parameter estimates in the M step of the EM algorithm are straightforward. Unlike the maximum likelihood approach, random effects are estimated in the M step instead of the E step of the EM algorithm and do not require the computation of high-dimensional integrals.

In an example, parameter estimates obtained using informatively censored data were generally quite similar to estimates obtained using complete data (i.e., data that were generated before being artificially censored). The extent of agreement was usually greater if censoring was correctly assumed to be informative than when it was assumed to be fixed. Parameter estimates obtained from the complete data set using maximum likelihood estimation and maximum *a posteriori* estimation were also quite similar with the exception of the within-subject variance components.

5.2 Future Research

Procedures have been developed for maximum *a posteriori* estimation for a General Linear Mixed Model for the analysis of censored normal or lognormal data, whether or not the censoring is informative. These procedures have led to a great reduction in computational complexities in comparison to previously available methods for data with noninformative or fixed censoring mechanisms. Parameter estimates obtained using informatively censored data were found to be similar to estimates obtained using complete data. Simulation studies are needed to assess large- and small-sample properties of the parameter estimates. Exact asymptotic distributions of the parameter estimates under general regularity conditions also need to be derived. This could include an evaluation of some approximate F statistics and their small sample distributions.

"Reduced-biased" estimates of the within-subject variance components were proposed in this dissertation and used to compute approximate standard errors of fixed effects. It would be useful to derive approximate asymptotic estimates of the standard errors of random effects and variance components and to derive restricted maximum likelihood or analogous "restricted maximum a posteriori" estimates of variance components.

The methods developed in this dissertation are applicable to longitudinal studies where the effects of serial correlation are negligible. These methods are applicable when $V(\underset{\sim}{e}_i)=\underset{\sim}{I}\,\sigma_e^2$. As an extension, one could attempt to model other types of covariance structures that occur in studies of longer duration (e.g., irregularly-timed, inconsistently-timed longitudinal data with $V(\underset{\sim}{e}_i)$ having AR(1) covariance structures).

Zeger, Liang, and Albert (1988) describe how generalized estimating equations can be used to analyze uncensored longitudinal data. They consider two approaches: the first approach is the "subject-specific" approach where the covariance structure is explicitly modeled, while in the second approach, the "population-averaged" approach, the marginal expectation is the focus and the covariance matrix is regarded as a nuisance. It may be possible to extend these approaches so they can be used for parameter estimation in general linear mixed models with censored data.

Appendix A

Acute Effects of Nisoldipine on Chronic Stable Angina Treadmill Exercise Data (Lam et al. 1985)

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
1	1	10	540	600	
	2	20	590	660	
	3	0	480	510	
	4	5	505	530	
2	1	0	395	460	
	2	10	440	540	
	3	5	480	570	
	4	20	480	540	
3	1	20	405	525	
	2	5	345	390	
	3	10	240	285	
	4	0	260	310	
4	1	5	675	675	
	2	0		690	Fatigue
	3	20		660	Fatigue
	4	10		700	Fatigue
5	1	10		540	Fatigue
	2	20		585	Fatigue
	3	0	480	480	
	4	5	550	570	

Appendix A
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data (Lam et al. 1985)

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
6	1	0	90	120	
	2	10	110	120	
	3	5	90	120	
	4	20	190	210	
7	1	5	450	510	
	2	0	360	420	
	3	20	450	510	
	4	10	480	510	
8	1	20	420	450	
	2	5	370	420	
	3	10	490	530	
	4	0	240	330	
9	1	0	335	335	
	2	10	455	480	
	3	5	420	450	
	4	20	420	490	
10	1	20	265	360	
	2	5	290	350	
	3	10	285	335	
	4	0	310	370	

Appendix A
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data (Lam et al. 1985)

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
11	1	5	540	660	
	2	0	420	630	
	3	20	360	660	
	4	10	600	650	
12	1	10	900	900	
	2	20	780	780	
	3	0	600	660	
	4	5	840	840	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
1	1	0	538	556	
	2	10	572	555	Fatigue
	4	20	706	669	Fatigue
2	1	20	276	402	
	2	5	383	496	
	3	10	361	488	
	4	0	283	355	
3	1	20	345	445	
	2	5	260	354	
	3	10	402	410	
	4	0	234	268	
4	1	10	432	515	
	2	20	303	480	
	3	0	365	473	
	4	5	433	462	
5	1	20	264	351	
	3	10	224	273	
	4	0	167	179	
6	1	20	421	477	
	2	5	393	537	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
	4	0	248	367	
7	1	10	398	420	
	3	0	288	434	
	4	5	333	422	
8	1	10	458	527	
	2	20	547	562	Angina
	3	0	354	466	
9	1	20	333	498	
	2	5	294	439	
	3	10	324	431	
	4	0	227	342	
10	1	0	437	478	
	2	10	360	468	
	3	5	346	430	
11	1	20	350	437	
	2	5	348	372	
	3	10	310	330	
	4	0	225	319	
12	1	0	364	398	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
	3	5	379	483	
	4	20	328	475	
13	1	0	405	471	
	2	10	467	512	
	3	5	463	546	
	4	20	435	445	
14	1	20	543	566	Angina
	2	5	482	495	
	3	10	392	441	
15	1	20	551	609	Angina
	3	10	545	517	Fatigue
	4	0	473	467	
16	1	5	478	542	
	2	0	426	459	
	4	10	461	506	
17	1	0	475	529	
	2	10	459	452	
	3	5	372	434	
	4	20	421	466	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
18	1	20	555	549	Fatigue
	2	5	424	448	
	4	0	451	413	
19	1	0	375	380	
	2	10	426	357	
	3	5	440	321	
	4	20	374	274	
20	1	0	594	669	Angina
	2	10	627	632	Angina
	4	20	702	743	Angina
21	1	20	428	491	
	3	10	475	482	
	4	0	434	452	
22	1	0	286	270	
	2	10	441	332	
	3	5	340	296	
	4	20	442	436	
23	1	0	449	467	
	3	5	582	579	Fatigue
	4	20	507	445	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
24	1	20	402	500	
	2	5	470	564	Angina
	3	10	571	543	Fatigue
	4	0	367	448	
25	2	0	229	334	
	3	20	242	354	
	4	10	332	394	
26	1	10	549	509	Fatigue
	2	20	602	580	Fatigue
	3	0	376	410	
	4	5	478	438	
27	1	20	229	336	
	2	5	95	293	
	3	10	30	231	
	4	0	31	150	
28	1	5	152	284	
	2	0	26	112	
	4	10	179	179	
29	1	20	491	534	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
	2	5	540	512	
	3	10	542	559	Angina
	4	0	469	484	
30	1	0	423	539	
	2	10	594	635	Angina
	3	5	412	598	Angina
	4	20	380	537	
31	1	0	275	377	
	3	5	190	286	
	4	20	395	468	
32	1	5	276	317	
	2	0	116	209	
	3	20	252	361	
	4	10	270	291	
33	2	5	777	745	Fatigue
	3	10	636	604	Fatigue
	4	0	678	645	Fatigue
34	1	0	179	200	
	2	10	161	179	
	3	5	101	236	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
35	1	20	530	652	Angina
	2	5	600	697	Angina
	3	10	552	652	Angina
	4	0	495	592	Angina
36	1	0	380	491	
	4	20	473	623	Angina
37	1	5	742	701	Fatigue
	2	0	653	661	Angina
	3	20	806	811	Angina
	4	10	741	711	Fatigue
38	1	10	503	545	
	2	20	474	591	Angina
	3	0	299	424	
39	1	20	317	347	
	2	5	348	362	
	3	10	473	398	
	4	0	385	387	
40	1	20	343	468	
	2	5	462	454	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
	3	10	468	453	
	4	0	320	318	
41	2	5	535	613	Angina
	3	10	414	542	
	4	0	354	434	
42	1	0	553	580	Angina
	2	10	549	533	Fatigue
	3	5	324	464	
	4	20	518	609	Angina
43	1	10	509	544	
	2	20	405	490	
44	2	10	575	473	Fatigue
	3	5	391	386	
	4	20	614	624	Angina
45	1	20	457	595	Angina
	2	5	475	529	
	3	10	542	655	Angina
	4	0	559	616	Angina
46	1	10	593	592	Fatigue

Appendix B

Acute Effects of Nisoldipine on Chronic Stable Angina

Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
	3	0	493	446	
	4	5	617	600	Fatigue
47	1	10	462	495	
	3	0	401	476	
48	1	20	588	657	Angina
	2	5	664	678	Angina
	3	10	480	489	
	4	0	664	685	Angina
49	1	10	312	347	
	2	20	237	315	
	3	0	104	196	
	4	5	202	274	
50	1	20	91	187	
	2	5	45	44	
	3	10	150	173	
	4	0	73	40	
51	1	5	513	603	Angina
	2	0	466	529	
	3	20	575	642	Angina

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
52	1	10	544	515	Fatigue
	2	20	385	491	
	3	0	323	428	
53	1	20	548	551	
	2	5	409	421	
	4	0	436	408	
54	1	20	438	535	
	4	0	195	281	
55	1	0	324	404	
	2	10	542	539	Fatigue
	3	5	480	526	
56	1	20	773	814	Angina
	2	5	623	744	Angina
	3	10	633	689	Angina
	4	0	591	644	Angina
57	1	20	301	414	
	2	5	379	451	
	3	10	272	355	
	4	0	265	432	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
58	1	0	367	399	
	2	10	518	612	Angina
	4	20	600	596	Fatigue
59	1	0	412	491	
	2	10	483	542	
	3	5	455	508	
	4	20	428	462	
60	1	20	326	409	
	3	10	412	460	
	4	0	341	474	
61	1	20	268	397	
	2	5	291	394	
	3	10	349	458	
	4	0	210	354	
62	1	20	575	690	Angina
	2	5	678	782	Angina
	3	10	645	693	Angina
	4	0	553	608	Angina
63	1	20	503	549	
	2	5	516	499	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
	3	10	438	422	
	4	0	421	403	
64	1	20	431	521	
	2	5	560	607	Angina
	3	10	414	487	
	4	0	502	533	
65	1	0	645	722	Angina
	2	10	691	726	Angina
	3	5	594	670	Angina
	4	20	617	707	Angina
66	1	5	252	305	
	2	0	215	242	
67	1	10	492	537	
	2	20	488	487	
	3	0	358	389	
	4	5	494	462	
68	1	20	546	672	Angina
	3	10	477	572	Angina
	4	0	520	578	Angina

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
69	1	5	532	468	
	2	0	359	415	
	4	10	621	554	Fatigue
70	1	5	593	636	Angina
	2	0	469	553	
	3	20	507	618	Angina
71	1	0	550	541	Fatigue
	2	10	534	609	Angina
	3	5	495	550	
	4	20	619	680	Angina
72	1	0	541	540	Fatigue
	2	10	650	602	Fatigue
	3	5	548	547	Fatigue
	4	20	585	556	Fatigue
73	1	0	565	590	Angina
	2	10	540	547	
	3	5	548	556	
	4	20	506	513	
74	1	10	395	448	
	2	20	398	479	

Appendix B
Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
	3	0	203	329	
	4	5	458	436	
75	1	5	402	424	
	2	0	297	318	
	3	20	253	379	
	4	10	415	428	
76	1	0	309	382	
	2	10	354	404	
	3	5	178	334	
77	1	10	432	355	
	2	20	434	425	
	3	0	296	287	
	4	5	392	412	
78	1	0	374	359	
	2	10	333	352	
	3	5	41	151	
	4	20	346	302	
79	3	5	382	489	
80	1	10	812	788	Fatigue

Appendix B

Acute Effects of Nisoldipine on Chronic Stable Angina
Treadmill Exercise Data generated from Population Parameter Estimates

Subject	Visit	Dose (mg)	Onset of Anginal Pain (seconds)	Cessation of Exercise (seconds)	Reason for Discontinuation
	2	20	733	721	Fatigue
	3	0	570	583	Angina

REFERENCES

Aitkin, M. (1981), "A Note on the Regression Analysis of Censored Data," Technometrics, 23(2), 161-163.

Batchelor, J. R., and Hackett, M. (1970), "HL-A Matching in Treatment of Burned Patients With Skin Allografts," The Lancet, 2, 581-583.

Beck, J. V., and Arnold, K. J. (1977), Parameter Estimation in Engineering and Science, New York: Wiley.

Bissette, J., Carr, G., Koch, G. G., Adams, K. F., and Sheps, D. S. (1986), "Analysis of (Events/Time at Risk) Ratios From Two-Period Crossover Studies," Proceedings of the Biopharmaceutical Section of the American Statistical Association.

Breslow, N. (1974), "Covariance Analysis of Censored Survival Data," Biometrics, 30, 89-99.

Carriquiry, A. L. (1985), "The Analysis of a Censored Multivariate Normal Distribution: Onset of Puberty in Mice," unpublished M.S. thesis, University of Illinois, Department of Animal Sciences.

Carriquiry, A. L., Gianola, D., and Fernando, R. L. (1987), "Mixed-Model Analysis of a Censored Normal Distribution With Reference to Animal Breeding," Biometrics, 43, 929-939.

Chaitman, B. R., Wagniart, P., Pasternac, A., Brevers, G., Scholl, J., Lam, J., Methe, M., Ferguson, R., and Bourassa, M. G. (1984), "Improved Exercise Tolerance After Propranolol, Diltiazem, or Nifedipine in Angina Pectoris: Comparison at 1, 3, and 8 Hours and Correlation with Plasma Drug Concentration," The American Journal of Cardiology, 53(1), 1-9.

Chatterjee, S., and McLeish, D. L. (1986), "Fitting Linear Regression Models to Censored Data by Least Squares and Maximum Likelihood Methods," Communications in Statistics - Theory and Methods, 15(11), 3227-3243.

Cox, D. R., and Oakes, D. (1984), Analysis of Survival Data, New York: Chapman and Hall.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society, B, 39, 1-38.

Elandt-Johnson, R. C., and Johnson, N. L. (1980), Survival Models and Data Analysis, New York: Wiley.

Fairclough, D. L., and Helms, R. W. (1984), "Mixed Effects Model Analyses of Incomplete Longitudinal Pulmonary Function Measurements in Children," Institute of Statistics Mimeo Series No. 1470, University of North Carolina at Chapel Hill.

Harville, D. A. (1974), "Bayesian Inference for Variance Components Using Only Error Contrasts," Biometrika, 61, 383-385.

---- (1976), "Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects," The Annals of Statistics, 4(2), 384-395.

---- (1977), "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," Journal of the American Statistical Association, 72 (358), 320-338.

Helms, R. W. (1988), "Manipulating Statistical Hypotheses and Tests as Boolean Functions," The American Statistician, 42, 253-256.

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

Kotz, S., and Johnson, N. L. (1983), Encyclopedia of Statistical Sciences (Vol. 4), New York: Wiley.

Laird, N. M., and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," Biometrics, 38, 963-974.

Lam, J., Chaitman, B. R., Crean, P., Blum, R., and Waters, D. D. (1985), "A Dose-Ranging, Placebo-Controlled, Double-Blind Trial of Nisoldipine in Effort Angina: Duration and Extent of Antianginal Effects,"Journal of the American College of Cardiology, 6(2), 447-452.

Lawless, J. F. (1982), Statistical Models and Methods for Lifetime Data, New York: Wiley.

Little, R. J. A., and Rubin, D. B. (1987), Statistical Analysis with Missing Data, New York: Wiley.

McCarroll, K., and Helms, R. W. (1987), "An Evaluation of Some Approximate F Statistics and Their Small Sample Distributions for the Mixed Model With Linear Covariance Structure," Institute of Statistics Mimeo Series No. 1838T, University of North Carolina at Chapel Hill.

Murray, S. C., and Helms, R. W. (1990), "Linear Models with Generalized AR(1) Covariance Structure for Irregularly-Timed Data," Institute of Statistics Mimeo Series No. 1876, University of North Carolina at Chapel Hill.

Naughton, J., and Haider, R. (1973), Methods of Exercise Testing. In: Naughton, J. P., Hellerstein, H. K., and Mohler, I. C., eds. Exercise Testing and Exercise Training in Coronary Heart Disease, New York: Academic Press, 79-91.

Neyman, J., and Scott, E. L. (1948), "Consistent Estimates Based on Partially Consistent Observations," Econometrica 16(1), 1-32.

Pettitt, A. N. (1986), "Censored Observations, Repeated Measures and Mixed Effects Models: An Approach Using the EM Algorithm and Normal Errors," Biometrika 73(3), 635-643.

Sampford, M. R., and Taylor, J. (1959), "Censored Observations in Randomized Block Experiments," Journal of the Royal Statistical Society, B, 21, 214-237.

Schluchter, M. D. (1991), "Methods for the Analysis of Informatively Censored Longitudinal Data," presented at the Symposium on Longitudinal Data Analysis, Sponsored by the Biopharmaceutical Section of the American Statistical Association, Fort Lauderdale, Florida.

Schmee, J., and Hahn, G. J. (1979), "A Simple Method for Regression Analysis With Censored Data," Technometrics, 21(4), 417-432.

Searle, S. R. (1982), Matrix Algebra Useful for Statistics, New York: Wiley.

Swan, A. V. (1969a), "Maximum Likelihood Estimation From Grouped and Censored Normal Data," Algorithm AS 16, Applied Statistics, 18, 110-114.

---- (1969b), "The Reciprocal of Mills's ratio," Algorithm AS 17, Applied Statistics, 18, 115-116.

---- (1977), "A Remark on AS 16," Applied Statistics, 26, 122.

Timm, N. H. (1970), "The Estimation of Variance-Covariance and Correlation Matrices from Incomplete Data," Psychometrika, 35, 417-437.

Ware, J. H. (1985), "Linear Models for the Analysis of Longitudinal Studies," The American Statistician, 39(2), 95-101.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989), "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions," Journal of the American Statistical Association, 84(408), 1065-1073.

Wolynetz, M. S. (1974), "Analysis of Type I Censored - Normally Distributed Data," unpublished Ph.D. dissertation, University of Waterloo, Department of Statistics and Actuarial Science.

---- (1979a), "Maximum Likelihood Estimation From Confined and Censored Normal Data," Algorithm AS 138, Applied Statistics, 28, 185-195.

---- (1979b), "Maximum Likelihood Estimation in a Linear Model From Confined and Censored Normal Data," Algorithm AS 139, Applied Statistics, 28, 195-206.

Wolynetz, M. S., and Binns, M. R. (1983), "Stayablilty of Dairy Cattle: Models with Censoring and Covariates," Journal of Dairy Science, 66, 935-942.

Wu, M. C., and Bailey, K. R. (1989), "Estimation and Comparison of Changes in the Presence of Informative Right Censoring: Conditional Linear Model," Biometrics, 45, 939-955.

---- (1989), "Analysing Changes in the Presence of Informative Right Censoring Caused by Death and Withdrawal," Statistics in Medicine, 7, 337-346.

Wu, M. C., and Carroll, R. J. (1988), "Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process," Biometrics, 44, 175-188.

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988), "Models for Longitudinal Data: A Generalized Estimating Equation Approach," Biometrics, 44, 1049-1060.

Zellner, A. (1971), An Introduction to Bayesian Inference in Econometrics, New York: Wiley.