Friedl, Kauermann:

Standard Errors for EM Estimates in Variance Component Models

Sonderforschungsbereich 386, Paper 145 (1999)

Online unter: http://epub.ub.uni-muenchen.de/

Standard Errors for EM Estimates in Variance Component Models

Herwig Friedl* Göran Kauermann †

April 8, 1999

Abstract

A procedure is derived for computing standard errors in random intercept models for estimates obtained from the EM algorithm. We discuss two different approaches: a Gauß-Hermite quadrature for Gaussian random effect models and a nonparametric maximum likelihood estimation for an unspecified random effect distribution. An approximation of the expected Fisher information matrix is proposed which is based on an expansion of the EM estimating equation. This allows for inferential arguments based on EM estimates, as demonstrated by an example and simulations.

Keywords: EM algorithm, Gauß-Hermite Quadrature, Nonparametric Maximum Likelihood, Estimating Equation.

*Institute of Statistics; Technical University Graz; Lessingstr. 27; 8010 Graz, Austria

†Institute of Statistics; Ludwig-Maximilians-University Munich; Akademiestr. 1; 80796 Munich, Germany

1 Introduction

We consider a generalized variance component model for n clusters of independent response variables $y_i = (y_{i1}, \ldots, y_{in_i})^t$, $i = 1, \ldots, n$. The response y_i is assumed to depend on the covariate matrix $x_i = (x_{i1}, \ldots, x_{in_i})^t$ and the unobservable random effect z_i. The mean response is modeled by the conditional generalized linear model $E(y_{ij}|z_i, x_{ij}) = \mu_{ij} = h(x_{ij}^t\beta + z_i)$ where $h(\cdot)$ is the inverse link function and β denotes the associated p dimensional parameter vector of interest. Conditional on the random effect, the components of y_i are assumed to be independent, i.e. $f(y_i|z_i;\beta) = \prod_j f(y_{ij}|z_i,\beta)$ where the density $f(\cdot|\cdot)$ is assumed to be of exponential family form. For $n_i > 1$ the model can be used for equi-correlated dependent observations, see e.g. Diggle et al. (1994, chapter 9) or Breslow & Clayton (1993) and references given there. If $n_i = 1$ the model is known as random effect model which provides a general and convenient way for modeling overdispersion, see e.g. Aitkin (1996).

The random effects z_i, $i = 1, \ldots, n$, are assumed to be independently and identically distributed with density $f(z_i)$. Since the z_i are unobserved this leads to the observed (marginal) log likelihood $l(\beta) = \sum_{i=1}^n \log \int f(y_i|z_i)f(z_i)dz_i$ which can be maximized by the iterative expectation maximization (EM) algorithm (see Dempster et al., 1977). In the t-th step this gives the function

$$Q(\beta|\beta^{(t)}) = \sum_{i=1}^{n} \frac{\int \log\{f(y_i|z_i;\beta)f(z_i)\}\, f(y_i|z_i;\beta^{(t)})f(z_i)dz_i}{\int f(y_i|z_i;\beta^{(t)})f(z_i)dz_i}, \qquad (1)$$

which has to be maximized in β with $\beta^{(t)}$ held fixed. It is a traditional and also convenient approach to assume normally distributed random effects, which allows to approximate $l(\beta)$ by a Gauß-Hermite (GH) quadrature. More flexibility is however achieved by treating the effect distribution as unknown, as suggested by Aitkin

(1996, 1999). This leads to nonparametric maximum likelihood (NPML) estimation as introduced by Laird (1978). In both settings the EM algorithm can be applied.

A criticism of the EM algorithm is, that it does not automatically provide estimates for the variance-covariance matrix of the EM estimate $\hat{\beta}$. As pointed out in McLachlan & Krishnan (1997) for a variety of examples, this point is closely related to the problem of slow convergence. If the maximization of (1) is done by a Newton-Raphson procedure based on $\partial^2 Q(\beta|\beta^{(t)})/\partial\beta\partial\beta^t$, directly applying successive E and M steps will provide an estimate of the complete information at convergence. However, this does not account for the missing information on the unobservable random effects. Louis (1982) provides a very general derivation of the observed information matrix and shows that this matrix can be rewritten as a difference of the complete and the missing information. Oakes (1999) discusses a formula for the observed information which depend only on derivatives of $Q(\cdot)$. We suggest a simplified version of this approach yielding the appropriate measure of information – the estimated a priori expected information – which takes the special structure of the considered models into account, see also Meilijson (1989).

In Section 2 we embed the EM algorithm in the framework of estimating equations. For both settings, GH approximation and NPML estimation, we expand the EM estimating equations. The first order derivative provides an approximative variance for the estimates. In Section 3 we apply this variance approximation in a data example and investigate its small sample behavior by a simulation study.

2 EM Estimating Equations

2.1 Gaussian Effects

Assuming $z_i \overset{iid}{\sim} N(0,1)$, we can model $\mu_{ij} = h(x_{ij}^t\beta + z_i\sigma_z)$ which allows to approximate the integrals in $l(\beta)$ by a quadrature formula like GH. Hinde (1982) uses this technique in random effect models, Anderson & Aitkin (1985) apply it to variance component models. The quadrature yields the approximation

$$f(y_i;\theta) \;=\; \int f(y_i|z_i;\theta)\varphi(z_i)dz_i \approx \sum_{k=1}^{K} f(y_i|\zeta_k;\theta)\pi_k \;=:\; f_K(y_i;\theta), \tag{2}$$

where φ denotes the standard normal density function, $\theta = (\beta^t, \sigma_z)^t$ and K is the number of approximation points. Note that for given K, the masspoints ζ_k and their associated masses π_k are known and available from tables. Applying this quadrature also to the nominator in (1), approximates $Q(\cdot)$ by

$$Q_K(\theta|\theta^{(t)}) = \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}^{(t)} \left\{\log f(y_i|\zeta_k;\theta) + \log \pi_k\right\} \tag{3}$$

with weights $w_{ik}^{(t)} = f(y_i|\zeta_k;\theta^{(t)})\pi_k / f_K(y_i;\theta^{(t)})$. These weights can thereby be seen as masses for the masspoints ζ_k corresponding to the posterior distribution $f(z_i|y_i)$. Formula (3) represents the E-step of the underlying EM algorithm. The M-step is given by the $p+1$ dimensional estimating equation

$$\frac{\partial Q_K(\theta|\theta^{(t)})}{\partial\theta} = \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}^{(t)} s_{ik} = 0. \tag{4}$$

with $s_{ik} = s_{ik}(\theta) = \partial \log f(y_i|\zeta_k;\theta)/\partial\theta$ denoting the i-th score contribution, given $z_i = \zeta_k$.

We embed the EM algorithm into the concept of estimating equations by defining

$$g_\theta(\theta) = \left.\frac{\partial Q_K(\tilde{\theta}|\theta)}{\partial\tilde{\theta}}\right|_{\tilde{\theta}=\theta}$$

as estimating function for θ (see also Oakes, 1999). It is easily seen that the EM estimate $\hat{\theta}$ solves $g_\theta(\hat{\theta}) = 0$. Let now θ denote the vector of true parameter values in the approximating density (2). This means $E_K\{g_\theta(\theta)\} = 0$, where subscript K indicates, that the expectation is calculated using the density $f_K(\cdot)$. As in the usual likelihood theory, we can expand $g_\theta(\hat{\theta})$ about θ and find in first order approximation

$$\hat{\theta} - \theta = -\left(\frac{\partial g_\theta(\theta)}{\partial \theta^t}\right)^{-1} g_\theta(\theta).$$

Differentiation of $g_\theta(\theta)$ has to take into account that the weights ω_{ik} in (3) depend on θ. Assuming $f(y|z,\theta)$ to be of exponential family form one gets $\partial\omega_{ik}/\partial\theta = \omega_{ik}(s_{ik} - \sum_l \omega_{il}s_{il})$. This yields

$$\frac{\partial g_\theta(\theta)}{\partial \theta^t} = \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}\Big(s_{ik}s_{ik}^t + \frac{\partial s_{ik}}{\partial \theta^t}\Big) - \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{l=1}^{K} w_{ik}w_{il}s_{ik}s_{il}^t. \qquad (5)$$

Using density $f_K(\cdot)$ we find $E_K(w_{ik}s_{ik}s_{ik}^t) = -E_K(w_{ik}\partial s_{ik}/\partial\theta^t)$ so that the first component in (5) has zero expectation. Hence, we can approximate (5) in the usual likelihood fashion by the Fisher type matrix

$$F_K(\theta) := E_K\left(-\frac{\partial g_\theta(\theta)}{\partial \theta^t}\right) = \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{l=1}^{K} E_K\Big(w_{ik}w_{il}s_{ik}s_{il}^t\Big) = E_K\Big(g_\theta(\theta)g_\theta^t(\theta)\Big). \qquad (6)$$

In first order approximation one has $var_K(\hat{\theta}) = F_K^{-1}(\theta)$ where the variance of the regression coefficient $\hat{\beta}$ is obtained by extracting the corresponding submatrix of $F_K^{-1}(\theta)$. One should note that this implicitly takes the variability due to the estimation of the random effect variance σ_z^2 into account.

Formula (5) can be related to the results given in Louis (1982). We can rewrite the observed information (5) as the difference of the complete and the missing information, again using the approximate density $f_K(\cdot)$. Direct calculation provides the complete information $\sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}\partial s_{ik}/\partial\theta^t$, whereas the remaining components in (5) give the missing information. Important for this assignment is the

property that the random effect distribution does not depend on θ. Therefore, the conditional scores s_{ik} can be also defined as the complete scores evaluated at $z_i = \zeta_k$. Hence, $s_i(\theta) := \sum_k w_{ik} s_{ik}$ gives an approximation of the i-th observed score component, which is the conditional expectation of the corresponding complete score, given the data. Rewriting (6) as the sum of observed score variances, i.e. $F_K(\theta) = \sum_{i=1}^n var_K\{s_i(\theta)\}$, leads to the idea suggetsed in Meilijson (1989). He uses the empirical Fisher information matrix to estimate the Fisher information for identically distributed variates. The arguments above do not hold, if the effect density is totally unknown. We show however that the estimating equation approach directly provides variance estimation also in this setting.

2.2 Unknown Effect Distribution

Let now $f(z)$ be unknown so that β and $f(z)$ have to be estimated simultaneously by the NPML approach as suggested by Aitkin & Francis (1995) or Aitkin (1996, 1999), see also Laird (1978). This approach directly generalizes (3), however now $\zeta = (\zeta_1, \ldots, \zeta_K)^t$ and $\pi = (\pi_1, \ldots, \pi_K)^t$ are both treated as unknown and are estimated from the data. Like in the previous section, $Q(\cdot|\cdot)$ is used to approximate (1) with $\theta = (\beta^t, \zeta^t)^t$ and π as unknown parameters. We get (4) as M-step for θ and the masses are obtained from $\pi_k^{(t)} = \sum_{i=1}^n n_i w_{ik}^{(t)} / \sum_{i=1}^n n_i$. For the following expansion of the EM estimating equations it is helpful to reparameterize π by the canonical multinomial representation $\pi_k = \exp\{\vartheta_k - \kappa(\vartheta)\}$ for $k = 1, \ldots, K-1$. This guarantees $\pi_k > 0$ for all $\vartheta = (\vartheta_1, \ldots, \vartheta_{K-1})^t \in \Re^{K-1}$ where $\partial\kappa(\vartheta)/\partial\vartheta_k = \pi_k$. The resulting $K-1$ dimensional estimating equation for ϑ is then

$$g_\vartheta(\theta, \vartheta) := \sum_{i=1}^{n} \sum_{k=1}^{K-1} n_i \tilde{e}_k (w_{ik} - \pi_k), \tag{7}$$

where $\tilde{e}_k$ is a vector of dimension $K-1$ that consists of zeros except of a 1 at the k-th position. Moreover, as in Section 2.1 we have the $p+K$ dimensional estimating

equation for θ

$$g_\theta(\theta,\vartheta) = \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}\, s_{ik} \tag{8}$$

where the weights $w_{ik} = w_{ik}(\theta,\vartheta)$ now depend on both parameters, whereas $s_{ik} = s_{ik}(\theta)$ solely depends on θ. The EM estimates are defined through $g(\hat\theta,\hat\vartheta) = 0$ with $g(\cdot) = \{g_\theta(\cdot)^t, g_\vartheta(\cdot)^t\}^t$. A linear expansion about the true parameter (θ,ϑ) fulfilling $E_K\{g(\theta,\vartheta)\} = 0$ gives in first order approximation

$$\begin{pmatrix} \hat\theta - \theta \\ \hat\vartheta - \vartheta \end{pmatrix} = -\begin{pmatrix} \dfrac{\partial g_\theta(\theta,\vartheta)}{\partial\theta^t} & \dfrac{\partial g_\theta(\theta,\vartheta)}{\partial\vartheta^t} \\ \dfrac{\partial g_\vartheta(\theta,\vartheta)}{\partial\theta^t} & \dfrac{\partial g_\vartheta(\theta,\vartheta)}{\partial\vartheta^t} \end{pmatrix}^{-1} \begin{pmatrix} g_\theta(\theta,\vartheta) \\ g_\vartheta(\theta,\vartheta) \end{pmatrix}.$$

The derivative $\partial g_\theta(\theta,\vartheta)/\partial\theta$ is found as in (5). Differentiation of the weights gives $\partial\omega_{ik}/\partial\vartheta_l = \omega_{ik}\delta(k=l) - \omega_{ik}\omega_{il}$ with $\delta(k=l) = 1$ for $k = l$ and 0 otherwise. This in turn provides the derivatives of (7) and (8) by

$$\begin{aligned}
\frac{\partial g_\theta(\theta,\vartheta)}{\partial\vartheta^t} &= \sum_{i=1}^{n}\sum_{k=1}^{K-1} w_{ik}\, s_{ik}\tilde e_k^t - \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{l=1}^{K-1} w_{ik}\, w_{il}\, s_{ik}\tilde e_l^t = \frac{\partial g_\vartheta^t(\theta,\vartheta)}{\partial\theta} \\
\frac{\partial g_\vartheta(\theta,\vartheta)}{\partial\vartheta^t} &= \sum_{i=1}^{n}\sum_{k=1}^{K-1} (w_{ik} - \pi_k) n_i \tilde e_k \tilde e_k^t - \sum_{i=1}^{n}\sum_{k=1}^{K-1}\sum_{l=1}^{K-1} (w_{ik}w_{il} - \pi_k\pi_l) n_i \tilde e_k \tilde e_l^t.
\end{aligned}$$

Let $F_K(\theta,\vartheta) = -E_K\{\partial g(\theta,\vartheta)/\partial(\theta,\vartheta)\}$ denote the entire Fisher matrix which consists of the submatrices $F_{K,\theta\theta}(\cdot)$, $F_{K,\theta\vartheta}(\cdot)$ and $F_{K,\vartheta\vartheta}(\cdot)$. As before, one gets with (6) the entries for $F_{K,\theta\theta}(\cdot)$ and using $E_K(w_{ik}) = \pi_k$ we find for the remaining components

$$\begin{aligned}
F_{K,\theta\vartheta}(\theta,\vartheta) &= E_K\left(-\frac{\partial g_\theta(\theta,\vartheta)}{\partial\vartheta^t}\right) = E_K\left(-\frac{\partial g_\vartheta^t(\theta,\vartheta)}{\partial\theta}\right) \\
&= E_K\Big(g_\theta(\theta,\vartheta) g_\vartheta^t(\theta,\vartheta)\Big) = \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{l=1}^{K-1} E_K\Big(w_{ik}w_{il}s_{ik}\Big)\tilde e_l^t, \\
F_{K,\vartheta\vartheta}(\theta,\vartheta) &= E_K\left(-\frac{\partial g_\vartheta(\theta,\vartheta)}{\partial\vartheta^t}\right) = E_K\Big(g_\vartheta(\theta,\vartheta) g_\vartheta^t(\theta,\vartheta)\Big) \\
&= \sum_{i=1}^{n}\sum_{k=1}^{K-1}\sum_{l=1}^{K-1} n_i \tilde e_k \tilde e_l^t E_K\Big((w_{ik} - \pi_k)(w_{il} - \pi_l)\Big).
\end{aligned} \tag{9}$$

The inverse of $F_K(\cdot)$ provides a first order approximation for the variance of $(\hat\theta,\hat\vartheta)$. The variance for $\hat\beta$, the parameter of interest, is obtained by extracting the corre-

sponding submatrix of $F_K^{-1}(\cdot)$. It should be noted that this automatically takes the variability stemming from the estimation of $\hat{\zeta}$ and $\hat{\pi}$ into account.

Though $F_K(\cdot)$ in both settings above has a simple structure, analytic calculation is not directed. This is because the weights w_{ik} depend on y_i which makes analytic integration complicated. To overcome this point one can use Monte Carlo integration by drawing n random effects z_i^* from the discrete distribution with masspoints $\hat{\zeta}$ and masses $\hat{\pi}$, where $\hat{\zeta} = \zeta$ and $\hat{\pi}_k = \pi_k$ are fixed for Gaussian random effects. Given z_i^* we draw y_i^* from $f(y_i|z_i^*;\hat{\theta})$ and calculate the simulated Fisher matrix by replacing the expectations in $F_K(\cdot)$ by the simulated empirical moments. Taking average about several simulations provides the Monte Carlo estimate $\tilde{F}_K$, say. It is moreover advisable to correct the variance for a degree of freedom of the estimates, in particular if NPML estimation is used. In simulations we found that the correction factor $n/(n-df)$ appears as appropriate choice, where df is the number of estimated parameters, i.e. the dimension of $F_K(\cdot)$.

3 Example and Simulation

3.1 Example

We demonstrate the variance approximation with a data example taken from the literature. Thall & Vail (1990) analyze data observed at 59 epileptic patients (see also Breslow & Clayton, 1993, or Diggle et al, 1994). The response y_{ij} gives the number of epileptic seizures of patient i during the observation period j where $j = 1, \ldots, 5$. The first period has length 8 weeks denoted by $l_1 = 8$ while the remaining periods have length 2 weeks, i.e. $l_j = 2$ for $j = 2, \ldots, 5$. Between the first and the remaining periods the patients received a therapy t, which was randomized to treatment ($t = 1$) or placebo ($t = 0$). The focus of interest is on inference about

the treatment effect. A detailed description and analysis of the data is found in Diggle et al. (1994, pages 13-16, 183-185). They suggest the mixed effect model

$$E(y_{ij}|z_i, t_i) \quad = \quad l_j \exp\{z_i + \delta(j > 1)\beta_1 + t_i\beta_t + t_i\delta(j > 1)\beta_{t1}\} \tag{10}$$

where $\delta(j > 1)$ is an indicator for the period, i.e. $\delta(j > 1) = 1$ for $j > 1$ and zero otherwise. In (10) β_{t1} gives the therapy effect, β_1 adjusts for a placebo effect while β_t copes for a possible randomization effect. We fit model (10) by NPML estimation which suggests $K = 9$ distinct masspoints. If a model with $K > 9$ masspoints is fitted, the resulting additional masspoints either do not differ from these 9 or have negligible masses. The resulting distribution function from $\hat{\zeta}$ and $\hat{\pi}$ is plotted in Figure 1 and shows a uniform shape. Table 1 gives the estimates for the regression coefficients with variances calculated by Monte Carlo integration. We also used a Gauß Hermite quadrature (K=12) to fit the data where the corresponding fitted random effect distribution is also shown in Figure 1. The GH estimates hardly differ from the NPML estimates. Also the inference allows for similar interpretations. The estimated variances resemble those given in Diggle et al. which are based on a normal approximation of the likelihood as suggested by Breslow & Clayton (1993).

(*Table 1 and Figure 1 about here*)

3.2 Simulation

We run a simulation study to investigate the small sample behavior of the suggested variance approximation. We consider the model $y_{ij}|z_i \sim \text{Poisson}\{\mu_i = 1 + z_i + x_{ij}\beta\}$, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$ with $\beta = 1$, $n = 40$ and $n_i = 2$. The covariate x_{ij} is taken as binary factor, i.e. $x_{ij} \in \{0, 1\}$, with balanced design in the sense $x_{i1} = x_{i2}$

and $x_{ij} = 1$ for half of the data. The z_i's are drawn from the three settings:

$$\begin{aligned}
&\text{a)} \quad \text{normal:} && z_i \sim N(0, 0.5^2), \\
&\text{b)} \quad \text{mixed:} && z_i \sim \begin{cases} N(0, 0.3^2) & \text{with probability } 0.5 \\ N(1, 0.3^2) & \text{with probability } 0.5, \end{cases} \\
&\text{c)} \quad \text{contaminated:} && z_i \sim \begin{cases} N(0, 0.3^2) & \text{with probability } 0.9 \\ N(1.5, 0.1^2) & \text{with probability } 0.1. \end{cases}
\end{aligned}$$

We fit the model by NPML estimation starting with $K = 8$ masspoints and reducing K until all masspoints are different. We also use a GH quadrature with $K = 16$ masspoints. Table 2 shows the mean and standard deviation of 2-0 simulated estimates. Both quadrature formulae provide unbiased estimates and for settings a) and b) they show the same variability. In setting c) however the NPML estimate is clearly less variable than a GH estimate. In general, NPML estimation shows to be not less efficient than GH estimation, even if random effects are normally distributed where the GH procedure gives the right quadrature. Moreover the NPML approach can cope for non-normality of the random effect distribution.

In Table 2 we also report the coverage probability of confidence intervals based on the suggested standard errors. The variance approximations show to work reasonably well with a slightly liberal character though. In the contaminated case on the other hand, the NPML confidence bands are conservative. In general, confidence bands based on NPML estimates behave rather promissing in all three settings.

(*Table 2 about here*)

4 Results and Conclusions

The above results suggest a variance approximation of EM estimates in random effect models based on quadrature formulae. Assuming the differences between the

density $f(\cdot)$ of the random effect and its approximation $f_K(\cdot)$ to be negligible, we can use Fisher type matrices for variance estimation. The same arguments used above also allows to examine differences between masspoints ζ_j and ζ_k or the relevance of the masses π_k. This indirectly gives an exploratory procedure to evaluate the number K of masspoints used and comply with the proposals in Laird (1978).

References

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, **6**, 251-262.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, **55**, 218-234.

Aitkin, M., Francis, B.J. (1995). Fitting overdispersed generalized linear models by nonparametric maximum likelihood. *The GLIM Newsletter*, **25**, 37-45.

Anderson, D.A., Aitkin, M. (1985). Variance component models with binary response: interviewer variability. *Journal of the Royal Statistical Society, Series B*, **47**, 203-210.

Breslow, N.E., Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9-25.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. (with discussion). *Journal of the Royal Statistical Society B*, **39**, 1-38.

Diggle, P.J., Liang, K.-Y., Zeger, S.L. (1994). *Analysis of longitudinal data*. Oxford University Press, Oxford

Hinde, J. (1982). Compound Poisson regression models. In: *GLIM 82* (R. Gilchrist, ed.), 109-121. Springer-Verlag, New York.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, **73**, 805-811.

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society B*, **44**, 226-233.

McLachlan, G.J., Krishnan, T. (1997). *The EM Algorithm and Extensions*. John Wiley & Sons, Inc., New York.

Meilijson, I. (1989). A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society B*, **51**, 127-138.

Oakes, D. (1999). Direct calculation of the information matrix via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **61**, 479-482.

Thall, P.F., Vail, S.C. (1990). Some Covariance Models for Longitudinal Count Data With Overdispersion. *Biometrics*, **46**, 537–547.

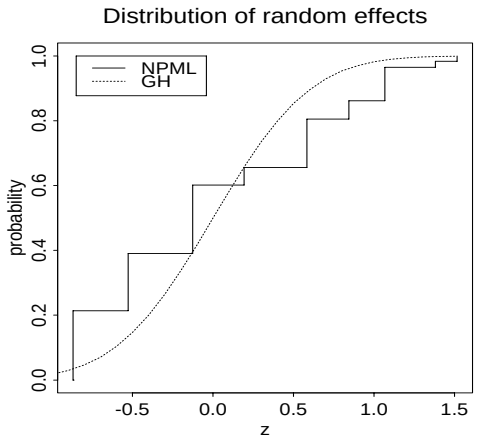

Figure 1: Estimated random effect distribution for epileptic seizure data

	NPMLE Fit		Gauß-Hermite Fit	
effect	$\hat{\beta}$	$\sqrt{\text{var}(\hat{\beta})}$	$\hat{\beta}$	$\sqrt{\text{var}(\hat{\beta})}$
β_t	-.126	.194	.048	.070
β_1	.121	.084	.122	.052
β_{t1}	-.295	-.079	-.298	.083

Table 1: Estimates and standard errors for epileptic seizure data

random effect	mean($\hat{\beta}$)	s.e.($\hat{\beta}$)	coverage 90 %	95 %
NPMLE				
normal	0.98	.21	87.1	92.6
mixed	1.02	.22	83.3	87.9
contaminated	1.00	.07	93.8	98.1
GH				
normal	.99	.22	87.0	91.5
mixed	.98	.20	84.8	91.4
contaminated	1.00	.23	84.5	91.0

Table 2: Mean and standard error of EM estimates and the resulting coverage probability of confidence intervals based on 200 simulations.