The **R** Package **geepack** for Generalized Estimating Equations

Ulrich Halekoh
Danish Institute of Agricultural Sciences

Søren Højsgaard
Danish Institute of Agricultural Sciences

Jun Yan
University of Iowa

Abstract

This paper describes the core features of the R package **geepack**, which implements the generalized estimating equations (GEE) approach for fitting marginal generalized linear models to clustered data. Clustered data arise in many applications such as longitudinal data and repeated measures. The GEE approach focuses on models for the mean of the correlated observations within clusters without fully specifying the joint distribution of the observations. It has been widely used in statistical practice. This paper illustrates the application of the GEE approach with **geepack** through an example of clustered binary data.

Keywords: generalized estimating equation, random effect, mixed model, quasi-likelihood.

1. Introduction

Generalized Estimating Equations (GEE) (Liang and Zeger 1986) are a general method for analyzing data collected in clusters where 1) observations within a cluster may be correlated, 2) observations in separate clusters are independent, 3) a monotone transformation of the expectation is linearly related to the explanatory variables and 4) the variance is a function of the expectation. It is essential to note that the expectation and the variance referred to in points 3) and 4) are conditional given cluster-level or individual-level covariates.

There are several approaches to extend generalized linear models to clustered data. Mixed effect models and transition models (Diggle, Liang, and Zeger 1994, Chapter 7, 9-10) fully specify the joint distribution within clusters via latent variables or conditional dynamics. With the presence of random effects, likelihood estimation necessitates the integration over the random effects distributions, which may be numerically intractable. Lee and Nelder (1996,

2001) introduced hierarchical generalized linear models and showed that the integration may be avoided by working on the h-likelihood. Compared to these approaches, the method of GEE fits marginal mean models with the advantage that only correct specification of marginal means is needed for the parameter estimator to be consistent and asymptotically normal. This approach has become an important tool in analyzing longitudinal data or repeated measures arising in a wide variety of applications. For a discussion on the relation between marginal and mixed effects models, see Heagerty and Zeger (2000) and Nelder and Lee (2004).

Several implementations of GEE have become available (Horton and Lipsitz 1999). The basic approach of Liang and Zeger (1986) is available in SAS (SAS Institute Inc. 1999, **proc genmod**), Stata (StataCorp LP 2005), XLISP-STAT (Lumley 1996) and in S-PLUS by the packages **oswald** (Smith 1998) and **gee** or **yags** (Carey 2002, 2004). The last two packages have been ported to R (R Development Core Team 2005). The R package **geepack** implements the basic approach and some extensions (Yan 2002; Yan and Fine 2004). Three features of **geepack** distinguish it from other implementations: 1) There is an interface function `geeglm` which is designed to be as similar to `glm` as possible; 2) A jackknife variance estimator is available as an alternative to the sandwich estimator; and 3) Covariates can be incorporated into the scale and correlation parameters in a similar fashion to the mean modeling. In this paper, we illustrate the aspects of **geepack** with the focus on the first two features.

The paper is organized as follows. In Section 2 we introduce an example dataset on repeated measures of binary data. In Section 3 we outline the GEE approach and in Section 4 we describe the features of the `geeglm` function that implements the approach in **geepack**. We close the article with an analysis of the data in Section 5 and a conclusion in Section 6.

2. An example data set

To illustrate the type of problems well suited for the GEE approach we consider a data set on respiratory illness. The data is provided in **geepack** and detailed information about the data can be found in Koch, Carr, Amara, Stokes, and Uryniak (1990). Briefly, the data comes from a clinical study in which the effect of a treatment of patients with respiratory illness was

Baseline	Visit / Treat	Response pattern															Sum
	1	0	1	0	0	0	1	1	1	0	0	1	1	1	0	1	
	2	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	
	3	0	0	0	1	0	0	1	0	1	1	1	0	1	1	1	
	4	0	0	0	0	1	0	0	1	0	1	0	1	1	1	1	
0	A	7	2	2	2	1	0	1	0	1	0	1	2	0	4	7	30
	P	18	1	0	2	1	2	0	0	1	0	0	1	2	0	3	31
1	A	0	0	0	0	0	0	1	1	0	0	4	0	1	0	17	24
	P	1	4	1	0	0	0	0	1	1	3	1	1	2	1	10	26
Sum		26	7	3	4	2	2	2	2	3	3	6	4	5	5	37	111

Table 1: Number of patients for the different response patterns classified by baseline response and treatment. The patterns are ordered according to increasing numbers of positive responses.

		Number of positive reponses				
		0	1	2	3	4
Sex	F	7	3	3	3	7
	M	19	13	9	17	30
Center	1	18	9	6	11	12
	2	8	7	6	9	25

Table 2: Number of patients for the number of positive responses across the four visits for sex and center.

examined. A total of 111 patients from two clinical centers were randomized to receive either placebo or an active treatment. At four examination visits, the respiratory state of a patient was classified as good (=1) or poor (=0). Variables characterizing a patient were: `center` (1,2), treatment `treat` (A=active, P=placebo), `sex` (M=male, F=female), `age` (in years) at baseline and `baseline` respiratory state which is also binary. The values of the covariates were constant for the repeated elementary observations on each patient.

Table 1 shows the number of patients for the response patterns across the four visits classified by baseline status and treatment. Patients with baseline respiratory status equal to 0 seem either to have a low or large number of positive responses whereas patients with a baseline of 1 tend to respond positively. Table 2 describes the distribution of the number of positive responses per patient for sex and center.

Figure 1 presents the plot of age against the proportion of positive responses for each patient. It indicates a quadratic relationship between the proportions and the age. We fit a logistic model to the data (which would be appropriate if there were no time effect and no spread in the response probabilities for patients with the same covariate values):

```
R> m.glm <- glm(outcome ~ baseline + center + sex + treat + age + I(age^2),
+     data = respiratory, family = binomial)
```

The correlation matrix of the Pearson residuals within a patient based on the `glm`-fit is shown in Table 3 and indicates an appreciable correlation within patient measurements.

	visit1	visit2	visit3	visit4
visit1	1.00	0.35	0.24	0.30
visit2	0.35	1.00	0.34	0.28
visit3	0.24	0.34	1.00	0.36
visit4	0.30	0.28	0.36	1.00

Table 3: Correlation matrix for the measurements at different visits based on the Pearson residuals from the logistic model.

3. Theory of GEE

For the regression analysis of correlated observations, Liang and Zeger (1986) introduced the GEE approach. This approach generalized the estimation method of quasi-likelihood of

Figure 1: Relation of `age` to the proportion of positive responses. The smooth line is a spline.

Wedderburn (1974) to correlated data. An alternative generalization was proposed by Lee and Nelder (1996, 2001). An extensive review of the development of the GEE approach is given by Ziegler, Kastner, and Blettner (1998).

Consider a sample of $i = 1, \ldots, K$ independent multivariate observations $Y_i = (Y_{i1}, \ldots, Y_{it}, \ldots, Y_{in_i})$. Here i may represent a cluster with n_i observations. The expectations $\mathsf{E}(Y_{it}) = \mu_{it}$ are related to the p dimensional regressor vector x_{it} by the mean-link function g

$$g(\mu_{it}) = x_{it}^\top \beta. \tag{1}$$

Let

$$\mathsf{VAR}(Y_{it}) = \phi a_{it}, \tag{2}$$

where ϕ is a common scale parameter and $a_{it} = a(\mu_{it})$ is a known variance function. Let $R_i(\alpha)$ be a working correlation matrix completely described by the parameter vector α of length m. Let

$$V_i = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2}$$

be the corresponding working covariance matrix of Y_i, where A_i is the diagonal matrix with entries a_{it}. For given estimates $(\hat{\phi}, \hat{\alpha})$ of (ϕ, α) the estimate $\hat{\beta}$ is the solution of the equation

$$\sum_{i=1}^{K} \frac{\partial \mu_i^\top}{\partial \beta} V_i^{-1}(Y_i - \mu_i) = 0.$$

Liang and Zeger (1986) suggest to use consistent moment estimates for ϕ and α. This yields an iterative scheme which switches between estimating β for fixed values of $\hat{\phi}$ and $\hat{\alpha}$ and estimating (ϕ, α) for fixed values of $\hat{\beta}$. This scheme yields a consistent estimate for β. Moreover, $K^{1/2}(\hat{\beta} - \beta)$ is asymptotically multivariate normally distributed with zero mean and

name	function $v(\mu)$
`gaussian`	identity
`binomial`	$\mu(1-\mu)$, $\mu \in (0,1)$
`poisson`	μ, $\mu > 0$
`gamma`	μ^2, $\mu > 0$

Table 4: Variance functions in `geeglm`.

covariance matrix $\Sigma = \lim\limits_{K \to \infty} K\Sigma_0^{-1}\Sigma_1\Sigma_0^{-1}$ where

$$\Sigma_0 = \sum_{I=1}^{K} \frac{\partial \mu_i^\top}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta^\top}, \quad \Sigma_1 = \sum_{i=1}^{K} \frac{\partial \mu_i^\top}{\partial \beta} V_i^{-1} \mathsf{COV}(Y_i) V_i^{-1} \frac{\partial \mu_i}{\partial \beta^\top}. \tag{3}$$

Replacing β, ϕ and α by consistent estimates and the covariance matrix $\mathsf{COV}(Y_i)$ by $(Y_i - \mu_i)(Y_i - \mu_i)^\top$ in (3) yields a so called sandwich estimate $\hat{\Sigma}$ of Σ. The estimate $\hat{\Sigma}$ is a consistent estimate of Σ even if the working correlation matrices $R_i(\alpha)$ are misspecified.

4. The `geeglm` function

The `geeglm` function largely follows the syntax of the `glm` function and many of the methods available for `glm` objects are also available for `geeglm` objects. We discuss in the following the most important arguments of the `geeglm` function and the `anova` method for comparing models by Wald tests. These will be exemplified in Section 5.

4.1. Variance and link functions (`family`)

The variance function is specified by the `family` argument in `geeglm` and is identified by the name of the corresponding distribution in a generalized linear model. The available variance functions are given in Table 4. The available link functions for the mean are the same as those in `glm` with the exception of the `cauchit` link for the `binomial` family.

4.2. Working correlation (`corstr`, `zcor`)

Four pre-defined working correlation structures are available and are specified via the argument `corstr` (Table 5). It is also possible to provide a correlation matrix the entries of which remain fixed under the computation. This may be necessary if the estimate for α does not yield a positive definite $R(\alpha)$ (Chaganty 1997). Using a fixed working correlation (`corstr="fixed"`) may still yield efficient estimates (Chaganty and Joe 2004) for β. Additionally, a user-defined correlation structure (`corstr="userdefined"`) can be provided by expressing the correlation parameters as linear combinations of covariates. Given a $n_i \times m$ matrix X_i of covariates, the upper diagonal correlations parameters $r_i = (r_{i,12}, r_{i,13}, \ldots, r_{i,1n_i}, r_{i,23}, \ldots, r_{i,n_{i-1}n_i})$ of the working correlation matrix $R_i(\alpha)$ can be written as $r_i = X_i\alpha$. The `zcor` argument takes the concatenated matrices $(X_1^\top, \ldots, X_K^\top)^\top$ as the design matrix for the working correlation. For the fixed correlation matrix one has simply $m = 1$ and $X_i = r_i$. Some useful correlation structures are defined as simple linear restrictions of the unstructured working correlation matrix for which the corresponding design matrix is provided by the `genZcor` function of the

name	$R(\alpha)$
independence	$\mathrm{COR}(Y_{it}, Y_{it'}) = 0, \quad t \neq t'$
exchangeable	$\mathrm{COR}(Y_{it}, Y_{it'}) = \alpha, \quad t \neq t'$
ar1	$\mathrm{COR}(Y_{it}, Y_{it'}) = \alpha^{|t-t'|}$
unstructured	$\mathrm{COR}(Y_{it}, Y_{it'}) = \alpha_{tt'}, t \neq t'$

Table 5: Working correlations in `geeglm`.

geepack package. For example, the entries in each off-diagonal of a Toeplitz correlation matrix are equal, $r_{i,kl} = r_{i|k-l|}$. Hence, the corresponding design matrix is obtained by adding those columns in the unstructured design matrix with the same difference in the k, l indices.

4.3. Missing values (`waves`)

In case of missing values, the GEE estimates are consistent if the values are missing completely at random (Rubin 1976). The `geeglm` function assumes by default that observations are equally separated in time. Therefore, one has to inform the function about different separations if there are missing values and other correlation structures than the independence or exchangeable structures are used. The `waves` arguments takes an integer vector that indicates that two observations of the same cluster with the values of the vector of k respectively l have a correlation of r_{kl}.

4.4. Jackknife variance estimates (`std.err`)

For a small number of clusters ($K \leq 30$) the sandwich variance estimator exhibits bias and Paik (1988) recommended using the jackknife variance estimator. It is defined as

$$\frac{K-p}{K} \sum_{i=1}^{K} (\hat{\beta}_{-i} - \hat{\beta})(\hat{\beta}_{-i} - \hat{\beta})^\top,$$

where p is the number of parameters in the mean structure and $\hat{\beta}_{-i}$ are the estimates of β leaving out the ith cluster. By default the `geeglm` function returns the sandwich estimates. Specifying `std.err="fij"` the fully iterated jackknife estimate is returned. Additionally, the computationally less demanding approximate jackknife estimate (`std.err="jack"`) or a one-step jackknife estimate (`std.err="j1s"`) can be obtained. Simulation studies (Ziegler, Kastner, Brunner, and Blettner 2000; Yan and Fine 2004) indicate that the approximate jackknife estimates are in many cases in good agreement with the fully iterated estimates.

4.5. `anova` method

The `anova` method allows either to produce a table of tests for sequentially adding terms to a model or to compare two nested models. The test statistic for testing the difference between model M1 and a submodel M2 is the Wald test statistic. If β is the parameter of model M1 and the submodel M2 is obtained from M1 by setting $C\beta = 0$, where C is a rank q contrast matrix and $\hat{\Sigma}_{\hat{\beta}}$ is a consistent estimate of the covariance matrix of $\hat{\beta}$, the test statistic is $\hat{\beta}^\top C^\top (C\hat{\Sigma}_{\hat{\beta}} C^\top)^{-1} C\hat{\beta} \sim \chi_q^2$.

5. Analysis of the respiratory data

In the initial description of the respiratory data we saw appreciable within-patient correlation. We now fit the logistic model with the same mean structure and the binomial variance function $a(\mu_{it}) = \mu_{it}(1 - \mu_{it})$ using the GEE approach. We estimate β under the four pre-defined working correlations and a user-defined Toeplitz working correlation. For example, the fit with the exchangeable working correlation is obtained by

```
R> m.ex <- geeglm(outcome ~ baseline + center + sex + treat + age + I(age^2),
+     data = respiratory, id = interaction(center, id),
+     family = binomial, corstr = "exchangeable")
```

The design matrix for the Toeplitz working correlation was constructed by first obtaining the design matrix for the unstructured correlation using the `genZcor` function and then adding the appropriate columns. The use of the `waves` argument is not necessary because there are no missing observations.

```
R> zcor <- genZcor(clusz = c(xtabs(~ id + center, data = respiratory)),
+     waves = respiratory$visit, corstrv = 4)
R> zcor.toep <- matrix(NA, nrow(zcor), 3)
R> zcor.toep[,1] <- apply(zcor[,c(1,4,6)], 1, sum)
R> zcor.toep[,2] <- apply(zcor[,c(2,5)], 1, sum)
R> zcor.toep[,3] <- zcor[,3]
```

The fit with the Toeplitz working correlation is obtained by

```
R> m.toep <- geeglm(outcome ~ baseline + center + sex + treat + age + I(age^2),
+     data = respiratory, id = interaction(center, id),
+     family = binomial, corstr = "userdefined", zcor = zcor.toep)
```

	Est.GLM	SE.GLM	Est.Ind	SE.Ind	Est.Exc	SE.Exc
(Intercept)	3.87	0.96	3.87	1.31	3.87	1.31
baseline1	1.89	0.25	1.89	0.38	1.89	0.38
center2	0.51	0.25	0.51	0.38	0.51	0.38
sexM	-0.45	0.32	-0.45	0.48	-0.45	0.48
treatP	-1.32	0.24	-1.32	0.38	-1.32	0.38
age	-0.21	0.05	-0.21	0.06	-0.21	0.06
age^2/100	0.26	0.06	0.26	0.08	0.26	0.08
	Est.Ar1	SE.Ar1	Est.Un	SE.Un	Est.Toep	SE.Toep
(Intercept)	3.48	1.29	3.86	1.30	3.83	1.31
baseline=1	1.90	0.37	1.92	0.37	1.89	0.37
center=2	0.59	0.37	0.50	0.38	0.52	0.38
sex=M	-0.42	0.48	-0.44	0.48	-0.45	0.48
treat=P	-1.25	0.37	-1.32	0.38	-1.32	0.37
age	-0.19	0.06	-0.21	0.06	-0.21	0.06
age^2/100	0.24	0.08	0.26	0.08	0.25	0.08

Table 6: Comparison of parameter estimates.

	sand	fij	j1s	jack
(Intercept)	1.31	1.39	1.35	1.35
baseline=1	0.38	0.39	0.38	0.38
center=2	0.38	0.39	0.39	0.39
sex=M	0.48	0.51	0.50	0.50
treat=P	0.38	0.39	0.38	0.38
age	0.06	0.06	0.06	0.06
$age^2/100$	0.08	0.08	0.08	0.08

Table 7: Standard errors for the parameters from the exchangeable working correlation fit for the different variance estimators (sand: sandwich, fij: fully iterated jackknife, j1s: 1-step jackknife; jack: approximate jackknife).

A table of the estimated parameters, their standard deviations, and Wald tests is obtained by

```
R> coef(summary(m.ex))
```

Table 6 presents the results from the fit of the simple generalized linear model and the fits obtained by using GEE with the different working correlations. In all cases the estimates of β are in close agreement. The standard errors under all five working correlation assumptions are practically identical and about 30% larger than under the assumption of independent observations.

To illustrate the facility of different variance estimates, Table 7 presents a comparison of the sandwich variance estimate and various jackknife variance estimates (Yan and Fine 2004).

For this particular dataset, all these variance estimates are virtually the same.

To test whether the effect of `age` may be removed, we first fit a model without the `age` terms

```
R> m.ex.0 <- update(m.ex, . ~ . - age - I(age^2))
```

and then compare this model to the full model using the `anova` method. This yields a p-value that underlines the importance of the `age` effect.

```
R> anova(m.ex, m.ex.0)

Analysis of 'Wald statistic' Table
Model 1 outcome ~ baseline + center + sex + treat + age + I(age^2)
Model 2 outcome ~ baseline + center + sex + treat
  Df     X2 P(>|Chi|)
1  2 12.403     0.002
```

6. Conclusion

The R package **geepack** provides a flexible estimating equations approach for estimating the covariate depending mean, scale and correlations parameters of correlated observations. In

this paper we described the `geeglm` function of the package that is close in syntax to the `glm` function and implements the estimating equation approach of Liang and Zeger (1986). Additionally, the function allows to use a fixed user-defined correlation structure and several jackknife variance estimators. The latter are preferable to the sandwich estimator in case of a small number of clusters.

geepack has additional features which allow the scale parameters and the parameters of the working correlation matrix to be modeled as functions of explanatory variables, see Yan (2002) and Yan and Fine (2004). Using additional estimating equations for these parameters the approach yields consistent estimates for β, the scale and correlation parameters. The function `ordgee` of **geepack** allows the analysis of ordinal data according to the method proposed by Heagerty and Zeger (1996).

References

Carey VJ (2002). **gee**: *Generalized Estimation Equation Solver*. R package version 4.13-10; Ported from S-PLUS to R by Thomas Lumley (versions 3.13 and 4.4) and Brian Ripley (version 4.13).

Carey VJ (2004). **yags**: *Yet Another GEE Solver*. R package version 4.0-1, URL http://www.biostat.harvard.edu/~carey/.

Chaganty NR (1997). "An Alternative Approach to the Analysis of Longitudinal Data via Generalized Estimating Equations." *Journal of Statistical Planning and Inference*, **63**, 39–54.

Chaganty NR, Joe H (2004). "Efficiency of Generalized Estimating Equations for Binary Responses." *Journal of the Royal Statistical Society B*, **66**(4), 851–860.

Diggle PJ, Liang KY, Zeger SL (1994). *Analysis of Longitudinal Data*. Oxford University Press, New York, first edition. ISBN 0 19 852284 3.

Heagerty PJ, Zeger SL (1996). "Marginal Regression Models for Clustered Ordinal Measurements." *Journal of the American Statistical Association*, **91**, 1024–1036.

Heagerty PJ, Zeger SL (2000). "Marginalized Multilevel Models and Likelihood Inference." *Statistical Science*, **15**(1), 1–26.

Horton NJ, Lipsitz SR (1999). "Review of Software to Fit Generalized Estimating Equation Regression Models." *The American Statistician*, **53**, 160–169.

Koch GG, Carr GJ, Amara IA, Stokes ME, Uryniak TJ (1990). "Categorical Data Analysis." In DA Berry (ed.), "Statistical Methodology in the Pharmaceutical Sciences," pp. 389–473. Marcel Dekker, New York. ISBN 0 8247 8117 1. The data is copyright by the authors. Reproduced by permission of Routledge/Taylor & Francis Group, LLC.

Lee Y, Nelder JA (1996). "Hierarchical Generalized Linear Models." *Journal of the Royal Statistical Society B*, **58**(4), 619–678.

Lee Y, Nelder JA (2001). "Hierarchical Generalised Linear Models: A Synthesis of Generalised Linear Models, Random-Effect Models and Structured Dispersions." *Biometrika*, **88**(4), 987–1006.

Liang KY, Zeger S (1986). "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, **73**(1), 13–22.

Lumley T (1996). "XLISP-STAT Tools for Building Generalised Estimating Equation Models." *Journal of Statistical Software*, **1**(3), 1–20.

Nelder JA, Lee Y (2004). "Conditional and Marginal Models: Another View." *Statistical Science*, **19**(2), 219–238.

Paik MC (1988). "Repeated Measurement Analysis for Nonnormal Data in Small Samples." *Communications in Statistics - Simulation and Computation*, **17**, 1155–1171.

R Development Core Team (2005). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rubin DB (1976). "Inference and Missing Data." *Biometrika*, **63**(3), 581–592.

SAS Institute Inc (1999). *SAS/STAT User's Guide, Version 8.* SAS Institute Inc., Carry, NC, United States of America.

Smith DM (1998). "**oswald**: Object Oriented Software for the Analysis of Longitudinal Data in S-PLUS." S-PLUS package version 3.4, URL http://www.maths.lancs.ac.uk/Software/Oswald/.

StataCorp LP (2005). *Stata 9 Longitudinal/Panel Data Reference Manual.* StataCorp LP, Collage Station, Texas.

Wedderburn RWM (1974). "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method." *Biometrika*, **61**(3), 439–447.

Yan J (2002). "**geepack**: Yet Another Package for Generalized Estimating Equations." *R News*, **2**, 12–14.

Yan J, Fine J (2004). "Estimating Equations for Association Structures." *Statistics in Medicine*, **23**, 859–874.

Ziegler A, Kastner C, Blettner M (1998). "The Generalized Estimation Equations: An Annotated Bibliography." *Biometrical Journal*, **40**, 115–139.

Ziegler A, Kastner C, Brunner D, Blettner M (2000). "Familial Association of Lipid Profile: a Generalized Estimation Equation Approach." *Statistics in Medicine*, **19**, 3345–3357.

Affiliation:

Ulrich Halekoh
Unit of Statistics and Decision Analysis
Department of Genetics and Biotechnology
Danish Institute of Agricultural Sciences
8830 Tjele, Denmark
E-mail: ulrich.halekoh@agrsci.dk
URL: http://www.agrsci.dk/afdelinger/forskningsafdelinger/gbi/medarbejdere/uhh

Søren Højsgaard
Unit of Statistics and Decision Analysis
Department of Genetics and Biotechnology
Danish Institute of Agricultural Sciences
8830 Tjele, Denmark
E-mail: Soren.Hojsgaard@agrsci.dk
URL: http://www.agrsci.dk/afdelinger/forskningsafdelinger/gbi/medarbejdere/shd

Jun Yan
Department of Statistics and Actuarial Science
University of Iowa
Iowa City, IA 52242-1409, United States of America
E-mail: jyan@stat.uiowa.edu
URL: http://www.stat.uiowa.edu/~jyan/