MISSPECIFICATION OF FRAILTY RANDOM EFFECTS IN A CLUSTERED SURVIVAL DATA.

John Kiche[1], Oscar Ngesa[2], George Orwa[3]
[1] Pan Africa University Institute for Basic Sciences, Technology and Innovation
Email:kicheopiyo07@gmail.com
[2]Taita Taveta University
Email:oscanges@ttu.ac.ke
[3]Jomo Kenyatta University of Agriculture and Technology
Email:gorwa@fsc.ac.ke

September 28, 2018

Abstract

Survival Analysis models the time it takes until an event occurs. The prototypical event is death, from which the name Survival Analysis is derived.Accordingly, each time Survival Analysis is studied, aspects of some selected rates or reliability of some study are usually considered. Frailty modelling has been used in this study as the statistical tool for analysing the time-event data. Parametric and non-parametric models and the frailty models are fitted to help derive the required conclusions.The impact of misspecification of frailty random effects in a survival data using parametric frailty modelling approach were determined during this research study.It is expected that these approaches would produce less bias estimates compared to the results achieved of the estimates when the misspecification of the frailty random effects are ignored.

Keywords: misspecification, frailty modeling, clustered data

I. Introduction

Study of clustered survival data has become one of the most important research issues of the developing countries. Some of the reviewed literature based on frailty models and misspecification of the frailty random effects are discussed as follows. Correlated or clustered failure time data occur in many fields such as medical studies (Cai and Prentice, 1995; Kalbfleisch and Prentice, 2002). In many cases the failure times of interest may not be observed exactly but are known only to belong to certain intervals. Such data are usually referred to as interval-censored failure time data, and they could arise naturally in, for example, periodic follow-up studies where each study subject is observed only at discrete time points (Finkelstein and Wolfe, 1986; Sun, 2006; Wang et al., 2006). Regression analysis of clustered interval-censored data where the failure times of

interest may be related to the cluster size. In other words, the cluster size may be informative for the failure times of interest and thus obviously needs to be taken into account in the analysis. The general idea of frailty gives a possible way of introducing unobserved heterogeneity and associations into models for a given survival data, upon which, the chance of misspecification when using the resulting frailty model is high. More so,tackling the foregoing misspecification by use of parametric frailty model approaches has not been thoroughly explored.

Parameter estimation is done by maximising the marginal log-likelihood. Survival analysis generally constitute methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. The event can be death, occurrence of a disease, marriage, divorce, etc. The time to event or survival time can be measured in days, weeks, years, etc Data that measures lifetime or the length of time until the occurrence of an event are called failure time, lifetime or survival data. For example variables of interest might be the survival time for patients. A frailty model is a random effects model for time variables, where the random effect (the frailty) has a multiplicative effect on the hazard. It can be used for univariate (independent) failure times, that is to describe the influence of unobserved covariates in a proportional hazard model. Frailty models are the survival data analog to regression models, which account for heterogeneity and random effects.Frailty random effects need to have spatial components because they vary from one region to another.

Study of clustered survival data has become one of the most important research issues of the developing countries. Some of the reviewed literature based on frailty models and misspecification of the frailty random effects are discussed as follows. Correlated or clustered failure time data occur in many fields such as medical studies (Cai and Prentice, 1995; Kalbfleisch and Prentice, 2002). In many cases the failure times of interest may not be observed exactly but are known only to belong to certain intervals. Such data are usually referred to as interval-censored failure time data, and they could arise naturally in, for example, periodic follow-up studies where each study subject is observed only at discrete time points (Finkelstein and Wolfe, 1986; Sun, 2006; Wang et al., 2006). Regression analysis of clustered interval-censored data where the failure times of interest may be related to the cluster size. In other words, the cluster size may be informative for the failure times of interest and thus obviously needs to be taken into account in the analysis.

In essence, the frailty concept goes back to work of Greenwood and Yule(1920) on accident proneness. The concept of frailty was introduced by Vaupel and Stallard (1979) showing that some individuals are more frail or susceptible or at risk than others although they may appear to be similar while considering the observable or measurable attributes like sex, age and weight. Frailty models are extensions of the proportional hazards model which is best known as the Cox model(Cox,1972),the most popular model in survival analysis. Log-normal frailty models are especially useful in modelling dependence structures innmultivariate frailty models, for example in McGilchrist and Aisbett (1991), McGilchrist(1993), Lillard (1993), Lillard et al. (1995), Xue and Brookmeyer (1996), Sastry (1997),Gustafson (1997), Ripatti and Palmgren (2000); Ripatti *et al*. (2002), Huang and Wolfe (2002). However, the log-normal distribution has also been applied in univariate cases, for example by Flinn and Heckman (1982).

The inverse Gaussian (inverse normal) distribution was introduced as an alternative to the gamma distribution by Hougaard (1984) and has been used for example by Manton et al. (1986), Klein *et al*. (1992), Keiding et al. (1997) and Price and Manatunga (2001). Positive stable frailty model was introduced as a frailty distribution by Hougaard (1986b) and applied for example by Wang *et al*. (1995) and Manatunga and Oakes (1999). Fine *et al*. (2003) and Martinussen and Pipper (2005) recently suggested new estimation procedures in the shared positive stable frailty model. It was further extended by Hougaard's power variance function distribution (Hougaard, 1986a) and Aalen's compound Poisson distribution (Aalen 1988, 1992). All moments of this distribution

are infinite. This result is important with respect to identifiability issues treated by Elbers and Ridder(1982). They found that a finite mean of the frailty distribution is one condition (among others) for identifiability of univariate frailty models.

II. Methods

i. Baseline hazards

The Weibull hazard(dist="weibull") is:

$$h(t;\rho,\lambda = \rho\lambda t^{\rho-1}) \tag{1}$$

with $\rho, \lambda > 0$.

The inverse Weibull (or Frechet) hazard (dist="inweibull" or dist="frechet") is :

$$h(t;\rho,\lambda) = \frac{\rho\lambda t^{-p-1}}{exp(\lambda t^{-p} - 1)} \tag{2}$$

with $\rho, \lambda > 0$.

The exponential hazard(dist="exponential") is :

$$h(t;\lambda) = \lambda \tag{3}$$

with $\lambda > 0$.

The Gompertz hazard(dist="gompertz") is;

$$h(t;\gamma,\lambda) = \lambda exp(\gamma t) \tag{4}$$

with $\gamma, \lambda > 0$.

The lognormal hazard (dist="lognormal") is :

$$h(t;\mu,\sigma) = \frac{\phi(\frac{log(t)-\mu}{\sigma})}{\sigma t[1 - \Phi(\frac{log(t)-\mu}{\sigma})]} \tag{5}$$

with $\mu \epsilon R, \sigma > 0$, and where ϕ and Φ are the density and distribution functions of a standard Normal random variable.

III. Frailty modelling

Frailty is an unobserved random proportionality factor that modifies the hazard function of an individual or a group of related individuals. The notion of frailty provides a convenient way to introduce random effects, association and unobserved heterogeneity into models for survival data.

Frailty models are survival models with at least one random effect. For example, a proportional hazards model can be written as:

$$\lambda_j(t) = \lambda_0(t)e^{\beta' X_j(t)} \tag{6}$$

The subscript j indexes the individual. This model becomes a frailty model by adding a random effect as in:

$$\lambda_{ij}(t) = \lambda_0(t)e^{\beta' X_{ij}(t)+\omega_i} \tag{7}$$

The subscript i indexes a cluster of individuals and ω is the random effect.

This model now becomes a spatial frailty model by adding a spatial structure as in:

$$\lambda_{ij}(t) = \lambda_0(t)e^{\beta' X_{ij}(t)+\omega_i+\nu_i} \tag{8}$$

The subscript i indexes a cluster of individuals and ω_i is the random effect while ν_i is the spatial structured heterogeneity.

i. Univariate frailty models

In a univariate model, the random effect pertains to an individual with one observation and accounts for unobserved heterogeneity at the level of the individual. For example, a common response of interest in oncology is disease free survival (DFS). A common scenario in clinical trials in oncology is one in which each patient's tumors are removed at the start of the trial (baseline), a treatment is administered, and then the patient's disease status is recorded at pre-specified intervals. The time to the first recurrence of the disease is the response of interest.

ii. Positive stable frailty model

A distribution is called positive stable if the appropriately normalized sum of n independent random variables from this distribution has the same distribution. The normalizationis given by $n^{1/\gamma}$, where the index γ must be in the range of $(0; 1]$to get a distribution on positive numbers. Despite the fact that no closed form expressions exist for the probability density or the survival function of a random variable with positive stable distribution, the Laplace transform has a very simple form:

$$\mathbf{L}(s) = e^{\frac{-ks^\gamma}{\gamma}} \tag{9}$$

For reasons of identiability, we restrict the two-parameter frailty distribution to the case of $k = \gamma$.

Consequently,

$$s(t) = \mathbf{L}(\Lambda)_0(t)) = e^{-\Lambda_0(t)^\gamma} \tag{10}$$

$$f(t) = \gamma\lambda_0(t)\Lambda_0(t)^{\gamma-1}e^{-\Lambda_0(t)^\gamma} \tag{11}$$

$$\gamma(t) = \gamma\lambda_0(t)\Lambda_0(t)^{\gamma-1} \tag{12}$$

iii. PVT frailty model

An extended family of frailty distributions, including gamma, inverse Gaussian as well as positive stable distributions, is the family of power variance function distributions,suggested by Tweedy (1984) and later derived independently by Hougaard (1986a). This is a three parameter family denoted by $PVF(\gamma, k, \lambda)$. The Laplace transform is

$$\mathbf{L}(s) = e^{\frac{-k}{\gamma}((\lambda+s)^{\gamma}-\lambda^{\gamma})}$$

.
Expectation and variance of a PVF distributed random variable Z are

$$\mathbf{E}(Z) = k\lambda^{\gamma-1}$$

and

$$\mathbf{V}(Z) = k(1-\gamma)\lambda^{\gamma-2}$$

The resulting survival function is given by

$$S(t) = e^{\frac{-k}{\gamma}((\lambda+\Lambda_0(t))^{\gamma-\lambda\gamma})}$$

and the unconditional hazard function is

$$\lambda(t) = k\lambda_0(t)(\lambda+\Lambda_0(t))^{\gamma-1}$$

iv. Compound Poisson frailty model

The notation $cP(\gamma, k, \lambda)$ is used for a compound Poisson distribution.

The marginal survival and hazard function in case of a compound Poisson frailty model is given by:

$$S(t) = e^{\frac{-1-\gamma}{\gamma\sigma^2}((1+\frac{\sigma^2}{1-\gamma}\Lambda_0(t))^{\gamma}-1)}$$

and

$$\lambda(t) = \frac{\lambda_0(t)}{(1+\frac{\sigma^2}{1-\gamma}\Lambda_0(t))^{1-\gamma}}$$

v. Log-normal models

Two variants of the log-normal model exist. We assume a normally distributed random variable W to generate frailty as $Z = e^W$. The two variants of the model are given by the restrictions $\mathbf{E}W = 0$ and $\mathbf{E}Z = 1$, where the first one is much more popular in the literature. Unfortunately, no explicit form of the unconditional likelihood exists. Consequently, estimation strategies based on numerical integration in the maximum likelihood approach are required.

vi. Multivariate frailty models

A second important application of frailty models is in the field of multivariate survival data. Such kind of data occurs for example if lifetimes (or times of onset of a disease) of relatives (twins, parent-child) or recurrent events like infections in the same individual are considered. In such cases independence between the clustered survival times can not be assumed. Multivariate models are able to account for the presence of dependence between these event times. A commonly used and very general approach is to specify independence among observed data items conditional on a set of unobserved or latent variables (Hougaard, 2000). The dependence structure in the multivariate model arises from a latent variable in the conditional models for multiple observed survival times, for example let $S(t_1|Z, X_1)$ and $S(t_2|Z, X_2)$ be the conditional survival functions of two related individuals with different vectors of observed covariates X_1 and X_2 , respectively.

Averaging over an assumed distribution for the latent variables (e.g., using a gamma, log-normal, stable distribution) then induces a multivariate model for the observed data. In the case of paired observations, the two-dimensional survival function is of the form

$$S(t_1, t_2) = \int_0^\infty S(t_1|Z, X_1)S(t_2|Z, X_2)g(z)dz \tag{13}$$

where g denotes the density of the frailty Z . In the case of twins, $S(t_1, t_2)$ denotes the fraction of twins pairs with twin 1 surviving t_1 and twin 2 surviving t_2 . Frailty models for multivariate survival data are derived under conditional independence assumption by specifying latent variables that act multiplicatively on the baseline hazard.

IV. Semi-Parametric model.

.

In statistics, a semi parametric model is a model that has parametric and non-parametric components. A parametric model is one in which the indexing parameter is a finite-dimensional vector (in k -dimensional Euclidean space for some integer k; i.e. the set of possible values for θ is a subset of R^k, or $\Theta \subset R^k$ In this case we say thatθ is finite-dimensional. In non-parametric models, the set of possible values of the parameter θ is a subset of some space, not necessarily finite-dimensional. For example, we might consider the set of all distributions with mean 0. Such spaces are vector spaces with topological structure, but may not be finite-dimensional as vector spaces.

Thus,$\Theta \subset F$ for some possibly infinite-dimensional space F in semi parametric models, the parameter has both a finite-dimensional component and an infinite-dimensional component (often a real-valued function defined on the real line). Thus the parameter space Θ in a semi parametric model satisfies $\Theta \subset R^k XF$, where F is an infinite-dimensional space.

It may appear at first that semi parametric models include non-parametric models, since they have an infinite-dimensional as well as a finite-dimensional component. However, a semi parametric model is considered to be "smaller" than a completely non-parametric model because we are often interested only in the finite-dimensional component of θ . That is, we are not interested in estimating the infinite-dimensional component. In non-parametric models, by contrast, the primary interest is in estimating the infinite-dimensional parameter.

Thus the estimation task is statistically harder in nonparametric models.These models often use smoothing or kernels. A well-known example of a semi parametric model is the Cox proportional hazards model. If we are interested in studying the time T to an event such as death due to cancer

or failure of a light bulb, the Cox model specifies the following distribution function for T :

$$F(t) = 1 - exp(-\int_0^t \lambda_0(u)e^{\beta' x}du) \tag{14}$$

where x is the covariate vector, and β and$\lambda_0(u)$ are unknown parameters. $\theta = (\beta, \lambda_0(u))$. Hereβ is finite-dimensional and is of interest;$\lambda_0(u)$ is an unknown non-negative function of time (known as the baseline hazard function) and is often a nuisance parameter. The collection of possible candidates for $\lambda_0(u)$ is infinite-dimensional.

.

i. The Cox model

.

Let Y_i denote the observed time (either censoring time or event time) for subject i. Let C_i be the indicator that the time corresponds to an event (i.e. if $C_i = 1$ the event occurred and if$C_i = 0$ the time is a censoring time).

Let $X_i = \{X_{i1}, \ldots X_{ip}$ be the realized values of the covariates for subject i. The hazard function for the Cox proportional hazard model has the form

$$\lambda(t/X_i) = \lambda_0(t)exp(\beta_1 X_{i1} + \ldots + \beta_p X_{ip}) = \lambda_0(t)exp(X_i\beta) \tag{15}$$

This expression gives the hazard rate at time t for subject i with covariate vector (explanatory variables) X_i

Ignoring ties for the moment, conditioned upon the existence of a unique event at some particular time t the probability that the event occurs in the subject i for which $C_i = 1$ and$Y_i = t$ is

$$L_i(\beta) = \frac{\theta_i}{\sum_{j:Y_{j\geq Y_i}} \theta_j} \tag{16}$$

Where $\theta_j = exp(X_j.\beta)$. Observe that the factors of $\lambda_0(t)$ that would be present in both the numerator and denominator have canceled out.

Treating the subjects' events as if they were statistically independent, the joint probability of all realized events conditioned upon the existence of events at those times is the partial likelihood:

$$L(\beta) = \prod_{i:C_{i=1}} \frac{\theta_i}{\sum_{j:Y_{j\geq Y_i}} \theta_j} \tag{17}$$

The corresponding log partial likelihood is

$$l(\beta) = \sum_{i:C_{i=1}} \left(X_i\beta - log \sum_{j:Y_{j\geq Y_i}} \theta_j \right) \tag{18}$$

This function can be maximized over β to produce maximum partial likelihood estimates of the model parameters.

The partial score function is

$$l^{'}(\beta) = \sum_{i:C_{i=1}} \left(X_i - \frac{\sum_{j:Y_{j\geq Y_i}} \theta_j X_j}{\sum_{j:Y_{j\geq Y_i}} \theta_j} \right) \tag{19}$$

and the Hessian matrix of the partial log likelihood is

$$l^{''}(\beta) = -\sum_{i:C_{i=1}} \left(\frac{\sum_{j:Y_{j\geq Y_i}} \theta_j X_j X_j^{'}}{\sum_{j:Y_{j\geq Y_i}} \theta_j} - \frac{\left[\sum_{j:Y_{j\geq Y_i}} \theta_j X_j\right]\left[\sum_{j:Y_{j\geq Y_i}} \theta_j X_j^{'}\right]}{\left[\sum_{j:Y_{j\geq Y_i}} \theta_j\right]^2} \right) \tag{20}$$

Using this score function and Hessian matrix, the partial likelihood can be maximized using the Newton-Raphson algorithm.

The inverse of the Hessian matrix, evaluated at the estimate of β , can be used as an approximate variance-covariance matrix for the estimate, and used to produce approximate standard errors for the regression coefficients.

.

ii. Tied times.

.

Several approaches have been proposed to handle situations in which there are ties in the time data.Breslow's method describes the approach in which the procedure described above is used unmodified, even when ties are present.

An alternative approach that is considered to give better results is Efron's method.

Lett_j denote the unique times, letH_j denote the set of indices i such that $Y_i = t_j$ and $C_i = 1$, and let $m_j = |H_j|$

Efron's approach maximizes the following partial likelihood.

$$L(\beta) = \prod_j \frac{\prod_{i\epsilon H_j} \theta_i}{\prod_{l=0}^{m-1} \left[\sum_{j:Y_{j\geq t_j}} \theta_i - \frac{l}{m} \sum_{i\epsilon H_j} \theta_i\right]} \tag{21}$$

The corresponding log partial likelihood is

$$l(\beta) = \sum_j \left(\sum_{i\epsilon H_i} X_i \beta - \sum_{l=0}^{m-1} log \left(\sum_{i:Y_{i\geq t_j}} \theta_i - \frac{l}{m} \sum_{i\epsilon H_j} \theta_i \right) \right) \tag{22}$$

the score function is

$$l^{'}(\beta) = \sum \left(\sum_{i\epsilon H_i} X_i - \sum_{l=0}^{m-1} \frac{\sum_{i:Y_{i\geq t_j}} \theta_i X_i - \frac{l}{m} \sum_{i\epsilon H_j} \theta_i X_i}{\sum_{i:Y_{i\geq t_j}} \theta_i - \frac{l}{m} \sum_{i\epsilon H_j} \theta_i} \right) \tag{23}$$

$$_j$$

.

and the Hessian matrix is

$$l''(\beta) = -\sum_{j}\sum_{l=0}^{m-1}\left(\frac{\sum_{i:Y_{i\geq Y_j}}\theta_i X_i X_i' - \frac{l}{m}\sum_{i\epsilon H_j}\theta_i X_i X_i'}{\phi_{j,l,m}} - \frac{Z_{j,l,m}Z_{j,l,m}'}{\phi_{j,l,m}^2}\right) \tag{24}$$

Where

$$\phi_{j,l,m} = \sum_{i:Y_{i\geq t_j}}\theta_i - \frac{l}{m}\sum_{i\epsilon H_j}\theta_i \tag{25}$$

$$Z_{j,l,m} = \sum_{i:Y_{i\geq t_j}}\theta_i X_i - \frac{l}{m}\sum_{i\epsilon H_j}\theta_i X_i \tag{26}$$

Note that when H_j is empty (all observations with timet_j are censored), the summands in these expressions are treated as zero.

iii. Generation of the data

.

The function genfrail in the package frailtySurv can be used to generate survival times under a wide variety of conditions. The survival function at time t of the j^{th} observation of cluster i , when given time-independent covariate Z_{ij} and frailty variate ω_i, is given by

$$S_{ij}(t/Z_{ij},\omega_i) = exp\{-\Lambda_0(t)\omega_i e^{\beta^T Z_{ij}}\} \tag{27}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ is the unspecified cumulative baseline hazard function.

.

V. The Log-likelihood approach of the frailty models

.

For many applications, the natural logarithm of the likelihood function, called the log-likelihood, is more convenient to work with. Because the logarithm is a strictly increasing function, the logarithm of a function achieves its maximum value at the same points as the function itself, and hence the log-likelihood can be used in place of the likelihood in maximum likelihood estimation and related techniques. Finding the maximum of a function often involves taking the derivative of a function and solving for the parameter being maximized, and this is often easier when the function being maximized is a log-likelihood rather than the original likelihood function.

Likelihood function can be written as

$$L(\beta,\theta,\Lambda_0) = \prod_{i=1}^{n}\int\prod_{j}^{m_i}\{\lambda_{ij}(T_{ij}/Z_{ij},\omega)\}^{\delta_{ij}}S_{ij}(T_{ij}/Z_{ij},\omega)f(\omega)d\omega \tag{28}$$

$$= \prod_{i=1}^{n}\prod_{j=1}^{m_i}\{\lambda_0(T_{ij})e^{\beta^T Z_{ij}}\}^{\delta_{ij}}\prod_{i=1}^{n}L^{(N_i.(\tau))}\{H_i.(\tau)\} \tag{29}$$

.
where τ is the end of follow-up period, f is the frailty's density function,
$N_{ij}(t) = \delta_{ij} I(T_{ij} \leq t)$, $N_i.(t) = \sum_{j=1}^{m_i} N_{ij}(t)$, $H_{ij}(t) = \Lambda_0(T_{ij} \bigwedge t) e^{\beta^T Z_{ij}}$, and
$H_i.(t) = \sum_{j=1}^{m_i} H_{ij}(t), j = 1, \cdots, m_j, i = 1, \cdots, n.$
.

We note that the m^{th} derivative of the Laplace transform evaluated at $H_i.(\tau)$ equals $\int \omega^{N_i.(\tau)} exp\{-\omega H_i.(\tau)\} f(\omega) d\omega,$ i
$1, \cdots, n$..

The log-likelihood equals

$$\iota(\beta, \theta, \Lambda_0) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \delta_{ij} log\{\lambda_0(T_{ij}) e^{\beta^T Z_{ij}}\} + \sum_{i=1}^{n} log L^{\{N_i.(\tau)\}}\{H_i.(\tau)\} \tag{30}$$

.

i. Score Equations

. In some cases , we can solve the score equations instead of maximizing the log-likelihood .
The score function with respect to β can be given by ;

$$U_\beta = \frac{\partial}{\partial \beta} \ell(\beta, \theta, \Lambda_0) = \sum_{i=1}^{n} \{ \sum_{j=1}^{m_i} \delta_{ij} Z_{ij} + \frac{\frac{\partial}{\partial \beta} H_i.(\tau) \frac{\partial}{\partial H_i.(\tau)} L^{N_i.(\tau)}(H_i(\tau))}{L^{N_i.(\tau)}(H_i.(\tau))} \} \tag{31}$$

$$= \sum_{i=1}^{n} \{ \sum_{j=1}^{m_i} \delta_{ij} Z_{ij} + \sum_{j=1}^{m_i} H_{ij}(T_{ij}) Z_{ij} \frac{L^{N_i.(\tau)+1}(H_i(\tau))}{L^{N_i.(\tau)}(H_i.(\tau))} \} \tag{32}$$

Note that

$$L^{(N_i.(\tau)+1)} H_i.(\tau) / L^{(N_i.(\tau))} H_i.(\tau) \tag{33}$$

corresponds to ψ_i in Gorfine *et al.* (2006).

The score function with respect to θ is given by

$$U_\theta = \frac{\partial}{\partial \theta} \ell(\beta, \theta, \Lambda_0) = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial \theta} L^{N_i.(\tau)}(H_i(\tau))}{L^{N_i.(\tau)}(H_i.(\tau))} \} \tag{34}$$

.
The score equations are given by $U(\beta, \theta, \Lambda_0) = (U_\beta, U_\theta) = 0$ and the estimator of $\gamma = (\beta^T, \theta)$ is defined as the value of β^T, θ that solves the score equations for any given Λ_0 .

VI. Major Frailty distributions

. The frailty distributions have the support $\omega \epsilon (0, \infty)$ The gamma and Power variance function have a closed-form analytic expression for the Laplace transform , but the log-normal and inverse

Gaussian Laplace transforms must be evaluated numerically . For the gamma , log-normal , and inverse Gaussian , there is a positive relationship between the distribution parameter θ and the strength of dependence between cluster members. As θ increases , intra-cluster failure -times dependency increases. The opposite is true for the PVF(Power variance function), and as θ increases , the dependence between failure-times of the cluster's members decreases.

i. Log-normal

.
The log-normal distribution is denoted by $LN(\theta)$ and with density function

$$f(\omega);\theta) = \frac{1}{\omega\sqrt{\theta 2\pi}} exp\{\frac{-(ln\omega)^2}{2\theta}\} \tag{35}$$

so the mean and variance are $exp(\theta/2)$ and $exp(2\theta) - exp(\theta)$, respectively.
The Laplace transform and its derivatives equal

$$L^m(s) = \int_0^\infty (-\omega)^m e^{-s\omega} f(\omega;\theta) d\omega; m = 0, 1, 2, \cdots \tag{36}$$

Similar to the gamma distribution, the special case of $\theta = 0$ implies that $\omega \equiv 1$
The densit's partial derivative with respect to θ is given by

$$\frac{\partial}{\partial\theta} f(\omega;\theta) = \frac{ln^2(\omega) exp(\frac{-ln^2\omega}{2\theta})}{2\sqrt{2\pi}\theta^{5/2}\omega} - \frac{exp(\frac{-ln^2\omega}{2\theta})}{2\sqrt{2\pi}\theta^{3/2}\omega} \tag{37}$$

.

ii. Inverse Gaussian

.

The inverse Gaussian distribution is denoted by $IG(\theta)$, with mean 1 and variance θ. .

The density is given by

$$f(\omega;\theta) = (2\pi\theta\omega^3)^{-1/2} exp\{\frac{-(\omega-1)^2}{2\theta\omega}\} \tag{38}$$

where $\theta > 0$.
The Laplace transform and its derivatives equal

$$L^m(s) = \int_0^\infty (-\omega)^m e^{-s\omega} f(\omega;\theta) d\omega; m = 1, 2, \cdots \tag{39}$$

Similar to the gamma and log-normal , $\omega \equiv 1$ when $\theta = 0$
The partial derivative of the density function with respect to θ is given by

$$\frac{\partial}{\partial\theta} f(\omega;\theta) = \frac{(\omega-1)^2 exp\{-\frac{(\omega-1)^2}{2\theta\omega}\}}{2\sqrt{2\pi}\theta^2\omega\sqrt{\theta\omega^3}} - \frac{\omega^3 exp\{-\frac{(\omega-1)^2}{2\theta\omega}\}}{2\sqrt{2\pi}(\theta\omega^3)^{3/2}} \tag{40}$$

.

iii. Power Variance Function

.

The power variance function distribution is denoted by $PVF(\theta, \delta, \mu)$ and with density

$$f(\omega;\theta,\delta,\mu) = exp(-\mu\omega + \frac{\delta^\theta}{\theta})\frac{1}{\pi}\sum_{k=1}^{\infty}\frac{\Gamma(k\theta+1)}{k!}(-\frac{1}{\omega})^{\theta k+1}sin(\theta k\pi) \tag{41}$$

where $0 < \theta \leq 1$, $\mu \geq 0$, $\delta > 0$
To avoid identifiability problems , we let $\delta = \mu = 1$ as in Hanagal (2009), and get a one-parameter PVF density

$$f(\omega;\theta) = exp(-\omega + \theta^{-1})\frac{1}{\pi}\sum_{k=1}^{\infty}\frac{\Gamma(k\theta+1)}{k!}(-\frac{1}{\omega})^{\theta k+1}sin(\theta k\pi) \tag{42}$$

When $\theta = 1$, the degenerate distribution with $\omega \equiv 1$ is obtained.
PVF has expectation 1 and variance $1 - \theta$.

The Laplace transform is given by

$$L(s) = exp[-\{(1+s)^\theta - 1\}/\theta] \tag{43}$$

The Laplace transform derivatives are given by .

$$L^m(s) = (-1)^m L(s)\sum_{j=1}^{m} c_{m,j}(\theta)(1+s)^{j\theta-m}; m = 1, 2, \cdots \tag{44}$$

with coefficients
$c_{m,m}(\theta) = 0.$

$c_{m,1}(\theta) = \dfrac{\Gamma(m-\theta)}{\Gamma(1-\theta)}.$

$c_{m,j}(\theta) = c_{m-1,j-1}(\theta) + c_{m-1,j}(\theta)\{(m-1) - j\theta\}.$

The partial derivative of the Laplace transform with respect to θ are given by

$$\frac{\partial}{\partial\theta}L^m(s) = \frac{\partial}{\partial\theta}[(-1)^m L(s)\sum_{j=1}^{m} c_{m,j}(\theta)(1+s)^{j\theta-m}] \tag{45}$$

$$= (-1)^m\{\frac{\partial}{\partial\theta}L(s)\}\sum_{j=1}^{m}c_{m,j}(\theta)(1+s)^{j\theta-m} + (-1)^m L(s)\sum_{j=1}^{m}\{\frac{\partial}{\partial\theta}c_{m,j}(\theta)(1+s)^{j\theta-m} + c_{m,j}(\theta)j(1+s)^{j\theta-m}ln(1+s)\} \tag{46}$$

where

$$\frac{\partial}{\partial\theta}L(s) = exp\{\frac{1-()s+1)^{\theta}}{\theta}\}\{-\frac{-1-(s+1)^{\theta}}{\theta^2} - \frac{(s+1)^{\theta}log(s+1)}{\theta}\} \tag{47}$$

and the partial derivatives of the coefficients are

$\frac{\partial}{\partial\theta}c_{m,m}(\theta) = 0.$

$\frac{\partial}{\partial\theta}c_{m,1}(\theta) = \frac{\Gamma(m-\theta)\{\psi^{(0)}(1-\theta) - \psi^0(m-\theta)\}}{\Gamma(1-\theta)}.$

$\frac{\partial}{\partial\theta}c_{m,j}(\theta) = \frac{\partial}{\partial\theta}c_{m-1,j-1}(\theta) + \frac{\partial}{\partial\theta}c_{m-1,j}(\theta)\{(m-1) - j\theta\} - jc_{m-1,j}(\theta)$

.

iv. Expectation-maximization (EM) Algorithm

Expectation-maximization (EM) is an iterative method used to find maximum likelihood estimates of parameters in probabilistic models, where the model depends on unobserved, also called latent, variables. EM alternates between performing an expectation (E) step, which computes an expectation of the likelihood by including the latent variables as if they were observed, and a maximization (M) step, which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found in the E step. The parameters found on the M step are then used to start another E step, and the process is repeated until some criterion is satisfied. EM is frequently used for data clustering like for example in Gaussian mixtures or in the Baum-Welch training of a Hidden Markov Model.

v. Model Selection

There are different methods for selecting the most appropriate model in statistical analysis. The most commonly used methods include information and likelihood based criteria. To compare the different frailty models used in the study and the corresponding baseline hazard functions , the information based criteria is applied . The most commonly used model selection criteria are the Akaike information criterion (AIC) and Bayesian information criterion (BIC). AIC is given by the expression,

$$AIC = -2log(L) + 2k \tag{48}$$

where L is the maximized likelihood value and k is the number of parameters in the model. BIC is given by the expression

$$BIC = -2log(L) + kln(N) \tag{49}$$

,

where N is the total sample size.

The model with the smallest AIC value is considered a better fit.

RESULTS

	coef	se(coef)	Chisq	p
age	0.00318	0.0111	0.0814	0.775
sex	-1.48	0.358	0.171	0.000035
diseaseGN	0.088	0.406	0.0468	0.829
diseaseAN	0.351	0.400	0.770	0.023
diseasePKD	-1.43	0.631	5.14	0.023
frailty(id)			0.0000271	0.933

Table 1: *Table Showing frailty results*

	coef	se(coef)	Chisq	p
age	0.005	0.02	0.107	0.744
sex	-1.697	0.461	13.56	0.00023
diseaseGN	0.18	0.545	0.109	0.741
diseaseAN	0.393	0.545	0.770	0.471
diseasePKD	-1.14	0.825	0.52	0.19
frailty(id, dist = "gauss			12.1	

Table 2: *Table of frailty results with Gaussian*

AIC	gamma	ingau	possta	lognor
exponential	674	676	682	675
weibull	674	677	682	676
inweibull	692	692	692	692
loglogistic	685	685	686	685
lognormal	679	679	681	679
logskewnormal	681	681	682	681

Table 3: *Table comparing AIC of the models used.*

BIC	gamma	ingau	possta	lognor
exponential	684	685	692	685
weibull	686	688	694	687
inweibull	703	702	703	702
loglogistic	697	697	697	696
lognormal	691	691	692	691
logskewnormal	695	695	696	695

Table 4: *Table of the BIC of the models used.*

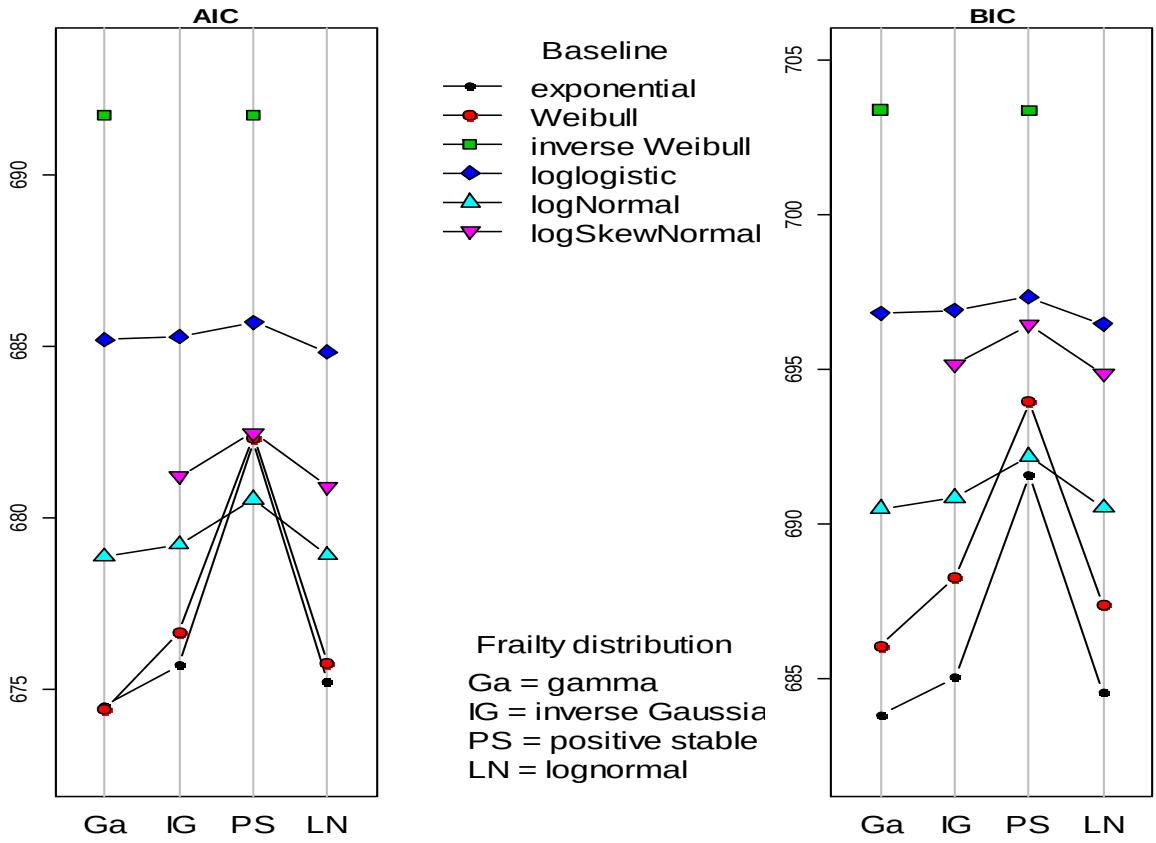
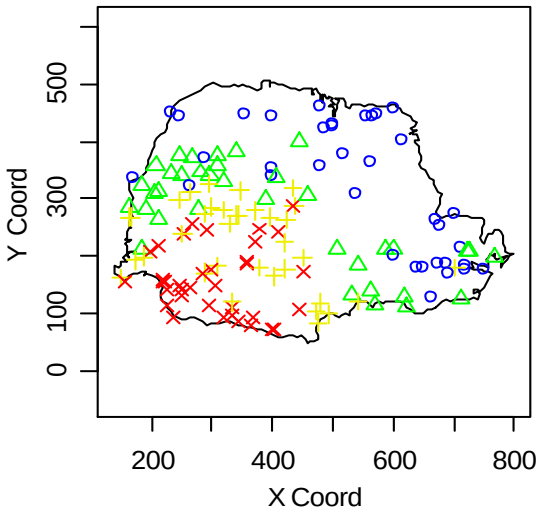
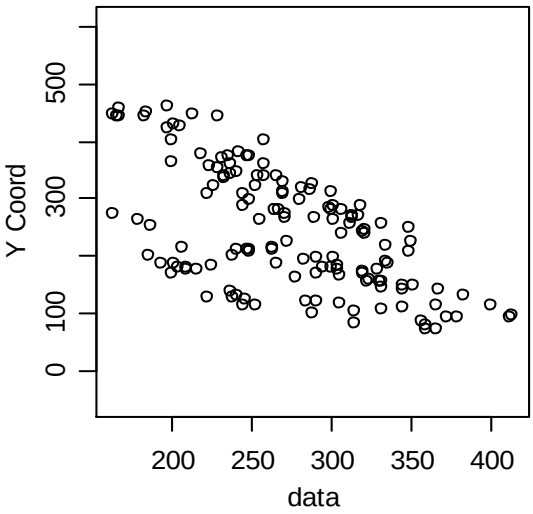
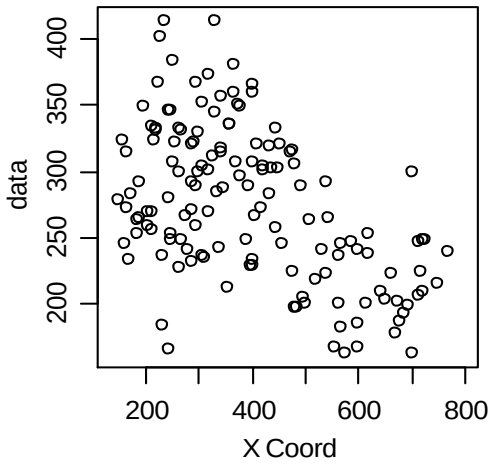
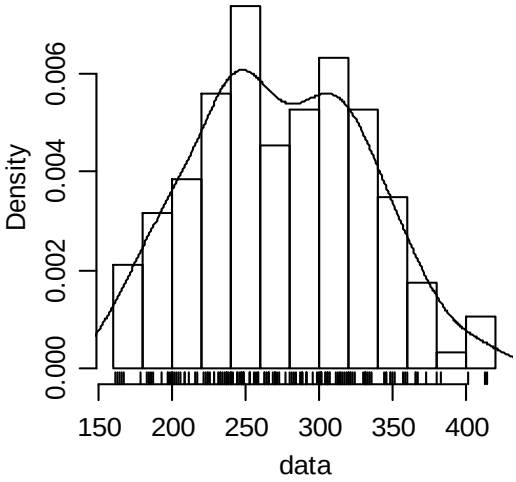


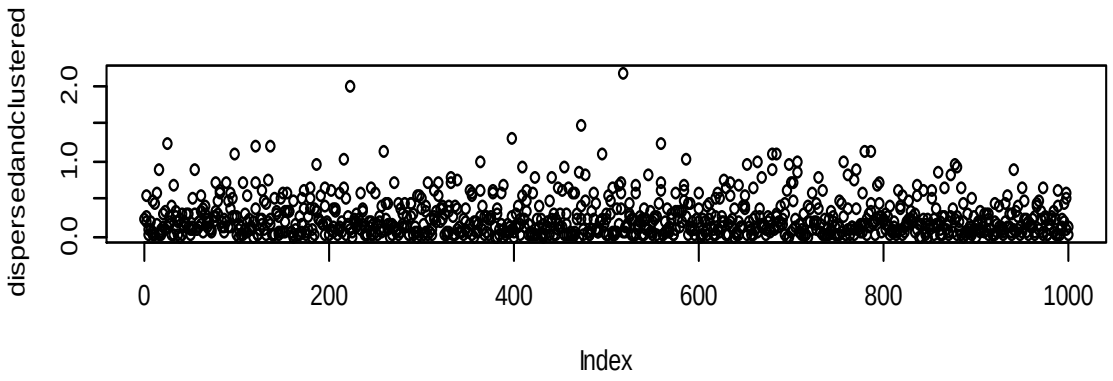
Figure 1: *Plots of AIC and BIC values*

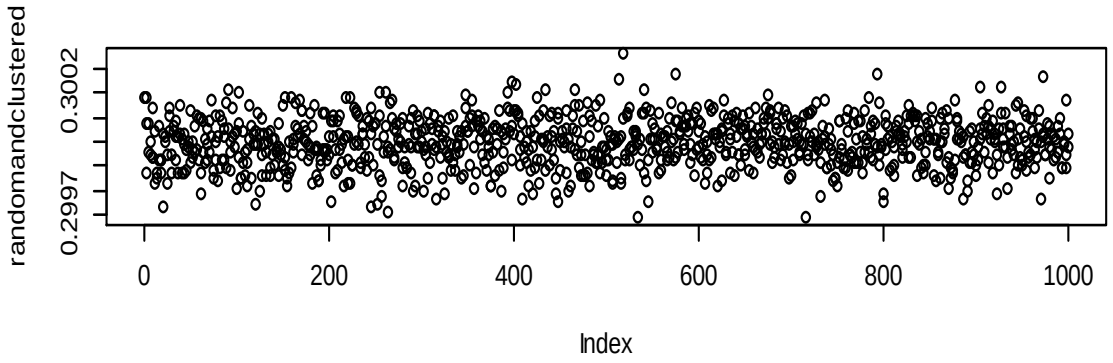

Y Coord
0
100
300
500
200
400
600
800
X Coord

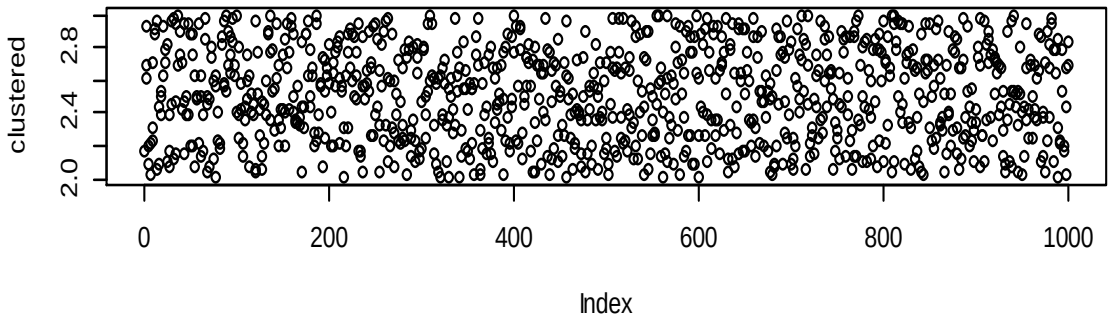

Y Coord
0
100
300
500
200
250
300
350
400
data


data
200
250
300
350
400
200
400
600
800
X Coord


Density
0.000
0.002
0.004
0.006
150
200
250
300
350
400
data


dispersedandclustered
0.0
1.0
2.0
0
200
400
600
800
1000
Index


randomandclustered
0.2997
0.3002
0
200
400
600
800
1000
Index


clustered
2.0
2.4
2.8
0
200
400
600
800
1000
Index

	ESTIMATE	SE	p-val
theta	0.301	0.156	
lambda	0.025	0.014	
sex	-1.485	0.396	<.001 ***
age	0.005	0.011	0.657

Table 5: *Table of Gamma frailty and Exponential as the baseline hazard distribution.*

Frailty distribution: Gamma
Baseline hazard distribution: Exponential
Loglikelihood: -333.248

Kendall's Tau: 0.131

	ESTIMATE	SE	p-val
theta	0.342	0.197	
lambda	0.020	0.011	
sex	-1.356	0.382	<=001 ***
age	0.005	0.011	0.679

Table 6: *Table showing results of Lognormal frailty and Exponential baseline hazard distribution*

Frailty distribution: Lognormal
Baseline hazard distribution: Exponential
Loglikelihood: -333.606

VII. Discussion

i. Model Section Result.

- The best combination of the model to be used is when the frailty distribution is gamma and the baseline hazard distribution is exponential since it gives the minimum AIC value of 674 .
- We can also make use of the frailty distribution as gamma while the baseline hazard ditribution is weibull since it is also giving us a minimum value of AIC of 674.
- However from the BIC results , we can make use of the frailty distribution as gamma and the baseline hazard distribution as exponential since it gives a lower value of BIC as 684 as compared to the one of Weibull distribution as the baseline hazard fuction.
- Standard errors are computed as the square roots of the diagonal elements of the observed information matrix. According to this model, sex has a significant impact on the hazard of infection while it is not affected by age. Conditional on the patient's frailty and on the age, the hazard of infection for a female at any time tis estimated to be exp(-1.485) =0.227 times that of a male, with Wald confdence interval

ii. Choice of Frailty model.

Importance of quality control is well known as one of the areas in statistical analysis hence the need need to deal with missing data and the misspecification of frailty random effects in a given study. The choice of the correct frailty model remains a main area of concern hence the need of this study which will give comprehensive results and the required policies to the practitioners on how to deal with missing data effects and on the misspecification of the frailty random effects.

References

Duchateau L., Janssen P. (2008). The frailty model. Springer. New York: Springer-Verlag.

Munda M, Rotolo F, Legrand C (2012). parfm: Parametric Frailty Models in R. Journal of Statistical Software, 51(11), 1-20.

Wienke A (2010). Frailty Models in Survival Analysis. Chapman and Hall/CRC biostatistics series.

Glidden D, Vittinghoff E (2004). Modelling Clustered Survival Data From Multicentre Clinical Trials. Statistics in medicine, 23(3), 369-388.

Kendall, MG. (1938) A new measure of rank correlation. Biometrika 30(1/2), 81-93.

Kadane ,J. and Lazar, N.(2001) Methods and Criteria for Model Selection. Technical Report,759,Carnegie Mellon University.

Garg, M. B. Rao and C. Redmond 1970. Maximum likelihood estimation of the parameters. Journal of the Royal Statistics Society.

E.J. Bedrick and C.L. Tsai, Model Selection for Multivariate Regression in Small Samples,Biometrics, 50 (1994), 226-231.

Johnson, Norman L.; Kotz, Samuel; Balakrishnan, N. (1994), Continuous univariate distributions. Vol. 1, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics (2nd ed.), New York.

Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasilikelihood estimation in models with surrogate predictor. J. Amer. Statist. Assoc. 85, 652-663.