International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 4 Issue 4 || April. 2016 || PP-4101-05

Gamma and inverse Gaussian frailty models: A comparative study

Samia A. Adham, Amani A. AlAhmadi
(Department of Statistics/ Faculty of Science/ King Abdulaziz University/ Saudi Arabia)

ABSTRACT:*Frailty models have become very popular during the last three decades and their applications are numerous. The main goal of this manuscript is to compare two frailty models (gamma frailty model and inverse Gaussian frailty model) each of which has a log-logistic distribution to be its baseline hazard function. A real data set is applied for the two considered frailty models in order to deal with models comparison. It has been concluded that the gamma frailty model is the best model fits this data set. Then the inverse Gaussian frailty model, which provides a better fit of the considered data set than the Cox's model.*

KEYWORDS–*Proportional hazards; Heterogeneity;Shared Frailty Models; Survival Analysis.*

I. INTRODUCTION

In applications of survival analysis, usually only a few covariates such as age, sex, severity of disease or laboratory data are known. It is known that there are many other factors that can influence survival, including health status, life style, smoking, occupation and genetic risk factors. In many applications, the population under study cannot be assumed to be homogeneous but must be considered as a heterogeneous. A popular regression model for the analysis of survival data is the Cox proportional hazards regression model. It allows testing for differences in survival times of two or more groups.

The frailty approach is a statistical modeling concept which aims to account for heterogeneity, caused by unmeasured covariates. The frailty model is a random effect model, where the random effect (the frailty) has a multiplicative effect on the baseline hazard function. This random effect explains the dependence in the frailty models. The term frailty was first suggested by [1] in the context of mortality studies. [2]suggested a random effects model in order to account for the unobserved heterogeneity due to unobserved covariates and introduced the model to the literature of economics and the model is called the mixed proportional hazards model. [3][4] and [5] considered distributions for the frailty model to find the best model. [6]used frailty model to explain the deviant behavior of mortality rates at advanced ages.

There are many applications of the gamma frailty model. [7]studied the expulsion of intrauterine contraceptive devices. [8]studied recidivism among criminals using gamma-Weibull model. [9]used the gamma frailty model to check the proportional hazards assumptions in his study of malignant melanoma. A formal of the goodness-of-fit tests for the gamma frailties was constructed by [10]. They also construct a new class of frailty models that extend the gamma frailty model by using certain polynomial expansions that are orthogonal with respect to the gamma density. For that extended family, they obtained an explicit expression for the marginal likelihood of the data. The order selection test is based on finding the best fitting model in such a series of expanded models. A bootstrap was used to obtain p-values for the tests. Simulations and data examples illustrated the test's performance. [11]considered gamma distribution as frailty distribution and the log-logistic distribution as baseline distribution for bivariate survival times. Because this distribution has the advantage of having simple algebraic expressions for its survivor and hazard functions and a closed form for its distribution function. [12]studied the case of severe acute malnutrition (SAM) in developing countries. Then, they used exponential, Weibull and log-logistic as baseline hazard functions and the gamma as well as inverse Gaussian for the frailty distributions and then based on AIC criteria, all models were compared for their performance.

In this manuscript the analysis of the right censored survival data are considered. A real data example is applied for illustration. Section (2) concerns with maximum likelihood approach to the shared frailty models. Reconstitution data set: Reconstitution of blood–milk barrier after mastitis is presented in (3), in details. Finally, Section (4) discusses some important conclusions.

II. LIKELIHOOD APPROACH TO SHARED FRAILTY MODELS

In this section, the shared frailty model, which we consider in this manuscript, is explained. Then the likelihood function according for the right censored data is presented. gamma frailty and the inverse Gaussian frailty modelsSuppose that we have a data set of n individuals from some population and $i = 1,2, \dots, C$ subgroups or clusters. Each subgroup consists of $n_i \geq 1$ individuals. The individuals in

each subgroup have dependent event times due to unobserved frailty u_i. This frailty term may represent aggregate effect of common genes or shared environmental effect on survival of members of a given family, such as siblings, husband and wives. The goal is to estimate the frailty variance, θ. The variance of the frailty distribution is used to determine the degree of heterogeneity in the study population. The frailty model is given as

$$h_{ij}(t|u,\beta,Z) = h_o(t)u_i e^{\beta Z_{ij}}, u > 0. (\beta, Z \in R) \quad (1)$$

where $h_o(t)$ is a common baseline hazard function, β is a vector of unknown regression coefficients and, for $i = 1,2,\dots,C$ and $j = 1,2,$ Z_{ij} is a vector of the observable covariates. The frailties u_i are unobserved (random) common risk factor shared by all subjects in cluster i assumed to be identically and independently distributed random variables with a common density function $f(U,\theta)$, where θ is the parameter of the frailty distribution. The value of the frailty u_i is common to all individuals in the cluster. In the literature, different frailty distributions have been proposed, such as gamma distribution, inverse Gaussian distribution, positive stable distribution, power variance function distribution, compound Poisson distribution and lognormal distribution. A more detailed presentation of the shared frailty models can be found in [13].

The likelihood function for right censored survival data is given by

$$L = \prod_{j=1}^{n} \left[\left(1 - G_j(t)\right) f_j(t)\right]^{\delta_j} \left[\left(1 - F_j(t)\right) g_j(t)\right]^{1-\delta_j}, \quad (2)$$

where δ_j is the censoring indicator, g and G are the density function and the cumulative distribution function of the censoring time, respectively; and f and F are, respectively, the density function and the cumulative distribution function of the event time.

The distribution of censoring times in the likelihood function can be ignored because it does not depend on the parameters of interest related to the survival function. Therefore, assuming right censoring, the likelihood function given by (2) can be rewritten as

$$L = \prod_{j=1}^{n} (f_j(t))^{\delta_j} (S_j(t))^{1-\delta_j}, \quad (3)$$

where $S_j(t) = 1 - F_j(t)$ is the survival function of the event time. Considering the shared frailty model presented above, the likelihood function for the j^{th} subject in the i^{th} subgroup is given by

$$L_i = \prod_{j=1}^{n_i} \left(f_{ij}(t)\right)^{\delta_{ij}} (S_{ij}(t))^{1-\delta_{ij}}. \quad (4)$$

Since $h_{ij}(t) = \frac{f_{ij}(t)}{S_{ij}(t)}$, then the likelihood function in (4) reduces to

$$L_i = \prod_{j=1}^{n_i} \left(h_{ij}(t)\right)^{\delta_{ij}} S_{ij}(t). \quad (5)$$

The conditional likelihood function for the i^{th} subgroup is then given by

$$L_i(\psi,\beta|u_i) = \prod_{j=1}^{n_i} (h_o(t)u_i e^{\beta Z_{ij}})^{\delta_{ij}} e^{H_o(t)u_i e^{\beta Z_{ij}}},$$

where, ψ is a vector of parameters of the baseline hazard function. It follows that, the marginal likelihood function for the i^{th} subgroup is

$$L_i(\psi,\theta,\beta) = \prod_{j=1}^{n_i} \int_0^{\infty} (h_o(t)u_i e^{\beta Z_{ij}})^{\delta_{ij}} e^{H_o(t)u_i e^{\beta Z_{ij}}} g_k(u_i) du$$

where $g_k(u_i)$ is the probability density function of the frailty $u_i, k = 1,2,$ $i = 1,2,\dots,C.$ For $k = 1,2,$ the probability density functions of frailties in the gamma frailty model and the inverse Gaussian frailty model are, respectively, given by

$g_1(u_i) = \frac{u_i^{\frac{1}{\theta}-1}\exp(-\frac{u_i}{\theta})}{\Gamma(\frac{1}{\theta})\theta^{\frac{1}{\theta}}}$ and $g_2(u_i) = \left[\frac{1}{2\pi\theta}\right]^{\frac{1}{2}} u_i^{-\frac{3}{2}} e^{-\frac{(u_i-1)^2}{2u_i\theta}}, (u_i, \theta > 0).$

Where $u_i > 1$indicates that individuals in group$i$ are frail, whereas $u_i < 1$ indicates that individuals are strong and have lower risk. The log-logistic hazard function and cumulative hazards functions with parameters $\psi = (\alpha, \kappa)$are, respectively, given by

$h_o(t) = \frac{\exp(\alpha)\kappa t^{\kappa-1}}{1+\exp(\alpha)t^{\kappa}}$and $H_o(t) = ln\left[\frac{\exp(\alpha)\kappa t^{\kappa-1}}{1+\exp(\alpha)t^{\kappa}}\right].$

Hence, the marginal log likelihood function for the gamma frailty with the log-logistic baseline hazard function is

$$l(\psi,\theta,\beta) = \sum_{i=1}^{G}\left[d_i\log(\theta) - \log\left(\Gamma\left(\frac{1}{\theta}\right)\right) + \log\left(\Gamma\left(\frac{1}{\theta}+d_i\right)\right) - \left(\frac{1}{\theta}+d_i\right)\log\left(1+\theta\sum_{j=1}^{n_i}\ln[1+\exp(\alpha)t^{\kappa}]\ e^{\beta Z_{ij}}\right) + \sum_{j=1}^{n_i}\delta_{ij}\left(\beta Z_{ij} + \log\left(\frac{\exp(\alpha)\kappa t^{\kappa-1}}{1+\exp(\alpha)t^{\kappa}}\right)\right)\right]$$

And the marginal log likelihood function for the inverse Gaussian frailty model with the log-logistic baseline hazard function is

$$l(\psi,\theta,\beta) = \sum_{i=1}^{G}\left[\delta_{ij}\log\left(\frac{\exp(\alpha)\,\kappa t^{\kappa-1}}{1+\exp(\alpha)\,t^{\kappa}}\right) - \frac{1}{2}\log\left(2\sum_{j=1}^{n_i}\ln[1+\exp(\alpha)\,t^{\kappa}] + \theta\right) + \frac{1}{2}\log(\theta) - \theta\left(2\sum_{j=1}^{n_i}\ln[1+\exp(\alpha)\,t^{\kappa}] + \theta\right)^{\frac{1}{2}} + \left(2\sum_{j=1}^{n_i}\ln[1+\exp(\alpha)\,t^{\kappa}] + \theta\right)^{\frac{1}{2}}\right]$$

By maximizing the log likelihood function for each of the two frailty models proposed above, one can obtain the maximum likelihood estimates for the parameters$\psi, \theta$and β , see [13]and [14].

III. NUMERICAL ILLUSTRATION

In this sectionwe applied a real data set to compare the Cox proportional hazards model with its extensions, the gamma frailty model and the inverse Gaussian frailty model, each of which has the log-logistic distribution as a baseline hazard function. The comparisons between the Cox proportional hazard model and the considered gamma and inverse Gaussian frailty models are based on the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC). The effect of the considered frailty on the coefficients of the treatment effects is examined. The real right censored data set, called Reconstitution data set: Reconstitution of blood–milk barrier after mastitis, is applied for the required comparisons. (Source of the data set: http://www.vetstat.ugent.be/research/frailty/datasets/)

In order to perform the required computations, the statistical package R is used. This study assumes two covariates Drug and heifer with coefficients $\beta_1$and β_2, respectively. However, the two treatment times (active compound and placebo) for each cow is assumed independent.Then, the frailty modelcan be written as:

$h_{ij}(t) = h_{\circ}(t)u_i \exp(\beta_1 drug + \beta_2 heifers).$

Hence,thegamma and the inverse Gaussian frailty models with the log-logistic hazard function are applied to this data setin order to compare the effects of these two frailty models. In addition, the Cox's model is also applied to this data set to compare it with the considered two frailty models.

The maximum likelihood estimate of the parameters of the gamma and the inverse Gaussian frailty models and the Cox's modelare computed. Furthermore, the comparison of the gamma and the invers Gaussian frailty modelsare considered according to the AIC and BIC.

Table 1: Parameters estimates

parameter	Gamma	Inverse Gaussian	Cox
$\theta\ (SE)$	0.307 (0.151)	0.347(0.229)	_____
$\beta_1\ (SE)$(p-value)	0.453(0.172) (0.009)	0.454 (0.173) (0.009)	0.223(0.165)
$\beta_2(SE)$(p-value)	0.378(0.231) (0.102)	0.336 (0.210) (0.109)	0.340(0.145)
AIC	729.882	729.907	738.068
BIC	743.075	743.1	744.481

Table (1) provides the maximum likelihood estimates of the parameter θ and the regression parametersβ_1 and β_2 of the gamma frailtyand inverse Gaussian frailty models the with log-logistic baseline hazard function. In the case of not including frailty (Cox's model), it is clear that the regression coefficients (β_1, β_2) of the effect of the two covariate heifer and Drug are biased down. Whereas, for the gamma and the invers Gaussian frailty models, the regression estimates and their standard errors (SE) increase, which is predictable because the frailty variable u_i is included in the model. The p-value of theregression coefficients $(\beta_1, \beta_2)$of drug and heifer are 0.009 and 0.102, respectively; which indicates that $\beta_1$is significant while β_2 is not significant for the gamma frailty and the same results are founded for the invers Gaussian frailty.

It is clear that from Table (1) the parameter θ of the gamma frailty is less than the parameter θofthe invers Gaussian frailty, which indicates that including gamma frailty gives a better fit than the invers Gaussian frailty. The estimate of the variance of frailty term θ equal 0.307 and 0.347 for the gamma and the invers Gaussian frailties, respectively. The p-value ofθ equal 0.005 and 0.004, respectively, for the gamma and theinvers Gaussian frailty models. That means that the heterogeneity parameter θ is significant in the two considered frailty models.

The AIC and BIC value are computed for the gamma frailty and the inverse Gaussian frailty models with the log-logistic baseline hazard function and for the Cox's model. The smallest AIC and BIC values suggest the model that gives better fit for the data than other models. One can see from Table (1), that the gamma frailty model gives the best fit to this data set then the invers Gaussian frailty model is better than Cox's model.

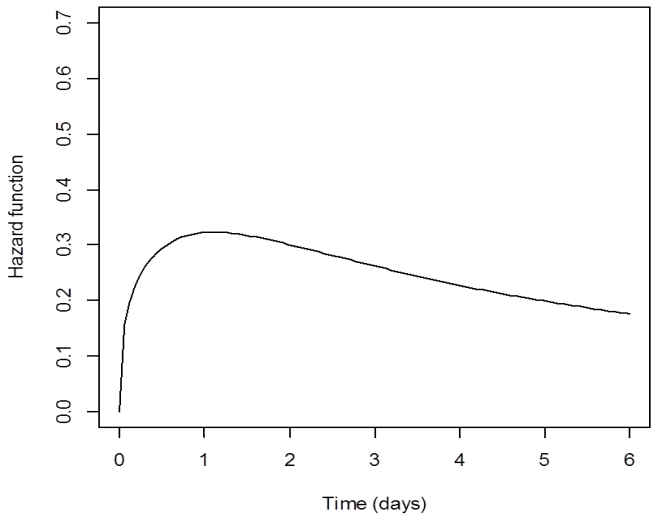

Figure 1: The hazard functions for invers Gaussian frailty and gammafrailty with thelog-logistic baseline hazard function.

The hazard functions for bothgamma and inverse Gaussian frailty models are increasing and then decreasing and this is expected. It is clear from Fig.(1) that the two curves of the hazard functions of the gamma frailty model and the invers Gaussian frailty model are compatible.

The estimated values of the parameters of the log-logistic baseline hazard functions are given in Table (2).

Table 2: The value of the parameter estimate of the log-logistic baseline hazard function

parameter	gamma frailty	invers Gaussian frailty
α	-1.293	-1.152
κ	1.349	1.367

For the log-logistic baseline hazard function, it is known that,thenegative sign of the parameterα indicates that the hazard function is increasing and then decreasing.

IV. CONCLUSIONS

This study compares the gamma and the inverse Gaussian frailty models when assuming the log-logistic distribution as their baseline hazard function. The maximum likelihood estimation method is considered to estimate the parameters of the considered models in order to compare them through estimation and testing the significance of the parameters of the models under consideration. A real data set called Reconstitutiondata set is applied to compare the two frailty models. The AIC and BIC were computed to assess the considered frailty models which of them gives the best fit to this data set. It has been found that the gamma frailty model is the best model that fits this data set among the other two models. Then, the invers Gaussian frailty model fits the databetter than the Cox's model. Furthermore, it has been found that, the heterogeneity parameter θ is significant in both gamma and inverse Gaussian frailty models.

REFERENCES

[1] J.W. Vaupel,K.G. Manton, and E. Stallard, The impact of heterogeneity in individual frailty on the dynamics of mortality,*Demography16,*1979, 439-454.

[2] T. Lancaster, *Analysis of Transition Data* (University of Cambridge, New work, 1990).

[3] P. Hougaard, Life Table Methods for Heterogeneous Populations: Distributions Describing Heterogeneity.*Biometrika,71,* 1984, 75-83.

[4] P. Hougaard, Survival models for heterogeneous populations derived from stable distributions. *Biometrika73,* 1986 a, 387-396.

[5] P. Hougaard, A class of multivariate failure time distributions. *Biometrika73,* 1986 b,671-678.

[6] J.W. Vaupel, and A.I. Yashin, Heterogeneity's ruses: some surprising effects of selection on population dynamics. *American Statistician,39,* 1985,176-185.

[7] O.O. Aalen, Two examples of modelling heterogeneity in survival analysis. *ScandinavianJournalofStatistics14,* 1987, 19-25.

[8] R. Ellermann, S. Pasquale, and J. M. Tien,An Alternative Approach to Modeling Recidivism Using Quantile Residual Life Functions. *OperationsResearch*40, 1992, 485-504.

[9] P.K. Andersen, O. Borgan, R.D. Gill, and N. Keiding, *Statistical models based on counting processes* (Springer Verlag, New York, 1993).

[10] C. Geerdens,G. Claeskens, and P. Janssen, Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty, *Biostatistics*, 2012, 1-14.

[11] D.D. Hanagal, and R. Sharma, Modeling Heterogeneity for Bivariate Survival Data by Shared Gamma Frailty Regression Model, *Model Assisted Statistics and Applications, 8,* 2013, 85-102.

[12] Banbeta.A , D. Seyoum, T. Belachew, B. Birlie, and Y. Getachew, Modeling time-to-cure from severe acute malnutrition application of various parametric frailty models. *Archives of Public Health, 73(6),* 2015. Doi: *10.1186/2049-3258-73-6.*

[13] LDuchateau, and P. Janssen,*The Frailty Model.* (Springer-Verlag, New York, 2008).

[14] U.A. Abdulkarimova, *Frailty Models for Modelling Heterogeneity*, McMaster University, MSc, 2013.