Simulation study for misspecifications on a frailty model

Andréa Ferreira and Nancy L. Garcia *

Universidade Estadual de Campinas

Abstract. Nielsen, G.G.; Gill, R.D.; Andersen, P.K.; Sorensen, T.I.A. (1992) propose a consistent and asymptotically normal estimator for the variance of the frailty distribution under gamma assumption. A simulation study shows that this estimator is asymptotically biased for log-normal and absolute value of normal frailty distributions.

Key words: Frailty models, simulation study.

1 Introduction

In the last two decades there was a big leap in terms of survival analysis methodology. This type of analysis has been applied not only in the traditional areas of biology and engineering, but also in demography, social sciences and economy. One of the biggest problems in survival analysis is related to the presence of populational heterogeneity. A very well-known method of analysis is the method of partial likelihood based on the Cox's proportional hazard model (Cox, 1972). In this model it is assumed that the heterogeneity can be measured through the observable covariates and they were all included in the model. It is a semi-parametric model since it considers the hazard function to be unknown but models covariate variables through a regression model. However, it is possible that there are factors that are influencing the variable of interest and cannot be measured. This unobserved covariable can lead to very different conclusions, biased estimators and reduced efficiency of the model (Heckman and Singer, 1982; Vaupel and Yashin; 1985; Trussel and Rodrigues, 1990). Several analysis in epidemiology and prognostic studies require the inclusion of non-observable covariates. For example, studies about incidence of colon cancer can depend on familiar variable or genetic factors, also it can depend on environment factors shared by elements of the same family or

*Corresponding author: IMECC, UNICAMP - Caixa Postal 6065, 13081-970 - Campinas, SP - BRAZIL - email: nancy@ime.unicamp.br

living in a close neighborhood. This heterogeneity is frequently called biological variation and it is recognized as one of the most important source of variations in medicine and biology (Aalen, 1988).

Frailty models were introduced by Vaupel, Manton and Stallard (1979) as a generalization of the Cox's proportional hazards models allowing a random effect due to the unobserved heterogeneity of each individual (or group). Estimation for the frailty model has been studied by several authors. In particular, when the frailty has gamma distribution Nielsen *et al.* (1992) proposed a maximum likelihood estimator for cumulative hazard function and the variance of the random effect which is consistent and asymptotically normal (Murphy, 1994, 1995). The gamma distribution of the frailty has been used by several authors who justify this choice based on its analytic simplicity and its variety of forms as the parameters vary. Obviously, the gamma distribution is not the only choice for the distribution of the frailty. Several other parametric distributions have been suggested such as absolute value of normal (to be called from here on absolute normal), log-normal, beta among others (Heckman and Singer, 1982; Hougaard, 1984, 1986; Vaupel, 1990, b; Aalen, 1989). However, the frailty is an unobservable variable and it is extremely important to study robust properties of estimators under misspecification of the frailty distribution. For gamma frailty, there is the problem of identifiability (Hoem, 1990). However, making the restriction of unit expectation and letting the variance be the unknown parameter leads to a nice interpretation. The variance models the heterogeneity, and when it vanishes the frailty is identically one for all the subjects (no heterogeneity). For example, to have the asymptotic distribution of the variance estimator allows to design hypothesis testing for the heterogeneity of the model. The objective of this work is to show in a simulation study that asymptotic properties of the estimator for the variance proposed by Nielsen *et al.* (1992) under non-gamma distributions are not robust.

The outline of the paper is as follows. In Section 2, we describe for completeness, the estimator under study and its asymptotic properties under gamma frailty, most of this results can be found in Nielsen *et. al.* (1992) and Murphy (1994, 1995). Section 3 presents the simulation study under three distributions: gamma, log-normal and absolute normal and we can see that although the estimator performs well for the gamma case, it lacks consistency for the other distributions.

2 Maximum likelihood estimation under gamma distribution

Nielsen *et al.* (1992) proposed a counting process approach for estimation in frailty models assuming that the random effect follows a gamma distribution. Let $\mathbf{N}$ be a multivariate counting

process with components N_{ih} where the components with the same value of the index i share the same frailty variable Z_i. Usually the index i refers to a group, and h to stratum (treatment). The intensity of the process N_{ih} is denoted by λ_{ih}. Consider Y_i an non-negative observable predictable process and α the basic unknown risk function. The random effects Z_i are i.i.d. gamma distributed random variables. As pointed before, in order to deal with the identifiability problem (Hoem, 1990) we are going to make the restriction $\mathbb{E}(Z_i) = 1$ and $\mathrm{Var}\,(Z_i) = \theta$ and work a single parameter θ. If $\theta = 0$ then $Z_i \equiv 1$ and there is no heterogeneity in the model. We are going to concentrate on the semi-parametric model:

$$\lambda_{ih}(t) = Z_i Y_{ih}(t)\alpha_h(t) \tag{2.1}$$

where the basic risk functions α_h are unknown and need to be estimated. That is, the goal is to jointly estimate θ and the accumulated risk functions $A_h(t) = \int_0^t \alpha_h(u)du$, based solely in the observations of $(\mathbf{N}, \mathbf{Y})$. In this case, it is possible to write the joint likelihood of $(\mathbf{N}, \mathbf{Y})$. First, write the joint distribution of $(\mathbf{N}, \mathbf{Y})$ given $\mathbf{Z} = \mathbf{z}$

$$f_{(\mathbf{N},\mathbf{Y})\ |\ \mathbf{Z}=\mathbf{z}}(\mathbf{n}, \mathbf{y}) = \prod_h \prod_t (z_i Y_{ih}(t)\alpha_h(t))^{\Delta N_{ih}(t)} \exp\Big[-z_i \int_0^\tau Y_{ih}(u)\, dA_h(u)\Big], \tag{2.2}$$

where τ denotes the end of the observation period. Multiplying the conditional density (2.2) by the gamma density of $\mathbf{Z}$ we obtain the joint distribution of $(\mathbf{N}, \mathbf{Y}, \mathbf{Z})$ as

$$f_{\mathbf{N},\mathbf{Y},\mathbf{Z}}(\mathbf{n}, \mathbf{y}, \mathbf{z}) = \prod_h \prod_t (z_i Y_{ih}(t)\alpha_h(t))^{\Delta N_{ih}(t)} \exp\Big[-z_i \int_0^\tau Y_{ih}(u)\, dA_h(u)\Big] \frac{(\frac{1}{\theta})^{\frac{1}{\theta}} (z_i)^{\frac{1}{\theta}-1}}{\Gamma(\frac{1}{\theta})} \exp\Big[-\frac{1}{\theta}\, z_i\Big]. \tag{2.3}$$

Integrating over $\mathbf{z}$ the complete density given by (2.3), we obtain the joint density of $(\mathbf{N}, \mathbf{Y})$

$$f_{\mathbf{N},\mathbf{Y}}(\mathbf{n}, \mathbf{y}) = \frac{(\frac{1}{\theta})^{\frac{1}{\theta}}}{\Gamma(\frac{1}{\theta})} \prod_h \prod_t (Y_{ih}(t)\alpha_h(t))^{\Delta N_{ih}(t)} \frac{\Gamma(\sum_u \Delta N_i h(u) + \frac{1}{\theta})}{[\frac{1}{\theta} + \int_0^\tau Y_{ih}(u) dA_h(u)]^{\sum_u \Delta N_{ih}(u) + \frac{1}{\theta}}}. \tag{2.4}$$

Dividing the joint density of $(\mathbf{N}, \mathbf{Y}, \mathbf{Z})$ by the marginal density of $(\mathbf{N}, \mathbf{Y})$ we have that the conditional distribution of $\mathbf{Z}$ given $(\mathbf{N}, \mathbf{Y})$ is product of gamma densities with mean

$$\frac{\sum_u \Delta N_i h(u) + \frac{1}{\theta}}{[\frac{1}{\theta} + \int_0^\tau Y_{ih}(u) dA_h(u)]} \tag{2.5}$$

and variance

$$\frac{\sum_u \Delta N_i h(u) + \frac{1}{\theta}}{[\frac{1}{\theta} + \int_0^\tau Y_{ih}(u) dA_h(u)]^2}. \tag{2.6}$$

In this case, it is possible to jointly estimate (θ, A) using the EM algorithm (Dempster *et al.*, 1977). The **E**-step consists in estimating the value of the $\mathbf{Z}_i$'s by their conditional expectation given $(\mathbf{N}, \mathbf{Y})$

$$\mathrm{E-step}: \qquad \hat{Z}_i = \frac{\sum_h \Delta N_{ih}(\tau) + \frac{1}{\theta}}{[\frac{1}{\theta} + \sum_h \int_0^\tau Y_{ih}(u) dA_h(u)]} \tag{2.7}$$

and the M-step is given by the Nelson-Aalen estimator of A given by

$$\mathrm{M-step}: \qquad \hat{A}_h(t) = \int_0^t \frac{dN_{\cdot h}(u)}{\sum_i \hat{Z}_i Y_{ih}(u)} \tag{2.8}$$

where $N_{\cdot h} = \sum_i N_{ih}$.

Given the estimates $\hat{Z}_i$ and $\hat{A}(t)$, we can estimate the hazard function for each individual as

$$\hat{\Lambda}_i(t) = \hat{Z}_i \int_0^t Y_i(u)\, \hat{\alpha}(u)\, du = \hat{Z}_i\, \hat{A}(\max\{t; Y_i(t) = 1\}). \tag{2.9}$$

The likelihood $L(\theta)$ was based on the joint distribution of (N, Y) and the EM algorithm to obtain the estimates $\hat{A}_1$ and $\hat{A}_2$ of the risk function for each stratum. Let

$$L(\theta) := \prod_{i=1}^{n} \prod_{h=1}^{2} \left\{ \frac{\left(\frac{1}{\theta}\right)^{\frac{1}{\theta}}}{\Gamma\left(\frac{1}{\theta}\right)} \prod_t (Y_{ih}(t) dA_h(t))^{\triangle N_{ih}(t)} \frac{\Gamma\left(N_{ih}(\tau) + \frac{1}{\theta}\right)}{\left[\frac{1}{\theta} + \int_0^\tau Y_{ih}(u) dA_h(u)\right]^{N_{ih}(\tau) + \frac{1}{\theta}}} \right\}. \tag{2.10}$$

Let $\hat{\theta}$ denote the maximum likelihood estimator as obtained by Nielsen *et al.* as the argument that maximizes $L(\theta)$. However, since θ represents the variance of the frailty distribution, we consider the parameter space to be $[0, \infty)$ and compute the maximum likelihood estimator of θ as

$$\tilde{\theta} := \arg\max_{\theta \geq 0} L(\theta). \tag{2.11}$$

2.1 Asymptotic results under gamma distribution

In this section, we state the results of Murphy (1994, 1995) for the estimator of the variance θ. Call θ_0 the true value of the variance. The results for $\tilde{\theta}$ follow immediately from these ones.

2.1.1 Consistency and asymptotic normality

Theorem 2.12 (Murphy (1994, 1995)) *i)* $\hat{\theta} \xrightarrow{P} \theta_0$ *and*
ii) $\sqrt{n}\ (\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$
as $n \to \infty$ *where* σ^2 *is computed through the Fisher information matrix*

$$\sigma^2 = \left[E\left(- \frac{\partial^2 L(\theta, A)}{\partial \theta^2}\bigg|_{(\theta_0, A_0)} \right) \right]^{-1}. \tag{2.13}$$

In the following, we are going to detail some of the calculation necessary to obtain σ^2. In order to compute the second partial derivative of $L(\theta, A)$ with respect to θ, we can use the following equivalent definition

$$\begin{aligned} L(\theta, A) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \ln\left(1+\theta N_i(u-)\right) dN_i(u) - \left(\theta^{-1}+N_i(\tau)\right)\ln\left(1+\theta\int_0^\tau Y_i(u)dA(u)\right) \\ &\quad + \int_0^\tau \ln\left(Y_i(u)\triangle A(u)\right) dN_i(u). \end{aligned} \tag{2.14}$$

Therefore,

$$\begin{aligned} -\left.\frac{\partial^2 L(\theta, A)}{\partial\theta^2}\right|_{(\theta_0, A_0)} &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \left(\frac{N_i(t-)}{1+\theta_0 N_i(t-)}\right)^2 dN_i(t) - N_i(\tau)\left(\frac{\int_0^\tau Y_i dA_0}{1+\theta_0\int_0^\tau Y_i dA_0}\right)^2 + \\ &\quad 2\theta_0^{-3}\left[\log\left(1+\theta_0\int_0^\tau Y_i dA_0\right) - \frac{\theta_0\int_0^\tau Y_i dA_0}{1+\theta_0\int_0^\tau Y_i dA_0} - \frac{1}{2}\left(\frac{\theta_0\int_0^\tau Y_i dA_0}{1+\theta_0\int_0^\tau Y_i dA_0}\right)^2\right], \end{aligned}$$

when $\theta_0 = 0$, the last term is defined by its limit $\frac{2}{3}\left(\int_0^\tau Y_i dA_0\right)^3$.

Notice that $\tilde{\theta} = \hat{\theta}\mathbf{1}_{\hat{\theta}\geq 0}$ and we have

Proposition 2.15 *The restricted estimator $\tilde{\theta}$ satisfies:*

i) $\tilde{\theta} \xrightarrow{P} \theta_0$;

ii) $\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{\mathcal{D}} G$.

as $n \to \infty$ where G is a random variable with cumulative distribution function given by

$$F_G(u) = (1/2)\mathbf{1}_{[\theta_0=0]}\mathbf{1}_{\{0\}}(u) + \Phi(u/\sigma)\mathbf{1}_{(0,\infty)}(u) \tag{2.16}$$

where Φ is the cumulative distribution function of the standard normal and σ is given by expression (2.13).

2.1.2 Asymptotic variance σ^2

Case 1: $\theta_0 = 0$. In this case, all $Z_i \equiv 1$ and there is unobserved heterogeneity and σ^2 can be obtained through the following expected value:

$$\begin{aligned} &E\left\{n^{-1}\sum_{i=1}^{n}\sum_{h=1}^{2}\left[\int_0^\tau (N_{ih}(t-))^2 dN_{ih}(t) - N_{ih}(\tau)\left(\int_0^\tau Y_{ih}dA_{0h}\right)^2 + \frac{2}{3}\left(\int_0^\tau Y_{ih}dA_{0h}\right)^3\right]\right\} \\ &\quad = n^{-1}\sum_{i=1}^{n}\sum_{h=1}^{2}\left\{E\left[\int_0^\tau (N_{ih}(t-))^2 dN_{ih}(t)\right] - E\left[N_{ih}(\tau)\left(\int_0^\tau Y_{ih}dA_{0h}\right)^2\right]\right. \\ &\quad\quad \left. + \frac{2}{3}E\left[\left(\int_0^\tau Y_{ih}dA_{0h}\right)^3\right]\right\}. \end{aligned} \tag{2.17}$$

Noting that $\alpha_{0h}(t) = \lambda$, $N_{ih}(\tau) = 1$ and $N_{ih}(t-) = 0$ since before time t there is no failure,

$$E\left[\left(\int_0^\tau Y_{ih} dA_{0h}\right)^3\right] = E\left[(\lambda \max\{t : Y_{ih}(t) = 1\})^3\right], \tag{2.18}$$

$$E\left[N_{ih}(\tau)\left(\int_0^\tau Y_{ih} dA_{0h}\right)^2\right] = E\left[(\lambda \max\{t : Y_{ih}(t) = 1\})^2\right] \tag{2.19}$$

$$E\left[\int_0^\tau (N_{ih}(t-))^2 dN_{ih}(t)\right] = E\left[\sum_{t>0}^{\tau} (N_{ih}(t-))^2 \triangle N_{ih}(t)\right] = 0. \tag{2.20}$$

Therefore, for non-censored data we have

$$\sigma^2 = \left\{ n^{-1} \sum_{i=1}^{n} \sum_{h=1}^{2} \left[\frac{2}{3} E\left[(\lambda \max\{t : Y_{ih}(t) = 1\})^3\right] - E\left[(\lambda \max\{t : Y_{ih}(t)\})^2\right]\right]\right\}^{-1}. \tag{2.21}$$

If there are censored data, we have to modify (2.19) as

$$\begin{aligned} E\left[N_{ih}(\tau)\left(\int_0^\tau Y_{ih} dA_{0h}\right)^2\right] &= E\left[E\left[N_{ih}(\tau)\left(\int_0^\tau Y_{ih} dA_{0h}\right)^2 | N_{ih}(\tau)\right]\right] \\ &= E\left[N_{ih}(\tau) E\left[(\lambda \max\{t : Y_{ih}(t) = 1\})^2\right]\right]. \end{aligned} \tag{2.22}$$

Using the fact that $N_{ih}(\tau) = 1$ if, and only if there is failure we have

$$\sigma^2 = \left\{ n^{-1} \sum_{i=1}^{n} \sum_{h=1}^{2} \left[\frac{2}{3} E\left[(\lambda \max\{t : Y_{ih}(t) = 1\})^3\right] - E\left[(\lambda \max\{t : Y_{ih}(t) = 1\})^2\right] P(\text{failure})\right]\right\}^{-1} \tag{2.23}$$

Case 2: $\theta_0 > 0$. In this case, the computation of the expression (2.13) is much more difficult. A much simpler approach is to use the observed Fisher information number given by

$$\begin{aligned} I\left(\hat{\theta}\right) = n^{-1} \sum_{i=1}^{n} \sum_{h=1}^{2} \Bigg\{ & \int_0^\tau \left(\frac{N_{ih}(t-)}{1 + \hat{\theta} N_{ih}(t-)}\right)^2 dN_{ih}(t) \\ & - N_{ih}(\tau) \left(\frac{\int_0^\tau Y_{ih} dA_{0h}}{1 + \hat{\theta} \int_0^\tau Y_{ih} dA_{0h}}\right)^2 + 2\hat{\theta}^{-3} \Bigg[\log\left(1 + \hat{\theta} \int_0^\tau Y_{ih} dA_{0h}\right) \\ & - \frac{\hat{\theta} \int_0^\tau Y_{ih} dA_{0h}}{1 + \hat{\theta} \int_0^\tau Y_{ih} dA_{0h}} - \frac{1}{2}\left(\frac{\hat{\theta} \int_0^\tau Y_{ih} dA_{0h}}{1 + \hat{\theta} \int_0^\tau Y_{ih} dA_{0h}}\right)^2\Bigg]\Bigg\}. \end{aligned}$$

Working out the following integrals:

$$\int_0^\tau Y_{ih} dA_{0h} = \int_0^\tau Y_{ih}(t) \alpha_{0h}(t) dt = A_{0h}(\max\{t : Y_{ih}(t) = 1\}) \tag{2.24}$$

$$\int_0^\tau \left(\frac{N_{ih}(t-)}{1+\hat{\theta}N_{ih}(t-)}\right)^2 dN_{ih}(t) \quad = \quad \sum_{t>0}\left(\frac{N_{ih}(t-)}{1+\hat{\theta}N_{ih}(t-)}\right)^2 \triangle N_{ih}(t) \quad = \quad 0. \tag{2.25}$$

we obtain the following expression:

$$\begin{aligned} I(\hat{\theta}) \quad = \quad n^{-1} \sum_{i=1}^{n} \sum_{h=1}^{2} \Bigg\{ & -N_{ih}(\tau)\left(\frac{A_{0h}\ \left(\max\{t : Y_{ih}(t)=1\}\ \right)}{1+\hat{\theta}A_{0h}\left(\max\{t : Y_{ih}(t)=1\}\right)}\right)^2 \\ & + 2(\hat{\theta})^{-3}\Bigg[\log\left(1+\hat{\theta}A_{0h}\left(\max\{t : Y_{ih}(t)=1\}\right)\right) - \frac{\hat{\theta}A_{0h}\left(\max\{t : Y_{ih}(t)=1\}\right)}{1+\hat{\theta}A_{0h}\left(\max\{t : Y_{ih}(t)=1\}\right)} \\ & -\frac{1}{2}\left(\frac{\hat{\theta}A_{0h}\left(\max\{t : Y_{ih}(t)=1\}\right)}{1+\hat{\theta}A_{0h}\left(\max\{t : Y_{ih}(t)=1\}\right)}\right)^2\Bigg]\Bigg\}. \end{aligned} \tag{2.26}$$

3 Simulation studies

In the following, we will follow the simulation procedure of Nielsen *et al* (1992) and concentrate on the two sample case, that is $h = 1, 2$. In this case, we are assuming that there are two individuals sharing the same frailty, for example, brothers. For simplicity, we take $\alpha_1(t) = \alpha_2(t) \equiv 1$, however in the analysis, A_1 and A_2 are estimated non-parametrically. For selected values of the variance parameter θ of the frailty density, we generate m datasets of n independent pairs (t_{i1}, t_{i2}) of survival times in the following way using S-Plus to generate the random variables:
$v_{ih}, i = 1, \ldots, n,\ h = 1, 2$ independent exp(1) random variables;
$z_i, i = 1, \ldots, n$, independent and identically distributed random variables with mean one and variance θ (to be generated using gamma, log-normal and absolute normal distributions);

$$t_{ih} = v_{ih}/z_i. \tag{3.1}$$

The data were analyzed twice, one time without censoring. The second time a $U(0, 8)$ censoring variable was used. That is, let $c_{ih}, i = 1, \ldots, n,\ h = 1, 2$ be independent and identically distributed $U(0, 8)$ random variables and let

$$t_{ih} = \min\{v_{ih}/z_i, c_{ih}\}. \tag{3.2}$$

3.1 Estimation of θ_0

For all of the cases, the tables present the average and standard deviation for several simulation studies. For all cases, the sample size n are 100, 200, 500 and 1000. The estimator $\tilde{\theta}$ were computed

using the procedure described above. These values can be compared with the values $\hat{\theta}$ presented in Nielsen *et al.* (1992). In all cases, we have $m = 200$ repetitions of the experiment. All simulations were carried using S-plus running on a PC. Matlab for Windows was used for the maximization and iteration procedure.

3.1.1 Gamma frailty

Table 1 presents the average and standard deviation when the data was generated by equations (3.1) or (3.2) and $z_1, \ldots, z_n$ are independent and identically distributed gamma random variables with mean one and variance θ. Also, we present in this table the standard deviation of the estimator using expression (2.21) and (2.23) for $\theta_0 = 0$. For the case, $\theta_0 > 0$ only the variance based on the observed Fisher information given by (2.26) was computed to obtain $\hat{\sigma}(\tilde{\theta})$. As expected the estimative $\tilde{\theta}$ are very close to the true parameter value θ_0. Also, as n increases the approximation gets better. In fact, it is closer for censored data, although the standard deviation also increases.

Figure 1 presents the histograms of the simulated values. We can observe that for $\theta_0 = 0$ we have a mixture of a normal random variable and a discrete variable with mass concentrated at 0. For $\theta_0 > 0$ we can see that as n increases the approximation to a normal variable is attained. These conclusions were expected in view of the results of Murphy (1994, 1995), however they were included for the sake of comparison with the log-normal and normal case.

3.1.2 Log-normal and absolute normal frailty

Table 2 presents the average and standard deviation when the data was generated by equations (3.1) or (3.2) and $z_1, \ldots, z_n$ are independent and identically distributed log-normal random variables and $z_1 = |w_1|, \ldots, z_n = |w_n|$, where $w_1, \ldots, w_n$ are independent and identically distributed normal random variables with mean one and variance θ. In these cases, the expressions for the observed Fisher information are not so easily obtained as in (2.26) and are not presented. Notice that for the log-normal case the values of $\tilde{\theta}$ underestimates the true variance, whereas for the absolute normal case it overestimates it, even for very large sample $n = 1000$ it has a very big bias. On the other hand, Figures 2and 3 present the histograms of the simulated values and we can see that, although the mean do not approach the true value, the curves approaches a normal curve.

		Uncensored data				Censored data			
n	θ_0	mean	SD	$\sigma(\tilde{\theta})$	$\hat{\sigma}(\tilde{\theta})$	mean	SD	$\sigma(\tilde{\theta})$	$\hat{\sigma}(\tilde{\theta})$
	0.0	0.0201	0.0471	0.0500	0.0415	0.0276	0.0503	0.0471	0.0501
100	0.2	0.1476	0.1188	—	0.0967	0.1614	0.1398	—	0.1248
	0.4	0.3133	0.1634	—	0.1337	0.3200	0.1739	—	0.1679
	0.0	0.0170	0.0320	0.0353	0.0289	0.0175	0.0366	0.0333	0.0352
200	0.2	0.1639	0.0883	—	0.0658	0.1653	0.0917	—	0.0862
	0.4	0.3460	0.1181	—	0.1258	0.3519	0.1316	—	0.1357
	0.0	0.0110	0.0201	0.0223	0.0183	0.0119	0.0219	0.0210	0.0222
500	0.2	0.1816	0.0592	—	0.0475	0.1855	0.0719	—	0.0568
	0.4	0.3713	0.0749	—	0.0958	0.3832	0.0877	—	0.0978
1000	0.2	0.1884	0.0468	—	0.0353	—	—	—	—
	0.4	0.3756	0.0585	—	0.0693	—	—	—	—

Table 1: Mean and standard deviation for 200 replication of estimate of θ_0, $\sigma(\tilde{\theta})$ and $\sigma(\tilde{\theta})$ are the standard deviation of the estimate computed using true and observed Fisher information number respectively, under gamma frailty

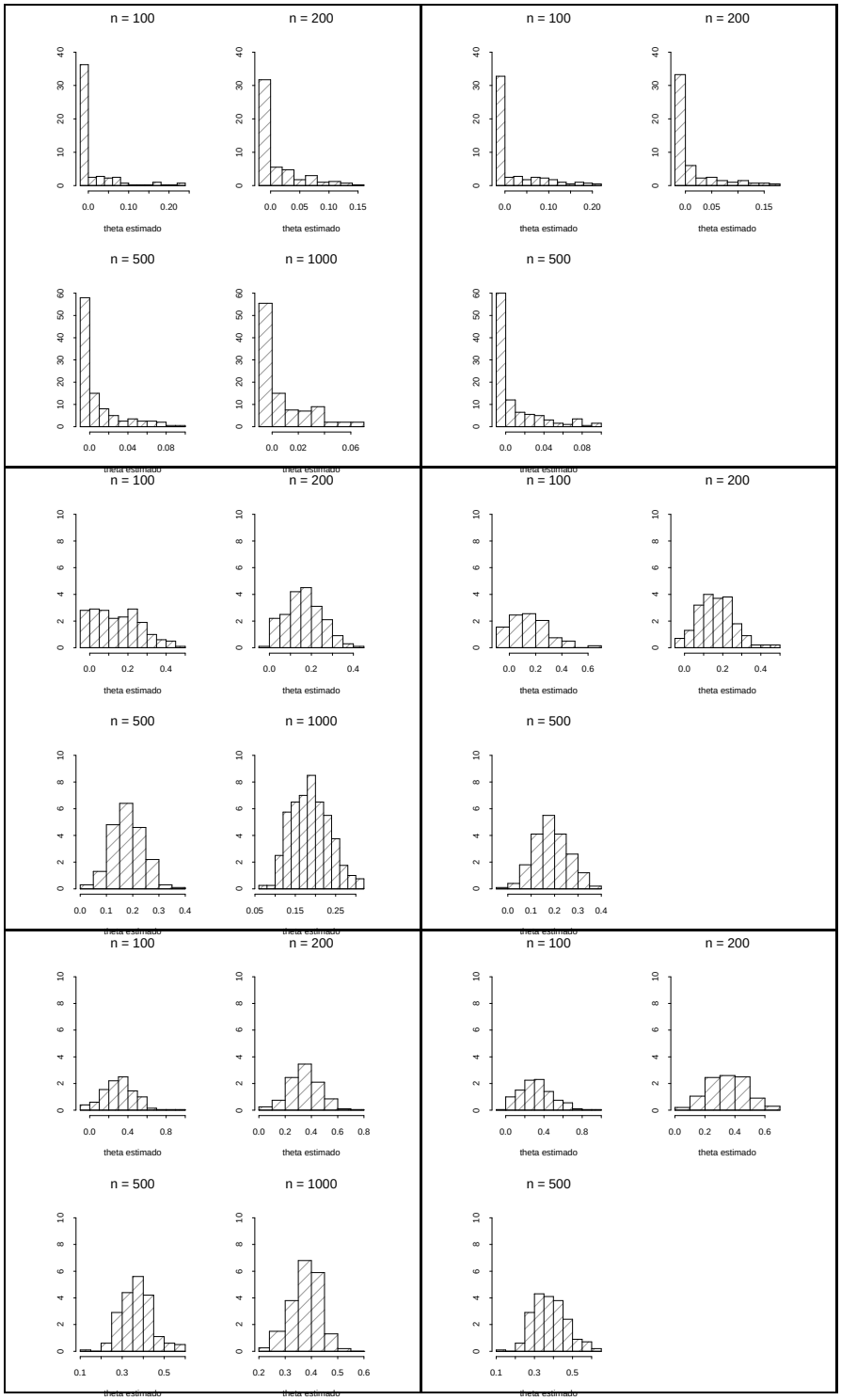
		log-normal				absolute normal			
		Uncensored		Censored		Uncensored		Censored	
n	θ_0	mean	SD	mean	SD	mean	SD	mean	SD
100	0.2	0.1035	0.0938	0.1091	0.1074	0.3526	0.1844	0.3028	0.1809
	0.4	0.1862	0.1326	0.2136	0.1568	0.5395	0.1817	0.5199	0.1977
200	0.2	0.1195	0.0866	0.1292	0.0950	0.3255	0.1135	0.2931	0.1221
	0.4	0.2280	0.1063	0.2533	0.1225	0.5749	0.1340	0.5454	0.1430
500	0.2	0.1361	0.0574	0.1388	0.0644	0.3470	0.0680	0.3072	0.0705
	0.4	0.2407	0.0624	0.2573	0.0722	0.6076	0.0907	0.5691	0.0976
1000	0.2	0.1388	0.0374	—	—	0.3546	0.0515	—	—
	0.4	0.2573	0.0469	—	—	0.6275	0.0675	—	—

Table 2: Average and standard deviation for 200 replication of estimate of θ_0 under log-normal and absolute normal frailty

$\theta_0 = 0$

$\theta_0 = 0.2$

$\theta_0 = 0.4$

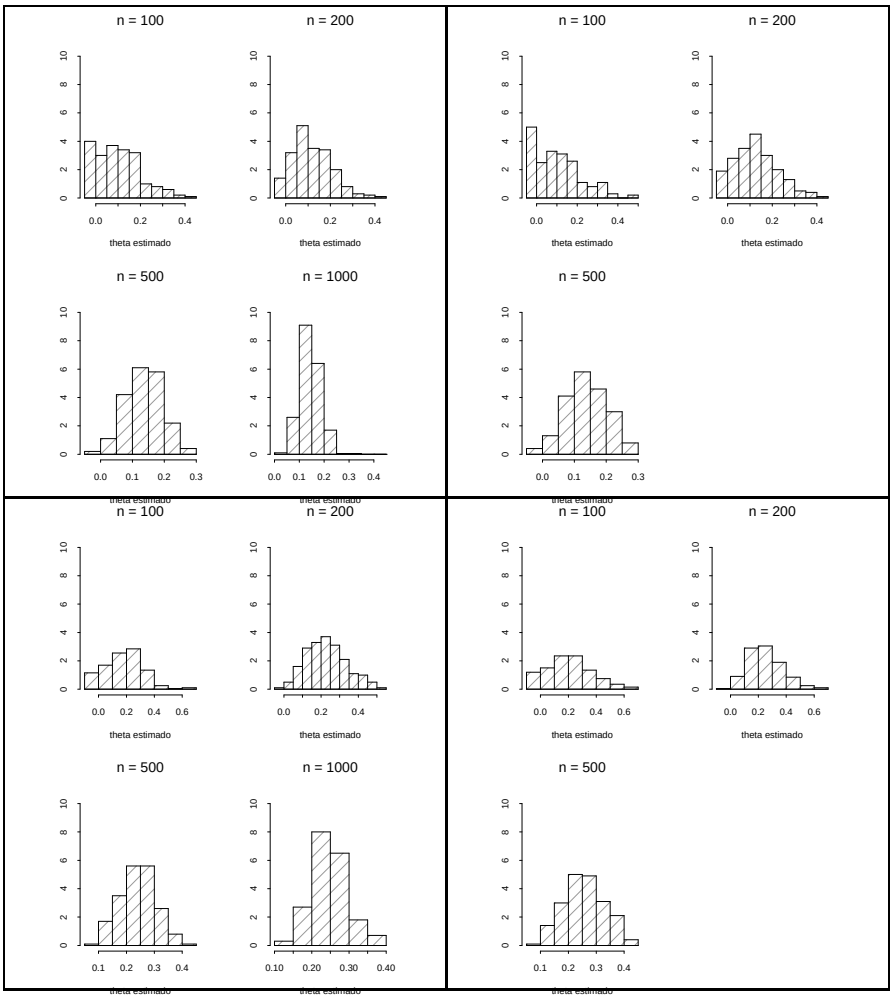


Uncensored data Censored data

Figure 1: Histogram for 200 simulation of the estimator $\tilde{\theta}$ under gamma frailty

$\theta_0 = 0.2$

$\theta_0 = 0.4$

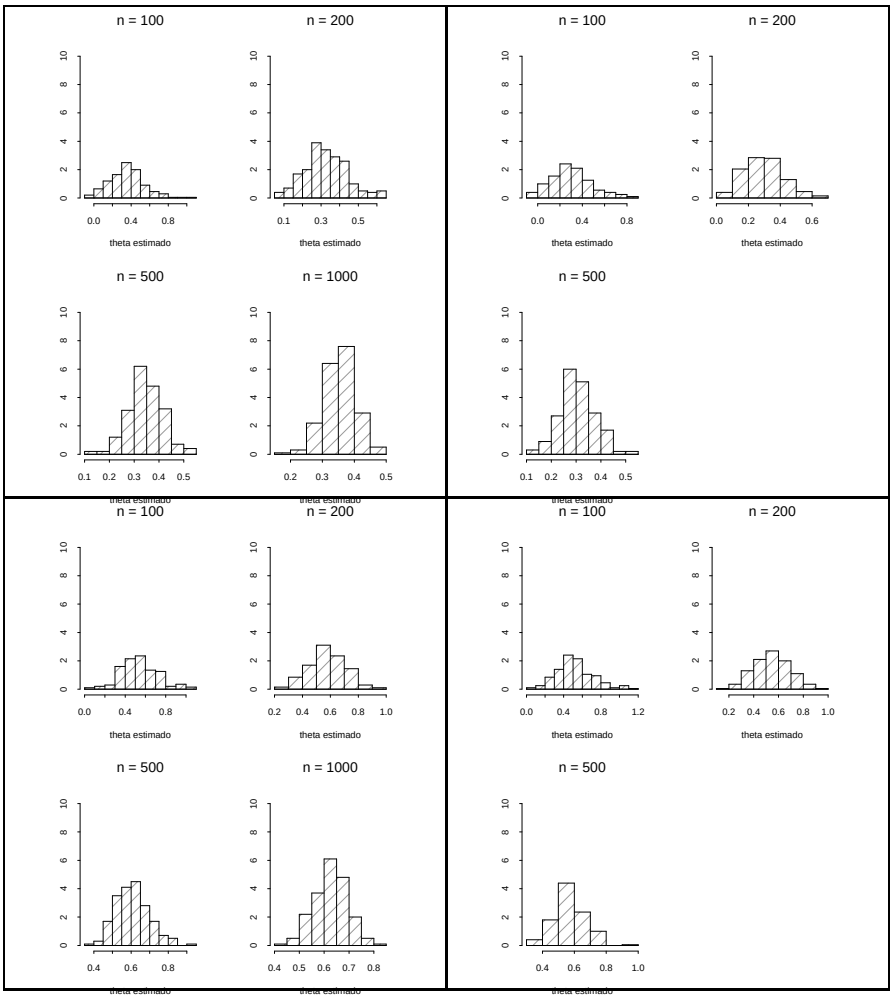


Uncensored data Censored data

Figure 2: Histogram for 200 simulation of the estimator $\tilde{\theta}$ under log-normal frailty

$\theta_0 = 0.2$

$\theta_0 = 0.4$



Uncensored data Censored data

Figure 3: Histogram for 200 simulation of the estimator $\tilde{\theta}$ under absolute normal frailty

4 Conclusion

The estimator proposed by Nielsen *et al.* (1992) based on the EM-algorithm is very good and it has optimal asymptotic properties for the case that it was designed for, unobserved frailty variable with gamma distribution (Murphy, 1994, 1995). However, as the simulation study shows, the estimator is not consistent when the gamma assumption fails. It underestimates the variance for log-normal frailty and overestimates it in the absolute normal case. On the other hand, the histograms show that, although the estimator is not consistent, it still follows asymptotically a normal distribution and maybe it could be possible to find a non-parametric correction for the bias. This is not, however the objective of this work. We would like to stress that, the frailty variable is not observable, therefore it cannot be tested to check whether it satisfies the distributional assumption. Consequently, caution must be taken when using a parametric procedure, since a misspecification on the hypothesis can lead to a very strong bias.

Acknowledgments This work was partially supported by CAPES and FAEP and it is part of AF's monograph written for fulfillment of the requirements for the Master degree in Statistics at University of Campinas. This manuscript was written while NLG was Visiting Scholar at Department of Statistics - University of California - Berkeley and it was partially supported by CNPq Grant 301054/93-2 and FAPESP Grant99/00260-3.

References

Aalen, O.O. (1988) Heterogeneity in survival analysis. *Statistics in Medicine*, **7**: 1121–1137.

Aalen, O.O. (1989) A linear-regression model for the analysis of life times, *Statistics in Medicine*, **8**(8): 907–925.

Cox, D.R. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society B*, **34**:187–202.

Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* **38**:1–38.

Greenwood, P.E.; Wefelmeyer, W. (1990) Efficiency of estimators for partially specified filtered models. *Stochastic Process. Appl.* **36**:353–370.

Heckman, J.; Singer, B. (1982) Population heterogeneity in demographic models. In: Land, K.; Rogers, A. (Ed.) *Multidimensional Mathematical Demography.* New York, Academic Press. pp. 567–599.

Hoem, J.M. (1990) Identifiability in hazards models with unobserved heterogeneity: The compatibility of two apparently contradictory results. *Theoretical Population Biology* **37**: 124–128.

Hougaard, P. (1984) Life table methods for heterogeneous populations – Distributions describing the heterogeneity. *Biometrika* **71**:75–83.

Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**:387–396.

Murphy, S.A. (1994) Consistency in a Proportional Hazards Model Incorporating a Random Effect. *The Annals of Statistics* **22**:712-731.

Murphy, S.A. , (1995) Asymptotic Theory for the Frailty Model. *The Annals of Statistics* **23**:182–198.

Nielsen, G.G.; Gill, R.D.; Andersen, P.K.; Sorensen, T.I.A. (1992) A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models. *Scandinavian Journal of Statistics* **19**:25–43.

Trussel, J.; Rodrigues, G. (1990) Heterogeneity in demographic research. In: Adams, J. *et al.* (Ed.) *Convergent Questions in Genetics an Demography. New York, Oxford University Press.* pp. 111–132.

Vaupel, J.W. (1990) Relative Risks – Frailty models of life-history data. Theor. Popul. Biol. **37**(1): 220-234.

Vaupel,J.W.; Manton, K.G.; Stallard, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**:439–454.

Vaupel,J.W.; Yashin, I.A. (1985) Heterogeneity′s ruses: some surprising effects of selection on population dynamics. *American Statistician* **39**:176–185.