Degradation data analysis using the Probability Integral Transformation and a linear degradation path model with Weibull distributed random effects

Clódio de Almeida
Departamento de Estatística - UFMG
Av. Presidente Antônio Carlos, 6627. Belo Horizonte - MG,Brasil.
clodioalm@gmail.com

Marta Afonso Freitas
DEP-Laboratório de Apoio à Decisão e Confiabilidade (LADEC)-UFMG
Av. Presidente Antônio Carlos, 6627. Belo Horizonte - MG,Brasil.
marta.afreitas@gmail.com

Enrico Antônio Colosimo
Departamento de Estatística -UFMG
Av. Presidente Antônio Carlos, 6627. Belo Horizonte - MG,Brasil.
enricoc@est.ufmg.br

RESUMO

Dados de degradação são, em geral, modelados utilizando um modelo de regressão não linear com coeficientes aleatórios. A estimação por máxima verossimilhança (MV) é frequentemente implementada nesses casos utilizando a suposição de normalidade dos efeitos aleatórios é frequentemente utilizada. Infelizmente esta suposição pode não ser realista ou válida na prática. A má-especificação da distribuição dos efeitos aleatórios leva a estimativas viciadas de características da confiabilidade tais como quantis da distribuição do tempo até a falha e do tempo médio até a falha. Neste artigo utilizamos a *Probability Integral Transformation (PIT)* proposta por Nelson (2006) para obter estimativas de MV quando se assume que os efeitos aleatórios têm distribuição não normal. Em particular, o método é utilizado para um modelo de perfil linear de degradação com efeitos aleatórios com distribuição Weibull. O método é aplicado a uma situação real.

PALAVRAS CHAVE. Dados de degradação. Modelo de efeitos aleatórios. Confiabilidade.

MAIN AREA: 1) ESTATÍSTICA; 2) MODELOS PROBABILÍSTICOS.

ABSTRACT

Degradation data are usually modeled by a nonlinear regression model with random coefficients. Maximum Likelihood (ML) estimation is often implemented in those cases using normality assumption of random effects. Unfortunately this assumption may be unrealistic or invalid in practice. Misspecification of the random effects distribution leads to biased estimates of the reliability figures of interest such as quantiles of the time to failure distribution and the mean time to failure. In this paper we use the Probability Integral Transformation (PT) proposed by Nelson et al. (2006) for obtaining ML estimates when the random effects are assumed to have non-normal distribution. In particular, the method is used for a linear path degradation model with Weibull distributed random effects. The method is applied to a real situation.

KEYWORDS. Degradation data. Random effects model. Reliability.

MAIN AREA: 1) STATISTICS; 2) PROBABILITY MODELS .

1. Introduction

In modern high-reliability applications, manufacturers have to face short product development times and, consequently, the reliability tests must be conducted with severe time constraints. In addition, high-reliability components or products may last for many years so one might not expect to see failures in the reliability testing or very few, if any, on the field in a reasonable amount of time. This will result in limited information about reliability needed for product design and reliability improvement. Even using the technique of accelerating the life by testing at higher levels of stress (accelerated life testes) provide little help, because no failures are likely to occur in a reasonable amount of time. If, however, we could monitor, over time, a degradation (or a performance) variable that is closely related to failure (e.g., length of a fatigue crack, light output of a laser or the amount of wear of automobile tires) on all the test units, there would be a large amount of reliability information.

Usually, in order to conduct a degradation test, one has to prespecify a threshold level of degradation, obtain measurements of degradation at different fixed times, and define that failure occurs when the amount of degradation for a test unit exceeds that level. Thus, these degradation measurements may provide some useful information to assess reliability even when failures do not occur during the test period.

In the literature, there are basically two major approaches of modeling for degradation data. One approach is to use stochastic models. Whitmore and Shenkelbert (1997) and Patterson and Thompson (1971) for example, used Wiener diffusion processes with a drift to explain degradation paths, with the advantage being that a time-to-failure distribution is readily available, i.e., the inverse Gaussian distribution.

An alternative approach is to consider more general statistical models. Degradation in these models is modeled by a function of time and some possibly multidimensional random variables. These models are called *general degradation path models*. In this case the degradation measurements of the individual units are modeled using the same functional form and differences between individual units using random effects. These models take the general form:

$$Y_{ij} = D(t_{ij}; a; \beta_i) + \varepsilon_{ij} \qquad (1)$$

where $Y_{ij}\,(i = 1,2,...,n;\ j = 1,2,...,m_i)$ is the random variable representing the amount of degradation of the i^{th} unit at a prespecified time t_{ij}; m_i is the number of repeated degradation measurements taken on the i^{th} unit; $D(t_{ij}; a; \beta_i)$ is the actual degradation path of the i^{th} unit ; $a = (a_1, a_2, ..., a_p)^t$ is a $p\times 1$ vector of fixed effects that describes population characteristics; $\beta_i = (\beta_{i1},..., \beta_{ik})^t$ is a $k\times 1$ vector or the i^{th} unit random effects that represent an individual unit´s characteristics; ε_{ij} is the random error associated to the i^{th} unit at time t_{ij}. The deterministic form of $D(t_{ij}; a; \beta_i)$ might be based on empirical analysis of the degradation process under study, but whenever possible it should be based on the physical-chemical phenomenon associated with it. It is generally assumed that: 1) the ε_{ij} are independent and identically distributed (*iid)* as $N(0;\sigma_\varepsilon^2)$, a normal distribution with mean zero and variance σ_ε^2 (fixed and unknown); 2) the vectors of random effects $\beta_i = (\beta_{i1},..., \beta_{ik})^t\,(i = 1,..., n)$ are *iid* as $\Lambda(\beta\,|\,\theta)$, where $\Lambda(\beta\,|\,\theta)$ is a multivariate distribution function, which may depend on an unknown (fixed) $q\times 1$ parameter vector θ that must be estimated from the data and 3) the $\{\varepsilon_{ij}\}$´s and $\{\beta_i\}$´s are independent of each other $(i = 1,2,..., n;\ j = 1,2,..., m_i)$.

The distinguishing feature of the degradation data analysis for reliability assessment is the need to be implemented in two stages. In the first one, a mixed model (linear or nonlinear) is fitted to the repeated degradation measurements data and the vector of (fixed) parameters $(a^t; \theta^t; \sigma_\varepsilon^2)^t$ is estimated. In fact, this first stage is a longitudinal data analysis (Verbeke and

Molengberghs, 2000). Once the model parameters are estimated, the second stage comprises the estimation of the failure time distribution F(t). In the literature a good amount of papers have been dealing with the problem of parameter estimation. For example, Lu and Meeker (1993) developed statistical methods using degradation measures to estimate a time-to-failure distribution. They considered a nonlinear mixed-effects model and proposed a two-stage method to obtain point estimates and confidence intervals of percentiles of the failure time distribution. The authors assumed that the vector of random effects β, or some appropriate reparameterization followed a Multivariate Normal Distribution (*MVN*) with vector mean μ_β and variance-covariance matrix Σ_β. Other authors worked on approximations for the likelihood function in order to obtain the restricted maximum likelihood estimators, all under the normality assumption of the random effects. Examples are the works by Lindstrom and Bates (1990), Pinheiro and Bates (1995, 2000) among others. A good reference on degradation models under the normality assumption is Meeker and Escobar (1998).

Nevertheless, in many situations, there is a need for suitable methods to fit mixed models with non-normal random effects. Verbeke and Lesafre (1996), Molenberghs and Verbeke (2005, chap. 23) for example, proposed mixed models with mixtures of normals as random effects of the so-called *heterogeneity model*. However the implementation of these methods is challenging in practical data analysis. The difficulties related to parameter estimation in models with non-normal random effects lead to the development of alternative computational techniques in particular, two methods are available that employ transformation results. For mixed models with non-normal random effects, Nelson et al. (2006) proposed a simple computational method that uses the probability integral transformation (PIT) method to obtain the maximum likelihood estimates (MLE). It can conveniently be implemented using Gaussian quadrature tools already available in commercial statistical *softwares*, thus providing a practical estimation solution to mixed models with non-normal random effects. The authors used the procedure Proc NLMIXED available in the statistical software SAS/STAT(SAS Institute Inc.) in all the examples discussed in their paper. But the method can also be used with other commercial *softwares* such as MATLAB (MathWorks, Inc.), R (general public licence; www.r-project.org) and S-Plus (TIBCO Software Inc.).

In this paper we present the preliminary results of a research work related to the use of linear and nonlinear models with non-normal random effects to analyze degradation data. We show here how the Probability Integral Transformation (PIT) method can be used for fitting degradation path models with non-normal random effects, in particular, INA A model with Weibull distributed random effects. A simple linear path model is assumed. The approach is also applied to the wheel degradation data described in Freitas et al. (2009).

The outline of the paper is as follows. Section 2 describes briefly the motivating situation. Section 3 presents the PIT method proposed by Nelson et.al (2006). Section 4 shows the details of the likelihood maximization for Weibull random effects in a linear degradation path model where the PIT transformation is used . Section 5 shows the results of the re-analysis of the wheel degradation data and a comparison to the ones obtained by Freitas et al. (2009). Finally, Section 6 presents final remarks and suggests directions for future work.

2. Motivating situation

We use the train wheel degradation data from Freitas et al. (2009) as the motivating example. The complete database includes, among other information, the diameter measurements of train wheels, taken at equally spaced inspection times (in Km): $t_1 = 0, t_2 = 50{,}000, \ldots, t_{13} = 600{,}000$. The wheels are labeled according to their working positions in a given car using a three-dimension indicator vector, representing in this order: position (side) within an axle (left=0; right=1); truck position (front=0; back=1) and axle position within a truck (outer=0; inner=1). The diameter of a new wheel is 966 mm. When the diameter reaches 889 mm the wheel is replaced by a new one.``Failure" of the wheel is then defined to occur when the degradation ((i.e., 966-[observed diameter measure at time t]) reaches the threshold level $D_f = 77mm$. As in Freitas et al. (2009) we use in this paper the degradation measurements of the wheels labeled [0,0,0] of the locomotive cars only (a total of 14 units). One of the purposes of the study is to estimate the mean distance to failure (MTTF) and time to failure some quantiles of the failure time distribution.

3. The Probability Integral Transformation (PIT) method.

We briefly present the PIT (Probability Integral Transformation) method by Nelson et al. (2006). First recall that in order to estimate the parameters of the general degradation path model (1) using the maximum likelihood method, one has to maximize the (marginal) likelihood function given by (Meeker and Escobar, 1998):

$$L(a,\theta,\sigma_\varepsilon^2) = f(y\,|\,a,\theta,\sigma_\varepsilon^2) = \prod_{i=1}^{n} L_i = \prod_{i=1}^{n}\left\{\int_{\Xi_{\beta_i}} f(y_i\,|\,a,\beta_i,\sigma_\varepsilon^2)\, f(\beta_i\,|\,\theta)\,d\beta_i\right\} \qquad (2)$$

where Ξ_{β_i} denotes the limit of the multiple integral in β_i; $y_i = (y_{i1},\dots,y_{i,m_i})^t$ is the vector of degradation measurements for the i^{th} unit and $y = (y_1^{\,t},\dots,y_n^t)^t$ is the vector comprising all the degradations measurements (for all the n units) and $f(\beta_i\,|\,\theta)$ denotes the density of the random effects vector β_i, typically non-normal. Although the random effect can be vector valued, let us illustrate the method for the scalar case (i.e., β_i has dimension 1, for i=1,…,n). Therefore we can write the contribution of the i^{th} profile (unit) to the marginal likelihood as:

$$L_i = f(y_i\,|\,a,\theta,\sigma_\varepsilon^2) = \int_{\Xi_{\beta_i}}\left\{\prod_{j=1}^{m_i} f(y_{ij}\,|\,a,\beta_i,\sigma_\varepsilon^2)\, f(\beta_i\,|\,\theta)\right\} d\beta_i \qquad (3)$$

In order to accommodate non-normal random effects within the numerical integration techniques available in commercial *softwares*, Nelson (2006) used the probability integral transform. The procedure is described next.

Suppose that random effects (assumed continuous) have a non-normal distribution $f(\beta_i\,|\,\theta)$. Let η_i be a random effect distributed as a Standard normal, that is, $\eta_i :\ N(0,1)$. Then, using the probability integral transform, $\upsilon_i = \Phi(\eta_i)$ has a $Uniform(0,1)$ distribution. Here, is the standard normal cumulative distribution function (CDF). Applying the probability integral transform once more, $F(\beta_i\,|\,\theta)$ has also a $Uniform(0,1)$ distribution, where $F(\beta_i\,|\,\theta)$ is the CDF of β_i with parameter θ. Consequently, $\beta_i = F^{-1}(\upsilon_i\,|\,\theta)$ has density $f(\beta_i\,|\,\theta)$, where $F^{-1}(.\,|\,\theta)$ is the inverse CDF of β_i. Consequently $\beta_i = F^{-1}(\upsilon_i\,|\,\theta) = F^{-1}(\Phi(\eta_i)\,|\,\theta)$ has the non-normal distribution of interest. Therefore using probability theory for transformations, the authors re-wrote the contribution of the i^{th} profile to the marginal likelihood (equation (3)) in terms of the normal random effect η_i as :

$$L_i = f(y_i\,|\,a,\theta,\sigma_\varepsilon^2) = \int_{\Xi_{\eta_i}}\left\{\prod_{j=1}^{m_i} f(y_{ij}\,|\,a, F^{-1}(\Phi(\eta_i)\,|\,\theta),\sigma_\varepsilon^2)\,\phi_{NOR}(\eta_i)\right\} d\eta_i \qquad (4)$$

where ϕ_{NOR} is the standard normal probability density function. Numerical integration techniques can be used to approximate the marginal likelihood in (4) and other nonlinear mixed-effects models when no closed-form solution exists. In addition, standard maximization methods (*e.g.* Newton-Raphson, quasi-Newton) can be used to solve for $(a^t;\theta^t;\sigma_\varepsilon^2)^t$. The authors used the Gaussian quadrature technique, which approximates the function in (4) by a weighted average of the integrand evaluated at a number, *Q*, of predetermined abscissas (quadrature points) $d_q(q=1,\dots,Q)$ over the random effects η_i (Pinheiro and Bates (1995); Davidian and Gallant (1993); Lesaffre and Spiessens (2001)). The authors implemented the methodology in two examples using the PROC NLMIXED available in the software SAS. The PIT method can be extended in a straightforward manner to two or more independent non-normal random effects, *i.e.*, cases where $\beta_i=(\beta_{i1},\dots,\beta_{ik})^t\,(i=1,\dots,n)$. However, when correlation is present between two or more non-normal random effects, the PIT method becomes more complicated and requires the use of multivariate probability integral transform approach (Genest and Rivest, 2001).

4. Likelihood maximization for Weibull random effects in a linear degradation path model: numerical solution base on the PIT method.

In this section we present the details of a numerical solution based on the PIT method to estimate the mean time to failure and other reliability figures for the degradation model proposed for the motivating situation described in Section 2. Freitas et al. (2009) used a linear (straight line) degradation path model to analyze the train wheel degradation data. First, recall that under the general degradation path model described in Sections 1 and 3 the likelihood function takes the form of the equation (2). In order to get the maximum likelihood estimates $\hat{\alpha}_{MV},\hat{\theta}_{MV},\hat{\sigma}_\varepsilon^2$ of $\alpha,\theta,\sigma_\varepsilon^2$ respectively, it is necessary to maximize equation (2) or equivalently, the log-likelihood which, given the model assumptions can be written as:

$$\log L=\log f(y\,|\,a,\theta,\sigma_\varepsilon^2)=\sum_{i=1}^{n}\log\left\{\int_{\Xi_{\beta_i}}\left[\prod_{j=1}^{m_i}\frac{1}{\sigma_\varepsilon}\phi_{NOR}(z_{ij})\right]f(\beta_i\,|\,\theta)d\beta_i\right\} \qquad (5)$$

where $z_{ij}=\dfrac{[y_{ij}-D(t_{ij},\alpha,\beta_i)]}{\sigma_\varepsilon}$; ϕ_{NOR} is the standard normal probability density function and $f(\beta_i\,|\,\theta)$ is the probability density function of the (non-normal) random effects vector β_i .

Freitas et al. (2009) used the following degradation path model in the analysis of the wheel degradation data:

$$Y_{ij}=D(t_{ij};\beta_i)+\varepsilon_{ij}=\frac{1}{\beta_i}t_{ij}+\varepsilon_{ij}\qquad i=1,\dots,n\,(\text{units});\quad j=1,\dots,m_i\ (\text{measurements}) \qquad (6)$$

(here, $n=14$ wheels and $m_i=13$, for i=1,2,…,14). In addition to the general degradation path model assumptions (Section 1) , we assume that the $\beta_i' s$ are *iid*. according to a Weibull (α,δ) distribution with probability density function given by:

$$f(\beta\,|\,\alpha,\delta)=\frac{\alpha}{\delta}\left(\frac{\beta}{\delta}\right)^{\alpha-1}\exp\left\{-\left(\frac{\beta}{\delta}\right)^{\alpha}\right\}\quad(\alpha>0,\ \delta>0,\beta>0) \qquad (7)$$

In this case, it is possible to prove that the failure time T has also a Weibull distribution, in other words, with the model (7) and a specific D_f

$$\beta_i\sim Weibull(\alpha,\delta)\Leftrightarrow T\sim Weibull(shape,scale)=(\alpha,D_f\,\delta).$$

In addition, here $\theta = (\alpha, \delta)^t$ and with the general model assumptions, we have:

$$Y_i \mid \beta_i, \sigma_\varepsilon^2 \overset{indep}{:} N(\mu_i, \Sigma_{Y_i});\ \mu_i = (\frac{1}{\beta_i} t_{i1} \Sigma .., \frac{1}{\beta_i} t_{im_i})^t;\ \Sigma_{Y_i} = \sigma_\varepsilon^2 I_{m_i}$$

(where I_{m_i} is the identity matrix $m_i \times m_i$. Therefore, the log-likelihood (5) takes the form:

$$\log f(y \mid a, \theta, \sigma_\varepsilon^2) = \sum_{i=1}^{n} \log \left\{ \int_{\Xi_{\beta_i}} \left[\prod_{j=1}^{m_i} \frac{\exp\{-(y_{ij} - (t_{ij}/\beta_i))^2 / (2\sigma_\varepsilon^2)\}}{\sqrt{2\pi}\sigma_\varepsilon} \right] f(\beta_i \mid \theta) d\beta_i \right\} \qquad (8)$$

where $\quad f(\beta_i \mid \theta) = f(\beta_i \mid \alpha, \delta) = \frac{\alpha}{\delta} \left(\frac{\beta_i}{\delta} \right)^{\alpha\text{-}1} \exp\left\{ -\left(\frac{\beta_i}{\delta} \right)^{\alpha} \right\}$.

In order to calculate the integrals in (8) we use the PIT method. Therefore the log-likelihood (8) is re-written as:

$$\log f = \sum_{i=1}^{n} \log \left\{ \int_{\Xi_{\eta_i}} \left[\prod_{j=1}^{m_i} \frac{\exp\{-[y_{ij} - (t_{ij}/\delta(-\log(1-\Phi(\eta_i)))^{(1/\alpha)})]^2 / (2\sigma_\varepsilon^2)\}}{\sqrt{2\pi}\sigma_\varepsilon} \right] \Phi(\eta_i) d\eta_i \right\} \qquad (9)$$

where $\Phi(\eta_i)$ and $\phi(\eta_i)$ are the standard normal cumulative (CDF) and probability density functions (PDF) respectively; $F^{-1}(\Phi(\eta_i) \mid \theta) = \delta(-\log(1-\Phi(\eta_i)))^{1/\alpha}$ is the Weibull inverse CDF evaluated at ; and $\Xi_{\eta_i} = (-\infty, \infty)$ defines the limits of integration. Note that the integrand in the log-likelihood (9) is multiplied by $\phi(\eta_i)$, consequently the significant mass points are concentrated in the rage (-3.5; 3,5). We use the optimization algorithm based on the adaptative barrier (Lang, 2010) and the software R (version.14.1) in all calculations.

5. Motivating situation revisited

We need to estimate the distribution of the *time to failure* –the amount of time (in Km) it takes to achieve a degradation threshold level of $D_f = 77mm$. In the analysis of the same data set, Freitas et.al (2009) came up to the conclusion that the random effects normality assumption was inadequate. The authors based their conclusion on probability plots constructed for the pseudo-failure times calculated via the approximate method of analysis (Meeker and Escobar, 1998). Weibull and lognormal probability plots indicated that both distributions were suitable to describe the wheels failure times. Since linear degradation path models like the one postulated in (6) were used, it was possible to conclude that Weibull and lognormal wer also appropriate to describe the random effects distribution. Therefore they used these two distributions in (1) a Failure Time Analysis (FTA), performed on the censored and observed failure times) and (2) an approximate analysis, performed on the pseudo failure times. A third analysis was implemented using a LME (linear mixed effects model-LME). In that case, parameters estimates were obtained by the maximum likelihood method and the traditional random effects normality assumption.

Here, we fit the Weibull distributed random effects degradation path model (6) to the wheel degradation data using the PIT transformation, as briefly described in Section 3 and 4 (LME+PIT). In addition we compare the results to the ones obtained previously by Freitas et al. (2009) for the Weibull case. The results are summarized in Table 1.

Table 1 . Interval and point estimates obtained by each method

Method	Distribution [a]	Estimates $(\times 10^3 Km)$		
		MTTF	$t_{0.10}$	$t_{0.50}$
LME+PIT	Weibull	1062.14 [789;1362] [b]	383.78 [210;694]	995.42 [714;1321]
Approximate	Weibull	1060.88 [804; 1400] [c]	383.35 [208; 707]	994.25 [727; 1359]
LME	Normal	1914.15 [850;9913] [c]	400.00 [311; 534]	701.33 [540; 980]
Failure Time Analysis(FTA)	Weibull	971.77 [422;2239] [c]	471.60 [228;766]	937.13 [459;1912]

a: random effects distribution; b: 95% C.I – parametric Bootstrap;
c: 95% CI – asymptotic normality ML estimators.

The main observations from Table 1 are:

1. The point estimates obtained through the maximization of the transformed likelihood (i.e., LME model + PIT transformation) are very similar to the ones obtained with the approximate method. On the other hand, the confidence intervals constructed with the former have smaller widths than the ones provided by the latter, suggesting a higher precision of the estimates obtained with the method here implemented. This was already expected. The fitted model is very simple (a straight line) and the pseudo-failure times obtained through least squares estimation were well explained by a Weibull distribution.
2. The point estimates obtained with the random effects normality assumption (LME-normal) and the ones obtained with the other two methods, namely, the LME-PIT and the approximate are very different, with the mean covered distance (MTTF) being the worst case. On the other hand, the confidence intervals provided by the LME-normal fit have the smallest width among all the other methods used (including the FTA). This can be explained by the fact that the main effect of the misspecification of the random effects distribution is on the bias of the estimates, not on their precision.
3. Finally, as it was expected, the confidence intervals based on the traditional FTA have larger widths than the ones provided by the other three approaches discussed here. This is probably a consequence of the large number of censoring observations. Only 3 out of 14 units achieved the threshold level during the observation period.

6. Concluding remarks and future work

In this paper we show how the Probability Integral Transformation can be used as a estimation method for mixed models with non-normal random effects, In particular, we implemented the PIT method to fit a linear degradation path model with Weibull random effects, This model is used to analyze the wheel degradation data described originally by Freitas et al. (2009). We highlight some of the main points next:

1. The results obtained with the LME+PTT and the approximate method were quite similar. This is due to the simplicity of the degradation path model postulated for the degradation data (a straight line). The approximate method first fits the straight line models by least squares to each one of the wheels profiles independently. Then, each fitted curve is used to find *the pseudo-failure time*, the time when the fitted profile crosses the threshold level. A traditional failure time analysis is then performed on those pseudo failure times as if they were actually the *observed* failure times. For the particular case under study, the level of extrapolation was low and the pseudo failure times were well explained by a Weibull distribution. Although the point estimates of those two methods are similar, the PIT method provided estimates with higher precision. It should be emphasized that when the approximate method is used, the prediction error (due to the estimation of the pseudo failure times) is not taken into account since the pseudo failure times as used as if they were the observed data. Therefore, the confidence intervals shown in Table 1 should be in fact wider them the ones provided.
2. Even for this simple model, we can see the impact of the misspecification of the random effects distribution on the parameter estimates. The point estimates obtained using the LME-normal, in other words, ML estimation using the normality assumption of the random effects are quite different from the ones obtained with the LME+PIT and the approximate method.
3. Since the PIT method can be conveniently implemented in standard software (e.g., R, SAS, S-PLUS, Matlab), it can be used with more complex degradation path models where the normality assumption cannot be used.
4. The degradation path model used for this data set was quite simple. Consequently, it was possible to verify the inadequacy of the normality assumption and identify appropriate candidates. Unfortunately, in most of the practical problems, this assumption is hard to verify, especially in cases for which the profile equation has more than one random effect (in other words, the random effect is actually a vector valued random variable).

There are still important questions that should be addressed in future work. I would be interesting to implement also the Likelihood Reformulation Method (LR) (Liu and Yu, 2008) and study both PIT and LR though a simulation study, using also vector valued random effects. Other distributions should also be studied. In addition, it is important to study the performance of the PIT method in the case of vector valued random effects. As far as the real data is concerned, it is important to model and understand the effect of the position on the wheel degradation mechanism. Therefore, it is necessary to model all the wheels jointly.

Acknowledgments

The authors are grateful to CNPQ/Brazil, CAPES/Brazil and Fapemig/MG/Brasil for the financial support of this research.

References

Davidian, M. and Gallant, A. (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, 80:475–488.

Freitas, M. A., Toledo, M. L. G., Colosimo, E. A., and Pires, M. C. (2009). Using degradation data to assess reliability: A case study on train wheel degradation. *Quality and Reliability Engineering International*, 25:607–629.

Genest, C. and Rivest, L. (2001). On the multivariate probability integral transformation. *Statistics and Probability Letters*, pages 391–399.

Lang, K. (2010). *Numerical Analysis for Statisticians*. Springer, New York, NY.

Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: An example. *Applied Statistics*, 50:325–335. 14

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46:673–687.

Liu, L. and Yu, Z. (2008). A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine*, 27:3105–3124.

Lu, C. J. and Meeker, W. Q. (1993). Using degradation measurements to estimate a time-to-failure distribution. *Technometrics*, 35:161–174.

Meeker, W.Q, Escobar, L.A. (1998). Statistical Methods for Reliability Data. Wiley, New York, NY.

Molengerghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer-Verlag, New York.

Nelson, K., Lipsitz, S. R., Fitzmaurice, G., Ibrahim, J., Parzen, M., and Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with Non-normal random effects. *Journal of Computational and Graphical Statistics*, 15(1):89–99.

Pinheiro, J. C. and Bates, D. C. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J. of Computational and Graphical Statistics*, 1(4):12–35.

Pinheiro, J. C. and Bates, D. C. (2000). *Mixed effects models in S and S-Plus*. Springer, New York, 1st edition.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association*, 92:217–221.

Verbeke, G. and Molengerghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York, 2nd. edition.

Whitmore, G. A. and Shenkelberg, F. (1997). Modelling accelerated degradation data using wiener diffusion with a time scale transformation. *Lifetime Data Analysis*, 3:27–45.