Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v9n2p371

Chi-square tests for generalized exponential AFT distributions with censored data

By Aidi, Seddik-Ameur

Published: 14 October 2016

Electronic Journal of Applied Statistical Analysis
Vol. 09, Issue 02, October 2016, 371-384
DOI: 10.1285/i20705948v9n2p371

Chi-square tests for generalized exponential AFT distributions with censored data

K. Aidi[*] and N. Seddik-Ameur

Laboratory of probability and statistics LaPS
Badji mokhtar university Annaba - Algeria

Published: 14 October 2016

Generalized exponential models have numerous applications particularly in reliability studies. Using the approach proposed by Bagdonavicius and Nikulin for censored data, we propose the construction of modified chi-square goodness-of-fit tests for the generalized exponentiated exponential model (GEE) and an accelerated failure time model with the generalized exponentiated exponential distribution as the baseline ($AFT - GEE$). Based on maximum likelihood estimators on initial data, these statistics recover the information lost while grouping data and follow chi-square distributions. The elements of the criteria tests are given explicitly. Numerical examples from simulated samples and real data have been presented to illustrate the feasibility of the proposed tests.

keywords: Accelerated failure time models- Chi-square test- Maximum likelihood estimation- Reliability.

1 Introduction

Testing the fit of parametric models to experimental data is a crucial problem because the choice of the hypothesized model affect seriously the results of the analysis and the conclusions may be invalid if the model chosen is not appropriate. At that end, several techniques may be used. However, in presence of censorship, classical procedures become more complex and are not sufficiently developed. Modifications of some existing tests

[*]Corresponding author: khaoula.aidi@yahoo.fr

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

were proposed like chi-square tests based on the difference between Kaplan-Meyer and maximum likelihood estimators of the cumulative distribution function (Habib et al., 1986), or the differences between observed and expected numbers of failure in time intervals (Hjort, 1990; Hollander and Pena, 1992).

Using the approach proposed by Bagdonavicius and Nikulin (2011), we construct modified chi-square goodness-of-fit tests to assess the adequacy of generalized exponentiated exponential distributions (GEE) and the corresponding accelerated failure time model ($AFT-GEE$), when data are right censored and the parameters are unknown. Based on maximum likelihood estimators on initial data, these test statistics recover information lost while grouping data and follow chi-square distributions.

The first test concerns the GEE distribution proposed by Gupta and Kundu (1999). This distribution which has a great interest in simulations, is more flexible than gamma, Weibull and log-normal distributions (Gupta and Kundu, 2007; Kundu and Gupta, 2007) and attracts the interest of researchers till now. For complete samples, and using Monte Carlo simulations and Pearson system techniques, Hassan (2005) created tables of goodness-of-fit critical values for this distribution for the classical statistics. Rao (2012) considered the estimation of reliability in multicomponent stress-strength model when variates are given by this distribution with different shape parameters; and Achcar et al. (2015) developed the bayesian estimation using MCMC methods. Goodness-of-fit tests for censored data are not investigated except likelihood ratio against alternative hypothesis.

The second test concerns the corresponding accelerated failure time model $AFT-GEE$, where the baseline is a GEE distribution. As we know, the lifetimes of the devices under normal use conditions are often very long, so life testing is the best option for engineers to perfect the reliability of their products and to eliminate the causes of failure in a short time. Accelerated failure time (AFT) models which relate survival times to covariates are the most appropriated in these cases and the commonly used statistical tests are the likelihood ratio, the Wald and score tests. In censored data case, Bagdonavičius et al. (2013) constructed modified chi-square tests for Weibull, loglogistic, lognormal AFT models. Galanova et al. (2012) developed modifications of the wellknown statistics Kolmogorov, Cramer-Von Mises-Smirnov, and Anderson-Darling statistics which take into account the unknown parameters and censorship for exponential, gamma, Weibull, Lognormal and generalized Weibull AFT models. These statistics are based on Kaplan-Meyer estimate instead of the empirical distribution function. Also, Balakrishnan et al. (2013) considered statistic tests based on residuals and investigated the statistical distributions of the classical test statistics for various models. Goual and Seddik-Ameur (2014) constructed a modified chi-square test for the generalized inverse Weibull aft model. Recently, Medeiros et al. (2014) evaluated the performance of the gradient test in comparison with the likelihood ratio test in Weibull, log-normal and log-logistic AFT models.

Here, we consider an $AFT-GEE$ model which can be very interesting to improve the performance of real systems. Maximum likelihood estimation of the unknown parameters and fitting test when data are right censored are developed and the results are applied on simulated and reliability data sets.

2 Statistic test for right censored data

Let $T_1, ..., T_n$ be n i.i.d. random variables grouped into k classes I_j. To assess the adequacy of a parametric model F_0

$H_0 : \quad P(T_i \le t \mid H_0) = F_0(t;\theta), t \ge 0, \quad \theta = (\theta_1, ..., \theta_s)^T \in \Theta \subset R^s$

when data are right censored and the parameter vector θ is unknown, Bagdonavicius and Nikulin (2011) proposed a statistic test Y^2 based on the vector

$$Z_j = \frac{1}{\sqrt{n}}(U_j - e_j)\ , \qquad j = 1, 2, ..., k \quad , \text{ with } k \succ s.$$

This one represents the differences between observed and expected numbers of failures (U_j and e_j) to fall into these grouping intervals $I_j = (a_{j-1}, a_j]$ with $a_0 = 0$, $a_k = \tau$, where τ is a finite time. The authors considered a_j as random data functions such as the k intervals chosen have equal expected numbers of failures e_j.

The statistic test Y^2 is defined by

$$Y^2 = Z^T \widehat{\Sigma}^- Z \ = \sum_{j=1}^{k} \frac{(U_j - e_j)^2}{U_j} + Q$$

where $\widehat{\Sigma}^-$ is, if necessary, a generalized inverse of the matrix Σ, and $Z = (Z_1, ..., Z_k)^T$.

$$\begin{aligned}
Q &= W^T \widehat{G}^- W & \widehat{A}_j &= U_j/n, & U_j &= \sum_{i:t_i \in I_j} \delta_i, \\
W &= (W_1, ..., W_s)^T, \quad \widehat{G} = [\widehat{g}_{ll'}]_{s\times s}, & \widehat{g}_{ll'} &= \widehat{i}_{ll'} - \sum_{j=1}^{k} \widehat{C}_{lj} \widehat{C}_{l'j} \widehat{A}_j^{-1}, \\
\widehat{C}_{lj} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \frac{\partial}{\partial \theta} \ln h(t_i, \widehat{\theta}), & \widehat{i}_{ll'} &= \frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{\partial \ln h(t_i, \widehat{\theta})}{\partial \theta_l} \frac{\partial \ln h(t_i, \widehat{\theta})}{\partial \theta_{l'}}, \\
\widehat{W}_l &= \sum_{j=1}^{k} \widehat{C}_{lj} \widehat{A}_j^{-1} Z_j, \qquad l, l' = 1, ..., s
\end{aligned}$$

and $\widehat{\theta}$ is the maximum likelihood estimator of θ on initial non-grouped data.

Under the null hypothesis H_0 , the limit distribution of the statistic Y^2 is a chi-square with $r = rank(\Sigma)$ degrees of freedom. The description and applications of modified chi-square tests are discussed in Balakrishnan et al. (2013).

3 Goodness-of-fit test for generalized exponentiated exponential distribution GEE

Because of its several applications, the generalized exponentiated exponential model GEE proposed by Gupta and Kundu (1999) is extensively studied by the authors and others. Nevertheless, goodness-of-fit tests have not been sufficiently investigated. In this section, we construct a modified chi-square test based on the Y^2 statistic, defined above, for the GEE distribution with density and hazard and functions

$$f(t;\theta) = \alpha\lambda(1-e^{-\lambda t})^{\alpha-1}e^{-\lambda t} \quad ; \; t \geq 0$$

$$h(t;\theta) = \frac{\alpha\lambda(1-e^{-\lambda t})^{\alpha-1}e^{-\lambda t}}{1-(1-e^{-\lambda t})^{\alpha}}$$

where $\alpha > 0$ and $\lambda > 0$, are the shape and the scale parameters.

We consider $T = (T_1, T_2, ..., T_n)^T$ an i.i.d. random variables from GEE distribution. And suppose that we observe the right censored sample $t = (t_1, t_2, ..., t_n)^T$ with fixed censoring time τ. Each t_i can be written as $t_i = (T_i, \delta_i)$ where T_i is a failure time and δ_i a censoring indicator.

As random grouping intervals are considered such as expected failures to get into each interval $I_j = (a_{j-1}, a_j]$ must be the same for any j, so the estimated limit intervals $\widehat{a}_j$, $(0 < \widehat{a}_1 < \widehat{a}_2 < ... < \widehat{a}_k = \tau)$ are obtained as

$$\hat{a}_j = -\frac{1}{\lambda}\ln\left(1-\left[1-\exp\left(\frac{\sum_{l=1}^{i-1} H\left(t_l,\hat{\theta}\right)-E_j}{n-i+1}\right)\right]^{\frac{1}{\alpha}}\right) \quad ; \; j = 1,..,k-1, \quad \hat{a}_k = t_{(n)}$$

where H represents the GEE cumulative hazard function and $\hat{\theta} = \left(\hat{\alpha},\hat{\lambda}\right)^T$, the maximum likelihood estimator of the unknown parameter vector θ.

3.1 Estimated matrix $\hat{W}$

We first compute the elements of the estimated matrix $\hat{C} = (C_{ij})_{2\times k}$

$$\begin{aligned}
\hat{C}_{1j} &= \frac{1}{n}\sum_{i:t_i\in I_j}^{n}\delta_i\left[\frac{1}{\hat{\alpha}}+\frac{\ln(1-e^{-\hat{\lambda}t_i})}{1-(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}}\right] \\
\hat{C}_{2j} &= \frac{1}{n}\sum_{i:t_i\in I_j}^{n}\delta_i\left[\frac{1}{\hat{\lambda}}+\frac{t_i\left[\hat{\alpha}e^{-\hat{\lambda}t_i}+(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}-1\right]}{(1-e^{-\hat{\lambda}t_i})\left(1-(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}\right)}\right]
\end{aligned}$$

and the matrix $\hat{W}$ can be deduced

$$\hat{W}_l = \sum_{j=1}^{r}\hat{C}_{lj}A_j^{-1}Z_j, \quad l = 1,2 \qquad j = 1,..,k$$

3.2 Estimated matrix $\hat{G}$

The estimated matrix $\hat{G} = [\hat{g}_{ll'}]_{2\times 2}$, is derived from the estimated information matrix $\widehat{I}$ and the estimated matrix $\hat{C}$.

$$\hat{g}_{11} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left[\frac{1}{\hat{\alpha}}+\frac{\ln(1-e^{-\hat{\lambda}t_i})}{1-(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}}\right]^2 - \frac{A_j^{-1}}{n^2}\sum_{j=1}^{k}\left[\sum_{i:t_i\in I_j}^{n}\delta_i\left(\frac{1}{\hat{\alpha}}+\frac{\ln(1-e^{-\hat{\lambda}t_i})}{1-(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}}\right)\right]^2$$

$$\hat{g}_{12} = \hat{g}_{21}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\frac{1}{\hat{\alpha}}+\frac{\ln(1-e^{-\hat{\lambda}t_i})}{1-(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}}\right)\left(\frac{1}{\hat{\lambda}}+\frac{t_i\left[\hat{\alpha}e^{-\hat{\lambda}t_i}+(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}-1\right]}{(1-e^{-\hat{\lambda}t_i})\left(1-(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}\right)}\right)$$
$$-\frac{A_j^{-1}}{n^2}\sum_{j=1}^{k}\left[\sum_{i:t_i\in I_j}^{n}\delta_i\left(\frac{1}{\hat{\alpha}}+\frac{\ln(1-e^{-\hat{\lambda}t_i})}{1-(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}}\right)\right]\left[\sum_{i:t_i\in I_j}^{n}\delta_i\left(\frac{1}{\hat{\lambda}}+\frac{t_i\left[\hat{\alpha}e^{-\hat{\lambda}t_i}+(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}-1\right]}{(1-e^{-\hat{\lambda}t_i})\left(1-(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}\right)}\right)\right]$$

$$\hat{g}_{22} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left[\frac{1}{\hat{\lambda}}+\frac{t_i\left[\hat{\alpha}e^{-\hat{\lambda}t_i}+(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}-1\right]}{(1-e^{-\hat{\lambda}t_i})\left(1-(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}\right)}\right]^2$$
$$-\frac{A_j^{-1}}{n^2}\sum_{j=1}^{k}\left[\sum_{i:t_i\in I_j}^{n}\delta_i\left(\frac{1}{\hat{\lambda}}+\frac{t_i\left[\hat{\alpha}e^{-\hat{\lambda}t_i}+(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}-1\right]}{(1-e^{-\hat{\lambda}t_i})\left(1-(1-e^{-\hat{\lambda}t_i})^{\hat{\alpha}}\right)}\right)\right]^2$$

Therefore, we obtain the explicit form of the criteria test Y_n^2 to fit data from a GEE distribution when data are right censored

$$Y_n^2 = \sum_{j=1}^{k}\frac{(U_j-e_j)^2}{U_j} + \widehat{W}^T\hat{G}^{-}\widehat{W}$$

4 Construction of the $AFT-GEE$ model

An accelerated failure time model (AFT) with a baseline distribution $F_0(t)$ is defined from its survival function by

$$S(t) = S_0\left(te^{-\beta^T z}\right)$$

where $z = (z_0, z_1, ..., z_m)^T$ is the vector of the covariables related to the different stresses and $\beta = (\beta_0, \beta_1, ..., \beta_m)^T$ their coefficients, so the survival function of the $AFT-GEE$ distribution is obtained as follows:

$$S(t) = 1 - \left(1 - \exp\left\{-\lambda t e^{-\beta^T z}\right\}\right)^{\alpha}$$

and its pdf and hazard functions are deduced

$$f(t) = \alpha\lambda e^{-\beta^T z} \exp\left\{-\lambda t e^{-\beta^T z}\right\} \left(1 - \exp\left\{-\lambda t e^{-\beta^T z}\right\}\right)^{\alpha-1}$$

$$h(t) = \frac{\alpha\lambda e^{-\beta^T z} \exp\left\{-\lambda t e^{-\beta^T z}\right\} \left(1 - \exp\left\{-\lambda t e^{-\beta^T z}\right\}\right)^{\alpha-1}}{1 - \left(1 - \exp\left\{-\lambda t e^{-\beta^T z}\right\}\right)^{\alpha}}$$

Depending on the values of the shape parameter, this distribution can be unimodal or decreasing and the rate failure can be increasing, constant or decreasing which enable it to be used to describe the reliability of many real systems.

4.1 Maximum likelihood estimation in censored data case

Let us consider $T = (T_1, T_2, ..., T_n)^T$ a sample from an $AFT - GEE$ distribution with the parameter vector $\theta = (\alpha, \lambda, \beta_0, \beta_1, ..., \beta_m)^T$ with right censored data.

The log-likelihood function can be written as

$$\begin{aligned} L(t,\theta) \quad &= \quad \sum_{i=1}^{n} \delta_i [\ln(\alpha\lambda) - \beta^T z_i - t_i u_i + (\alpha - 1)\ln\left(1 - e^{-t_i u_i}\right) \\ &\quad - \ln\left(1 - \left(1 - e^{-t_i u_i}\right)^{\alpha}\right)] + \sum_{i=1}^{n} \ln\left(1 - \left(1 - e^{-t_i u_i}\right)^{\alpha}\right) \end{aligned}$$

where $u_i = \lambda e^{-\beta^T z_i}$

The maximum likelihood estimators $\widehat{\alpha}$, $\widehat{\lambda}$, $\widehat{\beta}_0$, $\widehat{\beta}_1$, ..., $\widehat{\beta}_m$ of the unknown parameters α, λ, β_0, $\beta_1, ..., \beta_m$ are obtained by equaling the following score equations to zero.

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{n} \delta_i \left[\frac{1}{\alpha} + \frac{\ln\left(1 - e^{-t_i u_i}\right)}{1 - (1 - e^{-t_i u_i})^{\alpha}}\right] - \sum_{i=1}^{n} \frac{\left(1 - e^{-t_i u_i}\right)^{\alpha} \ln\left(1 - e^{-t_i u_i}\right)}{1 - (1 - e^{-t_i u_i})^{\alpha}}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^{n} \delta_i \left[\frac{1}{\lambda} - \frac{t_i e^{-\beta^T z_i}\left[1 - \alpha e^{-t_i u_i} - \left(1 - e^{-t_i u_i}\right)^{\alpha}\right]}{\left(1 - e^{-t_i u_i}\right)\left(1 - \left(1 - e^{-t_i u_i}\right)^{\alpha}\right)}\right] - \sum_{i=1}^{n} \frac{\alpha t_i e^{-\beta^T z_i} \exp\left\{-t_i u_i\right\} \left(1 - e^{-t_i u_i}\right)^{\alpha-1}}{1 - \left(1 - e^{-t_i u_i}\right)^{\alpha}}$$

$$\frac{\partial L}{\partial \beta_0} = \sum_{i=1}^{n} \delta_i \left[-1 + \frac{t_i u_i \left[1 - \alpha e^{-t_i u_i} - \left(1 - e^{-t_i u_i}\right)^{\alpha}\right]}{\left(1 - e^{-t_i u_i}\right)\left(1 - \left(1 - e^{-t_i u_i}\right)^{\alpha}\right)}\right] + \sum_{i=1}^{n} \frac{\alpha t_i u_i e^{-t_i u_i} \left(1 - e^{-t_i u_i}\right)^{\alpha-1}}{1 - \left(1 - e^{-t_i u_i}\right)^{\alpha}}$$

$$\frac{\partial L}{\partial \beta_1} = \sum_{i=1}^{n} \delta_i \left[-z_{i1} + \frac{t_i z_{i1} u_i \left[1 - \alpha e^{-t_i u_i} - \left(1 - e^{-t_i u_i}\right)^{\alpha}\right]}{\left(1 - e^{-t_i u_i}\right)\left(1 - \left(1 - e^{-t_i u_i}\right)^{\alpha}\right)} \right] + \sum_{i=1}^{n} \frac{\alpha t_i z_{i1} u_i e^{-t_i u_i} \left(1 - e^{-t_i u_i}\right)^{\alpha-1}}{1 - \left(1 - e^{-t_i u_i}\right)^{\alpha}}$$

$$\frac{\partial L}{\partial \beta_m} = \sum_{i=1}^{n} \delta_i \left[-z_{im} + \frac{t_i z_{im} u_i \left[1 - \alpha e^{-t_i u_i} - \left(1 - e^{-t_i u_i}\right)^{\alpha}\right]}{\left(1 - e^{-t_i u_i}\right)\left(1 - \left(1 - e^{-t_i u_i}\right)^{\alpha}\right)} \right] + \sum_{i=1}^{n} \frac{\alpha t_i z_{im} u_i e^{-t_i u_i} \left(1 - e^{-t_i u_i}\right)^{\alpha-1}}{1 - \left(1 - e^{-t_i u_i}\right)^{\alpha}}$$

Their analytical forms cannot be obtained, so we use iterative methods like Newton Raphson method.

5 Goodness-of-fit test for the $AFT - GEE$ model in censored data case

For testing the hypothesis H_0 that the distribution of $T = (T_1, ..., T_n)^T$ belongs to the parametric model $AFT - GEE$ defined above, we construct the statistic test Y^2. For this, we group data into k intervals I_j, such as the numbers of theoretical failure times e_j are the same, so the limits a_j of the grouping intervals are obtained as follows:

$$\hat{a}_j = -\frac{1}{\hat{\lambda}} e^{\hat{\beta}^T z} \ln \left(1 - \left[1 - \exp\left(\frac{\sum_{l=1}^{i-1} H(t_i, \hat{\theta}) - E_j}{n - i + 1} \right) \right]^{\frac{1}{\hat{\alpha}}} \right)$$

where

$$E_j = \frac{j}{k-1} \sum_{i=1}^{n} H(t_i, \hat{\theta}) = -\frac{j}{k-1} \sum_{i=1}^{n} \ln \left(1 - \left(1 - \exp\left\{ -\hat{\lambda} t_i e^{-\hat{\beta}^T z_i} \right\} \right)^{\hat{\alpha}} \right), \quad j = 1, ..k-1$$

With this choice $e_j = \frac{E_k}{k}$ for any $j = 1, ..., k$ and $E_k = \sum_{i=1}^{n} H\left(t_i, \widehat{\theta}\right)$.

5.1 Estimated matrix $\hat{C}$

To obtain the estimated vector $\hat{W}$, we must compute the elements of the estimated matrix $\hat{C}$:

$$\hat{C}_{1j} = \frac{1}{n} \sum_{i:t_i \in I_j}^{n} \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\ln\left(1 - e^{-t_i u_i}\right)}{1 - \left(1 - e^{-t_i u_i}\right)^{\hat{\alpha}}} \right]$$

$$\hat{C}_{2j} = \frac{1}{n} \sum_{i:t_i \in I_j}^{n} \delta_i \left[\frac{1}{\hat{\lambda}} - \frac{t_i e^{-\hat{\beta}^T z_i} \left[1 - \hat{\alpha} e^{-t_i u_i} - \left(1 - e^{-t_i u_i}\right)^{\hat{\alpha}}\right]}{\left(1 - e^{-t_i u_i}\right)\left(1 - \left(1 - e^{-t_i u_i}\right)^{\hat{\alpha}}\right)} \right]$$

$$\hat{C}_{3j} = \frac{1}{n}\sum_{i:t_i\in I_j}^{n}\delta_i\left[-1+\frac{t_iu_i\left[1-\hat{\alpha}e^{-t_iu_i}-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right]$$

$$\hat{C}_{4j} = \frac{1}{n}\sum_{i:t_i\in I_j}^{n}\delta_i\left[-z_{i1}+\frac{t_iz_{i1}u_i\left[1-\hat{\alpha}e^{-t_iu_i}-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right]$$

$$\hat{C}_{mj} = \frac{1}{n}\sum_{i:t_i\in I_j}^{n}\delta_i\left[-z_{im}+\frac{t_iz_{im}u_i\left[1-\hat{\alpha}e^{-t_iu_i}-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right]$$

5.2 Estimated matrix $\hat{I}$

After simplifications, the components of the symetric estimated matrix $\hat{I}$ defined by

$$\widehat{i}_{ll'} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{\partial\ln h(t_i,\widehat{\theta})}{\partial\theta_l}\frac{\partial\ln h(t_i,\widehat{\theta})}{\partial\theta_{l'}}$$

are obtained as follows:

$$\hat{i}_{11} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\frac{1}{\hat{\alpha}}+\frac{\ln\left(1-e^{-t_iu_i}\right)}{1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}}\right)^2$$

$$\hat{i}_{22} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\frac{1}{\hat{\lambda}}-\frac{t_ie^{-\hat{\beta}^Tz_i}\left[1-\hat{\alpha}e^{-t_iu_i}-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)^2$$

$$\hat{i}_{33} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(-1+\frac{t_iu_i\left[1-\hat{\alpha}e^{-t_iu_i}-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)^2$$

$$\hat{i}_{44} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(-z_{i1}+\frac{t_iz_{i1}u_i\left[1-\hat{\alpha}e^{-t_iu_i}-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)^2$$

$$\hat{i}_{12} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\frac{1}{\hat{\alpha}}+\frac{\ln\left(1-e^{-t_iu_i}\right)}{1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}}\right)\left(\frac{1}{\hat{\lambda}}-\frac{t_ie^{-\hat{\beta}^Tz_i}\left[1-\hat{\alpha}e^{-t_iu_i}-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)$$

$$\hat{\imath}_{13} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\frac{1}{\hat{\alpha}} + \frac{\ln\left(1-e^{-t_iu_i}\right)}{1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}}\right)\left(-1 + \frac{t_iu_i\left[1-\hat{\alpha}e^{-t_iu_i} - \left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)$$

$$\hat{\imath}_{14} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\frac{1}{\hat{\alpha}} + \frac{\ln\left(1-e^{-t_iu_i}\right)}{1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}}\right)\left(-z_{i1} + \frac{t_iz_{i1}u_i\left[1-\hat{\alpha}e^{-t_iu_i} - \left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)$$

$$\hat{\imath}_{23} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\frac{1}{\hat{\lambda}} - \frac{t_ie^{-\hat{\beta}^Tz_i}\left[1-\hat{\alpha}e^{-t_iu_i} - \left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)\left(-1 + \frac{t_iu_i\left[1-\hat{\alpha}e^{-t_iu_i} - \left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)$$

$$\hat{\imath}_{24} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(\frac{1}{\hat{\lambda}} - \frac{t_ie^{-\hat{\beta}^Tz_i}\left[1-\hat{\alpha}e^{-t_iu_i} - \left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)\left(-z_{i1} + \frac{t_iz_{i1}u_i\left[1-\hat{\alpha}e^{-t_iu_i} - \left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)$$

$$\hat{\imath}_{34} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left(-1 + \frac{t_iu_i\left[1-\hat{\alpha}e^{-t_iu_i} - \left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)\left(-z_{i1} + \frac{t_iz_{i1}u_i\left[1-\hat{\alpha}e^{-t_iu_i} - \left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right]}{\left(1-e^{-t_iu_i}\right)\left(1-\left(1-e^{-t_iu_i}\right)^{\hat{\alpha}}\right)}\right)$$

Therefore the quadratic form $\widehat{Q}$ is obtained and the test statistic can be deduced easily:

$$Y_n^2\left(\hat{\theta}\right) = \sum_{j=1}^{k}\frac{\left(U_j-e_j\right)^2}{U_j} + \hat{W}^T\left[\hat{\imath}_{ll'} - \sum_{j=1}^{k}\hat{C}_{lj}\hat{C}_{l'j}\hat{A}_j^{-1}\right]^{-1}\hat{W}$$

6 Simulations

In this section, an important simulation study is carried out to show the feasability of the proposed tests. Data from GEE distribution with $(\alpha = 2, \lambda = 0.8)$ and $AFT-GEE$ model with the following values of the parameters $\alpha = 3$, $\lambda = 2$, $\beta_0 = 0.5, \beta_1 = 0.8$ are generated $M = 10\,000$ times with sample sizes $n_1 = 15$, $n_2 = 25$, $n_3 = 50$, $n_4 = 130$, $n_5 = 350$ and $n_6 = 500$. Using the Newton Raphson method, the values of the mean simulated MLEs $\hat{\alpha}, \hat{\lambda}, \hat{\beta}_0, \hat{\beta}_1$ of the $AFT - GEE$ parameters and their mean square error are given in table 1. In the second time, and using R software, we calculate the values of the criteria test Y^2 for the two models. Tables 2 and 3 give the empirical levels of significance comparing to the corresponding theoretical levels of significance γ $(\gamma = 0.01, \gamma = 0.05, \gamma = 0.1)$. Note that the MLes of the GEE parameters are not reported here.

$M = 10\ 000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha}$	2.8854	2.8985	2.9254	2.9523	2.9785	3.0026
$S.M.E$	0.0040	0.0032	0.0028	0.0021	0.0016	0.0004
$\hat{\lambda}$	1.8654	1.8925	1.9625	1.9745	1.9894	1.9998
$S.M.E$	0.0035	0.0029	0.0025	0.0019	0.0013	0.0008
$\hat{\beta}_0$	0.5721	0.5654	0.5425	0.5254	0.5102	0.5019
$S.M.E$	0.0045	0.0039	0.0026	0.0019	0.0012	0.0006
$\hat{\beta}_1$	0.8423	0.8325	0.8214	0.8154	0.8045	0.8012
$S.M.E$	0.0037	0.0025	0.0017	0.0012	0.0009	0.0005

Table 1. Mean simulated values of MLEs $\hat{\alpha}, \hat{\lambda}, \hat{\beta}_0, \hat{\beta}_1$ and their corresponding square mean errors

$M = 10\ 000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\gamma = 1\%$	0.0034	0.0045	0.0054	0.0061	0.0078	0.0089
$\gamma = 5\%$	0.0159	0.0298	0.0312	0.0421	0.0453	0.0463
$\gamma = 10\%$	0.0236	0.0320	0.0371	0.0442	0.0489	0.0521

Table 2. Simulated levels of significance for Y^2 test for $AFT - GEE$ model against their theoretical values ($\gamma = 0.01, \gamma = 0.05, \gamma = 0.1$)

$M = 10\ 000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\gamma = 1\%$	0.0023	0.0054	0.0084	0.0091	0.0112	0.0128
$\gamma = 5\%$	0.0034	0.0072	0.0097	0.0108	0.0152	0.0163
$\gamma = 10\%$	0.0073	0.0120	0.0139	0.0260	0.0282	0.0319

Table 3. Simulated levels of significance for Y^2 test for GEE model against their theoretical values ($\gamma = 0.01, \gamma = 0.05, \gamma = 0.1$)

The maximum likelihood estimated parameter values, presented in Table 1, agree closely with the true parameter values. The values of the criteria test Y^2 of the proposed models $AFT - GEE$ and GEE, obtained for different simulated levels of significance (Table 2, Table 3) give good results for all sample sizes.

7 Applications

To show the usefulness of the proposed tests, we apply our results to two real data sets from survival analysis and reliability.

Example 1:

We consider sample data of 51 patients with advanced acute myelogenous leukemia reported to the International Bone Marrow Transplant Registry. These patients had received an autologous (auto) bone marrow transplant in which, after high doses of chemotherapy, their own marrow was reinfused to replace their destroyed immune system.

Leukemia free-survival times (in months) for Autologous Transplants:

0.658, 0.822, 1.414, 2.5, 3.322, 3.816, 4.737, 4.836*, 4.934, 5.033, 5.757, 5.855, 5.987, 6.151, 6.217, 6.447*, 8.651, 8.717, 9.441*, 10.329, 11.48 , 12.007, 12.007*, 12.237, 12.401*, 13.059*, 14.474*, 15*, 15.461, 15.757, 16.48, 16.711, 17.204*, 17.237, 17.303*, 17.664*, 18.092, 18.092*, 18.75*, 20.625*, 23.158, 27.73*, 31.184*, 32.434*, 35.921*, 42.237*,44.638*, 46.48*, 47.467*, 48.322*, 56.086.

* means censored data.

We apply the test proposed above, for testing the null hypothesis H_0, for which these data fit a generalized exponentiated exponential distribution GEE. We calculate the maximum likelihood estimators $\hat{\alpha}, \hat{\lambda}$ of the unknown parameters α, λ:

$$\hat{\theta} = \left(\hat{\alpha}, \hat{\lambda}\right)^T = (0.9190, 0.03)^T$$

We choose $k = 5$ classes, the results necessary for the calcul of the criteria test Y_n^2 , are given in Table 4.

$\hat{a}_j$	3.0890	7.1303	12.5666	22.8197	56.0860
U_J	4	10	6	6	2
e_j	5.5863	5.5863	5.5863	5.5863	5.5863
$\hat{C}_{1j}$	−0.1917	−0.2469	−0.0924	−0.0628	−0.0095
$\hat{C}_{2j}$	2.4278	6.1542	3.7326	3.7633	1.2740

Table 4. values of $\widehat{a}_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}$

and the estimated matrix $\hat{G}$

$$\hat{G}_{2\times 2} = \begin{pmatrix} 0.1198 & -3.7226 \\ -3.7226 & 177.7709 \end{pmatrix}$$

the estimated vector $\hat{W}$:

$$\hat{W} = \left(\hat{W}_1, \hat{W}_2\right)^T = \begin{pmatrix} -0.1259 & 8.2881 \end{pmatrix}^T$$

So, we obtain the value of the criteria test statistic Y_n^2 :

$$Y_n^2 = X^2 + Q = 9.0652 + 0.4406 = 9.5059$$

For significance level $\gamma = 0.05$, the critical value of the chi-square distribution is $\chi_5^2 = 11.07050$. As $\chi_5^2 > Y_n^2$, so, we can say that these data can be modeled by the GEE distribution.

Example 2:

We consider a data set from Cox et al. (1984) which represents life-testing results of springs given by Mr W. Armstrong. We want to verify if these data can be modeled by an $AFT - GEE$ distribution.

Springs are tested under cycles of repeated loading and failure time is the number of cycles to failure, it being convenient to take 10 cycles as the unit of ' time '. Here 60 springs were allocated, 10 to each of six different stress levels. At the lower stress levels, where failure time is long, some springs are censored, i.e. testing is abandoned before failure has occurred.

Stress										
0.950	0.225	0.171	0.198	0.189	0.189	0.135	0.162	0.135	0.117	0.162
0.900	0.216	0.162	0.153	0.216	0.225	0.216	0.306	0.225	0.243	0.189
0.850	0.324	0.321	0.432	0.252	0.279	0.414	0.396	0.379	0.351	0.333
0.800	0.627	1.051	1.434	2.020	0.525	0.402	0.463	0.431	0.365	0.715
0.750	3.402	9.417	1.802	4.326	11.52*	7.152	2.969	3.012	1.550	11.211
0.700	12.51*	12.505*	3.027	12.505	6.253	8.011	7.795	11.604*	11.604*	12.47*

$*$ Censored

Table 5. Failure times of springs

Using Newton Raphson method, we compute the maximum likelihood estimators of the unknown parameters

$$\hat{\theta} = \left(\hat{\alpha}, \hat{\lambda}, \hat{\beta}_0, \hat{\beta}_1\right)^T = (3.0663,\ 0.0624, 11.6171, -18.1301)^T.$$

The results necessary for the calcul of the criteria test Y_n^2 , with $k = 6$ grouping intervals, are given in Table 6.

$\hat{a}_j$	0.190	0.296	0.445	1.356	6.124	12.510
U_J	11	10	12	5	9	7
e_j	9.1682	9.1682	9.1682	9.1682	9.1682	9.1682
$\hat{C}_{1j}$	−0.0169	−0.0285	−0.0752	−0.0375	−0.0362	−0.0103
$\hat{C}_{2j}$	0.9726	0.0707	0.9864	0.6638	0.9867	0.4858
$\hat{C}_{3j}$	−0.2478	−0.2540	−0.3735	−0.1662	−0.2487	−0.1551
$\hat{C}_{4j}$	−0.2362	−0.2373	−0.3418	−0.1495	−0.2121	−0.1226

Table 6. values of $\widehat{a}_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}, \hat{C}_{3j}, \hat{C}_{4j}$

So, we obtain the value of the criteria test Y_n^2

$$Y_n^2 = X^2 + Q = 5.1920 + 7.1284 = 12.3204$$

For significance level $\gamma = 0.05$, and as the value of $Y_n^2 = 12.3204$ is inferior than the critical value $\chi_6^2 = 12.5915$; so we can say that the proposed model $AFT - GEE$ fit these data.

Conclusion:

Papers related to the GEE distribution showed that this one can be used more frequently than Weibull and gamma distributions. So, observed data from reliability and survival analysis can be suitably modeled by GEE and $AFT - GEE$ distributions. We hope that the results obtained through this study will be used by practionnars in several fields.

Acknowledgment
The authors are grateful to the referees for their constructive comments and suggestions which helped improve the presentation of this paper.

References

Achcar, J., Moala, F., and Boleta, J. (2015). Generalized exponential distribution: A bayesian approach using mcmc methods. *International Journal of Industrial Engineering Computations*, 6(1):1–14.

Bagdonavicius, V. and Nikulin, M. (2011). Chi-squared goodness-of-fit test for right censored data. *International Journal of Applied Mathematics and Statistics*, 24(SI-11A):30–50.

Bagdonavičius, V. B., Levuliene, R. J., and Nikulin, M. S. (2013). Chi-squared goodness-of-fit tests for parametric accelerated failure time models. *Communications in Statistics-Theory and Methods*, 42(15):2768–2785.

Balakrishnan, N., Voinov, V., and Nikulin, M. S. (2013). *Chi-squared goodness of fit tests with applications*. Academic Press.

Cox, D., Hinkley, D., Rubin, D., and Silverman, B. (1984). *Monographs on statistics and applied probability*. Springer.

Galanova, N., Lemeshko, B. Y., and Chimitova, E. (2012). Using nonparametric goodness-of-fit tests to validate accelerated failure time models. *Optoelectronics, Instrumentation and Data Processing*, 48(6):580–592.

Goual, H. and Seddik-Ameur, N. (2014). Chi-squared type test for the aft-generalized inverse weibull distribution. *Communications in Statistics-Theory and Methods*, 43(13):2605–2617.

Gupta, R. D. and Kundu, D. (1999). Theory & methods: Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, 41(2):173–188.

Gupta, R. D. and Kundu, D. (2007). Generalized exponential distribution: Existing results and some recent developments. *Journal of Statistical Planning and Inference*, 137(11):3537–3547.

Habib, M., Thomas, D., et al. (1986). Chi-square goodness-if-fit tests for randomly censored data. *The Annals of Statistics*, 14(2):759–765.

Hassan, A. S. (2005). Goodness-of-fit for the generalized exponential distribution. *Interstat electronic Journal: http://interstat. statjournals. net/YEAR/2005/abstracts/0507001. php.*

Hjort, N. L. (1990). Goodness of fit tests in models for life history data based on cumulative hazard rates. *The Annals of Statistics*, pages 1221–1258.

Hollander, M. and Pena, E. A. (1992). A chi-squared goodness-of-fit test for randomly censored data. *Journal of the American Statistical Association*, 87(418):458–463.

Kundu, D. and Gupta, R. D. (2007). A convenient way of generating gamma random variables using generalized exponential distribution. *Computational Statistics & Data Analysis*, 51(6):2796–2802.

Medeiros, F. M., da Silva-Júnior, A. H., Valença, D. M., and Ferrari, S. L. (2014). Testing inference in accelerated failure time models. *International Journal of Statistics and Probability*, 3(2):121.

Rao, G. S. (2012). Estimation of reliability in multicomponent stress-strength based on generalized exponential distribution. *Revista Colombiana de Estadística*, 35(1):67–76.