Tests d’adéquation à la loi de Weibull : données complètes et censurées

Goodness-of-fit tests for the Weibull distribution: complete and censored data

Meryam Krit and Emmanuel Remy
EDF R&D
6 quai Watier, 78401 Chatou

Olivier Gaudoin
Université Grenoble Alpes
Laboratoire Jean Kuntzmann, 38041 Grenoble Cedex 9

Résumé
La loi de Weibull est utilisée couramment dans de nombreux domaines comme les sciences de l’environnement, la géologie, la chimie, l’économie et la géographie. Cette loi est aussi largement utilisée en fiabilité puisqu’elle permet de modéliser les durées de vie de systèmes ayant un taux de défaillance croissant (systèmes qui s’usent) ou décroissant (systèmes qui s’améliorent), contrairement à la loi exponentielle qui est caractérisée par un taux de défaillance constant.

Bien que la loi de Weibull soit largement utilisée, la vérification de sa pertinence pour un jeu de données est souvent faite par des outils élémentaires comme le graphe de probabilités. Il existe des techniques plus sophistiquées, les tests d’adéquation, qui ont pour vocation de déterminer si un modèle aléatoire est adapté ou non à un jeu de données. Plusieurs tests d’adéquation à la loi de Weibull ont été développés au cours des années, qui permettent d’aider à la prise de décision quant à la pertinence de cette loi. Mais aucune étude comparative exhaustive n’a été effectuée et il n’existe aucun consensus sur les tests les plus efficaces.

L’objectif de cet article est de présenter une revue des tests d’adéquation à la loi de Weibull à deux paramètres, pour des données complètes et censurées. Cette revue comprend des tests usuels ainsi que de nouveaux tests, qui s’avèrent très compétitifs. Une étude comparative complète est présentée, qui permet d’identifier les tests les plus performants. Enfin, on fournit des recommandations d’utilisation de ces tests en fonction des caractéristiques des jeux de données à étudier.

Summary
The Weibull distribution has attracted considerable attention in many fields such as the environmental sciences, geology, chemistry, economics and geography. It is also used in reliability engineering since it allows to model systems with increasing hazard rate (aging systems) or decreasing hazard rate (improving systems), unlike the exponential distribution that assumes that the system failure rate is constant.

Although the Weibull distribution is widely used, the checking of its relevance for a given data set is not always done or done by elementary techniques such as Weibull plots. There exist more sophisticated techniques which aim to determine if a given model is adapted to a given data set, the goodness-of-fit (GOF) tests. Many GOF tests have been developed over the years, but there is no consensus on the most efficient methods.

The aim of this paper is to present an up-to-date review of the GOF tests for the two-parameter Weibull distribution for complete and censored data. This review includes a state of the art of usual tests and the introduction of competitive new tests. An extensive comparison study is presented and it allows to identify the most powerful tests. Finally, some guidance is given on the selection of the most appropriate tests based on the features of the data studied.

1-Industrial context and objectives
Risk management of industrial facilities, such as EDF’s power plants, needs to accurately anticipate the failures and to predict the system reliability in precise way. This requires, as a first step, the building of relevant probabilistic models in order to reflect the randomness of the occurrence of failures. These models should be complex enough to be able to represent the way the systems are used. In a second step, statistical inference of the developed models must be made, based on the available knowledge such as the operation feedback data. When these two steps are carried out, a final step, as important as the previous ones, consists in, firstly to validate the fitted models using statistical criteria and secondly to compare the different competing models. The presented work in this paper falls within this step of model validation.

The Weibull distribution has attracted a huge attention in many fields such as the environmental sciences, geology, chemistry, physics, it allows to model some physical quantities. It is also widely used in reliability engineering since it allows decreasing, constant and increasing failure rates unlike the exponential distribution that makes the assumption of a constant hazard rate. That is why having suitable and efficient techniques to check the relevance of the Weibull distribution for a given data set is of great importance.

Goodness-of-fit (GOF) tests are a useful tool to check the validity of models. They are generally based on the measure of the difference between an empirical quantity computed from the data and a corresponding theoretical one computed from the tested model. Mathematically, the difficulty is to determine the distribution of the test statistic under the hypothesis that the data come from the tested model. The performance of a given test is given by its power, that assesses its capacity to detect the fact that the observations are not issued from the given model.

There is a wide literature on GOF tests for the exponential distribution. Reviews on these tests are given by Spurrier (1984) and Henze & Meintanis (2005). For the Weibull distribution, it is common to validate the choice of this distribution by using graphical methods, such as the Weibull plot. There exists some GOF tests for the Weibull distribution but they are quite old. The most known ones are those based on the empirical distribution function, such as the Cramer-Von Mises and Anderson-Darling tests (D’Agostino & Stephens, 1986). Some tests are based on the usual probability plot (Evans, Johnson & Green, 1989) or the stabilized one (Coles, 1989). Some tests are based on the normalized spacings (Tiku & Singh, 1981,
The two-parameter Weibull distribution $\mathcal{W}(\eta, \beta)$ is defined by its cumulative distribution function (cdf):

$$F(x; \eta, \beta) = 1 - \exp \left( - \frac{x}{\eta} \right)^{\beta}, x \geq 0, \eta > 0, \beta > 0. \quad (1)$$

When $X_1, \ldots, X_n$ is a sample of the $\mathcal{W}(\eta, \beta)$ distribution, its probability density function (pdf) is:

$$f(x; \eta, \beta) = \frac{\beta}{\eta} \left( \frac{x}{\eta} \right)^{\beta-1} e^{-\left( \frac{x}{\eta} \right)^{\beta}}, x \geq 0, \eta > 0, \beta > 0. \quad (2)$$

We have the following results and notations:

- For all $i$, $\ln X_i$ have the extreme value distribution $\mathcal{E}(\ln 1/\beta)$ with the cdf $G(y; \ln 1/\beta) = 1 - \exp(-\exp(\beta(y - \ln \eta)))$ and its pdf is $g(y; \ln 1/\beta) = \beta \exp(\beta(y - \ln \eta)) - \exp(\beta(y - \ln \eta)))$, $y \in \mathbb{R}$.

- Three estimation methods of the parameters $\eta$ and $\beta$ from the sample $X_1, \ldots, X_n$ are considered: the maximum likelihood, least squares and moment methods.

  - The maximum likelihood estimators (MLEs) of $\eta$ and $\beta$, $\hat{\eta}_n$ and $\hat{\beta}_n$, are solutions of the following equations:

    $$\hat{\eta}_n = \left( \frac{1}{n} \sum_{i=1}^{n} X_i^{\hat{\beta}_n} \right)^{1/\hat{\beta}_n} \quad (3)$$

    $$\frac{n}{\hat{\beta}_n} + \sum_{i=1}^{n} \ln X_i - \frac{n}{\hat{\beta}_n} \sum_{i=1}^{n} X_i^{\hat{\beta}_n} \ln X_i = 0. \quad (4)$$

  - The least squares estimators (LSEs) based on the Weibull probability plot, $\tilde{\eta}_n$ and $\tilde{\beta}_n$, are defined as follows (Liao & Shimokawa, 1999):

    $$\tilde{\beta}_n = \frac{\sum_{i=1}^{n} (c_i - \bar{c})^2}{\sum_{j=1}^{n} (\ln X_i - \ln \bar{X})(c_i - \bar{c})} \quad (5)$$
\[ \ln \tilde{\eta}_h = \ln \bar{x} - \frac{\bar{c}}{\hat{\beta}_h} \]  \hspace{1cm} (6)

where \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} \ln X_i \), \( c_i = \ln [-\ln (1 - (i - 0.5)/n)] \) and \( \bar{c} = \frac{1}{n} \sum_{i=1}^{n} c_i \).

- The moment estimators (MEs), \( \hat{\eta}_h \) and \( \hat{\beta}_h \), are defined as follows:
\[ \hat{\beta}_h = \frac{\pi}{\sqrt{6}} \left[ \frac{1}{n-1} \sum_{i=1}^{n} (\ln X_i - \ln \bar{x})^2 \right]^{-1/2} \hspace{1cm} (7)\]
\[ \ln \tilde{\eta}_h = \ln \bar{x} + \frac{\gamma}{\hat{\beta}_h} \hspace{1cm} (8) \]

where \( \gamma = 0.577\ldots \) is the Euler constant.

- For all \( i \), let \( \hat{Y}_i = \ln \left( \frac{X_i}{\hat{\eta}_h} \right) \). Antle & Bain (1976) proved that the distribution of \( (\hat{Y}_1, \ldots, \hat{Y}_n) \) does not depend on \( \eta \) and \( \beta \). We have proved the same result for the two samples \( \hat{Y}_i = \ln \left( \frac{X_i}{\hat{\eta}_h} \right) \) and \( \tilde{Y}_i = \ln \left( \frac{X_i}{\tilde{\eta}_i} \right) \), \( i = 1, \ldots, n \) (Krit, Gaudoin, Remy & Xie, 2014). This property is essential since it is what allows to build GOF tests. The new developed statistics will be functions of either the sample \( \hat{Y}_i, \tilde{Y}_i \) or \( \hat{Y}_i \) or \( \tilde{Y}_i \).

### 3-State of the art of GOF tests for the Weibull distribution

#### 3.1 Tests based on probability plots

For the Weibull distribution, the usual probability plot, or Weibull plot, is the set of points \( (\ln X^*_i, \ln [-\ln (1 - (i - 0.5)/n)]) \), \( i \in \{1, \ldots, n-1\} \). Under the Weibull assumption, these points should be approximately on a straight line.

Evans, Johnson and Green (1989) proposed a GOF test based on a statistic similar to the coefficient of determination of the regression on the Weibull plot:
\[ R^2_{EJG} = \frac{\sum_{i=1}^{n} (\ln X^*_i - \ln X^*_0) M_i^2}{\sum_{i=1}^{n} (\ln X^*_i - \ln X^*_0)^2 \sum_{i=1}^{n} (M_i - M_0)^2} \hspace{1cm} (9) \]

where \( M_i = \frac{1}{\hat{\beta}_h} \ln \left( \frac{1 - (i - 0.3175)/n}{1 + 0.6306} \right) \) and \( \ln X^*_0 = \frac{1}{n} \sum_{i=1}^{n} \ln X_i \).

Ozturk & Korukoglu (1988) adapted the idea of Shapiro-Wilk GOF test of exponentiality. Their idea is to compute the ratio of two estimators of \( \beta \). The corresponding statistic has the following expression:
\[ OK_n = \ln 2 (n - 1) \frac{\sum_{i=1}^{n} (0.6079 w_{n+1} - 0.257 w_i) \ln X^*_i}{\sum_{i=1}^{n} (2i - n - 1) \ln X^*_i} \hspace{1cm} (10) \]

where the \( w_i \) are defined in D’Agostino (1971). They have recommended to use the following standardized statistic that improves the performance of the test:
\[ OK^*_n = \frac{OK_n - 1 - 0.13/\sqrt{n} + 1.18/n}{0.49/\sqrt{n} - 0.36/n} \hspace{1cm} (11) \]

There exist other GOF tests based on probability plots such as the test \( SB_n \) (Smith & Bain, 1976) and the test \( SPP_n \) (Coles, 1989). For all these GOF tests, the rejection of \( H_0 \) is pronounced for too small or too large values of the statistics.

#### 3.2 Tests based on the empirical distribution function

These tests are based on a measure of the departure between the empirical cdf of the \( \ln X_i \): \( G_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i \leq x} \) and the theoretical cdf using the MLEs \( \hat{G}(y) = \hat{G}(y, \ln \hat{\eta}_h, 1/\hat{\beta}_h) \). The null hypothesis is rejected when this difference is too large. The best known statistics are:

- Kolmogorov-Smirnov statistic (KS):
Ziku & Singh (1981) suggested to reject the Weibull hypothesis for both large and small values of the statistic:

\[ \left\lfloor \frac{1}{n} - \hat{U}_i, i = 1 : n \right\rfloor, \max \left( \hat{U}_R - \frac{i-1}{n}, i = 1 : n \right) \]

Liao & Shimokawa (1999) combined the ideas of Kolmogorov-Smirnov and Anderson-Darling statistics with the LSEs instead of the MLEs. They proposed the statistic:

\[ \hat{\eta}_n = \frac{1}{n} \sum_{i=1}^{n} \left( \hat{U}_i - \frac{2i-1}{2n} \right) \]

Both tests \( T_S \) and \( MSF \) reject the Weibull assumption for large and small values of the statistics, while the test \( LOS \) rejects it for only large values.

3-4 Other GOF tests families

There exist other families of GOF tests for the Weibull distribution such as:

\[ KS = \sqrt{n} \sup_{y \in \mathbb{R}} \left| \left[ G_0(y) - \hat{G}_0(y) \right] \right| = \sqrt{n} \max \left( \frac{i}{n} - \hat{U}_i, i = 1 : n \right), \max \left( \hat{U}_R - \frac{i-1}{n}, i = 1 : n \right) \]

Cramer-von Mises statistic (CM):

\[ CM = n \int_{-\infty}^{+\infty} \left[ G_0(y) - \hat{G}_0(y) \right]^2 d\hat{G}_0(y) = \sum_{i=1}^{n} \left( \hat{U}_i - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n} \]

Anderson-Darling statistic (AD):

\[ AD = n \int_{-\infty}^{+\infty} \left[ G_0(y) - \hat{G}_0(y) \right]^2 d\hat{G}_0(y) = -n + \frac{1}{n} \sum_{i=1}^{n} \left[ (2i-1 - 2n) \ln(1 - \hat{U}_i) - (2i-1) \ln(\hat{U}_i) \right] \]

where \( \hat{G}_0(y) = G(y; \ln \tilde{\beta}_n, 1/\tilde{\beta}_n) = 1 - e^{-e^{-\tilde{\beta}_n(y+\ln\tilde{\beta}_0)}} \) and \( \tilde{U}_i = \hat{G}_0(\ln X_i) \). The null hypothesis \( H_0 \) is rejected for large values of the statistics.

3-3 Tests based on the normalized spacings

This family of the GOF tests are based on the normalized spacings defined by:

\[ E_i = \frac{\ln X_{i+1} - \ln X_i}{\beta \hat{E} \left[ \ln X_{i+1} - \ln X_i \right]} \]

Any statistic of the form \( \sum a_i E_i / \sum b_j E_j \) can be used to build a GOF test.

Mann, Schreier & Fertig (1973) used the fact that for \( i \in \{1, \ldots, n-1\} \), the \( \beta E_i \) are asymptotically independent and distributed according to a standard exponential distribution. The authors proposed the statistic:

\[ MSF = \frac{\sum_{i=1}^{n-1} \frac{E_j}{|j+1|}}{\sum_{j=1}^{n-1} E_j} \]

where \( |x| \) is the floor of \( x \).

Tiku & Singh (1981) suggested to reject the Weibull hypothesis for both large and small values of the statistic:

\[ TS = \frac{2 \sum_{i=1}^{n-2} (n-1-i) E_i}{(n-2) \sum_{i=1}^{n-1} E_i} \]

Lockhart, O’Reilly & Stephens (1986), used the random variables \( Z_j = \frac{1}{|j+1|} \sum_{i=1}^{j} E_i / \sum_{i=1}^{n} E_i, j = 1, \ldots, n-2 \). Under \( H_0 \), the \( Z_j \) are approximately distributed as the order statistics of the uniform distribution \( U_0(0,1) \). Then, Lockhart et al proposed a GOF test based on the Anderson-Darling statistic computed for the \( Z_j \):

\[ LOS = 2 - n - \frac{1}{n-2} \sum_{i=1}^{n-2} \left( (2i-n+3) \ln(1-Z_i) - (2i-1) \ln Z_i \right) \]

Both tests \( TS \) and \( MSF \) reject the Weibull assumption for large and small values of the statistics, while the test \( LOS \) rejects it for only large values.
Inspired from this previous work, we have shown that the use of the MLEs instead of MEs keeps the same asymptotic convergence results of the integral \[ \int_1^\infty Y_n(t) \, dt \] has no simplified explicit expression. We can compute it using Simpson or Monte Carlo integration or we can simply discretize it. All the previous GOF tests are included in our comparison study.

### 4-New GOF tests for the Weibull distribution

Cabaña & Quiroz (2005), used the Laplace transform to build GOF tests for the Weibull and type I extreme value distributions. Their statistic is based on the closeness between the empirical and theoretical Laplace transform, which is measured by the empirical moment generating process \[ \hat{\varphi}_n(s) = \sqrt{n} \sum_{j=1}^n e^{-Y_j s} - \Gamma(1-s) \]. Its distribution, under \( H_0 \), does not depend on the Weibull parameters \( \eta \) and \( \beta \). They proved the asymptotic convergence, under \( H_0 \), of \( \hat{\varphi}_n(s), s \in J \) to a Gaussian process \( \hat{G}_\theta(s) \) and they suggested the following test statistic based on the MEs:

\[
CQ_n = \hat{\varphi}_{n,S} V^{-1}(S)(\hat{\varphi}_{n,S})^\top
\]

where \( \hat{\varphi}_{n,S} = (\hat{\varphi}_n(s_1), \ldots, \hat{\varphi}_n(s_k)), S = \{s_1, \ldots, s_k \} \subset J \) and \( V(S) \) is the limiting covariance matrix of \( \hat{\varphi}_{n,S} \) given in Cabaña & Quiroz (2005).

Inspired from this previous work, we have shown that the use of the MLEs instead of MEs keeps the same asymptotic convergence results of \( \hat{\varphi}_n(s) = \sqrt{n} \sum_{j=1}^n e^{-Y_j s} - \Gamma(1-s) \) to a Gaussian process denoted \( \hat{G}_\theta(s) \) with specific covariance matrix \( \hat{V} \) defined in Krit (2014). Hence we have the statistic:

\[
\hat{CQ}_n = \hat{\varphi}_{n,S} \hat{V}^{-1}(S)(\hat{\varphi}_{n,S})^\top
\]

The statistic \( \hat{CQ}_n \) has a limiting chi-squared distribution with \( k \) degrees of freedom, when \( n \to +\infty \). The covariance matrix \( \hat{V} \) can be substituted with any non singular matrix \( A \). The distribution of the statistic is still independent of the parameters and the new statistic is denoted \( \hat{CQ}^* \):

\[
\hat{CQ}^* = \hat{\varphi}_{n,S} A (\hat{\varphi}_{n,S})^\top.
\]

The advantage of this new statistic is the fact that there is no need to compute the inverse of the covariance matrix and the possibility to adjust the values of the covariances the way it maximizes the performance of the test.

In Krit (2014), other GOF tests were suggested. These tests combine both approaches: the one of Henze (1993) based on the weighted \( L^2 \) norm and the one of Cabaña & Quiroz (2005) based on the difference between the Laplace transform and its empirical version, measured by \[ \hat{\varphi}_n(s) = \sqrt{n} \sum_{j=1}^n e^{-Y_j s} - \Gamma(1-s) \]. A statistic similar to the statistic of Henze can be written as an integral:

\[
\int_I \nu^2(t) w_\alpha(t) \, dt
\]

where \( I \subset [-\infty, 1] \) is a bounded interval and \( w_\alpha \) is a weight function which depends on a parameter \( \alpha \) that can be chosen to obtain the best performance of the test. It is common in the tests of Cramer-Von-Mises and Anderson-Darling to use the pdf of the tested distribution as a weight function. That is why we use the pdf of the \( \delta \nu(0, 1) \) dilated with a parameter \( \alpha \): \( w_\alpha(t) = e^{\alpha t^2} \).

The integral[23] has no simplified explicit expression. We can compute it using Simpson or Monte Carlo integration or we can simply discretize it on an appropriately chosen range of values of \( t \). With a discretization on \([0,1]\), we obtain the following test statistic:

\[
LT_{\alpha,m} = n \sum_{k=-m}^{1} \left( \frac{1}{n} \sum_{j=1}^n e^{-Y_j k/m} - \Gamma(1-k/m) \right)^2 w_\alpha(k/m).
\]
This statistic $LT_{a,m}$ can be computed using either the sample $\tilde{Y}_1, \ldots, \tilde{Y}_n$ or $\tilde{Y}_1, \ldots, \tilde{Y}_n$ or $\tilde{Y}_1, \ldots, \tilde{Y}_n$ instead of $Y_1, \ldots, Y_n$ depending on the chosen estimation method. We denote respectively the corresponding versions $LT_{a,m}$, $LT_{a,m}$ and $LT_{a,m}$. We have the convergence of, respectively, the statistics $LT_{a,m}$ and $LT_{a,m}$, under $H_0$, to the distribution of $\sum_{k=1}^{m} G_p(k/m)w_i(k/m)$ and $\sum_{k=1}^{m} G_p(k/m)w_i(k/m)$.

The Weibull assumption is rejected at level $\alpha$ if the statistic is greater than the quantile of order $1 - \alpha$ of the distribution of the chosen statistic under $H_0$. The quantiles are obtained by simulation.

4-2 **Likelihood based tests**

Likelihood based GOF tests for the two-parameter Weibull distribution were proposed in Krit, Gaudoin, Remy & Xie (2014). The principle of these tests consists in nesting the Weibull distribution in three-parameter generalized Weibull families and testing the value of the third parameter by using the Wald, score and likelihood ratio tests. The building approach of these test statistics relies on getting rid of the nuisance parameters, using the three estimation methods previously defined. Since the $Y_i = \beta \ln(X_i/\eta)$ are a sample of the $\mathcal{E} \mathcal{F}_1(0, 1)$ distribution, this distribution has no unknown parameter. Using the idea of likelihood based tests, the $\mathcal{E} \mathcal{F}_1(0, 1)$ distribution can be included in a larger family with only one parameter $\theta$. It is possible to derive likelihood based tests of $H_0$: “$\theta = \theta_0$” vs $H_1$: “$\theta \neq \theta_0$” in these families.

This approach is summarized in the following steps:

- Choose a generalized Weibull family $\mathcal{G}W/(\theta, \beta)$,
- Compute the pdf of $Y = \ln X$ when $\eta = \beta = 1, g(.; \theta)$,
- Compute the likelihood $L(\theta) = \prod_{i=1}^{n} g(y_i; \theta)$ and the MLE of $\theta$, $\hat{\theta}_n$,
- Compute the score $U(\theta) = \frac{\partial \ln L(\theta)}{\partial \theta}$ and the observed information $I(\theta) = \frac{\partial^2 \ln L(\theta)}{\partial \theta^2}$,
- The likelihood based statistics are:

$$W = I(\theta_0)(\hat{\theta}_n - \theta_0)^2,$$

$$Sc = \frac{U^2(\hat{\theta}_n)}{I(\hat{\theta}_n)}$$

$$LR = -2 \ln \frac{L(\hat{\theta}_n)}{L(\theta_0)}$$

- Many Generalized Weibull families were used to develop the new GOF tests. For each family, three GOF tests are developed (based on Wald, the score and the likelihood ratio procedures). Each test statistic has three versions, derived either from $\tilde{Y}_i$, $\hat{Y}_i$ or $\hat{F}_i$. In total 63 GOF tests have been developed and compared in Krit, Gaudoin, Remy & Xie (2014).

Among the generalized Weibull families that were used in building these tests, let us mention only those which lead to particularly powerful GOF tests:

- Exponentiated Weibull (Madhokar & Srivastava 1993), whose Wald statistic is given by:

$$EW = n(\hat{\theta}_n - 1)^2$$

where $\hat{\theta}_n = -n/\sum_{i=1}^{n} \ln(1 - e^{-x_i})$.

- Generalized Gamma (Stacy, 1962), whose likelihood ratio based statistic is:
\[ GG = (2\hat{k}_n - 1)n\ln\hat{k}_n - 2n\ln\Gamma(\hat{k}_n) + 2(\sqrt{\hat{k}_n} - 1) \sum_{i=1}^{n} Y_i + 2 \sum_{i=1}^{n} e^{Y_i} - 2\hat{k}_n \sum_{i=1}^{n} e^{Y_i} \left[ \sum_{i=1}^{n} e^{Y_i} - \frac{n}{\hat{k}_n} \right] \]  

(29)

where \( \hat{k}_n \) verifies the equation: 
\[ \varphi(\hat{k}_n) = \frac{1}{n} \sum_{i=1}^{n} Y_i \quad \text{and} \quad \varphi = \int_{0}^{1} \] is the digamma function.

- **Marshall- Olkin Extended Weibull** (Marshall & Olkin, 1997), whose Wald statistic is:

\[ MO = (\hat{\alpha}_n - 1)^2 \left( n - 2 \sum_{i=1}^{n} e^{-2\rho_i} \right) \]

(30)

where \( \hat{\alpha} \) verifies the equation:
\[ \alpha - 2 \sum_{i=1}^{n} \frac{e^{-\rho_i}}{1 - (1 - \alpha)e^{-\rho_i}} = 0 \]

- **Modified Weibull** (Lai & Xie, 2003), whose Wald statistic is:

\[ MW = \beta_n^2 \left[ \sum_{i=1}^{n} e^{2\rho_i} + \sum_{i=1}^{n} e^{3\rho_i} \right] \]

(31)

where \( \beta_n \) verifies:
\[ \sum_{i=1}^{n} e^{\rho_i} + \sum_{i=1}^{n} \beta_n e^{\rho_i} = \sum_{i=1}^{n} e^{2\rho_i} + \rho_n e^{\rho_i}. \]

- **Power Generalized Weibull** (Haghighi & Nikulin, 2006), whose Wald statistic is:

\[ PGW = (\hat{\nu}_n - 1)^2 \left[ -n + 2 \sum_{i=1}^{n} \ln(1 + e^{\rho_i}) e^{\rho_i} + \sum_{i=1}^{n} \left( \ln(1 + e^{\rho_i}) \right)^2 (1 + e^{\rho_i}) \right] \]

(32)

where: \( \hat{\nu}_n = \frac{1}{n} \sum_{i=1}^{n} \ln(1 + e^{\rho_i}) \left[ (1 + e^{\rho_i}) \right] \frac{1}{n} - 1 \).

Theoretical results about the asymptotic distribution of some of these tests were established. But these asymptotic results are not practical in the case of small samples. That is why the quantiles have to be assessed using Monte-Carlo simulations. The rejection of Weibull assumption is done for large values of the statistics.

### 4-3 Combined GOF tests

Complementary behaviors of some likelihood based GOF tests have been mentioned in Kritt, Gaudoin, Remy & Xie (2014). For instance, when the statistic \( PGW \) has a very low power against a fixed alternative, the statistic \( MW \) has, conversely, very high power against the same alternative and vice versa. Building a GOF test that combines both statistics might help to get rid of the bias and give a global good performance for almost all the tested alternatives. This approach can be applied to any tests statistics with complementary behaviors.

For instance, we combine here the two test statistics \( MW \) and \( PGW \). In order to keep the same order of magnitude, we center each statistic \( MW \) and \( PGW \) by its mean value (respectively \( \hat{MW} \) and \( \hat{PGW} \)) and normalize it by its standard deviation (respectively \( sd(MW) \) and \( sd(PGW) \)), these two depending on the size of the tested data.

\[ T_1 = \max \left( \frac{1}{sd(MW)} |\hat{MW} - \hat{MW}|, \frac{1}{sd(PGW)} |\hat{PGW} - \hat{PGW}| \right) \]

(33)

\[ T_2 = \frac{1}{sd(MW)} |\hat{MW} - \hat{MW}| + \frac{1}{sd(PGW)} |\hat{PGW} - \hat{PGW}|. \]

(34)
The Weibull assumption is rejected for large values of the statistics. The quantiles of the distributions of \( T_1 \) and \( T_2 \), under \( H_0 \), are given by Monte-Carlo simulations and can be applied to any sample size.

### 4-4 GOF tests for censored samples

There are several kinds of censorings. Among them, type II simple censoring occurs when a predetermined number of lifetimes is not observed; the type II left censoring occurs when the \( r_1 \) smallest observations are missing and the right censoring is when the \( r_2 \) largest observations are missing. The type I multiple censoring occurs when each observation has a specific censoring time.

All the GOF tests based on the MLEs can be adapted to simply censored samples on both sides. Indeed, the MLEs \( \hat{\eta}_i \) and \( \hat{\beta}_i \) can be computed in the case of type II simple censoring. Let \( x_i \leq \ldots \leq x_i \leq \ldots \leq x_{n-r_2} \leq \ldots \leq x_n \) be a simple left and right censored sample, where the \( r_1 \) smallest and \( r_2 \) largest values are missing. The MLEs of the Weibull parameters distribution maximize the likelihood function:

\[
\mathcal{L}(\mathbf{x}, \eta, \beta) = [F(x_{i+1})]^{r_1} \prod_{i=1}^{n-r_2} f(x_i) [1 - F(x_{n-r_2})]^{r_2}
\]

The MLEs of \( \eta \) and \( \beta \) verify the two equations:

\[
-r_1 \left( \frac{\hat{\beta}_i}{\hat{\eta}_i} \right) \hat{\beta}_i \left( \frac{x_i+1}{\hat{\eta}_i} \right) \hat{\eta}_i \exp\left( -\left( x_i+1 / \hat{\eta}_i \right) \hat{\beta}_i \right) + r_2 \left( \frac{\hat{\beta}_i}{\hat{\eta}_i} \right) \hat{\eta}_i \exp\left( -\left( x_{n-r_2} / \hat{\eta}_i \right) \hat{\beta}_i \right) - (n - r_1 - r_2) \left( \frac{\hat{\beta}_i}{\hat{\eta}_i} \right) + \left( \frac{\hat{\beta}_i}{\hat{\eta}_i} \right) \sum_{i=r_1+1}^{n-r_2} \left( \frac{x_i}{\hat{\eta}_i} \right) = 0 \quad (36)
\]

\[
r_1 \left( \frac{x_i+1}{\hat{\eta}_i} \right) \hat{\beta}_i \ln \left( \frac{x_i+1}{\hat{\eta}_i} \right) \exp\left( -\left( x_i+1 / \hat{\eta}_i \right) \hat{\beta}_i \right) - r_2 \ln \left( \frac{x_{n-r_2}}{\hat{\eta}_i} \right) + \left( n - r_1 - r_2 \right) \left( \frac{\hat{\beta}_i}{\hat{\eta}_i} \right) + \sum_{i=r_1+1}^{n-r_2} \left( 1 - \left( \frac{x_i}{\hat{\eta}_i} \right) \right) \ln \left( \frac{x_i}{\hat{\eta}_i} \right) = 0 \quad (37)
\]

After deriving the MLEs, we can compute \( \hat{Y} = \hat{\beta}_i \ln \left( \frac{x_i}{\hat{\eta}_i} \right) \). The property of independence, under \( H_0 \), of the distribution of the \( \hat{Y}_i \) from the Weibull parameters is remained in the case of simple censored samples (Antle & Bain, 1969). Then, all the previous GOF tests based on \( \hat{Y}_i \) can be applied in the case of type II simple censoring.

The type I multi-censoring does not keep this independence property and in this case, the distribution of the sample \( \hat{Y}_i \) depends on the parameters \( \eta \) and \( \beta \) and on the censoring times. We have proved this result by simulation. Then, the presented GOF tests can not be adapted in the case of multi-censoring.

### 5-Comparison study

The purpose of this section is to use intensive Monte Carlo simulations in order to assess the performance of all the presented tests and to compare them all. The algorithms have been written in R and a special package, called EWGoF, has been developed. It allows to use any of the Weibull GOF tests presented in this paper and others (86 GOF tests). Also a broad range of Exponential GOF tests is given in the same package (20 GOF tests).

The comparison study is done using a broad class of alternative distributions. For each distribution, we simulate 50000 complete samples of size \( n \in \{10, 20, 50\} \). All the GOF tests are applied with a significance level set to \( \alpha = 5\% \). The power of the tests is assessed by the percentage of rejection of the null hypothesis. First, Weibull samples are simulated in order to check that the percentage of rejection is close to the nominal significance level 5\%. For the other simulations, we have chosen a broad range of alternative distributions: with increasing hazard rate (IHR), decreasing hazard rate (DHR), bathtub hazard rate (BT) and upside-down hazard rate (UBT).

From the analysis of the powers of each test, we have concluded that the performances of the tests depend on the monotony of the underlying hazard rate. A non parametric estimation of the hazard rate can be done, in order to assess its monotony. In the case this monotony is known, we recommend the following GOF tests:

- For IHR alternatives: \( \hat{L}_S \).
- For UBT alternatives: \( \hat{S}_T_4 \).
- For DHR-BT alternatives: \( \hat{T}_1 \).

If no information is known about the hazard rate, we have the following conclusions:

- Among all the usual tests, the tests \( T_S \) and \( OK^* \) are very powerful. The test \( OK^* \) has the simplest expression which is more convenient for practical purposes.
- The test \( \hat{T}_1 \) is the most powerful test, because it combines two complementary GOF tests statistics with two different methods of estimating the parameters.
6-Case study

The case study deals with the mechanical performance of a passive component within a power plant. The reliability of the component depends on two main characteristics: the length of the defects and the toughness of the material. Under severe stress conditions, the preexisting flaws, which uneventfully remain non-progressive through the operation of the structure, might initiate if its toughness is not high enough. Examinations have been performed, resulting in a hundred measures of the length of the defects and of the toughness under a fixed temperature. The question to be answered is whether or not the Weibull distribution is adapted for both studied variables (length of the defects and the toughness). Graphically in figure 1, the two histograms and the estimated pdfs (kernel non-parametric estimation) of the two data sets, show that the Weibull distribution can be a candidate distribution.

![Length of defects and Toughness](image)

**Figure 1.** Histograms and estimated pdfs of the two data sets

We have applied the best of the previous GOF tests and computed their corresponding p-values. The p-values are given in tables 1 and 2:

**Table 1.** P-values of the best GOF tests of Weibull distribution applied to the length of the defects

<table>
<thead>
<tr>
<th>GOF tests</th>
<th>AD</th>
<th>OK*</th>
<th>SPP</th>
<th>TS</th>
<th>ST*</th>
<th>CQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-values</td>
<td>6.3%</td>
<td>30.2%</td>
<td>13.2%</td>
<td>13.5%</td>
<td>15.1%</td>
<td>30%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GOF tests</th>
<th>GG</th>
<th>EW</th>
<th>PGW</th>
<th>MO</th>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-values</td>
<td>22.6%</td>
<td>25.2%</td>
<td>28.1%</td>
<td>9.6%</td>
<td>9.5%</td>
<td>22.4%</td>
</tr>
</tbody>
</table>

**Table 2.** P-values of the best GOF tests of Weibull distribution applied to the toughness values

<table>
<thead>
<tr>
<th>GOF tests</th>
<th>AD</th>
<th>OK*</th>
<th>SPP</th>
<th>TS</th>
<th>ST*</th>
<th>CQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-values</td>
<td>77.3%</td>
<td>94.8%</td>
<td>88.9%</td>
<td>87%</td>
<td>87.1%</td>
<td>92.2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GOF tests</th>
<th>GG</th>
<th>EW</th>
<th>PGW</th>
<th>MO</th>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-values</td>
<td>95%</td>
<td>97%</td>
<td>99%</td>
<td>95.8%</td>
<td>52%</td>
<td>51%</td>
</tr>
</tbody>
</table>
For the toughness data, the p-values are very high. Then, the Weibull distribution is clearly not rejected in this case. For the length of defects data, the p-values are not too high especially for some GOF tests such as $AD$, $MO$ and $T_1$, but still large enough to not reject the Weibull distribution. The p-value of $TS$ is also not so high, this can be explained by the fact that the tested data have a lot of ties. Since the statistic $TS$ is based on the spacings, a large number of these spacings is null that is why the p-value is not too high, but it is still high enough to lead to the same conclusion. Then, the Weibull distribution can be used for both data sets.

7-Conclusion

A broad range of GOF tests for the Weibull distribution has been studied. An intensive comparison study led to identify the most powerful tests. If nothing particular is known about the alternative distribution, we recommend the use of $OK^*$ and $TS$. The $OK^*$ test is the most interesting test, in the sense that it has a simple expression that makes it of simple use, alike the $TS$ test. Also the $OK^*$ test can be applied for any kind of samples even samples with ties. These tests are not widely used in practice and sometimes they are not even known by the practitioner. The combined tests $T_1$ and $T_2$ are also recommended. It seems that any judicious combination of complementary tests can improve the GOF tests performances. An R package (EWGoF) has been developed to apply all the presented GOF tests. Some of the presented tests can also be applied to type II simple censored data sets.

An important prospect is to develop GOF tests in the case of multi-censored data. Moreover, the presented tests deal only with non repairable systems. For repairable systems, we have developed new exact GOF tests based on the conditional sampling given sufficient statistics.

8-References


Cabaña A. & Quiroz A.J. (2005), Using the empirical moment generating function in testing the Weibull and type 1 Extreme Value distributions, Test, 14(2), 417-431.


Coles S.J. (1989), On goodness-of-fit tests for the two-parameter Weibull distribution derived from the stabilized probability plot, Biometrika, 76(3), 593-598.


