Goodness-of-fit tests for the Weibull distribution with censored data

Florian PRIVÉ[1], Olivier GAUDOIN[1] & Emmanuel REMY[2]

1. Univ. Grenoble Alpes, Laboratoire Jean Kuntzmann, France
2. EDF R&D, Industrial Risk Management, Chatou, France

23 June 2016

EDF & Laboratoire Jean Kuntzmann (LJK)

Table of contents

Table of contents

Industrial context

Risk management of industrial facilities, such as EDF's (major French electric utility) power plants, needs to accurately predict system reliability:

- Building of relevant probabilistic models,
- Statistical inference of the developed models,
- Validation of the fitted models using statistical criteria such as **goodness-of-fit tests**.

The most usual models for lifetimes are the exponential and Weibull distributions.

Previous work

- Krit, Gaudoin & Remy (2014) have made an exhaustive study in order to identify the best:
  - Goodness-of-fit tests for the Exponential distribution, for complete and censored data,
  - Goodness-of-fit tests for the Weibull distribution, for complete samples only.
- The present work focuses on:
  - Goodness-of-fit tests for the Weibull distribution **for censored samples**,
  - type-II censoring only.

Table of contents

Censored samples

Let X be a sample of n times before failures:

$$X_1, \dots, X_n,$$

The **ordered** sample is:

$$X_1^*, \dots, X_n^*.$$

In the case of type-II censoring, **we observe only**:

$$X_1^*, \dots, X_m^*$$

where $m < n$ (for example, $m = 15$ and $n = 351$).

Goodness-of-fit test

Goodness-of-fit test

Null hypothesis:
H_0 : "$X_1, \dots, X_n$ is a sample from the Weibull distribution."
Alternative:
H_1 : "$X_1, \dots, X_n$ is not a sample from the Weibull distribution."

Note that

- Only the part $X_1^*, \dots, X_m^*$ of $X_1, \dots, X_n$ is observed.
- We test the assumption that the sample comes from the family of Weibull distributions (with **unknown parameters**), NOT that it comes from a fully specified Weibull distribution.

Definition and Property

Cumulative density function of the two-parameter Weibull distribution $\mathcal{W}(\eta, \beta)$:

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

If $X \sim \mathcal{W}(\eta, \beta)$, then $\ln(X) \sim \mathcal{EV}_1(\mu, \sigma)$ where $\mu = \ln(\eta)$ and $\sigma = \frac{1}{\beta}$ are respectively **location and scale parameters**:

$$\boxed{Y = \beta \ln\left(\frac{X}{\eta}\right) = \frac{\ln(X) - \mu}{\sigma} \sim \mathcal{EV}_1(0, 1)} \qquad (2)$$

where $\mathcal{EV}_1$ is the type-I Extreme Value distribution.

Estimators

$$\forall i, \text{ let } \hat{Y}_i = \hat{\beta}_{m,n} \ln\left(\frac{X_i}{\hat{\eta}_{m,n}}\right) \text{ and } \tilde{Y}_i = \tilde{\beta}_{m,n} \ln\left(\frac{X_i}{\tilde{\eta}_{m,n}}\right), \text{ where:}$$

- $\hat{\eta}_{m,n}$ and $\hat{\beta}_{m,n}$ are the **Maximum Likelihood** Estimators (MLEs),
- $\tilde{\eta}_{m,n}$ and $\tilde{\beta}_{m,n}$ are the **Least Squares** Estimators (LSEs). They are based on the Weibull Probability Plot, defined right after,

Independence from parameters

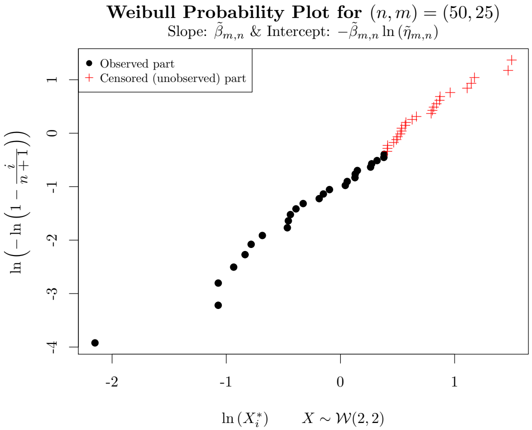
The distributions of $\hat{Y}_i$ and $\tilde{Y}_i$ are

- expected not to be far from the one of Y: $\mathcal{EV}_1(0,1)$,
- independent from both η and β (or μ and σ), respectively proved by Antle & Bain (1969) for the MLE and adapted from Liao & Shimokawa (1999) for the LSE.

Table of contents

Weibull Probability Plot

$$\ln\left(-\ln\left(1 - F(x;\eta,\beta)\right)\right) = \beta(\ln(x) - \ln(\eta)) \quad (3)$$



Test based on the WPP

Smith & Bain (1976) used the statistic $Z^2 = n(1 - R^2_{SB})$, where $R^2_{SB} =$

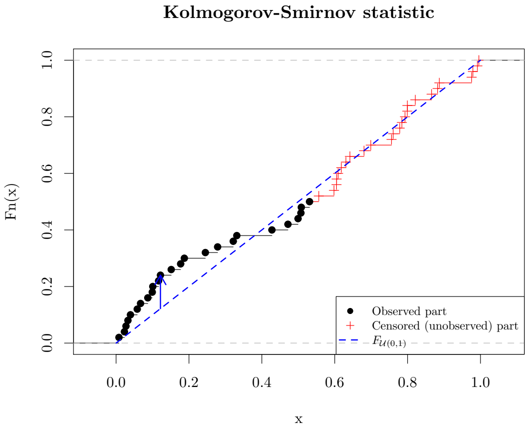
$$\frac{\left[\sum_{i=1}^{m}(\ln(X_i^*) - \overline{\ln(X^*)})(c_i - \overline{c})\right]^2}{\sum_{i=1}^{m}\left(\ln\left(X_i^*\right) - \overline{\ln(X^*)}\right)^2 \sum_{i=1}^{m}(c_i - \overline{c})^2} = \frac{\left[\sum_{i=1}^{m}(Y_i^* - \overline{Y^*})(c_i - \overline{c})\right]^2}{\sum_{i=1}^{m}(Y_i^* - \overline{Y^*})^2 \sum_{i=1}^{m}(c_i - \overline{c})^2} \quad (4)$$

with $c_i = \ln(-\ln(1 - p_i))$ and $p_i = \frac{i}{n+1}$ (mean ranks).

The test rejects the null hypothesis when Z^2 is too high. This test can be adapted with other plotting positions p_i.

Tests based on the empirical distribution function: a demonstrative example for the Kolmogorov-Smirnov statistic

We use either the $\hat{U}_i^* = F_{\mathcal{EV}_1(0,1)}(\hat{Y}_i^*) = 1 - \exp\left(-\exp\left(\hat{Y}_i^*\right)\right)$ or the $\tilde{U}_i^*$ and then compare their distribution to $\mathcal{U}(0,1)$ distribution.



Tests based on the empirical distribution function

D'Agostino & Stephens (1986):

- Kolmogorov-Smirnov (**KS**):

$$KS_{m,n} = \max_{1 \leq i \leq m} \left\{ \frac{i}{n} - \hat{U}_i^*, \hat{U}_i^* - \frac{i-1}{n} \right\} \tag{5}$$

- Cramer-von Mises (**CM**), Anderson-Darling (**AD**) and Watson (**W**)

Liao-Shimokawa's statistic (1999) (**LS**), adapted for censored data:

$$LS_{m,n} = \sum_{i=1}^{m} \frac{\max \left\{ \frac{i}{n} - \tilde{U}_i^*, \tilde{U}_i^* - \frac{i-1}{n} \right\}}{\sqrt{\tilde{U}_i^*(1 - \tilde{U}_i^*)}} \tag{6}$$

EDF & Laboratoire Jean Kuntzmann (LJK)

Tests based on the normalized spacings - 1

Normalized spacings are defined as: $\forall i \in 1, \ldots, m-1$,

$$E_i = \frac{\ln\left(X_{i+1}^*\right) - \ln\left(X_i^*\right)}{\mathrm{E}\left[\frac{\ln\left(X_{i+1}^*\right) - \mu}{\sigma}\right] - \mathrm{E}\left[\frac{\ln\left(X_i^*\right) - \mu}{\sigma}\right]} = \sigma \frac{Y_{i+1}^* - Y_i^*}{\mathrm{E}\left[Y_{i+1}^* - Y_i^*\right]}. \tag{7}$$

Thus, the E_i are independent from μ and directly proportional to σ. Therefore, any statistic which can be written as

$$\sum_i a_i E_i \, / \sum_j b_j E_j \tag{8}$$

is independent from both parameters and can be used as a test statistic for the Weibull distribution hypothesis.

Tests based on the normalized spacings - 2

- 3 test statistics have been proposed, among them Tiku-Singh (1981):

$$TS_m = \frac{2\sum_{i=1}^{m-2}(m-i-1)E_i}{(m-2)\sum_{j=1}^{m-1} E_j} \qquad (9)$$

- **A new idea** is to use either Spearman or Kendall trend tests on the E_i.

Simplified likelihood based tests - 1

These tests, adapted from the corresponding ones for complete data (Krit et al, 2016), are based on **generalized Weibull distributions (with 3 parameters)**. For instance:

- $\mathcal{EW}(\theta, \eta, \beta)$, whose cdf is:

$$\left[1 - e^{-(x/\eta)^{\beta}}\right]^{\theta} \tag{10}$$

  then we can test $\theta = 1$ (particular case of Weibull with only 2 parameters) vs $\theta \neq 1$.

- $\mathcal{AW}(\xi, \eta, \beta)$ whose cdf is:

$$1 - e^{-\xi x - (x/\eta)^{\beta}}. \tag{11}$$

  This time, we test $\xi = 0$ vs $\xi \neq 0$.

Simplified likelihood based tests - 2

The score and observed information are:

$$U(\theta) = \frac{\partial \ln(L)}{\partial \theta}(\theta), \tag{12}$$

$$I(\theta) = -\frac{\partial^2 \ln(L)}{\partial \theta^2}(\theta). \tag{13}$$

The likelihood based statistics are:

$$W = I(\theta_0)(\hat{\theta}_{m,n} - \theta_0)^2, \tag{14}$$

$$Sc = \frac{U^2(\theta_0)}{I(\theta_0)}, \tag{15}$$

$$LR = -2 \ln\left(\frac{L(\theta_0)}{L(\hat{\theta}_{m,n})}\right). \tag{16}$$

Other tests

- Shapiro-Wilk type tests, based on the ratio of two linear estimators of $\sigma = 1/\beta$,
- Tests based on the Kullback-Leibler information,
- Others not presented here.

The goal was to be as thorough as possible in order to obtain the best performing test statistics.

The powers of a total of 75 goodness-of-fit tests were investigated.

Table of contents

Simulations

These tests are NOT asymptotic.

We used:

- 500 000 simulations in order to estimate the quantiles of the tests statistics distribution under H_0,
- 100 000 simulations in order to estimate the power of a given test statistic for a given alternative (16 different ones).
- simulations for $n = 50$ and $m \in \{25, 50\}$.
- and others for different values of n and m.

Alternatives

We have chosen usual alternatives of the Weibull distribution:

- the Gamma distribution $\mathcal{G}(k, \theta)$,
- the Lognormal distribution $\mathcal{LN}(\mu, \sigma)$,
- the Inverse-Gamma distribution $\mathcal{IG}(\alpha, \beta)$,

but also new ones introduced in Krit et al (2014):

- several $\mathcal{GW}$ distributions: $\mathcal{AW}(\xi, \eta, \beta)$, $\mathcal{EW}(\theta, \eta, \beta)$ and $\mathcal{GG}(k, \eta, \beta)$,
- the distributions I and II of Dhillon: $\mathcal{D}1(\beta, b)$ and $\mathcal{D}2(\lambda, b)$,
- the Inverse Gaussian distribution $\mathcal{IS}(\mu, \lambda)$,
- Chen's distribution $\mathcal{C}(\lambda, \beta)$,

Alternatives

We grouped them according to the shape of their hazard rate:

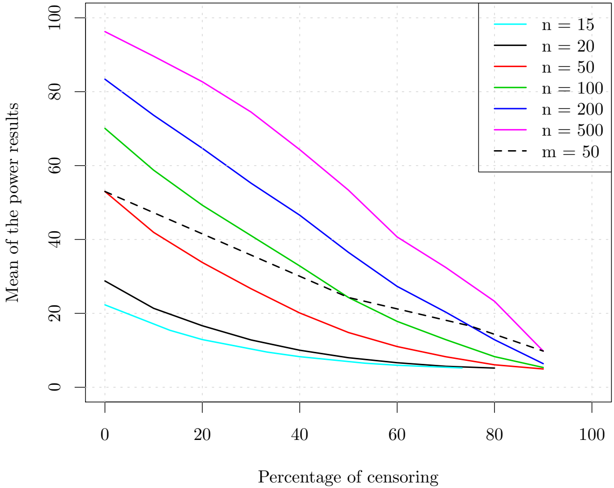
- increasing hazard rate (IHR)
- upside-down bathtub-shaped hazard rate (UBT)
- decreasing hazard rate (DHR)
- bathtub-shaped hazard rate (BT)

Weibull	$Exp(1)$	$\mathcal{W}(1, 0.5)$	$\mathcal{W}(1, 3)$
IHR	$\mathcal{G}(2, 1)$ $\mathcal{D}2(1, 2)$	$\mathcal{G}(3, 1)$	$\mathcal{AW}(10, 0.02, 5.2)$
DHR	$\mathcal{G}(0.2, 1)$	$\mathcal{AW}(2, 20, 0.1)$	$\mathcal{EW}(0.5, 1, 0.95)$
BT	$\mathcal{GG}(0.1, 1, 4)$ $\mathcal{D}1(1, 0.8)$	$\mathcal{GG}(0.2, 1, 3)$	$\mathcal{C}(2, 0.4)$
UBT	$\mathcal{LN}(0, 0.8)$ $\mathcal{IS}(1, 0.25)$	$\mathcal{IG}(3, 1)$ $\mathcal{IS}(1, 4)$	$\mathcal{EW}(4, 12, 0.6)$

Power results

		n = 50 — m = 50 (complete)						n = 50 — m = 25 (censored)					
		R_{SB}	$\widetilde{KS}$	$\widehat{AD}$	TS	$\widehat{EW}_S$	$\widetilde{MW}_S$	R_{SB}	$\widetilde{KS}$	$\widehat{AD}$	TS	$\widehat{EW}_S$	$\widetilde{MW}_S$
Weibull	Exp(1)	5,1	5,1	5,1	5,0	5,1	5,0	5,1	4,9	5,0	5,1	4,9	4,9
	Weibull(1, 0.5)	5,0	5,0	5,0	5,0	5,1	5,1	5,1	5,0	5,0	5,0	5,1	5,0
	Weibull(1, 3)	5,1	5,1	5,0	5,0	5,0	4,9	5,0	4,9	4,9	5,1	4,9	5,0
IHR	Gamma(2, 1)	2,3	6,2	8,5	11,2	10,4	13,9	3,3	5,8	4,6	5,4	3,5	7,0
	Gamma(3, 1)	2,4	8,5	13,2	18,5	17,1	22,5	2,7	6,7	5,0	6,3	3,2	8,3
	AW(10, 0.02, 5.2)	80,3	79,6	71,0	82,1	83,5	81,3	43,2	50,5	58,2	60,1	67,6	46,8
	Dhillon2(1, 2)	2,5	8,0	12,8	17,6	17,4	21,9	2,9	6,3	4,8	5,9	3,2	7,8
UBT	Lnorm(0, 0.8)	22,4	34,3	55,6	71,4	64,7	75,2	1,9	14,9	9,2	13,3	5,9	19,1
	InvGamma(3, 1)	76,4	74,4	91,7	96,8	93,3	97,3	3,7	27,2	17,7	25,1	12,7	34,7
	EW(4, 12, 0.6)	5,2	15,6	26,4	37,2	34,5	42,5	2,1	8,7	5,9	7,9	3,5	11,3
	InvGauss(1, 0.25)	73,9	71,8	89,3	95,8	87,4	96,4	7,0	37,6	26,7	35,8	20,1	48,0
	InvGauss(1, 4)	24,2	36,3	56,6	73,7	64,7	77,2	2,2	18,4	11,1	16,5	7,5	23,6
DHR	Gamma(0.2, 1)	23,7	31,7	45,8	55,2	56,6	40,4	7,6	6,0	8,7	8,1	10,8	4,7
	AW(2, 20, 0.1)	85,8	99,3	100,0	99,9	99,8	97,7	6,8	8,4	20,4	17,9	20,0	7,7
	EW(0.5, 1, 0.95)	12,4	12,2	13,8	17,9	19,7	12,8	6,7	5,4	6,8	6,4	8,3	4,6
BT	GG(0.1, 1, 4)	29,6	46,7	67,8	74,1	73,7	55,6	7,7	6,1	8,8	8,4	11,1	4,9
	GG(0.2, 1, 3)	23,5	31,8	45,7	55,1	56,7	40,3	7,7	6,0	8,8	8,2	10,9	4,7
	Chen(2, 0.4)	10,0	9,7	11,9	14,9	16,1	9,8	6,1	5,0	6,1	5,6	6,9	4,5
	Dhillon1(1, 0.8)	14,8	16,1	20,1	26,4	28,7	18,6	6,9	5,6	7,2	6,9	8,8	4,5
	Mean	30,6	36,4	45,6	53,0	51,5	50,2	7,4	13,7	13,1	14,8	12,7	15,1

Abacus of power results for the Tiku-Singh test statistic



EDF & Laboratoire Jean Kuntzmann (LJK)

Analysis of a real data set

$$(X_1^*, \ldots, X_{15}^*) = (8, 13, 14, 18, 23, 27, 31, 33, 40, 41, 41, 41, 42, 42, 45).$$

- lifetimes of components in EDF hydropower plants,
- this sample is censored at more than 95%: it is only the $m = 15$ first failure times of $n = 351$ components.

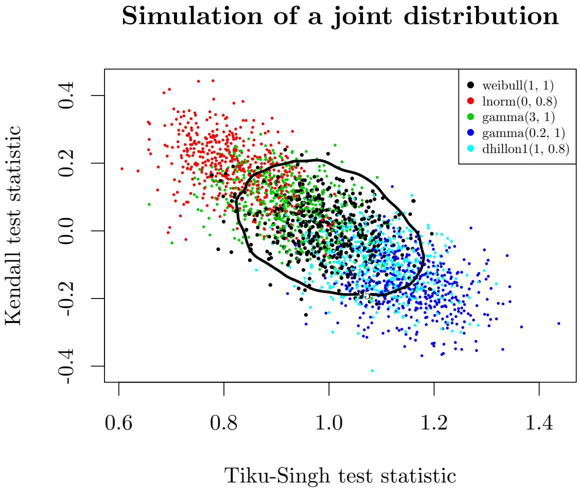
If we run a Tiku-Singh test on these values, we get a p-value of 46.4%.

The Weibull assumption can't be rejected for those data.

Table of contents

Conclusion and perspectives

- For small sample sizes and, especially, for strong censoring, the powers of the tests are quite small.
- Tiku-Singh is the best of the 75 studied tests (fast, powerful and unbiased).
- Other types of censoring should be investigated.
- The combination of tests should be explored.



Thanks for your attention !

Any questions ?