Parametric bootstrap goodness-of-fit tests for imperfect maintenance models

Cécile Chauvel[1], Jean-Yves Dauxois[2], Laurent Doyen[1] and Olivier Gaudoin[1]

1: Laboratoire Jean Kuntzmann, Université Grenoble Alpes
2: Institut de Mathématiques de Toulouse, Université de Toulouse-INSA

4th AMMSI Workshop, Grenoble, January 2016

Context

- Modelling of the failure process of one repairable system.
- Only corrective maintenances, which may be imperfect.

Many imperfect maintenance models have been proposed.

To analyze a dataset, it is necessary to check whether these models are adapted or not $\Longrightarrow$ **Goodness-of-fit (GoF) tests**

GoF tests are well known for simple models AGAN and ABAO.

Very few work exist for testing the fit of imperfect maintenance models.

Aim of this work

Develop a methodology for testing the fit of any imperfect maintenance model.

Notations and assumptions

- $T_1, T_2, \ldots, T_n$ the n first **failure times** of the system, with $T_0 = 0$.
- Repair duration is considered negligible or not taken into account.
- $N = (N_t)_{t \geq 0}$ the failure **counting process**, characterized by its intensity

$$\lambda_t = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\left(N_{t+\Delta t} - N_{t^-} = 1 \mid \mathcal{H}_{t^-}\right), \quad t \geq 0,$$

where $\mathcal{H}_{t^-}$ is the past of the process just before t.

An **imperfect maintenance model** is composed of two parts:

- The **initial intensity** expresses the intrinsic wear before the first maintenance.
- A model for the **effect of maintenances**.

Initial intensity

$h(t)$ is the failure rate of the first failure time.

Usual models :

- **Power Law Process (PLP)** : $h(t) = abt^{b-1}, \quad t \geq 0, \; a, b > 0.$
  - $b > 1$: the system wears.
  - $b = 1$: the system is stable.
  - $b < 1$: the system improves.
- **Log Linear Process (LLP)**: $h(t) = \exp(a + bt), \quad t > 0, \; a, b \in \mathbb{R}.$
  - $b > 0$: the system wears.
  - $b = 0$: the system is stable.
  - $b < 0$: the system improves.

Initial intensity

$h(t)$ is the failure rate of the first failure time.

Usual models :

- **Power Law Process (PLP)** : $h(t) = abt^{b-1}, \quad t \geq 0, \; a, b > 0.$
  - $b > 1$: the system wears.
  - $b = 1$: the system is stable.
  - $b < 1$: the system improves.
- **Log Linear Process (LLP)**: $h(t) = \exp(a + bt), \quad t > 0, \; a, b \in \mathbb{R}.$
  - $b > 0$: the system wears.
  - $b = 0$: the system is stable.
  - $b < 0$: the system improves.

In this presentation, we consider wearing systems.

Maintenance effect

- **As Good As New** (perfect maintenance):
  N is a renewal process and

$$\lambda_t = h\left(t - T_{N_{t^-}}\right),\ t \geq 0$$

- **As Bad As Old** (minimal maintenance):
  N is a Non-Homogeneous Poisson Process (NHPP) and

$$\lambda_t = h(t),\ t \geq 0$$

- **Imperfect maintenance**: between ABAO and AGAN.

Some imperfect maintenance models

Brown-Proschan (BP) model (Brown & Proschan, 1983)

Let $p \in [0,1]$. Each maintenance is:

- AGAN with probability p,
- ABAO with probability $1 - p$.

The intensity is

$$\lambda_t = h\left(t - T_{N_{t^-}} + \sum_{j=1}^{N_{t^-}} \left(\prod_{k=j}^{N_{t^-}} (1 - B_k)\right) (T_j - T_{j-1})\right), \quad t \geq 0.$$

Quasi-renewal (QR) model (Wang & Pham, 1996)

Let $(Y_i)_{i=1,\ldots,n}$ be a sequence of i.i.d. random variables and $q > 0$ a parameter characterizing the repair effect. Then, under the QR model,

$$T_i - T_{i-1} = q^{i-1} Y_i, \quad i \in \mathbb{N}^*.$$

The times between two successive failures are independent and the counting process N is a **geometric process** (Lam, 1988).

The intensity is

$$\lambda_t = q^{-N_{t-}} h\left(q^{-N_{t-}} \left(t - T_{N_{t-}}\right)\right), \quad t \geq 0.$$

- $q = 1$: AGAN maintenance.
- $q \in]0, 1[$: stochastically decreasing inter-failure times.
- $q > 1$: system improvement.

$\Longrightarrow$ Geometrical growth of the inter-failure times is a **strong condition**.

Extended Geometric Process (EGP) (Bordes & Mercier, 2013)

Let $(Y_i)_{i=1,\dots,n}$ be a sequence of i.i.d. random variables and $q > 0$ a parameter characterizing the repair effect. Then, under the model,

$$T_i - T_{i-1} = q^{b_i} Y_i,$$

where $(b_i)_{i\in\mathbb{N}^*}$ is a non-decreasing sequence of non-negative real numbers such that

- $b_1 = 0$
- $\lim_{i\to\infty} b_i = \infty$.

For instance, for $i \in \mathbb{N}^*$,

- $b_i = i - 1$ (quasi-renewal case),
- $b_i = \sqrt{i-1}$,
- $b_i = \log(i)$.

Virtual age models (Kijima, 1989)

Let $(A_i)_{i=1,\ldots,n}$ be a sequence of positive random variables.

Model assumption : After the i^{th} maintenance, the system behaves like a new system which has not failed until A_i.

$\Longrightarrow$ The variables A_i are called **effective ages**.

The intensity is

$$\lambda_t = h\left(A_{N_{t^-}} + t - T_{N_{t^-}}\right), \ t \geq 0.$$

A virtual age model is defined by a particular expression of the effective ages.

ARA_∞ model (Doyen & Gaudoin, 2004) or Kijima type II

The repair is supposed to reduce the **effective age** by a factor $\rho \leq 1$:

$$\forall i \in \mathbb{N}^*, \quad A_i = (1-\rho)(A_{i-1} + T_i - T_{i-1}),$$

where $A_{i-1} + T_i - T_{i-1}$ is the age of the system just before the i^{th} maintenance and $A_0 = 0$.

The intensity is $\lambda_t = h\left(t - \rho \sum_{j=0}^{N_{t^-}-1} (1-\rho)^j T_{N_{t^-}-j}\right), \ t \geq 0.$

ARA_1 model (Doyen & Gaudoin, 2004) or Kijima type I

The **supplement of effective age** since the last failure is reduced by a factor $\rho \leq 1$:

$$\forall i \in \mathbb{N}^*, \quad A_i = A_{i-1} + (1-\rho)(T_i - T_{i-1}),$$

and $A_0 = 0$.

The intensity is $\lambda_t = h\left(t - \rho T_{N_{t^-}}\right), \ t \geq 0.$

Usually, parameter estimation is done by **likelihood maximization**.

$\Longrightarrow$ Assessment of both the **intrinsinc ageing** and the **repair effect**.

Likelihood function:

$$L_n = \left(\prod_{i=1}^{n} \lambda_{T_i} \right) \exp \left(- \int_0^{T_n} \lambda_t dt \right).$$

Construction of a goodness-of-fit test

Let $\mathcal{C} = \left\{\lambda(\theta), \theta \in \Theta \subset \mathbb{R}^d\right\}$ be an imperfect maintenance model, where θ is the model parameter.

Is $\mathcal{C}$ a relevant model for the observed data $T_1, \ldots, T_n$?

$\Longrightarrow$ **Goodness-of-fit test**: statistical test of

$$H_0 : \text{“}\lambda \in \mathcal{C}\text{''} \text{ vs } H_1 : \text{“}\lambda \notin \mathcal{C}\text{''}$$

Construction of a GoF test

1. **Find a statistic** expressing the gap between the data and the model.
2. **Determine the distribution** of the statistic under H_0.
3. **Compare** the observed statistic with a quantile of this distribution.

We propose 2 families of GoF tests, based on:

- Martingale residuals.
- Probability integral transforms.

Martingale

Let $\Lambda = (\Lambda_t)_{t \geq 0}$ be the cumulative intensity of the failure process N, with

$$\Lambda(t) = \int_0^t \lambda_s ds, \;\; t \geq 0.$$

Definition

The process $M = (M_t)_{t \geq 0}$ defined by $M = N - \Lambda$ is a zero mean **martingale**.

N is close to Λ in the sense that the expectation of their difference is null.

Martingale residuals

In our setting, the intensity has a parametric form $\lambda(\theta)$ with $\theta \in \Theta \subset \mathbb{R}^p$. Let us denote

- $\Lambda(\theta) = (\Lambda_t(\theta))_{t \geq 0}$ the cumulative intensity of the process, where

$$\Lambda_t(\theta) = \int_0^t \lambda_s(\theta) ds, \ \ t \geq 0.$$

- $\widehat{\theta}$ the maximum likelihood estimator of θ.

Definition

The **martingale residuals** are the random variables $(\widehat{M}_i)_{i=1,\ldots,n}$ such that

$$\widehat{M}_i = N(T_i) - \Lambda_{T_i}(\widehat{\theta}) = i - \Lambda_{T_i}(\widehat{\theta}).$$

When estimating θ, the martingale property is lost but N is still expected to be close to $\Lambda(\widehat{\theta})$.

Example: ARA_∞-PLP model

- initial intensity PLP: $h(t) = abt^{b-1}$.
- maintenance effect ARA_∞: $A_i = (1-\rho)(A_{i-1} + T_i - T_{i-1})$.

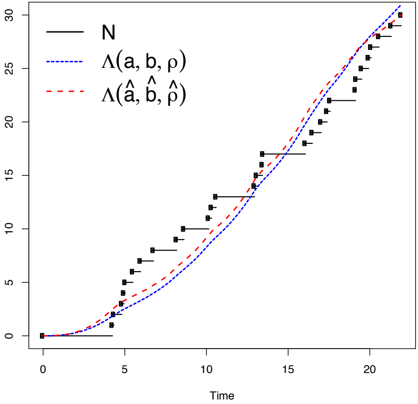
Cumulative intensity:

For $t \geq 0$,

$$\Lambda_t(a, b, \rho) = a \sum_{i=1}^{N_t+1} \left\{ \left(T_i - \rho \sum_{j=0}^{i-2} (1-\rho)^j T_{i-j-1} \right)^b - \left(T_{i-1} - \rho \sum_{j=0}^{i-2} (1-\rho)^j T_{i-j-1} \right)^b \right\},$$

where we set $T_{N_t+1} = t$.

3 parameters: $a > 0$, $b > 1$ and $\rho \in [0, 1]$.



Simulated dataset with $n = 30$ failure times from the ARA_∞-PLP model, with $a = 0.05$, $b = 2.5$ and $\rho = 0.1$.

The estimated cumulative intensity $\Lambda(\hat{\theta})$ is as close to the counting process as the real cumulative intensity $\Lambda(\theta)$.

Test statistics

We consider 3 statistics which measure discrepancies between N and $\Lambda(\hat{\theta})$.

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

$$KS_m(\hat{\theta}) = \sup_{i=1,\ldots,n} \left| \widehat{M}_i \right| = \sup_{i=1,\ldots,n} \left| i - \Lambda_{T_i}(\hat{\theta}) \right|.$$

- **Cramér-von Mises (CvM)**:

$$CvM_m(\hat{\theta}) = \int_0^{T_n} \left(N_t - \Lambda_t(\hat{\theta}) \right)^2 d\Lambda_t(\hat{\theta}).$$

One can show that:

$$CvM_m(\hat{\theta}) = -\frac{1}{3} \sum_{i=1}^{n} \left\{ \left(i - 1 - \Lambda_{T_i}(\hat{\theta}) \right)^3 - \left(i - 1 - \Lambda_{T_{i-1}}(\hat{\theta}) \right)^3 \right\}.$$

- **Anderson-Darling (AD)**:

$$AD_m(\hat{\theta}) = \int_0^{T_n} \frac{\left(N_t - \Lambda_t(\hat{\theta})\right)^2}{\Lambda_t(\hat{\theta})\left(n+1-\Lambda_t(\hat{\theta})\right)} d\Lambda_t(\hat{\theta}).$$

One can show that:

$$AD_m(\hat{\theta}) = \frac{1}{n+1}\sum_{i=2}^{n}\left\{(i-1)^2 \log\left(\frac{\Lambda_{T_i}(\hat{\theta})}{\Lambda_{T_{i-1}}(\hat{\theta})}\right) - (n+2-i)^2 \log\left(\frac{n+1-\Lambda_{T_i}(\hat{\theta})}{n+1-\Lambda_{T_{i-1}}(\hat{\theta})}\right)\right\}$$
$$+ (n+1)\log\left(1 - \frac{\Lambda_{T_1}(\hat{\theta})}{n+1}\right) - n.$$

Probability Integral Transform

Under H_0 : "$\lambda \in \mathcal{C}$", for $i = 0, \ldots, n-1$, the random variables

$$\Lambda_{T_{i+1}}(\theta) - \Lambda_{T_i}(\theta)$$

are i.i.d. and follow the $\mathcal{E}(1)$ distribution.

$\longrightarrow$ **Transformation into uniform variables.**

Definition

For $i = 0, \ldots, n-1$, let $S(\cdot \mid \mathbf{T}_i; \theta)$ denote the **reliability function** of the inter-failure time $T_{i+1} - T_i$ conditionally to $\mathbf{T}_i = (T_1, T_2, \ldots, T_i)$:

$$\begin{aligned} S(s \mid \mathbf{T}_i; \theta) :=& P(T_{i+1} - T_i > s \mid \mathbf{T}_i; \theta) \\ =& \exp\left(-\Lambda_{T_i+s}(\theta) + \Lambda_{T_i}(\theta)\right), \quad \text{for } s \geq 0. \end{aligned}$$

Definition - conditional Probability Integral Transform (PIT)

When applying this reliability function to the observed inter-failure times, we define the variables

$$U_i(\theta) = S(T_{i+1} - T_i \mid \mathbf{T}_i; \theta), \; i = 0, \dots, n-1.$$

This transform of the variables $T_{i+1} - T_i$ is also known as Rosenblatt's transform.

Under H_0, the U_i's are **i.i.d. with standard uniform distribution.**

In practice, θ is estimated and we will test the uniformity of $U_0(\hat{\theta}), \dots, U_{n-1}(\hat{\theta})$.

Test statistics based on the U_i

Let $F_{n,S}$ be the empirical c.d.f. of the random variables $U_i(\hat{\theta})$ and

$$U_{(0)}(\hat{\theta}) \leq U_{(1)}(\hat{\theta}) \leq \cdots \leq U_{(n-1)}(\hat{\theta}).$$

We propose 3 test statistics:

- **Kolmogorov-Smirnov (KS)**

$$KS_u(\hat{\theta}) = \sqrt{n} \sup_{x \in [0,1]} |F_{n,S}(x) - x|$$
$$= \sqrt{n} \max \left\{ \max_{i=1,\ldots,n} \left(\frac{i}{n} - U_{(i-1)}(\hat{\theta}) \right), \max_{i=1,\ldots,n} \left(U_{(i-1)}(\hat{\theta}) - \frac{i-1}{n} \right) \right\}.$$

- **Cramér-von Mises (CvM)**

$$CvM_u(\hat{\theta}) = n \int_0^1 \left(F_{n,S}(x) - x\right)^2 dx$$
$$= \sum_{i=1}^{n} \left(U_{(i-1)}(\hat{\theta}) - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n}.$$

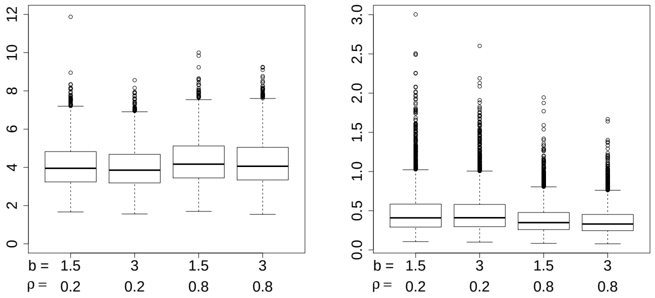
- **Anderson-Darling (AD)**

$$AD_u(\hat{\theta}) = n \int_0^1 \frac{\left(F_{n,S}(x) - x\right)^2}{x(1-x)} dx$$
$$= -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \left\{ \log \left(U_{(i-1)}(\hat{\theta}) \right) + \log \left(1 - U_{(n-i)}(\hat{\theta}) \right) \right\}.$$

Liu et al (Liu, Huang, & Zhang, 2012) proposed to perform a GoF test by comparing the value of $KS_u(\hat{\theta})$ to critical values that can be found in classical tables for testing the uniformity of a sample.

We believe that this approach is questionable because the estimation of θ should be taken into account: even under H_0, the $U_i(\hat{\theta})$ are neither independent nor uniformly distributed.

So do the distributions of the test statistics under H_0 depend on the model parameters?



(a) $KS_m(\hat{\theta})$ (b) $AD_u(\hat{\theta})$

Figure: Boxplots of 4000 simulated test statistics under the ARA_∞-PLP model for different values of b and ρ, $a = 0.05$ and $n = 30$.

This experiment seems to indicate that **the distributions of the test statistics under H_0 depend on the model parameters**.

So the usual approach can not be used to perform the tests.
$\Longrightarrow$ It is necessary to use a **parametric bootstrap** approach.

Parametric bootstrap

Let $Z(\hat{\theta})$ denote a generic test statistic. We use the **closeness of θ and $\hat{\theta}$** to approximate the distribution of $Z(\hat{\theta})$.

Algorithm

1. Compute the MLE $\hat{\theta}$ of θ and the statistic $Z(\hat{\theta})$ on the dataset $T_1, \ldots, T_n$.

2. For $i = 1$ until L,
   a. Generate $T^*_{1,i}, T^*_{2,i}, \ldots, T^*_{n,i}$ under the model of intensity $\lambda(\hat{\theta})$.
   b. Compute $\hat{\theta}^*_i$ the MLE of $\hat{\theta}$ from $T^*_{1,i}, \ldots, T^*_{n,i}$.
   c. Compute the statistic $Z^*_i = Z^*_i(\hat{\theta}^*_i)$ from $T^*_{1,i}, \ldots, T^*_{n,i}$ and $\hat{\theta}^*_i$.

3. The hypothesis H_0 is rejected at significance level α if $Z(\hat{\theta})$ is higher than the empirical quantile of order $1 - \alpha$ of $Z^*_1, \ldots, Z^*_L$.

- The asymptotic validity of the KS and CvM parametric bootstrap goodness-of-fit tests has been proved in the classical framework of **i.i.d. random variables** (Stute, Manteiga, & Quindmil, 1993).

- Results extended to the case of **independent random vectors** $\Longrightarrow$ Goodness-of-fit tests for copula models (Genest & Rémillard, 2008).

- In our case, theoretical results are difficult to obtain because of the **dependence** between $T_1, \ldots, T_n$.
  $\Longrightarrow$ Assessment of the validity of the approach by a simulation study.

Simulation design

The tests are performed on a huge number of simulated datasets. The power of a test is estimated by the percentage of rejection of the null hypothesis.

- Level of significance $\alpha = 0.05$.
- Tested null hypotheses H_0 : ARA_∞-PLP, ARA_1-PLP and QR-PLP.
- $n = 30$ failures.
- $M = 1500$ simulated datasets for each of the models:
  - **Brown-Proschan**, $p \in \{0.2, 0.8\}$,
  - **Extended Geometrical Process**, $b_i = \sqrt{i-1}$ and $q \in \{0.8, 0.9, 0.95\}$,
  - **Quasi-Renewal**, $q \in \{0.8, 0.9, 0.95\}$,
  - **ARA_1** or **ARA_∞**, $\rho \in \{0.2, 0.8\}$,
- with **PLP initial intensity**: $a = 0.05$, $b \in \{1.5, 2, 2.5, 3\}$ or **LLP initial intensity**: $a = -5$, $b \in \{0.005, 0.01, 0.05, 0.1\}$
- $L = 1000$ bootstrap repetitions.

Test of H_0 : ARA$_\infty$-PLP

Data simulated under the ARA$_\infty$-PLP model (H_0).

ρ	b	KS_m	CvM_m	AD_m	KS_u	CvM_u	AD_u
0.2	1.5	6.1	4.5	3.9	6.1	6.7	6.3
	2	4.9	3.8	3.9	6.1	5.9	6.1
	2.5	4.9	3.9	4.4	5.9	5.6	5.0
	3	4.9	4.3	4.5	5.5	5.9	5.7
0.8	1.5	5.1	4.8	4.6	3.5	3.7	4.1
	2	5.3	5.2	5.5	3.7	4.1	3.5
	2.5	4.5	5.0	5.1	6.1	5.6	5.4
	3	3.9	4.1	4.3	5.9	4.5	4.9

The empirical levels are close to the theoretical level $\alpha = 5\%$.

Test of H_0 : ARA$_\infty$-PLP

Data simulated under the EGP-PLP model.

b	q	KS$_m$	CvM$_m$	AD$_m$	KS$_u$	CvM$_u$	AD$_u$
1.5	0.8	22.2	32.3	**33.3**	17.5	**22.7**	20.7
	0.9	13.9	21.3	**21.9**	6.2	**7.2**	6.0
	0.95	8.6	11.5	**11.9**	3.3	3.7	**3.2**
2	0.8	58.2	70.9	**71.9**	22.1	**25.3**	24.9
	0.9	29.2	41.0	**41.6**	**6.7**	5.9	5.9
	0.95	11.4	**17.6**	17.4	9.5	**10.6**	10.4
2.5	0.8	81.7	89.7	**90.1**	17.5	19.9	**22.2**
	0.9	30.4	42.8	**43.4**	16.2	19.9	**21.1**
	0.95	9.9	16.4	**16.8**	26.3	32.1	**33.0**

- The results depend strongly on the value of the parameters. When q tends to 1, the model gets closer to a renewal process, so H_0 is less rejected.
- Some tests are biased.
- AD$_m$ is clearly the best test.

Conclusion of the simulation study

- Recommendation: perform the **AD_m** and **AD_u** tests.
- **Difficulty to distinguish** between the ARA_∞ and the BP models that are close to a renewal process (Last & Szekli, 1998).
- The ARA_1-PLP and QR-PLP models seem to be **very flexible** and are **hardly ever rejected**.
- Tests of the ARA-PLP models are not powerful when data are simulated with an initial intensity of type LLP
$\Longrightarrow$ Tests more able to **detect a discrepancy in the repair effect** than in the shape of the intrinsic wear.
- Previous remark does not hold for QR models.
- On the whole, the powers are not very high but $n = 30$. We have observed an **increase in power** when setting $n = 100$.

EDF data

- 4 identical and independent systems in 4 different EDF coal-fired power stations: S_1, S_2, S_3 and S_4.
- Maintenance times collected during 9 years.
- The tables give the p-values of the tests.
- We have also computed the AIC criterion:

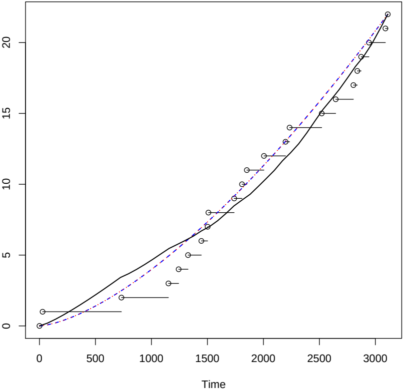
$$AIC = -2 \max_{\theta \in \Theta} \log \left(L_n(\theta) \right) + 2d.$$

where d is the number of estimated parameters (here $d = 3$). The best model among those considered is the one that minimizes the AIC.

EDF data - Study of S_1

$n = 22$

Model	ARA_∞	ARA_1	QR
$\hat{a}$	1.2×10^{-4}	1.2×10^{-4}	0.0013
$\hat{b}$	1.51	1.51	1.18
$\hat{\rho}/\hat{q}$	7.6×10^{-6}	7.3×10^{-5}	0.94
KS_m	0.95	0.94	0.47
CvM_m	0.82	0.60	0.46
AD_m	0.58	0.41	0.38
KS_u	0.48	0.44	0.09
CvM_u	0.55	0.51	0.11
AD_u	0.57	0.55	0.21
AIC	264.64	264.64	**263.69**

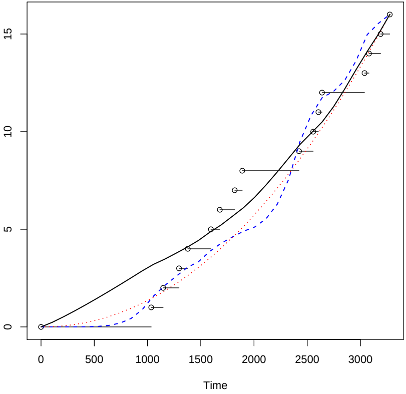


$\Longrightarrow$ **None of the 3 models is rejected.** QR is considered as the best model.

EDF data - Study of S_3

$n = 16$

Model	ARA_∞	ARA_1	QR
$\hat{a}$	9.2×10^{-19}	7.6×10^{-7}	0.0012
$\hat{b}$	6.14	2.08	1.14
$\hat{\rho}/\hat{q}$	0.16	5.0×10^{-6}	0.92
KS_m	0.13	0.33	0.38
CvM_m	0.39	0.63	0.43
AD_m	0.34	0.65	0.31
KS_u	0.17	0.4	0.10
CvM_u	0.05	0.06	0.01
AD_u	0.09	0.06	0.01
AIC	**194.82**	201.43	205.00



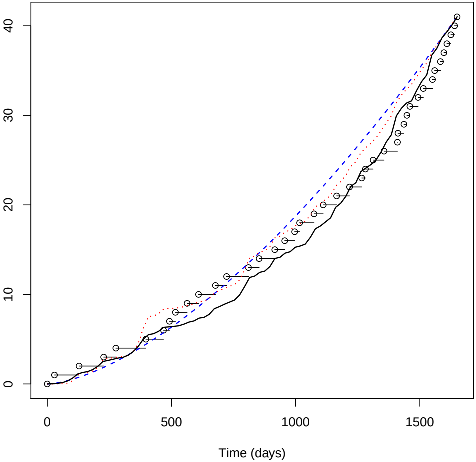
$\Longrightarrow$ **All models are rejected.** Possible explanation: data exhibit long time periods without any failure. This phenomenon can not be captured by these models.

Photocopier data

- Maintenance times of a photocopier (Murthy, Xie, & Jiang, 2003).
- $n = 42$ maintenances during the first 4 and a half year of service.

Model	ARA_∞	ARA_1	QR
$\hat{a}$	3.9×10^{-4}	3.7×10^{-9}	2.3×10^{-5}
$\hat{b}$	1.6	4.3	2.3
$\hat{\rho}/\hat{q}$	1.0×10^{-6}	0.95	0.96
KS_m	0.31	0.19	0.44
CvM_m	0.26	0.12	0.38
AD_m	0.24	0.09	0.31
KS_u	< 0.01	0.42	0.45
CvM_u	< 0.01	0.54	0.61
AD_u	< 0.01	0.60	0.58
AIC	380.33	357.52	**349.74**

$\Longrightarrow$ Rejection of ARA$_\infty$. We recommend to use QR.



Conclusion

- We have proposed **2 classes of goodness-of-fit tests**, based on:
  - martingale residuals
  - conditional PIT of the interfailure times.

  Evaluation of the quantiles of the test statistics under the null hypothesis are made with a **parametric bootstrap** procedure.[1]

- In both families, the **Anderson-Darling tests** performed well in most simulated cases and we recommend their use in practice.

- Methodology completely general: can be applied on **any imperfect maintenance model**.

- Very difficult to discriminate between the usual imperfect maintenance models. Models are **very flexible** and can be adapted to a **broad range of data**. $\Longrightarrow$ **Moderate powers** of the tests.

[1] Chauvel C., Dauxois, J.-Y., Doyen, L. and Gaudoin, O. *Parametric bootstrap goodness-of-fit tests for imperfect maintenance models.* Submitted. 2015.

Future work

- Obtain a **theoretical validation** of the bootstrap procedure.
  Method validated by a simulation study.

- Build a test of which statistic is a **combination** of the AD_m and AD_u statistics, such as the maximum or a weighted sum.

- Use **other statistics** than Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling statistics.

- Find other tests that do not require the bootstrap procedure.

- Obtain **asymptotic results** for the statistics in order to know their distribution functions and **apply other kinds of tests**.

- Incorporate **preventive maintenances** as well as study the case where **several identical systems** are observed in parallel.

References I

Bordes, L., & Mercier, S. (2013). Extended geometric processes: semiparametric estimation and application to reliability. *Journal of the Iranian Statistical Society*, *12*, 1-34.

Brown, F., & Proschan, F. (1983). Imperfect repair. *Journal of Applied Probability*, *20*, 851-859.

Doyen, L., & Gaudoin, O. (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering and System Safety*, *84*, 45-56.

Genest, C., & Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'IHP*, *44*, 1096-1127.

Kijima, M. (1989). Some results for repairable systems with general repair. *Journal of Applied Probability*, *26*, 89-102.

Lam, Y. (1988). Geometric processes and replacement problem. *Acta Mathematicae Applicatae*, *4*, 366-377.

References II

Last, G., & Szekli, R. (1998). Asymptotic and monotonicity properties of some repairable systems. *Advances in Applied Probability*, *30*, 1089-1110.

Liu, Y., Huang, H.-Z., & Zhang, X. (2012). A data-driven approach to selecting imperfect maintenance models. *IEEE Transactions on reliability*, *61*, 101-112.

Murthy, D. N. P., Xie, M., & Jiang, R. (2003). *Weibull models*. Wiley.

Stute, W., Manteiga, W. G., & Quindmil, M. P. (1993). Bootstrap based goodness-of-fit-tests. *Metrika*, *40*, 243-256.

Wang, H., & Pham, H. (1996). A quasi renewal process and its application in imperfect maintenance. *International Journal of Systems Science*, *27*, 1055-1062.

Thank you for your attention.