Classical and Bayesian goodness-of-fit tests for the exponential model: A comparative study

Maria J. Polidoro[a,d], Fernando J. Magalhães[b,d] and M. A. Amaral Turkman[c,d]

[a]*ESTGF, Polytechnic Institute of Porto, Portugal*;
[b]*ISCAP, Polytechnic Institute of Porto, Portugal*;
[c]*Department of Statistics and Operations Research, Faculty of Sciences, Lisbon University, Portugal*;
[d]*CEAUL-Center of Statistics and Aplications of Lisbon University, Portugal*

Abstract

Most common statistical methodologies assume a parametric model for the data and inference is made based on that assumption. If the model does not fit the data, the resulting inference will be mislead. Thus, evaluation of the fitting of a proposed parametric statistical model to a given dataset becomes an important issue.

In several practical situations, namely in reliability and life sciences problems, the exponential model has been widely used and several classical tests were already proposed for that purpose. In this work we suggest two Bayesian tests when an exponential model is proposed to describe the data, and using a simulation study, we compare their power with the classical ones.

Keywords: goodness-of-fit test; Bayesian nonparametric model; Bayes factor; mixture of finite Polya trees; power of test

AMS Subject Classification: 62G10; 62F15

1 Introduction

In statistical methodology it is often necessary to find a suitable model for the data under study. An important issue in modeling is to evaluate the fitting of the proposed parametric statistical model to the given dataset.

In many problems, namely in reliability engineering and life sciences, the exponential model is widely used; however, a wrong choice of the parametric model may mislead the statistical inference and hence it is important to find the best available test to evaluate the fit.

In a classical approach, formal methods, often called goodness-of-fit tests, involve the test of a null hypothesis where the parametric model is defined without the specification of an alternative hypothesis or alternative model. There are many kinds of goodness-of-fit tests in literature (see, e.g., D'Agostino and Stephens (1986)). Some of them are very specific, but others are quite broad tests which are applicable to general cases. The most commonly used are Pearson's chi-squared test and those based on the empirical distribution function.

In the last decades, the interest in the problem of goodness-of-fit test for the exponential case has increased and new test statistics emerged (see, e.g., Baringhaus and Henze (1991, 2000); Choi et al. (2004); Henze and Meintanis (2005); Grané and Fortiana (2011), and references therein).

The Bayesian foundation for fit evaluation is conveyed by the predictive distribution through exploratory methods or by using formal posterior predictive model checks, like Bayesian p-values (see, e.g., Gelman et al. (1996); Robins et al. (2000); Bayarri and Berger (2000); Hjort et al. (2006)). More recently, Johnson (2004, 2007) proposed a Bayesian chi-squared goodness-of-fit test of a parametric model by generalizing the classical Pearson's chi-squared statistic and discussed the use of pivotal test statistics.

An enhanced Bayesian nonparametric alternative consists on embedding the proposed parametric model (H_0) in an alternative nonparametric model (H_1). To validate the proposed model the parametric fit is compared with the nonparametric one, using the Bayes factor as a measure of evidence against H_0, based on the observed values.

These testing problems require the formulation of Bayesian nonparametric models (see, for example, the book by Hjort et al. (2010), for some discussion on the subject). Bayesian literature on nonparametric goodness-of-fit tests is still scarce. For a continuous density function, particularly for the normal density, Verdinelli and Wasserman (1998), Berger and Guglielmi (2001) and Tokdar and Martin (2011) proposed a Bayesian nonparametric goodness-of-fit test, which assigns a mixture of Gaussian processes, a mixture of Polya trees and a Dirichlet process mixture, respectively, for the alternative model. All the three tests allow calculation of the Bayes factor, however only the alternative model based on a mixture of Polya trees is computationally more accessible and intuitively simple.

In this work we suggest the Bayesian nonparametric goodness-of-fit test of Berger and Guglielmi (2001) and the Bayesian chi-squared test of Johnson (2004), to evaluate the goodness-of-fit in the exponential case and we compare the power of these tests with some classical test statistics.

The paper is organized as follows: in Section 2 we review some definitions on nonparametric Bayesian statistical models as well as the Bayesian nonparametric test and the Bayesian chi-squared goodness-of-fit test. In Section 3 we describe some of the classical test statistics and the Bayesian tests for the problem of testing the adequacy of an exponential model. In Section 4 we carry out a simulation study to compare the power of the different proposed tests. Finally, in Section 5, conclusions are stated based on the obtained simulation results.

2 Bayesian approach

Let $(X_1, X_2 \ldots, X_n)$ be a vector of continuous, identically distributed, conditionally independent observations drawn from a probability density function $f(x|\theta)$ defined with respect to Lebesgue measure and indexed by an s-dimensional parameter vector $\theta \in \Theta \subset \mathbb{R}^s$, which is unknown. Our goal is to test the adequacy of the assumed model $f(x|\theta)$ based on the observed data $(x_1, x_2 \ldots, x_n)$. The assumed model $f(x|\theta)$ is the "null" model or "null" hypothesis, represented by H_0. For the Bayesian formulation, if an "alternative" hypothesis, H_1 is to be specified, it will be associated to a nonparametric Bayesian model named by G.

2.1 Nonparametric Bayesian model

In a nonparametric Bayesian context, the random sample is defined by an unknown random probability measure G, and the goal is to place a prior directly on the class of random probability measures. Lavine (1992, 1994) proposed Polya trees as an useful nonparametric prior distribution for random probability measures G on the sample space Ω of the random variable X. Reference papers on Polya trees are also Hanson and Johnson (2002) and Hanson (2006). The Polya tree can be easily constructed as follows.

A finite Polya tree for a distribution G is built by dividing the sample space Ω, into a sequence of ever finer partitions. Let $\{B_0, B_1\}$ be a measurable partition of Ω at the first level; then, let $\{B_{00}, B_{01}\}$ and $\{B_{10}, B_{11}\}$ be measurable partitions of B_0 and B_1, respectively, at the second level; continue in this way until M levels (i.e. $m = 1, 2, \ldots, M$) are achieved and call the set of measurable partitions a finite binary tree partition of Ω. Let, at the m-th level, $\varepsilon_{1:m} = \varepsilon_1\varepsilon_2 \cdots \varepsilon_m$, with each $\varepsilon_j \in \{0, 1\}$, $j = 1, 2, \ldots, m$, so that each $\varepsilon_{1:m}$ defines a unique subset $B_{\varepsilon_{1:m}}$. It is clear that the number of partitions at the m-th level is 2^m and that $B_{\varepsilon_{1:m}}$ will split into $B_{\varepsilon_{1:m}0}$ and $B_{\varepsilon_{1:m}1}$ at level $(m + 1)$. The finite binary tree partition structure of the Polya tree can be denoted by $\Pi = \{B_{\varepsilon_{1:m}}, m = 1, 2, \ldots, M\}$.

In order to define random measures on Ω we construct random measures on the sets $B_{\varepsilon_{1:m}}$ for $m = 1, 2, \ldots, M$. Starting at Ω, we can move into B_0, with probability Y_0, or into B_1, with probability $Y_1 = 1 - Y_0$. Generally, on entering $B_{\varepsilon_{1:m}}$, we can either move into $B_{\varepsilon_{1:m}0}$, with conditional probability $Y_{\varepsilon_{1:m}0}$ or into $B_{\varepsilon_{1:m}1}$ with conditional probability $Y_{\varepsilon_{1:m}1} = 1 - Y_{\varepsilon_{1:m}0}$. The marginal probability of a subset in the m-th partition is

$$G(B_{\varepsilon_{1:m}}) = \left(\prod_{j=1,\varepsilon_j=0}^{m} Y_{\varepsilon_1\cdots\varepsilon_{j-1}0}\right)\left(\prod_{j=1,\varepsilon_j=1}^{m} (1 - Y_{\varepsilon_1\cdots\varepsilon_{j-1}0})\right)$$

with the marginal probability for the first level, i.e. for $j = 1$, being given by Y_0 or $1 - Y_0$.

For instance, for $m = 2$, $G(B_{00}) = Y_0Y_{00}$, $G(B_{01}) = Y_0(1-Y_{00})$, $G(B_{10}) = (1-Y_0)Y_{10}$ and $G(B_{11}) = (1 - Y_0)(1 - Y_{10})$. In Polya trees, these probabilities are random and independent Beta variables, i.e., $Y_0 \sim \mathrm{Beta}(\alpha_0, \alpha_1)$ and for every $\varepsilon_{1:m}$, $Y_{\varepsilon_{1:m}0} \overset{\text{ind}}{\sim} \mathrm{Beta}(\alpha_{\varepsilon_{1:m}0}, \alpha_{\varepsilon_{1:m}1})$, with nonnegative parameters $\alpha_{\varepsilon_{1:m}0}$ and $\alpha_{\varepsilon_{1:m}1}$. Denoting the collection of parameters α's by $\mathcal{A} = \{\alpha_{\varepsilon_{1:m}}, m = 1, 2, \ldots, M\}$, the particular finite Polya tree distribution with M levels is defined by the partitions in Π and the Beta parameters in $\mathcal{A}$, and is denoted by $G \sim \mathrm{FPT}_M(\Pi, \mathcal{A})$.

The parameters of the Polya tree can be chosen such that G is absolutely continuous with probability one. In particular, any $\alpha_{\varepsilon_{1:m}} = \rho(m)$ such that $\sum_{m=1}^{\infty} \rho(m)^{-1} < \infty$ guarantees G to be absolutely continuous (Schervish, 1995). For example, Berger and Guglielmi (2001) considered $\alpha_{\varepsilon_{1:m}} = c\rho(m)$, $c > 0$ and $\rho(m) = m^2$, m^3, 2^m, 4^m and 8^m.

By defining Π and $\mathcal{A}$ the Polya tree can be centered on some particular parametric distribution $f(x|\theta)$ on Ω, so that $E[G(B_{\varepsilon_{1:m}})|\theta] = F_\theta(B_{\varepsilon_{1:m}}) = \Pr(X \in B_{\varepsilon_{1:m}}|\theta)$, where θ is the parameter vector of the parametric distribution. One way of doing it is to fix a nested partition sequence Π, not depending on θ, and then choose the parameters $\alpha_{\varepsilon_{1:m}}$'s depending on θ (Berger and Guglielmi (2001)).

For instance, if $X \in \mathbb{R}^+$, define $B_0 = (0, F_{\hat{\theta}}^{-1}(0.5)]$, $B_1 = (F_{\hat{\theta}}^{-1}(0.5), +\infty)$ and more generally, at level m,

$$B_{\varepsilon_{1:m}} = \left\{\left(F_{\hat{\theta}}^{-1}\left(\frac{k-1}{2^m}\right), F_{\hat{\theta}}^{-1}\left(\frac{k}{2^m}\right)\right],\ m = 1, 2, \ldots, M,\ k = 1, 2, \ldots, 2^m\right\},$$

where $F^{-1}_{\hat{\theta}}(\cdot)$ are quantiles of $f(x|\theta)$ substituting θ by its m.l.e. vector, $\hat{\theta}$. Then, for $h > 0$, define

$$\alpha_{\varepsilon_{1:m-1}0}(\theta) = h^{-1}\rho(m)\left(\frac{F_\theta(B_{\varepsilon_{1:m-1}0})}{F_\theta(B_{\varepsilon_{1:m-1}1})}\right)^{1/2} \tag{1}$$

and

$$\alpha_{\varepsilon_{1:m-1}1}(\theta) = h^{-1}\rho(m)\left(\frac{F_\theta(B_{\varepsilon_{1:m-1}1})}{F_\theta(B_{\varepsilon_{1:m-1}0})}\right)^{1/2}. \tag{2}$$

For example, since $G(B_0|\theta) = Y_0 \sim \text{Beta}(\alpha_0(\theta), \alpha_1(\theta))$, then

$$E[G(B_0)|\theta] = E[Y_0|\theta] = \frac{\alpha_0(\theta)}{\alpha_0(\theta) + \alpha_1(\theta)} = F_\theta(B_0),$$

and for any $B_{\varepsilon_{1:m}} \in \Pi$, $E[G(B_{\varepsilon_{1:m}})|\theta] = F_\theta(B_{\varepsilon_{1:m}})$.

The function $\rho(\cdot)$ controls the smoothness of the distribution upon which the Polya tree distribution concentrates its mass and h refers to an overall scale factor which, in some sense, controls the overall variance of the Polya tree about its mean which is the parametric distribution. Further details on these two measures will be referred to later.

Finally, uncertainty about θ can also be modeled as $\pi(\theta)$, generating a mixture of finite Polya trees distribution for G. The notation $G|\Pi, \mathcal{A}_\theta \sim \text{MFPT}_M(\Pi, \mathcal{A}_\theta)$ is used to denote that G has a mixture of finite Polya trees prior distribution with M levels, fixed partition Π and the remaining Polya tree parameters, $\mathcal{A}_\theta$, are updated throughout the procedure.

2.2 Berger and Guglielmi's Bayesian nonparametric goodness-of-fit test

Berger and Guglielmi's Bayesian nonparametric goodness-of-fit test is defined by $H_0 : X \sim f(x|\theta), \theta \in \Theta$, versus $H_1 : X \sim G|\Pi, \mathcal{A}_\theta, \theta \in \Theta$, with a prior density, $\pi(\theta)$, usually noninformative. The test is performed based on the Bayes factor for H_0 against H_1, given by

$$\text{BF}_{01}(\boldsymbol{x}) = \frac{p(\boldsymbol{x}|H_0)}{p(\boldsymbol{x}|H_1)} = \frac{\int_\Theta f(\boldsymbol{x}|\theta)\pi(\theta)d\theta}{\int_\Theta p(\boldsymbol{x}|\theta)\pi(\theta)d\theta}, \tag{3}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, $f(\boldsymbol{x}|\theta) = \prod_{i=1}^n f(x_i|\theta)$, and $p(\boldsymbol{x}|\theta)$ is the marginal density of the sample under a Polya tree scheme, which following Lavine (1992), is given by

$$p(\boldsymbol{x}|\theta) = f(\boldsymbol{x}|\theta)\psi(\theta), \tag{4}$$

with

$$\psi(\theta) = \prod_{j=2}^{n}\prod_{m=1}^{m^*(x_j)} \frac{\alpha'_{\varepsilon_{1:m}(x_j)}(\theta)\left(\alpha_{\varepsilon_{1:m-1}0(x_j)}(\theta) + \alpha_{\varepsilon_{1:m-1}1(x_j)}(\theta)\right)}{\alpha_{\varepsilon_{1:m}}(x_j)(\theta)\left(\alpha'_{\varepsilon_{m-1}0(x_j)}(\theta) + \alpha'_{\varepsilon_{m-1}1(x_j)}(\theta)\right)}, \tag{5}$$

where $\varepsilon_{1:m}(x_j)$ is the index $\varepsilon_1\varepsilon_2\cdots\varepsilon_m$ such that x_j belongs to $B_{\varepsilon_1\cdots\varepsilon_m}$, for each level m, and $\alpha'_{\varepsilon_{1:m}(x_j)}(\theta)$ is equal to $\alpha_{\varepsilon_{1:m}(x_j)}(\theta)$ plus the number of observations among $\{x_1, x_2, \ldots, x_{j-1}\}$ which belong to $B_{\varepsilon_1\cdots\varepsilon_m(x_j)}$. For each x_j, the product in Equation (5) is up to the smallest level $m^*(x_j)$, such that no x_i, $i < j$, belongs to $B_{\varepsilon_{1:m}(x_j)}$.

The Bayes factor measures the evidence in favor of the null model against the alternative, based on observed values $\boldsymbol{x}$, i.e., values of $\text{BF}_{01}(\boldsymbol{x})$ larger or smaller than one are interpreted as evidence given by the data, respectively, in favor or against H_0. More formally, let c_{BF} be a fixed threshold that controls type I error rate; then for any given data set $\boldsymbol{x}$, if $\mathrm{BF}_{01}(\boldsymbol{x}) > c_{BF}$ ($<$) then we accept (reject) the null model.

The computation of the Bayes factor is simplified because by (4), (3) can be written as

$$\mathrm{BF}_{01}(\boldsymbol{x}) = \left(\int_{\Theta} \psi(\theta)\pi(\theta|\boldsymbol{x})d\theta \right)^{-1},$$

where $\pi(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)\pi(\theta)/p(\boldsymbol{x}|H_0)$, i.e., the Bayes factor can be written as the inverse of a posterior expectation of $\psi(\theta)$ under H_0. Thus, generating a sample $(\theta_1, \theta_2, \ldots, \theta_L)$ from $\pi(\theta|\boldsymbol{x})$, an approximation of the Bayes factor is given by

$$\widehat{\mathrm{BF}}_{01}(\boldsymbol{x}) = \frac{L}{\sum_{l=1}^{L} \psi(\theta_l)}. \tag{6}$$

Berger and Guglielmi (2001) proposed examining a plot of the Bayes factor as a function of the scale factor h, because it determines how concentrated the defined mixture of finite Polya trees prior distribution (MFPT) is about its mean, F_θ. As $h \to 0$, the MFPT concentrates both in terms of similarity in shape and distance from the fixed F_θ, and the Bayes factor will converge to one. As $h \to \infty$ the MFPT will, usually, be dispersed from the fixed F_θ, and the Bayes factor will be quite large. Between these two extremes, the Bayes factor will sometimes increase with h, but also will first decrease and then increase. Thus, they minimize over h to make a conservative choice of the Bayes factor (Tokdar et al. (2010)).

2.3 Johnson's Bayesian chi-squared test

Denote by $F(x|\theta)$ the cumulative distribution corresponding to the density $f(x|\theta)$. Let $\check{\theta}$ be a random observation from the posterior distribution of θ, and let $0 \equiv a_0 < a_1 < \ldots < a_k \equiv 1$ with $p_k = a_k - a_{k-1}$, $k = 1, 2, \ldots, K$ be a partition of $[0, 1]$. Usually, a noninformative prior distribution is specified on θ and is recommended to use $K \simeq n^{0.4}$ bins. Then Johnson's Bayesian chi-squared test statistic is given by

$$R_n^B(\check{\theta}) = \sum_{k=1}^{K} \frac{(m_k(\check{\theta}) - np_k)^2}{np_k}, \tag{7}$$

where $m_k(\check{\theta})$ represents the number of observations that fell into the k-th bin, i.e., the number of x_i's satisfying $F(x_i|\check{\theta}) \in (a_{k-1}, a_k]$, $i = 1, 2, \ldots, n$. Johnson (2004) proves that under the null hypothesis, $R_n^B(\check{\theta})$ is asymptotically distributed as $\chi^2_{(K-1)}$, independently of the dimension of the underlying parameter vector θ.

3 Tests for exponentiality

In this Section, preliminary notation is provided and classical and Bayesian tests for exponentiality are introduced.

A non-negative random variable X denoting the time to failure (or lifetime) of some item of interest has an exponential distribution, $X \sim \mathrm{Exp}(\lambda)$, if its density is given by

$$f(x|\lambda) = \lambda \exp(-\lambda x), \quad x \geq 0$$

where $\lambda > 0$ defines the unknown failure rate.

The problem consists in testing the null hypothesis

$$H_0 : X \sim \text{Exp}(\lambda) \text{ for some } \lambda > 0,$$

against the general alternative that X is not exponentially distributed, based on a sample $(x_1, x_2 \ldots, x_n)$ of identical distributed and conditional independent observations of X,.

3.1 Classical tests

A large number of classical tests for exponentiality have been proposed in the literature. The tests are based on different characteristics of the exponential distribution, and can be classified into several categories. According to Henze and Meintanis (2005) there are: tests based on the empirical distribution function; tests based on the integrated empirical distribution function; tests based on spacings and the Gini index; tests based on the entropy characterization; tests based on the statistic of Cox and Oakes; tests based on a characterization via the mean residual life function; tests statistics derived from the empirical Laplace transform; test statistics derived from the empirical characteristic function, among others. Henze and Meintanis (2005) compared twenty one test statistics for the exponentiality against eighteen alternative distributions. This exhaustive study concludes that there is no test statistic which is better than the others in terms of power. However, the study indicates that the Cox and Oakes (1984) statistic CO_n, the Epps and Pulley (1986) statistic EP_n, the modified Cramér-von Mises type statistic $\overline{\text{CM}}_n$ of Baringhaus and Henze (2000), based on a characterization of the mean residual life function, and the Baringhaus and Henze (1991) classical test statistic BH_n, based on the empirical Laplace transform, are among the most powerful test statistics and they are quite easy to evaluate. Note that the latter test statistic depends on an arbitrary constant a which, according to the studies, is now taken as $a = 1$. The Anderson and Darling (1954) test statistic AD_n was not considered in the study by Henze and Meintanis (2005) but it is included in our study because of its wide use in literature.

In this work, the above mentioned five classical test statistics will be used and compared in terms of power, with the Bayesian tests referred to in the previous Section. With that purpose in mind, each one of these classical test statistics is defined following closely the notation in Henze and Meintanis (2005). We just present the expressions of the corresponding tests statistics used to evaluate the adequacy of the exponential distribution; all theoretical details can be found in the literature.

Let $Y_i = X_i/\bar{X}$ and $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$. Then,

1. The test statistic of Cox and Oakes (1984) is given by

$$\text{CO}_n \quad = n + \textstyle\sum_{i=1}^{n} (1 - Y_i) \log(Y_i) \ .$$

2. The normalized Epps and Pulley (1986) test statistics is defined by

$$\text{EP}_n \quad = (48n)^{1/2} + \left[\frac{1}{n} \sum_{i=1}^{n} \exp(-Y_i) - \frac{1}{2} \right] \ .$$

3. The Cramér-von Mises type statistic $\overline{\text{CM}}_C$ of Baringhaus and Henze (2000) is computed as

$$\overline{\text{CM}}_n \quad = \frac{1}{n}\sum_{i,k=1}^{n}\left(2-3\text{e}^{-\min(Y_i,Y_k)}-2\min(Y_i,Y_k)(\text{e}^{-Y_i}+\text{e}^{-Y_k})+2\text{e}^{-\max(Y_i,Y_k)}\right) .$$

4. Baringhaus and Henze (1991) test statistic is defined by

$$\text{BH}_n \quad = \frac{1}{n}\sum_{i,k=1}^{n}\left[\frac{(1-Y_i)(1-Y_k)}{Y_i+Y_k+1}-\frac{Y_i+Y_k}{(Y_i+Y_k+1)^2}+\frac{2Y_iY_k}{(Y_i+Y_k+1)^2}+\frac{2Y_iY_k}{(Y_i+Y_k+1)^3}\right] .$$

5. The Anderson and Darling (1954) test statistic is given by

$$\text{AD}_n \quad = -n-\frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log(W_{(i)})+\log(1-W_{(n-i+1)})\right] ,$$

where $W_{(i)} = 1-\exp(-Y_{(i)})$, $1 \leq i \leq n$, and $Y_{(1)} \leq Y_{(2)} \cdots \leq Y_{(n)}$ are the order statistics of $(Y_1, Y_2, \ldots, Y_n)$.

For the first two test statistics, Henze and Meintanis (2005) proved that under the null hypothesis, EP_n and $\text{CO}_n^* = \left(\frac{6}{n}\right)^{1/2}\left(\frac{\text{CO}_n}{\pi}\right)$ are assymptotically standard normal distributed, therefore we reject the null hypothesis (that the observations are from an exponential distribution) for large values of $|\text{EP}_n|$ and $|\text{CO}_n^*|$. For the other three test statistics, the distribution under the null hypothesis has not been obtained analytically. To determine the critical values (empirical quantiles) of the respective test statistic distribution, Monte Carlo simulations are employed. The null hypothesis is rejected if the observed value of each one of these test statistics exceeds the corresponding critical value.

In order to obtain the empirical quantiles of each test statistic, 100,000 random samples of size n are generated from the standard exponential distribution and the value of the test statistic is calculated. The sample sizes considered are $n = 25$, $n = 50$ and $n = 100$, and the chosen significance levels are $\alpha = 0.1$, $\alpha = 0.05$ and $\alpha = 0.025$. The empirical critical value for the three classical statistics, $\overline{\text{CM}}_n$, BH_n and AD_n, presented in Table 1 for each n, is determined with the quantile $(1-\alpha)\times 100\%$ from the corresponding empirical distribution (for BH_n, the empirical critical values are very close to the ones obtain in Baringhaus and Henze (2000)).

Table 1: Empirical critical values of $\overline{\text{CM}}_n$, BH_n and AD_n

	$\overline{\text{CM}}_n$			BH_n			AD_n		
α	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
$n = 25$	0.341	0.451	0.564	0.219	0.304	0.385	1.042	1.292	1.562
$n = 50$	0.344	0.455	0.567	0.220	0.305	0.396	1.053	1.312	1.573
$n = 100$	0.348	0.464	0.583	0.221	0.311	0.401	1.058	1.325	1.595

3.2 Bayesian tests

Let us begin by defining the test based on Berger and Guglielmi (2001) which assumes two models, a parametric one and a nonparametric one, such that the former is embedded in the latter. The parametric Bayesian model is given by

$$X_i|\lambda \overset{\text{iid}}{\sim} \text{Exp}(\lambda), \text{ for } i = 1, 2, \ldots, n,$$
$$\lambda \sim \pi(\lambda)$$

and the nonparametric Bayesian model is

$$X_1, X_2, \ldots, X_n|G \overset{\text{iid}}{\sim} G$$
$$G|\Pi, \mathcal{A}_\lambda \sim \text{MFPT}_M(\Pi, \mathcal{A}_\lambda).$$
$$\lambda \sim \pi(\lambda)$$

where $\text{MFPT}_M(\Pi, \mathcal{A}_\lambda)$ defines a mixture of finite Polya trees prior distribution, with parameters $(\Pi, \mathcal{A}_\lambda)$ and M pre-specified levels. We consider a conjugate noninformative prior distribution for the unknown parameter λ, i.e., $\lambda \sim \text{Gamma}(a, b)$, with $a, b \to 0$.

The binary partitions, which are fixed not depending on λ, are given by

$$B_{\varepsilon_{1:m}} = \left\{ \left(F_{\hat{\lambda}}^{-1}\left(\frac{k-1}{2^m}\right), F_{\hat{\lambda}}^{-1}\left(\frac{k}{2^m}\right) \right], \; m = 1, 2, \ldots, M, \; k = 1, 2, \ldots, 2^m \right\},$$

where $F_{\hat{\lambda}}^{-1}(\cdot)$ defines the quantiles of the exponential distribution taking as parameter value its m.l.e., $\hat{\lambda} = 1/\bar{x}$, with $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$.

The values for the parameters $\alpha_{\varepsilon_{1:m}}(\lambda)$ are obtained using Equations (1) and (2) with $\rho(m) = 4^m$. As h must take a range of values near to and far from zero, a suggestion by Tokdar and Martin (2011) is considered. This suggestion proposes evaluating Bayes factor for 13 values of h in the interval $[2^{-6}, 2^6]$.

For the Johnson's Bayesian chi-squared test we use the test statistic defined in Equation (7) and a fixed significance level $\alpha = 0.05$. We substitute $\check{\theta}$ by $\check{\lambda}$, a random observation from the posterior distribution considering a conjugate noninformative prior distribution, $\text{Ga}(a, b)$, $a, b \to 0$. This test is applied to binned data, and the author recommended to use $K \simeq n^{0.4}$ bins. We also tried different number of bins, but the results are very similar. In the simulation study it is considered $K = 4$, 5 and 6 for $n = 25$, 50 and 100, respectively.

4 Simulation study

In order to investigate the power of the classical and Bayesian tests, a simulation study is carried out. The goal is to test $H_0 : X \sim \text{Exp}(\lambda)$, where $\lambda > 0$ is unknown.

Several different alternative distributions are considered in this work and they are summarized in Table 2. These distributions were chosen such that the most common time to failure distributions are included and a wide variety of distributions with different failure rates and other characteristics is covered. For example, the Gamma and Weibull distributions have increasing failure rate (IFR) for $a > 1$ and decreasing failure rate (DFR) for $0 < a < 1$. For $a = 1$, they both reduce to standard exponential distribution with constant failure rate. The failure rate for the LogNormal distribution initially increases over time and then decreases (non-monotonous). Half-Normal distribution has IFR and $\chi^2(1)$ distribution has DFR. Half-Cauchy is an heavy tailed distribution.

Table 2: Alternative distributions.

Distribution	Notation	Density
Gama	$\text{Ga}(a, 1)$	$\Gamma(a)^{-1}x^{a-1}\exp(-x)$
Weibull	$\text{Wei}(a, 1)$	$ax^{a-1}\exp(-x^a)$
LogNormal	$\text{LN}(0, 1)$	$x^{-1}(2\pi)^{-1/2}\exp(-\log^2(x)/2)$
Half-Normal	$\text{HN}(0, 1)$	$(2/\pi)^{1/2}\exp(-x^2/2)$
Qui-quadrado	$\chi^2(a)$	$1/(2^{a/2}\Gamma(a/2))x^{a/2-1}\exp(-x/2)$
Half-Cauchy	$\text{HCa}(0, 1)$	$(2/\pi)(x^2 + 1)$

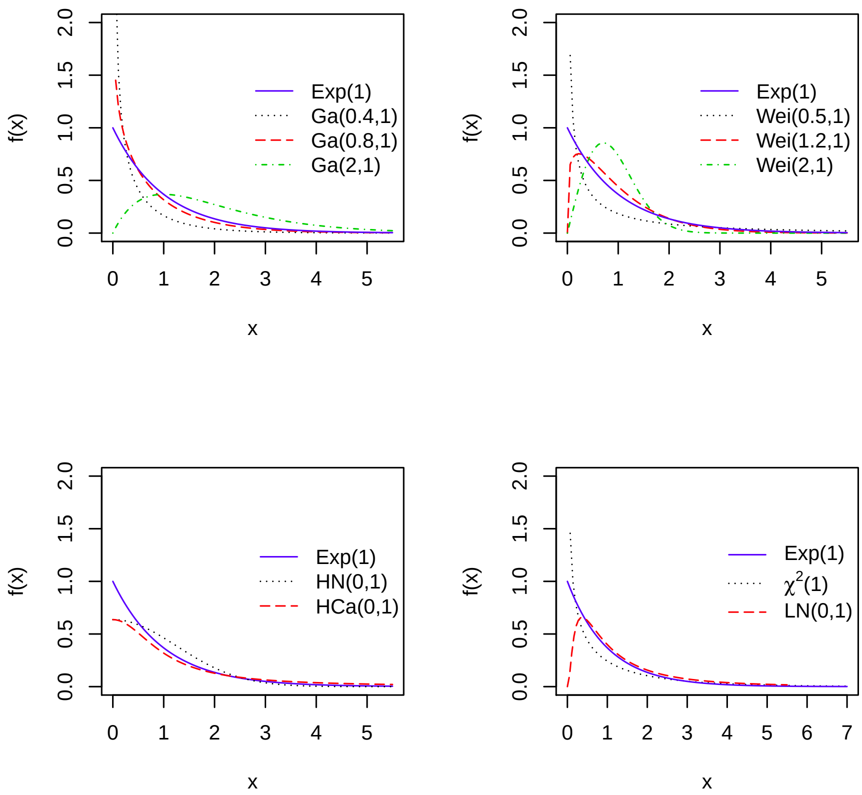

Figure 1: Density plots

The choice of the parameters of the different alternative distributions is done in such a way that the density gradually differs from a standard exponential distribution ($\text{Exp}(\lambda = 1)$). All these distributions are depicted in Figure 1 as well as the standard exponential distribution, in order to allow visualization of the density curve.

Bayes factor is approximated via direct Monte Carlo method as defined in Equation (6), for each one of the h values, based on $L = 20,000$ simulated values from the posterior distribution and by defining a MFPT with $M = 8$ levels. The final estimate for the Bayes factor is obtained by evaluating the minimum value of the thirteen estimates for all values of h. We decided to reject the exponential model whenever the final estimate for the Bayes factor is less than the critical threshold c_{BF}.

To determine the critical threshold, we use the same idea behind the frequentist type I error. We generate 500 samples of sizes n from the standard exponential distribution and compute the Bayes factor $\mathrm{BF}_{01}(\boldsymbol{x})$ for each sample size n. The empirical critical threshold for the Bayes factor and for every n ($n = 25$, $n = 50$ and $n = 100$) is determined by the quantile $\alpha \times 100\%$ from the empirical distribution of $\widehat{\mathrm{BF}}_{01}(\boldsymbol{x})$. In Table 3 we present the empirical critical threshold, obtained by simulation, necessary to achieve a type I error rate of 0.05.

Table 3: Empirical critical threshold for $\widehat{\mathrm{BF}}_{01}(\boldsymbol{x})$.

n	25	50	100
c_{BF}	0.4876	0.5697	0.7117

A comparative study is carried out as follows. A total of 500 random samples with different sizes n equal to 25, 50 and 100 are generated from each of the 10 alternative distributions, namely Ga(0.4,1), Ga(0.8,1), Ga(2,1), Wei(0.5,1), Wei(1.2,1), Wei(2,1), LN(0,1), HN(0,1), $\chi^2(1)$ and HCa(0,1). The Bayes factor, the Bayesian chi-squared test and all the five classical test statistics are calculated from each of the simulated samples. Table 4 shows the average (and standard deviation) of the empirical power based on the proportion of correct rejections.

The main conclusions that can be drawn from the results of the simulation study are:

1. In general, the Bayesian nonparametric test does very well against most alternative distributions;
2. The Bayesian chi-squared test has a very poor performance in terms of empirical power;
3. The power of the Bayesian nonparametric test is higher than the power of the classical tests when the alternative distribution has IFR;
4. When the simulated samples are drawn from alternative distributions with DFR, as it is the case of Ga(0.4,1) and Wei(0.5,1), the Bayesian nonparametric test is at least as powerful as the classical tests; however, it is less powerful than some classical tests against the remaining DFR distributions (Ga(0.8,1) and $\chi^2(1)$) which are closer to a standard exponential distribution;
5. For the Half-Cauchy distribution (heavy tail) the power of the Bayesian nonparametric test is comparable with the classical tests, and for the LogNormal distribution, particularly when samples are smaller, the Bayesian nonparametric test dominates;
6. If the alternative distribution is distinctly different from the exponential one, the power values are above 0.80, and sometimes equal to 1 for large sample size ($n = 100$), except for the Bayesian chi-squared test.

Table 4: Average (and standard deviations) of the empirical proportion of correct H_0 rejections out of 500 samples, for 3 different samples sizes n and for 7 different tests. For the first distribution, Exp(1), this corresponds to empirical type I error rate, and for the alternative distributions it represents the empirical power.

		Test						
Distribution	n	$\widehat{\text{BF}}_{01}(\boldsymbol{x})$	$R_n^B(\tilde{\theta})$	EP_n	CO_n	$\overline{\text{CM}}_n$	BH_n	AD_n
Exp(1)	25	0.050 (0.028)	0.052 (0.028)	0.038 (0.022)	0.034 (0.021)	0.048 (0.030)	0.042 (0.024)	0.046 (0.023)
	50	0.050 (0.030)	0.052 (0.041)	0.034 (0.017)	0.042 (0.031)	0.044 (0.023)	0.044 (0.026)	0.052 (0.027)
	100	0.050 (0.031)	0.052 (0.041)	0.056 (0.032)	0.050 (0.033)	0.050 (0.027)	0.054 (0.031)	0.048 (0.037)
Ga(0.4, 1)	25	0.898 (0.040)	0.404 (0.084)	0.820 (0.042)	0.956 (0.025)	0.824 (0.042)	0.886 (0.041)	0.942 (0.026)
	50	0.990 (0.011)	0.856 (0.037)	0.984 (0.021)	0.998 (0.006)	0.984 (0.021)	0.996 (0.008)	0.998 (0.006)
	100	1	1	1	1	1	1	1
Ga(0.8, 1)	25	0.110 (0.046)	0.042 (0.024)	0.114 (0.049)	0.148 (0.052)	0.110 (0.044)	0.136 (0.047)	0.146 (0.052)
	50	0.214 (0.071)	0.080 (0.048)	0.212 (0.067)	0.266 (0.068)	0.202 (0.068)	0.232 (0.077)	0.248 (0.078)
	100	0.318 (0.083)	0.310 (0.054)	0.348 (0.063)	0.452 (0.095)	0.328 (0.079)	0.380 (0.074)	0.392 (0.093)
Ga(2, 1)	25	0.696 (0.060)	0.310 (0.054)	0.588 (0.088)	0.578 (0.066)	0.592 (0.088)	0.650 (0.077)	0.554 (0.076)
	50	0.989 (0.019)	0.630 (0.064)	0.924 (0.025)	0.984 (0.025)	0.918 (0.030)	0.956 (0.033)	0.928 (0.021)
	100	1	0.932 (0.035)	0.996 (0.008)	0.998 (0.006)	0.996 (0.008)	0.998 (0.006)	0.998 (0.006)
$\chi^2(1)$	25	0.576 (0.068)	0.25 (0.057)	0.626 (0.067)	0.810 (0.043)	0.614 (0.075)	0.704 (0.079)	0.782 (0.049)
	50	0.880 (0.034)	0.592 (0.074)	0.882 (0.033)	0.982 (0.020)	0.872 (0.025)	0.926 (0.028)	0.962 (0.028)
	100	1	0.928 (0.030)	0.992 (0.014)	1	0.988 (0.017)	1	1
Wei(0.5, 1)	25	0.990 (0.024)	0.664 (0.071)	0.938 (0.037)	0.976 (0.025)	0.940 (0.035)	0.966 (0.023)	0.974 (0.019)
	50	1	0.986 (0.019)	0.998 (0.006)	1	1	1	1
	100	1	1	1	1	1	1	1
Wei(1.2, 1)	25	0.208 (0.040)	0.084 (0.049)	0.150 (0.041)	0.124 (0.048)	0.164 (0.041)	0.164 (0.042)	0.144 (0.047)
	50	0.294 (0.038)	0.132 (0.065)	0.270 (0.039)	0.260 (0.057)	0.276 (0.056)	0.284 (0.071)	0.224 (0.053)
	100	0.635 (0.029)	0.242 (0.038)	0.598 (0.033)	0.594 (0.057)	0.574 (0.042)	0.612 (0.044)	0.530 (0.024)
Wei(2, 1)	25	0.990 (0.019)	0.742 (0.045)	0.982 (0.020)	0.972 (0.021)	0.980 (0.019)	0.978 (0.017)	0.970 (0.017)
	50	1	0.994 (0.009)	1	1	1	1	1
	100	1	1	1	1	1	1	1
HCa(0, 1)	25	0.766 (0.030)	0.514 (0.059)	0.764 (0.039)	0.742 (0.033)	0.770 (0.036)	0.758 (0.035)	0.754 (0.038)
	50	0.968 (0.020)	0.752 (0.039)	0.950 (0.025)	0.928 (0.030)	0.956 (0.023)	0.946 (0.021)	0.936 (0.025)
	100	0.998 (0.004)	0.958 (0.017)	0.998 (0.006)	0.998 (0.006)	0.998 (0.006)	0.998 (0.006)	0.998 (0.006)
LN(0, 1)	25	0.270 (0.038)	0.114 (0.031)	0.134 (0.049)	0.090 (0.037)	0.170 (0.040)	0.122 (0.050)	0.170 (0.052)
	50	0.352 (0.042)	0.210 (0.057)	0.168 (0.067)	0.140 (0.057)	0.266 (0.075)	0.206 (0.057)	0.342 (0.078)
	100	0.740 (0.038)	0.446 (0.064)	0.234 (0.048)	0.184 (0.063)	0.446 (0.074)	0.296 (0.062)	0.704 (0.047)
HN(0, 1)	25	0.267 (0.036)	0.118 (0.054)	0.232 (0.042)	0.160 (0.056)	0.246 (0.041)	0.246 (0.037)	0.198 (0.044)
	50	0.586 (0.040)	0.192 (0.065)	0.500 (0.085)	0.372 (0.079)	0.514 (0.071)	0.462 (0.087)	0.408 (0.073)
	100	0.866 (0.038)	0.398 (0.043)	0.848 (0.048)	0.722 (0.068)	0.864 (0.044)	0.810 (0.061)	0.792 (0.057)

7. The classical tests present some instability, in terms of power, as it was already mentioned by Henze and Meintanis (2005). The same happens with the Anderson-Darling test.

8. For the considered alternatives, the number of correct rejections increases, as expected, as the sample size increases, independently of the test.

5 Conclusion

In this work, we suggested two Bayesian goodness-of-fit test for exponentiality. A Monte Carlo simulation study was carried out for power comparisons with some of the most powerful classical tests proposed in literature and for some close alternative distributions.

The simulation results showed a poor performance of the Johnson's Bayesian chi-squared test. The Bayesian nonparametric test showed good discriminatory power for most of the alternative distributions under consideration, particular for those with IFR, and it showed to be at least as powerful as the classical tests for most of the remaining alternative distributions with DFR.

Acknowledgements: Research partially supported by National Funds through FCT, project PEst-OE/MAT/UI0006/2011 and grant SFRH/BD/36869/2007, and by Polytechnic Institute of Porto through PFAD.IPP grant.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR Portugal

References

Anderson, T. W., & Darling, D. A. (1954). A test of goodness-of-fit. *Journal of American Statistical Association*, *49*, 765-769.

Baringhaus, L., & Henze, N. (1991). A class of consistent tests for exponentiality based on the empirical Laplace transform. *Annals of the Institute of Statistical Mathematics*, *43*, 551-564.

Baringhaus, L., & Henze, N. (2000). Tests of fit for exponentiality based on a characterization via the mean residual life function. *Statistical Papers*, *41*, 225-236.

Bayarri, M. J., & Berger, J. O. (2000). P-values for composite null models. *Journal of the American Statistical Association*, *95*, 1127-1142.

Berger, J. O., & Guglielmi, A. (2001). Bayesian Testing of a Parametric Model versus Nonparametric Alternatives. *Journal of the American Statistical Association*, *96*, 174-184.

Choi, B., Kim, K., & Song, S. H. (2004). Goodness of fit test for exponentiality based on Kullback-Leibler information. *Communications in Statistics - Simulation and Computation*, *33(2)*, 525-536.

Cox, D. R., & Oakes, D. (1984). *Analysis of Survival Data.* New York: Chapman and Hall.

D'Agostino, R. B., & Stephens, M. A. (1986). *Goodness-of-Fit Techniques.* New York: Marcel Dekker Inc.

Epps, T. W., & Pulley, L. B. (1986). A test for exponentiality vs. monotone hazard alternatives derived from the empirical characteristic function. *Journal of the Royal Statistical Society, Series B*, *48(2)*, 206-213.

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, *6*, 733-807.

Grané, A., & Fortiana, J. (2011). A directional test of exponentiality based on maximum correlations. *Metrika*, *73*, 255-274.

Hanson, T. (2006). Inference for Mixtures of Finite Polya Tree Models. *Journal of the American Statistical Association*, *101*, 1548-1565.

Hanson, T., & Johnson, W. O. (2002). Modeling regression errors with a mixture of Polya trees. *Journal of the American Statistical Association*, *97*, 1020-1033.

Henze, N., & Meintanis, S. G. (2005). Recent and classical tests for exponentiality: A partial review with comparisons. *Metrika*, *61*, 29-45.

Hjort, N. L., Dahl, F. A., & Steinbakk, G. H. (2006). Post-processing posterior predictive p-values. *Journal of the American Statistical Association*, *101*, 1157-1174.

Hjort, N. L., Holmes, C., Müller, P., & Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambridge University Press.

Johnson, V. E. (2004). A Bayesian chi-squared test for goodness-of-fit. *The Annals of Statistics*, *32:6*, 2361-2384.

Johnson, V. E. (2007). Bayesian model assessment using pivotal quantities. *Bayesian Analysis*, *2*, 719-734.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modeling. *The Annals of Statistics*, *20*, 1222-1235.

Lavine, M. (1994). More aspects of Polya tree distributions for statistical modeling. *The Annals of Statistics*, *22*, 1161-1176.

Robins, J. M., van der Vaart, A., & Ventura, V. (2000). Asymptotic distribution of p-values in composite null models. *Journal of the American Statistical Association*, *95*, 1143-1159.

Schervish, M. J. (1995). *Theory of Statistics*. New York: Springer.

Tokdar, S. T., Chakrabarti, A., & Ghosh, J. K. (2010). Bayesian nonparametric goodness of fit tests. In M. Chen, D. Dey, P. Ḿuller, D. Sun, & K. Ye (Eds.), *Frontiers of statistical decision making and bayesian analysis* (p. 185-194). Frontiers of Statistical Decision Making and Bayesian Analysis, Springer.

Tokdar, S. T., & Martin, R. (2011). *Bayesian test of normality versus a Dirichlet process mixture alternative* (Tech. Rep.). Manuscript Submitted for Publication, 2011.

Verdinelli, I., & Wasserman, L. A. (1998). Bayesian goodness-of-fit testing using infinite-dimensional exponential families. *The Annals of Statistics*, *20*, 1203-1221.