Model Checking with Right Censored Data Using Relative Belief Ratio

Luai Al-Labadi 1,*, Ayman Alzaatreh 2 and Mark Asuncion 1

1 Department of Mathematical & Computational Sciences, University of Toronto Mississauga, Mississauga, ON L5L 1C6, Canada
2 Department of Mathematics and Statistics, American University of Sharjah, Sharjah 61174, United Arab Emirates
* Correspondence: luai.allabadi@utoronto.ca

Abstract: Model checking is a topic of special interest in statistics. When data are censored, the problem becomes more difficult. This paper employs the relative belief ratio and the beta-Stacy process to develop a method for model checking in the presence of right-censored data. The proposed method for the given model of interest compares the concentration of the posterior distribution to the concentration of the prior distribution using a relative belief ratio. We propose a computational algorithm for the method and then illustrate the method through several data analysis examples.

Keywords: beta-Stacy process; model checking; nonparametric Bayesian statistics; relative belief inferences; right-censored data

1. Introduction

An important problem in statistics is to check whether a chosen statistical model is in agreement with observed data. This problem is known as model checking. The tests used to describe how well a statistical model fits a set of observations are goodness-of-fit tests. If it is determined that the observed data do not contradict the model, the true values of the parameters can be inferred. If the model fails to pass its checks, there is cause for concern about the accuracy of the inferences used in analysis. Thus, checking a proposed model on the basis of observed data is a critical component of statistical inference.

The time until an event occurs is the variable of interest in survival analysis. It is common to refer to this time as the survival time. In this case, the observed data are typically right censored. That is, at the right side, the exact survival time can become incomplete, for example, when a subject leaves the study before an event occurs, or when the study ends before the event occurs, right censoring occurs. This type of data are common in medical studies where the variable of interest is the time to death or time to relapse of a disease, in reliability studies where the variable of interest is the time until a machine part fails, and in social sciences where the variable of interest is the lifetime of the elderly in certain social programs [1]. While there are several well-known methods for model checking, the approach we want to take is Bayesian in nature. There is interest in developing nonparametric Bayesian procedures for hypothesis testing. The majority of this work focuses on uncensored data. The proposed model was embedded as a null hypothesis into a larger family of distributions, which is the main approach of model testing. Then priors were placed on the null and the alternative, and a Bayes factor was computed (see, for example, [2–6]). A second important approach for model testing is to apply a prior to the true distribution generating the data, and then measure the distance between the posterior distribution and the hypothesized distribution (see, for example, [7–9]). A third and more recent approach combines the second approach with relative belief ratio ([10]) (for instance, see the work of [11–13]).
When data are censored, goodness-of-fit tests become a challenging problem. This explains why there are few and scattered studies on goodness-of-fit tests for right-censored data. In particular, Bayesian goodness-of-fit tests are rarely covered. Some exceptions include the work of [14–16] considered the two-sample problem by computing the Bayesian factors and beta process priors ([17]). Al-Labadi and Zarepour [15] used the beta-Stacy process and the Kolmogorov–Smirnov distance to test simple hypotheses. Chen and Hanson [16] used Polya tree priors and computed the Bayesian factor. To compute the posteriors, the MCMC procedure was used. Permutation tests were used to compute \( p \)-values. These approaches are computationally intensive.

The main problem addressed in this paper is model checking for lifetime distribution functions. A general description of the approach developed in this paper is provided along with a discussion of the benefits that the approach brings to the problem of model checking. The beta-Stacy process \( \text{BSP}(k(t), F_0(t)), t \geq 0 \) is considered as a prior on on the space of cumulative distribution functions ([18]). Here, \( k(\cdot) > 0 \) is a function defined on \( \mathbb{R}^+, \) and \( F_0 \) is a cumulative distribution function (CDF) on \( \mathbb{R}^+ = [0, \infty) \) and it represents the prior guess for the beta-Stacy process. How closely realizations from the process are clustered around \( F_0 \) is determined by the value of \( k(\cdot) \). The use of the beta-Stacy process is justified because, unlike the Dirichlet process, it is conjugate to both exact and right-censored observations ([19]). The proposed method then compares the posterior and prior distribution concentrations about the model of interest. If the posterior is more concentrated on the model than the prior, this is evidence in favor of the model; if the posterior is less concentrated, this is evidence against the model. This comparison is produced using a relative belief ratio that measures the evidence in the observed data for or against the model and provides a measure of the strength of this evidence; thus, the methodology is based on a direct measure of statistical evidence. The approach is fairly simple to implement and does not necessitate obtaining a closed form of the relative belief ratio. Theoretical results are developed to support appropriate hyperparameter selections \( k(\cdot) \) and \( F_0 \).

The rest of the paper is structured as follows. Section 2 provides the background on the definitions and generic properties of the beta-Stacy process, the relative belief ratio, and the Cramér–von Mises distance between probability measures. Section 3 discusses the proposed methodology and the selection of relevant values for the beta-Stacy process’s hyperparameters. Section 4 develops a computational algorithm for implementing the approach. Section 5 examines the performance of the methodology through a number of examples. This method works well in general and can be used for both censored and uncensored observations. Section 6 wraps up the paper with a summary of the findings. Lastly, datasets and figures are provided in the Appendix A.

2. Notations and Background

Let the random sample \( X_1, \ldots, X_n \) be drawn from an unknown CDF \( F \) defined on \( \mathbb{R}^+ \). Let \( (T, \delta) = (T_1, \delta_1), \ldots, (T_n, \delta_n) \) be observed, where \( T_i = \min(X_i, C_i) \), \( \delta_i = \mathbb{I}(X_i \leq C_i) \) and \( C_1, \ldots, C_n \) are the censoring times. When \( \delta_i = 1 \), \( X_i \) is observed, and when \( \delta_i = 0 \), \( X_i \) is right censored; this type of data can occur, for instance, when \( X_i \) is the lifetime of a patient who enrolls in a study of a certain disease. If death occurs from the disease before the study ends, \( X_i \) is recorded; however, if the patient is still alive, withdraws, or dies for another reason at the end of the study, \( C_i \) records the end of the period during which the patient was under observation ([20]).

In survival analysis, CDF \( F \) is frequently referred to as the lifetime distribution function. We use the same notation throughout this paper for the probability measure and its corresponding cumulative distribution function, i.e., \( F(t) = F([-\infty, t]) \).

2.1. Beta-Stacy Process

Walker and Muliere [18] introduced the beta-Stacy process, which is a nonparametric prior that is widely used in survival analysis. In this section, a relevant introduction about the beta-Stacy process is provided.
The beta-Stacy process is defined by another process known as the log-beta process ([18]). Let \( \beta(\cdot) \) be a positive function defined on \( \mathbb{R}^+ \), and \( \alpha \) be a measure concentrated on \( \mathbb{R}^+ \) that is absolutely continuous with respect to the Lebesgue measure such that \( \int_0^\infty (\beta(z))^{-1} \alpha(dz) = \infty \). On the basis of [18], \( \{Z(t)\}_{t \geq 0} \) is a log-beta process with parameters \( \alpha \) and \( \beta \), written as \( Z(t) \sim \log \text{BSP}(\alpha(t), \beta(t)) \), if \( \{Z(t)\}_{t \geq 0} \) is a Lévy process with Lévy measure defined, for \( s > 0 \), by

\[
L_t(ds) = \left[ \frac{1}{1 - e^{-s}} \int_0^t e^{-s\beta(z)} \alpha(dz) \right] ds
\]

and has the following moment generating function:

\[
\log E e^{-uZ(t)} = \int_0^\infty (e^{-us} - 1)L_t(ds), \quad t \geq 0, u \in \mathbb{R}.
\]

In order to define the beta-Stacy process, let positive function \( k(\cdot) \) be defined on \( \mathbb{R}^+ \) and the absolutely continuous cdf \( F_0 \) be defined on \( \mathbb{R}^+ \). We call \( \{F(t)\}_{t \geq 0} \) a beta-Stacy process with parameters \( k(\cdot) \) and \( F_0 \), written as \( F(t) \sim \text{BSP}(a(t), \beta(t)) \), if \( F(t) = 1 - e^{-Z(t)} \), and \( \{Z(t)\}_{t \geq 0} \) is a log-beta process with parameters

\[
a(dz) = k(z)F_0(dz) \quad \text{and} \quad \beta(z) = k(z)F_0([z, \infty)).
\]

Walker and Muliere [18] showed that

\[
E[F(t)] = F_0(t) = 1 - \exp\left( -\int_0^t (\beta(z))^{-1} \alpha(dz) \right),
\]

rendering \( F_0 \) the prior guess. The expression in (1) explains the need for assumption \( \int_0^\infty (\beta(z))^{-1} \alpha(dz) = \infty \). The beta-Stacy process includes various neutral-to-right processes proposed in the literature. For instance, the beta-Stacy process reduces to the Dirichlet process when \( k(t) = k \) for all \( t \geq 0 \). The beta-Stacy process also reduces to a simple homogenous process ([21]) when \( \beta(\cdot) \) is constant.

Next, we describe the posterior distributions of \( \{Z(t)\}_{t \geq 0} \) and \( \{F(t)\}_{t \geq 0} \). Let random sample \( X_1, \ldots, X_n \) be drawn from \( F \), and \( (T, \delta) = (T_1, \delta_1), \ldots, (T_n, \delta_n) \) be observed, where \( T_i = \min(X_i, C_i), \delta_i = I(X_i \leq C_i) \) and \( C_1, \ldots, C_n \) are censoring times. Define the counting process \( N \) by

\[
N(t) = \sum_{i=1}^n I(T_i \leq t \text{ and } \delta_i = 1)
\]

and the (left-continuous) at-risk process \( Y \) by

\[
Y(t) = \sum_{i=1}^n I(T_i \geq t),
\]

where \( I \) is the indicator function. In particular, let \( N(t) = N(t) - N(t^-) \) (i.e., \( N\{t\} \) is the number of observed \( X_i \)’s at the exact position \( t \)). Let \( \{Z(t)\}_{t \geq 0} \sim \log \text{BSP}(a(t), \beta(t)) \) and \( F(t) = 1 - \exp(-Z(t)) \). Given the data \( (T, \delta) \), the posterior distribution of \( Z \) is a log-beta process ([18]) with parameters

\[
a^*(t) = a(t) + N(t)
\]

and

\[
\beta^*(t) = \beta(t) + Y(t) - N(t).
\]

The posterior Lévy measure for \( Z(t) \) is given by

\[
L^*_t(ds) = \left[ \frac{1}{1 - e^{-s}} \int_0^t e^{-s(\beta(z) + Y(z))} \alpha(z) \right] ds.
\]
There are fixed points of discontinuity in the posterior process. These extra points appear at the exact (uncensored) observations. If $t_i$ is an exact observation with corresponding jump $S_i$, then

$$1 - \exp(-S_i) \sim \text{beta}(N\{t_i\}, \beta(t_i) + Y(t_i) - N\{t_i\}).$$

(4)

If $N\{t_i\} = 1$, then the random jump $S_i$ follows an exponential density with mean $[\beta(t_i) + Y(t_i) - N\{t_i\}]^{-1}$.

Let $F(t) \sim \text{BSP}(k(t), F_0(t))$, $t \geq 0$. Given the data $(T, \delta)$, the posterior distribution of $F$ is a beta-Stacy process with parameters

$$k^*(t) = \frac{\beta^*(t)}{F_0^*[t, \infty]} = \frac{\beta(t) + Y(t) - N\{t\}}{F_0^*[t, \infty]},$$

(5)

and

$$F_0^*(t) = 1 - \prod_{0 \leq j} \left(1 - \frac{d\alpha^*(z)}{\beta^*(z) + \alpha^*(z)}\right) = 1 - \prod_{0 \leq j} \left(1 - \frac{k(z)dF_0(z) + dN(z)}{k(z)F_0(z, \infty) + Y(z)}\right),$$

(6)

with $\alpha^*(s)$ and $\beta^*(s)$ are defined in (2) and (3), and $\prod$ stands for the product integral. Note that, as $k(\cdot)$ tends to zero, the nonparametric Kaplan–Meier estimator of the distribution function is obtained. On the other hand, $F^*$ becomes the prior guess $F_0$ as $k(\cdot)$ grows large. Thus, parameter $k(t)$ can be viewed as the concentration parameter. The posterior consistency of the beta-Stacy process was addressed by Kim and Lee [22].

Algorithms A and B in Appendix A can be used to sample from prior and posterior beta-Stacy processes. These algorithms were developed by Al-Labadi and Zarepour [15] (see also Lee and Kim [23]).

2.2. Relative Belief Ratio

Evans’ (2015) relative belief ratio has become a widely used tool in statistical hypothesis testing theory. Assume that we have a statistical model defined by the pdf $\{f_\theta : \theta \in \Theta\}$ with respect to the Lebesgue measure on the parameter space $\Theta$. Let $\pi(\theta)$ to be a prior on $\Theta$. After observing the data $(T, \delta)$, the posterior distribution of $\theta$ is given by the conditional density function

$$\pi(\theta | (T, \delta)) = \frac{f_\theta((T, \delta))\pi(\theta)}{\int_{\Theta} f_\theta((T, \delta))\pi(\theta) d\theta'}.$$

Assume that we are interested in inferring regarding the parameter $\theta$. Let $\pi$ and $\pi(\cdot | (T, \delta))$ be continuous at $\theta$. Then, the relative belief ratio for a hypothesized value $\theta_0$ of $\theta$ is given by

$$RB_{\Theta}(\theta_0 | (T, \delta)) = \pi(\theta_0 | (T, \delta)) / \pi(\theta_0),$$

(7)

The posterior density to the prior density ratio at $\theta_0$, that is, $RB(\theta_0 | (T, \delta))$, measures how beliefs about $\theta_0$ changed from a priori to a posteriori. When $\pi$ and $\pi(\cdot | (T, \delta))$ are discrete, the relative belief ratio is defined through limits. For more information, see Evans (2015).

Quantity $RB_{\Theta}(\theta_0 | (T, \delta))$ is a measure of evidence that $\theta_0$ is the true value. If $RB_{\Theta}(\theta_0 | (T, \delta)) > 1$, then the probability of $\theta_0$ being the true value increases from a priori to a posteriori; thus, there is evidence based on the data that $\theta_0$ is the true value, and there is, hence, evidence in favor of $\theta_0$. If $RB_{\Theta}(\theta_0 | (T, \delta)) < 1$, then, the probability of $\theta_0$ being the true value decreases from a priori to a posteriori. As a result, the data provide evidence that $\theta_0$ is not the true value. Case $RB_{\Theta}(\theta_0 | (T, \delta)) = 1$ implies there is no evidence in either direction.

The ability to calibrate relative belief ratios is an appealing feature that renders it desirable in hypothesis testing problems. After calculating the relative belief ratio, it is critical to determine whether the result represents strong or weak evidence for or against
 entropy

\[ H_0 : \theta = \theta_0. \] A typical calibration \( RB(\theta_0 \mid (T, \delta)) \) is obtained by computing the tail probability (Evans, 2015)

\[ \text{Str}_{\Theta}(\theta_0 \mid (T, \delta)) = \Pi(\text{RB}_{\Theta}(\theta \mid (T, \delta)) \leq \text{RB}_{\Theta}(\theta_0 \mid (T, \delta))) (T, \delta), \tag{8} \]

where \( \Pi(\cdot \mid (T, \delta)) \) is the posterior distribution of the posterior density \( \pi(\cdot \mid (T, \delta)) \). The posterior probability that the true value of \( \theta \) has a relative belief ratio no greater than that of the hypothesized value \( \theta_0 \) is represented by Equation (8). When \( \text{RB}_{\Theta}(\theta_0 \mid (T, \delta)) < 1 \), there is evidence against \( \theta_0 \), and a small value for \( \text{Str}_{\Theta}(\theta_0 \mid (T, \delta)) \) represents a large posterior probability that the true value has a relative belief ratio greater than \( \text{RB}_{\Theta}(\theta_0 \mid (T, \delta)) \) and hence there is strong evidence against \( \theta_0 \). Similarly when \( \text{RB}_{\Theta}(\theta_0 \mid (T, \delta)) > 1 \), indicates a strong evidence in favour of \( \theta_0 \), while a small value of \( \text{Str}_{\Theta}(\theta_0 \mid (T, \delta)) \) indicates weak evidence in favour of \( \theta_0 \). Figure 1 illustrates the strength of the evidence for both cases; \( \text{RB}_{\Theta}(\theta_0 \mid (T, \delta)) < 1 \) and \( \text{RB}_{\Theta}(\theta_0 \mid (T, \delta)) > 1 \).

![Figure 1. Strength of evidence. The shaded area represents \( \text{Str}_{\Theta}(\theta_0 \mid (T, \delta)) \). RB represents the relative belief ratio \( \text{RB}_{\Theta}(\theta_0 \mid (T, \delta)) \). The smaller the shaded area, the stronger the evidence for case (a). The larger the shaded area, the stronger the evidence for case (b).](image)

2.3. Cramér–Von Mises Distance

Let \( F \) and \( G \) be two cdfs; Cramér–von Mises distance between \( F \) and \( G \) is defined as

\[ d(F, G) = \int_{-\infty}^{\infty} (F(x) - G(x))^2 G(dx). \]

Other distances can be used (see Gibbs and Su (2002)), but \( d \) has some computational advantages. The formula for the distance between the beta-Stacy process and a continuous cdf is provided in the following result.

**Lemma 1.** Let \( F_0 \) be a continuous cdf and \( F_e = 1 - \exp(-Z_e(t)) \), where \( Z_e(t) = \sum_{i=1}^{M} I_{i} \delta_{\theta_i}([0, t]) \), where \( t \geq 0 \) with \( \theta_1, \ldots, \theta_M \overset{i.i.d.}{\sim} F_0 \). Let \( \theta_{(1)} \leq \ldots \leq \theta_{(M)} \) denote the order statistics of \( \theta_1, \ldots, \theta_M \) and \( f_1', \ldots, f_M' \) be the associated jump sizes such that \( f_i = f_i' \) when \( \theta_i = \theta_{(i)} \). Then

\[
d(F_e, F_0) = \frac{1}{3} - \sum_{i=0}^{M} F_e(\theta_{(i)}) \left[ (F_0(\theta_{(i+1)}))^2 - (F_0(\theta_{(i)}))^2 \right] + \sum_{i=0}^{M} (F_e(\theta_{(i)}))^2 \left[ F_0(\theta_{(i+1)}) - F_0(\theta_{(i)}) \right].
\]

where \( F_e(\theta_{(i)}) = 1 - \exp\left(-\sum_{k=1}^{i} f_k'\right) \).

**Proof.** Note that

\[
F_e(x) = \begin{cases} 
0 & \text{if } x < \theta_{(1)} \\
F_e(\theta_{(i)}) & \text{if } \theta_{(i)} \leq x < \theta_{(i+1)} \ (i = 1, \ldots, M - 1), \\
1 & \text{if } x \geq \theta_{(M)}.
\end{cases}
\]
Le \( \theta(0) = 0 \) and \( \theta(M+1) = +\infty \). Then,

\[
d(F_{\epsilon}, F_0) = \int_{\theta(0)}^{\theta(\epsilon+1)} [F_{\epsilon}(x) - F_0(x)]^2 f_0(x) dx
\]

\[
= \sum_{i=0}^{M} \int_{\theta(i)}^{\theta(i+1)} [F_{\epsilon}(\theta(i)) - F_0(x)]^2 f_0(x) dx.
\]

Substituting \( y = F_0(x) \) and \( U_{(i)} = F_0(\theta(i)) \) gives

\[
d(F_{\epsilon}, F_0) = \sum_{i=0}^{M} \frac{U_{(i+1)}^{2}}{U_{(i)}} [F_{\epsilon}(\theta(i)) - y]^2 dy
\]

\[
= \frac{1}{3} \sum_{i=0}^{M} \left\{ [F_{\epsilon}(\theta(i)) - U_{(i)}]^{3} - [F_{\epsilon}(\theta(i)) - U_{(i+1)}]^{3} \right\}.
\]

\[
= \frac{1}{3} \sum_{i=0}^{M} \left\{ U_{(i+1)}^{3} - U_{(i)}^{3} \right\} - \sum_{i=0}^{M} F_{\epsilon}(\theta(i)) \left( U_{(i+1)}^{2} - U_{(i)}^{2} \right)
\]

\[
+ \sum_{i=0}^{M} F_{\epsilon}^{2}(\theta(i)) \left( U_{(i+1)} - U_{(i)} \right)
\]

\[
= \frac{1}{3} \sum_{i=0}^{M} F_{\epsilon}(\theta(i)) \left[ (F_{\epsilon}(\theta(i)) - F_0(\theta(i)))^2 - (F_{\epsilon}(\theta(i)))^2 \right]
\]

\[
+ \sum_{i=0}^{M} (F_{\epsilon}(\theta(i)))^2 \left[ F_0(\theta(i+1)) - F_0(\theta(i)) \right]
\]

\[
+ \sum_{i=0}^{M} (F_{\epsilon}(\theta(i)))^2 \left[ (F_{\epsilon}(\theta(i)))^2 - (F_0(\theta(i)))^2 \right]
\]

\[
= \frac{1}{3} \sum_{i=0}^{M} F_{\epsilon}(\theta(i)) \left[ (F_0(\theta(i)))^2 - (F_0(\theta(i)))^2 \right]
\]

\[
+ \sum_{i=0}^{M} (F_{\epsilon}(\theta(i)))^2 \left[ F_0(\theta(i+1)) - F_0(\theta(i)) \right].
\]

When considering the prior and posterior distributions of the Cramér–von Mises distance, the following lemma allows for the use of the approximation to \( BSP(k(\cdot), F_0) \).

**Lemma 2.** If \( F \sim BSP(k(\cdot), F_0) \) and \( F_{\epsilon} \) is given by Algorithm A, then \( d(F_{\epsilon}, F_0) \overset{a.s.}{\rightarrow} d(F, F_0) \) as \( \epsilon \to 0 \).

**Proof.** By Kim and Lee (2001), \( F_{\epsilon}(x) \overset{a.s.}{\rightarrow} F(x) \), where \( a.s. \) stands for almost surely convergence. As \( (F_{\epsilon}(x) - F_0(x))^2 \leq f_0(x) \), where \( f_0(x) \) is integrable, the proof follows from the dominated convergence theorem. \( \square \)

### 3. Model Checking Using the Relative Belief

Consider the statistical model \( \{ F_{\theta} : \theta \in \Theta \} \) of continuous cdf on \( \mathbb{R}^+ \). Let \( X_1, \ldots, X_n \) be a random sample drawn from an unknown cdf \( F \) defined on \( \mathbb{R}^+ \). Assume that \( \langle T, \delta \rangle = ((T_1, \delta_1), \ldots, (T_n, \delta_n)) \) is observed, where \( T_i = \min(X_i, C_i) \), \( \delta_i = I(X_i \leq C_i) \) and \( C_1, \ldots, C_n \) are censoring times. The aim is to test the hypothesis \( H_0 : F \in \{ F_{\theta} : \theta \in \Theta \} \). Let \( BSP(k(\cdot), F_0) \) be the prior on \( F \) for some choice of \( k(\cdot) \) and \( F_0 \). Then \( F(t) \mid (T, \delta) \sim BSP(k^{*}(t), F_0^{*}(t)) \), where \( k(\cdot) \) and \( F_0 \) are defined in (5) and (6), respectively. If \( H_0 \) is true, then the posterior distribution of the distance between \( F \) and the proposed model should be more concentrated around 0 than the prior distribution. As a result, this test involves comparing the concentrations of the prior and the posterior distributions of \( d \) (see Lemma 1) about 0 using the relative belief ratio with the interpretation as discussed in Section 2.2.

To perform this test, we must measure the distance and then set relevant values for \( k(\cdot) \) and \( F_0 \). To calculate the distance, similar to Al-Labadi and Evans [24], we computed \( d(F, F_{\hat{\theta}}) \), where \( \hat{\theta} \) is the relative belief estimate of \( \theta \), which is always the same as the maximal
likelihood estimate (MLE) for the full model parameter. In terms of hyperparameters, we set \( F_0 = F_{\tilde{\theta}} \) and \( k(t) = k \) for all \( t \). There are numerous advantages in setting \( F_0 = F_{\tilde{\theta}} \). First, it avoids prior-data conflict, a possible contradiction between the data and the prior. This typically happens when the prior places its mass in a region of the parameter space where the data suggest the true value does not lie \(([25,26])\). In the context of the approach considered in this paper, prior-data conflict arises whenever there is a small overlap between the effective support regions of \( F \) and \( F_{\tilde{\theta}} \). Note that, by Lemma 1, the distance \( d(F, F_{\tilde{\theta}}) \) depends on the prior guess \( F_0 \) through the jump points \( \theta_i \). If the \( \theta_i \) lay in one tail of \( F_{\tilde{\theta}} \), then we get prior-data conflict between \( F \) and \( F_{\tilde{\theta}} \) because \( F_0 \) and \( F \) had the same effective support. To avoid this, it is required that \( \theta_i \) are selected in a region that includes most of the mass of \( F_{\tilde{\theta}} \). When \( F_0 = F_{\tilde{\theta}} \), then \( F_{\tilde{\theta}} \) is the prior mean of \( F \); thus, both share the same effective support, which renders it a reasonable choice to avoid prior-data conflict. We refer the reader to Example 1 of Al Labadi and Evans (2018) for an interesting discussion about prior-data conflict. Nevertheless, the choice of \( F_0 = F_{\tilde{\theta}} \) should also avoid any impacts due to the “double use of the data”. This means that the approach becomes conservative in detecting the model failure when \( H_0 \) is false. Although setting \( F_0 = F_{\tilde{\theta}} \) appears to induce a data-dependent prior distribution for \( \delta \), the following lemma implies that this is not the case; thus, the approach is prior distribution-free with this choice.

Lemma 3. If \( F \sim \text{BSP}(k(\cdot), F_{\tilde{\theta}}) \), then the distribution of \( d(F, F_{\tilde{\theta}}) \) does not depend on \( F_{\tilde{\theta}} \).

Proof. Using Lemma 1, since \((\theta_i)_{1 \leq i \leq M}\) is a sequence of i.i.d. random variables with continuous distribution \( F_{\tilde{\theta}} \), for \( i \geq 1 \), we have \( U_i \overset{d}{=} F_{\tilde{\theta}}(\theta_i) \), where \((U_i)_{1 \leq i \leq M}\) is a sequence of i.i.d. random variables follow a uniform distribution on \([0,1]\). Thus,

\[
d(F, F_{\tilde{\theta}}) = \frac{1}{3} - \sum_{i=0}^{M} F_{\tilde{\theta}}(\theta(i)) \left[ (U(i+1))^2 - (\theta(i))^2 \right] + \sum_{i=0}^{M} (F_{\tilde{\theta}}(\theta(i))^2 \left[ U(i+1) - U(i) \right],
\]

where \( U(i) \) is the \( i \)-th order statistic for \((U_i)_{1 \leq i \leq M}\) i.i.d. uniform \([0,1]\). Now, as \( \epsilon \to 0 \), by Lemma 2, we conclude that the distribution of \( d(F, F_{\tilde{\theta}}) \) does not depend on \( F_{\tilde{\theta}} \). \( \square \)

The following results shows that setting \( F_0 = F_{\tilde{\theta}} \) prevents any effect due to the double use of the data. Specifically, as the sample size increases, the posterior distribution of \( d(F, F_{\tilde{\theta}}) \) becomes concentrated around 0 if and only if \( H_0 \) is true. For the proof, see Al-Labadi and Evans (2018).

Lemma 4. Let \((T, \delta) = ((T_1, \delta_1), \ldots, (T_n, \delta_n)) \sim F \), where \( F \sim \text{BSP}(k(\cdot), F_{\tilde{\theta}}) \). Suppose that \( \hat{\theta} \overset{a.s.}{\to} \theta_0 \), \( \sup_{F \in [0,1]} |F_{\tilde{\theta}}(y) - F_{\theta_0}(y)| \overset{a.s.}{\to} 0 \) as \( n \to \infty \).

(i) If \( H_0 \) is true, then, as \( n \to \infty \), \( d(F|T, \delta), F_{\theta_0} \) \( \overset{a.s.}{\to} 0 \).

(ii) If \( H_0 \) is false, then, as \( n \to \infty \), \( \lim \inf d(F|T, \delta), F_{\theta_0} \) \( \overset{a.s.}{\to} > 0 \).

Now, concerning the choice of \( k(\cdot) \), we considered \( k(t) = k \) for all \( t > 0 \). In general, larger values of \( k \) must be chosen to identify smaller deviations. Consequently, it is possible to consider multiple values of \( k \). One way to perform that is, for instance, to start with \( k = 1 \). If a larger (smaller) value of \( k \) renders the relative belief ratios to be below (above) 1, \( H_0 \) is rejected (accepted). As Section 5 shows, when the null hypothesis is correct (not correct), the relative belief ratio always remains above (below) 1 when larger (smaller) values of \( k \) are considered. When using the Dirichlet process, Al-Labadi and Zarepour [27] advised using \( k \leq 0.5n \) for complete data to avoid the prior becoming too influential. Setting \( k \) between 1 and 10 is satisfactory for most purposes. As indicated in the introduction, when \( k(t) = k \), the beta-Stacy process turns into the Dirichlet process. However, in the presence
of right censored data, the posterior distribution of the Dirichlet process becomes beta-Stacy process ([19]). This justifies the necessity of using the beta-Stacy process in the approach.

4. Computational Algorithm

Closed forms of the prior and posterior densities of \( D = d(F, F_{\hat{\theta}}) \) are required to compute the relative belief ratio as in (7). This is not usually available. As a result, the relative belief ratio must be approximated through simulation. A particular problem of computing (7) arises here when both \( \pi_D(0) \) and \( \pi_D|_{(T, \delta)}(0) \) are close to 0, where \( \pi_D(\cdot) \) and \( \pi_D|_{(T, \delta)}(\cdot) \) denote the pdf’s of \( D \) and \( D|_{(T, \delta)} \), respectively. In such a case, determining \( RB_D(0 | (T, \delta)) \) is difficult. The formal definition of the relative belief ratio, as discussed in Section 2.2, is as a limit that can be approximated at zero by

\[
\hat{RB}_D(0 | T, \delta) = \frac{\Pi_D([0, d_c])}{\Pi_{D|_{(T, \delta)}}([0, d_c])},
\]

for a suitably small value \( d_c \), where \( \Pi_D(\cdot) \) and \( \Pi_{D|_{(T, \delta)}}(\cdot) \) denote the cdfs of \( D \) and \( D|_{(T, \delta)} \), respectively. From Equation (8), the strength of the evidence based on the relative belief ratio \( RB_D(0 | T, \delta) \) can be computed using

\[
Str_D(0 | (T, \delta)) = \Pi_{D|_{(T, \delta)}}(RB_D(0 | (T, \delta)) \leq RB_D(0 | (T, \delta))).
\]

Appendix B provides computational Algorithm C for assessing \( H_0 \) on the basis of estimates of (9) and (10). A similar algorithm for complete data based on the Dirichlet process was developed by Al-Labadi and Evans [24].

5. Examples

The approach is demonstrated by two main examples in this section. Throughout this section, let Exp \((\lambda)\), Weibull \((k, \lambda)\), and Lognormal \((\mu, \sigma)\) denote the exponential distribution with mean \(1/\lambda\), the Weibull distribution with shape parameter \(k\) and scale parameter \(\lambda\), and the log-normal distribution with mean \(\mu\) and standard deviation \(\sigma\), respectively. In all examples, the sensitivity to choosing \(k\) is investigated. We set \(\epsilon = 0.01, i_0 = 1, M_0 = 20, \) and \(r_1 = r_2 = 1000\) in Algorithms A, B, and C, though other values are also possible. R package \texttt{parmsurvfit} was used to compute the MLE of the distribution parameters.

Example 1. A real dataset from Lee and Wang (2003) based on the remission times (in months) of cancer patients. The dataset is given in Appendix C. We tested hypothesis

\[ H_0 : F, \text{ given in Table 1, is the underlying distribution of the observed data,} \]

where \(F\) could be a family of distributions (composite hypothesis) or a specific distribution with known parameters (simple hypothesis). Various values of \(k\) were considered to investigate the approach’s sensitivity to concentration parameter selection. The p-value of the (frequentist) log-rank test was also computed for comparison purposes. Table 1 summarizes the findings. When \(H_0\) is true, we want \(RB > 1\) and a strength close to 1, and when \(H_0\) is false, we want \(RB < 1\) and a strength close to 0. According to Table 1, the proposed test performed well in this example.
Table 1. Model checking of Example 1.

| $F$                      | $F_0$                      | $k$       | $RB(Str)$ | $p$-Value |
|--------------------------|----------------------------|-----------|-----------|-----------|
| Exp($\lambda$)           | Exp($\hat{\lambda} = 0.0962$) | 1         | 20(1)     | 0.6521    |
|                          |                            | 5         | 19.84(1)  |           |
|                          |                            | 10        | 15.98(1)  |           |
| Weibull($k, \lambda$)    | Weibull($\hat{k} = 1.0536, \hat{\lambda} = 10.1901$) | 1         | 20(1)     | 0.9986    |
|                          |                            | 5         | 19.76(1)  |           |
|                          |                            | 10        | 17.58(1)  |           |
| Lognormal($\mu, \sigma$) | Lognormal($\hat{\mu} = 1.8198, \hat{\sigma} = 1.0912$) | 1         | 20(1)     | 0.7550    |
|                          |                            | 5         | 18.48(1)  |           |
|                          |                            | 10        | 15.52(1)  |           |
| Exp($\lambda = 2$)       | Exp($\lambda = 2$)         | 1         | 0(0)      | 0         |
|                          |                            | 5         | 0(0)      |           |
|                          |                            | 10        | 0(0)      |           |
| Weibull(2,0.5)           | Weibull(2,0.5)             | 1         | 0(0)      | 0         |
|                          |                            | 5         | 0(0)      |           |
|                          |                            | 10        | 0(0)      |           |
| Lognormal(3,1)           | Lognormal(2,1)             | 1         | 0(0)      | 0         |
|                          |                            | 5         | 0(0)      |           |
|                          |                            | 10        | 0(0)      |           |

Example 2. Simulated data. The primary purpose of this dataset is to investigate how the proposed test performs as the sample size increases. We considered data $(T_1, \delta_1), \ldots, (T_n, \delta_n)$, where $T_i = \min(X_i, C_i)$ where the survival times $(X_i)_{1 \leq i \leq n}$ were generated from Lognormal$(1, 4)$, while the censored time $(C_i)_{1 \leq i \leq n}$ are generated from Lognormal $(4, 1)$.

$H_0 : F$ in Table 2 is the underlying distribution of the observed data.

Table 2 summarizes the results that show that the selected models were accepted. Figures A1–A3 (see Appendix D) give the plots of $F_0 = F_\theta$ and 5 sample paths each for the prior beta-Stacy process and the posterior beta-Stacy process for each case in Table 2. These figures clearly show that the plots of the sample paths for the posterior process moved toward the plot of $F_0$. This supports the previous conclusion regarding the null hypothesis. Furthermore, in this case, the $p$-values of the (frequentist) log-rank test support the conclusion that the null hypothesis should not be rejected.

Table 2. Model checking of Example 2.

| $F$     | $n$ | $F_0$                      | $c$       | $RB(Str)$ | $p$-Value |
|---------|-----|----------------------------|-----------|-----------|-----------|
| Exp($\lambda$) | 20  | Exp($\hat{\lambda} = 0.03$) | 1         | 3.4(0.83) | 0.9995    |
|         |     |                            | 5         | 0.76(0.08)|           |
|         |     |                            | 10        | 0.36(0)   |           |
|         |     |                            | 50        | 1         | 0.24(0.02)| 0.9977    |
|         |     |                            | 5         | 0.02(0)   |           |
|         |     |                            | 10        | 0(0)      |           |
|         |     |                            | 100       | 1         | 0.2(0.02) | 0.9992    |
|         |     |                            | 5         | 0(0)      |           |
|         |     |                            | 10        | 0(0)      |           |
| Weibull($k, \lambda$) | 20  | Weibull($\hat{k} = 0.38, \hat{\lambda} = 20.88$) | 1         | 18.12(1)  | 1         |
|         |     |                            | 5         | 9.78(1)   |           |
|         |     |                            | 10        | 5.22(0.739)|          |
|         |     |                            | 50        | 1         | 19.88(1)  | 0.9992    |
|         |     |                            | 5         |           |           |
|         |     |                            | 10        |           |           |
Table 2. Cont

| $F$  | $n$ | $F_0$                  | $c$     | $RB(\text{Str})$ | $p$-Value |
|------|-----|------------------------|---------|-------------------|-----------|
| 100  | 5   | 15.98(1)               |         |                   | 0.9991    |
|      | 10  | 11.12(1)               |         |                   |           |
|      |     | **Weibull($\hat{k} = 0.38, \hat{\lambda} = 14.10$)** | 1       | 20(1)             |           |
|      | 5   | 19.48(1)               |         |                   |           |
|      | 10  | 16.46(1)               |         |                   |           |
|      |     | **Lognormal($\mu, \sigma$)** | 20      | 11.68(0.416)     | 0.9867    |
|      | 5   | 3.52(1)                |         |                   |           |
|      | 10  | 1.88(1)                |         |                   |           |
|      |     | **Lognormal($\hat{\mu} = 1.14, \hat{\sigma} = 2.40$)** | 50      | 2.0(1)            | 0.9454    |
|      | 5   | 16.4(1)                |         |                   |           |
|      | 10  | 12.4(1)                |         |                   |           |
|      |     | **Lognormal($\hat{\mu} = 1.34, \hat{\sigma} = 3.44$)** | 100     | 20(1)             | 0.9373    |
|      | 5   | 19.36(1)               |         |                   |           |
|      | 10  | 16.12(1)               |         |                   |           |

6. Concluding Remarks

The beta-Stacy process and relative belief ratio were used to propose a general approach for model checking. This method could be used for both complete and right-censored data. It could also be used to test composite or simple hypotheses. Several examples demonstrated that the approach works very well.

Though the Cramér–von Mises distance was used here, other distance measures such as the Kolmogorov–Smirnov and Anderson–Darling distances are viable alternatives. Testing for families of multivariate distributions is an important extension of the approach presented in this paper. While conceptually similar, computational and inferential issues must be addressed. This problem can be addressed in future work.

Author Contributions: Formal analysis, A.A.; Investigation, L.A.-L. and M.A.; Methodology, L.A.-L. and A.A.; Project administration, L.A.-L.; Software, L.A.-L. and M.A.; Writing – review & editing, A.A. and M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are gratefully acknowledge that the work in this paper was partly supported by Faculty Research grant FRG21-S-S03 and the Open Access Program from the American University of Sharjah.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Simulation Algorithms

Algorithm A: Simulation algorithm to approximate the prior of beta-Stacy process.

1. Generate log-beta process with parameters $\alpha(dz) = k(z)F_0(dz)$ and $\beta(z) = k(z)F_0[z, \infty)$.
   The following steps are required to accomplish this:
   (a) Fix a small positive number $\epsilon$.
   (b) Generate the total number of jumps $M \sim \text{Poisson}(\lambda_\epsilon)$, where $\lambda_\epsilon = \alpha([0, \infty))/\epsilon$.
   (c) For $i = 1, \ldots, M$, generate the jump times $\theta_i$ from the probability density function $\alpha(dz)/\alpha([0, \infty))$.
   (d) Generate the jump sizes $J_1, \ldots, J_M$ such that $1 - \exp(-J_i)|\theta_i \sim \text{Beta}(\epsilon, \beta(\theta_i))$. 
1. Generate the posterior log-beta process. The following steps are required to accomplish this:

(a) Generate the process \( \{Z_k^*(t)\}_{t \geq 0} \) on the basis of Algorithm A with \( \beta(\theta_i) \) replaced by \( \beta(\theta_i) + Y(\theta_i) \).

(b) Generate the process \( Z_j^*(t) = \sum_{i=1}^l S_i^* I(T_i \leq t) \), where the random jumps \( (S_i^*)_{1 \leq i \leq l} \) are associated with the fixed points of discontinuity from the distribution given in (4), where \( l \leq n \). These random jumps occur at points where the data are not right-censored.

(c) The approximated posterior log-beta process is given by:

\[
Z_k^*(t) = Z_k(t) \mid (T, \delta) = Z_k^*(t) + Z_j^*(t) = \sum_{i=1}^M I(\theta_i \leq t) + \sum_{i=1}^l S_i^* I(t_i \leq t).
\]

2. The approximate prior beta-Stacy process is \( F_c(t) = F_c(t) \mid (T, \delta) = 1 - \exp(-Z_k^*(t)) \).

Algorithm B: Simulation algorithm to approximate the posterior of beta-Stacy process.

1. Generate the posterior log-beta process. The following steps are required to accomplish this:

(i) Generate a sample from \( F_c \), where \( F_c \) is an approximation of \( F \sim \text{BSP}(k(\cdot), F_\beta) \). See Algorithm A.

(ii) Compute \( \hat{d}(\text{pri}) = d(F_c, F_\beta) \) as described in Lemma 2.

(iii) Repeat Steps (i) and (ii) to obtain a sample of \( r_1 \) values from the prior of \( D \).

(iv) Generate a sample from \( F_c \mid (T, \delta) \), where \( F_c \mid (T, \delta) \) an approximation of \( F \). See Algorithm B.

(v) Compute \( \hat{d}(\text{pos}) = d(F_c \mid (T, \delta), F_\beta) \) as described in Lemma 2.

(vi) Repeat Steps (iv) and (v) to obtain a sample of \( r_2 \) values from the posterior of \( D \).

(vii) For a fixed positive number \( M_0 \), let \( \hat{F}_D \) denote the empirical cdf of \( D \) based on the sample in (iii) and for \( i = 0, \ldots, M_0 \), let \( \hat{d}_{i/M_0}(\text{pri}) \) be the estimate of \( d_{i/M_0}(\text{pri}) \), the \((i/M_0)\)-th quantile of \( D \). Here \( \hat{d}_{0}(\text{pri}) = 0 \), and \( \hat{d}_{1}(\text{pri}) \) is the largest value of \( d(\text{pri}) \).

Let \( \hat{F}_D(\cdot \mid (T, \delta)) \) denote the empirical cdf of \( D \) based on the posterior sample of \( d(\text{pos}) \) in (vi). For \( d \in [\hat{d}_{i/M_0}(\text{pri}), \hat{d}_{(i+1)/M_0}(\text{pri})] \), estimate \( \hat{R}_D(d \mid (T, \delta)) \) by the ratio of the estimates of the posterior and prior contents of \([\hat{d}_{i/M_0}(\text{pri}), \hat{d}_{(i+1)/M_0}(\text{pri})]\). Specifically,

\[
\hat{R}_D(d \mid (T, \delta)) = M_0 \hat{F}_D(\hat{d}_{(i+1)/M_0}(\text{pri}) \mid (T, \delta)) - \hat{F}_D(\hat{d}_{i/M_0}(\text{pri}) \mid (T, \delta)).
\]

Moreover, estimate (9) by \( \hat{R}_D(0 \mid (T, \delta)) = M_0 \hat{F}_D(\hat{d}_{i_0/M_0}(\text{pri}) \mid (T, \delta)) \) where \( i_0 \) is chosen so that \( i_0/M_0 \) is not too small (typically \( i_0/M_0 = 0.05 \)).

(viii) Estimate (10) by the finite sum

\[
\sum_{A} (\hat{F}_D(\hat{d}_{(i+1)/M_0}(\text{pri}) \mid (T, \delta)) - \hat{F}_D(\hat{d}_{i/M_0}(\text{pri}) \mid (T, \delta))),
\]

where \( A = \{ i \geq i_0 : \hat{R}_D(\hat{d}_{i/M_0}(\text{pri}) \mid (T, \delta)) \leq \hat{R}_D(0 \mid (T, \delta)) \} \).

By Al-Labadi and Evans (2018), for fixed \( M_0 \), as \( r_1 \to \infty, r_2 \to \infty \), (A2) converge a.s. to \( R_B(d \mid (T, \delta)) \) and (A3) converge a.s. to \( D(\hat{R}_D(d \mid (T, \delta)) \leq \hat{R}_D(0 \mid (T, \delta)) \mid (T, \delta)) \).
Appendix C. Example 1 Dataset

Table A1. Remission times (months) of 137 cancer patients of the dataset used in Example 1. + denotes a right-censored observation.

|        | 4.5  | 19.13 | 14.24 | 7.87  | 5.49  | 2.02  | 9.22  | 3.82  | 26.31  | 4.65+ |
|--------|------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
|        | 2.62 | 0.90  | 21.73 | 0.87+ | 0.51  | 3.36  | 43.01 | 0.81  | 3.36   | 1.46  |
| 24.80+ | 10.86+| 17.14 | 15.96 | 22.69 | 4.33  | 7.28  | 2.46  | 3.48   | 4.23   |
| 6.54   | 8.65 | 5.41  | 2.23  | 4.34  | 32.15 | 4.87  | 5.71  | 7.59   | 3.02   |
| 4.51   | 1.05 | 9.47  | 79.05 | 2.02  | 4.26  | 11.25 | 10.34 | 10.66  | 12.03  |
| 2.64   | 14.74| 1.19  | 8.66  | 14.83 | 5.62  | 18.10 | 25.74 | 17.36  | 1.35   |
| 9.02   | 6.94 | 7.26  | 4.70+ | 3.70  | 3.64  | 3.57  | 11.64 | 6.25   | 25.82  |
| 3.88   | 3.02+| 19.36+| 20.28 | 46.12 | 5.17  | 0.20  | 36.66 | 10.06  | 4.94   |
| 5.06   | 16.62| 12.07 | 6.97  | 0.08  | 1.40  | 2.75  | 7.32  | 1.26   | 6.76   |
| 8.60+  | 7.62 | 3.52  | 9.74  | 0.40  | 5.41  | 2.54  | 2.69  | 8.26   | 0.50   |
| 5.32   | 5.09 | 2.09  | 7.93  | 12.02 | 13.08 | 5.85  | 7.09  | 5.32   | 4.33+  |
| 2.83   | 8.37 | 14.77 | 8.53  | 11.98 | 1.76  | 4.40  | 34.26 | 2.07   | 17.12  |
| 12.63  | 7.66 | 4.18  | 13.29 | 23.63 | 3.25  | 7.63  | 2.87  | 3.31   | 2.26   |
| 2.69   | 11.79| 5.34  | 6.93  | 10.75 | 13.11 | 7.39  |       |        |        |

Appendix D. Example 2 Figures

Figure A1. Plots of the sample paths for the model, prior and posterior of $BS(x = 0.01, f_0)$ with different choices of $n$ and $k$ in Example 2.
Figure A2. Plots of the sample paths for the model, prior and posterior of $BS(a = 0.01, F_0)$ with different choices of $n$ and $k$ in Example 2.

Figure A3. Cont.
Figure A3. Plots of the sample paths for the model, prior and posterior of $BS(a = 0.01, F_0)$ with different choices of $n$ and $k$ in Example 2.

References
1. Lawless, J.F. Statistical Models and Methods for Lifetime Data, 2nd ed.; Wiley & Sons, Inc.: Hoboken, NJ, USA, 2003.
2. Florens, J.P.; Richard, J.F.; Rolin, J.M. Bayesian Encompassing Specification Tests of a Parametric Model against a Nonparametric Alternative; Technical Report 9608; Universite Catholique de Louvain, Institut de Statistique: Neuchatel, Switzerland, 1996.
3. Carota, C.; Parmigiani, G. On Bayes Factors for Nonparametric Alternatives; In Bayesian Statistics 5; Bernardo, J.M., Berger, J., Dawid, A.P., Smith, A.F.M., Eds.; Oxford University Press: London, UK, 1996.
4. Verdinelli, I.; Wasserman, L. Bayesian goodness-of-fit testing using finite-dimensional exponential families. *Ann. Stat.* 1998, 26, 1215–1241. [CrossRef]
5. Berger, J.O.; Guglielmi, A. Bayesian testing of a parametric model versus nonparametric alternatives. *J. Am. Stat. Assoc.* 2001, 96, 174–184. [CrossRef]
6. McVinish, R.; Rousseau, J.; Mengersen, K. Bayesian goodness of fit testing with mixtures of triangular distributions. *Scand. J. Stat.* 2009, 36, 337–354. [CrossRef]
7. Hsieh, P. A nonparametric assessment of model adequacy based on Kullback-Leibler-divergence. *Stat. Comput.* 2011, 23, 149–162. [CrossRef]
8. Swartz, T.B. Nonparametric goodness-of-fit. *Commun. Stat. Theory Methods* 1999, 28, 2821–2841. [CrossRef]
9. Viele, K. Evaluating Fit Using Dirichlet Processes; Technical Report 384; University of Kentucky, Dept. of Statistics: Lexington, KY, USA, 2000.
10. Evans, M. Measuring Statistical Evidence Using Relative Belief; Monographs on Statistics and Applied Probability 144, CRC Press, Taylor & Francis Group: London, UK, 2015.
11. Abdelrazeq, I.; Al-Labadi, L.; Alzaatreh, A. On one-sample Bayesian tests for the mean. *Statistics* 2020, 54, 424–440. [CrossRef]
12. Al-Labadi, L.; Patel, V.; Vakiloroayaei, K.; Wan, C. Kullback-Leibler Divergence for Bayesian Nonparametric Model Checking. *J. Korean Stat. Soc.* 2021, 50, 272–289. [CrossRef]
13. Evans, M.; Tomal, J. Measuring statistical evidence and multiple testing. *FACET* 2018, 3, 563–583. [CrossRef]
14. Damien, P.; Walker, S. A Bayesian non-parametric comparison of two treatments. *Scand. J. Stat.* 2002, 29, 51–56. [CrossRef]
15. Al-Labadi, L.; Zarepour, M. A Bayesian nonparametric goodness of fit test for right censored data based on approximate samples from the beta-Stacy process. *Can. J. Stat.* 2013, 41, 466–487. [CrossRef]
16. Chen, Y.; Hanson, T. Bayesian nonparametric k-sample tests for censored and uncensored data. *Comput. Stat. Data Anal.* 2014, 71, 335–346. [CrossRef]
17. Hjort, N.L. Nonparametric Bayes estimators based on Beta processes in models for life history data. *Ann. Stat.* 1990, 18, 1259–1294. [CrossRef]
18. Walker, S.; Muliere, P. Beta-stacy processes and a generalisation of the polya-urn scheme. *Ann. Stat.* 1997, 25, 1762–1780. [CrossRef]
19. Phadia, E.G. *Prior Processes and Their Applications*; Springer: Berlin/Heidelberg, Germany, 2013.
20. D’Agostino, R.B.; Stephens, M.A. *Goodness-of-Fit Techniques*; Marcel Dekker Inc.: New York, NY, USA, 1986.
21. Ferguson, T.S.; Phadia, E.G. Bayesian nonparametric estimation based on censored data. *Ann. Stat.* 1979, 7, 163–186. [CrossRef]
22. Kim, Y.; Lee, J. On posterior consistency of survival models. *Ann. Stat.* 2001, 29, 666–686. [CrossRef]
23. Lee, J.; Kim, Y. A new algorithm to generate beta processes. *Comput. Stat. Data Anal.* 2004, 25, 401–405. [CrossRef]
24. Al-Labadi, L.; Evans, M. Prior based model checking. *Can. J. Stat.* 2018, 46, 380–398. [CrossRef]
25. Evans, M.; Moshonov, H. Checking for prior-data conflict. *Bayesian Anal.* 2006, 1, 893–914. [CrossRef]
26. Al-Labadi, L.; Evans, M. Optimal robustness results for relative belief inferences and the relationship to prior-data conflict. *Bayesian Anal.* 2017, 12, 705–728. [CrossRef]
27. Al-Labadi, L.; Zarepour, M. Two-sample Bayesian nonparametric goodness-of-fit test. *Math. Methods Stat.* 2017, 26, 212–225. [CrossRef]