Statistical Analysis for Weibull Distributions in Presence of Right and Left Censoring

Adriano Polpo
Dept. of Statistics / UFSCar
São Carlos, Brazil
polpo@ufscar.br

Marcos Coque-Jr.
Bradesco
Osasco, Brazil
4223.mcoque@bradesco.com.br

Carlos Pereira
IME / USP
São Paulo, Brazil
cpereira@ime.usp.br

Abstract—**Series system reliability is based on the minimum life time of its components. Its dual, the parallel system, is based on maximum. Here, we consider the statistical analysis of both, series and parallel, systems where the components follow the Weibull parametric model. Our perspective is Bayesian. Due to the mathematical complexity, to obtain the posterior distribution we use the Metropolis-Hasting simulation method. Based on this posterior, we evaluated the evidence of the Full Bayesian Significance Test (FBST) for comparing the reliabilities of the components. The reason for using FBST is the fact that we are testing precise hypotheses. We also compute the probability of a particular component be responsible for the system failure. An example illustrates the methodology.**

Keywords-Metropolis-Hasting, evidence, significance test

I. INTRODUCTION

We address the case of k (≥ 2) components for both, series and parallel, systems. In other words, we deal with both, k-out-of-k and 1-out-of-k, reliability systems. Recall that for the series system to work all k components must be working -- k-out-of-k, on the other hand, for the parallel system to work at least 1 component must be working -- 1-out-of-k. In this paper, all the components are considers to follow the Weibull survival model. That is, the lives of all the components have two-parameter Weibull distributions. Our main objective is the estimation of all parameters -- of the Weibull distributions -- involved in the whole system.

Consider a system with k components and let X_j, $j=1,...,k$, and denoting the failure time of the j-th component. The first assumption is that $X_1,...,X_k$ are statically independent Weibull random variables. The series system fails as soon as one component fail and the parallel system fails only when all components fail. The sample observation of a system is random vector, (T,δ), with $T = min(X_1,...,X_k)$ for the series system and $T = max(X_1,...,X_k)$ for parallel system, and $\delta = j$ if $T = X_j$, $j=1,...,k$. Note that δ indicates the component that is responsible for the system failure.

Consider now a sample of n systems (all series or all parallels) that are independent and identically distributed. The set of n sample observations is $(\mathbf{T},\boldsymbol{\delta}) = \{(T_i,\delta_i),\ i=1,...,n\}$. Note that, to obtain the sample we have considered for the i-th observation a vector $(X_{1i},...,X_{ki})$, $i=1,...,n$, of latent or invisible observations of all j components. In fact, we only record the minimum (maximum) life time of all components of this i-th series (parallel) system. In addition, we record which component produce that value of T. For example, suppose that we have 1-out-of-3 system and have the observation $(T=59\ min, \delta=2)$. That is, the second component was the last to fail and its failure time was 59 minutes. Although, the other two components fail before the second, their failure time could not be recorded. Suppose the data above was from 3-out-of-3 system. In this case the second component was the first to fail and the other two will fail after 59 minutes.

The unknown reliability function of the j-th component is $R_j(t) = Pr(X_j > t)$, for $j=1,...,k$. Consequently the system reliability function is given by $R(t) = \prod_{j=1}^{k} R_j(t)$, for the series system and $R(t) = 1 - \prod_{j=1}^{k}[1 - R_j(t)]$, for the parallel system.

The present paper is the parametric Weibull counterpart of the nonparametric papers of Salinas-Torres, Pereira and Tiwari [1], Salinas-Torres, Pereira and Tiwari [2] and Polpo and Pereira [3]. Coque Jr. [4] developed the parametric estimator under Weibull model for the two component series system. On the other hand, the two component Weibull parallel system was introduced by Polpo, Coque Jr. and Pereira [5]. The use of Weibull model was motivated by the work of Irony, Lauretto, Pereira and Stern [6].

In the next section we establish the likelihood function for the Weibull model and the posterior distribution for its parameters. In the sequel we show how to use the Metropolis-Hasting to perform the parameters estimation. A simulated example to show how the estimation process works is then considered. For Hypothesis testing, we use the FBST to compare the parameters of the components. The FBST is based on the evidence given by Pereira and Stern [7]. At the end we discuss the future projects of the authors for general coherent system.

We end this section with the notation used.

- $f(t \mid \theta)$ density function with parameter vector θ at point t.
- $R(t \mid \theta)$ reliability function with parameter vector θ at point t.
- $L(\theta \mid t)$ likelihood function at point θ.
- $II(.)$ unit function: $II(TRUE) = 1$, $II(FALSE) = 0$.
- Δ the responsible component to fail.

The first author have received partial support of FAPESP and the third author have received partial support of CNPq -- a Brazilian agency for research.

978-1-4244-4905-7/09/$25.00©2009 IEEE

$R(.)$ reliability function of the system.
$R_j(.)$ reliability function of the j-th component.
$max(a,b)$ maximum between a and b.
$min(a,b)$ minimum between a and b.
N sample size, number of systems observed.
$\Pr(E)$ probability of event E.
T system failure survival time.
$(\mathbf{T}, \delta)$ $= \{(T_i, \delta_i), i=1,...,n\}$, random sample to be observed.
X_j j-th component failure time.

II. Likelihoods, Priors and Posteriors

Considering now the Weibull model with standard parameters $\theta_j = (\beta_j, \eta_j)$, the reliability function of a random variable X is

$$\Pr(X > x \mid \theta) = R(x \mid \theta) = \exp\left\{-\left(\frac{x}{\eta}\right)^{\beta}\right\} \tag{1}$$

$x > 0$, shape $\beta > 0$ and scale $\eta > 0$. The hazard function, the mean and the variance of this Weibull are respectively:

$h(x \mid \theta) = \frac{\beta}{\eta^{\beta}} x^{\beta-1}$, $\mathrm{E}(X \mid \theta) = \eta\Gamma\left(1+\frac{1}{\beta}\right)$ and $\mathrm{Var}(X \mid \theta) = \eta^2\left\{\Gamma\left(1+\frac{2}{\beta}\right) - \left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2\right\}$

A. Likelihoods

The likelihood function of the system sample is as follows:

series system

$$L(\theta \mid \mathbf{t}, \delta) \propto \prod_{j=1}^{k}\prod_{i=1}^{n}\left[f_j(t_i \mid \theta_j)\right]^{\mathbb{I}(\delta_i=j)}\left[R_j(t_i \mid \theta_j)\right]^{1-\mathbb{I}(\delta_i=j)} \tag{2}$$

parallel system

$$L(\theta \mid \mathbf{t}, \delta) \propto \prod_{j=1}^{k}\prod_{i=1}^{n}\left[f_j(t_i \mid \theta_j)\right]^{\mathbb{I}(\delta_i=j)}\left[1 - R_j(t_i \mid \theta_j)\right]^{1-\mathbb{I}(\delta_i=j)} \tag{3}$$

where f is for densities, $\boldsymbol{\theta} = (\theta_1,..., \theta_k)$, $\theta_j = (\beta_j, \eta_j)$, $j=1,...,k$, and II_A is 1 if A occurs and 0 otherwise. Recall that here t is the minimum (maximum) for series (parallel) system.

B. Priors

Jefrey's noninformative distribution was the prior chosen here. For standard Weibull model, Equation (1), Jeffrey's improper prior is $\frac{1}{\beta\eta}$. This prior was used for both, series and parallel, system; that is

$$\pi(\theta) \propto \prod_{j=1}^{k}\frac{1}{\beta_j\eta_j} \tag{4}$$

C. Posteriors

The posterior distribution is obtained by the normalized product of the prior times the likelihood. We simple multiply the prior (4) by the likelihood functions given in Equations (2) and (3).

series system

$$\pi(\theta \mid \mathbf{t}, \delta) \propto \prod_{j=1}^{k}\frac{1}{\beta_j\eta_j}\prod_{i=1}^{n}\left[\frac{\beta_j}{\eta_j^{\beta_j}}t_i^{1-\beta_j}\exp\left\{-\left(\frac{t_i}{\eta_j}\right)^{\beta_j}\right\}\right]^{\mathbb{I}(\delta_i=j)} \times\left[\exp\left\{-\left(\frac{t_i}{\eta_j}\right)^{\beta_j}\right\}\right]^{1-\mathbb{I}(\delta_i=j)} \tag{5}$$

parallel system

$$\pi(\theta \mid \mathbf{t}, \delta) \propto \prod_{j=1}^{k}\frac{1}{\beta_j\eta_j}\prod_{i=1}^{n}\left[\frac{\beta_j}{\eta_j^{\beta_j}}t_i^{1-\beta_j}\exp\left\{-\left(\frac{t_i}{\eta_j}\right)^{\beta_j}\right\}\right]^{\mathbb{I}(\delta_i=j)} \times\left[1-\exp\left\{-\left(\frac{t_i}{\eta_j}\right)^{\beta_j}\right\}\right]^{1-\mathbb{I}(\delta_i=j)} \tag{6}$$

III. Estimation and FBST

A. Estimation steps

For parameter estimation we use the posterior mean, although there is no closed form for it. The Metropolis-Hasting method should be appropriate for our solutions.

Here we use, as the starting distribution, the gamma distribution for the parameters θ_j, $j=1,...,k$. The posterior distributions obtained from this method is used to obtain the estimates (the means of these distributions) of the parameters.

B. FBST

The Full Bayesian Significance Test (FBST) reviewed by Pereira, Stern and Wechsler [8] is the testing procedure for comparing the reliabilities of the components. The following null hypothesis are the ones of interest:

- H_0: $[(\beta_1 = ... = \beta_k)$ and $(\eta_1 = ... = \eta_k)]$.
- H_0: $[E(X_1) = ... = E(X_k)]$.

The FBST consists of two steps:

1. to evaluate the tail of the posterior distribution up to the manyfold defined by H_0;
2. to decide if the evidence, the volume of the tail, is large or small, exactly as we do with the standard p-values.

Next section presents an example to illustrate the use of the procedure described above.

C. Example: Simulated Observations

Consider a simulated sample of size n = 100 of a three component's system. In fact we have simulated a hundred observations for each one of the following distributions.

component 1: $X_1 \sim$ *Weibull*, with $E(X_1) = 2$ and $sd(X_1) = 2$

component 2: $X_2 \sim$ *gamma*, with $E(X_2) = 2$ and $sd(X_2) = 0.816$

component 3: $X_3 \sim$ *log-normal*, with $E(X_3) = 2.014$ and $sd(X_3) = 2.639$

We also evaluated the probability of each one of the components to be responsible for the system failure, for both systems (series and parallel).

The goal of such example is to evaluate the quality of the Bayesian estimation. Since we have fixed the distributions, we can check how good are the estimates comparing them with the true values of the parameters. We used the same components in both, series and parallel, system.

All chains produced by the Metropolis-Hasting method converged. We use a burn-in of size *10000*, a jump of *10* simulated points of the posterior and the number of generate points from the Metropolis-Hasting method to built the posterior distributions was *10000*.

Table 1 and Table 2 list the estimates and the standard deviation of each parameter. Also, we present the higher posterior (HPD) credible interval of *95%* for these parameters. These intervals were obtained from the marginal posterior distributions.

TABLE I. SERIES: PARAMETER ESTIMATES

		Mean	SD	Credible Interval 95% HPD	
Component 1	β_1	1.1577	0.1258	0.9257	1.4143
	η_1	1.6785	0.1990	1.3223	2.0836
Component 2	β_2	3.2504	0.4531	2.3615	4.1349
	η_2	2.3400	0.1806	1.9743	2.6766
Component 3	β_3	1.4808	0.1766	1.1318	1.8180
	η_3	2.0092	0.2382	1.5665	2.4871

TABLE II. PARALLEL: PARAMETER ESTIMATES

		Mean	SD	Credible Interval 95% HPD	
Component 1	β_1	0.6547	0.0849	0.4870	0.8178
	η_1	1.2287	0.2516	0.7474	1.7176
Component 2	β_2	2.0303	0.1947	1.6596	2.4120
	η_2	2.1938	0.1312	1.9487	2.4597
Component 3	β_3	0.6987	0.0737	0.5627	0.8488
	η_3	1.8249	0.2901	1.2968	2.4189

For the series system, Figure 1, Figure 2 and Figure 3 present the estimates of the reliability functions for components 1, 2 and 3 respectively. Figure 4 illustrates how good is the system estimate, compared with the actual distribution, even with the presence of censored data. Also, for the series system we have that $\rho_1 = Pr(X_1 < min(X_2, X_3)) = 0.5050$, $\rho_2 = Pr(X_2 < min(X_1, X_3)) = 0.1420$ and $\rho_3 = Pr(X_3 < min(X_1, X_2)) = 0.3531$. Figure 5 shows the estimates of the probabilities of component Cj be responsible for the system failure at a given time t.

Now, for the parallel system, Figures 6, 7, 8 and 9 present the corresponding estimates of the components and system distribution functions. The figures here are $\rho_1 = Pr(X_1 > max(X_2, X_3)) = 0.2252$, $\rho_2 = Pr(X_2 > max(X_1, X_3)) = 0.4477$ and $\rho_3 = Pr(X_3 > max(X_1, X_2)) = 0.3271$. Figure 10, for parallel systems, is the correspondent to Figure 5, for series systems. In all figures the dash line represents the true distribution and the gray line represents the estimate distribution.

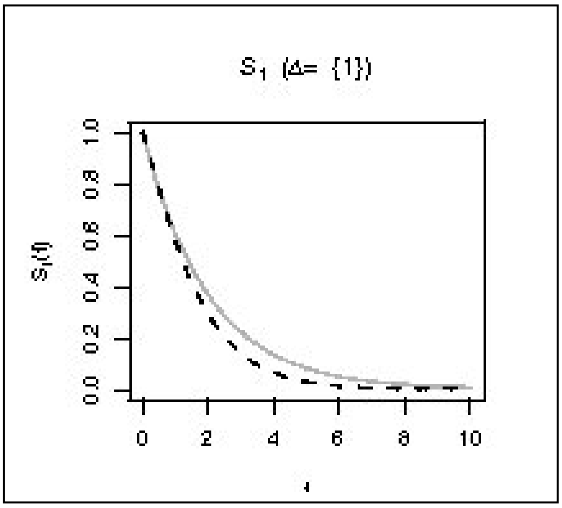

Figure 1. Series: Estimates of component 1 reliability function.

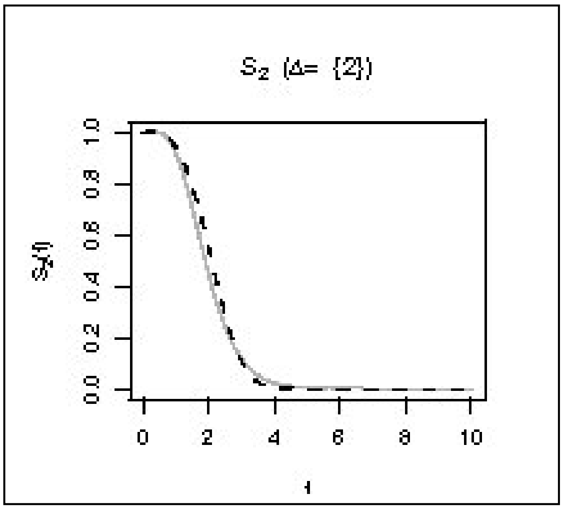

Figure 2. Series: Estimates of component 2 reliability function.

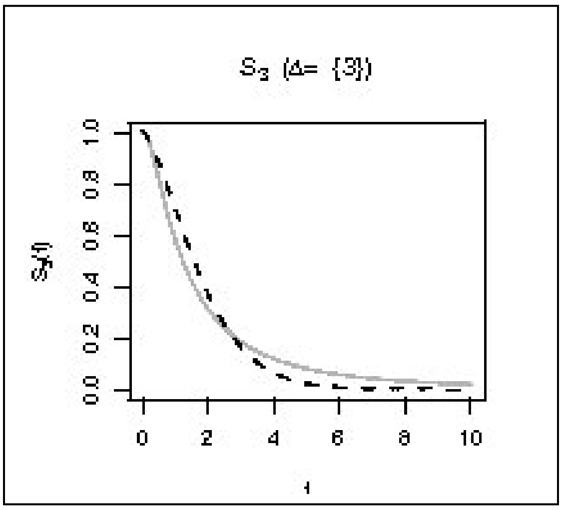

Figure 3. Series: Estimates of component 3 reliability function.

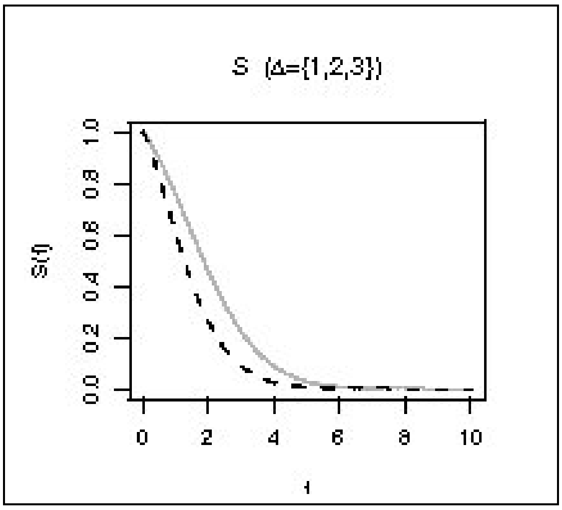

Figure 4. Series: Estimates of system reliability function.

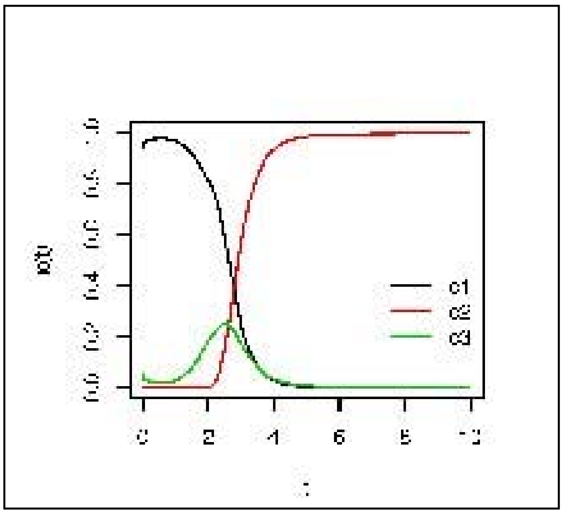

Figure 5. Series: Probability of C_i, $i=1,2,3$, be responsible of the system failure.

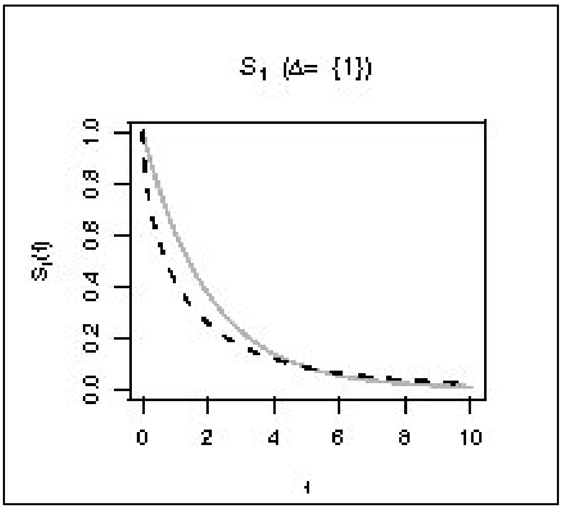

Figure 6. Parallel: Estimates of component 1 reliability function.

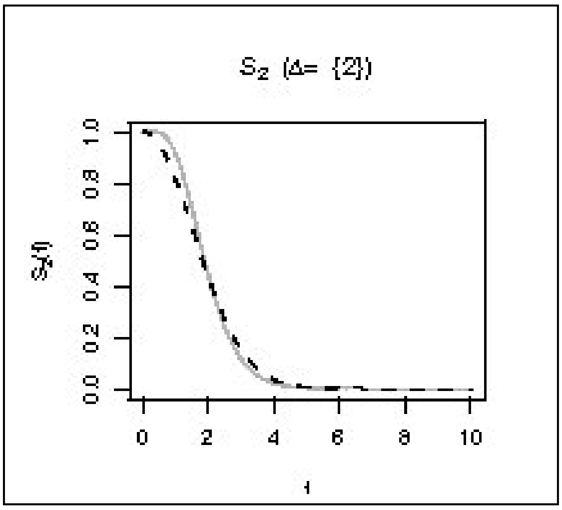

Figure 7. Parallel: Estimates of component 2 reliability function.

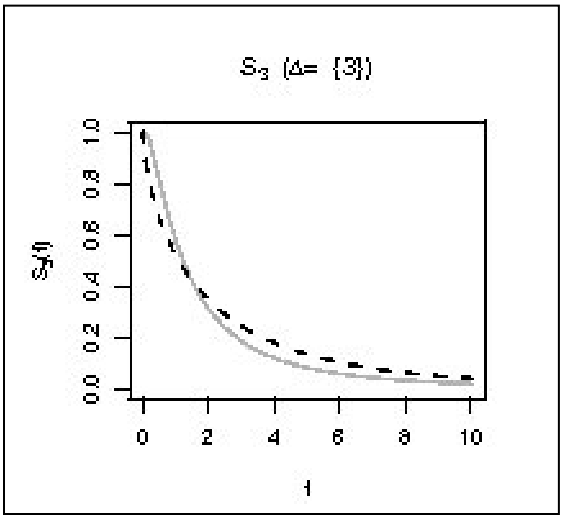

Figure 8. Parallel: Estimates of component 3 reliability function.

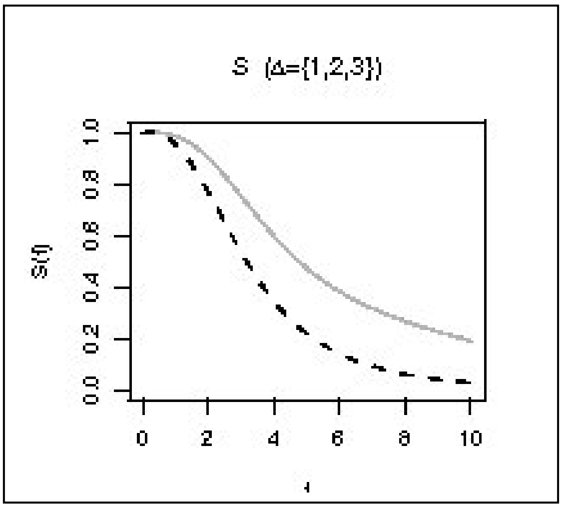

Figure 9. Parallel: Estimates of system reliability function.

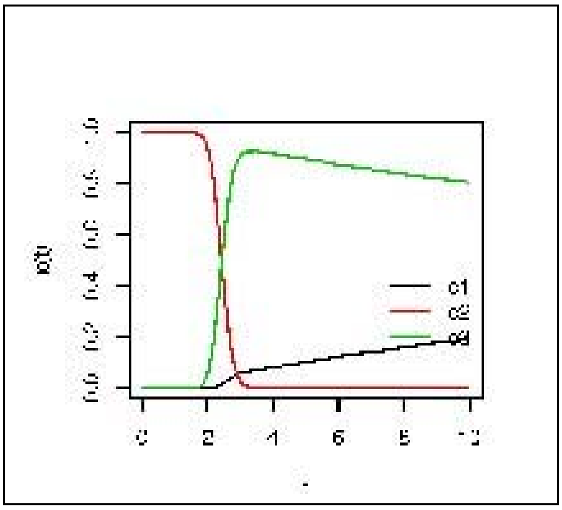

Figure 10. Parallel: Probability of C_i, $i=1,2,3$, be responsible of the system failure.

Figure 11. SERIES FBST

H_0	e-value
$E(X_1) = E(X_2) = E(X_3)$	0.3646
$(\beta_1 = \beta_2 = \beta_3) \wedge (\eta_1 = \eta_2 = \eta_3)$	0

TABLE III. PARALLEL FBST

H_0	e-value
$E(X_1) = E(X_2) = E(X_3)$	0.1898
$(\beta_1 = \beta_2 = \beta_3) \wedge (\eta_1 = \eta_2 = \eta_3)$	0

To evaluate the evidence index from FBST procedure we compute probability of the tangential set to the hypothesis, describe in Section 3.2.

A set $\mathcal{T}$ is said to be tangential to H_0 if all his points have density higher than the density of any parameter point satisfying H_0. The evidence in favoring H_0 is 1-$\mathcal{T}$. Tables 3 and 4 present the evidence values for the hypothesis of equal expected lives of the components. These null hypothesis, as expected, must not be rejected. Recall that in our simulated samples the original generator life distributions have closed expected values. On the other hand, we could not accept the equivalence among the parameters of those distributions. This conclusion also agrees with the generator distributions.

IV. FINAL REMARKS

In this paper we present statistical analysis of series and parallel systems. The example presented shows the force of the method. Data were generated from different known distributions. However, in real life one does not known the distributions that better fits the problem and consequently the observations. The family of Weibull distributions is very rich and accommodates many kinds of situations. There is still a larger family that is the Weibull with three parameters considered in Irony, Lauretto, Pereira and Stern [6]. This parametric family is still more closed to produce solutions as the nonparametric ones. It would not be an absurd to name this last family as a semi-parametric distribution family for survival random variables.

With the nonparmetric solutions together with the Weibull ones for the parallel and series systems, it should be interesting to look for equivalent solutions for some coherent systems like the bridge one, see Barlow [9] and Barlow and Proschan [10].

REFERENCES

[1] Salinas-Torres, V., Pereira, C. & Tiwari, R. “Convergence of dirichlet measures arising in context of bayesian analysis of competing risks models,” Journal of Multivariate Analysis 62, 24–35, 1997.

[2] Salinas-Torres, V., Pereira, C. & Tiwari, R. “Bayesian nonparametric estimation in a competing risks model or a series system,” Journal of Nonparametric Statistics 14, 449–458, 2002.

[3] Polpo, A. & Pereira, C. “Reliability nonparametric bayesian estimation in parallel systems,” IEEE Transactions on Reliability, in press.

[4] Coque Jr., M. “Competing risk under weibull model: A bayesian perspective (in portuguese),” Master’s thesis, USP, 2004.

[5] Polpo, A., Coque-Jr, M. & Pereira, C. “Parallel system using the weibull model,” in M. Lauretto, C. Pereira & J. Stern, eds, ‘Bayesian Inference and Maximum Entropy Methods in Science and Engineering - 28’, American Institute of Physics Press, Melville, pp. 215-231, 2008.

[6] Irony, T., Lauretto, M., Pereira, C. & Stern, J. “System and Bayesian Reliability: Essays in Honor of Professor Richard E. Barlow,” Quality, Reliability and Engineering Statistics, chapter A weibull wearout test: full bayesian approach, 2002.

[7] Pereira, C. & Stern, J. “Evidence and credibility: a full bayesian test of precise hypothesis,” Entropy 1, 104–115, 1999.

[8] Pereira, C., Stern, J. & Wechsler, S. “Can a significance test be genuinely baysian?,” Bayesian Analysis 3(1), 19–100, 2008.

[9] Barlow, R. “Engineering Reliability,” SIAM and ASA, Philadelphia, PA, 1998.

[10] Barlow, R. & Proschan, F. “Mathematical Theory of Reliability,” SIAM, Philadelphia, PA, 1981