

Non-linear failure rate: A comparison of the Bayesian and frequentist approaches to estimation

Tien Thanh Thach^{1,*} and Radim Briš¹

¹Department of Applied Mathematics, VŠB – Technical University of Ostrava, The Czech Republic

Abstract. In this article, a new generalization of linear failure rate called non-linear failure rate is developed, analyzed, and applied to a real dataset. A comparison of Bayesian and frequentist approaches to the estimation of parameters and reliability characteristics of non-linear failure rate is investigated. The maximum likelihood estimators are obtained using the cross-entropy method to optimize the log-likelihood function. The Bayes estimators of parameters and reliability characteristics are obtained via Markov chain Monte Carlo method. A simulation study is performed in order to compare the proposed Bayes estimators with maximum likelihood estimators on the basis of their biases and mean squared errors. We demonstrate that the proposed model fits a well-known dataset better than other mixture models.

1 Introduction

The Weibull failure rate,

$$h(t) = \frac{a}{b} \left(\frac{t}{b}\right)^{a-1} \tag{1}$$

equals 0 at time $t = 0$. This failure rate model might be only appropriate for modeling some physical systems that do not suffer from random shocks. The model, where initial failure rate is negligible might be inappropriate for modeling some physical systems that require initial positive failure rate. That is the physical systems that suffer from random shocks and also from wear out failures. The linear failure rate (LFR),

$$h(t) = a + bt \tag{2}$$

provides partly a remedy for this problem.

The LFR was originally proposed in [1], and has been studied in [2] as a special case of polynomial failure rate model for type II censoring. It is a generalization of an exponential distribution model ($b = 0$) and the Rayleigh distribution model ($a = 0$). The LFR model can also be considered as a mixture of failure rates of an exponential distribution and a Rayleigh distribution. Because the Rayleigh distribution also has a lot of limitations, as well as the LFR, that is not flexible enough to represent the most common types of failure rates, new generalizations of LFR must be developed.

In our study, we bring a new non-linear failure rate (NLFR) model, which is a generalized version of the LFR. We introduce it as a mixture of failure rates of the exponential and Weibull

*e-mail: tien.thach.thanh@vsb.cz

distributions. One of the first attempt to solve the non-linear failure rate model provides [3] but solutions were hard to find due to computational difficulties. In addition, the non-linear failure rate generalized in [3] differs from our new model in other details.

The remainder of this article is organized as follows. Section 2 introduces the intended non-linear failure rates model including basic characteristics of the corresponding lifetime distribution as well. Section 3 brings maximum likelihood estimators for the three unknown parameters of the model on one hand and the adaptive Markov chain Monte Carlo (MCMC) algorithm based on MCMC method on the other hand. In addition, Bayesian methodology resulting from the special likelihood function of the mixture model is demonstrated bringing formulas for alternative estimators in comparison with MLE. Section 4 reports simulation study on specially selected situations where both MCMC and MLE are investigated. Section 5 shows an illustrative example to demonstrate our developed estimation procedures on a real dataset. Finally, Section 6 brings conclusions.

2 The model

2.1 Non-linear failure rate model

Let the failure rate function of the system be in either of the following two states: (1) Initially, it experiences a constant failure rate model. (2) When the system wears out, it also experiences a wear out failure model. That is

$$h(t) = a + bt^{k-1}, \quad a, b, k, t > 0 \tag{3}$$

This model allows an initial positive failure rate, $h(0) = a$, whereas $h(0) = 0$ for most other increasing failure rate function, such as with the Weibull distribution. As mentioned in [2] for the LFR model, this type of situation would exist if failures result at random and also from wear out or deterioration.

Figure 1 shows that non-linear failure rate in equation (3) characterizes three of the most common types of failure rate functions which have been described in [4]:

1. *Increasing failure rate (IFR)*: the instantaneous failure rate increases as a function of time. We expect to see an increasing number of failures for a given period of time.
2. *Decreasing failure rate (DFR)*: the instantaneous failure rate decreases as a function of time. We expect to see the decreasing number of failures for a given period of time.
3. *Constant failure rate (CFR)*: the instantaneous failure rate is constant for the observed lifetime. We expect to see a relatively constant number of failures for a given period of time.

2.2 Characteristics of the lifetime distribution

Using the relationship between $h(t)$ and $R(t)$ we have

$$\begin{aligned} R(t) &= \exp \left\{ - \int_0^t h(s) ds \right\} \\ &= \exp \left\{ - \int_0^t (a + bs^{k-1}) ds \right\} \\ &= \exp \left\{ - \left(at + \frac{b}{k} t^k \right) \right\} \end{aligned} \tag{4}$$

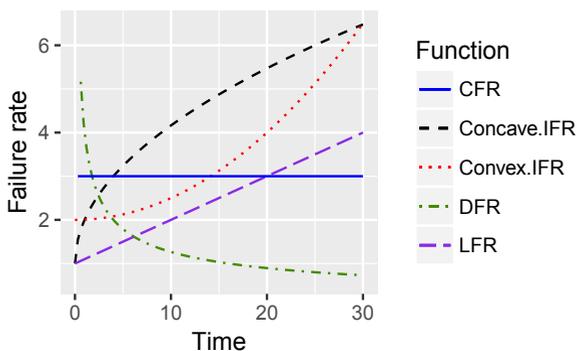


Figure 1. LFR and NLFR models

Then, the probability density function is given as

$$\begin{aligned}
 f(t) &= h(t)R(t) \\
 &= \left(a + bt^{k-1}\right) \exp \left\{ - \left(at + \frac{b}{k} t^k \right) \right\}
 \end{aligned}
 \tag{5}$$

The cumulative failure rate is given by

$$\begin{aligned}
 H(t) &= \int_0^t h(s)ds \\
 &= at + \frac{b}{k} t^k
 \end{aligned}
 \tag{6}$$

And the mean time to failure (MTTF) is given by

$$\begin{aligned}
 MTTF &= \mathbb{E}(T) \\
 &= \int_0^\infty R(t)dt \\
 &= \int_0^\infty \exp \left\{ - \left(at + \frac{b}{k} t^k \right) \right\} dt
 \end{aligned}
 \tag{7}$$

This integral can be obtained by using some suitable numerical methods.

3 Estimation of parameters and reliability characteristics

Let $D : t_1, \dots, t_n$ be the random failure times of n devices under test whose failure time distribution is given as in equation (5) and $\theta = (a, b, k)$. If there is no censoring, the likelihood function takes the general form

$$L(D|\theta) = \prod_{i=1}^n f(t_i|\theta)
 \tag{8}$$

The log-likelihood function can be written as

$$\log L(D|\theta) = \sum_{i=1}^n \log f(t_i|\theta)
 \tag{9}$$

If some observations are right censored, we have to make an adjustment to this expression. For an observation of an observed failure, we put in the pdf as above. But for a right-censored observation, we put in the reliability function, indicating that observation is known to exceed a particular value. The likelihood function in general then takes the form

$$L(D|\theta) = \prod_{i=1}^n f(t_i|\theta)^{\delta_i} R(t_i|\theta)^{1-\delta_i} \tag{10}$$

This expression means that when t_i is an observed failure, the censoring indicator is $\delta_i = 1$, and we enter a pdf factor. When t_i is a censored observation, we have $\delta_i = 0$, and we enter a reliability factor [5].

The log-likelihood function for censoring case is

$$\log L(D|\theta) = \sum_{i=1}^n \delta_i \log f(t_i|\theta) + \sum_{i=1}^n (1 - \delta_i) \log R(t_i|\theta) \tag{11}$$

3.1 The cross-entropy method for continuous multi-extremal optimization

The main idea for the cross-entropy (CE) method for optimization can be stated as follows: Suppose we wish to maximize some “performance” function $S(\mathbf{x})$ over all elements/states \mathbf{x} in some set $\mathcal{X} \subset \mathbb{R}^n$. Let us denote the maximum by γ^* , thus

$$S(\mathbf{x}^*) = \gamma^* = \max_{\mathbf{x} \in \mathcal{X}} S(\mathbf{x}) \tag{12}$$

To proceed with CE, we first randomize our deterministic problem by defining a family of pdfs $\{f(\cdot; \mathbf{v}), \mathbf{v} \in \mathcal{V}\}$ on the set \mathcal{X} . Next, we associate with equation (12) the estimation of

$$\ell(\gamma) = \mathbb{P}_u(S(\mathbf{X}) \geq \gamma) = \mathbb{E}_u I_{\{S(\mathbf{X}) \geq \gamma\}} \tag{13}$$

the so-called associated stochastic problem (ASP). Here, \mathbf{X} is a random vector with pdf $f(\cdot; \mathbf{u})$, for some $\mathbf{u} \in \mathcal{V}$ (for example, \mathbf{X} could be a normal random vector) and γ is a known or unknown parameter. Note that there are in fact two possible estimation problems associated with equation (13). For a given γ we can estimate ℓ , or alternatively, for a given ℓ we can estimate γ , the root of equation (13). Let us consider the problem of estimating ℓ for a certain γ close to γ^* . Then, typically $\{S(\mathbf{X}) \geq \gamma\}$ is a rare event, and estimation of ℓ is a non-trivial problem. The CE method solves this efficiently by making adaptive changes to the probability density function according to the Kullback-Leibler CE, thus creating a sequence $f(\cdot; \mathbf{u}), f(\cdot; \mathbf{v}_1), f(\cdot; \mathbf{v}_2), \dots$ of pdfs that are “steered” in the direction of the theoretically optimal density $f(\cdot; \mathbf{v}^*)$ corresponding to the degenerate density at an optimal point. In fact, the CE method generates a sequence of tuples $\{(\gamma_t, \mathbf{v}_t)\}$, which converges quickly to a small neighborhood of the optimal tuple (γ^*, \mathbf{v}^*) . More specifically, we initialize by setting $\mathbf{v}_0 = \mathbf{u}$, choosing a not very small quantity ϱ , say $\varrho = 10^{-2}$, and then we proceed as follows:

Adaptive updating of γ_t . For a fixed \mathbf{v}_{t-1} , let γ_t be the $(1 - \varrho)$ -quantile of $S(\mathbf{X})$ under \mathbf{v}_{t-1} . That is, γ_t satisfies

$$\mathbb{P}_{\mathbf{v}_{t-1}}(S(\mathbf{X}) \geq \gamma_t) \geq \varrho \tag{14}$$

$$\mathbb{P}_{\mathbf{v}_{t-1}}(S(\mathbf{X}) \leq \gamma_t) \geq 1 - \varrho \tag{15}$$

where $\mathbf{X} \sim f(\cdot; \mathbf{v}_{t-1})$

Algorithm Generic CE Algorithm for Optimization

- 1: Choose some $\hat{\mathbf{v}}_0$. Set $t = 1$.
- 2: Generate a sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ from the density $f(\cdot; \hat{\mathbf{v}}_{t-1})$ and compute the sample $(1 - \varrho)$ -quantile $\hat{\gamma}_t$ of the performances according to equation (16).
- 3: Use the same sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ and solve the stochastic program equation (18). Denote the solution by $\tilde{\mathbf{v}}_t$.
- 4: Apply equation (19) to smooth out the vector $\tilde{\mathbf{v}}_t$.
- 5: Repeat steps 2-4 until a pre-specified stopping criterion is met.

A simple estimator of γ_t , denoted $\hat{\gamma}_t$, can be obtained by drawing a random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ from $f(\cdot; \mathbf{v}_{t-1})$ and evaluating the sample $(1 - \varrho)$ -quantile of the performances as

$$\hat{\gamma}_t = S_{\lceil(1-\varrho)N\rceil} \tag{16}$$

Adaptive updating of \mathbf{v}_t . For fixed γ_t and \mathbf{v}_{t-1} , derive \mathbf{v}_t from the solution of the program

$$\max_{\mathbf{v}} D(\mathbf{v}) = \max_{\mathbf{v}} \mathbb{E}_{\mathbf{v}_{t-1}} I_{\{S(\mathbf{X}) \geq \gamma_t\}} \ln f(\mathbf{X}; \mathbf{v}) \tag{17}$$

The stochastic counterpart of equation (17) is as follows: for fixed $\hat{\gamma}_t$ and $\hat{\mathbf{v}}_{t-1}$ (the estimate of \mathbf{v}_{t-1}), derive $\hat{\mathbf{v}}_t$ from the following program

$$\max_{\mathbf{v}} \hat{D}(\mathbf{v}) = \max_{\mathbf{v}} \frac{1}{N} \sum_{i=1}^N I_{\{S(\mathbf{X}_i) \geq \hat{\gamma}_t\}} \ln f(\mathbf{X}_i; \mathbf{v}) \tag{18}$$

Instead of updating the parameter vector \mathbf{v} directly via the solution of equation (18) we use the following smoothed version

$$\hat{\mathbf{v}}_t = \alpha \tilde{\mathbf{v}}_t + (1 - \alpha) \hat{\mathbf{v}}_{t-1}, \quad i = 1, \dots, n \tag{19}$$

where $\tilde{\mathbf{v}}_t$ is the parameter vector obtained from the solution of equation (18), and α is called the smoothing parameter, with $0.7 < \alpha \leq 1$.

Using normal updating, the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ are sample from an n -dimensional multivariate normal distribution with independent components, $\mathcal{N}(\hat{\boldsymbol{\mu}}_{t-1}, \hat{\boldsymbol{\sigma}}_{t-1}^2)$. While applying CE algorithm, the mean vector $\hat{\boldsymbol{\mu}}_t$ should converge to \mathbf{x}^* and the vector of standard deviations $\hat{\boldsymbol{\sigma}}_t$ converge to the zero vector. In short, we should obtain a degenerated pdf with all mass concentrated in the vicinity of the point \mathbf{x}^* . More detail for this explanation can be found in [6, 7], and a short introduction can also be found in [8].

3.2 Maximum likelihood estimation

In our study, we maximize the log-likelihood function given in equation (9) in case of non-censored data or equation (11) in case of censored data using CE algorithm and obtain the maximizer $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{k})$. By using the invariance property of MLE's,

1. The MLE for $R(t)$, say $\hat{R}(t)$, will be

$$\hat{R}(t) = \exp \left\{ - \left(\hat{a}t + \frac{\hat{b}}{\hat{k}} t^{\hat{k}} \right) \right\} \tag{20}$$

2. The MLE for $h(t)$, say $\hat{h}(t)$, will be

$$\hat{h}(t) = \hat{a} + \hat{b}t^{\hat{k}-1} \tag{21}$$

3. The MLE for $H(t)$, say $\hat{H}(t)$, will be

$$\hat{H}(t) = \hat{a}t + \frac{\hat{b}}{\hat{k}}t^{\hat{k}} \tag{22}$$

4. The MLE for MTTF will be

$$M\hat{T}T\hat{F} = MTTF(\hat{a}, \hat{b}, \hat{k}) \tag{23}$$

which can be obtained by installing into formula (7) and integrating.

3.3 The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm is the most popular MCMC method. The basic idea of the algorithm is to generate data from some generic distribution, say target distribution. Suppose that our target distribution has density π with respect to some reference measure (usually d -dimensional Lebesgue measure). Then, given \mathbf{X}_n , a “proposed value” \mathbf{Y}_{n+1} is generated from some pre-specified density $q(\mathbf{X}_n, \mathbf{Y})$, and then accepted with probability

$$\alpha(\mathbf{x}, \mathbf{y}) = \begin{cases} \min \left\{ \frac{\pi(\mathbf{y}) q(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x}) q(\mathbf{x}, \mathbf{y})}, 1 \right\}, & \pi(\mathbf{x})q(\mathbf{x}, \mathbf{y}) > 0 \\ 1, & \pi(\mathbf{x})q(\mathbf{x}, \mathbf{y}) = 0 \end{cases} \tag{24}$$

If the proposed value is accepted, we set $\mathbf{X}_{n+1} = \mathbf{Y}_{n+1}$; otherwise, we set $\mathbf{X}_{n+1} = \mathbf{X}_n$. The function $\alpha(\mathbf{x}, \mathbf{y})$ is chosen, of course, precisely to ensure that the Markov chain $\mathbf{X}_0, \mathbf{X}_1, \dots$ is reversible with respect to target density $\pi(\mathbf{y})$, so that the target density is stationary for the chain. If the proposed is symmetric, then this reduces to

$$\alpha(\mathbf{x}, \mathbf{y}) = \begin{cases} \min \left\{ \frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}, 1 \right\}, & \pi(\mathbf{x})q(\mathbf{x}, \mathbf{y}) > 0 \\ 1, & \pi(\mathbf{x})q(\mathbf{x}, \mathbf{y}) = 0 \end{cases} \tag{25}$$

This particular case of the Metropolis-Hastings algorithm is called the Metropolis algorithm.

It has long been recognized that the choice of the proposed density $q(\mathbf{x}, \mathbf{y})$ is crucial to the success (e.g. rapid convergence) of the Metropolis-Hastings algorithm. The most common case involves a symmetric random-walk Metropolis algorithm (RWM) in which the proposed value is given by $\mathbf{Y}_{n+1} = \mathbf{X}_n + \mathbf{Z}_{n+1}$, where the increments $\{\mathbf{Z}_n\}$ are i.i.d. from some fixed symmetric distribution (e.g. $N(0, \sigma^2 I_d)$). In this case, the crucial issue become how to scale the proposal (e.g. how to choose σ): too small and the chain will move to slowly; to large and the proposals will usually be rejected. Instead, we must avoid both extremes [9].

3.4 Adaptive MCMC

The search for improved proposal distribution is often done manually, through trial and error, though this can be difficult, especially in high dimensions. An alternative approach is adaptive MCMC, which ask the computer to automatically “learn” better parameter value “on the fly” – that is, while algorithm runs.

Suppose $\{P_\gamma\}_{\gamma \in \mathcal{Y}}$ is a family of Markov chain kernels, each having stationary distribution π . (For example, perhaps P_γ corresponds to RWM algorithm with increment distribution

$N(0, \gamma^2 I_d)$.) An adaptive MCMC would randomly update the value of γ at each iteration, in an attempt to find the best value.

Let Γ_n be the chosen kernel choice at the n th iteration, so

$$\mathbb{P}(\mathbf{X}_{n+1} \in A | \mathbf{X}_n = \mathbf{x}, \Gamma_n = \gamma, \mathbf{X}_{n-1}, \dots, \mathbf{X}_0, \Gamma_{n-1}, \dots, \Gamma_0) = P_\gamma(\mathbf{x}, A) \tag{26}$$

for $n = 0, 1, 2, \dots$. Here the $\{\Gamma_n\}$ are updated according to some adaptive algorithm. In principal, the choice of Γ_n could depend on the entire history $\mathbf{X}_{n-1}, \dots, \mathbf{X}_0, \Gamma_{n-1}, \dots, \Gamma_0$, though in practice it is often the case that the pairs process $\{(\mathbf{X}_n, \Gamma_n)\}_{n=0}^\infty$ is Markovian. In general the algorithms are quite easy to implement, requiring only moderate amounts of extra computer programming [9].

3.5 Bayesian estimation

For our mixture model, the Bayesian model is constructed by specifying a prior distribution for $\theta = (a, b, k)$, and then multiplying with the likelihood function to obtain the posterior distribution.

Denote the prior distribution of θ as $\pi(\theta)$, the posterior distribution of θ given $D : t_1, \dots, t_n$ is given by

$$\pi(\theta|D) = \frac{L(D|\theta)\pi(\theta)}{\int L(D|\theta)\pi(\theta)d\theta} \tag{27}$$

Because the denominator in equation (27) is a normalizing constant, Bayes' theorem is often expressed as:

$$\pi(\theta|D) \propto L(D|\theta)\pi(\theta) \tag{28}$$

As mentioned in [11], the prior distribution is given beforehand, usually based on prior information of the parameters, such as that from historical data, previous experiences, expert suggestions, even wholly subjective suppositions, or simply from the point of mathematical conveniences.

Here, adopted [12], we assume that a, b , and k each have independent $\text{gamma}(\alpha_1, \beta_1)$, $\text{gamma}(\alpha_2, \beta_2)$, and $\text{gamma}(\alpha_3, \beta_3)$ priors respectively, i.e.

$$\pi_1(a) \propto a^{\alpha_1-1} \exp\{-\beta_1 a\}, \alpha_1, \beta_1 > 0 \tag{29}$$

$$\pi_2(b) \propto b^{\alpha_2-1} \exp\{-\beta_2 b\}, \alpha_2, \beta_2 > 0 \tag{30}$$

$$\pi_3(k) \propto k^{\alpha_3-1} \exp\{-\beta_3 k\}, \alpha_3, \beta_3 > 0 \tag{31}$$

If $\alpha_1 = \alpha_2 = \alpha_3 = 1, \beta_1 = \beta_2 = \beta_3 = 0$ we have a generalized uniform distribution on \mathbb{R}^+ [11] or a diffuse prior, and if $\alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$, we have a non-informative prior.

Based on the assumption of independence of individual priors, the joint prior of θ is given by

$$\pi(\theta) = \pi_1(a)\pi_2(b)\pi_3(k) \tag{32}$$

Then, under the square error loss function, the Bayes estimator of a, b, k , failure rate function $h(t)$, reliability function $R(t)$ and cumulative failure rate function are given by

$$a^* = \mathbb{E}(a|D) = \int_\theta a\pi(\theta|D)d\theta \tag{33}$$

$$b^* = \mathbb{E}(b|D) = \int_\theta b\pi(\theta|D)d\theta \tag{34}$$

$$k^* = \mathbb{E}(k|D) = \int_{\theta} k\pi(\theta|D)d\theta \tag{35}$$

$$h^*(t) = \mathbb{E}(h(t; \theta)|D) = \int_{\theta} h(t; \theta)\pi(\theta|D)d\theta \tag{36}$$

$$R^*(t) = \mathbb{E}(R(t; \theta)|D) = \int_{\theta} R(t; \theta)\pi(\theta|D)d\theta \tag{37}$$

$$H^*(t) = \mathbb{E}(H(t; \theta)|D) = \int_{\theta} H(t; \theta)\pi(\theta|D)d\theta \tag{38}$$

In our study, we use adaptive MCMC sampling to generate sample $\theta_i = (a_i, b_i, k_i), i = 1, \dots, n$ from the posterior distribution $\pi(\theta|D)$. Then, Monte Carlo integration estimates $a^*, b^*, k^*, h^*(t), R^*(t)$ and $H^*(t)$ by calculating the means:

$$a^* \approx \frac{1}{n} \sum_{i=1}^n a_i \tag{39}$$

$$b^* \approx \frac{1}{n} \sum_{i=1}^n b_i \tag{40}$$

$$k^* \approx \frac{1}{n} \sum_{i=1}^n k_i \tag{41}$$

$$h^*(t) \approx \frac{1}{n} \sum_{i=1}^n h(t; \theta_i) \tag{42}$$

$$R^*(t) \approx \frac{1}{n} \sum_{i=1}^n R(t; \theta_i) \tag{43}$$

$$H^*(t) \approx \frac{1}{n} \sum_{i=1}^n H(t; \theta_i) \tag{44}$$

4 Monte Carlo simulations

A Monte Carlo simulation study is conducted to compare the performance of CE and MCMC estimators for the parameters of the non-linear failure rate. For each of the following choice of parameters, we simulate 1000 datasets with sample size $n = 25, 50, 100$ and 200 , respectively, and based on each dataset we computed the CE and MCMC estimator for the model parameters. The datasets are generated from the failure distribution Eq.(5) using the accept-reject method. A clear explanation and proof of this method can be found in [13]. In order to obtain MCMC estimators, we have implemented the adaptive MCMC algorithm to construct a Markov chain of length 20,000 with burn-in of 1000 and reduced autocorrelation by retaining only every 5 iterations of the chain and obtain 3801 samples. For prior information we used the generalized uniform distribution on \mathbb{R}^+ , i.e. $\alpha_1 = \alpha_2 = \alpha_3 = 1, \beta_1 = \beta_2 = \beta_3 = 0$. Note here that when $k = 1$, our model coincides with the exponential model with constant failure rate $\lambda = a + b$.

1. $a = 0.03, b = 0.07$, and $k = 0.5$
2. $a = 0.03, b = 0.07$, and $k = 2$
3. $a = 0.03, b = 0.07$, and $k = 3$

Table 1. Comparison of CE and MCMC when $a = 0.03, b = 0.07$, and $k = 0.5$

n	Method	Bias a	MSE a	Bias b	MSE b	Bias k	MSE k
25	CE	-0.0018	0.0004	0.0011	0.0015	0.0546	0.0466
	MCMC	0.0042	0.0002	-0.0021	0.0008	0.0761	0.0260
50	CE	-0.0023	0.0002	0.0014	0.0009	0.0241	0.0225
	MCMC	0.0003	0.0001	-0.0004	0.0005	0.0383	0.0123
100	CE	-0.0025	0.0002	0.0017	0.0006	0.0215	0.0140
	MCMC	-0.0019	0.0001	0.0016	0.0004	0.0302	0.0079
200	CE	-0.0020	0.0001	0.0022	0.0003	0.0170	0.0068
	MCMC	-0.0031	0.0001	0.0037	0.0002	0.0279	0.0054

Table 2. Comparison of CE and MCMC when $a = 0.03, b = 0.07$, and $k = 2$

n	Method	Bias a	MSE a	Bias b	MSE b	Bias k	MSE k
25	CE	-0.0063	0.0013	0.0108	0.0014	-0.0019	0.0781
	MCMC	0.0464	0.0030	-0.0064	0.0005	0.1947	0.1919
50	CE	-0.0084	0.0009	0.0111	0.0009	-0.0253	0.0498
	MCMC	0.0284	0.0013	-0.0079	0.0005	0.1896	0.1260
100	CE	-0.0071	0.0006	0.0084	0.0005	-0.0309	0.0247
	MCMC	0.0178	0.0007	-0.0081	0.0004	0.1502	0.0768
200	CE	-0.0083	0.0004	0.0103	0.0004	-0.0543	0.0187
	MCMC	0.0092	0.0004	-0.0034	0.0003	0.0716	0.0377

Tables 1-3 list the results of the simulation study. Bias is calculated as the mean of 1000 estimates minus the true value, and the mean square error (MSE) is calculated as the mean of the squared differences between 1000 estimators and the true value.

Figure 2-7 visualize the results from the Tables 1-3. From these figures, we see that the biases and MSEs of a and b are very small. So what we are interesting here is the biases and MSEs of the shape parameter k . For all three special cases of the shape parameter k :

- In most of the cases, the CE estimator seem to have less biases than the MCMC estimator (see right panel of Figures 2, 4 and 6).
- In case of decreasing failure rate, i.e. $k < 1$, the MSEs of MCMC estimator seem to be always smaller than MSEs of CE estimator (see right panel of Figure 3).
- In case of increasing failure rate, i.e. $k > 1$, the MSEs of CE estimator seem to be always smaller than MSEs of MCMC estimator (see right panel of Figures 5 and 7).

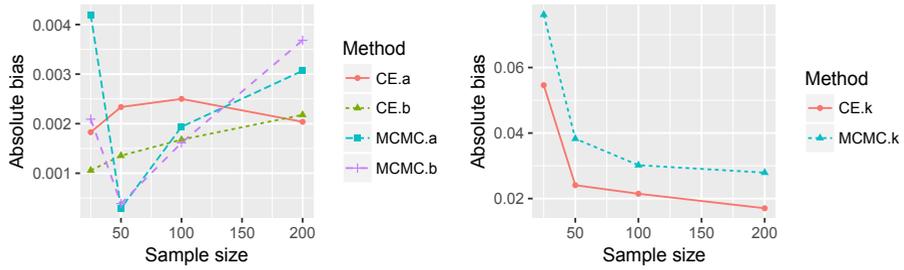


Figure 2. Biases of CE and MCMC estimators for a, b and k when $k = 0.5$

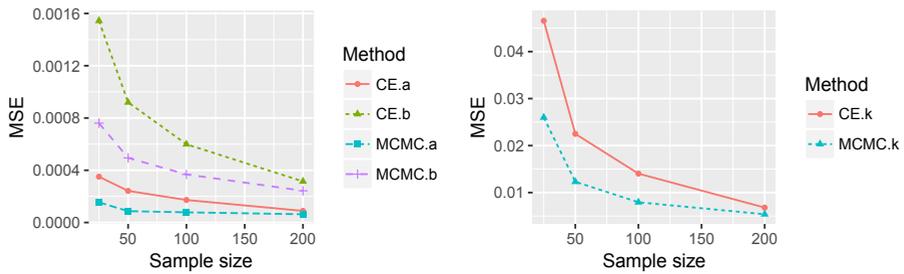


Figure 3. MSEs of CE and MCMC estimators for a, b and k when $k = 0.5$

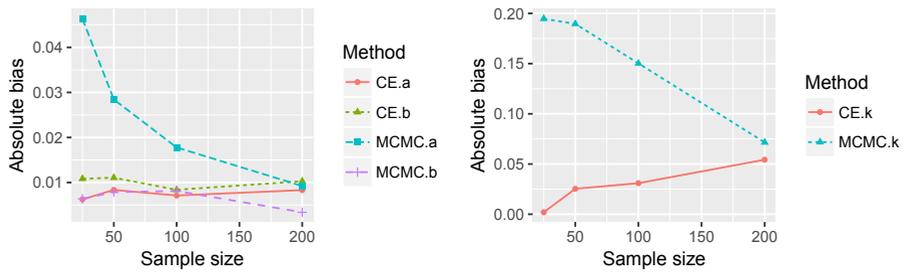


Figure 4. Biases of CE and MCMC estimators for a, b and k when $k = 2$

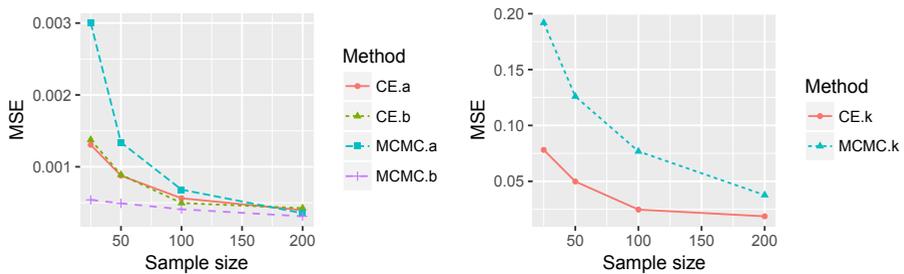


Figure 5. MSEs of CE and MCMC estimators for a, b and k when $k = 2$

Table 3. Comparison of CE and MCMC when $a = 0.03, b = 0.07$, and $k = 3$

n	Method	Bias a	MSE a	Bias b	MSE b	Bias k	MSE k
25	CE	-0.0088	0.0012	0.0211	0.0015	-0.1453	0.1193
	MCMC	0.0389	0.0023	0.0061	0.0011	0.2035	0.3075
50	CE	-0.0098	0.0007	0.0202	0.0010	-0.1602	0.0833
	MCMC	0.0216	0.0010	0.0019	0.0007	0.1591	0.1999
100	CE	-0.0096	0.0005	0.0201	0.0007	-0.1866	0.0694
	MCMC	0.0115	0.0005	0.0010	0.0005	0.0847	0.1072
200	CE	-0.0084	0.0003	0.0183	0.0005	-0.1770	0.0517
	MCMC	0.0064	0.0003	0.0002	0.0003	0.0564	0.0566

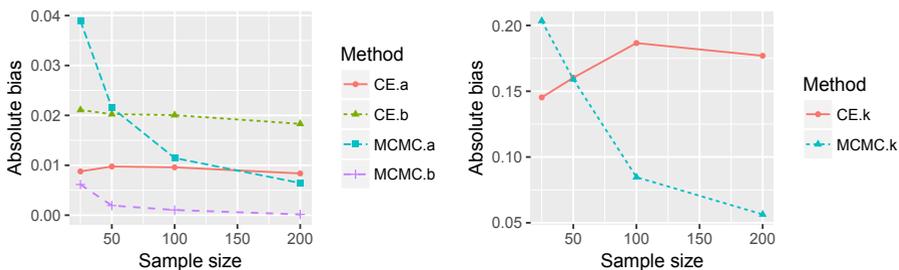


Figure 6. Biases of CE and MCMC estimators for a, b and k when $k = 3$

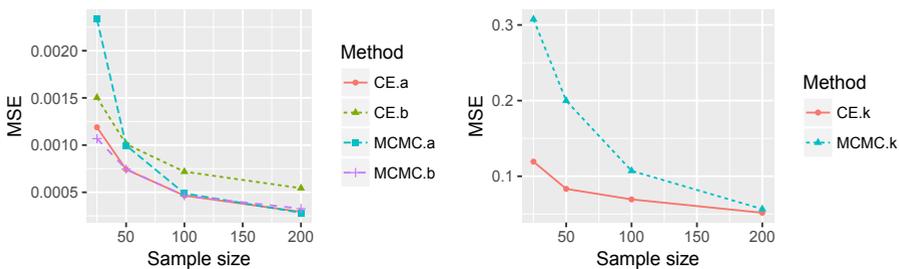


Figure 7. MSEs of CE and MCMC estimators for a, b and k when $k = 3$

5 Illustrative examples

We consider the aircraft windshield failure data given in Table 4. The data consist of 153 observations. Among them 88 are classified as failed windshields, and the remaining 65 are censored time, means that had not failed at the time of observation. The unit for measurement is 1000 hours [14].

In this study, we use both CE and MCMC methods to estimate the parameters and reliability characteristics. In order to obtain MCMC estimators, we have implemented the adaptive

Table 4. Aircraft windshield failure data

	Failure Times			Service Times		
0.040	1.866	2.385	3.443	0.046	1.436	2.592
0.301	1.876	2.481	3.467	0.140	1.492	2.600
0.309	1.899	2.610	3.478	0.150	1.580	2.670
0.557	1.911	2.625	3.578	0.248	1.719	2.717
0.943	1.912	2.632	3.595	0.280	1.794	2.819
1.070	1.914	2.646	3.699	0.313	1.915	2.820
1.124	1.981	2.661	3.779	0.389	1.920	2.878
1.248	2.010	2.688	3.924	0.487	1.963	2.950
1.281	2.038	2.823	4.035	0.622	1.978	3.003
1.281	2.085	2.890	4.121	0.900	2.053	3.102
1.303	2.089	2.902	4.167	0.952	2.065	3.304
1.432	2.097	2.934	4.240	0.996	2.117	3.483
1.480	2.135	2.962	4.255	1.003	2.137	3.500
1.505	2.154	2.964	4.278	1.010	2.141	3.622
1.506	2.190	3.000	4.305	1.085	2.163	3.665
1.568	2.194	3.103	4.376	1.092	2.183	3.695
1.615	2.223	3.114	4.449	1.152	2.240	4.015
1.619	2.224	3.117	4.485	1.183	2.341	4.628
1.652	2.229	3.166	4.570	1.244	2.435	4.806
1.652	2.300	3.344	4.602	1.249	2.464	4.881
1.757	2.324	3.376	4.663	1.262	2.543	5.140
1.795	2.349	3.385	4.694	1.360	2.560	

Table 5. Point estimates and two-sided 90% and 95% HPD intervals for a, b, k and $MTTF$

	MCMC	90% HPD	95% HPD
a	0.0348	[0.0070, 0.0661]	[0.0025, 0.0722]
b	0.0716	[0.0307, 0.1077]	[0.0274, 0.1201]
k	2.9227	[2.4056, 3.4397]	[2.3223, 3.5605]
$MTTF$	3.0351	[2.8177, 3.2302]	[2.7918, 3.2895]

MCMC algorithm to construct a Markov chain of length 50,000 with burn-in of 20,000 and reduced the autocorrelation by retaining only every 5 iterations of the chain and obtain 6001 samples. In fact, it is not necessary to set the long length of burn-in of 20,000 because the adaptive MCMC rapidly converges to the stationary distribution. For prior information we used the generalized uniform distribution on \mathbb{R}^+ , i.e. $\alpha_1 = \alpha_2 = \alpha_3 = 1, \beta_1 = \beta_2 = \beta_3 = 0$. Table 5 shows our MCMC point estimates and two-sided 90% and 95% highest posterior density (HPD) intervals for a, b, k and $MTTF$, and Table 6 shows our CE point estimates and two-sided 90% and 95% bootstrap percentile confident intervals (PERC) for a, b, k and $MTTF$.

Figure 8 shows posterior distributions and trace plots of each parameter of the Bayesian model obtained by MCMC algorithm, and Figures 9-11 show time courses of all relevant functions obtained by both CE and MCMC methods, i.e. reliability, failure rate and cumulative failure rate functions.

Table 6. Point estimates and two-sided 90% and 95% bootstrap percentile confident intervals for a, b, k and MTTF

	CE	90% PERC	95% PERC
a	0.0269	[0.0000, 0.0569]	[0.0000, 0.0626]
b	0.0692	[0.0375, 0.1127]	[0.0321, 0.1244]
k	2.9284	[2.5112, 3.4443]	[2.4331, 3.5554]
$MTTF$	3.0535	[2.8478, 3.2726]	[2.8105, 3.2996]

Table 7. Estimated AIC for nine mixture models for aircraft windshield failure data

Mixture models	Models forms and parameters	AIC
1. Weibull-Weibull	$p \times \text{Weib}(\beta_1, \alpha_1) + (1 - p) \times \text{Weib}(\beta_2, \alpha_2)$	350.159
2. Weibull-Exponential	$p \times \text{Weib}(\beta_1, \alpha_1) + (1 - p) \times \text{Exp}(\lambda)$	348.260
3. Weibull-Normal	$p \times \text{Weib}(\beta_1, \alpha_1) + (1 - p) \times \text{Nor}(\mu, \sigma)$	349.710
4. Weibull-Lognormal	$p \times \text{Weib}(\beta_1, \alpha_1) + (1 - p) \times \text{Lnor}(\mu, \sigma)$	351.235
5. Normal-Exponential	$p \times \text{Nor}(\mu, \sigma) + (1 - p) \times \text{Exp}(\lambda)$	351.513
6. Normal-Lognormal	$p \times \text{Nor}(\mu, \sigma) + (1 - p) \times \text{Lnor}(\mu_2, \sigma_2)$	351.579
7. Lognormal-Exponential	$p \times \text{Lnor}(\mu, \sigma) + (1 - p) \times \text{Exp}(\lambda)$	349.301
8. Linear failure rate	$h(t) = a + bt$	359.106
9. Non-linear failure rate	$h(t) = a + bt^k$	347.371

Table 8. MLEs of parameters for nine mixture models

Mixture models	MLEs of parameters
1. Weibull-Weibull	$\hat{\beta}_1 = 1.249, \hat{\alpha}_1 = 0.245, \hat{\beta}_2 = 2.777, \hat{\alpha}_2 = 3.485, \hat{p} = 0.017$
2. Weibull-Exponential	$\hat{\beta} = 2.768, \hat{\alpha} = 3.484, \hat{\lambda} = 4.052, \hat{p} = 0.983$
3. Weibull-Normal	$\hat{\beta} = 7.359, \hat{\alpha} = 4.481, \hat{\mu} = 2.303, \hat{\sigma} = 0.868, \hat{p} = 0.387$
4. Weibull-Lognormal	$\hat{\beta} = 1.246, \hat{\alpha} = 0.395, \hat{\mu} = 1.075, \hat{\sigma} = 0.440, \hat{p} = 0.029$
5. Normal-Exponential	$\hat{\mu} = 3.053, \hat{\sigma} = 1.220, \hat{\lambda} = 24.003, \hat{p} = 0.995$
6. Normal-Lognormal	$\hat{\mu}_1 = 0.302, \hat{\sigma}_1 = 0.182, \hat{\mu}_2 = 1.073, \hat{\sigma}_2 = 0.443, \hat{p} = 0.027$
7. Lognormal-Exponential	$\hat{\mu} = 1.080, \hat{\sigma} = 0.435, \hat{\lambda} = 1.729, \hat{p} = 0.965$
8. Linear failure rate	$\hat{a} = 0.004, \hat{b} = 0.161$
9. Non-linear failure rate	$\hat{a} = 0.027, \hat{b} = 0.069, \hat{k} = 2.928$

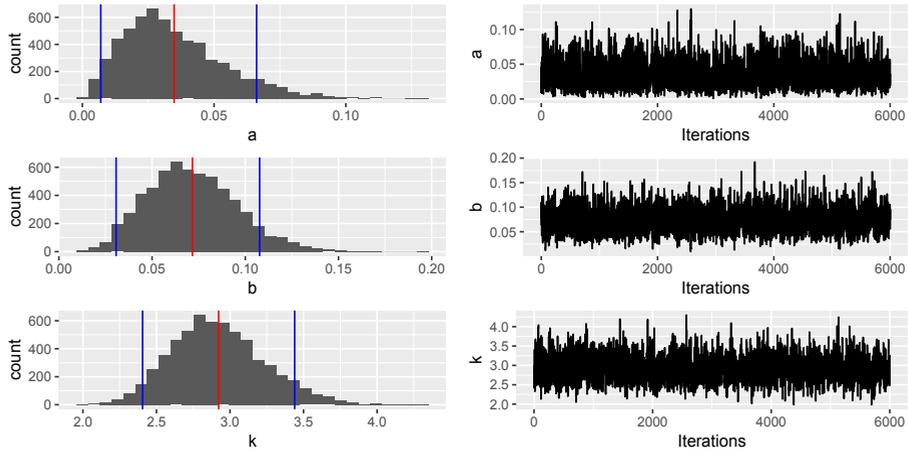


Figure 8. Posterior densities and trace plots of each parameter of the Bayesian model

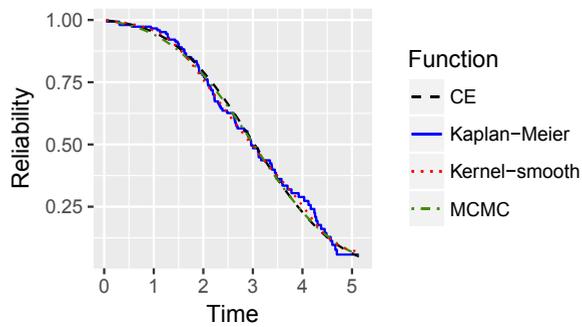


Figure 9. The time courses of the reliability functions.

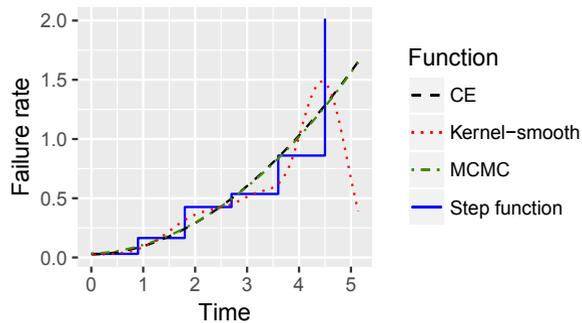


Figure 10. The time courses of the failure rate functions.

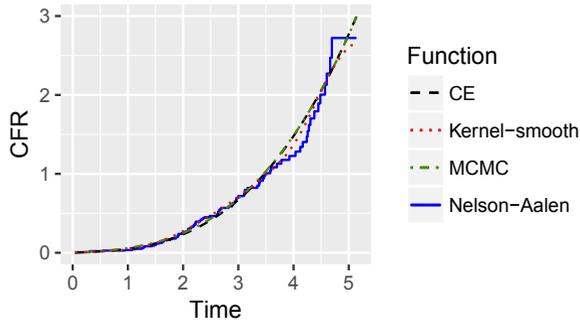


Figure 11. The time courses of the cumulative failure rate functions.

Tables 7-8 represent the Akaike Information Criterion (AIC) values and MLEs of parameters for mixture models, respectively. The AIC values and MLEs from model 1 to 7 have been obtained in [14]. In comparison with these results, Table 7 shows that non-linear failure rate has smallest AIC value. A model for which the AIC is the minimum is considered to be the best approximating model among a set of alternative models.

For this right censored dataset, both CE and MCMC methods give almost the same results. The MLEs are easily obtained by using CE method. This fact is due to the stochastic search of CE method. To improve the result of CE algorithm, one can choose the initial value(s) of (vector) standard deviation $\hat{\sigma}_0$ (sometimes changing the initial value(s) of (vector) mean $\hat{\mu}_0$ is needed) through trial and error until we get the smallest optimum value which yields the optimal estimator(s). Nevertheless, Bayes credible intervals are easily obtained by using MCMC method.

6 Conclusions

The benchmark dataset, the aircraft windshield failure data, shows that the innovative non-linear failure rate might be more appropriate than some mixtures of distributions for modeling the data which might be originated from more than one failure mode.

We recommend using the CE and MCMC methods as the mutual add-in tools for parameter estimation. MLEs are easily obtained by using CE method. The CE method is not so sensitive to the choice of the initial values of the algorithm in comparison with other alternative optimization methods. However, Bayes credible intervals are easily achieved by using MCMC method.

In our study, we consider only the sum of two components in the non-linear failure rate which results from the assumption of two independent competing risks, i.e. the random shock and the wear out. The case of dependent competing risks will be considered next. And we only use the diffuse priors on the parameters for Bayesian estimation, so this study can also be continued with other alternative priors.

Acknowledgement

This work was supported by the European Regional Development Fund in the Research Centre of Advanced Mechatronic Systems project, project number CZ.02.1.01/0.0/0.0/16_019/0000867 within the Operational Programme Research, Development and Education, and partially by the VŠB – Technical University of Ostrava in the SGS project number SP2018/68.

References

- [1] Kodlin D., A new response time distribution (*Biometrics*) **2**, 227-239 (1967).
- [2] Bain L. J., Analysis for the linear failure-rate life-testing distribution (*Technometrics*) **16**, 551-559 (1974).
- [3] Salem S. A., Bayesian estimation of a non-linear failure rate from censored samples type II (*Microelectronics and Reliability*) **32**, 1385-1388 (1992).
- [4] Hamada M. S., Wilson A. G., Martz H. F., Reese C. S., *Bayesian reliability* (Springer, 2008).
- [5] Moore D. F., *Applied survival analysis using R* (Springer, 2016).
- [6] Kroese D. P., Porotsky S., Rubinstein R. Y., The cross-entropy method for continuous multi-extremal optimization (*Methodology and Computing in Applied Probability*) **8**, 383-407 (2006).
- [7] Rubinstein R. Y., Kroese D. P., *The cross-entropy method: a unified approach to combinatorial optimization, Monte Carlo simulation and machine learning* (Springer-Verlag, New York, 2004).
- [8] Zio E., *The Monte Carlo simulation method for system reliability and risk analysis* (Springer, 2013).
- [9] Brooks S., Gelman A., Jones G., Meng X. L., *Handbook of Markov Chain Monte Carlo* (CRC Press, 2011).
- [10] Canfield R., A Bayesian approach to reliability estimation using a loss function (*IEEE Transactions on Reliability*) **19**, 13-16 (1970).
- [11] Jiang H., Xie M., Tang L. C., Markov chain Monte Carlo methods for parameter estimation of the modified weibull distribution (*Journal of Applied Statistics*) **35**(6), 647-658 (2008).
- [12] Kundu D., Howlader H., Bayesian inference and prediction of the inverse Weibull distribution for type-II censored data. (*Computational Statistics and Data Analysis*) **54**(6), 1547-1558 (2010).
- [13] Robert C. P., Casella G., *Introducing Monte Carlo methods with R* (Springer, 2010).
- [14] Blischke W. R., Karim M. R., Murthy D. P., *Warranty data collection and analysis* (Springer, 2011).