Non-linear failure rate: A Bayes study using Hamiltonian Monte Carlo simulation

Tien T. Thach[a], Radim Bris[a,*], Petr Volf[b], Frank P. A. Coolen[c]

[a] *Department of Applied Mathematics, VSB - Technical University of Ostrava, The Czech Republic*
[b] *Institute of Information Theory and Automation, Prague, The Czech Republic*
[c] *Department of Mathematical Science, Durham University, Durham, UK*

Abstract

A generalization of the linear failure rate called non-linear failure rate is introduced, analyzed, and applied to real data sets for both censored and uncensored data. The Hamiltonian Monte Carlo and cross-entropy methods have been exploited to empower the traditional methods of statistical estimation. We have obtained the Bayes estimators of parameters and reliability characteristics using Hamiltonian Monte Carlo and these estimators are considered under both symmetric and asymmetric loss functions. Additionally, the maximum likelihood estimators of parameters are obtained by using the cross-entropy method to optimize the log-likelihood function. The superiority of the proposed model and estimation procedures are demonstrated on real data sets adopted from references.

Keywords: Non-linear failure rate, Bayesian estimators, loss functions, Hamiltonian Monte Carlo, model selection criteria, cross-entropy method, maximum likelihood estimators.

1. Introduction

The exponential distribution is often used in reliability studies as the model for the time until failure of a device. The lack of memory property of the exponential distribution implies that the device does not wear out. That is, regardless of how long the device has been operating, the probability of a failure in the next 1000 hours is the same as the probability of a failure in the first 1000 hours of operation. The lifetime of a device with failures caused by random shocks might be appropriately modeled as an exponential distribution. However, the lifetime T of a device that suffers slow mechanical wear, such as bearing wear, is better modeled by a distribution such that the probability $\mathbb{P}(T < t + \Delta t | T > t)$ increases with t. Distributions such as the Weibull distribution are often used in practice to model the failure time of this type of device [1].

However, the failure rate of Weibull distribution, $h(t) = bt^{k-1}$, equals 0 at time $t = 0$. This model might only be suitable for modeling some physical systems that suffer only from random shock or wear out. For physical systems that suffer from both random shock and wear out failures, the Weibull model might be inappropriate. In this regard, the linear failure rate (LFR), $h(t) = a + bt$, provides partly a solution. The LFR was first introduced in [2] and had been studied in [3] as a special case of polynomial failure rate model for type II censoring. It is a combination of exponential and Rayleigh distributions. It can also be considered as a generalization of the exponential model ($b = 0$) or the Rayleigh model ($a = 0$). The Bayesian estimation technique of this LFR model has been given in [4]. Because of the limitation of the Rayleigh failure rate, as well as the LFR, which is only linearly increasing, new generalizations of LFR must be developed.

In this paper, we introduce a generalized version of the LFR called non-linear failure rate (NLFR). It is considered as a combination of the exponential and Weibull distributions. This NLFR model not only allows

*Corresponding author: Applied Mathematics, Technical University of Ostrava, 17 Listopadu 15, Ostrava-Poruba 708 33, The Czech Republic.
Email address: radim.bris@vsb.cz (Radim Bris)

for an initial positive failure rate but also takes into account all shapes of the Weibull failure rate. The first research work which attempts to solve the NLFR is given in [5] but solutions were hard to obtain due to computational difficulties. Besides, our generalization differs from [5] in other details. This model was also introduced in [6], but with a different name, motivation, parametrized form, model explanation, and purpose. Compare to our previous conference proceedings [7], we have provided here a different parameterization of the NLFR model; properties of the model have been studied; symmetric and asymmetric loss functions have been considered; recent advance Markov chain Monte Carlo (MCMC) method, called Hamiltonian Monte Carlo (HMC) has been employed; model selection criteria have been applied; two more data sets have been added and analyzed.

In addition to the model, we exploit the modern computational methods known as the cross-entropy (CE) and HMC methods. Mixture models often result in too many parameters. For example, the models in this study and [5, 6] have 3 parameters; the model introduced in [8] has 4 parameters; the models in [9, 10] have 5 parameters; the model in [11] has 6 parameters. The maximum likelihood estimates (MLEs) of such model parameters are based on the log-likelihood function. However, traditional methods of maximization of a log-likelihood function of such mixture models sometimes do not provide the expected result due to multiple optimal points. For mixture modes, we usually rely on the expectation and maximization (EM) algorithm [12]. However, the EM is a local search procedure and therefore there is no guarantee that it converges to the global maximum. As an alternative to EM, we use the CE algorithm which most of the time will provide the global maximum [13].

Bayesian estimation is considered under the symmetric and asymmetric loss functions. Bayesian approach to parameter estimation of lifetime distributions has been considered by many authors such as [14, 15, 16, 17]. Likewise, MCMC methods make Bayesian inference for such models easier and more practical. However, the efficiency of some MCMC algorithms relies on a good choice of the proposal distribution. Here we apply the HMC simulation method. HMC is a Markov chain Monte Carlo algorithm that avoids the random walk behavior and sensitivity to correlated parameters that plague many MCMC methods by taking a series of steps informed by first-order gradient information. More specifically, it is a generalization of the Metropolis algorithm that includes 'momentum' variables so that each iteration can move further in parameter space, thus allowing faster mixing and moving much more rapidly through the target distribution, especially in high dimensions. These features allow it to converge to high-dimensional target distribution much more quickly than simpler methods such as random walk Metropolis or Gibbs sampling [18].

The benchmark data sets, e.g. the aircraft windshield failure data [19], the lifetime data of male mice exposed to 300 rads of radiation [20], and the U.S.S. Halfbeak Diesel engine data [21] are examples showing that the NLFR model is more appropriate than LFR model, Weibull model and some mixture models for modeling these data sets, especially when in data files occur failures with more than one failure mode.

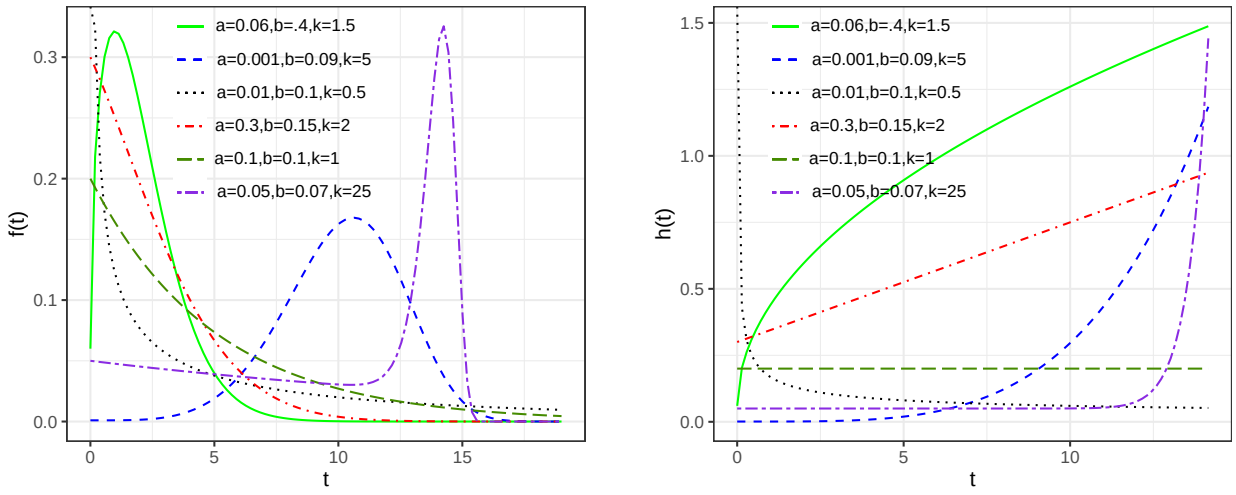
The paper is organized as follows. The intended NLFR model is introduced in Section 2 along with its reliability characteristics. The properties of the NLFR model including the moments and order statistics are given in Section 3. Section 4 specifies the parameters that need to be estimated. Section 5 introduces the CE method for the optimization and MLEs of the unknown parameters of the model. The HMC methods, loss functions and model selection criteria for Bayesian inference are given Section 6. Section 7 provides the applications of the NLFR model to some real data sets. Finally, Sections 8 brings conclusions.

2. The NLFR model and its reliability characteristics

2.1. The NLFR model

From the beginning and early age of operation, a physical system suffers only from random shocks which means that the failure rate experience a constant failure mode. When the system wears out, due to mechanical wear, it also experiences wear-out failure mode. So let the failure rate function of the system in these two situations be in either of the following two states: (1) initially, it experiences a constant failure model, and (2) when the system wears out, it also experiences wear-out failure model. That is

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



(a) **(b)**

Figure 1: (a) Probability density functions and (b) the corresponding failure rate functions of the NLFR.

This model allows an initial positive failure rate, $h(0) = a > 0$, whereas $h(0) = 0$ for most other increasing failure rate function. As mentioned in [3] for the LFR, this type of situation would exist if failures result from random as well as wear out or deterioration mode. This model is also useful for modeling a series system with two independent components. One component follows the exponential distribution and the other component follows the Weibull distribution.

The second component in Eq. (1) is the Weibull failure rate and it has a different parameterization compared to the model in our conference proceedings [7]. We chose this new parameterization because as we have demonstrated in [22] the Weibull model with failure rate function form $h(t) = bt^{k-1}$ shows high correlated parameters which affect Bayes estimators. Fig. 1 shows that the NLFR possesses various shapes of the probability density function (PDF) and the corresponding failure rate function which characterizes three of the most common types of the failure rate function that is increasing, decreasing and constant.

2.2. The reliability characteristics of NLFR model

Here we derive the reliability characteristics of the NLFR model. Using the relationship between the reliability and failure rate functions, the reliability/survival function is given by

$$R(t) = \exp\left\{-at - (bt)^k\right\}$$

Then, the PDF function is given as

$$f(t) = \left(a + kb(bt)^{k-1}\right) \exp\left\{-at - (bt)^k\right\} \tag{2}$$

The cumulative failure rate (CFR) function is given by

$$H(t) = at + (bt)^k$$

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

$$MTTF = \int_0^\infty \exp\left\{-at - (bt)^k\right\} dt \tag{3}$$

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

3. Properties of the NLFR model

3.1. The moments

The rth non-central moment or the rth moment about the origin of the NLFR model can be derived as follows by using the Taylor expression of e^x.

$$\begin{aligned}\mu'_r &= \int_0^{+\infty} x^r dF(x) = -\int_0^{+\infty} x^r de^{-ax-(bx)^k} \\ &= \frac{r}{kb^r}\sum_{n=0}^{+\infty}\frac{(-a)^n}{n!\,b^n}\Gamma\left(\frac{n+r}{k}\right) \end{aligned} \tag{4}$$

for $r = 1, 2, \ldots$, where $\Gamma(\cdot)$ is the gamma function.

3.2. Order statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample from the NLFR distribution and $X_{i:n}$ is the ith order statistic of the sample, then the PDF of $X_{i:n}$ is given by

$$f_{i:n}(x) = \frac{1}{B(i, n-i+1)}F(x)^{i-1}(1-F(x))^{n-i}f(x)$$

where $B(\cdot,\cdot)$ is the beta function.

Since we have

$$(1-F(x))^{n-i} = e^{-(n-i)H(x)}$$

and

$$F(x)^{i-1} = (1-e^{-H(x)})^{i-1} = \sum_{l=0}^{i-1}\binom{i-1}{l}(-1)^l e^{-lH(x)}$$

Therefore,

$$\begin{aligned} f_{i:n}(x) &= \frac{1}{B(i, n-i+1)}\sum_{l=0}^{i-1}\binom{i-1}{l}(-1)^l h(x)e^{-(n+l+1-i)H(x)} \\ &= n\binom{n-1}{i-1}\sum_{l=0}^{i-1}\binom{i-1}{l}\frac{(-1)^l}{n+l+1-i}f(x; a', b', k) \end{aligned}$$

where $a' = (n+l+1-i)a$ and $b' = b\sqrt[k]{n+l+1-i}$

Using (4), the rth non-central moment of the ith order statistics $X_{i:n}$ is

$${\mu'}_r^{(i:n)} = \frac{nr}{kb'^r}\binom{n-1}{i-1}\sum_{m=0}^{+\infty}\sum_{l=0}^{i-1}\binom{i-1}{l}\frac{(-1)^l(-a'^m)}{(n+l+1-i)m!\,b'^m}\Gamma\left(\frac{m+r}{k}\right)$$

4. Target parameters to estimate

Let $\mathcal{D}$: $t_1, ..., t_n$ be a random sample from the NLFR model with parameter $\boldsymbol{\theta} = (a, b, k)$. If there is no censoring, the likelihood function takes the form

$$\begin{aligned} L(\mathcal{D}|\boldsymbol{\theta}) &= \prod_{i=1}^{n} f(t_i|\boldsymbol{\theta}) \\ &= \left[\prod_{i=1}^{n}\left(a + kb(bt_i)^{k-1}\right)\right]\exp\left\{-\sum_{i=1}^{n}\left(at_i + (bt_i)^k\right)\right\} \end{aligned} \tag{5}$$

Then, the corresponding log-likelihood function can be derived as

$$\log L(\mathcal{D}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left(a + kb(bt_i)^{k-1}\right) - \sum_{i=1}^{n} \left(at_i + (bt_i)^k\right) \tag{6}$$

If some observations are right-censored, the likelihood function takes the form

$$\begin{aligned} L(\mathcal{D}|\boldsymbol{\theta}) &= \prod_{i=1}^{n} f(t_i|\boldsymbol{\theta})^{\delta_i} R(t_i|\boldsymbol{\theta})^{1-\delta_i} \\ &= \prod_{i=1}^{n} h(t_i|\boldsymbol{\theta})^{\delta_i} R(t_i|\boldsymbol{\theta}) \\ &= \left[\prod_{i=1}^{n} \left(a + kb(bt_i)^{k-1}\right)^{\delta_i}\right] \exp\left\{-\sum_{i=1}^{n} \left(at_i + (bt_i)^k\right)\right\} \end{aligned} \tag{7}$$

where $\delta_i = 1$ if t_i is an observed failure, and $\delta_i = 0$ if t_i is a censored observation. And in this case, the log-likelihood function for right-censored case is

$$\log L(\mathcal{D}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \log\left(a + kb(bt_i)^{k-1}\right) - \sum_{i=1}^{n} \left(at_i + (bt_i)^k\right) \tag{8}$$

The target here is to find an estimate of the vector parameter $\boldsymbol{\theta}$ that maximizes the log-likelihood function, and estimates of the MTTF, reliability and failure rate functions.

5. Numerical approach 1

5.1. The cross-entropy method for continuous multi-extremal optimization

Suppose we want to maximize a function $S(\mathbf{x})$, $\mathbf{x} \in \mathcal{X}$. We denote γ^* as the maximum, therefore

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

This deterministic problem is then randomized by setting a family of PDFs $\{f(\cdot;\mathbf{v}), \mathbf{v} \in \mathcal{V}\}$ on $\mathcal{X}$. Then the associated stochastic problem is defined as the estimation of

$$\ell(\gamma) = \mathbb{P}\left(S(\mathbf{X}) \geq \gamma\right) = \mathbb{E} I_{\{S(\mathbf{X})\geq\gamma\}}$$

for a given γ, where $\mathbf{X} \sim f(\cdot;\mathbf{u})$, $\mathbf{u} \in \mathcal{V}$ (e.g. $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$) and I is an indicator function. When γ is chosen close to γ^*, $\{S(\mathbf{X}) \geq \gamma\}$ is a rare event, and estimation of ℓ is non-trivial. The CE method solves this by making adaptive changes to the PDF according to the Kullback-Leibler CE, thus creating a sequence $f(\cdot;\mathbf{u}), f(\cdot;\mathbf{v}_1), f(\cdot;\mathbf{v}_2), \ldots$ 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 $(\gamma^*, \mathbf{v}^*)$ [23]. The generic CE algorithm for optimization can be summarized as follows:

1: Choose some $\hat{\mathbf{v}}_0$. Set $t = 1$.

2: Generate a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from the density $f(\cdot;\hat{\mathbf{v}}_{t-1})$ and compute the sample $(1-\varrho)$-quantile $\hat{\gamma}_t$ of the performances as $\hat{\gamma}_t = S_{\lceil(1-\varrho)N\rceil}$, where the quantity ϱ is choosing not very small, say $\varrho = 10^{-2}$.

3: Use the same sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and solve the stochastic program

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

4: Smooth out the vector $\tilde{\mathbf{v}}_t$ by applying

$$\hat{\mathbf{v}}_t = \alpha\tilde{\mathbf{v}}_t + (1-\alpha)\hat{\mathbf{v}}_{t-1}$$

where α is called the smoothing parameter, $0.7 < \alpha \leq 1$.

5: Repeat steps 2-4 until a pre-specified stopping criterion is met.

When the multivariate normal distribution with independent components is used for updating, i.e. $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_{t-1}, \hat{\boldsymbol{\sigma}}^2_{t-1})$, then $\hat{\boldsymbol{\mu}}_t \to \mathbf{x}^*$ and $\hat{\boldsymbol{\sigma}}_t \to \mathbf{0}$. More detail for such explanation can be found in [23]. The R code for this optimization method is given in Appendix A.

5.2. Maximum likelihood estimation

The MLE of the parameter $\boldsymbol{\theta} = (a, b, k)$, say $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{k})$, is obtained by using the CE algorithm to optimize the the log-likelihood function Eq. (6) for non-censored data or Eq. (8) for censored data. And using the invariance property of MLE,

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

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

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

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

(3) The MLE for MTTF will be $\hat{MTTF} = MTTF(\hat{a}, \hat{b}, \hat{k})$ which can be obtained by installing into formula (3) and integrating.

6. Numerical approach 2

6.1. Hamiltonian Monte Carlo

We provide a brief introduction of Hamiltonian Monte Carlo based on what was introduced in [18, 24, 25]. MCMC was first introduced in [26] as a method for simulating the distribution of states for a system of idealized molecules. Later, another approach to molecular simulation was introduced in [27], in which the motion of the molecules was deterministic, following Newton's law of motion, which have an elegant formalization as Hamiltonian dynamics. In 1987, Duane and others merged the MCMC and molecular dynamics together [28]. The method was named "hybrid Monte Carlo," which abbreviates to "HMC,". However the name "Hamiltonian Monte Carlo," retaining the abbreviation, is more specific and descriptive. Duane and others applied HMC not to molecular simulation, but to lattice field theory simulations of quantum chromodynamics. Statistical applications of HMC began with [29] as using it for neural network models. Apparently, HMC seems to be under-appreciated by statisticians, and perhaps also by physicists outside the lattice field theory community [24].

Hamiltonian dynamics

In physics, Hamiltonian dynamics is used to describe how objects move throughout a system. Hamiltonian dynamics describes an object's motion in terms of its location $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ and momentum $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_d)$ (object's mass times its velocity) at certain time t. For each location the object takes, there is an associated potential energy $U(\boldsymbol{\theta})$ (the height of the surface at a given position), and for each momentum there is an associated kinetic energy $K(\boldsymbol{\phi})$. The total energy of the system, known as Hamiltonian $H(\boldsymbol{\theta}, \boldsymbol{\phi})$, is constant and defined as the sum of the potential and kinetic energies:

$$H(\boldsymbol{\theta}, \boldsymbol{\phi}) = U(\boldsymbol{\theta}) + K(\boldsymbol{\phi})$$

The partial derivatives of the Hamiltonian determine how $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ change over time t according to Hamiltonian equations:

$$\frac{\partial \theta_i}{\partial t} = \frac{\partial H}{\partial \phi_i} = \frac{\partial K(\boldsymbol{\phi})}{\partial \phi_i}, \quad i = 1, \ldots, d \tag{9}$$

$$\frac{\partial \phi_i}{\partial t} = -\frac{\partial H}{\partial \theta_i} = -\frac{\partial U(\boldsymbol{\theta})}{\partial \theta_i}, \quad i = 1, \ldots, d \tag{10}$$

Based on the expression of $\frac{\partial U(\boldsymbol{\theta})}{\partial \theta_i}$ and $\frac{\partial K(\boldsymbol{\phi})}{\partial \phi_i}$ and the initial location $\boldsymbol{\theta}_0$ and initial momentum $\boldsymbol{\phi}_0$ of an object at time t_0, we can predict the location and momentum of the object at any time $t = t_0 + T$ by simulating these dynamics for a duration T.

The leapfrog method for simulating Hamiltonian dynamics

The leapfrog method for simulating Hamiltonian dynamics for a duration T is performed by updating the location and momentum variables. A leapfrog step updating the location variable $\boldsymbol{\theta}$ and the momentum variable $\boldsymbol{\phi}$ over ϵ units time, starting at time t, is as follows:

(1) Take a half time-step to update the momentum variable

$$\phi_i(t + \epsilon/2) = \phi_i(t) - \frac{\epsilon}{2}\frac{\partial U}{\partial \theta_i}(\theta_i(t))$$

(2) Take a full time-step to update the position variable

$$\theta_i(t + \epsilon) = \theta_i(t) + \epsilon\frac{\partial K}{\partial \phi_i}(\phi_i(t + \epsilon/2))$$

(3) Take the remaining half time-step to finish updating the momentum variable

$$\phi_i(t + \epsilon) = \phi_i(t + \epsilon/2) - \frac{\epsilon}{2}\frac{\partial U}{\partial \theta_i}(\theta_i(t + \epsilon))$$

We can run L leapfrog steps to simulate Hamiltonian dynamics over ϵL units of time.

Potential energy, kinetic energy and the target distribution

The target distribution $\pi(\boldsymbol{\theta}|\mathcal{D})$ (posterior distribution) that we wish to sample can be related to a potential energy function via the concept of a canonical distribution from statistical mechanics. Given some energy function $E(x)$ for the state x of some physical system, the canonical distribution over states has probability density function

$$\pi(x) = \frac{1}{c}e^{-E(x)/T}$$

where T is the temperature of the system and c is the normalizing constant needed for this function to sum or integrate to one. For Hamiltonian Monte Carlo simulation, we choose $T = 1$.

The Hamiltonian $H(\boldsymbol{\theta}, \boldsymbol{\phi})$ is an energy function for the joint state of location $\boldsymbol{\theta}$ and momentum $\boldsymbol{\phi}$. Therefore, following the canonical distribution for energy function, a joint distribution for them is defined as follows:

$$\begin{aligned}\pi(\boldsymbol{\theta}, \boldsymbol{\phi}) &\propto e^{-H(\boldsymbol{\theta},\boldsymbol{\phi})} \\ &= e^{-U(\boldsymbol{\theta})}e^{-K(\boldsymbol{\phi})} \\ &\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi})\end{aligned}$$

We see that $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are independent and each has canonical distribution, with energy functions $U(\boldsymbol{\theta})$ and $K(\boldsymbol{\phi})$. The potential energy $U(\boldsymbol{\theta})$ will be defined to be minus the log pdf of the target distribution for $\boldsymbol{\theta}$

that we wish to sample. The kinetic energy $K(\boldsymbol{\phi})$ is usually defined as minus of the log pdf of the zero-mean multivariate normal distribution with covariance matrix M. Therefore,

$$U(\boldsymbol{\theta}) = -\log \pi(\boldsymbol{\theta}|\mathcal{D})$$
$$K(\boldsymbol{\phi}) = \frac{\boldsymbol{\phi}^T M^{-1} \boldsymbol{\phi}}{2}$$

Here M is a symmetric, positive definite mass matrix, which is typically diagonal, and is often a scalar multiple of the identity matrix. With these forms for U and K, Hamiltonian equations (9) and (10) can be rewritten as follows, for $i = 1, \ldots, d$:

$$\frac{\partial \theta_i}{\partial t} = [M^{-1}\boldsymbol{\phi}]_i$$
$$\frac{\partial \phi_i}{\partial t} = \frac{\partial \log \pi(\boldsymbol{\theta}|\mathcal{D})}{\partial \theta_i}$$

Hamiltonian Monte Carlo algorithm

HMC uses Hamiltonian dynamics rather than a probability distribution as a proposal function to propose future states for Markov chain in order to explore the target distribution more effectively. Starting with the current state $(\boldsymbol{\theta}, \boldsymbol{\phi})$, we simulate Hamiltonian dynamics for a short time using the leapfrog method. Then the final state of the position and momentum variables of the simulation are used as the proposal states $(\boldsymbol{\theta}^*, \boldsymbol{\phi}^*)$ for Markov chain. The proposed state is accepted according to an update rule which is similar to the Metropolis acceptance rule. Specifically if the probability of the proposed state after Hamiltonian dynamics

$$\pi(\boldsymbol{\theta}^*, \boldsymbol{\phi}^*) \propto e^{-U(\boldsymbol{\theta}^*) - K(\boldsymbol{\phi}^*)} \tag{11}$$

is greater than probability of the state prior to the Hamiltonian dynamics

$$\pi(\boldsymbol{\theta}, \boldsymbol{\phi}) \propto e^{-U(\boldsymbol{\theta}) - K(\boldsymbol{\phi})} \tag{12}$$

then the proposed state is accepted, otherwise, the proposed state is accepted randomly. If the proposed state is rejected, the next state is the same as the current state. The HMC algorithm for drawing N samples from the target distribution is described as follows:

(1) Given a starting position state $\boldsymbol{\theta}^{(0)}$

(2) For $i = 0$ to $N - 1$

1: Draw a momentum variable $\boldsymbol{\phi}^{(i)} \sim \pi(\boldsymbol{\phi})$

2: Run the leapfrog algorithm starting at $(\boldsymbol{\theta}^{(i)}, \boldsymbol{\phi}^{(i)})$ for L steps with step-size ϵ to obtain proposed state $(\boldsymbol{\theta}^*, \boldsymbol{\phi}^*)$

3: Calculate the acceptance probability

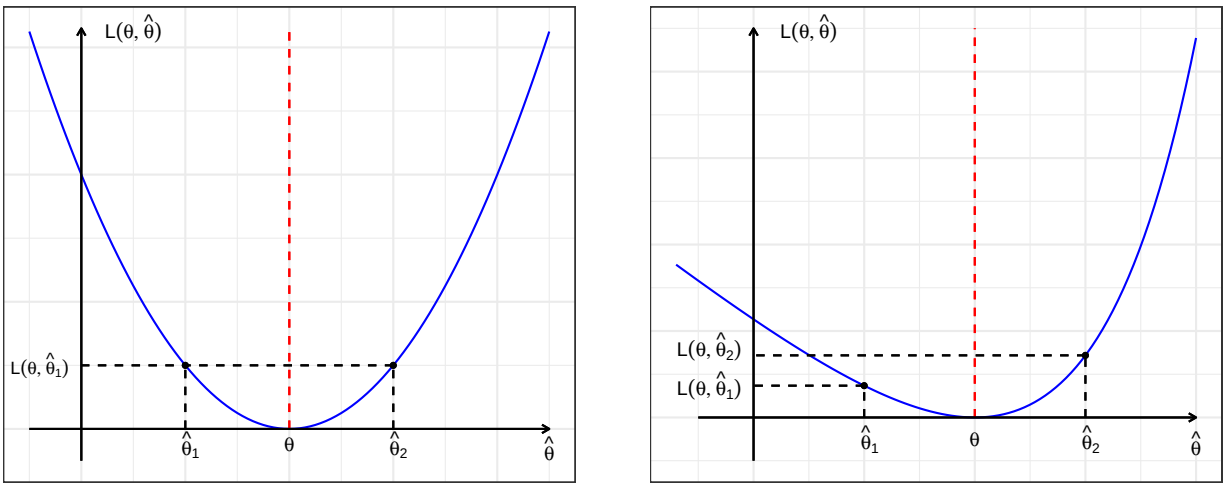
$$r = \min\left(1, e^{U(\boldsymbol{\theta}^{(i)}) - U(\boldsymbol{\theta}^*) + K(\boldsymbol{\phi}^{(i)}) - K(\boldsymbol{\phi}^*)}\right)$$

4: Draw $u \sim U(0, 1)$

5: Set

$$\boldsymbol{\theta}^{(i+1)} = \begin{cases} \boldsymbol{\theta}^* \text{ if } u \leq r \\ \boldsymbol{\theta}^{(i)} \text{ otherwise} \end{cases}$$

There is no need to keep track of $\boldsymbol{\phi}^{(i)}$ after the accept/reject step because we do not care about it in itself, and it immediately get updated at the beginning of the next iteration.



(a) **(b)**

Figure 2: (a) Squared error loss function and (b) linear exponential loss function.

Restricted parameters and areas of zero density

HMC is build to work with all continuous positive target densities. If the algorithm reaches a point of zero density at any point during an iteration, the stepping will be stopped and given up, spending another iteration at the previous value of location $\boldsymbol{\theta}$. This resulting algorithm allows the chain stays in the positive area and preserves detailed balance. An alternative way is to check if the density is positive after each step and, otherwise, change the sign of the momentum to return to the direction in which it came. Another usual way to handle bounded parameters is to use transformation, e.g. taking the logarithm of a positive parameter or the logit for a parameter restricted to fall between 0 and 1, or more complicated joint transformation [18].

Setting the tuning parameters and the No-U-Turn sampler

Besides the choice of the momentum distribution (usually a multivariate normal distribution with mean 0 and covariance set to a prespecified 'mass matrix' M), the efficiency of the HMC depends also on the choice of the scaling factor ϵ of the leapfrog step and the number of leapfrog steps L per iteration. Recently, a variance of HMC called No-U-Turn sampler [30] has been proposed to automatically update these parameters during the burn-in (or warm-up) period and then held fixed during the later iteration. Notice that the HMC method can only be used for continuous distributions. The R code for HMC simulation is given in Appendix B.

6.2. Loss functions

In estimation problems, it is natural for the loss function to be a function of the distance between the true value of the parameter θ and its estimated value, $\hat{\theta}$. The most widely used loss criterion in one parameter estimation problems is squared error loss (SEL), that is,

$$L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$$

SEL is a symmetric function (see Fig. 2a) that penalizes overestimation and underestimation equally, and takes the value zero when the estimate is right on target [31].

The Bayes estimator of θ under the SEL function is the value $\hat{\theta}$ which minimizes $\mathbb{E}\left[(\hat{\theta} - \theta)^2 | \mathcal{D}\right]$. That is

$$\hat{\theta}_{BS} = \mathbb{E}\left[\theta | \mathcal{D}\right]$$

where $\mathbb{E}(\cdot|\mathcal{D})$ denotes the posterior expectation with respect to the posterior density of θ.

The use of symmetric loss function may be inappropriate in the estimation of reliability function as has been recognized in [32]. Overestimate of reliability function or mean failure time is usually much more serious than underestimate of reliability function or mean failure time. Also, an underestimate of the failure rate results in more serious consequences than an overestimate of the failure rate. For example, in the disaster of the space shuttle [33] the management underestimated the failure rate and therefore overestimated the reliability of solid-fuel rocket booster [34].

Varian [35] motivated the use of asymmetric loss functions in estimation problems arising in real estate assessment, where the overestimation of a property's value might cause it to remain on the market unsold for an extended period, ultimately costing the seller inordinate and unnecessary expenses. The estimation of peak water flow in the construction of dams or levees has asymmetric consequences; overestimation might lead to increased construction costs while underestimation might lead to the much more serious consequence of subsequent overflows which can seriously threaten lives and property in adjacent communities [31].

There are many forms of asymmetric loss functions. One of the more widely use versions of asymmetric loss is the linear exponential loss (LINEX loss or LEL) function which can be expressed as

$$L(\theta, \hat{\theta}) = e^{c(\hat{\theta}-\theta)} - c(\hat{\theta} - \theta) - 1, \qquad c \neq 0$$

In case $c > 0$, this function rises exponentially when $\hat{\theta} > \theta$ and approximately linearly when $\hat{\theta} < \theta$ (see Fig. 2b). The sign and magnitude of the parameter c represents the direction and degree of symmetry, respectively. If $c > 0$, the overestimation is more serious than underestimation, and vice versa. For c close to zero, the LINEX loss is approximately SEL and therefore almost symmetric [31].

The posterior expectation of the LINEX loss function is

$$\mathbb{E}\left[L(\theta, \hat{\theta})|\mathcal{D}\right] = e^{c\hat{\theta}}\mathbb{E}\left[e^{-c\theta}|\mathcal{D}\right] - c\left(\hat{\theta} - \mathbb{E}[\theta|\mathcal{D}]\right) - 1 \tag{13}$$

The Bayes estimator of θ under the LINEX loss function is the value $\hat{\theta}$ which minimizes Eq. (13). That is

$$\hat{\theta}_{BL} = -\frac{1}{c}\log\left\{\mathbb{E}\left[e^{-c\theta}|\mathcal{D}\right]\right\}$$

provided that the expectation $\mathbb{E}\left[e^{-c\theta}|\mathcal{D}\right]$ exists and is finite.

The modified LINEX loss is the general entropy loss (GEL) function, defined as:

$$L(\theta, \hat{\theta}) = \left(\frac{\hat{\theta}}{\theta}\right)^c - c\log\left(\frac{\hat{\theta}}{\theta}\right) - 1$$

The Bayes estimator of θ under the GEL function is given as

$$\hat{\theta}_{BG} = \left(\mathbb{E}\left[\theta^{-c}|\mathcal{D}\right]\right)^{-\frac{1}{c}}$$

provided that the expectation $\mathbb{E}\left[\theta^{-c}|\mathcal{D}\right]$ exists and is finite. It can be shown that, when $c = 1$, this loss becomes the entropy loss and the Bayes estimator $\hat{\theta}_{BG}$ coincides with the Bayes estimator under the weighted squared-error loss function. Similarly, when $c = -1$ the Bayes estimator $\hat{\theta}_{BG}$ coincides with the Bayes estimator under squared error loss function [36].

6.3. Bayesian estimation

In Bayesian inference, we consider the unknown vector parameter $\boldsymbol{\theta} = (a, b, k)$ as a random variable and derive its probability distribution based on prior knowledge and available data. Using Bayes' rule, the posterior distribution of $\boldsymbol{\theta}$ given the data $\mathcal{D}$: $t_1, \ldots, t_n$ is derived as

$$\pi(\boldsymbol{\theta}|\mathcal{D}) = \frac{L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{14}$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution for $\boldsymbol{\theta}$ and $L(\mathcal{D}|\boldsymbol{\theta})$ is the likelihood function given in Eq. (5) or (7). Since the denominator in Eq. (14) is a normalizing constant and not necessary for Bayesian inference using MCMC methods, the posterior distribution is often expressed as:

$$\pi(\boldsymbol{\theta}|\mathcal{D}) \propto L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{15}$$

As mentioned in [37], 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 we assume the independence gamma prior distributions for a, b and k, i.e.

$$\begin{aligned}
\pi_1(a) &\propto a^{\alpha_1-1}\exp\{-\beta_1 a\}\,,\ \alpha_1,\beta_1>0\\
\pi_2(b) &\propto b^{\alpha_2-1}\exp\{-\beta_2 b\}\,,\ \alpha_2,\beta_2>0\\
\pi_3(k) &\propto k^{\alpha_3-1}\exp\{-\beta_3 k\}\,,\ \alpha_3,\beta_3>0
\end{aligned}$$

Since we use proper prior distributions, the posterior distribution is also proper. As we can see, the prior distributions also depend on their parameters, called hyperparameters. In Bayesian analysis, the hyperparameters are usually assumed to be known in advance. These hyperparameters can also be estimated from the given data by using the empirical Bayes method. Here we choose the hyperparameters for gamma priors such that their means approximate the MLEs of the parameters. These prior distributions are also applied for the LFR and Weibull models in a later section.

Then, under the SEL function, the Bayes estimators of a, b, k, failure rate function $h(t)$ and reliability function $R(t)$ are given by

$$\begin{aligned}
a^*_{BS} &= \mathbb{E}(a|\mathcal{D}) = \int a\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\\
b^*_{BS} &= \mathbb{E}(b|\mathcal{D}) = \int b\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\\
k^*_{BS} &= \mathbb{E}(k|\mathcal{D}) = \int k\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\\
h^*_{BS}(t) &= \mathbb{E}(h(t;\boldsymbol{\theta})|\mathcal{D}) = \int h(t;\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\\
R^*_{BS}(t) &= \mathbb{E}(R(t;\boldsymbol{\theta})|\mathcal{D}) = \int R(t;\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}
\end{aligned}$$

Suppose that $\{\boldsymbol{\theta}_i, i=1,\ldots,N\}$ is generated from the posterior distribution $\pi(\boldsymbol{\theta}|\mathcal{D})$ using HMC simulation method. Then when i is sufficiently large (say, bigger than n_0), $\{\boldsymbol{\theta}_i, i=n_0+1,\ldots,N\}$ is a (correlated) sample from the true posterior. Then, the approximate Bayes estimates of a^*_{BS}, b^*_{BS}, k^*_{BS}, $h^*_{BS}(t)$ and $R^*_{BS}(t)$ are obtained by calculating the means:

$$\begin{aligned}
a^*_{BS} &\approx \frac{1}{N-n_0}\sum_{i=n_0+1}^{N} a_i\\
b^*_{BS} &\approx \frac{1}{N-n_0}\sum_{i=n_0+1}^{N} b_i\\
k^*_{BS} &\approx \frac{1}{N-n_0}\sum_{i=n_0+1}^{N} k_i\\
h^*_{BS}(t) &\approx \frac{1}{N-n_0}\sum_{i=n_0+1}^{N} h(t;\boldsymbol{\theta}_i)
\end{aligned}$$

$$R^*_{BS}(t) \approx \frac{1}{N-n_0}\sum_{i=n_0+1}^{N} R(t;\boldsymbol{\theta}_i)$$

However, in practice, we usually run m parallel chains (say, $m = 3, 4$ or 5), instead of only 1, for assessing sampler convergence. Then the posterior means are obtained as follows

$$\begin{aligned}
a^*_{BS} &\approx \frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} a_{i,j}\\
b^*_{BS} &\approx \frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} b_{i,j}\\
k^*_{BS} &\approx \frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} k_{i,j}\\
h^*_{BS}(t) &\approx \frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} h(t;\boldsymbol{\theta}_{i,j})\\
R^*_{BS}(t) &\approx \frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} R(t;\boldsymbol{\theta}_{i,j})
\end{aligned}$$

The Bayes estimators of the parameters, failure rate and reliability functions under LEL function are given respectively by

$$\begin{aligned}
a^*_{BL} &= -\frac{1}{c}\log\left(\int e^{-ca}\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\right)\\
b^*_{BL} &= -\frac{1}{c}\log\left(\int e^{-cb}\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\right)\\
k^*_{BL} &= -\frac{1}{c}\log\left(\int e^{-ck}\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\right)\\
h^*_{BL}(t) &= -\frac{1}{c}\log\left(\int e^{-ch(t;\boldsymbol{\theta})}\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\right)\\
R^*_{BL}(t) &= -\frac{1}{c}\log\left(\int e^{-cR(t;\boldsymbol{\theta})}\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\right)
\end{aligned}$$

The approximate Bayes estimates of a^*_{BL}, b^*_{BL}, k^*_{BL}, $h^*_{BL}(t)$ and $R^*_{BL}(t)$ are given respectively by

$$\begin{aligned}
a^*_{BL} &\approx -\frac{1}{c}\log\left(\frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} e^{-ca_{ij}}\right)\\
b^*_{BL} &\approx -\frac{1}{c}\log\left(\frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} e^{-cb_{ij}}\right)\\
k^*_{BL} &\approx -\frac{1}{c}\log\left(\frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} e^{-ck_{ij}}\right)\\
h^*_{BL}(t) &\approx -\frac{1}{c}\log\left(\frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} e^{-ch(t;\boldsymbol{\theta}_{ij})}\right)
\end{aligned}$$

$$R^*_{BL}(t) \approx -\frac{1}{c}\log\left(\frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} e^{-cR(t;\boldsymbol{\theta}_{ij})}\right)$$

The Bayes estimators of the parameters, failure rate and reliability functions under GEL function are given respectively by

$$\begin{aligned}
a^*_{BG} &= \left(\int a^{-c}\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\right)^{-\frac{1}{c}}\\
b^*_{BG} &= \left(\int b^{-c}\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\right)^{-\frac{1}{c}}\\
k^*_{BG} &= \left(\int k^{-c}\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\right)^{-\frac{1}{c}}\\
h^*_{BG}(t) &= \left(\int h(t;\boldsymbol{\theta})^{-c}\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\right)^{-\frac{1}{c}}\\
R^*_{BG}(t) &= \left(\int R(t;\boldsymbol{\theta})^{-c}\pi(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}\right)^{-\frac{1}{c}}
\end{aligned}$$

The approximate Bayes estimates of a^*_{BG}, b^*_{BG}, k^*_{BG}, $h^*_{BG}(t)$ and $R^*_{BG}(t)$ are given respectively by

$$\begin{aligned}
a^*_{BG} &\approx \left(\frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} a_{ij}^{-c}\right)^{-\frac{1}{c}}\\
b^*_{BG} &\approx \left(\frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} b_{ij}^{-c}\right)^{-\frac{1}{c}}\\
k^*_{BG} &\approx \left(\frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} k_{ij}^{-c}\right)^{-\frac{1}{c}}\\
h^*_{BG}(t) &\approx \left(\frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} h(t;\boldsymbol{\theta}_{ij})^{-c}\right)^{-\frac{1}{c}}\\
R^*_{BG}(t) &\approx \left(\frac{1}{m(N-n_0)}\sum_{j=1}^{m}\sum_{i=n_0+1}^{N} R(t;\boldsymbol{\theta}_{ij})^{-c}\right)^{-\frac{1}{c}}
\end{aligned}$$

6.4. Model selection criteria

Akaike information criterion (AIC). The AIC is defined as

$$AIC = -2\log L(\mathcal{D}|\tilde{\boldsymbol{\theta}}) + 2k$$

where $\tilde{\boldsymbol{\theta}} = \mathbb{E}[\boldsymbol{\theta}|\mathcal{D}]$ is an estimate of the parameter vector $\boldsymbol{\theta}$ based on it posterior distribution, k is the dimension of $\boldsymbol{\theta}$. A model with the minimum value of AIC is considered to be the best-approximating model among a set of alternative models.

Bayesian information criterion (BIC). The BIC is defined by

$$BIC = -2\log L(\mathcal{D}|\tilde{\boldsymbol{\theta}}) + \log(n)k$$

where n is the number of observations. The definition of the BIC is almost the same as the AIC except that the penalty factor for using k parameters is $\log(n)$. When comparing several models, the model with the lowest BIC fits best.

Corrected Akaike information criterion (AICc). The AICc is the AIC with a correction for small sample sizes. The formula for AICc is as follows

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1}$$

As $n \to \infty$, the AICc converges to AIC.

Deviance information criterion (DIC). The DIC is proposed based on the principle DIC = 'goodness of fit'+'complexity' [38]. The DIC is defined as

$$DIC = \overline{D} + p_D$$

where $\overline{D}$ is a measure of fit

$$\overline{D} = \mathbb{E}_{\boldsymbol{\theta}|\mathcal{D}} \left[-2 \log L(\mathcal{D}|\boldsymbol{\theta})\right]$$

and p_D is a measure of "effective number of parameters"

$$p_D = \mathbb{E}_{\boldsymbol{\theta}|\mathcal{D}} \left[-2 \log L(\mathcal{D}|\boldsymbol{\theta})\right] + 2 \log L(\mathcal{D}|\tilde{\boldsymbol{\theta}})$$

Here $\tilde{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$ that can be usually chosen as $\tilde{\boldsymbol{\theta}} = \mathbb{E}\left[\boldsymbol{\theta}|\mathcal{D}\right]$. DIC is easily computed via MCMC methods. A model with the smallest value of DIC is considered to be the best-approximating model among a set of alternative models (see also [39]).

7. Applications

In this section, we bring three examples to illustrate the proposed model and estimation procedures discussed in this paper. Here we present mainly the results of the Bayes estimators under the SEL function. In case we have good knowledge about the consequences of overestimation and underestimation of a particular problem which we are concerning, then we can choose the appropriate loss function for the estimation.

7.1. Aircraft windshield failure data

Table 1 contains the failure data of aircraft windshields which are given in [19, p. 574]. Among 153 observations, there are 88 failed observed windshields and 65 censored observations. The unit of measurement of this data is 1000 hours. It has been shown in our previous study [7] that the NLFR fits this data set better than many mixture distributions in terms of the AIC measure. In this study, we provide a Bayes estimation of the NLFR when fitted to this data set and give a comparison with the Weibull and LFR models.

To obtain the Bayes estimates of the parameters and reliability characteristics, we implemented the HMC algorithm to simulate samples from the posterior distribution (15). We constructed 4 parallel Markov chains with different starting points and each of length 2000 with burn-in (warm-up) of 1000 and we obtained final posterior sample of size 1000 for each chain.

Fig. 3 shows the trace plots after removing the burn-in period and density estimates of the parameters obtained by HMC algorithm. The trace plots show that the 4 parallel chains for each parameter produced by HMC algorithm converge quickly to the same target distribution. The densities are distributed approximately symmetrically around the central values which means that the simulated samples provide good Bayes estimates under SEL function. The scatter plot matrix of HMC output shows the posterior correlations between the parameters (Fig. 4). The graph shows that all pairs of parameters have very small correlation.

Table 2 shows the MLE and Bayes point estimates and two-sided 90% and 95% highest posterior density (HPD) intervals for a, b, k and MTTF. From the table, we can see that the estimate of a is positive and far from zero and its credible intervals also support for this claim. Fig. 5 displays the Bayes estimates of

Table 1: 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	

reliability and failure rate functions when fitting NLFR, LFR and Weibull models to the windshield failure data. It is easy to see that the NLFR fits to the data better than LFR and Weibull models. Table 3 provides the values of the log-likelihood (Log-lik), AIC, BIC, AICc and DIC for fitting the NLFR, LFR and Weibull models to the windshield failure data. As we can see NLFR has largest Log-lik value and smallest values for the other criteria. Therefore, it is considered as the best model for this data set among the given models. However, the LFR and Weibull models can also be considered as appropriate candidates for this data set.

7.2. Mice data

Data in Table 4 represent the days until death for male mice exposed to 300 rads of radiation [20, p. 389]. The unit for measurement is day. We consider here only the group was maintained in a germ-free environment and the causes of death is due to the effect of other causes. The new feature of male mice data is that more than one failure mode occurs [20].

Table 2: MLEs, Bayes estimates and HPD intervals for the parameters and MTTF for fitting NLFR to windshield data.

Parameter	MLE	Bayes	90% HPD	95% HPD
a	0.0268	0.0268	$[0.0207, 0.0331]$	$[0.0197, 0.0344]$
b	0.2785	0.2776	$[0.2607, 0.2958]$	$[0.2562, 0.2989]$
k	2.9260	2.9092	$[2.5193, 3.2860]$	$[2.4964, 3.4198]$
$MTTF$	3.0519	3.0646	$[2.8850, 3.2607]$	$[2.8441, 3.2967]$

MLE: Maximum likelihood estimate; HPD: Highest posterior density; MTTF: mean time to failure

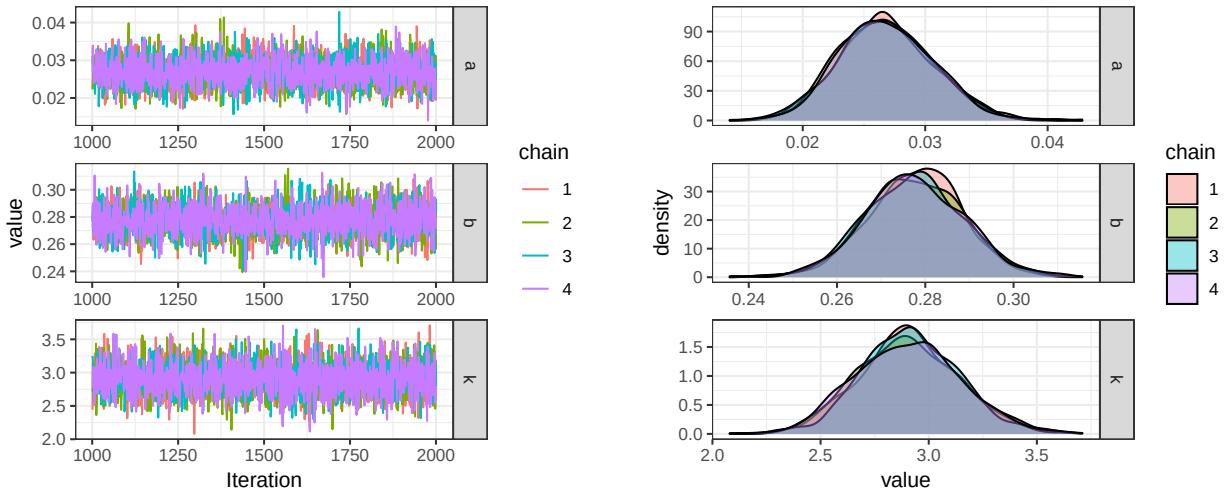


Figure 3: (a) Trace plots and (b) approximate posterior densities of the parameters produced by HMC sampling with 4 parallel chains when fitting NLFR to windshield data.

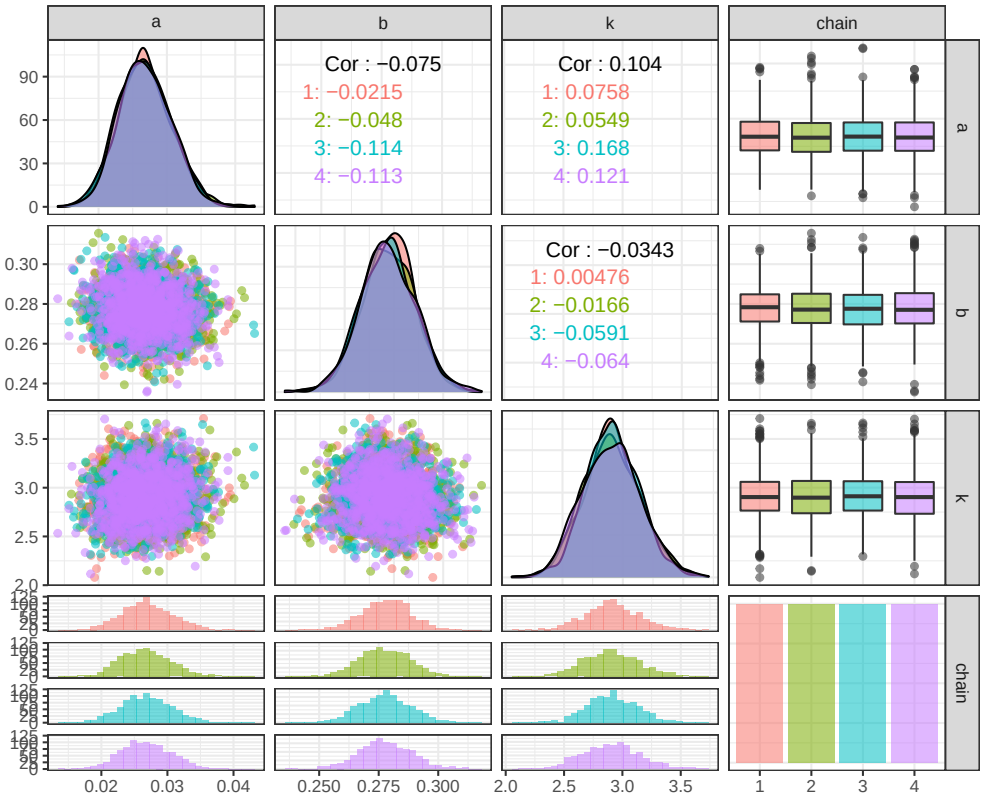
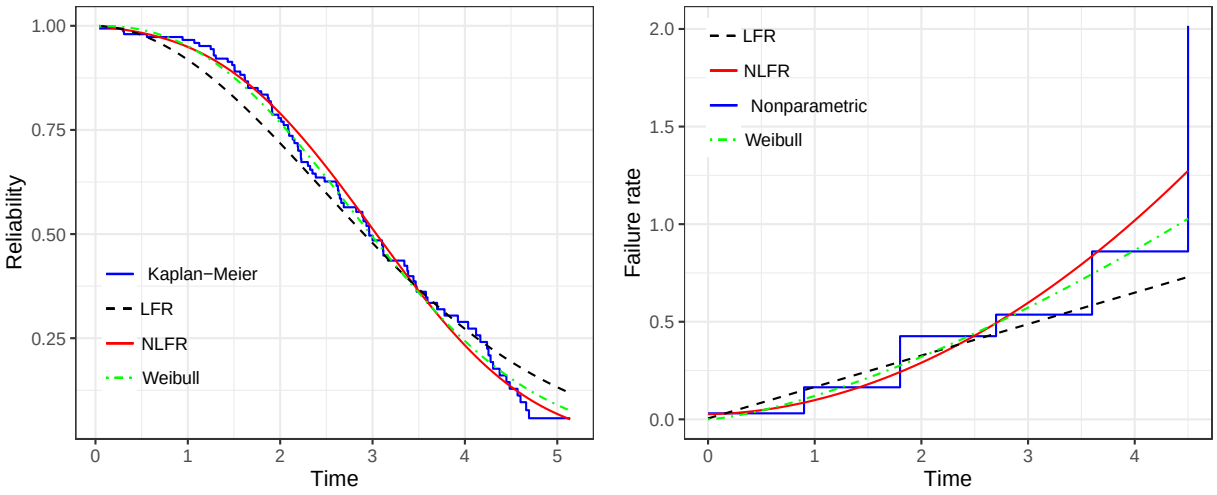


Figure 4: Scatter plot matrix of HMC output with 4 parallel chains obtained by fitting NLFR to windshield data.



(a) (b)

Figure 5: Bayes estimates of the (a) reliability function and (b) failure rate function when fitting NLFR, LFR and Weibull models to windshield data. Kaplan-Meier is the nonparametric estimate of the reliability function for right censored data.

Table 3: Log-likelihood, AIC, BIC, AICc and DIC when fitting NLFR, LFR and Weibull distributions to windshield data.

Model	Log-lik	AIC	BIC	AICc	DIC
NLFR	−170.69	347.38	356.47	347.54	344.71
LFR	−176.55	357.11	363.17	357.18	354.39
Weibull	−174.06	352.11	358.18	352.20	351.29

Log-lik: log-likelihood; AIC: Akaike information criterion; BIC: Bayesian information criterion; AICc: corrected Akaike information criterion; DIC: deviance information criterion; NLFR: nonlinear failure rate; LFR: linear failure rate

Table 4: Male mice exposed to 300 rads of radiation (other causes in germ-free group)

136	246	255	376	421	565	616
617	652	655	658	660	662	675
681	734	736	737	757	769	777
800	807	825	855	857	864	868
870	870	873	882	895	910	934
943	1015	1019				

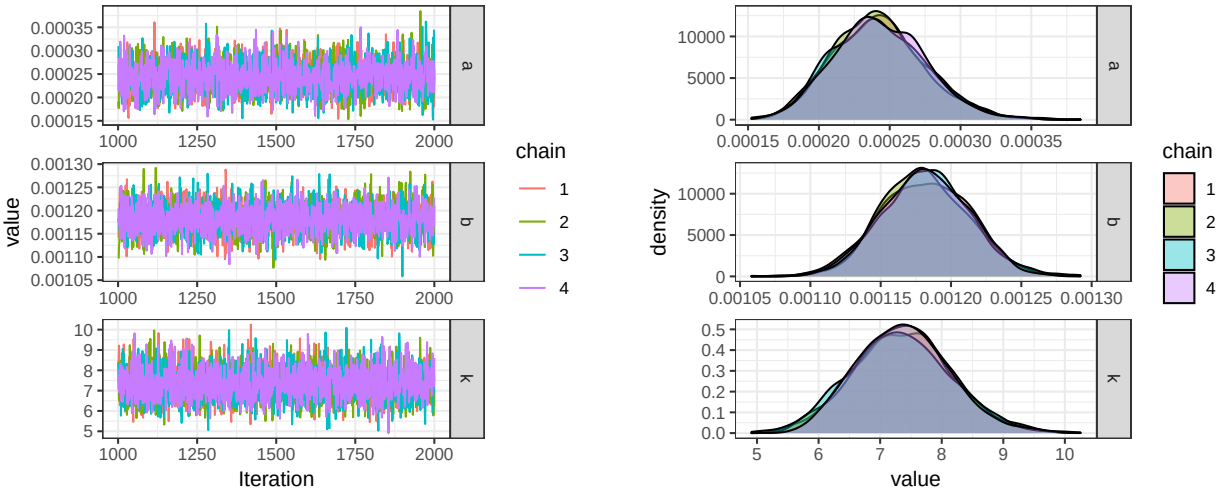


Figure 6: (a) Trace plots and (b) approximate posterior densities of the parameters produced by HMC sampling with 4 parallel chains when fitting NLFR to mice data.

For this data set, we used the same procedure described in Section 7.1 for HMC simulation from posterior distribution. The trace plots and density estimates of the parameters of HMC output are given in Fig. 6. The trace plots also show that the 4 parallel chains for each parameter converged quite fast and converged to the same target distribution. The densities are also distributed approximately symmetrically around the central values. The scatter plot matrix of HMC output shows very small posterior correlations between the parameters (Fig. 7), thanks to the reparameterization.

The MLE and Bayes point estimates and two-sided 90% and 95% HPD intervals for a, b, k and MTTF are shown in Table 5. The estimate of a is still positive but smaller than the previous case. Fig. 8 displays the Bayes estimates of the reliability and failure rate functions when fitting NLFR, LFR and Weibull models to the mice data. In this case, the NLFR also fits to the data better than the LFR and Weibull models. The values of the Log-lik, AIC, BIC, AICc and DIC when fitting the NLFR, LFR and Weibull models to the data are provided in Table 6 which shows that NLFR has largest Log-lik value and smallest values for the other criteria. As we can see here, the LFR model is not appropriate for modeling this data set.

7.3. USS-halfback Diesel Engine data

Table 7 gives times of unscheduled maintenance actions for the U.S.S. Halfbeak number 4 main propulsion diesel engine over 25, 518 operating hours [21, p. 395]. The unit for measurement is hour. The data are times of recurrent events on one machine, hence not typical lifetime data. Here, we assume that the time from one maintenance to the following one is as lifetime data to demonstrate our proposed model for uncensored data.

Here we also used the same procedure described in Section 7.1 for HMC simulation. The trace plots and density estimates of the parameters obtained by HMC algorithm are given in Fig. 9. The trace plots also

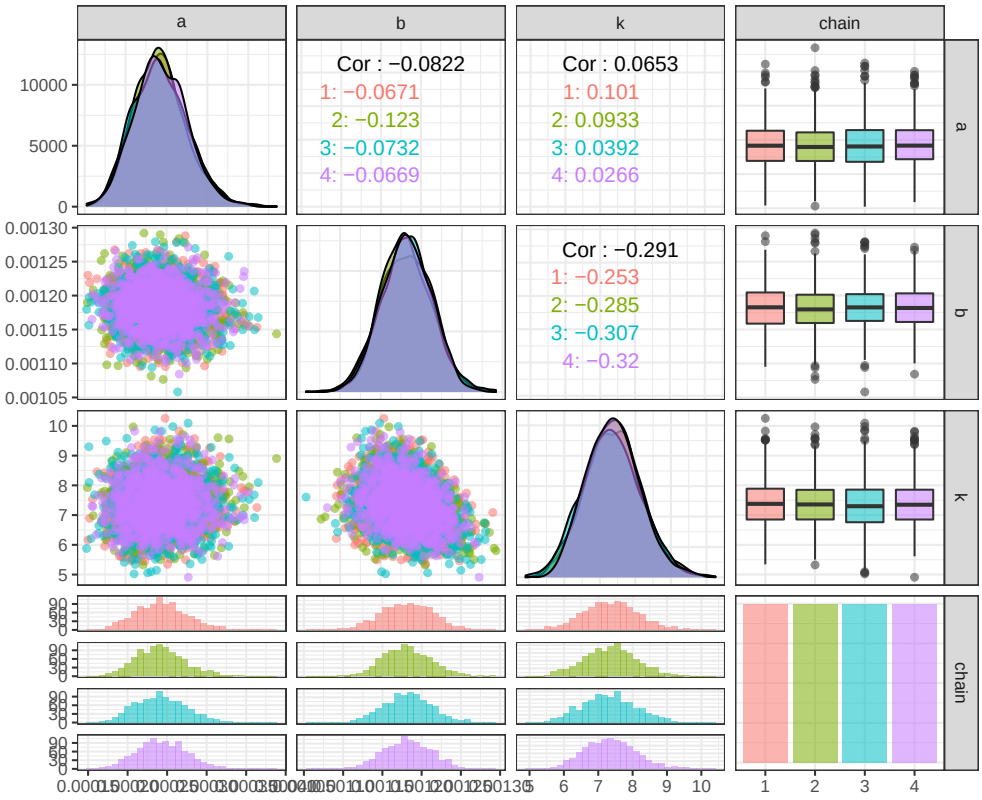


Figure 7: Scatter plot matrix of HMC output with 4 parallel chains obtained by fitting NLFR to mice data.

Table 5: MLEs, Bayes estimates and HPD intervals for the parameters and MTTF for fitting NLFR to mice data

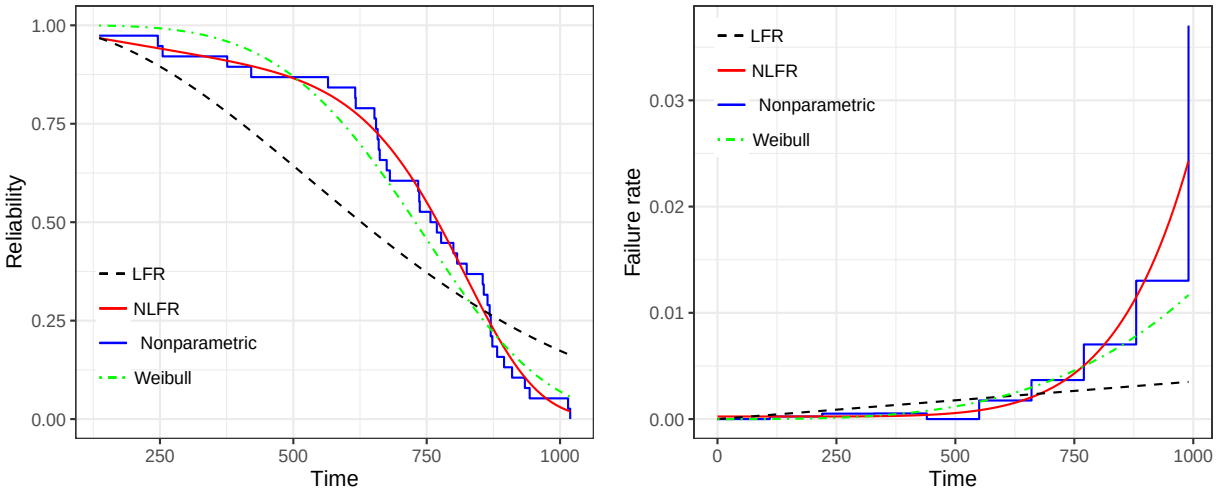
Parameter	MLE	Bayes	90% HPD	95% HPD
a	0.0002	0.0002	$[0.0002, 0.0003]$	$[0.0002, 0.0003]$
b	0.0012	0.0012	$[0.0011, 0.0012]$	$[0.0011, 0.0012]$
k	7.4383	7.3629	$[6.1454, 8.6862]$	$[5.8989, 8.9388]$
$MTTF$	720.20	720.39	$[689.07, 753.79]$	$[682.66, 759.64]$

MLE: Maximum likelihood estimate; HPD: Highest posterior density; MTTF: mean time to failure

Table 6: Log-likelihood, AIC, BIC, AICc and DIC when fitting NLFR, LFR and Weibull distributions to mice data.

Model	Log-lik	AIC	BIC	AICc	DIC
NLFR	−250.09	506.18	511.10	506.89	503.06
LFR	−267.26	538.51	541.79	538.85	535.35
Weibull	−255.59	515.18	518.46	515.53	514.16

Log-lik: log-likelihood; AIC: Akaike information criterion; BIC: Bayesian information criterion; AICc: corrected Akaike information criterion; DIC: deviance information criterion; NLFR: nonlinear failure rate; LFR: linear failure rate



(a) (b)

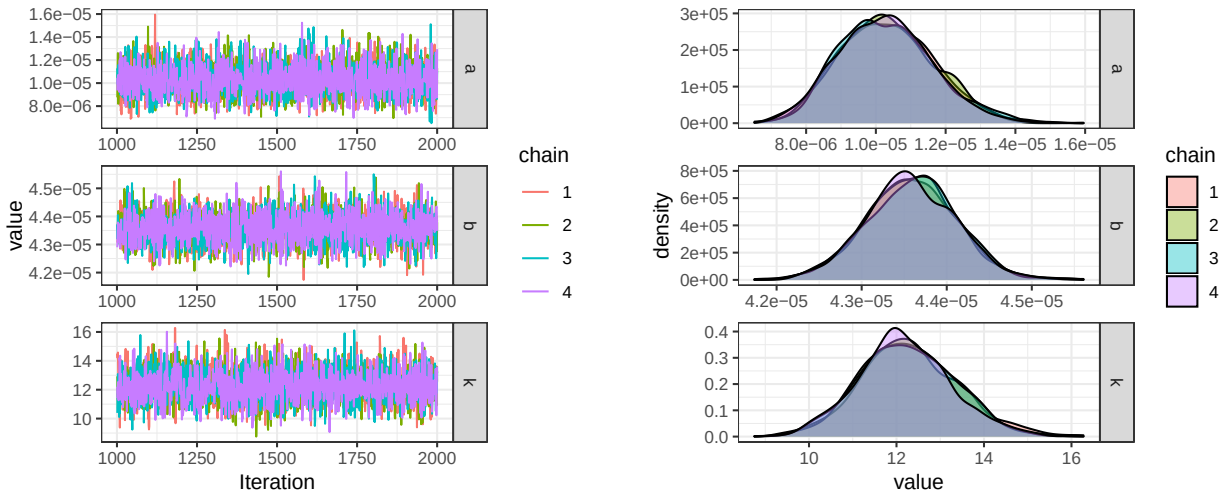
Figure 8: Bayes estimates of the (a) reliability function and (b) failure rate function when fitting NLFR, LFR and Weibull models to mice data.

Table 7: U.S.S. Halfbeak Diesel Engine Data

1382	2990	4124	6827	7472	7567
8845	9450	9794	10848	11993	12300
15413	16497	17352	17632	18122	19067
19172	19299	19360	19686	19940	19944
20121	20132	20431	20525	21057	21061
21309	21310	21378	21391	21456	21461
21603	21658	21688	21750	21815	21820
21822	21888	21930	21943	21946	22181
22311	22634	22635	22669	22691	22846
22947	23149	23305	23491	23526	23774
23791	23822	24006	24286	25000	25010
25048	25268	25400	25500	25518	

show that the 4 parallel chains for each parameter produced by HMC algorithm converge quickly to the same target distribution. The densities are distributed approximately symmetrically around the central values. The scatter plot matrix of HMC output in Fig. 10 also shows very small posterior correlations between the parameters.

Table 8 shows the MLE and Bayes point estimates and two-sided 90% and 95% HPD intervals for a, b, k and MTTF. The estimate of a is still positive but close to zero and the estimate of k shows the shape change in the wear out phase. Fig. 11 displays the Bayes estimates of reliability and failure rate functions when fitting NLFR, LFR and Weibull models to the mice data. In this case, the NLFR fits to the data much better than the LFR and Weibull models. Table 9 provides the values of the Log-lik, AIC, BIC, AICc and DIC when fitting the NLFR, LFR and Weibull models to the data. Again, NLFR has largest Log-lik value and smallest values for the other criteria. For this data set, the NLFR clearly outperform the LFR and Weibull models in term of the AIC, BIC, AICc and DIC measures. Once again, the LFR model is not an appropriate model for the data set in which the failure rate has a shape change in the wear out phase.



(a) **(b)**

Figure 9: (a) Trace plots and (b) approximate posterior densities of the parameters produced by HMC sampling with 4 parallel chains when fitting NLFR to USS-halfback Diesel engine data.

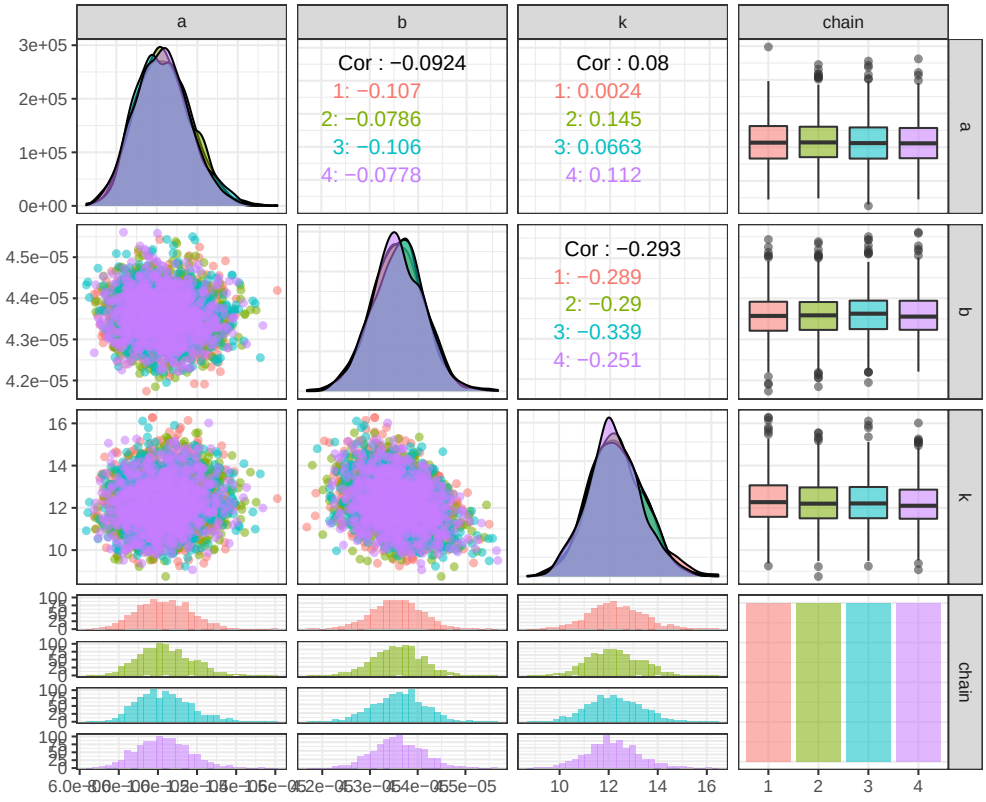
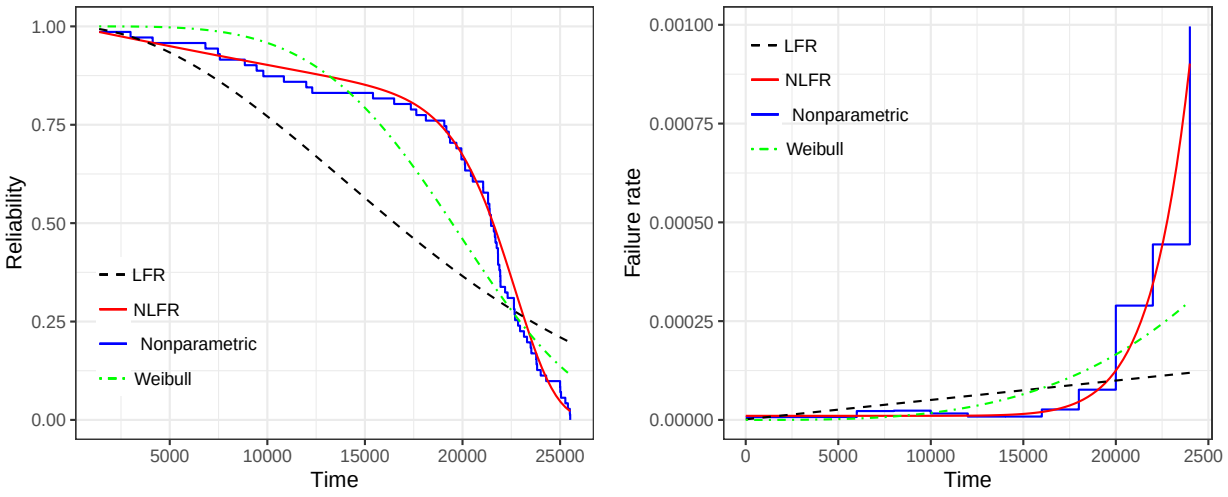


Figure 10: Scatter plot matrix of HMC output with 4 parallel chains obtained by fitting NLFR to USS-halfback Diesel engine data.



(a) (b)

Figure 11: Bayes estimates of the (a) reliability function and (b) failure rate function when fitting NLFR, LFR and Weibull models to USS-halfback Diesel engine data.

Table 8: MLEs, Bayes estimates and HPD intervals for the parameters and MTTF for fitting NLFR to USS-halfback Diesel engine data

Parameter	MLE	Bayes	90% HPD	95% HPD
a	0.00001	0.00001	$[0.000008, 0.000013]$	$[0.000008, 0.000013]$
b	0.00004	0.00004	$[0.000043, 0.000044]$	$[0.000043, 0.000045]$
k	12.4316	12.2552	$[10.47271, 14.00529]$	$[9.952805, 14.21508]$
$MTTF$	19342.6	19666.8	$[19089.23, 20221.54]$	$[19010.52, 20372.36]$

MLE: Maximum likelihood estimate; HPD: Highest posterior density; MTTF: mean time to failure

Table 9: Log-likelihood, AIC, BIC, AICc and DIC when fitting NLFR, LFR and Weibull distributions to USS-halfback Diesel engine data.

Model	Log-lik	AIC	BIC	AICc	DIC
NLFR	-684.73	1375.47	1382.25	1375.82	1373.06
LFR	-735.60	1475.21	1479.73	1475.38	1472.38
Weibull	-716.24	1436.49	1441.01	1436.66	1435.65

Log-lik: log-likelihood; AIC: Akaike information criterion; BIC: Bayesian information criterion; AICc: corrected Akaike information criterion; DIC: deviance information criterion; NLFR: non-linear failure rate; LFR: linear failure rate

8. Conclusions

In this paper, we have proposed the NLFR model with a new parameterization which differs from all previous studies. The new paramterization shows vary small correlated parameters which is good for Bayesian estimation using the posterior mean. The properties and reliability characteristics of the model have been studied. The cross-entropy method for continuous multi-extremal optimization has been used to produce MLE of the parameters. Bayes estimators are obtained under symmetric and asymmetric loss functions. The HMC simulation method has been employed to simulate data from the posterior distribution. Three benchmark data sets have been used to illustrate the superiority of the proposed model and the Bayesian estimation method. Since the NLFR model fits to the data sets better than the Weibull model, future work could be focused on proportional hazard model based on NLFR instead of the Weibull model.

Acknowledgement

The authors would like to express thanks to the anonymous referees for their constructive comments and valuable suggestions. This work was partly 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 partly by the VSB-Technical University of Ostrava, project "Applied Statistics" No. SP2019/16.

Appendix A. R code for cross-entropy optimization method

In this section we provide R code for cross-entropy optimization method using the "CEoptim" package [40]. We provide here the R code for the mice data analyzed in Section 7.2.

```
#input the mice data into R: For simplicity we use the scan() function here.
survtime <- scan()
136 246 255 376 421 565 616 617 652 655
658 660 662 675 681 734 736 737 757 769
777 800 807 825 855 857 864 868 870 870
873 882 895 910 934 943 1015 1019
#define the log-likelihood function
LogLikelihood <- function(x,t) {
  sum(log(x[1]+x[3]*x[2]*(x[2]*t)^(x[3]-1))-(x[1]*t+(x[2]*t)^x[3]))
}
#define initial values for the cross-entropy algorithm
mu0 <- c(0,0,10)
sigma0 <- c(0.001,0.001,10)
#restrict the parameter space
A=diag(-1,3)
B=c(0,0,0)
#invoke the CEoptim library
library(CEoptim)
#execute the CEoptim function
opt <- CEoptim(LogLikelihood, f.arg = list(t=survtime), maximize = T,
                    continuous = list(mean=mu0,sd=sigma0, conMat=A, conVec=B,
                       smoothMean=0.7, smoothSd=0.7, sdThr=0.00001), rho = 0.01, N=1000)
#optimum value
opt$optimum
#optimizer (MLE)
CE <- opt$optimizer$continuous
```

Appendix B. R code for HMC simulation via Stan

In this study, we use Rstan [41] to sample from posterior distribution. Rstan is the R interface to Stan [42] which provides full Bayesian inference using the No-U-Turn sampler (NUTS) [30], a variant of HMC. The Stan code for our mode for the mice data is given below:

```
functions{
  real model_log(vector x, real a, real b, real k){
    vector[num_elements(x)] term;
    real loglik;
    for (i in 1:num_elements(x)) {
      term[i] = log(a+k*b*(b*x[i])^(k-1))-(a*x[i]+(b*x[i])^k);
    }
    loglik = sum(term);
    return loglik;
  }
}
data{
  int N;
  vector[N] Y;
}
parameters{
  real<lower=0> a;
  real<lower=0> b;
  real<lower=0> k;
}
model{
  a ~ gamma(50,2.064566e+05);
  b ~ gamma(50,4.227379e+04);
  k ~ gamma(50,6.721977e+00);
  Y ~ model(a,b,k);
}
```

To implement the model we need to create a Stan file with the above code and save the file as `NLFRmice.stan`.

```
#after installing the rstan package, we invoke it by
library("rstan")
#run the Stan model
fit = stan(file = "NLFRmice.stan",
         data   = list(Y=survtime,N=38),
         iter   = 2000,
         chains = 4)
#model summary
print(fit, digits_summary = 4)
#samples vs. iteration plot
traceplot(fit, nrow=3)
traceplot(fit, nrow=3, inc_warmup=TRUE)
#posterior samples for the 3 parameters.
dat = as.matrix(fit)[,1:3]
#Bayes estimates for the parameters
est = round(apply(dat, 2, mean),digits = 4)
#invoke package that contains HPDinterval function
library(coda)
```

```
#90% HPD intervals for the parameters
round(HPDinterval(as.mcmc(dat), prob=0.90), digits = 4)
#95% HPD intervals for the parameters
round(HPDinterval(as.mcmc(dat), prob=0.95), digits = 4)
```

References

[1] Montgomery DC, Runger GC. Applied statistics and probability for engineers. 7th ed. Wiley; 2018.
[2] Kodlin D. A new response time distribution. Biometrics 1967;2:227-239.
[3] Bain LJ. Analysis for the linear failure-rate life-testing distribution. Technometrics 1974;16:551-559.
[4] Pandey A, Singh A, Zimmer WJ. Bayes estimation of the linear hazard-rate model. IEEE Transactions on Reliability 1993;42:636-640.
[5] Salem SA. Bayesian estimation of a non-linear failure rate from censored samples type II. Microelectronics and Reliability 1992;32:1385-1388.
[6] Bousquet N, Bertholon H, Celeux G. An alternative competing risk model to the Weibull distribution for modeling aging in lifetime data analysis. Lifetime Data Analysis 2006;12:481-504.
[7] Thach TT, Bris R. Non-linear failure rate: A comparison of the Bayesian and frequentist approaches to estimation, Proceedings of International Conference on Mathematics ICM 2018, ITM Web of Conferences 20, paper 03001 (2018) https://doi.org/10.1051/itmconf/20182003001, Open Access, ISBN: 978-7598-9058-3. Edited by: Radim Bris, Gyu Whan Chang, Chu Duc Khanh, Mohsen Razzaghi, Krzysztof Stempak, Phan Thanh Toan.
[8] Xie M, Lai CD. Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. Reliability Engineering & System Safety 1995;52:87-93.
[9] Almalki SJ, Yuan J. A new modified Weibull distribution. Reliability Engineering & System Safety 2013;111:164–70.
[10] He B, Cui W, Du X. An additive modified Weibull distribution. Reliability Engineering & System Safety 2016;145:28-37.
[11] Wang FK. A new model with bathtub-shaped failure rate using an additive Burr XII distribution. Reliability Engineering & System Safety 2000;70(3):305-312.
[12] McLachlan GJ, Krishnan T. The EM algorithm and extensions. 2nd ed. Wiley-Interscience; 2008.
[13] Rubinstein RY, Kroese DP. The cross-entropy method: a unified approach to combinatorial optimization, Monte Carlo simulation and machine learning. Springer-Verlag, New York; 2004.
[14] Gupta A, Mukherjee B, Upadhyay SK. Weibull extension model: A Bayes study using Markov chain Monte Carlo simulation. Reliability Engineering & System Safety 2008;93(10):1434–1443.
[15] Upadhyay SK, Gupta A, A Bayesian analysis of modified Weibull distribution using Markov Chain Monte Carlo simulation, Journal of Statistical Computation and Simulation 2010;80(3):241–254.
[16] Soliman AA, Abd-Ellah AH, Abou-Elheggag NA, Essam AA. Modified Weibull model: A Bayes study using MCMC approach based on progressive censoring data. Reliability Engineering & System Safety 2012;100:48-57.
[17] Kundu D, Howlader H. Bayesian inference and prediction of the inverse Weibull distribution for type-II censored data. Computational Statistics and Data Analysis 2010;54(6):1547-1558.
[18] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3rd ed. CRC Press; 2014.
[19] Blischke WR, Karim MR, Murthy DP. Warranty data collection and analysis.
[20] Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. 2nd ed. Wiley, New York; 2002.
[21] Meeker WQ, Escobar LA. Statistical methods for reliability data. Wiley; 1998.
[22] Thach TT and Bris R. Reparameterized Weibull distribution: a Bayes study using Hamiltonian Monte Carlo, Proceedings of the 29th European safety and reliability conference. Singapore: Research Publishing, 2019, pp.997–1004. Edited by: Beer M and Zio E.
[23] Kroese DP, Porotsky S, Rubinstein RY. The cross-entropy method for continuous multi-extremal optimization. Methodology and Computing in Applied Probability 2006;8:383-407.
[24] Neal RM. MCMC using Hamilton dynamics. In: Brooks S, Gelman A, Jones G, Meng XL, editors. Handbook of Markov chain Monte Carlo. CRC Press; 2011.
[25] Stansbury DE. MCMC: Hamiltonian Monte Carlo (a.k.a. Hybrid Monte Carlo) (2012). `https://theclevermachine.wordpress.com/2012/11/18/mcmc-hamiltonian-monte-carlo-a-k-a-hybrid-monte-carlo/`
[26] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics 1953;21:1087-1092.
[27] Alder B, Wainwright T. Studies in molecular dynamics I. General method. Journal of Chemical Physics 1959;31(2):459-466.
[28] Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. Physics Letters B 1987;195(2):216-222.
[29] Neal RM. Bayesian Learning for Neural Networks, Lecture Notes in Statistics 118. Springer, New York; 1996.
[30] Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research 2014;15:1593–1623.
[31] Samaniego FJ. A comparison of the Bayesian and Frequentist approaches to estimation. Springer; 2010.
[32] Canfield R. A Bayesian approach to reliability estimation using a loss function. IEEE Transactions on Reliability 1970;19:13-16.
[33] Feynman RP. Mr. Feynman goes to Washington. Engineering and Science, California Institute of Technology, Pasadena, CA 1987;6-22.
[34] Basu AP, Ebrahimi N. Bayesian approach to life testing and reliability estimation using asymmetric loss function. Journal of Statistical Planning and Inference 1991;29:21-31.

[35] Varian HR. A bayesian approach to real estate assessment. In: L. J. Savage, S. E. Feinberg and A. Zellner, Eds., Studies in Bayesian Econometrics and Statistics: In Honor of L. J. Savage, North-Holland Pub. Co., Amsterdam 1975;195-208.
[36] Singh PK, Singh SK, Singh U. Bayesian estimator of inverse Gaussian parameters under general entropy loss function using Lindley's approximation, Communication in Statistics-Simulation and Computation 2010;37(9):1750-1762.
[37] Jiang H, Xie M, Tang LC. Markov chain Monte Carlo methods for parameter estimation of the modified weibull distribution. Journal of Applied Statistics 2008;35(6):647-658.
[38] Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. The deviance information criterion: 12 years on. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2014;76(3):485–493.
[39] Banerjee S, Carlin BP, Gelfand AE. Hierarchical Modeling and Analysis for Spatial Data. 2nd ed. CRC press; 2014.
[40] Benham T, Duan Q, Kroese DP, Liquet B. CEoptim: Cross-Entropy R Package for Optimization. Journal of Statistical Software 2017;76(8):1-29.
[41] Stan Development Team. RStan: the R interface to Stan. R package version 2.18.2. `http://mc-stan.org/rstan/`.
[42] The Stan Development Team. Stan Modeling Language User's Guide and Reference Manual. `http://mc-stan.org/`.