Reliability Engineering and System Safety 167 (2017) 95–104

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

A nonparametric Bayesian modeling approach for heterogeneous lifetime data with covariates

Mingyang Li [a,*], Hongdao Meng [b], Qingpeng Zhang [c]

[a] Department of Industrial and Management Systems Engineering, University of South Florida, Tampa, FL, United States
[b] School of Aging Studies, University of South Florida, Tampa, FL, United States
[c] Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong SAR, China

ARTICLE INFO

Keywords:
Bayesian modeling
Mixture model
Lifetime regression
Gibbs sampler
Model selection

ABSTRACT

Lifetime data collected at product design and production stage or field operational stage often exhibit heterogeneity patterns, making the homogeneity assumption in conventional statistical lifetime models invalid. Mixture models are important modeling approaches that account for data heterogeneity. However, existing mixture models are constrained by assuming an known number of sub-populations. This paper proposes a new Bayesian statistical model to analyze heterogeneous lifetime data by assuming an unknown number of sub-populations. Each sub-population is characterized by an accelerated failure time model to quantify the effects of possible reliability impact factors. The proposed model allows simultaneous identification of the number of sub-populations and the model parameters of sub-populations. Convenient sampling strategies are further proposed to address the challenges of model estimation. Both numerical case study and real case study are provided to illustrate the proposed approach and demonstrate its validity.

1. Introduction

Statistical modeling and analysis of lifetime data is an essential component for reliability assessment and prediction. A typical assumption of conventional statistical reliability models is to assume that product population is homogeneous. However, in many reliability engineering applications, such homogeneity assumption may not be valid. In the modern semiconductor industry, due to the non-uniformity of oxide film thickness resulting from poor wafer uniformity and oxide growth control, some Metal-Oxide-Semiconductor field effect transistor (MOSFET) units with thinner films will tend to breakdown earlier than other units with thicker films under the same voltage stress [1]. In the automobile industry, due to the unknown change in raw material properties or supplier quality and improperly unverified design changes, early failures are often reported in the warranty databases from a large number of standard products in the field [2]. As also reported in [3,4], the heterogeneity issue of product lifetime becomes even more obvious for many immature manufacturing processes with evolving technologies, such as new drilling products with evolving Micro-Electro-Mechanical System (MEMS) technology developed by Baker Hughes Corporation. Neglecting the product lifetime heterogeneity may greatly affect the performance of various reliability assurance tasks, such as inaccurate reliability assessment, inappropriate testing designs and less cost-effective maintenance planning.

To account for the heterogeneity issue of product reliability, mixture models are important modeling approaches in analyzing the heterogeneous reliability data and have been investigated to address various reliability engineering problems. For instance, Majeske [5] proposed a Weibull-Uniform mixture distribution to characterize heterogeneous lifetime data from the automobile warranty claims database by explicitly taking into account a fraction of vehicles containing manufacturing or assembly defects when leaving the assembly plant. Attardi et al. [6] considered a Weibull mixture model with two sub-populations to quantify a heterogeneous population of vehicles based on field failure records. In [7], a mixture model is adopted to evaluate reliability of space systems operated in remote environments. Yuan and Ji [8] proposed a Bayesian mixture model to characterize the heterogeneous degradation paths of a laser device sample. In [9], a mixture model with two sub-populations is formulated to investigate the heterogeneous remaining useful lifetime of lead-acid batteries. Additional discussion and comparison of mixture models can be also found in [10–12] and references therein.

Existing mixture models in analyzing heterogeneous lifetime data have several limitations. For instance, many existing mixture models assume a known number of sub-populations before model estimation based on prior knowledge [5,6] or graphical visualization [13,14]. Such

* Corresponding author.
E-mail address: mingyangli@usf.edu (M. Li).

http://dx.doi.org/10.1016/j.ress.2017.05.029
Received 29 February 2016; Received in revised form 1 May 2017; Accepted 18 May 2017
Available online 23 May 2017
0951-8320/

determination of the number of sub-populations is often subjective. To objectively determine the number of sub-populations from data, several model selection techniques, such as likelihood ratio tests [15,16] and information criterion [17,18], are adopted. These methods involve a two-step procedure (i.e., model estimation and model selection) and require estimating and comparing multiple models. It will be desirable to objectively determine the number of sub-populations with a one-step procedure and estimate only a single model. Another limitation of heterogeneity modeling under the mixture model framework is that mixture of distributions (e.g., Weibull distributions [10,19], Gamma distributions [20], etc.) are widely investigated without taking into account the possible reliability impact factors, such as stress factors or environmental factors. It will be beneficial to incorporate such reliability impact factors and consider the mixture of regression models.

With the advancements in Bayesian nonparametric statistics, it becomes possible to develop advanced data-driven models with more flexibility and less assumption through integration of statistical models and stochastic processes [21]. Dirichlet process has been embedded into the mixture models to address the aforementioned limitations of assuming a known number of sub-populations in conventional mixture models [22,23]. In [22], Dirichlet process is embedded into the normal distribution and a marginal sampling approach is proposed for model estimation. Motivated by the success in [22], Dirichlet process is further embedded into lifetime distributions, such as Weibull distribution [24] and lognormal distribution [25], to analyze heterogeneous lifetime data. To further incorporate possible reliability impact factors in the context of accelerated life tests, Weibull accelerated failure time (AFT) model is further considered in [26]. Lognormal AFT model is another important class of AFT model but has not been addressed yet.

Moreover, existing model estimation strategies in [24–26] consider the marginal sampling approach. As pointed out by Ishwaran and James [23] and Papaspiliopoulos and Roberts [27], the marginal sampling approach may cause slow mixing of Markov chain and involve numerical integrations. To overcome such limitations, a conditional sampling approach is proposed in [23] and further extended by Papaspiliopoulos and Roberts [27] and Walker [28]. But existing conditional sampling approaches are mainly constrained to distribution-based formulation, such as the normal distribution, and are developed in the context of density estimation [27,28]. In this paper, a Bayesian nonparametric mixture of lognormal AFT model is proposed to analyze both the complete and right-censored heterogeneous lifetime data. Moreover, a new Bayesian estimation strategy based on the conditional sampling approach is proposed. It allows convenient samplings from common distributions and exhibits faster and higher quality of convergence than existing estimation methods in [24–26], which will facilitate the practical implementations for practitioners.

This paper proposes a new Bayesian modeling and estimation approach to analyze heterogeneous lifetime data with covariates. Compared to existing literatures, the contributions of the proposed work include (i) analyzing heterogeneous lifetime data without predetermining a fixed number of sub-populations; (ii) jointly identifying the number of sub-populations and estimating model parameters within each sub-population; (iii) incorporating possible reliability impact factors and quantifying their effects on heterogeneous lifetime data; (iv) convenient Bayesian sampling strategies for both complete and right-censored lifetime data. The remainder of this paper is organized as follows: Section 2.1 introduces the proposed model formulation; Section 2.2 addresses the corresponding model estimation challenges and the proposed estimation strategies; Section 3 illustrates the proposed work and demonstrates its validity through both a numerical case study and a real case study. Section 4 draws the conclusions.

2. Methodology

This section first introduces the conventional mixture model framework and describes its transition and connection to the proposed model formulation. Then, detailed estimation procedure are described to address the model estimation challenges.

2.1. Model formulation

Consider a heterogeneous population made up of multiple homogeneous sub-populations and assume each homogeneous sub-population can be modeled by a lognormal regression. When product unit i is known to belong to sub-population k, its lifetime observation, t_i, can be modeled explicitly by an AFT model as

$$\log(t_i|\boldsymbol{\beta}_k, \mathbf{x}_i) = \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i + \epsilon_i, i = 1, \ldots, n_k, \tag{1}$$

where $\mathbf{x}_i = [1, x_{i1}, \ldots, x_{ip_k}]^{\mathrm{T}}$ is a $(p_k + 1)$-dimensional vector of covariates representing the possible reliability impact factors for sub-population k, such as stress factors, environmental factors, etc. $\boldsymbol{\beta}_k = [\beta_{k0}, \ldots, \beta_{kp_k}]^{\mathrm{T}}$ is the corresponding $(p_k + 1)$-dimensional covariate coefficient vector to quantify the influence of possible covariates. ϵ_i is a random error term. Different specification of error term yields different AFT models. For instance, if error term is specified as the extreme value distribution, Eq. (1) becomes Weibull AFT model. Such model can characterize a variety of failures, such as dielectric breakdown, ball bearing failures, damage in laminated composites, etc., and has meaningful interpretation of failure mechanism based on the extreme value theory[10]. If error term is specified as the normal distribution, Eq. (1) becomes lognormal AFT model. Such model is applicable to characterizing other failures due to a degradation process, such as corrosion, material diffusion, crack growth propagation [29]. In this paper, lognormal AFT is considered and assume $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \sigma_k^2), i = 1, \ldots, n_k$, where $\mathrm{N}(\cdot)$ denotes the univariate normal distribution, σ_k^2 is the variance parameter and n_k is the sample size of sub-population k.

The heterogeneous patterns of lifetime observations mainly result from the influence of different covariates throughout the multi-stage product lifecycle, such as different design settings in the product design and development stage, different production settings in the manufacturing stage and different operating conditions in the field deployment stage. In practice, it is difficult, if not impossible, to record and observe information of all covariates since product lifecycle tracking and complete root cause analysis may be costly, time-consuming or unavailable due to limited resources and inadequate data collecting, storing and integrating capabilities. In addition, much covariates information, such as random manufacturing errors and defects, cannot be observed. Therefore, after considering the influence of observed covariates, lifetime data heterogeneity may still exist and is defined as latent heterogeneity. The latent heterogeneity is caused by unobserved/unavailable covariates and the mixture model formulation by assuming a heterogeneous population aims to quantify such latent heterogeneity.

Given the covariates, $\mathbf{x}$, lifetime distribution of sub-population k can be essentially characterized by a lognormal probability density function, $f_k(t|\mathbf{x}, \boldsymbol{\theta}_k)$, as

$$f_k(t|\mathbf{x}, \boldsymbol{\theta}_k) = \frac{1}{t\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\log(t) - \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x})^2}{2\sigma_k^2}\right), \tag{2}$$

where $\boldsymbol{\theta}_k$ represents a set of those unknown parameters, i.e., $\boldsymbol{\theta}_k = \{\boldsymbol{\beta}_k, \sigma_k^2\}$. In reality, the sub-population to which t_i belongs to is generally unknown and thus, the membership of t_i becomes an unknown latent variable. If assume the heterogeneous population is composed of K homogeneous sub-populations, t_i has a prior probability belief, denoted as w_k, to be categorized into sub-population k and $\sum_{k=1}^{K} w_k = 1$. Therefore, the probability density function of the heterogeneous population, $f(t|\mathbf{x}, \boldsymbol{\Theta})$, can be given by

$$f(t|\mathbf{x}, \boldsymbol{\Theta}) = \sum_{k=1}^{K} w_k f_k(t|\mathbf{x}, \boldsymbol{\theta}_k), \tag{3}$$

where $\boldsymbol{\Theta}$ represents a collection of all unknown parameters in Eq. (3), i.e., $\boldsymbol{\Theta} = \{w_k, \boldsymbol{\theta}_k\}_{k=1}^{K}$.

A major limitation of this formulation is that the total number of sub-populations, K, needs to be assumed as a known and fixed quantity before model estimation procedure can be performed. While the total number of sub-populations can often be estimated based on domain expert knowledge or graphical visualization, such estimation of K is relatively subjective. A more objective and justifiable alternative is to determine K through a two-step procedures, namely model estimation and model selection. Specifically, a series of candidate models with different fixed and hypothetical values of K are first constructed and estimated. Based on the estimation results, model selection techniques, such as statistical tests (e.g., likelihood ratio test) and information criterion (e.g., Akaike information criterion (AIC)), are employed to choose the best model with the most appropriate K. From a practical point of view, the two-step procedure of constructing and estimating a series of models separately and then selecting the best model is cumbersome. It will be desirable to develop the one-step procedure by jointly estimating model parameters and selecting K.

To overcome these limitations, a Bayesian nonparametric model formulation is present in this paper to (i) assume an unknown number of sub-populations before model estimation; (ii) objectively determine K from data; and (iii) jointly estimate K and $\boldsymbol{\Theta}$ through a one-step procedure. The proposed model formulation can be written in a hierarchical structure as

$$\begin{aligned} &t_i|\mathbf{x}, \boldsymbol{\theta}_{(i)} \sim f(\cdot|\mathbf{x}, \boldsymbol{\theta}_{(i)}), i = 1, \ldots, n, \\ &\boldsymbol{\theta}_{(i)}|P \overset{\text{i.i.d.}}{\sim} P, i = 1, \ldots, n, \\ &P|\gamma, P_0 \sim \text{DP}(\gamma, P_0), \end{aligned} \tag{4}$$

where n is the total number of product units, P is a random distribution assigned by prior of a Dirichlet process (DP), denoted as DP(·), with a positive scalar γ and a base distribution P_0. $f(\cdot|\mathbf{x}, \boldsymbol{\theta}_{(i)})$ characterizes the homogeneous sub-population associated with t_i of product unit i and $\boldsymbol{\theta}_{(i)}$ is a collection of corresponding unknown parameters. In this paper, lognormal AFT model is considered for each sub-population due to its popularity in lifetime modeling as well as its elegant model structure which allows convenient sampling strategies to be developed in Section 2.2. $\boldsymbol{\theta}_{(i)}$ can be the same or different among different product units. To illustrate the connection and difference of the proposed formulation in Eq. (4) compared to the conventional formulation in Eq. (3), denote $\boldsymbol{\theta}_k$ as different values among $\boldsymbol{\theta}_{(i)}$ and further define discrete variable z_i for each t_i as, $z_i = k$ when $\boldsymbol{\theta}_{(i)} = \boldsymbol{\theta}_k$. A more constructive representation of DP(·) [30] is employed, and the realization of DP(·), namely P, can be written as

$$P = \sum_{k=1}^{\infty} w_k \delta_{\boldsymbol{\theta}_k}, \tag{5}$$

where $\delta_{\boldsymbol{\theta}_k}$ is the Dirac delta measure with a point mass of 1 at $\boldsymbol{\theta}_k$. $w_1 = v_1, w_k = \prod_{k'=1}^{k-1}(1 - v_{k'})v_k, \forall k \geq 2$, where $v_k \sim \text{beta}(1, \gamma)$, $\forall k$, and beta(·) denotes the beta distribution. $\boldsymbol{\theta}_k$ can be drawn independently from the base distribution of a Dirichlet process, P_0. With the representation in Eqs. (5) and (4) can be rewritten as

$$\begin{aligned} &t_i|\mathbf{x}, \boldsymbol{\theta}_{z_i} \sim f(\cdot|\mathbf{x}, \boldsymbol{\theta}_{z_i}), i = 1, \ldots, n, \\ &\boldsymbol{\theta}_{z_i} \sim \sum_{k=1}^{\infty} w_k \delta_{\boldsymbol{\theta}_k}, i = 1, \ldots, n. \end{aligned} \tag{6}$$

Based on (6), the probability density function of the heterogeneous population, $g(t|\mathbf{x}, \boldsymbol{\Theta})$, can be given by

$$g(t|\mathbf{x}, \boldsymbol{\Theta}) = \sum_{k=1}^{\infty} w_k f_k(t|\mathbf{x}, \boldsymbol{\theta}_k). \tag{7}$$

Compared to Eq. (3), which is a finite mixture of lognormal AFT model, the proposed model formulation can be viewed as an infinite mixture of lognormal AFT model. With the proposed formulation, there is no restriction on the number of sub-populations assumed before model estimation. The actual value of K will be learned objectively from data. The estimation procedure will be elaborated in the next section.

2.2. Model estimation

To estimate model parameters and identify the number of sub-populations in Eq. (4), suppose n observations of lifetime data with covariates are available from either laboratory tests or field operations. Denote the available data as, $\mathbf{D} = \{t_i, \Delta_i, \mathbf{x}_i\}$, where Δ_i is a right-censored indicator for observation i. $\Delta_i = 1$ if the failure of unit i is observed and $\Delta_i = 0$ if it is censored. Let $\pi(\boldsymbol{\Theta})$ denote the prior density of a collection of all unknown parameters, i.e., $\boldsymbol{\Theta} = \{w_k, \boldsymbol{\beta}_k, \sigma_k^2\}_{k=1}^K$, where $\pi(\cdot)$ represents an arbitrary probability density function. As shown in (5), $\boldsymbol{w}$ and $\boldsymbol{\theta}_k$ are generated independently to construct a realization of a Dirichlet process, where $\boldsymbol{w} = \{w_k\}_{k=1}^{\infty}$ and $\boldsymbol{\theta}_k = \{\boldsymbol{\beta}_k, \sigma_k^2\}$. Thus, joint prior density can be further expressed as, $\pi(\boldsymbol{\Theta}) = \pi(\boldsymbol{w}) \cdot \prod_{k=1}^{\infty} \pi(\boldsymbol{\beta}_k, \sigma_k^2)$. The joint posterior density, therefore, can be given by

$$\begin{aligned} \pi(\boldsymbol{\Theta}|\mathbf{D}) \propto L(\boldsymbol{\Theta}|\mathbf{D}) \cdot \pi(\boldsymbol{\Theta}) = \prod_{i=1}^{n} \Bigg[&\left(\sum_{k=1}^{\infty} w_k f_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2) \right)^{\Delta_i} \\ &\cdot \left(\sum_{k=1}^{\infty} w_k R_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2) \right)^{1-\Delta_i} \Bigg] \cdot \pi(\boldsymbol{w}) \cdot \prod_{k=1}^{\infty} \pi(\boldsymbol{\beta}_k, \sigma_k^2), \end{aligned} \tag{8}$$

where $R_k(t|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2))$ is the reliability function of sub-population k, i.e., $R_k(t|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)) = 1 - \int_0^t f_k(s|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2))\text{d}s$ and $L(\boldsymbol{\Theta}|\mathbf{D})$ is the joint likelihood function.

Under the Bayesian framework, parameter estimation requires to compute the marginal posteriors of each individual unknown parameter. Both conventional sampling techniques, such as inversion sampling [31], and conventional numerical integration techniques, such as Gaussian quadrature approximation [32], fail since computing the marginal posterior requires high dimensional integration of the joint posterior. For instance, the marginal posterior of w_1 can be expressed as, $\pi(w_1|\mathbf{D}) = \int_{\boldsymbol{\Theta}^{(-w_1)}} \pi(\boldsymbol{\Theta}|\mathbf{D})\text{d}\boldsymbol{\Theta}^{(-w_1)}$, where $\boldsymbol{\Theta}^{(-w_1)}$ denotes a collection of parameters by excluding w_1, i.e., $\boldsymbol{\Theta}^{(-w_1)} = \boldsymbol{\Theta}\backslash\{w_1\}$. To overcome such limitations, Markov Chain Monte Carlo (MCMC) sampling methods can be employed. However, the joint likelihood function, $L(\boldsymbol{\Theta}|\mathbf{D})$, is in complex form, and this poses challenges for conventional MCMC sampling methods. The first challenge is that all unknown parameters are highly dependent among each other. Such high dependency will result in slow or even failed convergence of the conventional MCMC. To further explain such high dependency, the full conditional posterior of $(\boldsymbol{\beta}_k, \sigma_k^2)$ can be expressed as

$$\begin{aligned} \pi(\boldsymbol{\beta}_k, \sigma_k^2|\boldsymbol{\Theta}^{(-(\boldsymbol{\beta}_k, \sigma_k^2))}, \mathbf{D}) \propto \prod_{i=1}^{n} \Bigg[&\left(\sum_{k=1}^{\infty} w_k f_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2) \right)^{\Delta_i} \\ &\cdot \left(\sum_{k=1}^{\infty} w_k R_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2) \right)^{1-\Delta_i} \Bigg] \cdot \pi(\boldsymbol{\beta}_k, \sigma_k^2). \end{aligned} \tag{9}$$

As shown in Eq. (9), $(\boldsymbol{\beta}_k, \sigma_k^2)$ depend on all the remaining unknown parameters, i.e., $\boldsymbol{\Theta}^{(-(\boldsymbol{\beta}_k, \sigma_k^2))} = \boldsymbol{\Theta}\backslash\{\boldsymbol{\beta}_k, \sigma_k^2\}$. The second challenge is that there is an infinite number of unknown parameters to be estimated in Eq. (8). It makes the model estimation procedure computationally formidable.

To address the first challenge of high dependency among unknown parameters, the key is to reduce the complexity of $L(\boldsymbol{\Theta}|\mathbf{D})$. Augment $\mathbf{Z} = \{z_i\}_{i=1}^n$ into the likelihood function, $L(\boldsymbol{\Theta}|\mathbf{D})$, where z_i are discrete variables introduced in Section 2.1. z_i can be interpreted as the latent membership of t_i. When $z_i = k$, it indicates the t_i of product unit i belongs to sub-population k. With augmented $\mathbf{Z}$, the joint likelihood function, $L(\boldsymbol{\Theta}|\mathbf{Z}, \mathbf{D})$, is given by

$$L(\boldsymbol{\Theta}|\mathbf{Z}, \mathbf{D}) = \prod_{i=1}^{n} \prod_{k=1}^{\infty} (f_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)^{\Delta_i} \cdot R_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)^{1-\Delta_i})^{\mathbf{I}(z_i = k)}, \tag{10}$$

where $\mathbf{I}(\cdot)$ is an indicator function. To demonstrate that such augmentation will reduce complexity of the high dependency structure, the full

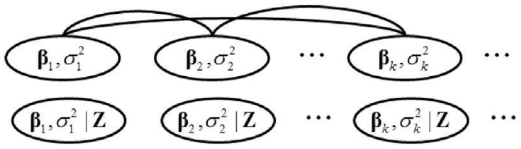


Fig. 1. Descriptive diagram of dependency structure and its simplification.

conditional posterior of $(\boldsymbol{\beta}_k, \sigma_k^2)$, conditioned on augmented $\mathbf{Z}$, can be expressed as

$$\pi(\boldsymbol{\beta}_k, \sigma_k^2 | \boldsymbol{\Theta}^{(-(\boldsymbol{\beta}_k, \sigma_k^2))}, \mathbf{Z}, \mathbf{D}) \propto \prod_{i=1}^{n} (f_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)^{\Delta_i} R_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)^{1-\Delta_i})^{\mathbf{I}(z_i=k)} \cdot \pi(\boldsymbol{\beta}_k, \sigma_k^2). \tag{11}$$

It can be observed from Eq. (11) that $(\boldsymbol{\beta}_k, \sigma_k^2)$ no longer depend on the remaining unknown parameters, i.e., $\boldsymbol{\Theta}^{(-(\boldsymbol{\beta}_k, \sigma_k^2))} = \emptyset$. Fig. 1 gives further illustration that augmented variables $\mathbf{Z}$ simplify the parameters' dependency structure. Nodes represent parameters and arcs represent existing dependency relationships. Conditioned on $\mathbf{Z}$, $(\boldsymbol{\beta}_k, \sigma_k^2)$ become conditional independent among each other with all arcs disconnected.

To further compute Eq. (11), the prior density, $\pi(\boldsymbol{\beta}_k, \sigma_k^2)$, needs to be explicitly specified. A popular prior specification is to assume multivariate normal prior for $\boldsymbol{\beta}_k$ and inverse gamma prior for σ_k^2, i.e., $\boldsymbol{\beta}_k \sim \text{MVN}_{p+1}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ and $\sigma_k^2 \sim \text{IG}(a_k, b_k)$. Thus, $\pi(\boldsymbol{\beta}_k, \sigma_k^2)$ can be explicitly written as

$$\pi(\boldsymbol{\beta}_k, \sigma_k^2) \propto |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp(-\tfrac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)) \cdot (\sigma_k^2)^{-a_k - 1} \exp\left(-\frac{b_k}{\sigma_k^2}\right). \tag{12}$$

The prior parameters in Eq. (12) can be elicited based on the available prior knowledge. For instance, $\boldsymbol{\mu}_k$ quantifies the average prior belief of $\boldsymbol{\beta}_k$ and diagonal values of $\boldsymbol{\Sigma}_k$ control the confidence of such prior belief. Larger values of $\boldsymbol{\Sigma}_k$ correspond to lower levels of confidence and vice versa. a_k and b_k can be also elicited based on the average prior belief and the prior variance using moment matching methods. When the prior knowledge is not available, less informative/non-informative priors can be specified. In this paper, less informative priors are considered.

When the available data is complete lifetime data, i.e., $\Delta_i = 1, \forall i = 1, \ldots, n$, the full conditional posterior distribution of $\boldsymbol{\beta}_k$ can be written as (see the Appendix A for proof)

$$\boldsymbol{\beta}_k | \boldsymbol{\Theta}^{(-\boldsymbol{\beta}_k)}, \mathbf{Z}, \mathbf{D} \sim \text{MVN}_{p+1}(\boldsymbol{\mu}_{k,\text{new}}, \boldsymbol{\Sigma}_{k,\text{new}}), \tag{13}$$

where $\boldsymbol{\Sigma}_{k,\text{new}} = (\frac{1}{\sigma_k^2}\mathbf{X}_k^{\mathrm{T}}\mathbf{X}_k + \boldsymbol{\Sigma}_k^{-1})^{-1}$ and $\boldsymbol{\mu}_{k,\text{new}} = \boldsymbol{\Sigma}_{k,\text{new}}(\frac{1}{\sigma_k^2}\mathbf{X}_k^{\mathrm{T}}\mathbf{Y}_k + \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k)$. $\mathbf{X}_k = [x_{ij}], i \in \mathbf{S}_k, j = 1, \ldots, (p+1)$ and $\mathbf{Y}_k = [y_{i0}], i \in \mathbf{S}_k$, where $y_{i0} = \log(t_i)$ and $\mathbf{S}_k$ is an index set defined as $\mathbf{S}_k = \{i : z_i = k, \forall i = 1, \ldots, n\}$. The full conditional posterior distribution of σ_k^2 can be written as (see the Appendix A for proof)

$$\sigma_k^2 | \Theta^{(-\sigma_k^2)}, \mathbf{Z}, \mathbf{D} \sim \text{IG}(a_{k,\text{new}}, b_{k,\text{new}}), \tag{14}$$

where $a_{k,\text{new}} = \frac{|\mathbf{S}_k|}{2} + a_k$, $b_{k,\text{new}} = \frac{1}{2}(\mathbf{Y}_k - \mathbf{X}_k\boldsymbol{\beta}_k)^{\mathrm{T}}(\mathbf{Y}_k - \mathbf{X}_k\boldsymbol{\beta}_k) + b_k$ and $|\cdot|$ is the cardinality measure of a set.

When the available data includes right-censored observations, i.e., $\exists i, \Delta_i = 0$, denote index sets $\mathbf{S}_{1k}$ and $\mathbf{S}_{0k}$ as $\mathbf{S}_{1k} = \{i : z_i = k, \Delta_i = 1, \forall i = 1, \ldots, n\}$ and $\mathbf{S}_{0k} = \{i : z_i = k, \Delta_i = 0, \forall i = 1, \ldots, n\}$, respectively. Augment latent variables ξ_i's, $i \in \mathbf{S}_{0k}$, to represent the true but unobserved lifetime observations for right-censored data. Based on such augmentation, $\boldsymbol{\beta}_k$ and σ_k^2 can be generated similarly using Eqs. (13) and (14) with different $\mathbf{Y}_k^* = [y_{i0}^*], i \in \mathbf{S}_k$, where $y_{i0}^* = \log(t_i), \forall i \in \mathbf{S}_{1k}$ and $y_{i0}^* = \log(\xi_i), \forall i \in \mathbf{S}_{0k}$. Each latent variable ξ_i can be further generated by a truncated lognormal distribution as

$$\xi_i | \mathbf{Z}, \mathbf{D} \sim \text{LN}_{(t_i, \infty)}(\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i, \sigma_k^2), \forall i \in \mathbf{S}_{0k}. \tag{15}$$

Based on the inverse transform sampling method (see the Appendix B for proof), ξ_i can be sampled as $\xi_i = \exp[\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i + \sigma_k \Phi^{-1}(q + (1-q)\Phi(\frac{t_i - \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i}{\sigma_k})]$, where $q \sim \text{Unif}(0, 1)$, $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are cumulative distribution function (cdf) and inverse cdf of standard normal distribution.

To address the second challenge of an infinite number of parameters, a slice sampling technique introduced by Walker [28] is employed. Compared to the distribution-based formulation for improving density estimation in [28], this paper considers a regression-based formulation to analyze heterogeneous complete and right-censored lifetime data for product reliability modeling. The regression-based formulation allows quantifying the effects of covariates, such as stress factors and environmental factors, on product lifetime. It also allows differentiating whether the lifetime heterogeneity can be captured by the observed covariates, the unobserved covariates, or both. The issue of an infinite number of parameters is caused by an infinite number of choices of k, i.e., $k = 1, \ldots, \infty$. For instance, the conditional posterior density of $z_i = k$ is given by

$$\pi(z_i = k | \boldsymbol{\Theta}, \mathbf{D}) \propto \prod_{k=1}^{\infty} f_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)^{\Delta_i} \cdot R_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)^{1-\Delta_i}. \tag{16}$$

To perform sampling for $z_i | \boldsymbol{\Theta}, \mathbf{D}$, there is an infinite number of possible k values. To address this issue, the slice sampling technique employed allows a finite number of k to be generated in actual implementation. The rationale is to further introduce augmented variables $\mathbf{U} = \{u_i\}_{i=1}^n$, where $u_i | z_i$, $\boldsymbol{w} \sim \text{Unif}(0, w_k)$ and $\text{Unif}(\cdot)$ denotes the uniform distribution. Recall $\Pr(z_i = k) = w_k$, then $\pi(u_i, z_i = k | \boldsymbol{w})$ is given by

$$\pi(u_i, z_i = k | \boldsymbol{w}) = \mathbf{I}(k \in A(u_i)), \tag{17}$$

where $A(u_i)$ is a set defined as $A(u_i) = \{k : w_k > u_i\}$. With augmented $\mathbf{U}$, the joint likelihood function, $L(\Theta | \mathbf{Z}, \mathbf{U}, \mathbf{D})$, is given by

$$L(\Theta | \mathbf{Z}, \mathbf{U}, \mathbf{D}) = \prod_{i=1}^{n} \prod_{k=1}^{\infty} [\mathbf{I}(k \in A(u_i)) \cdot f_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)^{\Delta_i} \cdot R_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)^{1-\Delta_i}]^{\mathbf{I}(z_i=j)}. \tag{18}$$

Based on (18), the full conditional posterior of $z_i = k$ can be expressed as

$$\pi(z_i = k | u_i, \Theta, \mathbf{D}) \propto \mathbf{I}(k \in A(u_i)) \cdot f_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)^{\Delta_i} \cdot R_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2)^{1-\Delta_i}. \tag{19}$$

As compared to Eqs. (16) and (19) shows that conditioned on u_i, z_i can only take possible values from the set $A(u_i)$. It can be shown that the cardinality of this set is always nonempty and finite, i.e., $0 < |A(u_i)| < \infty$ (see the Appendix C for proof). Thus, with help of augmented variable u_i, z_i only needs to be sampled from a finite number of possible values. Furthermore, the maximum number of k values required, denoted as K^*, is also finite since $K^* = |\cup_{i=1}^n A(u_i)|$. K^* can be further explicitly computed as [28]

$$K^* = \min\{K : \sum_{k=1}^{K} w_k > 1 - \min\{u_i\}_{i=1}^n\}. \tag{20}$$

Based on Eq (20), only a finite number $w_k | \mathbf{Z}$ are required to be generated during the model estimation. As shown in Eq. (5), $w_k | \mathbf{Z}$ can be obtained from $v_k | \mathbf{Z}$, where $v_k | \mathbf{Z}$ is given by Li et al. [33]

$$v_k | \mathbf{Z} \sim \text{beta}(1 + \sum_{i=1}^{n} \mathbf{I}(z_i = k), n + \gamma - \sum_{l=1}^{k} \sum_{i=1}^{n} \mathbf{I}(z_i = l)), \tag{21}$$

where beta$(\cdot)$ represents the beta distribution. Notice that sampling v_k in Eq. (21) is different from that in [28], where $v_k | v_1, \ldots, v_{k-1}, u_i$ is sampled sequentially from the truncated beta distribution. It is more straightforward to sample v_k's directly from the beta distribution in this paper.

In summary, the proposed estimation procedure can be described in the Appendix D. Notice that for the marginal sampling approach considered in [24–26], Dirichlet process will be marginalized out and the

Table 1
Simulation settings.

No.	Ground-truth values				Prior settings	
k	w_k	β_{0k}	β_{1k}	σ_k^2	$\pi(\beta_{0k}, \beta_{1k})$	$\pi(\sigma_k^2)$
1	0.4	9	−0.5	0.04	MVN_2 ($(0, 0)^T$, diag(100, 100))	IG(0.01,0.01)
2	0.6	12	−1	0.16		

Note: $\text{MVN}_p(\cdot)$: p-dimensional multivariate normal distribution; IG($\cdot$): inverse gamma distribution

numerical integration needs to be performed to calculate $\int f(t_i)\mathrm{d}P_0$ (if $\delta_i = 1$) or $\int R(t_i)\mathrm{d}P_0$ (if $\delta_i = 0$) for each observation $t_i, i = 1, \ldots, n$, where $f(\cdot)$ and $R(\cdot)$ are the probability density function and reliability function, respectively, and P_0 is the base distribution of the Dirichlet process. Based on the proposed estimation procedures described above, all samples can be drawn from the common distributions and no numerical integration is required. Most computer programs and packages have routines to generate random numbers from such common distributions, which significantly improve the convenience in actual implementation for practitioners.

3. Case studies

3.1. Numerical case study

To demonstrate the effectiveness of the proposed methodology, a mixture of two lognormal AFT model is assumed without losing generality. Simulation study is considered since ground-truth quantities, such as model parameters, the number of sub-populations and the reliability curve, can be pre-specified. They will serve as benchmark quantities to comprehensively and rigorously evaluate the performance of the proposed work and compare its performance with alternative methods. It is also noticed that in the numerical case study, both scenarios of the complete and right-censored lifetime data with covariates are illustrated and investigated. For simplicity, univariate covariate is considered to represent a single stress factor. Table 1 summarizes the predetermined ground-truth values for model parameters and describes the assumed prior settings. Less-informative priors are considered to mimic the scenario when prior knowledge is limited or even absent.

Based on ground-truth settings in Table 1, a complete lifetime dataset with a sample size of 100 is simulated. Univariate covariate x is generated from uniform distribution, e.g., $x \sim \text{Unif}(0, 3)$. The positive scalar in the Dirichlet process is set as $\gamma = 5$. Given the simulated complete lifetime observations $\{t_i, \delta_i = 1, x_i\}_{i=1}^{100}$, the proposed sampling algorithm is implemented according to Appendix D. In step 0, all observations are assigned to a single sub-population, i.e., $\{z_i = 1\}_{i=1}^{100}$ and $w_1 = v_1 \sim \text{Beta}(101, 5)$. In step 1, the augmented variable u_i for each observation is sampled as $u_i \sim \text{Unif}(0, w_1), i = 1, \ldots, 100$. In step 2, according to Eq. (20), the maximum number of sub-populations required, K^*, is calculated based on $\{u_i\}_{i=1}^{100}$. When $K^* > 1$, follow stick-breaking procedures to generate $v_k \sim \text{Beta}(1, 5)$, compute $w_k = \prod_{k'=1}^{k-1}(1 - v_{k'})v_k$ and generate $(\beta_{k,0}, \beta_{k,1})$ and σ_k^2 from $\pi(\beta_{k,0}, \beta_{k,1})$ and $\pi(\sigma_k^2)$, respectively, $\forall k > 1$. In step 3, update membership values z_i's by Eq. (19). Since lifetime data is complete, step 4a is carried out to update (β_{0k}, β_{1k}) and σ_k^2 using Eqs. (13) and (14). In step 5, based on the updated z_i's, (β_{0k}, β_{1k})'s and σ_k^2's, v_k's are updated using Eq. (21). The above steps 1–5 will be repeated until the maximum number of iteration is achieved.

Fig. 2 shows the trace-plot of the number of sub-populations generated against MCMC iterations and the corresponding histogram. Typically, posterior mode is selected as the most appropriate number of sub-populations [27,28]. In Fig. 2, it is observed that the posterior of m is highly concentrated on $m = 2$ and thus, 2 sub-populations are identified. The identified number of sub-populations is identical to the ground-truth number of sub-populations assumed in Table 1. To make inference of sub-population specific parameters, sampling iterations with $K = 2$ are first extracted. Since the posterior density is invariant to the permutation in the labeling of sub-population parameters, ordering constraint of $\beta_{(1),0} < \beta_{(2),0}$ is then considered to address the label-switching issue [34], where $\beta_{(k),0}$ represents the k^{th} smallest intercept coefficient among $\beta_{k,0}$'s. The ordering constraint of intercept coefficient is considered under the assumption that different sub-populations can be uniquely identified based on their corresponding $\beta_{(k),0}, k = 1, \ldots, K$. As $\beta_{k,0}$ represents the central location of lifetime distribution of sub-population k in the absence of influence of covariates, it is often true in practice that different sub-populations with different failure mechanisms, e.g., infant mortality failure and wear-out failure, will inherently have different expected lifetime spans and their $\beta_{k,0}$'s can be uniquely ordered. For cases when ordering based on $\beta_{(k),0}$ cannot yield unique identification, ordering constraints with multiple parameters or relabeling algorithms can be further considered [35]. Table 2 summarizes the estimation results of the corresponding model parameters.

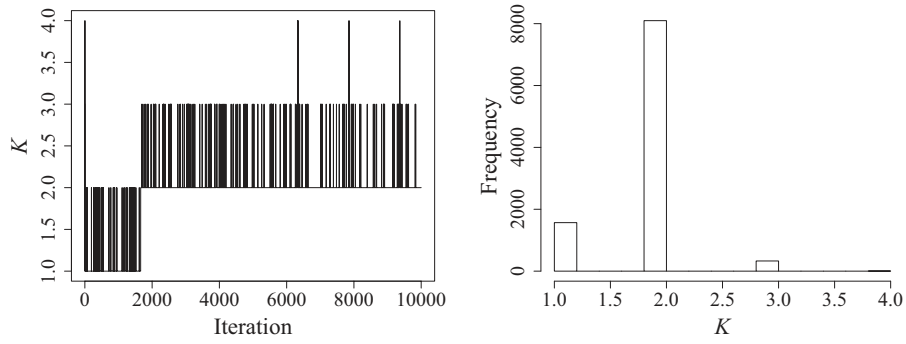
Table 2
Estimation results of complete lifetime data.

Parameter	Ground-truth value	MLE[a]	Posterior mean	Posterior median	99% Credible interval
w_1	0.4	0.48	0.45	0.45	[0.33,0.58]
$\beta_{1,0}$	9	9.03	9.04	9.03	[8.93,9.19]
$\beta_{1,1}$	−0.5	−0.55	−0.55	−0.55	[−0.66, −0.48]
σ_1^2	0.04	0.04	0.04	0.04	[0.02,0.07]
w_2	0.6	0.52	0.50	0.50	[0.38,0.62]
$\beta_{2,0}$	12	11.96	11.98	11.96	[11.72,12.38]
$\beta_{2,1}$	−1	−1.01	−1.02	−1.01	[−1.34, −0.83]
σ_2^2	0.16	0.15	0.16	0.16	[0.10,0.28]

[a] Maximum likelihood estimates assuming known number of sub-populations and known membership for each observation.

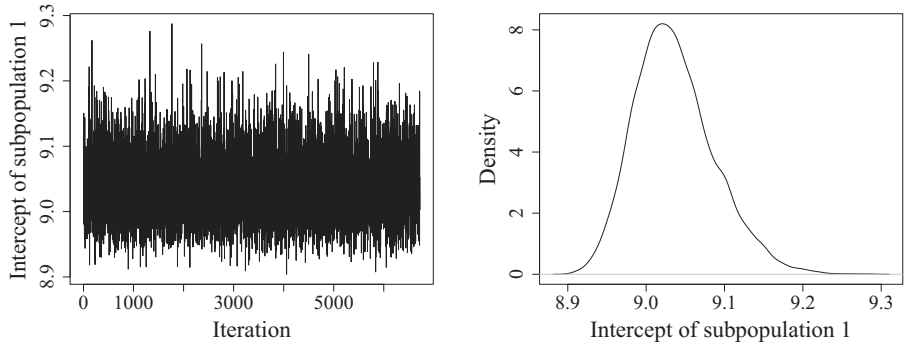
As shown in Table 2, Bayesian point estimates (e.g., posterior mean, posterior median) are close to the predefined ground-truth values. Furthermore, Bayesian interval estimates (e.g., 99% credible intervals) fully cover the ground-truth values. It is worth noting that Bayesian parameter estimation results shown above are obtained by assuming both an unknown number of sub-populations and an unknown sub-population membership for each observation. If assuming the true number of sub-populations and true sub-population membership for each observation, maximum likelihood estimation (MLE) is carried out. There is slight discrepancy between MLE estimation results and ground-truth values. It is because of the variability of the finite samples. Bayesian estimation results are also close to MLE estimation results since non-informative priors are considered. Figs. 3 and 4 show the posterior samples of $\beta_{1,0}$ and σ_2^2. Left figures show that the mixing of MCMC performs well. It indicates the convergence of MCMC method. Right figures show the corresponding posterior densities. Posterior samples of other estimated parameters exhibit similar results and thus are omitted here.

Another aspect is to investigate the influence of hyper-parameter, γ, in the DP process. Fig. 5 shows scenarios of decreasing and increasing γ to 1 and 10, respectively. It can be shown in Fig. 5a that when γ increases, it is more likely to explore a larger number of sub-populations and thus, requires less burn-in iterations to identify the correct number of sub-populations. It can be explained by the γ's influence on DP. When γ is large, DP is more likely to generate different values of $\boldsymbol{\beta}$ and σ^2. It allows the estimation method to explore a larger space of the possible number of sub-populations. On the contrary, when γ is small, DP



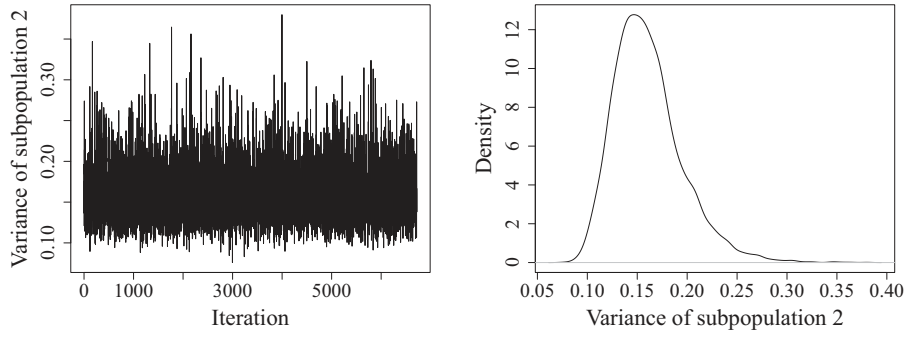
(a) Traceplot of K (b) Histogram of K

Fig. 2. Trace-plot and histogram of K ($\gamma = 5$).



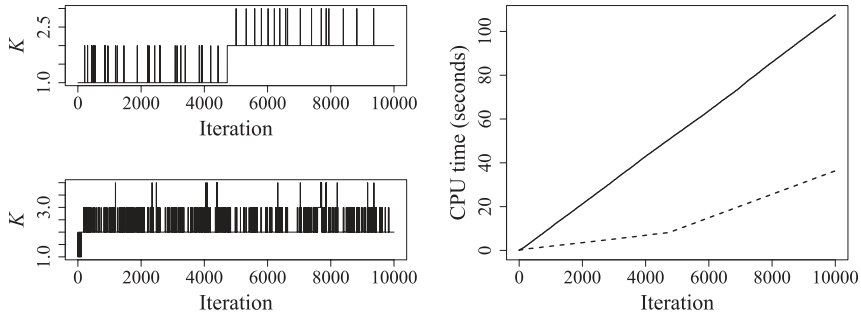
(a) Trace-plot of $\beta_{1,0}$ (b) Posterior density of $\beta_{1,0}$

Fig. 3. Estimation results of $\beta_{1,0}$ ($\gamma = 5$).



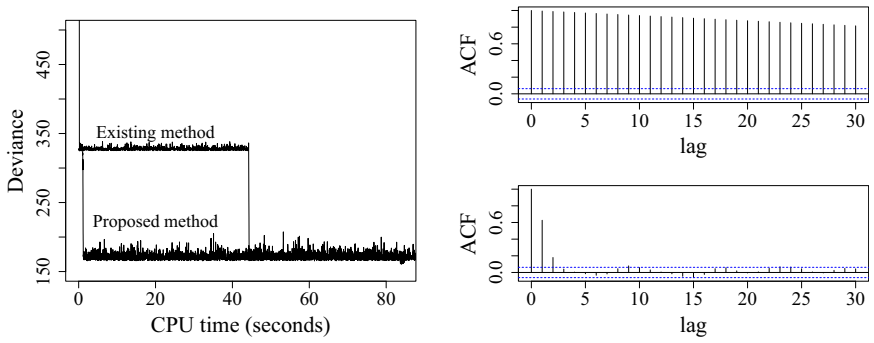
(a) Trace-plot of σ_2^2 (b) Posterior density of σ_2^2

Fig. 4. Estimation results of σ_2^2 ($\gamma = 5$).



(a) Comparison of trace-plots of K (b) Comparison of computation time

Fig. 5. Trace-plots and computation time comparisons at $\gamma = 1$ (top in Fig. 5a and dashed line in Fig. 5b) and $\gamma = 10$ (bottom in Fig. 5a and solid line in Fig. 5b).



(a) Comparison of convergence time (b) Comparison of ACF plots

Fig. 6. Computation efficiency and mixing quality comparisons between the existing method (top in Fig. 6b) and the propose method (bottom in Fig. 6b).

is more likely to generate the same values of $\boldsymbol{\beta}$ and σ^2. The estimation method will be less likely to explore a larger number of subpopulations and requires more burn-in iterations to identify the correct number of sub-populations. However, more iterations do not indicate more computational time. In fact, although larger γ requires less burn-in iterations, computational time per iteration is higher than that given by smaller γ. It is because at each iteration, the dimensionality of parameter space increases and sampling additional parameters will require additional computation time. As shown in Fig. 5b, sampling 10,000 iterations under settings of $\gamma = 1$ (dashed line) and $\gamma = 10$ (solid line) require about 36 and 107 s, respectively. The corresponding burn-in iterations 4700 and 300 require approximately 8 and 3 s, respectively. Therefore, although the number of burn-in iterations required has significant difference with the order of magnitude 1, the computation time has no significant difference in the order of magnitude and both settings can be considered in the actual implementation with the similar computation efficiency.

The proposed estimation method is based on the stick-breaking representation of DP. It will be also desirable to compare its performance with the estimation method based on the Polya urn representation in [22,24,25,36]. Fig. 6 shows the comparison results. In Fig. 6a, deviance measure is considered to represent the goodness-of-fit and monitor the convergence of both estimation algorithms [28,37]. When both methods converge, they exhibit similar goodness-of-fit. However, the proposed method converges after about 3 s while the existing method converges after about 45 s. The proposed work has significant computational efficiency and is faster than the existing work in the order of magnitude 1. All computation is carried out in a computer with 64 bit Intel dual-core processor @ 2.60 GHz and 16GB of RAM. Fig. 6b shows the comparison of mixing performance between two methods. A good sampling algorithm is expected to have a low auto-correlation function (ACF) value, which can further demonstrate the quality of convergence [38]. The proposed work generates less correlated samples and has much better mixing of Markov chains than the existing method. As shown in Fig. 6, both estimation methods exhibit similar modeling accuracy, but the existing method requires more computation time and has slower mixing performance than the proposed method. It is mainly because the existing method based on Polya urn representation has less efficient one-coordinate-at-a-time sampling strategy and often requires numerical integrals due to its nature of marginalization [23,27].

To evaluate the accuracy of predicted reliability functions, an independent test sample of additional 50 lifetime observations at $x = 0.5$ is simulated and the corresponding Kaplan–Meier curve is established. Fig. 7 compare the ground-truth reliability function, K-M curve and predicted reliability functions at $x = 0.5$ based on the proposed model considering the heterogeneous population as well as conventional models (e.g., lognormal AFT and Weibull AFT) assuming the homogeneous population. The proposed heterogeneous model has the satisfactory prediction accuracy of reliability function and is closer to both K-M curve and

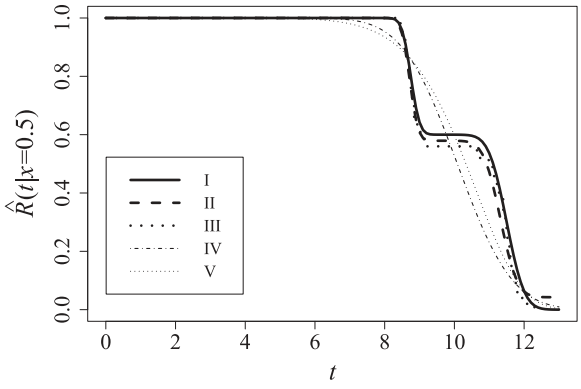


Fig. 7. Comparison of $\hat{R}(t|x = 0.5)$ among ground-truth (I), proposed model w/ considering heterogeneity (II), K-M (III), parametric lognormal AFT (IV) and parametric Weibull AFT (V).

ground-truth reliability function than the conventional homogeneous models.

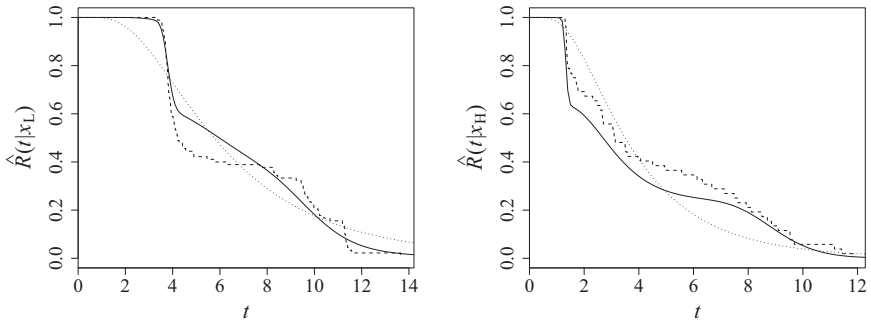
To further demonstrate the estimation results of heterogeneous lifetime data with right-censored observations, two scenarios of 5% and 30% right-censored data are created from the previous simulated 100 complete lifetime data observations. Lifetime observations which exceed the termination time are treated as right-censored observations. Similar estimation procedures in Appendix D are implemented except that in step 4, step 4a is replaced with step 4b. Specifically, for each right-censored observation i with $\delta_i = 0$, generate latent variable ξ_i according to Eq (15). Based on the updated ξ_i's, (β_{0k}, β_{1k}) and σ_k^2 are then updated using Eqs. (13) and (14). Table 3 summarizes the estimation results. When the proportion of right-censored observations is small (e.g., 5%), the estimation results are similar to the complete data estimation results in Table 2. When the proportion of right-censored observations is moderate (e.g., 30%), the estimation results of sub-population 2 become inaccurate but those of sub-population 1 are still similar to the previous estimation results. It is because sub-population 2 consists of lifetime observations with larger values. As the right-censored proportion increases, more complete data observations from sub-population 2 become right-censored observations and there is less data available for accurate estimation of sub-population 2.

3.2. Real case study

To further demonstrate the applicability of the proposed methodology in real practice, a real case study of assembly data collected from an automatic robotic assembly process is provided. Assembly time for each assembly product unit is recorded and can serve as an important

Table 3
Estimation results of right-censored lifetime data.

Parameter	5% right-censored observations			30% right-censored observations		
	Posterior mean	Posterior median	99% Credible interval	Posterior mean	Posterior median	99% Credible interval
w_1	0.45	0.45	[0.30,0.57]	0.40	0.40	[0.27,0.54]
$\beta_{1,0}$	9.04	9.03	[8.93,9.20]	9.03	9.03	[8.92,9.17]
$\beta_{1,1}$	−0.56	−0.55	[−0.69, −0.47]	−0.55	−0.55	[−0.67,−0.45]
σ_1^2	0.04	0.04	[0.02,0.07]	0.03	0.03	[0.02,0.06]
w_2	0.51	0.50	[0.38,0.65]	0.55	0.55	[0.42,0.68]
$\beta_{2,0}$	12.17	12.10	[11.66,13.69]	14.71	14.57	[12.63,18.21]
$\beta_{2,1}$	−1.18	−1.12	[−2.24,−0.83]	−2.48	−2.42	[−4.41,−1.33]
σ_2^2	0.43	0.19	[0.11,5.76]	2.01	1.46	[0.30,9.76]



(a) Comparisons at $x = x_L$ (b) Comparisons at $x = x_H$

Fig. 8. Comparisons of $\hat{R}_{KM}(t)$, $\hat{R}_{Homo}(t)$ and $\hat{R}_{Hete}(t)$. $\hat{R}_{KM}(t)$: the benchmark reliability function based on K-M (dashed curve); $\hat{R}_{Homo}(t)$: the estimated reliability function assuming the homogeneous assumption (dot curve); $\hat{R}_{Hete}(t)$: the estimated reliability function of the proposed model assuming the heterogeneous assumption (solid curve).

productivity measure for the assembly process. Such data are essentially lifetime data and "lifetime" can be interpreted as the "time duration" of a product unit in the assembly process. Due to the product quality variability (e.g., material quality, dimension variation), assembly time data exhibit heterogeneity. There is also a controllable robotic variable, denoted as x, may affect the assembly time. It consists of two settings, namely low-level and high-level robotics settings, denoted as x_L and x_H. To take into account the heterogeneity and quantify the influence of robotic settings on assembly time, the proposed model is employed to analyze such assembly time data. The same less-informative prior settings in Table 1 are assumed since no prior knowledge is available. Fig. 8 shows the comparison results of estimated reliability curves, $R(t|x)$, at a specific robotic setting, x. $R(t|x)$ can be interpreted as the probability of not having finished the assembly task till time t and a lower value indicates a better productivity.

Unlike the simulation study, there is no ground-truth reliability function in real data. As shown in Fig. 7 of the simulation study, K-M curve can serve as a good surrogate of the ground-truth reliability function. Since K-M curve is calculated from lifetime data rather than lifetime data with covariates, real data is first stratified into two subsets with covariate values x_L and x_H, respectively. Within each subset data, the corresponding K-M curve is then calculated and denoted as $\hat{R}_{KM}(t|x_L)$ or $\hat{R}_{KM}(t|x_H)$. Notice that K-M curves (i.e., dashed curves in Fig. 8) are step functions with discontinuities at observed time observations t_i's.

To compare the performance between the proposed model with considering heterogeneity and the convectional model without considering heterogeneity at x_L (or x_H), their corresponding estimated reliability functions, denoted as $\hat{R}_{Hete}(t|x_L)$ (or $\hat{R}_{Hete}(t|x_H)$) and $\hat{R}_{Homo}(t|x_L)$ (or $\hat{R}_{Homo}(t|x_H)$), will be compared with $\hat{R}_{KM}(t|x_L)$ (or $\hat{R}_{KM}(t|x_H)$). A model whose estimated reliability function is closer to K-M curve has better performance. Let $d(\hat{R}, \hat{R}_{KM}|x) = \sum_i |\hat{R}(t_i|x) - \hat{R}_{KM}(t_i|x)|$ quantify the distance between the estimated reliability function and K-M curve at x. When $x = x_L$, $d(\hat{R}_{Homo}, \hat{R}_{KM}|x_L) = 8.23$ and $d(\hat{R}_{Hete}, \hat{R}_{KM}|x_L) = 4.72$; when $x = x_H$, $d(\hat{R}_{Homo}, \hat{R}_{KM}|x_H) = 4.61$ and $d(\hat{R}_{Hete}, \hat{R}_{KM}|x_H) = 3.56$. Thus, the proposed model gives smaller distance values and shows better performance. Fig. 8 also graphically confirms such results. At both $x = x_L$ and $x = x_H$, the estimated reliability functions based on the proposed method considering heterogeneity (i.e., solid curves) are closer to K-M curves and better capture their trends than those neglecting such heterogeneity (i.e., dot curves). It can be also seen that high-level robotic setting gives a lower $\hat{R}(t|x)$ value at fixed time t, indicating a better productivity. It is also noticed that the estimated reliability functions based on the proposed model or the conventional model are smooth and continuous functions of t.

4. Conclusion

A Bayesian nonparametric model for heterogeneous lifetime data modeling and quantification is proposed in this paper. It releases the assumption of per-sepecifying the number of sub-populations and quantify the effects of possible covariates on product lifetime. Specifically, the Dirichlet process is embedded into lognormal AFT model to account for lifetime latent heterogeneity and incorporate possible covariates. This novel formulation allows the number of sub-populations to be identified objectively from data. Moreover, such identification is realized jointly with estimating model parameters for all sub-populations in a single step. The resulting model estimation challenges are further addressed and resolved by employing a series of sampling techniques into the MCMC. In particular, when lifetime data is complete, this paper analytically shows that conjugate priors are available for model estimation and it provides great convenience in practical implementation for practitioners.

In this paper, normal-disturbed error term is assumed for modeling lifetime data of each sub-population. Other error terms structures can be further investigated in the future, such as the extreme value distribution and the logistic distribution. The mixing proportion of the proposed model is assumed to be independent of covariates. For future works, the proposed sampling approach can be extended to estimate models with

the Dirichlet process variants, such as the dependent Dirichlet process [36,39], where covariates could have both effects on the mixing proportion and the model parameters of each sub-population.

Acknowledgments

This work was supported in part by National Science Foundation under Grant BCS-1638301, in part by University of South Florida Research & Innovation Internal Awards Program under Grant No. 0114783, and in part by National Natural Science Foundation of China under Grants Nos. 71402157 and 71672163.

Appendix A. Proofs of Eqs. (13) and (14)

When $\Delta_i = 1, \forall i$, $\pi(\boldsymbol{\beta}_k, \sigma_k^2)$ is specified in (12) and let $\mathbf{S}_k = \{i : z_i = k, \forall i = 1, \ldots, n\}$, the full conditional posteriors for $\boldsymbol{\beta}_k$, $\pi(\boldsymbol{\beta}_k|\Theta^{(-(\boldsymbol{\beta}_k))}, \mathbf{Z}, \mathbf{D})$, is given by

$$\begin{aligned}\pi(\boldsymbol{\beta}_k|\Theta^{(-(\boldsymbol{\beta}_k))}, \mathbf{Z}, \mathbf{D}) &\propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)\right) \\ &\cdot \prod_{i\in\mathbf{S}_k} f_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)\right. \\ &\left. -\frac{\sum_{i\in\mathbf{S}_k}(\log(t_i) - \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x})^2}{2\sigma_k^2}\right) \end{aligned} \tag{22}$$

Denote $\mathbf{X}_k = [x_{ij}], i \in \mathbf{S}_k, j = 1, \ldots, (p+1)$ and $\mathbf{Y}_k = [y_{i0}], i \in \mathbf{S}_k$, where $y_{i0} = \log(t_i)$, $\sum_{i\in\mathbf{S}_k}(\log(t_i) - \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x})^2$in (22) can be written as $(\mathbf{Y}_k - \mathbf{X}_k\boldsymbol{\beta}_k)^{\mathrm{T}}(\mathbf{Y}_k - \mathbf{X}_k\boldsymbol{\beta}_k)$. Thus, (22) can be further simplified as

$$\begin{aligned}\pi(\boldsymbol{\beta}_k|\Theta^{(-(\boldsymbol{\beta}_k))}, \mathbf{Z}, \mathbf{D}) &\propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k) + \frac{1}{2\sigma_k^2}(\mathbf{Y}_k - \mathbf{X}_k\boldsymbol{\beta}_k)^{\mathrm{T}}\right. \\ &\left.(\mathbf{Y}_k - \mathbf{X}_k\boldsymbol{\beta}_k)\right) \propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}_k^{\mathrm{T}}(\tfrac{1}{\sigma_k^2}\mathbf{X}_k^{\mathrm{T}}\mathbf{X}_k + \boldsymbol{\Sigma}_k^{-1})\boldsymbol{\beta}_k\right. \\ &\left.+(\frac{1}{\sigma_k^2}\mathbf{X}_k^{\mathrm{T}}\mathbf{Y}_k + \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k)^{\mathrm{T}}\boldsymbol{\beta}_k\right) \\ &\propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_{k,\mathrm{new}})^{\mathrm{T}}\boldsymbol{\Sigma}_{k,\mathrm{new}}^{-1}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_{k,\mathrm{new}})\right) \end{aligned} \tag{23}$$

where $\boldsymbol{\Sigma}_{k,\mathrm{new}} = (\frac{1}{\sigma_k^2}\mathbf{X}_k^{\mathrm{T}}\mathbf{X}_k + \boldsymbol{\Sigma}_k^{-1})^{-1}$ and $\boldsymbol{\mu}_{k,\mathrm{new}} = \boldsymbol{\Sigma}_{k,\mathrm{new}}(\frac{1}{\sigma_k^2}\mathbf{X}_k^{\mathrm{T}}\mathbf{Y}_k + \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k)$. Therefore, based on (23), $\boldsymbol{\beta}_k|\Theta^{(-(\boldsymbol{\beta}_k))}, \mathbf{Z}, \mathbf{D} \sim \mathrm{MVN}_{p+1}(\boldsymbol{\mu}_{k,\mathrm{new}}, \boldsymbol{\Sigma}_{k,\mathrm{new}})$.

Similarly, the full conditional posteriors for σ_k^2, $\pi(\sigma_k^2|\Theta^{(-(\sigma_k^2))}, \mathbf{Z}, \mathbf{D})$, is given by

$$\begin{aligned}\pi(\sigma_k^2|\Theta^{(-(\sigma_k^2))}, \mathbf{Z}, \mathbf{D}) &\propto \left(\sigma_k^2\right)^{-a_k-1}\exp(-\frac{b_k}{\sigma_k^2}) \cdot \prod_{i\in\mathbf{S}_k} f_k(t_i|\mathbf{x}, \boldsymbol{\beta}_k, \sigma_k^2) \\ &\propto (\sigma_k^2)^{-a_k-1-\frac{|\mathbf{S}_k|}{2}}\exp\left(-\frac{b_k}{\sigma_k^2} - \frac{\sum_{i\in\mathbf{S}_k}(\log(t_i) - \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x})^2}{2\sigma_k^2}\right) \\ &\propto (\sigma_k^2)^{-a_{k,\mathrm{new}}-1}\exp\left(-\frac{b_{k,\mathrm{new}}}{\sigma_k^2}\right) \end{aligned} \tag{24}$$

where $a_{k,\mathrm{new}} = \frac{|\mathbf{S}_k|}{2} + a_k$, $b_{k,\mathrm{new}} = \frac{1}{2}(\mathbf{Y}_k - \mathbf{X}_k\boldsymbol{\beta}_k)^{\mathrm{T}}(\mathbf{Y}_k - \mathbf{X}_k\boldsymbol{\beta}_k) + b_k$. Therefore, based on (24), $\sigma_k^2|\Theta^{(-(\sigma_k^2))}, \mathbf{Z}, \mathbf{D} \sim \mathrm{IG}(a_{k,\mathrm{new}}, b_{k,\mathrm{new}})$.

Appendix B. Sampling of ξ_i

Given a random variable $q \sim \mathrm{Unif}(0, 1)$, sampling ξ_i is equivalently to solve $q = \int_{t_i}^{\xi_i} g(s|\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i, \sigma_k^2)\mathrm{d}s / \int_{t_i}^{\infty} g(s|\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i, \sigma_k^2)\mathrm{d}s$, where $g(\cdot|\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i, \sigma_k^2)$ is the probability density function of normal distribution with mean $\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i$ and variance σ_k^2. Thus, based on normalization, there is $q = [\Phi(\frac{\xi_i - \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i}{\sigma_k}) - \Phi(\frac{t_i - \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i}{\sigma_k})]/[1 - \Phi(\frac{t_i - \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i}{\sigma_k})]$. It can be further simplified as $\xi_i = \exp[\boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i + \sigma_k\Phi^{-1}(q + (1-q)\Phi(\frac{t_i - \boldsymbol{\beta}_k^{\mathrm{T}}\mathbf{x}_i}{\sigma_k})]$

Appendix C. Proof of $0 < |A(u_i)| < \infty$

Recall that the augmented variable u_i is generated through $u_i|z_i, \boldsymbol{w} \sim \mathrm{Unif}(0, w_k)$, there is $u_i < w_k$. Since $A(u_i)$ is defined as $A(u_i) = \{k' : w_{k'} > u_i\}$, $k \in A(u_i)$ or equivalently $|A(u_i)| \geq 1 > 0$. Suppose $|A(u_i)| \to \infty$, $\sum_{k'\in A(u_i)} w_{k'} > |A(u_i)|u_i \to \infty$. However, $\sum_{k'\in A(u_i)} w_{k'} \leq \sum_{k=1}^{\infty} w_{k'} = 1$. Based on such contradiction, $|A(u_i)| < \infty$. Thus, $A(u_i)$ is a finite and nonempty set, i.e., $0 < |A(u_i)| < \infty$.

Appendix D. Summary of the sampling algorithm

Step 0: Initialize $K^{(0)} = 1$, and $z_i^{(0)} = 1, \forall i = 1, \ldots, n$. Generate $(\boldsymbol{\beta}_k^{(0)}, \sigma_k^{2(0)})$ from $\pi(\boldsymbol{\beta}_k, \sigma_k^2)$ in Eq. (12) and let $w_1^{(0)} = v_1^{(0)} \sim \mathrm{Beta}(n+1, \gamma)$. For iteration $\tau = 1, \ldots, \tau_{\max}$, repeat **Steps 1–5**;

Step 1: $u_i^{(\tau)}|z_i^{(\tau-1)} = k, \boldsymbol{w}^{(\tau-1)} \sim \mathrm{Unif}(0, w_{z_i^{(\tau-1)}}^{(\tau-1)})$;

Step 2: Determine K^* by Eq. (20). If $K^* > K^{(\tau-1)}$, set $K^{(\tau)} = K^*$; otherwise, $K^{(\tau)} = K^{(\tau-1)}$. Generate $v_k \sim \mathrm{Beta}(1, \gamma)$, $(\boldsymbol{\beta}_k, \sigma_k^2) \sim \pi(\boldsymbol{\beta}_k, \sigma_k^2)$ and compute $w_k = \prod_{k'=1}^{k-1}(1 - v_{k'})v_k, \forall k > K^{(\tau-1)}$;

Step 3: Update $z_i^{(\tau)} = k|u_i^{(\tau)}, \boldsymbol{\Theta}, \mathbf{D}$ by Eq. (19);

Step 4: Update $(\boldsymbol{\beta}_k^{(\tau)}, \sigma_k^{2(\tau)})|\mathbf{D}, \mathbf{Z}$ based on the following two cases:

Step 4-a: If lifetime data is complete without right-censored observations, update $(\boldsymbol{\beta}_k^{(\tau)}, \sigma_k^{2(\tau)})|\mathbf{D}, \mathbf{Z}$ by Eqs. (13) and (14);

Step 4-b: If lifetime data is right-censored, generate latent variables ξ's by Eq (15) and then update $(\boldsymbol{\beta}_k^{(\tau)}, \sigma_k^{2(\tau)})|\mathbf{D}, \mathbf{Z}$ by Eqs. (13) and (14) with ξ's;

Step 5: Update $v_k^{(\tau)}|\mathbf{Z}$ by Eq. (21) and compute $w_1^{(\tau)} = v_1^{(\tau)}$, $w_k^{(\tau)} = \prod_{k'=1}^{k-1}(1 - v_{k'}^{(\tau)})v_k^{(\tau)}, \forall k > 2$.

References

[1] Yuan T, Zhu X. Reliability study of ultra-thin dielectric films with variable thickness levels. IIE Trans 2012;44(9):744–53.
[2] Wu H, Meeker WQ. Early detection of reliability problems using information from warranty databases. Technometrics 2002;44(2):120–33.
[3] Xiang Y, Coit DW, Feng Q. n subpopulations experiencing stochastic degradation: reliability modeling, burn-in, and preventive replacement optimization. IIE Trans 2013;45(4):391–408.
[4] DiGiovanni AA, Sullivan EC. Apparatus and methods for detecting performance data in an earth-boring drilling tool. Technical Report. Houston, TX: Baker Hughes Incorporated; 2014.
[5] Majeske KD. A mixture model for automobile warranty data. Reliab Eng Syst Saf 2003;81(1):71–7.
[6] Attardi L, Guida M, Pulcini G. A mixed-weibull regression model for the analysis of automotive warranty data. Reliab Eng Syst Saf 2005;87(2):265–73.
[7] Castet J-F, Saleh JH. Single versus mixture weibull distributions for nonparametric satellite reliability. Reliab Eng Syst Saf 2010;95(3):295–300.
[8] Yuan T, Ji Y. A hierarchical bayesian degradation model for heterogeneous data. IEEE Trans Reliab 2015;64(1):63–70.
[9] Kontar R, Son J, Zhou S, Sankavaram C, Zhang Y, Du X. Remaining useful life prediction based on the mixed effects model with mixture prior distribution. IISE Trans 2017;49(7):682–97.
[10] Murthy DNP, Xie M, Jiang R. Weibull models. John Wiley & Sons; 2004.
[11] Li M, Liu J. Bayesian hazard modeling based on lifetime data with latent heterogeneity. Reliab Eng Syst Saf 2016;145:183–9.
[12] Zhang Z, Si X, Hu C, Kong X. Degradation modeling–based remaining useful life estimation: a review on approaches for systems with heterogeneity. Proc Inst Mech Eng Part O 2015;229(4):343–55.
[13] Jiang R, Murthy DNP. Modeling failure-data by mixture of 2 weibull distributions: a graphical approach. Reliab IEEE Trans 1995;44(3):477–88.
[14] Jiang R, Murthy DNP. Impact of quality variations on product reliability. Reliab Eng Syst Saf 2009;94(2):490–6.
[15] Hartigan J.A.. A failure of likelihood asymptotics for normal mixtures1985; 2:807–810.
[16] Adler RJ. An introduction to continuity, extrema, and related topics for general gaussian processes. Lect Notes-Monogr Ser 1990:i155.
[17] Akaike H. Information theory and an extension of the maximum likelihood principle. In: Selected Papers of Hirotugu Akaike. Springer; 1998. p. 199–213.
[18] Koehler AB, Murphree ES. A comparison of the akaike and schwarz criteria for selecting model order. Appl Stat 1988:187–95.
[19] Tsionas EG. Bayesian analysis of finite mixtures of weibull distributions. Commun Stat Theory Methods 2002;31(1):37–48.
[20] Wiper M, InInsua DR, Ruggeri F. Mixtures of gamma distributions with applications. J Comput Graph Stat 2001;10(3).
[21] Warr RL, Collins DH. Bayesian nonparametric models for combining heterogeneous reliability data. Proc Inst Mech Eng Part O 2014;228(2):166–75.

[22] Escobar MD, West M. Bayesian density estimation and inference using mixtures. J Am Stat Assoc 1995;90(430):577–88.
[23] Ishwaran H, James LF. Approximate dirichlet process computing in finite normal mixtures. J Comput Graphical Stat 2002;11(3).
[24] Kottas A. Nonparametric bayesian survival analysis using mixtures of weibull distributions. J Stat Plann Inference 2006;136(3):578–96.
[25] Cheng N, Yuan T. Nonparametric bayesian lifetime data analysis using dirichlet process lognormal mixture model. Naval Res Logist (NRL) 2013;60(3):208–21.
[26] Yuan T, Liu X, Ramadan SZ, Kuo Y. Bayesian analysis for accelerated life tests using a dirichlet process weibull mixture model. Reliab IEEE Trans 2014;63(1):58–67.
[27] Papaspiliopoulos O, Roberts GO. Retrospective markov chain monte carlo methods for dirichlet process hierarchical models. Biometrika 2008;95(1):169–86.
[28] Walker SG. Sampling the dirichlet mixture model with slices. Commun Stat 2007;36(1):45–54.
[29] Tobias PA, Trindade D. Applied reliability. CRC Press; 2011.
[30] Sethuraman J. A constructive definition of dirichlet priors. DTIC Document; 1991.
[31] Gilks WR. Markov chain monte carlo. Wiley Online Library; 2005.
[32] Stroud AH. Approximate calculation of multiple integrals. Prentice Hall; 1971.
[33] Li M, Han J, Liu J. Bayesian nonparametric modeling of heterogeneous time-to-event data with an unknown number of sub-populations. IISE Trans 2017;49(5):481–92.
[34] Richardson S, Green PJ. On bayesian analysis of mixtures with an unknown number of components (with discussion). J R Stat Soc 1997;59(4):731–92.
[35] Jasra A, Holmes CC, Stephens DA. Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. Stat Sci 2005:50–67.
[36] De Iorio M, Johnson WO, Müller P, Rosner GL. Bayesian nonparametric nonproportional hazards survival modeling. Biometrics 2009;65(3):762–71.
[37] Green PJ, Richardson S. Modelling heterogeneity with and without the dirichlet process. Scand J Stat 2001;28(2):355–75.
[38] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis, 2. FL, USA: Chapman & Hall/CRC Boca Raton; 2014.
[39] MacEachern S. Dependent dirichlet processes. Technical report. Department of Statistics,The Ohio State University; 2000.