STATISTICS IN TRANSITION new series, December 2025
Vol. 26, No. 4, pp. xx–XX, https://doi.org/10.59139/stattrans-2025-xxx
Received – xx.xx.xxxx; accepted – xx.xx.xxxx

Bayesian nonparametric model for weighted data using mixture of Burr XII distributions

Soleiman Khazaei[1], **Soghra Bohlouri Hajjar**[2]

Abstract

In this paper, we develop a Bayesian nonparametric approach for analyzing weighted survival data. Specifically, we employ the Dirichlet Process Burr XII Mixture Model (DPBMM) to estimate the underlying density and survival functions when the observed data are weighted. Parameters are inferred using Markov chain Monte Carlo (MCMC) methods, and the Metropolis-Hastings algorithm is applied to obtain de-biased samples from the weighted observations. Numerical illustrations are provided using both simulated and real lifetime data, including the presence of censored observations. The performance of the proposed method is compared with classical kernel density estimates to demonstrate its flexibility in modeling complex and heavy-tailed distributions.

Key words: Bayesian nonparametric, weighted data, Dirichlet process, mixture model, Burr XII distribution, survival data.

1. Introduction

Building upon the foundational ideas introduced by Fisher (1934), the concept of weighted distributions has been further developed. Rao (1985) and Rao et al. (1915) recognized the importance of a unified framework and identified a variety of sampling scenarios that could be effectively described using weighted distributions Patil (1978). Consider a non-negative random variable X with a natural density function $f(x;\theta)$, where $\theta \in \Theta$ denotes the natural parameter and Θ is the parameter space. A new random variable is then defined with a density function $g(x;\theta)$, specified as follows:

$$g(x;\theta) = \frac{w(x;\theta) f(x;\theta)}{E[w(X;\theta)]}, \quad E[w(X;\theta)] < \infty, \quad x \geq 0, \tag{1}$$

This new variable is referred to as a weighted random variable with respect to X, and $g(x;\theta)$ is called the weighted density function corresponding to $f(x;\theta)$. The function $w(x;\theta)$ is a non-negative function of x, and $E[w(X;\theta)]$ denotes its mathematical expectation under the distribution of X.

If $w(x;\theta) = x$, the resulting weighted distribution is known as the length-biased distribution. For instance, studies involving family size as a sampling factor often produce length-biased samples. In Zelen and Feinleib (1969), this distribution is applied to the early

[1]Department of Statistics, Razi University, Kermanshah, Iran. E-mail: s.khazaei@razi.ac.ir.
ORCID: https://orcid.org/0000-0003-2537-9232.
[2]Department of Statistics, Razi University, Kermanshah, Iran. E-mail: bohlurihajjar.soghra@razi.ac.ir.
ORCID: https://orcid.org/0000-0001-9803-5823.

detection of breast cancer. Similarly, Patil and Rao (1977) used the length-biased distribution to study human family structures and wildlife populations. Later, Patil and Rao (1978) introduced a broader class of distributions of the form given in Equation 1, incorporating arbitrary non-negative weight functions $w(x;\theta)$, and provided several practical examples. For additional examples of weighted distributions and their applications, see Blumenthal (1967), Gupta and Kirmani (1990), Mahafoud and Patil (1982), Patil and Rao (1977), Patil and Rao (1978). This paper adopts a Bayesian nonparametric framework to model weighted data derived from such distributions. Specifically, we use the Dirichlet Process Mixture Model (DPMM), a popular Bayesian nonparametric approach, and apply it to survival analysis.

Burr (1942) introduced a family of distributions, from which twelve distinct types (named Burr distributions Type I to XII) can be derived as special cases. Among them, the Burr Type XII (Burr XII) distribution is widely used in survival studies.

Let $\kappa_B(t|c,k)$ and $\mathscr{K}_B(t|c,k)$ denote the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of the Burr XII distribution, respectively. These functions, which will be employed in our mixture model, are defined as:

$$\kappa_B(t|c,k) = ck\frac{t^{c-1}}{(1+t^c)^{k+1}}, \quad c,k>0, \quad t>0 \tag{2}$$

$$\mathscr{K}_B(t|c,k) = 1-(1+t^c)^{-k} \tag{3}$$

with the parameter space

$$\Theta = \{(c,k); 0<c<\infty, 0<k<\infty\}.$$

In Hatjispyros et al. (2017) , the Dirichlet Process Mixture Model (DPMM) is employed for density estimation under length-biased data. The authors use a log-normal distribution as the kernel, with a fixed distribution assigned to its shape parameter.

In contrast, we consider a DPMM with the Burr Type XII (Burr XII) distribution as the kernel, which includes two shape parameters treated as random variables in the model. Despite the increasing adoption of nonparametric methods in data analysis, many existing approaches struggle to handle weighted data effectively. Key limitations include inflexibility in capturing complex distributional shapes, poor modeling of heavy-tailed behavior, and a lack of adaptability to hidden heterogeneity. Traditional parametric and classical nonparametric models often fail to accurately represent the underlying structure of such data, particularly when distributions exhibit skewness or heavy tails. The Burr XII distribution is highly flexible, making it well-suited for modeling diverse distributional forms, especially those with heavy tails. However, its integration into mixture models, particularly within a Bayesian nonparametric framework, has received limited attention. This research addresses this methodological gap by introducing a Bayesian nonparametric model based on a Dirichlet Process Mixture of Burr XII distributions. The proposed framework offers enhanced flexibility and robustness for analyzing weighted data, enabling more accurate characterization of the complex, heterogeneous structures frequently encountered in real-world applications.

The Burr XII distribution has support on $\mathbb{R}^+$ and serves as a generalization of both the log-normal and Weibull distributions. These characteristics make it particularly suitable for survival analysis (Bohlouri Hajjar and Khazaei (2018), Lanjoni et al. (2016), Rao et al. (2015) and Rodriguez (1977)).

In Bohlouri Hajjar and Khazaei (2018), the Burr XII distribution is used as the kernel in a DPMM framework, where the survival function and hazard rate are computed for both simulated and real-world datasets. In Joudaki et al. (2024), a Dirichlet Process Mixture Model (DPMM) with a three-parameter Burr XII kernel was considered. Their study investigates survival analysis using this flexible modeling approach, demonstrating its effectiveness in capturing complex features of survival data. Bayesian estimation methods for hybrid censored data from the Burr XII distribution using various loss functions were discussed in Hassan (2021). These methods were particularly valuable for managing complex censoring scenarios. In Nurul et al. (2024), a DPMM-based approach was proposed for clustering mixed-type data with cluster-specific covariance matrices, effectively addressing intricate data structures. Moreover, Michael et al. (2023) applied DPMMs to longitudinal data involving repeated attempts, successfully modeling the complexities inherent in such datasets.

In the next section, we present the preliminary concepts and methodology, followed by a detailed introduction of the proposed model. Section 4 describes the use of Gibbs sampling to estimate the original (unweighted) distributions from their corresponding weighted forms. Section 5 demonstrates the application of our approach to both simulated and real datasets. Finally, Section 6 summarizes our findings and conclusions.

2. Preliminary and Methodology

We aim to estimate the density and survival functions by considering a general case of the weight function $w(x;\theta)$. To avoid computing the often intractable normalizing constant, our strategy is to model $g(x;\theta)$ directly and then infer $f(x;\theta)$, using the fact that $g(x;\theta) \propto w(x;\theta)f(x;\theta)$.

If we assume that $f(x;\theta)$ belongs to a parametric family, then both $f(x;\theta)$ and $g(x;\theta)$ are known up to the normalizing constant, which may not be analytically tractable.

Let $w(\cdot;\theta)$ be a general weight function; an essential condition for modelling $F(\cdot;\theta)$ through $G(\cdot;\theta)$ $\big(F(\cdot;\theta)$ and $G(\cdot;\theta)$ denote the distribution functions corresponding to $f(\cdot;\theta)$ and $g(\cdot;\theta)$, respectively$\big)$ is

$$\int_0^\infty w(x)^{-1} g(x)dx < \infty, \tag{4}$$

because f is a distribution function.

Through the invertibility implied by Equation (4), it becomes possible to reconstruct the distribution function F from G. In the Bayesian nonparametric framework, we place a suitable nonparametric prior on g, relying on the relationship defined by Equation (4).

The key question, however, is how the posterior structure derived from modeling g directly can be transformed into the corresponding posterior structure for f.

The first step involves developing a method to convert a weighted sampler into an unweighted sample. Once this conversion is achieved, inference about the posterior distributions can be made.

The Markov Chain Monte Carlo (MCMC) approach is an indirect method for simulating samples from complex probability distributions. One of the key MCMC methods is the Metropolis-Hastings algorithm Hatjispyros et al. (2017), which generates samples from a target distribution by utilizing its full joint density function along with proposal distributions for each of the variables of interest.

Algorithm 1: Metropolis-Hastings algorithm

1. **Initialize with** $x^{(0)} \sim q(x)$

2. **for** $i = 1, 2, \ldots$ **do**

Propose $x^{cand} \sim q(x^{(i)}|x^{(i-1)})$

Calculate the acceptance probability:

$$\alpha(x^{cand}|x^{(i-1)}) = \min\left\{1, \frac{q(x^{(i-1)}|x^{cand})\pi(x^{cand})}{q(x^{cand}|x^{(i-1)})\pi(x^{(i-1)})}\right\}$$

Generate $u \sim \text{Uniform}(0,1)$

if $u < \alpha$ **then**

$x^{(i)} \leftarrow x^{cand}$

else

$x^{(i)} \leftarrow x^{(i-1)}$

end if

end for.

Here, $q(x)$ represents the weighted distribution. Hatjispyros et al. (2017) demonstrated how the Metropolis-Hastings algorithm can be used to convert a length-biased sample into an unbiased one. Following a similar strategy, we aim to apply this algorithm using a general weight function, as defined in Equation (4), to transform samples from a weighted distribution into their unweighted counterparts. This methodology is particularly important when dealing with complex models where the weighted distribution arises due to inherent sampling bias.

Suppose that $y_1, y_2, \ldots, y_N$ denote a random sample from g. The Metropolis-Hastings algorithm is used to convert this sample into a sample from $f(x;\theta) \propto w(x;\theta)^{-1} g(x;\theta)$. In the algorithm, we assume that $g(\cdot)$ is replaced by $q(\cdot)$ in Algorithm 1 with the acceptance probability

$$\min\left\{1, \frac{w^{-1}(y_{j+1})}{w^{-1}(x_j)}\right\}.$$

If x_j denotes the current sample from $f(x)$, then

$$\begin{aligned} x_{j+1} &= y_{j+1} \quad \text{with probability} \quad \min\left\{1, \frac{w^{-1}(y_{j+1})}{w^{-1}(x_j)}\right\}, \\ x_{j+1} &= x_j \quad \text{otherwise.} \end{aligned} \tag{5}$$

The transition density is

$$P(x_{j+1}|x_j) = \min\left\{1, \frac{w^{-1}(y_{j+1})}{w^{-1}(x_j)}\right\} g(x_{j+1}) + \{1 - r(x_j)\} 1(x_{j+1} = x_j),$$

where

$$r(x) = \int \min\left\{1, \frac{w^{-1}(x^*)}{w^{-1}(x)}\right\} g(x^*) dx^*.$$

We can outline the general methodology as follows:

1. *Sample Generation*: Consider $(y_1, \ldots, y_n)$ as a sample from g, to which we assign a suitable nonparametric prior.

2. *Posterior Inference*: Using MCMC methods, posterior samples from the random measure $\Pi(dg|y_1, \ldots, y_n)$ and other relevant parameters are obtained. Consequently, a sequence $\{y_{n+1}^l\}, l = 1, 2, \ldots$ from the posterior predictive density $g(y|y_1, \ldots, y_n)$ will be generated.

3. *Weighted Proposal Values*: The sequence $\{y_{n+1}^l\}$ serves as proposal values in a Metropolis-Hastings chain whose stationary distribution is the weighted posterior predictive, i.e.,
$$\{y_{n+1}^l\} \propto w(y)^{-1} g(y|y_1, \ldots, y_n).$$
Using Equation (5), we generate the corresponding values $\{x_{n+1}^l\}$ at iteration l.

4. *Final Sample*: The resulting sequence $\{x_{n+1}^l\}$ constitutes a sample from the posterior predictive distribution f, which corresponds to the unweighted density.

3. The model and inference

In this section, we aim to model $g(x; \theta)$. Modeling the weighted distribution $g(x)$ within the Bayesian nonparametric framework is based on an infinite mixture model Lo (1984), which takes the following form:

$$g_P(y) = \int \kappa(y; \theta) P(d\theta), \tag{6}$$

where P is a discrete probability measure and $\kappa(y; \theta)$ is a kernel density defined on $(0, \infty)$ for all θ in the parameter space. This kernel satisfies the condition

$$\int_0^\infty w^{-1}(y; \theta) \kappa(y; \theta) dy < \infty.$$

By choosing the Burr(XII) density (with parameters c and k) as the kernel of the mixture model, we obtain

$$g_{c,k,P}(y) = \int_{\mathbb{R}} \kappa_B(y|c,k)P(dc,dk),$$

where $\kappa_B(y|c,k)$ is the Burr(XII) density and P is a discrete random probability measure. Suppose that

$$P \sim DP(\upsilon, P_0),$$

where $DP(\upsilon, P_0)$ denotes the Dirichlet process with precision parameter $\upsilon > 0$ and base measure P_0 Ferguson (1983). We refer to this mixture model as the Dirichlet Process Burr(XII) Mixture Model (DPBMM).

The hierarchical representation of the DPBMM can be expressed as follows:

$$\begin{aligned} y|c,k &\sim \kappa_B(y|c,k), \\ (c,k)|P &\sim P, \\ P|\upsilon, P_0 &\sim DP(\upsilon, P_0). \end{aligned} \tag{7}$$

Suppose that the base distribution P_0 is the prior distribution for the joint distribution of c and k. By choosing Burr(XII) distribution as the kernel, P_0 that yields a closed-form expression for $\int \kappa_B(.|c,k)P_0(dc,dk)$ is not available. Moreover, we choose multiple distributions of Uniform$(0,\phi)$ and Exponential with the parameter γ for P_0, i.e.,

$$P_0(c,k|\phi,\gamma) = \text{Unif}(c|0,\phi) \times \text{Exp}(k|\gamma). \tag{8}$$

This choice achieves the modeling goals. Considering the hyperparameters γ and ϕ as random, we choose prior distributions $Pareto(a_\phi, b_\phi)$ and $IGamma(a_\gamma, b_\gamma)$ for them, respectively. We set $a_\phi = a_\gamma = d$ and chose $d = 2$, which makes the variance of the Pareto distribution infinite. This allows the distribution to accommodate a wide range of values. The parameters b_ϕ and b_γ are determined by the data Bohlouri Hajjar and Khazaei (2018).

Finally, for any t_i, $i = 1,...,n$, representing lifetime data in a sample of n observations, by considering DPBMM and selecting priors for parameters of the model we have

$$\begin{aligned} t_i|c_i,k_i &\sim \kappa_B(t_i|c_i,k_i), \quad i = 1,...,n, \\ (c_i,k_i)|P &\sim P, \\ P &\sim DP(\nu, P_0), \\ P_0|\gamma,\phi &\sim \text{Unif}(c|0,\phi) \times \text{Exp}(k|\gamma), \\ \nu,\gamma,\phi &\sim \text{Gamma}(a_\nu, b_\nu) \times IGamma(a_\gamma, b_\gamma) \times Pareto(a_\phi, b_\phi). \end{aligned} \tag{9}$$

After determining the model, we want to formulate how to sample from DPMMs by Gibbs sampling. According to Kottas (2006), Gibbs sampling for drawing a sample from

$$[(\theta_1,\ldots,\theta_n), \upsilon, \ldots \,|\, t]$$

is based on the following full conditional distributions (brackets are used to indicate conditional and marginal distributions):

$$\begin{aligned}
&(1) \quad [(\theta_i)|(\theta_{-i}, z_{-i}), \upsilon, ..., t], \quad \text{for } i = 1, ..., n \\
&(2) \quad [(\theta_j^*)|z, n^*, \upsilon, ..., t], \quad \text{for } j = 1, ..., n^* \\
&(3) \quad [\upsilon|\{(\theta_j^*), j = 1, ..., n^*\}, n^*, t], [...|\{(\theta_j^*), j = 1, ..., n^*\}, n^*, t].
\end{aligned} \tag{10}$$

Here, t is the vector of failure time data. The θ_i's are the parameters of the kernel in DPMMs that will be analyzed.

Model (8) and the discreteness property of the Dirichlet process exhibit clustering in the θ's. We present n^* as the number of clusters among the θ_i's, denoted by θ_j^*'s. The vector of indicators $z = (z_1, ..., z_n)$ indicates the clustering configuration such that $z_i = j$ when $\theta_i = \theta_j^*$.

Also, θ_{-i}, which is used in (8), is defined as $\theta_{-i} = (\theta_1, \theta_2, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_n)$. Model (12) is the same as model (8), with the difference that the vector of indicators $z = (z_1, ..., z_n)$ indicates the clustering configuration such that $z_i = j$ when $\theta_i = \theta_j^*$. Also, θ_{-i}, which is used in (8), is defined as $\theta_{-i} = (\theta_1, \theta_2, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_n)$.

4. Modeling

We apply the following algorithm to model the unweighted density $f(x)$ from the weighted density $g(x)$. First, to generate a sample from $g(x)$, the model's parameters need to be estimated. To this aim, we draw a sample from (c_i, k_i) and update z_i for each t_i.

In simulation-based parameter estimation, we use the Gibbs sampler, which includes two steps to reach the goal.

Algorithm 2 : Gibbs sampler

1. Initialize with $\theta^{(0)} \sim f(\theta)$
2. For $i = 1, 2, ...$ do

$$\begin{aligned}
&\theta_1^{(i)} \sim f(\theta_1|\theta_2^{(i-1)}, \theta_3^{(i-1)}, ..., \theta_d^{(i-1)}, D), \\
&\vdots \\
&\theta_d^{(i)} \sim f(\theta_d|\theta_1^{(i)}, \theta_2^{(i)}, ..., \theta_{d-1}^{(i)}, D),
\end{aligned}$$

where $\theta_1, ..., \theta_d$ are model parameters, and D is the vector of observations. The values at iteration i are sampled from the conditional distributions using the most recent values of the other parameters.

Now, the model will be applied to lifetime data with right-censored observations, which are very common in survival studies. To calculate the related distributions, the data are divided into uncensored and censored observations.

1- Uncensored data: For uncensored data (t_{io}), the conditional posterior density of (c_i, k_i) is a mixture distribution Neal (2003):

$$f(c_i,k_i \mid \{(c_{i'},k_{i'}); i \neq i'\}, \nu, \gamma, \phi, t_{io}) = \frac{q_0^o h^o(c_i,k_i \mid \phi,\gamma,t_{io}) + \sum_{j=1}^{n^{*(i)}} n_j^{*(i)} q_j^o \delta_{c_j^*,k_j^*}}{q_0^o + \sum_{j=1}^{n^{*(i)}} n_j^{*(i)} q_j^o},$$

where $q_j^o = k_B(t_{io} \mid c_j^*, k_j^*)$, and

$$\begin{aligned} q_0^o &= \nu \int_0^{\phi} \int_0^{\infty} k_B(t_{io} \mid c,k) G_0(c,k)\, dc\, dk \\ &= \frac{\nu}{\phi} \int_0^{\phi} \frac{c t_{io}^{c-1}}{(1+t_{io}^c)} \left(\int_0^{\infty} \frac{k e^{-k/\gamma}}{(1+t_{io}^c)^k}\, dk \right) dc \\ &= \frac{\nu}{\phi} \int_0^{\phi} \frac{c t_{io}^{c-1}}{(1+t_{io}^c)\left(\ln(1+t_{io}^c) + \frac{1}{\gamma}\right)}\, dc, \end{aligned}$$

where the last integration can be computed numerically, and

$$h^o(c_i,k_i \mid \gamma,\phi,t_{io}) \propto k_B(t_{io} \mid c_i,k_i) P_0(c_i,k_i \mid \gamma,\phi) \propto [c_i \mid \gamma,\phi,t_{io}][k_i \mid c_i,\gamma,\phi,t_{io}],$$

with

$$[c_i \mid \gamma,\phi,t_{io}] \propto c_i t_{io}^{c_i-1} I_{(0,\phi)}(c_i), \quad i = 1,\ldots,n,$$

and

$$[k_i \mid c_i,\gamma,\phi,t_{io}] \propto \text{Gamma}\left(\cdot \mid 2, \frac{1}{\frac{1}{\gamma} + \ln(1+t_{io}^{c_i})} \right).$$

2- Right censored data: For right-censored data (t_{ic}), the conditional posterior density of (c_i, k_i) is

$$f(c_i,k_i \mid \{(c_i,k_i); i \neq i'\}, \nu, \gamma, \phi, t_{ic}) = \frac{q_0^c h^c(c_i,k_i \mid \phi,\gamma,t_{ic}) + \sum_{j=1}^{n^{*(i)}} n_j^{*(i)} q_j^c \delta_{c_j^*,k_j^*}}{q_0^c + \sum_{j=1}^{n^{*(i)}} n_j^{*(i)} q_j^c},$$

where $q_j^c = 1 - K_B(t_{ic} \mid c_j^*, k_j^*)$, and

$$\begin{aligned} q_0^c &= \nu \int_0^{\phi} \int_0^{\infty} \big(1 - K_B(t_{ic} \mid c,k)\big) G_0(c,k)\, dc\, dk \\ &= \frac{\nu}{\phi\gamma} \int_0^{\phi} \int_0^{\infty} \frac{e^{-k/\gamma}}{(1+t_{ic}^c)^k}\, dk\, dc \\ &= \frac{\nu}{\phi\gamma} \int_0^{\phi} \left(\frac{1}{\gamma} + \ln(1+t_{ic}^c) \right) dc, \end{aligned}$$

where the last integration can be computed numerically. Using the property of censored

data, we have

$$\begin{aligned} h^c(c_i,k_i \mid \gamma,\phi,t_{ic}) &\propto \big(1-K_B(t_{ic} \mid c_i,k_i)\big)G_0(c_i,k_i) \\ &\propto [c_i \mid \gamma,\phi,t_{ic}][k_i \mid c_i,\gamma,\phi,t_{ic}] \\ &= \frac{I_{(0,\phi)}(c_i)}{\phi\gamma}\frac{1}{\frac{1}{\gamma}+\ln(1+t_{ic}^{c_i})}k_i \exp\left\{-k_i\Big(\frac{1}{\frac{1}{\gamma}+\ln(1+t_{ic}^{c_i})}\Big)\right\} \\ &= \frac{I_{(0,\phi)}(c_i)}{\phi\gamma}\frac{1}{\frac{1}{\gamma}+\ln(1+t_{ic}^{c_i})} \times \text{Gamma}\Big(k_i \mid 2, \frac{1}{\frac{1}{\gamma}+\ln(1+t_{ic}^{c_i})}\Big). \end{aligned}$$

We use the slice sampling method to sample from the first part of the above expression. Therefore, using this MCMC approach, a sample from $h^c(c_i,k_i \mid \phi,\gamma,t_{ic})$ can be obtained. Now, for both observed and censored data, (c_i,k_i) for $i=1,\ldots,n$ can be updated iteratively and improved. In a general form, (c_j^*,k_j^*) can be updated conditional on ϕ,γ, and t as follows:

$$\begin{aligned} f(c_j^*,k_j^* \mid \phi,\gamma,t,n^*) &\propto G_0(c_j^*,k_j^* \mid \gamma,\phi)\prod_{\{io:s_{io}=j\}}k_B(t_{io} \mid c_j^*,k_j^*)\prod_{\{ic:s_{ic}=j\}}\big(1-K_B(t_{ic} \mid c_j^*,k_j^*)\big) \\ &\propto [c_j^* \mid \gamma,\phi,t_{io}][k_j^* \mid c_j^*,\gamma,\phi,t_{ic}]\prod_{\{ic:s_{ic}=j\}}\frac{1}{(1+t_{ic}^{c_j^*})^{k_j^*}} \\ &\propto c_j^{*n_j^o}I_{(0,\phi)}(c_j^*)\prod_{\{io:s_{io}=j\}}\frac{t_{io}^{c_j^*-1}}{1+t_{io}^{c_j^*}} \times \text{Gamma}(n_j^o+1,B^*), \end{aligned} \tag{11}$$

where

$$B^* = \sum_{\{io:s_{io}=j\}}\left(\frac{1}{\gamma}+\ln(1+t_{io}^{c_j^*})\right)+\sum_{\{ic:s_{ic}=j\}}\ln(1+t_{ic}^{c_j^*}),$$

and n_j^o is the number of observed data points in cluster j.

The key task in generating a sample from Equation (13) is drawing from the first part of the equation. Sampling from the gamma distribution is straightforward. To sample from

$$[c_j^* \mid \phi,\gamma,t] \propto c_j^{*n_j^o}I_{(0,\phi)}(c_j^*)\prod_{\{io:s_{io}=j\}}\frac{t_{io}^{c_j^*-1}}{1+t_{io}^{c_j^*}},$$

auxiliary variables $W=\{w_{io}:\{io:s_{io}=j\}\}$ are introduced such that

$$[c_j^*,W \mid \phi,t_{io}] = c_j^{*n_j^o}I_{(0,\phi)}(c_j^*)\prod_{\{io:s_{io}=j\}}I_{(0,\frac{t_{io}^{c_j^*-1}}{1+t_{io}^{c_j^*}})}(w_{io}).$$

By marginalization over W, $[c_j^* \mid \phi,t_{io}]$ is obtained for $j=1,\ldots,n^*$. Moreover, w_{io} are

uniform variables on $(0, \frac{t_{io}^{c_j^*-1}}{1+t_{io}^{c_j^*}})$. Therefore,

$$[c_j^* \mid \phi, t] = c_j^{*n_j^o} I_{(B,\phi)}(c_j^*),$$

where $B = \max\{0, \frac{\ln(w_{io})}{1+t_{io}}\}$. Drawing from $[c_j^* \mid \phi, t]$ is now straightforward.

Subsequently, following the approach in Escobar and West (1995), ϕ, γ, and ν are updated. Introducing a latent variable u such that

$$[u \mid \nu, t] = \text{Beta}(\nu + 1, n),$$

we have

$$[\nu \mid u, n^*, t] = p\,\text{Gamma}(a_\nu + n^*, b_\nu - \ln(u)) + (1-p)\,\text{Gamma}(a_\nu + n^* - 1, b_\nu - \ln(u)),$$

where

$$p = \frac{a_\nu + n^* - 1}{n(b_\nu - \ln(u)) + a_\nu + n^* - 1}.$$

To update ϕ, we have

$$[\phi \mid c^*, k^*] = [\phi][c^*, k^* \mid \phi] = \frac{2b_\phi^2}{\phi^3} I_{(b_\phi,\infty)}(\phi) \prod_{j=1}^{n^*} \frac{1}{\phi} I_{(0,\phi)}(c_j^*) = \frac{2b_\phi^2}{\phi^{n^*+3}} I_{(b^*,\infty)}(\phi),$$

where $b^* = \max\{b_\phi, \max_{1 \le j \le n^*} c_j^*\}$, implying

$$[\phi \mid c^*, k^*] = \text{Pareto}(\phi \mid 2 + n^*, b^*).$$

Repeating this technique, γ is updated as

$$[\gamma \mid c^*, k^*] = [\gamma] \prod_{j=1}^{n^*} [k_j^* \mid \gamma] = \text{IGamma}(n^* + 2, b_\gamma + \sum_{j=1}^{n^*} k_j^*).$$

Thus, all conditional distributions required for Equation (8) can now be computed.

5. Data illustrations

We evaluate the performance of the proposed model by applying it to three simulated datasets, each with a sample size of $n = 200$. For model comparison, we consider alternative approaches from the literature. Gibbs sampling is implemented with a total of $N = 15,000$ iterations, discarding the initial 1,000 as burn-in, and applying thinning by retaining every 20th iteration. Assuming a sufficiently large sample size, we place a $Gamma(1, 0.001)$ prior on the concentration parameter α, allowing the model to infer an appropriate number of clusters based on the data.

To quantify the deviation between estimated and true models, we compute numerical indices using both Euclidean and Hellinger distance metrics that have been widely used in

similar studies:

$$d_E(f,\hat{f}) = \sqrt{\sum_{i=1}^{n} \left(f(t_i) - \hat{f}(t_i)\right)^2} \tag{12}$$

$$d_H(f,\hat{f}) = \sqrt{\sum_{i=1}^{n} \left(\sqrt{f(t_i)} - \sqrt{\hat{f}(t_i)}\right)^2}. \tag{13}$$

Here, f denotes the true function, which may be a density, survival, or hazard function, while $\hat{f}$ represents its estimate obtained from the MCMC output.

We generate samples of size $n = 200$ from the following distributions:

(i) Mixture of Burr distributions with 2 parameters (B2M):

$$0.4 \times \text{Burr}(c = 25, k = 10) + 0.6 \times \text{Burr}(c = 7, k = 4).$$

(ii) Mixture of lognormal distributions (LNM):

$$0.2 \times \text{LN}(\mu_1 = 0, \sigma_1^2 = 0.15) + 0.3 \times \text{LN}(\mu_2 = 1, \sigma_2^2 = 0.02) + 0.5$$

$$\times \text{LN}(\mu_3 = 2, \sigma_3^2 = 0.04).$$

(iii) Mixture of Weibull distributions (WM):

$$0.4 \times \text{Weibull}(\alpha = 1, \lambda = 0.25) + 0.6 \times \text{Weibull}(\alpha = 6, \lambda = 0.5).$$

Here, $\text{Burr}(c,k)$ denotes the Burr Type XII distribution with shape parameter c and scale parameter k; $\text{LN}(\mu,\sigma^2)$ denotes the lognormal distribution with scale parameter μ and shape parameter σ; and $\text{Weibull}(\alpha,\lambda)$ refers to the Weibull distribution with shape parameter α and scale parameter λ.

Table 1 presents the computed index values for the density, survival, and hazard functions under the Dirichlet Process Mixture Model (DPMM) with different kernel choices, evaluated across the three simulated datasets. Details of the DPWM, DPLNM, and DPB2M models are provided in references Hassan et al. (2021) and Cheng and Yuan (2013). The results in Table 1 indicate that the DPB2M model achieves a better fit compared to the other models based on the simulated datasets.

Table 1. Deviance metrics d_E ($d_H(\cdot,\hat{\cdot})$) for DPMMs with different kernels. Values in parentheses are Hellinger distances.

Dataset	Metric	DPWM	DPLNM	DPB2M
I	$f(t)$	1.553 (1.126)	1.019 (0.522)	6.307 (4.015)
	$S(t)$	0.602 (0.467)	0.295 (0.213)	3.276 (1.658)
	$h(t)$	23.097 (4.031)	10.160 (1.473)	33.559 (8.834)
II	$f(t)$	1.448 (1.224)	0.373 (0.309)	1.901 (1.709)
	$S(t)$	0.650 (0.471)	0.195 (0.144)	1.692 (1.088)
	$h(t)$	5.542 (2.489)	1.788 (0.668)	7.407 (3.832)
III	$f(t)$	2.218 (0.669)	4.762 (1.184)	2.607 (0.767)
	$S(t)$	0.399 (0.213)	0.321 (0.198)	0.368 (0.206)
	$h(t)$	8.281 (1.053)	13.017 (1.695)	13.809 (1.558)

Note that the values in parentheses represent Hellinger distances.

5.1. Simulated data

In this subsection, we illustrate the capability of the DPB2M model to effectively model weighted data and recover the corresponding unweighted distribution. For this purpose, we consider two types of datasets: simulated data and real data.

Assume we are given a sample $(x_1,\ldots,x_n)$, and our objective is to estimate its density function. To evaluate the goodness of fit of the proposed model, we compare the resulting density estimate with the following two density estimators, which are used in Hatjispyros et al. (2017):

i) **The classical kernel density estimate:**

$$\tilde{g}_h(x;(x_1,\ldots,x_n)) \propto \frac{1}{n}\sum_{j=1}^{n} N(x \mid x_j,h^2)\,1_{(0,+\infty)}(x)$$

ii) **The kernel density estimate for indirect data (Jones' kernel density estimate):**

$$\hat{f}_{J,h}(x;(x_1,\ldots,x_n)) \propto \frac{1}{n}\hat{\mu}\sum_{j=1}^{n} x_j^{-1} N(x \mid x_j,h^2)\,1_{(0,+\infty)}(x)$$

where $\hat{\mu}$ is the harmonic mean of the sample $(x_1,\ldots,x_n)$.

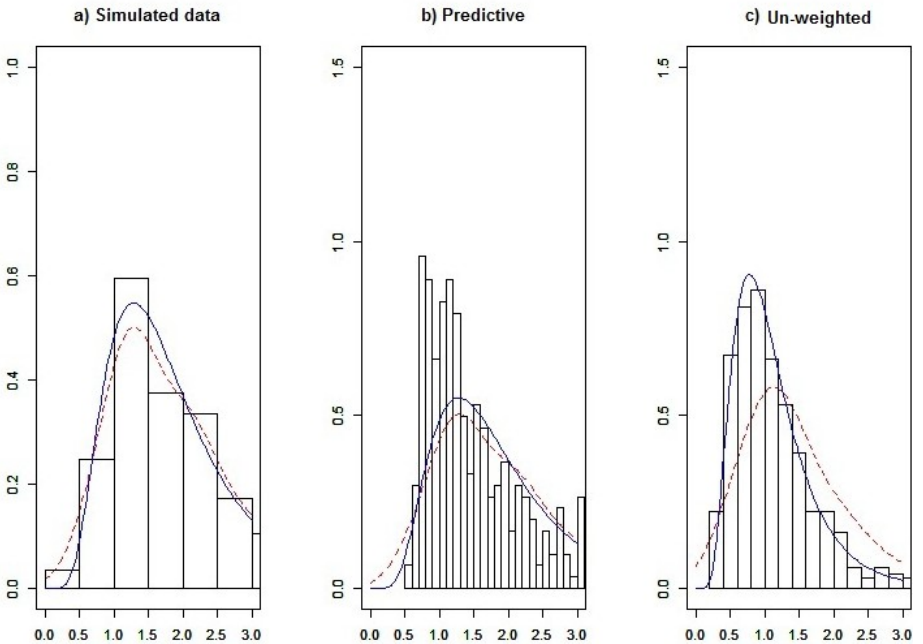

Figure 1. Simulated data from the log-normal distribution with parameters $(0.5, 0.5)$ and a sample size of $n = 100$. In each figure, the true densities are shown with a solid line, and the kernel density estimates $\tilde{g}_h$ (a),(b) and $\hat{f}_{J,h}$ (c) with a dashed line.

These estimators are among the best-known nonparametric methods and provide a good fit for both the weighted and unweighted data in our analysis. For the simulated datasets, the Metropolis-Hastings algorithm was run for over 50,000 iterations, while the Gibbs sampler was executed for 60,000 iterations, with a burn-in period of 10,000 iterations.

5.1.1 Length biased distribution of log-normal

Here, the first dataset is simulated from the log-normal distribution with parameters $(\mu, \sigma^2) = (0.5,\ 0.5)$. We know that the length-biased distribution of a log-normal with parameters $\mu + \sigma^2$ and σ^2 is again a log-normal with parameters μ and σ^2 Patil and Rao (1978).

By choosing the log-normal distribution as the kernel, we can illustrate the model's preference and the algorithm. This model was tested in Hatjispyros et al. (2017) for length-biased data using simulated data from the Gamma distribution with DPMM when the Gamma distribution was considered the kernel.

The results of the simulated data are shown in Figure 1. In panel (a), the histogram of simulated log-normal $(0.5, 0.5)$ data is presented, and the true density curve is depicted with a solid line, while the kernel density estimate $\tilde{g}_h$ is shown with a dashed line. The estimate $\tilde{g}_h$ closely approximates the true underlying density.

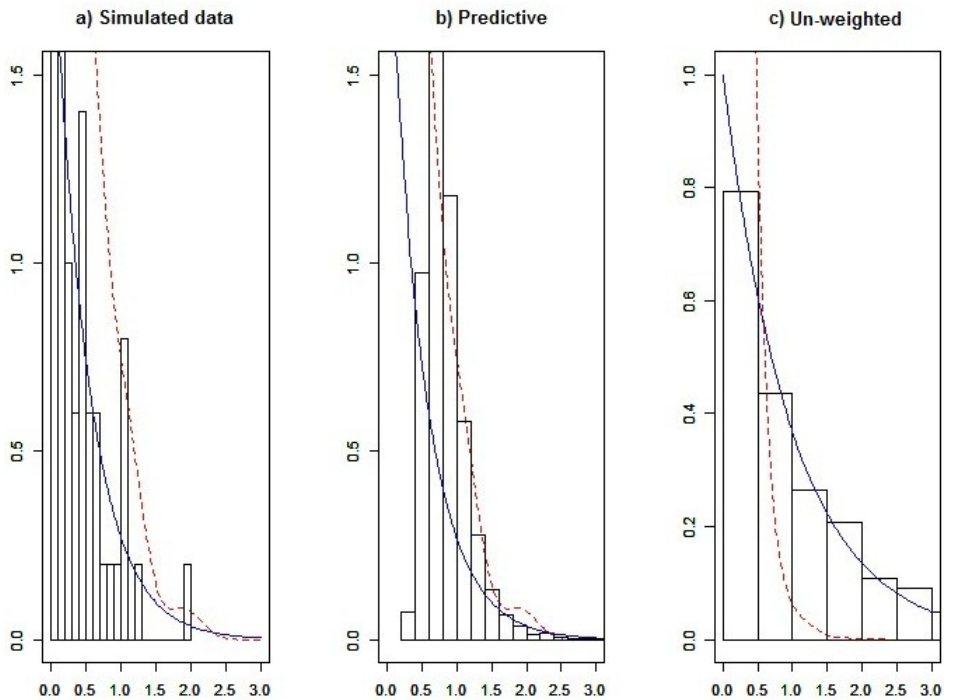

Figure 2. Simulated data from the Weibull distribution with parameters $(\alpha = 1, \lambda = 2)$ and sample size $n = 100$. True densities are shown with a solid line, and the kernel density estimates $\tilde{g}_h$ and $\hat{f}_{J,h}$ with a dashed line.

In panel (b), the histogram of the posterior predictive distribution of the data is shown along with the true density curve (solid line). The real data here follows a log-normal distribution with parameters $(0, 0.5)$, and Jones' density estimate $\hat{f}_{J,h}$ is represented by a dashed line.

Panel (c) depicts the histogram of the transformed data to the unweighted scale, corresponding to the indirect data estimate, which is shown with a dashed line. The distribution of the unweighted data is also close to the true distribution, a log-normal with parameters $(0, 0.5)$, demonstrating the model's ability to recover the underlying density effectively.

5.1.2 Weighted distribution of Gamma

Here, we consider a $Gamma(\alpha, \beta)$ distribution with the weight function $w(x|a,b) = x^a \exp(-x/b)$. By applying the weighting function, we obtain a new distribution that is again a Gamma distribution, but with updated parameters $(\alpha + a, (\beta + b)/(b\beta))$.

The dataset is simulated from a weighted Gamma distribution, specifically $Gamma(1,2)$ using the weight function $w(x) = \exp(-x)$, which corresponds to $a = 0$ and $b = 1$. The corresponding unweighted version of this distribution is a Gamma distribution with parameters $\alpha = 1$ and $\beta = 1$.

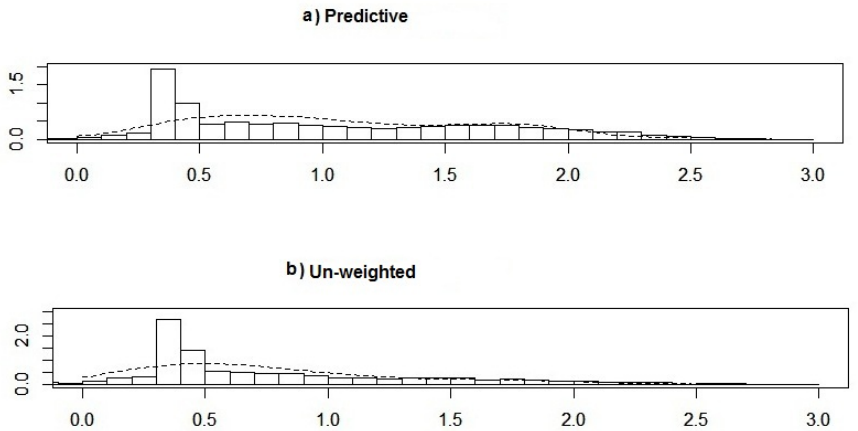

Figure 3. Real dataset of the widths of shrubs with size $n = 46$. (a) Histogram of data and estimated posterior predictive density by DPBMM, and (b) histogram of de-biased data using the Metropolis-Hastings algorithm and $\hat{f}_{J,h}$.

In Figure 2, panel (a), we present the histogram of the simulated data along with its true density curve and the kernel density estimate $\tilde{g}_h$. Panel (b) displays the histogram of the predictive values, the corresponding kernel estimate $\tilde{g}_h$, and the true density curve. In panel (c), we present the histogram of the unweighted distribution obtained using the Metropolis-Hastings algorithm, along with the estimate $\hat{f}_{J,h}$ for this data.

5.2. Real data

In the previous section, Dirichlet Process Bayesian Mixture (DPB2M) models demonstrated a good fit to the simulated data. Therefore, we now apply the flexibility of this Bayesian nonparametric model to real datasets, namely the shrub width data and the bladder cancer data.

5.2.1 Widths of shrubs data

We consider the data that can be found in Muttlak and McDonald (1990) for applications with real data. This data consists of 46 measurements of the width of shrubs that are sampled by line-transect. In this sampling method, the probability of inclusion in the sample is proportional to the width of the shrub, making it a case of length-biased sampling.

We can see the predictive values of the DPB2M model with the histogram and $\tilde{g}_h$ with the dashed line in panel (a) of Figure 3, and also the unweighted version of the data values and $\hat{f}_{J,h}$ depicted in panel (b).

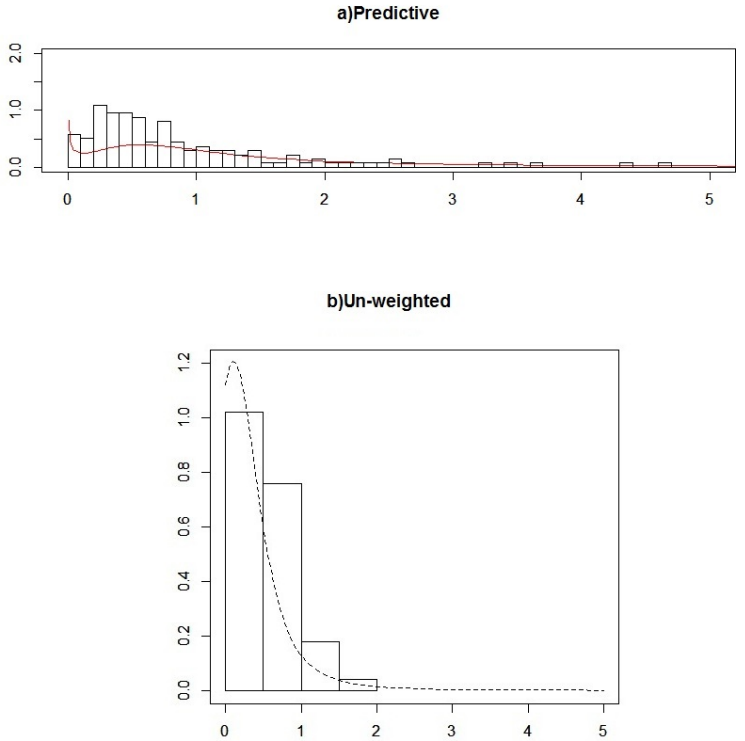

Figure 4. Real dataset of bladder cancer patients with size $n = 137$. (a) Histogram of data and estimated posterior predictive density by DPBMM, and (b) histogram of de-biased data using the Metropolis-Hastings algorithm and $\hat{f}_{J,h}$.

5.2.2 Bladder cancer data

The next real dataset is survival data which includes censored values. This data is taken from Lee and Wang (2003), page 231, which corresponds to remission times (in months) of a random sample of bladder cancer patients. Properties of this data are: the total number of observations 137, censored data 9, largest observation 46.12, and smallest observation 0.08. For easier model fitting, we divided the data into 10 intervals.

In Ahmad et al. (2016), a parametric model was fitted to this dataset without considering censored observations. This model is referred to as the length-biased weighted Lomax distribution. The Lomax distribution is a special case of the Burr(XII) distribution. In this section, we apply the DPBM model, which is a Bayesian nonparametric model with Burr(XII) distribution as the kernel of a mixture model.

In Figure 4, panel (a), we show the histogram of the bladder cancer data, along with the estimated density function based on the DPBM model. Since this dataset comes from a weighted distribution, a histogram of the unweighted values obtained using the Metropolis-Hastings method and the corresponding curve is presented in panel (b) of Figure 4.

6. Conclusion

This article uses the Bayesian nonparametric approach to model weighted data. We use the Dirichlet process mixture model (DPMM) with Burr(XII) distribution as the kernel function in mixing models. We assumed weighted distribution with an arbitrary weight function that satisfies equation (2). Using the Metropolis-Hastings algorithm, the weighted distribution converted to the unweighted one. We fit the DPMM with the different kernels and weight functions for real and simulated data sets as an application. As an application in the survival study, a real lifetime dataset containing censored observations is used, and density and survival functions are estimated.

7. Discussion

While the proposed Bayesian nonparametric model, the Dirichlet Process Mixture of Burr XII distributions, provides substantial flexibility and robustness for modeling weighted data, several limitations should be noted. First, the model's computational complexity can be high, particularly with large datasets or high-dimensional covariates, potentially hindering its scalability in practice. Second, the selection of hyperparameters and prior distributions may significantly affect performance and inference, necessitating careful tuning and sensitivity analysis. Third, although the Burr XII distribution is flexible, there may be cases where other kernel distributions might better capture certain data characteristics.

Future research could address these limitations by extending the model to incorporate covariate information more explicitly, such as through hierarchical or dependent Dirichlet processes, enhancing its utility in complex data settings. Additionally, developing more efficient computational algorithms, such as variational inference or scalable MCMC methods, could improve the model's feasibility for big data applications. Exploring integration with other flexible distributions or developing multivariate extensions may further broaden its scope and practical impact.

Acknowledgements

The author thanks Razi University for facilities provided during this research.

References

Ahmad, A., Ahmad, S. P. and Ahmed, A., (2016). Length-biased weighted Lomax distribution: statistical properties and application. *Pakistan Journal of Statistics and Operation Research*, 12(2), pp. 245–255.

Bohlourihajjar, S., Khazaei, S., (2018). Bayesian nonparametric survival analysis using mixture of Burr XII distributions. *Communications in Statistics-Simulation and Computation*, 47(9), pp. 2724–2738.

Blumenthal, S., (1967). Proportional sampling in life length studies. *Technometrics*, 9(2), 205–218.

Burr, I. W., (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics*, 13(2), 215–232.

Cheng, N., Yuan, T., (2013). Nonparametric Bayesian lifetime data analysis using Dirichlet process lognormal mixture model. *Naval Research Logistics (NRL)*, 60(3), pp. 208–221.

Damien, P., Walker, S., (2002). A Bayesian Non-parametric Comparison of Two Treatments. *Scandinavian Journal of Statistics*, 29(1), pp. 51–56.

Escobar, M. D., West, M., (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430), pp. 577–588.

Ferguson, T. S., (1983). Bayesian density estimation by mixtures of normal distributions. In Recent advances in statistics, pp. 287–302, *Academic Press*.

Fisher, R. A., (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics*, 6(1), pp. 13–25.

Ghosh, J. K., Ramamoorthi, R. V., (2003). *Bayesian nonparametrics*. Springer Series in Statistics. Springer-Verlag, New York.

Gupta, R. C., Kirmani, S. N. U. A., (1990). The role of weighted distributions in stochastic modeling. *Communications in Statistics-Theory and Methods*, 19(9), pp. 3147–3162.

Hassan, et al., (2021). E-Bayesian estimation of Burr Type XII model based on adaptive Type-II progressive hybrid censored data. *International Journal of Computing Science and Mathematics*, 14(3), pp. 233–248.

Hatjispyros, S. J., Nicoleris, T., Walker, S. G., (2017). Bayesian nonparametric density estimation under length bias. *Communications in Statistics-Simulation and Computation*, 46(10), pp. 8064–8076.

Kilany, N. M., (2016). Weighted Lomax distribution. *SpringerPlus*, 5(1), p. 1862.

Hajji Joudaki, et al., (2024). Survival Analysis Using Dirichlet Process Mixture Model with a Three-Parameter Burr XII Kernel. *Communications in Statistics - Simulation and Computation*, 53(5), pp. 2406–2424.

Kottas, A., (2006). Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *Journal of Statistical Planning and Inference*, 136(3), 578–596.

Lanjoni, B. R., Ortega, E. M. and Cordeiro, G. M., (2016). Extended Burr XII regression models: theory and applications. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(1), pp. 203–224.

Lee, E. T., Wang, J., (2003). *Statistical methods for survival data analysis* (Vol. 476). John Wiley & Sons.

Lo, A. Y., (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, pp. 351–357.

Mahafoud, M., Patil, G. P., (1982). On weighted distributions. In *Statistics and Probability: Essays in honor of C. R. Rao*, pp. 383–405.

McLachlan, G., Peel, D., (2004). *Finite mixture models*. John Wiley & Sons.

Michael, J., et al., (2023). Dirichlet Process Mixture Models for the Analysis of Repeated Attempt Designs. *Biometrics*, 79(4), 3907–3915.

Müller, P., Quintana, F. A., (2004). Nonparametric Bayesian data analysis. *Statistical Science*, pp. 95–110.

Muttlak, H. A., McDonald, L. L., (1990). Ranked set sampling with size-biased probability of selection. *Biometrics*, pp. 435–445.

Neal, R. M,. (2003). Slice sampling. *The Annals of Statistics*, 31(3), pp. 705–767.

Nurul, A. B., et al., (2024). Clustering Mixed-Type Data via Dirichlet Process Mixture Models with Cluster-Specific Covariance Matrices. *Symmetry*, 16(6), pp. 712–732.

Patil, G. P., Rao, C. R., (1977). The weighted distributions: A survey of their applications. *Applications of Statistics*, 383.

Patil, G. P., Rao, C. R., (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, pp. 179–189.

Rao, C. R., (1985). Weighted distributions arising out of methods of ascertainment: What population does a sample represent? In *A celebration of statistics*, pp. 543–569. Springer, New York, NY.

Rao, C. R., (1965). On discrete distributions arising out of methods of ascertainment. *Sankhya: The Indian Journal of Statistics, Series A*, pp. 311–324.

Rao, G. S., Aslam, M. and Kundu, D., (2015). Burr-XII distribution parametric estimation and estimation of reliability of multicomponent stress-strength. *Communications in Statistics - Theory and Methods*, 44(23), pp. 4953–4961.

Rodriguez, R. N., (1977). A guide to the Burr type XII distributions. *Biometrika*, 64(1), pp. 129–134.

Sethuraman, J., (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, pp. 639–650.

Walker, S. G., (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation*, 36(1), pp. 45–54.

Zelen, M., Feinleib, M., (1969). On the theory of screening for chronic diseases. *Biometrika*, 56(3), pp. 601–614.