BAYESIAN NONPARAMETRICS FORRELIBALITY STUDIES

Bayesian Modelling and Inference
CSE 5539 Project Report

Ganga Reddy Tankasala (tankasala.1@osu.edu)

Contents

I Introduction

The Objective of this project is to study the application of bayesian nonparametrics(BNP) in estimating reliability of complex systems. Determining accurate reliability is of utmost importance in fields like engineering (Nuclear power plants, etc…) , medicine , automotives and other automation systems. Moreover reliability is closely correlated to quality of service and any business needs to maintain its reliability levels using maintenance services which are constrained by cost. This project explores the application of bayesian Nonparametrics in estimating the lifetime distributions using limited failure data and their results are compared against conventional method described later in this report.

II Limitations of the Conventional methods

The biggest challenge is identifying the reliability of a complex system composed of multiple distinct elements. Today's conventional methods does not provide accurate estimates of the lifetime distributions which might hinder the operations and services of businesses and the cost of mis calculations often leads to failure of system where the loss in capital will be much higher.

However. This is mainly because of the following reasons

1) Limited failure data
2) Limited / Zero knowledge of components of the heterogenous mixture
3) Identification of individual components lifetime distribution
4) Total number of components

Failure time distribution can be approximated to mixture of failure time distributions of its components. The above idea stems from the fact that any complex system can be safely assumed to be composition of unknown number of distributions. Moreover mixture distribution is more realistic than a single distribution model like weibull, Gaussian , log normal etc.. This provides us the motivation to use a mixture model. Since the number of components/subsystems in a complex system are unknown , the most reasonable option to choose is self adaptive non parametric model. Summing up the above inferences , application of Bayesian Nonparametric Mixture model seems a good alternative to conventional model at least theoretically.

III Reliability and distributions

The concept of reliability is closely associated with survival times very much like to the distribution of survival time and failure rates in in most domains.
"Reliability of a component in a system is defined as the probability that it can perform adequately under the expected operation condition over a specific period of time"

This probability can be expressed as

$$R(t) = p(\ T > t) = \int_0^t f(x)dx \qquad (1)$$

where f(x) is the Probability Density Function (PDF) (or time-of-failure density function) of failure time, is a random variable denoting the failure time, and R(t) is the reliability function. The failure rate (or hazard rate) is defined as the rate of possible failures for the survivors to time. It is denoted as

$$h(t) = f(t) / \left(\int_0^t f(t)dt \right) \qquad (2)$$

where F(t) is the Cumulative Distribution Function (CDF), which is the probability that the component may fail within time , i.e

$$F(t) = 1- R(t) = p(\ T <= t) = 1- \int_0^t f(t)dt \qquad (3)$$

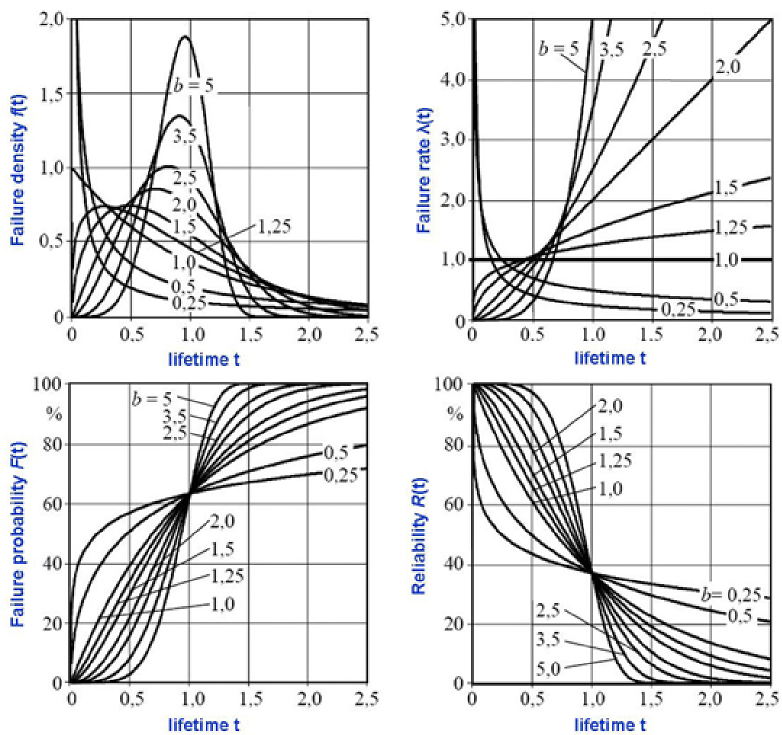

Figure showing probability distributions of failure density, failure rate , failure and reliability

From the above equations one can easily conclude that the if one of the above distribution is predicted an be, the others can be established using simple mathematical relation (like cumulative integral , derivative , etc..) they share.

To estimate the above mentioned unknown quantities, various statistical methods have been proposed and used. As for a multi component , complex system homogenous behavior of a failure time distribution is not an ideal choice. As mentioned above , the component lifetime in complex system is likely best captured by a multi-component lifetime distribution. Such distribution is often called mixture distribution which is a weighted sum of distributions, such weighted by probability sum of one . A mixture distribution can be mathematically written as

$$f(t) \sim \sum_{k=1}^{K} \pi_k f_k(t|\theta_k) \tag{4}$$

Where *f(t)* can be considered as failure/lifetime distribution of complex system, i.e composition of a finite number of distributions, f_k ,each specified by parameters Θ_k are commonly called mixture components. often it is the case that same component distribution with different parameters is used, i.e.$(f_1 = f_2 = f_3 \ldots = f_k)$, but Θ_k are different. The parameter π_k is the k-th mixture coefficient or probability of the influence weight of the k-th mixture component.

IV Parametric and Nonparametric Model

The above defined mixture model is called a finite mixture model if k, the number of components K has to be fixed or estimated. The equation (4) represents a parametric mixture model. These models are simple to understand and apply for any known system and the literature has shown that they promise better results for known systems. On the other hand they cannot completely handle the concerns of a complex system. The disadvantages of using a parametric model are mentioned below.

IV a. Disadvantages of parametric model

- They rely on restrictive parametric assumptions of f(t)
- Sufficient data may not be available to establish the parameters
- Failure mechanism of some subsystems/components is not known
- Failure distribution of some components is difficult to model

Most often it is impossible to assign a parametric distribution for a complex system specially for those systems whose data availability grows with respect to time. The above limitations inspires for Non parametric model which comes the various advantages compared to parametric models.

IV b. Relative Advantages of NonParametric model

- They are Flexible and generic in nature
- Offer Significant modeling flexibility to capture a complex system
- Handles smaller sample sizes
- MCMC has aided for practical applications of bayesian Nonparametrics
- Precise information about k (no of components) is not longer required.

This project mainly focuses mainly on the Dirichlet Process Mixture Model (DPMM) with Bayesian nonparametric perspective that adjusts automatically for the unknown. This project aims to prove the robustness of DPMM in capturing the components of a mixture failure distribution.

Without any informative priors , likelihood of the distribution dominates the analysis which is similar to any non bayesian approaches where parameter selection is determined by likelihood of the function. Thus bayesian nonparametric models which offers the flexibility of having infinite parameters is presented by avoiding critical dependency on prior parameters. The procedure of nonparametric data analysis is often considered as a distribution-free method and the distribution of the data may be based on the rank of which the distribution is most influenced by the data.

IV c. Choosing Parametric and Nonparametric ?

When both parametric and nonparametric methods are applicable to a problem, in general the parametric method is generally preferred due to its efficiency and simplicity. Once the assumptions for the parametric method are questionable, it makes more sense to approach nonparametric methods.

V Dirichlet Distribution

In order to get a good understanding of Dirichlet distribution , lets first consider a simple binomial distribution which is mathematically described as

$$p(X = x \mid n, p) = (n! / x! . (n-x)!)\ p^x . (1- p)^{n-x} \qquad (6)$$

Binomial distributions examples include probability that number of successful events (yes) or unsuccessful events (no) when a bernoulli trial is made. The Binomial distribution is limited to only two(bi) states (Heads / tails of coin), however generalization of Binomial distribution to multiple output states (six states/faces of a dice) rather than two states gives us a distribution named multinomial whose distribution is

$p(x_1 , x_2 ,x_3 ... ,x_n | n, ,p_1 , p_2 ,... ,p_k) = (n! / \Pi (x_i !)) \Pi (p_i^{xi})$ (7)

Parameters of Binomial distribution as defined in (6) is p, i.e probability of success / probability of heads (H). Choosing p often become the challenge and considering the conjugacy, a prior beta distribution on p will result in a posterior distribution of p after one/more observations. In other words beta distribution is conjugate prior for binomial. The beta distribution is given by

$p (p | \alpha , \beta) = B(\alpha , \beta)^{-1} . p^{(\alpha-1)} . p^{(\beta-1)}$ where (8)

$B(\alpha , \beta) = \Gamma(\alpha). \Gamma(\beta) / \Gamma (\alpha + \beta)$ (9)

A Multinomial parameterized by vector of probabilities (p_1,p_2,p_2,,..p_n) such that sum of the vector of probabilities should equate to one. In other words (p_1+ p_2 + p_2 +..+ p_n) = 1. Choosing this vector of probabilities, a conjugate prior of multinomial distribution is considered which happens to be a dirichlet distribution which is a generalization of beta.

$$p(P = \{p_i\} \mid \alpha_i) = \frac{\prod_i \Gamma(\alpha_i)}{\Gamma(\sum_i \alpha_i)} \prod_i p_i^{\alpha_i - 1}$$

(10)

VI Dirichlet Process Mixture Model (DPMM)

Dirichlet process can be considered as a flexible continuous case of the Dirichlet distribution. The definition of the DP includes two base parameters. The first one is a positive scalar (concentration) parameter, which expresses the belief towards and the second one is a probability base distribution (G_o) , which is a nonparametric distribution (empirical distribution). The G is said to follow a Dirichlet Process, which can be written as

$G = DP (vG_o)$

(11)

G is a discrete distribution containing values drawn from (G_o), while (G_o) is a base distribution instead of a base measure (M). DP is also known as " distribution over distributions" because of the fact that any sample from DP itself is a distribution.

In a Bayesian framework, it is necessary to specify a prior distribution to obtain, via Bayes' theorem, the posterior distribution on which statistical inference on the data is based on. The

Dirichlet Process prior fits rich classes of Bayesian nonparametric models with its fulfillment of the two properties namely

1) It is flexible in support of prior distributions and the posteriors can be analyzed easily
2) It can capture the number of unknown mixture components.

In order to implement the DPMM, we apply a mixture model, as in (4), on the DP. For general DP mixture models, when the data size grows and the data becomes more complicated, the theory dictates to assign an infinite number of mixture components and parameters growing with the data, so

$$t_i = f_i\,(\theta_i)\ ,\ i = 1,2,3\ldots \qquad (12)$$

If θ_i is drawn from G or G is set as prior , then the model becomes Dirichlet process Mixture Model (DPMM).

VI a. Hierarchical form of DPMM

$$t_i = f_i\,(\theta_i) \qquad (12)$$
$$\theta_i = G(.) \qquad (13)$$
$$G = DP(vG_o) \qquad (14)$$

G_o is an infinite-dimensional distributional parameter, which makes the DPMM a nonparametric method. Behaviour of the above model is highly sensitive to the choice of the kernel distribution (f_i).

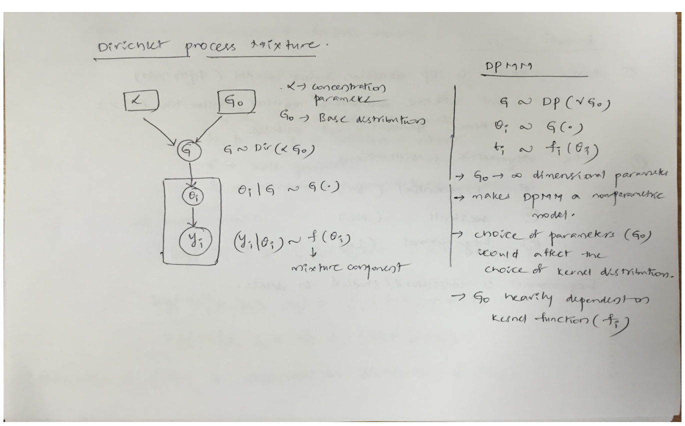

Fig 1. Figure showing the graphical model of DPMM

The basic choice or rather an easy choice of the model is to consider f_i to be a gaussian distribution i.e as the kernel distribution because of its flexibility in handling applications in both conjugate and non-conjugate distributions. However, as the property of lifetime data requires the failure time/ life time to be positive ($f(t) > 0$), Dirichlet Process Gaussian Mixture Model with the data range over all real numbers ($-\infty$, $+\infty$) is not suitable for reliability analysis. Hence to account for the failure/lifetime distributions other asymmetric kernel distributions such as the Exponential, Weibull and Lognormal distribution, are considered for their flexibility and efficiency to fit mixture models. In this project we have used log normal kernel as the base distribution. However to present a brief idea for choosing a kernel distribution, various failure models are presented below.

VII Failure Models

Two major failure models which covers most of the spectrum of lifetime distributions of objects of interest are discussed in this report are namely

1) Accelerated Failure Model (AFT)
2) Proportional Hazard model (PHM)

VII a. Accelerated Failure Model (AFT)

$$\mathbf{S1(t) = S2(c{*}t) \quad \forall\ t \geq 0}$$

In simple definition, if we consider S1 and S2 as survival times of two populations (1 & 2), then AFT model states that aging rate of population 1 is c times that of c2.

Let us consider S1 be survival time of a dog and S2 be survival time of human.

$$S1(t) = S2(c{*}t) \quad \forall\ t \geq 0 \tag{15}$$

From general wisdom, we know that a year for dog is equivalent to 7 years of human. Applying the above model we get $c = 7$ which implies dogs age 7 times faster than humans.

$$\log(T_i) = (\beta_o + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \dots + \beta_p Z_{pi}) + (\sigma\epsilon_i) \tag{16}$$

In a given population of size n, for a given subject , i = 1,2,3,...n,

Z_{1i} are the observed covariates
T_i are observed lifetimes
σ is scale parameter
$\beta_0, \beta_1, \beta_2, \beta_3, \dots \beta_p$ are coefficients of interest
ϵ_i are random disturbance terms, assumed to be independent and identically distributed with a probability density distribution function $f(\epsilon)$.

Distribution of ϵ determines the distribution of T and under various assumptions of ϵ, various AFT models can be derived. Some of them are mentioned below.

For example if ϵ_i is normal , then T_i will be

$$T_i = \exp(\beta_o + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \ldots + \beta_p Z_{pi}) * \exp(\sigma\epsilon_i) \quad \text{(A)}$$
$$\log(T_i) = (\beta_o + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \ldots + \beta_p Z_{pi}) + \sigma\epsilon_i \quad \text{(B)}$$
$$\log(T_i \mid \beta_0, .. \beta_p, Z_i) = (\quad k1 \quad) + \sigma\epsilon_i \quad \text{(C)}$$

The above equation implies that T_i follows a lognormal distribution , similarly for the various assumptions ϵ_i the lifetime distribution will be as follows.

	Distribution of ϵ_i	Distribution of T_i
1.	Normal	Lognormal
2.	Logistic	Log Logistic
3.	Log gamma	Gamma
4.	Extreme values(1 params)	Exponential
5.	Extreme values(2 params)	Weibull

VII b. Proportional Hazard Model (PHM)

Also known as Cox Model , it is a statistical model which defines the model in terms of hazard rate as follows.

$$\lambda(t \mid Z, \beta) = \lambda_o(t) * \exp(\beta_o + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \ldots + \beta_p Z_{pi}) \quad (17)$$

$\lambda(t)$ is the hazard function
Z_{1i} are the observed covariates
$\beta_0, \beta_1, \beta_2, \beta_3, \ldots \beta_p$ are coefficients of interest

It implies that hazard rate of a system increases / decreases proportionally to the effect of covariates with respect to a baseline distribution, hence the name Proportional hazard model.In this project we have considered only log normal as the kernel distribution. However after going through the results one can easily deduce that any logical distribution can be substituted in place of log normal distribution.

VIII Methodology

The Dirichlet Process depends heavily on its base distribution G_o. Based on the nonparametric statistical properties, an empirical cumulative density function (CDF) of is obtained as well as its probability density function (PDF). Consequently, the estimated CDF, PDF and hazard rate function (HRF) of lifetime data can be obtained applying equations (2,3,4,5). In this section, a procedure for estimating the mixture density function of lifetime data using DPMM based on the Lognormal kernel function is given

a. Model Specification

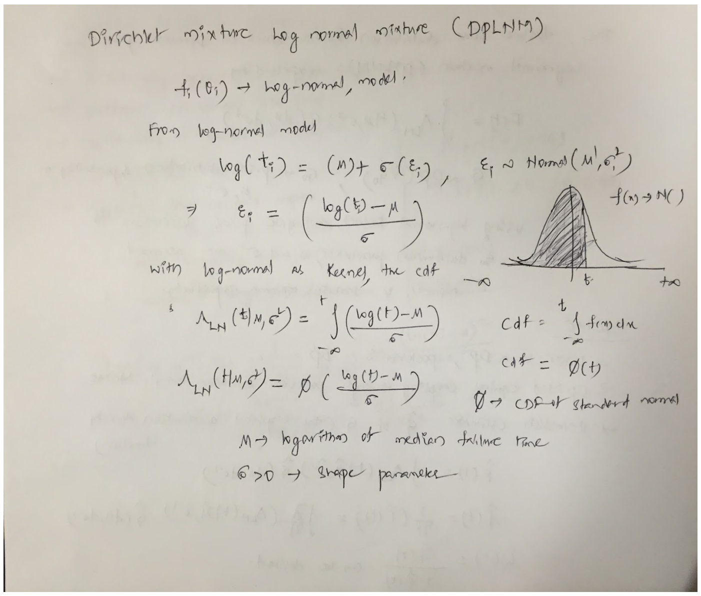

Fig2 Figure showing the cdf of DPLNM model

A log normal distribution has the cdf of the form of (by following equations A,B,C)

$$\Lambda_{LN}(t|\mu, \sigma^2) = \Phi\left[\frac{\log(t) - \mu}{\sigma}\right] I(t \geq 0) \tag{18}$$

where Φ is the CDF of the standard normal distribution (standard Gaussian) is the logarithm of the median time of failure time and is the shape parameter (deviation parameter). The failure time distribution function based on the Dirichlet Process Lognormal Mixture (DPLNM) is expressed by

$$F(t) = \int \Lambda_{LN}(t|\mu, \sigma^2) G(d\mu, d\sigma^2) \tag{19}$$

Using the Lognormal kernel distribution, the conjugate prior distributions for distribution parameters, and, are assumed Normal and Inverse-Gamma (IG), respectively. for the DPLNM depends on hyper parameters which are incorporated into the hierarchical form of the DPLNM model.

b. MCMC Algorithm

Based on the conditional distributions and the prior distributions, we can obtain the full conditional posterior distribution of the Lognormal kernel distribution, which depends on the prior distributions of the hyper parameters. As the prior distribution is a DP the obtained posterior distribution is also a DP as well. The DP mixture model requires sampling through Markov-Chain-Monte-Carlo (MCMC) methods, which is computationally intensive because updating the Markov Chain process required for each replication is necessary until convergence is achieved. Finally, an estimate , $F^{\wedge}(t)$ of the empirical distribution density function is obtained. Using and Bayes law, the cumulative distribution function of the failure time is estimated by

$$\hat{F}(t) = \int \Lambda_{LN}(t|\hat{\mu}, \hat{\sigma}^2)\hat{G}(d\mu, d\sigma^2) \tag{20}$$

The PDF of failure time is

$$\hat{f}(t) = \frac{\partial \hat{F}(t)}{\partial t} = \int \frac{\partial \Lambda_{LN}(t|\hat{\mu}, \hat{\sigma}^2)}{\partial t} \hat{G}(d\mu, d\sigma^2) \tag{21}$$

and Hazard rate and reliability are calculated using equations (3,4,5).

The nonparametric Bayesian DPMM using the MCMC procedure of the hierarchical structure can be summarized in Figure shown below.

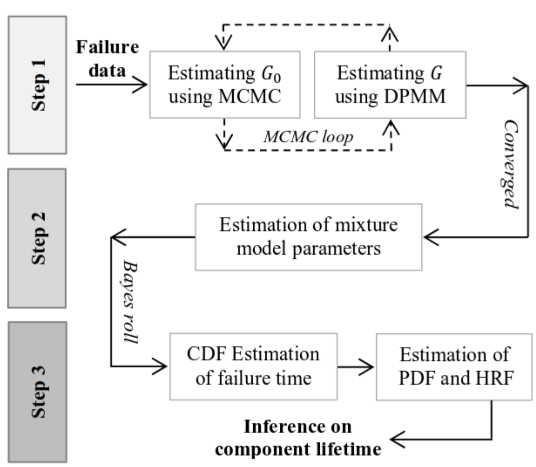

Fig.1 The process of nonparametric Bayesian analysis based on DPMM using MCMC algorithm

IX Results

In order to demonstrate the effectiveness of the proposed models and determine their suitability for reliability analysis of a complex system , a simulation test is undertaken to study the estimation of the lifetime distributions by the proposed models against a known failure data distribution.

In this test, an artificially generated lifetime data set is provided as reference data in a model-based simulation as its very hard to get failure data because of the reason that most systems are replaced before their lifetime in order to avoid untimely failure.

Generally, in a mixture model each mixture component density represents the probability distribution function of a group of individuals in the whole population. Based on the nature of the lifetime type data, a mixture of long tail distributions is assumed. A mixture of Lognormal (LN) distribution and Inverse-Gaussian (IGN) distribution is considered for the creating artificial data . Based on the mixture model definition , the LN-ING mixture model is shown as below

$$f(t) = p.\ LN\ (4, 0.16) + (1-p)\ .\ IGN\ (8,0.49) \qquad (22)$$

The performance of parametric mixture (PARMIX), Exponential kernel (DPEM), Weibull kernel (DPWM) and Lognormal kernel (DPLNM) are compared with those of the actual model. The following figures shows that PDF, CDF and hazard rate capture by each model.

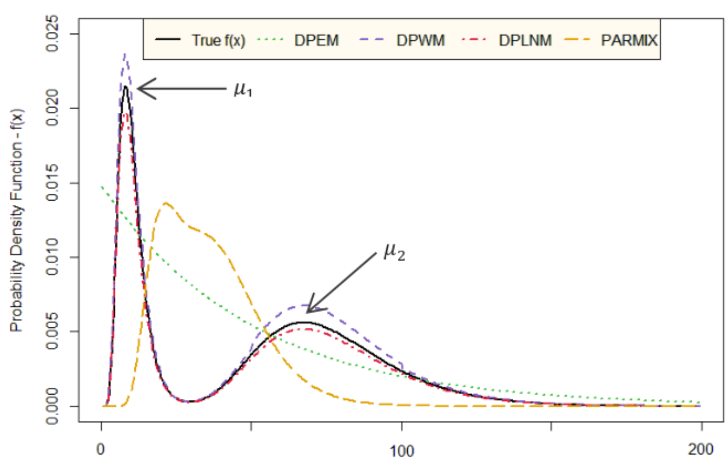

Fig 3. Figure showing PDF of original and mixture models

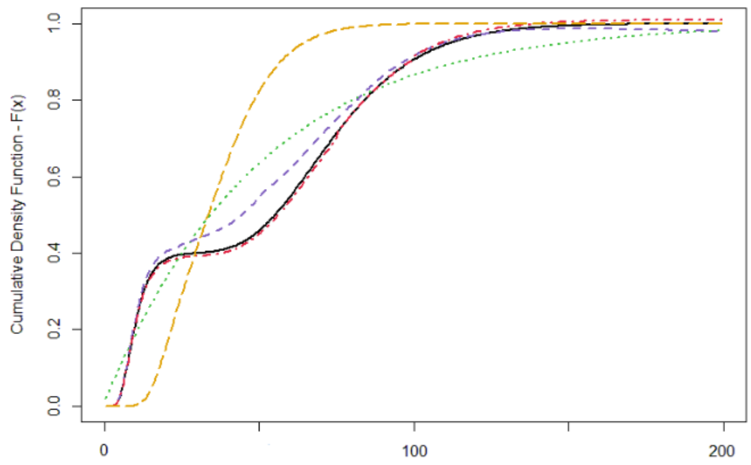

Fig 4. Figure showing CDF of original and mixture models

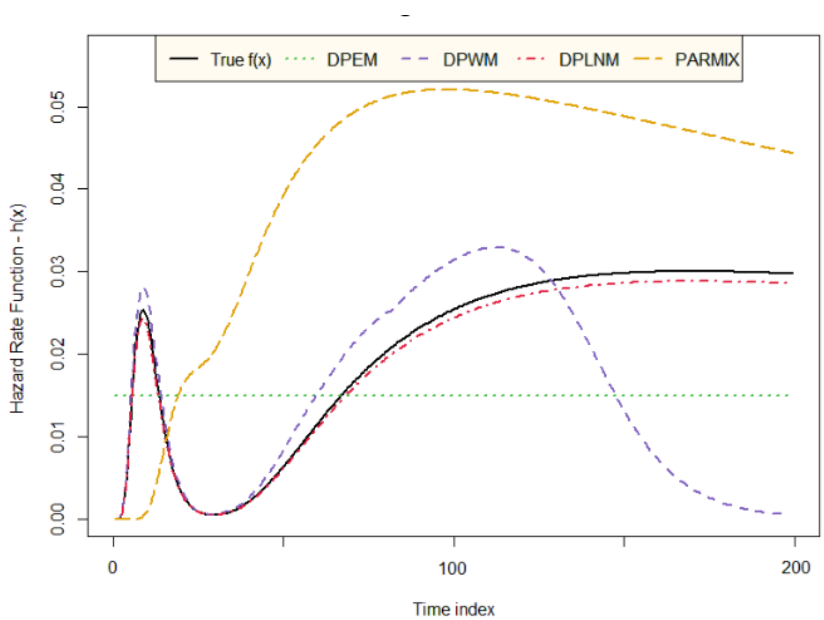

Fig 5. Figure showing Hazard rate of original and mixture models

A quantitative comparison is made using Kolmogorov-Smirnov test, a nonparametric test for goodness-of-fit, to determine the closeness of the mixture models against the original mixture model. The results are shown in table below

Kolmogorov-Smirnov goodness-of-fit test of failure time cumulative density and hazard rate function estimation

Models	Kolmogorov-Smirnov test			
	$F(x)$		$h(x)$	
	Test Stats	*P-value*[a]	*Test Stats*	*P-value*[a]
PARMIX	0.1652	0.004	0.2193	0.000
DPEM	0.1946	0.002	0.2164	0.000
DPWM	0.1106	0.106	0.0694	0.039
DPLNM	0.02681	0.868	0.0451	0.667

[a.] The significant level is 0.05

.

Table1 Kolmogorov-Smirnov goodness-of-fit for CDF and Hazard rate

X Conclusions

The following observations are noted after a close inspection of the figures and table

1) Exponential kernel and PARMIX are not capable of capturing the generated mixture distribution with long tail.
2) Though pdf, cdf of DPWM match closely to the original data, hazard rate of DPMW is different from the original. We can conclude that weibull is not a good choice inspite of weibull being a long tail model like the original model.
3) The estimated PDF, CDF and hazard rate function of the generated data using DPLNM show a close match. Hence Log normal models are an ideal choice for mixture type distributions.

XI Future work

The results convinces one of scope of bayesian nonparametric mixture models in reliability analysis. The possible extensions of this project could be

1) The work can can be extended to other failure models such as PHM
2) The work can be extended to other degradation models like Gompertz, Extreme value distributions, and other distributions.
3) The results can be verified using failure distribution of a real time complex system (like nuclear power plants, medical diagnosis etc…)

XII References

1) Bayesian Nonparametric Reliability Analysis for a Railway System at Component Level (http://ro.uow.edu.au/cgi/viewcontent.cgi?article=2915&context=eispapers)
2) A tutorial on Bayesian nonparametric models (http://web.mit.edu/sjgershm/www/GershmanBlei12.pdf)
3) Bayesian Parametric Analysis for Reliability Study of Locomotive Wheels (http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6517760&tag=1)

4) Non parametric and semi parametric Bayesian reliability analysis
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.330.5607&rep=rep1&type=pdf
5) Lecture notes -- Bayesian Modelling and Inference (CSE 5539)
(http://web.cse.ohio-state.edu/~kulis/teaching/5539_sp15/index.htm)
6) Wikipedia