

A BAYESIAN ANALYSIS ON WEIBULL MODEL ALLOWING NEARLY INSTANTANEOUS FAILURES

K. Muralidharan¹, Rajiv Parikh² and C. D. Lai³

•
¹Department of Statistics, Faculty of Science,
The M. S. University of Baroda, Vadodara- 390 002, India.
Email: muralikustat@gmail.com

²Department of Statistics,
Bhavnagar University, Bhavnagar- 364 002, India.
Email: parikh.rajeev@gmail.com

³Institute of Fundamental Sciences,
Massey University, Palmerston North, New Zealand
c.lai@massey.ac.nz

Abstract: In this article, we study, the maximum likelihood as well as Bayes estimation on parameters of Mixture of Weibull with ‘nearly instantaneous failure’ as introduced in Lai et.al. (2007). For Maximum likelihood estimation, the EM algorithm is used. For Bayes estimation of parameters, we used three different algorithms namely, Population Monte Carlo method (PMC), Mixture version of Metropolis-Hasting and Gibbs sampler. The methods are compared using a simulation study. A numerical example is also discussed at the end of the paper.

Key Words: Maximum Likelihood, Expectation-Maximization, Metropolis-Hasting, Gibbs, Population Monte Carlo method, Bayes Estimation, Mixture of Non-Degenerate and Weibull Model.

1. Introduction

The mixed failure time distribution arises frequently in many different contexts in statistical literature. For instance, when we put units in a life testing experiment, then some of the units fail instantaneously and thereafter the life time of units follow a distribution such as exponential, Weibull, gamma etc. Such situations may be represented as a mixture of singular distribution at zero and a positive continuous random variable distribution. Lai et. al. (2007) has proposed a model as a mixture of generalized Dirac delta function and the 2-parameter Weibull having a closed form expression for survival function and hazard rate. The density of such a mixture can be shown as

$$f(x) = p\delta_d(x - x_0) + q\alpha\lambda^\alpha x^{\alpha-1} e^{-(\lambda x)^\alpha}, \quad p+q=1, q=1-p, 0 < p < 1 \quad (1)$$

Where

$$\delta_d(x - x_0) = \begin{cases} \frac{1}{d}, & x_0 \leq x \leq x_0 + d \\ 0, & \text{otherwise} \end{cases}$$

For sufficiently small d . Here $p > 0$ is the mixing proportion. We note that

$$\delta(x - x_0) = \lim_{d \rightarrow 0} \delta_d(x - x_0),$$

where $\delta(\cdot)$ is the Dirac delta function that is well known in mathematical analysis. One can view the Dirac delta function as a normal distribution having a zero mean and standard deviation that tends to 0. For a fixed value of d , the distribution becomes a mixture of a Weibull with Uniform distribution.

As a special case, the model presented in (1) for $x_0 = 0$ becomes a mixture of Weibull with a uniform distribution (See Muralidharan and Lathika, 2007). The pdf of mixture of Weibull with ‘nearly instantaneous failure’ occurring uniformly over $[0, d]$ can be given as

$$f(x) = \frac{p}{d} + (1-p)\alpha\lambda^\alpha x^{\alpha-1} e^{-(\lambda x)^\alpha}, \quad 0 < p < 1 \quad (2)$$

Taking $f_1 \sim U(x_0, x_0 + d)$ and $f_2 \sim \text{Weibull}(\alpha, \lambda)$, then (2) can be written as

$$f(x) = p f_1(x_i | d) + (1-p) f_2(x_i | \alpha, \lambda) \quad (3)$$

If X_1, X_2, \dots, X_n are the random sample of size n from (3), then the likelihood function is given by

$$L(p, \alpha, \lambda, d | \underline{x}) = \prod_{i=1}^n f(x_i)$$

Introducing Latent variable Z_i in the likelihood function, we have

$$L(p, \alpha, \lambda, d | \underline{x}, \underline{z}) = \prod_{i=1}^n \{p f_1(x_i | d)\}^{z_i} \{(1-p) f_2(x_i | \alpha, \lambda)\}^{1-z_i} \quad (4)$$

where

$$P(z_i = 1) = p f_1(x_i | d), \quad P(z_i = 0) = (1-p) f_2(x_i | d)$$

After some simplification, the Log-likelihood can be written as

$$\ln(L(p, \alpha, \lambda, d | \underline{x}, \underline{z})) = \sum_{i=1}^n z_i \ln \left(\left[\frac{p}{1-p} \right] \left[\frac{f_1(x_i | d)}{f_2(x_i | \alpha, \lambda)} \right] \right) + \ln((1-p) f_2(x_i | \alpha, \lambda)) \quad (5)$$

For further development in the study, we make use of the likelihood given in (5). One may refer to Lai et.al. (2007) paper for various characteristics of the model and the parametric estimation of the model.

Here, we compute the Maximum likelihood estimator using EM algorithm (see section-2), while in section-3, the Bayes estimators of the parameters are calculated. The posterior samples have been generated using three different approaches namely, Population Monte Carlo, Mixture

version of Metropolis-Hasting and Gibbs sampler. The performance of the algorithms is evaluated using a simulation study. The last section discusses a practical example.

2. Maximum Likelihood Estimators

To find MLE of the parameters of an underlying distribution we use EM algorithm given by Dempster, Laird and Rubin (1977). The EM algorithm has two advantages here: The first occurs when the data indeed has missing values, due to problems with or limitations of the observation process. The second occurs when optimizing the likelihood function is analytically intractable but when the likelihood function can be simplified by assuming the existence of and values for additional but missing (or hidden) parameters. There is much literature devoted to extensions and applications of the EM algorithm, and this is summarized in McLachlan & Krishnan (1997).

The expectation step or E-step computes the expected likelihood for the complete data. Let θ be the complete collection of parameters occurring in the mixture, i.e. $\theta = (d, \alpha, \lambda)$.

E-Step: We take expectation and get Q function

$$Q(\theta|\theta^{(t)}) = E[\ln(L(p, \theta | Z, X)) | X, \theta^{(t)}]$$

$$= \sum_{i=1}^n E(Z_i | x_i, \theta^{(t)}) \ln \left(\left[\frac{p}{1-p} \right] \left[\frac{f_1(x_i | d)}{f_2(x_i | \alpha, \lambda)} \right] \right) + \ln((1-p) f_2(x_i | \alpha, \lambda)) \quad (6)$$

where

$$E(Z_i | x_i, \theta^{(t)}) = P(Z_i = 1 | x_i, \theta^{(t)})$$

$$= \frac{p f_1(x_i | d^{(t)})}{p f_1(x_i | d^{(t)}) + (1-p) f_2(x_i | \alpha^{(t)}, \lambda^{(t)})}$$

$$= \frac{\frac{p^{(t)}}{d^{(t)}}}{\frac{p^{(t)}}{d^{(t)}} + (1-p^{(t)}) \alpha^{(t)} \lambda^{(t)\alpha^{(t)}} x_i^{\alpha^{(t)}-1} e^{-(\lambda^{(t)} x_i)^{\alpha^{(t)}}}} \quad (7)$$

M-Step: Here we maximize the expectation, i.e., the Q-function that we computed in the E-step. The two steps may be repeated as per requirement

Maximizing $Q(\theta|\theta^{(t)})$ w.r.t. θ yields the update equations

$$\frac{\partial Q}{\partial p} = 0 \Rightarrow p^{(t+1)} = \frac{1}{n} \sum_{i=1}^n E(Z_i) \quad (8)$$

$$\frac{\partial Q}{\partial d} = 0 \Rightarrow -\frac{1}{d} \sum_{i=1}^n E(z_i) = 0 \quad (9)$$

$$\frac{\partial Q}{\partial \lambda} = 0 \Rightarrow \lambda^{(t+1)} = \left[\frac{\sum_{i=1}^n (1 - E(z_i))}{\sum_{i=1}^n (1 - E(z_i)) x_i^{\alpha^{(t)}}} \right]^{1/\alpha^{(t)}} \quad (10)$$

$$\frac{\partial Q}{\partial \alpha} = 0 \Rightarrow \sum_{i=1}^n (1 - E(z_i)) \left[\frac{1}{\alpha^{(t)}} + \ln(\lambda^{(t)}) + \ln(x_i) - (\lambda^{(t)} x_i)^{\alpha^{(t)}} \ln(\lambda^{(t)} x_i) \right] = 0 \quad (11)$$

Since equation (9) cannot be solved analytically, we use $d^{(t+1)} = \text{Max}(x_i)$ since $x_0 \leq x_i \leq x_0 + d$. To obtain $\alpha^{(t+1)}$, we solve equation (11) numerically using Newton-Raphson method.

The EM algorithm starts by assigning initial values to all parameters to be estimated. It then iteratively alternates between two steps, the E-step and M-step. The E-step computes the expected likelihood and the M-step re-estimates all the parameters by maximizing the Q-function. If the convergence criteria is not met, then the parameters p, α, λ and d are updated. We can repeat E-step followed by the M-step until the likelihood converges. Every iteration is guaranteed to increase the log-likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function.

3. Bayesian Estimate when all parameters are unknown

Let the prior distributions be

$$\pi(p) = \frac{1}{B(\eta, \beta)} p^{\eta-1} (1-p)^{\beta-1}, \quad \eta > 0, \beta > 0$$

$$\pi(d) = \frac{1}{\xi}, \quad 0 < d < \xi$$

$$\pi(\alpha) = \theta e^{-\alpha\theta}, \quad \theta > 0$$

and

$$\pi(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{-(a+1)} e^{-b/\lambda}$$

Then the joint pdf becomes

$$f(\underline{x}; p, d, \alpha, \lambda) = \prod_{i=1}^n f(x_i) \pi(p) \pi(d) \pi(\alpha) \pi(\lambda)$$

$$= \prod_{i=1}^n \left(\frac{p}{d} + (1-p)\alpha \lambda^\alpha x_i^{\alpha-1} e^{-(\lambda x_i)^\alpha} \right) \frac{1}{\xi} \theta e^{-\alpha\theta} \frac{1}{B(\eta, \beta)} p^{\eta-1} (1-p)^{\beta-1} \frac{b^a}{\Gamma(a)} \lambda^{-(a+1)} e^{-b/\lambda} \quad (12)$$

Thus, the posterior density is proportional to

$$\Pi(p, d, \alpha, \lambda | \underline{x}) \propto \prod_{i=1}^n \left(\frac{p}{d} + (1-p)\alpha \lambda^\alpha x_i^{\alpha-1} e^{-(\lambda x_i)^\alpha} \right) e^{-\alpha\theta} p^{\eta-1} (1-p)^{\beta-1} \lambda^{-(a+1)} e^{-b/\lambda} \quad (13)$$

To generate sample from (13), we use Population Monte Carlo (PMC) algorithm and Mixture version of Metropolis-Hasting algorithm. The PMC algorithm is an iterated importance sampling scheme, it is an adaptive algorithm that calibrates the proposal distribution to the target distribution at each iteration by learning from the performance of the previous proposal distributions. A complete detail about PMC can be found in Capp'e et al. (2004). Letting $\theta = (p, d, \alpha, \lambda)$, the PMC algorithm can be given as

Algorithm-1

Initialize $\theta_0^{(1)}, \dots, \theta_0^{(M)}$; set $t=1$.

Step – 1 For $i=1, \dots, M$

1.1.1 Generate $\theta_{(t)}^{(i)} \sim q_{it}(\theta_{(t-1)}^{(i)})$

1.1.2 Compute

$$\rho^{(i)} = \frac{\Pi(\theta_{(t)}^{(i)})}{q_{it}(\theta_{(t-1)}^{(i)})}$$

1.2 Compute $w^{(i)} = \frac{\rho^{(i)}}{\sum_{l=1}^M \rho^{(l)}}$

1.3 Resample M values with replacement from the $\theta_{(t)}^{(i)}$'s using the weights $w^{(i)}$.

Step – 2 Repeat step – 1 for $t=2$ to N

Below we present Metropolis-Hasting (MH) algorithm corresponding to the mixture distribution shown in (13).

Algorithm-2

Initialize $\theta^{(0)}$; set $t=0$.

Step 1: Generate $\tilde{\theta} \sim q(\theta | \theta^{(t)})$ and $u \sim \text{Uniform}(0, 1)$

Step 2: Compute

$$r = \frac{f(\underline{x}|\tilde{\theta}) \pi(\tilde{\theta}) q(\theta^{(t)}|\tilde{\theta})}{f(\underline{x}|\theta^{(t)}) \pi(\theta^{(t)}) q(\tilde{\theta}|\theta^{(t)})}$$

Step 3: If $u < r$ then

$$\theta^{(t+1)} = \tilde{\theta}$$

else

$$\theta^{(t+1)} = \theta^{(t)}$$

Step 4: Repeat steps 1 to 2 for $t = 1, 2, \dots, N$ and return the values $\theta^{(1)}, \dots, \theta^{(T)}$.

It is to be noted that in both the algorithms, the instrumental distributions q , for (p, d, α, λ) is taken as their prior distributions. As seen above, Population Monte Carlo method is a combination of the Markov Chain Monte Carlo algorithms (for the construction of the proposal), Importance Sampling (for the construction of appropriate estimators), Sampling Importance Resampling (for sample equalization) and Iterated particle systems (for sample improvement). Thus the Population Monte Carlo (PMC) algorithm is in essence an iterated importance sampling scheme that simultaneously produces, at each iteration, a sample approximately simulated from a target distribution and provide unbiased estimates for that distribution. The sample is constructed using sample dependent proposals for generation and importance sampling weights for pruning the proposed sample.

Now, we define another approach to carry out Bayes estimation in mixture context. Introducing latent in the structure, Posterior density can be written as

$$\prod_{i=1}^n \left\{ \frac{p}{d} \right\}^{z_i} \left\{ (1-p) \alpha \lambda^\alpha x_i^{\alpha-1} e^{-(\lambda x_i)^\alpha} \right\}^{1-z_i} \theta e^{-\alpha \theta} p^{\eta-1} (1-p)^{\beta-1} \lambda^{-(a+1)} e^{-b/\lambda} \tag{14}$$

Then the full conditional distribution of (p, d, α, λ) is given by

$$\Pi(d|z) \propto d^{-n_1},$$

$$\Pi(p|z) \propto p^{n_1+\eta-1} (1-p)^{n_2+\beta-1},$$

$$\Pi(\alpha|z, \lambda) = \alpha^{n_2} e^{-\alpha \theta} \prod_{\{i: z_i=0\}} x_i^{\alpha-1} e^{-(\lambda x_i)^\alpha}$$

and

$$\Pi(\lambda|z, \alpha) = \lambda^{n_2\alpha-a-1} e^{-b/\lambda} \prod_{\{i: z_i=0\}} e^{-(\lambda x_i)^\alpha} \tag{15}$$

Where $n_1 = \sum_{i=1}^n I_{z_i=1}$, $n_2 = \sum_{i=1}^n I_{z_i=0}$ and $n = n_1 + n_2$

For generating a sample from this full conditional distribution, we use following algorithm.

Algorithm-3

Initialize $p^{(0)}$, $d^{(0)}$, $\alpha^{(0)}$ and $\lambda^{(0)}$; Set $t=1$.

Step – 1 Generate $z_i^{(t)}$ ($i=1, \dots, n$) from

$$P(z_i^{(t)}=1) \propto p^{(t-1)} / d^{(t-1)}$$

$$P(z_i^{(t)}=0) \propto (1 - p^{(t-1)}) \alpha^{(t-1)} \lambda^{(t-1)} \alpha^{(t-1)} x_i^{\alpha^{(t-1)} - 1} e^{-(\lambda^{(t-1)} x_i)^{\alpha^{(t-1)}}}$$

Step – 2 Compute $n_1^{(t)} = \sum_{i=1}^n I_{z_i^{(t)}=1}$ and $n_2^{(t)} = \sum_{i=1}^n I_{z_i^{(t)}=0}$

Step – 3 Generate $p^{(t)} \sim \text{Beta}(n_1^{(t)} + \eta - 1, n_2^{(t)} + \beta - 1)$

Step – 4 Generate $d^{(t)} \sim \text{Uniform}(d^{(t-1)})$

Step – 5 Generate $\alpha^{(t)} \sim \Pi(\alpha^{(t)} | n_2^{(t-1)}, \lambda^{(t-1)})$ using M-H algorithm.

Step – 6 Generate $\lambda^{(t)} \sim \Pi(\lambda^{(t)} | n_2^{(t-1)}, \alpha^{(t-1)})$ using M-H algorithm.

Step – 7 Repeat step-1 to 6 for $t = 2, 3, \dots, N$

To implement the M-H algorithm, it is necessary that a suitable candidate-generating density be specified. If the domain explored is too small, compared with the range of f , the Markov chain will have difficulties in exploring this range and thus will converge very slowly so q may be (for both α and λ , candidate-generating density is taken as exponential distribution) chosen in such a way that it converge to target distribution. Wide range of choices of q has been given in the literature; see Tierney (1994) and Chib and Greenberg (1995) and references contained therein.

4. Simulation Study

Here we have used the entire three algorithms to generate a posterior sample. A simulation study is carried out to compare the performance of different algorithms. A sample of size 50 was drawn from the population by taking different values of the parameters. On the basis of this sample, Bayes estimates are calculated using these algorithms. It was seen that, the number of iterations necessary to reach convergence for PMC is below 5000, while for M-H, it is below 15000 and for Gibbs, and it is above 20,000. A deeper difficulty in implementing Algorithm 3 is the existence of computational trapping states. Table-1 shows the ML estimators and Bayes estimators using three different algorithms.

Table 1 MLE and Bayes estimates

Parameter	MLE	PMC	M-H	Gibbs
p=0.2	0.2567	0.2084	0.2156	0.1692
d=0.05	0.0486	0.0467	0.0532	0.0594
$\alpha=2.0$	1.8345	1.8857	1.8678	1.8523
$\lambda=1.5$	1.4976	1.3885	1.3563	1.6974
p=0.6	0.5678	0.6132	0.5817	0.6679
d=0.05	0.0497	0.0522	0.0423	0.0526
$\alpha=2.0$	1.8754	1.8804	1.7923	1.9069
$\lambda=1.5$	1.6925	1.4760	1.5398	1.6856

5. Application to a real life data

Here, we consider the wood dryness data of 40 boards analyzed in Lai et.al. (2007) with $t_i=0, i=1,2,\dots,28$ and the other positive observations are 0.0463741, 0.0894855, 0.4, 0.42517, 0.623441, 0.6491, 0.73346, 1.35851, 1.77112, 1.86047, 2.12125, 2.12. Here we spread the zeros uniformly over an interval taking $d=0.042$ so that $t_1 = 0, t_2 = 0.0015, t_3 = 0.003, \dots, t_{28} = 0.042$. We obtain MLE estimates using EM algorithm as $\hat{p}_{MLE} = 0.7625, \hat{d} = 0.0418, \hat{\alpha}_{MLE} = 1.5689$ and $\hat{\lambda}_{MLE} = 1.2458$. PMC method is used to find Bayesian estimates. Below we present two Bayes estimates related to same data set but with different prior values.

- (i) $\hat{p}_{BAYES} = 0.6899, \hat{d}_{BAYES} = 0.0404, \hat{\alpha}_{BAYES} = 0.9363$ and $\hat{\lambda}_{BAYES} = 1.9704$ for $\theta = 1.6, \eta = 1.1, \xi = 0.9, \beta = 2.3, a = 1.5, b = 1.9$
(ii) $\hat{p}_{BAYES} = 0.7187, \hat{d}_{BAYES} = 0.0649, \hat{\alpha}_{BAYES} = 0.9015$ and $\hat{\lambda}_{BAYES} = 1.9704$ for $\theta = 2.5, \eta = 4.0, \xi = 0.05, \beta = 2.3, a = 2.0, b = 2.0$

6. Sensitivity analysis.

As can be seen from tables 2 and 3 that the estimate of the estimates depends on the choice of d . In practice, the value of d can be manually estimated quite accurately from the dataset. Even if the value of d is inside the peak of the target distribution (here it is Weibull), we are still able to estimate all four parameters of the model. The estimates are consistent for both parametric and Bayesian set up.

Table 2. Uniform spread of “nearly instantaneous failures” with $d = 0.135$

	\hat{p}	\hat{d}	$\hat{\alpha}$	$\hat{\lambda}$
Maximum likelihood estimates	0.78960	0.135	1.6852	1.2398
Bayes Estimates using PMC	0.8729	0.1265	1.2351	1.4388

Table 3. Uniform spread of “nearly instantaneous failures” with $d = 0.2$

	\hat{p}	\hat{d}	$\hat{\alpha}$	$\hat{\lambda}$
Maximum likelihood estimates	0.6125	0.2000	1.2341	1.3563
Bayes Estimates using PMC	0.7142	0.2764	1.5434	1.7649

References

- Cappe’ O., Guillin, A. Marin, Jean-M., Robert, C. P. (2004). Population Monte Carlo. *J. Comp. Graph. Statist.*, 13,907-930.
- Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *Journal of the American Statistical Association*, Vol. 46 (3) 167-174.
- Celeux, G. and Diebolt, J. (1985). The SEM Algorithm: a Probabilistic Teacher Algorithm derived from the EM algorithm for the Mixture problem. *Computational Statistics*, Quarterly, 2, 73-92.
- Celeux, G., Hurn, M. and Robert, C. P. (2000). Computational and Inferential Difficulties with Mixture Posterior distributions. *Journal of the American Statistical Association*, 95(451), 957-970.
- Celeux, G., Hurn, M. and Robert, C. P. (2000). Computational and Inferential Difficulties with Mixture Posterior distributions. *Journal of the American Statistical Association*, 95(451), 957-970.

-
- Chan, K. and Geyer, C. (1994). Discussion of “Markov Chains for Exploring Posterior Distribution” *The Annals of Statistics*, 22, 1747-1758.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis – Hastings Algorithm, *The American Statistician*, 49(4), 327-335.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39, 1-38.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (2000). *Markov Chain Monte Carlo in practice*. Chapman & Hall/ CRC, London.
- Lai, C. D., Khoo, B. C., Muralidharan, K. and Xie, M. (2007). Weibull model allowing nearly instantaneous failures. *Journal of Applied Mathematics and Decision Sciences*, Article ID 90842, 11 pages.
- Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24, 101-121.
- Muralidharan, K. (1999). Tests for the mixing proportion in the mixture of a degenerate and exponential distribution. *Journal of Indian Statistical Association*, Vol. 37, issue 2.
- Muralidharan, K. (2000). The UMVUE and Bayes estimate of reliability of mixed Failure time distribution, *Communication in Statistics- Simulations & computations*, 29(2), 603-619.
- Muralidharan, K. and Lathika, P. (2006). Analysis of instantaneous and early failures in Weibull distribution, *Metrika*, 64(3), 305-316.
- Raftery, A. and Lewis, S. (1992). How many Iterations in the Gibbs Sampler? In Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds.), *Bayesian Statistics*, 4, 763-773, Oxford: Oxford University Press.
- Ritter, C. and Tanner, M. (1992). Facilitating the Gibbs sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler, *Journal of the American Statistical Association*, 87(419), 861-868.