

## **MODELING THE PRESENCE OF LONG-TERM SURVIVORS USING GENERALIZED BURR XII MIXTURE MODEL**

**M. E. GHITANY, F. A. AL-AWADHI and S. A. AL-AWADHI**

Department of Statistics and Operations Research, Faculty of Science  
Kuwait University, P. O. Box 5969, Safat 13060, Kuwait

### **Abstract**

In this paper, we propose a generalized Burr XII mixture survival model in which the probability of long-term survivorship and the timing of event occurrence are modeled jointly. Both the maximum likelihood and Markov Chain Monte Carlo methods are used to make inference about the parameters of the model. Application of the proposed model to survival data of leukaemia patients is given and its goodness-of-fit is demonstrated.

### **1. Introduction**

Common parametric survival models, such as the exponential, Weibull and Gamma, are appropriate for data in which the event of interest (death or relapse, for example) is certain if we observe population individuals for long enough period of time. However, this implication is unrealistic in many contexts, since it does not take into consideration the existence of long-term survivors (cured or non-relapse individuals, for example). Numerous survival data in medicine and elsewhere have led to a consideration of mixture models which include a proportion of the population's individuals who will never fail (long-term survivors). For

2000 Mathematics Subject Classification: Primary 62F03, 62F10, 62F15.

Key words and phrases: parametric mixture models, random censoring, Kaplan-Meier estimator, maximum likelihood, Markov Chain Monte Carlo, goodness-of-fit.

Received September 20, 2003

references to work in different areas of applications, including medicine, integrated circuit reliability and criminology, see Ghitany [3].

Let  $B_i, 1 \leq i \leq n$ , be a sequence of binary random variables, where  $B_i = 1$  indicates that the  $i$ -th individual from a certain population is at risk to some type of failure and  $B_i = 0$  indicates that there is no risk of failure. Also, assume that  $P(B_i = 1) = p$ . We do not observe  $B_i$ , since we do not know whether an individual is subject to failure or not.

The variable of interest is the time to failure for the individuals of the population. Individuals in the risk group (or susceptible group) may have failure times or not (i.e., censored times) while individuals in the long-term survivors group will always have censored times. Conditional on  $B_i = 1$ , i.e., for individuals in the risk group, we assume that individuals have independent and identically distributed failure times with probability density function (p.d.f.)

$$f_0(x; a, c, k) = \frac{ck}{a} \frac{\left(\frac{x}{a}\right)^{c-1}}{\left[1 + \left(\frac{x}{a}\right)^c\right]^{k+1}}, \quad x > 0, \quad a, c, k > 0. \quad (1)$$

This is known as the generalized Burr XII (GBXII) distribution with cumulative distribution function (c.d.f.)

$$F_0(x; a, c, k) = 1 - \frac{1}{\left[1 + \left(\frac{x}{a}\right)^c\right]^k}, \quad x > 0, \quad a, c, k > 0. \quad (2)$$

The GBXII distribution includes several important distributions as special cases such as the Lomax ( $c = 1$ ), the log-logistic ( $k = 1$ ) and Weibull ( $a = bk^{1/c}, b > 0, k \rightarrow \infty$ ) distributions.

In this paper, we propose the GBXII mixture model with c.d.f.

$$F(t; a, c, k, p) = \begin{cases} pF_0(t; a, c, k), & \text{for } t > 0 \\ 1 - p, & \text{for } t = \infty \end{cases} \quad (3)$$

for estimating the proportion,  $p$ , of the long-term survivors in a population as well as estimating the parameters,  $a, c, k$ , of the failure time distribution for those individuals in the risk group. Both the maximum likelihood (ML) and Markov Chain Monte Carlo (MCMC) methods are used to make inference about the parameters of the GBXII mixture model. Application of the proposed model to survival data of leukaemia patients is given and its goodness-of-fit is demonstrated.

## 2. Maximum Likelihood Estimation

Given the survival data  $\mathbf{D}_n = \{(t_i, \delta_i) : i = 1, 2, \dots, n\}$ , where  $t_i$  is the survival time of individual  $i$  and  $\delta_i$  is the corresponding censor indicator, i.e.,  $\delta_i = 1(0)$  if  $t_i$  is failure (censored), and assuming that the failure distribution for the individuals in the risk group has the GBXII distribution with p.d.f. (1) and c.d.f. (2), the likelihood function is given by

$$L(a, c, k, p | \mathbf{D}_n) = \prod_{i=1}^n [pf_0(t_i; a, c, k)]^{\delta_i} [1 - pF_0(t_i; a, c, k)]^{1-\delta_i}. \quad (4)$$

The maximum likelihood estimates (MLEs)  $\hat{a}, \hat{c}, \hat{k}, \hat{p}$  of the parameters  $a, c, k, p$  can be obtained by maximizing the log-likelihood  $\log L(a, c, k, p | \mathbf{D}_n)$ . Iterative method, such as the Newton-Raphson technique, is needed to find the MLEs of the parameters numerically. Of course, a crucial point for such iterative methods is the initial values of the parameters to obtain the MLEs. For our GBXII mixture model, we use the MATHEMATICA software package to maximize the log-likelihood. Several initial points of the parameters are used to ensure that the obtained estimates are indeed the MLEs.

The asymptotic (observed) variances and covariances of the MLEs  $\hat{a}, \hat{c}, \hat{k}, \hat{p}$  are obtained by inverting the matrix of negative second derivatives of the log-likelihood with respect to the parameters, evaluated at the MLEs of the parameters.

To test for the nested distributions in the GBXII family, e.g., testing the null hypothesis  $H_0 : k = 1$  (log-logistic distribution), we use the

likelihood ratio statistic

$$\Delta_n = -2[\log L(\tilde{\theta} | \mathbf{D}_n) - \log L_n(\hat{\theta} | \mathbf{D}_n)], \quad (5)$$

where  $\hat{\theta}$  is the “unrestricted” MLE of  $\theta$  and  $\tilde{\theta}$  is the “restricted” MLE of  $\theta$  under  $H_0$ . For large samples,  $\Delta_n$  is distributed approximately as chi-square with degrees of freedom equal to the number of restrictions on  $\theta$  in  $H_0$ . Small values of  $\Delta_n$  suggest that  $H_0$  is plausible, whereas large values of  $\Delta_n$  suggest that  $H_0$  is implausible.

### 3. Bayesian Analysis

The application of MCMC techniques to survival models is now wide spread. The MCMC techniques are particularly attractive because they allow for uncertainty associated with the parameters to be explored.

The MCMC method simulates samples from the posterior distribution from which the exact complicated posterior distribution can be approximated. The output chain of the MCMC can be used to mimic a random sample from the required posterior distribution. For further discussion of MCMC methods, (see, e.g., Besag and Green [1], Tierney [8]).

We consider now a Bayesian approach based on the MCMC methodology for approximating the posterior distributions of the parameters of interest from the GBXII mixture model. The parameters  $a, c, k, p$  are considered to be independent with joint density given by

$$\pi(a, c, k, p) = \lambda_0^3 \exp[-\lambda_0(a + c + k)], \quad a, c, k > 0, 0 < p < 1. \quad (6)$$

That is, each of the parameters  $a, c, k$ , follows an exponential distribution with parameter  $\lambda_0 = 2 \times 10^{-8}$ , i.e., exponential prior with large variance, and the parameter  $p$  follows a uniform distribution over  $(0, 1)$ . It follows that the joint posterior density of  $a, c, k, p$  is given by

$$\pi(a, c, k, p | \mathbf{D}_n) \propto \pi(a, c, k, p) L(a, c, k, p | \mathbf{D}_n). \quad (7)$$

In the MCMC methodology, we consider the Gibbs within Metropolis sampler, which requires a complete set of conditional posterior

distributions, i.e.,  $\pi(a|c, k, p, \mathbf{D}_n)$ ,  $\pi(c|a, k, p, \mathbf{D}_n)$ ,  $\pi(k|a, c, p, \mathbf{D}_n)$ ,  $\pi(p|a, c, k, \mathbf{D}_n)$ . Since these conditional posteriors do not have standard distributional forms, the use of Metropolis-Hastings sampler is required.

For our GBXII mixture model, we use the WinBUGS software package to approximate the posterior distributions of the parameters. Our program, using WinBUGS software, is run for 10,000 iterations. The first 1,000 (10%) of the iterations are discarded due to the burn-in time. Convergence is monitored by three different approaches:

- (i) by inspecting the time series plots of the MCMC iterations;
- (ii) by using different runs from overspread starting values and compare the resulting fitted parameters;
- (iii) by using these different runs to calculate what is called the *potential scale reduction factor*,  $R$ , Gelman [2]. The potential scale reduction factor compares the variance between chains and within chain to monitor the convergence. It estimates the factor by which the variance of the summary of interest might be if the simulations were to be continued. Estimated potential scale reduction factor value closer to 1 indicates good convergence.

#### 4. Application

As an application for the proposed generalized Burr XII mixture model, we consider the Kersey *et al.* [6] data on the relapse times (in years) of Group 2 (autologous transplants) of 44 leukaemia patients. The data is given in the first two columns of Table 1. Maller and Zhou [7] reported that the Weibull mixture model provides poor fit for this data.

The MLEs of the parameters  $a$ ,  $c$ ,  $k$ ,  $p$  and their standard errors (in parentheses) of the considered data are given by:

$$\hat{a} = 0.232(0.077), \hat{c} = 3.225(1.036), \hat{k} = 0.649(0.442), \hat{p} = 0.799(0.061).$$

The maximized log-likelihood function is  $-13.490$ .

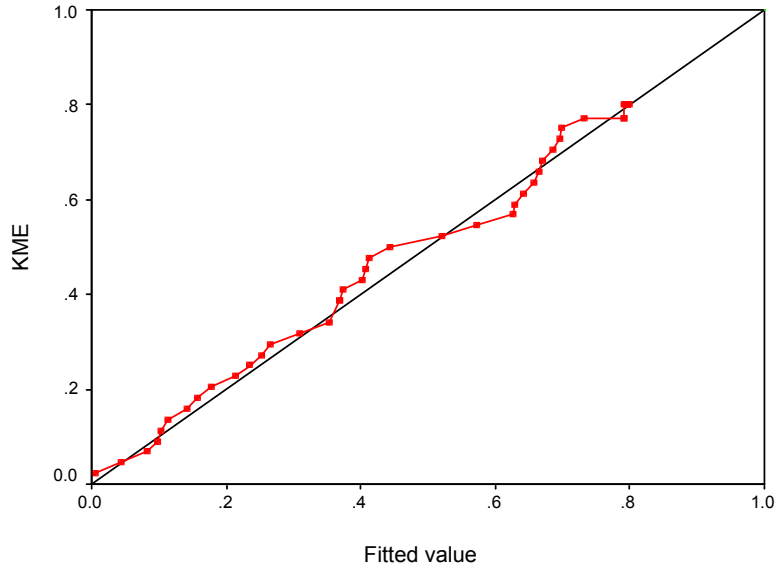
Let  $t_{(1)}, \dots, t_{(n)}$  be the ordered survival times and  $\delta_{(1)}, \dots, \delta_{(n)}$  be the corresponding censor indicators. The Kaplan-Meier estimator (KME),

Kaplan and Meier [5], of a distribution function  $F(t)$  is a non-parametric estimator given by

$$\tilde{F}(t) = 1 - \prod_{i:t_{(i)} \leq t} \left(1 - \frac{\delta_{(i)}}{n - i + 1}\right). \quad (8)$$

See Ghitany *et al.* [4] for using the  $\tilde{F}(t_{(n)})$ , where  $t_{(n)}$  is censored observation, as an estimate for the proportion of long-term survivors.

Figure 1 shows the probability plot of the KME, third column of Table 1, versus the fitted value  $\hat{p}F_0(t; \hat{a}, \hat{c}, \hat{k})$  of the GBXII mixture model, fourth column of Table 1. The correlation coefficient between the KME and the fitted GBXII mixture model is 0.996, judged to be a good fit for the given data.



**Figure 1.** Kaplan-Meier estimator versus fitted value of the generalized Burr XII mixture model.

**Table 1.** Kaplan-Meier estimator (KME) and fitted generalized Burr XII (GBXII) mixture model.

$t_{(i)}$	$\delta_{(i)}$	KME	GBXII mixture	GBXII mixture ( $k = 1$ )
0.0575	1	0.023	0.006	0.010
0.1096	1	0.045	0.043	0.053
0.1370	1	0.068	0.083	0.092
0.1452	1	0.091	0.097	0.106
0.1479	1	0.114	0.102	0.110
0.1534	1	0.136	0.112	0.120
0.1671	1	0.159	0.140	0.146
0.1753	1	0.182	0.158	0.162
0.1836	1	0.205	0.177	0.179
0.2000	1	0.227	0.215	0.214
0.2082	1	0.250	0.234	0.231
0.2164	1	0.273	0.253	0.249
0.2219	1	0.295	0.266	0.261
0.2411	1	0.318	0.310	0.302
0.2603	1	0.341	0.352	0.342
0.2685	1	0.386	0.369	0.359
0.2685	1	0.386	0.369	0.359
0.2712	1	0.409	0.375	0.364
0.2849	1	0.432	0.402	0.391
0.2877	1	0.455	0.407	0.396
0.2904	1	0.477	0.413	0.401
0.3068	1	0.500	0.442	0.431
0.3589	1	0.523	0.521	0.513
0.4027	1	0.545	0.571	0.567
0.4685	1	0.568	0.627	0.629
0.4712	1	0.591	0.629	0.631
0.4904	1	0.614	0.641	0.645
0.5178	1	0.636	0.657	0.662
0.5342	1	0.659	0.665	0.671
0.5452	1	0.682	0.671	0.677
0.5836	1	0.705	0.687	0.695
0.6110	1	0.727	0.697	0.705
0.6137	1	0.750	0.698	0.706
0.7589	1	0.773	0.733	0.744
1.9836	0	0.773	0.790	0.793
1.9973	0	0.773	0.790	0.793
2.0110	1	0.801	0.790	0.793
2.8849	0	0.801	0.795	0.796
2.9973	0	0.801	0.795	0.796
3.2658	0	0.801	0.796	0.796
4.0411	0	0.801	0.797	0.796
4.2055	0	0.801	0.797	0.796
4.2055	0	0.801	0.797	0.796
5.0548	0	0.801	0.798	0.797

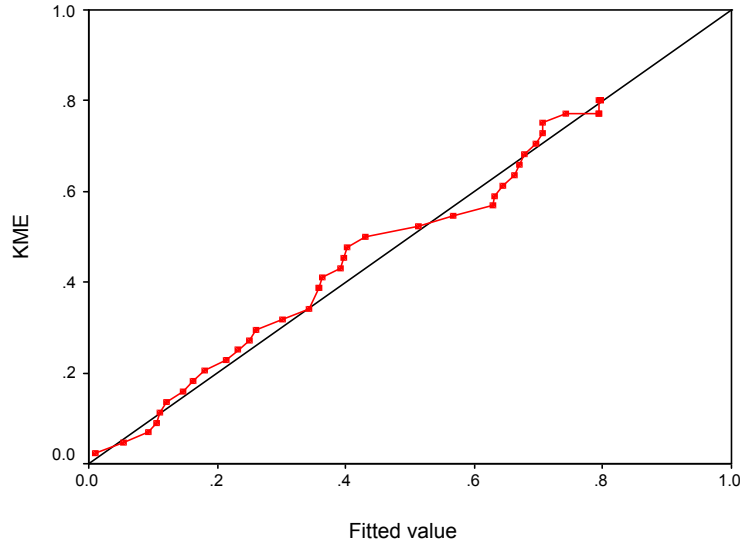
The above MLEs show that  $\hat{c}$  is much greater than 1 ( $\hat{k}$  is small), indicating that the Lomax (Weibull) mixture model is not suitable for the given data. On the other hand,  $k$  is close to 1, suggesting that the log-logistic mixture model may be suitable for such data. To verify such possibility, we test the null hypothesis  $H_0 : k = 1$ . The summary fit in this case is as follows:

$$\tilde{\alpha} = 0.289(0.031), \quad \tilde{c} = 2.727(0.555), \quad \tilde{p} = 0.797(0.061).$$

The maximized log-likelihood function is  $-13.677$ . Hence, the likelihood ratio statistic is

$$\Delta_n = -2[-13.677 - (-13.490)] = 0.374.$$

Which is smaller than the tabulated value  $\chi^2_{1;0.05} = 3.841$ . Hence, we cannot reject  $H_0$ . Figure 2 shows the probability plot of the KME versus the fitted value  $\tilde{p}F_0(t; \tilde{\alpha}, \tilde{c}, 1)$  of the GBXII mixture model with  $k = 1$ , fifth column of Table 1. The correlation coefficient between the KME and the fitted generalized Burr XII model with  $k = 1$  is 0.995, judged to be a good fit for the given data.



**Figure 2.** Kaplan-Meier estimator versus fitted value of the generalized Burr XII mixture model with  $k = 1$ .



We now consider the Bayesian analysis of the given data. Table 2 shows posterior summaries for the GBXII mixture model, including the estimated potential scale reduction,  $\hat{R}$ , which for all parameters is close to 1, indicating good convergence.

**Table 2.** Posterior summaries for the generalized Burr XII mixture model

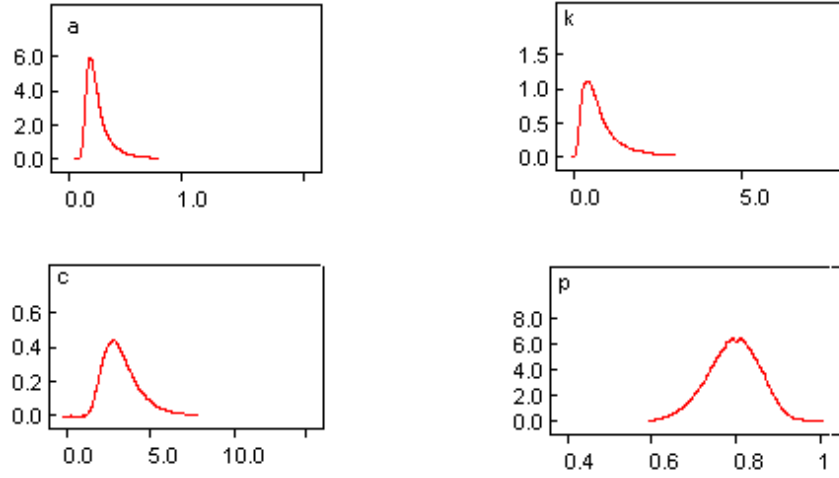
Parameter	Mean	Median	SD	MC error	Credible 95%	$\hat{R}$
$a$	0.279	0.234	0.166	0.003	(0.138, 0.714)	0.980
$c$	3.285	3.082	1.118	0.017	(1.717, 6.021)	1.003
$k$	0.943	0.661	0.969	0.020	(0.180, 3.223)	1.001
$p$	0.792	0.795	0.062	0.000	(0.661, 0.906)	1.001

Estimates of the posterior densities of the parameters  $a, c, k, p$  using the simulated samples are shown in Figure 3 and can be compared with the values in Table 2.

Table 3 shows the posterior summaries for the GBXII mixture model with  $k = 1$ , including the estimated potential scale reduction,  $\hat{R}$ , which for the parameters is close to 1, indicating good convergence.

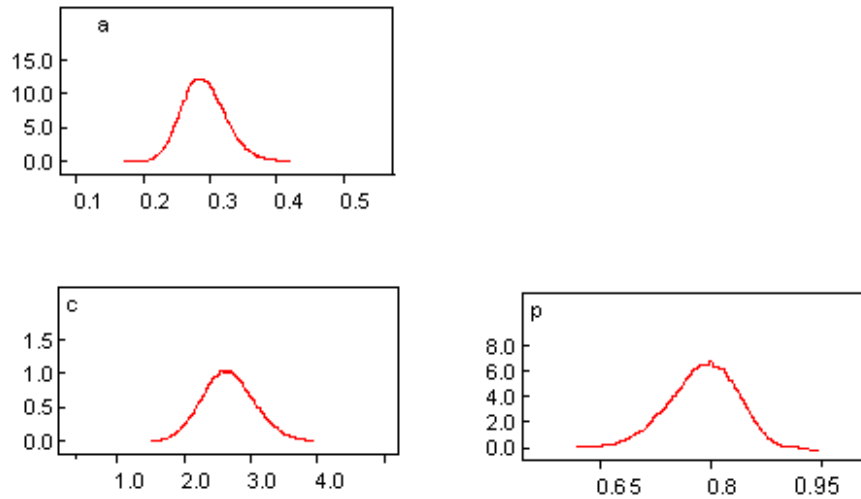
**Table 3.** Posterior summaries for the generalized Burr XII mixture model with  $k = 1$  (log-logistic mixture model)

Parameter	Mean	Median	SD	MC error	Credible 95%	$\hat{R}$
$a$	0.292	0.290	0.033	0.000	(0.232, 0.362)	1.003
$c$	2.668	2.653	0.394	0.003	(1.942, 3.486)	1.004
$p$	0.785	0.789	0.061	0.001	(0.656, 0.891)	1.003



**Figure 3.** Approximate marginal posterior densities for  $a, c, k, p$  of the generalized Burr XII mixture model.

Estimates of the posterior densities of the parameters  $a, c, p$  using the simulated samples are shown in Figure 4 and can be compared with the values in Table 3.



**Figure 4.** Approximate marginal posterior densities for  $a, c, p$  of the generalized Burr XII mixture model with  $k = 1$ .

## 5. Conclusion

It is shown that the GBXII mixture model can be useful for modeling survival data with long-term survivors. Also, for the data set considered in this paper, it is shown that the log-logistic mixture model (GBXII mixture model with  $k = 1$ ) provides almost equal good fit for the data as the full GBXII mixture model. Finally, the good fit of the GBXII mixture model (with  $k = 1$  or not) can be attributed to the fact that the MLE of  $c$  is greater than 1, implying that the estimated hazard rate function of the GBXII failure model, i.e., the estimated function

$$h_0(x; a, c, k) = \frac{f_0(x; a, c, k)}{1 - F_0(x; a, c, k)} = \frac{ck}{a} \frac{\left(\frac{x}{a}\right)^{c-1}}{1 + \left(\frac{x}{a}\right)^c}, \quad x > 0, a, c, k > 0$$

is upside-down bathtub shaped, i.e., increases initially and then decreases. This hazard rate feature is not covered by the usual Weibull failure model.

## Acknowledgement

The authors are grateful to Professor Ross Maller for helpful discussions.

## References

- [1] J. Besag and P. J. Green, Spatial statistics and Bayesian computation, J. Roy. Statist. Soc. B 55 (1993), 25-37.
- [2] A. Gelman, Inference and monitoring convergence, W. R. Gilks, et al., eds., Markov Chain Monte Carlo in Practice, Chapman and Hall, London, 1996, pp. 131-143.
- [3] M. E. Ghitany, On the information matrix of the exponential mixture models with long-term survivors, Biometrical J. 35 (1993), 15-27.
- [4] M. E. Ghitany, R. Maller and S. Zhou, Estimating the proportion of immunes in censored samples: a simulation study, Statistics in Medicine 14 (1995), 39-49.
- [5] E. L. Kaplan and P. Meier, Nonparametric estimation from incomplete observations, J. Amer. Stat. Assoc. 53 (1958), 457-481.
- [6] J. H. Kersey, D. Wiedorf, M. E. Nesbit, T. W. LeBien, W. G. Woods, P. B. McGlave, T. Kim, D. A. Vallera, A. I. Goldman, B. Bostrom, D. Hurd and N. K. C. Ramsay,

Comparison of autologous and allogeneic bone marrow transplantation for treatment of high-risk refractory acute lymphoblastic leukaemia, *New England J. Medicine* 317 (1987), 461-467.

- [7] R. A. Maller and S. Zhou, Testing for the presence of immune or cured individuals in censored survival data, *Biometrics* 51 (1995), 1197-1205.
- [8] L. Tierney, Markov chains for exploring posterior distributions, *Ann. Statist.* 22 (1994), 1701-1762.

