

ESTIMATION OF HAZARD AND SURVIVAL FUNCTION FOR COMPETING RISKS USING KERNEL AND MIXTURE MODEL IN BIMODAL SETUP

¹A. M. Rangoli, ²A. S. Talawar

•

¹Research Scholar, Department of Statistics, Karnatak University, Dharwad. India.

ajmr1008@gmail.com

²Professor, Department of Statistics, Karnatak University, Dharwad. India.

astalawar@kud.ac.in

Abstract

Aim of the present paper is to find suitable model for bimodal data. We have modelled mixture of two Weibull distributions in the presence of competing risks and also used Epanechnikov kernel to estimate hazard and survival functions. We considered prostate cancer data for application of the mixture model and kernel. We used maximum likelihood estimation (MLE) to estimate parameters of the mixture model, as the equations have no closed form, so we considered expectation-maximization (EM) algorithm. The mixture model and kernel gave good fit to the bimodal data. The prostate cancer data consists of three causes, we have estimated hazard function for these three causes using mixture model and kernel. The asymptotic confidence interval for the parameters of mixture model to all three causes were estimated. Also compared survival curve of mixture model with kernel and Kaplan-Meier survival curves for all the three causes.

Keywords: Weibull mixture model, EM algorithm, kernel, hazard, bimodal.

I. Introduction

The general Statistical analyses were different from survival analysis because of the presence of censoring. Basically, censoring means incomplete data. In survival analysis or medical studies, it is quite common that more than one cause of failure may be directed to a subject at the same time. It is required for an investigator to estimate a specific risk in the presence of other risk factors. In statistical literature, this process is known as the analysis of competing risks model. It is assumed, in the analysis of competing risks model, that data consist of a time to failure and an indicator denoting the cause of failure. In survival analysis our main objective is to estimate survival and hazard functions. Survival analysis can be done using parametric, non-parametric and Bayesian methods. For parametric approach we generally consider Weibull distribution because it has increasing, decreasing and constant hazard rate, but this distribution can be used when data is

unimodal. When we have bimodal data we cannot use the standard parametric lifetime distributions, in that case we can use mixture of distributions. For nonparametric approach we can consider the kernel method to estimate the hazard and survival functions.

Many authors work on mixture of distributions and kernel based methods to estimate hazard and survival functions. Modelling of mixture of gamma, mixture of Weibull and mixture of log normal distributions for analysing the heterogeneous survival data was considered by [1], for that they have considered mice data and Lung cancer dataset. The mixture of two and three Weibull distributions was modelled and estimated the parameters using MLE and tested for best fit of the models by [2], and used five different examples to show the hazard and survival functions. A parametric mixture model of three different distributions was used to analyse heterogeneous survival data by [3]. They have simulated the data and estimated the parameters using expectation-maximization (EM) algorithm and also compared individual distribution like exponential, gamma and Weibull with the mixture of these three distributions. Similarly, many authors have worked on mixture models ([4], [5], [6]). Estimation of the hazard function and its associated factors in gastric cancer patients using Wavelet and kernel smoothing methods was carried out by [7]. Repeated time to event models to characterize the repeated occurrence of clinical events and visualization of kernel based hazard with comparison to Weibull and Gompertz models was considered by [8]. (see also [9], [10]).

In present paper, we are considering estimation of density, hazard and survival functions using Epanechnikov kernel and mixture of two Weibull distributions. For estimation we are consider the prostate cancer data which is bimodal given in [11]. Generally, for bimodal data it is not appropriate to use standard parametric lifetime distributions but mixtures of those distributions are suitable for bimodal. Here we are considering two cases, case-I consists of estimation of hazard and survival function using mixture of two Weibull distributions and case-II considers kernel method of estimation.

II. Methods

2.1 Case-I: Mixture of two Weibull distribution

Now we are considering a parametric approach using mixture of two Weibull distributions. The study considers fitting of bimodal data to the mixture of Weibull distributions in presence of competing risks and calculation of the hazard and survival functions. The functional form of mixture distribution is given below.

Let T_1, T_2, \dots, T_n be the failure time of n patients where $T_i \in (0, t], i = 1, 2, \dots, n$ if we consider k competing events then $T_{ij} \rightarrow$ failure time of i^{th} patient with j^{th} cause. Each patient fail due to only one cause $T_i = \min(T_{i1}, T_{i2}, \dots, T_{ik})$. Let us consider C be the censoring time such that $T_{ij} = \min(T_{ij}, C)$. Let $F(t)$ be the cumulative distribution function (CDF) and $f(t)$ be probability distribution function and $h(t)$ and $S(t)$ be hazard and survival function at time t .

$$F(t) = 1 - (\pi_1 e^{-\alpha t^\gamma} + \pi_2 e^{-\beta t^\lambda}) \quad (1)$$

$$f(t) = \pi_1 \alpha \gamma t^{\gamma-1} e^{-\alpha t^\gamma} + \pi_2 \beta \lambda t^{\lambda-1} e^{-\beta t^\lambda} \quad (2)$$

$$S(t) = 1 - F(t)$$

$$S(t) = \pi_1 e^{-\alpha t^\gamma} + \pi_2 e^{-\beta t^\lambda} \quad (3)$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{\pi_1 \alpha \gamma t^{\gamma-1} e^{-\alpha t^\gamma} + \pi_2 \beta \lambda t^{\lambda-1} e^{-\beta t^\lambda}}{\pi_1 e^{-\alpha t^\gamma} + \pi_2 e^{-\beta t^\lambda}} \quad (4)$$

Here π_1 and π_2 be the weights such that $\pi_1 + \pi_2 = 1$

And α and γ be scale and shape parameters with weight π_1 and β and λ be scale and shape parameters with weight π_2 .

Likelihood function L is given as

$$L = \prod_{i=1}^n f(x_i)$$

$$L = \prod_{i=1}^n (\pi_1 \alpha \gamma t_i^{\gamma-1} e^{-\alpha t_i^\gamma} + \pi_2 \beta \lambda t_i^{\lambda-1} e^{-\beta t_i^\lambda})$$

Now the likelihood function in terms of competing risks can be given as

$$\delta_{ij} = \begin{cases} 1 & i^{\text{th}} \text{ subject fail due to } j^{\text{th}} \text{ cause, } j = 1, 2, \dots, k \\ 0 & i^{\text{th}} \text{ subject does not fail due to } j^{\text{th}} \text{ cause (Censored)} \end{cases}$$

$$L = \prod_{i=1}^n \prod_{j=1}^k (f_j(t_i))^{\delta_{ij}} S(t_i)^{1-\delta_{ij}}$$

$$L = \prod_{i=1}^n \prod_{j=1}^k (h_j(t_i))^{\delta_{ij}} S(t_i) \quad (5)$$

Now the log likelihood of equation (5) can be written as,

$$l = \log L = \sum_{i=1}^n \sum_{j=1}^k \left((\delta_{ij} * \log(h_j(t_i))) + \log(S(t_i)) \right)$$

$$l = \sum_{i=1}^n \sum_{j=1}^k \left(\left(\delta_{ij} * \log \left(\frac{\pi_1 \alpha_j \gamma_j t_i^{\gamma_j-1} e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 \beta_j \lambda_j t_i^{\lambda_j-1} e^{-\beta_j t_i^{\lambda_j}}}{\pi_1 e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 e^{-\beta_j t_i^{\lambda_j}}} \right) \right) + \log \left(\pi_1 e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 e^{-\beta_j t_i^{\lambda_j}} \right) \right)$$

$$l = \sum_{i=1}^n \sum_{j=1}^k \left(\left(\delta_{ij} \left(\log \left(\pi_1 \alpha_j \gamma_j t_i^{\gamma_j-1} e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 \beta_j \lambda_j t_i^{\lambda_j-1} e^{-\beta_j t_i^{\lambda_j}} \right) - \log \left(\pi_1 \alpha_j \gamma_j t_i^{\gamma_j-1} e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 \beta_j \lambda_j t_i^{\lambda_j-1} e^{-\beta_j t_i^{\lambda_j}} \right) \right) \right) + \log \left(\pi_1 e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 e^{-\beta_j t_i^{\lambda_j}} \right) \right) \quad (6)$$

For cause j we can write log likelihood as

$$l_j = \sum_{i=1}^{n_j} \left(\log \left(\pi_1 \alpha_j \gamma_j t_i^{\gamma_j-1} e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 \beta_j \lambda_j t_i^{\lambda_j-1} e^{-\beta_j t_i^{\lambda_j}} \right) \right)$$

$$- \sum_{i=1}^{n_j} \log \left(\pi_1 \alpha_j \gamma_j t_i^{\gamma_j-1} e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 \beta_j \lambda_j t_i^{\lambda_j-1} e^{-\beta_j t_i^{\lambda_j}} \right) + \sum_{i=1}^{n_j} \log \left(\pi_1 e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 e^{-\beta_j t_i^{\lambda_j}} \right) \quad (7)$$

Now estimating the parameter values using MLE, and it is obtained by first order partial derivatives with respect to each parameter and equating to zero. But these equations do not have closed form, so to estimate the parameter values we consider numerical estimation that is Newton-Raphson method or we can use expectation-maximization (EM) algorithm. The first order partial derivatives are given in Appendix.

2.1.1 Expectation–Maximization (EM) Algorithm

The EM algorithm is a powerful iterative method used for finding maximum likelihood estimates or maximum a posteriori estimates in statistical models where the data is incomplete or contains latent variables. The EM algorithm consists of an-expectation step (E-step), and a maximization step (M-step). The advantage of the EM algorithm is that it solves a difficult incomplete-data problem by constructing two easy steps. The E-step only needs to compute the conditional expectation of the log-likelihood with respect to the incomplete data, given the observed data. The M-step needs to find the maximizer of this expected likelihood. An additional advantage of this method compared to other optimization techniques is that it is very simple, and it converges reliably [12]. Let Y be the observed data and X be the missing data. We also write l , l_c and l_m for the log-likelihoods based on the observed, complete and missing data distributions respectively. The EM algorithm consists of iterating two steps. First is the expectation, or “E”, step, in which an objective function is constructed from the complete data likelihood. Second is the maximization, or “M”, step, in which the previously computed objective function is maximized. These two steps are then alternated until some convergence criterion is met [13]. Whatever value of θ the algorithm converges to and is used as our parameter estimate.

The E-step of the EM algorithm consists of computing the conditional expectation of the complete data likelihood, given the observed data. That is, the objective function at iteration k is given by

$$Q(\theta|\theta_{k-1}) = E_{\theta_{k-1}}(l_c(\theta; y, X)|Y = y) \quad (8)$$

Where θ_{k-1} is the parameter estimate obtained from the previous iteration.

The M-step of the EM algorithm consists of maximizing the objective function constructed in the previous E-step. That is, we define $\theta_k = \operatorname{argmax}_{\theta} Q(\theta|\theta_{k-1})$. Typically, this optimization must be performed numerically via, e.g., gradient ascent or the Newton-Raphson algorithm. In fact, it is possible to divide the set of parameters into groups (possibly with each group containing a single parameter) and optimize over each group individually with the others held fixed. This is called the Expectation-Conditional Maximization, or ECM, algorithm. Notationally, we can combine the E and M-steps of the EM algorithm into a single “update function”. We write $M(\theta_{k-1}) = \operatorname{argmax}_{\theta} Q(\theta|\theta_{k-1})$. The EM algorithm can thus be viewed as the iterative application of this update function, M .

2.1.2 Asymptotic Confidence Bounds

The MLE’s do not have closed form to know the distribution to calculate confidence intervals, in such a case we go with asymptotic distribution of the MLE of the parameters [14]. It is known that the asymptotic distribution of the MLE $\hat{\theta}$ is given by

$$(\hat{\theta} - \theta) \rightarrow N_4(0, I^{-1}(\theta))$$

Where $I^{-1}(\theta) \rightarrow$ Fisher information matrix of the unknown parameters

$$\theta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \lambda_1, \lambda_2, \lambda_3) .$$

The elements of the 4 X 4 matrix of $I^{-1}(\cdot)$, are approximated by $I_{ij}(\hat{\theta})$,

where

$$I_{ij}(\hat{\theta}) = - \frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta = \hat{\theta}}$$

Where, $\hat{\theta} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3)$ estimated parameters.

Now information matrix can be written as,

$$\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j} = \begin{bmatrix} \frac{\partial^2 l}{\partial \alpha_j^2} & \frac{\partial^2 l}{\partial \alpha_j \partial \beta_j} & \frac{\partial^2 l}{\partial \alpha_j \partial \gamma_j} & \frac{\partial^2 l}{\partial \alpha_j \partial \lambda_j} \\ \frac{\partial^2 l}{\partial \alpha_j \partial \beta_j} & \frac{\partial^2 l}{\partial \beta_j^2} & \frac{\partial^2 l}{\partial \beta_j \partial \gamma_j} & \frac{\partial^2 l}{\partial \beta_j \partial \lambda_j} \\ \frac{\partial^2 l}{\partial \alpha_j \partial \gamma_j} & \frac{\partial^2 l}{\partial \beta_j \partial \gamma_j} & \frac{\partial^2 l}{\partial \gamma_j^2} & \frac{\partial^2 l}{\partial \gamma_j \partial \lambda_j} \\ \frac{\partial^2 l}{\partial \alpha_j \partial \lambda_j} & \frac{\partial^2 l}{\partial \beta_j \partial \lambda_j} & \frac{\partial^2 l}{\partial \gamma_j \partial \lambda_j} & \frac{\partial^2 l}{\partial \lambda_j^2} \end{bmatrix} \quad (9)$$

The elements of the fisher information matrix are given in Appendix.

Therefore, the approximate $100(1 - \gamma)\%$ two-sided, confidence interval for θ is given by

$$\hat{\theta} \pm Z_{\gamma/2} \sqrt{I^{-1}(\hat{\theta})} \quad (10)$$

Here $Z_{\gamma/2}$ is the upper $\gamma/2$ th percentile of a standard normal distribution.

2.2 Case-II: Kernel density Estimation

A kernel is a weight function of observation on x and scaling parameter h which is called as the Bandwidth. The scaled distances obtained at a point x is used to compute kernel density at that point. The kernel density function is regarded as probability density [8]. The estimator to estimate density is given by,

$$\hat{f}(u) = \frac{1}{n} \sum_{i=1}^n K(u)$$

Where, $K(\cdot) \rightarrow$ Kernel Function

The kernel function has the following properties viz.,

$$K(u) \geq 0, \text{ for all } u$$

$$\int K(u) du = 1 \text{ (normalization)}$$

$$K(-u) = K(u) \text{ (symmetry)}$$

$$\int u K(u) du = 0$$

$$\text{And } \int u^2 K(u) \neq 0$$

Using the kernels we can estimate the survival and hazard functions. In general, to obtaining pattern for rate of failure the hazard curve is more obvious than survival curve. Hazard rate functions can be used for several statistical analysis in medicine, engineering and economics. For instance, hazard function commonly used when presenting results in clinical trials involving survival data. Several methods for hazard function estimation have been considered in the literature

[7], [9],[10]). Hazard function estimation by nonparametric methods has an advantage in flexibility because no formal assumptions are made about the mechanism that generates the sample order or the randomness [15]. There are many kernels in the literature say, Uniform, Triangle, Epanechnikov, Quartic, Triweight, Gaussian, Cosine etc. Now for our study we consider the Epanechnikov kernel [16].

$$K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$$

By considering $u = \frac{t-X_i}{b}$

Then kernel becomes

$$K_b(t) = \frac{3}{4b} \left(1 - \left(\frac{t-X_i}{b} \right)^2 \right) I\left(\left| \frac{t-X_i}{b} \right| \leq 1 \right) \quad (11)$$

Where, b is the bandwidth, n is the number of observation, X_i is the given observation and t is the point where kernels are calculated.

From this kernel many bumps are formed and summing the bumps gives us the density function.

The kernel density function is given by,

$$\hat{f}_b(t) = \frac{1}{nb} \sum_{i=1}^n K\left(\frac{t-X_i}{b}\right)$$

$$\hat{f}_b(t) = \frac{1}{nb} \sum_{i=1}^n \frac{3}{4} \left(1 - \left(\frac{t-X_i}{b} \right)^2 \right) I\left(\left| \frac{t-X_i}{b} \right| \leq 1 \right)$$

The estimation of the CDF, \widehat{F}_b is constructed by integrating \hat{f}_b . That is

$$\widehat{F}_b(t) = \int_{-\infty}^t \hat{f}_b(x) dx = \frac{1}{n} \sum_{i=1}^n \bar{K}\left(\frac{t-X_i}{b}\right)$$

Where, $\bar{K}(t) = \int_{-\infty}^t K(x) dx$

Estimation of hazard function using kernel [17],

$$\hat{h}(t) = \frac{1}{b} \sum_{i=1}^n K\left(\frac{t-X_i}{b}\right) \Delta \widehat{\Lambda}(t_i) \quad (12)$$

Where, n is the number of failure times, b is the bandwidth, $\widehat{\Lambda}(t)$ is the Nelson-Aalen estimator of the cumulative hazard function.

2.2.1 Nelson-Aalen estimator of cumulative hazard function

Let the hazard function be,

$$h(t) = \lim_{l \rightarrow 0} \frac{1}{l} P(t \leq T < t+l | T \geq t) = \frac{f(t)}{S(t)}$$

where $f(t)$ be density function and $S(t)$ be survival function, and the survival function in terms of hazard function can be expressed as,

$$S(t) = e^{-\int_0^t h(u) du}$$

Now the cause specific hazard function be given by,

$$h_j(t) = \lim_{l \rightarrow 0} \frac{1}{l} p(t \leq T < t + l, J = j / T \geq t)$$

$$h(t) = \sum_{j=1}^J h_j(t)$$

And the cumulative cause specific hazard function be given by

$$\Lambda(t) = \int_0^t h(u) du$$

And cause specific Nelson-Aalen estimator of the cumulative hazard [18] is given by,

$$\Lambda_j(t) = \sum_{k=1}^K \frac{\text{Number of individuals observed to fail due to cause } j \text{ at } t_k}{\text{Number of individuals at risk just prior to } t_k}$$

2.2.2 Selection of the Bandwidth

Important thing in the kernel density estimation is selection of the bandwidth. We calculate bandwidth using Silverman's Rule [19]. That is

$$\hat{h} = \frac{1.06 \cdot \hat{\sigma}}{n^{\frac{1}{5}}} \quad (13)$$

Where, $\hat{\sigma}$ is the sample standard deviation.

2.2.3 Kaplan-Meier (K-M) Estimator

The Kaplan-Meier estimator known as the product limit estimator is a non-parametric statistic used to estimate the survival function from lifetime data [20]. An important benefit of the Kaplan-Meier curve is that, the method can take into account some types of censored data, particularly right-censoring, which occurs if a patient withdraws from a study, or is lost due to follow-up, or is alive without event incidence at last follow-up. The Kaplan-Meier estimate is an easiest way of computing survival over time. The Kaplan Meier estimator of survival function is defined as

$$\hat{S}(t) = \prod_{i: t_i < t} \left(1 - \frac{d_i}{n_i}\right)$$

Where t_i is the failure time, d_i is the number of events that occurs at time t_i and n_i is the number individuals at risk of experiencing the event immediately prior to t_i .

III. Results

The prostate cancer data consists of 489 patients, 25.5% of them are failed due to cancer, 19% failed due to CVD and 25.74% failed due to other causes and rest of the data were censored. Median failure

time for cancer patients is 23 months and for CVD patients 20.5 months and for other causes 24 months.

From figure 1 we can see that our data is bimodal, to fit this data, we used mixture of two Weibull distributions (black line) and the kernel density estimation (red line). Figure 2 explains the hazard curves for three different causes cancer, CVD and other causes. Here we can see that for cause cancer and CVD the hazard initially increases till 30 months then decreases. For other causes the hazard increases-decreases-increases, so we can say that, the hazard function is non-monotonic. Figure 3 explains the survival curve of three causes using kernel, mixture model and Kaplan-Meier survival functions. Here we can observe that kernel and mixture model survival curves are close to each other. For all three causes Kaplan-Meier survival curve has less probability of surviving as compared to the kernel and mixture model. Table 1 shows the estimated parameter values using MLE by considering EM algorithm. Table 2 gives the estimated parameter values with their corresponding standard error and confidence limits.

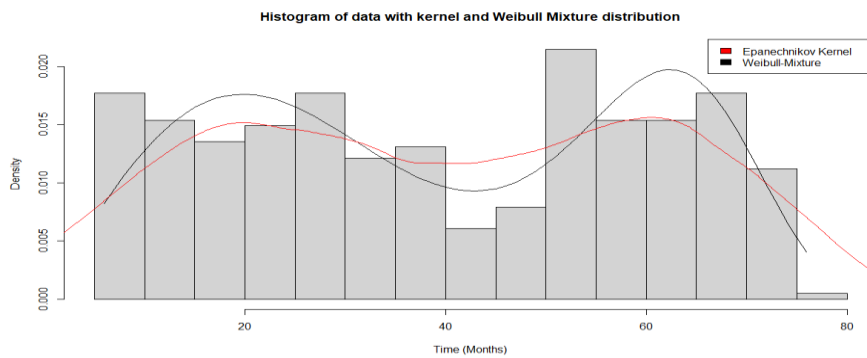


Figure 1: Histogram of the data with fitted Mixture of Weibull distribution and Epanechnikov kernel.

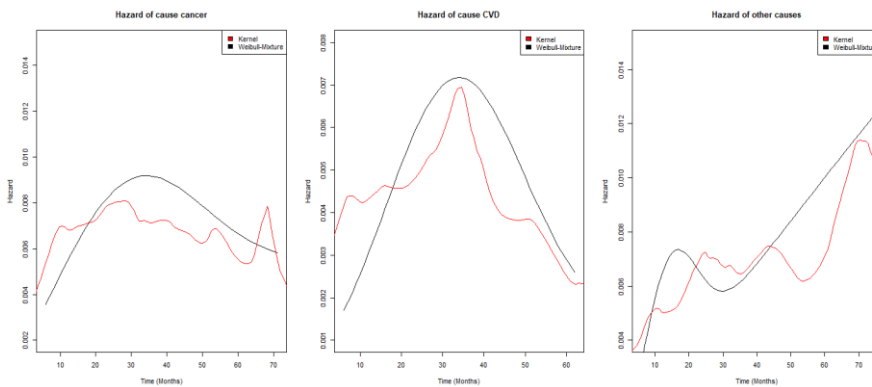


Figure 2: Hazard curves using mixture model and kernel for three causes

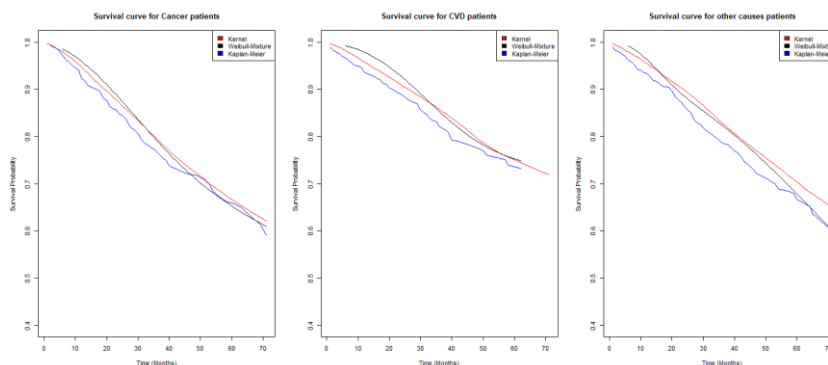


Figure 3: Survival curves using mixture model, kernel and Kaplan-Meier for three causes

Table 1: Estimated parameter values for three causes

Parameter	Cancer	CVD	Other Causes
π_1	0.8155268	0.810495	0.9224448
π_2	0.1844732	0.1895035	0.07755524
α	0.001494359	0.001097538	0.000103001
β	0.000370462	0.000076065	0.000848062
γ	1.241094000	1.058928000	1.954219000
λ	2.180196000	2.600613000	2.469943000

Table 2: Estimated parameter values, standard error (SE), lower control limit (LCL) and upper control limit (UCL)

Causes	Parameters	SE	LCL	UCL
Cancer	$\alpha = 0.001494359$	0.000059337	0.001378061	0.001610657
	$\beta = 0.000370462$	0.000006754	0.000357225	0.000383700
	$\gamma = 1.241094000$	0.004894528	1.231500902	1.250687098
	$\lambda = 2.180196000$	0.000420388	2.179372055	2.181019945
CVD	$\alpha = 0.001097538$	0.000139156	0.000824797	0.001370279
	$\beta = 0.000076065$	0.000001373	0.000073374	0.000078756
	$\gamma = 1.058928000$	0.000000469	1.058927081	1.058928919
	$\lambda = 2.600613000$	0.000187876	2.600244771	2.600981229
Other causes	$\alpha = 0.000103001$	0.000003077	0.000096969	0.000109032
	$\beta = 0.000848062$	0.000022916	0.000803148	0.000892976
	$\gamma = 1.954219000$	0.000666012	1.952913641	1.955524359
	$\lambda = 2.469943000$	0.011053927	2.448277701	2.491608299

IV. Discussions

From the study we conclude that, in real life situations with competing risks data, if data is bimodal we can use mixture of distributions or kernel methods to estimate hazard and survival functions. In this paper we have considered both approach to estimate hazard and survival function in presence of the competing risks. To estimate the parameters of the mixture models we have used MLE as it does not have closed form so we considered EM algorithm to estimate parameters of all three causes. We have also calculated the standard error and asymptotic confidence interval for all the parameters. All the estimated parameters are statistically significant at 5% level of significance. Here we can see that hazard function initially increases, then decreases and increases. For survival curve we have compared kernel, mixture model and Kaplan-Meier methods. So, when we have bimodal density, and having competing risks approach, the mixture model (as a parametric approach) is more appropriate or kernel method (as a nonparametric approach) is more appropriate to estimate hazard and survival functions.

Acknowledgment: The first author is thankful to Department of Science and Technology, innovation in science pursuit for inspired research (DST-INSPIRE) for financial support (Fellowship/2021/210203).

References

- [1] Erişoğlu, Ü., Erişoğlu, M., and Erol, H. (2011). A mixture model of two different distributions approach to the analysis of heterogeneous survival data. *International Journal of Computational and Mathematical Sciences*, 5(2), 75-79.
- [2] Razali, A. M., and Al-Wakeel, A. A. (2013). Mixture Weibull distributions for fitting failure times data. *Applied Mathematics and Computation*, 219(24), 11358-11364.
- [3] Mohammed, Y. A., Yatim, B., and Ismail, S. (2013). A simulation study of a parametric mixture model of three different distributions to analyze heterogeneous survival data. *Modern Applied Science*, 7(7), 1-9.
- [4] Elmahdy, E. E. (2015). A new approach for Weibull modeling for reliability life data analysis. *Applied Mathematics and computation*, 250, 708-720.
- [5] Larson, M. G., and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 34(3), 201-211.
- [6] Enogwe, S. U., Okereke, E. W., and Ibeh, G. C (2023). A Bimodal Extension of Suja Distribution with Applications. *Statistics and Applications* 21(2), pp 155-173.
- [7] Ahmadi, A., Roudbari, M., Gohari, M. R., and Hosseini, B. (2012). Estimation of hazard function and its associated factors in gastric cancer patients using wavelet and kernel smoothing methods. *Asian Pacific journal of cancer prevention*, 13(11), 5643-5646.
- [8] Goulooze, S. C., Väitalo, P. A., Knibbe, C. A., and Krekels, E. H. (2018). Kernel-based visual hazard comparison (kbVHC): a simulation-free diagnostic for parametric repeated time-to-event models. *The AAPS journal*, 20, 1-11.
- [9] Hess, K. R., Serachitopol, D. M., and Brown, B. W. (1999). Hazard function estimators: a simulation study. *Statistics in medicine*, 18(22), 3075-3088.
- [10] Klein, J. P., and Bajorunaite, R. (2003). Inference for competing risks. *Handbook of statistics*, 23, 291-311.
- [11] Andrews, D. F., and Herzberg, A. M. (2012). *Data: a collection of problems from many fields for the student and research worker*. Springer Science and Business Media.
- [12] Park, C. (2005). Parameter estimation of incomplete data in competing risks using the EM algorithm. *IEEE Transactions on Reliability*, 54(2), 282-290.
- [13] Ruth, W. (2024). A review of Monte Carlo-based versions of the EM algorithm. arXiv preprint arXiv:2401.00945.
- [14] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York.
- [15] Klein, J. P., and Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data* (Vol. 1230). New York: Springer.
- [16] Guedes, D. G. P., Cunha, E. E., & Lima, G. F. C. (2017). Genetic evaluation of age at first calving from Brown Swiss cows through survival analysis. *Archivos de zootecnia*, 66(254), 247-255.
- [17] Heisey, D. M., and Patterson, B. R. (2006). A review of methods to estimate cause-specific mortality in presence of competing risks. *The Journal of Wildlife Management*, 70(6), 1544-1555.
- [18] Beyersmann, J., Allignol, A., and Schumacher, M. (2011). *Competing risks and multistate models with R*. Springer Science and Business Media.
- [19] Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.
- [20] Kaplan, E. L. and Meier, P. (1958). Non-parametric estimation from incomplete observation. *Journal of American Statistical Association*, 53, 457-481.

Appendix:

The log likelihood function (7) for the mixture of two Weibull distribution in presence of competing risks is given as

$$l_j = \sum_{i=1}^{n_j} \left(\log \left(\pi_1 \alpha_j \gamma_j t_i^{\gamma_j-1} e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 \beta_j \lambda_j t_i^{\lambda_j-1} e^{-\beta_j t_i^{\lambda_j}} \right) \right) - \sum_{i=1}^{n_j} \log \left(\pi_1 \alpha_j \gamma_j t_i^{\gamma_j-1} e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 \beta_j \lambda_j t_i^{\lambda_j-1} e^{-\beta_j t_i^{\lambda_j}} \right) + \sum_{i=1}^n \log \left(\pi_1 e^{-\alpha_j t_i^{\gamma_j}} + \pi_2 e^{-\beta_j t_i^{\lambda_j}} \right)$$

Here we are considering first with cause 1, that is G1 stands for failure times for cause 1, so here $j = 1$. Similarly, we can consider cause 2 and 3 as G2 and G3 respectively.

$$ea = e^{-(\alpha*(G1^\gamma))}; eb = e^{-(\beta*(G1^\lambda))}; eaa = e^{-(\alpha*(x^\gamma))}; ebb = e^{-(\beta*(x^\lambda))};$$

$$\log x2 = \log(G1^2); x2g = G1^{2*\gamma}; x2l = G1^{2*\lambda}; xg1 = G1^{\gamma-1}; xl1 = G1^{\lambda-1}$$

$$deno1 = (p1 * \alpha * \gamma * xg1 * ea) + (p2 * \beta * \lambda * xl1 * eb)$$

$$deno2 = p1 * ea + p2 * eb$$

$$deno3 = p1 * eaa + p2 * ebb$$

$$nume1 = p1 * \gamma * xg1 * ea * (1 - \alpha * (G1^\gamma))$$

$$nume2 = p1 * (G1^\gamma) * ea$$

$$nume3 = p1 * (x^\gamma) * eaa$$

$$numeb1 = p2 * \lambda * xl1 * eb * (1 - \beta * (G1^\lambda))$$

$$numeb2 = p2 * (G1^\lambda) * eb$$

$$numeb3 = p2 * (x^\lambda) * ebb$$

$$numeg1 = p1 * \alpha * \gamma * xg1 * ea * (\gamma * \log(G1) + 1 - \alpha * \gamma * (G1^\gamma) * \log(G1))$$

$$numeg2 = p1 * \alpha * (G1^\gamma) * \log(G1) * ea$$

$$numeg3 = p1 * \alpha * (x^\gamma) * \log(x) * eaa$$

$$numel1 = p2 * \beta * \lambda * xl1 * eb * (\lambda * \log(G1) + 1 - \beta * \lambda * (G1^\lambda) * \log(G1))$$

$$numel2 = p2 * \beta * (G1^\lambda) * \log(G1) * eb$$

$$numel3 = p2 * \beta * (x^\lambda) * \log(x) * ebb$$

$$numea1 = p1 * \gamma * x2g * \alpha * ea$$

$$numea2 = p1 * \gamma * xg1 * ea$$

$$numea3 = p1 * (G1^\gamma) * ea$$

$$numea4 = p1 * (x^\gamma) * eaa$$

$$numeb11 = p2 * \lambda * x2l * \beta * eb$$

$$numeb22 = p2 * \lambda * xl1 * eb$$

$$numeb33 = p2 * (G1^\lambda) * eb$$

$$numeb44 = p2 * (x^\lambda) * ebb$$

$$\frac{\partial \log l_j}{\partial \alpha_j} = \sum_{i=1}^{n_j} \left(\frac{nume1}{deno1} \right) + \sum_{i=1}^{n_j} \left(\frac{nume2}{deno2} \right) - \sum_{i=1}^{n_j} \left(\frac{nume3}{deno3} \right)$$

$$\begin{aligned} \frac{\partial^2 \log l_j}{\partial \alpha_j^2} &= \sum_{i=1}^{n_j} \left(\frac{(deno1 * p1 * \gamma * x2g * ea * (-\alpha - 2)) - (nume1 * nume1)}{deno1^2} \right) - \\ &\quad \sum_{i=1}^{n_j} \left(\frac{(deno2 * p1 * ea * (-G1^{2*\gamma})) - (nume2 * nume2)}{deno2^2} \right) + \\ &\quad \sum_{i=1}^n \left(\frac{(deno3 * p1 * eaa * (-x^{2*\gamma})) - (nume3 * nume3)}{deno3^2} \right) \\ \frac{\partial \log l_j}{\partial \beta_j} &= \sum_{i=1}^{n_j} \left(\frac{numeb1}{deno1} \right) + \sum_{i=1}^{n_j} \left(\frac{numeb2}{deno2} \right) - \sum_{i=1}^n \left(\frac{numeb3}{deno3} \right) \\ \frac{\partial^2 \log l_j}{\partial \beta_j^2} &= \sum_{i=1}^{n_j} \left(\frac{(deno1 * p2 * \lambda * x2l * eb * (-\beta - 2)) - (numeb1 * numeb1)}{deno1^2} \right) - \\ &\quad \sum_{i=1}^{n_j} \left(\frac{(deno2 * p2 * eb * (-G1^{2*\lambda})) - (numeb2 * numeb2)}{deno2^2} \right) + \\ &\quad \sum_{i=1}^n \left(\frac{(deno3 * p2 * ebb * (-x^{2*\lambda})) - (numeb3 * numeb3)}{deno3^2} \right) \\ \frac{\partial \log l_j}{\partial \gamma_j} &= \sum_{i=1}^{n_j} \left(\frac{numeg1}{deno1} \right) + \sum_{i=1}^{n_j} \left(\frac{numeg2}{deno2} \right) - \sum_{i=1}^n \left(\frac{numeg3}{deno3} \right) \\ \frac{\partial^2 \log l_j}{\partial \gamma_j^2} &= \sum_{i=1}^{n_j} \left(\frac{((p1 * deno1 * \alpha * ea * \log(G1) * xg1 * (\gamma * \log(G1) + 2 - \gamma * \log(G1) * \alpha * (G1^\gamma) - \alpha * (G1^\gamma) - \alpha * \gamma * (G1^\gamma) * \log x2 - \alpha * (G1^\gamma) - \alpha * \gamma * x2g)) - (numeg1 * numeg1))}{deno1^2} \right) + \\ &\quad \sum_{i=1}^{n_j} \left(\frac{(p1 * deno2 * \alpha * ((\log(G1))^2) * (G1^\gamma) * ea * (1 - \alpha * \gamma * (G1^{2*\gamma})) - (numeg2 * numeg2))}{deno2^2} \right) - \\ &\quad \sum_{i=1}^n \left(\frac{(p1 * deno3 * \alpha * ((\log(x))^2) * (x^\gamma) * eaa * (1 - \alpha * \gamma * (x^{2*\gamma})) - (numeg3 * numeg3))}{deno3^2} \right) \\ \frac{\partial \log l_j}{\partial \lambda_j} &= \sum_{i=1}^{n_j} \left(\frac{numel1}{deno1} \right) + \sum_{i=1}^{n_j} \left(\frac{numel2}{deno2} \right) - \sum_{i=1}^n \left(\frac{numel3}{deno3} \right) \\ \frac{\partial^2 \log l_j}{\partial \lambda_j^2} &= \sum_{i=1}^{n_j} \left(\frac{(p2 * deno1 * \beta * eb * \log(G1) * x1 * (\lambda * \log(G1) + 2 - \lambda * \log(G1) * \beta * (G1^\lambda) - \beta * (G1^\lambda) - \beta * \lambda * (G1^\lambda) * \log x2 - \beta * (G1^\lambda) - \beta * \lambda * x2l)) - (numel1 * numel1))}{deno1^2} \right) + \\ &\quad \sum_{i=1}^{n_j} \left(\frac{(p2 * deno2 * \beta * ((\log(G1))^2) * (G1^\lambda) * eb * (1 - \beta * \lambda * (G1^{2*\lambda})) - (numel2 * numel2))}{deno2^2} \right) - \\ &\quad \sum_{i=1}^n \left(\frac{(p2 * deno3 * \beta * ((\log(x))^2) * (x^\lambda) * ebb * (1 - \beta * \lambda * (x^{2*\lambda})) - (numel3 * numel3))}{deno3^2} \right) \\ \frac{\partial^2 \log l_j}{\partial \alpha_j \partial \gamma_j} &= \frac{\partial^2 \log l_j}{\partial \gamma_j \partial \alpha_j} \\ &= - \sum_{i=1}^{n_j} \left(\frac{(deno1 * p1 * \alpha * x2g * ea * (\gamma * \log(G1^2) + 1 - \alpha * \gamma * (G1^\gamma) * \log(G1))) - (numea1 * numeg1)}{deno1^2} \right) + \\ &\quad \sum_{i=1}^{n_j} \left(\frac{(p1 * deno1 * xg1 * ea * (\gamma * \log(G1) + 1 - \alpha * \gamma * (G1^\gamma) * \log(G1))) - (numea2 * numeg1)}{deno1^2} \right) + \end{aligned}$$

$$\sum_{i=1}^{n_j} \left(\frac{(p1 * deno2 * (G1^{\gamma}) * \log(G1) * ea * (1 - \alpha * (G1^{\gamma}))) - (numea3 * numeg2)}{deno2^2} \right) -$$

$$\sum_{i=1}^n \left(\frac{(p1 * deno3 * (x^{\gamma}) * \log(x) * eaa * (1 - \alpha * (x^{\gamma}))) - (numea4 * numeg3)}{deno3^2} \right)$$

$$\frac{\partial^2 \log l_j}{\partial \lambda_j \beta_j} = \frac{\partial^2 \log l_j}{\partial \beta_j \lambda_j}$$

$$= - \sum_{i=1}^{n_j} \left(\frac{(deno1 * p2 * \beta * x2l * eb * (\lambda * \log(G1^2) + 1 - \beta * \lambda * (G1^{\lambda}) * \log(G1))) - (numeb11 * numel1)}{deno1^2} \right) +$$

$$\sum_{i=1}^{n_j} \left(\frac{(p2 * deno1 * x1l * eb * (\lambda * \log(G1) + 1 - \beta * \lambda * (G1^{\lambda}) * \log(G1))) - (numeb22 * numel1)}{deno1^2} \right) +$$

$$\sum_{i=1}^{n_j} \left(\frac{(p2 * deno2 * (G1^{\lambda}) * \log(G1) * eb * (1 - \beta * (G1^{\lambda}))) - (numeb33 * numel2)}{deno2^2} \right) -$$

$$\sum_{i=1}^n \left(\frac{(p2 * deno3 * (x^{\lambda}) * \log(x) * ebb * (1 - \beta * (x^{\lambda}))) - (numeb44 * numel3)}{deno3^2} \right)$$

$$\frac{\partial^2 \log l_j}{\partial \alpha_j \beta_j} = \frac{\partial^2 \log l_j}{\partial \beta_j \alpha_j} = \sum_{i=1}^{n_j} \left(\frac{nume1 * numeb1}{deno1^2} \right) - \sum_{i=1}^{n_j} \left(\frac{nume2 * numeb2}{deno2^2} \right) +$$

$$\sum_{i=1}^n \left(\frac{nume3 * numeb2}{deno3^2} \right)$$

$$\frac{\partial^2 \log l_j}{\partial \alpha_j \lambda_j} = \frac{\partial^2 \log l_j}{\partial \lambda_j \alpha_j} = \sum_{i=1}^{n_j} \left(\frac{nume1 * numel1}{deno1^2} \right) - \sum_{i=1}^{n_j} \left(\frac{nume2 * numel2}{deno2^2} \right) +$$

$$\sum_{i=1}^n \left(\frac{nume3 * numel3}{deno3^2} \right)$$

$$\frac{\partial^2 \log l_j}{\partial \beta_j \gamma_j} = \frac{\partial^2 \log l_j}{\partial \gamma_j \beta_j} = \sum_{i=1}^{n_j} \left(\frac{numeb1 * numeg1}{deno1^2} \right) - \sum_{i=1}^{n_j} \left(\frac{numeb2 * numeg2}{deno2^2} \right) +$$

$$\sum_{i=1}^n \left(\frac{numeb3 * numeg3}{deno3^2} \right)$$

$$\frac{\partial^2 \log l_j}{\partial \gamma_j \lambda_j} = \frac{\partial^2 \log l_j}{\partial \lambda_j \gamma_j} = \sum_{i=1}^{n_j} \left(\frac{numeg1 * numel1}{deno1^2} \right) + \sum_{i=1}^{n_j} \left(\frac{numeg2 * numel2}{deno2^2} \right) -$$

$$\sum_{i=1}^n \left(\frac{numeg3 * numel3}{deno3^2} \right)$$