International Journal of Scientific Research in Mathematical and Statistical Sciences **Research Paper**
Vol.6, Issue.2, pp.262-266, April (2019) **E-ISSN:** 2348-4519
DOI: https://doi.org/10.26438/ijsrmss/v6i2.262266

Linear estimation models by polynomial hazard functions

S.G. Parekh[1,*] D.K. Ghosh[2], S.R. Patel[3]

[1*]Faculty of Business Administration, Dharmsinh Desai University, Nadiad, India
[2]Department of Statistics, Saurastra University, Rajkot, India
[3]Department of Statistics, Sardar Patel University, Vallabh Vidyanagar, India

[*]Corresponding Author: sg_parekh@yahoo.com

Available online at: www.isroset.org

Received: 21/Apr/2019, Accepted: 30/Apr/2019, Online: 30/Apr/2019

Abstract**—**The paper deals with the estimation of survival function with the use of polynomial hazard function as frailty and general baseline distribution. Considering Cox PH regression model with some frailty distribution the estimation of regressors and frailty parameter is done. Particular cases for parabolic hazard function have been considered with different base line distributions. Further with linear hypothesis of regressors and frailty parameter the linear estimation is carried out by Gauss-Marcoff model. Theory has been supported by suitable example illustrated by taking kidney infection data.

Keywords**—**Linear Models; Polynomial Hazard; Frailty variable; Gauss-Markov model; kidney infection.

I. INTRODUCTION

Several survival models has been introduced by many authors, amongst which Cox [1] PH model is well-known using regression analysis for censored data. The study of this model needs the co-variates, which should be defined. In many cases it is not possible to include all covariates, for example genetic factor comprising of all possible genes influencing survival. These unknown unobservable unmeasurable factors termed as heterogeneity or frailties are included in the model as random effect term or frailty which extends the Cox PH model. The term frailty is used first by Vaupel et al [2] and was separately used by Clayton [3].

Several authors like Ibrahim et al. [4], Sahu et al. [5], Yu [6], Santos and Achcar [7] and Hanagal [8] used Weibull distribution when hazard function is linear function of frailty parameter. We take a more general hazard function as parabolic in frailty parameter θ. Parekh and Patel [9, 10] have discussed some univariate continuous frailty models and univariate Bayesian frailty models.

Generally hazard function is constant (Exponential distribution) or log-linear function (Weibull distribution) for the life time data. Extending this idea for the hazard function as polynomial such as

$$h(t) = \alpha_0 + \alpha_1 t + \cdots + \alpha_{m-1} t^{m-1}\ ,\ t \geq 0 \qquad (1.1)$$

the model corresponding to this type of hazard function gives shapes of hazard function as U shape in contrast to linear line; also these models sometimes provide distribution free methods of estimating survival or hazard function.

The cumulative hazard function is of the form

$$H(t) = \gamma_1 t + \gamma_2 t^2 + \cdots + \gamma_m t^m \qquad (1.2)$$

where $\gamma_j = \frac{\alpha_{j-1}}{j}, j = 1,2, \ldots, m$

The survival function, $S(t)$ can be expressed as function of $H(t)$ such as

Since hazard function, $h(t) = \frac{f(t)}{S(t)}$ and

$$H(t) = \int \frac{f(t)}{S(t)} dt = -logS(t)$$

$$\Rightarrow \quad S(t) = e^{-H(t)} \qquad (1.3)$$

and the p.d.f.

$$f(t) = h(t).S(t) \qquad (1.4)$$

and the hazard function given in (1.1) with m = 1, $h(t) = a_0$ corresponding to exponential distribution and for m = 2, the log linear hazard rate corresponds to Weibull distribution. Distribution with $m > 2$ are sometimes useful though it is rarely used and for m $\geq$ 4 has theoretical importance but there is no practical use. Distribution with $m > 2$ have non-monotone hazard function.

Instead of constant or linear hazard, we consider the hazard rate as a polynomial of order 2 such as

$$h(t) = a_0 + a_1 t + a_2 t^2$$

which is non-monotone hazard functions.

And $H(t) = \int h(t)dt = a_0 t + \frac{a_1 t^2}{2} + \frac{a_2 t^3}{3}$ which ensures the sufficient condition $H(0) = 0$ and $H(\infty) = \infty$ for any polynomial to be taken as hazard function.

Since hazard function for the Weibull base line distribution is a log linear function, one has to obtain the estimation of hazard function, survival function and probability density function by obtaining the estimator of θ and by using (1.3) and (1.4) we get the estimators of survival function $S(t)$, density function $f(t)$. Let us use Cox PH model in this context. As defined by Cox the model is of this form. If T denotes survival time, $h(t)$ denotes hazard rate and $\underline{\beta}$ the vector of parameters and $\underline{X}$ be vector of covariates then Cox proportional hazard function will be

$T = h_0(t)e^{\underline{X}'\underline{\beta}+\gamma_0 t}$

$Y = \log(T) = \alpha + \underline{X}'\underline{\beta} + \theta^*,$

where $\gamma_0 t = e^{\theta} = \theta^*$, known as frailty variable.

In most general form the Cox PH model is

$$\eta = \underline{\delta}'\underline{\beta} + \xi\theta^* \tag{1.5}$$

where θ^* is associated with hazard rate $h(\theta^*) = a_0 + a_1\theta^* + a_2\theta^{*2}$ in this article. The main problem of this paper is to estimate θ^* and hazard rate, $h(\theta^*)$ hence that of survival function using Cox PH model. We use least square theory to estimate θ^* and thereby to estimate survival function in section II. Particular cases have been discussed in section III. In section IV we have given the application of the theory to kidney infection data.

II. LINEAR ESTIMATION

In this section we convert the Cox PH model introduced in section 1 by using error term as frailty variable. The Gauss-Markoff model is considered as random effect model. Then we obtain the estimators of θ^* and $\underline{\beta}$.

With usual notations considering Gauss-Markoff linear model

$$\underline{Y} = X\underline{\beta} + \underline{e} \tag{2.1}$$

where $\underline{Y}$ is nx1 vector, $(Y_1, Y_2, \dots Y_n)$, X is nxp matrix $(X_{ij}, i = 1,2,\dots,n; j = 1,2,\dots p)$ of known constants, $\underline{\beta}$ is px1 vector of parameters $\beta_1, \beta_2, \dots \beta_p$ and $\underline{e} = (e_1, e_2, \dots, e_p)'$ is error vector responsible for frailty model. $E(\underline{\epsilon}) = \underline{1}\theta^*$ and $E(\underline{\epsilon}\,\underline{\epsilon}') = h(\theta^*)I$, where $\underline{1}$ is the n-vector of one's and $h(\theta^*) = a_0 + a_1\theta^* + a_2\theta^{*2}$ and reconstructing it as

$$\underline{Y} = X\underline{\beta} + \underline{1}\theta^* + \underline{\epsilon}^* \tag{2.2}$$

where,

$\underline{\epsilon}^* = \underline{\epsilon} - \underline{1}\theta^*$, such that $E(\underline{\epsilon}^*) = \underline{1}\theta^*$and $E(\underline{\epsilon}^*\underline{\epsilon}^{*\prime}) = h(\theta^*)I$.

We are interested in the simultaneous estimation of linear function of β and θ^* as they are associated for shared frailty model with multivariate exponential, multivariate Weibull and other such distributions. We consider this reconstructed linear model (2.2) which contains variance as a polynomial of mean θ^* and to avoid $h(\theta^*)$ one has to minimize

$$\phi(\underline{\beta}, \theta^*) = \left(\underline{Y} - X\underline{\beta} - \underline{1}\theta^*\right)' [h(\theta^*)]^{-1} \left(\underline{Y} - X\underline{\beta} - \underline{1}\theta^*\right), \quad \forall\, \theta^* \in R^1 \text{ and } \underline{\beta} \in R^p$$

Minimization of $\phi(\underline{\beta}, \theta^*)$ leads to the following two equations

$$X'X\beta + X'\underline{1}\theta^* = X'\underline{Y} \tag{2.3}$$

$$(a_1 + 2a_2\theta^*)\left(\underline{Y} - X\underline{\beta} - \underline{1}\theta^*\right)'\left(\underline{Y} - X\underline{\beta} - \underline{1}\theta^*\right) + 2\underline{1}'\left(\underline{Y} - X\underline{\beta} - \underline{1}\theta^*\right)\left(a_0 + a_1\theta^* + a_2\theta^{*2}\right) = 0 \tag{2.4}$$

Now obtaining $\underline{\beta}$ in terms of θ^* from equation (2.3) and solving resultant equation for θ^*, the estimation of $\hat{\underline{\beta}}, \widehat{\theta^*}$ of $\underline{\beta}, \theta^*$ are obtained as

$$\widehat{\theta^*} = \frac{\left(a_0\underline{1}'A\underline{1} - a_2\underline{y}'A\underline{y}\right) + \sqrt{M}}{\left(a_1\underline{1}'A\underline{1} + 2a_2\underline{1}'A\underline{y}\right)}, a_1 = a_2 \neq 0 \tag{2.5}$$

and

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{Y} - (X'X)^{-1}X'\underline{1}\left[\frac{\left(a_0\underline{1}'A\underline{1} - a_2\underline{y}'A\underline{y}\right) + \sqrt{M}}{\left(a_1\underline{1}'A\underline{1} + 2a_2\underline{1}'A\underline{y}\right)}\right] \tag{2.6}$$

where,

$A = (I - X(X'X)^{-1}X')$,

$M = (a_0\underline{1}'A\underline{1} - a_2\underline{y}'A\underline{y})^2 + \left(a_1\underline{1}'A\underline{1} + 2a_2\underline{1}'A\underline{y}\right)\left(a_1\underline{y}'A\underline{y} + 2a_0\underline{1}'A\underline{y}\right)$

and $X'X$ is taken as positive definite. If $X'X$ is not positive definite then instead of $(X'X)^{-1}$ we use the generalized inverse $(X'X)^{-}$ and in this case $\widehat{\theta^*}$ and $\hat{\underline{\beta}}$ will not be uniquely determined.

Using $\widehat{\theta^*}$ one can obtain the estimators of hazard rate, $h(t)$; Survival function, $S(t)$ and probability density function, $f(t)$.

Further with the use of estimators $\hat{\underline{\beta}}$ of $\underline{\beta}$ one gets the estimators of regressors. We have derived the estimators of regressors which are generally used in cox PH model. Thus the estimator is used for the main variables affecting hazard function and thereby affecting p.d.f. $f(t)$.

III. PARTICULAR CASES OF THE ESTIMATORS

In this section we consider some of the particular cases of these estimators obtained in (2.5) and (2.6). For hazard function $h(\theta^*) = a_0 + a_1\theta^* + a_2\theta^{*2}$, some special cases like (i) $a_2 = 0$ (ii) $a_0 = a_2 = 0$ and $a_1 = 1$ (iii) $a_0 = a_1 = 0$ and $a_2 = 1$ (iv) $h(\theta^*) = a_0$ have been discussed as under.

Case I: Let $a_2 = 0$ then the hazard rate function being linear function of θ^*, it corresponds to the hazard function with baseline Weibull distribution. The estimators $\widehat{\theta^*}$ and $\hat{\underline{\beta}}$ obtained in (2.5) and (2.6) respectively will reduce to

$$\widehat{\theta^*} = \frac{\left(a_0\underline{1}'A\underline{1}\right) \pm \sqrt{\mathrm{N}}}{\left(a_1\underline{1}'A\underline{1}\right)}$$

and

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y} - (X'X)^{-1}X'\underline{1}\left[\frac{\left(a_0\underline{1}'A\underline{1}\right) \pm \sqrt{\mathrm{N}}}{\left(a_1\underline{1}'A\underline{1}\right)}\right]$$

where,

$A = (I - X(X'X)^{-1}X')$,

$N = (a_0\underline{1}'A\underline{1})^2 + \left(a_1\underline{1}'A\underline{1}\right)\left(a_1\underline{y}'A\underline{y} + 2\mathrm{a}_0\underline{1}'A\underline{y}\right)$

Case II: If $\mathrm{a}_0 = \mathrm{a}_2 = 0$ and $\mathrm{a}_1 = 1$ which corresponds to the hazard rate of special type of Weibull distribution as base line distribution, then the estimator of $\widehat{\theta^*}$ and $\underline{\hat{\beta}}$ obtained in (2.5) and (2.6) respectively will reduce to

$$\widehat{\theta^*} = \pm\sqrt{\frac{\underline{y}'A\underline{y}}{\underline{1}'A\underline{1}}}$$

and

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y} - (X'X)^{-1}X'\underline{1}\sqrt{\frac{\underline{y}'A\underline{y}}{\underline{1}'A\underline{1}}}$$

Case III: If $\mathrm{a}_0 = \mathrm{a}_1 = 0$ and $\mathrm{a}_2 = 1$ and which is the case for parabolic hazard rate, then the estimator of $\widehat{\theta^*}$ and $\underline{\hat{\beta}}$ obtained in (2.5) and (2.6) respectively will reduce to

$$\widehat{\theta^*} = -\frac{\underline{y}'A\underline{y}}{\underline{1}'A\underline{y}}$$

and

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y} + (X'X)^{-1}X'\underline{1}\frac{\underline{y}'A\underline{y}}{\underline{1}'A\underline{y}}$$

Case IV: Let $h(\theta^*) = \mathrm{a}_0$, that is, $\mathrm{a}_1 = \mathrm{a}_2 = 0$ in $h(\theta^*) = a_0 + a_1\theta^* + a_2{\theta^*}^2$. In this case hazard rate is constant which means that the base line distribution is exponential.

Considering Normal equations given in (2.3) and (2.4) and using the above technique by substituting $\underline{\beta}$ from (2.3) in (2.4) it reduces to quadratic equation in θ^* such as

$$a{\theta^*}^2 + 2b\theta^* - c = 0 \quad (2.7)$$

where $a = a_1\underline{1}'A\underline{1} + 2\mathrm{a}_2\underline{1}'A\underline{y}$,

$b = a_0\underline{1}'A\underline{1} - \mathrm{a}_2\underline{y}'A\underline{y}$, $c = a_1\underline{y}'A\underline{y} + 2\mathrm{a}_0\underline{1}'A\underline{y}$

since $\mathrm{a}_1 = \mathrm{a}_2 = 0$, $a = 0$ and the above equation (2.7) reduces to a linear equation

$$2b\theta^* - c = 0$$

in which $b = a_0\underline{1}'A\underline{1}$ and $c = 2\mathrm{a}_0\underline{1}'A\underline{y}$

so that $\widehat{\theta^*} = \frac{\underline{1}'A\underline{y}}{\underline{1}'A\underline{1}}$

and $\hat{\beta} = (X'X)^{-1}X'\underline{Y} - (X'X)^{-1}X'\underline{1}\widehat{\theta^*}$

Remark: We note that $(X'X)^{-1}X'\underline{Y}$ is the usual regression estimator of $\underline{\beta}$. Due to shared frailty, the estimator is reduced by the quantity $(X'X)^{-1}X'\underline{1}\ \frac{\underline{1}'A\underline{Y}}{\underline{1}'A\underline{1}}$ which shows that the frailty has negative effect in the estimation of primary covariates. Particularly, when $\widehat{\theta^*} = \overline{Y}$, the shared frailty is neglected and in this case the regression frailty is usual regression estimator, $\hat{\beta} = (X'X)^{-1}X'\underline{Y}$, no new covariates are necessary to explain hazard rate or survivor function. This is due to exponential base line distribution.

IV. APPLICATION TO KIDNEY INFECTION PATIENTS.

For illustration we use the kidney infection data (see Appendix A) given by McGilchrist and Aisbett [11] of 38 kidney Patients using portable dialysis. A catheter is inserted to each admitted kidney Patient and keeps it until infection is removed, that is, infection is cleared. After some time again catheter is inserted and removed when infection occurs due to kidney or some other reasons. Time (in number of days) recorded when catheter is reinserted. The purpose is to note the incidence rate of infection to risk variables (Age, Sex and disease type of kidney disorder, Glomerulo Neptiritis(GN),Acute Neptiritis(AN) and Polycyatic Neptiritis(PKD). The time interval recorded is considered as random variable and according to Cox PH model we have illustrated the estimation by MLE, using the best linear unbiased prediction (BLUP). Estimation of the problem is described in McGilchrist and Aisbett [11].

4.1: We obtain the estimates of the the parameters $\underline{\beta}$ and θ for the kidney infected patients for the data given by McGilchrist and Aisbett (1991) which is shown in Appendix A, with the use of the estimators of θ^* and $\underline{\beta}$ obtained in (2.5) and (2.6) respectively as under.

$$\widehat{\theta^*} = \frac{(2.876295a_0 - 90.9593\mathrm{a}_2) \pm \sqrt{\mathrm{M}}}{(2.876295a_1 + 20.73458\mathrm{a}_2)} \quad (4.1)$$

where

$M = 8.273073\mathrm{a}_0{}^2 + 8273.594256\ \mathrm{a}_2{}^2 + 261.6257\mathrm{a}_1{}^2 + 62.566865\mathrm{a}_0\mathrm{a}_1 + 1885.961413\mathrm{a}_1\mathrm{a}_2 - 93.328752\mathrm{a}_0\mathrm{a}_2$

and Y is $logT$, the frailty parametric value should be in terms of $\hat{\theta} = ln\,\theta^*$

$$\underline{\hat{\beta}} = \begin{bmatrix} 0.049381 - 0.016642\hat{\theta} \\ 2.681690 - 0.329215\hat{\theta} \\ -0.44970 - 0.002971\hat{\theta} \\ -1.41220 + 0.125860\hat{\theta} \\ 0.364796 - 0.063681\hat{\theta} \end{bmatrix} \quad (4.2)$$

Some of the interesting cases are as under

Case (1): $a_0 = a_2 = 0\ and\ a_1 = 1$, associated with the hazard rate θ^* of Weibull distribution and estimates given in (4.1) and (4.2) of θ^* and $\underline{\beta}$ will be

$$\hat{\theta} = ln(\widehat{\theta^*}) = ln(5.623506) = 1.726955$$

and the value of the regression coefficient vector is

$$\underline{\hat{\beta}} = \begin{bmatrix} 0.020641 \\ 2.113151 \\ -0.45483 \\ -1.19484 \\ 0.25497 \end{bmatrix}.$$

For this case the estimated values of hazard rate, cumulated hazard rate, survival time and density function at time, t are respectively, Estimated hazard rate, $h(\hat{\theta}) = \hat{\theta} = 1.726955$

$H(\hat{\theta}) = \frac{\hat{\theta}^2}{2} = 1.491187$

Estimated survival time, $S(t) = e^{-H(t)}$

$= e^{-1.491187}$

$= 0.225105$

Estimated p.d.f., $f(t) = 7.671775$

Thus the estimated values of hazard rate is 1.726955, estimated survival time is 0.225105 and estimated value of density function at that time is 7.671775

Case (2): $a_0 = a_1 = 0 \; and \; a_2 = 1$, associated with the hazard rate θ^{*2} of weibull distribution and estimates given in (4.1) and (4.2) of θ^* and $\underset{\sim}{\beta}$ will be

$$\hat{\theta} = ln(\widehat{\theta^*}) = ln(8.773681) = 2.171756$$

and

$$\hat{\underset{\sim}{\beta}} = \begin{bmatrix} 0.013238 \\ 1.966715 \\ -0.45615 \\ -1.13886 \\ 0.226633 \end{bmatrix}$$

For this case the estimated values of hazard rate, cumulated hazard rate, survival time and density function at time, t are respectively,

Estimated hazard rate, $h(\hat{\theta}) = \hat{\theta}^2 = 4.716524$

$H(\hat{\theta}) = \frac{\hat{\theta}^3}{3} = 3.414380$

Estimated survival time, $S(t) = e^{-H(t)}$

$= e^{-3.414380}$

$= 0.032896$

Estimated p.d.f, $f(t) = 143.368616$

Thus the estimated values of hazard rate is 4.716524, estimated survival time is 0.032896 and estimated value of density function at that time is 143.368616

V. CONCLUSION and Future Scope

In cox PH model considering hazard function as parabolic, the estimators of regression coefficients in presence of frailty parameters have reduction effect than the usual estimator of the regression coefficients which also suggest that new covariates are either necessary to explain hazard rate or not. As a special case parabolic hazard function has been considered and the estimation has been done. Further estimation for some special cases of parabolic hazard function have been discussed. One may take general polynomial hazard function and derive the estimators of the Cox model and study its special cases.

APPENDIX A

Patient	Time(T)	Y=lnT	Age	Sex	GN	AN	PKD
1	8	2.07944	28	0	0	0	0
2	23	3.13549	48	1	1	0	0
3	22	3.09104	32	0	0	0	0
4	447	6.10256	31.5	1	0	0	0
5	30	3.4012	10	0	0	0	0
6	24	3.17805	16.5	1	0	0	0
7	7	1.94591	51	0	1	0	0
8	511	6.23637	55.5	1	1	0	0
9	53	3.97029	69	1	0	1	0
10	15	2.70805	51.5	0	1	0	0
11	7	1.94591	44	1	0	1	0
12	141	4.94876	34	1	0	0	0
13	96	4.56435	35	1	0	1	0
14	149	5.00395	42	1	0	1	0
15	536	6.28413	17	1	0	0	0
16	17	2.83321	60	0	0	1	0
17	185	5.22036	60	1	0	0	0
18	292	5.67675	43.5	1	0	0	0
19	22	3.09104	53	1	1	0	0
20	15	2.70805	44	1	0	0	0
21	152	5.02388	46.5	0	0	0	1
22	402	5.99645	30	1	0	0	0
23	13	2.56495	62.5	1	0	1	0
24	39	3.66356	42.5	1	0	1	0
25	12	2.48491	43	0	0	1	0
26	113	4.72739	57.5	1	0	1	0
27	132	4.8828	10	1	1	0	0
28	34	3.52636	52	1	0	1	0
29	2	0.69315	53	0	1	0	0
30	130	4.86753	54	1	1	0	0
31	27	3.29584	56	1	0	1	0
32	5	1.60944	50.5	1	0	1	0
33	152	5.02388	27	1	0	0	1
34	190	5.24702	44.5	1	1	0	0
35	119	4.77912	22	1	0	0	0
36	54	3.98898	42	1	0	0	0
37	6	1.79176	52	1	0	0	1
38	63	4.14313	60	0	0	0	1

REFERENCES

[1] Cox, D., "Regression models and life tables (with discussion)". Journal of the Royal Statistical Society,Series B, Vol.**34** , pp.**187-200**, **1972.**

[2] Vaupel, J. S., "The impact of heterogeneity in individual frailty on the dynamics of mortality", *Demography , Vol.* ***16***, pp.**439-454, 1979**.

[3] Clayton, D, "A model for association in bivariate life tables and its applications to epidemiological studies of familial tendency in chronic disease incidence," *Biometrica ,Vol.**65***, pp.**141-151, 1978.**

[4] Ibrahim, J. G., Chen, M., and Sinha, D., "*Bayesian Survival Analysis"*, New York: Springer, **2001**.

[5] Sahu, K.S., Dey, D.K., Aslanidou, H., Sinha, D., "A Weibull regression model with Gamma frailties for multivariate survival data", Lifetime Data Analysis, Vol.**3**, pp.**123 – 137**, **1997.**

[6] Yu, B.,"Estimation of shared gamma frailty models by a modified EM algorithm" Computational Statistics and Data Analysis , Vol. **50**, pp.**463-474, 2006**.

[7] Santos dos, C. A. and Achcar, J. A. "A Bayesian analysis for multivariate survival data in the presence of covariates" Journal of Statistical Theory and Applications, Vol. **9, pp.233-253, 2010**.

[8] Hangal, D., "*Modeling survival data using frailty models*", CRC Press, **2011**.

[9] Parekh S.G. and Patel S.R., "Some Univarite Continuous Frailty Models", International Journal of Computational and Theoretical Statistics, Vol.**5**, Issue.**1**, pp.**29-37, 2018a**.

[10] Parekh S.G. and Patel S.R., "On estimation of Univariate Bayesian frailty models," International Journal of Statistics and Applied Mathematics, Vol.**3**, Issue.**2**, pp.**199-205, 2018b**

[11] McGilchrist, C.A., Aisbett, C.W., "Regression with frailty in survival analysis" Biometrics, Vol.**47**, pp.**461 - 466, 1991**.

