THE EGYPTION STATSTICAL JOURNAL
ISSR, CAIRO UNIV., VOL . 44, NO. 2, 2000

MODELING COMPETING RISKS DATA USING A DAGUM TYPE I DISTRIBUTION

Adel A. N. Al-Hessainan and Yousef Al-Zalzalah

Department of Quantitative Methods and Information Systems
Kuwait University
Kuwait

ABSTRACT

An introduction to the theory of competing risks is given, then the formulation of competing risks models with the mathematical relationships between Dagum type-I and Beta-P models are derived. The formulation of the Dagum type-I model within the context of competing risks is given. Hence Dagum type-I probability distribution is used to fit competing risks data. The fitted model is compared to the Weibull propability distribution. Finally the model derived is used to fit the survival times of cancer patients and the maximum likelihood method is used to estimate the parameters of the model and their covariance structure.

Key Words and Phrases: lifetimes' models; Income models; Burr distributions; Pareto distributions; Parametric modeling; Cancer data.

1. Introduction

The theory of competing risks is concerned with the assessment of specific risks. In this study we focus our work on the special case where there are two types of causes of death (Failure) that are assumed to be of interest to the researcher, namely cancer and the other being a non-cancer cause of failure. The time till death (failure time) of a cancer patient is $Z = \min(Y_1, Y_2)$, where Y_1 is the lifetime of a cancer patient that has died because of cancer, Y_2 is the lifetime of a cancer patient that has died of other causes. We note that Y_1 and Y_2 are the theoretical random variables with unknown probability density functions.

2. Formulation of the competing risks model

In practice the time till failure of a system $Z = \min Y_l$, l=1,2 and the cause of failure c_l are the only observable variables.

Let π_1 be the probability that the cause of death is cancer, hence π_2 is the probability that the cause of death is non-cancer.

$$\pi_i = P\left\{Y_i = \min_l Y_l\right\} > 0, \pi_1 + \pi_2 = 1. \qquad (2.1)$$

Also let $f_i(x)$ be the probability density of failure time (death) given that c_i ; i=1, 2 is the cause of failure.

$$\text{Then } f_i(x) = \frac{1}{\pi_i} g_i(x)[1 - 2G_i(x)]I(x > 0), i = 1,2 \qquad (2.2)$$

Where g_i is the probability density function, and G_i is the cumulative distribution function, of the failure times Y_i , i = 1, 2.

It can be shown that the cumulative distribution function of failure time given that c_i is the cause of failure is $F_i(x)$ *where*

$$F_i(x) = \frac{1}{\pi_i}\left[G_i(x) - G_i^2(x)\right] \quad I(x > 0) \qquad i = 1,2 \qquad (2.3)$$

Naturally the conditional survival function is $R_i(x)$ = 1-$F_i(x)$, where $F_i(x)$ is given by (2.3).

Furthermore, $R_i(x)$ is the proportion of cancer patients that would survive past time x, given that the cause of failure is c_i.

The conditional hazard function is thus derived to be of the form

$$H_i(x) = \frac{g_i(x)[1 - 2G_i(x)]}{\pi_i + G_i^2(x) - G_i(x)} \quad I(x > 0) \quad ; i = 1, 2 \tag{2.4}$$

The hazard function (2.4) is interpreted as the proportion of cancer patients that die within the neiboughrhood, Δx, of the survived time x, given that the cause of death is c_i.

Now for n independent cancer patients let N_1 denote the random number of patients that died from cancer and let N_2 denote the random number of cancer patients that died from other causes. Then the joint probability density function of N_1 and N_2 is

$$p(n_1, n_2) = \frac{n!}{n_1!\, n_2!} \pi_1^{n_1} \pi_2^{n_2} \ , n = n_1 + n_2 \tag{2.5}$$

Then the conditional joint distribution of the failure times $x_{11}, x_{12}, .., x_{1n_1}$; $x_{21}, x_{22}, .., x_{2n_2}$ has the form

$$f(x_{11}, x_{12}, .., x_{1n_1}; x_{21}, x_{22}, .., x_{2n_2}) = \prod_{i=1}^{2} \frac{1}{\pi_i^{n_i}} \prod_{j=1}^{n_i} g_i(x_{ij}) \prod_{k=1}^{n_i} [1 - 2G_i(x_{ik})]; i \neq l \tag{2.6}$$

Where x_{ij} refers to the lifetime of the jth system failing from cause i.

Hence the likelihood of the observed cause of failure and associated lifetimes is

$$L = \frac{n!}{n_1! n_2!} \prod_{i=1}^{2} \prod_{j=1}^{n_i} g_i(x_{ij}) \prod_{k=1}^{n_2} [1 - 2G_i(x_{ik})]; \ i \neq l \tag{2.7}$$

Now rearranging the above likelihood function we get

$$L = \frac{n!}{n_1!\, n_2!} L_1 L_2 \tag{2.8}$$

Where

$$L_1 = \left[\prod_{j=1}^{n_1} g_1(x_{1j})\right]\cdot\left[\prod_{k=1}^{n_2}[1-2G_1(x_{1k})]\right] \quad , \quad L_2 = \left[\prod_{k=1}^{n_2} g_2(x_{1k})\right]\cdot\left[\prod_{j=1}^{n_1}[1-2G_2(x_{1j})]\right] \tag{2.9}$$

We note here that the parameters of L_1 and L_2 would be estimated by maximizing L_1 and L_2 separately. Herman and Patel (1971), Moeschberger and David (1971) and Hoel (1977) first pointed out this simplification.

3. Dagum type-I and the Beta-P models

Al-Zalzalah (1993) proposed the Beta-P model to represent the lifetime distribution of cancer patients, $g_1(x)$. He has shown that the Beta-P model provides a good fit for competing risks data. Milke and Johnson (1974) discussed the Beta-P distribution and they have shown that this distribution appears to provide a reasonable distribution of commonly encountered types of measurements when they are heavy tailed distributed and in addition posses very desirable computational properties.

The Beta-P comulative distribution function is given by,

$$U(y) = 1-\left[1+\left(\frac{y}{\beta}\right)^{\theta}\right]^{-\alpha} I(y>0) \tag{3.1}$$

Where $0<\alpha, 0<\beta$ and $0<\theta$. Milke and Johnson (1974) give other properties of the Beta-P distribution function.

In this paper we propose the use of the Dagum type-I distribution to model the lifetime probability distribution within the context of competing risks theory.

Dagum (1977) proposed the Dagum type-I model, to model personal income distributions. Our interest in using the Dagum type-I model stems from the results found, in Al-Hessainan (1994), about the mathematical relationships between this

distribution and other members of the generalized logistic-Burr system of distributions. The Dagum type-I model is actually found to be a generalization of the Burr type III and Burr type XII distributions. The Burr type XII distribution has been devised as a lifetime distribution model by many authors. Of which are Burr (1942), Burr and Cislak (1968), Papadopoulos (1978), Tadikamalla (1980), Lewis (1981), Wingo (1983), Evans and Ragab (1983), Al-Marzoug and Ahmed (1985), Khan and Khan (1987), Cordasevschi (1987), Nigm (1988), Al-Hussaini and Jaheen (1992) and Al-Hussaini, Mousa and Jaheen (1992). Another important property of the Dagum type-I distribution is that it has a closed mathematical form of its cumulative distribution function as well as its reliability and hazard functions.

For a random variable X with a Dagum type-I probability distribution $g_X(x)$ with parameters $(\alpha, \theta, \lambda)$, we find the random variable $Y = \frac{1}{X}$ has a Beta-P distribution $U_Y(y)$ with parameters (α, θ, β), where $\lambda = \beta^{-\theta}$. The respective forms of the probability density functions are

$$g_X(x;\alpha,\theta,\lambda) = \alpha\theta\lambda x^{-(\theta+1)}(1+\lambda x^{-\theta})^{-(\alpha+1)} I(x>0);\ \alpha,\theta,\lambda \geq 0 \qquad (3.2)$$

$$u_Y(y;\alpha,\theta,\beta) = \frac{\alpha\theta}{\beta}\left(\frac{y}{\beta}\right)^{\theta-1}\left(1+\left(\frac{y}{\beta}\right)^{\theta}\right)^{-(\alpha+1)} I(y>0);\alpha,\theta,\beta \geq 0 \qquad (3.3)$$

Where α and θ are shape parameters and λ and β are scale parameters.

The cumulative distribution function for the Dagum type-I model has the closed form

$$G_X(x;\alpha,\theta,\lambda) = (1+\lambda x^{-\theta})^{-\alpha} I(x>0);\alpha,\theta,\lambda \geq 0 \qquad (3.4)$$

Thus we find the survival function to be of the form

$$R_X(x;\alpha,\theta,\lambda) = 1-(1+\lambda x^{-\theta})^{-\alpha} \tag{3.5}$$

And the hazard function to be of the form

$$H_X(x;\alpha,\theta,\lambda) = \frac{\alpha\theta\lambda x^{-(\theta+1)}}{\left(1+\lambda x^{-\theta}\right)\left((1+\lambda x^{-\theta})^{\alpha}-1\right)} \tag{3.6}$$

4. The Dagum type-I distribution as a competing risks model

Now by assuming that the lifetimes of cancer patients, Y_i, follow the Dagum type-I distribution (3.2), the competing risks likelihood model (2.9), introduced in section 2, is found to be of the form

$$L_i = \left\{\prod_{j=1}^{n_i} \alpha \quad \theta \quad \lambda x_{ij}^{-(\theta+1)}(1+\lambda \quad x_{ij}^{-\theta})^{-(\alpha+1)}\right\}\left\{\prod_{k=1}^{n_l}\left(1-2(1+\lambda \quad x_{lk}^{-\theta})^{-\alpha}\right)\right\}; i=1,2, i \neq$$

(4.1)

By taking the natural log of (4.1), we obtain the log-likelihood function that has the form

$$\begin{aligned}\ln L_i = {} & n_i \ln(\alpha\theta\lambda) - (\theta+1)\sum_{j=1}^{n_i}\ln(x_{ij}) - (\alpha+1)\sum_{j=1}^{n_i}\ln(1+\lambda x_{ij}^{-\theta}) \\ & + \sum_{k=1}^{n_l}\ln\left(1-(1+\lambda x_{lk}^{-\theta})^{-\alpha}\right); \qquad i=1,2,\ i\neq l\end{aligned} \tag{4.2}$$

One can see here that a numerical estimation method is needed to obtain the maximum likelihood estimators of the parameters (α, θ, λ) of (4.2).

Once the competing risks model parameters are estimated, we can compute the conditional survival and hazard functions of the cancer patients.

From the competing risks cumulative distribution function (2.3), and assuming a Dagum type-I lifetime model, we find the conditional survival function to be of the form

$$R_i(x) = 1 - \frac{1}{\pi_i}\left[1 - \left(1 - (1+\lambda x^{-\theta})^{-\alpha}\right)^2\right] I(0 < x < v_i) \quad ; i = 1,2 \qquad (4.3)$$

The conditional hazard function for the cancer patients, based on a Dagum type-I lifetime model, is given by (4.4).

$$H_i(x) = \frac{2\alpha\theta\lambda x^{-\theta}(1+\lambda x^{-\theta})^{-\alpha}\left[1-(1+\lambda x^{-\theta})^{-\alpha}\right]}{\left[1-(1+\lambda x^{-\theta})^{-\alpha}\right]^2 - (1-\pi_i)} I(0 < x < v_i);\ i = 1,2 \quad (4.4)$$

It is worth noting here that the limiting lifetime, given that the cause of failure is c_i is given by

$$v_i = \frac{\lambda^{\frac{1}{\theta}}}{\left[\left(1-\sqrt{1-\pi_i}\right)^{\frac{-1}{\alpha}} - 1\right]^{\frac{1}{\theta}}} \quad ; i = 1,2 \qquad (4.5)$$

Function (4.5) could be interpreted as the maximum lifetime anticipated for a cancer patient that dies because of c_i. On the other hand it could be interpreted as the censored time of the observations of cancer patients lifetimes.

5. An application example: Survival times of cancer patients

Al-Zalzalah (1993) used the Beta-P competing risks model to fit cancer data obtained from Bogu (1949). Moeschberger and David (1971) proposed a Weibull probability distribution model to fit the same data set. The data set represents the survival times of 96 cancer patients treated for breast cancer in one particular

hospital. For the purpose of our study the patients are classified into two groups, patients who died from cancer and patients who died from other causes. For more details of the experiment we refer the reader to Al-Zalzalah (1993), David and Moeschberger (1976) and Bogu (1949).

Table 5.1 gives the survival times, in months, of patients in the two groups, according to cause of death.

Table 5.1: Survival times, in months, for cancer patients

Patients died with cancer present						Patients died from other causes	
0.3	12.2	15.5	28.2	41.0	78.0	0.3	110
5.0	12.3	17.9	21.1	42.0	80.0	4.0	111
5.6	13.5	19.8	30.0	44.0	84.0	7.4	112
6.2	14.4	20.4	31.0	46.0	87.0	15.5	132
6.3	14.4	20.9	31.0	48.0	89.0	23.4	162
6.6	14.8	21.0	32.0	48.0	90.0	46.0	
6.8	15.7	21.0	35.0	51.0	97.0	46.0	
7.5	16.2	21.1	35.0	51.0	98.0	51.0	
8.4	16.3	23.0	38.0	52.0	100	65.0	
8.4	16.5	23.6	39.0	54.0	114	68.0	
10.3	16.8	24.0	40.0	56.0	126	83.0	
11.0	17.2	24.0	40.0	60.0	131	88.0	
11.8	17.3	27.9	41.0	78.0	174	96.0	

To get the maximum likelihood estimates of the parameters of the Dagum competing risks model, for the cancer patients, we wrote a program in FORTRAN code. The program utilizes the optimization subroutine UMIAH in the IMSL MATH/STAT library (1987a) and (1987b). The estimated values of the parameters of L_1, and their standard deviations are listed in table 5.2.

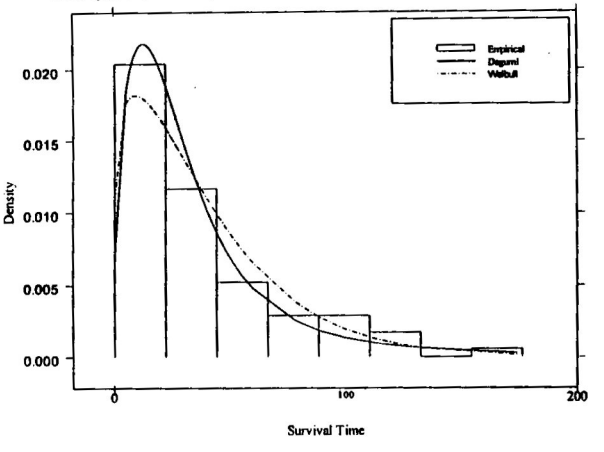
Table 5.2: MLEs for Dagum type-I competing risks model

Parameter	Estimated value	Standard deviation
α	0.6665	0.2145
0	2.0492	0.0105
λ	1504.1132	286.1819

The maximum likelihood estimates for the shape and the scale parameters of the Weibull distribution were found to be 1.197301 and 41.443228 respectively.

Figure 5.1 shows the empirical density of the survival times of the cancer patients along with the fitted line using the Dagum type-I and the Weibull probability density functions.

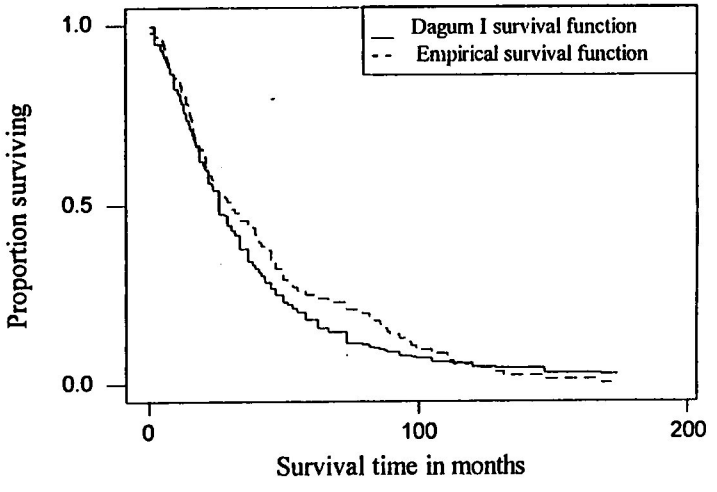
Figure 5.1: Empirical pdf and Dagum type-I fit vs Weibull fit for survival times of Breast cancer patents



It can be seen that the Dagum type-I model fits the data better than the Weibull model. Kolmogorov-Smirnov goodness of fit tests have shown that the Dagum type-I model is a good fit for this data at the $\alpha = 0.05$ level, and that it is a better fit than the Weibull model. The D-statistics for the two models are 0.12248 and 0.16605 repectively.

Figure 5.2 shows the fit of the Dagum type-I survival function (3.5). For example, a breast cancer patient is estimated to have a 54% chance of surviving for more than 24 months. She has a chance of 18% to survive for more than 60 months.

Figure 5.2: Dagum type-I survival function fit to breast cancer data.



6. Conclusion

In this study we found that the Dagum type-I model is a suitable model to fit survival times of cancer patients as well as competing risks models. The model has shown a better fit, to the breast cancer data, than the Weibull distribution. Hence we recommend its use in modeling lifetimes and competing risks, especially for two competing risks models. In future work the aouthors would conduct a comparative study of the useage of the Dagum type-I functional form to other models used in the survival modeling literature. The exponential, gamma, and Pearson type-VI are few examples of such forms. In addition the authors are interested in modeling lifetimes data, solicited from local health institutions.

As discussed in Al-Hessainan (1994), there are close mathematical relationships between the Dagum type-I model and the other members of the

generalized logistic-Burr system of distributions. This keeps the selection doors open for a host of different parametric families of probability models to fit competing risks data.

We also intend to investigate the possibility of the use of other members of the generalized logistic-Burr system of distributions to fit survival times and competing risks data.

Acknowledgment

Research supported by grant CS040, awarded by Kuwait University.

The authors would like to express gratitude to the two anonymous referees for their sound comments and suggestions that improved the paper.

BIBLIOGRAPHY

Al-Hessainan, A. N. (1994). *The Lorenz curve and its applications; Bayesian estimation of the parameters of the generalized logistic-Burr system of distributions.* Ph.D. dissertation. University of California, Riverside.

Al-Hussaini, E. K. and Jaheen, Z. F. (1992). Bayesian estimation of the parameters, reliability and failure rate functions of the Burr type XII failure model. *Journal of Statistical Computation and Simulation,* Volume 41, 31-40.

Al-Hussaini, E. K., Mousa M. A. M. A. and Jaheen, Z. F. (1992). Estimation under the Burr type XII failure model based on censored data: a comparative study. *Test,*Volume 1, No. 1, 47-60.

Al-Marzoug, A. M. and Ahmed, M. (1985). Estimation of parameters of Burr probability model using fractional moments. *Pakistan J. Statist.*, 1, 67-77.

Al-Zalzalah, Y. S. H. (1993). *Analysis of competing independent and dependent risks.* Ph.D. dissertation. Colorado State University, Fort Collins.

Boug, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. R. Stat. Soc.* B11, 15-53.

Burr, I. W. (1942). Cumulative frequency functions. *Ann. Math. Statist.*, 13, 215-232.

Burr, I. W. and Cislak, P. J. (1968). On a general system of distributions: I. Its curve-shaped characteristics; II. The sample median. *Journal of the American Statistical Association,* 63, 627-635.

Cordasevschi, C. (1987). Tolerance limits for a distribution of Burr type. *Stud. Cerc. Mat.*, 39, 489-492.

Dagum, C. (1977). A new model of personal income distribution: specifications and estimation. *Economie Appliquee*, 30(3), 413-436.

David, H. A. and Moeschberger, M. L. (1976). *The theory of competing risks.* Macmillam publishing co., Inc. New York.

Evans, I. G. and Ragab, A. S. (1983). Bayesian inferences given a type-2 censored sample from Burr distribution. *Communication in Statistics, Theory and Methods*, 12, 1569-1580.

Herman, R. J. and Pattel, R. K. N. (1971). Maximum likelihood estimation for multi-risk model. *Technometrics*, 13, 358-396.

Hoel, D. G. (1977). A representation of mortality data by competing risks. Biometrics, 28, 475-488.

IMSL Problem-Solving Software Systems (1987a). *IMSL User's Manual, MATH/LIBRARY*, Volume 3, Page 817. Houston.

IMSL Problem-Solving Software Systems (1987b). *IMSL User's Manual, STAT/LIBRARY*, Volume 3, Page 963. Houston.

Khan, A. H. and Khan, A. I. (1987). Moments of order statistics from Burr's distribution and its characterization, *Metron*, 45, 21-29.

Lewis, A. W. (1981). *The Burr distribution as a general parametric family in survivorship and reliability theory applications*. Ph.D. dissertation. Department of Biostatistics, University of North Carolina, Chapel Hill.

Milke, P. W. and Johnson, E. S. (1974). Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. Water resource research, 10, No. 2, 223-226.

Moeshberger, M. L. and David, H. A. (1971). Life tests under competing causes of failure and the theory of competing risks. Biometrics, 27, 909-933.

Nigm, A. M. (1988). Prediction bounds for the Burr model. *Communication in Statistics, Theory and Methods*, 17, 287-296.

Papadopoulos, A. S. (1978). The Burr distribution as a failure model from a Bayesian approach. *IEEE, Trans. Rel.*, R-5, 369-371.

Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *International Statistical Review*, 48, 337-344.

Wingo, D. R. (1983). Maximum likelihood methods for fitting the Burr type XII distribution to lift data. *Biometrical Journal*, 25, 77-84.

Dr. Adel Al-Hessainan and Dr. Yousef Al-Zalzalah

Assistant Professors of statistics
Department of Quantitative Methods and Information Systems
College of administrative sciences
Kuwait University
P.O. Box 5486
Safat, 13055
Kuwait

E-mail:

Hussainan@kuc01.kuniv.edu.kw

Zalzalah@kuc01.kuniv.edu.kw