Journal of Statistics and Mathematics
ISSN: 0976-8807 & E-ISSN: 0976-8815 Volume 3, Issue 1, 2012, pp.-74-77.
Available online at http://www.bioinfo.in/contents.php?id=85

GUMBEL DISTRIBUTION MODEL FOR THE BREAST CANCER SURVIVAL DATA USING MAXIMUM LIKELIHOOD METHOD

KHAN K.H.*

Department of Mathematics, College of Science and Humanities, Salman Bin Abdulaziz University, Al-Kharj, Kingdom of Saudi Arabia
*Corresponding Author- kizarkhan@yahoo.com

Received: March 11, 2012; Accepted: April 09, 2012

Abstract -The survival rate estimates for the breast cancer censored data have been considered for the 254 patients. The data [10] was treated at the chemotherapy department, Bradford Royal Infirmary for ten years. Here in this paper Gumbel probability distribution (see [3], [4], [5]) model is used to obtain the survival rates of the patients. Maximum likelihood method [9] has been used through unconstrained optimization method [12, 13] (DFP-Davidon-Fletcher-Powell) to find the parameter estimates and variance-covariance matrix for the Gumbel distribution model. Finally the survivor rate estimates for the parametric (Gumbel) probability model has been compared with the non-parametric (Kaplan-Meier) [7] method.
Keywords- Gumbel distribution model, Censoring, Breast Cancer Data sets, DFP-unconstrained optimization method, Maximum likelihood function and Kaplan-Meier survivor rate estimates.

Citation: Khan K.H. (2012) Gumbel Distribution Model for the Breast Cancer Survival Data Using Maximum Likelihood Method. Journal of Statistics and Mathematics, ISSN: 0976-8807 & E-ISSN: 0976-8815 Volume 3, Issue 1, pp.-74-77.

Introduction

Breast cancer is a systemic disease [1,10,15] until proved otherwise. When the treatment is stopped the disease progresses with uniform 'velocity' v through a fixed 'distance'd in the disease to recurrence point. In this paper, we find the parameter estimates, survival rate estimates, variance covariance matrix for the Gumbel probability distribution model using maximum likelihood function using breast cancer data [14]. For the survival of the patient with the breast cancer, a statistical approach is considered; wihich is based on two parameters refered as scale and shape parameters respectively of the said distributions. Further work on probabilistic approach has been done by Khan, K.H. [14]. using Inverse Guassioan distribution model. The survivor rate estimates for the Gumbel probability distribution has also been compared with the non-parametric model [7].

The Gumbell Model and Estimation of Parameters

The data regarding survival analysis generally falls in two classes- (i) the failure time of items, which actually fail during the experiment, (ii) the survival times of items which, actually survive with the experiment.

These classes are generally separated statistically by the use of censoring, for detail see Cox, [11]. In parametric models the pdf of lifetime 'T' has form with survival function, where is a vector of parameters. The contribution to the likelihood of an item that fails at time t is and an item that survives beyond time is. Thus, according to the Lawless [8], using the Gumbel distribution models, the likelihood function when the time is divided into intervals is given as

$$L=\prod_{i=1}^{NG}\left[F(t_i)-F(t_{i-1})\right]^{f_i}\left(1-F(T_n)\right)^{N-F} \qquad (2.1)$$

where NG, f_i , N and F are the number of recurrence groups, number of failures (recurrences) in the ith year, sample size and total number of recurrences in 10 years respectively.

The maximum likelihood estimates can be obtained by taking the log-likelihood function. Since the probability of no failure until time t is defined by, then the log-likelihood function can be written as

$$\ln L = \sum_{i=1}^{10} f_i\left[\ln\left(R(t_{i-1}) - R(t_i)\right)\right] + (N-F)\ln\left(R(T_n)\right) \qquad (2.2)$$

To find the parameter estimates we used the unconstrained optimization method '(DFP)' developed by Davidon and amended by Fletcher and Powell (see. [12],[13]). The DFP method (Quasi-Newton-Method) is an iterative method, which minimizes the objective function and requires only first partial derivatives in addition to the function values. So, the log-likelihood function to be maximized is equivalent to the minus times the log-likelihood function to be minimized. $(\ell = -\ln L)$. Therefore the required form for the estimation parameters is . . The variance-covariance matrix of estimates $\hat{a}\text{and } \hat{b}$

$$\begin{bmatrix} \frac{\partial^2 \ell}{\partial a^2} & \frac{\partial^2 \ell}{\partial a \partial b} \\ \frac{\partial^2 \ell}{\partial a \partial b} & \frac{\partial^2 \ell}{\partial b^2} \end{bmatrix}^{-1}$$

is calculated automatically and numerically as a part of these optimization procedures, and without any direct evaluation of the second derivatives of which would be very complicated. Since the article is concerned with the use of Gumbel distribution [6] so the pdf and are respectively as under.

$$f(t) = \frac{1}{\lambda}\exp\left[\frac{t-\mu}{\lambda} - \exp\left\{\frac{t-\mu}{\lambda}\right\}\right], \quad -\infty < t < \infty \qquad \lambda > 0 \quad (2.3)$$

where a scale parameter and is the location parameter.

For reparameterization,we take $a = \frac{1}{\lambda}$ and $b = \frac{1}{\lambda}\exp\left(\frac{-\mu}{\lambda}\right)$ and so the above *pdf* becomes

$$f(t) = b\exp(at)\exp\left[-\frac{b}{a}\exp\{at\}\right] \qquad (2.4)$$

$$R(t) = \int_t^{\infty} f(x)dx$$

Now the survivor function is

$$\Rightarrow R(t) = \exp\left[-\left(\frac{b}{a}\right)\exp(at)\right] . \qquad (2.5)$$

$$h(t) = \frac{f(t)}{R(t)}$$

The hazard function or the failure rate is

$$h(t) = b\exp(at)) \qquad (2.6)$$

The hazard rate/failure rate is proportional to the scale parameter and time as it passes.

Now the likelihood function can be written as

$$\ell = -\sum_{i=1}^{10} f_i\left[\ln\left(R(t_{i-1}) - R(t_i)\right)\right] - (N-F)\ln\left(R(T_n)\right) \qquad (2.6)$$

Where $R(t) = 0$ as $t \to \infty$ and at $t = 0,\ R(t) = \exp\left(-\frac{b}{a}\right) = k = 1$ (Max.)

The maximum likelihood estimates for (scale and shape parameters) $\hat{a}$ and $\hat{b}$, are the values of a and b which maximize L , or, equivalently, which minimize $\ell = -\ln L$. Thus, we have

$$\ell = -f_1\ln(k - R(t_1)) - \sum_{i=2}^{10} f_i\left[\ln\left(R(t_{i-1}) - R(t_i)\right)\right] - (N-F)\ln\left(R(T_n)\right) \qquad (2.7)$$

where $R(t) = \exp\left[-\frac{b}{a}\exp(at)\right]$ .The partial derivatives of ℓ w.r.t. a and b are

$$\frac{\partial \ell}{\partial a} = \frac{f_1}{(k - R(t_1))}\frac{\partial R(t_1)}{\partial a} - \sum_{i=2}^{10}\frac{f_i}{\Delta_i}\left(\frac{\partial R(t_{i-1})}{\partial a} - \frac{\partial R(t_i)}{\partial a}\right) - \frac{(N-F)}{R(T_n)}\frac{\partial R(T_n)}{\partial a} \qquad (2.8)$$

$$\frac{\partial \ell}{\partial b} = \frac{f_1}{(k - R(t_1))}\frac{\partial R(t_1)}{\partial b} - \sum_{i=2}^{10}\frac{f_i}{\Delta_i}\left(\frac{\partial R(t_{i-1})}{\partial b} - \frac{\partial R(t_i)}{\partial b}\right) - \frac{(N-F)}{R(T_n)}\frac{\partial R(T_n)}{\partial b} \qquad (2.9)$$

where $\frac{\partial R(t)}{\partial a} = -\frac{b}{a^2}e^{at}(at-1)R(t)$, $\frac{\partial R(t)}{\partial b} = -\frac{1}{a}e^{at}R(t)$ and $\Delta_i = R(t_{i-1}) - R(t_i)$.

Using eq.(2.7), eq.(2.8) and eq.(2.9) in the DFP optimization method, we can find the parameter estimates, survivor rate estimates, variance covariance matrices and maximum likelihood function.

Application

We considered the data of 254 patients surviving with breast cancer. These patients were initially treated at the department of chemotherapy department, Bradford Royal Infirmary, [15], England, thirty five years ago. Each patient was treated for a period of ten years or until death. The patients surviving with breast cancer were between 23 and 82 years old (Hancock et al. [10]). The patients were classified into four diffrenet stages using TNM (Tumor Nodes Metastases) system and clinically staged accordingly.

Out of 254 patients, 100 patients were premenopausal and 154 were postmenopausal. A woman was considered to be postmenopausal when 2 years had elapsed since her last menstrual period. The two main categories are premenopausal and postmenopausal. Note that Stages I & II for premenopausal and postmenopausal were each combined together. In the light of Table-1 and Table -2 the survival related to the clinical stage (%age) over ten years is given in Table-3.

Table 1- Age Distribution Related to Clinical Stage and Menopasal Status

Patient	Stage I		Stage II		Stage III		Stage IV	
Age	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
21-30	-	-	-	-	2	-	1	-
31-40	6	-	1	-	12	-	11	-
41-50	16	4	8	2	17	3	16	7
51-60	1	13	-	3	5	29	4	16
61-70	-	12	-	1	-	27	-	24
71-80	-	3	-	1	-	4	-	4
81-90	-	-	-	1	-	-	-	-

Table 2- Survivals and Failures Related to Clinical Stage and Menopausal Status

Stage	Meno-pausal Status	Surviv-ing with Cancer	Surviving with Re-currence	Dying without Cancer	Dying with Recur-rence	Dying with Cancer	Patients in each Stage
Stage I	Pre-	16	4	1	0	2	23
	Post-	8	5	4	1	14	32
Stage II	Pre-	5	1	0	0	3	9
	Post-	1	0	1	1	5	8
Stage III	Pre-	6	2	1	1	26	36
	Post-	5	4	2	7	45	63
Stage IV	Pre-	0	0	0	0	32	32
	Post-	1	1	0	3	46	51

Table 3- Survival Related to Clinical Stage (%age) over ten years

Satges	Pre-menopausial		Post-menopausal	
	Survival (% age)	Failure (% age)	Survival (% age)	Failure (% age)
I & II	81.25	18.75	35	65
III	22.22	77.78	14.29	85.71
IV	0%	100	3.92	96.08

Table 5-Estimates of Parameters amd ML-Function for Gumbel Distribution Model

Esti-mates	Pre-menopausal			Post-menopausal		
	Satge-I&II	Satge-III	Stage-IV	Satge-I&II	Satge-III	Stage-IV
$\hat{a}$	0.24132	0.27192	0.251214	0.26447	0.19508 3	0.32319 22
$\hat{b}$	0.01305	0.039898	0.011651 4	0.02049	0.04610 88	0.13861 43
MLF	30.3748 498	84.64368 85	105.9951 4	87.3541 861	156.286 883	63.3680 32

Table 6- Estimates of Variance-Covariance Matrix and Gradient vector for the Gumbel Model

Pre-Menopausal Stages			
Variance Covar-iance Matrix	Stage-I & II 0.0088010 - 0.0002268 -0.0002268 0.0000095	Stage-III 0.0011877 - 0.0001362 -0.0001362 0.0000905	Stage-IV 0.0019266 - 0.0003123 -0.0003123 0.0006691
Gradient Vector	-0.6894E-07 - 0.1612E-05	-0.31356E-07 - 0.13696E-05	-0.7231E-09 - 0.9972E-08
Post-Menopausal Stages			
Variance Covar-iance Matrix	Stage-I&II 0.0022478 - 0.00023554 -0.00023554 0.00004086	Stage-III 0.0009655 - 0.0001578 -0.0001578 0.0000554	Stage-IV 0.0009237 - 0.000091 -0.000091 0.0002653
Gradient Vector	0.26865E-08 0.40733E-07	-0.26702E-07 - 0.67640E-07	0.61286E-06 - 0.54996E-05

Table 7-Survival Proportion for Pre-menopausal Stages

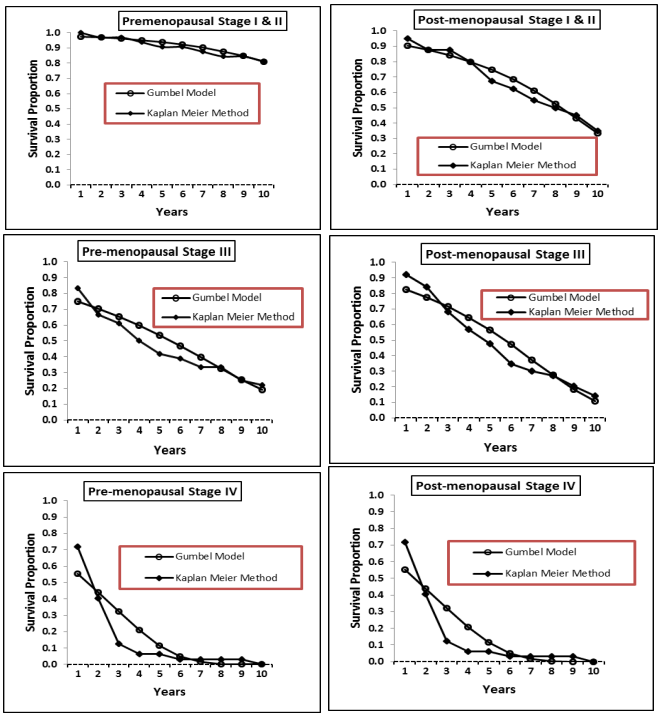
Time (Years)	Stage I & II		Stage III		Stage - IV	
	Kaplan-Mier	Gumbel	Kaplan-Mier	Gumbel	Kaplan-Mier	Gumbel
1	1.00000	0.975314	0.833333	0.750310	0.71875	0.552928
2	0.96875	0.968825	0.666666	0.705287	0.40625	0.441051
3	0.96875	0.960665	0.611111	0.654188	0.12500	0.322739
4	0.93750	0.950423	0.499999	0.597039	0.06250	0.209634
5	0.90625	0.937604	0.416666	0.534258	0.06250	0.115499
6	0.90625	0.921609	0.388888	0.466772	0.03125	0.050690
7	0.87500	0.901733	0.333333	0.396120	0.03125	0.016248
8	0.84375	0.877164	0.333333	0.324484	0.03125	0.003374
9	0.84375	0.846992	0.249999	0.254624	0.03125	0.000385
10	0.81250	0.810247	0.222222	0.189637	0.00000	0.000019

Table 8- Survival Proportion for Post-menopausal Stages

Year	Stage I & II		Stage III		Stage - IV	
	Kaplan-Mier	Gumbel	Kaplan-Mier	Gumbel	Kaplan -Mier	Gumbel
1	0.9500	0.904015	0.9206	0.824830	0.6862	0.550869
2	0.8750	0.876816	0.8412	0.776657	0.4313	0.464617
3	0.8750	0.842607	0.6825	0.717675	0.2156	0.373270
4	0.8000	0.800035	0.5714	0.647005	0.1372	0.281708
5	0.6750	0.747785	0.4761	0.564702	0.0784	0.196187
6	0.6250	0.684801	0.3492	0.472353	0.0784	0.123216
7	0.5500	0.610639	0.3015	0.373664	0.0784	0.067761
8	0.5000	0.525937	0.2698	0.274718	0.0391	0.031414
9	0.4500	0.432963	0.2063	0.183463	0.0391	0.011693
10	0.3500	0.336041	0.1428	0.107999	0.0391	0.003282

Graphical Representation of Survivor Rate Estimates for Different Stages of Breast Cancer Using Gumbel Distribution Model

The Graphical comparisons of Gumbel model (parametric) survivor-rate with Kaplan-Meier (non- parametric) model survivor-rate estimates given in the following figures.



Conclusions

Analysis shows that the Gumbel distribution is a reasonable model to describe the progression of breast cancer and finding survivor rates for 254 patients. Using Maximum likelihood method through unconstrained optimization method (DFP-Davidon-Fletcher-Powell) the parameter estimates and variance-covariance matrix for the Gumbel distribution model were found.

However unlike a number of two-parameter distributions which are used in survivor studies it does have some beaming on the physical process being described.

Acknowledgment

The author (Khizar H.Khan) thankfully acknowledges the support provided by the Department of Mathematics, College of Science

and Humanities, Salman Bin Abdulaziz University, Al-Kharj, and Ministry of Education, Saudi Arabia for providing the facilities and an environment to perform the research work.

References

[1] Boag, J.W. (1949) *J.R. Stat. Soc. Series* B, 1(11), 15 - 53.
[2] Bain, L.J.,Englehardt, M. (1991) *Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods,* 2, Marcel Dekker.
[3] Coles S. (2001) *An Introduction to Statistical Modelling of Extreme Values.* Springer-Verlag.
[4] Gupta R.D., Kundu D. (2007) *Journal of Statistical Planning and Inference* 137,3537-3547.
[5] Kotz S., Nadarajah S. (2000) *Extreme Value Distributions: Theory and Applications.* Imperial College Press.
[6] Nadarajah S. (2006) *Environmetrics* 17,13 - 23
[7] Kaplan E.L.,Meier P.P. (1958) *J. Amer. Statist. Assoc*., 53, 457-481.
[8] Lawless J.F. (1982) *Statistical Models and Methods for lifetime Data,* John Wiley and Sons.
[9] Meeker W.Q., Escober L.A., *Stanford J.* and Vardeman S. (1994)
[10]Hancock K., Peet B.G.,Price J.,Watson G. W., Stone J., Turner R. L. (1977) *British Journal of Surgery*, 64, 134 - 138.
[11]Cox D.R.,Oaks D. (1984) *Analysis of Survival Data.* London Chapman and Hall.
[12]Davidon W.C. (1959). *Mathematical Programming*, 9,1-30.
[13]Fletcher R.,Powell M.J.D. (1963) *The Computer Journal,* 6, 163-168.
[14]Khan K.H.,Zafar Mehmud (2002) *An International Journal*,1 (2),201-209.
[15]Watson G. W.,Turner R. L. (1959) *British Medical Journal*, 1, 1315 - 1320.

Table 4-Data for Stages I to IV over the ten years

Time (Years)	Stage-I &II Pre- menopausal		Stage-I & II Post-menopausal		Stage-III Pre-menopausal		Stage-III Post-menopausal		Stage-IV Pre-menopausal		Stage-IV Post-menopausal	
	Survivers	Failures	Survivers	Failures	Survivers	Failures	Survivers	Failures	Survivers	Failures	Survivers	Failures
0	32	0	40	0	36	0	63	0	32	0	51	0
1	32	0	38	2	30	6	58	5	23	9	35	16
2	31	1	35	3	24	6	53	5	13	10	22	13
3	31	0	35	0	22	2	43	10	4	9	11	11
4	30	1	32	3	18	4	36	7	2	2	7	4
5	29	1	27	5	15	3	30	6	2	0	4	3
6	29	0	25	2	14	1	22	8	1	1	4	0
7	28	1	22	3	12	2	19	3	1	0	4	0
8	27	1	20	2	12	0	17	2	1	0	2	2
9	27	0	18	2	9	3	13	4	1	0	2	0
10	26	1	14	4	8	1	9	4	0	1	2	0