Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.3, No.14, 2013

A Comparison Of Kaplan-Meier Product Limit Estimation And Markov Process In Estimating Survival Time Of A Cohort Of Dogs With Rabies

Thomas Adidaumbe Ugbe[1*] Stephen Sabastian Akpan[2]

1. DEPARTMENT OF MATHEMATICS/STATISTICS AND COMPUTER SCIENCE UNIVERSITY OF CALABAR, CALABAR, CROSS RIVER STATE, NIGERIA
2. DEPARTMENT OF MATHEMATICS/STATISTICS AND COMPUTER SCIENCE UNIVERSITY OF CALABAR, CALABAR,CROSS RIVER STATE,NIGERIA

***** Email of the corresponding author**:** **ugbe_thomas@yahoo.com**

ABSTRACT
In this research two methods were used to compute survival time of a cohort of 84 dogs with rabies. In carrying out Kaplan-Meier analysis, euthanasia was equated with death. Death (including subjects) recorded were treated as occurring prior to the exact time they occurred while censored (subjects lost to follow-up) were treated as occurring later than they time they actually occurred. Survival probability estimates and variance were calculated. In Markov process method, a 5 state time homogeneous Markov chain was used, the fundamental matrix which was obtained by summing counts recorded based on the number of dogs making respective transition was used to obtain probability matrix. The N matrix was calculated as well as its variance. The variances of the two methods computed served as a basis for comparing the efficiency of the two methods. Hence, the Kaplan-Meier product limit estimation with the smaller variance is more efficient in estimating survival time
Keywords: Transition probability; Efficiency; Censored; Euthanasia; Fundamental matrix;

Introduction

The Kaplan-Meier estimator is named after Edward L. Kaplan and Paul Meier. It dates back to 1952 when Paul Meier at John Hopkins University (now University of Chicago) stumbled on Green wood's paper on the duration of cancer. Since then many researchers have brought forward different ideas that has made the Kaplan-Meier estimator what it is today. Hillis et al (1986) defined Kaplan-Meier product limit model of survival analysis as a simple stochastic process that is defined by a set of transition matrices that contain probabilities of transition from state (alive) to state (dead). Over the years investigators almost exclusively use Kaplan-Meier product limit estimation for evaluation of time-event data; Thompson and Fugent (1992) ,Al-sarraf et al. (1996), Khanna et al. (1998). Veterinary studies include a high frequency of euthanized subjects, sometimes over 50% of the total observations are from euthanized animals. Cox et al. (1991), Berg et al. (1992), Zwahlen et al. (1998). While death remains the outcome of primary interest in such studies, investigators are inconsistent in their attention to euthanasia. Investigators have ignored deleted, censored or simply equated with death any observation that terminates in euthanasia Al-sarraf et al. (1996), Khanna et al. (1998). Other investigators acknowledge that euthanasia posed an analytical problem ,Slater et al. (2001), staatz et al. (2002). But choose to ignore the problem and equated euthanasia with death.

An alternative strategy to describe and evaluate time event data is the use of Markov models.Markov models have been used to describe human disease processes such as the evaluation and description of diabetics, retinopathy, renal disease, papilloma virus and human immunodeficiency virus Hendriks et al. (1996), Markov models can be used to describe disease as a series of probable transitions between health states. This methodology has considerate appeal for use in veterinary clinical studies since it offers a method to evaluate multiple health states simultaneously. In addition it potentially offers a method to accommodate observations from euthanized animals by recognizing euthanasia as a concurrent outcome of interest. In this work interest is in comparing Kaplan-Meier product Limit estimation and markov process in estimating survival time of a cohort of Dogs with rabies.

2. MATERIALS AND METHODS

2.1 Introduction

Kaplan-Meier survival analysis also known as the Kaplan-Meier product limit estimate can be used to estimate survival. The Kaplan-Meier method involves tracking the fates of individuals over time and estimating how long it

takes for death to occur. This method has been applied broadly to measure how long it takes for any specific event to occur; such as the time until a cancer patient recovers from a treatment, the time until an infection appears, the time until pollination occurs, and so on. The Kaplan-Meier method is conceptually similar to life tables calculations because you keep track of the number of individuals active and the number of deaths that occur over intervals of time. Specifically, you count the number of individuals who die at a certain time and divide that number by the number of individuals that are "at risk" (alive and part of the study) at that time.
In this research two methods were used to compute survival time of a cohort of 84 dogs with rabies. The data used for the study was got from Kolay's veterinary services located at No.14 Barracks Road Calabar, Cross River State (Southern Nigeria), Nigeria.

2.2 Kaplan-Meier Product Limit Formula

Let t_1, t_2, t_3 …… denoted the actual times of death of the individuals in the cohort. Also let $d_1,d_2,d_3,$…… denote the number of deaths that occur at each of these times, and let n_1, n_2, n_3…….. be the corresponding number of patients remaining in the cohort.
Note that $n_2 = n_1 - d_1$ …… …. (1)
$n_3=n_2-d_2$, etc…… (2)
Then, loosely speaking,
$S(t_2)$= P(T>t_2)= "probability of surviving beyond time t_2"…… …... (3)
depends conditionally on,
$S(t_1)$= P (T>t_1)= "probability beyond time t_1" …… . (4)
Likewise,
$S(t_3)$=P(T>t_3)="probability of surviving beyond time t_3"…… (5)
depends conditionally on,
$S(t_2)$=P(T>t_2)="probability of surviving beyond time t_2" etc…… …(6)
By using this recursive idea, we can iteratively build a numerical estimate $\hat{S}(t)$ of the true survival function S(t). Specifically; for any time $T \in (o,t_1)$ we have
S(t)= P(T>t)= "probability of surviving beyond time t" = 1,…… (7)
Because no deaths have yet occurred. Therefore, for all t in this interval, let $\hat{S}(t)=1$…… (8)
We know from elementary probability that P(A and B) = P(A) x P(B/A)…… (9)
Let A= "survive from time t_1" and B= "survive from time t_1" to beyond some time t before t_2". Having both events occurs is therefore equivalent to the event "A and B"= "survive to beyond time t before t_2" ie "T>t". Hence, the following holds;

(i) For ant time $T \in (t_1,t_2)$, we have;
S(t) = P(T>t) = P[survive in (o,t_1)] x P[survive in (t_1,t) / survive in (o,t_1)]

i.e $\hat{S}(t) = 1 \quad \frac{n_1 - d_1}{n_1}$, or

$\hat{S}(t) = 1 - d_1$ (10)

Similarly, for ant time $t \in (t_2, t_3)$, we have ;
S(t) = P(T>t) = P[survive in (t_1, t_2) x P[survive in (t_2, t)/ survive in (t_1,t_2)]

i.e $\hat{S}(t) = \left(1-\frac{d_1}{n_1}\right)$ x $\left(1-\frac{d_2}{n_2}\right)$ or

$\hat{S}(t) = \left(1-\frac{d_1}{n_1}\right)\left(1-\frac{d_2}{n_2}\right)$, etc (11)

In general, for $t \in [t_j, t_j+1]$, j= 1,2,3……,we have ;

$$\hat{S}(t) = \left(1-\frac{d_1}{n_1}\right)\left(1-\frac{d_2}{n_2}\right)\ldots\ldots\ldots\ldots\ldots\left(1-\frac{d_j}{n_j}\right)$$

$$= \prod_{i=1}^{j}\left(1-\frac{d_i}{n_i}\right) \qquad (12)$$

The Kaplan-Meier estimator $\hat{S}(t)$ can be regarded as a point estimate of the survival function S(t) at any time t. In order to calculate the variance of the Kaplan-Meier product limit estimator, the Greenwood's formula is used and is given by

$$\text{var}\,[S(t)] = \hat{S}^2(t) = \sum_{t_i < t} \frac{d_i}{n_i(n_i - d_i)} \qquad (13)$$

where n_i is the number of subjects at risk at the beginning of the period t_i
d_i is the number of subjects who die during the time period t_i Greenwood (1926).

2.3 Markov Modelling And Analysis

Matrix solution provides an exact solution of the time spent in each state conditional on the entry state in which an individual enters the model. Matrix solution is restricted to time homogeneous Markov chain. The transition probability matrix of a chain that contains absorbing states is divided into four sections. Q contains transition probabilities between transient states; R contains transition probabilities from transient to absorbing states; O is a zero matrix, and I is an identity matrix Brown and Brown (1990a).

Table 1

From \ To	Transient state	Absorbing state
Transient state	Q	R
Absorbing state	O	I

The average number of cycles that a subject resides in transient state before absorption, given a specified starting state is estimated from the fundamental matrix (N). Calculating N is the inverse of the transient probabilities in Q Brown and Brown (1990b). The N matrix specifies the average number of cycles that a subject reside in transient state such that $N= (1-Q)^{-1}$ where 1 is identify matrix and Q is the square matrix of the transient probabilities within P. multiplication of the number of cycles by the length of the cycle gives the expected duration in each state conditional on a starting state. The sum of these durations gives an estimate of expected survival conditional on a starting state Beck and Pauker (1983).

Theorem**:** If survival is considered as a two state Markov chain, and P is the probability of death, 1-P is the probability of surviving during the cycle. Than the variance of the N matrix with more than two states is $V(N)= N(2N-1) –N^2$

Proof**:** If P is the probability of death and 1-P is the probability of surviving during that cycle. The transition matrix is $\begin{bmatrix} 1-p & p \\ 0 & 1 \end{bmatrix}$.

Since survival is the waiting time for the first occurrence of death, the number of cycles that a subject survives can follow a geometric distribution (Hogg and Craig 1978).

For geometric distribution; P(surviving X cycles = X;p) = $P(1-P)^{x-1}$

where 1,2,3,4……

and E(x) = $\frac{1}{p}$, var(x) = $\frac{1-p}{p^2}$

$$\Rightarrow \text{Var}(x) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1}{p^2}(2-p) - \frac{1}{p^2}.$$

Since the expected survival time (time to death) is given by the inverse of the probability of death, this can be likened to taking the Q matrix to determine the N matrix Beck and Pauker (1983).

If N = $\frac{1}{p}$, then

$\text{Var}(N) = N^2 (2-N^{-1}) - N^2$

$= N(2N-1) - N^2$

3. Results And Discussion

In the context of the Kaplan-Meier procedure, euthanasia is equated with death and the subjects who become unreliable are spoken of as being censored, these are subjects which were lost to follow-up before the end of the study. It is impossible to know whether or not censored subjects survived or died hence omitting them from the analysis will amount to losing valuable information. We recognized that any attempt to salvage information from subjects which were lost to follow-up would involve a certain amount of "fudging", hence, we proposed that subjects who became unavailable during a given time period can be counted among those who survive through the end of that period but then deleted from the number who are at risk for the next time period. That is, deaths recorded at time t are treated as if they occurred slightly before t and loss recorded as of time t are treated as occurring slightly after t. In this way fudging is kept conceptual, systematic and automatic.

The table below shows a vivid explanation of the proposition

Table 2 showing the Kaplan-Meier proposition

Time of events (days)	**At risk**	**Censored**	**Died**	**Survived**
6	84	1	5	79
13	78	2	8	70
21	68	1	13	55
27	54	1	5	49
32	48	1	2	46
39	45	1	4	41
43	40	0	3	37
89	37	1	12	25
92	24	1	3	21

From table 2, 84 subjects (Dogs) who were at risk at the beginning of the study, 1 becomes unreliable (censored) during the first 6 days of the study, and 5 died. The number surviving the first 6 days is therefore 79 and the total number at risk after this period is 78. Another 2 subjects became unreliable (censored) during the next 7 days and other 8 died. Hence the number surviving the next seven days is 70 and the number at risk at the end of the period is 68. The same applies for the other time periods shown. Next we calculate the survival probability estimates for each of the time periods. Apart from the first period the rest will be calculated as a compound conditional probability. Table 3 illustrates this.

Table 3 showing computation of Kaplan-Meier survival probability estimate $\hat{S}(t)$

Time of event (days)	**At risk ($n_i+1=n_i-d_i-c_i$)**	**Censored (c_i)**	**Died (d_i)**	**Survival (n_i+1-d_i)**	**Survival probability $\frac{n_i+1-d_i}{N_i+1}$**	**Kaplan-Meier probability estimate $\hat{S}(t)$**
6	84	1	5	79	79/84	0.9405
13	78	2	8	70	70/78	0.8440
21	68	1	13	55	55/68	0.6826
27	54	1	5	49	49/54	0.6194
32	48	1	2	46	46/48	0.5936
39	45	1	4	41	41/45	0.5408
43	40	0	3	37	37/40	0.5002
89	37	1	12	25	25/37	0.3380
92	24	1	3	21	21/24	0.2958

The above table (table 3) illustrate the logic of the procedure. The Kaplan-Meier product estimation using the formula given in section 2.2 is 0.0057 and variance of the Kaplan-Meier estimator also given in section 2.2 is 0.0000010.

In Markov modeling, the sum of the expected duration time spent in transient state before absorption is the expected survival of the cohorts of subjects. For the purpose of this study five health states were define as follows.

Table 4 definition of health states to describe a cohort of Dogs with rabies

State	**Definition**	**Type**
WELL	WELL	TRANSIENT
TOXIC	TOXIC	TRANSIENT
LTF	LOST TO FOLLOW UP	ABSORBING
DEAD	DEAD	ASORBING
EUTH	EUTHANIZED	ABSORBING

Using the fundamental matrix solution, for each cycle, a count was recorded based on the number of dogs making the respective transitions. The counts were summated to give the overall summation matrix (S) and the probability matrix was obtained thereafter. The two matrices are displayed below in table 5a and 5b

Table 5a Summation Matrix

	TOXIC	WELL	EUTH	DEATH	LTF
TOXIC	318	39	9	44	9
WELL	8	61	0	2	0
EUTH	0	0	0	0	0
DEAD	0	0	0	0	0
LTF	0	0	0	0	0

Table 5b Probability Matrix

	TOXIC	WELL	EUTH	DEATH	LTF
TOXIC	0.7589	0.0931	0.0215	0.1050	0.0215
WELL	0.1127	0.8592	0	0.0282	0
EUTH	0	0	1	0	0
DEAD	0	0	0	1	0
LTF	0	0	0	0	1

The variance of N matrix stated earlier in section 2.3 was 0.0269

4. Conclusion

From the result of the variances obtained, we see that the variance of Kaplan-Meier estimate (0.0000010) is less than the variance of N matrix (0.0269) this shows that Kaplan-Meier estimate with smaller variance is more efficient and a better estimator of the survival time. See, for example, Hogg and Craig (1978) on efficiency of estimators. Also, the Kaplan-Meier estimators provides estimates of survival time that can be partitioned according to the reason for loss and according to the health state of the animals, and data used has low concentration of censored and euthanized subjects.

References

Al-sarraf, R., Mauldin, G. N., Patnaik, A. K.and Meleo ,K. A. (1996). "A prospective study of radiation therapy for the treatment of grade 2 mast cell tumors in 32 dogs," *J. Vet Int. Med*. **10,** 376-378.

Beck, J.R. and Pauker, S. G. (1983). "The Markov process in Medical prognosis," *medical decision making*, **3**, 419-458.

Berg, J., Weinstein, J., Schelling S.H.and Rand, W.M. (1992). "Treatment of Dogs withOsterosarcomo by administration of Cisplatin after computation or limb-sparing surgery," (1987- 1990). *J. AM Vet med. Assoc***.,** 2001, 2003- 2008.

Brown, R. F.and Brown, B.W. (1990a), "Absorbing Markov Chains. In: essentials of finite mathematics: Matrics, Linear programming, probability, Markov chains," *Ardsley House, New York*, 399-412.

Brown, R. F.and Brown, B. W. (1990b), " The fundamental fundamental matrix. In: Essentials of finite mathematics: Matrices, Linear programming, probability, markov chains," *Ardsley House New York*, 412-423..

Cox, N. R., Brawner, W. R., Powers, R. D.and Wright, J. C. (1991), " Tumors of the nose and paranasal sinuses in cats: 32 cases with comparison to a national database (1977 through 1987)," *J. AM Anim Hosp Assoc*., **27**, 339-347.

Greenwood, M. (1926), "The natural duration of cancer reports on public health and medical subjects 33".

Hendriks, J.C.M., Satten,,G.A., Longin, I. M., Van, Druten H.A.M., Schellekens, P.T.A., Coutinho, R. A.and Van, greensven G.J.P (1996), " Use of immunological markers and continuous-time Markov models to estimate progression of HIV infection in homosexual men," *AIDS*, **10**,649-656.

Hillis A., Maguire, M., Hawkins, B. S.and Newhouse, M. M. (1986), " The Markov process as a general method for nonparametric analysis of right-censored medical data, " *J. chronic disease*, **39**, 595-604.

Hogg, R. V.and CRAIG, A.T. (1978), " Introduction to mathematical statistics, "*4th edition, new York: Macmillan publishing Co.Inc*.

Khanna, C., Lund, E.M., Redic, K.A., Hayden, D.W., BELL, F.W. Goulland, E.L. and Klausner, J.S. (1998), " Randomized controlled trial of doxonubicin versus dactinomycin in a multiagent protocol for treatment of dogs with malignant lymphoma," *J.AM. Vet. Med. Assoc*., **213**, 985-990.

Slater, M.R., Geller, S. and Rogers, K. (2001), " Long-term health and predictors of survival for hyperthyroid cats tested with iodine 131," *J Vet Int. Med*. **15**,47-51.

Staatz, A.J., Monnet, E.and Seim, H. B. (2002), " Open peritoneal drainage versus primary closure for the treatment of septic peritonitis in dogs and cats: 42 cases (1993-1999)," *Vet surg*, **31**, 174-180.

Thompson, J.P.and Fugent, M. J. (1992), " Evaluation of survival times after limb amputation, with and without subsequent administration of cisplatin, for treatment of appendicular osterosarcoma, in dogs: 30 cases (1979-1990)." *J.AM Vet Med. Assoc*. **200**, 531-533.

Zwahlen,C.H.,Lucroy,M.D.,Kraegel,S.A.,madewell,B.R.(1998), " Results of chemotherapy for cats with alimentary malignant lymphoms: 21 cases (1993-1997)," *J. AM Vet med Assoc*. **213**, 1144-1149