

Some Methods for Summarizing Survivorship Data in Nonstandard Situations

A. P. GORE*¹

Department of Mathematics and Statistics, University of Calgary, Canada

SHARAYU PARANJAPE, M. B. RAJARSHI

Department of Statistics, University of Poona, India

MADHAV GADGI

Centre for Ecological Studies, Indian Institute of Science

Summary

One difficulty in summarising biological survivorship data is that the hazard rates are often neither constant nor increasing with time or decreasing with time in the entire life span. The promising Weibull model does not work here. The paper demonstrates how bath tub shaped quadratic models may be used in such a case. Further, sometimes due to a paucity of data actual lifetimes are not as certainable. It is shown how a concept from queuing theory namely first in first out (FIFO) can be profitably used here. Another nonstandard situation considered is one in which lifespan of the individual entity is too long compared to duration of the experiment. This situation is dealt with, by using ancillary information. In each case the methodology is illustrated with numerical examples.

Key words: Survivorship data; Hazard rates; Weibull model; Bath tub shape, Chi-square test.

1. Introduction

One problem that has interested ecologists considerably is that of modelling life data. Often this is done by fitting a suitable probability distribution to such data. Alternatively a smooth function can be chosen to fit the age specific mortality rates of the so called hazard rates. Three commonly visualized varieties of hazard rates are (i) increasing with age (ii) decreasing with age (iii) constant or age independent. It has been noticed that the Weibull distribution model for life data is flexible enough to accommodate all these variations with suitable choices of parameter values and hence has been successfully deployed to summarise many data sets. (PINDER et al., 1978).

There are two difficulties likely to be encountered in this approach. Firstly, in the biological realm often hazard rates are not monotone. There is high mortality among infants while the death rate is quite low among young adults. It increases again as old age sets in. A so called bath tub shape model for hazard rates is more

* On leave of absence from University of Poona.

¹ First author's research supported in part by an NSERC grant of Dr. J. K. WANI.

appropriate in such situations. (Strictly speaking hazard rates become very very high as the chance of survival reduces to zero in extreme old age. Hence they cannot assume a symmetric shape like the cross section of a bath-tub. However, the nomenclature appears to be well established now.) Weibull model will work in such cases only if early deaths are ignored, as done by Pinder et al. They further state that "combining juveniles and adults usually produces a two phased survivorship curve . . . (which) cannot be modelled by a simple distribution". This seems to be true. But it is sometimes possible to directly fit a suitable hazard function instead of a probability distribution. An example of this type is presented below.

Another difficulty in modelling life is either a paucity of data or availability of only limited information. This may happen for various reasons. Lifespan may be too long compared to the duration of experiment. Specific age may not be ascertainable and only life stages can be identified. Further, there are situations in which ages of individuals are not available so that usual methods of analysis cannot be applied. In such cases an alternative approach may be necessary.

Our note consists of three examples. The first gives a data set with usual type of observations and shows that the Weibull model gives a poor fit while a quadratic hazard model provides a considerable improvement. In the second example sample plots are observed once every day and increase or decrease in the number of heaps of deer droppings is recorded. It is not possible to count the number of days in which an individual pellet heap gets decomposed. However, here the duration of the experiment is much longer than the typical "life span" of an individual pellet heap. In the third example, the age of the individual culm of bamboo cannot be ascertained but only arrivals and stages in life can be observed. Further, the duration of the experiment is much shorter than a typical life span.

2. Bath Tub Model for Survival of Dall's Mountain Sheep

We consider the data on life length of 464 sheep (whose exact age at death could be ascertained) given by MURIE (1944, p. 123). Weibull distribution was fitted to these observations using the methods given by JOHNSON and KOTZ (1970). The location parameter of the Weibull distribution, in this case is zero while the estimates of scale and shape parameters respectively are $\hat{\alpha} = 7.9355$ and $c = 1.47$.

For the Weibull distribution the differences between observed and expected frequencies are very large and goodness of fit chi-square is 678 (8 d.f.) which is unacceptable.

To fit the quadratic hazard model, we adopt the method described in detail by LAWLESS (1982). Here we first calculate the empirical survival function $S_n(t)$ evaluated at upper class limits, $S_n(t)$ being the sample proportion of individuals alive at time t , assuming that we began with n individuals all born at time zero. From $S_n(t)$ we obtain the empirical cumulative hazard function $H_n(t) = -\log S_n(t)$.

Table 1

Fitting of Weibull and quadratic hazard model to Dall's sheep data

| age at death | observed frequency | expected frequency (Weibull) | expected frequency (quadratic hazard) |
|--------------|--------------------|------------------------------|---------------------------------------|
| 0-2 | 121 | 21.5 | 115.3 |
| 2-3 | 7 | 55.9 | 18.2 |
| 3-4 | 8 | 42.8 | 7.9 |
| 4-5 | 7 | 42.9 | 5.4 |
| 5-6 | 18 | 41.6 | 10.3 |
| 6-7 | 28 | 38.8 | 20.9 |
| 7-8 | 29 | 35.1 | 35.1 |
| 8-9 | 42 | 31.6 | 48.3 |
| 9-10 | 47 | 27.4 | 55.7 |
| 10-11 | 66 | 23.6 | 52.7 |
| 11-15 | 90 | 103 | 94.4 |

A cubic function $at + b\frac{t^2}{2} + c\frac{t^3}{3}$ is fitted to the values of $H_n(t)$ by least squares.

Derivative of this function (i.e. $a + bt + ct^2$) is our model for the age specific hazard rate. This model must therefore satisfy the universal properties of hazard rates namely that they are nonnegative at all ages and extremely large at very old ages. For this, a and c must be both nonnegative and b must be at least $-2\sqrt{ac}$. If the fitted value of b is not large enough, it is replaced by the above lower bound as suggested by BAIN (1978). For our present data the fitted values are $\hat{a} = 0.2246$, $\hat{b} = -0.0968$ and $\hat{c} = 0.01122$. Having obtained the hazard function, the survival function can be estimated working backwards. Expected frequencies are obtained as $n(\hat{S}_n(t-1) - \hat{S}_n(t))$ for the age class $(t-1)$ to (t) , taking $\hat{S}_n(0) = 1$. The expected values obtained are remarkably better than those under the Weibull model. In fact the difference between observed and expected frequencies (in absolute values) is less for the quadratic model, in every age class. The goodness of fit chi-square is 23.2 (7 d.f.) which is very much smaller than the corresponding value for the Weibull model (though it is still beyond the 1% value namely 18.3). The quadratic hazard model seems to perform better.

This example and the preceding discussion suggest that fitting a quadratic hazard function is computationally simpler than fitting the Weibull model. It is also conceptually more appropriate.

3. Estimating survival distribution for pellet heaps of deer assuming FIFO

Our second example concerns the population dynamics of heaps of pellets of Chital or axis deer (*Axis axis*). Pellet heap counts have often been used in the estimation of population size (GATES, 1979; SEBER, 1973). This suggests the study of (i) standing crop, (ii) rate of production and (iii) rate of decomposition or morta-

lity pattern of heaps. In a study conducted by one of us (MG) in the Bandipur Tiger Reserve in India, 20 sampling plots of 30 meters square each were observed once daily for 120 days in July–October 1977. Plots had been cleared prior to the experiment but no pellets were removed from plots once the experiment started. Daily records were maintained of the number of pellet heaps deposited in a plot over the previous 24 hours as well as the number of old pellet heaps still remaining in the plot. The initial plan was to estimate the daily output of pellet heaps. Subsequently it was desired to model the time to decay but the values of this variable for individual pellet heaps had not been observed.

The analysis began with the study of arrival pattern. The average number of arrivals in a plot per day varied from 0.066 to 0.697. This variation was expected as the deer did not spend their time uniformly in different parts of the reserve. For each plot a Poisson distribution was fitted for the number of arrivals in a day and a chi-square goodness of fit statistic was calculated. In 17 plots of 20, the p -value was more than 0.1. In all 20 plots p -value was more than 0.01. Daily arrival of pellet heaps on sample plots are thus satisfactorily described by the Poisson distribution with mean varying from plot to plot.

Perhaps the simplest model here would be a Poisson-Markov model (BARTLETT 1978). It assumes that (i) arrivals have a Poisson distribution, (ii) arrivals are density independent (iii) survival times are independent of one another and (iv) age specific death rates are constant. It can be shown by standard techniques (see HOEL, PORT and STONE (1972) pp. 53–56) that under this model, given an equilibrium condition the total occupancy (number of old as well as new pellet heaps in a plot on a day) has a Poisson distribution with mean λ/q where λ is the arrival rate and q is the constant probability of departure on any day. This was again subjected to chi-square tests for goodness of fit. The result was extremely high chi-square values in all cases. It was therefore felt that the assumption (iv) asserting a constant hazard rate had to be replaced. However, no direct observations on the length of life were available to try any alternative model for hazard rate. This difficulty was resolved by making one more assumption that pellet heap life lengths follow a First In First Out (FIFO) discipline. This simply means that when we notice that the number of old pellet heaps in a plot is reduced by 1 we assume that the heap that entered the plot earliest among the present ones, has disintegrated.

Assuming FIFO, pseudolifetimes were calculated. This essentially brings us back to the standard problem. Now a quadratic hazard function $h(t) = a + bt + ct^2$ can be fitted to the values. This was done for all 20 plots following the approach used in the first example. The chi-square values indicated a reasonably good fit in at least 14 out of 20 plots. A typical data set and associated analysis are given in Table 2 and 3.

Here the computed chi-square value (11.86) was not too large for 6 degrees of freedom ($p > .05$).

It should be pointed out that the question of whether the hazard function in

Table 2

Arrival (1) and Departure (2) Dates and Life Times (3) (FIFO) of Chital Pellets in a Typical Observation Plot

| 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
|----|----|----|----|----|----|-----|-----|----|
| 3 | 8 | 5 | 30 | 58 | 28 | 72 | 90 | 18 |
| 4 | 30 | 26 | 35 | 58 | 23 | 72 | 90 | 18 |
| 5 | 30 | 25 | 43 | 60 | 17 | 78 | 91 | 13 |
| 8 | 32 | 24 | 44 | 60 | 16 | 79 | 91 | 12 |
| 13 | 35 | 22 | 46 | 61 | 15 | 81 | 92 | 11 |
| 18 | 40 | 22 | 48 | 61 | 13 | 85 | 95 | 10 |
| 18 | 40 | 22 | 54 | 66 | 12 | 86 | 95 | 9 |
| 18 | 48 | 30 | 55 | 67 | 12 | 87 | 96 | 9 |
| 28 | 48 | 20 | 56 | 70 | 14 | 88 | 96 | 8 |
| 28 | 48 | 20 | 57 | 71 | 14 | 88 | 102 | 14 |
| 28 | 48 | 20 | 58 | 74 | 16 | 89 | 102 | 13 |
| 29 | 48 | 19 | 59 | 74 | 15 | 92 | 102 | 10 |
| 30 | 58 | 28 | 60 | 74 | 14 | 94 | 103 | 9 |
| 30 | 58 | 28 | 61 | 82 | 21 | 94 | 105 | 11 |
| 30 | 58 | 28 | 62 | 82 | 20 | 100 | 116 | 16 |
| 30 | 58 | 28 | 65 | 87 | 22 | 101 | 116 | 15 |
| 30 | 58 | 28 | 65 | 89 | 24 | 101 | 120 | 19 |
| 30 | 58 | 28 | 69 | 89 | 20 | 104 | 120 | 16 |

Table 3

Fitting Quadratic Hazard Function to Life Time (FIFO) of Pellets
 Hazard Function Fitted: $h(t) = 0.0297 - 0.0058t + 0.0006t^2$
 (1) Class interval (days). (2) Observed freq. (3) Expected freq.

| 1 | 2 | 3 | 1 | 2 | 3 |
|-------|----|------|-------|---|------|
| 0-2 | 0 | 3.55 | 15-17 | 8 | 8.02 |
| 3-5 | 1 | 2.24 | 18-20 | 8 | 8.79 |
| 6-8 | 1 | 2.47 | 21-23 | 6 | 7.88 |
| 9-11 | 7 | 3.93 | 24-26 | 5 | 5.70 |
| 12-14 | 10 | 6.08 | 27+ | 8 | 5.35 |

fact has a bath tub shape is also answered in the foregoing analysis. Thus if b and c are both close to zero, it suggests that hazard function is constant and death or decay is attributable to factors unrelated to age. If a , b , and c are all positive we know that hazard rate strictly increases with age. Of the 14 plots for which the quadratic hazard function gave a satisfactory fit 5 turned out to have a bath-tub shaped hazard while in case of 9 the hazard function was increasing with age.

4 Implications of FIFO assumption

It will be noticed that the foregoing analysis of the pellet heap counts is critically dependent upon the FIFO assumption. It seems reasonable to contend that other things remaining constant, an individual born earlier may die earlier. However, this is at best a general tendency and need not hold good in specific cases. The FIFO assumption may therefore introduce some distortion. To illustrate the point consider the following hypothetical data.

Table 4
Effect of FIFO Assumption on Estimates of Lifetimes

| Ser. No. | Arrival dates | true departure dates | true life times | ordered departure dates | FIFO lifetime |
|----------|---------------|----------------------|-----------------|-------------------------|---------------|
| 1 | 1 | 10 | 9 | 10 | 9 |
| 2 | 8 | 16 | 8 | 15 | 7 |
| 3 | 12 | 15 | 3 | 16 | 4 |
| 4 | 15 | 25 | 10 | 20 | 5 |
| 5 | 16 | 20 | 4 | 25 | 9 |
| | | | Total 34 | Total 34 | |
| | | | Average 6.8 | Average 6.8 | |

It can be seen that the FIFO assumption attributed the latest departure date to the last arrival thus shrinking the longest true lifetime value. Similarly the shortest lifetime value is slightly increased. However, the average lifetime was unaffected. All these features can be easily shown to hold in general. Thus FIFO assumption is quite acceptable if the main aim is to estimate the mean but it reduces the variability in the data. However, if we group the data the difference in the two frequency distributions is somewhat dampened. Thus for data in Table 4 if we use class intervals 0-5 and 6-10 then the class frequencies of true lifetimes (as well as pseudo lifetimes) are 2 and 3 respectively.

It appears that while FIFO assumption may affect lifetimes, it leaves the mean unaffected while variability may be reduced. However, an analytic assessment of the impact of FIFO assumption on estimated lifetimes, remains an open problem.

5. The life distribution of bamboo culms

The last example concerns a study of regeneration of bamboo *Bambusa arundinacea* in a particular locality in the state of Karnataka, India. The life span of an individual bamboo plant (clump) which is a cluster of shoots (culms) is four or five decades and can be ascertained through historical records of its once-in-life gregarious flowering just before death. The life distribution of individual culms is, however, not easy to record or study. The life span is of the order of 15 years and most field studies are of a much shorter duration. Further the age of an individual culm cannot be precisely ascertained. One can only classify a culm as being new (roughly 1 year old) or old (2 or more years in age). Table 5 gives observations on a typical transect for over four years.

We note that in this problem, methods adopted for the analysis of pellet heaps are inadequate. The number of departures is quite small and we do not have a

Table 5

Annual Clumpwise Record of Old and New Shoots of *Bambusa arundinacea* in a Transect in Karnataka (India)

| Clump No. | 1975 | | 1976 | | 1977 | | 1978 | |
|-----------|------|-----|------|-----|------|-----|------|-----|
| | New | Old | New | Old | New | Old | New | Old |
| 1 | 1 | 7 | 1 | 8 | 0 | 8 | 2 | 8 |
| 2 | 3 | 12 | 0 | 15 | 0 | 15 | 0 | 15 |
| 3 | 0 | 8 | 1 | 8 | 1 | 9 | 1 | 10 |
| 4 | 6 | 8 | 1 | 14 | 2 | 15 | 4 | 17 |
| 5 | 0 | 5 | 1 | 5 | 1 | 6 | 0 | 6 |
| 6 | 2 | 3 | 0 | 5 | 0 | 5 | 1 | 5 |
| 7 | 1 | 23 | 2 | 24 | 0 | 24 | 2 | 24 |
| 8 | 2 | 11 | 0 | 13 | 4 | 13 | 0 | 17 |
| 9 | 0 | 6 | 1 | 6 | 6 | 7 | 3 | 13 |
| 10 | 2 | 8 | 0 | 10 | 0 | 10 | 0 | 10 |
| 11 | 0 | 5 | 0 | 5 | 0 | 5 | 0 | 5 |
| 12 | 1 | 8 | 1 | 9 | 0 | 9 | 0 | 9 |
| 13 | 0 | 6 | 0 | 6 | 0 | 6 | 0 | 6 |
| 14 | 1 | 6 | 1 | 7 | 0 | 7 | 0 | 7 |
| 15 | 0 | 4 | 0 | 4 | 1 | 4 | 0 | 5 |
| 16 | 1 | 4 | 0 | 5 | 0 | 5 | 0 | 5 |
| 17 | 1 | 4 | 0 | 5 | 0 | 5 | 0 | 5 |
| 18 | 1 | 3 | 2 | 4 | 2 | 6 | 3 | 8 |
| 19 | 1 | 4 | 0 | 5 | 0 | 5 | 1 | 5 |
| 20 | 1 | 13 | 0 | 14 | 0 | 14 | 2 | 14 |

clean state to being with. We first check whether the arrivals follow a Poisson distribution. Table 6 gives the details for the year 1975.

Table 6

Testing Whether the Number of New Culms in a Clump Follows a Poisson Distribution for the Typical Transect in the Year 1975.

n (number of observations) = 20

λ (estimated average number of new culms per clump) = 1.2

| No. of new arrivals | Probability under Poisson Distribution ($\lambda=1.2$) | expected frequency | observed frequency |
|---------------------|----------------------------------------------------------|--------------------|--------------------|
| 0 | 0.3012 | 6.024 | 6 |
| 1 | 0.3614 | 7.228 | 9 |
| 2 | 0.2169 | 4.338 | 3 |
| 3 or more | 0.1205 | 2.410 | 2 |

chi-square = 0.92 $p > 0.1$

Similar results were obtained for different years and over different transects. Emergence of new culms in a clump is thus well described by a Poisson distribution. The average number of new culms (λ) varied from one transect to another and for a fixed transect from one year to another. In addition to this we only know the number of old culms in each clump. We must, therefore, come up with a model that will involve life distribution of culms and will predict the number of culms per clump. The observed and expected values (not frequencies) can then be compared using suitable tests. Notice that if in an ongoing system, we know the arrival distribution, the life distribution and the length of time for which the system is in operation we can predict the number of individuals present in the system.

SHARMA and TOMAR (1964) (also see PRASAD and GADGIL, 1981, p. 215) followed for 16 years a cohort of 439 individual culms of *Dendrocalamus strictus* another bamboo species, and recorded their ages at death. The resulting empirical life table is given in Table 7.

This could be used as a working hypothesis for the life distribution. For the means of Poisson arrival distributions and for the length of time for which the system is in operation, no such aid could be found and recourse was taken to seeking informed opinion of a naturalist working on bamboo growth in the concerned area. This revealed that clumps began to form around 1972 and typical number of culms per clump was about 3. The corresponding values for 1973 and 1974 were calculated by linear interpolation.

The expected number of culms in a clump in a year is calculated using these arrival rates and the empirical survival function. Thus the number of old culms in 1975 is the sum total of survivors from those which arrived in 1972, 73 and 74. Chance of death in the first year is .002. So, of the 1.80 that arrived a year before, 1.8 (1-.002) = 1.796 survive in 1975. Hence the expected number of old culms in

Table 7

Empirical Life Table of a Cohort of 439 Individual Culms of *D. strictus*. All Culms were 1 Year Old at the Beginning of Observation (adapted from SHARMA and TOMAR, 1964)

| Age | no. of survivors | no. of deaths | proportion of deaths | Age | no. of survivors | no. of deaths | proportion of deaths |
|-----|------------------|---------------|----------------------|-----|------------------|---------------|----------------------|
| 1 | 439 | 0 | 0 | 9 | 229 | 57 | 0.129 |
| 2 | 438 | 1 | 0.002 | 10 | 168 | 61 | 0.138 |
| 3 | 433 | 5 | 0.011 | 11 | 106 | 62 | 0.141 |
| 4 | 427 | 6 | 0.013 | 12 | 46 | 60 | 0.136 |
| 5 | 410 | 17 | 0.038 | 13 | 20 | 26 | 0.059 |
| 6 | 380 | 30 | 0.068 | 14 | 9 | 11 | 0.025 |
| 7 | 340 | 40 | 0.091 | 15 | 1 | 8 | 0.018 |
| 8 | 286 | 54 | 0.123 | 16 | 0 | 1 | 0.002 |

Table 8

Fitting of Survival Distribution of *D. strictus* to Data on *B. arundinacae*

| Year | estimated rate of new arrivals (λ) | Mean number of culms per clump | | $F(4, 16)$ |
|------|----------------------------------------------|--------------------------------|----------|------------|
| | | expected | observed | |
| 1972 | *3 | | | |
| 1973 | γ 2.4 | | | |
| 1974 | γ 1.8 | | | |
| 1975 | 1.2 | 7.09 | 7.4 | |
| 1976 | 0.55 | 8.12 | 8.6 | 0.44 |
| 1977 | 0.85 | 8.34 | 8.9 | |
| 1978 | 0.95 | 8.65 | 9.7 | |

* prior assumption γ linear interpolation

1975 is given by $3(1 - 0.026) + 2.4(1 - 0.013) + 1.8(1 - 0.002) = 7.09$. Thus we have a vector of expected number of old culms per clump (for years 1975-78). Corresponding observed values are simply the averages of 20 values for the 20 clumps. Each clump constitutes an independent experimental unit for the present. To apply the Hotelling's T^2 test (see KSHIRSAGAR, 1972, p. 142) we need estimates of variances and covariances (covariance matrix) of the sample averages. These can be obtained using the four sets of old culm counts (one set for each year). Results of these computations are given in Table 8.

The calculated F statistic is quite small indicating a good agreement between observed and expected values. Our conclusion is that the empirical life table given by Sharma and Tomar is reasonably good to describe the survival of culms of *B. arundinacae*. If the F statistic had been too large, no useful conclusion would have emerged. For it would simply mean that one or more of the premises needed modification. If on the other hand one knew for certain that the empirical life table was the only unsubstantiated assumption, attempts could be made to change it suitably. We would of course not recommend this approach if it were possible to extend the observations sufficiently so that direct values of lifetimes could be obtained.

References

- BAIN, L. J., 1978: Statistical analysis of reliability and life testing models. Marcel Dekker, N. Y.
- BARTLETT, M. S., 1978: An introduction to stochastic processes with special reference to methods and application. Cambridge University Press.
- GATES, C. E., 1979: Line transects and related issues. pp. 71-154 in R. M. CORMAC, G. P. PATIL and D. S. ROBSON (Eds.) Sampling biological populations. International Cooperative Publishing House, Fairland, Md.
- HOEL, P. G., PORT, S. C., STONE, C. J., 1972: Introduction to stochastic processes. Houghton Mifflin, Boston.
- JOHNSON, N. L. and KOTZ, S., 1970: Distributions in statistics, 1. Continuous univariate distributions. Houghton Mifflin, Boston.
- KSHIRSAGAR, A. M., 1972: Multivariate analysis. Marcel Dekker, N. Y.
- LAWLESS, J. F., 1982: Statistical models and methods for lifetime data. Wiley, N. Y.
- MURIE, A., 1944: The wolves of Mount McKinley Park. U.S. Dept. of Int. Nat. Park Service, Washington, D. C.
- PINDER, J. E. III, WIENER, J. G. and SMITH, M. H., 1978: The Weibull distribution: a new method of summarising survivorship data. *Ecology* **59**, 175-179.
- PRASAD, S. N. and GADGIL, MADHAV, 1981: Conservation of bamboo sources in Karnataka. Tech. Rep. Indian Institute of Science, Bangalore, India. 275 pages.
- SEBER, G. A. F., 1973: The estimation of animal abundance. Hafner, N. Y.
- SHARMA, N. K. and TOMAR, M. S., 1964: Bamboo forests of Madhya Pradesh. Proceedings of All India Bamboo Study Tour & Symposium. Forest Research Institute, Dehra Dun, India.

Manuscript received: Sept, 9, 1984

Authors' addresses:

Dr. A. P. GORE
The University of Calgary
Faculty of Science
Department of Mathematics & Statistics
2500 University Drive N. W.
Calgary, Alberta, Canada T2N 1N4

SHARAYU PARANJAPE, M. B. RAJARSHI
University of Poona
Department of Statistics
Pune - 411007, India

MADHAV GADGIL
Centre for Ecological
Studies, Indian Institute
of Science
Bangalore, India