B.A.J. **2**, I, 129–155 (1996)

AN ACTUARIAL SURVEY OF STATISTICAL MODELS FOR DECREMENT AND TRANSITION DATA

I: MULTIPLE STATE, POISSON AND BINOMIAL MODELS

BY A. S. MACDONALD, B.Sc., Ph.D., F.F.A.

ABSTRACT

This paper surveys some statistical models of survival data. A basic model of a random lifetime is defined, and censoring is introduced. Methods based on observations of small segments of lifetimes are compared. Markov and semi-Markov (multiple state) models are recommended as well-understood and flexible models well suited to actuarial data. A Poisson model is discussed as an approximation to a two state model, while traditional Binomial-type models are shown to be more restricted and less tractable than multiple state models.

KEYWORDS

Actuarial Estimate; Balducci Assumption; Censoring; Markov Models; Multiple State Models; Semi-Markov Models

1. INTRODUCTION

This paper surveys, in three parts, some work on survival models of current or potential interest to actuaries. It is not a review article in the sense of covering new research work; it includes work which is far from recent and there is more explanatory material than would be usual. Its purpose is to make modern tools in mortality analysis more accessible to actuaries, and to indicate some recent lines of development. Nothing in the paper is original, indeed much of it is elementary. It, nevertheless, seems useful to try to show how several apparently separate aspects of the subject form a coherent whole.

Part I deals with the type of data most familiar to actuaries, namely observations of small segments of lifetimes, typically single years of age. It includes the statistical treatment of multiple state models, of which multiple decrement models are a special case. Part II describes competing risks models and compares them with multiple state models, and describes the non-parametric and semi-parametric approaches to analysing observations of complete lifetimes, with emphasis on the Kaplan–Meier estimate and the Cox model. In Part III we describe some modern counting process models which provide a unifying framework for almost all of the preceding material.

The list of references for all three parts is given at the end of Part I.

Part II appears in *British Actuarial Journal*, Volume **2**, Part II and Part III in *B.A.J.* **2**, III.

Much of the long history of mortality analysis can be traced to the actuarial applications of life tables. Like most of science and mathematics,

however, by far the larger part of the *corpus* is of recent origin. Haycocks & Perks (1955) introduced their textbook in the following terms:

> "This book is about the principles and methods of actuarial statistics. By 'actuarial statistics' we mean the statistics that actuaries compile for the purposes of their professional work. Our subject is thus a severely practical one, and the methods used are such as are sufficient for the practical purposes to be served. Elaborate theoretical development would be inappropriate for our purpose; utilitarianism is the keynote and approximation pervades the whole subject. The modern developments of mathematical statistics have made hardly any impact in this field ..."

Since then, mathematical statistics has made an enormous impact in this field, and survival analysis has developed rapidly into a major branch of statistics, but it has received little attention from actuaries. It is worth considering why, and whether or not, it matters.

The analysis of mortality in generally healthy populations is rather insensitive to the methods employed. Any of the models discussed here, applied to typical actuarial mortality data, will usually give numerical results so similar as to be the same for all practical purposes, particularly if the results are to be graduated. It could, therefore, be argued that actuaries have no practical need of the more considered analyses made by statisticians. However, to regard only the *computational* aspects of mortality analysis would be to miss the point. Hoem & Funck-Jensen (1982) emphasised the following distinct aspects of a statistical model:

(a) *A probabilistic model.* By this is meant the collection of *assumptions* underlying the model, such as the statistical independence of events befalling different lives.

(b) *The numerical methods used to compute some model quantities from others (for instance the transition probabilities from the transition intensities).* Data analysis yields estimates of some model quantities. For example, the Continuous Mortality Investigation Bureau (CMIB), in their most recent investigations, used a model in which μ_x was estimated; numerical methods were then used to compute q_x (CMIR **9**, 1988).

(c) *Problems and methods of statistical inference applied when the model is used to analyse real data.* Statistical methods are appropriate when the model includes stochastic elements or errors of measurement. The relationships between model quantities, described in (b), are often useful in devising statistical methods.

Consider a practical problem — for example, to calculate an annuity rate. To solve it a computational tool is needed, such as a life table, but the life table must be chosen by reference to suitable data. The *statistical model* provides the link between the data ((c) above) and the computational tool ((b) above); both are aspects of the modelling process.

The life table l_x is often cited as a deterministic model, in Gerber's (1990) terms, the "somewhat embarrassing deterministic model". This viewpoint

leads to confusion when data are considered; it obscures the need for a statistical model to link data and applications, and it deflects attention from the nature of the process being modelled onto computational issues. Exposed to risk, for example, is an arithmetical problem arising in estimation. In a traditional actuarial approach, exposed to risk, practically, *is* the estimation procedure. This is putting the cart before the horse.

There are good reasons why the insensitivity of estimated rates q_x to the particular model assumptions should not lead actuaries to dismiss the framework of statistical models:

(a) Actuarial work is moving in new directions (for example, PHI and health care insurance), for which new models are needed in order to fashion the appropriate computational tools.
(b) Different models suit different data. An actuary who is given no choice of data must be able to select an appropriate model; a more fortunate actuary will sensibly specify the data in the context of a model.
(c) Statistical models of failure times are common currency outside the United Kingdom actuarial profession. They appear in medical statistics ('survival analysis'), engineering ('reliability theory'), demography ('event history analysis') and econometrics (see, for example, Lancaster, 1990). Actuaries should be aware of these links with other fields.

In Section 2 a stochastic model of mortality is described, and some questions of statistical inference are raised. In Section 3 we introduce the very general and flexible class of models known as multiple state models, which seem to be naturally adapted to actuarial use. Actuaries, in the past, have been more familiar with models based on the Binomial distribution, motivated as it is by the life table. Section 4 discusses the Binomial and Poisson models as alternatives to the multiple state model.

The surveys by Andersen & Borgan (1985) and Clayton (1988) go into more detail on some of these topics, and are useful sources of further work and examples. Jewell (1980) gives a wide-ranging survey of the uses of mathematical models in actuarial science.

2. Modelling the Time to Death

2.1 *A Simple Model*

This section describes a simple model of the mortality of a single life. We begin with this model because:

(a) it is identical to the models of failure times which are in common use elsewhere; and
(b) it leads to the treatment of life contingencies found in modern textbooks such as Bowers *et al.* (1986) or Gerber (1990).

The model assumption is that the time from birth to death can be

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

represented as a continuous random variable $\mathbf{T}$, having a value in the interval $(0, \omega)$, where ω is some limiting age. Let:

$$F(t) = \mathrm{P}[\mathbf{T} \leq t] \tag{1}$$

be the distribution function of $\mathbf{T}$, and define:

$$S(t) = 1 - F(t) = \mathrm{P}[\mathbf{T} > t] \tag{2}$$

to be the survivor function of $\mathbf{T}$. Then define the force of mortality to be:

$$\mu_t = \lim_{dt \to 0^+} \frac{1}{dt} \mathrm{P}[t < \mathbf{T} \leq t + dt \,|\, \mathbf{T} > t] \tag{3}$$

assuming the limit exists. The following synonyms for 'force of mortality' are often found: 'hazard rate'; 'force of transition'; 'transition intensity'. To deal with ages $x > 0$, we define a family of random variables $\{\mathbf{T}_x\}_{x=0}^{x=\omega}$. $\mathbf{T}_x$ is defined as the future lifetime after age x, conditional on having survived to age x; obviously $\mathbf{T}_0 = \mathbf{T}$. The distribution and survivor functions of $\mathbf{T}_x$ are denoted $F_x(t)$ and $S_x(t)$ respectively, and satisfy:

$$F_x(t) = \mathrm{P}[\mathbf{T} \leq x + t \,|\, \mathbf{T} > x] \quad \text{and} \quad S_x(t) = \mathrm{P}[\mathbf{T} > x + t \,|\, \mathbf{T} > x] \tag{4}$$

and all the usual relationships of the life table can be shown to hold, see for example Bowers *et al.* (1986). It is easily shown that the force of mortality based on $\mathbf{T}_x$, namely:

$$\lim_{dt \to 0^+} \frac{1}{dt} \mathrm{P}[t < \mathbf{T}_x \leq t + dt \,|\, \mathbf{T}_x > t] \tag{5}$$

is equal to μ_{x+t}. In actuarial notation, $F_x(t)$ is ${}_tq_x$, and $S_x(t)$ is ${}_tp_x$, but the statistical notation helps us to keep in mind the nature of the model. The force of mortality can be interpreted through the approximate relationship, valid for small dt:

$${}_{dt}q_x \approx \mu_x \, dt. \tag{6}$$

It will be helpful later on if we make equation (6) precise. A function $g(t)$ is said to be '$o(t)$' ('little-oh-of-t') if:

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

$$\lim_{t \to 0} \frac{g(t)}{t} = 0$$

in other words, if $g(t)$ tends to zero sufficiently faster than t itself. It is easy to see that the sum of a finite number of $o(t)$ functions is again $o(t)$, as is the product of any $o(t)$ function and a bounded function. Then we can show from the definition of μ_{x+t} that:

$$_{dt}q_x = \mu_x dt + g(dt) \tag{7}$$

where $g(t)$ is some function which is $o(t)$. We usually just write:

$$_{dt}q_x = \mu_x dt + o(dt) \tag{8}$$

since the precise form of $g(t)$ is of no concern.

The integrated hazard (to give it its usual name) often arises in survival analysis because it turns out to be a natural function to estimate. It is defined by:

$$\Lambda_t = \int_0^t \mu_s ds. \tag{9}$$

The probability density function of $\mathbf{T}_x$, denoted $f_x(t)$, is defined in terms of the force of mortality:

$$\begin{aligned} f_x\partial(t) &= \frac{\partial}{\partial t} F_x\partial(t) \\ &= \lim_{dt \to 0^+} \frac{F_x(t+dt) - F_x(t)}{dt} \\ &= \frac{S(x+t)}{S(x)} \times \lim_{dt \to 0^+} \frac{F(x+t+dt) - F(x+t)}{S(x+t)dt} \\ &= {}_tp_x \mu_{x+t}. \end{aligned} \tag{10}$$

Since $F_x(t) = 1 - {}_tp_x$, this can be rewritten in the familiar form:

$$\frac{\partial}{\partial t} {}_tp_x = - {}_tp_x \mu_{x+t} \tag{11}$$

which, when integrated with the boundary condition ${}_0p_x = 1$, gives the important formula:

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

$$ {}_tp_x = \exp\left(-\int_0^t \mu_{x+s} ds \right). \tag{12} $$

At this point it is worth recalling Hoem & Funck-Jensens' analysis (Section 1). Equation (11) allows us to calculate μ_x, given the probabilities ${}_tp_x$; equation (12) allows us to do the reverse. These are examples of "numerical methods used to compute some model quantities from others". In practical terms, we can estimate either probabilities ${}_tp_x$ or forces μ_x from the data, and then compute any other quantities which we need. The question of *what* to estimate from the data can, therefore, be decided on statistical grounds.

2.2 *Questions of Inference*

We now turn to statistical inference. Given some mild conditions on the distribution of **T**, we can obtain all information by estimating $F(t)$, $S(t)$, $f(t)$ or μ_t for all $t \geq 0$.

The simplest experiment would be to observe a large number of new-born lives; the proportion alive at age $t > 0$ would furnish an estimate of $S(t)$. The estimate would be a step function, and the larger the sample the closer to a smooth function we would expect it to be. For use in applications it could be smoothed further. We need not assume that **T** is a member of any parametric family; this is a *non-parametric* approach to estimation. Clearly, there are some practical problems:

(a) Even if a satisfactory group of lives could be found, the experiment would take about 100 years to complete.

(b) The observational plan requires us to observe the deaths of *all* the lives in the sample. In practice many would be lost to the investigation, for one reason or another, and to exclude these from the analysis might bias the result. The statistical term for this problem is *censoring*. All we know in respect of some lives is that they died after a certain age.

In medical statistics, where the lifetimes are often shorter, non-parametric estimation is very important. In Part II, Section 6 we show how the experiment above can be amended to allow for censoring. Otherwise, we must use a different observational plan, and base inference on data gathered over a shorter time; for example, 3 years (the ELT tables) or 4 years (the CMIB tables). A consequence is that we no longer observe the same cohort throughout their joint lifetimes, so we might not be sampling from the same distribution. It might be sensible to widen the model assumption, so that the mortality of lives born in year y is modelled by a random variable $\mathbf{T}^y$, for example. In practice we usually divide the investigation up into single years of age.

Observing lives between (say) integer ages x and $x+1$, and limiting the period of investigation, are also forms of censoring. Censoring might still

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

occur at unpredictable times — by lapsing a life policy for example — but survivors will certainly be lost to observation at a known time, either on attaining age $x+1$ or when the investigation ends.

2.3 *Censoring Mechanisms*

Censoring is the key feature of survival data (indeed survival analysis might be defined as the analysis of censored data) and the mechanisms which give rise to censoring play an important part in statistical inference. Some of the commonest censoring assumptions are these (they are not all mutually exclusive):

(a) *Right-censoring.* Data are right-censored if the censoring mechanism cuts short observations in progress. An example is the ending of an investigation on a fixed date.

(b) *Left-censoring.* Data are left-censored if the censoring mechanism prevents us from knowing when entry into the state which we wish to observe took place. An example arises in medical studies in which patients are subject to regular examinations. Discovery of a condition tells us only that the onset fell in the period since the previous examination; the time elapsed since onset has been left-censored.

(c) *Interval-censoring.* Data are interval-censored if the observational plan only allows us to say that an event of interest fell within some interval of time. An example arises in actuarial investigations, where we might know only the calendar year of death.

(d) *Random censoring.* If censoring is random, then the time $\mathbf{C}_i$ (say) at which observation of the i^{th} lifetime is censored is a random variable. The observation will be censored if $\mathbf{C}_i < \mathbf{T}_i$, where $\mathbf{T}_i$ is the (random) lifetime of the i^{th} life. The case in which the censoring mechanism is a second decrement of interest gives rise to multiple decrement models; see Section 3 and Part II, Section 5.

(e) *Non-informative censoring.* Censoring is non-informative if it gives no information about the lifetimes $\{\mathbf{T}_i\}$. In the case of random censoring, the independence of each pair $\mathbf{T}_i$, $\mathbf{C}_i$ is sufficient to ensure that the censoring is non-informative. Informative censoring is more difficult to analyse, essentially because the resulting likelihoods cannot usually be factorised.

(f) *Type I censoring.* If the censoring times $\{\mathbf{C}_i\}$ are known in advance (a degenerate case of random censoring), then the mechanism is called 'Type I censoring'.

(g) *Type II censoring.* If observation is continued until a predetermined number of deaths has occurred, then 'Type II censoring' is said to be present. This can simplify the analysis, because then the number of events of interest is non-random.

It is obvious that the observational plan is likely to introduce censoring of

some kind, and consideration should be given to the effect on the analysis in specifying the observational plan. Censoring might also depend on the results of the observations to date; for example if strong enough evidence accumulates during the course of a medical experiment, the investigation might be ended prematurely, so that the better treatment can be extended to all the subjects under study, or the inferior treatment withdrawn. Andersen *et al.* (1993) gave a comprehensive account of censoring schemes.

3. Multiple State Models

Multiple state models have been described by Sverdrup (1965) and Waters (1984), among others. In this section we describe multiple state models briefly, using the simplest example of a two state model to illustrate general points. Our main purpose is to emphasise:

(a) some advantages of multiple state models over some others often used in actuarial work, which are described in Section 4; and
(b) connections with other parts of survival analysis.

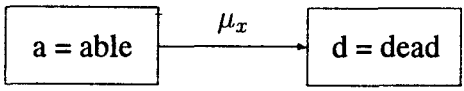

Figure 1. A two state model of mortality

3.1 *The Two State Model (I) — Assumptions*

The two state model is illustrated in Figure 1. There is an alive state and a dead state, with transitions in one direction only. The probability that a life alive at a given age should be dead at any subsequent age is governed by the age-dependent transition intensity $\mu_{x+t}(t \geq 0)$, in a way made precise by Assumption 2 below.

Assumption 1. The probabilities that a life at any given age will be found in either state at any subsequent age depend only on the ages involved and on the state currently occupied. This is the *Markov* assumption.

Assumption 2. ${}_{dt}q_{x+t} = \mu_{x+t}dt + o(dt) \quad (t \geq 0)$.

Assumption 2 simply takes equation (8) as a starting point, because it is mathematically convenient to do so. Assumption 1 is more subtle. The past history of an individual — for example, current state of health, spells of sickness, occupation — is excluded from the model. If we knew these factors, we could:

(a) treat each combination of factors as a separate model; in other words, *stratify* the problem; or

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

(b) specify a model which took them into account; in other words, treat the problem as one of *regression.*

This would take us into the area treated in Part II, Section 7. Here, only the currently occupied state is relevant.

For the purposes of inference, we restrict our attention to ages between x and $x+1$, and introduce a further assumption:

Assumption 3. μ_{x+t} is a constant μ for $0 \leq t < 1$.

It is important to emphasise that this two state model is *not* the same as the model of Section 2; we start with different assumptions. One model is formulated in terms of a random variable **T** representing the lifetime, the other in terms of a transition intensity between states. It is easy to impose some mild conditions under which the models are equivalent, but when we consider more than one decrement these two formulations lead in different directions (see Part II, Section 5).

3.2 *The Two State Model (II) — Probabilities*

Since we have specified the model in terms of a transition intensity, we consider briefly how to compute probabilities. Consider the survival probability ${}_{t+dt}p_x$, and condition on the state occupied at age $x+t$. By the Markov assumption, nothing else affects the probabilities of death or survival after age $x+t$:

$$\begin{aligned} {}_{t+dt}p_x &= {}_tp_x \times \mathrm{P}[\text{Alive at } x+t+dt \,|\, \text{Alive at } x+t] \\ &\quad + {}_tq_x \times \mathrm{P}[\text{Alive at } x+t+dt \,|\, \text{Dead at } x+t] \\ &= {}_tp_x \times {}_{dt}p_{x+t} + {}_tq_x \times 0 \\ &= {}_tp_x \times (1 - \mu_{x+t}dt + o(dt)). \end{aligned}$$

Therefore:

$$\begin{aligned} \frac{\partial}{\partial t}{}_tp_x &= \lim_{dt \to 0^+} \frac{{}_{t+dt}p_x - {}_tp_x}{dt} \\ &= -{}_tp_x\mu_{x+t} + \lim_{dt \to 0^+} \frac{o(dt)}{dt} \\ &= -{}_tp_x\mu_{x+t} \end{aligned} \tag{13}$$

which is the same as equation (11), so ${}_tp_x$ can be computed from equation (12). The important point is that it has been derived here strictly from the assumptions of the two state model, and that the method is easily extended

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

to models with more states. In the Markov framework, equation (13) is an example of the *Kolmogorov forward equations.* It is interesting to note that the Markov assumption is implicit in the traditional development of life table probabilities; it is introduced in the definitions (4).

3.3 *The Two State Model (III) — Statistics*

Next we define our observations. We will describe two slightly different approaches. Both suppose that we observe a total of N lives during some finite period of observation, between the ages of x and $x+1$. We could suppose that lives were observed, or not, as a result of some random mechanism (not depending on any parameter of interest), but, in this paper, we suppose that data are analysed retrospectively, so we regard N as a non-random quantity. We do not assume that we observe the N lives simultaneously, nor do we assume that we observe each life for the complete year of age. We do assume that all N lives are identical and statistically independent.

For simplicity we consider Type I censoring, and we use the notation of Broffit (1984). For $i=1,\ldots,N$ define $x+a_i$ to be the age at which observation of the i^{th} life starts, and let $x+b_i$ be the age at which observation of the i^{th} life must cease if the life survives to that age. $x+b_i$ will be either $x+1$, or the age of the i^{th} life when the investigation ends, whichever is smaller. The important point is that a_i and b_i are known in advance.

Under Type I censoring, statistics in respect of different lives are not identically distributed, which is sometimes inconvenient. An alternative, essentially due to Sverdrup (1965) is to let the entry age $x+\mathbf{E}_i$ and the age at censoring $x+\mathbf{C}_i$ be (dependent) random variables with densities $f_E(e_i)$ and $f_{C|E}(c_i \mid e_i)$ respectively, the same for all i. Mixed distributions, giving probability masses at the points $\mathbf{E}_i=0$ and $\mathbf{C}_i=1$ are easily accommodated.

The constants a_i and b_i, or the random variables $\mathbf{E}_i$ and $\mathbf{C}_i$, define a mechanism for starting and ending observation of the i^{th} life, distinct from the decrement(s) which we are observing. Multiple decrements can be regarded as a set of mutually censoring processes, but the observational plan might still allow for censoring by some mechanism other than the decrements; that is the purpose of the definitions above. To avoid repetition, the following is given in terms of Type I censoring; it will be clear how to modify it for the random censoring.

Define a random variable $\mathbf{D}_i$ as follows:

$$\mathbf{D}_i = \begin{cases} 1 \text{ if the } i^{th} \text{ life is observed to die} \\ 0 \text{ if the } i^{th} \text{ life is not observed to die.} \end{cases}$$

$\mathbf{D}_i$ is an example of an *indicator* random variable; it indicates the occurrence of death. Define a random variable $\mathbf{T}_i$ as follows:

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

$x+\mathbf{T}_i$ = the age at which observation of the i^{th} life ends.

Notice that $\mathbf{D}_i$ and $\mathbf{T}_i$ are not independent, since:

$$\mathbf{D}_i=0 \Leftrightarrow \mathbf{T}_i=b_i, \qquad \mathbf{D}_i=1 \Leftrightarrow a_i<\mathbf{T}_i<b_i. \tag{14}$$

It will often be useful to work with the time spent under observation, so define $\mathbf{V}_i=\mathbf{T}_i-a_i$. $\mathbf{V}_i$ is called the *waiting time* and it is seen to be related to the 'central exposed to risk', but with an important difference: $\mathbf{V}_i$ is a random variable. It has a mixed distribution, with a probability mass at the point b_i-a_i.

The pair $(\mathbf{D}_i, \mathbf{V}_i)$ comprise a *statistic*, meaning that the outcome of our observation is a sample (d_i, v_i) drawn from the distribution of $(\mathbf{D}_i, \mathbf{V}_i)$. Let $f_i(d_i, v_i)$ be the joint distribution of $(\mathbf{D}_i, \mathbf{V}_i)$. It is easily written down by considering the two cases $\mathbf{D}_i=0$ and $\mathbf{D}_i=1$:

$$f_i(d_i, v_i)=\begin{cases} {}_{b_i-a_i}p_{x+a_i} & (d_i=0) \\ {}_{v_i}p_{x+a_i}\mu_{x+a_i+v_i} & (d_i=1) \end{cases}$$

$$=\begin{cases} \exp\left(-\int_0^{b_i-a_i}\mu_{x+a_i+t}dt\right) & (d_i=0) \\ \exp\left(-\int_0^{v_i}\mu_{x+a_i+t}dt\right)\mu_{x+a_i+v_i} & (d_i=1) \end{cases}$$

$$=\exp\left(-\int_0^{v_i}\mu_{x+a_i+t}dt\right)\mu_{x+a_i+v_i}^{d_i} \tag{15}$$

where the last step follows from equation (14). Now assume that μ_{x+t} is a constant μ for $0\leq t<1$ (this is the first time we have needed this assumption) and $f_i(d_i, v_i)$ takes on the simple form:

$$f_i(d_i, v_i)=e^{-\mu v_i}\mu^{d_i}. \tag{16}$$

The joint probability function of all the $(\mathbf{D}_i, \mathbf{V}_i)$, by independence, is:

$$\prod_{i=1}^{i=N} e^{-\mu v_i}\mu^{d_i}=e^{-\mu(v_1+\ldots+v_N)}\mu^{d_1+\ldots+d_N}$$

$$=e^{-\mu v}\mu^{d} \tag{17}$$

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

where $d=\sum_{i=1}^{i=N} d_i$ and $v=\sum_{i=1}^{i=N} v_i$. In other words, define random variables $\mathbf{D}$ and $\mathbf{V}$ to be the total number of deaths and the total waiting time, respectively, and the joint probability function of all the $(\mathbf{D}_i, \mathbf{V}_i)$ can be simply expressed in terms of $\mathbf{D}$ and $\mathbf{V}$.

3.4 *The Two State Model (IV) — the Maximum Likelihood Estimator*

The probability function immediately furnishes the likelihood for μ:

$$L(\mu; d, v) = e^{-\mu v}\mu^d \tag{18}$$

which yields the maximum likelihood estimate (MLE) for μ:

$$\hat{\mu} = d/v. \tag{19}$$

This is intuitively satisfying and even obvious. Seal (1977) pointed out that Sprague suggested $\hat{\mu}$ as an estimate of μ as early as 1879. Notice that the *estimate* $\hat{\mu}$, being a function of the sample values d and v, can itself be regarded as a sample drawn from the distribution of the corresponding *estimator*:

$$\tilde{\mu} = \mathbf{D}/\mathbf{V}. \tag{20}$$

The distinction between the estimator and the estimate is often ignored, as doing so causes little confusion. It is useful to maintain the distinction here, though, because it will help to make clear the nature of the so-called Poisson model of mortality in Section 4. We will use boldface capitals for random variables, and lower case letters for samples.

It is important, in applications, to be able to estimate the moments of the estimator $\tilde{\mu}$, for example, in order to compare the experience with that of a standard table. At least, we need to estimate $\mathrm{E}[\tilde{\mu}]$ and $\mathrm{Var}[\tilde{\mu}]$. It is a standard result of maximum likelihood theory that the asymptotic distribution of $\tilde{\mu}$ is Normal, with mean μ and variance $\mu/\mathrm{E}[\mathbf{V}]$, but it is useful to describe the approach of Sverdrup (1965), partly because it is based on the following elegant results which will be of further interest in Part III, Section 8:

$$\mathrm{E}[\mathbf{D}_i - \mu\mathbf{V}_i] = 0 \tag{21}$$

$$\mathrm{Var}[\mathbf{D}_i - \mu\mathbf{V}_i] = \mathrm{E}[\mathbf{D}_i]. \tag{22}$$

Note that equation (21) can also be written as $\mathrm{E}[\mathbf{D}_i] = \mu\mathrm{E}[\mathbf{V}_i]$. In the case that the $\{a_i\}$ and $\{b_i\}$ are known constants, this follows from integrating/

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

summing the probability function of $(\mathbf{D}_i, \mathbf{V}_i)$ over all possible events to obtain:

$$\int_0^{b_i - a_i} e^{-\mu v_i} \mu dv_i + e^{-\mu(b_i - a_i)} = 1 \tag{23}$$

and then differentiating with respect to μ, once to obtain equation (21) and twice to obtain equation (22). Under the random censoring defined above, we replace equation (23) with:

$$\int_0^1 \left(\int_{e_i}^1 e^{-\mu(c_i - e_i)} \mu f_{C|E}(c_i \mid e_i) dc_i \right) f_E(e_i) de_i = 1 \tag{24}$$

and proceed as before. To find the asymptotic distribution of $\tilde{\mu}$, consider:

$$\frac{1}{N}(\mathbf{D} - \mu\mathbf{V}) = \frac{1}{N} \sum_0^N (\mathbf{D}_i - \mu\mathbf{V}_i).$$

It is reasonable to assume that, as $N \to \infty$, the empirical frequencies of the $\{a_i\}$ and the $\{b_i\}$ converge to some distribution, so that, asymptotically, we can disregard the difference between the fixed and random censoring. Then note that:

$$\lim_{N \to \infty} (\tilde{\mu} - \mu) = \lim_{N \to \infty} \frac{N}{\mathbf{V}} \left(\frac{\mathbf{D}}{N} - \frac{\mu\mathbf{V}}{N} \right). \tag{25}$$

By the law of large numbers, $\mathbf{V}/N \to \mathrm{E}[\mathbf{V}_i]$ in probability, and by the Central Limit Theorem:

$$\frac{1}{N}(\mathbf{D} - \mu\mathbf{V}) \sim \text{Normal}\left(0, \frac{\mathrm{E}[\mathbf{D}]}{N^2}\right) \tag{26}$$

so, asymptotically:

$$\tilde{\mu} \sim \text{Normal}\left(\mu, \frac{\mu}{\mathrm{E}[\mathbf{V}]}\right). \tag{27}$$

The derivation given above is somewhat heuristic. Conditions under which

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

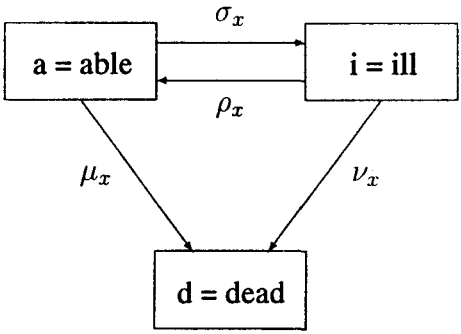

Figure 2. An illness-death model

the result is valid were discussed by Sverdrup (1965) and Borgan (1984). Simulation studies suggest that the asymptotic distribution gives good results if $\mathrm{E}[\mathbf{D}] \geq 10$ (Schou & Væth, 1980).

3.5 *The General Markov Model*

The two state model above can be extended to any number of states, with arbitrary transitions between them, including increments and repeated transitions. An example having these features which has been applied in actuarial work is the three state illness-death model, shown in Figure 2.

Following Waters (1984), we use the following notation. Let g and h denote any two states. If states g and h are distinct, let μ^{gh}_{x+t} be the transition intensity from state g to state h at age $x+t$. Define:

$$ {}_tp^{gh}_x = \mathrm{P}[\text{In state } h \text{ at age } x+t \,|\, \text{In state } g \text{ at age } x] \tag{28} $$

where now g and h need not be distinct. The event whose probability is defined by equation (28) does not specify what must happen between age x and age $x+t$; in particular, if $g=h$, it does not require that the life remains in state g between these ages. So for any state g, define:

$$ {}_tp^{\overline{gg}}_x = \mathrm{P}[\text{In state } g \text{ from age } x \text{ to } x+t \,|\, \text{In state } g \text{ at age } x]. \tag{29} $$

If return to state g is impossible, then ${}_tp^{\overline{gg}}_x = {}_tp^{gg}_x$, but this is not true (for example) in the case of states a and i in the illness-death model above. We enlarge Assumption 2 of the two state model as follows:

*Assumption 2**. For any two distinct states g and h, ${}_{dt}p^{gh}_{x+t} = \mu^{gh}_{x+t}dt + o(dt)$ $(t \geq 0)$, *and* the probability that a life makes any two or more transitions in time dt is $o(dt)$.

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

Then, by the same method as before, we can derive the Kolmogorov forward equations:

$$\frac{\partial}{\partial t}\,{}_tp_x^{gh} = \sum_{j \neq h} {}_tp_x^{gj}\mu_{x+t}^{jh} - {}_tp_x^{gh}\mu_{x+t}^{hj} \tag{30}$$

$$\frac{\partial}{\partial t}\,{}_tp_x^{\overline{gg}} = -\,{}_tp_x^{\overline{gg}} \sum_{j \neq g} \mu_{x+t}^{gj}. \tag{31}$$

The calculation of probabilities from the estimated intensities requires the solution of these equations, which is a straightforward numerical exercise; see for example Jones (1994) and Kulkarni (1995). If the intensities are piecewise constant, then probabilities can be computed analytically (Ramsay, 1989), and this is also possible in certain other cases (CMIR **13**, 1993). Markov models also lead directly and naturally to life insurance mathematics (Hoem, 1969, 1988), which lies outside the scope of this review. For inference, we do need the solution of equation (31), but since this is of the same form as equation (11), and satisfies a similar boundary condition, its solution is just:

$${}_tp_{x+t}^{\overline{gg}} = \exp\left(-\int_0^t \sum_{j \neq g} \mu_{x+s}^{gj}\,ds\right). \tag{32}$$

We now return to inference, for the age interval x to $x+1$, taking the illness-death model as an example. The observations in respect of a single life are now:

(a) the times between successive transitions; and
(b) the numbers of transitions of each type.

If the transition intensities are constant, equation (32) shows that each spell of length t in the able or ill states contributes a factor of the form $e^{-(\mu+\sigma)t}$ or $e^{-(\nu+\rho)t}$, respectively, to the likelihood, so it suffices to record the total waiting time spent in each state. Then defining:

$\mathbf{V}_i$ = Waiting time of the i^{th} life in the able state
$\mathbf{W}_i$ = Waiting time of the i^{th} life in the ill state
$\mathbf{S}_i$ = Number of transitions able → ill by the i^{th} life
$\mathbf{R}_i$ = Number of transitions ill → able by the i^{th} life
$\mathbf{D}_i$ = Number of transitions able → dead by the i^{th} life
$\mathbf{U}_i$ = Number of transitions ill → dead by the i^{th} life

and defining totals $\mathbf{V} = \sum_1^N \mathbf{V}_i$ (and so on), and using lower case symbols for

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

the observed samples as usual, it is easily shown that the likelihood for the four parameters μ, ν, σ, ρ, given the data, is:

$$L(\mu, \nu, \sigma, \rho) = e^{-(\mu+\sigma)v} e^{-(\nu+\rho)w} \mu^d \nu^u \sigma^s \rho^r. \tag{33}$$

This factorises into functions of each parameter of the form $e^{-\mu v}\mu^d$, so the maximum likelihood estimators are:

$$\tilde{\mu} = \frac{\mathbf{D}}{\mathbf{V}}, \quad \tilde{\nu} = \frac{\mathbf{U}}{\mathbf{W}}, \quad \tilde{\sigma} = \frac{\mathbf{S}}{\mathbf{V}}, \quad \tilde{\rho} = \frac{\mathbf{R}}{\mathbf{W}}. \tag{34}$$

The asymptotic properties of these estimators follow from results similar to equations (21) and (22), and the fact that the random variables $(\mathbf{D}_i - \mu\mathbf{V}_i)$, $(\mathbf{U}_i - \nu\mathbf{W}_i)$, $(\mathbf{S}_i - \sigma\mathbf{V}_i)$, $(\mathbf{R}_i - \rho\mathbf{W}_i)$ are uncorrelated, that is:

$$\mathrm{E}[(\mathbf{D}_i - \mu\mathbf{V}_i)(\mathbf{U}_i - \nu\mathbf{W}_i)] = 0 \quad \text{etc.} \tag{35}$$

The estimators are not independent: $\mathbf{D}_i$ and $\mathbf{U}_i$ are both 0 or 1, but $\mathbf{D}_i\mathbf{U}_i \neq 1$, while (assuming that the i^{th} life starts in the able state) $\mathbf{S}_i = \mathbf{R}_i$ or $\mathbf{R}_i + 1$. They are, however, asymptotically independent: the same argument as in the two state model shows that the vector $(\tilde{\mu}, \tilde{\nu}, \tilde{\sigma}, \tilde{\rho})$ has an asymptotic multivariate Normal distribution; that each component has a marginal asymptotic distribution of the same form as in equation (27); and that, asymptotically, the components are uncorrelated because of equation (35) and so independent (being Normal).

Equations (21), (22) and (35) are particularly noteworthy; only slightly generalised, they are the key to the application, described in Part III, Section 8, of powerful martingale methods to many areas of survival analysis.

The calculation of the estimates $\hat{\mu}$, etc. requires the central exposed-to-risk to be computed. This can be done exactly in some circumstances, but if the exposure data are in census form (as in the CMIB investigations), the usual census formulae provide estimates. Multiple state models are, therefore, especially well suited to the data available in many actuarial investigations.

3.6 *The CMIB Illness-Death Model*

A large scale application of a multiple state model is the illness-death model developed by the CMIB (Waters, 1991a). This differs from the model shown in Figure 2 in that the transition intensities out of the ill state depend on the duration of sickness z as well as age x. The Markov property no longer holds if the state occupied by the individual is just able, ill or dead as before, but, if the idea of 'state' is extended to include the time spent in the current state, a Markov model can be assumed.

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

In principle, estimation of the intensities proceeds just as above, over rectangles representing pairs of age and duration intervals, chosen so that $\nu_{x,z}$ and $\rho_{x,z}$ can reasonably be assumed to be constant. However, the data related to policies with *deferred periods*; that is, no benefit is paid until the duration of sickness exceeds the deferred period. The only available data in respect of sickness and recovery were the starting and ending times of *claims*; transitions to and from the ill state were not observed if no claim was made. Moreover, for deferred periods longer than 1 week, $\rho_{x,z}$ displayed a 'run-in' period of roughly 4 weeks from the end of the deferred period, during which it was lower than might have been expected on the basis of the intensities for deferred period 1 week. The CMIB suggested that this might be caused by policyholders sometimes not claiming after the deferred period when recovery was imminent. As a result, certain adjustments were necessary in the calculation of the sickness inception rates (Waters, 1991b):

(a) The exposed-to-risk in the able state was reduced to allow for a proportion of lives who were sick, but not claiming. The latter proportion was estimated numerically, by an integration involving both σ_x and $\rho_{x,z}$.
(b) For deferred periods longer than 1 week, the rates of recovery $\rho_{x,z}$ in the 4 weeks after the deferred period were adjusted to remove the effect of the run-in period.
(c) An iterative approach to estimating σ_x was used, because the graduated values of σ_x were needed in adjustment (a) above, and because the estimates $\hat{\sigma}_x$ also figured in the estimated variance of $\tilde{\sigma}_x$.

Unobserved transitions are not unusual in applications of multiple state models, usually because the analyst has no control over the data. Lindsey & Ryan (1993) discussed a model for cancer in rodents in which the existence of a tumour can only be established after death, which is analagous to the time between onset of sickness and claiming under a PHI policy being unknown. They used a standard iterative method (the EM algorithm) to overcome the problem.

The CMIB also dealt at length with computational procedures for deriving probabilities and financial functions from the intensities; in particular, it was shown that well-known tools such as the Manchester Unity functions, claim inception rates and claim annuities could all be computed. Here again, we should be reminded of Hoem & Funck-Jensens' description of the modelling process. The procedures relied on numerical (recursive) integration of the integro-differential equations corresponding to the Kolmogorov forward equations of a Markov model (Waters & Wilkie, 1987). In principle this was identical to the approach used by Forfar, McCutcheon & Wilkie (1988) for the production of life tables, but the intensities were much less well-behaved over very short time periods than is the force of mortality typical of life tables.

3.7 *Further Topics*

Here we make some further remarks about multiple state models:

(a) *Semi-Markov models.* A model in which transition probabilities depend on the past history in the current state, but not on anything occurring prior to entering the current state, is a *semi-Markov* model. The CMIB illness-death model is an example. Wilkie (1988b) suggested a marriage and mortality model along similar lines, in which the ages of both spouses and the duration of marriage were included.

In principle, estimation is no more difficult than in a Markov model, provided that the quantities upon which intensities depend can be partitioned finely enough for constant intensities to be assumed, but the calculation of probabilities and other functions of the estimated intensities is more laborious. As long as numerical answers are sufficient, however, any equations which arise should be soluble by standard methods.

(b) *Non-constant intensities.* In Part II, Section 6 we will consider non-parametric estimation of the lifetime distribution or integrated hazard. These methods can be applied to the transitions of a multiple state model. Note, however, that methods used to estimate lifetime distributions in a decrement model do not have the same interpretation in a model with repeated transitions, essentially because the probability of being in the initial state after some time is not the same as the probability of having remained in the initial state. Non-parametric methods give an estimate of the former probabilities. Since both can be computed from the intensities, the latter might be a better target for estimation.

The assumption of constant intensities is mainly useful at the estimation stage. It does not materially aid the calculation of probabilities and related functions, especially where numerical methods are used. Where some of the transitions are modelled by known or assumed intensities, these can take on any form. For example, Wilkie (1988a) proposed a model for AIDS in which several intensities were given by parametric functions.

(c) *Regression models.* In Part II, Section 7 we look at regression problems, in which the hazard rate depends on covariates whose values describe each life. The same methods can be applied to multiple state models. For applications, however, the transition intensities are needed, so, in the terminology of Part II, Section 7, all the baseline intensities have to be estimated. This can be a useful way to study a heterogeneous population, for which we cannot simply aggregate the observations on different lives. A simple demographic example was given by Wood *et al.* (1994).

Some instructive applications of multiple state models to AIDS can be found in Daykin *et al.* (1988), Wilkie (1988a) and Ramsay (1989).

4. Binomial and Poisson Models

4.1 *Binomial-Type Models*

Much of the motivation for the analysis of mortality data is provided by the following thought-experiment: observe N identical, independent lives aged x exactly for one year, and record the number d who die. Then d is a sample value of a random variable **D**. If we suppose that each life dies with probability q_x and survives with probability $1-q_x$, then **D** has a Binomial distribution with parameters N and q_x. The intuitive estimate of q_x is $\hat{q}_x = d/N$, and this is also the maximum likelihood estimate. The corresponding estimator $\tilde{q}_x$ has mean q_x and variance $q_x(1-q_x)/N$. This is the Binomial model of mortality.

The direct connection with the life table quantities l_x and d_x is obvious, and the Binomial model is often cited in textbooks when the stochastic aspects of data analysis are introduced; see for example Benjamin & Pollard (1980). It was discussed by Forfar, McCutcheon & Wilkie (1988), who compared it with a Poisson model, of which more later. The main point to be made here is that the Binomial model leads to problems if the observations are more realistic:

(a) we might not observe all lives over the same interval of age; and
(b) there will usually be decrements other than death, and sometimes increments as well.

In terms of the data defined in Section 3, the $\{a_i\}$ and the $\{b_i\}$ are, in general, not all the same. Considering the i^{th} life, we have:

$$\mathrm{P}[\mathbf{D}_i = d_i] = {}_{b_i-a_i}q_{x+a_i}{}^{d_i}(1-{}_{b_i-a_i}q_{x+a_i})^{1-d_i} \quad (d_i = 0, 1). \tag{36}$$

In respect of this individual, equation (36) makes a contribution to the total likelihood, in which ${}_{b_i-a_i}q_{x+a_i}$ appears as a parameter and d_i as the observed statistic. Defining the vector quantities:

$$\vec{q} = ({}_{b_1-a_1}q_{x+a_1},\ {}_{b_2-a_2}q_{x+a_2}, \ldots,\ {}_{b_N-a_N}q_{x+a_N})$$
$$\vec{d} = (d_1, d_2, \ldots, d_N)$$

we can write the total likelihood as:

$$L(\vec{q}; \vec{d}) = \prod_{i=1}^{i=N} {}_{b_i-a_i}q_{x+a_i}{}^{d_i}(1-{}_{b_i-a_i}q_{x+a_i})^{1-d_i}. \tag{37}$$

We have to find the value of the vector $\vec{q}$ — in general N numbers — which maximises the likelihood. The dimension of the problem might be reduced if

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

some of the $\{a_i\}$ and the $\{b_i\}$ are equal, but the usual approach is to make an assumption about the distribution of $\mathbf{T}_x$ in the age range $[x, x+1]$ which allows us to express any ${}_{b_i-a_i}q_{x+a_i}$ in terms of q_x, making the likelihood a function of one parameter again. Possible assumptions are:

(a) *uniform distribution of deaths:* ${}_tq_x = tq_x$ $(0 \le t \le 1)$

(b) *the Balducci assumption:* ${}_{1-t}q_{x+t} = (1-t)q_x$ $(0 \le t \le 1)$

(c) *constant force of mortality:* ${}_tq_x = 1 - e^{-\mu t}$ $(0 \le t \le 1)$.

(Note that the Balducci assumption implies a decreasing force of mortality between integer ages.) The resulting maximum likelihood estimators were treated in detail by Broffit (1984). None of them is particularly attractive. The question of how Binomial-type models might be generalised to multiple decrements is considered in Part II, Section 5.

4.2 *The Actuarial Estimate*

The Balducci assumption has been used in an attempt to provide a theoretical justification of the initial exposed to risk in the traditional actuarial estimate of q_x; see, for example, Batten (1978). Under the Balducci assumption:

$$\begin{aligned} \mathrm{E}[\mathbf{D}] &= \sum_{i=1}^{i=N} {}_{b_i-a_i}q_{x+a_i} \\ &= \sum_{i=1}^{i=N} {}_{1-a_i}q_{x+a_i} - \sum_{i=1}^{i=N} {}_{b_i-a_i}p_{x+a_i}\ {}_{1-b_i}q_{x+b_i} \\ &= \sum_{i=1}^{i=N} (1-a_i)q_x - \sum_{i=1}^{i=N} (1-\mathrm{E}[\mathbf{D}_i])(1-b_i)q_x. \end{aligned} \tag{38}$$

For simplicity we are assuming that the $\{a_i\}$ and $\{b_i\}$ are known, and that death is the only decrement. Substituting the observed number of deaths d on the left side of equation (38) would usually give the moment estimate of q_x. However, the right side of equation (38) also involves expected deaths, and, as Dorrington & Slawski (1993) pointed out, in such a way that it is impossible to extract all the terms in $\mathrm{E}[\mathbf{D}]$ and the $\{\mathrm{E}[\mathbf{D}_i]\}$ on one side and all the terms in q_x on the other. Summing the last term of equation (38) over the observed rather than the expected survivors, we obtain:

$$\mathrm{E}[\mathbf{D}] \approx \sum_{i=1}^{i=N} (1-a_i)q_x - \sum_{i=1}^{i=N} (1-d_i)(1-b_i)q_x \tag{39}$$

leading to the estimate:

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

$$\hat{q}_x = \frac{d}{\sum_{i=1}^{i=N}(1-a_i) - \sum_{i:D_i=0}(1-b_i)} \tag{40}$$

in which the denominator is the traditional initial exposed to risk, counting the deaths as exposed to risk until the end of the year of age. Under the crude assumption that deaths occur, on average, at age $x+1/2$, and ignoring the awkward possibility that $a_i > 1/2$, we obtain the well-known formula:

$$\hat{q}_x = \frac{d}{E_x^c + d/2}. \tag{41}$$

This is known to statisticians as the *actuarial estimate*. Note that equation (40) is *not* the moment estimate of q_x; Broffit (1984) called it a modified method of moments estimator. Hoem (1984) pointed out that equation (39) rests on an incorrect treatment of exits, in which the probability that a life aged $x+a_i$ dies between ages $x+t_i$ and $x+1$ is not conditioned upon survival to age $x+t_i$. The correct moment estimate under the Balducci assumption (Hoem, 1984) is the solution of:

$$d = \sum_{i=1}^{i=N} \frac{b_i - a_i}{1-(1-b_i)q_x^*} q_x^*. \tag{42}$$

On the basis of equation (39), Dorrington & Slawski (1993) mounted a defence of the actuarial estimate against the criticisms of Hoem (1984). In one sense it needs no defence, since it is plainly true that the actuarial estimate is an estimate of q_x; but so, for example, is the proportion of inhabitants of London who work in Paris. The actuarial estimate appears, of course, more sensible, but both estimates are obtained by stepping outside the model framework. Given the insensitivity of results to methods, in the case of small rates of decrement and no increments, the actuarial estimate has worked well enough for simple problems. Elandt-Johnson & Johnson (1980) compared five methods of estimating q_x, and said (we have changed their notation to agree with ours):

> "In conclusion, we may say that for sufficiently large sample sizes and small q_x (<0.3), one can use any of the estimators $\hat{q}_x^{(1)}$ through $\hat{q}_x^{(5)}$. Of course, the simplest is the actuarial estimator ... and we would recommend it for use. It is a *good* and *robust* estimator of q_x."

The Binomial model, and the actuarial estimate, are not without strengths. The actuarial estimate avoids numerical solution of equations such as equation (42), and it might be used if there is a compelling reason to estimate q_x instead of something else; and, as we shall see in Part II, Section 6, the

Binomial model can be generalised simply to give a non-parametric estimate — the Kaplan–Meier estimate — which is widely used in survival analysis. In the case of grouped data, the actuarial estimate can be similarly extended, and is similarly useful.

However, it cannot be said that the actuarial estimate is any simpler than the estimates based on multiple state models in Section 3; indeed, if the exposure data are of the census type the need to compute an initial exposed to risk is a pointless complication. Nor are estimates always so insensitive as those of q_x seem to be. Hoem & Funck-Jensen (1982) gave examples in which a Binomial-type model leads to absurd conclusions, such as negative transition intensities. Crucially, the Binomial model is not so easily generalised to settings with more than one decrement. Even the simplest case of two decrements gives rise to difficult problems (see Part II, Section 5); the introduction of repeated transitions such as sickness and recovery is more difficult still. Extension of models in these directions is much simpler within the multiple state framework.

4.3 *Poisson Models*

The Poisson distribution is used to model the number of 'rare' events occurring during some period of time, for example the number of particles emitted by a radioactive source in a minute. Such analogies suggest the Poisson distribution as a model for the number of deaths among a group of lives, given the time spent exposed to risk.

In this section we will let E_x^c denote the total central exposed to risk; in terms of our previous notation $E_x^c = v$, the realised value of the total random waiting time $\mathbf{V}$. If we assume that we observe N individuals as before, and that the force of mortality is a constant μ, then a Poisson model is given by the assumption that $\mathbf{D}$ has a Poisson distribution with parameter μE_x^c. That is:

$$\mathrm{P}[\mathbf{D} = d] = \frac{e^{-\mu E_x^c}(\mu E_x^c)^d}{d!}. \tag{43}$$

This model was described by Forfar, McCutcheon & Wilkie (1988), and it is of direct interest, as it was adopted by the CMIB for the graduation of the extensive standard tables based on the 1979–82 experiences (CMIR **9**, 1988).

Under the observational plan described above, the Poisson model is not an exact model, since it allows a non-zero probability of more than N deaths, but it is often a good approximation. Alternatively, we might adjust the observational plan so that the Poisson model is exact. Examples of suitable observational plans are:

(a) to continue observation until the waiting time reaches a pre-determined value; or

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

(b) to replace each life who dies with an identical and independent life at the moment of death (Scott, 1982).

Clearly plan (b) is impractical. Plan (a) is practical, but most actuarial investigations are not of this type. Sverdrup (1965) pointed out that under plan (a) above, **D** has a Poisson (μE_x^c) distribution (in our notation), and that the model extends to multiple decrements, but that:

> "... in most practical cases neither the period of experimentation, nor the number of persons involved or the waiting time would be fixed in advance."

Scott (1982) noted that under plan (b) the assumption of a constant force of mortality can be dropped, and that the estimator $\tilde{\mu}_x$ below then estimates the integrated hazard $\int_0^1 \mu_{x+t}\,dt$ (which, somewhat confusingly, he called m_x). Estimation of the integrated hazard is of great interest in survival analysis (see Part II, Section 6), but usually with even stronger assumptions of piecewise constancy than we have used here.

The Poisson likelihood leads to the following estimator of (constant) μ:

$$\tilde{\mu} = \frac{\mathbf{D}}{E_x^c} \tag{44}$$

with the following properties:

$$\mathrm{E}[\tilde{\mu}] = \mu \qquad \mathrm{Var}[\tilde{\mu}] = \frac{\mu}{E_x^c} \tag{45}$$

and, in practice, we will substitute $\hat{\mu}$ for μ to estimate these from the data. Under the two state model, $\mathrm{E}[\tilde{\mu}] = \mu$ and $\mathrm{Var}[\tilde{\mu}] = \mu/\mathrm{E}[\mathbf{V}]$, but the true values of μ and $\mathrm{E}[\mathbf{V}]$ are unknown and must be estimated from the data as $\hat{\mu}$ and E_x^c respectively. So, although the estimators are different, we obtain the same numerical estimates of the parameter and of the moments of the estimator, in either case. Furthermore, there is a Poisson central limit theorem (Hoem, 1987) which shows that the asymptotic distribution of **D** is Poisson. These considerations might tempt the pragmatic actuary to turn aside from the deeper mathematics of the two state approach, which would be a mistake, for the two state model is easily extended to arbitrary transitions while the Poisson model is not.

A short and clear summary of estimation in decrement models can be found in Gerber (1990, Chapter 11).

4.4 *Estimation of Central Rates of Decrement*

Traditionally, d/E_x^c would be taken to estimate m_x, the central rate of mortality. In a statistical framework, we ask what is the underlying model, in which central rates arise as the quantities which are estimated by the statistic

https://doi.org/10.1017/S1357321700003366 Published online by Cambridge University Press

D/**V** or perhaps **D**/E_x^c? And if such a model can be found, what are the statistical properties of these estimators? We emphasise that we have no argument with the statement that d/E_x^c can be assumed to estimate m_x; anything at all can be assumed to estimate m_x; we are concerned to provide a statistical rationale.

In a probabilistic model, both forces of decrement and probabilities of decrement can be interpreted sensibly, even if we consider a single life. From such a starting point, we have been able to derive estimators for μ_x or q_x based on a finite sample of lives; but the very concept of a central rate of decrement is based on a highly idealised population of lives, so that m_x is unlikely to emerge as an observable quantity in a realistic setting. Further, the population in question is the stationary distribution of lives whose mortality is governed by the force of decrement, and it is unlikely that the observed lives will be so distributed. It is simplest to regard the forces as the fundamental model quantities, and to obtain m_x as a derived quantity.

Forfar, McCutcheon & Wilkie (1988) did offer m_x as an alternative to μ_x for the construction of graduated estimates, but, unlike their discussion of q_x and μ_x, the treatment of m_x was not carried out within the framework of a model.

Acknowledgements

I would like to thank Philip Cooper, Howard Waters and David Wilkie for helpful comments on earlier drafts of this paper.

References

Aalen, O.O. (1978). Non-parametric inference for a family of counting processes. *The Annals of Statistics*, **6**, 701–726.

Aalen, O.O. (1987). Dynamic modelling and causality. *Scandinavian Actuarial Journal*, **1987**, 177–190.

Andersen, P.K. & Borgan, Ø. (1985). Counting process models for life history data: a review. *Scandinavian Journal of Statistics*, **12**, 97–158.

Andersen, P.K., Borgan, Ø., Gill, R.D. & Keiding, N. (1993). *Statistical models based on counting processes*. Springer-Verlag, New York.

Andersen, P.K. & Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, **10**, 1100–1120.

Arnold, B.C. & Brockett, P.L. (1983). Identifiability for dependent multiple decrement/competing risks models. *Scandinavian Actuarial Journal*, **1983**, 117–127.

Batten, R.W. (1978). *Mortality table construction*. Prentice-Hall, Englewood Cliffs, N.J.

Bailey, W.G. & Haycocks, H.W. (1946). *Some theoretical aspects of multiple decrement tables*. T. and A. Constable Ltd.

Bailey, W.G. & Haycocks, H.W. (1947). A synthesis of methods of deriving measures of decrement from observed data (with discussion). *J.I.A.* **73**, 179–212.

Benjamin, B. & Pollard, J.H. (1980). *The analysis of mortality and other actuarial statistics*. Heinemann, London.

BORGAN, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics*, **11**, 1–16.

BOWERS, N.L., GERBER, H.U., HICKMAN, J.C., JONES, D.A. & NESBITT, C.J. (1986). *Actuarial mathematics*. Society of Actuaries, Itasca, Il.

BRESLOW, N.E. (1974). Covariance analysis of censored survival data. *Biometrics*, **30**, 89–100.

BRESLOW, N.E. & CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, **2**, 437–453.

BRESLOW, N.E. (1993). Introduction to Kaplan & Meier (1958) 'Nonparametric estimation from incomplete observations', in KOTZ, S. & JOHNSON, N.L. (1993). *Breakthroughs in statistics II: methodology and distribution*. Springer-Verlag, New York, 319–337.

BROFFIT, J.D. (1984). Maximum likelihood alternatives to actuarial estimators of mortality rates (with discussion). *Transactions of the Society of Actuaries*, **XXXVI**, 77–142.

CARRIÈRE, J.F. (1994). Dependent decrement theory. To appear in *Transactions of the Society of Actuaries*, **XLVI**.

CLARKE, R.D. (1978). Mortality of impaired lives (with discussion). *J.I.A.* **105**, 15–46.

CLAYTON, D. (1988). The analysis of event history data: a review of progress and outstanding problems. *Statistics in Medicine*, **7**, 819–841.

COLLETT, D. (1994). *Modelling survival data in medical research*. Chapman & Hall, London.

CONTINUOUS MORTALITY INVESTIGATION BUREAU (CMIB) (1988). *CMIR* **9**.

CONTINUOUS MORTALITY INVESTIGATION BUREAU (CMIB) (1993). Calculation of continuation tables and allowance for non-recorded claims based on the PHI experience 1975–78. *CMIR* **13**, 123–130.

COX, D.R. (1972). Regression models and life-tables (with discussion). *J.R.S.S. B*, **34**, 187–220.

COX, D.R. (1975). Partial likelihood. *Biometrika*, **62**, 269–276.

COX, D.R. & HINKLEY, D.V. (1974). *Theoretical statistics*. Chapman & Hall, London.

COX, D.R. & OAKES, D. (1984). *Analysis of survival data*. Chapman & Hall, London.

CROWDER, M.J. (1991). On the identifiability crisis in competing risks analysis. *Scandinavian Journal of Statistics*, **18**, 223–233.

CROWDER, M. (1994). Identifiability crises in competing risks. *International Statistical Review*, **62**, 379–391.

DAYKIN, C.D., CLARK, P.N.S., EVES, M.J., HABERMAN, S., LE GRYS, D.J., LOCKYER, J., MICHAELSON, R.W. & WILKIE, A.D. (1988). The impact of HIV infection and AIDS on insurance in the United Kingdom. *J.I.A.* **115**, 727–838.

DAVID, H.A. & MOESCHBERGER, M.L. (1978). *The theory of competing risks*. Griffin, London.

DORRINGTON, R.E. & SLAWSKI, J.K. (1993). A defence of the conventional actuarial approach to the estimation of the exposed-to-risk. *Scandinavian Actuarial Journal*, **1993**, 107–113.

ELANDT-JOHNSON, R.C. & JOHNSON, N.L. (1980). *Survival models and data analysis*. John Wiley, New York.

ENGLAND, P.D. (1993). *Statistical modelling of excess mortality of medically impaired insured lives*. Ph.D. thesis, City University, London.

FLEMING, T.R. & HARRINGTON, D.P. (1991). *Counting processes and survival analysis*. John Wiley, New York.

FORFAR, D.O., MCCUTCHEON, J.J. & WILKIE, A.D. (1988). On graduation by mathematical formula (with discussion). *J.I.A.* **115**, 1–149 and 693–698, and *T.F.A.* **41**, 97–269.

GAIL, M. (1975). A review and critique of some models used in competing risk analysis. *Biometrics*, **31**, 209–222.

GEHAN, E.A. (1965). A generalised Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, **52**, 203–223.

GERBER, H.U. (1990). *Life insurance mathematics (English edition)*. Springer-Verlag, Berlin, and the Swiss Association of Actuaries, Zurich.

GILL, R.D. (1980). Censoring and stochastic integrals. *Mathematical Centre Tracts*, **124**, Mathematisch Centrum, Amsterdam.

GILL, R.D. (1984). Understanding Cox's regression model: a martingale approach. *Journal of the American Statistical Association*, **79**, 441–447.

GILL, R.D. & JOHANSEN, S. (1990). A survey of product-integration with a view toward application in survival analysis. *The Annals of Statistics*, **18**, 1501–1555.

GREENWOOD, M. (1926). The errors of sampling of the survivorship tables. *Reports on Public Health and Statistical Subjects*, **33**, Appendix 1. H.M.S.O., London.

HAYCOCKS, H.W. & PERKS, W. (1955). *Mortality and other investigations, Vol. 1.* Cambridge University Press.

HECKMAN, J.J. & HONORE, B.E. (1989). The identifiability of the competing risks model. *Biometrika*, **76**, 325–330.

HOEM, J.M. (1969). Markov chain models in life insurance. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik*, **9**, 91–107.

HOEM, J.M. (1976). The statistical theory of demographic rates (with discussion). *Scandinavian Journal of Statistics*, **3**, 169–185.

HOEM, J.M. (1984). A flaw in actuarial exposed-to-risk theory. *Scandinavian Actuarial Journal*, **1984**, 187–194.

HOEM, J.M. (1987). Statistical analysis of a multiplicative model and its application to the standardization of vital rates: a review. *International Statistical Review*, **55**, 119–152.

HOEM, J.M. (1988). The versatility of the Markov chain as a tool in the mathematics of life insurance. *Transactions of the 23rd International Congress of Actuaries, Helsinki* **S**, 171–202.

HOEM, J.M. & FUNCK-JENSEN, U. (1982). Multistate life table methodology: a probabilist critique. In *Multidimensional mathematical demography*, eds. LAND, K.C. & ROGERS, A., 155–264, Academic Press.

HOGG, R. & KLUGMAN, S. (1984). *Loss distributions.* John Wiley, New York.

JACOBSEN, M. (1982). Statistical analysis of counting processes. *Lecture Notes in Statistics*, **12**, Springer-Verlag, New York.

JEWELL, W.S. (1980). Generalized models of the insurance business (life and/or non-life insurance). *Transactions of the 21st International Congress of Actuaries, Zurich & Lausanne*, **S**, 87–141.

JONES, B.L. (1994). Actuarial calculations using a Markov model. To appear in *Transactions of the Society of Actuaries*, **XLVI**.

KALBFLEISCH, J.D. & PRENTICE, R.L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, **60**, 267–278.

KALBFLEISCH, J.D. & PRENTICE, R.L. (1980). *The statistical analysis of failure time data.* John Wiley, New York.

KAPLAN, E.L. & MEIER, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481. Reprinted in KOTZ, S. & JOHNSON, N.L. (1993). *Breakthroughs in statistics II: methodology and distribution.* Springer-Verlag, New York, 319–337.

KARR, A.F. (1991). *Point processes and their statistical inference (2nd edition).* Marcel Dekker, New York.

KULKARNI, V.G. (1995). *Modeling and analysis of stochastic systems.* Chapman & Hall, London.

LANCASTER, T. (1990). *The econometric analysis of transition data.* Cambridge University Press.

LINDSEY, J.C. & RYAN, L.M. (1993). A three-state multiplicative model for rodent tumorigenicity experiments. *Applied Statistics*, **42**, 283–300.

MAKEHAM, W.M. (1874). On an application of the theory of the composition of decremental forces. *J.I.A.* **18**, 317–322.

MARSHALL, A.W. & OLKIN, I. (1967). A bivariate exponential distribution. *Journal of the American Statistical Association*, **62**, 30–44.

NEILL, A. (1977). *Life contingencies.* Heinemann, London.

PAPATRYANDAFYLOU, A. & WATERS, H.R. (1984). Martingales in life insurance. *Scandinavian Actuarial Journal*, **1984**, 210–230.

PETO, R., PIKE, M.C., ARMITAGE, P., BRESLOW, N.E., COX, D.R., HOWARD, S.V., MANTEL, N., MCPHERSON, K., PETO, J. & SMITH, P.G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *British Journal of Cancer*, **35**, 51–67.

PRENTICE, R.L., KALBFLEISCH, J.D., PETERSON, A.V., JR., FLOURNOY, N.S., FAREWELL, V.T. & BRESLOW, N.E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541–554.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics*, **11**, 453–466.

RAMSAY, C.M. (1989). AIDS and the calculation of life insurance functions (with discussion). *Transactions of the Society of Actuaries*, **XLI**, 393–422.

RENSHAW, A.E. (1988). Modelling excess mortality using GLIM. *J.I.A.* **115**, 299–315.

SCHOU, G. & VÆTH, M. (1980). A small sample study of occurrence/exposure rates for rare events. *Scandinavian Actuarial Journal*, **1980**, 209–205.

SCOTT, W.F. (1982). Some applications of the Poisson distribution in mortality studies. *T.F.A.* **38**, 255–263.

SEAL, H.L. (1977). Studies in the history of probability and statistics. XXXV Multiple decrements or competing risks. *Biometrika*, **64** 429–439.

SMITH, A.D. (1991). The use of martingales in actuarial work. *Transactions of the 2nd A.F.I.R. International Colloquium, Brighton*, **4**, 39–81.

SPRAGUE, T.B. (1879). On the construction of a combined marriage and mortality table from observations made as to the rates of marriage and mortality among any body of men. *J.I.A.* **21**, 406–452.

SVERDRUP, E. (1965). Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between states of health. *Skandinavisk Aktuaritidskrift*, **48**, 184–211.

TSIATIS, A.A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences, U.S.A.*, **72**, 20–22.

WATERS, H.R. (1984). An approach to the study of multiple state models. *J.I.A.* **111**, 363–374.

WATERS, H.R. (1991a). A multiple state model for permanent health insurance. *Continuous Mortality Investigation Bureau Report*, **12**, 5–20.

WATERS, H.R. (1991b). Computational procedures for the model. *Continuous Mortality Investigation Bureau Report*, **12**, 79–96.

WATERS, H.R. & WILKIE, A.D. (1987). A short note on the construction of life tables and multiple decrement tables. *J.I.A.* **114**, 569–580.

WHITEHEAD, J. (1980). Fitting Cox's regression model to survival data using GLIM. *Applied Statistics*, **29**, 268–275.

WILKIE, A.D. (1988a). An actuarial model for AIDS. *J.I.A.* **115**, 839–853.

WILKIE, A.D. (1988b). Markov models for combined marriage and mortality tables. *Transactions of the 23rd International Congress of Actuaries, Helsinki*, **3**, 473–486.

WOOD, J.W., HOLMAN, D.J., YASHIN, A.I., PETERSON, R.J., WEINSTEIN, M. & CHANG, M.-C. (1994). A multistate model of fecundability and sterility. *Demography*, **31**, 403–426.