MULTISTATE LIFE-TABLES AS REPEATED MEASURE MODELS

C.R. HEATHCOTE

A multistate life-table typically describes the evolution of a cohort from birth at time $t = 0$ to death at or before some maximum age $t = m$. The entries in the table are the counts of the number of individuals in each of several states for ages $t = 0, 1,\dots,m,$ with the states falling into two categories, an "alive" or transient set numbered say $1, 2, \dots,a,$ and a "death" or absorbing set numbered $a+1, a+2,\dots, a+d$. The a transient states may for example be a different states of health and the d absorbing states different risks of death. For a cohort of $\ell(0)$ individuals we then have $\ell(0)$ repeated counting measures over $[0, m]$ in the sense that the data consist of individual life indicators for $t = 0, 1,\dots,m$. Relevant references are Chiang (1968) and Chapter 8 of Crowder and Hand (1990). We shall be concerned with probabilistic structure and not inference, and hope to show how life tables can be placed into the context of the workshop.

Life tables are usually displayed as a series of rows indexed by age t but it is convenient for our purposes to stand this arrangement on its side and interpret the life table counts as a realisation of a stochastic process with the usual convention of (discrete) time running from left to right as in Table 1. For convenience time and age are measured on the same scale and from the same origin.

Table 1. A multistate life table depicted as a stochastic process. For $t = 0, 1,\ldots,m$, $\ell_i(t)$ = no. in transient state i, $c_h(t)$ = cumulative count in absorbing state h.

Absorbing states	a+d	0	$c_{a+d}(1)$	- - - - - - -	$c_{a+d}(m)$
		.	.		
		.	.		
	a+1	0	$c_{a+1}(1)$	- - - - - - -	$c_{a+1}(m)$
Transient states	a	$\ell_a(0)$	$\ell_a(1)$	- - - - - - -	0
		.	.		
		.	.		
	2	$\ell_2(0)$	$\ell_2(1)$	- - - - - - -	0
	1	$\ell_1(0)$	$\ell_1(1)$	- - - - - - -	0
		0	1	Time t →	m

Table 1 illustrates the north-easterly drift of the initial frequencies $\ell_1(0),\ldots,\ell_a(0)$ of the transient states,

$$\sum_{i=1}^{a} \ell_i(0) = \ell(0) = \text{size of cohort},$$

to the cumulative frequencies $c_{a+1}(m),\ldots,c_{a+d}(m)$ of the absorbing states at $t = m$,

$$\sum_{h=a+1}^{a+d} c_h(m) = \ell(0).$$

Movement between the transient states is permitted but this of course is not the case with the absorbing states.

For a cohort of $\ell(0)$ individuals the counts in Table 1 can be viewed as resulting from the superposition of $\ell(0)$ stochastic processes, each of which trace the life history of a particular individual. If also, as is usually assumed, the individuals are independent we have the superposition of $\ell(0)$ independent processes.

The special feature of the individual processes is that they are sequences of indicator variables as follows. For a given individual let $X(t)$ denote the state occupied at time t so that $\{X(t), t = 0, 1,\ldots,m\}$ is a stochastic process with state space $\{1, 2,\ldots,a,a+1,\ldots,a+d\}$. Introduce

$$I_{ik}(0,t) = \begin{cases} 1 \text{ if } X(t) = k \text{ given } X(0) = i \\ 0 \text{ if } X(t) \neq k \text{ given } X(0) = i \end{cases} .$$

Put $\underset{\sim}{I}_i'(0,t) = (I_{i1}(0, t), I_{i2}(0, t),\ldots,I_{ia+d}(0, t))$. Then with initial condition

$$I_{ik}(0,0) = \delta_{ik} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{otherwise} \end{cases}$$

and $i \in \{1, 2,\ldots,a\}$, the vector process $\{\underset{\sim}{I}_i(0, t), t = 0, 1,\ldots,m\}$ describes the life history of an individual. The point of view adopted here is that most, if not all, life table functions of interest can be conveniently studied through such sequences of indicators. For example, of the $\ell_i(0)$ individuals in state i at $t = 0$,

$$\ell_{ik}(t) = \sum_{j=1}^{\ell_i(0)} I_{ik}^{(j)}(0,t)$$

is the total number alive in state k at time t, where $I_{ik}^{(j)}(0,t), j = 1,2,\ldots,\ell_i(0)$ are identically distributed as $I_{ik}(0,t)$. Also, with

$$I_k(t) = \sum_{i=1}^{a} I_i(0)\, I_{ik}(0,t)$$

the indicator of the event "individual is in state k at time t", the sum

$$\ell_k(t) = \sum_{j=1}^{\ell(0)} I_k^{(j)}(t), \quad k \in \{1, 2, \ldots, a\},$$

is the number alive and in state k at t. Their sum

$$\ell(t) = \sum_{k=1}^{a} \ell_k(t)$$

is the total number alive at t. As is well known (Chiang, 1968) $\ell(t)$ is distributed as the convolution of multinomials for a cohort of independent individuals.

More generally for $s \le t$ let

$$I_{ik}(s,t) = \begin{cases} 1 \text{ if } X(t) = k & \text{given } X(s) = i \\ 0 \text{ if } X(t) \ne k & \text{given } X(s) = i \end{cases}.$$

Then

$$W_{ih}(s,t) = \sum_{\ell=1}^{a} I_{i\ell}(s,t\text{-}1)\, I_{\ell h}(t\text{-}1,t)$$

is the indicator of the event "death occurs at time t of risk h given $X(s) = i$". The joint distribution of $W_{ih}(s,t)$, $h = a+1,\ldots,a+d$; $t = s+1,\ldots,m$ is clearly of interest in the study of competing risks. The basic probability structure is multinomial, with a convenient normal approximation, again provided that the individuals are independent.

To illustrate the results that can be obtained consider the total occupation time of state k by an individual of state i,

$$V_{ik} = \sum_{t=0}^{m} I_{ik}(0,t).$$

If $V_{ik}^{(j)}$, $j = 1, 2,\ldots,\ell_i(0)$, are independent and identically distributed as V_{ik}, then

$$L_{ik} = \frac{1}{\ell_i(0)} \sum_{j=1}^{\ell_i(0)} V_{ik}^{(j)}$$

is the average occupation time of state k, commonly called the *life expectancy* of state k of an individual born into state i. Under the assumption of independence

$$E \exp[\xi\sqrt{\ell_i(0)}\,(L_{ik} - EL_{ik})] = E \exp\left[\left(\frac{\xi}{\sqrt{\ell_i(0)}} \sum_{j=1}^{\ell_i(0)} (V_{ik}^{(j)} - EV_{ik})\right]\right.$$

$$= \{\{E \exp \frac{\xi}{\sqrt{\ell_i(0)}}(V_{ik} - EV_{ik})\}^{\ell_i(0)}$$

$$= \{1 + \frac{\xi^2}{2\ell_i(0)} \text{ Var } V_{ik} + o(\ell_i(0))\}^{\ell_i(0)} .$$

Therefore

$$\lim_{\ell_i(0)\to\infty} E \exp\{\xi\sqrt{\ell_i(0)}\,(L_{ik} - EL_{ik}) = \exp \frac{\xi^2}{2} \text{ Var } V_{ik}.$$

With

$$p_{ik}(0,t) = \Pr(X(t) = k \mid X(0) = i)$$

it follows that

$$\text{Var } I_{ik}(0,t) = p_{ik}(0,t)\{1 - p_{ik}(0,t)\}$$

$$\text{Cov}(I_{ik}(0,t), I_{ik}(0,s)\,) = \Pr(X(t) = k, X(s) = k \mid X(0) = i) - p_{ik}(0,t)\; p_{ik}(0,s), \quad t < s$$

$$= p_{ik}(0,t)\; \{p_{kk}(t,s) - p_{ik}(0,s)\} ,$$

the last relation following if $\{X(t)\}$ is Markov. Also

$$EL_{ik} = EV_{ik} = \sum_{t=0}^{m} p_{ik}(0,t) ,$$

the expected occupation time of state k. Hence for large $\ell_i(0)$, the life expectancy L_{ik} is approximately normally distributed with this mean and variance

$$\frac{1}{\ell_i(0)}\{\sum_{t=1}^{m} \text{Var } I_{ik}(0,t) + \sum_{t} \sum_{< s} \text{Cov}(I_{ik}(0,t), I_{ik}(0,s))\}.$$

Whilst the transition probabilities can be easily estimated from longitudinal data the situation is more complicated if only aggregate counts are available (see e.g. Ch. 8 of Crowder and Hand, 1990).

In summary, the relevant point about life tables is that they provide an example of repeated categorical observations over a fixed time interval with multinomials and their normal approximations as the basic probabilistic tools. Particular features are the presence of absorbing states and the sorts of questions asked. The inferential problems are the familiar ones arising with aggregated count data.

REFERENCES

Chiang, C.L. (1968). *Introduction to stochastic processes in biostatistics.* Wiley. New York.

Crowder, M.J. and Hand, D.J. (1990). *Analysis of repeated measures.* Chapman and Hall. London.

Department of Statistics
Faculty of Economics and Commerce
GPO Box 4
Canberra ACT 2601
Australia