

## UNIT 7 — BINOMIAL AND POISSON MODELS

- Syllabus objectives*
- (iv) Derive maximum likelihood estimators for the transition intensities in models of transfers between states with piecewise constant transition intensities.
4. Describe the Poisson approximation to the estimator in 3. in the case of a single decrement and its advantages and disadvantages.
- (v) Describe the Binomial model of mortality, derive a maximum likelihood estimator for the probability of death and compare the Binomial model with the multiple state models.
1. Describe the Binomial model of the mortality of a group of identical individuals subject to no other decrements between two given ages.
  2. Derive the maximum likelihood estimator for the rate of mortality in the Binomial model and its mean and variance.
  3. Describe the advantages and disadvantages of the multiple state model and the Binomial model, including consistency, efficiency, simplicity of the estimators and their distributions, application to practical observational plans and generality.

### 1 Binomial-type models

1.1 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.

1.2 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.

- 1.3 In terms of the data defined in Unit 5, the  $\{a_i\}$  and the  $\{b_i\}$  are in general not all the same. Considering the  $i^{\text{th}}$  life, we have:

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

In respect of this individual, the above 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:

$$\vec{q} = ({}_{b_1 - a_1}q_{x+a_1}, {}_{b_2 - a_2}q_{x+a_2}, \dots, {}_{b_N - a_N}q_{x+a_N})$$

$$\vec{d} = (d_1, d_2, \dots, d_N)$$

we can write the total likelihood as:

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

- 1.4 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 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*:  ${}_tq_x = tq_x \quad (0 \leq t \leq 1)$
- (b) *the Balducci assumption*:  ${}_{1-t}q_{x+t} = (1-t)q_x \quad (0 \leq t \leq 1)$
- (c) *constant force of mortality*:  ${}_tq_x = 1 - e^{-\mu t} \quad (0 \leq t \leq 1).$

(Note that the Balducci assumption implies a decreasing force of mortality between integer ages.)

## 2 The actuarial estimate

2.1 Under the Balducci assumption:

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

2.2 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 would usually give the moment estimate of  $q_x$ . However, the right side also involves expected deaths in such a way that it is impossible to extract all the terms in  $E[\mathbf{D}]$  and the  $\{E[\mathbf{D}_i]\}$  on one side and all the terms in  $q_x$  on the other.

2.3 Summing the last term over the observed rather than the expected survivors, we obtain:

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

leading to the estimate:

$$\hat{q}_x = \frac{d}{\sum_{i=1}^N (1 - a_i) - \sum_{i:\mathbf{D}_i=0} (1 - b_i)}$$

in which the denominator is called the 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  $x + \frac{1}{2}$ , and ignoring the awkward possibility that  $a_i > \frac{1}{2}$ , we obtain the formula:

$$\hat{q}_x = \frac{d}{E_x^c + \frac{1}{2}d} .$$

(Recall that  $E_x^c = v$ , the observed total waiting time at age  $x$ .)

Note that it is only an approximate moment estimate of  $q_x$ .

2.4 This is known to statisticians as the *actuarial estimate*. The Binomial model, and the actuarial estimate, are not without strengths. The actuarial estimate avoids numerical solution of equations and it might be used if there is a compelling reason to estimate  $q_x$  instead of something else. And, as we have seen, the Binomial model can be generalised

simply to give a non-parametric estimate — the Kaplan-Meier estimate — which is widely used in survival analysis.

- 2.5** However, it cannot be said that the actuarial estimate is any simpler than the estimates based on multiple state models. Indeed, if the exposure data are of the census type (see Unit 10), the need to compute an initial exposed-to-risk is a pointless complication. 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; 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.

### 3 Poisson models

- 3.1** 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.

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

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

- 3.3** 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 (but not necessarily practicable) observational plans are:

- (a) to continue observation until the waiting time reaches a pre-determined value; or
- (b) to replace each life who dies with an identical and independent life at the moment of death.

**3.4** The Poisson likelihood leads to the following estimator of (constant)  $\mu$ :

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

with the following properties:

$$E[\tilde{\mu}] = \mu$$

$$\text{Var}[\tilde{\mu}] = \frac{\mu}{E_x^c}$$

and, in practice, we will substitute  $\hat{\mu}$  for  $\mu$  to estimate these from the data. Under the two-state model,  $E[\tilde{\mu}] = \mu$  and  $\text{Var}[\tilde{\mu}] = \mu / E[V]$ , but the true values of  $\mu$  and  $E[V]$  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.

## 4 Comparison of multiple-state, Binomial and Poisson models

**4.1** When we compare models, we distinguish three aspects:

- (a) how well each model represents the process we are trying to model;
- (b) how easy it is to find, characterise and use the model parameters, given the data with which we must work; and
- (c) how easily each model is extended to problems other than the study of human mortality.

**4.2** The underlying process we take to be the time(s) of death of one or more lives, considered to be indistinguishable, except in respect of their deaths. If death is the only decrement, this leads to the two specifications:

- (a) representing the time of death by the random variable  $T_x$ ; or
- (b) the two-state model parameterised by  $\mu_{x+t}$ .

They can usually be taken to be equivalent, since under reasonable conditions we can derive the force of mortality starting with (a), while we can obtain the distribution of the time to death starting with (b).

It is evident that the two-state model represents the process closely (in fact, almost by definition), while the Binomial model represents a restricted view of the process, since it represents only the year of death, and not the time of death. This suggests that if sufficient data are available to use the two-state model, then using the Binomial model instead will not make the fullest use of the information. This turns out to be the case.

Unless the waiting times  $E_x^c$  are fixed in advance, which would be unusual in actuarial work, the Poisson model is an approximation to the multiple-state model, in which  $E_x^c$  is regarded as non-random. This is acceptable if  $\mu$  is small.

Note that both formulations above are non-parametric in the usual sense, although both can be regarded as parameterised by a function:  $\mu_{x+t}$  in (b) and  $F_x(t)$  or  $f_x(t)$  in (a). Direct estimation of these functions (integrated in the case of  $\mu_{x+t}$ ) leads to the Kaplan-Meier or Nelson-Aalen estimates.

#### 4.3 Parametric models are obtained if we restrict attention to single years of age, and in the two-state model also assume a constant transition intensity.

The form and statistical properties of the parameter estimates (and how easy they are to find and to use) depend on the form of the likelihoods, which in turn depend on the available data.

If the exact dates of birth, entry to and exit from observation, and death (if observed) are all known, then:

- (a) we can calculate exactly the MLE of  $\mu$  in the two-state model;
- (b) the Binomial model (or more accurately, Bernoulli model) based on individual lives is complicated, and further assumptions (such as the Balducci assumption) are needed to get results.

The consequence of (b) above is that the Binomial estimate of  $q_x$  has a higher variance than the estimate  $\hat{q}_x = 1 - \exp(\hat{\mu})$  obtained from the two-state model. However, the difference is tiny unless  $\mu$  is extremely high. (A rule of thumb, due to Sverdrup, is that if  $\mu$  is very small, most of the information is in the number of deaths, while if  $\mu$  is very large, most of the information is in the times of death.) Likewise, when  $\mu$  is very small, the actuarial estimate  $\hat{q}_x = d_x / (E_x^c + \frac{1}{2}d_x)$  provides acceptable results.

Often, not all the dates of the relevant events are known, and then the MLE of  $\mu$  in the two-state model must also be approximated. In at least one important case (the Continuous Mortality Investigation Bureau studies) the data allow easy approximation of the waiting times  $E_x^c$  (see Unit 10).

In terms of computation, therefore, the two-state model is preferred if complete life histories are available; otherwise both the two-state and Binomial models require some

degree of numerical approximation at the estimation stage. There is no difference in practice between the two-state and Poisson models, because the maximum likelihood estimates are the same (though the estimators are not).

**4.4** The statistical properties of the MLEs in the various models differ slightly.

- (a) In the multiple-state model, the MLE is consistent and asymptotically unbiased; the variance of the estimator is also only available asymptotically. Simulation experiments suggest that the results are reasonable if  $d_x \leq 10$ .
- (b) In the Poisson model, the MLE is consistent and unbiased. Its mean and variance are available exactly in terms of the true  $\mu$ , but are estimated from the data by the same expressions as estimate the asymptotic mean and variance in the two-state model.
- (c) In the “naive” Binomial model, in which  $N$  identical lives are observed for exactly one year, the MLE is consistent and unbiased, and the exact mean and variance can be obtained in terms of the true  $q_x$ . In practice, the data rarely conform to the “naive” model, so only approximate results are available.

When  $\mu$  is very small, there are few reasons to prefer any one of these models on the basis of the statistical properties of the MLEs alone.

**4.5** Finally we consider the generality of the models, chiefly how easily they can be extended to more complicated processes than one decrement, and how effective they are when forces of transition are high compared with typical human mortality.

- (a) The Markov multiple-state model is extended very simply as we have seen. No matter how complex the model, the estimators have the same simple form and statistical properties, depend only on data that will often be available exactly or approximately, and the apparatus needed in applications (such as the Kolmogorov equations) carries over without difficulty. Further extensions are possible, which complicate the calculation of probabilities but not the estimation of parameters, for example semi-Markov models.
- (b) The Poisson model extends just as easily to multiple decrements, but not to processes with increments.
- (c) There are considerable difficulties in extending the Binomial model even to multiple decrements. It is relatively simple to extend the ordinary life table to multiple decrement tables, and these have long been used by actuaries. However, extending the life table (essentially a computational tool) is very far from extending any underlying probabilistic model, and, when the matter was investigated, it was found not to be a simple task (we omit details). Extension of the life table to increments is also not too hard, but extension of the Binomial model is harder still.

If transition intensities are high, the loss of information (times of transitions) under the Binomial model becomes more serious, while the Poisson model becomes a poorer approximation to the multiple-state model (because there is more randomness in the waiting times).

#### 4.6

In conclusion, when studying ordinary human mortality, transition intensities are so low that none of the models considered stands out on statistical grounds alone. This is why actuaries have used life tables so successfully for so long.

However, when we must model more complicated processes or higher transition intensities, which is increasingly the case as new insurance products are developed, the multiple state approach appears to offer significant advantages.

It may still be the case that a simplified approach is ultimately adopted, for example for calculations to be made by office staff, but it is best to begin with a specification which most nearly represents the process being modelled, and then make approximations as required for estimation and in applications.

**END**