University of Illinois   Department of Economics
Fall 2001   Roger Koenker

Economics 478
Lecture 8
Introduction to Multi-State Survival Models

As an introduction to multi-state survival models I thought it might be useful to consider the problem of constructing a data structure and survival model for a problem that is (supposedly) well understood, the academic publishingly process. I believe that this is a useful illustrative exercise and will help in digesting the material in Chapter 8 of T&G.

Figure 1 gives a concise schematic of the process. Papers are submitted from state -1, are refereed in state 0, from which there are three destinations: accept, reject or revise. Papers designated as revise enter state 1 and emerge eventually to return to state 0 to be rerefereed.

In Table 1 we illustrate 3 representative examples of sample paths for the process.
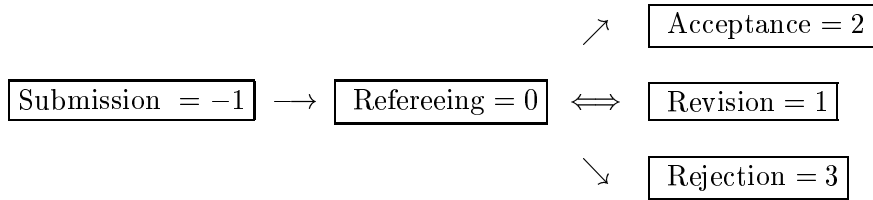
id	t_0	s_{-1}	s_0	s_1	t_1
1	t_0	-1	0	3	t_1
2	t_0	-1	0	1	t_1
2	t_1	0	1	1	t_2
3	t_0	-1	0	1	t_1
3	t_1	0	1	0	t_2
3	t_2	1	0	1	t_3
3	t_3	0	1	0	t_4
3	t_4	1	0	1	t_5
3	t_5	0	1	0	t_6
3	t_6	1	0	2	t_7

In case 1 we have a paper rejected after first round of refereeing. In the second the paper enters the revision state but never emerges due to censoring. In the third the paper undergoes 3 revisions before it is finally accepted – this means that it is refereed 4 times. Of course it may have

1

already been revised several times before submission. The big question here is how much of the history should enter the modeling.

First consider the stratification. We have 3 possible destination states from the refereeing state: accept, reject, revise. Of these, two are obviously terminal and one is not.[1] From the revision state there is only one outcome (of course, we can have censoring that indicates a revision wasn't completed in the span of the sample period. We count 4 possible transitions so in T&G terminology we have 4 strata. Now in the counting process formulation we have for each transition from the revision state we have one record from each refereeing state we have 3 records. This implies, perhaps not entirely in accord with intuition, that we consider a paper in the refereeing state 0 as "at risk" for all three outcomes. When a decision is reached we record two censored event times and one observed event time.

The next stage of the modeling exercise would be to decide how the co-variates would be handled. In particular, this would entail deciding how much of the prior history was incorporated into the parametric component of the model. Should the number of prior rounds be used, or the time in prior rounds, should we distinguish revision and referee time in the latter case, etc. etc. But these are not dissimilar to questions that we are familiar with from previous linear modeling exercises so we will leave them for class discussion.

---

[1]this is a bit like academic heaven, hell, and purgatory.