A.3 Markov models, censoring and Kaplan-Meier estimator

1. A company manufactures electrical devices which have two main components. These two components are believed to fail independently and at constant rates while the devices are in operation. In order to study whether these rates might depend on environmental conditions, the company tested a large number of devices for 1000 hours of operation under three different conditions. For each device, it was recorded whether the machine broke down at some point, and if so which component caused the failure.

   The following table shows the number of devices tested in each environment, and the numbers of failures due to component 1 or due to component 2.

environment	n	$d^{(1)}$	$d^{(2)}$	$q^{(1)}$	$q^{(2)}$	$\hat{\mu}^{(1)}$	$\hat{\mu}^{(2)}$
A (dry)	200	18	34	0.090	0.170	0.104	0.197
B (humid)	500	31	138	0.062	0.276	0.076	0.337
C (normal)	200	18	42				

   (a) Interpret the remaining columns, and explain how the rates $\hat{\mu}^{(1)}$ and $\hat{\mu}^{(2)}$ were obtained. Complete the third line of the table. Explain why $\hat{\mu}_C^{(1)} > \hat{\mu}_A^{(1)}$ even though $d_C^{(1)} = d_A^{(1)}$.

   (b) Test the hypothesis that the failure rates of the components do not depend on environmental conditions.

2. Consider a single-server queueing system, started empty at time 0, observed until time t.

   (a) Denote the arrival rate by λ, the service rate by μ. Write down the likelihood function and determine the maximum likelihood estimators $(\hat{\lambda}, \hat{\mu})$;
   Derive approximations for (i) a $(1-\alpha)$-CI for λ; (ii) a $(1-\alpha)$-CI for μ;
   (iii) a joint $(1-\alpha)$-confidence region for (λ, μ) with minimal area.

   (b) Suppose that the queue length cannot increase beyond m and the length of the queue has an impact on the arrival rates (but not on the service times).

      i. How do you model this situation?

      ii. Derive maximum likelihood estimators.

      iii. Suppose $m = 2$. If $\lambda_0 = \lambda_1$, what is the large-sample distribution of

$$\frac{\hat{\lambda}_0 - \hat{\lambda}_1}{\sqrt{\mathrm{Var}(\hat{\lambda}_0) + \mathrm{Var}(\hat{\lambda}_1)}} \quad \text{or} \quad \frac{(\hat{\lambda}_0 - \hat{\lambda}_1)^2}{\mathrm{Var}(\hat{\lambda}_0) + \mathrm{Var}(\hat{\lambda}_1)}?$$

      iv. Construct an approximate test for $H_0 : \lambda_0 = \lambda_1$ vs $H_1 : \lambda_0 \neq \lambda_1$.

3. A life office uses the three-state healthy-sick-dead model in the pricing of its long term sickness policies. The transition rates are assumed to be constant. Denote the state space by $\mathbb{S} = \{H, S, \Delta\}$ and transition rates by $\sigma = q_{HS}$, $\rho = q_{SH}$, $\delta = q_{H\Delta}$, $\gamma = q_{S\Delta}$.

   For a group of policy holders, over a one-year period the following data were recorded:

transition from	number
H to S	150
H to Δ	60
S to H	50
S to Δ	10

The total times spent in states H and S were 6250 years and 350 years, respectively.

(a) Write down the likelihood function for this model and show that this is maximized when $\sigma = 0.024$.

(b) Construct an approximate 95% confidence interval for σ.

(c) Policyholders pay contributions at rate C when in state H and receive benefits at rate B when in state S. No death benefit is payable. The life office uses the model to set the ratio of contributions to benefits. In terms of σ, ρ, δ and γ, what is the value of the ratio C/B of contribution rate to benefit rate under which the company would break even on average in the long-run? Construct an approximate confidence interval for this value.

4. Public health officials often compare the effects of different changes in population health by considering the resulting change in life expectancy. But what about the personal and societal costs of illness and disability? This has led to the notion of "Quality Adjusted Life Years" (QALYs), in which years are weighted by some measure of the "quality of life". Suppose we have a Markov representation of lifetimes, in which there are m non-absorbing "alive" states, and 1 absorbing "dead" state. The "quality" of a year of life in state i is w_i. Individuals begin in state i with probability p_i.

   (a) Show that the expected total value of a life is given by $-p^T Q_*^{-1} w$, where T means "transpose", and Q_* is the submatrix of Q corresponding to the non-absorbing states.

   Suppose the model is the simple Healthy-Sick-Dead model described in Question 3, with $\sigma = 0.1$, $\rho = 0$, $\delta = 0.01$, and $\gamma = 0.2$.

   (b) What is the life expectancy of someone initially healthy?

   Suppose we weight a year of being sick so as to be equivalent to half of a year of health.

   (c) What is the expected number of QALYs in the lifetime of someone initially healthy? What about the number remaining to someone who has just become sick?

   Suppose that a cure now raises ρ to 0.2.

   (d) How many QALYs would be saved by the cure? Suppose the alternative were a treatment that would lower γ by some amount. Could you achieve the same QALY savings by such a treatment? How low would γ have to go?

   (e) What fraction of individuals are sick right before they die?

   (f) Suppose after 30 years the healthy survivors move into "old age", in which the rate of becoming sick σ rises to 0.5. What is the effect on QALYs?

5. Explain the types of censoring and truncation which are relevant in the following situations:

   (a) Children are tested at monthly intervals from 12 to 18 months of age. On each occasion one of the tests examines the acquisition of the ability to perform a simple task involving tool use. The distribution of the age at which this skill is acquired is studied.

   (b) A study examines the time until the onset of a second episode of clinical depression, in individuals who have suffered a previous episode. An individual can be enrolled in the study if they have suffered precisely one such episode in the past. The study lasts for four years.

(c) Patients who first experience symptoms of malaria after returning to the UK from a malarial region are observed. The period of their stay in the region and the time of onset of symptoms is obtained, and used to study the incubation time of malaria (i.e. the time from infection to symptoms), which is typically on the order of a few weeks.

6. An example of "Type II censoring" is as follows: failure times of n machines are observed, all started at the same moment. Once k machines have failed, observation ceases. The observed data can be written as $x_1, x_2, \ldots, x_k, x_k+, x_k+, \ldots, x_k+$, where $x_1 \leq x_2 \cdots \leq x_k$ and where $+$ indicates a right-censored observation.

   Write down the likelihood function in terms of the underlying lifetime distribution.

   What form will the Kaplan-Meier estimate of the distribution take in this case?

7. Show that when no censoring or truncation occurs, the Kaplan-Meier estimator corresponds to the empirical distribution function.