A.2 Estimation of lifetime distributions

Please hand in your work by Monday 8 February 2008, 4pm, at the Department of Statistics.

1. The survival times (in days after transplant) for the original $n = 69$ members of the Stanford Heart Transplant Program were as follows:

Survival time after heart transplant (days)							
15	3	624	46	127	64	1350	280
23	10	1024	39	730	136	1775	1
836	60	1536	1549	54	47	51	1367
1264	44	994	51	1106	897	253	147
51	875	322	838	65	815	551	66
228	65	660	25	589	592	63	12
499	305	29	456	439	48	297	389
50	339	68	26	30	237	161	14
167	110	13	1	1			

The aim of this exercise is to construct the associated lifetable.

(a) Complete the following table of counts d_x of associated curtate residual lifetimes (in years=365 days), counts ℓ_x of subjects alive exactly x years after their transplant, total time $\tilde{\ell}_x$ spent alive between x and $x+1$ years after their transplant, by all subjects:

x	0	1	2	3	4
d_x			8	4	3
ℓ_x					
$\tilde{\ell}_x$		19.148	10.203	4.937	1.315

(b) Calculate the maximum likelihood estimators $\hat{q}_x^{(0)}$ and $\hat{q}_x$ for q_x, $x = 0, \ldots, 4$, based on the discrete and continuous method, respectively.

(c) Calculate the maximum likelihood estimates. Comment on the differences.

(d) Estimate the probability to survive for 3 months

i. assuming fractional and integer parts of lifetimes are independent, and the fractional part is uniform;

ii. assuming the force of mortality is constant over the first year;

iii. directly from the data (the total time spent alive until three months after the transplant is 12.584 years). Hint: You may, of course, guess formulas to test your intuition, but you should then state your assumptions and apply the discrete and/or continuous method to justify your estimates as maximum likelihood estimates.

2. We have discussed the 22 skeletons of *A. sarcophagus* analysed by Erickson *et al.*. The observed (curtate) ages at death in years were 2,4,6,8,9,11,12,13,14,14,15,15,16, 17,17,18,19,19,20,21,23,28.

(a) Estimate directly the life expectancy of this population.

(b) Estimate a 95% confidence interval for the life expectancy.

(c) We estimated survival probabilities by $\hat{q}^d_x = d_x/\ell_x$. Show that the life expectancy predicted from this estimated distribution must be the same as that computed directly from the observed lifetimes.

(d) Estimate survival probabilities ($\hat{q}^c_x$) for this population, using the continuous method, grouping the lifetime by periods of five years. (The relevant mathematical generalisation of the continuous method is explored in the problem below.) Based on this estimated life-table, estimate the life expectancy for the population. Why is it different from the life expectancy estimated above?

(e) As we explained in the lecture notes, it is reasonable to add 1/2 year to the life expectancy estimated from averaging curtate lifetime observations to estimate the full life expectancy $\mathring{e}_x$, on the assumption that individuals who died between age x and $x+1$ probably lived on average an extra half year. We have estimated a hazard rate ("force of mortality") of 0.333 for ages 20 and above. Supposing this is true, what is the true expected length of life of an individual whose curtate lifetime is reported as 25 years? What does this suggest about the validity of the $+\frac{1}{2}$ rule when the mortality is high.

(f) Suppose a population has Gompertz hazard rate given by $h(x) = Be^{\theta x}$ at age x, for $x \geq 0$, where B and θ are assumed nonnegative. We observe n individuals, with deaths at ages $x_1, \ldots, x_n$. Define $Q(\theta) := \frac{1}{n}\sum e^{\theta x_i}$, $\bar{x} := \frac{1}{n}\sum x_i$. The equation

$$\frac{Q'(\hat{\theta})}{Q(\hat{\theta})-1} - \frac{1}{\hat{\theta}} = \bar{x}$$

has a unique solution (optional extra: Find conditions under which such a solution must exist). Show that $\hat{\theta}$ and $\hat{B} := \hat{\theta}/(Q(\hat{\theta})-1)$ are the maximum-likelihood estimator for (θ, B). Compute the maximum likelihood estimate for fitting Gompertz parameters to the dinosaur population. Use the asymptotic theory to compute a 95% confidence interval for B, assuming the Gompertz model. (As an extra optional challenge: Use the bivariate distribution to compute a 95% confidence interval for the mortality rate at age 20.)

(g) The exponential integral functions are defined by

$$E_n(z) := \int_z^\infty t^{-n}e^{-t}dt.$$

Express the life expectancy at age x of this population, in terms of the function E_1. Using either a table or computer software, estimate the life expectancy for the dinosaur population from the Gompertz model. (For instance, `Maple` uses the command `Ei`.)

3. In a certain population, the force of mortality of lifetimes T is believed to be constant over ages $x_{j-1} \leq x < x_j$, $j \geq 1$, where $x_0 = 0$. Denote these unknown constants by γ_j, $j \geq 1$. You observe n full lifetimes $T^{(1)}, \ldots, T^{(n)} \sim T$ sampled from this population.

   (a) Determine the likelihood function of the sample, in terms of the parameters γ_j, $j \geq 1$.

   (b) Let L_j be the total time spent alive between ages x_{j-1} and x_j. Express L_j explicitly in terms of $T^{(1)}, \ldots, T^{(n)}$.

(c) Show that a maximum likelihood estimator for $\gamma_j \in [0, \infty]$, $j \geq 1$, is given by

$$\hat{\gamma}_j = \begin{cases} \frac{D_j}{L_j} & \text{if } L_j > 0, \\ \infty & \text{otherwise,} \end{cases}$$

where D_j is the number of deaths between ages x_{j-1} and x_j.

(d) Determine the maximal age under the lifetime distribution determined by such a maximum likelihood estimate. Discuss briefly possible modifications.

(e) Denote $h_j = \mathbb{P}(T \leq x_j | T > x_{j-1})$, $j \geq 1$. Express h_j, $j \geq 1$, in terms of γ_j, $j \geq 1$ and deduce maximum likelihood estimators for the new parameters.

(f) Discretize $K = \sup\{x_j : j \geq 0, x_j \leq T\}$ and express the probability mass function p_K of K in terms of h_j, $j \geq 1$.

(g) Derive maximum likelihood estimators for h_j based on the discrete likelihood function.