A.1 Revision, lifetime distributions, Lexis diagrams and the census approximation

Questions 1–3 are to be done for discussion in class. Questions 4–6 are to be handed in for marking.

1. (a) Let $L_1, \dots, L_n$ be independent $\mathrm{Exp}(\lambda)$ random variables. Show that the maximum likelihood estimator for λ is given by

$$\hat{\lambda} = \frac{n}{L_1 + \ldots + L_n}. \tag{A.1}$$

(b) The following data resulted from a life test of refrigerator motors (hours to burnout):

Hours to burnout				
104.3	158.7	193.7	201.3	206.2
227.8	249.1	307.8	311.5	329.6
358.5	364.3	370.4	380.5	394.6
426.2	434.1	552.6	594.0	691.5

i. Assuming refrigerator motors have $\mathrm{Exp}(\lambda)$ lifetimes, give the maximum likelihood estimate for λ.

ii. Still assuming $\mathrm{Exp}(\lambda)$ lifetimes, calculate the Fisher information and construct approximate 95% confidence intervals for λ and $1/\lambda$ using the approximate Normal distribution of the maximum likelihood estimator.

iii. Still assuming $\mathrm{Exp}(\lambda)$ lifetimes, show that $2n\lambda/\hat{\lambda} \sim \chi^2_{2n}$. Let a be such that $\mathbb{P}(2n\lambda/\hat{\lambda} \le a) = \alpha/2$ and b such that $\mathbb{P}(2n\lambda/\hat{\lambda} \ge b) = \alpha/2$. Deduce an exact 95% confidence interval for $1/\lambda$.

iv. Produce a histogram of the data and comment.

v. Merge columns of your histogram appropriately to test whether the hypothesis of $\mathrm{Exp}(\lambda)$ lifetimes can be rejected. Use a χ^2 goodness of fit test.

2. (a) Let $T_1, \dots, T_m$ be independent continuous nonnegative random variables with hazard functions $h_1(\cdot), \dots, h_m(\cdot)$. Let $T = \min(T_1, \dots, T_m)$ and $K = \arg\min(k \mapsto T_k)$. (That is, K is the index of the smallest time, the value of k such that $T = T_k$.) Show that

i. T has hazard function $h_1(\cdot) + \ldots + h_m(\cdot)$;

ii. Conditioned on $\{T = t\}$ K has probability mass function

$$p_k = \frac{h_k(t)}{h_1(t) + \cdots + h_m(t)};$$

iii. If the T_k are exponentially distributed then T and K are independent.

(b) Let $T_1, \dots T_m$ be independent random variables with Weibull distributions with rate parameters $k_1, \dots k_m$ and common exponent n. Show that $T = \min(T_1, \dots, T_m)$ also has a Weibull distribution with exponent n.

(c) Calculate the hazard function of the truncated exponential distribution with maximal age ω, and calculate the limit (in distribution) as $\lambda \downarrow 0$.

3. Let λ be any positive function and $\Lambda(t) = \int_0^t \lambda(s)ds$. Suppose X is a random variable with exponential distribution with parameter 1. Show that $\Lambda^{-1}(X)$ is a random variable with hazard rate $\lambda(t)$.

4. 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 life table.

(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, $x = 0, \dots, 4$, 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 12.58 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.

5. Review the material in on the *census approximation* (section 2.10.2). For purposes of your sketches, you may assume the "census date" is the start of the year (1 January). Suppose we have census counts from the years $K, K+1, \ldots, K+N$.

   (a) Denote by $P_{k,t}$ the number of lives under observation, aged k (last birthday), at any time t.

      i. On a Lexis diagram, sketch the region where you would find the deaths of individuals who died in year t aged k years old; and the region where you would find the deaths of individuals who died in year $t+1$ aged $k+1$ years old. Also sketch sample lifelines for such people, indicating when these individuals might have been born.

      ii. Given n individuals at risk and aged k between times a_i and b_i, show that the total time at risk is

$$E_k^c = \sum_{i=1}^{n}(b_i - a_i) = \int_K^{K+N} P_{x,t}dt.$$

      iii. Assume that $P_{k,t}$ is linear between census dates $t = K, K+1, \ldots, K+N$. Calculate E_k^c in terms of $P_{x,t}$, $t = K, K+1, \ldots, K+N$. Explain why the assumption cannot hold exactly.

   (b) Depending on the records available, you may not know the exact age of individuals at death. You may only know the calendar year of birth and the calendar year of death. Thus, instead of $d_k^{(1)}$ = # deaths aged k at last birthday before death, you will have $d_k^{(2)}$ = # deaths in calendar year of the k-th birthday.

      i. On a Lexis diagram, sketch the region where you would find the deaths of individuals who died in year t whose k-th birthday was in the same year; and the region where you would find the deaths of individuals who died in year $t+1$ with $k+1$-th birthday was in the same year. Also sketch sample lifelines for such people, indicating when they might have been born.

      ii. Describe the resulting estimate of the force of mortality. Explain the definition that you need to use for $P_{k,t}$. State any further assumptions you make.

      iii. What function of μ_t is being estimated? Assuming mortality rates are changing with age, explain why this calculation is estimating something slightly different than the calculation in the previous part.

6. Below is an excerpt from a cohort life table for men in England and Wales born in 1894, including curtate life expectancies. (Data from the Human Mortality Database.) Using the given data:

   (a) Estimate the change to e_0, the curtate life expectancy at birth, if the mortality rate in the first two years of life were reduced to modern-day levels (say $q_0 = 0.005$, $q_1 = 0.0004$).

   (b) Make a rough estimate of the change to e_0 if the increases in mortality due to the 1914-18 war and the 1918-19 influenza pandemic had not occurred.

Age x	l_x	q_x	e_x
0	100000	0.16134	44.82
1	83866	0.05398	52.39
⋮	⋮	⋮	⋮
14	74067	0.00220	45.99
15	73904	0.00237	45.09
16	73729	0.00260	44.20
17	73538	0.00301	43.31
18	73316	0.00313	42.44
19	73087	0.00787	41.57
20	72512	0.01836	40.90
21	71181	0.03218	40.65
22	68890	0.04424	40.98
23	65842	0.06194	41.86
24	61764	0.02088	43.59
25	60474	0.00551	43.51
26	60141	0.00385	42.75
27	59910	0.00384	41.91
28	59680	0.00391	41.07
29	59446	0.00377	40.23
30	59222	0.00386	39.38
31	58994	0.00367	38.53
32	58777	0.00380	37.67
33	58554	0.00399	36.81
34	58320	0.00445	35.96
35	58061	0.00460	35.11