

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

1. (a) The log likelihood function is given by

$$\ell(\lambda) = \log \left( \prod_{k=1}^n (\lambda e^{-\lambda L_k}) \right) = n \log \lambda - \lambda \sum_{k=1}^n L_k$$

We differentiate w.r.t.  $\lambda$  to get

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum_{k=1}^n L_k, \quad \ell''(\lambda) = -\frac{n}{\lambda^2} < 0.$$

$\ell'$  has its unique zero for

$$\lambda = \hat{\lambda} = \frac{n}{L_1 + \dots + L_n}$$

and this is a maximum of  $\ell$ , since  $\ell'' < 0$ . Therefore  $\hat{\lambda}$  maximizes  $\ell$ .

- (b) i. Just apply (a) to the data and get

$$\hat{\lambda} = \frac{20}{l_1 + \dots + l_{20}} = 0.002917.$$

- ii. The Fisher information is given by

$$I_n(\lambda) = -\mathbb{E}(\ell''(\lambda)) = \frac{n}{\lambda^2} \quad (\text{B.1})$$

so that, approximately  $\hat{\lambda} \sim \mathcal{N}(\lambda, \lambda^2/n) \approx \mathcal{N}(\lambda, \hat{\lambda}^2/n)$ , therefore

$$\begin{aligned} .95 &= \mathbb{P}(|Z| < 1.96) \approx \mathbb{P}\left(\left|\frac{\hat{\lambda} - \lambda}{\hat{\lambda}/\sqrt{n}}\right| < 1.96\right) \\ &= \mathbb{P}(\hat{\lambda} - 1.96\hat{\lambda}/\sqrt{n} < \lambda < \hat{\lambda} + 1.96\hat{\lambda}/\sqrt{n}) \\ &= \mathbb{P}\left(\frac{1}{\hat{\lambda} + 1.96\hat{\lambda}/\sqrt{n}} < \frac{1}{\lambda} < \frac{1}{\hat{\lambda} - 1.96\hat{\lambda}/\sqrt{n}}\right), \end{aligned}$$

so that  $(\hat{\lambda} - 1.96\hat{\lambda}/\sqrt{n}, \hat{\lambda} + 1.96\hat{\lambda}/\sqrt{n}) = (0.001638, 0.004195)$  is an approximate 95% confidence interval for  $\lambda$ , and  $(238.4, 610.3)$  an approximate 95% confidence interval for  $1/\lambda$ .

- iii. Since  $L_1 + \dots + L_n \sim \Gamma(n, \lambda)$ , we have  $2\lambda(L_1 + \dots + L_n) \sim \Gamma(n, \frac{1}{2}) = \chi_{2n}^2$  and so for  $2n = 40$ , and the lower and upper 2.5% quantiles

$$\begin{aligned} 0.95 &= \mathbb{P}(24.43 < |X^2| < 59.34) = \mathbb{P}(24.43 < 2\lambda n/\hat{\lambda} < 59.34) \\ &= \mathbb{P}\left(\frac{2n}{59.34\hat{\lambda}} < \frac{1}{\lambda} < \frac{2n}{24.43\hat{\lambda}}\right), \end{aligned}$$

so the exact 95% confidence interval for  $\frac{1}{\lambda}$  is  $(231.1, 561.3)$ .

- iv. We count  $(0, 3, 4, 8, 2, 2, 1)$  in  $(100k, 100(k+1))$ ,  $k = 0, \dots, 6$ . Histogram;-)  
v. Expected numbers under  $\text{Exp}(\hat{\lambda})$  are  $(e^{-100k} - e^{-100(k+1)})n$ , i.e.

$$(5.1, 3.8, 2.8, 2.1, 1.6, 1.2, 0.9) \quad \text{and } 2.6 \text{ for } > 700.$$

For the  $\chi^2$  test we require expected numbers above 5, so we keep the first bin, merge the next three to get 8.7 and the remainder to get 6.2 (alternatively merge next two and remainder). The data then is

Bin	0-100	100-400	400-∞	total
observed	0	15	5	20
expected	5.1	8.7	6.2	20

and we calculate the  $\chi^2_{3-2} = \chi^2_1$  test statistic

$$\sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i} = 9.84 \Rightarrow |Z| = \sqrt{\sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i}} = 3.14 \gg 1.96$$

so there is strong evidence against exponentiality.

2. (a) Identify the survival function of  $T$  as

$$\begin{aligned} \mathbb{P}(T > t) &= \mathbb{P}(T_1 > t, \dots, T_m > t) = \mathbb{P}(T_1 > t) \dots \mathbb{P}(T_m > t) \\ &= \exp\left\{-\int_0^t h_1(s) ds\right\} \dots \exp\left\{-\int_0^t h_m(s) ds\right\} \\ &= \exp\left\{-\int_0^t (h_1(s) + \dots + h_m(s)) ds\right\}. \end{aligned}$$

- (b) By (a), the hazard function of  $T$  now is  $k_1 t^n + \dots + k_m t^n$ , so  $T$  has a Weibull distribution with rate parameter  $k = k_1 + \dots + k_m$  and exponent  $n$ .  
 (c) We first calculate the survival function and let  $\lambda \rightarrow 0$  to get

$$\bar{F}(t) = \mathbb{P}(T > t | T \leq \omega) = \frac{e^{-\lambda t} - e^{-\lambda \omega}}{1 - e^{-\lambda \omega}} \rightarrow \frac{\omega - t}{\omega},$$

which is the survival function of the uniform distribution on  $[0, \omega]$ . This is not surprising since the exponential density for small  $\lambda$  is very flat initially, also after truncation and renormalisation.

We calculate the hazard function of the truncated exponential distribution via the density

$$f(t) = \frac{\lambda e^{-\lambda t}}{1 - e^{-\lambda \omega}} \Rightarrow h(t) = \frac{\lambda}{1 - e^{-\lambda(\omega-t)}}. \tag{B.2}$$

3. We have  $X \sim \exp(1)$ . We want the distribution of  $Y = \Lambda^{-1}(X)$ .

If we let  $F_X$  and  $F_Y$  be the corresponding cdfs, we have  $F_X(x) = \mathbb{P}\{X \leq x\} = 1 - e^{-x}$ , so

$$\begin{aligned} F_Y(y) &= \mathbb{P}\{Y \leq y\} \\ &= \mathbb{P}\{\Lambda^{-1}(X) \leq y\} \\ &= \mathbb{P}\{X \leq \Lambda(y)\} \text{ because } \Lambda \text{ is strictly increasing} \\ &= 1 - e^{-\Lambda(y)}, \end{aligned}$$

which is the cdf of a random variable with hazard rate  $\lambda$ .

4. (a) The full table of deaths  $d_x$ , lives at risk  $\ell_x$  and total time at risk  $\tilde{\ell}_x$  aged  $x$  is

$x$	0	1	2	3	4
$d_x$	45	9	8	4	3
$\ell_x$	69	24	15	7	3
$\tilde{\ell}_x$	35.63	19.15	10.20	4.94	1.32

- (b) The discrete method based on curtate lifetimes  $K^{(1)}, \dots, K^{(n)}$ ,  $n = 69$ , factorises the likelihood

$$\prod_{j=1}^{69} p_K(K^{(j)}) = \prod_{x=0}^{\infty} (1 - q_x)^{\ell_x - d_x} q_x^{d_x} \tag{B.3}$$

and differentiation of each factor leads to maximum likelihood estimators  $\hat{q}_x^{(0)} = d_x/\ell_x$ . The continuous method based on  $T^{(1)}, \dots, T^{(n)}$ ,  $n = 69$ , and the assumption of constant forces of mortality between integer ages, factorises the likelihood

$$\prod_{j=1}^{69} f_T(T^{(j)}) = \prod_{x=0}^{\infty} \mu_{x+\frac{1}{2}}^{d_x} \exp\left\{-\tilde{\ell}_x \mu_{x+\frac{1}{2}}\right\} \tag{B.4}$$

and differentiation of each factor leads to maximum likelihood estimators  $\hat{q}_x = 1 - \exp\{-d_x/\tilde{\ell}_x\}$ .

- (c) From the formulas obtained in (b) we calculate

$x$	0	1	2	3	4
$\hat{q}_x^{(0)}$	0.65	0.38	0.53	0.57	1
$\hat{q}_x$	0.717	0.375	0.543	0.555	0.898

$\hat{q}_0^{(0)} < \hat{q}_0$  since the total time  $\tilde{\ell}_0$  spent at risk is very short. We can see this directly from the data. Most subject dying in the first year die very early (e.g. three subjects die the day after their transplant). This actually suggests that the force of mortality is not constant over the first year, but much higher initially.

$\hat{q}_4 < \hat{q}_4^{(0)} = 1$  allows survival beyond the maximal observed age under the continuous method. The specification of the distribution estimate is not complete, but with no data we get no estimate. Some methods of graduation will allow to extrapolate beyond maximal age.

By both methods, the one-year survival probabilities indicate a bathtub behaviour, decreasing initially and then increasing.

- (d) i. Under the estimates from curtate lifetimes and the assumption of independent uniform fractional part,

$$\begin{aligned} \mathbb{P}(T > 0.25) &= \mathbb{P}(T > 1) + \mathbb{P}(K = 0, S > 0.25) \\ &\text{is estimated by } (1 - \hat{q}_0^{(0)}) + \hat{q}_0^{(0)} \frac{3}{4} = 0.837. \end{aligned}$$

- ii. Under the estimates from continuous lifetimes and the assumption of constant force of mortality between integer ages

$$\begin{aligned} \mathbb{P}(T > 0.25) &= \exp\left\{-\int_0^{0.25} \mu_t dt\right\} \\ &\text{is estimated by } \exp\{-0.25\hat{\mu}_{0+\frac{1}{2}}\} = (1 - \hat{q}_0)^{0.25} = 0.729. \end{aligned}$$

- iii. Again we can apply the discrete or continuous method (formally for units of three months). The continuous method assumes constancy of forces of mortality over each three-month period and gives an estimate

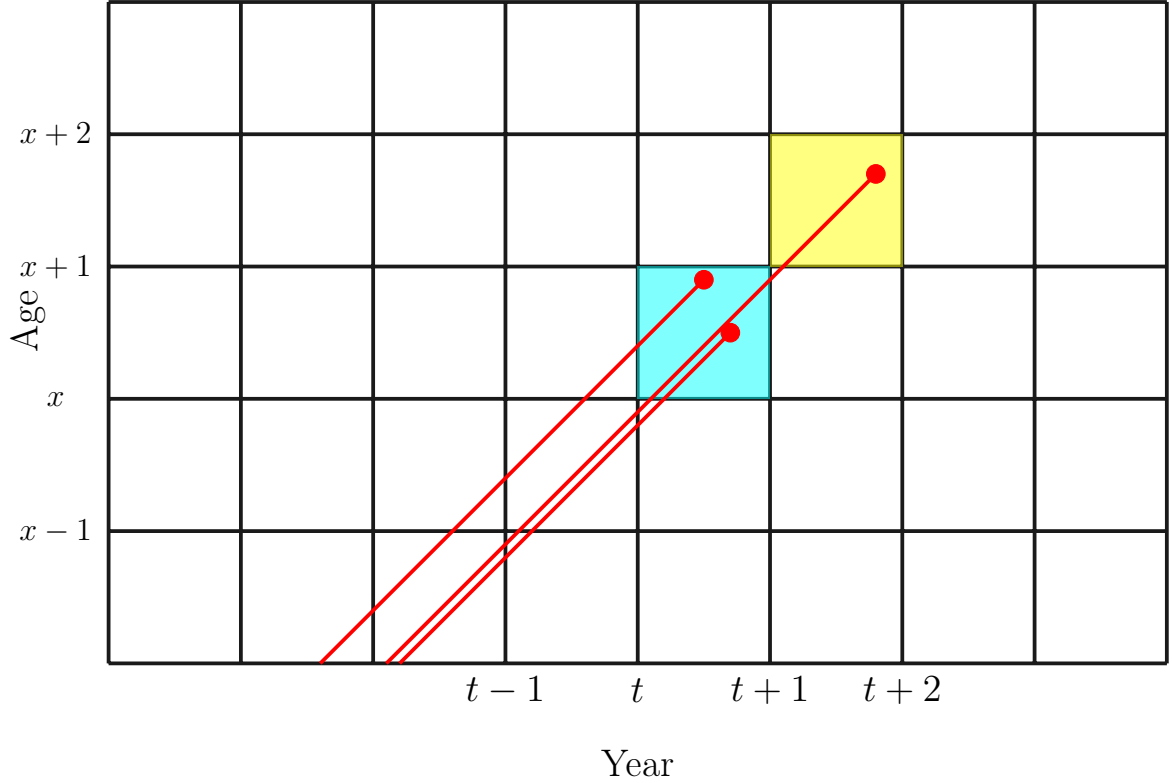
$$\exp\left\{-\frac{d}{4\tilde{\ell}}\right\} = \exp\left\{-\frac{31}{4 \times 12.584}\right\} = 0.540.$$

Here,  $4\tilde{\ell}$  is the total number of time units (as calculated from years  $\tilde{\ell}$ ) at risk during first three-month unit.

The discrete method is based on one-unit death probabilities and gives  $1 - d/\ell = 1 - 31/69 = 0.551$  as an estimate for the first-unit survival probability.

These estimates are much smaller reflecting a higher risk to die initially. In fact, this suggests that neither assumption i. nor ii. is optimal. An initially decreasing force of mortality would be better.

5. (a) i. The turquoise region corresponds to age  $x$  in year  $t$ . The same individuals are age  $x + 1$  in year  $t + 1$ , and this portion of their lifelines falls in the yellow region.



- ii. Just write the quantities as integrals and sums and interchange the order of integration and summation:

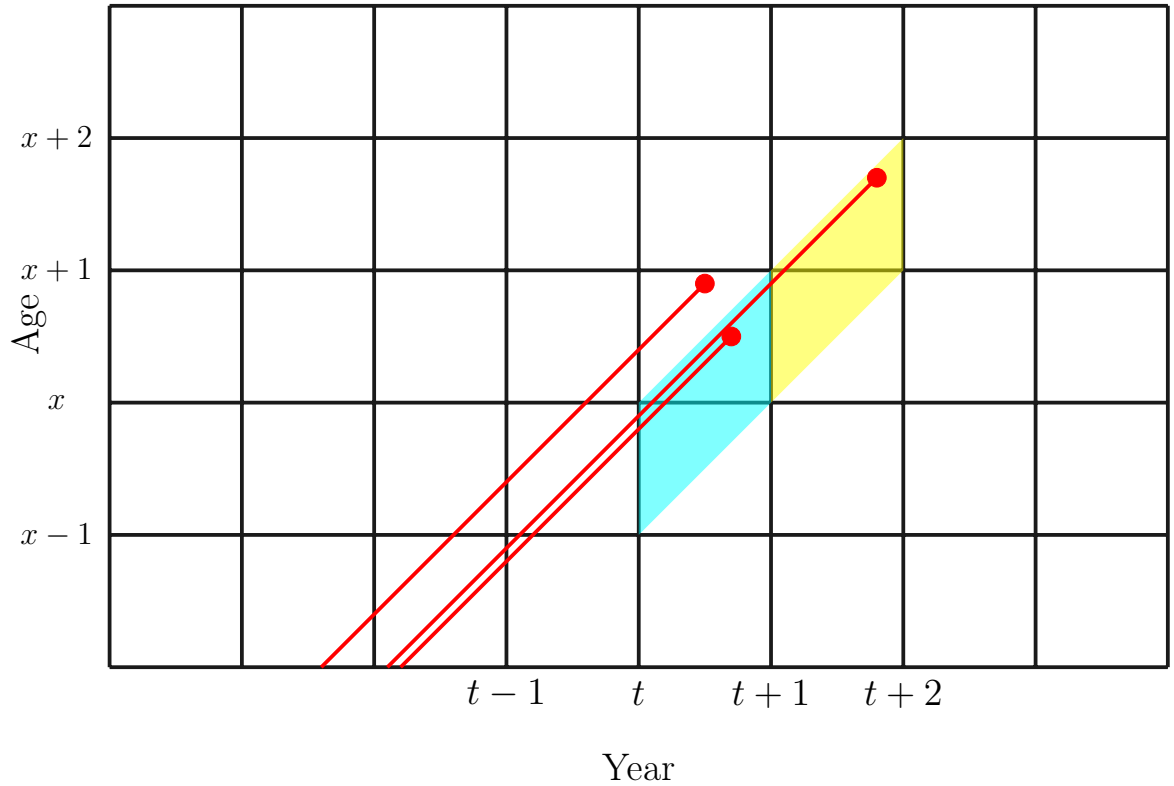
$$E_x^c = \sum_{i=1}^n \int_{a_i}^{b_i} dt = \int_K^{K+N+1} \sum_{i=1}^n 1_{\{a_i \leq t \leq b_i\}} dt = \int_K^{K+N+1} P_{x,t} dt. \quad (\text{B.5})$$

- iii. Under the assumption of piecewise linearity we calculate

$$E_x^c = \sum_{k=K}^{K+N} \int_0^1 (rP_{x,k} + (1-r)P_{x,k+1}) dr = \sum_{k=K}^{K+N} \frac{P_{x,k} + P_{x,k+1}}{2}. \quad (\text{B.6})$$

The assumption of piecewise linear  $P_{x,t}$  cannot hold exactly since  $P_{x,t} \in \mathbb{N}$ , but for large  $n$  this is negligible.

- (b) i. The turquoise region corresponds to age  $x$  in year  $t$ . The same individuals are age  $x + 1$  in year  $t + 1$ , and this portion of their lifelines falls in the yellow region.



ii. We can define

$$P_{x,t}^{(2)} = \# \text{ lives at risk at time } t \text{ with } x\text{th birthday in calendar year } [t], \quad (\text{B.7})$$

but also ought to adjust

$$E_x^{c,2} = \int_K^{K+N+1} P_{x,t}^{(2)} dt \approx \sum_{k=K}^{K+N} \frac{P_{x,k} + P_{x+1,k+1}}{2} \quad (\text{B.8})$$

since it is more natural to assume that the cohort of lives  $P_{x,k}$  with  $x$ -th birthday in calendar year  $k$  changes linearly to  $P_{x+1,k+1}$  since this will count the same people (being age  $k$  at the start of year  $x$  and age  $k+1$  at the start of year  $x+1$ ).

iii. The first estimate is approximating  $\int_x^{x+1} \mu_s ds$ ; that is, the average of  $\mu_s$  over the age interval from  $x$  to  $x+1$ , assuming age-specific mortality doesn't change with time. (This is just  $\mu_x$  if  $\mu_s$  is assumed constant on  $[x, x+1]$ .) The second estimate clearly includes the experience of individuals at ages between  $x-1$  and  $x+1$ , but weighted toward the middle (in proportion to the width of the parallelogram in the above figure.) In fact, the estimate  $\tilde{\mu}_x = d_x^{(2)}/E_x^{c,2}$  will be approximating

$$\int_{x-1}^{x+1} (1 - |s - x|) \mu_s ds.$$