B.2 Estimation of lifetime distributions

1. (a) The full table of deaths d_x, lives at risk ℓ_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
ℓ_x	69	24	15	7	3
$\tilde{\ell}_x$	35.627	19.148	10.203	4.937	1.315

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

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

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

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

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
$\hat{q}_x^{(0)}$	0.652	0.375	0.533	0.571	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 actualy 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,

$$\mathbb{P}(T > 0.25) = \mathbb{P}(T > 1) + \mathbb{P}(K = 0, S > 0.25)$$

$$\text{is estimated by} \qquad (1 - \hat{q}_0^{(0)}) + \hat{q}_0^{(0)} \frac{3}{4} = 0.837.$$

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

$$\mathbb{P}(T > 0.25) = \exp\left\{-\int_0^{0.25} \mu_t dt\right\}$$

$$\text{is estimated by} \qquad \exp\{-0.25\hat{\mu}_{0+\frac{1}{2}}\} = (1 - \hat{q}_0)^{0.25} = 0.729.$$

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 for i. nor for ii. are appropriate. One should rather model a decreasing force of mortality, initially.

2. We call the random variable of a random dinosaur lifetime X, and the observations $x_1, \ldots, x_{22}$. We do the estimates two different ways. In black we do the estimates purely on curtate lifespans, in red we include estimates for the fractional part of lifespan, which we assume to be uniform on $[0, 1]$ and independent of the integer part.

   (a) The life expectancy is $\mathbb{E}[X]$, which we estimate by $\frac{1}{22}\sum x_i = 14.6$ years. The correction is $\frac{1}{2}$, so we estimate the total life expectancy to be 15.1 years.

   (b) The observed curtate lifespans have mean $\bar{x} = 14.6$ years and SD

   $$\hat{\sigma} = \sqrt{\frac{1}{n-1}\left(\sum (x_i - \bar{x})^2\right)} = 6.28 \text{ years}.$$

   Using Student's T distribution (with 21 degrees of freedom) to estimate the distribution of $(\bar{x} - \mathbb{E}[X])/\hat{\sigma}$, we estimate a 95% confidence interval for $\mathbb{E}[X]$ to be

   $$\bar{x} \pm t_{21}(.975) \cdot \frac{\hat{\sigma}}{\sqrt{n}} = 14.6 \pm 2.8 \text{ years} = (11.8, 17.4) \text{ years}.$$

   where $t_{21}(.975) = 2.08$ is the quantile function of a Student T random variable with 21 degrees of freedom; that is, $t_d(\alpha)$ satisfies $P\{T \le t_d(\alpha)\} = \alpha$ when T is a Student T random variable with d degrees of freedom.
   We impute a true average observed lifespan of $\bar{x}^* = 15.1$ years; if we assume in addition that the fractional part is independent of the integer part, the variance of the fractional part is 1/12, which must be added to the variance of the curtate lifespan, yielding $\hat{\sigma}^* = \sqrt{\hat{\sigma}^2 + \frac{1}{12}}$, which differs from $\hat{\sigma}$ only in the fourth decimal place. This results in a confidence interval of $(12.3, 17.9)$ years.

   (c) The formula for life expectancy (ignoring the correction for partial years) is

   $$e_0 = \sum_{x=1}^{\infty} \mathbb{P}\{\text{reach age } x\} = \sum_{x=1}^{\infty} (1-q_0)(1-q_1)\cdots(1-q_{x-1}).$$

   Since $1 - \hat{q}_x^d = (\ell_x - d_x)/\ell_x = \ell_{x+1}/\ell_x$, and $\ell_0 = n$, the size of the starting population, we see that

   $$e_0 = \frac{1}{n}\sum_{x=1}^{\infty} \ell_x = \frac{1}{n}\sum_{x=1}^{\infty} \#\{x_i \ge x\} = \frac{1}{n}\sum_{i=1}^{n} T_i.$$

   (d) Counting only the curtate lifespans, we observe in age range 0–4 there are 20 who survive the full 5 years, one who survives 2 years, and one who survives 4 years, yielding a total of 106 lived years. Continuing through all the age ranges:

age	$\tilde{\ell}_x$	d_x	$\hat{\mu}_x$	$\hat{q}^c_x = 1 - e^{-\hat{\mu}_x}$
0–4	106	2	.019	.019
5–9	93	3	.032	.032
10–14	74	5	.068	.065
15–19	36	8	.22	.20
20–24	9	3	.33	.28
25–29	3	1	.33	.28

The life expectancy of this life table is given by

$$e_0 = \int_0^\omega \exp\left\{-\int_0^x \mu_s ds\right\} dx,$$

where μ_s is the step function of estimated mortality rates, and ω is the maximum age, which we may take to be 29 or 30, or ∞ if we choose to model mortality as constant forever after age 25. If you have constantant mortality μ_i over time intervals $[t_{i-1}, t_i)$, taking $t_0 = 0$, this comes out to

$$e_0 = \frac{1}{\mu_1}\left(1 - e^{-t_1\mu_1}\right) + \frac{1}{\mu_2} e^{-t_1\mu_1}\left(1 - e^{-(t_2-t_1)\mu_2}\right) + \cdots + \frac{1}{\mu_k} e^{-t_1\mu_1 - \cdots - (t_{k-1}-t_{k-1})\mu_{k-1}}\left(1 - e^{-(t_k - t_{k-1})\mu_k}\right).$$

Plugging in the above estimates of μ, we get $e_0 = 14.5$ years (regardless of which endpoint we choose, since the difference between an endpoint of 30 and an endpoint of ∞ is only 0.04 years in life expectancy).

The difference from the result of part (b) comes from the constraint that we have forced upon the data, that mortality rates are constant over periods of 5 years. That is, we have not directly averaged the data, but found the average that is closest to fitting this assumption.

We may estimate life expectancy as

$$\hat{e}_0 = \frac{1}{2} + \sum_{x=1}^{\infty} (1 - \hat{q}^c_0) \cdots (1 - \hat{q}^c_{x-1}) = 14.5 \text{ years}.$$

A better estimate comes from including half of the integer year in which the individual died in the count of years at risk, we replace the estimate of total years lived by $\tilde{\ell}^*_x = \tilde{\ell}_x + d_x/2$. We end up with an alternative estimate $\hat{e}^*_0 = 14.8$ years.

age	$\tilde{\ell}_x$	d_x	$\hat{\mu}_x$	$\hat{q}^c_x = 1 - e^{-\hat{\mu}_x}$
0–4	107	2	.019	.019
5–9	94.5	3	.032	.031
10–14	76.5	5	.065	.063
15–19	40	8	.20	.18
20–24	10.5	3	.29	.25
25–29	3.5	1	.29	.25

(e) The lifetime is reported at 25 years. That is to say, we know the lifetime X is between 25 and 26 years. Thus $Z := X - 25$ has the distribution of an exponential random variable conditioned on $X < 1$. If the rate is μ, this yields

$$\mathbb{E}[Z \,|\, Z < 1] = \frac{\int_0^1 \mu e^{-\mu z} z dz}{\int_0^1 \mu e^{-\mu z} dz} = \frac{e^\mu - 1 - \mu}{\mu(e^\mu - 1)} \approx \frac{1}{2} - \frac{\mu}{12}.$$

When $\mu = 0.333$ this evaluates to 0.472 (and the linear approximation is accurate to 5 decimal places), giving us an average time of death of 25.472 years. One lesson is that the approximation that adds 1/2 year to each curtate lifetime is fairly close, even when the mortality rates are quite high.

(f) The log likelihood is

$$\ell(B,\theta) = \sum_i [\ln h(x_i) - H(x_i)] = n\ln B + \sum_i \left[\theta x_i - \frac{B}{\theta}\left(e^{\theta x_i} - 1\right)\right].$$

The maximum of ℓ could come either at an interior point or at a boundary point. The likelihood goes to 0 as B goes to 0, so the maximum on the boundary is attained along $\theta = 0$, and at an interior point in B.

The partial derivatives are

$$\frac{\partial \ell}{\partial B} = \frac{n}{B} - \sum_i \left[\frac{1}{\theta}\left(e^{\theta x_i} - 1\right)\right] = -\frac{n}{B} + \frac{n}{\theta}\left(Q(\theta) - 1\right)$$
$$\frac{\partial \ell}{\partial \theta} = -n\|x\|_1 + nB\frac{\theta Q'(\theta) - Q(\theta) + 1}{\theta^2}.$$

For any given choice of $\hat\theta$, the maximum likelihood estimate of B must be $\hat B = \hat\theta/(Q(\hat\theta) - 1)$. Define

$$\ell_*(\theta) := \ell(\theta/(Q(\theta)-1), \theta) = -\log\frac{Q(\theta)-1}{\theta} + \|x\|_1\theta - 1.$$

The maximum-likelihood estimate is $(\hat\theta/(Q(\hat\theta)-1), \hat\theta)$, where $\hat\theta$ maximises ℓ_*. We note that

$$\ell_*'(\theta) = -\frac{Q'(\theta)}{Q(\theta)-1} + \frac{1}{\theta} + \|x\|_1,$$
$$\ell_*''(\theta) = \frac{Q'(\theta)^2 - Q''(\theta)Q(\theta) + Q''(\theta)}{(Q(\theta)-1)^2} - \theta^{-2}$$

Critical points for ℓ_* will thus be solutions to

$$\frac{Q'(\theta)}{Q(\theta)-1} - \frac{1}{\theta} = \|x\|_1 \qquad (1)$$

The optional question asked why there should be a unique maximum which is also a critical point. Define $\|x\|_k = (x_1^k + \cdots + x_n^k)^{1/k}$. Then

$$\lim_{\theta\to\infty}\frac{Q'(\theta)}{Q(\theta)-1} = \max x_i \ge \|x\|_1,$$

and by L'Hôpital's rule,

$$\begin{aligned}\lim_{\theta\downarrow 0}\frac{Q'(\theta)}{Q(\theta)-1} - \frac{1}{\theta} &= \lim_{\theta\downarrow 0}\frac{\theta Q'(\theta) - Q(\theta) + 1}{\theta(Q(\theta)-1)}\\ &= \lim_{\theta\downarrow 0}\frac{\theta Q''(\theta)}{Q(\theta) - 1 + \theta Q'(\theta)}\\ &= \frac{Q''(0)}{2Q'(0)}\\ &= \frac{\|x\|_2^2}{2\|x\|_1}.\end{aligned}$$

If this is below $\|x\|_1$ (meaning that $2\|x\|_1^2 > \|x\|_2^2$) then there must be a solution to (1), which will be a solution to $\ell_*'(\hat\theta) = 0$. On the other hand, if $2\|x\|_1^2 \le \|x\|_2^2$ then ℓ_*' starts below 0 and ends below 0. In this case, the maximum may be expected to lie at the boundary point $(0, 1/\|x\|_1)$; the proof that the likelihood does not come back up is not easy, though, and will not be required or given here.

In the former case, the first critical point along the curve $B = \theta/(Q(\theta)-1)$ will be a local maximum, unless it is a saddle point; we do need to check the second derivative to be sure.

Solving these equations numerically, we find the solution $(\hat{\theta}, \hat{B}) = (0.145, .0130)$. The second derivative at this point, in particular, may be computed by $Q(\hat{\theta}) = 12.2$, $Q'(\hat{\theta}) = 241$, $Q''(\hat{\theta}) = 5152$, implying that $\ell^*(\hat{\theta}) = -35.8$. (In fact, the second derivative is always negative, which implies that this is the only critical point. This is not easy to prove, but it can be checked numerically.)
Note: We compute $Q'(\theta) = n^{-1}\sum x_i e^{\theta x_i}$, and $Q''(\theta) = n^{-1}\sum x_i^2 e^{\theta x_i}$.
We have

$$\mathcal{I}(\theta, B) = \begin{pmatrix} B(2\theta^{-3}q_0(\theta) - \theta^{-2}q_1(\theta) + \theta^{-1}q_2(\theta)) & \theta^{-1}q_1(\theta) - \theta^{-2}q_0(\theta) \\ \theta^{-1}q_1(\theta) - \theta^{-2}q_0(\theta) & B^{-2} \end{pmatrix}$$

As usual, we estimate the information at the true parameter by substituting the estimated parameters, yielding

$$\mathcal{I}(\hat{\theta}, \hat{B}) \approx \begin{pmatrix} 408 & 1068 \\ 1068 & 5917 \end{pmatrix}$$

It follows that $(\hat{\theta} - \theta, \hat{B} - B)$ is approximately normal with covariance matrix

$$n^{-1}\mathcal{I}^{-1} \approx \begin{pmatrix} 2.1 \times 10^{-4} & -3.9 \times 10^{-5} \\ -3.9 \times 10^{-5} & 1.5 \times 10^{-5} \end{pmatrix}.$$

Thus a 95% confidence interval for B is about $0.0130 \pm 1.96 \cdot \sqrt{1.5 \times 10^{-5}} = (.0054, .0206)$.
The optional question asks for a 95% confidence interval for $h(20) = Be^{20\theta}$. We treat B and θ as normal random variables, using the asymptotic theory. Note that

$$\mathbb{E}[Q(\theta)] = \mathbb{E}\left[e^{\theta X}\right],$$

where X is chosen from a Gompertz distribution with parameters (B, θ). We define

$$q_k(\theta) := \mathbb{E}[Q^{(k)}(\theta)] = e^{B/\theta}\theta^{1-k}B^{-1}\int_{B/\theta}^{\infty}(\ln u + \ln\theta/B)^k u e^{-u} du.$$

While $q_0(\theta) = 1 + \theta/B$, the others have no simple closed-form solutions. Still, there is no difficulty in computing them. We get $q_0(\hat{\theta}) = 12.2$, $q_1(\hat{\theta}) = 239$, and $q_2(\hat{\theta}) = 5020$.
We can represent $(\hat{\theta} - \theta, \hat{B} - B) = (\alpha Z_1 + \beta Z_2, \gamma Z_1 + \delta Z_2)$, where (Z_1, Z_2) are independent standard normal random variables and

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = (n\mathcal{I})^{-1/2} = \begin{pmatrix} 1.4 \times 10^{-2} & -2.2 \times 10^{-3} \\ -2.2 \times 10^{-5} & 3.2 \times 10^{-3} \end{pmatrix}.$$

(Note: This is the square root of the matrix $(n\mathcal{I})^{-1}$. How do we turn this into an estimate for $h(20)$? We use a Taylor series to approximate

$$\begin{aligned} h(20; \theta, B) - h(20; \hat{\theta}, \hat{B}) &\approx \frac{\partial h(20)}{\partial \theta}(\theta - \hat{\theta}) + \frac{\partial h(20)}{\partial B}(B - \hat{B}) \\ &= 20Be^{20\theta}(\theta - \hat{\theta}) + e^{20\theta}(B - \hat{B}) \\ &\approx 4.7(\theta - \hat{\theta}) + 18(B - \hat{B}) \\ &\approx 0.028Z_1 + .047Z_2, \end{aligned}$$

which is normal with mean 0 and variance $\sigma^2 = .0030$. Thus, we estimate a 95% confidence interval to be

$$h(20; \hat{\theta}, \hat{B}) \pm 1.96\sigma = .23 \pm .11.$$

(g) The survival function for a Gompertz population is

$$\exp\left\{-\frac{B}{\theta}\left(e^{\theta x}-1\right)\right\},$$

so the life expectancy is

$$e_0=\int_0^\infty \exp\left\{-\frac{B}{\theta}\left(e^{\theta x}-1\right)\right\}dx.$$

Substituting $u=\frac{B}{\theta}e^{\theta x}$ yields

$$e_0=\frac{e^{B/\theta}}{\theta}\int_{B/\theta}^\infty e^{-u}\frac{du}{u}=\frac{e^{B/\theta}}{\theta}\,\mathrm{E}_1(B/\theta).$$

Plugging in the estimates yields $\hat{e}_0=14.5$ years.

3. (a) Since $f_T(t)=h_T(t)\exp\{-\int_0^t h_T(s)ds\}$, we get as likelihood function

$$\prod_{i=1}^n f_T(T^{(i)})=\prod_{j=1}^\infty \gamma_j^{D_j}\exp\left\{-\gamma_j L_j\right\}$$

where L_j is the total time spent alive between ages x_{j-1} and x_j, and D_j is the number of deaths between ages x_{j-1} and x_j.

(b) Explicitly, we get

$$L_j=\sum_{i=1}^n\left[(x_j-x_{j-1})1_{\{T^{(i)}\geq x_j\}}+(T^{(i)}-x_{j=1})1_{\{x_j>T^{(i)}\geq x_{j-1}\}}\right].$$

(c) The likelihood function factorises. Therefore, we maximise separately for each $j\geq 1$

$$l_j(\gamma_j)=\log\left(\gamma_j^{D_j}\exp\left\{-\gamma_j L_j\right\}\right)=D_j\log(\gamma_j)-\gamma_j L_j$$

by differentiation, equating to zero get $l_j'(\gamma_j)=D_j/\gamma_j-L_j=0$, i.e. $\hat{\gamma}_j=D_j/L_j$, since also $l_j''(\gamma_j)=-D_j/\gamma_j^2<0$.

If $L_j=0$, then no life was observed as old as x_{j-1}. The likelihood does not depend on γ_j, and putting $\hat{\gamma}_j=\infty$ completes the definition of $\hat{\gamma}_j$ to a maximum likelihood estimator, as any other value would have done.

(d) There isn't exactly a maximum age, but we can't analyse ages past the interval of the maximal observed age. A Gompertz-Makeham hazard, i.e. exponentially increasing, would have been another option, which would then permit extrapolation into the unobserved range. Parameters could also be fitted from a standard lifetable or based on other large samples that the population studied here may be comparable with.

(e) We calculate

$$h_j=\mathbb{P}(T\leq x_j|T>x_{j-1})=1-\exp\left\{-(x_j-x_{j-1})\gamma_j\right\},$$

so that

$$\hat{h}_j=1-\exp\left\{-(x_j-x_{j-1})\hat{\gamma}_j\right\}=\exp\left\{-(x_j-x_{j-1})D_j/L_j\right\}$$

if $L_j>0$, $\hat{h}_j=1$ otherwise.

(f) $\mathbb{P}(K=x_{j-1})=(1-h_1)\times\ldots\times(1-h_{j-1})h_j$.

(g) The discrete likelihood function is

$$\prod_{i=1}^{n} p_K(K^{(i)}) = \prod_{j=1}^{\infty} (1-h_j)^{N_j - D_j} h_j^{D_j},$$

where $N_j = n - D_1 - \ldots - D_{j-1}$ is the number of people alive at x_{j-1}, i.e. at risk for the jth period.

The likelihood function fully factorises, and we differentiate w.r.t. h_j to get, after cancellation of all common terms

$$-(N_j - D_j)h_j + D_j(1-h_j) = 0 \quad \Rightarrow \quad \hat{h}_j^{(0)} = \frac{D_j}{N_j}$$

since also the likelihood is zero at the boundary $h_j = 1$ and/or $h_j = 0$, provided $N_j > 0$. If $N_j = D_j$, then $\hat{h}_j^{(0)} = 1$, so that further specification of $\hat{h}_{j+1}^{(0)}, \ldots$ does not change the distribution specified by the maximum likelihood estimates.