B.1 Revision, lifetime distributions

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

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

We differentiate w.r.t. λ to get

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

ℓ' has its unique zero for

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

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

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

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

ii. The Fisher information is given by

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

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 λ, and $(238.4, 610.3)$ an approximate 95% confidence interval for $1/\lambda$.

iii. Since $L_1 + \ldots + L_n \sim \Gamma(n, \lambda)$, we have $2\lambda(L_1 + \ldots + L_n) \sim \Gamma(n, \frac{1}{2}) = \chi^2_{2n}$ 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 $(100k, 100(k+1))$, $k = 0, \ldots, 6$. Histogram;-)

v. Expected numbers under $\mathrm{Exp}(\hat{\lambda})$ are $(e^{-100k} - e^{-100(k+1)})n$, i.e.

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

For the χ^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

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

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 $\lambda \to 0$ to get

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

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

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

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

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 λ.

4. (a) We focus on continuous M. The discrete case is analogous.

$$\mathbb{E}(T) = \int_0^\infty \mathbb{E}(T | M = \lambda) f_M(\lambda) d\lambda = \int_0^\infty \frac{1}{\lambda} f_M(\lambda) d\lambda = \mathbb{E}\left(\frac{1}{M}\right).$$

Also, since $\mathbb{E}(T^2 | M = \lambda) = \text{Var}(T | M = \lambda) + (\mathbb{E}(T | M = \lambda))^2 = 2\lambda^{-2}$,

$$\mathbb{E}(T^2) = \int_0^\infty \frac{2}{\lambda^2} f_M(\lambda) d\lambda = 2\mathbb{E}\left(\frac{1}{M^2}\right)$$

and hence

$$\mathrm{Var}(T) = \mathbb{E}(T^2) - (\mathbb{E}(T))^2 = 2\mathbb{E}\left(\frac{1}{M^2}\right) - \left(\mathbb{E}\left(\frac{1}{M}\right)\right)^2.$$

Finally, by Tonelli's theorem

$$\bar{F}_T(t) = \int_t^\infty \int_0^\infty \lambda e^{-\lambda s} f_M(\lambda) d\lambda ds = \int_0^\infty e^{-\lambda t} f_M(\lambda) d\lambda = \mathcal{M}_M(-t).$$

(b) We calculate from the definition

$$f_T(t) = \int_0^\infty \lambda e^{-\lambda t} \frac{\nu^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\nu\lambda} d\lambda = \frac{\alpha\nu^\alpha}{(t+\nu)^{\alpha+1}},$$

having identified the Gamma density for parameters $\alpha + 1$ and $t + \nu$ under the integral, and also using $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$.
Now we deduce

$$\bar{F}_T(t) = \int_t^\infty \frac{\alpha\nu^\alpha}{(s+\nu)^{\alpha+1}} = \frac{\nu^\alpha}{(t+\nu)^\alpha}$$

and

$$h_T(t) = \frac{f_T(t)}{\bar{F}_T(t)} = \frac{\alpha}{t+\nu}.$$

This is called the *Pareto distribution.*

5. (a) Given that there are $\ell_x = n$ independent subjects at risk, each one of them will die with probability q_x by the end of the year. Therefore, d_x has a binomial distribution with parameters n and q_x as its conditional distribution, given $\ell_x = n$. Recallthat for a binomial distribution with parameters (n, q), the expectation is nq, the variance $nq(1 - q)$, and $\mathbb{E}[X^2] = nq(1 - q) + (nq)^2$.
Now we condition on ℓ_x to get

$$\begin{aligned}
\mathbb{E}[d_x] &= \sum_{n=0}^{\ell_0} \mathbb{P}(\ell_x = n)\mathbb{E}\left[d_x \,\middle|\, \ell_x = n\right] \\
&= \sum_{n=0}^{\ell_0} \mathbb{P}\big\{\ell_x = n\big\} n q_x \\
&= q_x \mathbb{E}[\ell_x].
\end{aligned}$$

Similarly,

$$\mathrm{Var}\left(d_x - q_x\ell_x\right) = \mathbb{E}\left[(d_x - q_x\ell_x)^2\right] = \mathbb{E}\left[d_x^2\right] - 2q_x\mathbb{E}\left[d_x\ell_x\right] + q_x^2\mathbb{E}\left[\ell_x^2\right]$$

where

$$\begin{aligned}
\mathbb{E}\left[d_x^2\right] &= \sum_{n=0}^{\ell_x} \mathbb{P}\big\{\ell_x = n\big\}\left(nq_x(1-q_x) + n^2q_x^2\right) \\
&= q_x(1-q_x)\mathbb{E}[\ell_x] + q_x^2\mathbb{E}\left[\ell_x^2\right], \\
\mathbb{E}\left[d_x\ell_x\right] &= \sum_{n=0}^{\ell_x} \mathbb{P}\big\{\ell_x = n\big\} n^2 q_x, \\
&= q_x\mathbb{E}\left[\ell_x^2\right]
\end{aligned}$$

and therefore

$$\begin{aligned}
\mathrm{Var}(d_x - q_x\ell_x) &= q_x(1-q_x)\mathbb{E}\left[\ell_x\right] + q_x^2\mathbb{E}\left[\ell_x^2\right] - 2q_x^2\mathbb{E}\left[\ell_x^2\right] + q_x^2\mathbb{E}\left[\ell_x^2\right] \\
&= q_x(1-q_x)\mathbb{E}\left[\ell_x\right].
\end{aligned}$$

(b) Since ℓ_0 is non-random, we obtain

$$\mathbb{E}(\hat{q}_0^{(0)}) = \frac{\mathbb{E}(d_0)}{\ell_0} = q_0,$$

so that $\hat{q}_0^{(0)}$ is unbiased. We also calculate

$$\mathrm{Var}(\hat{q}_0^{(0)}) = \frac{\mathrm{Var}(d_0)}{\ell_0^2} = \frac{q_0(1-q_0)}{\ell_0} \approx \frac{d_0(\ell_0 - d_0)}{\ell_0^3}.$$

Since ℓ_1 is a random variable, the argument breaks down (the expectation of the quotient of two random variables is not the quotient of expectations!). Moreover, we need a convention for the case $\ell_1 = 0$, which happens with positive probability. We might put $\hat{q}_1^{(0)} = 1$ if $\ell_1 = 0$. Then

$$\mathbb{E}\left(\hat{q}_1^{(0)}\right) = \mathbb{P}(\ell_1 = 0) + \sum_{n=1}^{\infty} \mathbb{P}(\ell_1 = n)\mathbb{E}\left(\frac{d_1}{\ell_1}\middle| \ell_1 = n\right) = q_0^{\ell_0} + (1 - q_0^{\ell_0})q_1$$

so that $\hat{q}_1^{(0)}$ is biased (except if indeed $q_1 = 1$).

For the variance, we can write down the analogue to $\mathrm{Var}(\hat{q}_0^{(0)})$:

$$\mathrm{Var}(\hat{q}_1^{(0)}) \approx \frac{d_1(\ell_1 - d_1)}{\ell_1^3}.$$

This can be justified asymptotically, as $\ell_0 \to \infty$, since maximum likelihood estimators are asymptotically Normal, and the Fisher information matrix here, based on the log likelihood

$$\sum_{x \in \mathbb{N}} (\ell_x - d_x)\log(1 - q_x) + d_x \log(q_x)$$

(with zero off-diagonal second derivatives), has diagonal entries

$$I_{xx}(q.) = \mathbb{E}\left(\frac{\ell_x - d_x}{(1-q_x)^2}\right) + \mathbb{E}\left(\frac{d_x}{q_x^2}\right) = \mathbb{E}(\ell_x)\left(\frac{1-q_x}{(1-q_x)^2} + \frac{q_x}{q_x^2}\right) = \frac{\mathbb{E}(\ell_x)}{q_x(1-q_x)},$$

so that

$$\mathrm{Var}(\hat{q}_x^{(0)}) \approx \frac{1}{I_{xx}(q.)} = \frac{q_x(1-q_x)}{\mathbb{E}(\ell_x)} \approx \frac{d_x(\ell_x - d_x)}{\ell_x^3}.$$

6. (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

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

and differentiation of each factor leads to maximum likelihood estimators $\hat{q}_x^{(0)} = d_x/\ell_x$.

The continuous method based on $T^{(1)}, \ldots, T^{(n)}$, $n = 69$, and the assumption of constant forces of mortality between integer ages, factorises the likelihood

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

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
$\hat{q}_x^{(0)}$	0.652	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 i. nor ii. is optimal. An initially decreasing force of mortality would be better.