B.1 Lifetime distributions

1. Lifetimes of humans, animals, ...; lifetime of a machine before breakdown; of a car or aeroplane before withdrawal from service; lifetime of a lightbulb before burn-out; time a new employee spends at a firm before retirement/resignation/dismissal; survival time of patients after beginning treatment; time from infection by a disease until symptoms appear; time from a couple starting to try to conceive and first pregnancy; time to weaning of newborn babies,....

2. (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. \( \lambda \) 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
\[ \lambda = \hat{\lambda} = \frac{n}{L_1 + \cdots + L_n} \]

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
\[ \hat{\lambda} = \frac{20}{l_1 + \cdots + l_{20}} = 0.002917. \]

ii. The Fisher information is given by
\[ I_n(\lambda) = -\mathbb{E}(\ell''(\lambda)) = \frac{n}{\lambda^2} \]

so that, approximately \( \hat{\lambda} \sim N(\lambda, \lambda^2/n) \).

In the standard way, we can proceed to approximate the variance \( \lambda^2/n \) by \( \hat{\lambda}^2/n \), so that \( \hat{\lambda} \approx N(\lambda, \hat{\lambda}^2/n) \). From this we have
\[ .95 = \mathbb{P}(|Z| < 1.96) \approx \mathbb{P} \left( \left| \frac{\hat{\lambda} - \lambda}{\hat{\lambda}/\sqrt{n}} \right| < 1.96 \right) \]
\[ = \mathbb{P}(\hat{\lambda} - 1.96\hat{\lambda}/\sqrt{n} < \lambda < \hat{\lambda} + 1.96\hat{\lambda}/\sqrt{n}) \]
\[ = \mathbb{P} \left( \frac{1}{\lambda + 1.96\lambda/\sqrt{n}} < \frac{1}{\lambda} < \frac{1}{\lambda - 1.96\lambda/\sqrt{n}} \right), \]

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 \( 1/\lambda \).
Alternatively, in the exponential case one can also proceed without approximating the variance, since we have

\[
.95 = P(|Z| < 1.96) \approx P \left( \left| \frac{\lambda - \lambda}{\lambda \sqrt{n}} \right| < 1.96 \right) \\
= P \left( \frac{1}{\lambda} \left( 1 - \frac{1.96}{\sqrt{n}} \right) < \frac{1}{\lambda} < \frac{1}{\lambda} \left( 1 + \frac{1.96}{\sqrt{n}} \right) \right)
\]

which gives instead an interval of (0.002028, 0.005193) for \( \lambda \) and (192.6, 493.1) for \( 1/\lambda \).

iii. Since \( L_1 + \cdots + L_n \sim \Gamma(n, \lambda) \), we have \( 2\lambda(L_1 + \cdots + L_n) \sim \Gamma(n, \frac{1}{2}) \).

To show that \( \Gamma(n, \frac{1}{2}) \) is the same as \( \chi^2_{2n} \), it’s enough to show that \( \Gamma(\frac{1}{2}, \frac{1}{2}) \) is the same as \( \chi^2_1 \), since we know that the sum of \( 2n \) independent copies of a \( \Gamma(\frac{1}{2}, \frac{1}{2}) \) has \( \Gamma(n, \frac{1}{2}) \) distribution, while the sum of \( 2n \) independent copies of a \( \chi^2_1 \) has \( \chi^2_{2n} \) distribution.

That is, we want to show that if \( X \sim N(0, 1) \), then \( X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2}) \).

Let \( Y = X^2 \). Since \( X \) has density function \( f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \), a change of variables gives the density function for \( Y \):

\[
f_Y(y) = \frac{2}{\sqrt{2\pi}} \exp \left( -\frac{y}{2} \right) \left( \frac{dy}{dx} \right) \\
\propto \frac{1}{\sqrt{y}} \exp \left( -\frac{y}{2} \right)
\]

which is proportional to the density function of the \( \Gamma(\frac{1}{2}, \frac{1}{2}) \) distribution, as desired.

Taking the lower and upper 2.5% quantiles from a table of the \( \chi^2_{40} \) distribution, we have

\[
0.95 = P(24.43 < |X^2| < 59.34) = P(24.43 < 2\lambda n/\hat{\lambda} < 59.34) \\
= P \left( \frac{2n}{59.34\hat{\lambda}} < \frac{1}{\hat{\lambda}} < \frac{2n}{24.43\lambda} \right),
\]

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

iv. We count (0, 3, 4, 8, 2, 2, 1) in \((100k, 100(k+1)), k = 0, \ldots, 6\). If “fridgelifetimes” is a vector in R containing the data, then the command \texttt{hist("fridgelifetimes", xlab="hours")} produces the following histogram:
Note there is an initial interval $[0, 100]$ not shown in the range. The shape does not look consistent with a constant failure rate (which is what corresponds to the exponential distribution). There is far too little mass on small values, and too much of a peak in the middle. For example, 0 out of 20 machines fail before time 100, but 8 out of the 13 surviving machines fail between times 300 and 400.

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

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

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

<table>
<thead>
<tr>
<th>Bin</th>
<th>0-100</th>
<th>100-400</th>
<th>400-∞</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>observed</td>
<td>0</td>
<td>15</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>expected</td>
<td>5.1</td>
<td>8.7</td>
<td>6.2</td>
<td>20</td>
</tr>
</tbody>
</table>

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

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

so there is strong evidence against exponentiality.

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

$$P(T > t) = P(T_1 > t, \ldots, T_m > t) = P(T_1 > t) \cdots P(T_m > t)$$

$$= \exp \left\{ -\int_0^t h_1(s)ds \right\} \cdots \exp \left\{ -\int_0^t h_m(s)ds \right\}$$

$$= \exp \left\{ -\int_0^t (h_1(s) + \cdots + h_m(s))ds \right\}.$$  

(b) By (a), the hazard function of $T$ now is $k_1t^n + \cdots + k_m t^n$, and since hazard functions determine distributions, $T$ has a Weibull distribution with rate parameter $k = k_1 + \cdots + k_m$ and exponent $n$. 
(c) The exponential distribution truncated at $\omega$ is simply the distribution of an exponential random variable conditioned not to exceed $\omega$. We first calculate the survival function and let $\lambda \to 0$ to get

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

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

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

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

4. Prior:

$$f_{\mu}^{\text{prior}}(m) \propto m^{\alpha - 1} e^{-\beta m}.$$

Likelihood:

$$L(m; T_1, \ldots, T_n) = \prod_{i=1}^{n} me^{-m T_i} = m^n e^{-m \sum T_i}.$$

Since the posterior density is proportional to the product of the prior density and the likelihood, the posterior density is given by

$$f_{\mu}(m) \propto m^{\alpha + n - 1} e^{-(\beta + \sum T_i)m}.$$

So the posterior distribution is $\Gamma(\alpha + n, \beta + \sum T_i)$.

5. (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

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

Finally, by Tonelli’s theorem

$$F_T(t) = \int_t^\infty \int_0^\infty e^{-\lambda s} f_M(\lambda)d\lambda ds = \int_0^\infty e^{-\lambda t} f_M(\lambda)d\lambda = M_M(-t).$$
(b) Following the hint, we calculate the moment generating function of $\tilde{M} \sim \text{Poi}(\mu)$

$$M_{\tilde{M}}(-t) = \mathbb{E}(e^{-t\tilde{M}}) = \sum_{k=0}^{\infty} e^{-kt} \frac{\mu^k}{k!} e^{-\mu} = \exp \left\{ -\mu(1 - e^{-t}) \right\}.$$ 

To apply (a), we compare this with the survival function of the Gompertz-Makeham variable $T$

$$\bar{F}(t) = \exp \left\{ - \int_0^t h(s)ds \right\} = \exp \left\{ -\rho_0 t + \frac{\rho_1}{\rho_2} (1 - e^{\rho_2 t}) \right\}.$$ 

If $\rho_0 = 0$, $\rho_2 = -1$, then $\mu = \rho_1$ works. To get a factor $e^{-\rho_0 t}$ we just take $M_{\rho_0 + \tilde{M}}(-t)$, to get a coefficient $e^{-\rho_2 t}$ for $t$, we take $M_{\rho_0 - \rho_2 \tilde{M}}(-t) = \mathbb{E} \left( \exp \left\{ -t\rho_0 + t\rho_2 \tilde{M} \right\} \right) = \exp \left\{ -\rho_0 t - \mu(1 - e^{\rho_2 t}) \right\},$

so $\mu = -\rho_1/\rho_2$. Hence $M = \rho_0 - \rho_2 \tilde{M}$, i.e.

$$\mathbb{P}(M = \rho_0 - \rho_2 k) = \left( -\frac{\rho_1}{\rho_2} \right)^k \frac{k!}{\rho_2^k}, \quad k \geq 0.$$

6. (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, given $\ell_x$, $d_x \sim \text{Bin}(\ell_x, q_x)$, so that given $\ell_x$, $d_x - q_x \ell_x$ has mean 0 and variance $\ell_x q_x(1 - q_x)$.

Hence we have

$$\mathbb{E}(d_x - q_x \ell_x) = \mathbb{E} \left( \mathbb{E}(d_x - q_x \ell_x | \ell_x) \right) = \mathbb{E}(q_x \ell_x - q_x \ell_x) = 0,$$

and

$$\text{Var}(d_x - q_x \ell_x) = \mathbb{E} \left( (d_x - q_x \ell_x)^2 \right) = \mathbb{E} \left( \mathbb{E} \left( (d_x - q_x \ell_x)^2 | \ell_x \right) \right) = \mathbb{E}(\ell_x q_x(1 - q_x)) = q_x(1 - q_x)\mathbb{E}\ell_x.$$ 

(b) The likelihood of the data is given by

$$\prod_i q_i^{d_i} (1 - q_i)^{\ell_i - d_i}.$$ 

This factorizes so we can calculate each $\hat{q}_x$ separately.

To maximize $\hat{q}_0^{d_0} (1 - \hat{q}_0)^{\ell_0 - d_0}$ we take $\hat{q}_0 = d_0/\ell_0$ (the familiar form of the MLE for the binomial distribution).

Since $\ell_0$ is deterministic, we have

$$\mathbb{E}\hat{q}_0 = \frac{\mathbb{E}d_0}{\ell_0} = \frac{q_0 \ell_0}{\ell_0} = q_0,$$
so indeed $\hat{q}_0$ is an unbiased estimator of $q_0$.

Its variance is

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

Given the data, we can estimate this by

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

We might proceed in the same way for $x \geq 1$, but now $\ell_x$ is random and there is the complication that it may take the value 0.

If $\ell_x > 0$, then $\hat{q}_x = d_x/\ell_x$ in a similar way, but with positive probability $\ell_x = 0$, in which case the data gives us no information about $q_x$, and the value of $q_x$ makes no difference to the likelihood.

We can put $\hat{q}_x$ equal to some constant $c_x$ in the case $\ell_x = 0$ (for example, $c_x = 1$). Then

$$\mathbb{E}\hat{q}_x = c_x \mathbb{P}(\ell_x = 0) + \sum_{n \geq 1} \mathbb{P}(\ell_x = n) \mathbb{E}\left(\frac{d_x}{\ell_x} \mid \ell_x = n\right)$$

$$= c_x \mathbb{P}(\ell_x = 0) + q_x \mathbb{P}(\ell_x > 0).$$  

Whatever value we take for $c_x$, this is not in general equal to $q_x$, so the estimator $\hat{q}_x$ is not unbiased.

For an asymptotic estimate of the variance, we can calculate the Fisher information matrix by looking at the second derivatives of the log likelihood ratio which is

$$\sum_{x \geq 0} (\ell_x - d_x) \log(1 - q_x) + d_x \log q_x.$$  

The off-diagonal entries are 0, and the diagonal entries are given by

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

so that

$$\text{Var}(\hat{q}_x^{(0)}) \approx \frac{1}{I_{xx}(q)} = \frac{q_x(1 - q_x)}{\mathbb{E}(\ell_x)} \approx \frac{\hat{q}_x(1 - \hat{q}_x)}{\ell_x} = \frac{d_x(\ell_x - d_x)}{\ell_x^2}.$$