Appendix B

Solutions

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

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

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

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

so there is strong evidence against exponentiality.

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

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

(b) By (a), the hazard function of T now is $k_1 t^n + \ldots + k_m t^n$, so T has a Weibull distribution with rate parameter $k = k_1 + \ldots + k_m$ 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

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

3. (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 $\mathbb{E}(T^2|M=\lambda) = \mathrm{Var}(T|M=\lambda) + (\mathbb{E}(T|M=\lambda))^2 = 2\lambda^{-2}$,

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

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 distribution is called Pareto distribution.

(c) Following the hint, we calculate the moment generating function of $\tilde{M} \sim \mathrm{Poi}(\mu)$

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

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

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

If $\rho_0 = 0$, $\rho_2 = -1$, then $\mu = \rho_1$ works. To get a factor $e^{-\rho_0 t}$ we just take $\mathcal{M}_{\rho_0+\tilde{M}}(-t)$, to get a coefficient ρ_2 for t, we take

$$\mathcal{M}_{\rho_0-\rho_2\tilde{M}}(-t) = \mathbb{E}\left(\exp\left\{-t\rho_0 - t\rho_2\tilde{M}\right\}\right) = \exp\left\{-\rho_0 t - \mu(1-e^{\rho_2 t})\right\},$$

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

$$\mathbb{P}(M = \rho_0 - \rho_2 k) = \frac{(-\rho_1/\rho_2)^k}{k!} e^{\rho_1/\rho_2}.$$

4. (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 $\ell_x = n$.

Now we condition on ℓ_x to get

$$\begin{aligned}\mathbb{E}(d_x) &= \sum_{n=0}^{\ell_0} \mathbb{P}(\ell_x = n)\mathbb{E}(d_x|\ell_x = n) \\ &= \sum_{n=0}^{\ell_0} \mathbb{P}(\ell_x = n) n q_x \\ &= q_x \mathbb{E}(\ell_x).\end{aligned}$$

Similarly,

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

where now

$$\begin{aligned}\mathbb{E}(d_x^2) &= \sum_{n=0}^{\ell_0} \mathbb{P}(\ell_x = n)\left(\mathrm{Var}(d_x|\ell_x = n) + (\mathbb{E}(d_x|\ell_x = n))^2\right) \\ &= \sum_{n=0}^{\ell_x} \mathbb{P}(\ell_x = n)(nq_x(1-q_x) + n^2q_x^2) = q_x(1-q_x)\mathbb{E}(\ell_x) + q_x^2\mathbb{E}(\ell_x^2),\end{aligned}$$

and therefore

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

(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

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