Spring 2005

**Actuarial Mathematics and Statistics**
**Statistics 5 Part 2: Statistical Inference**
**Tutorial Solutions**

1. (a) Can interpreted as a frequency. *Repeatable expt:* Repeatedly roll the die and record the sequence of score.

   (b) Can interpreted as a frequency. *Repeatable expt:* Select individual randomly from all 20-yr-old males then record lifespan. Note that this interpretation assumes an infinite pool of people to choose from - not entirely realistic.

   (c) Cannot be interpreted as a frequency. There's only one Wayne Rooney.

   (d) Cannot be interpreted as a frequency.

   (e) Can be interpreted as a frequency if you regard the next pint as a random draw from the population of pints of milk in supermarkets.

2. (a) Bin(10, $p$) form some unknown $p$.

   (b) Geometric($p$) might be plausible if results of throws are independent and probability of success $p$ is constant over throws. Not entirely realistic given $p$ might increase with practice.

   (c) Poisson($\lambda$) might be appropriate here.

   (d) A normal distribution: N($\mu$, $\sigma^2$).

   (e) Possibly a $\Gamma(\alpha, \beta)$ distribution or a normal distribution would be appropriate.

3. (a)

$$\begin{aligned}
\text{MSE}(g(\underline{X})) &= E((g(\underline{X}) - \theta)^2) \\
&= E((g(\underline{X}))^2) - 2E(g(\underline{X}))\theta + \theta^2 \\
&= [E((g(\underline{X}))^2) - (E(g(\underline{X})))^2] + [(E(g(\underline{X})))^2 - 2E(g(\underline{X}))\theta + \theta^2] \\
&= \text{Var}(g(\underline{X})) + [E(g(\underline{X})) - \theta]^2 \\
&= \text{Var}(g(\underline{X})) + (\text{bias}(g(\underline{X})))^2.
\end{aligned}$$

   (b)

(i) $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, so $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^n E(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \mu$.

(ii) $\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \sigma^2/n$.

(iii)

$$\begin{aligned}
S^2 &= \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^n X_i^2 - \frac{2}{n-1}\bar{X}\sum_{i=1}^n X_i + \frac{n}{n-1}(\bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^n X_i^2 - \frac{2n}{n-1}(\bar{X})^2 + \frac{n}{n-1}(\bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^n X_i^2 - \frac{n}{n-1}(\bar{X})^2.
\end{aligned}$$

Recall that $E(X^2) = Var(X) + (E(X))^2 = \sigma^2 + \mu^2$ and $E(\bar{X}^2) = Var(\bar{X}) + (E(\bar{X}))^2 = \frac{\sigma^2}{n} + \mu^2$. So

$$\begin{aligned}
E(S^2) &= \frac{1}{n-1}\sum_{i=1}^n E(X_i^2) - \frac{n}{n-1}E(\bar{X}^2) \\
&= \frac{n}{n-1}(\sigma^2 + \mu^2) - \frac{n}{n-1}(\frac{\sigma^2}{n} + \mu^2) \\
&= \sigma^2.
\end{aligned}$$

4. (a) Yes. $E(\bar{X}) = \text{E(X)} = \lambda$, therefore $\text{Bias}(\bar{X}) = 0$.

   (b) $\text{MSE} = \text{Var}(\bar{X}) = \frac{\lambda}{n}$, since $Var(X) = \lambda$.

   (c) For Poisson($\lambda$), $f_X(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0, 1, \ldots$, so that $\log(f_X(X, \lambda) = X\log(\lambda) - \lambda - \log(X!)$ and hence,

$$\frac{\partial}{\partial\lambda}\log(f_X(X;\lambda)) = \frac{X-\lambda}{\lambda}.$$

   It follows that

$$\begin{aligned} E((\frac{\partial}{\partial\lambda}\log(f_X(X;\lambda))^2) &= E(\frac{(X-\lambda)^2}{\lambda^2}) \\ &= \frac{\text{Var}(X)}{\lambda^2} \\ &= \frac{1}{\lambda}. \end{aligned}$$

   From the definition (see lecture notes!) it follows that the CRLB is equal to $\frac{\lambda}{n} = \text{MSE}(\bar{X})$, so that $\bar{X}$ is indeed the most efficient unbiased estimator.

   (d) We have that

$$\begin{aligned} \text{MSE}(a\bar{X}) &= \text{Bias}^2 a\bar{X} + \text{Var}(a\bar{X}) \\ &= (a-1)^2\lambda^2 + \frac{a^2\lambda}{n} \end{aligned}$$

   This is a quadratic in $a$ which can easily be shown to be minimised by $a = \frac{n\lambda}{n\lambda+1}$. Note that we would need to know $\lambda$ in order to obtain the optimal $a$ and that the optimal value of $a \to 1$ as the sample size $n$ increases.

5. (a) Let

$$f(a,b) = MSE(a,b) = a^2\sigma_1^2 + b^2\sigma_2^2 + \theta^2(a+b-1)^2$$

   then

$$\begin{aligned} \frac{\partial f}{\partial a} &= 2\sigma_1^2 a + 2\theta^2(a+b-1) \\ \frac{\partial f}{\partial a} &= 2\sigma_2^2 b + 2\theta^2(a+b-1). \end{aligned}$$

Setting these equal to zero we obtain the simultaneous linear equations.

$$\begin{aligned}(\sigma_1^2 + \theta^2)a + \theta^2 b &= \theta^2 \\ \theta^2 a + (\sigma_2^2 + \theta^2)b &= \theta^2.\end{aligned}$$

These can be solved to obtain the expressions for $a^*$ and $b^*$ given in the lectures. To check we its a maximum we apply the second derivative test to the matrix of 2nd derivatives. Note that $\frac{\partial^2 f}{\partial a^2} = \sigma_1^2 + \theta^2$, $\frac{\partial^2 f}{\partial b^2} = \sigma_2^2 + \theta^2$, $\frac{\partial^2 f}{\partial a \partial b} = \theta^2$. Application of the test is straightforward.

(b) The MSE for the optimal unbiased estimator is 0.5. For the optimal estimator $a^* = 0.49998$ and $b^* = 0.49998$ and we obtain an MSE of 0.49998. Clearly there is little difference between the optimal unbiased and the optimal estimators in this case.

6. (a)

From the results in question 3, we have that $E(\bar{X}) = \mu = \theta$, $\text{Var}(\bar{X}) = \frac{\sigma^2}{n} = \frac{\theta^2}{n}$, and $\text{MSE}(\bar{X}) = \frac{\theta^2}{n}$.

(b) Now consider the estimator $Y = a\bar{X}$. $E(Y) = E(a\bar{X}) = a\theta$, so bias($Y$)=$E(Y) - \theta = (a-1)\theta$, and Var($Y$)= Var($a\bar{X}$)= $a^2 \cdot \frac{\theta^2}{n}$. Thus,

$$\text{MSE}(Y) = a^2 \cdot \frac{\theta^2}{n} + (a-1)^2\theta^2 = \frac{\theta^2}{n}(a^2 + n(a-1)^2).$$

To minimize MSE($Y$), we must solve

$$\frac{d}{da}\text{MSE}(Y) = \frac{\theta^2}{n}(2a + 2n(a-1)) = 0.$$

We get the solution $a = \frac{n}{n+1}$. Note: You must check that MSE($Y$) is *minimized* at this value of $a$. Argue directly, or check that $\frac{d^2}{da^2}\text{MSE} > 0$. Let $Y^* = \frac{n}{n+1}\bar{X}$, then

$$\text{MSE}(Y^*) = \frac{\theta^2}{n}\left(\left(\frac{n}{n+1}\right)^2 + n\left(\frac{-1}{n+1}\right)^2\right) = \frac{\theta^2}{n+1} < \frac{\theta^2}{n} = \text{MSE}(\bar{X}).$$

Thus, $Y^*$ is more efficient than $\bar{X}$. For small values of $n$ this difference, proportionally, is quite substantial, but as $n \to \infty$, $\bar{X}$ becomes nearly as efficient as $Y^*$.

(c) Using Chebyshev's inequality we obtain

$$P(|\bar{X} - \theta| > 0.1\theta) \leq \frac{MSE(\bar{X})}{0.01\theta^2} = \frac{1}{2}.$$

.

7. Suppose the population distribution is $\exp(1/\theta)$ and the size of the random sample is $n$.

Then $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$ and $\log f(x;\theta) = -\log\theta - \frac{x}{\theta}$.

Differentiate to get $\frac{d}{d\theta}\log f(x;\theta) = \frac{-1}{\theta} + \frac{x}{\theta^2} = \frac{1}{\theta^2}(x-\theta)$.

So $(\frac{d}{d\theta}\log f(x;\theta))^2 = \frac{(x-\theta)^2}{\theta^4}$ and

$$E((\frac{d}{d\theta}\log f(x;\theta))^2) = \frac{1}{\theta^2}E((X-\theta)^2) = \frac{\text{Var}(X)}{\theta^4} = \frac{1}{\theta^2}.$$

Thus the Cramer-Rao lower bound is given by $\frac{1}{n\cdot\frac{1}{\theta^2}} = \frac{\theta^2}{n}$. Since $\bar{X}$ is unbiased and $\text{Var}(\bar{X}) = \theta^2/n$, $\bar{X}$ is the most efficient unbiased estimator BUT it is NOT the most efficient estimator (cf question 6.)

8. We need to check that the MSE $\to 0$ as $n \to \infty$. It is easily shown that $E(\frac{X}{n}) = p$ and $\text{Var}(\frac{X}{n}) = \frac{p(1-p)}{n}$. Hence $\frac{X}{n}$ is unbaised and its MSE is equal to its variance. Clearly the variance tends to zero as $n \to \infty$ and it follows that $\frac{X}{n}$ is a consistent estimator for $p$.

9. (a) The density for the $N(0,\sigma^2)$ distribution is given by $f(x;\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{x^2}{2\sigma^2})$. The parameter is $\theta = \sigma^2$. Now we compute the Cramer-Rao lower bound.

$$\begin{aligned}
\log f(x;\theta) &= \log(1/\sqrt{2\pi}) - \frac{1}{2}\log(\theta) - \frac{1}{2}\frac{x^2}{\theta} \\
\frac{d}{d\theta}\log f(x;\theta) &= -\frac{1}{2\theta} + \frac{x^2}{2\theta^2} \\
\frac{d^2}{d\theta^2}\log f(x;\theta) &= \frac{1}{2\theta^2} - \frac{x^2}{\theta^3} \\
E(\frac{d^2}{d\theta^2}\log f(x;\theta)) &= \frac{1}{2\theta^2} - \frac{E(X^2)}{\theta^3}.
\end{aligned}$$

Since $\mu = 0$, $E(X^2) = \text{Var}(X) = \sigma^2 = \theta$. Thus

$$E(\frac{d^2}{d\theta^2} \log f(x;\theta)) = \frac{1}{2\theta^2} - \frac{\theta}{\theta^3} = \frac{-1}{2\theta^2} = \frac{-1}{2\sigma^4}.$$

So the Cramer-Rao lower bound is given by $\frac{2\sigma^4}{n}$.

(b) Since $E(S^2) = \sigma^2$ (see question 2, above), $S^2$ is an **unbiased** estimator for $\sigma^2$. However, $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$. So $S^2$ is **not** the *most efficient* unbiased estimator, although, as $n \to \infty$, it 'becomes' most efficient.

In the special case when $\underline{X}$ is a random sample from $N(0, \sigma^2)$, we have $E(X^2) = \text{Var}(X) = \sigma^2$. Thus, another unbiased estimator of $\sigma^2$ is given by $S_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2$, with $E(S_n^2) = \frac{1}{n}\sum_{i=1}^n E(X_i^2) = \sigma^2$.

Now we compute the $\text{MSE}(S_n^2) = \text{Var}(S_n^2)$: First note that

$$\text{Var}(S_n^2) = \frac{1}{n^2}\sum_{i=1}^n \text{Var}(X_i^2) = \frac{1}{n^2}\sum_{i=1}^n 2\sigma^4 = \frac{2\sigma^4}{n}.$$

Since $\text{Var}(S_n^2)$ attains the Cramer-Rao lower bound and $S_n^2$ is an unbiased estimator of $\sigma^2$, we conclude that $S_n^2$ is the **most efficient unbiased estimator** of $\sigma^2$.

Now consider $S_{n+2}^2 = \frac{1}{n+2}\sum_{i=1}^n X_i^2 = \frac{n}{n+2}S_n^2$. We have $E(S_{n+2}^2) = \frac{n}{n+2}E(S_n^2) = \frac{n}{n+2}\sigma^2$, so $S_{n+2}^2$ is biased with

$$\text{bias}(S_{n+2}^2) = \frac{n}{n+2}\sigma^2 - \sigma^2 = -\frac{2}{n+2}\sigma^2.$$

Next, $\text{Var}(S_{n+2}^2) = \frac{n^2}{(n+2)^2}\frac{2\sigma^2}{n}$, so

$$\text{MSE}(S_{n+2}^2) = \text{Var}(S_{n+2}^2) + (\text{bias}(S_{n+2}^2))^2 = \frac{2\sigma^4}{n+2}.$$

Therefore, although $S_n^2$ is the most efficient *unbiased* estimator, it is not the most efficient estimator as $S_{n+2}^2$ is more efficient.

10. (a) Since the $n$ individuals are chosen randomly without replacement the *ith* individual selected is equally likely to be any one of the 100 individuals in the populaton. Thus $P(Y_i = 1) = p$, so that $Y_i \sim Bernoulli(p)$ as required.

(b) The $Y_i$ are not independent since we are selecting without replacement. Thus, for example, if $i \neq j$, $P(Y_i = 1|Y_j = 0) > p$.

(c) Note that $\hat{p}$ is unbiased since $E(\hat{p}) = p$. From the formula given in the lecture notes we have that

$$\text{Var}(\hat{p}) = \frac{(N-n)}{(N-1)} \frac{\text{Var}(Y_i)}{n} = \frac{50}{99} \frac{p(1-p)}{50} = 0.0024$$

(d) $P(|\hat{p} - p| > 0.05) < \frac{MSE}{0.05^2} = \frac{0.0024}{0.0025} = 0.96$.

This suggests that Chebyshev's inequality is giving a fairly loose upper bound. In reality the frequency will be considerably less than this.

11. (a) Since observations are independent we have that

$$L(\lambda) = \frac{e^{-\lambda}\lambda^3}{3!} \times \frac{e^{-\lambda}\lambda^6}{6!} \times \frac{e^{-\lambda}\lambda^2}{2!} = \frac{e^{-3\lambda}\lambda^{11}}{8640}$$

The log of the likelihood is given by

$$l(\lambda) = 11 \log \lambda - 3\lambda - \log 8640.$$

(b) Differentiate, equate to zero, solve + check maximised to obtain the MLE of $\lambda$ as $\hat{\lambda} = \frac{11}{3} = 3.67$.

(c) For the Poisson($\lambda$) model the population mean is equal to $\lambda$. We obtain the method of moments estimate by equating the sample mean to population mean and solving for $\lambda$. This gives the same estimate as (b).

12. (a) First, $\mu = np$, so set $\bar{X} = X_1 = np$ and solve to get $\tilde{p} = X_1/n$=MME. To find the MLE, maximize the log-likelihood function:

$$\begin{aligned} L(p) &= \binom{n}{x} p^x (1-p)^{n-x} \\ \log L(p) &= \log \binom{n}{x} + x \log p + (n-x)\log(1-p) \\ \frac{d}{dp} \log L(p) &= \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}. \end{aligned}$$

Solve $\frac{d}{dp} \log L(p) = \frac{x-np}{p(1-p)} = 0$ to get $\hat{p} = X_1/n$=MLE.

**Note:** You should check that the *maximum* occurs at $\hat{p}$, i.e. compute the $2^{nd}$derivative of $\log L(p)$.

**(b)** For the exponential distribution, exp($\lambda$), we have the equation $\mu = 1/\lambda$. So to find the MME, set $\bar{X} = 1/\lambda$ and solve for $\lambda$ to get $\tilde{\lambda} = 1/\bar{X} = MME$.

To find the MLE, maximize the log-likelihood function:

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \\ &= \lambda^n e^{-\lambda \Sigma x_i} \\ \log L(\lambda) &= n \log \lambda - \lambda \sum_{i=1}^{n} x_i \\ \frac{d}{d\lambda} \log L(\lambda) &= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i. \end{aligned}$$

Solve $\frac{d}{d\lambda} \log L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$ to get $\hat{\lambda} = \frac{n}{\Sigma x_i} = 1/\bar{X}$=MLE.

**Note:** $\frac{d^2}{d\lambda^2} \log L(\lambda) = -n/\lambda^2 < 0$, so log-likelihood is *maximized* at $\hat{\lambda}$.

**(c)** To find the MME, note that $\mu = \frac{\theta}{\theta+1}$ (check this by computing the expectation directly). Therefore, set $\bar{X} = \frac{\theta}{\theta+1}$ and solve for $\theta$ to get $\tilde{\theta} = \bar{X}/(1-\bar{X}) = MME.$

To find the MLE, maximze the log-likelihood function:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n (\prod_{i=1}^{n} x_i)^{\theta-1} \\ \log L(\theta) &= n \log \theta + (\theta - 1) \log(\Pi x_i) \\ \frac{d}{d\theta} \log L(\theta) &= \tfrac{n}{\theta} + \log(\Pi x_i). \end{aligned}$$

Solve $\frac{d}{d\theta} \log L(\theta) = 0$ to get $\hat{\theta} = -n/(\log(\Pi X_i)) = -n/\Sigma \log(X_i) = -1/\overline{\log X}$=MLE.

Log-likelihood is *maximized* at $\hat{\theta}$, since $\frac{d}{d\theta} \log L(\theta) = -n/\theta^2 < 0$.

13. (a) There must be at least 2 and at most 8 black balls in the bag (i.e. $2 \le r \le 8$).

(b)

$$L(r) = \frac{\binom{r}{2} \times \binom{10-r}{2}}{\binom{10}{4}} \propto r(r-1)(10-r)(10-r-1).$$

(c) $L(r)$ is maximised for $r = 5$ when $r(r-1)(10-r)(10-r-1) = 400$. This is not surprising given that half the balls in the sample were black.

14. To find the MLE, first determine the likelihood function:

$$L(\theta) = \frac{4}{\theta^2}\left(1 - \frac{x_1}{\theta}\right)\left(1 - \frac{x_2}{\theta}\right),$$

for $\theta \geq \max\{x_1, x_2\}$ and is equal to 0, otherwise. Check a plot of $L(\theta)$.

The MLE should be one of the critical points for the *full* function $G(\theta) = \frac{4}{\theta^2}\left(1 - \frac{x_1}{\theta}\right)\left(1 - \frac{x_2}{\theta}\right)$. So find the critical points of $G(\theta)$ and then work out which one is the MLE $\hat{\theta}$.

$$\begin{aligned} G(\theta) &= 4\left(\frac{1}{\theta^2} - \frac{(x_1+x_2)}{\theta^3} + \frac{x_1x_2}{\theta}\right) \\ \frac{d}{d\theta}G(\theta) &= 4\left(-\frac{2}{\theta^3} + \frac{3(x_1+x_2)}{\theta^4} - \frac{4x_1x_2}{\theta^5}\right) \\ &= -\frac{4}{\theta^5}\left(2\theta^2 - 3(x_1+x_2)\theta + 4x_1x_2\right). \end{aligned}$$

Note that $\frac{d}{d\theta}G(\theta) = 0$ iff $2\theta^2 - 3(x_1+x_2)\theta + 4x_1x_2 = 0$. Using the quadratic formula, we get:

$$\theta = \frac{3(x_1+x_2) \pm \sqrt{9(x_1+x_2)^2 - 32x_1x_2}}{4}.$$

Which root is the MLE?

Consider the graph of the full function $G(\theta) = \frac{4}{\theta^2}\left(1 - \frac{x_1}{\theta}\right)\left(1 - \frac{x_2}{\theta}\right)$ (you should sketch this - see tutorial). Note that $G(\theta) = 0$ at $\theta = x_{(1)}, x_{(2)}$.

Since the minimum occurs in the interval $(x_{(1)}, x_{(2)})$ and $G$ has a maximum at some value greater than $x_{(2)}$, the MLE must be the larger root of the quadratic. So

$$\hat{\theta} = \frac{3(x_1+x_2) + \sqrt{9(x_1+x_2)^2 - 32x_1x_2}}{4}.$$

The MME is much easier to find! First find the theoretical mean of the population:

$$\mu = E(X) = \int_0^\theta x\frac{2}{\theta}\left(1-\frac{x}{\theta}\right)dx = \frac{2}{\theta}\int_0^\theta (x-\frac{x^2}{\theta})dx = \frac{\theta}{3}.$$

Solve $\bar{X} = \theta/3$ to get $\theta = 3\bar{X}$=MME.

15. (a) It is the region

$$\{(a,b)|a \leq x_{(1)}, b \geq x_{(n)}\}.$$

(b) For $(a,b)$ in this region we have that

$$L(a,b) = \frac{1}{(b-a)^n}.$$

It is maximised by selecting the values of $a$ and $b$ that minimise $|b-a|$ over this region i.e. at $(\hat{a},\hat{b}) = (x_{(1)}, x_{(n)}$.

(b) No - in the particular case when $n=1$, $\hat{a}=\hat{b}$ for any sample so they are certainly not independent in this (rather artificial) case. Let $n=2$, $a=0$ and $b=1$ and let $\underline{Y} = (X_{(1)}, X_{(2)})$. Then the range of $\underline{Y}$ is $\{(y_1,y_2)|0 \leq y_1 \leq y_2 \leq 1\}$. If you look at the solution to Qu. 3 in Tutorial 7 of Statistics IV last term you will see that the density of $\underline{Y}$ is constant over this region and that $Cov(Y_1,Y_2) = \frac{1}{36}$.

16. (a) The probability that there are no cars speeding cars on a given day is given by $e^{-\lambda}$. It follows that

$$L(\lambda) = e^{-4\lambda}(1 - e^{-\lambda})^6.$$

Taking logs, differentiating and equating to 0, we find that the MLE of $\lambda$ is given by

$$\hat{\lambda} = -\log 0.4 = 0.916.$$

(b)It follows a Bin(10, $p$) distribution where $p = e^{-\lambda}$.

(c) We can find the MLE of $\lambda$ by using the invariance of MLE's under 1-1 transformations. We know that the MLE of $p$ is $\frac{4}{10}$ so that $e^{-\hat{\lambda}} = 0.4$.

17. The probability that a player loses is $q = (1-p)^3$ (a 1-1 function of $p$, note). It is clear that the MLE of $q$ is $\hat{q} = \frac{1}{2}$. From this it follows from invariance property that $1 - \hat{p} = \hat{q}^{\frac{1}{3}}$ so that

$$\hat{p} = 1 - \frac{1}{2^{\frac{1}{3}}} = 0.206.$$

18. (a) Let $X$ denote the diameter of a tuber. Then the probability that a tuber goes through both grids is $P(X < 3) = 1 - e^{-3\lambda}$. The probability that it goes through *neither* is $P(X > 6) = e^{-6\lambda}$ while the probability that it goes through the 1st but not the second is $P(3 < X < 6) = e^{-3\lambda} - e^{-6\lambda} = e^{-3\lambda}(1 - e^{-3\lambda})$. These probabilities define the factors in the likelihood for each of the tubers in the sample. n the three groups listed above there are 50, 20 and 30. This gives a likelihood

$$L(\lambda) = (1 - e^{-3\lambda})^{50}(e^{-6\lambda})^{20}(e^{-3\lambda}(1 - e^{-3\lambda}))^{30}$$

giving the likelihood given in the question.

(b) The likelihood can we written as $p^{70}(1-p)^{80}$ this is maximised by $\hat{p} = \frac{7}{15}$. Using the invariance property we obtain the MLE of $\lambda$ as

$$\hat{\lambda} = \frac{(\log 15 - \log 7)}{3} = 0.254$$

19. For the Poisson distribution with parameter $\mu$, we have $P(X = 0) = e^{-\mu}, P(X = 1) = \mu e^{-\mu}$, and $P(X > 1) = 1 - (1 + \mu)e^{-\mu}$. So the likelihood for our data is given by:

$$\begin{aligned} L(\mu) &= (e^{-\mu})^{n_0}(\mu e^{-\mu})^{n_1}(1 - (1+\mu)e^{-\mu})^{n-n_0-n_1} \\ \log L(\mu) &= -n_0\mu + n_1 \log \mu - n_1\mu + (n - n_0 - n_1)\log(1 - (1+\mu)e^{-\mu}) \\ \frac{d}{d\mu}\log L(\mu) &= -(n_0 + n_1) + \frac{n_1}{\mu} + (n - n_0 - n_1)\frac{(1+\mu)e^{-\mu} - e^{-\mu}}{1 - (1+\mu)e^{-\mu}}. \end{aligned}$$

Solving $\frac{d}{d\mu}\log \text{Ł}(\mu) = 0$ is equivalent to solving

$$-(n_0+n_1)\mu(1-(1+\mu)e^{-\mu})+n_1(1-(1+\mu)e^{-\mu})+(n-n_0-n_1)\mu^2 e^{-\mu} = 0,$$

which reduces to

$$n_1 - (n_0 + n_1)\mu + e^{-\mu}(n\mu^2 + n_0\mu - n_1) = 0.$$

Substituting in $n = 20, n_0 = 8, n_1 = 7$ we get

$$7 - 15\mu + e^{-\mu}(20\mu^2 + 8\mu - 7) = 0.$$

So the problem reduces to that of finding the solution to $f(\mu) = 0$ where $f(\mu) = 7 - 15\mu + e^{-\mu}(20\mu^2 - 8\mu - 7)$. Use the *Newton-Raphson method*: Start with a value $\mu_0$, and generate iterations using

$$\mu_{r+1} = \mu_r - \frac{f(\mu_r)}{f'(\mu_r)}, \qquad r = 0, 1, 2..$$

To choose a starting value, note that $\frac{n_0}{n} = 0.4$ and $\frac{n_0+n_1}{n} = 0.75$. According to the Poisson tables, if $X \sim Po(0.95)$ then $P(X = 0) = 0.3867$ and $P(X \leq 1) = 0.7541$, so start with $\mu_0 = 0.95$. Iterating, we get $\hat{\mu} = 0.9415$, to 4 decimal places.

20. First step, find the likelihood function:

$$\begin{aligned} L(t, \lambda) &= \prod_{i=1}^{n} \frac{\lambda^t}{\Gamma(t)} x_i^{t-1} e^{-\lambda x_i} = \frac{\lambda^{nt}}{(\Gamma(t))^n}(\prod_{i=1}^{n} x_i)^{t-1} e^{-\lambda \Sigma x_i} \\ \log L(t, \lambda) &= nt \log \lambda - n \log \Gamma(t) + (t-1)\log(\Pi x_i) - \lambda \Sigma x_i. \end{aligned}$$

Therefore, we must solve the following equations simultaneously:

$$\begin{aligned}
\frac{d}{dt}\log L(t,\lambda) &= n\log\lambda - n\frac{\Gamma'(t)}{\Gamma(t)} + \log(\Pi x_i) = 0 \\
\frac{d}{d\lambda}\log L(t,\lambda) &= \frac{nt}{\lambda} - \sum_{i=1}^{n} x_i = 0.
\end{aligned}$$

The second equation yields: $\lambda = nt/\Sigma x_i = t/\bar{x}$, which can be substitued into the first equation to get

$$n\log(t/\bar{x}) - n\frac{\Gamma'(t)}{\Gamma(t)} + \log(\Pi x_i) = 0.$$

This equation must be solved numerically for $\hat{t}$, then set $\hat{\lambda} = \hat{t}/\bar{x}$.

To find the MME, first note that $\mu = \frac{t}{\lambda}$ and $\sigma^2 = \frac{t}{\lambda^2}$. So we need to solve the following equations simultaneously:

$$\begin{aligned}
\bar{x} &= \frac{t}{\lambda} \\
s^2 &= \frac{t}{\lambda^2}.
\end{aligned}$$

We get $\lambda = t/\bar{x}$ from the first equation, and substituting this into the second equation yields $t = \bar{x}^2/s^2$. So the MME's for $t$ and $\lambda$, respectively, are $\bar{X}^2/S^2$ and $\bar{X}/S^2$.

The MME's are much simpler to obtain than the MLE's. The values of the MME's could be used as the starting values for the numerical evaluation of the MLE's.

21. To find the MLE's, first determine the likelihood function:

$$\begin{aligned}
L(\mu_1,\mu_2,\sigma^2) &= \prod_{i=1}^{n_1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2}\left(\frac{x_{1i}-\mu_1}{\sigma}\right)^2)\cdot\prod_{i=1}^{n_2}\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2}\left(\frac{x_{2i}-\mu_2}{\sigma}\right)^2) \\
\log L(\mu_1,\mu_2,\sigma^2) &= (n_1+n_2)\left(\log(\frac{1}{\sqrt{2\pi}}) - \frac{1}{2}\log(\sigma^2)\right) \\
&\quad - \frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(x_{1i}-\mu_1)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n_2}(x_{2i}-\mu_2)^2.
\end{aligned}$$

Then simultaneously solve the following equations:

$$\begin{aligned}
\frac{d}{d\mu_1}\log L &= \frac{1}{\sigma^2}\sum_{i=1}^{n_1}(x_{1i}-\mu_1)=0 \\
\frac{d}{d\mu_2}\log L &= \frac{1}{\sigma^2}\sum_{i=1}^{n_2}(x_{2i}-\mu_2)=0 \\
\frac{d}{d\sigma^2}\log L &= -\frac{1}{2}(n_1+n_2)\frac{1}{\sigma^2}+\frac{1}{2\sigma^4}\left(\sum_{i=1}^{n_1}(x_{1i}-\mu_1)^2+\sum_{i=1}^{n_2}(x_{2i}-\mu_2)^2\right)=0.
\end{aligned}$$

Solve the first two equations to get $\hat{\mu}_1 = \bar{X}_1$ and $\hat{\mu_2} = \bar{X}_2$. Substitute these solutions into the third equation to get $\hat{\sigma}^2 = \frac{1}{n_1+n_2}\left((n_1-1)S_1^2+(n_2-1)S_2^2\right)$

22. To find the MLE, first determine the likelihood function:

$$\begin{aligned}
L(\mu) &= \prod_{i=1}^{n}\left(\frac{1}{1-\Phi(-\mu)}\right)\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(x_i-\mu)^2) \\
&= (2\pi)^{-n/2}(1-\Phi(-\mu))^{-n}\exp(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2) \\
\log L(\mu) &= -\frac{n}{2}\log(2\pi)-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2-n\log(1-\Phi(-\mu)).
\end{aligned}$$

Therefore, we must solve

$$\frac{d}{d\mu}\log L(\mu)=\sum_{i=1}^{n}(x_i-\mu)-n\frac{\phi(-\mu)}{1-\Phi(-\mu)}=0,$$

which is equivalent (divide both sides by $n$) to solving

$$\bar{x}-\mu-\frac{\phi(-\mu)}{1-\Phi(-\mu)}=0.$$

In order to find the MME, we need to find an equation for $E(X)$ in terms of the parameter $\mu$. So compute $E(X)$ first:

$$\begin{aligned}
E(X) &= \frac{1}{1-\Phi(-\mu)} \int_0^\infty \frac{x}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} dx \\
&= \frac{1}{1-\Phi(-\mu)} \int_{-\mu}^\infty (z+\mu) e^{-z^2/2} dz \qquad \text{(change of variables } z = x-\mu) \\
&= \frac{1}{1-\Phi(-\mu)} \left( \int_{-\mu}^\infty \frac{z}{\sqrt{2\pi}} e^{-z^2/2} dz + \mu \int_{-\mu}^\infty \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \right) \\
&= \frac{1}{1-\Phi(-\mu)} \left(\left(\frac{-e^{-z^2/2}}{\sqrt{2\pi}}\right)\Big|_{-\mu}^\infty + \mu(1-\Phi(-\mu)\right) \\
&= \mu + \frac{\phi(-\mu)}{1-\Phi(-\mu)}.
\end{aligned}$$

So the MME is given by the solution to $\bar{x} = \mu + \frac{\phi(-\mu)}{1-\Phi(-\mu)}$, (and MLE=MME).

We can use the normal tables to find $\hat{\mu}$ when $\bar{x} = 1.42$, i.e. try to find a value for $\mu$ from the tables so that $\mu + \frac{\phi(-\mu)}{1-\Phi(-\mu)} = 1.42$. By inspection, we see that $\hat{\mu} = 1.2$ works!

23. (a) The probability that a patient doesn't relapse in the 6-month trial is given by

$$1 - \int_0^6 \beta^2 t e^{-\beta t} dt = (1 + 6\beta)e^{-6\beta}.$$

*Use integration by parts!*

(b) Each patient who does not relapse contributes a factor to the likelihood given by the probability in (a). Each patient whose relapse time is less that 6 months (and is therefore precisely measured contributes a factor equal to the p.d.f. $\beta^2 t_i e^{-\beta t_i}$.

This leads naturally to the likelihood

$$L(\beta) \propto \beta^{12}(1 + 6\beta)^4 e^{-37.2\beta}$$

(c) To find the MLE of $\beta$, we take the log of the likelihood:

$$l(\beta) = 12 \log \beta + 4 \log(1 + 6\beta) - 37.2\beta$$

Now differentiate and equate to zero to give

$$\frac{12}{\beta} + \frac{24}{1 + 6\beta} - 37.2 = 0.$$

On multiplying through by $-\beta(1 + 6\beta)$ we obtain

$$223.2\beta^2 - 58.8\beta - 12 = 0.$$

This is solved by $\beta = -0.13, 0.40$. Since $\beta$ must be positive the MLE is 0.40.