MATH/STAT 3460, Intermediate Statistical Theory
Winter 2014
Toby Kenney

Sample Midterm Examination
Model Solutions

This Sample Midterm has more questions than the actual midterm, in order to cover a wider range of questions.

Basic Questions

1. *The number of hours of sunlight on a given day of the year is believed to follow a Normal distribution with mean 6 and variance σ^2. Over a series of 10 years, the number of hours of sunlight observed on this day is 4.9, 6.8, 4.6, 8.7, 9.9, 2.3, 3.1, 6.4, 8.9, and 4.2. What is the maximum likelihood estimate for σ?*

   The log-likelihood function is given by $l(\sigma) = -10\log(\sigma) - \frac{\sum_{i=1}^{10}(X_i-6)^2}{2\sigma^2} = -10\log(\sigma) - \frac{60.22}{2\sigma^2}$. Differentiating this gives $S(\sigma) = \frac{60.22}{\sigma^3} - \frac{10}{\sigma} = 0$, which gives $\sigma^2 = \frac{60.22}{10} = 6.022$, so $\sigma = 2.454$.

2. *The remaining lifetime (in years) of a patient undergoing a certain kind of treatment is exponentially distributed with parameter λ. In a study which follows 10 patients for a period of 3 years, seven of the patients have lifetimes: 0.3, 0.8, 0.9, 1.4, 1.8, 2.5, and 2.9, while the remaining three patients survive to the end of the three-year period.*

   (a) Show that $\frac{7}{19.6}$ is the maximum likelihood estimate for λ.

   The log-likelihood is given by $7\log(\lambda) - (0.3+0.8+0.9+1.4+1.8+2.5+2.9+3\times 3.0)\lambda = 7\log(\lambda) - 19.6\lambda$. Setting the derivative to zero gives $\frac{7}{\lambda} - 19.6 = 0$, so $\lambda = \frac{7}{19.6}$.

   (b) Show that $[0.14, 0.73]$ is a 10% likelihood interval for λ.

   The relative likelihood of λ is

   $$R(\lambda) = \frac{\lambda^7 e^{-19.6\lambda}}{\left(\frac{7}{19.6}\right)^7 e^{-7}}$$

   This gives $R(0.14) = 0.100$ and $R(0.73) = 0.100$, so $[0.14, 0.73]$ is a 10% likelihood interval.

3. *A team of doctors wants to determine how common a certain disease is. They test 1000 individuals at random, and find that 5 of them have the disease. Use a normal approximation to find a 10% likelihood interval for the probability that an individual has the disease.*

The maximum likelihood estimate is $p = 0.005$. The log-likelihood function is $l(p) = 5\log(p) + 995\log(1-p)$. The score function is $S(p) = \frac{5}{p} - \frac{995}{(1-p)}$, and the observed information is therefore $\mathcal{I}(p) = \frac{5}{p^2} + \frac{995}{(1-p)^2}$, which evaluated at $p = 0.005$ gives $\mathcal{I}(0.005) = 201005$, so the relative log-likelihood function is approximately given by $r(p) = -\frac{201005}{2}(p-0.005)^2$. The 10% likelihood interval is given by setting this to at least $\log(0.1)$, so $(p-0.005)^2 < \frac{2\log(10)}{201005} = 0.0000229$, so $|p - 0.05| < 0.00479$, so $[0.00021, 0.00979]$ is a 10% likelihood interval for p.

4. *The number of people visiting a doctor's office on a given day is believed to be a Poisson distribution with parameter $\sqrt[3]{a^2+5}$. Over a series of 10 days, the total number of people who visit the office is 841. What is the MLE of a?*

   Let $\lambda = \sqrt[3]{a^2+5}$. The log-likelihood is $841\log(\lambda) - 10\lambda$, so the score is $\frac{841}{\lambda} - 10$, which gives the MLE is $\lambda = 84.1$. By the invariance principle, the MLE $\hat{a}$ satisfies $\sqrt[3]{\hat{a}^2+5} = 84.1$, so $\hat{a} = \sqrt{84.1^3 - 5} = 771.24$.

5. *The length of time between eruptions of a certain volcano in years is believed to follow an exponential distribution with parameter λ. The following observations are made:*

Length of time (years)	*frequency*
0–100	*23*
100–200	*36*
200–400	*45*
over 400	*26*

   Use Newton's method to find the MLE for λ. Start with an estimate of 0.003, and perform two steps.

   The likelihood of the data is

$$(1-e^{-100\lambda})^{23}(e^{-100\lambda} - e^{-200\lambda})^{36}(e^{-200\lambda} - e^{-400\lambda})^{45}e^{-26\times 400\lambda}$$

   The score function is therefore

$$S(\lambda) = 23\frac{100e^{-100\lambda}}{1-e^{-100\lambda}} + 36\frac{200e^{-200\lambda} - 100e^{-100\lambda}}{e^{-100\lambda} - e^{-200\lambda}} + 45\frac{400e^{-400\lambda} - 200e^{-200\lambda}}{e^{-200\lambda} - e^{-400\lambda}} - 10400$$

   and the observed information is

$$\mathcal{I}(\lambda) = 23\frac{10000e^{-100\lambda}}{(1-e^{-100\lambda})^2} + 36\frac{10000e^{-300\lambda}}{(e^{-100\lambda} - e^{-200\lambda})^2} + 45\frac{40000e^{-600\lambda}}{(e^{-200\lambda} - e^{-400\lambda})^2}$$

   We calculate the following steps:

λ	$S(\lambda)$	$\mathcal{I}(\lambda)$
0.003	−1786	8822828
0.002798	131.8	10172158
0.00281		

   So the maximum likelihood estimate is 0.00281.

6. *In a trial for a new drug, the probability of a response to dose d is assumed to be $1-\frac{1}{1+e^{\alpha+\beta d}}$ for some α and β. The data from a study of the drug are given in the following table:*

dose	*-1*	*0*	*1*	*2*
number	*24*	*26*	*23*	*21*
number of responses	*2*	*8*	*15*	*18*

(a) Show that $\alpha=-0.85106$ and $\beta=1.39655$ is the maximum likelihood estimate for α and β, and calculate the observed information matrix at these values.

The likelihood of the data is

$$\left(1-\frac{1}{1+e^{\alpha-\beta}}\right)^2\left(\frac{1}{1+e^{\alpha-\beta}}\right)^{22}\left(1-\frac{1}{1+e^{\alpha}}\right)^8\left(\frac{1}{1+e^{\alpha}}\right)^{18}\left(1-\frac{1}{1+e^{\alpha+\beta}}\right)^{15}\left(\frac{1}{1+e^{\alpha+\beta}}\right)^{8}\left(1-\frac{1}{1+e^{\alpha+2\beta}}\right.$$

so the score function is

$$\frac{\partial l}{\partial\alpha}=2\frac{1}{1+e^{\alpha-\beta}}-22\frac{e^{\alpha-\beta}}{1+e^{\alpha-\beta}}+8\frac{1}{1+e^{\alpha}}-18\frac{e^{\alpha}}{1+e^{\alpha}}+15\frac{1}{1+e^{\alpha+\beta}}-8\frac{e^{\alpha+\beta}}{1+e^{\alpha+\beta}}+18\frac{1}{1+e^{\alpha+2\beta}}-3\frac{e^{\alpha+2\beta}}{1+e^{\alpha+2\beta}}$$

$$\frac{\partial l}{\partial\beta}=-2\frac{1}{1+e^{\alpha-\beta}}+22\frac{e^{\alpha-\beta}}{1+e^{\alpha-\beta}}+15\frac{1}{1+e^{\alpha+\beta}}-8\frac{e^{\alpha+\beta}}{1+e^{\alpha+\beta}}+36\frac{1}{1+e^{\alpha+2\beta}}-6\frac{e^{\alpha+2\beta}}{1+e^{\alpha+2\beta}}$$

Evaluating when $\alpha=-1.7314$, $\beta=0.6899$ we get $\frac{\partial l}{\partial\alpha}=-3.625\times10^{-6}$ and $\frac{\partial l}{\partial\beta}=9.894\times10^{-9}$, so this is the maximum likelihood estimate.
The second derivatives are

$$\frac{\partial^2 l}{\partial\alpha^2}=-2\frac{e^{\alpha-\beta}}{(1+e^{\alpha-\beta})^2}-22\frac{e^{\alpha-\beta}}{(1+e^{\alpha-\beta})^2}-8\frac{e^{\alpha}}{(1+e^{\alpha})^2}-18\frac{e^{\alpha}}{(1+e^{\alpha})^2}-15\frac{e^{\alpha+\beta}}{(1+e^{\alpha+\beta})^2}-8\frac{e^{\alpha+\beta}}{(1+e^{\alpha+\beta})^2}-18\frac{e^{\alpha+2\beta}}{(1+e^{\alpha+2\beta})^2}$$

$$\frac{\partial^2 l}{\partial\alpha\partial\beta}=2\frac{e^{\alpha-\beta}}{(1+e^{\alpha-\beta})^2}+22\frac{e^{\alpha-\beta}}{(1+e^{\alpha-\beta})^2}-15\frac{e^{\alpha+\beta}}{(1+e^{\alpha+\beta})^2}-8\frac{e^{\alpha+\beta}}{(1+e^{\alpha+\beta})^2}-36\frac{e^{\alpha+2\beta}}{(1+e^{\alpha+2\beta})^2}-6\frac{e^{\alpha+2\beta}}{(1+e^{\alpha+2\beta})^2}$$

$$\frac{\partial^2 l}{\partial\beta^2}=-2\frac{e^{\alpha-\beta}}{(1+e^{\alpha-\beta})^2}-22\frac{e^{\alpha-\beta}}{(1+e^{\alpha-\beta})^2}-15\frac{e^{\alpha+\beta}}{(1+e^{\alpha+\beta})^2}-8\frac{e^{\alpha+\beta}}{(1+e^{\alpha+\beta})^2}-72\frac{e^{\alpha+2\beta}}{(1+e^{\alpha+2\beta})^2}-12\frac{e^{\alpha+2\beta}}{(1+e^{\alpha+2\beta})^2}$$

So the observed information matrix is

$$\begin{pmatrix}15.172 & 7.875\\ 7.875 & 16.631\end{pmatrix}$$

(b) Use a normal approximation to calculate a 10% likelihood region for (α,β).

The normal approximation gives

$$r(\alpha,\beta) = -\frac{15.172}{2}(\alpha+0.85106)^2 - 7.875(\alpha+0.85106)(\beta-1.39655) - \frac{16.631}{2}(\beta-1.39655)^2$$

So a 10% likelihood region is given by

$$7.586(\alpha+0.85106)^2 - 7.875(\alpha+0.85106)(\beta-1.39655) - 8.315(\beta-1.39655)^2 < 2.302585093$$

7. *We observe two samples from a Poisson distribution with parameter λ. If the true value of λ is 1.3, what is the probability that this value lies within a 10% likelihood interval?*

   If the values are X_1 and X_2, the log-likelihood function is $(X_1+X_2)\log(\lambda) - 2\lambda$, and is maximised by $\lambda = \frac{X_1+X_2}{2}$, so the relative likelihood of 1.3 is $\left(\frac{2.6}{X_1+X_2}\right)^{X_1+X_2} e^{\frac{X_1+X_2}{2}-1.3}$.

   We want to find when this is more than 0.1. We calculate this for various values of $X_1 + X_2$:

$X_1 + X_2$	$R(1.3)$
0	0.272
10	0.000057
7	0.0088
5	0.126
6	0.036

   So we see that 1.3 is in the 10% likelihood interval, whenever $X_1+X_2 \leqslant 5$. X_1+X_2 has a Poisson distribution with parameter 2.6, so the probability that it is at most 5 is given by $e^{-2.6}\left(1+2.6+\frac{2.6^2}{2}+\frac{2.6^3}{6}+\frac{2.6^4}{24}+\frac{2.6^5}{120}\right) = 0.95$.

8. *Let $X_1,\ldots,X_{100}$ be samples from a normal distribution with mean 0 and variance σ^2. If $\sum {X_i}^2 = 180$, show that using a chi-square approximation, $[1.175, 1.551]$ is a 95% confidence interval for σ.*

   By the chi-squared approximation, the 95% confidence interval is a 14.65% likelihood interval, so the relative log-likelihood is at least -1.92072941. The log-likelihood is given by $-100\log(\sigma) - \frac{\sum {X_i}^2}{2\sigma^2}$, and is therefore maximised by $\sigma^2 = \frac{180}{100} = 1.8$. The relative log-likelihood is given by $-100\log(\sigma) + 50\log(1.8) + 50 - \frac{90}{\sigma^2}$. We evaluate the relative log-likelihood at $\sigma = 1.175$ and $\sigma = 1.551$, and get -1.925 and -1.913 respectively, so this is a 95% confidence interval.

9. *The lifetime of a particle (in nanoseconds) is exponentially distributed with parameter λ. However, checking whether the particle has decayed will influence the lifetime. A scientist plans an experiment to keep 1000 particles for a fixed time period T, and after that time, to measure how many particles have decayed. Show that $T = \frac{1.5936}{\lambda}$ gives the highest expected information about λ.*

Let N be the number of particles which decay. The log-likelihood of λ is $l(\lambda) = N\log(1-e^{-\lambda T}) - (1000-N)\lambda T$. The score function is

$$S(\lambda) = NT\frac{\lambda e^{-\lambda T}}{(1-e^{-\lambda T})} - (1000-N)T$$

and the information function is

$$\mathcal{I}(\lambda) = NT^2\frac{e^{-\lambda T}}{(1-e^{-\lambda T})^2}$$

Since the expected value of N is $1000(1-e^{-\lambda T})$, we get

$$\mathcal{I}_E(\lambda) = 1000T^2\frac{e^{-\lambda T}}{(1-e^{-\lambda T})}$$

This is maximised when $T^2\left(\frac{1}{(1-e^{-\lambda T})}-1\right)$ is maximised, which happens when

$$2T\left(\frac{1}{(1-e^{-\lambda T})}-1\right) - T^2\frac{\lambda e^{-\lambda T}}{(1-e^{-\lambda T})^2} = 0$$

We rearrange this to get:

$$\begin{aligned}
2(1-e^{-\lambda T}) - 2(1-e^{-\lambda T})^2 - \lambda T e^{-\lambda T} &= 0\\
2e^{-\lambda T}(1-e^{-\lambda T}) - \lambda T e^{-\lambda T} &= 0\\
2(1-e^{-\lambda T}) - \lambda T &= 0\\
\lambda T &= 2(1-e^{-\lambda T})
\end{aligned}$$

We test $\lambda T = 1.5936$ and get $2(1-e^{-\lambda T}) = 1.5936$, so this is indeed the value of T which gives most information about λ.

10. *Let $X_1, \ldots, X_{40}$ be samples from a uniform distribution on $[0, a]$. The MLE for a is $\max(X_1, \ldots, X_{40})$. What is the bias of this estimate? [Hint: the expected value of a positive random variable with cumulative distribution function $F(x)$ is $\int_0^\infty (1-F(x))dx$.]*

    Let $M = \max(X_1, \ldots, X_{40})$, then the cumulative distribution of M is $F(x) = \left(\frac{x}{a}\right)^{40}$ for $0 \leqslant x \leqslant a$. Therefore, $\mathbb{E}(M) = \int_0^a (1-\left(\frac{x}{a}\right)^{40}dx = \left[x - \frac{x^{41}}{41a^{40}}\right]_0^a = a - \frac{a}{41}$, so the bias is $-\frac{a}{41}$.