Problem Sheet 4

Remember: when online, you can access the Statistics 1 data sets from an **R** console by typing
`load(url("http://www.stats.bris.ac.uk/%7Emapjg/Teach/Stats1/stats1.RData"))`

1. Find an expression for the maximum likelihood estimate of θ in terms of the observed values $x_1, \ldots, x_n$ of a random sample of size n from a Poisson(θ) distribution with probability mass function
$$p(x;\theta) = \begin{cases} e^{-\theta}\theta^x/x! & x = 0, 1, 2, \ldots \\ 0 & \text{otherwise,} \end{cases}$$
where $\theta > 0$ is an unknown parameter.

*2. Find an expression for the maximum likelihood estimate of θ in terms of the observed values $x_1, \ldots, x_n$ of a random sample of size n from a continuous distribution with probability density function
$$f_X(x;\theta) = \begin{cases} \theta x^{\theta-1} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$
where $\theta > 0$ is an unknown parameter.

In a random sample of size $n = 5$ from this distribution, the observed values of $x_1, x_2, \ldots, x_5$ were $0.07, 0.29, 0.95, 0.51$ and 0.50 respectively. Compute the value of the maximum likelihood estimate of θ for this set of observations.

3. (a) Find an expression for the maximum likelihood estimate of θ in terms of the observed values $x_1, \ldots, x_n$ of a random sample of size n from a Binomial(K, θ) distribution [so the Binomial parameter is K and the sample size is n] with probability mass function
$$p(x;\theta) = \begin{cases} \begin{pmatrix} K \\ x \end{pmatrix} \theta^x (1-\theta)^{K-x} & x = 0, 1, \ldots, K \\ 0 & \text{otherwise,} \end{cases}$$
where K is known and where θ is an unknown parameter such that $0 < \theta < 1$.

(b) The award of a safety qualification for a particular type of outdoor activity is partly based on success in a written examination with 10 multiple choice questions. A score of at least 9 out of 10 is required to pass the examination. In seven mock examinations a trainee has obtained scores of 9, 8, 10, 5, 8, 10 and 7. Assuming that the trainee has independent probability θ of correctly answering each question in the examination, and assuming these seven scores are a random sample of size seven from a Binomial$(10, \theta)$ distribution with the same value of θ, find the maximum likelihood estimate of the probability that the trainee will pass the examination.

(c) Can you see a connection between the answer to (a) and the problem described in handout section 3.1: data 4, 5 and 1 from a geometric distribution leading to a mle of $3/10 = 0.3$?

*4. Consider the following data, recording the failure time, in hours, for a batch of 25 lamps. The data are contained in the Statistics 1 data set `lamp`.

```
5.5  3.8  8.0  7.8  9.3  4.7  4.0  0.3  4.6  0.6  7.9  1.8  4.0
0.7  4.0  1.6  2.6  0.7  0.2  3.1  1.0  3.4  3.7 10.8  1.2
```

Assuming an Exponential distribution with parameter θ is an appropriate model, find the maximum likelihood estimate of θ based on the above data. Hence find the maximum likelihood estimates of:

(a) the median of the distribution of the lifetimes of lamps in the population;

(b) the probability of a randomly chosen lamp surviving beyond 10 hours.

Compare these to appropriate simple estimates calculated directly from the data, without assuming an Exponential distribution.

5. Let $X_1, \ldots, X_n$ be a random sample of size n from a $N(\mu_X, \sigma^2)$ distribution and let $Y_1, \ldots, Y_m$ be a random sample of size m from a $N(\mu_Y, \sigma^2)$ distribution, and assume the samples are independent of each other. Note that the means of the two distributions are assumed to be possibly different, but the variances are assumed to be the same.

(a) Since all $n + m$ random variables are independent, their joint probability density function is just the product of all the $n + m$ individual (marginal) probability density functions. Show that the loglikelihood function for μ_X, μ_Y and σ^2 based on all $n + m$ observations is given by $\sum_{i=1}^{n} \log f_X(x_i; \mu_X, \sigma^2) + \sum_{j=1}^{m} \log f_Y(y_j; \mu_Y, \sigma^2)$.

(b) Hence, by adapting the method used in section 3.9 of the lectures, explain why the likehihood equations for this problem are

$$\begin{aligned} \frac{\partial}{\partial \mu_X} l(\mu_X, \mu_Y, \sigma^2) &= \frac{\sum_{i=1}^{n} x_i - n\mu_X}{\sigma^2} = 0 \\ \frac{\partial}{\partial \mu_Y} l(\mu_X, \mu_Y, \sigma^2) &= \frac{\sum_{j=1}^{m} y_j - m\mu_Y}{\sigma^2} = 0 \\ \text{and} \quad \frac{\partial}{\partial \sigma} l(\mu_X, \mu_Y, \sigma^2) &= -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i - \mu_X)^2}{\sigma^3} - \frac{m}{\sigma} + \frac{\sum_{j=1}^{m}(y_j - \mu_Y)^2}{\sigma^3} = 0 \end{aligned}$$

(c) Hence find expressions in terms of $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ for the joint maximum likelihood estimates of μ_X, μ_Y and σ^2 from the combined sample.

(d) In a study of the relative size of secretarial starting salaries in the public and private sector (a few years ago!), 9 private sector posts and 10 public sector posts were chosen at random from jobs advertised on the web. The table below shows the advertised stating salaries (in £1,000). You may assume that the population variances are the same in the private and public sectors. Assuming the model above is appropriate, compute the maximum likelihood estimates of μ_X (Private sector), μ_Y (Public sector) and the common variance σ^2. The data are contained in the Statistics 1 data sets `secretaries`.

Private sector	12.1	13.4	11.3	10.6	9.7	12.5	9.6	13.6	11.2	
Public sector	9.3	8.5	8.2	13.1	8.8	11.9	10.1	9.8	12.2	10.4

*6. The following example shows that it is possible to make sensible estimates on a sensitive subject (Tax evasion, Aids testing etc.) while maintaining individual confidentiality.

Assume n subjects are each given an envelope. Half the envelopes contains the instructions "Tick box 1 if you have ever cheated on your tax return and tick box 0 otherwise"; the others contain the instructions "Toss a coin and tick box 1 if it falls heads and box 0 if it falls tails". The two sets of instructions are allocated to the envelopes at random and only the subject knows which set of instructions applied to him (or her). Assume that all subjects follow the instructions in their envelope honestly and correctly.

(a) Assume the probability that any given subject cheated on their tax return is τ and let θ denote the probability that a randomly chosen subject, following the procedure above, will tick box 1. Express τ in terms of θ. What are the possible values for θ?

(b) Assume 8 out of a sample of 20 subjects ticked box 1; find the maximum likelihood estimate of θ based on these data. *Hence* estimate τ.