Answer four questions. All questions carry equal marks. All books and notes may be used. Calculators may be used.

1. Let $X_1, \ldots, X_n$ be an independent and identically distributed random sample from $X \sim N(\mu, \sigma^2)$. We would like to use

$$T_n = \frac{1}{n} \sqrt{\frac{n}{2}} \sum_{i=1}^{n} |X_i - \mu|$$

as an estimator for $\sigma$.

(a) Show that when $Y$ is a continuous random variable $f_{|Y|}(y) = f_Y(y) + f_Y(-y)$, with $y > 0$.

(b) Derive $E|Y|$ and $Var|Y|$ when $Y \sim N(0,1)$.

(c) Is $T_n$ unbiased? Is $T_n$ consistent? Explain your reasoning in both cases. Calculate the efficiency of $T_n$. Discuss the estimation of $\sigma$ as fully as you can.

2. Assume that the speed of a molecule with mass $m$ is a random variable $X$ with density
\[ f_X(x) = \sqrt{\frac{2}{\pi}} x^2 e^{-(x^2/2)} \]

where \( x > 0 \). The kinetic energy of the molecule is \( Y = \frac{1}{2} mx^2 \). Derive the density of \( Y \). Suppose we have a random sample \( Y_1, \ldots, Y_n \) from \( Y \). Discuss the estimation of the mass \( m \) of the molecule as fully as possible.

3. Characteristics of individuals in a population are controlled by genes. These entities are capable of assuming one from a number of possible forms, called alleles. An example may be found in the blood group of an individual. This is controlled by a gene with three alleles, denoted as A, B and O. Alleles always occur in pairs and hence, there are six blood groups, which we may denote as AA, AO, BB, BO, AB and OO. Combinations such as AO and OA are regarded as equivalent and similarly for BO and OB as well as AB and BA. The above six genotypes are not directly observable. The allele O is recessive and hence, there are only four observable blood groups: A, B, AB and O. This is because AA and AO reduce to A, BB and BO to B while we identify OO with O.

An experiment is carried out which collects a sample of 345 individuals and verifies the blood group of each individual. The number of people with blood group A is 150, 29 people are in blood group B, 6 in blood group AB and 160 have blood type O. Let \( p, q \) and \( r \) denote the proportions in which the three alleles A, B, and O occur in the population.

(a) Describe in as much detail as possible how we can apply likelihood theory in this example to estimate the proportions \( p, q \) and \( r \) from the data. If you conclude that Newton-Raphson optimisation will be required to solve the problem then describe in words a proposal of how the computations may be implemented. Do not carry out these computations themselves.

Suppose that the maximum likelihood solutions to the above problem are obtained for \( p = 0.26076, \) \( q = 0.05226 \) and \( r = 0.68698 \). The observed information matrix at the optimum was
The inverse of the observed information matrix is

\[
10^{-4} \begin{pmatrix} 3.2930 & -0.1958 \\ -0.1958 & 0.7377 \end{pmatrix}.
\]

(b) Carry out a likelihood-based test of hypothesis to assess whether \( p = 0.25, \ q = 0.05 \) and \( r = 0.70 \). Derive at least two likelihood-based confidence intervals and compare the results. Comment on the validity of your results.

4. The data in the table below was obtained from a carefully controlled study on the effect of the rate and volume of air inspired by humans on the occurrence of a transient condition known as vaso-constriction in the skin of the fingers. For each individual, the volume and rate of inspiration is given, together with the response which is an indicator with a value of one when constriction occurs and zero otherwise. There were 39 people in the study.

<table>
<thead>
<tr>
<th>vol</th>
<th>rate</th>
<th>resp</th>
<th>vol</th>
<th>rate</th>
<th>resp</th>
<th>vol</th>
<th>rate</th>
<th>resp</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.70</td>
<td>0.82</td>
<td>1.00</td>
<td>3.50</td>
<td>1.09</td>
<td>1.00</td>
<td>1.25</td>
<td>2.50</td>
<td>1.00</td>
</tr>
<tr>
<td>0.80</td>
<td>3.20</td>
<td>1.00</td>
<td>0.70</td>
<td>3.50</td>
<td>1.00</td>
<td>0.60</td>
<td>0.75</td>
<td>0.00</td>
</tr>
<tr>
<td>0.90</td>
<td>0.75</td>
<td>0.00</td>
<td>0.90</td>
<td>0.45</td>
<td>0.00</td>
<td>0.80</td>
<td>0.57</td>
<td>0.00</td>
</tr>
<tr>
<td>0.60</td>
<td>3.00</td>
<td>0.00</td>
<td>1.40</td>
<td>2.33</td>
<td>1.00</td>
<td>0.75</td>
<td>3.75</td>
<td>1.00</td>
</tr>
<tr>
<td>0.20</td>
<td>1.60</td>
<td>1.00</td>
<td>0.85</td>
<td>1.42</td>
<td>1.00</td>
<td>1.70</td>
<td>1.06</td>
<td>0.00</td>
</tr>
<tr>
<td>0.40</td>
<td>2.00</td>
<td>0.00</td>
<td>0.95</td>
<td>1.36</td>
<td>0.00</td>
<td>1.35</td>
<td>1.35</td>
<td>0.00</td>
</tr>
<tr>
<td>1.60</td>
<td>1.78</td>
<td>1.00</td>
<td>0.60</td>
<td>1.50</td>
<td>0.00</td>
<td>1.80</td>
<td>1.50</td>
<td>1.00</td>
</tr>
<tr>
<td>1.90</td>
<td>0.95</td>
<td>1.00</td>
<td>1.60</td>
<td>0.40</td>
<td>0.00</td>
<td>2.70</td>
<td>0.75</td>
<td>1.00</td>
</tr>
<tr>
<td>1.10</td>
<td>1.83</td>
<td>0.00</td>
<td>1.10</td>
<td>2.20</td>
<td>1.00</td>
<td>1.20</td>
<td>2.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.95</td>
<td>1.90</td>
<td>0.00</td>
<td>0.75</td>
<td>1.90</td>
<td>0.00</td>
<td>1.30</td>
<td>1.63</td>
<td>1.00</td>
</tr>
</tbody>
</table>

A statistical model was fitted to describe the dependence of the response on the rate and volume of inspiration. Some output from the analysis is shown overleaf.
> model_glm(response ~ rate + vol, family = binomial(link = logit), data = Vaso)

Call: glm(formula = response ~ rate + vol, family = binomial(link = logit), data = Vaso)

Deviance Residuals:
Min       1Q   Median       3Q      Max
-1.45085  -0.7575138  0.0373873  0.5280191  2.27158

Coefficients:
Value Std. Error t value
(Intercept) -9.114104 3.1523248 -2.891232
rate  2.470525 0.8730222  2.829853
vol  3.876099 1.4332810  2.704354

(Dispersion Parameter for Binomial family taken to be 1)

Null Deviance: 54.03984 on 38 degrees of freedom

Residual Deviance: 31.26918 on 36 degrees of freedom

Number of Fisher Scoring Iterations: 5

Correlation of Coefficients:

(Intercept) rate vol
rate -0.9230832
vol -0.9436642 0.7789512

> anova(model)

Analysis of Deviance Table

Binomial model

Response: response
Terms added sequentially (first to last)

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Deviance</th>
<th>Resid. Df</th>
<th>Resid. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rate</td>
<td>1</td>
<td>3.51100</td>
<td>37</td>
<td>50.52884</td>
</tr>
<tr>
<td>vol</td>
<td>1</td>
<td>19.25965</td>
<td>36</td>
<td>31.26918</td>
</tr>
</tbody>
</table>

The following command produces the observed inverse Fisher information matrix at the computed optimum.

```r
> summary(model)$cov.unscaled
          (Intercept)       rate       vol
(Intercept)    10.050525 -2.5689787 -4.3134688
rate            -2.568979  0.7697767  0.9869073
vol            -4.313469  0.9869073  2.0768246
```

(a) Describe in detail the model which was fitted to the data. Without addressing the issue of the fit of the model to the data, do you believe that a model with such a structure could provide an adequate description of the variability of the response? Explain your answer. Implement a statistical test of hypothesis to test that rate and volume of inspiration do not affect the response observed. Justify your procedure. Implement a statistical test of hypothesis to assess whether volume may be removed as a predictor from a model with both rate and volume as predictor variables. Justify your procedure.

The statistician decided to drop rate as a predictor from the model.

```r
> model2_glm(response ~ vol, family = binomial(link = logit), data = Vaso,
>              anova(model2))
```

Analysis of Deviance Table

Binomial model

Response: response

Terms added sequentially (first to last)

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Deviance</th>
<th>Resid. Df</th>
<th>Resid. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vol</td>
<td>1</td>
<td>7.050458</td>
<td>37</td>
<td>46.98938</td>
</tr>
</tbody>
</table>
(b) Apparently, the removal of rate has caused an increase of 15.72 in the residual deviance. Is this not in contradiction to the above output? Should we remove rate from the model? Explain your answer.

(c) Finally, the above model with both rate and volume as predictors (full model) has a residual deviance of 31.27 with 36 degrees of freedom. Suppose we were to fit a maximal model which fits a parameter for every observation. Such a model would obviously fit the data exactly and have a residual deviance of zero. In effect, this maximum model is equivalent to fitting 39 separate models, each for a sample of size one. So, if we were to apply the results from large-sample theory for maximum likelihood estimation, then we would treat the residual deviance of the full model as approximately chi-squared distributed with 36 degrees of freedom. We could then use this as a test of goodness-of-fit of the model. Comment and explain your answer.

5. (a) You are designing a new telephone exchange. From long experience it is known that, over one month, the maximum number of lines (in thousands) used in the area that the exchange covers is Poisson distributed with an unknown mean \( \lambda \) (i.e. the average maximum is \( \lambda \times 1000 \)).

   (i) In talking to the telephone engineer, you find out that she expects the average maximum number of lines used to be about 7000, but that she is quite unsure about this and feels it might be anywhere between 5000 and 9000. Form a gamma prior distribution on \( \lambda \) that reflects the opinion of the engineer.

   (ii) The telephone company decides to collect some data on the maximum number of lines used in a month in the area that the exchange serves. It observes the following maximum number of lines used (in thousands): 8.8, 10.2, 6.4, 6.7, 9.5, 8.4. Calculate the posterior distribution of \( \lambda \) on the basis of these data and the prior you specified in part (i).

   Contd.
(iii) What is the posterior mean of $\lambda$?

(iv) Compute the posterior predictive distribution for the maximum number of lines that will be used.

(v) The company will use the posterior predictive distribution of $\lambda$ to decide how many lines to put into the exchange. Let $X$ be the number of lines put into the exchange and $Y$ be the maximum number used per month. The company feels that it loses more if it does not make a big enough exchange than if it makes one too big; thus, it defines its loss function

$$L(X,Y) = \begin{cases} c_1(X - Y), & \text{if } X \geq Y, \\ c_2(Y - X), & \text{if } X < Y, \end{cases}$$

where $c_2 > c_1$ to reflect the company's feelings about the difference in the two types of loss. Show that the expected loss associated with creating an exchange of $X$ lines is:

$$(c_1 + c_2) \left( X P(Y \leq X | D) - \sum_{y=0}^{X} y P(Y = y | D) \right) + c_2 (E(Y | D) - X),$$

where $D$ denotes the data.

(b) "Bayesian statistics is of no use because it cannot be implemented in practical problems". Briefly discuss.

6. (a) I have four coins in a box. Coins A and B are unbiased (that is, $P(\text{head}) = P(\text{tail}) = 0.5$), coin C is biased (with $P(\text{head}) = 0.25$) and coin D has 2 tails.

(i) I pick a coin randomly from the box and then flip it. What is the probability I get a head?
(ii) I put the coin back in the box and then pick randomly again. I flip the coin and it lands heads. What is the probability that I picked coin A?

(iii) Repeat the calculation that you have done in (ii), working out the probability for each of the other three coins given a head was obtained.

(iv) Now, I take the coin that I took out in part (ii) and that landed heads, and flip it again. Now what is the probability that it lands heads? You may assume that, given the coin that has been picked, flips are independent.

(b) The strength of a new type of steel foundation pile is known, from long experience, to be normally distributed with an unknown mean $\mu$ and a known variance $\sigma^2$. You have been asked to work with a structural engineer to determine the mean strength.

After consultation with the civil engineer, you decide that there is little knowledge on the mean strength, and so will use a uniform prior on the range $[200, 3000]$ tonnes.

(i) Give a reason why the conjugate prior for the mean of a normal distribution may not be appropriate to describe mean strength.

(ii) Data are available on the strength of 10 specimen foundations; the sum of the strengths in the data is 5000 tonnes and the sum of the squares is 260,000 tonnes squared. Write down the functional form of the posterior distribution for $\mu$ given the uniform prior and these data.

(iii) Devise a Monte Carlo scheme, using importance sampling, to generate samples from the posterior mean of $\mu$.

(iv) Devise a Monte Carlo scheme for evaluating the normalising constant for the posterior distribution of $\mu$. 

© UNIVERSITY OF DUBLIN 1998.