**1.** Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution with p.d.f.

$$f(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{otherwise}. \end{cases}$$

Show that, if $X_{(r)}$ is the $r^{th}$ order statistic, then

$$E\left(X_{(r)}\right) = \frac{r}{n+1}, \quad V\left(X_{(r)}\right) = \frac{r}{(n+1)(n+2)}\left(1 - \frac{r}{n+1}\right).$$

Define the median of the random sample, distinguishing between the two cases $n$ odd and $n$ even. Show that the median has expected value $\frac{1}{2}$ if the random sample is drawn from a uniform distribution on $(0,1)$. Find its variance in the particular case when $n$ is odd. What is the expected value of the median if the random sample is drawn from a uniform distribution on $(a,b)$?

**2.** Let $X$ be a continuous random variable with cumulative distribution function $F$ which is strictly increasing. If $Y = F(X)$, show that $Y$ is uniformly distributed on the interval $(0,1)$.

The *Weibull distribution* with index $\alpha$ has c.d.f.

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - \exp\left(-\left(x/\lambda\right)^{\alpha}\right), & x \geq 0, \end{cases} \qquad \lambda, \alpha > 0.$$

It is typically used in industrial reliability studies in situations where failure of a system comprising many similar components occurs when the weakest component fails; it is also used in modelling survival.

Explain why a probability plot for the Weibull distribution may be based upon plotting the logarithm of the $r^{th}$ order statistic against $\log\left[-\log\left(1 - \frac{r}{n+1}\right)\right]$ and give the slope and intercept of such a plot.

**3.** Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables from a distribution with p.d.f.

$$f(x, \theta_1, \theta_2) = \frac{1}{\theta_2}\exp\left(-\frac{(x-\theta_1)}{\theta_2}\right), \quad x > \theta_1, \quad (\theta_1\theta_2) \in \mathbb{R} \times \mathbb{R}^+.$$

Find maximum likelihood estimators of $\theta_1$ and $\theta_2$.

**4.** Find the likelihood for $\theta$ based on $X$, a binomially distributed random variable with probability function

$$p_X(x) = \binom{m}{x}\theta^x(1-\theta)^{m-x}, \quad x = 0, 1, \ldots m.$$

Show that the maximum likelihood estimate of $\theta$ is $\widehat{\theta} = x/m$ and that the observed and expected information both have maximum likelihood estimates equal to $m \Big/ \left(\widehat{\theta}(1-\widehat{\theta})\right)$.

**5.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a gamma distribution with probability density function

$$f_X(x; \lambda, \alpha) = \Gamma(\alpha)^{-1} \lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}, \quad x > 0, \quad \lambda, \alpha > 0,$$

so that $E(X) = \alpha / \lambda$. Show that the expected information matrix for $\lambda$ and $\alpha$ is

$$n \begin{pmatrix} \alpha / \lambda^2 & -1/\lambda \\ -1/\lambda & \dfrac{\partial^2 \log \Gamma(\alpha)}{\partial \alpha^2} \end{pmatrix}.$$

**6.** (Harder question.) According to genetic linkage theory, observed frequencies of four phenotypes resulting from crossing tomato plants are in the ratio $\frac{9}{16} + \theta : \frac{3}{16} - \theta : \frac{3}{16} - \theta : \frac{1}{16} + \theta$. In 1931, J.W. MacArthur reported the following frequencies:

| **Phenotype** | **Observed frequency** |
|---|---|
| Tall, cut-leaf | 926 |
| Tall, potato-leaf | 288 |
| Dwarf, cut-leaf | 293 |
| Dwarf, potato-leaf | 104 |
| **Total** | 1611 |

Write down the likelihood of $\theta$ given these observations. Find the maximum likelihood estimate of $\theta$, use it to calculate expected frequencies for the four phenotypes and compare them with the observed frequencies. Does genetic linkage theory look plausible?

## Optional Computer Exercises

**1.** The data set *memories.txt*, which you will find on the website, contains recall times (in seconds) of pleasant and unpleasant memories [Dunn, G. and Master, D. (1982), *Psychological Medicine*, **12**]. Download the data set into one of your directories (say, `c:\Data`), read it into R and get it ready for use with the commands

```
> memories <- read.table("c:/Data/memories.txt", header = TRUE)
> attach(memories)
```

If you enter `memories` you will be able to look at the data.

Try looking at boxplots of these data with the command `boxplot(memories)`. What conclusions can you draw?

The apparent skewness revealed by the boxplots along with the fact that these are measurements of waiting times might lead you to suspect that the data are exponentially distributed. Order the vector `Pleasant` and assign it to `y`, construct a vector `x` and plot `y` against `x` with the sequence of commands below.

```
y <- sort(Pleasant)
x <- -log(1 - c(1:20)/21)
```

```
plot(x,y)
```

Explain why this plot allows you to assess whether the pleasant memories are exponentially distributed. What do you conclude? Repeat this procedure for the unpleasant memories and state your conclusions about the adequacy of an exponential model.

**2.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a truncated Poisson distribution with probability mass function

$$p(x;\theta) = \frac{\theta^x e^{-\theta}}{x!(1-e^{-\theta})}, \quad x = 1, 2, \ldots, \quad \theta > 0.$$

Find the likelihood and obtain the score function and the observed information.

On a spring afternoon in Portland, Oregon, data on the sizes of different groups observed in public places were collected by Coleman and James:

| Group size: | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Frequency: | 1486 | 694 | 195 | 37 | 10 | 1 |

A possible model for these data is a truncated Poisson distribution. Show that the log-likelihood is maximised by the value of $\theta$ which maximises

$$3663 \log\theta - 2423\theta - 2423 \log(1 - e^{-\theta}).$$

Using the sample mean as a starting value, apply the Newton-Raphson method to obtain the maximum likelihood estimate of $\theta$ to 3 places of decimals. Obtain an approximate 95% confidence interval for $\theta$.