MTH6134 Statistical Modelling II
Exercises

Autumn 2023

Exercises built upon a list provided by Dr S Coad (formerly of QMUL).

1. Suppose that $Y_i \sim \text{Bin}(r_i, \pi)$ for $i = 1, 2, \ldots, n$, all independent, where the r_i are known.

   1. Write down the likelihood for the data $y_1, \ldots, y_n$.

   Solution: The likelihood is

$$\begin{aligned} L(\pi; \mathbf{y}) &= \prod_{i=1}^{n} \binom{r_i}{y_i} \pi^{y_i} (1-\pi)^{r_i - y_i} \\ &= \left\{ \prod_{i=1}^{n} \binom{r_i}{y_i} \right\} \pi^{\sum_{i=1}^{n} y_i} (1-\pi)^{\sum_{i=1}^{n}(r_i - y_i)}. \end{aligned}$$

   2. Find the maximum likelihood estimator $\hat{\pi}$ of π.

   Solution: The log-likelihood is

$$\ell(\pi; \mathbf{y}) = \sum_{i=1}^{n} \log \binom{r_i}{y_i} + \sum_{i=1}^{n} y_i \log \pi + \sum_{i=1}^{n} (r_i - y_i) \log(1-\pi).$$

   Thus, we have

$$\frac{d\ell}{d\pi} = \frac{\sum_{i=1}^{n} y_i}{\pi} - \frac{\sum_{i=1}^{n}(r_i - y_i)}{1-\pi}.$$

   Setting this derivative to zero, we obtain

$$\sum_{i=1}^{n} y_i (1-\hat{\pi}) - \sum_{i=1}^{n} (r_i - y_i)\hat{\pi} = 0,$$

   which yields the maximum likelihood estimate

$$\hat{\pi} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} r_i}.$$

3. Prove that $\hat{\pi}$ is an unbiased estimator of π.

Solution: We can write

$$\begin{aligned} E(\hat{\pi}) &= \frac{1}{\sum_{i=1}^{n} r_i} E\left(\sum_{i=1}^{n} Y_i\right) \\ &= \frac{1}{\sum_{i=1}^{n} r_i} \sum_{i=1}^{n} E(Y_i) \\ &= \frac{1}{\sum_{i=1}^{n} r_i} \sum_{i=1}^{n} r_i \pi = \pi, \end{aligned}$$

and so $\hat{\pi}$ is an unbiased estimator of π.

2. Suppose you have the following binomial data from a single binomial sample: $r = 15, y = 7$.

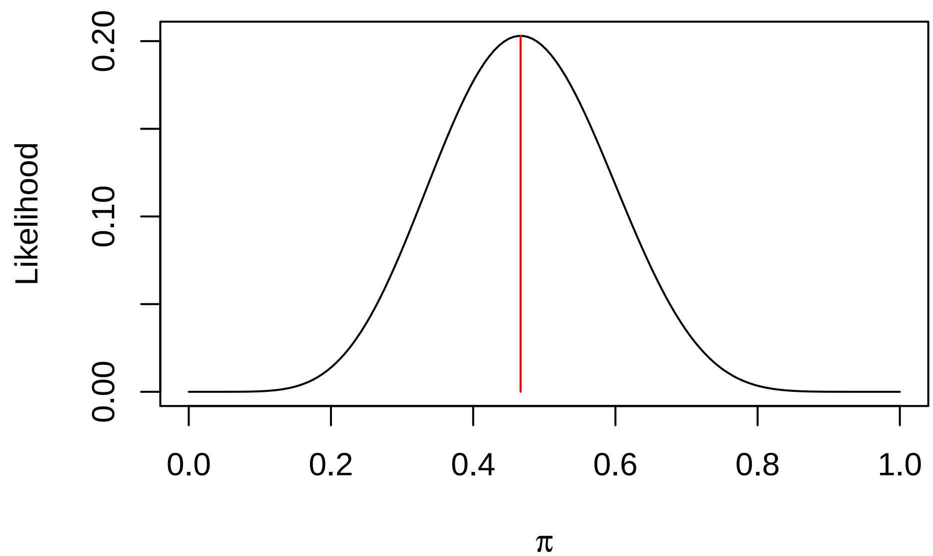
1. Write down the likelihood for the data y.
2. Find the maximum likelihood estimator $\hat{\pi}$ of π.
3. Using `R`, make a plot of the likelihood function $L(\pi)$. Examine and describe this function.

Solution: Here is the the plot of the likelihood. The line indicates the position of maximum likelihood estimate $\hat{\pi}$ (x-coordinate), and the line joins with the value $L(\hat{\pi})$ (at the top of the curve). Recall that for the code below to work, you need to define quantities, e.g. `ri<-15` and `yi<-7`.

```
par(mar=c(4,4,1,1)); (phat<-yi/ri)

## [1] 0.4666667

L<-function(p) dbinom(x=yi,size=ri,prob=p)
curve(expr=L,from=0,to=1,n=150,xlab=expression(pi),ylab="Likelihood")
lines(phat*c(1,1),c(0,L(p=phat)),col="red")
```



The assymmetry of $L(\pi)$ indicates that higher values of π than the mle are more likely than smaller values of π, for the data given.

4. Consider the following binomial sample: $r = 105, y = 49$. Repeat the computation of the likelihood $L(\pi)$, the maximum likelihood estimate $\hat{\pi}$ and the plot of $L(\pi)$. Compare the results with those of the original data and comment.

3. Consider the following binomial data pairs (r, y): $(60, 19), (70, 25), (30, 15), (40, 14), (20, 9)$.

   1. Repeat the computations of steps 1-3 of the previous question (problem 2). In this case, consider and analyze each data pair separately.
   2. Analyze the data jointly, using the result of the problem 1.
   3. Compare the results of the two analyses. Are the estimates that you obtained related?

4. Suppose that $Y_i \sim \text{Poisson}(\mu)$ for $i = 1, 2, \ldots, n$, all independent.

   1. Write down the likelihood for the data $y_1, \ldots, y_n$.

Solution: The likelihood is

$$L(\mu; \mathbf{y}) = \prod_{i=1}^{n} \frac{\mu^{y_i} e^{-\mu}}{y_i!} = \frac{\mu^{\sum_{i=1}^{n} y_i} e^{-n\mu}}{\prod_{i=1}^{n} y_i!}.$$

   2. Find the maximum likelihood estimator $\hat{\mu}$ of μ.

Solution: The log-likelihood is

$$\ell(\mu; \mathbf{y}) = \sum_{i=1}^{n} y_i \log \mu - n\mu - \sum_{i=1}^{n} \log(y_i!).$$

Thus, we have

$$\frac{d\ell}{d\mu} = \frac{\sum_{i=1}^n y_i}{\mu} - n.$$

Setting this derivative to zero, we obtain

$$\sum_{i=1}^n y_i - n\hat{\mu} = 0,$$

which yields the maximum likelihood estimate

$$\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n} = \overline{y}.$$

3. Prove that $\hat{\mu}$ is an unbiased estimator of μ.

Solution: We can write

$$\begin{aligned} E(\hat{\mu}) &= \frac{1}{n} E\left(\sum_{i=1}^n Y_i\right) = \frac{1}{n}\sum_{i=1}^n E(Y_i) \\ &= \frac{1}{n} n\mu = \mu, \end{aligned}$$

and so $\hat{\mu}$ is an unbiased estimator of μ.

5. The following count data 5, 1, 3, 5, 5, 4, 3, 2, are assumed to be a series of independent realizations of Poisson(μ).

   1. Write down the likelihood for the data $y_1, \ldots, y_n$.
   2. Find the maximum likelihood estimator $\hat{\mu}$ of μ.
   3. Plot the likelihood function $L(\mu)$ with `R`. Examine and describe this function.

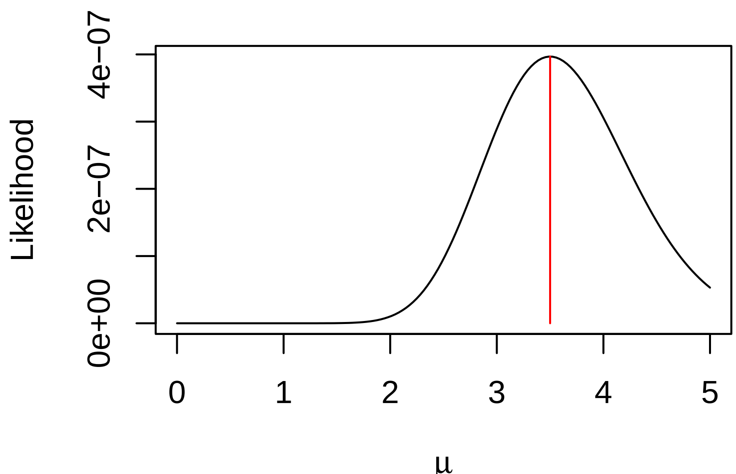
   Solution: Here I give the `R` code for the likelihood. Remember you need to define the variable `y<-c(5, 1, 3, 5, 5, 4, 3, 2)`.

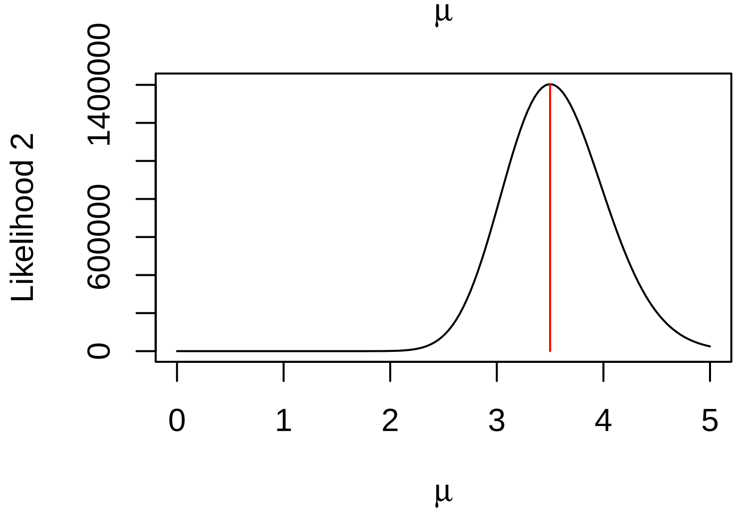
```
Lp<-function(mu){
  res<-1
  for(i in y) res<-res*dpois(x=i,lambda=mu)
  return(res)
}
```

   4. Now suppose that you have a sample of Poisson data with the same sample value $\bar{y}$ as with the data above, but with $n = 16$. Redo the plot of $L(\mu)$, compare with the first plot and comment.

Solution:
Note that if the mean is to be the same as before ($\bar{y} = 3.5$), this implies that $\sum_{i=1}^{n} y_i = 56$ and we need to define these quantities for the code below, e.g. `s<-56` for the sum and `n<-16`. I give the two plots so that they can be compared. The line indicates the location of the maximum likelihood estimate $\hat{\mu}$.

```
par(mar=c(4,4,1,1),mfrow=c(2,1))
curve(expr = Lp,from=0,to=5,n=150,xlab=expression(mu),
      ylab="Likelihood")
lines(x=mean(y)*c(1,1),c(0,Lp(mu=mean(y))),col="red")
Lp2<-function(mu)  mu^s*exp(-n*mu)
## the likelihood ignores the constant factor! <- explain why.
curve(expr = Lp2,from=0,to=5,n=150,xlab=expression(mu),
      ylab="Likelihood 2")
lines(x=mean(y)*c(1,1),c(0,Lp2(mu=mean(y))),col="red")
```





What are the similarities and differences between the likelihood plots of the two data sets?

6. Consider the count data $47, 40, 46, 41, 40$. Repeat the computations of items 1-3 of problem 5.

7. Suppose that $Y_i \sim \text{N}(\beta x_i, \sigma^2)$ for $i = 1, 2, \ldots, n$, all independent, where x_i is a known covariate.

   1. Write down the likelihood for the data $y_1, \ldots, y_n$.

Solution: The likelihood is

$$\begin{aligned} L(\beta, \sigma^2; \mathbf{y}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \beta x_i)^2}{2\sigma^2}\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2\right\}. \end{aligned}$$

   2. Find the maximum likelihood estimators $\hat{\beta}$ and $\hat{\sigma}^2$ of β and σ^2.

Solution: The log-likelihood is

$$\ell(\beta, \sigma^2; \mathbf{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2.$$

Thus, we have

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i(y_i - \beta x_i)$$

and

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \beta x_i)^2.$$

Setting the first derivative to zero, we obtain

$$\sum_{i=1}^{n} x_i y_i - \hat{\beta}\sum_{i=1}^{n} x_i^2 = 0,$$

which yields the maximum likelihood estimate

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

Similarly, setting the second derivative to zero, we obtain

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta} x_i)^2.$$

   3. Prove that $\hat{\beta}$ is an unbiased estimator of β.

Solution: We can write

$$\begin{aligned} E(\hat{\beta}) &= \frac{1}{\sum_{i=1}^n x_i^2} E\left(\sum_{i=1}^n x_i Y_i\right) = \frac{1}{\sum_{i=1}^n x_i^2} \sum_{i=1}^n x_i E(Y_i) \\ &= \frac{1}{\sum_{i=1}^n x_i^2} \sum_{i=1}^n \beta x_i^2 = \beta, \end{aligned}$$

and so $\hat{\beta}$ is an unbiased estimator of β.

8. In this problem we study properties of the link function $g(u) = log(u)$.

   1. Determine the domain and range of $g(u)$.

      Solution: This function is defined for positive values of u and takes real values, i.e. its domain is $\mathbb{R}_{>0} = (0, \infty)$ and the range is all of $\mathbb{R}$.

   2. For which type of response is the link $g(u)$ function most suitable?

      Solution: This link is most suitable for positive responses.

   3. Invert $g(u)$ and compute directly the derivative of the inverse $g^{-1}(u)$, i.e. $\frac{d}{du} g^{-1}(u)$.

      Solution: The inverse $g^{-1}(\cdot)$ is the function such that $g^{-1}(g(u)) = u$. For the logarithm, the inverse is $g^{-1}(u) = \exp(u)$, and the derivative of the inverse is the standard result $\frac{d}{du} g^{-1}(u) = \exp(u)$.

   4. Find out about the inverse function theorem. Use the inverse function theorem to compute the derivative of the inverse $g^{-1}(u)$.

      Solution: The inverse theorem states that if $g^{-1}(u)$ is the inverse of $g(u)$, then the derivative of the inverse satisfies $\frac{d}{du} g^{-1}(u) = 1/\left(g'(g^{-1}(u))\right)$, where $g'(u) = \frac{d}{du} g(u)$. In our case, we have $g'(u) = \frac{d}{du} \log(u) = 1/u$ so that $\frac{d}{du} g^{-1}(u) = 1/\left(1/\exp(u)\right) = \exp(u)$ that coincides with the earlier computation.

   5. Repeat the steps and computations above for the following link functions:
      (a) The identity link $g(u) = u$.
      (b) The inverse quadratic link $g(u) = u^{-2}$.
      (c) The square root link $g(u) = u^{-1/2} = \sqrt{u}$.
      (d) The logit link $g(u) = \log(u/(1-u))$.
      (e) The complementary log-log link $g(u) = \log(-\log(1-u))$.
      (f) The Cauchy link $g(u) = \Phi^{-1}(u)$, where $\Phi(u) = \frac{1}{2} + \frac{1}{\pi} \arctan(u)$.

(g) **(Medium)** The Gumbel link $g(u) = \Phi^{-1}(u)$, where $\Phi(u) = \exp(-\exp(-u))$ is the cumulative Gumbel distribution. Discuss one potential disadvantage of this link.

(h) **(Hard)** The probit link $g(u) = \Phi^{-1}(u)$, where $\Phi(u)$ is the cumulative distribution of the standard normal random variable.

9. Suppose that $Y_i \sim \mathrm{N}(\mu, \sigma^2)$ for $i = 1, 2, \ldots, n$, all independent,

   1. Write down the likelihood for the data $y_1, \ldots, y_n$. **Hint**: Try to reuse the equations in lecture notes (ditto for the second item).
   2. Determine analytically the maximum likelihood estimates.

   Solution: You should obtain the maximum likelihood estimates of the sample mean for μ and the (biased) sample variance for σ^2.

   3. Find the Fisher information matrix.

10. The observations 6.3, 4.2, 6.02, 4.32, 4.04, 3.95 are assumed to be independent realizations of the normal model $\mathrm{N}(\mu, \sigma^2)$.

   1. Using `R`, compute the likelihood estimates with formulæ $\hat{\mu} = \bar{y}$ and $\hat{\sigma}^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 / n$.

   Solution: Recall that you need to define the variables `y<-c(6.3, 4.2, 6.02, 4.32, 4.04, 3.95)` and `n<-6` for the analysis below.

```
mean(y) ## easy

## [1] 4.805

var(y) ## not quite, as this is the unbiased version

## [1] 1.12575
```

   The above computation with `var` is not the correct maximum likelihood estimate and here we compute the correct $\hat{\sigma}^2$ in two ways:

```
var(y)*(1-1/n) ## rectify the unbiased to be mle

## [1] 0.938125

mean( (y-mean(y))^2   )  ## the definition above

## [1] 0.938125
```

   2. Formulate the estimation of μ, σ like a linear regression in `R` and compute the estimates $\hat{\mu}, \hat{\sigma}^2$. In other words, use the function `lm` and process its output.

   Solution: To formulate as regression, just note that there is no "explanatory variable", so the model has only the intercept.

```
lm(y~1)->M
```

From which we retrieve $\hat{\mu}$ (the mean) as the intercept

```
M$coefficients

## (Intercept)
##       4.805
```

and then estimate $\hat{\sigma}^2$ in two ways. First the usual way with the residuals of this very simple model, then by exploiting the function `summary`:

```
mean( M$residuals^2 )

## [1] 0.938125

summary(M)->S
S$sigma^2*(1-1/n) ## corrected to obtain the mle

## [1] 0.938125
```

11. The Federal Trade Commission measured the numbers of milligrammes of tar (x) and carbon monoxide (y) per cigarette for all domestic filtered and mentholated cigarettes of length 100 millimetres. A sample of 12 brands yielded the following data:

Brand	x	y
Capri	9	6
Carlton	4	6
Kent	14	14
Kool Milds	12	12
Marlboro Lights	10	12
Merit Ultras	5	7
Now	3	4
Salem	17	18
Triumph	6	8
True	7	8
Vantage	8	13
Virginia Slims	15	13

1. Calculate the least squares regression line for these data.

Solution: The data give $\sum_{i=1}^{12} x_i = 110$, $\sum_{i=1}^{12} y_i = 121$, $\sum_{i=1}^{12} x_i y_i = 1,294$ and $\sum_{i=1}^{12} x_i^2 = 1,234$. So we have

$$\hat{\beta}_1 = \frac{1,294 - 110 \times 121/12}{1,234 - 110^2/12} = 0.8191$$

and

$$\hat{\beta}_0 = \frac{121}{12} - 0.8191 \times \frac{110}{12} = 2.575.$$

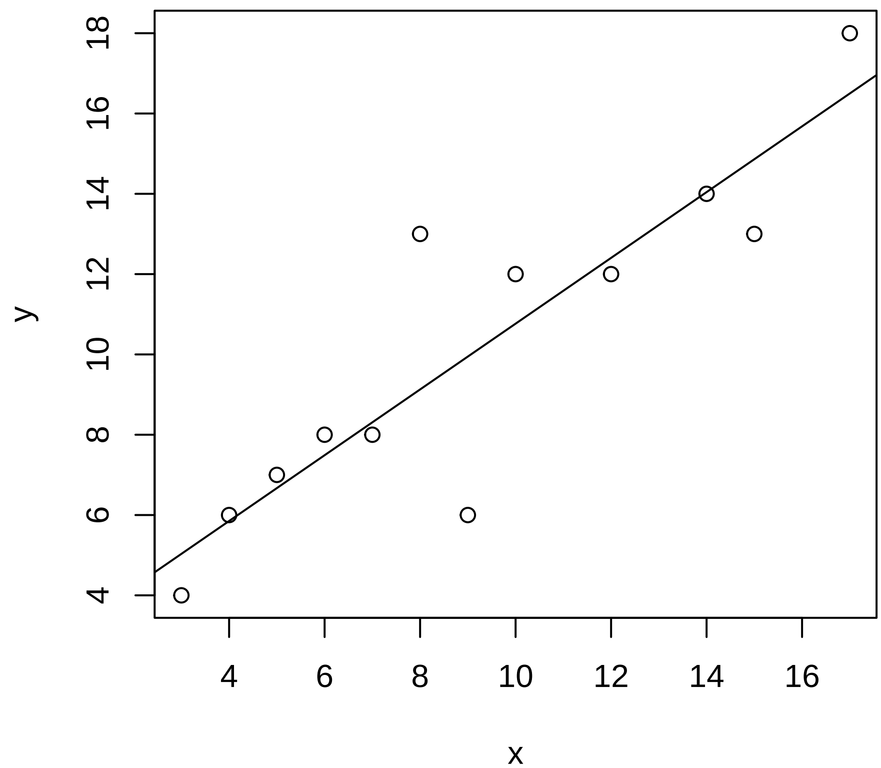
It follows that the least squares regression line is

$$\hat{y}_i = \hat{\mu}_i = 2.575 + 0.8191 x_i.$$

2. Plot the points and the least squares regression line on the same graph.

Solution: Here is the `R` code and the plot with data and fitted line.

```
DAT<-matrix(ncol=2,byrow=TRUE,c(9,6,4,6,14,14,12,12,
10,12,5,7,3,4,17,18,6,8,7,8,8,13,15,13))
colnames(DAT)<-c("x","y")
plot(DAT) ## The data
lm(y~x,data=as.data.frame(DAT))->M
abline(a=M$coefficients[1],b=M$coefficients[2])
```




```
M$coefficients

## (Intercept)           x
##   2.5753323   0.8190547
```

3. Find an unbiased estimate of σ^2.

Solution: An unbiased estimate of σ^2 is

$$s^2 = \frac{1}{10}\sum_{i=1}^{12}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 3.953.$$

As an extra output, here is the computation using `R`:

```
summary(M)$sigma^2 ## using summary
## [1] 3.952806
sum(M$residuals^2)/(M$df.residual) ## The traditional formula
## [1] 3.952806
```

12. Consider the data on manatees in Practical 1. Use `R` to answer the questions below.

    1. Produce a scatterplot of the data. Does the relationship between y and x seem to be linear?

    Solution: Note that this analysis was covered in the lab so here we only give additional comments.

    Concerning the scatterplot, there is some evidence of a linear relationship between y and x. However, it is not a particularly strong relation.

    2. Fit a simple linear regression model to the data. Give the values of $\hat{\beta}_0$ and $\hat{\beta}_1$, and test $H_0 : \beta_1 = 0$.

    Solution: The fitted linear regression model is

    $$\hat{y}_i = \hat{\mu}_i = -38.29 + 0.1187x_i.$$

    Since the p-value for the test of $H_0 : \beta_1 = 0$ is $P < 0.001$, there is very strong evidence that $\beta_1 \neq 0$.

    3. By examining the residual plots, comment on whether there is any reason to doubt the assumptions of the model.

    Solution: There is no reason to doubt the assumptions of linearity or constant variance. Although the first and last points on the Q-Q plot are a little away from the line, there does not seem much reason to doubt the normality assumption either.

13. Suppose that $Y_i \sim \mathrm{N}(\beta x_i, \sigma^2)$ for $i = 1, 2, \ldots, n$, all independent, where x_i is a known covariate.

    1. Find the Fisher information matrix. **Hint**: Try to reuse the equations in the lecture notes.

Solution: In this case, the Fisher information matrix V is 2×2. We have

$$\frac{\partial^2 \ell}{\partial \beta^2} = -\frac{1}{\sigma^2}\sum_{i=1}^n x_i^2, \qquad \frac{\partial^2 \ell}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^n x_i(y_i - \beta x_i)$$

and

$$\frac{\partial^2 \ell}{\partial \sigma^4} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n (y_i - \beta x_i)^2.$$

Computing expectations and reversing the sign, it follows that $v_{11} = \sum_{i=1}^n x_i^2/\sigma^2$, $v_{12} = 0$ and $v_{22} = n/(2\sigma^4)$. Thus, we obtain

$$V = \begin{pmatrix} \frac{\sum_{i=1}^n x_i^2}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}.$$

2. State the asymptotic distributions of the maximum likelihood estimators $\hat{\beta}$ and $\hat{\sigma}^2$ of β and σ^2.

Solution: We have

$$V^{-1} = \begin{pmatrix} \frac{\sigma^2}{\sum_{i=1}^n x_i^2} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}.$$

This shows that, for large n, $\hat{\beta} \sim \mathrm{N}(\beta, \sigma^2/\sum_{i=1}^n x_i^2)$ and $\hat{\sigma}^2 \sim \mathrm{N}(\sigma^2, 2\sigma^4/n)$.

3. Explain why the distribution of $\hat{\beta}$ is exact.

Solution: The distribution of $\hat{\beta}$ is exact because $\hat{\beta}$ is a linear combination of normal random variables.
To be sure, the maximum likelihood estimator is $\hat{\beta} = \left(\sum_{i=1}^n x_i y_i\right) / \left(\sum_{i=1}^n x_i^2\right)$, which is easily rewritten as $\hat{\beta} = \sum_{i=1}^n c_i y_i$ where the c_i are constants $c_i = x_i / \left(\sum_{j=1}^n x_j^2\right)$. That is, $\hat{\beta}$ is a linear combination of the y_i. When we considere these as normal random variables Y_i, we use a known result that a linear combination of normal random variables has a normal distribution, hence it is exact.

14. Consider the manatees' data again and a regression model that passes through the origin.

1. Explain in simple terms what does a model going through the origin imply for the manatees' data.
2. Using the data, compute with the help of `R` an estimate of the matrix V and then give its inverse V^{-1} which is an estimation of the variance-covariance matrix for the model parameters.

Solution: We use the formulæ directly. Initial calculations are:

```
## Manatee data
x <- c(447,460,481,498,513,512,526,559,585,614,645)
y <- c(13,21,24,16,24,20,15,34,33,33,39)
M<-lm(y~x-1) ## the regression through the origin
M$coefficients ## this is \hat\beta_1

##          x
## 0.04749354

sigma2<-mean(M$residuals^2) ## this is \hat\sigma^2
sigma2

## [1] 37.65413
```

We are not ready to compute the elements of V:

```
v11<-sum(x^2)/sigma2
v22<-length(x)/(2*sigma2^2)
```

so that the estimate of V is $\hat{V} = \begin{pmatrix} 8.34036 \times 10^4 & 0 \\ 0 & 0.0038792 \end{pmatrix}$ and an estimate of the variance-covariance matrix is $\hat{V}^{-1} = \begin{pmatrix} 1.198989 \times 10^{-5} & 0 \\ 0 & 257.7879122 \end{pmatrix}$.

The estimated standard errors of model parameters are given by square roots of the previous matrix, i.e. the standard error of $\hat{\beta}_1$ is 0.0035 and for $\hat{\sigma}^2$ is 16.0558.

3. Briefly comment upon your results.

Solution: Let us look at a 95% confidence intervals for each parameter, for β_1 we have $(0.0407, 0.0543)$, and for σ^2 we have $(6.1854, 69.1229)$.

The interval for σ^2, does not including zero but it is quite wide if we compare it with $(3.1191, 34.8563)$ which is the interval for σ^2 from the model including intercept. This is a consequence of the increased variability $\hat{\sigma} = 16.0558$ of the no intercept model relative to the variability $\hat{\sigma} = 8.0964$ of the model with intercept.

We may remove the intercept, but the price to pay is increased variability by about a factor of two. This is perhaps too high a price to pay.

15. Consider the model $Y_i \sim \mathrm{N}(\mu, \mu^2)$ for $i = 1, 2, \ldots, n$, all independent,

    1. Write down the likelihood for the data $y_1, \ldots, y_n$.
    2. Determine analytically the maximum likelihood estimate.
    3. Find the Fisher information matrix.

16. Suppose that $Y_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ for $i = 1, 2, \ldots, n$, all independent, where $\mu_i = \mathbf{x}_i\beta$ and the σ_i are **known**.

    1. Write down the likelihood for the data $y_1, \ldots, y_n$.

Solution:
The likelihood is

$$\begin{aligned} L(\beta; \mathbf{y}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(y_i - \mathbf{x}_i\beta)^2}{2\sigma_i^2}\right\} \\ &= \left(\prod_{i=1}^{n} 2\pi\sigma_i^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - \mathbf{x}_i\beta)^2}{\sigma_i^2}\right\} \\ &= (2\pi)^{-\frac{n}{2}} \left(\prod_{i=1}^{n} \sigma_i^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - \mathbf{x}_i\beta)^2}{\sigma_i^2}\right\}. \end{aligned}$$

2. Show that $\hat{\beta} = (\mathbf{X}^\top \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Sigma^{-1} \mathbf{Y}$ is the maximum likelihood estimator of β. Here $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$.

Solution: Since

$$\sum_{i=1}^{n} \frac{(y_i - \mathbf{x}_i\beta)^2}{\sigma_i^2} = (\mathbf{y} - \mathbf{X}\beta)^\top \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta),$$

where $\mathbf{X}$ is the $n \times p$ design matrix with ith row $\mathbf{x}_i^\top$ and $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, we write

$$L(\beta; \mathbf{y}) = (2\pi)^{-\frac{n}{2}} \left(\prod_{i=1}^{n} \sigma_i^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^\top \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta)\right\}.$$

So the log-likelihood is

$$\ell(\beta; \mathbf{y}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n} \log(\sigma_i^2) - \frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^\top \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta).$$

Thus, we have

$$\frac{\partial \ell}{\partial \beta} = \mathbf{X}^\top \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta).$$

Assume that $\text{rank}(\mathbf{X}) = p$, so that the $p \times p$ matrix $\mathbf{X}^\top \Sigma^{-1} \mathbf{X}$ is non-singular. Then, setting the above derivatives to zero, we obtain

$$\hat{\beta} = (\mathbf{X}^\top \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Sigma^{-1} \mathbf{y}.$$

3. Find the Fisher information matrix.

Solution: We have

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta^\top} = -\mathbf{X}^\top \Sigma^{-1} \mathbf{X}.$$

It follows that the Fisher information matrix is

$$V = \mathbf{X}^\top \Sigma^{-1} \mathbf{X}.$$

17. Consider the following data (-1,3.1), (-1,2.1), (0,5.4), (0,4.2), (1,6), (1,6), which is given as pairs (x_i, y_i). Implement in `R` the results of the model $Y_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ for $i = 1, 2, \ldots, n$, all independent, where $\mu_i = \beta_0 + \beta_1 x_i$. The σ_i are known as $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_3^2 = \sigma_4^2 = 2$, $\sigma_5^2 = \sigma_6^2 = 4$.

    In particular, compute the maximum likelihood estimate $\hat{\beta}$ and its asymptotic variance-covariance matrix.

Solution: The maximum likelihood estimate is the $\hat{\beta} = (\mathbf{X}^\top \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Sigma^{-1} \mathbf{y}$ which to be implemented, requires declaring the variables `x,y` with `R` commands `x<-c(-1, -1, 0, 0, 1, 1)` and `y<-c(3.1, 2.1, 5.4, 4.2, 6, 6)` and variances `sigma2i<-c(1, 1, 2, 2, 4, 4)`.

```
X<-cbind(1,x)  ## Design model matrix: intercept and variable x
S<-diag(sigma2i)  ## Matrix of variances
FIM<- t(X)%*%solve(S)%*%X; ## Fisher information matrix
betah<-solve(FIM)%*%t(X)%*%solve(S)%*%y
```

We have the following results $\hat{\beta}$, Fisher information matrix $\mathbf{X}^\top \Sigma^{-1} \mathbf{X}$ and asymptotic variance-covariance matrix $(\mathbf{X}^\top \Sigma^{-1} \mathbf{X})^{-1}$.

```
betah ## mle

##         [,1]
##    4.492308
## x 1.815385

FIM ## Fisher information matrix

##            x
##    3.5 -1.5
## x -1.5  2.5

solve(FIM) ## Asymptotic variance-covariance matrix

##                     x
##    0.3846154 0.2307692
## x 0.2307692 0.5384615
```

18. Suppose that $Y_i \sim \mathrm{Bin}(r_i, \pi_i)$ for $i = 1, 2, \ldots, n$, all independent, where the r_i are known, $\pi_i = \beta_0 + \beta_1 x_i$ and x_i is a known covariate.

1. Write down the likelihood for the data $y_1, \ldots, y_n$.

Solution: The likelihood is

$$\begin{aligned} L(\beta_0, \beta_1; \mathbf{y}) &= \prod_{i=1}^{n} \binom{r_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{r_i - y_i} \\ &= \prod_{i=1}^{n} \binom{r_i}{y_i} (\beta_0 + \beta_1 x_i)^{y_i} (1 - \beta_0 - \beta_1 x_i)^{r_i - y_i}. \end{aligned}$$

2. Obtain the likelihood equations.

Solution: The log-likelihood is

$$\ell(\beta_0, \beta_1; \mathbf{y}) = \sum_{i=1}^{n} \log \binom{r_i}{y_i} + \sum_{i=1}^{n} y_i \log(\beta_0 + \beta_1 x_i) + \sum_{i=1}^{n} (r_i - y_i) \log(1 - \beta_0 - \beta_1 x_i).$$

Thus, we have

$$\frac{\partial \ell}{\partial \beta_0} = \sum_{i=1}^{n} \frac{y_i}{\beta_0 + \beta_1 x_i} - \sum_{i=1}^{n} \frac{r_i - y_i}{1 - \beta_0 - \beta_1 x_i}$$

and

$$\frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^{n} \frac{x_i y_i}{\beta_0 + \beta_1 x_i} - \sum_{i=1}^{n} \frac{x_i (r_i - y_i)}{1 - \beta_0 - \beta_1 x_i}.$$

Setting these derivatives to zero yields the likelihood equations.

3. Find the Fisher information matrix.

Solution: In this case, the Fisher information matrix V is 2×2. We have

$$\frac{\partial^2 \ell}{\partial \beta_0^2} = -\sum_{i=1}^{n} \frac{y_i}{(\beta_0 + \beta_1 x_i)^2} - \sum_{i=1}^{n} \frac{r_i - y_i}{(1 - \beta_0 - \beta_1 x_i)^2},$$

$$\frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} = -\sum_{i=1}^{n} \frac{x_i y_i}{(\beta_0 + \beta_1 x_i)^2} - \sum_{i=1}^{n} \frac{x_i (r_i - y_i)}{(1 - \beta_0 - \beta_1 x_i)^2}$$

and

$$\frac{\partial^2 \ell}{\partial \beta_1^2} = -\sum_{i=1}^{n} \frac{x_i^2 y_i}{(\beta_0 + \beta_1 x_i)^2} - \sum_{i=1}^{n} \frac{x_i^2 (r_i - y_i)}{(1 - \beta_0 - \beta_1 x_i)^2}.$$

Concerning the expectations required, since $E(Y_i) = r_i \pi_i = r_i(\beta_0 + \beta_1 x_i)$, it follows that

$$v_{11} = \sum_{i=1}^{n} \frac{r_i}{\pi_i (1 - \pi_i)}, \qquad v_{12} = \sum_{i=1}^{n} \frac{x_i r_i}{\pi_i (1 - \pi_i)}$$

and

$$v_{22} = \sum_{i=1}^{n} \frac{x_i^2 r_i}{\pi_i (1 - \pi_i)}.$$

Thus, we obtain

$$V = \begin{pmatrix} \sum_{i=1}^n \frac{r_i}{\pi_i(1-\pi_i)} & \sum_{i=1}^n \frac{x_i r_i}{\pi_i(1-\pi_i)} \\ \sum_{i=1}^n \frac{x_i r_i}{\pi_i(1-\pi_i)} & \sum_{i=1}^n \frac{x_i^2 r_i}{\pi_i(1-\pi_i)} \end{pmatrix}.$$

19. Suppose that $Y_i \sim$ Poisson(μ_i) for $i = 1, 2, \ldots, n$, all independent, where $\mu_i = \beta_0 + \beta_1 x_i$ and x_i is a known covariate.

1. Write down the likelihood for the data $y_1, \ldots, y_n$.

Solution: The likelihood is

$$L(\beta_0, \beta_1; \mathbf{y}) = \prod_{i=1}^n \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} = \frac{\prod_{i=1}^n (\beta_0 + \beta_1 x_i)^{y_i} e^{-\sum_{i=1}^n (\beta_0 + \beta_1 x_i)}}{\prod_{i=1}^n y_i!}.$$

2. Obtain the likelihood equations.

Solution: The log-likelihood is

$$\ell(\beta_0, \beta_1; \mathbf{y}) = \sum_{i=1}^n y_i \log(\beta_0 + \beta_1 x_i) - n\beta_0 - \beta_1 \sum_{i=1}^n x_i - \sum_{i=1}^n \log(y_i!).$$

Thus, we have

$$\frac{\partial \ell}{\partial \beta_0} = \sum_{i=1}^n \frac{y_i}{\beta_0 + \beta_1 x_i} - n$$

and

$$\frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^n \frac{x_i y_i}{\beta_0 + \beta_1 x_i} - \sum_{i=1}^n x_i.$$

Setting these derivatives to zero yields the likelihood equations.

3. Find the Fisher information matrix.

Solution: In this case, the Fisher information matrix V is 2×2. We have

$$\frac{\partial^2 \ell}{\partial \beta_0^2} = -\sum_{i=1}^n \frac{y_i}{(\beta_0 + \beta_1 x_i)^2},$$

$$\frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} = -\sum_{i=1}^n \frac{x_i y_i}{(\beta_0 + \beta_1 x_i)^2}$$

and

$$\frac{\partial^2 \ell}{\partial \beta_1^2} = -\sum_{i=1}^n \frac{x_i^2 y_i}{(\beta_0 + \beta_1 x_i)^2}.$$

For the expectation, since $E(Y_i) = \mu_i = \beta_0 + \beta_1 x_i$, it follows that

$$v_{11} = \sum_{i=1}^{n} \frac{1}{\mu_i}, \qquad v_{12} = \sum_{i=1}^{n} \frac{x_i}{\mu_i}$$

and

$$v_{22} = \sum_{i=1}^{n} \frac{x_i^2}{\mu_i}.$$

Thus, we obtain

$$V = \begin{pmatrix} \sum_{i=1}^{n} \frac{1}{\mu_i} & \sum_{i=1}^{n} \frac{x_i}{\mu_i} \\ \sum_{i=1}^{n} \frac{x_i}{\mu_i} & \sum_{i=1}^{n} \frac{x_i^2}{\mu_i} \end{pmatrix}.$$

20. Consider the data on diabetics in Practical 2. Use `R` to answer the questions below.

    1. Produce scatterplots of y against each of the explanatory variables. Does y appear to be linearly related to them?
    2. Fit a multiple linear regression model to the full data. Give the values of the estimated regression coefficients and test $H_0 : \beta_1 = 0$.
    3. Remove x_1 from the model. By examining the residual plots, comment on whether there is any reason to doubt the assumptions of the reduced model.

21. Suppose that $Y_i \sim \text{Poisson}(\mu)$ for $i = 1, 2, \ldots, n$, all independent, and consider testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, where μ_0 is known.

    1. Write down the restricted maximum likelihood estimate $\hat{\mu}_0$ of μ under H_0 and the maximum likelihood estimate $\hat{\mu}$.

    Solution: The restricted maximum likelihood estimate of μ under H_0 is $\hat{\mu}_0 = \mu_0$ and the maximum likelihood estimate of μ is $\hat{\mu} = \overline{y}$.

    2. Obtain the generalised likelihood ratio.

    Solution: The generalised likelihood ratio is

$$\begin{aligned} \Lambda(\mathbf{y}) &= \frac{L(\hat{\mu}_0; \mathbf{y})}{L(\hat{\mu}; \mathbf{y})} \\ &= \frac{\hat{\mu}_0^{\sum_{i=1}^n y_i} e^{-n\hat{\mu}_0}}{\prod_{i=1}^n y_i!} \frac{\prod_{i=1}^n y_i!}{\hat{\mu}^{\sum_{i=1}^n y_i} e^{-n\hat{\mu}}} \\ &= \frac{\mu_0^{\sum_{i=1}^n y_i} e^{-n\mu_0}}{\overline{y}^{\sum_{i=1}^n y_i} e^{-n\overline{y}}} \\ &= \left(\frac{\mu_0}{\overline{y}}\right)^{\sum_{i=1}^n y_i} e^{n(\overline{y} - \mu_0)}. \end{aligned}$$

3. Use Wilks' theorem to find the critical region of a test with approximate significance level α for large n.

Solution: We have

$$\begin{aligned} -2\log\{\Lambda(\mathbf{y})\} &= -2\left\{n\overline{y}\log\left(\frac{\mu_0}{\overline{y}}\right)+n(\overline{y}-\mu_0)\right\} \\ &= 2n\left\{\overline{y}\log\left(\frac{\overline{y}}{\mu_0}\right)+\mu_0-\overline{y}\right\}. \end{aligned}$$

Here, $p=1$ and $p_0=0$, so that $s=1$. Therefore, by Wilks' theorem, when H_0 is true and n is large,

$$2n\left\{\overline{Y}\log\left(\frac{\overline{Y}}{\mu_0}\right)+\mu_0-\overline{Y}\right\}\sim\chi^2_1.$$

Hence, for a test with approximate significance level α, we reject H_0 if and only if

$$2n\left\{\overline{y}\log\left(\frac{\overline{y}}{\mu_0}\right)+\mu_0-\overline{y}\right\}>\chi^2_{1,\alpha}.$$

22. Consider the data 5, 1, 3, 5, 5, 4, 3, 2 which are assumed to be independent realizations of the Poisson distribution with expectation μ. We want to test $H_0:\mu=\mu_0$ with $\mu_0=3$.

1. Obtain the numerical value of the generalised likelihood ratio $\Lambda(\mathbf{y})$ and discuss about the distribution of this statistic to perform the test $H_0:\mu=\mu_0$.

Solution:
After defining `y` with commands `y<-c(5, 1, 3, 5, 5, 4, 3, 2)`, here is the observed value of the ratio

```
LLobs<-prod(dpois(x=y,lambda=3)/dpois(x=y,lambda=mean(y)))
LLobs

## [1] 0.7288998
```

Make sure you can explain what the above code means.
As for the distribution and the test, we reject H_0 if the observed value of $\Lambda(y)$ is smaller than the α lower quantile of the distribution of $\Lambda(y)$ under the null hypothesis.
We can use simulation, as in lectures of week 4 to obtain samples of this null distribution of Λ, see the code below.

```
NN<-25000 ## number of simulations
LLsim<-c(matrix(ncol=NN))
for(i in 1:NN){
  set.seed(i)
  y<-rpois(n=n,lambda=3) ## Simulate under the null
  lambdares<-3 ## Restricted
  lambdamle<-mean(y) ## unrestricted
  LL<-prod(dpois(x=y,lambda=lambdares)/dpois(x=y,lambda=lambdamle))
  LLsim[i]<-LL
}
```

After running the code, plot a histogram of `LLsim` and see where is the value `LLobs` located. You can compute the p-value with `mean(LLsim<LLobs)`. What do you conclude?

2. Use Wilk's theorem to test $H_0 : \mu = \mu_0$, perform the test and write your conclusions.

Solution: Here is the observed statistic and the 5% upper quantile of the χ^2_1 distribution.

```
-2*log(LLobs)

## [1] 0.6324381

qchisq(p=0.05,df=1,lower.tail = FALSE)

## [1] 3.841459
```

What do you conclude?

3. Using the normal approximation to the data, perform the test $H_0 : \mu = \mu_0$ and compare with the earlier results.

23. Suppose for $i = 1, 2, \ldots, n$, we have independent $Y_i \sim \text{Bin}(r_i, p)$, where r_i is known. Using data $y_1, \ldots, y_n$, consider testing $H_0 : p = p_0$ against $H_1 : p \neq p_0$, where p_0 is known.

1. Write down the restricted maximum likelihood estimate $\hat{p}_0$ of p under H_0 and the maximum likelihood estimate $\hat{p}$.
2. Obtain the generalised likelihood ratio for this test.
3. Use Wilks' theorem to find the critical region of a test with approximate significance level α, for large n.
4. The following (25,10), (15,6), (30,10) are data pairs (r_i, y_i) from acceptance sampling in textile industry. Apply your results to build the generalized likelihood ratio and use Wilks' theorem with $\alpha = 0.05$ to test $H_0 : p = 0.3$.

Solution: The maximum value of the likelihood restricted to H_0 is 0.0019, computed as the product of binomial probabilities using the value $\hat{p}_0 = 0.3$; the unrestricted maximum value of the likelihood is 0.0043, obtained as the product of binomial probabilities using the maximum likelihood estimate $\hat{p} = 0.3714$. Thus the ratio is $\Lambda(\mathbf{y}) = 0.4417$.

To apply Wilks' theorem, we compute $-2\log\Lambda(\mathbf{y}) = 1.6344$, which is to be compared with the critical value 3.8415, where the χ^2 distribution has one degree of freedom. As result of the comparison, we do not reject H_0.

24. Suppose that $Y \sim \text{Bin}(r, \pi)$, where r is known.

   1. Show that this distribution is a member of the exponential family.

Solution: We write

$$\begin{aligned} f_Y(y;\pi) &= \binom{r}{y}\pi^y(1-\pi)^{r-y} \\ &= \exp\left\{\log\binom{r}{y} + y\log\pi + (r-y)\log(1-\pi)\right\} \\ &= \exp\left\{y\log\left(\frac{\pi}{1-\pi}\right) + r\log(1-\pi) + \log\binom{r}{y}\right\}. \end{aligned}$$

Thus, we have $a(y) = y$, $b(\pi) = \log\{\pi/(1-\pi)\}$, $c(\pi) = r\log(1-\pi)$ and $d(y) = \log\binom{r}{y}$.

   2. Explain why the distribution is in canonical form and write down the natural parameter.

Solution: Since $a(y) = y$, the distribution is in canonical form. The natural parameter is $\log\{\pi/(1-\pi)\}$.

   3. Use the general results for $E\{a(Y)\}$ and $\text{Var}\{a(Y)\}$ to verify that $E(Y) = r\pi$ and $\text{Var}(Y) = r\pi(1-\pi)$.

Solution: We have $b'(\pi) = 1/\{\pi(1-\pi)\}$, $b''(\pi) = (2\pi-1)/\{\pi(1-\pi)\}^2$, $c'(\pi) = -r/(1-\pi)$ and $c''(\pi) = -r/(1-\pi)^2$. So we obtain

$$E(Y) = -\frac{-r/(1-\pi)}{1/\{\pi(1-\pi)\}} = r\pi$$

and

$$\begin{aligned} \text{Var}(Y) &= \frac{\frac{2\pi-1}{(\pi(1-\pi))^2} \times \frac{-r}{1-\pi} + \frac{r}{(1-\pi)^2} \times \frac{1}{\pi(1-\pi)}}{\frac{1}{(\pi(1-\pi))^3}} \\ &= -r\pi(2\pi-1) + r\pi^2 = r\pi(1-\pi), \end{aligned}$$

25. Suppose that $Y \sim \mathrm{N}(\mu, \sigma^2)$, where σ^2 is known.

   1. Show that this distribution is a member of the exponential family.

> **Solution:** We write
>
> $$\begin{aligned} f_Y(y;\mu) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y-\mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{-\frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right\} \\ &= \exp\left\{\frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2}\right\}. \end{aligned}$$
>
> Thus, we have $a(y) = y$, $b(\mu) = \mu/\sigma^2$, $c(\mu) = -\mu^2/(2\sigma^2)$ and $d(y) = -\log(2\pi\sigma^2)/2 - y^2/(2\sigma^2)$.

   2. Explain why the distribution is in canonical form and write down the natural parameter.

> **Solution:** Since $a(y) = y$, the distribution is in canonical form. The natural parameter is μ/σ^2.

   3. Use the general results for $E\{a(Y)\}$ and $\mathrm{Var}\{a(Y)\}$ to verify that $E(Y) = \mu$ and $\mathrm{Var}(Y) = \sigma^2$.

> **Solution:** We have $b'(\mu) = 1/\sigma^2$, $b''(\mu) = 0$, $c'(\mu) = -\mu/\sigma^2$ and $c''(\mu) = -1/\sigma^2$. So we obtain
>
> $$E(Y) = -\frac{-\mu/\sigma^2}{1/\sigma^2} = \mu$$
>
> and
>
> $$\mathrm{Var}(Y) = \frac{(0 \times -\mu/\sigma^2) - (-1/\sigma^2 \times 1/\sigma^2)}{1/\sigma^6} = \sigma^2.$$

26. Consider a sequence of independent Bernoulli trials, where each trial has success probability p. The number of failures observed until we obtain r successes is a negative binomial random variable $X \sim \mathrm{NB}(r,p)$ with probability mass function $\Pr(X = x) = \binom{x+r-1}{x} p^r (1-p)^x$.

   1. Show that this distribution is a member of the exponential family.
   2. Is the distribution in canonical form? Which is the natural parameter?
   3. Using exponential family results, show that $E(X) = r(1-p)/p$.
   4. Consider data $x_1, x_2, \ldots, x_n$. Determine the maximum likelihood estimate $\hat{p}$. Is this estimator unbiased? Justify your answer.
   5. Compute the Fisher information number for estimating p.

27. Consider the random variable $Y \sim \text{Ber}(p)$.

    1. Show that this distribution is a member of the exponential family.
    2. Determine if the distribution is in canonical form and write down the natural parameter.
    3. Use the general results for $E\{a(Y)\}$ and $\text{Var}\{a(Y)\}$ to determine that $E(Y)$ and $\text{Var}(Y)$.

28. Repeat the calculations of Exercise 27 for the following distributions: a) binomial, b) geometric, c) exponential, d) gamma, e) lognormal and f) chi-squared.

29. Consider the mean $\mu = E(Y)$ and variance $\sigma^2 = V(Y)$ of the random variable $Y \sim \text{Ber}(p)$. Determine if the variance is a function of the mean and if so, give its explicit formula $\sigma^2 = f(\mu)$.

30. Repeat the calculations of Exercise 29 for the following distributions of the exponential family: a) binomial, b) geometric, c) negative binomial, d) Poisson, e) exponential, f) chi-squared, g) gamma and h) lognormal.

31. Suppose that $Y_i \sim \text{Bin}(r_i, \pi_i)$ for $i = 1, 2, \ldots, n$, all independent, where the r_i are known, $\log\{\pi_i/(1-\pi_i)\} = \beta_0 + \beta_1 x_i$ and x_i is a known covariate.

    1. Find the Fisher information matrix.

Solution: We know that $\mu_i = r_i \pi_i$. It follows that

$$\eta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \log\left(\frac{\mu_i}{r_i - \mu_i}\right)$$

and

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{r_i}{\mu_i(r_i - \mu_i)} = \frac{1}{r_i \pi_i (1-\pi_i)}.$$

Thus, since $\text{Var}(Y_i) = r_i \pi_i (1-\pi_i)$, the Fisher information matrix is

$$V = \begin{pmatrix} \sum_{i=1}^n r_i \pi_i (1-\pi_i) & \sum_{i=1}^n x_i r_i \pi_i (1-\pi_i) \\ \sum_{i=1}^n x_i r_i \pi_i (1-\pi_i) & \sum_{i=1}^n x_i^2 r_i \pi_i (1-\pi_i) \end{pmatrix}.$$

    2. Obtain the asymptotic distributions of the maximum likelihood estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1.

Solution: We have

$$V^{-1} = \frac{1}{|V|}\begin{pmatrix} \sum_{i=1}^n x_i^2 r_i \pi_i (1-\pi_i) & -\sum_{i=1}^n x_i r_i \pi_i (1-\pi_i) \\ -\sum_{i=1}^n x_i r_i \pi_i (1-\pi_i) & \sum_{i=1}^n r_i \pi_i (1-\pi_i) \end{pmatrix},$$

where

$$|V| = \sum_{i=1}^n r_i \pi_i (1-\pi_i) \sum_{i=1}^n x_i^2 r_i \pi_i (1-\pi_i) - \left\{\sum_{i=1}^n x_i r_i \pi_i (1-\pi_i)\right\}^2.$$

This shows that, for large n, $\hat{\beta}_0 \sim \text{N}(\beta_0, v^{11})$ and $\hat{\beta}_1 \sim \text{N}(\beta_1, v^{22})$, where $v^{11} = \sum_{i=1}^n x_i^2 r_i \pi_i (1-\pi_i)/|V|$ and $v^{22} = \sum_{i=1}^n r_i \pi_i (1-\pi_i)/|V|$.

3. State the approximate standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$.

Solution: The approximate standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$ are $\sqrt{\hat{v}^{11}}$ and $\sqrt{\hat{v}^{22}}$, respectively.

32. Suppose that the continuous random variables $Y_1, \ldots, Y_n$ have distributions depending on the parameters $\theta_1, \ldots, \theta_p$ and that their ranges do not depend on the parameters. Let $L(\theta; \mathbf{y})$ and $l(\theta; \mathbf{y})$ denote the likelihood and log-likelihood of the parameter vector θ, respectively.

1. Show that
$$\frac{\partial l(\theta; \mathbf{y})}{\partial \theta_j} = \frac{1}{L(\theta; \mathbf{y})} \frac{\partial L(\theta; \mathbf{y})}{\partial \theta_j}.$$

Solution: Since $l(\theta; \mathbf{y}) = \log L(\theta; \mathbf{y})$, we can write
$$\frac{\partial l(\theta; \mathbf{y})}{\partial \theta_j} = \frac{1}{L(\theta; \mathbf{y})} \frac{\partial L(\theta; \mathbf{y})}{\partial \theta_j}.$$

2. Prove that
$$E\left\{\frac{\partial l(\theta; \mathbf{Y})}{\partial \theta_j}\right\} = 0.$$

Solution: Let S be the sample space. Then we have
$$\begin{aligned} E\left\{\frac{\partial l(\theta; \mathbf{Y})}{\partial \theta_j}\right\} &= \int_S \frac{\partial l(\theta; \mathbf{y})}{\partial \theta_j} L(\theta; \mathbf{y}) d\mathbf{y} \\ &= \int_S \frac{\partial L(\theta; \mathbf{y})}{\partial \theta_j} d\mathbf{y} \\ &= \frac{\partial}{\partial \theta_j} \int_S L(\theta; \mathbf{y}) d\mathbf{y} \\ &= \frac{\partial}{\partial \theta_j}(1) = 0. \end{aligned}$$

3. By differentiating the identity in part 1 with respect to θ_k, prove that
$$E\left\{-\frac{\partial^2 l(\theta; \mathbf{Y})}{\partial \theta_j \partial \theta_k}\right\} = E\left\{\frac{\partial l(\theta; \mathbf{Y})}{\partial \theta_j} \frac{\partial l(\theta; \mathbf{Y})}{\partial \theta_k}\right\}.$$

Solution: Differentiating the identity in part 1 with respect to θ_k, we obtain
$$\begin{aligned} \frac{\partial^2 l(\theta; \mathbf{y})}{\partial \theta_j \partial \theta_k} &= \frac{1}{L(\theta; \mathbf{y})} \frac{\partial^2 L(\theta; \mathbf{y})}{\partial \theta_j \partial \theta_k} - \frac{1}{\{L(\theta; \mathbf{y})\}^2} \frac{\partial L(\theta; \mathbf{y})}{\partial \theta_j} \frac{\partial L(\theta; \mathbf{y})}{\partial \theta_k} \\ &= \frac{1}{L(\theta; \mathbf{y})} \frac{\partial^2 L(\theta; \mathbf{y})}{\partial \theta_j \partial \theta_k} - \frac{\partial l(\theta; \mathbf{y})}{\partial \theta_j} \frac{\partial l(\theta; \mathbf{y})}{\partial \theta_k}. \end{aligned}$$

It follows that

$$E\left\{-\frac{\partial^2 l(\theta;\mathbf{Y})}{\partial\theta_j\partial\theta_k}\right\} = -\int_S \frac{\partial^2 L(\theta;\mathbf{y})}{\partial\theta_j\partial\theta_k}d\mathbf{y} + E\left\{\frac{\partial l(\theta;\mathbf{Y})}{\partial\theta_j}\frac{\partial l(\theta;\mathbf{Y})}{\partial\theta_k}\right\}$$
$$= -\frac{\partial^2}{\partial\theta_j\partial\theta_k}(1) + E\left\{\frac{\partial l(\theta;\mathbf{Y})}{\partial\theta_j}\frac{\partial l(\theta;\mathbf{Y})}{\partial\theta_k}\right\}$$
$$= E\left\{\frac{\partial l(\theta;\mathbf{Y})}{\partial\theta_j}\frac{\partial l(\theta;\mathbf{Y})}{\partial\theta_k}\right\}.$$

33. Consider the data on beetles in Practical 3. answer the questions below. Fit the logistic, probit and extreme value models in R. Which of these provides the best description of the data? Present the results in a clear and concise table.

Solution: Note that this is indeed what the lab covers, so this exercise is about writing your summary of what you obtained in the lab session. As some help for your analysis, here are the deviances for the different link functions suggested, together with their p-values. A small p-value suggests that the model (link) is inadequate for the data. This information should be also compared with plots as obtained in the lab.

```
beetle<-read.csv("beetle.csv")
x<-beetle[,1]
r<-beetle[,2]
y<-beetle[,3]
p <- y/r
candidates<-c("logit", "probit", "cloglog")
deviance<-c()
for(link in candidates){
   beetle <- glm(p ~ x,family=binomial(link=link),weights=r)
        deviance<-c(deviance,beetle$deviance)
   }
deviance<-rbind(deviance,pchisq(q=deviance,df=beetle$df.residual,
                                lower.tail=FALSE))
colnames(deviance)<-candidates
rownames(deviance)<-c("deviance","p-value")
round(deviance,4)

##             logit  probit cloglog
## deviance 11.2322 10.1198  3.4464
## p-value   0.0815  0.1197  0.7511
```

34. Suppose that $Y_i \sim \text{Poisson}(\mu_i)$ for $i = 1, 2, \ldots, n$, all independent, where $\log(\mu_i) = \beta_0 + \beta_1 x_i$ and x_i is a known covariate.

   1. Find the Fisher information matrix.

Solution: We know that $\eta_i = \log(\mu_i)$. It follows that $\partial\eta_i/\partial\mu_i = 1/\mu_i$. Thus, since $\text{Var}(Y_i) = \mu_i$, the Fisher information matrix is

$$V = \begin{pmatrix} \sum_{i=1}^n \mu_i & \sum_{i=1}^n x_i\mu_i \\ \sum_{i=1}^n x_i\mu_i & \sum_{i=1}^n x_i^2\mu_i \end{pmatrix}.$$

2. Obtain the asymptotic distributions of the maximum likelihood estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1.

Solution: We have

$$V^{-1} = \frac{1}{|V|}\begin{pmatrix} \sum_{i=1}^n x_i^2\mu_i & -\sum_{i=1}^n x_i\mu_i \\ -\sum_{i=1}^n x_i\mu_i & \sum_{i=1}^n \mu_i \end{pmatrix},$$

where

$$|V| = \sum_{i=1}^n \mu_i \sum_{i=1}^n x_i^2\mu_i - \left(\sum_{i=1}^n x_i\mu_i\right)^2.$$

This shows that, for large n, $\hat{\beta}_0 \sim \text{N}(\beta_0, v^{11})$ and $\hat{\beta}_1 \sim \text{N}(\beta_1, v^{22})$, where $v^{11} = \sum_{i=1}^n x_i^2\mu_i/|V|$ and $v^{22} = \sum_{i=1}^n \mu_i/|V|$.

3. State the approximate standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$.

Solution: The approximate standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$ are $\sqrt{\hat{v}^{11}}$ and $\sqrt{\hat{v}^{22}}$, respectively.

35. Suppose that $Y_i \sim \text{Bin}(r_i, \pi_i)$ for $i = 1, 2, \ldots, n$, all independent, where the r_i are known, $\log\{\pi_i/(1-\pi_i)\} = \beta_0 + \beta_1 x_i$ and x_i is a known covariate.

1. Show that the maximum likelihood estimate of π_i in the maximal model is y_i/r_i.

Solution: Let $\pi = (\pi_1, \ldots, \pi_n)^\top$. Then the likelihood is

$$L(\pi; \mathbf{y}) = \prod_{i=1}^n \binom{r_i}{y_i} \pi_i^{y_i}(1-\pi_i)^{r_i - y_i},$$

and so the log-likelihood is

$$l(\pi; \mathbf{y}) = \sum_{i=1}^n \log\binom{r_i}{y_i} + \sum_{i=1}^n y_i \log(\pi_i) + \sum_{i=1}^n (r_i - y_i)\log(1-\pi_i).$$

Thus, we have

$$\frac{\partial l}{\partial \pi_i} = \frac{y_i}{\pi_i} - \frac{r_i - y_i}{1-\pi_i}.$$

Setting this derivative to zero, we obtain

$$y_i(1-\hat{\pi}_{i,\max}) - (r_i - y_i)\hat{\pi}_{i,\max} = 0,$$

which yields the maximum likelihood estimate $\hat{\pi}_{i,\max} = y_i/r_i$.

2. Obtain the generalised likelihood ratio.

Solution: The generalised likelihood ratio is

$$\begin{aligned}\Lambda(\mathbf{y}) &= \frac{L(\hat{\beta};\mathbf{y})}{L(\hat{\beta}_{\max};\mathbf{y})} \\ &= \prod_{i=1}^{n}\left(\frac{\hat{\pi}_i}{\hat{\pi}_{i,\max}}\right)^{y_i}\left(\frac{1-\hat{\pi}_i}{1-\hat{\pi}_{i,\max}}\right)^{r_i-y_i},\end{aligned}$$

where $\hat{\pi}_i$ is the maximun likelihood estimate of π_i in the two-parameter model.

3. Use Wilks' theorem to find the critical region of a test with approximate significance level α for large n.

Solution: We have

$$\begin{aligned}-2\log\{\Lambda(\mathbf{y})\} &= -2\sum_{i=1}^{n}\left\{y_i\log\left(\frac{\hat{\pi}_i}{\hat{\pi}_{i,\max}}\right) + (r_i-y_i)\log\left(\frac{1-\hat{\pi}_i}{1-\hat{\pi}_{i,\max}}\right)\right\} \\ &= 2\sum_{i=1}^{n}\left[y_i\log\left(\frac{y_i}{r_i\hat{\pi}_i}\right) + (r_i-y_i)\log\left\{\frac{r_i-y_i}{r_i(1-\hat{\pi}_i)}\right\}\right].\end{aligned}$$

Here, $p=n$ and $p_0=2$, so that $s=n-2$. Therefore, by Wilks' theorem, when H_0 is true and n is large,

$$2\sum_{i=1}^{n}\left[Y_i\log\left(\frac{Y_i}{r_i\hat{\pi}_i}\right) + (r_i-Y_i)\log\left\{\frac{r_i-Y_i}{r_i(1-\hat{\pi}_i)}\right\}\right] \sim \chi^2_{n-2}.$$

Hence, for a test with approximate significance level α, we reject H_0 if and only if

$$2\sum_{i=1}^{n}\left[y_i\log\left(\frac{y_i}{r_i\hat{\pi}_i}\right) + (r_i-y_i)\log\left\{\frac{r_i-y_i}{r_i(1-\hat{\pi}_i)}\right\}\right] > \chi^2_{n-2,\alpha}.$$

36. In lectures we have surveyed the logistic, probit and extreme value (complementary log-log) links which are used for the analysis of proportions (binomial data). In principle, a link for proportion data is any continuous function that transforms $(0,1)\to\mathbb{R}$. For example, the probit link $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function.

   1. Do some research about the link using the inverse cauchy distribution; write its explicit expression and show that it satisfies the transformation $(0,1)\to\mathbb{R}$.

2. Plot the link transformation and compare with other links mentioned. Can you see some advantages or drawbacks of the Cauchy?
3. Analyze the beetle data of Practical 3 using the link `cauchit`. Compare what you obtain with the earlier results. Does it improve over these? Write your comments.

37. A researcher wishes to know if consumption of caffeine improves performance on a memory test. There were 30 volunteers for each dose of caffeine (x), in milligrammes, and the number of volunteers who achieved a grade A in the memory test (y) is recorded. Below are the results.

x	0	50	100	150	200	250	300	350	400	450	500
y	10	13	17	15	10	5	4	3	3	1	0

1. Fit a logistic regression model to the data. Give the values of the estimated regression coefficients and assess the goodness of fit of the model.

Solution: The fitted logistic regression model is

$$\hat{\pi}_i = \frac{\exp(0.2385 - 0.0064x_i)}{1 + \exp(0.2385 - 0.0064x_i)}$$

and the deviance is $D = 18.625$. Since $\chi^2_{9,0.05} = 16.92$ and $\chi^2_{9,0.01} = 21.67$, the p-value is $0.01 < P < 0.05$, and so there is moderate evidence that this model does not fit the data particularly well. These results can be obtained in `R` as follows.

```r
x <- c(0,50,100,150,200,250,300,350,400,450,500)
y <- c(10,13,17,15,10,5,4,3,3,1,0)/30
mod1 <-glm(y~x, family=binomial, weights = rep(30,11))
summary(mod1)
```

```
##
## Call:
## glm(formula = y ~ x, family = binomial, weights = rep(30, 11))
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.238469   0.226199   1.054    0.292
## x           -0.006442   0.001009  -6.381 1.75e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 69.358  on 10  degrees of freedom
## Residual deviance: 18.625  on  9  degrees of freedom
## AIC: 55.87
##
## Number of Fisher Scoring iterations: 4
```

```r
qchisq(0.95,df = 9)
```

```
## [1] 16.91898
```

```r
qchisq(0.99,df = 9)
```

```
## [1] 21.66599
```

```r
pchisq(q = 18.625, df = 9,lower.tail = FALSE)
```

```
## [1] 0.02857711
```

2. Add x^2 to the model. Is there evidence that this model is an improvement over the two-parameter one?

Solution: The new model is

$$\hat{\pi}_i = \frac{\exp(-0.3974 + 0.0046x_i - 0.000028x_i^2)}{1 + \exp(-0.3974 + 0.0046x_i - 0.000028x_i^2)}$$

and the deviance is $D = 7.6639$. Since $\chi^2_{8,0.1} = 13.36$, the p-value is $P > 0.1$, and so there is no evidence that this model does not fit the data well.

```
mod2 <-glm(y~x+I(x^2), family=binomial, weights = rep(30,11))
summary(mod2)
```

```
##
## Call:
## glm(formula = y ~ x + I(x^2), family = binomial, weights = rep(30,
##     11))
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.974e-01  3.021e-01  -1.315  0.18836
## x            4.600e-03  3.633e-03   1.266  0.20538
## I(x^2)      -2.762e-05  9.257e-06  -2.984  0.00285 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 69.3577  on 10  degrees of freedom
## Residual deviance:  7.6639  on  8  degrees of freedom
## AIC: 46.909
##
## Number of Fisher Scoring iterations: 5
```

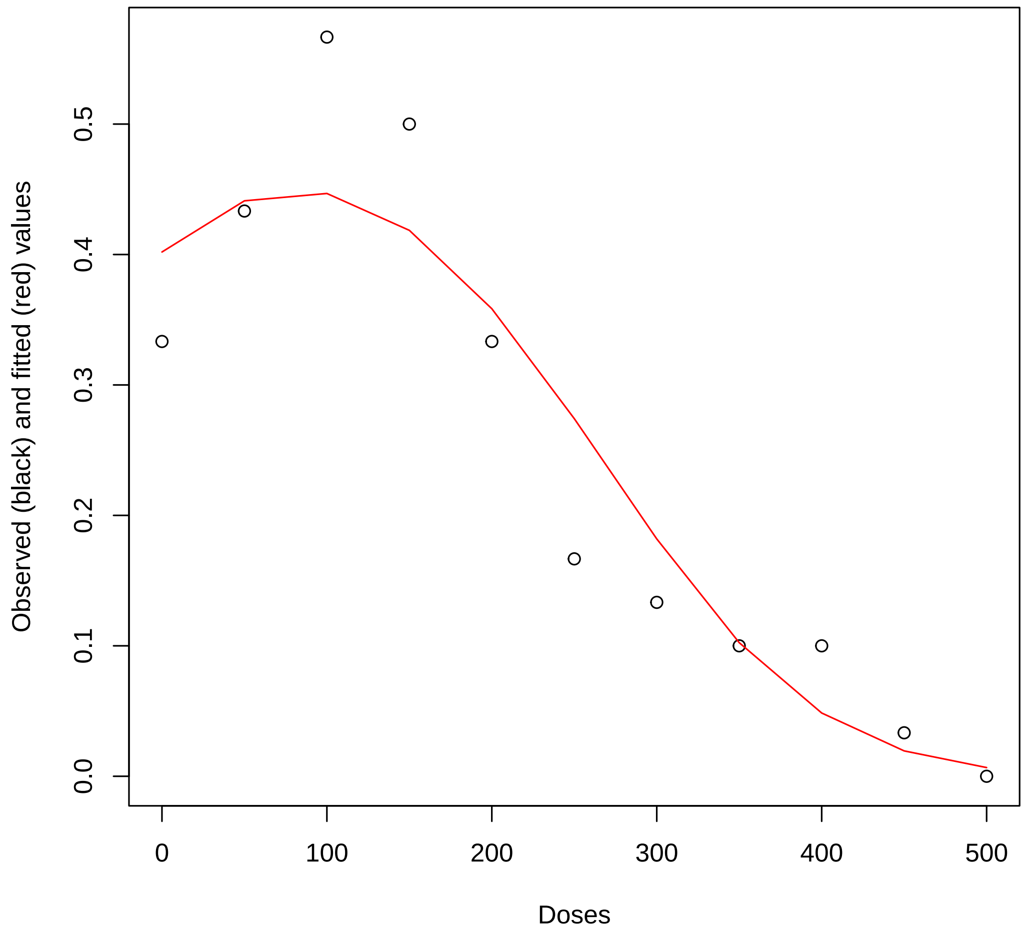
```
pchisq(q = deviance(mod2), df = mod2$df.residual, lower.tail = FALSE)
```

```
## [1] 0.4669742
```

3. Obtain the fitted values of the new model. Plot both the proportions and the fitted values against the doses.

Solution: A plot of both the proportions and the fitted values of the new model against the doses is given below.

```
plot(x,y, xlab="Doses",
     ylab="Observed (black) and fitted (red) values")
lines(x, fitted(mod2), col="red")
```



This shows that the new model captures the data quite well.

38. Suppose that $Y_i \sim \text{Poisson}(\mu_i)$ for $i = 1, 2, \ldots, n$, all independent, where $\log(\mu_i) = \beta_0 + \beta_1 x_i$ and x_i is a known covariate.

1. Show that the maximum likelihood estimate of μ_i in the maximal model is y_i.

Solution: Let $\mu = (\mu_1, \ldots, \mu_n)^\top$. Then the likelihood is

$$L(\mu; \mathbf{y}) = \prod_{i=1}^{n} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!},$$

and so the log-likelihood is

$$\ell(\mu; \mathbf{y}) = \sum_{i=1}^{n} y_i \log(\mu_i) - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} \log(y_i!).$$

Thus, we have

$$\frac{\partial \ell}{\partial \mu_i} = \frac{y_i}{\mu_i} - 1.$$

Setting this derivative to zero yields the maximum likelihood estimate $\hat{\mu}_{i,\max} = y_i$.

2. Obtain the generalised likelihood ratio.

Solution: The generalised likelihood ratio is

$$\begin{aligned}\Lambda(\mathbf{y}) &= \frac{L(\hat{\beta}; \mathbf{y})}{L(\hat{\beta}_{\max}; \mathbf{y})} \\ &= \prod_{i=1}^{n} \left(\frac{\hat{\mu}_i}{\hat{\mu}_{i,\max}} \right)^{y_i} e^{-\hat{\mu}_i + \hat{\mu}_{i,\max}},\end{aligned}$$

where $\hat{\mu}_i$ is the maximum likelihood estimate of μ_i in the two-parameter model.

3. Use Wilks' theorem to find the critical region of a test with approximate significance level α for large n.

Solution: We have

$$\begin{aligned}-2\log\{\Lambda(\mathbf{y}\} &= -2\sum_{i=1}^{n} \left\{ y_i \log\left(\frac{\hat{\mu}_i}{\hat{\mu}_{i,\max}} \right) - \hat{\mu}_i + \hat{\mu}_{i,\max} \right\} \\ &= 2\sum_{i=1}^{n} \left\{ y_i \log\left(\frac{y_i}{\hat{\mu}_i} \right) - y_i + \hat{\mu}_i \right\}.\end{aligned}$$

Here, $p = n$ and $p_0 = 2$, so that $s = n - 2$. Therefore, by Wilks' theorem, when H_0 is true and n is large,

$$2\sum_{i=1}^{n} \left\{ Y_i \log\left(\frac{Y_i}{\hat{\mu}_i} \right) - Y_i + \hat{\mu}_i \right\} \sim \chi^2_{n-2}.$$

Hence, for a test with approximate significance level α, we reject H_0 if and only if

$$2\sum_{i=1}^{n} \left\{ y_i \log\left(\frac{y_i}{\hat{\mu}_i} \right) - y_i + \hat{\mu}_i \right\} > \chi^2_{n-2,\alpha}.$$

39. Suppose that $Y_i \sim \mathrm{N}(\beta x_i, \sigma^2)$ for $i = 1, 2, \ldots, n$, all independent, where x_i is a known covariate and σ is known. The fitted values are $\hat{\mu}_i = \hat{\beta} x_i$ and the variance of Y_i is $V(\hat{\mu}_i) = \sigma^2$, and we have $V(x) = \sigma^2$. .

1. Write down the Pearson residual e_i^P.

Solution: The Pearson residual is

$$e_i^P = \frac{y_i - \hat{\mu}_i}{\sigma}.$$

2. Find the transformation $A(x)$.

Solution: The Anscombe residuals are

$$A(x) = \int \sigma^{-\frac{2}{3}} dx = \sigma^{-\frac{2}{3}} x + c,$$

where c is an arbitrary integration constant.

3. Obtain the Anscombe residual e_i^A.

Solution: Since $A'(x) = \sigma^{-\frac{2}{3}}$, the Anscombe residual is

$$e_i^A = \frac{y_i - \hat{\mu}_i}{\sigma},$$

which is the same as the Pearson residual.

40. In an experiment designed to assess the potency of two test preparations of an insecticide relative to a standard, 60 aphids were placed on each of 12 cabbage plants. The three insecticides (w) were then applied in various doses (x), in milligrammes per litre of water, to each of four plants. The number of aphids still alive after three days (y) is determined and the results are as follows:

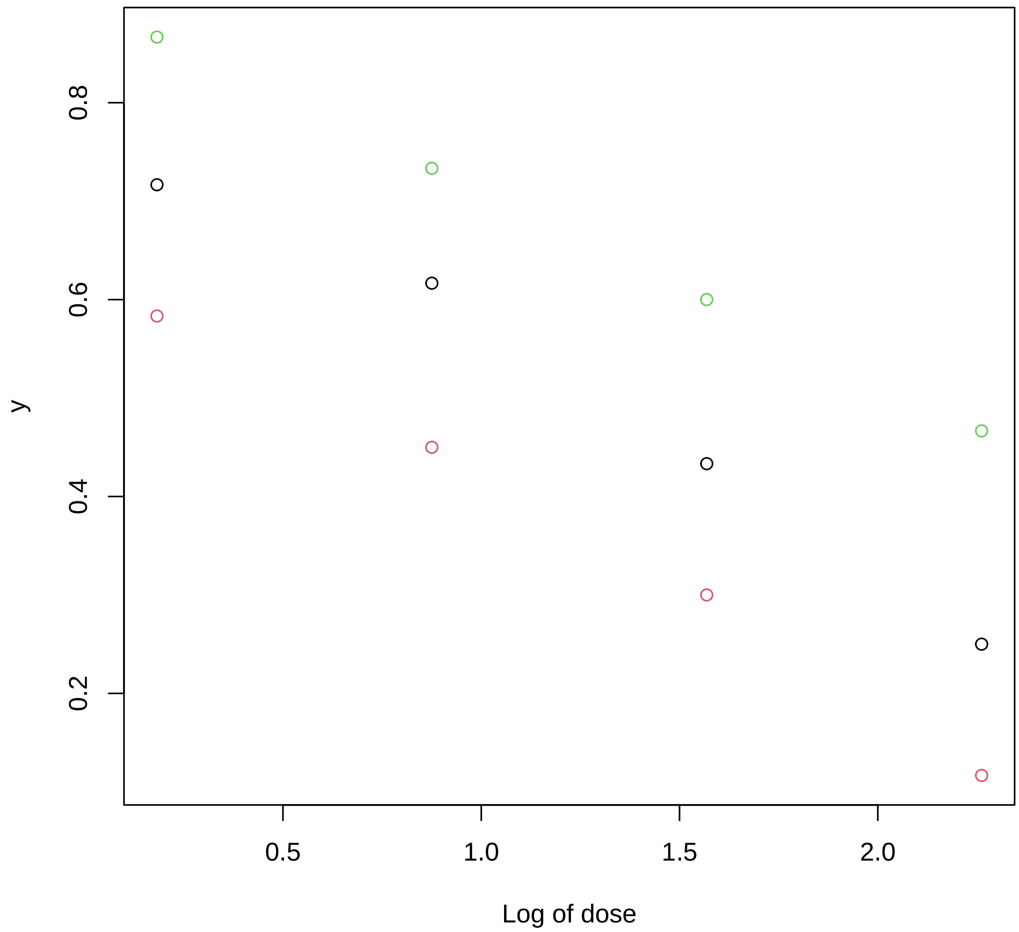
x	1.2	2.4	4.8	9.6	1.2	2.4	4.8	9.6	1.2	2.4	4.8	9.6
w	1	1	1	1	2	2	2	2	3	3	3	3
y	43	37	26	15	35	27	18	7	52	44	36	28

Analyse the data by fitting probit regression models in which the probit of the proportion of aphids killed by the insecticide is related to the logarithm of the dose.

1. Plot the proportions against the logarithms of the dose by insecticide. What are your conclusions?

Solution: A plot of the proportions against the logarithms of the dose by insecticide is given below.

```
x <- c(1.2,2.4,4.8,9.6,1.2,2.4,4.8,9.6,1.2,2.4,4.8,9.6)
w <- as.factor(c(1,1,1,1,2,2,2,2,3,3,3,3))
y <- c(43,37,26,15,35,27,18,7,52,44,36,28)/60
plot(log(x), y, col = w, xlab="Log of dose")
```



This shows that the regression lines for the three insecticides are roughly parallel.

2. By comparing the deviance for the model which allows a different intercept and slope for each insecticide with that for one in which the slopes are the same, test whether the regression lines are parallel.

Solution: The deviance for the model which allows a different intercept and slope for each insecticide is 1.6008 on six degrees of freedom, while that for the one in which the slopes are the same is 1.8554 on eight degrees of freedom. Thus, the difference in the deviances is 0.2546 on two degrees of freedom. Clearly, the p-value is $P > 0.1$, and so there is no evidence that the regression lines are not parallel.

```r
# same slope
mod1 <- glm(y ~ log(x) + w, family = binomial(link = "probit"),
            weights = rep(60,12))
summary(mod1)
```

```
##
## Call:
## glm(formula = y ~ log(x) + w, family = binomial(link = "probit"),
##     weights = rep(60, 12))
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.74989    0.11669   6.426 1.31e-10 ***
## log(x)      -0.60522    0.06602  -9.167  < 2e-16 ***
## w2          -0.40415    0.12069  -3.349 0.000812 ***
## w3           0.46753    0.12148   3.849 0.000119 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 135.7846  on 11  degrees of freedom
## Residual deviance:   1.8554  on  8  degrees of freedom
## AIC: 61.818
##
## Number of Fisher Scoring iterations: 4
```

```r
# different slope
mod2 <- glm(y ~ log(x) * w, family = binomial(link = "probit"),
            weights = rep(60,12))
summary(mod2)
```

```
##
## Call:
## glm(formula = y ~ log(x) * w, family = binomial(link = "probit"),
##     weights = rep(60, 12))
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.75193    0.16045   4.686 2.78e-06 ***
## log(x)      -0.60689    0.11154  -5.441 5.30e-08 ***
## w2          -0.36110    0.22364  -1.615   0.1064
## w3           0.40877    0.23889   1.711   0.0871 .
## log(x):w2   -0.03891    0.16106  -0.242   0.8091
## log(x):w3    0.04398    0.16058   0.274   0.7842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 135.7846  on 11  degrees of freedom
## Residual deviance:   1.6008  on  6  degrees of freedom
## AIC: 65.563
##
## Number of Fisher Scoring iterations: 4
```

```
# difference in deviance
anova(mod2)

## Analysis of Deviance Table
##
## Model: binomial, link: probit
##
## Response: y
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                         11    135.785
## log(x)    1   82.436         10     53.349
## w         2   51.494          8      1.855
## log(x):w  2    0.255          6      1.601

# critical value
qchisq(0.95,2)

## [1] 5.991465
```

3. Test whether there are differences between the insecticides.

Solution: The deviance for the model in which both the intercept and the slope are the same for each insecticide is 53.349 on 10 degrees of freedom. Thus, the difference in the deviances for this model and the previous one is 51.494 on two degrees of freedom. Since $\chi^2_{2,0.001} = 13.82$, the p-value is $P < 0.001$, and so there is overwhelming evidence of differences between the insecticides.

```
# critical region
qchisq(0.999,2)

## [1] 13.81551
```

41. Consider again the cloth data in Practical 6. Show that the estimate of the dispersion parameter ψ is $\hat{\psi} = 2.194$.

Solution: The fitted values are $\hat{\mu}_i = 0.0151x_i$ and the original variance of Y_i is $V(\hat{\mu}_i) = \hat{\mu}_i$. So the estimate of ψ is

$$\hat{\psi} = \frac{1}{31}\sum_{i=1}^{32}\frac{(y_i - 0.0151x_i)^2}{0.0151x_i} = 2.194.$$

This can be computed in R as follows.

```r
cloth <- read.csv("cloth.csv")
mod1 <- glm(y ~ x - 1,family=poisson(identity), data=cloth)
n <- dim(cloth)[1]
param <- mod1$coefficients[1]
dispersion<-sum(((cloth$y-param*cloth$x)^2)/(param*cloth$x))/(n-1)
print(dispersion)

## [1] 2.194371
```

42. The following relationships can be described by generalized linear models. For each one, identify the response variable and the explanatory variables, select a probability distribution for the response (justifying your choice) and write down the linear component.

    1. The effect of age, sex, height, mean daily food intake and mean daily energy expenditure on a person's weight.

       Solution: Response: Y_i = weight - continuous scale, possibly Normally distributed; Explanatory variables: x_{1i} = age, x_{2i} = sex (indicator variable), x_{3i} = height, x_{4i} = mean daily food intake, and x_{5i} = mean daily energy expenditure; $E(Y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i}; Y_i \sim \text{Normal}(\mu_, \sigma^2)$

    2. The proportions of laboratory mice that became infected after exposure to bacteria when five different exposure levels are used and 20 mice are exposed at each level.

       Solution: Response: Y_i = number of mice infected in each group of $n = 20$ mice; Explanatory variables: $x_{1i}, \ldots, x_{5i}$, as indicator variables for exposure levels; $Y_i \sim \text{Binomial}(n, \pi_i)$ because 'infection' is a binary outcome (but the plausibility of the assumption of independence of infection for mice depends on the experimental conditions); $g(\pi_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i}$

    3. The relationship between the number of trips per week to the super- market for a household and the number of people in the household, the household income and the distance to the supermarket.

       Solution: Response: Y_i = number of trips per week; Explanatory variables: x_{1i} = number of people in the household, x_{2i} = household income, x_{3i} = distance to supermarket; $Y_i \sim \text{Poisson}(\lambda_i)$ is a simple model for count data with $\log(\lambda_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$

43. In a cross-sectional study of skin cancer, the site of the tumour and its histological type were recorded for 400 patients. The contingency table below shows the number of patients (y) with each combination of tumour type and site.

Histological Type	Site Head and Neck	 Trunk	 Extremities	Total
Hutchinson's melanotic freckle	22	2	10	34
Superficial spreading melanoma	16	54	115	185
Nodular	19	33	73	125
Indeterminate	11	17	28	56
Total	68	106	226	400

The null hypothesis is that tumour type and site are independent.

1. Express the null hypothesis as a log-linear model, explaining your notation and any additional constraints.

Solution: The null hypothesis may be expressed as the log-linear model

$$\log\{E(Y_{jk}|N=n)\} = \mu + \alpha_j + \beta_k$$

with $\sum_{j=1}^{4} \alpha_j = 0$ and $\sum_{k=1}^{3} \beta_k = 0$, where μ is the overall effect, α_j is the effect of tumour type j and β_k is the effect of site k.

2. Obtain the expected values under the null hypothesis. Compare these with the observed values.

Solution: First of all, we input the data in R.

```
counts <- c(22,16,19,11,2,54,33,17,10,115,73,28)
site <- gl(3,4, length = 12)
type <- gl(4,1, length=12)
```

The expected values under the null hypothesis are $e_{jk} = y_{j.}y_{.k}/n$, where $n = 400$, and they can be computed as follows in R.

```
cancer <- glm(counts ~ site + type, family = poisson)
```

Look at the table with data and fitted values, written as column vectors.

```r
cbind(site, type, counts,fitted(cancer))
```

```
##    site type counts        
## 1     1    1     22   5.780
## 2     1    2     16  31.450
## 3     1    3     19  21.250
## 4     1    4     11   9.520
## 5     2    1      2   9.010
## 6     2    2     54  49.025
## 7     2    3     33  33.125
## 8     2    4     17  14.840
## 9     3    1     10  19.210
## 10    3    2    115 104.525
## 11    3    3     73  70.625
## 12    3    4     28  31.640
```

The column of fitted can be reformatted as matrix:

```r
matrix(nrow=4,fitted(cancer))
```

```
##       [,1]   [,2]    [,3]
## [1,]  5.78  9.010  19.210
## [2,] 31.45 49.025 104.525
## [3,] 21.25 33.125  70.625
## [4,]  9.52 14.840  31.640
```

Thus, we obtain the following table:

	Site			
Histological Type	Head and Neck	Trunk	Extremities	Total
Hutchinson's melanotic freckle	5.780	9.010	19.210	34
Superficial spreading melanoma	31.450	49.025	104.525	185
Nodular	21.250	33.125	70.625	125
Indeterminate	9.520	14.840	31.640	56
Total	68	106	226	400

By comparing these with the observed values, we see that cell $(1, 1)$ accounts for much of the lack of fit. Hutchinson's melanotic freckle is more common on the head and neck than under the assumed model.

3. Find the deviance and the value of Pearson's goodness-of-fit test statistic. What is your conclusion?

Solution: The deviance is

$$D = 2\sum_{j=1}^{4}\sum_{k=1}^{3} y_{jk}\log\left(\frac{y_{jk}}{e_{jk}}\right) = 51.795$$

and the value of Pearson's goodness-of-fit test statistic is

$$X^2 = \sum_{j=1}^{4}\sum_{k=1}^{3} \frac{(y_{jk} - e_{jk})^2}{e_{jk}} = 65.813.$$

These can be computed in R as follows.

```
# deviance, computed in two ways
deviance(cancer)

## [1] 51.79501

2*sum(counts*log(counts/fitted(cancer)))

## [1] 51.79501

# pearson's goodness of fit test, computed in two ways as well
sum((residuals(cancer,type="pearson")^2))

## [1] 65.81293

sum( (counts-fitted(cancer))^2/fitted(cancer)  )

## [1] 65.81293

# critical region
qchisq(0.999,6)

## [1] 22.45774

# or alternatively p-value
pchisq(q = 65.81,df = 6, lower.tail = FALSE)

## [1] 2.947238e-12
```

Since $\chi^2_{6,0.001} = 22.46$, the p-value is $P < 0.001$, and so there is overwhelming evidence that tumour type is not independent of site.

44. In a prospective study on a new treatment for pneumonia, patients were randomly allocated to two groups each of 40 patients. One group received the new treatment and the other the standard one, and the responses were the time taken to recover. The contingency table below shows the number of patients (y) with each combination of treatment and time taken to recover.

	Time to Recover			
	Short	Medium	Long	Total
Standard	6	15	19	40
New	10	21	9	40

The null hypothesis is that the time taken to recover is the same for each treatment group.

1. Express the null hypothesis as a log-linear model, explaining your notation and any additional constraints.

Solution: The null hypothesis may be expressed as the log-linear model

$$\log\{E(Y_{jk}|Y_{j.} = y_{j.})\} = \mu + \alpha_j + \beta_k$$

with $\sum_{j=1}^{2} \alpha_j = 0$ and $\sum_{k=1}^{3} \beta_k = 0$, where μ is the overall effect, α_j is the effect of treatment j and β_k is the effect of time to recover k.

2. Obtain the expected values under the null hypothesis. Compare these with the observed values.

Solution: The expected values under the null hypothesis are $e_{jk} = y_{j.}y_{.k}/n$, where $n = 80$. Thus, in R we obtain the following table:

	Time to Recover			
	Short	Medium	Long	Total
Standard	8	18	14	40
New	8	18	14	40

By comparing these with the observed values, we see that the time taken to recover is longer for the standard treatment. The R commands to obtain these result are the following.

```
counts <- c(6,10,15,21,19,9)
time <- gl(3,2, length = 6)
treatment <- gl(2,1, length=6)
mod1 <- glm(counts ~ time + treatment, family = poisson)
mod2 <- glm(counts ~ time, family = poisson)
fitted(mod2)
```

```
##  1  2  3  4  5  6
##  8  8 18 18 14 14
```

```
anova(mod1)
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: counts
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev
## NULL                          5    13.6602
## time       2   7.9934         3     5.6669
## treatment  1   0.0000         2     5.6669
```

3. Find the deviance and the value of Pearson's goodness-of-fit test statistic. What is your

conclusion?

Solution: The deviance is

$$D = 2\sum_{j=1}^{2}\sum_{k=1}^{3} y_{jk}\log\left(\frac{y_{jk}}{e_{jk}}\right) = 5.6669$$

and the value of Pearson's goodness-of-fit test statistic is

$$X^2 = \sum_{j=1}^{2}\sum_{k=1}^{3}\frac{(y_{jk}-e_{jk})^2}{e_{jk}} = 5.5714.$$

Since $\chi^2_{2,0.1} = 4.605$ and $\chi^2_{2,0.05} = 5.991$, the p-value is $0.05 < P < 0.1$, and so there is weak evidence that the time taken to recover is not the same for each treatment group. In R, this is as follows.

```
deviance(mod1)

## [1] 5.666881

sum(residuals(mod1, type="pearson")^2)

## [1] 5.571429

qchisq(0.9,2)

## [1] 4.60517

qchisq(0.95,2)

## [1] 5.991465

pchisq(q = deviance(mod1), df = mod1$df.residual, lower.tail = FALSE)

## [1] 0.05881016

pchisq(q = sum(residuals(mod1, type="pearson")^2), df = mod1$df.residual, lower.tail

## [1] 0.06168501
```

45. Suppose that $T_i \sim \text{Exp}(\lambda_i)$ for $i = 1, 2, \ldots, n$, all independent. Consider model M1, in which all the λ_i are parameters; and model M2 for which $\lambda = \lambda_1 = \cdots = \lambda_n$. That is, M1 and M2 are the maximal and the null models, respectively.

    1. Derive the corresponding maximum likelihood estimates for M1 and M2.
    2. Using your estimates, determine the maximum value of the (log) likelihood in each case, i.e. $\hat{l}_{max} = \hat{l}_{M1} = l(\hat{\lambda}_1, \ldots, \hat{\lambda}_n; y)$ and $\hat{l}_{null} = \hat{l}_{M2} = l(\hat{\lambda}; y)$.
    3. Derive an expression for the Fisher information matrix for the parameters in each of M1 and M2, and write formulæ for confidence intervals for parameters estimates in each case.
    4. Repeat all the computations above for the case when $T_i \sim \text{Exp}(\theta_i)$. Your analysis will consider models 1 and 2. Compare which results and comment on the similarities when they appear. **Hint**. When using $\text{Exp}(\lambda_i)$ in the first part of this problem, you will use the

density $f_T(t) = \lambda_i \exp(-\lambda_i t)$ for $t \geq 0$, while for $\text{Exp}(\theta_i)$, you'll use $f_T(t) = \theta_i^{-1} \exp(-t/\theta_i)$ for $t \geq 0$. In both cases the density is zero for negative t.

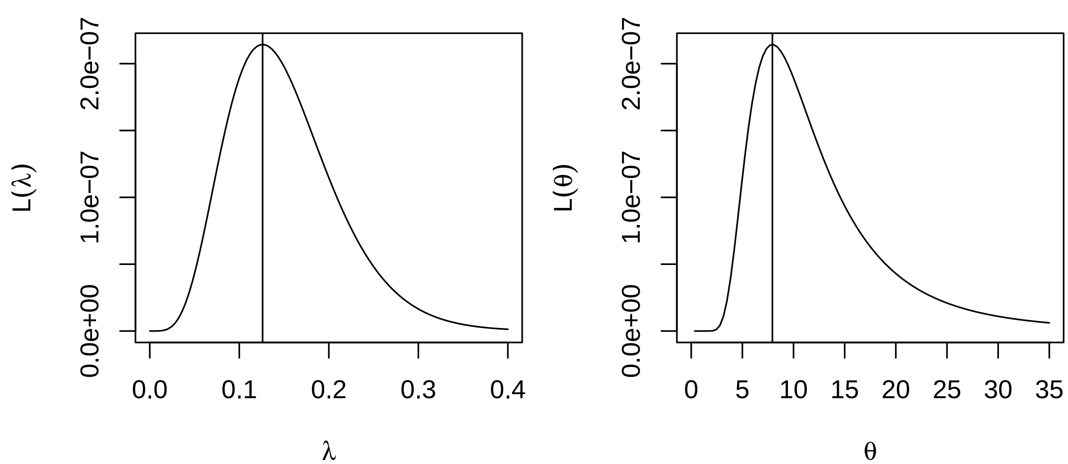
Solution: The likelihood estimates satisfy a useful property: if we have a maximum likelihood estimate $\hat{\beta}$, then the maximum likelihood estimate of a one to one transformation $\gamma = f(\beta)$ is $\hat{\gamma} = f(\hat{\beta})$. In this example, we have that parameters λ, θ are related as $\lambda = 1/\theta$, and we have $\hat{\lambda} = n/\sum_{i=1}^n t_i$ whose reciprocal is directly the mle $\hat{\theta} = 1/\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n t_i$. Additionally, the maximum value attained by the likelihood, that is, $L(\hat{\lambda}; y)$ is exactly equal to $L(\hat{\theta}; y)$.

46. The following values are lifetimes of electronic components: 1.86, 12.96, 13.74, 8.57, 2.54. Using the results of Exercise 45 or otherwise,

    1. Plot the likelihood function of the data under the model $T_i \sim \text{Exp}(\lambda)$. Do another plot of the likelihood using the model $T_i \sim \text{Exp}(\theta)$.

Solution: Assuming you have already loaded the data, i.e. `y<-c(1.86, 12.96, 13.74, 8.57, 2.54)`, the following commands will be used.

```
n<-length(y)
## likelihood for lambda
l1<-function(x) n*log(x)-x*sum(y); L1<-function(x) exp(l1(x=x))
## likelihood for theta
l2<-function(x) -n*log(x)-sum(y)/x; L2<-function(x) exp(l2(x=x))
par(mar=c(4,4,1,1),mfrow=c(1,2))
curve(expr=L1,from=0,to=0.4,xlab=expression(lambda),ylab=expression(L(lambda)))
abline(v=1/mean(y)) ## mle
curve(expr=L2,from=0,to=35,xlab=expression(theta),ylab=expression(L(theta)))
abline(v=mean(y)) ## mle
```



Note that the value $\hat{l}_{null}$ is exactly the same in both cases:

```r
l1(x=1/mean(y)) ## for the model with lambda
## [1] -15.35579
l2(x=mean(y)) ## for the model with theta
## [1] -15.35579
```

2. Give confidence intervals for each case. Use $\alpha = 0.05$.

Solution: Here is the confidence interval for λ, together with estimate.

```r
qnorm(p=c(0.025,0.5,0.975),mean=1/mean(y),sd=1/(mean(y)*sqrt(n)))
## [1] 0.01556308 0.12603983 0.23651658
```

And the confidence interval for θ, together with estimate

```r
qnorm(p=c(0.025,0.5,0.975),mean=mean(y),sd=mean(y)/sqrt(n))
## [1]  0.9796702  7.9340000 14.8883298
```

In the likelihood plots, the assymetry of the likelihood suggests that we should be cautious when using e.g. confidence intervals because higher values of the parameter are more likely than smaller values.

47. Suppose that $T_i \sim \text{Exp}(\lambda_i)$ for $i = 1, 2, \ldots, n$, all independent, where $\lambda_i = \beta x_i$ and x_i is a known covariate.

1. Write down the likelihood for the data $t_1, \ldots, t_n$.

Solution: The likelihood is

$$\begin{aligned} L(\beta; \mathbf{t}) &= \prod_{i=1}^{n} \beta x_i e^{-\beta x_i t_i} \\ &= \beta^n \left(\prod_{i=1}^{n} x_i \right) e^{-\beta \sum_{i=1}^{n} x_i t_i}. \end{aligned}$$

2. Show that the maximum likelihood estimator of β is $\hat{\beta} = n / \sum_{i=1}^{n} x_i T_i$.

Solution: The log-likelihood is

$$\ell(\beta; \mathbf{t}) = n \log \beta + \sum_{i=1}^{n} \log(x_i) - \beta \sum_{i=1}^{n} x_i t_i.$$

Thus, we have

$$\frac{d\ell}{d\beta} = \frac{n}{\beta} - \sum_{i=1}^{n} x_i t_i.$$

Setting this derivative to zero, we obtain

$$n - \hat{\beta} \sum_{i=1}^{n} x_i t_i = 0,$$

which yields the maximum likelihood estimate

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} x_i t_i}.$$

3. Find the Fisher information.

Solution: We have

$$\frac{d^2\ell}{d\beta^2} = -\frac{n}{\beta^2}.$$

It follows that the Fisher information is

$$v = \frac{n}{\beta^2}.$$

48. The break strength t_i in MPa was recorded for $n = 5$ industrial ceramic components. It is assumed that the distribution of strength is associated with porosity index x_i, which is a quantity controlled in the manufacturing process. The data is (1,6.798), (3,21.223), (3,1.873), (5,0.1), (5,0.398), which is given as pairs (x_i, t_i).

1. Using the model of Exercise 47, compute the maximum likelihood estimate of β.

Solution: The mle of β is

$$\hat{\beta} = \frac{5}{6.798 + 63.669 + 5.619 + 0.5 + 1.99} = 0.0636.$$

2. Compute its observed Fisher information.

Solution: The observed Fisher information is obtained by evaluating the Fisher information using the mle $\hat{\beta}$.

$$v = \frac{5}{0.004049} = 1234.8376.$$

3. Using your results and $\alpha = 0.05$, give a confidence interval for $\hat{\beta}$.

> **Solution:** The approximate variance of $\hat{\beta}$ is the reciprocal of the observed Fisher information, i.e. $v^{-1} = 8.1 \times 10^{-4}$; and the approximated standard error is $v^{-1/2} = 0.02846$. The confidence interval for $\hat{\beta}$ is thus $\hat{\beta} \mp 1.96 v^{-1/2} = (0.0079, 0.1194)$.

49. Suppose that the survival time $T > 0$ of a patient has a Weibull distribution with probability density function

$$f(t) = \alpha \lambda t^{\alpha - 1} \exp(-\lambda t^{\alpha}),$$

where $\alpha > 0$ and $\lambda > 0$.

1. Show that the survivor function is $S(t) = \exp(-\lambda t^{\alpha})$.

> **Solution:** The survivor function is
>
> $$S(t) = \int_t^{\infty} \alpha \lambda u^{\alpha - 1} \exp(-\lambda u^{\alpha}) du = \exp(-\lambda t^{\alpha}).$$

2. Obtain the hazard function.

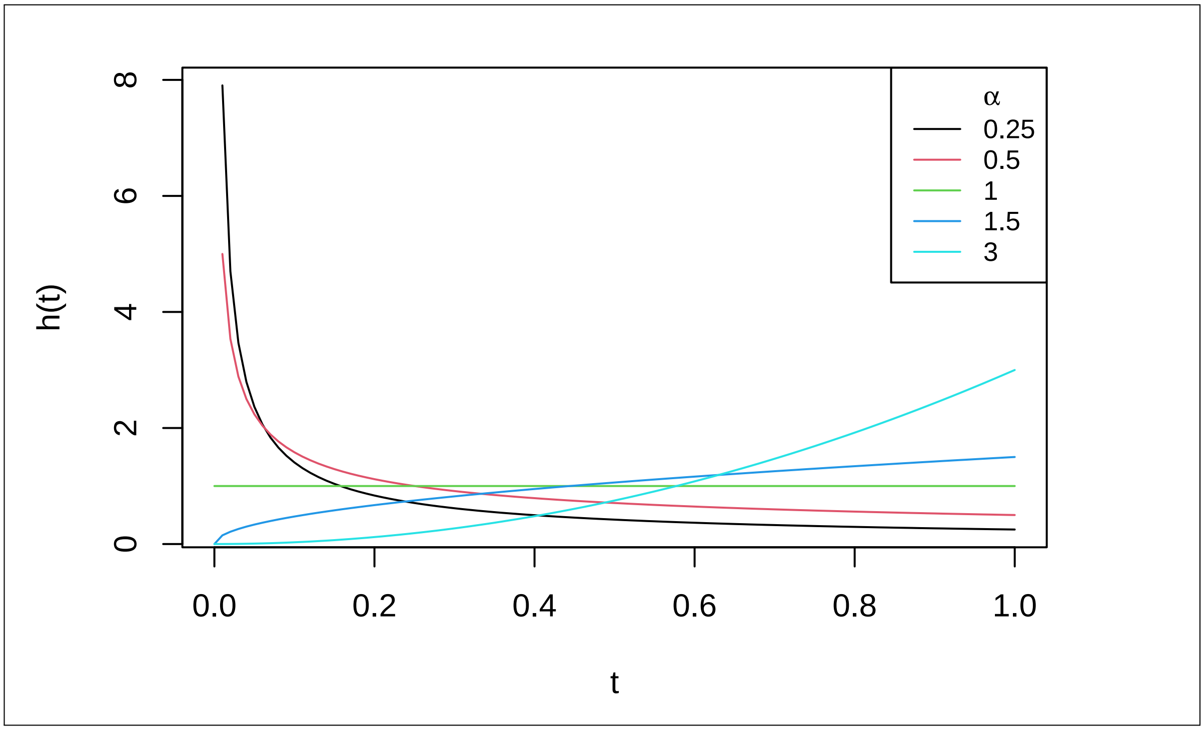
> **Solution:** The hazard function is
>
> $$h(t) = \frac{\alpha \lambda t^{\alpha - 1} \exp(-\lambda t^{\alpha})}{\exp(-\lambda t^{\alpha})} = \alpha \lambda t^{\alpha - 1}.$$

3. Explain how the hazard function behaves for different values of α.

> **Solution:** If $\alpha = 1$, the hazard function is constant. However, if $\alpha > 1$, it is monotonically increasing in t, whereas, if $\alpha < 1$, it is monotonically decreasing in t.

50. Using `R`, plot the hazard function for the Weibull distribution considering the scale parameter $\lambda = 1$ and the the following cases for the shape parameter $\alpha = 0.25, 0.5, 1, 1.5, 3$.

> **Solution:** For creating the plot, consider that the Weibull distribution implemented in `R` has two parameters: `scale` and `shape`. The scale parameter is σ which is defined as the reciprocal of λ as was used in Exercise 49. In this exercise we set this scale parameter to one, and vary the shape parameter as requested to produce the plot below for the range $t \in [0, 1]$. We see the monotonic behaviour of $h(t)$ as described in Exercise 49.



51. Consider a set of censored life times observations (δ_i, t_i) for $i = 1, \ldots, n$. Here δ_i indicates censoring, i.e. if $\delta_i = 1$ we observed $T_i = t_i$ and if $\delta_i = 0$ then we had $T_i > t_i$. Assume that the times t_i follow an exponential distribution.

    1. Using the null model, derive an expression for the mle of λ.
    2. Compute the Fisher information number and by plugging-in the mle $\hat{\lambda}$, give a formula for the observed Fisher information.
    3. Using the earlier results give a formula for the estimated standard error of $\hat{\lambda}$ and for a $100(1-\alpha)$ confidence interval for $\hat{\lambda}$.

52. Consider the data (1,0.62), (0,4.32), (1,4.58), (1,2.86), (1,0.85), (0,5.28), (1,0.1), (1,2.27), (1,21.22), (0,1.87), where the pairs are (δ_i, t_i) as above. Using the results of Exercise 51, estimate λ, compute its estimated variance and give a confidence interval for $\hat{\lambda}$ using $\alpha = 0.05$.

53. Consider the data (1,63.67), (0,5.62), (1,0.5), (1,1.99), (1,10.09), (0,13.44), (1,41.13), (1,28.24), (0,39.36), (1,17.59), (1,15.98), (0,13.05), where the pairs are (δ_i, t_i) as above. Using the results of Exercise 51, estimate λ, compute its estimated variance and give a confidence interval for $\hat{\lambda}$ using $\alpha = 0.05$.