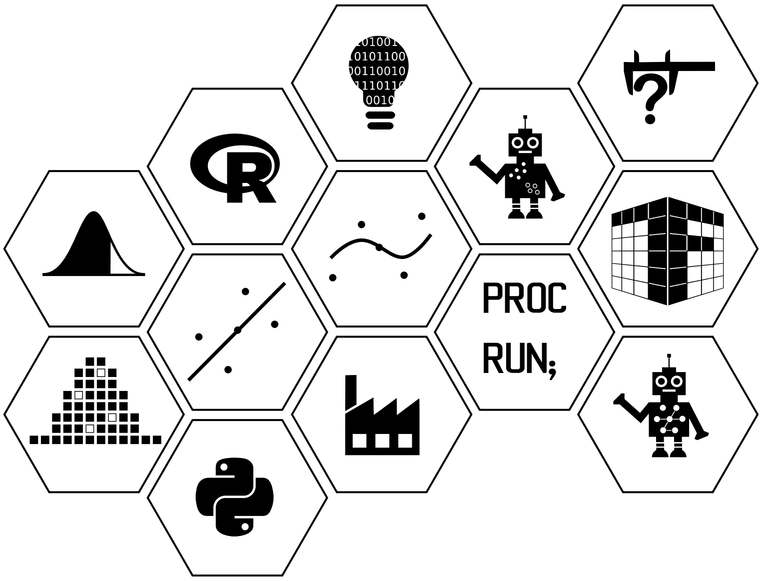
Learning from Data/Data Science Foundations

Week 7: Interval estimation for likelihood



DATA ANALYTICS GLASGOW

Interval Estimation for Likelihood

As we have already seen, point estimates obtained from different samples of data will generally take different values, and the point estimate calculated from a particular sample will not generally be equal to the unknown population parameter (though we might anticipate it being fairly close). Therefore, instead of relying on one number (i.e. our maximum likelihood estimate) as our point estimate, we can also use the sample data to identify a range of plausible values for the unknown population parameter. In this week's material we will consider how to construct an interval estimate using likelihood, and this provides a framework to generalise the ideas that we met in week 4, to enable us to construct confidence intervals for parameters of any known distribution.

Week 7 learning material aims

The material in week 7 covers:

- definitions of relative likelihood and relative log-likelihood;
- likelihood intervals;
- large sample properties to obtain confidence intervals with approximate coverage for MLEs;
- interpreting the results of these intervals.

The first video for this week provides an explanation and illustration of relative likelihood and uses the idea of relative log-likelihood to introduce the ideas of a likelihood interval and Wilks interval:

Video

Relative likelihood and likelihood/Wilks intervals

Duration 8:44

Relative likelihood

Suppose that the maximum likelihood estimate for θ is $\hat{\theta}_{MLE}$. Relative plausibilities of other θ values may be found by comparing the likelihood of those other values with the likelihood of $\hat{\theta}_{MLE}$.

For example, the figure below shows the likelihood for different values of θ and the MLE for an arbitrary example. We might be interested in how plausible a value of 0.5 or 0.7 is for θ?

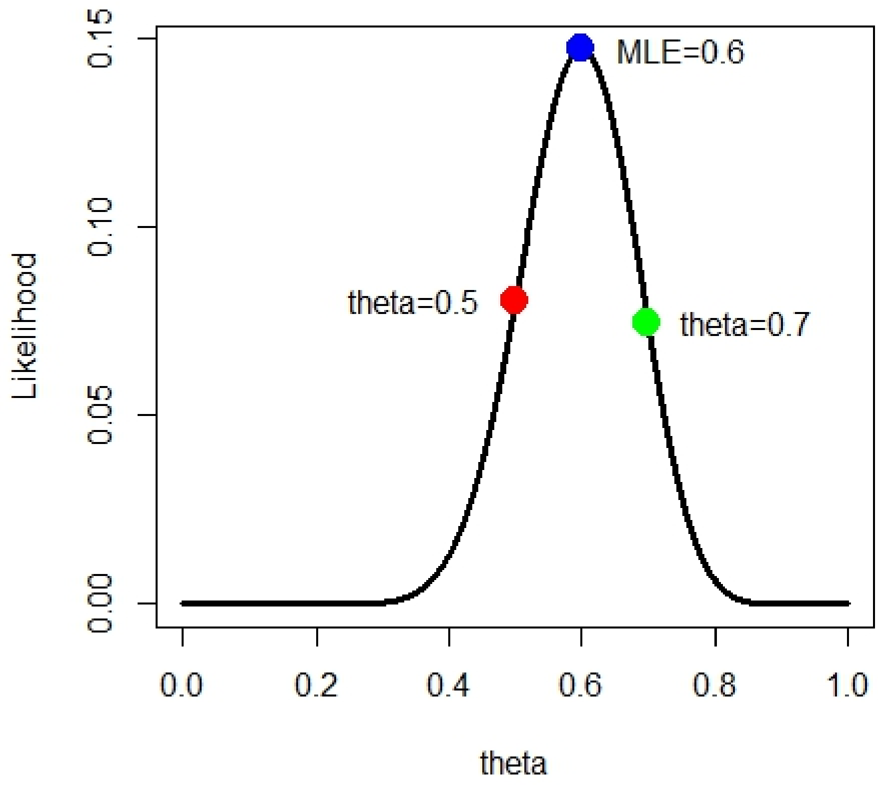


Figure 1

In order to consider the relative plausibility of other values of θ we define **relative likelihood** and the **relative log-likelihood**:

Definition 1

Relative likelihood

The **relative likelihood function** $(R(\theta))$ and the **relative log-likelihood function** $(r(\theta))$ are defined as

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta}_{MLE})}.$$

$$r(\theta) = \log_e \frac{L(\theta)}{L(\hat{\theta}_{MLE})} = \ell(\theta) - \ell(\hat{\theta}_{MLE}).$$

Example 1

Let's write down the relative likelihood for example 1 from the learning material in week 6. The details were:

A business wishes to monitor the useage of their website and so they record the time in days between hits on their website for 6 months:

1, 5, 15, 2, 3, 45, 13, 3, 3, 16, 23, 42, 4, 7, 4

Assuming that the data are independent observations from an exponential distribution with parameter θ; we found in week 6 that $\hat{\theta}_{MLE} = 0.081$.

Likelihood: for data $x_1, \ldots, x_n$

$$L(\theta) \propto \prod_{i=1}^{n} \theta\, e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}.$$

The **relative likelihood** function is defined by:

$$\begin{aligned} R(\theta) &= \frac{L(\theta)}{L(\hat{\theta}_{MLE})} \\ &= \frac{\theta^n e^{-\theta \sum x_i}}{\hat{\theta}^n_{MLE} e^{-\hat{\theta}_{MLE} \sum x_i}} = \frac{\theta^{15} e^{-186\theta}}{0.081^{15} e^{-0.081 \times 186}}. \end{aligned}$$

Task 1

Write down the relative log-likelihood for example 1.

Likelihood intervals

In week 4 we constructed confidence intervals for situations where we could assume normality or approximate normality for the distribution. We can use likelihood to construct (approximate) confidence intervals when we are working with any known distribution more generally. The starting point here is the idea of a likelihood interval.

In order to obtain an interval estimate for θ, we shall use the relative likelihood function, $R(\theta)$.

Since $\hat{\theta}_{MLE}$ is defined to be that value of θ that maximises $L(\theta)$, then $R(\theta)$ must also reach its maximum at $\hat{\theta}_{MLE}$.

We can use the relative likelihood to define an interval estimate for θ. A range of plausible values for θ consists of all θ such that:

$$L(\theta) \geq p \times L(\hat{\theta}_{MLE}),$$

i.e.

$$\frac{L(\theta)}{L(\hat{\theta}_{MLE})} \geq p,$$

i.e.

$$R(\theta) \geq p.$$

For p between 0 and 1, this is known as a **$100\mathbf{p}$% likelihood interval for $\boldsymbol{\theta}$**.

It is often easier to determine likelihood intervals using the relative log-likelihood function, $r(\theta)$.

In terms of $r(\theta)$, a **$100\mathbf{p}$% likelihood interval for $\boldsymbol{\theta}$** is defined by

$$r(\theta) \geq \log_e(p).$$

The question here is what is a suitable value for p and hence what is the interpretation?

In principle, $100p$% intervals for any p in the range 0 to 1 could be used. Let's just select any value for p for now to illustrate an example here. So suppose we select 0.5 for a 50% likelihood interval.

In particular, a 50% likelihood interval is defined by:

$$r(\theta) \geq \log_e(0.5) = -0.693.$$

Example 2

Plot of relative log-likelihood function

Let's plot the relative log-likelihood function for example 1/task 1 on times between hits on a website, and we'll add to the plot a line to help us obtain a 50% likelihood interval.

In `R`:

```
## The data
hittime <- c(1, 5, 15, 2, 3, 45, 13, 3, 3, 16, 23, 42, 4, 7, 4)

## The number of observations is given by:
n <- length(hittime)

## The maximum likelihood estimate (MLE) was found to be:
thetahat <- 1/mean(hittime)

## To plot the log-likelihood set up a sequence of values for theta around the MLE
theta <- seq(0.01, 0.2, length = 50)

## The log-likelihood can be found using:
loglik <- n * log(theta) - theta * sum(hittime)

## Relative log-likelihood can be found using:
relloglik <- n * log(theta) - theta * sum(hittime)-(n * log(thetahat) - thetahat * sum(hittime))

## This can then be plotted against theta:
plot(theta, relloglik, type = "l", lwd=3, ylab="relative log-likelihood")
abline(v=thetahat, lwd=3, col="yellow")
axis(3, at=c(thetahat))

## 50% likelihood interval
abline(h=-0.693, lwd=3, col="green", lty=2)
axis(4, at=c(-0.693))
```

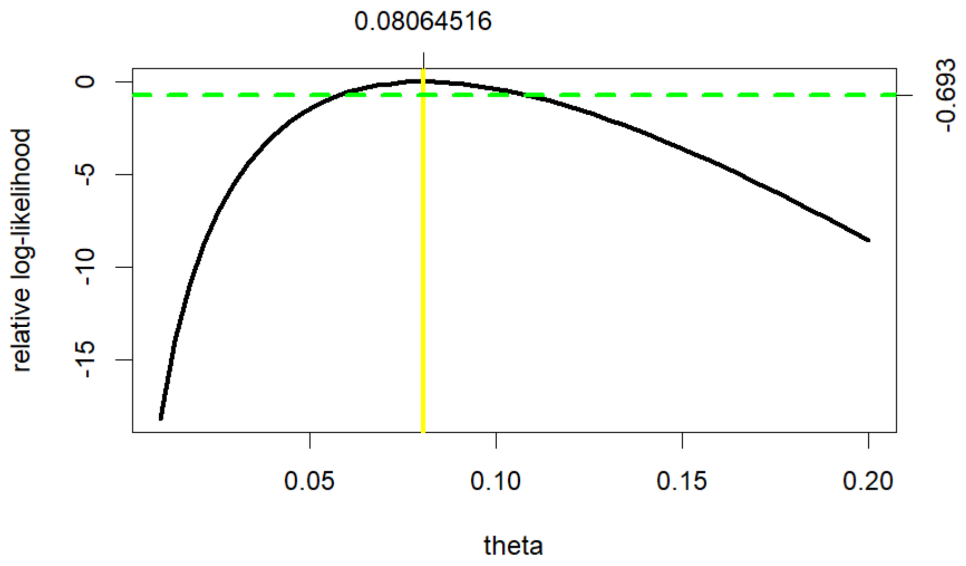


Figure 2

A plot of the relative log-likelihood function is displayed in the figure above. This illustrates that $\hat{\theta}_{MLE} = \frac{1}{\bar{x}} = 0.081$. The figure also has a horizontal line at -0.693, which can be used to obtain a 50% likelihood interval for θ. This interval can be found by obtaining the values of θ where the horizontal dashed line at -0.693 meets the relative log-likelihood function i.e. where $r(\theta) = -0.693$.

For example, it appears as though **a 50% likelihood interval for θ** is approximately: (0.055, 0.105).

This example, was just to illustrate the ideas of a likelihood interval. At this stage, we have not identified any particular reason for using $100p\%$ likelihood intervals for one value of p rather than another. Notice that, as p increases the width of the $100p\%$ likelihood interval decreases. We need additional properties to provide further guidance on an appropriate choice of p and to develop (approximate) confidence intervals, to enable us to interpret the intervals in a similar way to that of week 4.

Intervals based on approximate confidence

In the next two sections, we'll outline and apply results based on large sample properties of maximum likelihood estimators (i.e. based on the limiting behaviour as $n \to \infty$). These approximate distributional results will enable us to produce approximate 95% confidence intervals for parameters from any known statistical model/distribution. **We will use these results without proof in this week** to introduce the general ideas and we'll return to the properties behind these theoretical results in supplementary material for later weeks of this course.

Initially we can use the construction of a likelihood interval combined with large sample properties to introduce a **Wilks interval**.

Wilks Intervals

For a general parameter θ, when θ is set to its true value θ_T then, approximately,

$$2[\ell(\hat{\theta}_{MLE}) - \ell(\theta_T)] \sim \chi^2_1.$$

This means that the quantity $2[\ell(\hat{\theta}_{MLE}) - \ell(\theta_T)]$ is an approximate **pivotal quantity** for θ. The large sample distribution of $2[\ell(\hat{\theta}_{MLE}) - \ell(\theta_T)]$ is approximately χ^2.

A confidence interval stems from the result that

$$P\{2[\ell(\hat{\theta}_{MLE}) - \ell(\theta_T)] \leq \chi^2_1(c)\} \approx c,$$

where $\chi^2_1(c)$ denotes the cth quantile of the χ^2_1 distribution i.e. a χ^2_1 random variable will be less than $\chi^2_1(c)$ with probability c.

An approximate $100c\%$ confidence interval for θ is then defined by

$$\{\theta : 2[\ell(\hat{\theta}_{MLE}) - \ell(\theta)] \leq \chi^2_1(c)\}.$$

Notice that we can also define the confidence interval in terms of the relative log-likelihood function as

$$\{\theta : -2r(\theta) \leq \chi^2_1(c)\},$$

$$\{\theta : r(\theta) \geq -\frac{1}{2}\chi^2_1(c)\}.$$

This approximation is valid provided the sample size (i.e. the number of independent observations) is large.

For **95% confidence** this result is:

$$\{\theta : -2r(\theta) \leq \chi^2_1(0.95)\},$$

$$\{\theta : -2r(\theta) \leq 3.84\},$$

$$\{\theta : r(\theta) \geq -1.92\},$$

Note the construction here and connection to the likelihood interval, so we have:

$$\{\theta : r(\theta) \geq \log_e(p)\},$$

where p here is 0.1465.

Confidence intervals obtained by this method usually have to be found by numerical search. Such intervals are sometimes known as **Wilks** intervals, after the person who discovered the distributional result.

Example 3

Wilks interval

We will continue example 1, on times between hits on a website, to estimate the Wilks interval from a plot of the relative log-likelihood function.

In `R`:

```
## Using the R commands from  example 1, we can plot the relative log-likelihood against theta:
plot(theta, relloglik, type = "l", lwd=3, ylab="relative log-likelihood")
abline(v=thetahat, lwd=3, col="yellow")
axis(3, at=c(thetahat))

## Wilks approximate 95% confidence interval
abline(h=-1.92, lwd=3, col="cyan")
axis(4, at=c(-1.92) )
```

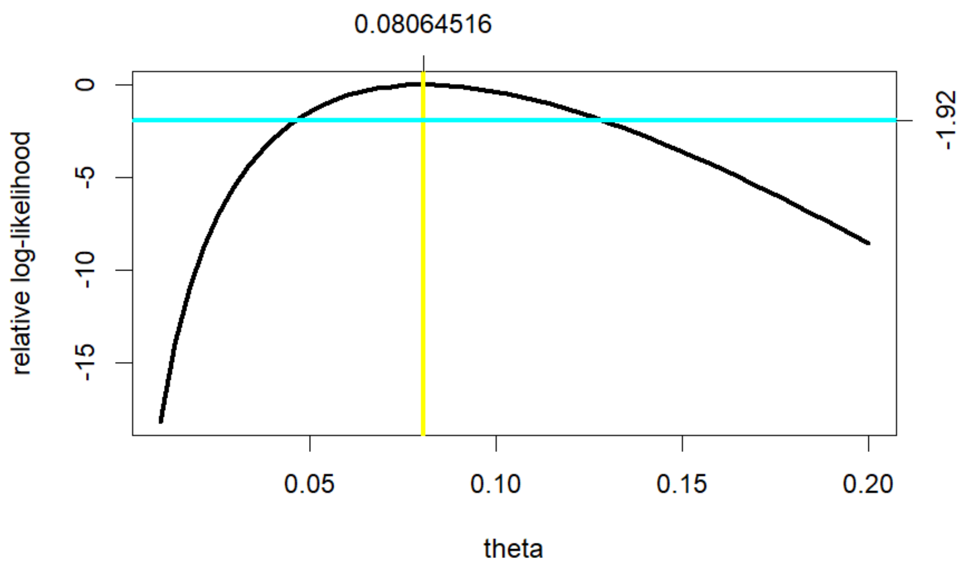


Figure 3

It can be seen from the figure that the **Wilks interval is approximately: (0.048, 0.130)**. This can be found from obtaining the values of θ where the horizontal line meets the function i.e. where $r(\theta) = -1.92$.

To obtain the Wilks interval with approximate confidence 0.95 the Newton-Raphson algorithm can be used, for example. Starting values can be estimated from the plot above for the lower and upper bounds of the interval i.e. 0.048 and 0.130 respectively.

For example, the following iterative algorithm could be used to find the bounds (B) for the interval, where θ_L is the lower bound and θ_U is the upper bound:

$$\theta_B^{(j+1)} = \theta_B^{(j)} - \frac{g(\theta_B^{(j)})}{g'(\theta_B^{(j)})},$$

with $g(\theta_B) = r(\theta) + 1.92$.

Supplement 1

One option for computing the bounds of the Wilks interval in `R` is to use the package `rootSolve` and the function `uniroot`. For example,

```
## Construct the function g() of interest
hittime <- c(1, 5, 15, 2, 3, 45, 13, 3, 3, 16, 23, 42, 4, 7, 4)

gfunction <- function(theta,ndat,thetahat,y){
  ndat * log(theta) - theta * sum(y)-(ndat * log(thetahat) -
thetahat * sum(y))+1.92

}
library(rootSolve)
uniroot.all(gfunction, interval=c(0.01, 0.15), ndat=15,
thetahat=0.081, y=hittime)
```

R Console

```
[1] 0.0464000 0.1286012
```

```
## Arguments here are: the function, g(),
## the end points for the interval to search over,
```

```
## and the data for the remaining arguments in the function,
where here sample size = ndat.
```

In this course we will just estimate Wilks intervals from the plot of the relative log-likelihood function.

Task 2

For the following example from week 5:

1. plot the relative log-likelihood function;
2. estimate the Wilks interval from the plot.

The numbers of car accidents at a fixed point on a motorway were recorded for 20 consecutive months. The results are as shown below, with x_i being the number of accidents at the fixed point in the ith month.

Month	**1**	**2**	**3**	**4**	**5**	**6**	**7**
No. of accidents	2	2	1	1	0	4	2

Month	**8**	**9**	**10**	**11**	**12**
No. of accidents	1	2	1	1	1

Month	**13**	**14**	**15**	**16**
No. of accidents	3	1	2	2

Month	**17**	**18**	**19**	**20**
No. of accidents	3	2	3	4

You can assume that these data follow a Poisson(λ) distribution. We are interested in the mean number of car accidents.

Note: we already found the MLE for λ by hand and derivations and plots of the likelihood and log-likelihood functions for λ in `R` in example 6 of week 5.

Wald Intervals

Wald intervals are generally much easier to compute.

The large sample distribution of a maximum likelihood estimate $\hat{\theta}$ is approximately normal with mean θ_T and variance $1/k(\mathbf{x})$. This result holds as the sample size tends to infinity. (Again, note that we'll look at more details for a proof of this in later weeks).

This result states that when θ_T is the true value of θ then, approximately,

$$\hat{\theta}_{MLE} \sim N(\theta_T, 1/k(\mathbf{x})).$$

Recall that $k(\mathbf{x})$ denotes the sample information, defined as $-l''(\hat{\theta}_{MLE})$. This means that, approximately,

$$\frac{\hat{\theta}_{MLE} - \theta_T}{\sqrt{1/k(\mathbf{x})}} \sim N(0,1),$$

and so the quantity $(\hat{\theta}_{MLE} - \theta_T)/\sqrt{1/k(\mathbf{x})}$ is an approximate pivotal quantity for θ.

This means that

$$P\{-z \leq \frac{\hat{\theta}_{MLE} - \theta_T}{\sqrt{1/k(\mathbf{x})}} \leq z\} \approx c,$$

where z denotes the point in a standard normal distribution below which lies probability $1-(1-c)/2$, namely $\Phi^{-1}(1-\frac{(1-c)}{2})$.

An approximate $100c$% confidence interval for θ is therefore given by

$$\left(\hat{\theta}_{MLE} - z\sqrt{1/k(\mathbf{x})},\ \hat{\theta}_{MLE} + z\sqrt{1/k(\mathbf{x})}\right).$$

Typically $c = 0.95$ and hence $z = \Phi^{-1}(0.975) = 1.96$.

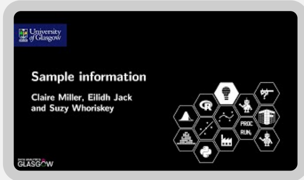
These are called **Wald** intervals, again after the person who proposed this approach.

Note:

$$\text{Var}\{\hat{\theta}_{MLE}\} \approx \frac{1}{k(\mathbf{x})}.$$

The second video for this week provides an explanation for the sample information, $k(\mathbf{x})$:

Video



Sample information

Duration 6:17

Wald intervals are usually very easy to obtain relative to Wilks intervals, but their properties are less satisfactory for finite sample sizes. In particular the intervals are not invariant, so that different parameterizations of a model lead to fundamentally different intervals. A related property is that the intervals are always symmetric, so that unless the log-likelihood is symmetric we will actually end up including some parameter values in the interval that are *less likely* than some values outside the interval. However, the Wald interval can be so much easier to obtain than the Wilks interval that it is sensible to use the approach.

An overview of the Wald interval and a comparison with the Wilks interval is provided in the video below:

Video

Wald intervals

Duration 5:08

Example 4

Wald interval

Let's find a 95% approximate Wald interval for the paramater θ in example 1, which was the example on length of time between hits on a website. We assumed that the data were independent observations from an exponential distribution with parameter θ.

Log-likelihood:

$$\ell(\theta) = n \log_e \theta - \theta \sum_{i=1}^{n} x_i,$$

$$\ell'(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i,$$

$$\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}},$$

$$\ell''(\theta_{MLE}) = -\frac{n}{\theta^2},$$

and this is < 0 for all $\theta > 0$.

So, we have

$$k(\mathbf{x}) = -l''(\hat{\theta}_{MLE}) = \frac{n}{\hat{\theta}^2}.$$

Since $\hat{\theta}_{MLE} = 0.081$, the **standard error** of $\hat{\theta}_{MLE}$ is

$$\sqrt{1/k(\mathbf{x})} = \sqrt{\frac{\hat{\theta}^2}{n}} = \sqrt{\frac{0.081^2}{15}} = 0.021.$$

The 0.975 quantile of the standard normal distribution is 1.96, as we saw in the week 4 learning material.

An approximate 95% confidence interval for θ is then

$$\left(\hat{\theta} - 1.96 \times 0.021, \hat{\theta} + 1.96 \times 0.021\right),$$

$$(0.040, 0.122).$$

It is therefore highly likely that θ lies in the range 0.040 to 0.122.

If we compare this to our estimates for the Wilks interval (from the plot of the relative log-likelihood) of 0.048 to 0.130 we can see that there is a bit of a difference between the two intervals - likely down to the slight assymetry of the distribution. However, the results are similar enough for the Wald interval to be useful.

Task 3

Compute an approximate 95% Wald interval for the car accidents example in task 2.

The numbers of car accidents at a fixed point on a motorway were recorded for 20 consecutive months. We were interested in the population mean number of accidents at this point in a month.

Assume these data follow a Poi(λ) distribution, and use the summaries below, which you already found in example 6 of week 5.

$\hat{\lambda}_{MLE} = \bar{x}$, $\ell''(\lambda) = -\frac{\sum_{i=1}^n x_i}{\lambda^2}$, $\bar{x} = 1.9$, $\sum_{i=1}^n x_i = 38.$

Supplement 2

You can explore visualisation for these intervals further by using the `rpanel` package. For example, for the website hit times example:

```
library(rpanel)
rp.likelihood("sum(log(dexp(data, theta)))", hittime, 0.01,
0.20)
```

The last two terms here are user selected lower and upper limits for the parameter of interest, here θ.

After running this, the first 4 radio buttons should be self explanatory. The 'threshold proportion' adds a horizontal line to the plot at $\log(p)$, you can change the value on the slider to illustrate different $100p$% likelihood intervals. The 'ci' radio button adds a Wilks interval and the 'quadratic approximation' radio button adds a Wald interval. (Note: you might have to click an icon that appears on your task bar to view the panel).

Learning outcomes for week 7

- define the terms *relative likelihood* and *relative log-likelihood*;
- define a likelihood interval;
- define a Wilks interval;
- provide approximate Wilks intervals from a plot of the relative log-likelihood, (which would be provided);
- define a Wald interval;
- compute Wald intervals;
- interpret the results of Wilks and Wald intervals.

Review exercises, selected video solutions and written answers to all tasks/review exercises are provided below.

Task 4

You will find data that are believed to follow a Poisson distribution, $X_i \sim \text{Po}(\lambda)$ for $i = 1, \ldots 100$, at http://www.stats.gla.ac.uk/~claire/RData_2021.html in Week 7 - review exercises (`R` Data file) in the object `poidata`.

- Plot the relative log-likelihood function in `R`. You can assume that a plausible range of values for λ is (0.5, 1.5).
- Estimate an approximate 95% Wilks interval using a plot of the relative log-likelihood.
- You can use the `optim` function in `R` to check your MLE for λ (optional).

Task 5

Given data $x_1, \ldots, x_n$ obtained as a random sample from the $\text{Ga}(2, \theta)$ distribution, in week 6 task 6 you were asked to find the maximum likelihood estimate of θ. Use the results from that task to derive an approximate 95% confidence interval for θ using the Wald method.

Task 6

Looking back to week 6 review task 4, we had data from price changes for 20 companies represented on the FTSE 100 index. We assumed that these data are from an independent random sample, and follow a $\mathrm{N}(0, \phi)$ distribution.

In week 6, we found the maximum likelihood estimate of ϕ to be,

$$\hat{\phi} = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

which was evaluated to be $\hat{\phi} = 0.144$.

Now, compute approximate 95% confidence intervals for ϕ by both the Wilks (only do this approximately using the figure below) and the Wald methods. For the Wald interval, use $\sum_{i=1}^{20} x_i^2 = 2.8865$ and $\ell''(\phi) = \frac{-n}{2\phi^2}$. Compare your answers to both intervals.

The data can be found at http://www.stats.gla.ac.uk/~claire/RData_2021.html in Week 7 - review exercises (`R` Data file) in the object `FTSE`.

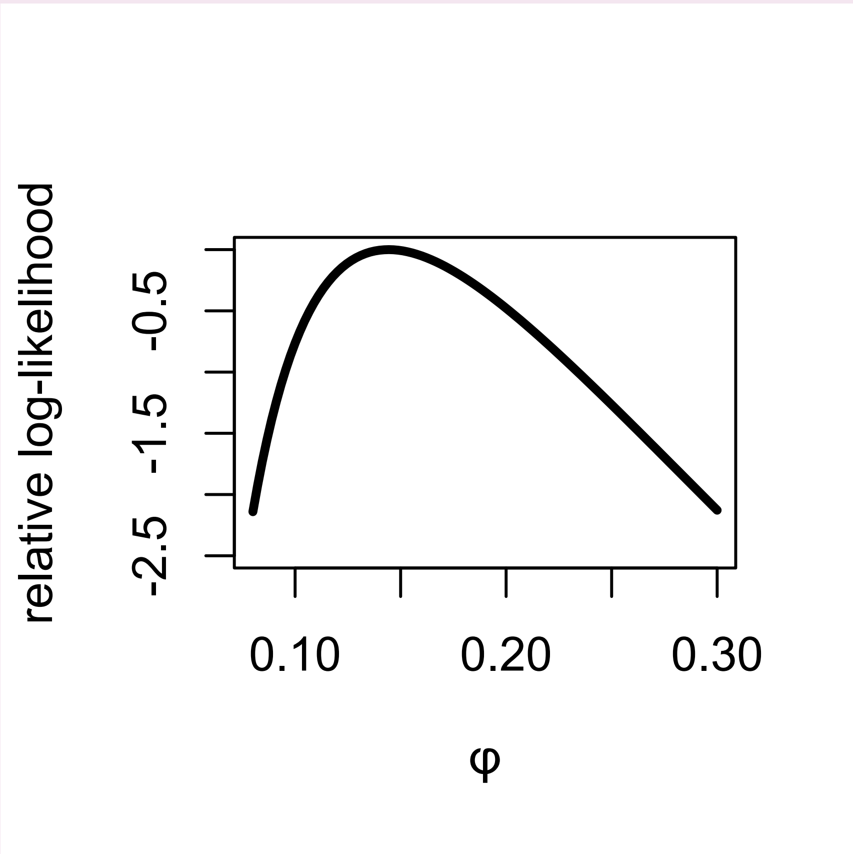


Figure 6

Answer 1

Log-likelihood: $\ell(\theta) = n \log_e \theta - \theta \sum_{i=1}^{n} x_i,$

The relative log-likelihood is given by:

$$
\begin{aligned}
r(\theta) &= \log_e \frac{L(\theta)}{L(\hat{\theta}_{MLE})}, \\
r(\theta) &= n \log_e(\theta) - \theta \sum_{i=1}^{n} x_i - n \log_e(\hat{\theta}_{MLE}) + \hat{\theta}_{MLE} \sum_{i=1}^{n} x_i, \\
r(\theta) &= 15 \log_e(\theta) - 186\theta - 15 \log_e(0.081) + 0.081 \times 186, \\
&= 15 \log_e(\theta) - 186\theta + 52.766.
\end{aligned}
$$

Answer 2

Car accidents:

```
## create a plausible sequence of values for the population
parameter lambda
lambda = seq(1,3,length=1000)

## the data - number of accidents in each month
x <- c(2,2,1,1,0,4,2,1,2,1,1,1,3,1,2,2,3,2,3,4)

## the sample size
n <- length(x)

## the likelihood function
lik <- (lambda^(sum(x))*exp(-n*lambda))/prod(factorial(x))

## the log-likelihood function
loglik <- sum(x)*log(lambda)-n*lambda

## the maximum likelihood estimate
lambdahat <- mean(x)
```

```
## the relative log-likelihood function
relloglik <- sum(x)*log(lambda)-n*lambda-
(sum(x)*log(lambdahat)-n*lambdahat)

## plotting the log-likelihood function, with a line added at
the MLE, and for 95% Wilks
plot(lambda,relloglik, ylab="relative log-likelihood",
     xlab=expression(lambda), type="l", lwd=3, cex.lab=1.5)
abline(v=lambdahat, col="yellow", lwd=3)
abline(h=-1.92, lwd=3, col="cyan")
```

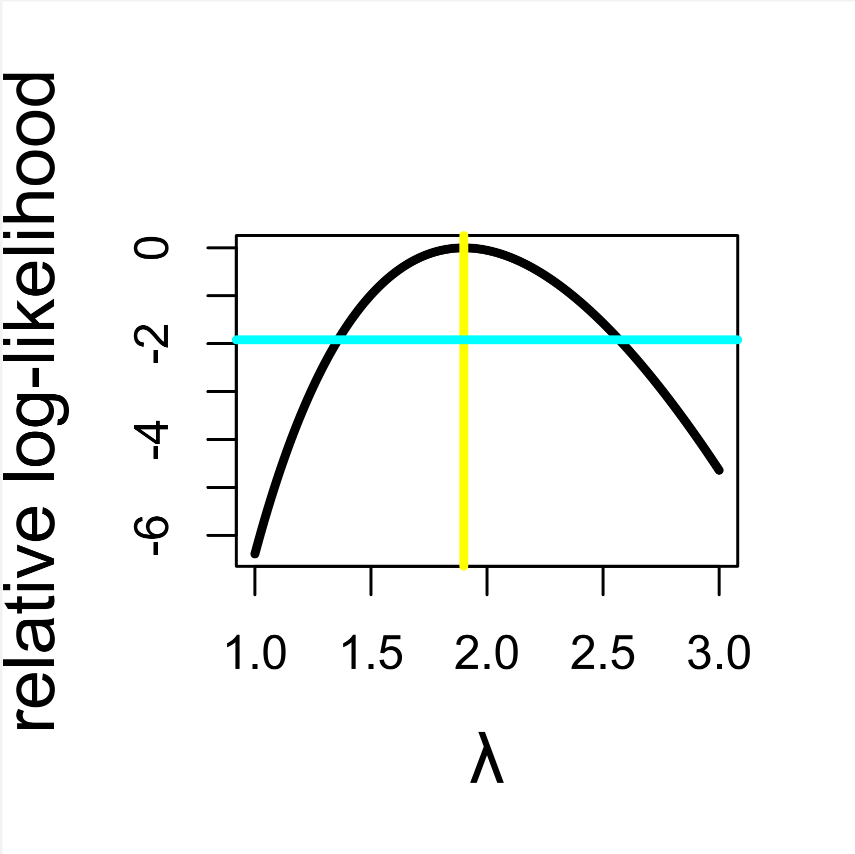


Figure 4

```
## estimated lower and upper bounds for Wilks CI
min(lambda[(relloglik>=-1.92)])
```

R Console

```
[1] 1.358358
```

```
max(lambda[(relloglik>=-1.92)])
```

R Console

```
[1] 2.56957
```

An approximate 95% Wilks interval is (1.36, 2.57). Therefore, it is highly likely that λ (the population mean number of accidents at the fixed point in a month) lies in the range 1.36 to 2.57.

```
## If you want to compute this interval in R
gfunction <- function(lambda,ndat,lambdahat,x){
  sum(x)*log(lambda)-ndat*lambda-(sum(x)*log(lambdahat)-
ndat*lambdahat)+1.92

}
library(rootSolve)
uniroot.all(gfunction, interval=c(1.3, 2.6), ndat=20,
lambdahat=1.9, x=x)
```

R Console

```
[1] 1.358299 2.569591
```

Answer 3

$$k(\mathbf{x}) = -l''(\hat{\lambda}_{MLE}) = \frac{\sum_{i=1}^{n} x_i}{\hat{\lambda}^2}$$

Since $\hat{\lambda}_{MLE} = 1.9$, the **standard error** of $\hat{\lambda}_{MLE}$ is

$$\sqrt{1/k(\mathbf{x})} = \sqrt{\frac{\hat{\lambda}^2}{\sum_{i=1}^{n} x_i}} = \sqrt{\frac{1.9^2}{38}} = 0.30822$$

The 0.975 quantile of the standard normal distribution is 1.96.

An approximate 95% confidence interval for θ is then

$$(\hat{\lambda} - 1.96 \times 0.30822, \hat{\lambda} + 1.96 \times 0.30822)$$

$$(1.9 - 1.96 \times 0.30822, 1.9 + 1.96 \times 0.30822)$$

$$(1.296, 2.504)$$

It is therefore highly likely that λ (and hence the population mean number of accidents in a month at this fixed point) lies in the range 1.30 to 2.50.

Answer 4

Poisson data example:

```
## The data
poidata <- read.csv("poidata.csv")[2]
## The number of observations is given by:
n <- dim(poidata)[1]
## The maximum likelihood estimate (MLE) was found to be:
lambdahat <- mean(poidata[,1])
## To plot the log-likelihood set up a sequence of values for
theta
## around the MLE
lambda <- seq(0.5, 1.5, length = 100)
## The log-likelihood can be found using:
loglik <- (-n * lambda) + (sum(poidata) * log(lambda))
## Relative log-likelihood can be found using:
relloglik <- (-n * lambda) + (sum(poidata) * log(lambda)) -
                   ((-n * lambdahat) + (sum(poidata) *
log(lambdahat)))
## This can then be plotted against lambda:
plot(lambda, relloglik, type = "l", lwd=3, ylab="relative log-
likelihood",
     xlab=expression(lambda))
abline(v=lambdahat, lwd=3, col="yellow")
## 95\% likelihood interval
abline(h=-1.92, lwd=3, col="green")
```

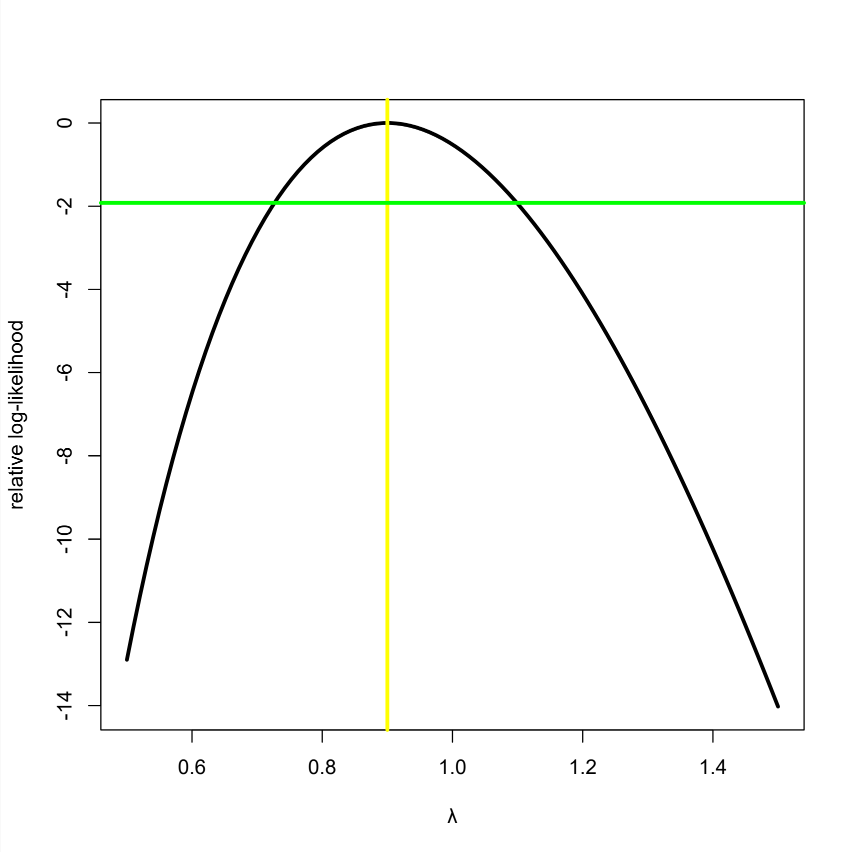


Figure 5

An estimate of a 95% Wilks Confidence Interval is (0.75, 1.1).

```
## Construct a function for the log-likelihood:
myfunction <- function(lambda, y, n){
  (-n * lambda) + (sum(y) * log(lambda))
}
## Optimise the function to estimate the parameter:
optim(par=0.9, fn=myfunction,
method="BFGS",control=list(fnscale= -1),
          y=poidata, n=100)
```

R Console

```
$par
[1] 0.9

$value
[1] -99.48245

$counts
function gradient
```

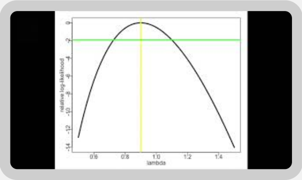
```
7        1

$convergence
[1] 0

$message
NULL
```

The MLE of λ is 0.9, which lies in our estimated Wilks interval.

Video

Video model answers for task 4

Duration 2:31



Answer 5

Model: $X_1, X_2, \ldots, X_n$ independent, with each $X_i \sim \mathrm{Ga}(2, \theta)$

Data: $x_1, x_2, \ldots, x_n$

$L(\theta) \propto \prod_{i=1}^n \theta^2 x_i \, e^{-\theta x_i} = \theta^{2n} \, e^{-\theta \sum_{i=1}^n x_i} \prod_{i=1}^n x_i$

$\ell(\theta) = 2n \log_e \theta - \theta \sum_{i=1}^n x_i + K$

$\ell'(\theta) = \frac{2n}{\theta} - \sum_{i=1}^n x_i$

$\ell'(\theta) = 0 \text{ when } \theta = \frac{2n}{\sum_{i=1}^n x_i} = \frac{2}{\bar{x}}$

$\ell''(\theta) = -\frac{2n}{\theta^2}$

and this is < 0 for all $\theta > 0$.

Therefore, $\hat{\theta}_{MLE} = \frac{2}{\bar{x}}$

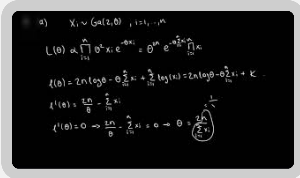
The sample information is $k(\mathbf{x}) = \frac{2n}{\hat{\theta}^2}$.

An approximate 95% Wald interval is therefore given by

$\hat{\theta} \pm 1.96\sqrt{\hat{\theta}^2/(2n)}$

i.e. $\hat{\theta}\left(1 \pm 1.96 \sqrt{1/(2n)}\right)$

Video



Video model answers for task 5

Duration 4:38

Answer 6

From the plot of $r(\phi)$, an approximate 95% Wilks interval is $(0.085, 0.29)$.

An approximate 95% Wald interval is given by

$\hat{\phi} \pm 1.96\sqrt{2\hat{\phi}^2/n}$

i.e. $0.144 \pm 1.96\sqrt{2(0.021)/20}$

i.e. 0.144 ± 0.0898

i.e. $(0.054, 0.234)$.

It is highly likely that ϕ will lie between 0.05 and 0.234.

The Wald interval is slightly lower than the Wilks one. This is because of the asymmetry in the log-likelihood function. We would probably therefore prefer the Wilks interval because it sticks more closely to the information in the log-likelihood function.