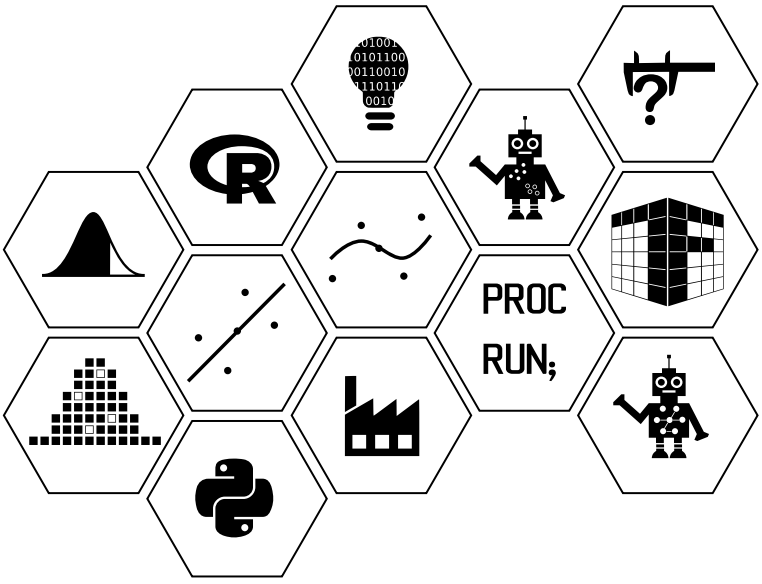
Learning from Data/Data Science Foundations

Week 10: Hypothesis testing using likelihood II



DATA ANALYTICS
GLASGOW

Hypothesis testing using likelihood II

In this final week of our introduction to likelihood we will explore large sample results for generalised likelihood ratio tests (GLRTs), the application and interpretation of such tests and the relationship of these tests to the classical test for multinomial data known as Pearson's χ^2 test.

Week 10 learning material aims

The material in week 10 covers:

- large sample properties for a GLRT;
- applying, and interpreting results from, a GLRT;
- deriving a GLRT for the multinomial distribution;
- the link between a GLRT and a classical χ^2 test.

The generalized likelihood ratio test (GLRT)

We described in week 9, that for simple hypotheses it is possible to find the exact distribution of the likelihood ratio statistic $\Lambda = L(\theta_{H_1})/L(\theta_{H_0})$ (or more usually $\lambda = \log_e \Lambda$), and hence calculate a p-value exactly. This is referred to as a **Likelihood Ratio Test**.

However, in most circumstances it is not possible to find the exact distribution and we have to use approximate p-values based on large sample results. Fortunately, a very general result exists for the log of the generalized likelihood ratio test statistic. Consider an observation $\mathbf{x}$ on a random vector of dimension n with p.d.f. (or p.m.f.) $f(\mathbf{x}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector. Suppose that we want to test,

$$H_0 : \mathbf{B}(\boldsymbol{\theta}) = \mathbf{0} \quad \text{vs.} \quad H_1 : \mathbf{B}(\boldsymbol{\theta}) \neq \mathbf{0},$$

where $\mathbf{B}$ is a vector valued function of $\boldsymbol{\theta}$ (e.g. $\theta_1 = \theta_2$, i.e. $\theta_1 - \theta_2 = 0$) such that H_0 imposes r restrictions on the parameter vector.

The **test statistic** here is: $\Lambda = L(\hat{\boldsymbol{\theta}}_{H_1})/L(\hat{\boldsymbol{\theta}}_{H_0})$,

i.e.

$$\lambda = \log_e \Lambda = \ell(\hat{\boldsymbol{\theta}}_{H_1}) - \ell(\hat{\boldsymbol{\theta}}_{H_0}).$$

If H_0 is true then in the limit as $n \to \infty$ (i.e. as the sample size tends to infinity)

$$2\lambda = 2(\ell(\hat{\boldsymbol{\theta}}_{H_1}) - \ell(\hat{\boldsymbol{\theta}}_{H_0})) \sim \chi^2_r,$$

where ℓ is the log-likelihood function, $\hat{\boldsymbol{\theta}}_{H_1}$ is the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ (under H_1) and $\hat{\boldsymbol{\theta}}_{H_0}$ is the value of $\boldsymbol{\theta}$ satisfying $\mathbf{B}(\boldsymbol{\theta}) = \mathbf{0}$ which maximises the likelihood (i.e. the restricted MLE).

We reject the null hypothesis for values of 2λ greater than $\chi^2_r(c)$, where c is our level of confidence between 0 and 1.

This result uses the same approximation as the Wilks interval described in weeks 7 and 9 for comparing the plausibility of different values of $\boldsymbol{\theta}$.

The first video for this week, explores the result above and describes how to determine the number of restrictions imposed on our parameters under H_0.

Video

Restrictions under H

Duration 7:40

Example 1

A simple null hypothesis and composite alternative hypothesis

Internet Banking

The number of people that log on to internet banking in 1 second (x_i) was recorded for one bank over 200, 1 second, intervals. The data are:

```
8 12  6 11  3  9  9  8  5  4  6 11  6 14  3  5 15 11  7  6  9
9 14 13
6 11  . . . . . . . . . . . . . . . . . . . .  9  8  5  8
9 14 14
```

The sum of the counts gives us $\sum_{i=1}^{200} x_i = 1800$.

Suppose that the banking industry mean count rate in a 1 second interval is 10. If the bank's customers are performing in a similar way to those of other banks then θ, the population mean count rate for the bank of interest, should be 10.

Is the bank performing in a similar way to other banks?

The counts can be treated as observations of independent identically distributed $\mathrm{Poi}(\theta)$ random variables with p.m.f.

$$p(x_i, \theta) = \frac{\theta^{x_i} e^{-\theta}}{x_i!}, \quad x_i \geq 0, \theta > 0.$$

We would test

$$H_0 : \theta = 10 \text{ versus } H_1 : \theta \neq 10.$$

The test can be performed using a **generalized likelihood ratio test**.

The **likelihood** is:

$$L(\theta) = \frac{\theta^{\sum x_i} e^{-n\theta}}{\prod_{i=1}^{n} x_i!},$$

and the **log-likelihood** is

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

where $n = 200$ and $\sum x_i = 1800$.

In this case H_0 completely specifies the value of θ, and **the maximised log-likelihood under $\mathbf{H_0}$** is (ignoring the constant K):

$$\ell(\hat{\theta}_{H_0}) = \ell(10) = \log_e(10) \sum_{i=1}^{n} x_i - n \times 10 = \log_e(10) \times 1800 - 200 \times 10 = 2144.653.$$

For the **maximum of the log-likelihood under $\mathbf{H_1}$** we have

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

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

which is < 0 for all $\theta > 0$ and so we have a maximum.

Therefore,

$$\hat{\theta}_{MLE} = \bar{x} = 1800/200 = 9,$$

and the **maximised log-likelihood under $\mathbf{H_1}$** is (ignoring the constant K)

$$\begin{aligned}\ell(\hat{\theta}_{H_1}) = \ell(\hat{\theta}_{MLE}) &= \log(\bar{x}) \sum x_i - \sum x_i \\ &= \log(9) \times 1800 - 1800 = 2155.004.\end{aligned}$$

Now calculate the **log-likelihood ratio test statistic**:

$$2\lambda = 2(\ell(\hat{\theta}_{H_1}) - \ell(\hat{\theta}_{H_0})),$$

$$2\lambda = 2(2155.004 - 2144.653) = 20.70.$$

The **rejection region** (RR) of the approximate test (at $c = 0.95$) is then:

$$RR = \{\mathbf{x} : 2\lambda > \chi^2_1(0.95)\}.$$

Remember: the rejection region is the set of values of the test statistic for which we reject H0.

H0 imposes 1 restriction on the parameter values, since $\theta = 10$, and hence $r = 1$ (where r is the number of restrictions on the parameters). Given that a χ^2_1 random variable has a 5% chance of being greater than or equal to 3.84, would you accept or reject H_0?

Conclusion:

We reject H_0 because this hypothesis does not look plausible in the light of the observed data.

The p-value is (from `R`):

```
1-pchisq(20.7, df=1)
```

```
R Console
[1] 5.3716e-06
```

Therefore, the bank customers do not appear to be operating in a similar way to customers from other banks.

The second video [1] for this week provides examples of how to compute the maximised log-likelihood for the situations of: (1) simple H0 and composite H1 and (2) composite H0 and H1.

Video

Computing the maximised log-likelihood under H

Duration 5:36

Task 1

How many restrictions under H_0 are imposed in the following scenarios?

Part 1

We are interested in time until the occurence of an event for 3 different companies, and we model this using:

$$X_1 \sim \text{Expo}(\gamma_1), \quad X_2 \sim \text{Expo}(\gamma_2) \quad X_3 \sim \text{Expo}(\gamma_3).$$

To compare these populations we want to test the following hypotheses:

H_0: $\gamma_1 = \gamma_2 = \gamma_3$

H_1: γ_i not all equal.

Part 2

For the example above, if the null hypothesis was:

H_0: $\gamma_1 = \gamma_2 = \gamma_3 = 4$.

Example 2

Composite null and alternative hypotheses

The following example was introduced in a task in week 8 and extended to find a confidence interval for a difference in the population proportions in week 9.

Patients with high blood pressure were randomly allocated to receive one of two treatments. The patients had been treated with cognitive behaviour therapy (CBT) and were

then given no further treatment (NFT) or CBT and beta-blockers (BB). Six weeks later it was noted whether or not each of the patients showed a decrease in their blood pressure. For the NFT group 15 out of 50 patients showed a decrease after 6 weeks, whereas this happened for 26 out of 50 of the BB patients.

Is there a difference in the population proportions experiencing an improvement in condition between the two treatment groups?

An appropriate model for the data is

$$X_1 \sim \text{Bi}(50, \theta_1),$$

$$X_2 \sim \text{Bi}(50, \theta_2),$$

where X_1 and X_2 denote the numbers of patient experiencing a decrease in the NFT and BB groups respectively.

The **hypotheses** are

$H_0 : \theta_1 = \theta_2,$

$H_1 : \theta_1 \neq \theta_2.$

The **likelihood function** is

$$L(\theta_1, \theta_2; \mathbf{x}) = K_1 \theta_1^{15} (1-\theta_1)^{35} K_2 \theta_2^{26} (1-\theta_2)^{24},$$

where K_1 and K_2 are constants.

The **log-likelihood** function is

$$\ell(\theta_1, \theta_2; \mathbf{x}) = 15 \log_e(\theta_1) + 35 \log_e(1-\theta_1) + 26 \log_e(\theta_2) + 24 \log_e(1-\theta_2) + K.$$

We found in week 8, that the maximum likelihood estimates are:

$$\hat{\theta}_1 = 15/50 = 0.30,$$

$$\hat{\theta}_2 = 26/50 = 0.52.$$

Therefore, the **maximised log-likelihood** under H1, ignoring the constant K, is

$$\ell(\theta_1, \theta_2; \mathbf{x}) = 15 \log_e(\hat{\theta}_1) + 35 \log_e(1 - \hat{\theta}_1) + 26 \log_e(\hat{\theta}_2) + 24 \log_e(1 - \hat{\theta}_2),$$

$$\ell(\theta_1, \theta_2; \mathbf{x}) = 15 \log_e(0.3) + 35 \log_e(0.7) + 26 \log_e(0.52) + 24 \log_e(0.48) = -65.16.$$

Under H_0 set θ_1 and θ_2 to be θ,

$$\ell(\theta; \mathbf{x}) = 15 \log_e(\theta) + 35 \log_e(1 - \theta) + 26 \log_e(\theta) + 24 \log_e(1 - \theta) = 41 \log_e(\theta) + 59 \log_e(1 - \theta),$$

$$\ell'(\theta) = \frac{41}{\theta} - \frac{59}{1 - \theta} = 0.$$

The MLE is $\hat{\theta}_{MLE} = \frac{15+26}{59+41} = 0.41$ and the **maximised log-likelihood** under H0, ignoring the constant K , is

$$\ell(\hat{\theta}_{H_0}) = 41 \log_e(\hat{\theta}) + 59 \log_e(1 - \hat{\theta}) = 41 \log_e(0.41) + 59 \log_e(0.59) = -67.69.$$

The **rejection region** of the test is

$$RR = \{\mathbf{x} : 2(\ell(\theta_{H_1}) - \ell(\theta_{H_0})) > \chi^2_r(1 - \alpha)\}.$$

Since there is one restriction i.e. $\theta_1 = \theta_2$, working at a significance level of $\alpha = 0.05$, we have that the **observed value of the test statistic** is

$$2(-65.16 - (-67.69)) = 5.06,$$

and, $\chi^2_1(0.95) = 3.84$.

Conclusion:

The observed value of our test statistic of 5.06 lies in the rejection region and so we reject the null hypothesis. We therefore conclude that, the population proportion experiencing an improvement under the two treatment groups is highly likely not to be the same.

An alternative way of analysing the data is to construct a confidence interval for the difference between the population means as we saw in week 9. This has the advantage of giving an indication of the size of any effect.

Supplement 1

We could find the maximised log-likelihood for H_0 and H_1 and the p-value for example 2 in R using the following code:

```
## The data
x1=15; x2=26; n1=50; n2=50

## Function to compute the log-likelihood under H0 and H1
loglik.bm <- function(par,n1,n2,x1,x2) {
  if (length(par)==1) {
    theta1 <- par; theta2 <- par
    } # under H_0
  else {
    theta1 <- par[1]; theta2 <- par[2]
    } # under H_1
  ll <- x1*log(theta1) + (n1-x1)*log(1-theta1) + x2*log(theta2)
+ (n2-x2)*log(1-theta2)
}

## Maximised log-likelihood under H1
m1 <- optim(c(0.3,0.52), loglik.bm, method = "BFGS", control =
list(fnscale=-1),
            n1=n1, n2=n2, x1=x1, x2=x2)

## Maximised log-likelihood under H0
m0 <- optim(0.41, loglik.bm, method = "BFGS", control =
list(fnscale=-1),
            n1=n1, n2=n2, x1=x1, x2=x2)

## GLRT statistic and comparison to chi-squared distribution
lambda <- m1$value - m0$value
p.value <- 1 - pchisq(2*lambda, df = 1)
p.value
```

R Console

```
[1] 0.02461766
```

Therefore, the p-value of 0.025 provides the same conclusion as the rejection region (working at 5% significance).

Task 2

Scientists were interested in comparing the conditions that experiments were being conducted in for two labs. In order to do this the number of bacteria in each of 12 samples from lab 1 were counted and the number of bacteria in each of 17 samples from lab 2 were counted. The results are shown below:

Lab 1 (X): `1164 1119 1415 964 711 773 579 977 792 1244 1115 1186`

Lab 2 (Y): `1382 883 1040 729 651 1227 1195 1370 1079 1089 1175 900 821 1075 1256 1134 918`

Are the labs operating under the same conditions?

As the data are counts, a Poisson probability model would be plausible. We therefore assume that

$$X_1, \ldots, X_{12} \sim \text{Poi}(\theta_1),$$

$$Y_1, \ldots, Y_{17} \sim \text{Poi}(\theta_2),$$

where all the observations are assumed to be independent.

Use a GLRT to compare θ_1 and θ_2 and hence comment on whether or not the labs appear to be working under the same conditions on the basis of these data.

GLRT for the Multinomial Distribution

The following describes data where a suitable model would be a multinomial probability distribution:

- each member of a population may be classified into one, and only one, category, according to some characteristic such as hair colour, eye colour, etc (i.e. categorical data);
- data are available in the form of frequencies which give the number of members of the sample which fall into each category;
- we wish to test a hypothesis which proposes a specific form for the underlying probabilities associated with the categories.

We will now develop a GLRT for this situation.

GLRT for a Multinomial Distribution

Data: $x_1, \ldots, x_k,$ where $\left(\sum_{i=1}^{k} x_i = n\right)$.

Model: $\mathbf{X} = \{X_1, \ldots, X_k\} \sim \text{Mu}(n; \theta)$ $\left(\sum_{i=1}^{k} \theta_i = 1\right)$.

A simple null hypothesis in this situation is:

$$H_0 : \theta_i = 1/k,$$

with H_1 : θ_i do not all equal $1/k$.

(More generally, H_0: The parameters $\theta_1, \ldots, \theta_k$ can be expressed in terms of p unknown parameters, $\phi_1, \ldots, \phi_p$, where $p < k - 1$, the number of unknown parameters in the general model above. The alternative hypothesis H_1 is expressed by the general multinomial model above.)

Under the general multinomial model, the **likelihood function** is

$$L(\boldsymbol{\theta}) = K \prod_{i=1}^{k} \theta_i^{x_i}.$$

The **log-likelihood function** is

$$\ell(\theta) = \log_e(K) + \sum_{i=1}^{k} x_i \log_e(\theta_i).$$

Under the general hypothesis H_1, and the constraint that $\sum_{i=1}^{n} \theta_i = 1$, it can be shown that this is maximised at

$$\hat{\theta}_i = x_i/n.$$

Supplement 2

To illustrate finding the MLE for the multinomial example above, consider the situation of 2 categories, we have x_1, and x_2 and $x_1 + x_2 = n$, θ_1 and θ_2 and $\theta_1 + \theta_2 = 1$.

The log-likelihood (ignoring the constant K) is:

$$\ell(\theta) = x_1 \log_e(\theta_1) + x_2 \log_e(\theta_2).$$

Differentiating and setting equal to zero, we get:

$$\frac{\partial \ell}{\partial \theta_1} = \frac{x_1}{\theta_1} = 0,$$

$$\frac{\partial \ell}{\partial \theta_2} = \frac{x_2}{\theta_2} = 0.$$

Therefore,

$$\frac{\partial \ell}{\partial \theta_2} = \frac{x_2}{\theta_2} = \frac{x_2}{1-\theta_1} = 0,$$

and hence,

$$\frac{x_1}{\theta_1} = \frac{x_2}{1-\theta_1}.$$

Therefore,

$$x_1 = (x_1 + x_2)\theta_1 = n\theta_1,$$

and

$$\hat{\theta}_1 = x_1/n,$$

and similary for $\hat{\theta}_2$.

Suppose we denote the ML estimates under H_0 as $\hat{\theta}_{01}, \ldots, \hat{\theta}_{0k}$.

The maximised log-likelihood under H_0 is then

$$\ell(\hat{\boldsymbol{\theta}}_{\mathbf{0}}) = \log_e(K) + \sum_{i=1}^{k} x_i \log_e(\hat{\theta}_{0i}).$$

The **GLRT statistic** is, as usual, twice the difference between the maximised log-likelihoods, namely

$$2\left(\sum_{i=1}^{k} x_i \log_e(x_i/n) - \sum_{i=1}^{k} x_i \log_e(\hat{\theta}_{0i})\right) = 2\sum_{i=1}^{k} x_i \log_e(x_i/(n\hat{\theta}_{0i})).$$

This can be written in the form

$$2\sum_{i=1}^{k} O_i \log_e\left(\frac{O_i}{E_i}\right),$$

where O_i and $E_i = n\hat{\theta}_i$ denote the observed frequencies and the expected frequencies respectively. The expected frequencies are computed under the assumption that the null hypothesis H_0 is true.

The large sample distribution of this test statistic is thus a χ^2 distribution with $k-1-p$ degrees of freedom. In the degrees of freedom, k refers to the number of levels (or categories) and since there is a constraint on the parameters under the model (i.e. $\sum \theta_i = 1$) we subtract 1. We also subtract the number of parameters estimated under H_0, denoted as p.

The **rejection region** of the approximate test is then

$$RR = \left\{\mathbf{x} : 2\sum_{i=1}^{k} O_i \log_e\left(\frac{O_i}{E_i}\right) > \chi^2_{k-1-p}(1-\alpha)\right\}.$$

Example 3

GLRT for Multinomial

Sales

An interest rate product sales person has 5 products to offer customers who have agreed to buy a product from that bank. The sales person met 30 customers over a month and the number of customers who bought each product was recorded. The data are displayed in the table below i.e. 6 customers bought product 1, and 2 customers bought product 5. The sales person is told to promote all products equally.

Product	**1**	**2**	**3**	**4**	**5**
Frequency	6	14	4	4	2

The null hypothesis of customers randomly deciding to buy a particular product leads to the assumption that the cell probabilities are all $\theta_i = 1/5$.

Do any of the products perform better than random chance?

The observed and expected cell counts are therefore

Product	1	2	3	4	5
Frequency	6	14	4	4	2
Expected	6	6	6	6	6

The **test statistic** is:

$$2\sum_{i=1}^{k} O_i \log_e\left(\frac{O_i}{E_i}\right),$$

and hence the **observed value of the test statistic** is:

$$2(6\log_e(6/6) + 14\log_e(14/6) + 4\log_e(4/6) + 4\log_e(4/6) + 2\log_e(2/6)) = 12.84.$$

Rejection Region

Since there are 5 categories and 0 parameters have been estimated under H0 (the values of θ are completely specified under H0), the critical value is $\chi^2_{5-1-0}(0.95) = 9.49$, for $\alpha = 0.05$.

Conclusion

Since $12.84 > 9.49$, we therefore have sufficient evidence to reject the null hypothesis that the products are being bought at random. This suggests that some products are performing better than by chance.

For info, the p-value for this example is:

```
1 - pchisq(12.84, df = 4)
```

R Console

```
[1] 0.01208463
```

Task 3

Gene Frequencies

Suppose that we have blood types for 1029 people coded as below:

Blood Type	M	MN	N	Total
Observed Frequency	342	500	187	1029

Use the GLRT statistic for multinomial data above to investigate the following hypotheses:

H_0: θ_i = 1/3,

H_1: θ_i do not all equal 1/3

What can you conclude about the distribution of blood types?

Pearson's χ^2 test

The GLRT statistic is in fact approximately equivalent to what is often regarded as the *usual* or classical statistic for tests on categorical data, Pearson's χ^2 statistic. This is defined by:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i},$$

where O_i is what we observe and E_i is what we expect under H_0. As with the GLRT, this test statistic converges in distribution to a χ^2_{k-1-p} random variable, where k=no. of intervals/categories and p=no. of parameters estimated under H_0.

Pearson's statistic and the likelihood ratio statistic are asymptotically equivalent under H_0. Pearson's test has been more commonly used than the likelihood ratio test, since it is somewhat easier to compute by hand, and is the basis of a classical χ^2 test.

Example 4

Pearson's χ^2 statistic

Let's compute the Pearson's χ^2 statistic for example 3.

The observed and expected cell counts are therefore

Product	1	2	3	4	5
Frequency	6	14	4	4	2
Expected	6	6	6	6	6

The **observed value of the test statistic** is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i},$$

$$\chi^2 = \frac{(6-6)^2}{6} + \frac{(14-6)^2}{6} + \frac{(4-6)^2}{6} + \frac{(4-6)^2}{6} + \frac{(2-6)^2}{6} = 10.67 + 0.67 + 0.67 + 2.67 = 14.68.$$

Rejection Region

Since there are 5 categories and 0 parameters have been estimated under

H0, the critical value is $\chi^2_{5-1-0}(0.95) = 9.49$, for $\alpha = 0.05$ and we have the same conclusion as before, with p-value:

```
1-pchisq(14.68, df=4)
```

R Console

```
[1] 0.005413081
```

Supplement 3

The above test can be carried out in `R` using the following commands:

```
Obs <- c(6, 14, 4, 4, 2)
chisq.test(Obs)
```

R Console

```
Chi-squared test for given probabilities

data:  Obs
X-squared = 14.667, df = 4, p-value = 0.005445
```

Pearson's χ^2 test is more classically used for contingency tables where you might be interested in testing:

(a) for association between 2 categorical variables e.g. is there an association between hair colour and eye colour?; (b) if the population proportions are the same across multiple populations for the same categorical variable i.e. an extension of the proportion test that we used in week 4.

Supplement 4

For more details on the tests in (a) and (b) above, which were mentioned briefly in week 4 as well, see examples in Chapter 13 of **Statistical inference: a short course (Panik, 2012)** and the function `chisq.test` in `R`.

Summary of weeks 5-10

In weeks 5 to 10 for this course we have introduced the ideas of likelihood, and interval estimation and hypothesis testing for likelihood. The main aim here is that if we can assume a known probability distribution for our data then we know everything about our data apart from the parameters of the statistical distribution, and hence statistical model, that we have assumed. The problem is then reduced to estimating and making inferences on these population parameters. We have illustrated that the method of maximum likelihood is an effective way to obtain point estimates for these parameters and that exact inference can be performed for the normal distribution to produce 95% confidence intervals.

The large sample properties and subsequent results that we have introduced have enabled us to develop approximate 95% confidence intervals and perform hypothesis testing using likelihood for any know statistical distribution/model.

These ideas enable the results that you have used so far in *predictive modelling* (which have assumed normality) to be developed for more sophisticated models in later parts of that course and in the course on *advanced predictive models* to follow. We have focussed so far on the classical statistical approach often referred to as **frequentist statistical inference**. However, please see the **supplementary material** on **Bayesian inference** which was released in Week 7 if you are interested in an introduction to the philosophy and preliminary ideas for Bayesian inference, for which likelihood also plays a pivotal role.

Learning outcomes for week 10

- construct, use and interpret results from the generalised likelihood ratio test;
- state and use the generalised likelihood ratio test statistic for a multinomial distribution;
- frame statistical conclusions from hypothesis tests clearly.

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

Review exercises

Task 4

Returning to Task 7 in Week 5, we were looking at a problem involving a Geiger counter. This is a device which is used to measure radioactivity. In order to check that the device is calibrated correctly, a measurement can be taken from a source of known radioactive strength. The counts recorded by the Geiger counter over 200 one second intervals were recorded and these can be represented by the random variables $X_1, \ldots, X_{200}$.

For this particular radioactive source, if the Geiger counter is functioning correctly then the mean count in a one second interval should be 20. The sum of the observed counts was $\sum_{i=1}^{200} x_i = 3654$.

In Week 5, we assumed that the counts $X_1, \ldots, X_{200}$ each followed a $\mathrm{Po}(\theta)$ distribution and we found that the log-likelihood function for θ was

$$L(\theta) = \prod_{i=1}^{n} \theta^{x_i} e^{-\theta} / x_i! = \theta^{\sum_i x_i} e^{-n\theta} / \prod_i x_i!$$

and the maximum likelihood estimate of θ was shown to be $\hat{\theta}_{MLE} = \frac{3654}{200} = 18.27$.

- Define the *sample information* in the context of estimating θ. Use this to construct an approximate 95% confidence interval by the Wald method.

- Consider the hypotheses $H_0 : \theta = 20$ and $H_1 : \theta \neq 20$. Carry out a generalised likelihood ratio test (GLRT) for the hypotheses, compute the associated *p*-value in `R`, and report your conclusions. Hint: Use `1 - pchisq(GLRT value, df)` to calculate your *p*-value.

- Comment on the relative merits of checking the calibration of the Geiger counter by a confidence interval and by a hypothesis test. Which would you recommend?

Task 5

A team of geneticists, who were interested in the effect of the chemical DES on the development of abnormal cells in humans, grew three different independent samples in tissue culture, each at a different dose level. At metaphase, the number of chromosomes in each cell was counted and each cell was classified as normal or abnormal.

The following data were recorded.

Dose (μg/dl)	Number of abnormal cells	Sample size
5	21	99
15	31	100
25	42	99

Use the fact that the experiments are independent to combine the likelihoods for each sample: i.e. dose 5, 15 and 25 (note that cells are classified as either normal **or** abnormal at each dose and hence we have three independent Binomial experiments). Hence, construct a GLRT of the null hypothesis that the population proportions of cells which are abnormal are equal across the three doses, against the alternative hypothesis that there are differences. Apply your test to the data provided, compute the associated p-value in `R`, and report your conclusions.

Task 6

A company is considering the marketing strategy for a newly developed product. Part of this involves the choice of a particular design for the packaging of the product. Five different designs have been constructed and tested on potential customers. Each of 100 customers was asked to select their preferred design. The results are shown below.

Design	**1**	**2**	**3**	**4**	**5**
Number of times selected	18	14	24	29	15

Since the data are categorical, use the test statistic:

$$2\sum_{i=1}^{k} O_i \log_e\left(\frac{O_i}{E_i}\right)$$

to construct a GLRT of the null hypothesis that a customer is equally likely to choose any of the designs (i.e. probability of selecting a design = 1/5) against the alternative hypothesis that the probabilities of selection for the designs are not all equal. (You may assume that the data $(X_1, X_2, X_3, X_4, X_5)$ arise from a multinomial distribution).

Apply your test to the data, compute the associated *p*-value in `R`, and report your conclusions.

Answer 1

Number of restrictions:

Part 1

$\gamma_1 = \gamma_2 = \gamma_3$

Therefore, $\gamma_1 - \gamma_3 = \gamma_2 - \gamma_3 = 0$

i.e.

$$\mathbf{B}(\boldsymbol{\theta}) = \begin{pmatrix} \gamma_1 - \gamma_3 \\ \gamma_2 - \gamma_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Therefore, we have two restrictions under the null hypothesis.

Part 2

We now have: $\gamma_1 = \gamma_2 = \gamma_3 = 4$

Therefore, $\gamma_1 - 4 = \gamma_2 - 4 = \gamma_3 - 4 = 0$

i.e.

$$\mathbf{B}(\boldsymbol{\theta}) = \begin{pmatrix} \gamma_1 - 4 \\ \gamma_2 - 4 \\ \gamma_3 - 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Therefore, we have three restrictions under the null hypothesis.

Answer 2

As the two labs are independent, the **likelihood function** is

$$L(\theta_1, \theta_2) = \prod_{i=1}^{12} \frac{\theta_1^{x_i} e^{-\theta_1}}{x_i!} \prod_{i=1}^{17} \frac{\theta_2^{y_i} e^{-\theta_2}}{y_i!}$$

$$L(\theta_1, \theta_2) = K e^{-12\theta_1} \theta_1^{\sum_{i=1}^{12} x_i} e^{-17\theta_2} \theta_2^{\sum_{i=1}^{17} y_i}$$

The **log-likelihood** is then

$$\ell(\theta_1, \theta_2) = \log_e(K) - 12\theta_1 + (\sum_{i=1}^{12} x_i) \log_e(\theta_1) - 17\theta_2 + (\sum_{i=1}^{17} y_i) \log_e(\theta_2)$$

The **hypotheses** are:

$H_0 : \theta_1 = \theta_2$

$H_1 : \theta_1 \neq \theta_2$

Also,

$\sum_{i=1}^{12} x_i = 12039$ and $\sum_{i=1}^{17} y_i = 17924$

We saw in example 1 that for Poisson likelihoods $\hat{\theta}_{MLE_i} = \sum_{i=1}^{n_j} x_i/n_j, \quad j = 1, 2.$

Therefore, under H_1, the MLE's are

$\hat{\theta}_1 = 12039/12 = 1003.25$ and

$\hat{\theta}_2 = 17924/17 = 1054.35$

and the **maximised log-likelihood**, under H_1, ignoring the constant $\log(K)$, is

$$\ell(\hat{\theta}_{H_1}) = -12\hat{\theta}_1 + \sum x_i \log_e(\hat{\theta}_1) - 17\hat{\theta}_2 + \sum y_i \log_e \hat{\theta}_2$$

$\ell(\hat{\theta}_{H_1}) = -12 \times 1003.25 + 12039 \log_e(1003.25) - 17 \times 1054.35 + 17924 \log_e(1054.35) = 178001.8$

Under H_0,

$$\ell(\theta) = \log_e(K) - 12\theta + (\sum_{i=1}^{12} x_i) \log_e(\theta) - 17\theta + (\sum_{i=1}^{17} y_i) \log_e(\theta)$$

$$\ell(\theta) = \log_e(K) - 29\theta + (\sum_{i=1}^{12} x_i + \sum_{i=1}^{17} y_i) \log_e(\theta)$$

The MLE is

$\hat{\theta}_{MLE} = \frac{\sum x_i + \sum y_i}{12+17} = 1033.21$

and the **maximised log-likelihood**, under H_0, ignoring the constant $\log(K)$, is

$$\ell(\hat{\theta}_{H_0}) = -12\hat{\theta} + \sum x_i \log_e(\hat{\theta}) - 17\hat{\theta} + \sum y_i \log_e \hat{\theta}$$

$\ell(\hat{\theta}_{H_0}) = -12 \times 1033.21 + 12039 \log_e(1033.21) - 17 \times 1033.21 + 17924 \log_e(1033.21) = 177992.9$

The **rejection region** of the test is

$RR = \{\mathbf{x} : 2(\ell(\theta_{H_1}) - \ell(\theta_{H_0})) > \chi^2_r(1-\alpha)\}$

Since there is one restriction i.e. $\theta_1 = \theta_2$, working at a significance level of $\alpha = 0.05$, we have that the **observed value of the test statistic** is

$2(178001.8 - 177992.9) = 17.8$ and,

$\chi^2_1(0.95) = 3.84.$

Conclusion:

The value of 17.8 for the GLRT statistic lies in the rejection region and so we reject the null hypothesis. We therefore conclude that, the labs appear not to be operating under the same conditions and the mean number of bacteria present in each lab are highly likely to be different.

An alternative way of analysing the data is to construct a confidence interval for the difference between the population means, or for the ratio. This has the advantage of giving an indication of the size of any effect. This would be done using the results from week 8 for a confidence interval for a linear function of population parameters.

Answer 3

Gene Types answer:

Blood Type	M	MN	N	Total
Observed Frequency	342	500	187	1029
Expected Frequency	343	343	343	

The likelihood ratio test statistic is:

$$2\sum_{i=1}^{3} O_i \log_e\left(\frac{O_i}{E_i}\right) = 2\left(342\log_e\left(\frac{342}{343}\right) + 500\log_e\left(\frac{500}{343}\right) + 187\log_e\left(\frac{187}{343}\right)\right) = 2(-0.999 + 188.44 - 113.4) = 148.1.$$

There are 2 degrees of freedom since there are 3 categories and 0 parameters have been estimated from the data**. Comparing this to a $\chi^2_2(0.95) = 5.99$ the null hypothesis is rejected, and we can conclude that it is highly likely that the population proportions with each blood type are not all equal to 1/3.

**We can also think of the degrees of freedom as being formed as follows.

H0: $\theta_1 = \theta_2 = \theta_3 = 1/3$

In a standard GLRT this would be 3 restrictions. However, because of the constraint on the parameters for the multinomial distribution we reduce this by 1.

Answer 4

The sample information is $k(\mathbf{x}) = -\ell''(\hat{\theta})$.

An approximate 95% confidence interval for θ is then given by

$$\hat{\theta}_{MLE} \pm 1.96\sqrt{1/k(\mathbf{x})}$$

$$\hat{\theta} \pm 1.96\sqrt{\hat{\theta}^2/(\sum_{i=1}^{n} x_i)}$$

$$18.27 \pm 1.96\sqrt{18.27^2/3654}$$

$$18.27 \pm 0.592$$

$$(17.68, 18.86)$$

It is highly likely that the population mean count is between 17.7 and 18.9.

$\mathsf{H}_0 : \theta = 20$

$\mathsf{H}_1 : \theta \neq 20$

$$\ell(\hat{\theta}_{H_1}) = \ell(\hat{\theta}_{MLE}) = (\sum_{i=1}^{n} x_i)\log_e \hat{\theta} - n\hat{\theta} + K = 3654\log_e(18.27) - 200 \times 18.27 + K = 6961.821 + K$$

$$\ell(\hat{\theta}_{H_0}) = \ell(20) = (\sum_{i=1}^{n} x_i)\log_e 20 - n \times 20 + K = 3654\log_e(20) - 200 \times 20 + K = 6946.406 + K$$

The GLRT statistic is then

$$2(l(\hat{\theta}_{H_1}) - l(20)) = 2(6961.821 - 6946.406) = 30.83$$

This is referred to a χ^2 distribution with 1 degree of freedom (since there is 1 restriction on $\mathsf{H}_0 : \theta = 20$), whose upper 5% point is 3.84. Since 30.83 $>$ 3.84, there is therefore highly significant evidence that the mean count is not 20.

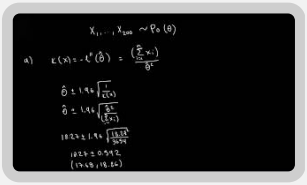
The p-value was calculated to be < 0.001, and so the hypothesis of the population mean count being equal to 20 is rejected, and we can conclude that there is evidence, again, that the population mean count is not equal to 20. This agrees with our result from the GLRT, as expected.

The advantage of the confidence interval is that it gives an indication of where the true mean lies, not just that it is significantly different from 20. The confidence interval highlights that the true mean is highly likely to lie between 17.7 and 18.9 counts in one second.

Video

Video model answers

Duration 7:10



Answer 5

We have data from three independent Binomial experiments. The sample proportions are $21/99$, $31/100$ and $42/99$. at dose levels 10, 15 and 25 respectively. The hypotheses are:

$H_0 : \theta_1 = \theta_2 = \theta_3$

H_1 : the θ_i are not all equal.

The likelihood function is

$$L(\boldsymbol{\theta}; \mathbf{x}) = K\theta_1^{21}(1-\theta_1)^{78}\theta_2^{31}(1-\theta_2)^{69}\theta_3^{42}(1-\theta_3)^{57}$$

which is maximised at $\hat{\theta}_1 = 21/99$, $\hat{\theta}_2 = 31/100$ and $\hat{\theta}_3 = 42/99$.

The maximised log-likelihood (under H_1) is

$$\ell(\hat{\theta}_{H_1}) = \ell(\hat{\theta}_{MLE}) = \log_e(K) + 21\log_e(\hat{\theta}_1) + 78\log_e(1-\hat{\theta}_1) + 31\log_e(\hat{\theta}_2) + 69\log_e(1-\hat{\theta}_2) + 42\log_e(\hat{\theta}_3) + 57\log_e(1-\hat{\theta}_3) = \log(K) - 180.55$$

Under H_0, the likelihood function is

$$L(\theta; \mathbf{x}) = K\theta^{94}(1-\theta)^{204}$$

which is maximised at $\hat{\theta} = 94/298$. The value of the maximised log-likelihood is

$$\ell(\hat{\theta}_{H_0}) = \log_e(K) + 94\log_e(\hat{\theta}) + 204\log_e(1-\hat{\theta}) = \log_e(K) - 185.8$$

The observed value of the GLRT is then

$$2(\ell(\hat{\theta_{H_1}}) - \ell(\hat{\theta}_{H_0})) = 2(-180.55 + 185.8) = 10.5$$

The critical value for this test statistic is $\chi^2(2; 0.95) = 5.99$, since there are 2 restrictions on the null hypothesis.

So, we reject the null hypothesis and conclude that, in the wider population of similar human cells, it is highly likely that the proportions of abnormal cells likely to be found at the three different dose levels are not all the same.

The *p*-value was calculated to be 0.005, and so we reject the null hypothesis ($\theta_1 = \theta_2 = \theta_3$), and come to the same conclusion as we did with our GLRT, i.e. that the θ_i's are not all equal.

It appears (informally) that the proportions of abnormal cells increases with dose level.

Answer 6

Data: $x_1, \ldots, x_k$, where $\left(\sum_{i=1}^k x_i = n\right)$

Model: $\mathbf{X} = \{X_1, \ldots, X_k\} \sim \text{Mu}(n; \theta) \quad \left(\sum_{i=1}^k \theta_i = 1\right)$

Under the general multinomial model, the likelihood function is

$$L(\theta) = K\prod_{i=1}^k \theta_i^{x_i}$$

The log-likelihood function is

$$\ell(\theta) = \log(K) + \sum_{i=1}^k x_i \log(\theta_i)$$

Under the general hypothesis H_1, and the constraint that $\sum_{i=1}^n \theta_i = 1$, it can be shown that this is maximised at

$$\hat{\theta}_i = x_i/n$$

The test statistic is:

$$2\sum_{i=1}^{k} O_i \log_e\left(\frac{O_i}{E_i}\right)$$

Since there are 5 different options then there is a probability of 1/5 (θ) that each option will be selected, which for a sample size of 100 gives an expected frequency of 20 in each option.

$H_0 : \theta = 1/5$

$H_1 : \theta \neq 1/5$

This gives us that:

Design	**1**	**2**	**3**	**4**	**5**
Number of times selected (observed)	18	14	24	29	15
Number of times selected (expected)	20	20	20	20	20

The observed value of the test statistic is:

$$2(18\log_e(18/20) + 14\log_e(14/20) + 24\log_e(24/20) + 29\log_e(29/20) + 15\log_e(15/20)$$

$$= 2 \times 3.94 = 7.88$$

RR = $\left\{2\sum_{i=1}^{k} O_i \log_e\left(\frac{O_i}{E_i}\right) > \chi^2_{(k-1-p)}(0.95)\right\}$

RR = $\left\{2\sum_{i=1}^{k} O_i \log_e\left(\frac{O_i}{E_i}\right) > \chi^2_{(5-1-0)}(0.95)\right\}$

RR = $\left\{2\sum_{i=1}^{k} O_i \log_e\left(\frac{O_i}{E_i}\right) > \chi^2_{4}(0.95)\right\}$

RR = $\left\{2\sum_{i=1}^{k} O_i \log_e\left(\frac{O_i}{E_i}\right) > 9.49\right\}$

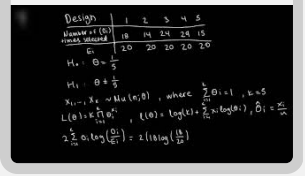
Since the observed value of the test statistic is 7.88. We do not have significant evidence that the pattern of preferences across designs is anything other than uniform.

The *p*-value was calculated to be 0.0961 (`1-pchisq(7.88, df=4)`) and so in this case we do not have sufficient evidence to reject H_0.

Video

Video model answers

Duration 6:57



Footnotes

1. Many thanks to Suzy Whoriskey for all her contributions to the development of the course material ↩