1 Likelihood Ratio Tests

The Neyman Pearson approach is optimal but in rather limited circumstances. We can devise more widely applicable methods by extending the use of the likelihood ratio to test

1. of a composite hypothesis against an alternative composite hypothesis or
2. of constructing a test of a simple hypothesis against an alternative composite hypothesis when a UMP test does not exist.

Suppose we have a sample $x_1, x_2, \cdots, x_n$ from a distribution with density $f(x, \boldsymbol{\theta})$ where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \cdots, \theta_k\}$. We are interested in some test of a hypothesis

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \text{ against } H_1 : \boldsymbol{\theta} \in \Theta_1.$$

The only restriction being that H_0 is a simplified version of H_1.

For the Neyman Person we considered

$$\lambda = \frac{\mathscr{L}(H_0)}{\mathscr{L}(H_1)}$$

and we can do the same again for composite hypotheses. Of course there may be unspecified parameters so we choose to consider

$$\lambda = \frac{\max_{\boldsymbol{\theta} \in \Theta_0} \mathscr{L}}{\max_{\boldsymbol{\theta} \in \Theta_1} \mathscr{L}}$$

That is we take the ratio

$$\lambda = \frac{\mathscr{L}(H_0)}{\mathscr{L}(H_1)}$$

where assume that we have used the maximum likelihood estimates (under each hypothesis) for the unspecified parameters.

As we have required that H_0 is a special case of H_1 it follows that $0 \le \lambda \le 1$ and we can envisage a critical region of the form $\lambda \le$ constant. As you will see there are problems!

Example 1.1. *$X_1, X_2, \cdots, X_n$ is a random sample from a Normal distribution, say $N(\mu, \sigma^2)$. We wish to test*

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu \neq \mu_0$$

Both hypotheses are composite since σ^2 is unknown. Therefore we will apply the likelihood ratio test. The likelihood is

$$\mathscr{L}(\mu, \sigma^2 | x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\} \tag{1}$$

To find $\max_{\boldsymbol{\theta} \in \Theta_0} \mathscr{L}$ substitute in Eq. (1) μ with μ_0 and maximize with respect to σ^2, i.e., find the MLE of σ^2 when $\mu = \mu_0$ is known. So,

$$\log \mathscr{L}(x; \mu, \sigma^2) = -\log(2\pi\sigma^2)^{n/2} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu_0)^2$$

or

$$\frac{\partial \log \mathscr{L}(x;\mu_0,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2\sigma^4} = 0.$$

Therefore,

$$\hat{\sigma}^2 = \frac{1}{n}\sum (x_i-\mu_0)^2 \textit{ and } \max_{\boldsymbol{\theta}\in\Theta_0} \mathscr{L} = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^n \exp\left(-\frac{n}{2}\right)$$

To find $\max_{\boldsymbol{\theta}\in\Theta_1} \mathscr{L}$ *we are looking for the MLEs of* μ *and* σ^2*. It is known that the later are*

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum (x_i-\bar{x})^2 := s'^2.$$

Therefore

$$\max_{\boldsymbol{\theta}\in\Theta_1} \mathscr{L} = \left(\frac{1}{\sqrt{2\pi s'^2}}\right)^n \exp\left(-\frac{n}{2}\right)$$

Substituting into the likelihood ratio gives

$$\lambda = \left(\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{\sum_{i=1}^n (x_i-\mu_0)^2}\right)^{n/2}.$$

We note that (skipping some algebra)

$$\sum (x_i-\mu_0)^2 = \sum (x_i-\bar{x})^2 + n(\bar{x}-\mu_0)^2$$

which eventually gives

$$\lambda = \left(1+\frac{t^2}{n-1}\right)^{-n/2}$$

where

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \textit{ and } s^2 = \frac{1}{n-1}\sum (x_i-\bar{x})^2.$$

is the usual t *statistic. The critical region is*

$$\begin{aligned} C &= \left\{x : \lambda \le k\right\} \\ &= \left\{x : \left(1+\frac{t^2}{n-1}\right)^{-n/2} \le k\right\} \\ &= \left\{x : \left(1+\frac{t^2}{n-1}\right) \ge k^{-2/n}\right\} \\ &= \left\{x : t^2 \ge k_1\right\} \\ &= \left\{x : |t| \ge \sqrt{k_1} = k_2\right\} \end{aligned}$$

For the computation of k_2 we have that

$$\alpha = P(|t| \geq k_2 | \mu = \mu_0). \tag{2}$$

We know that $t \sim t_{n-1}$ under the null hypothesis H_0 and for the Eq. (2): $k_2 = t_{n-1,1-\alpha/2}$. Therefore the test with statistic t and critical region $\{x : |t| \geq t_{n-1,1-\alpha/2}\}$ is a likelihood ratio test of

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu \neq \mu_0.$$

2 Asymptotic likelihood ratio test

It should be apparent that finding the distribution of λ is complex and probably impossible to find in general. Our life is made very much easier by Wilks who proved that

$$\Lambda = -2\ln\lambda \tag{3}$$

has a χ^2_{r-s} distribution, where r is given the number of parameters estimated in H_1 and s the number of parameters estimated in H_0. This is a large sample approximation but enables us to produce tests in a wide variety of situations.

Example 2.1. *Suppose we have $X_1, X_2, \cdots, X_n$ a random sample from a Normal distribution, say $N(\mu_x, \sigma_x^2)$. We also have a second, independent, random sample $Y_1, Y_2, \cdots, Y_m$ from $N(\mu_y, \sigma_y^2)$. We wish to test*

$$H_0 : \sigma_x = \sigma_y \text{ against } H_1 : \sigma_x \neq \sigma_y$$

Both hypotheses are composite since $\sigma_x, \sigma_y, \mu_x, \mu_y$ is unknown. Therefore we will apply the likelihood ratio test. The likelihood since the two random samples are independent is,

$$\begin{aligned}
\mathscr{L}(\sigma_x, \sigma_y, \mu_x, \mu_y | x, y) &= \mathscr{L}_x(\sigma_x, \mu_x) \times \mathscr{L}_y(\sigma_y, \mu_y) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma_x^2}}\right)^n \exp\left\{-\frac{1}{2\sigma_x^2}\sum_{i=1}^{n}(x_i - \mu_x)^2\right\} \times \\
&\quad \left(\frac{1}{\sqrt{2\pi\sigma_y^2}}\right)^m \exp\left\{-\frac{1}{2\sigma_y^2}\sum_{i=1}^{m}(y_i - \mu_y)^2\right\}
\end{aligned}$$

The log-likelihood is,

$$\begin{aligned}
\log\mathscr{L}(\sigma_x, \sigma_y, \mu_x, \mu_y | x, y) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma_x^2 - \frac{1}{2\sigma_x^2}\sum_{i=1}^{n}(x_i - \mu_x)^2 \\
&\quad -\frac{m}{2}\log 2\pi - \frac{m}{2}\log\sigma_y^2 - \frac{1}{2\sigma_y^2}\sum_{i=1}^{m}(y_i - \mu_y)^2
\end{aligned}$$

To find $\max_{\boldsymbol{\theta}\in\Theta_0}\mathscr{L}$ *substitute in* $\log\mathscr{L}(\sigma_x,\sigma_y,\mu_x,\mu_y|x,y)$ $\sigma_x=\sigma_y=\sigma$ *and maximize with respect to* σ^2*, i.e., find the MLE of* σ^2*, with respect to* μ_x*, i.e., find the MLE of* μ_x *and with respect to* μ_y*, i.e., find the MLE of* μ_y*. With other words the MLEs will derived by solving the following system of equations,*

$$\begin{aligned}
\frac{\partial}{\partial\mu_x}\log\mathscr{L}(\sigma,\mu_x,\mu_y|x,y) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu_x)=0\\
\frac{\partial}{\partial\mu_y}\log\mathscr{L}(\sigma,\mu_x,\mu_y|x,y) &= \frac{1}{\sigma^2}\sum_{i=1}^{m}(y_i-\mu_y)=0\\
\frac{\partial}{\partial\sigma^2}\log\mathscr{L}(\sigma,\mu_x,\mu_y|x,y) &= -\frac{n}{\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu_x)^2\\
&\quad -\frac{m}{\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{m}(y_i-\mu_y)^2=0
\end{aligned}$$

After some algebra,

$$\hat{\mu}_x=\bar{x},\hat{\mu}_y=\bar{y},\hat{\sigma}^2=\frac{1}{n}\sum(x_i-\bar{x})^2+\frac{1}{m}\sum(y_i-\bar{y})^2$$

To find $\max_{\boldsymbol{\theta}\in\Theta_1}\mathscr{L}$ *we are looking for the MLEs of* $\mu_x,\sigma_x^2,\mu_y,\sigma_y^2$*. These are*

$$\hat{\mu}_x=\bar{x},\quad \hat{\sigma_x}^2=\frac{1}{n}\sum(x_i-\bar{x})^2:=s_x'^2.$$

$$\hat{\mu}_y=\bar{y},\quad \hat{\sigma_y}^2=\frac{1}{m}\sum(y_i-\bar{y})^2:=s_y'^2$$

Substituting into the likelihood ratio gives, after, some algebra reduces this to

$$\lambda=\left\{\frac{\hat{\sigma}^{-(n+m)}}{\hat{\sigma_x}^{-n}\hat{\sigma_y}^{-m}}\right\}$$

Taking logs

$$\Lambda=2n\log(\hat{\sigma_x})+2m\log(\hat{\sigma_y})-2(n+m)\log(\hat{\sigma}).$$

There are 4 parameters, all of these had to be estimated for H_1 *and 3 for* H_0*. It follows that* Λ *is* X_1^2*.*

3 Goodness-of-fit-tests

A common likelihood-ratio based test is the goodness-of-fit test.

3.1 Categories

Suppose we have an experiment which has k mutually exclusive outcomes which we will label $A_1, A_2, \cdots, A_k$. Suppose further we repeat our experiment n times and find that the number of outcomes in A_j is $n_j, \quad j = 1, 2, \cdots, k$. In addition we will write the probability of an outcome being in A_j is $p_j, \quad j = 1, 2, \cdots, k$. Clearly $\sum_{j=1}^{k} n_j = n$ and $\sum_{j=1}^{k} p_j = 1$.

This simple model has many applications, you could think of asking questions in a survey with the A_j as categories of answers, or the A_j could correspond to the bins of a histogram. To proceed any further we need a bit more theory.

3.2 The multinomial distribution

If we have the situation above of k mutually and exhaustive categories $A_j \quad j = 1, 2, \cdots, k$ and we have

n_1 observations in A_1 and P[fall in A_1] = p_1
n_2 observations in A_2 and P[fall in A_2] = p_2
...
n_j observations in A_j and P[fall in A_j] = p_j
...
n_k observations in A_k and P[fall in A_k] = p_k

Then

$$P[n_1 \text{ in } A_1, \cdots, n_k \text{ in } A_k] = \frac{n!}{n_1! n_2! \cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k} \tag{4}$$

It is reasonably clear that this is an extension of the Binomial and the distribution is known as the multinomial distribution.

3.3 Maximum likelihood estimators of multinomial

We can find the maximum likelihood estimators by maximizing the log likelihood

$$\ell(\boldsymbol{p}) = \sum_{j=1}^{k} n_j \log(p_j) + constant$$

subject to $\sum_{j=1}^{k} p_j = 1$. The constant is of course $\log(n!) - \sum_{j=1}^{k} \log(n_j!)$

Then consider the function

$$\Phi(\boldsymbol{p}, \lambda) = \sum_{j=1}^{k} n_j \log(p_j) + \lambda(\sum_{j=1}^{k} p_j - 1) + constant.$$

The MLEs will derived by solving the following system of equations,

$$\frac{\partial \Phi}{\partial p_1} = \frac{n_1}{p_1} + \lambda = 0$$

$$\frac{\partial \Phi}{\partial p_2} = \frac{n_2}{p_2} + \lambda = 0$$
$$\vdots$$
$$\frac{\partial \Phi}{\partial p_k} = \frac{n_k}{p_k} + \lambda = 0$$
$$\frac{\partial \Phi}{\partial p \lambda} = \sum p_j - 1 = 0$$

This is equivalent with

$$\frac{n_1}{p_1} = \frac{n_2}{p_2} = \ldots = \frac{n_k}{p_k} = \frac{\sum n_j}{\sum p_j} = \frac{n}{1}.$$

Hence it is easy to show that

$$\hat{p}_j = \frac{n_j}{n}.$$

Of course you could say that on either falls in A_j or not - a Binomial problem. The maximum likelihood estimate of p_j is just that for the Binomial probability $\hat{p}_j = \frac{n_j}{n}$.

3.4 Likelihood ratio for the multinomial

Suppose we now wish to test

$$H_0 : P(A_j) = p_j, \quad j = 1, 2, \cdots, k \text{ against } H_1 : \text{ probabilities are unspecified}$$

The likelihood ratio is in general

$$\lambda = \prod_{j=1}^{k} \hat{p}_j^{n_j} / \prod_{j=1}^{k} \breve{p}_j^{n_j} = \prod_{j=1}^{k} \left(\hat{p}_j / \breve{p}_j\right)^{n_j}$$

where the $\hat{p}_j$ are the estimates under H_0 and $\breve{p}_j$ are the estimates under H_1. Since we have unspecified probabilities, we need the estimates of these probabilities - however we know that

$$\breve{p}_j = n_j / n \qquad j = 1, 2, \cdots, n$$

so

$$\lambda = \prod_{j=1}^{k} \hat{p}_j^{n_j} / \prod_{j=1}^{k} \left(\frac{n_j}{n}\right)^{n_j} = \prod_{j=1}^{k} \left(n \hat{p}_j / n_j\right)^{n_j}$$

or

$$\Lambda = -2 \log \lambda = 2 \sum_{j=1}^{k} n_j \log\left(\frac{n_j}{n \hat{p}_j}\right).$$

Our statistic is often written as the asymptotically equivalent form

$$X^2 = \sum_{j=1}^{k} \frac{(n_j - \hat{e}_j)^2}{\hat{e}_j}, \text{ where } \hat{e}_j = n \hat{p}_j. \tag{5}$$

Note here that if the probabilities p_j under the null hypothesis are known and not need to be estimated then the chi-square statistic reduces to the form,

$$X^2 = \sum_{j=1}^{k} \frac{(n_j - e_j)^2}{e_j}, \text{ where } e_j = np_j. \tag{6}$$

Example 3.1. *An experimenter bred flowers and found the following numbers in each of the four possible classes*

Class	A_1	A_2	A_3	A_4
Number	*120*	*48*	*36*	*13*
Probability	*9/16*	*3/16*	*3/16*	*1/16*

The table also includes the probability of falling in each class - according to established theory. We aim to test

H_0 : *probabilities are as given by the table* H_1 : *probabilities are unspecified.*

In the following table we have calculated the expected frequencies:

class	p_j	n_j	$e_i = np_j$
1	9/16	120	122.06
2	3/16	48	40.69
3	3/16	36	40.69
4	1/16	13	13.56

The likelihood ratio, since the probabilities under the null hypothesis are given, is

$$\lambda = \prod_{j=1}^{k} \left(np_j / n_j\right)^{n_j}$$

and we find, after some arithmetic, using Eq. (6) that Λ *is approximately 1.9. We know that* $\Lambda = -2\log(\lambda)$ *is chi-squared. Here we have unspecified k-1 parameters under* H_1 *and none under* H_0 *so the number of degrees of freedom is 3. We will reject* H_0 *if* Λ *is large i.e. exceed the 95% point of* X_3^2 *which is 7.815 . In this case we accept* H_0.

3.5 Goodness of fit–Non multinomial distribution

We can use the ideas above for goodness-of-fit testing for various theoretical distributions. We need to split the x axis in k intervals (classes) $A_1, A_2, \dots, A_k$ and calculate $P(A_1), P(A_2), \dots, P(A_k)$ using the theoretical distribution.

Example 3.2. *Suppose we have 200 random numbers which are suppose to be from a* $U(0,1)$ *distribution. The numbers in the intervals (0-0.1), (0.1-0.2), e.t.c., are*

class	0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6
frequency	19	18	20	16	26	18
class		0.6-0.7	0.7-0.8	0.8-0.9	0.9-1	
frequency		19	19	23	22	

This is just a test of

$$H_0 : p_j = 0.1 \textit{ for all } j \textit{ against An unspecific alternative.}$$

We find that , using $\Lambda = \sum_{j=1}^{k} \frac{(n_j - e_j)^2}{e_j} = 3.69$, *where* $e_j = np_j$ *(the probabilities under the null hypotheses are given and not need to be estimated); see also Eq. 6. We know that* Λ *is* $X^2_{r-s} = X^2_{9-0}$ *and we accept* H_0 *at 5% since* Λ *does not lie in the critical region* $\Lambda \geq \chi^2_{9,0.95} = 16.919$ *and conclude that the distribution is indeed Uniform.*

Example 3.3. *A survey of families with 5 children gave rise the following distribution.*

No Boys	0	1	2	3	4	5	total
no families	8	40	88	110	56	18	320

One model for the number of boys is the binomial, specifically that the number of boys X *has the distribution*

$$P[X = x] = \binom{5}{x} p^x (1-p)^{5-x} \qquad x = 0, 1, 2, 3, 4, 5,$$

where $p = P[\textit{ boy }]$.

We can see if this distribution fits the data when $p = \frac{1}{2}$. *Calculating the expected probabilities,*

$$P[0 \textit{ births in } 5] = \binom{5}{0} \left(\frac{1}{2}\right)^5 = 1/32$$

$$P[1 \textit{ birth in } 5] = \binom{5}{1} \left(\frac{1}{2}\right)^5 = 5/32$$

$$\cdots,$$

we conclude to the following table,

No Boys	0	1	2	3	4	5	total
no families (n_j)	8	40	88	110	56	18	320
expected ($e_j = np_j$)	10	50	100	100	50	10	

The likelihood ratio statistic in Eq. (6) is $\Lambda = 11.096$ *and the degrees of freedom is 6-1=0. We only need to estimate 5 of the probabilities since the remaining one follows from the fact that they must add up to one. The upper point of* $\chi^2_5 = 11.07$ *so in this case we reject* H_0 *and conclude our model is wrong.*

We know however that in general $p[\textit{ boy }] > \frac{1}{2}$. *Indeed from our data the proportion of boys is* $\hat{p} = 0.5375$. *We can revisit out Binomial model but in this case we use* $\hat{p} = 0.5375$. *The expectations are harder - I get*

No Boys	*0*	*1*	*2*	*3*	*4*	*5*	*total*
no families (n_j)	*8*	*40*	*88*	*110*	*56*	*18*	*320*
expected ($\hat{e}_j = n\hat{p}_j$)	*6.8*	*39.3*	*91.5*	*106.3*	*61.8*	*14.4*	*320*

and $\Lambda = 1.03$. *We now have (6-1)-1 =4 degrees of freedom. We estimate 5 parameters for* H_1 *and one under* H_0. *The conclusion is that this fits the data very well.*

Example 3.4. *According to the data of the following table can we assume that the parent distribution is standard normal?*

classes	frequencies
≤ 39.5	6
39.5–44.5	13
44.5–49.5	40
49.5–54.5	65
54.5–59.5	52
≥ 59.5	24

For testing the null hypothesis that the data follow normal distribution against the alternative that follow any other distribution we need to:

1. *Derive the maximum likelihood estimators of* μ, σ^2, *i.e.,*

$$\mu = \bar{x} = \frac{1}{n}\sum n_i x_i, \quad \bar{\sigma}^2 = s'^2 = \frac{1}{n}\left(\sum n_i x_i^2 - \frac{(\sum n_i x_i)^2}{n}\right),$$

where x_i *is the center of each interval and* n_i *the observed frequency. The centers of the classes are 37, 42, 47, 52, 57, 62 (for eg., consider the class 44.5–49.5 we have [44.5+49.5]/2=47). After some calculations,*

$$\bar{\mu} = 52.4 \textit{ and } \bar{\sigma}^2 = 36.77 = 6.06^2.$$

2. *Calculate the estimated probabilities* p_j *from* $N(\mu, \sigma^2) = N(52.4, 36.77)$. *For e.g.,*

$$P(x \leq 39.5) = P\left(\frac{x - 52.4}{6.06} \leq \frac{39.5 - 52.4}{6.06}\right) = P(z \leq -2.13) = 0.5 - P(0 < z < 2.13) = 0.0166.$$

3. *Calculate the expected (estimated) frequencies* $\hat{e}_j = n\hat{p}_j$. *For e.g.,* $\hat{e}_1 = 200 \times 0.0166$.

So we derive the following table,

classes	n_j	$\hat{p}_j$	$\hat{e}_j = n\hat{p}_j$
≤ 39.5	6	0.0166	3.32
39.5–44.5	13	0.0802	16.04
44.5–49.5	40	0.2188	43.76
49.5–54.5	65	0.3212	64.24
54.5–59.5	52	0.2422	48.44
≥ 59.5	24	0.1210	24.20
total	200	1	200

One useful rule of thumb is to try and ensure that non of the expected values are less than 5%. This is a rather mysterious number but if we examine our chi-squared approximation carefully we see it breaks down when we have small expected values. Therefore we merge the classes ≤ 39 *and 39.5–44.5, i.e., we have finally 5 classes.*

According to the table we find that, using Eq. (5) $\Lambda = \sum_{j=1}^{5} \frac{(n_j - \hat{e}_j)^2}{\hat{e}_j} = 0.6021$. *We know that* Λ *is* $X^2_{r-s} = X^2_{4-2}$ *and we accept* H_0 *at 1% since* Λ *does not lie in the critical region* $\Lambda \geq \chi^2_{2,0.99} = 9.210$ *and conclude that the distribution is indeed normal.*