Topic 19: Extensions on the Likelihood Ratio*

April 19, 2010

We begin with a composite hypothesis test

\[ H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1 \]

with \( \Theta_0 \cap \Theta_1 = \emptyset \) and \( \Theta_0 \cup \Theta_1 = \Theta \), the parameter space. Let \( C \) be the critical region for an \( \alpha \) level test, i.e., we reject the null hypothesis whenever the data \( x \) fall in the critical region. Thus, the power function

\[ \pi(\theta) = P_{\theta}\{X \in C\} \]

has the property that

\[ \pi(\theta) \leq \alpha \quad \text{for all} \quad \theta \in \Theta_0 \]

and that \( \alpha \) is the maximum value of the power function on \( \Theta_0 \), the parameter values associated to the null hypothesis.

### 1 One-Sided Tests

Let's collect a simple random sample of independent normal observations with unknown mean and known variance \( \sigma_0^2 \). We noticed, in the case of a simple hypothesis test

\[ H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu = \mu_1 \]

that the critical region as determined by the Neyman-Pearson lemma depended only on whether or not \( \mu_1 \) was greater than \( \mu_0 \) or not. For example, if \( \mu_1 > \mu_0 \), then the critical region \( C = \{x; \bar{x} \geq \bar{x}_\alpha\} \) shows that we reject \( H_0 \) whenever the sample mean is higher than some threshold value \( \bar{x}_\alpha \), irrespective of the difference between \( \mu_0 \) and \( \mu_1 \).

We will examine the idea that if a test is most powerful against each possible alternative in a simple hypothesis test, when we can say that this test is in some sense best overall for a composite hypothesis. Stated in terms of the power function, we are asking if a test has the property that its power function \( \pi \) is greater for any value of \( \theta \) than the power function of any other test for any value of \( \theta \). Such a test is called uniformly most powerful. In general, a hypothesis will not have a uniformly most powerful test. However, we can hope for such a test for procedures involving simple hypotheses in which the test statistic that emerged from the likelihood test did not depend on the specific value of the alternative. This was seen in the example above using independent normal data. In this case, the power function

\[ \pi(\mu) = P_{\mu}\{\bar{X} \geq k_\alpha\} \]

increases as \( \mu \) increases and so the test has the intuitive property of becoming more powerful with increasing \( \mu \).

In general, we look for a test statistic \( T(x) \) (like \( \bar{x} \) in the example above). Next, we check that the likelihood ratio,

\[ \frac{L(\theta_2|x)}{L(\theta_1|x)} \quad \theta_1 < \theta_2. \]

depends on the data \( x \) only through the value of statistic \( T(x) \) and, in addition, this ratio is a monotone function of \( T(x) \). If these conditions hold, then for an appropriate value of \( k_\alpha \), \( C = \{x; T(x) \geq k_\alpha\} \) is the critical region for a uniformly most powerful \( \alpha \) level test for the one-sided alternative hypothesis

\[ H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0. \]

*© 2011 Joseph C. Watkins

237
A corresponding criterion holds for the one sided test with the inequalities reversed:

$$H_0 : \theta \geq \theta_0 \quad \text{versus} \quad H_1 : \theta < \theta_0.$$  

These conditions are satisfied for the case above as well as the tests for $p$, the probability of success in Bernoulli trials.

**Example 1** (One sample one proportion $z$-test). For $X = (X_1, \ldots, X_n)$ is a sequence of Bernoulli trials with unknown success probability $p$, we can have and the **one-sided tests** with the alternative is greater

$$H_0 : p \leq p_0 \quad \text{versus} \quad H_1 : p > p_0$$

or less

$$H_0 : p \geq p_0 \quad \text{versus} \quad H_1 : p < p_0.$$  

**Example 2.** We return to the example of the survivability of bee hives over a given winter. The probability of survival is $p_0 = 0.7$. The one-sided alternative for a mild winter is that this survival probability has increased. This leads us to consider the hypotheses

$$H_0 : p \leq 0.7 \quad \text{versus} \quad H_1 : p > 0.7.$$  

for a test of the probability that a feral bee hive survives a winter. The test statistic derived from the central limit theorem is

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

and for an $\alpha$ level test, the critical value is $z_\alpha$ where $\alpha$ is the probability that a standard normal is at least $z_\alpha$.

For this study, 112 colonies have been selected with 88 surviving. Thus $\hat{p} = 0.7875$ and $z = 1.979$. If the significance level is $\alpha = 0.05$, then we will reject $H_0$ because $z = 1.979 > 1.645 = z_0.05$. We also perform this test in R by entering

```r
> prop.test(88,112,0.7,alternative=c("greater"))
```

1-sample proportions test with continuity correction

data: 88 out of 112, null probability 0.7
X-squared = 3.5208, df = 1, p-value = 0.0303
alternative hypothesis: true p is greater than 0.7
95 percent confidence interval:
0.7107807 1.0000000
sample estimates:

p

0.7857143

**Exercise 3.** Is the test above significant at the 10% level? the 5% level?

A direct appeal to the central limit theorem gives a slightly different $p$-value, namely

```r
> 1-pnorm(1.979)
[1] 0.02390800
```

This is expressed in the R output as a **continuity correction**. In this case we are approximating probability for the outcome $P\{X \geq x\}$ for $X$ the number of successes in Bernoulli trials by $P\{Y \geq x + 1/2\}$ where $Y$ is a normal random variable. This correction can be seen by looking at the area associated to a histogram for the mass function for $X$ and the density function for $Y$.

Because $P\{X \geq x\} = P\{X > x - 1\} = 1 - P\{X \leq x - 1\}$, the R command for $P\{X \geq x\}$ is `1-pbinom(x-1,n,p)`. The table below compares computing the $P$-value using the binomial directly, binompvalue, the normal approximation normpvalue, and the normal approximation with the continuity correction normpvaluecc. The number for the test above are shown on line 9.
> x<-c(80:92)
> binompvalue<-1-pbinom(x-1,112,0.7)
> normpvalue<-1-pnorm(x,n*p,sqrt(n*p*(1-p)))
> normpvaluecc<-1-pnorm(x+0.5,n*p,sqrt(n*p*(1-p)))
> data.frame(x,binompvalue,normpvalue,normpvaluecc)

<table>
<thead>
<tr>
<th>x</th>
<th>binompvalue</th>
<th>normpvalue</th>
<th>normpvaluecc</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>0.4154655712</td>
<td>0.3707323063</td>
<td>0.4102833420</td>
</tr>
<tr>
<td>81</td>
<td>0.3367097745</td>
<td>0.2959409315</td>
<td>0.3325027711</td>
</tr>
<tr>
<td>82</td>
<td>0.2641118385</td>
<td>0.2289505277</td>
<td>0.2613434153</td>
</tr>
<tr>
<td>83</td>
<td>0.2000721958</td>
<td>0.1714364652</td>
<td>0.1989421672</td>
</tr>
<tr>
<td>84</td>
<td>0.1460628585</td>
<td>0.1241065395</td>
<td>0.1464910316</td>
</tr>
<tr>
<td>85</td>
<td>0.1025553369</td>
<td>0.0867731109</td>
<td>0.1042322648</td>
</tr>
<tr>
<td>86</td>
<td>0.0691426141</td>
<td>0.0585464089</td>
<td>0.0715976128</td>
</tr>
<tr>
<td>87</td>
<td>0.0446167293</td>
<td>0.0380903233</td>
<td>0.0474407060</td>
</tr>
<tr>
<td>88</td>
<td>0.0275341591</td>
<td>0.0238806213</td>
<td>0.0303009849</td>
</tr>
<tr>
<td>89</td>
<td>0.0162104857</td>
<td>0.0144193968</td>
<td>0.0186445384</td>
</tr>
<tr>
<td>90</td>
<td>0.0090854776</td>
<td>0.0083811681</td>
<td>0.0110460276</td>
</tr>
<tr>
<td>91</td>
<td>0.0048368618</td>
<td>0.0046873842</td>
<td>0.0062982479</td>
</tr>
<tr>
<td>92</td>
<td>0.0024402066</td>
<td>0.0025215113</td>
<td>0.0034547541</td>
</tr>
</tbody>
</table>

**Example 4.** For a test of hive survivability over a harsh winter, we have

\[ H_0 : p \geq 0.7 \quad \text{versus} \quad H_1 : p < 0.7. \]

If we have 26 observations, then we are reluctant to use the central limit theorem and appeal directly to the binomial distribution. If 16 hives survive, then we use the binomial test as follows.

> binom.test(16,26,0.7,alternative=c("less"))

Exact binomial test

data: 16 and 26
number of successes = 16, number of trials = 26, p-value = 0.2295
alternative hypothesis: true probability of success is less than 0.7
95 percent confidence interval:
0.0000000 0.7743001
sample estimates:
probability of success
0.6153846

and we do not reject for any significance level \( \alpha \) below 0.2295.

## 2 Likelihood Ratio Tests

The likelihood ratio test is a popular choice for composite hypothesis. \( \Theta_0 \) is a subspace of the whole parameter space. The basic idea is that the null hypothesis is unlikely to be true if the maximum likelihood on the null hypothesis is sufficiently smaller that the likelihood maximized over \( \Theta \), the entire parameter space. In symbols, let \( \hat{\theta}_0 \) be the parameter value that maximizes the likelihood for \( \theta \in \Theta_0 \) and \( \hat{\theta} \) be the parameter value that maximizes the likelihood for \( \theta \in \Theta \). Then the **likelihood ratio**

\[
\Lambda(x) = \frac{L(\hat{\theta}_0|x)}{L(\hat{\theta}|x)}.
\]

239
Note that this ratio is the reciprocal from the version given by the Neyman-Pearson lemma. Thus, the critical region consists of those values that are below a critical value.

The rejection region for an \( \alpha \)-level likelihood ratio test is

\[
\{ \Lambda(x) \leq \lambda_\alpha \}
\]

As with any \( \alpha \) level test, \( \lambda_\alpha \) is chosen so that

\[
P_\theta(\Lambda(X) \leq \lambda_\alpha) \leq \alpha \text{ for all } \theta \in \Theta.
\]

**Example 5.** Let \( \Theta = \mathbb{R} \) and consider the two-sided hypothesis

\[
H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0.
\]

Here the data are \( n \) independent \( N(\mu, \sigma^2) \) random variables \( X_1, \ldots, X_n \) with known variance \( \sigma^2 \). Then, \( \hat{\mu}_0 = \mu_0 \) and \( \hat{\mu} = \bar{x} \). Consequently,

\[
L(\hat{\mu}_0|x) = \left( \frac{1}{2\pi \sigma^2} \right)^n \exp -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu_0)^2,
\]

\[
L(\hat{\mu}|x) = \left( \frac{1}{2\pi \sigma^2} \right)^n \exp -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

and

\[
\Lambda(x) = \exp -\frac{1}{2\sigma^2} \left( \sum_{i=1}^{n} (x_i - \mu_0)^2 - (x_i - \bar{x})^2 \right) = \exp -\frac{n}{2\sigma^2} (\bar{x} - \mu_0)^2.
\]

Now notice that

\[
-2 \ln \Lambda(x) = \frac{n}{\sigma^2} (\bar{x} - \mu_0)^2 = \left( \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right)^2.
\]

Then, critical region

\[
\{ \Lambda(x) \leq \lambda_\alpha \} = \left\{ \left( \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right)^2 \geq -2 \ln \lambda_\alpha \right\}
\]

Because \( (\bar{X} - \mu_0) / (\sigma / \sqrt{n}) \) is a standard normal random variable, \( -2 \ln \Lambda(X) \) is the square of a single standard normal, hence, a \( \chi^2 \)-square random variable with 1 degree of freedom.

Naturally we can use both

\[
\left( \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right)^2 \quad \text{and} \quad \left| \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right|,
\]

as a test statistic. For the first, the critical value is just the square of the critical value for the second choice.

**Example 6 (Bernoulli trials).** Consider the two-sided hypothesis

\[
H_0 : p = p_0 \quad \text{versus} \quad H_1 : p \neq p_0.
\]

The likelihood is

\[
L(p|x) = (1 - p)^{n-\sum x_i} p^{\sum x_i}.
\]

Then, \( \hat{p}_0 = p_0 \) and \( \hat{p} \) is the sample proportion. Thus,

\[
\Lambda(x) = \frac{(1 - p_0)^{n(1 - \hat{p})} p_0^{n\hat{p}}}{(1 - \hat{p})^{n(1 - p_0)} \hat{p}^{n\hat{p}}}
\]

Let’s repeat the strategy that we used for normal data in the previous example:

\[
- \ln \Lambda(x) = n \left( (1 - \hat{p})(\ln(1 - \hat{p}) - \ln(1 - p_0)) + \hat{p}(\ln \hat{p} - \ln p_0) \right).
\]
Next, let’s replace the logarithms with their linear approximation:
\[
\ln(1 - \hat{p}) - \ln(1 - p_0) \approx -\frac{\hat{p} - p_0}{1 - p_0} \quad \ln \hat{p} - \ln p_0 \approx \frac{\hat{p} - p_0}{p_0}.
\]

Then,
\[
-\ln \Lambda(x) = n\left(\left(1 - \hat{p}\right)\left(-\frac{\hat{p} - p_0}{1 - p_0}\right) + \hat{p}\left(\frac{\hat{p} - p_0}{p_0}\right)\right) = n(\hat{p} - p_0) - \frac{1 - \hat{p}}{1 - p_0} + \frac{\hat{p}}{p_0}.
\]

Again, this is approximately the square of a standard normal. Thus, we can either use
\[
\frac{(\hat{p} - p_0)^2}{p_0(1 - p_0)/n} \quad \text{or} \quad \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}
\]
for the test statistic. Under the null hypothesis, these statistics have approximately a standard normal distribution.

For a two-sided test and a 98% confidence interval, we enter
\[
> \text{prop.test}(88,112,0.7,\text{alternative}=\text{c("two.sided")},\text{conf.level} = 0.98)
\]

1-sample proportions test with continuity correction

data: 88 out of 112, null probability 0.7
X-squared = 3.5208, df = 1, p-value = 0.0606
alternative hypothesis: true p is not equal to 0.7
98 percent confidence interval:
0.6785906 0.8652397
sample estimates:
p
0.7857143

to obtain the interval (0.676, 0.8652).

We can also use the central limit theorem and R to compute the power function, \(\pi\). For a one-sided test with a greater than alternative, the critical region is
\[
C = \left\{ x; \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \geq z_\alpha \right\}.
\]

To find \(\pi(p)\), we need to rewrite this expression so that we can create an expression
\[
z = \frac{\hat{p} - p}{\sqrt{p(1 - p)/n}}
\]
that is a standard normal under the parameter value \(p\). Beginning with the expression defining the critical region, we
have that

\[
\frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \geq z_\alpha
\]

\[
\hat{p} - p_0 \geq z_\alpha \sqrt{p_0(1 - p_0)/n}
\]

\[
\hat{p} - p \geq z_\alpha \sqrt{p_0(1 - p_0)/n} + p_0 - p
\]

\[
\frac{\hat{p} - p}{\sqrt{p(1 - p)/n}} \geq z_\alpha \sqrt{p_0(1 - p_0)/n} + \frac{p_0 - p}{\sqrt{p(1 - p)/n}}
\]

\[
z = \frac{\hat{p} - p}{\sqrt{p(1 - p)/n}} \geq z_\alpha \sqrt{p_0(1 - p_0)/p(1 - p)} + \frac{p_0 - p}{\sqrt{p(1 - p)/n}}
\]

To explore the properties of the power function in the case of overwintering of bee hives, we first keep the number of hives at \( n = 112 \) and considering increasing values \( p = 0.75, 0.80, 0.85, 0.90 \) for the alternative to see the power increase from about 30% to nearly 100%.

\[
> n<-112
> p0<-0.7
> zalpha<-qnorm(0.95)
> p<-c(0.75,0.80,0.85,0.90)
> power<-1-pnorm(zalpha*sqrt(p0*(1-p0)/(p*(1-p))) + (p0-p)/sqrt(p*(1-p)/n))
> data.frame(p,power)
\]

\[
\begin{array}{ll}
1 & 0.75 0.3019748 \\
2 & 0.80 0.7767714 \\
3 & 0.85 0.9902226 \\
4 & 0.90 0.9999972 \\
\end{array}
\]

Power increases with increasing sample size. Here we fix the alternative at 80% and choose \( n \) from 40 to 240. The power for these values increases from 38% to more than 97%.

\[
> n<-c(1:6)*40
> p0<-0.7
> zalpha<-qnorm(0.95)
> p<-0.8
> power<-1-pnorm(zalpha*sqrt(p0*(1-p0)/(p*(1-p))) + (p0-p)/sqrt(p*(1-p)/n))
> data.frame(n,power)
\]

\[
\begin{array}{ll}
1 & 40 0.3808391 \\
2 & 80 0.6374501 \\
3 & 120 0.8035019 \\
4 & 160 0.8993508 \\
5 & 200 0.9506427 \\
6 & 240 0.9766255 \\
\end{array}
\]

Exercise 7. For the two-sided two-sided \( \alpha \)-level proportion test

\[
H_0 : p_1 = p_2 \quad \text{versus} \quad H_1 : p_1 \neq p_2,
\]

based on \( n_1 \) Bernoulli trials, \( x_{1,1}, x_{1,2}, \ldots, x_{1,n_1} \) from the first population and \( n_2 \) Bernoulli trials, \( x_{2,1}, x_{2,2}, \ldots, x_{2,n_2} \) from the second, the likelihood ratio test is equivalent to the critical region

\[
|z| \geq z_{\alpha/2}
\]
where

\[ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0(1 - \hat{p}_0) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \]

with \( \hat{p}_i \), the sample proportion of successes from the observations from population \( i \) and \( \hat{p}_0 \), the pooled proportion

\[ \hat{p}_0 = \frac{1}{n_1 + n_2} \left( (x_{1,1} + \cdots + x_{1,n_1}) + (x_{2,1} + \cdots + x_{2,n_2}) \right) = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}. \]  

By a variant of the central limit theorem, \( z \) has, under the null hypothesis, approximately a standard normal random variable.

### 3 Chi-square Tests

This exact computation for normal data for a two-sided test of the mean shows that the test statistic has a \( \chi^2 \) distribution with 1 degree of freedom. We can extend the ideas using the delta method to show the asymptotic normality and efficiency of a maximum likelihood estimator to obtain the following very general \( \chi^2 \) test.

**Theorem 8.** Whenever the maximum likelihood estimate has an asymptotically normal distribution, let \( \Lambda(x) \) be the likelihood ratio criterion for an \( d \) dimensional parameter space:

\[ H_0 : \theta_1 = c_1 \text{ for all } i = 1, \ldots, k \quad \text{versus} \quad H_1 : \theta_1 \neq c_1 \text{ for some } i = 1, \ldots, k \]

Then under \( H_0 \),

\[ -2 \ln \Lambda_n(X) \]

converges in distribution to a \( \chi^2_{d-k+1} \) random variable.

More generally, if the test is based on a \( n \)-dimensional parameter space that has linear constraints on \( k \) of the variables. This can be changed to the test above by a linear change of variables.

**Exercise 9.** Use a second order Taylor series for \( \ln L(c|x) \) and the asymptotic normality of maximum likelihood estimators to outline the argument for the case \( d = k = 1 \).

The basic approach taken for this case extends to the general case. If we expand \( \ln L(c|x) \) in a Taylor series about the parameters \( \theta_1 = \hat{\theta}_1, \theta_2 = \hat{\theta}_2, \ldots, \theta_k = \hat{\theta}_k \) the maximum likelihood estimators, then the first order terms in the expansion of \( \ln L(c|x) \) vanish. The second order derivatives are the entries of the Fisher information matrix evaluated at the maximum likelihood estimator. These terms converge by the law of large numbers. A multidimensional central limit theorem applies to the vector of terms \( \sqrt{n}(\hat{\theta}_1 - c_1, \ldots, \hat{\theta}_k - c_k) \). The result is the Fisher information matrix and its inverse cancelling leaving the sum of the squares of \( d - k + 1 \) approximately normal random variables. This is the definition of a \( \chi^2_{d-k+1} \) distribution.

**Example 10.** During meiosis, paired chromosomes experience crossing over events in the formation of gametes. During prophase I, the four available chromatids (two from each parent) are in tightly aligned allowing breaks and reattachments of homologous sites on two chromatids. (See Figure 1.)

Recombination can occur with a small probability at any location along chromosome. As described in Topic 9 in the discussion that moved us from Bernoulli trials to a Poisson random variable, the number of crossing over events is modeled as Poisson random variable. The mean number of crosses over for a given chromosomal segment is called its genetic length with Morgans as the unit of measurement. This name is in honor of Thomas Morgan who won the Nobel Prize in Physiology or Medicine in 1933 for discoveries relating the role the chromosome plays in heredity.

We are now collecting whole genome sequences for trios - an individual along with both parents. Consequently, we can determine on both the father’s and the mother’s chromosome the number of crossing over events and address the
question: Are these processes different in the production of sperm and eggs? One simple question is: Are the number of crossing over events different in sperm and in eggs? This leads to the hypothesis

\[ H_0 : \lambda_m = \lambda_f \quad \text{versus} \quad H_1 : \lambda_m \neq \lambda_f \]

where \( \lambda_m \) and \( \lambda_f \) is the parameter in the Poisson random variable that gives the number of crossing over events in the human chromosome across all 23 autosomes. (We will not look at the X and Y chromosomes in this circumstance.)

The data are \( n_m, n_f \) the number of crossing over events for each parent’s chromosome. Thus the likelihood function is

\[
L(\lambda_m, \lambda_f | x_m, x_f) = \frac{\lambda_m^{n_m}}{n_m!} e^{-\lambda_m} \cdot \frac{\lambda_f^{n_f}}{n_f!} e^{-\lambda_f}.
\]

**Exercise 11.** Show that the maximum likelihood estimates for the likelihood function above is

\[
\hat{\lambda}_m = n_m \quad \text{and} \quad \hat{\lambda}_f = n_f.
\]

Thus,

\[
L(\hat{\lambda}_m, \hat{\lambda}_f | n_m, n_f) = \frac{n_m^{n_m}}{n_m!} \cdot \frac{n_f^{n_f}}{n_f!} e^{-(n_m+n_f)}.
\]

Under the null hypothesis, \( \lambda_m \) and \( \lambda_f \) have a common value. Let’s denote this by \( \lambda \). Then the likelihood function is

\[
L(\lambda | n_m, n_f) = \frac{\lambda^{n_m}}{n_m!} e^{-\lambda} \cdot \frac{\lambda^{n_f}}{n_f!} e^{-\lambda} = \frac{\lambda^{n_m+n_f}}{n_m! n_f!} e^{-2\lambda}.
\]

**Exercise 12.** Show that the maximum likelihood estimate for the likelihood function above is

\[
\hat{\lambda} = \frac{n_m + n_f}{2}.
\]
Thus,

\[ L(\lambda|x_m, x_f) = \frac{((n_m + n_f)/2)^{n_m + n_f}}{n_m! n_f!} e^{-(n_m + n_f)}. \]

The likelihood ratio, after cancelling the factorial and exponential factors, is

\[ \Lambda(n_m, n_f) = \frac{L(\lambda|n_m, n_f)}{L(\lambda, \lambda_f|n_m + n_f)} = \frac{(n_m + n_f)^{n_m + n_f}}{2^{n_m + n_f} n_m! n_f!}. \]

Our data for two individuals sharing the same parents are \( n_m = 56 \) and \( n_f = 107 \). Thus, \(-2 \ln \Lambda(n_m, n_f) = 16.228\). To compute the p-value

\[
> nm<56 \\
> nf<107 \\
> n<-nm+nf \\
> 1-pchisq(-2*(n*log(n)-nf*log(nf)-nm*log(nm)-n*log(2)),1) \\
[1] 5.615274e-05
\]

This very low p-value allow us to reject the null hypothesis.

## 4 Answer to Selected Exercises

3. The test is significant at the 10% level, but at 5%. The test statistic has the value \( z = -1.876403 \). The p-value, 0.0606, is indicated by the area under the curve and outside the blue vertical lines in the Figure 2. The critical region for the 10% and 5% significance tests are indicated by the inner (10%) and and outer (5%) pair of vertical red lines. Thus \( z \) is in the critical region for a 10% test but on for a 5% test.

6. For \( n_i \) Bernoulli trials, \( x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,n_i}), i = 1, 2 \), we have the likelihood

\[
L(p_1, p_2|x_1, x_2) = p_1^{x_{1,1}} (1 - p_1)^{1-x_{1,1}} \cdots p_1^{x_{1,n_1}} (1 - p_1)^{1-x_{1,n_1}} \cdot p_2^{x_{2,1}} (1 - p_2)^{1-x_{2,1}} \cdots p_2^{x_{2,n_2}} (1 - p_2)^{1-x_{2,n_2}}
\]

To find the maximum likelihood estimator, take logarithms and derivatives with respect to \( p_1 \) and \( p_2 \) to obtain

\[
\hat{p}_1 = \frac{1}{n_1} (x_{1,1} + \cdots + x_{1,n_1}) \quad \text{and} \quad \hat{p}_2 = \frac{1}{n_2} (x_{2,1} + \cdots + x_{2,n_2}).
\]

Then,

\[
L(\hat{p}_1, \hat{p}_2|x_1, x_2) = \hat{p}_1^{n_1} (1 - \hat{p}_1)^{n_1} \hat{p}_2^{n_2} (1 - \hat{p}_2)^{n_2}
\]

Under the null hypothesis, \( p_1 = p_2 \). We set this equal to \( p_0 \) to write the likelihood

\[
L(p_0|x_1, x_2) = p_0^{x_{1,1} + \cdots + x_{1,n_1}} (1 - p_0)^{n_1 - (x_{1,1} + \cdots + x_{1,n_1})} \cdot p_0^{x_{2,1} + \cdots + x_{2,n_2}} (1 - p_0)^{n_2 - (x_{2,1} + \cdots + x_{2,n_2})}
\]

\[
= p_0^{x_{1,1} + \cdots + x_{1,n_1} + x_{2,1} + \cdots + x_{2,n_2}} \cdot (1 - p_0)^{n_1 - (x_{1,1} + \cdots + x_{1,n_1}) + n_2 - (x_{2,1} + \cdots + x_{2,n_2})}
\]

\[
= p_0^{n_1 \hat{p}_1 + n_2 \hat{p}_2} (1 - p_0)^{n_1 (1 - \hat{p}_1) + n_2 (1 - \hat{p}_2)}
\]

\[ \text{Figure 2: The critical values and test statistic value used in Exercise 3.} \]
Again, take logarithms and derivatives with respect to $p_0$ to obtain $\hat{p}_0$ in equation (1), the proportion obtained by pooling the data. Here,

$$L(\hat{p}_0|x_1, x_2) = \frac{\hat{p}_0^{n_1\hat{p}_1} (1 - \hat{p}_0)^{n_1(1 - \hat{p}_1)}}{\hat{p}_0^{n_2\hat{p}_2} (1 - \hat{p}_0)^{n_2(1 - \hat{p}_2)}}$$

Thus, the likelihood ratio,

$$\Lambda(x_1, x_2) = \frac{\frac{\hat{p}_0^{n_1\hat{p}_1}(1 - \hat{p}_0)^{n_1(1 - \hat{p}_1)}}{\hat{p}_0^{n_2\hat{p}_2}(1 - \hat{p}_0)^{n_2(1 - \hat{p}_2)}}}{\Lambda(x_1, x_2)}$$

Again, replace the logarithms with their linear approximation:

$$\ln(1 - \hat{p}_i) - \ln(1 - \hat{p}_0) \approx -\frac{\hat{p}_i - \hat{p}_0}{1 - \hat{p}_0} \ln \hat{p}_i - \ln \hat{p}_0 \approx \frac{\hat{p}_i - \hat{p}_0}{\hat{p}_0}.$$

$$-\ln \Lambda(x_1, x_2) = n_1\hat{p}_1(\ln \hat{p}_1 - \ln \hat{p}_0) + n_1(1 - \hat{p}_1)(\ln(1 - \hat{p}_1) - \ln(1 - \hat{p}_0))$$

$$+ n_2\hat{p}_2(\ln \hat{p}_2 - \ln \hat{p}_0) + n_2(1 - \hat{p}_2)(\ln(1 - \hat{p}_2) - \ln(1 - \hat{p}_0))$$

$$\approx n_1\hat{p}_1 - n_1(1 - \hat{p}_1) + n_2\hat{p}_2 - n_2(1 - \hat{p}_2)$$

$$= n_1(\hat{p}_1 - \hat{p}_0) - n_1(1 - \hat{p}_1) + n_2(\hat{p}_2 - \hat{p}_0)$$

$$= n_1(\hat{p}_1 - \hat{p}_0)^2 + n_2(\hat{p}_2 - \hat{p}_0)^2$$

Now note that

$$n_1(\hat{p}_1 - \hat{p}_0)^2 = n_1 \left(\frac{(n_1 + n_2)\hat{p}_1 - (n_1\hat{p}_1 + n_2\hat{p}_2)}{n_1 + n_2}\right)^2 = n_1 n_2 \left(\frac{\hat{p}_1 - \hat{p}_2}{n_1 + n_2}\right)^2$$

. Perform a similar computation for the second term

$$-\ln \Lambda(x_1, x_2) \approx \frac{n_1 n_2^2 + n_2 n_1^2 ((\hat{p}_1 - \hat{p}_2)/n_1 + n_2)/n_2}{\hat{p}_0(1 - \hat{p}_0)} = \frac{n_1 n_2 (\hat{p}_1 - \hat{p}_2)^2/(n_1 + n_2)}{\hat{p}_0(1 - \hat{p}_0)}$$

$$= \frac{(\hat{p}_1 - \hat{p}_2)^2}{\hat{p}_0(1 - \hat{p}_0) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

9. For the case in which $d = k = 1$. Then, for $n$ observations,

$$\sqrt{n}(c - \hat{c})$$

converges in distribution to a normal random variable with variance $1/I(c)$, the reciprocal of the Fisher information. Thus,

$$\frac{c - \hat{c}}{1/\sqrt{nI(c)}} = \sqrt{n}(c - \hat{c})$$

246
has approximately a standard normal distribution and its square

\[ \frac{n(c - \hat{\theta})^2}{1/I(c)} \]

has a \( \chi^2_1 \) distribution.

We next apply Taylor’s theorem to obtain the quadratic approximation

\[ -2 \ln \Lambda_1(X) = -2 \ln L(c_1|X) + 2 \ln L(\hat{\theta}_1|X) \approx -2(c_1 - \hat{\theta}_1) \frac{d}{d\theta} \ln L(\hat{\theta}_1|X) - (c_1 - \hat{\theta}_1)^2 \frac{d^2}{d\theta^2} \ln L(c_1|X) \]

\[ = -n(c_1 - \hat{\theta}_1)^2 \frac{d^2}{d\theta^2} \ln f(X|c_1) \]

The linear term vanishes because \( \frac{d}{d\theta} \ln L(\theta_1|X) = 0 \) at \( \hat{\theta}_1 \).

We now apply the law of large numbers to see that

\[ \frac{1}{n} \frac{d^2}{d\theta^2} \ln L(c_1|X) = \frac{1}{n} \sum_{i=1}^{n} f_X(x_i|c_1) \to -I(c). \]

Thus,

\[ -2 \ln \Lambda_1(X) \approx n(c_1 - \hat{\theta}_1)^2 \times -I(c) = \frac{n(c - \hat{\theta})^2}{1/I(c)} \]

11. Taking logarithms, we find the score function,

\[ \ln L(\lambda_m, \lambda_f|x_m,x_f) = n_m \ln \lambda - \ln(n_m!) - \lambda_m + n_f \ln \lambda - \ln(n_f!) - \lambda_f. \]

The derivative with respect to \( \lambda_m \) is

\[ \frac{\partial}{\partial \lambda_m} \ln L(\lambda|x_m,x_f) = \frac{n_m}{\lambda_m} - 1. \]

Now set this equal to 0 and solve for \( \lambda_m \). Because the second derivative with respect to \( \lambda_m \) is negative, this is a maximum. A nearly identical computation can be used to find \( \lambda_f \).

12. Taking logarithms, we find the score function,

\[ \ln L(\lambda|x_m,x_f) = (n_m + n_f) \ln \lambda - \ln(n_m!n_f!) - 2\lambda. \]

The derivative with respect to \( \lambda \) is

\[ \frac{\partial}{\partial \lambda} \ln L(\lambda|x_m,x_f) = \frac{n_m + n_f}{\lambda} - 2. \]

Now set this equal to 0 and solve for \( \lambda \). Because the second derivative with respect to \( \lambda \) is negative, this is a maximum.