Likelihood-ratio test - Wikipedia, the free encyclopedia

In statistics, a likelihood ratio test is a statistical test used to compare the fit of two models, one of which (the null model) is a special case of the other (the alternative model). The test is based on the likelihood ratio, which expresses how many times more likely the data are under one model than the other. This likelihood ratio, or equivalently its logarithm, can then be used to compute a p-value, or compared to a critical value to decide whether to reject the null model in favour of the alternative model. When the logarithm of the likelihood ratio is used, the statistic is known as a log-likelihood ratio statistic, and the probability distribution of this test statistic, assuming that the null model is true, can be approximated using Wilks' theorem.

In the case of distinguishing between two models, each of which has no unknown parameters, use of the likelihood ratio test can be justified by the Neyman–Pearson lemma, which demonstrates that such a test has the highest power among all competitors.[1]

Use

Each of the two competing models, the null model and the alternative model, is separately fitted to the data and the log-likelihood recorded. The test statistic (often denoted by $D$) is twice the difference in these log-likelihoods:

$$D = -2 \ln \left( \frac{\text{likelihood for null model}}{\text{likelihood for alternative model}} \right)$$

$$= -2 \ln(\text{likelihood for null model}) + 2 \ln(\text{likelihood for alternative model})$$

model with more parameters will always fit at least as well (have a greater log-likelihood). Whether it fits significantly better and should thus be preferred is determined by deriving the probability or p-value of the difference $D$. Where the null hypothesis represents a special case of the alternative hypothesis, the probability distribution of the test statistic is approximately a chi-squared distribution with degrees of freedom equal to $\text{df}_2 - \text{df}_1$. [2] Symbols $\text{df}_1$ and $\text{df}_2$ represent the number of free parameters of models 1 and 2, the null model and the alternative model, respectively. The test requires nested models, that is: models in which the more complex one can be transformed into the simpler model by imposing a set of constraints on the parameters. [3]

For example: if the null model has 1 free parameter and a log-likelihood of −8024 and the
alternative model has 3 degrees of freedom and a LL of −8012, then the probability of this difference is that of chi-squared value of +2(8024 − 8012) = 24 with 3 − 1 = 2 degrees of freedom. Certain assumptions must be met for the statistic to follow a chi-squared distribution and often empirical p-values are computed.

**Background**

The likelihood ratio, often denoted by (the capital Greek letter lambda), is the ratio of the likelihood function varying the parameters over two different sets in the numerator and denominator. A likelihood-ratio test is a statistical test for making a decision between two hypotheses based on the value of this ratio.

It is central to the Neyman–Pearson approach to statistical hypothesis testing, and, like statistical hypothesis testing generally, is both widely used and much criticized.

This article needs additional citations for verification. (July 2012)

**Simple-versus-simple hypotheses**

Main article: Neyman–Pearson lemma

A statistical model is often a parametrized family of probability density functions or probability mass functions. A simple-vs-simple hypotheses test has completely specified models under both the null and alternative hypotheses, which for convenience are written in terms of fixed values of a notional parameter:

\[
\begin{align*}
H_0 : \theta &= \theta_0, \\
H_1 : \theta &= \theta_1.
\end{align*}
\]

Note that under either hypothesis, the distribution of the data is fully specified; there are no unknown parameters to estimate. The likelihood ratio test statistic can be written as:

\[
\Lambda(x) = \frac{L(\theta_0 | x)}{L(\theta_1 | x)} = \frac{f(x | \theta_0)}{f(x | \theta_1)}
\]

or

\[
\Lambda(x) = \frac{L(\theta_0 | x)}{\sup \{ L(\theta | x) : \theta \in \{\theta_0, \theta_1\} \}}
\]

where is the likelihood function. Note that some references may use the reciprocal as the definition. In the form stated here, the likelihood ratio is small if the alternative model is better than the null model and the likelihood ratio test provides the decision rule as:

If

\[
L(\theta | x)
\]

The likelihood ratio, often denoted by (the capital Greek letter lambda), is the ratio of the likelihood function varying the parameters over two different sets in the numerator and denominator. A likelihood-ratio test is a statistical test for making a decision between two hypotheses based on the value of this ratio.
The values are usually chosen to obtain a specified significance level, through the relation:

\[ q \cdot P(\Lambda = c \mid H_0) + P(\Lambda < c \mid H_0) = \alpha \]

[citation needed] The Neyman-Pearson lemma states that this likelihood ratio test is the most powerful among all level tests for this problem.[citation needed]

Definition (likelihood ratio test for composite hypotheses)

A null hypothesis is often stated by saying the parameter \( \theta \) is in a specified subset \( \Theta_0 \) of the parameter space \( \Theta \).

The likelihood function is

\( L(\theta|x) = f(x|\theta) \)

with

\( f(x|\theta) \)

\( \theta \)
A likelihood ratio test is any test with critical region (or rejection region) of the form

\[ \Lambda(x) = \frac{\sup \{ L(\theta \mid x) : \theta \in \Theta_0 \}}{\sup \{ L(\theta \mid x) : \theta \in \Theta \}} \]

where

\( \Lambda(x) \)

is any number satisfying

\[ \sup \{ x \mid \Lambda \leq c \} \leq c \leq 1 \quad \text{and} \quad x \]

Many common test statistics such as the Z-test, the F-test, Pearson's chi-squared test and the G-test are tests for nested models and can be phrased as log-likelihood ratios or approximations thereof.

**Interpretation**

Being a function of the data, the LR is therefore a statistic. The likelihood-ratio test rejects the null hypothesis if the value of this statistic is too small. How small is too small depends on the significance level of the test, i.e., on what probability of Type I error is considered tolerable ("Type I" errors consist of the rejection of a null hypothesis that is true).

The numerator corresponds to the maximum likelihood of an observed outcome under the null hypothesis. The denominator corresponds to the maximum likelihood of an observed outcome varying parameters over the whole parameter space. The numerator of this ratio is less than the denominator. The likelihood ratio hence is between 0 and 1. Lower values of the likelihood ratio mean that the observed result was much less likely to occur under the null hypothesis as compared to the alternative. Higher values of the statistic mean that the observed outcome was more than or equally likely or nearly as likely to occur under the null hypothesis as compared to the alternative, and the null hypothesis cannot be rejected.

**Distribution: Wilks' theorem**

If the distribution of the likelihood ratio corresponding to a particular null and alternative hypothesis can be explicitly determined then it can directly be used to form decision regions (to accept/reject the null hypothesis). In most cases, however, the exact distribution of the likelihood ratio corresponding to specific hypotheses is very difficult to determine. A convenient result, attributed to Samuel S. Wilks, says that as the sample size approaches , the test statistic for a nested model will be asymptotically distributed with degrees of freedom equal to the difference in dimensionality of

\[ n \]

and

\[ \sum \log(\Lambda) \]

This means that for a great variety of hypotheses, a practitioner can compute the likelihood ratio

\[ \chi^2 \]

for the data and compare
to the chi squared value corresponding to a desired statistical significance as an approximate statistical test.

**Examples**

**Coin tossing**

An example, in the case of Pearson's test, we might try to compare two coins to determine whether they have the same probability of coming up heads. Our observation can be put into a contingency table with rows corresponding to the coin and columns corresponding to heads or tails. The elements of the contingency table will be the number of times the coin for that row came up heads or tails. The contents of this table are our observation.

<table>
<thead>
<tr>
<th></th>
<th>Heads</th>
<th>Tails</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coin 1</td>
<td>$k_{1H}$</td>
<td>$k_{1T}$</td>
</tr>
<tr>
<td>Coin 2</td>
<td>$k_{2H}$</td>
<td>$k_{2T}$</td>
</tr>
</tbody>
</table>

Here consists of the parameters

\[ \Theta \]

\[ \Theta_0 \]

\[ \Lambda \]

\[ -2 \log(\Lambda) \]

which are the probability that coins 1 and 2 come up heads or tails. The hypothesis space is defined by the usual constraints on a distribution,

\[ H \]

\[ 0 \leq p_{ij} \leq 1 \]

\[ p_{iH} + p_{iT} = 1 \]

The null hypothesis is the sub-space where

In all of these constraints,
and

Writing

for the best values for

under the hypothesis

, maximum likelihood is achieved with

Writing

for the best values for

under the null hypothesis

, maximum likelihood is achieved with

which does not depend on the coin .

The hypothesis and null hypothesis can be rewritten slightly so that they satisfy the constraints for the logarithm of the likelihood ratio to have the desired nice distribution.

Since the constraint causes the two-dimensional
to be reduced to the one-dimensional

, the asymptotic distribution for the test will be

, the
distribution with one degree of freedom.

For the general contingency table, we can write the log-likelihood ratio statistic as

References


External links

- **Practical application of Likelihood-ratio test described**

- **Vassar College's Likelihood Ratio Given Sensitivity/Specificity/Prevalence Online Calculator**