Likelihood-ratio test

From Wikipedia, the free encyclopedia

In statistics, a **likelihood ratio test** is a statistical test used to compare the goodness of 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's 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]

Contents

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:

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

The model with more parameters will always fit at least as well (have an equal or 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 $df2 - df1$.[2] Symbols $df1$ and $df2$ represent the number of free parameters of models 1 and 2, the null model and the alternative model, respectively.

Here is an example of use. If the null model has 1 parameter and a log-likelihood of −8024 and the alternative model has 3 parameters and a log-likelihood of −8012, then the probability of this difference is that of chi-squared value of $+2 \cdot (8024 - 8012) = 24$ with $3 - 1 = 2$ degrees of freedom. Certain assumptions[3] must be met for the statistic to follow a chi-squared distribution, and often empirical p-values are computed.

The likelihood-ratio test requires nested models, i.e. models in which the more complex one can be transformed into the simpler model by imposing a set of constraints on the parameters. If the models are not nested, then a generalization of the likelihood-ratio test can usually be used instead: the relative likelihood.

Simple-vs-simple hypotheses

A statistical model is often a parametrized family of probability density functions or probability mass functions $f(x|\theta)$. A simple-vs-simple hypothesis 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 θ:

$$H_0 : \quad \theta = \theta_0,$$
$$H_1 : \quad \theta = \theta_1.$$

Note that under either hypothesis, the distribution of the data is fully specified; there are no unknown parameters to estimate. The likelihood ratio test is based on the **likelihood ratio**, which is often denoted by Λ (the capital Greek letter lambda). The likelihood ratio is defined as follows:[4][5]

$$\Lambda(x) = \frac{L(\theta_0|x)}{L(\theta_1|x)} = \frac{f(\cup_i x_i|\theta_0)}{f(\cup_i x_i|\theta_1)}$$

or

$$\Lambda(x) = \frac{L(\theta_0 \mid x)}{\sup\{ L(\theta \mid x) : \theta \in \{\theta_0, \theta_1\}\}},$$

where $L(\theta|x)$ is the likelihood function, and $\sup$ is the supremum function. Note that some references may use the reciprocal as the definition.[6] 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 follows:

If $\Lambda > c$, do not reject H_0;

If $\Lambda < c$, reject H_0;

Reject with probability q if $\Lambda = c$.

The values c, q are usually chosen to obtain a specified significance level α, through the relation $q \cdot P(\Lambda = c \mid H_0) + P(\Lambda < c \mid H_0) = \alpha$. The Neyman-Pearson lemma states that this likelihood ratio test is the most powerful among all level α tests for this problem.[1]

Definition (likelihood ratio test for composite hypotheses)

A null hypothesis is often stated by saying the parameter θ is in a specified subset Θ_0 of the parameter space Θ.

$$H_0 : \quad \theta \in \Theta_0$$
$$H_1 : \quad \theta \in \Theta_0^C$$

The likelihood function is $L(\theta|x) = f(x|\theta)$ (with $f(x|\theta)$ being the pdf or pmf), which is a function of the parameter θ with x held fixed at the value that was actually observed, *i.e.*, the data. The **likelihood ratio test statistic** is [7]

$$\Lambda(x) = \frac{\sup\{\, L(\theta \mid x) : \theta \in \Theta_0\,\}}{\sup\{\, L(\theta \mid x) : \theta \in \Theta\,\}}.$$

Here, the $\sup$ notation refers to the supremum function.

A **likelihood ratio test** is any test with critical region (or rejection region) of the form $\{x | \Lambda \le c\}$ where c is any number satisfying $0 \le c \le 1$. 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 x, the likelihood ratio 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. Low values of the likelihood ratio mean that the observed result was less likely to occur under the null hypothesis as compared to the alternative. High values of the statistic mean that the observed outcome was nearly as likely to occur under the null hypothesis as the alternative, and the null hypothesis cannot be rejected.

Distribution: Wilks's 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 n approaches ∞, the test statistic $-2\log(\Lambda)$ for a nested model will be asymptotically χ^2-distributed with degrees of freedom equal to the difference in dimensionality of Θ and Θ_0.[3] This means that for a great variety of hypotheses, a practitioner can compute the likelihood ratio Λ for the data and compare $-2\log(\Lambda)$ to the χ^2 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 X.

	Heads	Tails
Coin 1	k_{1H}	k_{1T}
Coin 2	k_{2H}	k_{2T}

Here Θ consists of the possible combinations of values of the parameters p_{1H}, p_{1T}, p_{2H}, and p_{2T}, which are the probability that coins 1 and 2 come up heads or tails. In what follows, $i = 1, 2$ and $j = H, T$. The hypothesis space H is constrained by the usual constraints on a probability distribution, $0 \le p_{ij} \le 1$, and $p_{iH} + p_{iT} = 1$. The space of the null hypothesis H_0 is the subspace where $p_{1j} = p_{2j}$. Writing n_{ij} for the best values for p_{ij} under the hypothesis H, the maximum likelihood estimate is given by

$$n_{ij} = \frac{k_{ij}}{k_{iH} + k_{iT}}.$$

Similarly, the maximum likelihood estimates of p_{ij} under the null hypothesis H_0 are given by

$$m_{ij} = \frac{k_{1j} + k_{2j}}{k_{1H} + k_{2H} + k_{1T} + k_{2T}},$$

which does not depend on the coin i.

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 H to be reduced to the one-dimensional H_0, the asymptotic distribution for the test will be $\chi^2(1)$, the χ^2 distribution with one degree of freedom.

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

$$-2 \log \Lambda = 2 \sum_{i,j} k_{ij} \log \frac{n_{ij}}{m_{ij}}.$$

References

1. Neyman, Jerzy; Pearson, Egon S. (1933). "On the Problem of the Most Efficient Tests of Statistical Hypotheses". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **231** (694–706): 289–337. Bibcode:1933RSPTA.231..289N (http://adsabs.harvard.edu/abs/1933RSPTA.231..289N). doi:10.1098/rsta.1933.0009 (https://dx.doi.org/10.1098%2Frsta.1933.0009). JSTOR 91247 (https://www.jstor.org/stable/91247).
2. Huelsenbeck, J. P.; Crandall, K. A. (1997). "Phylogeny Estimation and Hypothesis Testing Using Maximum Likelihood". *Annual Review of Ecology and Systematics* **28**: 437–466. doi:10.1146/annurev.ecolsys.28.1.437 (https://dx.doi.org/10.1146%2Fannurev.ecolsys.28.1.437).
3. Wilks, S. S. (1938). "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". *The Annals of Mathematical Statistics* **9**: 60–62. doi:10.1214/aoms/1177732360 (https://dx.doi.org/10.1214%2Faoms%2F1177732360).
4. Mood, A.M.; Graybill, F.A. (1963) *Introduction to the Theory of Statistics*, 2nd edition. McGraw-Hill ISBN 978-0070428638 (page 286)

5. Stuart, A.; Ord, K.; Arnold, S. (1999) *Kendall's Advanced Theory of Statistics, Volume 2A*, Arnold. (Chapter 22).
6. Cox, D. R.; Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall. p. 92. ISBN 0-412-12420-3.
7. Casella, George; Berger, Roger L. (2001). *Statistical Inference* (Second ed.). p. 375. ISBN 0-534-24312-6.

External links

- Practical application of likelihood ratio test described (http://www.itl.nist.gov/div898/handbook/apr/section2/apr233.htm)
- R Package: Wald's Sequential Probability Ratio Test (http://cran.r-project.org/web/packages/SPRT/SPRT.pdf)
- Richard Lowry's Predictive Values and Likelihood Ratios (http://faculty.vassar.edu/lowry/clin2.html) Online Clinical Calculator

Retrieved from "https://en.wikipedia.org/w/index.php?title=Likelihood-ratio_test&oldid=668781783"

Categories: Statistical ratios │ Statistical tests │ Statistical theory