

Additions

Kevin O'Brien

August 15, 2010

Contents

0.1	Estimation methods	1
0.2	Algorithms	2
0.3	Likelihood ratio tests	4

- LMEs
- Likelihood and and log likelihood functions
- Likelihood ratio test
- more on score functions etc
- MLEs
- Algorithms

0.1 Estimation methods

Likelihood is the hypothetical probability that an event that has already occurred would yield a specific outcome. Likelihood differs from probability in that probability refers to future occurrences, while likelihood refers to past known outcomes.

The Likelihood function is a fundamental concept in statistical inference. It indicates how likely a particular population is to produce an observed sample. The set of values that maximize the likelihood function are considered to be optimal, and are used as the estimates of the parameters.

Maximum likelihood (ML) estimation is a method of obtaining parameter estimates by optimizing the likelihood function. The likelihood function is constructed as a function of the parameters in the specified model.

Restricted maximum likelihood (REML) is an alternative methods of computing parameter estimated. REML is often preferred to ML because it produces unbiased estimates of covariance parameters by taking into account the loss of degrees of freedom that results from estimating the fixed effects in β .

REML estimation reduces the bias in the variance component, and also handles high correlations more effectively, and is less sensitive to outliers than ML. The problem with REML for model building is that the "likelihoods" obtained for different fixed effects

are not comparable. Hence it is not valid to compare models with different fixed effects using a likelihood ratio test or AIC when REML is used to estimate the model. Therefore models derived using ML must be used instead.

0.2 Algorithms

The most common methods of estimating is the Newton-Raphson method, the EM algorithm and the Fisher scoring methods.

The expectation-maximization (EM) algorithm is an iterative technique, alternating between performing an expectation (E) step and the maximisation (M) step. The ‘E’ step computes the expectation of the log-likelihood evaluated using the current estimate for the variables. In the ‘M’ step, parameters that maximise the expected log-likelihood, found on the previous ‘E’ step, are computed. These parameter estimates are then used to determine the distribution of the variables in the next ‘E’ step. The algorithm alternatives between these two steps until convergence is reached.

The main drawback of the EM algorithm is its slow rate of convergence.

The EM algorithm is rarely used entirely in LME estimation, instead providing an initial set of values that can be passed to other optimization techniques.

Newton Raphson methods are the most common techniques for ML and REML estimation. The NR algorithm minimizes an objective function defined as -2 times the log likelihood for the covariance parameters. At every iteration the NR algorithm requires the calculation of a vector of partial derivatives, known as the gradient, and the second derivative matrix with respect to the covariance parameters. This is known as the observed hessian matrix. the fisher scoring algorithm is an adaptation of the NR algorithm, using the expected hessian matrix rather than the observed one.

The optimization algorithms need to ensure that the D and R estimates are positive definite.

Assuming a statistical model $f_\theta(y)$ parameterized by a fixed and unknown set of parameters θ , the likelihood $L(\theta)$ is the probability of the observed data y considered as a function of θ (?).

The log likelihood $l(\theta)$

0.3 Likelihood ratio tests

Likelihood ratio tests are a class of tests based on the comparison of the values of the likelihood functions of two candidate models. LRTs can be used to test hypotheses about covariance parameters or fixed effects parameters in the context of LMEs.

The test statistic for the LRT is the difference of the log-likelihood functions, multiplied by -2 . The probability distribution of the test statistic is approximated by the χ^2 distribution with $(\nu_1 - \nu_2)$ degrees of freedom, where ν_1 and ν_2 are the degrees of freedom of models 1 and 2 respectively.

The score function $S(\theta)$ is the derivative of the log likelihood with respect to θ ,

$$S(\theta) = \frac{\partial}{\partial \theta} l(\theta),$$

and the maximum likelihood estimate is the solution to the score equation

$$S(\theta) = 0.$$

The Fisher information $I(\theta)$, which is defined as

$$I(\theta) = -\frac{\partial^2}{\partial \theta^2} l(\theta),$$

give rise to the observed Fisher information ($I(\hat{\theta})$) and the expected Fisher information ($\mathcal{I}(\theta)$).