Additions

Kevin O'Brien

September 9, 2015

Contents

0.1 Likelihood and estimation

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 $\boldsymbol{\beta}$.

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.

0.1.1 Likelihood Ratio Tests

The relationship between the respective models presented by **?** is known as "nesting". A model A to be nested in the reference model, model B, if Model A is a special case of Model B, or with some specific constraint applied.

A general method for comparing nested models fitted by ML is the ***likelihood ratio test*** (Cite: Lehmann 1986). Such a test can also be used for models fitted using REML, but only if both models have been fitted by REML, and if the fixed effects specification is the same for both models.

If k_i is the number of parameters to be estimated in model i, then the asymptotic, or "large sample", distribution of the LRT statistic, under the null hypothesis that the restricted model is adequate, is a χ^2 distribution with $k_2 - k_1$ degrees of freedom (**?**, pg.83).

We generally use LRTs to evaluate the significance of terms in the random effects structure, i.e. different nested models are fitted in which the random effects structure is changed.

A general method for comparing models with a nesting relationship is the likelihood ratio test (LRTs). LRTs are a family of tests used to compare the value of likelihood functions for two models, whose respective formulations define a hypothesis to be tested (i.e. the nested and reference model). The significance of the likelihood ratio test can be found by comparing the likelihood ratio to the χ^2 distribution, with the appropriate degrees of freedom.

When testing hypotheses around covariance parameters in an LME model, REML estimation for both models is recommended by West et al. REML estimation can be shown to reduce the bias inherent in ML estimates of covariance parameters (**?**). Conversely, **?** advises that testing hypotheses on fixed-effect parameters should be based on ML estimation, and that using REML would not be appropriate in this context.

0.2 Model Selection Using Likelihood Ratio Tests

An important step in the process of model selection is to determine, for a given pair of models, if there is a "nesting relationship" between the two.

We define Model A to be "nested" in Model B if Model A is a special case of Model B, i.e. Model B with a specific constraint applied.

One model is said to be *nested* within another model, i.e. the reference model, if it represents a special case of the reference model (**?**).

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 $(\mathcal{I}(\theta))$.

0.2.1 Likelihood Ratio Tests

The relationship between the respective models presented by **?** is known as "nesting". A model A to be nested in the reference model, model B, if Model A is a special case of Model B, or with some specific constraint applied.

A general method for comparing models with a nesting relationship is the likelihood ratio test (LRTs). LRTs are a family of tests used to compare the value of likelihood functions for two models, whose respective formulations define a hypothesis to be tested (i.e. the nested and reference model). The significance of the likelihood ratio test can be found by comparing the likelihood ratio to the χ^2 distribution, with the appropriate degrees of freedom.

When testing hypotheses around covariance parameters in an LME model, REML estimation for both models is recommended by West et al. REML estimation can be shown to reduce the bias inherent in ML estimates of covariance parameters (**?**). Conversely, **?** advises that testing hypotheses on fixed-effect parameters should be based on ML estimation, and that using REML would not be appropriate in this context.

0.3 Likelihood Ratio Tests

PB on LRTS for LMEs

Pinheiro & Bates (2000; p. 88) argue that Likelihood Ratio Test comparisons of models varying in fixed effects tend to be anticonservative i.e. will see you observe significant differences in model fit more often than you should.

I think they are talking, especially, about situations in which the number of model parameter differences (differences between the complex model and the nested simpler model) is large relative to the number of observations.

This is not really a worry for this dataset, but I will come back to the substance of this view, and alternatives to the approach taken here.

0.3.1 Pinheiro Bates

A general method for comparing nested models fitted by ML is the ***likelihood ratio test*** (Cite: Lehmann 1986). Such a test can also be used for models fitted using REML, but only if both models have been fitted by REML, and if the fixed effects specification is the same for both models.

If k_i is the number of parameters to be estimated in model i, then the asymptotic, or "large sample", distribution of the LRT statistic, under the null hypothesis that the restricted model is adequate, is a χ^2 distribution with $k_2 - k_1$ degrees of freedom (**?**, pg.83).

We generally use LRTs to evaluate the significance of terms in the random effects structure, i.e. different nested models are fitted in which the random effects structure is changed.

0.3.2 Empirical p-values of LRT tests

For both REML and ML estimates, the nominal $p-$values for the LRT statistics under a χ^2 distribution with 2 degrees of freedom are much greater than empirical values. A number of ways of dealing with this issues are discussed (**?**, pg.86).

One should be aware that these p-values may be conservative. That is, the reported p-value may be greater than the true p-value for the test and, in some cases, it may be much greater.(**?**, pg.87).

0.3.3 Other material

A general method for comparing nested models fit by maximum likelihood is the ***likelihood ratio test***. This test can be used for models fit by REML (restricted maximum liklihood), but only if the fixed terms in the two models are invariant, and both models have been fit by REML. Otherwise, the argument: method="ML" must be employed (ML = maximum likelihood).

- Example of a likelihood ratio test used to compare two models:
  `>anova(modelA, modelB)`

- The output will contain a p-value, and this should be used in conjunction with the AIC scores to judge which model is preferred. Lower AIC scores are better.

- Generally, likelihood ratio tests should be used to evaluate the significance of terms on the random effects portion of two nested models, and should not be used to determine the significance of the fixed effects.

- A simple way to more reliably test for the significance of fixed effects in an LME model is to use conditional F-tests, as implemented with the simple "anova" function. Example:
  `>anova(modelA)`

  will give the most reliable test of the fixed effects included in model1.

0.3.4 Nested and Reference Models

Hypotheses can be formulated in the context of a pair of models that have a nesting relationship [CITE: West et al].

LRTs are a class of tests used to compare the value of likelihood functions for two models defining a hypothesis to be tested (i.e. the nested and reference model).

The significance of the likelihood ratio test can be found by comparing it to the χ^2 distribution, with the appropriate degrees of freedom.

0.3.5 LRTs for covariance parameters

[cite: West et al] When testing hypotheses around covariance parameters in an LME model, REML estimation for both models is recommended by West et al. REML estimation can be shown to reduce the bias inherent in ML estimates of covariance parameters [cite: Morrel98]

0.3.6 Likelihood Ratio Tests

The relationship between the respective models presented by **?** is known as "nesting". A model A to be nested in the reference model, model B, if Model A is a special case of Model B, or with some specific constraint applied.

A general method for comparing models with a nesting relationship is the likelihood ratio test (LRTs). LRTs are a family of tests used to compare the value of likelihood functions for two models, whose respective formulations define a hypothesis to be tested (i.e. the nested and reference model). The significance of the likelihood ratio test can be found by comparing the likelihood ratio to the χ^2 distribution, with the appropriate degrees of freedom.

When testing hypotheses around covariance parameters in an LME model, REML estimation for both models is recommended by West et al. REML estimation can be shown to reduce the bias inherent in ML estimates of covariance parameters (**?**). Conversely, **?** advises that testing hypotheses on fixed-effect parameters should be based on ML estimation, and that using REML would not be appropriate in this context.

- LMEs

- Likelihood and and log likelihood functions

- Likelihood ratio test

- more on score functions etc

- MLEs

- Algorithms

The estimate for the fixed effects are referred to as the best linear unbiased estimates (BLUE). Henderson's estimate for the random effects is known as the best linear unbiased predictor (BLUP).

0.3.7 Likelihood Ratio Tests

L= - 2ln is approximately distributed as 2 under H_0 for large sample size and under the normality assumption.

The power of the likelihood ratio test may depends on specific sample size and the specific number of replications, and [Roy 2009] proposes simulation studies to examine this further.

0.3.8 Relevance of Estimation Methods

Nested LME models, fitted by ML estimation, can be compared using the likelihood ratio test [Lehmann (1986)]. Models fitted using REML estimation can also be compared, but only if both were fitted using REML, and both have the same fixed effects specifications.

Likelihood ratio tests are generally used to test the significance of terms in the random effects structure. Information Criteria Additionally nested models may be compared by using the Akaike Information Criterion,(AIC) and the Bayesian Information Criterion (BIC).

When comparing the respective scores for nested models, the model with the smaller score is considered to be the preferable model. ML / REML [Morrell 1998] The variance components in the LME model may be estimated by ML or REML. Maximum Likelihood estimates do not take into account the estimation of fixed effects and so are biased downwards. REML estimates accounts for the presence of these nuisance parameters by maximising the linearly independent error contrasts to obtain more unbiased estimates. Treatment of items as fixed effects [Pinheiro Bates 2000] addresses the issue of treating items as fixed effects. Such a specification is useful only for the specific sample of items, rather than the population of items, where the interest would naturally lie.

[Pinheiro Bates 2000] advises the specification of random effects to correspond to items; treating the item effects as random deviations from the population mean.

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 $l(\theta)$

? used ML estimation to estimate the true correlation between the variables when the measurements are linked over time. The methodology relies on the assumption that the two variables with repeated measures follow a multivariate normal distribution. The methodology currently does not extend to any more than two cases. The MLE of the correlation takes into account the dependency among repeated measures.

The true correlation ρ_{xy} is repeated measurements can be considered as having two components: between subject and within-subject correlation. The usefulness of estimating repeated measure correlation coefficients is the calculation of between-method and within-method variabilities are produced as by-products.

0.3.9 Other material

A general method for comparing nested models fit by maximum likelihood is the ***likelihood ratio test***. This test can be used for models fit by REML (restricted maximum liklihood), but only if the fixed terms in the two models are invariant, and both models have been fit by REML. Otherwise, the argument: method="ML" must be employed (ML = maximum likelihood).

- Example of a likelihood ratio test used to compare two models:
  `>anova(modelA, modelB)`

- The output will contain a p-value, and this should be used in conjunction with the AIC scores to judge which model is preferred. Lower AIC scores are better.

- Generally, likelihood ratio tests should be used to evaluate the significance of terms on the random effects portion of two nested models, and should not be used to determine the significance of the fixed effects.

- A simple way to more reliably test for the significance of fixed effects in an LME model is to use conditional F-tests, as implemented with the simple "anova" function. Example:
  `>anova(modelA)`

  will give the most reliable test of the fixed effects included in model1.

0.3.10 Nested and Reference Models

Hypotheses can be formulated in the context of a pair of models that have a nesting relationship [CITE: West et al].

LRTs are a class of tests used to compare the value of likelihood functions for two models defining a hypothesis to be tested (i.e. the nested and reference model).

The significance of the likelihood ratio test can be found by comparing it to the χ^2 distribution, with the appropriate degrees of freedom.

0.3.11 Test for inter-method bias

Bias is determinable by examination of the 't-table'. Estimate for both methods are given, and the bias is simply the difference between the two. Because the R implementation does not account for an intercept term, a $p-$value is not given. Should a $p-$value be required specifically for the bias, and simple restructuring of the model is required wherein an intercept term is included. Output from a second implementation will yield a $p-$value.