dentist: Computing uncertainty by sampling points around maximum likelihood estimates

James D. Boyko1\* and Brian O’Meara1

1Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA

\*To whom correspondence should be addressed (jboyko@utk.edu).

Running head: Computing MLE uncertainty

Key words: confidence intervals, phylogenetic comparative methods, likelihood surface, Hessian

Abstract

1. dentist is an R-package that “dents” the likelihood surface by sampling points a specified distance around the maximum likelihood estimate. This allows an estimate of confidence intervals around parameter estimates without an analytic solution to likelihood equations.

2. We describe the importance of estimating uncertainty around parameter estimates as well as demonstrate the ability of dentist to accurately estimate confidence intervals. We introduce several plotting tools to visualize the results of a dentist analysis.

3. dentist is freely available from https://github.com/bomeara/dentist, written in the R language, and can be used for any given likelihood function.

**Introduction**

At the core of the scientific method is determining whether a particular hypothesis fits with a set of observations better than competing theories. Understanding how observations can provide evidence for hypotheses and quantifying the level of belief in these hypotheses has been the subject of much statistical development over the past 100 years (refs.). In that time, statistics has played an increasingly important role in the methodology of scientific inquiry. One of the most important of these developments has been the theory of likelihood, which is a rigorous quantification of belief (Fisher 1955, Edwards, B&A). A set of parameters is used to describe our hypothesis in the form of a model. Our belief in the model is then quantified via likelihood (or other subsidiaries) when fit to data. In biological applications, these parameter values represent a description of ecological and evolutionary processes such as speciation rates, mutation rates, migration rates, transitions between discrete states, or rates of phenotypic evolution (refs). However, whether we should *act* as if a hypothesis is true has often been overshadowed by measuring whether we believe it *is* true (Neyman 1956, Pearson 1955).

Likelihood is a useful tool.

We need to think more about parameter uncertainty.

The two main methods are analytic and bootstrap.

But these are either computationally intensive or difficult for complex likelihoods.

Here we introduce Dentist which works by denting the likelihood surface.

We demonstrate its use and interpretation.

We end by discussing future work and extensions.

The goal of a likelihood is to… Likelihood approaches have three questions that must be answered how do we find the maximum, how do we compare across datasets, and how do we find the curve around that area. This third point is often ignored in our field of comparative methods. There are several programs such as GLM, or ttests which have well defined distributions for which the confidence intervals are analytically known. However, in comparative methods and other complicated likelihood functions there is no analytic solution. In these cases, it is still important to determine good estimates of the confidence intervals of our parameters. Furthermore, there is a growing concern about likelihood ridges and identifiability issues. These have come up across several comparative methods for discrete character evolution, continuous character evolution, and speciation extinction methods. In all cases, it can be beneficial to examine the likelihood surface to look for ridges and determine identifiability.

Take down of priors?

The importance of uncertainty estimates

Uncertainty, beyond identifying places where we cannot make estimates, gives us a way to discuss our results with differing confidence (O’Meara). It can often give a way to think and focus on particular results that are not only interesting from the apriori hypotheses, but gives authors a way to parse their results in an interesting way.

“We believe AIC is suitable for this selection purpose and that the only additional consideration is thus to get reliable unconditional sampling variances (and confidence intervals) for MLEs after model selection.” “The matter of a (1 − α)100% unconditional confidence interval is now consid- ered. We have two general approaches: the bootstrap (see, e.g., Buckland et al. 1997), or analytical formulas based on analysis results from just the one data set. The analytical approach requires less computing; hence we start with it.” “In the case that Q 􏲈 1 (θ is unique to one model in the set of R models), none of the above results can be used. In this case it seems that there may not be a direct way to include model selection uncertainty into the uncertainty about the value of θ. An approach we can envision here is to adjust upward the conditional sampling variance estimator, v􏱫ar(θ | gi ), by some variance inflation factor.” B&A

**Alternative approaches and benefits of dentist**

Profile likelihood is the likelihood of a subset of parameters with the remaining parameters being fixed at their maximum likelihood estimate given the former (Murphy and Van Der Vaart). Profile likelihoods have been leveraged to calculate confidence intervals (Venzon and Moolgavkar, Meyer and Hill).

The parametric bootstrap is able to produce confidence intervals for parameters without analytic solutions to maximum likelihood equations at the cost of increased computational effort (Efron). These methods have been applied in phylogenetic comparative methods to some degree, but their ease of use and interpretations begs for more application (Boettiger et al.).

Fisher Information and the second derivative and hessian matrix relies on the asymptotic normality of the maximum likelihood estimate. This is a measure which quantifies how much information a random variable carries about its unknown generating parameters.

**The underlying algorithm of dentist**

The m-unit support limits for a parameter are the two parameters astride the evaluate at which the support is m units less than the maximum. The m-unit support region for a number of parameters is the region in the parameter space bounded by the curve on which support is m units less than the maximum (Edwards).

This proposes new values using a normal distribution centered on the original parameter values.

**Example usage with known distributions**

This will show the univariate plots of the parameter values versus the likelihood as well as bivariate plots of pairs of parameters to look for ridges.

**Known distributions**

**Use code shown from O’meara.**

Implementation in other packages

OUwie and corHMM.

**Conclusions**

This "dents" the likelihood surface by reflecting points better than a threshold back across the threshold (think of taking a hollow plastic model of a mountain and punching the top so it's a volcano). It then uses essentially a Metropolis-Hastings walk to wander around the new rim. It adjusts the proposal width so that it samples points around the desired likelihood.

This is better than using the curvature at the maximum likelihood estimate since it can actually sample points in case the assumptions of the curvature method do not hold. It is better than varying one parameter at a time while holding others constant because that could miss ridges: if I am fitting 5=x+y, and get a point estimate of (3,2), the reality is that there are an infinite range of values of x and y that will sum to 5, but if I hold x constant it looks like y is estimated very precisely. Of course, one could just fully embrace the Metropolis-Hastings lifestyle and use a full Bayesian approach.

While running, it will display the current range of likelihoods in the desired range (by default, the best negative log likelihood + 2 negative log likelihood units) and the parameter values falling in that range. If things are working well, the range of values will stabilize during a search

The algorithm tunes: if it is moving too far away from the desired likelihoods, it will decrease the proposal width; if it staying in areas better than the desired likelihood, it will increase the proposal width. It will also expand the proposal width for parameters where the extreme values still appear good enough to try to find out the full range for these values.

In general, the idea of this is not to give you a pleasingly narrow range of possible values -- it is to try to find the actual uncertainty, including finding any ridges that would not be seen in univariate space.

**Table 1**. The main input variables for dent\_walk and a brief explanation for each.

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| par | Starting parameter vector, generally at the optimum. If named, the vector names are used to label output parameters. |
| fn | The likelihood function, assumed to return negative log likelihoods |
| best\_neglnL | The negative log likelihood at the optimum; other values will be greater than this. |
| delta | How far from the optimal negative log likelihood to focus samples |
| nsteps | How many steps to take in the analysis |
| print\_freq | Output progress every print\_freq steps |
| lower\_bound | Minimum parameter values to try. One for all or a vector of the length of par. |
| upper\_bound | Maximum parameter values to try. One for all or a vector of the length of par. |
| adjust\_width\_interval | When to try automatically adjusting proposal widths |
| badval | Bad negative log likelihood to return if a non-finite likelihood is returned |
| sd\_vector | Vector of the standard deviations to use for proposals. Generated automatically if NULL |
| restart\_after | Sometimes the search can get stuck outside the good region but still accept moves. After this many steps without being inside the good region, restart from one of the past good points |
| debug | If TRUE, prints out much more information during a run |
| ... | Other arguments to fn. |