Nested sampling algorithm

From Wikipedia, the free encyclopedia

The **nested sampling algorithm** is a computational approach to the problem of comparing models in Bayesian statistics, developed in 2004 by physicist John Skilling.[1]

Contents [hide]

Background [edit]

Bayes' theorem can be applied to a pair of competing models $M1$ and $M2$ for data D, one of which may be true (though which one is not known) but which both cannot simultaneously be true. The posterior probability for $M1$ may be calculated as follows:

$$\begin{aligned} P(M1|D) &= \frac{P(D|M1)P(M1)}{P(D)} \\ &= \frac{P(D|M1)P(M1)}{P(D|M1)P(M1)+P(D|M2)P(M2)} \\ &= \frac{1}{1+\frac{P(D|M2)}{P(D|M1)}\frac{P(M2)}{P(M1)}} \end{aligned}$$

Given no a priori information in favor of $M1$ or $M2$, it is reasonable to assign prior probabilities $P(M1) = P(M2) = 1/2$, so that $P(M2)/P(M1) = 1$. The remaining Bayes factor $P(D|M2)/P(D|M1)$ is not so easy to evaluate since in general it requires marginalization of nuisance parameters. Generally, $M1$ has a collection of parameters that can be lumped together and called θ, and $M2$ has its own vector of parameters that may be of different dimensionality but is still referred to as θ. The marginalization for $M1$ is

$$P(D|M1) = \int d\theta P(D|\theta, M1)P(\theta|M1)$$

and likewise for $M2$. This integral is often analytically intractable, and in these cases it is necessary to employ a numerical algorithm to find an approximation. The nested sampling algorithm was developed by John Skilling specifically to approximate these marginalization integrals, and it has the added benefit of generating samples from the posterior distribution $P(\theta|D, M1)$.[2] It is an alternative to methods from the Bayesian literature[3] such as bridge sampling and defensive importance sampling.

Here is a simple version of the nested sampling algorithm, followed by a description of how it computes the marginal probability density $Z = P(D|M)$ where M is $M1$ or $M2$:

```
Start with N points θ_1, ..., θ_N sampled from prior.
for i = 1 to j do        % The number of iterations j is chosen by guesswork.
    L_i := min(current likelihood values of the points);
    X_i := exp(-i/N);
    w_i := X_{i-1} - X_i;
    Z := Z + L_i * w_i;
    Save the point with least likelihood as a sample point with weight w_i.
    Update the point with least likelihood with some Markov Chain
```

```
        Monte Carlo steps according to the prior, accepting only steps that
        keep the likelihood above L_i.
    end
    return Z;
```

At each iteration, X_i is an estimate of the amount of prior mass covered by the hypervolume in parameter space of all points with likelihood greater than θ_i. The weight factor w_i is an estimate of the amount of prior mass that lies between two nested hypersurfaces $\{\theta|P(D|\theta,M)=P(D|\theta_{i-1},M)\}$ and $\{\theta|P(D|\theta,M)=P(D|\theta_i,M)\}$. The update step $Z:=Z+L_i*w_i$ computes the sum over i of L_i*w_i to numerically approximate the integral

$$\begin{aligned} P(D|M) &= \int P(D|\theta,M)P(\theta|M)d\theta \\ &= \int P(D|\theta,M)dP(\theta|M) \end{aligned}$$

The idea is to chop up the range of $f(\theta)=P(D|\theta,M)$ and estimate, for each interval $[f(\theta_{i-1}),f(\theta_i)]$, how likely it is a priori that a randomly chosen θ would map to this interval. This can be thought of as a Bayesian's way to numerically implement Lebesgue integration.

Implementations [edit]

- Simple example code written in C, R, or Python demonstrating this algorithm can be downloaded from John Skilling's website
- There is also a Haskell port of the above simple codes on Hackage
- An implementation in R originally designed for fitting of spectra is described at [1] and can be obtained on GitHub at [2]
- A highly modular Python parallel implementation of Nested Sampling for statistical physics and condensed matter physics applications is publicly available from GitHub [3].

Applications [edit]

Since nested sampling was proposed in 2004, it has been used in multiple settings within the field of astronomy. One paper suggested using nested sampling for cosmological model selection and object detection, as it "uniquely combines accuracy, general applicability and computational feasibility."[4] A refinement of the nested sampling algorithm to handle multimodal posteriors has also been suggested as a means of detecting astronomical objects in existing datasets.[5] Other applications of nested sampling is in the field of finite element updating where nested sampling is used to choose an optimal finite element model and this was applied to structural dynamics. [6]

See also [edit]

- Bayesian model comparison

References [edit]

1. ^ Skilling, John (2004). "Nested Sampling". *AIP Conference Proceedings* **735**: 395–405. doi:10.1063/1.1835238.
2. ^ Skilling, John (2006). "Nested Sampling for General Bayesian Computation". *Bayesian Analysis* **1** (4): 833–860. doi:10.1214/06-BA127.
3. ^ Chen, Ming-Hui, Shao, Qi-Man, and Ibrahim, Joseph George (2000). *Monte Carlo methods in Bayesian computation*. Springer. ISBN 978-0-387-98935-8.
4. ^ Mukherjee, P., Parkinson, D., and Liddle, A.R. (2006). "A Nested Sampling Algorithm for Cosmological Model Selection". *Astrophysical Journal* **638** (2): 51–54. arXiv:astro-ph/0508461. Bibcode:2005astro.ph..8461M. doi:10.1086/501068.
5. ^ Feroz, F., Hobson, M.P. (2008). "Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses". *MNRAS* **384** (2): 449–463. arXiv:0704.3704. Bibcode:2008MNRAS.384..449F. doi:10.1111/j.1365-2966.2007.12353.x.
6. ^ Mthembu, L., Marwala, T., Friswell, M.I. and Adhikari, S. 2011. Model selection in finite element model updating using the Bayesian evidence statistic.Mechanical Systems and Signal Processing, 25(7), 2399– 2412.

Categories: Bayesian statistics | Model selection | Statistical algorithms