A Simple Approach to Parallel Nested Sampling

R. Wesley Henderson and Paul M. Goggans

Department of Electrical Engineering
University of Mississippi

July 22, 2015
Clarkson University

Outline

Introduction

- Nested sampling's precision is limited by N.
- Increasing N requires more computation.
- Concurrently-run, independent nested sampling results can be combined to increase the effective value of N.
- Several examples demonstrate this technique's utility.

Outline

Parameter estimation and model selection

Parameter estimation

$$\Pr(\boldsymbol{\Theta}|\boldsymbol{D}, M) = \frac{\Pr(\boldsymbol{D}|\boldsymbol{\Theta}, M)\Pr(\boldsymbol{\Theta}|M)}{\Pr(\boldsymbol{D}|M)}$$

Parameter estimation and model selection

Parameter estimation

$$\Pr(\boldsymbol{\Theta}|\boldsymbol{D}, M) = \frac{\Pr(\boldsymbol{D}|\boldsymbol{\Theta}, M)\Pr(\boldsymbol{\Theta}|M)}{\Pr(\boldsymbol{D}|M)}$$

Abbreviations

$$\Pr(\boldsymbol{\Theta}|\boldsymbol{D}, M) \equiv \mathscr{P}(\boldsymbol{\Theta}) \qquad \Pr(\boldsymbol{D}|\boldsymbol{\Theta}, M) \equiv \mathscr{L}(\boldsymbol{\Theta})$$

$$\Pr(\boldsymbol{\Theta}|M) \equiv \pi(\boldsymbol{\Theta}) \qquad \Pr(\boldsymbol{D}|M) \equiv \mathscr{Z}$$

Parameter estimation and model selection

Parameter estimation

$$\Pr(\boldsymbol{\Theta}|\boldsymbol{D}, M) = \frac{\Pr(\boldsymbol{D}|\boldsymbol{\Theta}, M)\Pr(\boldsymbol{\Theta}|M)}{\Pr(\boldsymbol{D}|M)}$$

Abbreviations

$$\Pr(\boldsymbol{\Theta}|\boldsymbol{D}, M) \equiv \mathscr{P}(\boldsymbol{\Theta}) \qquad \Pr(\boldsymbol{D}|\boldsymbol{\Theta}, M) \equiv \mathscr{L}(\boldsymbol{\Theta})$$
$$\Pr(\boldsymbol{\Theta}|M) \equiv \pi(\boldsymbol{\Theta}) \qquad \Pr(\boldsymbol{D}|M) \equiv \mathscr{Z}$$

Model selection

$$\Pr(M|\boldsymbol{D}) \propto \Pr(\boldsymbol{D}|M)\Pr(M)$$

Model selection

Model posterior ratio

$$\frac{\Pr\left(\boldsymbol{M}_1|\boldsymbol{D}\right)}{\Pr\left(\boldsymbol{M}_2|\boldsymbol{D}\right)} = \frac{\Pr\left(\boldsymbol{D}|\boldsymbol{M}_1\right)}{\Pr\left(\boldsymbol{D}|\boldsymbol{M}_2\right)} \frac{\Pr\left(\boldsymbol{M}_1\right)}{\Pr\left(\boldsymbol{M}_2\right)}$$

Model selection

Model posterior ratio

$$\frac{\Pr\left(M_1|\boldsymbol{D}\right)}{\Pr\left(M_2|\boldsymbol{D}\right)} = \frac{\Pr\left(\boldsymbol{D}|M_1\right)}{\Pr\left(\boldsymbol{D}|M_2\right)}\frac{\Pr\left(M_1\right)}{\Pr\left(M_2\right)}$$

Evidence integral

$$\Pr(\boldsymbol{D}|M) = \int_{\boldsymbol{\Theta}} \Pr(\boldsymbol{D}|\boldsymbol{\Theta}, M)\Pr(\boldsymbol{\Theta}|M)\,\mathrm{d}\boldsymbol{\Theta}$$

THE UNIVERSITY of MISSISSIPPI

Nested sampling

Prior mass

$$X(L) = \int_{\{\Theta : \mathscr{L}(\boldsymbol{\Theta}) > L\}} \pi(\boldsymbol{\Theta})\, \mathrm{d}\boldsymbol{\Theta}$$

Nested sampling

Prior mass

$$X(L) = \int_{\{\Theta : \mathscr{L}(\boldsymbol{\Theta}) > L\}} \pi(\boldsymbol{\Theta})\, \mathrm{d}\boldsymbol{\Theta}$$

Alternate evidence integral

$$\mathscr{Z} = \int_0^1 L(X)\, \mathrm{d}X$$

Estimating the prior mass

For a set of parameters $\boldsymbol{\Theta}$, the likelihood can be computed exactly. Prior mass, however, generally must be estimated.

Estimating the prior mass

For a set of parameters $\boldsymbol{\Theta}$, the likelihood can be computed exactly. Prior mass, however, generally must be estimated.

Shrinkage

$$t_i = \frac{X_i}{X_{i-1}}$$

Estimating the prior mass

For a set of parameters $\boldsymbol{\Theta}$, the likelihood can be computed exactly. Prior mass, however, generally must be estimated.

Shrinkage

$$t_i = \frac{X_i}{X_{i-1}}$$

Shrinkage distribution

$$t_i \sim \text{Beta}(N, 1)$$

Using the prior mass estimate

$$X_i = \prod_{k=1}^{i} t_k$$

Using the prior mass estimate

$$X_i = \prod_{k=1}^{i} t_k$$

Estimating evidence with quadrature

$$Z \approx \sum_{i=1}^{m} \left(X_{i-1} - X_i\right) L_i$$

Using the prior mass estimate

$$X_i = \prod_{k=1}^{i} t_k$$

Estimating evidence with quadrature

$$Z \approx \sum_{i=1}^{m} \left(X_{i-1} - X_i \right) L_i$$

Uncertainty in evidence estimate

$$\exp\left(\pm\sqrt{H/N} \right)$$

Using the prior mass estimate

$$X_i = \prod_{k=1}^{i} t_k$$

Estimating evidence with quadrature

$$Z \approx \sum_{i=1}^{m} \left(X_{i-1} - X_i\right) L_i$$

Uncertainty in evidence estimate

$$\exp\left(\pm\sqrt{H/N}\right)$$

Larger N gives a smaller uncertainty in the evidence estimate. Raising N requires more computation time. Unless we can combine samples from independent runs!

Outline

THE UNIVERSITY of MISSISSIPPI

Combining independent chains – general case

General case: M ordered sets of discarded samples with associated prior mass X

$$\boldsymbol{X}^1 = \left\{ X_1^1, X_2^1, \cdots, X_{Q_1}^1 \right\}$$
$$\boldsymbol{X}^2 = \left\{ X_1^2, X_2^2, \cdots, X_{Q_2}^2 \right\}$$
$$\vdots$$
$$\boldsymbol{X}^M = \left\{ X_1^M, X_2^M, \cdots, X_{Q_M}^M \right\}$$

Combining independent chains – simplest case

Simplest case has two sets of one discarded sample each.

$$\boldsymbol{X}^1 = \left\{X_1^1\right\}$$
$$\boldsymbol{X}^2 = \left\{X_1^2\right\}$$

Combined set is

$$\hat{\boldsymbol{X}}^{1,2} = \left\{\hat{X}_1^{1,2}, \hat{X}_2^{1,2}\right\}$$

What is the distribution of the shrinkage from the larger member to the smaller member of $\hat{\boldsymbol{X}}^{1,2}$?

Combining independent chains – shrinkage

The shrinkage t is defined

$$t_i = \frac{X_i}{X_{i-1}}$$

For the first sample,

$$t_1 = \frac{X_1}{1} = X_1$$

In our example,

$$\hat{X}_1^{1,2} = \hat{t}_1^{1,2} \sim \mathrm{Beta}(N, 1)$$

Combining independent chains – highest order statistic

For a set with n members i.i.d. as $f(x)$ with CDFs $F(x)$, the density of the ith order statistic is

$$f_{\hat{X}_{(i)}}(x) = \frac{n!}{(n-i)!\,(i-1)!} f(x)[F(x)]^{i-1}[1-F(x)]^{n-i}$$

In our example,

$$f_{\hat{X}_{(2)}}(x) = 2f(x)[1-F(x)]$$

$$f_{t_i}(x) = Nx^{N-1} \qquad F_{t_i}(x) = x^N$$

THE UNIVERSITY of MISSISSIPPI

Combining independent chains – highest order statistic

For a set with n members i.i.d. as $f(x)$ with CDFs $F(x)$, the density of the ith order statistic is

$$f_{\hat{X}_{(i)}}(x) = \frac{n!}{(n-i)!\,(i-1)!} f(x)[F(x)]^{i-1}[1-F(x)]^{n-i}$$

In our example,

$$f_{\hat{X}_{(2)}}(x) = 2f(x)[1-F(x)]$$

$$f_{t_i}(x) = Nx^{N-1} \qquad F_{t_i}(x) = x^N$$

$$\boxed{f_{X_{(2)}}(x) = 2Nx^{2N-1}}$$

Generalization

- Show empirically that, in general, shrinkage in combined set is distributed as $\mathrm{Beta}(N \times M, 1)$
- Procedure
  - Generate 32 sets of 10,000 shrinkage samples from $Beta(100, 1)$.
  - Get prior mass values from cumulative product of shrinkage samples in each set.
  - Combine sets of prior mass samples, then sort by prior mass.
  - Compute actual shrinkage between each consecutive pair of samples.
  - Compare sample mean of $\log t$ with $\frac{1}{N}$ and $\frac{1}{N \times M}$.

Generalization results

- $\frac{1}{N} = 1/100 = 0.01$
- $\frac{1}{N \times M} = 1/3200 = 3.125 \times 10^{-4}$
- Log geometric mean of combined shrinkage samples: $3.200 \times 10^{-4} \pm 4.146 \times 10^{-5}$
- Relative error with $\frac{1}{N}$: 96.80%
- Relative error with $\frac{1}{N \times M}$: 2.393%

Implementation

Many ways to implement:

- Concurrent independent nested sampling runs using multiple supercomputer nodes, processor cores, GPU cores, etc.
- Concurrent or non-concurrent independent NS runs by different users, later combined to improve precision
- If original evidence estimate is not precise enough, subsequent NS runs can be used to improve precision

Advantages

- Runs are independent (i.e., no communication is required), so the speed-up is nearly ideal.
- Posterior distributions with multiple high-probability modes can be sampled effectively.
- Precision can be improved with subsequent runs.

Outline

Examples

- Eggcrate likelihood function
- Modified lighthouse problem
- Sum of sinusoidal signals in noise

Nested sampling implementation details

- Evidence is estimated by sampling shrinkage distribution instead of using geometric mean
- Log-evidence error bars are the standard deviations of the log-evidence samples

Eggcrate

Prior

$$\pi(\boldsymbol{\Theta}) = \left(\frac{1}{10\pi}\right)^2 \mathbb{1}_{[0,10\pi]}\left(\Theta_1\right) \mathbb{1}_{[0,10\pi]}\left(\Theta_2\right)$$

Eggcrate

Prior

$$\pi(\boldsymbol{\Theta}) = \left(\frac{1}{10\pi}\right)^2 \mathbb{1}_{[0,10\pi]}\left(\Theta_1\right) \mathbb{1}_{[0,10\pi]}\left(\Theta_2\right)$$

Likelihood

$$\mathscr{L}(\boldsymbol{\Theta}) = \exp\left\{\left[2 + \cos\left(\frac{\Theta_1}{2}\right)\cos\left(\frac{\Theta_2}{2}\right)\right]^5\right\}$$

Eggcrate

Prior

$$\pi(\boldsymbol{\Theta}) = \left(\frac{1}{10\pi}\right)^2 \mathbb{1}_{[0,10\pi]}\left(\Theta_1\right) \mathbb{1}_{[0,10\pi]}\left(\Theta_2\right)$$

Likelihood

$$\mathscr{L}(\boldsymbol{\Theta}) = \exp\left\{\left[2 + \cos\left(\frac{\Theta_1}{2}\right)\cos\left(\frac{\Theta_2}{2}\right)\right]^5\right\}$$

Posterior

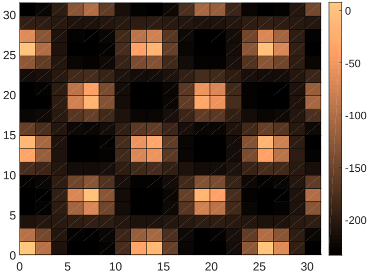
$$\mathscr{P}(\boldsymbol{\Theta}) = \frac{\left(\frac{1}{10\pi}\right)^2 \exp\left\{\left[2+\cos\left(\frac{\Theta_1}{2}\right)\cos\left(\frac{\Theta_2}{2}\right)\right]^5\right\} \mathbb{1}_{[0,10\pi]}(\Theta_1)\mathbb{1}_{[0,10\pi]}(\Theta_2)}{235.88}$$

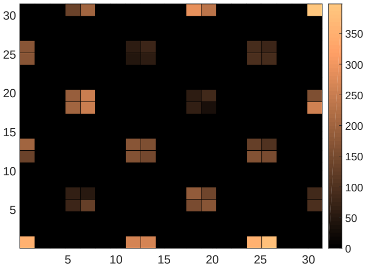
THE UNIVERSITY of MISSISSIPPI

Eggcrate parameters

- $N = 20$
- $M = 32$

Eggcrate results: log-posterior and histogram

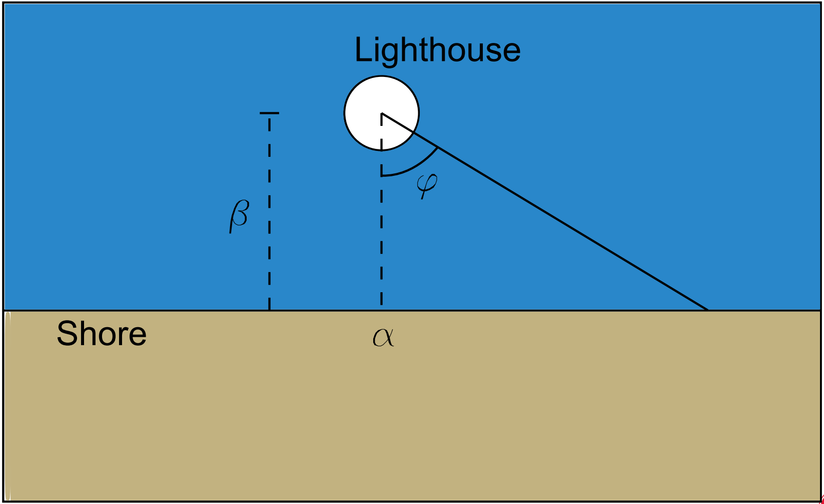




$$\log \mathscr{Z} = 235.1 \pm 1.336$$

Compare the above value with the value from Feroz, et al., 2009 computed using standard quadrature: 235.88.

Lighthouse problem



Lighthouse problem

Prior

$$\Pr\left(\alpha_j\right) = (1/200)\mathbb{1}_{[-100,100]}\left(\alpha_j\right)$$
$$\Pr\left(\beta_j\right) = (1/100)\mathbb{1}_{[0,100]}\left(\beta_j\right)$$

Lighthouse problem

Prior

$$\Pr\left(\alpha_j\right) = (1/200)\mathbb{1}_{[-100,100]}\left(\alpha_j\right)$$
$$\Pr\left(\beta_j\right) = (1/100)\mathbb{1}_{[0,100]}\left(\beta_j\right)$$

Original likelihood

$$\mathscr{L}(\boldsymbol{\Theta}) = \Pr\left(x_k|\alpha_j, \beta_j\right) = \frac{\beta_j}{\pi\left(\beta_j^2 + \left(x_k - \alpha_j\right)^2\right)}$$

Lighthouse problem extension

New likelihood for one observation

$$\mathscr{L}(\boldsymbol{\Theta}) = \Pr\left(x_k | \boldsymbol{\alpha}, \boldsymbol{\beta}\right) = \sum_{j=1}^{J} A_j \frac{\beta_j}{\pi\left(\beta_j^2 + \left(x_k - \alpha_j\right)^2\right)}$$

Lighthouse problem extension

New likelihood for one observation

$$\mathscr{L}(\boldsymbol{\Theta}) = \Pr\left(x_k|\boldsymbol{\alpha},\boldsymbol{\beta}\right) = \sum_{j=1}^{J} A_j \frac{\beta_j}{\pi\left(\beta_j^2 + \left(x_k - \alpha_j\right)^2\right)}$$

New likelihood for multiple observations

$$\mathscr{L}(\boldsymbol{\Theta}) = \Pr\left(\boldsymbol{x}|\boldsymbol{\alpha},\boldsymbol{\beta}\right) = \prod_{k=1}^{K}\sum_{j=1}^{J} A_j \frac{\beta_j}{\pi\left(\beta_j^2 + \left(x_k - \alpha_j\right)^2\right)}$$

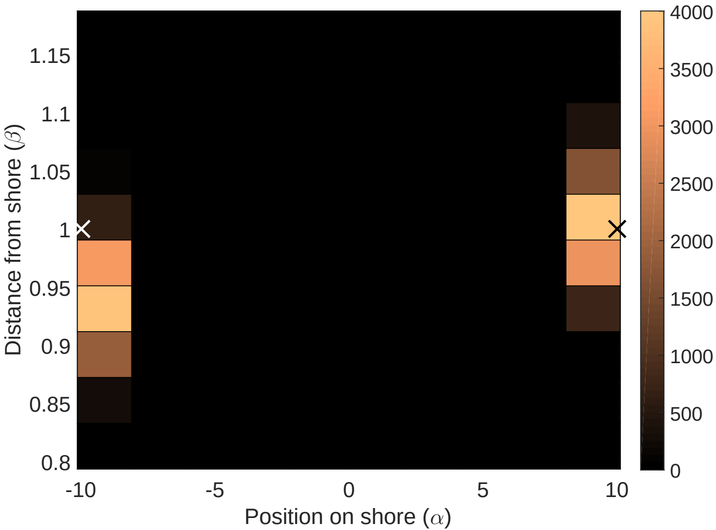
Lighthouse problem parameters

- 1000 observations of 2 lighthouses
- $N = 40$
- $M = 4$

Table: Lighthouse parameters

j	A_j	α_j	β_j
1	0.5	−10.0	1
2	0.5	10.0	1

Lighthouse results



Lighthouse results

Table: Lighthouse problem results

J	$\langle \log \mathscr{Z} \rangle$	$\mathrm{Stdev}(\log \mathscr{Z})$
1	−12680	1.160
2	−9254	1.107
3	−9260	1.445
4	−9300	1.377

Sinusoidal signals in noise

Signal model: sum of sinusoidal signals corrupted by additive white Gaussian noise (AWGN)

$$s(t) = \left[\sum_{j=1}^{J} A_j \cos\left(\omega_j t\right) + B_j \sin\left(\omega_j t\right) \right]$$

$$d(t) = s(t) + n(t)$$

Sinusoidal signals in noise

Joint prior

$$\pi(\boldsymbol{\Theta}) = \left(\frac{1}{(20)(20)(512\pi)} \right)^J \prod_{j=1}^{J} \mathbb{1}_{[-10,10]}\left(A_j\right) \mathbb{1}_{[-10,10]}\left(B_j\right) \mathbb{1}_{[0,512\pi]}\left(\omega_j\right)$$

Sinusoidal signals in noise

Joint prior

$$\pi(\boldsymbol{\Theta}) = \left(\frac{1}{(20)(20)(512\pi)}\right)^J \prod_{j=1}^{J} \mathbb{1}_{[-10,10]}\left(A_j\right) \mathbb{1}_{[-10,10]}\left(B_j\right) \mathbb{1}_{[0,512\pi]}\left(\omega_j\right)$$

For observed time series $d\left(t_k\right)$ such that $1 \leq k \leq K$,

Likelihood

$$\mathscr{L}(\boldsymbol{\Theta}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^K \exp\left\{-\left[\sum_{k=1}^{K}\left(s\left(t_k\right) - d\left(t_k\right)\right)^2\right] / \left(2\sigma^2\right)\right\}$$

Sinusoidal signals in noise

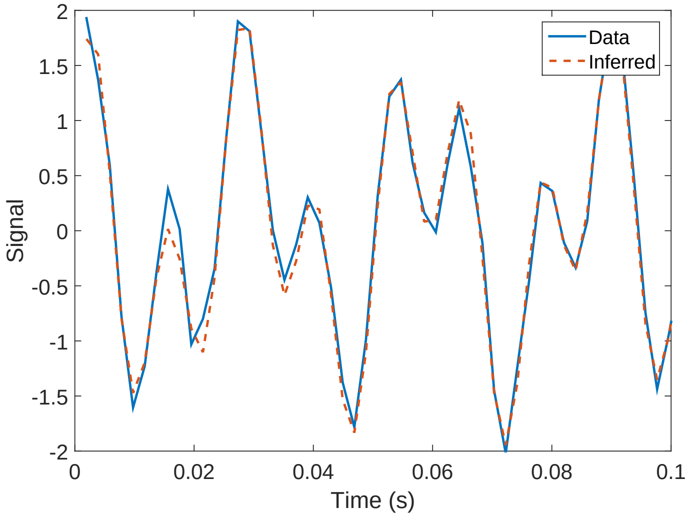
Test signal parameters:

- $J = 2$
- $f_s = 512$ samples/s
- $K = 1000$
- $\sigma^2 = 0.01$
- $N = 25$
- $M = 4$

Table: Sinusoidal signal parameters

j	A_j	B_j	ω_j
1	1.0	0.0	68π
2	0.0	1.0	160π

Sinusoidal signals in noise – truncated signal plot



Sinusoidal signals in noise – results

Table: Signal problem results

J	$\langle \log \mathcal{Z} \rangle$	Stdev($\log \mathcal{Z}$)
1	−2546	1.174
2	−553.7	1.349
3	−583.5	1.568

Outline

Conclusion

- Combining the results of independent nested sampling runs decreases the shrinkage between consecutive samples.
- This is demonstrated using an analytical example for the simplest case and a numerical test for a more general case.
- This technique is effective for determining the evidence in several example problems, including for distributions with several prominent modes.
- Simple, effective, extensible.