Diffusive Nested Ensemble Sampling

ABSTRACT

We proposed an affine invariant ensemble version of the Diffusive Nested Sampling Method and did some tests on it. (still under development)

Subject headings: methods: data analysis — methods: numerical — methods: statistical — bayesian

1. Introduction

In bayesian decision theory, one should return the best mixture of competing models instead of simply deciding which model is the best. To achieve this, the evaluation of evidence integral Z is required. Bayesian decision theory can be summarized as

$$P(\mathrm{Model}_j|\mathrm{Data}) = \frac{P(\mathrm{Data}|\mathrm{Model}_j)P(\mathrm{Model}_j)}{\sum_j P(\mathrm{Data}|\mathrm{Model}_j)P(\mathrm{Model}_j)}, \tag{1}$$

which is just the Bayes law. The *likelihood* in Eqn. (1) is in fact the evidence Z_j for model j.

$$Z_j = P(\mathrm{Data}|\mathrm{Model}_j) = \int P(\mathrm{Data}|\theta, \mathrm{Model}_j)P(\theta|\mathrm{Model}_j)\mathrm{d}\theta, \tag{2}$$

where $P(\mathrm{Data}|\theta, \mathrm{Model}_j)$ is the likelihood of parameter θ for model j and $P(\theta|\mathrm{Model}_j)$ is the prior of parameter θ for model j. To make notations simple, we will use $L(\theta)$ for the likelihood and $\pi(\theta)$ for the prior. We'll also drop the index j in Z_j because the discussion applies to all models. So the evidence can be written as

$$Z = \int L(\theta)\pi(\theta)\mathrm{d}\theta. \tag{3}$$

The prior $\pi(\theta)$ is normalized in the parameter space, that is

$$\int \pi(\theta)\mathrm{d}\theta = 1, \tag{4}$$

while the likelihood $L(\theta)$ is not.

However, evaluating or even estimating the evidence integral has always been challenging. Diffusive Nested Sampling proves to be an efficient and accurate method to evaluate or estimate the evidence (Brewer *et al.* 2011). We hope to take advantage of affine invariant ensemble sampler (Goodman *et al.* 2010) to make diffusive nested sampling even more efficient.

2. Diffusive Nested Sampling

In nested sampling, we change the variable in the evidence integral from parameter θ to the prior mass

$$M(L^*) = \int_{L(\theta)>L^*} \pi(\theta)\mathrm{d}\theta, \tag{5}$$

which is, in another word, cumulant prior mass covering the area whose likelihood values are greater than L^* (Skilling 2006). M is a monotonically decreasing function of L^* and it ranges from 0 to 1. And the mapping between M and L^* is a bijection. An infinitesimal increment of M is

$$\mathrm{d}M = \int_{L^*-\mathrm{d}L^*<L(\theta)<L^*} \pi(\theta)\mathrm{d}\theta = \pi(\theta) \times \text{volume of } \theta, \text{ satisfying } L^* - \mathrm{d}L^* < L(\theta) < L^*. \tag{6}$$

Multiply both sides with L^* and integrate. It is easy to see that

$$\int_0^1 L^*\mathrm{d}M = \int L(\theta)\pi(\theta)\mathrm{d}\theta. \tag{7}$$

so evidence Z can be expressed as

$$Z = \int_0^1 L^*(M)\mathrm{d}M. \tag{8}$$

So the integral Z is the area below the $L^*(M)$ curve. In most cases, it is impossible to know the function $L^*(M)$ analytically and evaluate the integral analytically. Note from Eqn. (8) that M can be viewed as a random variable with uniform distribution. Nested sampling takes advantage of this and proposes to build the $L^*(M)$ curve statistically.

2.1. Level and Constrained Prior

In nested sampling, we first try to find several points on the $L^*(M)$ curve, $\{(M_0, L_0^*), (M_1, L_1^*), \ldots\}$. We call these points levels. The L^* is called a level's likelihood threshold or just threshold. Each level defines a constrained prior,

$$p_{L_j^*}(\theta) = \frac{\pi(\theta)}{M_j}\mathbb{1}_{L(\theta)>L_j^*}, \tag{9}$$

where

$$\mathbb{1}_{L(\theta)>L_j^*} = \begin{array}{l} 1,\ L(\theta) > L_j^*, \\ 0,\ \text{otherwise}, \end{array}$$

and note that $p_{L_j^*}$ is properly normalized by M_j.

2.2. Setting Level Thresholds

We already know the right-most point on the curve, ($M_0 = 1$, $L_0^* = 0$), which is level 0, because there is no restriction on the likelihood function with $L_0^* = 0$ and from the definition of prior mass, Eqn. (5), $M(\theta)$ should cover the whole parameter space. And most likely another point on the left-most side of the curve, ($M_{max} = 0$, L_{max}), ~~because we must have found the optimal likelihood before we start to consider evaluating the evidence integral~~. I don't think you need to know L_{max} before constructing levels.

To find ~~a new level~~the first level L_1^*, we generate N samples from the prior density $\pi(\theta)$, θ_1, ..., θ_N. We calculate the likelihoods $L(\theta_k)$. We choose L_1^* so that the number of k with $L(\theta_k) > L_1^*$ is N/e. ~~We then rank the likelihoods in the chain in descending order and find the N/e-th likelihood, which we call L_1^*. This would give us a new point on the $L^*(M)$ curve, which is approximately $(1/e, L_1^*)$. This is level 1 and L_1^* is the likelihood threshold of level 1.~~ A better way to do this is with the algorithm called **quick find**. In the STL it is described here: /http://www.cplusplus.com/reference/algorithm/nth_element/. $M_1 \approx 1/e$ is the prior mass that level 1 covers (the prior mass of the parameters whose likelihoods are larger than L_1^*). M_1 is a random variable and its expectation and variance will be given below.

~~To find the next point, we generate N samples from the constrained prior density $p_{L_1^*}(\theta)$ defined in Eqn. (9). But in most cases, we~~ One way to estimate the next level L_2^* would be to generate N samples from the constrained prior density $p_{L_1^*}(\theta)$ defined in Eqn. (9). But instead (for reasons explained below??) we actually generate samples from a mixture of $p_{L_1^*}(\theta)$ and prior $\pi(\theta)$ until we have N samples with likelihood larger than the previous level's threshold L_1^*. We sample the mixture because the area covered by level 1 may be disconnected in paremeter space and only sampling the constrained prior $p_{L_1^*}(\theta)$ may get us stucked in only one or few of those disconnected areas. Like before, we get a chain of N likelihoods, rank these likelihoods in descending order and find the N/e-th likelihood, which we call L_2^*. This would give us another point on the $L^*(M)$ curve, which is approximately $(1/e^2, L_2^*)$. This is level 2 and L_2^* is the likelihood threshold of level 2. Again, $M_2 \approx 1/e^2$ is the prior mass that level 2 covers. Like M_1, M_2 is also a random variable. What are the mixture proportions? It is possible to make error bars for M_1 and M_2 even when the θ_k are from MCMC. It follows the error bar method for order statistics.

There is a simple stopping criterion. Suppose we have constructed levels $L_1^*, \ldots, L_k^*$ and that M_j is the probability for L_k^*. We took $M_j = e^{-j}$ above. The evidence integral corresponding to these levels is

$$Z_k = \sum_{j=0}^{k-1} \int_{M_j}^{M_{j+1}} L(M) dM$$

The remaining integral is

$$\int_{M_k}^{1} L(M) dM \ .$$

But $L(M) < L_{max}$ always, so the remain integral cannot be larger than $L_{max}(1 - M_k)$. We choose

a stopping point k so that $L_{max}(1 - M_k) \leq \epsilon Z_k$, with $\epsilon =$?? We keep going until new ~~levels'~~*level's* contribution to the integral is small compared to our requirement of precision. At the end, we get a series of points on the $L^*(M)$ curve, $\{(M_0, L_0^*), (M_1, L_1^*), (M_2, L_2^*), \ldots\}$. Correspondingly, we also have a series of mixture of constrained priors,

$$p(\theta) = \sum_{j=0} w_j p_{L_j^*}, \tag{10}$$

where $p_{L_j^*}$ is the constrained prior defined by level j,

$$p_{L_j^*}(\theta) = \frac{\pi(\theta)}{M_j} \mathbb{1}_{L(\theta)>L_j^*}, \quad j = 0, 1, 2, \ldots. \tag{11}$$

and w_j are the weights of each level which sum up to 1,

$$\sum_{j=0} w_j = 1. \tag{12}$$

The choice of weight may change according to different purposes. For example, when we are building a new level, we might want to put more weight on the last level. And in the final stage, when we sample all the levels together, we might want to have the same weight on all the levels.

2.3. Affine Invariant Stretch Move

To implement the affine invariant ensemble sampler to diffusive nested sampling, we assign different levels to different walkers in the ensemble. This probably would cause concern since all the walkers would be from different density and this might render stretch move invalid. But in the stretch move, the helper walker does not interfere with the walker which it helps, even if they are from different density distribution.

Our goal is to sample the mixture of the constrained priors Eqn. (10). The way we realize it is by updating the walkers according to their own constrained priors Eqn. (9) and updating the ~~indeces~~indices of the levels of those walkers according to their weights and other restrictions. So the algorithm consists of two part: updating the ensemble of walkers followed or preceded by updating all the level indeces of those walkers. The first part can be summarized as: probably $X_{new} = \cdots$, and $\alpha = z$. I don't think the metropolis part is stated correctly.

- randomly choose a helping walker Y
- propose a new walker with stretch move: $X_{new} \to Y + \alpha(X_{old} - Y)$, where α is a random variable from some distribution (Goodman *et al.* 2010).
- if the proposed X_{new} has a likelihood smaller than its current threshold L^*, reject the proposal.
- else, accept the proposed X_{new} with probability: $\max\left(z^{\dim-1}\frac{p_{L^*}(X_{new})}{p_{L^*}(X_{old})}, 1\right)$.

The second part can be summarized as (Brewer *et al.* 2011):

- propose a new level for a walker in the ensemble, with proposal probability, Eqn. (13): $i \to j$
- if $j \geq i$, accept the proposal if, in parameter space, the likelihood the walker is larger than L_j^*; reject if it is smaller.
- else, accept the proposal with probability $\frac{M_i}{M_j}\frac{w_j}{w_i}$. You have to be sure that the w_j are chosen so that these are ≤ 1.

The proposal probabilities for the second part, $T_{i\to j}$, can be written as entries of matrix T,

$$(T_{i\to j}) = (T_{ij}) \begin{pmatrix} 0.5 & 0.5 & & & & & & \\ 0.5 & 0 & 0.5 & & & & & \\ & 0.5 & 0 & 0.5 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & 0.5 & 0 & 0.5 & & \\ & & & & 0.5 & 0 & 0.5 & \\ & & & & & 0.5 & 0.5 & \end{pmatrix} \tag{13}$$

2.4. Refining Level Masses

Suppose we have J new levels in total, $\{(M_0, L_0^*), (M_1, L_1^*), (M_2, L_2^*), \ldots, (M_J, L_J^*)\}$, where (M_0, L_0^*) is not counted as a new level. However, the prior masses $\{M_0, M_1, \ldots, M_J\}$ are only approximations of the true prior masses $\{M_0^*, M_1^*, \ldots, M_J^*\}$. As a result, the normalization of constrained prior Eqn. (9) is also approximate. So when we sample the mixture of the constrained priors Eqn. (10), we actually sample the following distribution

$$p^*(\theta) = C \sum_{j=0}^{J} w_j \frac{M_j^*}{M_j} \frac{\pi(\theta)}{M_j^*} \mathbb{1}_{L(\theta)>L_j^*}, \tag{14}$$

where C is a normalization term and of course $M_0^* = M_0 = 1$. Let there be an extra level whose likelihood threshold L_{J+1}^* is the maximum likelihood and prior mass M_{J+1} is 0. Define the set of parameters whose likelihoods are sandwiched by level j and level $j+1$ as

$$B_j = \{\theta | L(\theta) \in [L_j^*, L_{j+1}^*]\}, \;\; j = 0, 1, 2, \ldots, J. \tag{15}$$

The theoretical probability that a samples is between level 0 and level 1 is (r is used for ratio.)

$$r_0^* = \text{Prob}(B_0) = C w_0 \frac{M_0^*}{M_0} \left(1 - \frac{M_1^*}{M_0^*}\right) = C \frac{w_0}{M_0}(M_0^* - M_1^*). \tag{16}$$

The theoretical probability that a samples is between level 1 and level 2 is Inconsistent numbering. 1 comes after 0, not 2. 1 comes before 2, not 0.

$$r_2^* = \text{Prob}(B_1) = Cw_0\frac{M_0^*}{M_0}\left(\frac{M_1^*}{M_0^*} - \frac{M_2^*}{M_0^*}\right) + Cw_1\frac{M_1^*}{M_1}\left(1 - \frac{M_2^*}{M_1^*}\right) = C\left(\frac{w_0}{M_0} + \frac{w_1}{M_1}\right)(M_1^* - M_2^*). \tag{17}$$

Similarly, the theoretical probability that a samples is between level j and level $j+1$ is

$$r_j^* = \text{Prob}(B_j) = C\left(\frac{w_0}{M_0} + \frac{w_1}{M_1} + \ldots + \frac{w_j}{M_j}\right)(M_j^* - M_{j+1}^*). \tag{18}$$

From the mixture of the constrained priors Eqn. (10), we get a likelihood chain of length n. Let the number of likelihood samples sandwiched between level j and level $j+1$ be n_j. And the actual percentage of samples between adjacent levels is $\{r_0, r_1, r_2, \ldots, r_J\}$ which is just the corresponding n_j devided by n. Matching these actual percentages with the theoretical ones r^*'s, we get a total of $J+1$ linear equations with $J+1$ unknowns $\{\frac{1}{C}, M_1^*, M_2^*, \ldots, M_J^*\}$,

$$AM^* = b, \tag{19}$$

where matrix A is

$$A = \begin{pmatrix} -r_0 & -\frac{w_0}{M_0} & 0 & 0 & \ldots & 0 \\ -r_1 & \left(\frac{w_0}{M_0} + \frac{w_1}{M_1}\right) & -\left(\frac{w_0}{M_0} + \frac{w_1}{M_1}\right) & 0 & \ldots & 0 \\ -r_2 & 0 & \left(\frac{w_0}{M_0} + \frac{w_1}{M_1} + \frac{w_2}{M_2}\right) & -\left(\frac{w_0}{M_0} + \frac{w_1}{M_1} + \frac{w_2}{M_2}\right) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ -r_J & 0 & 0 & 0 & 0 & \left(\frac{w_0}{M_0} + \frac{w_1}{M_1} + \cdots + \frac{w_J}{M_J}\right) \end{pmatrix} \tag{20}$$

and M^* is the vector of unknowns,

$$M^* = \begin{pmatrix} \frac{1}{C} & M_1^* & M_2^* & M_3^* & \ldots & M_J^* \end{pmatrix}^T, \tag{21}$$

and vector b is

$$b = \begin{pmatrix} -\frac{w_0}{M_0} & 0 & 0 & 0 & \ldots & 0 \end{pmatrix}^T. \tag{22}$$

We solve Eqn. (19) to get the refined prior masses M^*'s.

2.5. Computing Evidence

We take the mean of likelihoods sandwiched between two levels,

$$\bar{L}_j = \frac{1}{n_j}\sum_{L_j^* \le L(\theta) < L_{j+1}^*} L(\theta), \tag{23}$$

where θ represents samples. Again with the extra level whose likelihood threshold L^*_{J+1} is the maximum likelihood and prior mass M_{J+1} is 0, the evidence is

$$Z = \sum_{j=0}^{J} \bar{L}_j (M^*_j - M^*_{j+1}). \tag{24}$$

3. 2-d Gaussian Testing Case

The algorithm was tested on a 2-d gaussian likelihood and a 2-d uniform prior. The likelihood is

$$L(\theta_1, \theta_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\theta_1^2 + \theta_2^2}{2\sigma^2}\right),\ \sigma = 1, \tag{25}$$

where θ_1 and θ_2 are the parameters. The prior is

$$\pi(\theta_1, \theta_2) = \frac{1}{400},\ \theta_1 \in [-10,\ 10],\ \theta_2 \in [-10,\ 10], \tag{26}$$

and 0 otherwise. This prior basically is a square whose sides' length is 20 and whose area is 400. The evidence is approximately inverse of that area,

$$\text{evidence} \approx \frac{1}{400}, \tag{27}$$

where the approximation is equivalent to equality up to machine error because gaussian distribution has extremely thin tail. The likelihood threshold of any level can be analytically calculated in this model. As a matter of fact, the whole $L^*(M)$ curve can be built analytically,

$$\log L^*(M) = -\log 2\pi - \frac{200M}{\pi}, \tag{28}$$

where the number 200 comes from half the area that the prior covers. Note that $\log L^*$ is a linear function of M.

3.1. Testing Level Thresholds Setting

The test is to see if the algorithm can build levels matching the analytically calculated ones. Recall that to find a new level, one needs to generate a chain of N_1 likelihoods larger than previous level's threshold. (N_1 is used so not to be confused with N_2 used later.) Each new level requires N_1 likelihoods. Two N_1's are tested, $N_{1a} = 10,000$ and $N_{1b} = 100,000$. The larger N_{1b} should give a smaller variance than N_{1a}. For both N_{1a} and N_{1b}, 6 levels are built for $10,000$ times in order to check the statistical features of these levels.

For $N_{1a} = 10,000$, after ranking the chain of N_{1a} likelihoods in descending order, the $J_{1a} = 3,678$-th likelihood is picked as the new level's threshold. For $N_{1b} = 100,000$, after ranking the chain of N_{1b} likelihoods. the $J_{1b} = 36,787$-th likelihood is picked as the next level's threshold.

For N_{1a}, the expectation of the prior masses that each level covers are $\left\{\frac{J_{1a}}{N_{1a}+1}, \left(\frac{J_{1a}}{N_{1a}+1}\right)^2, \ldots\right\}$. And for N_{1b}, the expectation of the prior masses that each level covers are $\left\{\frac{J_{1b}}{N_{1b}+1}, \left(\frac{J_{1b}}{N_{1b}+1}\right)^2, \ldots\right\}$. The variance of the prior masses can also be easily calculated. For example, the variance of the 1st level's covered mass is $\frac{(J_{1a})(N_{1a}-J_{1a})}{(N_{1a}+1)^2(N_{1a}+2)}$ for N_{1a} and $\frac{(J_{1b})(N_{1b}-J_{1b})}{(N_{1b}+1)^2(N_{1b}+2)}$ for N_{1b}. The expection and variance of the corresponding (logarithm of) likelihood thresholds can then be calculated straightfowardly, because $\log L^*$ is a linear fuction of M (from Eqn. (28)). The mean values of the 6 levels' thresholds are listed in Tab. (1) for both N_{1a} and N_{1b} together with the corresponding analytical values of the thresholds. The standard deviations with their analytical values are listed in Tab. (2). The histograms of level1 and level 6 are visualized in Fig. (1)

3.2. Testing Constrained Prior Mixture

We draw samples from a mixture of all the constrained priors defined by true levels listed in Tab. (3). Every adjacent two levels define a bin. For each sample, we find two adjacent levels whose thresholds sandwich the likelihood of the sample and that sample can be put into the bin defined by those two levels. The prior masses of samples inside each bin should follow uniform distribution, which is consistent with our testing result, illustrated in Fig. (2). The variances of the likelihoods of samples can vary dramatically among different bins, Fig. (3).

4. High-Dimension Gaussian Testing Case

In 10-d case, the likelihood is

$$L(\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{10/2}} \exp\left(-\frac{\boldsymbol{\theta}^2}{2\sigma^2}\right),\ \sigma = 1, \tag{29}$$

where θ_1 and θ_2 are the parameters. The prior is

$$\pi(\boldsymbol{\theta}) = \frac{1}{20^{10}},\ \theta_j \in [-10,\ 10],\ j = 1, 2, \ldots, 10, \tag{30}$$

and 0 otherwise. We make 30 levels in 10-d case so that the last level will cover approximately $1/20^{10}$ of the total prior hyper-volume. To build each level, a likelihood chain of length N_1 is generated. After all the levels are built, we sample mixture of constrained priors for N_2 times, using these samples to refine the prior masses and evaluate the evidence.

4.1. Testing Prior Mass Refinement

We repeat the prior mass refinement and evidence evaluation for $1,000$ times for different N_1's and N_2's. The mean evidences for all N_1's and N_2's are very close to the true evidence 9.77×10^{-14}.

The standard deviations are summarized in Tab. (4).

REFERENCES

Goodman, J., Weare, J., 2010, Comm. App. Math. and Comp. Sci., 5, 65

Brewer, B. J., Pártay, L. B. & Csányi, G, 2011, Statistics and Computing, 21, 649

Skilling, J., 2006, Bayesian Analysis, 4, 833

	$N_{1a} = 10,000$		$N_{1b} = 100,000$	
Level	experimental mean	analytical value	experimental mean	analytical value
1	-25.2525	-25.2504	-25.2569	-25.2570
2	-10.4490	-10.4481	-10.4530	-10.4530
3	-5.00537	-5.00442	-5.00707	-5.00708
4	-3.00304	-3.00241	-3.00382	-3.00372
5	-2.26627	-2.26615	-2.26680	-2.26675
6	-1.99543	-1.99538	-1.99567	-1.99565

Table 1: The experiment was repeated $10,000$ times. The experimental mean values of the logarithm of the 6 levels' thresholds in the table are the mean of the $10,000$ repetitions. Notice that the experimental means for N_{1b} tend to be closer to analytical values than those for N_{1a} as expected.

	$N_{1a} = 10,000$		$N_{1b} = 100,000$	
Level	experimental std	analytical std	experimental std	analytical std
1	0.36	0.31	0.11	0.097
2	0.18	0.16	0.057	0.051
3	0.081	0.072	0.026	0.023
4	0.034	0.031	0.011	0.0097
5	0.014	0.013	0.0044	0.0040
6	0.0057	0.0051	0.0018	0.0016

Table 2: The experiment was repeated $10,000$ times. The experimental std of the logarithm of the 6 levels' thresholds in the table are the variance of the $10,000$ repetitions. The algorithm almost achieves the expected precision. Note that the std's for N_{1a} are approximately $\sqrt{10}$ times those for N_{1b}.

level index	log likelihood threshold	log prior mass
1	-29.8629798621251	-1
2	-15.0587589731369	-2
3	-9.61259046551731	-3
4	-7.60905703840871	-4
5	-6.87199828087570	-5
6	-6.60084951704394	-6
7	-6.50109946133118	-7
8	-6.46440346657875	-8
9	-6.45090376453600	-9
10	-6.44593750169253	-10

Table 3: True levels' thresholds and prior masses are listed in the table. 10 Levels are given. We keep many digits because these are true levels.

std for 10-d case ($\times 10^{-15}$)

	$N_1 = 10^4$	$N_1 = 3 \times 10^4$	$N_1 = 10^5$
$N_2 = 10^6$	3.1760	3.2465	3.1653
$N_2 = 3 \times 10^6$	1.8297	1.8816	1.8390
$N_2 = 10^7$	0.9887	0.9704	1.0238

Table 4: The standard deviations for different N_1's and N_2's. The experiments are repeated for $1,000$ times. Compared with evidence value 9.77×10^{-14}, all the standard deviations are reasonably small. But N_2 is clearly more important in reducing variance. Note that there are 30 levels in this case, $30 \times N_1$ and N_2 are actually comparable.

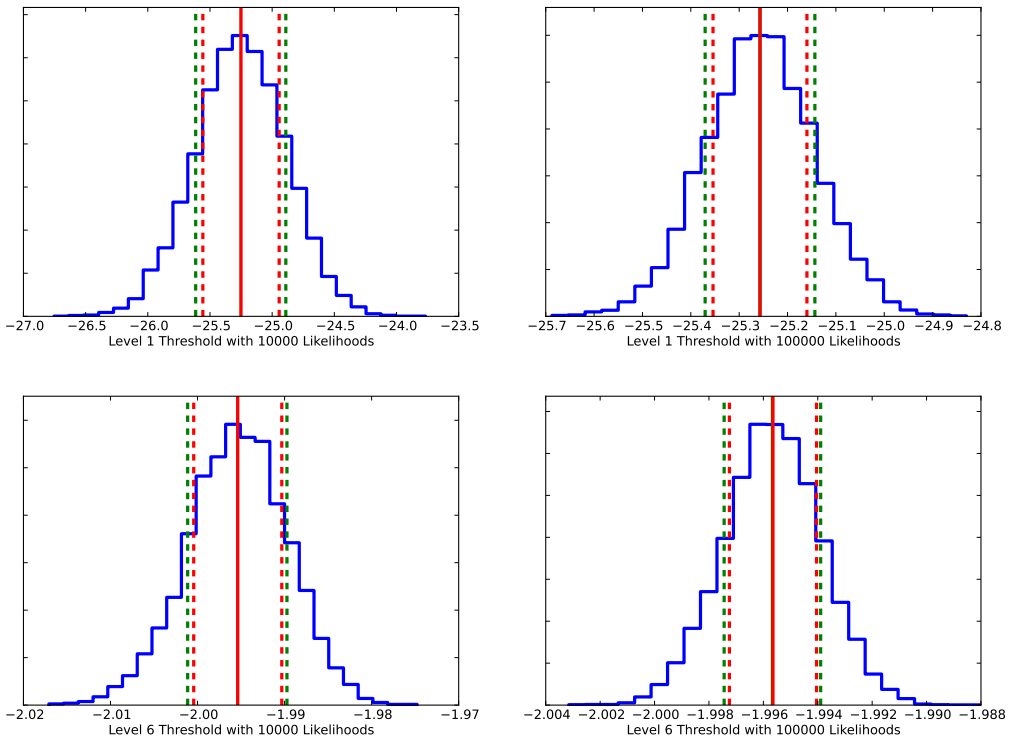

Fig. 1.— Level 1 (upper) and Level 6 (lower): Both histograms are plotted with 25 bins and 10,000 samples. (Here samples mean repetitions of the experiment, not the samples of likelihood to build every single level.) The left-hand side is the histogram of levels built with N_{1a} likelihoods and the right-hand side is the histogram of levels built with N_{1b} likelihoods. The red solid lines indicate the true values of the logarithm of the likelihood thresholds of levels and dark green solid lines indicate the experimental mean of the logarithm of the likelihood thresholds, which cannot be distinguished from the true values in these pictures. The red dashed lines indicate the theoretical standard deviation and the dark green dashed lines indicate the experimental standard deviation.

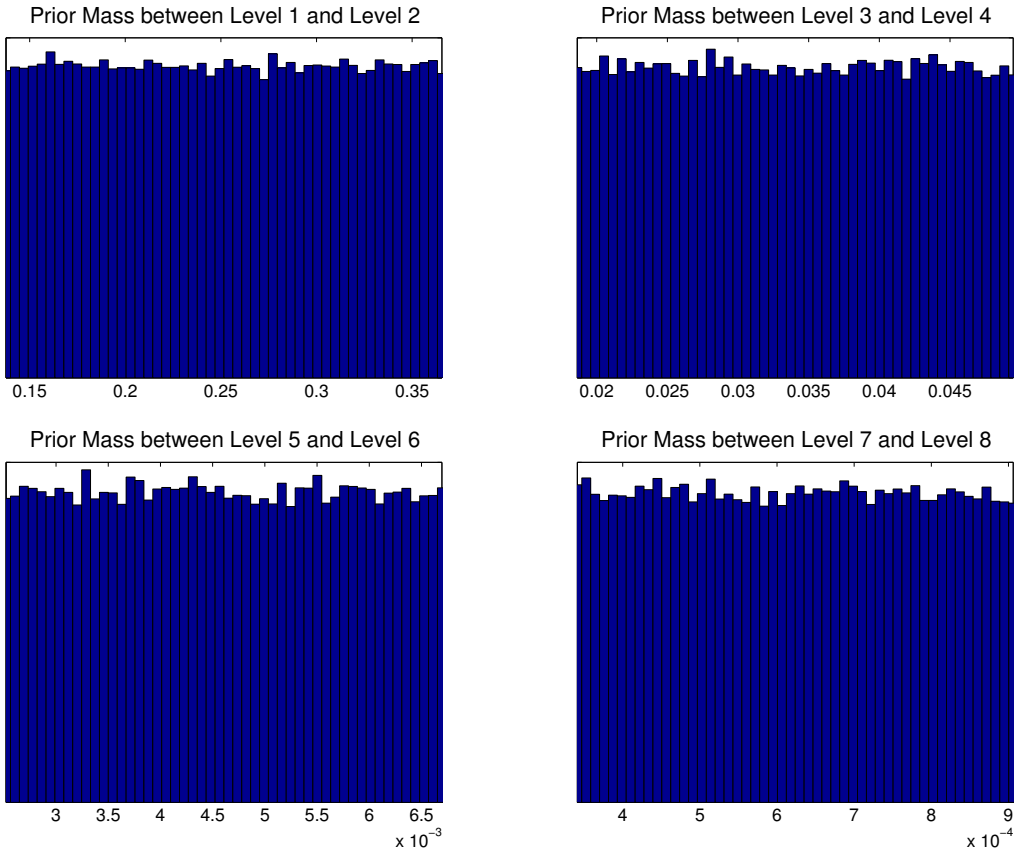

Fig. 2.— Histograms of prior masses of samples inside a bin sandwiched by two adjacent levels. 4 examples are given. All are approximately uniform distribution.

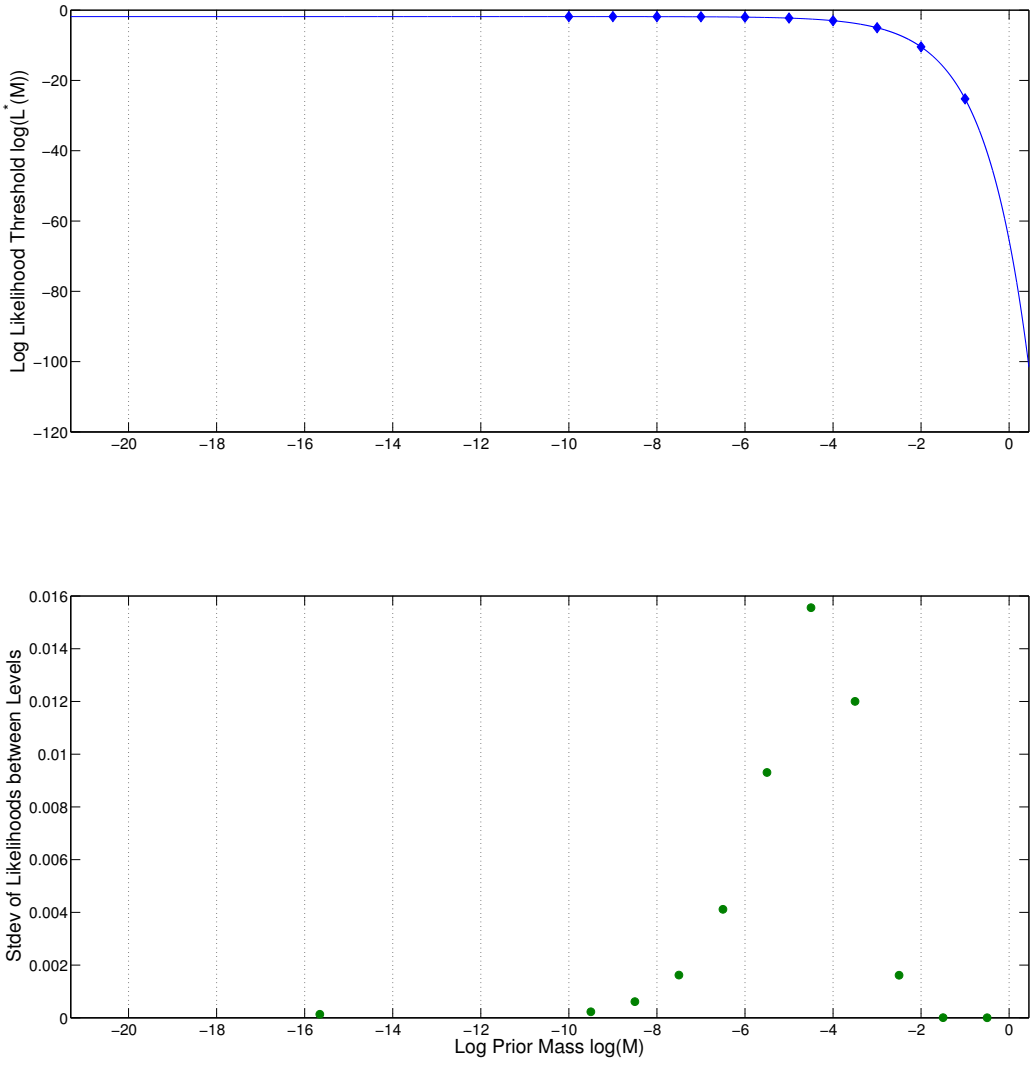

Fig. 3.— The 10 levels in the figure are the true levels summarized in Tab. (3). Notice that the standard deviation of likelihood samples between level 3 ($\log M = -3$) and level 6 ($\log M = -6$) will pretty much determine the standard deviation of the final result.