Nested Sampling over Discrete Structures

Bastiaan J. Braams
Centrum Wiskunde & Informatica (CWI), Amsterdam

John Skilling Celebration, Cambridge (UK), 2025-09-05

Introduction and Motivation

Best and Better Linear Unbiased Estimators

Estimators for an absorbing Markov chain

Nested Sampling over Discrete Structures

Closing Remarks

Introduction and Motivation

Discrete structure $(D, \preceq)$, a large finite set D with a preorder $\preceq$.

Nested sampling descends in $\preceq$ towards terminal nodes $T \subseteq D$.

We may count the number of visits $c(i)$ to any point $i \in D$.

Sparse region: $\mathcal{E}(c(i)) \ll 1$. Dense region: $1 \ll \mathcal{E}(\exp(c(i)))$.

In the continuous case all is sparse. The sample analysis proceeds from a global Z (normalization constant) down.

In the dense region repeated visits also inform about Z.

To be addressed here: How to arrange the descent (the outer loop) and how to analyze the data. Nothing about the inner loop.

Prelude on Best Linear Unbiased Estimators, and a view of the (outer loop of) nested sampling as a Markov process.

Some notation follows E. W. Dijkstra: $Z = \langle \sum i : i \in D : \pi(i) \rangle$.

Classical Best Linear Unbiased Estimator (BLUE)

Discrete space D.

Unnormalized probability distribution $p : D \to \mathbb{R}_+$.

(Unknown) normalization constant $Z = \langle \sum i : i \in D : p(i) \rangle$.

QoI $f : D \to \mathbb{R}$; seeking $\bar{f} = Z^{-1} \langle \sum i : i \in D : p(i) f(i) \rangle$.

Sample N times i.i.d. $i \sim p/Z$ *and obtain probability masses* $p(i)$.

$\rightsquigarrow$ count vector $c : D \to \mathbb{N}$. Let $S = \{ i : 1 \le c(i) \}$. (Hash table.)

Now $\langle \sum i : i \in S : c(i) \rangle = N$ and $\mathcal{E}(c/N) = p/Z$. Thus (BLUE):

$$\bar{f} \simeq N^{-1} \left\langle \sum i : i \in S : c(i) f(i) \right\rangle .$$

Next: seeking better BLUE with use of $\langle i : i \in S : p(i) \rangle$.

A Better Linear Unbiased Estimator

(Recall D, p, Z, N, counts $c : D \to \mathbb{N}$, and $S = \{i : 1 \le c(i)\}$.)
For $i \in D$, let $q(i) = 1 - (1 - p(i)/Z)^N$.
Given N, but before sampling, $\mathrm{Prob}(i \in S) = q(i)$.
Let $w(i) = p(i)/q(i)$ $(i \in S)$, $w(i) = 0$ $(i \notin S)$. $\mathcal{E}(w) = p$. Thus

$$\bar{f} \simeq Z^{-1} \left\langle \sum i : i \in S : w(i) f(i) \right\rangle \ .$$

Apply it to $f : [f = 1]$ to obtain nonlinear equation for Z:

$$Z = \phi(Z) \ , \quad \phi(Z) = \left\langle \sum i : i \in S : p(i)/(1 - (1 - p(i)/Z)^N) \right\rangle \ .$$

See [BJB, arXiv:2507.19294] for analysis of this estimator for $\bar{f}$ and of the iteration $Z \mapsto \phi(Z)$.

Variations on BLUE

(Recall D, p, Z, N, counts $c : D \to \mathbb{N}$, and $S = \{i : 1 \le c(i)\}$. Also: $q(i) = 1 - (1 - p(i)/Z)^N$ $(i \in D)$; $w(i) = p(i)/q(i)$ $(i \in S)$, $w(i) = 0$ $(i \in D \setminus S)$. $\mathcal{E}(w) = p$ and $\bar{f} \simeq Z^{-1}\langle \sum i : i \in S : w(i)f(i)\rangle$.)

Minor variation: N is drawn from $\langle N : N \in \mathbb{N} : \exp(-t)t^N/N!\rangle$ for some t; after observing the sample, set $t = N$ (observed).

Then $q(i) = 1 - \exp(-Np(i)/Z)$; all else as before.

Perspective from sufficient statistics: Data S and p_S are sufficient for c_S; i.e., whatever the underlying probability distribution for p, the probability distribution of c_S is known from S and p_S.

Stronger: Data S and p_S are complete for c_S (for rich enough class of p). $\rightsquigarrow$ Unique Minimal Variance Unbiased Estimator (UMVUE).

Common perspective for later: Separate sampling from analysis.

Estimators for an absorbing Markov Chain

(Outer iteration of nested sampling ~ absorbing Markov chain. Only for clarity, probability distributions are normalized here.)

Discrete space D. Starter probability distribution $\pi : D \to \mathbb{R}_+$.

Substochastic transition matrix $K : D \times D \to \mathbb{R}_+$; $\rho(K) < 1$.

N independent runs with starting positions drawn from π.

Hit counts $\langle i : i \in D : c(i) \rangle$. Let $S = \{ i : 1 \le c(i) \}$. (Hash table.)

Analysis in the manner of the earlier BLUE. Obtain hitting probabilities $q(i)$. Then weighted sums are approximated by:

$$\left\langle \sum i : i \in D : p(i) f(i) \right\rangle \simeq \left\langle \sum i : i \in S : (p(i)/q(i)) f(i) \right\rangle .$$

Next: nonlinear equations (concentrated on S) to obtain q.

Analysis for an absorbing Markov chain

(Recall D, π (normalized), K (substochastic), N, c, S. Seeking q.)

Green matrix $G = (\mathrm{Id} - K)^{-1}$. $G(i,j)$ is expected number of visits to j for a particle starting at i.

Let $s^T = \pi^T G$ (iteration $s^T = \pi^T + s^T K$). $s(i)$ is the expected number of hits of i for a run with starting position drawn from π.

Hitting probabilities for one run: $h(i) = s(i)/G(i,i)$.

For N runs, let $\langle i : i \in D : q(i) \rangle$ be the overall hitting probabilities.

$1 - q(i)$ is multiplicative: $q(i) = 1 - (1 - h(i))^N$ for $i \in D$.

Iteration on S: $s(j) = \pi(j) + \langle \sum i : i \in S : (s(i)/q(i))K(i,j) \rangle$.

Local relations: $h = s/\mathrm{diag}(G)$, $q = 1 - (1 - h)^N$.

...

...Analysis for an absorbing Markov chain

(Recall D, π, K, N, c, S. Seeking q. $G = (\mathrm{Id} - K)^{-1}$, $s^T = \pi^T G$, $h(i) = s(i)/G(i,i)$. Finally $q(i) = 1 - (1 - h(i))^N$.)

Iteration on S (think conjugate directions):

$$s(j) = \pi(j) + \left\langle \sum k : k \in S : (s(k)/q(k))K(k,j) \right\rangle$$
$$G(i,j) = \delta(i,j) + \left\langle \sum k : k \in S : (G(i,k)/q(k))K(k,j) \right\rangle$$

Iteration to obtain G is painful.

$(1 - K(i,i))^{-1} \leq G(i,i)$; equality for monotone Markov chains.

Hutchinson: $G(i,i) = \mathcal{E}(x(i)(Gx)(i))$ if $\mathcal{E}(x) = 0 \wedge \mathcal{E}(xx^T) = \mathrm{Id}$.

Caveat. Sums over D into sums over S needs sufficiently diffuse functions. It will not work for familiar local Markov kernels.

On to nested sampling...

Nested sampling over discrete structures

Notation throughout:

$(D, \leq)$ is a finite set with a preorder (reflexive and transitive).

$\pi : D \to \mathbb{R}_+$ is a probability distribution (possibly unnormalized).

$Z = \langle \sum i : i \in D : \pi(i) \rangle$ is the normalization constant.

For $i \in D$, $D_{\prec}(i) = \{j : j \in D \wedge j \prec i\}$ and $Z_{\prec}(i) = \langle \sum j : j \prec i : \pi(j) \rangle$.

To follow:

Description of the nested sampling outer loop.

Analysis of the output for case of a general preorder.

Specialization to case of a total preorder.

Outer loop for nested sampling

Given a positive integer N: the initial number of live elements.

Given a membership test for a downset $T \subseteq D$, called terminal elements, that includes at least all minimal elements of D.

Initialize $S := \{\}$; an empty bag of elements of D.

Initialize $R \subseteq D$; a bag of N elements drawn i.i.d. $\sim \pi/Z$.

(With each element $i \in S$ or $i \in R$ is associated its mass $\pi(i)$.)

While $R \neq \{\}$:

Identify i: any maximal element of R; move it from R to S.
If $i \notin T$: select $j \in D_{\prec}(i)$ random $\sim \pi/Z_{\prec}(i)$ and add it to R.

End while.

The result of the procedure is the bag S of elements from D.

Analysis of the output

Important for the analysis: the output is unchanged if we run N single-item chains independently and accumulate the results.

Procedures following the BLUE and absorbing Markov chain analysis, now exposition for unnormalized probabilities.

View S as a set with counts $c : S \to \mathbb{N}$ and masses $\pi : S \to \mathbb{R}_+$.

Before sampling, let $q(i) = \mathrm{Prob}(1 \le c(i))$ $(i \in D)$.

After sampling, weighted sums are approximated by:

$$\left\langle \sum i : i \in D : p(i) f(i) \right\rangle \simeq \left\langle \sum i : i \in S : (p(i)/q(i)) f(i) \right\rangle \; .$$

Next: nonlinear equations (concentrated on S) to obtain q.

...Analysis of the output

(Seeking q: $q(i) = \mathrm{Prob}(1 \le c(i))$ for $i \in D$.)

As previously, $1 - q$ is multiplicative: $q(i) = 1 - (1 - h(i))^N$, where $h(i)$ is the hitting probability of i for the case $N = 1$.

Strict monotone Markov chain: $G(i, i) = 1$, $h \equiv s$ and a recurrence:

$$\begin{aligned} h(j) &= \pi(j)/Z + \left\langle \sum i : i \in D \setminus T \wedge j \prec i : h(i)\pi(j)/Z_{\prec}(i) \right\rangle \\ &\simeq \pi(j)/Z + \left\langle \sum i : i \in S \setminus T \wedge j \prec i : (h(i)/q(i))\pi(j)/Z_{\prec}(i) \right\rangle \end{aligned}$$

Also $Z_{\prec}$ (and $Z_{\preceq}$) can be approximated by sums over S

$$Z_{\prec}(j) \simeq \left\langle \sum i : i \in S \wedge i \prec j : \pi(i)/q(i) \right\rangle$$

Iteration starts with q, upward recurrence for $Z_{\prec}$, Z, downward recurrence for h, recalculate q.

...Analysis of the output

(q; recurrence up $\rightsquigarrow Z_{\prec}$, Z; down $\rightsquigarrow h$; local algebra $\rightsquigarrow q$.)

$$
\begin{aligned}
Z_{\prec}(j) &\simeq \left\langle \sum i : i \in S \wedge i \prec j : \pi(i)/q(i) \right\rangle \\
h(j) &\simeq \pi(j)/Z + \left\langle \sum i : i \in S \setminus T \wedge j \prec i : (h(i)/q(i))\pi(j)/Z_{\prec}(i) \right\rangle \\
q(i) &= 1 - (1 - h(i))^N
\end{aligned}
$$

All on S. Initial guess may come from $\mathcal{E}(c/N) = h/q$ on S, but this will only work in the dense region, $\{i : 2 \leq c(i)\}$.

In the sparse region, $\{i : Nh(i) \ll 1\}$, $q(i)/h(i) \simeq N$, and then

$$
q(j)/\pi(j) \simeq N/Z + \left\langle \sum i : i \in S \setminus T \wedge j \prec i : 1/Z_{\prec}(i) \right\rangle \ .
$$

This makes a nice pair with the recurrence for $Z_{\prec}$. ...

...Analysis of the output, case of a total preorder

Recall the approximations in the sparse region, and note that a total preorder implies a proper total order in the sparse region.

$$Z_{\prec}(j) \simeq \left\langle \sum i : i \in S \wedge i \prec j : \pi(i)/q(i) \right\rangle$$
$$q(j)/\pi(j) \simeq N/Z + \left\langle \sum i : i \in S \setminus T \wedge j \prec i : 1/Z_{\prec}(i) \right\rangle$$

We make contact with the classical analysis of nested sampling.

Sort S ascending by $\prec$ to get $\langle m : 0 \leq m < M : i_m \rangle$, and simple recurrences. In the sparse region, for large M and large N:

$$Z_{\prec}(i_m) \simeq Z(1-1/N)^{M-m}, \simeq Z\exp((m-M)/N)$$
$$q(i_m)/\pi(i_m) \simeq \frac{N}{Z}(1-1/N)^{m+1-M}, \simeq \frac{N}{Z}\exp((M-m-1)/N)$$

Closing remarks

One loose end: If Z is known in advance then we have too much information. I propose to address it by treating N as an uncertain parameter in the expression $q(i) = 1 - (1 - h(i))^N$.

The BLUE work (first section of the talk) has a preliminary write-up [arXiv:2507.19294].

A nested sampling manuscript will follow. Primarily it still needs some (synthetic) numerical experiments or demonstrations.

Comments will be appreciated!