An Importance Sampling Algorithm for Models with Weak Couplings

Mehdi Molkaraie
ETH Zurich
mehdi.molkaraie@alumni.ethz.ch

Abstract—We propose an importance sampling algorithm to estimate the partition function of the Ising model and the q-state Potts model. The proposal (auxiliary) distribution is defined on a spanning tree of the Forney factor graph representing the model, and computations are done on the remaining edges. In contrast, in an analogous importance sampling algorithm in the dual Forney factor graph, computations are done on a spanning tree, and the proposal distribution is defined on the remaining edges.

I. INTRODUCTION

We consider the problem of estimating the partition function of the ferromagnetic Ising and q-state Potts models with spatially varying (bond-dependent) coupling parameters. The partition function is an important quantity in statistical physics [1], [2], in machine learning [3], and in information theory [4].

In general, the partition function is not available analytically but is only available as a summation with an exponential number of terms – making its exact computation intractable. Therefore, we rely on approximating the partition function [5] or deriving bounds on this quantity [6], [7].

In this paper, we first represent the models of interest with their modified Forney factor graphs (FFG), which are constructed via simple manipulations in the original FFG [8]. We then define a proposal distribution on a spanning tree in the modified FFG to propose an importance sampling algorithm for estimating the partition function. The algorithm can efficiently compute an estimate of the partition function when the coupling parameters associated with the edges that lie out of the spanning tree are weak.

In contrast, similar importance sampling algorithms can be designed in the dual FFG of the models by defining the proposal distribution on the edges that lie out of a spanning tree of the model. In this case, the partition function can be efficiently estimated when the coupling parameters on the spanning are strong [9]–[11].

The paper is organized as follows. In Section II we review the Ising model, the q-state Potts model, and their graphical model representations in terms of FFGs. The modified FFGs of the models are presented in Section III. In Section IV we describe the importance sampling algorithm for estimating the partition function. Contrast to analogous algorithms in the dual FFG is discussed in Section V.

II. THE MODELS

Let \( X_1, X_2, \ldots, X_N \) be a collection of discrete random variables. Suppose each random variable takes on values in a finite alphabet \( \mathcal{X} \), which in this context is equal to the abelian group \( \mathbb{Z}/q\mathbb{Z} = \{0, 1, \ldots, q-1\} \). Let \( x_i \) represent a possible realization of \( X_i \), \( x \) stand for a configuration \((x_1, x_2, \ldots, x_N)\), and \( X \) stand for \((X_1, X_2, \ldots, X_N)\).

For simplicity, we assume ferromagnetic models, with periodic boundaries, with pairwise interactions, and without an external magnetic field. Although some of our results are applicable to more general settings.

Let \( f: \mathcal{X}^N \rightarrow \mathbb{R}_{\geq 0} \) be a non-negative function, which factors into a product of local functions \( v_{k, \ell}: \mathcal{X}^2 \rightarrow \mathbb{R}_{\geq 0} \) as

\[
f(x) = \prod_{(k, \ell) \in E} v_{k, \ell}(x_k, x_\ell)
\]

where \( E \) contains all the unordered pairs \((k, \ell)\) with non-zero interactions. A real coupling parameter \( J_{k, \ell} \) is associated with the interacting pair \((x_k, x_\ell)\).

From (1), we define the following probability mass function (known as the Boltzmann distribution [1])

\[
p(x) \triangleq \frac{f(x)}{Z}
\]

Here, the normalization constant \( Z \) is the partition function given by

\[
Z = \sum_{x \in \mathcal{X}^N} f(x)
\]

A. The Ising Model

In the Ising model, \( q = 2 \) and

\[
v_{k, \ell}(x_k, x_\ell) = \begin{cases} e^{J_{k, \ell}}, & \text{if } x_k = x_\ell \\ e^{-J_{k, \ell}}, & \text{if } x_k \neq x_\ell \end{cases}
\]

The model is called ferromagnetic (resp. antiferromagnetic) if \( J_{k, \ell} > 0 \) (resp. \( J_{k, \ell} < 0 \)) for each \((k, \ell) \in E\). If the couplings can be both positive or negative, the model is known as an Ising spin glass.

B. The q-State Potts Model

In the Potts model, \( q > 2 \) and

\[
v_{k, \ell}(x_k, x_\ell) = \begin{cases} e^{J_{k, \ell}}, & \text{if } x_k = x_\ell \\ 1, & \text{if } x_k \neq x_\ell \end{cases}
\]

where \( J_{k, \ell} > 0 \) in a ferromagnetic model.
C. FFG of the Models

The factorization in (1) can be represented by a FFG, in which nodes represent the factors and edges represent the variables. The edge that represents some variable $x$ is connected to the node representing the factor $v_2(\cdot)$ if and only if $x$ is an argument of $v_2(\cdot)$. If a variable (an edge) appears in more than two factors, such a variable is replicated using equality indicator factors (9).

The FFG of the 2D Ising model with pairwise (nearest-neighbor) interactions is shown in Fig. 1, where the unlabeled boxes represent factors (4) and the boxes labeled “=” are equality indicator factors. E.g., in Fig. 1, for variables $X, X', X'',$ and $X'''$ the equality indicator factor is given by

$$
\Phi_\pm(x,x',x'',x''') = \delta(x-x') \cdot \delta(x-x'') \cdot \delta(x-x''')(6)
$$

where $\delta(\cdot)$ is the Kronecker delta function.

Similarly, Fig. 1 shows the FFG of the 2D Potts model with pairwise interactions, where the unlabeled boxes represent factors as in (5).

Note that in a 2D model with periodic boundary conditions

$$|E| = 2N \quad (7)$$

III. THE MODIFIED FFG

In this section, we present the modified FFG of the Ising and Potts models. Recall that all arithmetic manipulations are done modulo 2 in the case of the Ising model, and modulo $q$ in the case of the Potts model.

A. Modified FFG of the Ising Model

We note that each factor (4) is only a function of $x_k + x_\ell$, we can thus represent $v_{k,\ell}(\cdot)$ using only one variable $y_m$.

We thus let

$$
v_m(y_m) = \begin{cases} 
e^{J_m}, & \text{if } y_m = 0 \\ ne^{-J_m}, & \text{if } y_m = 1 \end{cases} \quad (8)
$$

Following the above observation, we can build the modified FFG of the 2D Ising model as shown in Fig. 2, where the unlabeled boxes represent (8) and boxes labeled “+” are mod2 indicator factors, which impose the constraint that all their incident variables sum to zero (modulo 2). E.g., in Fig. 2 for binary variables $X_1, X_2,$ and $Y_1$ the mod2 indicator factor is given by

$$
\Phi_+(y_1, x_1, x_2) = \delta(y_1 + x_1 + x_2) \quad (9)
$$

Let $Y = (Y_1, Y_2, \ldots, Y_{|E|})$ be the set of all the variables attached to the mod2 indicator factor factors and $y$ be a realization of $Y$. Here, $|E|$ denotes the cardinality of $E$, which is also equal to the number of unordered interacting pairs in the model.

Lemma 1. Consider a cycle of length $c$ in the modified FFG of the Ising model. For variables $Y_1, Y_2, \ldots, Y_c$ attached to the mod2 indicator factors in the cycle, it holds that

$$
\sum_{m=1}^c Y_m = 0 \quad (10)
$$

Proof. In (10), each $Y_m$ can be expanded as the symmetric difference of the corresponding adjacent variables $(X_k, X_\ell)$ in the cycle. Moreover, each variable appears twice in this expansion. We conclude that $\sum_{m=1}^c Y_m = 0$.

An example of a cycle is shown by thick edges in Fig. 2, where $Y_1, Y_2, \ldots, Y_c$ are marked by blue edges.

B. Modified FFG of the q-State Potts Model

In this case, each factor (5) is only a function of $x_k - x_\ell$. Similar to our approach in Section III-A, we represent (5) as

$$
v_k(y_k) = \begin{cases} \ne^{J_k}, & \text{if } y_k = 0 \\ 1, & \text{otherwise.} \end{cases} \quad (11)
$$

The modified FFG of the 2D Potts model is shown in Fig. 3, where the unlabeled boxes represent factors (11) and boxes labeled “+” are mod$_q$ indicator factors, which impose the
constraint that all their incident variables sum to zero (modulo $q$). E.g., in Fig. 4 for variables $X_1, X_2$, and $Y_1$ the mod$_q$ indicator factor is given by

$$\Phi_+(y_1, x_1, x_2, \ldots) = \delta(y_1 + x_1 - x_2)$$  \hspace{0.5cm} (12)

A sign inverter (depicted by a small circle) is inserted in one of the edges incident to the mod$_q$ indicator factors. However, the choice on which side to insert them can be made arbitrarily (because of the symmetry in the factors).

There is again a linear dependency among the variables $Y_1, Y_2, \ldots, Y_c$ in any cycle of length $c$ in the modified FFG of the Potts model. However, the dependency among variables is affected by the arrangement of the sign inverters. As an example, in Fig. 4 we have arranged the sign inverters in a way that the sum of the variables in any cycle of length four be zero: each $Y_m$ can be expanded as the difference between the corresponding adjacent variables $(X_k, X_l)$ in the cycle; furthermore, each variable appears twice in this expansion, once with a positive sign, and once with a negative sign.

C. Variables in the Modified FFG

We partition $E$ into two disjoint subsets $T$ and $S$, where $T$ is a spanning tree in the modified FFG. Thus $Y$ is also partitioned into $Y_T$ and $Y_S$. In such a partitioning, $Y_S$ can be computed as linear combination of $Y_T$ (cf., Lemma 1).

An example of a spanning tree in the modified FFG of the 2D Ising model is illustrated in Fig. 5 where the thick blue edges represent $Y_T$ and the thin red edges represent $Y_S$. Here, $Y_S$ is a linear combination of $Y_T$.

In a 2D grid with periodic boundary conditions, we have

$$|T| = N - 1$$ \hspace{0.5cm} (13)  
$$|S| = N + 1$$ \hspace{0.5cm} (14)

Accordingly, let

$$\Upsilon_T(y_T) = \prod_{m \in T} \upsilon_m(y_m)$$ \hspace{0.5cm} (15)

$$\Upsilon_S(y_S) = \prod_{m \in S} \upsilon_m(y_m)$$ \hspace{0.5cm} (16)

We define the following proposal probability mass function on a spanning tree in the modified FFG

$$q(y_T) \triangleq \frac{\Upsilon_T(y_T)}{Z_q}$$ \hspace{0.5cm} (17)

with

$$Z_q = \sum_{Y_T} \Upsilon_T(y_T)$$ \hspace{0.5cm} (18)

In this set-up, $Z_q$ is available in closed form. For the Ising model, we obtain

$$Z_q = \prod_{m \in T} (e^{J_m} + e^{-J_m})$$ \hspace{0.5cm} (19)

$$= 2^{|T|} \prod_{m \in T} \cosh J_m$$ \hspace{0.5cm} (20)

and for the $q$-state Potts model

$$Z_q = \prod_{m \in T} \sum_{t=0}^{q-1} \upsilon_m(t)$$ \hspace{0.5cm} (21)

$$= \prod_{m \in T} (e^{J_m} + q - 1)$$ \hspace{0.5cm} (22)

We also let

$$\Upsilon(y) = \prod_{m \in E} \upsilon_m(y_m)$$ \hspace{0.5cm} (23)

The global probability mass function in the modified FFG can then be defined as

$$p_M(y) \triangleq \frac{\Upsilon(y)}{Z_M}$$ \hspace{0.5cm} (24)
where $Z_M$ is the partition function of the modified FFG.

The partition functions $Z$ and $Z_M$ are closely related.

**Lemma 2.** In a ferromagnetic Ising model, the partition functions $Z$ and $Z_M$ are related to each other by

$$ Z = 2Z_M \quad (25) $$

**Proof.** Let $\bar{x}$ be the component-wise addition of $x$ and the all-ones vector, i.e., in $\bar{x}$, components of $x$ that are 0 become 1, and those that are 1 become 0. From each $x$, we can create a valid configuration $y$ in the modified NFG. But then, $\bar{x}$ will give rise to the same configuration $y$.

Hence, each valid configuration $y$ in the modified NFG corresponds to two configurations $x$ and $\bar{x}$. Due to the symmetry in the factors, these configurations contribute equally to the sum in (3); therefore, $Z = 2Z_M$. $\blacksquare$

**Lemma 3.** For a ferromagnetic $q$-state Potts model, the partition functions $Z$ and $Z_M$ are related to each other by

$$ Z = qZ_M \quad (26) $$

The proof follows along the same lines as in the proof of Lemma 2.

In Section [IV] we propose an importance sampling algorithm in the modified FFG of the Ising model and the $q$-state Potts model to estimate $Z_M$, which can then be used to compute an estimate $Z$.

**IV. Importance Sampling in the Modified FFG**

The importance sampling algorithm works as follows. We first draw independent samples $y_T^{(1)}, y_T^{(2)}, \ldots$ according to $q(y_T)$ in (17), and thereafter compute $y_S^{(1)}, y_S^{(2)}, \ldots$. These samples are then used to compute an estimate of $Z_M$.

Drawing independent samples according to $q(y_T)$ is straightforward. For the Ising model, the product form of (15) suggests that to draw $y_T^{(t)}$ we can do the following:

1: draw $u^{(t)}_1, u^{(t)}_2, \ldots, u^{(t)}_{|T|} \sim U[0, 1]$
2: for $m = 1$ to $|T|$ do
3: if $u_m^{(t)} < \frac{1}{1 + e^{-2J_m}}$ then
4: $y_m^{(t)} = 0$
5: else
6: $y_m^{(t)} = 1$
7: end if
8: end for

In Line 3,

$$ \frac{1}{1 + e^{-2J_m}} = \frac{v_m(0)}{v_m(0) + v_m(1)} \quad (27) $$

which is equal to $\text{sigm}(2J_m)$, where $\text{sigm}()$ denotes the sigmoid (logistic) function [3, Chapter 1].

Similarly, in the case of the $q$-state Potts model, we can apply the following subroutine to draw independent samples $y_T^{(t)}$ according to the corresponding proposal distribution.

After drawing $y_T^{(t)}$, we compute $y_S^{(t)}$. We then use the samples in the following importance sampling algorithm.

1: for $\ell = 1$ to $L$ do
2: draw $y_T^{(t)}$ according to $q(y_T)$
3: compute $y_S^{(t)}$
4: end for
5: compute

$$ \hat{Z}_M^{IS} = \frac{Z_q}{L} \sum_{\ell=1}^{L} \gamma_S(y_S^{(t)}) \quad (28) $$

We show that $\hat{Z}_M^{IS}$ is an unbiased estimator of $Z_M$.

$$ E_q[\hat{Z}_M^{IS}] = \frac{Z_q}{L} \sum_{\ell=1}^{L} E_q[\gamma_S(y_S)] = \sum_y \gamma_T(y_T) \gamma_S(y_S) = Z_M $$

**A. The Variance of $\hat{Z}_M^{IS}$**

For a finite-size model, the variance of $\hat{Z}_M^{IS}$ can be computed as

$$ \text{Var}[\hat{Z}_M^{IS}] = E \left[ \left( \hat{Z}_M^{IS} \right)^2 \right] - \left( E[\hat{Z}_M^{IS}] \right)^2 \quad (29) $$

$$ = \frac{1}{L} \left( Z_q^2 E_q[\gamma_S^2(y_S)] - Z_M^2 \right) \quad (30) $$

Fig. 5: A spanning tree in the modified FFG of the 2D Ising model. The thick blue edges represent $Y_T$ and the thin red edges represent $Y_S$. Here, $Y_S$ is a linear combination of $Y_T$. 

1: draw $u_1^{(t)}, u_2^{(t)}, \ldots, u_{|T|}^{(t)} \sim U[0, 1]$
2: for $m = 1$ to $|T|$ do
3: if $u_m^{(t)} < \frac{1}{1 + (q-1)e^{-J_m}}$ then
4: $y_m^{(t)} = 0$
5: else
6: draw $y_m^{(t)}$ randomly from $\{1, 2, \ldots, q-1\}$
7: end if
8: end for

After drawing $y_T^{(t)}$, we compute $y_S^{(t)}$. We then use the samples in the following importance sampling algorithm.
Hence
\[
\frac{L}{Z_M^2} \text{Var}[\tilde{Z}_M^{IS}] = \left( \frac{Z_M}{Z_M} \right)^2 \text{E}_q[\chi^2_{S}(Y_S)] - 1
\]
\[
= \sum_{\gamma} \frac{p_M(\gamma)}{q(\gamma)} - 1
\]
\[
= \chi^2(p_M, q)
\]
where \(\chi^2(\cdot, \cdot)\) denotes the chi-squared divergence, which is non-negative, with equality to zero if and only if its two arguments are equal [12, Chapter 4].

For simplicity, let us assume that for \(m \in S\), the coupling parameters of the model are constant denoted by \(J_m\). In the limit \(J_m \to 0\), we have
\[
\lim_{J_m \to 0} p_M(\gamma) = q(\gamma)
\]
Hence
\[
\lim_{J_m \to 0} \chi^2(p_M, q) = 0
\]
Therefore, \(Z_M\) can be estimated efficiently via the proposed importance sampling estimator when \(J_m\) is small for \(m \in S\).

V. IMPROBABILITY SAMPLING IN THE DUAL FFG

We briefly discuss an analogous importance sampling algorithm in the dual FFG of the \(q\)-state Potts model. In the dual FFG, we denote the partition function by \(Z_d\). In the dual domain, small circles attached to equality indicator factors denote sign inverters. The thick edges show a spanning tree in the dual FFG.

An example of such a partitioning is shown in Fig. 6 where \(\tilde{Y}_T\) is the set of all the variables associated with the thick edges and \(\tilde{Y}_S\) is the set of all the variables associated with the remaining thin edges.

Let
\[
\Gamma_S(\tilde{Y}_S) = \prod_{m \in S} \gamma_m(\tilde{y}_m)
\]
\[
\Gamma_T(\tilde{Y}_T) = \prod_{m \in T} \gamma_m(\tilde{y}_m)
\]
We define the following proposal probability mass function
\[
q_d(\tilde{Y}_S) \propto \frac{\Gamma_S(\tilde{Y}_S)}{Z_{d}}
\]
where \(Z_{d}\) is analytically available as
\[
Z_{d} = \sum_{\tilde{y}_S} \Gamma_S(\tilde{Y}_S)
\]
\[
= \prod_{m \in S} \sum_{t=0}^{q-1} \gamma_m(t)
\]
\[
q^{IS} \text{exp}\left( \sum_{m \in S} J_m \right)
\]
We let
\[
\Gamma(\tilde{y}) = \prod_{m \in E} \gamma_m(\tilde{y}_m)
\]
and define the global probability mass function in the dual FFG as
\[
p_d(\tilde{y}) \propto \frac{\Gamma(\tilde{y})}{Z_d}
\]
where $Z_d$ is the partition function of the dual FFG.

The importance sampling algorithm works as follows: at iteration $\ell$, we draw a sample $\tilde{y}^{(\ell)}_S$ according to the proposal distribution $\mathcal{G}^{(m)}$. The product form of (39) suggests that in order to draw $\tilde{y}^{(\ell)}_S$ we can apply the following subroutine [11].

1. draw $u_{1}^{(\ell)}, u_{2}^{(\ell)}, \ldots, u_{|S|}^{(\ell)} \sim \mathcal{U}[0, 1]$
2. for $m = 1$ to $|S|$ do
3.  if $u_{m}^{(\ell)} < \frac{1 + (q - 1)\exp(-J_m)}{q}$ then
4.  $\tilde{y}_{m}^{(\ell)} = 0$
5.  else
6.  draw $\tilde{y}_{m}^{(\ell)}$ randomly from $\{1, 2, \ldots, q - 1\}$
7. end if
8. end for

After drawing $\tilde{y}^{(\ell)}_S$, we compute $\tilde{y}^{(\ell)}_T$. Finally, we use the following importance sampling algorithm to estimate $Z_d$.

1. for $\ell = 1$ to $L$ do
2.  draw $\tilde{y}^{(\ell)}_S$ according to $q(\tilde{y}_S)$
3.  compute $\tilde{y}^{(\ell)}_T$
4.  end for
5. compute

$$\hat{Z}_d^{IS} = \frac{Z_d}{L} \sum_{\ell=1}^{L} \Gamma_T(\tilde{y}^{(\ell)}_T)$$  \hspace{1cm} (47)

Here, $\hat{Z}_d^{IS}$ is an unbiased estimator of $Z_d$, i.e.,

$$\mathbb{E}_{q(\cdot)}[\hat{Z}_d^{IS}] = Z_d$$  \hspace{1cm} (48)

see [10].

Similar to our approach in Section IV-A, we can show

$$\frac{L}{Z_d^2} \text{Var}[\hat{Z}_d^{IS}] = \chi^2(p_d, q_d)$$  \hspace{1cm} (49)

For simplicity, we assume that for $m \in T$, the coupling parameters of the model are constant denoted by $J_T$.

In the limit $J_T \to \infty$

$$\lim_{J_T \to \infty} p_d(\tilde{y}) = q_d(\tilde{y})$$  \hspace{1cm} (50)

Thus

$$\lim_{J_T \to \infty} \frac{L}{Z_d^2} \text{Var}[\hat{Z}_d^{IS}] = 0$$  \hspace{1cm} (51)

We conclude that $Z_d$ can be estimated efficiently via the importance sampling estimator when $J_m$ is large for $m \in T$.

For more details on constructing the dual FFG of the Ising model and the $q$-state Potts model, see [15, 16, 9–11].

VI. CONCLUSION

We proposed an importance sampling algorithm in the modified FFG of the Ising model and the $q$-state Potts model to estimate the partition function. The proposal distribution of the importance sampling algorithm is defined on a spanning tree of the model. The algorithm can efficiently compute an estimate of the partition function when the coupling parameters associated with the edges that lie out of the spanning tree are weak. In contrast, the proposal distribution for the analogous importance sampling algorithm in the dual FFG is defined on the edges that lie out of the spanning tree. In this case, accurate estimates of the partition function can be obtained when couplings associated with the edges of the spanning tree are strong.

The methods can handle more demanding cases when combined with annealed importance sampling [17].

REFERENCES

[1] J. M. Yeomans, Statistical Mechanics of Phase Transitions. Oxford University Press, 1992.
[2] R. J. Baxter, Exactly Solved Models in Statistical Mechanics. Dover Publications, 2007.
[3] K. P. Murphy, Machine Learning: A Probabilistic Perspective. The MIT Press, 2012.
[4] D. J. C. MacKay, Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 2003.
[5] G. Potamianos and J. Goutsias, “Stochastic approximation algorithms for partition function estimation of Gibbs random fields,” IEEE Trans. Information Theory, vol. 43, pp. 1984–1965, Nov. 1997.
[6] J. S. Yedidia, “An idiosyncratic journey beyond mean field theory,” in Advanced Mean Field Methods: Theory and Practice. The MIT Press, 2001, pp. 21–36.
[7] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, “A new class of upper bounds on the log partition function,” IEEE Trans. Information Theory, vol. 51, pp. 2313–2335, July 2005.
[8] G. D. Forney, Jr., “Codes on graphs: normal realization,” IEEE Trans. Information Theory, vol. 47, pp. 520–548, Feb. 2001.
[9] M. Molkaraie, “An importance sampling scheme for models in a strong external field,” Proc. 2015 IEEE Int. Symp. on Inf. Theory, Hong Kong, June 14–19, 2015, pp. 1179–1183.
[10] M. Molkaraie, “An importance sampling algorithm for the Ising model with strong couplings,” Proc. 2106 Int. Zurich Seminar on Communications (IZS), March 2–4, 2016, pp. 180–184.
[11] M. Molkaraie and V. Gómez, “Efficient Monte Carlo methods for the Potts model at low temperature,” arXiv:1506.07044
[12] I. Csiszár and P. C. Shields, Information Theory and Statistics: A Tutorial. now Publishers Inc., 2004.
[13] A. Al-Bashabsheh and Y. Mao, “Normal factor graphs and holographic transformations,” IEEE Trans. Information Theory, vol. 57, pp. 752–763, Feb. 2011.
[14] G. D. Forney, Jr., “Codes on graphs: duality and MacWilliams identities,” IEEE Trans. Information Theory, vol. 57, pp. 1382–1397, Feb. 2011.
[15] M. Molkaraie and H.-A. Loeliger, “Partition function of the Ising model via factor graph duality,” Proc. 2013 IEEE Int. Symp. on Information Theory. Istanbul, Turkey, July 7–12, 2013, pp. 2304–2308.
[16] A. Al-Bashabsheh and Y. Mao, “On stochastic estimation of the partition function,” Proc. 2014 IEEE Int. Symp. on Information Theory, Honolulu, USA, June 29–July 4, 2014.
[17] R. M. Neal, “Annealed importance sampling,” Statistics and Computing, vol. 11, pp. 125–139, 2001.