Estimating the Partition Function of 2-D Fields and the Capacity of Constrained Noiseless 2-D Channels Using Tree-Based Gibbs Sampling

Hans-Andrea Loeliger and Mehdi Molkaraie
ETH Zurich
Dept. of Information Technology & Electrical Engineering
8092 Zürich, Switzerland
Email: {loeliger, molkarai}@isi.ee.ethz.ch

***Abstract*—Tree-based Gibbs sampling (proposed by Hamze and de Freitas) is used to compute a Monte-Carlo estimate of the partition function of factor graphs with cycles. The proposed method can be used, in particular, to compute the capacity of noiseless constrained 2-D channels.**

I. Introduction

Let $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_N$ be finite sets, let $\mathcal{X}$ be the Cartesian product $\mathcal{X} \triangleq \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_N$, and let f be a nonnegative function $f : \mathcal{X} \to \mathbb{R}$. We are interested in computing (exactly or approximately) the quantity

$$Z \triangleq \sum_{x \in \mathcal{X}} f(x) \tag{1}$$

(or, equivalently, $\frac{1}{N} \log Z$) for cases where

- $\mathcal{X}_1, \ldots \mathcal{X}_N$ are "small" sets (e.g., $|\mathcal{X}_1| = |\mathcal{X}_2| = \ldots = 2$),
- N is large,
- and f has a "useful" factorization (as will be detailed below).

Note that

$$p(x) \triangleq \frac{1}{Z} f(x) \tag{2}$$

is a probability mass function on $\mathcal{X}$. We will also need the set

$$\mathcal{X}_{f+} \triangleq \{x \in \mathcal{X} : f(x) \neq 0\}. \tag{3}$$

The quantity (1) is known as the "partition function" in statistical physics (where it is considered as a function of a "temperature" parameter that is of no concern to us here). The computation of (1) is also the key to computing information rates of source / channel models with memory [1]–[3].

If f has a cycle-free factor graph with not too many states, then the sum (1) can be computed by sum-product message passing [1], [4]. In this paper, however, we consider the case where no such cycle-free factor graph exists. In particular, we are interested in examples of the following type.

Example 1 (Simple 2-D Constraint). Consider a grid of $N = M \times M$ binary (i.e., $\{0, 1\}$-valued) variables with the constraint that no two (horizontally or vertically) adjacent variables have both the value 1. Let f be the indicator function of this constraint, which can be factored into factors of the form

$$\kappa(x_k, x_\ell) = \begin{cases} 0, & \text{if } x_k = x_\ell = 1 \\ 1, & \text{else,} \end{cases} \tag{4}$$

with one such factor for each adjacent pair (x_k, x_ℓ).

The corresponding Forney-style factor graph of f is shown in Fig. 1, where the boxes labeled "=" are equality constraints [5]. (Fig. 1 may also be viewed as a factor graph as in [4] where the boxes labeled "=" are the variable nodes.)

This example is known as the 2-D $(1, \infty)$ constrained channel [6].

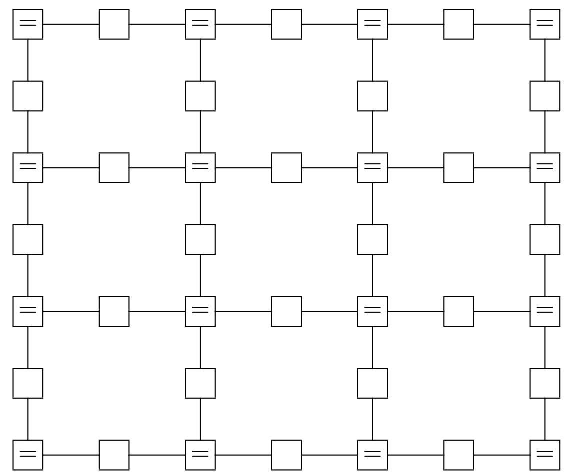

Fig. 1. Forney-style factor graph for Example 1. The unlabeled boxes represent factors as in (4).

Note that, in this example, $Z = |\mathcal{X}_{f+}|$. For this particular example, $\lim_{M\to\infty} \frac{1}{M^2} \log_2(Z) \approx 0.5879$ is known to nine decimal digits [7], [8]. However, the method proposed in this paper works also for various generalizations of this example for which this limit is not known to any useful accuracy [6].

A number of Monte-Carlo methods to estimate Z have been proposed, see [9], [10]. However, these methods assume that f is strictly positive, which excludes applications as in Example 1; more about this will be said in Section II.

In this paper, we propose a new Monte-Carlo method for the computation of Z that works also for Example 1. In

978-1-4244-4983-5/09/$25.00 © 2009 IEEE

constrast to the method of [3], the proposed method converges to the exact value of $\frac{1}{N}\log(Z)$ in the limit of infinitely many samples.

II. Estimating $1/Z$ Using Gibbs Sampling

One method to estimate $1/Z$ (and thus Z itself) goes as follows.

1) Draw samples $x^{(1)}, x^{(2)}, \ldots, x^{(K)}$ from $\mathcal{X}_{f+}$ according to $p(x)$ defined in (2).
2) Compute

$$\Gamma \triangleq \frac{1}{K|\mathcal{X}_{f+}|}\sum_{k=1}^{K}\frac{1}{f(x^{(k)})} \tag{5}$$

It is easily verified that $\mathrm{E}[\Gamma] = 1/Z$. This method was proposed in [11], see also [9].

However, there are two major issues with this method. First, it is usually assumed (as in [9], [11]) that f is strictly positive. In this case, $\mathcal{X}_{f+} = \mathcal{X}$ and $|\mathcal{X}_{f+}| = |\mathcal{X}|$ is known. However, this assumption excludes applications as in Example 1. (Indeed, in Example 1, we would have $f(x^{(k)}) = 1$ for all samples $x^{(k)}$, and $|\mathcal{X}_{f+}| = Z$ is the desired unknown quantity.) We will see how this issue is resolved by an idea from [12].

Second, there is the problem of generating the samples $x^{(1)}$, $x^{(2)}, \ldots, x^{(K)}$ according to $p(x)$. A standard general method is Gibbs sampling [9], [13], which, however, produces strongly dependent samples. In consequence, the required number of samples K is likely to exceed the limits of practicality. We will see how this issue is eased by tree-based Gibbs sampling as proposed by Hamze and de Freitas [14].

III. Proposed New Method

The proposed method combines tree-based Gibbs sampling from [14] with an idea from [12].

Let (A,B) be a partition of the index set $\{1,\ldots,N\}$ such that, (i) for fixed x_A, the factor graph of $f(x) = f(x_A, x_B)$ is a tree and (ii) for fixed x_B, the factor graph of $f(x) = f(x_A, x_B)$ is also a tree. An example of such a partition is shown in Fig. 2.

A. *Tree-Based Gibbs Sampling [14]*

Starting from some initial configuration $x^{(0)} = (x_A^{(0)}, x_B^{(0)})$, the samples $x^{(k)} = (x_A^{(k)}, x_B^{(k)})$, $k = 1, 2, \ldots$, are created as follows. First, $x_A^{(k)}$ is sampled according to

$$p(x_A|x_B = x_B^{(k-1)}) \propto f(x_A, x_B^{(k-1)}); \tag{6}$$

then $x_B^{(k)}$ is sampled according to

$$p(x_B|x_A = x_A^{(k)}) \propto f(x_A^{(k)}, x_B). \tag{7}$$

The point is that the sampling can be done very efficiently in both cases since the corresponding factor graphs are cycle-free; see the appendix for details.

Tree-based Gibbs sampling mixes much faster than naive Gibbs sampling [14].

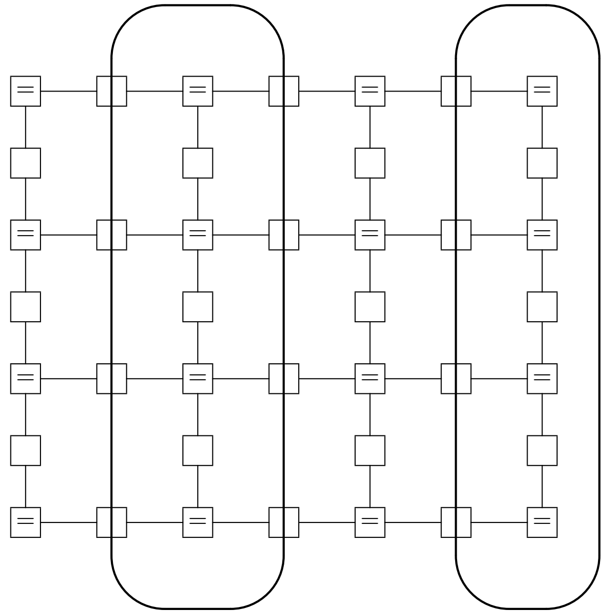

Fig. 2. Partition of Fig. 1 into two cycle-free parts (one part inside the two ovals, the other part outside the ovals).

B. *Tree-Based Estimation of $1/Z$ [12]*

Let

$$f_A(x_A) \triangleq \sum_{x_B} f(x_A, x_B) \tag{8}$$

and

$$f_B(x_B) \triangleq \sum_{x_A} f(x_A, x_B). \tag{9}$$

Since

$$\sum_{x_A} f(x_A) = \sum_{x_B} f(x_B) = \sum_{x} f(x) = Z, \tag{10}$$

we can estimate Z by applying the algorithm of Section II to f_A or to f_B (as noted in [12]). Specifically, an estimate Γ_A of $1/Z$ is formed as follows:

1) Draw samples $x_A^{(1)}, x_A^{(2)}, \ldots, x_A^{(K)}$ from $(\mathcal{X}_A)_{f_A^+}$ according to $p(x_A) \triangleq \sum_{x_B} p(x_A, x_B) = f_A(x_A)/Z$.
2)

$$\Gamma_A \triangleq \frac{1}{K|(\mathcal{X}_A)_{f_A^+}|}\sum_{k=1}^{K}\frac{1}{f(x_A^{(k)})} \tag{11}$$

where

$$(\mathcal{X}_A)_{f_A^+} \triangleq \{x_A : f_A(x_A) \neq 0\}. \tag{12}$$

By symmetry, we also have an analogous estimate Γ_B. The computation of

$$f(x_A^{(k)}) = \sum_{x_B} f(x_A^{(k)}, x_B), \tag{13}$$

which is required in (11), is easy since the corresponding factor graph is a tree.

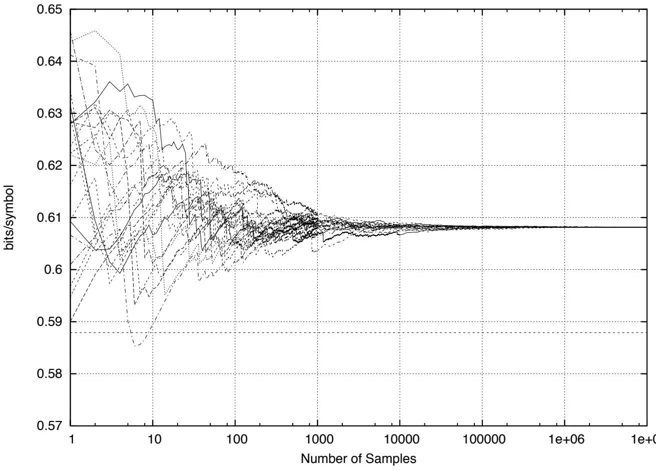

Fig. 3. Estimated capacity (in bits per symbol) vs. the number of samples K for a 10×10 grid with a $(1, \infty)$ constraint (Example 1). The plot shows 10 independent sample paths, each with two estimates, one from Γ_A and one from Γ_B.

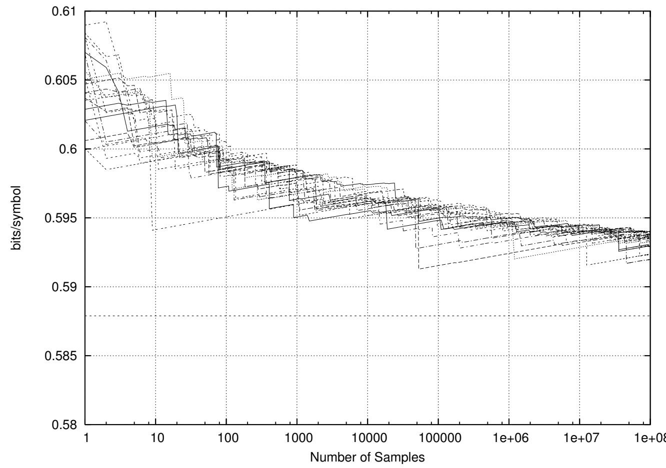

Fig. 4. Estimated capacity (in bits per symbol) vs. the number of samples K for a 60×60 grid with a $(1, \infty)$ constraint (Example 1). The plot shows 10 independent sample paths, each with two estimates, one from Γ_A and one from Γ_B.

The quantity $|(\mathcal{X}_A)_{f_A^+}|$ in (12) may be easy to determine even if f is not strictly positive. This applies, in particular, to Example 1 (and many similar examples) where

$$(\mathcal{X}_A)_{f_A^+} = \{x_A : f(x_A, 0) \neq 0\}. \tag{14}$$

In this case, $|(\mathcal{X}_A)_{f_A^+}| = \sum_{x_A} f(x_A, 0)$ is easily computed by sum-product message passing in the (cycle-free) factor graph of $f(x_A, 0)$.

C. A Happy Combination

It is now obvious to create the required samples $x_A^{(1)}$, $x_A^{(2)}$, $\ldots$, $x_A^{(K)}$ in (11) by means of tree-based Gibbs sampling as in Section III-A. The marginals (13) may then be obtained as a by-product of the tree-based sampling (see the appendix).

We thus obtain two estimates, Γ_A and Γ_B, as a by-product of tree-based Gibbs sampling with virtually no extra computations.

IV. Numerical Experiments

Some experimental results with the proposed method are shown in Figures 3 through 6. All figures refer to f as in Example 1 and show the quantity (the "capacity") $\frac{1}{N}\log_2(Z)$.

Figures 3 and 4 use a factor graph partition as in Fig. 2. In Fig. 3, we have $N = 10 \times 10$ and the estimated capacity is about 0.6082. In Fig. 4, we have $N = 60 \times 60$; for this size of grid there are issues with slow convergence.

To improve the convergence and to speed up the mixing, we can partition the factor graph (the extension of Fig. 1 to $N = 60 \times 60$) into "thicker" vertical strips. Such thick strips have cycles, but exact sum-product computation is still possible, e.g., by converting the strip into an equivalent cycle-free factor graph. The computation time is exponential in the thickness of the strip, but the faster mixing (as shown in Figures 5 and 6) results in a substantial reduction of total computation time for strips of moderate width.

From Fig. 6, the estimated capacity is about 0.5914.

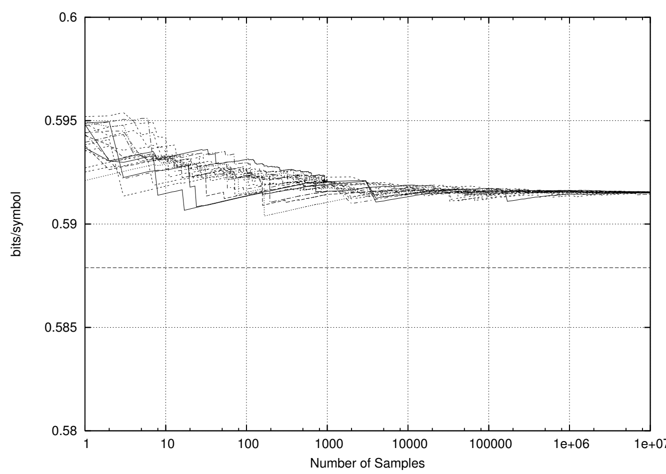

Fig. 5. Same conditions as in Fig. 4, but with strips of width two.

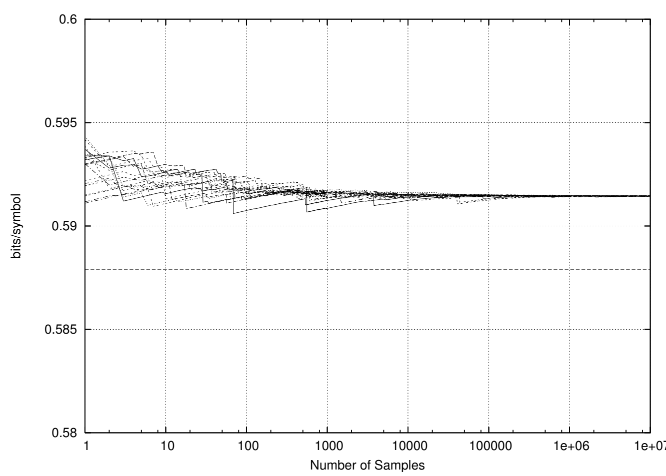

Fig. 6. Same conditions as in Fig. 4, but with strips of width three.

Also shown in the figures (as a horizontal dotted line) is the infinite-grid limit $\lim_{M\to\infty} \frac{1}{M^2}\log_2(Z) \approx 0.5879$, which is known for this simple example (see Section I).

All figures show the estimates from Γ_A and from Γ_B for several independent experiments.

V. Bounds for Infinite Grid

Let $C_M \triangleq \frac{1}{M^2}\log_2(Z)$ be the capacity of a constraint as in Example 1 for an $M \times M$ grid. It is clear (from tiling the whole plane with $M \times M$ squares) that $C_\infty \leq C_M$ for any finite M.

On the other hand, by tiling the plane with $M \times M$ squares separated by all-zero guard rows and all-zero guard columns, we obtain $C_\infty \geq C_M(\frac{M}{M+1})^2$.

In the example of Figures 4–6 (with $M = 60$), we thus obtain $0.5721 \leq C_\infty \leq 0.5914$.

VI. Concluding Remarks

We have shown that tree-based Gibbs sampling (as proposed by Hamze and de Freitas) can be used to compute an estimate of the partition function with virtually no extra computational cost. The proposed method can be used, in particular, to compute (a Monte Carlo estimate of) the capacity of noiseless constrained 2-D channels. Our preliminary numerical experiments are encouraging.

Appendix: Sampling from Markov Chains

We recall some pertinent facts about the simulation of Markov chains and cycle-free factor graphs. Let $p(x) = p(x_1,\ldots,x_n)$ be the probability mass function of a Markov chain. If $p(x)$ is given in the form

$$p(x) = p(x_1)\prod_{k=2}^{n} p(x_k|x_{k-1}), \tag{15}$$

then it is obvious how to create i.i.d. samples according to $p(x)$. Now consider the case where $p(x)$ is not given in the form (15), but in the more general form

$$p(x) \propto \prod_{k=2}^{n} g_k(x_{k-1}, x_k) \tag{16}$$

with general factors g_k. It is then still easy to create i.i.d. samples according to $p(x)$, which may be seen as follows. First, a probability mass function of the form (16) can be rewritten in the form (15) (which allows efficient simulation). Second, this reparameterization of $p(x)$ may be efficiently carried out by backward sum-product message passing, as will be detailed below. The resulting algorithm is know as "backward-filtering forward-sampling" (or, in a time-reversed version, as "forward-filtering backward-sampling") [15].

Specifically, let $\overleftarrow{\mu}_{X_k}$ be the backward sum-product message along the edge X_k in the factor graph of (16), as is illustrated in Fig. 7 (cf. [5]). We then have $\overleftarrow{\mu}_{X_n}(x_n) = 1$ and

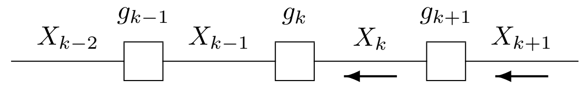

Fig. 7. Forney-style factor graph of (16) with messages $\overleftarrow{\mu}_{X_k}$ (17).

$$\overleftarrow{\mu}_{X_k}(x_k) \triangleq \sum_{x_{k+1}} g_{k+1}(x_k, x_{k+1})\overleftarrow{\mu}_{X_{k+1}}(x_{k+1}) \tag{17}$$

$$= \sum_{x_{k+1},\ldots,x_n} \prod_{m=k+1}^{n} g_m(x_{m-1}, x_m) \tag{18}$$

for $k = n-1, n-2, \ldots, 1$. Then

$$p(x_1) = \sum_{x_2,\ldots,x_n} p(x_1,\ldots,x_n) \tag{19}$$

$$\propto \overleftarrow{\mu}_{X_1}(x_1) \tag{20}$$

and

$$p(x_k|x_{k-1}) = \frac{g_k(x_{k-1}, x_k)\overleftarrow{\mu}_{X_k}(x_k)}{\overleftarrow{\mu}_{X_{k-1}}(x_{k-1})} \tag{21}$$

for $k = 2, \ldots, n$. The proof of (21) follows from noting that

$$p(x_{k-1}) = \gamma\overrightarrow{\mu}_{X_{k-1}}(x_{k-1})\overleftarrow{\mu}_{X_{k-1}}(x_{k-1}) \tag{22}$$

and

$$p(x_{k-1}, x_k) = \gamma\overrightarrow{\mu}_{X_{k-1}}(x_{k-1})g_k(x_{k-1}, x_k)\overleftarrow{\mu}_{X_k}(x_k) \tag{23}$$

where $\overrightarrow{\mu}_{X_{k-1}}$ is the forward sum-product message along the edge X_{k-1} and where γ is the missing scale factor in (16).

We also note that

$$\sum_{x_1} \overleftarrow{\mu}_{X_1}(x_1) = \sum_{x} g(x) \tag{24}$$

where $g(x)$ is defined as the right-hand side of (16). In this paper, this fact is used to compute the marginals (13) as a by-product of the sampling.

The generalization of all this to arbitrary factor graphs without cycles is straightforward.

Acknowledgement

The authors wish to thank David MacKay and Iain Murray for pointing out to us the work by Hamze and de Freitas [14].

References

[1] D. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3498–3508, August 2006.

[2] P. H. Siegel, "Information-theoretic limits of two-dimensional optical recording channels," in *Optical Data Storage 2006* (Proc. of SPIE, Vol. 6282, Eds. Ryuichi Katayama and Tuviah E. Schlesinger), Montreal, Quebec, Canada, April 23–26, 2006, pp. 62820W-1 – 2820W-13.

[3] O. Shental, N. Shental, S. Shamai (Shitz), I. Kanter, A. J. Weiss, and Y. Weiss, "Discrete-input two-dimensional Gaussian channels with memory: estimation and information rates via graphical models and statistical mechanics," *IEEE Trans. Inform. Theory,* vol. 54, pp. 1500–1513, April 2008.

[4] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inform. Theory,* vol. 47, pp. 498–519, Feb. 2001.

[5] H.-A. Loeliger, “An introduction to factor graphs,” *IEEE Signal Proc. Mag.,* Jan. 2004, pp. 28–41.

[6] K. Kato and K. Zeger, “On the capacity of two-dimensional run-length constrained channels,” *IEEE Trans. Inform. Theory,* vol. 45, pp. 1527–1540, July 1999.

[7] N. J. Calkin and H. S. Wilf, “The number of independent sets in a grid graph,” *SIAM J. Discr. Math.,* vol. 11, pp. 54–60, Feb. 1998.

[8] W. Weeks and R. E. Blahut, “The capacity and coding gain of certain checkerboard codes,” *IEEE Trans. Inform. Theory,* vol. 44, pp. 1193–1203, May 1998.

[9] G. Potamianos and J. Goutsias, “Stochastic approximation algorithms for partition function estimation of Gibbs random fields,” *IEEE Trans. Inform. Theory,* vol. 43, pp. 1984–1965, Nov. 1997.

[10] F. Huang and Y. Ogata, “Comparison of two methods for calculating the partition functions of various spatial statistical models,” *Australian & New Zealand Journal of Statistics,* vol. 43 (1), pp. 47–65, 2001.

[11] Y. Ogata and M. Tanemura, “Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure,” *Ann. Inst. Statist. Math.,* vol. 33, pp. 315–338, 1981.

[12] H.-A. Loeliger and M. Molkaraie, “Simulation-based estimation of the partition function and the information rate of two-dimensional models,” *Proc. 2008 IEEE Int. Symp. on Information Theory,* Toronto, Canada, July 6–11, 2008, pp. 1113–1117.

[13] D. J. C. MacKay, “Introduction to Monte Carlo methods,” in *Learning in Graphical Models,* M. I. Jordan, ed., Kluwer Academic Press, 1998, pp. 175–204.

[14] F. Hamze and N. de Freitas, “From fields to trees,” *Proc. Conf. on Uncertainty in Artificial Intelligence,* Banff, July 2004.

[15] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* 2nd ed., CRC Press, 2006.