2017

Fast spatial inference in the homogeneous Ising model

Alejandro Murua
Universite de Montreal

Ranjan Maitra
Iowa State University, maitra@iastate.edu

Fast spatial inference in the homogeneous Ising model

Abstract

The Ising model is important in statistical modeling and inference in many applications, however its normalizing constant, mean number of active vertices and mean spin interaction are intractable. We provide accurate approximations that make it possible to calculate these quantities numerically. Simulation studies indicate good performance when compared to Markov Chain Monte Carlo methods and at a tiny fraction of the time. The methodology is also used to perform Bayesian inference in a functional Magnetic Resonance Imaging activation detection experiment.

Keywords

finger-tapping, fMRI, hypergeometric distribution, path sampling, Swendsen-Wang algorithm, Stirling's approximation, Wang-Landau algorithm

Disciplines

Statistical Models | Statistics and Probability

Comments

This is a pre-print of the article Murua, Alejandro, and Ranjan Maitra. "Fast spatial inference in the homogeneous Ising model." *arXiv preprint arXiv:1712.02195* (2017). Posted with permission.

Fast spatial inference in the homogeneous Ising model

Alejandro Murua and Ranjan Maitra

The Ising model is important in statistical modeling and inference in many applications, however its normalizing constant, mean number of active vertices and mean spin interaction are intractable. We provide accurate approximations that make it possible to numerically calculate these quantities in the homogeneous case. Simulation studies indicate good performance when compared to Markov Chain Monte Carlo methods and at a tiny fraction of the time. The methodology is also used to perform Bayesian inference in a functional Magnetic Resonance Imaging activation detection experiment.

Index Terms

finger-tapping, fMRI, hypergeometric distribution, path sampling, Swendsen-Wang algorithm, Stirling's approximation, Wang-Landau algorithm

I. Introduction

A. The Ising model

Let $X = \{X_1, \ldots, X_n\}$ be n binary random variables with a conditional dependence structure specified via the neighborhood $\mathcal{N} = \{(i,j) : i \sim j, 1 \leq i < j \leq \ n\}$, where the notation $i \sim j$ means that variables X_i and X_j are neighbors. The celebrated Ising model of statistical physics specifies the joint probability mass function (p.m.f.) of X as

$$f_X(x; \alpha, \beta) = \mathrm{pr}(X = x) \propto \exp\big[\sum_i \alpha_i x_i + \sum_{i \sim j} \beta_{ij}\big\{x_i x_j + (1 - x_i)(1 - x_j)\big\}\big], \tag{I.1}$$

where it is assumed that $x_i \in \{0, 1\}$, $x = (x_1, \ldots, x_n)$, $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta = (\beta_{ij}; 1 \leq i < j \leq \ n)$. The parameter $\alpha_i \geq 0$ modulates the chance that $X_i = 1$ while the parameter β_{ij} specifies the strength of the interaction between X_i and X_j or equivalently, between the two neighboring sites i and j. The p.m.f. (I.1) has summation constant (also called partition function or phase transition structure) denoted by $Z(\alpha, \beta)$.

Model (I.1) was proposed by [1] to his student Ernst Ising as a way to characterize magnetic phase transitions or singularities in the partition function over a lattice graph. [2] published the model that bears his name and showed that in one dimension, that is, for a linear lattice graph, the phase transition structure is trivial with no singularities in the partition function. The distribution has applicability in disciplines beyond physics – indeed, one of its earliest uses in the statistical literature was as a prior model for a binary scene in image analysis [3]. Other applications include state-time disease surveillance [4], [5] and mapping [6]–[8]; modeling of protein hydrophobicity [9], genetic codon bias thermodynamics [10], DNA elasticity [11] or ion channel interaction [12] in statistical genetics; modeling of electrophysiological phenomena of the retina [13] and cortical recordings in neuroscience [14]–[16]; and modeling of biological evolution [17]. The Ising model has also been used to model voting patterns of senators in the US Congress [18] or

A. Murua is with the Department of Mathematics and Statistics at the Université de Montréal, CP 6128, succ. centre-ville, Montréal, Québec H3C 3J7, Canada.

R.Maitra is with the Department of Statistics, Iowa State University, Ames, Iowa, USA.

This research was supported in part by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health (NIH) under its Award No. R21EB016212, and by the Natural Sciences and Engineering Research Council of Canada (NSERC) through grant number 327689-06. The content of this paper however is solely the responsibility of the authors and does not represent the official views of either the NIBIB, the NIH or the NSERC.

behaviors on social networks [19], [20]. While many of these applications use a regular lattice structure, some (e.g. [18]–[20]) use more general non-lattice structures.

Parameter estimation in the Ising model is often challenging, especially in multi-dimensional lattices, and more generally for non-lattice conditional dependence graphs. This difficulty flows from the computational impracticality of obtaining exact closed-form expressions for the partition function in the presence of the interaction parameters β_{ij}s. Indeed, the computation of the partition function for such graphs with an external field has been shown to be NP-complete [21]. Even though the partition function is just a finite sum of exponential functions and can consequently be analytically expressed so that there is no phase transition for the finite graph Ising model, its computation is still intractable. Nevertheless, estimation of the parameters is needed in many applications, such as in performing posterior Bayesian inference in the functional Magnetic Resonance Imaging (fMRI) application of Section IV. In some cases, authors have eschewed parameter estimation in favor of approaches that are not always wholly satisfactory but obviate the need for an estimate of the partition function. For example, [7] would ideally have liked to have estimated the interaction parameter for assessing Medicare service area boundaries for competing hospice systems in Duluth, Minnesota, but they instead fixed the value for their study. Some authors have used empirical techniques such as pseudolikelihood (e.g., [3], [19], [20]), or written (I.1) in terms of an exponential family model and then used moment-matching or maximum entropy methods [14]–[16], [22]. Yet others [9]–[12], [23] have used a simple model structure, typically restricting to first order interactions, in order to employ a recursive algorithm to estimate the partition function [24] which is possible only with an Ising model with only nearest-neighbor (NN) structure (equivalently, first-order interactions). [25] used (I.1) in one dimension and with first-order neighborhood to signal if a probe (gene) is enriched or not, and specifically mentioned that they did not model more complex interactions because of the intractability of the partition function. Thus, there is need for a general method for estimating the partition function in several applications.

B. Background and previous work

Many diverse methods have been suggested to estimate the normalizing constant of intractable probability mass functions and densities. Some authors (e.g. [26]–[28]) have provided exact formulae for the partition function of the Ising model in a two-dimensional planar graph that also mostly assume that $\alpha_i = 0$ for all i in (I.1). Moreover, these calculations grow with the size of a graph and are not particularly applicable in the context of Bayesian inference when the partition function needs repeated evaluation. The popular method of path sampling [29]–[31] writes the normalizing constant as a function of an integral of an expectation. This expectation is then estimated by Monte Carlo or Markov Chain Monte Carlo (MCMC) sampling. For the case of the Ising model with $\alpha_i \equiv \alpha$ and isotropic dependence structure, that is $\beta_{ij} \equiv \beta$ for all $i, j \in \mathcal{N}$, we have

$$Z(\alpha,\beta) = \sum_{\{x_i\}} \exp\bigl\{\alpha \sum_i x_i - \beta \sum_{i\sim j} (1-\delta_{ij})\bigr\}, \tag{I.2}$$

where $\delta_{ij} = 1$ if $x_i = x_j$ and 0 otherwise. Let k be the degree or average number of neighbors of the graph. Then, writing $\alpha \sum_{i=1} x_i - \beta \sum_{i\sim j}(1-\delta_{ij}) = (\alpha - k\beta)\sum_{i=1}^n x_i + \beta \sum_{i,j} \eta_{ij} x_i x_j$, with $\eta_{ij} = 1$ if $i \sim j$ and 0 otherwise, we have $\partial/\partial\beta\{\log Z(\alpha,\beta)\} = Z^{-1}(\alpha,\beta)\partial/\partial\beta[Z(\alpha,\beta)] = E_{(\alpha,\beta)}\bigl(\sum_{i,j}\eta_{ij}x_i x_j\bigr) - kE_{(\alpha,\beta)}\bigl(\sum_{i=1} x_i\bigr)$. For fixed α, $\log\{Z(\alpha,\beta)/Z(\alpha,\beta_0)\} = \int_{\beta_0}^{\beta} E_{(\alpha,\beta')} \sum_{i\sim j}\delta_{ij} - m(\beta-\beta_0)$, with m the number of edges in the graph. For $\beta_0 = 0$, $Z(\alpha, 0) = \sum_{\{x_i\}} \exp(\alpha \sum_{i=1} x_i) = \sum_{\ell=0}^n \binom{n}{\ell} \exp(\alpha\ell) = [1+\exp(\alpha)]^n$. Therefore, by taking $\beta_0 = 0$, we get $\log Z(\alpha,\beta) = \int_{\beta_0}^{\beta} E_{(\alpha,\beta')}\bigl(\sum_{i\sim j}\delta_{ij}\bigr) + n\log(1+\exp(\alpha)) - m(\beta-\beta_0)$ with the integral term estimated via MCMC. Path sampling is usually implemented as a preprocessing step by evaluating $Z(\alpha,\beta)$ on a grid of parameter values.

A second approach estimates $Z(\alpha,\beta)$ together with the posterior distribution of the other parameters via a stochastic approach such as the Wang-Landau [32] or other flat-histogram [33], [34] algorithm. These algorithms replace $Z(\alpha, \beta)$ by a stochastic estimate $Z_j(\alpha, \beta)$ at each iteration and are usually applied to a finite grid of values of the parameters $\{(\alpha_r, \beta_r)\}_r$, but extensions [35] to the continuous analogue have also been developed. The sequence of estimates $\{Z_j(\alpha_r, \beta_r)\}_j$ is such that $Z_j(\alpha_r, \beta_r)$ converges, up to a constant, to $Z(\alpha_r, \beta_r)$, for all r, when the number of iterations $j \to +\infty$. Although originating in statistical physics [32], [36], the idea has lately also appeared in the statistical literature [35], [37], [38]. Other approaches also exist: for instance, [39] suggest using sequential Monte Carlo samplers to estimate the ratio of normalizing constants and using this estimate to perform Metropolis-Hastings sampling.

One major drawback of each of these stochastic and MCMC algorithms is that they need long sampling periods for both the burn-in and post-burn-in sampling phases. In small-scale problems, this may not be a major issue. However, in many cases, as in the showcase fMRI application of this paper, one may need to estimate the normalizing constant for several values of the parameters to the point of being computationally demanding. In this paper, we therefore investigate an approach to numerically approximate $Z(\alpha, \beta)$. Our approximations are derived in Section II and obviate the need for preprocessing to evaluate the constant on a grid before fitting a model such as is done in path sampling, or to run Wang-Landau for long iterations to ensure its convergence to the estimated value. We evaluate the performance of these approximations in Section III where our comparisons are against estimates obtained using MCMC and path sampling (for the normalizing constant). Section IV illustrates the utility and performance of these approximations in the context of using Bayesian methods for detecting activation in a fMRI motor task experiment. The paper concludes with some discussion.

II. Approximations in large isotropic Ising models

A. *The normalizing constant*

Our starting point is (I.1) under the assumption of isotropy (i.e., $\beta_{ij} \equiv \beta$ for all $i, j \in \mathcal{N}$). Using the same notation as (I.2), we can rewrite $p(x) = Z(\alpha, \beta)^{-1} \exp\{\alpha \sum_{i=1}^n x_i - \beta \sum_{i \sim j} (x_i - x_j)^2\}$. Let $G_{n,m} = (V, E)$ be the graph underlying the data, where m denotes the number of edges of the graph. The set of vertices is $V = \{x_1, \ldots, x_n\}$ and the set of edges is $E = \{(x_i, x_j) : i \sim j\}$. Let $\eta_{ij} = 1$ if $i \sim j$ and 0 otherwise. Since each x_i is either 0 or 1, $\sum_{i \sim j}(x_i - x_j)^2 = \sum_{i \sim j} x_i + x_j - 2x_i x_j = \frac{1}{2} \sum_{i,j=1}^n x_i \eta_{ij} + \frac{1}{2} \sum_{i,j=1}^n x_j \eta_{ij} - \sum_{i,j=1}^n x_i x_j \eta_{ij} = \sum_{i=1}^n x_i \sum_{j=1}^n \eta_{ij} - \sum_{i,j} x_i x_j \eta_{ij} = \sum_{i=1}^n k_i x_i - 2 \sum_{i \sim j} x_i x_j$, where k_i is the degree of vertex i, that is, the number of edges associated with the observation at the ith vertex, for $i = 1, \ldots, n$. In this paper, we suppose that $G_{n,m}$ is regular, that is, $k_i = k$ for all vertices. Consequently, the exponent in (I.2) can be rewritten as: $\alpha \sum_{i=1}^n x_i - \beta \sum_{i \sim j} (1 - \delta_{ij}) = \alpha' \sum_{i=1}^n x_i + \beta \sum_{i,j} \eta_{ij} x_i x_j$, where $\alpha' = \alpha - k\beta$. Further, for any fixed configuration of X, there exists an $\ell \geq 0$ so that $\sum_{i=1}^n x_i = \ell$. Let $M(\ell)$ be the set of sequences whose activation add up to ℓ. We may recast (I.2) as

$$Z(\alpha, \beta) == 1 + \exp(\alpha' n + \beta k n) + n \exp(\alpha') + \sum_{\ell=2}^{n-1} \exp(\alpha' \ell) \sum_{X \in M(\ell)} \exp\big(\beta \sum_{i,j} \eta_{ij} x_i x_j\big). \tag{II.1}$$

Our computations assume a graph $G_{n,m}$ with a total of $m = nk/2$ edges. For fixed ℓ, only ℓ observations, say $\tilde{X}_\ell = \{x_{i_1}, x_{i_2}, \ldots, x_{i_\ell}\}$, contribute to $\sum_{i,j} \eta_{ij} x_i x_j$, so that $\sum_{i,j} \eta_{ij} x_i x_j = \sum_{h=1}^{\ell} \sum_j \eta_{i_h,j} x_j$. The set $\tilde{X}_\ell$ can be thought of as a subgraph of the original graph, namely $G_\ell = (\tilde{X}_\ell, \tilde{E}_\ell)$ where the set of edges is a subset of all possible $\ell_2 = \ell(\ell-1)/2$ edges between vertices in $\tilde{X}_\ell$ that are present in $G_{n,m}$. In graph theory, the subgraphs G_ℓ are referred to as vertex-induced subgraphs. The vertex x_j contributes to the sum $\sum_{h=1}^{\ell} \sum_j \eta_{i_h,j} x_j$ only if x_j is a vertex of G_ℓ. This sum corresponds to twice the number of edges in G_ℓ. Computing the last sum in (II.1) corresponds to counting the number of subgraphs G_ℓ with a given number of edges. Our approximation of the partition function is based on the following

Proposition II.1. *Let $Y(s,\ell)$ be the number of vertex-induced subgraphs G_ℓ containing exactly s edges. The partition function can be written as*

$$Z(\alpha,\beta) = 1 + \exp(\alpha' n + \beta k n) + n\exp(\alpha') + \sum_{\ell=2}^{n-1} \binom{n}{\ell} \exp(\alpha'\ell)\mathcal{M}_{p(\cdot|\ell)}(2\beta),$$

where $\mathcal{M}_{p(\cdot|\ell)}(2\beta)$ is the moment generating function (m.g.f.) associated with the distribution $p(s|\ell) = Y(s,\ell)/\binom{n}{\ell}$, $s \in \{0,1,\dots,\ell k/2\}$, evaluated at $t = 2\beta$.

Proof. Write the last sum in (II.1) as

$$\sum_{X\in M(\ell)} \exp\bigl(\beta \sum_{i,j} \eta_{ij} x_i x_j\bigr) = \sum_{s=0}^{\ell_2} Y(s,\ell)\exp\bigl(2\beta s\bigr) = \binom{n}{\ell}\sum_{s=0}^{\ell_2}\exp\bigl(2\beta s\bigr)\,\frac{Y(s,\ell)}{\binom{n}{\ell}}.$$

The proportions $Y(s,\ell)/\binom{n}{\ell}$ define a distribution on s given $M(\ell)$. Because we suppose that the graph $G_{n,m}$ is regular, the support of $p(s|\ell)$ is over $\{0,1,\dots,\ell k/2\}$. □

Since s, the number of edges present in a given vertex-induced subgraph, is the sum of several sums, we may approximate the distribution of s given $M(\ell)$ for large ℓ by means of a Normal distribution, and replace $\mathcal{M}_{p(\cdot|\ell)}(\cdot)$ by the m.g.f. of a Normal distribution. We make this approach more formal below. For a given graph $G_{n,m}$, and s, $p(s|\ell)$ corresponds to the proportion of vertex-induced subgraphs with ℓ vertices and exactly s edges between the vertices. This is equivalent to computing the chance of obtaining a subgraph of ℓ vertices with exactly s edges, but this computation is straightforward. We may consider a related proportion, $p_e(s|\ell_2)$, of graphs with ℓ_2 "latent" edges, that contain exactly s edges of the original graph. It is easy to see that this proportion is given by the Hypergeometric distribution

$$p_e(s|\ell_2) = \frac{\binom{\ell_2}{s}\binom{n_2-\ell_2}{m-s}}{\binom{n_2}{m}} = \frac{\binom{m}{s}\binom{n_2-m}{\ell_2-s}}{\binom{n_2}{\ell_2}}, \qquad \text{for } s = s', s'+1, \dots, \min(\ell_2, m),$$

where $n_2 = \binom{n}{2}$ is the maximum number of edges for a graph with n vertices and $s' = \max(0, m-n_2+\ell_2)$. But these proportions are taken over a larger collection of graphs than $\{G_\ell\}$ so their distribution might differ significantly from the one we want to compute. Below, we see that although $p_e(s|\ell_2)$ has the right mean, its variance is much larger than that of $p(s|\ell)$.

Our approach consists of looking at the degree distribution of the subgraphs. Let $r_{\ell,h} = \sum_{j=1}^n \eta_{i_h,j} x_j$, for $h = 1,\dots,\ell$. These quantities are the observed degrees of the vertices $x_{i_1},\dots,x_{i_\ell}$. Let $r_\ell = \sum_{h=1}^\ell r_{\ell,h}$. Although we are just concerned with the distribution of the number of edges or sum of degrees, it is helpful to note that the quantities $r_{\ell,h}$ and r_ℓ are realizations from these distributions. That is, we may look at $r_{\ell,h}$ and r_ℓ as random variables. In particular, we are interested in finding their first two moments. The distribution of the number of vertex-induced subgraphs whose degrees add up to r has the form

$$p_d(r_\ell|\ell) = \sum_{r_{\ell,1}+r_{\ell,2}+\cdots r_{\ell,\ell}=r_\ell} p(r_{\ell,1}, r_{\ell,2},\dots,r_{\ell,\ell}).$$

Note that the support of this distribution lies over the even numbers $r_\ell = 2s$. The joint probability mass function $p(r_{\ell,1}, r_{\ell,2},\dots,r_{\ell,\ell})$ is also not straightforward to compute. However, the marginals are easily obtained for a regular graph with k edges for each vertex. In this case, the proportion of edges for a given vertex x_{i_h} in a subgraph of ℓ vertices has the hypergeometric distribution

$$p_{d,m}(r_{\ell,h}|\ell) = \frac{\binom{k}{r_{\ell,h}}\binom{n-1-k}{\ell-1-r_{\ell,h}}}{\binom{n-1}{\ell-1}}, \qquad r_{\ell,h} = r', r'+1, \dots, \min\{\ell-1, k\},$$

where $r' = \max(0, k-n+\ell)$. Therefore, the expectation of twice the number of edges is given by $\mu_\ell = E(r_\ell) = \ell E(r_{\ell,h}) = \ell(\ell-1)k/(n-1) = 2\ell_2\theta$, where $\theta = m/n_2 = k/(n-1)$ is the proportion of edges with respect to a complete graph. The variance depends on the dependency between the $r_{\ell,h}$s. We have the following

Proposition II.2. *Let* $y_{n,2} = (\ell-2)/(n-2)$, $\sigma_\ell^2 = 2\ell_2\theta(1-\theta)(1-y_{n,2})$, *and* $\rho_\ell = (\ell-1)(n-2k)/\{(n-2)(n-k-1)\}$. *We have* $\mathrm{Var}(r_\ell) = \sigma_\ell^2(1-\rho_\ell)$, *and* $\mathrm{Cov}(r_{\ell,t}, r_{\ell,h}) = -\sigma_\ell^2\rho_\ell/(2\ell_2) = \mathcal{O}(n^{-1})$. *In particular, for every* $\delta \in (0,1)$, *and* $\ell \leq n^\delta$, $\mathrm{Var}(r_\ell)/\sigma_\ell^2 \to 1$ *as* $n \to \infty$, *or equivalently* $\rho_\ell \to 0$ *as* $n \to \infty$, *uniformly on* ℓ.

Proof. See appendix. □

By Hoeffding's inequality [40, Section 6], each $r_{\ell,h}$ is concentrated about its mean $\mu_{\ell,1} = (\ell-1)\theta$, when n is large and ℓ is moderate to large. So we just need to study what happens with the distribution about its mean.

Proposition II.3. *Let* $y_{n,1} = (\ell-1)/(n-1)$, *and* $\sigma_{\ell,1}^2 = (\ell-1)(1-y_{n,1})\theta(1-\theta)$. *For* $\sqrt{n} \leq \ell \leq n - \sqrt{n}$,

$$p_{d,m}(r|\ell) = \left(2\pi\sigma_{\ell,1}^2\right)^{-1/2} \exp\left\{-\frac{(r-\mu_{\ell,1})^2}{2\sigma_{\ell,1}^2}\right\}\{1+\mathcal{O}(n^{-1/2})\}\{1+\omega(k,r)\},$$

with $\omega(k,r) = \mathcal{O}(1/(12r) + 1/\{12(k-r)\})$, *for all* $0 < r < k$ *such that* $|r - \mu_{\ell,1}| < \theta\sqrt{\ell-1}$.

Proof. See appendix. □

Proposition II.3 together with Hoeffding's inequality states that the variables $r_{\ell,h}$ behave like normal random variables with mean $\mu_{\ell,1}$, and variance $\sigma_{\ell,1}^2 \approx \sigma_\ell^2/\ell$, for values of $y_{n,1}$ far from 0 and 1. Moreover, because of Proposition II.2, these variables are weakly correlated, and hence, because of the near normality, nearly independent. This observation points to the use of the central limit theorem for large values of ℓ, to argue that $\mathcal{M}_{p_d(\cdot|\ell)}(\beta)$ should be well approximated by $\exp\{2\beta\ell_2\theta + \beta^2\ell_2\theta(1-y_{n,2})(1-\theta)(1-\rho_\ell)\}$. For small ℓ, the covariances are very small; indeed, the vertices are nearly independent. Hence, we can approximate $p_d(r)$ by

$$\tilde{p}_d(r|\ell) = \sum_{r_1+r_2+\cdots r_\ell = r} \prod_h p_{d,m}(r_h|\ell).$$

In this case, $\mathcal{M}_{\tilde{p}_d(\cdot|\ell)}(\beta) = \{\mathcal{M}_{p_{d,m}(\cdot|\ell)}(\beta)\}^\ell$, however we need to have $r = 2s$. That is, we need $\mathcal{M}_{\tilde{p}_d(even|\ell)}(\beta) = \sum_s \exp(2\beta s)\tilde{p}_d(2s|\ell)/\sum_s \tilde{p}_d(2s|\ell)$. Appendix B argues that we can approximate this latter quantity with $\mathcal{M}_{\tilde{p}_d(even|\ell)}(\beta) = \frac{2}{1+\cosh(\beta)}[\{\mathcal{M}_{p_{d,m}(\cdot|\ell)}(\beta)\}^\ell + \exp\{\beta(\ell k+1)\}\tilde{p}_d(\ell k|\ell)/2 + \exp(-\beta)\tilde{p}_d(0|\ell)/2]$.

1) The Normal approximations: This paper proposes an estimator of the partition function using the above results. Recall that when the m.g.f. of a sequence of random variables converges to the m.g.f. of a recognizable random variable, then the sequence of random variables converges in law, and hence weakly, to the recognizable random variable. Therefore, we may suppose that for large ℓ,

$$\sum_{s=s_\bullet}^{s^\bullet} p_d(r=2s|\ell)\exp(2\beta s) \approx \int_{2s_\bullet-1}^{2s^\bullet+1} \frac{1}{\sigma_\ell\sqrt{1-\rho_\ell}}\phi\left(\frac{x-\mu_\ell}{\sigma_\ell\sqrt{1-\rho_\ell}}\right)\exp(2\beta x)\,dx, \tag{II.2}$$

where $s_\bullet = \max(0, k-n+\ell)\,\ell/2$ and $s^\bullet = \min(\ell-1, k)\,\ell/2$. Set $w(\ell,s) = 2\ell_2\{(s+\frac{1}{2})/\ell_2 - \nu_\ell\}/(\sigma_\ell\sqrt{1-\rho_\ell})$, where $\nu_\ell = \theta + \beta\sigma_\ell^2(1-\rho_\ell)/(2\ell_2) = \theta\{1+\beta(1-\theta)(1-y_{n,2})(1-\rho_\ell)\}$. Let $\Delta_\Phi(\ell) = \Phi(w(\ell, s^\bullet+1/2)) - \Phi(w(\ell, s_\bullet-1/2))$, where $\Phi(\cdot)$ stands for the cumulative function of a standard normal distribution. A straightforward calculation shows that (II.2) simplifies to

$$\sum_{s=s_\bullet}^{s^\bullet} p_d(r=2s|\ell)\exp(2\beta s) \approx \exp(2\beta\theta\ell_2 + \beta^2\sigma_\ell^2(1-\rho_\ell)/2)\Delta_\Phi(\ell).$$

Next, setting $A_\phi(\alpha,\beta) = 1+\exp(\alpha' n+\beta kn)+n\exp(\alpha')$, we consider the function $g(\ell) = \alpha'\ell+2\beta\theta\ell_2\{1+(\beta/2)(1-y_{n,2})(1-\theta)(1-\rho_\ell)\}$, and define $\Sigma(h;\alpha,\beta) = \sum_{\ell=h}^{n-1}\binom{n}{\ell}\exp\{g(\ell)\}\Delta_\Phi(\ell)$. The partition function estimate $Z_{H.\phi}(\alpha,\beta)$, which we refer to as the Hyper-and-Normal edge proportion estimate is

$$Z_{H.\phi}(\alpha,\beta) = A_\phi(\alpha,\beta) + \sum_{\ell=2}^{\sqrt{n}-1}\binom{n}{\ell}\exp(\alpha'\ell)\mathcal{M}_{\tilde{p}_{d,m}(even|\ell)}(2\beta) + \Sigma(\sqrt{n};\alpha,\beta). \tag{II.3}$$

We observed in practice that $Z_{H.\phi}(\alpha,\beta)$ may be replaced by a simpler estimator, $Z_\phi(\alpha,\beta)$, based only on the Normal distribution without losing accuracy in the approximation to the partition function. This *normal edge proportion estimator* is

$$Z_\phi(\alpha,\beta) = A_\phi(\alpha,\beta) + \Sigma(2;\alpha,\beta). \tag{II.4}$$

The estimates $Z_{H.\phi}(\alpha,\beta)$ and $Z_\phi(\alpha,\beta)$ may be computed in $\mathcal{O}(n)$ operations. Although, this calculation is fast to compute, we investigate a further approximation to these estimates based on replacing the summation in the last terms of both (II.3) and (II.4) by an integral. Specifically, we regard the summation as a Riemann sum, and hence, as an approximation to the corresponding integral. The integral can be calculated as a new summation with number of terms much smaller than n, yielding a final approximation that can be computed much faster than $\mathcal{O}(n)$. In fact, the Euler-McLaurin formula [41] gives us

$$\Sigma(2;\alpha,\beta) \approx \int_2^{n-1}\Gamma(n+1)/\{\Gamma(x)\Gamma(n-x)\}\exp\{g(x)\}\Delta_\Phi(x)\,dx + B_\phi(\alpha,\beta),$$

with $B_\phi(\alpha,\beta) = (1/2)(n\exp\{g(n-1)\}\Delta_\Phi(n-1) + n_2\exp\{g(2)\}\Delta_\Phi(2))$. Using [42]'s approximation for the Gamma function, and writing $y = x/n = \ell/n$, in the above integral, we have $\Sigma(2;\alpha,\beta) \approx B_\phi(\alpha,\beta) + J_{n,m,1/2}(\alpha',\beta)$,where, for every $t\in\mathbb{R}$, we have the function

$$J_{n,m,t}(\alpha,\,\beta) = \sqrt{n/(2\pi)}\int_{2/n}^{1-1/n}(1-y)^{-n(1-y)-\frac{1}{2}}y^{-ny-t}\Delta_\Phi(ny)\exp\{g(ny)\}\,dx.$$

Let $h(y) = -(n(1-y)+0.5)\log(1-y) - (ny+0.5)\log(y)$. The derivative of the integrand in the above integral, that corresponds to the terms within the sum that this integral is approximating, is $d(y) = \exp\{h(y)+g(y)\}(\Delta_\Phi(ny)(h'(y)+g'(y)) + \Delta'_\Phi(ny))$. The integral approximation error is of order $\max_y|d(y)|/n$ which can be further shown to be of order $\max_y\exp(h(y)+g(y))$. This means that the relative error of the integral approximation is of order n^{-1} because most of terms in the sum are of the same order as $\max_y\exp(h(y)+g(y))$. We will denote the integral approximation to $Z_\phi(\alpha,\beta)$ by $\tilde{Z}_\phi(\alpha,\beta)$.

B. Mean number of active vertices

Having found approximations for $Z(\alpha,\beta)$, we now turn our attention to approximating $M = E(\hat{M}) \equiv E(\sum_{i=1}^n x_i)$, the expected number of active vertices. In addition to the setup in the previous section, let $\Delta_\phi(\ell) = \phi(w(\ell,s^\bullet+1/2)) - \phi(w(\ell,s_\bullet-1/2))$. Define $C_{\phi.M}(\alpha,\beta) = n_2(\exp\{g(n-1)\}\Delta_\Phi(n-1) + \exp\{g(2)\}\Delta_\Phi(2))$, where $\phi(\cdot)$ denotes the standard normal density. Using the same reasoning as before yields

$$M \approx M_\phi \doteq \frac{1}{Z_\phi}\left\{n\exp(\alpha' n+\beta kn) + n\exp(\alpha') + \sum_{\ell=2}^{n-1}\binom{n}{\ell}\ell\exp\{g(\ell)\}\Delta_\Phi(\ell)\right\}.$$

Using [42]'s approximation, the Euler-McLaurin expansion [41], and replacing ℓ by $n(\ell/n)$, yields an approximation of the series in the above by $C_{\phi.M}(\alpha,\beta) + nJ_{n,m,-1/2}(\alpha,\,\beta)$ to get

$$\tilde{M}_\phi = \frac{n}{\tilde{Z}_\phi}\left\{\,\exp\left(\alpha' n+\beta kn\right) + \exp(\alpha') + J_{n,m,-1/2}(\alpha,\beta) + C_{\phi.M}(\alpha,\beta)/n\right\}.$$

C. Mean spin interaction

We now turn our attention to approximating the expected number of matches of the Ising model or the mean spin interaction $S = E(\hat{S}) = E(\frac{1}{2}\sum_{i,j}\eta_{ij}x_ix_j)$. Let $Z = Z(\alpha,\beta)$. Proceeding as before, we get

$$
\begin{aligned}
S &= \frac{1}{Z}\sum_{\ell=0}^{n}\exp(\alpha'\ell)\sum_{X\in M(\ell)}\big(\frac{1}{2}\sum_{i,j}\eta_{ij}x_ix_j\big)\exp(\beta\sum_{i,j}\eta_{ij}x_ix_j)\\
&= \frac{m}{Z}\exp(\alpha' n+\beta kn) + \frac{1}{Z}\sum_{\ell=2}^{n-1}\binom{n}{\ell}\exp(\alpha'\ell)\sum_{s=s_\bullet}^{s^\bullet}p(s|\ell)\,s\exp(2\beta s).
\end{aligned}
$$

Using the weak convergence argument and similar reductions as for the case of the normalizing constant, the expectation of $s\exp(2\beta s)$ is approximately equal to

$$
\begin{aligned}
&\int_{2s_\bullet-1}^{2s^\bullet+1}\frac{1}{\sigma_\ell\sqrt{1-\rho_\ell}}\phi\left(\frac{x-\mu_\ell}{\sigma_\ell\sqrt{1-\rho_\ell}}\right)x\exp(2\beta x)\,dx\\
&= \frac{1}{2}\frac{d}{d\beta}\int_{2s_\bullet-1}^{2s^\bullet+1}\frac{1}{\sigma_\ell\sqrt{1-\rho_\ell}}\phi\left(\frac{x-\mu_\ell}{\sigma_\ell\sqrt{1-\rho_\ell}}\right)\exp(2\beta x)\,dx\\
&= \frac{1}{2}\exp(2\beta\theta\ell_2+\beta^2\sigma_\ell^2(1-\rho_\ell)/2)\Big[\{2\theta\ell_2+\beta\sigma_\ell^2(1-\rho_\ell)\}\Delta_\Phi(\ell)\\
&\qquad\qquad - \sigma_\ell\sqrt{1-\rho_\ell}\Delta_\phi(\ell)\Big].
\end{aligned}
$$

We thus have

$$
\begin{aligned}
E(\hat{S}) \approx S_\phi \doteq \frac{1}{Z_\phi}\bigg(&m\exp(\alpha' n+\beta kn) + \frac{1}{2}\sum_{\ell=2}^{n-1}\binom{n}{\ell}\Big[\{2\theta\ell_2+\beta\sigma_\ell^2(1-\rho_\ell)\}\Delta_\Phi(\ell)\\
&- \sigma_\ell\sqrt{1-\rho_\ell}\Delta_\phi(\ell)\Big]\exp(\alpha'\ell+2\beta\theta\ell_2+\beta^2\sigma_\ell^2(1-\rho_\ell)/2)\bigg). \quad \text{(II.5)}
\end{aligned}
$$

Let $\zeta_1(\ell) = \big\{1+\beta(1-\theta)(1-y_{n,2})(1-\rho_\ell)\big\}$, and $\zeta_2(\ell) = \sigma_\ell\sqrt{1-\rho_\ell}\Delta_\phi(\ell)$. Define

$$
\begin{aligned}
D_{\phi.M}(\alpha,\beta) = {} & (1/2)\{\zeta_1(2)\Delta_\Phi(2)-(2k)^{-1}(n-1)\zeta_2(2)\}\exp\{g(2)\}+\\
& (1/2)\{(n-2)\zeta_1(n-1)\Delta_\Phi(n-1)-k^{-1}\zeta_2(n-1)\}\exp\{g(n-1)\}.
\end{aligned}
$$

The expression in (II.5) can be further approximated in the same manner as before to obtain

$$
\begin{aligned}
S_\phi \approx \tilde{S}_\phi \doteq \frac{m}{\tilde{Z}_\phi}\bigg(&e^{\alpha' n+\beta kn} + D_{\phi.M}(\alpha,\beta)+\\
&\sqrt{\frac{n}{2\pi}}\int_{\frac{2}{n}}^{1-\frac{1}{n}}\Big[\zeta_1(ny)\Delta_\Phi(ny)-(kn)^{-1}y^{-2}\zeta_2(ny)\Big]\frac{\exp[\alpha' ny+\beta kny^2\zeta_1(ny)/2\}]}{(1-y)^{n(1-y)+\frac{1}{2}}y^{ny-\frac{3}{2}}}\,dy\bigg),
\end{aligned}
$$

upon replacing $kn/2$ by m, using [42]'s approximation to the binomial coefficients, and using the Euler-McLaurin approximation [41] for the sum.

III. Performance evaluations

We evaluated performance of the approximation formulae for the normalizing constant and the moments derived in Section II by comparing our analytical approximations with those obtained by simulation. The mean activation and spin interaction were estimated for each of a range of (α,β)-pairs using MCMC – these estimates were assumed to be the "gold standard" for our comparisons. However, since obtaining

TABLE I

DISCREPANCIES ASSOCIATED WITH THE APPROXIMATION OF THE LOGARITHM OF THE PARTITION FUNCTION. ALL DISCREPANCIES ARE COMPUTED AGAINST THE MCMC PATH SAMPLING ESTIMATES WHICH FORMS OUR "GOLD STANDARD" FOR COMPARISONS IN THESE EXPERIMENTS, EXCEPT FOR THE CASE $k = 1$, WHOSE DISCREPANCIES WERE COMPUTED USING THE KNOWN ASYMPTOTIC FORMULA FOR THE 1-NN GRAPH.

		Absolute and Relative Discrepancies					
		L_1		L_1/V_{MC}		R_1	
dimension	degree k	$\tilde{Z}_\phi$	Z_ϕ	$\tilde{Z}_\phi$	Z_ϕ	$\tilde{Z}_\phi$	Z_ϕ
116×152	2	26.72	26.73	0.0006	0.0006	0.009	0.009
116×152	4	270.48	270.55	0.006	0.006	0.032	0.032
116×152	8	437.97	438.00	0.010	0.010	0.047	0.047
116×152	24	373.57	373.59	0.008	0.008	0.044	0.044
64×64	2	6.20	6.20	0.0006	0.0006	0.009	0.009
64×64	4	62.86	62.93	0.006	0.006	0.032	0.032
64×64	8	100.81	100.81	0.010	0.010	0.047	0.047
64×64	24	88.24	88.24	0.009	0.009	0.044	0.044

MCMC simulation-based estimates for $Z(\alpha, \beta)$ is very difficult, we used path sampling [29], [30] as discussed in Section I-B to obtain its reference value.

Our approximation formulae apply to any regular graph but we restricted our attention to lattice graphs in this paper because of our particular interest in fMRI applications. Our simulation setup consisted of simulating realizations from Ising models on two lattice configurations and with three different neighborhood orders. (Because of edge effects, our lattice graphs are only approximately regular.) The two lattices had grids of sizes 116×152 and 64×64. The neighborhoods we chose for our simulations were of the first, second and fifth orders, corresponding to graphs of degree $k = 4, 8$ and 24, respectively. For each of the six combinations of grid sizes and graph degrees, we compared performance for 1,102 different pairs of values of the Ising parameters $(\alpha, \beta) \in [0, 5] \times [0.005, 10]$ (19 values for α, and 58 values for β). Note that there is no need to evaluate the approximations for negative α because $Z(-\alpha, \beta) = \exp(-\alpha n) Z(\alpha, \beta)$ for all pairs (α, β). (In particular, moments such as $E_{-\alpha,\beta}(\hat{M})$ can be easily obtained from $E_{\alpha,\beta}(\hat{M})$.) For each setting, we estimated the Ising moments and normalizing constant from samples obtained using the [43] algorithm with a burn-in period of 10,000 iterations and a sample size of 10,000 realizations from the post-burn-in iterations and used these estimates as the "gold standard" reference values. All calculations were done on a desktop computer with an 8-core Intel i7-3770 processor with clockspeed 3.40GHz with 32GB of RAM memory, and running Scientific Linux 6.4. For each moment estimate, we evaluated the performance of our approximations relative to the MCMC estimate by computing both the absolute value difference between the MCMC estimate m_{MC} and the analytical approximation given by the Normal edge proportion approximation m_N, and the relative absolute difference between these quantities $|m_{MC} - m_N|/m_{MC}$. The measures of absolute (L_1) and relative (R_1) discrepancy between all evaluations in the grid for (α, β) are given by the difference between the approximated and estimated surfaces $L_1(m_{MC}, m_N) = V^{-1} \int\int |m_{MC}(\alpha, \beta) - m_N(\alpha, \beta)| \, d\alpha \, d\beta$, and $R_1(m_{MC}, m_N) = V^{-1} \int\int |m_{MC}(\alpha, \beta) - m_N(\alpha, \beta)|/|m_{MC}(\alpha, \beta)| \, d\alpha \, d\beta$, with $V = \int\int d\alpha \, d\beta$. We also show the ratio of the absolute discrepancy to the mean volume of the region below the surface given by m_{MC}, $L_1(m_{MC}, m_N)/V_{MC}$, where $V_{MC} = (V^{-1} \int\int m_{MC} d\alpha \, d\beta)$. Table I shows the relative and absolute discrepancies between the analytical approximations of $\log Z(\alpha, \beta)$ and the path sampling estimates using the MCMC samples. The path sampling estimates were obtained using the estimate of the expected matches, that is $\log Z_{MC}(\alpha, \beta) = \int_0^\beta \mathrm{Mean}_{\alpha,b}(\sum_{i \sim j} \delta_{ij}^{(t)}) \, db + n \log\{1 + \exp(\alpha)\} - m\beta$, where $\delta_{ij}^{(t)}$ is the observed value of δ_{ij} in the tth sample generated by the Swendsen-Wang algorithm and, as before, m is the number of edges in the graph.

TABLE II
ABSOLUTE AND RELATIVE DISCREPANCIES BETWEEN MCMC ESTIMATES AND NUMERICAL APPROXIMATIONS.

			Absolute and Relative Discrepancies					
			L_1		L_1/V_{MC}		R_1	
Moment	dimension	degree k	$\tilde{M}_\phi$	M_ϕ	$\tilde{M}_\phi$	M_ϕ	$\tilde{M}_\phi$	M_ϕ
Vertex Activation $\hat{M}$	116×152	4	18.94	19.71	0.001	0.001	0.002	0.002
	116×152	8	8.42	8.81	0.0005	0.0005	0.0009	0.0009
	116×152	24	2.84	2.94	0.0002	0.0002	0.0003	0.0003
	64×64	4	4.14	4.32	0.001	0.001	0.002	0.002
	64×64	8	0.73	0.74	0.0002	0.0002	0.0003	0.0003
	64×64	24	0.25	0.25	0.00006	0.00006	0.0001	0.0001
			$\tilde{S}_\phi$	S_ϕ	$\tilde{S}_\phi$	S_ϕ	$\tilde{S}_\phi$	S_ϕ
Spin $\hat{S}$	116×152	4	39.18	41.47	0.001	0.001	0.002	0.002
	116×152	8	33.57	35.90	0.0005	0.0005	0.0009	0.001
	116×152	24	33.80	34.65	0.0002	0.0002	0.0003	0.0003
	64×64	4	9.27	9.35	0.001	0.001	0.002	0.002
	64×64	8	3.94	3.00	0.0003	0.0002	0.0004	0.0003
	64×64	24	3.85	2.66	0.00008	0.00006	0.0001	0.0001

Table I indicates that both the direct Normal edge proportion estimate and its counterpart that uses the Euler-McLaurin formula perform similarly. Thus there is almost no loss in accuracy when using the faster Euler-McLaurin based estimate. In order to further show the value of this simplified approximation, we also compared this analytical approximation with the theoretical asymptotic result for the logarithm of the normalizing constant for the 1-nearest-neighbor (1-NN) graph, that is, for $k = 2$. It can be shown that (see Appendix B)

$$\begin{aligned}\log Z(\alpha,\beta) \approx& (n/2)(\alpha-\beta)\\ &+ n\log\left[\exp(\beta/2)\cosh(\alpha/2) + \sqrt{\exp(\beta)\cosh^2(\alpha/2) - 2\sinh(\beta)}\right],\end{aligned} \tag{III.1}$$

after adapting the result of [44, Chapter 13, page 261] to the case of the {0,1}-statespace 1-NN Ising model. Performance evaluations for this case are also in Table I, with the results again indicating that the numerical approximation works very well even for the smallest possible value of k even though these approximations are based on moderate to large values of k.

Table II reports the values of the absolute value and relative discrepancies for the relevant moments (mean activation and spin interaction) of the Ising model. The results indicate good performance of our approximation formulae relative to the MCMC estimates. It is worth noting that the MCMC algorithm took about half a week to compute the 1,102 sets of moments for each combination of grid-size and graph degree combination, while our approximation formula took a total of less than six minutes to do the same calculations.

IV. APPLICATION TO ACTIVATION DETECTION IN FMRI EXPERIMENT

A. *Bayesian model for voxel activation*

We illustrate the use of our approximations in fully Bayesian inference for determining activation in a fMRI [45], [46] experiment. Our fMRI dataset is derived from images from the twelve replicated instances of a subject alternating between rest and also alternately tapping his right-hand and left-hand fingers [47]–[49]. For this illustration, we restrict attention only to the right-hand and the 20th slice, noting also that our derivations are general enough to extend to the other hand and the three-dimensional volume. Our data are in the form of p-values at each pixel that measure the significance of the positivity of the linear relationship between the pixelwise observed Blood-Oxygen-Level-Dependent (BOLD) time series response and the expected BOLD response obtained through a convolution of the input stimulus time-course with the Hemodynamic Response Function. Let $p_{r1}, p_{r2}, \ldots, p_{rn}$ be the observed p-values in the rth replication, where n is the number of pixels. Let $X_1, X_2, \ldots, X_n$ be indicator variables, with $X_i = 0$ or 1 depending on whether the ith pixel is truly active or not. Then, we may model p_{ri}, given the true state $X_i = x_i$ of the pixel as $f(p_{ri} \mid X_i = x_i) = \{U(p_{ri}; 0, 1)\}^{1-x_i} \{b(p_{ri}; a, b)\}^{x_i}$ where $U(p_{ri}; 0, 1)$ is the standard uniform density, and $b(p_{ri}; a, b)$ is the density of a Beta distribution with parameters (a, b), each evaluated at p_{ri}. To simplify the analysis, we reparametrize the $b(p_{ri}; a, b)$ parameters by $\psi = a + b$ and $\mu = a/(a + b)$. We assume a prior distribution on the X_i's in the form of (I.1) with homogeneous $\beta_{ij} \equiv \beta$ and a second-order neighborhood structure [50] of $k = 8$ neighbors for each interior pixel. We consider a standard uniform prior density for μ and a $\mathrm{Gamma}(\zeta, \theta)$ prior density for the parameter ψ. Specifically, the prior density for ψ is $\mathrm{gamma}(\psi; \zeta, \theta) \propto \psi^{\zeta-1} \exp(-\theta\psi)$. We assume uniform hyperprior densities for α and β. The posterior density of $\Theta = \{x_1, \ldots, x_n, \alpha, \beta, \psi, \mu\}$ is given by

$$\prod_{i=1}^{n}\prod_{r=1}^{R}\big\{b(p_{ri}; \mu\psi, \psi - \mu\psi)\big\}^{x_i} \frac{1}{Z(\alpha;\beta)} \exp\bigg\{ -\theta\psi + \alpha\sum_i x_i - \sum_{i\sim j}\beta(1-\delta_{ij})\bigg\} \psi^{\zeta-1}\mathbf{1}_{[0<\mu<1]},$$

where for any set S, $\mathbf{1}_S$ denotes the indicator function associated with S. Analytical inference being impractical to implement, we derive a Metropolis-Hastings algorithm to estimate the posterior densities of interest. Sampling from the above needs values of $Z(\alpha, \beta)$ for which approaches typically involve, among other strategies, offline estimation of the normalizing constant through tedious MCMC methods at some values and then interpolation at others (for example, [22]). Our approximations obviate the need for this offline approach and allow for the possibility of a direct approach. We now provide the MCMC framework for each parameter in Θ, after introducing additional notation for $\Theta_{-\eta} \equiv \Theta \setminus \{\eta\}$.

B. Posterior densities for MCMC simulations

1) **The parameter x_i's**: Let $x_{(-i)} = \{x_1, \ldots, x_n\} \setminus \{x_i\}$. We propose to update $x_i \in \{0, 1\}$ via Gibbs' sampling. The full conditional of the posterior distribution of x_i is given by

$$p(x_i \mid x_{(-i)}, \Theta_{-x}, p) \propto \Big\{\big(C_B\big)^R \prod_{r=1}^{R} p_{ri}^{\mu\psi}(1-p_{ri})^{(1-\mu)\psi}\Big\}^{x_i} \exp\{\alpha x_i - \beta\sum_{j\sim i}(1-\delta_{ij})\},$$

with $C_B = \Gamma(\psi)/\{\Gamma(\mu\psi)\Gamma((1-\mu)\psi)\}$. Let $A_i = \alpha + (\mu\psi - 1)\sum_{r=1}^{R}\log p_{ri} + \{(1-\mu)\psi - 1\}\sum_{r=1}^{R}\log(1-p_{ri}) + R\log C_B$. Then $\sum_{r=1}^{R}\log p_{ri}$ is the same as $R\log\check{p}_i$ where $\check{p}_i$ is the harmonic mean of $\{p_{ri} : r = 1, 2, \ldots, R\}$. Also, $\sum_{r=1}^{R}\log(1-p_{ri})/R \equiv \log\check{q}_i$ where $\check{q}_i$ is the harmonic mean of $\{1-p_{ri} : r = 1, 2, \ldots, R\}$. Then $A_i \equiv \alpha + (\mu\psi-1)R\log(\check{p}_i) + \{(1-\mu)\psi-1\}R\log(\check{q}_i) + R\log C_B$ and

$$\Pr(X_i = 1 | x_{(-i)}, \Theta_{-x}, p) \propto \frac{\exp(A_i - \sum_{j\sim i}\beta)}{\exp(A_i - \sum_{j\sim i}\beta) + \exp\{-2(\beta\sum_{j\sim i}x_j)\}}$$

2) **The parameter β:** We have $f(\beta|\Theta_{-\beta}, p) \propto \exp\{-\beta\sum_{i\sim j}(1-\delta_{ij})\}/Z(\alpha, \beta)$, where $Z(\alpha, \beta)$ is as in (I.2) and will be approximated numerically by $\tilde{Z}_\phi(\alpha, \beta)$ as defined in Section II-A1. We let

$\ddot{\beta}$ ∼Gamma(a, b) as our proposed update to β with $a/b = \beta$, and $a/b^2 = \gamma$, for some moderate $\gamma > 0$ (e.g., $\gamma = 1$). This yields the [51] acceptance ratio that is the minimum of 1 and

$$\frac{\tilde{Z}_\phi(\alpha,\beta)}{\tilde{Z}_\phi(\alpha,\ddot{\beta})}\frac{\Gamma(\beta^2/\gamma)}{\Gamma(\ddot{\beta}^2/\gamma)}\exp\big\{-(\ddot{\beta} - \beta)\sum_{i\sim j}(1 - \delta_{ij}) + \frac{1}{\gamma}(\ddot{\beta}^2 - \beta^2)\log(\beta\ddot{\beta}/\gamma)\big\}$$

which, upon applying [42]'s approximation for $\Gamma(z+1)$ simplifies to the minimum of 1 and

$$\frac{\tilde{Z}_\phi(\alpha,\beta)}{\tilde{Z}_\phi(\alpha,\ddot{\beta})}\frac{\ddot{\beta}}{\beta}\exp\big\{\frac{\beta^2+\ddot{\beta}^2}{\gamma}\log\frac{\beta}{\ddot{\beta}} + \frac{\ddot{\beta}^2-\beta^2}{\gamma} - (\ddot{\beta}-\beta)\sum_{i\sim j}(1-\delta_{ij})\big\},$$

3) **The parameter α:** The full conditional for α is given by

$$f(\alpha|\Theta_{-\alpha}, p) \propto \exp(\alpha\sum_{i=1}^{n} x_i)/Z(\alpha,\beta)$$

. Our proposed update to α is $\ddot{\alpha} \sim N(\alpha, \sigma_\alpha^2)$ with a moderate σ_α^2, (e.g., $\sigma_a^2 = 1$), which, after incorporating the approximation in Section II-A1. is accepted in a [52] step with probability given by $\min\Big\{1, \tilde{Z}_\phi(\alpha,\beta)\exp\{(\ddot{\alpha}-\alpha)\sum_{i=1}^n x_i\}/\tilde{Z}_\phi(\ddot{\alpha},\beta)\Big\}$.

4) **The parameter ψ:** The full conditional for ψ is given by

$$f(\psi|\Theta_{-\psi}, p) \propto \big[\Gamma(\psi)/\{\Gamma(\mu\psi)\Gamma((1-\mu)\psi)\}\big]^{R\sum_{i=1}^n x_i}\psi^{\zeta-1}\exp[\psi\{\mu\log A + (1-\mu)\log B - \theta\}],$$

where $A = \prod_{i=1}^n \breve{p}_i^{Rx_i}$, and $B = \prod_{i=1}^n \breve{q}_i^{Rx_i}$. Using [42]'s approximation to the factorial function, we have $\Gamma(\psi)/\{\Gamma(\mu\psi)\Gamma((1-\mu)\psi)\} \approx 1/\sqrt{2\pi}[\psi^{1/2}/\{\mu^{\mu\psi-0.5}(1-\mu)^{(1-\mu)\psi-0.5}\}]$. Therefore,

$$f(\psi|\Theta_{-\psi}, p) \approx\propto \psi^{\frac{1}{2}n_1+\zeta-1}\exp[-\{\theta + \mathrm{Ent}(\mu) - \mu\log A - (1-\mu)\log B\}\psi], \quad \text{(IV.1)}$$

where $n_1 = R\sum_{i=1}^n x_i$, and $\mathrm{Ent}(\mu) = \mu\log\mu + (1-\mu)\log(1-\mu)$ is the negative of the entropy associated with the probabilities $(\mu, 1-\mu)$. We update ψ with the proposal $\ddot{\psi}$ ∼Gamma$(\frac{1}{2}n_1+\zeta, \theta + \mathrm{Ent}(\mu) - \mu\log A - (1-\mu)\log B)$. More generally, we can reduce this updating step to a Gibbs' sampling step using the above approximated full conditional. The acceptance probability for $\ddot{\psi}$ is the minimum of 1 and

$$\left(\frac{\Gamma(\ddot{\psi})}{\Gamma(\mu\ddot{\psi})\Gamma((1-\mu)\ddot{\psi})}\frac{\Gamma(\mu\psi)\Gamma((1-\mu)\psi)}{\Gamma(\psi)}\right)^{R\sum_{i=1}^n x_i}$$
$$\times\left(\frac{\ddot{\psi}}{\psi}\right)^{\zeta-1}\exp\big\{(\ddot{\psi}-\psi)(\mu\log A + (1-\mu)\log B - \theta)\big\}$$
$$\times\left(\frac{\psi}{\ddot{\psi}}\right)^{\frac{1}{2}n_1+\zeta-1}\exp\big[-\{\theta - \mathrm{Ent}(\mu) - \mu\log A - (1-\mu)\log B\}(\psi-\ddot{\psi})\big],$$

which using (IV.1) yields the approximated acceptance ratio $\min\{1, \exp[(\psi-\ddot{\psi})n_1\{\mu\log\mu + (1-\mu)\log(1-\mu)\}]\}$, which is 1 whenever $\ddot{\psi} > \psi$ and so proposals $\ddot{\psi} > \psi$ are always accepted.

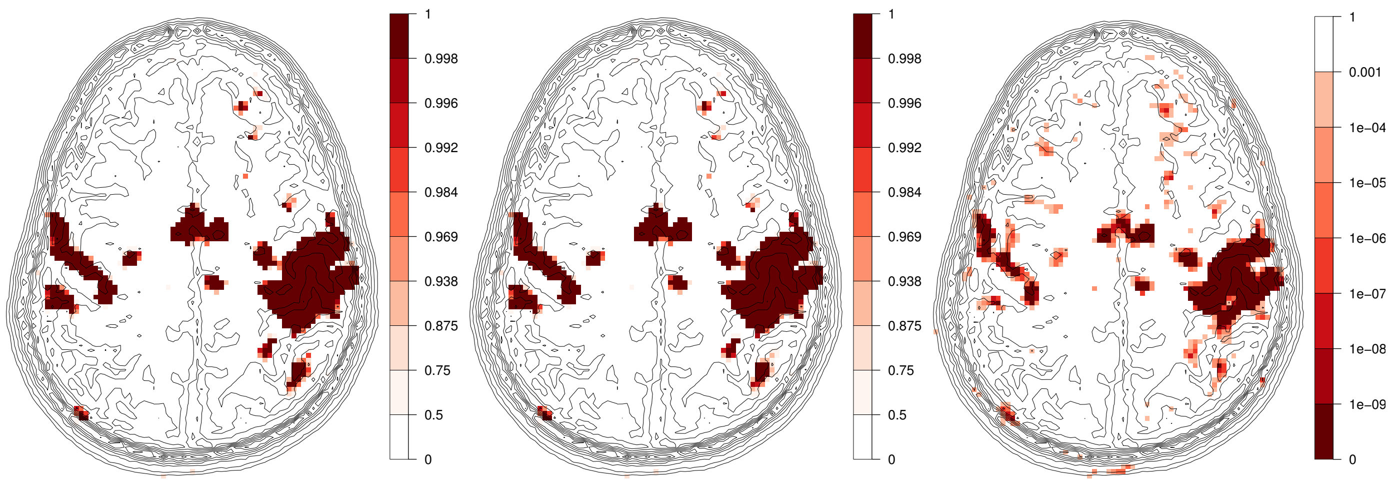
5) **The parameter μ:** Here, we use a random walk update $\ddot{\mu} \sim \mathrm{U}(0,1)$. This yields the Metropolis-Hastings acceptance ratio that is the minimum of 1 and

$$\left(\frac{\Gamma(\psi)}{\Gamma(\ddot{\mu}\psi)\Gamma((1-\ddot{\mu})\psi)}\frac{\Gamma(\mu\psi)\Gamma((1-\mu)\psi)}{\Gamma(\psi)}\right)^{n_1}$$
$$\times\exp\{\psi\ddot{\mu}\log A + \psi(1-\ddot{\mu})\log B - \psi\mu\log A - \psi(1-\mu)\log B\}$$
$$\approx \exp\big\{\psi(\ddot{\mu}-\mu)\log(A/B) + \psi n_1(\mathrm{Ent}(\mu) - \mathrm{Ent}(\ddot{\mu})) + \frac{n_1}{2}\log\left(\frac{\ddot{\mu}(1-\ddot{\mu})}{\mu(1-\mu)}\right)\big\},$$

where the last approximation follows by applying [42]'s approximation to the ratios of the $\Gamma()$ functions.

C. Results

The above posterior densities were used in the context of activation detection in the fMRI dataset. For this example, we set the hyperprior parameters $\theta = 1$ and $\zeta = 10$. This reflects our general a priori view that $\psi = a + b$ is large. We initialized our MCMC simulations with $(\alpha, \beta, \psi, \mu) = (0.001, 0.0025, 1, 0.001)$ and used collected a sample size of 10,000 realization after a burn-in period of the same number of iterations. The vertex values (for the autologistic variables) were updated (and initialized after a burn-in period of 3,000 iterations) with the [43] algorithm or the single-site updating given in point (1) of Section IV-B above. The posterior probability of activation, that is, the estimated posterior means of the x_is at each voxel, are displayed in Figure 1. As is customary in fMRI, the image is displayed using a radiological view. Thus, the right-hand side of the brain is imaged as the left-hand side. In the figure, we only display



(a) (b) (c)

Fig. 1. Leftmost and middle panels: the posterior probabilities of activation in the fMRI experiment using (a) Swendsen-Wang and (b) single-site updating. Rightmost panel (c): p-values after cluster-wise thresholding for p-values < 0.001. Displays are in radiological view in the log scale.

posterior probabilities that are greater than 0.5. It is clear that the posterior probabilities of activation using either Swendsen-Wang or single-site updating are essentially indistinguishable. For comparison, we have provided the results obtained upon using the commonly-used cluster-wise thresholding of the p-values of the test statistic. Here, activation regions are detected by drawing clusters of connected components each containing a pre-specified number of voxels with p-values below a specified threshold [53], [54]. To obtain our activation map, we choose a 2-D second-order neighborhood, a threshold of 0.001 for the p-values, following the recommendations of [55], and a minimum cluster size of 4 pixels, as optimally recommended by [56] using the AFNI software [57]–[59]. Although a detailed analysis of the results is beyond the purview of this paper, we note from the Bayesian model that there is very high posterior probability of activation in the left primary motor (M1) and pre-motor (pre-M1) cortices and the supplementary motor areas. There are also some areas on the right with high posterior probability of activation, perhaps as a consequence of the left-hand finger-tapping experiment that was also a part of the larger experiment. While the activation maps using cluster-wise thresholding are generally similar to those obtained using Bayesian inference, there are many stray pixels determined to be activated. Moreover, unlike cluster-wise thresholding, Bayesian methods provide us with the posterior probability of activation and this can be used in informing further decisions.

V. Discussion

In this paper, we provide explicit approximations for quantities derived from the homogeneous Ising model that are difficult to estimate. In particular, we develop approximations to the partition function that is otherwise intractable even for moderate numbers of observations. We showed that our approximation works well for very realistic lattice sizes and neighboring structures in the lattices. We stress that our approximations apply to general graphs, not necessarily lattices, and with general neighborhood structure. Indeed, its derivation does not use any special graph structure, and only supposes that the graph is regular, that is, that the degree k of each vertex is the same for the entire graph. Recall from the introduction that many researchers have had to simplify their models because of the difficulty in obtaining quantities such as the partition function of more realistic models. Further, unlike the restrictive special cases (such as those considered by [26]–[28]) where calculation of the exact partition function is possible, our approach does not grow with the size of the graph and can apply to all dimensions as long as the graph structure is regular. Therefore we expect that our approximation will facilitate the modeling of complex systems by allowing fast and reliable inference. An example of such a situation is a fully Bayesian approach to activation detection in fMRI, which we have demonstrated in Section IV can be done quite speedily using our approximations.

A few more comments are in order. The discrete Hypergeometric and Normal approximations of Section II can be computed in $\mathcal{O}(n)$ operations. However, further approximations obtained by replacing the sums by integrals may be computed much faster depending on the integration method used. We note that the analytical estimates for the mean activation and spin interaction work better for large k because the Normal approximation of the distribution of the mean of $r_h = \sum_j \eta_{i_h j} x_j$ is more suitable for large k.

A possible extension pertains to the case of approximations for the nonhomogeneous Ising model. We feel that it will be straightforward to generalize to a locally varying interaction parameter β_{ij} if we could somehow decompose the graph into regular components (homogeneous regions with approximately constant degree) where the β_{ij}s do not change much. In this case, the error in the approximation would probably depend on the size of each component, say n_c which are necessarily assumed to be large for our potential approximations to hold.

We close this section with one last comment on the fMRI application. One aspect that has so far not been invoked in fMRI is the fact that it is known that only a very small proportion of about 0.5-2 % of voxels are activated in a typical fMRI study [45]. However, this information has never been incorporated satisfactorily in the context of Bayesian activation detection of fMRI. Our approximations in Section II-B make it possible to perform Bayesian inference while constraining the prior parameters (α, β) so that the a priori proportion of expected activated voxels is satisfied. A thorough development and implementation of methodology for this application is beyond the scope of this paper, but such an approach would be of great practical interest for reliably assessing cognition. Thus, we see that while we have addressed an important problem in this paper, there remain issues meriting further attention.

APPENDIX

A. *Proof of Proposition 1*

We have $\mathrm{Var}(\sum_{h=1}^{\ell} r_{\ell,h}) = \sum_{h=1}^{\ell} (\ell-1)\theta(1-\theta)(1-y_{n,2}) + 2\sum_{t<h} \mathrm{Cov}(r_{\ell,t}, r_{\ell,h}) = \sigma_\ell^2 + 2\sum_{t<h} \mathrm{Cov}(r_{\ell,t}, r_{\ell,h})$, with $y_{n,2} = (\ell-2)/(n-2)$, and $\sigma_\ell^2 = 2\ell_2\theta(1-\theta)(1-y_{n,2})$. Let k_{th} be the number of neighbors in common between vertices t and h. In the notation that follows, the conditional expectation given η, means that the values of the couples (t,h) are fixed. Further,

$$
\begin{aligned}
&\mathrm{Cov}(r_{\ell,t}, r_{\ell,h}|\eta) \\
&\quad = \sum_{i=1}^{n}\sum_{j=1}^{n} \eta_{hi}\eta_{tj}\,\mathrm{Cov}(x_i, x_j) \\
&\quad = \frac{1}{(n-1)^2}\left\{\sum_{i=1}^{n} \eta_{hi}\eta_{ti}(\ell-1)(n-\ell) - 2\sum_{i<j} \eta_{hi}\eta_{tj}(\ell-1)(n-\ell)(n-2)\right\} \\
&\quad = \frac{(\ell-1)(n-\ell)}{(n-1)^2}\left(\sum_{i=1}^{n} \eta_{hi}\eta_{ti} - 2\sum_{i<j} \eta_{hi}\eta_{tj}/(n-2)\right)
\end{aligned}
$$

Note that $E(k_{th}) = nk(k-1)/\{(n-1)(n-2)\} = n\theta(k-1)/(n-2)$. Therefore

$$\begin{aligned}\operatorname{Cov}(r_{\ell,t}, r_{\ell,h}) &= \frac{(\ell-1)(n-\ell)}{(n-1)^2}\left\{\frac{n-1}{n-2}n\theta\frac{k-1}{n-2} - \frac{k^2}{n-2}\right\} \\ &= -\frac{(\ell-1)(n-\ell)}{(n-1)(n-2)}\frac{n-2k}{n-2}\theta.\end{aligned} \tag{A.1}$$

These covariances tend to zero uniformly on ℓ as $n \to +\infty$. For the variance, we have

$$\operatorname{Var}(r_\ell) = \sigma_\ell^2 - \frac{2\ell_2(\ell-1)(n-\ell)}{(n-1)(n-2)}\frac{n-2k}{n-2}\theta = \sigma_\ell^2\left\{1 - \frac{(\ell-1)(n-2k)}{(n-2)(n-k-1)}\right\}.$$

B. Proof of Proposition 2

For every pair of positive integers $h < g$, Stirling's formula [42], [60] yields

$$\binom{g}{h} = \frac{\sqrt{2\pi g}g^g}{2\pi\sqrt{h(g-h)}h^h(g-h)^{g-h}} \times \left\{1 + \mathcal{O}\left(\frac{1}{12h} + \frac{1}{12(g-h)}\right)\right\} \tag{A.2}$$

$$= \frac{1}{\sqrt{2\pi}}\frac{g^{g+1/2}}{(g-h)^{g-h+\frac{1}{2}}h^{h+\frac{1}{2}}}\{1 + \omega(g,h)\} \tag{A.3}$$

$$= \frac{1}{\sqrt{2\pi g}}\left(1 - \frac{h}{g}\right)^{-g+h-\frac{1}{2}}\left(\frac{h}{g}\right)^{-h-\frac{1}{2}}\{1 + \omega(g,h)\}, \tag{A.4}$$

where $\omega(g,h)$ denotes the error in Stirling's approximation. Let $p = (\ell-1)/(n-1)$. let $n_1 = n-1$, and $\bar\theta = 1-\theta$. Consider the first binomial term in the definition of $p_{d,m}(\cdot|\ell)$ Following the proof of the de-Moivre-Laplace's normal approximation to the binomial distribution [61], write $r = n_1p\theta + x\sqrt{n_1p\theta\bar\theta}$. Using (A.4), we have, for values of $r > 0$ and $r < k$,

$$\begin{aligned}\binom{k}{r} &= \binom{n_1\theta}{n_1p\theta + x\sqrt{n_1p\theta\bar\theta}} \\ &= (2\pi n_1p\theta(1-p))^{-1/2}\,p^{-n_1p\theta - x\sqrt{n_1p\theta\bar\theta}}(1-p)^{-n_1(1-p)\theta + x\sqrt{n_1p\theta\bar\theta}} \\ &\quad\times\{1+\omega(k,r)\}\big(1 + (x/p)\sqrt{p\bar\theta/(n_1\theta)}\big)^{-n_1p\theta - x\sqrt{n_1p\theta\bar\theta}} \\ &\quad\times\big(1 - \{x/(1-p)\}\sqrt{p\bar\theta/(n_1\theta)}\big)^{-n_1(1-p)\theta + x\sqrt{n_1p\theta\bar\theta}}\{1+\omega(k,r)\}.\end{aligned}$$

For $|z| < 1$, we have $\log(1-z) = -z - z^2/2 + O(z^3)$ and $\log(1+z) = z - z^2/2 + O(z^3)$. With these identities, we have

$$-n_1\theta(1-p)\log\left\{1 - \frac{x}{1-p}\sqrt{\frac{p\bar\theta}{n_1\theta}}\right\} = x\sqrt{n_1p\theta\bar\theta} + \frac{1}{2}\frac{px^2\bar\theta}{1-p} + R_1(x),$$

and

$$x\sqrt{n_1p\theta\bar\theta}\log\big\{1 - x\sqrt{p\bar\theta/(n_1\theta)}/(1-p)\big\} = -x^2p\bar\theta/(1-p) + R_2(x),$$

where the errors $R_1(x)$ and $R_2(x)$ are of order $\epsilon_1(x) = \mathcal{O}(x^3k^{-1/2}p^{3/2}(1-p)^{-2})$. That is,

$$\begin{aligned}\big(-n_1\theta(1-p) + x\sqrt{n_1p\theta\bar\theta}\big)\log\big(1 - \frac{x}{1-p}\sqrt{p\bar\theta/(n_1\theta)}\big) \\ = x\sqrt{n_1p\theta\bar\theta} - \frac{x^2}{2(1-p)}p\bar\theta + \epsilon_1(x).\end{aligned} \tag{A.5}$$

Similarly, $-n_1p\theta\log\{1+x\sqrt{\bar\theta/(n_1p\theta)}\} = -x\sqrt{n_1p\theta\bar\theta}+\frac{1}{2}x^2\bar\theta+R_3(x)$, and $-x\sqrt{n_1p\theta\bar\theta}\log\{1+x\sqrt{\bar\theta/(n_1p\theta)}\} = -x^2\bar\theta + R_4(x)$, where the errors $R_3(x)$ and $R_4(x)$ are of order $\epsilon_2(x) = \mathcal{O}(x^3k^{-1/2}p^{-1/2})$. Therefore,

$$\left(-n_1p\theta - x\sqrt{n_1p\theta\bar\theta}\right)\log\left\{1+\frac{x}{p}\sqrt{p\bar\theta/(n_1\theta)}\right\} = -x\sqrt{n_1p\theta\bar\theta} - \frac{x^2}{2}\bar\theta + \epsilon_2(x). \quad \text{(A.6)}$$

Combining (A.5) and (A.6) gives

$$\binom{k}{r} \approx (2\pi n_1p\theta(1-p))^{-1/2}p^{-n_1p\theta-x\sqrt{n_1p\theta\bar\theta}}(1-p)^{-n_1(1-p)\theta+x\sqrt{n_1p\theta\bar\theta}} \times \exp\left\{-\frac{x^2}{2(1-p)}\bar\theta\right\}.$$

Similarly we get

$$\begin{aligned}\binom{n-1-k}{\ell-1-r} &= \binom{n_1\bar\theta}{n_1p\bar\theta - x\sqrt{n_1p\theta\bar\theta}}\\ &\approx \{2\pi n_1p\bar\theta(1-p)\}^{-\frac{1}{2}}p^{-n_1p\bar\theta+x\sqrt{n_1p\theta\bar\theta}}\\ &\quad\times(1-p)^{-n_1(1-p)\bar\theta-x\sqrt{n_1p\theta\bar\theta}}\exp\left(-\frac{x^2}{2(1-p)}\theta\right),\end{aligned}$$

$$\text{and} \qquad \binom{n-1}{\ell-1} \approx (2\pi n_1p(1-p))^{-1/2}(1-p)^{-n_1(1-p)}p^{-n_1p}.$$

Further,

$$\begin{aligned}p_{d,m}(r|\ell) &\approx \{2\pi n_1p(1-p)\theta\bar\theta\}^{-1/2}\exp[-x^2/\{2(1-p)\}]\\ &= \{2\pi n_1p(1-p)\theta\bar\theta\}^{-1/2}\exp[-(r-n_1p\theta)^2/\{2n_1p(1-p)\theta\bar\theta\}].\end{aligned}$$

with the approximation being valid provided that $\{|x|/(1-p)\}\sqrt{p\bar\theta/k} < 1$ and $(|x|/p)\sqrt{p\bar\theta/k} < 1$. That is, $|x| < \sqrt{k/\bar\theta}\min\{\sqrt{p},(1-p)/\sqrt{p}\}$. If $\ell \in [an^\delta, n-an^\delta]$ for some constants $a>0$, and $0<\delta<1$, then $\min_p\min\{\sqrt{p},(1-p)/\sqrt{p}\} \geq (an^\delta-1)/(n-1) = \mathcal{O}(n^{\delta-1})$. In this case, $\epsilon_1(x) \leq (k/\bar\theta)^{3/2}k^{-1/2}an^{\delta-1} = \mathcal{O}(n^{\delta-1})$, and

$$\epsilon_2(x) \leq (k/\bar\theta)^{3/2}k^{-1/2}a^3n^{3(\delta-1)}\sqrt{(n-1)/(an^\delta-1)} = \mathcal{O}(n^{(5/2)(\delta-1)}).$$

For $\delta = 1/2$, $|x| < \mathcal{O}(\sqrt{\theta/\bar\theta})$, and $\epsilon_i(x) \leq \mathcal{O}(1/\sqrt{n})$, $i=1,2$.

Let $\mathcal{M}_{\tilde p_d(even|\ell)}(\beta) = \sum_s \exp(2\beta s)\tilde p_d(2s|\ell)/\sum_s\tilde p_d(2s|\ell)$. We may assume that $\sum_s\tilde p_d(2s|\ell) = 1/2$. We have

$$\begin{aligned}\mathcal{M}_{\tilde p_d(\cdot|\ell)}(\beta) &= \sum_{r=0}^{\ell k}\exp(\beta r)\tilde p_d(r|\ell)\\ &= \sum_{s=0}^{\ell k/2}\exp(\beta 2s)\tilde p_d(2s|\ell) + \sum_{s=0}^{\ell k/2-1}\exp\{\beta(2s+1)\}\tilde p_d(2s+1|\ell).\end{aligned}$$

But $\sum_s \exp\{\beta(2s+1)\}\tilde{p}_d(2s+1|\ell) = \exp(\beta)\sum_s \exp(2\beta s)\tilde{p}_d(2s+1|\ell)$ which is approximately equal to

$$\begin{aligned}
&\frac{1}{2}\exp(\beta)\sum_s \exp(2\beta s)\{\tilde{p}_d(2s|\ell)+\tilde{p}_d(2(s+1)|\ell)\} \\
&= \frac{1}{2}\exp(\beta)\left[\sum_{s=0}^{\ell k/2-1}\exp(2\beta s)\tilde{p}_d(2s|\ell)\right. \\
&\qquad\left. + \exp(-2\beta)\sum_{s=0}^{\ell k/2-1}\exp\{2\beta(s+1)\}\tilde{p}_d(2(s+1)|\ell)\right] \\
&= \frac{1}{2}\exp(\beta)\left[\{1+\exp(-2\beta)\}\sum_{s=0}^{\ell k/2}\exp(2\beta s)\tilde{p}_d(2s|\ell)\right. \\
&\qquad\left. - (\tilde{p}_d(\ell k|\ell)\exp(\beta\ell k) + \tilde{p}_d(0|\ell)\exp(-2\beta))\right] \\
&= \cosh(\beta)\sum_s \exp(2\beta s)\tilde{p}_d(2s|\ell) - \frac{1}{2}\tilde{p}_d(\ell k|\ell)\exp\{\beta(\ell k+1)\} - \frac{1}{2}\exp(-\beta)\tilde{p}_d(0|\ell).
\end{aligned}$$

Therefore, we estimate $\mathcal{M}_{p_d(even|\ell)}(\beta)$ by the quantity

$$\begin{aligned}
&\mathcal{M}_{\tilde{p}_d(even|\ell)}(\beta) \\
&= \frac{2}{1+\cosh(\beta)}\left\{\mathcal{M}_{\tilde{p}_d(\cdot|\ell)}(\beta) + \frac{\exp\{\beta(\ell k+1)\}}{2}\tilde{p}_d(\ell k|\ell) + \frac{\exp(-\beta)}{2}\tilde{p}_d(0|\ell)\right\} \\
&= \frac{2}{1+\cosh(\beta)}\left[\{\mathcal{M}_{p_{d,m}(\cdot|\ell)}(\beta)\}^\ell + \frac{\exp\{\beta(\ell k+1)\}}{2}\tilde{p}_d(\ell k|\ell) + \frac{\exp(-\beta)}{2}\tilde{p}_d(0|\ell)\right].
\end{aligned}$$

Consider the circular 1-NN Ising model with variables $\{v_1,\ldots,v_n\} \subset \{-1,1\}$ (in this case, v_n is a neighbor of v_1). The corresponding normalizing constant is given by $Z_1(L,K) = \sum_{\{v\}}\exp\{L\sum_i v_i + K(v_n v_1 + \sum_{i=1}^{n-1} v_i v_{i+1})\}$. A well-established result [44, Chapter 13, p. 261] said that for large n,

$$\log Z_1(L,K) \approx n\log\left[\exp(K)\cosh(L) + \sqrt{\exp(2K)\cosh^2(L) - 2\sinh(2K)}\right]. \quad \text{(A.7)}$$

In our setup, the variables $x_i \in \{0,1\}$. So we need to transform the variables v_i to x_i. It is easy to see that the corresponding transformation is $x_i = (1+v_i)/2$. so that

$$\begin{aligned}
&Z_1(L,K) \\
&= \sum_{\{x\}}\exp[L\sum_i(2x_i-1) + K\{(2x_n-1)(2x_1-1) + \sum_{i=1}^{n-1}(2x_i-1)(2x_{i+1}-1)\}] \\
&= \sum_{\{x\}}\exp\{2(L-2K)\sum_i x_i + 4K(x_n x_1 + \sum_{i=1}^{n-1} x_i x_{i+1})\}\exp\{n(K-L)\}.
\end{aligned}$$

From here, we get that $Z_1(K,L) = Z(2(L-2K),2K)\exp\{n(K-L)\}$, where the $4K$ is replaced by $2K$ because in our model the sum of over the neighboring vertices is multiplied by two. Therefore, using the fact that the one-nearest-neighbor graph is a regular graph with $k=2$, we get $K = \beta/2$, and $L = \alpha/2$. Consequently, $Z(\alpha,\beta) = \exp\{(\alpha-\beta)(n/2)\}Z_1(\alpha/2,\beta/2)$. Using (A.7), we obtain for large n,

$$\begin{aligned}
&\log Z(\alpha,\beta) \\
&\approx \frac{n(\alpha-\beta)}{2} + n\log\left[\exp\left(\frac{\beta}{2}\right)\cosh(\alpha/2) + \sqrt{\exp(\beta)\cosh^2\left(\frac{\alpha}{2}\right) - 2\sinh(\beta)}\right].
\end{aligned}$$

REFERENCES

[1] W. Lenz, "Beiträge zum verständnis der magnetischen eigenschaften in festen körpern," *Physikalische Zeitschrift*, vol. 21, pp. 613–615, 1920.

[2] E. Ising, "Beitrag zur theorie des ferromagnetismus," *Z. Physik*, vol. 31, pp. 253–258, 1925.

[3] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 259–302, 1986.

[4] C. Robertson, T. A. Nelson, Y. C. MacNab, and A. B. Lawson, "Review of methods for spacetime disease surveillance," *Spatial and Spatio-temporal Epidemiology*, vol. 1, pp. 105–116, 2010.

[5] E. Järpe, "Surveillance of the interaction parameter of the Ising model," *Communications in Statistics - Theory and Methods*, vol. 28, no. 12, pp. 3009–3027, 1999.

[6] D. Lee and R. Mitchell, "Boundary detection in disease mapping studies," *Biostatistics*, vol. 13, no. 3, pp. 415–426, 2012.

[7] H. Ma and B. Carlin, "Bayesian multivariate areal wombling for multiple disease boundary analysis," *Bayesian Analysis*, vol. 2, pp. 281–302, 2007.

[8] H. Ma, B. P. Carlin, and S. Banerjee, "Hierarchical and joint site-edge methods for Medicare hospice service region boundary analysis," *Biometrics*, vol. 66, no. 2, pp. 355–364, 2010.

[9] A. Irback, C. Peterson, and F. Potthast, "Evidence for nonrandom hydrophobicity structures in protein chains," *Proceedings of the National Academy of Sciences USA*, vol. 93, pp. 9533–9538, 1996.

[10] G. Rowe and L. Trainor, "A thermodynamic theory of codon bias in viral genes," *Journal of Theoretical Biology*, vol. 101, pp. 171–203, 1983.

[11] A. Ahsan, J. Rudnick, and R. Bruinsma, "Elasticity theory of the B-DNA to S-DNA transition," *Biophysical Journal*, vol. 74, pp. 132–137, 1998.

[12] Y. Liu and J. Dilger, "Application of the one- and two-dimensional Ising models to studies of cooperativity between ion channels," *Biophysical Journal*, vol. 64, pp. 26–35, 1993.

[13] E. Schneidman, M. Berry, R. Segev, and W. Bialek, "Weak pairwise correlations imply strongly correlated network states in a neural population," *Nature*, vol. 440, pp. 1007–1012, 2006.

[14] S. Yu, D. Huang, W. Singer, and D. Nikolic, "A small world of neuronal synchrony," *Cerebral Cortex*, vol. 18, pp. 2891–2901, 2008.

[15] L. Hamilton, J. Sohl-Dickstein, A. Huth, V. Carels, K. Deisseroth, and S. Bao, "Optogenetic activation of an inhibitory network enhances feedforward functional connectivity in auditory cortex," *Neuron*, vol. 80, no. 4, p. 10661076, 2013.

[16] E. Ganmor, R. Segev, and E. Schneidman, "Sparse low-order interaction network underlies a highly correlated and learnable neural population code," *Proceedings of the National Academy of Sciences USA*, vol. 108, pp. 9679–9684, 2011.

[17] E. Baake, M. Baake, and H. Wagner, "Ising quantum chain is equivalent to a model of biological evolution," *Physical Review Letters*, vol. 78, no. 3, p. 1782, 1997.

[18] O. Banerjee, L. E. Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," *Journal of Machine Learning Research*, vol. 9, pp. 485–516, 2008.

[19] S. Wasserman and P. Pattison, "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*," *Psychometrika*, vol. 61, no. 3, pp. 401–425, 1996.

[20] K. Klemm, V. M. Egu´luz, R. Toral, and M. San-Miguel, "Nonequilibrium transitions in complex networks: A model of social interaction," *Physical Review E*, vol. 67, p. 026120, 2003.

[21] F. Barahona, "On the computational complexity of ising spin glass models," *J. Phys. A: Math. Gen.*, vol. 15, pp. 3241–3253, 1982.

[22] R. Maitra and J. E. Besag, "Bayesian reconstruction in synthetic Magnetic Resonance Imaging," in *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE 1998) Meetings*, A. Mohammad-Djafari, Ed., vol. 3459, 1998, pp. 39–47.

[23] J. Majewski, H. Li, and J. Ott, "The Ising model in physics and statistical genetics," *The American Journal of Human Genetics*, vol. 69, pp. 853–862, 2001.

[24] L. Reichl, *A modern course in statistical physics*. Austin, TX: University of Texas Press, 1980.

[25] Q. Mo and F. Liang, "A hidden Ising model for ChIP-chip data analysis," *Bioinformatics*, vol. 26, no. 6, pp. 777–783, 2010.

[26] B. Kaufman, "Crystal statistics. ii. partition function evaluated by spinor analysis," *PHYSICAL REVIEW*, vol. 76, no. 8, pp. 1232–1243, 1949.

[27] N. N. Schraudolph and D. Kamenetsky, "Efficient exact inference in planar ising models," in *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds. Curran Associates, Inc., 2009, pp. 1417–1424.

[28] Y. M. Karandashev and M. Y. Malsagov, "Polynomial algorithm for exact calculation of partition function for binary spin model on planar graphs," *Optical Memory and Neural Networks*, vol. 26, no. 2, pp. 87–95, Apr 2017.

[29] Y. Ogata, "A Monte Carlo method for high dimensional integration," *Numerische Mathematik*, vol. 55, pp. 137–157, 1989.

[30] A. Gelman and X. L. Meng, "Simulating normalizing constants: from importance sampling to bridge sampling to path sampling," *Statistica Sinica*, vol. 13, pp. 163–185, 1998.

[31] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with an unknown number of components," *Journal of the Royal Statistical Society, Series B*, vol. 59, pp. 731–792, 1997.

[32] F. Wang and D. P. Landau, "Efficient, multiple-range random walk algorithm to calculate the density of states," *Physical Review Letter*, vol. 86, pp. 2050–2053, 2001.

[33] A. M. Ferrenberg and R. H. Swendsen, "Optimized Monte Carlo data analysis," *Physics Review Letters*, vol. 63, no. 12, p. 1195, 1989.

[34] S. Kumar, D. Bouzida, S. R. H., P. A. Kollman, and J. M. Rosenberg, "The weighted histogram analysis method for free-energy calculations on biomelecules. I. The method," *Journal of Computational Chemistry*, vol. 3, no. 8, pp. 1011–1021, 1992.

[35] Y. F. Atchade, N. Lartillot, and C. P. Robert, "Bayesian computation for statistical models with intractable normalizing constants," *Brazilian Journal of Probability and Statistics*, vol. 27, no. 4, pp. 416–436, 2013.

[36] D. P. Landau, S.-H. Tsai, and M. Exler, "A new approach to Monte Carlo simulations in statistical physics: Wang-Landau sampling," *American Journal of Physics*, vol. 72, pp. 1294–1302, 2004.

[37] F. Liang, "A generalized Wang-Landau algorithm for Monte-Carlo computation," *Journal of the American Statistical Association*, vol. 100, no. 472, pp. 1311–1327, 2005.

[38] Y. F. Atchade and J. S. Liu, "The Wang-Landau algorithm in general state spaces: applications and convergence analysis," *Statistica Sinica*, vol. 20, pp. 209–233, 2010.

[39] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *Journal of the Royal Statistical Society, Series B*, vol. 68, pp. 411–436, 2006.

[40] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.

[41] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Ninth Dover printing, tenth GPO printing ed. New York: Dover, 1964.

[42] J. Stirling, *Methodus Differentialis*, London, 1730.

[43] R. H. Swendsen and J.-S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," *Physical Review Letter*, vol. 58, pp. 86–88, 1987.

[44] S. Salinas, *Introduction to Statistical Physics*. Springer, 2001.

[45] N. A. Lazar, *The Statistical Analysis of Functional MRI Data*. Springer, 2008.

[46] M. A. Lindquist, "The statistical analysis of fmri data," *Statistical Science*, vol. 23, no. 4, pp. 439–464, 2008.

[47] R. Maitra, S. R. Roys, and R. P. Gullapalli, "Test-retest reliability estimation of functional mri data," *Magnetic Resonance in Medicine*, vol. 48, pp. 62–70, 2002.

[48] R. Maitra, "Assessing certainty of activation or inactivation in test-retest fMRI studies," *Neuroimage*, vol. 47, pp. 88–97, 2009.

[49] ——, "A re-defined and generalized percent-overlap-of-activation measure for studies of fmri reproducibility and its use in identifying outlier activation maps," *Neuroimage*, vol. 50, pp. 124–135, 2010.

[50] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 3rd ed. London: Springer-Verlag, 2009.

[51] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, p. 97109, 1970.

[52] N. Metropolis, A. W. Rosenbluth, M. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, p. 10871092, 1953.

[53] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of functional mri time-series," *Human Brain Mapping*, vol. 1, pp. 153–171, 1994.

[54] S. Hayasaka and T. E. Nichols, "Validating cluster size inference: random field and permutation methods," *Neuroimage*, vol. 20, p. 23432356, 2003.

[55] C.-W. Woo, A. Krishnan, and T. D. Wager, "Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations," *Neuroimage*, vol. 91, p. 412419, 2014.

[56] S. Forman, J. Cohen, M. Fitzgerald, W. Eddy, M. Mintun, and D. Noll, "Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold," *Magnetic Resonance in Medicine*, vol. 33, p. 636647, 1995.

[57] R. W. Cox, "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages," *Computers and biomedical research, an international journal*, vol. 29, no. 3, pp. 162–173, 1996.

[58] R. W. Cox and J. S. Hyde, "Software tools for analysis and visualization of fMRI data," *NMR in Biomedicine*, vol. 10, no. 4-5, pp. 171–178, 1997.

[59] R. W. Cox, "AFNI: What a long strange trip it has been," *NeuroImage*, vol. 62, pp. 743–747, 2012.

[60] A. de Moivre, "*Approximatio ad Summam Terminorium Binomii* $\overline{a+b}\backslash^{n}$*," in Seriem Expansi. reprinted in [62], 1733.*

[61] M. A. Pinsky, "The normal approximation to the hypergeometric distribution," 2006, northestern University https://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/pinsky-hypergeometric.pdf.

[62] R. C. Archibald, "A rare pamphlet of Moivre and some of his discoveries," Isis*, vol. 8, pp. 671–683, 1926.*