

# RESEARCH STATEMENT

KSHITIJ KHARE  
STANFORD UNIVERSITY

## 1. INTRODUCTION

My primary research interests lie broadly in statistical methodology and applied probability. In particular, my doctoral work includes research on graphical models and high-dimensional inference, convergence rates of Markov chains arising from Markov chain Monte Carlo (MCMC) simulation, and constructions of Gaussian fields using Markov chains and their connections to Bayesian statistics. The research undertaken in my thesis, together with future research plans, is described below.

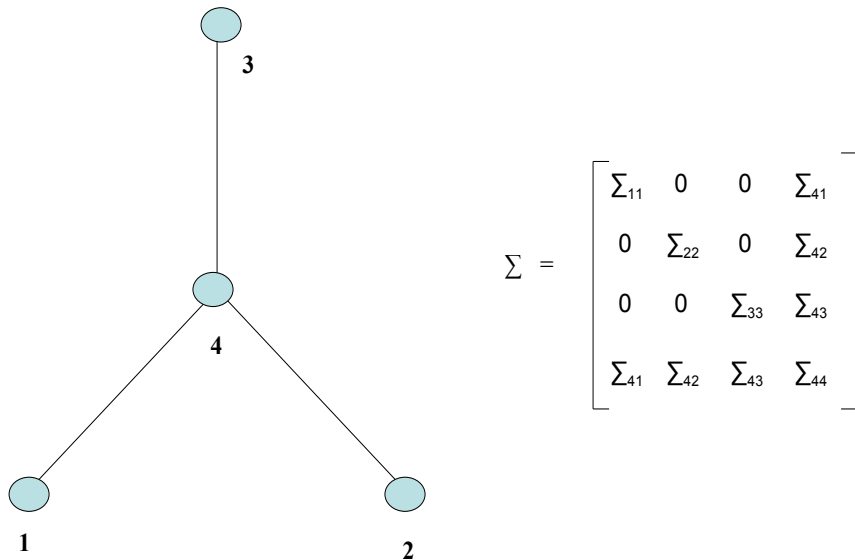
## 2. GRAPHICAL MODELS WITH APPLICATIONS TO HIGH-DIMENSIONAL INFERENCE

The ‘high-dimensional data’ revolution has left almost no field untouched. In recent years, due to advances in science and technology, high-throughput data from genomics, finance, environmental and marketing (among other) applications is being generated at a rapid pace. This has created an urgent need for methodology and tools for analyzing high-dimensional data. In biological applications, gene expression data where the number of genes is in the tens of thousands, and the number of available samples is in the hundreds or less is very common. In the environmental and climate sciences, measurements are often recorded at numerous locations on many variables. Social sciences survey data often record hundreds or thousands of characteristics on individuals and groups. Financial data such as stock and bond prices are often captured and recorded in real time, sometimes every minute or every second, and on tens of thousands of financial instruments from all over the world. Formulating correct statistical models that make sense of all the many complex relationships and multivariate dependencies that are in the data, investigating the properties of these models and developing inferential procedures has provided major challenges for statisticians, probabilists, and others working in this field.

A class of models that are well-suited for high-dimensional inference, and have found widespread applications, are ‘graphical models’. More specifically, a graphical model is a probability model that characterizes the marginal or conditional independence structure of a set of random variables by a graph. For Gaussian distributions, marginal and conditional independencies translate respectively into zeros in the covariance matrix and the inverse covariance matrix. Graphical models therefore have a natural place in high-dimensional inference since they provide a natural mechanism to reduce the parameter space, a fundamental step in modeling complex systems where the number of random variables often exceed the number of observations (the “large  $p$  small  $n$ ” problem).

A subclass of graphical models that has received recent attention in the literature are ‘covariance graph models’, formally introduced to the statistics community in [5]. Gaussian covariance graph models encode marginal independence among random variables that are represented by the vertices of a graph  $G$ . These marginal independencies translate into zeroes in the covariance matrix  $\Sigma$ . These models are distinctly different from the traditional concentration graph models (often also referred to by Gaussian graphical models or covariance selection models), where the zeroes are reflected in the precision matrix  $\Omega = \Sigma^{-1}$ . We now proceed to define covariance graph models formally. Consider an undirected graph  $G$  with a finite set of vertices  $V$  (of size  $p$ ) and a finite set  $E$  of edges between these vertices. The Gaussian covariance graph model corresponding to the graph  $G$  is the collection of  $p$ -variate Gaussian distributions with covariance matrix  $\Sigma$ , such that  $\Sigma_{ij} = 0$  whenever  $(i, j) \notin E$ . See below for an example when  $p = 4$ . In the frequentist setting, maximum likelihood estimation in covariance graph models has been a topic of interest in recent years. See [4, 15, 22, 37] and the references therein. These models have also been used in applications in [3, 19, 30]. My

doctoral work on developing a comprehensive Bayesian framework for analyzing covariance graph models, and also addressing outstanding issues in maximum likelihood estimation for these models, has been developed in a series of three papers [24, 25, 26]. A part of the work is presented below.



**The ‘butterfly’ graph with 4 vertices and the corresponding covariance matrix**

Although Gaussian covariance graph models are simple to understand, a comprehensive framework for Bayesian inference for these models is not available in the literature. The main difficulty being that these models give rise to curved exponential families. The zero restrictions on the entries of the covariance matrix  $\Sigma$  translate into complicated restrictions on the corresponding entries of the inverse covariance matrix  $\Sigma^{-1}$ , which is the canonical parameter. These complicated restrictions compound the problem of constructing a usable family of priors for covariance graph models - especially in high dimensions, where one cannot always use non-tractable priors and resort to MCMC. Moreover, no general theory for Bayesian inference, akin to the celebrated Diaconis-Ylvisaker conjugate theory [10] for natural exponential families, is available for curved exponential families.

An important requirement from a class of priors is that we should be able to compute quantities like the mean of the posterior distribution, either in closed form or by sampling from the posterior distribution by a simple mechanism. This is especially important in high dimensional situations, where complex computations can quickly become infeasible. A related desirable feature is conjugacy, i.e. the class of priors is such that the posterior distribution also belongs to the same class. The classes of Wishart distributions developed by Letac and Massam [29] and used later on in the concentration graph setting by Rajaratnam et al [34] seem promising as priors for covariance graph models. However, we establish that the posterior distribution fails to belong to the same class, and computing quantities like the posterior mean either in closed form or by sampling from the posterior distribution is intractable in high-dimensional situations.

It is not immediately clear how to construct a family of conjugate priors for our models, especially a family with properties that will easily enable inference in high dimensions. This problem, its importance and relevance, have been raised by many leaders in the graphical models community - yet a comprehensive framework for Bayesian inference is not available. We have addressed this problem and have developed a comprehensive framework for Bayesian inference for covariance graph models. We first identify that constructing a class of priors

on the same space on which our parameter of interest  $\Sigma$  lives, leads to priors which are at first intractable. We demonstrate that reparametrizing in terms of the modified Choleski decomposition of the covariance matrix  $\Sigma$ , together with a suitable ordering of the variables, leads to a family of conjugate priors with desirable properties. It is known that a given positive definite matrix  $\Sigma$  can be uniquely expressed as  $\Sigma = LDL^T$ , where  $L$  is a lower triangular matrix with diagonal entries equal to 1, and  $D$  is a diagonal matrix with positive diagonal entries.

This parametrization works because of three reasons. Firstly, it turns out that if the graph  $G$  is decomposable (see [28] for the definition), then there is a suitable ordering of the vertices such that the zero restrictions in  $\Sigma$  translate into zero restrictions in exactly the same places in the lower triangle of  $L$ . The unrestricted entries of  $L$  are now completely free parameters in the model. Secondly, our class of priors are conjugate, i.e., the posterior distribution lies in the same family as the prior distribution. Thirdly, for any distribution in our class, the conditional distribution of each column of  $L$  given the other parameters is multivariate normal, and the conditional distribution of each diagonal entry of  $D$  given the other parameters is inverse-gamma, thereby allowing us to construct a block Gibbs sampling procedure to sample from the posterior distribution. Our class of priors are specified only upto a normalizing constant. We establish conditions under which the normalizing constant exists. We also establish conditions under which the normalizing constant can be evaluated in closed form, thereby allowing us to do model selection by the method of marginal likelihood. In such cases, we derive other useful properties for our class of distributions, like hyper-Markov properties, existence of closed form Bayes estimators for two standard loss functions, and linearity of the posterior mean. We have had extremely positive feedback on our work from leaders in the field.

*Plans for future research.* Our work gives rise to further practical and theoretical questions. On the practical front, how should model selection be done for covariance graphs where the normalizing constant is not available? On the theoretical front, how should we choose priors to perform “objective Bayesian inference”? How should these ideas be extended to the case when  $G$  is not decomposable? This is partially addressed in ongoing work.

### 3. CONVERGENCE RATES OF MARKOV CHAINS

The ‘Markov chain Monte Carlo’ (MCMC) revolution in recent years has affected numerous scientific disciplines. Monte Carlo methods such as Gibbs sampling, Metropolis-Hastings algorithm, importance sampling and their variants are being widely used in statistics, statistical physics, finance and other areas. Sampling from a given probability density  $f$  (or estimating some quantity associated with it) is often required. However, it is not always possible to do so directly. The idea in Markov chain Monte Carlo is to use a Markov chain which converges to  $f$  and run this Markov chain long enough to approximately sample from  $f$ . An important question for the practitioner using this method is: what is the ‘burn-in’ period? That is, how long should the Markov chain be run in order for the approximation to be reasonably accurate? Finding ‘rates of convergence’ or providing ‘running time analyses’ of Markov chains, entails answering these questions as precisely and quantitatively as possible. Providing accurate answers to the questions above is often a hard task. My doctoral work pertaining to this area can be found in [7, 8, 27].

One of the most basic and popular Monte Carlo methods is the Gibbs sampler, also known as Glauber dynamics or the heat-bath algorithm. It is useful for sampling from multivariate probability distributions, perhaps known only upto a normalizing constant, by a sequence of one dimensional sampling problems. The algorithm was introduced in 1963 by Glauber [17] to do simulations for Ising models, and independently by Turcin [36]. See [7] for an extensive list of references. The Gibbs sampler is applicable when the conditional distributions of each coordinate given the others are easy to sample from. From  $(X_1, \dots, X_p)$

proceed to  $(X'_1, X_2, \dots, X_p)$  then  $(X'_1, X'_2, X_3, \dots, X_p), \dots, (X'_1, X'_2, \dots, X'_p)$  where at the  $i^{\text{th}}$  stage, the  $i^{\text{th}}$  coordinate is sampled from  $f$  with the other coordinates fixed. This is one pass. Continuing in this manner yields a Markov chain  $X, X', X'', \dots$ , which has  $f$  as stationary density under mild conditions discussed in [2, 35]. It is natural to ask how many steps are required for convergence. This is a challenging research area and currently available state-of-the-art techniques (eg. Harris recurrence) can give answers that are way off from the correct one, even for simple problems.

In [7], which is joint work with Persi Diaconis and Laurent Saloff-Coste, we derive sharp rates of convergence to the stationary distribution for some two-component Gibbs sampling Markov chains involving exponential families and their conjugate priors, where other standard techniques were giving inaccurate bounds. Our examples are illustrative, because it is easy to directly sample from the bivariate densities that we consider. However, they are a useful set of easy-to-understand standard statistical examples, and may be used to compare and benchmark more robust techniques, such as the Harris recurrence techniques in [20] or the spectral techniques in [1] and [40]. They may also serve as a base for the comparison techniques in [1, 9, 12]. Moreover, ideas and results from this work have been useful in analyzing more complicated Markov chains in [27]. The operators corresponding to the transition densities of these Markov chains are explicitly diagonalizable with classical orthogonal polynomials as eigenfunctions. There are interesting connections of this work to the subject of orthogonal polynomials and Lancaster families.

Here is an example taken from [7]. Let  $X \mid \theta \sim \text{Binomial}(n, \theta)$ , and suppose for simplicity we choose a uniform prior distribution on  $\theta$  (we can choose any Beta distribution as the prior and the analysis provided below goes through). The joint density of  $(X, \theta)$  is the bivariate Beta/Binomial density (uniform prior)

$$f(x, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x \in \{0, 1, \dots, n\}, \theta \in (0, 1).$$

The Gibbs sampler for the bivariate density  $f$  proceeds as follows. Suppose we start at  $(x, \theta)$ . Draw  $\theta'$  from the conditional distribution given  $x$ , which is  $\text{Beta}(x + 1, n - x + 1)$ . Draw  $x'$  from the conditional distribution given  $\theta'$ , which is  $\text{Binomial}(n, \theta')$ . The output is  $(x', \theta')$ . Let  $\tilde{K}(x, \theta; x', \theta')$  be the transition density for this chain.  $\tilde{K}$  has  $f$  as the stationary density. The theorem below shows that order  $n$  steps are necessary and sufficient for convergence. That is, for sampling from the Markov chain  $\tilde{K}$  to simulate from the bivariate probability distribution  $f$ , a small (integer) multiple of  $n$  steps suffice, while  $\frac{n}{2}$  steps do not. Let  $\tilde{K}_{n,\theta}^\ell$  denote the density of the bivariate chain after  $\ell$  steps starting from  $(n, \theta)$ .

**Theorem 1** ([7]). *For the bivariate chain  $\tilde{K}$ , for all  $\theta, n$  and  $\ell$ ,*

$$\frac{1}{2} \beta_1^\ell \leq \|\tilde{K}_{n,\theta}^\ell - f\|_{TV} \leq \frac{\beta_1^{\ell-1/2}}{1 - \beta_1^{2\ell-1}}, \quad \text{with } \beta_1 = 1 - \frac{2}{n+2}.$$

Here  $\|\cdot\|_{TV}$  denotes the total variation distance. The calculations work because the operator corresponding to the ‘ $x$ -chain’ (the Markov chain on  $\{0, 1, 2, \dots, n-1, n\}$  obtained by only looking at the  $x$ -component of the bivariate Markov chain) takes polynomials to polynomials, which can be used to explicitly identify its eigenvalues and eigenfunctions. We provide classes of examples with the same explicit behavior. These include the cases when  $X \mid \theta$  is distributed as one of the exponential families singled out by Morris [32, 33] (binomial, Poisson, negative binomial, normal, gamma, hyperbolic) with the prior on  $\theta$  being the conjugate prior, or  $X \mid \theta \sim g(\cdot - \theta)$  is a location family with a conjugate prior on  $\theta$ , and  $g$  being one of the six exponential families above.

In a companion paper [8], we develop a stochastic approach which uses one eigenfunction combined with coupling arguments. This is possible when the Markov chains are stochastically monotone. We show this is the case for all exponential families with any choice of prior,

and for location families where the density  $g$  is totally positive of order two. This lets us provide running time analyses for the examples in [7] for which higher moments do not exist (negative binomial, gamma, hyperbolic). In addition, location problems fall into the setting of iterated random functions so that backward iteration and coupling are available. We also provide an extension of the so-called Wilson’s lemma for lower bounds on the distance to stationarity, under weaker conditions.

In [27], we provide running time analyses of some standard multivariate Markov chains using spectral techniques. The class of examples includes the Moran’s model in genetics, the multivariate Bernoulli-Laplace process, a generalized Ehrenfest urn model, a generalized Polya urn model, the multivariate normal autoregressive process, and a multivariate hypergeometric random walk. These Markov chains arise from standard models in population genetics and image processing. It turns out that the operator corresponding to the transition densities of all the Markov chains in the list above takes polynomials to polynomials. This property enables us to get hold of the eigenvalues and eigenfunctions of the Markov chains. Using this information, we provide explicit non-asymptotic bounds on the distance to stationarity for every Markov chain, given that it starts at a natural initial state. These bounds are in terms of the chi-squared distance.

*Plans for future research.* The success we had in analyzing the Markov chains mentioned above is encouraging, but it is important to note that many Markov chains being used by practitioners can be much more complicated, and existing techniques are insufficient to provide useful convergence rates for these Markov chains. The main aim of my future research in this area is to enhance existing techniques to provide accurate running time analyses for these complicated Markov chains. One possibility is to extend techniques in [7, 8] to encompass multi-component Gibbs samplers. Another challenging direction of research which I plan to undertake is developing better techniques for analyzing non-reversible Markov chains.

#### 4. CONSTRUCTION OF GAUSSIAN FIELDS FROM MARKOV CHAINS

In various applications, it is useful to construct Gaussian fields on spaces such as the vertices of a graph, the plane or a manifold. Constructing positive definite covariance functions on such spaces can be challenging. On the other hand, it is usually easy to construct Markov chains. The Dynkin isomorphism (introduced in [13, 14]) associates a Gaussian field to a Markov chain. This gives fields with all positive covariances. A parallel construction of Diaconis-Evans [6] gives fields with all negative covariances. Both of these constructions provide simple recipes for obtaining Gaussian fields from Markov chains. These Gaussian fields have useful properties, which facilitate their use as priors for prediction and design problems [39]. Dynkin’s construction has been used for proving deep and interesting results in [16, 31] and many others. See [6] and [23] for a review of this literature.

In Dynkin’s construction, all covariances are non-negative while the Diaconis-Evans construction leads to all non-positive covariances. This is a limitation in applications where both positive and negative correlations naturally occur, for instance in the climate and environmental sciences. Through careful analysis, we extend these constructions to allow general covariance sign patterns.

Let  $\{X_t\}_{t \geq 0}$  be a reversible Markov chain with finite state space  $\mathcal{X}$ , and a symmetric generator matrix  $Q = ((q_{xy}))_{x,y \in \mathcal{X}}$ , such that  $Q^{-1}$  exists. This means  $Q$  is non-conservative and the process gets killed (goes to an absorbing “cemetery” not included in  $\mathcal{X}$ ) after a random time. Dynkin’s construction in this setting entails looking at a Gaussian field  $\{Z_x\}_{x \in \mathcal{X}}$  with covariance matrix  $\Sigma = -Q^{-1}$ . All elements of  $\Sigma$  are non-negative. This construction has various useful properties including the fact that the conditional expectations and variances for the Gaussian field  $\{Z_x\}_{x \in \mathcal{X}}$  with covariance matrix  $\Sigma$  can be elegantly expressed in terms of the Markov chain  $\{X_t\}_{t \geq 0}$ . These can potentially be used to estimate the conditional

means and variances by simulating from the Markov chain, if required. The Gaussian field  $\{Z_x\}_{x \in \mathcal{X}}$  is also a Markov random field with respect to the graph given by the zeroes in  $Q$ .

To extend Dynkin's construction, we introduce a sign matrix  $\mathcal{S}$  such that  $\mathcal{S}(x, x) = 1$ ,  $\mathcal{S}(x, y) = \mathcal{S}(y, x)$ ,  $\mathcal{S}(x, y) \in \{-1, 1\} \forall x \neq y \in \mathcal{X}$ . We say that the transition from  $x$  to  $y$  is positive if  $\mathcal{S}(x, y) = 1$ , and the transition from  $x$  to  $y$  is negative if  $\mathcal{S}(x, y) = -1$ . For each  $i \geq 1$ , let  $S_i$  be the random time when the  $i^{\text{th}}$  transition occurs for the Markov chain  $\{X_t\}_{t \geq 0}$ . Define the sign-chain  $\{H_t\}_{t \geq 0}$  by

$$H_t = \prod_{i=1}^{\infty} \mathcal{S}(X_{S_{i-1}}, X_{S_i}) 1_{\{S_i \leq t\}} \text{ (with } S_0 = 0\text{)}.$$

Hence,  $H_t$  is 1 if the number of negative transitions upto time  $t$  is even, and  $H_t$  is  $-1$  if the number of negative transitions upto time  $t$  is odd. Consider a Gaussian field  $\{Z_x^{\mathcal{S}}\}_{x \in \mathcal{X}}$  with mean zero and covariance matrix  $\Sigma^{\mathcal{S}} = (-Q \circ \mathcal{S})^{-1}$ , where  $\circ$  denotes the Hadamard (i.e. elementwise) product of two matrices. In this manner, we can get positive as well as negative elements in  $\Sigma^{\mathcal{S}}$ . We show in [23] that all properties of Dynkin's construction can be suitably extended for this more general construction. In particular, the conditional expectations and variances of the Gaussian field  $\{Z_x^{\mathcal{S}}\}_{x \in \mathcal{X}}$  can still be expressed explicitly in terms of the Markov chain  $\{X_t\}_{t \geq 0}$ , with appropriate modifications. It is remarkable that in these formulae, the presence of the matrix  $\mathcal{S}$  is reflected only through the sign-chain  $\{H_t\}_{t \geq 0}$ . Clearly, the presence of terms related to  $\{H_t\}_{t \geq 0}$  slightly complexifies these formulae, but it is quite straightforward to keep track of  $\{H_t\}_{t \geq 0}$  while simulating  $\{X_t\}_{t \geq 0}$ . Hence the presence of these terms makes little difference, atleast when we use simulation to estimate quantities related to the Markov chain  $\{X_t\}_{t \geq 0}$ .

Diaconis and Evans proposed a different construction by looking at a Gaussian field  $\{Z_x^E\}_{x \in \mathcal{X}}$  with covariance matrix  $\Sigma^E = -Q$ , which has negative individual covariances. We can generalize the Diaconis-Evans construction by associating with  $\{X_t\}_{t \geq 0}$  a Gaussian field  $\{Z_x^{E\mathcal{S}}\}_{x \in \mathcal{X}}$  with covariance matrix  $\Sigma^{E\mathcal{S}} = -Q \circ \mathcal{S}$ . The Diaconis-Evans construction has parallel properties to the ones described above for Dynkin's construction, and our generalization extends these properties. In [23], we provide a straightforward method to simulate  $\{Z_x^{E\mathcal{S}}\}_{x \in \mathcal{X}}$  when  $\mathcal{X}$  is finite.

*Plans for future research* The generalizations of Dynkin's and Diaconis-Evans' constructions, preserve to a large extent the useful properties of the original constructions. We are focusing on applying these constructions to prediction and design problems.

## REFERENCES

- [1] Amit, Y. (1996). Convergence properties of the Gibbs sampler for perturbations of Gaussians. *Ann. Statist.* **24**, 122–140.
- [2] Athreya, K., Doss, H. and Sethuraman, J. (1996). On the convergence of the Markov chain simulation method, *Ann. Statist.* **24**, 89–100.
- [3] Butte, A.J., Tamayo, P., Slonim, D., Golub, T.R. and Kohane, I.S. (2000). Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks, *Proc. Nat. Acad. Sci.* **97**, 12182-12186.
- [4] Chaudhuri, S., Drton, M. and Richardson, T.S. (2007). Estimation of a covariance matrix with zeroes, *Biometrika* **94**, 199-216.
- [5] Cox, D.R. and Wermuth, M. (1993). Linear dependencies represented by chain graphs (with Discussion), *Statist. Sci.* **8**, 204-218, 247-277.
- [6] Diaconis, P. and Evans, S. (2002). A different construction of Gaussian fields from Markov chains: Dirichlet covariances, *Ann. Inst. Henri Poincaré* **38**, 863-878.
- [7] Diaconis, P., Khare, K. and Saloff-Coste, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials, *Statistical Science* **23**, 151-178.
- [8] Diaconis, P., Khare, K. and Saloff-Coste, L. (2008). Gibbs sampling, conjugate priors and coupling, *submitted*.
- [9] Diaconis, P. and Saloff-Coste, L. (1993). Comparison theorems for Markov chains, *Annals of Applied Probability* **3**, 696-730.

- [10] Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families, *Ann. Statist.* **7**, 269-281.
- [11] Drton, M. and Richardson, T.S. (2008). Graphical methods for efficient likelihood inference in Gaussian covariance models, *Journal of Machine Learning Research* **9**, 893-914.
- [12] Dyer, M., Goldberg, L., Jerrum, M. and Martin, R. (2005). Markov chain comparison, *Probability Surveys* **3**, 89-111.
- [13] Dynkin, E.B. (1980). Markov processes and random fields, *Bulletin of the American Mathematical Society (N.S.)* **3**, 975-999.
- [14] Dynkin, E.B. (1983). Markov processes as a tool in field theory, *Journal of Functional Analysis* **50**, 167-187.
- [15] Edwards, D.M. (2000). *Introduction to Graphical Modelling*, 2nd ed. New York: Springer-Verlag.
- [16] Eisenbaum, N., Kaspi, H., Marcus, M., Rosen, J. and Shi, Z. (2000). A Ray-Knight theorem for symmetric Markov processes, *Annals of Probability* **28**, 1781-1796.
- [17] Glauber, R. (1963). Time dependent statistics of the Ising model, *Jour. Math. Phys.* **4**, 294-307.
- [18] Griffiths, R.C. (2006).  $n$ -kernel orthogonal polynomials in Dirichlet, Dirichlet-multinomial, Ewens' sampling distribution and Poisson Dirichlet process, Preprint, Department of Statistics, Oxford University.
- [19] Grzebyk, M., Wild, P. and Chouaniere, D. (2004). On identification of multifactor models with correlated residuals, *Biometrika* **91**, 141-151.
- [20] Jones, G. and Hobert, J. (2001). Honest exploration of intractable probability distributions via Markov chain Monte carlo, *Statistical Science* **16**, 312-334.
- [21] Karlin, S. and McGregor, J. (1965). Ehrenfest urn models, *Journal of Applied Probability* **2**, 352-376.
- [22] Kauermann, G. (1996). On a dualization of graphical Gaussian models, *Scand. J. Statist.* **23**, 105-116.
- [23] Khare, K. (2008) Dynkin's and Diaconis-Evans' constructions with sign structure, *working paper*.
- [24] Khare, K. and Rajaratnam, B. (2008). Maximum likelihood estimation in covariance graph models, *in process of submission*.
- [25] Khare, K. and Rajaratnam, B. (2008). Conjugate Wishart distributions for covariance graph models. Technical Report No. 2008-11, November 2008, Department of Statistics, Stanford University, *in process of submission*.
- [26] Khare, K. and Rajaratnam, B. (2008). Bayesian inference in covariance graph models, *working paper*.
- [27] Khare, K. and Zhou, H. (2008) Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions, to appear in the *Annals of Applied Probability*.
- [28] Lauritzen, S. (1996). *Graphical models*, Calderon Press, Oxford.
- [29] Letac, G. and Massam, H. (2007). Wishart distributions for decomposable graphs, *Ann. Statist.* **35**, 1278-1323.
- [30] Mao, Y., Kschischang, F.R. and Frey, B.J. (2004). Convolutional factor graphs as probabilistic models, In *Proceedings of the 20<sup>th</sup> Conference on Uncertainty in Artificial Intelligence*, Eds. M. Chickering and J. Halperin, pp. 374-381, Arlington, MA: AUAI Press.
- [31] Marcus, M.B. and Rosen, J. (2006). *Markov processes, Gaussian Processes and Local Times*, Cambridge University Press.
- [32] Morris, C. (1982). Natural exponential families with quadratic variance functions, *Annals of Statistics* **10**, 65-80.
- [33] Morris, C. (1983). Natural exponential families with quadratic variance functions: Statistical theory, *Annals of Statistics* **11**, 515-589.
- [34] Rajaratnam, B., Massam, H. and Carvalho, C. (2008). Flexible covariance estimation in graphical models. To appear in *Annals of Statistics*.
- [35] Tierney, L. (1994). Markov chains for exploring posterior distributions, (with discussion), *Ann. Statist.* **22**, 1701-1762.
- [36] Turcin, V. (1971). On the computation of multidimensional integrals by the Monte Carlo method, *Probab. Applications* **16**, 720-724.
- [37] Wermuth, M., Cox, D.R. and Marchetti, G.M. (2006). Covariance chains, *Bernoulli* **12**, 841-862.
- [38] Wilson, D.B. (2004). Mixing times of lozenge tiling and card shuffling Markov chains, *Annals of Applied Probability* **14**, 274-325.
- [39] Ylvisaker, D. (1987). Prediction and design, *Annals of Statistics* **15**, 1-19.
- [40] Yuen, W.K. (2000). Applications of geometric bounds on the convergence rates of Markov chains in  $\mathbb{R}^n$ , *Stochastic Processes and their Applications* **87**, 1-23.