Gibbs sampling, exponential families and orthogonal polynomials

Persi Diaconis
Departments of mathematics and Statistics
Stanford University

Kshitij Khare
Department of Statistics
Stanford University

Laurent Saloff-Coste*
Department of mathematics
Cornell University

October 16, 2006

Abstract

We give families of examples where a sharp analysis of the widely used Gibbs sampler is available. The examples involve standard exponential families and their conjugate priors. In each case, the transition operator is explicitly diagonalizable with classical orthogonal polynomials as eigenfunctions.

1 Introduction

The Gibbs sampler, also known as Glauber dynamics or the heat-bath algorithm, is a mainstay of scientific computing. It provides a way to draw samples from a multivariate probability density $f(x_1, x_2, \ldots, x_p)$, perhaps only known up to a normalizing constant, by a sequence of one dimensional sampling problems. From $(X_1, \ldots, X_p)$ proceed to $(X_1', X_2, \ldots, X_p)$ then $(X_1', X_2', X_3, \ldots, X_p), \ldots, (X_1', X_2', \ldots, X_p')$ where at the i-th stage, the coordinate is sampled from f with the other coordinates fixed. This is one pass. Continuing gives a Markov chain $X, X', X'', \ldots$, which has f as stationary density under mild conditions discussed below.

The algorithm was introduced in 1963 by Glauber [39] to do simulations for Ising models. It is still a standard tool of statistical physics, both for practical simulation (e.g., [61]) and as a natural dynamics (e.g., [9]). The basic Dobrushin uniqueness theorem showing existence of Gibbs measures was proved based on this dynamics (e.g., [41]). It was introduced as a base

*Research partially supported by NSF grant DMS 0102126

for image analysis by Geman and Geman [36]. Statisticians began to employ the method for routine Bayesian computations following the work of Tanner and Wong [69] and numerous papers by Allen Gelfand and Adrian Smith. Textbook accounts, with many examples from biology and the social sciences along with extensive references are in [37, 38, 54].

Despite heroic efforts by the applied probability community, useful running time analyses for Gibbs sampler chains is still a major research effort. An overview of available tools and results is given at the end of this introduction. The main purpose of the present paper is to give families of two component examples where a sharp analysis is available. These may be used to compare and benchmark more robust techniques. They may also serve as a base for the comparison techniques [21, 27].

Here is an example of our results. Let

$$f_\theta(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad \pi(d\theta) = \text{ uniform on } [0,1], \quad x \in \{0,1,2,\ldots,n\}.$$

These define the bivariate Beta/Binomial density (uniform prior)

$$f(x,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

with marginal density

$$m(x) = \int_0^1 f(x,\theta)d\theta = \frac{1}{n+1} \quad x \in \{0,1,2,\ldots,n\}.$$

The Gibbs sampler for $f(x,\theta)$ proceeds as follows:

- From x, draw θ' from Beta$(x, n-x)$.
- From θ', draw x' from Binomial(n,θ').

The output is (x',θ'). Let $\widetilde{K}(x,\theta;x',\theta')$ be the transition density for this chain. While $\widetilde{K}$ has $f(x,\theta)$ as stationary density, the $(\widetilde{K}, f)$ pair is not reversible. This blocks straightforward use of spectral methods. Jun Liu et al. [53] observed that the 'x-chain' with kernel

$$k(x,x') = \int_0^1 f_\theta(x')\pi(\theta|x)d\theta = \int_0^1 \frac{f_\theta(x)f_\theta(x')}{m(x)}\, d\theta$$

is reversible with stationary density $m(x)$. For the Beta/Binomial example

$$k(x,x') = \frac{2n}{2n+1}\frac{\binom{n}{x}\binom{n}{x'}}{\binom{2n}{x+x'}}, \quad 0 \le x, x' \le n. \tag{1.1}$$

The proposition below gives an explicit diagonalization of the x-chain and sharp bounds for the bivariate chain ($\widetilde{K}^{\ell}_{n,\theta}$ denotes the density of the distribution of the bivariate chain after ℓ steps starting from (n,θ)). It shows that order n steps are necessary and sufficient for convergence. The proof is given in Section 5.

Proposition 1.1 *For the Beta/Binomial example with uniform prior, we have:*

(a) The chain (1.1) *has eigenvalues*

$$\beta_0 = 1, \quad \beta_j = \frac{n(n-1)\cdots(n-j+1)}{(n+2)(n+3)\cdots(n+j+1)}, \quad 1 \le j \le n.$$

In particular, $\beta_1 = 1 - 2/(n+2)$. *The eigenfunctions are the discrete Tchebychev polynomials (orthogonal polynomials for* $m(x) = 1/(n+1)$ *on* $\{0,\dots,n\}$*).*

(b) For the bivariate chain $\widetilde{K}$*, for all* θ, n *and* ℓ*,*

$$\frac{1}{2}\beta_1^{\ell} \le \|\widetilde{K}^{\ell}_{n,\theta} - f\|_{\mathrm{TV}} \le 3\beta_1^{\ell-1/2}.$$

The calculations work because the operator with density $k(x,x')$ takes polynomials to polynomials. Our main results give two classes of examples with the same explicit behavior:

- $f_\theta(x)$ is one of the exponential families singled out by Morris [59, 60] (binomial, Poisson, negative binomial, normal, gamma, hyperbolic) with $\pi(\theta)$ the conjugate prior.
- $f_\theta(x) = g(x-\theta)$ is a location family with $\pi(\theta)$ conjugate and g belongs to one of the six exponential families above.

Section 2 gives background. In Section 2.1 the Gibbs sampler is set up more carefully both in systematic and random scan versions. Relevant Markov chain tools are collected in Section 2.2. Exponential families and conjugate priors are reviewed in Section 2.3. The six families are described more carefully in Section 2.4 which calculates needed moments. A brief overview of orthogonal polynomials is in Section 2.5.

Section 3 is the heart of the paper. It breaks the operator with kernel $k(x,x')$ into two pieces: $T : L^2(m) \to L^2(\pi)$ defined by

$$Tg(\theta) = \int f_\theta(x) g(x) m(dx)$$

and its adjoint T^*. Then k is the kernel of T^*T. Our analysis rests on a singular value decomposition of T. In our examples, T takes orthogonal polynomials for $m(x)$ into orthogonal polynomials for $\pi(\theta)$. This leads to explicit computations and allows us to treat the random scan, x-chain and θ-chain on an equal footing.

The x-chains and six θ-chains corresponding to the six classical exponential families are treated in Section 4. There are some surprises; while order n steps are required for the Beta/Binomial example above, for the parallel Poisson/Gamma example, $\log n + c$ steps are necessary and sufficient. The six location chains are treated in Section 5 where some standard queuing models emerge (e.g. the $M/M/\infty$ queue). All of the operators studied above turn out to be compact. In Section 6 we show this persists for more general families and priors. The final section points to other examples with polynomial eigenfunctions and other methods for studying present examples.

Our examples are just illustrative. It is easy to sample from any of the families $f(x, \theta)$ directly. Further, we do not see how to carry our techniques over to higher component problems. Basic convergence properties of the Gibbs sampler can be found in [4, 70]. Explicit rates of convergence appear in [64, 65]. These lean on Harris recurrence and require a drift condition of type $E(V(X_1)|X_0 = x) \leq aV(x) + b$ for all x. Also required are a minorization condition of the form $k(x, x') \geq \epsilon q(x')$ for $\epsilon > 0$, some probability density q, and all x with $V(x) \leq d$. Here d is fixed with $d \geq b/(1 + a)$. Rosenthal [64] then gives explicit upper bounds and shows these are sometimes practically relevant for natural statistical examples. Finding useful V and q is currently a matter of art. For example, a group of graduate students tried to use these techniques in the Beta/Binomial example treated above and found it difficult to make choices giving useful results. This led to the present paper. A marvelous expository account of this set of techniques with many examples and an extensive literature review is given by Jones and Hobart in [45]. In their main example an explicit eigenfunction was available for V; our Gamma/Gamma examples below generalize this. Some sharpenings are in [8] which also makes useful connections with classical renewal theory.

2 Background

This section gives needed background. The two component Gibbs sampler is defined more carefully in Section 2.1. Bounds on convergence using eigenvalues are given in Section 2.2. Exponential families and conjugate priors are reviewed in Section 2.3. The six families with variance a quadratic function of the mean are treated in Section 2.4. Finally, a brief review of orthogonal polynomials is in Section 2.5.

2.1 Two-Component Gibbs Samplers

Let $(\mathcal{X}, \mathcal{F})$ be a measurable space equipped with a σ-finite measure μ. Let $(\Theta, \mathcal{G})$ be a measurable space equipped with a probability measure π. Let $\{f_\theta(x)\}_{\theta\in\Theta}$ be a family of probability densities with respect to μ. These define a probability measure on $\mathcal{X} \times \Theta$ via

$$P(A \times B) = \int_B \int_A f_\theta(x)\mu(dx)\pi(d\theta) \quad A \in \mathcal{F}, B \in \mathcal{G}.$$

The marginal density on $\mathcal{X}$ is

$$m(x) = \int_\Theta f_\theta(x)\pi(d\theta) \quad (\text{so } \int_\mathcal{X} m(x)\mu(dx) = 1)$$

The posterior density is given by

$$\pi(\theta|x) = f_\theta(x)/m(x).$$

For simplicity, we assume that this formula defines a probability density with respect to $\pi(d\theta)$, for every $x \in \mathcal{X}$. In particular, we assume that $0 < m(x) < \infty$ for every $x \in \mathcal{X}$. The probability P splits with respect to $m(dx) = m(x)\mu(dx)$ in the form

$$P(A \times B) = \int_A \int_B \pi(\theta|x)\pi(d\theta)m(dx) \quad A \in \mathcal{F}, B \in \mathcal{G}.$$

The systematic scan Gibbs sampler for drawing from the distribution P proceeds as follows.

- Starting from (x, θ), first, draw x' from $f_\theta(\cdot)$; second, draw θ' from $\pi(\cdot|x')$.

The output is (x', θ'). This generates a Markov chain $(x, \theta) \to (x', \theta') \to \cdots$ having kernel

$$K(x, \theta; x', \theta') = f_\theta(x')f_{\theta'}(x')/m(x')$$

with respect to $\mu(dx')\pi(d\theta')$. A slight variant exchanges the order of the draws.

- Starting from (x, θ), first, draw θ' from $\pi(\cdot|x)$; second, draw x' from $f_{\theta'}(\cdot)$.

The output is (x', θ'). The corresponding Markov chain $(x, \theta) \to (x', \theta') \to \cdots$ has kernel

$$\widetilde{K}(x, \theta; x', \theta') = f_{\theta'}(x)f_{\theta'}(x')/m(x)$$

with respect to $\mu(dx')\pi(d\theta')$. Under mild conditions these two chains have stationary distribution P.

The "x-chain" (from x draw θ' from $\pi(\theta'|x)$ and then x' from $f_{\theta'}(x')$) has transition kernel

$$k(x, x') = \int_\Theta \pi(\theta|x)f_\theta(x')\pi(d\theta) = \int_\Theta \frac{f_\theta(x)f_\theta(x')}{m(x)}\pi(d\theta) \tag{2.1}$$

Note that $\int k(x, x')\mu(dx') = 1$ so that $k(x, x')$ is a probability density with respect to μ. Note further that $m(x)k(x, x') = m(x')k(x', x)$ so that the x chain has $m(dx)$ as a stationary distribution.

The "θ-chain" (from θ, draw x from $f_\theta(x)$ and then θ' from $\pi(\theta'|x)$) has transition density

$$k(\theta, \theta') = \int_\mathcal{X} f_\theta(x)\pi(\theta'|x)\mu(dx) = \int_\mathcal{X} \frac{f_\theta(x)f_{\theta'}(x)}{m(x)}\mu(dx). \tag{2.2}$$

Note that $\int k(\theta,\theta')\pi(d\theta) = 1$ and that $k(\theta,\theta')$ has $\pi(d\theta)$ as reversing measure.

Example (Poisson/Exponential) Let $\mathcal{X} = \{0,1,2,3,\ldots\}$, $\mu(dx)$ = counting measure, $\Theta = (0,\infty)$, $f_\theta(x) = e^{-\theta}\theta^x/x!$. Take $\pi(d\theta) = e^{-\theta}d\theta$. Then $m(x) = \int_0^\infty \frac{e^{-\theta}\theta^x}{x!}\, e^{-\theta}d\theta = 1/2^{x+1}$. The conditional density is $\pi(\theta|x) = f_\theta(x)/m(x) = 2^{x+1}e^{-\theta}\theta^x/x!$. Finally, the x-chain has kernel

$$k(x,x') = \int_0^\infty \frac{2^{x+1}\theta^{x+x'}e^{-2\theta}}{x!x'!}\, d\theta = \frac{2^{x+1}}{3^{x+x'+1}}\binom{x+x'}{x}, \quad 0 \le x, x' < \infty,$$

whereas the θ-chain has kernel

$$k(\theta,\theta') = 2e^{-\theta-\theta'}\sum_{x=0}^{\infty}\frac{(2\theta\theta')^x}{(x!)^2} = 2e^{-\theta-\theta'}I_0\left(\sqrt{4\theta\theta'}\right)$$

where I_0 is the classical modified Bessel function; see Feller [35, Sec. 2.7] for background.

A second construction called the random scan chain is frequently used. From (x,θ), pick a coordinate at random and update it from the appropriate conditional distribution. More formally, for $g \in L^2(P)$

$$\bar{K}g(x,\theta) = \frac{1}{2}\int_\Theta g(x,\theta')\pi(\theta'|x)\pi(d\theta') + \frac{1}{2}\int_{\mathcal{X}} g(x',\theta)f_\theta(x')\mu(dx'). \tag{2.3}$$

We note three things; First, $\bar{K}$ sends $L^2(P) \to L^2(P)$ and is reversible with respect to P. This is the usual reason for using random scans. Second, the right side of (2.3) is the sum of a function of x alone and a function of θ alone. That is $\bar{K} : L^2(P) \to L^2(m) + L^2(\pi)$ (the range of $\bar{K}$ is contained in $L^2(m) + L^2(\pi)$). Third, if $g \in (L^2(m) + L^2(\pi))^\perp$ (complement in $L^2(P)$), then $\bar{K}g = 0$ (Ker $\bar{K} \supseteq (L^2(m) + L^2(\pi))^\perp$). Indeed, for any $h \in L^2(P)$, $0 = \langle g, \bar{K}h\rangle = \langle \bar{K}g, h\rangle$. Thus $\bar{K}g = 0$. We diagonalize random scan chains in Section 3.

2.2 Bounds on Markov chains

2.2.1 General results

We briefly recall well-known results that will be applied to either our two-component Gibbs sampler chains or the x- and θ-chains. Suppose we are given a Markov chain described by its kernel $K(\xi,\xi')$ with respect to a measure $\mu(d\xi')$ (e.g., $\xi = (x,\theta)$, $\mu(d\xi) = \mu(dx)\pi(d\theta)$ for the two component sampler, $\xi = \theta$, $\mu(d\theta) = \pi(d\theta)$ for the θ-chain, etc.). Suppose further that the chain has stationary measure $m(d\xi) = m(\xi)\mu(d\xi)$ and write

$$\bar{K}(\xi,\xi') = K(\xi,\xi')/m(\xi'), \quad \bar{K}^\ell_\xi(\xi') = \bar{K}^\ell(\xi,\xi') = K^\ell(\xi,\xi')/m(\xi')$$

for the kernel and iterated kernel of the chain with respect to the stationary measure $m(d\xi)$. We define the chi-square distance between the distribution of the chain started at ξ after ℓ steps

and its stationary measure by

$$\chi^2_\xi(\ell) = \int |\bar{K}^\ell_\xi(\xi') - 1|^2 m(d\xi') = \int \frac{|K^\ell(\xi,\xi') - m(\xi')|^2}{m(\xi')} \mu(d\xi').$$

This quantity always yields an upper bound on the total variation distance

$$\|K^\ell_\xi - m\|_{\rm TV} = \frac{1}{2}\int |\bar{K}^\ell_\xi(\xi') - 1| m(d\xi') = \frac{1}{2}\int |K^\ell(\xi,\xi') - m(\xi')| \mu(d\xi'),$$

namely,

$$4\|K^\ell_\xi - m\|^2_{\rm TV} \le \chi^2_\xi(\ell).$$

Our analysis will be based on eigenvalue decompositions. Let us first assume that we are given a function ϕ such that

$$K\phi(\xi) = \int K(\xi,\xi')\phi(\xi')\mu(d\xi') = \beta\phi(\xi), \;\; m(\phi) = \int \phi(\xi) m(\xi) \mu(d\xi) = 0$$

for some (complex number) β. In words, ϕ is a generalized eigenfunction with eigenvalue β. We say "generalized" here because we have not assumed here that ϕ belongs to a specific L^2 space (we only assume we can compute $K\phi$ and $m(\phi)$). The second condition (orthogonality to constants in $L^2(m)$) will be automatically satisfied when $|\beta| < 1$. Such an eigenfunction yields simple lower bound on the convergence of the chain to its stationary measure.

Lemma 2.1 *Referring to the notation above, assume that $\phi \in L^2(m(d\xi))$ and $\int |\phi|^2 dm = 1$. Then*

$$\chi^2_\xi(\ell) \ge |\phi(\xi)|^2 |\beta|^{2\ell}.$$

Moreover, if ϕ is a bounded function, then

$$\|K^\ell_\xi - m\|_{\rm TV} \ge \frac{|\phi(\xi)||\beta|^\ell}{2\|\phi\|_\infty}.$$

Proof This follows from the the well-known results

$$\chi^2_\xi(\ell) = \sup_{\|g\|_{2,m} \le 1} \{|K^\ell_\xi(g) - m(g)|^2\} \tag{2.4}$$

and

$$\|K^\ell_\xi - m\|_{\rm TV} = \frac{1}{2} \sup_{\|g\|_\infty \le 1} \{|K^\ell_\xi(g) - m(g)|\}. \tag{2.5}$$

For chi-square, use $g = \phi$ as a test function. For total variation use $g = \phi/\|\phi\|_\infty$ as a test function. More sophisticated lower bounds on total variation are based on the second moment method (e.g., [66, 72]). □

To obtain upper bounds on the chi-square distance, we need much stronger hypotheses. Namely, assume that K is a self-adjoint operator on $L^2(m)$ and that $L^2(m)$ admits an orthonormal basis of real eigenfunctions φ_i with real eigenvalues $\beta_i \geq 0$, $\beta_0 = 1$, $\varphi_0 \equiv 1$, $\beta_i \downarrow 0$ so that

$$\int \bar{K}(\xi, \xi')\varphi_i(\xi')m(d\xi') = \beta_i \varphi_i(\xi).$$

Assume further that K acting on $L^2(m)$ is Hilbert-Schmidt (i.e., $\sum |\beta|_i^2 < \infty$). Then we have

$$\bar{K}^\ell(\xi, \xi') = \sum_i \beta_i^\ell \varphi_i(\xi)\varphi_i(\xi') \quad \text{(convergence in } L^2(m \times m))$$

and

$$\chi_x^2(\ell) = \sum_{i>0} \beta_i^{2\ell} \varphi_i^2(x). \tag{2.6}$$

2.2.2 Application to the two-component Gibbs sampler

All of the bounds in this paper are derived via the following route: bound L^1 by L^2 and use the explicit knowledge of eigenvalues and eigenfunctions to bound the sum in (2.6). This however does not apply directly to the two-component Gibbs sampler K (or $\widetilde{K}$) because these chains are not reversible with respect to their stationary measure. Fortunately, the x-chain and the θ-chain are reversible and their analysis yields bounds on the two component chain thanks to the following elementary observation. The x-chain has kernel $k(x, x')$ with respect to the measure $\mu(dx)$. It will also be useful to have $\bar{k}(x, x') = k(x, x')/m(x')$, the kernel with respect to the probability $m(dx) = m(x)\mu(dx)$. For $\ell \geq 2$, we let $k_x^\ell(x') = k^\ell(x, x') = \int k(x, y)k^{\ell-1}(y, x')\mu(dy)$ denote the density (w.r.t. $\mu(dx)$) of the distribution of the x-chain after ℓ-th and set $\bar{k}_x^\ell(x') = \bar{k}^\ell(x, x') = \int \bar{k}(x, y)\bar{k}^{\ell-1}(y, x')m(dy)$ (the density w.r.t. $m(dx)$).

Lemma 2.2 *Referring to the $K, \widetilde{K}$ two-compoent chains and x-chain introduced in* Section 2.1, *for any $p \in [1, \infty]$, we have*

$$\|(K_{x,\theta}/f) - 1\|_{p,P}^p \leq \int \left\|\bar{k}_z^{\ell-1} - 1\right\|_{p,m}^p f_\theta(z)\mu(dz) \leq \sup_z \left\|\bar{k}_z^{\ell-1} - 1\right\|_{p,m}^p$$

and

$$\|(\widetilde{K}_{x,\theta}/f) - 1\|_{p,P}^p \leq \left\|\bar{k}_x^{\ell-1} - 1\right\|_{p,m}^p .$$

Similarly, for the θ-chain, we have

$$\|(\widetilde{K}_{x,\theta}/f) - 1\|_{p,P}^p \leq \int \left\|k_\theta^{\ell-1} - 1\right\|_{p,\pi}^p \pi(d\theta) \leq \sup_\theta \left\|k_\theta^{\ell-1} - 1\right\|_{p,\pi}^p$$

and

$$\|(K_{x,\theta}/f) - 1\|_{p,P}^p \leq \left\|k_\theta^{\ell-1} - 1\right\|_{p,\pi}^p .$$

Proof We only prove the results involving the x-chain. The rest is similar. Recall that the bivariate chain has transition density

$$K(x,\theta;x',\theta') = f_\theta(x')f_{\theta'}(x')/m(x').$$

By direct computation

$$K^\ell(x,\theta;x',\theta') = \int f_\theta(z)k^{\ell-1}(z,x')\frac{f_{\theta'}(x')}{m(x')}\mu(dz).$$

For the variant $\widetilde{K}$, the similar formula reads

$$\widetilde{K}^\ell(x,\theta;x',\theta') = \int k^{\ell-1}(x,z)\frac{f_{\theta'}(z)}{m(z)}f_{\theta'}(x')\mu(dz).$$

These two bivariate chains have stationary density $f(x,\theta) = f_\theta(x)$ with respect to $\mu(dx)\pi(d\theta)$. So, we write

$$\frac{K^\ell(x,\theta;x',\theta')}{f(x',\theta')} - 1 = \int \left(\bar{k}^{\ell-1}(z,x') - 1\right) f_\theta(z)\mu(dz)$$

and

$$\frac{\widetilde{K}^\ell(x,\theta;x',\theta')}{f(x',\theta')} - 1 = \int \left(\bar{k}^{\ell-1}(x,z) - 1\right) f_{\theta'}(z)\mu(dz).$$

To prove the second inequality in the lemma (the proof of the first is similar), write

$$\begin{aligned}
\|(\widetilde{K}^\ell_{x,\theta}/f) - 1\|^p_{p,P} &= \int\int \left|\int \left(\bar{k}^{\ell-1}(x,z) - 1\right) f_{\theta'}(z)\mu(dz)\right|^p f_{\theta'}(x')\mu(dx')\pi(d\theta') \\
&\leq \int\int\int \left|\bar{k}^{\ell-1}(x,z) - 1\right|^p f_{\theta'}(z)\mu(dz)f_{\theta'}(x')\mu(dx')\pi(d\theta') \\
&\leq \int \left|\bar{k}^{\ell-1}(x,z) - 1\right|^p m(z)\mu(dz) = \int \left|\bar{k}^{\ell-1}(x,z) - 1\right|^p m(dz).
\end{aligned}$$

This gives the desired bound. □

To get lower bounds, we observe the following.

Lemma 2.3 *Let g be a function of x only (abusing notation, $g(x,\theta) = g(x)$). Then*

$$\widetilde{K}g(x,\theta) = \int k(x,x')g(x')\mu(dx').$$

If instead, g is a function of θ only then

$$Kg(x,\theta) = \int k(\theta,\theta')g(\theta')\pi(d\theta').$$

Proof Assume $g(x,\theta) = g(x)$. Then

$$\begin{aligned} \widetilde{K}g(x,\theta) &= \int\int \frac{f_{\theta'}(x) f_{\theta'}(x')}{m(x)} g(x')\mu(dx')\pi(d\theta') \\ &= \int k(x,x')g(x')d\mu(x'). \end{aligned}$$

The other case is similar. □

Lemma 2.4 *Let $\chi^2_{x,\theta}(\ell)$ and $\widetilde{\chi}^2_{x,\theta}(\ell)$ be the chi-square distances after ℓ steps for the K-chain and the $\widetilde{K}$-chain respectively, staring at (x,θ). Let $\chi^2_x(\ell)$, $\chi^2_\theta(\ell)$ be the chi-square distances for x-chain (starting at x) and the θ-chain (starting at θ), respectively. Then we have:*

$$\chi^2_\theta(\ell) \le \chi^2_{x,\theta}(\ell) \le \chi^2_\theta(\ell-1),$$

$$\|k^\ell_\theta - 1\|_{\text{TV}} \le \|K^\ell_{x,\theta} - f\|_{\text{TV}} \le \|k^{\ell-1}_\theta - 1\|_{\text{TV}},$$

and

$$\chi^2_x(\ell) \le \widetilde{\chi}^2_{x,\theta}(\ell) \le \chi^2_x(\ell-1),$$

$$\|k^\ell_x - m\|_{\text{TV}} \le \|\widetilde{K}^\ell_{x,\theta} - f\|_{\text{TV}} \le \|k^{\ell-1}_x - m\|_{\text{TV}}.$$

Proof This is immediate from Lemma 2.3 and (2.4)-(2.5). □

2.3 Exponential families and conjugate priors

Three topics are covered in this section: exponential families, conjugate priors for exponential families and conjugate priors for location families.

2.3.1 Exponential families

Let μ be a σ-finite measure on the Borel sets of the real line $\mathbb{R}$. Define $\Theta = \{\theta \in \mathbb{R} : \int e^{x\theta}\mu(dx) < \infty\}$. Assume that Θ is non-empty and open. Hölder's inequality shows that Θ is an interval. For $\theta \in \Theta$, set

$$M(\theta) = \log \int e^{x\theta}\mu(dx), \quad f_\theta(x) = e^{x\theta - M(\theta)}.$$

The family of probability densities $\{f_\theta,\ \theta \in \Theta\}$ is the exponential family through μ in its "natural parameterization". Allowable differentiations yield the mean $m(\theta) = \int x f_\theta(x)\mu(dx) = M'(\theta)$ and the variance $\sigma^2(\theta) = M''(\theta)$.

Statisticians realized that many standard families can be put in such form so that properties can be studied in a unified way. Standard references for exponential families include [7, 10, 48, 49, 50].

Example Let $\mathcal{X} = \{0, 1, 2, 3, \ldots\}$, $\mu(x) = 1/x!$. Then $\Theta = \mathcal{R}$, and $M(\theta) = e^\theta$,

$$f_\theta(x) = \frac{e^{x\theta - e^\theta}}{x!} \quad x = 0, 1, 2, \ldots \quad .$$

This is the Poisson(λ) distribution with $\lambda = e^\theta$.

2.3.2 Conjugate priors for exponential families

With notation as above, fix $n_0 > 0$ and $x_0 \in$ the interior of the convex hull of the support of μ. Define a prior density with respect to Lebesgue measure $d\theta$ by

$$\pi_{n_0, x_0}(d\theta) = z(x_0, n_0) e^{n_0 x_0 \theta - n_0 M(\theta)} d\theta$$

where $z(n_0, x_0)$ is a normalizing constant shown to be positive and finite in Diaconis and Ylvisaker (1979) which contains proofs of the assertions below. The posterior is

$$\pi(d\theta | x) = \pi_{n_0+1, \frac{n_0 x_0 + x}{n_0+1}}(d\theta).$$

Thus the family of conjugate priors is closed under sampling. This is sometimes used as the definition of conjugate prior. A central fact about conjugate priors is

$$E(m(\theta)|x) = ax + b.$$

This linear expectation property characterizes conjugate priors for families where μ has infinite support. Section 3 shows that linear expectation implies that the associated chain defined at (2.1) always has an eigenvector of the form $x - c$ with eigenvalue a, and c equal to the mean of the marginal distribution.

Example For the Poisson example above the conjugate priors are of form

$$z(n_0, x_0) e^{n_0 x_0 \theta - n_0 e^\theta} d\theta.$$

Setting $\lambda = e^\theta, \theta = \log \lambda, d\theta = d\lambda/\lambda$, the priors transform to

$$z(n_0, x_0) \lambda^{n_0 x_0 - 1} e^{-n_0 \lambda} d\lambda$$

and we see that $z(n_0, x_0) = n_0^{n_0 x_0}/\Gamma(n_0 x_0)$. This is the usual Gamma prior for Poisson(λ).

In the example, the Jacobian of the transformation $\theta \to m(\theta)$ blends in with the rest of the prior so that the same standard priors are used for the mean parameterization. In [18], this is shown to hold only for the six families discussed in Section 2.4 below. See [14, 42] for more on this.

2.3.3 Conjugate priors for location families

Let μ be Lebesgue measure on $\mathbb{R}$ or counting measure on $\mathbb{N}$. In this section we consider random variables of the form $Y = \theta + \epsilon$, with θ having density $\pi(\theta)$ and ϵ having density $g(x)$ (both with respect to μ). This can also be written as (densities w.r.t. $\mu(dx) \times \mu(d\theta)$)

$$f_\theta(x) = g(x-\theta), \quad f(x,\theta) = g(x-\theta)\pi(\theta)$$

In [24], a family of 'conjugate priors' π is suggested via posterior linearity. See [56] for further developments. The idea is to use the following well known fact: If X and Y are independent random variables with finite means and the same distribution, then $E(X|X+Y) = (X+Y)/2$. More generally, if X_r and X_s are random variables which are independent with X_r (resp X_s) having the distribution of the sum of r (resp s) independent copies of the same random variable Z then $E(X_r|X_r+X_s) = \frac{r}{r+s}(X_r+X_s)$. Here r and s may be taken as any positive real numbers if the underlying Z is infinitely divisible.

With this notation, take g as the density for X_r and π as the density for X_s and call these a *conjugate location pair*. Then the marginal density $m(y)$ is the convolution of g and π.

Example Let $g(x) = e^{-\lambda}\lambda^x/x!$ for $x \in \mathcal{X} = \{0,1,2,\ldots\}$. Take $\Theta = \mathcal{X}$ and let $\pi(\theta) = e^{-\eta}\eta^\theta/\theta!$. Then $m(x) = e^{-(\lambda+\eta)}(\lambda+\eta)^x/x!$ and

$$\pi(\theta|x) = \binom{x}{\theta}\left(\frac{\eta}{\lambda+\eta}\right)^\theta\left(\frac{\lambda}{\lambda+\eta}\right)^{x-\theta}, \quad 0 \le \theta \le x < \infty.$$

The Gibbs sampler (bivariate chain K) for this example becomes

- From x, choose θ from Binomial$(x, \lambda/(\eta+\lambda))$.
- From θ, choose $x = \theta + \epsilon$ with $\epsilon \sim$ Poisson(λ).

The x-chain may be represented as $X_{n+1} = S_{X_n} + \epsilon_{n+1}$ with $S_k \sim$ Binomial$(x, \lambda/(\eta+\lambda))$ and $\epsilon \sim$ Poisson(λ). This also represents the number of customers on service in a $M/M/\infty$ queue observed at discrete times: If this is X_n at time n, then S_{x_n} is the number served in the next time period and ϵ_{n+1} is the number of unserved new arrivals. The exlicit diagonalization of the $M/M/\infty$ chain, in continuous time, using Charlier polynomials appears in [3].

This same chain has yet a different interpretation: Let $f_\eta(j) = \binom{\eta}{j}p^j(1-p)^{\eta-j}$. Here $0 < p < 1$ is fixed and $\eta \in \{0,1,2,\ldots\}$ is a parameter. This model arises in under-reporting problems where the true sample size is unknown. See [58]. Let η have a Poisson(λ) prior. The Gibbs sampler for the bivariate distribution $f(j,\eta) = \binom{\eta}{j}p^j(1-p)^{\eta-j}e^{-\lambda}\lambda^\eta/\eta!$ goes as follows:

- From η, choose j from Bin(η, p)
- From j, choose $\eta = j + \epsilon$ with $\epsilon \sim$ Poisson$(\lambda(1-p))$.

Up to a simple renaming of parameters, this is the same chain discussed above. Similar 'translations' hold for any location problem where $\pi(\theta|x)$ has bounded range.

2.4 The six families

Morris [59, 60] has characterized exponential families where the variance $\sigma^2(\theta)$ is a quadratic function of the mean: $\sigma^2(\theta) = v_0 + v_1 m(\theta) + v_2 m^2(\theta)$. These six families have been characterized earlier by Meixner [57] in the development of a unified theory of orthogonal polynomials via generating functions. In [43] the same families are characterized in a regression context: For X_i independent with a finite mean, $\bar{X} = \frac{1}{n}\sum X_i, S_n^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$, one has

$$E(S_n^2 | \bar{X} = \bar{x}) = a + b\bar{x} + c\bar{x}^2$$

if and only if the distribution of X_i is one of the six families. In [30, 31], the six families are characterized by a link between orthogonal polynomials and martingales whereas [32] makes a direct link to Lie theory. Finally, Consonni and Veronese [18] find the same six families in their study of conjugate priors: The conjugate priors in the natural parameterization given above transform into the same family in the mean parameterization only for the six families.

Extensions are developed by Letac and Mora [52] and Casalis [14] who give excellent surveys of the literature. Still most useful, Morris [59, 60] gives a unified treatment of basic (and not so basic) properties such as moments, unbiased estimation, orthogonal polynomials and statistical properties. We give the six families in their usual parameterization along with the conjugate prior and formula for the moments $E_\theta(X^k)$, $E_x(\theta^k)$ of x and θ under $dP = f_\theta(x) d\mu(x) \pi(d\theta)$, given the value of the other.

Binomial: $\mathcal{X} = \{0, \ldots, n\}$, μ counting measure, $\Theta = [0, 1]$.

$$\begin{aligned}
f_\theta(x) &= \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad 0 < \theta < 1. \\
\pi(d\theta) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta, \quad 0 < \alpha, \beta < \infty. \\
E_\theta(X^k) &= \sum_{j=0}^{k} a_j \theta^j, \; a_j = n(n-1)\cdots(n-j+1), \; 0 \le k \le n. \\
E_x(\theta^k) &= \sum_{j=0}^{k} a_j x^j, \; a_j = [(\alpha+\beta+n)(\alpha+\beta+n+1)\ldots(\alpha+\beta+n+j-1)]^{-1}.
\end{aligned}$$

Poisson: $\mathcal{X} = \mathbb{N}$, μ counting measure, $\Theta = (0, \infty)$.

$$\begin{aligned}
f_\theta(x) &= \frac{e^{-\theta}\theta^x}{x!}, \quad 0 < \theta < \infty. \\
\pi(d\theta) &= \frac{\theta^{a-1} e^{-\theta/\alpha}}{\Gamma(a)\alpha^a} d\theta, \quad 0 < \alpha, a < \infty. \\
E_\theta(X^k) &= \sum_{j=0}^{k} a_j \theta^j, a_j = 1, \quad E_x(\theta^k) = \sum_{j=0}^{k} a_j x^j, \; a_j = \Big(\frac{\alpha}{\alpha+1}\Big)^j.
\end{aligned}$$

Negative Binomial: $\mathcal{X} = \mathbb{N}$, μ counting measure, $\Theta = [0,1]$.

$$\begin{aligned}
f_\theta(x) &= \frac{\Gamma(x+r)}{\Gamma(r)x!}\,\theta^x(1-\theta)^r, 0<\theta<1,\ r>0.\\
\pi(d\theta) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta,\ \ 0<\alpha,\ \beta<\infty.\\
E_\theta(X^k) &= \sum_{j=0}^{k} a_j\Big(\frac{\theta}{1-\theta}\Big)^j, a_k = r(r+1)\cdots(r+k-1).\\
E_x\left(\left(\frac{\theta}{1-\theta}\right)^k\right) &= \sum_{j=0}^{k} a_j x^j,\ \ a_k = [(\beta+r-1)(\beta+r-2)\cdots(\beta+r-k)]^{-1},\ \ k<\beta+r.
\end{aligned}$$

Normal: $\mathcal{X} = \Theta = \mathbb{R}$, μ Lebesgue measure.

$$\begin{aligned}
f_\theta(x) &= \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{1}{2}(x-\theta)/\sigma^2},\ \ 0<\sigma^2<\infty\\
\pi(d\theta) &= \frac{1}{\sqrt{2\pi\tau^2}}\,e^{-\frac{1}{2}(\theta-v)^2/\tau^2}d\theta,\ -\infty<v<\infty,\ 0<\tau<\infty.\\
E_\theta(X^k) &= \sum_{j=0}^{k} a_j\theta^j,\ \ a_k = 1\\
E_x(\theta^k) &= \sum_{j=0}^{k} a_j x^j,\ \ a_k = (\tau^2/(\tau^2+\sigma^2))^k.
\end{aligned}$$

Gamma: $\mathcal{X} = \Theta = (0,\infty)$, μ Lebesgue measure.

$$\begin{aligned}
f_\theta(x) &= \frac{x^{a-1}e^{-x/\theta}}{\theta^a\Gamma(a)},\ \ 0<a<\infty.\\
\pi(d\theta) &= \frac{c^b\theta^{-(b+1)}e^{-c/\theta}}{\Gamma(b)}d\theta,\ \ 0<b,c<\infty.\\
E_\theta(X^k) &= a\cdots(a+k-2)(a+k-1)\theta^k\\
E_x(\theta^k) &= \sum_{j=0}^{k} a_j x^j,\ \ a_k = [(a+b-1)(a+b-2)\cdots(a+b-k)]^{-1},\ 0\le k<a+b.
\end{aligned}$$

Hyperbolic: $\mathcal{X} = \Theta = \mathbb{R}$, μ Lebesgue measure.

$$\begin{aligned}
f_\theta(x) &= \frac{2^{r-2}}{\pi r(1+\theta^2)^{r/2}}\,e^{rx\tan^{-1}\theta}\beta\left(\frac{r}{2}+\frac{irx}{2},\ \frac{r}{2}-\frac{irx}{2}\right),\ \ r>0.\\
\pi(d\theta) &= \frac{\Gamma\left(\frac{\rho}{2}-\frac{\rho\delta i}{2}\right)\Gamma\left(\frac{\rho}{2}+\frac{\rho\delta i}{2}\right)}{\Gamma\left(\frac{\rho}{2}\right)\Gamma\left(\frac{\rho}{2}-\frac{1}{2}\right)\sqrt{\pi}}\ \frac{e^{\rho\delta\tan^{-1}\theta}}{(1+\theta^2)^{\rho/2}}d\theta, \qquad -\infty<\delta<\infty,\ \rho\ge 1.
\end{aligned}$$

$$E_\theta(X^k) = \sum_{j=0}^{k} a_j \theta^j, \quad a_k = k!.$$

$$E_x(\theta^k) = \sum_{j=0}^{k} a_j x^j, \quad a_k = r^k[(r+\rho-2)\cdots(r+\rho-(k+1))]^{-1}, \quad 0 < k \le r+\rho-1$$

Proof A unified way to prove the formulas involving $E_\theta(X^k)$ follows from Morris (1983, (3.4)). This says, for any of the six families with $m(\theta)$ the mean parameter and $p_k(x, m_0)$ the monic, orthogonal polynomials associated to the parameter θ_0,

$$E_\theta(p_k(x, m_0)) = b_k(m(\theta) - m(\theta_0))^k,$$

where, if the family has variance function $\sigma^2(\theta) = v_2 m^2(\theta) + v_1 m(\theta) + v_0$,

$$b_k = \prod_{i=0}^{k-1}(1 + iv_2).$$

For example, for the Binomial(n, p) family, $m(p) = np$, $\sigma^2(p) = np(1-p)$, so $v_2 = -1/n$ and

$$E_p(p_k(x, m_0)) = \Big\{ \prod_{i=0}^{k-1}(n-i) \Big\}(p - p_0)^k.$$

Comparing lead terms and using induction, gives the first binomial entry. The rest are similar; the values of v_2 are $v_2(\text{Poisson}) = 0$, $v_2(\text{NB}) = 1/r$, $v_2(\text{Normal}) = 0$, $v_2(\text{Gamma}) = 1/r$, $v_2(\text{Hyperbolic}) = 1$. Presumably, there is a unified way to get the $E_x(\theta^k)$ entries, perhaps using [60, Th 5.4]. This result shows that we get polynomials in x but the lead coefficients do not come out as easily. At any rate they all follow from elementary computations. □

Remarks 1. The moment calculations above are transformed into a singular value decomposition and an explicit diagonalization of the univariate chains (x-chain, θ-chain) in Section 3.

2. Note that not all moments are finite. Indeed, consider the geometric $f_\theta(x) = \theta^x(1-\theta)$ with a uniform prior. The marginal is $\int_0^1 \theta^x(1-\theta)d\theta = 1/(x+1)(x+2)$ on $0 \le x < \infty$. This admits no moments. None the less, the moments that are available are put to good use in [20].

3. The first five families are very familiar, the sixth family less so. As one motivation, consider the generalized arc sine densities

$$f_\theta(y) = y^{a-1}(1-y)^{(1-a)-1}\Gamma(a)\Gamma(1-a) \qquad 0 \le y,\ a < 1.$$

Transform these to an exponential family via $x = \log(y/(1-y))$, $\eta = \pi a - \pi/2$. This has density

$$g_\eta(x) = \frac{e^{x\eta + \log(\cos\eta)}}{2\cosh(\frac{\pi}{2}x)}, \quad -\infty < x < \infty, \quad -\frac{\pi}{2} < \eta < \frac{\pi}{2}.$$

The appearance of cosh explains the name hyperbolic. This density appears in [35, pg. 503] as an example of a density which is its own Fourier transform (like the normal). Many further references are in [28, 59, 60]. In particular, $g_0(x)$ is the density of $\frac{2}{\pi}\log|C|$ with C standard Cauchy. The mean of $g_\eta(x)$ is $\tan(\eta) = \theta$. Parameterizing by the mean leads to the density shown with $r = 1$. The average of r independent copies of independent variates with $r = 1$ gives the density with general r. The beta function is defined as usual; $\beta(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$. Because $\overline{\Gamma(a)} = \Gamma(\bar{a})$, the norming constant is real valued.

The conjugate prior for the mean parameter is of Pearson Type IV. When $\delta = 0$ this is a rescaled t density. For general δ the family is called the skew t in [28] which contains a wealth of information. Under the prior, the parameter θ has mean $\rho\delta/(\rho - 2)$ and satisfies

$$(\rho - (k + 2))E(\theta^{k+1}) = kE(\theta^{k-1}) + \rho\delta E(\theta^k), \qquad 1 \le k < \rho - 2.$$

This makes it simple to compute the $E_x(\theta^k)$ entry. Moments past ρ are infinite.

The marginal distribution $m(x)$ can be computed in closed form. Using Stirling's formula in the form $|\Gamma(\sigma + it)| \sim \sqrt{2\pi}\, e^{-\pi|t|/2}|t|^{\sigma-\frac{1}{2}}$. As $|t| \uparrow \infty$ shows that $m(x)$ has tails asymptotic to c/x^ρ. It thus has only finitely many moments, so the x-chain must be studied by non-spectral methods. Of course, the additive version of our set-up has moments of all order. We give a brief treatment in Section 6. The relevant orthogonal polynomials being Meixner-Pollaczek.

2.5 Some background on orthogonal polynomials

A variety of orthogonal polynomials are used crucially in the following sections. While we usually just quote what we need from the extensive literature, this section describes a simple example. Perhaps the best introduction is in [17]. We will make frequent reference to [44] which is through and up to date. The classical account [68] contains much that is hard to find elsewhere. The on line account [47] is very useful. For pointers to the literature on orthogonal polynomials and birth and death chains, see, e.g., [71].

As an indication of what we need, consider the Beta/Binomial example with a general Beta(α, β) prior. Then the stationary distribution for the x-chain on $\mathcal{X} = \{0, 1, 2, \ldots, n\}$ is

$$m(x) = \binom{n}{x}\frac{(\alpha)_x(\beta)_{n-x}}{(\alpha + \beta)_n} \qquad \text{where } (a)_x = \frac{\Gamma(a + x)}{\Gamma(a)} = a(a + 1)\cdots(a + x - 1).$$

The choice $\alpha = \beta = 1$ yields the uniform distribution while $\alpha = \beta = 1/2$ yields the discrete arc-sine density from [34, Chap. 3],

$$m(x) = \frac{\binom{2x}{x}\binom{2n-2x}{n-x}}{2^{2n}}.$$

The orthogonal polynomials for m are called Hahn polynomials. They are developed by [44, Sec. 6.2] who refers to the very useful treatment of Karlin and McGregor [46]. The polynomials are

given explicitly in [44, pg. 178–179]. Shifting parameters by one to make the classical notation match present notation, the orthogonal polynomials are

$$Q_\ell(x) = {}_3F_2\left(\begin{matrix} -\ell, \ell+\alpha+\beta-1, -x \\ \alpha, -n \end{matrix} \,\middle|\, 1 \right), \qquad 0 \le \ell \le n.$$

Here

$${}_rF_s\left(\begin{matrix} a_1 \dots a_r \\ b_1 \dots b_s \end{matrix} \,\middle|\, z \right) = \sum_{n=0}^{\infty} \frac{(a_1 a_2 \dots a_r)_n}{(b_1 b_2 \dots b_s)_n} \frac{z^n}{n!} \quad \text{with } (a_1 \dots a_r)_n = \prod_{i=1}^{r} (a_i)_n.$$

These polynomials satisfy

$$E_m(Q_\ell Q_m) = \delta_{\ell m} \frac{\ell!(n-\ell)!(\beta_\ell(\alpha+\beta+\ell-1))_{n+1}(\alpha+\beta)_n}{(n!)^3(\alpha+\beta+2\ell-1)(\alpha_\ell}$$

When $\alpha = \beta = 1$ these become the discrete Chebychev polynomials cited in Proposition 1.1. From our work in Section 2.2, we see we only need to know $Q_\ell(x_0)$ with x_0 the starting position. This is often available in closed form for special values, e.g., for $x_0 = 0$ and $x_0 = n$,

$$Q_\ell(0) = 1, \quad Q_\ell(n) = \frac{(-\beta-\ell)_\ell}{(\alpha+1)_\ell}, \;\; 0 \le \ell \le n. \tag{2.7}$$

For general starting values, one may draw on the extensive work on uniform asymptotics; see e.g. [68, Chap. 8] or [5].

We note that [59, Sect. 8] gives an elegant self-contained development of orthogonal polynomials for the six families. Briefly, if $f_\theta(x) = e^{x\theta - M(\theta)}$ is the density, then

$$p_k(x,\theta) = \sigma^{2k} \Big\{ \frac{d^k}{d^k m} f_\theta(x) \Big\} \Big/ f_\theta(x)$$

(derivatives with respect to the mean $m(\theta)$). If $\sigma^2(\theta) = v_2 m^2(\theta) + v_1 m(\theta) + v_0$ then

$$E_\theta(p_n p_k) = \delta_{nk} a_k \sigma^{2k} \text{ with } a_k = k! \prod_{i=0}^{k-1} (1 + i v_2).$$

We also find need for orthogonal polynomials for the conjugate priors $\pi(\theta)$.

3 A singular value decomposition

The results of this section show that the Gibbs sampler Markov chains associated to the six families have polynomial eigenvectors, with explicitly known eigenvalues. This includes the x-chain, θ-chain and the random scan chain. Analysis of these chains is in Sections 4 and 5. Section 6 explains the connection with the compactness of the associated operators. For

a discussion of Markov operators related to orthogonal polynomials, see, e.g., [6]. For closely related statistical literature, see [12] and the references therein.

Throughout, notation is as in Section 2.1. We have $\{f_\theta(x)\}_{\theta\in\Theta}$ a family of probability densities on the real line $\mathbb{R}$ with respect to a σ-finite measure $\mu(dx)$, for $\theta \in \Theta \subseteq \mathbb{R}$. Further, $\pi(d\theta)$ is a probability measure on Θ. These define a joint probability P on $\mathbb{R}\times\Theta$ with marginal density $m(x)$ (w.r.t. μ) and conditional density $\pi(\theta|x)(w.r.t.\pi)$ given by $\pi(\theta|x) = f_\theta(x)/m(x)$.

Let $c = \#\text{supp } m(x)$. This may be finite or infinite. For simplicity, throughout this section, we assume $\text{supp}(\pi)$ is infinite. Moreover we make the following hypotheses:

(H1) For some $\alpha_1, \alpha_2 > 0$, $\int e^{\alpha_1|x|+\alpha_2|\theta|}P(dx, d\theta) < \infty$.

(H2) For $0 \le k < c$, $E_\theta(X^k)$ is a polynomial in θ of degree k with lead coefficient $\eta_k > 0$.

(H3) For $0 \le k < \infty$, $E_x(\theta^k)$ is a polynomial in x of degree k with lead coefficient $\mu_k > 0$.

By (H1), $L^2(m(dx))$ admits a unique monic, orthogonal basis of polynomials p_k, $0 \le k < c$, with p_k of degree k. Also, $L^2(\pi(d\theta))$ admits a unique monic, orthogonal basis of polynomials q_k, $0 \le k < \infty$, with q_k of degree k. As usual, $\eta_0 = \mu_0 = 1$ and $p_0 \equiv q_0 \equiv 1$.

Theorem 3.1 *Assume* (H.1)-(H.3). *Then*

(a) The x-chain (2.1) *has eigenvalues* $\beta_k = \eta_k\mu_k$ *with eigenvectors* p_k, $0 \le k < c$.

(b) The θ-chain (2.2) *has eigenvalues* $\beta_k = \eta_k\mu_k$ *with eigenvectors* q_k *for* $0 \le k \le c$, *and eigenvalues zero with eigenvectors* q_k *for* $c < k < \infty$.

(c) The random scan chain (2.3) *has spectral decomposition given by*

$$\textit{eigenvalues } \frac{1}{2} \pm \frac{1}{2}\sqrt{\eta_k\mu_k}, \ \textit{eigenvectors } p_k(x) \pm \sqrt{\frac{\eta_k}{\mu_k}}q_k, \ 0 \le k < c$$

$$\textit{eigenvalues } \frac{1}{2}, \ \textit{eigenvectors } q_k \quad c \le k < \infty.$$

The proof will follow from a sequence of propositions. The first shows that the expectation operator with respect to f_θ takes orthogonal polynomials into orthogonal polynomials.

Proposition 3.2 $E_\theta[p_k(X)] = \eta_k q_k(\theta)$, $0 \le k < c$.

Proof For $k = 0, E_\theta[p_0] = 1 = \eta_0 q_0$. For $0 < k < c$, note that for $0 \le i < k$, the unconditional expectation is given by

$$E[\theta^i p_k(X)] = E[p_k(X)E(\theta^i|X)] = E[p_k(X)\widehat{p}(X)]$$

with $\widehat{p}$ a polynomial of degree $i < k$. By orthogonality, since $0 \le i < k < c$, $E[p_k(X)\widehat{p}(X)] = 0$. Thus $0 = E[\theta^i p_k(X)] = E[\theta^i E_\theta(p_k(X))]$. By assumption (H.2), $\eta_k^{-1}E_\theta[p_k(X)]$ is a monic polynomial of degree k in θ. Since it is orthogonal to all polynomials of degree less than k, we must have $E_\theta[p_k(X)] = \eta_k q_k(\theta)$. □

The second proposition is dual to the first.

Proposition 3.3 *$E_x[q_k(\theta)] = \mu_k p_k(x)$, $0 \le k < c$. If $c < \infty$, $E_x(q_k(\theta)) = 0$ for $k \ge c$.*

Proof The first part is proved as per Proposition 3.2. If $c < \infty$, and $k \ge c$, by the same argument we have, for $0 \le j < c$, $E[p_j(X)E_x[q_k(\theta)]] = 0$. But $\{p_j\}_{0 \le j < c}$ form a basis for $L^2(m(dx))$. and $E_x[q_k(\theta)] \in L^2(m(dx))$ since

$$E[(E_x(q_k(\theta))^2] \le E[q_k^2(\theta)] < \infty.$$

It follows that $E_x[q_k(\theta)] = 0$. □

Proof of Part (a) of Theorem 3.1 Suppose $0 \le k < c$. From the definitions, the x-chain operates on p_k as

$$E_x[E_\theta(p_k(X'))] = E_x[\eta_k q_k(\theta)] = \eta_k \mu_k p_k(x)$$

with equalities from Propositions 3.2, 3.3. Hence, $\eta_k \mu_k$ are eigenvalues of the x chain with p_k as eigenfunctions. This proves (a). □

Proof of Part (b) Suppose first $0 \le k < c$. Then, arguing as above, $\mu_k \eta_k$ are eigenvalues of the θ-chain with q_k as eigenvectors. If $c = \infty$, we are done. If $c < \infty$, then, for $k \ge c$, Proposition 3.3 shows that q_k is an eigenfunction for the θ-chain with eigenvalue zero. □

Proof of Part (c) From the development in Section 2.1, the random scan chain K takes $L^2(P)$ into $L^2(m) + L^2(\pi) \subseteq L^2(P)$ and ker K $\supseteq (L^2(m) + L^2(\pi))^\perp$. We have

$$Kg(X, \theta) = \frac{1}{2} E_x[g(x, \theta')] + \frac{1}{2} E_\theta[g(X', \theta)].$$

For $0 \le k < c$, consider K acting on $p_k(x) + \sqrt{\frac{\eta_k}{\mu_k}}\, q_k(\theta)$. The result is

$$\frac{1}{2}\Big(p_k(x) + E_x[q_k(\theta')]\sqrt{\frac{\eta_k}{\mu_k}}\Big) + \frac{1}{2}\Big(E_\theta[p_k(x)] + \sqrt{\frac{\eta_k}{\mu_k}}\, q_k(\theta)\Big) = \Big(\frac{1}{2} + \frac{1}{2}\sqrt{\eta_k \mu_k}\Big)\left(p_k(x) + \sqrt{\frac{\eta_k}{\mu_k}}\, q_k(\theta)\right).$$

Similarly,

$$K\Big(p_k - \sqrt{\frac{\eta_k}{\mu_k}}\, q_k\Big)(x, \theta) = \Big(\frac{1}{2} - \frac{1}{2}\sqrt{\eta_k \mu_k}\Big)\Big(p_k(x) - \sqrt{\frac{\eta_k}{\mu_k}}\, q_k(\theta)\Big).$$

Suppose first that $c < \infty$. For $k \ge c$, Proposition 3.3 shows $E_x q_k(\theta) = 0$ for all x. Thus $Kq_k(x, \theta) = \frac{1}{2} q_k(\theta)$. Further

$$\begin{aligned} &\text{span}\left\{p_k(x) \pm \sqrt{\tfrac{\eta_k}{\mu_k}}\, q_k(\theta) \quad 0 \le k < c,\ q_k(\theta) \quad c \le k < \infty\right\} \\ &= \text{span}\ \{p_k(x) \quad 0 \le k < c,\ q_k(\theta),\ 0 \le k < \infty\} = L^2(m) + L^2(\pi). \end{aligned}$$

It follows that K is diagonalizable with eigenvalues/eigenvectors

$$\frac{1}{2} \pm \frac{1}{2}\sqrt{\mu_k \eta_k}, \quad p_k(x) \pm \sqrt{\frac{\eta_k}{\mu_k}} q_k(\theta), \quad \text{for } 0 \le k < c,$$
$$\frac{1}{2}, \quad q_k(\theta), \text{ for } c \le k < \infty,$$

and $Kg = 0$ for $g \in (L^2(m) + L^2(\pi))^\perp$.

Suppose next that $c = \infty$, then, K is diagonalizable with eigenvalues/eigenfunctions

$$\Big(\frac{1}{2} \pm \sqrt{\eta_k \mu_k}\Big), \quad p_k(x) \pm \sqrt{\frac{\eta_k}{\mu_k}}\, q_k(\theta), \quad 0 \le k < \infty.$$

Again span $\Big\{p_k(x) \pm \sqrt{\frac{\eta_k}{\mu_k}}\, q_k(\theta) \quad 0 \le k < c\Big\}$ = span $\{p_k(x), q_k(\theta)\} = L^2(m) + L^2(\pi)$ and $Kg = 0$ for $g \in (L^2(m) + L^2(\pi))^\perp$. This completes the proof of (c). □

Remark The theorem holds with obvious modification if #supp $(\pi) < \infty$. This occurs for binomial location problems. It will be used without further comment in Section 5. Further, the arguments work to give some eigenvalues with polynomial eigenvectors when only finitely many moments are finite.

4 Exponential family examples

This section carries out the analysis of the Gibbs sampler for five examples. The x and θ chains for the beta/binomial, Poisson/Gamma and normal families. For each, we set up the results for general parameter values and carry out the bounds in some natural special cases.

4.1 Beta/Binomial

4.1.1 The x-chain for the Beta/Binomial

The Gibbs sampler for this chain was used as a simple expository example in [15]. The case of a uniform prior appears in Section 1 above. Fix $\alpha, \beta > 0$. On the state space $\mathcal{X} = \{0, 1, 2, \ldots, n\}$, let

$$\begin{aligned} k(x,y) &= \int_0^1 \binom{n}{y} \theta^{\alpha+x+y-1}(1-\theta)^{\beta+2n-(x+y)-1} \frac{\Gamma(\alpha+\beta+n)\, d\theta}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} \\ &= \binom{n}{y} \frac{\Gamma(\alpha+\beta+n)\Gamma(\alpha+x+y)\Gamma(\beta+2n-(x+y))}{\Gamma(\alpha+x)\Gamma(\beta+n-y)\Gamma(\alpha+\beta+2n)}. \end{aligned} \tag{4.1}$$

When $\alpha = \beta = 1$ (uniform prior), $k(x, x')$ is given by (2.1). For general α, β, the stationary distribution is the Beta/Binomial:

$$m(x) = \binom{n}{x} \frac{(\alpha)_x (\beta)_{n-x}}{(\alpha+\beta)_n} \quad \text{where} \quad (a)_j = \frac{\Gamma(a+j)}{\Gamma(a)} = a(a+1)\cdots(a+j-1).$$

From our work in previous sections we obtain the following result.

Proposition 4.1 *For $n = 1, 2, \ldots,$ and $\alpha, \beta > 0$, the Beta/Binomial x-chain* (4.1) *has:*

(a) Eigenvalues $\beta_0 = 1$ and $\beta_j = \frac{n(n-1)\cdots(n-j+1)}{(\alpha+\beta+n)_j} \quad 1 \le j \le n$.

(b) Eigenvectors Q_j, $0 \le j \le n$, the Hahn polynomials of Section 2.5.

(c) For any $\ell \ge 1$ and any starting state x

$$\chi_x^2(\ell) = \sum_{i=1}^{n} \beta_i^{2\ell} Q_i^2(x) z_i, \qquad z_i = \frac{(\alpha+\beta+2i-1)(\alpha+\beta)_n(\alpha)_i}{(\beta)_i(\alpha+\beta+i-1)_{n+1}} \binom{n}{i}.$$

We now specialize this to $\alpha = \beta = 1$ and prove the bounds announced in Proposition 1.1.

Proof of Proposition 1.1 From (a), $\beta_i = \frac{n(n-1)\cdots(n-i+1)}{(n+2)(n+3)\cdots(n+i+1)}$. From (2.7), $Q_i^2(n) = 1$. By elementary manipulations, $z_i = \beta_i(2i+1)$. Thus

$$\chi_n^2(\ell) = \sum_{y=0}^{n} \frac{(k^\ell(n,y) - m(y))^2}{m(y)} = \sum_{i=1}^{n} \beta_i^{2\ell+1}(2i+1).$$

We may bound $\beta_i \le \beta_1^i = \left(1 - \frac{2}{n+2}\right)^i$, and so

$$\chi_n^2(\ell) = \sum_{i=1}^{n} \beta_i^{2\ell+1}(2i+1) \le \sum_{i=1}^{n} \beta_1^{i(2\ell+1)}(2i+1)$$

Using $\sum_1^\infty x^i = 1/(1-x)$, $\sum_1^\infty i x^i = x/(1-x)^2$, we obtain

$$3\beta_1^{2\ell+1} \le \chi_n^2(\ell) \le \frac{3\beta_1^{2\ell+1}}{(1-\beta_1^{2\ell+1})^2} \le 27\beta_1^{2\ell+1}.$$

By Lemma 2.4, this gives (for the $\widetilde{K}$ chain)

$$3\beta_1^{2\ell+1} \le \widetilde{\chi}_{n,\theta}^2(\ell) \le 27\beta_1^{2\ell-1}.$$

For a lower bound in total variation, use the eigenfunction $\varphi_1(x) = x - \frac{n}{2}$. This is maximized at $x = n$ and the lower bound follows from Lemma 2.1. □

Remark Essentially, the same results hold for any Beta(α, β) prior in the sense that, for fixed α, β, starting at n, order n steps are necessary and sufficient for convergence.

4.1.2 The θ-Chain for the Beta/Binomial

Fix $\alpha, \beta > 0$. On the state space $[0,1]$, let

$$k(\theta,\eta) = \sum_{j=0}^{n} \binom{n}{j} \theta^j (1-\theta)^{n-j} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+j)\Gamma(\beta+n-j)}\ \eta^{\alpha+j-1}(1-\eta)^{\beta+n-j-1}. \tag{4.2}$$

This is a transition density with respect to Lebesgue measure $d\eta$ on $[0,1]$. It has stationary density

$$\pi(d\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\ \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta.$$

The relevant orthogonal polynomials are Jacobi polynomials $P_i^{a,b}$, $\alpha = a-1, \beta = b-1$, given on $[-1,1]$ in standard literature [47, 1.8]. We make the change of variables $\theta = (x+1)/2$ and write $p_i(\theta) = P_i^{\alpha-1,\beta-1}(2\theta-1)$. Then, we have

$$\int_0^1 p_j(\theta)p_k(\theta)\pi(\theta)d\theta = \frac{1}{2j+\alpha+\beta-1}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(j+\alpha)\Gamma(j+\beta)}{\Gamma(j+\alpha+\beta-1)j!}\delta_{jk} = z_i^{-1}\delta_{jk}. \tag{4.3}$$

This defines z_i.

Proposition 4.2 *For $\alpha, \beta > 0$, the θ-chain for the Beta/Binomial (4.2) has:*

(a) Eigenvalues $\beta_0 = 1$, $\beta_j = \frac{n(n-1)\cdots(n-j+1)}{(\alpha+\beta+n)_j}$ $\quad 1 \le j \le n, \quad \beta_j = 0$ for $j > n$.

(b) Eigenvectors p_j, the shifted Jacobi polynomials.

(c) With z_i from (4.3), for any $\ell \ge 1$ and any starting state $\theta \in [0,1]$,

$$\chi^2_\theta(\ell) = \sum_{i=1}^{n} \beta_i^{2\ell} p_i^2(\theta) z_i.$$

The following proposition gives sharp chi-square bounds, uniformly over α, β, n in two cases: (i) $\alpha \ge \beta$, starting from 1 (worst starting point), (ii) $\alpha = \beta$, starting from $1/2$ (heuristically, the most favorable starting point). The restriction $\alpha \ge \beta$ is not really a restriction because of the symmety $P_i^{a,b}(x) = (-1)^i P_i^{b,a}(-x)$. For $\alpha \ge \beta > 1/2$, it is known (e.g., [44, Lemma 4.2.1]) that

$$\sup_{[0,1]} |p_i| = \sup_{[-1,1]} |P_i^{\alpha-1,\beta-1}| = p_i(1) = \frac{(\alpha)_i}{i!}.$$

Hence, 1 is clearly the worst starting point from the viewpoint of convergence in chi-square distance, that is,

$$\sup_{\theta\in[0,1]} \{\chi^2_\theta(\ell)\} = \chi^2_1(\ell).$$

Proposition 4.3 *For $\alpha \geq \beta > 0$, $n > 0$, set $N = \log[(\alpha+\beta)(\alpha+1)/(\beta+1)]$. The θ-chain for the Beta/Binomial* (4.2) *satisfies:*

(i)
- $\chi_1^2(\ell) \leq 7e^{-c}$ *for* $\ell \geq \frac{N+c}{-2\log\beta_1}$, $c > 0$.
- $\chi_1^2(\ell) \geq \frac{1}{6}e^c$, *for* $\ell \leq \frac{N-c}{-2\log\beta_1}$, $c > 0$.

(ii) Assuming $\alpha = \beta > 0$,
- $\chi_{1/2}^2(\ell) \leq 13\beta_2^{2\ell}$ *for* $\ell \geq \frac{1}{-2\log\beta_2}$.
- $\chi_{1/2}^2(\ell) \geq \frac{1}{2}\beta_2^{2\ell}$, *for* $\ell > 0$.

Roughly speaking, part (i) says that, starting from 1, $\ell(\alpha,\beta,n)$ steps are necessary and sufficient for convergence in chi-square distance where

$$\ell(\alpha,\beta,n) = \frac{\log[(\alpha+\beta)(\alpha+1)/(\beta+1)]}{-2\log(1-(\alpha+\beta)/(\alpha+\beta+n))}.$$

Note that if $\alpha, n, n/\alpha$ tend to infinity and β is fixed,

$$\ell(\alpha,\beta,n) \sim \frac{n\log\alpha}{\alpha}, \quad \beta_1 \sim 1 - \frac{\alpha}{n}.$$

If $\alpha, n, n/\alpha$ tend to infinity and $\alpha = \beta$,

$$\ell(\alpha,\alpha,n) \sim \frac{n\log\alpha}{4\alpha}, \quad \beta_1 \sim 1 - \frac{2\alpha}{n}.$$

The result also says that, starting from 1, convergence occurs abruptly (i.e., with cutoff) at $\ell(\alpha,\beta,n)$ as long as α tends to infinity.

Part (ii) indicates a completely different behavior starting from 1/2 (in the case $\alpha = \beta$). There is no cutoff and convergence occurs at the exponential rate given by β_2 ($\beta_2 \sim 1 - \frac{4\alpha}{n}$ if n/α tends to infinity).

Proof of Proposition 4.3(i) We have $\chi_1^2(\ell) = \sum_1^n \beta_i^{2\ell} p_i(1)^2 z_i$ and

$$\begin{aligned}\frac{\beta_{i+1}^{2\ell} p_{i+1}(1)^2 z_{i+1}}{\beta_i^{2\ell} p_i(1)^2 z_i} &= \left(\frac{n-i}{\alpha+\beta+n+i}\right)^{2\ell} \frac{2i+\alpha+\beta+1}{2i+\alpha+\beta-1}\frac{i+\alpha+\beta-1}{i+1}\frac{i+\alpha}{i+\beta} \\ &\leq \frac{5}{6}\frac{(\alpha+\beta)(\alpha+1)}{\beta+1}\left(1-\frac{\alpha+\beta+2}{\alpha+\beta+n+1}\right)^{2\ell} \qquad (4.4)\end{aligned}$$

The lead term in $\chi_1^2(\ell)$ is

$$\left(\frac{(\alpha+\beta+1)\alpha}{\beta}\right)\beta_1^{2\ell}.$$

From (4.4), we get that for any

$$\ell \geq \frac{1}{-2\log\beta_1}\log[(\alpha+\beta)(\alpha+1)/(\beta+1))]$$

we have

$$\frac{\beta_{i+1}^{2\ell}p_{i+1}(1)^2 z_{i+1}}{\beta_i^{2\ell}p_i(1)^2 z_i} \leq 5/6.$$

Hence, for such ℓ,

$$\begin{aligned}\chi_1^2(\ell) &\leq \left(\frac{(\alpha+\beta+1)\alpha}{\beta}\right)\beta_1^{2\ell}\left(\sum_0^\infty (5/6)^k\right)\\ &\leq 5\left(\frac{(\alpha+\beta+1)\alpha}{\beta}\right)\beta_1^{2\ell}.\end{aligned}$$

With $N = \log[(\alpha+\beta)(\alpha+1)/(\beta+1)]$ as in the proposition, we obtain

$$\begin{aligned}\chi_1^2(\ell) &\leq 7e^{-c} \text{ for } \ell \geq \frac{N+c}{-2\log\beta_1},\ c>0;\\ \chi_\ell^2(1) &\geq \frac{1}{6}e^c \text{ for } \ell \leq \frac{N-c}{-2\log\beta_1},\ c>0.\end{aligned}$$

□

Proof of Proposition 4.3(ii) When $a = b$, the classical Jacobi polynomial $P_k^{a,b}$ is given by

$$P_k^{a,a}(x) = \frac{(a+1)_k}{(2a+1)_k}C_k^{a+1/2}(x)$$

where the C_k^ν's are the ultraspherical polynomials. See [44, (4.5.1)]. Now, formula [44, (4.5.16)] gives $C_n^\nu(0) = 0$ if n is odd and

$$C_n^\nu(0) = \frac{(2\nu)_n}{2^n(n/2)!(\nu+1/2)_{n/2}}$$

if n is even. Going back to the shifted Jacobi's, this yields $p_{2k+1}(1/2) = 0$ and

$$\begin{aligned}p_{2k}(1/2) &= \frac{(\alpha)_{2k}}{(2\alpha-1)_{2k}}C_{2k}^{\alpha-1/2}(0)\\ &= \frac{(\alpha)_{2k}}{(2\alpha-1)_{2k}}\frac{(2\alpha-1)_{2k}}{2^{2k}k!(\alpha)_k} = \frac{(\alpha+k)_k}{2^{2k}k!}\end{aligned}$$

We want to estimate

$$\chi^2_{1/2}(\ell) = \sum_1^{\lfloor n/2 \rfloor} \beta_{2i}^{2\ell} p_{2i} (1/2)^2 z_{2i}$$

and thus we compute

$$\begin{aligned}
\frac{\beta_{2(i+1)}^{2\ell} p_{2(i+1)} (1/2)^2 z_{2(i+1)}}{\beta_{2i}^{2\ell} p_{2i} (1/2)^2 z_{2i}} &= \left(\frac{(n-2i)(n-2i-1)}{(2\alpha+n+2i)(2\alpha+n+2i+1)} \right)^{2\ell} \\
&\times \frac{4i+2\alpha+1}{4i+2\alpha-1} \frac{2i+2\alpha-1}{2i+2\alpha+1} \frac{2i(2i+1)(2\alpha+2i+1)(2\alpha+2i)}{(2i+\alpha)^2(2i+\alpha+1)^2} \left(\frac{(\alpha+2i)(\alpha+2i+1)}{4(\alpha+i)(i+1)} \right)^2 \\
&\leq \frac{9}{5} \beta_2^{2\ell} \qquad (4.5)
\end{aligned}$$

Hence

$$\chi^2_{1/2}(\ell) \leq 10 \beta_2^{2\ell} p_2 (1/2)^2 z_2 \text{ for } \ell \geq \frac{1}{-2 \log \beta_2}.$$

As

$$p_2(1/2) = \frac{\alpha+1}{4} \text{ and } z_2 = \frac{4(2\alpha+3)}{\alpha(\alpha+1)^2},$$

this gives $\chi^2_{1/2}(\ell) \geq \frac{1}{2} \beta_2^{2\ell}$ and, assuming $\ell \geq \frac{1}{-2 \log \beta_2}$, $\chi^2_{1/2}(\ell) \leq 13 \beta_2^{2\ell}$. □

4.2 Poisson/Gamma

4.2.1 The x-Chain for the Poisson/Gamma

Fix $\alpha,\ a > 0$. For $x, y \in \mathcal{X} = \{0, 1, 2, \ldots\} = \mathbb{N}$, let

$$\begin{aligned}
k(x,y) &= \int_0^\infty \frac{e^{-\lambda(\alpha+1)/\alpha} \lambda^{a+x-1}}{\Gamma(a+x)(\alpha/(\alpha+1))^{a+x}} \frac{e^{-\lambda} \lambda^y}{y!} d\lambda \\
&= \frac{\Gamma(a+x+y)(\frac{\alpha}{2\alpha+1})^{a+x+y}}{\Gamma(a+x)(\frac{\alpha}{\alpha+1})^{a+x} y!}. \qquad (4.6)
\end{aligned}$$

The stationary distribution is the negative binomial

$$m(x) = \frac{(a)_x}{x!} \left(\frac{1}{\alpha+1} \right)^x \left(\frac{\alpha}{\alpha+1} \right)^a, \quad x \in \mathbb{N}.$$

When $\alpha = a = 1$, the prior is a standard exponential, an example given in Section 2.1. Then,

$$k(x,y) = \left(\frac{1}{3} \right)^{x+y} \binom{x+y}{x} \Big/ \left(1/2 \right)^x, \qquad m(x) = 1/2^{x+1}.$$

The orthogonal polynomials for the negative binomial are Meixner polynomials [47, (1.9)]: $M_j(x) = {}_2F_1\Big(\begin{matrix} -j-x \\ a \end{matrix} \Big| -\alpha \Big)$. These satisfy [47, (1.92)]

$$\sum_{x=0}^{\infty} M_j(x) M_k(x) m(x) = \frac{(1+\alpha)^j j!}{(a)_j}\ \delta_{jk}.$$

Our work in previous sections, together with basic properties of Meixner polynomials gives the following propositions.

Proposition 4.4 *For $a, \alpha > 0$ the Poisson/Gamma x-chain* (4.6) *has:*

(a) Eigenvalues $\beta_j = (\alpha/(1+\alpha))^j,\ 0 \le j < \infty$.

(b) Eigenfunctions $M_j(x)$, the Meixner polynomials.

(c) For any $\ell \ge 0$ and any starting state x

$$\chi_x^2(\ell) = \sum_{y=0}^{\infty} \frac{(k^\ell(x,y) - m(y))^2}{m(y)} = \sum_{i=1}^{\infty} \beta_i^{2\ell} M_i^2(x) z_i \qquad z_i = \frac{(a)_i}{(1+\alpha)^i i!}.$$

Proposition 4.5 *For $\alpha = a = 1$, starting at n,*

$$\begin{aligned} \chi_n^2(\ell) &\le 2^{-c} \text{ for } \ell = \log_2(1+n) + c,\ c > 0; \\ \chi_n^2(\ell) &\ge 2^{c} \text{ for } \ell = \log_2(n-1) - c,\ c > 0. \end{aligned}$$

Proof From the definitions, for all j and positive integer x

$$|M_j(x)| = |\sum_{i=0}^{j\wedge x} (-1)^i \binom{j}{i} x(x-1)\dots(x-i+1)| \le \sum_{i=0}^{j} \binom{j}{i} x^i = (1+x)^j.$$

Thus, for $\ell \ge \log_2(1+n) + c$,

$$\begin{aligned} \chi_n^2(\ell) &= \sum_{j=1}^{\infty} M_j^2(n) 2^{-j(2\ell+1)} \le \sum_{j=1}^{\infty} (1+n)^{2j} 2^{-j(2\ell+1)} \\ &\le \frac{(1+n)^2 2^{-(2\ell+1)}}{1-(1+n)^2 2^{-(2\ell+1)}} \le \frac{2^{-c-1}}{1-2^{-c-1}} \le 2^{-c}. \end{aligned}$$

The lower bound follows from using only the lead term. Namely

$$\chi_n^2(\ell) \ge (1-n)^2 2^{-2\ell} \ge 2^c \quad \text{for} \quad \ell = \log_2(n-1) - c.$$

□

Remark Note the contrast with the Beta/Binomial example above. There, order n steps are necessary and sufficient starting from n and there is no cutoff. Here, $\log_2 n$ steps are necessary and sufficient and there is a cutoff. See [21] for further discussion of cutoffs.

4.2.2 The θ-chain for the Poisson/Gamma

Fix $\alpha, a > 0$. For $\theta, \theta' \in \Theta = (0, \infty)$, let $\eta = (\alpha+1)\theta'/\alpha$ and write

$$\begin{aligned}
k(\theta, \theta') &= \sum_{j=0}^{\infty} \frac{e^{-\theta}\theta^j}{j!} \, \frac{e^{-\theta'(\alpha+1)/\alpha}(\theta')^{a+j-1}}{\Gamma(a+j)(\alpha/(\alpha+1))^{a+j}} \\
&= \frac{e^{-\theta-\eta}\ \eta^{a-1}}{\alpha/(\alpha+1)} \sum_{j=0}^{\infty} \frac{(\theta\eta)^j}{j!\Gamma(a+j)} \\
&= \frac{e^{-\theta-\eta}}{\alpha/(\alpha+1))} \Big(\frac{\eta}{\theta}\Big)^{\frac{a-1}{2}} \sum_{j=0}^{\infty} \frac{(\sqrt{\theta\eta})^{2j+a-1}}{j!\Gamma(a+j)} \\
&= \frac{e^{-\theta-\eta}}{\alpha/(1+\alpha)} \Big(\frac{\eta}{\theta}\Big)^{\frac{a-1}{2}} I_{a-1}(2\sqrt{\theta\eta}) \\
&= \frac{e^{-\theta-(\alpha+1)\theta'/\alpha}}{\alpha/(1+\alpha)} \Big(\frac{(\alpha+1)\theta'}{\alpha\theta}\Big)^{\frac{a-1}{2}} I_{a-1}(2\sqrt{(\alpha+1)\theta\theta'/\alpha}). \qquad (4.7)
\end{aligned}$$

Here I_{a-1} is the modified Bessell function. For fixed θ, $k(\theta, \theta')$ integrates to one as discussed in [35, pg. 58-59]. The stationary distribution of this Markov chain is the Gamma:

$$\pi(d\theta) = \frac{e^{-\theta/\alpha}\theta^{a-1}}{\Gamma(a)\alpha^a} d\theta.$$

To simplify notation, we take $\alpha = 1$ for the rest of this section. The relevant polynomials are the Laguerre polynomials [47, Sec. 1.11]

$$L_i(\theta) = \frac{(a)_i}{i!} \, {}_1F_1\Big(\begin{matrix} -i \\ a \end{matrix} \Big| \, \theta\Big) = \frac{1}{i!} \, \sum_{j=0}^{i} \frac{(-i)_j}{j!} \, (a+j)_{i-j}\theta^j.$$

Note that classical notation has the parameter a shifted by 1 whereas we have labelled things to mesh with standard statistical notation. The orthogonality relation is

$$\int_0^{\infty} L_i(\theta)L_j(\theta)\pi(\theta)d\theta = \frac{\Gamma(a+j)}{j!\Gamma(a)} \, \delta_{ij} = z_j^{-1}\delta_{ij}.$$

The multilinear generating function formula [44, Theorem 4.7.5] gives

$$\sum_{i=0}^{\infty} L_i(\theta)^2 z_i t^i = \frac{e^{-2t\theta/(1-t)}}{(1-t)^a} \sum_{0}^{\infty} \frac{1}{j!(a)_j} \left(\frac{\theta^2 t}{1-t^2}\right)^j.$$

Combining results, we obtain the following statements.

Proposition 4.6 *For $\alpha = 1$ and $a > 0$, the Markov chain with kernel (4.7) has:*

(a) *Eigenvalues* $\beta_j = \frac{1}{2^j} \quad 0 \le j < \infty$.

(b) *Eigenvectors* L_j *the Laguerre polynomials.*

(c) *For any* $\ell \ge 1$ *and any starting state* θ,

$$\chi^2_\theta(\ell) = \sum_{j=1}^{\infty} \beta_j^{2\ell} L_j^2(\theta) \frac{j!\Gamma(a)}{\Gamma(a+j)} = \frac{e^{-\frac{2^{-2\ell+1}\theta}{1-2^{-2\ell}}}}{(1-2^{-2\ell})^a} \sum_0^{\infty} \frac{1}{j!(a)_j} \left(\frac{\theta^2 2^{-2\ell}}{1-2^{-4\ell}} \right)^j - 1$$

Proposition 4.7 *For* $\alpha = 1$ *and* $a > 0$, *the Markov chain with kernel* (4.7) *satisfies*

- *For* $\theta > 0$, $\chi^2_\theta(\ell) \le e^2 2^{-c}$ *if* $\ell \ge \frac{1}{2}\log_2[2(1+a+\theta^2/a)] \ + c$, $c > 0$.
- *For* $\theta \in (0, a/2) \cup (2a, \infty)$, $\chi^2_\theta(\ell) \ge 2^c$ *if* $\ell \le \frac{1}{2}\log_2[\frac{1}{2}(\theta^2/a+a)] \ - c$, $c > 0$.

Proof For the upper bound, assuming $\ell \ge 1$, we write

$$\begin{aligned}
\chi^2_\theta(\ell) &= (1-4^{-\ell})^{-a} e^{-\frac{2\theta 4^{-\ell}}{1-4^{-\ell}}} \sum_0^{\infty} \frac{1}{j!(a)_j} \left(\frac{\theta^2 4^{-\ell}}{1-4^{-\ell}} \right)^j - 1 \\
&\le \frac{\exp\left((2\theta^2/a)4^{-\ell}\right)}{(1-4^{-\ell})^a} - 1 \\
&\le 2(\theta^2/a+a)4^{-\ell} \left(\frac{\exp\left(2(\theta^2/a)4^{-\ell}\right)}{(1-4^{-\ell})^{a+1}} \right)
\end{aligned}$$

For $\ell \ge \frac{1}{2}\left(\log_2[2(1+\theta^2/a+a)] + c\right)$, $c > 0$, we obtain $\chi^2_\theta(\ell) \le e^2 2^{-c}$.

The stated lower bound does not easily follows from the formula we just used for the upper bound. Instead, we simply use the first term in $\chi^2_\theta(\ell) = \sum_{j\ge 1} \beta_j^{2\ell} L_j^2(\theta) \frac{j!\Gamma(a)}{\Gamma(a+j)}$, that is, $a^{-1}(\theta-a)^2 4^{-\ell}$. This easily gives the desired result. □

Remark: It is not easy to obtain sharp formula starting from θ near a. For instance, starting at $\theta = a$, one gets a lower bounds by using the second term $\chi^2_\theta(\ell) = \sum_{j\ge 1} \beta_j^{2\ell} L_j^2(\theta) \frac{j!\Gamma(a)}{\Gamma(a+j)}$ (the first term vanishes at $\theta = a$). This gives $\chi^2_a(\ell) \ge [2a/(a+1)]4^{-2\ell}$. When a is large, this is significantly smaller than the upper bound proved above.

4.3 The Gaussian case

Here, the x-chain and the θ-chain are essentially the same and indeed the same as the chain for the additive models so we just treat the x-chain. Let $\mathcal{X} = \mathbb{R}$, $f_\theta(x) = e^{-\frac{1}{2}(x-\theta)^2/\sigma^2}/\sqrt{2\pi\sigma^2}$ and $\pi(d\theta) = \frac{e^{-\frac{1}{2}(\theta-\nu)^2/\tau^2}}{\sqrt{2\pi\tau^2}} d\theta$. The marginal density is Normal$(v, \sigma^2+\tau^2)$.

A stochastic description of the chain is

$$X_{n+1} = aX_n + \epsilon_{n+1} \qquad \text{with} \quad a = \frac{\tau^2}{\sigma^2+\tau^2}, \quad \epsilon \sim \text{Normal}\left(\frac{\sigma^2\nu}{\sigma^2+\tau^2}, \sigma^2\right). \tag{4.8}$$

This is the basic autoregressive (AR1) process. Feller [35, pg. 97-99] describes it as the discrete time Ornstein-Uhlenbeck process. The diagonalization of this Gaussian Markov chain has been derived by other authors in various contexts. Goodman and Sokal [40] give an explicit diagonalization of vector valued Gaussian autoregressive processes which specialize to (a), (b), (c) above. Donoho and Johnstone [26, Lemma 2.1] also specializes to (a), (b), (c) above. Both sets of authors give further references. Since it is so well studied, we will be brief and treat the special case with $\nu = 0, \sigma^2+\tau^2 = 1/2$. Thus the stationary distribution is Normal$(0, 1/2)$. The orthogonal polynomials are now Hermite polynomials [47, 1.13]. These are given by

$$H_n(y) = (2y)^n {}_2F_0\left(\begin{matrix} -n/2, -(n-1)/2 \\ -- \end{matrix} \,\middle|\, -\frac{1}{y^2}\right) = n! \sum_{k=0}^{[n/2]} \frac{(-1)^k (2y)^{n-2k}}{k!(n-2k)!}$$

They satisfy

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-y^2} H_m(y) H_n(y) dy = 2^n n! \delta_{mn}.$$

There is also a multilinear generating function formula which gives ([44, Example 4.7.3])

$$\sum_0^\infty \frac{H_n(x)^2}{2^n n!} t^n = \frac{1}{\sqrt{1-t^2}} \exp\left(\frac{2x^2 t}{1+t}\right).$$

Proposition 4.8 *For $\nu = 0$, $\sigma^2+\tau^2 = 1/2$, the Markov chain (4.8) has:*

(a) Eigenvalues $\beta_j = (2\tau^2)^j$ (as $\sigma^2+\tau^2 = 1/2$, we have $2\tau^2 < 1$).

(b) Eigenvectors the Hermite polynomials H_j.

(c) For any starting state x and all $\ell \geq 1$

$$\chi_x^2(\ell) = \sum_{k=1}^{\infty} (2\tau^2)^{2k\ell} H_k^2(x) \frac{1}{2^k k!} = \frac{\exp\left(\frac{2x^2(2\tau^2)^{2\ell}}{1+(2\tau^2)^{2\ell}}\right)}{\sqrt{1-(2\tau^2)^{4\ell}}} - 1.$$

The next proposition turns the available chi-square formula into sharp estimates when x is away from 0. Starting from 0, the formula gives $\chi_0^2(\ell) = (1-(2\tau^2)^{4\ell})^{-1/2} - 1$. This shows convergence at the faster exponential rate of $\beta_2 = (2\tau^2)^2$ instead of $\beta_1 = 2\tau^2$.

Proposition 4.9 *For $\nu = 0$, $\sigma^2 + \tau^2 = 1/2$, $x \in \mathbb{R}$, the Markov chain* (4.8) *satisfies:*

$$\begin{aligned}
\chi_x^2(\ell) &\leq 8e^{-c} \quad \textit{for } \ell \geq \frac{\log(2(1+x^2)) + c}{-2\log(2\tau^2)}, \ \ c > 0. \\
\chi_x^2(\ell) &\geq \frac{x^2 e^c}{2(1+x^2)} \quad \textit{for } \ell \leq \frac{\log(2(1+x^2)) - c}{-2\log(2\tau^2)}, \ \ c > 0. \\
\chi_0^2(\ell) &= (1-(2\tau^2)^{4\ell})^{-1} - 1 \geq (2\tau^2)^{4\ell}.
\end{aligned}$$

Proof For the upper bound, assuming

$$\ell \geq \frac{1}{-2\log(2\tau^2)} \left(\log(2(1+x^2)) + c\right), c > 0,$$

we have

$$(2\tau^2)^{2\ell} < 1/2, \ \ 2x^2(2\tau)^{2\ell} < 1$$

and it follows that

$$\begin{aligned}
\chi_x^2(\ell) &= \frac{\exp\left(\frac{2x^2(2\tau^2)^{2\ell}}{1+(2\tau^2)^{2\ell}}\right)}{\sqrt{1-(2\tau^2)^{4\ell}}} - 1 \leq \left(1 + 2(2\tau^2)^{4\ell}\right)\left(1 + 6x^2(2\tau^2)^{2\ell}\right) - 1 \\
&\leq 8(1+x^2)(2\tau^2)^{2\ell}.
\end{aligned}$$

For the lower bound, write

$$\chi_x^2(\ell) = \frac{\exp\left(\frac{2x^2(2\tau^2)^{2\ell}}{1+(2\tau^2)^{2\ell}}\right)}{\sqrt{1-(2\tau^2)^{4\ell}}} - 1 \geq \exp\left(x^2(2\tau^2)^{2\ell}\right) - 1 \geq x^2(2\tau^2)^{2\ell}.$$

□

5 Location families examples

In this section $f_\theta(x) = g(x-\theta)$ with g and π members of one of the six families of Section 2.4. To picture the associated Markov chains it is helpful to begin with the representation $x = \theta + \epsilon$. Here θ is distributed as π and ϵ is distributed as g. The x-chain goes as follows: from x, draw θ' from $\pi(\cdot|x)$ and then go to $x' = \theta' + \epsilon'$ with ϵ' independently drawn from g. It has stationary distribution $m(x)dx$, the convolution of π and g. For the θ-chain, starting at θ, set $x' = \theta + \epsilon$ and draw θ' from $\pi(\cdot|x')$. It has stationary distribution π.

$$E_\theta(X^k) = E_\theta((\theta+\epsilon)^k) = \sum_{j=0}^{k} \binom{k}{j} \theta^j E(\epsilon^{k-j}).$$

Thus (H.2) of Section 3 is satisfied with $\eta_k = 1$. To check the conjugate condition we may use results of [60, Sect. 4]. In present notation, Morris shows that if p_k is the monic orthogonal polynomial of degree k for the distribution π and p'_k the monic orthogonal polynomial of degree k for the distribution m, then

$$E_x(p_k(\theta)) = \left(\frac{n_1}{n_1+n_2}\right)^k b_k p'_k(x).$$

Here π is taken as the sum of n_1 copies and ϵ the sum of n_2 copies of one of the six families and

$$b_k = \prod_{i=0}^{k-1} \frac{1+\frac{ic}{n_1}}{1+\frac{ic}{n_1+n_2}}$$

where c is the coefficient for var $= a + b\mu + c\mu^2$ for the family. Comparing lead terms gives (H.3) with an explicit value of μ_k. In the present set-up, $\mu_k = \beta_k$ the k-th eigenvalue.

We now make specific choices for each of the six cases.

5.1 Binomial

For fixed $p, 0 < p < 1$ let $\pi = \text{Bin}(n_1, p), g = \text{Bin}(n_2, p)$. Then $m = \text{Bin}(n_1 + n_2, p)$ and

$$\pi(\theta|x) = \frac{\binom{n_1}{\theta}\binom{n_2}{x-\theta}}{\binom{n_1+n_2}{x}}$$

is hypergeometric. The θ-chain progresses as a population process on $0 \le \theta \le n_1$: from θ, there are ϵ new births and the resulting population of size $x = \theta + \epsilon$ is thinned down by random sampling. The x-chain has an autoregressive cast: From x, the process is decreased and then increased as

$$X_{n+1} = S_{X_n} + \epsilon_{n+1} \tag{5.1}$$

with S_x a hypergeometric with parameter n_1, n_2, X_n and ϵ_{n+1} drawn from $\text{Bin}(n_2, p)$.

For the binomial, the parameter c is $c = -1$ and the eigenvalues of the x-chain are

$$\beta_k = \frac{n_1(n_1-1)\cdots(n_1-k+1)}{(n_1+n_2)(n_1+n_2-1)\cdots(n_1+n_2-k+1)}, \quad 0 \le k \le n_1+n_2.$$

Note that $\beta_k = 0$ for $k \ge n_1 + 1$. The orthogonal polynomials are Krawtchouck polynomials ([47, 1.10], [44, pg 100])

$$k_j(x) = {}_2F_1\left(\begin{matrix}-j-x \\ -n\end{matrix}\,\middle|\,\frac{1}{p}\right)$$

which satisfy

$$\sum_{x=0}^{n}\binom{n}{x}p^x(1-p)^{n-x}k_j(x)k_\ell(x) = \binom{n}{j}^{-1}\left(\frac{1-p}{p}\right)^j\delta_{j\ell}.$$

Proposition 5.1 *Consider the chain* (5.1) *on* $\{0, \ldots, n_1 + n_2\}$ *with* $0 < p < 1$, *starting at* $x = 0$. *Set* $N = n_1 + n_2$, $q = p/(1-p)$. *Then we have*

$$e^{-c} \le \chi_0^2(\ell) \le e^{-c} e^{e^{-c}}$$

whenever

$$\ell = \frac{\log(qN) \ + c}{-2\log(1 - n_2/N)}, \quad c \in (-\infty, \infty).$$

Note two cases of interest: (i) For $p = 1/2$, the proposition shows that $\frac{\log(2N)}{-2\log(1-n_2/N)}$ steps are necessary and sufficient. There is a chi-square cutoff when N tends to infinity. (ii) For $p = 1/N$, there is no cutoff.

Proof We have $k_j^2(0) = 1$ for all j and the chi-square distance becomes

$$\chi_0^2(\ell) = \sum_{j=1}^{n_1} \beta_j^{2\ell} \binom{N}{j} q^j$$

with $N = n_1 + n_2$, $q = p/(1-p)$. For $j \le n_1$, the eigenvalues satisfy

$$\beta_j = \prod_{i=0}^{j-1} \left(1 - \frac{n_2}{N-i}\right) \le \left(1 - \frac{n_2}{N}\right)^j = \beta_1^j.$$

Hence, we obtain

$$\chi_0^2(\ell) \le \sum_{j=1}^{N} \binom{N}{j} (q\beta_1)^j = \left(1 + q\beta_1^{2\ell}\right)^N - 1 \le qN\beta_1^{2\ell} \left(1 + q\beta_1^{2\ell}\right)^{N-1}.$$

This gives the desired result since we also have $\chi_0^2(\ell) \ge Nq\beta_1^{2\ell}$. □

5.2 Poisson

Fix positive reals μ, n_1, n_2. Let $\pi = \text{Poisson}(\mu n_1)$, $g = \text{Poisson}(\mu n_2)$. Then

$$m = \text{Poisson}(\mu(n_1 + n_2)) \text{ and } \pi(\theta|x) = \text{Bin}\Big(x, \frac{n_1}{n_1 + n_2}\Big).$$

The x-chain is related to the $M/M/\infty$ queue and the θ-chain is related to Bayesian missing data examples in Section 2.3.3. Here, the parameter $c = 0$ so that

$$\beta_k = \left(\frac{n_1}{n_1 + n_2}\right)^k \qquad 0 \le k < \infty.$$

The orthogonal polynomials are Charlier polynomials ([47, 1.12], [44, pg. 177]):

$$C_j(x) = {}_2F_0\left(\begin{matrix} -j, -x \\ -- \end{matrix} \,\middle|\, -\frac{1}{\mu}\right), \quad \sum \frac{e^{-\mu}\mu^x}{x!} C_j(x) C_k(x) = j!\mu^{-j}\delta_{jk}.$$

We carry out a probabilistic analysis of this problem in [20].

5.3 Negative binomial

Fix p with $0 < p < 1$ and positive real n_1, n_2. Let $\pi = \text{NB}(n_1, p)$, $g = \text{NB}(n_2, p)$. Then $m = \text{NB}(n_1 + n_2, p)$ and

$$\pi(\theta|x) = \binom{x}{\theta} \frac{\Gamma(n_1+n_2)\Gamma(\theta+n_1)\Gamma(x-\theta-n_2)}{\Gamma(x+n_1+n_2)\Gamma(n_1)\Gamma(n_2)}, \qquad 0 \le \theta \le x$$

which is a negative hypergeometric. A simple example has $n_1 = n_2 = 1$ (geometric distribution) so $\pi(\theta|x) = 1/(1+x)$. The x-chain becomes: From x, choose θ uniformly in $0 \le \theta \le x$ and let $X' = \theta + \epsilon$ with ϵ geometric. The parameter $c = 1$ so that

$$\beta_0 = 1, \quad \beta_k = \frac{n_1(n_1+1)\cdots(n_1+k-1)}{(n_1+n_2)(n_1+n_2+1)\cdots(n_1+n_2+k-1)}, \qquad 1 \le k < \infty.$$

The orthogonal polynomials are Meixner polynomials discussed in Section 4.2 above.

5.4 Normal

Fix reals μ and $n_1, n_2, v > 0$. Let $\pi = \text{Normal}(n_1\mu, n_1 v)$, $g = \text{Normal}(n_2\mu, n_2 v)$. Then $m = \text{Normal}((n_1+n_2)\mu, (n_1+n_2)v)$ and $\pi(\theta|x) = \text{Normal}\Big(\frac{n_1}{n_1+n_2}x, \frac{n_1 n_2}{n_1+n_2}V\Big)$. Here $c = 0$ and

$$\beta_k = \left(\frac{n_1}{n_1+n_2}\right)^k \quad 0 \le k < \infty.$$

The orthogonal polynomials are Hermite, discussed in Section 4.3 above. Both the x and θ-chains are classical autoregressive processes as described in Section 4.3.

5.5 Gamma

Fix positive real n_1, n_2, α. Let $\pi = \text{Gamma}(n_1, \alpha), g = \text{Gamma}(n_2, \alpha)$. Then

$$m = \text{Gamma}(n_1+n_2, \alpha), \;\; \pi(\theta|x) = x \cdot \text{Beta}(n_1, n_2).$$

A simple case to picture is $\alpha = n_1 = n_2 = 1$. Then, the x-chain may be described as follows: From x, choose θ uniformly in $(0, x)$ and set $X' = \theta + \epsilon$ with ϵ standard exponential. This is simply a continuous version of the examples of Section 5.3. The parameter $c = 1$ and so

$$\beta_0 = 1, \beta_k = \frac{n_1(n_1+1)\cdots(n_1+k-1)}{(n_1+n_2)(n_1+n_2+1)\cdots(n_1+n_2+k-1)}, \quad 0 < k < \infty.$$

The orthogonal polynomials are Laguerre polynomials, discussed in Section 4.2 above.

5.6 Hyperbolic

The density of the sixth family is given in Section 2.3 in terms of parameters $r > 0$ and $|\theta| < \pi/2$. It has mean $\mu = r\tan(\theta)$ and variance $\mu^2/r + r$. See [59, Sect. 5]) or [28] for numerous facts and references. Fix real μ and positive n_1, n_2. Let the density π be hyperbolic with mean $n_1\mu$ and $r_1 = n_1(1+\mu^2)$. Let the density g be hyperbolic with mean $n_2\mu$ and $r_2 = n_2(1+\mu^2)$. Then m is hyperbolic with mean $(n_1+n_2)\mu$ and $r = (n_1+n_2)(1+\mu^2)$. The conditional density $\pi(\theta|x)$ is 'unnamed and apparently has not been studied' ([60, pg. 581]).

For this family, the parameter $c = 1$ and thus

$$\beta_0 = 1, \;\; \beta_k = \frac{n_1(n_1+1)\cdots(n_1+k-1)}{(n_1+n_2)\cdots(n_1+n_2+k-1)}.$$

The orthogonal polynomials are Meixner-Pollaczek polynomials ([68, pg. 395], [47, 1.7], [44, pg. 171]). These are given in the form

$$P_n^\lambda(x,\varphi) = \frac{(2\lambda)_n}{n!}{}_2F_1\left(\begin{matrix}-n, \lambda+ix \\ 2\lambda\end{matrix}\;\middle|\; 1-e^{-2i\varphi}\right)e^{in\varphi} \tag{5.2}$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{(2\varphi-\pi)x}|\Gamma(\lambda+ix)|^2 P_m^\lambda P_n^\lambda dx = \frac{\Gamma(n+2\lambda)}{n!(2\sin\varphi)^{2\lambda}}\;\delta_{mn}$$

Here $-\infty < x < \infty$, $\lambda > 0, 0 < \varphi < \pi$. The change of variables $y = \frac{rx}{2}$, $\varphi = \frac{\pi}{2} + \tan^{-1}(\theta)$ $\lambda = r/2$ transforms the density $e^{(2\varphi-\pi)x}|\Gamma(\lambda+ix)|^2$ to a constant multiple of the density $f_\theta(x)$ of Section 2.4.

We carry out one simple calculation. Let π, g have the density of $\frac{2}{\pi}\log|C|$, with C standard Cauchy. Thus

$$\pi(dx) = g(x)dx = \frac{1}{2\cosh(\pi x/2)}dx. \tag{5.3}$$

The marginal density is the density of $\frac{2}{\pi}\log|C_1C_2|$, that is,

$$m(x) = \frac{x}{2\sinh(\pi x/2)}.$$

Proposition 5.2 *For the additive walk based on* (5.3)*:*

(a) The eigenvalues are $\beta_k = \frac{1}{k+1}$, $0 \le k < \infty$.

(b) The eigenfunctions are the Pollaczek polynomials (5.2) *with* $\varphi = \pi/2$, $\lambda = 1$.

(c) $\chi_x^2(\ell) = 2\sum_{k=1}^{\infty}(k+1)^{-2\ell-1}\Big(P_k^1\Big(\frac{x}{2},\frac{\pi}{2}\Big)\Big)^2$.

Proof Using $\Gamma(z+1) = z\Gamma(z)$, $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$ we check that

$$|\Gamma(1+ix)|^2 = \Gamma(1+ix)\Gamma(1-ix) = (ix)\Gamma(ix)\Gamma(1-ix) = \frac{\pi(ix)}{\sin\pi(ix)} = \frac{\pi x}{\sinh(\pi x)}.$$

The result now follows from routine simplification. □

6 A Little Operator Theory

Most of the kernels studied in previous sections give compact operators on the associated L^2 spaces. In this section we give tools for explaining this and for proving compactness in less standard problems.

Throughout, with notation as in Section 2, let $L^2(m)$ and $L^2(\pi)$ be the usual L^2 spaces. For most of this section $\mathcal{X}$ and Θ can be general measurable spaces. Define

$$\begin{aligned} &T : L^2(m) \to L^2(\pi) && T^* : L^2(\pi) \to L^2(m) \\ &g(x) \longmapsto \int_{\mathcal{X}} f_\theta(x) g(x) \mu(dx) && h(\theta) \longmapsto \int_\Theta \pi(\theta|x) h(\theta) \pi(d\theta) \end{aligned}$$

It is straightforward to verify that T and T^* are bounded operators with norm 1. Further, T and T^* are adjoints (hence the notation):

$$\langle Tg, h \rangle_\pi = \langle g, T^* h \rangle_m = \int_{\mathcal{X} \times \Theta} g(x) h(\theta) f_\theta(x) \mu(dx) \pi(d\theta).$$

It follows that the eigenvalues of TT^* and T^*T are non-negative.

The mapping T corresponds to "choose x given θ from $f_\theta(x)$" while the mapping T^* corresponds to "choose θ given x from $\pi(\theta|x)$. Finally, T^*T has transition kernel $k(x, x')\mu(dx')$ from $L^2(m)$ to $L^2(m)$ (the x-chain) and TT^* has kernel $k(\theta, \theta')\pi(d\theta')$ from $L^2(\pi)$ to $L^2(\pi)$.

6.1 Compactness

This topic is treated in any graduate text on functional analysis. We have found the short treatment in [1, pg. 56-61]) particularly clear and useful. They call compact operators 'completely continuous'. The more comprehensive treatment of Ringrose [63] is also recommended.

Recall that if L, L' are Hilbert spaces, an operator $A : L \to L'$ is *compact* if for any $g_i \in L, \|g_i\| \leq 1$, there is a subsequence g_{i_j} such that Ag_{i_j} converges to a limit in L'. If $A^* : L' \to L$ is the adjoint of A it is known that

- A is compact iff A^*A, equivalently AA^*, equivalently A^* is compact.
- A^*A is compact iff there is an orthonormal basis g_i for L and real numbers $\beta_i \searrow 0$ so that $A^*Ag_i = \beta_i g_i$.
- In this case, set $h_k = Ag_i$. Then $\{g_i : \beta_i > 0\}$ form an orthonormal basis for $(\ker A)^\perp$, $\{h_i : \beta_i > 0\}$ form an orthogonal basis for the range of A with $\langle h_i, h_i \rangle = \beta_i^2$. We say $\{g_i, h_i\}$ give a *singular value decomposition* of A.

Example Consider the Poisson/Exponential example of Section 2.1. The measure $m(j) = 1/2^{j+1}$ is determined by its moments and $T^*T : L^2(m) \to L^2(m)$ has the Meixner polynomials as an orthonormal basis of $L^2(m)$ with $T^*TM_i = \beta_i M_i$, $\beta_i = 1/2^i$, $0 \leq i < \infty$. So T, T^*, TT^* as

well as T^*T are compact. All of the examples of Section 4 and 5 similarly arise from compact operators.

Example [1, pg. 60] shows that an infinite tri-diagonal matrix yields a compact operator if and only if the elements on, above, and below the diagonal tend to zero. Consider a birth and death chain with transition matrix K on $\{0, 1, 2, \ldots\}$ and stationary probability M. We can never have K compact. We may have $I - K$ compact (iff $K(i, i) \to 1$). Then, since $I - K$ has eigenvalues tending to zero, the operator K does not have a spectral gap. In Silver [67] the birth and death chain with $K(0, 0) = K(0, 1) = \frac{1}{2},\ K(i, i+1) = \frac{1}{3}\ K(i, i-1) = 2/3,\ 1 \leq i < \infty$ is studied. Aside from $\beta_0 = 1$, this chain has only continuous spectrum. The point of those examples is that compact operators do not occur easily when the state space is infinite.

The following proposition gives a simple sufficient condition for compactness in our setup.

Proposition 6.1 *Each of the following conditions implies that the operators T, T^*, T^*T, TT^* are compact.*

(a) $\sup_x \int_\Theta \pi^2(\theta|x)\pi(d\theta) < \infty$

(b) $\sup_\theta \int_\chi \frac{f_\theta^2(x)}{m(x)}\mu(dx) < \infty.$

(c) $\int \frac{f_\theta^2(x)}{m(x)}\pi(d\theta)\mu(dx) < \infty.$

Proof Condition (a) implies that T is a bounded operator from $L^2(m)$ to $L^\infty(\pi)$. By duality, T^* must be bounded from $L^1(\pi)$ to $L^2(m)$ and thus TT^* is bounded from $L^1(\pi)$ to $L^\infty(\pi)$. As TT^* has kernel $k(\theta, \theta')$ with respect to $\pi(d\theta')$, this implies

$$\sup_{\theta,\theta'}\{k(\theta, \theta')\} < \infty. \tag{6.1}$$

Recall that an operator with kernel k w.r.t. a measure π is trace class if $\int k(\theta, \theta)\pi(d\theta) < \infty$. Being trace class is a standard sufficient condition for compactness (the eigenvalues form a summable series). As π is a probability measure, (6.1) implies that TT^* is trace class hence compact. Exchanging the roles of x and θ, the same argument proves that T^*T is compact (in fact, trace class) starting from condition (b). Note that (a) is the same as

$$\sup_x\{\bar{k}(x, x)\} = \sup_x\{k(x, x)/m(x)\} < \infty$$

whereas (b) is the same as

$$\sup_\theta\{k(\theta, \theta)\} < \infty.$$

Condition (c) is weaker than (a) or (b) since

$$\int_{\mathcal{X}\times\Theta} \frac{f_\theta^2(x)}{m(x)}\pi(d\theta)\mu(dx) = \int_{\mathcal{X}} \frac{k(x,x)}{m(x)} m(dx) = \int_\Theta k(\theta,\theta)\pi(d\theta).$$

It exactly says that TT^* and T^*T are trace class. □

Example In the Poisson/Exponential example of Section 2.1, we have $\pi = e^{-\theta}d\theta$, $f_\theta(x) = \frac{e^{-\theta}\theta^x}{x!}$, $m(x) = 1/2^{x+1}$, $\pi(\theta|x) = f_\theta(x)/m(x)$ and

$$\frac{k(x,x)}{m(x)} = \int_\Theta \pi^2(\theta|x)\pi(d\theta) = \frac{2^{2(x+1)}}{(x!)^2} \int e^{-3\theta}\theta^{2x}d\theta = \frac{2^{2(x+1)}}{3^{2x+1}}\binom{2x}{x} \sim \frac{4(4/3)^{2x}}{3\sqrt{\pi x}}.$$

using $\binom{2x}{x} \sim 2^{2x}/\sqrt{\pi x}$. Condition (a) is not satisfied but condition (c) is since

$$\sum k(x,x) = \sum \frac{2^{2(x+1)}}{3^{2x+1}}\binom{2x}{x}2^{-x-1} < \infty.$$

Example Poisson/non-exponential. Let $f_\theta(x) = \frac{e^{-\theta}\theta^x}{x!}$, $\pi(d\theta) = ce^{-|\theta|^2}d\theta$ on $(0,\infty)$ with $c = 2/(\sqrt{\pi})$ (normalizing constant). In this case we have, for x large enough,

$$\begin{aligned} m(x) &= \int_\Theta f_\theta(x)\pi(d\theta) = \int_0^\infty \frac{e^{-\theta}}{x!}\theta^x e^{-\theta^2}d\theta \geq \frac{1}{e^{1/4}x!}\int_0^\infty e^{-(\theta+1/2)^2}\theta^x d\theta \\ &\geq \frac{1}{e^{1/4}2^{x+2}x!}\int_0^\infty e^{-u}u^{(x-1)/2}du = \frac{\Gamma((x+1)/2)}{e^{1/4}2^{x+2}x!} \end{aligned}$$

and

$$\begin{aligned} k(x,x) &= \frac{1}{(x!)^2 m(x)}\int_0^\infty e^{-2\theta}\theta^{2x}e^{-\theta^2}d\theta \leq \frac{1}{(x!)^2 m(x)}\int_0^\infty \theta^{2x}e^{-\theta^2}d\theta \\ &\leq \frac{1}{2(x!)^2 m(x)}\int_0^\infty \theta^{x-1/2}e^{-\theta}d\theta = \frac{\Gamma(x+1/2)}{2(x!)^2 m(x)} \\ &\leq \frac{e^{1/4}2^{x+1}\Gamma(x+1/2)}{x!\Gamma((x+1)/2)}. \end{aligned}$$

This shows that $k(x,x)$ is summable and thus T, T^*, TT^* and T^*T are compact.

We close this section by observing that, in contrast with the results obtained for the systematic scan Gibbs sampler, the operator $\bar{K}$ corresponding to the random scan Gibbs sampler is never compact when the state space is infinite. This easily follows from Theorem 3.1(c) which asserts that $1/2$ is an accumulation point for the eigenvalues of $\bar{K}$.

7 Other models, other methods

Even in the limited context of Markov chains with polynomial eigenfunctions, there are examples not treated here and further techniques for proving convergence. The present section gives brief pointers to these results.

7.1 Univariate Examples

Fix positive integers n, θ, N with $n, \theta \leq N$. Define

$$f_\theta(x) = \frac{\binom{\theta}{x}\binom{N-\theta}{n-x}}{\binom{N}{n}}, \qquad (n+\theta-N)_+ \leq x \leq \min(\theta, n).$$

This is the classical model for sampling without replacement from a population of size N containing θ-'reds' and $N-\theta$ 'blacks'. A sample of size n is chosen without replacement and x is the number of reds in the sample. A Bayesian treatment puts a prior $\pi(\theta)$ on θ. One standard choice is

$$\pi(\theta) = \frac{\binom{R}{\theta}\binom{M-R}{N-\theta}}{\binom{M}{N}}, \qquad (N+R-M)_+ \leq \theta \leq \min(R, N).$$

One may compute that the posterior $\pi(\theta|x)$ is again hypergeometric. In [23], it is shown that the x-chain and the θ-chain have polynomial eigenfunctions with simple eigenvalues. By passing to various limits, these authors show that this example includes various location models treated above (binomial; Poisson and normal). Further, the natural q-analog involving subspaces of a vector space gives some q-deformations of present results.

Markov chains with polynomial eigenfunctions have been extensively studied in the mathematical genetics literature. This work, which perhaps begins with [33], was unified in [13]. See [29] for a textbook treatment. Models of Fisher-Wright, Moran, Kimura, Karlin and McGregor are included. While many models are either absorbing, non-reversible, or have intractable stationary distributions, there are also tractable new models to be found. See the Stanford thesis work of Hua Zhou.

Further examples can be found in [12, Sec. 7-12]. In particular, one finds there a characterization of circulency symmetric bivariate measures where the Gibbs sampler has polynomial eigenfunctions. Many of these can be analysed by the methods of the present paper. Conversely, our examples give new and different examples for understanding the alternating conditional expectations that are the central focus of [12].

A rather different class of examples can be created using autoregressive processes. For definiteness, work on the real line $\mathcal{R}$. Consider processes of form, $X_0 = 0$, and for $1 \leq n < \infty$,

$$X_{n+1} = a_{n+1}X_n + \epsilon_{n+1},$$

with the pair independent and identically distributed. Under mild conditions on the distribution of (a_i, ϵ_i), the Markov chain X_n has a unique stationary distribution π which can be represented as the probability distribution of

$$X_\infty = \epsilon_0 + a_0\epsilon_1 + a_1a_0\epsilon_2 + \dots.$$

The point here is that for any k such that moments exist

$$E(X_1^k|X_0 = x) = E((a_1x + \epsilon_1)^k) = \sum_{i=0}^{k} \binom{k}{i} x^i E(a_1^i \epsilon_1^{k-i}).$$

If, for example, π has moments of all orders and is determined by those moments, then the Markov chain $\{X_n\}_{n=0}^\infty$ is generated by a compact operator with eigenvalues $E(a_1^i)$ $0 \le i < \infty$ and polynomial eigenfunctions.

We have treated the Gaussian case in Section 4.5. At the other extreme, take $|a| < 1$ constant and let ϵ_i take values ± 1 with probability $1/2$. The fine properties of π have been intensively studied as Bernoulli convolutions. See [19] and the references here. For example, if $a = 1/2$, then π is the usual uniform distribution on $[-1, 1]$ and the polynomials are Tchebychev polynomials. Unfortunately, for any value of $a \neq 0$, in the ± 1 case, the distribution π is known to be continuous while the distribution of X_n is discrete and so does not converge to π in L^1 or L^2. We do not know how to use the eigenvalues to get quantitative rates of convergence in one of the standard metrics for weak convergence.

As a second example take $(a, \epsilon) = (u, 0)$ with probability p and $(1 + u, -u)$ with probability $1 - p$ with u uniform on $(0, 1)$ and p fixed in $(0, 1)$. This Markov chain has a beta $(p, 1 - p)$ stationary density. The eigen values are $1/(k+1)$, $1 \le k < \infty$. It has polynomial eigenfunctions. Alas, it is not reversible and again we do not know how to use the spectral information to get usual rates of convergence. See [19] or [51] for more information about this so called "donkey chain".

7.2 Multivariate Models

The present paper and its companion paper have discussed univariate models. There are a number of multivariate models $f_\theta(x), \pi(\theta)$ with x or θ multivariate where the associated Markov chains have polynomial eigen functions. Some analogs of the six exponential families are developed in [14]. Preliminary thesis work of Khare and Zhou indicate that these exponential family chains have polynomial eigenfunctions.

An important special case – high dimensional Gaussian distributions, has been studied in [2, 40]. Here is a brief synopsis of these works. Let $m(x)$ be a p-dimensional normal density with mean μ and covariance Σ (i.e., $N_p(\mu, \Sigma)$). A Markov chain with stationary density m may be written as

$$X_{n+1} = AX_n + Bv + C\epsilon_{n+1}. \tag{7.2}$$

Here ϵ_n has a $N_p(0, I)$ distribution, $v = \Sigma^{-1}\mu$, and the matrices A, B, C have the form

$$A = -(D+L)^{-1}L^T,\ B = (D+L)^{-1},\ C = (D+L)^{-1}D^{1/2}$$

where D and L are the diagonal and lower triangular parts of Σ^{-1}. The chain (7.2) is reversible if and only if $\Sigma A = A^T\Sigma$. If this holds, A has real eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_p)$. In [40], Goodman and Sokal show that the Markov chain (7.2) has eigenvalues λ^K and eigenvectors H_K for $K = (k_1, k_2, \ldots, k_p)$ $k_i \geq 0$ with

$$\lambda^K = \prod_{i=1}^{p} \lambda_i^{k_i}, \qquad H_K(x) = \prod_{i=1}^{p} H_{k_i}(x_i)$$

where $H_k(x)$ are the usual one dimensional hermite polynomials. Goodman and Sokal show how a variety of stochastic algorithms, including the systematic scan Gibbs sampler for sampling from m, are covered by this framework. Explicit rates of convergence using these results remain to be carried out.

7.3 Conclusion

The present paper studies rates of convergence using spectral theory. In a companion paper we develop a stochastic approach which uses one eigen function combined with coupling. 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 $g(x)$ is totally positive of order two. This lets us give rates of convergence for the examples of Section 4 when 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. See [16, 19] for extensive references.

Acknowledgments We thank Jinho Baik, Alexi Borodin, Onno Boxma, Vlodic Bryc, Robert Griffiths, Len Gross, Susan Holmes, Murad Ismail, Christian Krattenthaler, Grigori Olshanski, Dennis Stanton and Hua Zhou for their enthusiastic help.

References

[1] Akhiezer, N. and Glazman, I. (1993). *Theory of Linear Operators in Hilbert Space*, Dover, N.Y.

[2] Amit, Y. (1996). Convergence properties of the Gibbs sampler for perturbations of Gaussians. *Ann. Statist.* **24**, 122–140.

[3] Anderson, W. (1991). *Continuous Time Markov Chains an Applications-Oriented Approach*, Springer, N.Y.

[4] Athreya, K., Doss, H. and Sethuraman, J. (1996). On the convergence of the Markov chain simulation method, *Ann. Statist.* **24**, 89–100.

[5] Baik, J., Kriecherbauer, T., McLaughlin, K. and Miller, P. (2006). Uniform asymptotics for polynomials orthogonal with respect to a general class of weights and universality results for associated ensembles. To appear, *Ann. Math. Studies*.

[6] Bakry, D. and Mazet, O. (2003). Characterization of Markov semigroups on $\mathbb{R}$ associated to some families of orthogonal polynomials, *Lecture Notes in Math*, Springer, New York.

[7] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

[8] Baxendale, P. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains, *Ann. Appl. Probab.* **15**, 700–738.

[9] Ben Arous, G., Bovier, A., and Gayard, V. (2003). Glauber dynamics of the random energy model, I, II, *Comm. Math Phys.* **235**, 379-425, **236**, 1–54.

[10] Brown, L. (1986). Fundamentals of statistical exponential families, Inst. Math. Statist., Hayward.

[11] Bryc, W. (2006). Approximation operators, exponential, and free exponential families. Preprint, Dept. of Math. Sci., University of Cincinnati.

[12] Buja, A.C. (1990) Remarks on functional canonical variates, alternating least square methods and ACE. *Ann. Statist.* **18**, 1032–1069.

[13] Cannings, C. (1974). The latent roots of certain Markov chains asrising in genetics: A new approach, I. Haploid Models, *Adv. Appl. Prob.* **6**, 260–290.

[14] Casalis, M. (1996). The $2d + 4$ simple quadratic families on $\mathbf{R}^d$, *Annal. Statist.* **24**, 1828–1854.

[15] Casella, G. and George, E. (1992). Explaining the Gibbs sampler, *Amer. Statistician* **46**, 167–174.

[16] Chamayou, J. and Letac, G. (1991). Explicit stationary distributions for compositions of random functions and products of random matrices, *Jour. Theoret. Probab.* **4**, 3–36.

[17] Chihara, T. (1978). *An Introduction to Orthogonal Polynomials*, Gordon and Breach, New York.

[18] Consonni, G. and Veronese, P. (1992). Conjugate priors for exponential families having quadratic variance functions, *Jour. Amer. Statist. Assoc.* **87**, 1123–1127.

[19] Diaconis, P. and Freedman, D. (1999). Iterated random functions, *SIAM Rev.* **41** 45–76.

[20] Diaconis, P. Khare, K., Saloff-Coste, L. (2006). Gibbs sampling, exponential families and coupling. Preprint, Dept. of Statistics, Stanford University.

[21] Diaconis, P. and Saloff-Coste, L. (1993). Comparison theorems for Markov chains, *Ann. Appl. Probab.* **3**, 696–730.

[22] Diaconis, P. and Saloff-Coste, L. (2006). Separation cut-offs for birth and death chains. To appear *Ann. Appl. Probab.*

[23] Diaconis, P. and Stanton, D. (2006). A hypergeometric walk. Preprint, Dept. of Statistics, Stanford University.

[24] Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families, *Ann. Statist.* **7**, 269–281.

[25] Diaconis, P. and Ylvisaker, D. (1985). Quantifying prior opinion. In *Bayesian Statistics* **2**, J Bernardo, et al. (editor), North Holland, Amsterdam.

[26] Donoho, D. and Johnstone, I. (1989). Projection-based approximation and a duality with kernel methods, *Ann. Statist.* **17**, 58–106.

[27] Dyer, M., Goldberg, L., Jerrum, M. and Martin, R. (2005). Markov chain comparison, *Probability Surveys* **3**, 89–111.

[28] Esch, D. (2003). The skew-t distribution: Properties and computations. Ph.D. Dissertation, Dept. of Statistics, Harvard University.

[29] Ewens, W. (2004), *Mathematical Population Genetics I. Theoretical Introduction*, 2nd Ed., Springer, N.Y.

[30] Feinsilver, P. (1986). Some classes of orthogonal polynomials associated with martingales, *Proc. Amer. Math. Soc.* **98**, 298–302.

[31] Feinsilver, P. (1991). Orthogonal polynomials and coherent states. In *Symmetries in Science*, **V**, Plenum Press, 159–172.

[32] Feinsilver, P. and Schott, R. (1993). *Representations and Probability Theory*, Kleiner Academic Press, Dordrecht.

[33] Feller, W. (1951). Diffusion processe: in Genetics, 2nd Berkeley Symposium on mathematical statistics, Univ. Calif. Press, Berkeley.

[34] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed., Wiley, N.Y.

[35] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd ed., Wiley, N.Y.

[36] Geman, S. and Geman, D. (1984). Stochastic relaxation Gibbs distributions and the Bayesian retoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

[37] Gilks, W. Richardson, S. and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

[38] Gill, J. (2002). Bayesian methods: A Social and Behavioral Sciences Approach, Chapman and Hall, Boca Raton.

[39] Glauber, R. (1963). Time dependent statistics of the Ising model, *Jour. Math. Phys.* **4**, 294–307.

[40] Goodman, J. and Sokal, A. (1984). Multigrid Monte Carlo method conceptual foundations, *Phys. Rev. D.* **40**, 2035–2071.

[41] Gross, L. (1979). Decay of correlations in classical lattice models at high temperatures, *Commun. Math. Phys.* **68**, 9–27.

[42] Gutierrez-Pena, E. and Smith, A. (1997). Exponential and Bayes in conjugate families: Review and extensions, *Test* **6**, 1–90.

[43] Harkness, W. and Harkness, M. (1968). Generlaized hyperbolic secant distributions, *Jour. Amer. Statist. Assoc.* **63**, 329–337.

[44] Ismail, M. (2005). *Classical and Quantum Orthogonal Polynomials*, Cambridge Press, Cambridge.

[45] Jones, G. and Hobart, J. (2001). Honest exploration of intractable probability distributions vis Markov chain monte carlo, *Statist. Sci.* **16**, 312–334.

[46] Karlin, S. and McGregor, J. (1961). The Hahn polynomials, formulas and application, *Scripta. Math.* **26**, 33–46.

[47] Koekoek, R. and Swarttouw, R. (1998). The Askey-scheme of hypergeometric orthogonal polynomials and its q-analog. http://math.nist.gov/opsf/projects/koekoek.html

[48] Jorgensen, C. (1997). *The Theory of Dispersion Models*, Chapman and Hall, London.

[49] Lehmann, E. and Romano, J. (2005). Testing Statistical Hypotheses, Springer, New York.

[50] Letac, G. (1992). *Lectures on Natural Exponential Families and Their Variance Functions.* Monografias de matematica, **50**, I.M.P.A., Rio de Janeiro.

[51] Letac, G. (2002) *Donkey walk and Dirichlet distributions.* Statist. Probab. Lett. 57 17–22.

[52] Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variance functions, *Ann. Statist.* **18**, 1–37.

[53] Liu, J., Wong, W. and Kong, A. (1995). Covariance structure and convergence rates of the Gibbs sampler with various scans, *Jour. Roy. Statist. Soc. B*, 157–169.

[54] Liu, J. (2001). *Monte Carlo Strategies in Scientific Compating*, Springer, New York,

[55] Malouche, D. (1998). Natural exponential families related to Pick functions, *Test*, 391–412.

[56] Meng, X.L., and Zaslavsky, A. (2002). Single observation unbiased priors, *Ann. Statisti.* **30**, 1345–1375.

[57] Meixner, J. (1934). Orthogonal polynom system mit einer Besonderth Gestalt der Erzengerder function, *Jour. London Math. Soc.* **9**, 6–13.

[58] Moreno, E. and Girón, F. (1998). Estimating with incomplete count data: A Bayesian approach, *Jour. Statist. Planning and Inference* **66**, 147–159.

[59] Morris, C. (1982). Natural exponential families with quadratic variance functions, *Ann. Statist.* **10**, 65–80.

[60] Morris, C. (1983). Natural exponential families with quadratic variance functions: Statistical theory, *Ann. Statist.* **11**, 515–589.

[61] Newman, M. and Barkema, G. (1999). *Monte Carlo Methodsin Statistical physics*, Oxford Press, oxford.

[62] Pommeret, D. (1996). Natural exponential families and Lie algebras, *Expo. Math.* **14**, 353–381.

[63] Ringrose, J. (1971) *Compact Non-Self-Adjoint Operators*, Van Nostrand, New York.

[64] Rosenthal, J. (1995) Minorization conditions and convergence rates for Markov chain monte carlo, *Jour. Amer. Statist. Assoc.* **90**, 558–566.

[65] Rosenthal, J. (1996). Analysis of the Gibbs sampler for a model related to James-Stein estimations, *Statist. Comput.* **6**, 269–275.

[66] Saloff-Coste, L. (2004) Total variation lower bounds for finite Markov chains: Wilson's lemma. In Random Walks and Geometry (V. Kaimanovich, W. Woess eds.), 515–532, de Gruyter, Berlin.

[67] Silver, J. S. (1996) *Weighted Poincaré and exhaustive approximation techniques for scaled Metropolis-Hasings Algorithms and spectral total variation convergence bounds in infinite commutable Markov chain theory.* PhD Thesis, Department of Mathematics, Harvard University.

[68] Szego, G. (1959) *Orthogonal Polynomials*. Revised edition, *Amer. Math. Soc.*, New York.

[69] Tanner, M. and Wong, W. (1987) The calculation of posterior distributions by data augmentation, (with discussion), *Jour. Amer. Statist. Assoc.* **82**, 528–550.

[70] Tierney, L. (1994) Markov chains for exploring posterior distributions, (with discussion), *Ann. Statist.* **22**, 1701–1762.

[71] Van Doorn, E.A. (2003) Birth-death processes and associated polynomials. In *Proceedings of the Sixth International Symposium on Orthogonal Polynomials, Special Functions and their Applications* (Rome, 2001). J. Comput. Appl. Math. **153**, 497–506.

[72] Wilson, D. (2004). Mixing times of Lozenge tiling and card shuffling Markov chains, *Ann. Appl. Probab.* **14**, 274–325.