– 371 (revised) –

Small sets and Markov transition densities

by

W.S. Kendall and G. Montana

February 2001

DEPARTMENT OF STATISTICS

UNIVERSITY OF WARWICK

Small sets and Markov transition densities

Wilfrid S Kendall [*] and Giovanni Montana [†]

18th July 2002

Abstract

The theory of general state-space Markov chains can be strongly related to the case of discrete state-space by use of the notion of *small sets* and associated *minorization conditions*. The general theory shows that small sets exist for all Markov chains on state-spaces with countably generated σ-algebras, though the minorization provided by the theory concerns small sets of order n and n-step transition kernels for some unspecified n. Partly motivated by the growing importance of small sets for Markov chain Monte Carlo and Coupling from the Past, we show that in general there need be no small sets of order $n = 1$ even if the kernel is assumed to have a density function (though of course one can take $n = 1$ if the kernel density is continuous). However $n = 2$ will suffice for kernels with densities (integral kernels), and in fact small sets of order 2 *abound* in the technical sense that the 2-step kernel density can be expressed as a countable sum of nonnegative separable summands based on small sets. This can be exploited to produce a representation using a latent discrete Markov chain; indeed one might say, inside every Markov chain with measurable transition density there is a discrete state-space Markov chain struggling to escape. We conclude by discussing complements to these results, including their relevance to Harris-recurrent Markov chains and we relate the counterexample to Turán problems for bipartite graphs.

Keywords: COUPLING FROM THE PAST, DATA-MINING, GRAPHICAL MODELS, LATENT DISCRETIZATION, MARKOV CHAIN MONTE CARLO, MINORIZATION CONDITION, PSEUDO-SMALL SETS, SMALL SETS, TRANSITION PROBABILITY DENSITY, TURÁN PROBLEM, ZARANKIEWICZ PROBLEM.

AMS Subject Classification: 60J05; 60J22, 65C40

1 Introduction

The notion of a small set was introduced to Markov chain theory by various writers (see for example [18]) and has been exploited to produce a reduction to the discrete case of Markov chain theory for general state-spaces (see Nummelin [17] and Meyn and Tweedie [14] for treatments in book form). The basic idea is to elicit a *minorization condition* for a given Markov chain:

[*] Supported by EPSRC grant GR/L56831 and EU network ERB-FMRX-CT96-0095.
Corresponding address: Statistics Department, University of Warwick, Coventry CV4 7AL, UK.
Email: w.s.kendall@warwick.ac.uk
[†] Supported by University of Warwick research studentship

Definition 1.1 *The transition probability kernel* $K(x, \cdot)$ *satisfies a* minorization condition (of order n) *if for some non-vanishing non-negative function* g *and some probability measure* μ *we have*

$$K^{(n)}(x, A) \quad \geq \quad g(x)\mu(A)$$

for all x*, all measurable* A*. In particular a set* C *is a* small set (of order n) *if its indicator function can occur together with a constant* $\rho \in (0, 1)$ *as* $g(x) = \rho\mathbb{I}_{[C]}$ *in a minorization condition of order* n*.*

The minorization can be used to produce the *split-chain construction* of Nummelin [16] – see also Athreya and Ney [1] where small sets are used for regeneration arguments – and hence to control convergence to equilibrium: as Nummelin wrote, "the 'elementary' techniques and constructions based on the notion of regeneration, and common in the study of discrete chains, can now be applied in the general case" [17, page *ix*]. More recently small sets have been used by Rosenthal [23] to establish rates of convergence for Markov chain Monte Carlo (see also the extended notion of *pseudo-small sets* described by Roberts and Rosenthal [20, 21]) and also (under the rubric of *gamma-coupling*) to produce effective Coupling from the Past (CFTP) constructions in the work of Murdoch and Green [11, 15] (see also some exciting new work on *catalytic perfect simulation* by Breyer and Roberts [5, 4]).

Closely related to the ideas presented here is the discretization proposed by Robert [19], originally devised for the purposes of Markov chain Monte Carlo convergence assessment. This discretization is based on sub-sampling of a discrete sequence derived from a continuous state-space Markov chain $\{X_n; n \geq 0\}$ depending on a sequence of renewals times, in the following way. Suppose that X_n possesses several disjoint small sets C_i, with $i = 1, \ldots, I$ for which the minorization condition of Definition 1.1 holds with constants ρ_i and measures μ_i. The C_i need not necessarily form a partition of the whole state-space. Suppose the above splitting construction is applied whenever X visits one of the C_i. Define the renewal times $\tau_0 = 1$ and τ_n, with $n \geq 1$ by:

$$\tau_n \quad = \quad \inf \Big\{ t > \tau_{n-1} : X_{t-1} \in C_i \text{ for some } i \in \{1, ..., I\} \\ \text{and regeneration occurs at time } t \Big\}$$

Robert shows that the finite valued sub-sequence η_n obtained from X_t by:

$$\eta_n = i \text{ if } X_{\tau_n - 1} \in C_i$$

is a homogeneous Markov chain defined on the finite state-space $\{1, \ldots, I\}$.

The theory of general Markov chains assures us of the existence of small sets, but gives no guarantees concerning the order. For the purposes of establishing convergence results this is of no great importance; however order 1 is required for current CFTP applications. This raises the question, for what sort of Markov chains can one guarantee existence of small sets of order 1? As a straightforward exercise in mathematical analysis at an advanced undergraduate level, one can show existence for state-space a smooth manifold when the kernel has a continuous density $p(x, y)$, and indeed then one can show small sets of order 1 *abound*, in the sense that they can be used to produce a representation:

$$p(x, y) \quad = \quad \sum_{i=1}^{\infty} f_i(x)g_i(y) \tag{1}$$

where the $f_i(x)$ are non-negative continuous functions supported on small sets, and the $g_i(y)$ are probability density functions. From this representation one can further deduce the existence of a *latent discrete Markov chain*: since $\int p(x, y)\,\mathrm{d}y = 1$ it follows that $\sum_i f_i(x) = 1$ for all x, and so $f_i(x)$ may be viewed as a transition probability density describing transitions from the state-space to a latent countable state-space $\{1, 2, \ldots\}$; and the entire stochastic dynamics of the original chain can be viewed as derived from a discrete state-space chain with transition probability matrix of entries

$$p_{ij} \quad = \quad \int g_i(y) f_j(y)\,\mathrm{d}y\,. \tag{2}$$

(*Finite* versions of such constructions, *finite-rank Markov chains*, are used to derive limit theorems in [25, 13]; see also [22].) We continue this line of enquiry in more detail in section §5.

However this particular representation fails hopelessly as soon as we move to the slightly more general category of Markov chains with measurable transition probability densities! Even the obvious step of allowing the f_i and g_i to be measurable is of no avail. For, as we show in the next section, there exist transition probability densities for which there are *no* non-trivial small sets of order 1. The construction is based on the construction of a Borel subset of the unit square with no non-null subsets of measurable rectangle form, and is related to a variant of the Turán problem from extremal graph theory.

However, and somewhat to our initial surprise, the cause of measurable transition densities is not entirely lost. As we show in section §3, so long as we move to order 2 we *can* construct non-trivial small sets (following known techniques for establishing the existence of small sets), and in fact they *abound* in the sense that one can build representations of the 2-step transition probability density $p^{(2)}(x, y)$ generalizing that of Eq. (1), and hence derive an interlacing latent discretization with transition matrix generalizing Eq. (2). Moreover this discretization uses only the measurable structure of the underlying space, rather than its topology: one need only suppose the state-space σ-algebra to be countably generated. In Section §4 we use the method of §3 to show that the weaker notion of *pseudo-small sets* [20, 21] results in the presence of many pseudo-small sets even at order 1; however this weaker notion is too weak to allow us to construct latent discretizations. In the concluding section §5 we discuss the latent discretization, and various complements including the extent to which the discretization can be generalized yet again, if one wishes to consider Markov chains whose kernels do not possess transition densities.

Acknowledgements

As well as the financial support noted above, we gratefully acknowledge Gareth Roberts' suggestion to consider pseudo-small sets, Jeff Rosenthal's pointer to the literature on finite rank Markov chains, and many other helpful conversations with our Warwick colleagues Saul Jacka, Jim Smith, and Jon Warren.

2 Measurable transition densities may have no non-null small sets of order 1

This section relies on a simple combinatorial calculation, known to graph theorists in a considerably refined form (see for example [9, 10]). We present a self-contained exposition, yielding as a first step a probabilistic construction of a measurable subset of $[0,1]^2$ which is "rectangle-free", which is to say, contains no non-null measurable rectangles. It should be clear to anyone who has studied measure theory that such sets must exist: however we have not been able to find a construction in the literature.

The combinatorial aspect concerns *arrays of cells*, $n \times n$ square lattices, the nodes of which are viewed as square cells of sidelength $\frac{1}{n}$, either filled or not, and arranged to pack the unit square. Unions of filled cells form *pixellated subsets* of $[0,1]^2$. We will be interested in whether we can find non-negligible *filled measurable rectangles*: pixellated subsets corresponding to unions of cells of the form

$$\{\text{ cell } (x_i, y_j) : i = 1, \ldots, r,\ j = 1, \ldots, s\}$$

defined by subsequences $x_1, \ldots, x_r$ and $y_1, \ldots, y_s$ where r and s amount to substantial fractions of n. The basic combinatorial argument constructs random subsets of arrays of cells which have low probability of containing measurable rectangles which are not very small. A Borel-Cantelli argument can then be applied to intersections of the corresponding pixellated subsets, so as to derive the following result.

Theorem 2.1 *There exist Borel measurable subsets $E \subset [0,1]^2$ of positive area which are* rectangle-free*, so that if $A \times B \subseteq E$ then* area$(A \times B) = 0$.

Proof:
Recall Stirling's asymptotic approximation:

$$n! \quad \sim \quad \exp\left(n\left(\log n - 1\right) + \frac{1}{2}\log(2\pi n)\right) \qquad \text{as } n \to \infty\,. \tag{3}$$

For fixed rational $\alpha \in (0,1)$ we apply Stirling's approximation to the formula for the mean number of $\lfloor \alpha n \rfloor \times \lfloor \alpha n \rfloor$ filled measurable rectangles to be found in an $n \times n$ array of cells of side-length $\frac{1}{n}$, such that cells are filled independently with fill probability p. (Here $\lfloor x \rfloor$ is the greatest integer smaller than x.) We obtain

$$\text{mean number of such measurable rectangles} \;=\; \binom{n}{\lfloor \alpha n \rfloor}^2 p^{\lfloor \alpha n \rfloor^2} \quad \sim$$
$$\exp\left(n^2\left(\alpha^2 \log p\right) - 2n\left(\alpha \log \alpha + (1-\alpha)\log(1-\alpha)\right) + \log\left(2\pi n \alpha(1-\alpha)\right)\right)$$

(at least for n running through the subsequence for which αn is an integer!).

We apply Markov's inequality to deduce that for fixed $\varepsilon > 0$ and $p \in (0,1)$

$$\begin{aligned}\mathbb{P}\left[\text{ at least one } \lfloor \alpha n \rfloor \times \lfloor \alpha n \rfloor \text{ filled measurable rectangle }\right] \quad &\leq \\ (1+\varepsilon) \times \exp\Big[n^2\left(\alpha^2 \log p\right) - 2n\left(\alpha \log \alpha + (1-\alpha)\log(1-\alpha)\right) & \\ + \log\left(2\pi n \alpha(1-\alpha)\right)\Big] &\end{aligned} \tag{4}$$

for all $n \geq N = N(\varepsilon, \alpha, p)$ such that αn is an integer. Clearly the upper bound tends to zero as $n \to \infty$ through the relevant subsequence. Moreover the mean area of the corresponding pixellated random set is given by $n^2 p / n^2 = p$.

We now construct a random subset Ξ of the unit square $[0,1]^2$ as the intersection

$$\Xi \quad = \quad H_{k_0} \cap H_{k_0+1} \cap \ldots$$

of a sequence $H_{k_0}, H_{k_0+1}, \ldots$ of such pixellated random sets. The set H_k is constructed as the union of filled cells in an $n_k \times n_k$ array of cells of side-length $\frac{1}{n_k}$, such that cells are filled independently with fill probability p_k. We fix $\varepsilon > 0$ and select

$$\begin{aligned} \alpha &= \alpha_k = \frac{1}{k} \\ p &= p_k = 1 - 2^{-k} \\ n &= n_k = \inf\left\{ r > 2^k \vee N(\varepsilon, \alpha_k, p_k) \ : \ \alpha r \text{ is an integer} \right\} . \end{aligned} \tag{5}$$

The mean area of Ξ is bounded below by

$$\mathbb{E}\left[\, \text{area}(\Xi) \,\right] \quad \geq \quad 1 - \sum_{k=k_0}^{\infty} (1 - \mathbb{E}\left[\, \text{area}(H_k) \,\right]) \quad = \quad 1 - 2^{1-k_0} \,,$$

and therefore Ξ has a positive chance of having positive area (at least if $k_0 > 1$).

On the other hand we may apply the first Borel-Cantelli lemma to show that all but finitely many of the events

$$R_k \quad = \quad \left\{ H_k \text{ contains no measurable rectangles of sidelength } \tfrac{1}{k} \text{ or greater} \right\}$$

must occur. For geometrical arguments show that the failure of R_k forces the corresponding cell array to contain at least one $\lfloor \alpha n \rfloor \times \lfloor \alpha n \rfloor$ filled measurable rectangle, and by the bound Eq. (4) the failure-probability of this event is therefore bounded above by

$$\text{constant} \times \left(1 - 2^{-k}\right)^{\frac{1}{2} n_k^2 / k^2} \ \leq \ \text{constant} \times e^{-2^{-k-1} n_k^2 / k^2} \ \leq \ \text{constant} \times e^{-2^{k-1}/k^2} \,.$$

This is summable, and so the first Borel-Cantelli lemma applies.

It follows that almost surely Ξ is rectangle-free, in the sense that if A and B are measurable subsets of $[0,1]$ with $A \times B \subseteq \Xi$ then area$(A \times B) = 0$. Figure 1 illustrates (an approximation of) this random construction. □

Remark 2.2 The above randomization argument can be replaced, at the price of more complexity, by a counting argument, demonstrating the existence of a counterexample $E \subset [0,1]^2$ of area prescribed to lie in the range $(0,1)$.

The indicator function for the random set Ξ nearly provides a Markov transition density under normalization, except that this normalization will fail when a slice along a fixed x has zero length. However this is easily fixed in any one of several ways, yielding the following corollary.

Corollary 2.3 *There exist measurable Markov transition densities for which there are no non-null small sets of order* 1*.*

Proof:
Suppose Ξ_1, Ξ_2, … are independent copies of Ξ as constructed in Theorem 2.1, but affinely transformed to fit into the rectangles

$$[0,1] \times [1/2, 1), [0,1] \times [1/4, 1/2), \ldots .$$

Consider the union

$$\Xi^* \quad = \quad \Xi_1 \cup \Xi_2 \cup \ldots \,,$$

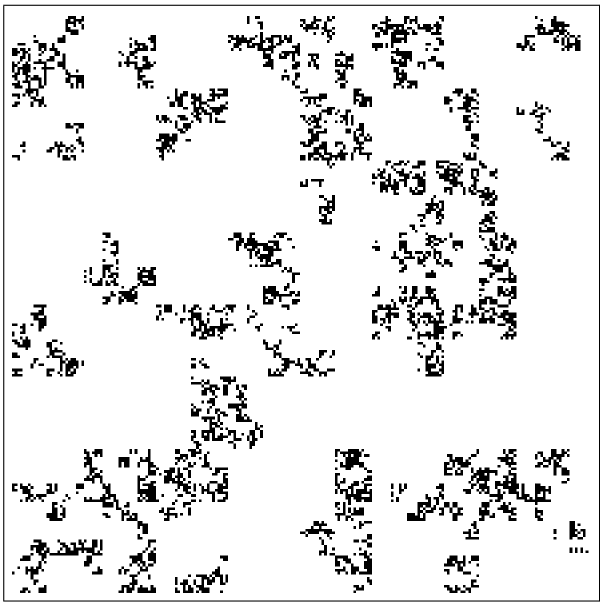

Figure 1: Example of rectangle-free random set Ξ.

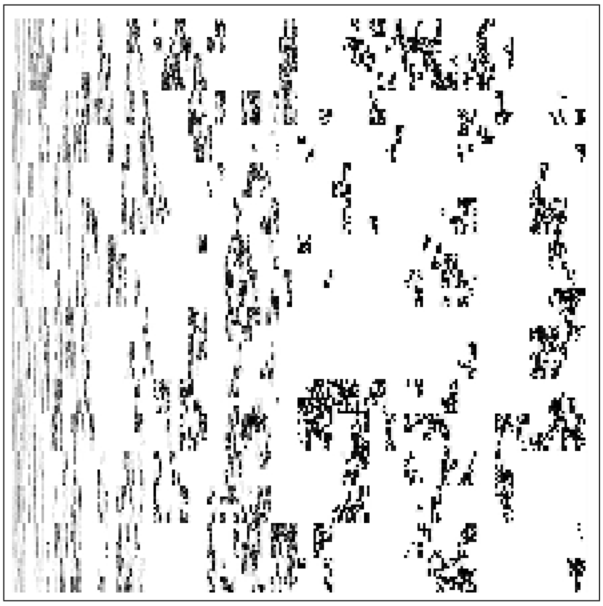

Figure 2: Example of rectangle-free random set Ξ^* with x-slices almost all of positive length.

as illustrated in Figure 2.

A slice of Ξ^* along fixed x (an *x-slice*) can have zero length only if its component x-slices along each of the Ξ_i have zero length. The component x-slices are independent and (saving only an exceptional null-set of x values corresponding to vertical cell boundaries) the chance of a component x-slice having non-zero length is positive and is the same for each component (by construction of the Ξ_i). Therefore independence shows that for non-exceptional x the x-slice of Ξ^* is almost surely of positive length.

Thus the following defines a Markov transition density for which there are no non-null small sets of order 1:

$$p(x, y) \quad = \quad \frac{\mathbb{I}_{[\Xi^*]}(x, y)}{\int_0^1 \mathbb{I}_{[\Xi^*]}(x, z)\, \mathrm{d}z} \tag{6}$$

where the ratio is taken to equal 1 for those x for which the denominator vanishes (only a null-set and therefore negligible). Existence of a non-null small set of order 1 would entail a lower bound

$$p(x, y) \quad \geq \quad \rho \mathbb{I}_{[B]}(y)$$

for all $x \in A$, for some positive ρ and non-null Borel sets A, $B \subset [0, 1]$. Hence (possibly reducing A somewhat) we would obtain a non-null measurable rectangle subset of Ξ, in contradiction to the assertion of Theorem 2.1. □

An alternative method of proof uses monotonic transformation of the x-axis to remove all but a null-set of coordinates at which x-slices have length-zero intersection with Ξ.

Remark 2.4 A refinement of this approach produces a rectangle-free symmetric subset $\Xi \subset [0, 1]^2$, symmetric in the sense that $(x, y) \in \Xi$ if and only if $(y, x) \in \Xi$. Simply modify the filling procedure of Theorem 2.1 so that cell (x, y) is filled if and only if cell (y, x) is filled, but otherwise cells are filled independently. The resulting random set Ξ is symmetric. Suppose $A \times B \subseteq \Xi$. Choose median values s, t such that $\text{length}(A \cap [0, s]) = \frac{1}{2}\,\text{length}(A)$, $\text{length}(B \cap [0, t]) = \frac{1}{2}\,\text{length}(B)$. If $s < t$ then $(A \cap [0, s]) \times (B \cap [t, 1])$ lies in the upper triangle $\Xi \cap \{(x, y) : x < y\}$; otherwise $(A \cap [s, 1]) \times (B \cap [0, t])$ lies in the lower triangle. Either way we exhibit a measurable rectangle subset of Ξ of measure $\frac{1}{4}\,\text{area}(A \times B)$ lying in a region which could have been produced by the original construction of Theorem 2.1 and therefore must have zero area. It follows that Ξ is not only symmetric but also rectangle-free.

Remark 2.5 Yet a further refinement can be used to produce a *reversible* Markov chain with no order-1 small sets, thus answering a question raised by Gareth Roberts. We sketch the construction of a transition density $p(x, y)$ on the unit square which is symmetric (hence doubly stochastic) and which takes only the values 0, 1, and 2.

We start with $p_0(x, y) \equiv 1$, and use the notation of Theorem 2.1, but increase the n_k if necessary so as to ensure they are all even. In order to maintain the doubly stochastic property we use moves developed for Markov chain Monte Carlo on contingency table configurations: at level k, independently with probability $1 - p_k = 2^{-k}$ for each of the $n_k^2/4$ cells of dimension $n_k^{-1} \times n_k^{-1}$ in the upper-left quadrant, if p_{k-1} is non-zero in that cell we reduce its value there to 0, add the removed mass uniformly over the cell which is its mirror image in $x = 1/2$, and alter p_{k-1} in the other two quadrants so as also to maintain mirror symmetry in the $y = 1/2$ axis. If on the other hand p_{k-1} is zero in the chosen cell then we perform the reverse move. We set p_k to be the result of these operations.

The support of p_k is similar to the set Ξ_k, except that, when proceeding from Ξ_k to Ξ_{k+1}, as far as the first quadrant is concerned, we add a union with $\Xi_k^c \setminus H_{k+1}$ as well as taking the intersection $\Xi_k \cap H_{k+1}$. The counting arguments are easily modified to take account of this, thus showing that the limiting support set is rectangle-free.

Finally we need to show that $p_k(x, y)$ converges to a limiting probability density. For any given point (x, y) the probability of $p_{k+1}(x, y) \neq p_k(x, y)$ is $1-p_k = 2^{-k}$. So by the first Borel-Cantelli lemma the sequence $\{p_k(x, y) : k = 1, 2, \ldots\}$ converges for almost all (x, y). Since p_k is bounded between 0 and 2, the limiting probability density $p_\infty(x, y)$ exists as a consequence of the Lebesgue dominated convergence theorem, and has the doubly stochastic property. By construction of the support set, it can have no non-trivial small sets of order 1.

3 Small sets of order 2 abound for measurable transition densities

A careful reading of the methods employed in the proof of the existence of small sets (see, *eg*, [17, §2.3], [14, §5.2] and also [18]) reveals that if a Markov chain with countably generated state-space σ-algebra has a measurable transition density then it possesses a small set of order 2. Here we give a variation on this proof which additionally shows that such small sets *abound*, in the sense that the 2-step transition density can be represented as a sum of non-negative separable terms involving small-set decompositions.

First note that the question posed (to show such Markov chains have small sets of order 2) is strictly measure-theoretic. Indeed we can suppose the reference probability measure to be atom-free (for otherwise we can immediately exhibit small sets based on the atoms). Furthermore we may identify states which are not separated by the σ-algebra. Any countable sequence of sets generating the state-space algebra can be used to map the state-space into the unit interval $[0, 1]$ in a standard way, expanding each $x \in [0, 1]$ in a dyadic expansion and mapping each state s to a dyadic expansion determined by which members of the countable generating sequence contain s. This map fails to be $1 : 1$ only at a countable number of $x \in [0, 1]$ where it will be $2 : 1$: we may delete the corresponding null-set from the state-space. We have thus reduced the state-space to the unit interval $[0, 1]$ furnished with a reference probability measure which is atom-free. Deleting a countable number of further null-sets, we may transform $[0, 1]$ using the distribution function for the reference probability measure so as to produce a state-space which is $[0, 1]$ furnished with Lebesgue measure.

In the remainder of this section we can therefore, without any loss of generality, confine our attention to the case of the unit interval furnished with Lebesgue measure as reference measure.

We begin with a general lemma, which uses Egoroff's theorem and the Lebesgue density theorem to establish near-L^1-continuity for functionals derived from L^1 functions on the unit square. Introduce the notation

$$p_x(\cdot) \quad = \quad p(x, \cdot)$$

and notice that by Fubini's theorem p_x may be viewed as a mapping from almost all $x \in [0, 1]$ into $L^1([0, 1])$.

Lemma 3.1 *Let $p(x, y)$ be an integrable function on $[0, 1]^2$. Then we can find subsets $A_\varepsilon \subset [0, 1]$, increasing as ε decreases, such that*

(a) for any fixed A_ε the "L^1-valued function" p_x is uniformly continuous on A_ε: for any $\eta > 0$ we can find $\delta > 0$ such that $|x - x'| < \delta$ and $x, x' \in A_\varepsilon$ implies

$$\int_0^1 |p_x(z) - p_{x'}(z)| \, \mathrm{d}z \quad < \quad \eta\,;$$

(b) every point x in A_ε is of full relative density: as $u, v \to 0$ so

$$\frac{\text{length}([x - u, x + v] \cap A_\varepsilon)}{u + v} \quad \to \quad 1\,.$$

Remark 3.2 In some sense this result must have been immediately accessible to early workers in the field: it bears a family resemblance to techniques used by Doob in [7, pages 199-202] for which Doob himself credits the essential idea to Doeblin [6]. However we have not been able to find in the literature anything resembling the application, Corollary 3.7.

Proof:
We use a modification of the celebrated consequence of Egoroff's theorem [12, §21, Theorem A], that every measurable function is "nearly" uniformly continuous, in the sense of being uniformly continuous off sets of arbitrarily small measure. This is usually stated for real-valued functions, but applies to such functions as p_x so long as we use L^1-continuity. For consider: we can L^1-approximate the underlying function $p(x, y)$ by a continuous function $f_1(x, y)$

$$\int_0^1 \int_0^1 |p(x, y) - f_1(x, y)| \, \mathrm{d}x \, \mathrm{d}y \quad < \quad \alpha\,.$$

for any fixed $\alpha \in (0, 1)$. Adding further continuous functions $f_2(x, y)$, ..., $f_n(x, y)$, ... we can require the approximation to improve geometrically:

$$\int_0^1 \int_0^1 |p(x, y) - (f_1(x, y) + \ldots f_n(x, y))| \, \mathrm{d}x \, \mathrm{d}y \quad < \quad \alpha^n\,.$$

By Markov's inequality, if

$$D_n \quad = \quad \{x : \int_0^1 |p(x, y) - (f_1(x, y) + \ldots f_n(x, y))| \, \mathrm{d}y > \alpha^{n/2}\}$$

then

$$\text{length}(D_n) \quad \leq \quad \alpha^{n/2}\,.$$

Thus off the union $D_k \cup D_{k+1} \cup \ldots$ we can approximate $p(x, y)$ uniformly by uniformly continuous functions. The total area of the union is at most $\alpha^k/(1 - \alpha)$, hence can be made arbitrarily small by increasing k.

Consequently for every $\varepsilon \in (0, 1)$ we can find a subset $A_\varepsilon \subseteq [0, 1]$ of measure at least $1 - \varepsilon$ and such that $x \mapsto p_x$ is uniformly L^1-continuous on A_ε. Moreover we may arrange for $A_\varepsilon \subseteq A_{\varepsilon'}$ whenever $\varepsilon > \varepsilon'$.

Now invoke the Lebesgue density theorem [24, Theorem 8.8]: the subset of points failing to have full relative density in a measurable subset is always of measure zero.

Since the above construction of A_ε actually only uses a countable number of set complements $(D_k \cup D_{k+1} \cup \ldots)^c$, we can simply remove all such points for each of the countably many complements. The lemma follows. $\square$

We now state and prove the central result of this section, establishing abundance of small sets in a rather specific fashion. We recall the discussion at the start of this section, demonstrating that this result will actually apply for any state-space with countably generated σ-algebra and atom-free reference probability measure: for the sake of simplicity we state it for the case of state-space $[0, 1]$ with Lebesgue measure as reference measure.

In the following we continue with the notation of Lemma 3.1, and note that $q_y(\cdot) = p(\cdot, y)$ possesses a similar property: let $\{B_\varepsilon : \varepsilon \in (0, 1)\}$ denote a corresponding monotone family of sets for which uniform continuity of q_y and full relative density hold.

Theorem 3.3 *Let $p(x, y)$, x, $y \in [0, 1]$, be a measurable probability transition density (so $\int_0^1 p(x, y)\,\mathrm{d}y = 1$ for all x) and let $\eta \in (0, 1)$. For almost all x, $y \in [0, 1]$ the two-step transition density*

$$p^{(2)}(x, y) \quad = \quad \int_0^1 p(x, z)p(z, y)\,\mathrm{d}z \quad = \quad \int_0^1 p_x(z)q_y(z)\,\mathrm{d}z$$

is subject to lower bounds of the form

$$p^{(2)}(x', y') \quad \geq \quad (1 - \eta)p^{(2)}(x, y)$$

for all $x' \in [x - u, x + u]$ save for a set of measure δu, all $y' \in [y - u, y + u]$ save for a set of measure δu, for all sufficiently small positive u (depending on η, δ in the range $(0, 1)$).

Remark 3.4 This result differs from the classic small-set existence result (*eg* [17, Thm. 2.1], [14, Thm. 5.2.1]) in showing that small-set minorization conditions for the 2-step transition density

$$p^{(2)}(x', y') \quad \geq \quad (1 - \eta)p^{(2)}(x, y)$$

can be established to hold for almost all x, y, over a suitable measurable rectangle near to (x, y) and for η arbitrarily close to 0. It is for this reason that we require Lemma 3.1 rather than the more direct methods of the classic result. We need the stronger result in order to obtain the "abundance" Corollary 3.7.

Remark 3.5 The result can be viewed as a Markov chain generalization of *Steinhaus' theorem* [2, Theorem 1.1.1], that $\{x - y : x, y \in E\}$ contains an open interval containing 0 if $E \subset \mathbb{R}$ is of positive Lebesgue measure.

Remark 3.6 In fact the proof remains valid if $p^{(2)}(x, y)$ is actually obtained as the convolution of two different probability transition densities $p(x, y)$ and $q(x, y)$. Moreover we use the normalization property $\int_0^1 p(x, y)\,\mathrm{d}y = 1$ simply to ensure non-triviality of p. Of course non-negativity is essential if the notion of small set is to make sense as stated in Definition 1.1.

Proof:
Consider $x \in A_\varepsilon$, $y \in B_\varepsilon$, set $\rho^{(2)} = p^{(2)}(x, y)$, and fix $\eta \in (0, 1)$. The result is immediate for $\rho^{(2)} = 0$. So suppose $\rho^{(2)} > 0$.

Neither p_x nor q_y need be bounded: however we can apply the monotone convergence theorem to deduce the existence of K such that

$$\rho^{(2)} \quad \geq \quad \int_0^1 (p_x(z) \wedge K)\,(q_y(z) \wedge K)\ \mathrm{d}z \quad > \quad \rho^{(2)}(1 - \eta/2)\,.$$

Now select u such that

(a) length$([x-u, x+u] \cap A_\varepsilon) > (1-\delta)u$, length$([y-u, y+u] \cap B_\varepsilon) > (1-\delta))u$,

(b) for $x' \in [x-u, x+u] \cap A_\varepsilon$, $y' \in [y-u, y+u] \cap B_\varepsilon$ we have

$$\int_0^1 |p_x(z) - p_{x'}(z)|\ \mathrm{d}z \ < \ \frac{\eta\rho^{(2)}}{4K}\,, \qquad \int_0^1 |q_y(z) - q_{y'}(z)|\ \mathrm{d}z \ < \ \frac{\eta\rho^{(2)}}{4K}\,.$$

Hence for $x' \in [x-u, x+u] \cap A_\varepsilon$, $y' \in [y-u, y+u] \cap B_\varepsilon$ we can deduce

$$\begin{aligned}\rho^{(2)}(1-\eta/2) \quad &< \quad \int_0^1 (p_x(z) \wedge K)\,(q_y(z) \wedge K)\ \mathrm{d}z \\ &\leq \quad \frac{\eta\rho^{(2)}}{2} + \int_0^1 p_{x'}(z) q_{y'}(z)\,\mathrm{d}z \quad = \quad \frac{\eta\rho^{(2)}}{2} + p^{(2)}(x', y')\,.\end{aligned}$$

Thus

$$p^{(2)}(x', y') \quad > \quad (1-\eta)\rho^{(2)} \tag{7}$$

for all $x' \in [x-u, x+u] \cap A_\varepsilon$, $y' \in [y-u, y+u] \cap B_\varepsilon$. This establishes the result for $x \in A_\varepsilon$, $y \in B_\varepsilon$. But

$$\text{area}(A_\varepsilon \times B_\varepsilon) \quad \geq \quad (1-\varepsilon)^2$$

so the result holds for almost all x, y by letting $\varepsilon \to 0$.

Note that an order 2 small-set minorization follows whenever $\rho^{(2)} > 0$ (this must hold for more than a null-set of y for each x if the 2-step transition density is to integrate to 1): if $x \in A_\varepsilon$, $y \in B_\varepsilon$ then for all sufficiently small u we have

$$p^{(2)}(x', y') \quad > \quad \text{positive constant}$$

for all $(x', y') \in [x-u, x+u] \cap A_\varepsilon \times [y-u, y+u] \cap B_\varepsilon$. Note that, say,

$$\text{length}([x-u, x+u] \cap A_\varepsilon), \text{length}([y-u, y+u] \cap B_\varepsilon) \quad > \quad u/2 \quad > \quad 0$$

for small enough u (apply the Lebesgue density condition *(b)* of Lemma 3.1), so the minorization is non-trivial! □

The construction has been designed to furnish a rich supply of small sets, and we can use this to obtain a representation of $p^{(2)}(x, y)$ as a sum of non-negative separable terms involving small-set decompositions. In the informal terminology of Section 1, small sets of order 2 *abound*.

Corollary 3.7 *If $p(x, y)$ is a measurable transition probability density then we can represent the 2-step transition probability density as follows:*

$$p^{(2)}(x, y) \quad = \quad \sum_{i=0}^{\infty} \beta_i \mathbb{I}_{[C_i]}(x) \mathbb{I}_{[D_i]}(y) \tag{8}$$

for positive β_i and subsets C_i, $D_i \subseteq [0, 1]$, holding for almost all x, $y \in [0, 1]$.

Remark 3.8 It is of course not possible in general to arrange for the $C_i \times D_i$ to be disjoint, for this would force $p^{(2)}(x, y)$ to have an essentially countable range.

Remark 3.9 As hinted in the introduction, the impact of a representation such as the above is clearer if we write it in the equivalent form

$$p^{(2)}(x, y) \quad = \quad \sum_{i=0}^{\infty} \beta(x, i) r_i(y) \tag{9}$$

where $\beta(x, i)$ is a transition probability density from $[0, 1]$ to the set of positive integers $\{1, 2, \ldots\}$ (so $\sum_i \beta(x, i) = 1$ for all $x \in [0, 1]$) and the $r_i(y)$ are probability densities on $[0, 1]$. We pursue this further in the concluding section.

Proof:
Let $\mathcal{S}$ be a countable sequence of functions enumerating all functions of the form

$$s(x, y) \quad = \quad \operatorname{ess\,inf} \left\{ p^{(2)}(u, v) \ : \ u \in C \, , v \in D \right\} \times \mathbb{I}_{[C]}(x) \mathbb{I}_{[D]}(y)$$

where C and D are restricted to be of the form of intersections of dyadic rational intervals with $A_{1/h}$, $B_{1/h}$:

$$\begin{aligned} C \quad &= \quad [r2^{-k}, (r+1)2^{-k}) \cap A_{1/h} \\ D \quad &= \quad [s2^{-k}, (s+1)2^{-k}) \cap B_{1/h} \, , \end{aligned}$$

for non-negative integers r, s, and positive integers k, h. Observe that the function $f_n(x, y)$ which is the pointwise maximum of the first n of the functions in the sequence $\mathcal{S}$ can be re-written in the form

$$f_n(x, y) \quad = \quad \sum_{i=0}^{m_n} \beta_i \mathbb{I}_{[C_i]}(x) \mathbb{I}_{[D_i]}(y) \, ,$$

for a *fixed* sequence of positive constants β_i and dyadic rational intervals C_i, D_i. This is because an addition of a further member of $\mathcal{S}$ to the computation of the maximum can be re-expressed as an addition of the excess in the form of a number of terms of the form $\beta_i \mathbb{I}_{[C_i]} \mathbb{I}_{[D_i]}$.

Letting $n \to \infty$ we obtain

$$f_\infty(x, y) \quad = \quad \sup_n f_n(x, y) \quad = \quad \sum_{i=0}^{\infty} \beta_i \mathbb{I}_{[C_i]}(x) \mathbb{I}_{[D_i]}(y) \, .$$

By construction and using Theorem 3.3 we can deduce that $f_n(x, y)$ increases monotonically and converges to $p^{(2)}(x, y)$ whenever $x \in \bigcup_\varepsilon C_\varepsilon$ and $y \in \bigcup_\varepsilon D_\varepsilon$. Thus the corollary follows by the Monotone Convergence Theorem. For by Theorem 3.3 it follows, for each fixed $\eta \in (0, 1)$, for each $\varepsilon > 0$, that

$$p^{(2)}(u, v) \quad \geq \quad (1 - \eta) p^{(2)}(x, y) \qquad \text{for all } u \in C \, , v \in D$$

whenever C, D are intersections with A_ε, B_ε of dyadic rational intervals of sufficiently small size such that $(x, y) \in C \times D$. Hence we can find

$$s \quad = \quad \operatorname{ess\,inf} \left\{ p^{(2)}(u, v) \ : \ u \in C \, , v \in D \right\} \times \mathbb{I}_{[C]} \mathbb{I}_{[D]} \quad \in \quad \mathcal{S}$$

such that $s(x, y) \geq (1 - \eta)p^{(2)}(x, y)$, and so $f_n(x, y) \uparrow p^{(2)}(x, y)$ for almost all x, $y \in [0, 1]$. □

Remark 3.10 If the reference measure has atoms then these may immediately be converted into small sets and removed from the step-2 kernel, after which the methods of Corollary 3.7 can be applied to the residual. It follows that the 2-step transition probability density representation Eq. (9) applies whenever the chain has a measurable transition density and the state-space has countably generated σ-algebra, regardless of whether the reference measure has atoms or not.

4 Pseudo-small sets

Roberts and Rosenthal [20, 21] introduced the idea of a *pseudo-small set*; Definition 1.1 of a small set is weakened to allow the common component of the $K(x, \cdot)$ to depend on pairs of states x, x' being considered.

Definition 4.1 *A subset C of state-space is* pseudo-small *of order n if there is $\alpha > 0$ such that for each pair x, $y \in C$ we may find a probability measure $\nu_{x,y}$ with*

$$K^{(n)}(x, \cdot), K^{(n)}(y, \cdot) \quad \geq \quad \alpha\nu_{x,y}(\cdot)\,.$$

For C to be a small set we would require $\nu_{x,y}$ not to depend on x, y.

Pseudo-smallness is well-suited to questions involving coupling, but not for coalescence (as would arise in Coupling from The Past algorithms such as in [11, 15]), and not for representations as described in Corollary 3.7 above.

Nevertheless we place on record here that any Markov chain with measurable transition density $p(x, y)$ on a state-space with countably generating σ-algebra must have an abundant supply of pseudo-small sets of order 1.

Just as in §3 we may reduce to the case of state-space $[0, 1]$ with Lebesgue measure as reference measure. Now Lemma 3.1 shows that for any given $\varepsilon > 0$ we may find a subset $A_\varepsilon \subseteq [0, 1]$ such that the "L^1-valued function" $p_x(\cdot) = p(x, \cdot)$ is uniformly continuous on A_ε. This means that for any δ we can divide A_ε into a finite collection of subsets C (by taking intersections with intervals) such that if x, $y \in C$ then

$$\int_0^1 |p_x(z) - p_y(z)|\,\mathrm{d}z \quad \leq \quad \delta\,.$$

A direct computation then shows that

$$\int_0^1 \min\{p_x(z), p_y(z)\}\,\mathrm{d}z \quad \geq \quad 1 - \delta/2\,.$$

Consequently C may be taken to be pseudo-small of order 1, with $\alpha = 1 - \delta/2$ and with $\nu_{x,y}$ of density

$$\frac{1}{\alpha}\min\{p_x(z), p_y(z)\}\,.$$

By using a countable sequence of A_ε, we may cover almost all the state-space with pseudo-small sets of order 1 with α fixed as close to 1 as desired.

5 Conclusion and complements

Properly considered, neither the counterexample given in Theorem 2.1 nor the abundance of order 2 small sets of Theorem 3.3 should come as a surprise. Were no counterexample to exist, the theory of Lebesgue-measurable subsets of $[0,1]^2$ would take on an appalling simplicity, since every such set would be expressible as the union of a null-set and a countable family of measurable rectangles. On the other hand, convolution of densities tends to force positivity: were we to convolve with itself a kernel density $p(x, y)$ which was just a constant times the indicator of a Borel subset of $[0, 1]^2$ then the result would have a zero at (x, y) only if $p(x, z)p(z, y)$ vanished for almost all $z \in [0, 1]$, which would clearly be hard to arrange for a substantial portion of the range of possible $(x, y) \in [0, 1]^2$. This intuition lies at the heart of all existence proofs for small sets.

We have mentioned in Section §2 that the counterexample is related to issues in graph theory. The relevant theory is that of the Zarankiewicz problem [3], a Turán problem for bipartite graphs. Given a bipartite graph G on r and s vertices, how large do s, r have to be before G can be guaranteed to contain a specified complete bipartite graph as subgraph? In our language, a bipartite graph G on m and n vertices corresponds to a filled subset of an $m \times n$ array of cells (cell (i, j) being filled if vertex i in the first vertex collection is connected to cell j in the second); subgraphs which are complete bipartite correspond to filled measurable rectangles. Detailed estimates, running well beyond our simple requirements, are to be found in [9, 10].

A major motivation for this work is the usefulness of order 1 small sets in CFTP constructions. Of course in specific CFTP problems one constructs such small sets directly, often aided by continuity of the transition density. However it seems worth knowing that for rather general Markov chains one can always construct order 2 small sets (thus just one step away from the realm of practical application). *Finding* such small sets is another matter entirely, since their definition involves exactly the kind of integration which Markov chain Monte Carlo (MCMC), and CFTP in particular, has been invented to avoid! It would be most interesting if one could devise situations in which the existence of order 2 small sets could be exploited in CFTP without requiring such explicit integrations. (Notice however that our theorem guarantees that small sets of order 1 abound for Markov chains arising as discrete-time samples of *continuous time* Markov processes with measurable transition densities on state-spaces with countably generated σ-algebras!)

There are other contexts in which the results of this paper may be of interest. For example in data-mining, methods of *automatic binning* attempt to determine whether a parameter-space region R of interest can be expressed as $R = \bigcup_{k=1}^{K} C_k$, where each C_k is a product set [8, § 5]. Thus in the two-dimensional context one would be interested in searching for subsets $A \times B$ of R. Our example is of course absurdly pathological for this application, but hints at possible difficulties such a search might face. It also indicates a useful direction for further research: it would be interesting to relate theoretical work on automatic binning to the question of finding efficient representations of the form Eq. (3.7).

In the area of statistics known as Graphical Models one views a collection of random variables $\{Y_i : i \in G\}$ as indexed by vertices i of a graph G satisfying the following property: two subcollections $\{Y_i : i \in A\}$, $\{Y_i : i \in B\}$ are conditionally independent given a third subcollection $\{Y_i : i \in C\}$ if the vertex set C separates A from B in the graph G. One can code $\{Y_i : i \in A\}$, $\{Y_i : i \in B\}$, $\{Y_i : i \in C\}$ as random variables X_1, X_2, X_3. Suppose X_1, X_2, X_3 possess a joint density; the prediction of X_3 given X_1 *without* knowledge of the intervening X_2 is given by a kernel to which the results of Theorem 3.3 (and hence the latent discrete structure of Eq. (9)) apply.

It may be worth being more explicit about the latent discretization represented by Eq. (9). What this says is that we may view any Markov chain $X = \{X_0, X_1, \ldots\}$ with measurable transition density $p^{(2)}(x, y)$ on $[0, 1]$ (or of course a state-space with countable generated σ-algebra) as being generated by a latent discrete Markov chain $Y = \{Y_1, Y_3, \ldots\}$ running in "odd time". If

$$p^{(2)}(x, y) \quad = \quad \sum_{i=0}^{\infty} \beta(x, i) r_i(y) \tag{10}$$

as in Eq. (9), then Y is governed by the transition probability matrix

$$p_{ij} \quad = \quad \int_0^1 r_i(z) \beta(z, j) \, \mathrm{d}z \, .$$

Furthermore, given $Y_{2n+1} = i_{2n+1}$ and $Y_{2n+3} = i_{2n+3}$, the conditional density of X_{2n+2} is proportional as a function of z to

$$r_{i_{2n+1}}(z) \beta(z, i_{2n+3}) \, ,$$

and does not further depend on other values of Y. If in addition we are given $X_{2n} = x_{2n}$ and $X_{2n+2} = x_{2n+2}$ then we may ask for the conditional density of X_{2n+1}. In fact there is some arbitrary aspect to this, depending on how we choose to couple the latent $Y_{2n+1} = i_{2n+1}$ to X_{2n+1}; however it can be chosen not to depend on anything but $X_{2n} = x_{2n}$, $Y_{2n+1} = i_{2n+1}$, and $X_{2n+1} = x_{2n+1}$. Given $X_{2n} = x$, $X_{2n+1} = x'$, one must choose a partition of the interval $[0, 1]$ into subsets $E_1(x, x')$, $E_2(x, x')$, ... such that

$$\int_{E_i(x,x')} p(x, w) p(w, x') \, \mathrm{d}w \quad = \quad \beta(x, i) r_i(x') \, .$$

That this is achievable follows because

$$\int_0^1 p(x, w) p(w, x') \, \mathrm{d}w \quad = \quad p^{(2)}(x, x') \quad = \quad \sum_i \beta(x, i) r_i(x') \, .$$

We may use this choice to define the conditional density of X_{2n+1} in a compatible way, as being proportional as a function of w to

$$p(x_{2n}, w) p(w, x_{2n+2}) \times \mathbb{I}_{[E_{i_{2n+1}}(x_{2n}, x_{2n+2})]}(w) \, .$$

Finally, many Markov chains in practice do not have transition densities, such as for example those which arise in Metropolis-Hastings MCMC. In the Metropolis-Hastings case the failure to have a transition density is rather a trivial matter, assuming that one is working with densities for proposal and acceptance kernels; and if one samples the chain whenever a proposal is accepted then the resulting sub-sampled chain does have a transition density, and Theorem 3.3 applies. It is pleasant to report that the same fix works in essentially every case where one might expect small sets to abound: one simply sub-samples at instances of stopping times such that the resulting chain has a transition density; we sketch the argument here.

Recall, as described for example in [17], that the Hopf decomposition theorem allows us to divide the study of irreducible Markov chains into *dissipative* cases (essentially transient) and *conservative* cases (essentially unions of recurrent classes). The dissipative case is hopeless: for example one can construct skew product Markov chains on $\mathbb{R}^2 \setminus \{(0,0)\}$ whose radial part is the exponential of a Gaussian random walk which drifts off to infinity, and whose angular parts jump so as to be replaced by uniformly random angles but at a rate depending on the radius and decreasing fast enough that there is a positive chance that such a jump may never happen. The chain is irreducible, and yet no matter what stopping time T may be chosen the distribution of X_T places a positive amount of probability on the ray running from $(0,0)$ through X_0.

Suppose on the other hand we consider a conservative chain. General theory (in fact using the existence of general small sets!) tells us we can find a maximal irreducibility measure ψ such that the chain is *Harris-recurrent* off a set N of ψ-measure zero: if $X_0 = x \notin N$ and A is a subset of state-space of positive ψ-measure then $\mathbb{P}\left[\, X \text{ hits } A | X_0 = x \,\right] = 1$. We suppose ψ to be diffuse and delete N from the state-space. Set S_x to be the countable union of ψ-null sets supporting the ψ-singular parts of the distributions of $X_1, X_2, \ldots$ conditional on $X_0 = x$, and define T_x to be the stopping time at which X first leaves S_x. Since $\psi(S_x) = 0$, Harris-recurrence shows that T_x must be finite. A calculation shows that the distribution of X_{T_x} has zero ψ-singular part, so a ψ-density exists for X_{T_x}. We can even show that T_x is essentially minimal for this property! By this means we construct a sub-sampled chain which has measurable ψ-density, for which the results of Theorem 3.3 apply.

References

[1] K.B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society*, 245:493–501, 1978.

[2] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*. Cambridge University Press, Cambridge, 1987.

[3] B. Bollobás. *Modern Graph Theory*. Springer-Verlag, New York, 1998.

[4] L.A. Breyer and G.O. Roberts. Some multi-step coupling constructions for Markov chains. Research report, University of Lancaster, 2000.

[5] L.A. Breyer and G.O. Roberts. Catalytic perfect simulation. *Methodology and computing in applied probability*, To appear, 2001.

[6] W. Doeblin. Sur les propriétés asymptotiques de mouvement régis par certains types de chaînes simples. *Bull. Math. Soc. Roum. Sci.*, 1,2(57-115,3-61), 1937.

[7] J.L. Doob. *Stochastic Processes*. John Wiley & Sons, New York, 1953.

[8] J.H. Friedman and N.I. Fisher. Bump hunting in high-dimensional data. *Statistics and Computing*, 9:123–143, 1999.

[9] Z. Füredi. New asymptotics for bipartite Turán numbers. *Journal of Combinatorial Theory. Series A*, 75:141–144, 1996.

[10] A.P. Godbole, B. Lamorte, and E.J. Sandquist. Threshold functions for the bipartite Turán property. *Electronic Journal of Combinatorics*, 4.1(R18), 1997.

[11] P.J. Green and D.J. Murdoch. Exact sampling for Bayesian inference: towards general purpose algorithms (with discussion). In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 6*, pages 301–321. The Clarendon Press Oxford University Press, 1999. Presented as an invited paper at the 6th Valencia International Meeting on Bayesian Statistics, Alcossebre, Spain, June 1998.

[12] P.R. Halmos. *Measure Theory*. Springer-Verlag, New York, 1974. (First edition, 1950, Litton, New York).

[13] A.H. Hoekstra and F.W. Steutel. Limit theorems for Markov chains of finite rank. *Linear Algebra and Applications*, 60:65–77, 1984.

[14] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, New York, 1993.

[15] D.J Murdoch and P.J. Green. Exact sampling from a continuous state space. *Scand. J. Stat.*, 25:483–502, 1998.

[16] E. Nummelin. A splitting technique for Harris-recurrent chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verve Gebiete*, 43:309–318, 1978.

[17] E. Nummelin. *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, Cambridge, 1984.

[18] S. Orey. *Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities*. Van Nostrand Rienhold, London, 1971.

[19] C.P. Robert. *Discretization and MCMC Convergence Assessment*. Lecture Notes 135. Springer-Verlag, New York, 1998.

[20] G.O. Roberts and J.S. Rosenthal. Quantitative bounds for convergence rates of continuous time Markov processes. *Electronic Journal of Probability*, 1:1–21, 1996. Paper 1.

[21] G.O. Roberts and J.S. Rosenthal. Small and pseudo-small sets for Markov chains. *Stochastic Models*, 17(2):to appear, 2000. Available at URL `ftp://markov.utstat.toronto.edu/jeff/psmall.ps.Z`.

[22] J.S. Rosenthal. Convergence of pseudo-finite Markov chains. Unpublished manuscript available at `http://markov.utstat.toronto.edu/jeff/research.html`, 1992.

[23] J.S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of American Statistical Association*, 90:558–566, 1995.

[24] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 1966.

[25] J.Th. Runnenburg and F.W. Steutel. On Markov chains the transition function of which is a finite sum of products of one variable. *Annals of Mathematical Statistics*, 33:1483–1484, 1962.

Other University of Warwick Department of Statistics
Research Reports authored or co–authored by W.S. Kendall.

161: The Euclidean diffusion of shape.
162: Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence.
172: A spatial Markov property for nearest–neighbour Markov point processes.
181: Convexity and the hemisphere.
202: A remark on the proof of Itô's formula for C^2 functions of continuous semimartingales.
203: Computer algebra and stochastic calculus.
212: Convex geometry and nonconfluent Γ-martingales I: Tightness and strict convexity.
213: The Propeller: a counterexample to a conjectured criterion for the existence of certain convex functions.
214: Convex Geometry and nonconfluent Γ-martingales II: Well–posedness and Γ-martingale convergence.
216: *(with E. Hsu)* Limiting angle of Brownian motion in certain two–dimensional Cartan–Hadamard manifolds.
217: Symbolic Itô calculus: an introduction.
218: *(with H. Huang)* Correction note to "Martingales on manifolds and harmonic maps."
222: *(with O.E. Barndorff-Nielsen and P.E. Jupp)* Stochastic calculus, statistical asymptotics, Taylor strings and phyla.
223: Symbolic Itô calculus: an overview.
231: The radial part of a Γ-martingale and a non-implosion theorem.
236: Computer algebra in probability and statistics.
237: Computer algebra and yoke geometry I: When is an expression a tensor?
238: *Itovsn3*: doing stochastic calculus with *Mathematica*.
239: On the empty cells of Poisson histograms.
244: *(with M. Cranston and P. March)* The radial part of Brownian motion II: Its life and times on the cut locus.
247: Brownian motion and computer algebra (Text of talk to BAAS Science Festival '92, Southampton Wednesday 26 August 1992, with screenshots of illustrative animations).
257: Brownian motion and partial differential equations: from the heat equation to harmonic maps (Special invited lecture, 49th session of the ISI, Firenze).
260: Probability, convexity, and harmonic maps II: Smoothness via probabilistic gradient inequalities.
261: *(with G. Ben Arous and M. Cranston)* Coupling constructions for hypoelliptic diffusions: Two examples.
280: *(with M. Cranston and Yu. Kifer)* Gromov's hyperbolicity and Picard's little theorem for harmonic maps.
292: Perfect Simulation for the Area-Interaction Point Process.
293: *(with A.J. Baddeley and M.N.M. van Lieshout)* Quermass-interaction processes.
295: On some weighted Boolean models.
296: A diffusion model for Bookstein triangle shape.
301: COMPUTER ALGEBRA: an encyclopaedia article.
308: Perfect Simulation for Spatial Point Processes.
319: Geometry, statistics, and shape.

321: From Stochastic Parallel Transport to Harmonic Maps.
323: *(with E. Thönnes)* Perfect Simulation in Stochastic Geometry.
325: *(with J.M. Corcuera)* Riemannian barycentres and geodesic convexity.
327: Symbolic Itô calculus in *AXIOM*: an ongoing story.
328: *Itovsn3* in *AXIOM*: modules, algebras and stochastic differentials.
331: *(with K. Burdzy)* Efficient Markovian couplings: examples and counterexamples.
333: Stochastic calculus in *Mathematica*: software and examples.
341: Stationary countable dense random sets.
347: *(with J. Møller)* Perfect Metropolis-Hastings simulation of locally stable point processes.
348: *(with J. Møller)* Perfect implementation of a Metropolis-Hastings simulation of Markov point processes
349: *(with Y. Cai)* Perfect simulation for correlated Poisson random variables conditioned to be positive.
350: *(with Y. Cai)* Perfect implementation of simulation for conditioned Boolean Model via correlated Poisson random variables.
353: *(with C.J. Price)* Zeros of Brownian Polynomials.
371: *(with G. Montana)* Small sets and Markov transition densities.

Also see the following related preprints

317: E. Thönnes: Perfect Simulation of some point processes for the impatient user.
334: M.N.M. van Lieshout and E. Thönnes: A Comparative Study on the Power of van Lieshout and Baddeley's J-function.
359: E. Thönnes: A Primer on Perfect Simulation.
366: J. Lund and E. Thönnes: Perfect Simulation for point processes given noisy observations.

If you want copies of any of these reports then please email your requests to the secretary using `statistics@warwick.ac.uk` (mail address: the Department of Statistics, University of Warwick, Coventry CV4 7AL, UK).