EXTREMAL INDICES, GEOMETRIC ERGODICITY OF MARKOV CHAINS, AND MCMC

by

Gareth O. Roberts Lancaster University

and

Jeffrey S. Rosenthal
University of Toronto

and

Johan Segers
Université catholique de Louvain

and

Bruno Sousa
Universidade do Minho

Technical Report No. 0611 September 1, 2006

TECHNICAL REPORT SERIES

University of Toronto
Department of Statistics

EXTREMAL INDICES, GEOMETRIC ERGODICITY OF MARKOV CHAINS, AND MCMC

Gareth O. Roberts*, Jeffrey S. Rosenthal†, Johan Segers‡, and Bruno Sousa§

Lancaster University, University of Toronto, Université catholique de Louvain, and Universidade do Minho

Abstract. We investigate the connections between extremal indices on the one hand and stability of Markov chains on the other hand. Both theories relate to the tail behaviour of stochastic processes, and we find a close link between the extremal index and geometric ergodicity. Our results are illustrated throughout with examples from simple MCMC chains.

Keywords. Extremal Index; Geometric Ergodocity; Markov Chains; MCMC.

1 Introduction

It is well-established that maxima of stationary processes with given fixed stationary distributions are affected by the dependence structure of the process. This dependence is effectively captured by the *extremal index*. However little is known about how the extremal index is related to more general mixing properties of stationary processes. The aim of this paper is to relate the extremal index to the concept of *geometric ergodicity* of Markov chains.

The extremal index, written as θ, takes values in $[0, 1]$, and can be interpreted as an indicator of extremal dependence, with $\theta = 1$ indicating asymptotic independence of extreme events. On the other hand $\theta = 0$ represents the case where we can expect strong clusterings of extreme events. In this case it is natural to expect excursions away from 'moderate' values to extreme regions to persist for random times which have heavy-tailed distributions. For Markov chains this behaviour is characteristic of non-geometrically ergodic Markov chains. Thus a natural question to ask is whether $\theta = 0$ is related to non-geometric ergodicity. In this paper we shall see that under certain extra conditions, the two conditions are equivalent.

A motivation for this work comes from MCMC, where commonly used algorithms (such as the Random Walk Metropolis method (RWM) on which

This version: April 15, 2008.

*Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, England; g.o.robert@lancaster.ac.uk

†Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3; jeff@math.toronto.edu; supported in part by NSERC of Canada.

‡Institut de statistique, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium; segers@stat.ucl.ac.be

§Departamento de Matemática para a Ciência e Tecnologia, Universidade do Minho, 4800-058 Guimarães, Portugal; bruno@mct.uminho.pt

we shall focus here) successfully identify modal regions of the target density but have a tendency to underestimate measures of its variation (perhaps its variance). For examples of this phenomenon, see Fearnhead and Meligotsidou (2004).

A natural explanation for this phenomenon comes from the following property of the extremal index. Suppose that a stationary sequence $\{X_n\}$ has extremal index θ. Suppose that u_n is a sequence of levels such that $\Pr(X_n > u_n)$ is asymptotically equivalent to $1/n$, and let T_n be the first hitting time of (u_n, ∞) by the process. Then $\lim_{n\to\infty} \Pr(T_n/n > x) = \mathrm{e}^{-\theta x}$ for positive x, so that the asymptotic expectation of T_n/n is equal to $1/\theta$ if $0 < \theta \leq 1$ and $\lim_{n\to\infty} \mathrm{E}(T_n/n) = \infty$ if $\theta = 0$.

In particular, the smaller the extremal index, the longer it will take before extreme levels are reached in comparison to independent sequences from the target density. Even for the geometrically ergodic RWM chains of Theorem 5.1, it will take more than twice as long as for an independent sequence to arrive at extreme levels. Moreover, if the extremal index is zero, it will even take an order of magnitude longer. Thus we believe the results of this paper go some way towards explaining empirically observed phenomena in MCMC output.

The outline of the paper is as follows. Section 2 contains a few preliminaries concerning extremes of stationary processes and Markov processes. In section 3 it is shown that Markov chains produced by the random walk Metropolis algorithm have extremal index equal to zero as soon as the stationary distribution has a long tail. Geometrically ergodic Markov chains, on the other hand, are shown in section 4 to have a positive extremal index as soon as the drift function satisfies a readily verifiable condition. These results are specialized to the random walk Metropolis algorithm in section 5 where explicit expressions for the extremal index are derived.

2 Preliminaries

Since this paper brings together two theories which have hitherto developed rather separately, we begin with a brief synopsis of the main concepts we shall require from both Markov chain theory and extremal indices.

Extremes of stationary processes. Let $\{X_n\}$ be a (strictly) stationary sequence of random variables, that is, a sequence such that for any integers $i \leq j$ the law of the random vector $(X_{i+k}, \dots, X_{j+k})$ does not depend on the integer k. Let F be the stationary marginal distribution function of the sequence, that is, $F(x) = \Pr(X \leq x)$. Further, define $M_n = \max(X_1, \dots, X_n)$. We are interested in the distribution of extremes of the process; see Leadbetter *et al.* (1983) for a detailed account of the theory of extremes of stationary processes.

The process $\{X_n\}$ has *extremal index* θ if for every $0 < \tau < \infty$ there exists a real sequence $\{u_n(\tau)\}$ such that

$$\lim_{n\to\infty} n\{1 - F(u_n(\tau))\} = \tau, \tag{2.1}$$

$$\lim_{n\to\infty} \Pr[M_n \le u_n(\tau)] = \mathrm{e}^{-\tau\theta}; \tag{2.2}$$

see Leadbetter (1983). A sufficient condition (but not a necessary one) for (2.1) is that F is continuous in a neighbourhood of its right end-point. Hence, the crux of the definition lies in (2.2).

If the random variables X_n are independent, then from elementary calculus

$$\Pr[M_n \le u_n(\tau)] = \left(1 - \frac{1}{n} n\{1 - F(u_n(\tau))\}\right)^n \to \mathrm{e}^{-\tau}, \qquad n \to \infty,$$

whence the extremal index is $\theta = 1$. Moreover, if $\{X_n\}$ has extremal index θ, then, since

$$\Pr[M_n > u_n(\tau)] \le n\{1 - F(u_n(\tau))\},$$

we have $1 - \mathrm{e}^{-\tau\theta} \le \tau$ for every τ, whence $0 \le \theta \le 1$.

Let $\{u_n\}$ be a real sequence. For integer $1 \le l < n$ define

$$\alpha_{n,l} = \max_{I,J} |\Pr[M(I) \le u, M(J) \le u] - \Pr[M(I) \le u] \Pr[M(J) \le u]|$$

where the maximum ranges over all non-empty subsets I and J of $\{1, \ldots, n\}$ with $\max I + l \le \max J$ and where $M(I) = \max\{X_i : i \in I\}$. Then *Leadbetter's condition* $D(u_n)$ is said to hold if there exists an integer sequence $l_n = o(n)$ such that $\alpha_{n,l_n} \to 0$ as $n \to \infty$. Leadbetter's condition is rather weak in that is implied by various mixing conditions on the sequence $\{X_n\}$; see also section 4.

Let u_n be such that $n\{1 - F(u_n)\}$ converges to some $0 < \tau < \infty$ as $n \to \infty$. If condition $D(u_n)$ is satisfied, then by the block-clipping technique (Loynes, 1965; Leadbetter, 1974),

$$\Pr[M_n \le u_n] = (\Pr[M_{r_n} \le u_n])^{n/r_n} + o(1), \qquad n \to \infty, \tag{2.3}$$

for integer sequences r_n such that $l_n = o(r_n)$ and $r_n = o(n)$. This property implies a generalization of the classical extremal types theorem for independent stationary sequences (Fisher and Tippett, 1928; Gnedenko, 1943): if the stationary sequence $\{X_n\}$ has extremal index θ, if Leadbetter's condition $D(u_n(\tau))$ holds for every $0 < \tau < \infty$, and if there exist real sequences $a_n > 0$ and b_n such that $a_n^{-1}(M_n - b_n)$ converges weakly to a non-degenerate law, then the limit law must be an extreme-value distribution (Leadbetter, 1983, Theorem 2.1).

Moreover, under Leadbetter's condition, the extremal index admits a number of intuitively appealing interpretations. Let u_n and r_n be as in (2.3) and assume $D(u_n)$ is satisfied. A comparison of (2.3) with the definition of the extremal index learns that

$$\mathrm{E}\left[\sum_{i=1}^{r_n} \mathbf{1}(X_i > u_n) \,\middle|\, \sum_{i=1}^{r_n} \mathbf{1}(X_i > u_n) > 0\right] = \frac{r_n\{1 - F(u_n)\}}{\Pr[M_{r_n} > u_n]} \to \frac{1}{\theta}$$

as $n \to \infty$, that is, the extremal index is the reciprocal of the limit of the expected size of clusters of exceedances over a high threshold. Moreover, by O'Brien (1987, Theorem 2.1),

$$\lim_{n\to\infty} \Pr[M_{r_n} \le u_n \mid X_0 > u_n] = \theta,$$

stating that the extremal index is the limiting probability that a high-threshold exceedance is followed by a run of non-exceedances. For characterizations of the extremal index in terms of point processes or inter-arrival times between exceedances, see Hsing *et al.* (1988), Ferro and Segers (2003), and Segers (2006). Extensions to higher dimensions are studied in Nandagopalan (1994), Perfekt (1997) and Segers (2006).

The extremal index of stationary Markov chains has been studied in Rootzén (1988), Smith (1992), Perfekt (1994) and Yun (1998). The crucial observation is that at extreme levels, Markov chains typically behave as a random walk or a monotone transformation thereof. The extremal index can be expressed in terms of this random walk. We will encounter this situation in section 5 for Markov chains arising from the random walk Metropolis algorithm.

Markov chains and geometric ergodicity. We shall restrict ourselves in this paper to Markov chains described by a transition kernel P so that $P(x, A) = \Pr(X_1 \in A | X_0 = x)$. Although the concepts we describe in this section are rather general, we shall ultimately apply them in the case where the Markov chain state space is a subset of $\mathbb{R}$. We refer to Meyn and Tweedie (1993) for details.

We shall assume that the chain is φ-irreducible, so that there exists a non-trivial measure φ such that for any Borel set A with $\varphi(A) > 0$,

$$\Pr(X_n \in A \text{ for some } n \mid X_0 = x) > 0, \qquad x \in \mathbb{R}.$$

We also assume aperiodicity: there does *not* exist a partition of the real line into d subsets, $D_0, D_1, \ldots, D_{d-1}$ ($d \ge 2$) with $\Pr(X_n \in D_i) = \mathbf{1}(i = n \bmod d)$.

An important concept for stability is the notion of a small set. A set C is *small* if there exists a positive integer n_0, and $\epsilon > 0$, and some probability measure ν concentrated on C, such that

$$\Pr(X_{n_0} \in \cdot \mid X_0 = x) \ge \epsilon\, \nu(\cdot), \qquad x \in C.$$

To make any sense out of stability, we shall require X to be positive recurrent. Thus we need that for some (and hence for all) small set C with $\varphi(C) > 0$ the expected return time to C from φ-almost all x is finite. Together with φ-irreducibity and aperiodicity, positive recurrence is sufficient to ensure the existence of a unique invariant and limiting probability measure, which we shall denote by π. However we shall largely consider a much stronger form of ergodicity: there exists a function $V \geq 1$, finite π-almost everywhere, and a small set C such that for some constants $\lambda < 1$ and b,

$$PV(x) := \int V(y)P(x, \mathrm{d}y) \leq \lambda V(x) + b\mathbf{1}(x \in C), \qquad x \in \mathbb{R}. \tag{2.4}$$

This geometric drift condition implies that the chain is *geometrically ergodic* (Mengersen and Tweedie, 1996, Theorem 1.4):

$$\|P^n(x, \cdot\,) - \pi\|_V \leq V(x)R\rho^n \tag{2.5}$$

for all positive integer n, all x, and some constants $0 < R < \infty$ and $0 < \rho < 1$, where $\|\mu\|_V = \sup_{|g|\leq V} |\int g(y)\mu(\mathrm{d}y)|$.

Random walk Metropolis algorithm. A famous class of Markov chains are those which are produced by the random walk Metropolis algorithm (Metropolis, 1953). Let π and q be probability densities on the real line, with q symmetric about zero. Let $\{X_n\}$ be the stationary Markov chain arising from the random walk Metropolis algorithm with stationary density π and with increments generated according to the proposal density q: given X_n, a random increment Z_{n+1} is drawn according to the density q; then X_{n+1} is equal to $Y_{n+1} = X_n + Z_{n+1}$ with probability $\alpha(X_n, Y_{n+1})$ and equal to X_n otherwise, where

$$\alpha(x, y) = \begin{cases} \min\{\pi(y)/\pi(x), 1\} & \text{if } \pi(x) > 0, \\ 1 & \text{if } \pi(x) = 0. \end{cases}$$

Formally, we can construct the chain starting at time 0 through the recursive equation

$$X_{i+1} = X_i + Z_{i+1}\mathbf{1}\{U_{i+1} \leq \alpha(X_i, X_i + Z_{i+1})\} \tag{2.6}$$

for integer $i \geq 0$; here $\{Z_i\}$ and $\{U_i\}$ are independent sequences of independent, identically distributed random variables, independent of X_0, and such that the probability density function of Z_i is q and the distribution of U_i is uniform on the interval $(0, 1)$. The process $\{X_n\}$ thus defined forms a Markov chain. Under mild regularity conditions, it is φ-irreducible, aperiodic and has stationarity density π (Roberts and Smith, 1997).

3 Metropolis chains with extremal index zero

Let $\{X_n\}$ be the stationary Markov chain (2.6) produced by the random walk Metropolis algorithm with stationary density π and proposal density q. Let F be the cumulative distribution function corresponding to π, so $F(x) = \int_{-\infty}^{x} \pi(z)dz = \Pr[X_n \le x]$.

Our first result states that if F has a long tail, the extremal index of the chain is equal to zero. The assumption is that the stationary distribution is such that the excess $X_0 - u$ conditionally on $X_0 > u$ converges to infinity as $u \to \infty$. Indeed, under this assumption, if $X_0 > u$ for some large u, it will take the algorithm a very large number of steps to bridge the gap $X_0 - u$ and to return to the region $(-\infty, u]$ again.

THEOREM 3.1. *Assume that the right end-point of F is infinity and that there exists $m > 0$ such that*

$$\lim_{u\to\infty} \frac{1 - F(u+m)}{1 - F(u)} = 1. \tag{3.1}$$

Then for every sequence real sequence $\{u_n\}$ such that $\limsup_{n\to\infty} n\{1 - F(u_n)\} < \infty$ we have

$$\lim_{n\to\infty} \Pr[M_n \le u_n] = 1.$$

In particular, the extremal index of the chain exists and is equal to zero.

Proof. Let $\varepsilon > 0$. Denote $\limsup_{n\to\infty} n\{1 - F(u_n)\} = C < \infty$. There exists $x > 0$ such that $\int_{-\infty}^{-x} q \le \varepsilon / C$. We have

$$\begin{aligned}
\Pr[M_n > u_n] &= \Pr[X_n > u_n] + \sum_{i=1}^{n-1} \Pr[X_i > u_n, \max(X_{i+1}, \ldots, X_n) \le u_n] \\
&\le \Pr[X_1 > u_n] + (n-1)\Pr[X_1 > u_n, X_2 \le u_n] \\
&\le \Pr[X_1 > u_n] + n\Pr[u_n < X_1 \le u_n + x] \\
&\quad + n\Pr[X_1 > u_n + x, X_2 \le u_n].
\end{aligned}$$

Equation (3.1) clearly implies that

$$\lim_{u\to\infty} \frac{1 - F(u+y)}{1 - F(u)} = 1$$

for every real y, whence

$$\begin{aligned}
&n\Pr[u_n < X_1 \le u_n + x] \\
&\quad = n\{1 - F(u_n)\}\left(1 - \frac{1 - F(u_n + x)}{1 - F(u_n)}\right) \to 0, \qquad n \to \infty.
\end{aligned}$$

Further, if, conditionally on $X_1 > u_n + x$, we have $X_2 \leq u_n$, then the proposed increment $X_2 - X_1$ must have been smaller than $-x$, so

$$\Pr[X_2 \leq u_n \mid X_1 > u_n + x] \leq \int_{-\infty}^{-x} q(z)\mathrm{d}z \leq \varepsilon/C.$$

Hence

$$\limsup_{n\to\infty} \Pr[M_n > u_n] \leq \varepsilon.$$

Since ε was arbitrary, we conclude that $\lim_{n\to\infty} \Pr[M_n > u_n] = 0$. □

Observe that condition (3.1) for a single positive m implies the same condition for all positive m. An interpretation is that, conditionally on $X > u$, the excess $X - u$ converges in probability to infinity as $u \to \infty$:

$$\lim_{u\to\infty} \Pr[X - u > m \mid X > u] = 1, \qquad m \geq 0.$$

In the following lemma, some simple sufficient conditions for condition (3.1), reproduced as condition (iii) in the lemma, are given in terms of the stationary density π.

Lemma 3.2. *Let π be a probability density on the real line such that $\pi(x) > 0$ for all sufficiently large x. Consider the following three conditions:*

(i) the function $\log \pi$ is absolutely continuous in a neighbourhood of infinity and $(\log \pi)'(x) \to 0$ as $x \to \infty$;

(ii) $\pi(u + m)/\pi(u) \to 1$ as $u \to \infty$ for every $m \geq 0$;

(iii) $\int_{u+m}^{\infty} \pi(z)dz / \int_u^{\infty} \pi(z)dz \to 1$ as $u \to \infty$ for every $m \geq 0$.

Then (i) implies (ii), and (ii) implies (iii), which is just (3.1) when the density of F is π.

Proof. (i) implies (ii). For $\varepsilon > 0$ we can find u_ε such that $|(\log \pi)'(u)| \leq \varepsilon$ for all $u \geq u_\varepsilon$. For such u and for $m \geq 0$,

$$\exp(-m\varepsilon) \leq \pi(u + m)/\pi(u) \leq \exp(m\varepsilon).$$

Since $\varepsilon > 0$ was arbitrary, we conclude that $\lim_{u\to\infty} \pi(u + m)/\pi(u) = 1$.

(ii) implies (iii). Fix $m > 0$ and $\varepsilon > 0$. By (ii), there exists u_0, depending on m and ε, such that $|\pi(u + m)/\pi(u) - 1| \leq \varepsilon$ for all $u \geq u_0$. For such u, we have

$$\int_{u+m}^{\infty} \pi(x)\mathrm{d}x = \int_u^{\infty} \frac{\pi(x + m)}{\pi(x)}\pi(x)\mathrm{d}x \leq (1 + \varepsilon) \int_u^{\infty} \pi(x)\mathrm{d}x$$

and, similarly, $\int_{u+m}^{\infty} \pi \geq (1 - \varepsilon) \int_u^{\infty} \pi$. Since ε was arbitrary, we arrive at (iii). □

Condition (ii) in Lemma 3.2 is satisfied if π behaves asymptotically as a power function, that is, if $\pi(x) \sim cx^{-\tau}$ for some $c > 0$ and $\tau > 0$. This includes for instance the Student t and the Pareto distributions. Condition (ii) in Lemma 3.2 is also satisfied if π behaves asymptotically as a Weibull density with a tail which is longer than the one of the Exponential distribution, that is, if $\pi(x) \sim c_1 \exp(-c_2 x^\beta)$ for $c_1 > 0$, $c_2 > 0$, and $0 < \beta < 1$.

In Mengersen and Tweedie (1996, Theorem 3.4), stationary densities π are considered for which π is positive and $\log \pi$ is absolutely continuous on a neighbourhood of infinity and for which the limit

$$\lim_{x \to \infty} (\log \pi)'(x) = -\eta \tag{3.2}$$

exists in $[-\infty, 0]$. Under some extra conditions, the Markov chain arising from the random walk Metropolis algorithm is found to be geometrically ergodic if and only if $\eta > 0$. By Theorem 3.1 and Lemma 3.2, in the complementary case, $\eta = 0$, the extremal index of the Markov chain is zero. A natural question is then whether the converse also holds: if (3.2) holds with $\eta > 0$, then is it true that the extremal index of the chain is positive? We come back to this question in section 5.

4 Geometrically ergodic Markov chains

In this section we consider a general Markov chain which is geometrically ergodic in the sense of equation (2.5); see Mengersen and Tweedie (1996, Theorem 1.4). We will show that under a mild condition on the drift function, V, the extremal index of the chain, provided it exists, must be positive. It remains an open question whether this extra condition on the drift function is also necessary for the extremal index to be positive.

Let $\{X_n\}$ be a stationary Markov chain with transition kernel P and stationary distribution π. Let x_+ be the right end-point of the stationary distribution, that is, $x_+ = \sup\{x : \pi(x, \infty) > 0\}$, and assume that $\pi(\{x_+\}) = 0$.

THEOREM 4.1. *Assume that $\{X_n\}$ is a stationary Markov chain, geometrically ergodic in the sense of (2.5) with drift function $V \geq 1$. If V is non-decreasing on $[x_0, x_+)$ for some $x_0 < x_+$ and if*

$$\limsup_{u \uparrow x_+} \mathrm{E}\left[\frac{V(X)}{V(u)} \,\middle|\, X > u \right] < \infty, \tag{4.1}$$

then the extremal index of the chain, provided it exists, is positive.

Proof. For $u \in [x_0, x_+)$ and for positive integer k, by equation (2.5),

$$\begin{aligned}
&\Pr[X_k > u \mid X_0 > u] \\
&= \int_u^{x_+} P^k(x, (u, \infty)) \frac{\pi(\mathrm{d}x)}{\pi(u, \infty)} \\
&\le \int_u^{x_+} \int_u^{x_+} \frac{V(y)}{V(x)} P^k(x, \mathrm{d}y) \frac{\pi(\mathrm{d}x)}{\pi(u, \infty)} \\
&\le \int_u^{x_+} \frac{1}{V(x)} \left(RV(x)\rho^k + \int_u^\infty V(y)\pi(\mathrm{d}y) \right) \frac{\pi(\mathrm{d}x)}{\pi(u, \infty)} \\
&\le R\rho^k + \int_u^{x_+} \frac{V(y)}{V(u)} \pi(\mathrm{d}x)
\end{aligned}$$

and thus

$$\Pr[X_k > u \mid X_0 > u] \le R\rho^k + \Pr[X > u]\mathrm{E}\left[\frac{V(X)}{V(u)} \,\middle|\, X > u \right]. \tag{4.2}$$

Denote

$$C := \limsup_{u \uparrow x_+} \mathrm{E}\left[\frac{V(X)}{V(u)} \,\middle|\, X > u \right] < \infty.$$

Let $\{u_n\}$ be a real sequence such that $\lim_{n\to\infty} n \Pr[X > u_n] = (2C)^{-1}$. Let k be a positive integer such that

$$R \sum_{i=k}^\infty \rho^i < \frac{1}{2}.$$

and denote $r_n = \lfloor n/k \rfloor$. We have

$$\begin{aligned}
&\Pr[M_n > u_n] \\
&\quad \ge \Pr\left[\max_{j=1,\dots,r_n} X_{jk} > u_n \right] \\
&\quad = \Pr[X_{r_n k} > u_n] + \sum_{j=1}^{r_n - 1} \Pr\left[X_{jk} > u_n, \max_{i=j+1,\dots,r_n} X_{ik} \le u_n \right] \\
&\quad \ge r_n \Pr[X > u_n] \left(1 - \Pr\left[\max_{i=k,\dots,n} X_i > u_n \mid X_0 > u_n \right] \right).
\end{aligned}$$

By equation (4.2),

$$\begin{aligned}
&\Pr\left[\max_{i=k,\dots,n} X_i > u_n \mid X_0 > u_n \right] \\
&\quad \le \sum_{i=k}^n \Pr[X_i > u_n \mid X_0 > u_n] \\
&\quad \le R \sum_{i=k}^\infty \rho^i + n \Pr[X > u_n]\mathrm{E}\left[\frac{V(X)}{V(u_n)} \,\middle|\, X > u_n \right].
\end{aligned}$$

Our choices for $\{u_n\}$ and k imply

$$\limsup_{n\to\infty} \Pr[\max_{i=k,\dots,n} X_i > u_n \mid X_0 > u_n] < 1.$$

This inequality in combination with the lower bound for $\Pr[M_n > u_n]$ above yields

$$\liminf_{n\to\infty} \Pr[M_n > u_n] > 0.$$

But since $\lim_{n\to\infty} \Pr[M_n > u_n] = 1 - \exp\{(2C)^{-1}\theta\}$, we conclude that $\theta > 0$. □

Note that in Theorem 4.1, Leadbetter's classical condition $D(u_n)$ is not explicitly assumed. However, this assumption is hidden in the assumption of geometric ergodicity, as can be seen as follows. From Meyn and Tweedie (1993, section 16.1.2), it follows that if a Markov chain is geometrically ergodic in the sense of (2.5), then there exists a constant $0 < \tilde{R} < \infty$ such that, with $0 < \rho < 1$ as in equation (2.5),

$$\operatorname{Cov}[f(X_0), g(X_k)] \le \tilde{R}\rho^k$$

for all measurable functions f and g with $|f| \le 1$ and $|g| \le 1$; see in particular Meyn and Tweedie (1993, equation (16.17)). By the Markov property, the inequality in the above display implies

$$\operatorname{Cov}[f(\dots, X_{-1}, X_0), g(X_k, X_{k+1}, \dots)] \le \tilde{R}\rho^k$$

for measurable functions f and g bounded in absolute value by 1. Taking f and g to be indicator functions shows that

$$|\Pr(A \cap B) - \Pr(A)\Pr(B)| \le \tilde{R}\rho^k$$

for all events $A \in \sigma(X_i : i \le 0)$ and $B \in \sigma(X_i : i \ge k)$; see also Doukhan (1994, p. 3). The above inequality in turn clearly implies Leadbetter's condition $D(u_n)$, with much to spare.

Given that the extremal index of a Markov chain is positive, it is of course of interest to find an explicit expression for it. In section 5 we will do this for Markov chains arising from the random walk Metropolis algorithm. The following auxiliary result, showing that the impact of an extreme value cannot last indefinitely long, will be useful there.

Lemma 4.2. *Under the conditions of Theorem 4.1, for all real sequences $\{u_n\}$ and all positive integer sequences $\{r_n\}$ such that $u_n < x_+$, $u_n \to x_+$, $r_n \to \infty$, and $r_n \Pr[X > u_n] \to 0$ as $n \to \infty$,*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \sum_{i=k}^{r_n} \Pr[X_i > u_n \mid X_0 > u_n] = 0. \tag{4.3}$$

Proof. By equation (2.5), we have for all n such that $u_n \geq x_0$,

$$\begin{aligned}
&\sum_{i=k}^{r_n} \Pr[X_i > u_n \mid X_0 > u_n] \\
&\quad \leq \quad R \sum_{i=k}^{\infty} \rho^i + r_n \Pr[X > u_n] \mathrm{E}\left[\frac{V(X)}{V(u_n)} \,\middle|\, X > u_n \right].
\end{aligned}$$

Let first n tend to infinity and then k tend to infinity to arrive at the stated equation. □

Typically, equation (4.3) is imposed as an extra condition (Smith, 1992; Perfekt, 1994) which in applications has to be verified on an *ad hoc* basis. So the merit of Lemma 4.2 is to show that the condition can be reduced to a readily verifiable assumption on the drift function.

5 Metropolis chains with positive extremal index

We come back to the Markov chain $\{X_n\}$ arising from the random walk Metropolis algorithm in equation (2.6). This time, we are interested in the case where the chain is geometrically ergodic and therefore, under the conditions of Theorem 4.1, its extremal index, provided it exists, is positive. In particular, we seek explicit expressions for the extremal index.

For simplicity, we restrict attention to the case where the stationarity density π is ultimately positive, $\log \pi$ is ultimately absolutely continuous, and there exists $0 < \eta \leq \infty$ such that

$$\lim_{x \to \infty} (\log \pi)'(x) = -\eta. \tag{5.1}$$

Note that if η in (5.1) would be equal to zero, then by Theorem 3.1 and Lemma 3.2, the extremal index of the chain would be equal to zero. This is why we force η in (5.1) to be positive.

There are two subcases: $\eta = \infty$ and $0 < \eta < \infty$. We concentrate on these two cases in Theorems 5.1 and 5.3, respectively. Examples of the first case are density functions $\pi(x)$ which are eventually proportional to $\exp(-cx^\beta)$ for $c > 0$ and $\beta > 1$; examples of the second case are density functions $\pi(x)$ which are eventually proportional to $\exp(-\eta x)$.

In the first case, $\eta = \infty$, the stationary distribution is such that the excess $X_i - u$ conditionally on $X_i > u$ converges in probability to zero as $u \to \infty$; see the proof of Theorem 5.1 below. Moreover, the acceptance probability $\alpha(x, y)$ satisfies $\alpha(x, x+z) \to 0$ if $z > 0$ and $\alpha(x, x+z) \to 1$ if $z < 0$ as $x \to \infty$. So for a high threshold u, conditionally on $X_i > u$, there are two equally likely things which can happen: either the proposed increment Z_{i+1} is positive, in which case the increment is rejected with very high probability, and thus $X_{i+1} = X_i$; or the proposed increment Z_{i+1} is negative, in which case the increment is accepted, and thus $X_{i+1} = X_i + Z_{i+1}$, which is then not larger than u_n with very high probability. Since in all cases the new value X_{i+1} is still large, the same reasoning applies for the next step. A more formal analysis then yields the following result.

THEOREM 5.1. *Let $\{X_n\}$ be the stationary Markov chain arising from the random walk Metropolis algorithm with stationary density π. Assume that the chain is geometrically ergodic with drift function V, that this V is non-decreasing on a neighbourhood of infinity, and that* (4.1) *holds. If $\eta = \infty$ in equation* (5.1), *then the extremal index of the chain exists and is equal to $\theta = 1/2$.*

Proof. Let $0 < \tau < \infty$ and let $\{u_n\}$ be a real sequence such that $n \Pr[X > u_n] \to \tau$ as $n \to \infty$; such a sequence always exists since the distribution of X is continuous. For positive integer k, denote $M_k = \max(X_1, \ldots, X_k)$.

By the discussion following Theorem 4.1, there exists a positive, finite constant $\tilde{R}$ such that

$$|\Pr(A \cap B) - \Pr(A)\Pr(B)| \leq \tilde{R}\rho^l$$

for all positive integer l and all events A and B for which there exists an integer j such that $A \in \sigma(X_i : i \leq j)$ and $B \in \sigma(X_i : i \geq j + l)$. This means we can apply Theorem 2.1 in O'Brien (1987), yielding

$$\lim_{n\to\infty} |\Pr[M_n \leq u_n] - \exp\{-n \Pr[X_0 > u_n] \Pr[M_{r_n} \leq u_n \mid X_0 > u_n]\}| = 0$$

for every positive integer sequence $\{r_n\}$ such that $\log n = o(r_n)$ and $r_n = o(n)$ as $n \to \infty$. So we only have to show that

$$\lim_{n\to\infty} \Pr[M_{r_n} \leq u_n \mid X_0 > u_n] = 1/2. \tag{5.2}$$

For positive integer k, obviously

$$\begin{aligned} &|\Pr[M_{r_n} \leq u_n \mid X_0 > u_n] - \Pr[M_k \leq u_n \mid X_0 > u_n]| \\ &\leq \ \Pr[\max(X_{k+1}, \ldots, X_{r_n}) > u_n \mid X_0 > u_n]. \end{aligned}$$

Hence, by Lemma 4.2, a sufficient condition for (5.2) is that

$$\lim_{n\to\infty} \Pr[M_k \leq u_n \mid X_0 > u_n] = 1/2 \tag{5.3}$$

for every positive integer k.

Without loss of generality, we assume the chain is constructed as in equation (2.6).

Fix an arbitrary $0 < \varepsilon < 1$. Let $0 < z_- < z_+ < \infty$ be such that $\Pr[z_- \leq |Z_1| \leq z_+] \geq 1 - \varepsilon$. From (5.1) with $\lambda = -\infty$, there exists u_ε such that $(\log \pi)'(u) \le \log(\varepsilon)/z_-$ for all $u \ge u_\varepsilon$. For $x \ge u_\varepsilon$ and for $z \ge z_-$ we have

$$\frac{\pi(x+z)}{\pi(x)} = \exp\left(\int_x^{x+z} (\log \pi)'(u)\mathrm{d}u\right) \le \exp\{(z/z_-)\log(\varepsilon)\} \le \varepsilon.$$

Hence for $u \ge u_\varepsilon$,

$$\begin{aligned}\Pr[X_0 > u + z_-] = \int_{u+z_-}^\infty \pi(x)\mathrm{d}x &= \int_u^\infty \pi(x+z_-)\mathrm{d}x \\ &\le \int_u^\infty \varepsilon\pi(x)\mathrm{d}x = \varepsilon \Pr[X_0 > u].\end{aligned}$$

Next, let n be large enough such that $u_n \ge u_\varepsilon + kz_+$.

It is an elementary exercise to verify that for arbitrary events A, B and C for which $\Pr(B) > 0$ and $\Pr(C) > 0$,

$$|\Pr(A \mid B) - \Pr(A \mid B \cap C)| \le \Pr(C^c \mid B) \tag{5.4}$$

where C^c denotes the complement of C. Consider the event

$$D = \{z_- \le |Z_i| \le z_+ \text{ and } U_i > \varepsilon \text{ for all } i = 1, \ldots, k\}.$$

By (5.4) applied to $C = D \cap \{X_0 \le u_n + z_-\}$,

$$\begin{aligned}&|\Pr[M_k \le u_n \mid X_0 > u_n] - \Pr[M_k \le u_n \mid \{u_n < X_0 \le u_n + z_-\} \cap D]| \\ &\le \Pr[\{X_0 > u_n + z_-\} \cup D^c \mid X_0 > u_n] \\ &\le \Pr[X_0 > u_n + z_- \mid X_0 > u_n] + \Pr(D^c) \\ &\le (2k+1)\varepsilon.\end{aligned}$$

Conditionally on $\{u_n < X_0 \le u_n + z_-\} \cap D$, we know that $X_i - X_{i-1} \ge -|Z_i| \ge -z_+$ for all $i = 1, \ldots, k$ and thus $X_i \ge u_n - kz_+ \ge u_\varepsilon$ for all $i = 0, \ldots, k$. Conditionally still on the same event, there are at each time $i = 0, \ldots, k$ two possibilities: $z_- \le Z_i \le z_+$ or $-z_+ \le Z_i \le -z_-$, each occurring with probability one half.

In the first case, $z_- \le Z_i \le z_+$, we have $\alpha(X_{i-1}, X_{i-1} + Z_i) = \pi(X_{i-1} + Z_i)/\pi(X_{i-1}) \le \varepsilon < U_i$. Hence, in this case, the proposed increment is rejected, so that $X_i = X_{i-1}$.

In the second case, $-z_+ \le Z_i \le z_i$, we have $\alpha(X_{i-1}, X_{i-1} + Z_i) = 1 \ge U_i$. Hence, in this case, the proposed increment is accepted, so that $X_i = X_{i-1} + Z_i \le X_{i-1} - z_-$.

On the one hand, if at time $i = 1$ the first possibility, $z_- \le Z_i \le z_+$, occurs, then $X_1 = X_0 > u_n$ and thus $M_k > u_n$. On the other hand, if at time $i = 1$ the second possibility, $-z_+ \le Z_i \le -z_-$, occurs, then $X_1 = X_0 + Z_1 \le X_0 - z_- \le u_n$ and also $X_j \le X_1 \le u_n$ for $j = 2, \ldots, k$, whence $M_k \le u_n$.

From these considerations we conclude that

$$\lim_{n\to\infty} \Pr[M_k \le u_n \mid \{u_n < X_0 \le u_n + z_-\} \cap D] = \frac{1}{2}.$$

Since ε was arbitrary, we arrive at (5.3), as was to be shown. □

Next we consider the case $0 < \eta < \infty$ in (5.1). The following lemma shows that for large threshold u, conditionally on $X_0 > u$, the asymptotic distribution of the chain is that of a random walk starting at u plus an exponential random variable with mean $1/\eta$ and with independent, identically distributed increments, the distribution of which is determined by η and the proposal density q. In fact, such kind of random walk behaviour is rather typical for Markov chains; indeed, Lemma 5.2 could also be derived from general results in e.g. Smith (1992), Perfekt (1994), and Yun (1998). However, the present short proof is instructive as well.

LEMMA 5.2. *Let $\{X_n\}$ be the stationary Markov chain in* (2.6) *with stationary density π and proposal density q. If* (5.1) *holds with $0 < \eta < \infty$, then for positive integer k*

$$\mathcal{L}(X_0 - u, X_1 - X_0, \ldots, X_k - X_{k-1} \mid X_0 > u) \to \mathcal{L}(E, A_1, \ldots, A_k)$$

as $u \to \infty$, where E is an exponential random variable with mean $1/\eta$, independent of the independent, identically distributed random variables $A_i = Z_i \mathbf{1}\{U_i \le \exp(-\eta Z_i)\}$.

Proof. For $z \ge 0$ we have $\pi(u+z)/\pi(u) \to \exp(-\eta z)$ and thus also $\Pr[X_0 - u > z \mid X_0 > u] \to \exp(-\eta z)$ as $u \to \infty$, that is,

$$\mathcal{L}(X_0 - u \mid X_0 > u) \to \mathcal{L}(E), \qquad u \to \infty.$$

Fix $0 < \varepsilon < 1$ and a positive integer k. Let u_ε be such that

$$1 - \varepsilon \le -\frac{(\log \pi)'(u)}{\eta} \le 1 + \varepsilon, \qquad u \ge u_\varepsilon.$$

For $x \ge u_\varepsilon$ and $z \ge u_\varepsilon - x$,

$$\exp\{-\eta z(1+\varepsilon)\} \le \frac{\pi(x+z)}{\pi(x)} \le \exp\{-\eta z(1-\varepsilon)\},$$

For real z, define $\alpha(z) = \min\{\exp(-\eta z), 1\}$. The inequalities in the above display imply that for $x \ge u_\varepsilon$ and $z \ge u_\varepsilon - x$,

$$|\alpha(x, x+z) - \alpha(z)| \le \varepsilon.$$

Further, let $z > 0$ be such that $\Pr[Z_i \le -z] \le \varepsilon$. Define the event

$$C = \{Z_i > -z \text{ and } |U_i - \alpha(Z_i)| > \varepsilon \text{ for } i = 1, \ldots, k\}$$

The event C is independent of X_0; moreover, $\Pr(C^c) \leq 3k\varepsilon$. Let $u \geq u_\varepsilon + kz$. On the event $C \cap \{X_0 > u\}$, we have $X_i - X_{i-1} > -z$ and thus $X_i > u - iz \geq u_\varepsilon$ for all $i = 1, \ldots, k$. Hence on the event $C \cap \{X_0 > u\}$ we have $|\alpha(X_{i-1}, X_{i-1} + Z_i) - \alpha(Z_i)| \leq \varepsilon$ and thus

$$\begin{aligned} X_i - X_{i-1} &= Z_i \mathbf{1}\{U_i \leq \alpha(X_{i-1}, X_{i-1} + Z_i)\} \\ &= Z_i \mathbf{1}\{U_i \leq \alpha(Z_i)\} = A_i \end{aligned}$$

for $i = 1, \ldots, k$. Since ε was arbitrary, we arrive at the stated result. □

THEOREM 5.3. *Let $\{X_n\}$ be the stationary Markov chain arising from the random walk Metropolis algorithm with stationary density π. Assume that the chain is geometrically ergodic with drift function V, that this V is non-decreasing on a neighbourhood of infinity, and that* (4.1) *holds. If $0 < \eta < \infty$ in equation* (5.1), *then the extremal index of the chain exists and is equal to*

$$\theta = \Pr\left[\max_{i \geq 1}(A_1 + \cdots + A_i) \leq -E\right] \tag{5.5}$$

with E and $A_1, A_2, \ldots$ as in Lemma 5.2. In particular, $0 < \theta < 1/2$.

Proof. Copying the line of reasoning in the proof of Theorem 5.1 up to equation (5.3), we see that it is sufficient to show that

$$\lim_{k\to\infty} \lim_{u\to\infty} \Pr[M_k \leq u \mid X_0 > u] = \theta. \tag{5.6}$$

But since

$$X_i - u = (X_0 - u) + \sum_{j=1}^{i} (X_j - X_{j-1}),$$

equation (5.6) is an immediate consequence of Lemma 5.2. □

EXAMPLE 5.4. Consider the random walk Metropolis algorithm with $\pi(x) = \mathrm{e}^{-x}$ for $0 < x < \infty$, and with $q(z) = 1/2$ for $-1 < z < 1$. Equation (5.1) is clearly satisfied with $\eta = 1$. Let $0 < \beta < 1$ and put $V(x) = \mathrm{e}^{\beta x}$. For $x > 1$, we have

$$\begin{aligned} PV(x) &= \mathrm{E}[V(X_1) \mid X_0 = x] \\ &= \frac{1}{2}\int_{-1}^{0} \mathrm{e}^{\beta(x+z)} \mathrm{d}z + \frac{1}{2}\int_0^1 \left(\mathrm{e}^{-z}\mathrm{e}^{\beta(x+z)} + (1 - \mathrm{e}^{-z})\mathrm{e}^{\beta x}\right) \mathrm{d}z \\ &= \frac{1}{2}\mathrm{e}^{\beta x}\left(1 + \int_0^1 \left(\mathrm{e}^{-\beta z} + \mathrm{e}^{-(1-\beta)z} - \mathrm{e}^{-z}\right) \mathrm{d}z\right). \end{aligned}$$

Since $a + b - ab < 1$ for $0 < a < 1$ and $0 < b < 1$, we find that there exists a constant $0 < \lambda < 1$ such that $PV(x) = \lambda V(x)$ for $x \geq 1$; that is, the chain satisfies the drift condition (2.4) with $C = [0, 1]$ as the small set. Since $\mathcal{L}(X - u \mid X > u) = \mathcal{L}(X)$ for $u > 0$, we also have

$$\mathrm{E}\left[\frac{V(X)}{V(u)} \,\middle|\, X > u\right] = \mathrm{E}\left[\frac{V(X+u)}{V(u)}\right] = \mathrm{E}[\mathrm{e}^{\beta X}] = \frac{1}{1-\beta}.$$

From Theorem 5.3 we conclude that the chain is geometrically ergodic and that its extremal index exists and is equal to the expression in (5.5); here E is a standard exponential random variable and $\{A_i\}$ is a sequence of independent and identically distributed random variables, independent of E, with $A_i = Z_i \mathbf{1}(U_i \leq \mathrm{e}^{-Z_i})$, where Z_i is uniform on $(-1, 1)$, U_i is uniform on $(0, 1)$, and Z_i and U_i are independent.

Acknowledgments. We thank Jon Tawn and Neal Madras for helpful discussions. We also thank the editor, Holger Rootzén for making a very helpful suggestion regarding an early version of the paper.

References

Doukhan, P. (1994) *Mixing, Properties and Examples*, Lecture Notes in Statistics **85**, Springer, New York.

Fearnhead, P.N. and Meligotsidou, L. (2004) “Exact filtering for partially observed continuous time models”, *Journal of the Royal Statistical Society, Series B*, **66**, 3, 771-789.

Ferro, C. A. T. and Segers, J. (2003) “Inference for Clusters of Extreme Values,” *Journal of the Royal Statistical Society, Series B*, **65**, 545–556.

Fisher, R. A. and Tippett, L. H. C. (1928) “Limiting forms of the frequency distribution of the largest or smallest member of a sample,” *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.

Gnedenko, B. V. (1943) “Sur la distribution limite du terme d’une série aléatoire,” *Annals of Mathematics*, **44**, 423–453.

Hsing, T., Hüsler, J., and Leadbetter, M. R. (1988) “On the exceedance point process for a stationary sequence,” *Probability Theory and Related Fields*, **78**, 97–112.

Loynes, R. M. (1965) “Extreme values in uniformly mixing stationary stochastic processes,” *Annals of Mathematical Statistics*, **36**, 993–999.

Leadbetter, M. R. (1974) “On extreme values in stationary sequences,” *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **28**, 289–303.

Leadbetter, M.R. (1983) “Extremes and Local Dependence in Stationary Sequences,” *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **65**, 291–306.

Leadbetter, M.R., Lindgren, G. and Rootzén (1983) *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

Mengersen, K.L. and Tweedie, R.L. (1996) "Rates of Convergence of the Hastings and Metropolis Algorithms," *The Annals of Statistics* **24**, 101–121.

Metropolis, N. Rosenbluth, A. Rosenbluth, M., Teller, A. and Teller, E. (1953) "Equations of state calculations by fast computing machines," *J. Chemical Physics* **21**, 1087–1091.

Meyn, S.P. and Tweedie, R.L. (1993) *Markov chains and stochastic stability*, Springer-Verlag, London. Available at probability.ca/MT.

Nandagopalan, S.] (1994) "On the multivariate extremal index," *J. Res. Natl. Inst. Stand. Technol.*, **99**, 543–550.

O'Brien, G. L. (1987) "Extreme values for stationary and Markov sequences," *Annals of Probability* **15**, 281–291.

Perfekt, R. (1994) "Extremal behaviour of stationary Markov chains with applications," *Annals of Applied Probability* **4**, 529–548.

Perfekt, R. (1997) "Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$," *Advances of Applied Probability* **29**, 138–164.

Roberts, G.O. and Smith A.F.M. (1997) "Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms," *Stoch. Proc. Appl.* **49**, 207–216.

Rootzén, H. (1988) "Maxima and exceedances of stationary Markov chains," *Advances in Applied Probability*, **20**, 371-390.

Segers, J. (2006) "Rare events, temporal dependence, and the extremal index," *Journal of Applied Probability*, 43(2), to appear.

Smith, R. L. (1992) "The extremal index for a Markov chain," *Journal of Applied Probability* **29**, 37–45.

Yun, S. (1998) "The extremal index of a higher-order stationary Markov chain," *Annals of Applied Probability* **8**, 408–437.