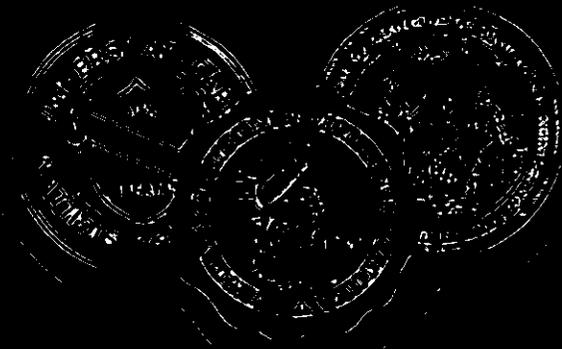


DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA

CHapel Hill, North Carolina



ON AN APPROXIMATION FORMULA FOR THE
EXTREMAL INDEX IN A MARKOV CHAIN

by

Richard L. Smith

October 1992

Mimeo Series #2037

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

The Library of the Department of Statistics
at
North Carolina State University

MIMMO
SERIES# ON AN APPROXIMATION
#2087 FORMULA FOR THE EXTREMAL
INDEX IN A MARKOV CHAIN

NAME Richard L. Smith
DATE

2036 FAN, J., TRUONG, Y. & WANG, Y.: Nonparametric variables, Sept. 1990. *Nonparametric Functional Estimation*, 1991, 618-627.

2037 JI, C.: Sieve estimators for pair-interaction potential fields, Oct. 1990. *Ann. Statist.*, to appear.

2038 CHAKRAVARTI, I.M.: A three-class association scheme and a (PBIB) design defined by the incidence of the Oct. 1990. *Discrete Math.*, to appear.

2039 CHAKRAVARTI, I.M.: Geometric construction of association schemes and codes from non-degenerate Discrete Math., to appear.

2040 CARLSTEIN, E.: Resampling techniques for stationarity, Oct. 1990. *New Directions in Time Series Analysis*, Murray Rosenblatt, (ed.) 1992, Springer-Verlag, IMA, to appear.

2041 FAN, J. & MARRON, S.: Best possible constant for *Statist.*, to appear.

2042 FAN, J., TRUONG, Y.K. & WANG, Y.: Measurement approach, Oct. 1990. (30 pages)

2043 REN, J.J.: On Hadamard differentiability and M-estimation in linear models, Sept. 1990. (dissertation)

2044 SEN, P.K. & SENGUPTA, D.: On characterizations of Pitman closeness of some shrinkage estimators, Dec. 1990. *Comm. Statist., Th & Methods*, 20, 1991, 9451-9450.

2045 TRUONG, Y.K. & FAN, J.: Deconvolution problems in time series, Dec. 1990. (19 pages)

2046 HU, T.-C. & FAN, J.: Bias correction and higher order kernel functions, Dec. 1990. *Statist. & Prob. Ltr.*, 18, 1992, 235-243.

2047 HU, T.-C. & WEBER, N.: On the invariance principle for exchangeable random variables, Jan. 1991. (6 pages)

2048 HU, T.-C.: A study on the law of the iterated logarithm for arrays of random variables, Jan. 1991. *Comm. Statist., Th & Meth.*, A20, 1991, 1989-1994.

2049 FAN, J.: Design-adaptive nonparametric regression, Feb. 1991. *J. Amer. Statist. Assoc.*, 4, 1992, to appear.

2050 SEN, P.K.: Some informative aspects of jackknifing and bootstrapping, Mar. 1991. *Order Statistics and Nonparametrics: Theory and Applications* (eds. I.A. Selamo & P.K. Sen), 1992, North Holland, to appear.

2051 HALL, P., HU, T.-C. & MARRON, S.: Improved variable window kernel estimates of probability densities, Apr. 1991. (12 pages)

2052 FAN, J. & GUBELS, L.: Variable Bandwidth and Local Linear Regression Smoothers, Apr. 1991. *Ann. Statist.*, to appear.

ed squared density derivatives, Apr. 1991. (dissertation)

ds for kernel estimators of the intensity function of a l. (dissertation)

thers in regression function estimation, May 1991.

tion of a bounded squared mean, May 1991. *Statist. &*

problem, May 1991. (17 pages)

tion and statistical estimation, May 1991. *Paperschrift in* 1992, 65-82.

s data, May 1991. (thesis)

design adaptive nonparametric function estimation: A

ssion estimation with errors-in-variables: Asymptotic *J. Multivariate Anal.*, to appear.

of Markov random field texture models, July 1991.

& JOHNSON, N.L.: Use of moments in distribution theory: A *J. Multivariate Anal.*, to appear.

2064 LUND, R.: On the seasonal cycle of hurricane frequency, Sept. 1991. (19 pages)

2065 KO, W.-J.: A sequential clinical trials model for determining the best among three treatments with normal responses, Oct. 1991. (dissertation)

2066 SMITH, R. & WEISSMAN, I.: Estimating the extremal index, Nov. 1991. (30 pages)

2067 SMITH, R.: Some comments on the relation between statistics and chaos, Nov. 1991. *Statist. Sc.*, 7, 1992, 102-113.

2068 SINGPURWALLA, N.D. & SMITH, R.L.: Boris Vladimirovich Gnedenko, Nov. 1991. *Statist. Sc.*, 7, 1992, 273-283.

2069 KRISHNAMOORTHY, G.: Statistical analysis of boundaries: A nonparametric approach, Jan. 1992. (dissertation)

2070 SHERMAN, M. & CARLSTEIN, E.: Method of moments estimation: Simple and complicated settings, Jan. 1992. (11 pages)

2071 GHOSH, J.K., SEN, P.K. & MICKERJEE, R.: Second order Pitman closeness and Pitman admissibility, Feb. 1992. (17 pages)

2072 JOHNSON, N.L., KOTZ, S. & PEARN, W.L.: Flexible process capability indices, Feb. 1992. *J. Quality Tech.*, 24, 216-231.

2073 JURECKOVA, J. & SEN, P.K.: Asymptotic equivalence of regression rank scores estimators and R-estimators in linear models, Feb. 1992. (16 pages)

**ON AN APPROXIMATION FORMULA FOR THE
EXTREMAL INDEX IN A MARKOV CHAIN**

by

RICHARD L. SMITH
University of North Carolina

Contact Address: Richard L. Smith, Department of Statistics, University of North Carolina, Chapel Hill, N.C. 27599-3260, U.S.A.

First version: October 26 1992

Abstract

In a previous paper, Smith and Weissman have considered estimators of the extremal index of a stationary stochastic process, and have shown that the theoretical properties of these estimators depend on an approximation formula for the extremal index. However, the derivation of this formula was entirely heuristic except for one very special case. In this paper, it is argued that the formula is valid for a broad class of processes, namely processes which are defined on a geometrically ergodic Harris Markov chain.

KEYWORDS AND PHRASES: extreme values in stationary processes, geometrically ergodic Markov chains.

1. INTRODUCTION

Let $\{X_n, n \geq 0\}$ denote a stationary Markov chain with marginal distribution function F . For integer $i \geq 2$ and a high threshold value u , define

$$\theta(i, u) = \Pr\{X_2 \leq u, X_3 \leq u, \dots, X_i \leq u | X_1 > u\} \quad (1.1)$$

O'Brien (1987) and Rootzén (1988) showed that for suitably large values of i and u , $\theta(i, u)$ approximates a parameter θ known as the *extremal index* of the process. This parameter characterises many of the important properties of extreme values in stochastic processes (Leadbetter, Lindgren and Rootzén, 1983). Smith (1992) showed that for a wide variety of Markov processes, the calculation of (1.1) reduces to solving a Wiener-Hopf integral equation.

Recently, Smith and Weissman (1992) have considered the problem of statistical estimation of an extremal index. The theoretical solution of this problem requires calculating the bias and variance of some empirical estimators, and they showed that this in turn requires an approximation to $\theta(i, u) - \theta$ for finite i and u . In particular, they proposed the approximation

$$\theta(i, u) - \theta \approx \theta\{F^{i\theta}(u) - 1\} + K\beta^i. \quad (1.2)$$

where K and $\beta \in (0, 1)$ are constants. However, the derivation of (1.2) was only given for one very special case, other cases being argued by heuristic reasoning supported by simulations.

In this paper, it is argued that (1.2) is in fact valid for a wide class of processes, namely whenever X_n is a function of a geometrically ergodic, Harris Markov chain.

As a specific example to which we want to apply this theory, consider the Markov chain whose joint distributions are specified by the formula

$$\Pr\{X_n \leq x, X_{n+1} \leq y\} = \exp\{-(e^{-Rx} + e^{-Ry})^{1/R}\}, \quad R \geq 1. \quad (1.3)$$

As a bivariate distribution function, this seems to be widely applicable in extreme value studies (Tawn 1988). It is therefore also a natural model for dependence in stochastic processes, and has previously been used as an example in the papers of Smith (1992) and Smith and Weissman (1992).

2. THE MAIN RESULT

The general theory of Markov chains in discrete time has been presented by Nummelin (1984). A Markov chain with state space (E, \mathcal{E}) (where \mathcal{E} denotes a class of measurable sets over E) is ψ -irreducible if every set of positive ψ -measure is hit with positive probability regardless of the starting point, and *Harris recurrent* if that probability is always 1; here ψ is a σ -finite measure on (E, \mathcal{E}) . For most of the discussion which follows, we can assume E to be the real line, \mathcal{E} the class of Borel sets, and ψ Lebesgue measure. If $C \in \mathcal{E}$ is such that $\psi(C) > 0$, ν is a probability measure on (E, \mathcal{E}) , and there exist a constant $\xi > 0$ and an integer $m_0 \geq 1$ such that

$$\Pr\{X_{m_0} \in A | X_0 = x\} \geq \xi\nu(A) \quad \text{for all } A \in \mathcal{E}, x \in C \quad (2.1)$$

then C is called a *small set* and ν a *small measure*. These always exist for ψ -irreducible processes; for a model like (1.3), in which the joint density is bounded and positive over finite regions of the plane, it suffices to take C any finite interval, ν the uniform distribution over any finite interval, and $m_0 = 1$. For Harris recurrent processes, C will be visited infinitely often and we can define *regeneration points* as follows: any time the process is in C , toss a coin so that with probability ξ it restarts with distribution ν ; otherwise the next value of the process is distributed according to $[\Pr\{X_{i+1} \in \cdot | X_i = x\} - \xi\nu(\cdot)]/(1 - \xi)$. Denote the regeneration times $0 < S_0 < S_1 < \dots$ and let $B_0 = \{X_i, 0 \leq i < S_0\}$, $B_1 = \{X_i, S_0 \leq i < S_1\}$, etc. Then B_0, B_1, B_2, \dots are independent regeneration cycles and B_1, B_2, \dots are identically distributed. Rootzén (1988), following Asmussen (1987), showed that a similar construction is possible when $m_0 > 1$ in (2.1) but in this case the cycles are only 1-dependent, rather than independent. We do not consider this situation any further here. Let μ denote the mean time between regenerations; when $\mu < \infty$ the process is *positive recurrent* and when $\mu = \infty$ it is *null recurrent*. When a process is positive recurrent then a stationary measure π exists, and if it is also *aperiodic* then $\Pr\{X_n \in A | X_0 = x\} \rightarrow \pi(A)$ as $n \rightarrow \infty$ for all $A \in \mathcal{E}$ and all $x \in E$, exactly as in the discrete-state case. Positive recurrent aperiodic Harris processes are also called *ergodic*.

For the following discussion we need to define a strengthening of the notion of an

ergodic process. A process is *geometrically ergodic* if convergence to π is geometrically fast; one of several equivalent characterisations (Nummelin 1984, Theorem 6.14, page 120) is that there exist a positive constant $\beta < 1$ and a positive function M with $\int_E M(x)\pi(dx) < \infty$ such that

$$|\Pr\{X_n \in A | X_0 = x\} - \pi(A)| \leq M(x)\beta^n \quad \text{for all } x \in E, A \in \mathcal{E}, n \geq 0. \quad (2.2)$$

There is an even stronger property known as *uniform ergodicity* in which the function $M(x)$ in (2.2) is replaced by a single constant M ; however this condition is not satisfied by the process (1.3) or any similar process, so we do not consider it further.

Given that the process is ergodic, a suitable condition for geometric ergodicity (Nummelin, page 90) is that there exist a small set C , a positive measurable function g and constants $k > 1, \gamma < 0, M_1 < \infty$ such that

$$\begin{aligned} \mathbf{E}\{kg(X_{i+1}) - g(X_i) | X_i = x\} &\leq \gamma \quad \text{for all } x \notin C, \\ \mathbf{E}\{g(X_{i+1})I(X_{i+1} \notin C) | X_i = x\} &\leq M_1 \quad \text{for all } x \in C. \end{aligned} \quad (2.3)$$

Here I denotes the indicator function. As an illustration of how this condition may be checked for processes of practical interest, we prove:

Lemma 1. The Markov chain defined by (1.3) is geometrically ergodic.

Proof. It follows quickly from (1.3) that

$$\Pr\{X_{i+1} - X_i \leq z | X_i = x\} = (1 + e^{-Rz})^{1/R-1} \exp[-e^{-x}\{(1 + e^{-Rz})^{1/R} - 1\}]. \quad (2.4)$$

From this we note two forms of limiting behaviour. As $x \rightarrow \infty$, the conditional distribution of $X_{i+1} - X_i$ converges to that of a random variable Z with distribution function $(1 + e^{-Rz})^{1/R-1}$; this random walk representation was the basis of Smith's (1992) method of calculating the extremal index. Since $\mathbf{E}(Z) < 0$ we can find $t > 0$ such that

$E\{e^{tZ}\} < 1$; then we will also have $E\{e^{t(X_{i+1}-X_i)}|X_i = x\} < \delta < 1$ for $x >$ some x^* . For $x \rightarrow -\infty$, write $z = -x/R + w$; then (2.4) converges to $\exp(-e^{-Rw}/R)$, i.e. there is a random variable W such that X_{i+1} behaves like $x(1 - 1/R) + W$ when $X_i = x \ll 0$. Define $C = (-c, c)$ for $c > x^*$, and let $g(x)$ be e^{tx} for $x > 0$ and $|x|$ for $x < 0$. Choose k so that $1 < k < \min\{1/\delta, R/(R - 1)\}$. Then for $x > c$,

$$E\{kg(X_{i+1}) - g(X_i)|X_i = x\} \leq (k\delta - 1)e^{tx}$$

while for $x < -c$

$$E\{kg(X_{i+1}) - g(X_i)|X_i = x\} \approx |x|\{k(1 - 1/R) - 1\} - kE(W)$$

both of which are negative for sufficiently large $|x|$ under the stated conditions on k . This establishes the first condition in (2.3), while the second is obvious. This completes the proof of Lemma 1.

We now return to the problem of the extremal index. Let ζ_0 denote the maximum of the process $\{X_i\}$ over the cycle C_0 , ζ_1 the maximum over C_1 , etc. Then $\zeta_0, \zeta_1, \zeta_2, \dots$, are independent and ζ_1, ζ_2, \dots , identically distributed. Define $G(u) = \Pr\{\zeta_1 \leq u\}^{1/\mu}$; Rootzén (1988) calls G the *phantom distribution function* associated with the Markov chain. Also let $\epsilon = -\log G(u)$; in the following we derive approximations under the assumption that ϵ is small. A key result is the following:

Lemma 2. Assume the Markov chain is stationary and geometrically ergodic. Then

$$\max_{n \geq 1} |\Pr\{\max(X_1, \dots, X_n) \leq u\} - G^n(u)| = O(\sqrt{\epsilon \log 1/\epsilon}).$$

Proof. Let $\nu_n = \min\{k \geq 0 : S_k > n\}$ denote the number of regeneration cycles up to and including time n . Let $\delta > 0$, $\delta' = \delta + 1/n$. Then

$$\begin{aligned} |\Pr\{\max(X_1, \dots, X_n) \leq u\} - G^n(u)| &\leq n\epsilon e^{-n\epsilon} \mu \delta' e^{\mu \delta' \epsilon} \\ &+ \Pr\{|\nu_n/n - 1/\mu| \geq \delta\} + \Pr\{\zeta_0 > u\}. \end{aligned} \tag{2.5}$$

This is similar to Theorem 3.1 of Rootzén (1988) but with two modifications. First, Rootzén's third term (involving ζ_0) is different from ours, but it is clear from Rootzén's proof that the term given here is equally valid. Second, for the first term on the right hand side, Rootzén gives just $\mu\delta$ as a bound on $|G(u)^{n+n\mu\delta} - G(u)^n|$. This can be strengthened by noting that for $0 < q < p < 1$ we have $p^n - q^n \leq n(p - q)p^{n-1}$; let $p = G(u) = e^{-\epsilon}$, $q = G(u)^{1+\mu\delta} = pe^{-\mu\delta\epsilon}$; then $p^n - q^n \leq ne^{-n\epsilon}(1 - e^{-\mu\delta\epsilon}) \leq n\epsilon e^{-n\epsilon}\mu\delta$. There is also a reverse inequality in which we must set $q = G(u)$, $p = G(u)^{1-\mu\delta'}$ and the same argument works but now there appears an additional factor $e^{\mu\delta'\epsilon}$. This establishes (2.5).

From the stationarity of the process it follows that the third term in (2.5) is $O(\epsilon)$ and we shall not consider this any further. Geometric ergodicity implies that the common distribution of $S_{i+1} - S_i$ for $i \geq 0$ has a finite moment generating function in a neighbourhood of the origin. Using this fact, a large deviations argument shows that there exists $K > 0$ such that $\Pr\{|\nu_n/n - 1/\mu| \geq \delta\} \leq e^{-n\delta^2/K}$. Set this equal to $\sqrt{\epsilon \log 1/\epsilon}$. Then it is readily verified that $\delta \sim \sqrt{(K \log 1/\epsilon)/(2n)}$. Since $\epsilon\sqrt{\log 1/\epsilon} \rightarrow 0$ as $\epsilon \rightarrow 0$ we have established that we may ignore the factor $e^{\mu\delta'\epsilon}$ in the first term of (2.5). If we split up $\delta' = \delta + 1/n$, then we may also ignore the term arising from $1/n$ since $\epsilon e^{-n\epsilon} = O(\epsilon)$. Finally we note that

$$n\epsilon e^{-n\epsilon}\delta = O(n^{1/2}\epsilon e^{-n\epsilon}\sqrt{\log 1/\epsilon}).$$

Maximising this with respect to n we get a bound of $O(\sqrt{\epsilon \log 1/\epsilon})$. Since this is the same order as the second term of (2.5), the proof of Lemma 2 is complete.

Now we consider how to evaluate the phantom distribution function. For the Markov chain starting in its stationary distribution, let S_0 denote the first recurrence time after time 0 as before, and let

$$\theta(u) = \Pr\{X_1 \leq u, \dots, X_{S_0-1} \leq u | X_0 > u\}. \quad (2.6)$$

By Rootzén's (1988) Corollary 4.2(i), $\theta(u) \rightarrow \theta$ as $u \rightarrow \infty$. The following argument strengthens this:

Lemma 3. $\Pr\{\zeta_1 > u\} = \mu\theta(u)(1 - F(u))$.

Proof. Considering the cycle B_1 and conditioning on the last time the process crosses u , we have

$$\begin{aligned}
& \Pr\{\zeta_1 > u\} \\
&= \int_{x=u}^{\infty} \sum_{i=0}^{S_1-S_0-1} \Pr\{X_{S_0+i} \in dx, X_{S_0+i+1} \leq u, \dots, X_{S_1-1} \leq u\} \\
&= \int_{x=u}^{\infty} \sum_{i=0}^{S_1-S_0-1} \Pr\{X_{S_0+i} \in dx\} \Pr\{X_{S_0+i+1} \leq u, \dots, X_{S_1-1} \leq u | X_{S_0+i} \in dx\} \\
&= \int_{x=u}^{\infty} \sum_{i=0}^{S_1-S_0-1} \Pr\{X_{S_0+i} \in dx\} \Pr\{X_1 \leq u, \dots, X_{S_0-1} \leq u | X_0 \in dx\} \tag{2.7}
\end{aligned}$$

where in the last line we have used the Markov property. However, by the “expectations over blocks” properties described in Section 5.3 of Nummelin (1984) (cf. equation (2.2) of Rootzén, 1988), and using the fact that F , the marginal distribution of the process, is necessarily also its stationary distribution, we have that

$$\sum_{i=0}^{S_1-S_0-1} \Pr\{X_{S_0+i} \in dx\} = \mu F(dx)$$

and this together with (2.7) establishes the Lemma.

From Lemma 3 it is easily seen that $|G(u) - F^{\theta(u)}(u)| = O(\epsilon^2)$ and hence that $|G^n(u) - F^{n\theta(u)}(u)| = O(\epsilon)$ uniformly over $n > 0$. Combined with Lemma 2, this shows that

$$\max_{n \geq 1} |\Pr\{\max(X_1, \dots, X_n) \leq u\} - F^{n\theta(u)}(u)| = O(\sqrt{\epsilon \log 1/\epsilon}). \tag{2.8}$$

This provides some argument for using $\theta(u)$ instead of θ in the approximation for the distribution of the maximum: if we use $\theta(u)$, then a bound on the error of approximation is given by (2.8) whereas there is, in general, no bound on the rate at which $\theta(u)$ must approach θ .

Finally we come to the question of how well $\theta(i, u)$ for fixed $i > 0$ approximates $\theta(u)$.

We have

$$\begin{aligned}
& \theta(u) - \theta(i, u) \\
&= \sum_{r=0}^{\infty} [\Pr\{X_1 \leq u, \dots, X_r \leq u, S_0 = r + 1 | X_0 > u\} \\
&\quad - \Pr\{X_1 \leq u, \dots, X_{i-1} \leq u, S_0 = r + 1 | X_0 > u\}] \\
&= \sum_{r=0}^{i-2} \Pr\{X_1 \leq u, \dots, X_r \leq u, \max_{r < j < i} X_j > u, S_0 = r + 1 | X_0 > u\} \\
&\quad - \sum_{r=i}^{\infty} \Pr\{X_1 \leq u, \dots, X_{i-1} \leq u, \max_{i \leq j \leq r} X_j > u, S_0 = r + 1 | X_0 > u\}. \tag{2.9}
\end{aligned}$$

The second term is bounded by $\Pr\{S_0 \geq i | X_0 > u\}$ and this, by geometric ergodicity, is at most $M\beta^i$ for some $\beta < 1$, where M will in general depend on u . The first term may be written in the form

$$\begin{aligned}
& \sum_{r=0}^{i-2} \Pr\{X_1 \leq u, \dots, X_r \leq u, S_0 = r + 1 | X_0 > u\} \{1 - F^{(i-1-r)\theta(u)}(u) + O(\sqrt{\epsilon \log(1/\epsilon)})\} \\
&= \left[\sum_{r=0}^{\infty} \Pr\{X_1 \leq u, \dots, X_r \leq u, S_0 = r + 1 | X_0 > u\} + O(\beta^i) \right] \\
&\quad \cdot \{1 - F^{i\theta(u)}(u) + O(\sqrt{\epsilon \log(1/\epsilon)})\}.
\end{aligned}$$

Thus we have established our main result:

Theorem. Suppose the chain is geometrically ergodic with geometric constant $\beta < 1$ and suppose $1 - F(u) = O(\epsilon)$. Then there is a constant M (in general, depending on u) such that

$$|\theta(u)F^{i\theta(u)}(u) - \theta(i, u)| \leq M\beta^i + O(\sqrt{\epsilon \log 1/\epsilon}). \tag{2.10}$$

3. DISCUSSION

The theorem does not establish that there is a specific constant K for which (1.2) holds. Nevertheless, having established the existence of a $O(\beta^i)$ bound, it seems natural to conjecture that there will be such a constant (depending on u , of course). With this modification, what (2.10) says is that (1.2) is valid as an approximation uniform over i for each u , the error being bounded by a term of $O(\sqrt{\epsilon \log 1/\epsilon})$.

Another feature of this result is that it gives some theoretical basis for considering the extremal index as a parameter depending on the threshold, in other words $\theta(u)$ instead of its limiting value θ . The practical need for this was suggested in a discussion contribution by Tawn (1990), and has been reinforced in subsequent studies including the data example of Smith and Weissman (1992). Of course, if $\theta(u)$ tends to θ fast enough, then we may substitute θ for $\theta(u)$ without affecting the overall error bound in (2.10). However, it is conceivable that the convergence of $\theta(u)$ to θ will be slow and that there is then a real gain from not making the substitution.

A final remark is that there is nothing in our derivation to require that the process $\{X_n\}$ itself be a Markov chain; it could equally well be one component of a multidimensional Markov chain, provided the assumed recurrence and geometric ergodicity properties hold. Indeed the example in Section 3 of Smith and Weissman (1992) is of this form. Other examples within this framework include k -dependent Markov chains (including general nonlinear autoregressive processes) and nonlinear moving average processes of finite extent (i.e. any process that can be represented in the form $X_n = f(Z_n, Z_{n-1}, \dots, Z_{n-k+1})$ with $\{Z_i\}$ independent) provided the resulting process is a geometrically ergodic Harris chain.

ACKNOWLEDGEMENT

This work was partially supported by the National Science Foundation.

REFERENCES

Asmussen, S. (1987), *Applied Probability and Queues*. Wiley, New York.

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

Nummelin, E. (1984), *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press.

O'Brien, G.L. (1987), Extreme values for stationary and Markov sequences. *Ann. Probab.* **15**, 281-291.

Rootzén, H. (1988), Maxima and exceedances of stationary Markov chains. *Adv. Appl. Probab.* **20**, 371-390.

Smith, R.L. (1992), The extremal index for a Markov chain. *J. Appl. Prob.* **29**, 37-45.

Smith, R.L. and Weissman, I. (1992), Estimating the extremal index. Tentatively accepted for *J.R. Statist. Soc. B*.

Tawn, J.A. (1988), Bivariate extreme value theory - models and estimation. *Biometrika* **75**, 397-415.

Tawn, J.A. (1990), Discussion of paper by Davison and Smith. *J.R. Statist. Soc. B* **52**, 428-249.