ON GENERATING POSITIVELY DEPENDENT SEQUENCES OF RANDOM VARIABLES WITH A GIVEN DISTRIBUTION, WITH APPLICATIONS TO EXTREME VALUES

Harry Jøe

Tech. Report No. 112

SIAM INSTITUTE FOR MATHEMATICS AND SOCIETY

HEALTH SCIENCES PROGRAM RAND CORPORATION

DEPARTMENT OF STATISTICS STANFORD UNIVERSITY

DEPARTMENT OF STATISTICS UNIVERSITY OF BRITISH COLUMBIA
On Generating Positively Dependent Sequences of Random Variables with a Given Distribution, with Applications to Extreme Values

Harry Joe

Tech. Report #112

October 1987

Department of Statistics, University of British Columbia, 2021 West Mall, Vancouver, B.C. V6T 1W5
On Generating Positively Dependent Sequences of Random Variables with a Given Distribution with Application to Extreme Values

Harry Joe

Department of Statistics

University of British Columbia

ABSTRACT

This paper is about generating random variables with a given distribution in such a way that the successive elements of the sequence will be positively dependent. Large values of a member of the sequence will tend to imply large values of its neighbours. Of particular concern will be sequences which have a very simple autocorrelation function which is of geometric form. Several methods of generating random sequences are considered because there is not a single method that will easily simulate random variables for all common distributions like the normal, the gamma, the Poisson and the Pareto. The dependent sequences provide a means of checking the sensitivity of any given procedure which assumes that the random variables are independently and identically distributed irrespective of how realistic this assumption might be in practice.

Little has been done on the problem of determining dependent structures for nonnormal random variables. And most research on the robustness of statistics dwells on their sensitivity to outliers or distortions to their distribution's shape rather than on the sensitivity of their distributions to a lack of independence. For inference based on extreme values, the tail of the distribution is of critical importance; the data are often positively dependent so that a study of dependent structures for nonnormal random variables is dictated. This is especially true since in many scientific and engineering areas extreme values of autocorrelated data are studied. For example, in the study of air or water quality, estimates of extreme concentrations of pollutants are compared with critical levels. These results are mainly about sequences of normally distributed random variables. The three methods described in this paper are based upon the self decomposability of the distribution of the random variables, the infinitely divisibility of this distribution, or the representation of this distribution as a mixture of distributions. These can be used to generate random variables for many of the common distributions, for example, normal, exponential, gamma, Poisson, negative binomial and pareto. Although no single method can be used for all of these distributions, the first method has been used for modelling nonnormal time series while the other two methods are new. The first method is not applicable for integer-valued random variables such as the Poisson distribution and it has a couple of other small disadvantages, so this is a reason for proposing other methods of obtaining geometrically correlated random variables. For example, the second method can be used to model Poisson counts that may be serially correlated in time or space. Since self-decomposable distributions are a subclass of the infinitely divisible distributions, the range of cases covered by the first two cases is limited and hence a third method is considered.

Over all the paper consists of two parts. The first addresses the problem of generating a stationary sequence of random variables with a geometric autocorrelation function, a problem of independent interest as indicated above. The second part of the paper is about the sensitivity of two inferential procedures for extreme values, to the assumption that the observations on which the procedures are based are independent when in fact they are positively correlated. The second part of this paper uses the results of the first part in reaching the conclusions obtained there. The first is encouraging; the procedures are not unduly sensitive to weak positive dependence. However, there are discernable trends related to increasing positive dependence. For example the usual standard error of the maximum likelihood estimate of a real valued function of a parameter computed assuming independence, may be too small when there is positive dependence; thus the reliability of the estimate could be seriously over estimated.

Key Words: autocorrelation, infinite divisibility, mixtures, associated random variables, extreme value theory, generalized extreme-value distribution, generalized pareto distribution.
1. Introduction. In this paper, we study methods of generating a stationary positively dependent sequence of random variables with a given distribution $F$. In particular, we will be concerned with a geometric autocorrelation function, the simplest realistic choice. Several methods are considered because there is not a single method that will easily simulate random variables for all common distributions, such as normal, gamma, Poisson and Pareto. The dependence models involved in these methods provide a means of checking the sensitivity of a procedure which assumes that random variables are independently and identically distributed (i.i.d.), irrespective of how realistic they might be in practice. In a similar vein, Gaver and Lewis (1980) mention that their correlated exponential sequences can be used to check sensitivity of queueing results to departures from independence.

There seems to be very little work on dependence structures for nonnormal random variables. Also, most research in robust statistics is on sensitivity to outliers or distortions to distribution shape. See, for example, Hampel et al. (1986), Chapter 1 and 8, for a discussion. For inference based on extreme values, the tail of a distribution is of main importance; the data are often positively dependent so that a study of dependence structures for nonnormal random variables is useful. Hirtzel, Corotis and Quon (1982), Hirtzel (1985), and Hirtzel and Chan (1987) mentioned several scientific and engineering areas where the study of properties of extreme values of autocorrelated data is relevant; one example is in the study of air or water quality where estimates of extreme concentrations of pollutants can be compared with critical levels. The results in these papers are mainly for normal sequences.

The three methods described in Section 2 of this paper are based on the self-decomposability of $F$, the infinite divisibility of $F$, or the representation of $F$ as a mixture of distributions. These methods can be used for many of the common distributions, for example, normal, exponential, gamma, Poisson, negative binomial and Pareto, although no single method can be used for all of these distributions. The first method has been used for modelling of nonnormal time series and the other two methods are new. The first method is not applicable for integer-valued random variables such as the Poisson distribution and has a couple of other small disadvantages, so this is a reason for proposing other methods of obtaining geometrically correlated random variables. For example, the second method can be used to model Poisson counts that may be serially correlated in time or space. Since self-decomposable distributions are a subclass of the infinitely divisible distributions, the cases covered by the first two methods is limited and hence a third method is considered.

In Section 3, it is shown that the methods lead to random variables that are positively dependent in the sense of association (Eary, Proschan and Walkup, 1967). The use of association leads to certain probability inequalities, for example, the maximum of one of these positively dependent sequences is stochastically smaller than that obtained under independence. In Section 4, the methods are used to check the sensitivity to positive dependence of the generalized extreme-value and generalized Pareto procedures for estimating tail probabilities.

In summary, this paper consists of two problems. The first problem is obtaining probabilistic methods for generating a stationary sequence of random variables with a geometric autocorrelation function. This
problem is of interest in itself, irrespective of applications. The second problem is determining the sensitivity
to an assumption of independent observations for two inferential procedures for extreme values. This is of
interest in itself; in this paper it is presented as an application of methods for generating positively correlated
random variables.

2. Methods for generating geometrically correlated random variables. In this section, we
consider methods for generating a stationary sequence of random variables \( X_1, X_2, \ldots \) with the common
distribution function \( F \) and autocorrelation function \( \rho(j) = c \theta^j, \ j \geq 1 \), where \( \rho(j) \) is the correlation of \( X_i \)
and \( X_{i+j} \) and \( c \in (0,1) \) is a constant.

Some notation used in this paper are the following. Exponential \((\mu)\) denotes the exponential distribution
with mean \( \mu \), Gamma \((\alpha, \lambda)\) denotes the gamma distribution with shape parameter \( \alpha \) and mean \( \alpha/\lambda \), Binomial
\((n, p)\) denotes the binomial distribution with parameters \( n \) and \( p \), Negative Binomial \((\alpha, p)\), where \( \alpha > 0 \)
can be a non-integer, denotes the negative binomial distribution with parameters \( \alpha \) and \( p \) and with support
on the nonnegative integers, Normal \((\mu, \sigma^2)\) denotes the normal distribution with mean \( \mu \) and variance \( \sigma^2 \),
Poisson \((\mu)\) denotes the Poisson distribution with mean \( \mu \), and Beta \((\alpha, \beta)\) denotes the beta distribution
with parameters \( \alpha \) and \( \beta \).

Method 1. This method is a generalization of the normal autoregressive process of order one; \( F \) must
be self-decomposable or of the class \( L \) (Gaver and Lewis, 1980; Feller, 1971, p. 588). The sequence \( \{X_i\} \) is
based on the recursion

\[
X_i = \theta X_{i-1} + \epsilon_i, \quad i > 1, \tag{2.1}
\]

where \( \{\epsilon_i\} \) is a sequence of i.i.d. random variables independent of the \( \{X_i\} \). (The definition of self-
decomposability of \( F \) is that if \( X \) is a random variable with distribution \( F \), then for every \( 0 < \theta < 1 \), there
is a random variable \( \epsilon \) independent of \( X \) such that \( \theta X + \epsilon \) has the distribution \( F \).) The autocorrelation can
easily be shown to be \( \rho(j) = \theta^j \). Special cases are the following.

(i) If \( F \) is the Normal \((0,1)\) distribution, then \( \epsilon_i \) is a Normal \((0,1 - \theta^2)\) random variable.

(ii) If \( F \) is the Exponential \((1)\) distribution, then, with probability \( \theta \), \( \epsilon_i \) is zero, and with probability
\( 1 - \theta \), \( \epsilon_i \) is an Exponential \((1)\) random variable.

(iii) If \( F \) is the Gamma \((\alpha, 1)\) distribution which \( \alpha \) is a positive integer, then, with probability \( \theta^\alpha \), \( \epsilon_i \) is
zero, and with probability \( p_j \), \( 1 \leq j \leq \alpha \) \( \epsilon_i \) is a Gamma \((j, 1)\) random variable, where \( p_j \) are probabilities
from a Binomial \((\alpha, 1 - \theta)\) distribution.

(iv) If \( F \) is the Gamma \((\alpha, 1)\) distribution, where \( \alpha \) may be a non-integer, then Bernier (1970) has
shown that with probability \( \theta^\alpha \), \( \epsilon_i \) is zero, and with probability \( p_j \) \( (j \geq 1) \), \( \epsilon_i \) is a Gamma \((j, \theta)\) random
variable, where \( p_j \) are probabilities from a Negative Binomial \((\alpha, \theta)\) distribution, that is,

\[
p_j = \frac{\Gamma(\alpha + j)}{\Gamma(\alpha) j!} \theta^\alpha (1 - \theta)^j, \quad j = 0, \ldots.
\]

Alternatively Lawrance (1978) has shown that \( \epsilon_i \) has a compound Poisson distribution.
The $\epsilon_i$ in (ii) to (iv) can be obtained via the equation from the moment generating functions or characteristic functions of (2.1). By the form of (2.1), $F$ must be a continuous distributions. Other self-decomposable distributions include the Pareto and lognormal distributions (Thorin, 1977a, 1977b), but $\epsilon_i$ do not have a simple form. The special case (ii) above results in the exponential first-order autoregressive sequence, denoted by EAR(1) (Lawrance and Lewis, 1980). The gamma first-order autoregressive process is studied in Gaver and Lewis (1980).

Gaver and Lewis (1980) note that when $F$ is either the exponential or gamma distribution function, (2.1) does not lead to a time-reversible process (in contrast with the normal case). The multivariate distribution of the sequence $X_1, \ldots, X_n$ from (2.1) has singular components on sets of the form $\{X_i = \theta X_{i-1}\}$ because $\epsilon_i$ can be zero with positive probability. This means that for a sample sequence, one would expect to see consecutive pairs of the form $(x, \theta x)$ a proportion $\theta^\alpha$ of the time. This might not be intuitively the behaviour expected for a positively dependent sequence.

A slight modification to get a sequence with an absolutely continuous distribution is:

$$X_i = X_i^0 + A_i, \quad A_i = \theta A_{i-1} + \epsilon_i, \quad i > 1,$$

where $\{X_i^0\}$ is a sequence of i.i.d. random variables independent of $\{A_i\}$ and $\{\epsilon_i\}$ is a sequence of i.i.d. random variables which are independent of $\{X_i^0\}$ and $\{A_i\}$. The autocorrelation function is $\rho(j) = c\theta^j$, where $c = \text{Var}(X_i^0)/\text{Var}(X_i)$. For example, if $F$ is the Gamma $(\alpha,1)$ distribution, then $X_i^0$ and $A_i$ can be chosen to be Gamma $(\alpha_0,1)$ and $(\beta,1)$ random variables respectively, where $\alpha_0 + \beta = \alpha$.

In summary, the method based on (2.1) has possible disadvantages of not being time-reversible, having a singular component on sets that do not suggest positive dependence, and not being usable for integer-valued distributions.

The next method leads to sequences that are time-reversible, but if $F$ is normal, the random variables are not jointly normal.

**Method 2.** Let $F$ be an infinitely divisible distribution. Let $M$ be a "large" positive integer and let $F_{1/M}$ be the distribution whose $M^{th}$ convolution is $F$. Let the sequence $\{X_i\}$ be defined by

$$X_i = X_{i1} + \cdots + X_{iM}, \quad i \geq 1,$$

where $X_{ij}$ are i.i.d. random variables with distribution $F_{1/M}$ for each $i$, and with probability $\theta$ ($0 \leq \theta < 1$), $X_{ij} = X_{i-1,j}$, and with probability $1 - \theta$, $X_{ij}$ is independent of all $X_{k\ell}$, $\ell \neq i$. It is easy to show that the autocorrelation function is $\rho(j) = \theta^j$. There are many infinitely divisible distributions but this method is only useful if $F_{1/M}$ has a simple form. Special cases are the following.

(i) If $F$ is the Gamma $(\alpha,1)$ distribution, then $F_{1/M}$ is the Gamma $(\alpha/M,1)$ distribution. To generate $\{X_i\}$, (a) let $L_i$ be a Binomial $(M, \theta)$ random variable, (b) if $L_i = 0$ then generate $X_i$ to have a Gamma $(\alpha,1)$ distribution independent of $X_{i-1}$, (c) if $L_i = M$ then $X_i = X_{i-1}$, and (d) if $0 < L_i < M$, let $B_i$ be a
Beta \((\alpha L_i / M, \alpha - \alpha L_i / M)\) random variable, let \(G_i\) be a Gamma \((\alpha - \alpha L_i / M, 1)\) random variable and let \(X_i = B_i X_{i-1} + G_i\).

(ii) If \(F\) is the Poisson \((\mu)\) distribution, then \(F_{1/M}\) is the Poisson \((\mu / M)\) distribution. To generate \(\{X_i\}\), (a) let \(L_i\) be a Binomial \((M, \theta)\) random variable, (b) if \(L_i = 0\) then generate \(X_i\) to have a Poisson(\(\mu\)) distribution independent of \(X_{i-1}\), (c) if \(L_i = M\) then \(X_i = X_{i-1}\), (d) if \(0 < L_i < M\), let \(B_i\) be a Binomial \((X_{i-1}, L_i / M)\) random variable, let \(P_i\) be a Poisson \((\mu - \mu L_i / M)\) random variable and let \(X_i = B_i + P_i\).

Note that the gamma random variables from (i) have a joint distribution with a singular component. The measure of the singular component can be made arbitrarily small by making \(M\) large. The singular component puts mass on events of the form \(\{X_i = \theta X_{i-1}\}\) rather than events of the form \(\{X_i = \theta X_{i-1}\}\) for Method 1.

This method basically requires infinite divisibility of \(F\) if \(M\) is large. It can be used for an arbitrary \(F\) if \(M = 1\), but this would be a more non-realistic dependence structure. The next method can be used for non-infinitely divisible distributions, but the autocorrelation function is \(\rho(j) = c \theta^j, j \geq 1\), where \(0 < c < 1\). We have not been able to come up with a method that has autocorrelation function \(\rho(j) = \theta^j\) and that in general has no singular component for a continuous \(F\).

**Method 3.** Let \(F\) be a distribution that can be written in the mixture form:

\[
F(x) = \int H(x; \lambda) dG(\lambda),
\]

where \(\{H(x; \lambda)\}\) is a parametric family of distribution functions and \(G\) is a distribution function. The sequence \(\{X_i\}\) is defined through the following. Let \(\{L_i\}\) be a sequence of random variables with the common distribution \(G\) such that with probability \(\theta\), \(L_i = L_{i-1}\), and with probability \(1 - \theta\), \(L_i\) is independent of \(L_k\), \(k \neq i\). Given \(L_i = \lambda_i\), \(X_i\) has distribution \(H(\cdot; \lambda_i)\). It can be shown that \(\rho(j) = c \theta^j\), where \(c = \Var(E(X|\Lambda))/\Var(X)\), \(X\) has distribution \(F\), \(\Lambda\) has distribution \(G\), and \(X\) given \(\Lambda = \lambda\) has distribution \(H(\cdot; \lambda)\). A dependent sequence of this form is a special case of the classes of multivariate distributions studied in Marshall and Olkin (1986). Special cases are the following.

(i) If \(F(x) = 1 - (1 + x)^{-\alpha}, x > 0, \alpha > 2,\) is the Pareto distribution with parameter \(\alpha\), then it can be written as a mixture with \(H(x, \lambda) = 1 - e^{-\lambda x}\) and \(G(\lambda)\) being the Gamma \((\alpha, 1)\) distribution. In the autocorrelation function, \(c = 1/\alpha\).

(ii) If \(F\) is the Negative Binomial \((\alpha, p)\) distribution with probability mass function \(f(x) = \Gamma(\alpha + x)p^\alpha(1 - p)^x/\Gamma(\alpha)x!, x = 0, 1, \ldots\), then it can be written as a mixture with \(H(x, \lambda)\) being the Poisson \((\lambda)\) distribution and \(G(\lambda)\) being the Gamma \((\alpha, \beta)\) distribution where \(\beta = p/(1 - p)\). In the autocorrelation function, \(c = 1 - p\).

(iii) If \(F\) is the Poisson \((\mu)\) distribution, then it can be written as a mixture with \(H(x, \lambda)\) being the Binomial \((\lambda, p)\) distribution and \(G(\lambda)\) being the Poisson \((\mu/p)\) distribution; \(p\) can be any real between 0 and 1. In the autocorrelation function, \(c = p\).
3. Association and inequalities. We will show that under some conditions the dependent sequences defined in Section 2 are positively dependent in the sense of Esary, Proschan and Walkup (1967). From this result, inequalities for probabilities involving the dependent sequences can be obtained. One consequence is that the maximum of the dependent sequence of length $N$ is stochastically smaller than the maximum of an independent sequence of the same length (compare Slepian’s lemma, Leadbetter, Lindgren and Rootzen, 1983). This result is discussed at the end of the section and leads into Section 4 on inference based on extreme values.

We start with some definitions of concepts of positive dependence.

**Definition 3.1.** Random variables $T_1, \ldots, T_n$ are associated if $\text{Cov}(g(T_1, \ldots, T_n), h(T_1, \ldots, T_n)) \geq 0$ for all functions $g, h$ nondecreasing in each argument such that the covariance exists.

**Definition 3.2.** Random variables $T_1, \ldots, T_n$ are conditionally increasing in sequence if $\text{Pr}(T_i > t | T_1 = t_1, \ldots, T_{i-1} = t_{i-1})$ is nondecreasing in $t_1, \ldots, t_{i-1}, i = 2, \ldots, n$.

An important relation between the above two definitions is that if a sequence of random variables is conditionally increasing in sequence then the sequence is associated (Barlow and Proschan, 1975). We will show that the sequences of random variables from (2.1), (2.2), (2.3) are associated by showing that they are conditionally increasing in sequence; this latter property is easier to show because the Markov property of the sequences means that calculations involve only 2 random variables at a time.

**Theorem 3.3.** Let $X_1, \ldots, X_n$ be random variables with a common distribution $F$. Suppose there are i.i.d. random variables $\epsilon_2, \ldots, \epsilon_n$ and a real number $\theta$ in $(0,1)$ such that $X_i = \theta X_{i-1} + \epsilon_i, i = 2, \ldots, n$. Then $X_1, \ldots, X_n$ are conditionally increasing in sequence.

**Proof.**

$$\text{Pr}(X_i > t | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) = \text{Pr}(X_i > t | X_{i-1} = x_{i-1}) = 1 - F_{\epsilon}(t - \theta x_{i-1})$$

is increasing in $x_{i-1}$, where $F_{\epsilon}$ is the distribution function of $\epsilon$ (compare Alam and Saxena, 1981, Section 3). $\square$

**Corollary 3.4.** Let $X_1, \ldots, X_n$ be random variables with a common distribution $F$. Suppose $X_i = X_i^0 + A_i, A_i = \theta A_{i-1} + \epsilon_i, i > 1$, where $\{X_i^0\}$ is a sequence of i.i.d. random variables independent of $\{A_i\}$ and $\{\epsilon_i\}$ is a sequence of i.i.d. random variables independent of $\{X_i^0\}$ and $\{A_i\}$. Then $X_1, \ldots, X_n$ are associated.

**Proof.** Let $g$ and $h$ be functions which are nondecreasing in the arguments $x_1, \ldots, x_n$. Suppose $\text{Cov}(g(X_1, \ldots, X_n), h(X_1, \ldots, X_n))$ exists. Then $\text{Cov}(g(X_1, \ldots, X_n), h(X_1, \ldots, X_n)) = E[\text{Cov}(g(X_1^0 + A_1, \ldots, X_n^0 + A_n), h(X_1^0 + A_1, \ldots, X_n^0 + A_n)|X_1^0, \ldots, X_n^0] + \text{Cov}(E(g(X_1^0 + A_1, \ldots, X_n^0 + A_n)|X_1^0, \ldots, X_n^0), E(h(X_1^0 + A_1, \ldots, X_n^0 + A_n)|X_1^0, \ldots, X_n^0))].$ From Theorem 3.3, $A_1, \ldots, A_n$ are associated so that $\text{Cov}(g(X_1^0 + A_1, \ldots, X_n^0 + A_n), h(X_1^0 + A_1, \ldots, X_n^0 + A_n)|X_1^0, \ldots, X_n^0)$ is nonnegative for each $X_1^0, \ldots, X_n^0$ and its expectation is nonnegative. $E(b(X_1^0 + A_1, \ldots, X_n^0 + A_n)|X_1^0, \ldots, X_n^0), b = g, h$ is nondecreasing in $X_1^0, \ldots, X_n^0$. Since $X_1^0, \ldots, X_n^0$ are independent random variables and hence associated, the covariance of the conditional expectations is nonnegative. Therefore, $X_1, \ldots, X_n$ are associated. $\square$
Theorem 3.5. Let $X_1, \ldots, X_n$ be random variables with a common distribution $F$ and characteristic function $\psi$. Suppose $X_i = X_{i1} + \cdots + X_{iM}$, $i \geq 1$, where $X_{ij}$ are i.i.d. random variables with distribution $F_{i/M}$ for each $i$, and with probability $\theta$, $X_{ij} = X_{i-1,j}$, and with probability $1 - \theta$, $X_{ij}$ is independent of all $X_{ik}$, $k \neq i$. Suppose that if $U$ and $V$ are independent random variables with respective characteristic functions $\psi^a$ and $\psi^b$, where $a$ and $b$ are positive multiples of $M^{-1}$ such that $a + b = 1$, then $\Pr(U > t|U + V = x)$ is increasing in $x$ for all $t$. Then $X_1, \ldots, X_n$ are conditionally increasing in sequence.

Proof. 

\[ \Pr(X_i > t|X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) = \Pr(X_i > t|X_{i-1} = x_{i-1}) \]

\[ = \sum_{\ell=0}^{M} \Pr(X_i > t|X_{i-1} = x_{i-1}, L = \ell) \Pr(L = \ell), \]

where $L$ is a binomial random variable with parameters $M$ and $\theta$. It suffices to show that $\Pr(X_i > t|X_{i-1} = x_{i-1}, L = \ell)$ is increasing in $x_{i-1}$ for each $\ell$. Letting $x = x_{i-1}$, this probability is equal to

\[ \Pr(U + V_2 > t|U + V_1 = x, L = \ell), \] (3.1)

where $U, V_1, V_2$ are independent, $U$ has distribution $F_{t/M}$ and $V_1, V_2$ have distribution $F_{1-t/M}$, where $F_0$ is the distribution with unit mass at 0 and, for $s > 0$, $F_s$ is the distribution with characteristic function $\psi^s$. (3.1) is increasing in $x$ if $\Pr(U > t|U + V_1 = x, L = \ell)$ is increasing in $x$, and this property follows from the given assumptions. \hfill \Box

Examples. For $F$ gamma and Poisson, the condition of Theorem 3.5 can easily be shown to hold.

Theorem 3.6. Let $X_1, \ldots, X_n$ be random variables with a common distribution $F$, where $F(x) = \int H(x; \lambda)dG(\lambda)$. Suppose $\{\Lambda_i\}$ is a sequence of random variables with the common distribution $G$ such that with probability $\theta$, $\Lambda_i = \Lambda_{i-1}$, and with probability $1 - \theta$, $\Lambda_i$ is independent of $\Lambda_k$, $k \neq i$, and that given $\Lambda_i = \lambda_i$, $X_i$ has distribution $H(\cdot; \lambda_i)$. Let $G(\lambda|x)$ be the conditional distribution of $\Lambda_i$ given $X_i = x$. If $H(x; \lambda)$ is stochastically increasing (decreasing) in $\lambda$ and $G(\lambda|x)$ is stochastically increasing (decreasing) in $x$, then $X_1, \ldots, X_n$ are conditionally increasing in sequence.

Proof. 

\[ \Pr(X_i > t|X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) = \Pr(X_i > t|X_{i-1} = x_{i-1}) \]

\[ = (1 - \theta) \Pr(X_i > t) + \theta \Pr(X_i > t|X_{i-1} = x_{i-1}, \Lambda_i = \Lambda_{i-1}). \]

It suffices to show that $\Pr(X_i > t|X_{i-1} = x, \Lambda_i = \Lambda_{i-1}) = \int [1 - H(t; \lambda)]dG(\lambda|x)$ is increasing in $x$. This will follow if $H(\cdot; \lambda)$ is stochastically increasing (decreasing) in $\lambda$ and $G(\lambda|x)$ is stochastically increasing (decreasing) in $x$. \hfill \Box

Examples. For the special cases given in Section 2, the condition of Theorem 3.6 are satisfied.

(i) $F$ Pareto: $H(x; \lambda) = 1 - e^{-\lambda x}$ is stochastically decreasing in $\lambda$, and $G(\lambda|x)$ is the Gamma $(\alpha+1, 1+x)$ distribution so that $G(\lambda|x)$ is stochastically decreasing in $x$. 


(ii) $F$ Negative Binomial: $H(z; \lambda)$ is the Poisson ($\lambda$) distribution so that $H(z; \lambda)$ is stochastically increasing in $\lambda$, and $G(\lambda|x)$ is the Gamma ($\alpha + x, 1/(1 - p)$) distribution.

(iii) $F$ Poisson: $H(z; \lambda)$ is the Binomial ($\lambda, p$) so that it is stochastically increasing in $\lambda$, and $G(\lambda|x)$ is the distribution of $x$ plus a Poisson ($(1 - p)/\mu/p$) random variable.

A consequence of association of the sequences (2.1), (2.2) and (2.3) is that the maximum of an associated sequence is stochastically smaller than the maximum if the sequence consisted of independent random variables. This is a known result and the derivation is outlined below for completeness.

If $X_1, \ldots, X_n$ are associated, then by taking $g(x_1, \ldots, x_n) = -I_A(x_1, \ldots, x_n)$ and $h(x_1, \ldots, x_n) = -I_B(x_1, \ldots, x_n)$ where $I$ is the indicator function of a set, $A = (\infty, t_1] \times \cdots \times (\infty, t_j] \times (\infty, \infty) \times \cdots \times (\infty, \infty)$ $B = (\infty, \infty) \times \cdots \times (\infty, \infty) \times (\infty, t_{j+1}] \times \cdots \times (\infty, t_n]$, for some $1 \leq j \leq n - 1$ and some reals $t_1, \ldots, t_n$, the inequality

$$\Pr(X_i \leq t_i, i = 1, \ldots, n) \geq \Pr(X_i \leq t_i, i = 1, \ldots, j) \Pr(X_i \leq t_i, i = j + 1, \ldots, n)$$

obtains. Since a subset of a set of associated random variables is associated,

$$\Pr(X_i \leq t_i, i = 1, \ldots, n) \geq \prod_{i=1}^n \Pr(X_i \leq t_i)$$

(3.2)

Random variables satisfying (3.3) are called lower positive orthant dependent (Dykstra, Hewett and Thompson, 1973).

If $X_1, \ldots, X_n$ are associated random variables with a common distribution $F$, then by taking $t_i = t$, $i = 1, \ldots, n$ in (3.2), we obtain

$$\Pr(\max(X_1, \ldots, X_n) \leq t) \geq F^n(t) \quad \text{for all } t. \quad (3.3)$$

That is, $\max(X_1, \ldots, X_n)$ is stochastically smaller than if $X_1, \ldots, X_n$ were i.i.d.

Let $F$ be a distribution and let $F_n$ be the distribution of the maximum of the dependent sequence $X_1, \ldots, X_n$ generated through (2.1), (2.2) or (2.3). From extreme value theory for weakly dependent stationary sequences (Leadbetter, Lindgren and Rootzen, 1983), the limiting extreme value distribution corresponding to $F_n$ is the same as the limiting extreme value distribution corresponding to $F^n$ (which is the distribution for the maximum of an independent sequence of length $n$ from the same distribution) provided a limiting extreme value distribution exists. Therefore $F_n$ will be "closer" to $F^n$ as $n$ increases with $F_n$ being stochastically smaller than $F^n$ according to the preceding paragraph. For example, for a normal autoregressive sequence (mean 0 and variance 1) with autocorrelation function $\rho(j) = 0.8^j$, the medians of $F^n$ for $n = 5, 10, 20, 50, 100, 500, 1000$ are 1.13, 1.50, 1.82, 2.20, 2.46, 2.99, 3.20 respectively. The corresponding estimated medians for $F_n$ are 0.63, 1.01, 1.41, 1.88, 2.22, 2.86, 3.10; these are based on sets of 1000 simulations and the standard errors of these estimates are less than 0.02.
4. Application to inference from extreme value theory. One reason for studying ways of generating dependent random variables with a given distribution is to check the sensitivity of an inference procedure to an assumption of independence. For example, Joe (1987) considers the problem of estimating or forecasting the quantiles of the maximum of $N$ observations (or equivalently estimating tail probabilities) when there are $n$ i.i.d. random variables from an unknown distribution $F$. For applications to extremes in hydrology or environmetrics, an assumption of weak positive dependence is more reasonable than an assumption of independence. However, it may be difficult to model the dependence and even if a model were possible the analysis would be harder than the analysis with an assumption of independence.

In Joe (1987) and Smith (1987) and references therein, inference procedures are based on two parametric approximations to $F^N$ from extreme value theory. These will be called the generalized Pareto and generalized extreme-value procedures. For the generalized Pareto procedure, the observations larger than a level $u$, with $u$ subtracted, are considered as a random sample from a generalized Pareto distribution (Pickands, 1975)

$$G(y; \delta, \gamma) = 1 - (1 + \gamma y/\delta)^{-1/\gamma},$$

which approximates $1 - \bar{F}(u + y)/\bar{F}(u)$, where $\bar{F} = 1 - F$. Based on maximum likelihood estimates of $\gamma$ and $\delta$, an estimate of the 100$p$-percentile of $F^N$ is

$$\hat{x}_{p,N} = u + \delta \{[mn^{-1}(1 - p^{1/n})]^{-\gamma} - 1\}/\gamma,$$  \hspace{1cm} (4.1)

where $m$ is the number of observations greater than $u$. For the generalized extreme-value procedure, the $n$ observations are divided into $m$ groups of $k$ each and the $m$ group maxima $Z_1, \ldots, Z_m$ are considered as random sample from the generalized extreme-value distribution (Prescott and Walden, 1980)

$$H(z; \gamma, \alpha, \beta) = \exp\{-[1 + \gamma(z - \beta)/\alpha]^{-1/\gamma}\},$$

(where $[w]_+ = \max[w, 0]$), which approximates $F^k(z)$. Based on maximum likelihood estimates of $\gamma, \alpha, \beta$, an estimate of the 100$p$-percentile of $F^N$ is

$$\hat{x}_{p,N} = \hat{\beta} + \hat{\alpha} \{[-kN^{-1} \log p]^{-\gamma} - 1\}/\gamma.$$  \hspace{1cm} (4.2)

The approximations are based on asymptotic theory for i.i.d. random variables from a distribution $F$ (that is in the domain of attraction of an extreme value distribution). Independence would usually be an approximation to weak positive dependence. In this section, we study the sensitivity of the generalized Pareto and generalized extreme-value procedures to weak positive dependence by simulating sequences from the three methods in Section 2. Normal random variables were generated using method 1, gamma and exponential random variables were generated using methods 1 and 2, and Pareto random variables were generated using method 3.

Some simulation results for the generalized Pareto and generalized extreme-value procedures are presented in Tables 1 and 2. For the simulations for the generalized Pareto procedure, $u = F^{-1}(0.9)$, $n = 600,
\( N = 1000, p = 0.5; \) for the simulations for the generalized extreme-value procedure, \( n = 600, m = 30, k = 20, N = 1000 \) and \( p = 0.5 \). These are some of the values of \( m, k, u \) used in the simulations in Joe (1987). The comparison here should only be made within each procedure since the simulations are intended only to show the effect of positive dependence. The simulations in Joe (1987) involved various different values of \( m, k, u \) to permit a comparison of the procedures. Some comments from the tables are:

1. The procedures are not sensitive to weak positive dependence. With weak dependence (\( \theta < 0.5 \)), the effect on the estimates is negligible. The dependence among group maxima or among observations above a certain level is, of course, less than the dependence among the observations. It takes strong serial dependence (\( \theta \) close to 0.8) to get estimates with averages quite different than would be obtained under independence; this dependence would be stronger than that in most real data sets with time dependence. Note that for the random sequences from the Pareto (5) distribution generated using method 3, the autocorrelation function is \( \rho(j) = 0.2\theta^j, j \geq 1 \), so that even \( \theta = 0.8 \) does not represent strong dependence here.

2. An explanation of the trends in the expected values of \( \hat{\beta}, \hat{z}_{p,N}, \hat{z}_{x,N} \) (from (4.1) and (4.2)) is the following. Let \( F_j \) denote the distribution of the maximum of a dependent sequence of length \( j \) of random variables with the common distribution \( F \). From the remark at the end of Section 3, \( F_j \) is closer to \( F^j \) as \( j \) increases. However, the estimates from on the generalized extreme-value procedure are based on the approximation

\[
F_N \approx (F_k)^{N/k}
\]

so that underestimation of quantiles of \( F_N \) for large \( N \) might be expected for strong dependence. This effect is seen in the decrease of the expected value of \( \hat{\beta} \) and \( \hat{z}_{x,N} \) as \( \theta \) increases for the generalized extreme-value procedure. From comparison of the remark at the end of Section 3 with Table 2, the underestimation for strong dependence seems to be negligible relative to the variability. There is also a decrease in the expected values of \( \hat{z}_{p,N} \) as \( \theta \) increases for the generalized Pareto procedure but this cannot be simply explained from the generalized Pareto approximation.

3. An explanation for the increase in the standard deviations of some of the parameter estimates is the following. If \( Y_1, \ldots, Y_m \) are positively dependent random variables, each with a density in the parametric family \( f(x; \eta) \), where \( \eta \) is an unknown vector of parameters, and \( \eta \) is estimated by maximum likelihood under the assumption that \( Y_1, \ldots, Y_m \) are independent, then, under some forms of positive dependence, the asymptotic covariance matrix of \( \hat{\eta} \) is larger (with respect to the ordering of positive definiteness) than the asymptotic covariance matrix under independence. This result is related to that in Gleser and Moore (1983), where a strong form of positive dependence is assumed. A consequence of this result is that the usual standard error of the maximum likelihood estimate of a real valued function of \( \eta \), computed assuming independence, may be too small when there is positive dependence. For the generalized Pareto and generalized extreme-value procedures, there is the extra complication of the parametric families being approximations. The simulations do show a increasing trend as \( \theta \) increases in the standard deviations of some of the parameter estimates and in the estimate of the median of \( F_N \). Because the standard errors based on the expected or
observed information matrix for the generalized Pareto or generalized extreme-value distributions assume independent observations, these might be a bit small when there is positive dependence (of the form in the simulations).

4. Finally note that as $\theta$ increases, the trend in $\hat{\gamma}$ is towards more bias (the normal, exponential and gamma distributions are in the domain of attraction of the extreme-value distribution with $\gamma = 0$ and the Pareto (5) distribution is in the domain of attraction of the extreme-value distribution with $\gamma = 0.2$).

Acknowledgements. This research was supported by the EPA through a cooperative research agreement with SIMS. The author thanks Jim Zidek for valuable comments and for pointing out an error in an earlier draft.

References.


North-Holland, Amsterdam.


### TABLE 1. Simulated expected values and standard deviations of estimates of parameters for generalized Pareto procedure; 3000 replications; $N = 1000$; $p = 0.5$; $n = 600$; $u = F^{-1}(0.9)$

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\delta}$</th>
<th>$\bar{x}_{p,N} \dagger$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal (0,1) - Method 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>-0.21(0.15)</td>
<td>0.58(0.11)</td>
<td>3.09(0.34)</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.22(0.15)</td>
<td>0.58(0.11)</td>
<td>3.08(0.36)</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.23(0.15)</td>
<td>0.58(0.11)</td>
<td>3.06(0.37)</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.26(0.17)</td>
<td>0.59(0.14)</td>
<td>2.99(0.45)</td>
</tr>
<tr>
<td></td>
<td>Exponential (1) - Method 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>-0.05(0.16)</td>
<td>1.05(0.21)</td>
<td>7.13(1.29)</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.05(0.15)</td>
<td>1.06(0.21)</td>
<td>7.06(1.22)</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.05(0.15)</td>
<td>1.05(0.20)</td>
<td>7.09(1.30)</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.09(0.16)</td>
<td>1.07(0.24)</td>
<td>6.89(1.70)</td>
</tr>
<tr>
<td></td>
<td>Gamma (0.5,1) - Method 2 with $M = 20$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.00(0.16)</td>
<td>0.85(0.17)</td>
<td>5.69(1.22)</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.02(0.17)</td>
<td>0.86(0.20)</td>
<td>5.60(1.33)</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.06(0.22)</td>
<td>0.89(0.28)</td>
<td>5.52(1.83)</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.20(0.41)</td>
<td>1.04(0.62)</td>
<td>5.34(4.01)</td>
</tr>
<tr>
<td></td>
<td>Pareto (5) - Method 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.16(0.18)</td>
<td>0.33(0.07)</td>
<td>3.29(1.13)</td>
</tr>
<tr>
<td>0.2</td>
<td>0.15(0.18)</td>
<td>0.33(0.07)</td>
<td>3.28(1.18)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.15(0.18)</td>
<td>0.33(0.07)</td>
<td>3.29(1.23)</td>
</tr>
<tr>
<td>0.8</td>
<td>0.14(0.18)</td>
<td>0.33(0.07)</td>
<td>3.20(1.39)</td>
</tr>
</tbody>
</table>

\* For $\theta = 0$, $F^{-1}(0.5^{0.001})$ for the four distributions are 3.20, 7.27, 5.77, 3.28 respectively.
TABLE 2. Simulated expected values and standard deviations of estimates of parameters for generalized extreme-value procedure; 3000 replications; \( N = 1000; p = 0.5; m = 30; k = 20; n = 600 \)

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>( \hat{\gamma} )</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\beta} )</th>
<th>( \hat{x}_{p, N} ) †</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (0,1) - Method 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>-0.17(0.16)</td>
<td>0.46(0.07)</td>
<td>1.66(0.10)</td>
<td>3.13(0.39)</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.19(0.16)</td>
<td>0.48(0.07)</td>
<td>1.63(0.10)</td>
<td>3.12(0.39)</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.21(0.16)</td>
<td>0.54(0.08)</td>
<td>1.52(0.12)</td>
<td>3.10(0.39)</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.26(0.16)</td>
<td>0.68(0.11)</td>
<td>1.16(0.15)</td>
<td>2.98(0.44)</td>
</tr>
<tr>
<td>Exponential (1) - Method 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.00(0.17)</td>
<td>0.93(0.15)</td>
<td>3.04(0.21)</td>
<td>7.36(1.49)</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.01(0.17)</td>
<td>0.96(0.15)</td>
<td>2.99(0.21)</td>
<td>7.33(1.53)</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.03(0.17)</td>
<td>1.05(0.17)</td>
<td>2.78(0.24)</td>
<td>7.28(1.49)</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.01(0.17)</td>
<td>1.14(0.18)</td>
<td>2.04(0.29)</td>
<td>7.15(1.85)</td>
</tr>
<tr>
<td>Gamma (0.5,1) - Method 2 with ( M = 20 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.05(0.17)</td>
<td>0.78(0.12)</td>
<td>1.97(0.17)</td>
<td>5.94(1.54)</td>
</tr>
<tr>
<td>0.2</td>
<td>0.07(0.17)</td>
<td>0.73(0.12)</td>
<td>1.70(0.16)</td>
<td>5.64(1.56)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.11(0.18)</td>
<td>0.64(0.11)</td>
<td>1.26(0.15)</td>
<td>5.04(1.47)</td>
</tr>
<tr>
<td>0.8</td>
<td>0.26(0.20)</td>
<td>0.45(0.08)</td>
<td>0.61(0.11)</td>
<td>4.70(2.17)</td>
</tr>
<tr>
<td>Pareto (5) - Method 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.21(0.18)</td>
<td>0.34(0.06)</td>
<td>0.83(0.08)</td>
<td>3.51(1.43)</td>
</tr>
<tr>
<td>0.2</td>
<td>0.22(0.19)</td>
<td>0.34(0.06)</td>
<td>0.83(0.08)</td>
<td>3.57(1.40)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.22(0.19)</td>
<td>0.34(0.06)</td>
<td>0.80(0.08)</td>
<td>3.51(1.39)</td>
</tr>
<tr>
<td>0.8</td>
<td>0.22(0.19)</td>
<td>0.33(0.06)</td>
<td>0.75(0.08)</td>
<td>3.42(1.44)</td>
</tr>
</tbody>
</table>

† For \( \theta = 0, F^{-1}(0.5^{0.001}) \) for the four distributions are 3.20, 7.27, 5.77, 3.28 respectively.

14