Holger Rootzén

The Rate of Convergence of Extremes of Stationary Normal Sequences

Preprint
June
1981

Institute of Mathematical Statistics
University of Copenhagen
Holger Rootzén

THE RATE OF CONVERGENCE OF EXTREMES
OF STATIONARY NORMAL SEQUENCES

Preprint 1981 No. 8

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

June 1981
Abstract

Let \{\xi_t\} be a stationary normal sequence with zero means, unit variances, and covariances \( r_t = \text{E} \xi_s \xi_{s+t} \), let \( \{\xi_t\} \) be independent and standard normal, and write \( M_n = \max_{1 \leq t \leq n} \xi_t \), \( \hat{M}_n = \max_{1 \leq t \leq n} \hat{\xi}_t \). In this paper we find bounds on \( |P(M_n \leq u) - P(\hat{M}_n \leq u)| \) which are roughly of the order

\[
\frac{1}{n} \left( \frac{1-\rho}{1+\rho} \right)
\]

where \( \rho \) is the maximal correlation, \( \rho = \sup \{0, r_1, r_2, \ldots \} \), and it is shown that, at least for m-dependent sequences, the bounds are of the right order. Further, bounds of the same order on the rate of convergence of the point processes of exceedances of one or several levels are obtained using a "representation" approach (which seems to be of rather wide applicability). As corollaries we obtain rates of convergence of several functionals of the point processes, including the joint distribution function of the k largest values amongst \( \xi_1, \ldots, \xi_n \).

AMS 1980 subject classification: Primary 60G15, secondary 60F05, 60G10, 60F17.

Key words and phrases: Maxima of normal sequences, extreme values, rate of convergence.

Research supported in part by the Office of Naval Research under contract N 0014-75-C-0809.
1. Introduction and discussion of the results

The asymptotic theory of extremes of independent and of stationary normal sequences has found many applications as testified e.g. by the books by Gumbel (1958) and Leadbetter, Lindgren & Rootzén (1982) and the references therein. However, for practical use of asymptotic theory, it is important to know the rate of convergence. The aim of this paper is to study in some detail the rate of convergence in extremal results for dependent stationary normal sequences. For the independent case, the reader is referred to the papers by Hall (1979) and Nair (1981).

Let \( \xi_1, \xi_2, \ldots \) be a stationary normal sequence, which for convenience will be assumed to have zero means and unit variances, and let \( r_t = E \xi_s \xi_{s+t} \) be its covariance function. Further, let \( \hat{\xi}_1, \hat{\xi}_2, \ldots \) be an "associated independent sequence", i.e. a sequence of independent standard normal variables, write \( M_n = \max_{1 \leq t \leq n} \xi_t \), \( \hat{M}_n = \max_{1 \leq t \leq n} \hat{\xi}_t \), and let \( \Phi \) be the standard normal distribution function. The first main result of this paper (Theorem 3.1) is that, for \( u \) given by

\[
\frac{n(1-\Phi(u))}{\Phi(u_n)} = K
\]

\( (1.1) \)

the following holds:

\[
\sup_{u \geq u_n} |P(M_n \leq u) - P(\hat{M}_n \leq u)| \leq 4R_n
\]

Here \( R_n \) depends on \( K \) and the covariances \( \{r_t\} \) in a rather complicated way (given by (2.3) below). The leading term of \( R_n \) is determined by the largest covariance \( \rho = \sup\{0, r_1, r_2, \ldots\} \) and for \( \rho > 0 \) it is of the order

\[
\left( \frac{1}{n} \right) \frac{(1-\rho)/(1+\rho)}{(\log n)^{-\rho/(1+\rho)}},
\]

while for \( \rho = 0 \) the order is
which is improved to the order $\frac{1}{n}$ in the case when in addition only finitely many of the $r_t$'s are non-zero. Next, it is shown that if only finitely many $r_t$'s are non-zero and if $u_n \to \infty$ in such a way that $P(M_n \leq u_n)$ converges to a non-trivial limit (or equivalently if $n(1 - \phi(u_n)) + K$, for some constant $K > 0$), then

$$P(M_n \leq u_n) - P(M_n \leq u_n) \sim e^{-K} R_n$$

if $\rho > 0$, and

$$P(M_n \leq u_n) - P(M_n \leq u_n) \sim -e^{-K} R_n$$

if $\rho = 0$. (Here $A \sim B$ has the standard meaning that $A = B(1+o(1))$.) In particular this shows that the bound in (1.1) is of the right order, at least in these cases.

It is often instructive to consider $u$ as a "level" and to study the exceedances of the level by $\{\xi_t\}$ - the connection with the maximum of course being that $M_n$ is less than $u$ if and only if there are no exceedances of $u$ by $\xi_1, \ldots, \xi_n$. More generally, we will consider the time-normalized point processes of exceedances of $r$ levels $u^{(1)}_{\geq \ldots \geq u^{(r)}}$ by $\xi_1, \ldots, \xi_n$, defined for $j = 1, \ldots, r$ as $N_{n}^{(j)}(B) = \#\{t \geq 1; t/n \in B, \xi_t > u^{(j)}\}$ for Borel sets $B \subset [0,1]$, where $\#\{\ldots\}$ is the cardinality of the set within brackets. The reader is referred to e.g. Kallenberg (1976) or to the appendix of Leadbetter, Lindgren & Rootzén (1979) for definitions and information about point processes. Further, we will write $N_n = (N_n^{(1)}, \ldots, N_n^{(r)})$ and will consider it as a random variable in the appropriate product space. The second main result (Theorem 4.1) is a representation theorem for $N_n$. Let $\hat{N}_n$ be defined from
\( \{\xi_t\} \) in the same way as \( N_n \) is defined from \( \{\xi_t\} \). Then we show, using an idea of Serfling (1976), that there exist versions of \( N_n \) and \( \hat{N}_n \) such that

\[
P(N_n + \hat{N}_n) \leq 16r^2 R_n,
\]

and similarly \( N_n \) is approximated by a vector of "successively more severely thinned Poisson processes". (It may deserve mention that this approach seems potentially useful also in connection with other problems than the one studied here.) One easy corollary concerns \( M_n^{(k)} \), the \( k \)-th largest among \( \xi_1, \ldots, \xi_n \) and is that

\[
\sup_{u(1) \geq \ldots \geq u(r) \geq u_n} |P(M_n \leq u(1), \ldots, M_n \leq u(r)) - P(M_n^{(1)} \leq u(1), \ldots, M_n^{(r)} \leq u(r))| \leq 16r^2 R_n,
\]

where of course \( M_n^{(k)} \) is the \( k \)-th largest among \( \hat{\xi}_1, \ldots, \hat{\xi}_n \).

Much of the interest in extremes of normal sequences has been centered on the double exponential limit of the distribution of \( M_n \), i.e. that, for \( a_n = (2 \log n)^{1/2}, b_n = a_n - \log \log n + \log 4\pi / (2a_n) \),

\[
P(a_n (M_n - b_n) \leq x) \to \exp(-e^{-x}), \quad n \to \infty,
\]

which is known to hold if \( r_n \log n \to 0 \), or in even more general circumstances, see Leadbetter, Lindgren & Rootzén (1978). The relation (1.3) can, to emphasize the connection with (1.1), be written as

\[
P(M_n \leq u_n) \to \exp(-e^{-x}), \quad n \to \infty,
\]

for \( u_n = u_n(x) = x/a_n + b_n \), and furthermore the same result holds also if \( a_n \) and \( b_n \) are replaced by different constants \( a'_n, b'_n \), pro-
vided \( a'_n/a_n + 1 \) and \( a_n (b'_n - b_n) \to 0 \). As was noted already by Fisher & Tippet (1928) the convergence in (1.3) is extremely slow in the independent case. This was made precise in Hall (1979) in the following way: if \( a_n \) and \( b_n \) are as above, then

\[
P(a_n (\hat{M}_n - b_n) \leq x) - \exp\{-e^{-x}\} \sim \frac{1}{16} e^{-x} \exp\{-e^{-x}\} \left(\frac{\log \log n}{\log n}\right)^2,
\]

while if \( a'_n = 1/b'_n \) and \( b'_n \) is chosen to be the solution of

\[
2\pi b'_n \exp(b'_n^2) = n^2
\]

then

\[
\frac{1}{\log n} \leq \sup_x |P(a'_n (\hat{M}_n - b'_n) \leq x) - \exp\{-e^{-x}\}| \leq C_2/\log n,
\]

for some constants \( 0 < C_1 < C_2 \geq 3 \) and \( n \geq 3 \). Further, Hall shows that the rate \( 1/\log n \) cannot be improved on by choosing \( a'_n, b'_n \) differently. This rate of convergence seems unfortunate if one e.g. in a statistical analysis wants to approximate the distribution of \( M_n \) by the limiting double exponential distribution. However, from a computational point of view it does not pose any problems, since of course \( P(M_n \leq u) = \phi(u)^n \) is quite simple to evaluate directly.

It is easily seen, by combining (1.1) with Hall's results (1.4), (1.5) that for dependent sequences the rate of convergence in (1.3), under appropriate conditions, is of the order \( (\log \log n)^2/\log n \) or \( 1/\log n \), i.e. equally slow as for independent sequences.

For dependent sequences the quantity \( P(M_n \leq u) \), however, is more difficult to evaluate, and perhaps the most interesting consequence of (1.1) is that it demonstrates that the approximation of \( P(M_n \leq u) \) by \( P(\hat{M}_n \leq u) = \phi(u)^n \) is reasonably accurate, at least when the maximal covariance \( \rho \) is not too close to one. Similarly, (1.2) measures how well quite complicated probabilities, concerning the
point processes, can be approximated by assuming independence.

The organization of the paper is as follows. Section 2 contains some notation and three "technical" lemmas in which most of the necessary estimates are proved. In Section 3 the elementary case, the speed of convergence of the distribution of the maximum, is treated in a fairly complete way. In the next section, Section 4, the representation theorem for the point processes of exceedances is established together with some corollaries, and finally, Section 5 contains a short discussion of possible avenues for finding improved approximations of the probabilities of interest.
2. Technical preliminaries

The estimates of this section contain some fairly involved constants, which for easy reference we will collect here. With notation as in the introduction, let

\[ \rho = \sup\{0, r_1, r_2, \ldots\} \]

and, in case \( \rho > 0 \), let \( \nu \) be the number of \( t \)'s such that \( r_t = \rho \).

We will throughout, without further comment, assume that the supremum is attained so that \( \nu \geq 1 \). In particular this is the case if \( r_t \to 0 \) as \( t \to \infty \), and then also \( \nu < \infty \). If \( \rho = 0 \) let \( \nu \leq \infty \) be the number of non-zero \( r_t \)'s. For the second order terms, define \( \rho' \) to be the supremum for \( t \geq 1 \) of the \( r_t \)'s which satisfy \( r_t \neq \rho \), if this quantity is positive, and zero otherwise, and let

\[ \varepsilon = 2\left(\frac{1}{1+\rho'} - \frac{1}{1+\rho}\right) = \frac{2(\rho-\rho')}{(1+\rho)(1+\rho')} . \]

Next, for \( \rho \neq 0 \), define

\[ c'(\rho) = \frac{(1+\rho)^{3/2}}{(1-\rho)^{1/2}} , \quad c''(\rho) = \frac{(2-\rho)(1+\rho)}{1-\rho} , \quad c(\rho) = c'(\rho)(4\pi)^{-\rho/(1+\rho)} \]

and put \( \delta = \sup\{|r_1|, |r_2|, \ldots\} \). The main factor, \( R_n \), in the bounds has a slightly different appearance in the three cases (i) \( \rho > 0 \), (ii) \( \rho = 0 \), \( \nu = \infty \), and (iii) \( \rho = 0 \), \( \nu < \infty \), and in addition depends on a constant \( K \), which will be introduced below,

\[ R_n = c(\rho)K^{2/(1+\rho)} \left(\frac{1}{n}\right)^{1-\rho}/(1+\rho) \left(\log n/K\right)^{-\rho/(1+\rho)} \{\nu + r_n\} \]

where \( r_n = 16K^{\varepsilon} (1-\delta^2)^{-1/2} \sum_{t=1}^{n} |r_t| (\frac{1}{n}) \varepsilon (\log n/K)^{1+\varepsilon/2} \), if \( \rho > 0 \),
(ii) \( R_n = 4K^2(1 - \delta^2)^{-1/2} \frac{1}{n} \log n \sum_{t=0}^{n} |r_t| \), if \( \rho = 0, \nu = \infty \)

(iii) \( R_n = K^2 \frac{1}{n} \nu \), if \( \rho = 0, \nu < \infty \).

As a starting point for the estimates we will use the important identity

\[
(2.4) \quad P(M_n \leq u) - P(M_n \leq u') = \sum_{1 \leq s < t \leq n} r_{s-t} \int_{-\infty}^{u} \int_{-\infty}^{u'} f_h(x_s = x_t = u) \, dx' \, dh,
\]

where \( f_h(x_s = x_t = u) \) is the function of \( n - 2 \) variables which is obtained by putting \( x_s = x_t = u \) in the density function of \( n \) stationary normal random variables with zero means, unit variances, and covariances \( h_{r_t} \), and where the "primes" signify that \( x_s \) and \( x_t \) are deleted from the integrations, the intervals of integration each being \((-\infty, u] \). The equation (2.4) is due, in various ways, to Slepian (1962), Berman (1964), and Cramér, and for a derivation of it see e.g. Leadbetter, Lindgren & Rootzén (1979), p. 45-47. It is useful to write it in a slightly different form.

Let

\[
\psi_r(u) = \frac{1}{2\pi(1-r^2)^{1/2}} \exp\left( -\frac{1}{2(1-r^2)} (u^2 - 2ru + u^2) \right)
\]

be the joint density of two standard normal variables with correlation \( r \), evaluated at the point \((u, u)\), and let

\[
f_h(x' | x_s = x_t = u) = f_h(x_s = x_t = u) / \psi_{hr_{s-t}}(u)
\]

be the conditional density, given that the \( s \)-th and \( t \)-th variables equal \( u \), in a \( n \)-dimensional normal distribution with zero means, unit variances and covariances \( h_{r_t} \). The identity (2.4) can then
be written as

$$P(M_n < u) - P(M_n < u) = \sum_{1 \leq s \leq t \leq n} r_{s-t} \int_{h=0}^{1} \int_{-\infty}^{\infty} f_h(x'|x_s = x_t = u) dx' \, dh.$$ 

Since \( f_h(x'|x_s = x_t = u) \) is a density function it is clear that

$$0 \leq \int_{-\infty}^{\infty} f_h(x'|x_s = x_t = u) dx' \leq 1.$$ 

The main proofs use the right-hand inequality in (2.6) to estimate the expression in (2.5). However, we will also see that often not much is lost by this.

**Lemma 2.1** Let \( u > 0 \), suppose \( r > 0 \), \(|r| < 1\), write \( \rho = \max\{0, r\} \) and let \( c, c' \) be given by (2.2). Then

(i) \[ \frac{1}{2\pi u^2 r} \left\{ c'(r) e^{-u^2/(1+r)} - e^{-u^2} \right\} / \{1 + \frac{c''(r)}{u^2}\} \leq \int_0^1 \psi_{hr}(u) \, dh \]

Suppose that furthermore \( u > 1 \). Then

(ii) \[ 0 \leq \int_0^1 \psi_{hr}(u) \, dh \]

\[ \leq 2^{2+\rho)/(1+\rho)} c(\rho) |r|^{-1} (1 - \phi(u)) / u^\rho \]

and, if \( r \leq \rho' \) for some constant \( 0 \leq \rho' < 1 \), then

(iii) \[ \int_0^1 \psi_{hr}(u) \, dh \]

\[ \leq 2^{2+\rho')/(1+\rho')} (4\pi)^{-\rho'/(1+\rho')} (1-r^2)^{-1/2} (1 - \phi(u)) / u^{2/(1+p')} \]
Proof By partial integration

\[(2.7) \quad 2\pi \int_0^1 \psi_{hr}(u) \, du = \int_0^1 (1 - h^2 r^2)^{-1/2} e^{-u^2/(1 + hr)} \, du = 1 \int_0^1 (2 - hr)(1 + hr)^{1/2} e^{-u^2/(1 + hr)} \, du , \]

and the second inequality in (i) follows at once, since the last integral in (2.7) is positive. Moreover, \((2 - hr)(1 + hr)^{1/2}(1 - hr)^{-3/2} \leq c''(\rho)(1 - h^2 r^2)^{-1/2}\), as is easily checked, and hence

\[(2.8) \quad 1 \int_0^1 (2 - hr)(1 + hr)^{1/2} e^{-u^2/(1 + hr)} \, du \leq 2\pi c''(\rho) \int_0^1 \psi_{hr}(u) \, du . \]

Inserting (2.8) into (2.7) we obtain that

\[\int_0^1 \psi_{hr}(u) \, du \{1 + c''(\rho)\} \geq \frac{1}{2\pi u^2 r} \{(c'(r)e^{-u^2/(1+r)} - e^{-u^2})\} , \]

which proves the first inequality in (i).

To prove (ii) we will use that

\[(2.9) \quad \sqrt{2\pi}(1 - \Phi(u)) > \exp\{-u^2/2\} \frac{u^2}{1+u^2} \geq \exp\{-u^2/2\} \]

for \(u \geq 1\). Thus, if \(r = \rho > 0\), by part (i),

\[\int_0^1 \psi_{hr}(u) \, du \leq \frac{c'(r)}{2\pi u^2 r} e^{-u^2/(1+r)} \]

\[\leq 2(2+\rho)/(1+\rho) (4\pi)^{-\rho/(1+\rho)} |r|^{-1} c'(\rho) [(1 - \Phi(u))/u^\rho]^2/(1+\rho) , \]

and similarly, for \(r < 0\), \(\rho = 0\),

\[\int_0^1 \psi_{hr}(u) \, du \leq \frac{1}{2\pi u^2 |r|} e^{-u^2/|r|} \leq \frac{4c'(0)}{|r|} (1 - \Phi(u))^2 , \]

and hence (ii) holds in either case. Finally, it is immediate that, for \(u \geq 1\),
\[
\frac{1}{2\pi(1-r^2)^{1/2}} \leq \frac{1}{2\pi(1-r^2)^{1/2}} e^{-u^2/(1+\rho')}
\]
\[
\leq 2^{(2+\rho')/(1+\rho')}(4\pi)^{-\rho'/(1+\rho')}\left(1-r^2\right)^{-1/2}(1-\Phi(u))^{-2/(1+\rho')},
\]
by (2.9), which proves (iii).

The main lemma now follows easily. In it we will only consider a restricted range of \(u\)-values (which may even be empty for small \(n\)). The remaining range of \(u\)'s of interest to us is easier to treat, as shown in the proof of Theorem 3.1 below.

**Lemma 2.2** Suppose that for some constant \(K > 0\),
\[
(2.10) \quad n(1-\Phi(u)) \leq K
\]
and that \(1 \leq u \leq 2(1+\rho)^{-1/2}(\log n/K)^{1/2}\). Then
\[
(2.11) \quad \sum_{t=1}^{n} |r_t| \int_{0}^{1} \psi_{hr_t}(u) \, dh \leq 4R_n,
\]
for \(R_n\) given by (2.3) (i), (ii), and (iii), respectively, for \(\rho > 0\), for \(\rho = 0, \nu = \infty\), and for \(\rho = 0, \nu < \infty\).

**Proof** First, by (2.9) and (2.10),
\[
\frac{n}{\sqrt{2\pi}} \frac{e^{-u^2/2}}{2u} \leq K,
\]
i.e.
\[
(2.12) \quad \log n/K \leq \frac{u^2}{2\left(1 + \frac{1}{u^2}\log 8\pi u^2\right)} \leq 2u^2,
\]
for \(u \geq 1\).

Now, suppose that \(\rho > 0\). Using Lemma 2.1 (ii) to bound summands with \(r_t = \rho\) and Lemma 2.1 (iii) to bound the remaining summands we have that
(2.13)  
\[ \sum_{t=1}^{n} |r_t| \int_{\psi t} \phi(u) \, du \leq n2^{2/(1+\rho)'/(1+\rho)} c(\rho) \frac{2/(1+\rho)}{(1+\rho)} \frac{2/(1+\rho)}{(1+\rho)} \sum_{t=1}^{n} \left( (1 - \phi(u)) u \right)^{2/(1+\rho)} \]

\[ + n2^{2/(1+\rho)'/(1+\rho)}(4\pi)^{-\rho'/2/(1+\rho')}_c \left( \frac{|r_t|}{1-r_t^2} \right)^{1/2} \left( (1 - \phi(u)) u \right)^{2/(1+\rho')} \]

where \( \Sigma' \) denotes summation over all \( t \in \{1, \ldots, n\} \) such that \( r_t \neq \phi \).

Since \( n(1 - \phi(u)) \leq K \) and \( (1/2 \log n/K) \leq u \leq 2(1 + \rho)^{-1/2}(\log n/K)^{1/2} \)

by assumption and (2.12), we have that

\[ \left( (1 - \phi(u)) u \right)^{2/(1+\rho)} \leq 2^{\rho/(1+\rho)} \left( \frac{K}{n} \right)^{2/(1+\rho)} \]

and that

\[ \left( (1 - \phi(u)) u \right)^{2/(1+\rho')} \leq 2^{\rho/(1+\rho')} \left( \frac{K}{n} \right)^{2/(1+\rho')} \]

Inserting this into (2.13) we obtain, with \( \delta = \sup \{|r_1|, |r_2|, \ldots\} \)

\[ \sum_{t=1}^{n} |r_t| \int_{\psi t} \phi(u) \, du \leq 4 c(\rho) K^{2/(1+\rho)} \left( 1 - \phi(u) \right)^{(1+\rho)/(1+\rho)} (1+\rho') (1+\rho') (\log n/K)^{-\rho/(1+\rho')} \]

and comparing with (2.3) (i), this proves (2.11) for the case \( \rho > 0 \).

Next, suppose \( \rho = 0 \). Then, using Lemma 2.1 (iii), similar calculations show that

\[ \sum_{t=1}^{n} |r_t| \int_{\psi t} \phi(u) \, du \leq 4 \sum_{t=1}^{n} \left( (1 - \phi(u)) u \right)^{2/(1+\rho)} \]

\[ \leq 16K^2 (1 - \delta^2)^{-1/2} \sum_{t=1}^{n} |r_t| \frac{1}{n} \log n/K \]

proving (2.11) for the case $p = 0, v = \infty$.

Finally, suppose $p = 0, v < \infty$, so that by Lemma 2.1 (ii)

$$
\sum_{t=1}^{n} \frac{1}{|r_{t}|} \psi_{r_{t}} (u) \leq 4 \sum_{t=1}^{n} \frac{|r_{t}|}{|r_{t}|} (1 - \phi(u))^{2} \leq 4K^{2} \frac{1}{n},
$$

which shows that (2.11) holds also in this case. $\Box$

Clearly (2.5) and Lemma 2.2 together will provide a bound for

$$|P(M_{n} \leq u) - P(M_{n} \leq u)|.$$ However, for the point processes of exceedances, some further estimates are needed. Let, as in the introduction, $u^{(1)} \geq \ldots \geq u^{(r)}$ be $r$ levels and define $I_{t}^{(i)} = I\{\xi_{t} > u^{(i)}\}$ where $I$ is the indicator function, i.e. $I_{t}^{(i)}$ is one if $\xi_{t} > u^{(i)}$ and zero otherwise. Further, let $\mathcal{B}_{0}$ be the trivial $\sigma$-algebra, and for $t \geq 1$ let $\mathcal{B}_{t} = \sigma(I_{s}^{(i)}; 1 \leq i \leq r, 1 \leq s \leq t)$ be the $\sigma$-algebra generated by the exceedances up to time $t$.

**Lemma 2.3** (i)

$$
\sup_{B \in \mathcal{B}_{t-1}} |P(I_{t}^{(i)} = 0)B) - P(I_{t}^{(i)} = 0)P(B)| \leq 2r \sum_{s-t}^{1} |r_{s-t}| \psi_{r_{s-t}} (u^{(r)}) dh.
$$

(ii) Suppose that $u = u^{(r)}$ satisfies the requirements of Lemma 2.2. Then

$$
\sum_{t=1}^{n} \sum_{i=1}^{r} E|P(I_{t}^{(i)} = 0 | B_{t-1}) - P(I_{t}^{(i)} = 0)| \leq 16r^{2} R_{n},
$$

with $R_{n}$ given by (2.3).

**Proof** (i) This follows from an extension of the proof of Lemma 3.2 of Watts, Rootzén & Leadbetter (1980). In fact, let in that proof $\ell'_{n} = 1, \ell_{n} = 0$, let $B \in \mathcal{B}_{t-1}$ and write
(2.14) \[ B = B_0 \{ I_s(\ell) = 0 \} \cup B_1 \{ I_s(\ell) = 1 \} \cup \ldots \cup B_r \{ I_s(\ell) = 0 \} \cup B_{r+1} \{ I_s(\ell) = 1 \} \]

(instead of \( B = B_0 \{ I_n(j) = 0 \} \cup B_1 \{ I_n(j) = 1 \} \) in the cited proof), where each of \( B_0, \ldots, B_{r+1} \) is a disjoint union of sets of the form \( I_\ell(j) = x_\ell j \) where each \( x_\ell j \) is zero or one and the intersection is over \( j = 1, \ldots, s \) and \( \ell = 1, \ldots, s-1, s+1, \ldots, t-1 \). Proceeding as in the cited reference, each term \( B_{j+k} \{ I_s(j) = k \} \) leads to a term

(2.15) \[ \int_{-\infty}^{\infty} \ldots \int f_h(x_s = u(j), x_t = u(i)) dx \]

in the estimation of the quantity \( F'(h) \) defined there, where \( f_h(x_s = u(j), x_t = u(i)) \) is the function of \( t-2 \) variables which is obtained by putting \( x_s = u(j), x_t = u(i) \) in the density function of \( t \) stationary normal random variables with zero means, unit variances and covariances \( h_{s-t} \). Now, (2.15) is just the density of two standard normal variables, with correlation \( h_{s-t} \) evaluated at \( (u(j), u(i)) \), and may easily be shown to be bounded by \( \psi_{h_{s-t}}(u(r)) \). Part (i) then follows at once, since there are \( 2r \) terms in (2.14) and since by construction \( \int_0^1 F'(h) dh \) is equal to \( P(\{ I_t^{(i)} = 0 \} B) - P(\{ I_t^{(i)} = 0 \} B) \).

(ii) This follows easily if we show that

\[ E(\text{P}(I_t^{(i)} = 0 | B_{t-1}) - P(I_t^{(i)} = 0)) \]

\[ \leq 2 \sup_{B \in B_{t-1}} |P(\{ I_t^{(i)} = 0 \} B) - P(I_t^{(i)} = 0) P(B) | \]

since then, by part (i) and Lemma 2.2,
However, for $B = \{ P(I_t^{(i)} = 0 \| B_{t-1}) > P(I_t^{(i)} = 0) \} \in \mathcal{B}_{t-1}$, by standard calculations

$$\sum_{t=1}^{n} \sum_{i=1}^{r} \mathbb{E}|P(I_t^{(i)} = 0 \| B_{t-1}) - P(I_t^{(i)} = 0)| \leq 2r \sum_{t=1}^{n} 2r \sum_{1 \leq s < t} 1 \int_{0}^{r_s-t} \psi_{r_s-t}(u(r)) \, dh$$

$$\leq 4r^2n \sum_{t=1}^{n} r_t \int_{0}^{r_t} \psi_{r_t}(u(r)) \, dh$$

$$\leq 16r^2 R_n .$$

which completes the proof of part (ii).
3. The rate of convergence of the maximum

The rate of convergence to zero of $P(M_n \leq u) - P(\hat{M}_n \leq u)$ now follows easily. To obtain efficient bounds we will, as in Lemma 2.2, restrict the domain of variation of $u$ by requiring that $n(1 - \Phi(u)) \leq K$, for some fixed $K > 0$, or equivalently that $u \geq u_n$, where $u_n$ is the solution to the equation $n(1 - \Phi(u_n)) = K$. Since

$$P(\hat{M}_n \leq u_n) = \Phi(u_n)^n = \{1 - \frac{1}{n}(n(1 - \Phi(u_n)))\}^n,$$

this clearly implies that

$$P(\hat{M}_n \leq u_n) \to e^{-K}, \quad n \to \infty,$$

and conversely, if (3.1) holds, then $n(1 - \Phi(u_n)) \to K$, as is easily seen. Moreover, if $P(M_n \leq u_n) - P(\hat{M}_n \leq u_n) \to 0$, then of course the same equivalence holds for $\hat{M}_n$ replaced by $M_n$.

Thus, since the bounds for the rate of convergence will be proved for $u \geq u_n$, they will apply to the upper part of the range of variation of $P(M_n \leq u)$, and by taking $K$ large an arbitrarily large part of this range is covered, but at the cost of a poorer bound.

Theorem 3.1 Let $\{\xi_t\}$ be stationary normal, with zero means, unit variances and covariances $r_t = E \xi_s \xi_{s+t}$. Let $\rho = \sup \{0, r_1, r_2, \ldots\}$ and let $v$ be the number of $r_t$'s, $t \geq 1$, with $r_t = \rho$, in case $\rho > 0$, and let $v$ be the number of non-zero $r_t$'s, for $t \geq 1$, otherwise. Further, let $\delta = \sup \{|r_1|, |r_2|, \ldots\}$ and let $\epsilon > 0$, $c$, $c'$, and $R_n$ be as defined in (2.1) - (2.3). Suppose that $u \geq 1$, and that

$$n(1 - \Phi(u)) \leq K,$$

for some constant $K$, with $n/K \geq \epsilon$. Then
(3.3) \[ |\mathbb{P}(M_n \leq u) - \mathbb{P}(\hat{M}_n \leq u)| \leq 4R_n, \]
or, more explicitly, writing \( \Delta_n = |\mathbb{P}(M_n \leq u) - \mathbb{P}(\hat{M}_n \leq u)| \), if \( \rho > 0 \) then
\[
\Delta_n \leq 4c'' \frac{(\rho)K^2/(1+\rho)}{n} \left\{ \frac{1}{n} (1-\rho)/(1+\rho) (\log n/K)^{-\rho/(1+\rho)} \right\} \{v + r_n\},
\]
with \( r_n \) given by (2.3,1), if \( \rho = 0 \) then
\[
\Delta_n \leq 16K^2 (1 - \delta^2)^{-1/2} \frac{1}{n} \log n/K \sum_{t=0}^{n} |r_t|,
\]
and if in addition \( v < \infty \) then
\[
\Delta_n \leq 4K^2 \frac{1}{n} \nu.
\]

**Proof** By (2.5) and (2.6)
\[
|\mathbb{P}(M_n \leq u) - \mathbb{P}(\hat{M}_n \leq u)| \leq \sum_{1 \leq s < t \leq n} |r_s-t| \int_{0}^{\psi r_{s-t}} (u)dh
\]
\[
\leq n \sum_{t=1}^{n} |r_t| \int_{0}^{\psi r_t} (u)dh,
\]
and it follows from Lemma 2.2 that (3.3) holds for \( u \) satisfying (3.2) and \( 1 \leq u \leq 2(1+\rho)^{-1/2} (\log n/K)^{1/2} \).

To complete the proof we will show that (3.3), rather trivially, is satisfied also for \( u > 2(1+\rho)^{-1/2} (\log n/K)^{1/2} \). In fact
\[
(3.4) \quad |\mathbb{P}(M_n \leq u) - \mathbb{P}(\hat{M}_n \leq u)| = |\mathbb{P}(M_n > u) - \mathbb{P}(\hat{M}_n > u)|
\]
\[
\leq \mathbb{P}(M_n > u) + \mathbb{P}(\hat{M}_n > u)
\]
\[
\leq 2n(1 - \phi(u)),
\]
by Boole's inequality. Since \( 1 - \phi(u) \leq (2\pi)^{-1/2} e^{u^2/2}/u \), we have for \( u \geq 2(1+\rho)^{-1/2} (\log n/K)^{1/2} \) that
\[ 1 - \Phi(u) \leq \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(1+p)^{-1/2}(\log n/K)^{1/2}\right\} (\log n/K)^{-1/2} \leq \frac{1}{\pi \sqrt{n}} \frac{1}{\log n} \exp\left\{-\frac{1}{2}(1+p)^{-1/2}(\log n/K)^{1/2}\right\} , \]

and hence, by (3.4),

\[ |P(M_n \leq u) - P(M_n \leq u)| \leq 4R_n , \]

by straightforward calculation.

As an easy corollary to the theorem we shall prove that an analogue of Hall's result (1.5) holds also for dependent sequences, under appropriate conditions.

**Corollary 3.2** Suppose that \( \{\xi_t\} \) is stationary normal, with zero means, unit variances, and covariances \( \{r_t\} \) such that

\[ (\log n)^2(\log \log n)^2 \left( \frac{1}{n} (1-\rho')/(1+\rho') \frac{1}{n} \sum_{t=1}^{n} |r_t| \right) \to 0 , \]

as \( n \to \infty \), where \( \rho' \) is defined on P. 6

Then for \( a'_n, b'_n, C_1, \) and \( C_2 \) satisfying (1.5)

\[ 0 < C_1 \leq \lim \inf_{n \to \infty} \sup_{x} \exp\{-e^{-x}\} \]

\[ \leq \lim \sup_{n \to \infty} \sup_{x} \exp\{-e^{-x}\} \leq C_2 \]

and the order \( 1/\log n \) of convergence cannot be improved by choosing other norming constants than \( a'_n, b'_n \). In particular, for

\( a_n = (2 \log n)^{1/2}, \ b_n = a_n - \{\log \log n - \log 4\pi\}/(2a_n) \),
\[ P(\alpha_n(M_n - b_n) \leq x) - \exp\{-e^{-x}\} \approx \frac{1}{16} e^{-x} \exp\{-e^{-x}\} \left(\log \log n\right)^2. \]

Proof By (1.5) and (1.4) it is sufficient to prove that

\[ \sup_{-\infty < u < \infty} |P(M_n \leq u) - P(\hat{M}_n \leq u)| = o(1/\log n). \]

We first note that (3.5) implies that \( \delta = \sup\{|r_1|, |r_2|, \ldots\} < 1 \) (since otherwise \( r_t \) would be periodic, which contradicts (3.5)). Now, it is straightforward to check that if the constant \( K \) in the bound \( R_n \) is chosen as \( K = K_n = 2 \log \log n \), and if (3.5) holds, then \( R_n = o(1/\log n) \), so that

\[ \sup_{u \geq u_n} |P(M_n \leq u) - P(\hat{M}_n \leq u)| = o(1/\log n), \]

for \( u_n \) given by \( n(1 - \phi(u_n)) = K_n = 2 \log \log n \).

Furthermore, for \( u \leq u_n \),

\[ |P(M_n \leq u) - P(\hat{M}_n \leq u)| \leq 2P(\hat{M}_n \leq u_n) + |P(M_n \leq u_n) - P(\hat{M}_n \leq u_n)| \]

\[ = 2\phi(u_n)^n + o(1/\log n), \]

and since

\[ \phi(u_n) = (1 - (1 - \phi(u_n))^n \]

\[ \leq e^{-n(1 - \phi(u_n))} \]

\[ = (\log n)^{-2}, \]

it follows that

\[ \sup_{u \leq u_n} |P(M_n \leq u) - P(\hat{M}_n \leq u)| = o(1/\log n), \]

which together with (3.7) proves (3.6).

From Theorem 3.1 follows that, supposing \( \Sigma_{t=1}^{n} |r_t| \) does not grow too rapidly, if \( \rho > 0 \) then the rate of convergence is at least of
the order
\[ \begin{align*} &\left(\frac{1}{n}\right)^{1-p/(1+p)} (\log n)^{-\rho/(1+p)}, \\
\end{align*} \]

and that if \( \rho = 0, \nu < \infty \), then the rate is at least of the order

\[ \frac{1}{n}. \]

We will now find a precise asymptotic expression for \( P(M_n \leq u) - P(\hat{M}_n \leq u) \) in the case when \( r_t = 0 \) if \( |t| > m \), for some constant \( m < \infty \), i.e. when the sequence is \( m \)-dependent. This will show that, at least for such sequences, these rates are of the right order.

If \( \rho = 0, \nu = \infty \), and \( \sum_{t=0}^{\infty} |r_t| < \infty \), the bound given by Theorem 3.1 is of the order

\[ \frac{1}{n} \log n. \]

It seems unlikely that this is the correct order, but the loss does not seem important, since clearly the rate of convergence cannot be better than \( 1/n \), in general.

**Theorem 3.3** Suppose \( \{\xi_t\} \) is stationary normal, with zero means, unit variances and covariances \( \{r_t\} \) such that \( r_t = 0 \) for \( |t| > m \), for some constant \( m < \infty \), and that

\[ n(1 - \Phi(u_n)) \to K > 0, \quad \text{as } n \to \infty. \]

Then, if \( \rho > 0 \)

\[ (3.8) \]

\[ P(M_n \leq u_n) - P(\hat{M}_n \leq u_n) \sim e^{-K} C(\rho) K^{2/(1+\rho)} \left( \frac{1}{n} \right)^{1-p/(1+p)} (\log n)^{-\rho/(1+p)} \nu \]

\[ \sim e^{-K} R_n, \]

\( K_n \)
and if \( \rho = 0 \), then

\[
(3.9) \quad P(M_n \leq u_n) - P(M_n \leq u_n) \sim -e^{-K} \frac{1}{n} \sqrt{r} e^{-K} n.
\]

**Proof** The essential part of the proof consists of a closer evaluation of the quantity

\[
\int_{-\infty}^{u'} \int_{-\infty}^{u'} f_h(x'|x_t = x_t = u) dx' = P, \quad \text{say},
\]

which was estimated by one in (2.6), and in the proof of Theorem 3.1. For this it is convenient to introduce a further stationary normal sequence, \( \{ \tilde{\xi}_t \} \) say, with means zero, variances one and covariance function \( h_{r_t} \). Let

\[
\tilde{M}_n = \max_{1 \leq t \leq n} \tilde{\xi}_t, \quad \tilde{M}_I = \max_{t \in I} \tilde{\xi}_t
\]

so that, for \( u = u_n \),

\[
P = P(\tilde{M}_n \leq u_n \mid \tilde{\xi}_s = \tilde{\xi}_t = u_n),
\]

and let \( I = \{ k \in [1,n]; |k-s| \leq m \text{ or } |k-t| \leq m \} \) and \( J = \{ 1, \ldots , n \} \cap I^c \). By Boole's inequality

\[
(3.10) \quad P(\tilde{M}_J \leq u_n \mid \tilde{\xi}_s = \tilde{\xi}_t = u_n) - \Sigma_{k \in I} P(\tilde{\xi}_k > u_n \mid \tilde{\xi}_s = \tilde{\xi}_t = u_n) \leq P
\]

\[
= P(\tilde{M}_J \leq u_n \mid \tilde{\xi}_s = \tilde{\xi}_t = u_n).
\]

Since \( \{ \tilde{\xi}_t \} \) is \( m \)-dependent, \( P(\tilde{M}_J \leq u_n \mid \tilde{\xi}_s = \tilde{\xi}_t = u_n) = P(\tilde{M}_J \leq u_n) \) and thus by a similar calculation

\[
P(\tilde{M}_n \leq u_n) \leq P(\tilde{M}_J \leq u_n \mid \tilde{\xi}_s = \tilde{\xi}_t = u_n)
\]

\[
\leq P(\tilde{M}_n \leq u_n) + \Sigma_{k \in I} P(\tilde{\xi}_k > u_n).
\]

Now, since \( \Sigma_{k \in I} P(\tilde{\xi}_k > u_n) \leq 4m(1 - \phi(u_n)) + 0 \), and since, by Theorem 3.1, \( P(\tilde{M}_n \leq u_n) \sim P(M_n \leq u_n) = \phi(u_n)^n e^{-K}, n \rightarrow \infty \), it follows that
uniformly in $s, t, \text{and } h$.

Next, given that $\xi_s = \xi_t = u_n$, $\xi_k$ is normal with variance not exceeding one and mean $u_n (r_k s - r_k t)/(1 + hr_{s-t})$. We will temporarily assume that

$$0 < \varepsilon \leq 1 - \max_{k \neq s, t} h (r_k s - r_k t)/(1 + hr_{s-t}) \tag{3.12}$$

for some constant $\varepsilon$ which does not depend on $k, s, t, \text{or } h$. Then

$$\sum_{k \in I} P (\xi_k > u_n \mid \xi_s = \xi_t = u_n) \leq 4m (1 - \Phi (\varepsilon u_n)) \to 0, \quad \text{as } n \to \infty,$$

and by (3.10) and (3.11),

$$P = \int_{-\infty}^{u_n} f_\varepsilon (x^\prime | x_s = x_t = u_n) \, dx' \to e^{-K}, \tag{3.13}$$

as $n \to \infty$, uniformly in $s, t, \text{and } h$.

Now, if $\rho = 0$ then (3.12) is satisfied, with $\varepsilon = 1$, and hence, by (2.5) and (3.13),

$$P(M_n \leq u_n) - P(M_n \leq u_n) = \sum_{1 \leq s < t \leq n} e^{-K} \psi_{r_s-t} (u) \int_0^{hr_{s-t}} e^{-\phi (\varepsilon u_n)} \, dh \tag{3.14}$$

$$\approx e^{-K} \frac{1}{n} \sum_{t=1}^{n-1} e^{-\phi (\varepsilon u_n)} \, dh$$

$$\approx e^{-K} \frac{1}{n^2} \int_0^{hr_t} e^{-\phi (\varepsilon u_n)} \, dh$$

$$\approx e^{-K} \frac{1}{n^2} e^{-\varepsilon^2 u_n^2}$$

$$\approx -e^{-K} \frac{1}{n} \varepsilon u_n^2.$$
where we have used Lemma 2.1 (i) in the fourth step and that 
\[(2\pi)^{-1/2} e^{-u_n^2/2}/u_n^2 \sim 1 - \phi(u_n) \sim K/n \] in the last step.

This proves (3.9), and we next suppose that \( \rho > 0 \) and let \( \Sigma'' \) denote the sum over \( s, t \) such that \( 1 \leq s < t \leq n \) and \( r_{s-t} = \rho \). In the same way as in (3.14) we then have that
\[
\sum_{s-t} \int \psi_{hr_{s-t}}(u_n) \int f_h(x'|x_s = x_t = u_n) dx' dh \\
\quad \quad \quad \quad \sim e^{-K c(\rho)(\frac{1}{n})(1-\rho)/(1+\rho)} (\log n)^{-\rho/(1+\rho)} \nu,
\]
since \( u_n \sim \sqrt{2 \log n} \), and since for \( s, t \) such that \( r_{s-t} = \rho \) the condition (3.12) is clearly satisfied for some suitable \( \varepsilon > 0 \). Since the sum of the remaining terms is \( o(n^{-(1-\rho)/(1+\rho)} (\log n)^{-\rho/(1+\rho)}) \), as was seen in the proof of Lemma 2.2, this shows that
\[
\sum_{s-t} \int \psi_{hr_{s-t}}(u_n) \int f_h(x'|x_s = x_t = u_n) dx' dh \\
\quad \quad \quad \quad \sim e^{-K c(\rho)(\frac{1}{n})(1-\rho)/(1+\rho)} (\log n)^{-\rho/(1+\rho)} \nu,
\]
and hence by (2.5) that (3.8) holds.

Comparing the asymptotic expressions for \( P(M_n \leq u_n) - P(M_n \leq u_n) \) with the bounds of Theorem 3.1 we see that the bounds asymptotically are too large by a factor \( 4 e^K \). Here the factor 4 is due to inaccuracies in the estimates (2.9) and (2.12) and could easily be reduced by restricting the range of \( u \) further. The factor \( e^K \) is due to the estimate (2.6), as was seen in the proof of Theorem 3.3, and could conceivably be reduced along similar lines as in that proof, but perhaps at the expense of a considerable increase in the complexity of proofs.
4. A representation for the point processes of exceedances

Let \( N_n = (N_n^{(1)}, \ldots, N_n^{(r)}) \) be the vector of time-normalized exceedances of the levels \( u^{(1)} \geq \ldots \geq u^{(r)} \), i.e. \( N_n^{(i)}(B) = \{ t; \xi_t > u^{(i)} \}, \) for any Borel set \( B \subseteq [0,1] \), and let \( \hat{N}_n = (\hat{N}_n^{(1)}, \ldots, \hat{N}_n^{(r)}) \) be defined similarly, with \( \{\xi_t\} \) replaced by the associated independent sequence \( \{\hat{\xi}_t\} \). It is known, see Leadbetter, Lindgren & Rootzén (1979), that under weak conditions (the same as those commonly used to establish (1.3)) \( N_n \) converges in distribution to a certain successively more severely thinned Poisson process (which will be described below). To formulate results about the rate of convergence of the distribution of \( N_n \), and, more generally, to find useful ways to measure the distance between the distributions of two point processes seems to be an interesting and non-trivial question, but here we will partly circumvent this issue by using a "representation" approach. More precisely, we will construct two processes which have the same distribution as \( N_n \) and \( \hat{N}_n \), respectively, and whose realizations are identical with high probability. Following common usage, we will refer to these processes as versions of \( N_n \) and \( \hat{N}_n \) and, since it does not lead to any confusion, we will use the same letter to denote processes which are versions of one another.

The limiting process \( N = (N^{(1)}, \ldots, N^{(r)}) \) can be described in the following way. Let \( 0 < \tau^{(1)} \leq \ldots \leq \tau^{(r)} \) be given parameters, and let \( N^{(r)} \) be a Poisson process in \([0,1]\), with parameter \( \tau^{(r)} \) and points \( \{\sigma_k\} \). Let \( \{\beta_k\} \) be independent random variables, independent also of \( N^{(r)} \) and taking values in \( 1, \ldots, r \) with probabilities

\[
P(\beta_k = i) = \frac{(\tau^{(r-i+1)} - \tau^{(r-i)})/\tau^{(r)}}{\tau^{(1)}/\tau^{(r)}}, \quad i = 1, \ldots, r-1
\]

\[
= \frac{1}{\tau^{(1)}}, \quad i = r.
\]
For each \( k \) such that \( \beta_k > 1 \) let \( N^{(r-1)}, \ldots, N^{(r-\beta_k+1)} \) have points at \( \sigma_k \) to complete the definition of \( N = (N^{(1)}, \ldots, N^{(k)}) \). Thus, in particular, each \( N^{(i)} \) is a Poisson process with intensity \( \tau^{(i)} \), but the dependence between the component processes does not have a Poisson character.

Since, for each \( i \), \( N^{(i)}_n \) is concentrated on the set \( \{1/n, \ldots, n/n\} \) while the probability is zero that \( N^{(i)} \) has a point in \( \{1/n, \ldots, n/n\} \), it is not possible to construct versions of \( N_n \) and \( N \) with realizations which are identical with a probability tending to one. However, such a construction is possible if \( N \) is first discretized as follows: for each \( i \), \( 1 \leq i \leq r \), let \( \hat{N}^{(i)}_n \) be concentrated on \( \{1/n, \ldots, n/n\} \) with \( \hat{N}^{(i)}_n((t/n)) = N^{(i)}_n(((t-1)/n, t/n]) \), \( t = 1, \ldots, n \). Thus \( \hat{N}^n \) is obtained from \( N \) by "discretizing" by placing all the points of \( N \) in the intervals \( ((t-1)/n, t/n] \) at the endpoints \( t/n \) of the intervals.

Theorem 4.1 Let \{\( \xi_t \)\} be stationary normal with zero means, unit variances and covariances \( \tau_s = \text{cov} \xi_s \xi_{s+t} \), let \( R_n \) be given by (2.3) and let \( u^{(1)} \geq \ldots \geq u^{(r)} \geq 1 \).

(i) If

\[
(4.1) \quad n(1 - \phi(u^{(r)})) \leq K,
\]

for some constant \( K \) with \( n/K \geq e \), then there exist versions \( N_n \) and \( \hat{N}_n \) of the vectors of time-normalized point-processes of exceedances of \( u^{(1)} \geq \ldots \geq u^{(r)} \) by \{\( \xi_t \)\} and \{\( \hat{\xi}_t \)\} respectively,

\[
(4.2) \quad P(N_n \neq \hat{N}_n) \leq 16r^2 R_n.
\]

(ii) Let \( N = (N^{(1)}, \ldots, N^{(r)}) \), be the thinned Poisson process described above, with parameters \( \tau^{(1)} \leq \ldots \leq \tau^{(r)} \) and let \( \hat{N}_n \) be obtained from \( N \) by "discretizing". Write \( \tau^{(i)}_n = n(1 - \phi(u^{(i)})) \),
i = 1, ..., r, and suppose that (4.1) is satisfied. Then there exist versions of \( N_n, \hat{N}_n \) such that

\[
P(N_n + \hat{N}_n) \leq 16r^2 n + \sum_{i=1}^{r} n^{-1} |\tau(i) - \tau(i)| + \frac{1}{n} \sum_{i=1}^{r} (\tau(i))^2.
\]

**Proof** (i) We will use the main idea of Serfling (1976) in the proof. Let \((\Omega, F, P)\) be a probability space which supports independent variables \( \eta_1, ..., \eta_n \) which are uniformly distributed on \([0,1]\), write \( p(i) = P(\xi_1 \leq u(i)) = \phi(u(i)), i = 1, ..., r \), and recall the notations \( I_{t(i)} = I\{\xi_t > u(i)\}, \hat{I}_{t(i)} = I\{\hat{\xi}_t > u(i)\}, \) and

\[
E_t = \sigma\{I_{k(i)}; 1 \leq i \leq r, 1 \leq k \leq t\}. \text{ Thus } P_t^{(i)} = P(I_t = 0 \mid E_{t-1}) \text{ is a function, } P_t^{(i)} = P_t^{(i)}(I_t^{(1)}, ..., I_t^{(r)}), \text{ of } I_t^{(1)}, ..., I_t^{(r)}. \text{ We will first use Serfling's construction to define suitable versions of the processes } \{I_t^{(i)}\} \text{ and } \{\hat{I}_t^{(i)}\} \text{ on } (\Omega, F, P). \text{ A version of the latter process is obtained by setting}
\]

\[
\hat{I}_t^{(i)} = I(\eta_t > p(i)), \ i = 1, ..., r, \ t = 1, ..., n,
\]

and further it is readily seen that by defining

\[
I_t^{(i)} = I(\eta_t > p_t^{(i)}) = I(\eta_t > p(i)), \ i = 1, ..., r,
\]

and then recursively, for \( t = 2, ..., n, \ i = 1, ..., r, \)

\[
I_t^{(i)} = I(\eta_t > p_t^{(i)}(I_t^{(1)}, ..., I_t^{(r)})),
\]

one obtains a version of the former process.

Hence, defining \( N_n^{(i)}, \hat{N}_n^{(i)} \) by \( N_n^{(i)}\{t/n\} = I_t^{(i)}, \hat{N}_n\{t/n\} = \hat{I}_t^{(i)}, \ i = 1, ..., r, \ t = 1, ..., n, \) and \( N_n^{(i)}\{1/n, ..., n/n\} = 0 \) it follows that \( N_n = (N_n^{(1)}, ..., N_n^{(r)}) \) and \( \hat{N}_n = (\hat{N}_n^{(1)}, ..., \hat{N}_n^{(r)}) \) have the same distributions as the point processes of exceedances of the levels \( u^{(1)}, ..., u^{(r)} \) by \( \{\xi_t\} \) and \( \{\hat{\xi}_t\} \) respectively. Furthermore,
\[ P(I_t^{(i)} + \hat{I}_t^{(i)}) \leq E\{E(\|I_t^{(i)} - \hat{I}_t^{(i)}\| \| \eta_1, \ldots, \eta_{t-1}\}) \]
\[ = E|p_t^{(i)} - p_t^{(i)}| \]
\[ = E|P(I_t^{(i)} = 0 \| \mathcal{B}_{t-1}^c) - P(I_t^{(i)} = 0)| \]

and since
\[ P(N_n \neq \hat{N}_n) \leq \sum_{t=1}^{r} \sum_{i=1}^{r} P(I_t^{(i)} \neq \hat{I}_t^{(i)}) \]
it follows by Lemma 2.3 (ii) that (4.2) holds for
\[ u(r) \leq 2(1+\rho)^{-1/2} \left( \log n/K \right)^{1/2}. \]
Further, for \( u(r) > 2(1+\rho)^{-1/2} \left( \log n/K \right)^{1/2} \)
\[ P(N_n \neq \hat{N}_n) \leq P(M_n > u(r)) + P(\hat{M}_n > u(r)) \]
\[ \leq 4R_n , \]
as was shown in the proof of Theorem 3.1, and hence (4.2) holds also in this case.

The proof of part (ii) is quite similar. For \( i = 1, \ldots, r \) let
\[ \pi_m^{(i)} = \sum_{k=m}^{\infty} \exp\{-\tau^{(i)}/n\} \cdot \left( \tau^{(i)}/n \right)^k/k! \]
be the probability that a Poisson variable with mean \( \tau^{(i)}/n \) is larger than \( m \), and define
\( \hat{T}_1^{(i)}, \ldots, \hat{T}_n^{(i)} \) by requiring that \( \hat{T}_t^{(i)} = m \) on the set \( \tau_t^{(i)} < \eta_t \leq \tau_t^{(i)} \).
It is then immediate that \( N_n = (\hat{N}_n^{(1)}, \ldots, \hat{N}_n^{(r)}) \) has the required distribution if \( N_n^{(i)} \) is defined by \( N_n^{(i)}(\{t/n\}) = \hat{N}_t^{(i)}, \ t = 1, \ldots, n, \)
and \( N_n^{(i)}(\{1/n, \ldots, n/n\})^c = 0. \) Furthermore,
\[ P(N_n^{(i)} \neq \hat{N}_n^{(i)}) \leq \sum_{t=1}^{n} P(I_t^{(i)} \neq \hat{I}_t^{(i)}) \]
\[ \leq n(|\tau_t^{(i)} - \tau_t^{(i)}| + \pi_2^{(i)}) \]
\[ \leq n(|\tau_t^{(i)}/n - \tau_t^{(i)}/n| + (\tau_t^{(i)}/n)^2) \]
\[ = |\tau_t^{(i)} - \tau_t^{(i)}| + \frac{1}{n}(\tau_t^{(i)})^2 , \]
where we have used that 

$$|\tau^{(i)}_1 - \rho^{(i)}| = |1 - e^{-\tau^{(i)}_1/n} - \tau^{(i)}_n/n| \leq |\tau^{(i)}_1/n - \tau^{(i)}_n/n| + (\tau^{(i)}_1/n)^2/2$$

and that 

$$\tau^{(i)}_2 \leq (\tau^{(i)}_1/n)^2/2$$

in the third step. The inequality (4.3) now follows at once from part (i), since 

$$P(N_n \neq \hat{N}_n) \leq P(N_n \neq \hat{N}_n) + \sum_{i=1}^{2/2} P(N_n \neq \hat{N}_n).$$

The inequality (4.3) now follows at once from part (i), since 

$$P(N_n \neq \hat{N}_n) \leq P(N_n \neq \hat{N}_n) + \sum_{i=1}^{2/2} P(N_n \neq \hat{N}_n).$$

The variation distance, $d$, between the distributions of two $r$-dimensional integrally valued random variables $X = \{X^{(k)}; 1 \leq k \leq r\}$ and $Y = \{Y^{(k)}; 1 \leq k \leq r\}$ is defined as

$$d(X,Y) = \frac{1}{2} \sum_{z \in Z^r} |P(X = z) - P(Y = z)|,$$

where $Z^r$ is the $r$-dimensional integer lattice. It is immediate that $d$ is a metric on the set of distributions on $Z^r$, and that it is a metric for convergence in distribution. The interest and usefulness of the metric can be seen from the easily obtained relations

$$d(X,Y) = \frac{1}{2} \sup_{|h| \leq 1} |E_h(X) - E_h(Y)|$$

$$= \sup_{A} |P(X \in A) - P(Y \in A)|$$

$$\geq \sup_{z \in Z^r} |P(X \leq z) - P(Y \leq z)|$$

The distance $d$ only depends on the marginal distributions of $X$ and $Y$, but if $X$ and $Y$ have a joint distribution it follows from the second inequality in (4.4) that

$$d(X,Y) \leq P(X \neq Y).$$

This at once gives the first part of the following corollary to theorem 4.1.

Corollary 4.2 Suppose that the hypothesis of Theorem 4.1 is satisfied, let $A_{ij}, j = 1, \ldots, k_i, i = 1, \ldots, r$ be Borel subsets of
[0,1], and write \( v_n = \{ N_n^{(i)}(A_{ij}) ; j = 1, \ldots, k_i, i = 1, \ldots, r \} \) and let \( \hat{v}_n, \tilde{v}_n, \nu \) be defined similarly, but with \( N \) replaced by \( \hat{N}_n, \tilde{N}_n, \) and \( N \), respectively.

(i) Then

\[
d(\hat{v}_n, \nu) \leq 16 r^2 R_n,
\]

\[
d(\tilde{v}_n, \nu) \leq 16 r^2 R_n + \sum_{i=1}^{r} \frac{1}{n} |\tau_n^{(i)} - \tau(i)| + \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{k_i} \tau(i)^2.
\]

and if furthermore the \( A_{ij} \)'s are intervals, then

\[
d(\nu_n, \nu) \leq 16 r^2 R_n + \sum_{i=1}^{r} \frac{1}{n} |\tau_n^{(i)} - \tau(i)| + \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{k_i} \tau(i)^2 + \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{k_i} 2k_i \tau(i)^2.
\]

(ii) Suppose \( r = 1 \), and write \( N_n = N_n^{(1)} \) for the time-normalized point process of exceedances of the level \( u = u^{(1)} \), let \( N \) be a Poisson process with parameter \( \tau \), and suppose that \( \tau_n = n(1 - \Phi(u)) \leq K, \) with \( n/K = e \). Further let \( h : [0,1] \to \mathbb{R} \) be bounded by one, \( \|h\| \leq 1 \), and have modulus of continuity \( \delta(\varepsilon) = \sup \{|h(t) - h(t')|; t, t' \in [0,1], |t - t'| \leq \varepsilon\} \). Then

\[
P(\|dN_n - dN\| > \varepsilon) \leq 16 r^2 R_n + |\tau_n - \tau| + \frac{\sum_{k=1}^{\infty} e^{-\tau/k!}}{k \varepsilon / \delta(1/n)} \cdot
\]

Proof (i) The first two bounds follow at once from Theorem 4.1 and (4.5). Furthermore,

\[
P(\tilde{v}_n + \nu) \leq \sum_{i=1}^{r} \sum_{j=1}^{k_i} P(\hat{N}_n(A_{ij}) + N(A_{ij}))
\]

\[
\leq \sum_{i=1}^{r} \sum_{j=1}^{k_i} 2(1 - e^{-\tau(i)/n})
\]

\[
\leq \sum_{i=1}^{r} 2k_i \tau(i),
\]

since clearly \( \hat{N}_n^{(i)}(A_{ij}) = N(A_{ij}) \) if \( N^{(i)} \) does not have any points in the two intervals \( ((t - 1)/n, t/n] \) which contain the endpoints of the interval \( A_{ij} \). The last inequality in part (i) is an
immediate consequence of (4.6), the second inequality in part (i), and the triangle inequality for the metric \( d \).

(ii) By (4.3) there exist versions of \( N_n, \tilde{N}_n \) such that

\[
(4.7) \quad P( N_n \neq \tilde{N}_n ) \leq 16R_n + |\tau_n - \tau| + \tau^2/n.
\]

Furthermore, recalling the definition of \( \tilde{N}_n \) from the Poisson process \( N \) with intensity \( \tau \), we have

\[
\int h d\tilde{N}_n - \int h dN \leq \sum_{t=1}^{n} \int_{s=(t-1)/n}^{t/n} |(h(t/n) - h(s))| dN(s)
\]

\[
\leq \sum_{t=1}^{n} \delta(1/n) \int_{s=(t-1)/n}^{t/n} dN(s)
\]

\[
= \delta(1/n)N([0,1]),
\]

and hence

\[
(4.8) \quad P( |\int h d\tilde{N}_n - \int h dN| > \varepsilon ) \leq \sum_{k>\varepsilon/\delta(1/n)} e^{-\tau k/k!}.
\]

Part (ii) now follows at once from (4.7) and (4.8).

As a last corollary we will give an approximation of the joint distribution of the \( M^{(k)}_n \)'s, the k-th largest of \( \xi_1, \ldots, \xi_n \). The proof is immediate from Corollary 4.2 (i) and the obvious relation

\[
\{ M^{(1)}_n \leq u^{(1)}_n, \ldots, M^{(r)}_n \leq u^{(r)}_n \} = \{ N^{(1)}_n ([0,1]) \leq 0, \ldots, N^{(r)}_n ([0,1]) \leq r - 1 \}
\]

and its counterpart for \( \hat{M}^{(k)}_n \), the k-th largest amongst \( \hat{\xi}_1, \ldots, \hat{\xi}_n \).

**Corollary 4.3** Suppose that the hypothesis of Theorem 4.1 is satisfied. Then
\[ |P(M_n^{(1)} \leq u^{(1)}, \ldots, M_n^{(r)} \leq u^{(r)}) - P(M_n^{(1)} \leq u^{(1)}, \ldots, M_n^{(r)} \leq u^{(r)})| \leq 16 r^2 R_n, \]

and

\[ |P(M_n^{(1)} \leq u^{(1)}, \ldots, M_n^{(r)} \leq u^{(r)}) - P(N^{(1)}([0,1]) \leq 0, \ldots, N^{(r)}([0,1]) \leq r-1)| \leq 16 r^2 R_n + \sum_{i=1}^{r} |\tau^{(i)} - \tau^{(i)}| + \frac{1}{n} \sum_{i=1}^{r} (\tau^{(i)})^2. \]

In particular, for \( \tau_n = n(1 - \Phi(u)) \leq K, \)

\[ |P(M_n^{(k)} \leq u) - \sum_{i=0}^{k-1} \frac{\tau_n^i}{i!} | \leq 16 R_n + \frac{1}{n} \tau_n^2. \]
5. A comment on improved approximations

As noted above, the approximation errors in Theorems 3.1 and 4.1 are roughly of the order \((1/n)(1-\rho)/(1+\rho)\), with \(\rho=\max(0,r_1,r_2,\ldots)\), provided \(r_t\) decreases sufficiently fast as \(t \to \infty\). For \(\rho\) small, this seems quite satisfactory, but for \(\rho\) close to one it would be useful to have more accurate approximations. Considering e.g. 
\[ P(M_n \leq u), \text{ and } P(\hat{M}_n \leq u) \] as a first order term in the approximation of it, one possibility would be to find a second order term, similar to the right hand side of (3.8), and then to find a bound for the error in the second order approximation of \(P(M_n \leq u)\). This, although perhaps feasible, seems likely to incur considerable extra complications to the already somewhat involved calculations of this paper.

Another possibility would be to approximate \(M_n\) by the maximum, say \(\bar{M}_n\), of some other dependent stationary sequence \(\{\xi_t\}\) which in some way is easier to handle than the original sequence \(\{\xi_t\}\). If \(\{\xi_t\}\) has zero means, unit variances, and covariances \(r_t = E\xi_s\xi_{s+t}\), an analogue of (2.4) is valid, namely 
\[ P(M_n \leq u) - P(\bar{M}_n \leq u) = \sum_{1 \leq s < t \leq n} \int f_{r,s-t}(x_s = x_t = u) \, dx \, dh, \]
where \(f_{r,s-t}(x_s = x_t = u)\) now is obtained by putting \(x_s = x_t = u\) in the density function of \(n\) standard normal variables with covariances 
\[ nr_t + (1-h)r_t, \] c.f. e.g. Leadbetter, Lindgren, & Rootzén (1979), p. 47. Let \(\rho, \rho'\) be as on p. 6 and define \(\bar{\rho}, \bar{\rho}'\) similarly from \(\{r_t\}\). Calculations parallel to those in Sections 2 and 3 then show that, under suitable conditions, the rate of convergence to zero of 
\[ P(M_n \leq u) - P(\bar{M}_n \leq u) \] is similar to that of 
\[ P(M_n \leq u) - P(\hat{M}_n \leq u) \]
but with \(\rho\) replaced by \(\max(\rho, \bar{\rho})\) if \(\rho > \bar{\rho}\), and by \(\max(\rho, \bar{\rho})\) if \(\rho < \bar{\rho}\).
(\rho', \overline{\rho}') if \rho = \overline{\rho}, \rho' \neq \overline{\rho}'. Thus, the order of the approximation is improved only if the maximal covariance \overline{\rho} in the approximating process is \textit{exactly} equal to \rho.

One consequence of the convergence of \( N_n \) for any \( r \), is that asymptotically the locations of the \( k \)-th largest values are uniformly distributed. An interesting question is to find a bound for the rate of this convergence, which is conjectured to be of the same order as the convergences treated in this paper.

\textbf{Acknowledgement} I want to thank Jan Lanke for many useful discussions on the subject of this paper, and Philip Hougaard for reading the manuscript.
References


No. 1 Olkin, Ingram and Væth, Michael: Maximum Likelihood Estimation in a Two-way Analysis of Variance with Correlated Errors in One Classification.

No. 2 Skovgaard, Ib M.: Transformation of an Edgeworth Expansion by a Sequence of Smooth Functions.

No. 3 Asmussen, Søren: Equilibrium Properties of the M/G/1 Queue.

No. 4 Johansen, Søren and Keiding, Susanne: A Family of Models for the Elimination of Substrate in the Liver.


No. 6 Tjur, Tue: A Connection Between Rasch's Item Analysis Model and a Multiplicative Poisson Model.

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Johansen, Søren: Asymptotic Inference in Random Coefficient Regression Models.

No. 2 Asmussen, Søren: On the Role of a Certain Eigenvalue in Estimating the Growth Rate of a Branching Process.


No. 4 Asmussen, Søren: Conditioned Limit Theorems Relating a Random Walk to its Associate with Applications to Risk Reserve Processes and the GI/G/1 Queue.

No. 5 Johansen, Søren: The Statistical Analysis of a Markov Branching Process.

No. 6 Asmussen, Søren: Time-Dependent Approximations in some Queueing Systems with Imbedded Markov Chains Related to Random Walks.


No. 8 The Rate of Convergence of Extremes of Stationary Normal Sequences.