The Annals of Probability
1990, Vol. 18, No. 2, 713–723

POISSON APPROXIMATION USING THE STEIN–CHEN METHOD AND COUPLING: NUMBER OF EXCEEDANCES OF GAUSSIAN RANDOM VARIABLES

BY LARS HOLST AND SVANTE JANSON

Royal Institute of Technology, Stockholm, and Uppsala University

Consider a family of (dependent) Gaussian random variables and count the number of them that exceed some given levels. An explicit upper bound is given for the total variation distance between the distribution of this number of exceedances and a Poisson distribution having the same mean. The bound involves only means and covariances of the indicators that the variables exceed the levels. The general result is illustrated by some examples from the extreme value theory of Gaussian sequences. The bound is derived as a special case of a result obtained by the Stein–Chen method for sums of dependent Bernoulli random variables. This general result requires the existence of a certain coupling, which in the Gaussian case follows by a correlation inequality.

1. Introduction. By a method due to Stein and Chen many new results have in recent years been obtained on Poisson approximation of random variables representing the number of occurrences of dependent events, see Arratia, Goldstein and Gordon (1989), Barbour and Holst (1989), Barbour, Holst and Janson (1988a, b), Chen (1975), Smith (1988), Stein (1986) and the references therein.

The main purpose of this paper is to use this method to study the distribution of the number of exceedances of (high) levels in a Gaussian vector. An upper bound on the variation distance between that distribution and a Poisson distribution with the same mean is obtained. This bound is useful both for numerical and theoretical purposes, for example, to study convergence rates for extremes in Gaussian sequences.

In order to accomplish this we will first derive a general upper bound for variation distances using the Stein–Chen method. We also need the existence of certain monotone couplings for multivariate normal distributions. This will be proved using a result on association between normal random variables.

The organization of the paper is as follows: In Section 2 the general results concerning the upper bound and couplings are derived. The number of exceedances in the Gaussian case is studied in Section 3.

2. Variation distance and coupling. In the following Γ stands for a *finite* index set; we set $\Gamma_\alpha = \Gamma - \{\alpha\}$. Let I_α, $\alpha \in \Gamma$, be Bernoulli random variables with

$$\mathbb{P}(I_\alpha = 1) = 1 - \mathbb{P}(I_\alpha = 0) = p_\alpha$$

Received October 1988.
AMS 1980 *subject classifications*. Primary 60F05; secondary 60G10, 60G15.
Key words and phrases. Convergence rates, coupling, extreme values, Gaussian distributions, Poisson approximation, Stein–Chen method.

and set

$$W = \sum_{\alpha \in \Gamma} I_\alpha, \qquad \lambda = \mathbb{E}W = \sum_{\alpha \in \Gamma} p_\alpha.$$

We say that X is $\mathrm{Po}(\lambda)$ if the random variable X has a Poisson distribution with mean λ. The total variation distance between X and W is defined as

$$d(W, X) = \sup_A |\mathbb{P}(W \in A) - \mathbb{P}(X \in A)|.$$

THEOREM 2.1. *Let, for each $\alpha \in \Gamma$, the random variables I_α, I_β, $J_{\beta\alpha}$, $\beta \in \Gamma_\alpha$ be defined on the same probability space with*

$$\mathscr{L}(J_{\beta\alpha}; \beta \in \Gamma_\alpha) = \mathscr{L}(I_\beta; \beta \in \Gamma_\alpha | I_\alpha = 1).$$

Then

(a) *for $V_\alpha = \sum_{\beta \in \Gamma_\alpha} J_{\beta\alpha}$,*

$$d(W, \mathrm{Po}(\lambda)) \le \frac{1 - e^{-\lambda}}{\lambda} \sum_{\alpha \in \Gamma} p_\alpha \mathbb{E}|W - V_\alpha|.$$

(b) *If there exists a partition $\Gamma_\alpha = \Gamma_\alpha^+ \cup \Gamma_\alpha^- \cup \Gamma_\alpha^0$ with $J_{\beta\alpha} \ge I_\beta$ for $\beta \in \Gamma_\alpha^+$ and $J_{\beta\alpha} \le I_\beta$ for $\beta \in \Gamma_\alpha^-$, then*

$$d(W, \mathrm{Po}(\lambda))$$
$$\le \frac{1 - e^{-\lambda}}{\lambda} \left\{ \sum_{\alpha \in \Gamma} p_\alpha^2 + \sum_{\substack{\alpha \ne \beta \\ \beta \in \Gamma_\alpha^+ \cup \Gamma_\alpha^-}} \sum |\mathrm{Cov}(I_\alpha, I_\beta)| + \sum_{\substack{\alpha \ne \beta \\ \beta \in \Gamma_\alpha^0}} \sum (\mathbb{E} I_\alpha I_\beta + p_\alpha p_\beta) \right\}.$$

REMARK. The case $\Gamma_\alpha^+ = \Gamma_\alpha^0 = \varnothing$ is considered in Barbour and Holst (1989). The general version is due to Barbour, Holst and Janson (1988b). To prove the theorem, the Stein–Chen method can be used. We include a proof for completeness.

PROOF. For a given set A and X $\mathrm{Po}(\lambda)$ define the function f_A by $f_A(0) = 0$ and

$$\lambda f_A(j + 1) - j f_A(j) = I(j \in A) - \mathbb{P}(X \in A).$$

Barbour and Eagleson (1983) proved that f_A is bounded with

$$\Delta f_A = \sup_{j \ge 0} | f_A(j + 1) - f_A(j) | \le \frac{1 - e^{-\lambda}}{\lambda}.$$

Hence we have

$$\begin{aligned}
|\mathbb{P}(W \in A) - \mathbb{P}(X \in A)| &= |\mathbb{E}(\lambda f_A(W + 1) - W f_A(W))| \\
&= \left| \sum p_\alpha \big(\mathbb{E}(f_A(W + 1)) - \mathbb{E}(f_A(W) \,|\, I_\alpha = 1)\big) \right| \\
&\le \sum p_\alpha |\mathbb{E}(f_A(W + 1) - f_A(V_\alpha + 1))| \\
&\le \sum p_\alpha \mathbb{E}| f_A(W + 1) - f_A(V_\alpha + 1)| \\
&\le \Delta f_A \sum p_\alpha \mathbb{E}|W - V_\alpha| \le \frac{1 - e^{-\lambda}}{\lambda} \sum p_\alpha \mathbb{E}|W - V_\alpha|,
\end{aligned}$$

proving (a).

Now for $\beta \in \Gamma_\alpha^+$,

$$0 \le p_\alpha \mathbb{E}|I_\beta - J_{\beta\alpha}| = p_\alpha \mathbb{E}(J_{\beta\alpha} - I_\beta) = p_\alpha \mathbb{E}(I_\beta | I_\alpha = 1) - p_\alpha p_\beta$$
$$= \mathbb{E}(I_\alpha I_\beta) - p_\alpha p_\beta = \operatorname{Cov}(I_\alpha, I_\beta).$$

Similarly for $\beta \in \Gamma_\alpha^-$,

$$0 \le p_\alpha \mathbb{E}|I_\beta - J_{\beta\alpha}| = -\operatorname{Cov}(I_\alpha, I_\beta).$$

Thus

$$\begin{aligned} &p_\alpha \mathbb{E}|W - V_\alpha| \\ &\quad = p_\alpha \mathbb{E}\left| I_\alpha + \sum_{\beta \in \Gamma_\alpha^+} (I_\beta - J_{\beta\alpha}) + \sum_{\beta \in \Gamma_\alpha^-} (I_\beta - J_{\beta\alpha}) + \sum_{\beta \in \Gamma_\alpha^0} (I_\beta - J_{\beta\alpha}) \right| \\ &\quad \le p_\alpha \left\{ \mathbb{E} I_\alpha + \sum_{\beta \in \Gamma_\alpha^+} \mathbb{E}|I_\beta - J_{\beta\alpha}| + \sum_{\beta \in \Gamma_\alpha^-} \mathbb{E}|I_\beta - J_{\beta\alpha}| + \sum_{\beta \in \Gamma_\alpha^0} \mathbb{E}(I_\beta + J_{\beta\alpha}) \right\} \\ &\quad = p_\alpha^2 + \sum_{\beta \in \Gamma_\alpha^+} \operatorname{Cov}(I_\alpha, I_\beta) + \sum_{\beta \in \Gamma_\alpha^-} \left|\operatorname{Cov}(I_\alpha, I_\beta)\right| + \sum_{\beta \in \Gamma_\alpha^0} (p_\alpha p_\beta + \mathbb{E} I_\alpha I_\beta). \end{aligned}$$

Hence (b) follows from (a). □

The crucial step when applying Theorem 2.1 in a specific situation is to construct an efficient coupling defining the partition of Γ_α and the J's; cf. the examples in Barbour and Holst (1989) and Barbour, Holst and Janson (1988a, b). However, it is not necessary to give an explicit construction; it suffices to know the existence of a suitable one. For this the following result will be used in Section 3.

THEOREM 2.2. *Let the random vector $(X_0, X_1, \dots, X_n)$ be Gaussian with* $\operatorname{Cov}(X_0, X_i) \ge 0$ *for* $1 \le i \le p$ *and* $\operatorname{Cov}(X_0, X_i) \le 0$ *for* $p < i \le n$. *Then, for any random variable T, which is independent of the vector, there exists a probability space with random variables $Y_1, \dots, Y_n, Z_1, \dots, Z_n$ such that*

$$\mathscr{L}(Y_1, \dots, Y_n) = \mathscr{L}(X_1, \dots, X_n),$$
$$\mathscr{L}(Z_1, \dots, Z_n) = \mathscr{L}(X_1, \dots, X_n | X_0 > T),$$
$$Y_i \le Z_i, \quad 1 \le i \le p, \quad Y_i \ge Z_i, \quad p < i \le n.$$

PROOF. As T and $(X_0, \dots, X_n)$ are independent, it follows from Corollary 3 of Joag-Dev, Perlman and Pitt (1983) that for any bounded real function increasing in all its arguments,

$$\begin{aligned} &\mathbb{E}\big(I(X_0 > T) f(X_1, \dots, X_p, -X_{p+1}, \dots, -X_n)\big) \\ &\quad = \mathbb{E}\mathbb{E}\big(I(X_0 > T) f(X_1, \dots, X_p, -X_{p+1}, \dots, -X_n) | T\big) \\ &\quad \ge \mathbb{E}\mathbb{E}(I(X_0 > T) | T) \mathbb{E} f(X_1, \dots, -X_n) \\ &\quad = \mathbb{E} I(X_0 > T) \mathbb{E} f(X_1, \dots, -X_n). \end{aligned}$$

Hence for all such f,

$$\mathbb{E}\big(f(X_1,\dots,X_p,-X_{p+1},\dots,-X_n)|X_0>T\big)\ge \mathbb{E}f(X_1,\dots,-X_n).$$

According to Theorem 1 in Kamae, Krengel and O'Brien (1977) or results in Section II.2 of Liggett (1985) this implies the existence of a probability space with random variables $Y_1,\dots,Y_n,Z_1,\dots,Z_n$ with

$$\mathscr{L}(Y_1,\dots,Y_p,-Y_{p+1},\dots,-Y_n)=\mathscr{L}(X_1,\dots,X_p,-X_{p+1},\dots,-X_n),$$

$$\mathscr{L}(Z_1,\dots,Z_p,-Z_{p+1},\dots,-Z_n)=\mathscr{L}(X_1,\dots,X_p,-X_{p+1},\dots,-X_n|X_0>T)$$

$$Y_i\le Z_i,\qquad 1\le i\le p,\qquad -Y_i\le -Z_i,\qquad p<i\le n,$$

which proves the assertion. □

3. Exceedances in the Gaussian case.

THEOREM 3.1. *Suppose X_α, $\alpha\in\Gamma$, are jointly normally distributed random variables. Let t_α, $\alpha\in\Gamma$, be real numbers and set*

$$I_\alpha=I(X_\alpha>t_\alpha),\qquad p_\alpha=\mathbb{P}(X_\alpha>t_\alpha),$$
$$W=\sum_{\alpha\in\Gamma}I_\alpha,\qquad \lambda=\mathbb{E}W=\sum_{\alpha\in\Gamma}p_\alpha.$$

Then

$$d(W,\mathrm{Po}(\lambda))\le\frac{1-e^{-\lambda}}{\lambda}\left(\sum_{\alpha\in\Gamma}p_\alpha^2+\sum_{\substack{\alpha,\beta\in\Gamma\\ \alpha\ne\beta}}\sum\left|\mathrm{Cov}(I_\alpha,I_\beta)\right|\right).$$

PROOF. For any fixed $\alpha\in\Gamma$, the existence of a partition of Γ_α with $\Gamma_\alpha^0=\varnothing$ in Theorem 2.1 follows from Theorem 2.2. From this the assertion follows. □

COROLLARY 3.2. *Suppose X_α, $\alpha\in\Gamma$, are jointly normally distributed random variables. Let t_α, $\alpha\in\Gamma$, be positive numbers and set*

$$I_\alpha=I(|X_\alpha|>t_\alpha),\qquad I_\alpha^+=I(X_\alpha>t_\alpha),\qquad I_\alpha^-=I(X_\alpha<-t_\alpha),$$
$$p_\alpha=\mathbb{P}(|X_\alpha|>t_\alpha),\qquad W=\sum_{\alpha\in\Gamma}I_\alpha,\qquad \lambda=\mathbb{E}W=\sum_{\alpha\in\Gamma}p_\alpha.$$

Then

$$d(W,\mathrm{Po}(\lambda))\le\frac{1-e^{-\lambda}}{\lambda}\Bigg(\sum_{\alpha\in\Gamma}p_\alpha^2+\sum_{\substack{\alpha,\beta\in\Gamma\\ \alpha\ne\beta}}\sum\Big(\left|\mathrm{Cov}(I_\alpha^+,I_\beta^+)\right|+\left|\mathrm{Cov}(I_\alpha^+,I_\beta^-)\right|$$
$$+\left|\mathrm{Cov}(I_\alpha^-,I_\beta^+)\right|+\left|\mathrm{Cov}(I_\alpha^-,I_\beta^-)\right|\Big)\Bigg).$$

PROOF. Set $X_\alpha^* = -X_\alpha$. Then

$$W = \sum_{\alpha\in\Gamma} I(X_\alpha > t_\alpha) + \sum_{\alpha\in\Gamma} I(X_\alpha^* > t_\alpha).$$

Theorem 3.1 applied to the variables $\{X_\alpha\} \cup \{X_\alpha^*\}$ (with two copies of Γ as index set) yields, if S denotes the double sum in the bound above,

$$d(W, \mathrm{Po}(\lambda)) \le \frac{1-e^{-\lambda}}{\lambda}\Bigg(\sum_{\alpha\in\Gamma} (\mathbb{E}I_\alpha^+)^2 + \sum_{\alpha\in\Gamma} (\mathbb{E}I_\alpha^-)^2 + \sum_{\alpha\in\Gamma} \big(|\mathrm{Cov}(I_\alpha^+, I_\alpha^-)| + |\mathrm{Cov}(I_\alpha^-, I_\alpha^+)|\big) + S\Bigg).$$

The assertion follows because $\mathrm{Cov}(I_\alpha^+, I_\alpha^-) = -\mathbb{E}I_\alpha^+\mathbb{E}I_\alpha^-$ and thus

$$(\mathbb{E}I_\alpha^+)^2 + (\mathbb{E}I_\alpha^-)^2 + 2|\mathrm{Cov}(I_\alpha^+, I_\alpha^-)| = (\mathbb{E}I_\alpha^+ + \mathbb{E}I_\alpha^-)^2 = p_\alpha^2. \qquad \square$$

We may replace the fixed levels t_α in Theorem 3.1 and Corollary 3.2 by random levels T_α that are independent of each other and of $\{X_\alpha\}$ (cf. Theorem 2.2). We leave the details to the reader.

REMARK. If λ is large, then $\mathrm{Po}(\lambda)$ may be approximated by $N(\lambda, \lambda)$. Hence the Poisson approximation result in Theorem 3.1 implies that W is asymptotically normally distributed in situations where $\lambda \to \infty$ while $\lambda^{-1}(\sum p_\alpha^2 + \sum\sum |\mathrm{Cov}(I_\alpha, I_\beta)|) \to 0$.

Let us take all t_α equal to t. Then the event $\{W = 0\}$ is the same as $\{\max X_\alpha \le t\}$. Consequently Theorem 3.1 gives as a corollary an upper bound for $|\mathbb{P}(\max X_\alpha \le t) - e^{-\lambda}|$ and, analogously, Corollary 3.2 gives an upper bound for $|\mathbb{P}(\max|X_\alpha| \le t) - e^{-\lambda}|$. Estimates on the distribution of the kth largest value are obtained similarly.

Theorem 3.1 is simply to apply numerically to specific examples as the bound only involves the one- and two-dimensional normal distribution functions. It is also easy to use Theorem 3.1 to obtain asymptotic results. We will illustrate that using the following estimates for the covariances:

Let Φ and φ be the standard normal distribution and density functions.

LEMMA 3.3. *Let (X, Y) be $N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & r\\ r & 1\end{pmatrix}\right)$. If $0 \le r < 1$, then for any real a and b,*

$$\begin{aligned}(1-\Phi(a))&\left(1-\Phi\left(\frac{b-ra}{\sqrt{1-r^2}}\right)\right)\\ &\le \mathbb{P}(X > a \text{ and } Y > b)\\ &\le (1-\Phi(a))\left[\left(1-\Phi\left(\frac{b-ra}{\sqrt{1-r^2}}\right)\right) + r\frac{\varphi(b)}{\varphi(a)}\left(1-\Phi\left(\frac{a-rb}{\sqrt{1-r^2}}\right)\right)\right].\end{aligned}$$

If $-1 < r \le 0$, the inequalities are reversed.

PROOF. By integration by parts we get

$$\mathbb{P}(X > a \text{ and } Y > b) = \int_a^\infty \varphi(x)\left(1 - \Phi\left(\frac{b - rx}{\sqrt{1 - r^2}}\right)\right) dx$$

$$= (1 - \Phi(a))\left(1 - \Phi\left(\frac{b - ra}{\sqrt{1 - r^2}}\right)\right)$$

$$+ \int_a^\infty (1 - \Phi(x))\varphi\left(\frac{b - rx}{\sqrt{1 - r^2}}\right)\frac{r}{\sqrt{1 - r^2}}\, dx.$$

Suppose $0 \le r < 1$. The lower bound then follows immediately. Next note that $(1 - \Phi(x))/(\varphi(x))$ is decreasing. Thus

$$\int_a^\infty (1 - \Phi(x))\varphi\left(\frac{b - rx}{\sqrt{1 - r^2}}\right)\frac{dx}{\sqrt{1 - r^2}}$$

$$\le \frac{1 - \Phi(a)}{\varphi(a)}\int_a^\infty \varphi(x)\varphi\left(\frac{b - rx}{\sqrt{1 - r^2}}\right)\frac{dx}{\sqrt{1 - r^2}}$$

$$= \frac{1 - \Phi(a)}{\varphi(a)}\int_a^\infty \varphi(b)\varphi\left(\frac{x - rb}{\sqrt{1 - r^2}}\right)\frac{dx}{\sqrt{1 - r^2}}$$

$$= (1 - \Phi(a))\frac{\varphi(b)}{\varphi(a)}\left(1 - \Phi\left(\frac{a - rb}{\sqrt{1 - r^2}}\right)\right),$$

giving the upper bound. For the case $-1 < r \le 0$ the same argument works with the inequalities reversed. □

LEMMA 3.4. *Let $\{z_n\}$ be a sequence such that $\lambda_n = n(1 - \Phi(z_n)) \le K$ for some constant K. Let (X_α, X_β) be $N\left(\binom{0}{0}, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right)$ and let $I_\alpha = I(X_\alpha > z_n)$, $I_\beta = I(X_\beta > z_n)$. Then, for some constants C depending on K only and all $n \ge 2$:*

(i) *If* $0 \le r < 1$,

$$0 \le \operatorname{Cov}(I_\alpha, I_\beta) \le C(1 - r)^{-1/2} n^{-2+2r/(1+r)}(\log n)^{-r/(1+r)}.$$

(ii) *If* $0 \le r \le 1$,

$$0 \le \operatorname{Cov}(I_\alpha, I_\beta) \le C\frac{r \log n}{n^2} e^{2r \log n}.$$

(iii) *If* $-1 \le r \le 0$,

$$0 \ge \operatorname{Cov}(I_\alpha, I_\beta) \ge -C\frac{|r|\log n}{n^2}.$$

(iv) *If* $-1 \le r \le 0$,

$$0 \ge \operatorname{Cov}(I_\alpha, I_\beta) \ge -C\frac{1}{n^2}.$$

PROOF. (i) We may assume that $n > 7K$ and thus $z_n > 1$. Lemma 3.3 yields

$$\begin{aligned}
0 \le \operatorname{Cov}(I_\alpha, I_\beta) &\le \mathbb{E}(I_\alpha I_\beta) \le (1 - \Phi(z_n))(1 + r)\left(1 - \Phi\left(\sqrt{\frac{1-r}{1+r}}\, z_n\right)\right) \\
&\le (1 + r) z_n^{-1}\varphi(z_n)\left(\sqrt{\frac{1-r}{1+r}}\, z_n\right)^{-1} \varphi\left(\sqrt{\frac{1-r}{1+r}}\, z_n\right) \\
&= (1 + r)^{3/2}(1 - r)^{-1/2}\left(\frac{\varphi(z_n)}{z_n}\right)^{1+(1-r)/(1+r)} z_n^{-2r/(1+r)}(2\pi)^{-r/(1+r)} \\
&\le C(1 - r)^{-1/2}\frac{1}{n}(\log n)^{-r/(1+r)}\left(\frac{1}{n}\right)^{(1-r)/(1+r)}.
\end{aligned}$$

In the last inequality we use the fact that the next to last term is a decreasing function of z_n; hence we may assume that $1 - \Phi(z_n) = K/n$ and thus $z_n \sim \sqrt{2\log n}$ and $(\varphi(z_n))/z_n \sim K/n$.

(ii) We may assume that $z_n > 0$. Then, by Lemma 3.3,

$$\begin{aligned}
\operatorname{Cov}(I_\alpha, I_\beta) &\le (1 - \Phi(z_n))\left((1 + r)\left(\Phi(z_n) - \Phi\left(\sqrt{\frac{1-r}{1+r}}\right)\right) + r(1 - \Phi(z_n))\right) \\
&\le (1 + r)\frac{\varphi(z_n)}{z_n}\left(z_n - \sqrt{\frac{1-r}{1+r}}\, z_n\right)\varphi\left(\sqrt{\frac{1-r}{1+r}}\, z_n\right) \\
&\quad + r(1 - \Phi(z_n))^2 \\
&\le 2r\varphi(z_n)\varphi\left(\sqrt{\frac{1-r}{1+r}}\, z_n\right) + C\frac{r}{n^2} \\
&= 2r z_n^{2/(1+r)}\left(\frac{\varphi(z_n)}{z_n}\right)^{1+(1-r)/(1+r)}(2\pi)^{-r/(1+r)} + C\frac{r}{n^2} \\
&\le Cr(\log n)^{1/(1+r)} n^{-2+2r/(1+r)} \\
&\le C\frac{1}{n^2} r(\log n) e^{2r\log n}.
\end{aligned}$$

(iii) We may assume that $z_n > 1$. Then Lemma 3.3 yields

$$0 \le -\operatorname{Cov}(I_\alpha, I_\beta) \le (1 - \Phi(z_n))\left((1+r)\left(\Phi\left(\sqrt{\frac{1-r}{1+r}}z_n\right) - \Phi(z_n)\right) - r(1 - \Phi(z_n))\right)$$

$$\le \frac{\varphi(z_n)}{z_n}(1+r)\left(\sqrt{\frac{1-r}{1+r}} - 1\right)z_n\varphi(z_n) + |r|\left(\frac{\varphi(z_n)}{z_n}\right)^2$$

$$\le |r|\left(\frac{\varphi(z_n)}{z_n}\right)^2\left(z_n^2 + 1\right) \le C|r|n^{-2}\log n.$$

(iv) $$\operatorname{Cov}(I_\alpha, I_\beta) \ge -\mathbb{E}I_\alpha \mathbb{E}I_\beta \ge -\left(\frac{K}{n}\right)^2. \qquad \square$$

REMARK. Of the estimates above, (ii) and (iii) are sharp (within a constant), when $|r|\log n$ is bounded and λ_n is bounded from below, while (i) and (iv) are sharp when both $|r|\log n$ and λ_n are bounded from below, provided $r \le r_0 < 1$. This can be proved using the left inequality in Lemma 3.3 and calculations similar to those used above.

As an example, let us consider a stationary Gaussian process. The following result contains results on the distribution of extremes of the process; cf. Leadbetter, Lindgren and Rootzén (1983), Chapter 4, where similar results on the distributional convergence of extremes are proved by different methods.

THEOREM 3.5. *Let $\{\xi_k\}$ be a standardized stationary normal sequence with covariances $\{r_k\}$ satisfying $r_k \le A/\log k$, $k \ge 2$, for some constant A. Let $\rho = \max(0, r_1, r_2, \dots)$. Let λ_n and z_n be real numbers such that $\lambda_n = n(1 - \Phi(z_n))$ and define $W_n = \sum_{k=1}^n I(\xi_k > z_n)$. Suppose that $\lambda_n \le B < \infty$, for some constant B. Then*

$$d(W_n, \operatorname{Po}(\lambda_n)) = O\left(n^{-(1-\rho)/(1+\rho)}(\log n)^{-\rho/(1+\rho)} + \frac{\log n}{n}\sum_{k=1}^n |r_k|\right) \quad \textit{as } n \to \infty.$$

Note that $0 \le \rho < 1$, since $r_k = 1$ for some $k \ge 1$ would imply $r_{mk} = 1$, $m \ge 1$, which contradicts $r_k \le A/\log k$.

PROOF. Using Theorem 3.1 we get, with $I_k = I(\xi_k > z_n)$,

$$d(W_n, \operatorname{Po}(\lambda_n)) \le \frac{1 - e^{-\lambda_n}}{\lambda_n}\left(\frac{\lambda_n^2}{n} + 2n\sum_{k=1}^{n-1}|\operatorname{Cov}(I_0, I_k)|\right).$$

If $\rho = 0$, i.e., $r_k \le 0$ for every $k > 0$, the result follows by Lemma 3.4(iii). Let us now assume that $\rho > 0$. Choose $\delta > 0$ with $3\delta < \rho/(1+\rho)$. We divide the sum $\sum |\mathrm{Cov}(I_0, I_k)|$ into four parts.

(i) Since $r_k \le A/\log k$, only a finite number of k have $r_k > \delta$. Each of these give by Lemma 3.4(i) a contribution

$$\mathrm{Cov}(I_0, I_k) \le Cn^{-2}\left(\frac{n^2}{\log n}\right)^{r_k/(1+r_k)} \le Cn^{-2}\left(\frac{n^2}{\log n}\right)^{\rho/(1+\rho)},$$

so their total contribution is $O(n^{-2}(n^2/\log n)^{\rho/(1+\rho)})$.

(ii) Next consider the terms with $0 \le r_k \le \delta$ and $k < n^\delta$. There are at most n^δ such terms, and by Lemma 3.4(i) each contributes

$$\mathrm{Cov}(I_0, I_k) \le Cn^{-2}\left(\frac{n^2}{\log n}\right)^{\delta/(1+\delta)} \le Cn^{-2+2\delta}.$$

Hence their total contribution is at most

$$Cn^{-2+3\delta} = o\left(n^{-2}\left(\frac{n^2}{\log n}\right)^{\rho/(1+\rho)}\right).$$

(iii) For the terms with $r_k \ge 0$ and $k \ge n^\delta$, $0 \le r_k \le A/\log k \le A/\delta \log n$. Hence Lemma 3.4(ii) yields $0 \le \mathrm{Cov}(I_0, I_k) \le Cr_k n^{-2}\log n$ for each such term and their total contribution is $O(n^{-2}\log n \sum_1^n |r_k|)$.

(iv) The remaining terms are those with $r_k < 0$ and Lemma 3.4(iii) shows that their total contribution is $O(n^{-2}\log n \sum_1^n |r_k|)$.

Consequently,

$$\frac{\lambda^2}{n} + 2n\sum_{k=1}^{n-1} |\mathrm{Cov}(I_0, I_k)| = O\left(n^{-1}\left(\frac{n^2}{\log n}\right)^{\rho/(1+\rho)} + \frac{\log n}{n}\sum_1^n |r_k|\right)$$

and the result follows. □

Remark 1. As a corollary we see that if $r_k \log k \to 0$ as $k \to \infty$ and $\lambda_n \to \lambda < \infty$, then $W_n \Rightarrow_d \mathrm{Po}(\lambda)$. In particular, $\mathbb{P}(\max_{k \le n}\xi_k \le z_n) \to e^{-\lambda}$ (a result first proved by Berman). Note that $r_k \log k = O(1)$ is not sufficient for this limit result; Mittal and Ylvisaker have shown that if $r_k \log k \to \gamma > 0$, then W_n converges to a mixture of Poisson distributions, see Leadbetter, Lindgren and Rootzén (1983), Section 6.5.

Remark 2. The assumption that $r_k \log k$ is bounded was used only to conclude that $\{k \le n: r_k > M/\log n\} \le n^\delta$ for some $M < \infty$; hence it could be replaced by, e.g., $\sum_1^\infty e^{-M/|r_k|} < \infty$, for some M or $\sum_1^\infty |r_k|^p < \infty$ for some $p < \infty$.

It is also easy to obtain limit theorems with explicit rates of convergence for the other conditions on r_k discussed in Leadbetter, Lindgren and Rootzén (1983), Section 4.5; we leave the details to the reader.

REMARK 3. In most applications, for example to ARMA processes where r_k decreases exponentially, $\sum_1^\infty |r_k| < \infty$. In this case we obtain from the theorem that $d(W_n, \mathrm{Po}(\lambda_n)) = O(n^{-(1-\rho)/(1+\rho)}(\log n)^{\rho/(1+\rho)})$ when $\rho > 0$ and $d(W_n, \mathrm{Po}(\lambda_n)) = O(\log n/n)$ when $\rho = 0$. [The latter can easily be improved to $o(\log n/n)$; when $\{\xi_k\}$ is m-dependent and $\rho = 0$ we get $O(1/n)$, using Lemma 3.4(iv).]

REMARK 4. It follows from Corollary 3.5 that if $|r_k| \le A/\log k$, the estimate in Theorem 3.5 holds also for $W_n = \sum_{k=1}^n I(|\xi_k| > z_n)$, with $\lambda_n = 2n(1 - \Phi(z_n))$ and $\rho = \max|r_k|$.

It should be obvious that the methods of this paper also apply to nonstationary sequences; this gives results similar to those in Leadbetter, Lindgren and Rootzén (1983), Sections 6.1–6.3, however with error estimates. We may also consider processes with other index sets. As an example we give a version of Theorem 3.5 for stationary processes with multidimensional indices. The proof is the same with n replaced by $|B_n|$.

THEOREM 3.6. *Let $\{\xi_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{Z}^d}$ be a standardized stationary normal process with covariances $\{r_{\mathbf{k}}\}$ satisfying $\sup r_{\mathbf{k}} \log|\mathbf{k}| < \infty$. Let $\rho = \max\{0, r_{\mathbf{k}}\colon \mathbf{k} \neq 0\}$. Let B_n be subsets of $\mathbb{Z}^d$ and let $\tilde{B}_n = \{\mathbf{k} - \mathbf{l}\colon \mathbf{k}, \mathbf{l} \in B_n\}$. Let λ_n and z_n be positive numbers such that $\lambda_n = |B_n|(1 - \Phi(z_n))$ and define $W_n = \sum_{\mathbf{k} \in B_n} I(\xi_{\mathbf{k}} > z_n)$. If $\{\lambda_n\}$ is bounded, then as $n \to \infty$,*

$$d(W_n, \mathrm{Po}(\lambda_n)) = O\left(|B_n|^{-(1-\rho)/(1+\rho)}(\log|B_n|)^{-\rho/(1+\rho)} + \frac{\log|B_n|}{|B_n|} \sum_{\mathbf{k} \in \tilde{B}_n} |r_{\mathbf{k}}|\right).$$

REFERENCES

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25.

BARBOUR, A. D. and EAGLESON, G. K. (1983). Poisson approximation for some statistics based on exchangeable trials. *Adv. in Appl. Probab.* **15** 585–600.

BARBOUR, A. D. and HOLST, L. (1989). Some applications of the Stein–Chen method for proving Poisson convergence. *Adv. in Appl. Probab.* **21** 74–90.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1988a). Poisson approximation in occupancy problems. Report No. 1988:5, Dept. Mathematics, Uppsala Univ.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1988b). Poisson approximation with the Stein–Chen method and coupling. Report No. 1988:10, Dept. Mathematics, Uppsala Univ.

CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.

JOAG-DEV, K., PERLMAN, M. D. and PITT, L. D. (1983). Association of normal random variables and Slepian's inequality. *Ann. Probab.* **11** 451–455.

KAMAE, T., KRENGEL, U. and O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5** 899–912.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

LIGGETT, T. (1985). *Interacting Particle Systems*. Springer, New York.

SMITH, R. (1988). Extreme value theory for dependent sequences via the Stein–Chen method of Poisson approximation. *Stochastic Process. Appl.* **30** 317–327.

STEIN, C. (1986). *Approximate Computation of Expectations*. IMS, Hayward, Calif.

DEPARTMENT OF MATHEMATICS
ROYAL INSTITUTE OF TECHNOLOGY
S-100 44 STOCKHOLM
SWEDEN

DEPARTMENT OF MATHEMATICS
THUNBERGSV. 3
S-752 38 UPPSALA
SWEDEN