On the Number of Exceedances in Gaussian and Related Sequences

Mikael Raab

Department of Mathematics
Stockholm, 1997

Mikael Raab
Department of Mathematics
Division of Mathematical Statistics
Royal Institute of Technology
S-100 44 Stockholm
SWEDEN

ISBN 91-7170-754-9

Abstract

Consider a finite sequence of random variables, and count the number of exceedances of some level a, i.e. the number of values exceeding the level. Let this level and the length of the sequence increase simultaneously so that the expected number of exceedances remains fixed. It is well-known that for Gaussian sequences the number of exceeding points converges in distribution to a Poisson distribution if the long-range dependence is not too strong. In this thesis we use the Stein-Chen method to investigate Poisson approximation for χ^2–sequences, i.e. sequences defined as sums of squares of independent and identically distributed Gaussian sequences. Conditions for Poisson convergence and bounds for the convergence rate are derived, which show similarities with the Poisson approximation for Gaussian sequences.

For Gaussian sequences with some large correlations, the Poisson convergence is slow due to clumping. Using Stein's method we show that, at least for m-dependent sequences, the rate of convergence is improved by using compound Poisson for approximating the distribution. An explicit bound for the convergence rate is derived for the compound Poisson approximation, and also for a subclass of the compound Poisson distribution, where only clumps of size two are considered.

Keywords and phrases: Convergence rates, Poisson approximation, compound Poisson approximation, Stein-Chen method, coupling, extreme values, χ^2–sequences, Gaussian sequences.

AMS subject classification: Primary 60F05. Secondary 60G10, 60G15, 60G70.

Acknowledgements

I thank my supervisor Lars Holst for his never-ending support and encouragement, and for sharing his broad knowledge of mathematical statistics with me. I am also grateful to my colleagues in the department for many inspiring discussions and helpful ideas.

Finally, I thank my father for always taking time discussing mathematical statistics with me, my mother for her steadfast belief in my ability and my whole family for their love, understanding and support.

Stockholm, Nov 1996

Mikael Raab

Contents

Chapter 1

Introduction

Assume $(Y_1, Y_2, \ldots, Y_n)$ is a sequence of random variables. What is the distribution of the number of Y's, W, exceeding some level a_n? In the thesis we will answer this question in terms of asymptotic results, showing that the distribution is close to a Poisson or a compound Poisson distribution, for certain sequences. Results on Poisson approximation are obtained for both Gaussian and χ^2–sequences, whereas compound Poisson approximation is investigated only for Gaussian sequences.

The exact distribution of W is, in general, untractable. The results in the thesis are upper bounds for the distance (defined in a suitable way) between the true and some approximating distribution, for finite (and possibly small) n, or asymptotic results valid for large n. Since increasing n will lead to an increasing number of exceeding points, we must scale the problem, i.e. let the level a_n depend on n. We let the mean number of exceeding points be held constant.

Almost throughout the whole thesis, stationarity of the sequence is assumed. However, this condition can be relaxed in a straightforward manner, and more general theorems can be obtained (at the expense of simplicity).

One classical book in extreme value theory is Leadbetter, Lindgren and Rootzén [12]. It treats random processes in both discrete and continuous time, mostly stationary Gaussian processes. Among other things, the convergence of the number of exceeding points (in the sense as defined above) for stationary sequences is considered. Sufficient conditions for convergence to the Poisson distribution are stated.

For general stationary sequences the concept of distributional mixing is used. The rate of convergence for Gaussian sequences is also derived. However, the question of convergence to the compound Poisson distribution is not treated in this book.

Little seems to have been written about the distribution of the number of exceeding points for χ^2–sequences. However, there are many papers on the related problem of the number of upcrossings for χ^2–processes in continuous time. Limit theorems for χ^2–processes can be found e.g. in Sharpe [23] or Lindgren [13]. More recent results concerning χ^2–processes are presented in a number of papers by Albin, see Albin [1] and the references therein.

Under fairly general conditions, the distribution of number of exceedances, W, converges to a Poisson distribution. Two conditions restricting the underlying sequence must be satisfied. Firstly, the "global" dependence must be limited, so that events related to extreme values are more or less asymptotically independent. Secondly, the "local" dependence must be controlled, thereby limiting the possibility of multiple extreme points, i.e. clustering effects. For stationary Gaussian sequences, a sufficient condition for Poisson convergence is the so called Berman condition,

$$\rho_k \ln k \to 0 \text{ as } k \to \infty,$$

where $\{\rho_k\}$ is the correlation function, see Berman [5]. This condition, which restricts large correlations between random variables at great distance, prevents the "global" dependence from being too strong. However, even with this condition satisfied, the convergence can be slow due to clumping – a phenomenon occurring when the "local" dependence is strong. It is then reasonable to try to find a distribution (close to Poisson) which approximates the true distribution better. A natural choice is the compound Poisson distribution. This distribution, having the Poisson distribution as a special case, can be viewed as a generalization where the number of clumps (of fixed sizes) are Poisson distributed. The latter half of this thesis treats compound Poisson approximation.

In 1970, Stein [24] introduced a powerful and general method providing a way to find approximations to the distribution of a random quantity Y, which at the same time gives estimates of the approximation error involved. The main idea for getting error bounds is to estimate the difference between $E(f(Y))$ and $\int f\,dP$, where P denotes the approximating distribution, for a suitable class of test functions f. The method was originally applied in the context of convergence to the standard normal distribution, being analogous to the Berry-Esseen theorem, but it is not restricted to normal approximation. In a paper by Chen [6], the method was applied to Poisson approximation, and hence it is known as the Stein-Chen method. See also the book by Stein [25].

The starting point for the Stein-Chen method is as follows. For any bounded function f and $Z \sim \mathrm{Po}(\lambda)$ we have

$$E(\lambda f(Z+1) - Zf(Z)) = 0. \tag{1.1}$$

If W, often a sum of (possibly dependent) indicator random variables, is almost $\mathrm{Po}(\lambda)$, then the left hand side of (1.1) should be close to zero. By a suitable choice of f, the expression can sometimes be efficiently bounded, implying good approximation if the bound is small.

The Stein-Chen method appears in two versions. In the "local approach" the dependence structure of the random variables is such that every pair of random variables can be classified as either strongly or weakly dependent. In the more recent "coupling approach", it is assumed that one can construct a probability space with monotone couplings, i.e. with relations arranging some of the random variables in a certain manner, independently of the outcome of the random variables. Many applications of this approach can be found in the book by Barbour, Holst and Janson [3].

The need for good approximation methods for the compound Poisson distribution has been mentioned above. There are some studies on the rate of convergence to compound Poisson distributions, see for instance Serfozo [22] or Wang [27]. In a paper by Arratia, Goldstein and Gordon [2], a method involving "declumping" is derived via the Stein-Chen method for sums of indicators .

In a paper from 1992, Barbour, Chen and Loh [4] generalized Stein's method by considering a compound Poisson identity analogous to (1.1). This avoided the process of "declumping". The method is not completely analogous to the Poisson case, since the Stein Operator associated with the compound Poisson identity is not necessarily the generator of a Markov process, and error bounds are not sharp (see [4]). In two papers by Roos [18, 19], the method was further developed, and applied to some specific problems outside the area of extreme value theory. Here we can also distinguish two different variants, the "local" approach and the "coupling" approach, analogous to the Poisson case.

As a final remark on Stein's method, we note that in most papers, the distance between the true and the approximating distribution is measured by the total variation distance. For two integer random variables, X and Y, we define this measure by

$$d(X,Y) = \sup_{A \subseteq N} |P(X \in A) - P(Y \in A)| = \frac{1}{2} \sum_k |P(X = k) - P(Y = k)|.$$

Hence the total variation distance has the convenient interpretation of measuring the amount of the "mass" of the distribution of X that must be moved to "construct" the distribution of Y. There are other measures, e.g. the Wasserstein and Hellinger distances, used in connection with Stein's method, see e.g. Barbour, Holst and Janson [3]. In the present thesis we will only consider the total variation distance.

1.1 An outline of the thesis

The thesis can be viewed as consisting of three parts. The first part treats Poisson approximation of the number of exceedances in χ^2–sequences. The next part improves some earlier results on Poisson approximation for Gaussian sequences. The final part, also treating Gaussian sequences, generalizes the results to approximations with the compound Poisson distribution.

In the *first part*, Chapter 2, the coupling version of the Stein-Chen method is used to investigate the properties of the Poisson approximation for χ^2–sequences. A prerequisite for the method to produce useful results is that the existence of a monotone coupling can be proved. After some basic definitions and notation introduced in Section 2.1, a proof of the existence of the coupling is given in Section 2.2, where a bound for the total variation distance between the true distribution of W, the number of exceeding points, and an approximating Poisson distribution is also derived. The bound consists only of first and second order moments of indicator random variables, one for each point of time, having the value one if an exceedance occurs at that particular time, and zero otherwise. These covariances (second order moments) can, at least in principle, be bounded, and thus the rate of convergence for the Poisson approximation can be determined. In Section 2.3 the convergence rate for $\chi^2(2)$–sequences is investigated. Finally some numerical results on the bound of the total variations distance are presented in Section 2.4.

The technique in Chapter 2 for approximating W is not restricted to χ^2–sequences. It is based on the fact that a χ^2–sequence can be represented as a *monotone* function of underlying Gaussian sequences. This is done by splitting the function into a sum of terms, each one of them being monotone. A more detailed explanation is found in Section 2.1.

In order for the text to have a reasonable length, some straightforward calculations in Section 2.3 have been omitted. A more detailed derivation of these results can be found in Raab [16].

The *second part*, chapter 3, contains some improvements on Poisson approximation for Gaussian sequences. The first section starts by showing that the bound for the rate of convergence can be improved for some sequences. In the same section we also show that this is only true for sequences with strong "global" dependence. For sequences with "local" dependence only, a lower bound is derived and shown to be of the same order as the upper bound, thus giving the right convergence rate.

Similarly to the case of χ^2–sequences, knowledge of the existence of a monotone coupling for Gaussian sequences is useful for Poisson approximation with the coupling version of the Stein-Chen method. In Section 3.2, we construct such a coupling explicitly for Gaussian sequences.

In the *third part*, Chapter 4, we continue to consider exceedances of Gaussian sequences, but from a somewhat different aspect. This chapter treats the fairly new area of compound Poisson approximation. The compound Poisson distribution, having the Poisson distribution as a special case, can be used to improve the approximation for sequences with strong "local" dependence, i.e. sequences where "clumping" of exceeding points occurs. Throughout the whole chapter, we assume that the sequence is m-dependent, i.e. the dependence among the variables only ranges over a finite number of points. This excludes sequences with “global” dependence.

In Section 4.1 the reader is introduced to Stein’s method for compound Poisson approximation, and some general theorems are stated. The topic of Section 4.2 is how to approximate the number of exceeding points with a suitable compound Poisson distribution. However, the parameters of this distribution are, in general, difficult to compute. It is therefore interesting to see if a “simpler” (in the sense defined in the beginning of Section 4.3) compound Poisson distribution can be used without much loss of speed of convergence compared to a general compound Poisson distribution. Section 4.3 is devoted to this. In the last section of Chapter 4, Section 4.4, numerical results for some MA-sequences are presented.

In Chapter 5, some general aspects on Poisson and compound Poisson approximation in the context of exceeding points in Gaussian and related sequences are discussed. Especially, the assumption of stationarity and, for Chapter 4, the restriction to m-dependent sequences are considered, and possible generalization of the results is suggested.

A slightly revised version of the content in Sections 2.1–2.3 will appear in *Stochastic Processes and their Applications*, see [15].

1.2 A popular description of the thesis

In many different areas in real life, extreme values are very important. For instance, extreme deviations from typical values for mechanical details could mean that a part fails, or at least tends to malfunction. Extreme loads on bridges can cause cracks to grow. Unusual and extreme conditions in nuclear reactors can cause security systems to activate. Extreme pulses in power supply networks can lead to malfunctioning of complicated control processes. These examples show that methods for estimating probabilities for different extreme events are important. However, a troublesome feature of extreme values is that they are, by definition, rare. Collecting data concerning extreme values is therefore often difficult, and thus the need for theoretical models is great.

Suppose as an example that we record the air temperature in Stockholm at noon every day. The values may look like the ones in Figure 1.1.

Now, this data series contains a lot of information. In this example let us assume we are only interested in the number of days with very high temperatures, e.g. temperatures exceeding $26°C$ (about $79°F$). Fix therefore a level at that temperature, as indicated in the figure, and count the points where the value exceeds that level. Denote this random number by W. From the figure, we see that from the 1st of June to the 15th of August, 7 days had temperatures above $26°C$, i.e. $W = 7$. Assume that it is important to find methods for calculating the probability of say y days or more of extreme temperature in a future month. In order to be able to do that, we first have to make certain assumptions.

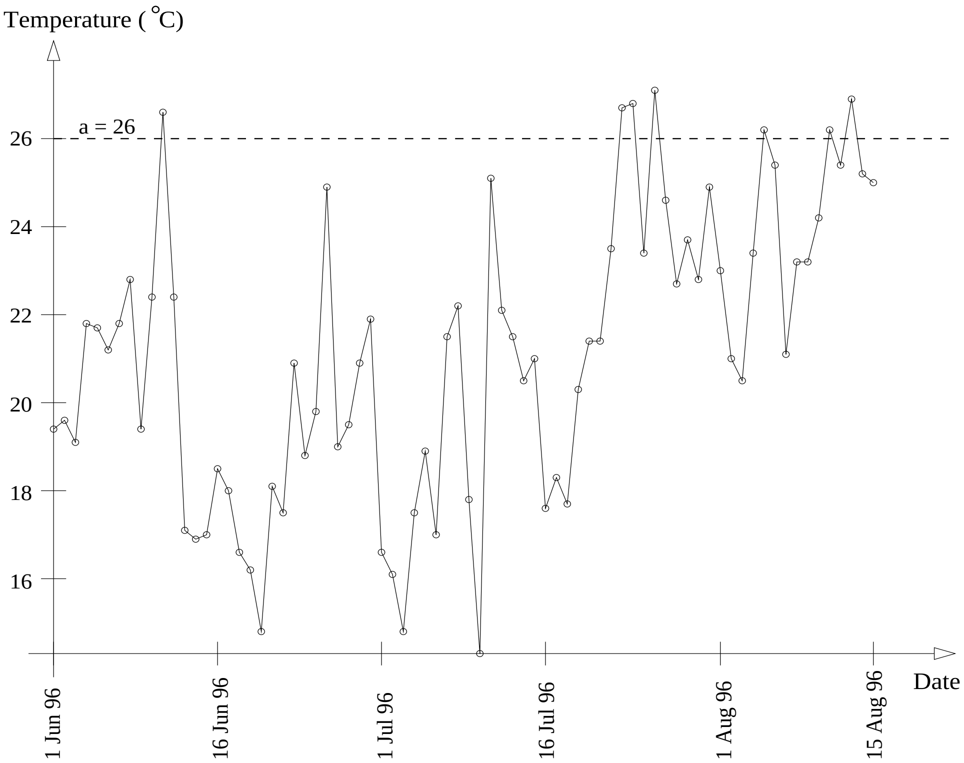

Figure 1.1: *Maximum day temperatures i Stockholm 1st Jun 96 – 15th Aug 96.*

First of all, we have to specify how likely the different temperatures are. Obviously, this will depend on geographic conditions. For instance, countries with an inland climate will probably experience more days with extreme temperature than countries with a coastal climate. Moreover, Spain will most likely have more days with temperatures above $26°C$ than Sweden. To summarize: we must specify the distribution of data. Very often the data series is assumed to be a Gaussian sequence, implying that all data are generated from the normal (Gaussian) distribution. Often stationarity is also assumed, which, among other things, means that trend or seasonal variations should be eliminated from the data. Suppose therefore that we only measure the temperature during summer.

Secondly, we have to specify the dependence structure between values from different days. We all know that the values are correlated: If the temperature has been very high for a couple of days, it is likely that we have a high pressure, and consequently a large probability of high temperature tomorrow too. On the other hand, several days with low temperatures are likely to be followed by more days with low temperatures (or at least not extremely high temperatures).

Let us start with the simplest possible model, assuming independence between the temperatures for different days. This means that we assume that the weather varies from day to day in an irregular manner, regardless of the weather the day before. From elementary statistics, we know that W is then *binomially* distributed, a well-known distribution from which we can compute any probability of interest, e.g. the probability of having more that 5 days with extreme temperature in the next 2 months.

From a meteorological point of view, the assumption of independence is of course unsatisfactory. We must include some kind of dependence structure in the model. This makes it much more difficult to compute the desired probabilities exactly. Instead, we will make use of approximations valid only for large data sets, so called *asymptotic approximations.*

An important theorem known at least since 1837, see Poisson [14], is that the binomial distribution tends to the *Poisson* distribution when the data set is large and the events considered are rare. This distribution is often used to describe the number of events occurring in some random process, where the events occur more or less independently of each other. The question for us is if this distribution also can be used for a more complex dependence structure when W is not binomial distributed.

Let us return for a moment to Figure 1.1. There we see a phenomenon which is typical for positively correlated (interacting) events. Days with extreme temperatures tend to cluster (see for instance the period 23. – 26. July). We do not only get single extreme days, but rather pairs of days, or even larger "clumps", with extreme temperatures. This phenomenon is *not* taken care of in the model with the Poisson distribution. If the dependence between adjacent days is so weak that the phenomenon is negligible then the Poisson distribution is still suitable. However, for strongly dependent values, the Poisson approximation is poor, and a better approximation is needed. Such a refinement can be achieved by the compound Poisson distribution. The idea behind the use of this distribution is as follows. Instead of counting the number of extreme points one by one, we count the number of *single* points, the number of "clumps" of size two, the number of clumps of size three and so on. In Figure 1.1, we have five single points and one clump of size two. The number of occurrences of each group of specified size can then be approximated by the Poisson distribution if the clumps occur sparsely, and therefore are more or less independent of each other. The number of days belonging to clumps of a certain size is easily calculated by multiplying the size of clumps with the number of points in that group. Finally, the desired total number of days with extreme temperature is computed by taking the sum of all the groups. This model might be unsatisfactory for meteorological forecasts, but it serves here to give an idea of possible applications.

Results on Poisson approximation for Gaussian sequences can be found in the literature, see e.g. Holst and Janson [9]. In the present thesis we will extend these result in two ways. In the first part of the thesis, we show how Poisson approximations can be performed for more general sequences than the Gaussian; particularly so called χ^2–*sequences* are investigated. In the latter part of the thesis we return to Gaussian sequences but develop the technique for taking care of the "clumping" as described above. The method will prove superior to ordinary Poisson approximation for all Gaussian sequences (except when independence is assumed so that no clumping occurs and the methods are equally good).

Once knowing that the approximation is good, we can use it, for instance, to compute approximately how probable no extreme temperatures during the next two months are, the probability of the case that at least four days of extreme temperatures will take place during the third month, and so on.

Chapter 2

Poisson approximation for χ^2–sequences

In this chapter, the coupling approach of the Stein-Chen method is used. After some preliminaries in Section 2.1, we derive in Section 2.2 a theorem proving the existence of a monotone coupling. This theorem is used to derive an upper bound for the distance between a Poisson distribution and the true distribution of the number of extreme points. In Section 2.3, the theorem is specialized to a $\chi^2(2)$–sequences and the rate of convergence to a Poisson distribution is given explicitly. Finally, Section 2.4 is devoted to some numerical results for $\chi^2(1)$– and $\chi^2(2)$–sequences.

2.1 Preliminaries

Let $(X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)})$, $(X_1^{(2)}, X_2^{(2)}, \ldots, X_n^{(2)})$, $\ldots$, $(X_1^{(f)}, X_2^{(f)}, \ldots, X_n^{(f)})$ be f independent and identically distributed standardized Gaussian sequences, i.e. Gaussian sequences with $E(X_k^{(i)}) = 0$, $V(X_k^{(i)}) = 1$, $i = 1, \ldots, f$, $k = 1, \ldots, n$. Set

$$Y_k = \sum_{i=1}^{f} (X_k^{(i)})^2.$$

The sequence $\{Y_k\}$ will be referred to as a χ^2–sequence since all the Y_k's are χ^2–distributed with f degrees of freedom.

Let $a_n > 0$ be a fixed positive number. Introduce indicators $\{I_k\}$ counting the number of exceedances of a_n^2 by the sequence $\{Y_k\}$,

$$I_k = \begin{cases} 1 & \text{if } Y_k > a_n^2 \\ 0 & \text{otherwise.} \end{cases}$$

Then $W = \sum_{k=1}^n I_k$ is the number of exceedances. Later we will let a_n and hence $I_k = I_k^n$ and $W = W_n$ depend on n. For the sake of simplicity, we will leave out the subscript n on the indicators as well as on W.

The idea is to show the existence of a certain monotone coupling for these indicators. Sharp bounds of the variation distance between W and a Poisson distributed random variable can then be derived using the coupling approach of the Stein-Chen method. In Section 2.2 we show the existence of a monotone coupling for increasing functions of Gaussian random variables. However, $Y_k = \sum_{i=1}^f (X_k^{(i)})^2$ is *not* an increasing function of $(X_k^{(1)}, \ldots, X_k^{(f)})$. We therefore introduce additional random variables and indicators, so that Y_k becomes an increasing function of these new variables. Define for each of the 2^f different quadrants in R^f,

$${}_jI_k = \begin{cases} 1 & \text{if } Y_k > a_n^2 \text{ and observation } k \text{ falls in quadrant } j \\ 0 & \text{otherwise,} \end{cases}$$

where we have assumed that the quadrants have been numbered in a suitable way (see Figure 2.1). More formally,

$${}_jI_k = I\Big\{h({}_jX_k^{(1)}, \ldots, {}_jX_k^{(f)}) > a_n^2\Big\}, \tag{2.1}$$

where the function h is defined by

$$h(x_1, x_2, \ldots, x_f) = (x_1^2 + x_2^2 + \ldots + x_f^2) I(x_1 \geq 0, x_2 \geq 0, \ldots, x_f \geq 0)$$

and

$$({}_jX_k^{(1)}, \ldots, {}_jX_k^{(f)}) = (\pm X_k^{(1)}, \ldots, \pm X_k^{(f)}), \qquad k = 1, \ldots, n,$$

so that each of the 2^f combinations of signs on the right hand side is represented when j runs over the different quadrants. The sum $\sum_{k=1}^n {}_jI_k$, which is increasing in $({}_jX_k^{(1)}, \ldots, {}_jX_k^{(f)})$, $k = 1, \ldots, n$, counts the number of exceeding points in quadrant j. Clearly we have

$$I_k = \sum_{j=1}^{2^f} {}_jI_k \quad \text{and} \quad W = \sum_{j=1}^{2^f} \sum_{k=1}^{n} {}_jI_k. \tag{2.2}$$

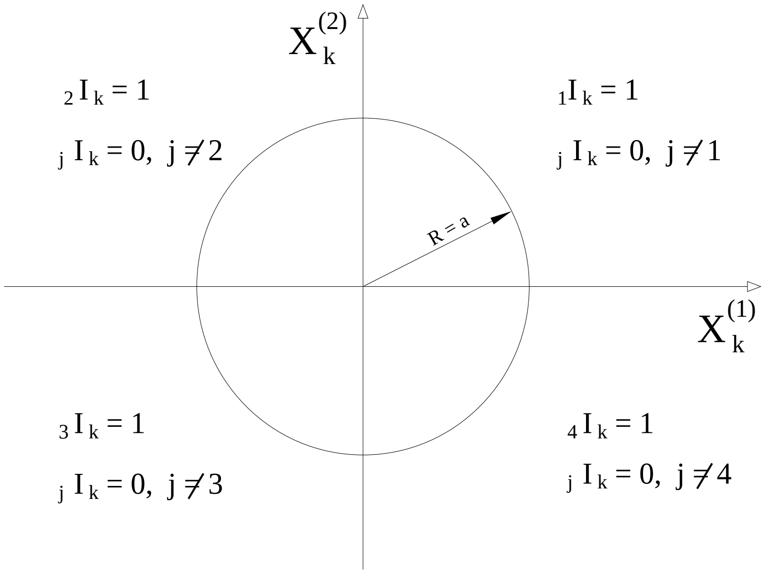

Figure 2.1: *Illustration of the definition of the indicators in two dimensions, corresponding to a $\chi^2(2)$–sequence.*

2.2 The Stein-Chen method and coupling

We begin this section by stating a general theorem for Poisson convergence under the assumption that there is a monotone coupling of the indicators. After that we prove that such a coupling really exists when considering extreme values of increasing functions of Gaussian random variables. Finally, these results are combined and applied to the number of extreme points of a χ^2–sequence. An upper bound for the total variation distance between the distribution of W and a Poisson distributed random variable with mean $E(W)$ is stated in terms of second order moments of the indicators.

Let Γ be a finite index set and put $\Gamma_i = \Gamma \backslash \{i\}$, $i \in \Gamma$. Let I_i, $i \in \Gamma$ be Bernoulli random variables with $P(I_i = 1) = 1 - P(I_i = 0) = p_i$, and set

$$W = \sum_{i \in \Gamma} I_i, \qquad \lambda = E(W) = \sum_{i \in \Gamma} p_i.$$

Define the total variation distance between two non-negative integer-valued random variables W and X as

$$d(W, X) = \sup_{A \subseteq N} |P(W \in A) - P(X \in A)|,$$

and let $\mathrm{Po}(\lambda)$ denote a random variable that is Poisson distributed with mean λ.

Theorem 2.1 *Let, for each $i \in \Gamma$, the random variables I_i, I_j, J_{ji}, $j \in \Gamma_i$ be defined on the same probability space with*

$$\mathcal{L}(J_{ji}\,;j \in \Gamma_i) = \mathcal{L}(I_j\,;j \in \Gamma_i \mid I_i = 1).$$

Assume that there exists a partition $\Gamma_i = \Gamma_i^+ \cup \Gamma_i^-$ with $J_{ji} \geq I_j$ for $j \in \Gamma_i^+$ and $J_{ji} \leq I_j$ for $j \in \Gamma_i^-$. Then

$$d(W, \mathrm{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \left\{ \sum_{i \in \Gamma} p_i^2 + \sum_{i \in \Gamma} \sum_{j \in \Gamma_i} |\mathrm{Cov}(I_i, I_j)| \right\}.$$

A more general form of the theorem, where not the whole Γ_i is assumed to be partitioned by Γ_i^+ and Γ_i^-, including a proof, can be found in Holst and Janson [9] or Barbour, Holst and Janson [3], Section 2.1.

As follows from the theorem, it is not necessary to give an explicit construction of the coupling. It suffices to know the existence of a suitable one. The next proposition, which is a generalization of a similar proposition in Holst and Janson [9], will prove useful.

Proposition 2.2 *Let* $(X_1^{(i)}, \ldots, X_n^{(i)})$, $i = 1, \ldots, f$, *be* f *independent copies of a Gaussian random vector with* $\mathrm{Cov}(X_1^{(i)}, X_k^{(i)}) \geq 0$ *for* $2 \leq k \leq p$ *and* $\mathrm{Cov}(X_1^{(i)}, X_k^{(i)}) \leq 0$ *for* $p < k \leq n$. *Let* h *be a real function, increasing in each argument. Then, for any* a, *there exists a probability space with random variables* $Y_2^{(1)}, \ldots, Y_n^{(f)}, Z_2^{(1)}, \ldots, Z_n^{(f)}$, *such that*

$$\begin{aligned}\mathcal{L}\left(Y_2^{(1)}, \ldots, Y_n^{(f)}\right) &= \mathcal{L}\left(X_2^{(1)}, \ldots, X_n^{(f)}\right), \\ \mathcal{L}\left(Z_2^{(1)}, \ldots, Z_n^{(f)}\right) &= \mathcal{L}\left(X_2^{(1)}, \ldots, X_n^{(f)} \,|\, h(X_1^{(1)}, \ldots, X_1^{(f)}) > a\right)\end{aligned}$$

and $Y_k^{(i)} \leq Z_k^{(i)}$, $2 \leq k \leq p$; $Y_k^{(i)} \geq Z_k^{(i)}$, $p < k \leq n$, *for* $i = 1, \ldots, f$.

Proof. Set

$$^*X_k^{(i)} = \begin{cases} X_k^{(i)} & k = 1, \ldots, p \\ -X_k^{(i)} & k = p+1, \ldots, n. \end{cases}$$

Let g be an arbitrary increasing function. Since both h and the indicator function $I(\,\cdot > a)$ are increasing functions, the composition $I(h(\cdot) > a)$ will also be increasing. Moreover, as all the covariances between $^*X_1^{(i)}$ and $^*X_k^{(j)}$ are non-negative, it follows from Corollary 3 of Joag-Dev, Perlman and Pitt [10] that

$$\begin{aligned}
& E\Big\{I\Big(h(^*X_1^{(1)}, \ldots, {}^*X_1^{(f)}) > a\Big)\Big\} \cdot E\Bigg\{I\Big(h(^*X_1^{(1)}, \ldots, {}^*X_1^{(f)}) > a\Big) \\
& \times g(^*X_2^{(1)}, \ldots, {}^*X_n^{(1)}, {}^*X_2^{(2)}, \ldots, {}^*X_n^{(f)}) \,|\, I\Big(h(^*X_1^{(1)}, \ldots, {}^*X_1^{(f)}) > a\Big) = 1\Bigg\} \\
& = E\Big\{I\Big(h(^*X_1^{(1)}, \ldots, {}^*X_1^{(f)}) > a\Big) \cdot g(^*X_2^{(1)}, \ldots, {}^*X_n^{(1)}, {}^*X_2^{(2)}, \ldots, {}^*X_n^{(f)})\Big\} \\
& \geq E\Big\{I\Big(h(^*X_1^{(1)}, \ldots, {}^*X_1^{(f)}) > a\Big)\Big\}\, E\Big\{g(^*X_2^{(1)}, \ldots, {}^*X_n^{(1)}, {}^*X_2^{(2)}, \ldots, {}^*X_n^{(f)})\Big\}.
\end{aligned}$$

This is equivalent to

$$\begin{aligned}
& E\Big\{g(^*X_2^{(1)}, \ldots, {}^*X_n^{(1)}, {}^*X_2^{(2)}, \ldots, {}^*X_n^{(f)}) \,|\, h(^*X_1^{(1)}, \ldots, {}^*X_1^{(f)}) > a\Big\} \\
& \qquad \geq E\Big\{g(^*X_2^{(1)}, \ldots, {}^*X_n^{(1)}, {}^*X_2^{(2)}, \ldots, {}^*X_n^{(f)})\Big\}.
\end{aligned}$$

According to Kamae, Krengel and O'Brien [11], Theorem 1, this implies the existence of a probability space with random variables $Y_2^{(1)},\ldots,Y_n^{(f)}$, $Z_2^{(1)},\ldots,Z_n^{(f)}$, such that

$$\begin{aligned}
&\mathcal{L}\Big(Y_2^{(1)},\ldots,Y_p^{(1)},-Y_{p+1}^{(1)},\ldots,-Y_n^{(1)},\ldots,Y_2^{(f)},\ldots,Y_p^{(f)},-Y_{p+1}^{(f)},\ldots,-Y_n^{(f)}\Big)\\
&\quad=\mathcal{L}\left({}^*X_2^{(1)},\ldots,{}^*X_n^{(1)},\ldots,{}^*X_2^{(f)},\ldots,{}^*X_n^{(f)}\right),\\
&\mathcal{L}\Big(Z_2^{(1)},\ldots,Z_p^{(1)},-Z_{p+1}^{(1)},\ldots,-Z_n^{(1)},\ldots,Z_2^{(f)},\ldots,Z_p^{(f)},-Z_{p+1}^{(f)},\ldots,-Z_n^{(f)}\Big)\\
&\quad=\mathcal{L}\left({}^*X_2^{(1)},\ldots,{}^*X_n^{(1)},\ldots,{}^*X_2^{(f)},\ldots,{}^*X_n^{(f)}\,|\,h({}^*X_1^{(1)},\ldots,{}^*X_1^{(f)})>a\right)
\end{aligned}$$

and $Y_k^{(i)}\le Z_k^{(i)}$, $2\le k\le p$; $-Y_k^{(i)}\le -Z_k^{(i)}$, $p<k\le n$, for $i=1,\ldots,f$, from which the statement follows. □

Proposition 2.2 combined with the fact that the indicators ${}_jI_k$ are increasing can now be used in Theorem 2.1, since $Y\le Z$ implies $I(Y>a)\le I(Z>a)$.

Theorem 2.3 *Suppose $(X_1^{(1)},\ldots,X_n^{(1)})$,…,$(X_1^{(f)},\ldots,X_n^{(f)})$ are f independent and identically distributed standardized Gaussian sequences. Let $Y_k=\sum_{i=1}^f (X_k^{(i)})^2$ for $k=1,\ldots,n$ and let the indicator random variables ${}_jI_k$, $j=1,\ldots,2^f$, $k=1,\ldots,n$, be defined by equation (2.1), where a_n is a real positive number. Finally, define W by equation (2.2) and let $\lambda=E(W)=n\cdot P(Y_1>a_n^2)$. Then*

$$d(W,\mathrm{Po}(\lambda))\le\frac{1-e^{-\lambda}}{\lambda}\Bigg\{nP(Y_1>a_n^2)^2+\sum_{i=1}^{2^f}\sum_{j=1}^{2^f}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\ne k}}^{n}|\mathrm{Cov}({}_jI_k,\,{}_iI_l)|\Bigg\}.$$

Theorem 2.3 can be used in two ways. In Subsection 2.3.2, we will use it together with approximations of the covariances in order to obtain the rate of convergence for the Poisson convergence for a $\chi^2(2)$–sequence. The theorem can also be combined with numerical calculations of the covariances to get an upper limit for $d(W,\mathrm{Po}(\lambda))$ for finite n, see Subsection 2.4.

Proof. Let $\Gamma = \{(j,k) : j = 1,\ldots,2^f,\ k = 1,\ldots,n\}$. For a fixed element $(j,k) \in \Gamma$, the existence of a monotone coupling in Theorem 2.1 follows from Proposition 2.2 applied to ${}_1X_1^{(i)},\ldots,\ {}_{2^f}X_n^{(i)}$, $i = 1,\ldots,f$.

Next note that ${}_iI_k \cdot {}_jI_k = 0$, $i \neq j$. The assertion now follows from Theorem 2.1 with indicators defined by (2.1) since

$$\sum_{i\in\Gamma} p_i^2 + \sum_{i\in\Gamma}\sum_{j\in\Gamma_i} |\mathrm{Cov}(I_i, I_j)|$$

$$= \sum_{j=1}^{2^f}\sum_{k=1}^{n} (E({}_jI_k))^2 + \sum_{i=1}^{2^f}\sum_{j=1}^{2^f}\sum_{\substack{k=1\\ \{j,l\}\neq\{i,k\}}}^{n}\sum_{l=1}^{n} |\mathrm{Cov}({}_jI_k,\ {}_iI_l)|$$

$$= 2^f n\,(E({}_1I_1))^2 + 2^f(2^f-1)n\,(E({}_1I_1))^2 + \sum_{i=1}^{2^f}\sum_{j=1}^{2^f}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n} |\mathrm{Cov}({}_jI_k,\ {}_iI_l)|$$

$$= n\cdot\left(P(Y_1 > a_n^2)\right)^2 + \sum_{i=1}^{2^f}\sum_{j=1}^{2^f}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n} |\mathrm{Cov}({}_jI_k,\ {}_iI_l)|\ .$$

□

Holst and Janson [9] deals with the number of exceeding points in a Gaussian sequence $\{X_k\}$, both in the "one-sided" case, $I(X_k > a_n)$, and in the "double-sided" case, $I(|X_k| > a_n)$. Since $I(|X_k| > a_n) = I(X_k^2 > a_n^2)$, the Poisson convergence of the extreme points of a $\chi^2(1)$–sequence is an immediate consequence of the results by Holst and Janson.

In the next section Theorem 2.3 will be used in order to get an upper limit for $d(W, \mathrm{Po}(\lambda))$ in the case of a $\chi^2(2)$–sequence. In principle, the case of a $\chi^2(f)$–sequence, where $f \geq 3$, can be treated in a similar way, although the amount of work will increase rapidly without adding any new basic understanding.

2.3 Poisson convergence in the $\chi^2(2)$ case

In Subsection 2.3.2, we state a main result, Theorem 2.7. First, in Subsection 2.3.1, we derive some estimates of the covariances needed for proving the theorem.

2.3.1 Asymptotic behavior of the covariances

Let $(X_1^{(1)},\ X_2^{(1)})$ and $(X_1^{(2)},\ X_2^{(2)})$ be two independent and identically distributed bivariate random variables with

$$\left(X_1^{(i)},\ X_2^{(i)}\right) \in N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right), \qquad i=1,2.$$

Let a be a real positive number and set for $i=1,2$,

$$I_i = I\left\{\left((X_i^{(1)})^2 + (X_i^{(2)})^2\right) I\left(X_i^{(1)} \geq 0, X_i^{(2)} \geq 0\right) > a^2\right\}. \qquad (2.3)$$

Finally, let φ and Φ be the density function and the cumulative distribution function respectively for a standardized, univariate Gaussian distribution.

Lemma 2.4 *Let I_1 and I_2 be defined by (2.3). Assume $|\rho|<1$. Then, for $\rho \geq 0$,*

$$0 \leq \mathrm{Cov}(I_1, I_2) \leq \frac{1}{16}\exp\left(-\frac{a^2}{2}\right)\exp\left(-\frac{a^2}{2}\frac{1-\rho}{1+\rho}\right) \times\left(1+\frac{2\rho}{1-\rho}+\frac{\rho}{a^2}\frac{1+\rho}{1-\rho}\right) - \frac{1}{16}\exp\left(-\frac{a^2}{2}\right)\exp\left(-\frac{a^2}{2}\right),$$

and for $\rho \leq 0$,

$$0 \leq -\mathrm{Cov}(I_1, I_2) \leq \frac{1}{16}\exp\left(-\frac{a^2}{2}\right)\left\{|\rho|\left(2+\frac{1-\rho^2}{a^2}\right) \times\exp\left(-\frac{a^2}{2(1-\rho^2)}\right) + \exp\left(-\frac{a^2}{2}\right) - (1-\rho^2)\exp\left(-\frac{a^2}{2}\frac{1-\rho}{1+\rho}\right)\right\}.$$

Proof. By symmetry,

$$E(I_1) = E(I_2) = \frac{1}{4}P((X_1^{(1)})^2 + (X_1^{(2)})^2 > a^2) = \frac{1}{4}\exp\left(-\frac{a^2}{2}\right). \qquad (2.4)$$

Since $X_2^{(i)} \mid X_1^{(i)} = x \in N(\rho x, 1-\rho^2)$, $i = 1, 2$, conditioning on $X_1^{(1)}$ and $X_1^{(2)}$ gives,

$$P\left(\left(\left(X_2^{(1)}\right)^2 + \left(X_2^{(2)}\right)^2\right) I\{X_2^{(1)} \geq 0, X_2^{(2)} \geq 0\} > a^2 \mid X_1^{(1)} = x, X_1^{(2)} = y\right)$$

$$= \iint_{\substack{u \geq 0,\ v \geq 0 \\ u^2+v^2 \geq a^2}} \frac{1}{2\pi} \frac{1}{(1-\rho^2)} \exp\left(-\frac{(u-\rho x)^2}{2(1-\rho^2)}\right) \exp\left(-\frac{(v-\rho y)^2}{2(1-\rho^2)}\right) du\, dv$$

$$= \int_a^\infty \int_0^{\pi/2} \frac{1}{2\pi} \frac{1}{(1-\rho^2)} \exp\left(-\frac{(r\cos\alpha - \rho x)^2 + (r \sin\alpha - \rho y)^2}{2(1-\rho^2)}\right) r\, d\alpha dr$$

$$= \frac{1}{2\pi} \int_0^{\pi/2} \left\{ \frac{\sqrt{2\pi}(\rho x \cos\alpha + \rho y \sin\alpha)}{\sqrt{1-\rho^2}} \left(1 - \Phi\left(\frac{a - (\rho x\cos\alpha + \rho y \sin\alpha)}{\sqrt{1-\rho^2}}\right)\right)\right.$$

$$\left. + \exp\left(-\frac{(a - (\rho x \cos\alpha + \rho y \sin\alpha))^2}{2(1-\rho^2)}\right)\right\} \exp\left(-\frac{(\rho x \sin\alpha - \rho y \cos\alpha)^2}{2(1-\rho^2)}\right) d\alpha.$$

Consequently,

$$E(I_1 I_2) = \iint_{\substack{x \geq 0,\ y \geq 0 \\ x^2+y^2 \geq a^2}} f_{X_1^{(1)},\ X_1^{(2)}}(x,y) P\left\{ \left(\left(X_2^{(1)}\right)^2 + \left(X_2^{(2)}\right)^2\right)\right.$$

$$\left. \times I\{X_2^{(1)} \geq 0, X_2^{(2)} \geq 0\} > a^2 \mid X_1^{(1)} = x, X_1^{(2)} = y \right\} dx dy$$

$$= \frac{1}{4\pi^2} \int_0^{\pi/2} \int_0^{\pi/2} \int_a^\infty \exp\left(-\frac{\rho^2 s^2 \sin^2(\alpha-\beta) + s^2(1-\rho^2)}{2(1-\rho^2)}\right)$$

$$\times \left\{ \frac{\sqrt{2\pi} s^2 \rho \cos(\alpha-\beta)}{\sqrt{1-\rho^2}} \left(1 - \Phi\left(\frac{a - \rho s\cos(\alpha-\beta)}{\sqrt{1-\rho^2}}\right)\right)\right.$$

$$\left. + s\exp\left(-\frac{(a-\rho s\cos(\alpha-\beta))^2}{2(1-\rho^2)}\right)\right\} ds\, d\alpha\, d\beta.$$

After integration by parts, the expression above can be rewritten as

$$E(I_1 I_2) = \frac{1}{4\pi^2} \int_0^{\pi/2} \int_0^{\pi/2} \frac{\sqrt{2\pi(1-\rho^2)}\rho\cos(\alpha-\beta)}{1-\rho^2\cos^2(\alpha-\beta)}$$

$$\times \left\{ a \exp\left(-\frac{a^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right) \left(1 - \Phi\left(\frac{a(1-\rho\cos(\alpha-\beta))}{\sqrt{1-\rho^2}}\right)\right)\right.$$

$$+\int_a^\infty \exp\left(-\frac{s^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right)\left(1-\Phi\left(\frac{a-\rho s\cos(\alpha-\beta)}{\sqrt{1-\rho^2}}\right)\right)\,ds\Bigg\}$$
$$+\frac{1}{1-\rho^2\cos^2(\alpha-\beta)}$$
$$\times\int_a^\infty s\exp\left(-\frac{(s-a\rho\cos(\alpha-\beta))^2+a^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right)\,ds\,d\alpha\,d\beta$$
$$=\frac{1}{4\pi^2}\int_0^{\pi/2}\int_0^{\pi/2}\frac{\sqrt{(1-\rho^2)}}{1-\rho^2\cos^2(\alpha-\beta)}\Bigg\{2\sqrt{2\pi}\rho a\cos(\alpha-\beta)$$
$$\times\exp\left(-\frac{a^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right)\left(1-\Phi\left(\frac{a(1-\rho\cos(\alpha-\beta))}{\sqrt{1-\rho^2}}\right)\right)$$
$$+\sqrt{(1-\rho^2)}\exp\left(-\frac{a^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right)\exp\left(-\frac{a^2(1-\rho\cos(\alpha-\beta))^2}{2(1-\rho^2)}\right)$$
$$+\sqrt{2\pi}\rho\cos(\alpha-\beta)\int_a^\infty\exp\left(-\frac{s^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right)$$
$$\times\left(1-\Phi\left(\frac{a-\rho s\cos(\alpha-\beta)}{\sqrt{1-\rho^2}}\right)\right)\,ds\Bigg\}\,d\alpha\,d\beta. \tag{2.5}$$

Assume now that $\rho\geq 0$. Integrating by parts again and using the fact that $(1-\Phi(x))/\varphi(x)$ is a decreasing function for $x>0$, the last integral in (2.5) can be rewritten as

$$\int_a^\infty \exp\left(-\frac{s^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right)\left(1-\Phi\left(\frac{a-\rho s\cos(\alpha-\beta)}{\sqrt{1-\rho^2}}\right)\right)\,ds$$
$$\leq\frac{\sqrt{2\pi}\sqrt{1-\rho^2}}{\sqrt{1-\rho^2\cos^2(\alpha-\beta)}}\left(1-\Phi\left(a\sqrt{\frac{1-\rho^2\cos^2(\alpha-\beta)}{1-\rho^2}}\right)\right)$$
$$\times\Bigg\{\left(1-\Phi\left(\frac{a(1-\rho\cos(\alpha-\beta))}{\sqrt{1-\rho^2}}\right)\right)+\frac{\rho\cos(\alpha-\beta)}{\sqrt{1-\rho^2}}$$
$$\times\left(\varphi\left(a\sqrt{\frac{1-\rho^2\cos^2(\alpha-\beta)}{1-\rho^2}}\right)\right)^{-1}\int_a^\infty\varphi\left(s\sqrt{\frac{1-\rho^2\cos^2(\alpha-\beta)}{1-\rho^2}}\right)$$

$$\times\varphi\left(\frac{a-\rho s\cos(\alpha-\beta)}{\sqrt{1-\rho^2}}\right)\,ds\Bigg\}$$

$$= \frac{\sqrt{2\pi}\sqrt{1-\rho^2}(1+\rho\cos(\alpha-\beta))}{\sqrt{1-\rho^2\cos^2(\alpha-\beta)}}\left(1-\Phi\left(a\sqrt{\frac{1-\rho^2\cos^2(\alpha-\beta)}{1-\rho^2}}\right)\right)$$

$$\times\left(1-\Phi\left(\frac{a(1-\rho\cos(\alpha-\beta))}{\sqrt{1-\rho^2}}\right)\right). \tag{2.6}$$

Replacing the last integral in (2.5) by (2.6) and using $1-\Phi(x)\le\varphi(x)/x$ for $x>0$, elementary approximations give

$$E(I_1I_2)\le\frac{1}{16}\Bigg\{\frac{2\sqrt{2\pi}\sqrt{1-\rho^2}\rho a}{1-\rho^2}\exp\left(-\frac{a^2}{2}\right)\frac{1}{\sqrt{2\pi}}\sqrt{\frac{1+\rho}{1-\rho}}\frac{1}{a}\exp\left(-\frac{a^2}{2}\frac{1-\rho}{1+\rho}\right)$$

$$+\exp\left(-\frac{a^2}{2}\right)\exp\left(-\frac{a^2}{2}\frac{1-\rho}{1+\rho}\right)+\frac{\sqrt{2\pi}\sqrt{1-\rho^2}\rho}{1-\rho^2}\frac{\sqrt{2\pi}\sqrt{1-\rho^2}(1+\rho)}{\sqrt{1-\rho^2}}$$

$$\times\frac{1}{\sqrt{2\pi}}\frac{1}{a}\exp\left(-\frac{a^2}{2}\right)\frac{1}{\sqrt{2\pi}}\sqrt{\frac{1+\rho}{1-\rho}}\frac{1}{a}\exp\left(-\frac{a^2}{2}\frac{1-\rho}{1+\rho}\right)\Bigg\}$$

$$=\frac{1}{16}\exp\left(-\frac{a^2}{2}\right)\exp\left(-\frac{a^2}{2}\frac{1-\rho}{1+\rho}\right)\left(1+\frac{2\rho}{1-\rho}+\frac{\rho}{a^2}\frac{1+\rho}{1-\rho}\right),$$

which combined with (2.4) prove the first part of the proposition.

Next we turn to the case $\rho\le 0$. Note that $(a-\rho s\cos(\alpha-\beta))/\sqrt{1-\rho^2}$ is increasing in s for $\rho\le 0$. Again using the fact that $(1-\Phi(\cdot))/\varphi(\cdot)$ is a decreasing function we get

$$\int_a^\infty\exp\left(-\frac{s^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right)\left(1-\Phi\left(\frac{a-\rho s\cos(\alpha-\beta)}{\sqrt{1-\rho^2}}\right)\right)\,ds$$

$$\le\left(1-\Phi\left(\frac{a(1-\rho\cos(\alpha-\beta))}{\sqrt{1-\rho^2}}\right)\right)\left(\varphi\left(\frac{a(1-\rho\cos(\alpha-\beta))}{\sqrt{1-\rho^2}}\right)\right)^{-1}\sqrt{1-\rho^2}$$

$$\times\exp\left(-\frac{a^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right)\left(1-\Phi\left(\frac{a(1-\rho\cos(\alpha-\beta))}{\sqrt{1-\rho^2}}\right)\right)$$

$$\leq \frac{1-\rho^2}{a} \exp\left(-\frac{a^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right)\left(1-\Phi\left(\frac{a(1-\rho\cos(\alpha-\beta))}{\sqrt{1-\rho^2}}\right)\right).$$

By substituting the last integral in (2.5) with the expression above, after some elementary approximations we arrive at

$$\mathrm{Cov}(I_1, I_2) \geq \rho \cdot \frac{2\sqrt{2\pi(1-\rho^2)}a}{16(1-\rho^2)} \exp\left(-\frac{a^2}{2}\right) \frac{\sqrt{1-\rho^2}}{\sqrt{2\pi}a} \exp\left(-\frac{a^2}{2(1-\rho^2)}\right)$$

$$+\rho \cdot \frac{\sqrt{2\pi(1-\rho^2)}}{16(1-\rho^2)} \frac{1-\rho^2}{a} \exp\left(-\frac{a^2}{2}\right) \frac{\sqrt{1-\rho^2}}{\sqrt{2\pi}a} \exp\left(-\frac{a^2}{2(1-\rho^2)}\right)$$

$$+\frac{1}{4\pi^2}\int_0^{\pi/2}\int_0^{\pi/2} \frac{1-\rho^2}{1-\rho^2\cos^2(\alpha-\beta)} \exp\left(-\frac{a^2(1-\rho^2\cos^2(\alpha-\beta))}{2(1-\rho^2)}\right)$$

$$\times \exp\left(-\frac{a^2(1-\rho\cos(\alpha-\beta))^2}{2(1-\rho^2)}\right)\, d\alpha\, d\beta - \frac{1}{16}\exp\left(-\frac{a^2}{2}\right)\exp\left(-\frac{a^2}{2}\right)$$

$$\geq \rho \cdot \frac{1}{16} \exp\left(-\frac{a^2}{2}\right) \exp\left(-\frac{a^2}{2(1-\rho^2)}\right)\left(2+\frac{1-\rho^2}{a^2}\right)$$

$$+\frac{1}{16}(1-\rho^2)\exp\left(-\frac{a^2}{2}\right)\exp\left(-\frac{a^2}{2}\frac{1-\rho}{1+\rho}\right) - \frac{1}{16}\exp\left(-\frac{a^2}{2}\right)\exp\left(-\frac{a^2}{2}\right).$$

$\square$

In the next proposition, the asymptotic behavior of the covariances is derived. The parameters n and a are restricted by the condition that the expected number of exceeding points, λ, is constant. This will determine $a = a_n$ as a function of n. The asymptotic behavior of the covariances as a function of n is then easily obtained.

Proposition 2.5 *Let $\{a_n\}$ be a sequence such that $n\cdot\exp(-a_n^2/2) = K$ for some constant K. Let I_1 and I_2 be defined by (2.3). Then for some constants C depending on K only and all $n \geq 2$:*

If $0 \leq \rho < 1$,

(i) $$0 \leq \mathrm{Cov}(I_1, I_2) \leq C\frac{1}{n}\left(\frac{1}{n}\right)^{\frac{1-\rho}{1+\rho}}\left(\frac{2\rho}{1-\rho}+1\right),$$

$$\text{(ii)} \qquad 0 \leq \operatorname{Cov}(I_1, I_2) \leq C \frac{1}{1-\rho} \rho \frac{1}{n^2} e^{2\rho \ln n} \ln n.$$

If $-1 < \rho \leq 0$,

$$\text{(iii)} \qquad 0 \geq \operatorname{Cov}(I_1, I_2) \geq -C|\rho| \frac{1}{n^2} \ln n.$$

We see that we have two possible approximations in the case where ρ is positive. The first is easier to obtain and is sufficient when ρ is large, while the second, being proportional to ρ, is used when ρ is small. Note also that the absolute value of the bound in (iii) is less than the bounds in (i) and (ii). This observation is used in the proof of Proposition 2.6.

Proof. First note that $\exp(-a_n^2/2) = K/n$ implies $a_n^2 \sim 2 \ln n$.
(i) The first assertion follows immediately from Lemma 2.4.
(ii) Note that

$$\begin{aligned} \exp\left(-\frac{a_n^2}{2}\frac{1-\rho}{1+\rho}\right) - \exp\left(-\frac{a_n^2}{2}\right) &= \int_{a_n^2 \frac{1-\rho}{1+\rho}}^{a_n^2} 1/2 e^{-x/2}\, dx \\ &\leq a_n^2 \frac{\rho}{1+\rho} \exp\left(-\frac{a_n^2}{2}\frac{1-\rho}{1+\rho}\right). \end{aligned} \tag{2.7}$$

By Lemma 2.4 and relation (2.7),

$$\begin{aligned} &\operatorname{Cov}(I_1, I_2) \leq \frac{1}{16} \exp\left(-\frac{a_n^2}{2}\right) \left(\left(1 + \frac{2\rho}{1-\rho} + \frac{\rho(1+\rho)}{a_n^2(1-\rho)}\right) \right. \\ &\left. \times \left(\exp\left(-\frac{a_n^2}{2}\frac{1-\rho}{1+\rho}\right) - \exp\left(-\frac{a_n^2}{2}\right)\right) + \left(\frac{2\rho}{1-\rho} + \frac{\rho(1+\rho)}{a_n^2(1-\rho)}\right) \exp\left(-\frac{a_n^2}{2}\right)\right) \\ &\leq \frac{1}{16} \exp\left(-\frac{a_n^2}{2}\right) \left(1 + \frac{2\rho}{1-\rho} + \frac{\rho(1+\rho)}{a_n^2(1-\rho)}\right) a_n^2 \frac{\rho}{1+\rho} \exp\left(-\frac{a_n^2}{2}\frac{1-\rho}{1+\rho}\right) \\ &+ \frac{\rho}{16} \left(\frac{2}{1-\rho} + \frac{1+\rho}{a_n^2(1-\rho)}\right) \exp\left(-\frac{a_n^2}{2}\right) \exp\left(-\frac{a_n^2}{2}\right) \\ &\leq C \frac{1}{n} \left(\frac{1}{n}\right)^{\frac{1-\rho}{1+\rho}} \frac{\rho \ln n}{1-\rho}. \end{aligned}$$

(iii) Obviously for $\rho \leq 0$,

$$\exp\left(-\frac{a_n^2}{2}\frac{1-\rho}{1+\rho}\right) \leq \exp\left(-\frac{a_n^2}{2(1-\rho^2)}\right) \leq \exp\left(-\frac{a_n^2}{2}\right). \tag{2.8}$$

Moreover, a similar argument as for (2.7) gives

$$(1+\rho)\left(\exp\left(-\frac{a_n^2}{2}\right) - \exp\left(-\frac{a_n^2}{2}\frac{1-\rho}{1+\rho}\right)\right) \leq |\rho| a_n^2 \exp\left(-\frac{a_n^2}{2}\right). \tag{2.9}$$

By combining Lemma 2.4 and inequalities (2.8)-(2.9), after some manipulation we end up with

$$\begin{aligned}
-\mathrm{Cov}(I_1, I_2) &\leq \frac{1}{16}\exp\left(-\frac{a_n^2}{2}\right)\left\{|\rho|\left(a_n^2+1\right)\exp\left(-\frac{a_n^2}{2}\right)\right.\\
&\left.+\rho^2\exp\left(-\frac{a_n^2}{2}\frac{1-\rho}{1+\rho}\right) + |\rho|\left(2+\frac{1-\rho^2}{a_n^2}\right)\exp\left(-\frac{a_n^2}{2(1-\rho^2)}\right)\right\}\\
&\leq C|\rho|\frac{1}{n^2}\ln n.
\end{aligned}$$

□

Finally we will consider the case when the sign of the covariance ρ is different for $(X_1^{(1)},\ X_2^{(1)})$ compared to $(X_1^{(2)},\ X_2^{(2)})$. The reason for this is that when we want to calculate the covariance of indicators in two different but adjacent quadrants the main problem is to estimate expressions of the form $E(I_1 I_3)$, where I_1 is as in (2.3) and

$$\begin{aligned}
I_3 &= I\left\{\left((X_2^{(1)})^2 + (X_2^{(2)})^2\right)\ I\{X_2^{(1)} \geq 0, X_2^{(2)} \leq 0\} > a^2\right\}\\
&= I\left\{\left((X_2^{(1)})^2 + (-X_2^{(2)})^2\right)\ I\{X_2^{(1)} \geq 0, -X_2^{(2)} \geq 0\} > a^2\right\},
\end{aligned}$$

implying that the same probability as in the previous proposition is to be estimated, but with different signs on the covariances $\mathrm{Cov}(X_1^{(1)}, X_2^{(1)})$ and $\mathrm{Cov}(X_1^{(2)}, X_2^{(2)})$. This is taken care of in Proposition 2.6.

As a remark, we note that when calculating the covariance of indicators in two "opposite" quadrants, the signs are reversed for both pairs $(X_1^{(1)}, X_2^{(1)})$ and $(X_1^{(2)}, X_2^{(2)})$. Proposition 2.5 can thus be used with ρ replaced by $-\rho$.

Proposition 2.6 *Let*

$$\left(X_1^{(1)},\ X_2^{(1)}\right) \in N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \tag{2.10}$$

$$\left(X_1^{(2)},\ X_2^{(2)}\right) \in N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \right), \tag{2.11}$$

be independent and $|\rho| < 1$. Define I_1 and I_2 by (2.3) combined with (2.10) and (2.11). Then, with the same condition on the sequence $\{a_n\}$ as in Proposition 2.5, i.e. $n \cdot \exp(-a_n^2/2) = K$ for some constant K, and $n \geq 2$,

$$\text{(i)} \qquad |\text{Cov}(I_1, I_2)| \leq C\frac{1}{n}\left(\frac{1}{n}\right)^{\frac{1-|\rho|}{1+|\rho|}} \left(\frac{2|\rho|}{1-|\rho|} + 1\right),$$

$$\text{(ii)} \qquad |\text{Cov}(I_1, I_2)| \leq C\frac{1}{n^2} e^{2|\rho| \ln n} \frac{|\rho| \ln n}{1-|\rho|}.$$

Proof. Assume, without loss of generality, that $\rho \geq 0$. Let I_1^+, I_2^+ be defined by (2.3) with distributions for both $(X_1^{(1)},\ X_2^{(1)})$ and $(X_1^{(2)},\ X_2^{(2)})$ given by (2.10). Similarly, let I_1^-, I_2^- be as in (2.3) with distributions for both pairs given by (2.11). By Slepian's inequality, see Tong [26] page 10,

$$E(I_1^- I_2^-) \leq E(I_1 I_2) \leq E(I_1^+ I_2^+).$$

Hence

$$|\text{Cov}(I_1, I_2)| \leq \max\left\{ |\text{Cov}(I_1^-, I_2^-)|, |\text{Cov}(I_1^+, I_2^+)| \right\}.$$

The statements now follow from Propositions 2.5 (i), (iii) and Propositions 2.5 (ii), (iii) respectively. ▫

2.3.2 Convergence of the number of exceedances

The main theorem will now be stated. In this theorem, stationarity is assumed. The assumption is used only to collect terms with equal covariances, and can easily be relaxed to cover non-stationary cases as well.

Theorem 2.7 *Suppose* $(X_1^{(1)},\ldots,X_n^{(1)}),(X_1^{(2)},\ldots,X_n^{(2)})$ *are two independent and identically distributed standardized stationary Gaussian sequences with correlation function* $\{\rho_k\}$. *Let* $\rho_{\max}=\max(|\rho_1|,|\rho_2|,\ldots)$ *and* $Y_k=(X_k^{(1)})^2+(X_k^{(2)})^2$, $k=1,\ldots,n$. *Assume* $\rho_{\max}>0$ *and suppose* a_n *are real positive numbers such that* $n\cdot P(Y_1>a_n^2)=\lambda$ *is constant. Set*

$$
\begin{aligned}
{}_1I_k &= I\{Y_k>a_n^2, X_k^{(1)}\geq 0, X_k^{(2)}\geq 0\},\\
{}_2I_k &= I\{Y_k>a_n^2, X_k^{(1)}\leq 0, X_k^{(2)}\geq 0\},\\
{}_3I_k &= I\{Y_k>a_n^2, X_k^{(1)}\geq 0, X_k^{(2)}\leq 0\},\\
{}_4I_k &= I\{Y_k>a_n^2, X_k^{(1)}\leq 0, X_k^{(2)}\leq 0\},
\end{aligned}
$$

$$W=\sum_{k=1}^{n}\sum_{j=1}^{4}{}_jI_k\ ,\qquad \lambda=E(W)=n\cdot P(Y_1>a_n^2).$$

If $|\rho_k|\leq A/\ln k$, $k\geq 2$, *for some constant* A, *then*

$$d(W,\mathrm{Po}(\lambda))=O\left(n^{-\frac{1-\rho_{\max}}{1+\rho_{\max}}}+\frac{\ln n}{n}\sum_{k=1}^{n}|\rho_k|\right),\qquad \textit{as } n\to\infty.$$

Note that if $\rho_{\max}=0$, the considered sequences are independent, identically distributed random variables. In fact we have
$W\sim \mathrm{Bin}(n,e^{-a_n^2/2})$, and

$$d(W,\mathrm{Po}(\lambda))=O(\frac{1}{n}).$$

In the Gaussian case, i.e. if $I_k=I\{X_k>a\}$, where $\{X_k\}$ is a stationary, standardized, Gaussian sequence, the corresponding result (see Section 3.1) is

$$d(W,\mathrm{Po}(\lambda))=O\left(n^{-\frac{1-\rho_{\max}}{1+\rho_{\max}}}(\ln n)^{-\frac{\rho_{\max}}{1+\rho_{\max}}}+\frac{1}{n}\sum_{k=1}^{n}|\rho_k|\right),\quad \text{as } n\to\infty.$$

We see that the convergence rate is almost the same in the $\chi^2(2)$ as in the Gaussian case. This is clearly the case if the extreme values in the χ^2–sequence do not "interfere", so that each extreme point in the χ^2–sequence corresponds to an extreme value in exactly one of the underlying Gaussian sequences. This seems to be true, at least in continuous time, since the realizations of a χ^2–process in continuous time at an extreme value are unimodal, pointing to the fact that only one of the underlying processes is extreme. Probably the same phenomenon occurs in discrete time.

Corollary 2.8 *Suppose $\rho_k \ln k \to 0$ as $k \to \infty$. Then*

$$W \xrightarrow{d} \mathrm{Po}(\lambda)\ , \qquad n \to \infty.$$

Proof. Since $\rho_k \ln k \to 0$ implies $\frac{\ln n}{n} \sum_{k=1}^{n} |\rho_k| \to 0$, it follows from Theorem 2.7 that $d(W, \mathrm{Po}(\lambda)) \to 0$, as $n \to \infty$, and hence $W \xrightarrow{d} \mathrm{Po}(\lambda)$.

□

Consequently the condition $\rho_k \ln k \to 0$, often called the Berman condition, is sufficient for Poisson convergence in the Gaussian case as well as in the $\chi^2(2)$–case.

Proof of Theorem 2.7. First note that $\rho_{\max} < 1$ since $|\rho_k| = 1$ for some $k \geq 1$ would imply that $|\rho_{mk}| = 1$, $m \geq 1$, which contradicts $|\rho_k| \leq A/\ln k$. By Theorem 2.3, and stationarity,

$$d(W, \mathrm{Po}(\lambda)) \leq \frac{1-e^{-\lambda}}{\lambda} \left\{ n \cdot \exp\left(-a_n^2\right) + 2n \sum_{i=1}^{4} \sum_{j=1}^{4} \sum_{k=2}^{n} |\mathrm{Cov}({}_jI_1,\ {}_iI_k)| \right\}.$$

The first term is of order $O(1/n)$. Next, divide $\sum_{k=2}^{n} |\mathrm{Cov}({}_jI_1,\ {}_iI_k)|$ into three parts.

(i) *Strong dependence but only a finite number of terms*

Let $0 < \delta < \frac{2}{3}\frac{\rho_{\max}}{1+\rho_{\max}}$. Since $|\rho_k| \leq A/\ln k$, only a finite number of terms have $|\rho_k| > \delta$. According to Propositions 2.5 (i), (iii) and

Proposition 2.6 (i), each contribution is equal to

$$\mathrm{Cov}({}_jI_1,\,{}_iI_k) \le C\frac{1}{n}\left(\frac{1}{n}\right)^{\frac{1-|\rho_k|}{1+|\rho_k|}}\frac{2|\rho_k|}{1-|\rho_k|} \le C\frac{1}{n}n^{-\frac{1-\rho_{\max}}{1+\rho_{\max}}}$$

and their total contribution is of the same order.

(ii) *Strong dependence*

Consider terms with $|\rho_k| \le \delta$ and $k < n^\delta$. Their contribution is, again from Propositions 2.5 (i), (iii) and Proposition 2.6 (i),

$$\mathrm{Cov}({}_jI_1,\,{}_iI_k) \le C\frac{1}{n^2}n^{\frac{2\delta}{1+\delta}}.$$

The total contribution is therefore

$$C\frac{1}{n^2}n^{\frac{2\delta}{1+\delta}}n^\delta \le C\frac{1}{n^2}n^{3\delta} = O\left(\frac{1}{n}n^{-\frac{1-\rho_{\max}}{1+\rho_{\max}}}\right).$$

(iii) *Weak dependence*

Finally, consider the remaining terms, i.e. terms with $k \ge n^\delta$, i.e. $|\rho_k| \le A/\ln k \le A/(\delta \ln n)$. Propositions 2.5 (ii), (iii) and Proposition 2.6 (ii) yield

$$\mathrm{Cov}({}_jI_1,\,{}_iI_k) \le C'\frac{1}{n^2}e^{2|\rho_k|\ln n}\frac{|\rho_k|\ln n}{1-|\rho_k|} \le C\frac{\ln n}{n^2}|\rho_k|$$

and the total contribution is given by

$$O\left(\frac{1}{n^2}\ln n\sum_k|\rho_k|\right).$$

Combining (i), (ii) and (iii) gives the desired result. ◻

Corollary 2.9 *If the underlying Gaussian sequence in Theorem 2.7 is a stationary ARMA sequence, then*

$$d(W,\mathrm{Po}(\lambda)) = O\left(n^{-\frac{1-\rho_{\max}}{1+\rho_{\max}}}\right), \qquad \textit{as } n\to\infty.$$

Proof. If $\rho_{\max} = 0$, the statement follows immediately, cf. the remark in connection with Theorem 2.7. Assume now that $\rho_{\max} > 0$. For a stationary ARMA sequence, $|\rho_k|$ decreases exponentially so that $\sum_{k=1}^{\infty} |\rho_k| < \infty$. Theorem 2.7 gives

$$d(W, \mathrm{Po}(\lambda)) = O\left(n^{-\frac{1-\rho_{\max}}{1+\rho_{\max}}} + \frac{\ln n}{n}\right) = O\left(n^{-\frac{1-\rho_{\max}}{1+\rho_{\max}}}\right), \qquad \text{as } n \to \infty.$$

□

2.4 Numerical examples for $\chi^2(1)$ and $\chi^2(2)$ sequences

In this last section of Chapter 2, we illustrate the Poisson convergence numerically. This is done for both the $\chi^2(1)$ and the $\chi^2(2)$ case, and the results are compared. As in Theorem 2.7, we assume that the underlying normal sequences are stationary and standardized with correlation function $\{\rho_k\}$.

2.4.1 The $\chi^2(2)$ case

Before we apply any numerical methods, we rewrite the expressions of the covariances in Subsection 2.3.1 so that the dimensions of the integrals are reduced. The expressions are then integrated numerically, using MATLAB.

From Theorem 2.3,

$$d(W, \mathrm{Po}(\lambda)) \leq \frac{1-e^{-\lambda}}{\lambda} \left\{ nP(Y_1 > a^2)^2 + \sum_{i=1}^{4} \sum_{j=1}^{4} \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} |\mathrm{Cov}({}_jI_k,\, {}_iI_l)| \right\},$$

where $P(Y_1 > a^2) = e^{-a^2/2}$. By symmetry and stationarity,

$$\sum_{i=1}^{4} \sum_{j=1}^{4} \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} |\mathrm{Cov}({}_jI_k,\, {}_iI_l)|$$

$$= 4 \cdot 2 \sum_{k=1}^{n} \sum_{l=k+1}^{n} |\mathrm{Cov}({}_1I_k,\, {}_1I_l)| + 2 \cdot 2 \sum_{i=1}^{4} \sum_{j=i+1}^{4} \sum_{k=1}^{n} \sum_{l=k+1}^{n} |\mathrm{Cov}({}_jI_k,\, {}_iI_l)|$$

$$= 8 \sum_{k=1}^{n-1} k |\mathrm{Cov}({}_1I_1,\, {}_1I_{n-k+1})| + 4 \sum_{i=1}^{4} \sum_{j=i+1}^{4} \sum_{k=1}^{n-1} k |\mathrm{Cov}({}_jI_1,\, {}_iI_{n-k+1})|.$$

Note that the sum over i and j in the last term runs over six terms, in four of which only one of the two underlying normal sequences change sign, and in two of which both sequences change sign. We get

$$d(W, \mathrm{Po}(\lambda)) \le \frac{1 - e^{-\lambda}}{\lambda} \Big\{ n \exp\big(-a^2\big) + 8 \sum_{k=1}^{n-1} k \cdot \Big(|\mathrm{Cov}({}_1I_1,\, {}_1I_{n-k+1})|$$

$$+ |\mathrm{Cov}({}_1I_1,\, {}_1I_{n-k+1})^{**}| + 2 |\mathrm{Cov}({}_1I_1,\, {}_1I_{n-k+1})^{*}| \Big) \Big\}, \qquad (2.12)$$

where the double star indicates change of signs on ρ, and the single star indicates different signs on the ρs (cf. the proof of Proposition 2.6).

Consider now the problem of calculating $E({}_1I_1 \cdot {}_1I_{n-k+1})$ for some fixed k. By using the substitution

$$\begin{cases} \alpha - \beta = u \\ \alpha = v \end{cases},$$

we get for any integrable function f,

$$\int_0^{\pi/2} \int_0^{\pi/2} f(\cos(\alpha - \beta))\, d\alpha\, d\beta = \int_{-\pi/2}^{\pi/2} f(\cos u)\, (\pi/2 - |u|)\; du$$

$$= 2 \int_0^{\pi/2} f(\cos u)\, (\pi/2 - u)\; du.$$

Applying this on equation (2.5) in the proof of Lemma 2.4 gives

$$\mathrm{Cov}({}_1I_1,\, {}_1I_{n-k+1}) = \frac{1}{2\pi^2} \int_0^{\pi/2} \frac{\sqrt{(1-\rho^2)}}{1 - \rho^2 \cos^2 u}$$

$$\times \Big\{ 2\sqrt{2\pi}\rho a \cos u \exp\Big(-\frac{a^2(1 - \rho^2 \cos^2 u)}{2(1-\rho^2)} \Big) \Big(1 - \Phi\Big(\frac{a(1 - \rho \cos u)}{\sqrt{1-\rho^2}} \Big) \Big)$$

$$+\sqrt{(1-\rho^2)}\exp\left(-\frac{a^2(1-\rho^2\cos^2 u)}{2(1-\rho^2)}\right)\exp\left(-\frac{a^2(1-\rho\cos u)^2}{2(1-\rho^2)}\right)$$
$$+\sqrt{2\pi}\rho\cos u\int_a^\infty \exp\left(-\frac{s^2(1-\rho^2\cos^2 u)}{2(1-\rho^2)}\right)\left(1-\Phi\left(\frac{a-\rho s\cos u}{\sqrt{1-\rho^2}}\right)\right)ds\Bigg\}$$
$$\times\left(\frac{\pi}{2}-u\right)du-\frac{1}{16}\exp\left(-\frac{a^2}{2}\right)\exp\left(-\frac{a^2}{2}\right), \tag{2.13}$$

where $\rho = \rho_{(n-k+1)-1} = \rho_{n-k}$. All terms of the form $|\mathrm{Cov}({}_1I_1,\ {}_1I_{n-k+1})|$ or $|\mathrm{Cov}({}_1I_1,\ {}_1I_{n-k+1})^{**}|$ can be calculated this way as double integrals with $\rho = \rho_{n-k}$ and $\rho = -\rho_{n-k}$ respectively.

Next, we turn to the problem of computing $|\mathrm{Cov}({}_1I_1,\ {}_1I_{n-k+1})^*|$. By calculations similar to those leading to equation (2.5), noting that the same arguments are valid with β changed to $-\beta$ (since $-y = s\sin(-\beta)$), we get

$$\mathrm{Cov}({}_1I_1,\ {}_1I_{n-k+1})^* = \frac{1}{4\pi^2}\int_0^{\pi/2}\int_0^{\pi/2}\int_a^\infty \frac{\sqrt{2\pi}\rho\cos(\alpha+\beta)}{\sqrt{1-\rho^2}}s^2$$
$$\times\exp\left(-\frac{s^2(1-\rho^2\cos^2(\alpha+\beta))}{2(1-\rho^2)}\right)\left(1-\Phi\left(\frac{a-\rho s\cos(\alpha+\beta)}{\sqrt{1-\rho^2}}\right)\right)$$
$$+s\cdot\exp\left(-\frac{(a^2-2as\rho\cos(\alpha+\beta)+s^2)}{2(1-\rho^2)}\right)ds\,d\alpha\,d\beta$$
$$-\frac{1}{16}\exp\left(-\frac{a^2}{2}\right)\exp\left(-\frac{a^2}{2}\right)$$
$$=\frac{1}{4\pi^2}\int_0^\pi\left[\frac{\sqrt{2\pi}\rho\cos u}{\sqrt{1-\rho^2}}\int_a^\infty s^2\exp\left(-\frac{s^2(1-\rho^2\cos^2 u)}{2(1-\rho^2)}\right)\right.$$
$$\times\left(1-\Phi\left(\frac{a-\rho s\cos u}{\sqrt{1-\rho^2}}\right)\right)ds+\Bigg\{\sqrt{2\pi(1-\rho^2)}a\rho\cos u$$
$$\left(1-\Phi\left(\frac{a(1-\rho\cos u)}{\sqrt{1-\rho^2}}\right)\right)+(1-\rho^2)\exp\left(-\frac{a^2(1-\rho\cos u)^2}{2(1-\rho^2)}\right)\Bigg\}$$

$$\times \exp\left(-\frac{a^2(1-\rho^2\cos^2 u)}{2(1-\rho^2)}\right)\Bigg] (\pi/2 - |\pi/2 - u|)\; du$$

$$-\frac{1}{16}\exp\left(-\frac{a^2}{2}\right)\exp\left(-\frac{a^2}{2}\right), \tag{2.14}$$

where the last equality follows from a substitution similar to that in (2.13).

Relation (2.12) and equations (2.13)–(2.14) are in a form suitable for computing in e.g. MATLAB. In Section 2.4.3, results from such calculations are shown together with similar results for the $\chi^2(1)$ case.

2.4.2 The $\chi^2(1)$ case

As we have remarked in Section 2.2, the investigation of a $\chi^2(1)$–sequence is best done by considering the underlying normal sequence instead. We immediately have the following result (Corollary 3.2 in Holst and Janson [9]):

Theorem 2.10 *Suppose $(X_1, \dots, X_n)$ is a standardized and stationary normal sequence with correlation function $\{\rho_k\}$. Let a be a real positive number and set for $k \in \{1, 2, \dots, n\}$,*

$$_1I_k = I\{X_k^2 > a^2, X_k \geq 0\} = I\{X_k > a\},$$

$$_2I_k = I\{X_k^2 > a^2, X_k \leq 0\} = I\{X_k < -a\},$$

$$W = \sum_{k=1}^{n}\sum_{j=1}^{2} {}_jI_k\,, \qquad \lambda = E(W) = n \cdot P(X_1^2 > a^2).$$

Then

$$d(W, \mathrm{Po}(\lambda)) \leq \frac{1-e^{-\lambda}}{\lambda}\Bigg\{ n \cdot \left(P(X_1^2 > a^2)\right)^2$$

$$+2\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n}\Big(|\mathrm{Cov}({}_1I_k,\, {}_1I_l)| + |\mathrm{Cov}({}_1I_k,\, {}_2I_l)|\Big)\Bigg\}, \tag{2.15}$$

where $P(X_1^2 > a^2) = P(|X_1| > a) = 2(1 - \Phi(a))$.

Using stationarity, (2.15) simplifies to

$$d(W, \mathrm{Po}(\lambda)) \leq \frac{1-e^{-\lambda}}{\lambda}\Bigg\{ n \cdot \Big(2(1-\Phi(a))\Big)^2 +$$

$$2\sum_{k=1}^{n-1} k\Big(|\mathrm{Cov}({}_1I_1,\ {}_1I_{n-k+1})| + |\mathrm{Cov}({}_1I_1,\ {}_2I_{n-k+1})|\Big)\Bigg\}. \qquad (2.16)$$

The covariances of the right hand side of (2.16) can easily be computed numerically as bivariate normal probabilities.

2.4.3 Results from the numerical calculations

Finally, we present the results from the numerical calculations. Three sequences are considered, namely where the underlying normal sequence is an AR(1) sequence, an MA(5) sequence and finally an ARMA(1,1) sequence. The results are shown in Table 2.2, 2.4 and 2.6.

The calculations have been performed in MATLAB, Version 4.2. The cumulative distribution function $\Phi(\ )$ has been approximated by a polynomial with an error of less than $7.5 \cdot 10^{-8}$. All covariances with $|\rho| < 10^{-10}$ were set to zero in order to speed up the calculations. The upper limit of the radial integral was set to 50. The parameter a was chosen as $a = \sqrt{-2\ln(\lambda/n)}$ in the $\chi^2(2)$ case and as $a = \Phi^{-1}(1-\lambda/(2n))$ in the $\chi^2(1)$ case.

1. AR(1).

The first example is an AR(1) sequence with parameter 0.5,

$$X_k = \frac{\sqrt{3}}{2}Z_k + 0.5X_{k-1}, \qquad \{Z_k\} \sim \mathrm{WN}(0,1). \qquad (2.17)$$

For this sequence, $\rho_k = 0.5^k$. The first six values of ρ_k are tabulated in Table 2.1. In Table 2.2, the results from the calculation of the total variation distance are shown.

k	1	2	3	4	5	6	...
ρ_k	0.5	0.25	0.125	0.0625	0.0313	0.0156	

Table 2.1: *Correlation function for the AR(1) sequence (2.17).*

		λ		
		1	5	10
n	100	0.15	0.51	0.74
	500	0.07	0.23	0.33
	1000	0.05	0.17	0.24

		λ		
		1	5	10
n	100	0.10	0.37	0.55
	500	0.05	0.16	0.23
	1000	0.03	0.11	0.16

Table 2.2: *Upper limits for $d(W, \mathrm{Po}(\lambda))$ for the AR(1) sequence (2.17) in the $\chi^2(2)$ case (left), and the $\chi^2(1)$ case (right).*

2. MA(5).

The next example is the MA(5) sequence

$$X_k = \frac{4}{\sqrt{21}}\Big(Z_k - 0.25Z_{k-1} - 0.25Z_{k-2} - 0.25Z_{k-3} - 0.25Z_{k-4} - 0.25Z_{k-5}\Big), \qquad \{Z_k\} \sim \mathrm{WN}(0,1). \tag{2.18}$$

Some values of the correlation function $\{\rho_k\}$ are found in Table 2.3. Note that the rate of convergence in this case is determined by by ρ_5. The upper limits for $d(W, \mathrm{Po}(\lambda))$ are shown in Table 2.4.

k	1	2	3	4	5	6	...
ρ_k	0	-0.048	-0.095	-0.143	-0.190	0	

Table 2.3: *Correlation function for the MA(5) sequence (2.18).*

		λ		
		1	5	10
n	100	0.04	0.21	0.35
	500	0.01	0.07	0.11
	1000	0.01	0.04	0.07

		λ		
		1	5	10
n	100	0.03	0.18	0.30
	500	0.01	0.05	0.09
	1000	0.01	0.03	0.05

Table 2.4: *Upper limits for $d(W, \mathrm{Po}(\lambda))$ for the MA(5) sequence (2.18) in the $\chi^2(2)$ case (left), and the $\chi^2(1)$ case (right).*

3. ARMA(1,1).

Our last example is an ARMA(1,1) sequence

$$X_k + 0.3X_{k-1} = \sqrt{\frac{91}{107}}\,(Z_k - 0.1Z_{k-1}), \quad \{Z_k\} \sim \mathrm{WN}(0,1), \quad (2.19)$$

with correlation function as in Table 2.5. In the last table, Table 2.6,

k	1	2	3	4	5	6	...
ρ_k	-0.385	0.115	-0.035	0.010	-0.003	0.001	

Table 2.5: *Correlation function for the ARMA(1,1) sequence (2.19).*

the results are presented.

We see that for the "kind" MA(5) sequence above, where the correlations are weak, the difference from a Poisson distribution measured by the total variation distance is just a few percent. For the AR(1) sequence, the distance between the true distribution and the corresponding Poisson distribution is larger. However, for low values of λ, the difference in total variation distance is low even for this sequence. For larger values of λ, W is more "Gaussian" and the distance from Poisson is larger.

		$\lambda = 1$	$\lambda = 5$	$\lambda = 10$
n	100	0.07	0.29	0.44
	500	0.03	0.11	0.17
	1000	0.02	0.08	0.12

		$\lambda = 1$	$\lambda = 5$	$\lambda = 10$
n	100	0.05	0.22	0.34
	500	0.02	0.08	0.12
	1000	0.01	0.05	0.08

Table 2.6: *Upper limits for* $d(W, \mathrm{Po}(\lambda))$ *for the ARMA(1,1) sequence (2.19) in the* $\chi^2(2)$ *case (left), and the* $\chi^2(1)$ *case (right).*

Chapter 3

Some comments on Poisson approximation for Gaussian sequences

This chapter is devoted to Poisson approximation of the number of exceedances of Gaussian sequences. In the paper by Holst and Janson [9], frequently referred to in the previous chapter, an upper bound for the rate of convergence is derived. In Section 3.1, we will show that this bound can be improved for sequences with strong "global" dependence. On the other hand, we will also show, that the bound is of the right order for sequences with correlation functions of finite support. Such sequences will be of primary interest later when compound Poisson approximation is considered.

As in the χ^2–case, the existence of a monotone coupling is also useful for Gaussian sequences, see [9]. For the latter case, the coupling will be constructed explicitly by elementary methods – a more straightforward way than the existence theorems used in Holst and Janson [9] or, for χ^2–sequences, in Section 2.2. In Section 3.2, such an explicit construction is carried out for Gaussian sequences. Unfortunately, the technique does not seem to work for χ^2–sequences.

For numerical results on Gaussian sequences, see the $\chi^2(1)$–case in Section 2.4.

3.1 Improvement of the bound for the convergence rate

We start this section by stating some of the results in Holst and Janson [9] for reference. These are then combined with sharper approximations yielding an improved bound for the convergence rate for Gaussian sequences.

The first theorem is a variant of Theorem 3.1 in [9].

Theorem 3.1 *Suppose $Y_1, Y_2, \ldots, Y_n$ are jointly normally distributed random variables. Let a be a real number and set for $k \in \{1, 2, \ldots, n\}$,*

$$I_k = I(Y_k > a), \qquad p_k = P(Y_k > a)$$

$$W = \sum_{k=1}^{n} I_k, \qquad \lambda = E(W) = \sum_{k=1}^{n} p_k.$$

Then

$$d(W, \mathrm{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \left(\sum_{k=1}^{n} p_k^2 + \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} |\mathrm{Cov}(I_k, I_l)| \right).$$

Next, the covariances of the indicators are estimated. For large correlations the following lemma is appropriate (Lemma 3.4 (i) in [9]).

Lemma 3.2 *Let $\{a_n\}$ be a sequence such that $n(1 - \Phi(a_n)) = K$ for some constant K. Let (X_1, X_2) be $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ and let $I_1 = I(X_1 > a_n)$, $I_2 = I(X_1 > a_n)$. If $0 \leq \rho < 1$, then for some constant C depending on K only and all $n \geq 2$,*

$$0 \leq \mathrm{Cov}(I_1, I_2) \leq C(1-\rho)^{-1/2} n^{-2+2\rho/(1+\rho)} (\ln n)^{-\rho/(1+\rho)}.$$

For weak correlations, Holst and Janson derive the following bound for the covariances,

$$0 \leq \mathrm{Cov}(I_1, I_2) \leq C \frac{\rho \ln n}{n^2} e^{2\rho \ln n}.$$

As follows from the next lemma, it is possible to sharpen this bound slightly. A similar improvement of the bound for $\rho \leq 0$ is also valid.

Lemma 3.3 *Let I_1 and I_2 be defined as in Lemma 3.2. Then,*

$$(i) \quad \text{if } 0 \le \rho < 1, \qquad 0 \le \operatorname{Cov}(I_1, I_2) \le C \frac{\rho}{\sqrt{1-\rho^2}} \left(\frac{1}{n}\right)^{2/(1+\rho)},$$

$$(ii) \quad \text{if } -1 < \rho \le 0, \qquad 0 \ge \operatorname{Cov}(I_1, I_2) \ge -C \frac{|\rho|}{\sqrt{1-\rho^2}} \frac{1}{n^2}.$$

Proof. From Cramér and Leadbetter [8] page 27,

$$\begin{aligned}
&\operatorname{Cov}(I_1, I_2) = P(X_1 > a_n, X_2 > a_n) - P(X_1 > a_n)P(X_2 > a_n) \\
&\quad = \frac{1}{2\pi} \int_0^\rho e^{-a_n^2/(1+z)} (1-z^2)^{-1/2}\, dz.
\end{aligned}$$

Assume first that $0 \le \rho < 1$. Then

$$0 \le \operatorname{Cov}(I_1, I_2) \le \frac{1}{2\pi} \int_0^\rho \frac{1}{\sqrt{1-\rho^2}} e^{-a_n^2/(1+\rho)}\, dz = \frac{\rho}{2\pi\sqrt{1-\rho^2}} e^{-a_n^2/(1+\rho)}.$$

Since $n(1-\Phi(a_n)) = K$ implies $a_n \sim \sqrt{2 \ln n}$, we get

$$\operatorname{Cov}(I_1, I_2) \le C \frac{\rho}{\sqrt{1-\rho^2}} e^{-2\ln n/(1+\rho)} = C \frac{\rho}{\sqrt{1-\rho^2}} \left(\frac{1}{n}\right)^{2/(1+\rho)},$$

proving (i).

Assume next that $-1 < \rho \le 0$. Similarly as for (i),

$$\begin{aligned}
&|\operatorname{Cov}(I_1, I_2)| \le \left| \int_0^\rho e^{-a_n^2/(1+z)} (1-z^2)^{-1/2}\, dz \right| \\
&= \left| \int_0^{-\rho} e^{-a_n^2/(1-z)} (1-z^2)^{-1/2}\, dz \right| \le \frac{|\rho|}{\sqrt{1-\rho^2}} e^{-a_n^2} \\
&\le C \frac{|\rho|}{\sqrt{1-\rho^2}} \frac{1}{n^2},
\end{aligned}$$

proving the second part of the statement. □

Finally we state the main result – a modified version of Theorem 3.5 in Holst and Janson [9].

Theorem 3.4 *Let $\{Y_k\}$ be a standardized, stationary Gaussian sequence with correlation function $\{\rho_k\}$ satisfying $\rho_k \leq A \ln k$, $k \geq 2$, for some constant A. Let $\rho = \max(0, \rho_1, \rho_2, \ldots)$. Let a_n be real numbers such that $n(1-\Phi(a_n)) = \lambda$ for some constant λ. Define $W = \sum_{k=1}^n I(Y_k > a_n)$. Then*

$$d(W, \mathrm{Po}(\lambda)) = O\left(n^{-(1-\rho)/(1+\rho)} (\ln n)^{-\rho/(1+\rho)} + \frac{1}{n}\sum_{k=1}^{n} |\rho_k|\right), \qquad (3.1)$$

as $n \to \infty$.

This theorem is slightly sharper than the corresponding theorem in [9] in the sense that a logarithmic factor in the last term of the convergence rate has been removed. However note that we, unlike Holst and Janson, have assumed that λ is constant.

Proof. Copying the proof in Holst and Janson [9], but using Lemma 3.3 above, instead of Lemma 3.4 (ii) and (iii) in that paper gives the desired result. □

Having dealt with the second term in the convergence rate (3.1), we will now concentrate on the first. Assume that $\{Y_k\}$ has correlation function $\{\rho_k\}$, and let a_n be defined as in Theorem 3.4. Assume further that the correlation function $\{\rho_k\}$ has finite support, i.e. $\rho_k = 0$ for all $k > q$ for some finite q. The rate of convergence is then determined by the first term in (3.1). We will now show that this term is of the right order, by using results for the convergence rate of the maxima of a stationary sequence. We have

$$\begin{aligned} d(W, \mathrm{Po}(\lambda)) &= \frac{1}{2}\sum_{j=0}^{\infty}\left|P(W=j) - \frac{\lambda^j}{j!}e^{-\lambda}\right| \\ &\geq \frac{1}{2}\left|P(W=0) - e^{-\lambda}\right| = \frac{1}{2}\left|P(\max_{1\leq k\leq n} Y_k \leq a_n) - e^{-\lambda}\right|. \end{aligned}$$

From Rootzén [20],

$$\left|P(\max_{1\leq k\leq n} Y_k \leq a_n) - \Phi(a_n)^n\right| \sim C\left(\frac{1}{n}\right)^{(1-\rho)/(1+\rho)} (\ln n)^{-\rho/(1+\rho)}$$

for some constant C. Since $1-\Phi(a_n)=\lambda/n$ and

$$\left|\left(1-\frac{\lambda}{n}\right)^n - e^{-\lambda}\right| = O(\frac{1}{n}),$$

the triangle inequality yields

$$\begin{aligned}\left|P(\max_{1\le k\le n} Y_k \le a_n) - e^{-\lambda}\right| &\ge \left|P(\max_{1\le k\le n} Y_k \le a_n) - \Phi(a_n)^n\right| \\ &- \left|\left((1-\frac{\lambda}{n}\right)^n - e^{-\lambda}\right| \sim \left(\frac{1}{n}\right)^{(1-\rho)/(1+\rho)} (\ln n)^{-\rho/(1+\rho)}.\end{aligned}$$

Thus

$$d(W,\mathrm{Po}(\lambda)) \ge C' \left(\frac{1}{n}\right)^{(1-\rho)/(1+\rho)} (\ln n)^{-\rho/(1+\rho)}.$$

Hence the rate of convergence in Theorem 3.4 is the right one for sequences with correlation functions having finite support.

3.2 Explicit construction of the coupling

In Holst and Janson [9], the existence of a monotone coupling between the Gaussian random variables was established. The proof is based on Strassen's theorem, and gives no indication of how to construct such a coupling.

In this section we construct the coupling explicitly. Although only the existence proved earlier is needed, an explicit construction may increase the understanding of the nature of the problem. Moreover, the construction is rather straightforward and "simple" in the sense that it is based only on the well-known Gram-Schmidt orthogonalization method.

Let, as in Proposition 2.2, $(X_1, X_2, \ldots, X_n)$ be a Gaussian random sequence with $\mathrm{Cov}(X_1, X_k) \ge 0$ for $2 \le k \le p$ and $\mathrm{Cov}(X_1, X_k) \le 0$ for $p < k \le n$. For simplicity, assume that all variables are standardized, i.e. $E(X_k) = 0$, $V(X_k) = 1$, $k = 1, 2, \ldots, n$. Denote the correlations by

$$\rho_{ij} = \mathrm{Cov}(X_i, X_j).$$

Assume further that $U_1, U_2, \ldots, U_n$ are independent and identically distributed standardized Gaussian random variables, $U_k \sim \mathrm{N}(0,1)$. Define $(Y_1, Y_2, \ldots, Y_n)$ by

$$\begin{aligned} Y_1 &= \alpha_{11} U_1 \\ Y_2 &= \alpha_{21} U_1 + \alpha_{22} U_2 \\ &\vdots \\ Y_n &= \alpha_{n1} U_1 + \alpha_{n2} U_2 + \ldots + \alpha_{nn} U_n, \end{aligned} \tag{3.2}$$

where $\{\alpha_{ij}\}$ are real coefficients. From the Gram-Schmidt orthogonalization method, see e.g Rao [17] p 498-501, it is well known that the coefficients can be chosen such that

$$\mathcal{L}(Y_1, Y_2, \ldots, Y_n) = \mathcal{L}(X_1, X_2, \ldots, X_n).$$

The choice of coefficients is not unique. The following algorithm can be used to determine the coefficients. Obviously $\alpha_{11} = 1$ since $V(Y_1) = 1$. Moreover, given the coefficients up to row $k-1$, all but the last coefficient in the k:th row are given by the $k-1$ linear equations specifying the correlations $\mathrm{Cov}(X_j, X_k)$, $j < k$. Finally α_{kk} is determined by the normalizing condition $\mathrm{Var}(X_k) = 1$. Note that the sign of α_{kk} is arbitrary (but will influence subsequent coefficients). Note also that if the distribution of $(X_1, X_2, \ldots, X_n)$ is singular, some of the coefficients can be chosen to be zero, so that not all the variables $U_1, U_2, \ldots, U_n$ are needed.

Taking the covariance between Y_1 and Y_k yields

$$\mathrm{Cov}(Y_k, Y_1) = \rho_{k1} = \alpha_{k1}. \tag{3.3}$$

These are the only coefficients we need explicitly.

We now turn to the conditional distributions. Let the random variable V be independent of all previous variables and have density function

$$f_V(x) = \begin{cases} \varphi(x)/(1-\Phi(a) & \text{if } x > a \\ 0 & \text{otherwise.} \end{cases}$$

Define

$$V_1 = \begin{cases} U_1 & \text{if } U_1 > a \\ V & \text{otherwise.} \end{cases}$$

Clearly

$$\mathcal{L}(V_1) = \mathcal{L}(U_1 | U_1 > a)$$

and

$$V_1 \geq U_1. \tag{3.4}$$

Next, define

$$\begin{aligned} Z_1 &= \alpha_{11} V_1 \\ Z_2 &= \alpha_{21} V_1 + \alpha_{22} U_2 \\ &\vdots \\ Z_n &= \alpha_{n1} V_1 + \alpha_{n2} U_2 + \ldots + \alpha_{nn} U_n \end{aligned} \tag{3.5}$$

with the same set of coefficients as for $Y_1, Y_2, \ldots, Y_n$. Due to independence between V_1 and $U_2, U_3, \ldots, U_n$,

$$\mathcal{L}(Z_1, Z_2, \ldots, Z_n) = \mathcal{L}(X_1, X_2, \ldots, X_n | X_1 > a).$$

Moreover,

$$Z_k - Y_k = \alpha_{k1}(V_1 - U_1) = \rho_{k1}(V_1 - U_1)$$

by (3.3). Since $V_1 - U_1$ is positive by (3.4), $Z_k \geq Y_k$ for $1 \leq k \leq p$ and $Z_k \leq Y_k$ for $p < k \leq n$, completing the final step in the construction of the coupling.

We summarize what we have achieved in this section in a theorem.

Theorem 3.5 *Let $(X_1, X_2, \ldots, X_n)$ be a standardized, Gaussian random sequence with $\mathrm{Cov}(X_1, X_k) \geq 0$ for $2 \leq k \leq p$ and $\mathrm{Cov}(X_1, X_k) \leq 0$ for $p < k \leq n$. Let $(Y_1, Y_2, \ldots, Y_n)$ and $(Z_1, Z_2, \ldots, Z_n)$ be defined by (3.2) and (3.5) respectively. Then, for any real a,*

$$\begin{aligned} \mathcal{L}\left(Y_1, Y_2, \ldots, Y_n\right) &= \mathcal{L}\left(X_1, X_2, \ldots, X_n\right), \\ \mathcal{L}\left(Z_1, Z_2, \ldots, Z_n\right) &= \mathcal{L}\left(X_1, X_2, \ldots, X_n \,|\, X_1 > a\right) \end{aligned}$$

and $Y_k \leq Z_k$, $1 \leq k \leq p$; $Y_k \geq Z_k$, $p < k \leq n$.

Chapter 4

Compound Poisson approximation for Gaussian sequences

In this chapter, we consider Gaussian sequences, concentrating on how to handle strong "local" dependence. If the dependence within the sequence is weak, a natural candidate for the approximation is the Poisson distribution, see Leadbetter, Lindgren and Rootzén [12] or Holst and Janson [9]; see also Chapter 3 above. It is well-known that if the Berman condition, $\rho_k \ln k \to 0$, as $k \to \infty$, is fulfilled (see Berman [5]; cf. Chapter 1 above) the number of exceedances W converges in distribution to Poisson. However, if the "local" dependence is strong, i.e. if some of the correlations in the correlation function are positive and large, then the convergence rate is slow. It is then reasonable to try to find a distribution, close to Poisson, which better approximates the true distribution. With large "local" dependence the extreme points tend to cluster. A natural choice is then the compound Poisson distribution (see the definition in Section 4.1 below). The question of how to approximate W with the compound Poisson distribution is the main issue in this chapter.

Throughout the whole of this chapter, we will assume that the sequence is stationary. However, the derived results can rather straightforwardly be generalized to cover non-stationary sequences too. In Chapter 5, we will briefly discuss which changes are necessary.

Section 4.1 contains a review of the local version of Stein's method for compound Poisson approximation. In Section 4.2 the theory is applied to the problem of exceedances of Gaussian sequences. Results are derived both for upper bounds for sequences of finite length, and for the asymptotic convergence rate as the length of the sequence grows. In the next section, a subclass of the compound Poisson distribution is investigated, where only the effect of small clumps are considered. Issues similar to those in Section 4.2 are treated and the results are compared. Comparisons are also made with the Poisson distribution. We show that the rate of convergence is increased as compared with the ordinary Poisson approximation, except for the trivial case of an independent sequence. In this case, no "clumping" occurs, and the extension to the compound Poisson distribution is superfluous. Finally, in Section 4.4, some numerical results are presented.

Throughout the whole chapter, the correlation function is assumed to have finite support. Without this assumption, which restricts the "global" dependence, the calculations are considerably more difficult. The purpose of this chapter is to show that the effect of clumping, arising from the "local" dependence, can be eliminated in the sense that the rate of convergence is increased as compared to ordinary Poisson approximation. It seems reasonable to assume that the improvement in using the compound Poisson distribution for sequences with strong "local" dependence is always the case, independently of whether the sequence has strong "global" dependence or not. The results can probably be generalized to hold under some mixing condition instead.

4.1 Stein's method for compound Poisson Approximation

As mentioned above, we begin this chapter with a review of the compound Poisson distribution, and with the question of how to combine it with Stein's method. We will state two equivalent definitions of the compound Poisson distribution; the first one is better known while the second one is more suitable for the present task.

Definition 4.1 *The random variable X is said to be integer-valued compound Poisson distributed if it has the same distribution as $\sum_{k=1}^{N} Z_k$, where N and $Z_1, Z_2, \ldots$ are independent, $N \sim Po(\mu)$ for some positive μ, and $Z_1, Z_2, \ldots$ are independent, identically distributed non-negative integer-valued random variables.*

Using characteristic functions, we can derive an equivalent definition for X. We have

$$\begin{aligned}\varphi_X(t) = E(e^{itX}) &= E(e^{it\sum_{k=1}^{N} Z_k}) \\ &= E\left(E(e^{itZ_1})^N\right) = g_N(\varphi_Z(t)),\end{aligned}$$

where φ_Z is the characteristic function of Z_1 and g_N the generating function of N,

$$g_N(s) = E(s^N) = e^{\mu(s-1)}.$$

Hence, letting $p_Z(k)$ be the probability function of Z_1,

$$\varphi_X(t) = \exp\left(\sum_{k=0}^{\infty} \mu e^{itk} p_Z(k) - \mu\right). \tag{4.1}$$

On the other hand, let $X_1, X_2, \ldots$ be independent, $X_k \sim \text{Po}(\lambda_k)$, and assume $\sum \lambda_k < \infty$. Let $Y = \sum_{k\geq 1} kX_k$. We have

$$\begin{aligned}\varphi_Y(t) = E(e^{it\sum_{k\geq 1} kX_k}) &= \prod_{k\geq 1} \exp\left(\lambda_k(e^{ikt} - 1)\right) \\ &= \exp\left(\sum_{k\geq 1} \lambda_k(e^{ikt} - 1)\right). \end{aligned} \tag{4.2}$$

Comparing (4.1) and (4.2), we see that the expressions are equivalent if we set $\lambda_k = \mu \cdot p_Z(k)$, $k \geq 1$ and $\mu = (\sum \lambda_k)/(1 - p_Z(0))$. This leads us to the following equivalent definition of the compound Poisson distribution.

Definition 4.2 *Let $\boldsymbol{\lambda}$ denote $\sum_{k\geq 1} \lambda_k \delta_k$, where δ_k is the unit mass at point k and $\lambda_1, \lambda_2, \ldots$ are non-negative real numbers. Assume $\sum_{k\geq 1} \lambda_k < \infty$. A random variable X is said to have an integer-valued compound Poisson distribution, written $X \sim CPo(\boldsymbol{\lambda})$, if X has the same distribution as $\sum_{k\geq 1} k \cdot N_k$, where $N_k \sim Po(\lambda_k)$, $k \geq 1$ are independent.*

Note that by choosing $\lambda_1 > 0$, $\lambda_k = 0$, $k \geq 2$, we get the ordinary Poisson distribution with mean λ_1.

The compound Poisson distribution is often an adequate distribution in models where occurrences of events are counted and the structure of dependence is such that there is a tendency of clumping among the events. The random variables N_k can then be interpreted as the number of clumps of size k, and consequently λ_k is the mean number of clumps of this size. The number of events in the clumps of size k is kN_k, and the total number of events is therefore $\sum_{k\geq 1} kN_k \sim \mathrm{CPo}(\boldsymbol{\lambda})$.

Let Γ denote an arbitrary finite index set and let $\Gamma_i = \Gamma\backslash\{i\}$, $i \in \Gamma$. Let I_i, $i \in \Gamma$ be Bernoulli random variables with

$$P(I_i = 1) = 1 - P(I_i = 0) = p_i$$

and define

$$W = \sum_{i\in\Gamma} I_i, \qquad \lambda_0 = E(W) = \sum_{i\in\Gamma} p_i.$$

Assume that, for each $i \in \Gamma$, we can divide Γ_i into three disjoint subsets Γ_i^{vs}, Γ_i^b and Γ_i^{vw} according to the following informal criteria:

$$\Gamma_i^{vs} = \{j \in \Gamma_i : I_j \text{ is very strongly dependent on } I_i\}$$

$$\Gamma_i^{vw} = \{j \in \Gamma_i : I_j \text{ is very weakly dependent on } \{I_k : k \in \{i\} \cup \Gamma_i^{vs}\}\}$$

$$\Gamma_i^b = \Gamma_i \backslash \{\Gamma_i^{vs} \cup \Gamma_i^{vw}\}.$$

Following the same notation as in Roos [19], set

$$U_i = \sum_{j\in\Gamma_i^{vs}} I_j, \qquad Z_i = I_i + U_i, \qquad X_i = \sum_{j\in\Gamma_i^b} I_j,$$

$$D = \max_{i\in\Gamma}\{|\Gamma_i^{vs}|\}. \tag{4.3}$$

Define

$$\lambda_k = \begin{cases} \dfrac{1}{k}\displaystyle\sum_{i\in\Gamma} E(I_i I(Z_i = k)) & k = 1, \ldots, D+1, \\ \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

and

$$\phi = \sum_{i\in\Gamma} \sum_{k=1}^{|\Gamma_i^{vs}|+1} \phi_{ik}, \tag{4.5}$$

for

$$\phi_{ik} = E\left(|E\left\{I_i I(Z_i = k)|(I_j : j \in \Gamma_i^{vw})\right\} - E\left\{I_i I(Z_i = k)\right\}|\right). \tag{4.6}$$

With these definitions, Z_k can be interpreted as being the size of the clump, if any, at point k, and $\sum E(I_i I(Z_i = k))$ as the mean number of points forming parts of clumps of size k. Consequently, λ_k is the mean number of clumps of size k.

Finally recall the definition of the total variation distance between two random variables W and X,

$$d(W, X) = \sup_A |P(W \in A) - P(X \in A)|,$$

and let $d(W, \mathrm{CPo}(\boldsymbol{\lambda}))$ be an abbreviation for $d(W, X)$, where $X \sim \mathrm{CPo}(\boldsymbol{\lambda})$.

The following theorem from Roos [19] will be of importance in the next section.

Theorem 4.1 *With the definitions above,*

$$d(W, CPo(\boldsymbol{\lambda})) \leq c_2(\boldsymbol{\lambda}) \sum_{i\in\Gamma} \Big((E(I_i))^2 + E(I_i)E(U_i + X_i)$$
$$+E(I_i X_i)\Big) + c_1(\boldsymbol{\lambda})\phi,$$

where $c_1(\boldsymbol{\lambda})$ and $c_2(\boldsymbol{\lambda})$ are constants depending on $\boldsymbol{\lambda}$ only.

From Barbour, Chen and Loh [4], see also Roos [19], we have the following bound valid for both constants $c_1(\boldsymbol{\lambda})$ and $c_2(\boldsymbol{\lambda})$. For any $\boldsymbol{\lambda}$,

$$c_1(\boldsymbol{\lambda}), c_2(\boldsymbol{\lambda}) \;\leq\; \min\left(1, \frac{1}{\lambda_1}\right) \exp\left(\sum_{k=1}^{\infty} \lambda_k\right) \tag{4.7}$$

Further, if $k\lambda_k$ is decreasing in k, the bound can be sharpened:

$$c_1(\boldsymbol{\lambda}) \leq \begin{cases} 1 & \text{if } \lambda_1 - 2\lambda_2 \leq 1, \\ \frac{1}{\sqrt{\lambda_1 - 2\lambda_2}}\left(2 - \frac{1}{\sqrt{\lambda_1 - 2\lambda_2}}\right) & \text{otherwise,} \end{cases} \tag{4.8}$$

$$c_2(\boldsymbol{\lambda}) \leq \min\left(1, \frac{1}{\lambda_1 - 2\lambda_2}\left(\frac{1}{4(\lambda_1 - 2\lambda_2)} + \max\left(0, \ln 2(\lambda_1 - 2\lambda_2)\right)\right)\right). \tag{4.9}$$

The calculation of $\lambda_1, \dots, \lambda_{D+1}$ is not always a straightforward and simple matter. Therefore, instead of using the compound Poisson distribution with parameters λ_k defined by (4.4), we can try to approximate W by a distribution with fewer non-zero parameters. Therefore define, for some integer $2 \leq l \leq D$,

$$\begin{array}{ll} \lambda_k^* = \lambda_k, & k = 2, 3, \dots, l \\ \lambda_k^* = 0, & k \geq l + 1 \\ \lambda_1^* = \lambda_1 + \sum_{j=l+1}^{D+1} j\lambda_j & \end{array} \tag{4.10}$$

This means that we ignore clumps of size $l + 1$ and larger (instead of size $D + 1$ as before) and adjust λ_1 so that the distribution still has the correct mean. The idea is to use this simpler distribution, hoping to keep the same rate of convergence (or suffer only a small loss of speed) as in Theorem 4.1. The next theorem, also from Roos [19], tells us how to bound the total variation distance if we use this new distribution.

Theorem 4.2 *Let λ_k and λ_k^* be defined by (4.4) and (4.10) respectively. Then*

$$d(W, CPo(\boldsymbol{\lambda}^*)) \leq c_1(\boldsymbol{\lambda}^*)\phi + c_2(\boldsymbol{\lambda}^*)$$
$$\times \left(\sum_{i\in\Gamma}\Big((E(I_i))^2 + E(I_i)E(U_i + X_i) + E(I_iX_i)\Big) + \sum_{j=l+1}^{D+1} j(j-1)\lambda_j\right),$$

where ϕ is defined by (4.5)–(4.6). The constants $c_1(\boldsymbol{\lambda}^*)$ and $c_2(\boldsymbol{\lambda}^*)$ are bounded in a similar way as $c_1(\boldsymbol{\lambda})$ and $c_2(\boldsymbol{\lambda})$ in Theorem 4.1.

4.2 Best possible compound Poisson approximation

In this section, we apply the general theorems in the previous section to stationary Gaussian sequences with dependence of finite range, i.e.

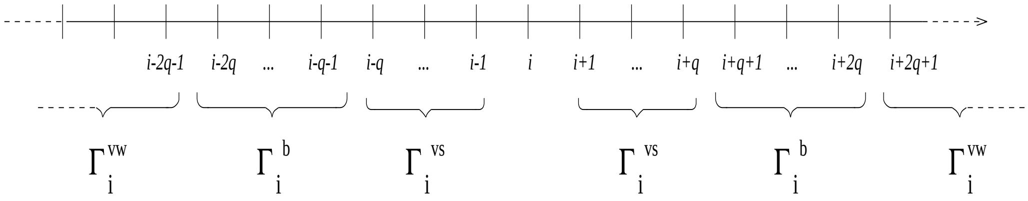

Figure 4.1: *Definition of* Γ_i^{vs}, Γ_i^b *and* Γ_i^{vw}.

m-dependent sequences. Typical examples of such m-dependent sequences are MA(q)-sequences. In Section 4.4, explicit calculations will be carried out numerically on two MA-sequences, and the results will be compared with ordinary Poisson approximation.

Assume $\{Y_i\}_{i=1}^n$ is a stationary Gaussian sequence with correlation function $\{\rho_k\}$, where

$$\rho_k = 0, \qquad k > q, \tag{4.11}$$

for some finite q. Assume further that the sequence is standardized, i.e. $E(Y_i) = 0$, $\mathrm{Var}(Y_i) = 1$. Let $\Gamma = \{1, 2, \dots, n\}$, $D = 2q$. For $i \in \Gamma$, set

$$\begin{aligned}
\Gamma_i &= \Gamma \backslash \{i\}, \\
\Gamma_i^{vs} &= \{j \in \Gamma_i : |i - j| \le q\}, \\
\Gamma_i^{b} &= \{j \in \Gamma_i : q < |i - j| \le 2q\}, \\
\Gamma_i^{vw} &= \Gamma_i \backslash \{\Gamma_i^{vs} \cup \Gamma_i^b\},
\end{aligned}$$

see Figure 4.1. Moreover, define

$$I_i = I(Y_i > a), \qquad W = \sum_{i \in \Gamma} I_i, \qquad \lambda_0 = E(W), \tag{4.12}$$

where a is a positive real number. Let Φ denote the cumulative distribution function for a standardized Gaussian random variable. Following the notation in Section 4.1,

$$\phi_{ik} = E\left(\left|E\left\{I_i I(Z_i = k) | (I_j : j \in \Gamma_i^{vw})\right\} - E\left\{I_i I(Z_i = k)\right\}\right|\right) = 0$$

due to condition (4.11), and consequently

$$\phi = 0.$$

Using stationarity,

$$\sum_{i\in\Gamma} E(I_i)^2 = nP(Y_1 > a)^2 = n(1-\Phi(a))^2,$$

$$\sum_{i\in\Gamma} E(I_i)E(U_i+X_i) = \sum_{i\in\Gamma}\sum_{j\in\Gamma_i^{vs}\cup\Gamma_i^b} E(I_i)E(I_j) \le n4q(1-\Phi(a))^2,$$

$$\sum_{i\in\Gamma} E(I_iX_i) = \sum_{i\in\Gamma}\sum_{j\in\Gamma_i^b} E(I_iI_j) \le n2q(1-\Phi(a))^2,$$

where the last inequality follows from the independence between I_i and I_j, $j\in\Gamma_i^b$ due to (4.11). Summing up,

$$\sum_{i\in\Gamma} E(I_i)^2 + E(I_i)E(U_i+X_i) + E(I_iX_i) \le (1+6q)n(1-\Phi(a))^2. \quad (4.13)$$

Finally, to specify the approximating compound Poisson distribution, we must determine the λ_k's. For $k\in\{1,2,\ldots,2q+1\}$,

$$\lambda_k = \frac{1}{k}\sum_{i\in\Gamma} E(I_iI(Z_i=k)) = \frac{1}{k}\sum_{i\in\Gamma} P(I_i=1\cap U_i=k-1), \qquad (4.14)$$

and $\lambda_k=0$, $k>2q+1$. However, for $i\in\{q+1,q+2,\ldots,n-q\}$, and with obvious modifications for points i near the endpoints of Γ,

$$P(I_i=1\,\cap\,U_i=k-1)$$
$$=\sum P(Y_i>a, Y_{\alpha_1}>a, Y_{\alpha_2}>a,\ldots,Y_{\alpha_{k-1}}>a, Y_{\alpha_k}<a,\ldots,Y_{\alpha_{2q}}<a), \qquad (4.15)$$

where the sum runs over all combinations $\alpha_1,\alpha_2,\ldots,\alpha_{2q}$ of length $2q$ of Γ_i^{vs}. In other words, the main part of the computation of the λ_k's consists of calculations of the form of (4.15). In Section 4.4, such calculations are carried out numerically.

Combining Theorem 4.1 and the derivation above gives the next theorem.

Theorem 4.3 *Let W and $\boldsymbol{\lambda}$ be defined by (4.12) and (4.14) respectively. Then*

$$d(W, CPo(\boldsymbol{\lambda})) \le c_2(\boldsymbol{\lambda})(1+6q)n(1-\Phi(a))^2,$$

where $c_2(\boldsymbol{\lambda})$ is bounded by (4.7). If $k\lambda_k$ is decreasing in k, the sharper bound (4.9) is valid.

4.2.1 An asymptotic bound for the total variation distance

Theorem 4.3 gives an upper bound for the total variation distance for any positive a and n. A natural thing to investigate is how this bound varies as n and a increase while retaining the mean value, λ_0, of W fixed. This restriction gives a unique relation between $a = a_n$ and n,

$$n(1 - \Phi(a_n)) = \lambda_0. \tag{4.16}$$

This implies $1 - \Phi(a_n) = \lambda_0/n$, and, when $n \to \infty$,

$$\varphi(a_n)/a_n \sim \lambda_0/n$$

$$a_n \sim \sqrt{2\ln n}$$

$$1 - \Phi(c_1 a_n) \sim \frac{\varphi(c_1 a_n)}{c_1 a_n} \sim \frac{\left(\sqrt{2\pi}\right)^{c_1^2-1}}{c_1}\left(\frac{\lambda_0}{n}\right)^{c_1^2}\left(\sqrt{2\ln n}\right)^{c_1^2-1} \tag{4.17}$$

$$\varphi(c_2 a_n) \sim \left(\sqrt{2\pi}\right)^{c_2^2-1}\left(\frac{\lambda_0}{n}\sqrt{2\ln n}\right)^{c_2^2} \tag{4.18}$$

where φ is the density function for a standardized Gaussian random variable, and $c_1 > 0$ and c_2 are constants. Relations (4.17) and (4.18) follow from the simple calculation

$$\frac{\varphi(c_1 a_n)}{c_1 a_n} = \frac{1}{c_1 a_n\sqrt{2\pi}} e^{-(c_1 a_n)^2/2} = \frac{1}{c_1}\left(\frac{1}{a_n\sqrt{2\pi}}\right)^{1+c_1^2-c_1^2}\left(e^{-a_n^2/2}\right)^{c_1^2}$$

$$= \frac{1}{c_1}\left(\frac{1}{a_n\sqrt{2\pi}}e^{-a_n^2/2}\right)^{c_1^2}\left(\frac{1}{a_n\sqrt{2\pi}}\right)^{1-c_1^2}$$

$$\sim \frac{\left(\sqrt{2\pi}\right)^{c_1^2-1}}{c_1}\left(\frac{\lambda_0}{n}\right)^{c_1^2}\left(\sqrt{2\ln n}\right)^{c_1^2-1}$$

Of course, if $c_1 < 0$, $1 - \Phi(c_1 a_n) \to 1$ and hence (4.17) does not hold.

Theorem 4.4 *Let $\{a_n\}$ be a real, positive sequence satisfying (4.16), where $\lambda_0 = E(W)$ is constant. Then*

$$d(W, CPo(\boldsymbol{\lambda})) = O\left(\frac{1}{n}\right)$$

as $n \to \infty$.

Proof. Since only a finite number of the elements in $\{\lambda_k\}$ are non-zero, the bound (4.7) of c_2 is finite. The theorem then follows from Theorem 4.3 and equation (4.16). □

4.3 Simplest compound Poisson approximation

Instead of using a compound Poisson distribution with all $\lambda_k > 0$, $k = 1, \ldots, 2q+1$, we will in this section only use the first two terms, i.e. a distribution of the form

$$X_1 + 2X_2, \tag{4.19}$$

where X_1 and X_2 are independent and $X_i \sim \mathrm{Po}(\lambda_i)$, $i = 1, 2$. This means that we assume that clumps of size three or larger are rare compared to single exceeding points and clumps of size two. The effectiveness of this approach is strongly dependent on how correlated the underlying random sequence is. Apart from the constant c_2, Theorem 4.1 differs from Theorem 4.2 only by the term

$$\sum_{j=l+1}^{D+1} j(j-1)\lambda_j. \tag{4.20}$$

It follows from Theorem 4.4 that the other terms are of order $1/n$. Consider now for instance the first omitted term, λ_3. As already noted in Section 4.2, λ_3 is (leaving out constants) a finite sum of terms of the form

$$\begin{aligned} &nP(\xi_1 > a, \xi_2 > a, \xi_3 > a, \xi_4 < a, \ldots, \xi_{2q+1} < a) \\ &\qquad \leq nP(\xi_1 > a, \xi_2 > a, \xi_3 > a), \end{aligned} \tag{4.21}$$

where $\xi_1 = Y_i$, $\xi_2 = Y_{\alpha_1}$, $\xi_3 = Y_{\alpha_2}, \ldots, \xi_{2q+1} = Y_{\alpha_{2q}}$ for some combination α_i defined as in (4.15). Assume now that (4.16) holds. If ξ_1, ξ_2 and ξ_3 are independent and $a = a_n$,

$$nP(\xi_1 > a, \xi_2 > a, \xi_3 > a) = nP(\xi_1 > a)^3 = n(1 - \Phi(a))^3 = O\left(\frac{1}{n^2}\right).$$

On the other hand, if $\xi_1 = \xi_2 = \xi_3$,

$$nP(\xi_1 > a, \xi_2 > a, \xi_3 > a) = nP(\xi_1 > a) = O(1).$$

For a strongly correlated sequence, λ_3 will dominate (i.e. approaches zero slower than) the terms in equation (4.13). In fact, even λ_4, $\lambda_5, \ldots$ can be more dominant than (4.13). If this is the case it means that large clumps are occurring frequently and the omission of λ_3, $\lambda_4, \ldots$ is unsuitable. All terms more dominant than $1/n^2$ must be taken into account and not left out in the approximating compound Poisson distribution in order to preserve the same rate of convergence as in Theorem 4.4. However, for a weakly or moderately correlated sequence (the exact condition will be stated in Subsection 4.3.1), omitting terms is appropriate and results in little or no loss in the speed of convergence.

In Subsection 4.3.1 an asymptotic bound for λ_3 as a function of the correlations is derived. The calculations are straightforward but tedious. Subsection 4.3.2 indicates how to find those correlations which really determine the speed of convergence. Finally, in Subsection 4.3.3, results from the two subsections just mentioned are combined with the results in the previous sections, and a bound for the total variation distance between W and a suitable compound Poisson distribution of the form (4.19) is given.

4.3.1 An asymptotic bound for the omitted terms

From (4.21), as a first step to bound λ_3, we derive an upper bound for probabilities of the form

$$P(\xi_1 > a, \xi_2 > a, \xi_3 > a), \tag{4.22}$$

where (ξ_1, ξ_2, ξ_3) is a Gaussian random vector. From Holst and Janson [9], the following lemma handles the case of only two random variables.

Lemma 4.5 *Let* (ξ_1, ξ_2) *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,

$$(1 - \Phi(a))\left(1 - \Phi\left(\frac{b - ra}{\sqrt{1 - r^2}}\right)\right) \le P(\xi_1 > a \cap \xi_2 > b)$$

$$\leq (1-\Phi(a))\left\{\left(1-\Phi\left(\frac{b-ra}{\sqrt{1-r^2}}\right)\right)+r\frac{\varphi(b)}{\varphi(a)}\left(1-\Phi\left(\frac{a-rb}{\sqrt{1-r^2}}\right)\right)\right\}.$$

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

Before stating a counterpart for three variables, we need some notations and supportive lemmas. Lemmas 4.6 and 4.7 present some preliminary results. By conditioning on one variable, Lemma 4.5 can also be used in the three-variable case. This leads to two complicated expressions (cf. the two terms in Lemma 4.5) which are treated in Lemma 4.8 and Lemma 4.9 respectively. The result is summarized in Proposition 4.10 and in some subsequent corollaries. Finally, the term (4.20) is estimated in Proposition 4.13.

Let (ξ_1, ξ_2, ξ_3) be a Gaussian random vector with non-singular covariance matrix

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}. \tag{4.23}$$

Assume, without loss of generality, that $\rho_{12} \leq \rho_{13} \leq \rho_{23}$. Let

$$r = \frac{\rho_{23} - \rho_{12}\rho_{13}}{\sqrt{(1-\rho_{12}^2)(1-\rho_{13}^2)}}. \tag{4.24}$$

Define, for $\rho_{12}, \rho_{13}, \rho_{23} \geq 0$ satisfying $r < 1$,

$$\gamma_1 = \frac{1}{\sqrt{1-\rho_{12}^2}}, \tag{4.25}$$

$$\gamma_2 = \frac{\rho_{12}}{\sqrt{1-\rho_{12}^2}}, \tag{4.26}$$

$$\begin{aligned}\gamma_3 &= \frac{1}{\sqrt{1-r^2}}\left(\frac{1}{\sqrt{1-\rho_{13}^2}} - \frac{r}{\sqrt{1-\rho_{12}^2}}\right) \\ &= \frac{1-\rho_{12}^2-\rho_{23}+\rho_{12}\rho_{13}}{(1-\rho_{12}^2)\sqrt{(1-r^2)(1-\rho_{13}^2)}},\end{aligned} \tag{4.27}$$

$$\gamma_4 = \frac{1}{\sqrt{1-r^2}}\left(\frac{\rho_{13}}{\sqrt{1-\rho_{13}^2}} - \frac{r\rho_{12}}{\sqrt{1-\rho_{12}^2}}\right)$$

$$= \frac{\rho_{13} - \rho_{12}\rho_{23}}{(1-\rho_{12}^2)\sqrt{(1-r^2)(1-\rho_{13}^2)}}, \tag{4.28}$$

$$\gamma_5 = \frac{1}{\sqrt{1-r^2}}\left(\frac{\rho_{12}}{\sqrt{1-\rho_{12}^2}} - \frac{r\rho_{13}}{\sqrt{1-\rho_{13}^2}}\right), \tag{4.29}$$

$$\gamma_6 = \frac{1}{\sqrt{1-r^2}}\left(\frac{1}{\sqrt{1-\rho_{12}^2}} - \frac{r}{\sqrt{1-\rho_{13}^2}}\right), \tag{4.30}$$

$$\gamma_7 = \sqrt{\frac{1-2\rho_{12}^2+\rho_{12}^2\rho_{13}^2}{(1-\rho_{12}^2)(1-\rho_{13}^2)}}, \tag{4.31}$$

$$\gamma_8 = \frac{1}{\gamma_7}\left(\frac{\rho_{13}}{1-\rho_{13}^2} - \frac{\rho_{12}}{1-\rho_{12}^2}\right), \tag{4.32}$$

$$\gamma_9 = \sqrt{\frac{2\rho_{12}(\rho_{13}-\rho_{12})}{1-2\rho_{12}^2+\rho_{12}^2\rho_{13}^2}}. \tag{4.33}$$

Note that (4.24)–(4.30) are well-defined since the non-singularity of (4.23) implies $\rho_{12}, \rho_{13}, \rho_{23} < 1$. Moreover $\rho_{13} \geq \rho_{12}$ implies $1 - 2\rho_{12}^2 + \rho_{12}^2\rho_{13}^2 \geq 1 - 2\rho_{12}^2 + \rho_{12}^4 = (1-\rho_{12}^2)^2 \geq 0$ so that the expressions inside the square roots in (4.31) and (4.33) are non-negative. In Lemma 4.6, it is shown that $\gamma_7 > 0$, making (4.32) well-defined.

Finally, define

$$h_1(\rho_{12},\rho_{13},\rho_{23}) = \Big(3 - 2(\rho_{12}+\rho_{13}+\rho_{23}) - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2$$
$$+2(\rho_{12}\rho_{13}+\rho_{12}\rho_{23}+\rho_{13}\rho_{23})\Big)\Big/\Big(1+2\rho_{12}\rho_{13}\rho_{23} - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2\Big), \tag{4.34}$$

$$h_2(\rho_{12},\rho_{13},\rho_{23}) = -\Big(\rho_{12}+\rho_{13}+\rho_{23} - \rho_{12}\rho_{13} - \rho_{12}\rho_{23} - \rho_{13}\rho_{23} - \rho_{12}^2$$
$$-\rho_{13}^2 - \rho_{23}^2 + 3\rho_{12}\rho_{13}\rho_{23}\Big)\Big/\Big(1+2\rho_{12}\rho_{13}\rho_{23} - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2\Big) \tag{4.35}$$

and

$$h_3(\rho_{12},\rho_{13},\rho_{23}) = 1+\rho_{12}-\rho_{13}-\rho_{12}^2-\rho_{23}^2-\rho_{12}\rho_{13}+\rho_{12}\rho_{23}+\rho_{13}\rho_{23}$$
$$-\rho_{12}^3-\rho_{12}\rho_{13}^2+\rho_{12}^2\rho_{23}-2\rho_{12}\rho_{23}^2+\rho_{12}\rho_{13}\rho_{23}+2\rho_{12}^2\rho_{13}\rho_{23}. \tag{4.36}$$

Lemma 4.6 *Let r, γ_1, γ_2, $\ldots$, γ_9 and $h_3(\rho_{12},\rho_{13},\rho_{23})$ be defined by (4.24)-(4.33) and (4.36) respectively, where $0 \le \rho_{12} \le \rho_{13} \le \rho_{23} < 1$ and $0 \le r < 1$. Then*

$$(i)\qquad 1+\gamma_4^2-\gamma_3\gamma_4 = \frac{h_3(\rho_{12},\rho_{13},\rho_{23})}{(1+\rho_{12})(1+2\rho_{12}\rho_{13}\rho_{23}-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2)} > 0,$$

$$(ii)\qquad 1+\gamma_4^2+\gamma_2^2-\gamma_3\gamma_4-\gamma_1\gamma_2 = \frac{(1-\rho_{23})(1+\rho_{23}-\rho_{12}-\rho_{13})}{1+2\rho_{12}\rho_{13}\rho_{23}-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2} > 0,$$

$$(iii)\qquad \gamma_7 = \sqrt{\frac{1-2\rho_{12}^2+\rho_{12}^2\rho_{13}^2}{(1-\rho_{12}^2)(1-\rho_{13}^2)}} \ge 1,$$

$$(iv)\qquad \gamma_7-\gamma_8 = \frac{(1+2\rho_{12}+\rho_{12}\rho_{13})\sqrt{(1-\rho_{12})(1-\rho_{13})}}{\sqrt{(1+\rho_{13})(1+\rho_{12})(1-2\rho_{12}^2+\rho_{12}^2\rho_{13}^2)}} > 0,$$

$$(v)\qquad \gamma_2^2+\gamma_7^2-\gamma_1\gamma_2-\gamma_7\gamma_8 = \frac{1}{1+\rho_{13}} \ge 1,$$

$$(vi)\qquad \gamma_6-\gamma_5 = \frac{(1+\rho_{13}-\rho_{12}-\rho_{23})\sqrt{1-\rho_{13}}}{\sqrt{(1+\rho_{13})(1+2\rho_{12}\rho_{13}\rho_{23}-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2)}} > 0,$$

$$(vii)\qquad \gamma_5^2+\gamma_7^2-\gamma_5\gamma_6-\gamma_7\gamma_8 = \frac{h_3(\rho_{12},\rho_{13},\rho_{23})}{(1+\rho_{12})(1+2\rho_{12}\rho_{13}\rho_{23}-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2)} > 0,$$

$$(viii)\qquad \gamma_5^2+\gamma_2^2+\gamma_7^2-\gamma_5\gamma_6-\gamma_1\gamma_2-\gamma_7\gamma_8 = \frac{(1-\rho_{23})(1+\rho_{23}-\rho_{12}-\rho_{13})}{1+2\rho_{12}\rho_{13}\rho_{23}-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2} > 0.$$

Proof. Only the inequalities are proved here. The equalities follow by simple, but tedious, calculations and can be verified by any computer program handling symbolic expressions (e.g. MAPLE).

(i) For fixed ρ_{12} and ρ_{13}, h_3 as a function of ρ_{23} is a quadratic expression with negative second order derivative, implying the lowest values of h_3 at the boundaries, i.e $\rho_{23} = \rho_{13}$, or

$$\rho_{23} = \rho_{12}\rho_{13} + \sqrt{(1-\rho_{12}^2)(1-\rho_{13}^2)}.$$

Assume first that $\rho_{13} = \rho_{23}$. Then h_3 simplifies to

$$h_3(\rho_{12}, \rho_{13}, \rho_{13}) = (1-\rho_{12})(1 + 2\rho_{12} + \rho_{12}^2 - \rho_{13} - \rho_{12}\rho_{13} - 2\rho_{12}\rho_{13}^2).$$

If $\rho_{13} \leq 1/\sqrt{2}$,

$$h_3(\rho_{12}, \rho_{13}, \rho_{13}) \geq (1-\rho_{12})(1 + 2\rho_{12} + \rho_{12}^2 - 1/\sqrt{2} - \rho_{12}/\sqrt{2} - \rho_{12}) > 0.$$

If $\rho_{13} > 1/\sqrt{2}$, the condition $r < 1$ implies $\rho_{13}^2 < (1+\rho_{12})/2$. Hence

$$\begin{aligned} h_3(\rho_{12}, \rho_{13}, \rho_{13}) &> (1-\rho_{12})\Big(1 + 2\rho_{12} + \rho_{12}^2 - \rho_{13} - \rho_{12}\rho_{13} \\ &-2\rho_{12}\frac{1+\rho_{12}}{2}\Big) = (1-\rho_{12})(\rho_{12} - \rho_{12}\rho_{13} + 1 - \rho_{13}) \geq 0. \end{aligned}$$

Next, turn to the case $\rho_{23} = \rho_{12}\rho_{13} + \sqrt{(1-\rho_{12}^2)(1-\rho_{13}^2)}$. Substituting ρ_{23} in h_3, the expression $1 + \rho_{12}$ factors out. Thus for

$$f(\rho_{12}, \rho_{13}) = h_3\left(\rho_{12}, \rho_{13}, \rho_{12}\rho_{13} + \sqrt{(1-\rho_{12}^2)(1-\rho_{13}^2)}\right)/(1+\rho_{12})$$

we have

$$\begin{aligned} f(\rho_{12}, \rho_{13}) = (\rho_{12} + \rho_{13} - 2\rho_{12}\rho_{13})&\sqrt{(1-\rho_{12}^2)(1-\rho_{13}^2)} \\ -(1-\rho_{13})(1-\rho_{12})&(\rho_{12} + \rho_{13} + 2\rho_{12}\rho_{13}). \end{aligned}$$

Since the first term in f is positive and the second is negative, it is enough to show that

$$f_1(\rho_{12}, \rho_{13}) - f_2(\rho_{12}, \rho_{13}) > 0,$$

where

$$f_1(\rho_{12},\rho_{13}) = (\rho_{12}+\rho_{13}-2\rho_{12}\rho_{13})^2\,(1-\rho_{12}^2)(1-\rho_{13}^2)$$

and

$$f_2(\rho_{12},\rho_{13}) = \left\{(1-\rho_{13})(1-\rho_{12})(\rho_{12}+\rho_{13}+2\rho_{12}\rho_{13})\right\}^2.$$

But $f_1(\rho_{12},\rho_{13}) - f_2(\rho_{12},\rho_{13})$ can be factorized as

$$2(1-\rho_{12})(1-\rho_{13})(\rho_{12}+\rho_{13})(\rho_{13}-\rho_{12})^2 \geq 0.$$

Since none of the points where $f_2 - f_1 = 0$ satisfies $r < 1$, the proof is complete.

(ii) Obviously the numerator is strictly positive. From (4.24), $r < 1$ implies $(\rho_{23}-\rho_{12}\rho_{13})^2 < (1-\rho_{12}^2)(1-\rho_{13}^2)$. Consequently

$$1-\rho_{12}^2-\rho_{13}^2-(\rho_{23}^2-2\rho_{12}\rho_{13}\rho_{23}) > 0,$$

so also the denominator is strictly positive.

(iii)

$$\frac{1-2\rho_{12}^2+\rho_{12}^2\rho_{13}^2}{(1-\rho_{12}^2)(1-\rho_{13}^2)} \geq \frac{1-2\rho_{12}^2+\rho_{12}^4}{(1-\rho_{12}^2)^2} = 1$$

(iv)-(viii) Follows immediately from the proofs of (i)-(iii).

□

Lemma 4.7 *For any β_1,β_2,β_4 and any $a>0$, $\beta_3>0$ such that $\beta_3 \geq \beta_4$,*

$$\begin{aligned}
&\int_a^\infty (1-\Phi(\beta_1 a-\beta_2 x))\varphi(\beta_3 x-\beta_4 a)\,dx \\
&\leq \frac{1}{\beta_3}(1-\Phi(a(\beta_1-\beta_2)))(1-\Phi(a(\beta_3-\beta_4))) \\
&+\frac{\beta_2}{\beta_3\sqrt{\beta_2^2+\beta_3^2}}\frac{1-\Phi(a(\beta_3-\beta_4))}{\varphi(a(\beta_3-\beta_4))}\varphi\left(a\frac{\beta_1\beta_3-\beta_2\beta_4}{\sqrt{\beta_2^2+\beta_3^2}}\right) \\
&\times\left(1-\Phi\left(a\left(\frac{\beta_2^2+\beta_3^2-\beta_1\beta_2-\beta_3\beta_4}{\sqrt{\beta_2^2+\beta_3^2}}\right)\right)\right)
\end{aligned}$$

Proof. Integrating by parts and observing that $(1-\Phi(x))/\varphi(x)$ is decreasing for $x \geq 0$,

$$\begin{aligned}
&\int_a^\infty (1-\Phi(\beta_1 a - \beta_2 x))\varphi(\beta_3 x - \beta_4 a)\,dx \\
&= \left[-\frac{1}{\beta_3}(1-\Phi(\beta_1 a - \beta_2 x))(1-\Phi(\beta_3 x - \beta_4 a))\right]_a^\infty \\
&+\frac{\beta_2}{\beta_3}\int_a^\infty \varphi(\beta_1 a - \beta_2 x)(1-\Phi(\beta_3 x - \beta_4 a))\,dx \\
&\leq \frac{1}{\beta_3}(1-\Phi(a(\beta_1-\beta_2)))(1-\Phi(a(\beta_3-\beta_4))) \\
&+\frac{\beta_2}{\beta_3}\frac{1-\Phi(a(\beta_3-\beta_4))}{\varphi(a(\beta_3-\beta_4))}\int_a^\infty \varphi(\beta_1 a - \beta_2 x)\varphi(\beta_3 x - \beta_4 a)\,dx \\
&\leq \frac{1}{\beta_3}(1-\Phi(a(\beta_1-\beta_2)))(1-\Phi(a(\beta_3-\beta_4))) \\
&+\frac{\beta_2}{\beta_3}\frac{1-\Phi(a(\beta_3-\beta_4))}{\varphi(a(\beta_3-\beta_4))}\varphi\left(a\frac{\beta_1\beta_3-\beta_2\beta_4}{\sqrt{\beta_2^2+\beta_3^2}}\right) \\
&\times\int_a^\infty \varphi\left(\frac{(\beta_2^2+\beta_3^2)x-(\beta_1\beta_2+\beta_3\beta_4)a}{\sqrt{\beta_2^2+\beta_3^2}}\right)\,dx,
\end{aligned}$$

from which the statement follows.

□

The next two lemmas treat the two expressions arising in Proposition 4.10 from the use of Lemma 4.5 on the three variables (ξ_1, ξ_2, ξ_3) in (4.22).

Lemma 4.8 *Let a be a real parameter defined by $n(1-\Phi(a)) = \lambda_0$ for some constant λ_0. Let h_1, h_2 and h_3 be defined by (4.34)-(4.36). If $0 \leq \rho_{12} \leq \rho_{13} \leq \rho_{23} < 1$ and $0 \leq r < 1$, then for all $n \geq 2$,*

$$\begin{aligned}
&\int_a^\infty (1-\Phi(\gamma_1 a - \gamma_2 x))\,(1-\Phi(\gamma_3 a - \gamma_4 x))\,\varphi(x)\,dx \\
&\leq C(1-\rho_{12})^{-1}(1-\rho_{23})^{-1}(1+\rho_{23}-\rho_{12}-\rho_{13})^{-1}(h_3(\rho_{12},\rho_{13},\rho_{23}))^{-1}
\end{aligned}$$

$$\times \left(\frac{1}{n}\right)^{g_1(\rho_{12},\rho_{13},\rho_{23})} (\ln n)^{g_2(\rho_{12},\rho_{13},\rho_{23})},$$

where

$$g_1(\rho_{12},\rho_{13},\rho_{23}) = \begin{cases} h_1(\rho_{12},\rho_{13},\rho_{23}) & \text{if } 1+\rho_{12}-\rho_{13}-\rho_{23}>0 \\ 2/(1+\rho_{12}) & \text{otherwise,} \end{cases}$$

$$g_2(\rho_{12},\rho_{13},\rho_{23}) = \begin{cases} h_2(\rho_{12},\rho_{13},\rho_{23}) & \text{if } 1+\rho_{12}-\rho_{13}-\rho_{23}>0 \\ -\rho_{12}/(1+\rho_{12}) & \text{otherwise,} \end{cases}$$

and C is a constant depending on λ_0 only.

Proof. First note the relations

$$\frac{\gamma_1}{\sqrt{1+\gamma_2^2}} = 1, \qquad \sqrt{1+\gamma_2^2} = \gamma_1. \tag{4.37}$$

By integrating by parts,

$$\int_a^\infty (1-\Phi(\gamma_1 a - \gamma_2 x))\,(1-\Phi(\gamma_3 a - \gamma_4 x))\,\varphi(x)\,dx$$

$$= (1-\Phi(a(\gamma_1-\gamma_2)))\,(1-\Phi(a(\gamma_3-\gamma_4)))\,(1-\Phi(a)) \tag{4.38}$$

$$+ \int_a^\infty \gamma_2\varphi(\gamma_1 a-\gamma_2 x)\,(1-\Phi(\gamma_3 a-\gamma_4 x))\,(1-\Phi(x))\,dx \tag{4.39}$$

$$+ \int_a^\infty \gamma_4\,(1-\Phi(\gamma_1 a-\gamma_2 x))\,\varphi(\gamma_3 a-\gamma_4 x)(1-\Phi(x))\,dx. \tag{4.40}$$

Using the fact that $(1-\Phi(x))/\varphi(x)$ is decreasing in x for $x \geq 0$, the sum of (4.39) and (4.40) equals

$$\int_a^\infty \gamma_2\varphi(\gamma_1 a-\gamma_2 x)\,(1-\Phi(\gamma_3 a-\gamma_4 x))\,(1-\Phi(x))\,dx$$

$$+ \int_a^\infty \gamma_4\,(1-\Phi(\gamma_1 a-\gamma_2 x))\,\varphi(\gamma_3 a-\gamma_4 x)(1-\Phi(x))\,dx$$

$$\leq \frac{\gamma_2(1-\Phi(a))}{\varphi(a)} \int_a^\infty \varphi(x)\varphi(\gamma_1 a-\gamma_2 x)(1-\Phi(\gamma_3 a-\gamma_4 x))\,dx$$

$$
\begin{aligned}
&+\frac{\gamma_4(1-\Phi(a))}{\varphi(a)}\int_a^\infty \varphi(x)(1-\Phi(\gamma_1 a-\gamma_2 x))\varphi(\gamma_3 a-\gamma_4 x)\,dx \\
=&\ \frac{\gamma_2(1-\Phi(a))}{\varphi(a)}\varphi\left(\frac{\gamma_1 a}{\sqrt{1+\gamma_2^2}}\right)\int_a^\infty \varphi\left(\frac{(1+\gamma_2^2)x-\gamma_1\gamma_2 a}{\sqrt{1+\gamma_2^2}}\right) \\
&\times(1-\Phi(\gamma_3 a-\gamma_4 x))\,dx \\
&+\frac{\gamma_4(1-\Phi(a))}{\varphi(a)}\varphi\left(\frac{\gamma_3 a}{\sqrt{1+\gamma_4^2}}\right)\int_a^\infty \varphi\left(\frac{(1+\gamma_4^2)x-\gamma_3\gamma_4 a}{\sqrt{1+\gamma_4^2}}\right) \\
&\times(1-\Phi(\gamma_1 a-\gamma_2 x))\,dx \\
=&\ \gamma_2(1-\Phi(a))\int_a^\infty \varphi(\gamma_1 x-\gamma_2 a)(1-\Phi(\gamma_3 a-\gamma_4 x))\,dx \\
&+\frac{\gamma_4(1-\Phi(a))}{\varphi(a)}\varphi\left(\frac{\gamma_3 a}{\sqrt{1+\gamma_4^2}}\right)\int_a^\infty \varphi\left(\frac{(1+\gamma_4^2)x-\gamma_3\gamma_4 a}{\sqrt{1+\gamma_4^2}}\right) \\
&\times(1-\Phi(\gamma_1 a-\gamma_2 x))\,dx
\end{aligned}
\tag{4.41}
$$

where in the last equality, the relations in (4.37) between γ_1 and γ_2 have been used. From (4.25)-(4.26), $\gamma_1 > 0$ and $\gamma_1 > \gamma_2$. Thus Lemma 4.7 can be applied on the first integral in (4.41). Moreover, since $\sqrt{1+\gamma_4^2} > 0$, and $1+\gamma_4^2 > \gamma_3\gamma_4$ by Lemma 4.6 (i), Lemma 4.7 can also be used on the second integral. We get

$$
\begin{aligned}
&\int_a^\infty (1-\Phi(\gamma_1 a-\gamma_2 x))\,(1-\Phi(\gamma_3 a-\gamma_4 x))\,\varphi(x)\,dx \\
&\le (1-\Phi(a))\Big\{(1+\rho_{12})(1-\Phi(a(\gamma_1-\gamma_2)))(1-\Phi(a(\gamma_3-\gamma_4))) \\
&+\frac{\rho_{12}\gamma_4}{\sqrt{\gamma_1^2+\gamma_4^2}}(1-\Phi(a(\gamma_1-\gamma_2)))\left(1-\Phi\left(a\left(\frac{\gamma_1^2+\gamma_4^2-\gamma_1\gamma_2-\gamma_3\gamma_4}{\sqrt{\gamma_1^2+\gamma_4^2}}\right)\right)\right) \\
&\times\varphi\left(a\sqrt{\frac{1-\rho_{23}}{1+\rho_{23}}}\right)\Big(\varphi(a(\gamma_1-\gamma_2))\Big)^{-1}+\frac{\gamma_4}{\sqrt{1+\gamma_4^2}}\varphi\left(\frac{a\gamma_3}{\sqrt{1+\gamma_4^2}}\right)\Big(\varphi(a)\Big)^{-1}
\end{aligned}
$$

$$
\times \left(1-\Phi\left(a\left(\frac{1+\gamma_4^2-\gamma_3\gamma_4}{\sqrt{1+\gamma_4^2}}\right)\right)\right)\big(1-\Phi(a(\gamma_1-\gamma_2))\big)
$$

$$
+\frac{\gamma_2\gamma_4}{\sqrt{(1+\gamma_4^2)(1+\gamma_4^2+\gamma_2^2)}}\varphi\left(\frac{a\gamma_3}{\sqrt{1+\gamma_4^2}}\right)\big(\varphi(a)\big)^{-1}
$$

$$
\times \left(1-\Phi\left(a\left(\frac{1+\gamma_4^2-\gamma_3\gamma_4}{\sqrt{1+\gamma_4^2}}\right)\right)\right)\varphi\left(a\left(\frac{\gamma_1+\gamma_1\gamma_4^2-\gamma_2\gamma_3\gamma_4}{\sqrt{(1+\gamma_4^2)(1+\gamma_4^2+\gamma_2^2)}}\right)\right)
$$

$$
\times\left(\varphi\left(a\left(\frac{1+\gamma_4^2-\gamma_3\gamma_4}{\sqrt{1+\gamma_4^2}}\right)\right)\right)^{-1}\left(1-\Phi\left(a\left(\frac{1+\gamma_4^2+\gamma_2^2-\gamma_3\gamma_4-\gamma_1\gamma_2}{\sqrt{1+\gamma_4^2+\gamma_2^2}}\right)\right)\right)\Bigg\}.
$$

Finally, our intention is to use the relations (4.17)-(4.18) on the expression above. Assume first that $\gamma_3-\gamma_4>0$. By Lemma 4.6 (ii) and (4.37), $1+\gamma_4^2+\gamma_2^2-\gamma_3\gamma_4-\gamma_1\gamma_2=\gamma_1^2+\gamma_4^2-\gamma_1\gamma_2-\gamma_3\gamma_4>0$. Moreover, $1+\gamma_4^2-\gamma_3\gamma_4>0$ by Lemma 4.6 (i). Thus the condition "$c_1>0$" in (4.17) is satisfied for all terms. Consequently,

$$
\int_a^\infty (1-\Phi(\gamma_1 a-\gamma_2 x))\,(1-\Phi(\gamma_3 a-\gamma_4 x))\,\varphi(x)\,dx
$$

$$
\leq C\left(\sqrt{\ln n}\right)^{-3}\left\{\frac{1}{(\gamma_1-\gamma_2)}\left(\frac{1}{n}\sqrt{\ln n}\right)^{(\gamma_1-\gamma_2)^2+(\gamma_3-\gamma_4)^2+1}\right.
$$

$$
+\frac{\gamma_4}{(\gamma_1-\gamma_2)(\gamma_1^2+\gamma_4^2-\gamma_1\gamma_2-\gamma_3\gamma_4)}\left(\frac{1}{n}\sqrt{\ln n}\right)^{\frac{(\gamma_1^2+\gamma_4^2-\gamma_1\gamma_2-\gamma_3\gamma_4)^2}{\gamma_1^2+\gamma_4^2}+\frac{1-\rho_{23}}{1+\rho_{23}}+1}
$$

$$
+\frac{\gamma_4}{(1+\gamma_4^2-\gamma_3\gamma_4)(\gamma_1-\gamma_2)}\left(\frac{1}{n}\sqrt{\ln n}\right)^{\frac{(1+\gamma_4^2-\gamma_3\gamma_4)^2}{1+\gamma_4^2}+(\gamma_1-\gamma_2)^2+\frac{\gamma_3^2}{1+\gamma_4^2}}
$$

$$
+\frac{\gamma_2\gamma_4}{(1+\gamma_4^2-\gamma_3\gamma_4)(1+\gamma_4^2+\gamma_2^2-\gamma_3\gamma_4-\gamma_1\gamma_2)}
$$

$$
\left.\times\left(\frac{1}{n}\sqrt{\ln n}\right)^{\frac{\gamma_3^2}{1+\gamma_4^2}+\frac{(1+\gamma_4^2+\gamma_2^2-\gamma_3\gamma_4-\gamma_1\gamma_2)^2}{1+\gamma_4^2+\gamma_2^2}+\frac{(\gamma_1+\gamma_1\gamma_4^2-\gamma_2\gamma_3\gamma_4)^2}{(1+\gamma_4^2)(1+\gamma_4^2+\gamma_2^2)}}\right\} \tag{4.42}
$$

$$
\leq C(1-\rho_{12})^{-1}(1-\rho_{23})^{-1}(1+\rho_{23}-\rho_{12}-\rho_{13})^{-1}
$$

$$\times(h_3(\rho_{12},\rho_{13},\rho_{23}))^{-1}\left(\frac{1}{n}\right)^{h_1(\rho_{12},\rho_{13},\rho_{23})}(\ln n)^{h_2(\rho_{12},\rho_{13},\rho_{23})}. \tag{4.43}$$

If $\gamma_3-\gamma_4\leq 0$, i.e. $1+\rho_{12}-\rho_{13}-\rho_{23}\leq 0$, the first term in (4.42) is replaced by

$$C\frac{1}{\gamma_1-\gamma_2}\left(\frac{1}{n}\sqrt{\ln n}\right)^{(\gamma_1-\gamma_2)^2+1}\left(\sqrt{\ln n}\right)^{-2},$$

and

$$\int_a^\infty(1-\Phi(\gamma_1 a-\gamma_2 x))\,(1-\Phi(\gamma_3 a-\gamma_4 x))\,\varphi(x)\,dx$$
$$\leq C(1-\rho_{12})^{-1}(1-\rho_{23})^{-1}(1+\rho_{23}-\rho_{12}-\rho_{13})^{-1}$$
$$\times(h_3(\rho_{12},\rho_{13},\rho_{23}))^{-1}\left(\frac{1}{n}\right)^{2/(1+\rho_{12})}(\ln n)^{-\rho_{12}/(1+\rho_{12})}. \tag{4.44}$$

From (4.43) and (4.44), the statement follows.

□

Lemma 4.9 *With the same assumptions as in Lemma 4.8,*

$$r\varphi(\gamma_9 a)\int_a^\infty(1-\Phi(\gamma_1 a-\gamma_2 x))(1-\Phi(\gamma_6 a-\gamma_5 x))\varphi(\gamma_7 x-\gamma_8 a)\,dx$$
$$\leq C(1+\rho_{13}-\rho_{23}-\rho_{12})^{-1}(1+\rho_{23}-\rho_{12}-\rho_{13})^{-1}(1-\rho_{23})^{-1}(1-\rho_{13})^{-1}$$
$$\times(1-\rho_{12})^{-3/2}(h_3(\rho_{12},\rho_{13},\rho_{23}))^{-1}\left(\frac{1}{n}\right)^{h_1(\rho_{12},\rho_{13},\rho_{23})}(\ln n)^{h_2(\rho_{12},\rho_{13},\rho_{23})}.$$

Proof. The proof is similar to the proof of the previous lemma, although the calculations become more involved, partly because the statement is more complicated than that in Lemma 4.8, partly because the calculations in the previous lemma were simplified by relation (4.37). Integrating by parts,

$$r\varphi(\gamma_9 a)\int_a^\infty(1-\Phi(\gamma_1 a-\gamma_2 x))(1-\Phi(\gamma_6 a-\gamma_5 x))\varphi(\gamma_7 x-\gamma_8 a)\,dx$$
$$=\frac{r}{\gamma_7}\varphi(\gamma_9 a)(1-\Phi(a(\gamma_1-\gamma_2)))(1-\Phi(a(\gamma_6-\gamma_5)))(1-\Phi(a(\gamma_7-\gamma_8)))$$

$$+r\varphi(\gamma_9 a)\int_a^\infty \frac{\gamma_2}{\gamma_7}(1-\Phi(\gamma_6 a-\gamma_5 x))(1-\Phi(\gamma_7 x-\gamma_8 a))\varphi(\gamma_1 a-\gamma_2 x)\,dx$$
$$+r\varphi(\gamma_9 a)\int_a^\infty \frac{\gamma_5}{\gamma_7}(1-\Phi(\gamma_1 a-\gamma_2 x))(1-\Phi(\gamma_7 x-\gamma_8 a))\varphi(\gamma_5 x-\gamma_6 a)\,dx$$
$$\le \frac{r}{\gamma_7}\varphi(\gamma_9 a)(1-\Phi(a(\gamma_1-\gamma_2)))(1-\Phi(a(\gamma_6-\gamma_5)))(1-\Phi(a(\gamma_7-\gamma_8)))$$
$$+\frac{r\gamma_2}{\gamma_7}\varphi(\gamma_9 a)\frac{1-\Phi(a(\gamma_7-\gamma_8))}{\varphi(a(\gamma_7-\gamma_8))}\int_a^\infty \varphi(\gamma_7 x-\gamma_8 a)(1-\Phi(\gamma_6 a-\gamma_5 x))$$
$$\times\varphi(\gamma_1 a-\gamma_2 x)\,dx$$
$$+\frac{r\gamma_5}{\gamma_7}\varphi(\gamma_9 a)\frac{1-\Phi(a(\gamma_7-\gamma_8))}{\varphi(a(\gamma_7-\gamma_8))}\int_a^\infty \varphi(\gamma_7 x-\gamma_8 a)(1-\Phi(\gamma_1 a-\gamma_2 x))$$
$$\times\varphi(\gamma_5 x-\gamma_6 a)\,dx$$
$$= \frac{r}{\gamma_7}\varphi(\gamma_9 a)(1-\Phi(a(\gamma_1-\gamma_2)))(1-\Phi(a(\gamma_6-\gamma_5)))(1-\Phi(a(\gamma_7-\gamma_8))) \tag{4.45}$$
$$+\frac{r\gamma_2}{\gamma_7}\varphi(\gamma_9 a)\frac{1-\Phi(a(\gamma_7-\gamma_8))}{\varphi(a(\gamma_7-\gamma_8))}\varphi\left(a\frac{\gamma_1\gamma_7-\gamma_2\gamma_8}{\sqrt{\gamma_2^2+\gamma_7^2}}\right)$$
$$\times\int_a^\infty(1-\Phi(\gamma_6 a-\gamma_5 x))\varphi\left(\frac{(\gamma_2^2+\gamma_7^2)x-(\gamma_1\gamma_2+\gamma_7\gamma_8)a}{\sqrt{\gamma_2^2+\gamma_7^2}}\right)\,dx \tag{4.46}$$
$$+\frac{r\gamma_2}{\gamma_7}\varphi(\gamma_9 a)\frac{1-\Phi(a(\gamma_7-\gamma_8))}{\varphi(a(\gamma_7-\gamma_8))}\varphi\left(a\frac{\gamma_6\gamma_7-\gamma_5\gamma_8}{\sqrt{\gamma_5^2+\gamma_7^2}}\right)$$
$$\times\int_a^\infty(1-\Phi(\gamma_1 a-\gamma_2 x))\varphi\left(\frac{(\gamma_5^2+\gamma_7^2)x-(\gamma_6\gamma_5+\gamma_7\gamma_8)a}{\sqrt{\gamma_5^2+\gamma_7^2}}\right)\,dx. \tag{4.47}$$

From Lemma 4.6 (iii), (v) and (vii), we have $\sqrt{\gamma_2^2+\gamma_7^2}>0$, $\sqrt{\gamma_5^2+\gamma_7^2}>0$, $\gamma_2^2+\gamma_7^2>\gamma_1\gamma_2+\gamma_7\gamma_8$ and $\gamma_5^2+\gamma_7^2>\gamma_6\gamma_5+\gamma_7\gamma_8$. Thus the prerequisites of Lemma 4.7 are fulfilled for both (4.46) and (4.47). Hence

$$r\varphi(\gamma_9 a)\int_a^\infty (1-\Phi(\gamma_1 a-\gamma_2 x))(1-\Phi(\gamma_6 a-\gamma_5 x))\varphi(\gamma_7 x-\gamma_8 a)\,dx$$

$$\leq \frac{r}{\gamma_7}\varphi(a\gamma_9)(1-\Phi(a(\gamma_7-\gamma_8)))\Big\{(1-\Phi(a(\gamma_1-\gamma_2)))(1-\Phi(a(\gamma_6-\gamma_5)))$$

$$+\frac{\gamma_2}{\sqrt{\gamma_2^2+\gamma_7^2}}\big(\varphi(a(\gamma_7-\gamma_8))\big)^{-1}\varphi\left(a\frac{\gamma_1\gamma_7-\gamma_2\gamma_8}{\sqrt{\gamma_2^2+\gamma_7^2}}\right)$$

$$\times\left(1-\Phi\left(a\frac{\gamma_2^2+\gamma_7^2-\gamma_1\gamma_2-\gamma_7\gamma_8}{\sqrt{\gamma_2^2+\gamma_7^2}}\right)\right)\Big[(1-\Phi(a(\gamma_6-\gamma_5)))$$

$$+\frac{\gamma_5}{\sqrt{\gamma_5^2+\gamma_2^2+\gamma_7^2}}\left(\varphi\left(a\frac{\gamma_2^2+\gamma_7^2-\gamma_1\gamma_2-\gamma_7\gamma_8}{\sqrt{\gamma_2^2+\gamma_7^2}}\right)\right)^{-1}$$

$$\times\varphi\left(a\frac{\gamma_6\gamma_2^2+\gamma_6\gamma_7^2-\gamma_5\gamma_1\gamma_2-\gamma_5\gamma_7\gamma_8}{\sqrt{\gamma_2^2+\gamma_7^2}\sqrt{\gamma_5^2+\gamma_2^2+\gamma_7^2}}\right)$$

$$\times\left(1-\Phi\left(a\frac{\gamma_5^2+\gamma_2^2+\gamma_7^2-\gamma_6\gamma_5-\gamma_1\gamma_2-\gamma_7\gamma_8}{\sqrt{\gamma_5^2+\gamma_2^2+\gamma_7^2}}\right)\right)\Big]$$

$$+\frac{\gamma_2}{\sqrt{\gamma_5^2+\gamma_7^2}}\big(\varphi(a(\gamma_7-\gamma_8))\big)^{-1}\varphi\left(a\frac{\gamma_6\gamma_7-\gamma_5\gamma_8}{\sqrt{\gamma_5^2+\gamma_7^2}}\right)$$

$$\times\left(1-\Phi\left(a\frac{\gamma_5^2+\gamma_7^2-\gamma_6\gamma_5-\gamma_7\gamma_8}{\sqrt{\gamma_5^2+\gamma_7^2}}\right)\right)\Big[(1-\Phi(a(\gamma_1-\gamma_2)))$$

$$+\frac{\gamma_2}{\sqrt{\gamma_2^2+\gamma_5^2+\gamma_7^2}}\left(\varphi\left(a\frac{\gamma_5^2+\gamma_7^2-\gamma_6\gamma_5-\gamma_7\gamma_8}{\sqrt{\gamma_5^2+\gamma_7^2}}\right)\right)^{-1}$$

$$\times\varphi\left(a\frac{\gamma_1\gamma_5^2+\gamma_1\gamma_7^2-\gamma_2\gamma_6\gamma_5-\gamma_2\gamma_7\gamma_8}{\sqrt{\gamma_5^2+\gamma_7^2}\sqrt{\gamma_2^2+\gamma_5^2+\gamma_7^2}}\right)$$

$$\times\left(1-\Phi\left(a\frac{\gamma_2^2+\gamma_5^2+\gamma_7^2-\gamma_1\gamma_2-\gamma_6\gamma_5-\gamma_7\gamma_8}{\sqrt{\gamma_2^2+\gamma_5^2+\gamma_7^2}}\right)\right)\Bigg]\Bigg\}.$$

As in the previous lemma, we use (4.17) and (4.18). From Lemma 4.6 (iv)–(viii), the condition $c_1 > 0$ for (4.17) is satisfied for all terms. Thus

$$r\varphi(\gamma_9 a)\int_a^\infty(1-\Phi(\gamma_1 a-\gamma_2 x))(1-\Phi(\gamma_6 a-\gamma_5 x))\varphi(\gamma_7 x-\gamma_8 a)\,dx$$

$$\leq C\frac{1}{\gamma_7-\gamma_8}\left(\frac{\sqrt{\ln n}}{n}\right)^{\gamma_9^2+(\gamma_7-\gamma_8)^2}\left(\sqrt{\ln n}\right)^{-3}$$

$$\times\Bigg\{\frac{1}{(\gamma_1-\gamma_2)(\gamma_6-\gamma_5)}\left(\frac{\sqrt{\ln n}}{n}\right)^{(\gamma_1-\gamma_2)^2+(\gamma_6-\gamma_5)^2}$$

$$+\frac{\gamma_2}{\gamma_2^2+\gamma_7^2-\gamma_1\gamma_2-\gamma_7\gamma_8}\left(\frac{\sqrt{\ln n}}{n}\right)^{-(\gamma_7-\gamma_8)^2+\frac{(\gamma_1\gamma_7-\gamma_2\gamma_8)^2}{\gamma_2^2+\gamma_7^2}+\frac{(\gamma_2^2+\gamma_7^2-\gamma_1\gamma_2-\gamma_7\gamma_8)^2}{\gamma_2^2+\gamma_7^2}}$$

$$\times\Bigg[\frac{1}{\gamma_6-\gamma_5}\left(\frac{\sqrt{\ln n}}{n}\right)^{(\gamma_6-\gamma_5)^2}+\frac{\gamma_5}{\gamma_5^2+\gamma_2^2+\gamma_7^2-\gamma_6\gamma_5-\gamma_1\gamma_2-\gamma_7\gamma_8}$$

$$\times\left(\frac{\sqrt{\ln n}}{n}\right)^{\frac{(\gamma_6\gamma_2^2+\gamma_6\gamma_7^2-\gamma_5\gamma_1\gamma_2-\gamma_5\gamma_7\gamma_8)^2}{(\gamma_2^2+\gamma_7^2)(\gamma_5^2+\gamma_2^2+\gamma_7^2)}-\frac{(\gamma_2^2+\gamma_7^2-\gamma_1\gamma_2-\gamma_7\gamma_8)^2}{\gamma_2^2+\gamma_7^2}+\frac{(\gamma_5^2+\gamma_2^2+\gamma_7^2-\gamma_6\gamma_5-\gamma_1\gamma_2-\gamma_7\gamma_8)^2}{\gamma_5^2+\gamma_2^2+\gamma_7^2}}\Bigg]$$

$$+\frac{\gamma_2}{\gamma_5^2+\gamma_7^2-\gamma_6\gamma_5-\gamma_7\gamma_8}\left(\frac{\sqrt{\ln n}}{n}\right)^{\frac{(\gamma_6\gamma_7-\gamma_5\gamma_8)^2}{\gamma_5^2+\gamma_7^2}-(\gamma_7-\gamma_8)^2+\frac{(\gamma_5^2+\gamma_7^2-\gamma_6\gamma_5-\gamma_7\gamma_8)^2}{\gamma_5^2+\gamma_7^2}}$$

$$\times\Bigg[\frac{1}{\gamma_1-\gamma_2}\left(\frac{\sqrt{\ln n}}{n}\right)^{(\gamma_1-\gamma_2)^2}+\frac{\gamma_2}{\gamma_2^2+\gamma_5^2+\gamma_7^2-\gamma_1\gamma_2-\gamma_6\gamma_5-\gamma_7\gamma_8}$$

$$\times\left(\frac{\sqrt{\ln n}}{n}\right)^{\frac{(\gamma_1\gamma_5^2+\gamma_1\gamma_7^2-\gamma_2\gamma_6\gamma_5-\gamma_2\gamma_7\gamma_8)^2}{(\gamma_5^2+\gamma_7^2)(\gamma_2^2+\gamma_5^2+\gamma_7^2)}-\frac{(\gamma_5^2+\gamma_7^2-\gamma_6\gamma_5-\gamma_7\gamma_8)^2}{\gamma_5^2+\gamma_7^2}+\frac{(\gamma_2^2+\gamma_5^2+\gamma_7^2-\gamma_1\gamma_2-\gamma_6\gamma_5-\gamma_7\gamma_8)^2}{\gamma_2^2+\gamma_5^2+\gamma_7^2}}\Bigg]\Bigg\}$$

$$\leq C(1+\rho_{13}-\rho_{23}-\rho_{12})^{-1}(1+\rho_{23}-\rho_{12}-\rho_{13})^{-1}(1-\rho_{23})^{-1}(1-\rho_{13})^{-1}$$

$$(1-\rho_{12})^{-3/2}\left(h_3(\rho_{12},\rho_{13},\rho_{23})\right)^{-1}\left(\frac{1}{n}\right)^{h_1(\rho_{12},\rho_{13},\rho_{23})}(\ln n)^{h_2(\rho_{12},\rho_{13},\rho_{23})}.$$

□

We are now ready to state the promised counterpart of Lemma 4.5 for three variables.

Proposition 4.10 *Let (ξ_1, ξ_2, ξ_3) be a Gaussian random vector with mean zero and non-singular covariance matrix (4.23). Assume, without loss of generality, that $\rho_{12} \leq \rho_{13} \leq \rho_{23}$. Define h_1, h_2 and h_3 by (4.34)-(4.36). If $\rho_{12} \geq 0$, then for any positive a related to n by $n(1-\Phi(a)) = \lambda_0$ for some constant λ_0, and $n \geq 2$,*

$$\begin{aligned} P(\xi_1 > a \cap \xi_2 > a \cap \xi_3 > a) &\leq C(1+\rho_{13}-\rho_{23}-\rho_{12})^{-1} \\ &\times(1+\rho_{23}-\rho_{12}-\rho_{13})^{-1}(1-\rho_{23})^{-1}(1-\rho_{13})^{-1}(1-\rho_{12})^{-3/2} \\ &\times (h_3(\rho_{12},\rho_{13},\rho_{23}))^{-1}\left(\frac{1}{n}\right)^{g_1(\rho_{12},\rho_{13},\rho_{23})}(\ln n)^{g_2(\rho_{12},\rho_{13},\rho_{23})}, \end{aligned}$$

where

$$g_1(\rho_{12},\rho_{13},\rho_{23}) = \begin{cases} h_1(\rho_{12},\rho_{13},\rho_{23}) & \text{if } 1+\rho_{12}-\rho_{13}-\rho_{23} > 0 \\ 2/(1+\rho_{12}) & \text{otherwise,} \end{cases}$$

$$g_2(\rho_{12},\rho_{13},\rho_{23}) = \begin{cases} h_2(\rho_{12},\rho_{13},\rho_{23}) & \text{if } 1+\rho_{12}-\rho_{13}-\rho_{23} > 0 \\ -\rho_{12}/(1+\rho_{12}) & \text{otherwise,} \end{cases}$$

and C is a constant depending on λ_0 only.

Proof. Let r be defined by (4.24). Since

$$(\xi_2, \xi_3 | \xi_1 = x) \sim \mathrm{N}\left(\begin{pmatrix} \rho_{12}x \\ \rho_{13}x \end{pmatrix}, \begin{pmatrix} 1-\rho_{12}^2 & \rho_{23}-\rho_{12}\rho_{13} \\ \rho_{23}-\rho_{12}\rho_{13} & 1-\rho_{13}^2 \end{pmatrix}\right),$$

see e.g. Cramér [7] page 306, we get

$$\begin{aligned} &P(\xi_1 > a \cap \xi_2 > a \cap \xi_3 > a) \\ &\qquad = \int_a^\infty P(\xi_2 > a \cap \xi_3 > a | \xi_1 = x)\varphi(x)\,dx \\ &\qquad = \int_a^\infty P\left(\eta_1 > \frac{a-\rho_{12}x}{\sqrt{1-\rho_{12}^2}} \cap \eta_2 > \frac{a-\rho_{13}x}{\sqrt{1-\rho_{13}^2}}\right)\varphi(x)\,dx, \end{aligned}$$

where

$$(\eta_1, \eta_2) \sim \mathrm{N}\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & r \\ r & 1 \end{array}\right)\right).$$

Note that $0 \leq r < 1$ because of the assumption of a non-singular covariance matrix. By Lemma 4.5,

$$P(\xi_1 > a \cap \xi_2 > a \cap \xi_3 > a) \leq \int_a^\infty f(x)\,dx + \int_a^\infty g(x)\,dx,$$

where

$$\begin{aligned} f(x) = & \left(1 - \Phi\left(\frac{a}{\sqrt{1-\rho_{12}^2}} - \frac{\rho_{12}x}{\sqrt{1-\rho_{12}^2}}\right)\right)\left(1 - \Phi\left(\frac{1}{\sqrt{1-r^2}}\right.\right. \\ & \left.\left.\times\left(\frac{a}{\sqrt{1-\rho_{13}^2}} - \frac{ra}{\sqrt{1-\rho_{12}^2}} - \frac{\rho_{13}x}{\sqrt{1-\rho_{13}^2}} + \frac{r\rho_{12}x}{\sqrt{1-\rho_{12}^2}}\right)\right)\right)\varphi(x) \\ & = (1 - \Phi(\gamma_1 a - \gamma_2 x))\,(1 - \Phi(\gamma_3 a - \gamma_4 x))\,\varphi(x) \end{aligned}$$

and

$$\begin{aligned} g(x) = & \left(1 - \Phi\left(\frac{a}{\sqrt{1-\rho_{12}^2}} - \frac{\rho_{12}x}{\sqrt{1-\rho_{12}^2}}\right)\right)\left(1 - \Phi\left(\frac{1}{\sqrt{1-r^2}}\right.\right. \\ & \left.\left.\times\left(\frac{a}{\sqrt{1-\rho_{12}^2}} - \frac{ra}{\sqrt{1-\rho_{13}^2}} - \frac{\rho_{12}x}{\sqrt{1-\rho_{12}^2}} + \frac{r\rho_{13}x}{\sqrt{1-\rho_{13}^2}}\right)\right)\right) \\ & \times r\varphi\left(\frac{a}{\sqrt{1-\rho_{13}^2}} - \frac{\rho_{13}x}{\sqrt{1-\rho_{13}^2}}\right)\left(\varphi\left(\frac{a}{\sqrt{1-\rho_{12}^2}} - \frac{\rho_{12}x}{\sqrt{1-\rho_{12}^2}}\right)\right)^{-1}\varphi(x) \\ & = r\varphi(\gamma_9 a)(1 - \Phi(\gamma_1 a - \gamma_2 x))(1 - \Phi(\gamma_6 a - \gamma_5 x))\varphi(\gamma_7 x - \gamma_8 a) \end{aligned}$$

The statement now follows from Lemma 4.8 and Lemma 4.9.

□

Corollary 4.11 *If* $\rho_{12} = \rho_{13} = \rho_{23} = \rho \geq 0$*, then*

$$P(\xi_1 > a \cap \xi_2 > a \cap \xi_3 > a) \leq C\,(1-\rho)^{-15/2}\left(\frac{1}{n}\right)^{3/(1+2\rho)}(\ln n)^{-3/(1+2\rho)}.$$

Corollary 4.12 *Let (ξ_1, ξ_2, ξ_3) be a Gaussian random vector with mean zero and non-singular covariance matrix (4.23). Assume, without loss of generality, that $\rho_{12} \leq \rho_{13} \leq \rho_{23}$. If $\rho_{12} < 0$, then for any positive a related to n by $n(1-\Phi(a)) = \lambda_0$ for some constant λ_0, and $n \geq 2$,*

$$P(\xi_1 > a \cap \xi_2 > a \cap \xi_3 > a) \leq C(1-\rho_{23})^{-4} \left(\frac{1}{n}\right)^2.$$

Proof. Replace the negative correlation(s) by zero. By Slepian's inequality (see Tong [26] page 10) this can only increase the probability in the statement. Increase (again supported by Slepian's inequality) ρ_{13} and ρ_{23} so that $1 - \rho_{13} - \rho_{23} \leq 0$. The result now follows from Proposition 4.10 with $\rho_{12} = 0$.

□

As an aside, we note that it is not necessary to increase ρ_{13} and ρ_{23} in the proof above, but the functions g_1 and g_2 in Proposition 4.10 are simpler when $1 + \rho_{12} - \rho_{13} - \rho_{23} \leq 0$, and the result still holds.

Recall from equations (4.14), (4.15) and (4.22) the computation of λ_3,

$$\begin{aligned} \lambda_3 &\leq \frac{1}{3} \sum_{i \in \Gamma} \sum_{\{\alpha_j\}} P(Y_i > a, Y_{\alpha_1} > a, Y_{\alpha_2} > a, Y_{\alpha_3} < a, \ldots, Y_{\alpha_{2q}} < a) \\ &\leq \frac{1}{3} \sum_{i \in \Gamma} \sum_{\alpha_1, \alpha_2} P(Y_i > a, Y_{\alpha_1} > a, Y_{\alpha_2} > a) \end{aligned}$$

Moreover, by redefining $\alpha_1, \alpha_2, \alpha_3$ as any combination of length three of Γ,

$$\lambda_3 \leq Cn \max_{\alpha_1, \alpha_2, \alpha_3} P(Y_{\alpha_1} > a, Y_{\alpha_2} > a, Y_{\alpha_3} > a). \tag{4.48}$$

Consequently, by Proposition 4.10,

$$\begin{aligned} &\lambda_3 \leq Cn(1 + \rho_{13} - \rho_{23} - \rho_{12})^{-1}(1 + \rho_{23} - \rho_{12} - \rho_{13})^{-1} \\ &\times (1-\rho_{23})^{-1}(1-\rho_{13})^{-1}(1-\rho_{12})^{-3/2} \left(h_3(\rho_{12}, \rho_{13}, \rho_{23})\right)^{-1} \\ &\times \left(\frac{1}{n}\right)^{g_1(\rho_{12}^*, \rho_{13}^*, \rho_{23}^*)} (\ln n)^{g_2(\rho_{12}^*, \rho_{13}^*, \rho_{23}^*)}, \end{aligned}$$

where g_1 and g_2 are as in Proposition 4.10 and ρ_{12}^*, ρ_{13}^* and ρ_{23}^* are the correlations of $Y_{\alpha_1}, Y_{\alpha_2}$ and Y_{α_3} respectively which maximize (4.48).

The combination $\{\alpha_i\}$ which actually maximizes the function in equation (4.48) depends of course on the underlying correlation function $\{\rho_k\}$. This will be discussed further in the next subsection.

We summarize the results of this subsection in Proposition 4.13, where we also show that the bound valid for λ_3 is also valid for λ_k, $k \geq 4$. In this proposition, one can get the impression that variables with negative correlations can limit the convergence rate. However, this is not the case. It is important to bear in mind that the best rate we can achieve is $O(1/n)$ (cf. Theorem 4.4). Hence the term in Proposition 4.13 derived from negatively correlated variables does not restrict the rate of convergence.

Proposition 4.13 *Assume $\{Y_k\}$ is a stationary Gaussian sequence with correlation function $\{\rho_k\}$ satisfying (4.11). Let D and λ_k, $k = 1, 2, \ldots$, be defined by (4.3) and (4.14) respectively, and let g_2 be as in Proposition 4.10. Then*

$$\sum_{j=3}^{D+1} j(j-1)\lambda_j = \begin{cases} O\left(\dfrac{1}{n^\nu}(\ln n)^\mu\right) & \textit{if } \min(\rho_1, \rho_2, \ldots) \geq 0, \\[2ex] O\left(\dfrac{1}{n} + \dfrac{1}{n^\nu}(\ln n)^\mu\right) & \textit{otherwise,} \end{cases}$$

as $n \to \infty$, where

$$\nu = \min_{\rho_{12}, \rho_{13}, \rho_{23}} f(\rho_{12}, \rho_{13}, \rho_{23})$$

for

$$f(\rho_{12}, \rho_{13}, \rho_{23}) =$$

$$= \begin{cases} 2\left(1 - \rho_{12} - \rho_{13} - \rho_{23} + \rho_{12}\rho_{13} + \rho_{12}\rho_{23} + \rho_{13}\rho_{23} - \rho_{12}\rho_{13}\rho_{23}\right) \\ \quad \times \left(1 + 2\rho_{12}\rho_{13}\rho_{23} - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2\right)^{-1} \\ & \textit{if } 1 + \rho_{12} - \rho_{13} - \rho_{23} > 0, \\ (1 - \rho_{12}/(1 + \rho_{12}) & \textit{otherwise,} \end{cases}$$

and

$$\mu = g_2(\rho_{12}^*, \rho_{13}^*, \rho_{23}^*). \tag{4.49}$$

The minimization is carried out for the sets of correlations $\rho_{12} \leq \rho_{13} \leq \rho_{23}$ corresponding to the different ways of choosing three of the n points in Γ satisfying $\rho_{12} \geq 0$, i.e., only non-negative correlations. The asterisks in (4.49) indicate the values of the variables that minimize ν.

Proof. We first prove, by contradiction, that the distribution of $(Y_1, Y_2, \ldots, Y_l)$ is non-singular for arbitrary l. Suppose the opposite, i.e. suppose that the distribution is singular. Then at least one of the components, assumed without loss of generality to be the last, is determined as a linear combination of the others. Thus Y_l is determined by $(Y_1, Y_2, \ldots, Y_{l-1})$. By stationarity, Y_{l+1} is, by the same reason, determined by $(Y_2, Y_2, \ldots, Y_l)$, and hence Y_{l+1} is also determined by $(Y_1, Y_2, \ldots, Y_{l-1})$. By induction, an arbitrary number of values are determined by $(Y_1, Y_2, \ldots, Y_{l-1})$, contradicting the assumption of only a finite number of non-zero elements of $\{\rho_k\}$. Hence all finite-dimensional distributions of $\{Y_k\}$ are non-singular.

If all correlations are positive, inequality (4.48) and Proposition 4.10 yields

$$\lambda_3 = O\left(\frac{1}{n^\nu}(\ln n)^\mu\right). \tag{4.50}$$

If at least one of the correlations is negative, the term $n \cdot 1/n^2 = 1/n$ from Corollary 4.12 must be included. Hence

$$\lambda_3 = O\left(\frac{1}{n} + \frac{1}{n^\nu}(\ln n)^\mu\right). \tag{4.51}$$

Finally, for $k \geq 4$,

$$\begin{aligned}\lambda_k &\leq \frac{1}{k}\sum_{i\in\Gamma}\sum P(Y_i > a, Y_{\alpha_1} > a, \ldots, Y_{\alpha_{k-1}} > a, Y_{\alpha_k} < a, \ldots, Y_{\alpha_{2q}} < a) \\ &\leq \frac{1}{k}\sum_{i\in\Gamma}\sum P(Y_{\alpha_1} > a, Y_{\alpha_2} > a, Y_{\alpha_3} > a).\end{aligned}$$

Thus λ_k, $k \geq 4$ is also bounded by the right hand side of equation (4.50) or (4.51). Hence the result follows since the sum in the statement is finite.

□

4.3.2 Finding the worst combination of correlations

The bound in Proposition 4.13 is not explicit in the sense that there is still a minimization that must be carried out. Since there is a finite number of possibilities, the minimization can be performed by evaluating the function for each case one by one, at least in principle. The solution does, of course, in general depend on the correlation function $\{\rho_k\}$. However, we can solve the problem explicitly in the important special case of an MA(q)-sequence with positive and decreasing coefficients.

Another approach used in the latter part of this subsection is to choose $\rho_{12} = \rho_{13} = \rho_{23} = \max(\rho_1, \rho_2, \ldots, \rho_q)$ in Proposition 4.13 and combine with Slepian's inequality. This gives an overestimation of the maximum value in (4.48) but is still useful.

Proposition 4.14 *Suppose $\{Y_k\}$ is a MA(q)-sequence with positive decreasing coefficients and correlation function $\{\rho_k\}$. Then*

$$\sum_{j=3}^{D+1} j(j-1)\lambda_j = O\left(\frac{1}{n^\nu}(\ln n)^\mu\right),$$

as $n \to \infty$, where

$$\nu = \begin{cases} 2(1-\rho_1)^2/(1+\rho_2-2\rho_1^2) & \text{if } 1+\rho_2 > 2\rho_1 \\ (1-\rho_2)/(1+\rho_2) & \text{otherwise} \end{cases}$$

and

$$\mu = \begin{cases} -(\rho_2+2\rho_1-3\rho_1^2)/(1+\rho_2-2\rho_1^2) & \text{if } 1+\rho_2 > 2\rho_1 \\ -\rho_2/(1+\rho_2) & \text{otherwise.} \end{cases}$$

Proof. Let $\{Y_k\}$ have the representation

$$Y_k = \theta_0\varepsilon_k + \theta_1\varepsilon_{k+1} + \ldots + \theta_q\varepsilon_{k+q},$$

$$\theta_i \geq 0, \qquad \theta_i \geq \theta_{i+1} \text{ for all } i. \tag{4.52}$$

From this it is a simple matter to calculate the covariance function,

$$\mathrm{Cov}(Y_k, Y_{k+j}) = \sum_{i=0}^{q-j} \theta_i \theta_{i+j}.$$

Consequently

$$\begin{aligned}
\rho_j - \rho_{j+1} &= \frac{1}{\rho_0}(\mathrm{Cov}(Y_k, Y_{k+j}) - \mathrm{Cov}(Y_k, Y_{k+j+1})) \\
&= \frac{1}{\rho_0}\left(\sum_{i=0}^{q-j} \theta_i \theta_{i+j} - \sum_{i=0}^{q-(j+1)} \theta_i \theta_{i+(j+1)}\right) \\
&= \frac{1}{\rho_0}\left(\theta_{q-j}\theta_q + \sum_{i=0}^{q-(j+1)} \theta_i \left(\theta_{i+j+1} - \theta_{i+j}\right)\right) \geq 0,
\end{aligned}$$

by (4.52). Hence it follows that $\{\rho_k\}$ is a positive decreasing sequence. Now choose (ξ_1, ξ_2, ξ_3) in Proposition 4.10 as consecutive variables from $\{Y_k\}$. Then the correlations will be $\rho_{12} = \rho_1$, $\rho_{13} = \rho_1$ and $\rho_{23} = \rho_2$. Since the correlation function is decreasing, the only triplet with larger values is one with all three values equal to ρ_1, but it is not possible to choose any variables (ξ_1, ξ_2, ξ_3) from $\{Y_k\}$ with this correlation structure. Hence the largest possible combination of correlations is the one chosen. By Slepian's inequality, this choice will lead to the maximum value of (4.48). The rest follows from Proposition 4.13 since, after some computation, $g_1(\rho_2, \rho_1, \rho_1)$ and $g_2(\rho_2, \rho_1, \rho_1)$ simplify to the expressions above for ν and μ.

□

The assumption of positive decreasing coefficients was used only to conclude that choosing consecutive variables yields the largest correlations. Hence it can be replaced by the weaker condition,

$$\rho_1 \geq \rho_2 \geq \rho_k, \qquad k = 3, 4, \ldots,$$

with no condition on the relation between ρ_k and ρ_l for $k, l \notin \{1, 2\}$.

We now aim at a general lower bound for the minimum value ν in Proposition 4.13 without posing any restrictions on the relative order of the terms in the correlation function.

Proposition 4.15 *Let $\{Y_k\}$ be a stationary Gaussian sequence with correlation function $\{\rho_k\}$ where $\rho_k = 0$, $k > q$. Let $\rho = \max(\rho_1, \rho_2, \ldots, \rho_q) \geq 0$. Then*

$$\sum_{j=3}^{D+1} j(j-1)\lambda_j = O\left(\left(\frac{1}{n}\right)^{2(1-\rho)/(1+2\rho)} (\ln n)^{-3\rho/(1+2\rho)}\right)$$

as $n \to \infty$.

Proof. As in Proposition 4.13, λ_3 is bounded by (4.48). Now replace the random variables in (4.48) by new ones, with all three correlations equal to ρ. By Slepian's inequality, the right hand side of (4.48) will then increase. The statement now follows from Corollary 4.11 and Proposition 4.13 (with ρ_{12}, ρ_{13} and ρ_{23} replaced by ρ).

□

As noted in the proof of Proposition 4.14, it is not possible to choose three random variables from $\{Y_k\}$ with all three correlations equal to ρ, where $\rho = \max(\rho_1, \rho_2, \ldots, \rho_q)$, unless the maximum value ρ is not unique. It is therefore possible to replace the exponents ν and μ in Proposition 4.15 by

$$\nu = \begin{cases} 2(1-\rho)^2/(1+\rho'-2\rho^2) & \text{if } 1+\rho' > 2\rho \\ (1-\rho')/(1+\rho') & \text{otherwise} \end{cases}$$

and

$$\mu = \begin{cases} -(\rho'+2\rho-3\rho^2)/(1+\rho'-2\rho^2) & \text{if } 1+\rho' > 2\rho \\ -\rho'/(1+\rho') & \text{otherwise,} \end{cases}$$

where ρ' is the second largest correlation (possibly equal to ρ) of $\{\rho_k\}$ (cf. Proposition 4.14). However, if ρ and ρ' are close, the gain is small.

4.3.3 Asymptotic bounds when omitting all but the first two terms

In Section 4.2, we showed that without omitting any terms in the approximating compound Poisson distribution, the convergence rate is

$O(1/n)$. In this subsection, we use the results in Subsections 4.3.1 and 4.3.2 to state theorems for the convergence rate when only the first two terms in the approximating compound Poisson distribution are used. Comparisons of the rates are made, not only for the two cases of compound Poisson approximation, but also for ordinary Poisson approximation.

Theorem 4.16 *Suppose $\{Y_k\}$ is a stationary Gaussian sequence with correlation function $\{\rho_k\}$ with $\rho_k = 0$, $k > q$. Assume further that $\{a_n\}$ is a real positive sequence satisfying $n(1-\Phi(a_n)) = \lambda_0$, where $\lambda_0 = E(W)$ is constant, and define $\boldsymbol{\lambda}^*$ by equation (4.10) for $l = 2$. Then,*

$$d(W, CPo(\boldsymbol{\lambda}^*)) = O\left(\frac{1}{n^\nu}(\ln n)^\mu\right)$$

as $n \to \infty$, where

$$\nu = \min_{\rho_{12},\rho_{13},\rho_{23}} f(\rho_{12},\rho_{13},\rho_{23})$$

for

$$f(\rho_{12},\rho_{13},\rho_{23}) =$$

$$= \begin{cases} \min\Big(1, 2\,(1-\rho_{12}-\rho_{13}-\rho_{23}+\rho_{12}\rho_{13}+\rho_{12}\rho_{23}+\rho_{13}\rho_{23}-\rho_{12}\rho_{13}\rho_{23}) \\ \quad \times\,(1+2\rho_{12}\rho_{13}\rho_{23}-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2)^{-1}\Big) \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } 1+\rho_{12}-\rho_{13}-\rho_{23} > 0, \\ (1-\rho_{12})/(1+\rho_{12}) \qquad\qquad\qquad \text{otherwise,} \end{cases}$$

and

$$\mu = \begin{cases} 0 & \text{if } \nu = 1 \\ h_2(\rho_{12}^*,\rho_{13}^*,\rho_{23}^*) & \text{if } 1+\rho_{12}^*-\rho_{13}^*-\rho_{23}^* > 0 \\ -\rho_{12}^*/(1+\rho_{12}^*) & \text{otherwise.} \end{cases} \tag{4.53}$$

The minimization is carried out for the sets of correlations $\rho_{12} \le \rho_{13} \le \rho_{23}$ corresponding to the different ways of choosing three of the n points in Γ satisfying $\rho_{12} \ge 0$, i.e. only non-negative correlations. The asterisks in (4.53) indicate the values of the variables that minimize ν.

Note that only positively correlated variables can reduce the speed of convergence from the "ideal" value of one.

Proof. From Theorem 4.2 and Theorem 4.4,

$$d(W, \mathrm{CPo}(\boldsymbol{\lambda}^*)) = O\left(\frac{1}{n} + \sum_{j=3}^{D+1} j(j-1)\lambda_j\right).$$

The result now follows from Proposition 4.13.

□

If also the second term in the approximating compound Poisson distribution is omitted then only a Poisson distributed random variable remains. By Slepian's inequality,

$$\lambda_2 \leq CnP(\xi_1 > a \cap \xi_2 > a),$$

where

$$(\xi_1, \xi_2) \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

and $\rho = \max(0, \rho_1, \rho_2, \ldots, \rho_q)$. Consequently, by Lemma 4.5 with $a = b$, and relation (4.17),

$$\begin{aligned} \lambda_2 &\leq Cn(1 - \Phi(a))\left(1 - \Phi\left(a\sqrt{\frac{1-\rho}{1+\rho}}\right)\right) \\ &\leq C\left(\frac{1}{n}\right)^{(1-\rho)/(1+\rho)} \left(\sqrt{\ln n}\right)^{-1+(1-\rho)/(1+\rho)}, \end{aligned}$$

and

$$d(W, \mathrm{Po}(E(W))) = O\left(\left(\frac{1}{n}\right)^{(1-\rho)/(1+\rho)} (\ln n)^{-\rho/(1+\rho)}\right), \qquad (4.54)$$

a result in accordance with Holst and Janson [9].

In ordinary Poisson approximation, a single large and positive value in the correlation function can reduce the rate of convergence substantially (note the dependence on $\max(0, \rho_1, \rho_2, \ldots, \rho_q)$ only). For the compound Poisson approximation, on the other hand, more than one value must be large for the convergence to be slow. For the two-term compound Poisson approximation, at least three values must be "large" (i.e. making ν in Theorem 4.16 small).

The use of Proposition 4.14 or Proposition 4.15 in the proof of the theorem above, instead of Proposition 4.13, immediately yields the following corollaries.

Corollary 4.17 *Suppose $\{Y_k\}$ is a MA(q)-sequence with positive decreasing coefficients and correlation function $\{\rho_k\}$. Under the assumptions of Theorem 4.16,*

$$d(W, CPo(\boldsymbol{\lambda}^*)) = O\left(\frac{1}{n^\nu}(\ln n)^\mu\right)$$

as $n \to \infty$, where

$$\nu = \begin{cases} \min\Bigl(1, 2(1-\rho_1)^2/(1+\rho_2-2\rho_1^2)\Bigr) & \text{if } 1+\rho_2 > 2\rho_1 \\ (1-\rho_2)/(1+\rho_2) & \text{otherwise} \end{cases}$$

and

$$\mu = \begin{cases} 0 & \text{if } \nu = 1 \\ -(\rho_2+2\rho_1-3\rho_1^2)/(1+\rho_2-2\rho_1^2) & \text{if } \nu < 1 \text{ and } 1+\rho_2 > 2\rho_1 \\ -\rho_2/(1+\rho_2) & \text{otherwise.} \end{cases}$$

Corollary 4.18 *Let $\rho = \max(\rho_1, \rho_2, \ldots, \rho_q)$. Under the assumptions of Theorem 4.16,*

$$d(W, CPo(\boldsymbol{\lambda}^*)) = O\left(\frac{1}{n^\nu}(\ln n)^\mu\right)$$

as $n \to \infty$, where

$$\nu = \begin{cases} 2(1-\rho)/(1+2\rho) & \text{if } \rho > 1/4 \\ 1 & \text{otherwise} \end{cases}$$

and

$$\mu = \begin{cases} -3\rho/(1+2\rho) & \text{if } \rho > 1/4 \\ 0 & \text{otherwise.} \end{cases}$$

To get a better understanding of the convergence rate, let us study Corollary 4.18 in more detail. In Figure 4.2, the exponent of $1/n$ in the convergence rate is plotted. If none of the terms in the approximating compound Poisson distribution is left out then the exponent in the convergence rate is always equal to one, see the dash-dotted line (and a part of the solid line). For weak correlations, clumps of size three or larger are negligible, and the convergence rate is not affected. On the other hand, for stronger correlated variables, the convergence is slowed down in comparison to the original and more complicated approximating distribution.

The point where even the simpler compound Poisson distribution reaches a convergence rate of $1/n$ is, by Corollary 4.18, not less than $\rho = 0.25$. Nevertheless, it is important to remember that the approximation in Subsection 4.3.2, which led to the expression now used, may not be sharp (see e.g. the remark after Proposition 4.15). Therefore, even for sequences with $\rho_{\max} > 0.25$, the convergence rate may still be close to (or equal to) one.

The dashed line in Figure 4.2 shows the exponent ν of $1/n$ of the convergence rate for ordinary Poisson approximation, cf. (4.54) or Holst and Janson [9]. Compared to the general compound Poisson approximation (the dash-dotted line), the increase in speed of convergence is substantial and in favour of the compound Poisson approximation. The greatest difference is for large correlations where "clumps" will disturb the Poisson approximation, while the compound Poisson approximation remains unaffected.

If we use the simpler compound Poisson distribution, the gain in speed compared to ordinary Poisson approximation remains the same for weakly correlated sequences. However, for strongly correlated sequences, clumps of size larger than two will affect not only the Poisson approximation but also the simple compound Poisson approximation. Although the rate of convergence is far from the "optimal" value one, it is still much better than ordinary Poisson approximation.

The three functions plotted in Figure 4.2 are in fact only upper bounds for the total variation distance for the three present distributions. However, from Section 3.1, we know that the bound for Poisson convergence is of the right order. Hence the difference in convergence rate for compound Poisson compared to Poisson is at least as shown in

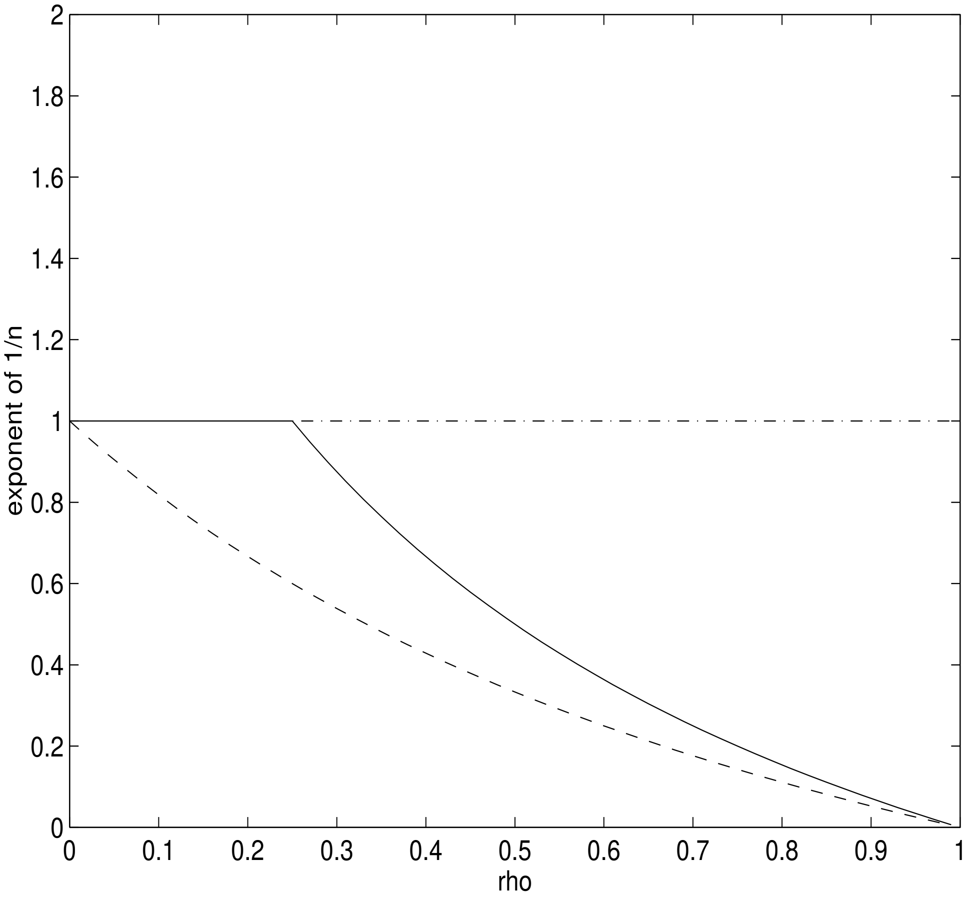

Figure 4.2: *Comparison of the convergence rate as a function of ρ for the three distributions: a general compound Poisson (dash-dotted), a two-term compound Poisson (solid) and the ordinary Poisson distribution (dashed).*

Figure 4.2.

To summarize, compound Poisson approximation will give a substantially faster convergence, unless there are only weak correlations, and therefore no clumping occurs. Compound Poisson approximation with only two terms works well for low and moderate correlations. The convergence rate is equal for Poisson and compound Poisson only in the trivial case of the Y_k's being independent.

Finally we note that, since the approximation in Subsection 4.3.2 in general is not sharp, the gain in the rate of convergence for the simple compound Poisson approximation is in fact even greater than that shown in Figure 4.2. Moreover, if we consider that the simple compound Poisson approximation is too crude, we can use e.g. one more term in the approximating distribution. A calculation similar to the one in Subsection 4.3.1 for λ_3 must be carried out for λ_4. Essentially, the same technique used in Subsection 4.3.1 can be used, although the amount of work will increase. Adding more terms to the approximating distribution will increase the rate of convergence until it reaches the "optimal" value of one.

4.4 Numerical results

In this last section of Chapter 4, we illustrate numerically the results from the previous sections. Two specific MA-sequences are examined, and the convergence of the number of extreme points to a certain distribution is compared for the three distributions:

- a general compound Poisson distribution,
- a compound Poisson distribution with only two terms,
- an ordinary Poisson distribution.

In Theorem 4.3, an upper limit for the total variation distance between W and a general compound Poisson distribution with $\boldsymbol{\lambda}$ defined by (4.14) is developed. By exchanging Theorem 4.1 for Theorem 4.2 in the derivation of Theorem 4.3, the following counterpart for the two-term compound Poisson distribution is obtained (if $l = 2$).

Theorem 4.19 *Let W and $\boldsymbol{\lambda}^*$ be defined by (4.12),(4.14) and (4.10). Then*

$$d(W, CPo(\boldsymbol{\lambda}^*)) \leq c_2(\boldsymbol{\lambda}^*) \left((1+6q)n(1-\Phi(a))^2 + \sum_{j=l+1}^{2q+1} j(j-1)\lambda_j \right).$$

Moreover, if $k\lambda_k^*$ is decreasing in k,

$$c_2(\boldsymbol{\lambda}^*) \leq \min\left(1, \frac{1}{\lambda_1^* - 2\lambda_2^*}\left(\frac{1}{4(\lambda_1^* - 2\lambda_2^*)} + \max\left(0, \ln 2(\lambda_1^* - 2\lambda_2^*)\right)\right)\right).$$

Finally, to be able to also make comparisons with ordinary Poisson approximation, the next theorem from Holst and Janson [9, Theorem 3.1], also stated in Section 3.1, is useful.

Theorem 4.20 *Let I_i, $i \in \Gamma$, W and λ_0 be defined by (4.12). Then*

$$d(W, Po(\lambda_0)) \leq \frac{1 - e^{-\lambda_0}}{\lambda_0}\left(\frac{\lambda_0^2}{n} + \sum_{i\in\Gamma}\sum_{j\in\Gamma_i}|\mathrm{Cov}(I_i, I_j)|\right).$$

1. MA(2).

Consider first the standardized MA(2)-sequence

$$Y_t = \frac{1}{\sqrt{2}}\left(\varepsilon_t + 0.8\varepsilon_{t-1} + 0.6\varepsilon_{t-2}\right), \qquad \{\varepsilon_t\} \sim \mathrm{WN}(0,1).$$

The correlation function is

$$\rho_k = \begin{cases} 0.64 & k = 1 \\ 0.30 & k = 2 \\ 0 & k \geq 3 \end{cases}$$

In Tables 4.1 – 4.4, the parameters of the approximating compound Poisson distributions are presented for four different values of n and four different values of $\lambda_0 = E(W)$. For the calculations of $\lambda_1, \ldots, \lambda_5$, i.e. evaluations of probabilities involving the multivariate normal distribution, the FORTRAN program RIND was used, see Rychlik and Lindgren [21].

In the Poisson case, the covariance terms in Theorem 4.20 can be rewritten as probabilities for bivariate normal distributions, and therefore computed either by RIND or by the routine “quad8” in MATLAB. The results presented here are based on calculations in MATLAB. A detailed description of the algorithm is found in Raab [16]; see also Section 2.4.

	λ_0			
	1	3	5	10
a	2.33	1.88	1.64	1.28
λ_1	0.58	1.35	1.88	2.69
λ_1^*	0.69	1.97	3.26	6.72
λ_2, λ_2^*	0.15	0.52	0.87	1.64
λ_3	0.034	0.18	0.38	1.01
λ_4	$2.00 \cdot 10^{-3}$	0.019	0.052	0.20
λ_5	$1.3 \cdot 10^{-4}$	$1.9 \cdot 10^{-3}$	$6.8 \cdot 10^{-3}$	0.038

Table 4.1: MA(2), $n = 100$.

	λ_0			
	1	3	5	10
a	2.88	2.51	2.33	2.05
λ_1	0.72	1.89	2.89	4.98
λ_1^*	0.77	2.14	3.45	6.65
λ_2, λ_2^*	0.12	0.43	0.77	1.68
λ_3	0.014	0.078	0.17	0.49
λ_4	$3.8 \cdot 10^{-4}$	$3.6 \cdot 10^{-3}$	0.010	0.042
λ_5	$1.4 \cdot 10^{-5}$	$2.0 \cdot 10^{-4}$	$6.9 \cdot 10^{-4}$	$3.7 \cdot 10^{-3}$

Table 4.2: MA(2), $n = 500$.

	λ_0			
	1	3	5	10
a	3.09	2.75	2.58	2.33
λ_1	0.76	2.08	3.23	5.78
λ_1^*	0.80	2.24	3.61	6.90
λ_2, λ_2^*	0.10	0.38	0.70	1.55
λ_3	$9.3 \cdot 10^{-3}$	0.053	0.12	0.34
λ_4	$1.8 \cdot 10^{-4}$	$1.7 \cdot 10^{-3}$	$5.0 \cdot 10^{-3}$	0.020
λ_5	$5.2 \cdot 10^{-5}$	$7.4 \cdot 10^{-5}$	$2.6 \cdot 10^{-4}$	$1.4 \cdot 10^{-3}$

Table 4.3: MA(2), $n = 1000$.

	λ_0			
	1	3	5	10
a	3.54	3.24	3.09	2.88
λ_1	0.84	2.37	3.82	7.19
λ_1^*	0.86	2.46	3.98	7.65
λ_2, λ_2^*	0.070	0.27	0.51	1.17
λ_3	$3.4 \cdot 10^{-3}$	0.020	0.046	0.14
λ_4	$3.0 \cdot 10^{-5}$	$3.1 \cdot 10^{-4}$	$9.0 \cdot 10^{-4}$	$3.8 \cdot 10^{-3}$
λ_5	$4.9 \cdot 10^{-7}$	$7.5 \cdot 10^{-6}$	$2.6 \cdot 10^{-5}$	$1.4 \cdot 10^{-4}$

Table 4.4: MA(2), $n = 5000$.

			λ_0			
			1	3	5	10
n	100	CPo($\boldsymbol{\lambda}$)	0.13	+	+	+
		CPo($\boldsymbol{\lambda}^*$)	0.36	+	+	+
		Po(λ)	0.32	0.67	0.80	0.96
	500	CPo($\boldsymbol{\lambda}$)	0.026	0.22	0.57	+
		CPo($\boldsymbol{\lambda}^*$)	0.11	0.66	+	+
		Po(λ)	0.20	0.42	0.51	0.63
	1000	CPo($\boldsymbol{\lambda}$)	0.013	0.10	0.25	0.86
		CPo($\boldsymbol{\lambda}^*$)	0.070	0.39	0.79	+
		Po(λ)	0.16	0.34	0.41	0.51
	5000	CPo($\boldsymbol{\lambda}$)	0.0026	0.018	0.042	0.12
		CPo($\boldsymbol{\lambda}^*$)	0.023	0.12	0.22	0.52
		Po(λ)	0.10	0.21	0.26	0.32

Table 4.5: *Upper bounds for the total variation distance for the MA(2)-sequence. The "+"-signs indicate values above 1.*

Table 4.5 shows the calculated upper bounds for the total variation distance for the same values of n and λ_0 as in the previous tables. For small n, the bounds are lower for Poisson approximation than for compound Poisson approximation. This is not a contradictory result since the computed values are upper bounds only, not the actual distance of variation. The estimates of the constants are sharper for Poisson approximation then for compound Poisson approximation. A numerical examination of this assertion is performed below.

For larger values of n, we observe that the bounds are substantially lower for compound Poisson approximation and the rate of convergence is better than for Poisson approximation.

Of course, the general compound Poisson approximation is always superior to the approximation with only two terms.

The worst combination of correlations (in the sense of minimizing ν in Theorem 4.16) is $(0, 0.30, 0.64)$, corresponding to consecutive variables. The exponent of $1/n$ of the convergence rate is 0.22 in the Poisson case, 0.54 for the two-term compound Poisson distribution and 1 for the general five-term compound Poisson distribution.

2. MA(3).

The second example is an MA(3)-sequence:

$$Y_t = \frac{2}{\sqrt{41}}\left(\varepsilon_t + 0.4\varepsilon_{t-1} + 0.3\varepsilon_{t-2} + 3\varepsilon_{t-3}\right), \qquad \{\varepsilon_t\} \sim \mathrm{WN}(0,1),$$

the correlation function of which is

$$\rho_k = \begin{cases} 0.1385 & k=1 \\ 0.1463 & k=2 \\ 0.2927 & k=3 \\ 0 & k \geq 4 \end{cases}$$

The parameters for the approximating compound Poisson distributions are presented in Tables 4.6 – 4.9 for the same values of n and λ_0 as in the previous example. In Table 4.10 the bounds for the total variation distance are shown. The conclusions from the first example are also valid here. The convergence rate (i.e. the exponent of $1/n$) is 0.55 for Poisson and 1 for both cases of compound Poisson approximation. In other words, using a two-term compound Poisson distribution is much better than a Poisson distribution, but adding more terms does not improve the approximation asymptotically.

We conclude this section by comparing "traditional" Poisson approximation as in Holst and Janson [9] with "one-term" compound Poisson approximation, i.e. also Poisson approximation. The calculations following Theorem 4.16, particularly equation (4.54), show that, asymptotically, the rate of convergence is the same for both methods. However, for fixed n, the methods differ as we shall now investigate.

Let $\rho = \max(\rho_1, \rho_2, \ldots, \rho_q)$. Assume for simplicity that $k\lambda_k$ is decreasing in k, $\min(\rho_1, \rho_2, \ldots, \rho_q) \geq 0$ and $n \gg q$. Assume also that a is large enough to make the approximation below in (4.56) valid. From Theorem 4.20,

$$d(W, \mathrm{Po}(\lambda_0)) \leq \frac{1 - e^{-\lambda_0}}{\lambda_0} \left(\frac{\lambda_0^2}{n} + \sum_{i \in \Gamma} \sum_{j \in \Gamma_i} |\mathrm{Cov}(I_i, I_j)| \right).$$

Since $\min(\rho_1, \rho_2, \ldots, \rho_q) \geq 0$,

$$\sum_{i \in \Gamma} \sum_{j \in \Gamma_i} |\mathrm{Cov}(I_i, I_j)| = \sum_{i \in \Gamma} \sum_{j \in \Gamma_i} \mathrm{Cov}(I_i, I_j)$$
$$= \sum_{k=1}^{n-1} 2(n-k)\mathrm{Cov}(I_1, I_{1+k}) = \sum_{k=1}^{q} 2(n-k)\mathrm{Cov}(I_1, I_{1+k})$$

	λ_0			
	1	3	5	10
a	2.33	1.88	1.64	1.28
λ_1	0.82	1.96	2.72	3.65
λ_1^*	0.84	2.18	3.36	6.32
λ_2, λ_2^*	0.081	0.41	0.82	1.84
λ_3	$5.9 \cdot 10^{-3}$	0.060	0.17	0.62
λ_4	$3.6 \cdot 10^{-4}$	$7.1 \cdot 10^{-3}$	0.027	0.16
λ_5	$1.8 \cdot 10^{-5}$	$6.4 \cdot 10^{-4}$	$3.3 \cdot 10^{-3}$	0.029
λ_6	$6.3 \cdot 10^{-7}$	$4.0 \cdot 10^{-5}$	$2.7 \cdot 10^{-4}$	$3.7 \cdot 10^{-3}$
λ_7	$1.1 \cdot 10^{-8}$	$1.2 \cdot 10^{-6}$	$1.1 \cdot 10^{-5}$	$2.4 \cdot 10^{-4}$

Table 4.6: MA(3), $n = 100$.

	λ_0			
	1	3	5	10
a	2.88	2.51	2.33	2.05
λ_1	0.93	2.59	4.08	7.22
λ_1^*	0.94	2.63	4.18	7.67
λ_2, λ_2^*	0.032	0.19	0.41	1.16
λ_3	$8.6 \cdot 10^{-4}$	$9.8 \cdot 10^{-3}$	0.030	0.13
λ_4	$2.2 \cdot 10^{-5}$	$4.6 \cdot 10^{-4}$	$1.9 \cdot 10^{-3}$	0.012
λ_5	$4.9 \cdot 10^{-7}$	$1.8 \cdot 10^{-5}$	$9.4 \cdot 10^{-5}$	$9.0 \cdot 10^{-4}$
λ_6	$7.8 \cdot 10^{-9}$	$4.8 \cdot 10^{-7}$	$3.3 \cdot 10^{-6}$	$4.5 \cdot 10^{-5}$
λ_7	$6.0 \cdot 10^{-11}$	$6.5 \cdot 10^{-9}$	$5.8 \cdot 10^{-8}$	$1.1 \cdot 10^{-6}$

Table 4.7: MA(3), $n = 500$.

	λ_0			
	1	3	5	10
a	3.09	2.75	2.58	2.33
λ_1	0.95	2.73	4.39	8.15
λ_1^*	0.96	2.75	4.44	8.35
λ_2, λ_2^*	0.021	0.12	0.28	0.82
λ_3	$3.7 \cdot 10^{-4}$	$4.3 \cdot 10^{-3}$	0.013	0.060
λ_4	$6.5 \cdot 10^{-6}$	$1.4 \cdot 10^{-4}$	$5.6 \cdot 10^{-4}$	$3.8 \cdot 10^{-3}$
λ_5	$1.0 \cdot 10^{-7}$	$3.7 \cdot 10^{-6}$	$2.0 \cdot 10^{-5}$	$1.9 \cdot 10^{-4}$
λ_6	$1.2 \cdot 10^{-9}$	$7.2 \cdot 10^{-8}$	$4.9 \cdot 10^{-7}$	$6.6 \cdot 10^{-6}$
λ_7	$6.4 \cdot 10^{-12}$	$6.8 \cdot 10^{-10}$	$6.0 \cdot 10^{-9}$	$1.2 \cdot 10^{-7}$

Table 4.8: MA(3), $n = 1000$.

	λ_0			
	1	3	5	10
a	3.54	3.24	3.09	2.88
λ_1	0.96	2.89	4.76	9.30
λ_1^*	0.98	2.91	4.79	9.35
λ_2, λ_2^*	$7.7 \cdot 10^{-3}$	0.046	0.11	0.32
λ_3	$4.8 \cdot 10^{-5}$	$5.8 \cdot 10^{-4}$	$1.8 \cdot 10^{-3}$	$8.7 \cdot 10^{-3}$
λ_4	$3.6 \cdot 10^{-7}$	$7.8 \cdot 10^{-6}$	$3.2 \cdot 10^{-5}$	$2.2 \cdot 10^{-4}$
λ_5	$2.6 \cdot 10^{-9}$	$9.6 \cdot 10^{-8}$	$5.1 \cdot 10^{-7}$	$4.9 \cdot 10^{-6}$
λ_6	$1.4 \cdot 10^{-11}$	$8.6 \cdot 10^{-10}$	$5.8 \cdot 10^{-9}$	$7.9 \cdot 10^{-8}$
λ_7	$3.3 \cdot 10^{-14}$	$3.7 \cdot 10^{-12}$	$3.2 \cdot 10^{-11}$	$6.1 \cdot 10^{-10}$

Table 4.9: MA(3), $n = 5000$.

			λ_0			
			1	3	5	10
n	100	CPo($\boldsymbol{\lambda}$)	0.19	+	+	+
		CPo($\boldsymbol{\lambda}^*$)	0.23	+	+	+
		Po(λ)	0.10	0.27	0.37	0.51
	500	CPo($\boldsymbol{\lambda}$)	0.037	0.25	0.57	+
		CPo($\boldsymbol{\lambda}^*$)	0.042	0.29	0.68	+
		Po(λ)	0.038	0.11	0.15	0.23
	1000	CPo($\boldsymbol{\lambda}$)	0.018	0.12	0.26	0.76
		CPo($\boldsymbol{\lambda}^*$)	0.020	0.14	0.31	0.91
		Po(λ)	0.025	0.073	0.10	0.16
	5000	CPo($\boldsymbol{\lambda}$)	0.0036	0.022	0.047	0.13
		CPo($\boldsymbol{\lambda}^*$)	0.0039	0.024	0.053	0.14
		Po(λ)	0.0095	0.028	0.040	0.061

Table 4.10: *Upper bounds for the total variation distance for the MA(3)-sequence. The "+"-signs indicate values above 1.*

$$\leq \sum_{k=1}^{q} 2(n-k)P(I_1 = 1, I_{1+k=1}) \leq 2n \sum_{k=1}^{q} P(I_1 = 1, I_{1+k} = 1).$$

Hence

$$d(W, \mathrm{Po}(\lambda_0)) \leq \frac{1-e^{-\lambda_0}}{\lambda_0} \left(\frac{\lambda_0^2}{n} + 2n \sum_{k=1}^{q} P(I_1 = 1, I_{1+k} = 1) \right). \quad (4.55)$$

We now turn to the bound in Theorem 4.19. Neglect edge effects. Inserting $l = 1$ in the theorem gives

$$c_2 \left((1+6q)n(1-\Phi(a))^2 + \sum_{j=l+1}^{2q+1} j(j-1)\lambda_j \right)$$

$$\geq c_2 \left((1+6q)n(1-\Phi(a))^2 + 2\lambda_2 \right)$$

$$= c_2 \left((1+6q)n(1-\Phi(a))^2 + \sum_{i\in\Gamma} E(I_i I(Z_i = 2)) \right)$$

$$\geq c_2 \left((1+6q)\frac{\lambda_0^2}{n} + \sum_{i\in\Gamma} \sum_{k=1}^{q} 2P(I_i = 1, I_{i+k} = 1, I_j = 0;\ j \in \Gamma_i^{vs} \backslash \{i+k\}) \right)$$

$$\approx c_2 \left((1+6q)\frac{\lambda_0^2}{n} + 2n \sum_{k=1}^{q} P(I_1 = 1, I_{1+k} = 1) \right), \quad (4.56)$$

where we define $I_{i+k} = I_{i-k}$ and $\Gamma_i^{vs} \backslash \{i+k\}$ as $\Gamma_i^{vs} \backslash \{i-k\}$ for $i+k \notin \Gamma_i^{vs}$, and

$$c_2 = \min\left(1, \frac{1}{\lambda_0}\left(\frac{1}{4\lambda_0} + \max(0, \ln(2\lambda_0))\right)\right).$$

Define

$$d_1(q) = \frac{c_2(1+6q)}{(1-e^{-\lambda_0})/\lambda_0}, \qquad d_2 = \frac{c_2}{(1-e^{-\lambda_0})/\lambda_0}.$$

Comparing expressions (4.55) and (4.56), we observe that the two coefficients in (4.55) are d_1 and d_2 times smaller respectively (at least for a large enough to make the approximation in (4.56) valid).

Some values of $d_1(q)$ and d_2 for different λ_0 are displayed in Table 4.11. We see that for λ_0 as in the examples above, d_1 is in the range of 1.5 to 3 (and d_2 even larger), explaining the apparent result in Table 4.5 and Table 4.10 that Poisson approximation is better than compound Poisson approximation. It seems important to improve the bound for c_2 in order to achieve a better bound for the total variation distance for compound Poisson approximation. However, the question of how to bound c_2 better is not simple – obviously more research is needed beyond the limit of the present work.

λ_0	1	3	5	10
$d_1(2)$	19.4	25.7	30.8	39.3
$d_1(3)$	28.3	37.5	45.0	57.4
d_2	1.49	1.97	2.37	3.02

Table 4.11: $d_1(q)$ and d_2 as a function of λ_0.

Chapter 5

Concluding remarks

The main task of this thesis has been to generalize previous results on Poisson approximation for Gaussian sequences. The research can be characterized by two simple questions:

- What happens if the sequence no longer is Gaussian?
- For sequences with substantial "local" dependence, Poisson approximation of the number of exceeding points, W, is poor. In what way and to what extent can the approximation of W be improved?

Seeking answers to the first question, sequences related to the Gaussian, primarily χ^2–sequences, have been discussed in Chapter 2. For χ^2–sequences, the results are similar to those in the Gaussian case as regards conditions for Poisson convergence as well as the rate of convergence. A basic requirement for the method, the coupling version of the Stein-Chen method, is that it must be possible to prove the existence of a monotone coupling of suitable indicators. For χ^2–sequences, Proposition 2.2 in Section 2.2 serves this purpose. We note that this proposition is by no means restricted to χ^2–sequences only. Any sequence built up as a monotone function of a Gaussian sequence, either directly or by introducing more variables, can be treated in a similar manner. Estimates of the covariances of the indicators will, on the other hand, depend on the underlying distribution. In Section 2.3, covariances for $\chi^2(2)$-sequences have been investigated.

Turning to the second question, some answers are given in Chapter 4. Using two terms in the approximating compound Poisson distribution gives a better approximation than ordinary Poisson approximation (except for white noise, where no clumping occurs at all). Adding more than two terms improves the convergence rate in some cases. The results are demonstrated in Figure 4.2, giving the rate of convergence as a function of the largest correlation in the sequence considered.

Many of the results in this thesis are valid for stationary as well as for non-stationary sequences. Those results that assume stationarity can be relaxed to cover non-stationary sequences as follows. In Chapter 2, stationarity is used only in Theorem 2.7 and the corollaries thereafter. The assumption is used only to collect terms with equal covariances. Nevertheless, some conditions must be imposed on the covariances in order to limit the "global" dependence.

Stationarity is not a crucial assumption in Chapter 4 either. We will now consider what changes are necessary to also allow non-stationary sequences. To begin with, condition (4.11) must be adapted to general sequences. A natural analogy to (4.11) is to assume that the points in the sequence considered, $\{Y_k\}$, are independent whenever the distance between them exceeds, say, q, i.e.

$$\operatorname{Cov}(Y_i, Y_j) = 0 \qquad \forall\, i, j : \; |i - j| > q. \tag{5.1}$$

Still assuming that $\{Y_k\}$ is standardized, the derivation of Theorem 4.3 and Theorem 4.4 works also under this condition. Equation (4.14) for computing the λ_k's is also applicable. Note, however, that the terms in the sum in (4.14) will in general not be equal, as in the case for stationary sequences (except for points near the endpoints of Γ).

In Section 4.3 statements involving the correlation function $\{\rho_k\}$ must be modified. Assume for simplicity that the vector $(Y_1, Y_2, \ldots, Y_n)$ has a non-singular Gaussian distribution, although it is a rather straightforward matter to generalize the statements to cover singular distributions also. In Proposition 4.13 and Theorem 4.16, the maximization should be carried out over the set of triplets of correlations (sorted in ascending order as in the stationary case) corresponding to all possible combinations of choosing three variables from $\{Y_k\}$ so that all three correlations are positive. Proposition 4.15 and Corollary 4.18 are also valid if we redefine ρ as

$$\rho = \max\left(0, \sup_{\substack{i,j\in\{1,\dots,n\}\\ i\neq j}} \{\mathrm{Cov}(Y_i, Y_j)\}\right),$$

i.e., as the largest correlation in the correlation matrix of the sequence.

Finally we comment on the assumption (4.11) in Chapter 4 for stationary sequences (and condition (5.1) in the non-stationary case). This condition eliminates the "global" dependence in the sequence. For Poisson approximation, we know that it is sufficient for convergence to the approximating distribution to impose the less restrictive Berman condition, $\rho_k \ln k \to 0$ as $k \to \infty$. The main reason for using the compound Poisson distribution is that the effect of clumping can be taken into account. It seems reasonable to think that this improvement in approximation error compared to the Poisson distribution is valid independently of the "global" dependence for the sequence. Furthermore, the longer the sequence is, the more the approximating compound Poisson distribution tends to the limiting Poisson distribution. Since the Berman condition ensures Poisson convergence, generalization of the results in Chapter 4 under this condition seems possible. However, an attempt to verify this will lead to substantial technical problems. Investigating how to overcome these difficulties would be a suitable topic for future research.

Bibliography

[1] J. M. P. Albin. On the general law of iterated logarithm with application to selfsimilar processes and to Gaussian processes in R^n and Hilbert space. *Stochastic Processes and their Applications*, 41:1–31, 1992.

[2] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: The Stein-Chen method. *Annals of Probability*, 17:9–25, 1989.

[3] A. Barbour, L. Holst, and S. Janson. *Poisson Approximation*. Oxford University Press, Oxford, 1992.

[4] A. D. Barbour, L. H. Y. Chen, and W. Loh. Compound Poisson approximation for nonnegative random variables via Stein's method. *Annals of Probability*, 20:1843–1866, 1992.

[5] S. M. Berman. Limit theorems for the maximum term in stationary sequences. *Annals of Mathematical Statistics*, 35:502–516, 1964.

[6] L. H. Y. Chen. Poisson approximation for dependent trials. *Annals of Probability*, 3:534–545, 1975.

[7] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, N.J., 1946.

[8] H. Cramér and M. R. Leadbetter. *Stationary and Related Stochastic Processes*. Wiley, 1967.

[9] L. Holst and S. Janson. Poisson approximation using the Stein-Chen method and coupling: Number of exceedances of Gaussian random variables. *Annals of Probability*, 18:713–723, 1990.

[10] K. Joag-Dev, D. Perlman, and L. Pitt. Association of normal random variables and Slepian's inequality. *Annals of Probability*, 11:451–455, 1983.

[11] T. Kamae, U. Krengel, and L. O'Brien. Stochastic inequalities on partially ordered spaces. *Annals of Probability*, 5:899–912, 1977.

[12] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York, 1983.

[13] G. Lindgren. Extreme values and crossings for the χ^2–process and other functions of multidimensional Gaussian processes with reliability applications. *Advances in Applied Probability*, 12:746–774, 1980.

[14] S. D. Poisson. *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités.* Bachelier, Paris, 1837.

[15] M. Raab. Poisson approximation of the number of exceedances of a discrete-time χ^2–process. *Stochastic Processes and their Applications (To appear).*

[16] M. Raab. Poisson approximation using the Stein-Chen method and coupling: Number of exceedances of a discrete-time χ^2-process. Technical Report TRITA-MAT-1995-MS-1, Department of Mathematics, Royal Institute of Technology, Stockholm, 1995.

[17] C. R. Rao. *Linear Statistical Inference and Its Applications.* Wiley, 1968.

[18] M. Roos. Compound Poisson approximation for the numbers of extreme spacings. *Advances in Applied Probability*, 25:847–874, 1993.

[19] M. Roos. Stein's method for compound Poisson approximation: The local approach. *Annals of Applied Probability*, 4:1177–1187, 1994.

[20] H. Rootzén. The rate of convergence of extremes of stationary normal sequences. *Advances in Applied Probability*, 15:54–80, 1983.

[21] I. Rychlik and G. Lindgren. CROSSREG - a technique for first passage and wave density analysis. *Probability in the Engineering and Informational Sciences*, 7:705–728, 1993.

[22] R. F. Serfozo. Compound Poisson approximations for sums of random variables. *Annals of Probability*, 14:1391–1398, 1986.

[23] K. Sharpe. Some properties of the crossing process generated by a stationary χ^2 process. *Advances in Applied Probability*, 10:373–391, 1978.

[24] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, 2:583–602, 1970.

[25] C. Stein. *Approximate computation of expectations, Institute of Mathematical Statistics Lecture Notes - Monograph Series, Vol. 7.* Hayward, California, 1986.

[26] Y. L. Tong. *Probability Inequalities in Multivariate Distributions.* Academic Press, 1980.

[27] Y. H. Wang. A compound Poisson convergence theorem. *Annals of Probability*, 19:452–455, 1991.