MODELING AND GENERATING DEPENDENT RISK PROCESSES FOR IRM AND DFA

BY

DIETMAR PFEIFER AND JOHANA NEŠLEHOVÁ

ABSTRACT

Modern Integrated Risk Management (IRM) and Dynamic Financial Analysis (DFA) rely in great part on an appropriate modeling of the stochastic behavior of the various risky assets and processes that influence the performance of the company under consideration. A major challenge here is a more substantial and realistic description and modeling of the various complex dependence structures between such risks showing up on all scales. In this presentation, we propose some approaches towards modeling and generating (simulating) dependent risk processes in the framework of collective risk theory, in particular w.r.t. dependent claim number processes of Poisson type (homogeneous and non-homogeneous), and compound Poisson processes.

1. INTRODUCTION

In the recent years it has become more and more popular for insurance companies and other financial institutions to use concepts like DFA (Dynamic Financial Analysis) for their risk management systems. The central idea behind this is to map all relevant assets and liabilities together with potential cash flows and other monetary transactions into a large-scale computer model in order to simulate the development of the balance sheet over some time. With such integrative approaches, not only potential effects of sophisticated strategic management decisions can be studied in vitro, but also negative business developments become visible early enough so that preventive measures could be taken in time.

The reliability of such systems, however, is largely determined by the precision with which the mostly random processes and their complex interrelationships involved are depicted here. For the common user of statistical methods, "correlation" is perhaps the first probabilistic concept to remember, in particular since this kind of "dependence" concept has introduced itself at large to the financial markets, supported by theoretical approaches like the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory (APT). This is due to the fact that these theories are founded on the "normal" world where in the multivariate setting, the joint distributions are indeed characterized by pairwise correlation (or covariance, resp.). This is not only no longer true outside

that world but also very misleading results can be produced when the concept of correlation is applied without caution, as has been pointed out in several papers in the recent past (see e.g. Embrechts et al. [8], [7], or Blum et al. [2]). The paper of Müller and Bäuerle [21] shows that this topic has meanwhile also raised sufficient interest in the insurance world.

2. COPULAS

When modeling multivariate distributions, one has to take into account the effects of the marginal distributions as well as of the dependence between them. This can be elegantly achieved by using the copula approach, which allows to deal with the margins and the dependence structure separately. The name "copula" has been first mentioned by Abe Sklar in 1959 as a function which couples a joint distribution function with its univariate margins. Since then, a great deal of papers has been published on this topic; a good introduction can be found in Nelsen [23].

Definition 2.1. A *copula* is a function C of n variables on the unit n-cube $[0,1]^n$ with the following properties:

1. the range of C is the unit interval $[0,1]$;
2. $C(\mathbf{u})$ is zero for all $\mathbf{u}$ in $[0,1]^n$ for which at least one coordinate equals zero;
3. $C(\mathbf{u}) = u_k$ if all coordinates of $\mathbf{u}$ are 1 except the k-th one;
4. C is n-increasing in the sense that for every $\mathbf{a} \le \mathbf{b}$ in $[0,1]^n$ the volume assigned by C to the n-box $[\mathbf{a}, \mathbf{b}] = [a_1, b_1] \times \cdots \times [a_n, b_n]$ is nonnegative. ∎

As can be easily seen, a copula is in fact a multivariate distribution function with univariate margins restricted to the n-cube. Copulas have many useful properties, such as uniform continuity and existence of all partial derivatives, just to mention a few. Moreover, it can be shown that every copula is bounded by the so-called *Fréchet-Hoeffding* bounds, i.e.

$$\max(u_1 + \cdots + u_n - n + 1, 0) \le C(u_1, \cdots, u_n) \le \min(u_1, \cdots, u_n)$$

which are commonly denoted by $\mathcal{W}$ and $\mathcal{M}$ in the literature. In two dimensions, both of the *Fréchet-Hoeffding* bounds are copulas themselves, but as soon as the dimension increases, the *Fréchet-Hoeffding* lower bound $\mathcal{W}$ is no longer n-increasing. However, the inequality on the left-hand side cannot be improved, since for any $\mathbf{u}$ from the unit n-cube, there exists a copula $C_{\mathbf{u}}$ such that $\mathcal{W}(\mathbf{u}) = C_{\mathbf{u}}(\mathbf{u})$ (see Nelsen [23], theorem 2.10.12).

Another well-known copula is the so-called *independence copula*, $\prod(\mathbf{u}) = \prod_{i=1}^{n} u_i$.

The following theorem due to Sklar justifies the role of copulas as dependence functions:

Theorem 2.2 (Sklar). Let H denote a n-dimensional distribution function with margins $F_1, \cdots, F_n$. Then there exists a n-copula C such that for all real $(x_1, \cdots, x_n)$,

$$H(x_1, \cdots, x_n) = C\big(F_1(x_1), \cdots, F_n(x_n)\big).$$

If all the margins are continuous, then the copula is unique, and is determined uniquely on the ranges of the marginal distribution functions otherwise. Moreover, the converse of the above statement is also true. If we denote by $F_1^{-1}, \cdots, F_n^{-1}$ the generalized inverses of the marginal distribution functions, then for every $(u_1, \cdots, u_n)$ in the unit n-cube,

$$C(u_1, \cdots, u_n) = H\big(F_1^{-1}(u_1), \cdots, F_n^{-1}(u_n)\big). \qquad \blacksquare$$

For a proof, see Nelsen [23], Theorem 2.10.9 and the references given therein.

Consequently, it is possible to construct a wide range of multivariate distributions by choosing the marginal distributions and a suitable copula. Moreover, the above theorem shows that copulas remain invariant under strictly increasing transformations of the underlying random variables. Besides this, they are uniformly continuous and all their partial derivatives exist almost everywhere, which is a useful property especially for computer simulations.

Since their discovery, various interesting properties of copulas have been extensively studied in the literature. However, most part of the research concentrates on the bivariate case, since multivariate extensions are generally less easily to be done. For further information on this subject, we refer to Nelsen [23], Joe [13] and the references mentioned therein.

Due to the Fréchet-Hoeffding inequality, every multivariate distribution function is non-trivially bounded, i.e. for every real $(x_1, \cdots, x_n)$ the following holds:

$$\mathcal{W}\big(F_1(x_1), \cdots, F_n(x_n)\big) \leq H(x_1, \cdots, x_n) \leq \mathcal{M}\big(F_1(x_1), \cdots, F_n(x_n)\big).$$

In the bivariate case, both the Fréchet-Hoeffding bounds as well as the independence copula have the following stochastic representation, as was already noticed by Hoeffding [11]:

Theorem 2.3. Let U and V be random variables uniformly distributed over the unit interval [0,1]. Then their joint distribution function restricted to the unit square $[0,1]^2$ is equal to

- $\mathcal{W}$, if and only if $U = 1 - V$ a.s.,
- Π, if and only if U and V are independent,
- $\mathcal{M}$, if and only if $U = V$ a.s. $\qquad \blacksquare$

In other words, every bivariate dependence structure lies somewhere between two extremes, the perfect negative and the perfect positive dependence. Because of this fact copulas can be partially ordered in the following way:

Definition 2.4. A copula C_1 is *smaller* than a copula C_2, in symbols: $C_1 \prec C_2$ if for any **u** in the unit square $[0,1]^2$, $C_1(\mathbf{u}) \leq C_2(\mathbf{u})$. ∎

As mentioned above, copulas reflect the dependence structure between the margins, and are therefore of great use in various dependence and association concepts. For instance, the well known bivariate concordance measures Kendall's τ and Spearman's ρ as well as the tail dependence and likelihood ratio dependence concepts can be expressed in terms of the underlying copula alone. However, the role played by copulas in the study of multivariate dependence is much more complex and far less well understood (for further details see Mari and Kotz [20], Nelsen [23] or Joe [13] and the references given therein).

Throughout this paper we will focus on one special bivariate dependence measure, the *linear correlation coefficient*,

$$\rho = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}.$$

It is easily seen that ρ does not depend on the underlying copula alone and will therefore be influenced by the marginal distributions as well. However, the following result from Hoeffding [11] suggests that the role played by copulas in this setting will nevertheless be important:

Theorem 2.5. Let (X, Y) be a bivariate random vector with a copula C and marginal distribution functions F and G such that $E(|X|) < \infty$, $E(|Y|) < \infty$ and $E(|XY|) < \infty$. Then the covariance between X and Y can be expressed in the following way:

$$\text{Cov}(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[C(F(x), G(y)) - F(x)G(y) \right] dxdy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[C(F(x), G(y)) - \prod(F(x), G(y)) \right] dxdy.$$

∎

This result together with the abovementioned Fréchet-Hoeffding inequality has the following consequence for the correlation coefficient:

Theorem 2.6. Let – under the conditions of Theorem 2.6 – (X, Y) be a bivariate random vector with *fixed* marginal distribution functions F and G. In case that the underlying copula is C the corresponding correlation coefficient will be denoted by ρ_C. Then

1. ρ_C is always bounded by the correlation coefficients corresponding to the Fréchet-Hoeffding bounds, $\rho_W \leq \rho_C \leq \rho_M$;
2. if C_1 and C_2 are copulas, then the relation $C_1 \prec C_2$ yields $\rho_{C_1} \leq \rho_{C_2}$. ∎

Cambanis et al. [3] proved that an extension in the following sense is possible:

Theorem 2.7. In the situation of the previous theorem, let $E_C(X,Y)$ denote the expectation corresponding to an underlying copula C. Moreover, let k denote a right continuous real function of two variables, satisfying the so-called *Monge condition*, i.e.

$$k(x,y) + k(x',y') \leq k(x,y') + k(x',y)$$

whenever $x \leq x'$ and $y \leq y'$. If either

1. k is symmetric and both the expectations $Ek(X,X)$ and $Ek(Y,Y)$ are finite, or
2. $Ek(X,y_0)$ and $Ek(x_0,Y)$ are finite for some x_0 and y_0,

and if C_1 and C_2 are copulas, than the relation $C_1 \prec C_2$ yields

$$E_{C_1} k(X,Y) \geq E_{C_2} k(X,Y). \qquad \blacksquare$$

A multivariate extension of this result is also possible, as well as another derivation which requires different assumptions. For more details on this subject, we refer to the work of Rachev and Rüschendorf [28], [29]. See also Mari and Kotz [20] where other aspects of the interplay between correlation and dependence are studied in more detail.

Consider now a bivariate random vector (X,Y) with fixed marginal distributions F and G such that both the correlation coefficients ρ_W and ρ_M exist and are finite. Then each number in the interval $[\rho_W, \rho_M]$ is equal to ρ_C for some copula C. Indeed, if we choose an α from $[0,1]$ and define a copula C_α by

$$C_\alpha := \alpha \cdot W + (1-\alpha) \cdot M,$$

then the corresponding correlation coefficient ρ_{C_α} is $\alpha \cdot \rho_W + (1-\alpha) \cdot \rho_M$. The so constructed one-parameter family of copulas includes the Fréchet-Hoeffding lower and upper bound and allows for both negative and positive correlation. However, it is worth noticing that $\alpha = \frac{\rho_M}{\rho_M - \rho_W}$ yields a zero correlation and a copula which does not correspond to the independence of X and Y. Moreover, the independence copula can never be constructed using the above method.

Other parametric families with similar properties which provide more tractable models have been described in the literature. We will mention the following few.

Example 2.8 (The Fréchet and Mardia families). Let α, β be in $[0,1]$ with $\alpha + \beta \leq 1$. The two-parameter family of copulas

$$C_{\alpha,\beta}(u,v) = \alpha \cdot M(u,v) + (1-\alpha-\beta)\prod(u,v) + \beta \cdot W(u,v)$$

is called the *Fréchet family*. This family includes the Fréchet-Hoeffding bounds as well as the independence copula and yields, upon setting $\alpha = 0$ and $\beta > 0$ a one-parameter family with negative correlation, and, upon setting $\alpha > 0$ and $\beta = 0$ a one-parameter family with positive correlation. In both cases the correlation is a function of the corresponding parameters.

Another one-parametric family can be built by setting $\alpha = \frac{\theta^2}{2}(1-\theta)$ and $\beta = \frac{\theta^2}{2}(1+\theta)$ for $\theta \in [-1,1]$:

$$C_\theta(u,v) = \frac{\theta^2}{2}(1-\theta)\,\mathcal{M}(u,v) + (1-\theta^2)\prod(u,v) + \frac{\theta^2}{2}(1+\theta)\cdot\mathcal{W}(u,v).$$

This family is the so-called *Mardia family* (see e.g. Nelsen [23], p. 12). Again, it includes the Fréchet-Hoeffding bounds as well as the independence copula; also, the correlation of its members ranges over the whole interval $[\rho_\mathcal{W}, \rho_\mathcal{M}]$.

However, neither the Fréchet family nor the Mardia family are ordered and the distribution functions constructed using them do not in general have a probability density.

Example 2.9 (The Frank family). Let θ be in $(-\infty,\infty)/\{0\}$. The one-parameter family defined by

$$C_\theta(u,v) = -\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$$

is called the *Frank family*. The Fréchet bounds are included as limiting cases, i.e. $C_{-\infty} = \mathcal{W}$ and $C_\infty = \mathcal{M}$. The limit $\theta \to 0$ yields the independence copula $\prod$. Moreover, the Frank family is positively ordered, that is, $C_\theta \prec C_{\theta'}$ as soon as $\theta \le \theta'$. From this it follows easily that putting $\theta \le 0$ yields a negative correlation and $\theta \ge 0$ a positive correlation, respectively.

The Frank family has an interesting application in insurance pricing, see e.g. Mari and Kotz [20], p. 78; for recent applications in finance, see e.g. Junker and May [15].

Members of the Frank family are all Archimedean copulas (see Nelsen [23], chapter 4) and the only ones in this class which are symmetric, i.e. for which

$$C(u,v) = u + v - 1 + C(1-u,\ 1-v) \text{ for all } (u,v) \in [0,1]^2$$

with the geometric interpretation that the rectangles $[0,u] \times [0,v]$ and $[1-u,1] \times [1-v,1]$ have the same C-volume (see e.g. Nelsen [23], p. 33).

Example 2.10 (The Plackett family). For $\theta > 0$, $\theta \ne 1$, the members of the *Plackett family* are defined by

$$C_\theta(u,v) = \frac{[1+(\theta-1)(u+v)] - \sqrt{[1+(\theta-1)(u+v)]^2 - 4uv\theta(\theta-1)}}{2(\theta-1)}$$

and for $\theta = 1$ by $C_1 := \prod$. They are absolutely continuous and again have the Fréchet-Hoeffding bounds as limiting cases, i.e. $C_0 = \mathcal{W}$ and $C_\infty = \mathcal{M}$. Like the Frank family, the Plackett family is positively ordered and symmetric, but no longer Archimedean. Consequently, putting $\theta < 1$ yields a negative correlation and $\theta > 1$ a positive correlation, respectively.

According to the Sklar's theorem, for any given fixed marginals each of the abovementioned families can be used to construct a parametric family of bivariate distributions, which contains both negatively and positively correlated members. Moreover, the minimum as well as the maximum correlation can be achieved. We will use this method in the next chapter and construct bivariate Poisson distributions, which are negatively correlated.

However, as we can see with the above examples, the linear correlation coefficient and the marginal distributions alone do not uniquely determine the underlying joint distribution. This is true only if we restrict ourselves to the multivariate normal or more generally, to the elliptical distributions, as pointed out for example by Embrechts et al. [8].

3. DEPENDENT POISSON RANDOM VARIABLES

The problem of constructing a bivariate Poisson distribution with negative correlation is an old one and has been studied by various authors (here we refer to a bivariate Poisson distribution as to any distribution with univariate Poisson margins), see e.g. Griffiths et al. [10] and Nelsen [23] and further references mentioned therein. According to the statements mentioned before, the correlation coefficient for such distributions satisfies $\rho_{\mathcal{W}} \leq \rho \leq \rho_{\mathcal{M}}$, where $\rho_{\mathcal{W}} < 0 < \rho_{\mathcal{M}}$ according to well-known Results due to Hoeffding and Fréchet (see e.g. Theorem 4 in Embrechts et al. [8]). Therefore, there exist bivariate Poisson distributions that have a negative correlation.

However, probably the best known bivariate Poisson distribution (e.g. Campbell [4] and Teicher [30] as well as Johnson et al. [14] and references given therein) has been derived as a limiting case of a bivariate binomial distribution and has a joint probability generating function of the form

$$\exp\bigl(\lambda_1(t_1-1) + \lambda_2(t_2-1) + a_{12}(t_1-1)(t_2-1)\bigr).$$

This distribution can be represented in the following way (for further details on the corresponding work of Dwass and Teicher see the references in Johnson et al. [14]):

If Y_1, Y_2 and Y_{12} are independent random variables distributed according to $\mathcal{P}(\lambda_1 - a_{12})$, $\mathcal{P}(\lambda_2 - a_{12})$ and $\mathcal{P}(a_{12})$ respectively, then the random vector $(Y_1 + Y_{12}, Y_2 + Y_{12})$ has the abovementioned bivariate Poisson distribution. ■

With this interpretation it is obvious, that only positive correlation can be achieved in this way. Furthermore, it can be shown that this holds for all infinitely divisible bivariate Poisson distributions, since their joint probability generating functions are all of the above form.

Another class of Poisson distributions can be constructed by mixing the aforementioned distribution with respect to the covariance, but, although such distributions are no longer infinitely divisible, again only non-negative correlation is permissible (see Griffiths et al. [10]). For further properties of such bivariate Poisson distributions, see Johnson et al. [14] and Kocherlakota & Kocherlakota [17].

Griffiths et al. [10] present a method how to construct bivariate Poisson distributions with negative correlation from an independent pair of Poisson distributed variables by shifting the probability mass under preserving the margins. Besides this, the authors present another technique by considering the following form of the joint probabilities:

$$P[X = i, Y = j] = P[X = i] P[Y = j] \big(1 + \alpha \cdot \psi(i) \varphi(j)\big)$$

where ψ and φ are maps from $\mathbb{N}_0$ to $\mathbb{R}$ with $E[\psi(X)] = E[\varphi(Y)] = 0$. If the marginals are Poisson with parameter 1, they set

$$\psi(k) = \varphi(k) = \frac{\big(\gamma^k e^{1-\gamma} - 1\big)}{1 + e^2}$$

with $\gamma \in [-1,1]$ which yields a negatively correlated distribution whenever $\alpha < 0$. A similar idea was used by Lee [19] and Lakshminarayana [18] who choose

$$\psi(i) = e^{-i} - e^{-\lambda_1(1-1/e)} \quad \text{and} \quad \varphi(j) = e^{-j} - e^{-\lambda_2(1-1/e)}$$

and construct a bivariate Poisson distribution with $\mathcal{P}(\lambda_1)$ and $\mathcal{P}(\lambda_2)$ margins belonging to the so-called *Sarmanov* family of distributions. Again, upon setting $\alpha < 0$ a negative correlation is achieved which, however, will tend to zero for increasing values of λ_1 and λ_2. Note that choosing the *Farlie-Gumbel-Morgenstern* family of copulas (for a definition see Nelsen [23]) can be viewed as a special case of the abovementioned construction by setting $\psi(i) = (1 - 2F(i))$ and $\varphi(j) = (1 - 2G(j))$.

In this paper we shall follow another scheme, however. According to the abovementioned theorem of Sklar, arbitrary multivariate Poisson distributions can be obtained from copulas by choosing univariate Poisson margins, as already pointed out e.g. by Joe [13]. Such models may not possess a stochastic interpretation, but can cover a wide range of dependence, including the desired negative correlation. Furthermore, many parametric families of copulas can be easily simulated.

The first case, where the copula approach yields an elegant solution, is a bivariate Poisson distribution H with minimum correlation as noted already by Griffiths et al. [10] as well as Nelsen [23]. As mentioned before, the correlation coefficient achieves its minimum when choosing the Fréchet-Hoeffding lower bound $\mathcal{W}$ as the underlying copula. Therefore, if we denote by F and G the distribution functions of $\mathcal{P}(\lambda_1)$ and $\mathcal{P}(\lambda_2)$ respectively and by $h(i,j)$ the probabilities assigned to the grid points $\{(i,j)\}$ (or, likewise, to the intervals $(i-1,i] \times (j-1,j]$), we get, upon setting

$$\begin{aligned} h(i,j) = {} & \mathcal{W}(F(i),G(j)) - \mathcal{W}(F(i),G(j-1)) - \mathcal{W}(F(i-1),G(j)) + \mathcal{W}(F(i-1),G(j-1)) \\ = {} & \max(F(i) + G(j) - 1,0) - \max(F(i) + G(j-1) - 1,0) \qquad (*) \\ & - \max(F(i-1) + G(j) - 1,0) + \max(F(i-1) + G(j-1) - 1,0), \end{aligned}$$

a bivariate Poisson distribution with $\mathcal{P}(\lambda_1)$ and $\mathcal{P}(\lambda_2)$ margins and the minimum possible (negative) correlation.

From this it can readily be deduced that at most a finite number of the probabilities $h(i,j)$ will have a nonzero value for $i,j > 0$. This follows from the fact that for i,j large enough we have

$$h(i,j) = \left(F(i) + G(j) - 1\right) - \left(F(i) + G(j-1) - 1\right) - \left(F(i-1) + G(j) - 1\right)$$
$$+ \left(F(i-1) + G(j-1) - 1\right) = 0.$$

Unfortunately, the structure of the above probability density function h will depend on the parameters λ_1 and λ_2 in a rather complex way, so that no close formula is obtained. Generally we can say that as the values of the parameters λ_1 and λ_2 increase, the number of cells $(i-1,i] \times (j-1,j])$ with $i,j > 0$ and nonzero probabilities rise and their positions move away from $(0,0)$.

This procedure can alternatively be visualized by a two-dimensional analogue for the generation of random variables from a discrete distribution, using the lower Fréchet-Hoeffding bound in terms of a pair $(U, 1-U)$ where U is a uniformly $\mathcal{R}[0,1]$-distributed random variable.

If we subdivide the x- and y-axis in the unit square by the values $F(k)$ and $G(k)$ for $k = 0, 1, 2, \ldots$ each where F and G as above denote the univariate cumulative distribution functions of a bivariate Poisson distribution with $\mathcal{P}(\lambda_1)$ and $\mathcal{P}(\lambda_2)$ margins, then the unit square is divided into a countably infinite collection of rectangles R_{ij} given by $R_{ij} := (F(i-1), F(i)] \times (G(j-1), G(j)]$ for $i,j > 0$. Generating a pair $(U, 1-U)$ then produces a random "point" in the unit square which with probability one hits exactly one of these rectangles R_{mn}, whose right upper corner coordinates $F(m)$ and $G(n)$ determine the values of the discrete random variables $X = m$ and $Y = n$ which are, by construction, $\mathcal{P}(\lambda_1)$ and $\mathcal{P}(\lambda_2)$-distributed, respectively and have minimal (negative) correlation. The following exemplifying graphs show how this works in the situation of a pair (X, Y) with $\mathcal{P}(1)$ and $\mathcal{P}(2)$ margins (left) and $\mathcal{P}(3)$ and $\mathcal{P}(5)$ margins (right).
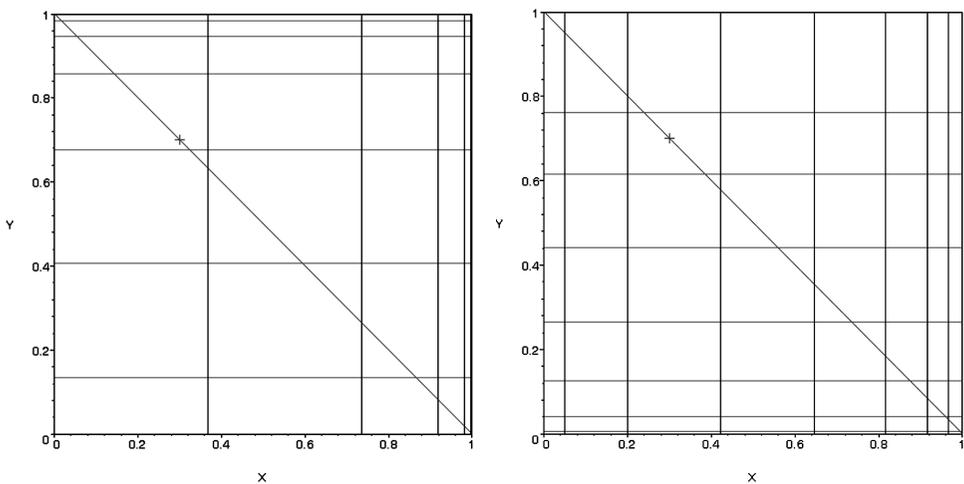

FIGURE 1.

The vertical lines here correspond to the subdivision of the *x*-axis according to *F*, the horizontal lines to the subdivision of the *y*-axis according to *G*. In the above example, we have $U = 0.3$, resulting in the simulated pair $(X, Y) = (0,3)$ [left] and $(X, Y) = (2,6)$ [right].
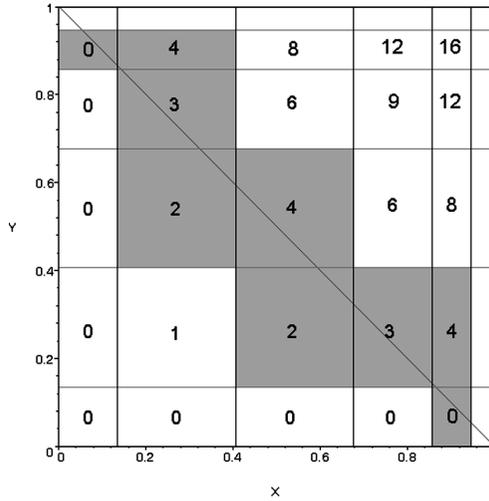

FIGURE 2.

The case of $\lambda_1 = \lambda_2 (= \lambda$, say) is of particular interest here because the calculation of the corresponding minimal correlation $\rho(\lambda)$ simplifies a little. Fig. 2 shows the situation for $\lambda = 2$; the numbers in the rectangles correspond to the values of $X \cdot Y$. The shaded rectangles denote those areas which are involved in the calculation of the minimal (negative) correlation (cf. also Nelsen [23] for some corresponding tabulated values).

Note that the minimal possible correlation here does not attain the value of -1 for any choice of λ_1 and λ_2, nor is it equal to another (negative) constant, but it in general depends both on λ_1 and λ_2. Nelsen [23] and Griffiths et al [10] use the following expression for the correlation coefficient (which immediately follows from Theorem 2.5):

$$\rho_{\mathcal{W}} = \frac{\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}P[X > i, Y > j] - \lambda_1\lambda_2}{\sqrt{\lambda_1\lambda_2}} = \frac{-\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\min(F(i) + G(j) - 1, 0) - \lambda_1\lambda_2}{\sqrt{\lambda_1\lambda_2}}$$

$$= \frac{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{u=i}^{\infty}\sum_{v=j}^{\infty}h(u, v) - \lambda_1\lambda_2}{\sqrt{\lambda_1\lambda_2}}.$$

Again, the above expression depends on the parameters λ_1 and λ_2 in a complex way and cannot easily be evaluated explicitly, in general.

However, as we will see in the sequel, both the probability function as well as the correlation can be computed using a (polynomial time) greedy algorithm.

This is due to the fact that the problem of constructing a bivariate discrete distribution with minimum correlation can be viewed as a basic transportation problem. Such problems, which are well-known in Operations Research (see e.g. Gass [9]), are of the following form:

$$Minimize \sum_{i=0}^{m} \sum_{j=0}^{n} c_{ij} x_{ij} \ under \ the \ conditions \ \sum_{i=0}^{m} x_{ij} = b_j, \ \sum_{j=0}^{n} x_{ij} = a_i \ and \ x_{ij} \geq 0$$

where a_i, b_j and c_{ij}, $i = 0, \cdots, m$ and $j = 0, \cdots, n$ are given non-negative real numbers with $m, n \in \mathbb{N}$ satisfying the side conditions $\sum_{i=0}^{m} a_i = \sum_{j=0}^{n} b_j$ and $c_{ij} + c_{k\ell} \geq c_{i\ell} + c_{kj}$ whenever $i \geq k$ and $j \geq \ell$, for all $i, k = 0, \cdots, m$ and $j, \ell = 0, \cdots, n$. If we interpret a_i as an amount of products manufactured in place i and b_j as an amount of products needed in place j, and the c_{ij} as the cost for shipping one product unit from i to j, then the x_{ij} solving the above minimization problem can be viewed as the corresponding optimal shipping strategy.

The algorithm which provides the optimal solution, known as the *north west corner rule*, is due to Hoffman [12]. However, the essential idea was first noticed by Monge in 1781 already. In our case, the x_{ij} correspond to the joint probabilities $h(i, j)$ defined before, and the "costs" are given by $c_{ij} = ij$. Since only a finite number of the probabilities $h(i, j)$ has a nonzero value for $i, j > 0$, the above problem of minimizing the correlation can be formulated in the following way:

NWC-Algorithm:

1. Find m and n such that
$$G(0) \geq 1 - F(m)$$
$$F(0) \geq 1 - G(n);$$

2. Set $h(i, j) = 0$ for $i > m$, $j > 0$ and for $i > 0$, $j > n$;

3. Set $h(i, 0) = \dfrac{e^{-\lambda_1}}{i!} \lambda_1^i$ for $i > m$ and $h(0, j) = \dfrac{e^{-\lambda_2}}{j!} \lambda_2^j$ for $j > n$;

4. Calculate the probabilities $h(i, j)$ as solutions of the transportation problem

$$Minimize \sum_{i=1}^{m} \sum_{j=1}^{n} ij \ h(i, j) \ under \ the \ conditions$$

$$\sum_{j=0}^{n} h(0, j) = F(0) + G(n) - 1,$$

$$\sum_{j=0}^{n} h(i, j) = \frac{e^{-\lambda_1}}{i!} \lambda_1^i, \quad i = 1, \ldots, n,$$

$$\sum_{i=0}^{m} h(i, 0) = G(0) + F(m) - 1,$$

$$\sum_{i=0}^{m} h(i, j) = \frac{e^{-\lambda_2}}{j!} \lambda_2^j, \quad j = 1, \ldots, m,$$

using the *north west corner rule*.

Note that for each selection of λ_1 and λ_2 we have to solve a different optimization problem. In particular, the numbers m and n depend on these parameters. Also, the optimal solution found by the north west corner rule corresponds exactly to (*) above. For further details and proofs, see Neslehova [24].

It is worth mentioning that the optimality of the north west corner rule algorithm is based on the same idea of shifting mass away from the main diagonal used by Griffiths et al. [10] in the sense that this algorithm represents the maximum possible shift.

In what follows we shall show how the above NWC algorithm based on the north west corner rule operates in situation of $\lambda_1 = \lambda_2 = \lambda$ say. For this purpose, we shall consecutively determine the various possible values for n in the corresponding first step above. For further details, see Neslehova [24].

Case $n = 0$:

Solve $F(0) = 1 - F(0)$ or, equivalently, $1 - 2e^{-\lambda} = 0$ giving $\lambda = \lambda_1 := \ln 2 = 0.6931$. Then according to step 1 of the NWC-algorithm, $F(0) \geq 1 - F(n)$ for $\lambda \in [\lambda_0, \lambda_1]$ with $\lambda_0 := 0$. Further, $h(i,j) = 0$ for $i > 0, j > 0$ and $h(0,i) = h(i,0) \, e^{-\lambda} \dfrac{\lambda^i}{i!}$ for $i > 0$, according to steps 2 and 3 of the NWC-algorithm. Note that in the particular case here, according to step 4 of the NWC-algorithm, no explicit minimization problem has to be solved; instead, we have directly

$$\rho(\lambda) = \rho_1(\lambda) := \frac{\sum\limits_{i=1}^{\infty} \sum\limits_{j=1}^{\infty} \sum\limits_{u=i}^{\infty} \sum\limits_{v=j}^{\infty} h(u,v) - \lambda^2}{\lambda} = \frac{0}{\lambda} - \lambda = -\lambda.$$

Case $n = 1$:

Solve $F(0) = 1 - F(1)$ or, equivalently, $1 - (2 + \lambda)e^{-\lambda} = 0$ giving the numerical solution $\lambda = \lambda_2 := 1.1461$. Then according to step 1 of the NWC-algorithm, $F(0) \geq 1 - F(n)$ for $\lambda \in [\lambda_1, \lambda_2]$. Further, $h(i,j) = 0$ for $i > 1, j > 0$ or $i > 0, j > 1$ and $h(0,i) = h(i,0) = e^{-\lambda} \dfrac{\lambda^i}{i!}$ for $i > 1$, according to steps 2 and 3 of the NWC-algorithm. It remains to solve

Minimize $h(1,1)$ under the conditions

$$h(0,0) + h(0,1) = p(0) \qquad h(0,0) + h(1,0) = q(0)$$
$$h(1,0) + h(1,1) = p(1) \qquad h(0,1) + h(1,1) = q(1)$$

where $p(0) = q(0) = F(0) + F(1) - 1 = e^{-\lambda}(2 + \lambda) - 1$ and $p(1) = q(1) = F(1) - F(0) = \lambda e^{-\lambda}$.

The solution is

$$h(0,0) = 0 \qquad h(0,1) = h(1,0) = e^{-\lambda}(2 + \lambda) - 1 \qquad h(1,1) = 1 - 2e^{-\lambda},$$

such that

$$\rho(\lambda) = \rho_2(\lambda) := \frac{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{u=i}^{\infty}\sum_{v=j}^{\infty} h(u,v) - \lambda^2}{\lambda} = \frac{h(1,1)}{\lambda} - \lambda = \frac{1 - 2e^{-\lambda}}{\lambda} - \lambda.$$

Case $n = 2$:

Solve $F(0) = 1 - F(2)$ or, equivalently, $2 - (4 + 2\lambda + \lambda^2)e^{-\lambda} = 0$ giving the numerical solution $\lambda = \lambda_3 := 1.5681$. Then according to step 1 of the NWC-algorithm, $F(0) \geq 1 - F(n)$ for $\lambda \in [\lambda_2, \lambda_3]$. Further, $h(i,j) = 0$ for $i > 2, j > 0$ or $i > 0, j > 2$ and $h(0,i) = h(i,0) = e^{-\lambda}\frac{\lambda^i}{i!}$ for $i > 2$, according to steps 2 and 3 of the NWC-algorithm. It remains to solve

Minimize $h(1,1) + 2h(1,2) + 2h(2,1) + 4h(2,2)$ under the conditions

$$h(0,0) + h(0,1) + h(0,2) = p(0) \qquad h(0,0) + h(1,0) + h(2,0) = q(0)$$
$$h(1,0) + h(1,1) + h(1,2) = p(1) \qquad h(0,1) + h(1,1) + h(2,1) = q(1)$$
$$h(2,0) + h(2,1) + h(2,2) = p(2) \qquad h(0,2) + h(1,2) + h(2,2) = q(2)$$

where

$$p(0) = q(0) = F(0) + F(2) - 1 = e^{-\lambda}\left(2 + \lambda + \frac{\lambda^2}{2}\right) - 1 \text{ and}$$

$$p(1) = q(1) = F(1) - F(0) = \lambda e^{-\lambda}, \qquad p(2) = q(2) = F(2) - F(1) = \frac{\lambda^2}{2}e^{-\lambda}.$$

The solution is

$$h(0,0) = h(0,1) = h(1,0) = 0 \qquad\qquad h(1,1) = 2e^{-\lambda}(1 + \lambda) - 1$$
$$h(0,2) = h(2,0) = e^{-\lambda}\left(2 + \lambda + \frac{\lambda^2}{2}\right) - 1$$
$$h(1,2) = h(2,1) = 1 - e^{-\lambda}(2 + \lambda) \qquad h(2,2) = 0$$

such that

$$\rho(\lambda) = \rho_3(\lambda) := \frac{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{u=i}^{\infty}\sum_{v=j}^{\infty} h(u,v) - \lambda^2}{\lambda} =$$

$$= \frac{h(1,1) + 2h(1,2) + 2h(2,1) + 4h(2,2)}{\lambda} - \lambda = \frac{3 - (6 + 2\lambda)e^{-\lambda}}{\lambda} - \lambda.$$

The following table contains an evaluation for $\rho(\lambda)$ in functional form, summarizing and continuing the results from the approach above. $\rho(\lambda)$ is given piecewise as $\rho_n(\lambda)$ in the interval $[\lambda_{n-1}, \lambda_n]$ for $n \geq 1$, where λ_n is the positive

solution of the equation $\rho_{n+1}(\lambda) - \rho_n(\lambda) = 0$. For a discussion of the existence of appropriate solutions, see Neslehova [24], Lemma 8.2.1.

n	$\rho_n(\lambda)$	λ_n	is the positive solution of
1	$-\lambda$	0.6931	$1 - 2e^{-\lambda} = 0$
2	$\dfrac{1 - 2e^{-\lambda}}{\lambda} - \lambda$	1.1461	$1 - (2 + \lambda)\,e^{-\lambda} = 0$
3	$\dfrac{3 - (6 + 2\lambda)\,e^{-\lambda}}{\lambda} - \lambda$	1.5681	$2 - \left(4 + 2\lambda + \lambda^2\right)e^{-\lambda} = 0$
4	$\dfrac{5 - \left(10 + 4\lambda + \lambda^2\right)e^{-\lambda}}{\lambda} - \lambda$	1.6783	$1 - (2 + 2\lambda)\,e^{-\lambda} = 0$
5	$\dfrac{6 - \left(12 + 6\lambda + \lambda^2\right)e^{-\lambda}}{\lambda} - \lambda$	1.9761	$2 - \left(4 + 2\lambda + \lambda^2 + \frac{1}{3}\lambda^3\right)e^{-\lambda} = 0$
6	$\dfrac{8 - \left(16 + 8\lambda + 2\lambda^2 + \frac{1}{3}\lambda^3\right)e^{-\lambda}}{\lambda} - \lambda$	2.1559	$2 - \left(4 + 4\lambda + \lambda^2\right)e^{-\lambda} = 0$
7	$\dfrac{10 - \left(20 + 12\lambda + 3\lambda^2 + \frac{1}{3}\lambda^3\right)e^{-\lambda}}{\lambda} - \lambda$	2.3761	$2 - \left(4 + 2\lambda + \lambda^2 + \frac{1}{3}\lambda^3 + \frac{1}{12}\lambda^4\right)e^{-\lambda} = 0$
8	$\dfrac{12 - \left(24 + 14\lambda + 4\lambda^2 + \frac{2}{3}\lambda^3 + \frac{1}{12}\lambda^4\right)e^{-\lambda}}{\lambda} - \lambda$	2.6082	$2 - \left(4 + 4\lambda + \lambda^2 + \frac{1}{3}\lambda^3\right)e^{-\lambda} = 0$

Surprisingly, the correlation $\rho(\lambda)$ is not a decreasing function of λ; however, as expected, we have $\lim_{\lambda \to \infty} \rho(\lambda)$ which will be proved in the subsequent Theorem 3.1. In Fig. 3, a plot of $\rho_n(\lambda)$ in the intervals $[\lambda_{n-1}, \lambda_n]$ for n between 1 and 7 is shown. Note that such a plot has been provided already by Griffiths et al. [10], however on a purely numerical basis only, except for the first interval.

Theorem 3.1. Let X_λ and Y_λ be Poisson distributed random variables with parameter $\lambda > 0$ each such that their correlation $\rho(\lambda) = \rho(X_\lambda, Y_\lambda)$ is minimal. Then

$$\lim_{\lambda \to \infty} \rho(\lambda) = -1.$$

Proof: Let F_λ denote the cumulative distribution function of X_λ and Y_λ, resp. By Skorohod's Embedding Theorem (see e.g. Billingsley [1], Theorem 25.6), we can assume that $X_\lambda := F_\lambda^{-1}(U)$ and $Y_\lambda := F_\lambda^{-1}(1 - U)$ with some random variable U being uniformly distributed over [0,1] where F_λ^{-1} denotes the pseudo-inverse
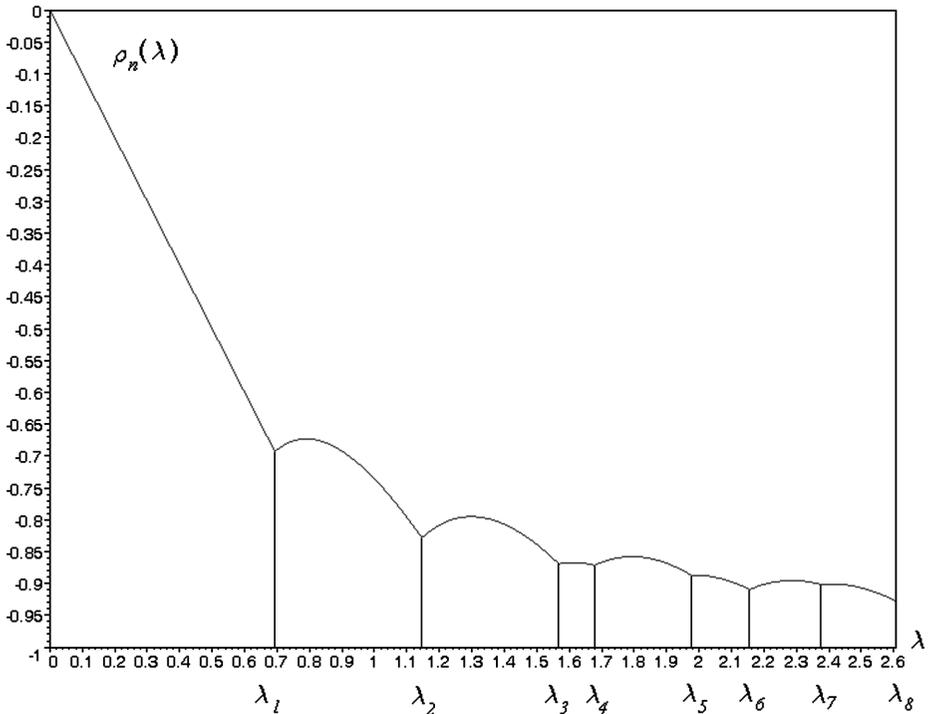
FIGURE 3.

of F_λ. Note that this particular choice also induces the minimal correlation $\rho(\lambda) = \rho(X_\lambda, Y_\lambda)$ according to what has been said above. Since

$$E\left(\frac{X_\lambda - \lambda}{\sqrt{\lambda}}\right)^4 = 3 + \frac{1}{\lambda} \leq 4 \text{ for } \lambda \geq 1,$$

an application of Markov's Inequality shows that we have, for $\lambda \geq 1$,

$$G_\lambda(x) := F_\lambda\left(x\sqrt{\lambda} + \lambda\right) = P\left(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \leq x\right) \geq \max\left\{0; 1 - \frac{4}{x^4}\right\} =: H_u(x), \; x > 0 \text{ and}$$

$$G_\lambda(x) = P\left(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \leq x\right) \leq \max\left\{1; \frac{4}{x^4}\right\} =: H_o(x), \; x < 0$$

such that

$$\left|G_\lambda^{-1}(u)\right| = \left|\frac{F_\lambda^{-1}(u) - \lambda}{\sqrt{\lambda}}\right| \leq \sqrt{2} \max\left(\frac{1}{\sqrt[4]{u}}, \frac{1}{\sqrt[4]{1-u}}\right) =: B(u) \text{ for } 0 < u < 1$$

whenever $\lambda \geq 1$.

Fig. 4 below visualizes these inequalities for the special choice $\lambda = 5$.
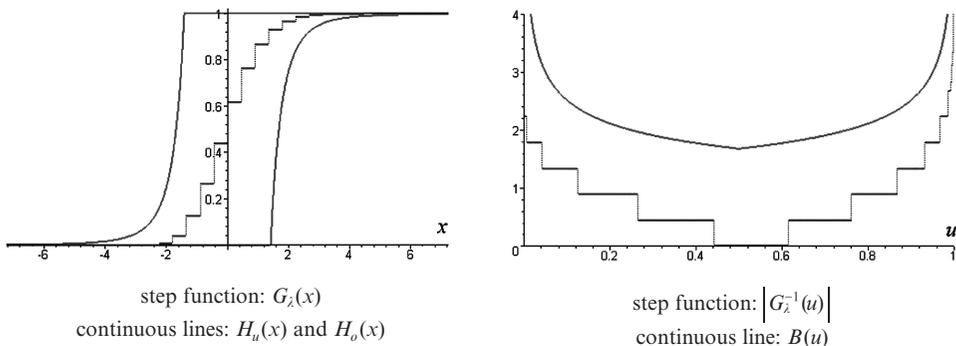

step function: $G_\lambda(x)$
continuous lines: $H_u(x)$ and $H_o(x)$

step function: $\left|G_\lambda^{-1}(u)\right|$
continuous line: $B(u)$

FIGURE 4.

It follows that

$$\left|\frac{X_\lambda - \lambda}{\sqrt{\lambda}}\right| \cdot \left|\frac{Y_\lambda - \lambda}{\sqrt{\lambda}}\right| \leq 2\max\left(\frac{1}{\sqrt{U}}, \frac{1}{\sqrt{1-U}}\right) =: W \text{ a.s., say,}$$

with $E(W) < \infty$. Now, by the Central Limit Theorem and Skorohod's Embedding Theorem again,

$$\lim_{\lambda \to \infty} \frac{X_\lambda - \lambda}{\sqrt{\lambda}} = \Phi^{-1}(U) = Z, \quad \lim_{\lambda \to \infty} \frac{Y_\lambda - \lambda}{\sqrt{\lambda}} = \Phi^{-1}(1-U) = 1 - Z \text{ a.s.}$$

where $Z := \Phi^{-1}(U)$ has a standard normal distribution. By Lebesgue's Dominated Convergence Theorem we thus find that

$$\lim_{\lambda \to \infty} \rho(\lambda) = \lim_{\lambda \to \infty} E\left(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \cdot \frac{Y_\lambda - \lambda}{\sqrt{\lambda}}\right) = E(Z \cdot (1 - Z)) = E(Z) - E(Z^2) = -1, \text{ a.s.}$$

which was to be proved. ∎

To complete the mathematical analysis made so far, we shall give a simple elementary proof for the minimality of the functions $\rho_n(\lambda)$ for $n = 1, 2, 3$ here. Note that this is equivalent to prove that

$$E(XY) \geq \begin{cases} 0, & 0 < \lambda \leq \lambda_1 \\ 1 - 2e^{-\lambda}, & \lambda_1 \leq \lambda \leq \lambda_2 \\ 3 - (6 + 2\lambda)e^{-\lambda}, & \lambda_2 \leq \lambda \leq \lambda_3. \end{cases}$$

Case 1: $0 < \lambda \leq \lambda_1$: trivial

<u>Case 2:</u> $\lambda_1 \leq \lambda \leq \lambda_2$: Let $A_0 := \{X = 0\}$, $B_0 := \{Y = 0\}$, then $P(A_0) = P(B_0) = e^{-\lambda}$, and hence, denoting by $\mathbf{1}_A$ the indicator random variable for the event A, since $XY \geq \mathbf{1}_{(A_0 \cup B_0)^c}$,

$$E(XY) \geq E(\mathbf{1}_{(A_0 \cup B_0)^c}) = P(A_0^c \cap B_0^c) = 1 - P(A_0 \cup B_0) \geq 1 - P(A_0) - P(B_0)$$
$$= 1 - 2e^{-\lambda}.$$

<u>Case 3:</u> $\lambda_2 \leq \lambda \leq \lambda_3$: $XY \geq \mathbf{1}_{(A_0 \cup B_0)^c} + \mathbf{1}_{(A_0 \cup A_1)^c \cap B_0^c} + \mathbf{1}_{(B_0 \cup B_1)^c \cap A_0^c}$ for $A_1 := \{X = 1\}$, $B_1 := \{Y = 1\}$, and hence

$$E(XY) \geq P\big((A_0 \cup B_0)^c\big) + P\big((A_0 \cup A_1)^c \cap B_0^c\big) + P\big((B_0 \cup B_1)^c \cap A_0^c\big)$$
$$= 1 - P(A_0 \cup B_0) + 1 - P(A_0 \cup A_1 \cup B_0) + 1 - P(B_0 \cup B_1 \cup A_0)$$
$$\geq 3 - 3P(A_0) - 3P(B_0) - P(A_1) - P(B_1) = 3 - 6e^{-\lambda} - 2\lambda e^{-\lambda},$$

as stated. ■

The case of different parameters λ_1 and λ_2 can in principle be tackled the same way, although more subtle cases have to be distinguished. For simplicity, we shall outline the analogues of cases 1 and 2 only, i.e.

$$E(XY) \geq \begin{cases} 0, & e^{-\lambda_1} + e^{-\lambda_2} \geq 1 \\ 1 - e^{-\lambda_1} - e^{-\lambda_2}, & e^{-\lambda_1} + e^{-\lambda_2} \leq 1 \text{ and} \\ & \min\{e^{-\lambda_1} + (1 + \lambda_2)e^{-\lambda_2}, e^{-\lambda_2} + (1 + \lambda_1)e^{-\lambda_1}\} \geq 1 \end{cases}$$

which follows immediately along the lines of proof above, or, in terms of the correlation $\rho(\lambda_1, \lambda_2)$:

$$\rho(\lambda_1, \lambda_2) \geq \begin{cases} -\sqrt{\lambda_1 \lambda_2}, & e^{-\lambda_1} + e^{-\lambda_2} \geq 1 \\ \dfrac{1 - e^{-\lambda_1} - e^{-\lambda_2}}{\sqrt{\lambda_1 \lambda_2}} - \sqrt{\lambda_1 \lambda_2}, & e^{-\lambda_1} + e^{-\lambda_2} \leq 1 \text{ and} \\ & \min\{e^{-\lambda_1} + (1 + \lambda_2)e^{-\lambda_2}, e^{-\lambda_2} + (1 + \lambda_1)e^{-\lambda_1}\} \geq 1 \end{cases}$$

which provides indeed the lower attainable bounds for the correlation in the range specified (see also Griffiths et al. [10] for tabulated values). Figure 5 shows a 3D-plot of this function.

   In a similar way, by choosing various other families of copulas, arbitrary joint Poisson distributions can be constructed. For instance, if we choose the Frank or the Plackett family introduced in the previous chapter, we get a one-parameter family of joint Poisson distributions, which are negatively correlated for a proper choice of the parameter. Moreover, the minimum possible negative
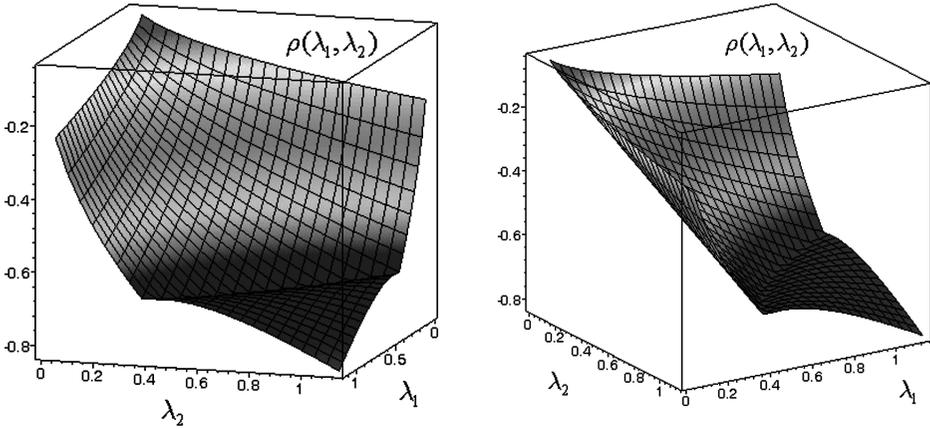
FIGURE 5.

correlation can be achieved as a limiting case. Figure 6 shows a simulation of 100 points distributed according to a Frank copula with $\theta = -10$, together with the grid generated by the distribution of a pair (X,Y) with $\mathcal{P}(1)$ and $\mathcal{P}(2)$ margins as in Fig. 1 [left].
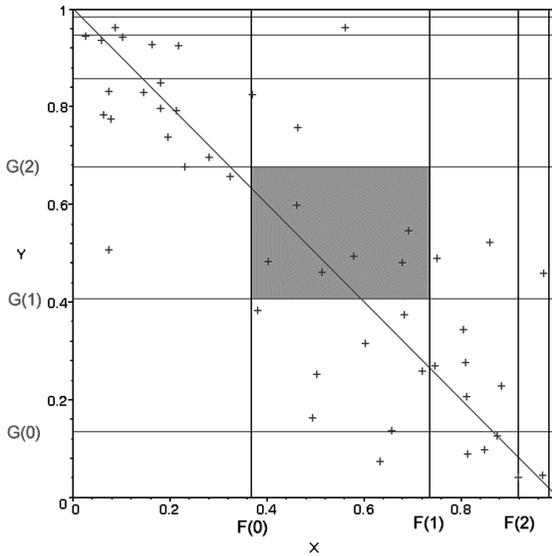

FIGURE 6.

For instance, all simulated random points falling into the shaded rectangle generate the (same) pair $(1,2)$.

Naturally, not only *negatively* correlated Poisson random variables can be constructed or simulated this way. Choosing the appropriate copulas (e.g. the Frank family with θ being arbitrary), all possible dependencies and correlations

can be achieved, even in a multi-dimensional setting. For the problem of *positive* correlation with different λ_1 and λ_2 (the only non-trivial case), see e.g. Nelsen [23] or Griffiths et al. [10].

The case of minimal (negative) correlation for identical parameters $\lambda_1 = \lambda_2$ ($= \lambda$, say) also plays a general role for the construction of negatively (and even positively) correlated Poisson distributed random variables with non-identical parameters. Suppose, for instance, that (X, Y) is a Poisson distributed pair with arbitrary dependence structure and margins $\mathcal{P}(\lambda)$ each. Let further V and W be Poisson distributed random variables, independent of each other and of (X, Y) with parameters μ and ν, respectively. Then the random variables $X + V$ and $Y + W$ are also Poisson distributed, with parameters $\lambda + \mu$ and $\lambda + \nu$ respectively, and negative correlation

$$\rho(X + V, Y + W) = \frac{\lambda \rho(X, Y)}{\sqrt{(\lambda + \mu)(\lambda + \nu)}} < 0.$$

If Z is another Poisson distributed random variable with parameter τ, independent of X, Y, V and W, then

$$\rho(X + V + Z, \ Y + W + Z) = \frac{\lambda \rho(X, Y) + \tau}{\sqrt{(\lambda + \mu + \tau)(\lambda + \nu + \tau)}}$$

which can even achieve positive values. This shows again that there are many ways to generate Poisson distributed random variables with the same parameters and the same correlation, but with different joint distributions. Likewise, if the pair (V, W) has itself a positive correlation $\rho = \rho(V, W) = -\rho(X, Y)$ and we have $\mu = \nu = \lambda$, then $X + V$ and $Y + W$ are indeed uncorrelated, but not independent. Note that such a pair (V, W) can easily be generated through the choice $V = S + T$, $W = S + U$ with independent S, T, U being Poisson distributed with parameters

$$\lambda_S = \rho\lambda, \ \lambda_T = \lambda_U = (1 - \rho)\lambda.$$

Hence there is obviously a continuum of possibilities to construct uncorrelated, but not independent Poisson pairs $X + V$ and $Y + W$ with the same marginal Poisson distribution.

The abovementioned negatively correlated Poisson distributions play an important role when constructing negatively correlated Poisson processes, as we shall see in the following section.

4. Dependent Poisson processes

A very elegant way to describe stochastic counting and – more generally – risk processes especially in insurance and finance is via the point process approach. There is a vast literature on this topic, see e.g. Daley and Vere-Jones [6], Kingman [16] or Reiss [25], to mention some monograph references. For a recent

contribution to this topic, see e.g. Rodrigues et al. [26]. Since in actuarial or financial applications, the assumption of a finite number of "events" (comprising both the temporal as well as the claims aspect) is sufficient for modeling in most cases, we will restrict ourselves to this situation. It is, however, easy to extend all considerations made here to infinitely many "events" by superposing an infinite family of point processes with a finite number of events each (particularly with the temporal component being dispersed in pairwise disjoint "time" intervals).

Definition 4.1. Let N be a non-negative integer-valued random variable and $\{\mathbf{X}_n\}_{n \in \mathbb{N}}$ be a familiy of i.i.d. random vectors with values in $\mathbb{R}^d$, independent of N, for some fixed dimension $d \in \mathbb{N}$. Then the random measure

$$\xi := \sum_{n=1}^{N} \varepsilon_{\mathbf{X}_n}$$

is called a (finite) *point process* with counting variable N and multiple event points $\{\mathbf{X}_n\}_{n \in \mathbb{N}}$. Here ε_x denotes the Dirac measure concentrated in the point $x \in \mathbb{R}^d$, i.e. we have

$$\varepsilon_x(A) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases} \text{ for all Borel sets } A \subseteq \mathbb{R}^d.$$

The (possibly infinite) measure $E\xi$, given by

$$E\xi(A) = E[\xi(A)] = E(N)Q(A) \text{ for all Borel sets } A \subseteq \mathbb{R}^d,$$

is called the *intensity measure* associated with ξ, where Q denotes the distribution of the event points $\{\mathbf{X}_n\}_{n \in \mathbb{N}}$.

The point process ξ is called *locally homogeneous* on some interval $I \in \mathbb{R}^d$ if the restriction of the intensity measure $E\xi$ to I is a (finite) multiple of the Lebesgue measure m^d over I or equivalently, if the conditional distribution $Q(\bullet|I)$ of an event point given it falls into I is a continuous uniform distribution over I, $\mathcal{R}(I)$. ■

The usual (time-oriented) representation as a counting process $\{N(t) | t \in I\}$ where $I = [0, T]$, say, for some $T > 0$, is obtained for $d = 1$ via

$$N(t) = \xi((0, t]) = \#\{n \in \mathbb{N} \mid 0 < X_n \leq t\}, \ 0 \leq t \leq T,$$

or, more generally, $\xi(A) = \#\{n \in \mathbb{N} \mid X_n \in A\}$ denotes the number of "events" generated by the $\{X_n\}_{n \in \mathbb{N}}$ within the "time set" $A \subseteq I$. The advantage of the measure theoretic approach, however, clearly lies in the fact that distributional properties of point processes and arithmetic and set theoretic operations on them can be described more simply. Note that no ordering assumptions on the event points $\{X_n\}_{n \in \mathbb{N}}$ are necessary, which makes it easy to generalize this theory to more than one dimension as is suggested by the above definition (for instance,

if occurrences of more than one type of claims have to be considered, say e.g. from windstorm, hailstorm, flooding, avalanches etc.). Such topics have already been discussed extensively in Cox and Lewis [5], however in the more "traditional" framework, a more modern approach is to be found in Griffiths et al. [10]; see also Rolski et al. [27]. We shall see in the subsequent chapter that more generally, also the standard model of *collective risk theory* is covered by this approach.

Definition 4.2. A point process ξ is called a finite *Poisson point process*, if the counting variable N is Poisson-distributed with mean $\lambda > 0$, i.e. $P^N = \mathcal{P}(\lambda)$. ∎

Note that for $d = 1$ and $I = [0, T]$ a locally homogeneous Poisson point processes ξ in our sense models exactly the first (random) number of "arrivals" of an ordinary time-homogeneous Poisson counting process with intensity λ which occur within the "time interval" $[0, T]$. If I is any other interval $I = [S, T]$ with $0 < S < T$ then ξ can be considered as a stochastic "cut" of the time window I from such a Poisson (counting) process with intensity λ.
　　The following theorems list some of the most important properties of Poisson point processes, which can e.g. be found in the references listed above.

Theorem 4.3. Let ξ_1 and ξ_2 be independent Poisson point processes with intensity measures $E\xi_1$ and $E\xi_2$, resp. Then $\xi = \xi_1 + \xi_2$ also is a Poisson point process with intensity measure $E\xi = E\xi_1 + E\xi_2$. In particular, if ξ_1 and ξ_2 are locally homogeneous on the same interval I, then $\xi = \xi_1 + \xi_2$ also is locally homogeneous on I. $\xi = \xi_1 + \xi_2$ is also called the *superposition* of ξ_1 and ξ_2. ∎

The proof of this theorem relies on the fact that the sum of two independent Poisson random variables is again a Poisson random variable. Note that in case of finite intensity measures, the new "event points" from the sum process ξ follow a distribution Q which can be expressed as a mixture of the original distributions Q_1 and Q_2 of the processes ξ_1 and ξ_2 through

$$Q = \frac{E(N_1)}{E(N_1) + E(N_2)} Q_1 + \frac{E(N_2)}{E(N_1) + E(N_2)} Q_2$$

where N_1 and N_2 denote the corresponding counting variables of ξ_1 and ξ_2.

Theorem 4.3 can easily be extended to an infinite superposition of Poisson point processes, with a proper limit if the resulting intensity measure is σ-finite. For instance, if $\{\xi_n\}_{n \in \mathbb{N}}$ is an independent family of Poisson point processes with N_n following a Poisson distribution with mean $\lambda > 0$ each and Q_n being a continuous uniform distribution over the interval $(n-1, n]$ for $n \in \mathbb{N}$, then $\xi = \sum_{n=1}^{\infty} \xi_n$ is some (infinite) point process (in a wider sense) which corresponds to the usual time-homogeneous Poisson counting process with intensity $\lambda > 0$ (cf. also Kingman [16], p. 16).

Theorem 4.4. Let $\xi = \sum_{n=1}^{N} \varepsilon_{\mathbf{X}_n}$ be some (finite) Poisson point process and $\{J_n\}_{n \in \mathbb{N}}$ be an i.i.d. family of $\{0,1\}$-valued random variables, with $P(J_n = 1) = p \in [0,1]$. Define the point processes ξ_1 and ξ_2 by

$$\xi_1 := \sum_{n=1}^{N} J_n \varepsilon_{\mathbf{X}_n}, \quad \xi_2 := \sum_{n=1}^{N} (1 - J_n) \varepsilon_{\mathbf{X}_n}.$$

Then we have $\xi = \xi_1 + \xi_2$ with ξ_1 and ξ_2 being independent Poisson point processes each with intensity measures $E\xi_1 = p \cdot E\xi$ and $E\xi_2 = (1-p) \cdot E\xi$. ■

We say that ξ_1 and ξ_2 are independent *decompositions* of ξ; ξ_1 and ξ_2 themselves are also called *thinnings* of ξ. W.r.t. simulations of such processes, ξ_1 and ξ_2 can be constructed from the starting Poisson point process ξ by independent marking of the event points (with marks "1" and "0" by a random number generator representing the J-sequence) and group the corresponding event points with identical marks.

Theorem 4.5. Let $\xi = \sum_{n=1}^{N} \varepsilon_{\mathbf{X}_n}$ be some finite Poisson point process with d-dimensional event points $\mathbf{X}_n = (X_n(1), \cdots, X_n(d))$. Then each of the point processes

$$\xi(k) = \sum_{n=1}^{N} \varepsilon_{X_n(k)}, \quad k = 1, \cdots, d$$

is a one-dimensional Poisson point process, with intensity measure $E\xi(k)$ given by

$$E\xi(k)(A) = E\xi\big(\mathbb{R}^{k-1} \times A \times \mathbb{R}^{d-k}\big) \text{ for all Borel sets } A \subseteq \mathbb{R}.$$

Conversely, if $\xi(k) = \sum_{n=1}^{N} \varepsilon_{X_n(k)}, k = 1, \cdots, d$ are one-dimensional Poisson point processes, then $\xi = \sum_{n=1}^{N} \varepsilon_{\mathbf{X}_n}$ with $\mathbf{X}_n = (X_n(1), \cdots, X_n(d))$ for all $n \in \mathbb{N}$ is a d-dimensional Poisson point process with intensity measure $E\xi = E(N) \cdot Q$ where Q denotes the joint distribution of the $\mathbf{X}_n$. ■

Note that a similar result holds if more than one-dimensional projections of the event points are considered.

The last theorem is one key result for the modeling and simulation of dependent Poisson processes, in the conventional sense as counting processes. Here, the $X_n(k)$ correspond to the times of occurrences of different kinds of "events", hence are typically concentrated on $\mathbb{R}^+$.

Construction Method I (fixed N). Let $\xi = \sum_{n=1}^{N} \varepsilon_{\mathbf{X}_n}$ be a Poisson point process with d-dimensional event-time points $\mathbf{X}_n = (X_n(1), \cdots, X_n(d))$ whose joint distributions

for each n are given through a copula $C_{\mathbf{X}}$. Then, according to Theorem 5.3, the processes $\xi(k) = \sum_{n=1}^{N} \varepsilon_{X_n(k)}, k = 1, \cdots, d$ are also Poisson, but dependent. For all Borel sets $A, B \subseteq \mathbb{R}$, the correlation between each two processes $\xi(i)$ and $\xi(j)$ (in the sense of Griffiths et al. [10])[1] is given by

$$\rho\big(\xi(i)(A), \xi(j)(B)\big) = \frac{Q_{ij}(A \times B)}{\sqrt{Q_1(A) Q_j(B)}}, \quad i, j \in \{1, \cdots, d\}$$

where Q_{ij} denotes the *joint* distribution of $X_n(i)$ and $X_n(j)$, and Q_i and Q_j denote the *marginal* distributions of $X_n(i)$ and $X_n(j)$, respectively. Note that the above correlation is necessarily non-negative. Moreover, when the marginal distributions of the $X_n(k)$ are fixed, the above expression depends only on the copula $C_{\mathbf{X}}$. In either case, the correlation does *not* depend on the distribution of the counting variable N, and there is a non-zero correlation also in case that $X_n(i)$ and $X_n(j)$ are *independent*.

A short proof of the correlation formula can be given as follows. Let $I_n(i) = \mathbf{1}_{\{X_n(i) \in A\}}$ and $I_n(j) = \mathbf{1}_{\{X_n(j) \in B\}}$. Then the $I_n(i)$ are independent and identically distributed as $I(i)$, say with a binomial distribution with $E(I(i)) = Q_i(A)$. Likewise, the $I_n(j)$ are independent and identically distributed as $I(j)$, say with a binomial distribution with $E(I(j)) = Q_j(B)$. Further, $\xi(i)(A) = \sum_{n=1}^{N} I_n(i)$, $\xi(j)(B) = \sum_{n=1}^{N} I_n(j)$ with

$$
\begin{aligned}
E\big(\xi(i)(A) \cdot \xi(j)(B)\big) &= E\left(\sum_{n=1}^{N} I_n(i) \cdot \sum_{n=1}^{N} I_n(j)\right) = E\left(\sum_{k=1}^{N} \sum_{m=1}^{N} I_k(i) \cdot I_m(j)\right) \\
&= E\left(\sum_{n=1}^{N} I_n(i) \cdot I_n(j)\right) + E\left(\sum_{\substack{k \neq m \\ l \leq k, m \leq N}} I_k(i) \cdot I_m(j)\right) \\
&= E(N) E\big(I(i) \cdot I(j)\big) + E(N(N-1)) E(I(i)) E(I(j)) \\
&= \lambda E\big(I(i) \cdot I(j)\big) + \lambda^2 E(I(i)) E(I(j))
\end{aligned}
$$

where $\lambda = E(N)$. It follows that

$$
\begin{aligned}
\mathrm{Cov}\big(\xi(i)(A), \xi(j)(B)\big) &= \lambda E\big(I(i) \cdot I(j)\big) + \lambda^2 E\big(I(i) \cdot I(j)\big) - \lambda E(I(i)) \cdot \lambda E(I(j)) \\
&= \lambda E\big(I(i) \cdot I(j)\big)
\end{aligned}
$$

---

[1] Note that for stochastic processes, there is no natural definition of correlation. Here, correlation of processes is understood as a set function, i.e. correlation between the random variables defined through the corresponding evaluation maps.

from which the correlation formula follows immediately:

$$\rho\left(\xi(i)(A),\xi(j)(B)\right) = \frac{\lambda E(I(i) \cdot I(j))}{\sqrt{\lambda E(I(i))\,\lambda E(I(j))}} = \frac{E(I(i) \cdot I(j))}{\sqrt{E(I(i))\,E(I(j))}} = \frac{Q_{ij}(A \times B)}{\sqrt{Q_1(A)\,Q_j(B)}}.$$

For related results, we refer to Griffiths et al. [10].

A typical application of such processes in insurance and finance occurs for instance if multiple events of different kind occur almost simultaneously or with short delay, or, on the contrary, exclude each other to a large extend. W.r.t. insurance, one could think of events that are triggered by the same climatic effect like a strong and geographically extended low pressure system, where buildings and their content might be affected by windstorm (perhaps with the delay of one or two days, if the area with insured property is large enough), and flooding or landslide could occur several days after the storm due to heavy rainfall. In these cases, copulas inducing high positive correlation are appropriate for modeling. The converse (negative correlation) might typically show up if certain events cannot occur simultaneously (see the subsequent section 5 for a more elaborate example). For instance, in finance, if a portfolio contains different types of options (e.g. a call and a put with the same strike price), then these will in many cases not be simultaneously executed upon movements of stock prices.

A "fine tuning" of such models is also possible, by superposition and thinning as indicated above. For instance, it might happen that flooding will not necessarily occur with each storm; then the corresponding marginal process for flooding can be thinned in an appropriate way. It could also be the case that some of the events are independent of each other; then additional components can be introduced via superposition of new (independent) processes. Also, it is possible to restrict the original process to a smaller time window (in particular, if the process is time-homogeneous), so that the number of occurrences for the different events in that time window is not the same (but has the same mean).

Construction Method II (variable N's). This approach is based on the results in section 3. Here we use different counting variables $N_1, N_2, \cdots, N_d$ having a joint distribution with Poisson margins, governed by an appropriate copula $C_{\mathbf{N}}$ and one-dimensional Poisson processes $\xi(k) = \sum_{n=1}^{N_k} \varepsilon_{X_n(k)}, k = 1, \cdots, d$ with one-dimensional occurrence time points $X_n(k)$ which themselves might be dependent, i.e. the components of $\mathbf{X}_n = (X_n(1), \cdots, X_n(d))$ might again be governed by a copula $C_{\mathbf{X}}$. In case that the $X_n(k)$ are totally independent the resulting correlation can be similarly expressed as

$$\rho\left(\xi(i)(A),\xi(j)(B)\right) = \rho\left(N_i, N_j\right) \cdot \sqrt{Q_1(A)\,Q_j(B)}, \quad i, j \in \{1, \cdots, d\}$$

so that the copula $C_{\mathbf{N}}$ plays the dominating role here. As above, we have, by the independence assumption,

$$E(\xi(i)(A) \cdot \xi(j)(B)) = E\left(\sum_{n=1}^{N_i} I_n(i) \cdot \sum_{n=1}^{N_j} I_n(j)\right) = E(N_i, N_j)Q_i(A)Q_j(B);$$

$$\mathrm{Cov}(\xi(i)(A), \xi(j)(B)) = E(\xi(i)(A) \cdot \xi(j)(B)) - E(\xi(i)(A))E(\xi(j)(B))$$

$$= \{E(N_i, N_j) - E(N_i)E(N_j)\}Q_i(A)Q_j(B) = \mathrm{Cov}(N_i, N_j) \cdot Q_i(A)Q_j(B);$$

$$\rho(\xi(i)(A), \xi(j)(B)) = \frac{\mathrm{Cov}(N_i, N_j) \cdot Q_i(A)Q_j(B)}{\sqrt{\lambda_i Q_i(A)}\sqrt{\lambda_j Q_j(B)}} = \rho(N_i, N_j) \cdot \sqrt{Q_i(A)Q_j(B)}$$

as stated. Note that unlike in Theorem 4.5, the joint d-dimensional point process is in general *not* a Poisson process any more. Note also that in contrast to construction method I, negative correlations can show up here, which is occasionally intended.

Applications of such processes in insurance and finance occur if different portfolios have to be modeled, which have potential influence on each other without being dependent on identical triggers. This kind of modeling is much more sophisticated and requires a great deal of prior background information.

5. DEPENDENT RISK PROCESSES

The classical model in the collective risk theory (see Rolski et al. [27]) is described by a counting process $\{N(t)|t \geq 0\}$ and a family $\{Z_n|n \in \mathbb{N}\}$ of independent non-negative random variables (claims), independent also of $\{N(t)|t \geq 0\}$. The risk process (aggregated claims process) $\{R(t)|t \geq 0\}$ is then defined by

$$R(t) = \sum_{n=1}^{N(t)} Z_n, \quad t \geq 0.$$

In case that the counting process is Poisson, this model can easily be subsumed under the approach provided in section 4. In the simplest case, we take $\xi = \sum_{n=1}^{N} \varepsilon_{\mathbf{X}_n}$ to be 2-dimensional, the first component of the $\mathbf{X}_n$ consisting of the occurrence times related to the counting process $\{N(t)|t \geq 0\}$, the second component consisting of the claims occurring at these times. Likewise, we could also consider a d-dimensional Poisson process again, where the last $d-1$ components of the $\mathbf{X}_n$ correspond to multiple claims at the same occurrence times. In any case, it is possible to adapt the construction methods I and II presented in the previous section in order to treat (dependent) occurrence times and (dependent) claims simultaneously.

The following picture shows a construction (simulation) of two dependent locally negatively correlated homogeneous Poisson processes $\xi(1)$ and $\xi(2)$ with intensity $\lambda = 1$ based on a subsample from the Frank copula in Fig. 6. On the r.h.s., the corresponding Poisson pairs are shown. The construction for the
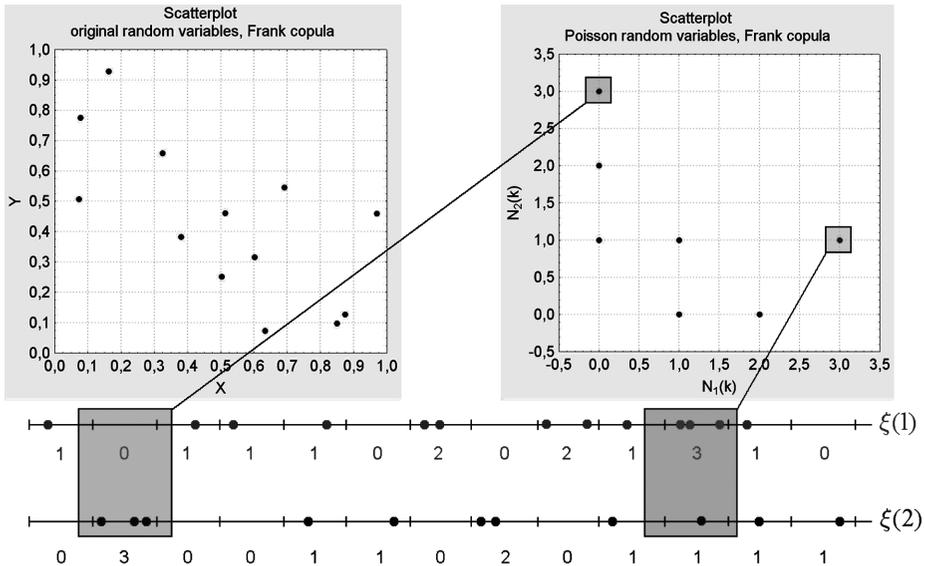
FIGURE 7.

two Poisson point processes $\xi(1)$ and $\xi(2)$ follows method II above, by super-position of indepent locally homogeneous Poisson point processes in the inter-vals $[k, k-1)$ for $k \in \{1, 2, \cdots, 13\}$.

6. SUMMARY

The modeling and simulation of dependence structures in insurance and finance portfolios is a task that cannot be achieved in a simple mechanical way. As was shown above it is for instance possible to present different constructions of Poisson and Poisson related processes which show identical behavior w.r.t. cor-relation, for instance. This means that knowledge of correlations alone is not sufficient to map the complicated processes observed in the real world into a (more or less sophisticated) mathematical model. It is therefore necessary to provide a certain minimal amount of information about the physical or economic conditions which rule the processes under consideration. With such informa-tion, however, it becomes possible to select from a broad palette of potential models based on the point process approach as outlined above, which are flex-ible enough to cover many important aspects of DFA and IRM, even if a "perfect" solution will perhaps never be possible.

ACKNOWLEDGEMENT

REFERENCES

[1] BILLINGSLEY, P. (1986) *Probability and Measure*. Wiley, N.Y.
[2] BLUM, P., DIAS, A. and EMBRECHTS, P. (2002) The ART of dependence modelling: the latest advances in correlation analysis. In: M. Lane (ed.): Alternative Risk Strategies, Risk Books, London.
[3] CAMBANIS, S., SIMONS, G. and STOUT, W. (1976) Inequalities for $\mathcal{E}k(X, Y)$ when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie u. verw. Gebiete* **36**, 285-294.
[4] CAMPBELL, J.T. (1934) The Poisson correlation function. *Proc. Edinb. Math. Soc., II. Ser.* **4**, 18-26.
[5] COX, D.R. and LEWIS, P.A.W. (1972) Multivariate point processes. Proc. 6th Berkeley Sympos. *Math. Statist. Probab.*, **3**, 401-448, University of California, 1970.
[6] DALEY, D.J. and VERE-JONES, D. (1988) *An Introduction to the Theory of Point Processes*. Springer, N.Y.
[7] EMBRECHTS, P., STRAUMANN, D. and MCNEIL, A.J. (2000) Correlation: pitfalls and alternatives. In: P. Embrechts (ed.): Extremes and Integrated Risk Management. Risk Books, London.
[8] EMBRECHTS, P., MCNEIL, A.J. and STRAUMANN, D. (2002) Correlation and dependence in risk management: properties and pitfalls. In: M.A.H. Dempster (ed.): Risk Management: Value at Risk and Beyond. Cambridge Univ. Press, Cambridge.
[9] GASS, S.I. (1969) Linear Programming: Methods and Applications. McGraw-Hill Book Company and Kogakusha Company, Ltd., Tokyo, 3rd ed.
[10] GRIFFITHS, R.C., MILNE, R.K. and WOOD, R. (1979) Aspects of correlation in bivariate Poisson distributions and processes. *Australian Journal of Statistics* **21(3)**, 238-255.
[11] HOEFFDING, W. Maßstabinvariante Korrelationstheorie. *Schriften des Mathematischen Seminars und des Instituts für Angewandte Mathematik* **5**, 181-233, Universität Berlin.
[12] HOFFMAN, A.J. (1963) On simple linear programming problems. In: Klee (ed.): Convexity, vol. 7, 317-327, Providence, R.I., Proc. Symp. Pure Math.
[13] JOE, H. (1997) Multivariate Models and Dependence Concepts. Chapman & Hall, London.
[14] JOHNSON, N.L., KOTZ, S. and BALAKRISHNAN, N. (1997) Discrete Multivariate Distributions. Wiley, N.Y.
[15] JUNKER, M. and MAY, A. (2002) *Measurement of aggregate risk with copulas*. Caesar preprint 021, Center of Advanced European Studies and Research, Bonn, Germany.
[16] KINGMAN, J.F.C. (1993) *Poisson Processes*. Clarendon Press, Oxford.
[17] KOCHERLAKOTA, S. and KOCHERLAKOTA, K. (1992) Bivariate Discrete Distributions. M. Dekker, N.Y.
[18] LAKSHMINARAYANA, J., PANDIT, S.N.N. and RAO, K.S. (1999) On a bivariate Poisson distribution. C*ommunications in Statistics – Theory and Methods* **28(2)**, 267-276.
[19] LEE, M.-L.T. (1996) Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics – Theory and Methods* **25(6)**, 1207-1222.
[20] MARI, D.D. and KOTZ, S. (2001) Correlation and Dependence. Imp. College Press, London.
[21] MÜLLER, A. and BÄUERLE, N. (1998) Modelling and comparing dependencies in multivariate risk portfolios. *ASTIN Bulletin* **28(1)**, 59-76.
[22] NELSEN, R.B. (1987) Discrete bivariate distributions with given marginals and correlation. *Communications in Statistics – Simulation* **16(1)**, 199-208.
[23] NELSEN, R.B. (1999) An Introduction to Copulas. Lecture Notes in Statistics 139, Springer, N.Y.
[24] NESLEHOVA, J. (2004) Dependence of Non-Continuous Random Variables. Ph.D. Thesis, University of Oldenburg.
[25] REISS, R.-D. (1993) A Course on Point Processes. Springer, N.Y.
[26] RODRIGUES, J., CID, J.E.R. and ACHCAR, J.A. (2002) Bayesian analysis for the superposition of two dependent nonhomogeneous Poisson processes. *Communications in Statistics – Theory and Methods* **31(9)**, 1467-1478.
[27] ROLSKI, T., SCHMIDLI, H., SCHMIDT, V. and TEUGELS, J. (1998) *Stochastic Processes for Insurance and Finance*. Wiley, N.Y.
[28] RACHEV, S.T. and RÜSCHENDORF, L. (1998) Mass Transportation Problems. Volume I: Theory. Springer, N.Y.

[29] RACHEV, S.T. and RÜSCHENDORF, L. (1998) *Mass Transportation Problems. Volume II: Applications.* Springer, N.Y.

[30] TEICHER, H. (1954) On the multivariate Poisson distribution. *Skandinavisk Aktuarietidskrift* **37**, 1-9.

PROF. DR. DIETMAR PFEIFER
Institut für Mathematik
Fakultät für Naturwissenschaften und Mathematik
Carl von Ossietzky Universität
D-26111 Oldenburg
Germany
Email: dietmar.pfeifer@uni-oldenburg.de
Web: www.mathematik.uni-oldenburg.de/personen/pfeifer/pfeifer.html

and

AON Jauch&Hübener
Heidenkampsweg 58
D-20059 Hamburg
Germany
Web: www.aon.com/de/ge/

JOHANA NEŠLEHOVÁ
RiskLab
ETH-Zentrum
CH-8092 Zürich
Switzerland
Email: neslehova@math.ethz.ch
Web: www.risklab.ch/