

# Multivariate counting processes: copulas and beyond

Nicole Bäuerle and Rudolf Grübel

*Universität Hannover*

Multivariate stochastic processes with Poisson marginals are of interest in insurance and finance; they can be used to model the joint behaviour of several claim arrival processes, for example. We discuss various methods for the construction of such models, with particular emphasis on the use of copulas. An important class of multivariate counting processes with Poisson marginals arises if the events of a background Poisson process with constant intensity are moved forward in time by a random amount and possibly deleted; here we think of the events of the background process as triggering later claims in different categories. We discuss structural aspects of these models, their dependence properties together with stochastic order aspects, and also some related computational issues. Various actuarial applications are indicated.

*Keywords and phrases.* Claim arrival processes, dependency, point processes, Poisson processes, stochastic modelling, stochastic ordering.

*AMS 1991 subject classifications.* Primary 91B30, 91B70; secondary 60G55.

**1. Introduction.** The standard model from risk theory assumes that claim arrival times form a (homogeneous) Poisson process and that claim sizes are independent and identically distributed; also, claim sizes and claim arrival times are independent. There is considerable interest in and need for models that take dependencies into account. One specific such question deals with the claim arrival times only: If we consider more than one type of claim then naturally the question arises as to how dependencies between the claim arrival times for the different claim types can be modelled. (Often, it is easy to extend such models so that they include claim sizes. Also, the counting processes can be regarded as representing the total claim size(s) in the degenerate situation where all claims have size 1.) A standard example for such possible interdependencies is that of a windstorm giving rise to immediate claims, but also leading to heavy rain and flooding with the corresponding later claims. Another important case arises in connection with claims originating at different spatial locations.

In the present paper we specifically consider models that have classical, but possibly dependent marginals. We are interested in the construction and analysis of multivariate stochastic processes  $X = (X_1, \dots, X_d)$  with the property that the one-dimensional marginal processes  $X_i$  are Poisson processes with constant rates  $\lambda_i$  respectively,  $i = 1, \dots, d$ . As  $X_i$  counts the claims of type  $i$  we are therefore

in the classical situation as far as the individual components are concerned. The emphasis is on modelling dependencies between the marginal processes.

We also assume stationarity in time; in order to avoid confusion with other stationarity concepts referring to space shifts in multivariate point processes, for example, we will use the term *time shift stationarity*. Informally, the component counting processes count events, such as claims of a particular type in the present application context. Formally, we regard a counting process  $X_i$  as a family of nonnegative, integer-valued random variables indexed by subsets  $A$  of the real line  $\mathbb{R}$ , where  $X_i(A)$  denotes the number of events of type  $i$  with ‘time stamp’ in  $A$ . If observations start at time 0, then it is customary to work with  $t \mapsto X_i([0, t])$  as the random measure  $A \mapsto X_i(A)$  can then conveniently be described by its ‘distribution function’. This leads to the familiar situation where the stochastic process is a family of random variables indexed by real numbers. However, the slightly more abstract view of counting processes as random measures is the key to a concise description and also, in our opinion, to a deeper understanding of our models. In particular, time shift stationarity can now formally be expressed as the requirement that the distribution of the random vector  $(X_1(A_1+t), \dots, X_d(A_d+t))$  does not depend on  $t \in \mathbb{R}$ . Here  $A_1, \dots, A_d$  are Borel subsets of the real line and we have written  $A_i + t$  for the shifted set  $\{x + t : x \in A_i\}$ .

The paper is organized as follows. In Section 2 we review two recent approaches to the construction of multivariate counting processes with Poisson process marginals. Both are related to the idea of modelling dependencies via copulas. In the first approach  $X$  is assumed to be a multivariate Lévy process which implies that the component processes can be regarded as thinnings of a basic (univariate) Poisson process; see Lindskog and McNeil (2003). In the second approach, due to Pfeifer and Nešlehová (2004), copulas are used to construct  $X((0, T])$  for fixed  $T > 0$ . In Section 3 we introduce models that incorporate thinning and shifts; we will use the acronym ‘TaS’ for this class. These models avoid some of the limitations of the constructions discussed in Section 2. We present some auxiliary (elementary) material from the theory of point processes and then obtain a structural result which is the basis for the further analysis, in particular for the investigation of dependence properties and stochastic order aspects of TaS models in Section 4. A key step is the interpretation of the thinning and shift operations as operations on the multivariate Poisson process that results if we attach suitable marks to the points of the initial univariate process of triggering events. In Section 5 we show that TaS models arise quite naturally from the assumption of time shift stationarity if the multivariate counting process  $X$  can be obtained from a multivariate Poisson process  $N$  by coordinate projections. We relate time shift stationarity of  $X$  to a spatial form of stationarity of  $N$  and an invariance property of the intensity measure associated with  $N$ . Interestingly, a variant of the copula idea of separating marginal distributions and dependence aspects reappears in a version for Poisson point processes. In Section 6 we list various actuarial applications. Section 7 deals with related computational issues; in this last section we also look more closely at the two-dimensional case, we work out an example in detail and we close with some summarizing comments.

The models discussed below also give rise to some interesting statistical problems, but these will be investigated in a separate paper.

**2. Copula-based models.** The distribution function  $F$  of a  $d$ -dimensional random vector  $\xi = (\xi_1, \xi_2, \dots, \xi_d)$  can be written in the form

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

where  $F_1, \dots, F_d$  are the one-dimensional distribution functions associated with the components  $\xi_1, \dots, \xi_d$  of the random vector and where  $C$  is a  $d$ -dimensional distribution function on the unit cube with uniform marginals, i.e.  $C(u, 1, \dots, 1) = \dots = C(1, \dots, 1, u) = u$  for all  $u$  in the unit interval  $[0, 1]$ . This decomposition separates the marginal distributions from the dependence structure; standard references are Nelsen (1999) and Joe (1997). Chapter 5 of a forthcoming book by McNeil, Frey and Embrechts (2004) reviews copulas in the context of risk theory and in particular discusses the use of the decomposition to explain the practically important phenomenon of tail clustering.

Suppose now that we start observations at time 0 so that our multivariate counting process can be indexed by  $\mathbb{R}_+$  as explained in the introduction. Then the transformation to uniform marginals can be applied to the individual random vectors  $(X_1(t), \dots, X_d(t))$  for each  $t \geq 0$ . This would result in a family  $(C_t)_{t \geq 0}$  of copulas. In comparison to the above static situation this has two disadvantages: First, a whole family of copulas is simply an unwieldy object. Secondly, as we will explain in Remark (d) at the end of Section 5, this family would not satisfy our demands: Whereas in the random vector case the distribution is completely specified by the copula and the marginal distributions, the family  $(C_t)_{t \geq 0}$  and the distributions of the component processes  $X_i$ ,  $1 \leq i \leq d$ , together do not determine the finite-dimensional distributions of the multivariate counting process  $X$ .

We are aware of two attempts to overcome these difficulties. For the first of these, we introduce some notation that will also be useful in the subsequent sections. Let  $\mathbb{D} := \{1, \dots, d\}$  and for all  $D \subset \mathbb{D}$ ,

$$e(D) = \begin{pmatrix} e_1(D) \\ \vdots \\ e_d(D) \end{pmatrix} \quad \text{with } e_i(D) = \begin{cases} 1, & \text{if } i \in D, \\ 0, & \text{otherwise} \end{cases}$$

(we switch between row and column vectors whenever typographically convenient). The condition that the marginal processes be constant rate Poisson processes implies that the paths of  $X$  are of pure jump type and that the jumps  $X(t) - X(t-)$  are of the form  $e(D)$  for some (non-empty)  $D \subset \mathbb{D}$ . A jump  $e(D)$  at time  $t$  means that the component processes  $X_i$  with  $i \in D$  (and only these) have a jump of height 1 at time  $t$ . If we assume in addition to the above that  $X$  is a Lévy process, then the Markov property and the homogeneity in time imply that the copula family introduced above can be described by its ‘infinitesimal generator’, the Lévy copula, which together with the other characteristics of a Lévy process is enough to specify the distribution of the whole multivariate process; see e.g. the recent monograph by Cont and Tankov (2004). In particular, for counting processes with paths of pure jump type and jumps of the form  $e(D)$  the Lévy assumption implies that the distribution of  $X$  is completely specified by (up to)  $2^d - 1$  numbers  $\lambda_D \geq 0$ ,  $\emptyset \neq D \subset \mathbb{D}$ : A counting process  $X$  of Lévy type can be represented as

$$X_i(t) = \sum_{D \ni i} N_D(t), \quad i = 1, \dots, d,$$

where  $N_D$  are independent Poisson processes with rate  $\lambda_D$ ,  $D$  as above. The rates  $\lambda_i$  of the marginal processes are related to the rates  $\lambda_D$  by  $\lambda_i = \sum_{D \ni i} \lambda_D$ ,  $i = 1, \dots, d$ . A similar result has been used in Lindskog and McNeil (2003). The processes  $N_D$  in turn can be obtained from one single (univariate) Poisson process  $N$  with rate  $\lambda = \sum_{D \subset \mathbb{D}} \lambda_D$  by independent marking of the points of  $N$  with probabilities  $p_D = \lambda_D/\lambda$  and then collecting the marked points into  $N_D$ . This may be easier to understand with the help of an example: Figure 1 shows a segment of a simulated path for  $d = 2$ , with  $\lambda_{\{1\}} = \lambda_{\{2\}} = \lambda_{\{1,2\}} = 1$ .

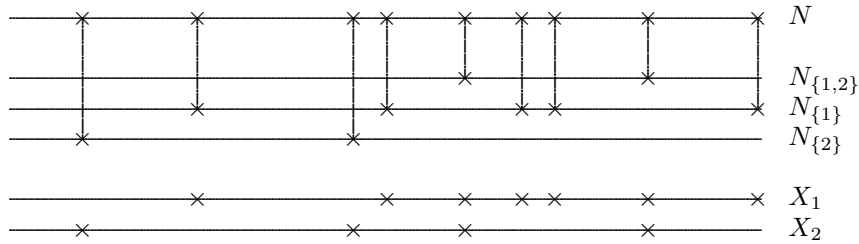


Figure 1: Representation of counting processes of Lévy type

Lévy type counting processes have a very specific dependence structure: In the random measure notation,  $X_i(A_i)$  and  $X_j(A_j)$  will always be independent if  $A_i \cap A_j = \emptyset$  because of the independence of the increments of  $X$ . Hence, for such processes, dependence of the marginal processes is only possible via the synchronicity of the jumps.

A completely different approach has recently been suggested by Pfeifer and Nešlehová (2004). In one of their models these authors consider the finite time interval  $[0, T]$  instead of  $\mathbb{R}_+$  and condition on the final random vector  $X(T) = (X_1(T), \dots, X_d(T))$  of the process. A static  $d$ -dimensional copula can be used, together with the condition that the components  $X_i(T)$  have a Poisson distribution with parameter  $\lambda_i T$ ,  $i = 1, \dots, d$ , to construct the law of the vector  $X(T)$ . Pfeifer and Nešlehová (2004) discuss in detail the use of standard copulas, including a description of an algorithm that implements the discretization step. To obtain  $(X(t))_{0 \leq t \leq T}$  from the final random vector  $X(T)$  one can use the familiar fact that, conditionally on their total number in  $[0, T]$  being equal to  $n$ , the  $n$  points of a Poisson process with constant rate are independent and uniformly distributed on  $[0, T]$ . Hence, in the  $d$ -dimensional case, we obtain  $(X(t))_{0 \leq t \leq T}$  from  $X(T)$  via

$$X_i(t) = \sum_{j=1}^{X_i(T)} 1_{[0,t]}(\xi_{ij}), \quad 1 \leq i \leq d, \quad 0 \leq t \leq T,$$

where  $\xi_{ij}$ ,  $1 \leq i \leq d$ ,  $1 \leq j \leq X_i(T)$ , are independent and uniformly distributed on  $[0, T]$ . Here  $1_A$  denotes the indicator function associated with the set  $A$ .

In this approach the dependence modelling of the component processes is thus based entirely on the dependence modelling of the total number of claims in the

period of interest. This means that this period has to be fixed in advance. In order to overcome the restriction to a finite time interval there are two strategies: First, one could simply take  $T$  large enough. The second method makes use of the fact that the superposition of independent Poisson processes is again a Poisson process. One could therefore extend the construction to arbitrary time intervals by patching together enough intervals of the form  $[k, k + 1)$ ,  $k = 0, 1, \dots$ ; see Section 5 in Pfeifer and Nešlehová (2004). Note, however, that with the second approach the dependence of the component processes is limited to the somewhat arbitrary base intervals: What happens in  $[k, k + 1)$  and  $[l, l + 1)$  is independent if  $k \neq l$ . Also, the result is not a time shift stationary process in the sense explained in the introduction: Whereas  $X_1(k + 1) - X_1(k + 1/2)$  and  $X_1(k + 1 + 1/2) - X_1(k + 1)$  will always be independent this is in general not the case with  $X_1(k + 1/2) - X_1(k)$  and  $X_1(k + 1) - X_1(k + 1/2)$ , if the base intervals are chosen as above; i.e., in the random measure notation, we may (and in general will) have that the distribution of  $(X_1(A_1), X_2(A_2))$  is not the same as the distribution of  $(X_1(A_1 + t), X_2(A_2 + t))$  if  $A_1 = A_2 = (0, 1/2]$  and  $t = 1/2$ .

If instead we choose  $T$  large enough another problem arises. The construction implies that the conditional distribution  $\mathcal{L}(X(t)|X(T))$  of  $X(t)$  given  $X(T)$  can be written as

$$\mathcal{L}(X(t)|X(T)) = \text{Bin}(X_1(T), t/T) \otimes \dots \otimes \text{Bin}(X_d(T), t/T),$$

where  $\text{Bin}(n, p)$  denotes the binomial distribution with parameters  $n$  and  $p$  and the ‘ $\otimes$ ’ is product measure, representing independence. From the familiar law of small numbers we know that  $\text{Bin}(\lambda_T, t/T)$  converges to the Poisson distribution with parameter  $t\lambda$  if  $\lim_{T \rightarrow \infty} \lambda_T/T = \lambda$ . The conditions on the marginal processes (together with the strong law of large numbers) imply that  $X_i(T)/T \rightarrow \lambda_i$  almost surely as  $T \rightarrow \infty$ , for  $i = 1, \dots, d$ . Hence the components of  $X(t)$  will asymptotically be independent, almost surely. This can easily be generalized to the finite-dimensional increments of  $X$  and leads to the overall conclusion that, asymptotically as  $T \rightarrow \infty$ , the copula chosen for the final value will be irrelevant to any fixed initial segment  $(X(s))_{0 \leq s \leq t}$  of the process and that the components of such a segment will become independent.

Thus, in summary, despite their intrinsic interest both constructions have shortcomings if we analyze the resulting models from the point of view that we adopt in the present paper. First, as an immediate consequence of their defining properties, Lévy type counting processes have no dependence ‘across time’, and the dependence ‘across components’ is of a very special nature. For this class of processes dependence modelling is reduced to the choice of thinning probabilities. Secondly, the models introduced by Pfeifer and Nešlehová (2004) are more flexible as they can make use of the whole range of (static) copulas, but they require the choice of a compact base interval  $[0, T]$ . If such an interval does not suggest itself from the application of interest, then both remedies known so far, letting either  $T$  tend to  $\infty$  or patching together several such intervals, have their disadvantages, the second getting into conflict with our assumption of time shift stationarity.

In the next section we introduce a family of models that extends the Lévy models by incorporating random shifts of the individual points. We will see that, at least to some extent, this can be used to overcome the difficulties that we have discussed in detail in this section.

**3. Models with thinning and shifts.** We first recall some elementary facts from the general theory of point processes; Daley and Vere-Jones (1988) and Resnick (1987) are standard references in this area. In order to keep the presentation compact we skip some technical details, such as measurability issues. In connection with our models the special class of Poisson point processes suffices: Given a  $\sigma$ -finite measure  $\nu$  on some space  $E$ , an integer-valued random measure  $N$  on  $E$  is a Poisson process with intensity measure  $\nu$ , which we abbreviate to  $N \sim \text{PP}(\nu)$ , if

(D1)  $N(A)$  has a Poisson distribution with mean  $\nu(A)$ ,

(D2) for pairwise disjoint  $A_1, \dots, A_k$ , the random variables  $N(A_1), \dots, N(A_k)$  are independent.

With  $E = \mathbb{R}$  and  $\nu = \rho\ell$ ,  $\rho > 0$  and  $\ell$  Lebesgue measure, we obtain the familiar Poisson process on the real line with constant rate (or intensity)  $\rho$ . Quite generally, if  $\nu$  is diffuse in the sense that  $\nu(\{x\}) = 0$  for all  $x \in E$ , in addition to being  $\sigma$ -finite, then  $N \sim \text{PP}(\nu)$  has countable support and no multiple points (with probability 1). This means that, with  $\delta_x$  the one-point mass in  $x \in E$ ,  $N = \sum_{i=1}^{\infty} \delta_{R_i}$ , with  $(R_i)_{i \in \mathbb{N}}$  a sequence of  $E$ -valued random quantities.

There are four basic operations on Poisson processes that we will need below, see e.g. Section 3.3.2 in Resnick (1987). First, a mapping  $T : E \rightarrow F$  transforms a point process  $N_E$  on  $E$  into a point process  $N_F$  on  $F$  via  $N_F(A) := N_E(T^{-1}(A))$ . The class of Poisson processes is stable under such mappings in the sense that  $N_E \sim \text{PP}(\nu_E)$  implies  $N_F \sim \text{PP}(\nu_F)$ , where  $\nu_F(A) = \nu_E(T^{-1}(A))$  is the measure-theoretical image of  $\nu_E$  under  $T$ , in the same way as  $N_F$  is the image of  $N_E$  under  $T$ . Secondly, if we split  $N$  independently such that each point of  $N$  becomes a point of  $N_1$  with probability  $p$  and of  $N_2$  with probability  $1 - p$ , then  $N_1$  and  $N_2$  are independent and  $N_1 \sim \text{PP}(p\nu)$ ,  $N_2 \sim \text{PP}((1 - p)\nu)$  if  $N \sim \text{PP}(\nu)$ . This splitting property may in fact be deduced from the marking property: If  $(Y_i)_{i \in \mathbb{N}}$  is a sequence of independent and identically distributed  $F$ -valued random variables and  $N = \sum_{i=1}^{\infty} \delta_{R_i} \sim \text{PP}(\nu)$ , then  $\tilde{N} := \sum_{i=1}^{\infty} \delta_{(R_i, Y_i)}$  is a Poisson process on  $E \times F$  with intensity measure  $\tilde{\nu} := \nu \otimes Q$ , where  $Q$  is the distribution of the  $Y_i$ 's. Finally, from (D1) and (D2) the superposition property follows easily: If  $N_1 \sim \text{PP}(\nu_1)$ ,  $N_2 \sim \text{PP}(\nu_2)$  are independent, both with the same state space  $E$ , then  $N_1 + N_2 \sim \text{PP}(\nu_1 + \nu_2)$ , where  $(N_1 + N_2)(A) = N_1(A) + N_2(A)$ . The splitting, marking and superposition properties extend to the case of more than two constituents by induction.

For the construction of the TaS models we start with a background Poisson process  $N$  on  $\mathbb{R}$  with constant intensity  $\lambda$  and a thinning mechanism described by a probability distribution  $(p_D)_{D \subset \mathbb{D}}$  on  $\mathcal{P}(\mathbb{D})$ , the power set (set of all subsets) of  $\mathbb{D}$ , as in the first model type discussed in Section 2. Additionally, we now have a sequence  $(Y_l)_{l \in \mathbb{Z}}$  of independent  $d$ -dimensional random vectors, the shifts, all with distribution  $Q$ . In the actuarial applications,  $Q$  will in general be concentrated on the non-negative orthant  $\mathbb{R}_+^d$ , but our results below do not require such a support condition. We assume that the Poisson process, the thinning mechanism and the shift sequence are independent; as a consequence, the stochastic structure of all subsequent constructions is fully determined by  $\lambda$ ,  $(p_D)_{D \subset \mathbb{D}}$  and  $Q$ . We may order the points  $(T_l)_{l \in \mathbb{Z}}$  of the background process in such a way that

$$-\infty < \dots < T_{-2} < T_{-1} < T_0 < 0 \leq T_1 < T_2 < \dots < \infty.$$

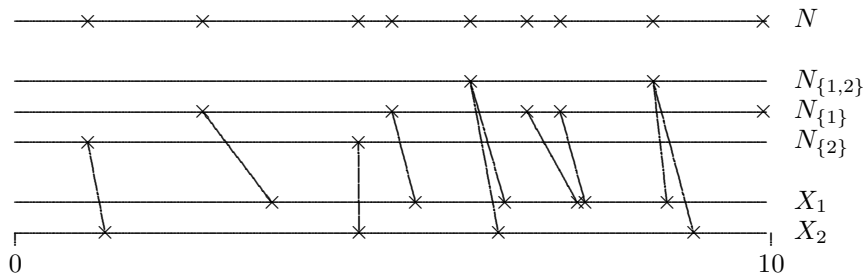


Figure 2: An example combining thinning and shifts

The points for component  $X_i$  of  $X$  are now constructed by shifting those  $T_l$  that are not deleted for this component (which happens with probability  $\sum_{D \ni i} p_D$ ) by the amount  $Y_{li}$ ,  $Y_l = (Y_{l1}, \dots, Y_{ld})$ . Again, it may be easier to understand this construction by way of an illustration: Figure 2 has the same basic parameters as used for Figure 1 but includes shifts simulated from a distribution  $Q$  that is the product of the uniform distribution on the unit interval and the exponential distribution with mean 1 (shifts with origin to the left or destination to the right of the plotting window are not shown).

$N$	$Y_1$	$Y_2$	$D$	$X_1$	$X_2$
0.960	0.316	0.230	{2}	—	1.190
2.481	0.916	0.206	{1}	3.397	—
4.543	0.660	0.008	{2}	—	4.551
4.987	0.308	0.463	{1}	5.295	—
6.022	0.455	0.369	{1, 2}	6.477	6.391
6.773	0.663	5.642	{1}	7.436	—
7.211	0.328	0.429	{1}	7.539	—
8.440	0.181	0.533	{1, 2}	8.621	8.973

Table 1: Numerical values for Figure 2

Table 1 displays the numerical values for the first eight claims in Figure 2. For example, an event occurring at time 0.960 gives rise to a claim of type 2 only; the corresponding delay is 0.230, so that the claim is registered at time 1.190. The event occurring at time 6.022 is the first to trigger claims of both types.

This example also helps to explain a technical point: Whereas, in order to emphasize the connection with the Lévy case, Figure 2 suggests that first the thinning is done and then the shifts are applied, Table 1 lists also those shifts that because of the thinning later become irrelevant, such as the value  $Y_1 = 0.316$  in the first line. Because of our basic independence assumptions the order of the two operations is irrelevant; in the proofs it will be more convenient to delete components in the second step. Indeed, we will think of the first four columns of the table as the points of a marked Poisson process in the proof of Theorem 1 below. With this order of the operations of shifting and thinning, we start with a simple one-dimensional

Poisson process with constant rate. With each event of this process we associate a  $d$ -dimensional shift vector and a subset  $D \subset \mathbb{D}$  of surviving components. Once the shifts are carried out a Poisson process on  $\mathbb{R}^d$  results whose points consist of the potential claim arrival times for each of the  $d$  different claim types, with one point for each of the triggering events (the time stamps of the triggering events are lost in the process). Each point of this  $d$ -dimensional process is additionally marked by the set  $D \subset \mathbb{D}$  of claim types that are going survive the thinning step. After the second (thinning) step, we have independent Poisson processes on  $\mathbb{R}^D$ , one for each  $D$  with  $p_D > 0$ . The number  $X_i(A)$  of claims of type  $i$  with time stamp in  $A$  is now obtained by taking the sum of the number of points in these individual processes that have their  $i^{\text{th}}$  coordinate in  $A$ .

It should be clear that this generalizes the Lévy model which we would obtain with  $Q$  concentrated on the zero vector. Also, this model is easy to simulate. We abbreviate the above by calling  $X$  a TaS model ('thin and shift') with parameters  $\lambda$ ,  $p$  and  $Q$ , where  $\lambda$  is the intensity of the Poisson base process,  $p = (p_D)_{D \subset \mathbb{D}}$  the thinning mechanism and  $Q$  the shift distribution. That these models satisfy our basic requirement of Poisson marginals is a standard fact from the general theory of point processes, see e.g. p.138 in Resnick (1987): Shifting the points of a constant rate Poisson process (on  $\mathbb{R}$ ) by i.i.d. amounts results in a constant rate Poisson process.

TaS models can be used to represent a mechanism underlying the generation of claims: There are triggering events, occurring with rate  $\lambda$ ; these may or may not generate claims (here  $p$  is used) at later instances ( $Q$  describes the delay distribution). Similar models are popular in quite diverse fields where they are used to explain clustering. They are special cases of the *Neyman-Scott* or *cluster processes*; see e.g. Sections 3.4 and 4.4 in Cox and Isham (1980). Here, however, we associate different component processes to the different shifts of one original point and we are interested in the dependence of the components, an aspect that is lost once the component processes have been merged into a point process on  $\mathbb{R}$ . (We will briefly discuss the merged processes in Section 6.3 below.)

Suppose now that the  $d$ -dimensional counting process  $X = (X_1, \dots, X_d)$  is of TaS type with parameters  $\lambda$ ,  $p$  and  $Q$ . The following is our basic structural result for these models; it gives a general description of the joint distribution of  $X_1(A_1), \dots, X_d(A_d)$  in terms of these parameters. We need two more definitions: For subsets  $A_1, \dots, A_d$  of  $\mathbb{R}$  and  $D, D' \subset \mathbb{D}$  with  $D \subset D'$  let

$$M(D, D'; A_1, \dots, A_d) := B_1 \times \dots \times B_d \quad \text{with} \quad B_i := \begin{cases} A_i, & \text{for } i \in D, \\ A_i^c, & \text{for } i \in D' \setminus D, \\ \mathbb{R}, & \text{otherwise.} \end{cases}$$

Further, with  $e = e(\mathbb{D}) = (1, \dots, 1)$  the  $d$ -dimensional vector that has all components equal to 1, we define a measure  $\nu(Q)$  on  $\mathbb{R}^d$  by

$$\nu(Q)(A) := \int Q(A - te) dt$$

for all  $d$ -dimensional Borel sets  $A$ ; here  $A - x$  denotes the set  $\{a - x : a \in A\}$ . This measure plays an important role throughout the sequel. It is obvious from



the result cited above on i.i.d. shiftings of constant rate Poisson processes that the marginal measures  $\nu(Q)^{\pi_i}$ ,  $1 \leq i \leq d$ , where  $\pi_i : \mathbb{R}^d \rightarrow \mathbb{R}$  denotes the projection on the  $i^{\text{th}}$  coordinate, are all equal to the Lebesgue measure. The following direct computation may be instructive:

$$\begin{aligned} \nu(Q)^{\pi_i}(B) &= \nu(Q)(\mathbb{R}^{i-1} \times B \times \mathbb{R}^{d-i-1}) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}^d} 1_{\mathbb{R}^{i-1} \times B \times \mathbb{R}^{d-i-1}}(x - te) Q(dx) dt \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} 1_{B-x_i}(t) dt Q(dx) \\ &= \ell(B) \end{aligned}$$

for all one-dimensional Borel sets  $B$ . Other properties of  $\nu(Q)$  will be given where they are needed.

**THEOREM 1** *Let  $X$  be a  $d$ -dimensional counting process of TaS type with base rate  $\lambda$ , thinning mechanism  $p = (p_D)_{D \subset \mathbb{D}}$  and shift distribution  $Q$ . Then, for any Borel subsets  $A_1, \dots, A_d$  of the real line, we have the following distributional representation:*

$$\begin{pmatrix} X_1(A_1) \\ \vdots \\ X_d(A_d) \end{pmatrix} =_{\text{distr}} \begin{pmatrix} \sum_{D \ni 1} \xi(D; A_1, \dots, A_d) \\ \vdots \\ \sum_{D \ni d} \xi(D; A_1, \dots, A_d) \end{pmatrix},$$

where the random variables  $\xi(D; A_1, \dots, A_d)$ ,  $\emptyset \neq D \subset \mathbb{D}$ , are independent and Poisson distributed with

$$E\xi(D; A_1, \dots, A_d) = \lambda \sum_{D' \supset D} p_{D'} \nu(Q)(M(D, D'; A_1, \dots, A_d)).$$

**PROOF:** Marking the events of the background process by their shift vectors and the respective set of surviving component indices leads to a Poisson process on  $\mathbb{R} \times \mathbb{R}_+^d \times \mathcal{P}(\mathbb{D})$  with intensity measure  $\lambda \ell \otimes Q \otimes (p_D)_{D \subset \mathbb{D}}$ . The transformation

$$\begin{aligned} T : \mathbb{R} \times \mathbb{R}_+^d \times \mathcal{P}(\mathbb{D}) &\rightarrow \mathbb{R}^d \times \mathcal{P}(\mathbb{D}), \\ (t, (y_1, \dots, y_d), D') &\mapsto ((t + y_1, \dots, t + y_d), D'), \end{aligned}$$

makes this into a Poisson process  $Z$  on  $\mathbb{R}^d \times \mathcal{P}(\mathbb{D})$  with intensity  $\lambda \nu(Q) \otimes (p_D)_{D \subset \mathbb{D}}$ ;  $Z$  can be regarded as a marked point process on  $\mathbb{R}^d$ , with marks from  $\mathcal{P}(\mathbb{D})$ .

Now let  $A_1, \dots, A_d$  be given. The vector  $X(A) := (X_1(A_1), \dots, X_d(A_d))$  can be written as a linear combination of the vectors  $e(D)$ ,  $D \subset \mathbb{D}$ , with non-negative integer coefficients: Each event of the background process contributes one such vector, with  $D = \emptyset$  in all but finitely many cases. A marked point  $(t, (y_1, \dots, y_d), D')$  contributes to  $e(D)$  if and only if  $D \subset D'$  and  $t + y_i \in A_i$  for  $i \in D$ ,  $t + y_i \notin A_i$  for  $i \in D' \setminus D$ , which translates into

$$D \subset D', \quad (t + y_1, \dots, t + y_d) \in M(D, D'; A_1, \dots, A_d)$$

for the  $Z$ -process. Now let  $\eta(D, D'; A_1, \dots, A_d)$  be the part of the coefficient of  $e(D)$  that comes from the points of  $Z$  with mark  $D'$ ; obviously,

$$X_i(A_i) = \sum_{D \ni i} \sum_{D' \supset D} \eta(D, D'; A_1, \dots, A_d) \quad \text{for } i = 1, \dots, d.$$

Using (D1) we see that  $\eta(D, D'; A_1, \dots, A_d)$  has a Poisson distribution with

$$E\eta(D, D'; A_1, \dots, A_d) = \lambda p_{D'} \nu(Q)(M(D, D'; A_1, \dots, A_d)).$$

If  $(D'_1, D_1) \neq (D'_2, D_2)$  then  $\eta(D_1, D'_1; A_1, \dots, A_d)$  and  $\eta(D_2, D'_2; A_1, \dots, A_d)$  refer to disjoint regions of the state space of  $Z$ , hence with

$$\xi(D; A_1, \dots, A_d) := \sum_{D' \supset D} \eta(D, D'; A_1, \dots, A_d)$$

an appeal to (D2) completes the proof.  $\square$

For some of the applications of the theorem the following interpretation of the variables in the representation, apparent from its proof, is useful: The random variable  $\xi(D; A_1, \dots, A_d)$  is the number of points  $t$  of the base process whose set  $D'$  of surviving components includes  $D$  and whose shifts  $t + y_i$  are in  $A_i$  for  $i \in D$  and in  $A_i^c$  for  $i \in D' \setminus D$ .

Despite its somewhat complicated appearance Theorem 1 has some useful qualitative and quantitative consequences. It displays, for example, the simple ‘multiplicative’ way that the parameters enter the dependency structure of the model. If  $Q$  is concentrated on the zero vector, in which case we have thinning only, then, for  $A = A_1 \times \dots \times A_d$ ,

$$\nu(Q)(A) = \int \delta_0(A - te) dt = \ell(\{t \in \mathbb{R} : te \in A\}) = \ell(A_1 \cap \dots \cap A_d),$$

which identifies  $\nu(\delta_0)$  as ‘Lebesgue measure on the diagonal’, i.e. the image of  $\ell$  under the transformation  $\mathbb{R} \ni x \mapsto (x, \dots, x) \in \mathbb{R}^d$ . As a consequence, components referring to disjoint time intervals become independent, as mentioned previously in Section 2.

**4. Dependence properties of TaS models.** The main conclusion with respect to the dependency of the component processes is the general structure of the components as overlapping sums of independent, Poisson distributed random variables. For example, for  $i \neq j$  we have

$$\begin{pmatrix} X_i(A_i) \\ X_j(A_j) \end{pmatrix} =_{\text{distr}} \begin{pmatrix} \xi_0 + \xi_1 \\ \xi_0 + \xi_2 \end{pmatrix}$$

with  $\xi_0, \xi_1, \xi_2$  independent and Poisson distributed. This type of multivariate Poisson distribution is discussed in Chapter 37 of Johnson, Kotz and Balakrishnan (1997). It follows immediately that

$$\text{cov}(X_i(A_i), X_j(A_j)) = \text{var}(\xi_0) = E\xi_0.$$

We can also use Theorem 1 to relate the covariance to the model parameters. For this, we define the tail measure  $q = (q_D)_{D \subset \mathbb{D}}$  associated with the thinning mechanism  $p = (p_D)_{D \subset \mathbb{D}}$  by

$$q_D := \sum_{D' \supset D} p_{D'} \quad \text{for all } D \subset \mathbb{D}.$$

In words:  $q_D$  is the probability that at least the components with indices in  $D$  survive, whereas  $p_D$  is the probability that exactly these survive. If the individual components survive independently with probability  $\rho$ , for example, then

$$p_D = \rho^{|D|}(1 - \rho)^{d - |D|}, \quad q_D = \rho^{|D|} \quad \text{for all } D \subset \mathbb{D}.$$

For  $D_0, D' \subset \mathbb{D}$  fixed with  $D_0 \subset D'$  the sets  $M(D, D'; A_1, \dots, A_d)$  with  $D_0 \subset D \subset D'$  are disjoint, and it is easy to check that

$$\sum_{D_0 \subset D \subset D'} M(D, D'; A_1, \dots, A_d) = B_1 \times \dots \times B_d \quad \text{with } B_i := \begin{cases} A_i, & \text{for } i \in D_0, \\ \mathbb{R}, & \text{otherwise.} \end{cases}$$

In particular, this union does not depend on  $D'$ . Hence, if  $\xi_0$  denotes as above the sum of the  $\xi$ -variables that are common to  $X_i(A_i)$  and  $X_j(A_j)$ ,

$$\begin{aligned} \text{cov}(X_i(A_i), X_j(A_j)) &= E\xi_0 \\ &= \sum_{D \supset \{i,j\}} \lambda \sum_{D' \supset D} p_{D'} \nu(Q)(M(D, D'; A_1, \dots, A_d)) \\ &= \lambda \sum_{D' \supset \{i,j\}} \sum_{\{i,j\} \subset D \subset D'} p_{D'} \nu(Q)(M(D, D'; A_1, \dots, A_d)) \\ &= \lambda \sum_{D' \supset \{i,j\}} p_{D'} \nu(Q) \left( \sum_{\{i,j\} \subset D \subset D'} M(D, D'; A_1, \dots, A_d) \right) \\ &= \lambda q_{\{i,j\}} \int P(t + Y_i \in A_i, t + Y_j \in A_j) dt, \end{aligned}$$

where  $Y = (Y_1, \dots, Y_d)$  is a random vector with distribution  $Q$  and  $P$  refers to the background probability space  $(\Omega, \mathcal{A}, P)$  on which all the random quantities are defined. Alternatively, we can write the integral as  $\nu(Q_{ij})(A_i \times A_j)$  if  $Q_{ij}$  denotes the distribution of the two-dimensional random vector  $(Y_i, Y_j)$ .

Obviously it follows that only non-negative correlations are possible in TaS models. We will now investigate the dependence structure in greater detail. It turns out that TaS models as a rule display some kind of strong positive dependence, namely association. Recall that a random vector  $\xi = (\xi_1, \dots, \xi_d)$  is called (positively) *associated* if

$$\text{cov}(f(\xi), g(\xi)) \geq 0$$

for all increasing functions  $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$ . We obtain the following result:

THEOREM 2 *Let  $X$  be a  $d$ -dimensional counting process of TaS type. Then*

- (i) *for any Borel subsets  $A_1, \dots, A_d$  of the real line,  $(X_1(A_1), \dots, X_d(A_d))$  is associated, and*
- (ii) *the components of  $X$  are independent if and only if  $p_D = 0$  for all  $D \subset \mathbb{D}$  with  $|D| \geq 2$ .*

PROOF: (i) The random variables  $\xi(D; A_1, \dots, A_d), \emptyset \neq D \subset \mathbb{D}$  are associated, since they are independent; see Theorem 2.1 in Esary et al. (1967).  $X_i(A_i)$  is a nondecreasing function (sum) of the random variables  $\xi(D; A_1, \dots, A_d), \emptyset \neq D \subset \mathbb{D}$ . Thus, according to property (P4) in Esary et al. (1967) the statement follows.

(ii) If  $p_D = 0$  for all  $D \subset \mathbb{D}$  with  $|D| \geq 2$ , the independence of the components follows from the construction. Now suppose there exists a  $D \subset \mathbb{D}$  with  $i, j \in D$  and  $p_D > 0$ . Then there also exist Borel sets  $A_i$  and  $A_j$  with  $\nu(Q_{ij})(A_i \times A_j) > 0$ . Thus,  $X_i(A_i)$  and  $X_j(A_j)$  are dependent.  $\square$

Theorem 2 now has the following implications. Suppose  $A_1, \dots, A_d$  are Borel subsets of the real line. The association property of the components yields

$$\begin{aligned} P(X_1(A_1) > x_1, \dots, X_d(A_d) > x_d) &\geq \prod_{i=1}^d P(X_i(A_i) > x_i) \\ &= P(X_1^\perp(A_1) > x_1, \dots, X_d^\perp(A_d) > x_d) \end{aligned}$$

for all  $x_1, \dots, x_d$ , where  $(X_1^\perp(A_1), \dots, X_d^\perp(A_d))$  has the same marginals as the random vector  $(X_1(A_1), \dots, X_d(A_d))$  but with independent components. Thus, a model with independent component processes underestimates the probability for a mutual occurrence of a large number of claims in arbitrary time intervals. Moreover, association implies *positive supermodular dependence* (see e.g. Christofides and Vaggelatos (2004)), i.e. we have

$$Ef(X_1(A_1), \dots, X_d(A_d)) \geq Ef(X_1^\perp(A_1), \dots, X_d^\perp(A_d))$$

for all functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  for which the expectations exist and which are supermodular. A function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is called *supermodular*, if

$$f(x) + f(y) \leq f(x \vee y) + f(x \wedge y)$$

for all  $x, y \in \mathbb{R}^d$  with  $x \vee y$  and  $x \wedge y$  denoting the componentwise maximum and minimum of  $x$  and  $y$  respectively. This property immediately implies some interesting ordering relations. To this end, let us recall some basic definitions from the theory of stochastic orders. For given random variables  $X, Y$  we say  $X \leq_{st} Y$  if  $Ef(X) \leq Ef(Y)$  for all increasing  $f : \mathbb{R} \rightarrow \mathbb{R}$  and  $X \leq_{cx} Y$  if  $Ef(X) \leq Ef(Y)$  for all convex  $f : \mathbb{R} \rightarrow \mathbb{R}$  whenever these expectations exist. For a comprehensive review we refer the reader to Müller and Stoyan (2002). For our counting variables we thus obtain

$$\sum_i X_i^\perp(A_i) \leq_{cx} \sum_i X_i(A_i)$$

which means that the situation is more risky measured in the convex ordering in case we have dependence. Moreover it holds that

$$\max\{X_i(A_i), i = 1, \dots, d\} \leq_{st} \max\{X_i^\perp(A_i), i = 1, \dots, d\}$$

which means that the maximum number of points in given sets is stochastically larger in the case with independent components. For these and further results see Chapter 3.9 in Müller and Stoyan (2002).

We now use Theorem 1 to obtain the global dependency structure of TaS models, by which we mean the distributional asymptotics of the random vector that counts the number of events of different type in a large time interval. We write  $N_d(\mu, \Sigma)$  for the  $d$ -dimensional normal distribution with mean vector  $\mu$  and covariance matrix  $\Sigma$ ; ' $\rightarrow_{\text{distr}}$ ' denotes convergence in distribution.

**THEOREM 3** *Let  $X = (X_1, \dots, X_d)$  be a  $d$ -dimensional counting process of TaS type with base rate  $\lambda$ , thinning mechanism  $p = (p_D)_{D \subset \mathbb{D}}$  and shift distribution  $Q$ . We assume that  $\int \|x\| Q(dx) < \infty$ . Then, as  $T \rightarrow \infty$ ,*

$$T^{-1/2} \begin{pmatrix} X_1([0, T]) - \lambda q_1 T \\ \vdots \\ X_d([0, T]) - \lambda q_d T \end{pmatrix} \rightarrow_{\text{distr}} N_d(0, \Sigma),$$

where  $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq d}$  is given by  $\sigma_{ij} = \lambda q_{\{i, j\}}$ ,  $1 \leq i, j \leq d$ .

**PROOF:** As in the proof of Theorem 1 let  $\eta(D, D'; A_1, \dots, A_d)$  be the number of events in the background process that contribute to the coefficient of  $e(D)$  in  $(X_1(A_1), \dots, X_d(A_d))$  and have survival mark  $D'$ . Let

$$\zeta_T(D, D') := \eta(D, D'; [0, T], \dots, [0, T]).$$

Then  $\zeta_T(D, D')$  is Poisson distributed with

$$E\zeta_T(D, D') = \lambda p_{D'} \nu(Q)(A_T(D, D'))$$

where

$$A_T(D, D') = B_1 \times \dots \times B_d \quad \text{with} \quad B_i := \begin{cases} [0, T], & \text{for } i \in D, \\ [0, T]^c, & \text{for } i \in D' \setminus D, \\ \mathbb{R}, & \text{otherwise.} \end{cases}$$

(We have  $A_T(D, D') = M(D, D'; [0, T], \dots, [0, T])$  in the notation introduced in Section 3.) Fubini's theorem gives

$$\nu(Q)(A_T(D, D')) = \int \int 1_{A_T(D, D')}(x - te) dt Q(dx).$$

In the case  $D = D'$  the value of the indicator function is equal to 1 if and only if  $-x_i \leq t \leq T - x_i$  for all  $i \in D$ . Hence the inner integral evaluates to

$$\int 1_{A_T(D, D')-te}(x) dt = T - \max_{i \in D} x_i + \min_{i \in D} x_i,$$

which, as a function of  $x \in \mathbb{R}_+^d$ , is bounded by  $T$ . We can therefore use dominated convergence to obtain that, for all  $D \subset \mathbb{D}$ ,  $D \neq \emptyset$ ,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \nu(Q)(A_T(D, D)) = 1.$$

The familiar asymptotics for Poisson distributions now provides

$$\frac{1}{T} (\zeta_T(D, D) - E\zeta_T(D, D)) \xrightarrow{\text{distr}} N(0, \lambda p_D).$$

In the case  $D \neq D'$  we have, because of  $D \neq \emptyset$ , an  $i \in D$  and a  $j \in D' \setminus D$ , so that  $x \in A_T(D, D') - te$  implies

$$-x_i \leq t \leq T - x_i \quad \text{and} \quad (t < x_j \quad \text{or} \quad t > T - x_j),$$

which leads to

$$\int \mathbf{1}_{A_T(D, D') - te}(x) dt \leq |x_i - x_j| \leq x_i + x_j \quad \text{for all } x_i, x_j \in \mathbb{R}_+.$$

Using the assumption that  $\int \|x\| Q(dx) < \infty$  and dominated convergence we obtain that  $\zeta_T(D, D')/T$  converges to 0 in probability.

We now regard  $\zeta_T = (\zeta_T(D, D'))_{\emptyset \neq D \subset D' \subset \mathbb{D}}$  as a random vector. The components of  $\zeta_T$  are independent and asymptotically normal in the sense that, after subtracting the mean and dividing by  $T$ , letting  $T \rightarrow \infty$  yields a limiting normal distribution, possibly degenerate (with variance 0). This implies that the vector as a whole is asymptotically normal. The vector  $(X_1([0, T]), \dots, X_d([0, T]))$  of interest is a (fixed and deterministic) linear transformation of  $\zeta_T$  by Theorem 1, hence also asymptotically normal.

Finally, using the calculation preceding the theorem and

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int P(t + Y_i \in [0, T], t + Y_j \in [0, T]) dt = \lim_{T \rightarrow \infty} \frac{1}{T} E\zeta_T(\{i, j\}, \{i, j\}) = 1$$

we obtain the asserted covariance structure.  $\square$

An important consequence of Theorem 3 is the fact that asymptotically the shift distribution  $Q$  does not matter, as long as its mean is finite. In particular, TaS models have the same global dependence structure as the pure thinning processes with  $Q = \delta_0$ , i.e. the Lévy models. The TaS models provide additional *local* features, but not *global* ones. In contrast, as explained in Section 2, in the construction by Pfeifer and Nešlehová (2004) we can have completely arbitrary dependence for the vector  $(X_1([0, T]), \dots, X_d([0, T]))$  but the local dependence is small if  $T$  is large.

**5. Towards Poisson process copulas.** In the previous sections we introduced TaS models as a class of multivariate counting processes  $X = (X_1, \dots, X_d)$  and discussed some of their properties. We now investigate the extent to which

such models arise naturally from the basic assumptions that the marginal processes  $X_1, \dots, X_d$  are Poisson processes with constant rate and that  $X$  is time shift stationary.

The definition of TaS models in Section 3 begins with a ‘parent’ Poisson process that describes the times of the triggering events. As explained in that section, we can alternatively regard a multivariate counting process  $X$  of TaS type as a superposition of independent Poisson processes, if we mark the time points of the parent process by the set  $D \subset \{1, \dots, d\}$  of coordinates of  $X$  where the triggering event results in a claim. The thinning part of the models thus leads to a superposition of independent Poisson processes; the dependence ‘across time’, which is our main interest, is entirely due to the shift part of the TaS models. In this section we therefore concentrate on one such process (no thinning, shifts only, so that in the marked Poisson process construction, all points have survival mark  $D = \mathbb{D}$ ) and we assume that  $X$  arises by projection from a Poisson point process  $N$  with state space  $\mathbb{R}^d$  in the sense that, for  $i = 1, \dots, d$ ,

$$X_i(B) = N(\pi_i^{-1}(B)) \quad \text{for all Borel sets } B \subset \mathbb{R}.$$

Here  $\pi_i : \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $x = (x_1, \dots, x_d) \mapsto x_i$ , denotes the projection onto the  $i^{\text{th}}$  coordinate. Let  $\nu$  be the intensity measure associated with  $N$ . What can be said about  $N$  and  $\nu$  if  $X$  is time shift stationary with constant rate marginals?

We assume throughout this section that the counting processes are locally finite with probability 1. In terms of insurance applications this means that we exclude the case where we could have an infinite number of claims in a finite time interval. This assumption implies that, if  $X$  arises by projection from a Poisson process with intensity measure  $\nu$ ,

$$\nu(\pi_i^{-1}([a_i, b_i])) < \infty \quad \text{for } i = 1, \dots, d, \quad a_i, b_i \in \mathbb{R}, \quad a_i < b_i.$$

For example, a  $d$ -dimensional Poisson process with constant positive intensity, i.e. with  $\nu = \lambda \ell^d$  where  $\ell^d$  denotes  $d$ -dimensional Lebesgue measure and  $\lambda > 0$ , would not generate a locally finite multivariate counting process if  $d > 1$ . To avoid trivialities at the other extreme we also assume that  $\nu \neq 0$ , i.e. with probability 1 there is at least one claim.

As  $X$  is a deterministic function of  $N$ , the distribution of  $N$  determines the distribution of  $X$ , hence the distribution of  $X$  is completely specified by  $\nu$ . Our first result in this section implies that, conversely, the distribution of  $X$  also specifies the intensity measure  $\nu$  and therefore the distribution of  $N$ . This aspect will be further expounded at the end of this section.

**THEOREM 4** *Let  $X = (X_1, \dots, X_d)$  and  $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_d)$  be multivariate counting processes that arise by projection from the Poisson processes  $N$  and  $\tilde{N}$  with intensity measures  $\nu$  and  $\tilde{\nu}$  respectively. Suppose that the random vectors*

$$(X_1(B_1), \dots, X_d(B_d)) \quad \text{and} \quad (\tilde{X}_1(B_1), \dots, \tilde{X}_d(B_d))$$

*have the same distribution for all sets of finite intervals  $B_i = [a_i, b_i]$ ,  $i = 1, \dots, d$ . Then  $N$  and  $\tilde{N}$  have the same distribution, i.e.  $\nu = \tilde{\nu}$ .*

PROOF: We show that the condition implies  $\nu(A) = \tilde{\nu}(A)$  for all  $A \in \mathcal{B}^d$  of the form  $A = B_1 \times \dots \times B_d$  with  $B_i = [a_i, b_i) \subset \mathbb{R}$ . Since this is a  $\cap$ -stable generator of  $\mathcal{B}^d$  the result then follows. Let  $\xi = (X_1(B_1), \dots, X_d(B_d))$ . We will show that  $\nu(A)$  can be obtained from the characteristic function  $\varphi_\xi$  associated with  $\xi$ . Equality in distribution implies equality of the characteristic functions, hence this would indeed lead to  $\nu(A) = \tilde{\nu}(A)$ .

It follows from the properties of a Poisson process similar as in the proof of Theorem 1 that there are disjoint sets  $A_D \in \mathcal{B}^d$ ,  $D \subset \mathbb{D}$  and  $D \neq \emptyset$ , which depend on  $B_1, \dots, B_d$ , such that

$$X_k(B_k) = \sum_{D \ni k} \xi_D \quad \text{for } k = 1, \dots, d,$$

where the  $\xi_D$  are Poisson distributed with respective (finite) parameters  $\lambda_D := \nu(A_D)$ ; in particular,  $A = A_{\mathbb{D}}$  and hence  $\nu(A) = \lambda_{\mathbb{D}}$ . Moreover, the  $\xi_D$  are mutually independent. We have to show that  $\lambda_{\mathbb{D}}$  can be obtained from the characteristic function  $\varphi_\xi$  of  $\xi$ . For  $t \in \mathbb{R}^d$  we obtain

$$t' \xi = \sum_{k=1}^d \sum_{k \in D} t_k \xi_D = \sum_{D \neq \emptyset} \xi_D \sum_{k \in D} t_k$$

and thus we have

$$\varphi_\xi(t) = E e^{it' \xi} = \prod_{D \neq \emptyset} E \exp\left(i \xi_D \sum_{k \in D} t_k\right) = \prod_{D \neq \emptyset} \varphi_{\xi_D}\left(\sum_{k \in D} t_k\right).$$

The characteristic function  $\varphi_{\xi_D}$  of the Poisson distributed random variable  $\xi_D$  is given by  $\varphi_{\xi_D}(\theta) = \exp(\lambda_D(e^{i\theta} - 1))$ ,  $\theta \in \mathbb{R}$ . Inserting we obtain

$$\varphi_\xi(t) = \exp\left(\sum_{D \neq \emptyset} \lambda_D (e^{i \sum_{k \in D} t_k} - 1)\right).$$

Differentiating  $\log \varphi_\xi(t)$  with respect to  $t_j$  gives

$$\frac{\partial}{\partial t_j} \log \varphi_\xi(t) = i \sum_{D \ni j} \lambda_D \exp\left(i \sum_{k \in D} t_k\right)$$

and we arrive at

$$\frac{\partial^d}{\partial t_d \dots \partial t_1} \log \varphi_\xi(t) = i^d \lambda_{\mathbb{D}} \exp\left(i \sum_{k=1}^d t_k\right).$$

Inserting  $t = 0$  we obtain the required formula for  $\lambda_{\mathbb{D}}$  in terms of  $\varphi_\xi$ .  $\square$

Note that there are no assumptions about the marginal processes in Theorem 4. We next relate the stationarity properties of  $X$  and  $N$  to each other. We recall that  $X$  is said to be time shift stationary if, for fixed  $B_1, \dots, B_d \in \mathcal{B}$ , the distribution of  $(X_1(B_1 + t), \dots, X_d(B_d + t))$  does not depend on  $t$ . If  $N$  is a  $d$ -dimensional point process we say that  $N$  is *spatially stationary with respect to diagonal shifts* if, for fixed  $A_1, \dots, A_k \in \mathcal{B}^d$ , the distribution of  $(N(A_1 + te), \dots, N(A_k + te))$  does not depend on  $t$ . Finally, a measure  $\nu$  on  $(\mathbb{R}^d, \mathcal{B}^d)$  is *invariant with respect to diagonal shifts* if  $\nu(A + te) = \nu(A)$  for all  $A \in \mathcal{B}^d$ .



**THEOREM 5** *Let  $X$  be a multivariate counting processes that arises by projection from a Poisson processes  $N$  with intensity measure  $\nu$ . Then the following statements are equivalent:*

- (i)  $X$  is time shift stationary,
- (ii)  $N$  is spatially stationary with respect to diagonal shifts,
- (iii)  $\nu$  is invariant with respect to diagonal shifts.

**PROOF:** Using (D1) and (D2) from Section 3 we see that (iii) implies that the distribution of the random vector  $(N(A_1 + te), \dots, N(A_k + te))$  does not depend on  $t$ , for fixed  $d$ -dimensional Borel sets  $A_1, \dots, A_k$ ,  $k \in \mathbb{N}$ . Because of the definition of  $X$  this in turn implies that the distribution of

$$(X_1(B_1 + t), \dots, X_d(B_d + t)) = (N(\pi_1^{-1}(B_1) + te), \dots, N(\pi_d^{-1}(B_d) + te))$$

does not depend on  $t$ , which is (i). Finally, to see that (i) implies (iii), we use Theorem 4 with  $X$  the original counting process and  $\tilde{X}$  the shifted process given by  $(\tilde{X}_1(B_1), \dots, \tilde{X}_d(B_d)) = (X_1(B_1 + t), \dots, X_d(B_d + t))$ .  $\square$

Our next result shows that intensity measures that are invariant with respect to diagonal shifts must be of a very specific form. As in Section 3, for a probability measure  $Q$  on  $\mathbb{R}^d$  we define the measure  $\nu(Q)$  by  $\nu(Q)(A) = \int Q(A - te) dt$ , a minor case of notational overloading in the present framework. Sometimes it is more convenient to work with integrals of functions than with measures of sets; the definition of  $\nu(Q)$  is equivalent to the requirement that

$$\int \phi(x_1, \dots, x_d) \nu(Q)(dx) = \int \int \phi(x_1 - t, \dots, x_d - t) Q(dx_1, \dots, dx_d) dt$$

for all non-negative and measurable functions  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ .

**THEOREM 6** *A measure  $\nu$  on  $(\mathbb{R}^d, \mathcal{B}^d)$  is invariant with respect to diagonal shifts if and only if it is of the form  $\nu = \lambda \nu(Q)$  for some  $\lambda > 0$  and some probability measure  $Q$  on  $\mathbb{R}^d$ .*

**PROOF:** Let  $\tilde{\nu}$  be the measure-theoretical image of  $\nu$  under the transformation

$$T : \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad (x_1, x_2, \dots, x_d) \mapsto (x_1, x_2 - x_1, \dots, x_d - x_1).$$

Let  $t \in \mathbb{R}$  be fixed and let  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  be a non-negative measurable function. Using  $\int \psi d\tilde{\nu} = \int \psi \circ T d\nu$  with  $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$  defined by  $\psi(x_1, \dots, x_d) = \phi(x_1 + t, x_2, \dots, x_d)$ , we obtain

$$\int \phi(x_1 + t, x_2, \dots, x_d) \tilde{\nu}(dx) = \int \phi(x_1 + t, x_2 - x_1, \dots, x_d - x_1) \nu(dx).$$

Further, the invariance of  $\nu$  implies that

$$\int \psi(x_1, \dots, x_d) \nu(dx) = \int \psi(x_1 - t, \dots, x_d - t) \nu(dx) \quad \text{for all } t \in \mathbb{R}$$

for all non-negative measurable functions  $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ . In particular, with  $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$  defined by  $\psi(x_1, \dots, x_d) = \phi(x_1 + t, x_2 - x_1, \dots, x_d - x_1)$ ,

$$\begin{aligned} \int \phi(x_1 + t, x_2 - x_1, \dots, x_d - x_1) \nu(dx) &= \int \phi((x_1 - t) + t, (x_2 - t) - (x_1 - t), \dots, (x_d - t) - (x_1 - t)) \nu(dx) \\ &= \int \phi(x_1, x_2 - x_1, \dots, x_d - x_1) \nu(dx) \\ &= \int \phi(x_1, x_2, \dots, x_d) \tilde{\nu}(dx). \end{aligned}$$

Putting these two arguments together we see that

$$\int \phi(x_1 + t, x_2, \dots, x_d) \tilde{\nu}(dx) = \int \phi(x_1, x_2, \dots, x_d) \tilde{\nu}(dx)$$

for all  $t \in \mathbb{R}$  and for all non-negative measurable functions  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ . This shows that  $\tilde{\nu}$  is invariant under shifts in the direction of the first coordinate.

Further, because of

$$\begin{aligned} \tilde{\nu}(A \times \mathbb{R}^{d-1}) &= \int 1_{A \times \mathbb{R}^{d-1}}(x_1, x_2, \dots, x_d) \tilde{\nu}(dx) \\ &= \int 1_{A \times \mathbb{R}^{d-1}}(x_1, x_2 + x_1, \dots, x_d + x_1) \tilde{\nu}(dx) \\ &= \int 1_{A \times \mathbb{R}^{d-1}}(x_1, x_2, \dots, x_d) \nu(dx) \\ &= \nu(A \times \mathbb{R}^{d-1}), \end{aligned}$$

the transformed measure  $\tilde{\nu}$  also has first marginal  $\tilde{\nu}_1 = \lambda\ell$ . For  $B \in \mathcal{B}_{d-1}$  fixed the measure  $A \mapsto \tilde{\nu}(A \times B)$  therefore has to be a multiple of Lebesgue measure, i.e. we have

$$\tilde{\nu}(A \times B) = \lambda\ell(A) Q_0(B)$$

where  $Q_0(B)$  is a function of  $B$  that has the property  $Q_0(\mathbb{R}^d) = 1$  because of the above statement on the first marginal measure  $\tilde{\nu}_1$  of  $\tilde{\nu}$ . Since  $B \rightarrow \tilde{\nu}([0, 1] \times B)$  also is a (finite, non-zero) measure, we see that  $Q_0$  is in fact a probability measure on  $(\mathbb{R}^{d-1}, \mathcal{B}^{d-1})$ .

Now let  $Q := \delta_0 \otimes Q_0$ ; we claim that  $\nu = \lambda\nu(Q)$ . For  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  measurable

and non-negative again, we have

$$\begin{aligned}
\int \phi d\nu &= \int \phi \circ T^{-1}(x) \tilde{\nu}(dx) \\
&= \int \phi(x_1, x_2 + x_1, \dots, x_d + x_1) \tilde{\nu}(dx) \\
&= \lambda \int \int \phi(x_1, x_2 + x_1, \dots, x_d + x_1) Q_0(dx_2, \dots, dx_d) dx_1 \\
&= \lambda \int \int \phi(-t, x_2 - t, \dots, x_d - t) Q_0(dx_2, \dots, dx_d) dt \\
&= \lambda \int \int \phi(x_1 - t, x_2 - t, \dots, x_d - t) (\delta_0 \otimes Q_0)(dx_1, dx_2, \dots, dx_d) dt \\
&= \lambda \int \int \phi(x_1 - t, x_2 - t, \dots, x_d - t) Q(dx_1, dx_2, \dots, dx_d) dt \\
&= \lambda \int \phi d\nu(Q),
\end{aligned}$$

which completes the first part of the proof.

Conversely, if  $\nu = \lambda\nu(Q)$ , then

$$\nu(A + te) = \lambda \int Q(A + te - t'e) dt' = \lambda \int Q(A - t'e) dt' = \nu(A),$$

where the second equality follows by substitution.  $\square$

Because of the cause-effect relationship between the background events and the claims of different types, it is natural in our application to think of shifts as moves forward in time. This leads to the requirement in the actuarial applications of the TaS models that  $Q$  be concentrated on the non-negative orthant  $\mathbb{R}_+^d$ , whereas the representation in Theorem 6 may involve shifts that can be negative. We now argue that the support condition on  $Q$  is a minor one in the sense that TaS models with this restriction on the shift distribution are dense in the full class. For this we need a concept of convergence for multivariate counting processes. We first recall some basic facts about the particular distance of probability distributions on which this concept relies. These can be found in many textbooks; see e.g. the appendix in Barbour, Holst and Janson (1992) for a comprehensive overview or Rachev (1991) for a thorough treatment in the larger context of probability metrics.

The total variation distance of two probability measures  $P$  and  $Q$  on some measurable space  $(E, \mathcal{A})$  is given by

$$d_{\text{TV}}(P, Q) := \sup_{A \in \mathcal{A}} |P(A) - Q(A)|.$$

For countable  $E$  we have

$$d_{\text{TV}}(P, Q) = \frac{1}{2} \sum_{x \in E} |P(\{x\}) - Q(\{x\})|.$$

If  $P, P_1, P_2, P_3, \dots$  are probability measures on  $(E, \mathcal{A})$  then we say that  $P_n$  converges to  $P$  in total variation as  $n \rightarrow \infty$  if  $\lim_{n \rightarrow \infty} d_{\text{TV}}(P_n, P) = 0$ , and then  $\lim_{n \rightarrow \infty} \int f dP_n = \int f dP$  for all bounded and measurable functions  $f : E \rightarrow \mathbb{R}$ . By a slight abuse of notation we will also apply this terminology to random variables and, for example, say that  $\xi_n$  converges to  $\xi$  in total variation if this holds for the respective distributions. For discrete real random variables, especially for counting variables which take their values in  $\mathbb{N}_0$ , convergence in total variation is equivalent to the familiar concept of convergence in distribution. In particular, if  $\xi, \xi_1, \xi_2, \dots$  are Poisson distributed with finite parameters  $\lambda, \lambda_1, \lambda_2, \dots$  respectively, then  $\lim_{n \rightarrow \infty} \lambda_n = \lambda$  implies convergence in total variation of  $\xi_n$  to  $\xi$  as  $n \rightarrow \infty$ . Finally, if  $\xi_{n1}, \dots, \xi_{nk}$  are independent and  $\xi_{ni}$  converges to  $\xi_i$  in total variation as  $n \rightarrow \infty$  for  $i = 1, \dots, k$ , then  $f(\xi_{n1}, \dots, \xi_{nk})$  converges in total variation to  $f(\xi_1, \dots, \xi_k)$  for any measurable function  $f : E^k \rightarrow \mathbb{R}$ .

**THEOREM 7** *Let  $X = (X_1, \dots, X_d)$  be a time shift stationary multivariate counting process that arises by projection from a Poisson point process with intensity measure  $\nu$ . Then there exists a sequence  $(Q_n)_{n \in \mathbb{N}}$  of probability measures on the non-negative orthant  $\mathbb{R}_+^d$  such that the TaS processes  $X^{(n)}$  with rate  $\lambda := \nu([0, 1] \times \mathbb{R}^{d-1})$ , thinning mechanism  $(p_D)_{D \subset \mathbb{D}}$  given by  $p_{\mathbb{D}} = 1$  (i.e., no thinning) and shift distribution  $Q_n$  converge to  $X$  as  $n \rightarrow \infty$  in the sense that, for all Borel subsets  $B_1, \dots, B_d$  of the real line with finite Lebesgue measure, the random vector  $(X_1^{(n)}(B_1), \dots, X_d^{(n)}(B_d))$  converges in total variation to  $(X_1(B_1), \dots, X_d(B_d))$ .*

**PROOF:** By Theorems 5 and 6 we have  $\nu = \lambda\nu(Q)$  for some probability measure  $Q$  on  $\mathbb{R}^d$ ; we may assume that  $\lambda = 1$ . Let  $\tilde{Q}_n$  and  $Q_n$  be the distributions of

$$(\xi_1 \vee (-n), \dots, \xi_d \vee (-n)) \quad \text{and} \quad (n + \xi_1 \vee (-n), \dots, n + \xi_d \vee (-n))$$

respectively where  $\xi = (\xi_1, \dots, \xi_d)$  is a random vector with distribution  $Q$ . Clearly,  $Q_n$  is concentrated on  $\mathbb{R}_+^d$ . From the properties of the operator  $Q \mapsto \nu(Q)$  we obtain  $\nu(Q_n) = \nu(\tilde{Q}_n)$ , as  $Q_n$  arises from  $\tilde{Q}_n$  by a diagonal shift. Also, by the construction of  $\tilde{Q}_n$  via  $\xi$ ,

$$|\tilde{Q}_n(A) - Q(A)| \leq 1 - Q([-n, \infty)^d) \quad \text{for all } A \in \mathcal{B}^d,$$

which implies that  $\tilde{Q}_n$  converges to  $Q$  in total variation as  $n \rightarrow \infty$ .

Now let  $B_1, \dots, B_d \subset \mathbb{R}$  be as in the statement of the theorem. As in the proofs of Theorem 1 and Theorem 4 we can write  $(X_1^{(n)}(B_1), \dots, X_d^{(n)}(B_d))$  as a fixed deterministic function of the random vector  $(N^{(n)}(A_D))_{D \subset \mathbb{D}, D \neq \emptyset}$ , with  $A_D \in \mathcal{B}^d$  disjoint and depending on  $B_1, \dots, B_d$ , and similarly for  $X$ . The components of these vectors are independent Poisson random variables with parameters given by  $\nu(Q_n)(A_D)$ ,  $n \in \mathbb{N}$ , and  $\nu(Q)(A_D)$  respectively. The sets  $A_D$  arising in the decomposition are  $d$ -dimensional rectangles  $C_1 \times \dots \times C_d$  where at least one of the  $C_j$ 's has finite Lebesgue measure. In particular,  $x \mapsto \int 1_{A_D - x}(te) dt$  is bounded by  $\ell(C_j)$ , hence the total variation convergence of  $\tilde{Q}_n$  to  $Q$  together with

$$\nu(Q_n)(A_D) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} 1_{A_D - x}(te) dt \tilde{Q}_n(dx)$$

yields the convergence of  $\nu(Q_n)(A_D)$  to  $\nu(Q)(A_D)$ . This in turn implies the convergence in total variation of  $N^{(n)}(A_D)$  to  $N(A_D)$  for all  $D \subset \mathbb{D}$ ,  $D \neq \emptyset$ . Using the last of the properties of total variation convergence given before the statement of the theorem we now obtain the assertion.  $\square$

In summary these results show that within the class of multivariate counting processes generated by Poisson point processes the requirements of constant rate marginals and time shift stationarity essentially lead to the class of TaS models.

We close this section with some remarks, related to the above material or making use of the ideas underlying the above proofs and constructions.

REMARKS (a) If  $Q$  has a density  $f$  with respect to  $d$ -dimensional Lebesgue measure  $\ell^d$  then  $\nu(Q)$  also has a density with respect to  $\ell^d$ . Suppose that the random vector  $Y = (Y_1, \dots, Y_d)$  has density  $f$  and let  $\tilde{f}$  be the density of the  $(d-1)$ -dimensional random vector  $(Y_2 - Y_1, \dots, Y_d - Y_1)$ . Then  $x = (x_1, \dots, x_d) \mapsto \tilde{f}(x_2 - x_1, \dots, x_d - x_1)$  is a density of  $\nu(Q)$ . To see this, let again  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  be measurable and non-negative. Then

$$\begin{aligned} & \int \phi d\nu(Q) \\ &= \int \cdots \int \phi(x_1 - t, \dots, x_d - t) f(x_1, \dots, x_d) dx_1 \cdots dx_d dt \\ &= \int \cdots \int \phi(x_1, \dots, x_d) \left( \int f(x_1 + t, \dots, x_d + t) dt \right) dx_1 \cdots dx_d, \\ &= \int \cdots \int \phi(x_1, \dots, x_d) \left( \int f(t, x_2 - x_1 + t, \dots, x_d - x_1 + t) dt \right) dx_1 \cdots dx_d. \end{aligned}$$

On the other hand, it is easy to see that

$$(z_2, \dots, z_d) \mapsto \int f(t, z_2 + t, \dots, z_d + t) dt$$

is a density for  $(Y_2 - Y_1, \dots, Y_d - Y_1)$ . This formula for the density of  $\nu(Q)$  can be used to simplify some calculations, such as in Section 7 below.

(b) The parametrization of TaS models by a base intensity  $\lambda$ , the survival probabilities  $p_D$ ,  $D \subset \mathbb{D}$ , and the shift distribution  $Q$  is convenient but somewhat redundant. For example, we could demand that  $p_\emptyset = 0$  since otherwise a proper change in the base rate together with the corresponding change in the survival probabilities would lead to the same stochastic process. Similarly, a shift of  $Q$  by a scalar multiple of  $e = (1, \dots, 1)$  leaves  $\nu(Q)$  unaffected and hence would also lead to the same multivariate counting process. Informally, the points of the background process may be moved backwards and forwards in time and this can be undone by a corresponding change in the delay times. This simple fact is the basis for the denseness property of TaS models formalized in Theorem 7. Such redundancies are also important in connection with statistical estimation of the model parameters; this aspect of the TaS models will be further investigated in a separate paper dealing with statistical issues.

(c) Formally, a counting process  $X$  on  $\mathbb{R}$  is a measurable function, defined on some probability space and with values in the space  $\mathcal{M}(\mathbb{R})$  of  $\mathbb{N}_0$ -valued, locally finite measures  $\mu$  on  $(\mathbb{R}, \mathcal{B})$ , where the  $\sigma$ -field on the latter is the one generated by the ‘projections’  $\mu \mapsto \mu(B)$ ,  $B \in \mathcal{B}$ . The finite-dimensional distributions of  $X$  are the distributions of the random vectors

$$(X(B_1), \dots, X(B_k)), \quad k \in \mathbb{N}, B_1, \dots, B_k \in \mathcal{B}.$$

These determine the distribution of  $X$ , see e.g. Section 6.2 in Daley and Vere-Jones (1988). To obtain multivariate counting processes we essentially ‘vectorize’ the one-dimensional concept. This means that the family of finite-dimensional distributions associated with the  $d$ -dimensional counting process  $X = (X_1, \dots, X_d)$  now relates to the distributions of

$$(X_1(B_{11}), \dots, X_1(B_{1l_1}), X_2(B_{21}), \dots, X_2(B_{2l_2}), \dots, X_d(B_{d1}), \dots, X_d(B_{dl_d}))$$

with  $B_{11}, \dots, B_{dl_d} \in \mathcal{B}$ . It is a consequence of Theorem 4 that in the case that  $X$  arises by projection from a multivariate Poisson process, it is enough to consider the random vectors  $(X_1(B_1), \dots, X_d(B_d))$ ,  $B_1, \dots, B_d \in \mathcal{B}$ ; in fact, we may take these sets to be intervals.

(d) It has already been mentioned at the beginning of Section 2 that the straightforward translation of the copula idea from random vectors to multivariate counting processes leads to two major problems. Using our general approach via Poisson point processes we now discuss the second of these. For simplicity we suppose that  $X = ((X_{1,t}, X_{2,t}))_{t \geq 0}$  is a two-dimensional counting process, where counting starts at time  $t = 0$  and  $X_{i,t}$  is the number of events of type  $i = 1, 2$  up to and including time  $t$ . Then, for each  $t \geq 0$ ,  $(X_{1,t}, X_{2,t})$  is an ordinary two-dimensional random vector; let  $C_t$  be an associated copula. We now show, by giving an example, that the family  $(C_t)_{t \geq 0}$ , together with the distribution of the marginal processes, does not suffice to determine the distribution of  $X$ . Our starting point is an ordinary static two-dimensional copula  $C$ , which we assume to be non-symmetric, i.e.  $C(a, b) \neq C(b, a)$  for some  $a, b \in (0, 1)$ . Let  $N$  be a Poisson point process on the unit square  $(0, 1) \times (0, 1)$  with intensity measure  $\nu$  given by

$$\nu((0, s] \times (0, t]) = C(s, t) \quad \text{for all } s, t \in (0, 1) \times (0, 1).$$

We define  $X = ((X_{1,t}, X_{2,t}))_{t \geq 0}$  by

$$X_{1,t} := N((0, t] \times (0, 1]), \quad X_{2,t} := N((0, 1] \times (0, t])$$

(the counting processes remain constant from time  $t = 1$  onwards). For the second process  $\tilde{X} = ((\tilde{X}_{1,t}, \tilde{X}_{2,t}))_{t \geq 0}$  we simply switch components, i.e.,  $\tilde{X}_{1,t} := X_{2,t}$  and  $\tilde{X}_{2,t} := X_{1,t}$  for all  $t \geq 0$ . We claim that  $X$  and  $\tilde{X}$  generate the same family of copulas in the above sense, but that  $X$  and  $\tilde{X}$  do not have the same distribution. To see this, we first fix  $t \in (0, 1)$ . Then, as in the proof of Theorem 1,

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} \xi_0 + \xi_1 \\ \xi_0 + \xi_2 \end{pmatrix}, \quad \begin{pmatrix} \tilde{X}_{1,t} \\ \tilde{X}_{2,t} \end{pmatrix} = \begin{pmatrix} \xi_0 + \xi_2 \\ \xi_0 + \xi_1 \end{pmatrix},$$

where  $\xi_0, \xi_1, \xi_2$  are independent Poisson random variables with respective parameters  $\lambda_0, \lambda_1, \lambda_2$  given by

$$\lambda_0 = C(t, t), \quad \lambda_1 = t - C(t, t) \quad \text{and} \quad \lambda_2 = t - C(t, t).$$

Because of  $\lambda_1 = \lambda_2$  this implies that the two random vectors  $(X_{1,t}, X_{2,t})$  and  $(\tilde{X}_{1,t}, \tilde{X}_{2,t})$  have the same distribution and therefore lead to the same copula. On the other hand,

$$\begin{pmatrix} X_{1,a} \\ X_{2,b} \end{pmatrix} = \begin{pmatrix} \xi_0 + \xi_1 \\ \xi_0 + \xi_2 \end{pmatrix}, \quad \begin{pmatrix} \tilde{X}_{1,a} \\ \tilde{X}_{2,b} \end{pmatrix} = \begin{pmatrix} \tilde{\xi}_0 + \tilde{\xi}_1 \\ \tilde{\xi}_0 + \tilde{\xi}_2 \end{pmatrix}.$$

Again,  $\xi_0, \xi_1, \xi_2$  are independent Poisson random variables, now with parameters given by

$$\lambda_0 = C(a, b), \quad \lambda_1 = a - C(a, b) \quad \text{and} \quad \lambda_2 = b - C(a, b),$$

and  $\tilde{\xi}_0, \tilde{\xi}_1, \tilde{\xi}_2$  are independent Poisson random variables with the respective parameters

$$\tilde{\lambda}_0 = C(b, a), \quad \tilde{\lambda}_1 = a - C(b, a) \quad \text{and} \quad \tilde{\lambda}_2 = b - C(b, a).$$

In particular,

$$\text{var}(X_{1,a} + X_{2,b}) = 2C(a, b) + a + b \neq 2C(b, a) + a + b = \text{var}(\tilde{X}_{1,a} + \tilde{X}_{2,b}),$$

which shows that the random vectors  $(X_{1,a}, X_{2,b})$  and  $(\tilde{X}_{1,a}, \tilde{X}_{2,b})$  have different distributions. In the terminology of the previous part we therefore see that a further reduction in the family of random vectors  $(X(B_1), \dots, X(B_d))$  to the sets  $B_1, \dots, B_d = (0, t], t \geq 0$ , would destroy the property that the distributions characterize the distribution of  $X$ , even in the case that  $X$  arises by projection from a Poisson process.

(e) Multivariate Poisson processes with constant intensity marginals are in a way analogous to multivariate random variables with uniform marginals, i.e. we could consider the corresponding intensity measure  $\nu$  as a Poisson process copula. Theorem 7 characterizes those copulas that lead to counting processes that are time shift stationary. It is well known that a non-homogeneous one-dimensional Poisson processes can be obtained from a homogeneous such process with a suitable time transformation, similar to the quantile transformation that underlies the usefulness of classical copulas. At present, this observation may be used to serve as a guideline in the construction of suitable models. Due to the fact that the transformations of the marginals refer to time, which will in general lead to different ‘speeds’ in the respective component processes, it seems that the practical consequences of this observation may be of limited value.

**6. Actuarial applications.** It seems that nowadays models in insurance are towards a broader perspective on assets and liabilities which particularly takes into account dependencies between processes. As possible applications of our general TaS models we suggest the following situations:

*6.1 Different lines of business.* Suppose the points of the base process represent time points of natural catastrophes like earthquakes, floods, hurricanes etc. An immediate damage to houses and cars which produces claims in the non-life insurance branch is then most likely followed by claims in health insurance due to epidemics or general bad health conditions after the catastrophe. Thus,  $X_1$  and  $X_2$  would count the claims in non-life and health insurance respectively. In the context of this application, an extension of our models, where a single event may generate more than one event in the individual component processes, may be of interest.

*6.2 Reinsurance.* Consider an insurance company and its main reinsurer which are connected via e.g. a stop-loss treaty. The reinsurer pays up to a specified limit, all claims which in total exceed a certain retention within one year. The points of the base process correspond here to the claim arrival time points for the insurer and these may also be claim arrival time points for the reinsurer. Thus,  $X_1$  and  $X_2$  would count the claims for the insurer and for the reinsurer respectively.

*6.3 Late claims.* A somewhat different application of the TaS model is as follows: Suppose we add up all components and obtain a new counting process  $\hat{X}$ , i.e.  $\hat{X} = \sum_i X_i$ . Note that  $\hat{X}$  is not a Poisson process anymore due to the dependence in the components. The interpretation is as follows: A point at time  $t$  with mark  $D$  in the base process corresponds to an event which triggers a series of claims which arrive at time points  $t + Y_{i_1}, \dots, t + Y_{i_{|D|}}$  where  $Y \sim Q$  and  $D = \{i_1, \dots, i_{|D|}\}$ . This situation is well-known from the IBNR (incurred but not reported) problem. Often insured events need several years until they are settled completely. Here we assume that at most  $d$  payments are needed in order to settle all claims arising from the same event.

For the expected total number  $E\hat{X}(A)$  of claims with time stamps in  $A$  the possible non-independence of the component processes is irrelevant and we simply obtain  $E\hat{X}(A) = \lambda d \ell(A)$  if there is no deletion. Beyond expected values, Theorem 1 yields the distributional representation

$$\hat{X}(A) \stackrel{= \text{distr}}{=} 1 \cdot Z_1 + 2 \cdot Z_2 + \dots + d \cdot Z_d,$$

where  $Z_1, \dots, Z_d$  are independent random variables, each with a Poisson distribution with parameter depending on  $A$ ; see Section 7 for an example.

*6.4 Distribution of the next claim arrival.* An interesting quantity for insurance applications is the time of the first claim arrival after time 0 say, which we denote by  $\tau$ . Let  $X = (X_1, \dots, X_d)$  be a  $d$ -dimensional counting process of TaS type with base rate  $\lambda$ , thinning mechanism  $p = (p_D)_{D \subset \mathbb{D}}$  and shift distribution  $Q$ . Formally, we define

$$\tau = \inf\{t \geq 0 : X_i((0, t]) > 0 \text{ for some } i \in \mathbb{D}\}.$$

The distribution of  $\tau$  can be expressed with the help of the model parameters as follows:

$$\begin{aligned} P(\tau > t) &= P(X_1([0, t]) = 0, \dots, X_d([0, t]) = 0) \\ &= P(\xi(D; [0, t], \dots, [0, t]) = 0, \forall D \subset \mathbb{D}, D \neq \emptyset) \\ &= \prod_{D \subset \mathbb{D}, D \neq \emptyset} P(\xi(D; [0, t], \dots, [0, t]) = 0) \end{aligned}$$



$$\begin{aligned}
&= \exp\left(-\lambda \sum_{D \subset \mathbb{D}, D \neq \emptyset} \sum_{D' \supset D} p_{D'} \nu(Q)(M(D, D'; [0, t], \dots, [0, t]))\right) \\
&= \exp\left(-\lambda \sum_{D' \subset \mathbb{D}, D' \neq \emptyset} \sum_{D \subset D', D \neq \emptyset} p_{D'} \nu(Q)(M(D, D'; [0, t], \dots, [0, t]))\right) \\
&= \exp\left(-\lambda \sum_{D' \subset \mathbb{D}, D' \neq \emptyset} p_{D'} \nu(Q)\left(\bigcup_{i \in D'} \pi_i^{-1}([0, t])\right)\right).
\end{aligned}$$

Note that  $\tau$  is in general not exponentially distributed; again, Section 7 contains an example.

For further applications, also to portfolio credit risk, we refer to Lindskog and McNeil (2003).

**7. Computational issues and examples.** Suppose we are given a TaS model  $X$  with parameters  $\lambda$ ,  $(p_D)_{D \subset \mathbb{D}}$  and  $Q$ . How can we obtain quantities of interest (expectations, probabilities etc.) for such a model? Under the heading ‘computational’ we summarize Monte Carlo simulations, standard numerical procedures but also algorithmic issues. With respect to the first, it should be clear how to simulate values from the random variables of interest once it is known how to simulate values from the shift distribution. As so often, stochastic simulation is easy to carry out and can be an invaluable tool for getting a ‘feel’ for the model, but its accuracy is limited. Also, the dependence of the values of interest on the model parameters is normally beyond the reach of plain Monte Carlo procedures. With standard numerical methods we may at least hope for better accuracy. A typical case is that of evaluating the probabilities associated with overlapping sums of independent Poisson variables that appear in Theorem 1 and its applications; this is discussed in Chapter 37 of Johnson, Kotz and Balakrishnan (1997).

First we consider the case  $d = 2$ ,  $Y = (Y_1, Y_2)$  with  $Y_1 \equiv 0$ ,  $Y_2 \geq 0$ , so that  $Q = \delta_0 \otimes Q_0$  with  $Q_0$  concentrated on  $[0, \infty)$ . In the actuarial framework this means that there are only two types of claims and that each claim of type 1 triggers a later claim of type 2. We now investigate the covariance of  $X_1([0, t])$  and  $X_2([0, t])$ . Using the formula in Section 4 we obtain

$$\begin{aligned}
\text{cov}(X_1([0, t]), X_2([0, t])) &= \lambda q_{\{1,2\}} \int P(x + Y_1 \in [0, t], x + Y_2 \in [0, t]) dx, \\
&= \lambda q_{\{1,2\}} \int_0^t P(x + Y_2 \in [0, t]) dx \\
&= \lambda q_{\{1,2\}} \int_0^t P(Y_2 \leq t - x) dx \\
&= \lambda q_{\{1,2\}} \int_0^t P(Y_2 \leq x) dx \\
&= \lambda q_{\{1,2\}} \left(t - \int_0^t P(Y_2 > x) dx\right).
\end{aligned}$$

From this representation of the covariance we immediately see that whenever  $Q_0 \leq_{st} Q'_0$  we obtain

$$\text{cov}(X_1([0, t]), X_2([0, t])) \geq \text{cov}(X_1([0, t]), X'_2([0, t])).$$

This means that stochastically larger shifts lead to a smaller covariance. This is of course what one expects. However, we also get the following somewhat more surprising result. Suppose that  $Q_0$  is a distribution with support  $[0, \infty)$  and finite first moment  $\mu$ . Then we have

$$\begin{aligned}\text{cov}(X_1([0, t]), X_2([0, t])) &= \lambda q_{\{1,2\}} \left( t - \int_0^t P(Y_2 > x) dx \right) \\ &= \lambda q_{\{1,2\}} \left( t - \mu + \int_t^\infty P(Y_2 > x) dx \right) \\ &= \lambda q_{\{1,2\}} (t - \mu + E(Y_2 - t)_+).\end{aligned}$$

Due to the properties of the convex order we thus obtain that  $Q_0 \leq_{cx} Q'_0$  implies

$$\text{cov}(X_1([0, t]), X_2([0, t])) \leq \text{cov}(X_1([0, t]), X'_2([0, t])).$$

This means that more variability in the shift distribution (leaving the mean unchanged) leads to a higher covariance.

We now consider a special parametric subclass of the general TaS model, without the above support assumption on  $Q$  but still in dimension  $d = 2$ , where some general computations can be carried out on the basis of the results of the previous sections. In this subclass we assume independent survival of the components, so that  $p_{\{1\}} = p_{\{2\}} = \rho(1 - \rho)$ ,  $p_{\{1,2\}} = \rho^2$  and  $p_\emptyset = (1 - \rho)^2$ . We further assume that the shifts are independent and both exponentially distributed with the same parameter  $\alpha$ , i.e. with  $\text{Exp}(\alpha)$  for the latter distribution we have  $Q = \text{Exp}(\alpha) \otimes \text{Exp}(\alpha)$ . Models from this parametric TaS class are therefore completely specified by the three (real) parameters  $\lambda$ ,  $\rho$  and  $\alpha$ ; our aim is to obtain some quantities of interest explicitly as functions of these.

It is known that, with  $Y_1, Y_2$  independent and  $\mathcal{L}(Y_1) = \mathcal{L}(Y_2) = \text{Exp}(\alpha)$ , the distribution of  $Y_1 - Y_2$  is two-sided exponential with parameter  $\alpha$ . Together with Remark (a) at the end of Section 5 this gives the density

$$(x_1, x_2) \mapsto \frac{\alpha}{2} e^{-\alpha|x_1 - x_2|}, \quad x_1, x_2 \in \mathbb{R},$$

for  $\nu(Q)$ . In particular, for all  $t > 0$ ,

$$\nu(Q)([0, t] \times [0, t]) = \int_0^t \int_0^t \frac{\alpha}{2} e^{-\alpha|x_1 - x_2|} dx_1 dx_2 = t - \frac{1 - e^{-\alpha t}}{\alpha}.$$

The formula for the covariance given at the beginning of Section 4 leads to

$$\text{cov}(X_1([0, t]), X_2([0, t])) = \lambda \rho^2 \left( t - \frac{1 - e^{-\alpha t}}{\alpha} \right).$$

We next consider the time  $\tau$  of the first claim arrival after time 0. Because of  $\mathbb{D} = \{1, 2\}$  the last sum in the formula for  $P(\tau > t)$  given in Section 6.4 reduces to the cases  $\{1\}$ ,  $\{2\}$  and  $\{1, 2\}$ . As both marginal measures of  $\nu(Q)$  are equal to  $\ell$ , we have

$$\nu(Q)(\pi_1^{-1}([0, t])) = \nu(Q)([0, t] \times \mathbb{R}) = t$$

and similarly  $\nu(Q)(\pi_2^{-1}([0, t])) = t$ . Further, with the computation used above for the covariance,

$$\begin{aligned} \nu(Q)(\pi_1^{-1}([0, t]) \cup \pi_2^{-1}([0, t])) \\ &= \nu(Q)([0, t] \times \mathbb{R}) + \nu(Q)(\mathbb{R} \times [0, t]) - \nu(Q)([0, t] \times [0, t]) \\ &= t + \frac{1 - e^{-\alpha t}}{\alpha}, \end{aligned}$$

hence

$$P(\tau > t) = \exp\left(-\lambda\left(2\rho t - \rho^2 t + \frac{\rho^2}{\alpha}(1 - e^{-\alpha t})\right)\right).$$

As announced at the end of Section 6.4, this is not an exponential distribution. Finally, we consider the total number  $\hat{X}([0, t]) = X_1([0, t]) + X_2([0, t])$  of claims in the time interval  $[0, t]$ . In the notation that we used at the beginning of Section 4 we can write this as  $\hat{X}([0, t]) = 2\xi_0 + \xi_1 + \xi_2$ , so that

$$\hat{X}([0, t]) = Z + 2Z'$$

with independent random variables  $Z$  and  $Z'$ . Here  $Z$  and  $Z'$  both have a Poisson distribution and, using similar calculations as for the covariance and the arrival time,

$$EZ' = \lambda\rho^2 t - \frac{\lambda\rho^2}{\alpha}(1 - e^{-\alpha t}), \quad EZ = 2\lambda\rho(1 - \rho)t + \frac{2\lambda\rho^2}{\alpha}(1 - e^{-\alpha t}).$$

It is known that distributions of this type are compound Poisson, which means that Panjer recursion or transform methods can be applied to obtain numerical values for the probabilities  $P(\hat{X}([0, t]) = k)$ ,  $k \in \mathbb{N}_0$ , for any specific values of the parameters  $\lambda$ ,  $\rho$  and  $\alpha$ .

In summary, we have introduced a flexible class of models for multivariate counting processes. The underlying structural assumptions of this class appear to be quite natural in a variety of insurance applications. Some general results have been obtained, in particular for the covariance of the component processes and the waiting time for the next claim. These aspects can be related to the stochastic ordering of the shift distributions that are part of the model specification. We have also shown that these models arise in a natural way in connection with time shift stationarity. Finally, at least in some simple but prototypical cases some explicit computations can be carried out that reveal the dependence of quantities of interest on the model parameters.

**Acknowledgement.** The second author is grateful to Paul Embrechts for stimulating discussions and to the ETH Zürich for hospitality and support. We are also grateful to two anonymous referees for their excellent work and their detailed suggestions, which considerably improved the paper.

## References

- Christofides, T.C. and Vaggelatos E.A. (2004) Connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis* **88**, 138-151.
- Cont, R. and Tankov, P. (2004) *Financial Modelling with Jump Processes*. Chapman and Hall, Boca Raton.
- Cox, D.R. and Isham, V. (1980) *Point Processes*. Chapman and Hall, London.
- Daley, D.J. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes*. Springer, New York.
- Esary, J.D., Proschan, F. and Walkup, D.W. (1967) Association of random variables with applications. *Ann. Math. Stat.* **38**, 1466-1474.
- Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997) *Discrete Multivariate Distributions*. Wiley, New York.
- Lindskog, F. and McNeil, A.J. (2003) Common Poisson shock models: applications to insurance and credit risk modelling. *ASTIN Bulletin* **33**, 209-238.
- McNeil, A.J., Frey, R. and Embrechts, P. (2004) *Quantitative Risk Management: Concepts, Techniques and Tools*. Book in preparation.
- Müller, A. and Stoyan D. (2002) *Comparison Methods for Stochastic Models and Risks*. Wiley, New York.
- Nelsen, R.B. (1999) *An Introduction to Copulas*. Lecture Notes in Statistics 139. Springer, New York.
- Pfeifer, D. and Nešlehová, J. (2004) Modeling and generating dependent risk processes for IRM and DFA. *ASTIN Bulletin* **34**, 333-360.
- Resnick, S.I. (1987) *Extreme Values, Regular Variation and Point Processes*. Springer, New York.

Nicole Bäuerle and Rudolf Grübel  
Institut für Mathematische Stochastik  
Universität Hannover  
Postfach 60 09  
D-30060 Hannover  
e-mail: [baeuerle@stochastik.uni-hannover.de](mailto:baeuerle@stochastik.uni-hannover.de)  
[rgrubel@stochastik.uni-hannover.de](mailto:rgrubel@stochastik.uni-hannover.de)