9

Correlated Poisson Processes and Their Applications in Financial Modeling

Alexander Kreinin
Risk Analytics, IBM, Canada

9.1 Introduction

Multivariate risk factor models set the foundation of financial risk measurement. The modern financial theory often recommends jump-diffusion models to describe dynamics of the individual risk factors, such as interest rates, foreign exchange rates, stock indices, and volatility surfaces (Kou, 2002; Lipton and Rennie, 2008; Merton, 1976; Musiela and Rutkowsky, 2008), One of the most popular model of jumps, the Poisson model, requires introduction of a codependence structure in the multivariate setting.

The multivariate Gaussian diffusion models are traditionally popular in financial applications (Musiela and Rutkowsky, 2008). In this class of models, the dynamics of the risk factors are described by the Gaussian stochastic processes. The calibration problem in this case can be reduced to the estimation of the drift vector and the diffusion matrix describing the covariance structure of the risk factor space. The only constraint imposed on the covariance matrix is nonnegativity of its eigenvalues.

It is very well known that calibration of the models for equity derivatives pricing can be performed satisfactorily in the class of jump-diffusion processes (Kou, 2002; Kyprianou *et al*. 2005; Merton, 1974). Once the jump processes are introduced, the calibration problem becomes more challenging and, technically, more demanding. In particular, an admissible set of model parameters is described by more sophisticated conditions: not only must an observed covariance matrix have nonnegative eigenvalues but the elements of this matrix should also obey additional inequalities depending on the intensities of the jump processes. The nature

Financial Signal Processing and Machine Learning, First Edition.
Edited by Ali N. Akansu, Sanjeev R. Kulkarni and Dmitry Malioutov.

of these new constraints and their computation is discussed in Griffiths *et al*.(1979), Whitt (1976), and Duch *et al*. (2014).

The computation of the aditional constraints on the elements of the covariance matrix appeared to be closely related to the analysis of the extreme joint distributions having maximal and minimal correlations. Our approach to the computation of the extreme joint distributions is, in spirit, very close to that developed by Frechet (1951) and Whitt (1976). The only difference is that we propose a pure probabilistic method for the computation of the joint probabilities, based on the Strong Law of Large Numbers and a well-known result on optimal permutations that can be derived from Hardy *et al*. (1952).

Practitioners often set the calibration problem as a matching of the intensities and correlations of the components. One of the difficulties of this problem for multivariate Poisson processes is presented by negative elements of the correlation matrix. We propose a solution to this problem having the following two ingredients:

1. A construction of the joint distribution with extreme correlations at some future point in time, usually at the end of time horizon, T.
2. A backward simulation of Poisson processes in the interval $[0, T]$.

The resulting multivariate Poisson process is a Markov process but not infinitely divisible.[1]

The chapter is organized as follows. In Section 9.2, we discuss some classes of risk factor models with jump processes and their existing and potential applications in different areas of financial risk management. In this section, we discuss some technical obstacles in calibration associated with negative correlations. In Section 9.3, the common shock model and the mixed Poisson model are presented in the bivariate case. The multivariate case of the CSM is discussed in Section 9.4. The first model, traditionally used in the actuarial science, introduces correlations by the simultaneous jumps in the components of the process. This model is too restrictive; it does not solve the problem with negative correlations. The second model introduces a much more general class of processes having time-dependent correlations between the components. This model addresses the negative correlation problems but might have a slow convergence to the stationary distribution, making the calibration problematic.

In Section 9.4, we describe the backward simulation algorithm and discuss the calibration problem for the parameters of the model. We also compare the backward simulation approach with the forward simulation. One of the ingredients of the solution, the computation of the lower and upper bounds for the correlation coefficient, is described in Section 9.5, where the computation of the extreme joint distribution and its support is presented. The computation of the extreme distribution allows one to find the boundaries for the correlation coefficient of the corresponding components of the multivariate Poisson process. We also describe a numerical scheme for the approximation of the joint distribution and computation of the boundaries for the correlation coefficients. These results, leading to the additional necessary conditions on the elements of the correlation matrix, are obtained with the help of a version of the Frechet–Hoeffding theorem[2] presented in Section 9.5.4.

Implementation of the algorithm for the computation of the extreme distributions requires continuity bounds of the approximation of the joint distribution by a distribution with bounded support. These bounds are obtained in Section 9.5.5.

[1] Despite that each component of the process is an infinitely divisible Poisson process.

[2] This version of the theorem is new, to the best of our knowledge.

In Section 9.6, we demonstrate a few patterns of the support of a bivariate Poisson distribution typical for maximal and minimal correlations. These examples are computed using approximation of the joint extreme distribution analyzed in Section 9.5.5. In Section 9.7, we extend the backward simulation (BS) approach to the processes having both Poisson and Wiener components. The idea behind this extension is to bring together, in a common framework, both Poisson and Gaussian processes and to develop a unified simulation algorithm for scenario generation with diffusion and pure jump components. In the Poisson–Wiener case, the time structure of correlations under BS simulation appeared to be linear, as in the case of the multivariate Poisson processes. This feature of the BS simulation put distance between our approach and the traditional forward simulation keeping the correlation constant over time.

The chapter concludes with comments on possible extensions of the BS an method and on the interplay between the probabilistic approach and the infinite-dimensional linear programming optimization problem describing the extreme measures.

9.2 Poisson Processes and Financial Scenarios

In this section, we review some existing and potential applications of the multivariate Poisson processes in financial modeling. The Poisson and the Gaussian risk factor models can be integrated into a general scenario generation framework for portfolio risk measurement. In this framework, the codependence structure of the risk factors is usually described by their correlation matrix.

The correlation matrix of the risk factors contains three submatrices: C_{gg} describing the Gaussian components, C_{pp} describing the Poisson component, and C_{gp} describing the cross-correlations of the risk factors.

Such a complex model for the risk factor space can be used in the integrated market–credit risk portfolio risk framework as well as for pricing of financial derivatives. For operational risk modeling, the matrix C_{pp} should be sufficient.

9.2.1 Integrated Market–Credit Risk Modeling

Monte Carlo methods form the industry standard approach to the computation of risk measures of credit portfolios. The conceptual model of the risk factors dynamics is a combination of the Merton model, developed in Merton (1974), and the conditional independence framework (Vasicek 1987, 2002).

The first dynamic integrated market–credit risk modeling framework was developed in Iscoe *et al*. (1999) for the Gaussian risk factors. Each credit-risky name in the portfolio is characterized by a credit-worthiness index described as a linear combination of the systemic components dependent on market risk factors, macroeconomic indices, and an idiosyncratic component. Calibration of the model is reduced to a series of the default boundary problems that must be solved for each name in the portfolio independently of the other names. This problem is equivalent to the generalized Shiryaev problem (Jaimungal *et al*., 2013) studied in the Gaussian case. In Jaimungal *et al*. (2013), it was demonstrated that adding a single random jump at time $t = 0$ to the Brownian motion process allows one to cover a very rich class of default time distributions including, in particular, finite mixture of gamma distributions.

Introduction of the general jump processes in the context of credit portfolio modeling leads to a nontrivial calibration problem streaming from heterogeneity of the risk factor space.

Pricing of credit derivatives is also based on the conditional-independence framework (Avellaneda and Zhu, 2001; Hull and White, 2001; Iscoe and Kreinin, 2006, 2007), where randomized, time-dependent default intensities of jumps describe the default process. The number of components of this process is equal to the number of names in the basket credit derivative.

9.2.2 Market Risk and Derivatives Pricing

Geometric Brownian motion has been widely used in the option-pricing framework to model returns of the assets. In Merton (1976), Merton proposed a jump-diffusion model to describe better a phenomenon called volatility smile in the option markets. However, he could not address the leptokurtic feature that the return distribution of assets may have a higher peak and two (asymmetric) heavier tails than those of the normal distribution.

To incorporate both of them, Kou (2002) proposed a double exponential jump-diffusion model. The model is simple enough to produce analytical solutions for a variety of option-pricing problems, including call and put options, interest rate derivatives, and path-dependent options. Detailed accounts of the development in the area of option pricing using jump models can be found in Cont and Tankov (2003, 2006), Boyarchenko and Levendorski (2002), and Carr and Madan (1998). An alternative class of models with non-Gaussian innovations is discussed in Barndorff-Nielsen and Shephard (2001) and Kyprianou *et al*. (2005).

Notice that risk-neutral pricing of basket derivatives with several underlying instruments in the jump-diffusion framework will require a model with several correlated jump processes.

9.2.3 Operational Risk Modeling

Dependent Poisson and compound Poisson processes find many applications in the area of operational risk (OR) modeling (see Aue and Kalkbrener, 2006); Badescu *et al*., 2013, 2015; Böcker and Klüppelberg, 2010; Chavez-Demoulin *et al*., 2006; Embrechts and Puccetti, 2006; Panjer, 2006; Peters *et al*., 2009; Shevchenko, 2011; and references therein). In the OR models, Poisson processes describe random operational events in the business units of a financial organization (Chavez-Demoulin *et al*., 2006; Lindskog and McNeil, 2001; Nešlehovà *et al*., 2006; Peters *et al*., 2009; Shevchenko, 2011). The result of each operational event is a random loss. Thus, the loss process is represented as a multivariate compound Poisson process.

While modeling credit events and credit derivatives pricing in the integrated market–credit risk framework are, probably, the most complicated tasks in financial risk measurement, the isolated operational risk-modeling problem is relatively simple. The operational risk framework is based on the assumption that operational losses of the jth business unit, ($j = 1, 2, \dots, J$), are described by a compound process

$$L_t^{(j)} = \sum_{k=1}^{N_t^{(j)}} \xi_k^{(j)}, \quad j = 1, 2, \dots, J, \tag{9.1}$$

where $N_t^{(j)}$ is the number of operational events occurred by time, t; J is the number of business units in the financial organization; and $\xi_k^{(j)}$ is the loss of the jth unit in the kth event. The problem is how to find distribution of the aggregated losses

$$L_t = \sum_{j=1}^{J} L_t^{(j)}, \quad t > 0,$$

and compute the risk measures of the loss process for the purpose of capital allocation.

Denote the arrival moments of the operational events in the jth unit by $T_k^{(j)}, (k = 1, 2, \ldots)$. Then the number of operational events is

$$N_t^{(j)} = \sum_{k=1}^{\infty} \mathbb{1}(T_k^{(j)} \le t),$$

where $\mathbb{1}(\cdot)$ is the indicator function.

There are two standard models for the processes $N_t^{(j)}$. The first is the classical Poisson model. The second is the negative binomial (NB) model generalizing the classical Poisson model (see Barndorff-Nielsen and Yeo, 1969). In this chapter, we do not discuss the technical details of the NB model, leaving this topic for future publications.

9.2.4 Correlation of Operational Events

Statistical analysis of the operational events indicates the presence of both positive and negative correlations in the multivariate arrival process (Bae, 2012).

In Duch *et al.*, (2014), we considered a fragment of the estimated annual correlation matrix of the operational events, originally studied in Bae, (2012). The estimated correlation matrix $\mathbf{C}_{pp} = \|\rho_{ij}\|$, where $\rho_{ij} = \mathrm{corr}(N_T^{(i)}, N_T^{(j)})$ is the correlation coefficient of the coordinates of the multivariate process, $\mathbf{N}_t$, looks as follows:

$$\mathbf{C}_{pp} = \begin{bmatrix} 1 & 0.14 & 0.29 & 0.32 & 0.15 & 0.16 & 0.03 & 0.05 & -0.06 \\ 0.14 & 1.0 & 0.55 & -0.12 & 0.49 & 0.52 & -0.16 & 0.2 & 0.02 \\ 0.29 & 0.55 & 1.0 & 0.11 & 0.27 & 0.17 & -0.31 & 0.05 & 0.08 \\ 0.32 & -0.12 & 0.11 & 1.0 & -0.12 & 0.23 & 0.19 & -0.18 & -0.11 \\ 0.15 & 0.49 & 0.27 & -0.12 & 1.0 & 0.49 & -0.17 & 0.44 & -0.03 \\ 0.16 & 0.52 & 0.17 & -0.23 & 0.49 & 1.0 & -0.02 & 0.13 & 0.29 \\ 0.03 & -0.16 & -0.31 & 0.19 & -0.17 & -0.02 & 1.0 & 0.32 & 0.5 \\ 0.05 & 0.2 & 0.05 & -0.18 & 0.44 & 0.13 & 0.32 & 1.0 & 0.16 \\ -0.06 & 0.02 & 0.08 & -0.11 & -0.03 & 0.29 & 0.5 & 0.16 & 1.0 \end{bmatrix}$$

One can notice that there are some correlation coefficients that cannot be ignored for the simulation purposes. In particular, $\rho_{32} = 0.55$ and $\rho_{73} = -0.31$ represent the extreme correlations in $\mathbf{C}_{pp}$. Practitioners working in the operational risk area often ignore the correlation structure of the process, $\mathbf{N}_t$, for the sake of simulation simplicity. We believe that such a simplification may result in inaccurate estimation of the operational losses.

9.3 Common Shock Model and Randomization of Intensities

Two possible extensions of the multivariate model with independent components are considered in this section. The first extension is the common shock model (CSM), which has traditional applications in insurance as well as in the area of operational risk modeling.

The second extension is randomization of intensities of the Poisson processes, leading to a more general class of stochastic processes usually called the mixed Poisson processes.

9.3.1 Common Shock Model

Correlation of Poisson processes can be controlled by various operations applied to independent processes. One of the most popular operations, often considered in actuarial applications, is a superposition of random processes, the CSM (Lindskog and McNeil, 2001; Powojovsky *et al*., 2002; Shevchenko, 2011). The idea of this model is described here in the bivariate case. Consider three independent Poisson processes, $\nu_t^{(1)}$, $\nu_t^{(2)}$, and $\nu_t^{(3)}$, and denote their intensities by λ_1, λ_2, and λ_3, respectively. Let us form two new processes using a standard superposition operation of the processes, $N_t^{(1)} = \nu_t^{(1)} \bigoplus \nu_t^{(2)}$ and $N_t^{(2)} = \nu_t^{(2)} \bigoplus \nu_t^{(3)}$, defined as follows. If $\nu_t^{(1)} = \sum_{i=0}^{\infty} \mathbb{1}(t_i^{(1)} \le t)$ and $\nu_t^{(2)} = \sum_{i=0}^{\infty} \mathbb{1}(t_i^{(2)} \le t)$, then the superposition of the processes, $(\nu^{(1)} \bigoplus \nu^{(2)})_t$, is defined by

$$\left(\nu^{(1)} \bigoplus \nu^{(2)}\right)_t := \sum_{j=1}^{2} \sum_{i=0}^{\infty} \mathbb{1}(t_i^{(j)} \le t).$$

Clearly, the processes $N_t^{(1)}$ and $N_t^{(2)}$ are dependent; their correlation coefficient,

$$\rho(N_t^{(1)}, N_t^{(2)}) = \frac{\lambda_2}{\sqrt{(\lambda_1 + \lambda_2)(\lambda_2 + \lambda_3)}},$$

does not depend on time (see Lindskog and McNeil, 2001; Powojovsky *et al*., 2002) and is always nonnegative.

If all the elements of the correlation matrix, C_{pp}, are positive, the CSM can be used for approximation of the multivariate Poisson processes (Powojovsky *et al*., 2002). However, if there are negative elements, the CSM has difficulties explaining these correlations and cannot be used unless the estimated negative elements are small.

9.3.2 Randomization of Intensities

One promising approach to the analysis and simulation of the jump processes with negative correlations is randomization of intensities. In this case, the resulting processes, usually called mixed Poisson processes, will have a more general probabilistic structure and nonstationary correlations.

Proposition 9.1 (Rolski *et al*., 1999) The correlation coefficient of the mixed Poisson processes is

$$\begin{aligned}\rho(\nu_t^{(1)}, \nu_t^{(2)}) &:= \frac{\mathbb{E}[\nu_t^{(1)} \cdot \nu_t^{(2)}] - \mathbb{E}[\nu_t^{(1)}]\mathbb{E}[\nu_t^{(2)}]}{\sigma(\nu_t^{(1)})\ \sigma(\nu_t^{(2)})} \\ &= \frac{\rho(\lambda_1(X), \lambda_2(X))}{\sqrt{1 + \frac{1}{t}\frac{\mathbb{E}[\lambda_1(X)]}{\sigma^2(\lambda_1(X))}} \cdot \sqrt{1 + \frac{1}{t}\frac{\mathbb{E}[\lambda_2(X)]}{\sigma^2(\lambda_2(X))}}},\end{aligned} \tag{9.2}$$

Randomization, clearly, increases applicability of the Poisson model. The only drawback of the model is that convergence to the limit value of the correlation coefficient may be very slow.

9.4 Simulation of Poisson Processes

Let us now discuss two approaches to simulation of the Poisson processes. The first method is the forward simulation of the arrival moments. It is, probably, the most natural simulation method for independent Poisson processes. The second method, discussed in this section, is the BS method described in the univariate case in Fox (1996) and in the multivariate case in Duch *et al*., (2014).

9.4.1 Forward Simulation

Let $\mathbf{N_t} = (N_t^{(1)}, \dots, N_t^{(j)}), t \geq 0$, be a J-dimensional Poisson process with independent components. Denote by λ_j the parameter of the jth coordinate of the process, $(j = 1, 2, \dots, J)$. Then,

$$\mathbb{E}[N_t^{(j)}] = \lambda_j t.$$

It is very well known that the interarrival times of the jth Poisson process, $\Delta T_k{}^{(j)} := T_k^{(j)} - T_{k-1}^{(j)}$, are mutually independent, for all k and j, and identically distributed, for each j random variables with exponential distribution,

$$\mathbb{P}(\Delta T_k{}^{(j)} \leq t) = 1 - \exp(-\lambda_j t), \quad t \geq 0, \quad j = 1, 2, \dots, J.$$

Probably, the most natural method to generate arrival moments of the Poisson processes is the recursive simulation

$$T_k^{(j)} = T_{k-1}^{(j)} + \Delta T_k{}^{(j)}, \quad k \geq 1, \quad T_0^{(j)} = 0. \tag{9.3}$$

The number of events, $N_t^{(j)}$, in the interval, $[0, t]$, is a stochastic process with independent increments such that

$$\mathbb{P}\left(\bigcap_{j=1}^{J} N_t^{(j)} = k_j\right) = \prod_{j=1}^{J} \exp\ (-\lambda_j t) \cdot \frac{(\lambda_j t)^{k_j}}{k_j!}, \quad j = 1, 2, \dots, J, \quad k_j \in \mathbb{Z}_+. \tag{9.4}$$

If one is only interested in the number of events in the interval, $[0, t]$, the random vector, $\mathbf{N_t}$, can be sampled directly from the joint distribution (9.4).

It is very well known that if the vectors of the interarrival times are independent, the process $\mathbf{N_t}$ is Markovian in the natural filtration, $\mathfrak{F}_t = \{\sigma(N_\tau)\}_{\tau \le t}$, generated by the process $\mathbf{N_t}$.

Denote by $P_t(n_1, \dots, n_J) = \mathbb{P}\left(\bigcap_{j=1}^{J} \{N_t^{(j)} = n_j\}\right)$ the probabilities of the states of the Markov process, $\mathbf{N_t}$. Then,

$$\frac{\partial}{\partial t} P_t(n_1, \dots, n_J) = -\Lambda P_t(n_1, \dots, n_J) + \sum_{j=1}^{J} \lambda_j P_t(n_1, \dots, n_j - 1, \dots, n_J), \tag{9.5}$$

where $\Lambda = \sum_{j=1}^{J} \lambda_j$.

Until now, we considered the processes with independent coordinates. In the general case, the following alternative, describing the correlation structure of the (multivariate) Poisson process, can be found in Shreve (2004) and Revus and Yor (1991). The key role in this proposition plays the natural filtration, $\mathfrak{F}_t$, described in Revus and Yor (1991). Recall that in the definition of the multivariate Poisson process given in Revus and Yor (1991) the increments of the components of the process are conditionally independent conditional on the natural filtration, $\mathfrak{F}_t$.

Proposition 9.2 Let $N_t^{(1)}$ and $N_t^{(2)}$ be two Poisson processes measurable with respect to the filtration $\mathfrak{F}_t$. Then, either these processes are independent or they have simultaneous jumps.

Remark 9.1 *If the processes $N_t^{(1)}$ and $N_t^{(2)}$ do not have simultaneous jumps, then the system of equations (9.5) has a product-form solution*

$$P_t(n_1, n_2) = P_t^{(1)}(n_1) \cdot P_t^{(2)}(n_2),$$

where

$$P_t^{(j)}(n_j) = \exp\ (-\lambda_j t) \frac{(\lambda_j t)^{n_j}}{n_j!}, \quad j = 1, 2.$$

The latter is equivalent to independence of $N_t^{(1)}$ and $N_t^{(2)}$. The probabilities $P_t^{(i)}(n)$ satisfy (9.5) with $J = 1$. If the processes $N_t^{(1)}$ and $N_t^{(2)}$ have simultaneous jumps with positive intensity, then the bivariate process has correlated coordinates.[3]

Remark 9.2 *The statement of Proposition 9.2 can be generalized to arbitrary local martingales and semi-martingales; see Revus and Yor (1991). Simulation of correlated processes using joint distribution of the interarrival times is not very convenient: one has to find relations between the correlations of the interarrival times and correlations of the number of events, $N_t^{(j)}$. One commonly used approach by practitioners is to exploit the CSM (Powojovsky* et al.*, 2002). Let us take M independent Poisson processes, $\eta_t^{(1)}, \eta_t^{(2)}, \dots, \eta_t^{(M)}$, and form the processes $N_t^{(j)}$ obtained by superposition of the subset of the processes $\eta_t^{(m)}$.*

[3] The latter situation is regarded as the CSM.

More precisely, let $T_l^{(j)}$ be a sequence of the arrival times of the jth process, $\eta_t^{(j)}$:

$$\eta_t^{(j)} = \sum_{l=1}^{\infty} \mathbb{1}(T_l^{(j)} \le t), \quad j = 1, 2, \dots, M.$$

Denote

$$\delta_{jk} = \begin{cases} 1, & \text{if } k\text{th process} \eta_t^{(k)} \text{ contributes to the superposition} \\ 0, & \text{otherwise,} \end{cases}$$

Then the jth process,

$$N_t^{(j)} = \bigoplus_{k=1}^{M} \delta_{jk} \eta_t^{(k)},$$

is

$$N_t^{(j)} := \sum_{k=1}^{M} \delta_{jk} \sum_{l=1}^{\infty} \mathbb{1}(T_l^{(k)} \le t), \quad j = 1, 2, \dots, J.$$

The intensity, λ_j, of the process $N_t^{(j)}$ satisfies

$$\lambda_j = \sum_{k=1}^{M} \lambda_k^* \delta_{jk}, \tag{9.6}$$

where λ_k^* is intensity of $\eta_t^{(k)}$.

The elements of the correlation matrix are

$$\rho_{ij} = \frac{\lambda_{ij}^c}{\sqrt{\lambda_i \ \lambda_j}}, \tag{9.7}$$

where

$$\lambda_{ij}^c = \sum_{k=1}^{M} \lambda_k^* \delta_{ik} \delta_{jk}.$$

The calibration problem for the CSM can be formulated as follows. Find M, the matrix of contributions, $\hat{\delta} = \|\delta_{jk}\|$, $(j, k = 1, 2, \dots, M)$, and a vector of intensities $(\lambda_1^*, \dots, \lambda_M^*)$ such that Equations (9.6) and (9.7) are satisfied, where λ_1, …, λ_J are estimated intensities of the multivariate process and ρ_{ij} are estimated correlation coefficients.

Proposition 9.3 The correlation coefficients, ρ_{ij}, satisfy the inequality

$$0 \le \rho_{ij} \le \min\left(\sqrt{\frac{\lambda_i}{\lambda_j}}, \sqrt{\frac{\lambda_j}{\lambda_i}}\right) \tag{9.8}$$

Proof. It immediately follows from (9.7) that the correlation coefficients must be nonnegative and must satisfy the inequality $\lambda_{ij}^c \le \min(\lambda_i, \lambda_j)$, implying the upper bound for the elements of the correlation matrix.

Proposition 9.3 demonstrates that the CSM is too restrictive and may lead to large errors if parameters of the model do not satisfy (9.8).

9.4.2 Backward Simulation

It is well known that the conditional distribution of the arrival moments of a Poisson process in an interval, $[0, T]$, conditional on the number of events in this interval[4] is uniform (Cont and Tankov, 2003). More precisely, let $\mathcal{T} = \{T_1, T_2, \dots, T_n\}$ be a sequence of n independent random variables with a uniform distribution in the interval, $[0, T]$.

Denote by τ_k the kth-order statistic of $\mathcal{T}$, $(k = 1, 2, \dots, n)$:

$$\tau_1 = \min_{1\le k\le n} T_k, \tau_2 = \min_{1\le k\le n} \{T_k : T_k > \tau_1\}, \dots, \tau_n = \max_{1\le k\le n} T_k.$$

Theorem 9.1 (Cont and Tankov, 2003; Rolski *et al*., 1999; Nawrotzki, 1962) *The conditional distribution of the arrival moments, $\tilde{T}_1 < \tilde{T}_2 < \cdots < \tilde{T}_n$, of a Poisson process, N_t, with finite intensity coincides with the distribution of the order statistics:*

$$\mathbb{P}(\tilde{T}_k \le t \mid N_T = n) = \mathbb{P}(\tau_k \le t), \quad t \le T, \quad k = 1, 2, \dots, n. \tag{9.9}$$

The converse statement, formulated and proved here, is a foundation of the BS method for Poisson processes. Consider a process, N_t, $(0 \le t \le T)$ defined as

$$N_t = \sum_{i=1}^{N_*} \mathbb{1}(T_i \le t),$$

where N_* is a random variable with a Poisson distribution; and the random variables, T_i, are mutually independent and independent of N_*, identically distributed, with the uniform distribution

$$\mathbb{P}(T_i \le t) = tT^{-1}, \quad i = 1, 2, \dots, 0 \le t \le T.$$

Theorem 9.2 *Let N_* have a Poisson distribution with parameter λT. Then, N_t is a Poisson process with intensity λ in the interval* $[0, T]$.

Proof. We give here a combinatorial proof of the statement exploiting the analytical properties of the binomial distribution and binomial thinning. First of all, we notice that $N_T = N_*$. Let us prove that

1. For any interval, $[s, s + t]$, of the length t, $(s + t < T)$, $\mathbb{P}(N_{s+t} - N_s = m) = e^{-\lambda t}\frac{(\lambda t)^m}{m!}$, $m = 0, 1, 2, \dots$.
2. For any m disjoint subintervals $[t_i, t_i + \tau_i] \in [0, T]$, $(i = 1, 2, \dots, m)$, the random variables $N_{t_i+\tau_i} - N_{t_i}$ are mutually independent, $m = 2, 3, \dots$.

Denote $\Delta N_s(t) := N_{s+t} - N_s$. By definition of N_t, we have $N_T = N_*$ and

$$\mathbb{P}(\Delta N_s(t) = k) = \sum_{m=0}^{\infty} \mathbb{P}(\Delta N_s(t) = k \mid N_* = k + m)\cdot\mathbb{P}(N_* = k + m). \tag{9.10}$$

[4] The arrival moments here are not sorted (in ascending order).

Notice that $\Delta N_s(t)$ is the number of events that have occurred in the interval $[s, s+t]$. The conditional distribution of the random variable $\Delta N_s(t)$ is binomial

$$\mathbb{P}(\Delta N_s(t) = k \mid N_* = k+m) = \binom{k+m}{k} \left(\frac{t}{T}\right)^k \cdot \left(1 - \frac{t}{T}\right)^m, \quad m = 0, 1, \ldots$$

and the probability generating function $\hat{p}(z) := \mathbb{E}[z^{N_T}]$ is[5] $\hat{p}(z) = \exp\ (\lambda T(z-1))$. The proof of the first statement of Theorem 9.2 is based on the following.

Lemma 9.1 *Let $\{p_k\}_{k=0}^{\infty}$ be a probability distribution. Denote its generating function by $\hat{p}(z)$:*

$$\hat{p}(z) = \sum_{k=0}^{\infty} p_k z^k, \quad |z| \le 1.$$

Consider a sequence

$$q_k = \sum_{m=0}^{\infty} p_{k+m} \binom{k+m}{k} x^k (1-x)^m, 0 \le x \le 1, k = 0, 1, \ldots. \tag{9.11}$$

The sequence $\{q_k\}_{k\ge 0}$ is a probability distribution with the generating function, $\hat{q}(z)$,

$$\hat{q}(z) = \hat{p}(1 - x + xz). \tag{9.12}$$

Applying Lemma 9.1, with $p_{k+m} = \mathbb{P}(\Delta N_s(t) = k \mid N_* = k+m)$ and satisfying equation (9.10) and $x = t \cdot T^{-1}$, we obtain that the generating function, $g(z) := \mathbb{E}[z^{\Delta N_s(t)}] = \exp\ (\lambda t(z-1))$. Therefore, the increments of the process have a Poisson distribution, $\Delta N_s(t) \sim \text{Pois}(\lambda t)$, and this distribution does not depend on s.

To prove the second statement of the theorem, we need the following generalization of Lemma 9.1. Consider a vector $\vec{x} = (x_1, x_2, \ldots, x_m)$, satisfying the conditions

$$x_j \ge 0, \quad j = 1, 2, \ldots, m, \quad \sum_{j=1}^{m} x_j < 1,$$

and denote

$$y = 1 - \sum_{j=1}^{m} x_j.$$

Denote $\nu_i := N_{t_i+\tau_i} - N_{t_i}, i = 1, 2, \ldots, m$. For an m-dimensional vector, $\vec{k} = (k_1, k_2, \ldots, k_m) \in \mathbb{Z}_+^m$, with nonnegative integer coordinates, $(k_j \ge 0)$, we define the norm of the vector

$$\|\vec{k}\| := \sum_{j=1}^{m} k_j.$$

For any m-dimensional vector, $\vec{x} = (x_1, x_2, \ldots, x_m)$, with nonnegative coordinates, $(x_j \ge 0)$, and $\vec{k} \in \mathbb{Z}_+^m$, we denote

$$\vec{x}^{\,\vec{k}} := \prod_{j=1}^{m} x_j^{k_j}.$$

[5] For the sake of brevity, we will call $\hat{p}(z)$ just generating function, in what follows.

We also introduce a combinatorial coefficient

$$\binom{\vec{k}+l}{\vec{k}} := \prod_{j=1}^{m} \binom{\sum_{i=j}^{m} k_i + l}{k_j}. \tag{9.13}$$

It is not difficult to see that this coefficient can be written in a more symmetric form as a multinomial coefficient

$$\binom{\vec{k}+l}{\vec{k}} = \frac{\left(\sum_{i=1}^{m} k_i + l\right)!}{l! \cdot \prod_{i=1}^{m} k_i!}.$$

Lemma 9.2 *Let $\{p_k\}_{k=0}^{\infty}$ be the probability distribution of a discrete random variable, ξ and let $\hat{p}(z) = \mathbb{E}[z^{\xi}]$ be its generating function. Let $\vec{k} \in \mathbb{Z}_+^m$ and the combinatorial coefficient, $\binom{\vec{k}+l}{\vec{k}}$, be defined by (9.13). Let $\pi : \mathbb{Z}_+^m \to \mathbb{R}$ be defined by*

$$\pi(\vec{k}) = \sum_{l=0}^{\infty} p_{\|\vec{k}\|+l} \binom{\vec{k}+l}{\vec{k}} \cdot \vec{x}^{\,\vec{k}} \cdot y^l, \tag{9.14}$$

and denote by $\hat{\pi}(\vec{z})$ the generating function

$$\hat{\pi}(\vec{z}) := \sum_{\vec{k} \in \mathbb{Z}_+^m} \pi(\vec{k}) \vec{z}^{\,\vec{k}}.$$

Then

$$\hat{\pi}(\vec{z}) = \hat{p}\left(1 - \sum_{j=1}^{m} x_j (1 - z_j)\right). \tag{9.15}$$

Lemma 9.2 is proved in the Appendix. Let us now finish the proof of the second statement.

Denote $p_j = \frac{\tau_j}{T}$, $(j = 1, 2, \ldots, m)$, $\vec{p} = (p_1, \ldots, p_m)$, and $q = 1 - \sum_{j=1}^{m} p_j$. Then we have

$$\mathbb{P}\left(\nu_1 = k_1, \ldots, \nu_m = k_m \mid X_T = \sum_{j=1}^{m} k_j + l\right) = \binom{\vec{k}+l}{\vec{k}} \cdot \vec{p}^{\,\vec{k}} \cdot q^l.$$

Therefore,

$$\mathbb{P}(\nu_1 = k_1, \ldots, \nu_m = k_m) = \sum_{l=0}^{\infty} \binom{\vec{k}+l}{\vec{k}} \cdot \vec{p}^{\,\vec{k}} \cdot q^l \cdot e^{-\lambda T} \frac{(\lambda T)^{\|\vec{k}\|+l}}{(\|\vec{k}\|+l)!}. \tag{9.16}$$

Consider the generating function

$$\pi(\vec{z}) := \mathbb{E}\left[\prod_{j=1}^{m} z_j^{\nu_j}\right], \quad |z_j| \le 1, \quad j = 1, 2, \ldots, m.$$

Applying Lemma 9.2 with $\hat{p}(z) := \mathbb{E}[z^{X_T}] = e^{\lambda T(z-1)}$, $x_j = p_j$ and $y = q$, we derive

$$\pi(\vec{z}) = \prod_{j=1}^{m} e^{\lambda \tau_j (z_j - 1)}.$$

The latter relation implies independence of the random variables $\nu_j, (j = 1, 2, \dots, m)$. Therefore, the process $N_t, (0 < t \le T)$ has independent increments. Theorem 9.2 is thus proved.

Remark 9.3 *The conditional uniformity of the (unsorted) arrival moments is a characteristic property of the more general class of mixed Poisson processes (Nawrotzki, 1962; Rolski* et al.*, 1999).*

Now we are in a position to formulate the BS method in the of case of $J = 1$.

Step 1. Generate a random number, n, having the Poisson distribution with parameter λT: $n \sim \text{Pois}(\lambda T)$, and assign $N_T := n$.
Step 2. Generate n uniformly distributed random variables in the interval $[0, T]$, $T_i \sim U([0, T])$, $i = 1, 2, \dots, n$.
Step 3. Sort the random variables, T_i, in the ascending order.
Step 4. Repeat Steps 1–3 n_{mc} times, where n_{mc} is the required number of scenarios.

Remark 9.4 *The BS method is applicable to the class of mixed Poisson processes with random intensity. It can also be used for simulation of jump processes represented as a time-transformed Poisson process (Feigin, 1979).*

Remark 9.5 *The BS of the Poisson processes can be implemented using quasi–Monte Carlo (QMC) algorithms (see Fox, 1996; Fox and Glynn, 1988). In Fox (1996), in Step 1, the stratified sampling was proposed for generation of the number of arrivals, N_T, and QMC for generation of arrival times. This numerical strategy significantly increases the rate of convergence.*

9.4.2.1 Backward Simulation: $J > 1$

It is easy to extend the BS algorithm to the multivariate Poisson processes. Suppose that marginal distributions of the random vector $\mathbf{N}_* = (N_*^{(1)}, N_*^{(2)}, \dots, N_*^{(j)})$ are Poisson, $N_*^{(j)} \sim \text{Pois}(\lambda_j T)$. Denote the correlation coefficient of $N_*^{(i)}$ and $N_*^{(j)}$ by ρ_{ij}.

Theorem 9.3 *Consider the processes*

$$N_t^{(j)} = \sum_{i=1}^{N_*^{(j)}} \mathbb{1}(T_i^{(j)} \le t), \quad j = 1, 2, \dots, J,$$

where the random variables, $T_i^{(j)}, (i = 1, 2, \dots, X_*^{(j)})$, are mutually independent and uniformly distributed in the interval $[0, T]$. Then, $N_t^{(j)}$ is a multivariate Poisson process in the interval $[0, T]$ and

$$corr(N_t^{(i)}, N_t^{(j)}) = \rho_{ij} t T^{-1}, \quad 0 \le t \le T. \tag{9.17}$$

Proof. We have already proved in Theorem 9.2 that $N_t^{(j)}$ is a Poisson process. Let us now formulate the auxiliary result used in the derivation of (9.17).

Consider a bivariate, integer-valued random vector, $\vec{\zeta} = (\zeta_1, \zeta_2)$, $\vec{\zeta} \in \mathbb{Z}_+^2$, and denote

$$p(k,l) = \mathbb{P}(\zeta_1 = k, \zeta_2 = l), \quad k,l = 0,1,2,\dots,$$

its probability distribution. By $\hat{p}(z,w)$, we denote the generating function

$$\hat{p}(z,w) := \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} p(k,l) z^k w^l, \quad |z| \le 1, \quad |w| \le 1.$$

Consider a random vector, $\xi = (\xi_1, \xi_2)$, such that for all $k = 0,1,2,\dots,k'$, and $l = 0,1,\dots,l'$

$$\mathbb{P}(\xi_1 = k, \xi_2 = l \mid \zeta_1 = k', \zeta_2 = l') = \binom{k'}{k} x^k (1-x)^{k'-k} \cdot \binom{l'}{l} y^l (1-y)^{l'-l}, \tag{9.18}$$

where $0 < x \le 1, 0 < y \le 1$. The components of the random vector ξ are conditionally independent. The joint probability, $q_{k,l} = \mathbb{P}(\xi_1 = k, \xi_2 = l)$, can be written as

$$q_{k,l} = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} p_{k+m,l+n} \cdot \binom{k+m}{k} x^k (1-x)^m \cdot \binom{l+n}{l} y^l (1-y)^n, \quad k,l = 0,1,2,\dots. \tag{9.19}$$

Lemma 9.3 *Suppose that the variance and the first moment of the random variables, ζ_i, $(i = 1,2)$, are equal:*[6] $\mathbb{E}[\zeta_i] = \sigma^2(\zeta_i)$. *The generating function $\hat{q}(z,w) := \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} q_{k,l} z^k w^l$, satisfies the relation*

$$\hat{q}(z,w) = \hat{p}(1 - x + xz, 1 - y + yw), \quad |z| \le 1, |w| \le 1. \tag{9.20}$$

The correlation coefficient of the random variables ξ_1 and ξ_2 is

$$\rho(\xi_1, \xi_2) = \sqrt{xy} \cdot \rho(\zeta_1, \zeta_2). \tag{9.21}$$

Lemma 9.3 is proved in the Appendix.

Let us apply Lemma 9.3 with $\vec{\zeta} = (N_*^{(i)}, N_*^{(j)})$ and $\vec{\xi} = (N_t^{(i)}, N_t^{(j)})$. The conditional probabilities, $\mathbb{P}((N_t^{(i)} = k, N_t^{(j)} = l) \mid (N_T^{(i)} = k', N_T^{(j)} = l'))$, satisfy (9.18) with $x = y = t\,T^{-1}$ and $\rho(\zeta_1, \zeta_2) = \mathrm{corr}(N_t^{(i)}, N_t^{(j)})$. Then, Lemma 9.3 implies (9.17).

Theorem 9.4 states the Markovian structure of the process $\vec{N}_t$. If $J = 1$, $\vec{N}_t$ is a Poisson process and, obviously, is Markovian in the natural filtration. If $J > 1$, the Markovian nature of the process is not obvious since it is constructed backward.

Theorem 9.4 *The process $\vec{N}_t = (N_t^{(1)}, N_t^{(2)}, \dots, N_t^{(j)})$ is a Markov process in the interval $[0,T]$.*

Proof. Let $0 = t_0 < t_1 < t_2 < \cdots < t_{m-1} < t_m < T$ be a partition of the interval $[0,T]$. Consider integer vectors, $\vec{n}_j \in \mathbb{Z}_+^m$, $\vec{n}_j = (n_1^{(j)}, n_2^{(j)}, \dots, n_{m-1}^{(j)}, n_m^{(j)})$, $(j = 1,2,\dots,J)$, and denote

$$P\left(\vec{n}_1, \vec{n}_2, \dots, \vec{n}_m\right) = \mathbb{P}\left(\bigcap_{k=1}^{m}\bigcap_{j=1}^{J} \{N_{t_k}^{(j)} = n_k^{(j)}\}\right).$$

[6] As in the case of the Poisson distributions.

By construction, the coordinates of $\vec{N}_t$ are conditionally independent, and the arrival moments have conditional uniform distributions. Therefore,

$$\mathbb{P}\left(\bigcap_{k=1}^{m}\bigcap_{j=1}^{J}\left\{N_{t_k}^{(j)}=n_k^{(j)}\right\}\middle|\ \vec{N}_T=\vec{k}\right)=\prod_{j=1}^{J}\mathbb{P}\left(\bigcap_{k=1}^{m}\left\{N_{t_k}^{(j)}=n_k^{(j)}\right\}\middle|\ N_T^{(j)}=k_j\right) \tag{9.22}$$

$$=\prod_{j=1}^{J}\left(\frac{k_j\ !}{(k_j-n_m^{(j)})!\cdot\prod_{k=1}^{m}\left(\Delta n_k^{(j)}\right)!}\cdot\prod_{k=1}^{m+1}\left(\frac{\tau_k}{T}\right)^{\Delta n_k^{(j)}}\right),$$

where $\Delta n_k^{(j)}=n_k^{(j)}-n_{k-1}^{(j)}$, $\tau_k=t_k-t_{k-1}$, $\tau_{m+1}=T-t_m$ and $n_0^{(j)}=0$.

Consider the conditional probabilities

$$P\left(\vec{n}_m\mid\vec{n}_{m-1},\ldots,\vec{n}_1\right):=\mathbb{P}\left(\bigcap_{j=1}^{J}\left\{N_{t_m}^{(j)}=n_m^{(j)}\right\}\middle|\bigcap_{j=1}^{J}\bigcap_{k=1}^{m-1}\left\{N_{t_k}^{(j)}=n_k^{(j)}\right\}\right).$$

We have

$$P(\vec{n}_m\mid\vec{n}_{m-1},\ldots,\vec{n}_1)=\frac{P(\vec{n}_1,\vec{n}_2,\ldots,\vec{n}_m)}{P(\vec{n}_1,\vec{n}_2,\ldots,\vec{n}_{m-1})} \tag{9.23}$$

$$=\frac{P(\vec{n}_{m-2},\ldots,\vec{n}_1\mid\vec{n}_m,\vec{n}_{m-1})P(\vec{n}_m,\vec{n}_{m-1})}{P(\vec{n}_1,\vec{n}_2,\ldots,\vec{n}_{m-1})}.$$

Equation (9.23) implies

$$P(\vec{n}_{m-2},\ldots,\vec{n}_1\mid\vec{n}_m,\vec{n}_{m-1})=P(\vec{n}_{m-2},\ldots,\vec{n}_1\mid\vec{n}_{m-1}). \tag{9.24}$$

Notice that $P(\vec{n}_m,\vec{n}_{m-1})=P(\vec{n}_m\mid\vec{n}_{m-1})P(\vec{n}_{m-1})$. Equations, (9.23) and (9.24) imply

$$P(\vec{n}_m\mid\vec{n}_{m-1},\ldots,\vec{n}_1)=P(\vec{n}_m\mid\vec{n}_{m-1}),$$

as was to be proved.

9.4.2.2 Generation of Poisson Random Vectors

Using the BS method, we reduce simulation of the multivariate Poisson processes to generation of the random integer-valued vector with marginal Poisson distribution. The following approach to generation of the Poisson random vectors, based on the idea of transformation of Gaussian random vectors with properly assigned correlations, is considered in Yahav and Shmueli (2011) and Duch *et al*. (2014).

Let us recall one simple, general property of random variables: if ξ is a continuous random variable with the cumulative distribution function (cdf), $F(x)$, then the random variable, $\zeta=F(\xi)$, has the standard uniform distribution, $U([0,1])$, in the unit interval.

Conversely, given a distribution function, F, and a sample from the uniform distribution, ζ_1, ζ_2, …, ζ_M, we obtain a new sample, $\xi_m=F^{(-1)}(\zeta_m)$, $(m=1,2,\ldots,M)$, from the distribution F (see Johnson and Kotz, 1969; Johnson *et al*., 1997).

Consider a random vector, $\vec{\eta} = (\eta_1, \dots \eta_J)$, with a mean-zero multivariate normal distribution, $\mathcal{N}(0, \rho)$, with the correlation matrix, ρ, and unit variances. The random vector, $\vec{\xi} = (\Phi(\eta_1), \dots, \Phi(\eta_m))$, where $\Phi(x)$ is the standard normal cdf, has a multivariate distribution with the standard uniform marginals. Let $[x]$ be the integer part of x, ($x \in \mathbb{R}$). Denote

$$P_\lambda(x) := \begin{cases} \sum_{k=0}^{[x]} e^{-\lambda} \frac{\lambda^k}{k!}, & \text{if } x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

the cdf of the Poisson random variable. Applying the inverse Poisson cdf to each coordinate of $\vec{\xi}$, we obtain a random vector $\vec{\zeta} = (\zeta_1, \dots, \zeta_J)$, where $\zeta_j = P^{-1}_{\lambda_j}(\Phi(\eta_j))$ has Poisson marginal distribution with parameter λ_j. The correlation coefficient, $\hat{\rho}_{k,l} := \text{corr}(\zeta_k, \zeta_l)$, is

$$\hat{\rho}_{k,l} = \frac{\mathbb{E}[\zeta_k \cdot \zeta_l] - \lambda_k \lambda_l}{\sqrt{\lambda_k \cdot \lambda_l}}, \quad k, l = 1, 2, \dots, J,$$

and

$$\mathbb{E}[\zeta_k \cdot \zeta_l] = \lambda_k \lambda_l + \hat{\rho}_{k,l} \cdot \sqrt{\lambda_k \cdot \lambda_l}. \tag{9.25}$$

9.4.2.3 Model Calibration

Let us now briefly discuss the calibration problem. In Duch *et al*. (2014), the calibration problem was set as a matching of the intensities, λ_j, and the correlation coefficients, $\hat{\rho}_{ij}$. The intensities, λ_j, can be found from the observations of the number of events in the interval, $[0, T]$. In Duch *et al*. (2014), it is shown that

$$\begin{aligned} \mathbb{E}[\zeta_k \cdot \zeta_l] &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} n \cdot m \mathbb{P}(\zeta_k = m, \zeta_l = n) \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} n \cdot m \mathbb{P}(u_m^{(k-1)} < \Phi(\eta_k) \leq u_m^{(k)}, u_n^{(l-1)} < \Phi(\eta_l) \leq u_n^{(l)}), \end{aligned} \tag{9.26}$$

where $u_j^{(i)} = P^{-1}_{\lambda_i}(j)$, ($i = 1, 2, \dots, J, j = 1, 2, \dots$).

The probabilities, $\mathbb{P}(u_m^{(k-1)} < \Phi(\eta_k) \leq u_m^{(k)}, u_n^{(l-1)} < \Phi(\eta_l) \leq u_n^{(l)})$, can be written as a linear combination of the bivariate normal distribution functions, $\Phi_2(\cdot, \cdot, \rho)$, with the arguments depending on the indices m and n. Then, from (9.25) and (9.26), we obtain an implicit equation for the correlation coefficient, $\hat{\rho}_{k,l}$.

A numerical scheme for the computation of the matrix $\hat{\boldsymbol{\rho}} = \|\hat{\rho}_{k,l}\|$ is considered in detail in Duch *et al*. (2014). There is, however, one delicate problem related to the existence of the correlation matrix $\boldsymbol{\rho}$. The problem is the sufficient conditions, for a given positive semidefinite matrix to be a correlation matrix of a random vector with the Poisson marginal distributions, depend on the parameters, λ_j, of these distributions. The latter property is in contrast with Gaussian random vectors. Thus, the admissibility test is required to verify if the calibration problem has a solution. It is shown in Duch *et al*. (2014) that the calibration problem for multivariate processes is decomposed into a series of bivariate problems for each correlation coefficient. This problem is analyzed in Section 9.5.

9.5 Extreme Joint Distribution

We would like to find the range, $[\rho^*, \rho^{**}]$, of the admissible values of the correlation coefficient, given the intensities, λ and μ, of the random variables, $\mathcal{N}$ and $\mathcal{N}'$, with Poisson marginal distributions. A straightforward observation is that this range is asymmetrical. Indeed, it is not difficult to see that the maximal correlation can be 1 in the case of equal intensities. On the other hand, it is impossible to obtain $\rho = -1$, simply because both $\mathcal{N}$ and $\mathcal{N}'$ are nonnegative, unbounded random variables and a linear relation, $\mathcal{N} = -b\mathcal{N}' + c$ with a constant $b > 0$ and an arbitrary finite c, contradicts the nonnegativity of the random variables.

9.5.1 Reduction to Optimization Problem

Denote the marginal probabilities of the vector $(\mathcal{N}, \mathcal{N}')$ by $p(i) := \mathbb{P}(\mathcal{N} = i)$ and $q(j) := \mathbb{P}(\mathcal{N}' = j)$, $(i, j = 0, 1, 2, \dots)$. The results of this subsection are derived in a general case and do not use the Poisson specification of the marginal probabilities.

Let us find the joint distributions of the random vector, $(\mathcal{N}, \mathcal{N}')$, such that the correlation coefficient $\rho = \rho(N, \mathcal{N}')$ takes an extreme value, $\rho \to \text{extr}$, and

$$\begin{aligned} \mathbb{P}(\mathcal{N} = i) &= p(i), i = 0, 1, 2, \dots, \\ \mathbb{P}(\mathcal{N}' = j) &= q(j), j = 0, 1, 2, \dots, \end{aligned} \tag{9.27}$$

where $\sum_{i=0}^{\infty} p(i) = \sum_{j=0}^{\infty} q(j) = 1$.

This problem was considered in Embrechts and Puccetti (2006), Griffiths *et al*. (1979), and Nelsen (1987). The coefficients $\rho^* = \min\rho(N, \mathcal{N}')$ and $\rho^{**} = \max\rho(N, \mathcal{N}')$ were numerically calculated in Griffiths *et al*. (1979) using reduction to the optimization problem discussed below. In that chapter, it was shown that both ρ^* and ρ^{**} are nonmonotone functions of the intensities, λ and μ. It was proved in Griffiths *et al*. (1979), using the nonmonotonicity property, that the distribution of the vector $(\mathcal{N}, \mathcal{N}')$ is not infinitely divisible.

Denote by $\mathfrak{p}$ the matrix of the joint probabilities

$$\mathfrak{p}_{i,j} = \mathbb{P}(\mathcal{N} = i, \mathcal{N}' = j), i = 0, 1, 2, \dots, j = 0, 1, \dots.$$

Let

$$f(\mathfrak{p}) := \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} ij\mathfrak{p}_{i,j}.$$

Then the solution to the problem

$$\begin{aligned} & f(\mathfrak{p}) \to \text{extr} \\ & \sum_{j=0}^{\infty} \mathfrak{p}_{i,j} = p(i), \quad i = 0, 1, \dots \\ & \sum_{i=0}^{\infty} \mathfrak{p}_{i,j} = q(j), \quad j = 0, 1, \dots \end{aligned} \tag{9.28}$$

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\mathfrak{p}_{i,j}=1$$

$$\mathfrak{p}_{i,j}\geq 0\quad i,j=0,1,\ldots$$

is the joint distribution maximizing (or minimizing) the correlation coefficient given marginal distributions of $\mathcal{N}$ and $\mathcal{N}'$. Note that the problem (9.28) belongs to the class of infinitely dimensional linear programming problems.[7]

9.5.2 Monotone Distributions

In this section, we describe the structure of the extreme joint distributions and derive formulae for the joint probabilities, $\mathfrak{p}^{*}_{i,j}$ and $\mathfrak{p}^{**}_{i,j}$ (see Theorem 9.5). We introduce two special classes of discrete, bivariate distributions that, according to Frechet (1951), we call co-monotone and anti-monotone distributions, and prove that, given marginal distributions, the solution to the extreme problem (9.28) belongs to these classes. We also demonstrate that in the case of discrete bivariate distributions, Theorem 9.5 is equivalent to the famous Frechet–Hoeffding theorem on extreme distributions, obtained in Hoeffding (1940).

Consider a set of points, (finite or infinite), $\mathcal{S}=\{s_n\}_{n\geq 1}$, where $s_n=(x_n,y_n)\in\mathbb{R}^2$. Consider the subsets $\mathcal{R}_+=\{(x,y)\in\mathbb{R}^2 : x\cdot y\geq 0\}$ and $\mathcal{R}_-=\{(x,y)\in\mathbb{R}^2 : x\cdot y\leq 0\}$.

Definition 9.1 *A set $\mathcal{S}=\{s_n\}_{n\geq 1}\in\mathbb{R}^2$ is called co-monotone if $\forall i,j$ the vector $s_i-s_j\in\mathcal{R}_+$. A set $\mathcal{S}\in\mathbb{R}^2$ is called anti-monotone if $\forall i,j\ s_i-s_j\in\mathcal{R}_-$.*

Figure 9.1 illustrates the notion of the co-monotone and anti-monotone sets introduced in Definition 9.1. Let $z\in\mathbb{R}^N$ and π be a permutation of N elements. The group of permutations of N elements is denoted by $\mathfrak{S}_N$. Finally, we introduce the notation $\pi z:=(z_{\pi(1)},z_{\pi(2)},\ldots z_{\pi(N)})$.

Suppose that $\mathcal{S}=\{(x_n,y_n)\}_{n=1}^{N}$ is a finite, co-monotone set in $\mathbb{R}^2$, $N\geq 2$. Consider the vectors, $x=(x_1,x_2,\ldots,x_N)$ and $y=(y_1,y_2,\ldots,y_N)$. Then there exists a permutation, π, such that both vectors πx and πy have monotonically increasing coordinates:

$$x_{\pi(1)}\leq x_{\pi(2)}\leq\cdots\leq x_{\pi(N)}\quad\text{and}\quad y_{\pi(1)}\leq y_{\pi(2)}\leq\cdots\leq y_{\pi(N)}.$$

If $\mathcal{S}$ is an anti-monotone set, then there exists a permutation, τ, such that τx has monotonically increasing coordinates and coordinates of τy are monotonically decreasing.

Let us now introduce the co-monotonicity of two-dimensional random vectors. Consider a random vector, (X,Y), where X and Y are discrete random variables. Denote by $\mathcal{S}:=\{s_n\}_{n=1}^{\infty}$, $s_n=(x_n,y_n)$, the support of its distribution. The probabilities, $\mathsf{P}(s_n)=\mathbb{P}(X=x_n,Y=y_n)$, satisfy

$$\sum_{n=1}^{\infty}\mathsf{P}(s_n)=1\quad\text{and }\mathsf{P}(s_n)>0\text{ for all }n,\ (n=1,2,\ldots).$$

Definition 9.2 *We call the distribution, P, co-monotone if its support is a co-monotone set. A discrete, bivariate distribution is called anti-monotone if its support is an anti-monotone set.*

[7] These problems are often numerically unstable.

Let a random vector (X, Y) take values on the lattice $\mathbb{Z}_+^{(2)} = \{(i,j) : i \geq 0, j \geq 0\}$. Denote $\mathfrak{p}_{i,j} := \mathbb{P}(X = i, Y = j)$, and introduce marginal cdf's $P_i := \mathbb{P}(X \leq i)$, and $Q_j := \mathbb{P}(Y \leq j)$, $(i, j = 0, 1, 2, \dots)$.

Suppose that $\mathsf{P} = \{p(n)\}_{n=0}^{\infty}$ and $\mathcal{Q} = \{q(n)\}_{n=0}^{\infty}$, are discrete distributions on $\mathbb{Z}_+$, and denote by $\mathfrak{D}(\mathsf{P}, \mathcal{Q})$ the class of discrete bivariate distributions,

$$\mathfrak{D}(\mathsf{P}, \mathcal{Q}) = \{f(i,j) : \sum_{j \in \mathbb{Z}_+} f(i,j) = p(i), \sum_{i \in \mathbb{Z}_+} f(i,j) = q(j), i, j \in \mathcal{S}\},$$

with the marginal distributions, P and Q, having finite first and second moments. If $\mathfrak{p} \in \mathfrak{D}(\mathsf{P}, \mathcal{Q})$ is a distribution of a random vector (X, Y), then $\mathbb{E}[XY] < \infty$. Obviously the Poisson distributions, P and Q, belong to $\mathfrak{D}(\mathsf{P}, \mathcal{Q})$.

The main result of this section is the following.

Theorem 9.5 *The distribution, $\mathfrak{p}^{**}$, maximizing the correlation coefficient of X and Y, given marginal distributions, is co-monotone. If a vector (i,j) belongs to the support of $\mathfrak{p}^{**}$, then the probability, $\mathfrak{p}_{i,j}^{**} := \mathbb{P}(X = i, Y = j)$, satisfies*

$$\mathfrak{p}_{i,j}^{**} = \min(P_i, Q_j) - \max(P_{i-1}, Q_{j-1}), i, j = 0, 1, 2, \dots, \tag{9.29}$$

$$\mathfrak{p}_{0,0}^{**} = \min(P_0, Q_0), \tag{9.30}$$

where the probabilities $P_{-1} = Q_{-1} = 0$.

The distribution, $\mathfrak{p}^{*}$, minimizing the correlation coefficient of X and Y is anti-monotone. If a vector (i,j) belongs to the support of $\mathfrak{p}^{*}$ then

$$\mathfrak{p}_{i,j}^{*} = \min(P_i, \bar{Q}_{j-1}) - \max(P_{i-1}, \bar{Q}_j), i, j = 0, 1, 2, \dots, \tag{9.31}$$

where $\bar{Q}_j = 1 - Q_j$ for $j = 0, 1, 2, \dots$ and $\bar{Q}_{-1} := 1$.

The extreme distributions, $\mathfrak{p}^{*}$ and $\mathfrak{p}^{**}$, are unique.

The rest of this subsection is dedicated to the proof of Theorem 9.5. At first, we recall one important result on the monotone sequences of real numbers, that can be found in Hardy *et al.* (1952). This result motivates introduction of the co-monotone or anti-monotone distributions. After that, we prove that the extreme distributions must be co-monotone and find their support and compute the joint probabilities. The reasoning described below is quite general and does not depend on the specific type of the marginal distributions.

Let $x_1 \leq x_2 \leq \dots \leq x_N$ be a monotone sequence of N real numbers, $x_k \in \mathbb{R}$, $(k = 1, 2, \dots, N)$, and $y_1, y_2, \dots, y_N$ be an arbitrary sequence of real numbers, $y_k \in \mathbb{R}$. Denote the inner product of two vectors, $x \in \mathbb{R}^N$, and $y \in \mathbb{R}^N$, by

$$\langle x, y \rangle := \sum_{k=1}^{N} x_k y_k, \quad x = (x_1, \dots, x_N), y = (y_1, \dots, y_N).$$

Lemma 9.4 (Hardy *et al.*, 1952) *For any monotonically increasing sequence, $x_1 \leq x_2 \leq \dots \leq x_N$, and a vector, $y \in \mathbb{R}^N$, there exist permutations, π_+ and π_-, solving the extreme problems*

$$\langle x, \pi_+ y \rangle = \max_{\pi \in \mathfrak{S}_N} \langle x, \pi y \rangle,$$

and

$$\langle x, \pi_{-} y\rangle = \min_{\pi\in\mathfrak{S}_N} \langle x, \pi y\rangle.$$

The permutation, π_+, sorts the vector y in ascending order; the permutation, π_-, sorts the vector y in descending order.

If x and y are co-monotone, then

$$\langle x, y\rangle = \max_{\pi,\tau\in\mathfrak{S}_N} \langle \pi x, \tau y\rangle. \tag{9.32}$$

Indeed,

$$\sum_{i=1}^{N} x_i y_i = \sum_{i=1}^{N} x_{\pi(i)} y_{\pi(i)}, \forall \pi \in \mathfrak{S}_N.$$

By the co-monotonicity assumption, there exists a permutation, π, such that both πx and πy are monotonically increasing. Then, Lemma 9.4 implies (3.32).

From now on, we will consider only discrete random vectors, (X, Y), defined on $\mathbb{Z}_+^2$, the positive quadrant of the two-dimensional lattice. The joint probabilities will be denoted by $\mathfrak{p}_{i,j} := \mathbb{P}(X = i, Y = j)$, $i, j = 0, 1, \ldots$ in what follows.

Consider two monotone sequences, $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$, both containing the set of all non-negative, integer numbers satisfying the inequalities $|x_{n+1} - x_n| \le 1$, $|y_{n+1} - y_n| \le 1$. Obviously, these sequences are co-monotone. Take the set $Z \in \mathbb{Z}_+^2$, $Z = \{(x_n, y_n)\}_{n=1}^{N}$, for some fixed N. Let us connect the nodes (x_n, y_n) and (x_{n+1}, y_{n+1}) by the arrows, $(n = 1, 2,...,N - 1)$. Then we obtain a directed path in $\mathbb{Z}_+^2$ (see Figure 9.1a). This path can be viewed as a graph of a monotonically increasing, multivalued function taking integer values.

In the case of the anti-monotone sequences, the path looks similar to that displayed in Figure 9.1b; it can be viewed as a graph of a monotonically decreasing, multivalued function taking integer values. To emphasize that the introduced path contains support of the distribution, we shall call it the *support path*, or, for the sake of brevity, the $\mathcal{S}$-path.

Lemma 9.5 *Suppose that a distribution $\mathfrak{p}$ solves the problem (9.28), $f(\mathfrak{p}) \to \max$. Then $\mathfrak{p}$ is co-monotone.*

Proof. Let (i, j) belong to the support of $\mathfrak{p}$. We shall prove that $\mathfrak{p}_{i+1,j} \cdot \mathfrak{p}_{i,j+1} = 0$. Indeed, the probabilities, $\mathfrak{p}_{i,j}$, satisfy the constraints

$$\mathfrak{p}_{i,j} + \mathfrak{p}_{i,j+1} = \hat{P}_i, \tag{9.33}$$

$$\mathfrak{p}_{i+1,j} + \mathfrak{p}_{i+1,j+1} = \hat{P}_{i+1}, \tag{9.34}$$

where $\hat{P}_i = \sum_{k\neq j,j+1} \mathfrak{p}_{i,k}$ and $\hat{P}_{i+1} = \sum_{k\neq j,j+1} \mathfrak{p}_{i+1,k}$. On the other hand,

$$\mathfrak{p}_{i,j} + \mathfrak{p}_{i+1,j} = \hat{Q}_j, \tag{9.35}$$

$$\mathfrak{p}_{i,j+1} + \mathfrak{p}_{i+1,j+1} = \hat{Q}_{j+1}, \tag{9.36}$$

where $\hat{Q}_j = \sum_{l\neq i,i+1} \mathfrak{p}_{l,j}$ and $\hat{Q}_{j+1} = \sum_{l\neq i,i+1} \mathfrak{p}_{l,j+1}$. Obviously,

$$\hat{P}_i + \hat{P}_{i+1} = \hat{Q}_j + \hat{Q}_{j+1}. \tag{9.37}$$

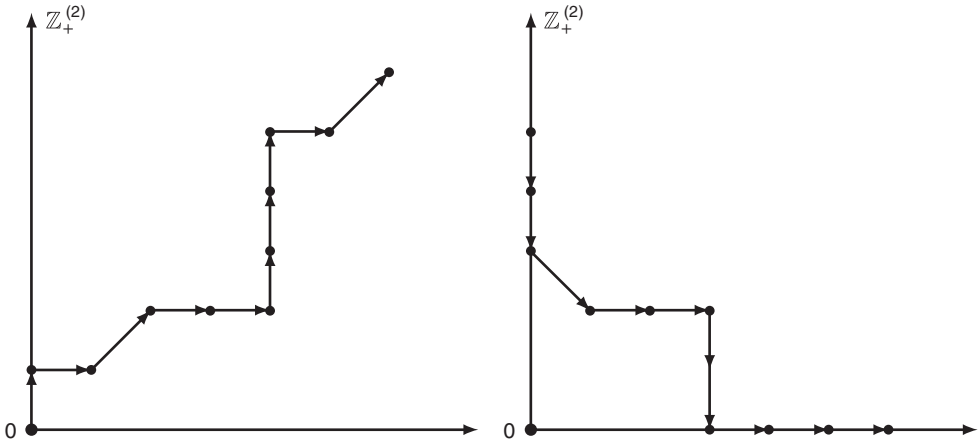

Figure 9.1 Typical monotone paths.

The function $f(\mathfrak{p})$ can be written as

$$f(\mathfrak{p}) = \sum_{\alpha} kl\mathfrak{p}_{k,l} + \sum_{\beta} kl\mathfrak{p}_{k,l},$$

where the set of indices, α, contains only four indices $\alpha = \{(i,j),(i,j+1),(i+1,j),(i+1, j+1)\}$; and β is the complementary set, $\beta = \mathbb{Z}_+^2 \backslash \alpha$.

Denote $t = \mathfrak{p}_{i+1,j+1}$, $t \geq 0$. Equations (9.33–9.37) imply

$$\begin{aligned} \mathfrak{p}_{i+1,j} &= \hat{P}_{i+1} - t, \\ \mathfrak{p}_{i,j} &= \hat{Q}_j - \hat{P}_{i+1} + t, \text{ and} \\ \mathfrak{p}_{i,j+1} &= \hat{P}_i - \hat{Q}_j + \hat{P}_{i+1} - t. \end{aligned}$$

We have $\max(0, \hat{P}_{i+1} - \hat{Q}_j) \leq t \leq \min(\hat{P}_{i+1}, \hat{Q}_{j+1})$. The objective function can be written as

$$f(\mathfrak{p}) = \sum_{(k,l)\in\beta} kl\mathfrak{p}_{k,l} + ij\cdot(\hat{P}_i + \hat{P}_{i+1}) + j\cdot\hat{P}_{i+1} + i\hat{Q}_{j+1} + t.$$

The solution to the problem, $f(\mathfrak{p}) \to \max$, is attained at the boundary, $t = t_{\max}$. In our case, $t_{\max} = \min(\hat{P}_{i+1}, \hat{Q}_{j+1})$. If $t_{\max} = \hat{P}_{i+1}$, then $\mathfrak{p}_{i+1,j} = 0$. If $t_{\max} = \hat{Q}_{j+1}$, then from (9.37) we obtain $\mathfrak{p}_{i,j+1} = 0$. If $\hat{P}_{i+1} = \hat{Q}_{j+1}$), then $\mathfrak{p}_{i+1,j} = \mathfrak{p}_{i,j+1} = 0$.

A similar statement can be proved for the minimization problem.

Lemma 9.6 *Let* $\mathfrak{p}$ *be a solution to problem (9.28),* $f(\mathfrak{p}) \to \min$. *Then* $\mathfrak{p}$ *is anti-monotone.*

Monotonicity of the distributions is directly related to monotonicity of the samples of random variables from these distributions.

Lemma 9.7 *Consider a finite random sample, $Z = \{Z_n\}_{n=1}^N$, of independent, two-dimensional vectors $Z_n = (X_n, Y_n)'$ from a co-monotone distribution. Then, Z is co-monotone.*

If a random sample of N independent two-dimensional vectors, Z, has co-monotone coordinates for any integer $N \geq 2$, then Z is a sample from a co-monotone distribution.

Proof. Consider an independent, finite sample, Z, from a co-monotone distribution. Let us find a permutation, π, ordering the first coordinate, $\{X_n\}_{n=1}^N$, and apply this permutation to the second coordinate. Then we obtain

$$\hat{X}_1 \leq \hat{X}_2 \leq \cdots \leq \hat{X}_N, \quad \text{where } \hat{X} = \pi X.$$

If the second coordinate, $\hat{Y} = \pi Y$, is also ordered, the proof is finished. Suppose, on the contrary, there is a couple of indices, $i < j$, such that $\hat{Y}_i > \hat{Y}_j$; then the first coordinate must satisfy the relation $\hat{X}_i = \hat{X}_j$, otherwise the co-monotonicity condition of the support is not satisfied. Then the transposition of the elements, $\hat{Y}_i$ and $\hat{Y}_j$, puts these elements in ascending order. After a finite number of transpositions, we find a permutation, τ, ordering the second coordinate and keeping the first coordinate without changes. Then the permutation, $\tau \cdot \pi$, makes Z monotone coordinate-wise.

To prove the converse statement, let us assume that support of the bivariate distribution is not co-monotone. Then there exist two elements of the support, $s_k = (x_k, y_k)$ and $s_l = (x_l, y_l)$, such that $x_k \leq x_l \quad \text{but} \quad y_k > y_l$. Since both elements of the support have positive probabilities, there exists sufficiently large N such that the sample of independent vectors contains both vectors, s_k and s_l. Then Z is not co-monotone.

Lemma 9.8 is a direct analog of Lemma 9.7 for the anti-monotone distributions.

Lemma 9.8 *Consider a finite random sample, $Z = (Z_1, \ldots, Z_N)$, of independent, two-dimensional vectors $Z_n = (X_n, Y_n)'$ from an anti-monotone distribution. Then, the set $\{Z_n\}_{n=1}^N$ is anti-monotone. If a random sample of N independent two-dimensional vectors, Z, has anti-monotone coordinates for any integer $N \geq 2$, then Z is a sample from an anti-monotone distribution.*

Let us now find the joint probabilities and compute the support of the extreme distributions. We will use a probabilistic argument based on the Strong Law of Large Numbers (SLLN) (Johnson *et al.*, 1997). Let us start with the distribution $\mathfrak{p}^{**}$, maximizing the correlation coefficient. Consider a sample, $\{(\tilde{X}_n, \tilde{Y}_n)\}_{n=1}^N$, of the size, N, from the distribution $\mathfrak{p}^{**}$. There exists a permutation, π, such that the vectors $X = \pi\tilde{X}$ and $Y = \pi\tilde{X}$ both have monotone coordinates. As $N \to \infty$, we obtain two sequences of increasing length,

$$\begin{aligned} X: \quad & \overbrace{0, 0, \ldots, 0}^{N_X(0)}, \overbrace{1, 1, \ldots, 1}^{N_X(1)}, \overbrace{2, 2, \ldots, 2}^{N_X(2)}, \ldots, \overbrace{k, k, \ldots, k}^{N_X(k)}, \ldots \\ Y: \quad & \underbrace{0, \ldots, 0}_{N_Y(0)}, \underbrace{1, \ldots, 1}_{N_Y(1)}, \underbrace{2, \ldots, 2}_{N_Y(2)}, \ldots, \underbrace{k, k, k, \ldots, k, k}_{N_Y(k)}, \ldots, \end{aligned} \tag{9.38}$$

with the number of elements, $N_X(k)$ and $N_Y(k)$, in the sequences satisfying, by the SLLN, the relations:

$$\lim_{N\to\infty}\frac{N_X(k)}{N}=p_k, k=0,1,2,\dots, \text{almost surely,} \tag{9.39}$$
$$\lim_{N\to\infty}\frac{N_Y(k)}{N}=q_k, k=0,1,2,\dots, \text{almost surely.}$$

The probabilities, $P_i=\mathbb{P}(X_k\le i)$ and $Q_j=\mathbb{P}(Y_k\le j)$, satisfy

$$P_i=\sum_{l=0}^{i}p_l,\quad Q_j=\sum_{l=0}^{j}q_l,\quad i,j=0,1,2,\dots.$$

Let us denote the number of pairs, (i,j), by N_{ij}. In particular, the number of pairs, $(0,0)$, is

$$N_{00}=\min(N_X(0),N_Y(0)). \tag{9.40}$$

Lemma 9.9 *The limits,*

$$\mathfrak{p}^{**}_{ij}=\lim_{N\to\infty}\frac{N_{ij}}{N}$$

exist almost surely for all $i,j=0,1,2,\dots$, as $N\to\infty$. They satisfy equation (9.29).

Proof. In the case of the state $(0,0)$, existence of the limit, $\mathfrak{p}^{**}_{0,0}$, follows from the continuity of the function min, equation (9.40), and existence of the limits in (9.39). Let us prove the statement of the lemma for $i>0$ and $j>0$. Consider the sequence of vectors, (X_k,Y_k), $(k=1,2,\dots)$. It is not difficult to see that N_{ij} satisfies the relation

$$N_{ij}=[\min(M_X(i),M_Y(j))-\max(M_X(i-1),M_Y(j-1))]^+, \tag{9.41}$$

where

$$M_X(i)=\sum_{l=0}^{i}N_X(l)\quad\text{and}\quad M_Y(i)=\sum_{l=0}^{i}N_Y(l),$$

and, as usual, $x^+:=\max(0,x)$. Indeed, the index, k, of the pair, (i,j), in the sequence satisfies the inequalities $k\ge\max(M_X(i-1),M_Y(j-1))$ and $k\le\min(M_X(i),M_Y(j))$. Equation (9.41) follows from these inequalities.

Passing to the limit as $N\to\infty$, we obtain that almost surely

$$\lim_{N\to\infty}\frac{M_X(i)}{N}=P_i,\quad\text{and}\quad\lim_{N\to\infty}\frac{M_Y(i)}{N}=Q_i,\quad i=0,1,\dots.$$

Existence of the limits, $\mathfrak{p}^{**}_{ij}$, now follows from (9.41) and continuity of the functions $\min(\cdot,\cdot)$ and $\max(\cdot,\cdot)$. From (9.41), we find that the probability, $\mathfrak{p}_{i,j}$, of the state (i,j) satisfies:[8]

$$\mathfrak{p}^{**}_{i,j}=\min(P_i,Q_j)-\max(P_{i-1},Q_{j-1}).$$

Equation (9.29) is thus proved.

[8] If $\min(P_i,Q_j)\le\max(P_{i-1},Q_{j-1})$, then the vector (i,j) does not belong to the support.

In the case of minimal correlation, for any sample, $\{(\tilde{X}_n, \tilde{Y}_n)\}_{n=1}^N$, there exists a permutation, π, sorting the first coordinate in ascending order and the second coordinate in descending order:

$$X_1 \le X_2 \le \ldots \le X_N, \quad \text{where } X = \pi\tilde{X},$$

and the second coordinate forms a decreasing sequence:

$$Y_1 \ge Y_2 \ge \ldots \ge Y_N, \quad \text{where } Y = \pi\tilde{X}.$$

As $N \to \infty$, we obtain two sequences,

$$\begin{aligned} X: \quad & \overbrace{0,\ldots,0}^{N_X(0)},\ldots\overbrace{i-1,\ldots,i-1}^{N_X(i-1)},\overbrace{i,i,\ldots,i}^{N_X(i)},\ldots,\overbrace{k,k,\ldots,k}^{N_X(k)},\ldots \\ Y: \quad & \ldots,\underbrace{j+1,j+1\ldots,j+1}_{N_Y(j+1)},\underbrace{j,\ldots,j}_{N_Y(j)},\underbrace{j-1,\ldots,j-1}_{N_Y(j-1)},\ldots,\underbrace{0,\ldots,0,0}_{N_Y(0)} \end{aligned} \tag{9.42}$$

with the number of elements, $N_X(k)$ and $N_Y(k)$, satisfying, by the SLLN, equation (9.39) almost surely.

Lemma 9.10 *The limits* $\mathfrak{p}^*_{i,j} = \lim_{N\to\infty} \frac{N_{i,j}}{N}$ *exist almost surely for all* $i, j = 0, 1, 2, \ldots$, *as* $N \to \infty$, *and satisfy equation (9.31).*

Proof. The proof is a minor modification of that of Lemma 9.9. Let for some k, $X_k = i$ and $Y_k = j$. Notice that $k \ge N - M_Y(j)$ (see (9.42)). Then we have

$$N_{i,j} = \min(M_X(i), N - M_Y(j-1)) - \max(M_X(i-1), N - M_Y(j)), \tag{9.43}$$

Passing to the limit in (9.43), as $N \to \infty$, we derive (9.31).

Proof of Lemma 9.10 completes the derivation of Equations (9.29)–(9.31). Uniqueness of the extreme distributions, $\mathfrak{p}^*$ and $\mathfrak{p}^{**}$, follows immediately from (9.29)–(9.31). Theorem 9.5 is thus proved.

9.5.3 Computation of the Joint Distribution

Equation (9.29), in principle, allows one to find all the elements of the support of the maximal distribution and to compute the probabilities. In practice, this computation can be done much more efficiently by the algorithm described in this section. The idea of the extreme joint distribution (EJD) algorithm is to walk along the $\mathcal{S}$-path, instead of considering all possible pairs, (i, j), satisfying Equation (9.29).

It is useful to look again at the structure of the sample, represented by (9.38). When N is fixed, we have segments of a random length containing $N_X(k)$ elements equal to k. As $N \to \infty$ the length of the segment, the random variables, $N_X(i)$ and $N_Y(j)$, also tend to infinity, but the ratios (9.39) converge to finite limits.

In the limit, the scaled diagram (9.38) will look like two partitions of the unit interval, $[0, 1]$ with the nodes, $\Pi_X = \{P_0, P_1, \ldots\}$ and $\Pi_Y = \{Q_0, Q_1, \ldots\}$, defined by the cdf's.

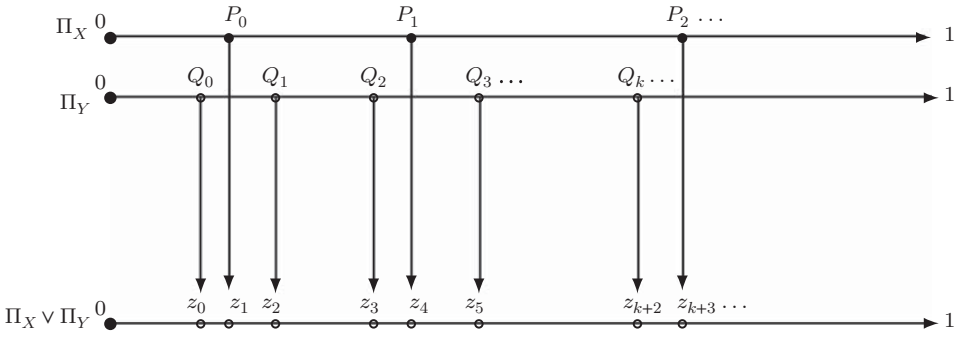

Figure 9.2 Partitions of the unit interval: $\text{corr}(X, Y) \to \max$.

The partition, $\Pi_Z = \Pi_X \vee \Pi_Y$, is formed by the union of the nodes of the partitions Π_X and Π_Y; see Figure 9.2. We denote the nodes of the partition, Π_Z, by $z_0 \le z_1 \le \ldots$.

It is convenient to assume that each node of Π_Z "remembers" the original partition it came from (Π_X or Π_Y), its index in the original partition, and its cumulative probability. The initial element of the support, $\{(x_k, y_k)\}_{k=0}^{\infty}$, is $(0, 0)$; the probability of this state is $z_0 = \min(P_0, Q_0)$. The algorithm scans the elements of the partition, Π_Z; finds the support of the extreme, bivariate distribution; and computes the joint probabilities, $\mathfrak{p}_{i,j}$.

9.5.3.1 EJD Algorithm

We shall start with the case of maximal correlation.

Step 0. $k := 0$; $x_k := 0, y_k := 0$; $\mathfrak{p}_{0,0} = z_0$.
Step 1. $k := k + 1$.
Step 2. If $z_{k-1} = P_i$ for some i and $z_{k-1} \neq Q_j$ for all j, then $x_k = i + 1$ and $y_k = y_{k-1}$. If $z_{k-1} = Q_j$ for some j and $z_{k-1} \neq P_i$ for all i, then $x_k = x_{k-1}$ and $y_k = j + 1$. If there exist such i and j that $z_{k-1} = P_i = Q_j$, then $x_k = i + 1$ and $y_k = j + 1$.
Step 3. Assign the kth element of the support, (x_k, y_k).
Step 4. $\mathfrak{p}^{**}_{x_k,y_k} := z_k - z_{k-1}$.
Step 5. Go to Step 1.

In the case of minimal correlation, we "flip" the distribution function and use the tail probabilities, $\ldots, 1 - Q_k, 1 - Q_{k-1}, \ldots, 1 - Q_0$, instead of Q_k (see Figure 9.3).

Remark 9.6 *One can think of the EJD algorithm as an observer moving along the $\mathcal{S}$-path in $\mathbb{Z}^2_+$; as such, the time to get to the kth node is z_k. Each time the observer's position is the kth node, (x_k, y_k), of the support, she records the coordinates of the node into the list and marks the node with the probability $\mathfrak{p}^{**}_{x_k,y_k}$.*

9.5.4 On the Frechet–Hoeffding Theorem

In the discrete case, the Frechet–Hoeffding theorem can be formulated as follows. Consider a space $\mathfrak{D}(\mathsf{P}, \mathcal{Q})$ of discrete bivariate distributions, with the marginal distributions, P and $\mathcal{Q}$.

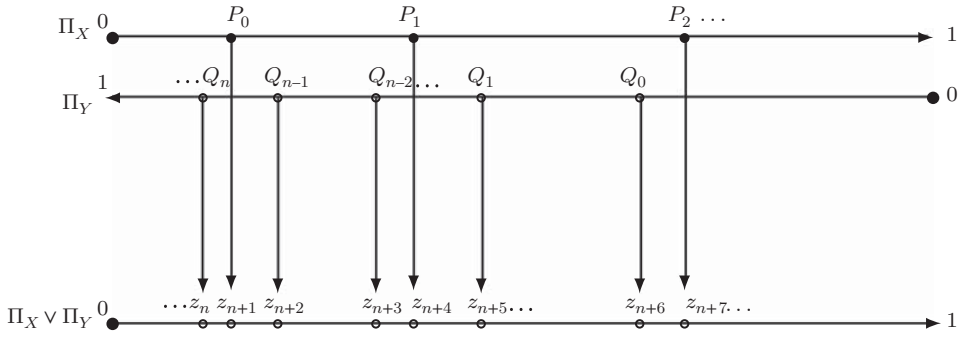

Figure 9.3 Partitions of the unit interval: $\text{corr}(X, Y) \to \min$.

Theorem 9.6 (Frechet, 1951; Hoeffding, 1940) *The bivariate cdf,* $H^*(i,j) := \mathbb{P}(X \le i, Y \le j)$*, maximizing the correlation of X and Y is*

$$H^*(i,j) = \min(P_i, Q_j), i,j \in \mathbb{Z}_+^2. \tag{9.44}$$

The bivariate cdf, $H_*(i,j)$*, minimizing the correlation of X and Y is*

$$H_*(i,j) = \max(0, P_i + Q_j - 1), i,j \in \mathbb{Z}_+^2. \tag{9.45}$$

Proposition 9.4 Theorem 9.5 is equivalent to the Frechet–Hoeffding theorem.

Proof. We shall only prove that Theorem 9.5 implies the Frechet–Hoeffding theorem in the case of maximimal correlation. It is enough to prove (9.44) in the case that the node, (i,j), belongs to the $\mathcal{S}$-path. We give the proof by induction on the index of the node.

Let us check the basis of induction. The first node is $(0,0)$. If $i = j = 0$, then $\mathbb{P}(X \le 0, Y \le 0) = \mathbb{P}(X = 0, Y = 0) = \min(P_0, Q_0)$. Thus, equation (9.44) is satisfied.

Suppose that the statement is true for the kth node, (i,j): $\mathbb{P}(X \le i, Y \le j) = \min(P_i, Q_j)$. Let us prove the statement for the $(k+1)$st node.

There are three possibilities: the $(k+1)$st node is $(i+1, j)$, the $(k+1)$st node is $(i+1, j+1)$, or the $(k+1)$st node is $(i, j+1)$. In the first case, using Equation (9.29), we find

$$\begin{aligned}\mathbb{P}(X \le i+1, Y \le j) &= \mathbb{P}(X \le i, Y \le j) + \mathfrak{p}_{i+1,j} \\ &= \min(P_i, Q_j) + \min(P_{i+1}, Q_j) - \max(P_i, Q_{j-1}).\end{aligned}$$

Since the $\mathcal{S}$-path contains the arrow from the node (i,j) to $(i+1, j)$, we have $P_i < Q_j$. On the other hand, since the second coordinate of the node is j, $P_i > Q_{j-1}$. Thus, $\max(P_i, Q_{j-1}) = P_i$. Then we obtain that

$$\min(P_i, Q_j) = \max(P_i, Q_{j-1}) = P_i.$$

Therefore,

$$\mathbb{P}(X \le i+1, Y \le j) = \min(P_{i+1}, Q_j),$$

as was to be proved in the first case.

In the second case, the $(k+1)$st node is $(i+1, j+1)$. The diagonal segment in the $\mathcal{S}$-path appears only if $P_i = Q_j$. Then we have

$$\begin{aligned}\mathbb{P}(X \le i+1, Y \le j+1) &= \mathbb{P}(X \le i, Y \le j) + \mathfrak{p}_{i+1,j+1} \\ &= \min(P_i, Q_j) + \min(P_{i+1}, Q_{j+1}) - \max(P_i, Q_j) \\ &= \min(P_{i+1}, Q_{j+1}).\end{aligned}$$

The last case is analogous to the first one.

9.5.5 *Approximation of the Extreme Distributions*

For practical computations, the marginal distributions, $\mathsf{P} = \{p_i\}_{i\ge 0}$ and $\mathcal{Q} = \{q_j\}_{j\ge 0}$, with two finite moments can be approximated by the distributions with finite support, $\tilde{\mathsf{P}} = \{\tilde{p}_i\}_{i\ge 0}$ and $\tilde{\mathcal{Q}} = \{\tilde{q}_j\}_{j\ge 0}$, such that

$$\max_{i\le I_*} |P_i - \tilde{P}_i| \le \epsilon, \quad 1 - P_{I_*} \le \epsilon,$$

$$\max_{j\le J_*} |Q_j - \tilde{Q}_j| \le \epsilon, \quad 1 - Q_{J_*} \le \epsilon,$$

where $P_n = \sum_{i=0}^{n} p_i$, $Q_m = \sum_{j=0}^{m} q_j$, $\tilde{P}_n = \sum_{i=0}^{n} \tilde{p}_i$, and $\tilde{Q}_m = \sum_{j=0}^{m} \tilde{q}_j$. Let $\mathfrak{p} \in \mathfrak{D}(\mathsf{P}, \mathcal{Q})$ be an extreme bivariate distribution and $\tilde{\mathfrak{p}} \in \mathfrak{D}(\tilde{\mathsf{P}}, \tilde{\mathcal{Q}})$ be the corresponding extreme bivariate distribution with the marginals, $\tilde{\mathsf{P}}$ and $\tilde{\mathcal{Q}}$. Then it follows from Theorem 9.5 that $\tilde{\mathfrak{p}}$ satisfies the inequality

$$\sup_{\substack{i\ge 0,\\ j\ge 0}} |\mathfrak{p}_{i,j} - \tilde{\mathfrak{p}}_{i,j}| \le 2\epsilon.$$

Similar inequality can be obtained for the approximation of the correlation coefficient. We shall formulate one sufficient condition that guarantees the approximation.

Proposition 9.5 Suppose that the first and the second moments of the marginal distributions are finite. Then, given $\epsilon > 0$, one can find the integer numbers, I_* and J_*, such that there exists a discrete bivariate distribution $\tilde{\mathfrak{p}}_{i,j}, 0 \le i \le I_*, 0 \le j \le J_*$ approximating the extreme distribution, $\mathfrak{p}_{i,j}$, in the following sense:

$$\sup_{i\ge 0} \sup_{j\ge 0} | \mathfrak{p}_{i,j} - \tilde{\mathfrak{p}}_{i,j} | < \epsilon, \tag{9.46}$$

$$\left| \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} ij \cdot (\mathfrak{p}_{i,j} - \tilde{\mathfrak{p}}_{i,j}) \right| < 3\epsilon. \tag{9.47}$$

Proof. We consider the case of the distribution, $\mathfrak{p} = \mathfrak{p}^{**}$, maximizing the correlation coefficient. Notice that the first two finite moments of a discrete random variable taking nonnegative,

integer values satisfy the relations

$$\mathbb{E}[\xi^2] = 2\sum_{k=1}^{\infty} k\mathbb{P}(\xi > k) + \mathbb{E}[\xi], \tag{9.48}$$

$$\mathbb{E}[\xi] = \sum_{k=0}^{\infty} \mathbb{P}(\xi > k). \tag{9.49}$$

Since $\mathbb{E}[\xi^2] < \infty$, Equations (9.48) and (9.49) imply convergence of the series $\sum_{k=1}^{\infty} k\mathbb{P}(\xi > k)$. Then we can find I_* and J_* such that

$$\sum_{i>I_*} i(1 - P_i) < \epsilon/2, \tag{9.50}$$

$$\sum_{j>J_*} j(1 - Q_j) < \epsilon/2. \tag{9.51}$$

Consider the distribution functions, $\tilde{P}_i$ and $\tilde{Q}_j$:

$$\tilde{P}_i = P_i, i = 0, 1, \dots, I_* - 1;\ \tilde{P}_i = 1, i \geq I_*, \tag{9.52}$$

$$\tilde{Q}_j = Q_j, j = 0, 1, \dots, J_* - 1;\ \tilde{Q}_j = 1, j \geq J_*. \tag{9.53}$$

Let $\tilde{\mathfrak{p}}$ be the extreme bivariate distribution, maximizing the correlation, corresponding to the marginals, $\tilde{\mathsf{P}}$ and $\tilde{\mathcal{Q}}$. It is not difficult to see that $\tilde{\mathfrak{p}}_{i,j} = 0$ for $i \geq I_*$ or $j > J_*$.

The first moments of the marginal distributions satisfy the inequalities

$$\left|\sum_{i=0}^{\infty}(1 - P_i) - \sum_{i=0}^{\infty}(1 - \tilde{P}_i)\right| < \frac{\epsilon}{I_*}, \text{ and } \left|\sum_{j=0}^{\infty}(1 - Q_j) - \sum_{j=0}^{\infty}(1 - \tilde{Q}_j)\right| < \frac{\epsilon}{J_*}.$$

Indeed,

$$\left|\sum_{i=0}^{\infty}(1 - P_i) - \sum_{i=0}^{\infty}(1 - \tilde{P}_i)\right| = \left|\sum_{i=I_*}^{\infty}(1 - P_i) - \sum_{i=I_*}^{\infty}(1 - \tilde{P}_i)\right|$$
$$\leq \frac{1}{I_*}\sum_{i=I_*}^{\infty} i\cdot(1 - P_i).$$

The second inequality is derived analogously. Consider the expected values

$$\mathcal{E} := \mathbb{E}[XY] = \sum_{i\geq 1}\sum_{j\geq 1} ij\mathfrak{p}^{**}_{i,j} \quad \text{and} \quad \tilde{\mathcal{E}} := \sum_{i\geq 1}\sum_{j\geq 1} ij\tilde{\mathfrak{p}}_{i,j}.$$

We have

$$\mathcal{E} - \tilde{\mathcal{E}} = \sum_{i=I_*}^{\infty}\sum_{j=J_*}^{\infty} ij\mathfrak{p}^{**}_{ij} + \sum_{i=I_*}^{\infty}\sum_{j=1}^{J_*} ij\mathfrak{p}^{**}_{ij} + \sum_{j=J_*}^{\infty}\sum_{i=1}^{I_*} ij\mathfrak{p}^{**}_{ij}$$

The first sum, $S_1 := \sum_{i=I_*}^{\infty} \sum_{j=J_*}^{\infty} ij\mathfrak{p}_{ij}^{**}$, satisfies the inequality

$$S_1^2 \le \sum_{i=I_*}^{\infty} i^2 \sum_{j=0}^{\infty} \mathfrak{p}_{i,j}^{**} \cdot \sum_{j=J_*}^{\infty} j^2 \sum_{i=0}^{\infty} \mathfrak{p}_{i,j}^{**} < \epsilon^2.$$

The second sum, $S_2 := \sum_{i=I_*}^{\infty} \sum_{j=1}^{J_*} ij\mathfrak{p}_{ij}^{**}$, satisfies the inequality

$$S_2 \le J_* \sum_{i=I_*}^{\infty} \sum_{j=0}^{\infty} i\mathfrak{p}_{ij}^{**} \le J_* \frac{\epsilon}{J_*} = \epsilon,$$

and the third sum, $S_3 := \sum_{j=J_*}^{\infty} \sum_{i=1}^{I_*} ij\mathfrak{p}_{ij}^{**}$, also satisfies $S_3 \le \epsilon$. Putting all the pieces together, we obtain $\mid \mathcal{E} - \tilde{\mathcal{E}} \mid \le 3\epsilon$.

Remark 9.7 *Similar reasoning allows us to prove continuity of the function* $f(\mathfrak{p}) = \sum_{i\ge1}\sum_{j\ge1} ij\mathfrak{p}_{i,j}$ *on the space* $\mathfrak{D}(P, Q)$.

9.6 Numerical Results

9.6.1 Examples of the Support

We shall start with the support of the joint distribution maximizing the correlation coefficient when $\lambda = \mu$. In this case, the support is a set of integer points, $\{(k,k)\}_{k=0}^{\infty}$, on the main diagonal; the correlation coefficient is $\rho^{**}(\lambda,\lambda) = 1$.

Figure 9.4 displays the support of the extreme distribution, $\mathfrak{p}^{**}$, with the intensities $\lambda = 3$ and $\mu = 4$. We apply the EJD algorithm with ϵ-approximation[9] of the marginal distributions, where $\epsilon = 10^{-5}$. It allows us to find (numerically) the joint probabilities, $\mathfrak{p}_{ij}^{**}$, and the correlation coefficient,

$$\rho^{**}(\lambda,\mu) = \frac{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} ij\mathfrak{p}_{ij}^{**} - \lambda\mu}{\sqrt{\lambda\mu}}. \tag{9.54}$$

The error of the approximation of the correlation coefficient $\epsilon(\rho) < 10^{-4}$. The maximal correlation $\rho^{**}(\lambda,\mu) = 0.979$ in this case. One can notice that the support of the joint distribution is located close to the main diagonal, but for large values of the first coordinate, some deviation from the diagonal pattern is observed. The reason for that is inequality of the intensities of the Poisson processes, $\lambda < \mu$, leading to a re-balance of the probability mass of the bivariate distribution maximizing the correlation coefficient.

In the case $\lambda = 3, \mu = 6$, the deviation of the support from the main diagonal increases as $k \to \infty$ (see Figure 9.5). Applying the EJD algorithm, we find the joint probabilities, $\mathfrak{p}_{ij}^{**}$, and from (9.54) the correlation coefficient, $\rho^{**}(3,6) = 0.977$.

Let us now consider an example of the support for the minimal (negative) correlation of the random variables. We take $\lambda = 3$ and $\mu = 6$. The joint probabilities, $\mathfrak{p}_{ij}^{*}$, can be found by the EJD algorithm. In this case, the minimal value of the correlation coefficient is $\rho_* = -0.944$. The geometric pattern of the support is different in this case because the minimal value of the correlation coefficient is negative.

[9] The same approximation of the marginal distributions is used in all examples considered in this section.

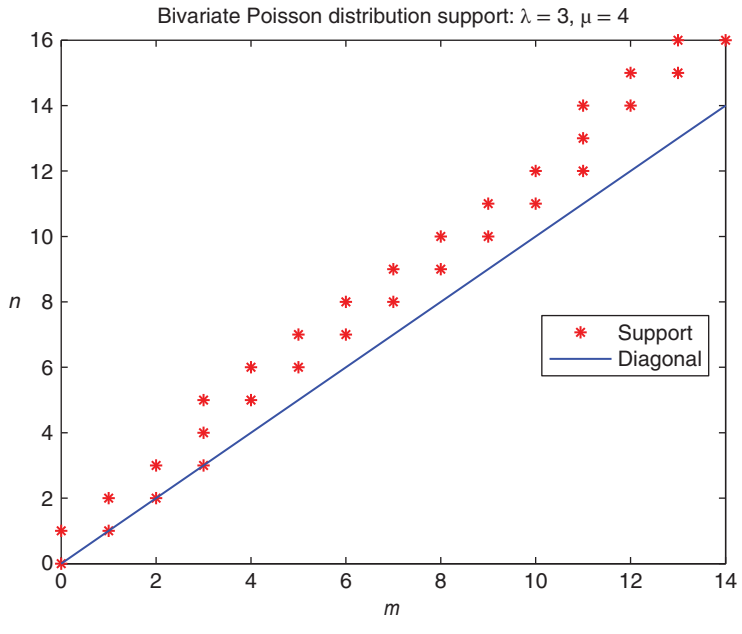

Figure 9.4 Support of the distribution $\mathfrak{p}^{**}$: $\lambda = 3, \mu = 4$.

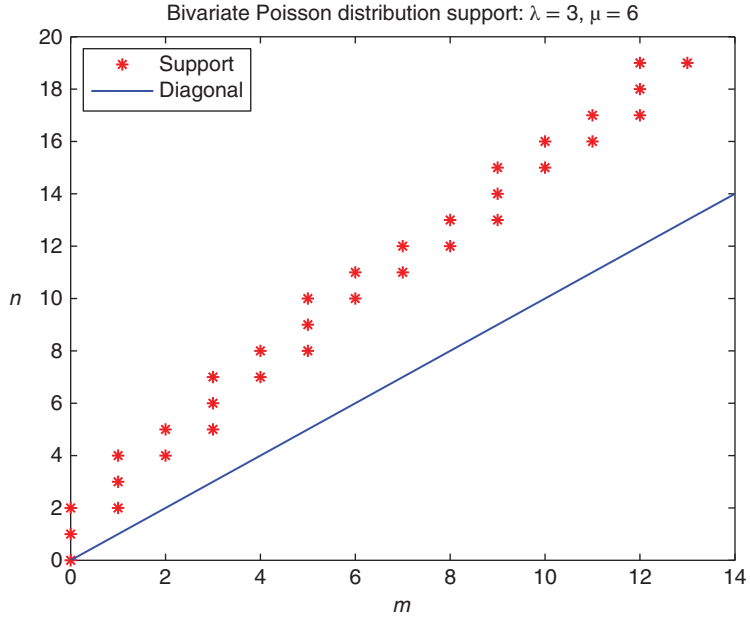

Figure 9.5 Support of the distribution $\mathfrak{p}^{**}$: $\lambda = 3, \mu = 6$.

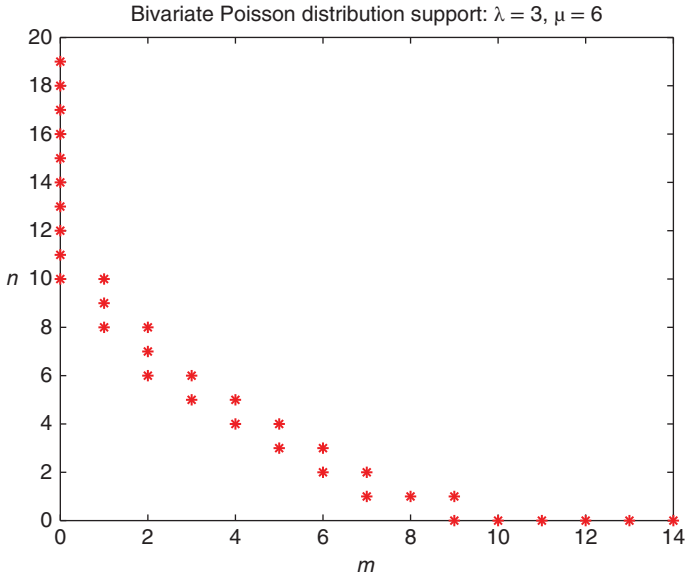

Figure 9.6 Support of the distribution $\mathfrak{p}^*$: $\lambda = 3$, $\mu = 6$.

9.6.2 *Correlation Boundaries*

The correlation coefficients, $\rho^*(\lambda, \mu)$ and $\rho^{**}(\lambda, \mu)$, are not monotone functions of λ as intensity μ is fixed.[10] The extreme correlation coefficients are shown in Figure 9.7 for $\mu = 3$. We also mention here that the lower and upper boundaries are not symmetrical. It is proved in Griffiths *et al*., (1979) that the two-dimensional process with variable correlation between the components cannot be infinitely divisible. The latter fact may look surprising because the marginals are the Poisson distributions belonging to the class of infinitely divisible distributions.

Let us now compare the correlation boundaries obtained using the traditional, forward simulation method and the BS method. We shall call the boundaries computed by the forward simulation (FS) the FS boundaries. The boundaries computed by the backward simulation method will be called the BS boundaries.

In Figure 9.8, the correlation boundaries are displayed for the intensities $\lambda = 8$ and $\mu = 4$. The time horizon $T = 3$. Recall that under the FS approach, we have to generate two sequences, ΔT_k and $\Delta T'_k$, such that

$$\mathbb{P}(\Delta T_k \le t) = 1 - e^{-\lambda t}, \quad \mathbb{P}(\Delta T'_k \le t) = 1 - e^{-\mu t}, \ \ t \ge 0,$$

and the random variables ΔT_k and $\Delta T'_k$ have minimal possible correlation for each k; at the same time, each of the sequences $\{\Delta T_k\}_{k\ge 0}$ and $\{\Delta T'_k\}_{k\ge 0}$ is formed by independent identically distributed random variables.

[10] This fact is established numerically. The joint probabilities should be computed with the error not exceeding 10^{-4} to provide proof of this fact.

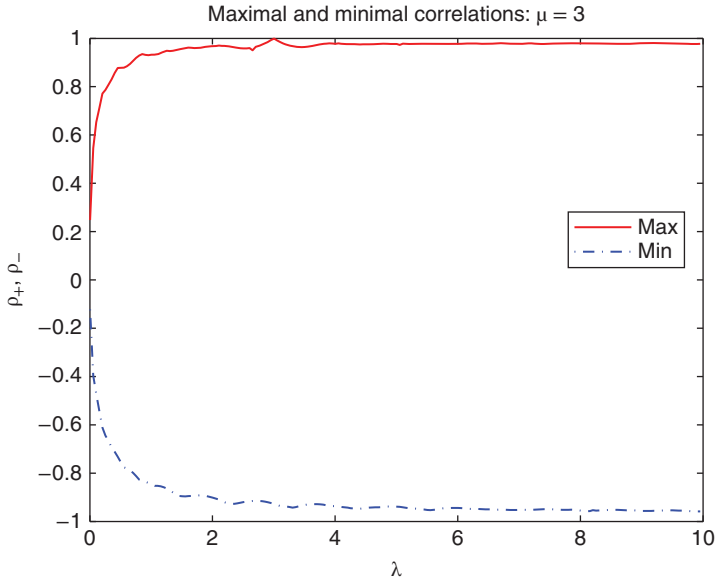

Figure 9.7 Correlation boundaries: $\mu = 3$.

According to the Frechet–Hoeffding theorem, the interarrival times, ΔT_k and $\Delta T'_k$, should satisfy the relation

$$\exp\ (-\lambda \cdot \Delta T_k) + \exp\ (-\mu \cdot \Delta T'_k) = 1, \quad k = 1, 2, \dots,$$

to provide minimal possible correlation between the interarrival times. The correlation coefficient, $\rho(t) = \text{corr}(N_t, N'_t)$, is computed using a Monte Carlo simulation with 900,000 Monte Carlo samples.

The computation of the FS upper boundary in Figure 9.8 is also based on the Frechet–Hoeffding theorem. According to the latter, the interarrival times, ΔT_k and $\Delta T'_k$, should satisfy the relation

$$\mu \Delta T_k = \lambda \Delta T'_k, k = 1, 2, \dots,$$

to provide maximal possible correlation between the arrival moments. One can see that the maximal correlation of the processes, N_t and N'_t, under the FS is a constant, while under the BS it is a linear function of time. In the case $\lambda \neq \mu$, the BS allows one to reach stronger correlations of the processes than the FS of the arrival moments.

9.7 Backward Simulation of the Poisson–Wiener Process

The BS approach is applicable to the class of multivariate stochastic process with the components formed by either Poisson or Gaussian processes. In this section, we analyze the basic bivariate process, $\boldsymbol{X_t} = (N_t, W_t)$, where N_t is a Poisson process with intensity μ; and W_t is the Wiener process.

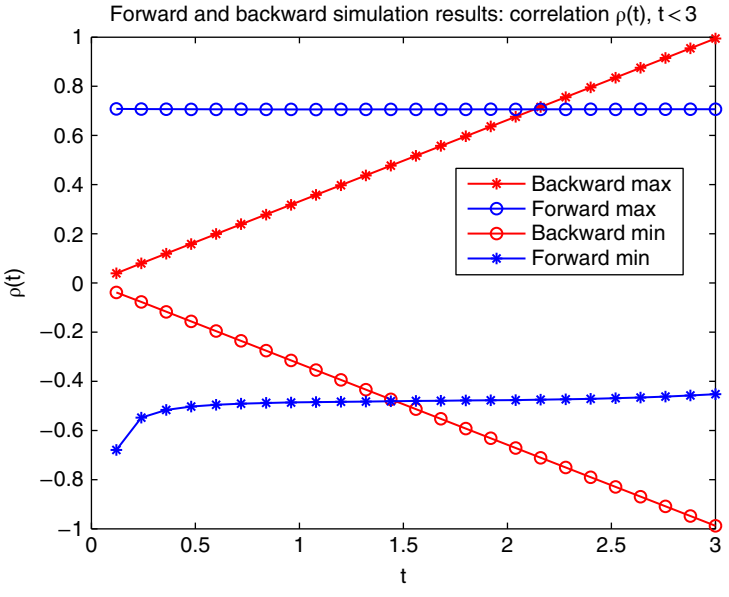

Figure 9.8 Comparison of correlation boundaries: $\lambda = 8, \mu = 4$.

Suppose that the joint distribution of $\boldsymbol{X_t}$ at time $t = T$ is known. Let $\{X(k)\}_{k=1}^{M}, X(k) = (N_T(k), W_T(k))$ be a random sample of the size M from this distribution. For each k, we apply the BS technique to simulate the first coordinate and the Brownian bridge construction to simulate the second coordinate.

More precisely, consider the kth random vector, $X(k) = (N_T(k), W_T(k))$, $(k = 1, 2, \ldots, M)$. Simplifying the notation, denote $n = N_T(k)$ and $x = W_T(k)$. Let us generate the arrival moments of the Poisson process, $0 \le T_1 < T_2 < \ldots < T_n \le T$. As before, the arrival moments are obtained by sorting the sequence of n independent identically distributed random variables, having a uniform distribution in the interval $[0, T]$, in ascending order.

The second coordinate of the process, $\boldsymbol{X_t}$, is generated recursively using the Brownian bridge construction. Suppose that the process W_t has already been generated at times $T_n > T_{n-1} > \ldots > T_k$ and $W_{T_k} = x_k$. Then, at time $t = T_{k-1}$, W_t has a normal distribution, $\mathcal{N}(m(t), \sigma^2(t))$, with the parameters

$$m(T_{k-1}) = \frac{x_k \cdot T_{k-1}}{T_k}, \quad \sigma^2(T_{k-1}) = \sqrt{\frac{T_{k-1} \cdot (T_k - T_{k-1})}{T_k}}.$$

Consider now the correlation function, $\rho(t) = \mathrm{corr}(N_t, W_t)$, of the process, $\boldsymbol{X_t}$, $(0 \le t \le T)$. Our first result is the computation of the extreme correlations of this process at time T. After that we will prove that, as in the case of the BS simulation of the Poisson processes, the correlation coefficient $\rho(t)$ in our case is the linear function of time, $\rho(t) = \frac{t}{T}\rho(T)$.

Let us now find the extreme values of the correlation coefficient, $\rho(T) = \mathrm{corr}(N_T, W_T)$. It is convenient to represent the random variable, $\zeta = N_T$, as a function of a normally distributed random variable, $\eta \sim \mathcal{N}(0, 1)$, $\zeta = P_\lambda^{-1}(\Phi(\eta))$, where $\lambda = \mu T$. The random variable W_T can be

represented as $W_T = \sqrt{T}\xi$, where $\xi \sim \mathcal{N}(0,1)$ is a random variable with the standard normal distribution. Then the covariance of the Poisson and Wiener processes at time T is

$$\mathrm{cov}(N_T, W_T) = \sqrt{T}\mathrm{cov}(\xi, \zeta) = \sqrt{T}\cdot\mathbb{E}[\xi\cdot\zeta].$$

Since the variance of W_T is $\sqrt{T}$, we find

$$\rho(T) = \frac{\sqrt{T}\mathbb{E}[\zeta\xi]}{\sqrt{T}\sqrt{\mu T}} = \mathrm{corr}(\xi, \zeta).$$

Denote by ρ_0 the correlation coefficient of ξ and η. Let us represent ξ as $\xi = \rho_0\eta + \sqrt{1-\rho_0^2}\eta'$, where η' is a standard normal random variable independent of η. We have

$$\mathbb{E}[\xi\cdot\zeta] = \rho_0\cdot\Psi(\lambda), \tag{9.55}$$

where $\Psi(\lambda) := \mathbb{E}[\eta P_\lambda^{-1}(\Phi(\eta))]$. Let us compute $\Psi(\lambda)$. The function $P_\lambda(x)$ is piecewise constant: $P_\lambda(x) = P_\lambda([x])$. Define $\beta_k := P_\lambda(k)$, $(k = 0, 1, \dots)$. If $\beta_{k-1} \le \Phi(\eta) < \beta_k$, then $P_\lambda^{-1}(\Phi(\eta)) = k-1$. Denote $\gamma_k(\lambda) := \Phi^{-1}(P_\lambda(k))$. Then we find

$$\Psi(\lambda) = \sum_{k=1}^{\infty}\int_{\gamma_k(\lambda)}^{\gamma_{k+1}(\lambda)} kx\varphi(x)\mathrm{d}x = \sum_{k=1}^{\infty} k\cdot(\varphi(\gamma_k) - \varphi(\gamma_{k+1})),$$

where $\varphi(x)$ is the standard normal density function. The function $\Psi(\lambda)$ is a smooth, monotone function of λ. Then, from (9.55) and the inequality, $|\rho_0| \le 1$, we obtain

Proposition 9.6 The correlation coefficient, $\rho(\xi, \zeta)$, satisfies the inequality

$$-\frac{1}{\sqrt{\lambda}}\Psi(\lambda) \le \rho(\xi, \zeta) \le \frac{1}{\sqrt{\lambda}}\Psi(\lambda), \quad \lambda > 0. \tag{9.56}$$

Figure 9.9 displays the low and upper bounds for the correlation coefficient $\rho(\xi, \zeta)$. Comparing Figures 9.7 and 9.9, we conclude that the correlation boundaries in Proposition 9.6 are smooth, monotone, and symmetric, as opposed to the boundaries in the case of the Poisson processes.

Let us now study the correlation coefficient, $\rho(t) = \mathrm{corr}(N_t, W_t)$.

Proposition 9.7 The correlation coefficient, $\rho(t)$, is a linear function of time:

$$\rho(t) = \frac{t}{T}\cdot\rho(T), 0 \le t \le T. \tag{9.57}$$

Proof. Denote by $p_t(x, k)$ the joint density function of the Brownian motion and the Poisson process at time t. We have

$$\int_{-\infty}^{\infty} p_t(x,k)\mathrm{d}x = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \quad \sum_{k=0}^{\infty} p_t(x,k) = \frac{e^{-\frac{x^2}{2t}}}{\sqrt{2\pi t}}.$$

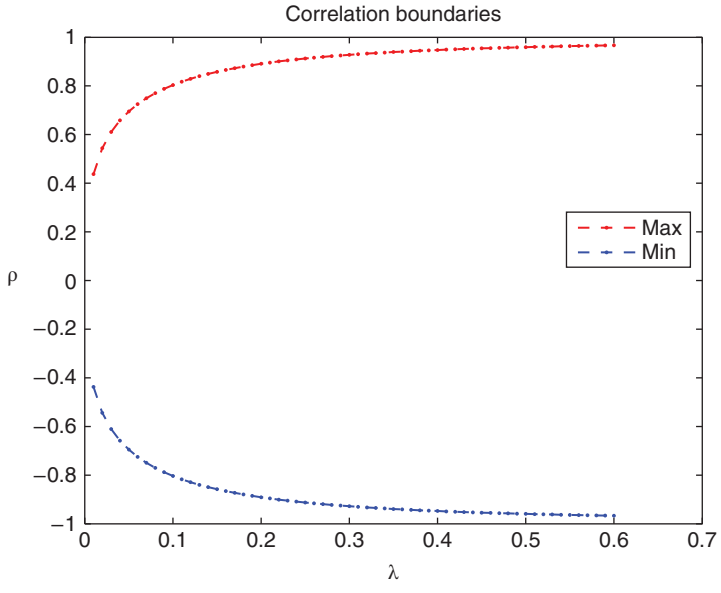

Figure 9.9 Correlation bounds.

Let

$$\hat{p}_t(x,z) := \sum_{n=0}^{\infty} p_t(x,n)z^n, \quad |z| \le 1,$$

be the generating function of the sequence $\{p_t(x,n)\}_{n=0}^{\infty}$. Consider the function

$$\Psi_t(u,z) := \mathbb{E}[\exp\ (iuW_t)z^{N_t}], \quad u \in \mathbb{R}, |z| \le 1.$$

The function, Ψ_t, is analytical in the area $|z| \le 1, t > 0$; its second derivative satisfies

$$\left.\frac{\partial^2 \Psi_t(u,z)}{\partial u \partial z}\right|_{(u,z)=(0,1)} = i \cdot \mathbb{E}[W_t \cdot N_t].$$

It is convenient to define the function

$$\psi_t(u,z) = -i \cdot \frac{\partial^2 \Psi_t(u,z)}{\partial u \partial z}.$$

Then we have

$$\psi_t(0,1) = \mathbb{E}[W_t \cdot N_t]. \tag{9.58}$$

Taking into account that

$$\mathbb{E}[W_t] = 0, \quad \sigma^2(W_t) = t, \quad \mathbb{E}[N_t] = \lambda t, \quad \sigma^2(N_t) = \lambda t, \quad 0 \le t \le T, \tag{9.59}$$

we derive from (9.58) and (9.59)

$$\psi_t(0,1) = \rho(t) \cdot \sqrt{\lambda} \cdot t. \tag{9.60}$$

Let us apply the BS algorithm to the process (W_t, N_t). The process N_t satisfies the relation

$$\mathbb{P}(N_t = m \mid N_T = n) = \binom{n}{m} q^m (1-q)^{n-m}, \quad m = 0, 1, \dots, n,$$

where $q = t/T$. The relation between $\Psi_t(u, z)$ and $\Psi_T(u, z)$ can be derived as follows. At time t, we have

$$\begin{aligned}\Psi_t(u, z) &= \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} e^{iux} z^n p_t(x, n)\, \mathrm{d}x \\ &= \int_{-\infty}^{\infty} e^{iux} \hat{p}_t(x, z)\, \mathrm{d}x.\end{aligned}$$

Lemma 9.11 *The generating function, $\hat{p}_t$, satisfies the equation*

$$\hat{p}_t(y, z) = \int_{-\infty}^{\infty} \hat{p}_T(x, 1-q+zq) \frac{1}{\sqrt{2\pi}\sigma(t)} e^{-\frac{(y-xt/T)^2}{2\sigma^2(t)}}\, dx, \quad 0 < t < T, \tag{9.61}$$

where $q = t/T$ and $\sigma(t) = \sqrt{\frac{t\cdot(T-t)}{T}}$.

Proof. The density, $p_t(y, m)$, satisfies

$$p_t(y, m) = \int_{-\infty}^{\infty} \sum_{n=m}^{\infty} \binom{n}{m} q^m (1-q)^{n-m} p_T(x, n) \frac{1}{\sqrt{2\pi}\sigma(t)} e^{-\frac{(y-xt/T)^2}{2\sigma^2(t)}}\, \mathrm{d}x.$$

We have

$$\begin{aligned}\sum_{m=0}^{\infty} z^m \sum_{n=m}^{\infty} \binom{n}{m} q^m (1-q)^{n-m} p_T(x, n) &= \sum_{n=0}^{\infty} p_T(x, n) \sum_{m=0}^{n} \binom{n}{m} (qz)^m (1-q)^{n-m} \\ &= \sum_{n=0}^{\infty} p_T(x, n)(1-q+qz)^n \\ &= \hat{p}_T(x, 1-q+qz).\end{aligned}$$

Therefore, the generating function, $\hat{p}_t(y, z)$, satisfies (9.61), as was to be proved.

Let us finish the proof of Proposition 9.7. The joint generating function, $\Psi_t(u, z)$, is the Fourier transform of $\hat{p}_t$. Therefore,

$$\begin{aligned}\Psi_t(u, z) &= \int_{-\infty}^{\infty} e^{iuy} \hat{p}_t(y, z)\, \mathrm{d}y \\ &= \int_{-\infty}^{\infty} e^{iuy} \int_{-\infty}^{\infty} \hat{p}_T(x, 1-q+zq) \frac{1}{\sqrt{2\pi}\sigma(t)} e^{-\frac{(y-xt/T)^2}{2\sigma^2(t)}}\, \mathrm{d}x\, \mathrm{d}y.\end{aligned}$$

Changing the order of integration, we obtain

$$\Psi_t(u,z) = \int_{-\infty}^{\infty} e^{iuy} \int_{-\infty}^{\infty} \hat{p}_T(x, 1-q+zq) \frac{1}{\sqrt{2\pi}\sigma(t)} e^{-\frac{(y-xt/T)^2}{2\sigma^2(t)}} \mathrm{d}x\, \mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \hat{p}_T(x, 1-q+zq) \int_{-\infty}^{\infty} e^{iuy} \frac{1}{\sqrt{2\pi}\sigma(t)} e^{-\frac{(y-xt/T)^2}{2\sigma^2(t)}} \mathrm{d}y\, \mathrm{d}x$$

The inner integral

$$\int_{-\infty}^{\infty} e^{iuy} \frac{1}{\sqrt{2\pi}\sigma(t)} e^{-\frac{(y-xt/T)^2}{2\sigma^2(t)}} \mathrm{d}y = e^{ixqu} \cdot e^{-\frac{u^2\sigma^2(t)}{2}}.$$

Therefore,

$$\Psi_t(u,z) = e^{-\frac{u^2\sigma^2(t)}{2}} \cdot \int_{-\infty}^{\infty} e^{ixqu} \cdot \hat{p}_T(x, 1-q+zq)\, \mathrm{d}x, \tag{9.62}$$

where $q = t/T$. Differentiating (9.62) twice and substituting $u = 0$, $z = 1$, we obtain

$$\psi_t(0,1) = \frac{t^2}{T^2} \cdot \int_{-\infty}^{\infty} x \frac{\partial}{\partial z} \hat{p}_T(x,1)\mathrm{d}x$$

$$= \frac{t^2}{T^2} \cdot \psi_T(0,1).$$

Then, the latter relation and (9.60) immediately imply Equation (9.57).

Remark 9.8 *We have considered the backward simulation of the process with Poisson and Wiener components. It is possible to extend this approach to the class of processes with mean-reverting components instead of Wiener processes. In this case, the process has a nonlinear time structure of correlations.*

9.8 Concluding Remarks

The backward simulation of the multivariate Poisson processes is based on the conditional uniformity of the unsorted arrival moments. The BS allows us to consider in a general framework both Gaussian and Poisson processes describing the dynamics of risk factors. The BS approach to the Poisson processes results in a bigger range of admissible correlations and simpler calibration algorithm.

The computation of the extreme distributions with maximal and minimal correlations of the Poisson processes is solved by the EJD algorithm, proposed in the present chapter, which has a simple probabilistic interpretation. This algorithm has an intimate connection to the class of infinite-dimensional linear optimization problems. It is interesting to understand its role in the analysis of this class of optimization problems.

The boundaries for the correlation coefficients of the multivariate Poisson process delivered by the EJD algorithm give necessary conditions for the existence of the process. It looks natural to try to prove that these conditions are also sufficient. A similar question can be formulated for the multivariate process with the Wiener and Poisson components.

The BS approach is applicable to the class of mixed Poisson processes. These models will be studied in our future publications.

Acknowledgments

I am very grateful to my colleagues Ian Iscoe, Yijun Jiang, Konrad Duch, Helmut Mausser, Oleksandr Romanko, and Asif Lakhany for stimulating discussions: without their support, this chapter would never have appeared.

Andrey Marchenko read the first draft of the paper and made a few important comments and suggestions, especially regarding monotone distributions. I would like to thank Isaac Sonin for his comments pointing at simplification of the proofs and Monique Jeanblanc and Raphael Douady for fruitful discussions during the 8th Bachelier Colloquium. Sebastian Jaimungal and Tianyi Jia made numerous comments and suggestions helping me to write this chapter. Tianyi also verified some of the proofs and derivations. I am very grateful to the referees for their comments. Finally, I would like to thank Dmitry Malioutov for his interest in this project.

Appendix A

A.1 Proof of Lemmas 9.2 and 9.3

A.1.1 Proof of Lemma 9.2

Consider the function, $\pi : \mathbb{Z}_+^m \to \mathbb{R}$,

$$\pi(\vec{k}) = \sum_{l=0}^{\infty} p_{\|\vec{k}\|+l} \binom{\vec{k}+l}{\vec{k}} \cdot \vec{x}^{\,\vec{k}} \cdot y^l, \tag{A.1}$$

and denote by $\hat{\pi}(\vec{z})$ the generating function

$$\hat{\pi}(\vec{z}) := \sum_{\vec{k}\in\mathbb{Z}_+^m} \pi(\vec{k}) \vec{z}^{\,\vec{k}}.$$

We have to show that

$$\hat{\pi}(\vec{z}\,) = \hat{p}\left(1 - \sum_{j=1}^{m} x_j(1-z_j)\right). \tag{A.2}$$

The generating function $\hat{\pi}(\vec{z}\,)$ can be written as

$$\hat{\pi}(\vec{z}\,) = \sum_{j=1}^{m} \sum_{k_j=0}^{\infty} \sum_{l=0}^{\infty} p_{\|\vec{k}\|+l} \vec{z}^{\,\vec{k}} \binom{\vec{k}+l}{\vec{k}} \cdot \vec{x}^{\,\vec{k}} \cdot y^l. \tag{A.3}$$

Denote $n = \sum_{j=1}^{m} k_j + l$. Let us also introduce the partial sums

$$K_J = \sum_{j=1}^{J} k_j, \quad J = 1, 2, \ldots, m.$$

Then, from Equation (A.3), we find

$$\hat{\pi}(\vec{z}\,) = \sum_{n=0}^{\infty} p_n \cdot \sum_{k_1=0}^{n} \binom{n}{k_1} (x_1 z_1)^{k_1} \sum_{k_2=0}^{n-k_1} \binom{n-k_1}{k_2} (x_2 z_2)^{k_2} \sum_{k_3=0}^{n-K_2} \binom{n-K_2}{k_3} (x_3 z_3)^{k_3} \cdot$$
$$\ldots \sum_{k_m=0}^{n-K_{m-1}} \binom{n-K_{m-1}}{k_m} (x_m z_m)^{k_m} \cdot \left(1 - \sum_{j=1}^{m} x_j\right)^{n-K_m}. \tag{A.4}$$

Financial Signal Processing and Machine Learning, First Edition.
Edited by Ali N. Akansu, Sanjeev R. Kulkarni and Dmitry Malioutov.

Consider the last sum,

$$S_m = \sum_{k_m=0}^{n-K_{m-1}} \binom{n-K_{m-1}}{k_m} (x_m z_m)^{k_m} \cdot \left(1 - \sum_{j=1}^{m} x_j\right)^{n-K_m}.$$

Since $K_m = K_{m-1} + k_m$, we obtain

$$S_m = \left(1 - \sum_{j=1}^{m-1} x_j - x_m(1-z_m)\right)^{n-K_{m-1}}.$$

Applying this transformation recursively to the sums over k_j, $(j = m-1, m-2, \dots, 1)$, in (A.4) we derive

$$\hat{\pi}(\vec{z}\,) = \sum_{n=0}^{\infty} p_n \cdot \left(1 - \sum_{j=1}^{m} x_j(1-z_j)\right)^n = \hat{p}\left(1 - \sum_{j=1}^{m} x_j(1-z_j)\right).$$

Lemma 9.2 is proved.

Remark A.1 *Notice that if* $\vec{z} = \vec{1}$, $\hat{\pi}(\vec{z}\,) = 1$. *The function* $\pi(\vec{k}) \geq 0$. *Therefore,* $\{\pi(\vec{k})\}_{\vec{k} \in \mathbb{Z}_+^m}$ *is a probability distribution on the m-dimensional integer lattice.*

A.1.2 Proof of Lemma 9.3

Derivation of Equation (9.20) is completely analogous to that of Equation (9.12) and, for this reason, is omitted. Let us derive (9.21).

From (9.20) we find, by differentiation $\left(\frac{\partial \hat{q}(z,w)}{\partial z} \mid_{z=w=1}, \frac{\partial \hat{q}(z,w)}{\partial w} \mid_{z=w=1}\right)$:

$$\mathbb{E}[\xi_1 \cdot \xi_2] = xy\mathbb{E}[\zeta_1\zeta_2], \quad \mathbb{E}[\xi_1] = x\mathbb{E}[\zeta_1], \quad \mathbb{E}[\xi_2] = y\mathbb{E}[\zeta_2],$$

and

$$\operatorname{cov}(\xi_1, \xi_2) = xy \cdot \operatorname{cov}(\zeta_1, \zeta_2).$$

Then we obtain

$$\sigma^2(\xi_1) = x^2 \cdot \sigma^2(\zeta_1) + \mathbb{E}[\zeta_1] \cdot (x - x^2),$$

and

$$\sigma^2(\xi_2) = y^2 \cdot \sigma^2(\zeta_2) + \mathbb{E}[\zeta_2] \cdot (y - y^2).$$

If the variance and the first moment of the random variables, ζ_i, $(i = 1, 2)$, are equal, then

$$\sigma(\xi_i) = \sqrt{x} \cdot \sigma(\zeta_i), \quad i = 1, 2.$$

Finally, we obtain

$$\rho(\xi_1, \xi_2) = \sqrt{xy} \cdot \rho(\zeta_1, \zeta_2),$$

and Lemma 9.3 is thus proved.

References

Aue, F. and Kalkbrener, M. (2006). LDA at work: Deutsche Banks approach to quantify operational risk. *Journal of Operational Risk*, 1(4), 49–95.

Avellaneda, M. and Zhu, J. (2001). Distance to default. *Risk*, 14(12), 125–129.

Badescu, A., Gang, L., Lin, S. and Tang, D. (2013). A mixture model approach to operational risk management. Working paper. Toronto: University of Toronto, Department of Statistics.

Badescu, A., Gang, L., Lin, S. and Tang, D. (2015). Modeling correlated frequencies with applications in operational risk management. *Journal of Operational Risk*, forthcoming.

Bae, T. (2012). A model for two-way dependent operational losses. Working paper. Regina, SK: University of Regina, Department of Mathematics and Statistics.

Barndorff-Nielsen, O. and Shephard, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society*, 63, 167–241.

Barndorff-Nielsen, O. and Yeo, G. (1969). Negative binomial processes. *Journal of Applied Probability*, 6(3), 633–647.

Böcker, K. and Klüppelberg, C. (2010). Multivariate models for operational risk. *Quantitative Finance*, 10(8), 855–869.

Boyarchenko, S. and Levendorski, S. (2002). *Non-Gaussian Merton-Black-Scholes theory*. River Edge, NJ: World Scientific.

Carr, P. and Madan, D. (1998). Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2, 61–73.

Chavez-Demoulin, V., Embrechts, P. and Nešlehovà, J. (2006). Quantitative models for operational risk: extremes, dependence and aggregation. *Journal of Banking and Finance*, 30(9), 2635–2658.

Cont, R. and Tankov, P. (2003). *Financial modeling with jump processes*. London: Chapman and Hall, CRC Press.

Cont, R. and Tankov, P. (2006). Retrieving Levy processes from option prices: regularization of an ill-posed inverse problem. *SIAM Journal on Control and Optimization*, 45, 1–25.

Duch, K., Kreinin, A. and Jiang, Y. (2014). New approaches to operational risk modeling. *IBM Journal of Research and Development*, 3, 31–45.

Embrechts, P. and Puccetti, G. (2006). Aggregating risk capital with an application to operational risk. *The Geneva Risk and Insurance Review*, 31(2), 71–90.

Feigin, P. (1979). On the characterization of point processes with the order statistic property. *Journal of Applied Probability*, 16(2), 297–304.

Fox, B. (1996). Generating Poisson processes by quasi-Monte Carlo. Working paper. Boulder, CO: SIM-OPT Consulting.

Fox, B. and Glynn, P. (1988). Computing Poisson probabilities. *Communication of the ACM*, 31, 440–445.

Frechet, M. (1951). Sur le tableaux de correlation dont les marges sont donnees. *Annales de l'Université de Lyon*, 14, 53–77.

Griffiths, R., Milne, R. and Wood, R. (1979). Aspects of correlation in bivariate Poisson distributions and processes. *Australian Journal of Statistics*, 21(3), 238–255.

Hardy, G., Littlewood, J. and Polya, G. (1952). *Inequalities*. Cambridge: Cambridge University Press.

Hoeffding, W. (1940). Masstabinvariante korrelations-theorie. *Schriften Math. Inst. Univ. Berlin*, 5, 181–233.

Hull, J. and White, A. (2001). Valuing credit default swaps II: modeling default correlations. *Journal of Derivatives*, 8(3), 29–40.

Iscoe, I. and Kreinin, A. (2006). Recursive valuation of basket default swaps. *Journal of Computational Finance*, 9(3), 95–116.

Iscoe, I. and Kreinin, A. (2007). Valuation of synthetic CDOs. *Journal of Banking and Finance*, 31, 3357–3376.

Iscoe, I., Kreinin, A. and Rosen, D. (1999). Integrated market and credit risk portfolio model. *Algorithmics Research Quarterly*, 2(3), 21–38.

Jaimungal, S., Kreinin, A. and Valov, A. (2013). The generalized Shiryaev problem and Skorokhod embedding. *Theory of Probabilities and Applications*, 58(3), 614–623.

Johnson, N. and Kotz, S. (1969). *Distributions in statistics: discrete distributions*. Boston: Houghton Mifflin.

Johnson, N., Kotz, S. and Balakrishnan, N. (1997). *Discrete multivariate distributions*. New York: John Wiley & Sons.

Kou, S. (2002). A jump diffusion model for option pricing. *Management Sciences*, 48(8), 1086–1101.

Kyprianou, A., Schoutens, W. and Wilmott, P. (2005). *Exotic option pricing and advanced Levy models*. Hoboken, NJ: John Wiley & Sons.

Lindskog, F. and McNeil, A.J. (2001). *Poisson shock models: applications to insurance and credit risk modeling*. Zurich: Federal Institute of Technology ETH Zentrum.

Lipton, A. and Rennie, A. (2008). *Credit correlation: life after copulas*. River Edge, NJ: World Scientific.

Merton, R. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29, 449–470.

Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3, 125–144.

Musiela, M. and Rutkowsky, M. (2008). *Martingale methods in financial modelling stochastic modelling*. New York: Springer-Verlag.

Nawrotzki, K. (1962). Ein grenwertsatz für homogene züffalige punktfolgen (verallgemeinerung eines satzes von A. Renyi). *Mathematische Nachrichten*, 24, 201–217.

Nelsen, R.B. (1987). Discrete bivariate distributions with givenmarginals and correlation. *Communications in Statistics –Simulation*, 16(1), 199–208.

Nešlehovà, J., Embrechts, P. and Chavez-Demoulin, V. (2006). Infinite mean models and the lda for operational risk. *Journal of Operational Risk*, 1(1), 3–25.

Panjer, H.H. (2006). *Operational risks: modeling analytics*. Hoboken, NJ: John Wiley & Sons.

Peters, G., Shevchenko, P. and Wüthrich, M. (2009). Dynamic operational risk: modeling dependence and combining different data sources of information. *The Journal of Operational Risk*, 4(2), 69–104.

Powojovsky, M., Reynolds, D. and Tuenter, H. (2002). Dependent events and operational risk. *Algorithmics Research Quarterly*, 5(2), 65–74.

Revus, D. and Yor, M. (1991). *Continuous Martingales and Brownian motion: a series of comprehensive studies in mathematics*, 2nd ed. New York: Springer.

Rolski, T., Schmidli, H., Schmidt, V. and Teugles, J. (1999). *Stochastic processes for insurance and finance*. Chichester, UK: John Wiley & Sons.

Shevchenko, P. (2011). *Modeling operational risk using modeling analytics*. New York: Springer.

Shreve, S. (2004). *Stochastic calculus for finance II: continuous-time models –a series of comprehensive studies in mathematics*. New York: Springer-Finance.

Vasicek, O. (1987). *Probability of loss on loan portfolio*. Working paper. San Francisco: KMW.

Vasicek, O. (2002). Loan portfolio value. *Risk*, 15, 160–162.

Whitt, W. (1976). Bivariate distributions with given marginals. *Annals of Statistics*, 4, 1280–1289.

Yahav, I. and Shmueli, G. (2011). On generating multivariate Poisson data in management science applications. *Applied Stochastic Models in Business and Industry*, 28(1), 91–102.