Jointly Poisson processes

Don H. Johnson and Ilan N. Goodman

Electrical & Computer Engineering Department, MS380

Rice University

Houston, Texas 77005–1892

{*dhj,igoodman*}@rice.edu

Abstract

What constitutes jointly Poisson processes remains an unresolved issue. This report reviews the current state of the theory and indicates how the accepted but unproven model equals that resulting from the small time-interval limit of jointly Bernoulli processes. One intriguing consequence of these models is that jointly Poisson processes can only be positively correlated as measured by the correlation coefficient defined by cumulants of the probability generating functional.

1 Introduction

To describe spike trains mathematically, particularly those that do not produce deterministic sequences of spikes, point process models are usually employed. From a mathematical viewpoint, the Poisson process is the simplest and therefore the model that has yielded the most results. Here, events occur randomly at a rate given by some function $\lambda(t)$ with no statistical dependence of one event's occurrence on the number and the timing of other events. Unfortunately, Poisson processes cannot accurately describe spike trains because of absolute and relative refractory effects. Here, the occurrence of a spike influences when the next one occurs. Some spike trains deviate even more from the Poisson model, with several spikes affecting subsequent ones in complicated ways. Modeling these falls under the realm of non-Poisson processes, which in many cases makes it very difficult to obtain analytic results. Consequently, the Poisson model is used to obtain predictions about the character of the spike train, like its information capacity, that are understood not to be precisely accurate for any realistic neural recording. In some cases, the Poisson process can be used to obtain bounds on performance that can be used as well-established guideposts for neural behavior.

When it comes to population models, in which several neurons presumably jointly encode information, we lack even a Poisson model for all but the simplest cases: the component point processes are either statistically independent or conditionally independent. Data show more complicated behavior since cross-correlation functions often show correlations among members of a population. Consequently, what is the generalization of the single Poisson process description to what could be termed the jointly Poisson model. Here, we seek to describe the joint statistics for several processes, each of which is Poisson (i.e., the marginal processes are Poisson).

2 Infinite Divisibility

From a probabilistic standpoint, specifying a unique joint probability distribution that has specified marginal distributions is ill-posed, since many joint distributions could conceivably work. The easiest way to show the ill-posed nature of this problem is to consider the situation for Gaussian

random variables. A set of random variables $\{X_1, \ldots, X_M\}$ is said to be jointly Gaussian if the joint probability density has the form

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{|2\pi\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{(\mathbf{x}-\mathbf{m})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})}{2} \right\}, \ \mathbf{X} = \{X_1, \ldots, X_M\}$$

Here, $\boldsymbol{\Sigma}$ is the covariance matrix, $|\cdot|$ represents the matrix determinant, $\mathbf{m}$ is the vector of means and $\mathbf{x}'$ represents the transpose of the vector $\mathbf{x}$. Each of the random variables has a Gaussian marginal probability distribution. One can also find a joint distribution *not* of this form that also has Gaussian marginals. For example, consider the two-dimensional case ($N = 2$) when the means are zero. Let the joint distribution be as written as above, but defined to be zero in the first and third quadrants. To obtain a valid joint distribution, we must multiply the above formula by two so that the total probability obtained by integration is one. This joint distribution yields marginal distributions no different from the jointly Gaussian case, but the random variables are *not* jointly Gaussian because the joint distribution does not have the form written above.

What makes the jointly Gaussian random vector special is the property of *infinite divisibility*: the random vector can be expressed as a sum of an arbitrary number of statistically independent random vectors (Daley and Vere-Jones, 1988). The probability distribution of the sum is the convolution of the individual probability distributions. Consequently, infinite divisibility demands that a probability distribution be expressed as the *n*-fold convolution of a density with itself. In special cases, like the Gaussian and the Poisson, each of the constituent random vectors has the same distributional form (i.e., they differ only in parameter values) as do their sum.

The characteristic function provides a more streamlined definition of what what infinite divisibility means. The characteristic function of a random vector $\mathbf{X}$ is defined to be

$$\Phi_{\mathbf{X}}(j\mathbf{u}) \triangleq \int p_{\mathbf{X}}(\mathbf{x}) e^{j\mathbf{u}'\mathbf{x}} d\mathbf{x}.$$

The characteristic function of a sum of statistically independent random vectors is the product of the individual characteristic functions.

$$\Phi_{\mathbf{Y}}(j\mathbf{u}) = \prod_{i=1}^{n} \Phi_{\mathbf{X}_i}(j\mathbf{u}) \quad \mathbf{Y} = \sum_{i=1}^{n} \mathbf{X}_i$$

Infinite divisibility demands that $\left[\Phi_{\mathbf{Y}}(j\mathbf{u})\right]^{1/n}$ also be a characteristic function for any positive integer value of *n*. If we express a characteristic function parametrically as $\Phi(j\mathbf{u}; \boldsymbol{\theta})$, with $\boldsymbol{\theta}$ denotes the probability distribution's parameters, the Gaussian case is special in that $\left[\Phi_{\mathbf{Y}}(j\mathbf{u}; \boldsymbol{\theta})\right]^{1/n} = \Phi_{\mathbf{Y}}(j\mathbf{u}; \boldsymbol{\theta}/n)$. For the jointly Gaussian case, these parameters are the mean and covariance matrix.

$$\Phi_{\mathbf{X}_i}(j\mathbf{u}; \mathbf{m}_i, \boldsymbol{\Sigma}_i) = \exp\left\{ j\mathbf{u}'\mathbf{m}_i - \mathbf{u}'\boldsymbol{\Sigma}_i\mathbf{u}/2 \right\}$$

Dividing these parameters by *n* does not affect the viability of the underlying Gaussian distribution,

which makes it an infinitely divisible random vector. The example given above of a bivariate distribution having Gaussian marginals is not infinitely divisible as its characteristic function does not have this property.

In the point process case, a single Poisson process is easily seen to be infinitely divisible since the superposition of Poisson processes is also Poisson. We must modify the just-presented mathematical formalism involving characteristic functions because we have a random process, not a random vector. The *probability-generating functional* is defined as

$$G[u(t)] \triangleq \mathsf{E}\left[\exp\left\{\int \log u(t)\, dN_t\right\}\right],$$

where the transform variable $u(t)$ is a real-valued function of time and N_t is the point process's counting function (the number of events that have occurred prior to time t). It has similar properties to the moment-generating function with one notable exception: it has no "inverse transform." However, the moment-generating function for the total number of counts in the interval implicit in the integral can be found from the probability generating function with the substitution $u(t) \to z$. Finding the probability distribution that underlies the expected value in the above formula requires a special series expansion. Interesting quantities, like moments can be found from the probability-generating functional by evaluating derivatives of its logarithm. For example, the formal derivative with respect to $u(\cdot)$ and evaluating the result at $u(\cdot) = 1$ yields the expected value.

$$\frac{d \log G[u(t)]}{du(t)} = \frac{1}{G[u(t)]}\,\mathsf{E}\left[\int \frac{1}{u(t)}\, dN_t \exp\left\{\int \log u(t)\, dN_t\right\}\right]\Bigg|_{u(t)=1} = \mathsf{E}\left[\int dN_t\right]$$

$$\frac{d \log G[u(t)]}{du(t_0)} = \frac{1}{G[u(t)]}\mathsf{E}\left[\frac{1}{u(t_0)} dN_{t_0} \exp\left\{\int \log u(t)\, dN_t\right\}\right]\Bigg|_{u(t)=1} = \mathsf{E}\left[dN_{t_0}\right]$$

The first of these is the total variation with respect to $u(t)$ and yields the expected number of events over the interval spanned by the integral. The second is the derivative at the time instant t_0, which yields the expected value of the process at that time instant.

Despite not being easily able to determine the probability distribution, showing infinite divisibility can be seen by inspection just as with characteristic functions. For a Poisson process, the probability-generating functional has the special form

$$G[u(t)] = \exp\left\{\int \big(u(t) - 1\big)\lambda(t)\, dt\right\}$$

To show infinite divisibility, we note that the only "parameter" of a Poisson process is its instantaneous rate function $\lambda(t)$. As the product of probability-generating functionals for Poisson processes yields the same form with the total rate equaling the sum of the component rates, the Poisson process is infinitely divisible.

What we seek here is a description of the joint probability distribution of several marginal Poisson processes so that the vector of Poisson processes is infinitely divisible. We exhibit here

what the probability generating functional for an infinitely divisible vector of Poisson processes must be and show how to use this quantity to derive some of its properties. In particular, we show that they can be constructed in a stereotypical way that elucidates the cross-correlation behavior required of jointly Poisson processes. Somewhat surprisingly, the range of correlation structures is quite limited, with values for the correlation parameters tightly intertwined with each other and with the dimensionality of the vector process. In particular, pairwise correlation coefficients cannot be negative for any pair and must decrease as the dimension increases.

3   Jointly Poisson Processes

The probability-generating functional for several point processes considered jointly has the simple form

$$G^{(M)}[\mathbf{u}(t)] \triangleq \mathsf{E}\left[\exp\left\{\sum_{m=1}^{M}\int \log u_m(t)\,dN_{m,t}\right\}\right] \tag{1}$$

where the expected value is computed with respect to the joint distribution of the point processes, which is the quantity we seek. The probability-generating functional of component process j can be found from this formula by setting $u_i(t) = 1$, $i \neq j$. If the processes are statistically independent, their joint probability functional equals the product of the marginal functionals. If the processes are added, the probability generating functional of the result equals the joint functional evaluated at a common argument: $G[u(t)] = G^{(M)}[u(t), u(t), \ldots, u(t)]$. These properties generalize those of moment generating functions. Furthermore, cross-*covariance* between two processes, i and j say, can be found by evaluating the second mixed partial of the log joint probability-generating functional:

$$\left.\frac{\partial^2 \log G^{(M)}[\mathbf{u}(t)]}{\partial u_i(t)\partial u_j(t)}\right|_{\mathbf{u}(t)=\mathbf{1}} = \mathsf{E}\left[\int dN_{i,t}\int dN_{j,t}\right] - \mathsf{E}\left[\int dN_{i,t}\right]\cdot\mathsf{E}\left[\int dN_{j,t}\right]$$

$$\left.\frac{\partial^2 \log G^{(M)}[\mathbf{u}(t)]}{\partial u_i(t_i)\partial u_j(t_j)}\right|_{\mathbf{u}(t)=\mathbf{1}} = \mathsf{E}\left[dN_{i,t_i}dN_{j,t_j}\right] - \mathsf{E}\left[dN_{i,t_i}\right]\cdot\mathsf{E}\left[dN_{j,t_j}\right]$$

Again, the first expression gives the cross-covariance of counts while the second gives the cross-covariance between the processes i, j at the times t_i, t_j.

Over thirty years ago, the probability-generating functional of two marginally Poisson processes that satisfied the infinite-divisibility condition was shown to have the unique form (Milne, 1974)

$$G^{(2)}[u_1(t), u_2(t)] = \exp\left\{\int \big(u_1(t)-1\big)v_1(t)\,dt + \int \big(u_2(t)-1\big)v_2(t)\,dt\right.$$
$$\left. + \iint \big(u_1(s)u_2(t)-1\big)v_c(\alpha,\beta)\,d\alpha\,d\beta\right\}. \tag{2}$$

This joint probability-generating functional is easily interpreted. First of all, by setting $u_2(t) = 1$, we obtain the marginal probability-generating functional of process 1, showing that it is a Poisson process having an instantaneous rate of $v_1(t) + \int v_c(t,\beta)\,d\beta$. Similarly, process 2 is also Poisson with a rate equal to $v_2(t) + \int v_c(\alpha,t)\,d\alpha$. Also, setting $v_c(s,t) = 0$ results in the product of the marginal

probability-generating functionals, corresponding to the case in which the processes are statistically independent. Thus, the "common rate" $v_c(\alpha, \beta)$ represents a joint rate variation that induces statistical dependence between the processes. The simplest example is $v_c(\alpha, \beta) = v_c(\beta)\delta(\alpha - \beta)$, indicating an instantaneous correlation at each moment in time. The resulting dependence term in the probability generating functional equals

$$\int\int \big(u_1(\alpha)u_2(\beta) - 1\big)v_c(\alpha, \beta)\,d\alpha\,d\beta = \int \big(u_1(t)u_2(t) - 1\big)v_c(t)\,dt\ .$$

Statistically dependent Poisson processes having an infinitely divisible joint probability distribution can be simply constructed by adding to statistically independent Poisson processes having rates $v_1(t)$ and $v_2(t)$ — what we call the *building-block* processes — a common Poisson process having rate $v_c(t)$ that is statistically independent of the others. This way of constructing jointly Poisson processes amounts to the construction described by Holgate (Holgate, 1964). An allowed variant is to delay the common process when it is added to one but not the other building-block process. Here, $v_c(\alpha, \beta) = v_c(\beta)\delta\big(\alpha - (\beta - t_0)\big)$. In this way, correlation can occur at a time lag other than zero, but still only at a single point.

More generally, $v_c(s, t)$ depends on its arguments in different ways that do not lead to a simple superposition of building-block Poisson processes. Using the probability generating function, you can show that the cross-covariance function between the two constructed processes equals the common rate: $\mathsf{cov}\big[dN_{1,t_1}, dN_{2,t_2}\big] = v_c(t_1, t_2)$. One would think that many common cross-covariances could be described this way. However, several important constraints arise.

- Cross-covariances must be non-negative. This condition arises because the common rate must be non-negative so that a valid probability generating functional results.

- For the constructed processes to be jointly (wide-sense) stationary, we must have constant rates and a cross-covariance function that depends only on the time difference. Here, the latter constraint means $v_c(s, t) = f(|s - t|)$. Milne and Westcott (Milne and Westcott, 1972) give more general conditions for the common rate function to be well-defined. Thus, correlation can extend continuously over some time lag domain. Consequently, the Holgate construction does not yield all possible jointly Poisson processes.

- It is not clear that the joint-rate characterization extends in its full generality to more than pairs of Poisson processes (Milne and Westcott, 1993) because the putative probability generating functional for the marginal process has not been shown to correspond to a Poisson's probability generating functional. However, the special case of the Holgate construction technique always works.

In sequel, we only consider jointly Poisson processes that can be constructed in Holgate's fashion as a superposition of building-block Poisson processes.

Calculating means and covariances from the probability generating functional for jointly Pois-

son processes is very revealing.

$$\text{Counts:} \qquad \mathsf{E}\left[\int dN_{i,t}\right] = \int \left(v_i(t) + v_c(t)\right) dt$$

$$\mathsf{cov}\left[\int dN_{1,t}, \int dN_{2,t}\right] = \int v_c(t)\, dt$$

$$\text{Instantaneous:} \qquad \mathsf{E}\left[dN_{i,t}\right] = v_i(t) + v_c(t)$$

$$\mathsf{cov}\left[dN_{1,t_1}, dN_{2,t_2}\right] = \begin{cases} 0, & t_1 \neq t_2 \\ v_c(t), & t_1 = t = t_2 \end{cases}$$

Since the variance of a Poisson process equals its mean, we find that the second-order correlation coefficient $\rho^{(2)}(t)$ equals

$$\text{Counts:} \qquad \rho^{(2)}(t) = \frac{\int v_c(t)\, dt}{\sqrt{\int \left(v_1(t) + v_c(t)\right) dt \cdot \int \left(v_2(t) + v_c(t)\right) dt}}$$

$$\text{Instantaneous:} \qquad \rho^{(2)}(t) = \begin{cases} 0, & t_1 \neq t_2 \\ \frac{v_c(t)}{\sqrt{(v_1(t) + v_c(t))(v_2(t) + v_c(t))}}, & t_1 = t = t_2 \end{cases}$$

Thus, the correlation coefficient between both the counts and the instantaneous values lies in the interval $[0, 1]$, with the maximal correlation occurring in the limit of large values for the common rate. However, note that correlation has *no* temporal extent and for some particular lag: given an event occurs in one process, it is correlated with the other process at the first process's event time and uncorrelated (statistically independent) at all others.

We can write the probability-generating functional in terms of the rates of the building-block processes, $v_i(t)$ and $v_c(t)$, or in terms of the rates of the constructed processes $\lambda_i(t) = v_i(t) + v_c(t)$ and the correlation coefficient $\rho^{(2)}(t)$ given above.

$$G^{(2)}[u_1(t), u_2(t)] = \exp\left\{\int \left(u_1(t) - 1\right) v_1(t)\, dt + \int \left(u_2(t) - 1\right) v_2(t)\, dt \right.$$

$$\left. + \int \left(u_1(t) u_2(t) - 1\right) v_c(t)\, dt \right\}$$

$$G^{(2)}[u_1(t), u_2(t)] = \exp\left\{\int \left(u_1(t) - 1\right) \lambda_1(t)\, dt + \int \left(u_2(t) - 1\right) \lambda_2(t)\, dt \right. \qquad (3)$$

$$\left. + \int \left(u_1(t) - 1\right) \left(u_2(t) - 1\right) \rho^{(2)}(t) \sqrt{\lambda_1(t) \lambda_2(t)}\, dt \right\}$$

We can extend this type of analysis to three Poisson processes constructed from six building-

block processes according to the following formulas for their rates.

$$\lambda_1(t) = \nu_1(t) + \nu_4(t) + \nu_5(t)$$
$$\lambda_2(t) = \nu_2(t) + \nu_4(t) + \nu_6(t)$$
$$\lambda_3(t) = \nu_3(t) + \nu_5(t) + \nu_6(t)$$

This generates pairwise-dependent processes with no third-order dependencies. The covariance between any pair is expressed by the building-block process rate they share in common. Consequently,

$$\rho_{1,2}^{(2)} = \frac{\nu_4(t)}{\sqrt{\lambda_1(t)\lambda_2(t)}} \ .$$

By letting $\nu_1 = \nu_2 \equiv \nu^{(1)}$ and $\nu_4 = \nu_5 = \nu_6 \equiv \nu^{(2)}$, we create what we term the *symmetric* case, in which we have only two separately adjustable rates that arise from the six statistically independent building-block processes. In this case, this cross-correlation simplifies to

$$\rho^{(2)} = \frac{\nu^{(2)}(t)}{\nu^{(1)}(t) + 2\nu^{(2)}(t)} \leq \frac{1}{2}, \ i \neq j \tag{4}$$

When a Poisson process having instantaneous rate $\nu^{(3)}(t)$ is added to all three building-block processes to create third-order dependence, the correlation coefficient becomes in the symmetric case

$$\rho^{(2)} = \frac{\nu^{(2)}(t) + \nu^{(3)}(t)}{\nu^{(1)}(t) + 2\nu^{(2)}(t) + \nu^{(3)}(t)}, \ i \neq j$$

Now, as the common process's rate grows, the pairwise correlation coefficient can approach one. If we define a third-order correlation coefficient according to

$$\rho^{(3)}[dN_{1,t}, dN_{2,t}, dN_{3,t}] \triangleq \frac{\left. \dfrac{\partial^3 \log G[u_1(t), u_2(t), u_3(t)]}{\partial u_1(t) \partial u_2(t) \partial u_3(t)} \right|_{\mathbf{u}=\mathbf{1}}}{\sqrt[3]{\mathrm{var}[dN_{1,t}]\mathrm{var}[dN_{2,t}]\mathrm{var}[dN_{3,t}]}} \ . \tag{5}$$

For the symmetric Poisson example, the third-order correlation coefficient is easily found to be

$$\rho^{(3)}(t) = \frac{\nu^{(3)}(t)}{\nu^{(1)}(t) + 2\nu^{(2)}(t) + \nu^{(3)}(t)}$$

Combining with the expression for the second-order correlation coefficient, we find the following bounds for the symmetric case relating the correlation quantities.

$$0 \leq \rho^{(3)} \leq 2\rho^{(2)} - \rho^{(3)} \leq 1$$

Note that this inequality chain indicates that $0 \leq \rho^{(3)} \leq \rho^{(2)} \leq 1$. The second-order correlation can be bigger than $\frac{1}{2}$, but only if $\rho^{(3)}$ increases as well in a manner defined by the inequality chain.

We need to extend this analysis to an arbitrary number of building block and constructed pro-cesses. We can form an arbitrary number of infinitely divisible, jointly defined Poisson processes by extending the two- and three-process Holgate construction technique. Given L statistically inde-pendent Poisson processes, we create a population of M statistically dependent Poisson processes according by superimposing L building-block processes according to the *construction matrix* $\mathbf{A}$: $\mathbf{N}_t = \mathbf{AB}_t$. Here, $\mathbf{N}_t$ and $\mathbf{B}_t$ represent column vectors of constructed and building-block Poisson processes of dimension M and $L > M$ respectively. The entries of the construction matrix are either 0 or 1. For example, the construction matrix underlying the two- and three-process examples are

$$M = 2: \mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$M = 3: \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

To introduce dependencies of all orders, $L \geq 2^M - 1$, and we concentrate on the case $L = 2^M - 1$ in sequel.

The probability generating functional $G^{(M)}[\mathbf{u}(t)]$ of $\mathbf{N}_t$ expressed in (1) can be written in matrix form as

$$G^{(M)}[\mathbf{u}(t)] = \mathsf{E}\left[\exp\left\{\int \log \mathbf{u}'(t)\, d\mathbf{N}_t\right\}\right]$$

where the logarithm of a vector is defined in the MATLAB sense (an element-by-element operation). Because $\mathbf{N}_t = \mathbf{AB}_t$, we have

$$G^{(M)}[\mathbf{u}(t)] = \mathsf{E}\left[\exp\left\{\int \log \mathbf{u}'(t)\mathbf{A}\, d\mathbf{B}_t\right\}\right]$$
$$= \mathsf{E}\left[\exp\left\{\int \left(\mathbf{A}'\log \mathbf{u}(t)\right)'\, d\mathbf{B}_t\right\}\right]$$

Each component of the vector $\mathbf{A}'\log \mathbf{u}(t)$ expresses which combination of components of $\mathbf{u}(t)$ are associated with each building block process. This combination corresponds to the constructed processes to which each building block process contributes. Since the building block processes are statistically independent and Poisson, we have

$$G^{(M)}[\mathbf{u}(t)] = \int \left[\exp\left\{\mathbf{A}'\log \mathbf{u}(t)\right\} - 1\right]'\mathbf{v}(t)\, dt$$

Expanding the vector notation for a moment, this result can also be written as

$$G^{(M)}[\mathbf{u}(t)] = \exp\left\{\sum_{l=1}^{L}\int\left(\left[\prod_{m=1}^{M} u_m^{A_{m,l}}(t)\right] - 1\right)v_l(t)\, dt\right\} \tag{6}$$

Here, $u_m^{A_{m,l}}(t)$ means $u_m(t)$ raised to the $A_{m,l}$ power. In other words, if $A_{m,l} = 1$, the term is included;

if $A_{m,l} = 0$ it is not. Thus, the probability generating functional consists of a sum of terms, one for each building block process, wherein the coefficient of each rate $v_l(t)$ is the product of arguments corresponding to those constructed process building block process l helped to build minus one. This form is what equation (2) describes.

However, we need to convert this result into the form of (3) so that the role of the cumulant correlation coefficients can come to light. We can view the cumulant moments, the mixed first partials of the logarithm of the probability generating functional, as coefficients of the multivariate Taylor series for $\log G^{(M)}[\mathbf{u}(t)]$ centered at the point $\mathbf{u}(t) = \mathbf{1}$. Because the mth term in (6) contains only multilinear combinations of u_m, second-order and higher derivatives of these terms are zero. Consequently, the Taylor series for $\log G^{(M)}[\mathbf{u}(t)]$ consists *only* of multilinear terms having $(u_m - 1)$ as its constituents with the cumulant moments as the series coefficients. Consequently, the jointly Poisson process can always be written in a form generalizing (3). This coefficient equals

$$\left. \frac{\partial^k \log G^{(M)}[\mathbf{u}(t)]}{\partial u_{m_1}(t) \dots \partial u_{m_k}(t)} \right|_{\mathbf{u}(t)=\mathbf{1}} = \sum_{l=1}^{L} \left(\prod_{m=m_1,\dots,m_k} A_{m,l} \right) v_l(t) \tag{7}$$

Because matrix $\mathbf{A}$ has only binary-valued entries, the product $\prod_m A_{m,l}$ equals either one or zero, bringing in the lth building block process only if it contributes to all of the constructed processes indexed by $m_1, \dots, m_k$. Note that the first partial derivative expresses the rate of each constructed process: $\lambda_m(t) = \sum_l A_{m,l} v_l(t)$.

We can normalize the Taylor series coefficient to obtain cumulant correlation coefficients by dividing by the geometric mean of the constructed process rates that enter into the partial derivative shown in (7).

$$\rho_{m_1,\dots,m_k}^{(k)}(t) \triangleq \frac{\left. \frac{\partial^k \log G^{(M)}[\mathbf{u}(t)]}{\partial u_{m_1}(t)\dots\partial u_{m_k}(t)} \right|_{\mathbf{u}(t)=\mathbf{1}}}{[\lambda_{m_1}(t) \cdots \lambda_{m_k}(t)]^{1/k}}$$
$$= \frac{\sum_{l=1}^{L} \left(\prod_{m=m_1,\dots,m_k} A_{m,l} \right) v_l(t)}{[\sum_l A_{m_1,l} v_l(t) \cdots \sum_l A_{m_k,l} v_l(t)]^{1/k}}$$

Because the numerator expresses which building block processes are in common with all the specified constructed processes, they and others are contained in each term in the denominator. This property means that each cumulant correlation coefficient is less than one and, since rates cannot be negative, greater than or equal to zero. Similar manipulations show that $\rho_{m_1,\dots,m_k}^{(k)}(t) \geq \rho_{m_1,\dots,m_k,m_{k+1}}^{(k+1)}(t)$: the size of the cumulant correlation coefficients cannot increase with order.

In the symmetric case, the expression for the cumulant correlation coefficients simplifies greatly.

$$\rho^{(k)}(t) = \frac{\sum_{l=k}^{M} \binom{M-k}{l-k} v^{(l)}(t)}{\sum_{l=1}^{M} \binom{M-1}{l-1} v^{(l)}(t)} \tag{8}$$

The denominator is the rate $\lambda(t)$ of each constructed process and the numerator is the sum of the rates of the processes that induce the dependence of the specified order. This result makes it easier

to see that the cumulant correlation coefficients cannot increase in value with increasing order: $0 \le \rho^{(k)}(t) \le \rho^{(k-1)}(t) \le 1$, $k = 3, \ldots, M$. Furthermore, more stringent requirements can be derived by exploiting the structure equation (8), showing that the cumulant correlation coefficients must obey the following two relationships in the symmetric case.

$$
\begin{aligned}
\sum_{k=2}^{M} \rho^{(k)} (-1)^k \binom{M-1}{k-1} &\le 1 \\
\sum_{k=m}^{M} \rho^{(k)} (-1)^{k+m} \binom{M-m}{k-m} &\ge 0, \quad m = 2, \ldots, M
\end{aligned}
\tag{9}
$$

For example, for four jointly Poisson processes, the cumulant correlation coefficients must satisfy the inequalities

$$
\begin{aligned}
3\rho^{(2)} - 3\rho^{(3)} + \rho^{(4)} &\le 1 \\
\rho^{(2)} - 2\rho^{(3)} + \rho^{(4)} &\ge 0
\end{aligned}
$$

4   Relations to Jointly Bernoulli Processes

Interestingly, this form of the jointly Poisson process can be derived as the limit of the jointly Bernoulli process when the event probability becomes arbitrarily small. First of all, a single Poisson process is defined this way, with the event probability equal to $\lambda(t)\Delta t$. To extend this approach to two jointly Poisson processes, we use the Sarmanov-Lancaster model for two jointly Bernoulli processes (Goodman, 2004). Letting X_1, X_2 be Bernoulli random variables with event probabilities p_1, p_2 respectively, the joint probability distribution is given by

$$
P(X_1, X_2) = P(X_1)P(X_2)\left[1 + \rho \frac{(X_1 - p_1)(X_2 - p_2)}{\sigma_1 \sigma_2}\right]
$$

where the standard deviation σ_i of each random variable equals $\sqrt{p_i(1-p_i)}$. The key to the derivation is to use the moment generating function, defined to be the two-dimensional z-transform of this joint distribution.

$$
\Phi(z_1, z_2) = \sum_{x_1} \sum_{x_2} P(x_1, x_2) z_1^{x_1} z_2^{x_2}
$$

Simple calculations show that for the jointly Bernoulli distribution given above, its moment generating function is

$$
\begin{aligned}
\Phi(z_1, z_2) &= [(1-p_1)(1-p_2) + \rho\sigma_1\sigma_2] + [p_1(1-p_2) - \rho\sigma_1\sigma_2]z_1 + [p_2(1-p_1) - \rho\sigma_1\sigma_2]z_2 \\
&\quad + [p_1 p_2 + \rho\sigma_1\sigma_2]z_1 z_2 \\
&= \big(1 + p_1(z_1 - 1)\big)\big(1 + p_2(z_2 - 1)\big) + (z_1 - 1)(z_2 - 1)\rho\sigma_1\sigma_2
\end{aligned}
$$

Letting event probabilities be proportional to the binwidth Δt, we evaluate this expression to first order in the event probabilities. Especially note that $\sigma_1\sigma_2 \approx \sqrt{\lambda_1\lambda_2}\Delta t$ as $\Delta t \to 0$ to first order.

Therefore, we have

$$\Phi(z_1, z_2) \xrightarrow{\Delta t \to 0} [1 + (z_1 - 1)\lambda_1 \Delta t][1 + (z_2 - 1)\lambda_2 \Delta t] + (z_1 - 1)(z_2 - 1)\rho \sqrt{\lambda_1 \lambda_2} \Delta t$$
$$= 1 + \lambda_1 \Delta t(z_1 - 1) + \lambda_2 \Delta t(z_2 - 1) + \rho \sqrt{\lambda_1 \lambda_2} \Delta t(z_1 - 1)(z_2 - 1)$$

Evaluating the natural logarithm and using the approximation $\log(1 + x) \approx x$ for small x, we find that

$$\log \Phi(z_1, z_2) \approx (z_1 - 1)\lambda_1 \Delta t + (z_2 - 1)\lambda_2 \Delta t + (z_1 - 1)(z_2 - 1)\rho \sqrt{\lambda_1 \lambda_2} \Delta t$$

If we sum the Bernoulli random variables in each process over a fixed time interval, say $[0, T]$, we obtain the number of events that occur in each process. The moment generating function of this sum is the product of the individual joint moment generating functions, which means its logarithm equals the sum of the logarithms of the individual functions. Since the number of random variables increases as the binwidth decreases (equal to $T/\Delta t$) and noting these terms are proportional to Δt, the sum becomes an integral to yield

$$\log \Phi(N_1, N_2) = (z_1 - 1)\int_0^T \lambda_1(t)\,dt + (z_2 - 1)\int_0^T \lambda_2(t)\,dt + (z_1 - 1)(z_2 - 1)\int_0^T \rho(t)\sqrt{\lambda_1(t)\lambda_2(t)}\,dt$$

If we let $\lambda_i(t) = \nu_i(t) + \nu_c(t)$ and substitute (4) for the definition of the correlation coefficient, we obtain the logarithm of the probability generating functional for two jointly Poisson processes constructed using Holgate's method in which $u_i(t) \to z_i$ as in equation (3).

Generalizing this result is tedious but straightforward: jointly Bernoulli processes converge in the limit of small event probabilities to jointly Poisson processes interdependent on each other at the same moment. An interesting sidelight is the normalization of the higher order dependency terms in the Sarmanov-Lancaster expansion demanded to make the correlation coefficient in the two models agree. In the Sarmanov-Lancaster expansion, the kth order term has the form exemplified by

$$\rho^{(k)} \frac{(X_1 - p_1)\cdots(X_k - p_k)}{C_k}$$

where C_k is the normalization constant that depends on correlation order and the specific choice of random variables in the term. Normally, Sarmanov-Lancaster expansions consist of products of orthonormal functions, which in this case would be $\prod(X_i - p_i)/\sigma_i$. This makes the putative normalization constant equal to $C_k = \prod \sigma_i$. However, the higher order correlation coefficients consequent of this definition have no guaranteed domains as does $\rho^{(2)}$. As described above, the jointly Poisson correlation coefficients defined via cumulants do have an orderliness. Associating the two demands that correlation coefficient be defined as

$$\rho^{(k)} \triangleq \frac{\mathsf{E}\big[(X_1 - p_1)\cdots(X_k - p_k)\big]}{\left(\prod_{i=1}^k \sigma_i^2\right)^{1/k}}$$

The normalization $\left(\prod_{i=1}^k \sigma_i^2\right)^{1/k}$ corresponds to the geometric mean of the variances found in the

definition (5) of correlation coefficients for Poisson processes. In the context of the Sarmanov-Lancaster expansion, we have

$$\rho^{(k)} = \rho^{(k)} \frac{\sigma_1^2 \cdots \sigma_k^2}{C_k \cdot \left(\prod \sigma_i^2\right)^{1/k}} \ .$$

Solving for C_k, we find that

$$C_k = \left(\sigma_1^2 \cdots \sigma_k^2\right)^{\frac{k-1}{k}} \ .$$

Using this normalization in the Sarmanov-Lancaster expansion now creates a direct relationship between its parameters and those of the jointly Poisson probability distribution. The inequality sets shown in (9) also guarantee existence of the Sarmanov-Lancaster model (Bahadur, 1961). This change does not affect the orthogonality so crucial in defining the Sarmanov-Lancaster expansion, only the normality.

Because of the correspondence between jointly Bernoulli processes and jointly Poisson processes, we can use the limit of the Sarmanov-Lancaster expansion to represent the joint distribution of jointly Poisson processes. In particular, we can evaluate information-theoretic quantities related to Poisson processes using this correspondence. Since entropy and mutual information are smooth quantities (infinitely differentiable), the small-probability limit can be evaluated *after* they are computed for Bernoulli processes.

References

R. R. Bahadur. A representation of the joint distribution of responses to *n* dichotomous items. In H. Solomon, editor, *Studies in Item Analysis and Prediction*, pages 158–168. Stanford University Press, California, 1961.

H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, 1946.

D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer-Verlag, New York, 1988.

I.N. Goodman. Analyzing statistical dependencies in neural populations. Master's thesis, Dept. Electrical & Computer Engineering, Rice University, Houston, Texas, 2004.

P. Holgate. Estimation for the bivariate Poisson distribution. *Biometrika*, 51:241–245, 1964.

R.K. Milne. Infinitely divisible bivariate Poisson processes. *Adv. Applied Prob.*, 6:226–227, 1974.

R.K. Milne and M. Westcott. Further results for Gauss-Poisson processes. *Adv. Applied Prob.*, 4:151–176, 1972.

R.K. Milne and M. Westcott. Generalized multivariate Hermite distributions and related point processes. *Ann. Inst. Statist. Math.*, 45:367–381, 1993.