A Multivariate Point Process Model for Neural Spike Trains

Reza Ramezan (rramezan@uwaterloo.ca)
Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. W.
Waterloo, ON N2L 3G1 Canada

Meixi Chen (meixi.chen@uwaterloo.ca)
Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. W.
Waterloo, ON N2L 3G1 Canada

Martin Lysy (mlysy@uwaterloo.ca)
Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. W.
Waterloo, ON N2L 3G1 Canada

Paul Marriott (pmarriott@uwaterloo.ca)
Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. W.
Waterloo, ON N2L 3G1 Canada

Abstract

The current state-of-the-art in neurophysiological data collection allows for simultaneous recording of tens to hundreds of neurons, for which point processes are an appropriate statistical modelling framework. However, existing point process models lack multivariate generalizations which are both flexible and computationally tractable. This paper introduces a multivariate generalization of the Skellam process with resetting (SPR), a point process tailored to model individual neural spike trains. The multivariate SPR (MSPR) is biologically justified as it mimics the process of neural integration. Its flexible dependence structure and a fast parameter estimation method make it well-suited for the analysis of simultaneously recorded spike trains from multiple neurons. The strengths and weaknesses of the MSPR are demonstrated through simulation and analysis of experimental data.

Keywords: multivariate point processes; Skellam process with resetting; records; spike trains; neurophysiology data

Introduction

The statistical modelling of neural spike trains falls under the general framework of point processes. Among point processes, the (inhomogeneous) Poisson process is a popular choice (Brown, Frank, Tang, Quirk, & Wilson, 1998; Ramezan, Marriott, & Chenouri, 2014), but its theoretical properties fail to address biological phenomena such as refractoriness and burst spiking activities (Ramezan, Marriott, & Chenouri, 2016). Furthermore, the Poisson process has very limited multivariate generalizations in terms of dependence structure (Marshall & Olkin, 1985; Kocherlakota & Kocherlakota, 1992; Johnson, Kotz, & Balakrishnan, 1997; Lakshminarayana, Pandit, & Rao, 1999; Karlis, 2003; Karlis & Meligkotsidou, 2005). Some alternative models include Integrate-and-Fire models (Gerstein & Mandelbrot, 1964; Paninski, Brown, Iyengar, & Kass, 2008), GLMs (Kass & Ventura, 2001; Dimatteo, Genovese, & Kass, 2001; Truccolo, Eden, Fellows, Donoghue,

& Brown, 2005; Wójcik, Mochol, Jakuczun, Wypych, & Waleszczyk, 2009), renewal processes (Koyama, 2015), and latent variable models (Shahbaba et al., 2014; Zhou, Moorman, Behseta, Ombao, & Shahbaba, 2016).

The Skellam process with resetting (SPR) is a point process recently developed to model spike trains collected from an individual neuron (Ramezan et al., 2016). A Skellam process $\{S(t), t \geq 0\}$ is defined as the difference between two independent Poisson processes, i.e. $S(t) = N_1(t) - N_2(t)$, where $N_i(t) \sim Poisson(\lambda_i t)$, $i = 1, 2$. Let $T_n = \min\{t : S(t) \geq n\}$. The SPR processs is then defined as $\tilde{S}(t) = \{S(t) - \sum_{j=1}^{\infty} \delta(t \geq T_j), \ t \geq 0\}$, where $\delta(\cdot)$ is the indicator function. Thus, resetting refers to bringing the path of the Skellam process back to state 0 after each spike, i.e. after each record of the process, to mimic the refractoriness of neurons (Ramezan et al., 2016). In this paper we generalize the results of (Ramezan et al., 2016) and introduce the multivariate Skellam process with resetting (MSPR) for the analysis of simultaneously recorded spike trains from multiple neurons.

Multivariate Skellam Process with Resetting

We first introduce the multivariate Skellam random vector, following a similar construction to that the multivariate Poisson random vector (Karlis, 2003; Karlis & Meligkotsidou, 2005).

Definition 1: Let $\mathbf{Z}$ be a p-variate random vector where

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_i \\ \vdots \\ Z_p \end{pmatrix} = \begin{pmatrix} Y_1 & + & \sum_{j=2}^{p} a_{1j} Y_{1j} \\ Y_2 & + & a_{21} Y_{12} & + & \sum_{j=3}^{p} a_{2j} Y_{2j} \\ \vdots \\ Y_i & + & \sum_{j=1}^{i-1} a_{ij} Y_{ji} & + & \sum_{j=i+1}^{p} a_{ij} Y_{ij} \\ \vdots \\ Y_p & & & + & \sum_{j=1}^{p-1} a_{pj} Y_{jp} \end{pmatrix},$$

where $a_{ij} \in \{-1, 0, 1\}$, $Y_i \sim Skellam(\lambda_{i1}, \lambda_{i2})$, and $Y_{ij} \sim Skellam(\gamma_{ij}, \gamma_{ij})$. All Skellam random variables Y_i and Y_{ij}, $\forall i, j$ are independent of one another, and also $a_{ij} = 0 \implies a_{ji} = 0$. The random vector $\mathbf{Z} \sim MSk(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\gamma})$ is called a *multivariate*

Skellam random vector with parameters $\boldsymbol{\lambda}_1 = (\lambda_{11}, ..., \lambda_{p1})$, $\boldsymbol{\lambda}_2 = (\lambda_{12}, ..., \lambda_{p2})$, and $\boldsymbol{\gamma} = (\gamma_{12}, ..., \gamma_{(p-1)p})$.

Definition 2: Let $\mathbf{S}(t) = \{S^{(i)}(t), t \geq 0, i = 1, ..., p\}$ be a p-dimensional Skellam processes, i.e., using Definition 1 with Skellam processes instead of random variables. Then for $T_n^{(i)} = \min\{t : S^{(i)}(t) \geq n\}$, the *multivariate Skellam process with resetting (MSPR)* is defined as

$$\tilde{\mathbf{S}}(t) = \left\{ S_0^{(i)}(t) - \sum_{j=1}^{\infty} \delta(t \geq T_j^{(i)}), \, t \geq 0, \, i = 1, ..., p \right\}.$$

The value of a p-variate Skellam process at time $t = t_0$ is a multivariate Skellam random vector, $\mathbf{S}(t_0) \sim MSk(\boldsymbol{\lambda}_1 t_0, \boldsymbol{\lambda}_2 t_0, \boldsymbol{\gamma} t_0)$. Within this process, the spike times of neuron i are modelled as the record times of the ith SPR within the multivariate Skellam process. These individual processes are potentially dependent to allow modelling the relationships between cells in a neuronal ensemble.

To the best of our knowledge, the MSPR is the first multivariate point process model for neural spike trains which allows for the full range of correlation coefficient. Indeed, neurons i and j are positively correlated if $sign(a_{ij}) = sign(a_{ji})$, negatively correlated if $sign(a_{ij}) = -sign(a_{ji})$, and uncorrelated if $a_{ij} = a_{ji} = 0$.

Parameter Estimation

We assume that the spike times from multiple independent trials are available for each of the simultaneously recorded neurons. The duration of each trial is assumed to be "long" relative to the average inter-spike interval so that asymptotic results apply.

Following Definitions 1 and 2, and the parameterization of (Ramezan et al., 2016), the two parameters of the marginal SPR for neuron i are $\lambda_1^{(i)} = \lambda_{i1} + \sum_j \gamma_{ij}$, $\lambda_2^{(i)} = \lambda_{i2} + \sum_j \gamma_{ij}$, for $i = 1, .., p$. We use the method of moments to estimate γ_{ij} using spike counts in each trial, i.e. $\widehat{\gamma}_{ij} = |0.5\widehat{\sigma}_{ij}|$ in which $\widehat{\sigma}_{ij}$ is the estimate of the covariance between spike counts of the two neurons i and j. We then plug in the moment estimate $\widehat{\gamma}_{ij}$ in the likelihood function of the marginal SPR processes and compute maximum profile likelihood estimates of parameters $\lambda_1^{(i)}$ and $\lambda_2^{(i)}$ for $i = 1, ..., p$ using constrained optimization. The standard error of estimates are obtained by boostrapping over the independent trials.

Data Analysis

We assessed the MSPR model using a 3-neuron simulation study as well as a 5-neuron experimental dataset from the prefrontal cortex of a mouse (3 in OFC and 2 in PL) in a classical conditioning study where the observation window is 10 seconds after a reward delivery (Kaminska, 2020).

Table 1 shows the results for 50 spike trains from a 3-neuron simulation study. Table 2 and Figure 1 present the results for the 5-neuron experiment. To put the parameter estimates into context, Table 2 displays the means and variances of the inter-spike intervals (ISI) as well as those estimated by the MSPR

Table 1: Parameter estimation for 50 repeated trials from an ensemble of three simulated neurons

	λ_{11}	λ_{12}	λ_{21}	λ_{22}	λ_{31}	λ_{32}	γ_{12}	γ_{13}	γ_{23}
True	15	10	20	15	10	7	5	15	10
Est.	17.8	11.6	19.7	12.9	15.2	10.9	5.0	11.5	8.3
SE	7.4	7.2	6.9	7.0	7.8	7.6	3.2	5.2	5.7

Table 2: Moment estimates of the ISIs from 13 repeated trials from an ensemble of five prefrontal cortex neurons

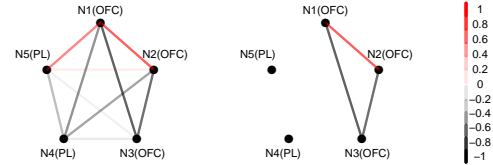|		Mean	Variance
Neuron# 1 (OFC)	Observed ISI (SE)	0.075 (0.035)	0.007 (0.008)
	Model ISI (SE)	0.121 (0.069)	0.015 (0.025)
Neuron# 2 (OFC)	Observed ISI (SE)	0.059 (0.011)	0.003 (0.002)
	Model ISI (SE)	0.096 (0.021)	0.009 (0.004)
Neuron# 3 (OFC)	Observed ISI (SE)	0.054 (0.011)	0.002 (0.001)
	Model ISI (SE)	0.083 (0.021)	0.007 (0.004)
Neuron# 4 (PL)	Observed ISI (SE)	0.076 (0.025)	0.009 (0.008)
	Model ISI (SE)	0.118 (0.047)	0.014 (0.014)
Neuron# 5 (PL)	Observed ISI (SE)	0.057 (0.021)	0.007 (0.006)
	Model ISI (SE)	0.080 (0.037)	0.006 (0.008)

Figure 1: Spike count correlation (left) and statistically significant correlations only (right).

model. Comparing the mean and variance estimates from the model to those of the empirical ISI data, it seems that a slight systematic upward bias exists in the model fit. While PP-plots for each neuron (not shown here) did not reveal a noticeable lack of fit, bias correction for the parameter estimation method could potentially be explored. Figure 1 suggests that all OFC neurons are correlated, whereas PL neurons are not. The correlations between OFC and PL neurons were not statistically significant.

Discussion

This paper introduces the multivariate Skellam process with resetting for the analysis of simultaneously recorded neural spike trains. Unlike other multivariate point processes, the MSPR has a flexible dependency structure and can be fit with relative ease. Future work includes the additional modelling of the crossing times $T_n^{(i)} = \min\{t : S^{(i)} \geq nk_i\}$ for $k_i, i = 1, ..., p$ to be estimated from the data, as was done in the univariate SPR model of (Ramezan et al., 2016). The challenge is to perform parameter estimation under this more complicated model without sacrificing computational efficiency.

Acknowledgments

References

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., & Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, *18*(18), 7411-7425.

Dimatteo, I., Genovese, C. R., & Kass, R. E. (2001). Bayesian curve-fitiing with free-knot splines. *Biometrika*, *88*(4), 1055-1071.

Gerstein, G. L., & Mandelbrot, B. (1964). Random walk models for the spike activity of a single neuron. *Biophysical Journal*, *4*(1), 41-68.

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

Kaminska, B. (2020). *Value and action encoding for natural outcomes and ethanol in the medial prefrontal and orbitofrontal cortex* (Doctoral dissertation, University of Massachusetts Amherst). doi: 10.7275/19095335

Karlis, D. (2003). Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *52*(3), 381-393.

Karlis, D., & Meligkotsidou, L. (2005). Multivariate Poisson regression with covariance structure. *Statistics and Computing*, *15*, 255-265.

Kass, R. E., & Ventura, V. (2001). A spike-train probability model. *Neural Computation*, *13*(8), 1713-1720.

Kocherlakota, S., & Kocherlakota, K. (1992). *Bivariate discrete distributions*. New York, NY: Marcel Dekker Inc.

Koyama, S. (2015, July). On the spike train variability characterized by variance-to-mean power relationship. *Neural Computation*, *27*(7), 1530-1548.

Lakshminarayana, J., Pandit, S. N. N., & Rao, K. S. (1999). On a bivariate Poisson distribution. *Communications in Statistics - Theory and Methods*, *28*(2), 267-276.

Marshall, A. W., & Olkin, I. (1985). A family of bivariate distributions generated by the bivariate Bernoulli distribution. *Journal of American Statistical Association*, *80*, 332-338.

Paninski, L., Brown, E. N., Iyengar, S., & Kass, R. E. (2008). Statistical models of spike trains. In C. Liang & G. Lord (Eds.), *Stochastic methods in neuroscience.* Oxford University Press.

Ramezan, R., Marriott, P., & Chenouri, S. (2014). Multiscale analysis of neural spike trains. *Statistics in Medicine*, *33*(2), 238-256. doi: 10.1002/sim.5923

Ramezan, R., Marriott, P., & Chenouri, S. (2016). Skellam process with resetting: A neural spike trains model. *Statistics in Medicine*, *35*(30), 5717-5729.

Shahbaba, B., Zhou, B., Lan, S., Ombao, H., Moorman, D., & Behseta, S. (2014). A semiparametric bayesian model for detecting synchrony among multiple neurons. *Neural Computation*, *26*(9).

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, *93*, 1074-1089.

Wójcik, D. K., Mochol, G., Jakuczun, W., Wypych, M., & Waleszczyk, W. (2009, August). Direct estimation of inhomogeneous markov interval models of spike trains. *Neural Computation*, *21*(8), 2105-2113. doi: 10.1162/neco.2009.07-08-828

Zhou, B., Moorman, D., Behseta, S., Ombao, H., & Shahbaba, B. (2016). A dynamic bayesian model for characterizing cross-neuronal interactions during decision-making. *Journal of the American Statistical Association*, *111*(514). doi: 10.1080/01621459.2015.1116988