

# DÉTECTION DE MOTIFS DE DÉPENDANCE AVEC DÉLAI

Julien Chevallier <sup>1</sup> & Thomas Laloë <sup>2</sup>

<sup>1</sup> *LJAD, Université de Nice Sophia-Antipolis, Parc Valrose, 06108 Nice  
julien.chevallier@unice.fr*

<sup>2</sup> *LJAD, Université de Nice Sophia-Antipolis, Parc Valrose, 06108 Nice  
thomas.laloe@unice.fr*

**Résumé.** La méthode des Unitary Events (UE) est couramment utilisée pour détecter des motifs de dépendance parmi les trains de spike de plusieurs neurones. Cette méthode a tout d’abord été introduite pour traiter des données binnées (temps discret) via le calcul du *nombre de coïncidence binné* (Grün, 1996). Cette méthode de comptage a récemment été transposée dans le cadre de données à temps continu (processus ponctuels) dans l’article de Tuleau-Malot et al. (2014). Ils y introduisent le *nombre de coïncidence avec délai* pour deux neurones. Le but de notre travail a été de généraliser cette méthode de comptage à un nombre quelconque de neurones. Sous l’hypothèse d’indépendance, il est possible de calculer l’espérance et la variance de ce nombre de coïncidence dans un cadre Poissonien et de dériver un test d’indépendance. Une étude empirique montre qu’il est possible d’affaiblir l’hypothèse Poissonienne.

**Mots-clés.** Motif de dépendance, Test d’indépendance, Processus de Poisson et de Hawkes.

**Abstract.** The Unitary Events (UE) method is a popular method used to detect dependence patterns of joint spike activity among simultaneously recorded neurons. The method based on binned coincidence count (discrete time) was first introduced in S. Grün’s PhD thesis (1996). This binned coincidence count has recently been transposed in the continuous time framework (point processes) by Tuleau-Malot and al. (2014) with the notion of delayed coincidence count for two neurons. The extension of this count to more than two neurons is the objective of the present work. Under a Poissonian assumption, we are able to compute the expectation and variance of this coincidence count under the independence hypothesis, which leads to an independence test. The possibility to weaken the Poissonian assumption is shown by an empirical study.

**Keywords.** Coincidence pattern, Independence tests, Poisson and Hawkes processes.

## 1 Contexte neurobiologique

La communication inter-neuronale est possible grâce à la capacité des neurones à générer et propager des signaux électriques. Les potentiels d’action (ou “spikes”), qui sont des

pics du potentiel électro-chimique de la membrane des neurones, sont caractéristiques de ce phénomène. C'est la raison pour laquelle le taux de décharge des neurones a largement été étudié depuis quelques années.

L'utilisation de micro électrodes permet l'enregistrement du potentiel au voisinage d'un petit nombre de neurones (Local Field Potential). C'est une des techniques d'enregistrement de données neuronales les plus populaires de nos jours. De plus, l'utilisation de tétrode permet un enregistrement redondant qui, via une triangulation spatiale, est essentiel à l'association des différents trains de spikes à leurs neurones respectifs (spike sorting). En effet, la génération d'un potentiel d'action par un neurone donné est un phénomène stéréotypé. Ainsi, les données qui nous intéressent dans la suite sont formées par un petit nombre  $n$  de trains de spikes (1 par neurone), i.e. d'ensembles de temps d'excitation. Nous supposons avoir accès à des observations i.i.d. de  $n$  processus ponctuels.

Il a été prouvé que les neurones peuvent se synchroniser et former ce que l'on appelle des assemblées neuronales. La compréhension et la détection de ce phénomène a nécessité le développement de méthodes d'analyses statistiques.

Le but de cet exposé est de tester  $\mathcal{H}_0$  : "Les temps d'émission de chaque neurone sont indépendants" contre son alternative.

Une des méthodes les plus populaires à ce jour est la méthode des *Unitary Events* (UE) introduite dans la thèse de S. Grün (1996). Cette méthode est basée sur l'idée qu'une dépendance excitatrice du neurone A vers le neurone B doit être caractérisée par une surabondance du motif suivant : spike du neurone A suivit (dans un délai très court de l'ordre de la milliseconde) d'un spike du neurone B. Si un tel motif est sur-représenté, alors il est qualifié de Unitary Event. Utilisables sur des données binnées, les différentes méthodes proposées par Grün et ses collaborateurs (par exemple, Grün et al. (1999)) ont récemment été généralisées au cadre des processus ponctuels dans un article de Tuleau-Malot et al. (2014). Ils proposent un test d'indépendance entre deux processus ponctuels. Nous présenterons la généralisation de ce test à  $n \geq 2$  processus ponctuels.

Nous présenterons tout d'abord la notion de coïncidence qui permet de donner un cadre mathématique à la détection des motifs décrits dans le paragraphe précédent. Afin d'en déduire un résultat statistique utilisable, il nous faut quantifier le nombre de coïncidences sous  $\mathcal{H}_0$ . Ceci fait, nous pouvons établir un test statistique dont le niveau asymptotique est connu.

## 2 Notion de coïncidence

On considère des enregistrements sur une fenêtre de temps  $[a, b]$  d'un nombre  $n \geq 2$  de neurones. Notons  $N_1, \dots, N_n$  des processus ponctuels sur  $[a, b]$ . La notion de coïncidence pour deux neurones introduites dans l'article de Tuleau-Malot et al. (2014) est la suivante :

Pour un paramètre  $\delta > 0$ , la variable

$$X = \int_{[a,b]^2} \mathbf{1}_{|x-y|\leq\delta} N_1(dx) N_2(dy) = \sum_{X \in N_1} \sum_{Y \in N_2} \mathbf{1}_{|X-Y|\leq\delta}$$

désigne le nombre de coïncidence de délai  $\delta$  entre les processus  $N_1$  et  $N_2$ .

La généralisation que nous proposons est donné par la définition qui suit.

**Définition 1.** Soit  $\mathcal{J} \subset \{1, \dots, n\}$ . On note  $J \geq 2$  son cardinal et  $\mathcal{J} = \{i_1 < \dots < i_J\}$ . Soit  $\delta > 0$ . Le nombre  $X$  de coïncidences de délai  $\delta$  entre les processus  $N_{i_1}, \dots, N_{i_J}$  est défini par :

$$X_{\mathcal{J}} = \int_{[a,b]^J} \mathbf{1}_{\left| \max_{i \in \{1, \dots, J\}} x_i - \min_{i \in \{1, \dots, J\}} x_i \right| \leq \delta} N_{i_1}(dx_1) \dots N_{i_J}(dx_J).$$

Notre manière de compter les coïncidences peut être résumé de la manière suivante :

1. Fixer un paramètre de délai  $\delta > 0$ ;
2. Compter le nombre de fois que tous les processus indicés par le sous-ensemble  $\mathcal{J}$  ont un point à distance moins de  $\delta$  les uns des autres.

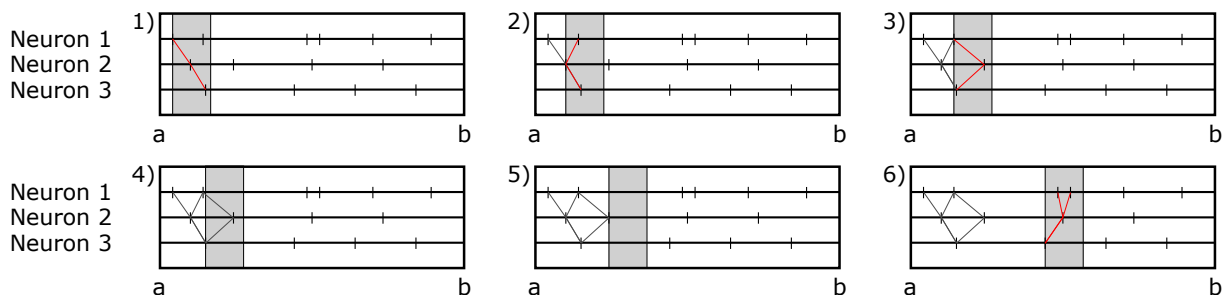


Figure 1: Six premières étapes du calcul dynamique du nombre de coïncidences sur  $[a, b]$  avec  $\mathcal{J} = \{1, 2, 3\}$ . Le rectangle grisé représente la fenêtre de taille  $\delta$ . Les lignes rouges montrent les motifs de coïncidence dénombrés à chaque étape. Les lignes grises montrent les motifs déjà dénombrés par une étape antérieures. A noter que deux motifs sont dénombrés à l'étape 6.

### 3 Test d'indépendance

Afin de construire un test statistique raisonnable, nous devons tout d'abord étudier le comportement du nombre de coïncidence sous  $\mathcal{H}_0$ . Cette étude est possible dans le cadre

de processus de Poisson homogènes. Ce cadre peut être considéré comme trop restrictif. En effet, les processus de Hawkes semblent plus adapté à la modélisation de neurones et ont passé des tests d'adéquation de loi sur données réelles - voir l'article de Reynaud-Bouret et al. (2014). Cependant, la validité de notre test est conservé dans le cadre des processus de Hawkes comme le montre notre étude empirique.

Pour calculer l'espérance et la variance du nombre de coïncidence, il faut tout d'abord considérer les quantités auxiliaires définies dans ce qui suit.

**Proposition 2.** *Soit un entier  $J \geq 2$ . Pour tout  $b > a$  et  $0 < \delta < b - a$  réels, notons pour tout  $k \in \{0, \dots, J\}$*

$$I(J, k) = \int_{[a, b]^{J-k}} \left( \int_{[a, b]^k} \mathbf{1}_{\left| \max_{i \in \{1, \dots, J\}} x_i - \min_{i \in \{1, \dots, J\}} x_i \right| \leq \delta} dx_1 \dots dx_k \right)^2 dx_{k+1} \dots dx_J,$$

avec la convention  $\int_{[a, b]^0} f(x) = f(x)$ . Alors,

- $I(J, J) = J^2 (b - a)^2 \delta^{2J-2} - 2J(J - 1) (b - a) \delta^{2J-1} + (J - 1)^2 \delta^{2J}$   
et pour tout  $k \in \{0, \dots, J - 1\}$ ,

- $I(J, k) = f(J, k) (b - a) \delta^{J+k-1} - h(J, k) \delta^{J+k}$

où  $f(J, k) = \frac{k(k+1) + J(J+1)}{J-k+1}$ ,

et  $h(J, k) = \frac{-k^3 + k^2(2+J) + k(5+2J-J^2) + J^3 + 2J^2 - J - 2}{(J-k+2)(J-k+1)}$ .

Ces quantités auxiliaires apparaissent dans le calcul des moments d'ordre 2 d'une somme définie à partir de processus de Poisson. Elles permettent de factoriser l'espérance et la variance du nombre de coïncidence  $X_{\mathcal{J}}$  comme énoncé dans le résultat suivant.

**Théoreme 3.** *Soient  $\mathcal{J}$  et  $X_{\mathcal{J}}$  comme dans la Définition 1. Sous l'hypothèse que  $N_1, \dots, N_n$  sont des processus de Poisson homogènes d'intensité respective  $\lambda_1, \dots, \lambda_n$  et sous  $\mathcal{H}_0$ , l'espérance et la variance du nombre  $X_{\mathcal{J}}$  de coïncidences sont données par :*

$$m_0 := \mathbb{E}[X_{\mathcal{J}}] = \left( \prod_{j=1}^J \lambda_{i_j} \right) I(J, 0)$$

et

$$v_0 = \text{Var}(X_{\mathcal{J}}) = m_0 + \sum_{k=1}^{J-1} \left( \sum_{\substack{\mathcal{L} \subset \mathcal{J} \\ \#\mathcal{L}=k}} \prod_{l \in \mathcal{L}} \lambda_{i_l}^2 \prod_{j \notin \mathcal{L}} \lambda_{i_j} \right) I(J, k)$$

De plus, si  $(N_1^{(k)}, \dots, N_n^{(k)})_{1 \leq k \leq M}$  désigne  $M$  essais i.i.d. de processus de Poisson homogènes d'intensité respective  $\lambda_1, \dots, \lambda_n$  sous  $\mathcal{H}_0$ , alors :

$$\sqrt{M} (\bar{m} - m_0) \xrightarrow[M \rightarrow \infty]{loi} \mathcal{N}(0, v_0),$$

où  $\bar{m} = \frac{1}{M} \sum_{k=1}^M X_{\mathcal{J}}^{(k)}$  et pour tout  $k = 1, \dots, M$ ,

$$X_{\mathcal{J}}^{(k)} = \int_{[a,b]^J} \mathbf{1}_{\left| \max_{i \in \{1, \dots, J\}} x_i - \min_{i \in \{1, \dots, J\}} x_i \right| \leq \delta} N_{i_1}^{(k)}(dx_1) \dots N_{i_J}^{(k)}(dx_J).$$

Les quantités  $m_0$  et  $v_0$  dépendent des intensités (non observables) des processus ponctuels. Le remplacement des intensités par leurs estimateurs usuels définis par  $\hat{\lambda}_i := \frac{1}{M(b-a)} \sum_{k=1}^M N_i^{(k)}([a, b])$  pour tout  $i = 1, \dots, n$ , ne compromet pas la nature asymptotiquement gaussienne de la variable  $X_{\mathcal{J}}$ , comme justifié par le résultat suivant.

**Remarque 4.**  $N_i^{(k)}([a, b])$  désigne le nombre de points de  $N_i^{(k)}$  présent dans l'intervalle  $[a, b]$ .

**Théoreme 5.** Avec les notations et hypothèses du Théoreme précédant, on a la convergence en loi suivante :

$$\sqrt{M} (\bar{m} - \hat{m}_0) \xrightarrow[M \rightarrow \infty]{loi} \mathcal{N}(0, \sigma^2),$$

où

$$\hat{m}_0 = \prod_{j=1}^J \hat{\lambda}_{i_j} \cdot I(J, 0)$$

et

$$\sigma^2 = v_0 - (b-a)^{-1} \mathbb{E}[X_{\mathcal{J}}]^2 \left( \sum_{j=1}^J \lambda_{i_j}^{-1} \right).$$

De plus,  $\sigma^2$  peut être estimé par

$$\hat{\sigma}^2 = \hat{v} - (b-a)^{-1} I(J, J) \prod_{j=1}^J \hat{\lambda}_{i_j}^2 \left( \sum_{k=1}^J \hat{\lambda}_{i_k}^{-1} \right),$$

où

$$\hat{v} = \hat{m}_0 + \sum_{k=1}^{J-1} \left( \sum_{\substack{\mathcal{L} \subset \mathcal{J} \\ \#\mathcal{L}=k}} \prod_{l \in \mathcal{L}} \hat{\lambda}_{i_l}^2 \prod_{j \notin \mathcal{L}} \hat{\lambda}_{i_j} \right) I(J, k),$$

et la convergence en loi suivante est vérifiée :

$$\sqrt{M} \frac{(\bar{m} - \hat{m}_0)}{\sqrt{\hat{\sigma}^2}} \xrightarrow[M \rightarrow \infty]{loi} \mathcal{N}(0, 1).$$

La preuve repose sur une simple application de la méthode delta.

Ce résultat permet donc la construction d'un test dont le niveau asymptotique est contrôlé sous l'hypothèse que les processus ponctuels sont Poisson homogènes.

## 4 Validation via simulation

L'utilisation de cette procédure sur des données simulées à permis de montrer que :

- l'approximation gaussienne est valide à partir d'un nombre d'essais  $M$  de l'ordre de 50;
- notre procédure reste raisonnable dans le cadre plus général des processus de Hawkes;
- la mise en place d'une procédure de test multiple sur les différents sous-ensembles  $\mathcal{J} \subset \{1, \dots, n\}$  est possible pour détecter les éventuelles connectivités fonctionnelles sous-jacentes.

## Bibliographie

Cet exposé est basé sur un travail en collaboration avec T. Laloë disponible sur <http://hal.inria.fr/hal-00998864>

[1] Tuleau-Malot, C., Rouis, A., Grammont, F., Reynaud-Bouret, P. (2014) , Multiple Tests Based on a Gaussian Approximation of the Unitary Events Method with delayed coincidence count, appearing in *Neural Computation* **26:7**.

[2] Grün, S. (1996) Unitary joint events in multiple neuron spiking activity: detection, significance, and interpretation, PhD thesis.

[3] Grün, S., Diesmann, M., Grammont, F., Riehle, A., Aertsen, A. (1999), Detecting unitary events without discretization of time, *Journal of neuroscience methods* **94**(1), 67–79.

[4] Reynaud-Bouret, P. ; Rivoirard, V. ; Grammont, F. ; Tuleau-Malot, C. (2014) Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis, *Journal of Mathematical Neuroscience*, **4:3**

[5] Benjamini, Y., Hochberg, Y. (1995), Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society. Series B. Methodological* **57**(1), 289–300.