

Université Paris 6

Master 2 Physique des Liquides et matière molle

Pierre Bohec

troisième année de magistère de physique fondamentale, Orsay

Sujet :

Analyse de l'activité de réseaux  
neuronaux par des modèles  
non-paramétriques : application au  
système olfactif de la *Periplaneta*  
*Americana*

Responsable du stage :

Christophe Pouzat

# Table des matières

<b>1</b>	<b>Présentation de la statistique de la décharge neuronale</b>	<b>3</b>
1.1	Processus stochastique ponctuel	3
1.2	Événement ponctuel et loi binomiale	5
1.3	Processus de renouvellement	5
1.3.1	Processus de Poisson homogène	6
1.3.2	<i>Hypothèse nulle</i> de renouvellement	6
1.3.3	Densité de probabilité des intervalles	7
1.4	Vers un modèle plus complet	10
1.4.1	Théorème de rééchantillonnage du temps	10
1.4.2	Les tests d'Ogata	11
1.4.3	Spline cubique de lissage	14
1.5	Processus de Poisson inhomogène	16
1.5.1	Histogramme temporel peri-Stimulus	16
1.5.2	<i>PSTH</i> et intensité conditionnelle	16

# Chapitre 1

## Présentation de la statistique de la décharge neuronale

Un train de potentiels d'action (PA) neuronal est une série de décharge de forme caractéristique (mesurée en mVolt) produites par un neurone. En neurophysiologie, il est possible de se rendre compte de l'activité neuronale en identifiant uniquement le temps correspondant au pic de chaque PA et en ignorant la forme du PA ou encore les dynamiques du potentiel de membrane sous le seuil de décharge. Le seul paramètre permettant de distinguer chaque PA est alors sa latence (cf figure 1.1). Cette section va présenter les divers traitements statistiques nécessaires à la pleine exploitation de ces mesures.

### 1.1 Processus stochastique ponctuel

Les processus stochastiques ponctuels sont des séries d'évènements identiques apparaissant à des latences aléatoires. Dans notre cas, ces évènements sont instantanés et se distinguent uniquement par leurs positions sur l'axe du temps. Dans le cadre de notre étude, nous nous intéressons non pas aux latences mais aux intervalles de temps séparant les potentiels d'action (Inter-Spike Interval ou *ISI*). L'étude probabiliste et statistique de tel processus utilise un formalisme basé sur les trois quantités suivantes [1] : le processus de dénombrement, l'histoire du processus, l'intensité conditionnelle.

**Le processus de dénombrement :** Soit les évènements  $\tau_j$  répartis aléatoirement le long de l'axe du temps (dans notre cas il s'agit des PAs), le processus de dénombrement  $N(t)$  correspond au nombre total d'évènements survenus dans l'intervalle de temps  $(0, t]$  :

$$N(t) = \#\{\tau_j \text{ tel que } 0 < \tau_j \leq t\} \quad (1.1)$$

Il est souvent utile de représenter le processus de dénombrement en traçant la fonction cumulative du nombre d'évènements en fonction du temps [11]. A chaque apparition d'un PA, représenté par un trait vertical, cette fonction est incrémentée d'une unité (cf figure 1.1). L'accélération (ou la décélération) de la décharge d'un neurone, qui correspond au nombre de PAs par unité de temps, est mise en évidence par la courbure convexe (respectivement concave) de la fonction. Une telle analyse fournit donc des informations générales sur le train de PAs.

**L'histoire du processus :** Pour décrire l'évolution d'un processus ponctuel à partir d'un instant  $t$ , il peut être nécessaire de prendre en compte les évènements antérieurs à  $t$ . Si le processus est composé de  $N$  PAs apparaissant aux instants  $\{\tau_j\}_{j=1}^N$ , l'histoire du processus au temps  $t$  est :

$$H_t = \{\tau_j \text{ tel que } \tau_j \leq t\} \quad (1.2)$$

**L'intensité conditionnelle :** Pour un processus composé de  $N$  événements et d'histoire  $H_t$ , l'intensité conditionnelle  $\lambda(t | H_t)$  au temps  $t$  représente la probabilité d'avoir un évènement par unité de temps. Plus précisément, elle est définie comme :

$$\lambda(t | H_t) = \lim_{h \rightarrow 0} \frac{Prob \{ \text{évènement } \epsilon(t, t+h] | H_t \}}{h} \quad (1.3)$$

De plus, il est possible de relier l'intensité conditionnelle  $\lambda(t | H_t)$  à la densité de probabilité  $f(t | H_t)$  d'apparition des  $N(t)$  événements ponctuels. En notant  $\tau_t$  le temps du dernier pontentiel d'action avec  $t > \tau_t$  on obtient [2] :

$$\lambda(t | H_t) = \frac{f(t | H_t)}{1 - \int_{\tau_t}^t f(t | H_t)} \quad (1.4)$$

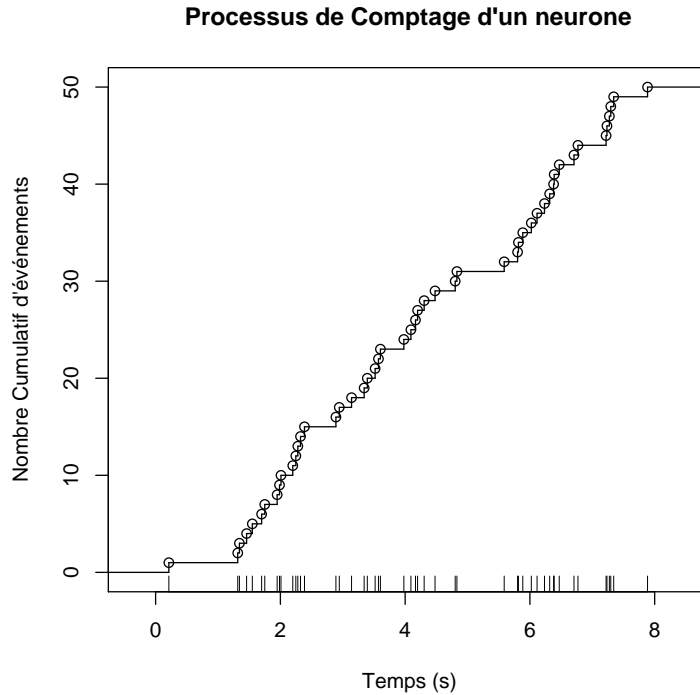


FIG. 1.1 – Exemple de trains de PAs issus d'enregistrements de l'activité d'un neurone du lobe olfactif de *Periplaneta Americana* ; processus de dénombrement en découlant

## 1.2 Événement ponctuel et loi binomiale

Maintenant que l'on a défini les grandeurs utiles à l'analyse statistique des trains de PAs, nous pouvons exprimer la probabilité jointe d'une succession d'intervalles donnée. En reprenant la définition de l'équation 1.3, on en déduit que pour  $\Delta t \rightarrow 0$  :

$$Prob\{\text{événement } \epsilon(t, t + \Delta t) \mid H_t\} \approx \lambda(t \mid H_t)\Delta t \quad (1.5)$$

En pratique, il est nécessaire de discrétiser l'axe des temps en pas  $\Delta t$  appelés *bins*. Le choix de la largeur du *bin* est cruciale mais il n'existe pas de méthode exacte pour la déterminer. La seule contrainte est qu'il y ait au plus un PA par *bin*. Sa valeur ne doit être ni trop grande pour ne pas perdre les informations correspondant aux variations rapides de l'intensité conditionnelle, ni trop petite pour éviter de prendre en compte les fluctuations statistiques. Le *bin* est un paramètre choisi "empiriquement" par l'utilisateur. Le résultat de la discrétisation de l'axe du temps en *bins* est une série de données binaires, les 0 correspondant aux *bins* ne contenant aucun PA et les 1 à ceux contenant un PA.

Afin de définir la probabilité jointe d'une succession d'intervalles, considérons  $N$  PAs aux instants  $\{\tau_j\}_{j \in [1, N]}$  et les  $N - 1$  intervalles en découlant  $\{ISI_j = \tau_{j+1} - \tau_j\}_{j \in [1, N-1]}$ . En supposant que l'axe de temps est composé de  $M$  *bins*, pour  $i \in [1, M]$ ,  $\lambda_i$  est la valeur de l'intensité conditionnelle dans le *bin*  $i$ . Si on note  $n_j = \frac{ISI_j}{h}$  le nombre de *bins* séparant les deux PAs et si on suppose que le *bin*  $i$  contient le temps  $\tau_j$ , la probabilité que le  $j^{\text{e}}$  intervalle soit égal à  $ISI_j$  est :

$$Prob(ISI_j) = \lambda_{i+n_j+1} \cdot \Delta t \prod_{k=1}^{n_j} (1 - \lambda_{i+k} \cdot \Delta t) \quad (1.6)$$

On reconnaît alors l'expression de la probabilité jointe (non normalisée) d'une variable aléatoire binomiale avec une probabilité de succès  $\lambda_i \cdot \Delta t$  et une probabilité d'échec  $1 - \lambda_i \cdot \Delta t$  d'une série de  $n_j + 1$  épreuves de Bernoulli de configuration :  $(\underbrace{0, 0, \dots, 0}_{n_j}, 1)$ .

La probabilité d'obtenir la suite  $\{ISI_j\}_{j \in [1, M-1]}$  est donc égal à :

$$Prob(\{ISI_j\}_{j \in [1, N-1]}) = \prod_{j=1}^{M-1} \left( \lambda_{i+n_j+1} \cdot \Delta t \prod_{k=1}^{n_j} (1 - \lambda_{i+k} \cdot \Delta t) \right) \quad (1.7)$$

La formule 1.7 est le cas le plus général pour étudier un train de PAs car nous n'avons pas encore fait d'hypothèses sur la forme de l'intensité conditionnelle. Par la suite nous allons simplifier le modèle en supposant d'abord que l'intensité  $\lambda$  est une constante puis qu'elle ne dépend que du temps écoulé depuis le dernier PA.

## 1.3 Processus de renouvellement

Un processus de renouvellement est un processus de dénombrement pour lequel les intervalles entre deux événements consécutifs sont des variables aléatoires indépendantes et identiquement distribuées (*iid*). Le terme "renouvellement" traduit le fait que la probabilité d'avoir un PA ne dépend que du temps écoulé depuis le dernier événement. Dans ce cas, seule la donnée du dernier événement n'est nécessaire pour décrire l'histoire du processus. L'étude d'un tel processus repose donc sur la distribution des intervalles entre PAs.

### 1.3.1 Processus de Poisson homogène

Un processus de Poisson de paramètre  $\lambda$  constant est un processus de renouvellement particulier dont la fonction de répartition des intervalles entre deux événements s'écrit :

$$f(ISI) = \begin{cases} \lambda e^{-\lambda ISI} & \text{si } ISI \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (1.8)$$

Le paramètre  $\lambda$  de la distribution représente le nombre moyen de PAs par *bin*. On peut montrer, sous l'hypothèse que les événements sont séparés dans le temps, que la probabilité d'apparition  $K$  événements entre  $t$  et  $t + T$  est égale à :

$$Prob(N(t + T) - N(t) = K) = \frac{(\lambda T)^K}{K!} e^{-\lambda T} \quad (1.9)$$

On remarque que cette formule n'est que le cas particulier de l'équation 1.7 où  $\lambda$  est constant.

Le processus de Poisson homogène n'est pas très efficace pour étudier les trains de potentiels d'action, mais comme on le verra plus tard (section 1.4.1) ses caractéristiques sont très utiles pour quantifier l'ajustement d'un modèle de décharge à des données expérimentales. Une propriété très utile concerne la fonction cumulative d'un processus de Poisson homogène [3]. En effet, si on observe  $K$  événements d'un processus de Poisson de paramètre  $\lambda = 1$  pendant un temps  $T$ , alors la fonction de distribution cumulative sera linéaire de pente  $\frac{1}{T}$  dans l'intervalle  $(0, T]$ , égale à 0 pour  $t \leq 0$  et à 1 pour  $t \geq T$ .

### 1.3.2 Hypothèse nulle de renouvellement

Bien que le processus de Poisson homogène ne traduise (presque) jamais nos décharges de neurones convenablement, il n'en demeure pas moins que nous pouvons vérifier *l'hypothèse nulle* (i.e les variables aléatoires sont *iid*). Pour cela, le processus doit passer différents tests avec succès : le test des rangs, le test d'autocorrélation [8] et le test du  $\chi^2$ . Afin d'illustrer tous ces tests, nous avons simulé un train de 1200 PAs tirés indépendamment et distribués identiquement selon une distribution *Log Normale* (figure 1.2).

#### Le test des rangs

Considérons les  $N$  événements définis dans la section 1.2 et les  $N - 1$  intervalles correspondants. Le rang  $O_j$  de l'intervalle  $ISI_j$  correspond à son indice dans la suite  $\{ISI_j\}_{j=1}^{N-1}$  ordonnée dans l'ordre croissant. Si les *ISIs* vérifient *l'hypothèse nulle* alors les couples de coordonnées  $(O_j, O_{j+k})$  ( $k \neq 0$ ) devrait être uniformément répartis sur le graphe  $O_{j+k}$  vs  $O_j$ . Nous traçons donc systématiquement les deux figures  $O_{j+1}$  vs  $O_j$  et  $O_{j+2}$  vs  $O_j$  appelé respectivement *lag 1* et *lag 2*. Ces deux graphes, représentés en haut de la figure 1.2, montrent bien une répartition uniforme des points.

#### Le test du $\chi^2$

En plus du test graphique sur les rangs, nous effectuons le test du  $\chi^2$ . En effet, sous *l'hypothèse nulle*, si l'on divise la surface des graphes précédents en plusieurs carrés d'aires  $d^2$  identiques, chaque carré possède le même nombre de points  $\frac{N}{d^2}$ . Donc si  $O_i$  est le nombre de points observé dans le carré  $i$  et  $E_i$  est le nombre de points attendus alors le  $\chi^2$  sera défini comme :

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \quad (1.10)$$

Ce test permet de quantifier l'uniformité du test des rangs et de définir un intervalle de confiance. Il est effectué sur les graphes de *lag 1* et *lag 2* présentés précédemment ainsi que sur les *lags* suivants. Nous tracons ensuite la valeur du  $\chi^2$  obtenue pour chaque lag (figure 1.2 en bas à droite). L'intervalle de confiance à 95% défini à partir de l'*hypothèse nulle* est indiqué en grisé. Nous voyons qu'aucun point ne sort de cette zone de confiance pour nos données simulées (i.e < 5%).

### Le test d'autocorrélation

Le dernier graphe correspond à un test d'autocorrélation sur les intervalles  $\{ISI_j\}_{j=1}^{N-1}$ . En notant le coefficient d'autocorrélation  $R$  on peut définir :

$$R(\text{lag } k) = \frac{\sum_{i=1}^{N-1} ISI_i \cdot ISI_{i+k}}{\left(\sum_{i=1}^{N-1} ISI_i^2 \cdot \sum_{i=1}^{N-1} ISI_{i+k}^2\right)^{\frac{1}{2}}} \quad (1.11)$$

Ce test mesure l'inter-dépendance des intervalles. Si les *ISIs* sont indépendants, la fonction d'autocorrélation prendra des valeurs faibles. Sous la désormais célèbre *hypothèse nulle*, nous pouvons définir un intervalle de 95% de confiance. Encore une fois, nous remarquons que l'intégralité des  $R$  se situe dans la bande de confiance à 95% pour les données simulées (cf figure 1.2 en bas à gauche).

### 1.3.3 Densité de probabilité des intervalles

Lorsque l'intensité conditionnelle  $\lambda$  d'un processus de renouvellement dépend du temps écoulé depuis le dernier PA, la distribution des ISIs est susceptible de suivre une loi de probabilité exponentielle. En plus de vérifier le processus de renouvellement, il est possible d'ajuster la distribution empirique des ISIs avec différentes densités de probabilités qui sont : *log-normale*, *Inverse gauss*, *Gamma*, *Weibull*, *Exponentielle réfractaire*, *log logistique*

Ceci est possible en estimant les paramètres de chaque distribution par maximisation de la vraisemblance. En effet si on veut vérifier que nos données suivent une certaine loi statistique de densité  $f(X | \theta)$ , où  $X$  est une variable aléatoire et  $\theta$  l'ensemble des paramètres de la loi, nous pouvons définir une grandeur  $L$  (pour *likelihood*) appelée vraisemblance. En supposant que les données sont composées de  $N - 1$  intervalles  $\{ISI_j\}_{j=1}^{N-1}$ , nous obtenons :

$$L\left(\{ISI_j\}_{j=1}^{N-1} | \theta\right) = \prod_{k=1}^{N-1} f(ISI_k | \theta) \quad (1.12)$$

Les paramètres de la distribution seront optimisés en maximisant la probabilité de trouver chaque réalisation et donc en maximisant la fonction de vraisemblance. Les maxima de vraisemblance des distributions *Log Normal*, *Inverse Gauss* et *Exponentielle réfractaires* peuvent être calculés algébriquement. Quant aux paramètres des distributions restantes, ils sont obtenus par optimisation numérique.

Afin de vérifier si les paramètres estimés sont corrects, nous effectuons un test appelé *Quantile-Quantile* en traçant les quantiles de la distribution empirique en fonction des quantiles de la distribution théorique. Le p-quantile d'une distribution de variable aléatoire *ISI* est défini comme la valeur  $X$  telle que :

$$Prob(ISI \leq X) = p \quad (1.13)$$

Le cas particulier  $p = \frac{1}{2}$  est communément appelé la médiane.

Il en résulte que si le graphe *Quantile-Quantile* est une droite de pente 1, alors les paramètres de la distribution sont correctement optimisés et la distribution de probabilité hypothétique traduit convenablement la distribution empirique. Afin de poursuivre avec l'exemple précédent, nous avons confronté le train de PAs simulé selon la loi *Log Normale* avec les six distributions évoquées ci-dessus

(cf figure 1.3). Les deux lignes en pointillés de chaque courbe correspondent aux intervalles de confiance à 95% et 99%. Visuellement, nous observons que seul le tracé *Quantile-Quantile* correspondant à la loi *Log Normale* respecte les intervalles de confiance.

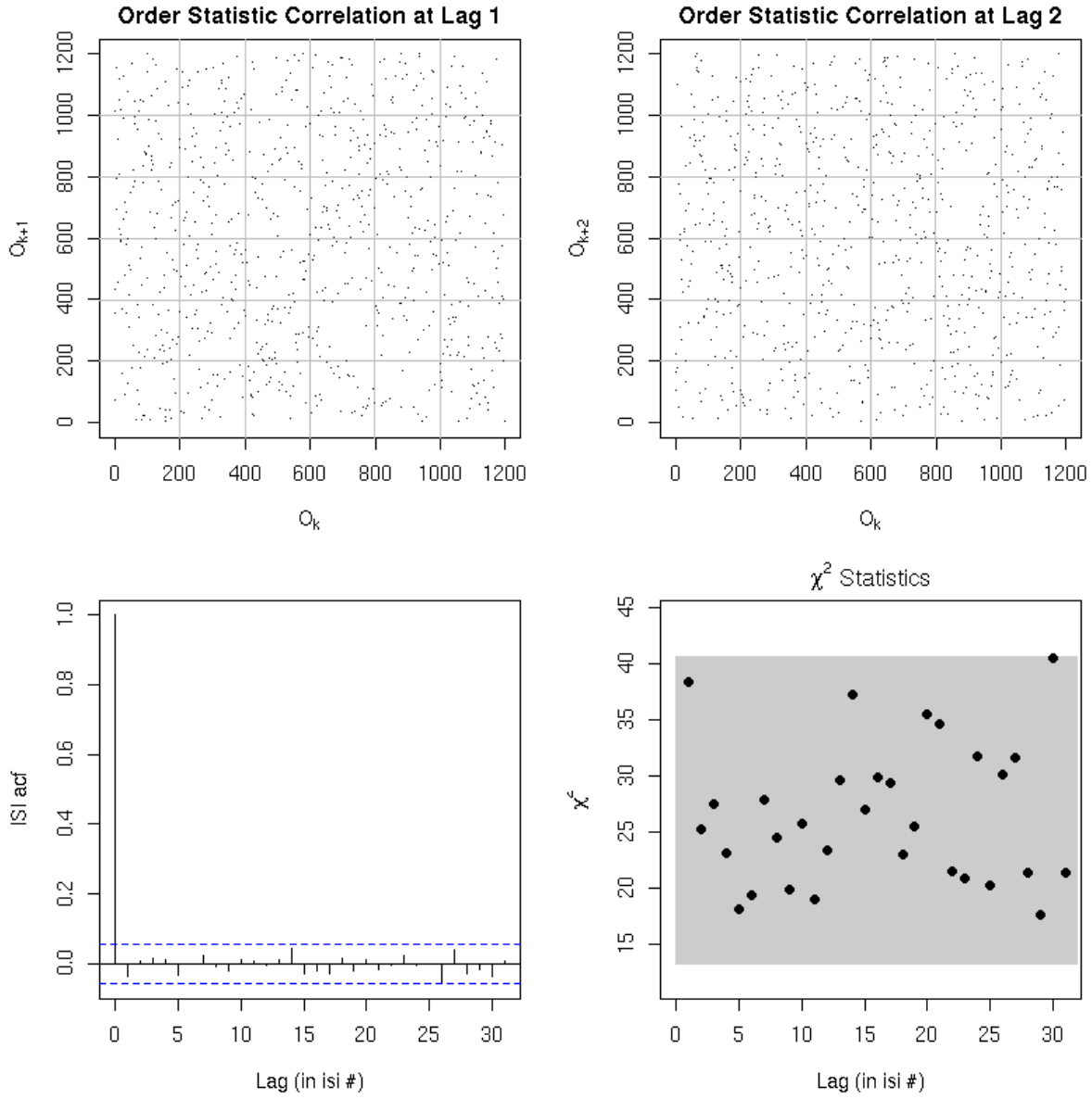


FIG. 1.2 – Le test de renouvellement sur des données simulées de 1200 intervalles entre potentiel d'action suivant une distribution *Log Normale*

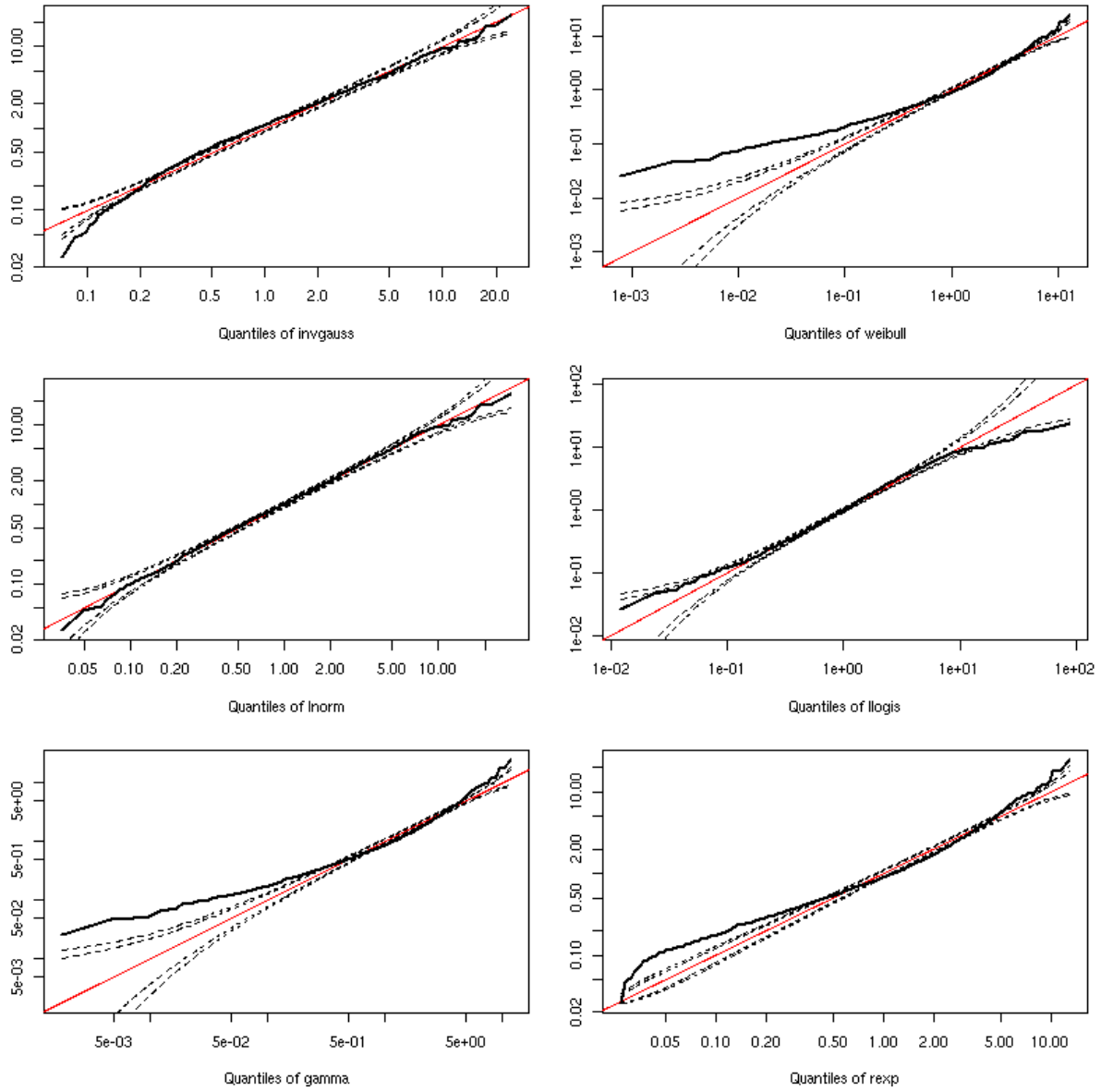


FIG. 1.3 – Graphe Quantile Quantile sur les données simulées de 1200 intervalles entre potentiels d'actions suivant une distribution Log Normal et ajusté en fonction de six distributions

## 1.4 Vers un modèle plus complet

Pour l'instant nous avons exprimé l'intensité conditionnelle en ne prenant en compte que les temps écoulés entre chaque potentiel d'action. Mais il se trouve que l'intensité peut prendre une forme plus complexe. Elle peut dépendre du temps écoulé depuis les deux ou trois derniers PAs. Les décharges nerveuses peuvent aussi être obtenues après l'application d'un stimulus et dans ce cas, l'intensité dépendra vraisemblablement du temps écoulé depuis le stimulus. Il est nécessaire de pouvoir complexifier notre modèle afin de prendre en compte plus de paramètres. De plus, il n'existe pas de modèle paramétrique viable pour décrire au mieux l'activité cérébrale. En effet ces modèles [10], comprenant des paramètres "bio-physiques" tel que les caractéristiques des membranes neuronales, sont basées sur la résolution de systèmes d'équation aux dérivées partielles non-linéaires du potentiel de membrane. Le nombre de paramètres en jeu est alors important, ce qui rend leur optimisation très difficile. Pour cette raison, d'autres chemins sont explorés depuis une trentaine d'années faisant appel à des modèles semi ou non-paramétriques [1, 6, 9].

Le modèle envisagé ici pour rendre compte de l'activité neuronale est un modèle non-paramétrique où l'estimation de l'intensité conditionnelle est effectuée par régression binomiale à l'aide des splines cubiques de lissages (section 1.4.3). L'intérêt des splines de lissages est la maîtrise de ses fondements mathématiques (noyaux reproduisant d'espace de hilbert ou RKHS [12], qui fournissent des preuves consistantes des méthodes d'estimation des paramètres de lissages (Validation croisée). Il en résulte une bonne compréhension et donc une bonne adaptation au niveau numérique avec des logiciels existant sur divers environnements [13, 4].

Les problèmes avec ces types de modèles sont de deux natures différentes. D'un coté la validation "statistique" de l'estimation de l'intensité conditionnelle et du modèle de décharge choisi, de l'autre l'interprétation biologique des résultats significatifs. Le premier problème trouve une réponse dans la littérature géophysique [7] avec l'étude de l'occurrence des tremblements de terres qui, au même titre que les trains de potentiels d'action, peuvent être considérés comme des événements ponctuels. On montre dans la section suivante que tout processus stochastique ponctuel d'intensité conditionnelle  $\lambda(t | H_t)$  (équation 1.3) est susceptible d'être ramené, via un artifice mathématique utilisant un rééchantillonnage du temps, à un processus de Poisson homogène d'intensité égal à 1. L'intérêt est bien sûr de se ramener à un cas beaucoup plus simple sur lequel des tests efficaces existent. Yoshihiko Ogata a établi une série de tests appelés avec originalité *tests d'Ogata*. Ces tests ont déjà été partiellement utilisés en neuroscience [5] et dernièrement l'article [2] fournit une preuve mathématique du *théorème de rééchantillonnage du temps*.

### 1.4.1 Théorème de rééchantillonnage du temps

Si on observe  $N$  événements ponctuels  $\{\tau_j\}_{j=1}^N$  dans l'intervalle  $(0, T]$  tels que  $0 < \tau_1 < \tau_2 < \dots < \tau_N \leq T$ , d'intensité conditionnelle intégrable  $\lambda(t | H_t)$ , on peut définir la fonction  $\Lambda$  pour  $j \in [1, N]$  :

$$\Lambda(\tau_j) = \int_0^{\tau_j} \lambda(u | H_t) du \quad (1.14)$$

D'après le *théorème de rééchantillonnage du temps* [2], la famille  $\{\Lambda(\tau_j)\}_{j=1}^N$  suit un processus de Poisson homogène d'intensité égale à 1. On peut illustrer ceci en prenant la limite  $h \rightarrow 0$  et  $n \rightarrow \infty$  de l'équation 1.6 puis en effectuant la transformation  $t \rightarrow \Lambda(t)$ . Alors la densité de probabilité  $p((\Lambda_{j+1}) - (\Lambda_j))$  tend vers :

$$p(\Lambda(\tau_{j+1}) - \Lambda(\tau_j)) \longrightarrow \exp(-(\Lambda(\tau_{j+1}) - \Lambda(\tau_j))) \quad (1.15)$$

On reconnaît la forme d'un processus de Poisson homogène défini dans l'équation 1.8. Cela implique que la famille de variable aléatoire  $\{\Lambda(\tau_{j+1}) - \Lambda(\tau_j)\}_{j=1}^{N-1}$  est indépendante et identiquement

distribuée suivant une loi exponentielle. Il est alors possible d'effectuer une batterie de tests utilisés normalement pour un processus de Poisson homogène sur n'importe quelle estimation d'une intensité conditionnelle  $\lambda(t | H_t)$ .

### 1.4.2 Les tests d'Ogata

Les tests d'Ogata peuvent être utilisés pour vérifier un processus de Poisson homogène. De plus, si une estimation de l'intensité conditionnelle  $\lambda(t | H_t)$  est disponible, la procédure pour effectuer les tests d'Ogata commence par une transformation recursive de chaque spike  $\{\tau_j\}_{j=1}^N$  observé entre  $(0, T]$  en  $\{\Lambda_j\}_{j=1}^{j=N}$ , soit :

$$\Lambda_1 = 0 \tag{1.16}$$

$$\Lambda_j = \Lambda_{j-1} + \int_0^{\tau_j - \tau_{j-1}} \lambda(t | H_t) dt \tag{1.17}$$

Après avoir effectué cette transformation de l'axe des temps, nous pouvons appliquer une série de 4 tests : le test de Kolmogorov-Smirnov, le test des intervalles transformés, le test de Berman, le test de la variance vs la moyenne. Comme dans la section précédente, nous allons illustrer les tests avec la simulation de trains de PAs selon une loi *log-normale*. La relation entre la densité de probabilité de la distribution d'intervalles et l'intensité conditionnelle est donnée par la formule 1.4.

#### Test de Kolmogorov-Smirnov

Le théorème énoncé précédemment implique que le nombre cumulatif d'événement en fonction du temps transformé devrait se situer sur une droite de pente 1 passant par l'origine (voir section 1.3). La déviation par rapport à cette droite est quantifiée avec la statistique de Kolmogorov-Smirnov. Ce test nous permet de comparer une fonction de répartition théorique à une fonction de répartition empirique. En haut à gauche de la figure 1.4 est représenté le nombre cumulatif de potentiels d'action des données simulées utilisées dans la section précédente. Les pointillés représentent les intervalles de confiance à 95% et 99%.

#### Test sur des intervalles transformés

Soit  $u_j, j > 1$ , défini par :

$$u_j = 1 - \exp(\Lambda_{j-1} - \Lambda_j) \tag{1.18}$$

Les  $\Lambda_{j-1} - \Lambda_j$  étant identiquement et indépendamment distribués selon une loi uniforme, il devrait en être de même pour les  $u_j$  sur l'intervalle  $]0, 1[$ . On se retrouve donc avec le même type de graphe que sur le test des rangs de la section 1.3.2. En haut à droite de la figure 1.4 est représenté  $u_j$  en fonction  $u_{j+1}$ .

#### Test de Berman

Pour ce test représenté en bas à gauche de la figure 1.4, nous traçons la fonction cumulative de la distribution empirique en fonction des  $u_j$  (définie dans le test précédent). Grâce au test de Kolmogorov-Smirnov et sous l'hypothèse nulle (*iid* et distribution uniforme des  $u_j$  sur  $]0, 1[$ ) nous pouvons tracer les intervalles de confiance à 95% et 99%.

## Test de la Variance vs la moyenne

Nous utilisons maintenant une propriété classique des processus de Poisson homogènes : l'égalité entre la variance et la moyenne. Pour ce test, nous divisons l'axe des temps transformés en  $K_w$  parties de taille  $w$ . Ensuite, nous comptons le nombre d'événements dans chaque partie et calculons la moyenne  $M_w$  et la variance  $V_w$  sur les  $K_w$ . Nous répétons cette opération sur des parties  $\{w_1, \dots, w_L\}$  de taille croissante. En bas à droite de la figure 1.4 est tracé la variance  $V_w$  en fonction de la moyenne  $M_w$ . Alors, si les événements transformés  $\{\Lambda_j\}_{j=1}^N$  suivent un processus de Poisson homogène d'intensité égale à 1, les points seront alignés selon une droite de pente 1.

Nous sommes maintenant armés d'une batterie de tests permettant de valider (ou pas) l'estimation non paramétrique d'une intensité conditionnelle. Le grand intérêt des tests de Ogata est leur universalité qui nous permet de vérifier des estimations de nature différente avec un seul et même outil.

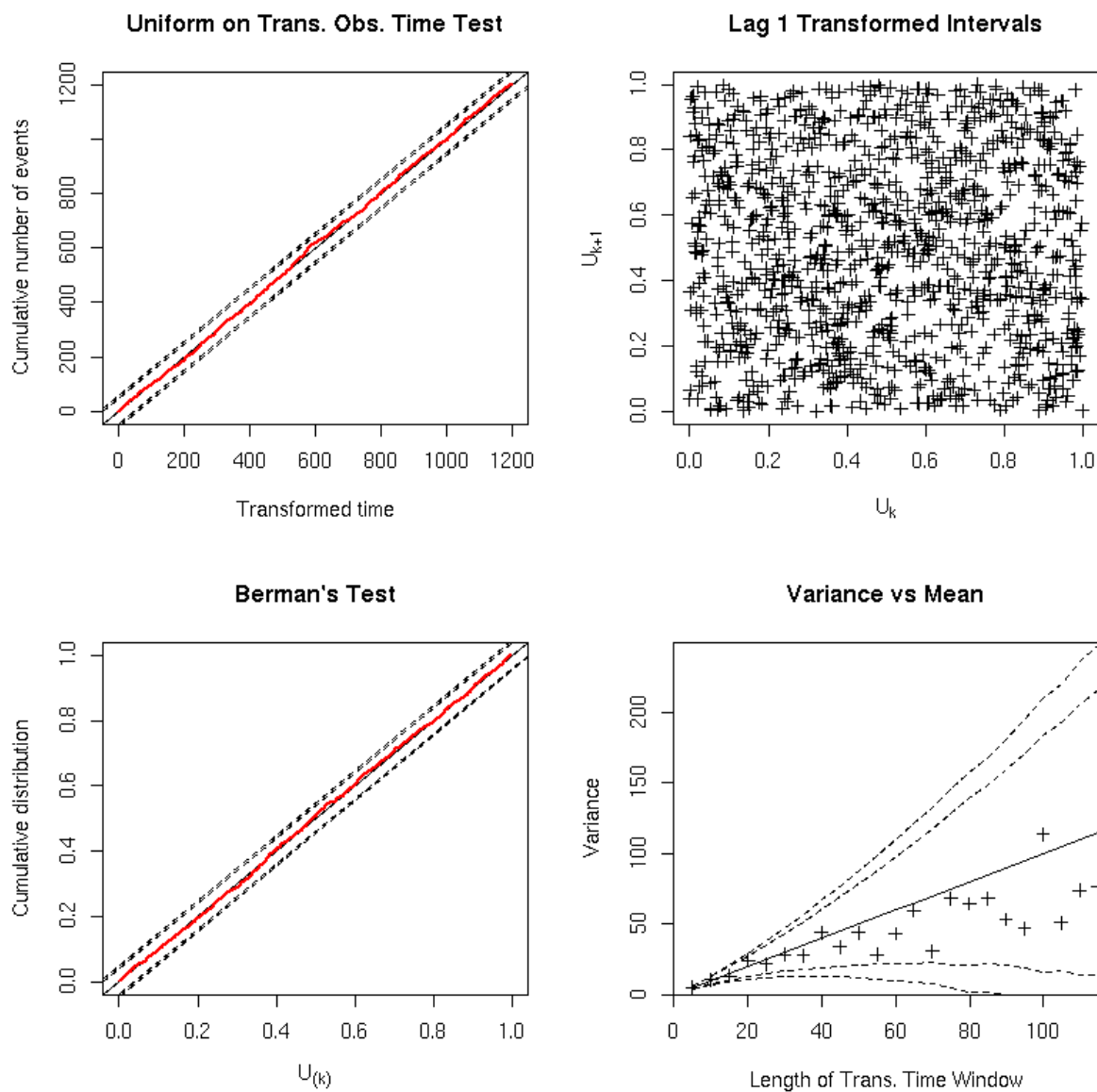


FIG. 1.4 – Test de Ogata sur des données simulées de 1200 intervalles entre potentiels d’actions suivant une distribution Log Normal et ajusté en fonction de six distributions

### 1.4.3 Spline cubique de lissage

Nous sommes amenés à utiliser des modèles non-paramétriques pour décrire des décharge neuronale. Pour comprendre le chemin que l'on emprunte, il faut se rappeler de la section 1.2 où l'on pouvait reconnaître dans l'équation 1.6 la probabilité jointe d'une variable aléatoire binomiale avec une probabilité de succès  $\lambda(t | H_t) \Delta t$ . Sous cette hypothèse de processus binomial, nous cherons à estimer l'intensité conditionnelle par une méthode de maximisation de vraisemblance pénalisée. Cette intensité sera approximée par des polynômes du troisième degré par morceaux appelés *splines cubiques*. L'intérêt de cette méthode est de pouvoir considérer plusieurs variables dans l'ajustement de l'intensité conditionnelle, ce qui permet de prendre en compte plus de données dans l'histoire du processus. Par exemple, si l'on note  $\{\tau_j\}_{j=1}^n$  les temps des différents PAs observés à l'instant  $t$ , on peut alors estimer l'intensité conditionnelle  $\lambda(t - \tau_1, t - \tau_2, \dots, t - \tau_n)$ . En pratique, le modèle le plus complexe est constitué de trois variables, on travaillera donc avec la fonction :  $\lambda(t - \tau_1, t - \tau_2, t - \tau_3)$ . Les estimations seront ensuite évaluées par les tests d'Ogata présentés précédemment.

#### Vraisemblance pénalisée

Soit  $\{Y_i\}_{i=1}^N$  la valeur de la variable de Bernoulli aux instants  $\{t_i\}_{i=1}^N$ , c'est à dire si  $t_i = \tau_j$  alors  $Y_i = 1$  sinon  $Y_i = 0$ . Afin de travailler sur un domaine où la fonction  $\lambda(t | H_t) \Delta t$  n'est pas bornée, nous effectuons un changement de variable avec le modèle logistique en posant :

$$f(t - \tau_1, t - \tau_2, t - \tau_3) = \log \left( \frac{\lambda(t - \tau_1, t - \tau_2, t - \tau_3)}{1 - \lambda(t - \tau_1, t - \tau_2, t - \tau_3)} \right) \quad (1.19)$$

Pour plus de clarté, nous allons travailler sur une forme simplifiée de  $f$  en ne prenant en compte que la dépendance à une variable  $\{x_i\}_{i=1}^n$ . On peut alors écrire l'opposé du logarithme de la vraisemblance pour une distribution binomiale :

$$-\frac{1}{n} \sum_{i=1}^n \left( Y_i f(x_i) - \log \left( 1 + e^{f(x_i)} \right) \right) \quad (1.20)$$

De plus nous allons pénaliser la courbure de l'ajustement en insérant un paramètre de lissage  $\rho$  tel que :

$$\rho \int_0^1 \left( \frac{d^2 f}{dx^2} \right)^2 dx \leq \eta \quad (1.21)$$

$\eta$  est en quelque sorte le degré de liberté du lissage. Mais il se trouve que la recherche du minimum de vraisemblance couplé avec la pénalisation est alors équivalente à la minimisation de la quantité suivante :

$$-\frac{1}{n} \sum_{i=1}^n \left( Y_i f(x_i) - \log \left( 1 + e^{f(x_i)} \right) \right) + \rho \int_0^1 \left( \frac{d^2 f}{dx^2} \right)^2 dx \quad (1.22)$$

Le cas où  $\rho = 0$  est une estimation non-pénalisée. Quand  $\rho \rightarrow \infty$  la fonction estimée est une ligne droite. L'influence de  $\rho$  est montrée figure 1.5 (où  $\rho$  est noté  $\lambda$ ). Nous sommes donc dans un cas très compliqué car on cherche à minimiser l'expression 1.22 qui a deux inconnues, la fonction  $f$  et le paramètre de lissage  $\rho$ . Il est alors possible de montrer (page 37 Gu [4]) que l'une des solutions est de décomposer  $f$  en une fonction polynomiale du troisième degré :

$$f(x) = \delta_1 + \delta_2 t + \sum_{j=1}^{q-2} \delta_{j+2} R(x, x_j) \quad (1.23)$$

où  $\{R(x, x_{j+2})\}_{j=1}^{q-2}$  sont des polynômes de degré 3 appelés splines cubiques. Les points  $x_j$  sont appelés les noeuds. La fonction utilisée nous permet de choisir le nombre de noeud. En pratique les noeuds sont placés uniformément dans la gamme des valeur de  $x$  observées.

## Validation croisée

Maintenant, nous avons déterminé la forme possible de  $f$ , il nous faut choisir le paramètre de lissage  $\rho$ . Le critère adapté peut-être la minimisation de la fonction  $M(\rho)$  :

$$M(\rho) = \frac{1}{n} \sum_{i=1}^n (f_\rho(x_i) - Y_i)^2 \quad (1.24)$$

En pratique nous minimisons la fonction  $V(\rho)$  :

$$V(\rho) = \frac{1}{n} \sum_{i=1}^n \left( f_\rho(x_i)^{[-i]} - Y_i \right)^2 \quad (1.25)$$

où  $f_\rho(x_i)^{[-i]}$  est la fonction prédite au point  $x_i$  à partir des données sans le point  $x_i$ . Cette méthode est appelée la *validation croisée*.

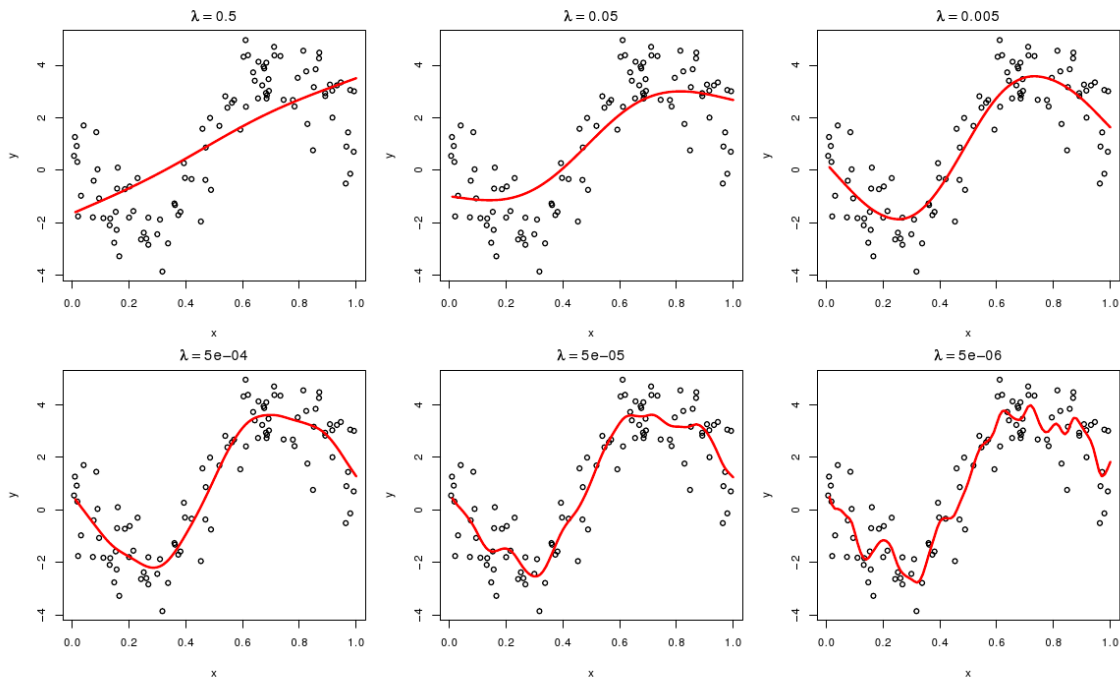


FIG. 1.5 – Influence du paramètre de lissage noté  $\lambda$  sur des données simulées

## 1.5 Processus de Poisson inhomogène

### 1.5.1 Histogramme temporel peri-Stimulus

Jusqu'à présent, les outils que nous avons considérés ne peuvent pas, pour diverses raisons, prendre en compte l'analyse de train de potentiels d'action acquis pendant la présentation d'un stimulus. La mesure de tels trains de potentiels d'action est généralement effectuée à plusieurs reprises sur des temps relativement courts (0.5s) en utilisant le même support de stimulation (contrairement à l'activité spontanée qui est mesurée une fois sur des durées de l'ordre de la minute). Nous avons représenté figure 1.8 les différents PAs (représentés par des traits) en fonction de chaque stimulation (*trial*).

L'utilisation de l'*Histogramme Temporel Peri-Stimulus* ou *PSTH* (figure 1.6) est alors préféré au graphe du processus de comptage représenté figure 1.1. Pour construire un *PSTH*, nous utilisons le temps auquel est appliqué le stimulus pour aligner les différentes mesures de décharges de neurones. Ensuite, nous divisons l'axe des temps en parties égales ou *bin* et nous comptons le nombre d'événements apparaissant dans chaque *bin*. Puis, nous représentons sous forme d'histogramme le nombre de PAs en fonction du temps.

Un *PSTH* est représenté figure 1.6 en prenant une largeur de *bin* de 20 ms. Le stimulus est représenté par une bande grise et dure 0.5seconde. En rouge est représenté l'intervalle de confiance à 95% sous l'hypothèse d'un processus de Poisson inhomogène. La figure 1.7 illustre le problème du choix du *bin* expliqué en section 1.2. En choisissant un *bin* trop petit (3ms), nous voyons apparaître le problème de fluctuations statistiques pour l'histogramme mais aussi pour les intervalles de confiance.

### 1.5.2 *PSTH* et intensité conditionnelle

Il est possible d'interpréter le *PSTH* comme le taux moyen de décharge par *bin* ou encore la probabilité non normalisée d'avoir un événement dans un *bin*. En reprenant l'idée de processus Poissonien, où les intervalles successifs ne sont pas corrélées, nous pouvons définir une intensité conditionnelle  $\lambda(t)$  ne dépendant que du temps (qui prend en réalité comme référence le temps auquel est appliqué le stimulus) et donc généraliser l'équation 1.9 à un processus de Poisson dit inhomogène :

$$Prob(N(t+T) - N(t) = K) = \frac{\left(\int_t^{t+T} \lambda(u) du\right)^K}{K!} \exp\left(-\int_t^{t+T} \lambda(u) du\right) \quad (1.26)$$

En pratique, les intégrales sont calculées numériquement avec un pas de temps égal à la largeur du bin.

Le *PSTH* produit une estimation brutale de l'intensité conditionnelle. Il serait mieux d'avoir une estimation de  $\lambda(t)$  à l'intérieur de chaque *bin*. Ceci est possible avec les techniques de lissage présentées précédemment. La méthode des splines cubiques de lissage permet de construire l'histogramme temporel lissé peri-stimulus (en anglais *SPSTH* ou *Smooth Peri Stimulus Histogram* cf courbe rouge de la figure 1.8). La qualité du *SPSTH* peut être jugée par les tests d'Ogata.

Afin d'avoir un modèle plus complet, il est possible de prendre en compte plus de paramètres pour estimer l'intensité conditionnelle d'une décharge neuronale stimulée. Comme dans la section 1.4.3, nous pouvons prendre en compte le temps des 3 derniers spikes, mais cette fois il faut aussi ajouter le temps du stimulus, qui par construction du *PSTH*, est une dépendance au temps. L'intensité conditionnelle aura la forme :  $\lambda(t, t - \tau_1, t - \tau_2, t - \tau_3)$ .

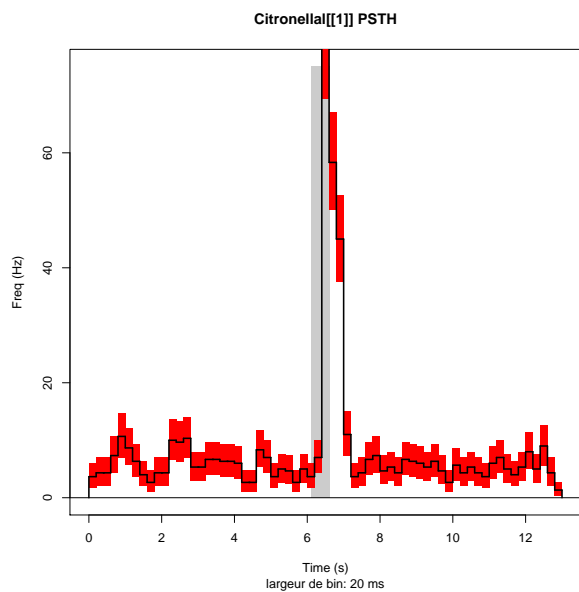


FIG. 1.6 – *PSTH* d'un neurone le stimulus olfactif Citronellal en utilisant un *bin* de 20 ms

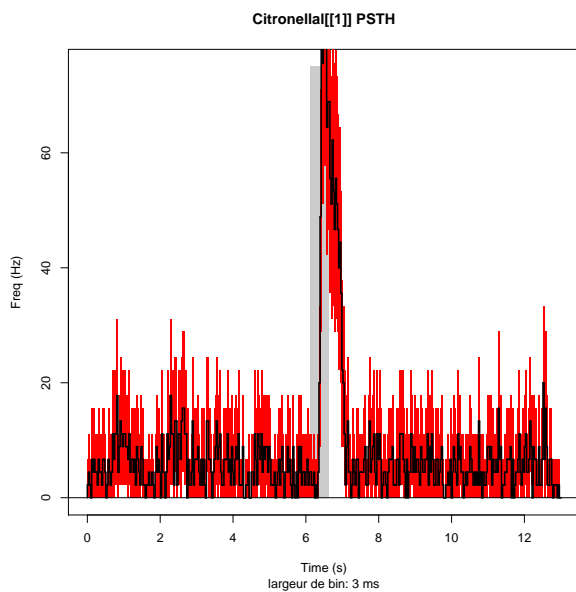


FIG. 1.7 – *PSTH* d'un neurone avec le stimulus olfactif Citronellal en utilisant un *bin* de 3 ms

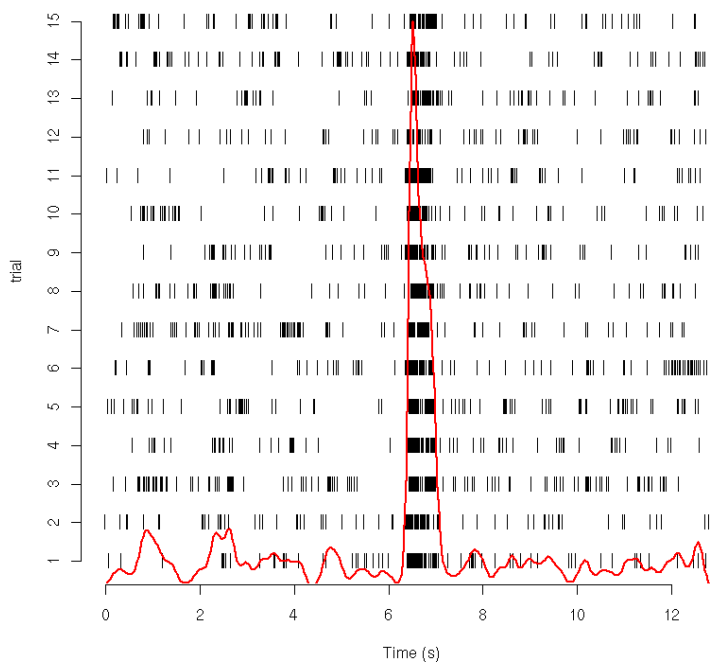


FIG. 1.8 – Représentation de l'occurrence des PAs d'un neurone par des traits noirs verticaux en fonction des répétitions de la stimulation olfactive; En rouge, le *SPSTH* du même neurone

# Bibliographie

- [1] D. R. Brillinger. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cybern*, 59(3) :189–200, 1988. 3, 10
- [2] Emery N Brown, Riccardo Barbieri, Valérie Ventura, Robert E Kass, and Loren M Frank. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput*, 14(2) :325–346, Feb 2002. Available from : <http://www.stat.cmu.edu/~kass/papers/rescaling.pdf>. 4, 10
- [3] D. R. Cox and P. A. W. Lewis. *The Statistical Analysis of Series of Events*. John Wiley & Sons, 1966. 6
- [4] Chong Gu. *Smoothing Spline Anova Models*. Springer, 2002. 10, 14
- [5] D.H. Johnson. Point process models of single-neuron discharges. *J. Computational Neuroscience*, 3(4) :275–299, 1996. 10
- [6] R. E. Kass and V. Ventura. A spike-train probability model. *Neural Comput.*, 13(8) :1713–1720, 2001. Available from : <http://www.stat.cmu.edu/~kass/papers/imi.pdf>. 10
- [7] Yoshihiko Ogata. Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical Association*, 83(401) :9–27, 1988. 10
- [8] D. H. Perkel, G. L. Gerstein, and G. P. Moore. Neuronal spike trains and stochastic point processes. I the single spike train. *Biophys. J.*, 7 :391–418, 1967. Available from : <http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=4292791>. 6
- [9] Wilson Truccolo, Uri T. Eden, Matthew R. Fellows, John P. Donoghue, and Emery N. Brown. A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble and Extrinsic Covariate Effects. *J Neurophysiol*, 93 :1074–1089, February 2005. Available at : <http://jn.physiology.org/cgi/content/full/93/2/1074>. 10
- [10] Henry C. Tuckwell. *Introduction to Theoretical Neurobiology*. Cambridge University Press, 1988. 10
- [11] Lon Turnbull, Emese Dian, and Guenter Gross. The string method of burst identification in neuronal spike trains. *J Neurosci Methods*, 145(1-2) :23–35, Jun 2005. 3
- [12] Grace Wahba. *Spline Models for Observational Data*. SIAM, 1990. 10
- [13] Simon N. Wood. *Generalized Additive Models. An Introduction with R*. Chapman & Hall/CRC, 2006. 10