

Le processus de Poisson comme outil de compréhension de la connectivité fonctionnelle du cerveau

Sahi Gonsangbeu
Yassine Obeïd

Polytech Nice-Sophia

2 mars 2022

Table des matières

- 1 Introduction
- 2 Outils utilisés dans MTGAUE
- 3 Focus sur les processus de Poisson homogène
- 4 Simulation de processus de Poisson homogène
- 5 Test sur nos données pour savoir si elles peuvent être modélisées par un processus de Poisson homogène

Introduction

Motivation

- Un enjeu majeur en neurosciences est d'identifier les connectivités fonctionnelles pouvant exister entre neurones au cours du temps.
- Les Processus de Poisson sont très utilisés à cet effet.

Introduction

Processus de Poisson

- Les processus de Poissons sont des processus stochastiques ponctuels
- Ils permettent de modéliser plusieurs types de phénomènes aléatoires dans tous les domaines.

Outils utilisés dans MTGAUE

Outils utilisés dans l'article MTGAUE

- La notion de coincidence par "Mutiple shift"
- Tests d'hypothèses, Tests multiples (procédure de Benjamini et Hochberg), False Discovery Rate (FDR)
- Processus de Poisson homogène

Outils utilisés dans MTGAUE

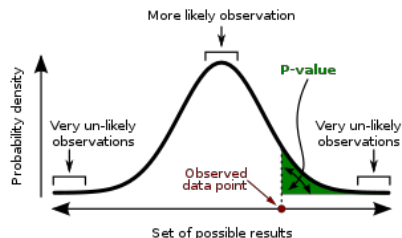
La notion de coincidence par "Multiple shift"

- Train de spike :
séquence de 0 et 1
notée $(H_n)_n$ où $H_i = 1$
traduit la présence d'un
spike sur l'intervalle
 $[ih - \frac{h}{2}, ih + \frac{h}{2}]$ où h est
résolution d'acquisition
des données (de l'ordre
de $10^{-3}, 10^{-4}$)
- Notion de coincidence
par "Multiple shift" :
Coincidence au temps
 ih sur la fenêtre W s'il
existe un "shift" j dans
 $\{-d, \dots, d\}$ où d est un
entier supérieur ou égal
à 1, tel que $H_i^1 = H_{i+j}^2 = 1$
avec $1 \leq i+j \leq n$

Outils utilisés dans MTGAUE

Tests multiples - Procédure de Benjamini-Höschberg

- Test d'hypothèse : procédure de décision entre deux hypothèses consistant à rejeter ou pas une hypothèse, appelée hypothèse nulle en fonction d'un échantillon de données.
- p-value : probabilité pour un modèle statistique donnée sous l'hypothèse nulle H_0 , d'obtenir la même valeur ou une valeur encore plus extrême que celle observée.



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

Outils utilisés dans MTGAUE

Tests multiples - Procédure de Benjamini-Höcherberg

- Principe : Corriger la liste des p-values établies lorsqu'on réalise un test sur plusieurs variables (ou lorsqu'on réalise plusieurs fois un même test) afin de rattraper les hypothèses H_0 rejetées à tort.
- Types de tests multiples basées sur deux taux d'erreur de type I :
 - Family Wise Error Rate : Estimation de la probabilité d'avoir au moins un faux positif.
 - False Discovery Rate : Taux moyen de faux positifs

NB : FP représente le nombre de faux positifs (le nombre de fois où on rejette H_0 alors que celle ci est vraie)

Outils utilisés dans MTGAUE

Tests multiples - Procédure de Benjamini-Höcherberg

- procédure de Benjamini-Höcherberg inventée par Yoav Benjamini et Yosef Höcherberg en 1995.
- Principe : Rejet de H_0^1, \dots, H_0^k pour le plus grand k tel que $p_k \leq \frac{k}{K} \cdot \alpha \Rightarrow \alpha \geq \frac{K}{k} \cdot p_k$. Où p_k représente la p-valeur associée à la statistique de test calculée, α le seuil de significativité fixé au préalable et K le nombre de tests.

Outils utilisés dans MTGAUE

Tests multiples - Procédure de Benjamini-Höschberg

Algorithme associé à la procédure :

- Ranger les p par ordre croissant
- $p_{K*} = p_K$
- $\forall k \in [K - 1, \dots, 1], p_{k*} = \min(p_{k+1*}, \frac{k}{K} \cdot p_k)$
- $p_* = [p_{1*}, \dots, p_{K*}]$

où les p_* sont les p-values corrigées.

Focus sur les processus de Poisson homogène

Processus de comptage

Definition

Désignons par $N(t)$ le nombre d'évènements se produisant dans l'intervalle de temps $[0, t]$, et supposons que $N(0) = 0$. Le processus $(N(t))_{t \geq 0}$ est appelé processus de comptage et vérifie :

- $\forall t \geq 0, N(t) \in \mathbb{N}$;
- $t \rightarrow N(t)$ est croissante ;
- $\forall 0 < a < b, N(b) - N(a)$ représente le nombre d'évènement se produisant dans l'intervalle de temps $]a, b]$.

Focus sur les processus de Poisson homogène

Processus de comptage

Theorem

Soit $(N(t))_{t \geq 0}$ un processus de comptage.

On a :

Si $a < b$ alors $N(a) \leq N(b)$

Focus sur les processus de Poisson homogène

Définition d'un processus de Poisson homogène

Definition

Un processus de comptage $\{N(t), t \geq 0\}$ est appelé processus de Poisson d'intensité $\lambda > 0$ si :

- $N(0) = 0$
- Le processus est à accroissements indépendants ie :
 $\forall 0 \leq t_1 \leq t_2 \cdots < t_n$, les variables aléatoires $N(t_i) - N(t_{i-1}), i = 1, \cdots, n$ sont globalement indépendantes ;
- Le nombre de tops se produisant dans un intervalle de temps de longueur $t \geq 0$ suit une loi de Poisson de paramètre λt , ie : $\forall h \geq 0, \forall t \geq 0, \forall n \in \mathbb{N}$,
$$\mathbb{P}(N(t+h) - N(t) = n) = \exp(-\lambda t) \frac{(\lambda t)^n}{n!}$$

Focus sur les processus de Poisson homogène

Temps d'inter-arrivées

Definition

Considérons un processus de Poisson $(N(t))_t$. Le premier saut arrive à un instant aléatoire T_1 , puis il faut attendre un temps T_2 avant le second saut, puis un temps T_3 ainsi de suite ... On note T_n le temps écoulé entre le $(n-1)$ -ième saut et le n -ième saut.

La suite des instants $(T_n)_{n \geq 1}$ est appelée suite des instants inter-arrivées.

Focus sur les processus de Poisson homogène

Temps d'inter-arrivées

Theorem

La suite des instants inter-arrivées $(T_n)_{n \geq 1}$ est une suite de variables aléatoires iid de loi exponentielle de paramètre λ .

Focus sur les processus de Poisson homogène

Définition du processus de Poisson à l'aide des temps d'inter-arrivées

Definition

On considère $(T_n)_{n \geq 1}$ une suite de variables aléatoires iid de loi exponentielle de paramètre λ . On définit $S_0 = 0$ et on pose

$$S_n = \sum_{i=1}^n T_i$$

Alors le processus $(N(t))_{t \geq 0}$ définit par

$$N(t) = \sum_{n=1}^{\infty} \mathbf{1}_{S_n \leq t} = \max\{n \geq 0 : S_n \leq t\}$$

est un processus de Poisson d'intensité λ .

Simulation de processus de Poisson homogène

Procédure

- Générer un vecteur de variables aléatoires indépendantes de loi exponentielle de paramètre λ . Il s'agit du vecteur des temps d'inter-arrivées $(T_n)_{n \geq 1}$
- on génère le vecteur des occurrences $(t_k)_{k \geq 1}$ par la formule :

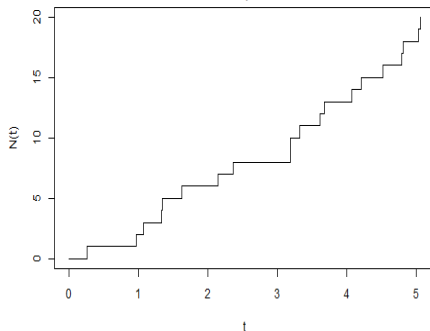
$$t_1 = T_1$$

$$t_k = t_{k-1} + T_k, \forall k > 1$$

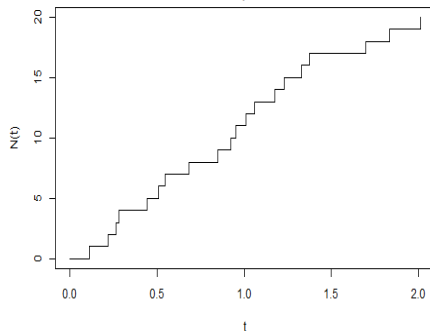
Simulation de processus de Poisson homogène

Résultats

Processus de Poisson de paramètre 5 et de taille 20



Processus de Poisson de paramètre 10 et de taille 20



└ Test sur nos données pour savoir si elles peuvent être modélisées par un processus de Poisson homogène

Test sur nos données pour savoir si elles peuvent être modélisées par un processus de Poisson homogène

Test de Kolmogorov-Smirnov

Principe du test :

- On teste l'hypothèse $H_0 : "F = F_0"$ contre l'hypothèse alternative $"F \neq F_0"$ au niveau α . α est notre seuil de significativité (On choisi 0.05)
- On accepte H_0 si :

$$\forall t \in R, |F_n(t) - F_0(t)| \leq \xi_{n,1-\alpha}$$

où $\xi_{n,1-\alpha}$ est le quantile d'ordre $1 - \alpha$ extrait de la table de Kolmogorov-Smirnov.

F_n est la fonction de répartition empirique

- └ Test sur nos données pour savoir si elles peuvent être modélisées par un processus de Poisson homogène

Test sur nos données pour savoir si elles peuvent être modélisées par un processus de Poisson homogène

Mise en oeuvre sur nos données avec le logiciel R

```
# Importation de nos données
```

```
data1 <- read.table("donnees/data1.txt")
```

```
data2 <- read.table("donnees/data2.txt")
```

```
data3 <- read.table("donnees/data3.txt")
```

```
data <- rbind(data1, data2)
```

```
# Concaténation de toutes les données ensemble
```

```
data <- rbind(data,data3)
```

```
data <- data$V1
```

└ Test sur nos données pour savoir si elles peuvent être modélisées par un processus de Poisson homogène

Test sur nos données pour savoir si elles peuvent être modélisées par un processus de Poisson homogène

Mise en oeuvre sur nos données avec le logiciel R

```
# Calcul des écarts
```

```
t <- diff(data)
```

```
lambda_t <- 1./mean(t)
```

```
# test de Kolmogorov-Smirnov
```

```
ks.test(t, "pexp", lambda_t) # ici on teste si les écarts
```

```
#entre nos données suivent bien une
```

```
#loi exponentielle de paramètre lambda_t
```

└ Test sur nos données pour savoir si elles peuvent être modélisées par un processus de Poisson homogène

Test sur nos données pour savoir si elles peuvent être modélisées par un processus de Poisson homogène

Mise en oeuvre sur nos données avec le logiciel R

Résultat du test :

One-sample Kolmogorov-Smirnov test

data : t

$D = 0.37095$, $p\text{-value} < 2.2e-16$

alternative hypothesis : two-sided

On obtient une p-value très négligeable. Donc nos données ne suivent pas un processus de Poisson homogène.

- └ Test sur nos données pour savoir si elles peuvent être modélisées par un processus de Poisson homogène

Merci de votre attention