Tests d'Indépendance par des Méthodes de Permutation et de Bootstrap et Applications en Neurosciences

Laboratoire de Mathématiques J.-A. Dieudonné, Université de Nice Sophia-Antipolis
Stage de Master 2 - Année 2011-2012

Mélisande Albert

Ce stage a été réalisé sous la direction de :

- Patricia Reynaud-Bouret (Laboratoire J.-A. Dieudonné, Université de Nice Sophia-Antipolis)
- Magalie Fromont (IRMAR / CREST-ENSAI, Rennes)

Je remercie le Laboratoire Jean-Alexandre Dieudonné de Nice Sophia-Antipolis et le Laboratoire de Statistique et de Modélisation du CREST (Rennes) de m'avoir permis de faire ce stage et de m'avoir si bien accueillie. Je tiens à remercier tout particulièrement Patricia Reynaud-Bouret et Magalie Fromont de m'avoir si bien encadrée tout au long de ce stage de Master 2.

Table des matières

1 Contexte scientifique **1**
1.1 Aspects biologiques . . . 1
1.1.1 Neurones et potentiels d'action . . . 1
1.1.2 Problématique . . . 1
1.1.3 Protocole expérimental . . . 1
1.2 Aspects mathématiques . . . 2
1.2.1 Travaux antérieurs . . . 2
1.2.2 Objectif du stage . . . 3

2 Tests d'indépendance et approches de bootstrap **3**
2.1 Tests d'indépendance . . . 3
2.2 Tests par permutation : un concept précurseur . . . 3
2.3 De la permutation au bootstrap . . . 4
2.4 Approches récentes . . . 8

3 Un nouveau test d'indépendance par permutation **8**
3.1 Problème statistique . . . 9
3.2 Notations et description du test . . . 9
3.3 Risque de première espèce du test . . . 10
3.4 Démonstration des Théorèmes 3 et 4 . . . 11
3.4.1 L'assemblage est uniforme . . . 11
3.4.2 Composition avec une permutation aléatoire . . . 14
3.5 Estimation par Monte Carlo . . . 15
3.5.1 Test avec estimation par Monte Carlo . . . 15
3.5.2 Monte Carlo d'un point de vue non-asymptotique . . . 15

4 Vers des tests d'indépendance bootstrap **17**
4.1 De la permutation au bootstrap d'Efron . . . 17
4.2 Du bootstrap d'Efron au bootstrap sauvage . . . 18

5 Étude de simulations **19**
5.1 Simulation des processus de Poisson homogènes . . . 19
5.2 Étude des fonctions de répartition empiriques . . . 19
5.3 Programmation des tests et estimation des niveaux . . . 22

6 Conclusion et Perspectives **23**

A Tests statistiques ; quelques rappels **25**

B Rappels sur les inverses généralisés **26**

C Rappels sur les fonctions de répartition et quantiles empiriques **26**

D Rappels sur les processus de Poisson homogènes **27**

E Rappels sur la distance de Wasserstein **29**

F Programmes sous Scilab **31**
F.1 Simulation des données . . . 31
F.2 Calcul des coïncidences . . . 31
F.3 Création de l'échantillon de la statistique calculée sur les permutées . . . 33
F.4 Calcul de la p-valeur . . . 33
F.5 Test par permutation et niveau empirique . . . 33

1 Contexte scientifique

1.1 Aspects biologiques

1.1.1 Neurones et potentiels d'action

Le néocortex constitue un réseau fortement connecté de plusieurs milliards de neurones. Ceux-ci communiquent entre eux grâce aux potentiels d'action :

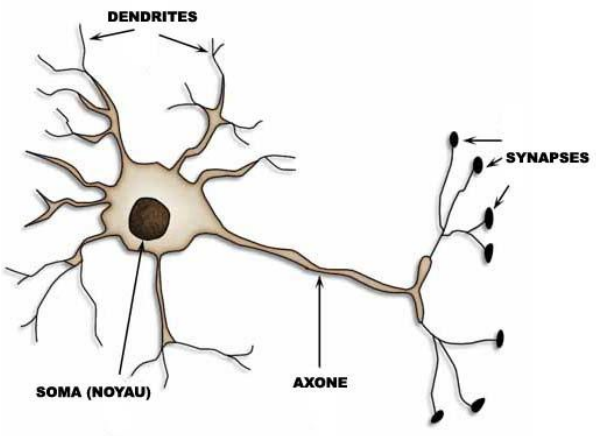

FIGURE 1 – Schéma type d'une cellule nerveuse.

Des signaux d'entrée se déplacent le long des dendrites, et si leur somme dépasse un certain seuil, alors un potentiel d'action se déclenche et se propage (sans perte d'amplitude) le long de l'axone. Lorsqu'il arrive au niveau des synapses chimiques, il libère les neurotransmetteurs qui iront se fixer sur les dendrites des neurones voisins.

Les potentiels d'action sont des phénomènes de type "tout ou rien", qu'on appellera spikes. On prendra donc comme modèle des processus ponctuels.

1.1.2 Problématique

On sait que la fréquence des potentiels d'action dépend de l'intensité de la stimulation membranaire. Un stimulus de forte intensité déclenche un grand nombre de potentiels d'action par unité de temps, ce qui va par exemple provoquer une réaction musculaire.

Par ailleurs, les neurones sont très fortement connectés (environ 10 000 à 40 000 connexions par neurone). La question naturelle est donc de se demander si la synchronisation de l'occurence des potentiels d'action entre les neurones joue un rôle dans la transmission du message nerveux, ou si au contraire ce n'est qu'une conséquence de la haute fréquence des potentiels d'action. Il a été montré par Grün [9] que ces coïncidences ne sont pas toujours dues au hasard, mais parfois à une dépendance ponctuelle entre neurones, ceci en fonction de la tâche effectuée. On se fixe donc ici comme objectif de pouvoir détecter les ruptures d'indépendance.

1.1.3 Protocole expérimental

Un singe (formé pour ce genre d'expérience) est en position assise sur une chaise en face d'un panneau vertical composé de sept touches lumineuses disposées de la manière suivante : six touches forment un cercle dont la septième est au centre.

Au début, la cible centrale s'allume, et l'expérience débute dès que le singe la touche. Après un délai fixe de 500 ms, une ou plusieurs touches périphériques s'allument en vert ; c'est le signal préparatoire (PS). Il indique à l'animal plus ou moins sur quelle touche il devra pointer. Ensuite, après une période préparatoire (PP) de 1s, la cible verte (ou une des cibles allumées en vert) devient rouge ; c'est le signal réponse. Il indique au singe quelle touche il doit pointer.

Chaque expérience donne lieu à un enregistrement temporel de l'activité neuronale se traduisant par l'enregistrement d'une séquence de potentiels d'action, simultanément sur quelques neurones. Pendant

une session de 150 essais environ, chacune des six directions de mouvement est présentée aléatoirement avec égale probabilité.

Au final, les données sont $\{(N_1^k, \cdots, N_m^k)_{1\leq k\leq n}\}$ où m représente le nombre de neurones enregistrés, n le nombre de fois que l'expérience a été réalisée et N_i^k représente l'activité du neurone i au cours de l'essai k (pour tous i et k, N_i^k est à valeurs dans $\mathbb{X}$, l'ensemble des sous-ensembles finis de $[a, b]$ contenant l'emplacement des spikes correspondants).

1.2 Aspects mathématiques

On considère deux neurones, par exemple le premier, et le deuxième. On suppose que les (N_1^k, N_2^k) sont indépendants et identiquement distribués (i.i.d. de même loi que (N_1, N_2). On souhaite répondre à la question suivante :

Peut-on considérer que les activités neuronales N_1 et N_2 associées à deux neurones distincts sont indépendantes ?

Mathématiquement, on cherche donc à contruire le test d'indépendance suivant :

$$\begin{cases} (H_0) & : & N_1 \text{ et } N_2 \text{ sont indépendants.} \\ (H_1) & : & \text{ils ne le sont pas.} \end{cases}$$

Des notions de base sur les tests statistiques sont rappelées en Annexe [A].

Dans des travaux antérieurs, des résultats ont déjà été obtenus sous des hypothèses plus ou moins réalistes :

($\mathcal{H}_{PH}$) L'activité neuronale peut être modélisée par un processus de Poisson homogène (un spike est associé à un point du processus), *i.e.* N_1 et N_2 sont des processus de Poisson homogènes.

($\mathcal{H}_{IID}$) Les essais successifs sont indépendants et identiquement distribués, *i.e.* $\mathbf{N} = (N_1^1, N_2^1), \ldots, (N_1^n, N_2^n)$ forme un n-échantillon de la loi de (N_1, N_2).

On supposera dans ce travail l'hypothèse ($\mathcal{H}_{IID}$) vérifiée, bien que ce ne soit pas complètement réaliste sur le plan physiologique.

Par contre, un des buts de ce travail est de s'affranchir de l'hypothèse de stationnarité des processus de Poisson, voire de l'hypothèse Poissonnienne. On aura donc comme objectif de remplacer ($\mathcal{H}_{PH}$) par :

($\mathcal{H}_{PP}$) N_1 et N_2 sont des processus ponctuels quelconques.

1.2.1 Travaux antérieurs

Sous les hypothèses ($\mathcal{H}_{PH}$) et ($\mathcal{H}_{IID}$), plusieurs tests ont déjà été mis au point par Christine Tuleau-Malot, Amel Rouis, Patricia Reynaud-Bouret et Franck Grammont (travail en cours de soumission).

Notations et définitions.

Définition.

> Étant donné $\delta \in]0, b-a[$, on parle de coïncidence lorsqu'un spike du premier neurone est à distance inférieure à δ d'un spike du second neurone.

Dans toute la suite, on prendra :

$$0 < \delta < \frac{b-a}{2}.$$

On note :

- δ le pas de temps servant à définir les coïncidences,
- $X^k = \int_a^b \int_a^b \mathbb{1}_{\{|x-y|\leq\delta\}} dN_1^k(x) dN_2^k(y)$ la variable aléatoire comptant le nombre de coïncidences entre les deux neurones sur la plage [a,b],
- $\bar{X} = \frac{1}{n}\sum_{k=1}^n X^k$ la moyenne empirique du nombre de coïncidences entre les neurones 1 et 2,
- $Y_i^k = N_i^k[a, b]$ le nombre de spikes sur l'intervalle $[a, b]$ pour le neurone i au $k^{\text{ème}}$ essai,
- $\bar{Y}_i = \frac{1}{n}\sum_{k=1}^n Y_i^k$ la moyenne empirique du nombre de spikes du neurone i,
- $\hat{\lambda}_i = \frac{\bar{Y}_i}{b-a}$ l'estimateur du l'intensité du processus de Poisson du neurone i.

Statistique de test. On définit la statistique de test suivante :

$$T_n = \frac{\sqrt{n}\left(\bar{X} - (2\delta(b-a) - \delta^2)\hat{\lambda}_1\hat{\lambda}_2\right)}{\sqrt{\hat{\lambda}_1\hat{\lambda}_2[2\delta(b-a) - \delta^2 + (\frac{2}{3}\delta^3 - \frac{\delta^4}{b-a})(\hat{\lambda}_1 + \hat{\lambda}_2)]}}.$$

Loi de la statistique sous (H_0).

Théorème 1 ([15]).
Si $(\mathcal{H}_{PH})$ et $(\mathcal{H}_{IID})$ sont vérifiées, alors sous l'hypothèse (H_0), T_n converge en loi vers une loi gaussienne centrée réduite lorsque n tend vers $+\infty$.

L'idée principale de la démonstration repose sur l'application d'une méthode Delta.

Test d'indépendance asymptotique. Étant donné $\alpha \in]0,1[$, on considère donc le test dont la région de rejet est définie par $|T_n| > z_{1-\frac{\alpha}{2}}$, où $z_{1-\frac{\alpha}{2}}$ est le $1-\frac{\alpha}{2}$ quantile de la loi $\mathcal{N}(0,1)$. Le test ainsi défini est asymptotiquement de niveau α, *i.e.*

$$\mathbb{P}_{(H_0)}\left(|T_n| > z_{1-\frac{\alpha}{2}}\right) \xrightarrow[n\to+\infty]{} \alpha.$$

1.2.2 Objectif du stage

Le test construit dans [15] se base donc entièrement sur l'hypothèse $(\mathcal{H}_{PH})$ *i.e.* "N_1 et N_2 sont des processus de Poisson homogènes". En outre, le résultat obtenu est asymptotique sur le nombre d'essais, alors que dans la pratique, l'expérience ne peut être réalisée qu'un nombre limité de fois pour des raisons de temps et de coût. De plus, lorsqu'on ne considère que les essais effectués dans une même direction, le nombre de données traitables peut chuter à une vingtaine. L'objectif que nous nous fixons est donc finalement double. Nous chercherons non seulement à nous affranchir de l'hypothèse Poissonnienne mais également à obtenir un résultat non asymptotique. L'approche qui s'impose est celle des tests par permutation ou plus généralement par bootstrap.

2 Tests d'indépendance et approches de bootstrap

2.1 Tests d'indépendance

L'indépendance figurant parmi les hypothèses de nombreux théorèmes, et étant souvent nécessaire dans des raisonnements probabilistes ou statistiques, il est fréquent qu'on ait besoin d'effectuer un test d'indépendance. Depuis les méthodes proposées par Pearson ou Hœffding dans [10] par exemple, de nombreuses statistiques de tests ont été introduites.

La plupart du temps, elles sont construites de sorte que leur loi sous l'hypothèse d'indépendance est connue ou asymptotiquement connue (c'est à dire qu'elles ne dépendent pas de la distribution des données). Cependant, dans certains cas, la loi asymptotique est éloignée de la vraie loi sous (H_0), voire elle peut dépendre de la distribution sous-jacente.

Afin de pallier à ces problèmes, de nouvelles méthodes telles que les tests par permutation ou par bootstrap ont été mises au point. Le but était de pouvoir explorer un jeu de données sans avoir à faire de suppositions sur leur distribution sous-jacente.

2.2 Tests par permutation : un concept précurseur

Les tests par permutation sont apparus en 1935, sous l'initiative de R.A. Fisher [6], avec E.J.G. Pitman et B.L. Welch et ont été largement étudiés depuis. Le principe de ces tests est de travailler conditionnellement aux données observées, de façon à eviter de formuler des hypothèses sur la distribution de la variable observée.

Étant donnée une statistique de test, on la calcule sur les données observées. Ensuite on permute ces dernières aléatoirement de façon à obtenir un nouvel échantillon sur lequel on recalcule la statistique de test. On confronte ensuite la statistique calculée sur les données observées à la distribution empirique des statistiques calculées sur les permutations. La validité de telles méthodes vient du fait que sous

l'hypothèse nulle, toute permutation des données observées est "également probable". Cette méthode sera mieux détaillée dans la suite, dans le cas particulier d'une statistique de test basée sur les coïncidences.

2.3 De la permutation au bootstrap

Cette idée de travailler conditionnellement aux données, d'abord développée pour des tests, a ensuite été généralisée à d'autres domaines statistiques tels que l'estimation ou la construction d'intervalles de confiance. Dans la lignée de ces travaux, M.H. Quenouille a introduit le Jackknife en 1948 [13] afin d'estimer le biais d'estimateurs sans formuler d'hypothèses sur la loi. Le bootstrap a ensuite été introduit par B. Efron à la fin des années 70 [5] comme généralisation et amélioration de ces méthodes.

L'idée originelle du bootstrap. On se donne un n-échantillon i.i.d. $X = (X_1, \ldots, X_n)$ de loi P inconnue. On s'intéresse à la loi d'une racine $R_n = R_n(X_1, \ldots, X_n; P)$, par exemple la différence entre la moyenne empirique et la vraie moyenne. L'idée est d'estimer la loi de R_n en remplaçant dans son expression la loi P par la mesure empirique $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ associée à l'échantillon X, à tous les niveaux. On estime donc la loi de R_n par celle de $R_n^* = R_n(X^*; P_n)$ conditionnellement à X, où $X^* = (X_1^*, \ldots, X_n^*)$ est un n-échantillon de la loi P_n, appelé échantillon bootstrap.

L'heuristique d'Efron est que la loi conditionnelle de $(R_n^* \mid X)$ est proche de celle de R_n. La justification est ce qui constitue la vraie difficulté. Il existe des théorèmes bootstrap "types" (c.f. par exemple M.A. Arcones et E. Giné, [1]) selon lesquels la loi conditionnelle converge vers la distribution asymptotique de R_n presque sûrement ou en probabilité. On cherchera dans ce mémoire à obtenir également des résultats non asymptotiques.

D'un point de vue pratique, étant donné que $\forall i, j \in \{1, \ldots, n\},\ \mathbb{P}\left(X_j^* = X_i \mid X\right) = 1/n$, on remarque que X^* est un tirage uniforme avec remise sur X. L'avantage ici est qu'on peut facilement simuler R_n^*. Les statisticiens approchent donc généralement les caractéristiques de la loi de la racine qui les intéressent par des méthodes de type Monte Carlo plutôt que de les calculer.

Exemple du bootstrap de la moyenne. Pour mieux comprendre cette approche, regardons un exemple simple détaillé par P.J. Bickel et D.A. Freedman dans [3].

Soit $X_1, \ldots, X_n$ des variables aléatoires réelles i.i.d. de loi P, avec P mesure de probabilité de moyenne μ inconnue et de variance σ^2 toutes deux supposées bien définies. On estime alors μ par la moyenne empirique usuelle $\mu_n = \frac{1}{n}\sum_{i=1}^n X_i$. On aimerait connaître l'erreur d'approximation que l'on fait en approchant μ par μ_n.

Généralement, on regarde le comportement de la racine

$$R_n = \sqrt{n}\,(\mu_n - \mu)\,.$$

On peut appliquer le théorème central limite, et on obtient

$$R_n \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Ainsi, dans ce cas particulier, on connaît la loi asymptotique de la racine. Néanmoins, il est intéressant d'un point de vue théorique de voir comment le bootstrap fonctionnerait ici.

Comme nous l'avons expliqué ci-dessus, l'approche bootstrap consiste à considérer $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ la mesure empirique de notre échantillon. L'étape suivante est de ré-échantillonner les données en prenant le m-échantillon $X_1^*, \ldots, X_m^*$ i.i.d. conditionnellement à $(X_1, \ldots, X_n)$, et de loi P_n. On estime donc notre racine par

$$R_m^* = \sqrt{m}\,(\mu_m^* - \mu_n) \quad \text{où} \quad \mu_m^* = \frac{1}{m}\sum_{j=1}^m X_j^*.$$

appelée racine bootstrappée.

- Prenons d'abord $m = n$. L'idée est que la racine bootstrappée $R_n^*|X$ se comporte comme R_n. Ainsi, on pourra la calculer à partir des données, et ensuite l'utiliser pour approcher la distribution inconnue de R_n. Cette approche mènerait à un intervalle de confiance pour μ même dans le cas où σ^2 est inconnue. Reste à démontrer que cette approximation par bootstrap est valide.

- L'autre cas est lorsque $m \neq n$. La taille du ré-échantillonnage a de l'importance d'un point de vue statistique, comme on le verra dans la preuve du théorème. Ils peuvent très bien varier de façon complètement indépendante. Mais, on voit que d'un point de vue mathématique, il y a un passage plutôt délicat lorsque $m = n$. Lorsqu'on compare la loi classique $\sqrt{n}\,(\mu_n - \mu)$ avec la loi bootstrappée $\sqrt{n}\,(\mu_n^* - \mu_n)$, on a une erreur due au fait que l'on remplace le paramêtre μ par μ_n. Mais ce changement est en fait compensé par l'erreur faite en remplaçant les X_i par les X_j^*.

Énonçons à présent le théorème justifiant l'approche bootstrap.

Théorème 2 ([3])**.**
Soit $X_1, \ldots, X_n$ des variables aléatoires réelles i.i.d. de loi P de moyenne μ et de variance $\sigma^2 < \infty$. Alors, presque sûrement,

$$\sqrt{m}\,(\mu_m^* - \mu_n) \;\mid (X_1, \ldots, X_n) \xrightarrow[n,m\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Pour démontrer ce théorème, on aura besoin d'introduire de nouvelles notations. On note

$$\Gamma_2 = \left\{ Q \in \mathcal{P}(\mathbb{R}) / \int_{\mathbb{R}} x^2 dQ(x) < \infty \right\}.$$

On définit alors sur Γ_2 la notion de convergence suivante :

$$Q_\alpha \Rightarrow Q \quad \underline{ssi} \quad Q_\alpha \rightharpoonup Q \;\text{ et }\; \int x^2 dQ_\alpha(x) \to \int x^2 dQ(x).$$

où " $\rightharpoonup$ " est la convergence faible * sur l'ensemble des mesures de probabilité sur $\mathbb{R}$. Notons que cette convergence n'est en fait rien d'autre que la convergence en loi.

Par le théorème de Glivenko-Cantelli (c.f. Théorème 6 en Annexe C) et la loi des grands nombres, on peut montrer que pour presque tout échantillon, la mesure empirique associée $P_n \Rightarrow P$, car $P \in \Gamma_2$ ($\sigma^2 < \infty$).

1ère étape : Montrons d'abord la convergence faible. Cela revient à prouver une convergence en loi, et donc la convergence des fonctions de répartitions. On veut montrer que pour presque tout échantillon $X_1, \ldots, X_n$ i.i.d. de distribution P, la mesure empirique P_n associée converge en loi vers P. Or la fonction de répartition de la mesure empirique n'est autre que la fonction de répartition empirique. On aimerait donc montrer que

$$p.s, \quad \forall t \in \mathbb{R},\; F_n(t) \xrightarrow[n\to+\infty]{} F(t).$$

Or ceci est directement obtenu par le théorème de Glivenko-Cantelli. Ainsi, on obtient bien que, pour presque tout $(X_1, \ldots, X_n)$ i.i.d. $\sim P$,

$$P_n \xrightarrow[n\to+\infty]{\mathcal{L}} P$$

d'où la convergence faible.

2ème étape : Pour montrer la convergence des moments d'ordre 2, on utilise cette fois-ci la loi des grands nombres.

$$\begin{aligned}
\int x^2 dP_n(x) &= \int x^2 \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(x) \\
&= \frac{1}{n} \sum_{i=1}^{n} X_i^2 \\
\text{(LGN)} \quad &\xrightarrow[n\to+\infty]{p.s.} \mathbb{E}\left[\, X_1^2 \,\right] \\
&= \int x^2 dP(x)
\end{aligned}$$

On peut l'appliquer ici car comme on a supposé que P est dans Γ_2, $\mathbb{E}\left[\, |X_i|^2 \,\right] < \infty$.

On a donc bien montré quelle que soit la loi $P \in \Gamma_2$, sa mesure empirique associée converge pour "$\Rightarrow$" vers P.

D'autre part, on peut mettre une distance sur cette notion de convergence. En effet, par la proposition 6 de l'Annexe E,

$$Q_\alpha \Longrightarrow Q \quad \underline{ssi} \quad d_2(Q_\alpha, Q) \to 0,$$

où d_2 désigne la distance de Wasserstein.

Soit maintenant $Z_1(Q), \ldots, Z_m(Q)$ des variables aléatoires i.i.d. de loi Q. Et soit $Q^{(m)}$ la distribution de

$$S_m(Q) = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \left[Z_j(Q) - \mathbb{E}\left[\ Z_j(Q)\ \right] \right].$$

En développant le carré, on voit bien que si $Q \in \Gamma_2$, alors $Q^{(m)} \in \Gamma_2$.

Commençons par prouver le lemme suivant :

Lemme 1.
Soit $Q, R \in \Gamma_2$. Avec les notations ci-dessus, on a

$$d_2(Q^{(m)}, R^{(m)}) \leq d_2(Q, R).$$

Preuve : Notons $\mu_Q = \mathbb{E}\left[\ Z_j(Q)\ \right]$ et $\mu_R = \mathbb{E}\left[\ Z_j(R)\ \right]$. On a

$$\begin{aligned}
\left[d_2(Q^{(m)}, R^{(m)})\right]^2 &= \left[d_2(S_m(Q), S_m(R))\right]^2 \\
&= \left[d_2 \left(\frac{1}{\sqrt{m}} \sum_{j=1}^{m} \left[Z_j(Q) - \mu_Q\right], \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \left[Z_j(R) - \mu_R\right] \right) \right]^2 \\
\text{(Propriété 3, Annexe E)} &= \frac{1}{m} \left[d_2 \left(\sum_{j=1}^{m} \underbrace{\left[Z_j(Q) - \mu_Q\right]}_{U_j}, \sum_{j=1}^{m} \underbrace{\left[Z_j(R) - \mu_R\right]}_{V_j} \right) \right]^2 \\
\text{(Propriété 4, Annexe E)} &\leq \frac{1}{m} \sum_{j=1}^{m} \left[d_2 \left(Z_j(Q) - \mu_Q, Z_j(R) - \mu_R \right) \right]^2 \\
&= \left[d_2 \left(Z_1(Q) - \mu_Q, Z_1(R) - \mu_R \right) \right]^2
\end{aligned}$$

En outre, on remarque que

$$\begin{aligned}
d_2^2(Z_1(Q), Z_1(R)) &= \inf_{X \sim Z_1(Q), Y \sim Z_1(R)} \mathbb{E}\left[\ |X - Y|^2\ \right] \\
&= \inf_{U \sim Z_1(Q) - \mu_Q, Y \sim Z_1(R) - \mu_R} \mathbb{E}\left[\ |(U - V) + (\mu_Q - \mu_R)|^2\ \right] \\
&= \inf_{U \sim Z_1(Q) - \mu_Q, Y \sim Z_1(R) - \mu_R} \left\{ \mathbb{E}\left[\ (U - V)^2\ \right] + 2(\mu_Q - \mu_R) \underbrace{\mathbb{E}\left[\ U - V\ \right]}_{0} + (\mu_Q - \mu_R)^2 \right\} \\
&= \inf_{U \sim Z_1(Q) - \mu_Q, Y \sim Z_1(R) - \mu_R} \left\{ \mathbb{E}\left[\ (U - V)^2\ \right] \right\} + (\mu_Q - \mu_R)^2 \\
&= d_2^2(Z_1(Q) - \mu_Q, Z_1(R) - \mu_R) + (\mu_Q - \mu_R)^2 \\
&\geq d_2^2(Z_1(Q) - \mu_Q, Z_1(R) - \mu_R)
\end{aligned}$$

En remplaçant dans l'inégalité précédente, on obtient bien que

$$\left[d_2(Q^{(m)}, R^{(m)})\right]^2 \leq \left[d_2(Z_1(Q), Z_1(R))\right]^2,$$

d'où le lemme. □

On peut donc à présent démontrer le Théorème 2.
Démonstration : Avec les notations introduites ci-dessus, on remarque que la distribution de

$$\sqrt{m}\,(\mu_m^* - \mu_n)\,|X = S_m(P_n)|X$$

est justement $P_n^{(m)}$. Et de même, on voit bien que $P^{(m)}$ est la loi de $\sqrt{m}\,(\mu_m - \mu) = S_m(P)$.

D'un autre côté, on a vu que $P_n \Rightarrow P$ d'où, par la proposition 6, que cela équivaut à $d_2(P_n, P) \to 0$. Ainsi,

$$\forall \varepsilon,\ \exists n_\varepsilon \in \mathbb{N}\ /\ \forall n \geq n_\epsilon,\ d_2(P_n, P) \leq \varepsilon.$$

On en déduit par le lemme 1 que,

$$\forall m \in \mathbb{N}, \forall n \geq n_\varepsilon,\ d_2(P_n^{(m)}, P^{(m)}) \leq d_2(P_n, P) \leq \varepsilon.$$

Ce résultat montre bien que les lois $\sqrt{m}\,(\mu_m^* - \mu_n)$ et $\sqrt{m}\,(\mu_m - \mu)$ sont proches si la taille de l'échantillon initial n est assez grande, et ce quel que soit m. Ainsi l'erreur qu'on fait en approchant la moyenne empirique μ_n par la bootstrappée μ_m^* est plus petite que celle de l'approximation de la vraie moyenne μ par celle empirique pour un échantillon de même taille m que le bootstrappé.
On remarque ici que n_ε ne dépend pas de m et le résultat reste valable dans le cas $m = n$.

Terminons à présent la démonstration.
Comme la convergence pour la distance de Wasserstein implique la convergence faible, on déduit que

$$P_n^{(m)} \xrightarrow[n\to+\infty]{\mathcal{L}} P^{(m)}$$

D'un autre côté, par le théorème central limite, on a

$$P^{(m)} \xrightarrow[m\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

Ainsi, on a les convergences en loi suivantes :

$$\lim_{n,m\to+\infty} P_n^{(m)} = \lim_{m\to+\infty} \left(\lim_{n\to+\infty} P_n^{(m)} \right) = \lim_{m\to+\infty} P^{(m)} = \mathcal{N}(0, \sigma^2)$$

ce qui achève la démonstration. □

Passons maintenant à des définitions plus générales de bootstrap.

Le bootstrap à poids. Si on introduit $\forall i \in \{1, \ldots, n\}$ la variable aléatoire $M_{n,i} = \sum_{j=1}^n \mathbb{1}_{\{X_i^j = X_i\}}$ comptant le nombre d'apparitions de X_i dans l'échantillon bootstrap, on remarque que la mesure empirique associée à l'échantillon bootstrap vérifie $P_n^* = \frac{1}{n}\sum_{j=1}^n \delta_{X_j^*} = \frac{1}{n}\sum_{i=1}^n M_{n,i}\delta_{X_i}$. L'idée (dont les premiers résultats généraux ont été montrés par J.T. Præstgaard et J.A. Wellner [12]) a été de remplacer les $M_{n,i}$ par d'autres poids $W_{n,i}$ échangeables. Diverses approches bootstrap ont vu le jour à partir de cette idée et de nombreux résultats ont été obtenus pour certains poids particuliers. Voici quelques exemples :

- *Poids déterministes : généralisation du Jackknife* (B. Efron (1982), C.F.J. Wu (1987), X. Shi (1991)). On considère un vecteur de $\mathbb{R}^n$ déterministe de poids positifs $w_n = (w_{n,1}, \ldots, w_{n,n})$ tels que

$\sum_{i=1}^n w_{n,i} = n$, et un permutation aléatoire σ_n tirée uniformément sur $\mathfrak{S}_n$, où $\mathfrak{S}_n$ désigne l'ensemble des permutations sur $\{1, 2, \ldots, n\}$. On note $W_{n,i} = w_{n,\sigma_n(i)}$. Ainsi, la mesure empirique associée est

$$P_n^W = \frac{1}{n}\sum_{i=1}^n W_{n,i}\delta_{X_i} = \frac{1}{n}\sum_{i=1}^n w_{n,\sigma_n(i)}\delta_{X_i}.$$

En prenant en particulier $w_n = (0, \underbrace{\frac{n}{n-1}, \ldots, \frac{n}{n-1}}_{n-1 \text{ fois}})$, on retrouve l'idée du Jackknife de M.H. Quenouille.

- *Poids Multinomiaux : bootstrap d'Efron et "m out of n bootstrap"* (Bretagnolle [4]). On remarque que les poids du bootstrap classique d'Efron $M_n = (M_{n,1}, \ldots, M_{n,n})$ suivent une loi multinomiale $\mathcal{M}(n; \frac{1}{n}, \ldots, \frac{1}{n})$. L'idée du "m out of n bootstrap" est alors de prendre comme échantillon bootstrap $X_m^* = (X_1^*, \ldots, X_m^*)$ avec $m < n$ et $\frac{m}{n} \to 0$ et les poids sont donc de la forme $W_m = \frac{n}{m}(M_{m,1}, \ldots, M_{m,m})$ où $(M_{m,1}, \ldots, M_{m,m})$ est de loi multinomiale $\mathcal{M}(m; \frac{1}{n}, \ldots, \frac{1}{n})$. En particulier, on obtient

$$P_n^W = \frac{1}{m}\sum_{i=1}^n M_{m,i}\delta_{X_i} = \frac{1}{n}\sum_{i=1}^n W_{m,i}\delta_{X_i}.$$

  Cette méthode a permis de corriger le bootstrap d'Efron dans des cas pathologiques.

- *Poids indépendants : bootstrap sauvage.* Dans les cas précédents, on impose toujours que $\sum_{i=1}^n W_{n,i} = n$ ce qui force une dépendance entre les poids. L'idée du bootstrap sauvage est de prendre des poids indépendants. Un cas particulier intéressant est le bootstrap d'Efron "Poissonisé". On reprend la même idée que ci-dessus, mais en prenant une taille d'échantillon bootstrap aléatoire. Le principe est de premièrement tirer aléatoirement N suivant une loi de Poisson de paramètre n. Ensuite, conditionnellement à N, on tire un N-échantillon $U = (U_1, \ldots, U_N)$ de loi uniforme $\mathcal{U}(\{1, \ldots, n\})$. Notre échantillon bootstrap est donc donné par $X_k^\star = X_{U_k}$ pour k allant de 1 à N. On condidère les poids $W_i^N = \sum_{k=1}^N \mathbb{1}_{\{U_k=i\}}$ qui représente le nombre de fois qu'est tiré chaque indice. Ainsi,

$$P_n^W = \frac{1}{n}\sum_{i=1}^n W_{n,i}\delta_{X_i} = \frac{1}{n}\sum_{i=1}^n \left(\sum_{k=1}^N \mathbb{1}_{\{U_k=i\}}\right)\delta_{X_i}.$$

  On peut montrer, en utilisant la propriété 2 décrite en Annexe D, que $(W_1^N, \ldots, W_n^N)$ est un n-échantillon i.i.d. de loi $\mathcal{P}(1)$ d'où le bootstrap sauvage. Ce point sera mieux détaillé pour notre cas particulier dans la sous-partie 4.2 de ce mémoire.

2.4 Approches récentes

Le coût algorithmique des tests par permutation étant très important, ils étaient peu utilisés dans la pratique. Mais grâce aux nouvelles technologies, ils sont plus couramment appliqués depuis quelques décennies dans de nombreux domaines en statistique. De plus le bootstrap a été bien étudié sur le plan théorique. Ainsi, une analyse détaillée de tests d'indépendance par permutation ou bootstrap plus récente a été faite par J.P. Romano [14] en 1989, ainsi que M.A. Arcones et E. Giné [1] en 1992 pour des théorèmes plus généraux ou A. Janssen et T. Pauls [11] en 2003 pour des tests plus particuliers. Enfin, dernièrement, on retrouve les travaux de S. Taskinen, H. Oja et R.H. Randles dans [17] en 2005, ceux de A. Gretton, K. Fukumizu dans [7] en 2008 et ceux de A. Gretton et L. Györfi dans [8].

3 Un nouveau test d'indépendance par permutation

Dans toute la suite, on ne se restreint plus au cadre de l'hypothèse $(\mathcal{H}_{PH})$ mais on se place dans le cadre de l'hypothèse $(\mathcal{H}_{PP})$.

3.1 Problème statistique

On envisage comme dans les travaux d'A. Rouis [15] une statistique basée sur le nombre de coïncidences, ici plus précisément

$$T = T(\mathbf{N}) = \sum_{k=1}^{n} X^k,$$

où $X^k = \int_a^b \int_a^b \mathbb{1}_{\{|x-y|\leq\delta\}} dN_1^k(x) dN_2^k(y)$.

Étant donné $\alpha \in]0,1[$, on cherche à construire un test rejetant (H_0) lorsque $T > q_{1-\alpha}$, où $q_{1-\alpha}$ sera une valeur critique à déterminer de telle sorte que le test soit exactement (non asymptotiquement) de niveau α sous l'hypothèse $(\mathcal{H}_{PP})$ seule.

La valeur critique que nous proposons dans cette partie est construite sur la base d'une approche de permutation comme décrite dans 2.2.

3.2 Notations et description du test

Notations. Pour mettre en œuvre l'approche de permutation, on aura besoin d'abord de connaître où sont les spikes pour chaque neurone individuellement, puis comment ils sont couplés. On introduit les notations suivantes :

- $\mathcal{N}_1 = (N_1^k)_{1\leq k\leq n} \in \mathbb{X}^n$ désigne l'ensemble des spikes du premier neurone (supposés ordonnés),
- $\mathcal{N}_2 = \{N_2^k\}_{1\leq k\leq n}$ désigne l'ensemble des spikes du second neurone (supposés non ordonnés), rendant, par un abus de notation, les répétitions distinguables (on notera bien que $\mathcal{N}_2$ est toujours de cardinal n, même en cas d'égalités de certains N_2^k),
- $\mathcal{S} = (\mathcal{N}_1; \mathcal{N}_2)$ code donc la position des spikes.
- L désigne l'application qui conditionnellement aux données $\mathcal{N}_1$ et $\mathcal{N}_2$ est définie par :

$$L : \left(\begin{array}{ccc} \{1,2,\ldots,n\} & \longrightarrow & \mathcal{N}_2 \\ i & \longmapsto & \text{le } N_2^i \text{ associé à } N_1^i \end{array} \right).$$

- L étant une bijection, on considère $\mathcal{L}$ l'ensemble des bijections entre $\{1,2,\ldots,n\}$ et $\mathcal{N}_2$.

Connaître la loi de $\mathbf{N}$ équivaut alors à connaître la loi de $\{\mathcal{S}, L\}$.

Statistique de test. Pour définir la statistique de test, on considère l'application (déterministe)

$$\text{Nbc} : (((u_k)_{1\leq k\leq n}, \{v_k\}_{1\leq k\leq n}), \ell) \mapsto \sum_{k=1}^{n} (\sum_{i,j} \mathbb{1}_{\{|u_k^i - \ell(k)^j|\leq\delta\}}).$$

qui compte le nombre de coïncidences entre les vecteurs $(u_1, \cdots, u_n)$ et $(\ell(1), \cdots, \ell(n))$ de $\mathbb{X}$, où ℓ est une bijection entre les ensembles $\{1,\ldots,n\}$ et $\{v_1,\ldots,v_n\}$.

Notre statistique de test peut ainsi s'écrire

$$T(\mathbf{N}) = \text{Nbc}(\mathcal{S}, L),$$

que l'on notera par souci de simplicité $T(\mathcal{S}, L)$.

Valeur critique. Étant donnée une variable aléatoire σ_n de loi uniforme sur $\mathfrak{S}_n$ indépendante de L et $\mathcal{S}$, on considère

$$T^{\sigma_n} = T(\mathcal{S}, L \circ \sigma_n).$$

Soit $F_{\mathcal{S},L}$ la fonction de répartition de la loi conditionnelle de T^{σ_n} sachant $\mathcal{S}$ et L. On choisit alors comme valeur critique

$$q_{1-\alpha} = (F_{\mathcal{S},L})^{-1}(1-\alpha).$$

Des rappels sur les notions de fonction de répartition et inverse généralisée sont faits en Annexe B et C.

3.3 Risque de première espèce du test

Niveau du test conditionnel. Commençons par un théorème que l'on montrera en section 3.4. On conditionne par rapport à l'ensemble des points des processus N_1^k et N_2^k de telle façon que le seul aléa restant réside dans le processus d'attribution des couples (N_1^k, N_2^k).

Théorème 3.
Quelle que soit la permutation $\sigma \in \mathfrak{S}_n$ (fixée), sous (H_0), on a l'égalité en lois :

$$L \circ \sigma \mid \mathcal{S} \sim L \mid \mathcal{S}.$$

On en déduit le résultat suivant.

Proposition 1.
Le test considéré est de bon niveau, c'est-à-dire que sous (H_0), il vérifie :

$$\mathbb{P}\left(T(\mathcal{S}, L) > (F_{\mathcal{S},L})^{-1}(1-\alpha) \mid \mathcal{S}\right) \leq \alpha.$$

Démonstration : Pour une meilleure lisibilité, notons $q(\mathcal{S}, L) := (F_{\mathcal{S},L})^{-1}(1-\alpha)$. Par le Théorème 3, on voit bien que quelle que soit la permutation $\sigma \in \mathfrak{S}_n$ fixée, sous l'hypothèse nulle,

$$\mathbb{P}\left(T(\mathcal{S}, L) > q(\mathcal{S}, L) \mid \mathcal{S}\right) = \mathbb{P}\left(T(\mathcal{S}, L \circ \sigma) > q(\mathcal{S}, L \circ \sigma) \mid \mathcal{S}\right),$$

avec $q(\mathcal{S}, L\circ\sigma) = (F_{\mathcal{S},L\circ\sigma})^{-1}(1-\alpha)$ est le $(1-\alpha)$-quantile de la loi conditionnelle $(T(\mathcal{S}, (L \circ \sigma) \circ \sigma_n) \mid \mathcal{S}, L)$. Or, il est clair que

$$\sigma_n \sim \mathcal{U}(\mathfrak{S}_n) \quad \Longleftrightarrow \quad \sigma \circ \sigma_n \sim \mathcal{U}(\mathfrak{S}_n).$$

On en déduit que

$$(T(\mathcal{S}, L \circ (\sigma \circ \sigma_n)) \mid \mathcal{S}, L) \;\sim\; (T(\mathcal{S}, L \circ \sigma_n) \mid \mathcal{S}, L).$$

D'où $q(\mathcal{S}, L \circ \sigma) = (F_{\mathcal{S},L\circ\sigma})^{-1}(1-\alpha) = (F_{\mathcal{S},L})^{-1}(1-\alpha) = q(\mathcal{S}, L)$.

Ainsi, on obtient

$$\begin{aligned}
\mathbb{P}\left(T(\mathcal{S}, L) > q(\mathcal{S}, L) \mid \mathcal{S}\right) &= \mathbb{P}\left(T(\mathcal{S}, L \circ \sigma) > q(\mathcal{S}, L) \mid \mathcal{S}\right) \\
\text{(ne dépend pas de } \sigma) &= \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \mathbb{P}\left(T(\mathcal{S}, L \circ \sigma_n) > q(\mathcal{S}, L) \mid \mathcal{S}, \sigma_n\right) \\
\text{(pour } \sigma_n \sim \mathcal{U}(\mathfrak{S}_n)) &= \sum_{\sigma \in \mathfrak{S}_n} \mathbb{P}\left(T(\mathcal{S}, L \circ \sigma_n) > q(\mathcal{S}, L) \mid \mathcal{S}, \sigma_n\right) \mathbb{P}\left(\sigma_n = \sigma\right) \\
&= \mathbb{E}_{\sigma_n}\left[\, \mathbb{P}\left(T(\mathcal{S}, L \circ \sigma_n) > q(\mathcal{S}, L) \mid \mathcal{S}, \sigma_n\right) \right] \\
&= \mathbb{E}_{\sigma_n}\left[\, \mathbb{P}_L\left(T(\mathcal{S}, L \circ \sigma_n) > q(\mathcal{S}, L)\right) \right] \\
&= \mathbb{E}_{\sigma_n}\left[\, \mathbb{E}_L\left[\, \mathbb{1}_{\{T(\mathcal{S}, L\circ\sigma_n) > q(\mathcal{S}, L)\}} \,\right] \right] \\
\text{(Fubini-Tonelli)} &= \mathbb{E}_L\left[\, \mathbb{E}_{\sigma_n}\left[\, \mathbb{1}_{\{T(\mathcal{S}, L\circ\sigma_n) > q(\mathcal{S}, L)\}} \,\right] \right] \\
&= \mathbb{E}_L\left[\, \mathbb{P}_{\sigma_n}\left(T(\mathcal{S}, L \circ \sigma_n) > q(\mathcal{S}, L)\right) \right]
\end{aligned}$$

où les notations $\mathbb{P}_Z(\cdot)$ et $\mathbb{E}_Z[\,\cdot\,]$ signifient que l'on "intègre" sur le seul aléa de Z, ou en d'autres termes, qu'on travaille conditionnellement à toutes les autres variables aléatoires.

Or,

$$\begin{aligned}
\mathbb{P}_{\sigma_n}\left(T(\mathcal{S}, L \circ \sigma_n) > q(\mathcal{S}, L)\right) &= \mathbb{P}\left(T(\mathcal{S}, L \circ \sigma_n) > (F_{\mathcal{S},L})^{-1}(1-\alpha) \mid \mathcal{S}, L\right) \\
&= 1 - \mathbb{P}\left(T(\mathcal{S}, L \circ \sigma_n) \leq (F_{\mathcal{S},L})^{-1}(1-\alpha) \mid \mathcal{S}, L\right) \\
&= 1 - F_{\mathcal{S},L}((F_{\mathcal{S},L})^{-1}(1-\alpha)) \\
\text{(Propriété 1 Annexe B)} &\leq 1 - (1-\alpha) \\
&\leq \alpha.
\end{aligned}$$

Ainsi, on a bien montré que

$$\mathbb{P}\left(T(\mathcal{S},L) > q(\mathcal{S},L) \mid \mathcal{S}\right) \;\leq\; \mathbb{E}_L\left[\,\alpha\,\right] = \alpha$$

d'où la proposition. □

Cette démonstration, inspirée de [2], repose sur une astuce qui n'est pas intuitive. On peut, en réalité, montrer un résultat plus fort et plus intuitif. Remarquons que ce résultat énoncé par exemple plus généralement par Romano (dans [14]) n'est en fait démontré nulle part.

Égalité des lois conditionnelles. On montre le résultat suivant.

Théorème 4.
Étant donnée une variable aléatoire σ_n de loi uniforme sur $\mathfrak{S}_n$ indépendante de L et $\mathcal{S}$, sous (H_0), on a l'égalité en loi :

$$T(\mathcal{S}, L\circ\sigma_n) \mid \mathcal{S}, L \;\sim T(\mathcal{S},L) \mid \mathcal{S}.$$

On retrouve bien le fait que le niveau conditionnel par rapport à $\mathcal{S}$ est bien inférieur à α.
En effet, si $G_{\mathcal{S}}$ désigne la fonction de répartition de la loi conditionnelle de $T(\mathcal{S},L) \mid \mathcal{S}$, par le Théorème 4, $G_{\mathcal{S}}$ et $F_{\mathcal{S},L}$ sont égales. Ainsi, par la propriété 1 en annexe B, on a bien que

$$\begin{aligned}\mathbb{P}\left(T(\mathcal{S},L) > (F_{\mathcal{S},L})^{-1}(1-\alpha) \mid \mathcal{S}\right) &= \mathbb{P}\left(T(\mathcal{S},L) > G_{\mathcal{S}}^{-1}(1-\alpha) \mid \mathcal{S}\right)\\ &\leq \alpha.\end{aligned}$$

Remarque **:** En particulier, ce théorème implique que permuter aléatoirement les N_2^k reviendra à simuler la loi conditionnelle de l'assemblage des points des neurones 1 et 2.

Remarque **:** En général, pour prouver qu'on peut appliquer les procédures par bootstrap, on cherche à montrer que la loi conditionnelle de la racine bootstrappée est "proche" (en un sens à définir) de celle de la vraie racine. Ici, par ce théorème, on a même montré que la distribution conditionnelle de la statistique permutée est égale à celle de la loi conditionnelle (par rapport aux spikes) de notre statistique de départ, ce qui justifie la permutation.

3.4 Démonstration des Théorèmes 3 et 4

3.4.1 L'assemblage est uniforme

Tout d'abord, on montre que L est un tirage uniforme (sans remise) sur $\mathcal{L}$ (conditionnellement aux données $\mathcal{S}$). On a d'abord besoin d'un lemme (sans conditionnement aux données).

Une égalité en lois non conditionnelle.

Lemme 2.
Soit un n-échantillon $(N_1^k, N_2^k)_{1\leq k\leq n}$ de même distribution que le couple (N_1, N_2) et soit une permutation $\sigma \in \mathfrak{S}_n$ (fixée ici).
Si N_1 et N_2 sont indépendants, alors on a l'égalité en loi :

$$\begin{pmatrix} (N_1^1, N_2^1)\\ (N_1^2, N_2^2)\\ \vdots\\ (N_1^n, N_2^n)\end{pmatrix} \sim \begin{pmatrix} (N_1^1, N_2^{\sigma(1)})\\ (N_1^2, N_2^{\sigma(2)})\\ \vdots\\ (N_1^n, N_2^{\sigma(n)})\end{pmatrix},$$

ce qui veut dire, en d'autres termes qu'on a :

$$(S,L) \sim (S, L\circ\sigma)$$

Preuve :

1ère étape : On commence par montrer que

$$\forall k \in \{1,\dots,n\}, \quad (N_1^k, N_2^k) \ \sim \ (N_1^k, N_2^{\sigma(k)}).$$

Si $k = \sigma(k)$, alors l'égalité en loi est évidente. Supposons donc que $k \neq \sigma(k)$.

Comme sous (H_0), on a $N_1 \perp\!\!\!\perp N_2$, alors $\mathbb{P}_{(N_1,N_2)} = \mathbb{P}_{N_1} \otimes \mathbb{P}_{N_2}$.
Ainsi, toujours sous (H_0),

$$\mathbb{P}_{(N_1^k,N_2^k)} = \mathbb{P}_{N_1^k} \otimes \mathbb{P}_{N_2^k} = \mathbb{P}_{N_1} \otimes \mathbb{P}_{N_2}.$$

Montrons donc que

$$\mathbb{P}_{(N_1^k,N_2^{\sigma(k)})} = \mathbb{P}_{N_1} \otimes \mathbb{P}_{N_2}.$$

Pour cela, commençons par montrer que $N_1^k \perp\!\!\!\perp N_2^{\sigma(k)}$:
Soit $A, B \subset \mathbb{X}$ mesurables, et notons $\begin{cases} \bar{A} = A \times \mathbb{X} \\ \bar{B} = \mathbb{X} \times B \end{cases}$

$$\begin{aligned} \mathbb{P}\left(\left\{N_1^k \in A\right\} \cap \left\{N_2^{\sigma(k)} \in B\right\}\right) &= \mathbb{P}\left(\left\{(N_1^k, N_2^k) \in \bar{A}\right\} \cap \left\{(N_1^{\sigma(k)}, N_2^{\sigma(k)}) \in \bar{B}\right\}\right) \\ &= \mathbb{P}\left((N_1^k, N_2^k) \in \bar{A}\right) \times \mathbb{P}\left((N_1^{\sigma(k)}, N_2^{\sigma(k)}) \in \bar{B}\right) \\ &= \mathbb{P}\left(N_1^k \in A\right) \times \mathbb{P}\left(N_2^{\sigma(k)} \in B\right). \end{aligned}$$

Ainsi, on a bien $N_1^k \perp\!\!\!\perp N_2^{\sigma(k)}$ d'où $\mathbb{P}_{(N_1^k,N_2^{\sigma(k)})} = \mathbb{P}_{N_1^k} \otimes \mathbb{P}_{N_2^{\sigma(k)}} = \mathbb{P}_{N_1} \otimes \mathbb{P}_{N_2}$.
En conclusion de la 1ère étape, on a bien pour tout $k \in \{1,\dots,n\}$,

$$\mathbb{P}_{(N_1^k,N_2^{\sigma(k)})} = \mathbb{P}_{(N_1^k,N_2^k)}.$$

2ème étape : Montrons maintenant le lemme.
On pose pour tout k, $Z_k = (N_1^k, N_2^k)$ et $\tilde{Z}_k = (N_1^k, N_2^{\sigma(k)})$.

Par la première étape, on a que $\forall k, Z_k \ \sim \tilde{Z}_k$.
De plus, par hypothèse, les Z_k sont indépendants car les essais le sont.

Montrons d'abord que les $\tilde{Z}_k$ sont indépendants :
Soit $I \subset \{1,\dots,n\}$ et $A_i, B_i \subset \mathbb{X}, \ i \in I$.

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i\in I}\left\{(N_1^i, N_2^{\sigma(i)}) \in A_i \times B_i\right\}\right) &= \mathbb{P}\left(\left\{\bigcap_{i\in I}(N_1^i \in A_i)\right\} \cap \left\{\bigcap_{i\in I}(N_2^{\sigma(i)} \in B_i)\right\}\right) \\ (N_1 \perp\!\!\!\perp N_2) \quad &= \mathbb{P}\left(\left\{\bigcap_{i\in I}(N_1^i \in A_i)\right\}\right) \times \mathbb{P}\left(\left\{\bigcap_{i\in I}(N_2^{\sigma(i)} \in B_i)\right\}\right) \\ (\text{essais } \perp\!\!\!\perp) \quad &= \prod_{i\in I} \mathbb{P}\left(N_1^i \in A_i\right) \times \prod_{i\in I} \mathbb{P}\left(N_2^{\sigma(i)} \in B_i\right) \\ &= \prod_{i\in I} \mathbb{P}\left(N_1^i \in A_i\right) \times \mathbb{P}\left(N_2^{\sigma(i)} \in B_i\right) \\ (N_1 \perp\!\!\!\perp N_2) \quad &= \prod_{i\in I} \mathbb{P}\left(\left\{N_1^i \in A_i\right\} \cap \left\{N_2^{\sigma(i)} \in B_i\right\}\right) \\ &= \prod_{i\in I} \mathbb{P}\left(\left\{(N_1^i, N_2^{\sigma(i)}) \in A_i \times B_i\right\}\right). \end{aligned}$$

Ainsi on a montré que

$$\mathbb{P}\left(\bigcap_{i\in I}\left\{\tilde{Z}_i \in A_i \times B_i\right\}\right) = \prod_{i\in I}\mathbb{P}\left(\left\{\tilde{Z}_i \in A_i \times B_i\right\}\right),$$

et ce quels que soient les sous-ensembles mesurables de $A_i \times B_i \subset \mathbb{X} \times \mathbb{X}$. On en déduit que $\forall\ C_i \subset \mathbb{X} \times \mathbb{X}$ mesurables, on a

$$\mathbb{P}\left(\bigcap_{i\in I}\left\{\tilde{Z}_i \in C_i\right\}\right) = \prod_{i\in I}\mathbb{P}\left(\left\{\tilde{Z}_i \in C_i\right\}\right),$$

et ce quel que soit $I \subset \{1,\dots,n\}$, c'est à dire que les $(\tilde{Z}_k)_k$ sont indépendants.

On en déduit que :

$$\mathbb{P}_{(\tilde{Z}_1,\dots,\tilde{Z}_n)} = \bigotimes_{k=1}^{n}\mathbb{P}_{\tilde{Z}_k} = \bigotimes_{k=1}^{n}\mathbb{P}_{Z_k} = \mathbb{P}_{(Z_1,\dots,Z_n)}.$$

□

Démonstration du Théorème 3. On voudrait montrer que pour toute fonction φ continue positive bornée, on a

$$\mathbb{E}\left[\ \varphi(L(1),\dots,L(n)) \mid \mathcal{S}\ \right] = \mathbb{E}\left[\ \varphi(L(\sigma(1)),\dots,L(\sigma(n))) \mid \mathcal{S}\ \right].$$

Soit φ continue bornée positive, et notons $g(\mathcal{S}) := \mathbb{E}\left[\ \varphi(L(1),\dots,L(n)) \mid \mathcal{S}\ \right]$.
On voudrait montrer $\mathbb{E}\left[\ \varphi(L(\sigma(1)),\dots,L(\sigma(n))) \mid \mathcal{S}\ \right] = g(\mathcal{S})$ *i.e.*

$$\forall h \in \mathcal{C}_b,\ \mathbb{E}\left[\ h(\mathcal{S})\varphi(L(\sigma(1)),\dots,L(\sigma(n)))\ \right] = \mathbb{E}\left[\ h(\mathcal{S})g(\mathcal{S})\ \right].$$

Soit $h \in \mathcal{C}_b$. Par définition de l'espérance conditionnelle,

$$\begin{aligned}\mathbb{E}\left[\ h(\mathcal{S})g(\mathcal{S})\ \right] &= \mathbb{E}\left[\ h(\mathcal{S})\ \mathbb{E}\left[\ \varphi(L(1),\dots,L(n)) \mid \mathcal{S}\ \right]\ \right] \\ &= \mathbb{E}\left[\ h(\mathcal{S})\ \varphi(L(1),\dots,L(n))\ \right].\end{aligned}$$

Calculons à présent le membre de gauche : on commence par introduire une nouvelle notation pour tout n-uplet $(X_1,Y_1),\dots,(X_n,Y_n)$:
$\Phi\begin{pmatrix}(X_1,Y_1)\\ \vdots \\ (X_n,Y_n)\end{pmatrix} = h(\mathcal{T})\ \varphi(L(1),\dots,L(n))$ (où $\mathcal{T} = ((X_i)_{1\le i\le n}, \{Y_i\}_{1\le i\le n})$ et $\forall i,\ L(i) = Y_i$).

Comme par la proposition 2 on a montré que $\begin{pmatrix}(N_1^1,N_2^1)\\ (N_1^2,N_2^2)\\ \vdots \\ (N_1^n,N_2^n)\end{pmatrix} \sim \begin{pmatrix}(N_1^1,N_2^{\sigma(1)})\\ (N_1^2,N_2^{\sigma(2)})\\ \vdots \\ (N_1^n,N_2^{\sigma(n)})\end{pmatrix}$,

on en déduit que $\Phi\begin{pmatrix}(N_1^1,N_2^1)\\ (N_1^2,N_2^2)\\ \vdots \\ (N_1^n,N_2^n)\end{pmatrix} \sim \Phi\begin{pmatrix}(N_1^1,N_2^{\sigma(1)})\\ (N_1^2,N_2^{\sigma(2)})\\ \vdots \\ (N_1^n,N_2^{\sigma(n)})\end{pmatrix}$.

D'où, comme dans $\mathcal{S}$, les N_2^i ne sont pas ordonnés, $h(((N_1^i)_{1\le i\le n}, \underbrace{\{N_2^{\sigma(i)}\}_{i\le i\le n}}_{\{N_2^i\}_{i\le i\le n}})) = h(\mathcal{S})$.

Ainsi,

$$\begin{aligned}
\mathbb{E}\left[\, h(\mathcal{S})\ \varphi(L(\sigma(1)),\dots,L(\sigma(n)))\,\right] &= \mathbb{E}\left[\,\Phi\begin{pmatrix}(N_1^1,N_2^{\sigma(1)})\\(N_1^2,N_2^{\sigma(2)})\\\vdots\\(N_1^n,N_2^{\sigma(n)})\end{pmatrix}\,\right]\\
&= \mathbb{E}\left[\,\Phi\begin{pmatrix}(N_1^1,N_2^{1})\\(N_1^2,N_2^{2})\\\vdots\\(N_1^n,N_2^{n})\end{pmatrix}\,\right]\\
&= \mathbb{E}\left[\, h(\mathcal{S})\ \varphi(L(1),\dots,L(n))\,\right].
\end{aligned}$$

Ainsi, on a bien démontré que

$$\mathbb{E}\left[\, h(\mathcal{S})\varphi(L(\sigma(1)),\dots,L(\sigma(n)))\,\right] = \mathbb{E}\left[\, h(\mathcal{S})\ \varphi(L(1),\dots,L(n))\,\right] = \mathbb{E}\left[\, h(\mathcal{S})g(\mathcal{S})\,\right].$$

et ce quelle que soit h, d'où

$$L\circ\sigma \mid \mathcal{S} \ \sim\ L \mid \mathcal{S}$$

ce qui achève la démonstration du Théorème 3.

L'assemblage est uniforme. On en déduit immédiatement la proposition suivante :

Proposition 2.
Conditionnellement à $\mathcal{S}$, L est un tirage uniforme sur $\mathcal{L}$.
Ainsi, une réalisation de L revient à tirer n fois sans remise uniformément dans $\mathcal{N}_2$.

Démonstration : En effet, soient ℓ et $\ell' \in \mathcal{L}$.
On pose $\sigma = \ell^{-1}\circ\ell'$ d'où $\ell' = \ell\circ\sigma$.

$$\begin{aligned}
\mathbb{P}\left(L=\ell \mid \mathcal{S}\right) &= \mathbb{E}\left[\,\mathbb{1}_{\{L=\ell\}} \mid \mathcal{S}\,\right]\\
&= \mathbb{E}\left[\,\mathbb{1}_{\{L=\ell'\circ\sigma^{-1}\}} \mid \mathcal{S}\,\right]\\
&= \mathbb{E}\left[\,\mathbb{1}_{\{L\circ\sigma=\ell'\}} \mid \mathcal{S}\,\right]\\
\text{(lemme 3)} \quad &= \mathbb{E}\left[\,\mathbb{1}_{\{L=\ell'\}} \mid \mathcal{S}\,\right]\\
&= \mathbb{P}\left(L=\ell' \mid \mathcal{S}\right).
\end{aligned}$$

□

Remarque **:** On en déduit ainsi que quelle que soit la permutation fixée $\sigma \in \mathfrak{S}_n$, $L\circ\sigma \mid \mathcal{S}$ est aussi un tirage uniforme sur $\mathcal{L}$. La question suivante est de savoir ce qui se passe lorsque la permutation est aléatoire.

3.4.2 Composition avec une permutation aléatoire

Proposition 3.
Si σ_n est une variable aléatoire uniforme sur $\mathfrak{S}_n$ indépendante de L et de $\mathcal{S}$, alors, conditionnellement à $\mathcal{S}$ et L, $L\circ\sigma_n$ est aussi uniforme sur $\mathcal{L}$.
Ainsi, on a l'égalité en loi :

$$L\circ\sigma_n \mid \mathcal{S}, L \ \sim\ L \mid \mathcal{S}.$$

Démonstration : En effet, soit $\ell \in \mathcal{L}$.

$$\begin{aligned}
\mathbb{P}\left(L\circ\sigma_n=\ell \mid \mathcal{S},L\right) &= \mathbb{P}\left(\sigma_n = L^{-1}\circ\ell \mid \mathcal{S},L\right)\\
&= \frac{1}{n!}.
\end{aligned}$$

car σ_n est uniforme sur $\mathfrak{S}_n$, et ce quel que soit $\ell \in \mathcal{L}$. □

L'application T étant déterministe, en appliquant directement la proposition 3, on a l'égalité

$$T(\mathcal{S}, L \circ \sigma_n) \mid \mathcal{S}, L \ \sim T(\mathcal{S}, L) \mid \mathcal{S}.$$

d'où le Théorème 4.

3.5 Estimation par Monte Carlo

3.5.1 Test avec estimation par Monte Carlo

En pratique, le quantile conditionnel $(F_{\mathcal{S},L})^{-1}(1-\alpha)$ est approché par une méthode de Monte Carlo classique. On note $F^B_{\mathcal{S},L}$ la fonction de répartition empirique obtenue à partir d'un B-échantillon $(\sigma_n^1, \ldots, \sigma_n^B)$ de la loi uniforme sur $\mathfrak{S}_n$, défini par

$$F^B_{\mathcal{S},L}(x) = \frac{1}{B}\sum_{j=1}^{n} \mathbb{1}_{\{T(\mathcal{S}, L\circ\sigma_n^j)\leq x\}}.$$

Des rappels sur les fonctions de répartition empiriques et les quantiles empiriques sont faits en Annexe C. Le quantile conditionnel $(F_{\mathcal{S},L})^{-1}(1-\alpha)$ est alors approché par $\left(F^B_{\mathcal{S},L}\right)^{-1}(1-\alpha)$. Le test que l'on veut mettre en œuvre en pratique est finalement donné par

$$\Phi_B = \mathbb{1}_{\{T(\mathcal{S},L) > \left(F^B_{\mathcal{S},L}\right)^{-1}(1-\alpha)\}}.$$

Si notre statistique de test était continue, on aurait pu, par des arguments classiques, montrer que le niveau asymptotique en B était exactement α. Le problème ici, est que cette statistique compte le nombre de coïncidences et n'a donc aucune chance d'être continue. On se doute bien qu'il est possible de démontrer un résultat du même genre dans le cas discret, mais il y a des chances que l'on perde l'égalité. On va donc plutôt se concentrer sur le niveau non asymptotique. Bien qu'on se doute que l'on puisse choisir un B aussi grand que possible, cela n'est cependant pas judicieux en termes de coût algorithmique. La question d'un choix adéquat de B est donc cruciale et justifie le point de vue non asymptotique.

3.5.2 Monte Carlo d'un point de vue non-asymptotique

Théorème 5.

Avec les mêmes notations que précédemment, pour tout $B \in \mathbb{N}^*$, le test Φ_B est exactement de niveau $\alpha_B = \frac{\lfloor B\alpha\rfloor+1}{B+1}$, i.e.

$$\mathbb{P}_{(H_0)}\left(T(\mathcal{S},L) > \left(F^B_{\mathcal{S},L}\right)^{-1}(1-\alpha)\right) \leq \frac{\lfloor B\alpha\rfloor+1}{B+1} \leq \alpha + \frac{1-\alpha}{B+1}.$$

Démonstration : (c.f. [2]) Dans toute cette démonstration, on suppose que (H_0) est vérifiée. Pour construire le quantile empirique, on commence par tirer uniformément et de façon indépendante un n-échantillon $(\sigma_1, \ldots, \sigma_B)$ de loi uniforme sur $\mathfrak{S}_n$.
Posons $\sigma_0 = L$. Pour une meilleure lecture, notons

$$\begin{cases} X_0^\star &=& T(\mathcal{S},L) & \text{la statistique de test calculée sur notre échantillon,} \\ X_k^\star &=& T(\mathcal{S}, L\circ\sigma_k) & \text{la statistique calculée sur l'échantillon permuté.} \end{cases}$$

Le quantile empirique (c.f. Annexe C) est obtenu par $\left(F^B_{\mathcal{S},L}\right)^{-1}(1-\alpha) = X^\star_{(\lceil B(1-\alpha)\rceil)} = X^\star_{(B-\lfloor B\alpha\rfloor)}$, où $X^\star_{(1)} \leq \cdots \leq X^\star_{(B)}$ représentent les statistiques d'ordre associées à $X_1^\star, \ldots, X_B^\star$.

Pour commencer, on s'intéresse à l'erreur conditionnelle $(\mathcal{EC}) = \mathbb{P}\left(T(\mathcal{S},L) > \left(F^B_{\mathcal{S},L}\right)^{-1}(1-\alpha) \mid \mathcal{S}\right)$. Avec nos nouvelles notations, on a

$$(\mathcal{EC}) = \mathbb{P}\left(X_0^\star > X^\star_{(B-\lfloor B\alpha\rfloor)} \mid \mathcal{S}\right).$$

Or on voit bien que, pour tout $x \in \mathbb{R}$ et pour tout $m \in [\![1;B]\!]$, $\{X^\star_{(m)} < x\}$ *ssi* le nombre de $\{X^\star_k < x, 1 \leq k \leq B\}$ est supérieur ou égal à m, d'où (pour $m = B - \lfloor B\alpha \rfloor$ et $x = X^\star_0$), on obtient :

$$\begin{aligned}
(\mathcal{EC}) &= \mathbb{P}\left(\#\left\{k \in [\![1;B]\!]; X^\star_k < X^\star_0\right\} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right) \\
&= \mathbb{P}\left(\sum_{k=1}^{B} \mathbb{1}_{\{X^\star_k < X^\star_0\}} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right) \\
&= \mathbb{P}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_k < X^\star_0\}} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right), \\
&= \mathbb{P}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_0\}} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right),
\end{aligned}$$

où $\#C$ désigne le cardinal d'un ensemble C. On peut faire démarrer la somme à 0 car $\mathbb{1}_{\{X^\star_0 < X^\star_0\}} = 0$. Or, par construction, les $X^\star_0, \ldots, X^\star_k$ sont indépendants et de même loi conditionnellement à $\mathcal{S}$ (par le Théorème 4). Et en particulier, $\forall u \in \{0, \ldots, B\}$, on a par symétrie

$$\mathbb{P}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_0\}} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right) = \mathbb{P}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(u)}\}} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right).$$

Prenons à présent $U \sim \mathcal{U}(\{0, \ldots, B\})$ (indépendant de tout le reste). Alors

$$\begin{aligned}
\mathbb{P}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(U)}\}} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right) &= \mathbb{E}\left[\, \mathbb{1}_{\{\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(U)}\}} \geq B - \lfloor B\alpha \rfloor\}} \mid \mathcal{S} \,\right] \\
&= \mathbb{E}\left[\, \sum_{u=0}^{B} \mathbb{1}_{\{U=u\}} \mathbb{1}_{\{\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(u)}\}} \geq B - \lfloor B\alpha \rfloor\}} \mid \mathcal{S} \,\right] \\
\text{(par indépendance)} \quad &= \sum_{u=0}^{B} \mathbb{P}(U = u)\, \mathbb{P}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(u)}\}} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right) \\
&= \frac{1}{B+1} \sum_{u=0}^{B} \mathbb{P}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_0\}} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right) \\
&= \mathbb{P}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_0\}} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right).
\end{aligned}$$

On obtient donc que

$$\begin{aligned}
(\mathcal{EC}) &= \mathbb{P}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(U)}\}} \geq B - \lfloor B\alpha \rfloor \mid \mathcal{S}\right) \\
&= \mathbb{P}_{\{\sigma_{(k)}\}_{0 \leq k \leq B}}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(U)}\}} \geq B - \lfloor B\alpha \rfloor\right)
\end{aligned}$$

Comme $(\mathcal{EC})$ ne dépend pas de U, lorsqu'on "intègre" sur U, on a

$$\begin{aligned}
(\mathcal{EC}) &= \mathbb{E}_U\left[\, \mathbb{P}_{\{\sigma_k\}}\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(U)}\}} \geq B - \lfloor B\alpha \rfloor\right) \,\right] \\
\text{(Fubini Tonelli)} \quad &= \mathbb{E}_{\{\sigma_k\}}\left[\, \mathbb{P}_U\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(U)}\}} \geq B - \lfloor B\alpha \rfloor\right) \,\right].
\end{aligned}$$

Or, comme précédemment,

$\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(U)}\}} \geq m$ <u>ssi</u> le nombre de $\{X^\star_{(k)} < X^\star_{(U)}\}$ est supérieur à m

<u>i.e.</u> il y a au moins m "$X^\star_{(k)}$" inférieurs strictement à $X^\star_{(U)}$

$\Rightarrow$ $U > m-1$ (car lorqu'ils sont ordonnés, on a $X^\star_{(0)} \leq X^\star_{(1)} \leq \cdots \leq X^\star_{(m-1)} \leq \cdots < X^\star_{(U)}$).

Donc au final,

$$\begin{aligned}
\mathbb{P}_U\left(\sum_{k=0}^{B} \mathbb{1}_{\{X^\star_{(k)} < X^\star_{(U)}\}} \geq B - \lfloor B\alpha \rfloor\right) &\leq \mathbb{P}_U\left(U \geq B - \lfloor B\alpha \rfloor\right) \\
&= \frac{B+1-(B-\lfloor B\alpha \rfloor)}{B+1} \\
&= \frac{\lfloor B\alpha \rfloor + 1}{B+1}.
\end{aligned}$$

d'où

$$\begin{aligned}
(\mathcal{EC}) &\leq \mathbb{E}_{\{\sigma_k\}}\left[\frac{(\lfloor B\alpha \rfloor + 1)}{B+1}\right] \\
&= \frac{\lfloor B\alpha \rfloor + 1}{B+1} \\
&= \alpha + \frac{1-\alpha}{B+1}.
\end{aligned}$$

En prenant l'espérance de $(\mathcal{EC})$, on obtient bien le théorème. □

4 Vers des tests d'indépendance bootstrap

4.1 De la permutation au bootstrap d'Efron

Notre statistique de test peut s'écrire sous de la forme

$$T = \sum_{k=1}^{n} F(N_1^k, N_2^k),$$

où F compte le nombre de coïncidences entre les deux processus ponctuels N_1^k et N_2^k. Pour appliquer notre test par permutation, on considère donc $T^{\sigma_n} = \sum_{k=1}^{n} F(N_1^k, N_2^{\sigma_n(k)})$ avec σ_n tiré uniformément sur $\mathfrak{S}_n$.

Ainsi, on obtient

$$\begin{aligned}
T^{\sigma_n} &= \sum_{i=1}^{n} F(N_1^i, N_2^{\sigma_n(i)}) \\
&= \sum_{i=1}^{n} F(N_1^i, \sum_{j=1}^{n} \mathbb{1}_{\{j=\sigma_n(i)\}} N_2^j) \\
&= \sum_{i,j=1}^{n} \underbrace{\mathbb{1}_{\{j=\sigma_n(i)\}}}_{W^P_{(i,j)}} F(N_1^i, N_2^j).
\end{aligned}$$

Une autre méthode possible serait d'appliquer le bootstrap d'Efron au lieu de la permutation : on commence par créer nos échantillons bootstrap N_1^* et N_2^* en tirant au hasard avec remise n points respectivement dans $(N_1^1, \ldots, N_1^n)$ et $(N_2^1, \ldots, N_2^n)$. La statistique bootstrappée devient donc

$$T^* = \sum_{k=1}^{n} F(N_1^{*k}, N_2^{*k}).$$

Cela revient à la construction suivante. Considérons à présent $(U_1, \dots, U_n)$ et $(V_1, \dots, V_n)$ les deux échantillons indépendants de loi uniforme sur $\{1, \dots, n\}$ correspondant aux tirages aléatoires précédents :

$$\left\{ \begin{array}{rcccl} N_1^{*k} & = & N_1^{U_k} & = & \sum_{i=1}^n N_1^i \mathbb{1}_{\{U_k=i\}} \\ N_2^{*k} & = & N_2^{V_k} & = & \sum_{j=1}^n N_2^j \mathbb{1}_{\{V_k=j\}}. \end{array} \right.$$

On obtient donc :

$$\begin{aligned} T^* &= \sum_{k=1}^n F(\sum_{i=1}^n N_1^i \mathbb{1}_{\{U_k=i\}}, \sum_{j=1}^n N_2^j \mathbb{1}_{\{V_k=j\}}) \\ &= \sum_{k=1}^n \sum_{i=1}^n \sum_{j=1}^n \mathbb{1}_{\{U_k=i\}} \mathbb{1}_{\{V_k=j\}} F(N_1^i, N_2^j) \\ &= \sum_{i,j=1}^n \underbrace{(\sum_{k=1}^n \mathbb{1}_{\{U_k=i\}} \mathbb{1}_{\{V_k=j\}})}_{W^B_{(i,j)}} F(N_1^i, N_2^j). \end{aligned}$$

On voit bien ici le lien entre la permutation et le bootstrap ; dans les deux cas, on obtient du bootstrap à poids avec les poids $W^P_{(i,j)} = \mathbb{1}_{\{j=\sigma_n(i)\}}$ pour la permutation et $W^B_{(i,j)} = \sum_{k=1}^n \mathbb{1}_{\{U_k=i\}} \mathbb{1}_{\{V_k=j\}}$ pour le bootstrap classique. Dans les deux cas, les poids définis précédemment ont l'inconvénient de ne pas être indépendants (dans les deux cas, la somme sur i et j de ces poids vaut n).

Lors de l'étude de la puissance des tests, cette non indépendance sera un obstacle à surmonter. Une des questions qui se posent alors est de savoir s'il est possible d'introduire du bootstrap sauvage ici.

4.2 Du bootstrap d'Efron au bootstrap sauvage

Comme on l'a vu dans la sous-partie 2.3, on propose de "Poissoniser" le bootstrap d'Efron classique. On tire au hasard avec remise N éléments parmi les $\{N_1^1, \dots, N_1^n\}$ et N éléments parmi les $\{N_2^1, \dots, N_2^n\}$ avec $N \sim \mathcal{P}(n)$. Cela revient à considérer deux N-échantillons $(U_1, \dots, U_N)$ et $(V_1, \dots, V_N)$ de loi uniforme sur $\{1, \dots, n\}$, et à prendre

$$\left\{ \begin{array}{rcccl} \left(N_1^W\right)^k & = & N_1^{U_k} & = & \sum_{i=1}^N N_1^i \mathbb{1}_{\{U_k=i\}} \\ \left(N_2^W\right)^k & = & N_2^{V_k} & = & \sum_{j=1}^N N_2^j \mathbb{1}_{\{V_k=j\}}. \end{array} \right.$$

Par un calcul analogue à celui pour les poids d'Efron, on obtient donc comme statistique bootstrappée :

$$T^W = \sum_{i,j=1}^n \underbrace{(\sum_{k=1}^N \mathbb{1}_{\{U_k=i\}} \mathbb{1}_{\{V_k=j\}})}_{W^N_{(i,j)}} F(N_1^i, N_2^j).$$

Afin de montrer qu'il s'agit bien de bootstrap sauvage, on doit vérifier que les poids $W^N_{(i,j)}$ sont bien indépendants. Pour cela on utilise un propriété des processus de Poisson.

On remarque tout d'abord que tirer aléatoirement les deux échantillons $(U_1, \dots, U_N)$ et $(V_1, \dots, V_N)$ de loi uniforme sur $\{1, \dots, n\}$ conditionnellement à N, revient à tirer $(Z_k = (X_k, Y_k))_{1 \le k \le N}$ uniformément sur $[0, n] \times [0, n]$ et de considérer $U_i = \lceil X_{i,j} \rceil$ et $V_j = \lceil Y_{i,j} \rceil$.

Ainsi, poissoniser le Bootstrap d'Efron revient à faire ce qui suit.
On commence par tirer aléatoirement la taille de notre échantillon $N \sim \mathcal{P}(n)$. Ensuite, on considère les Z_k pris uniformément sur $[0, n] \times [0, n]$, ce qui nous permet ensuite de déterminer les U_i et les V_j. Les poids $W_{i,j} = \sum_{k=1}^N \mathbb{1}_{\{U_k=i\}} \mathbb{1}_{\{V_k=j\}}$ comptent le nombre de Z_k appartenant au pavé $]i-1, i] \times]j-1, j]$. On voit bien que la propriété 2 de l'Annexe D, se généralise bien aux processus de Poisson en dimension supérieure. On l'applique alors aux processus en dimension 2 et on obtient que le processus de comptage des Z_k est un processus de Poisson homogène d'intensité λ, avec le paramètre de N vérifiant : $n = \lambda n^2$, d'où $\lambda = 1/n$. Les pavés $]i-1, i] \times]j-1, j]$ étant tous disjoints, par propriété d'indépendance des processus de Poisson, on déduit donc que les $(W^N_{i,j})_{(i,j)}$ sont i.i.d. et que $\forall (i,j), W^N_{i,j} \sim \mathcal{P}(1/n)$.

5 Étude de simulations

Les programmes de simulation sont donnés en Annexe F.

5.1 Simulation des processus de Poisson homogènes

Afin de pouvoir comparer les résultats obtenus pour le test par permutation construit dans ce mémoire à ceux déjà obtenus dans [15], on commence par simuler des données issues de processus de Poisson homogènes.

Pour cela, on utilise la propriété 2 détaillée en annexe D. Pour chaque neurone, on simule d'abord le nombre de spikes pour chaque essai suivant une loi de Poisson, et ensuite, conditionnellement à cet effectif, les spikes eux-mêmes uniformément sur $[a, b]$.

5.2 Étude des fonctions de répartition empiriques

Notons G la fonction de répartition empirique de la loi de notre statistique $T(\mathcal{S}, L)$.

Dans un premier temps, on a illustré sur la figure 2 le fait que les lois non conditionnelles de $T(\mathcal{S}, L)$ et $T(\mathcal{S}, L \circ \sigma_n)$ sont proches, pour σ_n permutation aléatoire uniforme de $\mathfrak{S}_n$. Pour cela, on a regardé les fonctions de répartitions empiriques de chacune des lois. On a donc commencé par créer deux échantillons i.i.d. de notre statistique de test. Le premier a été calculé à partir de données simulées suivant un processus de Poisson d'intensité `lambda=10` ; en rouge est tracée la fonction de répartition empirique de la loi $T(\mathcal{S}, L)$ associée à cet échantillon. Elle approche la vraie fonction de répartition empirique G. Ensuite, chaque donnée a été permutée aléatoirement afin de calculer une nouvelle fois notre statistique de test constituant notre deuxième échantillon dont la fonction de répartition empirique est tracée en vert. Elle approche la loi de $T(\mathcal{S}, L \circ \sigma_n)$.

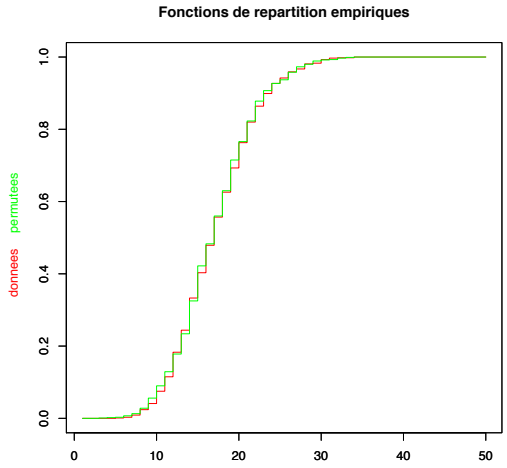

FIGURE 2 – Fonctions de répartition empiriques de la statistique calculée en les données et leur permutées.

On voit bien les deux courbes sont très proches. Ceci illustre une conséquence du lemme 2 prenant cette fois une permutation tirée uniformément sur $\mathfrak{S}_n$.

Ensuite, on a voulu voir comment se comportait la statistique lorsqu'on conditionne par rapport à l'ensemble des spikes $\mathcal{S}$ et l'attribution de départ L. On a donc simulé, toujours suivant un processus de Poisson homogène, une donnée que l'on a fixée. Puis on a tracé en bleu sur la figure 3, la fonction de répartition empirique $F^B_{\mathcal{S},L}$ calculée à partir de `B` permutées de notre donnée initiale. On retrouve en rouge la fonction de répartition empirique de la loi $T(\mathcal{S}, L)$ (la même que dans la figure 2 sans permutation).

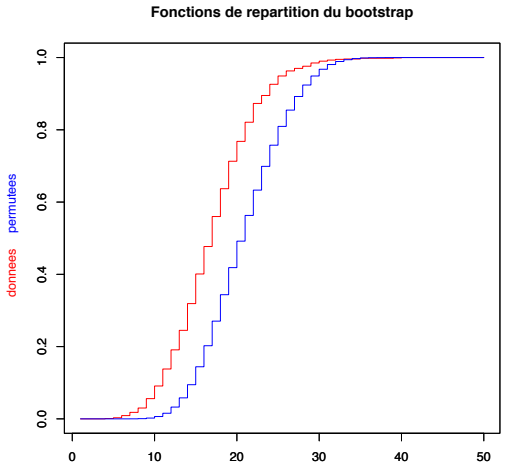

FIGURE 3 – Fonctions de répartition empiriques de la statistique calculée sur les permutées d'une donnée fixée.

On voit cette fois un gros écart entre les deux courbes. Celui-ci est dû au fait que l'on conditionne par rapport aux points et à l'attribution des couples de départ. Le seul aléa qui reste réside alors dans la permutation aléatoire des données. D'après le Théorème 4, cette courbe bleue approche la loi de la statistique $T(\mathcal{S}, L)|\mathcal{S}$ conditionnée par rapport aux points, et où l'aléa est seulement dans l'attribution des couples. On voit ici l'influence de ce conditionnement.

Pour voir si la différence n'est pas dûe à l'erreur d'approximation par Monte Carlo, on a itéré le raisonnement ; sur la figure 4, sont traçées plusieurs fonctions de répartition empiriques $F^B_{\mathcal{S},L}$ obtenues par permutation aléaloires uniformes d'une seule et même donnée.

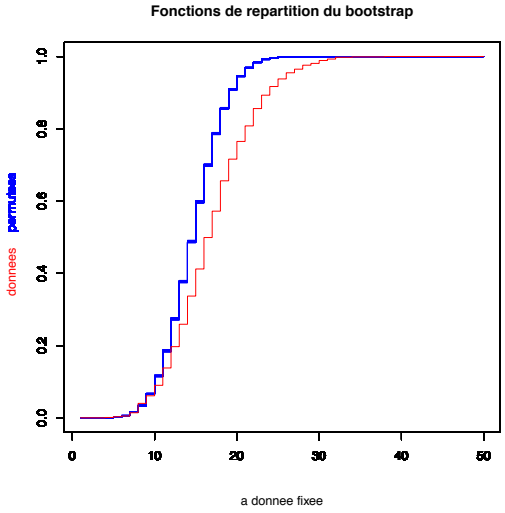

FIGURE 4 – Fonctions de répartition empiriques de la statistique calculée en les permutées d'une donnée fixée.

On voit bien que ces courbes se regroupent autour d'une autre courbe qui n'est pas la rouge. Encore ici, on voit l'influence du conditionnement par rapport aux positions des spikes. En fait, elles se regroupent autour de la fonction de répartition de la loi $T(\mathcal{S}, L) \mid \mathcal{S}$.

Finalement, pour se débarasser du conditionnement, on a tracé les fonctions de répartition empiriques obtenues par permutation simulées en changeant à chaque fois la donnée des spikes.

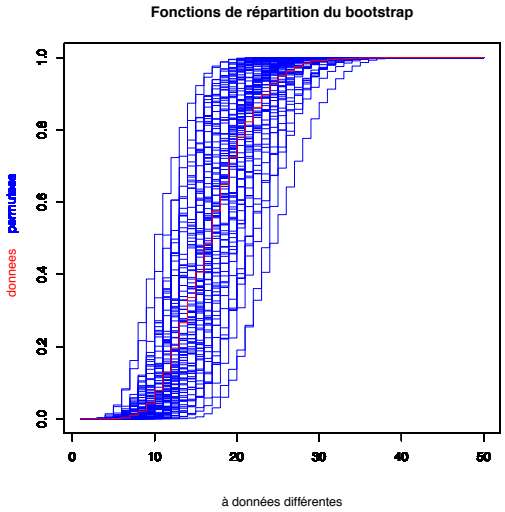

FIGURE 5 – Fonctions de répartition empiriques de la statistique calculée sur les permutées en variant les données.

On voit bien ici qu'elles sont différentes de la rouge (à cause du conditionnement), mais que par contre celle-ci est à peu près centrée parmi les bleues, d'où l'idée de regarder leur moyenne.

La figure 6 est obtenue en traçant la moyenne des fonctions de répartition empiriques obtenues par permutation en changeant la donnée à chaque fois.

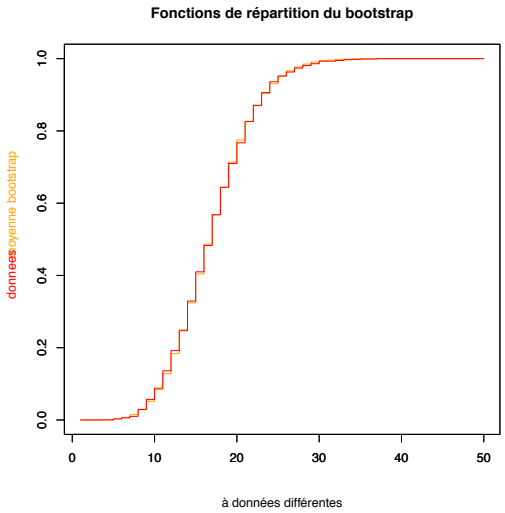

FIGURE 6 – Moyenne des fonctions de répartition des permutées en variant les données.

Ainsi, on trace en orange la moyenne des $F^B_{\mathcal{S},L}$ obtenue pour `M` répétitions $\frac{1}{M}\sum_{j=1}^{M} F^B_{\mathcal{S}_i,L_i}$ ce qui approche pour tout t, $\mathbb{E}\left[\ F_{\mathcal{S},L}(t)\ \right] = \mathbb{E}\left[\ \mathbb{P}\left(T(\mathcal{S},L) \leq t \mid \mathcal{S},L\right)\ \right] = \mathbb{P}\left(T(\mathcal{S},L) \leq t\right) = G(t)$. On retrouve exactement la fonction de répartition de la loi $T(\mathcal{S},L)$.

D'ailleurs, pour illuster ce résultat, on voit bien sur la figure 6 que les deux courbes se superposent presque. Cela illustre le fait que la fonction de répartition empirique bootstrap est très proche de celle de la vraie loi (sans conditionnement aux données), ce qui justifie la permutation.

5.3 Programmation des tests et estimation des niveaux

Afin de vérifier si notre test est bien de niveau convenable, on peut estimer ce dernier.

Tout d'abord, il faut construire le test. De manière classique, et dans le but de faire éventuellement des tests multiples, on préfère calculer la p-valeur, dont la définition est rappelée en Annexe A. Ici, on rappelle que notre test s'écrit

$$\Phi_B = \Phi_{B,\alpha}(\mathcal{S}, L) = \mathbb{1}_{\{T(\mathcal{S},L) > \left(F^B_{\mathcal{S},L}\right)^{(-1)}(1-\alpha)\}},$$

où $\left(F^B_{\mathcal{S},L}\right)^{(-1)}(1-\alpha)$ est calculée à partir d'un B-échantillon i.i.d. $\sigma_1, \dots, \sigma_B$ de permutations tirées uniformément sur $\mathfrak{S}_n$.

La p-valeur s'écrit donc dans notre cas,

$$\hat{\alpha}(\mathcal{S}, L) = \sup\left\{\alpha \ / \ \Phi_{B,\alpha}(\mathcal{S}, L) = 0\right\}.$$

Comme dans la démonstration du Théorème 5, on note $X^\star_k = T(\mathcal{S}, L \circ \sigma_k)$ la statistique calculée à partir de la $k^{\text{ème}}$ permutée. On pose aussi $X = T(\mathcal{S}, L)$ la vraie statistique (calculée sans permutation).

Remarque **:** Remarquons juste que lorsque α augmente, $(1-\alpha)$ diminue, donc $\lceil B(1-\alpha)\rceil$ aussi, et comme les $X^\star_{(k)}$ sont ordonnés, on en déduit que $\left(F^B_{\mathcal{S},L}\right)^{(-1)}(1-\alpha) = X^\star_{(\lceil B(1-\alpha)\rceil)}$ diminue.
En particulier, $\mathbb{P}\left(X \leq X^\star_{(\lceil B(1-\alpha)\rceil)}\right)$ diminue aussi, ce qui signifie que plus α augmente, à donnée fixée, plus la chance d'accepter l'hypothèse nulle diminue d'où le supremum dans la définition de la p-valeur.

Comment calculer la p-valeur ? Tout d'abord,

$$\begin{aligned}
\Phi_{B,\alpha}(\mathcal{S}, L) = 0 &\Leftrightarrow T(\mathcal{S}, L) \leq \left(F^B_{\mathcal{S},L}\right)^{(-1)}(1-\alpha) \\
&\Leftrightarrow X \leq X^\star_{(\lceil B(1-\alpha)\rceil)} \\
&\Leftrightarrow \#\left\{X^\star \geq X\right\} \geq B - (\lceil B(1-\alpha)\rceil - 1) = \lfloor \alpha B \rfloor + 1 \\
&\Leftrightarrow \lfloor \alpha B \rfloor \leq \#\left\{X^\star \geq X\right\} - 1
\end{aligned}$$

Or ce qui nous intéresse, c'est de minimiser le α vérifiant la condition ci-dessus. Posons

$$\tilde{\alpha} = \frac{\#\left\{X^\star \geq X\right\}}{B}.$$

Alors $\tilde{\alpha}B = \#\left\{X^\star \geq X\right\} \in \mathbb{N}$ d'où $\lfloor \tilde{\alpha}B \rfloor = \tilde{\alpha}B > \#\left\{X^\star \geq X\right\} - 1$.
D'un autre côté, $\forall \alpha < \tilde{\alpha}$, on a $\alpha B < \tilde{\alpha}B \in \mathbb{N}$ d'où $\lfloor \alpha B \rfloor < \lfloor \tilde{\alpha}B \rfloor$ On en déduit que

$$\begin{aligned}
\lfloor \alpha B \rfloor &\leq \lfloor \tilde{\alpha}B \rfloor - 1 \\
&= \tilde{\alpha}B - 1 \\
&= \#\left\{X^\star \geq X\right\} - 1
\end{aligned}$$

Et $\forall \alpha \geq \tilde{\alpha}$, on a par croissance de la partie entière

$$\lfloor \alpha B \rfloor \geq \lfloor \tilde{\alpha}B \rfloor \geq \#\left\{X^\star \geq X\right\} - 1$$

D'où on a montré que

$$\sup\left\{\alpha \ / \ \lfloor \alpha B \rfloor \leq \#\left\{X^\star \geq X\right\} - 1\right\} = \tilde{\alpha} = \frac{\#\left\{X^\star \geq X\right\}}{B}.$$

Ainsi la p-valeur est obtenue par la formule :

$$\hat{\alpha}(\mathcal{S}, L) = \frac{1}{B}\sum_{j=1}^{B} \mathbb{1}_{\{T(\mathcal{S},L) \leq T(\mathcal{S}, L\circ\sigma_k)\}}$$

Le test est donc construit de la manière suivante : on commence par simuler les permutations aléatoires, puis on calcule les nombres de coïncidences sur chaque donnée (permutée ou non). Ensuite on calcule la p-valeur pour finalement rejetter l'hypothèse nulle lorsque cette p-valeur est inférieure ou égale à α.

Estimation du niveau. Pour estimer le niveau, on répète `Nsimu` fois le test : à chaque itération, on simule une nouvelle donnée. Ensuite, on calcule la p-valeur associée à chaque nouvelle simulation des spikes que l'on garde en mémoire dans un vecteur. Afin de calculer le taux de rejets de l'hypothèse nulle (qui est pourtant vraie lorsqu'on simule les données initiales), on compte le nombre de fois que cette p-valeur est inférieure ou égale au seuil α fixé au préalable. Finalement, on estime le niveau par cette quantitée renormalisée par `Nsimu`.

Résultats pour le test d'indépendance par permutation. On a lancé les programmes de simulations détaillés en Annexe F en faisant varier plusieurs paramètres. Sachant que la méthode d'A. Rouis trouve ses limites lorsque les plages de temps $[a, b]$ sont très petites, on s'est fixé comme paramètres `a=1` et `b=1.1`, avec un `delta=0.01`. Ensuite, on a regardé les niveaux empiriques pour le test par permutation construit dans ce mémoire en faisant varier deux paramètres principaux : l'intensité des processus de Poisson homogènes `lambda` et le nombre d'essais `nessai`.

Voici les résultats que l'on a obtenu. D'une ligne à l'autre, c'est le paramètre `lambda` qui varie alors que d'une colonne à l'autre, c'est le nombre d'essais.

	`nessai` $= 10$	`nessai` $= 30$	`nessai` $= 50$	`nessai` $= 100$
`lambda` $= 10$	2.16%	3.32%	3.84%	4.16%
(IC)	$[1.76; 2.56]$	$[2.82; 3.82]$	$[3.31; 4.37]$	$[3.61; 4.71]$
`lambda` $= 20$	3.64%	4.04%	4.60%	4.06%
(IC)	$[3.12; 4.16]$	$[3.49; 4.59]$	$[4.02; 5.18]$	$[3.51; 4.61]$
`lambda` $= 50$	4.60%	4.78%	4.72%	5.10%
(IC)	$[4.02; 5.18]$	$[4.19; 5.37]$	$[4.13; 5.31]$	$[4.50; 5.71]$
`lambda` $= 80$	4.54%	5.08%	5.36%	5.06%
(IC)	$[3.96; 5.12]$	$[4.47; 5.69]$	$[4.74; 5.98]$	$[4.45; 5.67]$

On a regardé les intervalles de confiance (donnés en pourcentages) de niveau $\varepsilon = 5\%$ pour chacun des résultats obtenus. On aimerait montrer que notre test (prenant les valeurs 0 ou 1) se comporte sous l'hypothèse nulle comme une variable aléatoire de Bernouilli de paramètre $p \leq 5\%$. On peut alors construire un intervalle de confiance asymptotique de niveau ε (par une méthode utilisant le théorème cental limite) qui est de la forme

$$IC = \left[\hat{p} - q_{1-\varepsilon/2} \frac{\sqrt{\hat{p}\,(1-\hat{p})}}{\sqrt{\texttt{Nsimu}}} \;;\; \hat{p} + q_{1-\varepsilon/2} \frac{\sqrt{\hat{p}\,(1-\hat{p})}}{\sqrt{\texttt{Nsimu}}} \right],$$

où $q_{1-\varepsilon/2} = 1.96$ est le $(1-\varepsilon/2)$-quantile de la loi gaussienne centrée réduite.

On voit bien qu'en règle générale, le niveau empirique est toujours inférieur ou égal au seuil de 5% qu'on s'est fixé au départ. Les cas où cela dépasse légèrement ce seuil, on voit bien que le niveau espéré $p = 5\%$ est bien dans l'intervalle de confiance obtenu.

Lorsque le paramètre `lambda` est grand, on obtient un niveau empirique très proche de 5% ce qui est plutôt rassurant pour l'étude de la puissance. Cependant, lorsque `lambda` est petit, on obtient des niveaux empiriques largement inférieurs. Cela est du au fait que dans ces cas là, comme `lambda` représente l'intensité des processus de Poisson, on obtient très peu de points pour les données. Du coup, lorsqu'on permute, on observe très peu de coïncidences (pour chaque essai, on n'observe que 0 ou 1 coïncidence). Ainsi, la statistique calculée sur données permutées contient des atomes avec une masse importante, ce qui nous empêche d'avoir un niveau près de 5%. Cependant, ce qui est important c'est qu'il soit bien toujours inférieur ou égal à $\alpha = 5\%$ ce qui est bien le cas ici.

6 Conclusion et Perspectives

En conclusion, on a bien réussi à construire un test non asymptotique nous permettant de déterminer si deux neurones sont indépendants ou non sur une plage de temps $[a, b]$ sans formuler d'hypothèse sur leur activité. Dans le but de vraiment répondre à la problématique qui est de savoir à quels moments il y a rupture d'indépendance entre les neurones, la première perspective pour la fin du stage sera donc d'appliquer la théorie des tests multiples : on commencera par diviser la plage de temps correspondant à

l'expérience en petits intervalles et on regardera sur lesquels on rejette l'hypothèse nulle. L'étape naturelle qui suit est donc de pouvoir tester les vraies données fournies par Franck Grammont, neurophysiologiste au Laboratoire Jean-Alexandre Dieudonné.

D'un point de vue simulations, dans un but de comparer nos résultats à ceux des méthodes déjà existantes, pour l'instant, on a simulé les données selon des processus de Poisson homogènes. On pourra donc regarder ce qui se passe lorsque les données sont simulées d'abord suivant des processus de Poisson inhomogènes, et ensuite pour des processus ponctuels plus généraux, tels que les processus de Hawkes.

Pour ce qui est des perpectives plus éloignées, on souhaite étudier la puissance du test construit d'un point de vue non asymptotique, et regarder ce qui se passe pour les méthodes plus générales de bootstrap à poids. L'idée serait de trouver une sorte de distance entre la loi jointe et le produit des marginales telle que lorsqu'elle est supérieure à un certain seuil, on ait un contrôle de la puissance. On espère ainsi pouvoir comprendre ce qui se passe sous certaines alternatives.

Finalement, on pourra regarder si d'un point de vue théorique, on peut déduire de nouveaux tests pour des cas plus particuliers, comme par exemple des tests de nullité de fonctions d'interaction pour les processus de Hawkes.

Annexes

A Tests statistiques ; quelques rappels

Definitions. Voici les définitions de base concernant la théorie des tests statistiques.

Définition (Modèle statistique)**.**
Un modèle statistique est la donnée de

$$\{E, \mathcal{E}, (\mathbb{P}_\theta, \theta \in \Theta)\}.$$

avec
- E l'espace où vivent les observations,
- $\mathcal{E}$ une tribu sur E,
- $(\mathbb{P}_\theta, \theta \in \Theta)$ une famille de lois sur E paramétrées par θ.

Définition (Statistique)**.**
Une statistique T est une fonction mesurable des observations.

Définition (Test statistique)**.**
On suppose qu'on ait un modèle comme ci-dessus tel que Θ soit partagé en Θ_0 et Θ_1 avec $\Theta_0 \cap \Theta_1 = \varnothing$.

Tester "$(H_0) : \theta \in \Theta_0$" contre "$(H_1) : \theta \in \Theta_1$" au niveau α, revient à trouver une statistique de test Δ à valeurs dans $\{0, 1\}$ telle que

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta (\Delta = 1) \leq \alpha.$$

On appelle (H_0) l'hypothèse nulle et (H_1) l'alternative.
Δ est une fonction des observations de la forme $\mathbb{1}_{\{R\}}$. On appelle R la zone de rejet du test :
- si $\Delta = 1$, on rejette l'hypothèse nulle,
- et si au contraire $\Delta = 0$, on accepte l'hypothèse nulle.

Erreurs d'un test. On a deux moyens de calculer le risque de se tromper :

- Erreur de première espèce : elle donne le risque de rejeter l'hypothèse nulle à tort :

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta (\Delta = 1).$$

- Erreur de deuxième espèce : elle donne le risque de ne pas rejeter l'hypothèse nulle alors qu'elle est fausse :

$$\sup_{\theta \in \Theta_1} \mathbb{P}_\theta (\Delta = 0).$$

P-valeur d'un test.

Définition.
La p-valeur d'un test Δ_α est définie à partir des observations X par :

$$\hat{\alpha} = \sup \{\alpha \ / \ \Delta_\alpha(X) = 0\}.$$

C'est la plus grande valeur de α pour laquelle au vue de nos observations, on accepte l'hypothèse nulle (H_0).

On voit bien que la p-valeur nous permet de savoir si on accepte ou si on rejette le test. En effet, on voit bien que si la p-valeur $\hat{\alpha}(X) \leq \alpha$, alors on va rejetter le test (puisque le seuil α qu'on s'est fixé au départ est trop grand pour qu'on accepte (H_0) étant donné les observations). D'un autre côté, si $\hat{\alpha}(X) > \alpha$, on aura plutôt tendance à accepter l'hypothèse nulle. Ainsi, on peut dire que la p-valeur correspond en quelque sorte à la "crédibilité" de (H_0) au vue des observations.

B Rappels sur les inverses généralisés

Définition.

Soit $F : \mathbb{R} \to [0,1]$ une fonction croissante continue à droite et de limite finie à gauche (càdlàg) telle que $\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$.
On définit son inverse de Lévy par :

$$\forall t \in [0,1],\ F^{-1}(t) = \inf\{x \in \mathbb{R}; F(x) \geq t\}.$$

Lemme 3.

Soit F une fonction croissante càdlàg telle que $\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$.
Alors pour tous $x, y \in \mathbb{R}$,

$$F^{-1}(x) \leq y \quad \Leftrightarrow \quad F(y) \geq x.$$

Preuve :

$\Leftarrow$ Si $F(y) \geq x$, alors $y \in \{u; F(u) \geq x\}$ donc $y \geq \inf\{u; F(u) \geq x\} = F^{-1}(x)$.

$\Rightarrow$ Deux cas se présentent :

- Si $y > F^{-1}(x) = \inf\{u; F(u) \geq x\}$, alors $\exists v \in \{u; F(u) \geq x\}$ tel que $y > v$.
  Et par croissance de F, on obtient bien que $F(y) \geq F(v) \geq x$.
- Si $y = F^{-1}(x) = \inf\{u; F(u) \geq x\}$, alors $\exists (v_n)_n \subset \{u; F(u) \geq x\}$ telle que $\begin{cases} \forall n,\ v_n > y, \\ v_n \to y. \end{cases}$
  Par continuité de F à droite, $F(v_n) \to F(y)$ et comme $\forall n,\ F(v_n) \geq x$ par passage à la limite, $F(y) \geq x$.

□

Propriété 1.

Soit F fonction croissante càdlàg telle que $\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$.

$$\forall t \in [0,1],\ F(F^{-1}(t)) \geq t.$$

Démonstration : On applique le lemme (3) pour $x = t$ et $y = F^{-1}(t)$. □

C Rappels sur les fonctions de répartition et quantiles empiriques

Définition.

La fonction de répartition d'une variable aléatoire réelle X est la fonction F_X définie par

$$F_X : \begin{pmatrix} \mathbb{R} & \longrightarrow & [0,1] \\ x & \longmapsto & \mathbb{P}(X \leq x) \end{pmatrix}.$$

La fonction de répartition caractérise la loi de X.

En outre, on peut montrer que quelle que soit la variable aléatoire réelle X, $F_X : \mathbb{R} \to [0,1]$ est une fonction croissante càdlàg telle que $\lim_{-\infty} F_X = 0$ et $\lim_{+\infty} F_X = 1$. On peut en particulier calculer son inverse de Lévy.

Définition.

Si la loi μ a pour fonction de répartition F, on définit ses quantiles par :

$$q : \begin{pmatrix} [0,1] & \longrightarrow & \mathbb{R} \\ \alpha & \longmapsto & \inf\{t \in \mathbb{R}\ ;\ F(t) \geq \alpha\} \end{pmatrix}.$$

C'est, en d'autres termes, l'inverse de Lévy de la fonction F.

Définition.

Ayant un échantillon $(X_k)_{1\leq k\leq n}$ i.i.d de loi μ, on définit la fonction de répartition empirique par

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k\leq x\}}.$$

Celle-ci étant constante par morceaux, on voit bien que les quantiles empiriques sont donc définis par :

$$\hat{q}_n(\alpha) = X_{(\lceil n\alpha\rceil)} \begin{cases} X_{(n\alpha)} & \text{si } n\alpha \in \mathbb{N} \\ X_{([n\alpha]+1)} & \text{si } n\alpha \notin \mathbb{N}. \end{cases},$$

où l'on note avec les parenthèses $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ nos variables aléatoire ré-ordonnées, appelées statistiques d'ordre.

Théorème 6 (Glivenko-Cantelli**).**

Soit $X_1, \ldots, X_n$ n variables aléatoires réelles i.i.d. de fonction de répartition F.
Notons $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i\leq x\}}$ la fonction de répartition empirique associée. Alors,

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n\to+\infty]{p.s} 0.$$

D Rappels sur les processus de Poisson homogènes

Définition (Première définition d'un processus de Poisson homogène**).**

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

a) N(0)=0,

b) le processus est à accroissements indépendants,

c) le nombre de points se produisant dans un intervalle de temps de longueur $t \geq 0$ suit une loi de Poisson de paramètre λt, *i.e.*

$$\forall s, t \geq 0, \forall k \in \mathbb{N},\ \mathbb{P}\left(N(s+t) - N(s) = k\right) = e^{-\lambda t}\frac{\lambda^k}{k!}.$$

Les processus de Poisson peuvent être caractérisés comme suit :

Proposition 4 (Construction de processus d'un Poisson homogène**).**

Soit $(T_n)_{n\geq 0}$ une suite croissante de variables aléatoires réelles positives telles que $(T_1, T_2 - T_1, \ldots, T_n - T_{n-1}, \ldots)$ soient indépendants et identiquement distribués selon la loi exponentielle de paramètre λ. On lui associe le processus de comptage $\{N(t); t \geq 0\}$ avec $N(t) = \sum_{n\geq 1} \mathbb{1}_{\{T_n\leq t\}}$. Alors $\{N(t); t \geq 0\}$ est un processus de Poisson homogène d'intensité λ.

En outre la réciproque est vraie dans le sens où, si $\{N(t); t \geq 0\}$ est un processus de Poisson défini comme ci-dessus, alors les sauts $(X_n)_{n\geq 0}$ vérifient les hypothèse de la proposition 4 (où X_n est le $n^{\text{ème}}$ point du processus).

Propriété 2.

Soit $N(t) \sim \mathcal{P}(\lambda t)$ et supposons que conditionnellement à $\{N(t) = n\}$, $S_1, \cdots, S_n$ soient des variables aléatoires indépendantes, de loi uniforme sur le segment $[0, t]$. Alors le processus de comptage associé à ces points est un processus de Poisson homogène d'intensité λ sur $[0, t]$.

Démonstration : Notons $N_b - N_a$ le nombre de S_i apparaissant dans $]a, b]$.

1ère étape Montrons que $\forall\, 0 = r_0 < r_1 < \cdots < r_k = t$ subdivision de $[0, t]$ et $\forall\, (n_j)_{1\leq j\leq k}$ / $\sum_{j=1}^{k} n_j = n$, on a :

$$\mathbb{P}\left(N_{r_j} - N_{r_{j-1}} = n_j\ ;\ j \in [\![1; k]\!]\right) = \prod_{j=1}^{k} \frac{e^{-\lambda(r_j - r_{j-1})}[\lambda(r_j - r_{j-1})]^{n_j}}{n_j!}.$$

Soit $0 = r_0 < r_1 < \cdots < r_k = t$ une subdivision quelconque de $[0,t]$ et $(n_j)_{1\leq j\leq k}$ / $\sum_{j=1}^{k} n_j = n$. Alors

$$\mathbb{P}\left(\underbrace{N(r_j) - N(r_{j-1}) = n_j}_{n_j \text{ points dans }]r_{j-1},r_j]} \ ; \ j \in [\![1;k]\!] \mid N(t) = n\right) = n! \prod_{i=1}^{k} \frac{1}{n_j!} \left(\frac{r_j - r_{j-1}}{t}\right)^{n_j}.$$

En effet, puisque $S_1, \cdots, S_n$ i.i.d $\sim \mathcal{U}([0,t])$, alors, conditionnellement à $\{N(t) = n\}$, $(N_{r_1} - N_{r_0}, \cdots, N_{r_k} - N_{r_{k-1}})$ forme un vecteur de loi multinomiale de paramètres :

- $\star$ n (nombre de tirages indépendants) avec $\sum_{i=1}^{k} n_i = n$;
- $\star$ $p_j = \frac{r_j - r_{j-1}}{t}$ $(1 \leq j \leq k)$ (probabilité de la $j^{\text{ème}}$ issue n_j).

Pour se débarrasser du conditionnement, on multiplie par $\mathbb{P}(N(t) = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$.
Ainsi, on a :

$$\begin{aligned}
\mathbb{P}(N(r_j) - N(r_{j-1}) = n_j \ ; \ j \in \{1, n \ldots, k\}) &= \mathbb{P}\left(N(t) = \sum_{j=1}^{k} n_j\right).\mathbb{P}(N(r_j) - N(r_{j-1}) = n_j \ ; \ j \in [\![1;k]\!] \mid N(t \\
&= e^{-\lambda t}\frac{(\lambda t)^{(\sum_{j=1}^{k} n_j)}}{\not{n}!} \cdot \not{n}! \prod_{j=1}^{k} \frac{1}{n_j!}\left(\frac{r_j - r_{j-1}}{t}\right)^{n_j} \\
&= e^{-\lambda[\sum_{j=1}^{k}(r_j - r_{j-1})]} \prod_{j=1}^{k} \frac{(\lambda \not{t})^{n_j}}{n_j!} \frac{(r_j - r_{j-1})^{n_j}}{\not{t}^{n_j}} \\
&= \prod_{j=1}^{k} \frac{e^{-\lambda(r_j - r_{j-1})}[\lambda(r_j - r_{j-1})]^{n_j}}{n_j!}.
\end{aligned}$$

et ce $\forall\ 0 = r_0 < r_1 < \cdots < r_k = t$ subdivision de $[0,t]$, d'où la première étape.

2ème étape Le processus de comptage $(N_s)_{0\leq s\leq t}$ est un processus de Poisson homogène d'intensité λ.

- $N(0) = 0$ car $T_1 > 0$ presque sûrement d'où a).
- Par la première étape, $\forall\ u, v \geq 0$ tels que $u+v < t$ (en prenant $k = 2, r_0 = 0, r_1 = u, r_2 = u+v$), on a $\mathbb{P}(N(u) = l, N(u+v) - N(u) = m) = e^{-\lambda s}\frac{(\lambda u)^l}{l!} \cdot e^{-\lambda v}\frac{(\lambda v)^m}{m!}$
  et en sommant sur les l, on obtient :

$$\begin{aligned}
\mathbb{P}(N(u+v) - N(u) = m) &= \sum_{l\in\mathbb{N}} \mathbb{P}(N(u) = l, N(u+v) - N(u) = m) \\
&= \sum_{l\in\mathbb{N}} [e^{-\lambda u}\frac{(\lambda u)^l}{l!}] \cdot e^{-\lambda v}\frac{(\lambda v)^m}{m!} \\
&= e^{-\lambda v}\frac{(\lambda v)^m}{m!}.
\end{aligned}$$

  Finalement, $\forall s, t > 0,\ N(u+v) - N(u)$ suit une loi de Poisson de paramètre λv d'où c).
- Et de plus, on obtient que $\mathbb{P}\left(\bigcap_{1\leq j\leq k}\{N(r_j) - N(r_{j-1}) = n_j\}\right) = \prod_{j=1}^{k} \mathbb{P}(N(r_j) - N(r_{j-1}) = n_j)$.
  On en déduit donc que le processus est bien à accroissements indépendants, d'où b).

□

E Rappels sur la distance de Wasserstein

Dans toute cette partie, $(E, \|.\|)$ représente un espace de Banach séparable, et $\mathcal{P}(E)$ l'ensemble des mesures de probabilité sur E. On note

$$\Gamma_2 = \left\{ \gamma \in \mathcal{P}(E) \ / \ \int_X \|x\|^2 \gamma(dx) < +\infty \right\}$$

Définition.

On définit la distance Wasserstein entre deux mesures μ et $\nu \in \mathcal{P}(E)$ par

$$d_2(\mu, \nu) = \sqrt{\inf_{\gamma \in Adm(\mu,\nu)} \int_{E^2} \|x - y\|^2 d\gamma(x, y)},$$

où $Adm(\mu, \nu)$ désigne l'ensemble des mesures de probabilité sur E^2 de marginales μ et ν.
Dans une vision plus "probabiliste", on a

$$d_2(\mu, \nu) = \sqrt{\inf_{(X,Y) \ / \ X \sim \mu, Y \sim \nu} \mathbb{E}\left[\ \|X - Y\|^2 \ \right]}$$

Voici deux propositions importantes sur la distance de Wasserstein que nous admettrons ici.

Proposition 5.

Avec les mêmes notations que précédemment,

a) *Pour tous μ et ν probabilités sur E, l'infimum de la distance de Wasserstein entre μ et ν est atteint.*

b) *d_2 est une distance sur Γ_2.*

Proposition 6.

Soit α_n et $\alpha \in \mathcal{P}(E)$. Alors

$$d_2(\alpha_n, \alpha) \xrightarrow[n \to +\infty]{} 0 \quad \underline{ssi} \quad \alpha_n \rightharpoonup \alpha \text{ et } \int_E \|x\|^2 d\alpha_n(x) \xrightarrow[n \to +\infty]{} \int_E \|x\|^2 d\alpha(x)$$

D'autre part, on peut étendre cette définition aux variables aléatoires :

Définition.

Soit U et V des variables aléatoires à valeurs dans E. On note $\alpha \in \mathcal{P}(E)$ la loi de U et $\beta \in \mathcal{P}(E)$ celle de V. On pose

$$\begin{aligned} d_2(U, V) &= d_2(\alpha, \beta) \\ &= \sqrt{\inf_{X \sim U, Y \sim V} \mathbb{E}\left[\ \|X - Y\|^2 \ \right]} \end{aligned}$$

Propriété 3.

Avec les mêmes notations que précédemment, quelles que soient U et V des variables aléatoires à valeurs dans E, et pour tout scalaire a, on a

$$d_2(aU, aV) = |a| \ d_2(U, V)$$

Démonstration : On voit bien d'abord que si $a = 0$, le résultat est immédiat. Supposons à présent que $a \neq 0$. On remplace par la définition :

$$\begin{aligned} d_2(aU, aV) &= \sqrt{\inf_{X\sim aU, Y\sim aV} \mathbb{E}\,[\ \|X-Y\|^2\ ]} \\ &= \sqrt{\inf_{\tilde{X}\sim U, \tilde{Y}\sim V} \mathbb{E}\left[\ \|a\tilde{X}-a\tilde{Y}\|^2\ \right]} \\ &= |a|\sqrt{\inf_{\tilde{X}\sim U, \tilde{Y}\sim V} \mathbb{E}\left[\ \|\tilde{X}-\tilde{Y}\|^2\ \right]} \\ &= |a|\ d_2(U,V) \end{aligned}$$

□

Propriété 4.

On suppose ici que E est l'espace des réels $\mathbb{R}$ muni de sa norme usuelle $|\cdot|$.
Soit $(U_j)_{1\leq j\leq m}$ (respectivement $(V_j)_{1\leq j\leq m}$) des variables aléatoires i.i.d. de lois respectives supposées dans Γ_2 et telles que $\mathbb{E}\,[\ U_j\ ] = \mathbb{E}\,[\ V_j\ ]$. Alors,

$$d_2(\sum_{j=1}^m U_j, \sum_{j=1}^m V_j) \leq \sum_{i=1}^m d_2(U_j, V_j)$$

Démonstration : Commençons par considérer les variables aléatoires $X_1, \ldots, X_m$ i.i.d. de même loi que les U_j et $Y_1, \ldots, Y_m$ i.i.d. de même loi que les V_j telles que les couples $\{(X_j, Y_j)\}$ soient indépendants et optimaux pour $d_2(U_j, V_j)$.

En particulier, les m-uplets $(U_1, \ldots, U_m)$ et $(X_1, \ldots, X_m)$ ont même lois donc $\sum_{j=1}^m U_j \sim \sum_{j=1}^m X_j$. Et de même, $\sum_{j=1}^m V_j \sim \sum_{j=1}^m Y_j$.

La distance d_2 sur les variables aléatoires étant un infimum sur les couples ayant les mêmes marginales, on a :

$$\begin{aligned} d_2^2\left(\sum_{j=1}^m U_j, \sum_{j=1}^m V_j\right) &\leq \mathbb{E}\left[\ \left|\sum_{j=1}^m X_j - \sum_{j=1}^m Y_j\right|^2\ \right] \\ &= \mathbb{E}\left[\ \left(\sum_{j=1}^m (X_j - Y_j)\right)^2\ \right] \end{aligned}$$

Or

$$\left(\sum_{j=1}^m (X_j - Y_j)\right)^2 = \sum_{j=1}^m (X_j - Y_j)^2 + 2 * \sum_{j\neq k} (X_j - Y_j)(X_k - Y_k).$$

Mais par indépendance des couples (X_j, Y_j), on a $\forall j \neq k,\ (X_j - Y_j) \perp\!\!\!\perp (X_k - Y_k)$ d'où

$$\mathbb{E}\,[\ (X_j - Y_j)(X_k - Y_k)\ ] = \mathbb{E}\,[\ X_j - Y_j\ ]\,\mathbb{E}\,[\ X_k - Y_k\ ] = 0,$$

car $\mathbb{E}\,[\ X_j - Y_j\ ] = \mathbb{E}\,[\ X_j\ ] - \mathbb{E}\,[\ Y_j\ ] = \mathbb{E}\,[\ U_j\ ] - \mathbb{E}\,[\ V_j\ ] = 0$.

Ainsi, tous les doubles produits sont d'espérance nulle, d'où

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j=1}^{m}(X_j-Y_j)\right)^2\right] &= \mathbb{E}\left[\sum_{j=1}^{m}(X_j-Y_j)^2\right]\\ &= \sum_{j=1}^{m}\mathbb{E}\left[(X_j-Y_j)^2\right]\\ &= \sum_{j=1}^{m}d_2(U_j,V_j),\end{aligned}$$

par définition des couples (X_j, Y_j), d'où la propriété 4. □

F Programmes sous Scilab

F.1 Simulation des données

Voici le programme qui nous a permis de simuler les données des deux neurones à tester selon des processus de Poisson homogènes.
Les paramètres d'entrée `a` et `b` délimitent la plage de temps sur laquelle on travaille, `lambda` est l'intensité du processus de Poisson homogène modélisant les spikes, et `nessai` représente le nombre de fois qu'est effectuée l'expérience.
Cette fonction renvoie deux matrices (chacune correspondant à un neurone) comprenant sur chaque ligne, le nombre de spikes par essai dans la première colonne, puis l'emplacement de ces spikes (complétée par des zéros).

```
function Y=simulPP(a,b,lambda,nessai)
ntops=grand(nessai,2,'poi',lambda*(b-a));

s1=max(ntops(:,1));
s2=max(ntops(:,2));

M1=zeros(nessai,s1+1);
M2=zeros(nessai,s2+1);

M1(:,1)=ntops(:,1);
M2(:,1)=ntops(:,2);

for i=1:nessai
if ntops(i,1)<>0 then
M1(i,2:ntops(i,1)+1)=gsort(grand(1,ntops(i,1),'unf',a,b),'g','i');
end
if ntops(i,2)<>0 then
M2(i,2:ntops(i,2)+1)=gsort(grand(1,ntops(i,2),'unf',a,b),'g','i');
end
end
Y=list(M1,M2)
endfunction
```

F.2 Calcul des coïncidences

On commence par regarder une fonction qui compte le nombre de coïncidences (de paramètre `delta`) entre un essai `X` du neurone 1 et un `Y` du neurone 2.

```
function count=count_coincidences(X,Y,delta)
count = 0; // nombre de coïncidences
```

```
//=================================
// on récupère les données (sans les zéros à la fin)
//=================================
Nx = X(1); // nombre de spikes dans X
Ny = Y(1); // nombre de spikes dans Y
NxMax = Nx+1; // indice max pour X
NyMax = Ny+1; // indice max pour Y
j = 2; // indice initial pour la boucle dans Y

//=================================
// boucle extérieure : pour chaque point de X
//=================================
for i=2:NxMax
Xi = X(i); // valeur courante de X
Xlo = Xi-delta; // borne inférieure courante
//---------------
// on cherche la première valeur de Y supérieure ou égale à la borne inférieure
//---------------
while( (j<=NyMax) & (Y(j)<Xlo) )
j=j+1;
end

//----------------
// a-t-on besoin de faire le test ?
//----------------
if(j>NyMax) then break end;
//----------------
// si oui, on test à partir de l'indice j
//----------------
Xup = Xi + delta; // borne supérieure courante
k = j;
while( (k<=NyMax) & (Y(k)<=Xup) )
count = count+1;
k = k+1;
end

end
endfunction
```

Ensuite, on calcule le nombre total de coïncidences. Les paramètres d'entrée sont : `data1` et `data2` les matrices des spikes de chaque neurone, `delta` le paramètre définissant les coïncidences et un vecteur `permut` correspondant à la permutation du second neurone.

```
function count=coincidence_total(data1,data2,delta,permut)
ntrials=size(data1,1);
count=0;
for i=1:ntrials
count=count+count_coincidences(data1(i,:),data2(permut(i),:),delta);
end
endfunction
```

F.3 Création de l'échantillon de la statistique calculée sur les permutées

On crée à présent notre B-échantillon de la loi conditionnelle $T(\mathcal{S}, L \circ \sigma_n) \mid \mathcal{S}, L$. (On garde les mêmes notations pour les données d'entrée que pour les fonctions précédentes.)

```
function ech=permutation_bootstrap(data1,data2,delta,B)
ntrials=size(data1,1); // nombre d'essais
ech=zeros(1,B); // échantillon permuté
for b=1:B
permut = grand(1,'prm',(1:ntrials)'); // permutation tirée uniformément
ech(b) = coincidence_total(data1,data2,delta,permut);
end
endfunction
```

F.4 Calcul de la p-valeur

Ensuite, on regarde la fonction qui va calculer la p-valeur du test par permutation (toujours avec les mêmes notations que précédemment pour les varibles d'entrée) :

```
function p=permutation_pvalue(data1,data2,delta,B)
ntrials=size(data1,1);

L=[1:ntrials]; // la permutation identité

True_Coinc=coincidence_total(data1,data2,delta,L);
ech=permutation_bootstrap(data1,data2,delta,B);

p=sum(ech>=True_Coinc)/B;

endfunction
```

F.5 Test par permutation et niveau empirique

Programme du test par permutation. On rejette l'hypothèse nulle lorsque la p-valeur est inférieure au niveau `alpha` $= \alpha$ fixé au départ.

```
function t=permutation_test(data1,data2,delta,B,alpha)
p=permutation_pvalue(data1,data2,delta,B);
t=sum(p<=alpha);
endfunction
```

Niveau empirique. On regarde la proportion de rejets à tort lorsqu'on répète `Nsimu` fois le test en simulant à chaque fois une nouvelle donnée (vérifiant l'hypothèse nulle).

```
function niveau=level(a,b,lambda,nessai,delta,B,alpha,Nsimu)
p=zeros(1:Nsimu)
for n=1:Nsimu
donnees=simulPP(a,b,lambda,nessai);
p(n)=permutation_pvalue(donnees(1),donnees(2),delta,B);
end
niveau=sum(p<=alpha)/Nsimu;
endfunction
```

Références

[1] M.A. Arcones and E. Giné. *On the Bootstrap of U and V Statistics*. The Annals of Statistics, Vol. 20, No. 2, p. 655-674, 1992.

[2] S. Arlot, G. Blanchard and E. Roquain. *Resampling-Based Confidence Regions and Multiple Tests for a Correlated Random Vector*. COLT'07 Proceedings of the 20th annual conference on Learning theory, p. 127-141, 2007.

[3] P.J. Bickel and D.A. Freedman. *Some Asymptotic Theory for the Bootstrap*. The Annals of Statistics, Vol. 9, No. 6, p. 1196-1217, 1981.

[4] J. Bretagnolle. *Lois limites du Bootstrap de certaines fonctionnelles*. Annales de l'I. H. P., section B, tome 19, no 3, p. 281-296, 1983.

[5] B. Efron. *Bootstrap Methods : Another Look at the Jakknife*. The Annals of Statistics, Vol. 7, No. 1, p. 1-26, Janvier 1979.

[6] R.A. Fisher. *The Design of Experiment*. Oxford, England : Oliver & Boyd, 1935.

[7] A. Gretton, K. Fukumizu, C.-H. Teo, L. Song, B. Schölkopf, and A. Smola. *A Kernel Statistical Test of Independence*. In Advances in Neural Information Processing Systems 20, p. 585-592, Cambridge, MA, 2008. MIT Press, 2008.

[8] A. Gretton and L. Györfi. *Consistent Nonparametric Tests od Independence*. Journal of Machine Learning Research 11, p. 1391-1423, 2010.

[9] S. Grün, M. Diesmann, F. Grammont, A. Riehle, A. Aertsen. *Detecting unitary events without discretization in time*. Journal of Neurosciences methods 94, 1999.

[10] W. Hoeffding. *A Non-Parametric Test of Independence*. The Annals of Mathematical Statistics, Vol. 19, No. 4, p. 546-557, 1948.

[11] A. Janssen and T. Pauls. *How to do Bootstrap and Permutation Tests Work ?*. The Annals of Statistics, Vol. 31, No. 3, p. 768-806, 2003.

[12] J.T. Præstegaard and J.A. Wellner. *Exchangeably weighted bootstraps of the general empirical process*. The Annals of Probability, Vol. 21, No. 4, 1993.

[13] M.H. Quenouille. *Approximate tests of correlation in time series*. Journal of the Royal Statistical Society, Series B 11, p. 68-84, 1949.

[14] J.P. Romano. *Bootstrap and Randomization Tests of some Nonparametric Hypotheses*. The Annals of Statistics, Vol. 17, No. 1, p. 141-159, Mars 1989.

[15] A. Rouis. *Tests mutliples d'indépendance pour des données neuronales unitaires relatives à une tâche senseorimotrice*. Mémoire de stage de Master 2 IMEA, Université de Nice Sophia-Antipolis, 2011.

[16] C. Suquet *Simultation*. Cours Agrégation Externe, Université Lille 1, 2005 - 2006.

[17] S. Taskinen, H. Oja and R.H. Randles. *Multivariate Nonparametric Tests of Independence*. Journal of the American Statistical Association, Vol. 100, No. 471, p. 916-925, Septembre 2005