

Les séquences de potentiels d'action neuronaux
comme observations de processus de
comptage : estimation de l'intensité par
vraisemblance pénalisée

Christophe Pouzat

GTS vendredi 17 juin 2011

Outline

Les données

Régime non-homogène

Intensité conditionnelle

Transformation du temps

Un test basé sur le théorème de Donsker

Estimation de λ

Retour aux données

Où en est-on ?

Les données

Régime non-homogène

Intensité conditionnelle

Transformation du temps

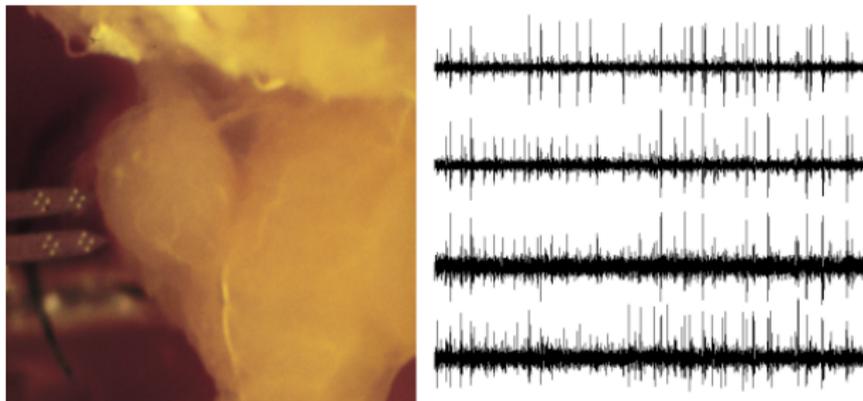
Un test basé sur le théorème de Donsker

Estimation de λ

Retour aux données

L'origine des données

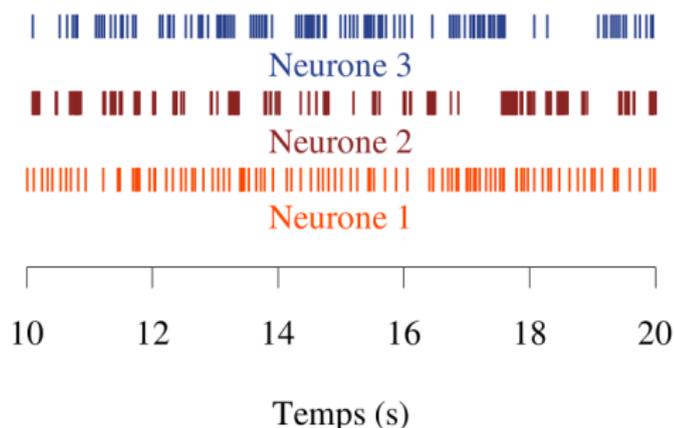
Vue de l'extérieur, l'activité des neurones se manifeste par l'émission d'impulsions électriques très brèves : **les potentiels d'action**.



A gauche, le cerveau et la sonde d'enregistrement qui comporte 16 électrodes (les points brillants). La largeur d'une branche de la sonde est de $80 \mu m$. A droite, 1 sec d'enregistrement sur 4 électrodes. Les pics sont des potentiels d'action.

Exemple de séquences de potentiels d'action

Grappe en raster



Après une étape de pré-traitement « assez lourde » appelée **tri des potentiels d'action** (PAs), on obtient les séquences ou trains de PAs.

Pourquoi et comment modéliser les trains de PAs ?

- ▶ Une hypothèse de travail centrale en neurosciences est que les temps d'apparition des PAs, par opposition à leurs formes, sont le seul support de transmission de l'information entre régions du cerveau.

Pourquoi et comment modéliser les trains de PAs ?

- ▶ Une hypothèse de travail centrale en neurosciences est que les temps d'apparition des PAs, par opposition à leurs formes, sont le seul support de transmission de l'information entre régions du cerveau.
- ▶ Cette hypothèse légitime l'étude des trains de PAs en tant que séquences de points sur la demi-droite réelle (représentant le temps) sans nécessairement tenir compte des mécanismes biophysiques qui les génèrent.

Pourquoi et comment modéliser les trains de PAs ?

- ▶ Une hypothèse de travail centrale en neurosciences est que les temps d'apparition des PAs, par opposition à leurs formes, sont le seul support de transmission de l'information entre régions du cerveau.
- ▶ Cette hypothèse légitime l'étude des trains de PAs en tant que séquences de points sur la demi-droite réelle (représentant le temps) sans nécessairement tenir compte des mécanismes biophysiques qui les génèrent.
- ▶ Dans ce qui suit nous allons assimiler un train de PAs à un **processus ponctuel** auquel nous allons associer un **processus de comptage** (ou compteur).

Processus de comptage ou compteur

- ▶ À des points $\{t_i\}_{i \in \mathbb{N}^*} \subset \mathbb{R}^{+*}$, tels que $t_j > t_i$ pour $j > i$, on associe le **processus de comptage** ou **compteur** défini par :

$$N(t) \equiv \#\{t_i \mid t_i \leq t\}$$

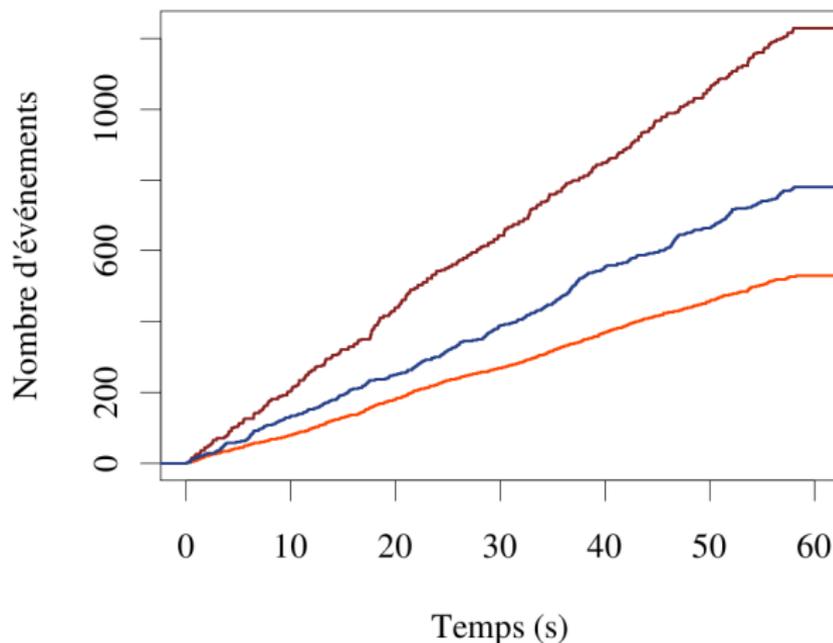
- ▶ Lorsque plusieurs neurones sont enregistrés simultanément on introduit un compteur, $N^{(k)}(t)$, par neurone. Ici l'exposant (k) est un indice de neurone. Alors pour que le processus multidimensionnel, $(N^{(1)}(t), \dots, N^{(K)}(t))$, reste « simple », on rajoute la condition :

$$t_i^{(k)} \neq t_j^{(l)}, \quad \forall i, j \in \mathbb{N}^* \quad \text{et} \quad \forall k \neq l, \quad k, l \in \{1, \dots, K\}$$

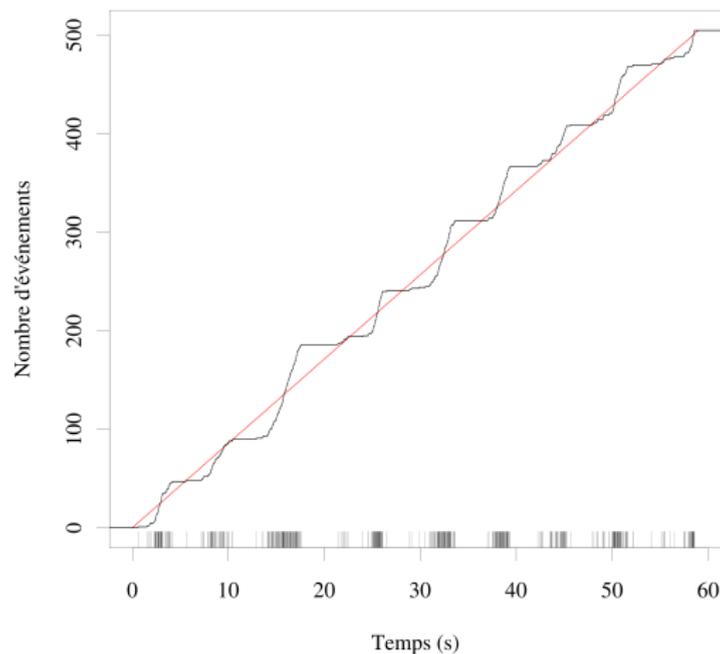
où $t_i^{(k)}$ est le temps du i -ème PA du neurone k .

Exemples de compteurs observés : régime spontané

La « dynamique de décharge » change d'un neurone à l'autre.

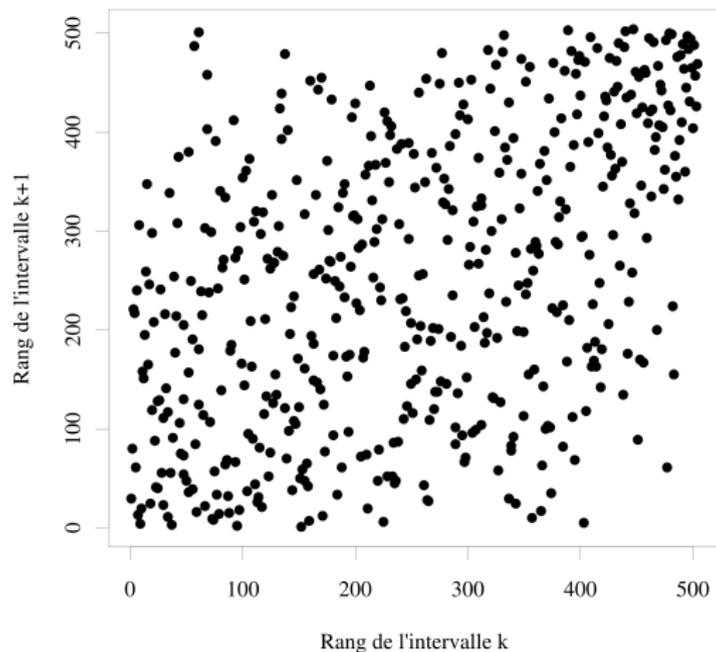


Un cas compliqué (1)



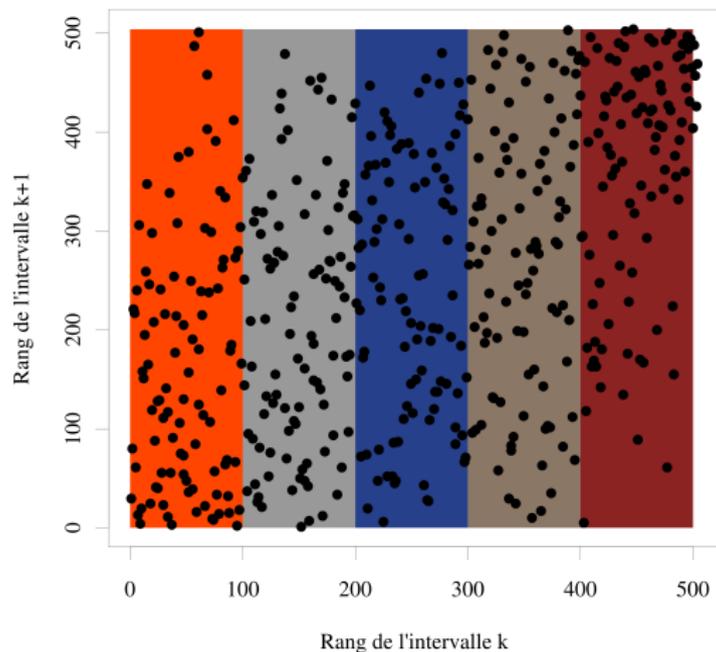
L'espérance d'un processus de Poisson de même fréquence moyenne est figurée en rouge (pointillés). Elle est utilisée comme « test » de stationnarité.

Un cas compliqué (2)



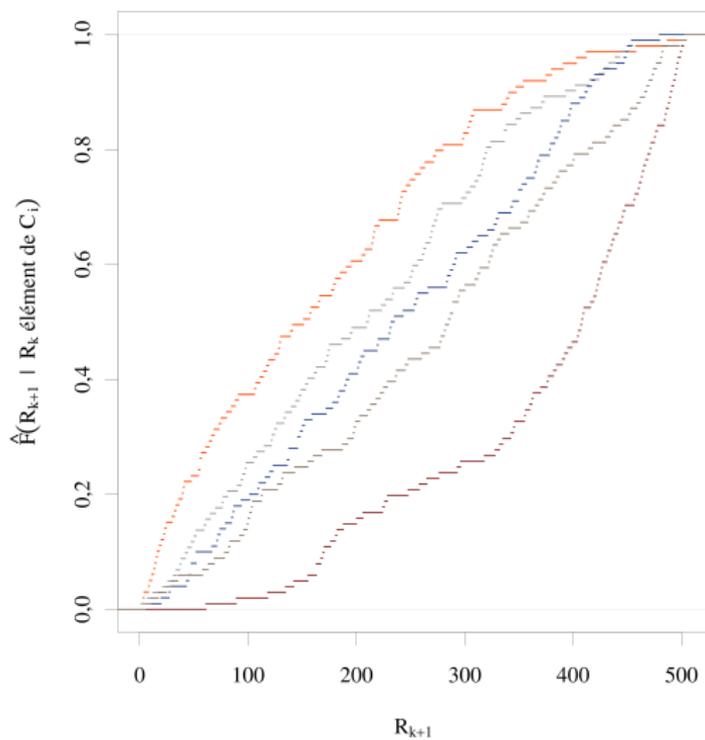
Les rangs des intervalles entre PAs successifs sont corrélés.

Un cas compliqué (2')



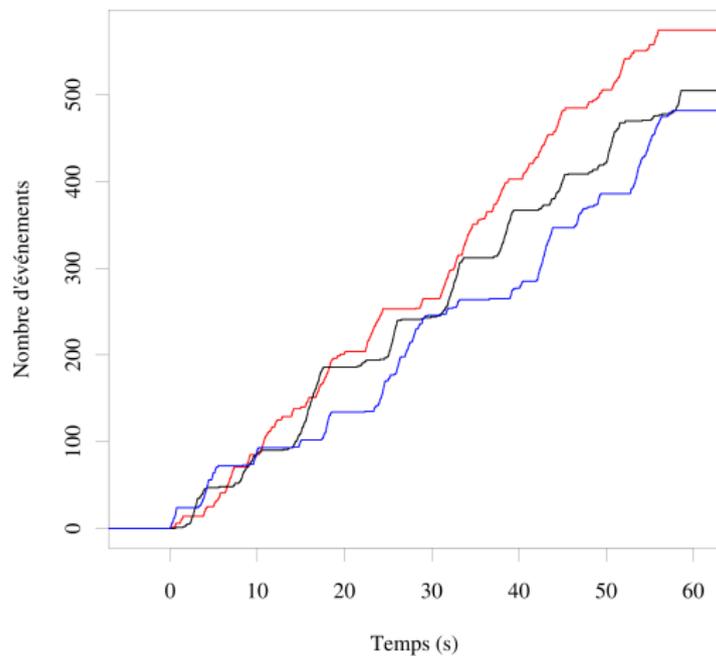
Les rangs des intervalles entre PAs successifs sont corrélés.

Un cas compliqué (3)



Fonctions de répartition empiriques du rang de l'intervalle $k+1$ suivant la classe à laquelle le rang de l'intervalle k appartient. Une partition en 5 classes, C_i , de tailles (presques) égales a été construite. Des bandes de confiance à 95 % auraient ici une « épaisseur » de 0,14.

Un cas compliqué (4)

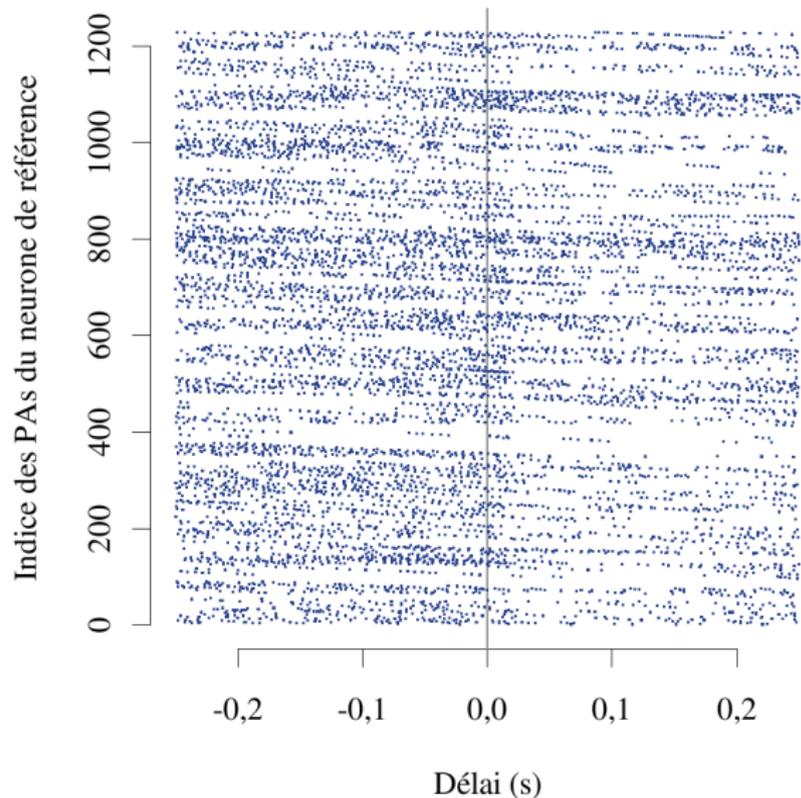


Un processus de renouvellement n'est pas un modèle adéquate. En rouge, le compteur observé obtenu par ré-échantillonnage de *bootstrap* des intervalles entre PAs. En bleu, le compteur observé obtenu par ré-échantillonnage de *bootstrap* « stratifié » des intervalles entre PAs – stratification suivant la partition précédente en 5 classes.

Premier bilan

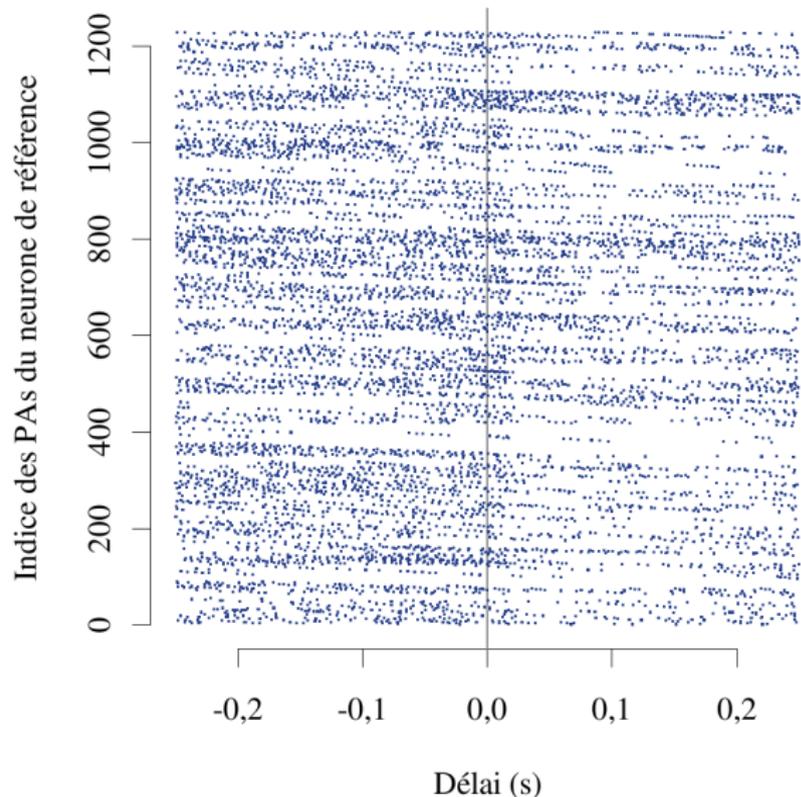
Même pour des neurones considérés isolément en régime stationnaire (homogène), un processus de renouvellement va s'avérer inadéquate, en général, pour modéliser la décharge (séquence de potentiels d'actions).

Interactions entre neurones



Temps des PAs d'un neurone test (neurone 3) relatifs aux temps des PAs d'un neurone de référence (neurone 2).

Interactions entre neurones



Temps des PAs d'un neurone test (neurone 3) relatifs aux temps des PAs d'un neurone de référence (neurone 2).
Conclusion : notre modèle de décharge devrait être capable de prendre en compte des interactions.

Où en est-on ?

Les données

Régime non-homogène

Intensité conditionnelle

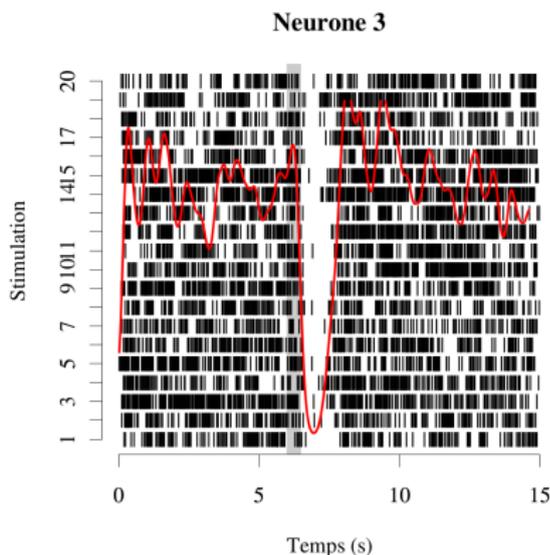
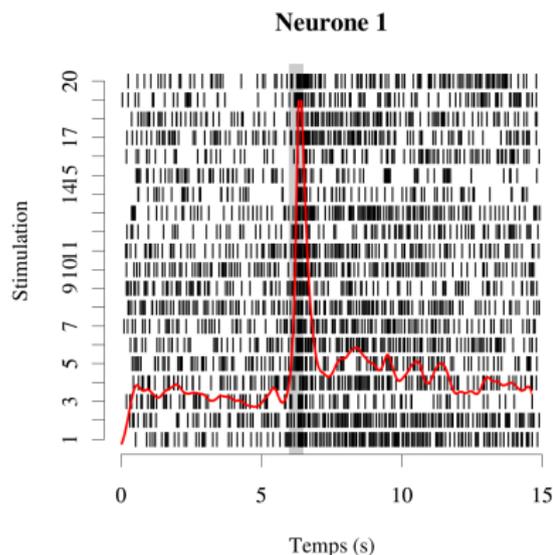
Transformation du temps

Un test basé sur le théorème de Donsker

Estimation de λ

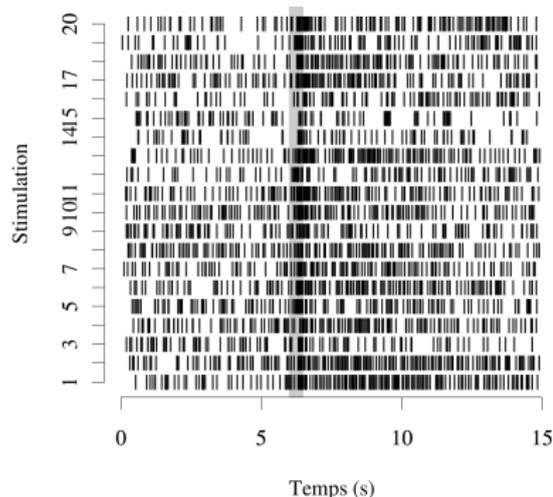
Retour aux données

Régime non-homogène : réponses aux odeurs (1)



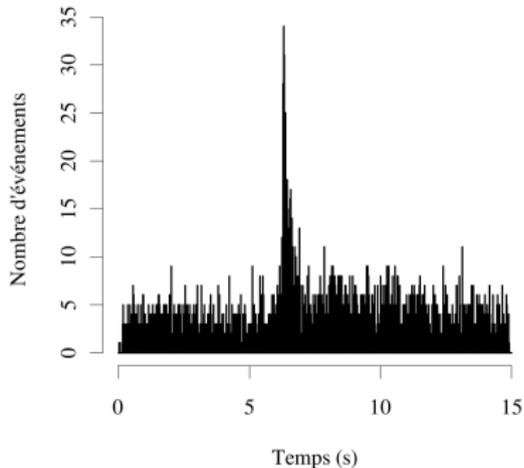
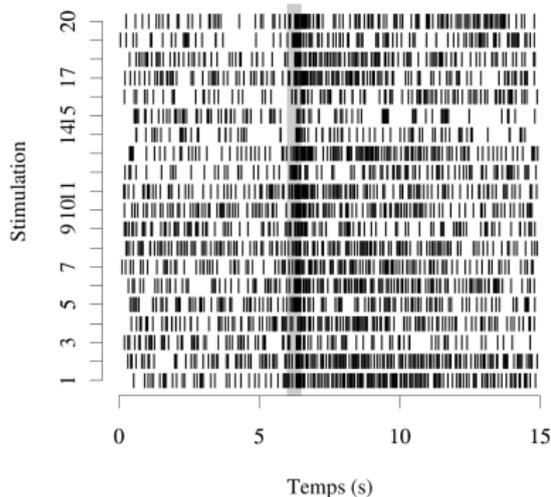
20 réponses à des stimulations de 500 ms avec du citronnellal. En rouge : une estimation « lisse » de la fréquence de décharge moyennée sur les stimulations.

Régime non-homogène : estimation de la fréquence moyennée (2)



On passe des données brutes...

Régime non-homogène : estimation de la fréquence moyennée (3)



On passe des données brutes à un histogramme construit avec un petit pas de temps (25 ms) ; ce qui nous donne, à une normalisation près, un estimateur peu biaisé mais de forte variance de la fréquence moyennée.

Régime non-homogène : estimation de la fréquence moyennée (4)

- ▶ On modélise le « processus moyenné » comme un **processus de Poisson inhomogène** d'intensité $\nu(t)$.
- ▶ L'histogramme que nous avons construit peut alors être considéré comme l'observation d'une collection de variables aléatoires de Poisson, $\{Y_1, \dots, Y_k\}$, de paramètres :

$$N_s \int_{t_i - \delta/2}^{t_i + \delta/2} \nu(u) du \approx N_s \nu(t_i) \delta, i = 1, \dots, k,$$

où t_i est le temps central d'une classe, δ est la largeur des classes, N_s est le nombre de stimulations et k est le nombre de classes.

Régime non-homogène : estimation de la fréquence moyennée (5)

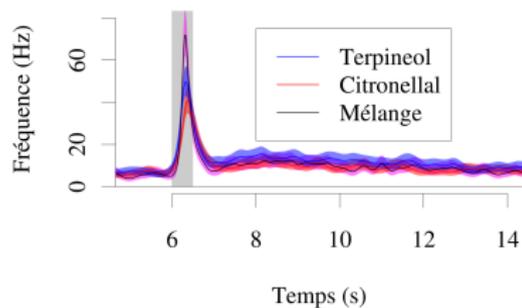
- ▶ Nous estimons en fait directement : $\eta() = \log(N_s \nu() \delta)$ par régression de Poisson.
- ▶ Comme nous ne voulons pas faire d'hypothèses « fortes » sur η nous choisissons comme estimateur, $\hat{\eta}$, la fonction qui minimise la vraisemblance pénalisée :

$$-\sum_{i=1}^k \left(y_i \eta(t_i) - \exp(\eta(t_i)) \right) + \lambda \int \left(\frac{d^2 \eta(u)}{dt^2} \right)^2 du.$$

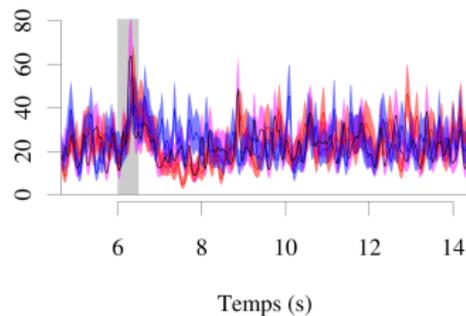
- ▶ Le paramètre de lissage, λ , est obtenu en minimisant le **critère de validation croisée généralisée**.
- ▶ Tous les calculs sont effectués avec le paquet *gss* (*general smoothing splines*) développé par Chong Gu pour le logiciel R.

Régime non-homogène : réponses aux odeurs (5)

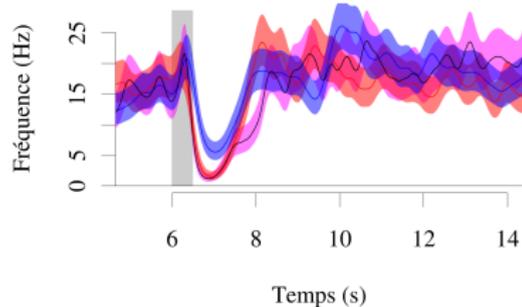
Neurone 1



Neurone 2



Neurone 3



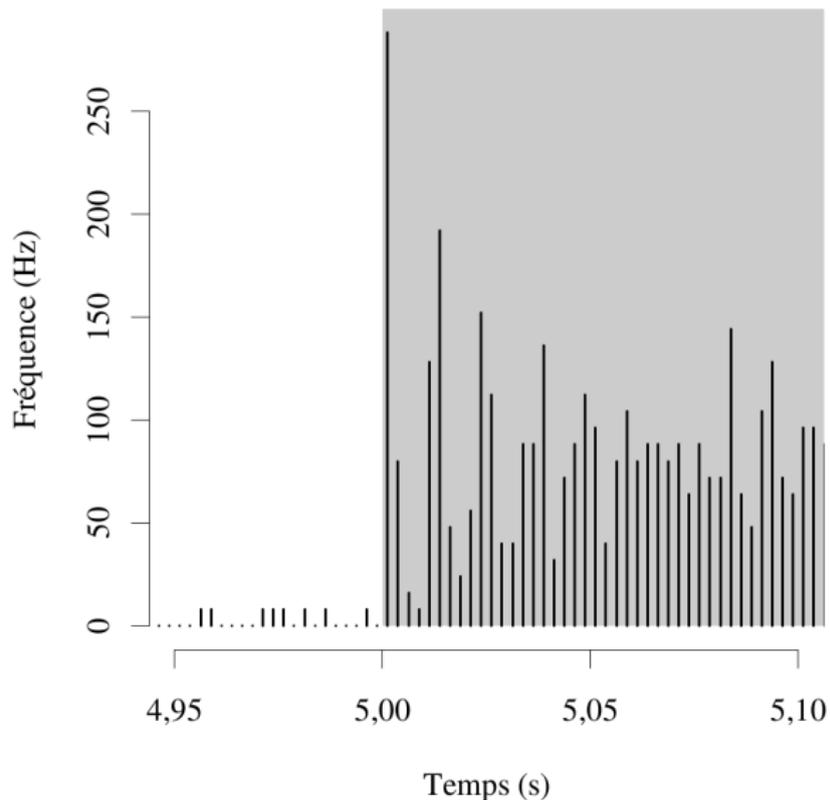
Pourquoi la fréquence moyennée est-elle insuffisante ? (1)

- ▶ Les neurophysiologistes rapportent habituellement les réponses aux stimulations sous forme d'histogrammes – normalisés de façon appropriée – appelés dans le domaine des « histogrammes des temps autour du stimulus » (*peri-stimulus time histograms, PSTH*).
- ▶ C'est ce que nous venons de faire en raffinant l'histogramme habituel par son estimation « lisse » et en complétant celle-ci par des intervalles de confiances.
- ▶ Mais ce type de résumé peut être assez ambiguë, comme une petite simulation – inspirée de situations réelles rencontrées dans le système auditif – va le montrer.
- ▶ Nous allons considérer un neurone dont la décharge spontanée correspond à un processus de renouvellement avec une distribution d'intervalles exponentielle *après une période réfractaire* de durée ρ .

Pourquoi la fréquence moyennée est-elle insuffisante ? (2)

- ▶ L'effet du stimulus sur la décharge va être modélisé « à la Cox » ; c.-à-d. que nous allons multiplier par un facteur constant le risque instantané pendant la durée de la stimulation.
- ▶ Concrètement nous allons prendre une période réfractaire ρ de 10 ms – ce qui est un peu exagéré par-rapport aux cas réels, une fréquence de décharge spontanée hors période réfractaire de 2,5 Hz et un facteur 200 correspondant à la stimulation, appliquée pendant 1 s.
- ▶ Nous allons simuler 50 stimulations avec la méthode de Lewis, plus communément appelée méthode de l'éclaircissage – au sens de l'éclaircissage d'une forêt – ou *thinning method* (Lewis et Shedler, 1979 ; Ogata 1981).

Pourquoi la fréquence moyennée est-elle insuffisante ? (3)



Histogramme des temps péri-stimulus du cas simulé. **Remarquez les oscillations au début de la réponse.**

Où en est-on ?

Les données

Régime non-homogène

Intensité conditionnelle

Transformation du temps

Un test basé sur le théorème de Donsker

Estimation de λ

Retour aux données

Cahier des charges du modèle

Notre modèle devrait nous permettre de travailler avec :

- ▶ le temps écoulé depuis le dernier PA du neurone (suffisant pour un processus de renouvellement homogène) ;
- ▶ des variables correspondant à l'histoire de la décharge du neurones, comme la durée de l'intervalle entre les deux derniers PAs ;
- ▶ des variables associées au temps écoulé depuis le dernier PA d'un neurone « fonctionnellement associé » ;
- ▶ le temps écoulé par-rapport au début d'une stimulation.

Quelques définitions

- ▶ Nous allons continuer en introduisant quelques définitions (un peu) plus « formelles », essentiellement pour clarifier nos notations.
- ▶ Pour des définitions correctes voir par exemple :
 - ▶ Rebolledo (1979) *La méthode des martingales appliquée à l'étude de la convergence en loi des processus*. Mémoires de la Société Mathématiques de France, **62**, pp 102–114.
 - ▶ Fleming & Harrington (2005) *Counting Processes and Survival Analysis*. Wiley

Filtration, histoire et intensité conditionnelle

- ▶ Les probabilistes des processus introduisent la **filtration** ou l'**histoire** : une famille de tribus croissantes, $(\mathcal{F}_t)_{0 \leq t \leq \infty}$, telle que toute l'information relative au processus au temps t , peut-être représentée par un élément de \mathcal{F}_t .
- ▶ L'**intensité conditionnelle** du compteur $N(t)$ est alors définie par :

$$\lambda(t | \mathcal{F}_t) \equiv \lim_{h \downarrow 0} \frac{\text{Prob}\{N(t+h) - N(t) = 1 | \mathcal{F}_t\}}{h}$$

Remarques

- ▶ En principe on s'attend à ce que \mathcal{F}_t fasse intervenir les trains de PAs de **tous** les neurones qui influent sur le neurone étudié.
- ▶ En pratique on a toujours enregistré au mieux qu'une toute petite fraction des ces neurones...
- ▶ On est alors « forcé » de considérer des approximations de $\lambda(t | \mathcal{F}_t)$ du type :

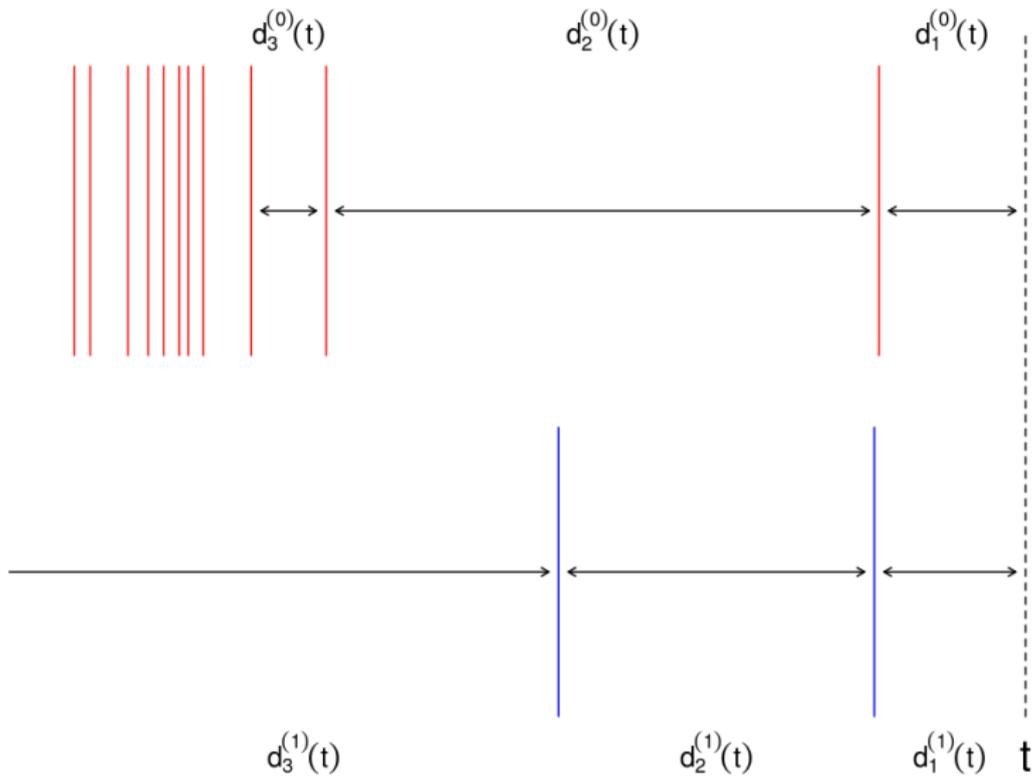
$$f\left(d_1^{(0)}(t), \dots, d_{m_0}^{(0)}(t), d_1^{(1)}(t), \dots, d_{m_1}^{(1)}(t), \dots, d_1^{(K)}(t), \dots, d_{m_K}^{(K)}(t)\right)$$

où, (0) est l'indice du neurone « étudié » et (1), ..., (K), sont les indices de K autres neurones enregistrés, où :

$$d_i^{(k)}(t) \equiv d_{i-1}^{(k)}(t) - \max\left\{t_j^{(k)} \mid t_j^{(k)} < d_{i-1}^{(k)}(t)\right\}, \quad \text{et} \quad d_0^{(k)}(t) = t.$$

On a de plus supposé implicitement que le processus est stationnaire.

Les $d_i^{(k)}(t)$ illustrés



Remarques sur notre approximation de λ

- ▶ Lorsque nous employons pour $\lambda(t | \mathcal{F}_t)$ l'approximation :

$$f\left(d_1^{(0)}(t), \dots, d_{m_0}^{(0)}(t), \dots, d_1^{(K)}(t), \dots, d_{m_K}^{(K)}(t)\right)$$

nous devons décider **quels neurones inclure** parmi ceux enregistrés et, **pour chaque neurone $k \in \{0, \dots, K\}$, quelle valeur de m_k** utiliser. Nous sommes donc immédiatement confronté à un problème de choix de « modèle ».

- ▶ Une fois ces choix effectués, il nous faut trouver la fonction f .
- ▶ La seule information directement exploitable fournie par des années de recherche en biophysique des neurones est que :

$$\lim_{d_1^{(0)}(t) \downarrow 0} f\left(d_1^{(0)}(t), \dots, d_{m_0}^{(0)}(t), \dots, d_1^{(K)}(t), \dots, d_{m_K}^{(K)}(t)\right) = 0$$

Les deux problèmes à résoudre

Maintenant que nous avons adopté un formalisme basé sur l'intensité conditionnelle, nous devons :

- ▶ trouver un estimateur $\hat{\lambda}$ de λ ;
- ▶ trouver des tests de qualité d'ajustement de nos modèles à nos données.

Où en est-on ?

Les données

Régime non-homogène

Intensité conditionnelle

Transformation du temps

Un test basé sur le théorème de Donsker

Estimation de λ

Retour aux données

Interprétation de λ

- ▶ Nous avons défini l'intensité du compteur $N(t)$ par :

$$\lambda(t | \mathcal{F}_t) \equiv \lim_{h \downarrow 0} \frac{\text{Prob}\{N(t+h) - N(t) = 1 | \mathcal{F}_t\}}{h}$$

- ▶ Ce que nous interprétons comme suit :

$$\begin{aligned} \text{Prob}\{N(t+h) - N(t) = 0 | \mathcal{F}_t\} &= 1 - \lambda(t | \mathcal{F}_t) h + o(h) \\ \text{Prob}\{N(t+h) - N(t) = 1 | \mathcal{F}_t\} &= \lambda(t | \mathcal{F}_t) h + o(h) \\ \text{Prob}\{N(t+h) - N(t) > 1 | \mathcal{F}_t\} &= o(h) \end{aligned}$$

où $o(h)$ est telle que :

$$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$$

Probabilité de ne pas avoir d'événement sur $]t, t + s]$

- ▶ On cherche la probabilité, $P_0(t, s)$, de ne pas avoir de PA sur l'intervalle $]t, t + s]$ lorsque s n'est pas infiniment petit.
- ▶ Nous divisons $]t, t + s]$ en sous intervalles disjoints :

$$\mathcal{I}_j =]t + jh, t + (j + 1)h], \quad j = 0, \dots, k-1 \quad \text{et} \quad h = s/k$$

- ▶ Alors :

$$P_0(t, s) = \prod_{j=0}^{k-1} (1 - \lambda(t + jh | \mathcal{F}_{t+jh}) h + o(h))$$

$$P_0(t, s) = \exp \left[\sum_{j=0}^{k-1} \log(1 - \lambda(t + jh | \mathcal{F}_{t+jh}) h + o(h)) \right]$$

$$P_0(t, s) = \exp \left[- \sum_{j=0}^{k-1} \lambda(t + jh | \mathcal{F}_{t+jh}) h + o(kh) \right]$$

$$P_0(t, s) \xrightarrow{k \rightarrow \infty} \exp \left(- \int_t^{t+s} \lambda(u | \mathcal{F}_u) du \right)$$

Densité des intervalles et intensité cumulée

- ▶ La fonction $P_0(t, s)$ dont nous venons d'obtenir l'expression est une **fonction de survie** à partir de laquelle on obtient la densité des intervalles entre PAs :

$$\begin{aligned} p(t, s) &= -\frac{dP_0(t, s)}{ds} \\ &= \lambda(t + s | \mathcal{F}_{t+s}) \exp\left(-\int_t^{t+s} \lambda(u | \mathcal{F}_u) du\right) \end{aligned}$$

- ▶ Si maintenant nous introduisons l'**intensité cumulée** :

$$\Lambda(t) \equiv \int_0^t \lambda(u | \mathcal{F}_u) du$$

- ▶ On voit que :

$$p(t, s) ds = d\Lambda(t + s) \exp - (\Lambda(t + s) - \Lambda(t))$$

Transformation du temps

- ▶ Si $\lambda(u | \mathcal{F}_u) > 0$ alors le processus Λ est strictement croissant et la **transformation du temps** : $t \in \mathbb{R}^{+*} \mapsto \Lambda(t) \in \mathbb{R}^{+*}$ admet une transformation inverse.
- ▶ La densité des intervalles entre PAs successifs après transformation du temps :
 $\{t_1, \dots, t_n\} \mapsto \{\Lambda(t_1) \equiv \Lambda_1, \dots, \Lambda(t_n) \equiv \Lambda_n\}$ est alors donnée, d'après la dernière équation de la diapo précédente, par :

$$p(\Lambda_{i+1} - \Lambda_i) = \exp -(\Lambda_{i+1} - \Lambda_i)$$

C'est une densité exponentielle de paramètre 1.

- ▶ Le processus ponctuel observé $\{\Lambda_1, \dots, \Lambda_n\}$ est donc une observation d'un **processus de Poisson de paramètre 1**.

Illustration de la transformation du temps sur données simulées (1)

Processus d'intensité et séquence d'événements

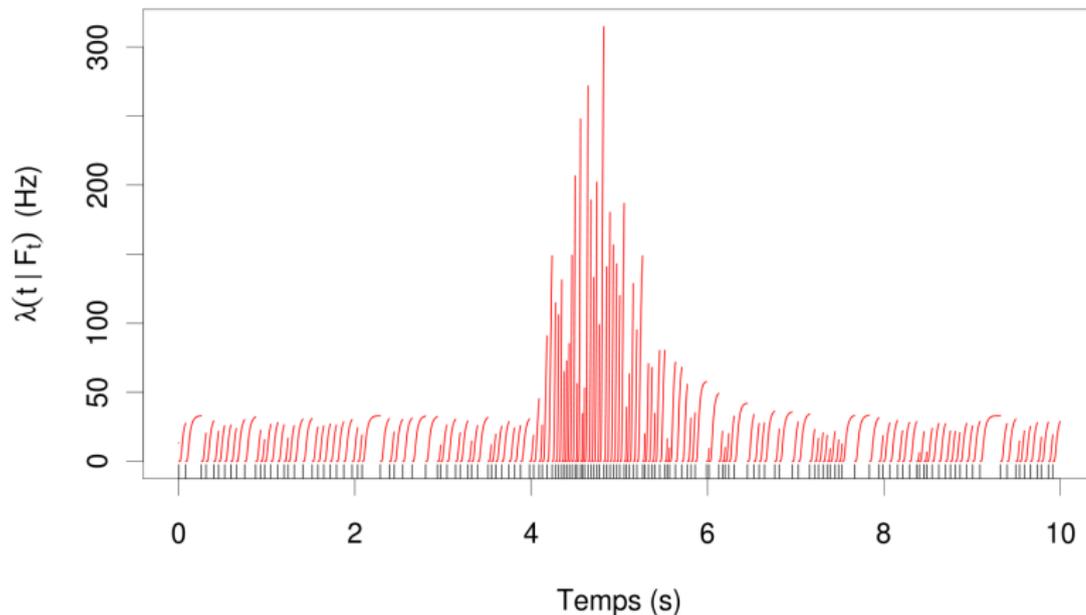


Illustration de la transformation du temps sur données simulées (2)

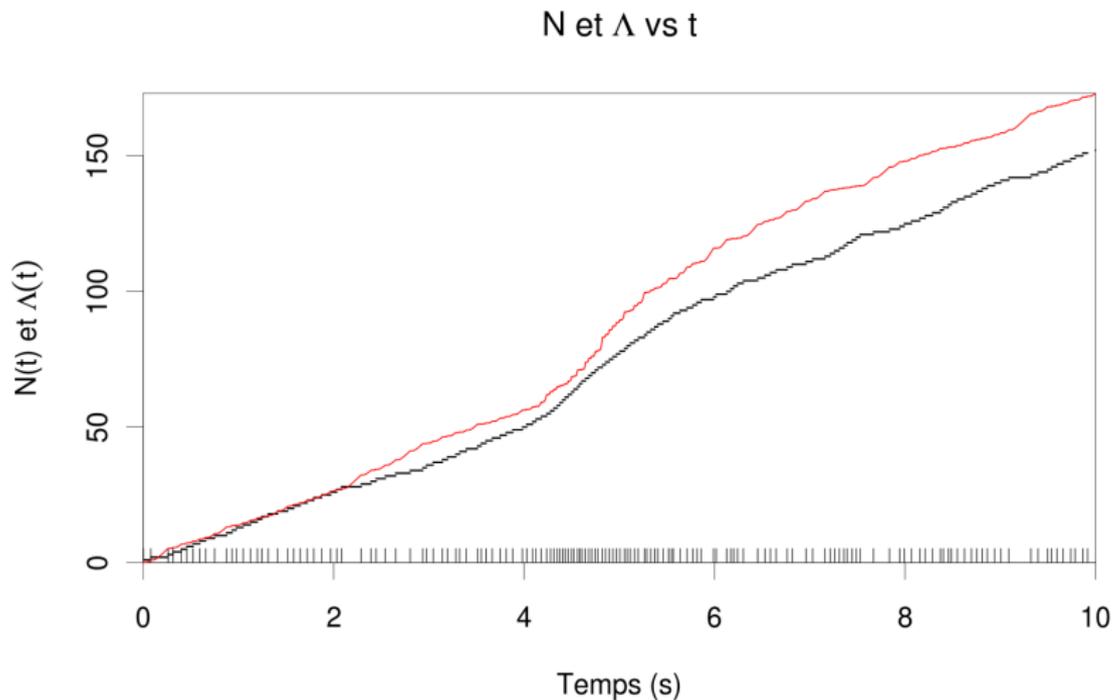
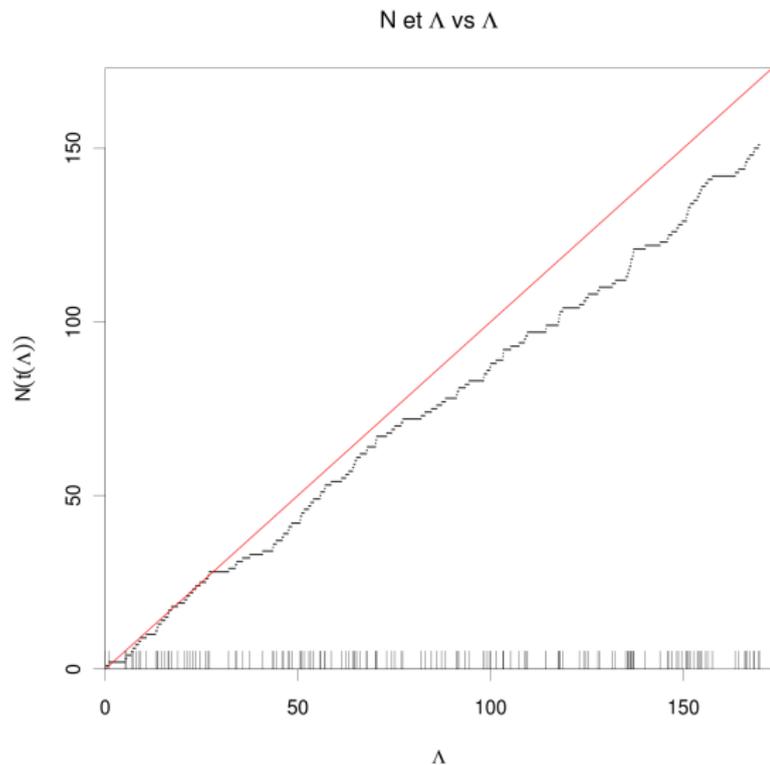
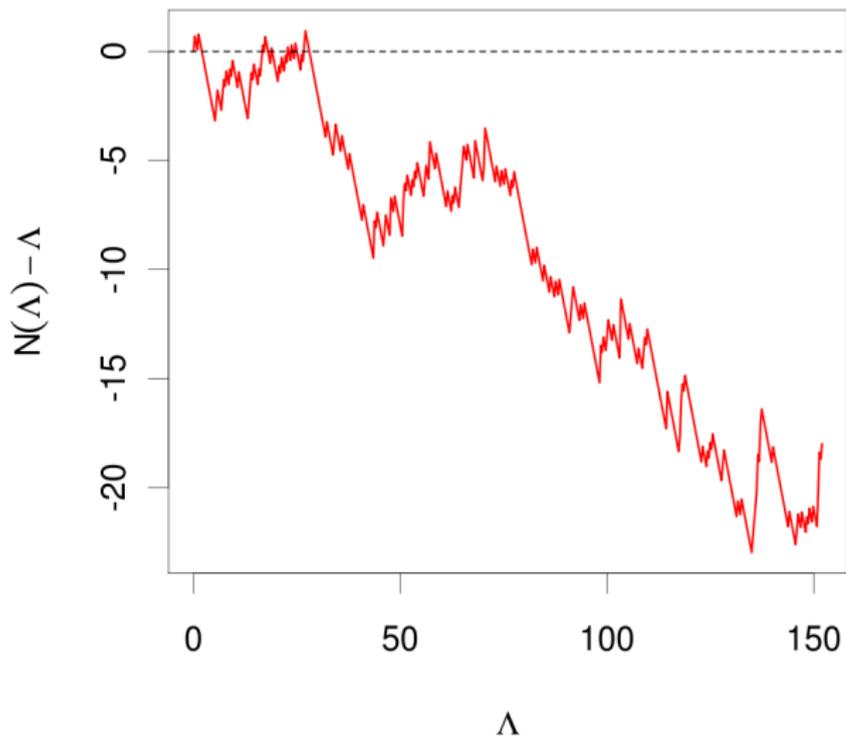


Illustration de la transformation du temps sur données simulées (3)



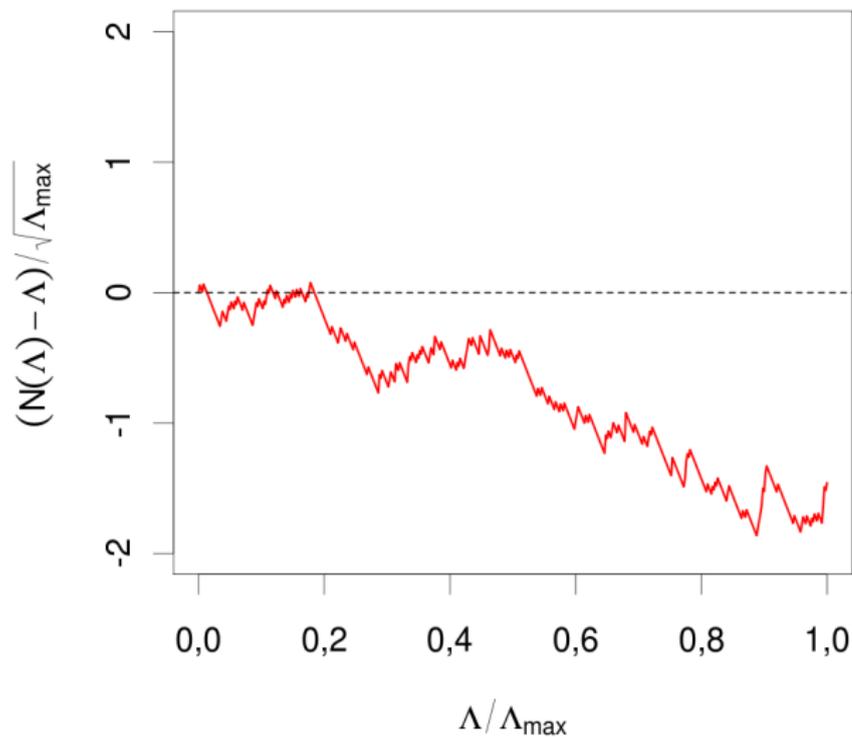
Un mouvement brownien ? (1)

$N(\Lambda) - \Lambda$ vs Λ



Un mouvement brownien ? (2)

$$(N(\Lambda) - \Lambda) / \sqrt{\Lambda_{\max}} \text{ vs } \Lambda / \Lambda_{\max}$$



Où en est-on ?

Les données

Régime non-homogène

Intensité conditionnelle

Transformation du temps

Un test basé sur le théorème de Donsker

Estimation de λ

Retour aux données

Théorème de Donsker

- ▶ Soit X_1, X_2, \dots une suite de variables aléatoires IID telles que $EX_i = 0$ et $EX_i^2 = 1$, alors la suite de processus :

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{i=0}^{\lfloor nt \rfloor} X_i, \quad 0 \leq t \leq 1, \quad X_0 = 0$$

converge en loi vers le mouvement brownien canonique.

- ▶ Pour une preuve du théorème de Donsker voir :
 - ▶ R Durrett (2009) *Probability : Theory and Examples*. CUP. Sec. 7.6, pp 323-329.
 - ▶ P Billingsley (1999) *Convergence of Probability Measures*. Wiley. p 121.

Application du théorème de Donsker (1)

- ▶ Si notre modèle est bon, alors les $\{\Lambda_1, \Lambda_2, \dots, \Lambda_n\}$ constituent une observation d'un processus de Poisson de paramètre 1.
- ▶ Les X_i de notre énoncé du théorème peuvent alors correspondre aux $\Lambda_{i+1} - \Lambda_i - 1$ (du fait de la loi exponentielle de paramètre 1 des intervalles).
- ▶ Alors le graphe de :

$$\frac{(\Lambda(t) - \Lambda(t_1)) - (N(t) - 1)}{\sqrt{N(t_{max}) - 1}} \text{ vs } \frac{\Lambda(t) - \Lambda(t_1)}{\Lambda(t_{max}) - \Lambda(t_1)}, \text{ pour } t_1 \leq t \leq t_{max}$$

va « ressembler de plus en plus » au chemin du mouvement brownien canonique sur $[0, 1]$.

Application du théorème de Donsker (2)

- ▶ Les X_i de l'énoncé du théorème peuvent encore être donnés par : $N(k+1) - N(k) - 1$, où $k \in \mathbb{N}$ et où le processus de comptage est défini sur l'échelle de temps transformé (du fait de la loi de Poisson de paramètre 1 suivie par le nombre d'événements générés par un processus de Poisson de paramètre 1 sur une durée de 1).
- ▶ Alors le graphe de :

$$\frac{N(t) - \Lambda(t)}{\sqrt{[\Lambda(t_{max})]}} \text{ vs } \frac{\Lambda(t)}{[\Lambda(t_{max})]} \quad \text{pour } 0 \leq t \leq \max\{t \mid \Lambda(t) \leq [\Lambda(t_{max})]\}$$

va « ressembler de plus en plus » au chemin du mouvement brownien canonique sur $[0, 1]$.

Remarques

- ▶ Si pour $0 \leq t \leq \max\{t \mid \Lambda(t) \leq \lfloor \Lambda(t_{max}) \rfloor\}$ nous définissons :

$$n_{max} \equiv \lfloor \Lambda(t_{max}) \rfloor, \quad x \equiv \frac{\Lambda(t)}{n_{max}} \quad \text{et} \quad \tilde{N}(u) \equiv N(\Lambda^{-1}(u))$$

- ▶ Alors notre dernier résultat devient :

$$\frac{\tilde{N}(n_{max}x) - n_{max}x}{\sqrt{n_{max}}} \xrightarrow{\text{Loi}} B(t)$$

où B désigne le mouvement brownien canonique.

- ▶ C'est un résultat classique qu'on trouve par exemple dans :
 - ▶ Rebolledo (1979) p 107.
 - ▶ Andersson & Britton (2000) *Stochastic Epidemic Models and Their Statistical Analysis*. Springer. p 39.

Alors quoi de neuf?

- ▶ Notre proposition, qui semble être originale, est d'utiliser la convergence en loi du processus de comptage « compensé » et exprimé sur une nouvelle échelle de temps vers un mouvement brownien canonique **quand le modèle est bon** pour tester si notre modèle est effectivement bon.
- ▶ Pour cela il nous faut être capables de reconnaître un mouvement brownien canonique sur $[0, 1]$ lorsque nous en voyons un...

Approche de la « région de prédiction »

- ▶ Soit $B(t)$ un mouvement brownien canonique et soit $f(t)$ une « frontière ». On définit le **temps de premier passage** $T(f)$ par :

$$T(f) \equiv \inf\{t \mid B(t) \geq f(t)\}$$

- ▶ Si on arrive à exprimer (ou estimer) la loi de $T(f)$ « en fonction de » f , on peut chercher $f_{0.975}$ telle que :
 $\text{Prob}(T(f_{0.975}) > 1) = 0.975$.
- ▶ Il suffit alors de définir :

$$\tilde{T}(f_{0.975}) \equiv T(f_{0.975}) \wedge T(-f_{0.975})$$

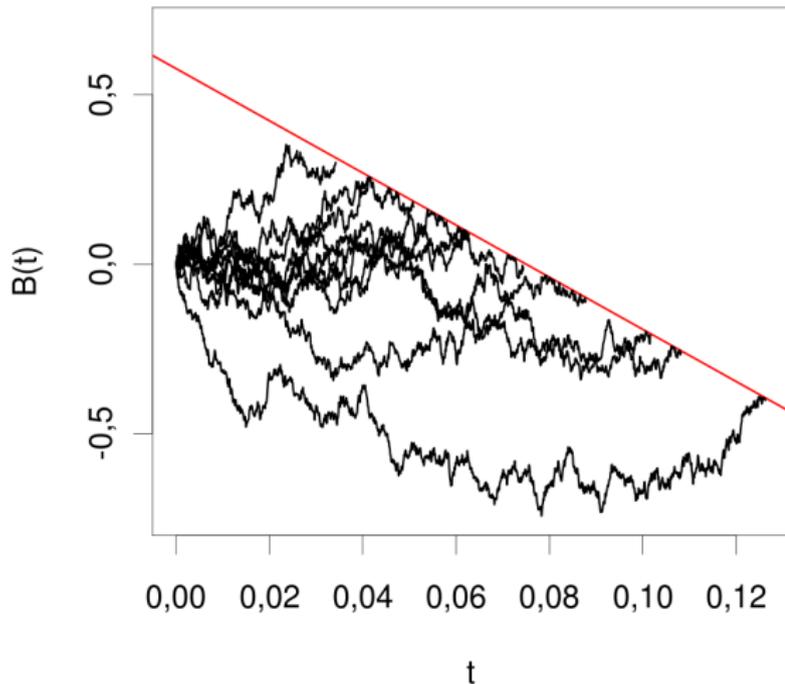
et la propriété de symétrie du mouvement brownien nous donne :

$$\text{Prob}(\tilde{T}(f_{0.975}) > 1) = 0.95$$

- ▶ Alors le domaine $[0, 1] \times [-f_{0.975}, f_{0.975}]$ contiendra **dans leur totalité** 95% des mouvements browniens observés sur $[0, 1]$.

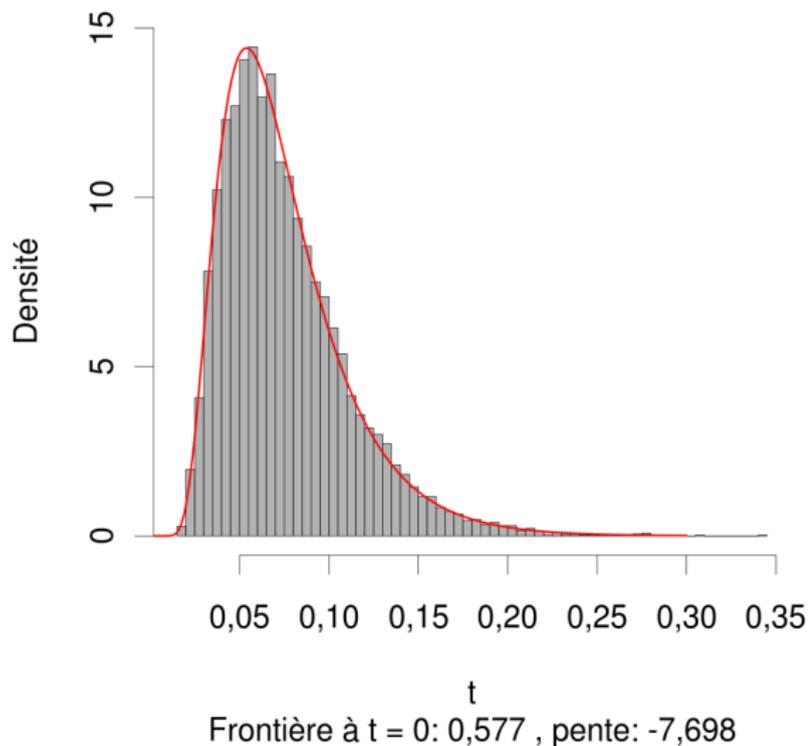
Illustration du temps de premier passage

Frontière linéaire



Loi d'un temps de premier passage

Densité du temps de premier passage



Loi gaussienne inverse d'un temps de premier passage

- ▶ Whitmore et Sheshadri montrent dans un élégant papier de 1987 (*The American Statistician* **41** : 280-281) que la loi du premier temps de passage à travers une frontière :

$$f(t) = a + bt, \quad \text{avec } a > 0 \quad \text{et} \quad b < 0$$

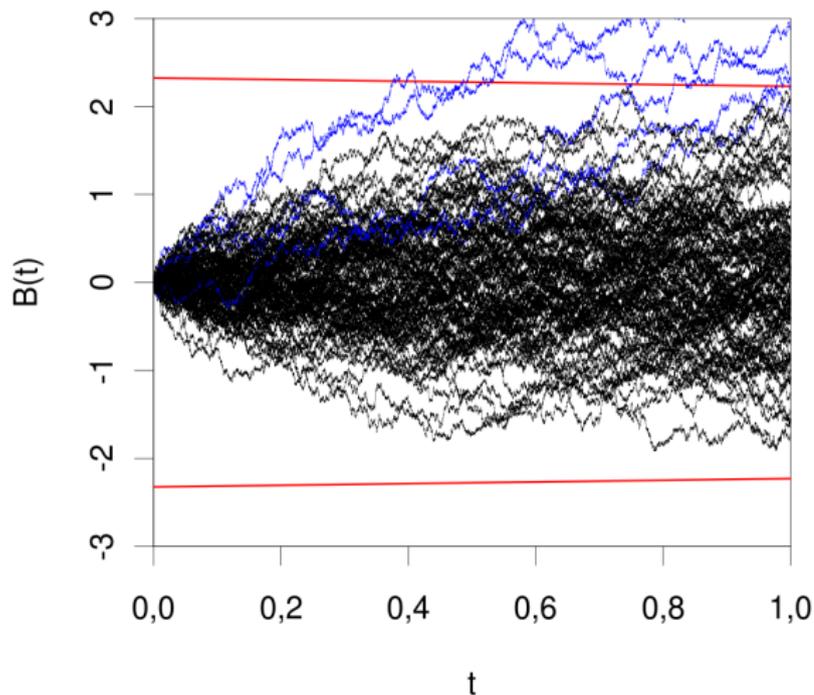
est une loi gaussienne inverse de densité :

$$g(t) = \frac{a}{\sqrt{2\pi t^3}} \exp\left(-\frac{(a+bt)^2}{2t}\right)$$

- ▶ Une optimisation numérique montre alors que $G(1) = 0.025$ pour $a = 2.324$ et $b = -0.095$, G désignant ici la fonction de répartition de la loi gaussienne inverse.

Région de prédiction à frontières linéaires

n = 100



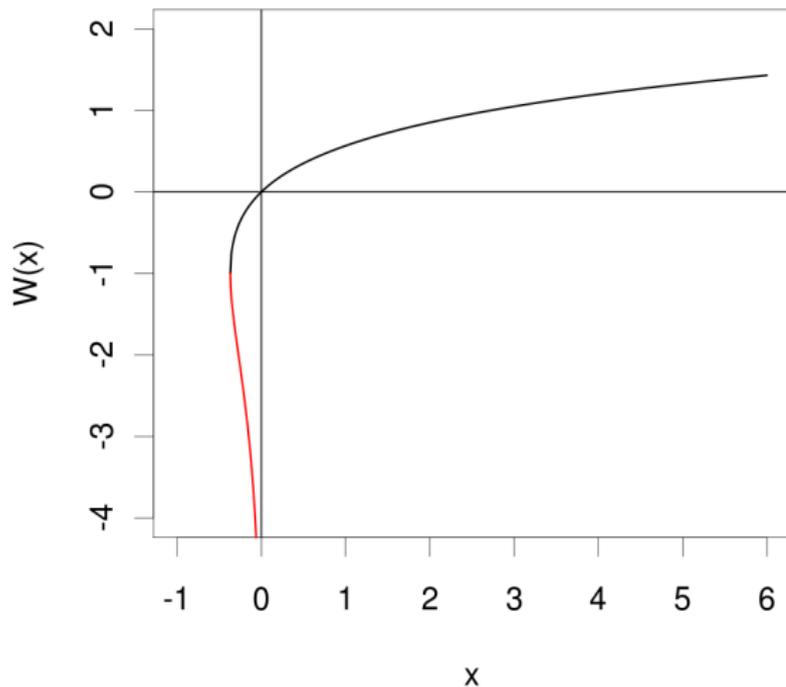
Région de prédiction de surface minimale

- ▶ Dès lors qu'une approche basée sur une région de prédiction est adoptée, il est raisonnable pour ne pas dire souhaitable d'utiliser une **région de surface minimale**.
- ▶ Dans un article (pas franchement trivial) de 2007, Kendall, Marin et Robert (Brownian Confidence Bands on Monte Carlo Output, *Statistics and Computing*, **17** : 1-10) ont montré que la frontière « supérieure » de cette surface minimale est donnée par :

$$u^*(t) \equiv \sqrt{-W_{-1}(-(\kappa t)^2)} \sqrt{t}, \quad \text{pour } \kappa t \leq 1/\sqrt{e}.$$

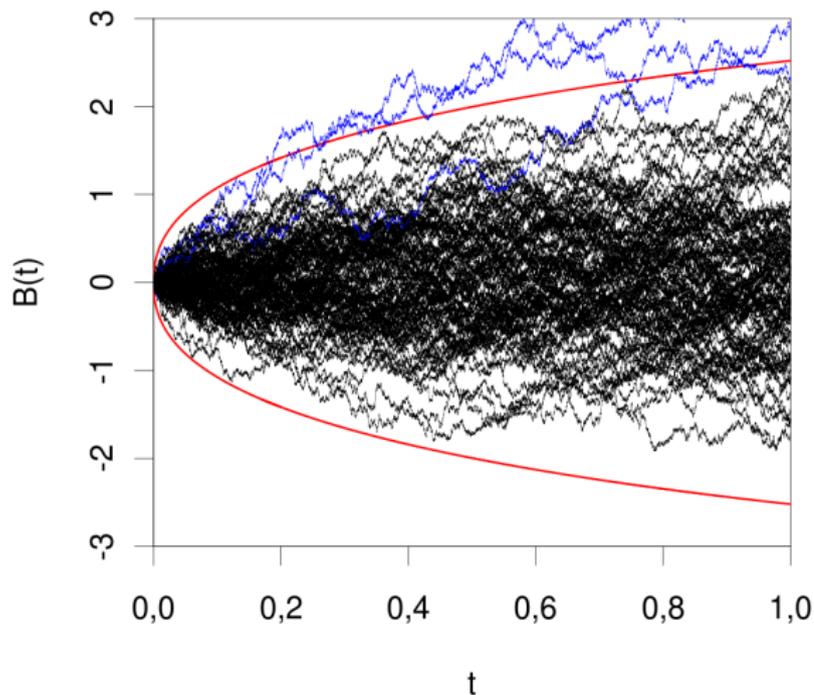
où W_{-1} est la « branche réelle secondaire » de la fonction W de Lambert définie comme la solution de : $W(z) \exp W(z) = z$ et où κ est un paramètre à ajuster pour obtenir la probabilité de recouvrement désirée.

Les deux branches de la fonction de Lambert



Région de prédiction (de surface) minimale à 95%

n = 100



Comment trouver κ pour avoir une probabilité de recouvrement donnée ?

- ▶ La mise en œuvre pratique de tests de qualité d'ajustement basés sur une région de prédiction requiert que nous soyons capable de construire une région dont nous contrôlons la probabilité de recouvrement.
- ▶ Dit autrement, il faut que je vous explique comment j'ai pu trouver le κ de la dernière figure qui donne une probabilité de 95%.
- ▶ La solution (numérique) à ce problème est donnée dans un article de Loader et Deely (1987, Computations of boundary crossing probabilities for the Wiener process. *J. Statist. Comput. Simulation* **27** : 95-105).

Une équation intégrale de Volterra

Dans leur article, Loader et Deely montrent que la fonction de répartition $G(T)$ du temps de premier passage à travers une frontière $f(t)$ est solution d'une équation intégrale de Volterra de première espèce :

$$H(t) = \int_0^t K(t, u) dG(u)$$

où :

$$H(t) = \Phi\left(-\frac{f(t)}{\sqrt{t}}\right) + \exp\left(-2(f(t) - t)\right) \Phi\left(\frac{-f(t) + 2t}{\sqrt{t}}\right)$$

Φ désignant la fonction de répartition d'une loi normale réduite, et :

$$K(t, u) = \Phi\left(\frac{f(u) - f(t)}{\sqrt{t - u}}\right) + \exp\left(-2(f(t) - f(u) - (t - u))\right) \Phi\left(\frac{f(u) - f(t) + 2(t - u)}{\sqrt{t - u}}\right)$$

Solution numérique par la méthode du point milieu

- ▶ Cette équation intégrale de Volterra peut-être résolue numériquement par la méthode du point milieu.
- ▶ On procède en choisissant : $0 = t_0 < t_1 < \dots < t_n = t$, $t_j = jh$ pour un $h > 0$.
- ▶ On définit : $t_{j-1/2} = (t_j + t_{j-1})/2$.

▶ Alors :

$$H(t_j) = \sum_{i=1}^j K(t_j, t_{i-\frac{1}{2}}) \Delta_i, \quad j = 1, \dots, n$$

où $\Delta_i = G(t_i) - G(t_{i-1})$.

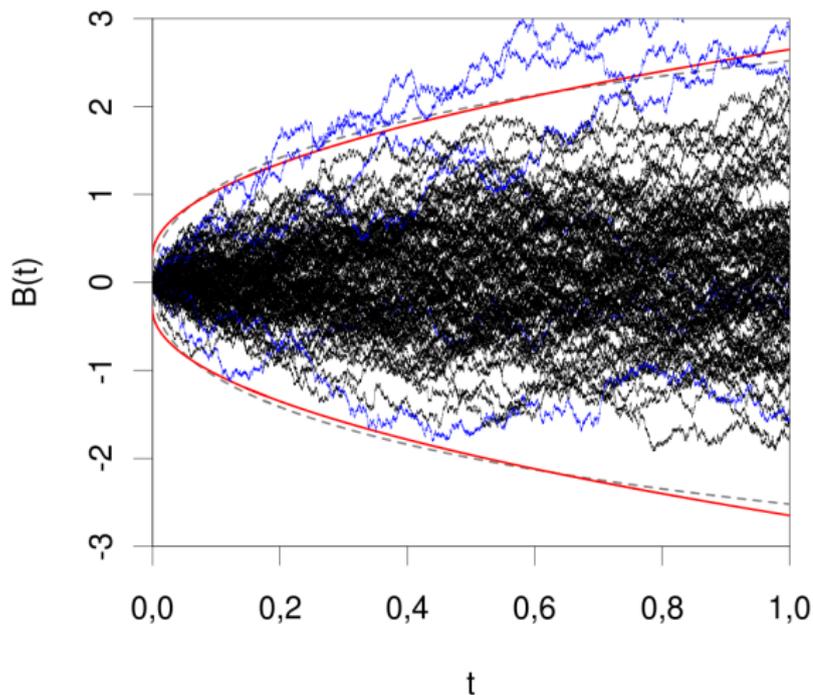
- ▶ On a alors un **système triangulaire inférieur** à résoudre ce qui est facile et rapide.
- ▶ Il est de plus possible, dans les cas qui nous intéressent, de borner l'erreur absolue de cette solution numérique.

Une région de prédiction presque minimale

- ▶ Un problème potentiel de l'approche de la région de prédiction de surface minimale est que votre logiciel favori ne comporte pas nécessairement la fonction W_{-1} (le paquet `gsl` de R la contient).
- ▶ Dans ce cas on peut utiliser un résultat supplémentaire de Kendall, Marin et Robert : une frontière du type $f(t) = a + b\sqrt{t}$ est presque de surface minimale.

Région de prédiction presque minimale à 95%

n = 100



Où en est-on ?

Les données

Régime non-homogène

Intensité conditionnelle

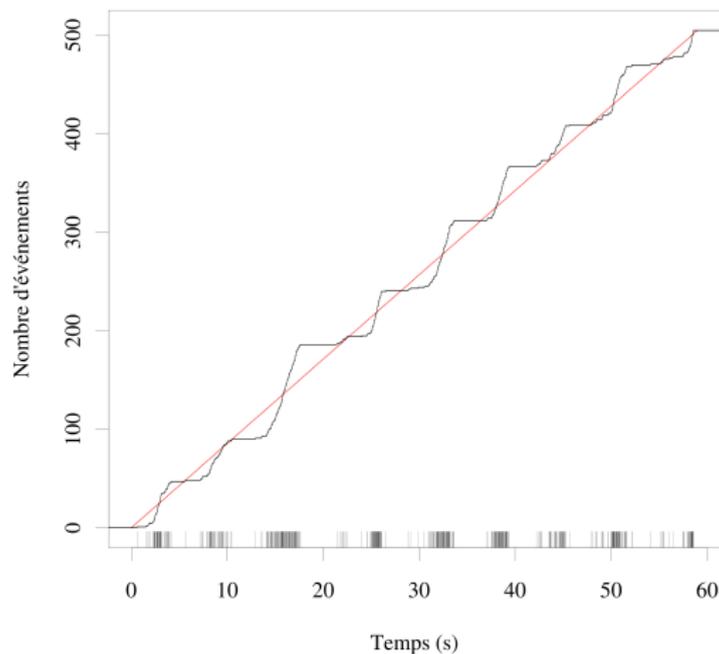
Transformation du temps

Un test basé sur le théorème de Donsker

Estimation de λ

Retour aux données

Retour sur le cas « compliqué »



L'analyse exploratoire précédente nous impose de considérer un modèle minimal du type :
 $f(d_1^{(0)}(t), d_2^{(0)}(t))$.

Approche de David Brillinger

- ▶ Nous suivons Brillinger (1988) qui commence par discrétiser l'axe des temps en blocs de durée h ; h étant suffisamment petit pour avoir au plus un PA par bloc.
- ▶ Nous sommes ainsi ramenés à un problème de **régression binomiale**.
- ▶ Les « données discrétisées » sont alors considérées comme une observation d'une collection de variables aléatoires de Bernoulli $\{Y_1, \dots, Y_k\}$ de paramètres : $f(d_1^{(0)}(t), d_2^{(0)}(t)) h$.
- ▶ Nous allons directement estimer :

$$\log \left(\frac{f(d_1^{(0)}(t), d_2^{(0)}(t)) h}{1 - f(d_1^{(0)}(t), d_2^{(0)}(t)) h} \right) = \eta(d_1^{(0)}(t), d_2^{(0)}(t))$$

Les données discrétisées

	event	time	neuron	lN.1	i1
14604	0	58.412	1	0.012	0.016
14605	1	58.416	1	0.016	0.016
14606	0	58.420	1	0.004	0.016
14607	1	58.424	1	0.008	0.016
14608	0	58.428	1	0.004	0.008
14609	0	58.432	1	0.008	0.008
14610	1	58.436	1	0.012	0.008
14611	0	58.440	1	0.004	0.012

event est la séquences de PAs discrétisée, **time** est le centre des blocs, **neuron** est le neurone auquel event « appartient », **lN.1** est $d_1^{(0)}(t)$ et **i1** est $d_2^{(0)}(t)$. Ici, h est égal à 4 ms.

Splines de lissage

- ▶ Comme la biophysique cellulaire ne nous donne que peu d'informations sur la forme fonctionnelle de η , nous allons employer des fonctions splines et nous allons pénaliser la vraisemblance comme précédemment (Wahba, 1990 ; Green et Silverman, 1994 ; Eubank, 1999 ; Gu, 2002).
- ▶ Nous continuons à effectuer les calculs avec le paquet gss de Chong Gu – ainsi qu'avec l'aide directe de ce dernier.
- ▶ $\eta(d_1, d_2)$ est décomposée de façon unique en :

$$\eta(d_1, d_2) = \eta_\emptyset + \eta_1(d_1) + \eta_2(d_2) + \eta_{1,2}(d_1, d_2)$$

où d_1 est une version transformée de $d_1^{(0)}(t)$ telle que $d_1 \in [0, 1]$; de même pour d_2 vis à vis de $d_2^{(0)}(t)$.

- ▶ La décomposition est unique parce-qu'on impose des conditions du type : $\int_0^1 \eta_i = 0$.

Où en est-on ?

Les données

Régime non-homogène

Intensité conditionnelle

Transformation du temps

Un test basé sur le théorème de Donsker

Estimation de λ

Retour aux données

Retour aux données

- ▶ La prochaine diapo va comporter la séquence « compliquée » que vous avez vue plusieurs fois.

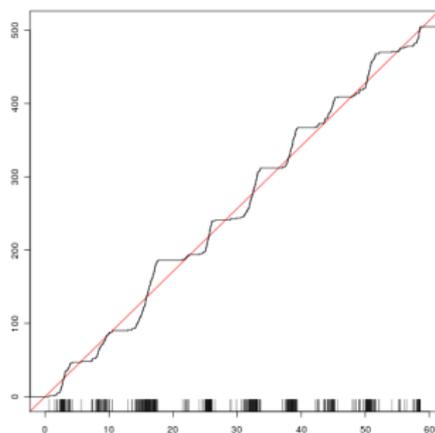
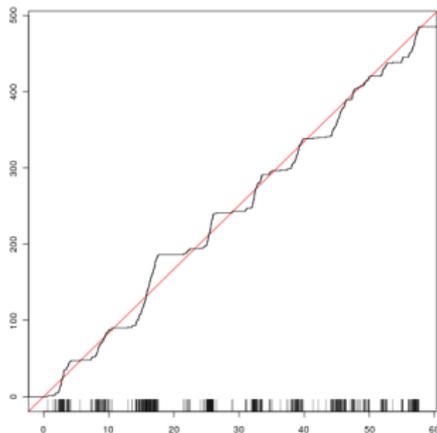
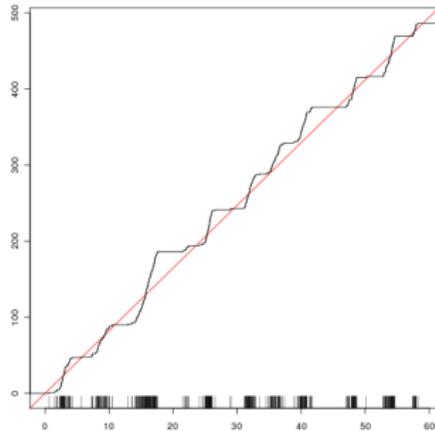
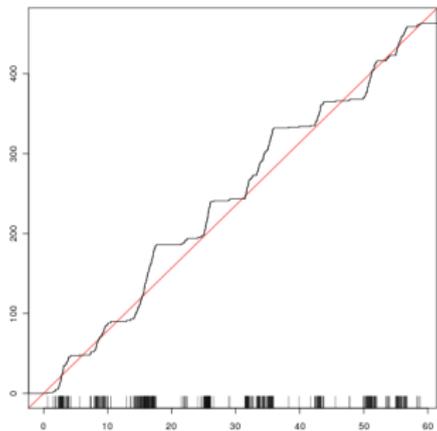
Retour aux données

- ▶ La prochaine diapo va comporter la séquence « compliquée » que vous avez vue plusieurs fois.
- ▶ Trois autres séquences vont aussi être montrées. La seconde moitié de ces dernières ($t \geq 29.5$) a été simulée.

Retour aux données

- ▶ La prochaine diapo va comporter la séquence « compliquée » que vous avez vue plusieurs fois.
- ▶ Trois autres séquences vont aussi être montrées. La seconde moitié de ces dernières ($t \geq 29.5$) a été simulée.
- ▶ La simulation a été faite avec le **même** modèle ajusté sur la première moitié du jeu de données.

Où est la vraie séquence ?



Réponse

La vraie séquence est située **en bas à droite**.

Trouver le « bon » modèle (1)

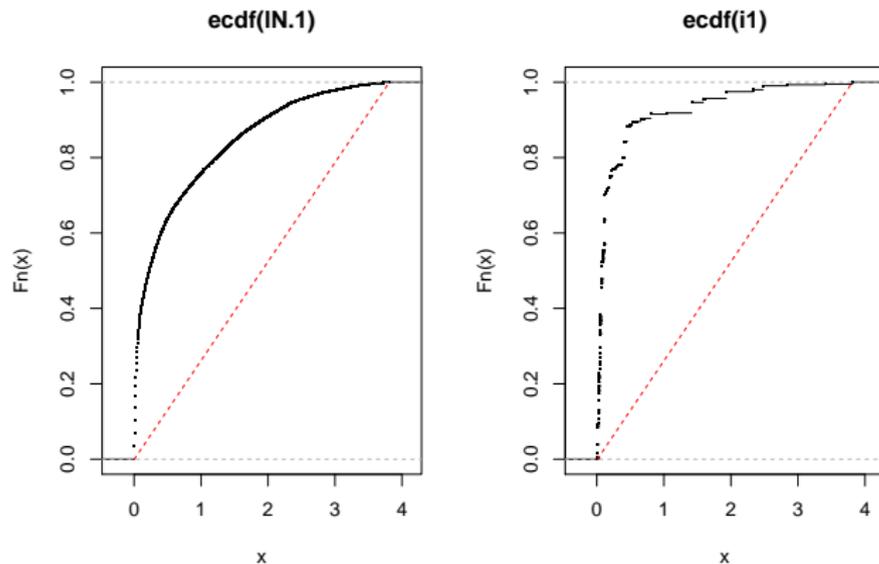
- ▶ Nous avons annoncé que nous commencerions avec un modèle à 2 variables :

$$\eta(d_1, d_2) = \eta_\emptyset + \eta_1(d_1) + \eta_2(d_2) + \eta_{1,2}(d_1, d_2) .$$

- ▶ Comme nous utilisons des modèles non paramétriques, nous ne devrions pas appliquer nos tests aux données utilisées pour ajuster les modèles...
- ▶ Nous procédons donc systématiquement en « coupant » les jeux de données en deux. Le modèle est ajusté sur une moitié et la qualité d'ajustement est testée sur l'autre moitié.

Un « détail » important (1)

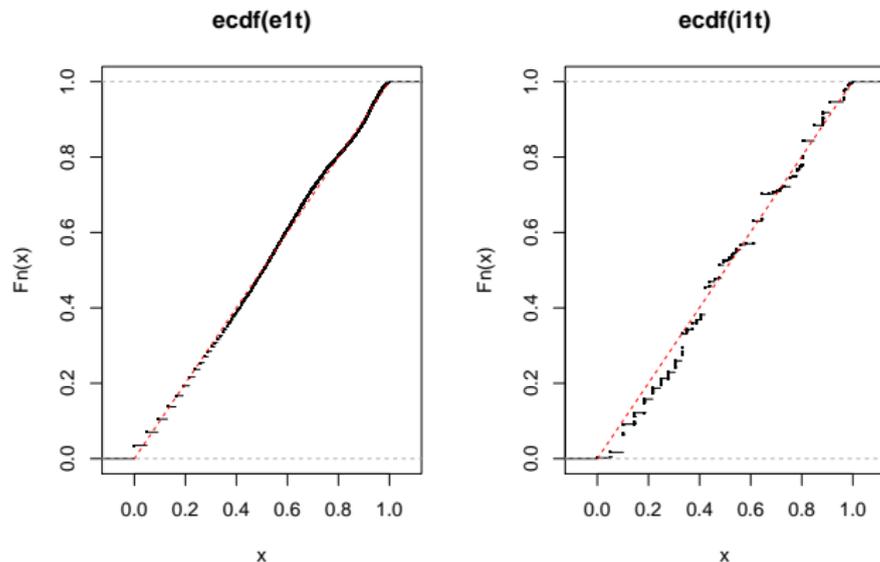
Les distributions de nos variables de départ, $d_1^{(0)}$ et $d_2^{(0)}$, sont (très) loin d'être uniformes :



Pour des raisons que je ne comprends pas bien, les ajustements sont *bien meilleurs* si je les transforme en variables uniformes.

Un « détail » important (2)

Je transforme donc les variables en utilisant une estimation lisse des fonctions de répartition empiriques de celles-ci, estimées sur la première moitié du jeu de données.



Ces fonctions de répartition empiriques des variables « uniformisées » sont construites sur le jeu de données complet.

Trouver le « bon » modèle (2)

Nous allons en fait ajuster deux modèles aux données :

- ▶ Modèle 1 :

$$\eta(d_1, d_2) = \eta_\emptyset + \eta_1(d_1) + \eta_2(d_2) + \eta_{1,2}(d_1, d_2) ,$$

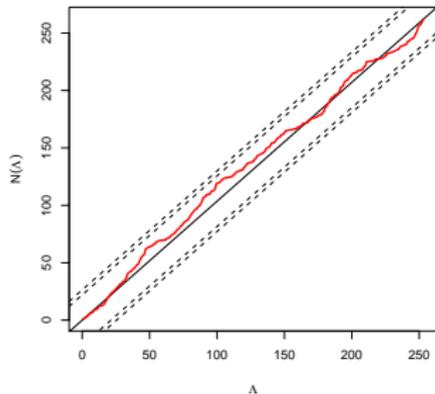
- ▶ Modèle 2 :

$$\eta(d_1, d_2) = \eta_\emptyset + \eta_1(d_1) + \eta_2(d_2) ,$$

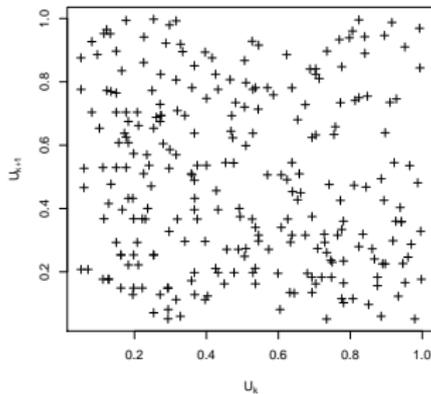
un modèle additif.

Modèle 1 ajusté au début et testé sur la fin

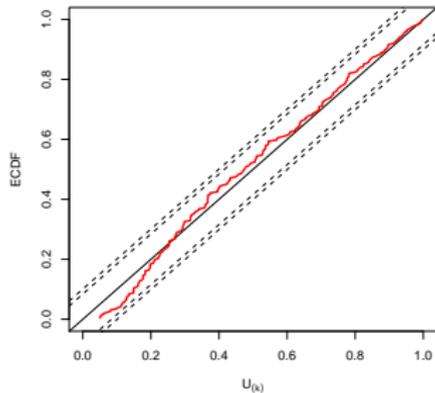
Uniform on Λ Test



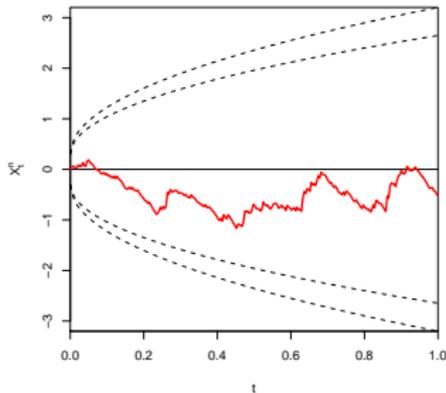
U_{k+1} vs U_k



Berman's Test

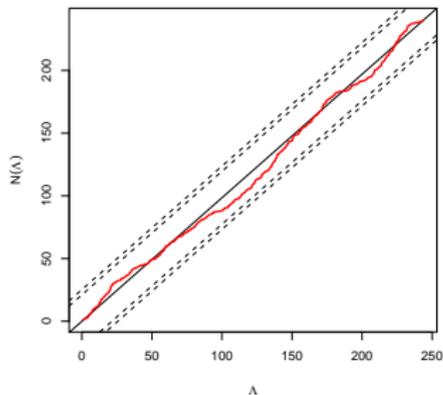


Wiener Process Test

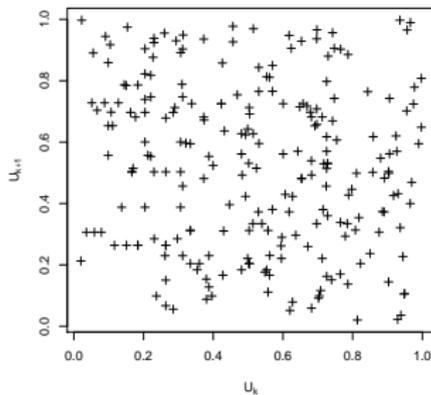


Modèle 1 ajusté à la fin et testé sur le début

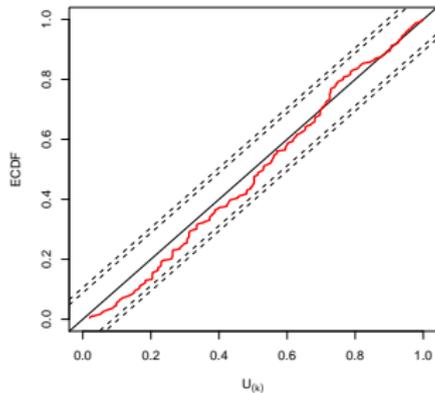
Uniform on Λ Test



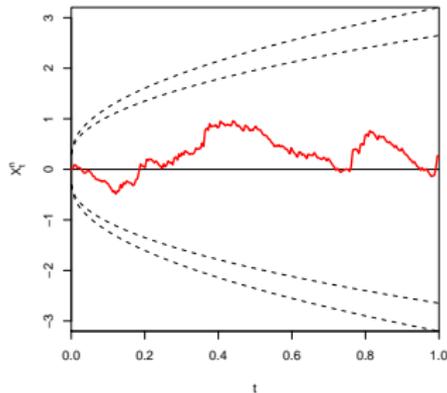
U_{k+1} vs U_k



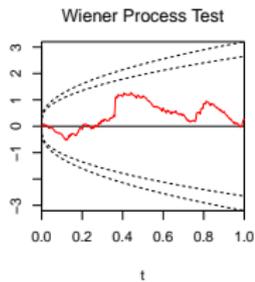
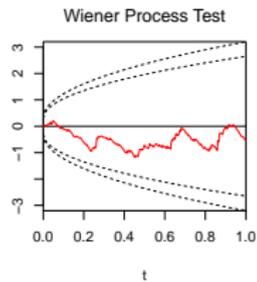
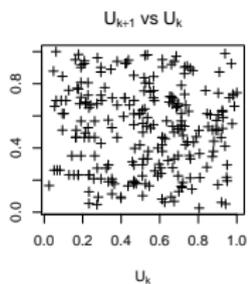
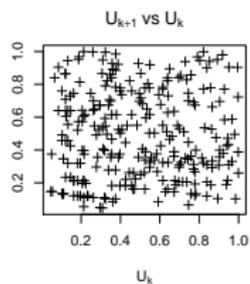
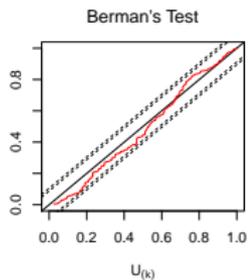
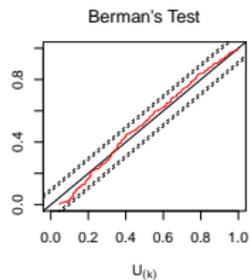
Berman's Test



Wiener Process Test



Modèle 2 « des deux façons »



Trouver le « bon » modèle (3)

- ▶ Nous avons donc deux modèles qui passent les tests. Lequel choisir ?

Trouver le « bon » modèle (3)

- ▶ Nous avons donc deux modèles qui passent les tests. Lequel choisir ?
- ▶ Nous pourrions décider que comme le second modèle est plus simple, nous le gardons.

Trouver le « bon » modèle (3)

- ▶ Nous avons donc deux modèles qui passent les tests. Lequel choisir ?
- ▶ Nous pourrions décider que comme le second modèle est plus simple, nous le gardons.
- ▶ Nous pourrions aussi comparer les probabilités attribuées aux données par chacun des modèles.

Trouver le « bon » modèle (3)

- ▶ Nous avons donc deux modèles qui passent les tests. Lequel choisir ?
- ▶ Nous pourrions décider que comme le second modèle est plus simple, nous le gardons.
- ▶ Nous pourrions aussi comparer les probabilités attribuées aux données pas chacun des modèles. Soient $p_{1,i}$ et $p_{2,i}$ les probabilités d'avoir un PA dans le bloc i suivant les modèles 1 et 2. Alors,

$$\text{Prob}\{Y_i = y_i \mid \text{Model } k\} = p_{k,i}^{y_i} (1 - p_{k,i})^{1-y_i}$$

Trouver le « bon » modèle (3)

- ▶ Nous avons donc deux modèles qui passent les tests. Lequel choisir ?
- ▶ Nous pourrions décider que comme le second modèle est plus simple, nous le gardons.
- ▶ Nous pourrions aussi comparer les probabilités attribuées aux données pas chacun des modèles. Soient $p_{1,i}$ et $p_{2,i}$ les probabilités d'avoir un PA dans le bloc i suivant les modèles 1 et 2. Alors,

$$\text{Prob}\{Y_i = y_i \mid \text{Model } k\} = p_{k,i}^{y_i} (1 - p_{k,i})^{1-y_i}$$

Nous pouvons ainsi calculer le log d'une probabilité pour notre séquence de PAs discrétisée

Trouver le « bon » modèle (3)

- ▶ Nous avons donc deux modèles qui passent les tests. Lequel choisir ?
- ▶ Nous pourrions décider que comme le second modèle est plus simple, nous le gardons.
- ▶ Nous pourrions aussi comparer les probabilités attribuées aux données pas chacun des modèles. Soient $p_{1,i}$ et $p_{2,i}$ les probabilités d'avoir un PA dans le bloc i suivant les modèles 1 et 2. Alors,

$$\text{Prob}\{Y_i = y_i \mid \text{Model } k\} = p_{k,i}^{y_i} (1 - p_{k,i})^{1-y_i}$$

Nous pouvons ainsi calculer le log d'une probabilité pour notre séquence de PAs discrétisée et nous obtenons : **-918.517**
avec le premier modèle

Trouver le « bon » modèle (3)

- ▶ Nous avons donc deux modèles qui passent les tests. Lequel choisir ?
- ▶ Nous pourrions décider que comme le second modèle est plus simple, nous le gardons.
- ▶ Nous pourrions aussi comparer les probabilités attribuées aux données pas chacun des modèles. Soient $p_{1,i}$ et $p_{2,i}$ les probabilités d'avoir un PA dans le bloc i suivant les modèles 1 et 2. Alors,

$$\text{Prob}\{Y_i = y_i \mid \text{Model } k\} = p_{k,i}^{y_i} (1 - p_{k,i})^{1-y_i}$$

Nous pouvons ainsi calculer le log d'une probabilité pour notre séquence de PAs discrétisée et nous obtenons : -918.517 avec le premier modèle et **-925.393 avec le second.**

Trouver le « bon » modèle (3)

- ▶ Nous avons donc deux modèles qui passent les tests. Lequel choisir ?
- ▶ Nous pourrions décider que comme le second modèle est plus simple, nous le gardons.
- ▶ Nous pourrions aussi comparer les probabilités attribuées aux données pas chacun des modèles. Soient $p_{1,i}$ et $p_{2,i}$ les probabilités d'avoir un PA dans le bloc i suivant les modèles 1 et 2. Alors,

$$\text{Prob}\{Y_i = y_i \mid \text{Model } k\} = p_{k,i}^{y_i} (1 - p_{k,i})^{1-y_i}$$

Nous pouvons ainsi calculer le log d'une probabilité pour notre séquence de PAs discrétisée et nous obtenons : -918.517 avec le premier modèle et -925.393 avec le second.

- ▶ Ces nombres sont obtenus avec les données (y_i) de la seconde moitié et les modèles $(p_{k,i})$ ajustés sur la première.

Trouver le « bon » modèle (4)

- ▶ Favoriser la simplicité nous ferait choisir le second modèle alors que maximiser la probabilité nous ferait choisir le second.

Trouver le « bon » modèle (4)

- ▶ Favoriser la simplicité nous ferait choisir le second modèle alors que maximiser la probabilité nous ferait choisir le second.
- ▶ La résolution de ce dilemme passe par la « construction » d'un mètre étalon sur l'échelle des probabilités. Cela nous permettra de savoir si la différence de 7 que nous venons de trouver est grande ou petite.

Trouver le « bon » modèle (4)

- ▶ Favoriser la simplicité nous ferait choisir le second modèle alors que maximiser la probabilité nous ferait choisir le second.
- ▶ La résolution de ce dilemme passe par la « construction » d'un mètre étalon sur l'échelle des probabilités. Cela nous permettra de savoir si la différence de 7 que nous venons de trouver est grande ou petite.
- ▶ Nous allons essayer de calibrer notre échelle avec un *bootstrap* paramétrique (Davison et Hinkley, 1997).

Trouver le « bon » modèle (4)

- ▶ Favoriser la simplicité nous ferait choisir le second modèle alors que maximiser la probabilité nous ferait choisir le second.
- ▶ La résolution de ce dilemme passe par la « construction » d'un mètre étalon sur l'échelle des probabilités. Cela nous permettra de savoir si la différence de 7 que nous venons de trouver est grande ou petite.
- ▶ Nous allons essayer de calibrer notre échelle avec un *bootstrap* paramétrique (Davison et Hinkley, 1997).
 - ▶ Nous supposons que le modèle k ajusté à la première moitié des données est correcte.

Trouver le « bon » modèle (4)

- ▶ Favoriser la simplicité nous ferait choisir le second modèle alors que maximiser la probabilité nous ferait choisir le second.
- ▶ La résolution de ce dilemme passe par la « construction » d'un mètre étalon sur l'échelle des probabilités. Cela nous permettra de savoir si la différence de 7 que nous venons de trouver est grande ou petite.
- ▶ Nous allons essayer de calibrer notre échelle avec un *bootstrap* paramétrique (Davison et Hinkley, 1997).
 - ▶ Nous supposons que le modèle k ajusté à la première moitié des données est correcte.
 - ▶ Nous simulons 500 séquences de PAs correspondant à la seconde moitié (avec un « éclaircissage »).

Trouver le « bon » modèle (4)

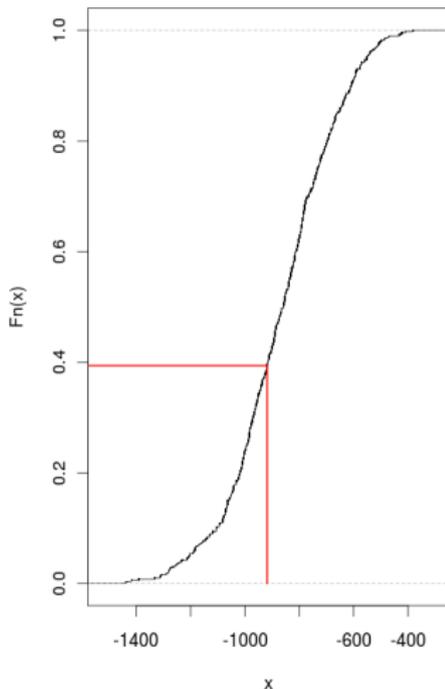
- ▶ Favoriser la simplicité nous ferait choisir le second modèle alors que maximiser la probabilité nous ferait choisir le second.
- ▶ La résolution de ce dilemme passe par la « construction » d'un mètre étalon sur l'échelle des probabilités. Cela nous permettra de savoir si la différence de 7 que nous venons de trouver est grande ou petite.
- ▶ Nous allons essayer de calibrer notre échelle avec un *bootstrap* paramétrique (Davison et Hinkley, 1997).
 - ▶ Nous supposons que le modèle k ajusté à la première moitié des données est correcte.
 - ▶ Nous simulons 500 séquences de PAs correspondant à la seconde moitié (avec un « éclaircissage »).
 - ▶ Nous calculons le log de la probabilité **avec chacun des deux modèles.**

Trouver le « bon » modèle (4)

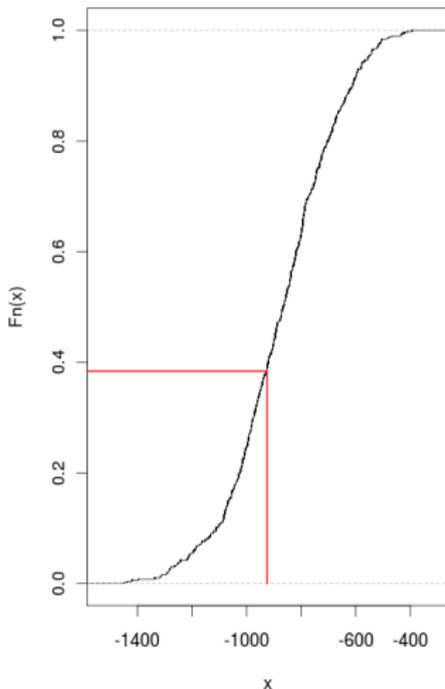
- ▶ Favoriser la simplicité nous ferait choisir le second modèle alors que maximiser la probabilité nous ferait choisir le second.
- ▶ La résolution de ce dilemme passe par la « construction » d'un mètre étalon sur l'échelle des probabilités. Cela nous permettra de savoir si la différence de 7 que nous venons de trouver est grande ou petite.
- ▶ Nous allons essayer de calibrer notre échelle avec un *bootstrap* paramétrique (Davison et Hinkley, 1997).
 - ▶ Nous supposons que le modèle k ajusté à la première moitié des données est correcte.
 - ▶ Nous simulons 500 séquences de PAs correspondant à la seconde moitié (avec un « éclaircissage »).
 - ▶ Nous calculons le log de la probabilité avec chacun des deux modèles.
 - ▶ Nous « résumons ».

Log proba quand le premier modèle est vrai

log prob with Model 1 when Model 1 is true

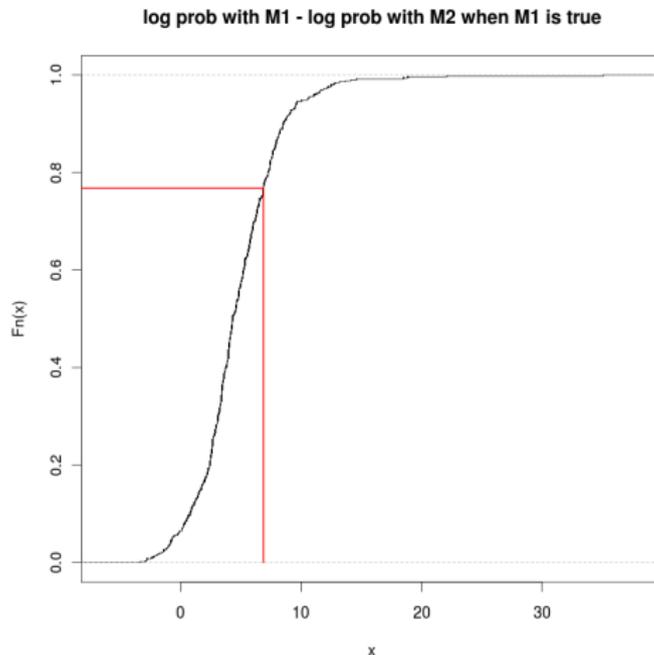


log prob with Model 2 when Model 1 is true



Les lignes rouges correspondent aux valeurs observées (sur les vraies données).

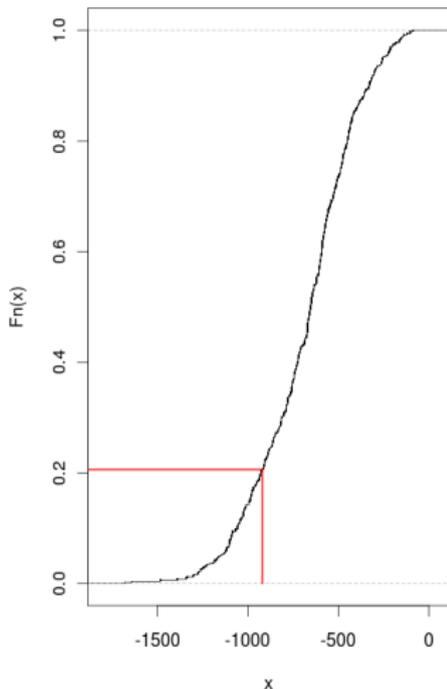
Différence des log proba quand 1 est vrai



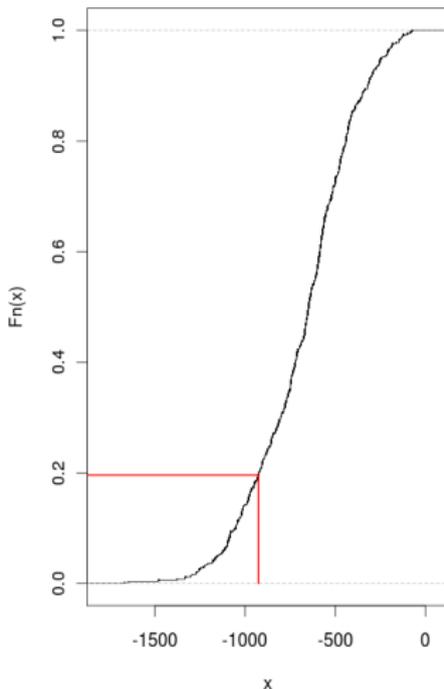
En rouge, la valeur observée. La valeur moyenne de cette différence, 4.78 ± 0.16 , est une estimation de la divergence de Kullback-Leibler entre les modèles 1 et 2.

Log proba quand le second modèle est vrai

log prob with Model 1 when Model 2 is true

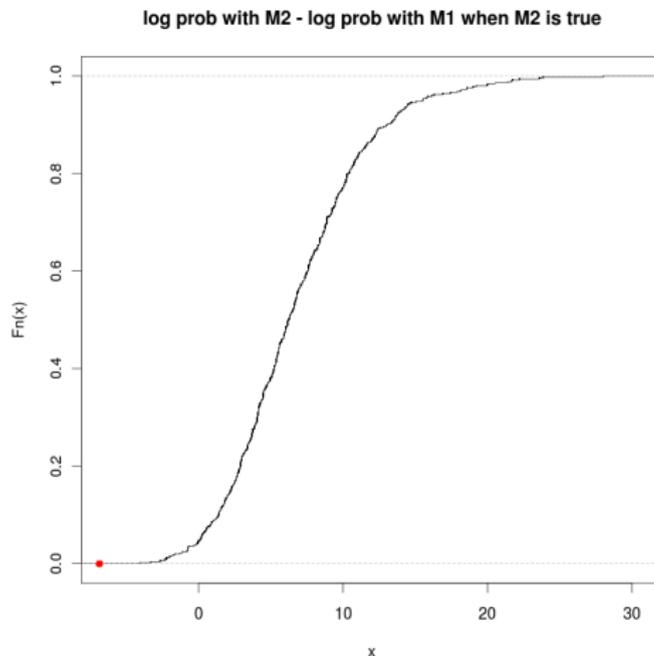


log prob with Model 2 when Model 2 is true



Les lignes rouges correspondent aux valeurs observées (sur les vraies données).

Différence des log proba quand 2 est vrai



En rouge, la valeur observée. La valeur moyenne de cette différence, 6.85 ± 0.22 , est une estimation de la divergence de Kullback-Leibler entre les modèles 2 et 1.

Trouver le « bon » modèle (5)

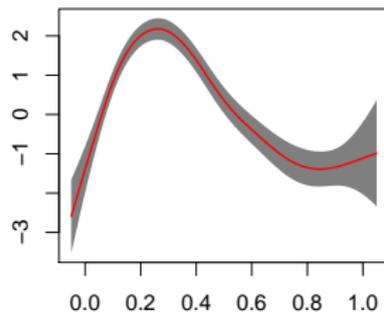
- ▶ Notre « calibration de *bootstrap* » exclut clairement le second modèle.
- ▶ Nous restons donc avec le modèle le plus complexe :

$$\eta(d_1, d_2) = \eta_{\emptyset} + \eta_1(d_1) + \eta_2(d_2) + \eta_{1,2}(d_1, d_2)$$

- ▶ Les figures qui suivent ont été obtenues en ajustant ce modèle à l'ensemble des données.

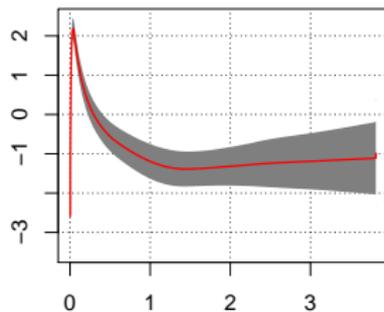
Formes fonctionnelles des termes univariés

Elapsed time since last spike



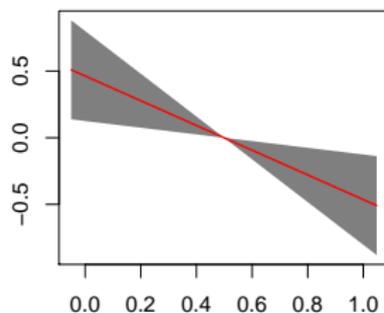
Probability scale

Elapsed time since last spike



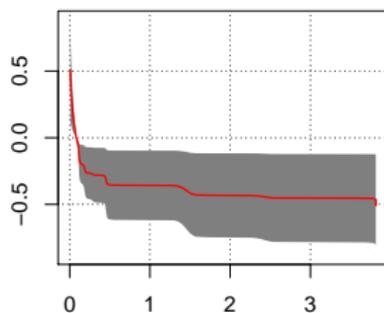
Time (s)

Last ISI



Probability scale

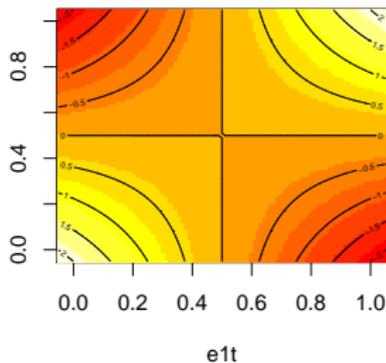
Last ISI



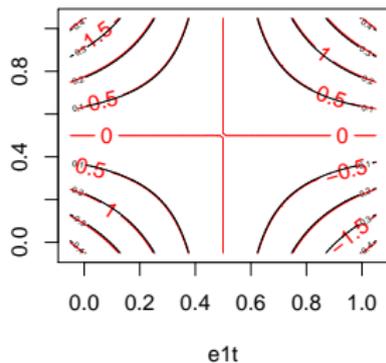
Time (s)

Forme fonctionnelle du terme d'interaction

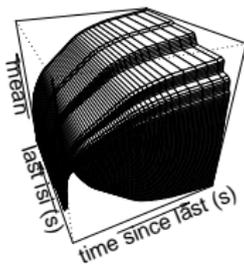
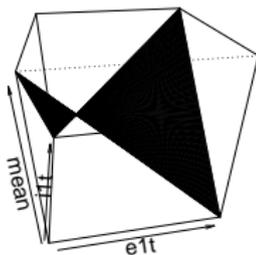
term e1t:i1t



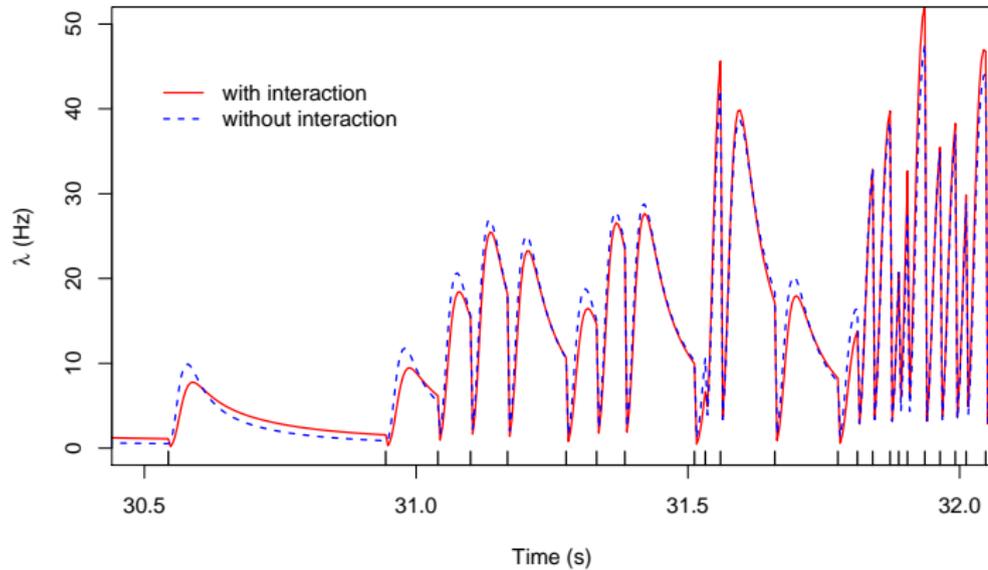
term e1t:i1t



Mean of term e1t:i1t



Intensités des deux modèles



Remerciements

- ▶ Merci pour votre attention.
- ▶ Merci à Antoine Chaffiol pour les données et le tri des PAs.
- ▶ Merci à Vilmos Prokaj, Jonathan Touboul et Olivier Faugeras pour m'avoir « signalé » le théorème de Donsker.