

Deux stratégies non-paramétriques pour estimer le taux de saut d'un processus markovien déterministe par morceaux

Romain Azaïs (Inria Nancy)

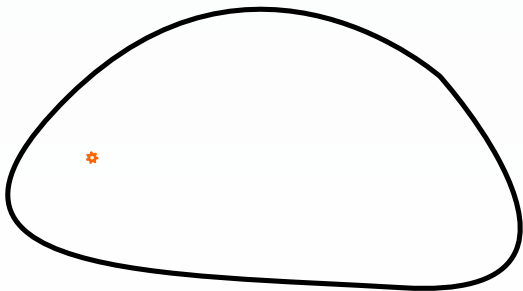
Marseille – 24 juin 2016

Processus markoviens déterministes par morceaux (PDMP) :

- Introduits dans les années 80 par Davis
- The class is “wide enough to include as special cases virtually all the non-diffusion models of applied probability”
- MC, CT-MC, M/G/1 queue, G/G/1 queue \subset PDMP
- Motivation initiale : optimisation

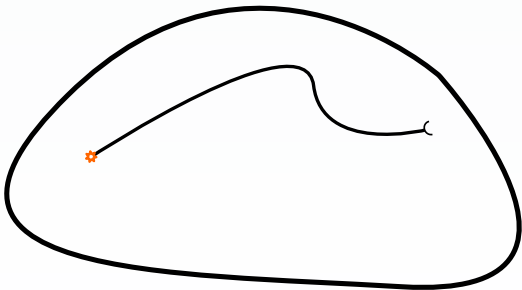
Processus markoviens déterministes par morceaux (PDMP) :

- Introduits dans les années 80 par Davis
- The class is “wide enough to include as special cases virtually all the non-diffusion models of applied probability”
- MC, CT-MC, M/G/1 queue, G/G/1 queue \subset PDMP
- Motivation initiale : optimisation



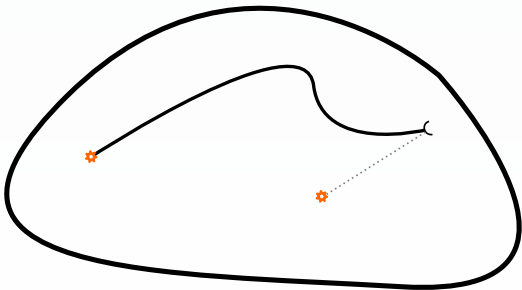
Processus markoviens déterministes par morceaux (PDMP) :

- Introduits dans les années 80 par Davis
- The class is “wide enough to include as special cases virtually all the non-diffusion models of applied probability”
- MC, CT-MC, M/G/1 queue, G/G/1 queue \subset PDMP
- Motivation initiale : optimisation



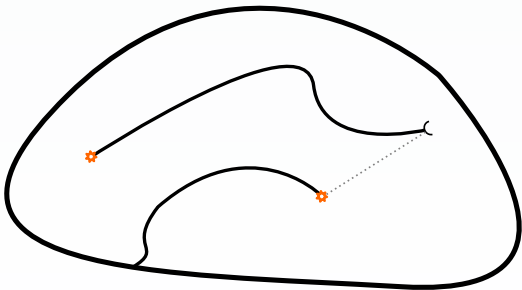
Processus markoviens déterministes par morceaux (PDMP) :

- Introduits dans les années 80 par Davis
- The class is “wide enough to include as special cases virtually all the non-diffusion models of applied probability”
- MC, CT-MC, M/G/1 queue, G/G/1 queue \subset PDMP
- Motivation initiale : optimisation



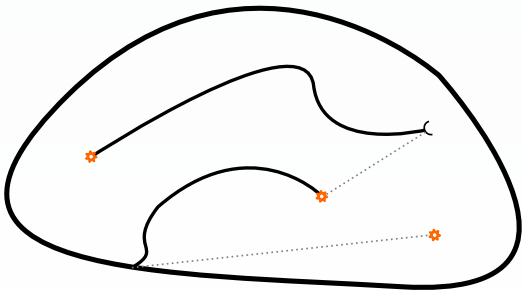
Processus markoviens déterministes par morceaux (PDMP) :

- Introduits dans les années 80 par Davis
- The class is “wide enough to include as special cases virtually all the non-diffusion models of applied probability”
- MC, CT-MC, M/G/1 queue, G/G/1 queue \subset PDMP
- Motivation initiale : optimisation



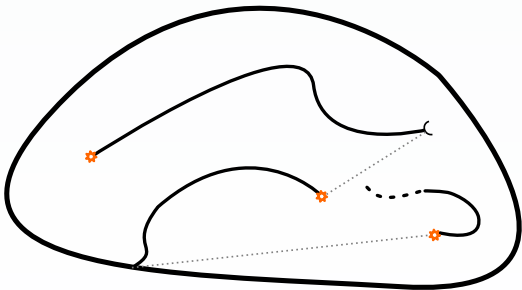
Processus markoviens déterministes par morceaux (PDMP) :

- Introduits dans les années 80 par Davis
- The class is “wide enough to include as special cases virtually all the non-diffusion models of applied probability”
- MC, CT-MC, M/G/1 queue, G/G/1 queue \subset PDMP
- Motivation initiale : optimisation



Processus markoviens déterministes par morceaux (PDMP) :

- Introduits dans les années 80 par Davis
- The class is “wide enough to include as special cases virtually all the non-diffusion models of applied probability”
- MC, CT-MC, M/G/1 queue, G/G/1 queue \subset PDMP
- Motivation initiale : optimisation



Un PDMP (X_t) sur $(E, \mathcal{B}(E))$ est défini à partir de 3 caractéristiques locales :

- le flot déterministe $\Phi : \mathbb{R} \times E \rightarrow E$
- le taux de saut $\lambda : E \rightarrow \mathbb{R}_+$
- le noyau de transition $Q : (\mathcal{B}(E), E) \rightarrow [0, 1]$

Un PDMP (X_t) sur $(E, \mathcal{B}(E))$ est défini à partir de 3 caractéristiques locales :

- le flot déterministe $\Phi : \mathbb{R} \times E \rightarrow E$
- le taux de saut $\lambda : E \rightarrow \mathbb{R}_+$
- le noyau de transition $Q : (\mathcal{B}(E), E) \rightarrow [0, 1]$

Temps déterministe de sortie de E suivant le flot :

$$t^+(x) = \inf\{t > 0 : \Phi(t|x) \in \partial E\}$$

Un PDMP (X_t) sur $(E, \mathcal{B}(E))$ est défini à partir de 3 caractéristiques locales :

- le flot déterministe $\Phi : \mathbb{R} \times E \rightarrow E$
- le taux de saut $\lambda : E \rightarrow \mathbb{R}_+$
- le noyau de transition $Q : (\mathcal{B}(E), E) \rightarrow [0, 1]$

Temps déterministe de sortie de E suivant le flot :

$$t^+(x) = \inf\{t > 0 : \Phi(t|x) \in \partial E\}$$

Conditions :

$$\int_0^\varepsilon \lambda(\Phi(t|x)) dt < \infty \quad \text{et} \quad Q(E \setminus \{x}|x) = 1$$

no explosion *jumps are true jumps*

Un PDMP (X_t) sur $(E, \mathcal{B}(E))$ est défini à partir de 3 caractéristiques locales :

- le flot déterministe $\Phi : \mathbb{R} \times E \rightarrow E$
- le taux de saut $\lambda : E \rightarrow \mathbb{R}_+$
- le noyau de transition $Q : (\mathcal{B}(E), E) \rightarrow [0, 1]$

Temps déterministe de sortie de E suivant le flot :

$$t^+(x) = \inf\{t > 0 : \Phi(t|x) \in \partial E\}$$

Conditions :

$$\int_0^{t^+(x)} \lambda(\Phi(t|x)) dt < \infty \quad \text{et} \quad Q(E \setminus \{x\}|x) = 1$$

no explosion *jumps are true jumps*

L'espace d'état typique est $E = \bigcup_{m \in M} \{m\} \times E_m$, avec $M \subset \mathbb{N}$ et $E_m \subset \mathbb{R}^{d_m}$

Partant de la condition initiale $X_0 = x$, la loi de T_1 est donnée par

$$\mathbb{P}(T_1 > t) = \exp\left(-\int_0^t \lambda(\Phi(s|x)) ds\right) \mathbb{1}_{\{0 \leq t < t^+(x)\}}$$

Partant de la condition initiale $X_0 = x$, la loi de T_1 est donnée par

$$\mathbb{P}(T_1 > t) = \exp\left(-\int_0^t \lambda(\Phi(s|x)) ds\right) \mathbb{1}_{\{0 \leq t < t^+(x)\}}$$

Entre les temps 0 et T_1 , le processus évolue selon le flot,

$$\forall 0 \leq t < T_1, X_t = \Phi(t|x)$$

Partant de la condition initiale $X_0 = x$, la loi de T_1 est donnée par

$$\mathbb{P}(T_1 > t) = \exp\left(-\int_0^t \lambda(\Phi(s|x)) ds\right) \mathbb{1}_{\{0 \leq t < t^+(x)\}}$$

Entre les temps 0 et T_1 , le processus évolue selon le flot,

$$\forall 0 \leq t < T_1, X_t = \Phi(t|x)$$

A l'instant T_1 le processus "saute" selon le noyau Q ,

$$\mathbb{E}[\varphi(X_{T_1}) | \Phi(T_1|x)] = \int \varphi(u) Q(du | \Phi(T_1|x))$$

Partant de la condition initiale $X_0 = x$, la loi de T_1 est donnée par

$$\mathbb{P}(T_1 > t) = \exp\left(-\int_0^t \lambda(\Phi(s|x)) ds\right) \mathbb{1}_{\{0 \leq t < t+(x)\}}$$

Entre les temps 0 et T_1 , le processus évolue selon le flot,

$$\forall 0 \leq t < T_1, X_t = \Phi(t|x)$$

A l'instant T_1 le processus "saute" selon le noyau Q ,

$$\mathbb{E}[\varphi(X_{T_1}) | \Phi(T_1|x)] = \int \varphi(u)Q(du|\Phi(T_1|x))$$

Et ainsi de suite, on détermine :

- le temps de parcours $S_n = T_n - T_{n-1}$ partant de $X_{T_{n-1}}$
- $X_{T_{n-1}+s} = \Phi(s|X_{T_{n-1}})$ pour $0 \leq s < S_n$
- X_{T_n} selon $Q(\cdot|\Phi(S_n|X_{T_{n-1}}))$

Résultats récents en probabilités et probabilités numériques :

- Ergodicité, stabilité : Costa & Dufour (2008), Last (2004)
- Vitesse de convergence à l'équilibre (modèles particuliers) : Chafaï *et al.* (2010), Fontbona *et al.* (2012), etc
- Méthodes numériques (via quantification optimale) : de Saporta *et al.* (2010), Brandejsky *et al.* (2012)
- Contrôle optimal : Costa & Dufour (2013), etc
- En dimension infinie (avec des applications en neurosciences) : Riedler *et al.* (2012), Genadot & Thieullen (2012, 2014)

Résultats récents en probabilités et probabilités numériques :

- Ergodicité, stabilité : Costa & Dufour (2008), Last (2004)
- Vitesse de convergence à l'équilibre (modèles particuliers) : Chafaï *et al.* (2010), Fontbona *et al.* (2012), etc
- Méthodes numériques (via quantification optimale) : de Saporta *et al.* (2010), Brandejsky *et al.* (2012)
- Contrôle optimal : Costa & Dufour (2013), etc
- En dimension infinie (avec des applications en neurosciences) : Riedler *et al.* (2012), Genadot & Thieullen (2012, 2014)

- Statistiques (cadre "général") : avant les années 2010, écriture de la vraisemblance (Jacobsen, 2006)

Motivation

Proposer des méthodes statistiques pour les PDMP dans le cadre “le plus général possible” :

- Bien comprendre les quantités statistiques en jeu
- Méthodes applicables dans des contextes variés

Motivation

Proposer des méthodes statistiques pour les PDMP dans le cadre “le plus général possible” :

- Bien comprendre les quantités statistiques en jeu
- Méthodes applicables dans des contextes variés



Souvent il vaut mieux exploiter les caractéristiques particulières du modèle considéré : Soubeyrand *et al.* (2009), Krell (2016), Hodara *et al.* (2016)

La question a-t-elle un intérêt ?

- Tout l'aléa est contenu dans la chaîne de Markov (Z_n, S_{n+1})
 - estimation pour des chaînes de Markov sur $E \times \mathbb{R}_+$
 - interprétabilité ?

La question a-t-elle un intérêt ?

- Tout l'aléa est contenu dans la chaîne de Markov (Z_n, S_{n+1})
 - estimation pour des chaînes de Markov sur $E \times \mathbb{R}_+$
 - interprétabilité ?
- Sauts forcés à la frontière
 - temps S_{n+1} censurés par une horloge déterministe fonction de Z_n
 - $\Phi(S_{n+1}|Z_n)$ peut se situer sur la frontière de l'espace d'état (contrairement à Z_n)

La question a-t-elle un intérêt ?

- Tout l'aléa est contenu dans la chaîne de Markov (Z_n, S_{n+1})
 - estimation pour des chaînes de Markov sur $E \times \mathbb{R}_+$
 - interprétabilité ?
- Sauts forcés à la frontière
 - temps S_{n+1} censurés par une horloge déterministe fonction de Z_n
 - $\Phi(S_{n+1}|Z_n)$ peut se situer sur la frontière de l'espace d'état (contrairement à Z_n)
- Conditionnellement à Z_n , la variable aléatoire $\Phi(S_{n+1}|Z_n)$ vit dans un espace (curviligne) de dimension 1
 - estimation de $Q \neq$ estimation du noyau de transition d'une MC

La question a-t-elle un intérêt ?

- Tout l'aléa est contenu dans la chaîne de Markov (Z_n, S_{n+1})
 - estimation pour des chaînes de Markov sur $E \times \mathbb{R}_+$
 - interprétabilité ?
- Sauts forcés à la frontière
 - temps S_{n+1} censurés par une horloge déterministe fonction de Z_n
 - $\Phi(S_{n+1}|Z_n)$ peut se situer sur la frontière de l'espace d'état (contrairement à Z_n)
- Conditionnellement à Z_n , la variable aléatoire $\Phi(S_{n+1}|Z_n)$ vit dans un espace (curviligne) de dimension 1
 - estimation de $Q \neq$ estimation du noyau de transition d'une MC
- Conditionnellement à Z_n , Z_{n+1} contient de l'information sur S_{n+1}
 - modèle à intensité multiplicative non vérifié pour l'estimation de λ

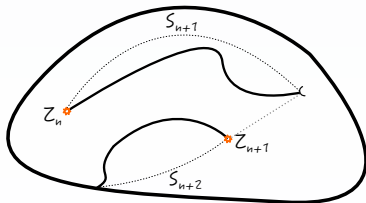
La question a-t-elle un intérêt ?

- Tout l'aléa est contenu dans la chaîne de Markov (Z_n, S_{n+1})
 - estimation pour des chaînes de Markov sur $E \times \mathbb{R}_+$
 - interprétabilité ?
- Sauts forcés à la frontière
 - temps S_{n+1} censurés par une horloge déterministe fonction de Z_n
 - $\Phi(S_{n+1}|Z_n)$ peut se situer sur la frontière de l'espace d'état (contrairement à Z_n)
- Conditionnellement à Z_n , la variable aléatoire $\Phi(S_{n+1}|Z_n)$ vit dans un espace (curviligne) de dimension 1
 - estimation de $Q \neq$ estimation du noyau de transition d'une MC
- Conditionnellement à Z_n , Z_{n+1} contient de l'information sur S_{n+1}
 - modèle à intensité multiplicative non vérifié pour l'estimation de λ
- $\lambda(\Phi(t|\xi)) = \lambda(x)$ dès lors que $\Phi(t|\xi) = x$
 - classe indénombrable d'estimateurs de $\lambda(x)$

Aujourd'hui

Estimation de la loi des temps inter-sauts \simeq estimation du taux de saut λ

- Nonparametric estimation of the conditional distribution of the inter-jumping times for PDMP's (avec Dufour & Gégout-Petit, 2014)
- Optimal choice among a class of nonparametric estimators of the jump rate for PDMP's (avec Muller-Gueudin, 2015)
- A new characterization of the jump rate for PDMP's with discrete transitions (avec Genadot, 2016)



Avant de commencer (1/2)

Fonction de survie conditionnelle :

$$\mathbb{P}(S_{n+1} > t | Z_n = x) = \underbrace{\exp\left(-\int_0^t \lambda(\Phi(s|x)) ds\right)}_{=G(t|x)} \mathbb{1}_{\{0 \leq t < t^+(x)\}}$$

Densité conditionnelle sur $[0, t^+(x))$:

$$\begin{aligned} f(t|x) &= -\partial_t G(t|x) \\ &= \lambda(\Phi(t|x))G(t|x) \end{aligned}$$

$$\text{NB : } \int_0^{t^+(x)} f(t|x) dt + G(t^+(x)|x) = 1$$

Avant de commencer (2/2)

Hypothèse principale : ergodicité du processus à temps discret sous-jacent (Z_n) , assurée par la condition de Döblin,

$$R(dy|x) \geq \varepsilon \eta(dy)$$

→ En particulier, on pourra appliquer le théorème ergodique presque sûr

Hypothèses de régularité peu restrictives (Lipschitz uniforme)...



mais t^+ est supposée bornée (sauf dans la seconde partie de l'exposé)

Stratégie 1

Appliquer le modèle à intensité multiplicative à tout prix

Modèle multiplicatif d'Aalen

$N(t)$ processus de comptage tel que

$$M(t) = N(t) - \int_0^t \lambda(s)Y(s)ds$$

est une martingale. Dans ce cadre,

$$\hat{\Lambda}(t) = \int_0^t \frac{dN(s)}{Y(s)}$$

est l'estimateur de Nelson-Aalen de $\Lambda(t) = \int_0^t \lambda(s)ds$.

Résultat

Pour tout entier i , le processus

$$t \mapsto \mathbb{1}_{\{S_{i+1} \leq t\}} - \int_0^t \lambda(\Phi(s|Z_i)) \mathbb{1}_{\{S_{i+1} \geq s\}} ds$$

est une martingale pour la filtration $(\sigma(Z_i) \wedge \mathcal{F}_t^{i+1})_{0 \leq t < t^+(Z_i)}$,

où (\mathcal{F}_t^{i+1}) est la filtration naturelle associée au processus

$$t \mapsto \mathbb{1}_{\{S_{i+1} \leq t\}}$$

Suite naturelle : sommer ces martingales de telle sorte que la somme reste une martingale

Résultat

Sous l'hypothèse $\mathbb{E}[\varphi(Z_{n+1})|Z_n, S_{n+1}] = \mathbb{E}[\varphi(Z_{n+1})|Z_n]$, le processus

$$t \mapsto \sum_{i=0}^{n-1} \left[\mathbb{1}_{\{S_{i+1} \leq t\}} - \int_0^t \lambda(\Phi(s|Z_i)) \mathbb{1}_{\{S_{i+1} \geq s\}} ds \right]$$

est une martingale pour la filtration

$$\left(\sigma(Z_0, \dots, Z_{n-1}) \vee \bigvee_{i=0}^{n-1} \mathcal{F}_t^{i+1} \right)$$

Résultat

Sous l'hypothèse $\mathbb{E}[\varphi(Z_{n+1})|Z_n, S_{n+1}] = \mathbb{E}[\varphi(Z_{n+1})|Z_n]$, le processus

$$t \mapsto \sum_{i=0}^{n-1} \left[\mathbb{1}_{\{S_{i+1} \leq t\}} - \int_0^t \lambda(\Phi(s|Z_i)) \mathbb{1}_{\{S_{i+1} \geq s\}} ds \right]$$

est une martingale pour la filtration

$$\left(\sigma(Z_0, \dots, Z_{n-1}) \vee \bigvee_{i=0}^{n-1} \mathcal{F}_t^{i+1} \right)$$

Cas typiques pour lesquels l'hypothèse est vérifiée

- Processus de renouvellement \simeq "flot constant"
- $Q(\cdot|x) \equiv \mu(\cdot)$

Cas général

La connaissance de Z_{n+1} peut apporter de l'information sur S_{n+1} .

Exemple typique : $Q(\cdot|x) = \delta_{\{x\}}$ et flot injectif. Sachant Z_n et Z_{n+1} , la loi de S_{n+1} est une Dirac en l'unique t tel que

$$\Phi(t|Z_n) = Z_{n+1}$$

→ Le processus “somme” n'est en général pas une martingale

→ Etudier S_{n+1} conditionnellement à Z_n et Z_{n+1}

Hypothèse : Q admet une densité par rapport à une certaine mesure μ

Résultat

On a

$$\mathbb{P}(S_{n+1} > t | Z_n, Z_{n+1}) = \tilde{G}(t | Z_n, Z_{n+1}) \mathbb{1}_{\{0 \leq t < t^+(Z_n)\}},$$

où $\tilde{G}(t|x, y)$ est la fonction de survie associée au taux

$$\tilde{\lambda}(t|x, y) = \frac{f(t|x)Q(y|\Phi(t|x))}{\int_t^{t^+(x)} f(t|x)Q(y|\Phi(t|x)) + G(t^*(x)|x)Q(y|\Phi(t^*(x)|x))}$$

Et on a la structure de martingale recherchée : le processus

$$t \mapsto \sum_{i=0}^{n-1} \left[\mathbb{1}_{\{S_{i+1} \leq t\}} - \int_0^t \tilde{\lambda}(s | Z_i, Z_{i+1}) \mathbb{1}_{\{S_{i+1} \geq s\}} ds \right]$$

est une martingale dans la filtration

$$\left(\sigma(Z_0, \dots, Z_n) \vee \bigvee_{i=0}^{n-1} \mathcal{F}_t^{i+1} \right)$$

Et on a la structure de martingale recherchée : le processus

$$t \mapsto \sum_{i=0}^{n-1} \left[\mathbb{1}_{\{S_{i+1} \leq t\}} - \int_0^t \tilde{\lambda}(s|Z_i, Z_{i+1}) \mathbb{1}_{\{S_{i+1} \geq s\}} ds \right]$$

est une martingale dans la filtration

$$\left(\sigma(Z_0, \dots, Z_n) \vee \bigvee_{i=0}^{n-1} \mathcal{F}_t^{i+1} \right)$$

Corollaire

S'il existe \mathcal{A} tel que $\#\mathcal{A} < \infty$ et pour tout x , $Q(\mathcal{A}|x) = 1$, alors, pour $(x, y) \in \mathcal{A}^2$, le compensateur du processus de comptage

$$t \mapsto \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \leq t\}} \mathbb{1}_{\{Z_i = x\}} \mathbb{1}_{\{Z_{i+1} = y\}}$$

est

$$t \mapsto \int_0^t \tilde{\lambda}(s|x, y) \left[\sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \geq s\}} \mathbb{1}_{\{Z_i = x\}} \mathbb{1}_{\{Z_{i+1} = y\}} \right] ds$$

Modèle à intensité multiplicative vérifié lorsque les transitions sont discrètes

→ Estimateur de Nelson-Aalen de $\tilde{\Lambda}(t|x, y)$

→ “Lissage” pour estimer $\tilde{\lambda}(t|x, y)$

→ Retour au processus de départ

Modèle à intensité multiplicative vérifié lorsque les transitions sont discrètes

→ Estimateur de Nelson-Aalen de $\tilde{\Lambda}(t|x, y)$

→ “Lissage” pour estimer $\tilde{\lambda}(t|x, y)$

→ Retour au processus de départ

Deux techniques pour “lisser” l’estimateur de Nelson-Aalen

- Méthode à noyau proposée par Ramlau-Hansen (83) : appliquée dans le cas général (transitions quelconques)
- Décomposition de $\tilde{\lambda}$ dans une base : seulement pour les transitions discrètes

Transitions discrètes : décomposition de $\tilde{\lambda}$

Hypothèse additionnelle : t^+ est bornée sur \mathcal{A}

On décompose $\tilde{\lambda}(\cdot|x, y)$ dans une base orthonormée de $\mathbb{L}_{[0,1]}^2$:

$$\tilde{\lambda}(st^+(x)|x, y) = \sum_{p \geq 0} \theta_p(x, y) B_p(s)$$

Résultat

$$\lambda(x) = \sum_{p \geq 0} B_p(0) \sum_{y \in \mathcal{A}} R(\{y\}|x) \theta_p(x, y),$$

où R est le noyau de transition de la MC (Z_n)

Estimation

- R est estimé par son équivalent empirique :

$$\widehat{R}^n(\{y\}|x) = \frac{\#\{Z_i = x, Z_{i+1} = y\}}{\#\{Z_i = x\}}.$$

- $\theta_p(x, y)$ est estimé par :

$$\widehat{\theta}_p^n(x, y) = t^+(x)^{-1} \int_0^1 B_p(s) d\widehat{\Lambda}^n(s|x, y)$$

Estimation

- R est estimé par son équivalent empirique :

$$\widehat{R}^n(\{y\}|x) = \frac{\#\{Z_i = x, Z_{i+1} = y\}}{\#\{Z_i = x\}}.$$

- $\theta_p(x, y)$ est estimé par :

$$\widehat{\theta}_p^n(x, y) = t^+(x)^{-1} \int_0^1 B_p(s) d\widehat{\Lambda}^n(s|x, y)$$

Résultat

$$\widehat{\theta}_p^n(x, y) \xrightarrow{\mathbb{P}} \theta_p(x, y)$$

$$\sqrt{n} \left(\widehat{\theta}_p^n(x, y) - \theta_p(x, y) \right) \xrightarrow{d} \mathcal{N}(0, \sigma_p^2(x, y)).$$

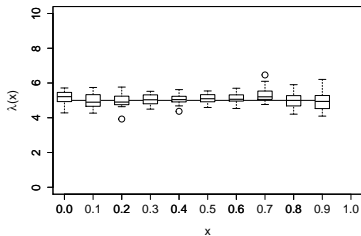
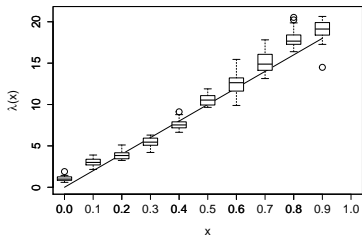
En particulier, on en déduit un test de nullité de $\theta_p(x, y)$

Estimation de λ en seuillant la série :

$$\hat{\lambda}^n(x) = \sum_{p=0}^{\tau_n} B_p(0) \sum_{y \in \mathcal{A}} \hat{R}^n(\{y\}|x) \hat{\theta}_p^n(x, y)$$

Résultat

$$\sup_{x \in \mathcal{A}} \left| \hat{\lambda}^n(x) - \lambda(x) \right| \xrightarrow{\mathbb{P}} 0$$



Transitions quelconques : lissage à noyau

Approximation $l(s|A, B)$ de $\tilde{\lambda}(s|x, y)$ pour $x \in A$ et $y \in B$:

$$M_{A,B}^n(t) = \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \leq t\}} \mathbb{1}_{\{Z_i \in A, Z_{i+1} \in B\}} - \int_0^t l(s|A, B) Y_{A,B}^n(s) ds \\ - \int_0^t \sum_{i=0}^{n-1} \left(\tilde{\lambda}(s|Z_i, Z_{i+1}) - l(s|A, B) \right) \mathbb{1}_{\{S_{i+1} \geq s\}} \mathbb{1}_{\{Z_i \in A, Z_{i+1} \in B\}} ds,$$

$$\text{où } Y_{A,B}^n(s) = \sum_{i=0}^{n-1} \mathbb{1}_{\{S_{i+1} \geq s\}} \mathbb{1}_{\{Z_i \in A, Z_{i+1} \in B\}}$$

→ Modèle à intensité multiplicative avec un terme de reste

→ Comment choisir l ?

Choix de l'approximation $l(t|A, B)$

- Le terme de reste doit aller vers 0
- $l(t|A, B)$ doit approcher la fonction qui nous intéresse $\tilde{\lambda}(t|x, y)$

Choix de l'approximation $l(t|A, B)$

- Le terme de reste doit aller vers 0
- $l(t|A, B)$ doit approcher la fonction qui nous intéresse $\tilde{\lambda}(t|x, y)$

$$l(t|A, B) = \frac{\int_{A \times B} \tilde{f}(t|x, y) \tilde{\nu}(dx \times dy)}{\int_{A \times B} \tilde{G}(t|x, y) \tilde{\nu}(dx \times dy)},$$

où $\tilde{\nu}$ est la loi invariante du couple (Z_n, Z_{n+1})

Il s'agit d'une "moyennisation spatiale" de $\tilde{\lambda}(t|x, y)$ contre la loi invariante

Estimation de l'approximation $l(t|A, B)$

On passe par l'estimation de $L(t|A, B) = \int_0^t l(s|A, B)ds$,

$$\widehat{L}^n(t|A, B) - L^{n,*}(t|A, B) = \widetilde{M}_{A,B}^n(t) + a^n(t),$$

où $\widehat{L}^n(t|A, B)$ est l'estimateur de Nelson-Aalen

Estimation de l'approximation $l(t|A, B)$

On passe par l'estimation de $L(t|A, B) = \int_0^t l(s|A, B)ds$,

$$\widehat{L}^n(t|A, B) - L^{n,*}(t|A, B) = \widetilde{M}_{A,B}^n(t) + a^n(t),$$

où $\widehat{L}^n(t|A, B)$ est l'estimateur de Nelson-Aalen

Résultat

Lissage à noyau de $\widehat{L}^n(t|A, B)$ pour obtenir $\widehat{l}^n(t|A, B)$:

$$\sup_{r_1 \leq t \leq r_2} \left| \widehat{l}^n(t|A, B) - l(t|A, B) \right| \xrightarrow{\mathbb{P}} 0$$

Retour au processus initial

Le lien entre f et l est donné par

$$\left| f(t|x) - \sum_k l(t|A, B_k)P(t, B_k|A) \right| \leq C \max_k \text{diam}(A \times B_k),$$

où (B_k) est une partition (finie) de E ,

et où $P(t, B_k|A)$ est définie comme la probabilité conditionnelle

$$\mathbb{P}_\nu(S_1 > t, Z_1 \in B_k | Z_0 \in A)$$

sous la mesure invariante ν de (Z_n)

On approche $f(t|x)$ par $f(t|A) = \sum_k l(t|A, B_k) P(t, B_k|A)$

et on estime $f(t|A)$ par $\hat{f}^n(t|A) = \sum_k \hat{l}^n(t|A, B_k) \hat{P}^n(t, B_k|A)$,

où

$$\hat{P}^n(t, B_k|A) = \frac{\#\{S_{i+1} > t, Z_{i+1} \in B_k\}}{\#\{Z_i \in A\}}$$

estime la probabilité conditionnelle $\mathbb{P}_\nu(S_1 > t, Z_1 \in B_k | Z_0 \in A)$

On approche $f(t|x)$ par $f(t|A) = \sum_k l(t|A, B_k) P(t, B_k|A)$
et on estime $f(t|A)$ par $\hat{f}^n(t|A) = \sum_k \hat{l}^n(t|A, B_k) \hat{P}^n(t, B_k|A)$,

où

$$\hat{P}^n(t, B_k|A) = \frac{\#\{S_{i+1} > t, Z_{i+1} \in B_k\}}{\#\{Z_i \in A\}}$$

estime la probabilité conditionnelle $\mathbb{P}_\nu(S_1 > t, Z_1 \in B_k | Z_0 \in A)$

Résultat

Convergence (uniforme) en probabilité de $\hat{f}^n(t|A)$ vers $f(t|A)$ qui est une approximation aussi précise que l'on veut de $f(t|x)$

Stratégie 2

Estimer $\lambda(\Phi(t|x))$ comme $f(t|x)/G(t|x)$

Estimateurs à noyau

$$\widehat{\mathcal{F}}^n(t|x) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{v_i^d w_i} \mathbb{K}_d \left(\frac{Z_i - x}{v_i} \right) \mathbb{K}_1 \left(\frac{S_{i+1} - t}{w_i} \right)$$

$$\widehat{\mathcal{G}}^n(t|x) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{v_i^d} \mathbb{K}_d \left(\frac{Z_i - x}{v_i} \right) \mathbb{1}_{\{S_{i+1} > t\}}$$

$$\widehat{\mathcal{V}}^n(x) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{v_i^d} \mathbb{K}_d \left(\frac{Z_i - x}{v_i} \right)$$

Fenêtres : $v_k = v_0(k+1)^{-\alpha}$ et $w_k = w_0(k+1)^{-\beta}$ avec $\alpha, \beta > 0$

\mathbb{K}_p est un noyau sur \mathbb{R}^p , $p \in \{1, d\}$

Ensemble admissible pour les paramètres :

$$\{(\alpha, \beta) : \alpha > 0, \beta > 0, \alpha d + \beta < 1, \alpha d + \beta + 2 \min(\alpha, \beta) > 1\} \neq \emptyset$$

Résultat

Pour $x \in E$ et $0 \leq t < t^+(x)$, on a

$$\begin{bmatrix} \widehat{\mathcal{F}}^n(t|x) \\ \widehat{\mathcal{G}}^n(t|x) \\ \widehat{\nu}^n(x) \end{bmatrix} \xrightarrow{p.s} \begin{bmatrix} \nu(x)f(t|x) \\ \nu(x)G(t|x) \\ \nu(x) \end{bmatrix}$$

Et

$$n^{\frac{1-\alpha d-\beta}{2}} \left(\begin{bmatrix} \widehat{\mathcal{F}}^n(t|x) \\ \widehat{\mathcal{G}}^n(t|x) \\ \widehat{\nu}^n(x) \end{bmatrix} - \begin{bmatrix} \nu(x)f(t|x) \\ \nu(x)G(t|x) \\ \nu(x) \end{bmatrix} \right) \xrightarrow{d} \mathcal{N}(0_3, \Sigma(x, t, \alpha, \beta)),$$

où la matrice de covariance $\Sigma(x, t, \alpha, \beta)$ est dégénérée avec un seul terme nul en position (1, 1)

Démonstration

$$\begin{bmatrix} \widehat{\mathcal{F}}^n(t|x) \\ \widehat{\mathcal{G}}^n(t|x) \\ \widehat{\mathcal{V}}^n(x) \end{bmatrix} - \begin{bmatrix} \nu(x)f(t|x) \\ \nu(x)G(t|x) \\ \nu(x) \end{bmatrix} = \frac{\mathcal{M}_n}{n} + \mathcal{R}_n$$

- Vitesse du terme de reste : propriété de “Lipschitz mixing”
- Comportement asymptotique du crochet de (\mathcal{M}_n) :

$$\frac{\langle \mathcal{M} \rangle_n}{n^{1+\alpha d}} \sim \begin{bmatrix} n^\beta \Delta_1 & 0 & 0 \\ 0 & \Delta_2 & 0 \\ 0 & 0 & \Delta_3 \end{bmatrix}$$

Estimation de la densité conditionnelle f

$$\hat{f}^n(t|x) = \frac{\hat{\mathcal{F}}^n(t|x)}{\hat{\nu}^n(t|x)}$$

Résultat

Pour $x \in E$ et $0 < t < t^+(x)$, on a

$$\hat{f}^n(t|x) \xrightarrow{p.s.} f(t|x)$$

Et

$$n^{\frac{1-\alpha d-\beta}{2}} \left(\hat{f}^n(t|x) - f(t|x) \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{\tau_1^2 \tau_d^2 f(t|x)}{(1 + \alpha d + \beta)\nu(x)} \right)$$

Estimation de la fonction de survie conditionnelle G

$$\widehat{G}^n(t|x) = \frac{\widehat{\mathcal{G}}^n(t|x)}{\widehat{\nu}^n(x)}$$

Résultat

Pour $x \in E$ et $0 < t < t^+(x)$, on a

$$\widehat{G}^n(t|x) \xrightarrow{p.s.} G(t|x)$$

Et

$$n^{\frac{1-\alpha d-\beta}{2}} \left(\widehat{G}^n(t|x) - G(t|x) \right) \xrightarrow{\mathbb{P}} 0$$

Mieux, on a aussi

$$n^{\frac{1-\alpha d}{2}} \left(\widehat{G}^n(t|x) - G(t|x) \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{\tau_d^2 G(t|x)}{(1+\alpha d)\nu(x)} \right)$$

Estimation de la fonction composée $\lambda \circ \Phi$

$$\widehat{\lambda \circ \Phi}^n(t|x) = \frac{\widehat{\mathcal{F}}^n(t|x)}{\widehat{\mathcal{G}}^n(t|x)}$$

Résultat

Pour $x \in E$ et $0 < t < t^+(x)$, on a

$$\widehat{\lambda \circ \Phi}^n(t|x) \xrightarrow{p.s.} \lambda \circ \Phi(t|x)$$

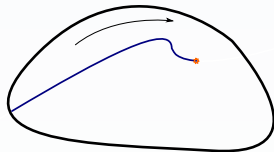
Et

$$n^{\frac{1-\alpha d-\beta}{2}} \left(\widehat{\lambda \circ \Phi}^n(t|x) - \lambda \circ \Phi(t|x) \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{\tau_1^2 \tau_d^2 \lambda \circ \Phi(t|x)}{(1+\alpha d+\beta)\nu(x)G(t|x)} \right)$$

Estimation du taux de saut λ

Estimation de $\lambda(x)$ pour $x \in E$ fixé

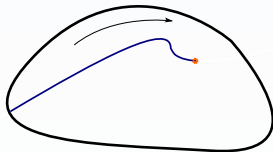
$$\mathcal{C}_x = \{\Phi(-t|x) : t \geq 0\} \cap E$$



Estimation du taux de saut λ

Estimation de $\lambda(x)$ pour $x \in E$ fixé

$$\mathcal{C}_x = \{\Phi(-t|x) : t \geq 0\} \cap E$$



Pour tout $\xi \in \mathcal{C}_x$, il existe un unique $t = \tau_x(\xi)$ tel que $\Phi(\tau_x(\xi)|\xi) = x$

En particulier, $\lambda \circ \Phi(\tau_x(\xi)|\xi) = \lambda(x)$.

Donc, pour tout $\xi \in \mathcal{C}_x$, $\widehat{\lambda}_\xi^n(x) = \widehat{\lambda} \circ \widehat{\Phi}^n(\tau_x(\xi)|\xi)$ est un “bon” estimateur de $\lambda(x)$ (convergence presque sûre et normalité asymptotique)

On propose de choisir l'estimateur de variance asymptotique minimale :

$$\frac{\tau_1^2 \tau_d^2 \lambda \circ \Phi(\tau_x(\xi)|\xi)}{(1 + \alpha d + \beta)\nu(\xi)G(t|\xi)} \propto (\nu(\xi)G(\tau_x(\xi)|\xi))^{-1} = \kappa_x(\xi)^{-1}$$

$$\widehat{\lambda}^n(x) = \widehat{\lambda}_{\xi^*}^n(x) \quad \text{où } \xi^* = \arg \max_{\xi \in \mathcal{C}_x} \kappa_x(\xi)$$

On propose de choisir l'estimateur de variance asymptotique minimale :

$$\frac{\tau_1^2 \tau_d^2 \lambda \circ \Phi(\tau_x(\xi)|\xi)}{(1 + \alpha d + \beta)\nu(\xi)G(t|\xi)} \propto (\nu(\xi)G(\tau_x(\xi)|\xi))^{-1} = \kappa_x(\xi)^{-1}$$

$$\widehat{\lambda}^n(x) = \widehat{\lambda}_{\xi^*}^n(x) \quad \text{où } \xi^* = \arg \max_{\xi \in \mathcal{C}_x} \kappa_x(\xi)$$

$$\widehat{\lambda}^n(x) = \widehat{\lambda}_{\xi^*}^n(x) \quad \text{où } \xi^* = \arg \max_{\xi \in \mathcal{C}_x} \widehat{\mathcal{G}}^n(\tau_x(\xi)|\xi)$$

→ Choix crucial des paramètres de lissage pour $\widehat{\mathcal{G}}^n$ le long de \mathcal{C}_x

Critère ISE le long de \mathcal{C}_x :

$$\begin{aligned} \text{ISE}_{\kappa}^n(\alpha) &= \int_{\mathcal{C}_x} \left(\widehat{\mathcal{G}}^n(\tau_x(\xi)|\xi) - \kappa_x(\xi) \right)^2 d\xi \quad (\text{intégrale curviligne}) \\ &= \int_{\mathcal{C}_x} \kappa_x(\xi)^2 d\xi + \varepsilon_{\kappa}^n(\alpha) \end{aligned}$$

où

$$\varepsilon_{\kappa}^n(\alpha) = \int_{\mathcal{C}_x} \widehat{\mathcal{G}}^n(\tau_x(\xi)|\xi)^2 d\xi - 2 \int_{\mathcal{C}_x} \widehat{\mathcal{G}}^n(\tau_x(\xi)|\xi) \kappa_x(\xi) d\xi$$

Mais $\kappa_x(\xi)$ est inconnu \rightarrow méthode de validation-croisée

Supposons $d = 1$ (E est un intervalle de \mathbb{R})

$$\frac{1}{\tilde{n}} \sum_{k=0}^{\tilde{n}-1} \hat{\mathcal{G}}^n(\tau_x(\tilde{Z}_k) | \tilde{Z}_k) \xrightarrow{p.s.} \int \hat{\mathcal{G}}^n(\tau_x(\xi) | \xi) \nu(\xi) \quad d\xi,$$

où $(\tilde{Z}_k, \tilde{S}_{k+1})$ est la chaîne immergée d'un autre PDMP, indépendant du premier, et généré selon les mêmes paramètres

Supposons $d = 1$ (E est un intervalle de \mathbb{R})

$$\frac{1}{\tilde{n}} \sum_{k=0}^{\tilde{n}-1} \widehat{\mathcal{G}}^n(\tau_x(\tilde{Z}_k) | \tilde{Z}_k) \mathbb{1}_{\mathcal{C}_x}(\tilde{Z}_k) \\ \xrightarrow{p.s.} \int_{\mathcal{C}_x} \widehat{\mathcal{G}}^n(\tau_x(\xi) | \xi) \nu(\xi) \quad d\xi,$$

où $(\tilde{Z}_k, \tilde{S}_{k+1})$ est la chaîne immergée d'un autre PDMP, indépendant du premier, et généré selon les mêmes paramètres

Supposons $d = 1$ (E est un intervalle de \mathbb{R})

$$\frac{1}{\tilde{n}} \sum_{k=0}^{\tilde{n}-1} \widehat{\mathcal{G}}^n(\tau_x(\tilde{Z}_k) | \tilde{Z}_k) \mathbb{1}_{\mathcal{C}_x}(\tilde{Z}_k) \mathbb{1}_{(\tau_x(\tilde{Z}_k), +\infty)}(\tilde{S}_{k+1})$$

$$\xrightarrow{p.s.} \int_{\mathcal{C}_x} \widehat{\mathcal{G}}^n(\tau_x(\xi) | \xi) \nu(\xi) G(\tau_x(\xi) | \xi) d\xi,$$

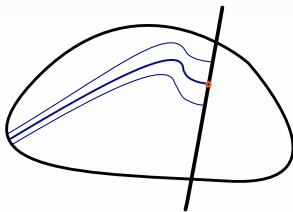
où $(\tilde{Z}_k, \tilde{S}_{k+1})$ est la chaîne immergée d'un autre PDMP, indépendant du premier, et généré selon les mêmes paramètres

Si $d > 1$, on utilise l'approximation suivante :

$$\frac{1}{\lambda_{d-1}(\mathbb{D}_{x,\rho})} \int_{\mathbb{T}_{x,\rho}} \varphi \longrightarrow \int_{\mathcal{C}_x} \varphi,$$

où

- $\mathbb{H}_x = \left\{ y \in \mathbb{R}^d : y - x \perp \frac{\partial \Phi}{\partial t}(x, 0) \right\}$
- $\mathbb{D}_{x,\rho} = B_d(x, \rho) \cap \mathbb{H}_x$
- $\mathbb{T}_{x,\rho} = \bigcup_{y \in \mathbb{D}_{x,\rho}} \mathcal{C}_y$



Approximation du critère ISE (à une constante additive près) :

$$\int_{\mathcal{C}_x} \widehat{\kappa}_x^n(\xi)^2 d\xi - \frac{2\Gamma\left(\frac{d-1}{2} + 1\right)}{\tilde{n} \pi^{\frac{d-1}{2}} \rho^{d-1}} \sum_{k=0}^{\tilde{n}-1} \widehat{\mathcal{G}}^n\left(\theta_x(\tilde{Z}_k) | \tilde{Z}_k\right) \mathbb{1}_{\mathbb{T}_{x,\rho}}(\tilde{Z}_k) \mathbb{1}_{(\theta_x(\tilde{Z}_k), +\infty)}(\tilde{S}_{k+1})$$

$$\xrightarrow{p.s.} \text{ISE}_{\kappa}^n(\alpha) \quad \text{quand } \tilde{n} \rightarrow +\infty, \rho \rightarrow 0$$

Approximation du critère ISE (à une constante additive près) :

$$\int_{\mathcal{C}_x} \widehat{\kappa}_x^n(\xi)^2 d\xi - \frac{2\Gamma\left(\frac{d-1}{2} + 1\right)}{\tilde{n} \pi^{\frac{d-1}{2}} \rho^{d-1}} \sum_{k=0}^{\tilde{n}-1} \widehat{\mathcal{G}}^n\left(\theta_x(\tilde{Z}_k) | \tilde{Z}_k\right) \mathbb{1}_{\mathbb{T}_{x,\rho}}(\tilde{Z}_k) \mathbb{1}_{(\theta_x(\tilde{Z}_k), +\infty)}(\tilde{S}_{k+1})$$

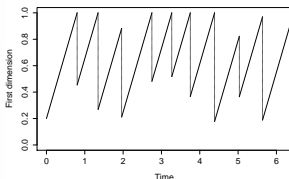
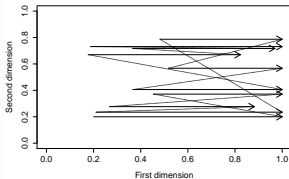
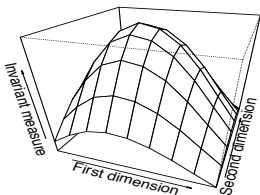
$$\xrightarrow{p.s.} \text{ISE}_{\kappa}^n(\alpha) \quad \text{quand } \tilde{n} \rightarrow +\infty, \rho \rightarrow 0$$

Rappelons que $\widehat{\lambda}_{\xi}^n(x) = \frac{\widehat{\mathcal{F}}^n(\tau_x(\xi) | \xi)}{\widehat{\mathcal{G}}^n(\tau_x(\xi) | \xi)}$

- α choisi pour $\widehat{\mathcal{G}}^n(\tau_x(\xi) | \xi)$ réutilisé pour calculer le dénominateur de $\widehat{\lambda}_{\xi}^n(x)$
- Validation-croisée pour choisir α et β dans $\widehat{\mathcal{F}}^n(\tau_x(\xi) | \xi)$

Une variante du modèle TCP

- $E = (0, 1)^2$
- $\Phi(t|(x_1, x_2)) = (x_1 + t, x_2)$
- $Q(\cdot|(x_1, x_2)) = \beta(2, 2/x_1) \otimes \beta(2, 2)$
- $\lambda((x_1, x_2)) = x_1 + x_2$

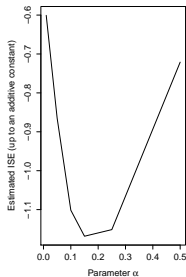


Simulations numériques

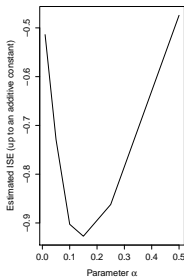
- **Objectif** : estimer λ en $x = (0.75, 0.5)$
- $\mathcal{C}_x = (0, 0.75] \times \{0.5\}$
- Les 10 000 premiers sauts de la MC (Z_n, S_{n+1}) sont observés
- Les 1 000 premiers sauts d'une autre MC $(\tilde{Z}_n, \tilde{S}_{n+1})$ sont observés pour la validation-croisée
- Boxplots calculés sur 100 répliques

1^{ère} étape : validation-croisée pour choisir α dans $\widehat{\mathcal{G}}^n(\tau_x(\xi)|\xi)$
(dénominateur)

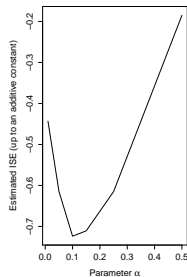
$\rho = 0.005$



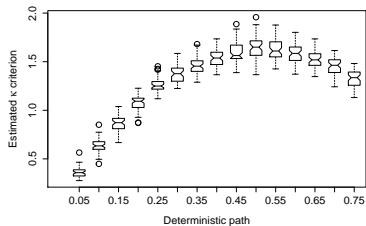
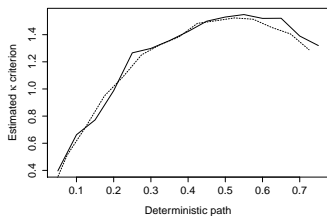
$\rho = 0.01$



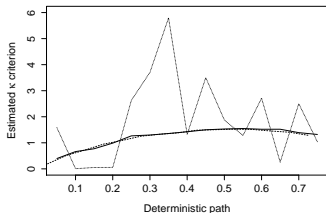
$\rho = 0.02$



2^{ème} étape : calculer le point $\xi^* \in \mathcal{C}_x$ qui maximise $\widehat{\mathcal{G}}^n(\tau_x(\xi)|\xi)$

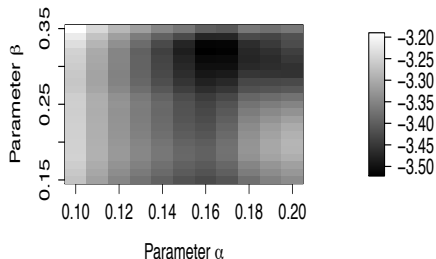
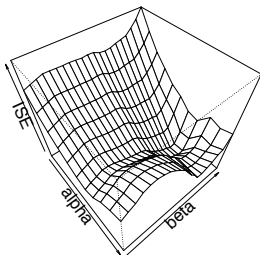


2^{ème} étape : calculer le point $\xi^* \in \mathcal{C}_x$ qui maximise $\widehat{\mathcal{G}}^n(\tau_x(\xi)|\xi)$

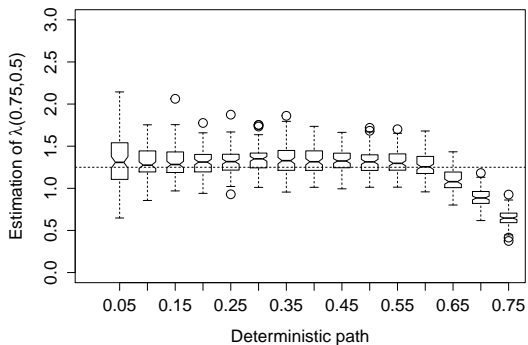


Mauvais choix de α ($\alpha = 0.04$)

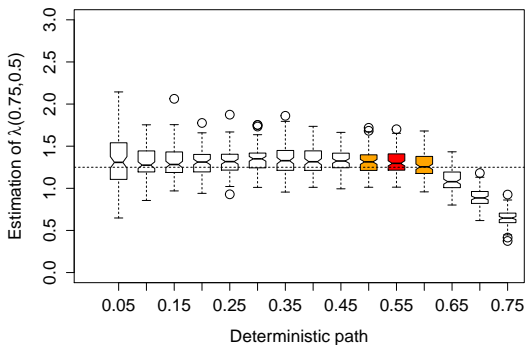
3^{ème} étape : validation-croisée pour choisir α et β dans $\hat{\mathcal{F}}^n(\tau_x(\xi)|\xi)$
(numérateur)



4^{ème} étape : calculer $\widehat{\lambda}_{\xi^*}^n(x)$ avec les paramètres optimaux α et β



4^{ème} étape : calculer $\widehat{\lambda}_{\xi^*}^n(x)$ avec les paramètres optimaux α et β



Merci de votre attention