

Modèle dynamique d'exposition à un risque alimentaire

Application au cas du méthyle mercure

En collaboration avec Patrice Bertail (Modal'X - Univ. Paris X),
Stéphan Cléménçon (TSI - Telecom Paris) et Philippe Verger (INRA - Mét@risk)

Jessica Tressou

INRA - Mét@risk - Chargée de Recherche
(Méthodologies d'analyse de risque alimentaire, Paris)

HKUST - ISMT - Visiting Scholar
(Hong Kong University of Science and Technology, Information & Systems Management)

GT "*Modélisation du Vivant*" CMAP - Ecole Polytechnique
Jeudi 17 janvier 2008

- 1 AQR : Analyse Quantitative de Risque (Alimentaire)
- 2 KDEM : Kinetic Dietary Exposure Model
- 3 Valeurs Extrêmes de KDEM

- 1 AQR : Analyse Quantitative de Risque (Alimentaire)
 - Principe et exemples
 - Cadre statique - Approche usuelle
 - Limites et solution proposée
- 2 KDEM : Kinetic Dietary Exposure Model
- 3 Valeurs Extrêmes de KDEM

① Appréciation Quantitative de Risque

- Identification et Caractérisation du Danger
- **Evaluation de l'exposition**
- **Caractérisation du risque**

② Différents Types de Risques

- **Chimique** / Microbiologique / Nutritionnel (+ Bénéfice)
- Aigu / **Chronique**

③ Exemple : **Méthyle Mercure** (MeHg)

- Présent dans les produits de la mer (poissons, crustacés, etc.)
- Danger : altérations du système nerveux, pb. neurologiques
- Population sensible : les foetus, via l'exposition de leur mère

Estimation de l'exposition au risque

- Exposition très rarement mesurée directement
- Évaluée à partir de données
 - 1 de consommation $Q = (Q_1, \dots, Q_P)$ (corrélées)
 - 2 et de contamination $K = (K_1, \dots, K_P)$
(indép. mais censurées à gauche)
- Exposition/Apport/Intake : $U = \sum_{p=1}^P K_p Q_p$
(Unité : $\mu\text{g}/\text{kgpc}/\text{s}$)

Caractérisation du risque

- Comparaison Dose "usuelle" / "DHT" Dose Hebdom. Tolérable
- Travaux de thèse :
 - Estimateur Plug-in de $\mathbb{P}(U > DHT)$: une U -statistique généralisée
 - Estimateur Valeurs-Extrêmes par ajustement de type Pareto

① Limites de l'approche statique

- Accumulation dans l'organisme ?
- Elimination/Absorption
- Pharmacocinétique lente : le méthyle mercure

② Intégration du temps

→ Approche dynamique : un modèle simple

- Accumulation par voie alimentaire

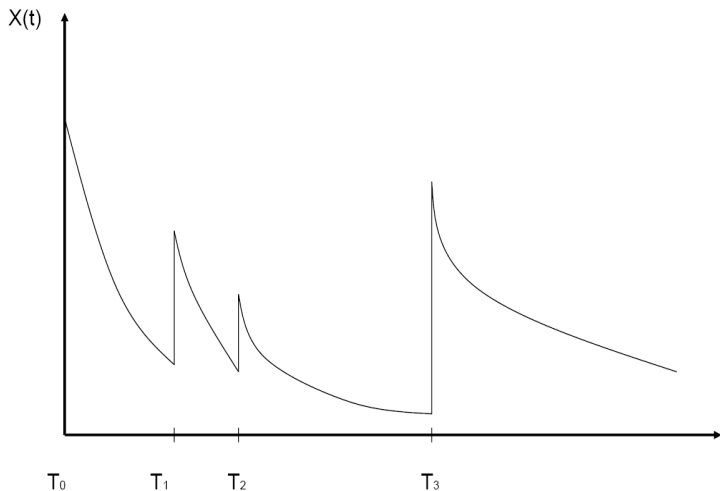
Processus Markovien à sauts

- Elimination progressive (selon une EDO)

"Pharmacocinétique linéaire"

- Déf. d'une dose tolérable de long terme

Kinetic Tolerable Intake (KTI) : $DHT / (1 - 2^{-1/DV})$



Kinetic Dietary Exposure Model (KDEM)

Un processus Markovien, déterministe par morceaux

- 1 AQR : Analyse Quantitative de Risque (Alimentaire)
- 2 KDEM : Kinetic Dietary Exposure Model
 - Définition du modèle
 - Stabilité stochastique
 - Estimation par simulation
- 3 Valeurs Extrêmes de KDEM

NOTATION

① **Pharmacocinétique** : $\dot{x}(t) = -r(x(t), \theta)$

Exemple : $r(x, \theta) = \theta \times x$,

avec $\theta = \ln(2)/DV$ (DV =Demie-Vie biologique)

NOTATION

- 1 **Pharmacocinétique** : $\dot{x}(t) = -r(x(t), \theta)$
Exemple : $r(x, \theta) = \theta \times x$,
avec $\theta = \ln(2)/DV$ (DV =Demie-Vie biologique)
- 2 **Historique des apports** : un proc. ponctuel marqué
(T_n, Q_n, K_n)
 - T_1, \dots, T_n : dates des prises alimentaires
 - $U_1 = \langle K_1, Q_1 \rangle, \dots$: apports aux dates T_1, \dots ,
 - Indépendance entre les $(\Delta T_n)_n$ et les $(U_n)_n$

NOTATION

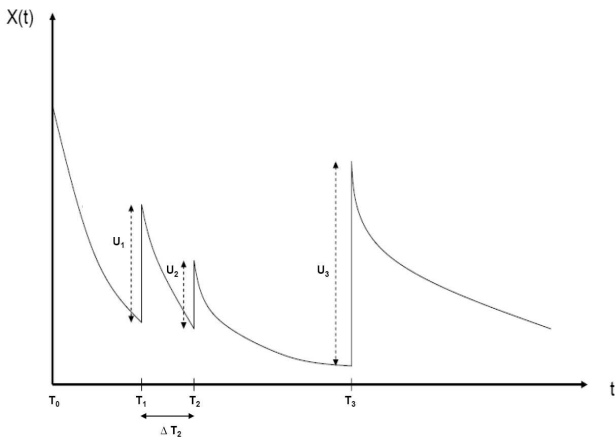
① **Pharmacocinétique** : $\dot{x}(t) = -r(x(t), \theta)$ Exemple : $r(x, \theta) = \theta \times x$,avec $\theta = \ln(2)/DV$ (DV =Demie-Vie biologique)② **Historique des apports** : un proc. ponctuel marqué
(T_n, Q_n, K_n)

- T_1, \dots, T_n : dates des prises alimentaires
- $U_1 = \langle K_1, Q_1 \rangle, \dots$: apports aux dates T_1, \dots ,
- Indépendance entre les $(\Delta T_n)_n$ et les $(U_n)_n$

Le processus d'exposition $X(t)$

$$X(t) = X(0) + \sum_{n=1}^{N(t)} U_n - \sum_{n=1}^{N(t)+1} \int_{T_{n-1}}^{T_n \wedge t} r(X(s), \theta_n) ds,$$

où θ_n est le taux d'élimination entre les dates T_n et T_{n+1} .



Kinetic Dietary Exposure Model (KDEM)

Le modèle KDEM est défini par :

- 1 la distrib. F_U des apports $(U_n)_{n \geq 1}$ (iid)
- 2 la distrib. G des durées entre apports $(\Delta T_{n+1})_{n \geq 1}$ (iid)
(voire celle du taux d'élimination notée H , pour les $(\theta_n)_{n \geq 1}$)
- 3 la vitesse d'élimination $r(\cdot, \theta)$ supposée linéaire

Le modèle KDEM est défini par :

- 1 la distrib. F_U des apports $(U_n)_{n \geq 1}$ (iid)
- 2 la distrib. G des durées entre apports $(\Delta T_{n+1})_{n \geq 1}$ (iid)
(voire celle du taux d'élimination notée H , pour les $(\theta_n)_{n \geq 1}$)
- 3 la vitesse d'élimination $r(\cdot, \theta)$ supposée linéaire

On peut l'étudier

- en temps discret : $\tilde{X} = (X(T_n))_n$
Proc. Autorég. à coef. aléat. : $X_{n+1} = e^{-\theta \Delta T_{n+1}} X_n + U_{n+1}$
- en temps continu : $X = (X(t))_{t \geq 0}$
Proc. Markovien bivarié : $(X(t), A(t))_{t \geq 0}$, avec
 $X(t) = e^{-\theta A(t)} X_{N(t)}$

Sous des hypothèses de régularité de F_U et G ,

Stabilité de la chaîne discrète

- Densité explicite de la transition de la chaîne discrète

$$\pi(x, y) = \int_{t=\frac{1}{\theta} \log(1 \vee \frac{x}{y})}^{\infty} f_U(y - xe^{-\theta t}) G(dt)$$

- Existence d'une unique loi stationnaire $\tilde{\mu}$
- Convergence vers ces lois à vitesse géométrique.

$$\sup_{\{\psi, |\psi(z)| \leq 1 + z^\gamma\}} \left| \int_{y=0}^{\infty} \psi(y) \Pi^n(x, dy) - \tilde{\mu}(\psi) \right| \leq R(1 + x^\gamma) r^{-n}$$

Preuve de ce type de résultat :
Conditions de drift (Foster-Lyapounov) et de minorisation,
cf. MEYN & TWEEDIE (1996).

- Minorisation sur $S = [0, s]$:

$$\Pi(x, \cdot) \geq \delta_s \cdot \mathbb{U}_s(\cdot),$$

avec $\delta_s = s \times \inf_{y \in [0, s]} f_U(y) > 0$.

- Drift : $\Delta V(x) = \mathbb{E}[V(X_1) | X_0 = x] - V(x)$

$$\Delta V(x) \leq -f(x) + b\mathbb{1}_{\{x \in S\}}$$

Ex : $V(x) = 1 + x$, $f(x) = cV(x)/2$, $b = \mathbb{E}[U_1] + c/2$ et

$$c = 1 - \mathbb{E}[e^{-\theta \Delta T_2}] > 0$$

Sous des hypothèses de régularité de F_U et G ,

Stabilité de la chaîne continue

- Générateur explicite de la chaîne bivariée

$$\mathcal{G}\phi(x, t) = \frac{g(t)}{\bar{G}(t)} \int_{u=0}^{\infty} \{\phi(x+u, 0) - \phi(x, t)\} F_U(du) \\ - r(x, \theta) \partial_x \phi(x, t) + \partial_t \phi(x, t)$$

- Existence de la loi stationnaire μ pour X
- Convergence vers cette lois à vitesse géométrique.

$$\sup_{\psi(z) \leq 1+z^\gamma} |\mathbb{E}_{x,a}[\psi(X_t)] - \mu(\psi)| \leq B_a(1+x^\gamma)\beta^t.$$

- Preuve : même principe.

Les lois stationnaires $\tilde{\mu}$ et μ héritent de propriétés de F_U

- Existence de moments

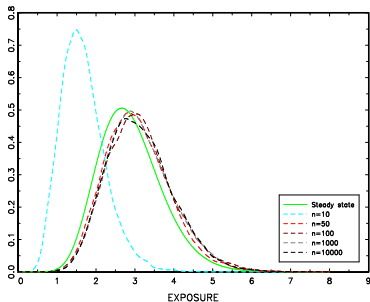
$$\mathbb{E}[U_1^\gamma] < \infty \Rightarrow \int x^\gamma \tilde{\mu}(dx) < \infty \text{ et } \int x^\gamma \mu(dx) < \infty$$

- Comportement de queue (domaine d'attraction du maximum)

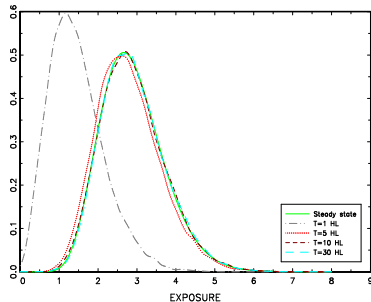
$$\bar{F}_U(x) \propto x^{-\kappa} \Rightarrow \tilde{\mu}(]x, \infty[) \propto x^{-\kappa} \text{ et } \mu(]x, \infty[) \propto x^{-\kappa}$$

Temps d'atteinte de l'état stationnaire ?

- Modèle : $\text{Exp}(m_{F_U})$ - $\text{Exp}(m_G)$
- Limite : $\mu = \Gamma(DV/(m_G \ln 2), m_{F_U})$.



Convergence vers $\tilde{\mu}$



Convergence vers μ

En théorie, des bornes pour la vitesse... cf. ROBERTS & TWEEDIE (1999)

Quantités d'intérêt pour les toxicologues ?

$$M_T = \sup_{0 \leq t \leq T} X(t) \quad m_T = T^{-1} \int_{t=0}^T X(t) dt$$

$$p_{u,T} = T^{-1} \int_{t=0}^T \mathbb{1}_{\{X(t) \geq u\}} dt \quad \tau_u^+ = \inf\{t \geq 0, X(t) \geq u\}$$

PB1 :

- On n'observe pas les trajectoires $(X(t))_t$
- On observe seulement des $(U_n)_n$ et des (ΔT_{n+1})

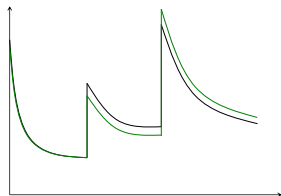
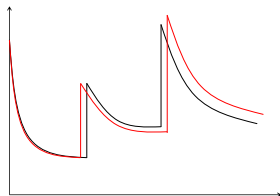
PB2 :

Qtés d'intérêt = fonctionnelles complexes de F_U et G
 \Rightarrow Pas d'estimateur Plug-In

IDÉE :

- Estimation de F_U et G par $\hat{F}_{U,n}$ et \hat{G}_n , et de θ par $\hat{\theta}_n$.
- Construction du proc. associé : $\hat{X}^{(n)}$
- Estimation par simulation de MC de $M_T, m_T, p_{u,T}, \tau_u^+ \dots$

Convergence de ces estimateurs ?



Topologie dans l'espace des proc. d'exposition :

La topologie induite par la distance de Hausdorff entre les graphes complétés

Résumé des principaux résultats, **obtenu par couplage**
Sous hyp. de consistance (L_1 -Mallows) des estim. $\hat{F}_{U,n}$ et \hat{G}_n
et de consistance forte de $\hat{\theta}_n$

Horizon fini : $T < \infty$

- Convergence en **distrib.** des fonctionnelles **continues**
Ex : Moyenne du proc. d'exposition, tps au dessus d'un seuil u
- Convergence en **moyenne** pour les fonctionnelles **Lipschitz**
Ex : Maximum d'exposition
- Par intégrabilité uniforme, convergence en moyenne pour la moyenne et le temps moyen au dessus de u

Résumé des principaux résultats, **obtenu par couplage**
Sous hyp. de consistance (L_1 -Mallows) des estim. $\hat{F}_{U,n}$ et \hat{G}_n
et de consistance forte de $\hat{\theta}_n$

$$(M_1(F_1, F_2) = \int |F_1(x) - F_2(x)| dx)$$

Etat stationnaire : $T \rightarrow \infty$

Des conditions supplémentaires :

- $T^2 \times (M_1(\hat{G}^{(n)}, G) + |\hat{\theta}_n - \theta|) \rightarrow 0$ p.s.
- $T \times M_1(\hat{F}_U^{(n)}, F_U) \rightarrow 0$ p.s.

\Rightarrow Convergence en **distrib.** et en **moyenne** des estimateurs MC :

$$\hat{m}_T \rightarrow m_\mu \quad \hat{p}_{u,T} \rightarrow \mu(]u, +\infty[)$$

Lois asymptotiques non calculables explicitement
⇒ Intervalles de confiance Bootstrap

① Algorithme

- Rééchantillonnage des observations des $(U_n)_n$ et des $(\Delta T)_n$
- Estimation de F_U et G à partir de ces obs.
- Simulation de Monte Carlo ⇒ \hat{X}^*
- Calcul des stat. d'intérêt T_b pour chaque rééchantillonnage $b = 1, \dots, B$
- IC : à partir des percentiles de $\{T_b, b = 1, \dots, B\}$.

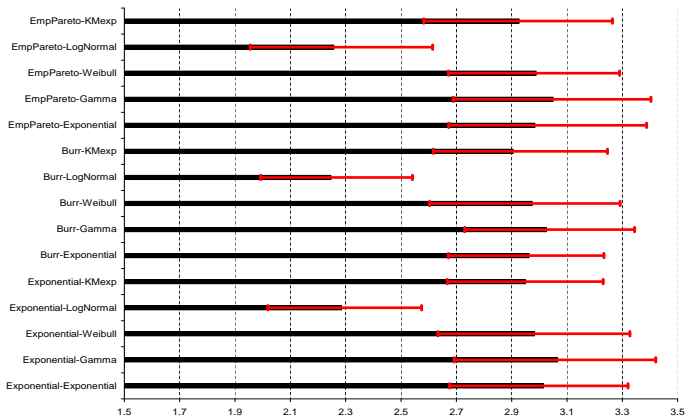
② Validité asymptotique

APPLICATION MEHG : le cas des femmes en âge de procréer.

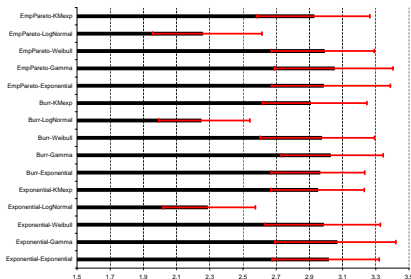
- Données INCA : consommations individuelles détaillées
- Population : Femmes de 15 à 45 ans ($n = 638$)
- Aliments : Poissons et produits de la mer
- Contamination des aliments : moyenne
- Calcul des apports par prise alimentaire... les U_n ($n_{F_U} = 1088$)
- Temps entre intakes... censurés ($n_G = 1214$)
- Choix de distributions
 - F_U : Exponential, Burr, EmpPareto
 - G : Exponential, Gamma, Weibull, LogNormal, KM-Exp

Moyenne à l'état stat. pour les femmes (15-45).

(Unité : $\mu\text{g}/\text{kgpc}$)

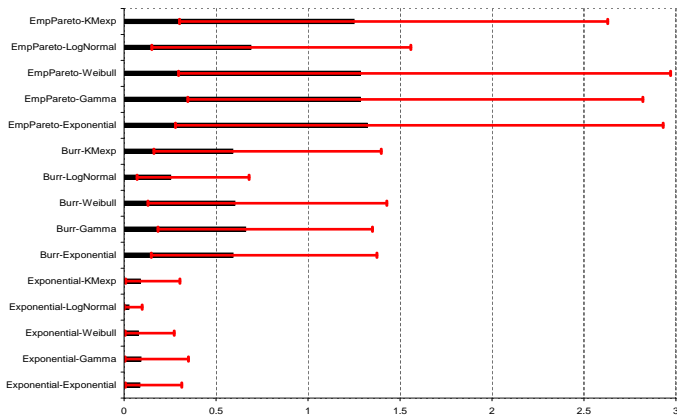


Moyenne à l'état stat. pour les femmes (15-45).

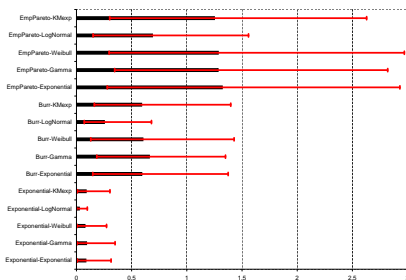


Pour une DV de 6 semaines,
A comparer à $KTI = 14.67 \mu\text{g}/\text{kgpc}$ pour $DHT = 1.6 \mu\text{g}/\text{kgpc}/\text{s}$
ou à $KTI = 6.41 \mu\text{g}/\text{kgpc}$ pour $DHT = 0.7 \mu\text{g}/\text{kgpc}/\text{s}$

Probab. stat. de dépasser $KTI = 6.41 \mu\text{g}/\text{kgpc}$.
(Femmes, 15-45)

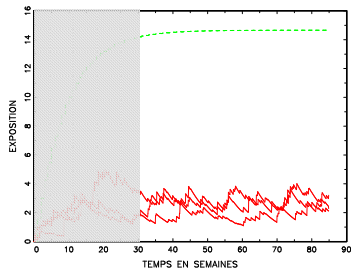
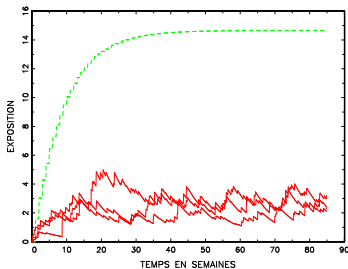


Probab. stat. de dépasser $KTI = 6.41 \mu\text{g}/\text{kgpc}$.



Simulation MC naïve non pertinente pour $KTI = 14.67 \mu\text{g}/\text{kgpc}$

Probab. stat. de dépasser $KTI = 14.67 \mu\text{g}/\text{kgpc}$.
 (Femmes, 15-45)



- 1 AQR : Analyse Quantitative de Risque (Alimentaire)
- 2 KDEM : Kinetic Dietary Exposure Model
- 3 Valeurs Extrêmes de KDEM
 - Echantillonnage d'importance
 - Filtrage Particulaire
 - Approche Régénérative

$$\tau_u^+(X) \leq T \Leftrightarrow \max_{t \in [0, T]} X_t > u \Leftrightarrow \exists t \leq T, X_t > u$$

Principe :

- Dét. des distrib. \tilde{F}_U et \tilde{G} qui augmentent l'occurrence de

$$A_{u, T} = \{\tau_u^+(X) \leq T\}$$

- Pondérer les trajectoires \tilde{X} par le rapport de vraisemblance

$$L_T(X) = \frac{d\mathbb{P}_{x_0}}{d\tilde{\mathbb{P}}_{x_0}} = \frac{1 - G(T - T_{N(T)})}{1 - \tilde{G}(T - T_{N(T)})} \times \prod_{k=1}^{N(T)} \frac{f_U(U_k)}{\tilde{f}_U(U_k)} \cdot \frac{g(\Delta T_k)}{\tilde{g}(\Delta T_k)}.$$

- Estimateur MC de $\mathbb{P}_{x_0}(A_{u, T})$:
 - générer M trajectoires sous $\tilde{\mathbb{P}}_{x_0}$
 - et calculer $\frac{1}{M} \sum_{m=1}^M L_T(X^{(m)}) \mathbf{1}_{\{\tau_u^+(X^{(m)}) \leq T\}}$

Probab. stat. de dépasser $KTI = 14.67 \mu\text{g}/\text{kgpc}$.
(Femmes, 15-45)

- Modèle Burr-Gamma :

$$\mathbb{P}_{x_0} = \text{Burr}(0.9456 ; 4.935) \otimes \Gamma(1.066 ; 117.2)$$

- $M = 1000$, $T = 20$ ans, $u = 14.67$, $x_0 = 3$:
 - Recherche très difficile de la "bonne" distribution
 - Exemple : pour $\tilde{\mathbb{P}}_{x_0} = \text{Burr}(0.9 ; 4.5) \otimes \Gamma(1.01 ; 115)$

$$\mathbb{P}_{x_0}(A_{u,T}) \approx 10^{-5}$$

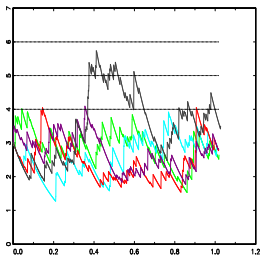
Estimation de $\mathbb{P}_{x_0}(\tau_u^+ \leq T)$ pour u grand et $T < \infty$.

Cf. Cérou, Del Moral, LeGland & Lezaud (2006)

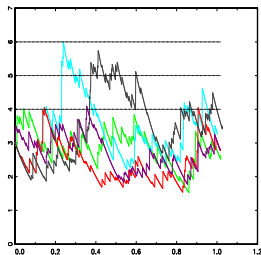
- 1 m niveaux intermédiaires $u_1 \leq \dots \leq u_m \leq u_{m+1} = u$.
- 2 N particules à générer de $X(0) = x_0 < u_1$ à T .
- 3 Pour chaque niveau u_j ($j = 1, \dots, m$) :
 $I_{1,j} = \{\text{particules ayant atteint } u_j \text{ avant } T\}$
 $I_{0,j} = \{\text{particules n'ayant pas atteint } u_j \text{ avant } T\}$
 - Pour les particules de $I_{0,j}$,
Sélection : Choisir une particule de $I_{1,j}$
Mutation : La dédoubler à partir de $\tau_{u_j}^+$ jusqu'à T .
 - Calculer $P_j = |I_{1,j}|/N$ et passer au niveau suivant u_{j+1} .

Estimation de $\mathbb{P}_{x_0}(\tau_u^+ \leq T) : P_1 \times \dots \times P_{m+1}$

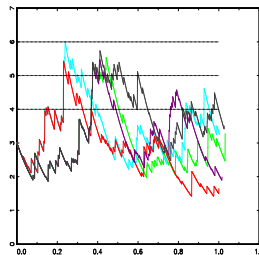
Illustration



Init. : $N = 5$



Iteration 1 : $u = 4$



Iteration 2 : $u = 5$

Modèle Exponentiel-Exponentiel, $N = 5$ particules, $x_0 = 3$, $m = 2$ niveaux intermédiaires ($u = (4, 5, 6)$), horizon $T = 1$ an.

Application : $u = KTI = 14.67 \mu\text{g}/\text{kgpc}$, $T = 20$ ans

Niveaux intermédiaires déterminés par l'algorithme adaptatif de Cérou & Guyader (2005), $u = 7.82, 8.68, 9.57, 10.35, 11.09, 12.18, 12.85, 14.12$

$$\mathbb{P}_{2.99}(\tau_{KTI}^+ \leq 20) \approx 0.918\%.$$

Valeurs Extrêmes : une approche régénérative

Principe :

Les propriétés de stabilité stochastique de KDEM impliquent que

- $S = [0, s]$ est un *petit ensemble*

$$\exists \phi, \delta > 0, \forall x \in S, \pi(x, y) \geq \delta \phi(y), \lambda(dy)\text{-p.s..}$$

- Sachant $X_i \in S$, la chaîne régénère en X_{i+1} avec prob. $\frac{\delta \phi x_i}{\pi X_i, X}$
- En estimant π par π_n , on construit des blocs pseudo-régénératifs
- Par Markov FORT et des arguments de couplage, ils sont i.i.d.
- Application de techniques de valeurs extrêmes pour v.a. i.i.d. à ces blocs (en cours)

- Bonnes propriétés stochastiques du modèle KDEM
- Des solutions pour l'estimation des quantités d'intérêt pour les toxicologues
- Cas d'événements rares : des solutions difficiles à mettre en oeuvre

encore du travail...

- Bonnes propriétés stochastiques du modèle KDEM
- Des solutions pour l'estimation des quantités d'intérêt pour les toxicologues
- Cas d'événements rares : des solutions difficiles à mettre en oeuvre

encore du travail...





Extensions envisagées...

- Bonnes propriétés stochastiques du modèle KDEM
- Des solutions pour l'estimation des quantités d'intérêt pour les toxicologues
- Cas d'événements rares : des solutions difficiles à mettre en oeuvre

encore du travail...

Extensions envisagées...

- 1 Population hétérogène, issue d'un mélange ?
- 2 Introduction de dépendance dans le proc. d'apports alim.
Plusieurs aliments complémentaires, substituts
Utilisation des données d'achats des ménages
- 3 Pharmaco-cinétique : autre modèle plus proche des données (à trouver !)
- 4 Epidémiologie : modéliser la relation entre l'exposition et le développement d'un effet

-  Asmussen, S. (2003).
Applied Probability and Queues.
Springer-Verlag, New York.
-  Cérou, F., Moral, P. D., LeGland, F. and Lezaud, P. (2006).
Genetic genealogical models in rare event analysis.
Alea 1, 183–196.
-  Meyn, S. and Tweedie, R. (1996).
Markov Chains and Stochastic Stability.
Springer-Verlag.
-  Whitt, W. (2002).
Stochastic-Process Limits. An Introduction to Stochastic-Process Limits and their Application to Queues.
Springer.

-  Bertail, P., Feinberg, M., Tressou, J. and Verger, P. (2006).
Analyse des Risques Alimentaires.
TEC&DOC, Lavoisier, Paris.
-  Bertail, P. and Tressou, J. (2006).
Incomplete generalized U-Statistics for food risk assessment.
Biometrics **62**, 66–74.
-  Tressou, J. (2005).
Méthodes statistiques pour l'évaluation du risque alimentaire.
PhD thesis, Université Paris X.
-  Tressou, J. (2006).
Non parametric modelling of the left censorship of analytical data in
food risk exposure assessment.
J. Amer. Stat. Assoc. **101**, 1377–1386.
-  Tressou, J., Crépet, A., Bertail, P., Feinberg, M. H. and Leblanc, J. C.
(2004).
Probabilistic exposure assessment to food chemicals based on extreme
value theory. application to heavy metals from fish and sea products.
Food Chem. Toxicol. **42**, 1349–1358.



Bertail, P., Cl emen on, S. and Tressou, J. (2006).

A storage model with random release rate for modeling exposure to food contaminants.

A para tre dans *Math. Biosc. Eng.*

<http://hal.archives-ouvertes.fr/hal-00138279>.



Bertail, P., Cl emen on, S. and Tressou, J. (2007).

Extreme values statistics for Markov chains via the (pseudo-)regenerative method.

Soumis pour publication,

<http://hal.archives-ouvertes.fr/hal-00165652>.



Bertail, P., Cl emen on, S. and Tressou, J. (2008).

Statistical analysis of a dynamic model for food contaminant exposure with applications to dietary methylmercury contamination.

En pr eparation.



Verger, P., Tressou, J. and Cl emen on, S. (2007).

Integration of time as a description parameter in risk characterisation : application to methylmercury.

Regul. Tox. Pharm. **49**, 25–30.