Simulation estimators in a dynamic model for food contaminant exposure

Joint work with P. Bertail (CREST LS and Paris X)

and Stéphan Clémençon (INRA Mét@risk and Paris X)

Jessica Tressou

INRA-Mét@risk, Food Risk Analysis Methodologies, Paris
Hong Kong University of Science and Technology, Information and Systems ManagemenT

Workshop: New Directions in Monte Carlo Simulations, Fleurance, June 25-29, 2007

Outline of the talk

1. **Food risk assessment**
2. **Kinetic Dietary Exposure Model (KDEM)**
3. **Stochastic stabilility of the process**
   - Discrete time
   - Continuous time
4. **Simulation based estimators**
   - Convergence of the Monte Carlo estimators
   - Bootstrap confidence intervals
5. **Rare Event Analysis: a particle filtering algorithm**
6. Conclusions, Perspectives, Other interests in Monte Carlo methods

Basic facts about Food Risk Analysis

1. **Food Risk Assessment**
   - Hazard identification and characterization
   - Exposure Assessment
   - Risk Characterization
2. **Different kind of risks**
   - Chemical / Microbiological / Nutritional (+ Benefit)
   - Acute / Chronic
3. **Studied example: methylmercury (MeHg)**
   - Present in fish and seafood
   - Hazard: brain damage in foetus, young children

1. **Exposure/intake assessment**
   - Not observed directly
   - Combination of consumption Q and contamination K data
   - Correlation in the Q_p's, left censorship in K_p
   - Parametric / non parametric models:

$$U = \sum_{p=1}^{P} K_p Q_p$$

   - Unit: μg/kgbw/w

2. **Risk characterization**
   - Comparison of the usual intake with the Provisional Tolerable Weekly Intake "$PTWI$"
   - Estimation of $\mathbb{P}(U > PTWI)$?
     - Plug-in estimator
     - Extreme Value-based estimator

Accumulation, Elimination?

1. **Exposure/intake assessment**
   - Not observed directly
   - Combination of consumption Q and contamination K data
   - Correlation in the Q_p's, left censorship in K_p
   - Parametric / non parametric models:

$$U = \sum_{p=1}^{P} K_p Q_p$$

   - Unit: μg/kgbw/w

2. **Risk characterization**
   - Comparison of the usual intake with the Provisional Tolerable Weekly Intake "$PTWI$"
   - Estimation of $\mathbb{P}(U > PTWI)$?
     - Plug-in estimator
     - Extreme Value-based estimator

Accumulation, Elimination?

1. **Pharmacokinetics:** elimination in between intakes at rate

$$\dot{x}(t) = -r(x(t), \theta)$$

Example: $r(x, \theta)$, $\theta = \ln(2)/HL$ (Biological half-life)

2. **Intake history:** a marked point process (T_n, Q_n, K_n)
   - $T_1, \ldots, T_n$ are the intake times,
   - $U_1 =< K_1, Q_1 >, \ldots$ are the intakes at times $T_1, \ldots,$
   - Independence between the $(\Delta T_n)_n$ and the $(U_n)_n$

The exposure process $X(t)$

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

where θ_n is the elimination rate between time T_n and T_{n+1}.

1. **Pharmacokinetics:** elimination in between intakes at rate

$$\dot{x}(t) = -r(x(t), \theta)$$

Example: $r(x, \theta)$, $\theta = \ln(2)/HL$ (Biological half-life)

2. **Intake history:** a marked point process (T_n, Q_n, K_n)
   - $T_1, \ldots, T_n$ are the intake times,
   - $U_1 =< K_1, Q_1 >, \ldots$ are the intakes at times $T_1, \ldots,$
   - Independence between the $(\Delta T_n)_n$ and the $(U_n)_n$

The exposure process $X(t)$

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

where θ_n is the elimination rate between time T_n and T_{n+1}.

1. **Pharmacokinetics:** elimination in between intakes at rate

$$\dot{x}(t) = -r(x(t), \theta)$$

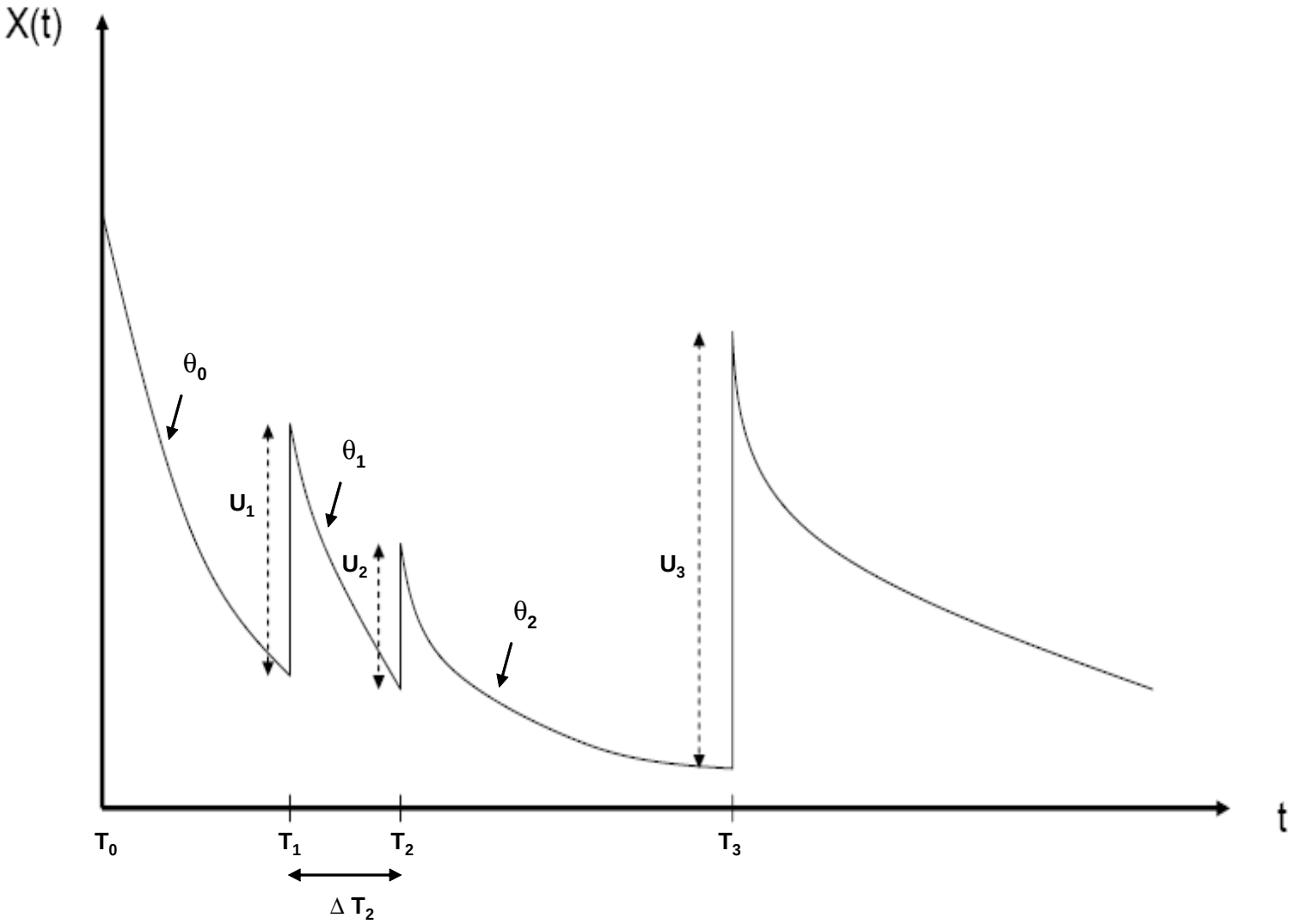
Example: $r(x, \theta)$, $\theta = \ln(2)/HL$ (Biological half-life)

2. **Intake history:** a marked point process (T_n, Q_n, K_n)
   - $T_1, \ldots, T_n$ are the intake times,
   - $U_1 =< K_1, Q_1 >, \ldots$ are the intakes at times $T_1, \ldots,$
   - Independence between the $(\Delta T_n)_n$ and the $(U_n)_n$

The exposure process $X(t)$

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

where θ_n is the elimination rate between time T_n and T_{n+1}.



Kinetic Dietary Exposure Model (KDEM)

A piecewise deterministic Markov process

Kinetic Dietary Exposure Model (KDEM)

A KDEM is defined by

1. **the set of independent distributions** (F_U, G, H)

$$(U_n)_{n\geq 1} \sim_{\text{iid}} F_U(du), \quad (\Delta T_{n+1})_{n\geq 1} \sim_{\text{iid}} G(ds), \quad (\theta_n)_{n\geq 1} \sim_{\text{iid}} H(d\theta)$$

2. **the release rate** $r(., \theta)$

General release rate / linear release rate $r(x, \theta) = \theta \times x$

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

Discrete time: $\tilde{X} = (X(T_n))_n$

Autoregressive with random coef.

$$X_{n+1} = e^{-\theta_n \Delta T_{n+1}} X_n + U_{n+1}$$

Continuous time: $X = (X(t))_{t\geq 0}$

A trivariate Markov process

$$(X(t), \theta(t), A(t))_{t\geq 0}$$

$$X(t) = e^{-\theta(t) A(t)} X_{N(t)}$$

Stochastic stability of the KDEM

Under regularity assumptions on F_U, G and H,

Positive recurrence and geometric ergodicity

- Explicit transition density of the discrete chain

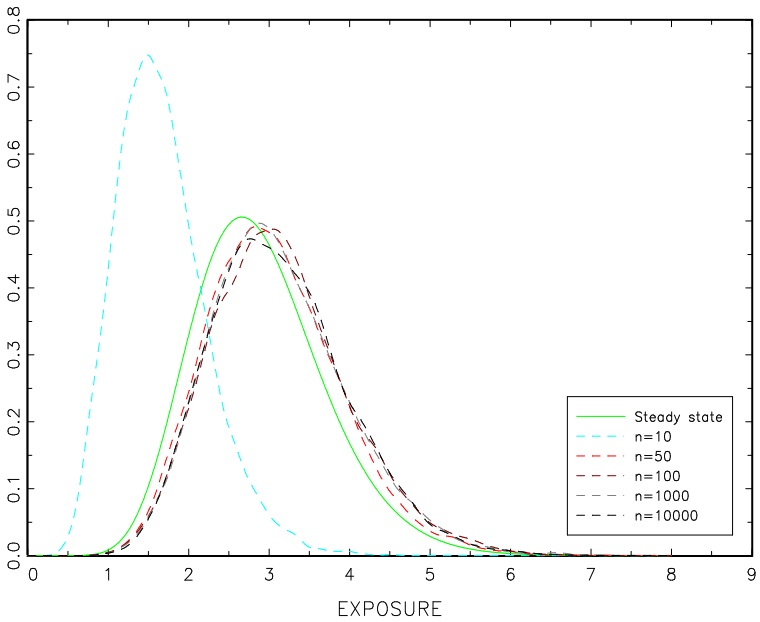
$$\pi(x,y) = \int_{\theta\in\Theta} \int_{t=\frac{1}{\theta}log(1\vee\frac{x}{y})}^{\infty} f_U(y - xe^{-\theta t})G(dt)H(d\theta)$$

- Similarly, explicit generator of the trivariate Markov process
- Existence of the stationary distributions $\tilde{\mu}$ and μ
- Convergence at geometric rate in both cases.
- Proof based on minorization and drift conditions.

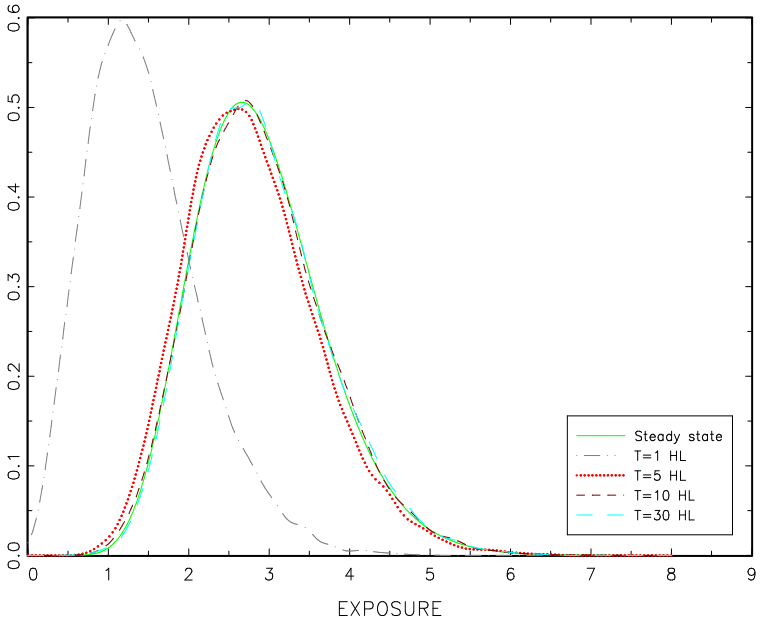
The stationary distributions $\tilde{\mu}$ and μ inherit the moment and heavy tail properties of the intake distribution F_U

Time to steady state
Illustration

Model: Exp(m_{F_U})-Exp(m_G)-Dirac, μ is a $\Gamma(HL/(m_G \ln 2), m_{F_U})$.



Convergence to $\tilde{\mu}$



Convergence to μ

Computable bounds for the time to steady state... theoretically!

Simulation-based estimators

1. Toxicologists'interest: distrib. or expectation of

   Maximum exposure value $\longrightarrow \sup_{0\leq t\leq T} X(t)$

   Mean exposure value $\longrightarrow T^{-1}\int_{t=0}^{T} X(t)dt$

   Average time spent over u $\longrightarrow T^{-1}\int_{t=0}^{T} \mathbb{I}_{\{X(t)\geq u\}}dt$

   First passage times $\longrightarrow \tau_x^+ = \inf\{t\geq 0, X(t)\geq x\}$

2. The trajectories of $X(t)$ are not observed
3. Obs.: $(U_n), (\Delta T_{n+1}), (\theta_n)$
4. Estimators $\hat{F}_{U,n}, \hat{G}_n, \hat{H}_n$ for the 3 distrib. F_U, G, and H
   Associated stochastic exposure process: $\hat{X}^{(n)}$.
5. Quantities of interest = complex functionals of F_U, G, and H

   $\Rightarrow$ No Plug-In estimator $\Rightarrow$ Simulation
6. Convergence of the simulation-based estimators?

Topology in the space of exposure processes: the one induced by the Hausdorff distance between completed graphs

Simulation-based estimators

1. Toxicologists'interest: distrib. or expectation of

   Maximum exposure value $\longrightarrow \sup_{0\leq t\leq T} X(t)$

   Mean exposure value $\longrightarrow T^{-1}\int_{t=0}^{T} X(t)dt$

   Average time spent over u $\longrightarrow T^{-1}\int_{t=0}^{T} \mathbb{I}_{\{X(t)\geq u\}}dt$

   First passage times $\longrightarrow \tau_x^+ = \inf\{t\geq 0, X(t)\geq x\}$

2. The trajectories of $X(t)$ are not observed
3. Obs.: $(U_n), (\Delta T_{n+1}), (\theta_n)$
4. Estimators $\hat{F}_{U,n}, \hat{G}_n, \hat{H}_n$ for the 3 distrib. F_U, G, and H
   Associated stochastic exposure process: $\hat{X}^{(n)}$.
5. Quantities of interest = complex functionals of F_U, G, and H

   $\Rightarrow$ No Plug-In estimator $\Rightarrow$ Simulation
6. Convergence of the simulation-based estimators?

Topology in the space of exposure processes: the one induced by the Hausdorff distance between completed graphs

Simulation-based estimators

1. Toxicologists'interest: distrib. or expectation of

   Maximum exposure value $\longrightarrow \sup_{0\leq t\leq T} X(t)$

   Mean exposure value $\longrightarrow T^{-1}\int_{t=0}^{T} X(t)dt$

   Average time spent over u $\longrightarrow T^{-1}\int_{t=0}^{T} \mathbb{I}_{\{X(t)\geq u\}}dt$

   First passage times $\longrightarrow \tau_x^+ = \inf\{t\geq 0, X(t)\geq x\}$

2. The trajectories of $X(t)$ are not observed
3. Obs.: $(U_n), (\Delta T_{n+1}), (\theta_n)$
4. Estimators $\hat{F}_{U,n}, \hat{G}_n, \hat{H}_n$ for the 3 distrib. F_U, G, and H
   Associated stochastic exposure process: $\hat{X}^{(n)}$.
5. Quantities of interest = complex functionals of F_U, G, and H

   $\Rightarrow$ No Plug-In estimator $\Rightarrow$ Simulation
6. Convergence of the simulation-based estimators?

Topology in the space of exposure processes: the one induced by the Hausdorff distance between completed graphs

Simulation-based estimators

1. Toxicologists'interest: distrib. or expectation of

   Maximum exposure value $\longrightarrow \sup_{0\leq t\leq T} X(t)$

   Mean exposure value $\longrightarrow T^{-1}\int_{t=0}^{T} X(t)dt$

   Average time spent over u $\longrightarrow T^{-1}\int_{t=0}^{T} \mathbb{I}_{\{X(t)\geq u\}}dt$

   First passage times $\longrightarrow \tau_x^+ = \inf\{t\geq 0, X(t)\geq x\}$

2. The trajectories of $X(t)$ are not observed
3. Obs.: $(U_n), (\Delta T_{n+1}), (\theta_n)$
4. Estimators $\hat{F}_{U,n}, \hat{G}_n, \hat{H}_n$ for the 3 distrib. F_U, G, and H
   Associated stochastic exposure process: $\hat{X}^{(n)}$.
5. Quantities of interest = complex functionals of F_U, G, and H

   $\Rightarrow$ No Plug-In estimator $\Rightarrow$ Simulation
6. Convergence of the simulation-based estimators?

Topology in the space of exposure processes: the one induced by the Hausdorff distance between completed graphs

Simulation-based estimators

1. Toxicologists'interest: distrib. or expectation of

   Maximum exposure value $\longrightarrow \sup_{0\leq t\leq T} X(t)$

   Mean exposure value $\longrightarrow T^{-1}\int_{t=0}^{T} X(t)dt$

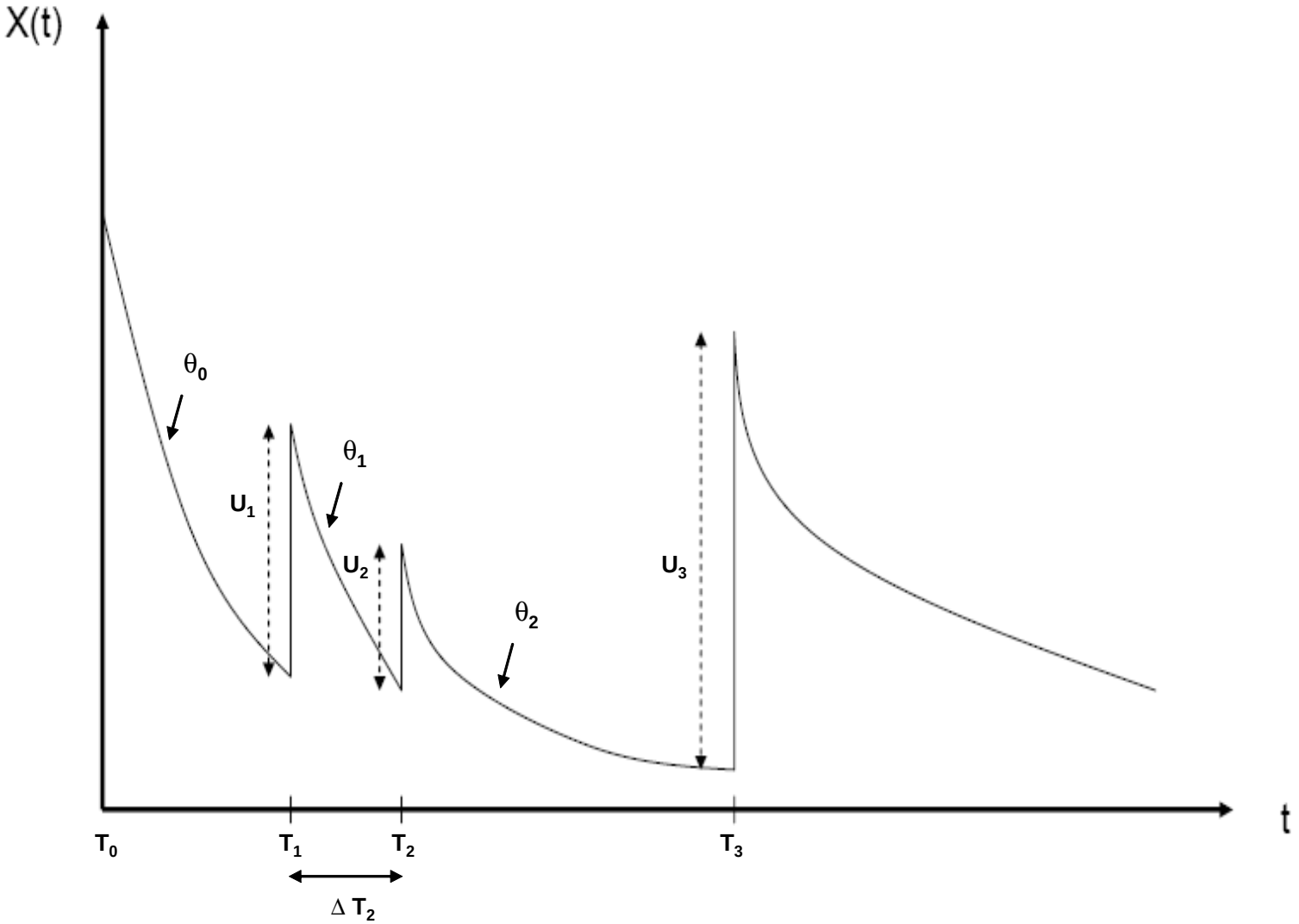
   Average time spent over u $\longrightarrow T^{-1}\int_{t=0}^{T} \mathbb{I}_{\{X(t)\geq u\}}dt$

   First passage times $\longrightarrow \tau_x^+ = \inf\{t\geq 0, X(t)\geq x\}$

2. The trajectories of $X(t)$ are not observed
3. Obs.: $(U_n), (\Delta T_{n+1}), (\theta_n)$
4. Estimators $\hat{F}_{U,n}, \hat{G}_n, \hat{H}_n$ for the 3 distrib. F_U, G, and H
   Associated stochastic exposure process: $\hat{X}^{(n)}$.
5. Quantities of interest = complex functionals of F_U, G, and H

   $\Rightarrow$ No Plug-In estimator $\Rightarrow$ Simulation
6. Convergence of the simulation-based estimators?

Topology in the space of exposure processes: the one induced by the Hausdorff distance between completed graphs


X(t)
θ0
θ1
θ2
U1
U2
U3
T0
T1
T2
T3
t
Δ T2

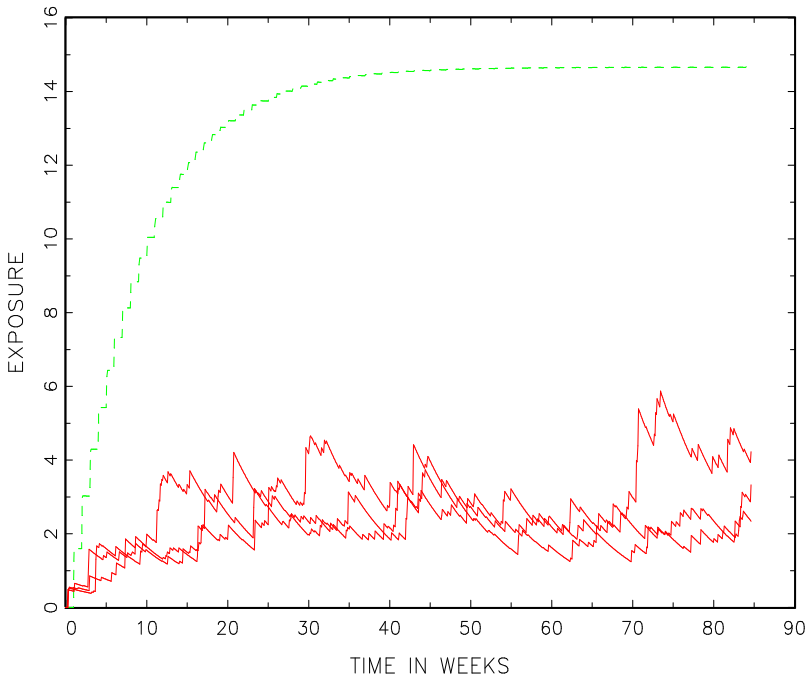
Under Mallows-L_1 consistency of the estimators $\hat{F}_{U,n}, \hat{G}_n, \hat{H}_n$

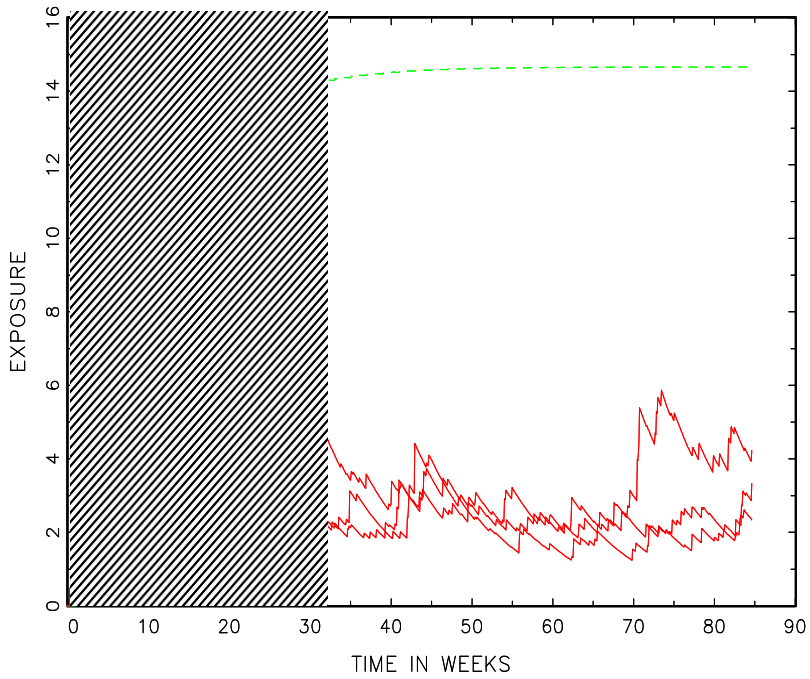
1. **Finite horizon:** $T < \infty$
   - Convergence in distribution for continuous functionals
     Ex: mean exposure, mean time above u
   - Convergence in mean for Lipschitz functionals
     Ex: Maximum exposure
   - Uniform integrability arguments yield convergence in mean for the mean exposure, mean time above u

2. **Steady state quantities:** $T \to \infty$
   Additional conditions:
   - $T^2 \times (M_1(\hat{G}^{(n)}, G) + M_1(\hat{H}^{(n)}, H)) \to 0,$ a.s.
   - $T \times M_1(\hat{F}_U^{(n)}, F_U) \to 0,$ a.s.
     $\Rightarrow$ Convergence in distrib. and in mean of the MC estimation of the mean exposure and the time spent over u.

3. **+ bootstrap confidence intervals:** $T < \infty$

Application to the French data on MeHg

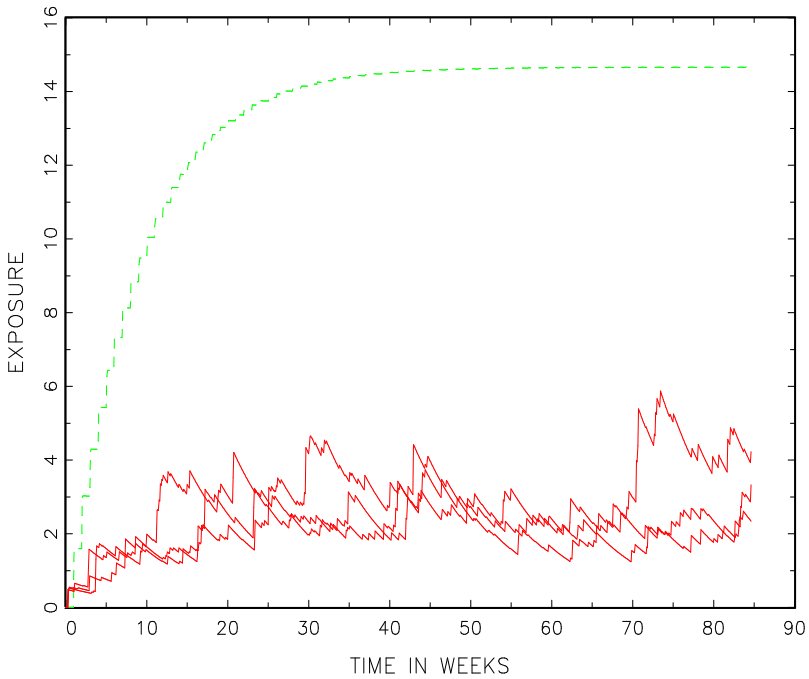


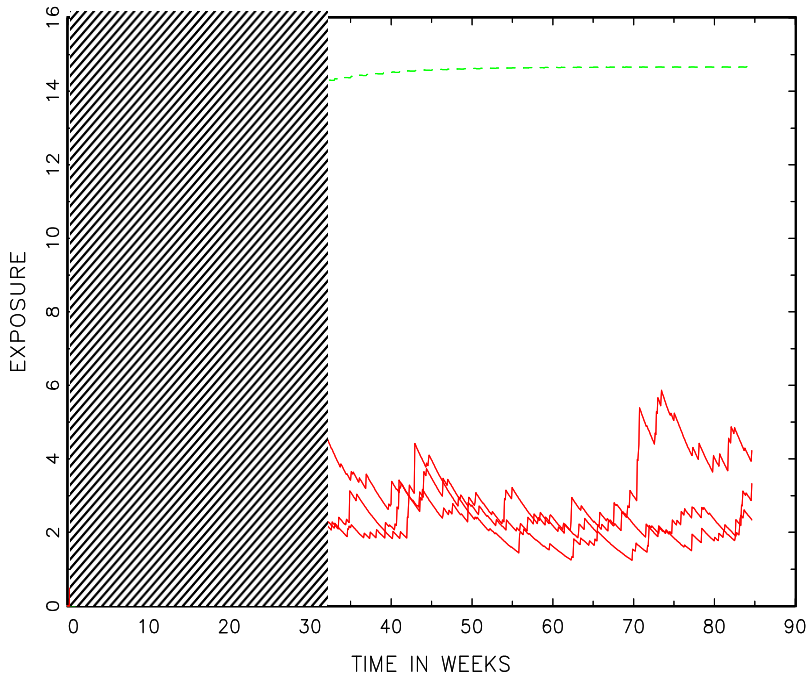


Def. of the Kinetic Tolerable Intake (KTI) based on the Provisional Tolerable Weekly Intake (PTWI) and the half life (HL) of the contaminant: $KTI = PTWI/(1 - 2^{-1/HL})$

$\mu(KTI, +\infty)$?

Application to the French data on MeHg

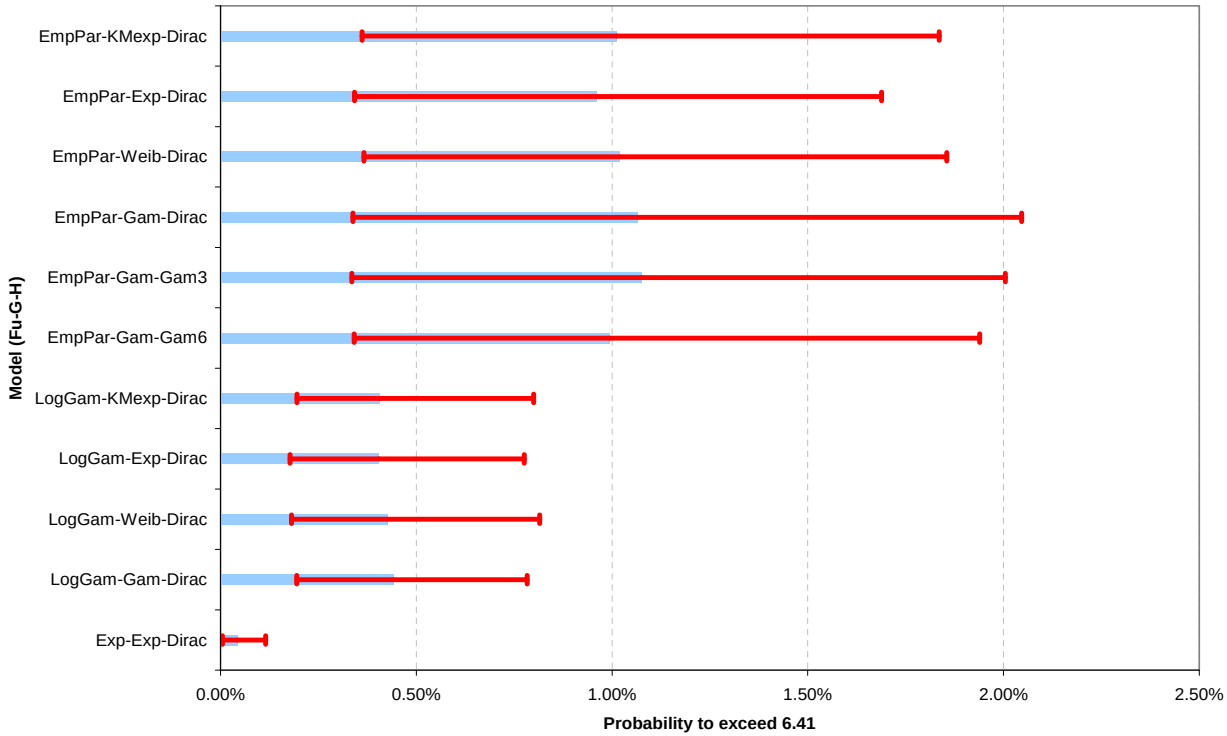




Def. of the Kinetic Tolerable Intake (KTI) based on the Provisional Tolerable Weekly Intake (PTWI) and the half life (HL) of the contaminant: $KTI = PTWI/(1 - 2^{-1/HL})$

$$\mu(KTI, +\infty)?$$

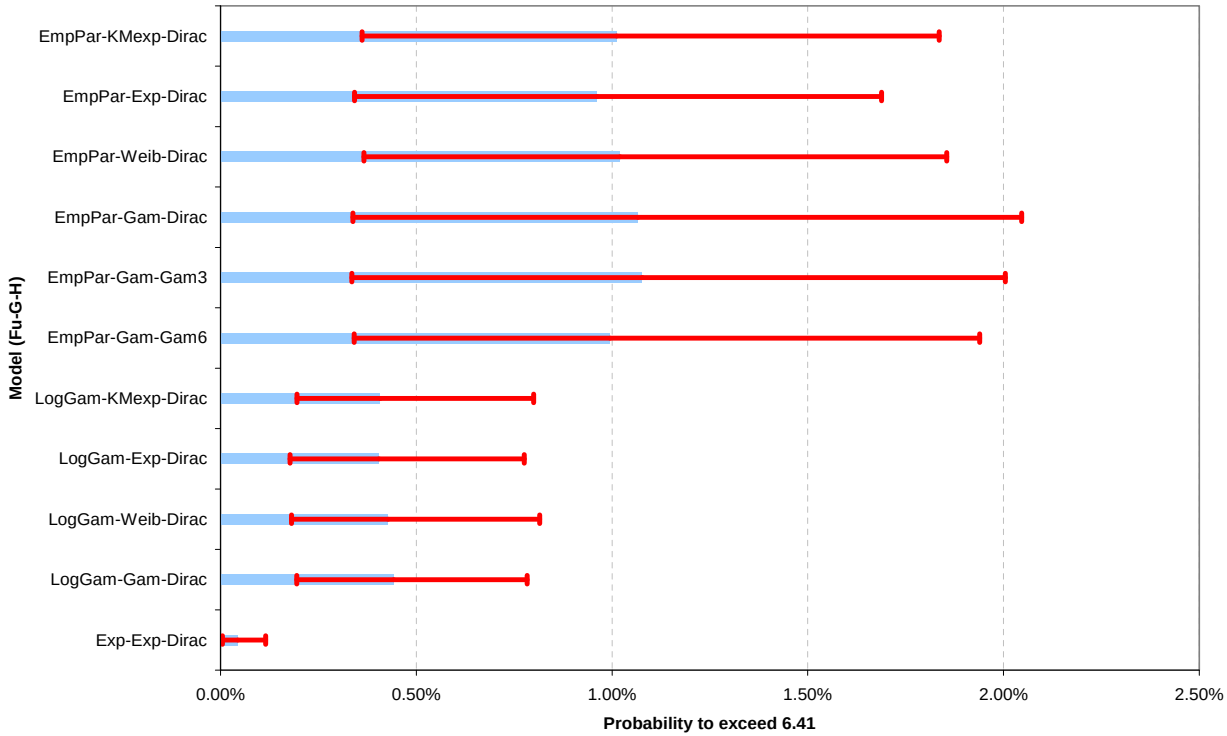
Bootstrap (95%-)confidence intervals for the simulation based estimators



Probab. of exceeding the KTI based on the (U.S.) National Research Council reference dose of $0.7\ \mu g/w/kgbw$, that is $6.41\ \mu g/kgbw$)

Naive MC simulation is not sufficient for estimating the probability of exceeding the KTI based on the PTWI $(14.67\ \mu g/kgbw)$

Bootstrap (95%-)confidence intervals for the simulation based estimators



Probab. of exceeding the KTI based on the (U.S.) National Research Council reference dose of $0.7\ \mu g/w/kgbw$, that is $6.41\ \mu g/kgbw$)

Naive MC simulation is not sufficient for estimating the probability of exceeding the KTI based on the PTWI $(14.67\ \mu g/kgbw)$

Rare event analysis: a particle filtering algorithm

Estimation of $\mathbb{P}_{x_0}(\tau_u^+ \leq T)$ with u large and a finite horizon T. Based on Cérou, Del Moral, LeGland & Lezaud (2006)

1. m intermediary levels $u_1 \leq \ldots \leq u_m \leq u_{m+1} = u$.
2. N particles from $X(0) = x_0 < u_1$ to T.
3. For each level u_j $(j = 1, \ldots, m)$:

   $I_{1,j} = \{$particles that reached u_j before T$\}$
   $I_{0,j} = \{$particles that have not reached u_j at T$\}$

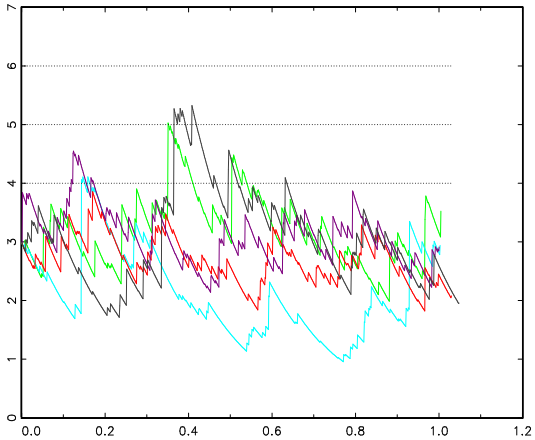
   - For particles in $I_{0,j}$,

     Selection: Randomly select a particle in $I_{1,j}$
     Mutation: Create an offspring from $\tau_{u_j}^+$ to T.

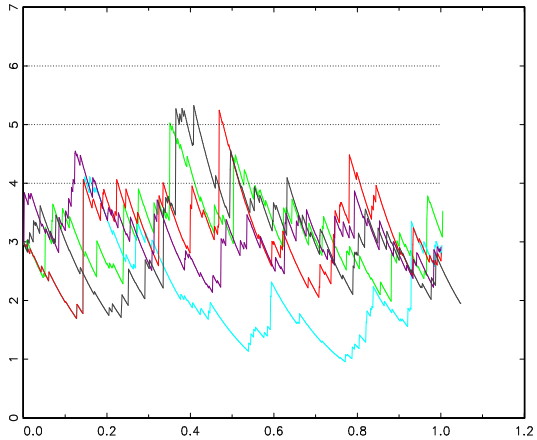
   - $P_j = |I_{1,j}|/N$ and pass onto next level u_{j+1}.

Estimator of $\mathbb{P}_{x_0}(\tau_u^+ \leq T)$: $P_1 \times \ldots \times P_{m+1}$

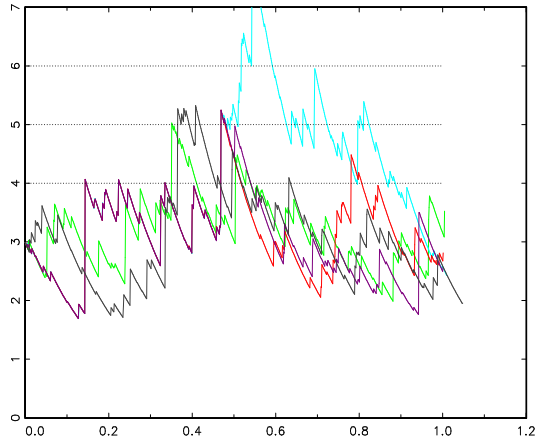
Multilevel Splitting Algorithm



Init.: $N = 5$



Iteration 1: $u = 4$



Iteration 2: $u = 5$

Illustration for $N = 5$ particules starting from $x_0 = 2.99$ with $m = 2$ intermediary levels ($u = (4, 5, 6)$) and a horizon T equal to one year.
(Women subpopulation, Log-Gamma/Gamma/Dirac model)

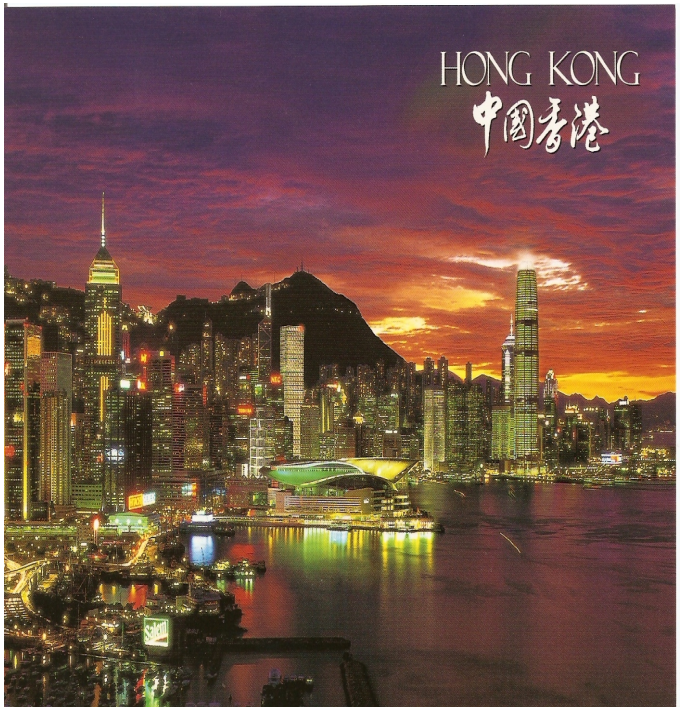
Application: $u = KTI = 14.67\ \mu g/kgbw$,
Intermediary levels from the adaptive algorithm of Cérou & Guyader (2005),

$\mathbb{P}_{2.99}(\tau^+_{KTI} \leq 20y) \approx 0.5\%$, $m_{20y} = 8$; $\mathbb{P}_{2.99}(\tau^+_{KTI} \leq 10y) \approx 0.08\%$, $m_{10y} = 12$.

Conclusion and Perspectives

Good properties of the KDEM model,
Some solutions for the estimation
of quantities of interest for toxicologists
BUT

1. Validation of the multilevel splitting on the basic Exp-Exp-Dirac model
   Kella & Stadje (2001): explicit Laplace Transform for τ_u^+
2. Pseudo-regenerative Extreme Value Estimators
   Comparison with the results of particle filtering
3. Multivariate marked point process



Other work with Monte carlo methods

- Bayesian nonparametrics
- Dirichlet process, Chinese Restaurant Process
- Lancelot F. James, Albert Y. Lo (HKUST-ISMT)

Thank you for your attention

Main References

- S. Asmussen (2003). *Applied Probability and Queues*, Springer-Verlag, New York, Second Edition.
- P. Bertail, S. Clémençon & J. Tressou (2006). A storage model for modelling exposure to food contaminants. https://hal.archives-ouvertes.fr/hal-00138279
- P. Bertail, S. Clémençon & J. Tressou (2007). Statistical analysis of a dynamic model for food contaminant exposure with applications to dietary methylmercury contamination.
- F. Cérou, P. Del Moral, F. LeGland & P. Lezaud (2006). Genetic Genealogical Models in Rare Event Analysis. *Alea* **1**, 183-196.
- S.P. Meyn & R.L. Tweedie (1996). *Markov Chains and Stochastic Stability*, Springer-Verlag.
- G.O. Roberts & R.L. Tweedie (2000). Rates of convergence of stochastically monotone and continuous time Markov models. *J. Appl. Probab.* **37**, 359-373.
- P. Verger, J. Tressou & S. Clémençon (2007). Integration of time as a description parameter in risk characterisation: application to methyl mercury. *Regulatory Toxicology and Pharmacology*.
- W. Whitt (2002). *Stochastic-Process Limits. An Introduction to Stochastic-Process Limits and their Application to Queues*, Springer.