Filtrage particulaire : une introduction avec applications

P. Del Moral

INRIA Centre Bordeaux-Sud Ouest

Journées GDR Statistique et Santé, octobre 2009
Univ. Paris-Descartes

Outline

En qq mots :

- 2 processus aléatoires : X Signal $\oplus$ Y Observation.
- Objectif $\rightsquigarrow$ Calculer les lois conditionnelles de X sachant Y

$$\mathrm{Loi}(X \mid Y)$$

Vocabulaire "équivalent" :

- *Assimilation de données (météo).*
- Chaines de Markov cachées (statistique fréquentiste).
- *Calcul de lois a posteriori (a priori*$=\mathrm{Loi}(X)$*)* (statistique baysienne).
- *Filtrage (proba. appliquées).*

Modèles de Filtrage de processus stochastiques

- Signal=Processus stochastique

ingénierie/physique/biologique/économique :

- Cibles non coopérative (militaire : missile, char, avion,...).
- Physiques (fluides : tornades, cyclones, modèles océano, pression/température/coeff de diffusivité,...).
- Financiers (actifs boursiers, portefeuilles, volatilités,...).
- Signaux (parole, codages/transmission informations,...)

Dynamiques et aléas :

- Equations d'évolution physiques (Ex. $\sum_i u_i \overrightarrow{F}_i = \overrightarrow{A}$)
- Perturbations et aléas :
  - Erreurs de modèles $\oplus$ Perturbations extérieures.
  - Contrôles et paramètres inconnus.
    $\rightsquigarrow$ lois a priori $\oplus$ (inconnues= réalisation de v.a.)

Modèles stochastiques de signaux (temps : $(n \in \mathbb{N}) \vee (t \in \mathbb{R}_+)$)

- Equation d'évolution : ($\in$ espace d'états E)

$$X_n = F_n(X_{n-1}, W_n) \quad \text{ou} \quad dX_t = A_t(X_t)\, dt + B_t(X_t)\, dW_t$$

- Interprétations et exemples :

$$\begin{array}{rcl} E & = & \mathbb{R},\ \mathbb{R}^d,\ \mathbb{R}^{\mathbb{Z}^2}, \text{espaces de trajectoires, excursions} \\ W_n & = & \text{Erreurs de modèles, perturbations extérieures} \\ & & \text{Contrôles et paramètres inconnus.} \end{array}$$

Note :

$$(E \rightsquigarrow E_n) \longrightarrow \text{mod. traject.} \quad X_n = (X'_0, \ldots, X'_n) \quad \text{non exclus}$$

Modèles de Filtrage de processus stochastiques

- Observations =Processus stochastique :

  ingénierie/physique/biologique/économique :

  - Ingénierie : Radar, Sonar, GPS, ...
  - Physiques (capteurs de pression/température/...).
  - Financiers (actifs boursiers, portefeuilles, volatilités,...).
  - Statistique (données réelles : médecine, pharmaceutiques, politique, éco-finances,...).

  Dynamiques et aléas :

  - Observations partielles : mélanges, coordonnées partielles.
  - Perturbations et aléas :
    - erreurs de mesures (bruits thermiques).
    - perturbations extérieures.
    - erreurs de modèle.

Modèles stoch. de capteurs/observations $((n \in \mathbb{N}) \vee (t \in \mathbb{R}_+))$

- Equation d'observation: ($\in$ espace d'états E')

$$Y_n = H_n(X_n, V_n) \quad \text{ou} \quad dY_t = H_t(X_t)\ dt + J_t(X_t)\ dV_t$$

- Interprétations et exemples :

$$\begin{array}{rcl} E & = & \mathbb{R},\ \mathbb{R}^d,\ \mathbb{R}^{\mathbb{Z}^2}, \text{espaces de trajectoires, excursions} \\ V_n & = & \text{Erreurs de mesures/modèles/extérieures} \end{array}$$

Objectifs

- Prédiction : $m < n \rightsquigarrow \mathrm{Loi}(X_n \mid (Y_0, \ldots, Y_m))$.
- Filtrage : $m = n \rightsquigarrow \mathrm{Loi}(X_n \mid (Y_0, \ldots, Y_n))$.
- Lissage : $m < n \rightsquigarrow \mathrm{Loi}(X_m \mid (Y_0, \ldots, Y_n))$.

Remarques

- Filtrage trajectoriel :

$$\rightsquigarrow \mathrm{Loi}((X_0, \ldots, X_n) \mid (Y_0, \ldots, Y_n))$$

- Modèle trajectoriel = Filtrage "simple" :

$$\mathcal{X}_n = (X_n, \ldots, X_n) \in E_n \quad \rightsquigarrow \quad \mathrm{Loi}(\mathcal{X}_n \mid (Y_0, \ldots, Y_n))$$

Les équations du filtrage

2 Etapes d'apprentissage/d'adpatation naturelles :

$$\mathrm{Loi}(X_n \mid (Y_0,\ldots,Y_n)) \xrightarrow{\text{prédiction}} \mathrm{Loi}(X_{n+1} \mid (Y_0,\ldots,Y_n))$$

$$\mathrm{Loi}(X_{n+1} \mid (Y_0,\ldots,Y_n)) \xrightarrow{\text{correction}} \mathrm{Loi}(X_{n+1} \mid ((Y_0,\ldots,Y_n),Y_{n+1}))$$

QQ interprétations utiles :

- Prédiction = Examen de la stat. d'évol. $X_n \rightsquigarrow X_{n+1}$ ($\perp Y$)
- Correction = Mise à jour des lois $\sim$ novel obs. Y_{n+1}

$\rightsquigarrow$ Potentiel de vraisemblance sur chaque site x :

$$G_{n+1}(x) = \text{"Proba de voir } Y_{n+1} \text{ si } X_{n+1} \text{ était en } x\text{"}$$

L'heuristique du filtrage particulaire

Objectif

Dynamique de population de N "individus"/particules" t.q.

$$(\widehat{\xi}_n^1, \ldots, \widehat{\xi}_n^N) \in E_n^N \rightsquigarrow \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{\widehat{\xi}_n^i} = \mathrm{Loi}(X_n \mid (Y_0, \ldots, Y_n))$$

2 Etapes d'apprentissage/d'adpatation naturelles :

$$(\widehat{\xi}_n^1, \ldots, \widehat{\xi}_n^N) \in E_n^N \xrightarrow{\text{exploration}} (\xi_{n+1}^1, \ldots, \xi_{n+1}^N) \in E_{n+1}^N$$

$$(\xi_{n+1}^1, \ldots, \xi_{n+1}^N) \in E_{n+1}^N \xrightarrow{\text{adaptation}} (\widehat{\xi}_{n+1}^1, \ldots, \widehat{\xi}_{n+1}^N) \in E_{n+1}^N$$

L'heuristique du filtrage particulaire

Interprétations $\oplus$ Correspondances

$$\begin{aligned}
\text{(prédiction,correction)} &= \text{(propositions,accept./rejet)} \\
&= \text{(mutation,selection)} \\
&= \text{(exploration,adaptation/branchements, \ldots)}
\end{aligned}$$

Méthode de simulation effective

- Prédiction $\rightsquigarrow$ simulation de N transitions locales du signal
- Correction$\rightsquigarrow$ processus de naissance et mort (taille fixe N)
  - On tue les individus sur des sites peu vraisemblables.
  - On multiplie les individus sur des sites plus vraisemblables.

$$\left.\begin{array}{ccc} \widehat{\xi}^1_n \in E_n & \xrightarrow{\text{exploration}} & \xi^1_{n+1} \in E_{n+1} \\ \vdots & & \vdots \\ \widehat{\xi}^i_n & \longrightarrow & \xi^i_{n+1} \\ \vdots & & \vdots \\ \widehat{\xi}^N_n & \longrightarrow & \xi^N_{n+1} \end{array}\right] \xrightarrow{\text{adaptation}} \begin{array}{c} \widehat{\xi}^1_{n+1} \in E_{n+1} \\ \vdots \\ \widehat{\xi}^i_{n+1} \\ \vdots \\ \widehat{\xi}^N_{n+1} \end{array}$$

- Exploration ($\sim$ simu. des aléas des transitions locales)

$$(\text{signal} \; : \; X_{n+1} = F(X_n, W_{n+1})) \rightsquigarrow \forall i \quad \xi^i_{n+1} = F(\widehat{\xi}^i_n, W^i_{n+1})$$

- Correction $\rightsquigarrow$ *Exemple : branchements multinomiaux*

$$(\widehat{\xi}^i_{n+1})_{1 \leq i \leq N} \text{ "iid" de loi discrète : } \sum_{i=1}^{N} \frac{G_{n+1}(\xi^i_{n+1})}{\sum_{j=1}^{N} G_{n+1}(\xi^j_{n+1})} \; \delta_{\xi^i_{n+1}}$$

Remarques importantes

Modèle trajectoriel $\mathcal{X}_n = (X_0, \ldots, X_n)$

- 1 particule=1 ligne ancestrale

$$\widehat{\xi}^i_n = (\widehat{\xi}^i_{0,n}, \widehat{\xi}^i_{1,n}, \ldots, \widehat{\xi}^i_{n,n})$$

$$p \in \{0, \ldots, n\} \to \widehat{\xi}^i_{p,n} \quad \text{ancêtre au niveau } p \text{ de l'individu} \quad \widehat{\xi}^i_{n,n}$$

- les mesures d'occupation des arbres généalogiques :

$$\frac{1}{N} \sum_{1 \le i \le N} \delta_{\widehat{\xi}^i_n} \quad \simeq_{N\uparrow\infty} \quad \mathrm{Loi}((X_0, \ldots, X_n) \mid (Y_0, \ldots, Y_n))$$

Exemple :

$$\frac{1}{N} \sum_{1 \le i \le N} \delta_{\widehat{\xi}^i_{0,n}} \simeq_{N\uparrow\infty} \mathrm{Loi}(X_0 \mid (Y_0, \ldots, Y_n))$$

Remarques importantes (suite)

Filtrage des bruits $\rightsquigarrow$ **Régulation optimale**

$$\frac{1}{N}\sum_{1\leq i\leq N}\delta_{(\widehat{W}^i_{0,n},\widehat{W}^i_{1,n},\ldots,\widehat{W}^i_{n,n})}\simeq_{N\uparrow\infty}\mathrm{Loi}((W_0,\ldots,W_n)\mid(Y_0,\ldots,Y_n))$$

Algorithmes stoch."équivalents" :

- Algorithmes génétiques, processus de branchements.
- Algorithmes de "population Monte Carlo", DMC, QMC, ...
- Processus d'adaptation :

  bootsrapping, selection, pruning-enrichment, reconfiguration, cloning, go with the winner, spawning, condensation, grouping, rejuvenations, harmony searches, biomimetics, ...

$$\Updownarrow$$

$$1950 \leq [\text{Méthodes de simulation heuristiques}] \leq 1996$$

Flot de mesures conditionnelles

- Flot de mesures de probabilités aléatoires : (ex.: prédicteurs)

$$n \in \mathbb{N} \mapsto \eta_n := \mathrm{Loi}(X_n \mid (Y_0, \ldots, Y_{n-1}))$$

- Système dynamique non linéaire : $\boxed{\eta_n = \Phi_n(\eta_{n-1})}$

$$\eta_n \xrightarrow{\text{Correction}} \widehat{\eta}_n = \Psi_n(\eta_n) \xrightarrow{\text{Prédiction}} \eta_{n+1} = \widehat{\eta}_n M_{n+1}$$

- **Transformation (non linéaire) de Boltzmann-Gibbs :**

$$\Psi_n(\eta_n)(dx) := \frac{1}{\eta_n(G_n)} \; G_n(x) \; \eta_n(dx) \quad (\text{"formule de Bayes"})$$

- **Transport linéaire markovien :**

$$(\widehat{\eta}_n M_{n+1})(dy) := \int \widehat{\eta}_n(dx) \; M_{n+1}(x, dy)$$

Chaines de Markov non linéaires

$\exists$ Markov $K_{n,\eta}(x, dy)$ (non unique) t.q.

$$\Phi_n(\eta_{n-1}) = \eta_{n-1} K_{n,\eta_{n-1}} = \text{Loi}\left(\overline{X}_n\right) = \eta_n$$

i.e. :

$$\mathbb{P}(\overline{X}_n \in dx_n \mid \overline{X}_{n-1}) = K_{n,\eta_{n-1}}(\overline{X}_{n-1}, dx_n)$$

Interpretation particulaire de type champ moyen :

- Chaine Markov $\xi_n = (\xi_n^1, \ldots, \xi_n^N) \in E_n^N$ t.q.

$$\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \simeq_{N \uparrow \infty} \eta_n$$

- Transitions locales "approchées" $(\forall 1 \leq i \leq N)$

$$\xi_{n-1}^i \rightsquigarrow \xi_n^i \sim K_{n,\eta_{n-1}^N}(\xi_{n-1}^i, dx_n)$$

QQ Avantages

- Modèle de champ moyen=linéarisation/perturbation stoch. :

$$\eta_n^N = \Phi_n(\eta_{n-1}^N) + \frac{1}{\sqrt{N}}\ W_n^N$$

avec $W_n^N \simeq W_n$ Champs gaussiens centrés et $\perp$.

- $\eta_n = \Phi_n(\eta_{n-1})$ stable $\Rightarrow$ non propagation des erreurs locales

$\Longrightarrow$ control uniforme des erreurs globales / temps

$\Rightarrow \mathrm{TCL} = \sum \text{des erreurs locales/champs gaussiens transportés}$

- "Pas besoin" d'étudier la cv à l'équilibre de modèles MCMC.
- Grille/Maillage stochastique adaptatif.
- Non linéarité du syst. $\rightsquigarrow$ interactions bénéfiques.
- Algo. naturel et facile a simuler, etc.

Modèles de Feynman-Kac

(Correction,Prédiction)=(Selection,Exploration)=(G_n, M_n)

$$\eta_n^N \simeq_{N\uparrow\infty} \eta_n = \Phi_n(\eta_{n-1}) := \Psi_{G_{n-1}}(\eta_{n-1})M_n$$

"Solution" : [X_n Markov $\sim$ transitions M_n]

$$\eta_n(f_n) = \frac{\gamma_n(f_n)}{\gamma_n(1)} \quad \text{with} \quad \gamma_n(f_n) = \mathbb{E}\left(f_n(X_n)\prod_{0\leq p<n} G_p(X_p)\right)$$

Formule $\prod$ $\rightsquigarrow$ est. non biaisée des Cts. de normalisation

$$\gamma_n(1) = \prod_{0\leq p<n}\eta_p(G_p) \simeq_{N\uparrow\infty} \gamma_n^N(1) := \prod_{0\leq p<n}\eta_p^N(G_p)$$

Fonctionnelles additives

- Modèles trajectoriels $\mathbb{P}_n := \mathrm{Loi}(X_0, \ldots, X_n)$

$$d\mathbb{Q}_n := \frac{1}{\mathcal{Z}_n} \left\{ \prod_{0 \leq p < n} G_p(X_p) \right\} d\mathbb{P}_n$$

- Hyp. : $M_n(x_{n-1}, dx_n) = H_n(x_{n-1}, x_n)\ \lambda_n(dx_n)$

$$\Longrightarrow \mathbb{Q}_n(d(x_0, \ldots, x_n)) = \eta_n(dx_n) \prod_{q=1}^{n} \mathcal{M}_{q,\eta_{q-1}}(x_q, dx_{q-1})$$

transitions rebours : $\mathcal{M}_{p+1,\eta}(x, dx') \propto G_p(x')\ H_{p+1}(x', x)\ \eta(dx')$

- Estimation particulaire $\sim$ arbre généalogique complet :

$$\mathbb{Q}_n^N(d(x_0, \ldots, x_n)) = \eta_n^N(dx_n) \prod_{q=1}^{n} \mathcal{M}_{q,\eta_{q-1}^N}(x_q, dx_{q-1})$$

Résumé : 4 Estimations particulaires

- Arbre généalogique complet $\Longrightarrow$ Mes. McKean $\oplus$ F-K trajectorielle

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{(\xi_0^i,\ldots,\xi_n^i)} \simeq_N \mathrm{Loi}(\overline{X}_0,\ldots,\overline{X}_n) \quad \& \quad \mathbb{Q}_n^N \simeq_N \mathbb{Q}_n$$

- Arbre généalogique (simple) $\Longrightarrow$ Mesure de F-K trajectorielle

$$\eta_n^N = \frac{1}{N}\sum_{i=1}^{N}\delta_{(\xi_{0,n}^i,\xi_{1,n}^i,\ldots,\xi_{n,n}^i)} \simeq_N \mathbb{Q}_n = \eta_n$$

- Constantes de normalisation

$$\gamma_n^N(1) := \prod_{0\leq p<n}\eta_p^N(G_p) \simeq_N \gamma_n(1) = \mathbb{E}\left(\prod_{0\leq p<n}G_p(X_p)\right)$$

QQ estimations non asymptotiques (+ de détails = réfs)

- Processus empiriques

$$\sup_{n\geq 0}\sup_{N\geq 1}\sqrt{N}\,\mathbb{E}(\|\eta_n^N-\eta_n\|_{\mathcal{F}_n}^p)<\infty$$

- Inégalités de concentration uniformes :

$$\sup_{n\geq 0}\ \mathbb{P}(|\eta_n^N(f_n)-\eta_n(f_n)|>\epsilon)\leq c\ \exp\left\{-(N\epsilon^2)/(2\sigma^2)\right\}$$

- Propagations du chaos (= biais+$\perp$): $\mathbb{P}_{n,q}^N:=\mathrm{Law}(\xi_n^{(1,N)},\ldots,\xi_n^{(q,N)})$

$$\mathbb{P}_{n,q}^N\simeq\eta_n^{\otimes q}+\frac{1}{N}\,\partial^1\mathbb{P}_{n,q}+\ldots+\frac{1}{N^k}\,\partial^k\mathbb{P}_{n,q}+\frac{1}{N^{k+1}}\,\partial^{k+1}\mathbb{P}_{n,q}^N$$

avec $\sup_{N\geq 1}\|\partial^{k+1}\mathbb{P}_{n,q}^N\|_{\mathrm{tv}}<\infty$ & $\sup_{n\geq 0}\|\partial^1\mathbb{P}_{n,q}\|_{\mathrm{tv}}\leq c\ q^2$.

Rmq : Arbres géalogiques simples $\rightsquigarrow$ idem sans les $\sup_n$

QQ estimations non asymptotiques (suite)

- *Est. des variances* (+ Cerou & Guyader `Hal-INRIA nov.08`) *:*

$$\mathbb{E}\left(\left[\gamma_n^N(1)-\gamma_n(1)\right]^2\right) \leq c\,\frac{n}{N} \times \gamma_n(1)^2$$

- *Modèles trajectoriels* (+ Doucet & Singh `Hal-INRIA juil.09`) *:*

$$F_n(x_0,\ldots,x_n)=\frac{1}{n+1}\sum_{0\leq p\leq n} f_p(x_p)$$

  - Estimations biais+$\mathbb{L}_p$ uniformes en temps + variance

$$N\,\mathbb{E}\left([(\mathbb{Q}_n^N-\mathbb{Q}_n)(F_n)]^2\right) \leq c \times (1/n+1/N)$$

  - Concentration sous gaussienne uniforme en temps :

$$\frac{1}{N}\log\sup_{n\geq 0}\mathbb{P}\left(\left|[\mathbb{Q}_n^N-\mathbb{Q}_n](F_n)\right| \geq \frac{b}{\sqrt{N}}+\epsilon\right) \leq -\epsilon^2/(2b^2)$$

"Idées intuitives" $\rightsquigarrow$ sg. non linéaire: $\eta_n = \Phi_n(\eta_{n-1}) = \Phi_{p,n}(\eta_p) = \eta_n$

Erreurs locales

$$W_n^N := \sqrt{N}\ \left[\eta_n^N - \Phi_n\left(\eta_{n-1}^N\right)\right] \simeq W_n\ \perp\ \text{champs gaussien}$$

Transport des erreurs locales :

$$\begin{array}{ccccccccc}
\eta_0 & \longrightarrow & \eta_1 = \Phi_1(\eta_0) & \longrightarrow & \eta_2 = \Phi_{0,2}(\eta_0) & \longrightarrow & \cdots & \longrightarrow & \Phi_{0,n}(\eta_0) \\
\Downarrow & & & & & & & & \\
\eta_0^N & \longrightarrow & \Phi_1(\eta_0^N) & \longrightarrow & \Phi_{0,2}(\eta_0^N) & \longrightarrow & \cdots & \longrightarrow & \Phi_{0,n}(\eta_0^N) \\
 & & \Downarrow & & & & & & \\
 & & \eta_1^N & \longrightarrow & \Phi_2(\eta_1^N) & \longrightarrow & \cdots & \longrightarrow & \Phi_{1,n}(\eta_1^N) \\
 & & & & \Downarrow & & & & \\
 & & & & \eta_2^N & \longrightarrow & \cdots & \longrightarrow & \Phi_{2,n}(\eta_2^N) \\
 & & & & & & & & \vdots \\
 & & & & & & \Downarrow & & \\
 & & & & & & \eta_{n-1}^N & \longrightarrow & \Phi_n(\eta_{n-1}^N) \\
 & & & & & & & & \Downarrow \\
 & & & & & & & & \eta_n^N
\end{array}$$

$\rightsquigarrow$ **Formule de décomposition** : *[Propagation des erreurs locales]+[ok pout* **tout** *schéma d'approximation]*

$$\begin{aligned}
\eta_n^N - \eta_n &= \sum_{q=0}^{n} [\Phi_{q,n}(\eta_q^N) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))] \\
&\simeq \frac{1}{\sqrt{N}} \sum_{q=0}^{n} W_q^N D_{q,n} \quad \text{decomp. 1er ordre } \Phi_{p,n}(\eta) - \Phi_{p,n}(\mu) \simeq (\eta - \mu) D_{p,n} + (\eta - \mu)^{\otimes 2} \ldots
\end{aligned}$$

$\Rightarrow$ Exemple estimation unif. + TCL en deux lignes : $\sqrt{N}\left[\eta_n^N - \eta_n\right] \simeq \sum_{q=0}^{n} W_q D_{q,n}$

Quelques réfs "récentes"

- Feynman-Kac formulae. Genealogical and interacting particle systems, Springer (2004) ⊕ Refs.
- joint work with Doucet A. and Singh S. S. A Backward Particle Interpretation of Feynman-Kac Formulae HAL-INRIA RR-7019 [30p], (July 2009).
- joint work with Doucet A. and Singh S. S. Forward Smoothing Using Sequential Monte Carlo Technical Report 638 (2009). Cambridge University Engineering Department.
- joint work with Doucet A., Jasra A. Sequential Monte Carlo Samplers. JRSS B (2006).