Utilisation d'échantillonnage pondéré dans les estimations répétées

Dorota Gajda[1 2]
Chantal Guihenneuc-Jouyaux [2 3]

[1]IGR [2]INSERM U1018 - EQ01PTB - CESP
[3]Université Paris Descartes

AppliBUGS, 20 décembre 2012

Étude empirique d'estimateurs par simulations

- Différentes valeurs des paramètres : L paramétrisations
- Simulation de K jeux de données sous le modèle d'intérêt
- Estimation des paramètres sur chaque jeu de données (K$\times$L)
- Etude des performances des estimations des paramètres (pour chaque paramétrisation)

Étude empirique d'estimateurs par simulations

- Différentes valeurs des paramètres : L paramétrisations
- Simulation de K jeux de données sous le modèle d'intérêt
- Estimation des paramètres sur chaque jeu de données (K×L)
- Etude des performances des estimations des paramètres (pour chaque paramétrisation)

Étude empirique d'estimateurs par simulations

- Différentes valeurs des paramètres : L paramétrisations
- Simulation de K jeux de données sous le modèle d'intérêt
- Estimation des paramètres sur chaque jeu de données (K$\times$L)
- Etude des performances des estimations des paramètres (pour chaque paramétrisation)

Étude empirique d'estimateurs par simulations

- Différentes valeurs des paramètres : L paramétrisations
- Simulation de K jeux de données sous le modèle d'intérêt
- Estimation des paramètres sur chaque jeu de données (K$\times$L)
- Etude des performances des estimations des paramètres (pour chaque paramétrisation)

Modèle Bayésien

Modèle paramétrique M_θ pour les données (X)

- idée a priori sur θ $\Rightarrow \pi(\theta)$
- information provenant des données $\Rightarrow \pi(X|\theta)$
- Combinaison des 2 sources d'information

(a priori + données)= loi a posteriori $\Rightarrow \pi(\theta|X)$

par Théorème de Bayes

$$\pi(\theta|X) = \frac{\pi(\theta, X)}{\pi(X)} = \frac{\pi(X|\theta) \cdot \pi(\theta)}{\pi(X)}$$

Estimations répétées

- Étude « empirique » d'estimateurs $\rightarrow$ via les simulations
- Nécessité d'un grand nombre de réplications de jeux de données
- Les résultats d'une approche bayésienne

$\rightarrow$ loi a posteriori (souvent multidimensionnelle)

ainsi que ses caractéristiques

(**en pratique approchées par algorithmes itératifs**)

- Temps de calcul parfois lourd
  - $\Rightarrow$ Diminution du nombre de réplications ou
  - $\Rightarrow$ Étude considérée non faisable

Étude de simulations dans le cadre Bayésien

- Simuler K jeux de données $X^{(K)} = (X_1, \dots X_N)$ sous le modèle M_θ en fixant les paramètres $\theta = \theta_0$

$$X_i \sim \pi(X|\theta_0)$$

- Fixer une même loi a priori sur les paramètres $\pi(\theta)$
- Calculer des espérances d'une fonction d'intérêt $g(\theta)$:

$$E_{[\theta|X]}[g(\theta)] = \int_\Theta g(\theta)\pi(\theta|x)d\theta$$

$$g(\theta) = \theta, \quad g(\theta) = \theta^2, \quad g(\theta) = \dots$$

- Étudier L scénarios pour les différents choix de θ_0

Problème calculatoire

- $\pi(\theta|X^{(K)})$ et/ou les intégrales de $\pi(\theta|X^{(K)})$ non explicites

Méthode classique : MCMC systématiquement

- Algorithme permettant de simuler une chaîne de Markov $\{\theta_1, ...\theta_N\}$ ayant pour loi stationnaire la loi a posteriori $\pi(\theta|X)$

  (N=nombre total d'itérations)

- Par théorème ergodique,

Si CM $\{\theta_1, ..., \theta_N\} \sim \pi(\theta|X)$ alors

$$\frac{1}{N}\sum_{i=1}^{N} g(\theta_i) \longrightarrow E_{[\theta|X]}[g(\theta)]$$

pour $g(\theta) = \theta, \quad g(\theta) = \theta^2, \quad g(\theta) = \dots$

Méthode itérative pour $K \times L$ estimations

$\Rightarrow$ procédure coûteuse en temps

Importance Sampling avec 2 jeux de données

- On a déjà une réalisation Markov. de la loi a posteriori $\pi(\theta|X^{(1)})$ (MCMC avec $X^{(1)}$)
- On veut estimer $E_{[\theta|\ X^{(2)}]}[g(\theta)]$ sans faire de MCMC

$$E_{[\theta|X^{(2)}]}[g(\theta)] = \int g(\theta)\pi(\theta|X^{(2)})d\theta = \int g(\theta)\frac{\pi(\theta|X^{(2)})}{\pi(\theta|X^{(1)})}\pi(\theta|X^{(1)})d\theta$$

Par th. ergodique,

si $\theta_1, ..., \theta_N \sim \pi(\theta|X^{(1)})$ alors

$$\frac{1}{N}\sum_{i=1}^{N} g(\theta_i)\frac{\pi(\theta_i|X^{(2)})}{\pi(\theta_i|X^{(1)})} \rightarrow E_{[\theta|X^{(2)}]}[g(\theta)]$$

$\pi(\theta|X^{(1)})$ **est appelée fonction d'importance**

IS combinée avec MCMC pour K=101

1$^{\text{ère}}$ stratégie : "Référence fixe"

1$\times$**MCMC** $+$ **100**$\times$**IS**

$\Rightarrow$ la même fonction d'importance $\pi(\theta|X^{(1)})$ obtenue via MCMC avec $X^{(1)}$

2$^{\text{ème}}$ stratégie : "Référence choisie"

10$\times$**MCMC** $+$ **91**$\times$**IS**

$\Rightarrow$ la fonction d'importance est différente $\pi(\theta|X^{(m_k)})$ où $X^{(m_k)}$ est choisie parmi $X^{(1)}, .., X^{(10)}$ via un des trois critères (Gajda et al. 2010) :

- critère 1 : minimisation de la norme L_1 : $||\pi(\theta^{(m_k)}|X^{(k)}) - \pi(\theta^{(m_k)}|X^{(m_k)})||_{L^1}$
- critère 2 : minimisation de la KL-divergence : $KL(\pi(\theta^{(m_k)}|X^{(k)}), \pi(\theta^{(m_k)}|X^{(m_k)}))$
- critère 3 : minimisation de la variance IS

3$^{\text{ème}}$ stratégie : "Mélange de densités"

10$\times$**MCMC** $+$ **91**$\times$**IS**

$\Rightarrow$ la fonction d'importance $\pi_{mix}(\theta) = \frac{1}{10}\sum_{j=1}^{10}\pi(\theta|X^{(j)})$ (Gajda et al. 2011)

IS combinée avec MCMC pour K=101

1$^{\text{ère}}$ stratégie : "Référence fixe"

1$\times$**MCMC** $+$ **100**$\times$**IS**

$\Rightarrow$ la même fonction d'importance $\pi(\theta|X^{(1)})$ obtenue via MCMC avec $X^{(1)}$

2$^{\text{ème}}$ stratégie : "Référence choisie"

10$\times$**MCMC** $+$ **91**$\times$**IS**

$\Rightarrow$ la fonction d'importance est différente $\pi(\theta|X^{(m_k)})$ où $X^{(m_k)}$ est choisie parmi $X^{(1)}, .., X^{(10)}$ via un des trois critères (Gajda et al. 2010) :

- critère 1 : minimisation de la norme L_1 : $||\pi(\theta^{(m_k)}|X^{(k)}) - \pi(\theta^{(m_k)}|X^{(m_k)})||_{L^1}$
- critère 2 : minimisation de la KL-divergence : $KL(\pi(\theta^{(m_k)}|X^{(k)}), \pi(\theta^{(m_k)}|X^{(m_k)}))$
- critère 3 : minimisation de la variance IS

3$^{\text{ème}}$ stratégie : "Mélange de densités"

10$\times$**MCMC** $+$ **91**$\times$**IS**

$\Rightarrow$ la fonction d'importance $\pi_{mix}(\theta) = \frac{1}{10}\sum_{j=1}^{10} \pi(\theta|X^{(j)})$ (Gajda et al. 2011)

IS combinée avec MCMC pour K=101

1ère stratégie : "Référence fixe"

1 $\times$ MCMC + 100 $\times$ IS

$\Rightarrow$ la même fonction d'importance $\pi(\theta|X^{(1)})$ obtenue via MCMC avec $X^{(1)}$

2ème stratégie : "Référence choisie"

10 $\times$ MCMC + 91 $\times$ IS

$\Rightarrow$ la fonction d'importance est différente $\pi(\theta|X^{(m_k)})$ où $X^{(m_k)}$ est choisie parmi $X^{(1)}, .., X^{(10)}$ via un des trois critères (Gajda et al. 2010) :

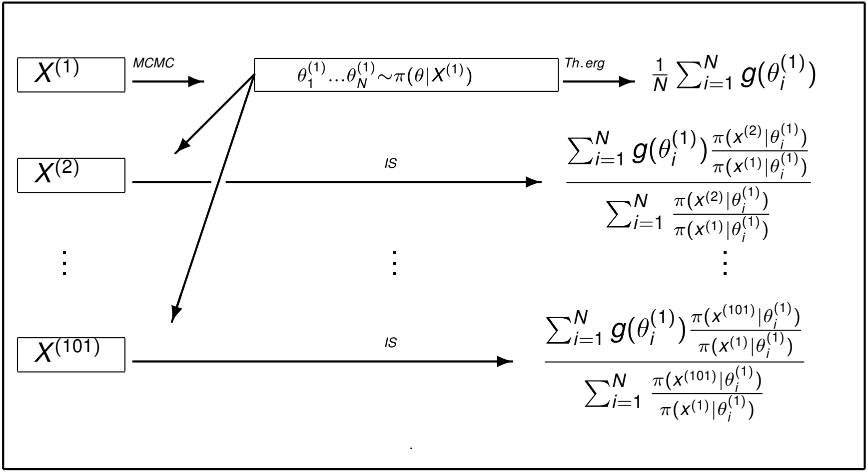
- critère 1 : minimisation de la norme L_1 : $||\pi(\theta^{(m_k)}|X^{(k)}) - \pi(\theta^{(m_k)}|X^{(m_k)})||_{L^1}$
- critère 2 : minimisation de la KL-divergence : $KL(\pi(\theta^{(m_k)}|X^{(k)}), \pi(\theta^{(m_k)}|X^{(m_k)}))$
- critère 3 : minimisation de la variance IS

3ème stratégie : "Mélange de densités"

10 $\times$ MCMC + 91 $\times$ IS

$\Rightarrow$ la fonction d'importance $\pi_{mix}(\theta) = \frac{1}{10}\sum_{j=1}^{10} \pi(\theta|X^{(j)})$ (Gajda et al. 2011)

1$^{\text{ère}}$ stratégie : "Référence fixe"



$2^{\text{ème}}$ stratégie : "Référence choisie" - objectif

$\operatorname{supp}\left\{g\pi(\cdot|X^{(k)})\right\} \subset \operatorname{supp}\left\{\pi(\cdot|X^{(1)})\right\}$???

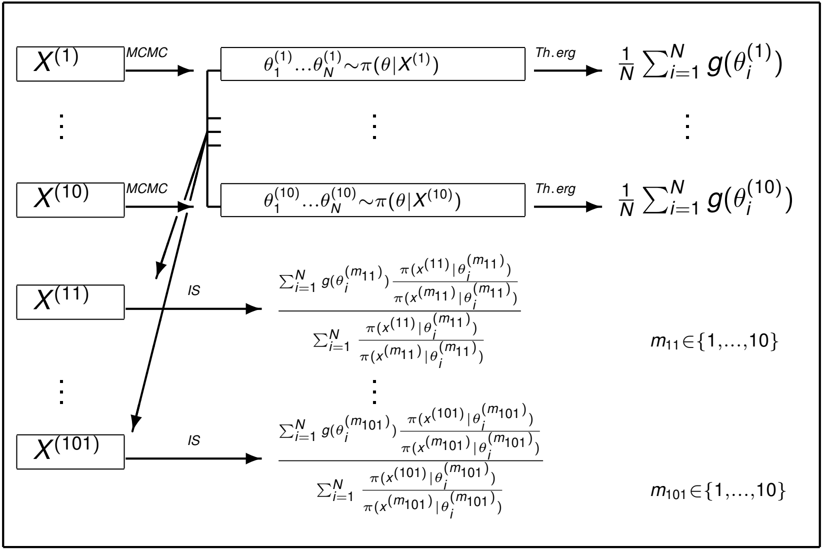
Objectif : Améliorer le choix de la fonction d'importance

$\Rightarrow$ Supports plus proches

$\Rightarrow$ Choix de la référence parmi 10 jeux de données présélectionnés pour chaque nouvelle estimation via IS

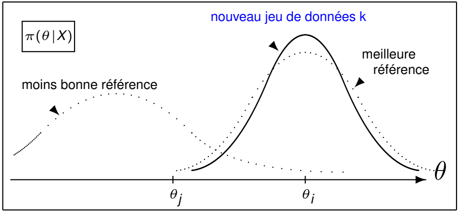
$\Rightarrow$ Étude de l'amélioration des performances par rapport à la $1^{\text{ère}}$ stratégie

$2^{\text{ème}}$ stratégie : "Référence choisie" - démarche



$2^{\text{ème}}$ stratégie : "Référence choisie" - critère 1

Critère 1 basé sur la minimisation de la norme L^1 de la différence entre deux densités a posteriori $||\pi(\theta^{(m_k)}|X^{(k)}) - \pi(\theta^{(m_k)}|X^{(m_k)})||_{L^1}$



avec $\tilde{w}_i(k, m_k) = \frac{w_i(k,m_k)}{\sum_{i=1}^N w_i(k,m_k)}$ et $w_i(k, m_k) = \frac{\pi(x^{(k)}|\theta_i^{(m_k)})}{\pi(x^{(m_k)}|\theta_i^{(m_k)})}$

$$\min_{m_k \in \{1,\ldots,M\}} \left\{ \sum_{i=1}^{N} \left| \tilde{w}_i(k, m_k) - \frac{1}{N} \right| \right\}.$$

$2^{\text{ème}}$ stratégie : "Référence choisie" - critère 2

Critère 2 basé sur la minimisation de la divergence de Kullback-Leibler entre deux densités a posteriori

$$KL(\pi(\theta|X^{(k)}), \pi(\theta|X^{(m_k)})) = \int \log\left(\frac{\pi(\theta|X^{(k)})}{\pi(\theta|X^{(m_k)})}\right)\pi(\theta|X^{(k)})d\theta$$

$\Rightarrow$ "contrôle" du comportement des queues de la distribution

avec $\tilde{w}_i(k, m_k) = \frac{w_i(k,m_k)}{\sum_{i=1}^N w_i(k,m_k)}$ et $w_i(k, m_k) = \frac{\pi(x^{(k)}|\theta_i^{(m_k)})}{\pi(x^{(m_k)}|\theta_i^{(m_k)})}$

$$\min_{m_k \in \{1,\dots,M\}} \left\{ \sum_{i=1}^N \tilde{w}_i(k, m_k) \ln w_i(k, m_k) - \frac{\ln \frac{1}{N}\sum_{i=1}^N w_i(k, m_k)}{\frac{1}{N}\sum_{i=1}^N w_i(k, m_k)} \right\}$$

$2^{\text{ème}}$ stratégie : "Référence choisie" - critère 3

Critère 3 basé sur la minimisation de la variance de l'estimateur IS

$\Rightarrow$ en pratique (Robert 2007) pour garantir que

$$E_{[\theta|X^{(m_k)}]}[(g(\theta)\frac{\pi(\theta|x^{(k)})}{\pi(\theta|x^{(m_k)})})^2] \text{ soit finie}$$

$$\Rightarrow |g(\theta)|\frac{\pi(x^{(k)}|\theta)}{\pi(x^{(m_k)}|\theta)} \text{ "le plus stable"}$$

$$\text{avec} \quad u(\theta_i^{(m_k)}) = |g(\theta_i^{(m_k)})| \cdot \pi(X^{(k)}|\theta_i^{(m_k)})/\pi(X^{(m_k)}|\theta_i^{(m_k)})$$

$$\min_{m_k \in \{1,\dots,M\}} \left\{ \frac{\max_{i=1,\dots,N} u(\theta_i^{(m_k)}) - \min_{i=1,\dots,N} u(\theta_i^{(m_k)})}{\sum_{i=1}^{N} u(\theta_i^{(m_k)})} \right\}$$

$3^{\text{ème}}$ stratégie : "Mélange de densités"

Mélange de densités a posteriori $\Rightarrow \pi_{mix}(\theta) = \frac{1}{10}\sum_{j=1}^{10} \pi(\theta|X^{(j)})$

- MCMC avec $X^{(1)}, .., X^{(10)}$ ($\leftarrow$ n=10N=nombre total d'itérations)
- Pour k=11, ...,101,
  IS pour $X^{(k)}$ avec la même fonction d'importance π_{mix}

$$\frac{1}{n}\sum_{j=1}^{10}\sum_{i=1}^{N} g(\theta_i^{(j)})\frac{\pi(\theta_i^{(j)}|X^{(k)})}{\pi_{mix}(\theta_i^{(j)})} \rightarrow E_{[\theta|X^{(k)}]}[g(\theta)]$$

avec $\left\{\theta_i^{(j)}\right\} \sim \pi(\theta|X^{(j)}), \quad j = 1, \ldots, 10$

- Problème : $\pi(X^{(j)})$ = constante de normalisation doit être connue sinon estimée (cf. Chen and Shao, 1997)

Modèles étudiés

- Modèle 1 (Modèle de Poisson) :

$$X_i|\lambda \sim \mathcal{P}(\lambda)$$

- Modèle 2 (Régression de Poisson) :

$$\begin{aligned} X_i|\lambda_i &\sim \mathcal{P}(\lambda_i) \\ log(\lambda_i) &= a + bZ_i \\ Z_i &\sim \mathcal{N}(0,1) \end{aligned}$$

- Modèle 3 (Régression de Poisson avec extravariabilité) :

$$\begin{aligned} X_i|\lambda_i &\sim \mathcal{P}(\lambda_i) \\ log(\lambda_i) &= a + \sum_{j=1}^{n} b_j Z_{ij} + \epsilon_i \end{aligned}$$

$$Z_{ij} \sim \mathcal{N}(0,1) \quad , \quad \epsilon_i \sim \mathcal{N}(0,\sigma_\epsilon^2)$$

Résultats avec le modèle 1

Simulation

$X_i|\lambda_0 \sim \mathcal{P}(\lambda_0)$ pour $\lambda_0 = 1$ ou $\lambda_0 = 20$

$X = (X_1, ..., X_N)$ pour $N = 20$

$\Rightarrow$ on réplique 101 fois : $X^{(1)}, ..., X^{(101)}$

Estimation

$X_i|\lambda \sim \mathcal{P}(\lambda)$, $\lambda \sim \mathcal{G}(\alpha, \beta)$ où $\alpha = 0.01$ et $\beta = 0.01$

$\Rightarrow$ on estime 101 espérances

$E_{[\theta|X^{(1)}]}(g(\theta)), ..., E_{[\theta|X^{(101)}]}(g(\theta))$ pour chaque fonction g

Loi a posteriori connue

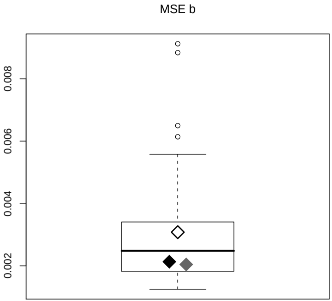
$\mathcal{G}(\sum_{i=1}^N X_i^{(k)} + \alpha, \beta + N) \Rightarrow E_{[\lambda|X^{(k)}]}[\lambda] = (\sum_{i=1}^N X_i^{(k)} + \alpha)/(\beta + N)$

Statégie		EQM ($\times 10^{05}$)
MCMC		9.1
Référence fixe	la pire	10823.4
Référence fixe	la meilleure réf. fixe	103.7
Référence choisie	critère 1 (norme L^1)	4.2
Référence choisie	critère 2 (Dv KL)	12.5
Référence choisie	critère 3 (Min var.)	4.5
Mélange †		4.0

†avec les constantes de normalisation estimées par « reverse logistic regression »

TABLE : Erreurs quadratiques moyennes entre estimation et vraie moyenne a posteriori dans le modèle de Poisson avec $\lambda = 20$.

Comparaison entre référence fixe et référence choisie



Modèle de régression de Poisson avec extravariabilité ($\sigma_\epsilon^2 = 1/2$)

Stratégie	EQM * ($\times 10^{-03}$)
La meilleure réf. "fixe"	1.2
La pire réf. "fixe"	9.1
Réf. "choisie"	
avec le critère 1	2.1
avec le critère 2	3.1
avec le critère 3	2.0

* EQM de *b* entre estimation via IS et MCMC

Conclusion

- Pour les modèles étudiés, IS donne des bonnes performances
- Importance de choix de la référence
  - référence fixe

    $\Rightarrow$ danger d'un choix maladroit
  - référence choisie

    $\Rightarrow$ amélioration des résultats
  - stratégie du mélange

    $\Rightarrow$ bons résultats

    $\Rightarrow$ fonction d'importance identique (+)

    $\Rightarrow$ estimation des constantes de normalisation (-)

Merci pour votre attention !