

Implémentation Efficace des Méthodes MCMC Utilisant Une Vraisemblance Simulée

Arnaud Doucet

Department of Statistics, University of Oxford
& Oxford-Man Institute

Collaboration avec Mike Pitt (Warwick) & Robert Kohn (UNSW)

Rennes, 26 Oct. 2012

- On veut simuler selon $\pi(d\theta) = \pi(\theta) d\theta$ tel que

$$\pi(\theta) = \frac{\gamma(\theta)}{\int \gamma(\theta') d\theta'}$$

où $\gamma : \Theta \rightarrow \mathbb{R}^+$ est connue ponctuellement mais son intégrale est inconnue.

- On veut simuler selon $\pi(d\theta) = \pi(\theta) d\theta$ tel que

$$\pi(\theta) = \frac{\gamma(\theta)}{\int \gamma(\theta') d\theta'}$$

où $\gamma : \Theta \rightarrow \mathbb{R}^+$ est connue ponctuellement mais son intégrale est inconnue.

- Inférence Bayésienne

$$\pi(\theta) = p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

où $\gamma(\theta) = p(y|\theta)p(\theta)$.

- On veut simuler selon $\pi(d\theta) = \pi(\theta) d\theta$ tel que

$$\pi(\theta) = \frac{\gamma(\theta)}{\int \gamma(\theta') d\theta'}$$

où $\gamma : \Theta \rightarrow \mathbb{R}^+$ est connue ponctuellement mais son intégrale est inconnue.

- Inférence Bayésienne

$$\pi(\theta) = p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

où $\gamma(\theta) = p(y|\theta)p(\theta)$.

- Les méthodes MCMC sont utilisées depuis 60 ans pour résoudre ce type de problèmes.

Algorithme de Metropolis-Hastings

- **Idée de Base:** Construire une chaîne de Markov ergodique $\{\theta_j\}$ de loi invariante $\pi(\theta)$.

Algorithme de Metropolis-Hastings

- **Idée de Base:** Construire une chaîne de Markov ergodique $\{\theta_j\}$ de loi invariante $\pi(\theta)$.
- **Algorithme de Metropolis-Hastings**

Algorithme de Metropolis-Hastings

- **Idée de Base:** Construire une chaîne de Markov ergodique $\{\theta_j\}$ de loi invariante $\pi(\theta)$.
- **Algorithme de Metropolis-Hastings**
- ① Etant donné ϑ_j , simuler $\vartheta' \sim q(\cdot | \vartheta_j)$.

Algorithme de Metropolis-Hastings

- **Idée de Base:** Construire une chaîne de Markov ergodique $\{\theta_j\}$ de loi invariante $\pi(\theta)$.
- **Algorithme de Metropolis-Hastings**

① Etant donné ϑ_j , simuler $\vartheta' \sim q(\cdot | \vartheta_j)$.

② Avec probabilité

$$1 \wedge \frac{\pi(\vartheta')}{\pi(\vartheta_j)} \frac{q(\vartheta_j | \vartheta')}{q(\vartheta' | \vartheta_j)}$$

$\vartheta_{j+1} = \vartheta'$, sinon $\vartheta_{j+1} = \vartheta_j$.

Algorithme de Metropolis-Hastings

- **Idée de Base:** Construire une chaîne de Markov ergodique $\{\theta_j\}$ de loi invariante $\pi(\theta)$.

- **Algorithme de Metropolis-Hastings**

- 1 Etant donné ϑ_j , simuler $\vartheta' \sim q(\cdot | \vartheta_j)$.

- 2 Avec probabilité

$$1 \wedge \frac{\pi(\vartheta') q(\vartheta_j | \vartheta')}{\pi(\vartheta_j) q(\vartheta' | \vartheta_j)}$$

$\vartheta_{j+1} = \vartheta'$, sinon $\vartheta_{j+1} = \vartheta_j$.

- Quand $\pi(\theta) = p(\theta | y)$, alors

$$\frac{\pi(\vartheta') q(\vartheta_j | \vartheta')}{\pi(\vartheta_j) q(\vartheta' | \vartheta_j)} = \frac{p(y | \vartheta') p(\vartheta') q(\vartheta_j | \vartheta')}{p(y | \vartheta_j) p(\vartheta_j) q(\vartheta' | \vartheta_j)}.$$

- Considérons maintenant le scénario où $p(y|\theta)$ ne peut pas être calculé; e.g.

$$p(y|\theta) = \int p(x, y|\theta) dx.$$

- Considérons maintenant le scénario où $p(y|\theta)$ ne peut pas être calculé; e.g.

$$p(y|\theta) = \int p(x, y|\theta) dx.$$

- Les approches MCMC classiques consistent à simuler selon

$$p(\theta, x|y) = \frac{p(x, y|\theta) p(\theta)}{p(y)}$$

en mettant à jour X conditionnellement à θ et θ conditionnellement à X .

- Considérons maintenant le scénario où $p(y|\theta)$ ne peut pas être calculé; e.g.

$$p(y|\theta) = \int p(x, y|\theta) dx.$$

- Les approches MCMC classiques consistent à simuler selon

$$p(\theta, x|y) = \frac{p(x, y|\theta) p(\theta)}{p(y)}$$

en mettant à jour X conditionnellement à θ et θ conditionnellement à X .

- Une alternative proposée récemment se base sur une approximation de $p(y|\theta)$.

- **Algorithme de Metropolis-Hastings**

- **Algorithme de Metropolis-Hastings**

- 1 Etant donné $(\vartheta_j, \hat{p}(y|\vartheta_j))$, simuler $\vartheta' \sim q(\cdot|\vartheta_j)$.

- **Algorithme de Metropolis-Hastings**

- 1 Etant donné $(\vartheta_j, \hat{p}(y|\vartheta_j))$, simuler $\vartheta' \sim q(\cdot|\vartheta_j)$.
- 2 Calculer une approximation $\hat{p}(y|\vartheta')$ de $p(y|\vartheta')$.

• Algorithme de Metropolis-Hastings

- 1 Etant donné $(\vartheta_j, \hat{p}(y|\vartheta_j))$, simuler $\vartheta' \sim q(\cdot|\vartheta_j)$.
- 2 Calculer une approximation $\hat{p}(y|\vartheta')$ de $p(y|\vartheta')$.
- 3 Avec probabilité

$$1 \wedge \frac{\hat{p}(y|\vartheta') p(\vartheta') q(\vartheta_j|\vartheta')}{\hat{p}(y|\vartheta_j) p(\vartheta_j) q(\vartheta'|\vartheta_j)}$$

$\vartheta_{j+1} = \vartheta'$, sinon $\vartheta_{j+1} = \vartheta_j$.

- **Algorithme de Metropolis-Hastings**

- 1 Etant donné $(\vartheta_j, \hat{p}(y|\vartheta_j))$, simuler $\vartheta' \sim q(\cdot|\vartheta_j)$.
- 2 Calculer une approximation $\hat{p}(y|\vartheta')$ de $p(y|\vartheta')$.
- 3 Avec probabilité

$$1 \wedge \frac{\hat{p}(y|\vartheta') p(\vartheta') q(\vartheta_j|\vartheta')}{\hat{p}(y|\vartheta_j) p(\vartheta_j) q(\vartheta'|\vartheta_j)}$$

$\vartheta_{j+1} = \vartheta'$, sinon $\vartheta_{j+1} = \vartheta_j$.

- **Proposition.** Si $\hat{p}(y|\theta)$ est un estimateur non-biaisé et non-négatif de $p(y|\theta)$ alors $p(\theta|y)$ est une marginale de la loi invariante associée au noyau de transition de $\{\vartheta_j, \hat{p}(y|\vartheta_j)\}$.

- Supposons que

$$p(y|\theta) = \int p(x, y|\theta) dx.$$

Echantillonnage d'Importance

- Supposons que

$$p(y|\theta) = \int p(x, y|\theta) dx.$$

- Soit $g(x|y, \theta)$ une loi d'importance alors

$$p(y|\theta) = \int \omega(x, \theta) g(x|y; \theta) dx.$$

où

$$\omega(x, \theta) = \frac{p(x, y|\theta)}{g(x|y; \theta)}.$$

- Supposons que

$$p(y|\theta) = \int p(x, y|\theta) dx.$$

- Soit $g(x|y, \theta)$ une loi d'importance alors

$$p(y|\theta) = \int \omega(x, \theta) g(x|y; \theta) dx.$$

où

$$\omega(x, \theta) = \frac{p(x, y|\theta)}{g(x|y; \theta)}.$$

- L'estimateur

$$\hat{p}(y|\theta) = \frac{1}{N} \sum_{k=1}^N \omega(X^k, \theta),$$

où $X^k \sim g(\cdot|y; \theta)$ est non-biaisé.

- Soit $\{X_t\}_{t \geq 0}$ un processus de Markov latent avec $X_0 \sim \mu(\cdot | \theta)$ et $X_{t+1} | (X_t = x) \sim f(\cdot | x_t, \theta)$. Les observations $\{y_t\}_{t \geq 1}$ sont supposées indépendantes conditionnellement de loi $Y_t | (X_t = x) \sim g(\cdot | x_t, \theta)$

$$p(y|\theta) = \int p(y|x, \theta) p(x|\theta) dx$$

où, en utilisant la notation $x := x_{0:T} := (x'_0, \dots, x'_T)'$

$$p(x|\theta) = \mu(x_0|\theta) \prod_{t=1}^T f(x_t|x_{t-1}, \theta),$$

$$p(y|x, \theta) = \prod_{t=1}^T g(y_t|x_t, \theta).$$

- Soit $\{X_t\}_{t \geq 0}$ un processus de Markov latent avec $X_0 \sim \mu(\cdot | \theta)$ et $X_{t+1} | (X_t = x) \sim f(\cdot | x_t, \theta)$. Les observations $\{y_t\}_{t \geq 1}$ sont supposées indépendantes conditionnellement de loi $Y_t | (X_t = x) \sim g(\cdot | x_t, \theta)$

$$p(y|\theta) = \int p(y|x, \theta) p(x|\theta) dx$$

où, en utilisant la notation $x := x_{0:T} := (x'_0, \dots, x'_T)'$

$$p(x|\theta) = \mu(x_0|\theta) \prod_{t=1}^T f(x_t|x_{t-1}, \theta),$$

$$p(y|x, \theta) = \prod_{t=1}^T g(y_t|x_t, \theta).$$

- Les méthodes de Monte Carlo séquentielles fournissent un estimateur non biaisé de $p(y|\theta)$ de variance relative $O(T/N)$ (Del Moral, 1996; Cérou, Del Moral, Guyader, 2010).

- Deux espèces X_t^1 (proie) et X_t^2 (prédateur)

$$\Pr \left(X_{t+dt}^1 = x_t^1 + 1, X_{t+dt}^2 = x_t^2 \mid x_t^1, x_t^2 \right) = \alpha x_t^1 dt + o(dt),$$

$$\Pr \left(X_{t+dt}^1 = x_t^1 - 1, X_{t+dt}^2 = x_t^2 + 1 \mid x_t^1, x_t^2 \right) = \beta x_t^1 x_t^2 dt + o(dt),$$

$$\Pr \left(X_{t+dt}^1 = x_t^1, X_{t+dt}^2 = x_t^2 - 1 \mid x_t^1, x_t^2 \right) = \gamma x_t^2 dt + o(dt),$$

observées a temps discret

$$Y_n = X_{n\Delta}^1 + W_n \text{ où } W_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2).$$

- Deux espèces X_t^1 (proie) et X_t^2 (prédateur)

$$\Pr \left(X_{t+dt}^1 = x_t^1 + 1, X_{t+dt}^2 = x_t^2 \mid x_t^1, x_t^2 \right) = \alpha x_t^1 dt + o(dt),$$

$$\Pr \left(X_{t+dt}^1 = x_t^1 - 1, X_{t+dt}^2 = x_t^2 + 1 \mid x_t^1, x_t^2 \right) = \beta x_t^1 x_t^2 dt + o(dt),$$

$$\Pr \left(X_{t+dt}^1 = x_t^1, X_{t+dt}^2 = x_t^2 - 1 \mid x_t^1, x_t^2 \right) = \gamma x_t^2 dt + o(dt),$$

observées a temps discret

$$Y_n = X_{n\Delta}^1 + W_n \text{ où } W_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2).$$

- On s'intéresse a $\theta = (\alpha, \beta, \gamma)$ de loi a priori (Boys et al., 2008; Kunsch 2011)

$$\alpha \sim \mathcal{G}(1, 10), \quad \beta \sim \mathcal{G}(1, 0.25), \quad \gamma \sim \mathcal{G}(1, 7.5).$$

- Deux espèces X_t^1 (proie) et X_t^2 (prédateur)

$$\Pr \left(X_{t+dt}^1 = x_t^1 + 1, X_{t+dt}^2 = x_t^2 \mid x_t^1, x_t^2 \right) = \alpha x_t^1 dt + o(dt),$$

$$\Pr \left(X_{t+dt}^1 = x_t^1 - 1, X_{t+dt}^2 = x_t^2 + 1 \mid x_t^1, x_t^2 \right) = \beta x_t^1 x_t^2 dt + o(dt),$$

$$\Pr \left(X_{t+dt}^1 = x_t^1, X_{t+dt}^2 = x_t^2 - 1 \mid x_t^1, x_t^2 \right) = \gamma x_t^2 dt + o(dt),$$

observées a temps discret

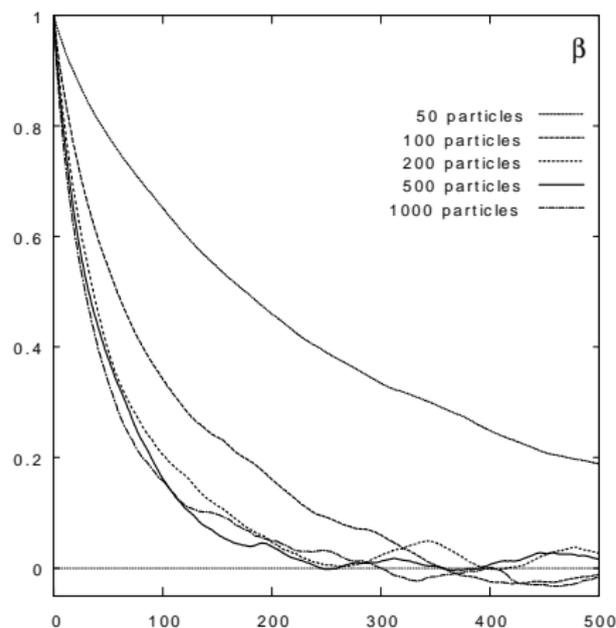
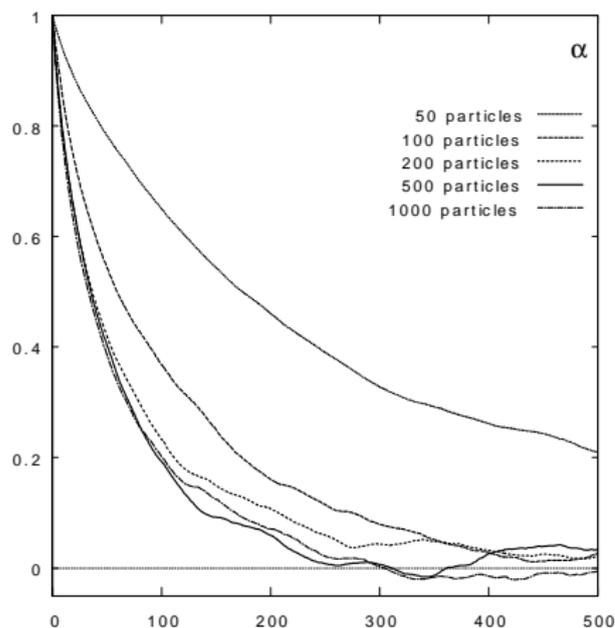
$$Y_n = X_{n\Delta}^1 + W_n \text{ où } W_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2).$$

- On s'intéresse à $\theta = (\alpha, \beta, \gamma)$ de loi a priori (Boys et al., 2008; Kunsch 2011)

$$\alpha \sim \mathcal{G}(1, 10), \quad \beta \sim \mathcal{G}(1, 0.25), \quad \gamma \sim \mathcal{G}(1, 7.5).$$

- On simule selon $p(\theta \mid y_{1:T})$ en utilisant Metropolis-Hastings où $p(y_{1:T} \mid \theta)$ est estimée en utilisant un filtre particulière.

Fonctions d'Autocorrélation



Autocorrélation de $\{\alpha_j\}$ (gauche) et $\{\beta_j\}$ (droite) pour différents N .

Comment sélectionner le nombre d'échantillons

- Si N est trop petit, alors la chaîne de Markov ne mélange pas rapidement et de très nombreuses itérations sont requises.

Comment sélectionner le nombre d'échantillons

- Si N est trop petit, alors la chaîne de Markov ne mélange pas rapidement et de très nombreuses itérations sont requises.
- Si N est trop large, chaque itération de MCMC est très chère.

Comment sélectionner le nombre d'échantillons

- Si N est trop petit, alors la chaîne de Markov ne mélange pas rapidement et de très nombreuses itérations sont requises.
- Si N est trop large, chaque itération de MCMC est très chère.
- **But:** Nous voudrions savoir comment sélectionner N de manière optimale.

MCMC avec Vraisemblance Intractable

- Soit $Z = \log \hat{p}_N(y|\vartheta) - \log p(y|\vartheta)$ l'erreur d'approximation pour la log-vraisemblance.

MCMC avec Vraisemblance Intractable

- Soit $Z = \log \hat{p}_N(y|\vartheta) - \log p(y|\vartheta)$ l'erreur d'approximation pour la log-vraisemblance.
- La loi de proposition pour Z est notée $g_N(z|\theta)$.

MCMC avec Vraisemblance Intractable

- Soit $Z = \log \hat{p}_N(y|\theta) - \log p(y|\theta)$ l'erreur d'approximation pour la log-vraisemblance.
- La loi de proposition pour Z est notée $g_N(z|\theta)$.
- La loi cible selon laquelle on simule est

$$\hat{\pi}_N(\theta, z) = \pi(\theta) \exp(z) g_N(z|\theta)$$

où la propriété de non-biais de $\hat{p}_N(y|\theta)$ implique

$$\int \exp(z) g_N(z|\theta) dz = 1$$

MCMC avec Vraisemblance Intractable

- Soit $Z = \log \hat{p}_N(y|\vartheta) - \log p(y|\vartheta)$ l'erreur d'approximation pour la log-vraisemblance.
- La loi de proposition pour Z est notée $g_N(z|\theta)$.
- La loi cible selon laquelle on simule est

$$\hat{\pi}_N(\theta, z) = \pi(\theta) \exp(z) g_N(z|\theta)$$

où la propriété de non-biais de $\hat{p}_N(y|\theta)$ implique

$$\int \exp(z) g_N(z|\theta) dz = 1$$

- L'algorithme précédent est $\vartheta' \sim q(\cdot|\vartheta)$, $Z' \sim g_N(z|\vartheta')$, accepte (ϑ', Z') avec proba

$$\alpha_Q(\vartheta, Z; \vartheta', Z') = \min \{1, \exp(Z' - Z) \omega(\vartheta'; \vartheta) / \omega(\vartheta; \vartheta')\},$$

où $\omega(\vartheta'; \vartheta) = \pi(\vartheta') / q(\vartheta'|\vartheta)$.

- Considérons une chaîne de Markov stationnaire $\{\theta_j\}$ de loi invariante $\pi(\theta)$ et $h : \Theta \rightarrow \mathbb{R}$ où $\mathbb{V}_\pi [h(\theta)] < \infty$. Définissons

$$\mu_h = \mathbb{E}_\pi [h(\theta)] \quad \text{et} \quad \hat{\mu}_{h,n} = n^{-1} \sum_{j=1}^n h(\vartheta_j).$$

alors

$$\lim_{n \rightarrow \infty} n \mathbb{V}_\pi (\hat{\mu}_{h,n}) = \mathbb{V}_\pi [h(\theta)] IF_h \quad \text{où} \quad IF_h = 1 + 2 \sum_{\tau=1}^{\infty} \rho_h(\tau),$$

où $\rho_h(\tau) = \mathbb{E}_\pi [h(\vartheta_0)h(\vartheta_\tau)]$.

- Considérons une chaîne de Markov stationnaire $\{\theta_j\}$ de loi invariante $\pi(\theta)$ et $h : \Theta \rightarrow \mathbb{R}$ où $\mathbb{V}_\pi [h(\theta)] < \infty$. Définissons

$$\mu_h = \mathbb{E}_\pi [h(\theta)] \quad \text{et} \quad \hat{\mu}_{h,n} = n^{-1} \sum_{j=1}^n h(\vartheta_j).$$

alors

$$\lim_{n \rightarrow \infty} n \mathbb{V}_\pi (\hat{\mu}_{h,n}) = \mathbb{V}_\pi [h(\theta)] IF_h \quad \text{où} \quad IF_h = 1 + 2 \sum_{\tau=1}^{\infty} \rho_h(\tau),$$

où $\rho_h(\tau) = \mathbb{E}_\pi [h(\vartheta_0)h(\vartheta_\tau)]$.

- Il est raisonnable de comparer différentes chaînes en se basant sur leurs inefficacités.

- Introduisons une chaîne alternative Q^* .

Une Chaîne Dominante

- Introduisons une chaîne alternative Q^* .
- Soit (ϑ, Z) l'état de la chaîne de Markov.

Une Chaîne Dominante

- Introduisons une chaîne alternative Q^* .
- Soit (ϑ, Z) l'état de la chaîne de Markov.
- 1 Simuler $\vartheta' \sim q(\cdot|\vartheta)$ et $Z' \sim g_N(\cdot|\vartheta')$.

Une Chaîne Dominante

- Introduisons une chaîne alternative Q^* .
 - Soit (ϑ, Z) l'état de la chaîne de Markov.
- 1 Simuler $\vartheta' \sim q(\cdot|\vartheta)$ et $Z' \sim g_N(\cdot|\vartheta')$.
 - 2 Accepter ϑ' avec proba. $\alpha_{Q^{\text{EX}}}(\vartheta; \vartheta') = \min \{1, \omega(\vartheta'; \vartheta) / \omega(\vartheta; \vartheta')\}$.

Une Chaîne Dominante

- Introduisons une chaîne alternative Q^* .
 - Soit (ϑ, Z) l'état de la chaîne de Markov.
- 1 Simuler $\vartheta' \sim q(\cdot|\vartheta)$ et $Z' \sim g_N(\cdot|\vartheta')$.
 - 2 Accepter ϑ' avec proba. $\alpha_{Q^{\text{EX}}}(\vartheta; \vartheta') = \min \{1, \omega(\vartheta'; \vartheta) / \omega(\vartheta; \vartheta')\}$.
 - 3 Accepter Z' avec proba. $\alpha_{Q^{\text{Z}}}(Z; Z') = \min \{1, \exp(Z' - Z)\}$.

Une Chaîne Dominante

- Introduisons une chaîne alternative Q^* .
 - Soit (ϑ, Z) l'état de la chaîne de Markov.
- 1 Simuler $\vartheta' \sim q(\cdot|\vartheta)$ et $Z' \sim g_N(\cdot|\vartheta')$.
 - 2 Accepter ϑ' avec proba. $\alpha_{Q^{\text{EX}}}(\vartheta; \vartheta') = \min \{1, \omega(\vartheta'; \vartheta) / \omega(\vartheta; \vartheta')\}$.
 - 3 Accepter Z' avec proba. $\alpha_{Q^{\text{Z}}}(Z; Z') = \min \{1, \exp(Z' - Z)\}$.
 - 4 (ϑ', Z') est accepté si et seulement si il y a acceptation dans les étapes (2) & (3).

- **Lemme:** La chaîne de Markov Q^* a les propriétés suivantes:

- **Lemme:** La chaîne de Markov Q^* a les propriétés suivantes:

- 1 On a

$$\alpha_Q(\theta, z; \theta', z') \geq \alpha_{Q^*}(\theta, z; \theta', z') = \alpha_{Q^{\text{EX}}}(\theta; \theta') \times \alpha_{Q^z}(z; z').$$

- **Lemme:** La chaîne de Markov Q^* a les propriétés suivantes:

- 1 On a

$$\alpha_Q(\theta, z; \theta', z') \geq \alpha_{Q^*}(\theta, z; \theta', z') = \alpha_{Q^{\text{EX}}}(\theta; \theta') \times \alpha_{Q^z}(z; z').$$

- 2 Q^* est une chaîne de Markov réversible de loi invariante $\hat{\pi}_N(\theta, z)$.

- **Lemme:** La chaîne de Markov Q^* a les propriétés suivantes:

- 1 On a

$$\alpha_Q(\theta, z; \theta', z') \geq \alpha_{Q^*}(\theta, z; \theta', z') = \alpha_{Q^{\text{EX}}}(\theta; \theta') \times \alpha_{Q^z}(z; z').$$

- 2 Q^* est une chaîne de Markov réversible de loi invariante $\hat{\pi}_N(\theta, z)$.
- 3 Pour toute fonction $h(\theta)$, on a $IF_h^{Q^*} \geq IF_h^Q$ (Peskun, 1973; Tierney 1998).

Hypothèses sur L'Estimateur de la Log-Vraisemblance

- **Hypothèse.** Soit $Z = \log \hat{p}_N (y|\vartheta) - \log p(y|\vartheta)$.

Hypothèses sur L'Estimateur de la Log-Vraisemblance

• **Hypothèse.** Soit $Z = \log \hat{p}_N(y|\vartheta) - \log p(y|\vartheta)$.

① On a

$$g_N(z|\theta) = \phi(z; -\gamma^2(\theta)/2N, \gamma^2(\theta)/N),$$
$$\pi_N(z|\theta) = \exp(z)g_N(z|\theta) = \phi(z; \gamma^2(\theta)/2N, \gamma^2(\theta)/N)$$

où $\phi(z; a, b^2)$ est la loi normale de moyenne a , variance b^2 .

Hypothèses sur L'Estimateur de la Log-Vraisemblance

• **Hypothèse.** Soit $Z = \log \hat{p}_N(y|\vartheta) - \log p(y|\vartheta)$.

① On a

$$\begin{aligned}g_N(z|\theta) &= \phi(z; -\gamma^2(\theta)/2N, \gamma^2(\theta)/N), \\ \pi_N(z|\theta) &= \exp(z)g_N(z|\theta) = \phi(z; \gamma^2(\theta)/2N, \gamma^2(\theta)/N)\end{aligned}$$

où $\phi(z; a, b^2)$ est la loi normale de moyenne a , variance b^2 .

② Pour une valeur donnée de σ^2 , on a $N = N_{\sigma^2}(\theta) = \gamma(\theta)^2/\sigma^2$.

Hypothèses sur L'Estimateur de la Log-Vraisemblance

- **Hypothèse.** Soit $Z = \log \hat{p}_N(y|\vartheta) - \log p(y|\vartheta)$.

- 1 On a

$$g_N(z|\theta) = \phi(z; -\gamma^2(\theta)/2N, \gamma^2(\theta)/N),$$
$$\pi_N(z|\theta) = \exp(z)g_N(z|\theta) = \phi(z; \gamma^2(\theta)/2N, \gamma^2(\theta)/N)$$

où $\phi(z; a, b^2)$ est la loi normale de moyenne a , variance b^2 .

- 2 Pour une valeur donnée de σ^2 , on a $N = N_{\sigma^2}(\theta) = \gamma(\theta)^2/\sigma^2$.
- Sous cette hypothèse, $g_N(z|\theta)$ et $\pi_N(z|\theta)$ sont seulement fonctions de σ^2

$$g(z|\sigma^2) = \phi(z; -\sigma^2/2, \sigma^2), \quad \pi(z|\sigma^2) = \phi(z; \sigma^2/2, \sigma^2).$$

et ϑ et Z sont indépendants sous $\hat{\pi}_N(\theta, z)$.

Loi Empirique vs Loi Asymptotique pour l'Estimateur

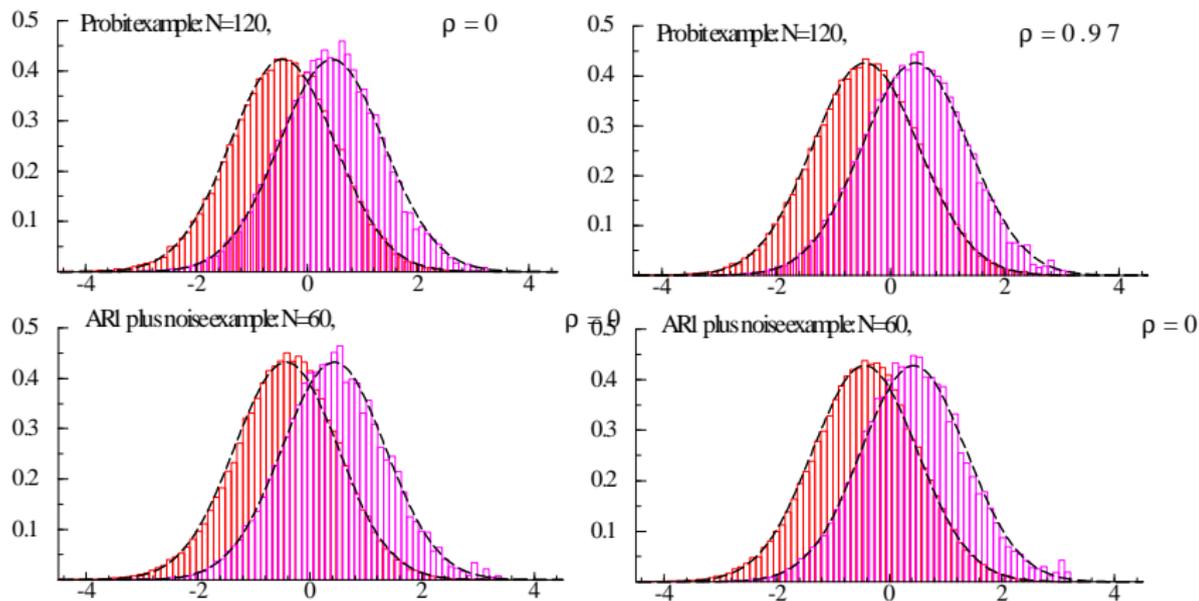


Figure: Histogrammes des valeurs proposées (rouge) et acceptées (roses) de Z . Les approximations Gaussiennes sont surimposées.

- **Lemme.** Sous l'hypothèse précédente, la chaîne Q^* utilisant la loi de proposition $q(\theta'|\theta) = \pi(\theta')$, notée Q^Z , vérifie

$$IF^Z(\sigma^2) = \int \frac{1 + p^*(w, \sigma)}{1 - p^*(w, \sigma)} \phi(w) dw,$$

où $p^*(w, \sigma) = \Phi(w + \sigma) - \exp(-w\sigma - \sigma^2/2)\Phi(w)$.

- **Lemme.** Sous l'hypothèse précédente, la chaîne Q^* utilisant la loi de proposition $q(\theta'|\theta) = \pi(\theta')$, notée Q^Z , vérifie

$$IF^Z(\sigma^2) = \int \frac{1 + p^*(w, \sigma)}{1 - p^*(w, \sigma)} \phi(w) dw,$$

où $p^*(w, \sigma) = \Phi(w + \sigma) - \exp(-w\sigma - \sigma^2/2)\Phi(w)$.

- En particulier, $IF^Z(\sigma^2)$ est indépendant de h .

- **Proposition.** Sous l'hypothèse précédente, on a

$$IF_h^Q(\sigma^2) \leq IF_h^{Q^*}(\sigma^2) \leq IF_h^U(\sigma^2),$$

$$IF_h^U(\sigma^2) = \frac{1}{2}(1 + IF_h^{EX})(1 + IF^Z(\sigma^2)) - 1.$$

- **Proposition.** Sous l'hypothèse précédente, on a

$$IF_h^Q(\sigma^2) \leq IF_h^{Q^*}(\sigma^2) \leq IF_h^U(\sigma^2),$$

$$IF_h^U(\sigma^2) = \frac{1}{2}(1 + IF_h^{\text{EX}})(1 + IF^Z(\sigma^2)) - 1.$$

- L'inégalité devient égalité quand $\sigma \rightarrow 0$ (car $IF^Z(\sigma^2) \rightarrow 1$ pour $\sigma \rightarrow 0$) et quand la proposition est parfaite (car $IF_h^{\text{EX}} = 1$ quand $q(\theta'|\theta) = \pi(\theta')$).

- **Proposition.** Sous l'hypothèse précédente, on a

$$IF_h^Q(\sigma^2) \leq IF_h^{Q^*}(\sigma^2) \leq IF_h^U(\sigma^2),$$

$$IF_h^U(\sigma^2) = \frac{1}{2}(1 + IF_h^{\text{EX}})(1 + IF^Z(\sigma^2)) - 1.$$

- L'inégalité devient égalité quand $\sigma \rightarrow 0$ (car $IF^Z(\sigma^2) \rightarrow 1$ pour $\sigma \rightarrow 0$) et quand la proposition est parfaite (car $IF_h^{\text{EX}} = 1$ quand $q(\theta'|\theta) = \pi(\theta')$).
- Pour la probabilité d'acceptation inconditionnelle, on a

$$P^Q(A|\sigma^2) \geq P^{Q^*}(A|\sigma^2) = 2\Phi(-\sigma/\sqrt{2}) P^{\text{EX}}(A).$$

- Le Computing Time (CT) pour Q est défini comme

$$CT_h^Q(\sigma^2) = IF_h^Q(\sigma^2) / \sigma^2$$

et satisfait

- Le Computing Time (CT) pour Q est défini comme

$$CT_h^Q(\sigma^2) = IF_h^Q(\sigma^2) / \sigma^2$$

et satisfait

- 1 $CT_h^Q(\sigma^2) \leq CT_h^U(\sigma^2)$ où $CT_h^U(\sigma^2) := IF_h^U(\sigma^2) / \sigma^2$.

- Le Computing Time (CT) pour Q est défini comme

$$CT_h^Q(\sigma^2) = IF_h^Q(\sigma^2) / \sigma^2$$

et satisfait

- $CT_h^Q(\sigma^2) \leq CT_h^U(\sigma^2)$ où $CT_h^U(\sigma^2) := IF_h^U(\sigma^2) / \sigma^2$.
- Si $IF_h^{\text{EX}} = 1$, alors $CT_h^U(\sigma^2)$ est minimisé à $\sigma_{opt}^U = 0.92$ et $IF^Z(\sigma_{opt}^U) = 4.54$, $P^Z(A|\sigma_{opt}^U) = 0.5153$.

- Le Computing Time (CT) pour Q est défini comme

$$CT_h^Q(\sigma^2) = IF_h^Q(\sigma^2) / \sigma^2$$

et satisfait

- $CT_h^Q(\sigma^2) \leq CT_h^U(\sigma^2)$ où $CT_h^U(\sigma^2) := IF_h^U(\sigma^2) / \sigma^2$.
- Si $IF_h^{\text{EX}} = 1$, alors $CT_h^U(\sigma^2)$ est minimisé à $\sigma_{opt}^U = 0.92$ et $IF^Z(\sigma_{opt}^U) = 4.54$, $P^Z(A|\sigma_{opt}^U) = 0.5153$.
- La valeur σ_{opt}^U augmente avec IF_h^{EX} et $\sigma_{opt}^U = 1.0206$ quand $IF_h^{\text{EX}} \rightarrow \infty$.

- Le Computing Time (CT) pour Q est défini comme

$$CT_h^Q(\sigma^2) = IF_h^Q(\sigma^2) / \sigma^2$$

et satisfait

- $CT_h^Q(\sigma^2) \leq CT_h^U(\sigma^2)$ où $CT_h^U(\sigma^2) := IF_h^U(\sigma^2) / \sigma^2$.
 - Si $IF_h^{\text{EX}} = 1$, alors $CT_h^U(\sigma^2)$ est minimisé à $\sigma_{opt}^U = 0.92$ et $IF^Z(\sigma_{opt}^U) = 4.54$, $P^Z(A|\sigma_{opt}^U) = 0.5153$.
 - La valeur σ_{opt}^U augmente avec IF_h^{EX} et $\sigma_{opt}^U = 1.0206$ quand $IF_h^{\text{EX}} \rightarrow \infty$.
- Le Relative Computing Time (RCT) pour la borne d'inefficacité $IF_h^U(\sigma^2)$ est défini comme

$$RCT_h^U(\sigma^2) = \frac{IF_h^U(\sigma^2)}{IF_h^{\text{EX}} \times \sigma^2}.$$

Relative Computing Time

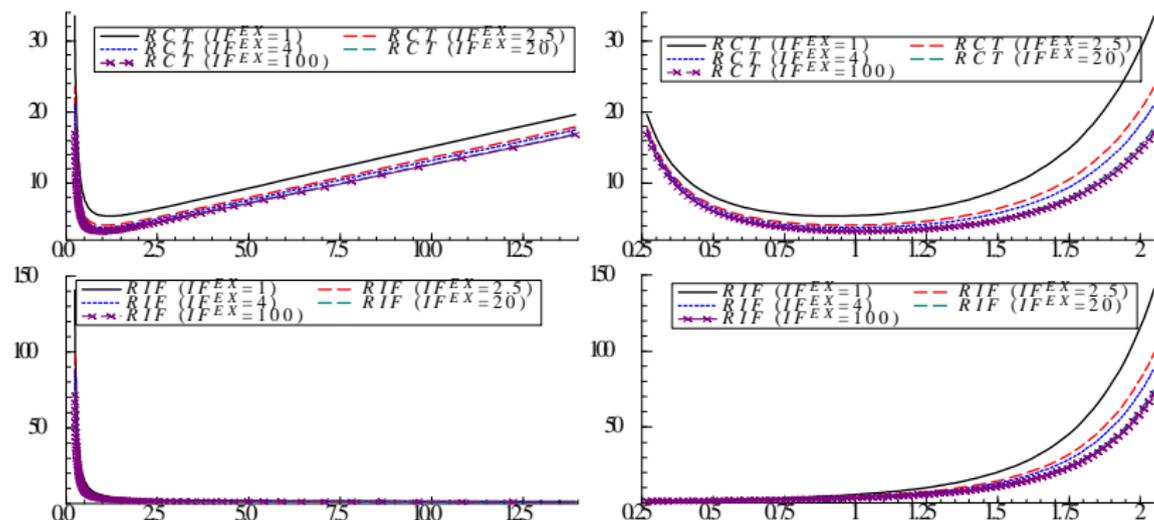


Figure: RCT_h^U (haut) et RIF_h^U (bas) vs $1/\sigma^2$ (gauche) et σ (droite).

Exemple 1: Modele Probit

- On utilise un simple modèle Probit, où pour $t = 1, \dots, T$

$$Y_t = \mathbb{I}(X_t > 0), \quad X_t \stackrel{iid}{\sim} N(\theta; 1).$$

Exemple 1: Modele Probit

- On utilise un simple modèle Probit, où pour $t = 1, \dots, T$

$$Y_t = \mathbb{I}(X_t > 0), \quad X_t \stackrel{iid}{\sim} N(\theta; 1).$$

- La vraisemblance est connue car $\Pr(y_t = 1 | \theta) = \Phi(\theta)$ mais est estimée

$$\hat{\Pr}(Y_t = 1) = \frac{1}{N} \sum_{k=1}^N \mathbb{I}(X_t^{(k)} > 0), \quad X_t^{(k)} \stackrel{iid}{\sim} N(\theta; 1)$$

et $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$ avec $\sigma_\theta^2 \gg 1$.

Exemple 1: Modele Probit

- On utilise un simple modèle Probit, où pour $t = 1, \dots, T$

$$Y_t = \mathbb{I}(X_t > 0), \quad X_t \stackrel{iid}{\sim} N(\theta; 1).$$

- La vraisemblance est connue car $\Pr(y_t = 1 | \theta) = \Phi(\theta)$ mais est estimée

$$\widehat{\Pr}(Y_t = 1) = \frac{1}{N} \sum_{k=1}^N \mathbb{I}(X_t^{(k)} > 0), \quad X_t^{(k)} \stackrel{iid}{\sim} N(\theta; 1)$$

et $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$ avec $\sigma_\theta^2 \gg 1$.

- Loi de proposition autorégressive

$$\theta' = \widehat{\theta} + \rho(\theta - \widehat{\theta}) + \sqrt{\frac{\widehat{\sigma}^2(1 - \rho^2)}{\nu - 2}} t_5,$$

où $\widehat{\theta}$ est le mode a posteriori. On considère $\rho \in \{0, 0.4, 0.6, 0.9, 0.97\}$.

Probabilités d'Acceptation

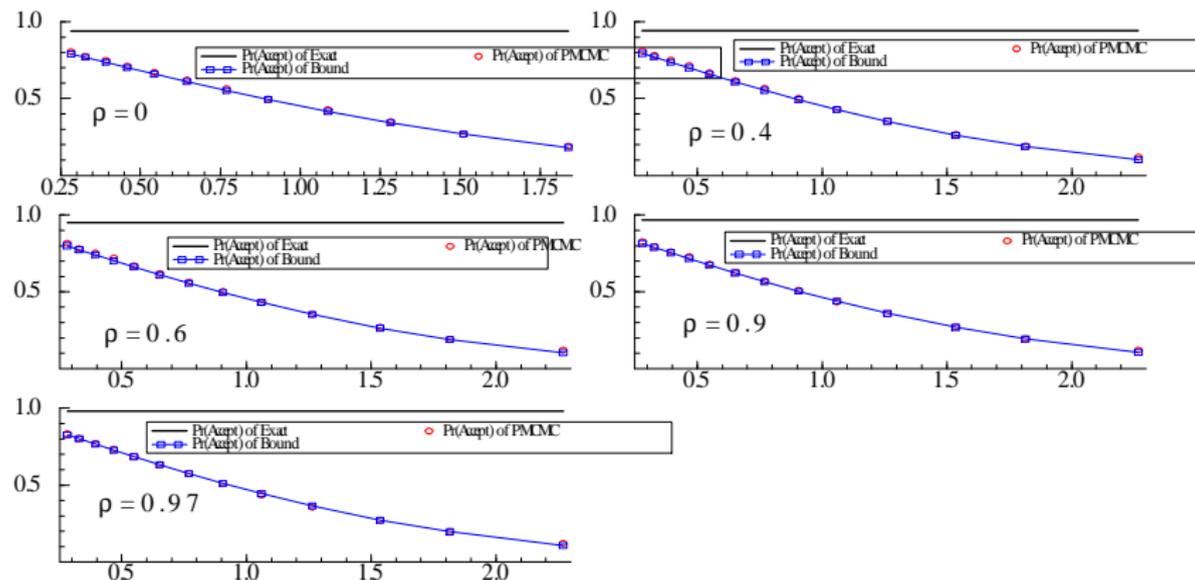


Figure: $T = 100$, $\theta = 0.5$. Proba accept. vs $\sigma(\bar{\theta})$. Proba estimées pour le MCMC exact (constant), proba pour Q (red) et borne inferieure (blue).

Relative Computing Time

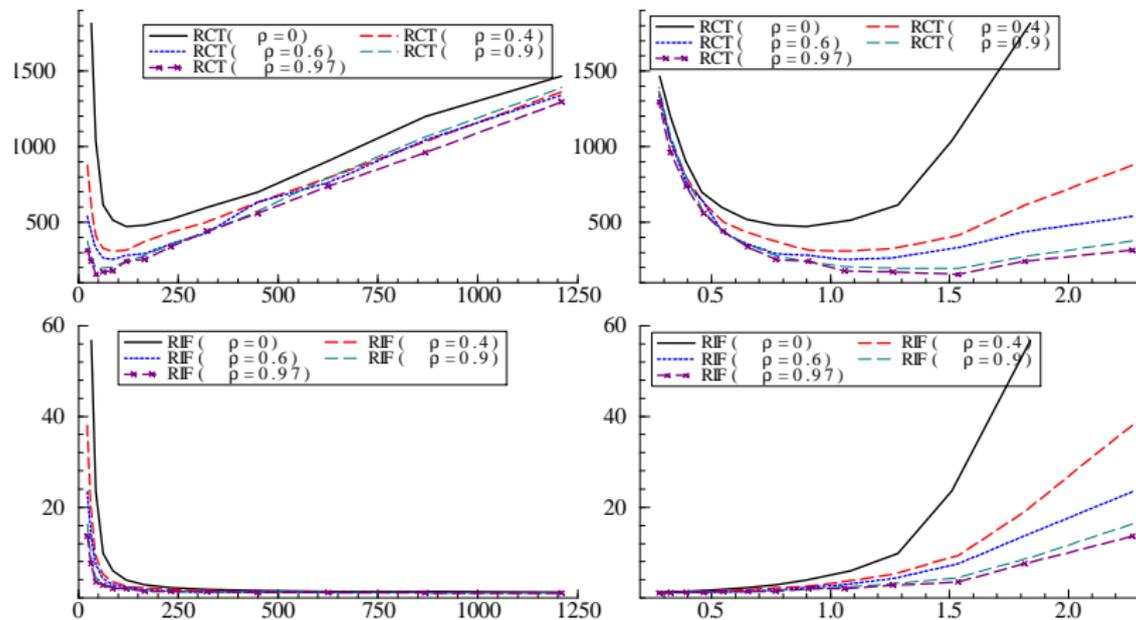


Figure: RCT_h^Q (haut) et RIF_h^Q (bas) vs N (gauche) et $\sigma(\bar{\theta})$ (droite).

Exemple 2: Modèle Autorégressif

- On a

$$Y_t = X_t + \sigma_\varepsilon \varepsilon_t \quad \text{et} \quad X_{t+1} = \mu(1 - \phi) + \phi X_t + \sigma_\eta \eta_t$$

où ε_t et η_t sont iid Gaussiennes et $\theta = (\phi, \mu, \sigma_\eta^2)$.

Exemple 2: Modèle Autorégressif

- On a

$$Y_t = X_t + \sigma_\varepsilon \varepsilon_t \quad \text{et} \quad X_{t+1} = \mu(1 - \phi) + \phi X_t + \sigma_\eta \eta_t$$

où ε_t et η_t sont iid Gaussiennes et $\theta = (\phi, \mu, \sigma_\eta^2)$.

- La vraisemblance peut être calculée en utilisant Kalman.

Exemple 2: Modèle Autorégressif

- On a

$$Y_t = X_t + \sigma_\varepsilon \varepsilon_t \quad \text{et} \quad X_{t+1} = \mu(1 - \phi) + \phi X_t + \sigma_\eta \eta_t$$

où ε_t et η_t sont iid Gaussiennes et $\theta = (\phi, \mu, \sigma_\eta^2)$.

- La vraisemblance peut être calculée en utilisant Kalman.
- Proposition autorégressif for θ basée sur une distribution t-multivariée.

Probabilités d'Acceptation

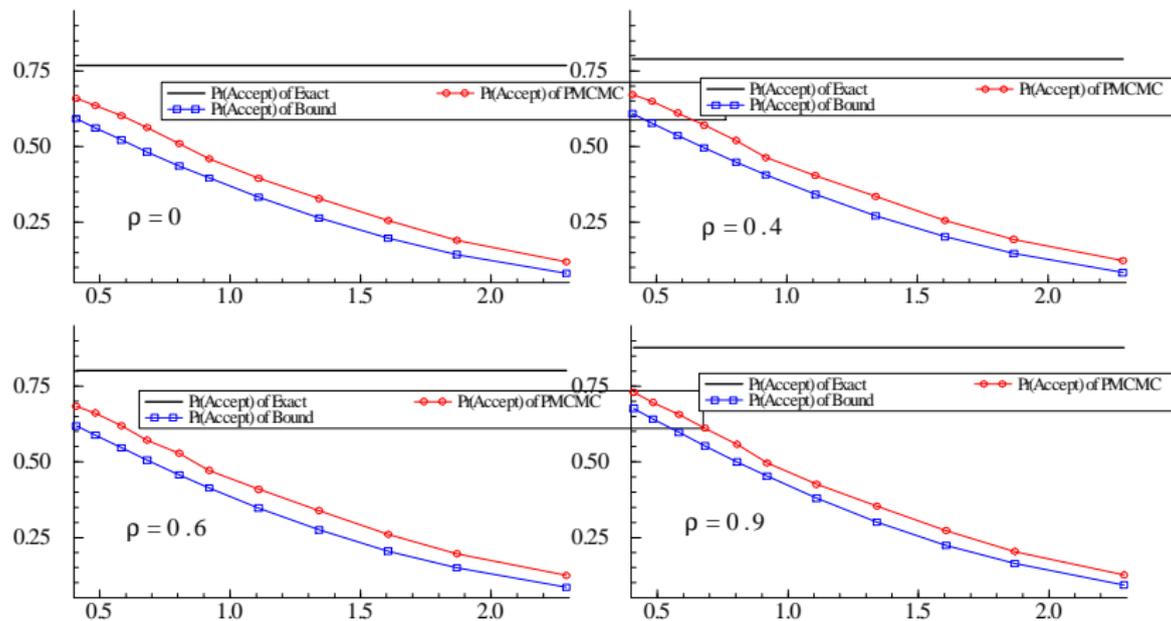


Figure: $T = 300$, $\phi = 0.8$, $\mu = 0.5$, $\sigma_{\eta}^2 = 1$, $\sigma_{\varepsilon}^2 = 0.5$. Proba d'acceptation vs $\sigma(\bar{\theta})$.

Relative Computing Time

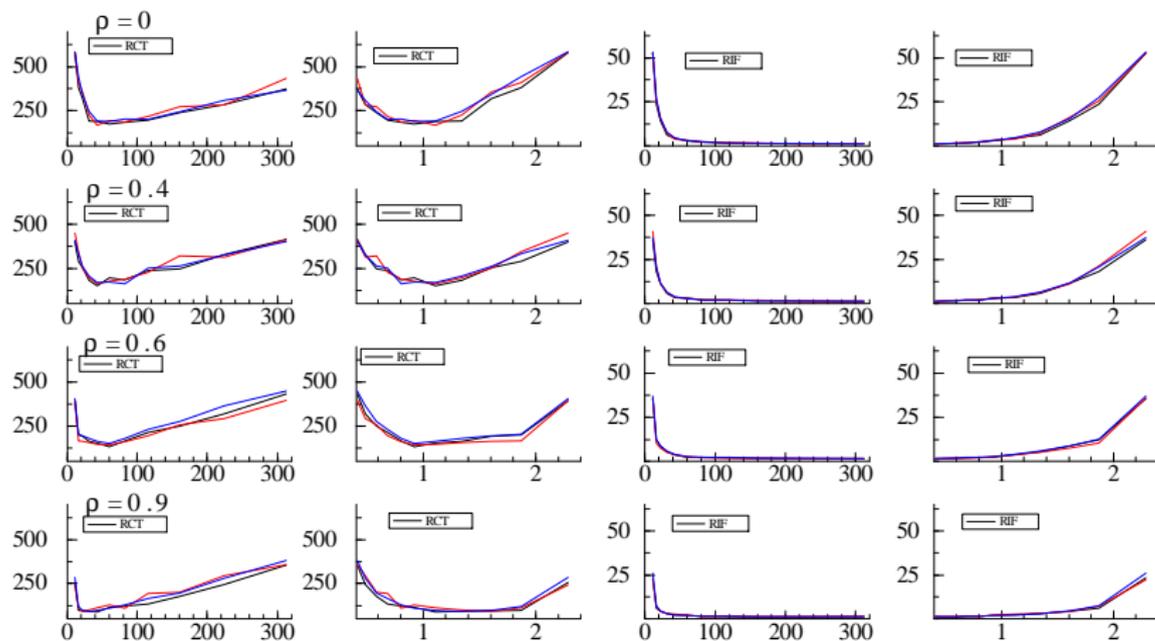


Figure: De gauche à droite: RCT_h^Q vs N , RCT_h^Q vs $\sigma(\bar{\theta})$, RIF_h^Q vs N et RIF_h^Q vs $\sigma(\bar{\theta})$.

- Pour une loi de proposition générale et sous des hypothèses réalistes sur la vraisemblance, il est raisonnable de sélectionner $\sigma \approx 1$.

- Pour une loi de proposition générale et sous des hypothèses réalistes sur la vraisemblance, il est raisonnable de sélectionner $\sigma \approx 1$.
- De meilleurs moyens d'adapter N en fonction de θ ont été récemment obtenus; e.g. Lee, Andrieu & D. 2011.

- Pour une loi de proposition générale et sous des hypothèses réalistes sur la vraisemblance, il est raisonnable de sélectionner $\sigma \approx 1$.
- De meilleurs moyens d'adapter N en fonction de θ ont été récemment obtenus; e.g. Lee, Andrieu & D. 2011.
- Il est possible d'utiliser les mêmes idées pour étudier des algorithmes qui ont de meilleures propriétés de convergence.

Setting the Number of Samples

- We want to assess experimentally whether our upper/lower bounds are sharp.

Setting the Number of Samples

- We want to assess experimentally whether our upper/lower bounds are sharp.
- We select N so as σ to be roughly equal to a pre-specified value.

Setting the Number of Samples

- We want to assess experimentally whether our upper/lower bounds are sharp.
- We select N so as σ to be roughly equal to a pre-specified value.
- ① Choose a large initial value for the number of samples, N_S .

Setting the Number of Samples

- We want to assess experimentally whether our upper/lower bounds are sharp.
- We select N so as σ to be roughly equal to a pre-specified value.
- ① Choose a large initial value for the number of samples, N_S .
- ② Run the MCMC scheme for a fixed number of iterates recording $\bar{\theta}$.

Setting the Number of Samples

- We want to assess experimentally whether our upper/lower bounds are sharp.
 - We select N so as σ to be roughly equal to a pre-specified value.
- 1 Choose a large initial value for the number of samples, N_S .
 - 2 Run the MCMC scheme for a fixed number of iterates recording $\bar{\theta}$.
 - 3 Record the estimated variance of the log of the likelihood estimator, $V(\bar{\theta}, N_S) = \hat{V} [\log \hat{p}_{N_S}(y|\bar{\theta}, u)]$.

Setting the Number of Samples

- We want to assess experimentally whether our upper/lower bounds are sharp.
 - We select N so as σ to be roughly equal to a pre-specified value.
- 1 Choose a large initial value for the number of samples, N_S .
 - 2 Run the MCMC scheme for a fixed number of iterates recording $\bar{\theta}$.
 - 3 Record the estimated variance of the log of the likelihood estimator, $V(\bar{\theta}, N_S) = \widehat{V} [\log \hat{p}_{N_S}(y|\bar{\theta}, u)]$.
 - 4 Set $N_{\bar{\theta}} = V(\bar{\theta}, N_S) / \sigma^2$.

Adaptive Choice of The Number of Samples

- Assume a consistent estimator $\hat{\gamma}^2(\theta; N)$ of $\gamma(\theta)$ is available.

Adaptive Choice of The Number of Samples

- Assume a consistent estimator $\hat{\gamma}^2(\theta; N)$ of $\gamma(\theta)$ is available.
- We select $N_{\bar{\theta}}$ so that $V(\bar{\theta}, N_{\bar{\theta}}) = \hat{V} \left[\log \hat{p}_{N_{\bar{\theta}}}(y|\bar{\theta}, u) \right] = \sigma^2$.

Adaptive Choice of The Number of Samples

- Assume a consistent estimator $\hat{\gamma}^2(\theta; N)$ of $\gamma(\theta)$ is available.
- We select $N_{\bar{\theta}}$ so that $V(\bar{\theta}, N_{\bar{\theta}}) = \hat{V} \left[\log \hat{p}_{N_{\bar{\theta}}}(y|\bar{\theta}, u) \right] = \sigma^2$.
- Select N_{θ} dynamically such that

$$\frac{\hat{\gamma}^2(\theta; N_{\bar{\theta}})}{N_{\theta}} = \sigma^2.$$

Distribution of the Standard Deviation of the Log-likelihood Estimator

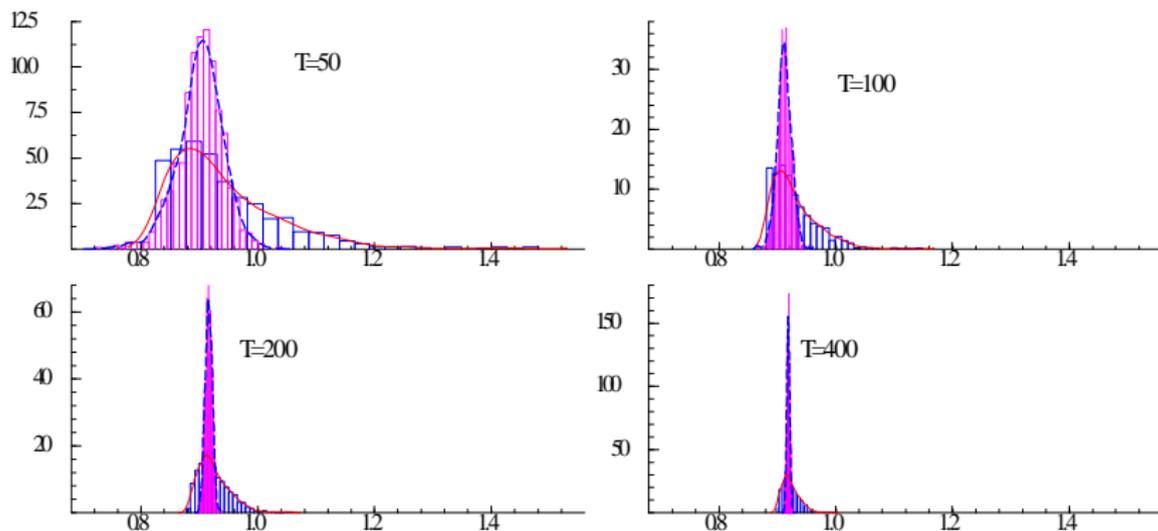


Figure: Probit model. Histograms for the standard deviation of the log-likelihood (over post of θ) when N is chosen statically (blue) for the proposal as $N_{\bar{\theta}}$ and when it is chosen dynamically (pink) as N_{θ} for $\sigma^2 = 0.92$.