

Introduction à l'économétrie bayésienne: Exemples d'applications

Guillaume Horny*

*Banque de France et UCLouvain
guillaume.horny@banque-france.fr

Plan de la présentation

- ➊ Régression Linéaire
- ➋ Modèle AR
- ➌ Panel Probit
- ➍ Mélange discret de normales
- ➎ DSGE

Partie I

Régression Linéaire

Plan de la partie

- 1 Modèle
- 2 A priori vague
- 3 A priori conjugué

Erreurs normales, indépendantes, homoscédastiques

Modèle: $y = X\beta + \epsilon$, où $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$.

Vraisemblance:

$$l(y|\beta, \sigma^2) \propto (\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - x_i\beta)^2 \right].$$

Choix de l'a priori

Besoin d'un **a priori** $\pi(\beta, \sigma^2)$

On va tester:

- 1 a priori vague
- 2 a priori informatif

Loi a posteriori (I)

A priori uniforme pour $(\beta, \ln \sigma^2)$:

$$\pi(\beta, \sigma^2) \propto 1/\sigma^2, -\infty < \beta < \infty, \sigma^2 > 0.$$

Loi a posteriori:

$$l(y|\beta, \sigma^2) \propto (\sigma^2)^{-1-n/2} \exp \left[-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right].$$

Illustration (I)

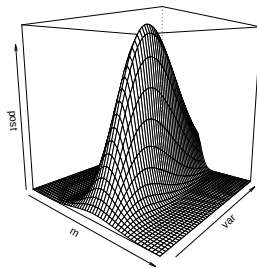
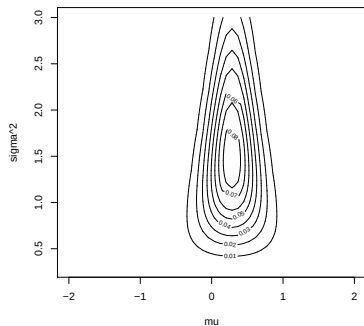
Soit $X\beta = \mu$ de dimension $(n \times 1)$

Loi a posteriori:

$$l(y|\beta, \sigma^2) \propto (\sigma^2)^{-1-n/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y - \mu)^2 \right].$$

$\Rightarrow$ loi bivariee

Illustration (II)



Mode: ($\mu = 0.28, \sigma^2 = 1.49$)
 Ce n'est pas une *normale* bivariée!

Loi a posteriori (II)

On peut réécrire la **loi a posteriori**:

$$p(\beta, \sigma^2 | y, X) \propto (\sigma^2)^{-1-n/2} \exp \left[-\frac{1}{2\sigma^2} (\beta - b)' X' X (\beta - b) \right] \\ \exp \left[-\frac{1}{2\sigma^2} e' e \right],$$

où $b = (X'X)^{-1}X'y$ et $e = y - Xb$.

Loi marginale a posteriori de σ^2 (I)

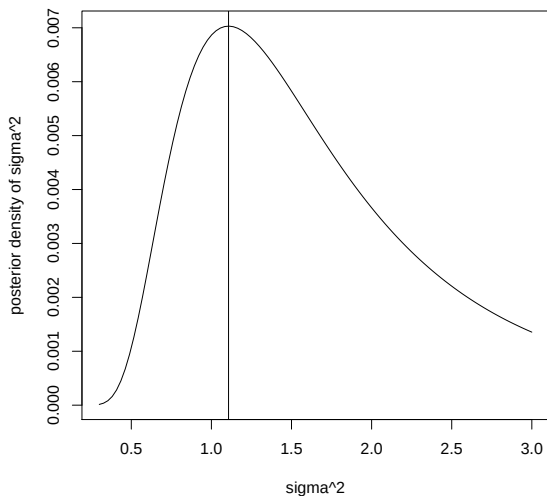
Loi marginale de σ^2 :

$$p(\sigma^2|y, X) \propto (\sigma^2)^{-v/2} \exp\left(-\frac{vs^2}{2\sigma^2}\right)$$

$$\Leftrightarrow 1/\sigma^2 \sim \gamma(v/2, vs^2/2),$$

où $v = n - k$ et $s^2 = (y - Xb)'(y - Xb)/v$.

Estimateur bayésien: $E(\sigma^2|y, X) = s^2$.

Loi marginale de σ^2 (II)

Loi marginale a posteriori β (I)

Loi marginale de β :

$$p(\beta|y, X) \propto \left(1 + \frac{s^2}{n-k} (\beta - b)' (X'X)^{-1} (\beta - b) \right)^{-n/2}$$

Estimateur bayésien: $E(\beta|y, X) = b$.

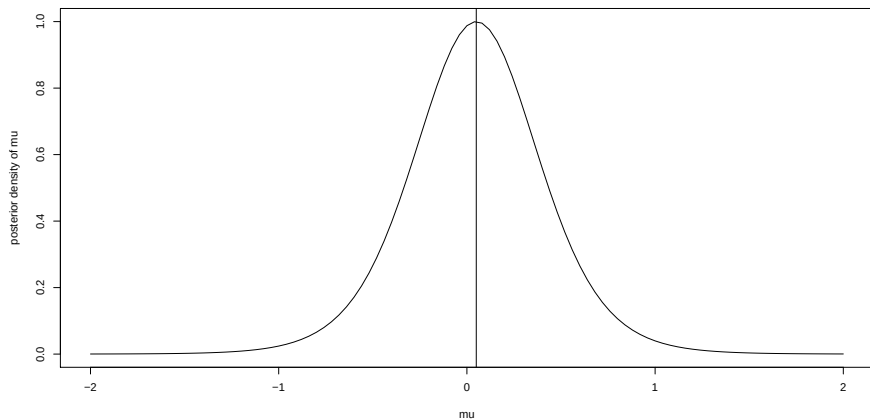
Pour chaque élément β_j de β , on a:

$$\frac{\beta_j - b_j}{s \sqrt{(X'X)^{-1}_{jj}}} \sim t_v,$$

où $(X'X)^{-1}_{jj}$ est le j -ème élément diagonal de $(X'X)^{-1}$.

Loi marginale de β (II)

Cas $X\beta = \mu$



Simulations de (β, σ^2)

Comme $\beta|\sigma^2 \sim \mathcal{N}(b, \sigma^2(X'X)^{-1})$, on échantillonne dans la loi a posteriori:

Algorithme

- 1 tirer σ^2 dans sa loi marginale a posteriori
- 2 tirer β sa loi conditionnelle à σ^2
- 3 répéter les étapes 1 et 2

Loi marginale a posteriori ϵ

Estimateur bayésien de ϵ :

$$E(\epsilon|y, X) = V(y - X\beta|y, X) = y - Xb.$$

A priori conjugué

A priori vague:

- permet de voir le rôle de l'estimateur MV
- rend les calculs faciles

Autre possibilité: a priori conjugué

Choix de l'a priori

Comme:

$$l(y, X|\beta, \sigma^2) \propto (\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} (\beta - b)' X' X (\beta - b) \right] \exp \left[-\frac{1}{2\sigma^2} e' e \right].$$

l'a priori conjugué est de la forme:

$$\pi(\beta, \sigma^2) \propto (\sigma^2)^{1-\alpha/2} \exp \left[-\frac{1}{2\sigma^2} (\beta - \beta_0)' A (\beta - \beta_0) \right] \exp \left[-\frac{1}{2\sigma^2} c \right].$$

Interprétation de l'a priori

D'où:

$$\begin{aligned}1/\sigma^2 &\sim \gamma(\alpha/2, c/2), \\ \beta|\sigma^2 &\sim \mathcal{N}(\beta_0, \sigma^2 A^{-1}).\end{aligned}$$

Estimateur bayésien de β

Estimateur bayésien de β :

$$E(\beta|Y, X) = (X'X + A)^{-1}(X'Xb + A\beta_0).$$

$\Rightarrow$ moyenne pondérée de l'estimateur MV et des paramètres a priori.

Partie II

Modèle AR

Plan de la partie

- 1 Modèle
- 2 A priori
- 3 Simulation de données
- 4 A priori excluant une racine unitaire
- 5 Résultats
- 6 A priori permettant une racine unitaire
- 7 Résultats
- 8 Extensions

Modèle

$\forall t = 1, 2, \dots, T:$

$$y_t = \alpha + \rho y_{t-1} + \epsilon_t,$$
$$\epsilon_t \sim \mathcal{N}(0, \sigma^2).$$

Vraisemblance

On note $\tau = 1/\sigma^2$

Vraisemblance:

$$p(y|y_0, \alpha, \rho, \tau) \propto \tau^{T/2} \exp \left[-\frac{\tau}{2} \sum_{t=1}^T (y_t - \alpha - \rho y_{t-1})^2 \right],$$

pour $(\alpha, \rho) \in \mathcal{R}^2, \tau > 0$.

Choix de l'a priori

- A priori excluant une racine unitaire:

Pour $-1 < \rho < 1 : p(\rho) \propto 1,$

$$p(\rho) \propto (1 - \rho^2)^{-1/2}$$

...

- A priori permettant une racine unitaire:

Pour $-\infty < \rho < \infty : p(\rho) \propto 1,$

A priori de Jeffrey : expression complexe fonction de T

Vraisemblance sous l'hypothèse de stationnarité

Sous l'hypothèse de stationnarité:

$$\begin{aligned}
 l(y, y_0 | \alpha, \rho, \tau) &= p(y | y_0, \alpha, \rho, \tau) p(y_0 | \alpha, \rho, \tau) \\
 &= \tau^{T/2} \exp \left[-\frac{\tau}{2} \sum_{t=1}^T (y_t - \alpha - \rho y_{t-1})^2 \right] \\
 &\quad \times \tau^{1/2} (1 - \rho^2) \exp \left[-\frac{\tau}{2} (1 - \rho^2) (y_0 - \mu)^2 \right],
 \end{aligned}$$

où $\mu = \alpha / (1 - \rho)$, valeur de y_t à l'état stationnaire

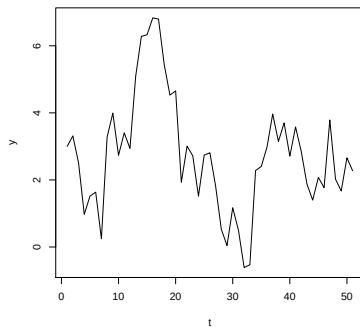
Relations avec le modèle linéaire

La vraisemblance est identique à celle du modèle linéaire!

⇒ sous des a priori diffus pour (α, ρ, τ) , l'estimateur bayésien est le même!

Données

Données simulées avec $t = 50$, $\alpha = 0.3$, $\rho = 0.9$ et $\tau = 1$

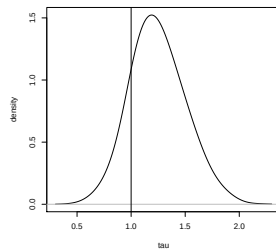
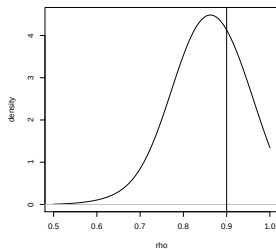
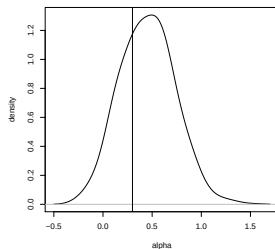


A priori:

- $\pi(\alpha)$ gaussienne, $\pi(\rho)$ uniforme sur $] -1, 1[$ et τ gamma
- peu informatifs: $\text{Var}(\alpha)=10^6$, $\text{Var}(\tau)=10^3$

Résultats

10 000 itérations, retient une itération sur 5 après les 5000 premières
Lois marginales a posteriori:

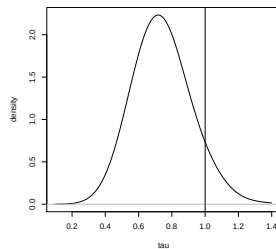
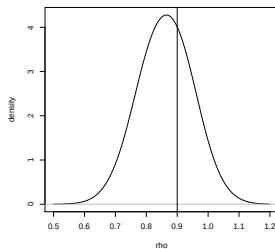
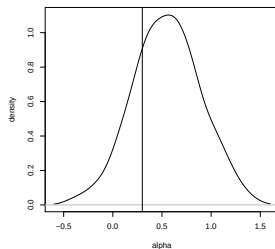


A priori:

- $\pi(\alpha)$ et $\pi(\rho)$ impropres, τ gamma
- peu informatifs: $\text{Var}(\tau)=10^3$

Résultats

10 000 itérations, retient une itération sur 5 après les 5000 premières
Lois marginales a posteriori:



Extension aux VAR

Cogley et Sargent (2001, 2002): dynamique de l'inflation aux Etats-Unis
Modèle:

$$Y_t = B_t Y_{t-1} + \epsilon_t,$$

$$B_t = B_{t-1} + \omega_t,$$

où $\epsilon_t \sim \mathcal{N}(0, \Sigma)$ et $\omega_t \sim \mathcal{N}(0, V)$.

Partie III

Mélange discret de normales

Plan de la partie

- 1 Modèle
- 2 Données simulées
- 3 Vraisemblance
- 4 Distributions a priori
- 5 Echantillonnage de Gibbs
- 6 Paramétrisation alternative
- 7 Algorithme du recuit simulé

Modèle

Les données comprennent 2 groupes, et on ne sait pas à quel groupe appartient l'individu i .

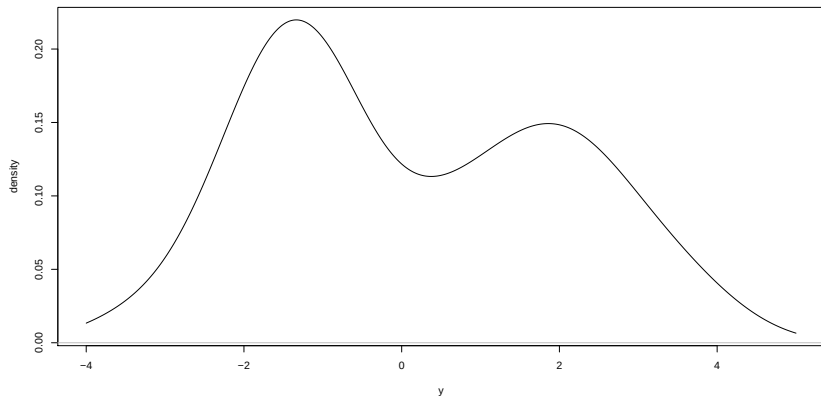
Loi marginale de y_i :

$$f(y_i|\mu_1, \mu_2, \alpha, \tau) = \alpha f(y_i|\mu_1, \tau) + (1 - \alpha)f(y_i|\mu_2, \tau),$$

où $\tau > 0$ et $0 \leq \alpha \leq 1$.

Observations issues d'un mélange de deux gaussiennes

$n = 200, \alpha = 0.4, \mu_1 = 2, \mu_2 = -1, \tau = 1.$



Vraisemblance

$$l(y_i|\mu_1, \mu_2, \alpha, \tau) = \prod_{i=1}^n \left(\alpha^{d_i} f(y_i|\mu_1, \tau)^{d_i} + (1 - \alpha)^{1-d_i} f(y_i|\mu_2, \tau)^{1-d_i} \right),$$

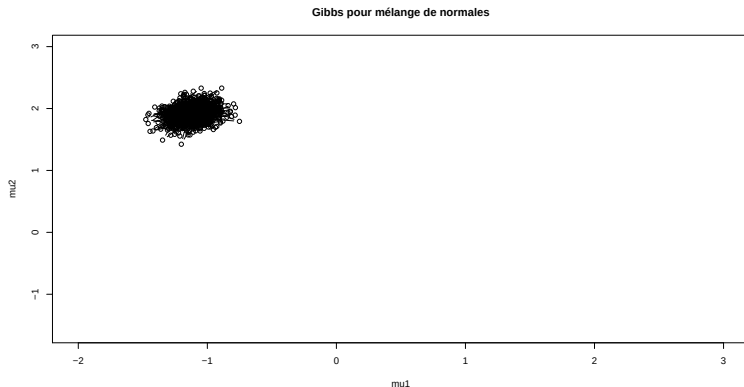
où d_i indique l'appartenance de l'individu i au groupe 1.

Distributions a priori

$$\begin{aligned}\alpha &\sim \mathcal{D}(1, 1), \\ \mu_1 &\sim \mathcal{N}(0, 10^6), \\ \mu_2 &\sim \mathcal{N}(0, 10^6), \\ \tau &\sim \gamma(0.001, 0.001).\end{aligned}$$

Echantillonnage de Gibbs (I)

10 000 itérations, retient une itération sur 5 après les 5000 premières

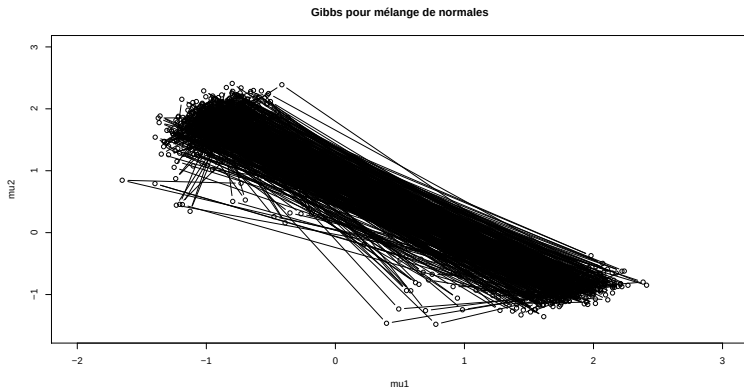


Echantillonnage de Gibbs (II)

Ré-essai avec mêmes données, valeurs initiales et paramètres des simulations:

Echantillonnage de Gibbs (II)

Ré-essai avec mêmes données, valeurs initiales et paramètres des simulations:



Echantillonnage de Gibbs (III)

- méthode stochastique!
- les deux composantes s'échangent leurs populations à chaque itération

Reparamétrisation

Robert (1994) propose de reparamétriser le modèle:

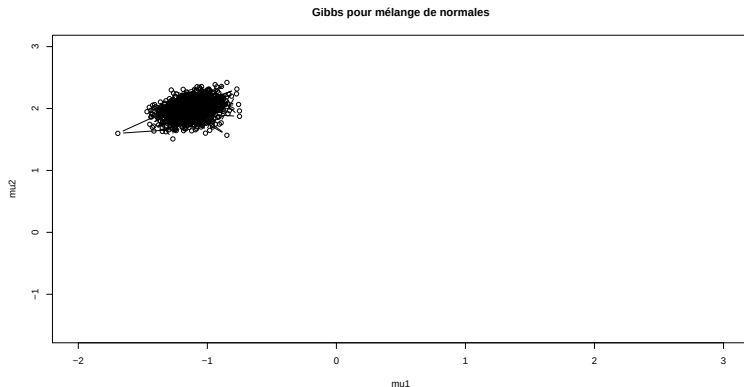
$$\mu_2 = \mu_1 + \theta,$$

où $\theta > 0$.

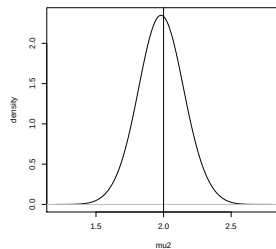
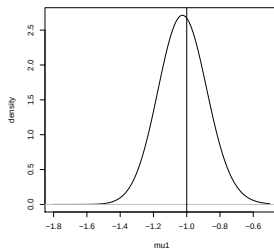
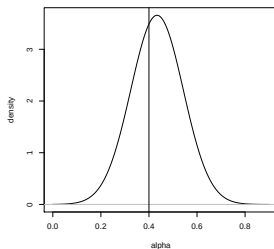
A priori: $\theta \sim \mathcal{TN}_{0,\infty}(0, 10^6)$

Echantillonnage de Gibbs (III)

10 000 itérations, retient une itération sur 5 après les 5000 premières



Lois marginales a posteriori



Méthode de recuit simulé

- Algorithme réputé pour optimiser les fonctions multimodales
- Issu d'un processus de métallurgie, où l'alternance des cycles de refroidissement lent et de réchauffage minimise l'énergie cinétique des molécules

Algorithme

Algorithme de recuit simulé

Initialiser x_0 et T_0 .

Itération (k):

❶ tirer x_c dans une loi symétrique $q(x^{(k)}, x_c)$

❷ tirer $u \sim \mathcal{U}_{[0,1]}$

❸ calculer:

$$r = \min \left(\exp \left[f(x_c) - f(x^{(k)}) / T^{(k)} \right], 1 \right).$$

❹

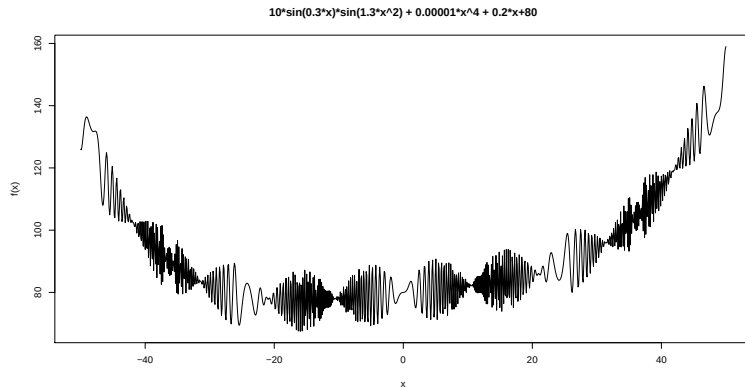
$$u \begin{cases} \leq r, & x^{(k+1)} = x_c \\ > r, & x^{(k+1)} = x^{(k)} \end{cases}$$

❺ répéter les étapes de 1 à 3 jusqu'à ce que $T^{(k)} \rightarrow 0$.

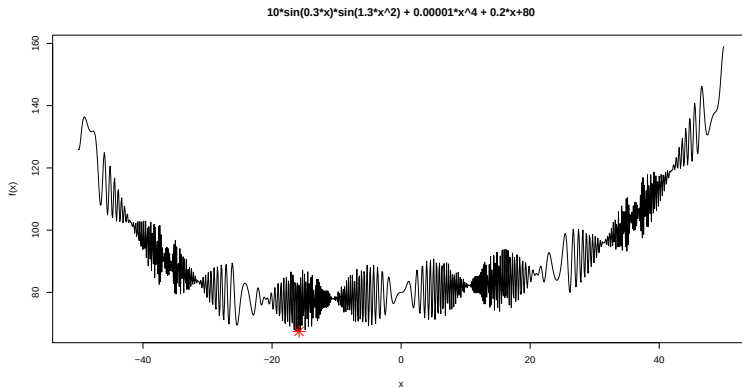
Intuitions

- le tirage est toujours accepté si $f(x_c) > f(x^{(k)})$ et $T < 1$
- le tirage est parfois accepté si $f(x_c) < f(x^{(k)})$ et $T < 1$, permettant d'échapper aux optima locaux
- lorsque $T \rightarrow 0$, $x^{(j+1)}$ est proche de $x^{(j)}$ et les tirages sont de plus en plus concentrés autour l'optimum local
- la littérature suggère généralement $T^{(j)} = 1/\ln j$.

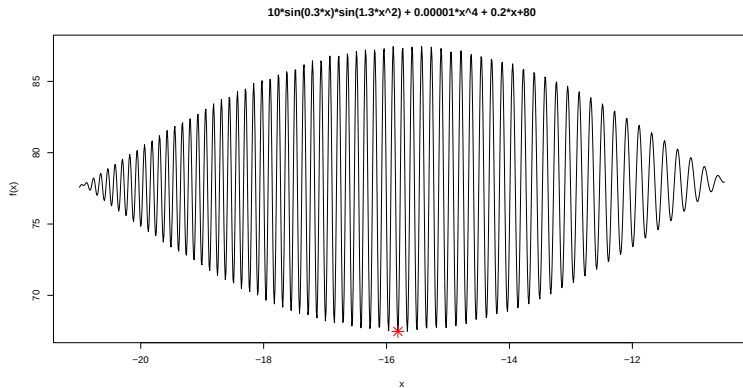
Exemple



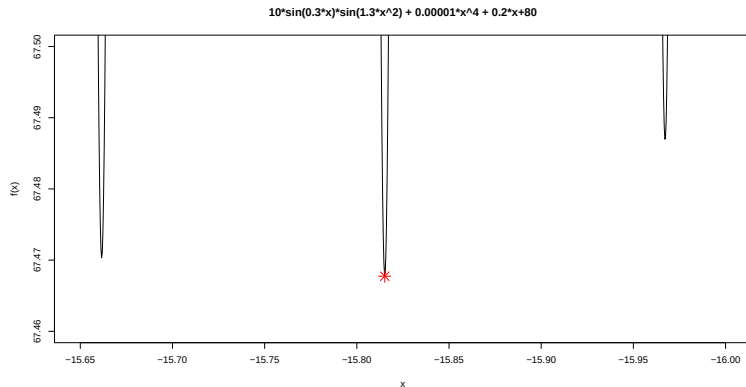
Exemple



Exemple



Exemple



Remarques

- le recuit simulé est une méthode Metropolis-Hastings (cf Robert, 1996)
- si $T(n)$, la chaîne de Markov est hétérogène

Partie IV

Panel Probit

Plan de la partie

- 1 Modèle
- 2 Loi a priori
- 3 Algorithme d'échantillonnage de Gibbs

Modèle

Soient $i = 1, \dots, n$, $t = 1, \dots, T$ (panel cylindré).

On note $y_i = (y_{i1}, \dots, y_{iT})$:

$$y_{it} = \begin{cases} 1 & \text{si } y_{it}^* > 0 \\ 0 & \text{si } y_{it}^* < 0 \end{cases},$$

$$y_{it}^* = x_{it}\beta + w_{it}b_i + \epsilon_{it},$$

$$b_i \sim \mathcal{N}(0, D),$$

$$\epsilon_{it} \sim \mathcal{N}(0, 1),$$

où b_i est un vecteur ($q \times 1$) d'effets aléatoires pour l'individu i .

Densité jointe des observations

Contribution de l'individu i :

$$\Pr(y_i|\beta, D) = \int_{\mathcal{R}^q} \left\{ \prod_{t=1}^T [\Phi(x_{it}\beta + w_{it}b_i)]^{y_{it}} [1 - \Phi(x_{it}\beta + w_{it}b_i)]^{1-y_{it}} \right\} \phi_q(b_i|0, D) db_i.$$

l'évaluation de la vraisemblance demande d'intégrer q fois

Au delà de $q = 3$, utiliser la quadrature de Gauss-Hermite est un acte de foi!

Approche Albert et Chib (1993)

Soit la variable latente z_{it} telle que:

$$z_{it}|b_i \sim \mathcal{N}(x_{it}\beta + w_{it}b_i, 1).$$

On en déduit:

$$y_{it} = \mathbb{1}(z_{it} > 0),$$
$$f(y_{it}|z_{it}, b_i) = \mathbb{1}(y_{it} = 1)\mathbb{1}(z_{it} > 0) + \mathbb{1}(y_{it} = 0)\mathbb{1}(z_{it} < 0).$$

Lois a posteriori

$$p(\beta, z, b|y) \propto \pi(\beta) \prod_{i=1}^n \prod_{t=1}^T f(y_{it}|z_{it}, b_i) f(z_{it}|b_i) f(b_i)$$

Loi a priori

$$\begin{aligned} D' &\sim \text{Wishart}(v_0^{-1} R_0, v_0), \\ \beta &\sim \mathcal{N}_p(\beta_0, B_0). \end{aligned}$$

Algorithme d'échantillonnage de Gibbs (I)

Algorithme

- 1 tirer β dans $\beta|y, z$,
- 2 tirer z dans $z|y, \beta$,
- 3 répéter les étapes 1 et 2.

Algorithme d'échantillonnage de Gibbs (II)

Lois marginales conditionnelles:

- la loi de $\beta|y, z, b$ est identique à la loi de $\beta|z, b$, cad gaussienne
- la loi de $z|y, \beta, b$ est le produit des lois des $z_{it}|y_{it}, \beta, b_i$, qui sont normales tronquées:

$$z_{it} \sim \begin{cases} \mathcal{TN}_{(0,\infty)}(x_{it}\beta + w_{it}b_i, 1) & \text{si } y_{it} = 1 \\ \mathcal{TN}_{(-\infty,0)}(x_{it}\beta + w_{it}b_i, 1) & \text{si } y_{it} = 0 \end{cases}$$

Algorithme d'échantillonnage de Gibbs (III)

On note pour l'individu i :

$$z_i = X_i\beta + W_ib_i + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, I_T).$$

Algorithme

1

$$\beta | \{z_{it}\}, D \sim \mathcal{N} \left(B \left[B_0^{-1}\beta_0 + \sum_{i=1}^n X_i' V_i^{-1} z_i \right], B \right),$$

$$\text{où } B = (B_0^{-1} + \sum_{i=1}^n X_i' V_i^{-1} X_i)^{-1}, V_i = I_T + W_i D W_i'.$$

Algorithme d'échantillonnage de Gibbs (IV)

Algorithme (suite)

2

$$z_{it} | z_{i(-t)}, y_{it}, \beta, D \sim \begin{cases} \mathcal{TN}_{(0,\infty)}(\mu_{it}, v_{it}) & \text{si } y_{it} = 1 \\ \mathcal{TN}_{(-\infty,0)}(\mu_{it}, v_{it}) & \text{si } y_{it} = 0 \end{cases}$$

où $\mu_{it} = \mathbb{E} [z_{it} | z_{i(-t)}, \beta, D]$, $v_{it} = \text{Var} [z_{it} | z_{i(-t)}, \beta, D]$.

Algorithme d'échantillonnage de Gibbs (V)

Algorithme (suite)

3

$$b_i | y, \beta, D \sim \mathcal{N}(D_i W_i' (z_i - X_i \beta, D_i)),$$

$$\text{où } D_i = (D^{-1} + W_i' W_i)^{-1}.$$

4

$$D^{-1} | y, \beta, \{b_i\} \sim \text{Wishart}(v_0 + n, R_0).$$

Partie V

DSGE

Plan de la partie

1 Méthode générale de résolution

2 Algorithme MH général

Méthode générale de résolution

- ① les choix des agents sont issus de programmes d'optimisations dynamiques
- ② caractérisation de l'équilibre de l'économie (état stationnaire)
- ③ (log-)linéarisation des conditions d'équilibre au voisinage de l'état stationnaire ...
- ④ ... pour en déduire un système d'équations dynamiques
- ⑤ recherche des racines stables

Algorithme MH général d'estimation de DSGE (I)

Algorithme MH

Lecture des données

Initialisation des paramètres en $\theta^{(0)}$

Itération (k) :

- 1 évaluation de $\pi(\theta^{(k)})$
- 2 résolution du système dynamique sachant $\theta^{(k)}$, recherche des racines stables
- 3 évaluation de $l(Y^T|\theta^{(k)})$ (généralement avec le filtre de Kalman)
- 4 tirage de θ_c dans $\mathcal{N}(\theta^{(k)}, \Sigma)$
- 5 répéter les étapes de 1 à 3 pour $\theta^{(k)} = \theta_c$.

Algorithme MH général d'estimation de DSGE (II)

Algorithme MH (suite)

6 évaluer:

$$r = \frac{I(Y^T | \theta_c) \pi(\theta_c)}{I(Y^T | \theta^{(k)}) \pi(\theta^{(k)})}$$

7

$$r \begin{cases} \geq 1, & \theta^{(k+1)} = \theta_c \\ < 1, & \begin{cases} \theta^{(k+1)} = \theta_c \\ \theta^{(k+1)} = \theta^{(k)} \end{cases} \end{cases} \quad \begin{array}{l} \text{avec probabilité } r \\ \text{avec probabilité } 1 - r \end{array}$$

8 répéter les étapes de 1 à 7

Partie VI

Références

Références transversales

Fernandez-Villaverde: Methods in Macroeconomic Dynamics, notes de cours.

Lancaster (2004): An introduction to modern bayesian econometrics, Blackwell Publishing.

Robert (1996): Méthodes de Monte Carlo par chaînes de Markov, Economica.

Bibliographie

Albert et Chib (1993): Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts, *Journal of Business and Economic Statistics*, 11, 1-15.

Cogley et Sargent (2001): Evolving post world war II US inflation dynamics, *NBER Macroeconomics annual*, 16, 331-373.

Chib et Carlin (1999): On MCMC sampling in hierarchical longitudinal models, *Statistics and Computing*, 9, 17-26.

Robert (1994): Discussion of "Markov Chains for exploring posterior distribution", *Annals of Statistics*, 22, 1742-1747.

Logiciels utilisés

- R
 bibliothèques: R2WinBUGS, BRugs
- Openbugs

Merci de votre attention!
guillaume.horny@banque-france.fr