

ÉCONOMÉTRIE BAYÉSIENNE

STÉPHANE ADJEMIAN

1. INTRODUCTION

Dans cette note je développe deux exemples d'estimation bayésienne. Nous avons déjà rencontré le premier exemple en cours, mais je n'avais pas donné les détails des calculs à l'époque. Il s'agit d'estimer l'espérance d'une variable aléatoire. Le second exemple est lui plus original par rapport au contenu du cours : il illustre l'approche bayésienne des modèles VAR.

2. ESTIMATION DE L'ESPÉRANCE

Le processus générateur des données est défini par :

$$(1a) \quad y_t = \mu + \varepsilon_t$$

$$(1b) \quad \varepsilon_t \underset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$$

où $t = 1, \dots, T$. On suppose que la variance σ_ε^2 est connue, nous cherchons à estimer l'espérance du processus $\{y_t\}$.

Notre croyance *a priori* sur μ est caractérisée par une loi normale centrée en μ_0 et de variance σ_μ^2 . La densité *a priori* est donc :

$$(2) \quad p_0(\mu) = (2\pi\sigma_\mu^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_\mu^2}(\mu-\mu_0)^2}$$

La densité postérieure de μ est proportionnelle au produit de la vraisemblance et de la densité *a priori*, nous devons donc commencer par écrire la vraisemblance.

2.1. Vraisemblance. La vraisemblance est la densité jointe de l'échantillon sachant les paramètres, μ et σ_ε^2 . D'après le processus générateur des données (1) nous savons que y_t est normalement distribué d'espérance μ et de variance σ_ε^2 . Comme par hypothèse y_t est indépendant de y_s pour tout $s \neq t$, la densité de l'échantillon $\mathcal{Y}_T = (y_1, \dots, y_T)$ est le produit des densités marginales :

$$(3) \quad p(\mathcal{Y}_T | \mu, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T (y_t - \mu)^2}$$

Notons que :

$$\begin{aligned} \sum_{t=1}^T (y_t - \mu)^2 &= \sum_{t=1}^T ([y_t - \hat{\mu}] - [\mu - \hat{\mu}])^2 \\ &= \sum_{t=1}^T (y_t - \hat{\mu})^2 + \sum_{t=1}^T (\mu - \hat{\mu})^2 - \sum_{t=1}^T (y_t - \hat{\mu})(\mu - \hat{\mu}) \\ &= \nu s^2 + T(\mu - \hat{\mu})^2 - \left(\sum_{t=1}^T y_t - T\hat{\mu} \right) (\mu - \hat{\mu}) \\ &= \nu s^2 + T(\mu - \hat{\mu})^2 \end{aligned}$$

où $\nu = T - 1$, $s^2 = (T - 1)^{-1}(y_t - \hat{\mu})^2$ est un estimateur de la variance de $\{y_t\}$ et $\hat{\mu} = T^{-1} \sum_{t=1}^T y_t$ est l'estimateur du maximum de vraisemblance de μ . Finalement, nous pouvons écrire la vraisemblance sous une forme équivalente :

$$(4) \quad p(\mathcal{Y}_T | \mu, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2}(\nu s^2 + T(\mu - \hat{\mu})^2)}$$

2.2. Densité postérieure. On obtient la densité postérieure, à une constante d'intégration près, en multipliant (2) par (4) :

$$p(\mu | \mathcal{Y}_T) \propto (2\pi\sigma_\mu^2)^{-\frac{1}{2}} (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2}(\nu s^2 + T(\mu - \hat{\mu})^2)} e^{-\frac{1}{2\sigma_\mu^2}(\mu - \mu_0)^2}$$

En éliminant les termes qui ne dépendent pas de μ , il vient :

$$p(\mu | \mathcal{Y}_T) \propto e^{-\frac{T}{2\sigma_\varepsilon^2}(\mu - \hat{\mu})^2} e^{-\frac{1}{2\sigma_\mu^2}(\mu - \mu_0)^2}$$

ou de façon équivalente :

$$p(\mu | \mathcal{Y}_T) \propto \exp \left\{ -\frac{1}{2} \left(\frac{T}{\sigma_\varepsilon^2} (\mu - \hat{\mu})^2 + \frac{1}{\sigma_\mu^2} (\mu - \mu_0)^2 \right) \right\}$$

Nous allons maintenant réécrire cette équation de façon à ne faire apparaître qu'une forme quadratique en μ . En développant les carrés, et en notant $A(\mu)$ le terme entre (grandes) parenthèses sous l'exponentielle, il vient :

$$\begin{aligned} A(\mu) &= \frac{T}{\sigma_\varepsilon^2} (\mu^2 + \hat{\mu}^2 - 2\mu\hat{\mu}) + \frac{1}{\sigma_\mu^2} (\mu^2 + \mu_0^2 - 2\mu\mu_0) \\ &= \left(\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2} \right) \mu^2 - 2\mu \left(\frac{T}{\sigma_\varepsilon^2} \hat{\mu} + \frac{1}{\sigma_\mu^2} \mu_0 \right) + \left(\frac{T}{\sigma_\varepsilon^2} \hat{\mu}^2 + \frac{1}{\sigma_\mu^2} \mu_0^2 \right) \\ &= \left(\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2} \right) \left[\mu^2 - 2\mu \frac{\frac{T}{\sigma_\varepsilon^2} \hat{\mu} + \frac{1}{\sigma_\mu^2} \mu_0}{\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2}} \right] + \left(\frac{T}{\sigma_\varepsilon^2} \hat{\mu}^2 + \frac{1}{\sigma_\mu^2} \mu_0^2 \right) \\ &= \left(\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2} \right) \left[\mu - \frac{\frac{T}{\sigma_\varepsilon^2} \hat{\mu} + \frac{1}{\sigma_\mu^2} \mu_0}{\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2}} \right]^2 + \left(\frac{T}{\sigma_\varepsilon^2} \hat{\mu}^2 + \frac{1}{\sigma_\mu^2} \mu_0^2 \right) - \frac{\left(\frac{T}{\sigma_\varepsilon^2} \hat{\mu} + \frac{1}{\sigma_\mu^2} \mu_0 \right)^2}{\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2}} \end{aligned}$$

En notant que les deux derniers termes ne dépendent pas de μ , il vient :

$$A(\mu) \propto \left(\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2} \right) \left[\mu - \frac{\frac{T}{\sigma_\varepsilon^2} \hat{\mu} + \frac{1}{\sigma_\mu^2} \mu_0}{\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2}} \right]^2$$

et finalement :

$$(5) \quad p(\mu | \mathcal{Y}_T) \propto \exp \left\{ -\frac{1}{2} \left(\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2} \right) \left[\mu - \frac{\frac{T}{\sigma_\varepsilon^2} \hat{\mu} + \frac{1}{\sigma_\mu^2} \mu_0}{\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2}} \right]^2 \right\}$$

On reconnaît, à une constante d'intégration près, la densité d'une loi normale. Ainsi, la distribution postérieure de μ est gaussienne d'espérance :

$$\mathbb{E}[\mu] = \frac{\frac{T}{\sigma_\varepsilon^2} \hat{\mu} + \frac{1}{\sigma_\mu^2} \mu_0}{\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2}}$$

et de variance :

$$\mathbb{V}[\mu] = \frac{1}{\frac{T}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\mu^2}}$$

On voit que si le prior est non informatif, au sens où la variance *a priori*, σ_μ^2 , tend vers l'infini, alors la variance *a posteriori* est la variance de l'estimateur du maximum de vraisemblance de μ , σ_ε^2/T . Dès lors que l'on apporte de l'information, la variance postérieure est plus faible que la variance de l'estimateur du MV. L'intérêt pratique de l'approche bayésienne est illustré par ce résultat. Dans un environnement où nous disposons de peu de données, l'échantillon est faiblement informatif, la prise en compte de nos croyances *a priori* permet d'accroître la précision de l'estimation. En spécifiant un prior, avec une variance finie plus ou moins importante, nous augmentons le degré de liberté.

L'espérance postérieure est une combinaison linéaire convexe de l'espérance *a priori* et de l'estimateur du maximum de vraisemblance (ici la moyenne empirique). Le mélange est défini par la variance *a priori* et la variance de l'estimateur du MV. Si l'information *a priori* est plus importante que l'information empirique (provenant de l'échantillon) alors l'espérance postérieure sera plus proche de l'espérance *a priori* que de l'estimateur du MV. En raisonnant à la limite :

- (i) Lorsque σ_0 tend vers 0, c'est-à-dire lorsque l'information *a priori* tend vers l'infini, l'espérance postérieure tend vers l'espérance *a priori*. On note que dans ce cas la variance postérieure tend vers zéro. On peut interpréter ce cas limite comme une calibration du modèle.
- (ii) Lorsque T tend vers l'infini, c'est-à-dire lorsque l'information empirique tend vers l'infini, l'espérance postérieure tend vers l'estimateur du MV. Dans le même temps la variance postérieure se rapproche de celle de l'estimateur du MV et tend finalement vers zéro. Plus généralement nous pourrions montrer que la densité postérieure hérite des propriétés de l'estimateur du maximum de vraisemblance¹).

La distribution postérieure est gaussienne tout comme la distribution *a priori* : nous avons choisi un prior conjugué.

2.3. Estimation ponctuelle. Dans la section précédente nous avons montré que la distribution postérieure de μ est gaussienne. Pour communiquer les résultats sous une forme plus synthétique on peut vouloir choisir un point dans la distribution *a posteriori*, c'est-à-dire proposer une estimation ponctuelle. Nous avons vu en cours que cela s'apparente à un problème de choix en univers incertain. Il est alors naturel de se donner une fonction de perte, $L(a, \mu)$, qui spécifie la perte occasionnée par le choix a alors que la vraie valeur est μ , et de minimiser l'espérance postérieure de la perte :

$$\mu^* = \arg \min_a \int_{\mathbb{R}} L(a, \mu) p(\mu | \mathcal{Y}_T) d\mu$$

Si la fonction de perte est quadratique, $L(a, \mu) = (a - \mu)^2$ alors on montre facilement que l'estimation ponctuelle, μ^* , est l'espérance postérieure de μ donnée plus haut. On obtient la même estimation ponctuelle avec la fonction de perte $(a, \mu) = |a - \mu|$, car dans le cas d'une distribution gaussienne il y a identité entre la médiane et l'espérance.

3. ESTIMATION D'UN VAR

Dans cette section, nous considérons un autre exemple où les résultats peuvent être obtenus « à la main ». Le modèle VAR gaussien se prête, comme tout modèle linéaire gaussien, à cet exercice et a l'avantage d'être un outil couramment utilisé

¹Nous avons vu en cours que, sous des conditions très générales, la distribution postérieure est normalement distribuée lorsque la dimension de l'échantillon tend vers l'infini (même propriété que l'estimateur du MV qui est asymptotiquement gaussien).

en macro-économie

Nous considérons un modèle VAR(p) pour caractériser le vecteur $1 \times m$ de variables endogènes y_t observées :

$$y_t = \sum_{i=1}^p y_{t-i} \mathbf{A}_i + \varepsilon_t$$

où $\{\mathbf{A}_i\}$ est une séquence de matrice $m \times m$ et ε_t est un bruit blanc gaussien, de dimension $1 \times m$ d'espérance nulle et de variance $\mathbb{V}[\varepsilon_t] = \Sigma$. Nous pourrions compléter le modèle avec des variables exogènes, une constante par exemple, mais nous allons à l'essentiel en omettant cette possibilité.

On note $\mathcal{Y}_T \equiv \{y_t\}_{t=-p+1}^T$ les données à notre disposition et on note z_t la concaténation horizontale des vecteurs lignes $y_{t-1}, y_{t-2}, \dots, y_{t-p}$. En concaténant verticalement les vecteurs lignes y_t, z_t et ε_t , pour $t = 1, \dots, T$, on obtient la représentation matricielle suivante du modèle VAR(p) :

$$Y = Z\mathcal{A} + E$$

où Y et E sont des matrices $T \times m$, Z est une matrice $T \times (mp)$ et $\mathcal{A} = (\mathbf{A}'_1, \mathbf{A}'_2, \dots, \mathbf{A}'_p)'$ la matrice $k \times m$ (avec $k = mp$) regroupant les coefficients auto-régressifs. La vraisemblance associée à ce modèle linéaire gaussien est donnée par :

$$\begin{aligned} \mathcal{L}(\mathcal{A}, \Sigma; \mathcal{Y}_T) &= (2\pi)^{-\frac{mT}{2}} |\Sigma|^{-\frac{T}{2}} \\ &\times e^{-\frac{1}{2} \text{tr}\{(Y-Z\mathcal{A})\Sigma^{-1}(Y-Z\mathcal{A})'\}} \end{aligned}$$

L'estimateur du maximum de vraisemblance (MCO) est défini par :

$$\hat{\mathcal{A}} = (Z'Z)^{-1}Z'Y$$

et

$$\hat{\Sigma} = T^{-1}(Y - Z\hat{\mathcal{A}})'(Y - Z\hat{\mathcal{A}})$$

Nous verrons plus loin qu'il est profitable de réécrire la vraisemblance en faisant apparaître l'estimateur des MCO :

$$\begin{aligned} \mathcal{L}(\mathcal{A}, \Sigma; \mathcal{Y}_T) &= (2\pi)^{-\frac{mT}{2}} \\ &\times |\Sigma|^{-\frac{k}{2}} e^{-\frac{1}{2} \text{tr}\{\Sigma^{-1}(\mathcal{A}-\hat{\mathcal{A}})'Z'Z(\mathcal{A}-\hat{\mathcal{A}})\}} \\ &\times |\Sigma|^{-\frac{T-k}{2}} e^{-\frac{1}{2} \text{tr}\{\Sigma^{-1}(Y-Z\hat{\mathcal{A}})'(Y-Z\hat{\mathcal{A}})\}} \end{aligned}$$

à des constantes d'intégration près on reconnaît les fonctions de densité de probabilité d'une gaussienne matricielle et d'une inverse Wishart (voir l'annexe A). On peut donc réécrire la vraisemblance sous la forme suivante :

$$\begin{aligned} \mathcal{L}(\mathcal{A}, \Sigma; \mathcal{Y}_T) &= (2\pi)^{-\frac{mT}{2}} \times (2\pi)^{\frac{km}{2}} |Z'Z|^{-\frac{m}{2}} \\ &\times f_{MN_{k,m}}(\mathcal{A}; \hat{\mathcal{A}}, (Z'Z)^{-1}, \Sigma) \\ &\times \frac{2^{\frac{\nu m}{2}} \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^m \Gamma\left(\frac{\nu+1-i}{2}\right)}{|\hat{S}|^{\frac{\nu}{2}}} \\ &\times f_{iW_m}(\Sigma; \hat{S}, \nu) \end{aligned}$$

avec $\nu = T - k - m - 1$ les degrés de liberté et $\hat{S} = T\hat{\Sigma}$. Cette écriture nous apprend que la vraisemblance du VAR(p) est proportionnelle au produit de la densité d'une

gaussienne matricielle et d'une inverse Wishart :

$$(6) \quad \mathcal{L}(\mathcal{A}, \Sigma; \mathcal{Y}_T) \propto f_{MN_{k,m}}(\mathcal{A}; \widehat{\mathcal{A}}, (Z'Z)^{-1}, \Sigma) \\ \times f_{i\mathcal{W}_m}(\Sigma; \widehat{S}, \nu)$$

Cette propriété va nous aider à poser une forme de la densité *a priori* telle que nous puissions obtenir une expression analytique de la densité postérieure.

3.0.1. *A priori non informatif.* Dans cette section nous allons nous supposons que nos croyances sont non informatives en adoptant un *a priori* à la Jeffrey :

$$(7) \quad p_0(\mathcal{A}, \Sigma) = |\Sigma|^{-\frac{m+1}{2}}$$

On note que dans le cas scalaire, $m = 1$, on retrouve le prior suggéré par Jeffrey ($1/\sigma^2$) décrit en cours. La densité *a posteriori* satisfait donc :

$$p(\mathcal{A}, \Sigma | \mathcal{Y}_T) \propto (2\pi)^{-\frac{mT}{2}} \times (2\pi)^{\frac{km}{2}} |Z'Z|^{-\frac{m}{2}} \\ \times f_{MN_{k,m}}(\mathcal{A}; \widehat{\mathcal{A}}, (Z'Z)^{-1}, \Sigma) \\ \times 2^{\frac{\nu m}{2}} \pi^{\frac{m(m-1)}{4}} |\widehat{S}|^{-\frac{\nu}{2}} \prod_{i=1}^m \Gamma\left(\frac{\nu+1-i}{2}\right) \\ \times f_{i\mathcal{W}_m}(\Sigma; \widehat{S}, \nu) \times |\Sigma|^{-\frac{m+1}{2}}$$

La densité jointe postérieure est donc proportionnelle au produit d'une gaussienne multivariée et d'une inverse Wishart :

$$(8) \quad p(\mathcal{A}, \Sigma; \mathcal{Y}_T) \propto f_{MN_{k,m}}(\mathcal{A}; \widehat{\mathcal{A}}, (Z'Z)^{-1}, \Sigma) \\ \times f_{i\mathcal{W}_m}(\Sigma; \widehat{S}, \tilde{\nu})$$

avec $\tilde{\nu} = T - k$. Ainsi, la densité postérieure peut s'écrire sous la forme suivante :

$$(9) \quad \mathcal{A} | \Sigma, \mathcal{Y}_T \sim MN_{k,m}(\widehat{\mathcal{A}}, \Sigma, (Z'Z)^{-1}) \\ \Sigma | \mathcal{Y}_T \sim i\mathcal{W}_m(\widehat{S}, \tilde{\nu})$$

Il n'est pas surprenant de constater que la distribution postérieure de \mathcal{A} (conditionnelle à la matrice de variance covariance) est centrée sur l'estimateur du maximum de vraisemblance, puisque notre *a priori* est non informatif. Nous pourrions montrer, en intégrant par rapport à Σ , que la distribution postérieure de \mathcal{A} est une version matricielle de la loi de Student. L'*a priori* de Jeffrey n'affecte que le degré de liberté de la distribution postérieure de \mathcal{A} . Dans cet exemple, nous pouvons caractériser la distribution postérieure « à la main ». Notons néanmoins que même si nous connaissons l'expression analytique de la distribution de \mathcal{A} et Σ , la construction des densités prédictives nécessite une approche par simulations, puisque les prévisions sont des fonctions non linéaires des matrices auto-régressives (dont nous connaissons la loi postérieure). L'intérêt pratique de l'approche bayésienne pourrait paraître peu évident dans ce cas, dans la mesure où la moyenne postérieure n'est pas différente de l'estimateur du maximum de vraisemblance.

3.0.2. *Un exemple d'a priori informatif.* Nous considérons maintenant un prior plus informatif qui va introduire un coin entre l'espérance postérieure et l'estimateur du maximum de vraisemblance ; dans un modèle linéaire gaussien, l'espérance postérieure est un mélange convexe de l'estimateur du maximum de vraisemblance et de l'espérance *a priori*. Afin d'aller à l'essentiel, nous adoptons une densité *a priori* dégénérée pour la matrice de variance-covariance des erreurs, en supposant que la

matrice Σ est connue (on posera $\Sigma = \widehat{\Sigma}$). Enfin nous spécifions le prior sur \mathcal{A} de la façon suivante :

$$(10) \quad p_0(\text{vec } \mathcal{A}) \sim \mathcal{N}(a_0, \Omega_0)$$

où Ω_0 est une matrice symétrique définie positive de dimension $mp \times mp$. En multipliant la vraisemblance par (10), on établit facilement que la distribution postérieure de $\text{vec } \mathcal{A}$ est gaussienne centrée en a_1 et de variance Ω_1 :

$$(11a) \quad \Omega_1 = (\Omega_0^{-1} + \Sigma^{-1} \otimes Z'Z)^{-1}$$

$$(11b) \quad a_1 = \Omega_1 \left[\Omega_0^{-1} a_0 + (\Sigma^{-1} \otimes Z'Z) \text{vec } \widehat{\mathcal{A}} \right]$$

Démonstration. La densité postérieure est proportionnelle au produit de la densité *a priori* et de la vraisemblance. Le noyau postérieur est donné par :

$$\begin{aligned} \mathcal{K}(\mathcal{A}|\mathcal{Y}_T) = \exp \left\{ -\frac{1}{2} \left[(\text{vec } \mathcal{A} - a_0)' \Omega_0^{-1} (\text{vec } \mathcal{A} - a_0) + \text{tr} \left(\Sigma^{-1} (\mathcal{A} - \widehat{\mathcal{A}})' Z'Z (\mathcal{A} - \widehat{\mathcal{A}}) \right) \right] \right\} \\ \times (2\pi)^{-\frac{km}{2}} |\Omega_0|^{-\frac{1}{2}} (2\pi)^{-\frac{mT}{2}} |\Sigma|^{-\frac{T}{2}} e^{-\frac{1}{2} \text{tr} \Sigma^{-1} \widehat{S}} \end{aligned}$$

Notons $a = \text{vec } \mathcal{A}$, $\widehat{a} = \text{vec } \widehat{\mathcal{A}}$ et $\mathcal{B}(a)$ le terme entre crochets sous la première exponentielle. En utilisant les propriétés des opérateurs vec , tr et du produit de kronecker nous avons :

$$\mathcal{B}(a) = (a - a_0)' \Omega_0^{-1} (a - a_0) + (a - \widehat{a})' (\Sigma^{-1} \otimes Z'Z) (a - \widehat{a})$$

En développant, il vient :

$$\begin{aligned} \mathcal{B}(a) = a' \Omega_0^{-1} a + a_0' \Omega_0^{-1} a_0 - 2a' \Omega_0^{-1} a_0 \\ + a' (\Sigma^{-1} \otimes Z'Z) a + \widehat{a}' (\Sigma^{-1} \otimes Z'Z) \widehat{a} - 2a' (\Sigma^{-1} \otimes Z'Z) \widehat{a} \end{aligned}$$

de façon équivalente il vient :

$$\begin{aligned} \mathcal{B}(a) = a' (\Omega_0^{-1} + \Sigma^{-1} \otimes Z'Z) a - 2a' (\Omega_0^{-1} a_0 + (\Sigma^{-1} \otimes Z'Z) \widehat{a}) \\ + a_0' \Omega_0^{-1} a_0 + \widehat{a}' (\Sigma^{-1} \otimes Z'Z) \widehat{a} \end{aligned}$$

En factorisant on trouve :

$$\mathcal{B}(a) = (a - a_1)' \Omega_1^{-1} (a - a_1) - a_1' \Omega_1^{-1} a_1 + a_0' \Omega_0^{-1} a_0 + \widehat{a}' (\Sigma^{-1} \otimes Z'Z) \widehat{a}$$

Finalement, en substituant dans l'expression du noyau, on peut réécrire celui-ci sous la forme suivante :

$$\begin{aligned} \mathcal{K}(\mathcal{A}|\mathcal{Y}_T) = \exp \left\{ -\frac{1}{2} (a - a_1)' \Omega_1^{-1} (a - a_1) \right\} \\ \times \exp \left\{ -\frac{1}{2} \left[a_0' \Omega_0^{-1} a_0 + \widehat{a}' (\Sigma^{-1} \otimes Z'Z) \widehat{a} - a_1' \Omega_1^{-1} a_1 \right] \right\} \\ \times (2\pi)^{-\frac{km}{2}} |\Omega_0|^{-\frac{1}{2}} (2\pi)^{-\frac{mT}{2}} |\Sigma|^{-\frac{T}{2}} e^{-\frac{1}{2} \text{tr} \Sigma^{-1} \widehat{S}} \end{aligned}$$

le premier terme correspond bien à l'expression (à une constante d'intégration près) d'une densité gaussienne pour a . En intégrant le noyau par rapport à a on obtient une expression analytique de la densité marginale :

$$\begin{aligned} p(\mathcal{Y}_T) &= \int \mathcal{K}(\mathcal{A}|\mathcal{Y}_T) d\mathcal{A} \\ &= (2\pi)^{\frac{km}{2}} |\Omega_1|^{-\frac{1}{2}} \\ &\times \exp \left\{ -\frac{1}{2} \left[a_0' \Omega_0^{-1} a_0 + \widehat{a}' (\Sigma^{-1} \otimes Z'Z) \widehat{a} - a_1' \Omega_1^{-1} a_1 \right] \right\} \\ &\times (2\pi)^{-\frac{km}{2}} |\Omega_0|^{-\frac{1}{2}} (2\pi)^{-\frac{mT}{2}} |\Sigma|^{-\frac{T}{2}} e^{-\frac{1}{2} \text{tr} \Sigma^{-1} \widehat{S}} \end{aligned}$$

□

La distribution postérieure de \mathcal{A} est donc gaussienne $\mathcal{N}(a_1, \Omega_1)$. L'interprétation de la distribution postérieure est directe. L'inverse de la variance postérieure (que l'on peut interpréter comme une quantification de l'information *a posteriori*) est égal à la somme de l'inverse de la variance *a priori* (l'information *a priori*) et de l'inverse de la variance de l'estimateur du maximum de vraisemblance de \mathcal{A} (l'information apportée par les données). *Ceteris paribus*, quand l'information *a priori* est importante, la matrice de variance-covariance Ω_0 est petite, la variance *a posteriori* est faible. L'espérance postérieure est un mélange de l'espérance *a priori*, a_0 , et de l'estimateur du maximum de vraisemblance, $\widehat{\mathcal{A}}$. Les pondérations respectives sont définies par le contenu informatif des croyances *a priori* et de l'échantillon. Lorsque l'information *a priori* tend vers l'infini, ie $\Omega_0 \rightarrow 0$, l'espérance postérieure tend vers l'espérance *a priori*. Lorsque l'information amenée par les données tend vers l'infini, ie $\Sigma^{-1} \otimes Z'Z \rightarrow 0$, l'espérance *a posteriori* tend vers l'estimateur du maximum de vraisemblance. On peut donc interpréter le paradigme bayésien comme un pont entre la calibration et l'estimation par maximum de vraisemblance. En notant que $Z'Z$ est généralement, si le modèle est stationnaire², un $\mathcal{O}(T)$, l'espérance postérieure tend vers l'estimateur du maximum de vraisemblance lorsque T tend vers l'infini.

ANNEXE A. DENSITÉS POUR LE MODÈLE BVAR

A.1. Distribution normale matricielle.

Définition 1. La matrice $p \times q$ aléatoire \mathbf{X} est distribuée conformément à une loi normale matricielle

$$\mathbf{X} \sim MN_{p,q}(\mathbf{M}, \mathbf{Q}, \mathbf{P})$$

où \mathbf{M} est une matrice $p \times q$, \mathbf{Q} et \mathbf{P} sont respectivement des matrices $p \times p$ et $q \times q$ symétriques et définies positives, si et seulement si $\text{vec}(\mathbf{X})$ est distribué comme une v.a. normale multivariée

$$\text{vec}(\mathbf{X}) \sim \mathcal{N}_{pq}(\text{vec}(\mathbf{M}), \mathbf{Q} \otimes \mathbf{P})$$

Ainsi, la fonction de densité associée à \mathbf{X} est donnée par :

$$f_{MN_{p,q}}(X; \mathbf{M}, \mathbf{P}, \mathbf{Q}) = (2\pi)^{-\frac{pq}{2}} |\mathbf{Q}|^{-\frac{p}{2}} |\mathbf{P}|^{-\frac{q}{2}} e^{-\frac{1}{2} \text{tr}\{\mathbf{Q}^{-1}(X-\mathbf{M})'\mathbf{P}^{-1}(X-\mathbf{M})\}}$$

A.2. Distributions de Wishart. La loi de Wishart est une version multivariée de la loi du χ^2 . Soit $\{X_i\}_{i=1}^\nu$ une suite de variables aléatoires gaussiennes indépendantes et identiquement distribuées $\mathcal{N}(0, Q)$, avec Q une matrice symétrique définie positive $q \times q$. Par définition $Y = \sum_{i=1}^\nu X_i X_i'$ est distribué selon une loi de Wishart. Les définitions suivantes caractérisent cette loi et la densité de l'inverse d'une v.a. de Wishart.

Définition 2. La matrice aléatoire, de dimension $q \times q$, symétrique et semi définie positive \mathbf{Y} est distribuée selon une loi de Wishart, $\mathbf{Y} \sim \mathcal{W}_q(\mathbf{Q}, \nu)$, si et seulement si sa densité est donnée par

$$f(Y; \mathbf{Q}, \nu) = \frac{|\mathbf{Q}|^{-\frac{\nu}{2}} |Y|^{\frac{\nu-q-1}{2}}}{2^{\frac{\nu q}{2}} \pi^{\frac{q(q-1)}{4}} \prod_{i=1}^q \Gamma\left(\frac{\nu+1-i}{2}\right)} e^{-\frac{1}{2} \text{tr}\{Y\mathbf{Q}^{-1}\}}$$

pour \mathbf{Q} une matrice symétrique semi définie positive, et $\nu \leq q$ le degré de liberté.

²La présence d'une racine unitaire ne ferait qu'accroître l'ordre de divergence, ce qui ne change pas qualitativement la conclusion.

Définition 3. Une matrice aléatoire, de dimension $q \times q$, \mathbf{X} est distribuée selon une loi inverse Wishart,

$$\mathbf{X} \sim i\mathcal{W}_q(\mathbf{Q}, \nu)$$

si et seulement si $\mathbf{X}^{-1} \sim \mathcal{W}_q(\mathbf{Q}^{-1}, \nu)$.

Ainsi la fonction de densité associée à \mathbf{X} est définie par :

$$f_{i\mathcal{W}_q}(X; \mathbf{Q}, \nu) = \frac{|\mathbf{Q}|^{\frac{\nu}{2}} |X|^{-\frac{\nu+q+1}{2}}}{2^{\frac{\nu q}{2}} \pi^{\frac{q(q-1)}{4}} \prod_{i=1}^q \Gamma\left(\frac{\nu+1-i}{2}\right)} e^{-\frac{1}{2}\text{tr}\{X^{-1}\mathbf{Q}\}}$$

ANNEXE B. RAPPELS D'ALGÈBRE POUR LE MODÈLE BVAR

B.1. L'opérateur vec. Soit X une matrice $m \times n$ formée en concaténant horizontalement les vecteurs colonnes x_1, x_2, \dots, x_n de dimensions $m \times 1$:

$$X = (x_1 | x_2 | \dots | x_n)$$

L'opérateur vec transforme une matrice en vecteur en concaténant verticalement les vecteurs colonnes formant cette matrice. Nous avons donc :

$$\text{vec } X = (x'_1, x'_2, \dots, x'_n)'$$

B.2. L'opérateur tr. Soit X une matrice carrée $m \times m$:

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & \dots & \dots & x_{1,m} \\ x_{2,1} & x_{2,2} & & & & x_{2,m} \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ x_{m,1} & \dots & \dots & \dots & \dots & x_{m,m} \end{pmatrix}$$

La trace d'une matrice carrée est la somme des scalaires sur sa diagonale. Ainsi, nous avons :

$$\text{tr } X = \sum_{i=1}^m x_{i,i}$$

B.2.1. propriété. Si A est un scalaire alors $\text{tr } A = A$.

B.2.2. propriété. Si A, B et C sont trois matrices de dimension $m \times p, p \times q$ et $q \times m$ alors $\text{tr } ABC = \text{tr } CAB$

B.3. Le produit de Kronecker. Soient A et B de matrices $m \times p$ et $n \times q$

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,p} \\ \vdots & & & \vdots \\ a_{m,1} & \dots & \dots & a_{m,p} \end{pmatrix}, \quad B = \begin{pmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,q} \\ \vdots & & & \vdots \\ b_{n,1} & \dots & \dots & b_{n,q} \end{pmatrix}$$

Le produit de kronecker de A par B est défini par :

$$A \otimes B = A = \begin{pmatrix} a_{1,1}B & a_{1,2}B & \dots & a_{1,p}B \\ \vdots & & & \vdots \\ a_{m,1}B & \dots & \dots & a_{m,p}B \end{pmatrix}$$

$A \otimes B$ est une matrice $mn \times pq$.

B.3.1. Propriété. Si A et B sont deux matrices carrées de plein rang, et donc inversibles, alors $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

B.3.2. Propriété. $\text{vec}(ABC) = (C' \otimes A)\text{vec } B$.

B.3.3. *Propriété.* Soient A , B , C et D , des matrices respectivement $m \times n$, $m \times p$, $p \times q$ et $n \times q$. On a alors $\text{tr}(A'BCD') = \text{vec}(A)'(D \otimes B)\text{vec}(C)$.

B.3.4. *Remarque.* Pour appliquer la dernière propriété, il est utile de noter que $\text{tr} A'BCD' = \text{tr} D'A'BC = \text{tr} CD'A'B = \text{tr} BCD'A'$ et que $\text{tr} A'BCD' = \text{tr}(A'BCD')'$ et donc que $\text{tr} A'BCD' = \text{tr} DC'B'A = \text{tr} ADC'B' = \text{tr} B'ADC' = \text{tr} C'B'AD$.