

Partie 2 : l'estimation bayésienne des DSGE

F. Karamé

TEPP (EPEE et CEE) & CEPREMAP (Dynare team)

Contribution à l'école ETEPP





Introduction : l'économétrie bayésienne des DSGE

1. Il est souvent difficile d'exprimer analytiquement la vraisemblance d'un DSGE mais l'évaluer numériquement est possible.

Il est difficile d'estimer un modèle DSGE par maximum de vraisemblance car les données sont peu informatives : limitées dans le temps, en nombre, ...

La vraisemblance peut se révéler plate dans certaines directions et cela peut poser des problèmes d'identification.

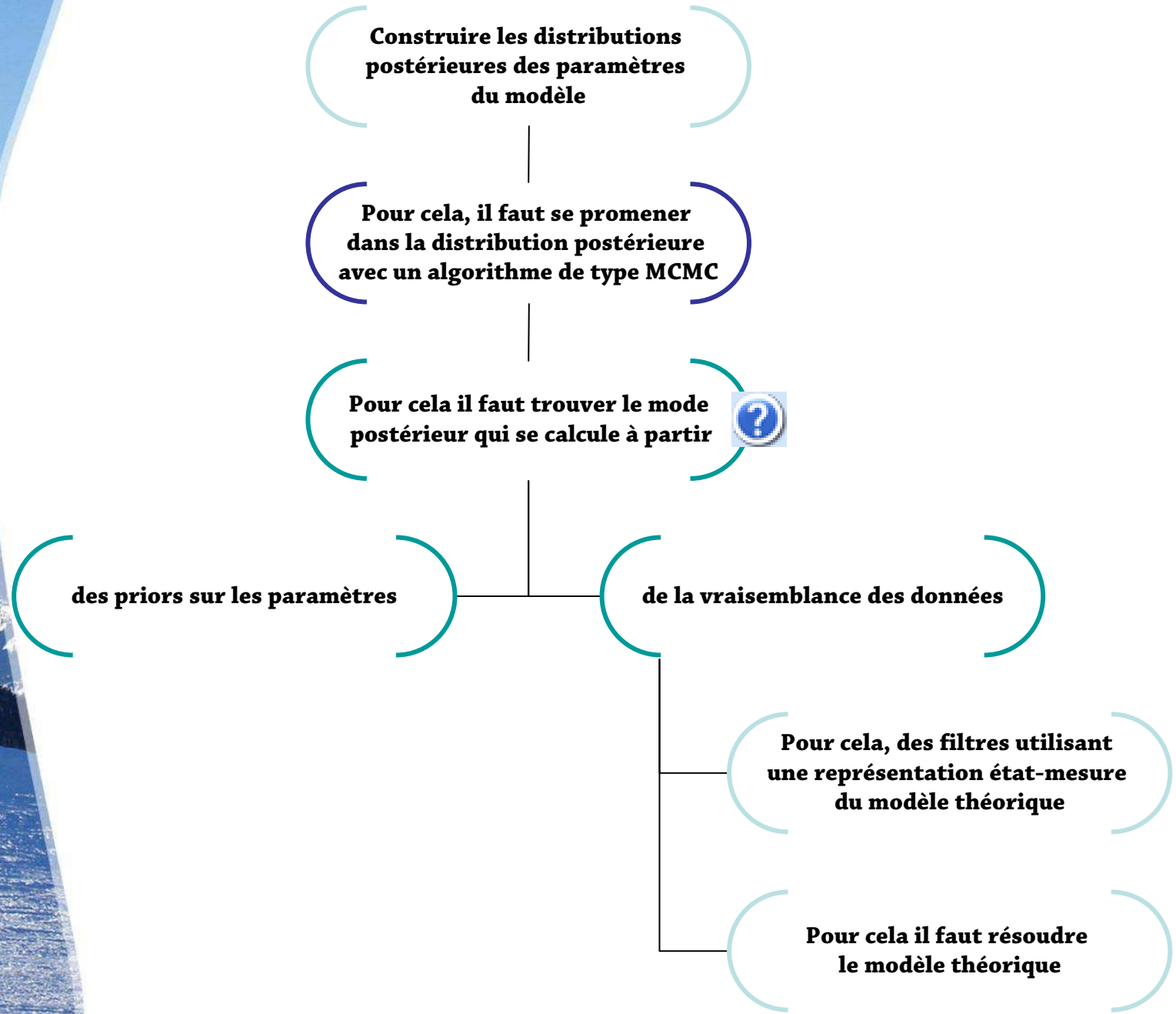
Il faut donc trouver d'autres sources d'information : l'utilisation de *priors* de l'approche bayésienne.

2. Les DSGE sont des modèles mal spécifiés.

Estimer un modèle mal spécifié de façon bayésienne avec des *priors* non informatifs (uniformes par exemple) peut souvent conduire à des résultats peu crédibles.

L'utilisation des *priors* peut donc modifier notre vision de la vraisemblance et conduire à des résultats plus interprétables en manipulant la distribution *a posteriori* des paramètres.

3. L'utilisation de *priors* informatifs permet de réduire l'incertitude *a posteriori* (en termes de variance).



Trouver le mode postérieur

On s'intéresse à la distribution *a posteriori* des paramètres du modèle selon la traditionnelle **formule de Bayes** :

$$p(\theta | \underline{y}_{1:T}) = \frac{L(\underline{y}_{1:T} | \theta) \cdot p(\theta)}{p(\underline{y}_{1:T})} \propto L(\underline{y}_{1:T} | \theta) \cdot p(\theta)$$

à une constante multiplicative près indépendante des paramètres $p(\underline{y}_{1:T}) = \int_{\theta_i} L(\underline{y}_{1:T} | \theta_i) \cdot p(\theta_i) d\theta_i$

L'estimation dite classique ou fréquentiste consiste à trouver les paramètres associés au maximum de vraisemblance :

$$\hat{\theta}_T^{ML} = \arg \max_{\theta} L(\underline{y}_{1:T} | \theta)$$

Une première estimation bayésienne consiste à déterminer **le mode de la distribution postérieure** des paramètres :

$$\begin{aligned} \hat{\theta}_T^{mod} &= \arg \max_{\theta} p(\theta | \underline{y}_{1:T}) \\ &= \arg \max_{\theta} \left\{ L(\underline{y}_{1:T} | \theta) \cdot p(\theta) \right\} \\ &= \arg \max_{\theta} \left\{ \ln L(\underline{y}_{1:T} | \theta) + \ln p(\theta) \right\} \end{aligned}$$



**Construire les distributions
postérieures des paramètres
du modèle**

**Pour cela, il faut se promener
dans la distribution postérieure
avec un algorithme de type MCMC**

**Pour cela il faut trouver le mode
postérieur qui se calcule à partir**



des priors sur les paramètres

de la vraisemblance des données

**Pour cela, des filtres utilisant
une représentation état-mesure
du modèle théorique**

**Pour cela il faut résoudre
le modèle théorique**



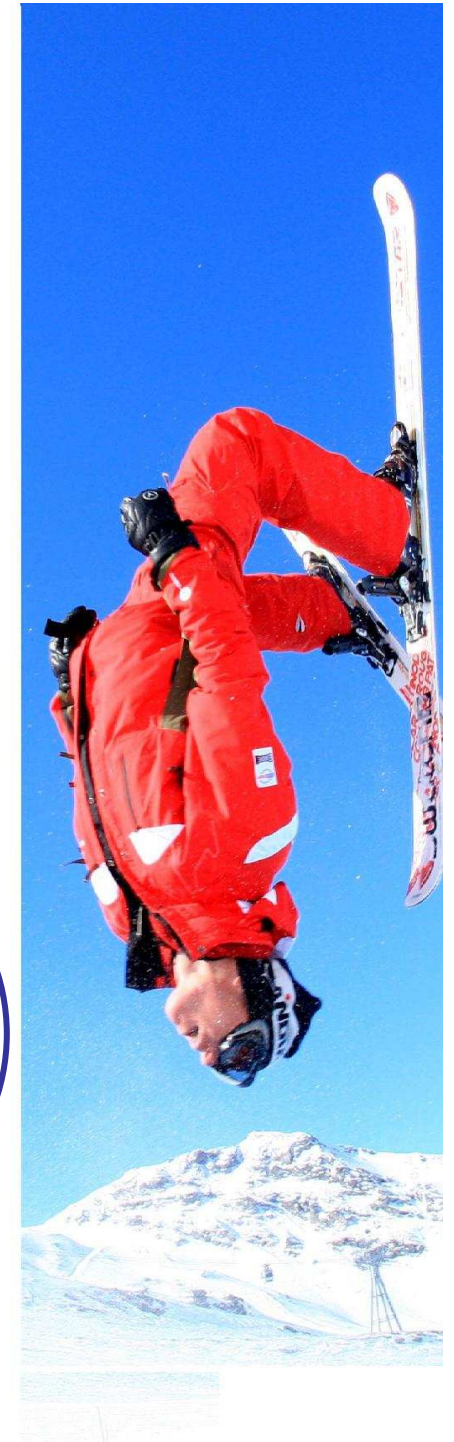
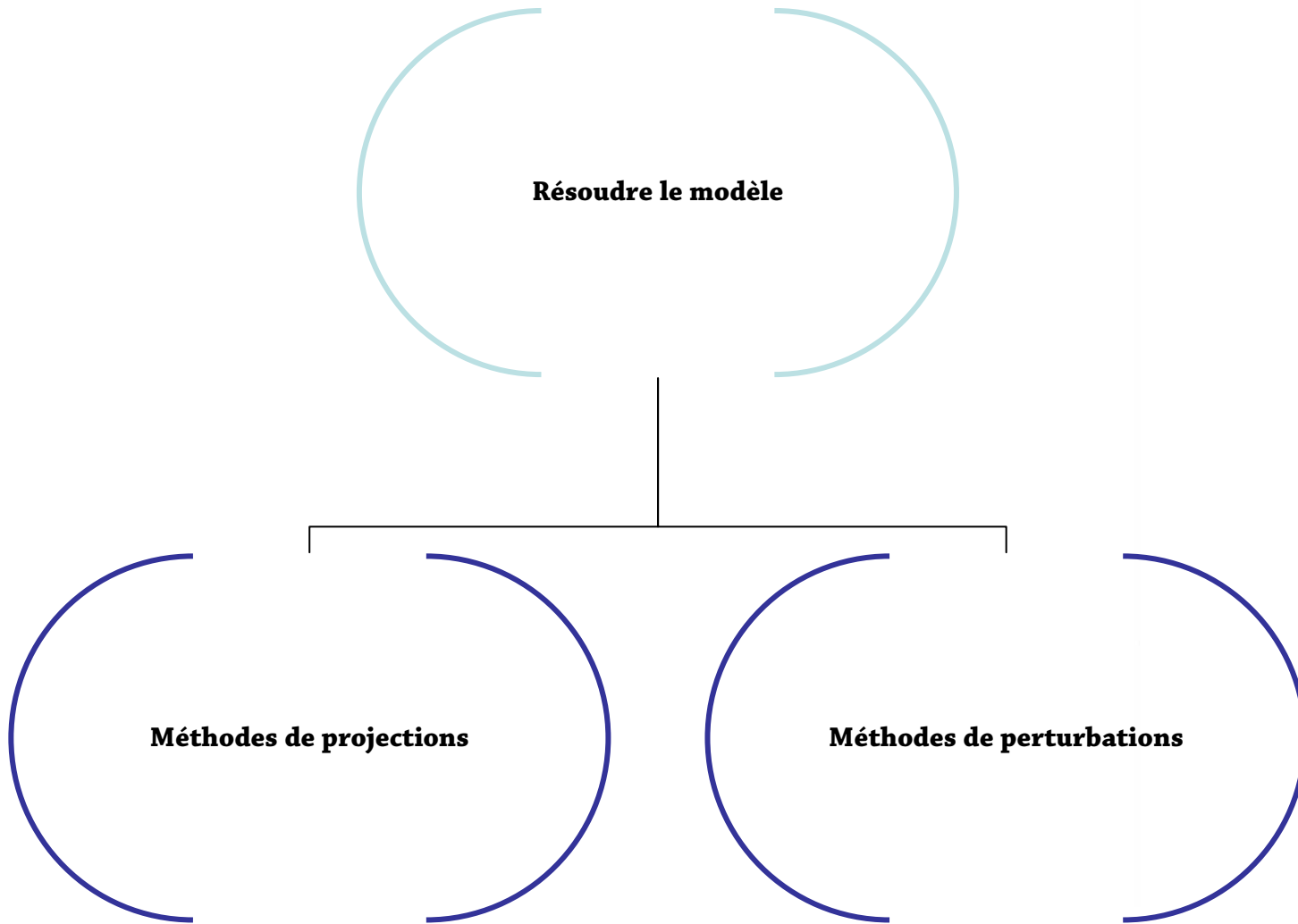


Illustration : le modèle de base

$$\max_{\{c_t, l_t\}_{t=0}^{\infty}} U = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, l_t) = E_0 \sum_{t=0}^{\infty} \beta^t \frac{(c_t^\theta (1-l_t)^{1-\theta})^{1-\tau}}{1-\tau}$$

subject to

$$y_t = c_t + i_t$$

$$y_t = e^{a_t} k_t^\alpha l_t^{1-\alpha}$$

$$k_{t+1} = i_t + (1-\delta)k_t$$

$$a_{t+1} = \rho a_t + e_{t+1} \quad \text{where } e_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_e).$$

$$k_{t+1} = e^{a_t} k_t^\alpha l_t^{1-\alpha} - c_t + (1-\delta)k_t$$

$$a_{t+1} = \rho a_t + \epsilon_t$$

$$\frac{(c_t^\theta (1-l_t)^{1-\theta})^{1-\tau}}{c_t} - \beta E_t \left(\frac{(c_{t+1}^\theta (1-l_{t+1})^{1-\theta})^{1-\tau}}{c_{t+1}} \left(1-\delta + \alpha \frac{e^{a_{t+1}} k_{t+1}^\alpha l_{t+1}^{1-\alpha}}{k_{t+1}} \right) \right) = 0.$$

$$\frac{1-\theta}{\theta} \frac{c_t}{1-l_t} - (1-\alpha) \frac{e^{a_t} k_t^\alpha l_t^{1-\alpha}}{l_t} = 0$$

$$k_t = e^{a_{t-1}} k_{t-1}^\alpha x^l(k_{t-1}, a_{t-1})^{1-\alpha} - x^c(k_{t-1}, a_{t-1}) + (1-\delta)k_{t-1}$$

$$a_t = \rho a_{t-1} + e_t.$$

$$y_t = e^{a_t} k_t^\alpha x^l(k_t, a_t)^{1-\alpha} + \epsilon_t^y$$

$$l_t = x^l(k_t, a_t) + \epsilon_t^l$$

$$i_t = e^{a_t} k_t^\alpha x^l(k_t, a_t)^{1-\alpha} - x^c(k_t, a_t) + \epsilon_t^i.$$

⇔

$$s_{t+1} = g(s_t, x^*(s_t), e_{t+1}) = g^*(s_t, e_{t+1})$$

$$y_t = m(s_t, \epsilon_t)$$

$$s_t \equiv \{k_t, a_t\} \quad e_t \equiv e_t$$

Les méthodes de perturbations (ou locales)

Idée : on approxime numériquement le modèle en faisant un développement de Taylor localement autour de l'état stationnaire déterministe.

Ecrivons le modèle sous la forme suivante (connue) :

$$E_t[f(y_{t+1}, y_t, y_{t-1}, u_t)] = 0$$

et posons la solution (inconnue) :

$$y_t = g(y_{t-1}, u_t, \sigma)$$

Le modèle et sa solution respectent l'état stationnaire :

$$f(\bar{y}, \bar{y}, \bar{y}, 0) = 0$$

$$\bar{y} = g(\bar{y}, 0, 0)$$

En substituant, on peut définir :

$$F_g(y_{t-1}, u_t, u_{t+1}, \sigma) = f(\overbrace{g(g(y_{t-1}, u_t, \sigma), u_{t+1}, \sigma)}^{y_{t+1}}, \overbrace{g(y_{t-1}, u_t, \sigma)}^{y_t}, y_{t-1}, u_t)$$

et donc réécrire le problème comme :

$$E_t[F_g(y_{t-1}, u_t, u_{t+1}, \sigma)] = 0$$

Plusieurs ordres d'approximation existent.

Le développement de Taylor à l'ordre 1 :

$$0 \approx F_g^{(1)}(y_-, u, u_+, \sigma) = f_{y+}(g_y(g_y \hat{y} + g_u u + g_\sigma \sigma) + g_u u_+ + g_\sigma \sigma) \\ + f_y(g_y \hat{y} + g_u u + g_\sigma \sigma) + f_{y-} \hat{y} + f_u u$$

avec :

$$\hat{y} = y_{t-1} - \bar{y} \quad u = u_t \quad u_+ = u_{t+1}$$

$$f_{y+} = \frac{\partial f}{\partial y_{t+1}} \quad f_y = \frac{\partial f}{\partial y_t} \quad f_{y-} = \frac{\partial f}{\partial y_{t-1}} \quad f_u = \frac{\partial f}{\partial u_t}$$

$$g_y = \frac{\partial g}{\partial y_t} \quad g_u = \frac{\partial g}{\partial u_t} \quad g_\sigma = \frac{\partial g}{\partial \sigma}$$

En prenant l'espérance conditionnelle et en regroupant, il vient :

$$0 \approx (f_{y+} g_y g_y + f_y g_y + f_{y-}) \cdot \hat{y} + (f_{y+} g_y g_u + f_y g_u + f_u) \cdot u \\ + (f_{y+} g_y g_\sigma + f_{y+} g_\sigma + f_y g_\sigma) \cdot \sigma$$

On est en présence d'un système de trois blocs d'équations à trois blocs d'inconnus qui est satisfait si :

$$\begin{cases} f_{y+} g_y g_y + f_y g_y + f_{y-} = 0 \\ f_{y+} g_y g_u + f_y g_u + f_u = 0 \\ f_{y+} g_y g_\sigma + f_{y+} g_\sigma + f_y g_\sigma = 0 \end{cases} \Leftrightarrow \begin{cases} f_{y+} g_y g_y + f_y g_y + f_{y-} = 0 \\ f_{y+} g_y g_u + f_y g_u + f_u = 0 \\ g_\sigma = 0 \end{cases} \Rightarrow \text{décomposition de} \\ \text{Schur généralisée} \rightarrow g_y(\theta) \text{ et } g_u(\theta)$$

$$y_t = \bar{y} + g_y(\theta) \cdot \hat{y}_{t-1} + g_u(\theta) \cdot u_t$$

A l'ordre 2 :

$$\begin{aligned} F_g^{(2)}(y_-, u, u_+, \sigma) &= F_g^{(1)}(y_-, u, u_+, \sigma) \\ &+ 0,5 \cdot [F_{y_- y_-}(\hat{y} \otimes \hat{y}) + F_{uu}(u \otimes u) + F_{u_+ u_+}(u_+ \otimes u_+) + F_{\sigma\sigma}\sigma^2] \\ &+ F_{y_- u}(\hat{y} \otimes u) + F_{y_- u_+}(\hat{y} \otimes u_+) + F_{y_- \sigma}\hat{y}\sigma \\ &+ F_{u u_+}(u \otimes u_+) + F_{u\sigma}u\sigma \\ &+ F_{u_+ \sigma}u_+\sigma \end{aligned}$$

avec $F_{..}$ des matrices de dérivées du vecteur de fonctions. En prenant l'espérance conditionnelle et en réorganisant, il vient :

$$\begin{aligned} 0 &\approx E_t[F_g^{(1)}(y_-, u, u_+, \sigma)] + 0,5 \cdot [F_{y_- y_-}(\hat{y} \otimes \hat{y}) + F_{uu}(u \otimes u) + (F_{u_+ u_+} \Sigma_\varepsilon + F_{\sigma\sigma})\sigma^2] \\ &+ F_{y_- u}(\hat{y} \otimes u) + F_{y_- \sigma}\hat{y}\sigma + F_{u\sigma}u\sigma \end{aligned}$$

satisfait si $F_{y_- y_-} = 0, F_{uu} = 0, (F_{u_+ u_+} \Sigma_\varepsilon + F_{\sigma\sigma}) = 0, F_{y_- u} = 0, F_{y_- \sigma} = 0$ et $F_{u\sigma} = 0$

Ces matrices dépendent des dérivées simples et secondes de f (**connues**) et de g (**inconnues**) et de **la solution à l'ordre 1**. Elles définissent 6 blocs d'équations à résoudre numériquement.

La solution à l'ordre 2 est alors de la forme :

$$\begin{aligned} y_t &= \bar{y} + 0,5 \cdot g_{\sigma\sigma}(\theta) \cdot \sigma^2 + g_y(\theta) \cdot \hat{y}_{t-1} + g_u(\theta) \cdot u_t \\ &+ 0,5 \cdot g_{yy}(\theta) \cdot (\hat{y}_{t-1} \otimes \hat{y}_{t-1}) + 0,5 \cdot g_{uu}(\theta) \cdot (u_t \otimes u_t) + 0,5 \cdot g_{yu}(\theta) \cdot (\hat{y}_{t-1} \otimes u_t) \end{aligned}$$

Les méthodes de projections (ou globales)

On veut approximer les règles de décision $f(\cdot)$ en fonction des variables d'état s .

On suppose la forme de la fonction permettant d'ajuster les règles de décision :

$$f(\tilde{s}) = T(s) \cdot c \quad \text{avec} \quad \begin{array}{l} \tilde{s} \in [s_{min}, s_{max}] \\ s \in [-1, 1] \end{array} \quad \text{et} \quad s = \varphi(\tilde{s})$$

Idée : on construit une grille multidimensionnelle sur un intervalle de valeurs des états et on cherche le vecteur c permettant d'ajuster les valeurs des règles en les points de la grille. Cela permettra ensuite d'avoir une fonction paramétrée pour extrapoler les règles de décision pour toute valeur des états.

Problème : on se donne des valeurs pour les états mais on n'a rien sur les règles. On sait seulement que lorsque les valeurs prises par les règles sont cohérentes avec celles des états, les CPO sont nulles.

La méthode de la **collocation orthogonale** :

$$r(s_i; \hat{c}) = 0 \quad i = 0, \dots, n$$

avec $r(\cdot, \cdot)$ le système des CPO du modèle calculé pour le point i de la grille (donc au total la dimension du problème est le nombre des CPO multiplié par le nombre de nœuds).

Critère : le vecteur solution fournit la valeur des règles qui annule les CPO du modèle en chaque point de la grille multidimensionnelle sur les états.

Les polynômes de Chebyshev : cas d'une seule variable d'état

$$\forall i=0, \dots, n$$

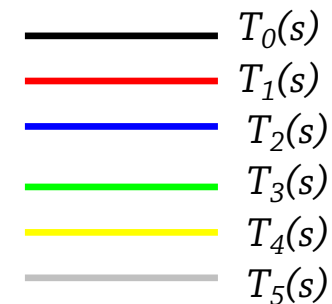
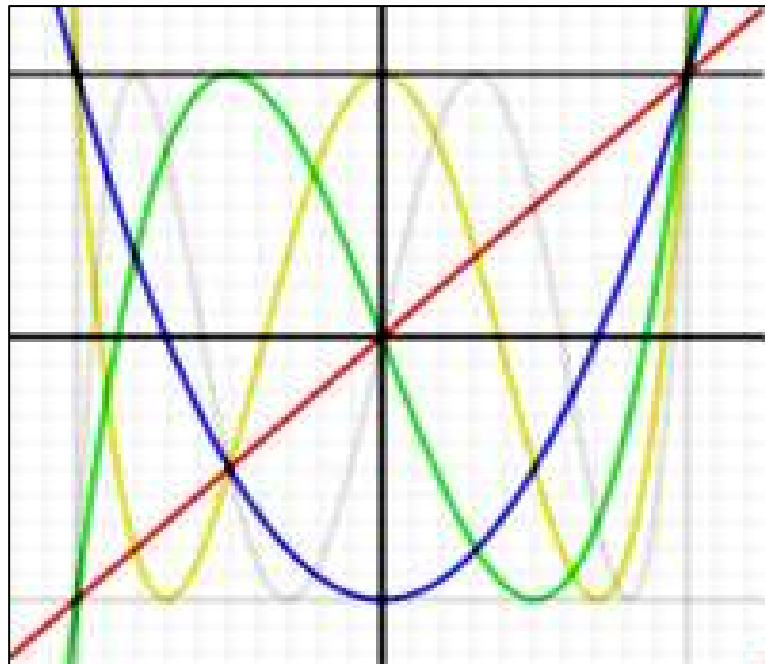
$$s_i = \cos\left(\frac{\pi(i+0.5)}{n+1}\right) \in [-1, 1]$$

$$\tilde{s}_i = \varphi^{-1}(s_i) \in [s_{\min}, s_{\max}]$$

$$\forall i=0, \dots, n$$

$$T_{i+1}(s) = 2sT_i(s) - T_{i-1}(s)$$

$$\text{avec } T_0(s) = 1 \text{ et } T_1(s) = s$$



$$\begin{bmatrix} T_0(s_0) & T_1(s_0) & T_2(s_0) \\ T_0(s_1) & T_1(s_1) & T_2(s_1) \\ T_0(s_2) & T_1(s_2) & T_2(s_2) \end{bmatrix} \cdot \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} f(\tilde{s}_0) \\ f(\tilde{s}_1) \\ f(\tilde{s}_2) \end{bmatrix} \Leftrightarrow T(s) \cdot c = f(\tilde{s}) \Leftrightarrow \begin{array}{l} c = T^{-1}(s) \cdot f(\tilde{s}) \\ \text{ou si } T(s) \text{ rectangulaire} \\ c = [T(s)' \cdot T(s)]^{-1} \cdot T(s)' \cdot f(\tilde{s}) \end{array}$$

Comment passer au cas multidimensionnel ?

Cela revient à choisir **un opérateur de combinaison** des nœuds et des polynômes de Chebychev.

Le plus général : le **produit tensoriel** :

$$\forall i = 0, \dots, n, \quad \forall j = 1, \dots, k$$

$$s_{i,j} = \cos\left(\frac{\pi(i+0.5)}{n+1}\right)$$

$$s = \left\{s_{i,1}\right\}_{i=0}^n \times \left\{s_{i,2}\right\}_{i=0}^n \times \dots \times \left\{s_{i,k}\right\}_{i=0}^n$$

$$\forall i = 0, \dots, n \quad \forall j = 1, \dots, k$$

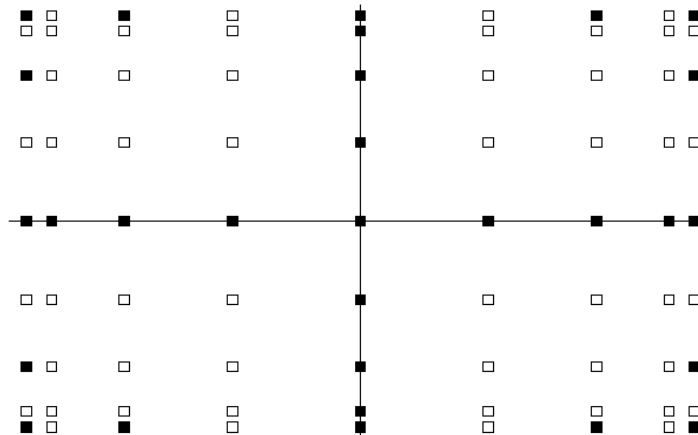
$$T_{i+1,j}(s_j) = 2s_j T_{i,j}(s_j) - T_{i-1,j}(s_j)$$

$$\text{avec } T_{0,j}(s_j) = 1 \text{ et } T_{1,j}(s_j) = s_j$$

$$T(s_i) = T_1(s_{i,1})T_2(s_{i,2})\dots T_k(s_{i,k})$$

Mais **'curse of dimensionality'** : la taille du problème à résoudre croît très vite.

Une alternative : **l'opérateur Smolyak** pour le choix des nœuds (-> grille creuse) et des puissances (-> polynômes complets). Par exemple, pour les nœuds :





Construire les distributions postérieures des paramètres du modèle

Pour cela, il faut se promener dans la distribution postérieure avec un algorithme de type MCMC

Pour cela il faut trouver le mode postérieur qui se calcule à partir



des priors sur les paramètres

de la vraisemblance des données

Pour cela, des filtres utilisant une représentation état-mesure du modèle théorique



Pour cela il faut résoudre le modèle théorique



Filtres bayésiens

Représentation
linéaire et
bruit gaussien

Filtre de Kalman
standard

Représentation
non linéaire
gaussienne

Filtre de Kalman
et méthodes de
Monte-Carlo

Filtre de Kalman et
quadratures

Représentation
non linéaire
non gaussienne

Avec
rééchantillonnage :
Filtre à particules

Sans
rééchantillonnage :
Filtre à particules
gaussien

La formule des probabilités totales (Chapman & Kolmogorov)

$$p(x_t) = \int p(x_t | x_{t-1}) \cdot p(x_{t-1}) \cdot dx_{t-1}$$



Comment calculer la vraisemblance ?

$$\begin{aligned} p(\underline{y_{1:T}} | \theta) &= \prod_{t=1}^T p(y_t | \underline{y_{1:t-1}}; \theta) \\ &= p(y_1 | \theta) \prod_{t=2}^T p(y_t | \underline{y_{1:t-1}}; \theta) \\ &= \int p(y_1 | s_0; \theta) \cdot p(s_0) \cdot ds_0 \prod_{t=2}^T \int p(y_t | s_t; \theta) \cdot p(s_t | \underline{y_{1:t-1}}; \theta) \cdot ds_t \end{aligned}$$

Hypothèse : on connaît la distribution initiale des états s_0 .

Constat : on peut calculer la vraisemblance si on connaît la distribution *a priori* des états à la date t .

La distribution *a priori* des états

$$\begin{aligned} p(s_t | \underline{y_{1:t-1}}) &= \int p(s_t, s_{t-1} | \underline{y_{1:t-1}}) \cdot ds_{t-1} \\ &= \int p(s_t | s_{t-1}, \underline{y_{1:t-1}}) \cdot p(s_{t-1} | \underline{y_{1:t-1}}) \cdot ds_{t-1} \\ &= \int \underline{p(s_t | s_{t-1})} \cdot p(s_{t-1} | \underline{y_{1:t-1}}) \cdot ds_{t-1} \end{aligned}$$

$s_t = g^*(s_{t-1}, e_t; \theta)$ **les équations d'état du modèle résolu**
+ des tirages dans la loi supposée des chocs

La distribution *a posteriori* des états

$$\begin{aligned} p(s_t | \underline{y_{1:t}}) &= p(s_t | y_t, \underline{y_{1:t-1}}) \\ &= \frac{p(y_t | s_t, \underline{y_{1:t-1}}) \cdot p(s_t | \underline{y_{1:t-1}})}{p(y_t | \underline{y_{1:t-1}})} \\ &= \frac{p(y_t | s_t) \cdot p(s_t | \underline{y_{1:t-1}})}{p(y_t | \underline{y_{1:t-1}})} \\ &= \frac{\underline{p(y_t | s_t)} \cdot p(s_t | \underline{y_{1:t-1}})}{\int \underline{p(y_t | s_t)} \cdot p(s_t | \underline{y_{1:t-1}}) \cdot ds_t} \end{aligned}$$

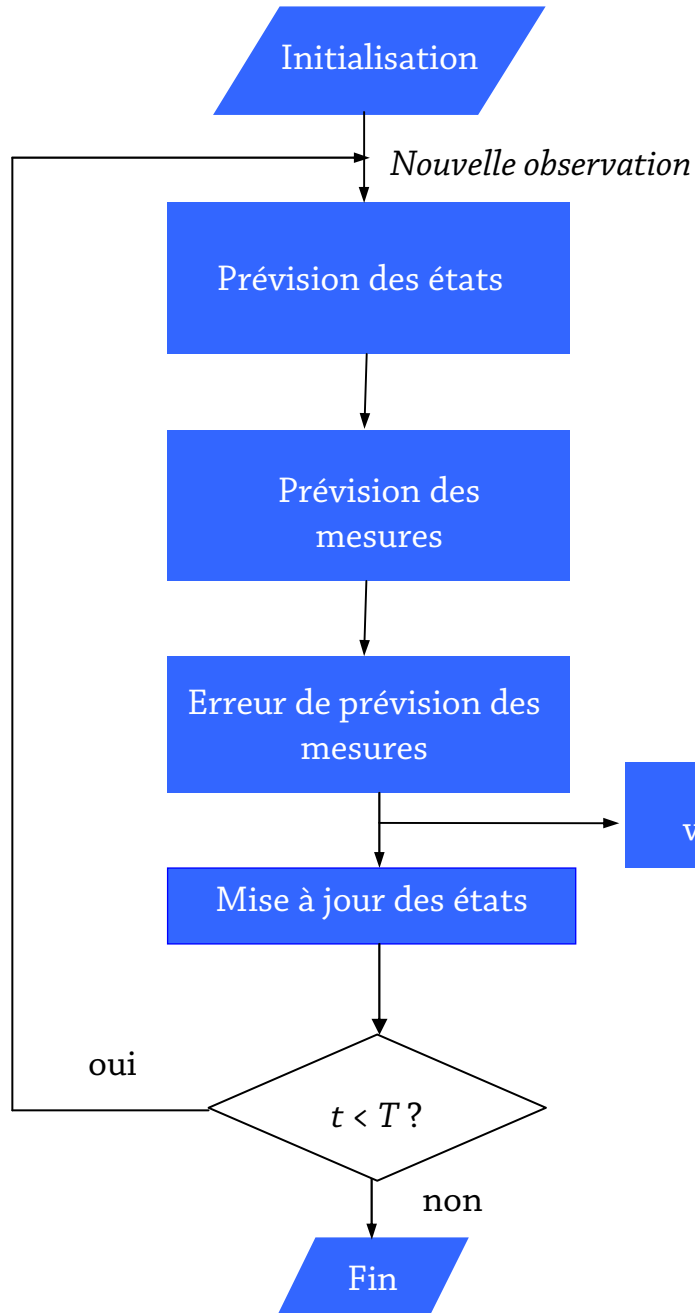
$y_t = m(s_t, \varepsilon_t; \theta)$ **les équations de mesure du modèle**
+ une hypothèse sur la loi des erreurs
pour écrire la vraisemblance

Sous l'hypothèse de linéarité du modèle et de normalité des chocs, tout se simplifie et donne le filtre de **Kalman**.

Implication sympathiques :

- Comme le modèle est linéaire, toutes les distributions sont gaussiennes. On manipule uniquement les résumés de la distribution des états et des mesures : l'espérance et la variance.
- Comme le modèle est linéaire, il suffit de dire que l'espérance des chocs est nulle.

Le filtre de Kalman



$$\begin{cases} s_t = g^*(s_{t-1}, e_t; \theta) \\ y_t = m(s_t, \varepsilon_t; \theta) \end{cases} \equiv \begin{cases} s_t = A s_{t-1} + e_t \\ y_t = Z s_t + \varepsilon_t \end{cases}$$

$$\begin{cases} \hat{s}_{t|t-1} = E[s_t | \underline{y}_{1:t-1}] = A \cdot \hat{s}_{t-1|t-1} \\ P_{t|t-1} = V[s_t | \underline{y}_{1:t-1}] = A \cdot P_{t-1|t-1} \cdot A' + Q \end{cases}$$

$$\begin{cases} \hat{y}_{t|t-1} = E[y_t | \underline{y}_{1:t-1}] = Z \cdot \hat{s}_{t|t-1} \\ \eta_t = y_t - \hat{y}_{t|t-1} \\ G_t = V[y_t | \underline{y}_{1:t-1}] = Z \cdot P_{t|t-1} \cdot Z' + Q \end{cases}$$

$$\ln L_t(y_t | \theta) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |G_t| - \frac{1}{2} \eta_t' \cdot G_t^{-1} \cdot \eta_t$$

$$\begin{cases} K_t = P_{t|t-1} \cdot Z' \cdot G_t^{-1} \\ \hat{s}_{t|t} = E[s_t | \underline{y}_{1:t}] = \hat{s}_{t|t-1} + K_t \cdot \eta_t \\ P_{t|t} = V[s_t | \underline{y}_{1:t}] = (I - K_t \cdot Z) \cdot P_{t|t-1} \end{cases}$$

Sans l'hypothèse de linéarité du modèle et de normalité des chocs, on doit utiliser un filtre à **particules**

Implications beaucoup moins sympathiques :

- on manipule **toute la distribution** des états et des mesures : revient à appliquer le modèle état-mesure à chaque particule qui constitue l'approximation de ces distributions pour les faire évoluer au cours du temps.
- on doit aussi **tirer les chocs des équations d'état** car dans un modèle non linéaire, il ne suffit plus de dire que l'espérance des chocs est nulle, il faut faire des tirages aléatoires...

Comment calculer des intégrales du type $E_p[g(x)] = \int g(x) \cdot p(x) \cdot dx$?

On approxime $p(\cdot)$ par une **mesure aléatoire**, c'est-à-dire on la discrétise par un plus ou moins grand nombre de noeuds x_i et de poids w_i :

$$\hat{p}(x) = \sum_{i=1}^N w_i \cdot \delta(x - x_i)$$

$$E_p[g(x)] \approx \sum_{i=1}^N w_i \cdot g(x_i)$$

Dans le cas d'une quadrature, les noeuds et les poids sont optimaux.

Dans le cas de tirages de Monte-Carlo, on tire N noeuds x_i avec des poids équiprobables $w_i = 1/N$.

L'échantillonnage d'importance

Que faire quand la loi $p(\cdot)$ n'est pas connue et/ou qu'on ne peut pas tirer dedans mais qu'on peut quand même l'évaluer ?

On suppose une **distribution d'importance** $q(\cdot)$ dans laquelle il est facile d'effectuer des tirages aléatoires.

Dès lors, l'intégrale peut s'écrire :

$$\begin{aligned} E_p [g(x)] &= \int g(x) \cdot p(x) \cdot dx \\ &= \int g(x) \cdot \frac{p(x)}{q(x)} \cdot q(x) \cdot dx \\ &= E_q \left[g(x) \cdot \frac{p(x)}{q(x)} \right] \end{aligned}$$

Si on peut approximer $q(\cdot)$ par quadrature ou par Monte-Carlo ($w_i = 1/N$)

$$\hat{q}(x) = \sum_{i=1}^N w_i \cdot \delta(x - x_i)$$

Alors il vient :

$$E_p [g(x)] \approx \sum_{i=1}^N w_i \cdot g(x_i) \cdot \frac{p(x_i)}{q(x_i)}$$

Si $p(\cdot)$ et $q(\cdot)$ ne somment pas à 1, on introduit une **normalisation** :

$$E_p[g(x)] = \frac{\int g(x) \cdot \tilde{p}(x) \cdot dx}{\int \tilde{p}(x) \cdot dx} = \frac{\int g(x) \cdot \frac{\tilde{p}(x)}{\tilde{q}(x)} \cdot \tilde{q}(x) \cdot dx}{\int \frac{\tilde{p}(x)}{\tilde{q}(x)} \cdot \tilde{q}(x) \cdot dx}$$

$$\approx \frac{\frac{1}{N} \sum_{i=1}^N g(x_i) \cdot \frac{\tilde{p}(x_i)}{\tilde{q}(x_i)}}{\frac{1}{N} \sum_{i=1}^N \frac{\tilde{p}(x_i)}{\tilde{q}(x_i)}} = \sum_{i=1}^N g(x_i) \cdot \tilde{w}_i$$

$$\tilde{w}_i = \frac{\frac{\tilde{p}(x_i)}{\tilde{q}(x_i)}}{\sum_{i=1}^N \frac{\tilde{p}(x_i)}{\tilde{q}(x_i)}}$$

Sequential Importance Sampling (SIS) Particle Filter

A la date t , on approxime les distributions *a priori* et *a posteriori* des états conditionnellement à l'information présente et passée par une mesure aléatoire

$$\{s_t^{(i)}, w_t^{(i)}\}_{i=1}^N$$

La distribution d'importance nous permettra de tirer les particules :

$$s_t^{(i)} \sim q(s_t^{(i)} | s_{t-1}^{(i)}, y_t)$$

Comment obtient-on les poids ?

$$\begin{aligned}w_t^{(i)} &= \frac{p(s_{0:t}^{(i)} | \underline{y_{1:t}})}{q(s_{0:t}^{(i)} | \underline{y_{1:t}})} \\&= p(y_t | \underline{y_{1:t-1}}, s_t^{(i)}) \cdot \frac{p(s_t^{(i)} | s_{0:t-1}^{(i)})}{q(s_t^{(i)} | s_{0:t-1}^{(i)}, \underline{y_{1:t}})} \cdot \frac{p(s_{0:t-1}^{(i)} | \underline{y_{1:t-1}})}{q(s_{0:t-1}^{(i)} | \underline{y_{1:t-1}})} \\&= w_{t-1}^{(i)} \cdot p(y_t | \underline{y_{1:t-1}}, s_t^{(i)}) \cdot \frac{p(s_t^{(i)} | s_{0:t-1}^{(i)})}{q(s_t^{(i)} | s_{0:t-1}^{(i)}, \underline{y_{1:t}})} \\&= w_{t-1}^{(i)} \cdot p(y_t | s_t^{(i)}) \cdot \frac{p(s_t^{(i)} | s_{t-1}^{(i)})}{q(s_t^{(i)} | s_{t-1}^{(i)}, \underline{y_{1:t}})}\end{aligned}$$

Le choix usuel pour la densité d'importance

$$q(s_t | s_{t-1}, \underline{y_{1:t}}) = p(s_t | s_{t-1}) \Rightarrow w_t^{(i)} = w_{t-1}^{(i)} \cdot p(y_t | s_t^{(i)})$$

la théorie

$$p(s_t | \underline{y_{1:t-1}}) = \int p(s_t | s_{t-1}) \cdot p(s_{t-1} | \underline{y_{1:t-1}}) \cdot ds_{t-1}$$

la contrepartie empirique

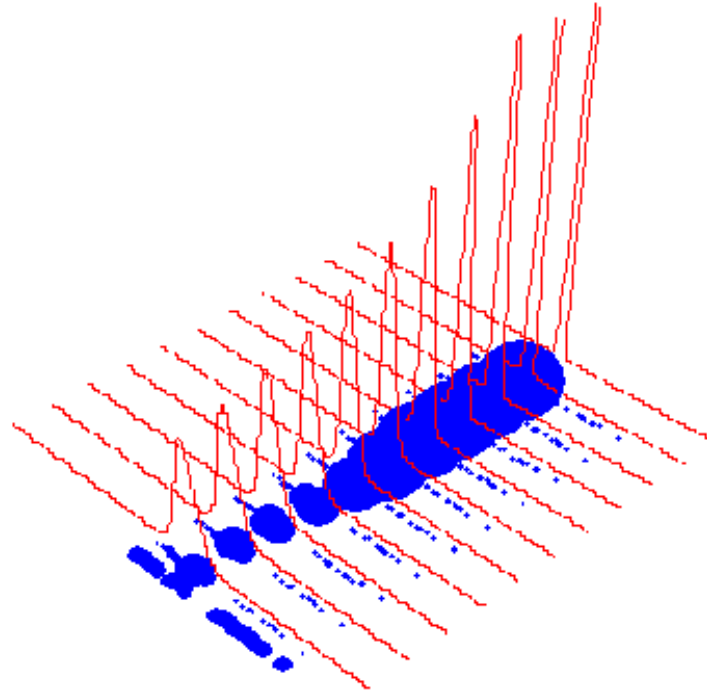
$$p(s_{t-1} | \underline{y_{1:t-1}}) \leftrightarrow \left\{ s_{t-1}^{(i)}, w_{t-1}^{(i)} \right\}_{i=1}^N$$

$$s_t^{(i)} \sim p(s_t | s_{t-1}) \leftrightarrow s_t^{(i)} = g^*(s_{t-1}^{(i)}, e_t; \theta)$$

$$w_t^{(i)} = w_{t-1}^{(i)} \cdot p(y_t | s_t^{(i)}) \leftrightarrow y_t^{(i)} = m(s_t^{(i)}, \varepsilon_t; \theta)$$

Problème : la dégénérescence du filtre

La variance des poids **augmente indéfiniment** avec le temps : au bout de quelques itérations temporelles, il ne reste plus qu'une seule particule avec un poids non nul !



$$N_{eff,t} = \frac{1}{\sum_{i=1}^N (w_t^{(i)})^2}$$

La solution d'usage : introduire une méthode de ré-échantillonnage (avec remise) dépendant du poids des particules pour éliminer celles à poids faible et multiplier celles à poids élevé.

Rééchantillonnage par la méthode inverse

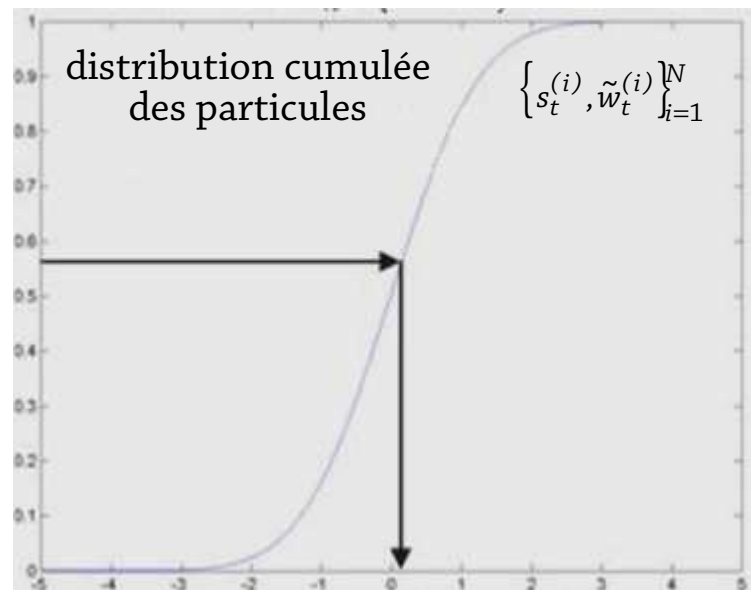
Soit u suit une loi uniforme $U_{[0,1]}$.

Soit $F^{-1}(\cdot)$ l'inverse de la fonction de répartition $F(\cdot)$.

$F^{-1}(u)$ suit la loi de F .

Tirage aléatoire de

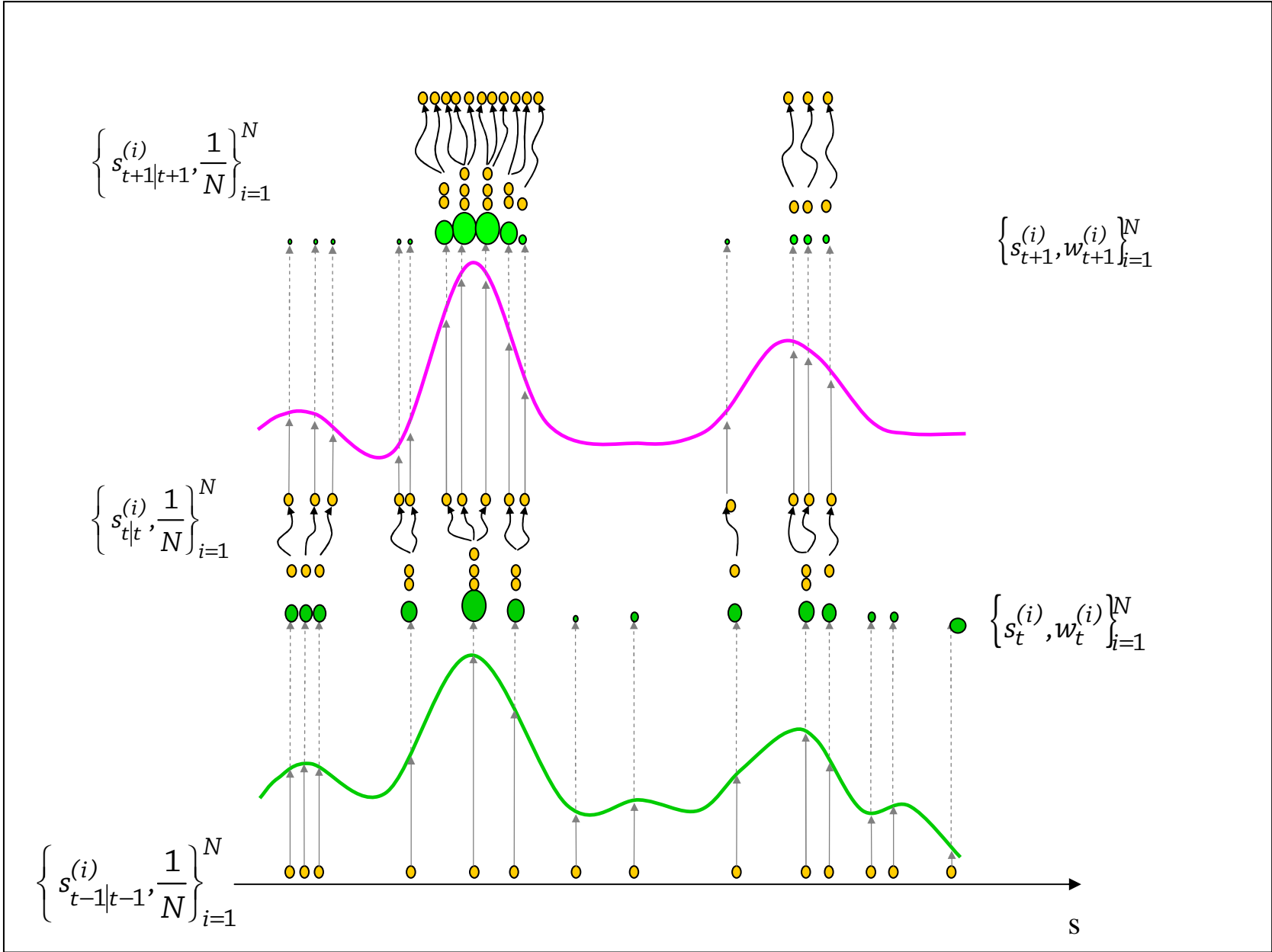
$$\left\{ u_t^{(i)} \right\}_{i=1}^N$$



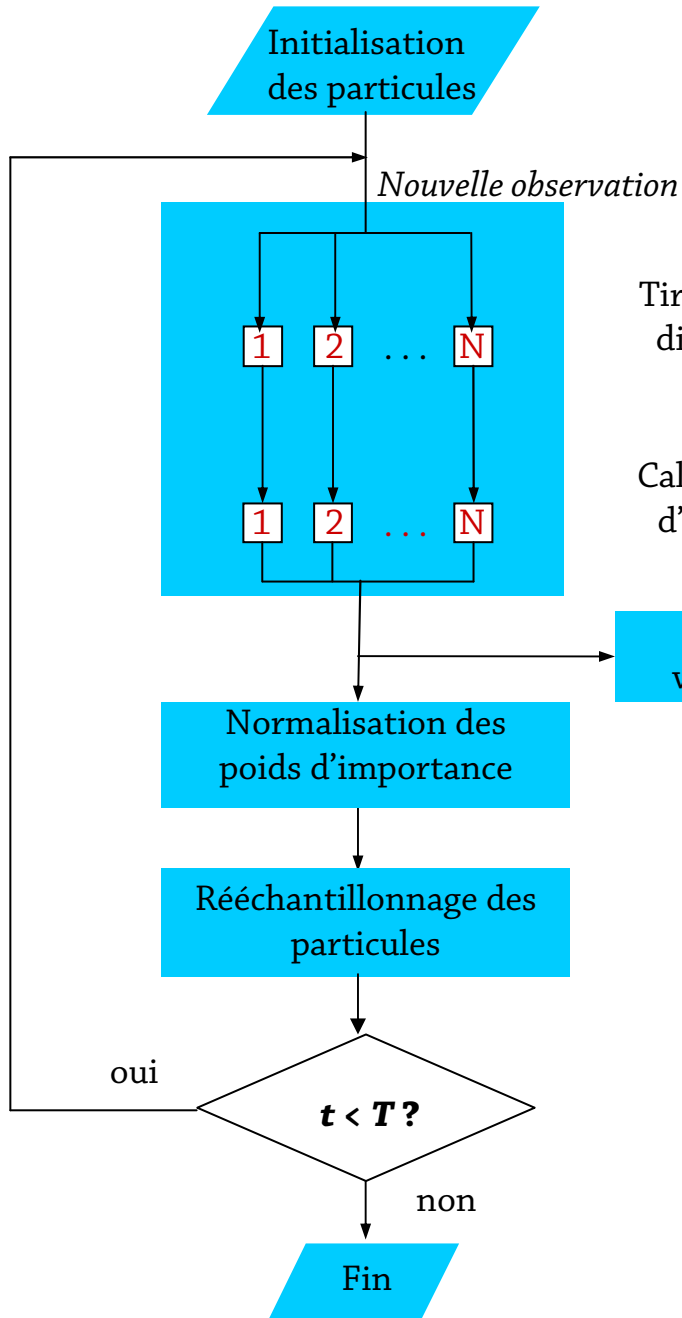
$$\tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}}$$

Particules rééchantillonnées avec remise

$$\left\{ s_{t|t}^{(i)}, \frac{1}{N} \right\}_{i=1}^N$$



Algorithme du Bootstrap Particle Filter



$$p(s_{t-1} | \underline{y}_{1:t-1}) \leftrightarrow \left\{ s_{t-1|t-1}^{(i)}, \frac{1}{N} \right\}_{i=1}^N$$

Tirage des particules dans la distribution d'importance

$$s_t^{(i)} \sim p(s_t | s_{t-1}) \leftrightarrow s_t^{(i)} = g^*(s_{t-1|t-1}^{(i)}, e_t; \theta)$$

Calcul des poids d'importance

$$w_t^{(i)} \propto p(y_t | s_t^{(i)}) \leftrightarrow y_t^{(i)} = m(s_t^{(i)}, \varepsilon_t; \theta)$$

Calcul de la vraisemblance

$$p(\underline{y}_{1:T} | \theta) \approx \prod_{t=1}^T \frac{1}{N} \sum_{i=1}^N p(y_t | s_t^{(i)}) = \prod_{t=1}^T \frac{1}{N} \sum_{i=1}^N w_t^{(i)}$$

Normalisation des poids d'importance

$$\tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}}$$

Rééchantillonnage des particules

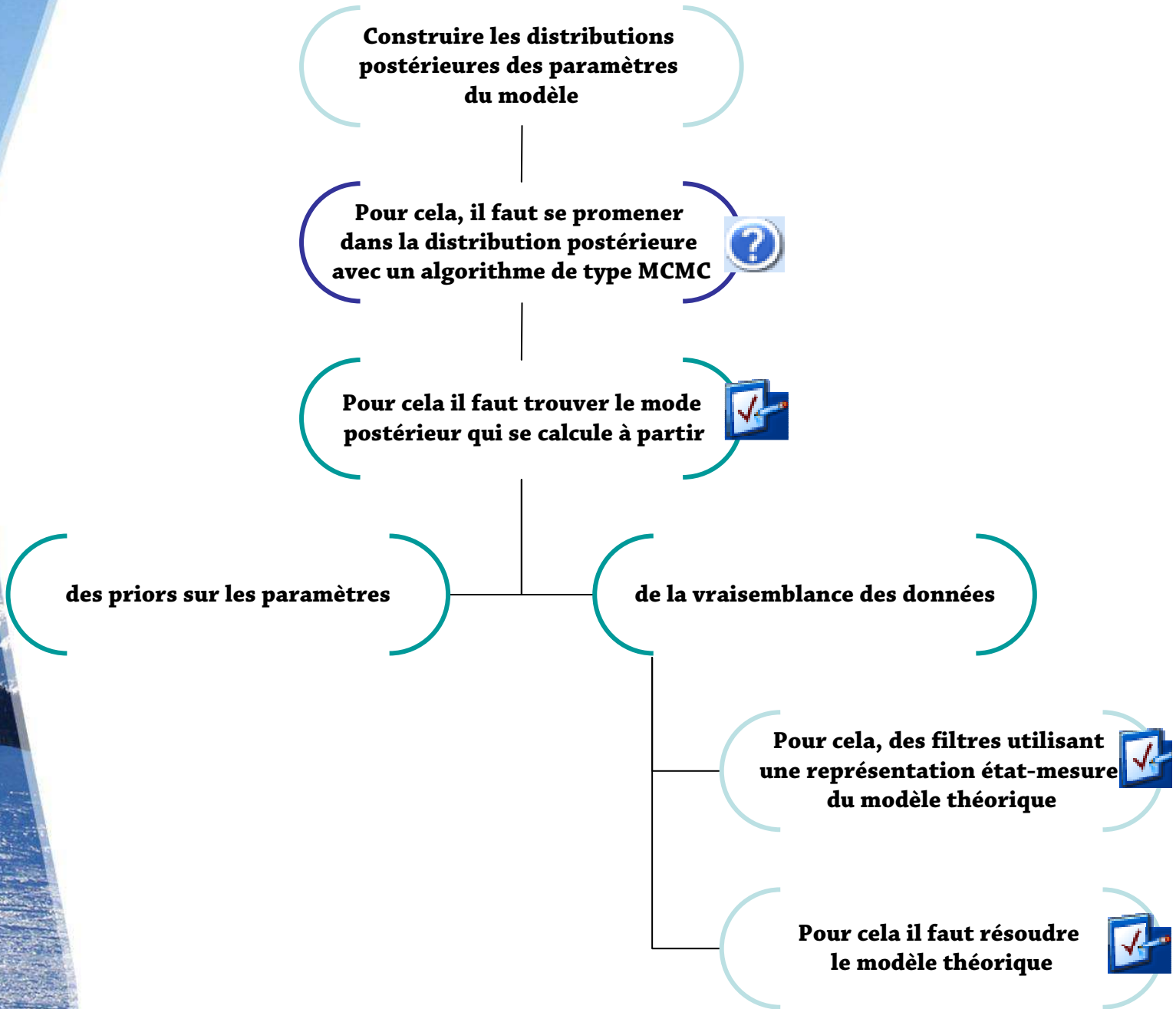
$$\left\{ s_t^{(i)}, \tilde{w}_t^{(i)} \right\}_{i=1}^N \rightarrow \left\{ s_{t|t}^{(i)}, \frac{1}{N} \right\}_{i=1}^N$$

oui

$t < T?$

non

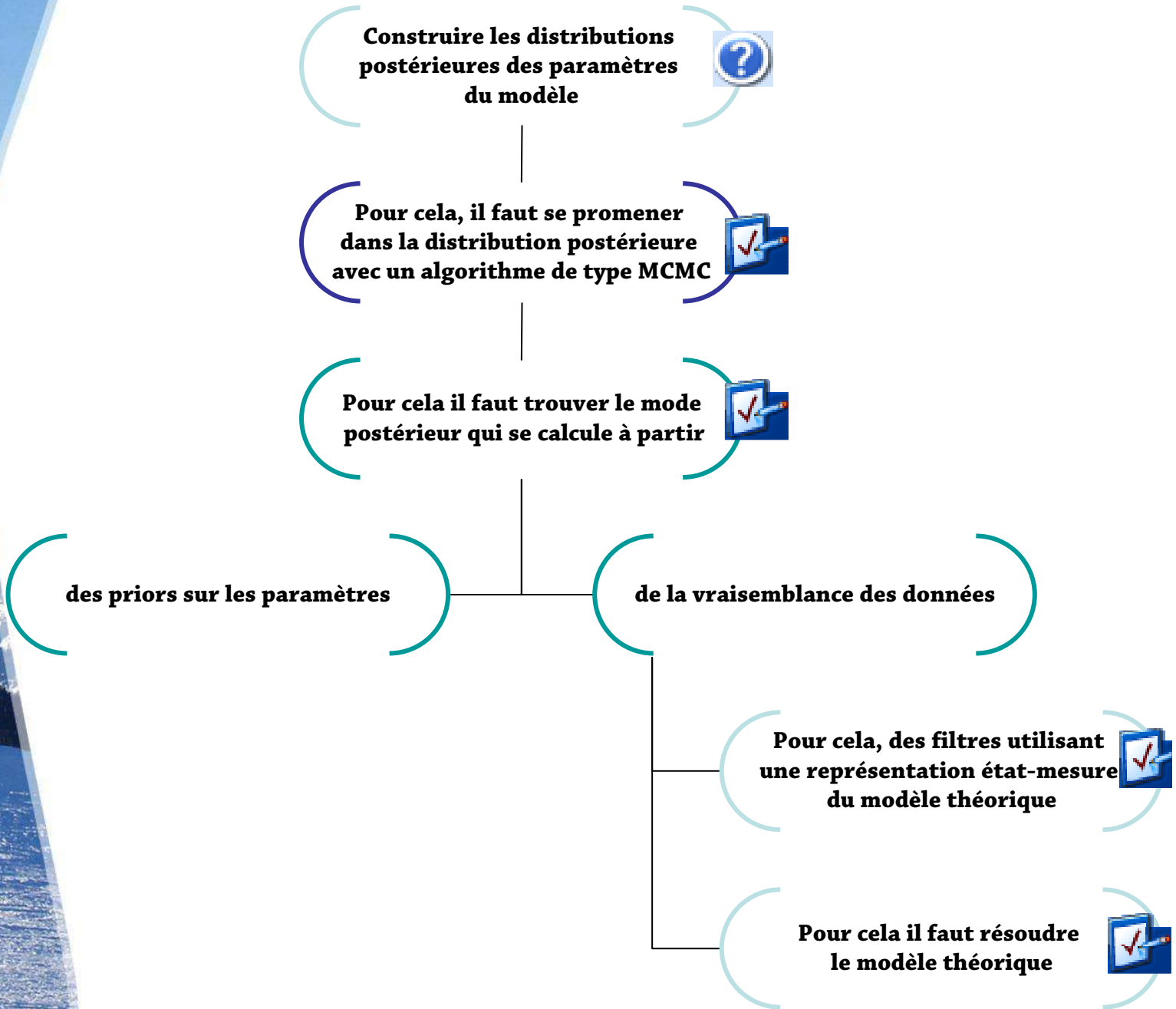
Fin



Explorer la densité postérieure : l'algorithme de Métropolis

$$\hat{\theta}_n^* = \hat{\theta}_{n-1} + v \quad v \approx N(0, \gamma^{RW} \hat{V}(\hat{\theta}_T^{mod})) \quad \hat{\theta}_0 = \hat{\theta}_T^{mod}$$
$$\hat{\theta}_n = \begin{cases} \hat{\theta}_n^* & \text{si } U(0,1) \leq \min \left\{ 1, \frac{p(\hat{\theta}_n^* | \underline{y}_{1:T})}{p(\hat{\theta}_{n-1} | \underline{y}_{1:T})} \right\} \quad (\text{acceptation}) \\ \hat{\theta}_{n-1} & \text{sinon (rejet)} \end{cases}$$

- L'algorithme à pas aléatoire peut être mono-chaîne ou multi-chaînes indépendantes.
- L'approche multi-chaînes permet de calculer des indicateurs de convergence vers la distribution stationnaire.
- La marche aléatoire s'interprète comme une distribution d'importance pour tirer les candidats.
- γ^{RW} est déterminé de façon à avoir 30% d'acceptation environ.
- Le stockage (après convergence) des candidats acceptés permet d'obtenir une approximation de la densité postérieure des paramètres du modèle.
- Les tirages des chocs exogènes et les probabilités uniformes du ré-échantillonnage sont conservés identiques pendant les itérations de l'algorithme de Métropolis.



Présentation construite en partie à partir de :

Adjemian S., 2007, Estimation of DSGE models, université du Mans (Gains) & CEPREMAP.

Miodrag B., University of Ottawa, Canada.

Références bibliographiques :

An S. & F. Schorfheide, 2006, Bayesian Analysis of DSGE Models, Econometric Reviews 26, 113-172.

Arulampalam A.S., S. Maskell, N. Gordon & T. Clapp, 2002, A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking, IEEE Transactions on Signal Processing 50, 174-188.

Doucet A., N. de Freitas & N. Gordon, 2001, Sequential Monte Carlo Methods in Practice, Springer Verlag.

Durbin J. & S.J. Koopman, 2002, Time Series Analysis by State-Space Methods, Oxford University Press.

Fernández-Villaverde J. & J.F. Rubio-Ramírez, 2008, How Structural are Structural Parameters?, NBER Macroeconomics Annual 2007, 83-137.

Fernandez-Villaverde J., 2009, The Econometrics of DSGE Models, Working paper, n°14677, NBER.

Harvey A.C., 1989, Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press.

Haug A.J., 2005, A Tutorial on Bayesian Estimation and Tracking Techniques Applicable to Nonlinear and Non-Gaussian Processes, MITRE Technical Report #W400.

Judd K., 1992, Projection Methods for Solving Aggregate Growth Models, Journal of Economic Theory 58, 410-452.

Schmitt-Grohé S. & M. Uribe, 2004, Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function, Journal of Economic Dynamics and Control 28, 755-775.

Winschel V. & M. Krätzig, 2008, Solving, Estimating and Selecting Nonlinear Dynamic Models without the Curse of Dimensionality, forthcoming Econometrica.