Markov Chain Monte Carlo.

6-601-09 Simulatio Monte Carlo

Geneviève Gauthier

HEC Montréal

Référence

Ces notes de cours sont inspirées de

1. S. M. Ross, *Simulation* (1997), Academic Press.
2. L. Stentoft (2009), Notes de cours 80-214-99 *Numerical methods in Finance*.

Pour une introduction aux chaînes de Markov, on peut consulter

1. G. Gauthier, Notes de cours 3-602-84 *Modèles probabilistes et stochastiques de la gestion*. http://neumann.hec.ca/~p240/

Chaîne de Markov I

Définitions

Definition

Un processus stochastique $\{X_t : t = 0, 1, 2, ...\}$ à temps discret et prenant les valeurs $1, ...m$ est une **chaîne de Markov homogène** si, quelque soit t,

$$P\left[X_{t+1} = j \,|X_t = 1\right] = q_{ij}.$$

Definition

La matrice $Q = (q_{ij})$ de format $n \times n$ est appelée la **matrice de transition**.

Chaîne de Markov II

Définitions

Definition

Une chaîne de Markov est dite **irréductible** si, pour toutes les paires (i, j), la probabilité de passé de l'état i à l'état j est strictement positive.

Definition

Pour une chaîne de Markov irréductible, il est possible de définir la **proportion à long-terme** π_j **du temps passé dans l'état** j,

1. $\sum_{j=1}^{m} \pi_j = 1$
2. $\pi_j = \sum_{i=1}^{m} q_{ij} \pi_i$.

Definition

Les probabilités π_j^* sont appelées les **probabilités stationnaires** si le fait d'entrer dans la chaîne avec la distribution π^* implique que X_t est de distribution π^* quelque soit t. Plus précisément, si

$$P\left(X_0 = j\right) = \pi_j^*, \ \forall j = 1, ..., m,$$

alors

$$\forall t = 1, 2, ..., \ P\left(X_t = j\right) = \pi_j^*.$$

1. Notons que

$$\begin{aligned}
\pi_j^* &= P\left(X_t = j\right) \\
&= \sum_{i=1}^{m} P\left(X_t = j \middle| X_{t-1} = i\right) P\left(X_{t-1} = i\right) \\
&= \sum_{i=1}^{m} q_{ij} \pi_i^* .
\end{aligned}$$

2. Constatons que les probabilités stationnaires satisfont le même système d'équations linéaires que les proportions à long terme du temps passé dans les états. Nous utiliserons donc la même notation pour les deux concepts.

Il existe une méthode qui permet d'obtenir les probabilité stationnaire autrement qu'en résolvant le système linéaire $\pi_j = \sum_{i=1}^m q_{ij}\pi_i$.

Theorem

Supposons qu'il existe des nombres positifs $x_1, ..., x_m$ satisfaisant $x_i q_{ij} = x_j q_{ji}$ et $\sum_i x_i = 1$ alors $\pi_j = x_j$.

Preuve.

$$\sum_{i=1}^m q_{ij} x_i = \sum_{i=1}^m q_{ij} \frac{x_j q_{ji}}{q_{ij}} = x_j \sum_{i=1}^m q_{ji} = x_j.$$

Puisque $\pi_1, ..., \pi_m$ est l'unique solution du système d'équations linéaires $\pi_j = \sum_{i=1}^m q_{ij}\pi_i$, alors $x_i = \pi_i$. $\square$

Une propriété importante de ce type de chaîne de Markov est:

Theorem

Pour toute fonction $h : \{1, ..., m\} \to \mathbb{R}$,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} h(X_t) = \sum_{i=1}^{m} h(i)\, \pi_i \text{ presque sûrement.}$$

Idée de la preuve.

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T} h(X_t) &= \sum_{i=1}^{m} h(i) \frac{\text{nb de périodes passées dans l'état } i}{T} \\ &\overset{T\to\infty}{\longrightarrow} \sum_{i=1}^{m} h(i)\, \pi_i . \; \square \end{aligned}$$

MCMC
G. Gauthier
Introduction
Chaîne de Markov
MCMC
Ex.: estimation
Algorithme de Hasting-Metropolis
Gibbs Sampler
Exemple

Definition

Une chaîne de Markov irréductible est **apériodique** si pour tout état $j = 1, ..., m$, il existe un instant T tel que

$$P\left(X_T = j \,|X_0 = j\,\right) > 0 \text{ et } P\left(X_{T+1} = j \,|X_0 = j\,\right) > 0$$

Lorsque la chaîne de Markov est irréductible et apériodique, la probabilité stationnaire est aussi la probabilité à long terme, c'est-à-dire que

$$\lim_{T\to\infty} P\left(X_T = j\right) = \pi_j, \qquad j = 1, 2, ..., m.$$

En général... I

1. Supposons que l'on veut générer une variable aléatoire X ayant pour distribution la fonction de masse π, c'est-à-dire que
$$P\left(X = x_i\right) = \pi_i.$$
Il serait possible de le faire en simulant la chaîne de Markov $\{X_t : t = 0, 1, 2, ...\}$ et conserver X_T pour T suffisamment grand.
2. Si le but est d'estimer $\theta = \mathrm{E}\left[h\left(X\right)\right]$, il sera possible de le faire en utilisant $\frac{1}{T}\sum_{t=1}^{T} h\left(X_t\right)$.
    1. Afin d'éliminer l'effet de la loi d'entrée dans la chaîne, il est aussi possible d'utiliser
$$\widehat{\theta} = \frac{1}{T - t_0}\sum_{t=t_0+1}^{T} h\left(X_t\right).$$

En général... II

3. Puisque les X_t ne sont plus indépendants, il n'est pas possible de calculer la marge d'erreur comme on le ferait avec un échantillon de variables aléatoires iid. Comment peut-on mesurer la précision de notre estimateur?

En général... III

- Le but est d'estimer l'espérance des erreur au carré (**M**ean **S**quare **E**rror)

$$\mathrm{MSE} = \mathrm{E}\left[\left(\widehat{\theta} - \theta\right)^2\right].$$

- L'idée principale est de briser l'échantillon de $T - t_0$ scénarios en s sous-échantillons de taille $n = \frac{T-t_0}{s}$.
- Poser $\widehat{\theta}_k = \frac{1}{n}\sum_{t=t_0+(k-1)n+1}^{t_0+kn} h\left(X_t\right)$.
- $\widehat{\theta}_1, ..., \widehat{\theta}_s$ seront traités comme s'ils étaient iid, même si ce n'est pas le cas (plus n est grand, plus cela est vraisemblable).
- La variance échantillonnale $\widehat{S}^2 = \frac{1}{s-1}\sum_{k=1}^{s}\left(\widehat{\theta}_k - \overline{\widehat{\theta}}\right)$ est alors un estimateur de $\mathrm{Var}\left[\widehat{\theta}_1\right]$.
- L'estimateur de MSE est $\widehat{\mathrm{MSE}} = \frac{\widehat{S}^2}{s}$.

Exemple d'application - Première partie I

Nous avons un échantillon $\mathbf{Y} = (Y_1, ..., Y_n)$ provenant d'une distribution et nous souhaitons estimer le vecteur de paramètres $\boldsymbol{\theta}$ conditionnellement aux observations $\mathbf{Y}$. La fonction de vraisemblance est

$$L(\boldsymbol{\theta} | Y) = f(Y | \boldsymbol{\theta})$$

où $f(Y; \boldsymbol{\theta})$ est la fonction de densité (masse) de l'échantillon, cette fonction dépend des paramètres $\boldsymbol{\theta}$.

Exemple d'application - Première partie II

1. Remarquons que la fonction densité (masse) conditionnelle des paramètres satisfait

$$f(\boldsymbol{\theta} | Y) = \frac{f(\boldsymbol{\theta}, Y)}{f(Y)} = \frac{f(Y | \boldsymbol{\theta}) f(\boldsymbol{\theta})}{f(Y)} \propto f(Y | \boldsymbol{\theta}) f(\boldsymbol{\theta}).$$

   Dans le cadre bayesien,

   1. $f(\boldsymbol{\theta} | Y)$ est la distribution à posteriori des paramètres;
   2. $f(Y | \boldsymbol{\theta})$ est la fonction de vraisemblance;
   3. $f(\boldsymbol{\theta})$ est la distribution à priori.

2. Maximiser la fonction de vraisemblance uniquement revient à supposer la distribution à priori constante.
3. Notons qu'il sera fort utile de pouvoir simuler une distribution connue à une constante près (proportionnelle à) puisque $f(Y)$ pourrait être difficile à calculer.

Algorithme de Hasting-Metropolis I

Le but de ce qui suit est de construire la chaîne de Markov irréductible et apériodique qui possède la loi stationnaire π désirée.

- Soit b_i, $i = 1, \ldots, m$ sont des nombres positifs, $B = \sum_{i=1}^{n} b_i$ de sorte

  $$\pi_i = \frac{b_i}{B},\ i = 1, 2, \ldots, n$$

  représente une fonction de masse.
- Supposons que B n'est pas disponible (requiert un temps de calcul trop long ou il est très compliqué de le calculer, etc.)

Algorithme de Hasting-Metropolis II

- Partons d'une matrice de transition Q
- Posons

$$\alpha_{ij} = \min\left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1\right) = \min\left(\frac{b_j q_{ji}}{b_i q_{ij}}, 1\right).$$

  - Puisque α_{ij} est un nombre positif inférieur à 1, on peut supposer qu'il représente une probabilité.
  - Notons que nous n'avons pas besoin d'évaluer B.
  - Si $\alpha_{ij} = \frac{\pi_j q_{ji}}{\pi_i q_{ij}}$ alors $\alpha_{ji} = 1$.

Algorithme de Hasting-Metropolis III

- Construisons une nouvelle chaîne de Markov X à partir de la matrice de transition Q et des probabilités α_{ij} :
- Sachant que $X_t = i$
  - Simuler une variable aléatoire intermédiaire X prenant la valeur j avec probabilité q_{ij}.
  - Sachant que $X = j$ alors $X_{t+1} = j$ avec probabilité α_{ij} et $X_{t+1} = i$ avec probabilité $1 - \alpha_{ij}$.

Algorithme de Hasting-Metropolis IV

- Les probabilités de transition de la chaine X sont

  Si $j \neq i$,

$$p_{ij} = P\left[X_{t+1} = j \,|X_t = i\right] = \alpha_{ij} q_{ij}.$$

Si $j = i$,

$$\begin{aligned} p_{ij} &= P\left[X_{t+1} = i \,|X_t = i\right] \\ &= 1 - \sum_{k \neq i} \alpha_{ik} q_{ik} \\ &= \sum_{k} q_{ik} - \sum_{k \neq i} \alpha_{ik} q_{ik} \\ &= q_{ii} + \sum_{k \neq i} \left(1 - \alpha_{ik}\right) q_{ik}. \end{aligned}$$

Algorithme de Hasting-Metropolis V

Theorem

La probabilité stationnaire de X est π.

Preuve. Il suffit de montrer que pour tout $i \neq j$,

$$\pi_i \alpha_{ij} q_{ij} = \pi_j \alpha_{ji} q_{ji}.$$

Si $\alpha_{ij} = \min\left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1\right) < 1$, alors $\alpha_{ji} = 1$ et

$$\pi_i \alpha_{ij} q_{ij} = \pi_i \frac{\pi_j q_{ji}}{\pi_i q_{ij}} q_{ij} = \pi_j q_{ji} = \pi_j \alpha_{ji} q_{ji}.$$

L'autre cas se traite de la même façon. $\square$

Algorithme de Hasting-Metropolis VI

Algorithme Hastings-Metropolis

- **Étape 0**. Choisir $X_0 \in \{1, 2, ..., m\}$.
- **Étape** $t+1$**.**
  1. Générer une v.a. X de fonction de masse $q_{X_t 1}, ..., q_{X_t m}$.
  2. Générer une v.a U de loi uniforme sur $[0, 1]$.
  3. Si $U < \alpha_{X_t, X}$ alors $X_{t+1} = X$
  4. Si $U \geq \alpha_{X_t, X}$ alors $X_{t+1} = X_t$

Gibbs Sampler I

C'est la version de l'algorithme Hastings-Metropolis la plus répandue.

Gibbs Sampler II

Notation

- $\mathbf{Z} = (Z_1, ..., Z_n)$ est un vecteur aléatoire dont la fonction de masse (densité dans le cas continu) conjointe est $p(\mathbf{z})$.
- Nous souhaitons générer un vecteur aléatoire dont la distribution est conditionnelle à l'appartenance à l'ensemble A, c'est-à-dire que la loi conjointe est

$$f(\mathbf{z}) = \frac{p(\mathbf{z})}{P[\mathbf{Z} \in A]}.$$

- Il faut faire l'hypothèse que pour n'importe laquelle des n variables, nous pouvons générer des valeurs à partir de

$$f_i(\mathbf{z}) = P(Z_i = z_i \,|Z_j = z_j, j \neq i).$$

Gibbs Sampler III

L'échantillonneur fonctionne à l'aide d'une chaîne de Markov dont l'espace d'états est $\mathbf{z} = (z_1, \ldots, z_i, \ldots, z_n) \in A$.
Il suffit alors d'appliquer l'algorithme de Hastings-Metropolis en construisant les probabilités de transition de la façon suivante :

1. $X_{t-1} = (z_1, \ldots, z_i, \ldots, z_n)$
2. Générer une coordonnée C choisie uniformément parmi les nombres entiers $1, 2, \ldots, n$.
3. Si $C = i$ alors une variable aléatoire X est générée à partir de la distribution f_i.

Gibbs Sampler IV

4. Si $X = y$, alors le vecteur $\mathbf{y} = (z_1, \ldots, z_{i-1}, y, z_{i+1}, \ldots, z_n)$ est considéré comme prochain état.

   En d'autres mots, l'échantillonneur de Gibbs utilise l'algorithme de Hastings-Metropolis avec les probabilités de transition

$$
\begin{aligned}
q_{\mathbf{zy}} &= \frac{1}{n} f_i(y) \\
&= \frac{1}{n} P(Z_i = y \,|\, Z_j = z_j, j \neq i) \\
&= \frac{1}{n} \frac{p(\mathbf{y})}{P(Z_j = z_j, j \neq i)}.
\end{aligned}
$$

Gibbs Sampler V

5. Parce que l'on veut simuler selon la loi $f(\mathbf{z}) = \frac{p(\mathbf{z})}{P[\mathbf{Z} \in A]}$, le nouveau vecteur $\mathbf{y}$ est accepté avec probabilité

$$\alpha_{\mathbf{zy}} = \min\left(\frac{f(\mathbf{y})\, q_{\mathbf{yz}}}{f(\mathbf{z})\, q_{\mathbf{zy}}}, 1\right).$$

Remarquons que si $\mathbf{z} \in A$ et $\mathbf{y} \in A$, alors $\alpha_{\mathbf{zy}} = 1$ puisque

$$\begin{aligned} f(\mathbf{z})\, q_{\mathbf{zy}} &= \frac{p(\mathbf{z})}{P[\mathbf{Z} \in A]} \frac{1}{n} \frac{p(\mathbf{y})}{P(Z_j = z_j, j \neq i)} = \frac{1}{n} \frac{p(\mathbf{z})\, p(\mathbf{y})}{P(Z_j = z_j, j \neq i)}, \\ f(\mathbf{y})\, q_{\mathbf{yz}} &= \frac{p(\mathbf{y})}{P[\mathbf{Y} \in A]} \frac{1}{n} \frac{p(\mathbf{z})}{P(Z_j = z_j, j \neq i)} = \frac{1}{n} \frac{p(\mathbf{z})\, p(\mathbf{y})}{P(Z_j = z_j, j \neq i)}. \end{aligned}$$

Ainsi $\mathbf{y}$ est conservé s'il appartient à A et il est rejeté sinon.

Gibbs Sampler VI

Algorithme de l'échantillonneur de Gibbs

1. $X_t = (z_1, \ldots, z_i, \ldots, z_n)$
2. $U \sim U(0,1)$, $C = \lfloor nU \rfloor + 1$.
3. Si $C = i$, $X = x$ est généré selon la loi f_i.
4. Si $y = (z_1, \ldots, z_{i-1}, y, z_{i+1}, \ldots, z_n) \in A$, $X_{t+1} = X$. Sinon, $X_{t+1} = X_t$.

Ex. Modèle à volatilité stochastique I

- Le modèle

$$
\begin{aligned}
\text{Rendement} &\quad : \quad r_t = \exp\left(\frac{v_t}{2}\right)\varepsilon_t \\
\text{Log variance} &\quad : \quad v_t = \alpha + \beta v_{t-1} + \sigma u_t
\end{aligned}
$$

  $\{\varepsilon_t, u_t : t = 1, 2, 3, \ldots\}$ est une suite iid de variables gaussiennes centrées et réduites.

- Seuls les rendements sont observés.
- Comment estime-t-on les paramètres du modèle ainsi que la série de log-volatilités?

Nous avons un échantillon $\mathbf{Y} = (Y_1, ..., Y_n)$ provenant d'une distribution donnée et nous souhaitons estimer le vecteur de paramètres $\boldsymbol{\theta}$ conditionnellement aux observations $\mathbf{Y}$.

La **fonction de vraisemblance** est

$$L(\boldsymbol{\theta} \,|\, Y) = f(Y \,|\boldsymbol{\theta})$$

où $f(Y; \boldsymbol{\theta})$ est la fonction de densité (masse) de l'échantillon, cette fonction dépend des paramètres $\boldsymbol{\theta}$.

1. Remarquons que la fonction densité (masse) conditionnelle des paramètres satisfait

$$f(\boldsymbol{\theta} \,|\, Y) = \frac{f(\boldsymbol{\theta}, Y)}{f(Y)} = \frac{f(Y \,|\boldsymbol{\theta})\, f(\boldsymbol{\theta})}{f(Y)} \propto f(Y \,|\boldsymbol{\theta})\, f(\boldsymbol{\theta})\,.$$

   Dans le cadre bayesien,

   1. $f(\boldsymbol{\theta} \,|\, Y)$ est la distribution à posteriori des paramètres;
   2. $f(Y \,|\boldsymbol{\theta})$ est la fonction de vraisemblance;
   3. $f(\boldsymbol{\theta})$ est la distribution à priori des paramètres.

2. Maximiser la fonction de vraisemblance revient à supposer la distribution à priori constante (absence d'information).
3. La distribution à posteriori est parfois difficile à calculer.

Il est possible de choisir la distribution à priori de sorte que $f(\boldsymbol{\theta})$ (à priori) et $f(\boldsymbol{\theta} \mid Y)$ (à posteriori) appartiennent à la même famille de distributions. Ce sont alors des distributions **conjuguées**.

Ex. Approche Bayesienne II
Distributions conjuguées

Theorem

Supposons que l'échantillon $Y_1, \ldots, Y_n$ est formé de variables aléatoires iid $N\left(\mu, \sigma^2\right)$ dont l'espérance est **inconnue** *et la variance est* **connue**.
Supposons aussi que la densité à priori $f\left(\mu\right)$ est la fonction de densité de la distribution normale d'espérance μ_0 et d'écart-type σ_0. Alors $f\left(\mu \,|\, Y\right)$ sera aussi de la famille gaussienne d'espérance μ^* et d'écart-type σ^* où

$$\mu^* = \frac{\sigma^2 \mu_0 + \sigma_0^2 \sum_{i=1}^n y_i}{\sigma^2 + n\sigma_0^2} \text{ et } \sigma^* = \sqrt{\frac{\sigma^2 \sigma_0^2}{\sigma^2 + n\sigma_0^2}}.$$

Preuve.

$$
\begin{aligned}
f(\mu) &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_0} \exp\left(-\frac{1}{2} \left(\frac{\mu - \mu_0}{\sigma_0} \right)^2 \right) \\
f(Y \mid \mu, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right) \\
&= (2\pi)^{-\frac{n}{2}} \frac{1}{\sigma^n} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right)
\end{aligned}
$$

Ex. Approche Bayesienne IV

Distributions conjuguées

Dans ce cas,

$$
\begin{aligned}
& f(\mu \,|\, Y, \sigma) \\
\propto \;& f(Y \,|\mu\, \sigma)\, f(\mu) \\
\propto \;& \frac{1}{\sigma^n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-\mu}{\sigma}\right)^2\right) \frac{1}{\sigma_0} \exp\left(-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right) \\
& \text{(Développer le carré et mettre sur le même dénominateur)} \\
= \;& \frac{1}{\sigma_0}\frac{1}{\sigma^n} \exp\left(-\frac{1}{2}\left(\frac{\mu^2 - 2\mu^*\mu + \frac{\sigma^2\mu_0^2+\sigma_0^2\sum_{i=1}^n y_i^2}{\sigma^2+n\sigma_0^2}}{(\sigma^*)^2}\right)\right) \\
& \text{(Compléter le carré et laisser de coté les termes} \\
& \text{qui ne dépendent que des quantités connues)} \\
\propto \;& \exp\left(-\frac{1}{2}\left(\frac{\mu-\mu^*}{\sigma^*}\right)^2\right). \;\square
\end{aligned}
$$

Ex. Approche Bayesienne V

Distributions conjuguées

Theorem

Supposons que l'échantillon $Y_1, ..., Y_n$ est formé de variables aléatoires iid $N\left(\mu, \sigma^2\right)$ dont l'espérance μ est **connue** *et la précision $\eta = \frac{1}{\sigma^2}$ est* **inconnue***.*
Supposons aussi que la densité à priori $f\left(\eta\right)$ est la fonction de densité de la distribution gamma de paramètres α_0 et β_0. Alors $f\left(\eta \,|\, Y\right)$ sera aussi de la famille gamma de paramètres α^* et β^* où

$$\alpha^* = \alpha_0 + \frac{n}{2} \text{ et } \beta^* = \beta_0 + \frac{1}{2}\sum_{i=1}^{n}\left(y_i - \mu\right)^2 .$$

Idée de la preuve[1].

$$\begin{aligned}
f(\eta) &= \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \eta^{\alpha_0 - 1} \exp(-\beta_0 \eta), \; \mathrm{E}[\eta] = \frac{\alpha_0}{\beta_0}, \; \mathrm{Var}[\eta] = \frac{\alpha_0}{\beta_0^2} \\
f(Y \mid \mu, \eta) &= (2\pi)^{-\frac{n}{2}} \eta^{\frac{n}{2}} \exp\left(-\frac{\eta}{2} \sum_{i=1}^{n} (y_i - \mu)^2 \right) \\
f(\eta \mid \mathbf{Y}, \mu) &= f(Y \mid \mu, \eta) f(\eta) \\
&\propto \eta^{\frac{n}{2}} \exp\left(-\frac{\eta}{2} \sum_{i=1}^{n} (y_i - \mu)^2 \right) \eta^{\alpha_0 - 1} \exp(-\beta_0 \eta) \\
&= \eta^{\alpha_0 + \frac{n}{2} - 1} \exp\left(-\left(\beta_0 + \frac{1}{2} \sum_{i=1}^{n} (y_i - \mu)^2 \right) \eta \right). \; \square
\end{aligned}$$

[1] Parfois notée $f(\eta) = \eta^{\alpha_0 - 1} \exp\left(-\frac{\eta}{\beta_0}\right) / \left(\Gamma(\alpha_0) \beta_0^{\alpha_0}\right)$.

Dans le cas continu, l'algorithme est souvent appliqué de la façon suivante :

1. Générer une valeur initiale $\boldsymbol{\theta}_0$ tel que $f\left(\boldsymbol{\theta}_0 \,|\, Y\right) > 0$ (c'est-à-dire que c'est une valeur possible selon la distribution à posteriori).
2. À l'étape $t+1$
   1. Générer $\boldsymbol{\theta}^*$ à partir d'une distribution symétrique.
   2. $\widetilde{\alpha} = \min\left(\frac{f(\boldsymbol{\theta}^* | Y)}{f(\boldsymbol{\theta}_t | Y)}, 1\right)$ (probabilité de conservation du scénario $\boldsymbol{\theta}^*$)
   3. $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}^*$ avec probabilité $\widetilde{\alpha}$ et $\boldsymbol{\theta}_t$ autrement.

Nous allons maintenant déterminer la fonction de densité à posteriori des paramètres à estimer.

Rappelons que le modèle est

$$\begin{aligned} \text{Rendement} &: \quad r_t = \exp\left(\frac{v_t}{2}\right)\varepsilon_t \\ \text{Log variance} &: \quad v_t = \alpha + \beta v_{t-1} + \sigma u_t . \end{aligned}$$

Exemple - densité à postériori II

Modèle à volatilité stochastique

$$
\begin{aligned}
& f\left(\alpha, \beta, \frac{1}{\sigma}, \mathbf{v} \,|\, \mathbf{r}\right) \\
= & \frac{f\left(\mathbf{r} \,|\alpha, \beta, \frac{1}{\sigma}, \mathbf{v}\right) f\left(\mathbf{v} \,|\alpha, \beta, \frac{1}{\sigma}\right) f\left(\alpha, \beta, \frac{1}{\sigma}\right)}{f\left(\mathbf{r}\right)} \\
= & \prod_{t=1}^{T} \frac{\exp\left(-\frac{v_t}{2}\right)}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{r_t}{\exp\left(\frac{v_t}{2}\right)}\right)^2\right) \\
& \times \prod_{t=1}^{T} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{v_t - \alpha - \beta v_{t-1}}{\sigma}\right)^2\right) \times \frac{f\left(\alpha, \beta, \frac{1}{\sigma}\right)}{f\left(\mathbf{r}\right)} \\
\propto & \frac{1}{\sigma^T} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} \frac{r_t^2}{\exp\left(v_t\right)} - \frac{1}{2}\sum_{t=1}^{T} v_t - \frac{1}{2}\sum_{t=1}^{T}\left(\frac{v_t - \alpha - \beta v_{t-1}}{\sigma}\right)^2\right)
\end{aligned}
$$

si la distribution à priori ne contient aucune information, c'est-à-dire que $f\left(\alpha, \beta, \frac{1}{\sigma}\right) \propto 1$.

Exemple - échantilonneur de Gibbs I

Modèle à volatilité stochastique

Dans ce cas particulier, nous utiliserons l'échantillonneur de Gibbs qui se décompose en deux étapes.

1. Générer un vecteur de log-variance $\mathbf{v}$ en supposant les autres paramètres connus (difficile)
2. Générer les paramètres α, β et σ en supposant le vecteur de log-variance connu (facile)

Rappel:

$$\text{Log variance} : v_t = \alpha + \beta v_{t-1} + \sigma u_t .$$

La partie facile est aussi exécutée en deux étapes.

1. Conditionnellement à $\sigma^2 = \frac{1}{\eta}$ et à la distribution gaussienne des termes de bruit, les paramètres α et β sont générés selon une distribution normale :

$$\alpha, \beta \,|\sigma, \mathbf{v}, \mathbf{r} \sim N\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{v}_{1:T}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

où

$$\mathbf{X} = \begin{pmatrix} 1 & v_0 \\ \vdots & \vdots \\ 1 & v_{T-1} \end{pmatrix} \text{ et } \mathbf{v}_{1:T} = \begin{pmatrix} v_1 \\ \vdots \\ v_T \end{pmatrix}.$$

Les habitués reconnaîtront la distribution conjointe des estimateurs des moindres carrés de l'ordonnée à l'origine et de la pente de la droite de régression linéaire simple.

2. Conditionnellement aux paramètres α et β ainsi qu'à la log-variance $\mathbf{v}$, le paramètre de précision est généré à l'aide de la distribution gamma :

$$\frac{1}{\sigma^2} \bigg| \alpha, \beta, \mathbf{v}, \mathbf{r} \sim \Gamma\left(\frac{T}{2}, \frac{1}{2} \sum_{i=1}^{T} \left(v_t - \alpha - \beta v_{t-1} \right)^2 \right)$$

   1. Rappelons que selon les densités conjuguées, les paramètres de la fonction de densité à posteriori sont

$$\alpha^* = \alpha_0 + \frac{T}{2} \text{ et } \beta^* = \beta_0 + \frac{1}{2} \sum_{i=1}^{T} \left(v_t - \alpha - \beta v_{t-1} \right)^2 .$$

   On remarque que la fonction de densité à priori ne contient aucune information puisque $\alpha_0 = 0$ et $\beta_0 = 0$.

- Il faut maintenant échantillonner les log-variances.
- Elle seront générées une à la fois, le critère d'acceptation étant basé sur la distribution à posteriori.
- Notation: $\mathbf{v}_{\backslash t} = \left(v_0, \ldots, v_{t-1}, v_{t+1}, \ldots, v_T \right)$.

Theorem

$$f\left(v_t \left| \mathbf{v}_{\backslash t}, \alpha, \beta, \sigma, \mathbf{r} \right. \right) \propto \exp\left(-\frac{v_t}{2} \right) \exp\left(-\frac{1}{2} \frac{r_t^2}{\exp\left(v_t \right)} \right) \frac{1}{\sigma^2}$$

$$\exp\left(-\frac{1}{2} \frac{\left(v_t - \alpha - \beta v_{t-1} \right)^2}{\sigma^2} \right) \exp\left(-\frac{1}{2} \frac{\left(v_{t+1} - \alpha - \beta v_t \right)^2}{\sigma^2} \right).$$

Idée de la preuve.

$$
\begin{aligned}
& f\left(v_t \middle| \mathbf{v}_{\backslash t}, \alpha, \beta, \sigma, \mathbf{r}\right) \\
= \; & \frac{f(\mathbf{v}, \alpha, \beta, \sigma, \mathbf{r})}{f\left(\mathbf{v}_{\backslash t}, \alpha, \beta, \sigma, \mathbf{r}\right)} \\
= \; & \frac{f(\mathbf{r} \,|\mathbf{v}, \alpha, \beta, \sigma)\, f(\mathbf{v} \,|\alpha, \beta, \sigma)\, f(\alpha, \beta, \sigma)}{f\left(\mathbf{r} \middle|\mathbf{v}_{\backslash t}, \alpha, \beta, \sigma\right) f\left(\mathbf{v}_{\backslash t} \,|\alpha, \beta, \sigma\right) f(\alpha, \beta, \sigma)} \\
= \; & \frac{f(r_t \,|v_t, \alpha, \beta, \sigma)\, f(v_t \,|v_{t-1}, \alpha, \beta, \sigma)\, f(v_{t+1} \,|v_t, \alpha, \beta, \sigma)}{\underbrace{f(r_t \,|v_{t-1}, \alpha, \beta, \sigma)\, f(v_{t+1} \,|v_{t-1}, \alpha, \beta, \sigma)}_{\text{Ne dépend pas de } v_t}} && (1) \\
& \text{(voir page suivante)} \\
\propto \; & f(r_t \,|v_t, \alpha, \beta, \sigma)\, f(v_t \,|v_{t-1}, \alpha, \beta, \sigma)\, f(v_{t+1} \,|v_t, \alpha, \beta, \sigma) && (2)
\end{aligned}
$$

puisque qu'au numérateur, nous trouvons

$$\begin{aligned} f\left(\mathbf{v}\,|\alpha,\beta,\sigma\right) &= \prod_{u=1}^{T} f\left(v_u\,|v_{u-1},\alpha,\beta,\sigma\right) \\ f\left(\mathbf{r}\,|\mathbf{v},\alpha,\beta,\sigma\right) &= \prod_{u=1}^{T} f\left(r_u\,|v_u,\alpha,\beta,\sigma\right) \end{aligned}$$

et au dénominateur, il y a

$$\begin{aligned} f\left(\mathbf{v}_{\setminus t}\,|\alpha,\beta,\sigma\right) &= \prod_{u=t+2}^{T} f\left(v_u\,|v_{u-1},\alpha,\beta,\sigma\right)\prod_{u=1}^{t-1} f\left(v_u\,|v_{u-1},\alpha,\beta,\sigma\right) \\ &\quad f\left(v_{t+1}\,|v_{t-1},\alpha,\beta,\sigma\right) \\ f\left(\mathbf{r}\,\middle|\mathbf{v}_{\setminus t},\alpha,\beta,\sigma\right) &= \prod_{u=t+1}^{T} f\left(r_u\,|v_{u-1},\alpha,\beta,\sigma\right)\prod_{u=1}^{t-1} f\left(r_u\,|v_u,\alpha,\beta,\sigma\right) \\ &\quad f\left(r_t\,|v_{t-1},\alpha,\beta,\sigma\right). \end{aligned}$$

Puisque

$$\begin{aligned} \text{Rendement} : r_t &= \exp\left(\frac{v_t}{2}\right)\varepsilon_t \\ \text{Log variance} : v_{t+1} &= \alpha + \beta v_t + \sigma u_{t+1}, \end{aligned}$$

alors

$$\begin{aligned} f\left(r_t \,|\, v_t, \alpha, \beta, \sigma\right) &= \frac{\exp\left(-\frac{v_t}{2}\right)}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{r_t^2}{\exp\left(v_t\right)}\right) \\ f\left(v_t \,|\, v_{t-1}, \alpha, \beta, \sigma\right) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{v_t - \alpha - \beta v_{t-1}}{\sigma}\right)^2\right). \end{aligned}$$

Remplacer le tout dans l'équation (2) complète la démonstration. $\square$

Rappelons qu'à l'étape $t+1$

1. Générer $\boldsymbol{\theta}^*$ à partir d'une distribution symétrique.
2. $\widetilde{\alpha} = \min\left(\frac{f(\boldsymbol{\theta}^*|Y)}{f(\boldsymbol{\theta}_t|Y)}, 1\right)$
(probabilité de conservation du scénario $\boldsymbol{\theta}^*$)
3. $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}^*$ avec probabilité $\widetilde{\alpha}$ et $\boldsymbol{\theta}_t$ autrement.

- La distribution à posteriori permet de calculer le ratio $\widetilde{\alpha}$.
- Afin de générer des scénarios pour v_t, la distribution normale d'espérance $\frac{v_{t-1}+v_{t+1}}{2}$ et d'écart-type σ est utilisée.
  - Puisque $v_1 = \alpha + \beta v_0 + \sigma u_1$, v_0 est généré à partir d'une $N\left(\frac{v_1-\alpha}{\beta}, \sigma^2\right)$.
  - Puisque $v_T = \alpha + \beta v_{T-1} + \sigma u_T$, v_T est généré à partir d'une $N\left(\alpha + \beta v_{T-1}, \sigma^2\right)$.

- **Efficacité numérique**
- L'acceptance est vérifier en log pour éviter de calculer les exponentielles.
- Les v_t sont simulés vectoriellement en deux étapes (les t pairs et les t impairs) plutôt qu'avec une boucle.