

ECN 6578A, Économétrie des marchés financiers, Hiver 2020

Cours 6

William McCausland

2021-03-07

Plan

1. Un modèle de volatilité stochastique
2. Inférence bayésienne : un peu de théorie
3. Inférence bayésienne : un peu de computation

Un modèle de volatilité stochastique

- Un modèle de volatilité stochastique simple

$$r_t = \mu + \sqrt{h_t} \epsilon_t,$$

$$\log h_t = \alpha_0 + \alpha_1 \log h_{t-1} + v_t.$$

$$\begin{bmatrix} \epsilon_t \\ v_t \end{bmatrix} \sim N \left(0, \begin{bmatrix} 1 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} \right).$$

- La volatilité n'est pas une fonction déterministe de rendements passés, comme dans les modèles (G)ARCH.
- Évaluer la vraisemblance est difficile :

$$f(r|\mu, \alpha_0, \alpha_1, \sigma_v^2) = \int f(h|\mu, \alpha_0, \alpha_1, \sigma_v^2) f(r|h, \mu, \alpha_0, \alpha_1, \sigma_v^2) dh,$$

où $r \equiv (r_1, \dots, r_T)$, $h \equiv (h_1, \dots, h_T)$.

- On peut introduire une corrélation négative entre ϵ_t et v_t pour capturer l'effet de levier.

Motivation des méthodes bayésiennes

- ▶ Analyse simple et élégante des modèles avec variables latentes : on peut le faire avec seulement des évaluations de $f(h, r|\mu, \alpha_0, \alpha_1, \sigma_v^2)$. On n'a pas besoin d'évaluer la vraisemblance $f(r|\mu, \alpha_0, \alpha_1, \sigma_v^2)$.
- ▶ En faisant des prévisions, on tient compte de l'incertitude sur
 - ▶ les paramètres,
 - ▶ les variables latentes,
 - ▶ les ordres (p et q par exemple) et
 - ▶ les modèles.
- ▶ Il n'y a pas de recours aux résultats asymptotiques en T .

Éléments de l'analyse bayésienne (modèle sans variables latentes)

- ▶ Quantités pertinentes :
 - ▶ θ , un vecteur de paramètres inconnus,
 - ▶ $y = (y_1, \dots, y_T)$, un vecteur aléatoire des variables observables, et
 - ▶ y° , le vecteur observé.
- ▶ Densités pertinentes :
 - ▶ $f(y|\theta)$, la densité conditionnelle des données (modèle),
 - ▶ $\mathcal{L}(\theta; y^\circ) = f(y^\circ|\theta)$, la vraisemblance,
 - ▶ $f(\theta)$, la densité *a priori*,
 - ▶ $f(\theta, y)$, la densité conjointe,
 - ▶ $f(\theta|y)$, la densité *a posteriori*,
 - ▶ $f(y)$, la densité marginale des données,
 - ▶ $f(y^\circ)$, la vraisemblance marginale (un nombre).

Éléments de l'analyse bayésienne (modèle avec variables latentes)

- ▶ Quantités pertinentes :
 - ▶ θ , un vecteur de paramètres inconnus,
 - ▶ $h = (h_1, \dots, h_T)$, un vecteur aléatoire des variables d'état,
 - ▶ $y = (y_1, \dots, y_T)$, un vecteur aléatoire des variables observables, et
 - ▶ y° , le vecteur observé.
- ▶ Densités pertinentes :
 - ▶ $f(h|\theta)$, la densité des variables d'état
 - ▶ $f(y|\theta, h)$, la densité conditionnelle des données (modèle),
 - ▶ $\mathcal{L}(\theta; y^\circ) = f(y^\circ|\theta)$, la vraisemblance,
 - ▶ $f(\theta)$, la densité *a priori*,
 - ▶ $f(\theta, h, y)$, la densité conjointe,
 - ▶ $f(\theta, h|y)$, la densité *a posteriori*,
 - ▶ $f(\theta|h, y)$ et $f(h|\theta, y)$ des densités *a posteriori* conditionnelles,
 - ▶ $f(y)$, la densité marginale des données,
 - ▶ $f(y^\circ)$, la vraisemblance marginale (un nombre).

Inférence bayésienne

- ▶ Par la règle de Bayes,

$$f(\theta|y^\circ) = \frac{f(\theta, y^\circ)}{f(y^\circ)} = \frac{f(\theta)f(y^\circ|\theta)}{f(y^\circ)} \propto f(\theta)f(y^\circ|\theta).$$

- ▶ $f(\theta)$ représente notre incertitude sur θ avant l'observation de y .
- ▶ $f(\theta|y^\circ)$ représente notre incertitude sur θ après qu'on observe $y = y^\circ$.
- ▶ Un point important à retenir : $f(\theta|y^\circ) \propto f(\theta, y^\circ)$.

Reprise et extension de l'exemple Bernoulli

- ▶ Si y_i est Bernoulli avec probabilité θ , $f(y|\theta) = \theta^{n_1}(1 - \theta)^{n_0}$.
- ▶ Mettons qu'on choisit la loi *a priori* $\theta \sim \text{Beta}(\alpha, \beta)$ sur $[0, 1]$:

$$f(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}.$$

- ▶ La densité conjointe est

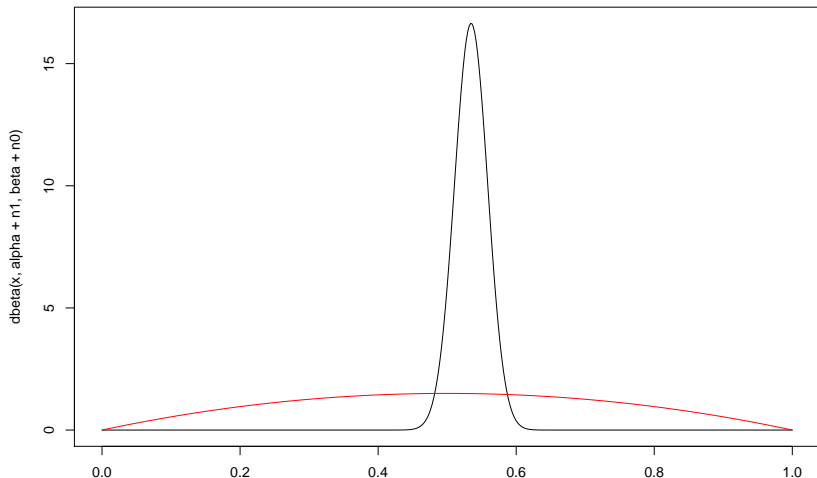
$$f(\theta, y) = f(\theta)f(y|\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha+n_1-1} (1 - \theta)^{\beta+n_0-1}.$$

- ▶ La loi *a posteriori* doit être $\theta \sim \text{Beta}(\alpha + n_1, \beta + n_0)$.
- ▶ La vraisemblance marginale est $f(\theta, y)/f(\theta|y)$:

$$f(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + n_1)\Gamma(\beta + n_0)}{\Gamma(\alpha + \beta + n)}.$$

Graphique pour l'exemple Bernoulli

```
n0 = 200; n1 = 230; alpha=2; beta=2  
x = seq(0, 1, by=0.002)  
plot(x, dbeta(x, alpha+n1, beta+n0), type='l')  
lines(x, dbeta(x, alpha, beta), col='red')
```



Exemple gaussien I

- Considérez le modèle $y_t \sim \text{iid } N(\mu, h^{-1})$.
- Le vecteur de paramètres est $\theta = (\mu, h)$.
- Le vecteur d'observables est $y = (y_1, \dots, y_T)$.
- La densité des données est

$$\begin{aligned} f(y|\theta) &= \prod_{t=1}^T \sqrt{\frac{h}{2\pi}} \exp \left[-\frac{h}{2} (y_t - \mu)^2 \right] \\ &= \left(\frac{h}{2\pi} \right)^{T/2} \exp \left[-\frac{h}{2} \sum_{t=1}^T (y_t - \mu)^2 \right]. \end{aligned}$$

Exemple gaussien II

- ▶ Mettons qu'on choisit une loi *a priori* où h et μ sont indépendents, avec

$$\mu \sim N(\bar{\mu}, \bar{\omega}^{-1}), \quad \bar{s}h \sim \chi^2(\bar{\nu}),$$

où $\bar{\mu}$, $\bar{\omega}$, \bar{s} et $\bar{\nu}$ sont des hyperparamètres constants choisis par l'investigateur.

- ▶ La densité *a priori* est

$$f(\theta) \propto \exp \left[-\frac{\bar{\omega}}{2} (\mu - \bar{\mu})^2 \right] \cdot h^{(\bar{\nu}-2)/2} \exp \left[-\frac{1}{2} \bar{s}^2 h \right].$$

- ▶ La densité conjointe est

$$f(\theta, y) \propto h^{(\bar{\nu}+T-2)/2} \exp \left[-\frac{\bar{\omega}}{2} (\mu - \bar{\mu})^2 - \frac{h}{2} \left(\bar{s}^2 + \sum_{t=1}^T (y_t - \mu)^2 \right) \right].$$

L'intégration et les objectifs de l'analyse bayésienne

- ▶ Plusieurs problèmes d'inférence bayésienne ont, comme solution, une intégrale par rapport à la densité *a posteriori*.
- ▶ Exemple 1, estimation ponctuelle de θ_k sous perte quadratique:

$$\hat{\theta}_k = E[\theta_k | y^\circ] = \int \theta_k f(\theta | y^\circ) d\theta.$$

- ▶ Exemple 2, quantification de l'incertitude sur θ_k :

$$\text{Var}[\theta | y^\circ] = E[(\theta_k - E[\theta_k | y^\circ])^2 | y^\circ].$$

- ▶ Exemple 3, densité prédictive (valeurs de y_{T+1} sur une grille) :

$$f(y_{T+1} | y^\circ) = E[f(y_{T+1} | \theta, y^\circ) | y^\circ].$$

Preuve de l'exemple 3

$$\begin{aligned} E[f(y_{T+1}|y_1, \dots, y_T, \theta)|y_1, \dots, y_T] \\ &= \int f(y_{T+1}|y_1, \dots, y_T, \theta) f(\theta|y_1, \dots, y_T) d\theta \\ &= \int f(y_{T+1}, \theta|y_1, \dots, y_T) d\theta \\ &= f(y_{T+1}|y_1, \dots, y_T) \end{aligned}$$

Méthodes pour trouver $E[g(\theta)|y^\circ]$

- ▶ Calcul analytique : élégant, exacte, presque toujours insoluble.
- ▶ Simulation Monte Carlo indépendante :
 - ▶ Si on peut simuler $\theta^m \sim \text{iid } \theta|y^\circ$,

$$\frac{1}{M} \sum_{m=1}^M g(\theta^m) \rightarrow_p E[g(\theta)|y^\circ].$$

- ▶ Cependant, cette simulation est rarement faisable.
- ▶ Simulation Monte Carlo chaîne de markov (MCMC) :
 - ▶ On choisit un processus markovien avec densité de transition $f(\theta^m|\theta^{m-1})$ telle que la loi *a posteriori* $\theta|y^\circ$ est la loi stationnaire du processus. C'est à dire :

$$\theta^{m-1} \sim f(\theta|y^\circ) \Rightarrow \theta^m \sim f(\theta|y^\circ).$$

- ▶ Sous quelques conditions techniques, la loi de θ^m converge à la loi *a posteriori* et

$$\frac{1}{M} \sum_{m=1}^M g(\theta^m) \rightarrow_p E[g(\theta)|y^\circ].$$

Exemple, densité de prévision

- ▶ L'objectif est la densité $f(y_{T+1}|y_1, \dots, y_T)$ sur une grille.
- ▶ Fixez une valeur y_{T+1} arbitraire sur une grille.
- ▶ On a vu (dans un modèle sans variables latentes)

$$f(y_{T+1}|y_1, \dots, y_T) = E[f(y_{T+1}|y_1, \dots, y_T, \theta)|y_1, \dots, y_T].$$

- ▶ Ici, $g(\theta) = f(y_{T+1}|y_1, \dots, y_T, \theta)$. Notez que les données observées y_1, \dots, y_T et le point de grille y_{T+1} sont fixes.
- ▶ Avec l'échantillon θ^m , $m = 1, \dots, M$, on calcule la quantité à gauche, qui converge à la quantité voulue à droite :

$$\frac{1}{M} \sum_{m=1}^M f(y_{T+1}|y_1, \dots, y_T, \theta^m) \rightarrow_p f(y_{T+1}|y_1, \dots, y_T).$$

- ▶ Répétez pour chaque point sur la grille.
- ▶ À noter : asymptotique en M et non en T .

MCMC 1 : marche aléatoire metropolis-hastings

- ▶ Pour tirer $\theta^m | \theta^{m-1}$:
 1. Tirer $\theta^* \sim N(\theta^{m-1}, \Sigma)$
 2. Calculer le ratio de Hastings :

$$H = \frac{f(\theta^* | y^\circ)}{f(\theta^{m-1} | y^\circ)}.$$

3. Accepter θ^* avec probabilité $\min(1, H)$.
- ▶ Accepter θ^* veut dire $\theta^m = \theta^*$; si on n'accepte pas, $\theta^m = \theta^{m-1}$.
 - ▶ On peut utiliser $f(\theta, y^\circ)$ au lieu de $f(\theta | y^\circ)$ parce que les constantes $f(y^\circ)$ s'annulent.
 - ▶ La convergence tient pour n'importe quelle Σ , mais il y a des choix qui sont meilleurs que d'autres.

Initialisation

```
# Vraies valeurs des paramètres  
vrai.mu = 6  
vrai.h = 0.04  
vrai.sigma = 1/sqrt(vrai.h)  
  
# Données simulées, statistiques suffisantes  
n = 10; set.seed(12345)  
y = rnorm(n, vrai.mu, vrai.sigma)  
y.bar = mean(y)  
y2.bar = mean(y^2)  
  
# Hyper-paramètres  
mu.bar = 10  
omega.bar = 0.01  
nu.bar = 4  
s2.bar = 0.01
```

Fonctions pour calculer des densités

```
# Densité a priori de mu
```

```
lnp.mu = function(mu) {  
  lnp = dnorm(mu,mu.bar,1/sqrt(omega.bar),log=TRUE)  
}
```

```
# Densité a priori de h
```

```
lnp.h = function(h) {  
  lnp = log(s2.bar) + dchisq(h*s2.bar,nu.bar,log=TRUE)  
}
```

```
# Densité des données
```

```
lnp.y..mu.h = function(mu, h) {  
  lnp = (n/2)*log(h) - (n/2)*log(2*pi)  
  lnp = lnp - 0.5*h*n * (y2.bar - 2*y.bar*mu + mu^2)  
}
```

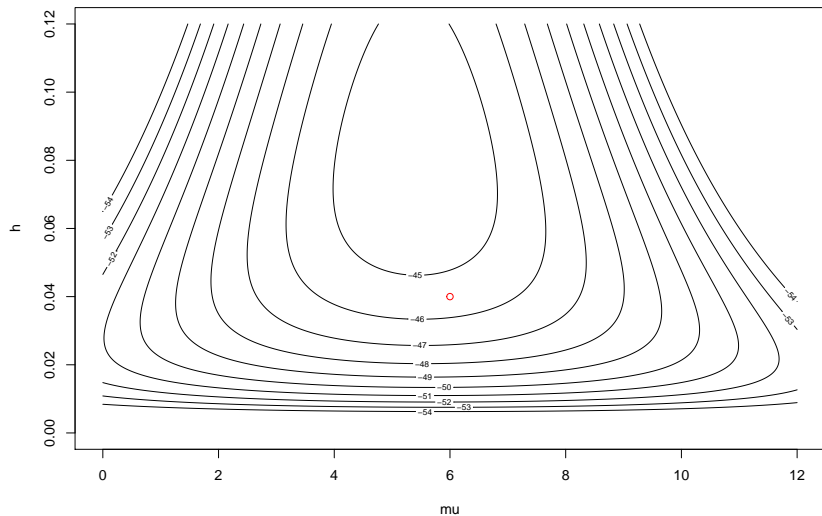
La densité conjointe

```
# Densité a posteriori de mu and h, pas normalisée  
lnp.mu.h..y = function(mu,h) {  
  lnp = lnp.mu(mu) + lnp.h(h) + lnp.y..mu.h(mu,h)  
}
```

```
# Fonction pour faire un graphique de la densité a posteriori  
do.plot = function() {  
  mu = seq(0, 12, by=0.01)  
  h = seq(0, 0.12, by=0.0001)  
  p = outer(mu, h, FUN=lnp.mu.h..y)  
  levels = seq(ceiling(max(p)), ceiling(max(p))-10, by=-1)  
  contour(mu, h, p, xlab='mu', ylab='h', levels=levels)  
  points(vrai.mu, vrai.h, col='red')  
}
```

Graphique de la densité conjointe

```
do.plot()
```

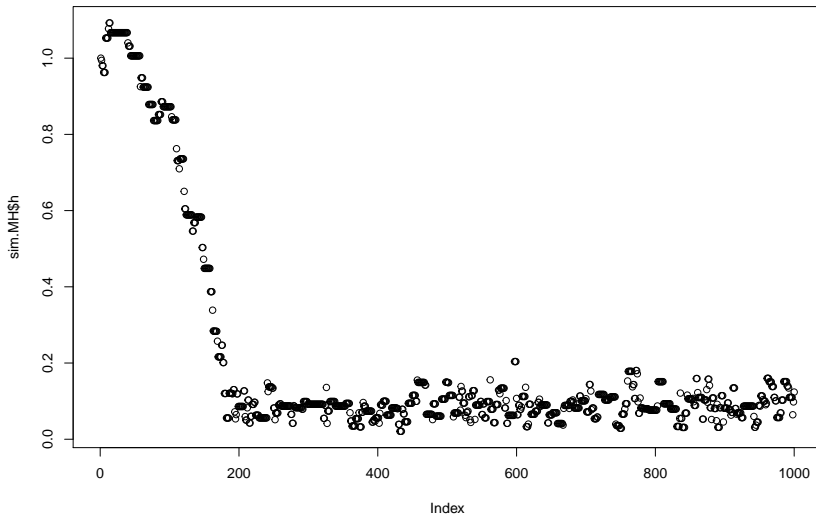


Code pour l'algorithme Metropolis Hastings

```
Metro.sim = function(M) {  
  mu = vector('numeric', M); h = vector('numeric', M)  
  mu[1] = 0; h[1] = 1  
  lnp = lnp.mu.h..y(mu[1], h[1])  
  for( m in seq(2, M) ) {  
    h.et=rnorm(1,h[m-1],0.05); mu.et=rnorm(1,mu[m-1],2.0)  
    if( h.et > 0.0 ) {  
      lnp.et = lnp.mu.h..y(mu.et, h.et)  
    } else lnp.et = -Inf  
    H = exp(lnp.et - lnp)  
    if( runif(1) < H ) {  
      h[m] = h.et; mu[m] = mu.et; lnp = lnp.et  
    } else {  
      h[m] = h[m-1]; mu[m] = mu[m-1]  
    }  
  }  
  list(mu=mu, h=h)  
}
```

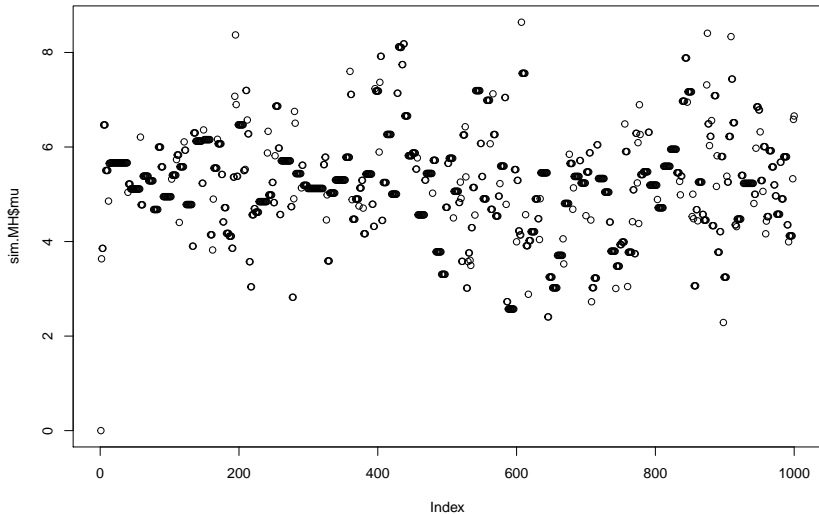
Trace de h

```
sim.MH = Metro.sim(1000)  
plot(sim.MH$h)
```



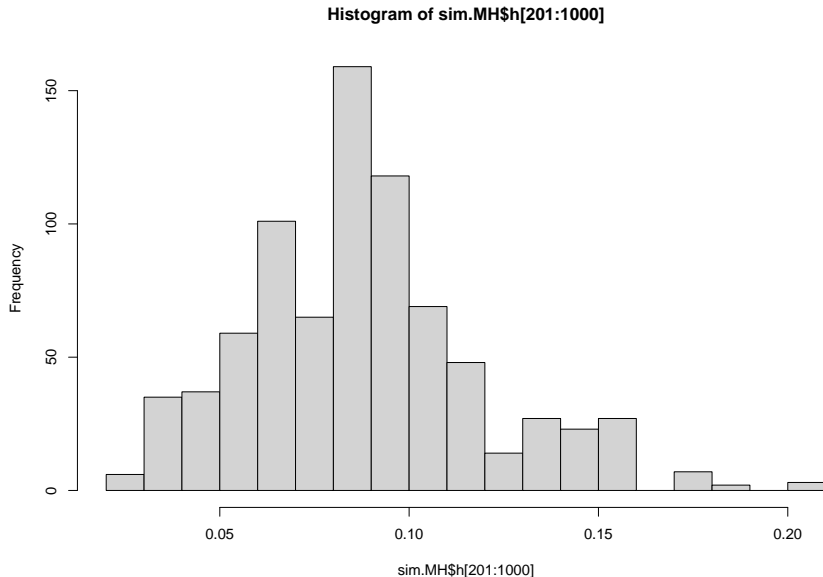
Trace de μ

```
plot(sim.MH$mu)
```



Histogramme de h

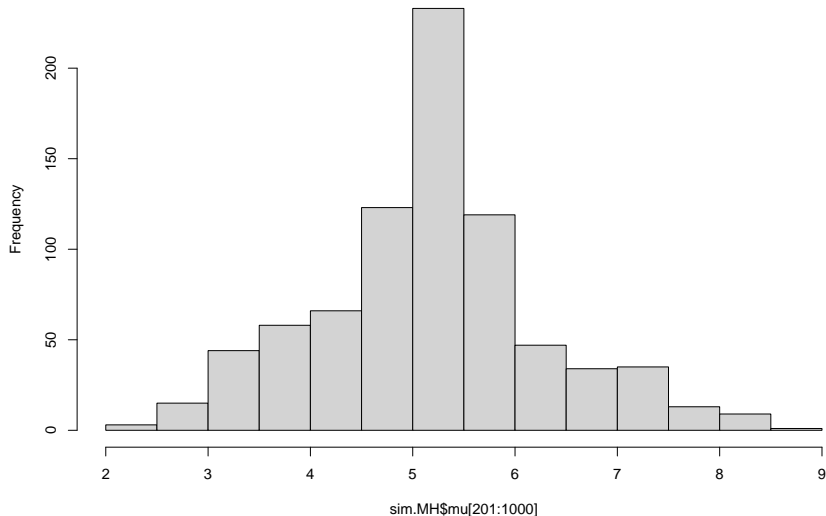
```
hist(sim.MH$h[201:1000], 20)
```



Histogramme de μ

```
hist(sim.MH$mu[201:1000], 20)
```

Histogram of sim.MH\$mu[201:1000]



MCMC 2 : échantillonnage de gibbs pour le modèle gaussien

- ▶ Considérez la densité de transition $f(\theta^m|\theta^{m-1})$ définie par
 1. $\mu^m \sim \mu|y = y^\circ, h = h^{m-1}$.
 2. $h^m \sim h|y = y^\circ, \mu = \mu^m$.
- ▶ Une preuve que $\theta|y^\circ$ est la loi stationnaire de cette loi de transition :
 - ▶ Mettons que la loi de $\theta^{m-1} = (\mu^{m-1}, h^{m-1})$ est la loi *a posteriori* $\theta|y = y^\circ$.
 - ▶ Alors la loi marginale de h^{m-1} est la loi $h|y = y^\circ$.
 - ▶ Après l'étape 1, la loi de (μ^m, h^{m-1}) est la loi *a posteriori*.
 - ▶ Alors la loi marginale de μ^m est la loi $\mu|y = y^\circ$.
 - ▶ Après l'étape 2, la loi de $\theta^m = (\mu^m, h^m)$ est la loi *a posteriori*.
- ▶ L'idée se généralise (diviser un problème en parties solubles)

Dérivation, loi *a posteriori* conditionnelle de $\mu|y = y^\circ, h$

- On peut écrire (c_1 et c_2 constants)

$$\begin{aligned} & -\frac{\bar{\omega}}{2}(\mu - \bar{\mu})^2 - \frac{h}{2} \sum_{t=1}^T (y_t - \mu)^2 \\ &= -\frac{\bar{\omega}}{2}(\mu - \bar{\mu})^2 - \frac{h}{2} \left[\sum_{t=1}^T (y_t - \bar{y})^2 + T(\mu - \bar{y})^2 \right] \\ &= c_1 - \frac{\bar{\omega}}{2}(\mu - \bar{\mu})^2 - \frac{hT}{2}(\mu - \bar{y})^2 \\ &= c_2 - \frac{\bar{\omega} + hT}{2} [\mu - (\bar{\omega}\bar{\mu} + hT\bar{y})/(\bar{\omega} + hT)]^2. \end{aligned}$$

- Dernière étape : **complétion du carré**
- Alors $f(\mu|y, h) \propto \exp \left[-\frac{\bar{\bar{\omega}}}{2}(\mu - \bar{\bar{\mu}})^2 \right]$, où
 - $\bar{\bar{\mu}} = \frac{\bar{\omega}\bar{\mu} + hT\bar{y}}{\bar{\omega} + hT},$
 - $\bar{\bar{\omega}} = \bar{\omega} + hT.$

MCMC 3 : échantillonnage de gibbs pour le modèle SV

Faire M fois :

1. Tirer α_0 et α_1 de la loi *a posteriori* conditionnelle $(\alpha_0, \alpha_1)|y, \sigma_v, h_1, \dots, h_T$.
2. Tirer $\omega_v = \sigma_v^{-2}$ de la loi *a posterior* conditionnelle $\omega_v|y, \alpha_0, \alpha_1, h_1, \dots, h_T$
3. Tirer h_1, \dots, h_T de la loi *a posteriori* conditionnelle $h_1, \dots, h_T|y, \alpha_0, \alpha_1, \omega_v$.

Notes :

- ▶ La loi $\alpha_0, \alpha_1|y, \omega_v, h_1, \dots, h_T$ est presque gaussienne. On peut tirer une proposition (α_0^*, α_1^*) de l'approximation gaussienne. Une étape "accepter ou rejeter" "corrige" l'approximation.
- ▶ Si la loi *a priori* ω_v est khi-carré avec un paramètre d'échelle, la loi conditionnelle *a posteriori* $\omega_v|y, \alpha_0, \alpha_1, h_1, \dots, h_T$ l'est aussi. On peut tirer de cette loi directement.
- ▶ Il y a plusieurs façons de tirer h de façon efficace. Les détails sont trop avancés pour le cours.