

ECN 6338 Cours 9

La génération de variables aléatoires multivariées

William McCausland

2025-11-06

Survol du cours 9

1. L'idée de l'exercice Monte Carlo, Monte Carlo avec chaînes de Markov (MCMC)
2. MCMC avec le marche aléatoire de Metropolis
3. Un modèle à deux paramètres
4. Marche aléatoire pour le modèle à deux paramètres
5. Échantillonnage de Gibbs
6. Exemple probit, participation des femmes dans la population active
7. Augmentation des données (data augmentation)

L'exercice Monte Carlo

Situations où l'intégration par Monte Carlo est convenable :

- ▶ problèmes de haute dimension (> 20)
- ▶ problèmes où on veut calculer plusieurs intégrales de la forme

$$E_p[g_i(\theta)] = \int_D p(\theta)g_i(\theta) d\theta, \quad i = 1, \dots, J,$$

où $p(\theta)$ est une densité commune, la *densité cible*. Souvent, p est la densité *a posteriori* $p(\theta|y)$ d'une analyse bayésienne, θ est un vecteur de paramètres d'un modèle et y est un échantillon.

L'exercice Monte Carlo est de tirer $\theta^{(m)}$, $m = 1, \dots, M$, tel que pour chaque $g(\theta)$ où l'espérance à droite est finie,

$$\frac{1}{M} \sum_{m=1}^M g(\theta^{(m)}) \xrightarrow{P} \int p(\theta)g(\theta) d\theta,$$

et un théorème central limite s'applique.

Monte Carlo avec chaînes markoviennes (MCMC)

- ▶ Le simple fait qu'on peut écrire la densité cible $p(\theta|y)$ ne veut pas dire qu'il est facile de tirer un échantillon iid de $\theta|y$.
- ▶ En fait, les tirages iid sont infaisables pour tous les modèles sauf les plus simples.
- ▶ En pratique, on se contente de simuler une chaîne markovienne dont la loi cible est la loi stationnaire de la chaîne :
 - ▶ Une chaîne markovienne sur Θ a une loi de transition $\theta^{(m+1)}|\theta^{(m)}$.
 - ▶ La loi cible avec densité $p(\theta|y)$ est une loi stationnaire de la chaîne si

$$\theta^{(m)} \sim p(\theta|y) \Rightarrow \theta^{(m+1)} \sim p(\theta|y).$$

- ▶ Quand la loi cible est une loi stationnaire, on peut dire que la transition markovienne préserve la loi cible.
- ▶ Si la chaîne est ergodique, la loi stationnaire est unique et il y a une loi de grands nombres et un théorème central limite qui s'appliquent.

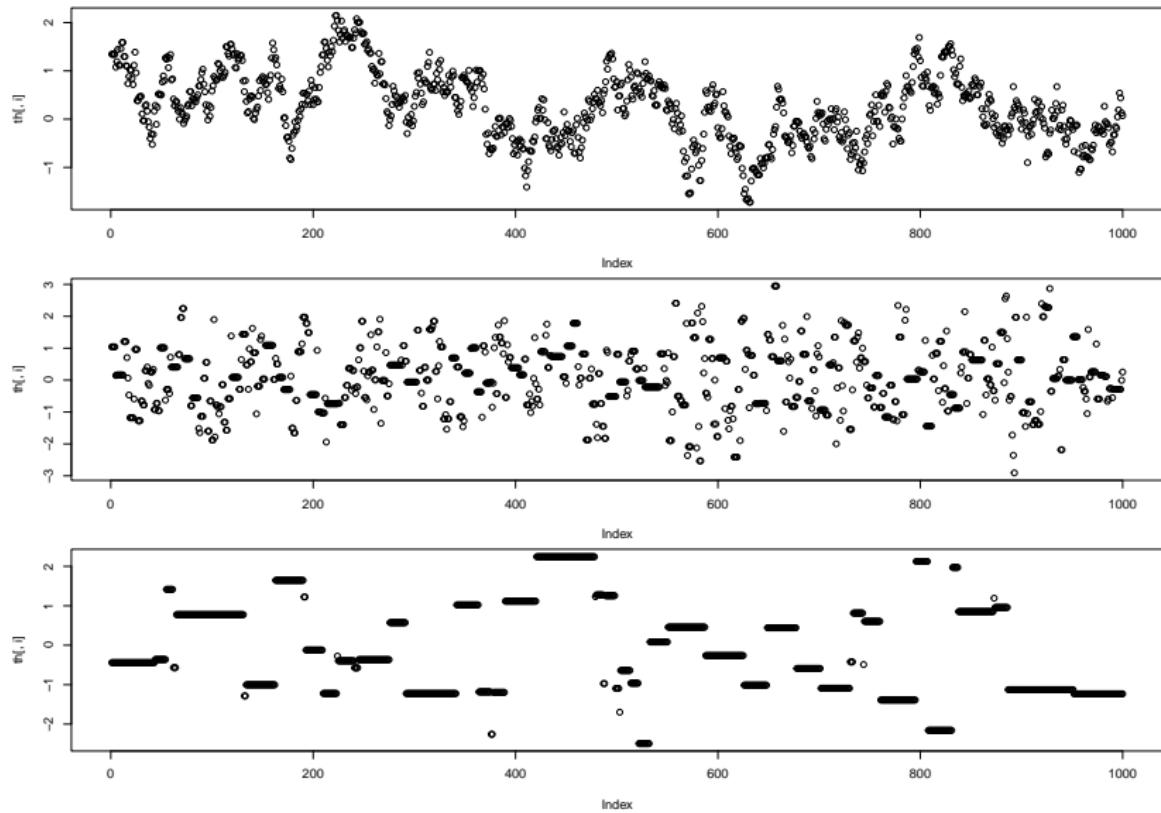
Mise à jour du type marche aléatoire de Metropolis

- ▶ Supposons que la loi cible $\theta|y$ ait un noyau de densité $k(\theta) \propto p(\theta|y)$. Par exemple, $k(\theta) \equiv p(\theta)p(y|\theta) = p(\theta|y)p(y)$
- ▶ La loi cible $\theta|y$ est l'unique loi invariante de la transition de Markov suivante, de $\theta^{(m)}$ à $\theta^{(m+1)}$, où Σ est n'importe quelle matrice définie positive :
 1. Tirer $\theta^* \sim N(\theta^{(m)}, \Sigma)$
 2. Tirer U de la loi uniform sur $[0, 1]$.
 3. Si $U \leq \frac{k(\theta^*)}{k(\theta^{(m)})}$ on fixe $\theta^{(m+1)} = \theta^*$, sinon on fixe $\theta^{(m+1)} = \theta^{(m)}$.
- ▶ La loi normale de l'étape 1 peut être remplacée par n'importe quelle loi symétrique autour de zéro.
- ▶ Elle peut être remplacée par une loi asymétrique, avec une modification appropriée du seuil à l'étape 3. (Mise à jour Metropolis-Hastings)
- ▶ Remarquez que le *ratio de Hastings* $k(\theta^*)/k(\theta^{(m)})$ égale à $p(\theta^*|y)/p(\theta^{(m)}|y)$ pour la densité cible normalisée $p(\theta|y)$.

Metropolis, $N(0, 1)$ target, $\sqrt{\Sigma} = 0.24, 2.4, 24$

```
set.seed(1234567890)
M = 1000
sigma = c(0.1, 1.0, 10.0) * 2.4 # 2.4 optimal dans le cas de l'application
th = array(0, dim=c(M, 3))
for (i in 1:3) {
  th[1, i] = rnorm(1); pth = dnorm(th[1, i])
  for (m in 2:M) {
    thst = th[m-1, i] + rnorm(1, 0, sigma[i])
    pthst = dnorm(thst)
    if (runif(1, 0, 1) < pthst/pth) {
      th[m, i] = thst; pth <- pthst
    }
    else
      th[m, i] = th[m-1, i]
  end
}
}
```

Graphiques, $\Sigma = 0.24, 2.4, 24$



L'écart-type et l'efficacité numérique de la moyenne

- ▶ Il y a un théorème central limite pour $M^{-1} \sum_{m=1}^M g(\theta^{(m)})$.
- ▶ Le paquet `mcmcse` calcule l'écart-type de cette moyenne.
- ▶ L'efficacité numérique relative donne le nombre de tirages iid qui équivaut à chaque tirage MCMC.
- ▶ Voici les calculs pour $M^{-1} \sum_{m=1}^M \theta^{(m)}$ (pour $g(\theta) = \theta$)

```
library(mcmcse); library(tidyverse); library(knitr)
mn <- rep(0, 3); sd <- rep(0, 3); nse <- rep(0, 3); rne <-
for (i in 1:3) {
  mc <- mcse(th[, i])
  mn[i] <- mc$est                      # Moyenne de l'échantillon
  nse[i] <- mc$se                       # Écart-type de la moyenne
  sd[i] <- sd(th[, i])                  # Écart-type de l'échantillon
  rne[i] <- (sd[i]/nse[i])^2/M          # Efficacité numérique
}
```

Résultats

```
tbl <- tibble(sigma=sigma, mn=mn, sd=sd, nse=nse, rne=rne)
kable(tbl)
```

sigma	mn	sd	nse	rne
0.24	0.2591129	0.7128065	0.1804449	0.0156046
2.40	0.0870327	0.9832893	0.0647938	0.2303010
24.00	-0.0946788	1.1585067	0.2616325	0.0196071

Un modèle à deux paramètres

Nous avons le modèle suivant : pour $i = 1, \dots, n$,

$$y_i = \mu + e_i, \quad e_i \sim \text{iid } N(0, h^{-1}).$$

La densité des données :

$$p(y_1, \dots, y_n | \mu, h) = \left(\frac{h}{2\pi} \right)^{n/2} \exp \left[-\frac{h}{2} \sum_{i=1}^n (y_i - \mu)^2 \right].$$

Une loi *a priori* pour le modèle à deux paramètres

Nous complétons le modèle avec la loi *a priori* où μ and h sont indépendants et

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

Ainsi

$$p(\mu, h) = p(\mu)p(h),$$

où

$$p(\mu) = \left(\frac{\bar{\omega}}{2\pi} \right)^{1/2} \exp \left[-\frac{1}{2} \bar{\omega} (\mu - \bar{\mu})^2 \right],$$

$$p(h) = [2^{\bar{\nu}/2} \Gamma(\bar{\nu}/2)]^{-1} (\bar{s}^2)^{\bar{\nu}/2} h^{(\bar{\nu}-2)/2} \exp \left[-\frac{1}{2} \bar{s}^2 h \right].$$

La loi *a posteriori* pour le modèle gaussien simple

La densité conjointe de μ , h et y :

$$\begin{aligned} p(\mu, h, y) &= p(\mu)p(h)p(y|\mu, h) \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{\bar{\omega}}{2\pi}\right)^{1/2} [2^{\bar{\nu}/2}\Gamma(\bar{\nu}/2)]^{-1} (\bar{s}^2)^{\bar{\nu}/2} \\ &\cdot h^{(\bar{\nu}+n-2)/2} \exp \left\{ -\frac{\bar{\omega}}{2}(\mu - \bar{\mu})^2 - \frac{h}{2} \left[\bar{s}^2 + \sum_{t=1}^n (y_t - \mu)^2 \right] \right\} \end{aligned}$$

La densité *a posteriori* est proportionnelle à la densité conjointe :

$$p(\mu, h|y) = \frac{p(\mu, h, y)}{p(y)} \propto p(\mu, h, y).$$

Ainsi

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

Données artificielles pour le modèle à deux paramètres

```
# Valeurs vraies des paramètres
vrai_mu <- 6
vrai_h <- 0.04
vrai_sigma <- 1/sqrt(vrai_h)

# Données artificielles, statistiques exhaustives
set.seed(123456789)
n <- 10
y <- rnorm(n, vrai_mu, vrai_sigma)
y_bar <- mean(y)
y2_bar <- mean(y^2)

# Hyper-paramètres
nu_bar <- 4
s2_bar <- 0.01
mu_bar <- 10
omega_bar <- 0.01
```

Fonctions de log-densité

Log-densité des données

```
lnp_y_mu_h = function(mu,h) {  
    lnp <- (n/2)*(log(h) - log(2*pi)) -  
            0.5*h*n*(y2_bar - 2*y_bar*mu + mu^2)  
}
```

Log-densités a priori de (μ , h)

```
lnp_mu <- function(mu, mu_bar=10, omega_bar=0.01) {  
    lnp <- dnorm(mu, mu_bar, 1/sqrt(omega_bar), log=T)  
}  
lnp_h <- function(h, nu_bar=4, s2_bar=0.01) {  
    lnp <- log(s2_bar) + dchisq(h*s2_bar, nu_bar, log=T)  
}
```

Log-densité a posteriori de (μ , h)/ y , pas normalisée

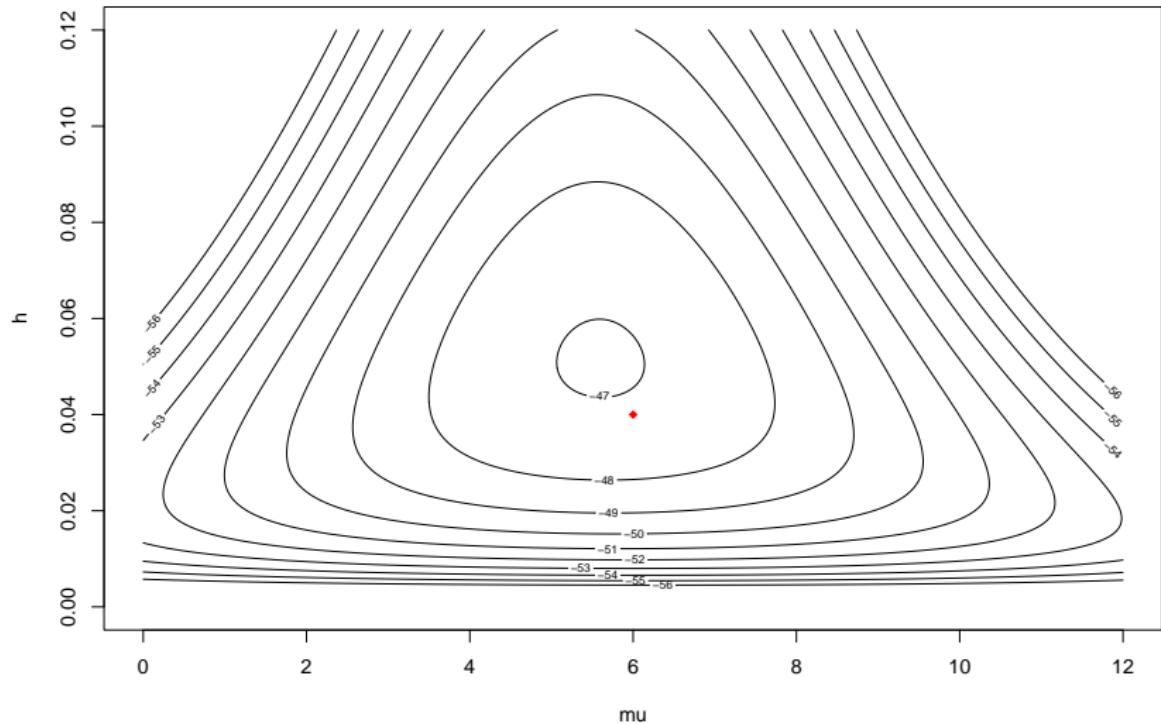
```
lnp_mu_h_y <- function(mu, h) {  
    lnp <- lnp_mu(mu) + lnp_h(h) + lnp_y_mu_h(mu, h)  
}
```

La densité a posteriori, code pour un graphique

```
# Faire le graphique de la log-densité a posteriori,  
# comme fonction de (mu, h)  
lnp_post <- function() {  
    mu <- seq(0, 12, by=0.01)  
    h <- seq(0, 0.12, by=0.0001)  
    lnp <- outer(mu, h, FUN=lnp_mu_h__y)  
    contour(mu, h, lnp, xlab='mu', ylab='h',  
            levels=seq(-56, -46))  
    points(vrai_mu, vrai_h, col='red', pch=18)  
}
```

La densité a posteriori, graphique

`lnp_post()`

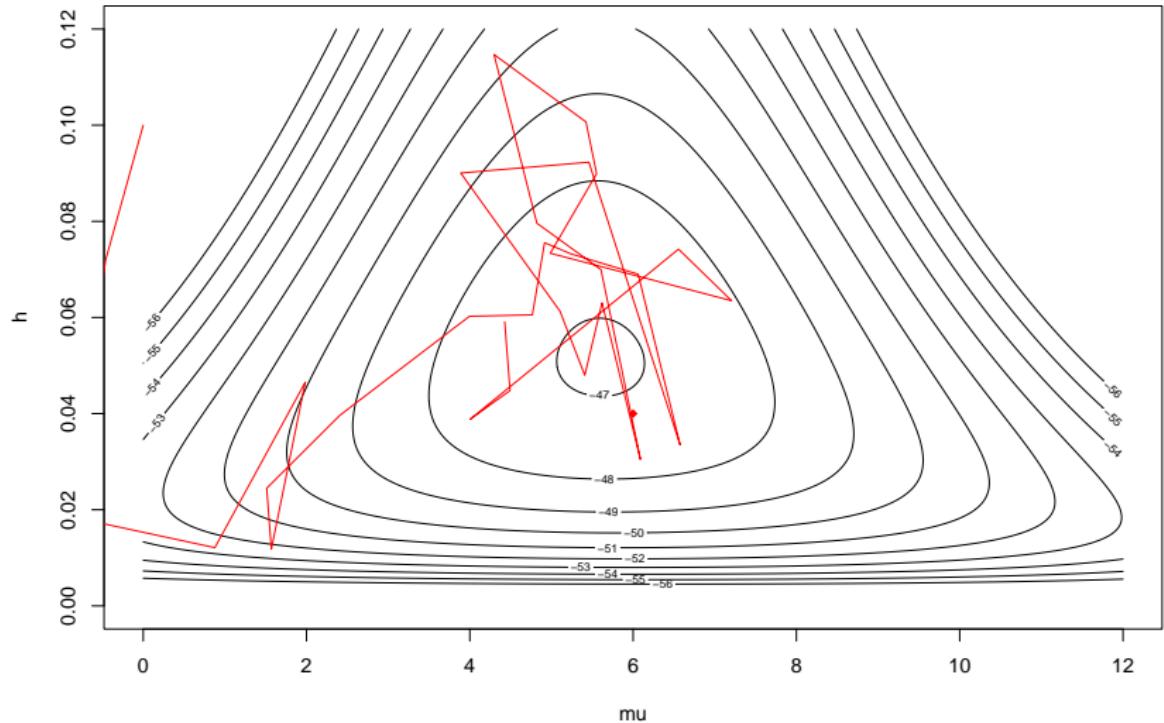


Code Metropolis pour le modèle à deux paramètres

```
Metro.sim = function(M) {  
  mu <- vector('numeric',M); mu[1] <- 0  
  h <- vector('numeric',M); h[1] <- 0.1  
  p <- ((nu_bar+n-2)/2)*log(h[1]) -  
    (omega_bar/2)*(mu[1]-mu_bar)^2 -  
    (h[1]/2)*(s2_bar+n*(y2_bar-2*mu[1]*y_bar+mu[1]^2))  
  for (m in seq(2, M)) {  
    h_et <- rnorm(1, h[m-1], 0.05)  
    mu_et <- rnorm(1, mu[m-1], 2.0)  
    p_et <- ((nu_bar+n-2)/2)*log(h_et) -  
      (omega_bar/2)*(mu_et-mu_bar)^2 -  
      (h_et/2)*(s2_bar + n*(y2_bar-2*mu_et*y_bar+mu_et^2))  
    if (runif(1) < exp(p_et - p) && (h_et > 0.0)) {  
      h[m] <- h_et; mu[m] <- mu_et; p <- p_et  
    }  
    else { h[m] <- h[m-1]; mu[m] <- mu[m-1] }  
  }  
  list(mu=mu, h=h)
```

Trajet Metropolis

```
lnp_post(); set.seed(123); sim <- Metro.sim(100); lines(sim)
```



Échantillonnage de Gibbs

- ▶ Décomposer le vecteur θ en blocs : $\theta = (\theta_1, \dots, \theta_J)$
- ▶ L'idée de base : une mise à jour de θ_i à θ'_i qui préserve la distribution conditionnelle $\theta_i | \theta_{-i}, y$ préserve la distribution $\theta | y$.
- ▶ Exemples :
 - ▶ Tirage directe de θ_i de la distribution $\theta_i | \theta_{-i}, y$,
 - ▶ marche aléatoire Metropolis pour la loi cible $\theta_i | \theta_{-i}, y$.
- ▶ Un balayage (sweep) qui préserve la loi cible $\theta | y$:
 - ▶ Tirer $\theta_1^{(m+1)}$ de la loi $\theta_1 | \theta_2^{(m)}, \dots, \theta_J^{(m)}, y$
 - ▶ Tirer $\theta_2^{(m+1)}$ de la loi $\theta_2 | \theta_1^{(m+1)}, \theta_3^{(m)}, \dots, \theta_J^{(m)}, y$
 - ▶ :
 - ▶ Tirer $\theta_J^{(m+1)}$ de la loi $\theta_J | \theta_1^{(m+1)}, \dots, \theta_{J-1}^{(m+1)}, y$
- ▶ On répète le balayage M fois pour obtenir un échantillon $\theta^{(m)}$, $m = 1, \dots, M$.
- ▶ “Diviser pour vaincre”, si l'ensemble d'étapes est plus facile que le problème entier.

Échantillonnage de Gibbs, modèle à deux paramètres

Si on connaissait $\textcolor{red}{h}$, tirer μ de la loi $\mu|h, y$ serait simple :

$$\begin{aligned} p(\mu|\textcolor{red}{h}, y) &\propto \exp \left[-\frac{\bar{\omega}}{2}(\mu - \bar{\mu})^2 - \frac{\textcolor{red}{h}}{2} \sum_{t=1}^n (y_t - \mu)^2 \right] \\ &= \exp \left\{ -\frac{1}{2} \left[\bar{\omega}(\mu^2 - 2\mu\bar{\mu} + \bar{\mu}^2) + \textcolor{red}{h}(n\bar{y}^{(2)} - 2\mu n\bar{y} + n\mu^2) \right] \right\} \end{aligned}$$

où $\bar{y}^{(2)} = n^{-1} \sum_{t=1}^n y_t^2$ et $\bar{y} = n^{-1} \sum_{t=1}^n y_t$. Alors

$$\begin{aligned} p(\mu|\textcolor{red}{h}, y) &\propto \exp \left\{ -\frac{1}{2} \left[(\bar{\omega} + \textcolor{red}{h}n)\mu^2 - 2(\bar{\omega}\bar{\mu} + \textcolor{red}{h}n\bar{y})\mu \right] \right\} \\ &\propto \exp \left[-\frac{1}{2}(\bar{\omega} + \textcolor{red}{h}n) \left(\mu - \frac{\bar{\omega}\bar{\mu} + \textcolor{red}{h}n\bar{y}}{\bar{\omega} + \textcolor{red}{h}n} \right)^2 \right]. \end{aligned}$$

Alors $\mu|\textcolor{red}{h}, y \sim N(\bar{\bar{\mu}}, \bar{\bar{\omega}}^{-1})$, où $\bar{\bar{\omega}} = \bar{\omega} + \textcolor{red}{h}n$ et

$$\bar{\bar{\mu}} = \frac{\bar{\omega}}{\bar{\omega} + \textcolor{red}{h}n}\bar{\mu} + \frac{\textcolor{red}{h}n}{\bar{\omega} + \textcolor{red}{h}n}\bar{y} = \bar{\bar{\omega}}^{-1}(\bar{\omega}\bar{\mu} + \textcolor{red}{h}n\bar{y}).$$

Échantillonnage de Gibbs, tirage de h

Si on connaissait μ , tirer h de $h|\mu, y$ serait simple :

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

Rappelons que $\bar{s}^2 h \sim \chi^2(\bar{\nu})$ et

$$p(h) = [2^{\bar{\nu}/2} \Gamma(\bar{\nu}/2)]^{-1} (\bar{s}^2)^{\bar{\nu}/2} h^{(\bar{\nu}-2)/2} \exp \left[-\frac{1}{2} \bar{s}^2 h \right].$$

Alors

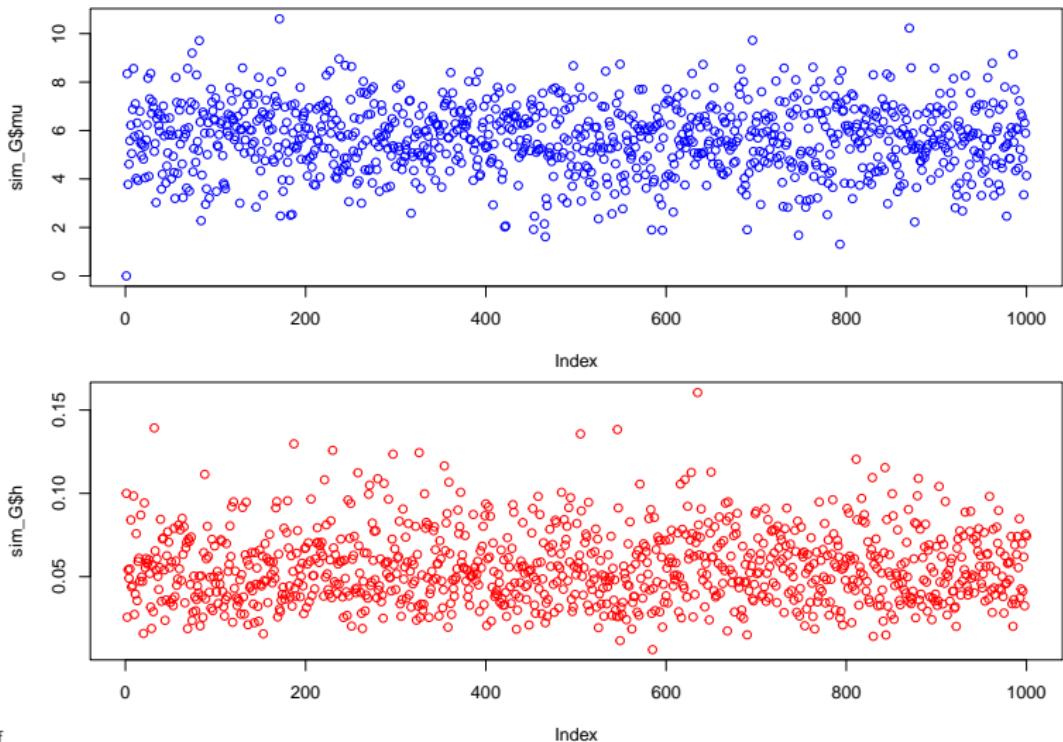
$$\bar{s}^2 h | \mu, y \sim \chi^2(\bar{\nu}),$$

où $\bar{s}^2 = \bar{s}^2 + \sum_{t=1}^n (y_t - \mu)^2$ et $\bar{\nu} = \bar{\nu} + n$.

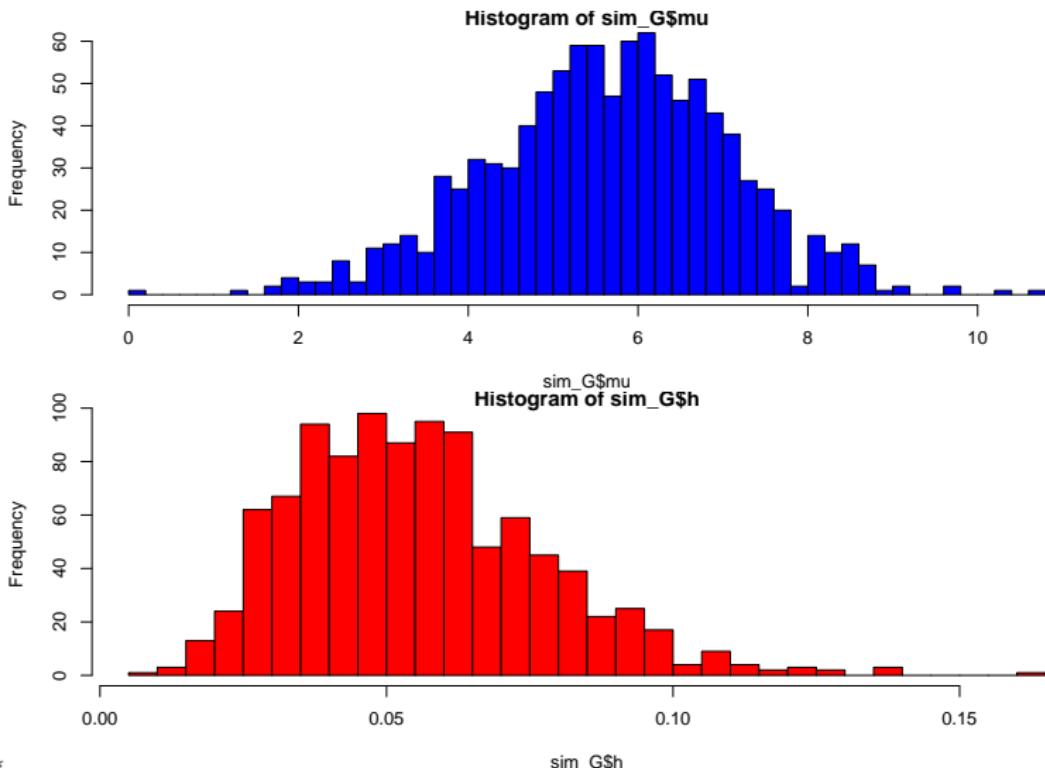
Code pour l'échantillonnage, modèle à deux paramètres

```
Gibbs.sim <- function(M) {  
  # Stockage, valeurs initiales  
  mu <- vector('numeric', M); mu[1] <- 0  
  h <- vector('numeric', M); h[1] <- 0.1  
  
  nu_bar.bar <- nu_bar + n; y.bar = mean(y)  
  for (m in seq(2,M)) {  
    s2_bar.bar <- s2_bar + sum((y-mu[m-1])^2)  
    h[m] <- rchisq(1, nu_bar.bar) / s2_bar.bar  
  
    omega_bar.bar <- omega_bar + h[m]*n  
    mu_bar.bar <- (omega_bar*mu_bar+h[m]*n*y.bar) /  
      omega_bar.bar  
    mu[m] <- rnorm(1, mu_bar.bar,  
                  1/sqrt(omega_bar.bar))  
  }  
  list(mu=mu, h=h)  
}
```

Résultats, trace



Résultats, histograms



Résultats, écarts-types numériques pour μ

```
library(mcmcse)
se <- mcse(sim_G$mu, size='sqroot', method="obm")
se
```

\$est [1] 5.6523

\$se [1] 0.03881128

```
std <- sd(sim_G$mu)
rne <- (std/se$se)^2/M
fmt <- "moy = %f, std = %f, se = %f, rne = %f\n"
cat(sprintf(fmt, se$est, std, se$se, rne))
```

moy = 5.652300, std = 1.417476, se = 0.038811, rne = 1.333877

Résultats, écarts-types numériques pour h

```
se <- mcse(sim_G$h, size='sqroot', method="obm")
std <- sd(sim_G$h)
rne <- (std/se$se)^2/M
cat(sprintf(fmt, se$est, std, se$se, rne))
```

moy = 0.055488, std = 0.021502, se = 0.000800, rne = 0.723202

Exemple économétrique (Exemple 17.1 de Wooldridge)

- ▶ Un modèle de la participation des femmes dans la population active (jeu de données MROZ)
- ▶ On observe un échantillon de $n = 753$ femmes mariées.
- ▶ La variable endogène est la décision d'être dans la population active ou non :

$$y_i = \begin{cases} 1 & \text{la femme } i \text{ est dans la population active} \\ 0 & \text{sinon.} \end{cases}$$

- ▶ Variables exogènes : $nwifeinc$, $educ$, $exper$, $exper^2$, age , $kidslt6$, $kidsge6$, $constant$, avec les données de la femme i organisées dans un vecteur x_i .
- ▶ Modèle probit : $\Pr[y_i = 1] = \Phi(x_i^\top \beta)$, où β est un vecteur de paramètres, $\Phi(\cdot)$ est la fonction de répartition de la loi $N(0, 1)$.
- ▶ Une loi *a priori* pour β : $\beta \sim N(\bar{\beta}, \bar{H}^{-1})$.

L'idée de l'augmentation des données

- ▶ Un exemple de « reculer pour mieux sauter ».
- ▶ Supposez qu'il soit difficile de simuler $\theta|y$.
- ▶ Toutefois, il y a une variable aléatoire z telle que on peut simuler facilement
 - ▶ $\theta|z, y$,
 - ▶ $z|\theta, y$.
- ▶ On utilise l'échantillonnage de Gibbs pour tirer un échantillon de $\theta, z|y$.
- ▶ On peut toujours ignorer z , mais quelquefois z est intéressant en soi.

Augmentation des données

- ▶ La densité *a posteriori* non-normalisée pour le modèle Probit :

$$f(\beta|y) \propto \exp\left[-\frac{1}{2}(\beta - \bar{\beta})^\top \bar{H}(\beta - \bar{\beta})\right] \prod_{i=1}^n \Phi(x_i \beta)^{y_i} (1 - \Phi(x_i \beta))^{1-y_i}.$$

- ▶ Un modèle d'utilité aléatoire : pour chaque femme i ,
 - ▶ y_i^* est la différence d'utilité entre participer dans la population active et ne pas participer.
 - ▶ Sachant x_i et β , $y_i^* = x_i \beta - u_i$ où $u_i \sim N(0, 1)$.
 - ▶ $y_i = 1$ si $y_i^* \geq 0$.
- ▶ Notez que

$$\begin{aligned}\Pr[y_i = 1|x_i, \beta] &= \Pr[y_i^* \geq 0|x_i, \beta] \\ &= \Pr[x_i \beta - u_i \geq 0|x_i, \beta] \\ &= \Pr[u_i \leq x_i \beta|x_i, \beta] \\ &= \Phi(x_i \beta).\end{aligned}$$

Augmentation des données

- ▶ Densité *a posteriori* non-normalisée pour le modèle augmenté :

$$\begin{aligned} f(\beta, y^* | y) \propto & \exp \left[-\frac{1}{2} (\beta - \bar{\beta})^\top \bar{H} (\beta - \bar{\beta}) \right] \\ & \cdot \exp \left[-\frac{1}{2} (y^* - X\beta)^\top (y^* - X\beta) \right] \\ & \cdot \prod_{i=1}^n [y_i 1_{[0, \infty)}(y_i^*) + (1 - y_i) 1_{(-\infty, 0)}(y_i^*)] \end{aligned}$$

- ▶ Maintenant on va dériver les lois conditionnelles *a posteriori*
 - ▶ $f(y_i^* | y_{-i}, \beta, y) \propto f(\beta, y^* | y)$, $i = 1, \dots, n$;
 - ▶ $f(\beta | y^*, y) \propto f(\beta, y^* | y)$.

Densité conditionnelle *a posteriori* de y_i^*

- ▶ Densité conditionnelle *a posteriori* :

$$f(y_i^* | y_{-i}^*, \beta, y) \propto \exp \left[-\frac{1}{2}(y_i^* - x_i \beta)^2 \right] \cdot [y_i 1_{[0, \infty)}(y_i^*) + (1 - y_i) 1_{(-\infty, 0)}(y_i^*)]$$

- ▶ La loi conditionnelle *a posteriori* suit une $N(x_i \beta, 1)$ tronquée à
 - ▶ $[0, \infty)$ si $y_i = 1$,
 - ▶ $(-\infty, 0)$ si $y_i = 0$.

Densité conditionnelle *a posteriori* de β

- ▶ Densité conditionnelle *a posteriori* :

$$f(\beta|y^*, y) \propto \exp \left\{ -\frac{1}{2} \left[(\beta - \bar{\beta})^\top \bar{H} (\beta - \bar{\beta}) + (y^* - X\beta)^\top (y^* - X\beta) \right] \right\}$$

- ▶ L'expression $[\cdot]$ entre crochets est quadratique en β , et

$$\begin{aligned} & (\beta - \bar{\beta})^\top \bar{H} (\beta - \bar{\beta}) + (y^* - X\beta)^\top (y^* - X\beta) \\ &= \beta^\top (\bar{H} + X^\top X) \beta^\top \\ &\quad - \beta^\top (\bar{H}\bar{\beta} + X^\top y^*) \\ &\quad - (\bar{\beta}^\top \bar{H} + (y^*)^\top X) \beta \\ &\quad + \bar{\beta}^\top \bar{H} \bar{\beta} + (y^*)^\top y^* \end{aligned}$$

- ▶ Comme $ax^2 + bx + c = a(x - h)^2 + k$, pour $h = -b/2a$ et $k = c - b^2/4a$ par la complétiion du carré, on peut exprimer la forme quadratique $[\cdot]$ dans la forme $(\beta - \bar{\beta})^\top \bar{H} (\beta - \bar{\beta}) + k$.

Densité conditionnelle *a posteriori* de β

- ▶ On obtient

$$\bar{\bar{H}} = \bar{H} + X^\top X$$

$$\bar{\bar{\beta}} = \bar{\bar{H}}^{-1}(\bar{H}\bar{\beta} + X^\top y^*) = \bar{\bar{H}}^{-1}(\bar{H}\bar{\beta} + X^\top X b),$$

où $b = (X^\top X)^{-1} X^\top y^*$.

- ▶ Puisque

$$f(\beta | y^*, y) \propto \exp \left[-\frac{1}{2} (\beta - \bar{\bar{\beta}})^\top \bar{\bar{H}} (\beta - \bar{\bar{\beta}}) \right]$$

$$\beta | y^*, y \sim N(\bar{\bar{\beta}}, \bar{\bar{H}}^{-1}).$$

Quatre femmes de l'échantillon

- ▶ Les 2 premières femmes ci-dessous font partie de la population active
 - ▶ la décision de la femme 26 est relativement attendue
 - ▶ la décision de la femme 119 est relativement inattendue
- ▶ Les 2 dernières ne font pas partie de la population active
 - ▶ la décision de la femme 502 est relativement inattendue
 - ▶ la décision de la femme 715 est relativement attendue

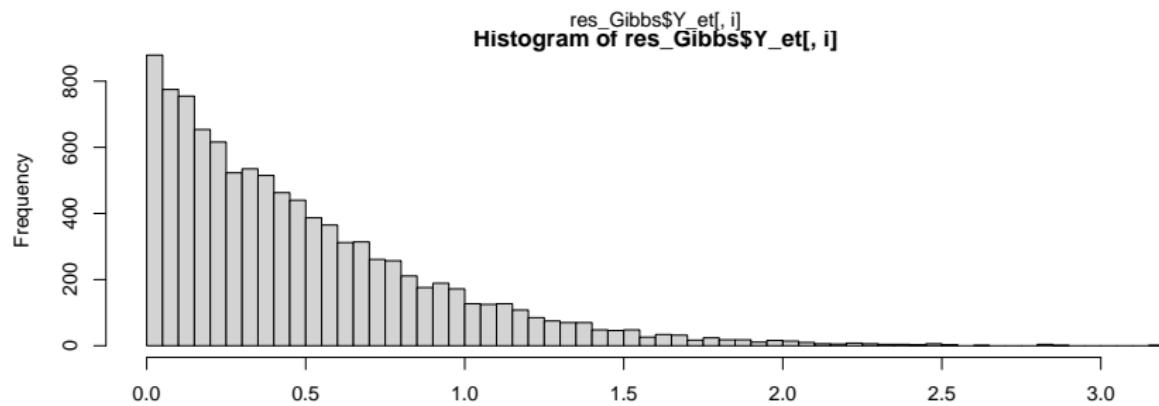
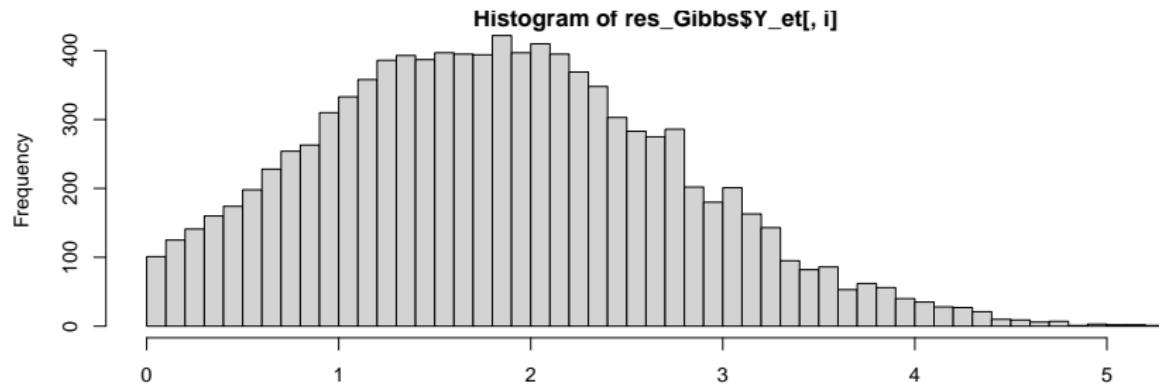
```
source('..../pop_active.R')
X_tbl[c(26, 119, 502, 715),]
```

##	nwifeinc	educ	exper	expersq	age	kidslt6	kidsge6	cons
## 26	27.34999	17	21	441	43	0		2
## 119	91.00000	17	1	1	38	1		3
## 502	24.00000	16	8	64	33	0		0
## 715	51.20000	15	5	25	31	3		0

- ▶ Voir les distributions des y_i^* pourrait être utile.

Deux femmes qui sont dans la population active

```
par(mfrow = c(2, 1), mar = c(4, 4, 0.5, 0.5))
for (i in c(26, 119)) {hist(res_Gibbs$Y_et[, i], 50)}
```



Deux femmes qui ne sont pas dans la population active

```
par(mfrow = c(2, 1), mar = c(4, 4, 0.5, 0.5))
for (i in c(502, 715)) {hist(res_Gibbs$Y_et[, i], 50)}
```

