

# ECN 6338 Cours 9

La génération de variables aléatoires multivariées

William McCausland

2022-03-19

## L'exercice Monte Carlo

Situations où l'intégration par Monte Carlo est convenable :

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

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

- ▶ Souvent,  $p$  est la densité *a posteriori*  $p(\theta|y)$  pour un modèle bayésien, où  $\theta$  est un vecteur de paramètres et  $y$  est un échantillon.

L'exercice Monte Carlo est de générer un échantillon  $\theta^{(m)}$ ,  $m = 1, \dots, M$  tel que

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

quand l'espérance à droite est finie.

# Monte Carlo avec chaînes markoviennes (MCMC)

- ▶ Le simple fait qu'on peut écrire la densité cible  $p(\theta|y)$  sur  $\Theta$  ne veut pas dire qu'il est facile de tirer un échantillon iid de la loi cible  $\theta|y$ .
- ▶ 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  $\Theta$  a une loi de transition  $\theta^{(m+1)}|\theta^{(m)}$ .
  - ▶ La loi cible est une loi stationnaire de la chaîne si  $\theta^{(m+1)}$  suit la loi cible quand  $\theta^{(m)}$  la suit.
  - ▶ 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.

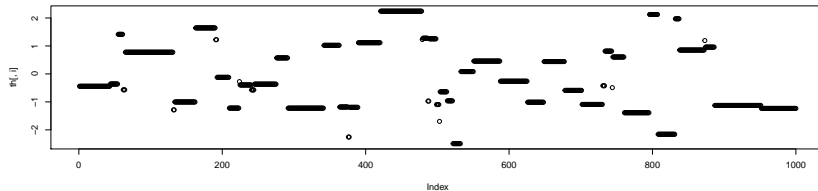
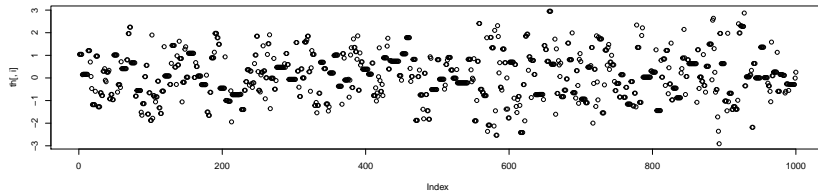
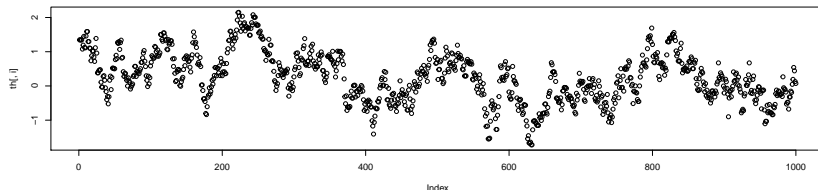
## Mise à jour du type marche aléatoire de Metropolis

- ▶ Supposons que la loi cible  $\theta|y$  ait un noyau de densité  $k(\theta) \propto p(\theta|y)$ .
- ▶ La loi cible est l'unique loi invariante de la transition de Markov suivante, de  $\theta^{(m)}$  à  $\theta^{(m+1)}$  :
  1. Tirer  $\theta^* \sim N(\theta^{(m)}, \Sigma)$ .
  2. Tirer  $U$  de la loi uniform sur  $[0, 1]$ .
  3. Si  $U \leq \frac{k(\theta^*)}{k(\theta^{(m)})}$  fixe  $\theta^{(m+1)} = \theta^*$ , sinon 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 remplacé par une loi asymétrique, avec une modification appropriée du seuil à l'étape 3. (Mise à jour Metropolis-Hastings)
- ▶ Notez que le *ratio de Hastings*  $k(\theta^*)/k(\theta^{(m)})$  égale à  $p(\theta^*|y)/p(\theta^{(m)}|y)$  pour la densité cible normalisée  $p(\theta|y)$ .

## Random walk Metropolis ( $N(0,1)$ target, $\sqrt{\Sigma} = 0.1, 1, 10$ )

```
set.seed(1234567890)
M = 1000
sigma = c(0.1, 1.0, 10.0) * 2.4 # 2.4 optimal dans le cas d
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]
  }
}
```

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



## 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  $\mu$  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  $\mu$ ,  $h$  et  $y$  :

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

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

```
# Vraies valeurs 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)
```

```
# Hyperparamètres
```

```
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 (mu, 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 (mu, h)|y, pas normalisée*

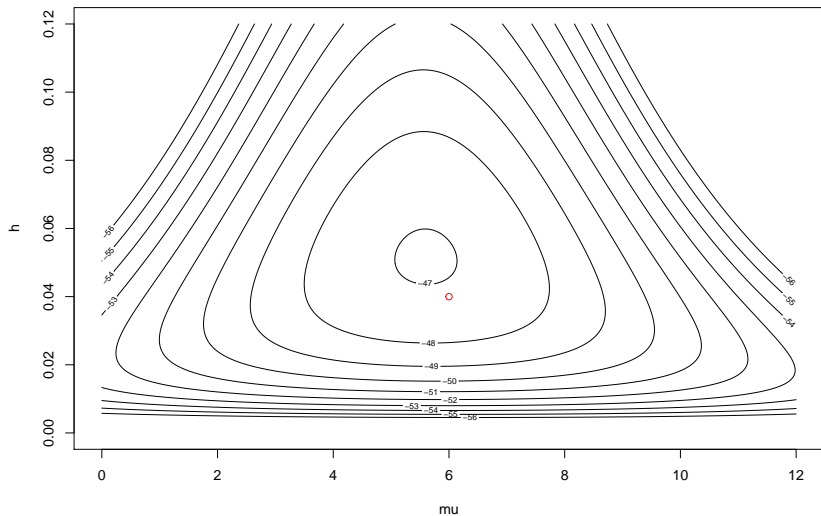
```
lnp_mu_h__y <- function(mu, h) {  
  lnp <- lnp_mu(mu) + lnp_h(h) + lnp_y__mu_h(mu, h)  
}
```

## La densité a posteriori, code pour un graphique

```
# Faire la 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')  
}
```

# La densité a posteriori, graphique

`lnp_post()`

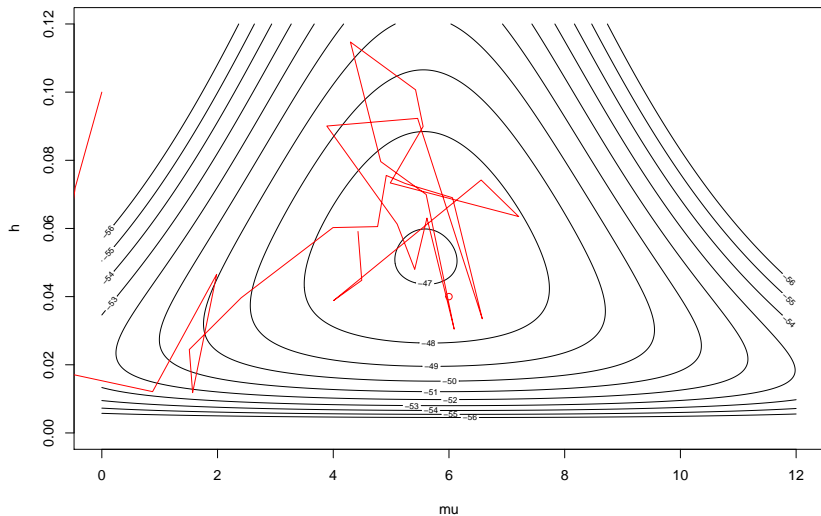


## Code Metropolis pour le modèle à deux paramètres

```
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  $\theta$  en blocs :  $\theta = (\theta_1, \dots, \theta_J)$
- ▶ L'idée de base : une mise à jour de  $\theta_i$  à  $\theta'_i$  qui préserve la distribution conditionnelle  $\theta_i | \theta_{-i}, y$  préserve la distribution  $\theta | y$ .
- ▶ Exemples :
  - ▶ Tirage direct de  $\theta_i$  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  $\theta_2 | \theta_1^{(m+1)}, \theta_3^{(m)}, \dots, \theta_J^{(m)}, y$
  - ▶  $\vdots$
  - ▶ Tirer  $\theta_J^{(m+1)}$  de la loi  $\theta_J | \theta_1^{(m+1)}, \dots, \theta_{J-1}^{(m+1)}, y$



## Échantillonnage de Gibbs pour modèle à deux paramètres

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

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

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

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

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

$$\bar{\bar{\mu}} = \frac{\bar{\omega}}{\bar{\omega} + hn} \bar{\mu} + \frac{hn}{\bar{\omega} + hn} \bar{y} = \bar{\bar{\omega}}^{-1} (\bar{\omega}\bar{\mu} + hn\bar{y}).$$

## Échantillonnage de Gibbs, tirage de $h$

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

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

Rapellons 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

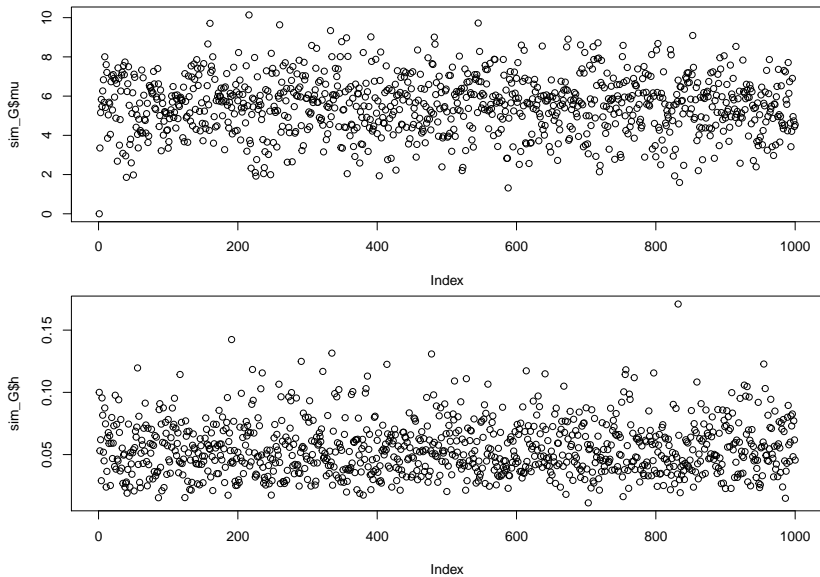
$$\bar{\bar{s}}^2 h|\mu, y \sim \chi^2(\bar{\bar{\nu}}),$$

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

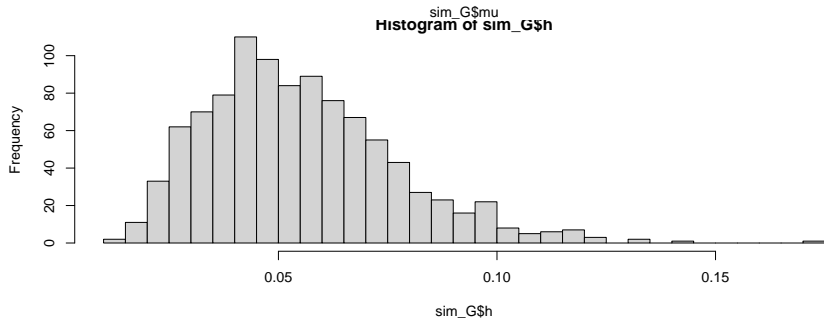
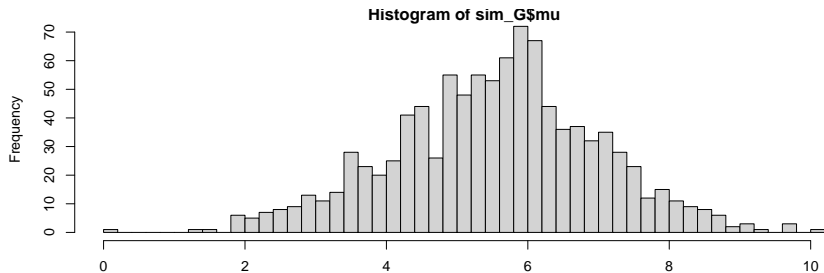
## Échantillonnage de Gibbs, code

```
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 $\mu$

```
library(mcmcse)  
mcse(sim_G$mu)
```

```
## $est  
## [1] 5.535818  
##  
## $se  
## [1] 0.04966814
```

```
sd(sim_G$mu)
```

```
## [1] 1.458128
```

## Résultats, écarts-types numériques pour $h$

```
library(mcmcse)  
mcse(sim_G$h)
```

```
## $est  
## [1] 0.05507961  
##  
## $se  
## [1] 0.0006833799
```

```
sd(sim_G$h)
```

```
## [1] 0.02161037
```

## Augmentation des données

- La densité *a posteriori* non-normalisé 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 où 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  $\beta$ ,  $y_i^* = x_i\beta - u$  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 :

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

- 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* est  $N(x_i \beta, 1)$  tronquée à
  - $[0, \infty)$  si  $y_i = 1$ ,
  - $(-\infty, 0)$  si  $y_i = 0$ .

## Densité conditionnelle *a posteriori* de $\beta$

- 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  $\beta$ , et

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

- Comme  $ax^2 + bx + c = a(x - h)^2 + k$ , pour  $h = -b/2a$  et  $k = c - b^2/4a$  par la complétion 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 $\beta$

- 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

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

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

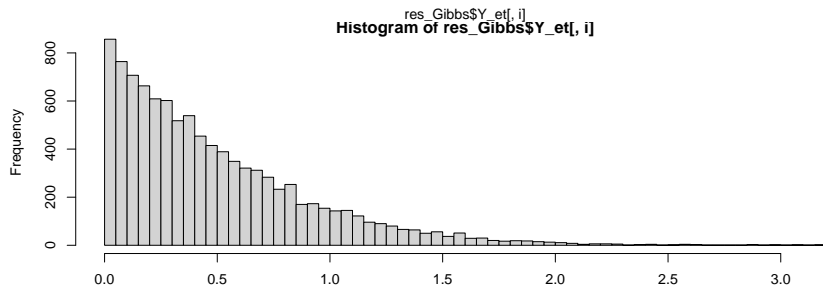
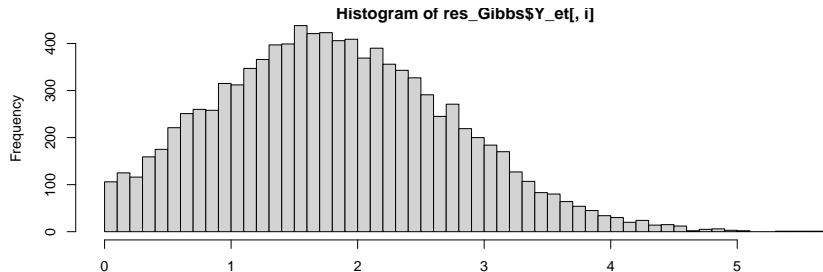
## Quatre femmes

```
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	

## 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)}
```

