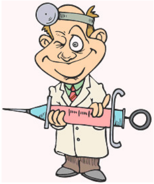
# MAPSI — cours 10 : Échantillonnage et MCMC

Pierre-Henri Wuillemin & Christophe Gonzales

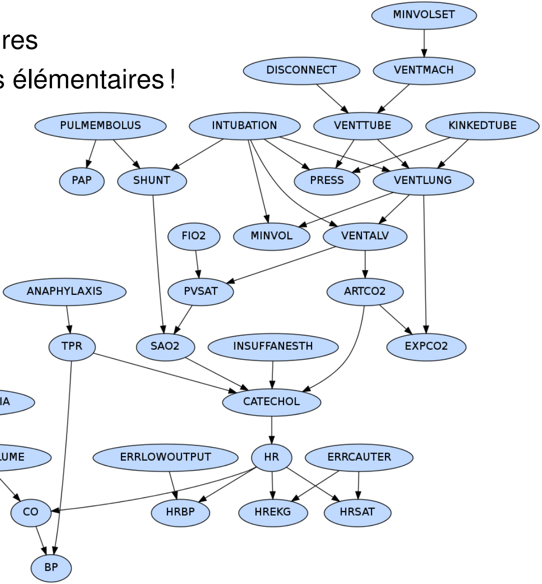
LIP6 / ISIR – Sorbonne Université, France



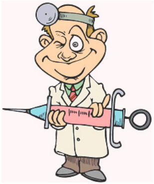
Réseau « alarm » :

37 variables aléatoires

$> 10^{16}$ événements élémentaires !



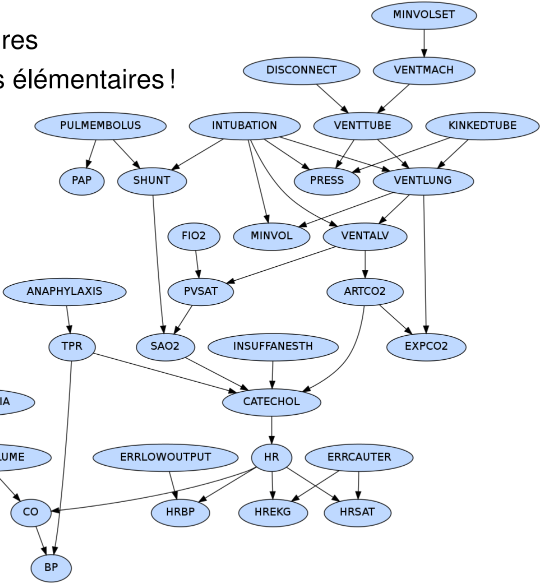
# Motivations : monitoring de patients



Réseau « alarm » :

37 variables aléatoires

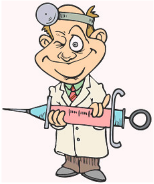
$> 10^{16}$ événements élémentaires !




**Problème :**

calculer $P(BP)$ ?

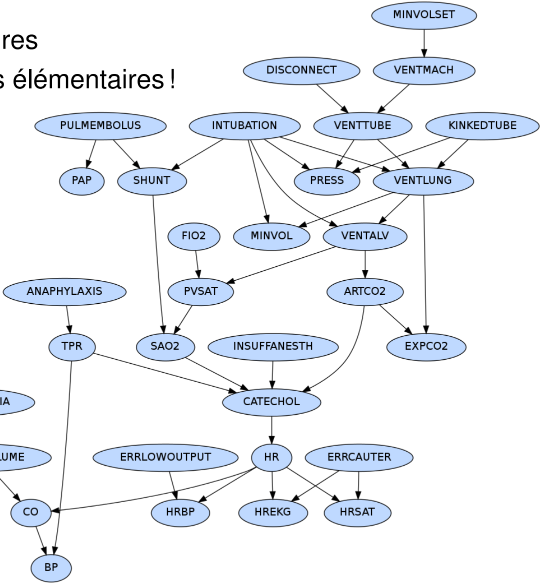
# Motivations : monitoring de patients



Réseau « alarm » :

37 variables aléatoires

$> 10^{16}$ événements élémentaires !

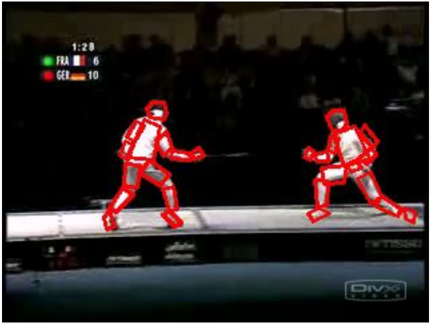


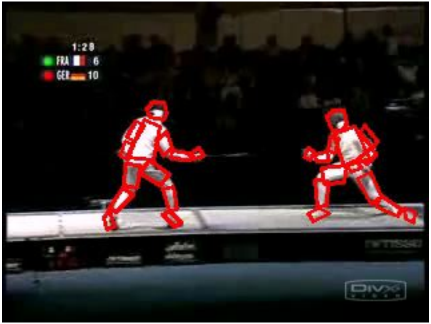

**Problème :**

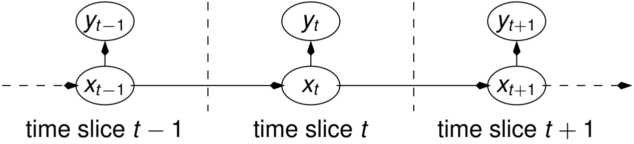
calculer $P(BP)$ ?

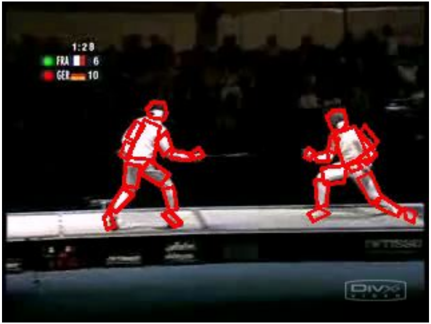
$\Longrightarrow$ calcul d'intégrales

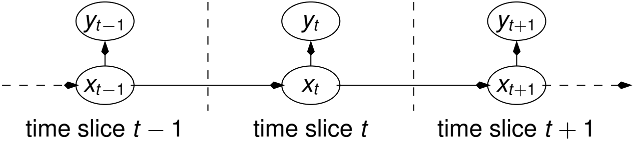
# Motivations : vidéo tracking



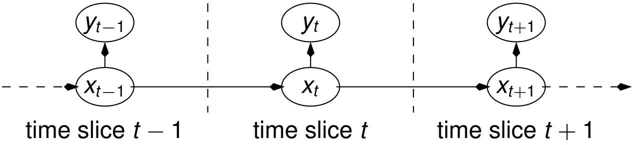

1:28
FRA 6
GER 10


$y_{t-1}$
$y_t$
$y_{t+1}$
$x_{t-1}$
$x_t$
$x_{t+1}$
time slice $t-1$
time slice $t$
time slice $t+1$




- **But :** Estimer $x_t$ sachant $y_{1:t}$ pour tout $t$ : $p(x_t|y_{1:t})$

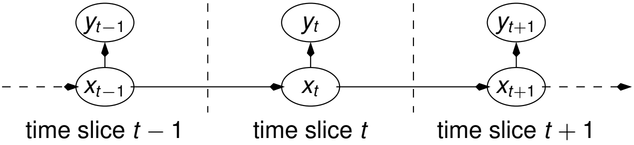



1. Prediction : $p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}$

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

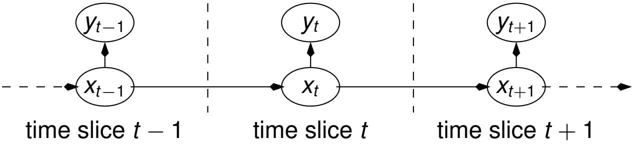
$$P(A,C|B) = \frac{P(A,C,B)}{P(B)} = \frac{P(A,C,B)}{P(C,B)}\frac{P(C,B)}{P(B)} = P(A|C,B)P(C|B)$$

$$P(A|B) = \int P(A,C|B)dC$$




1. Prediction : $p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}$
2. Correction : $p(x_t|y_{1:t}) \propto p(y_t|x_t)p(x_t|y_{1:t-1})$

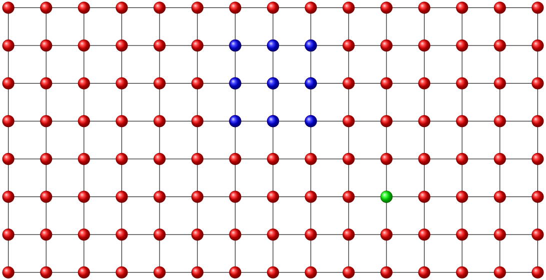
# Motivations : vidéo tracking



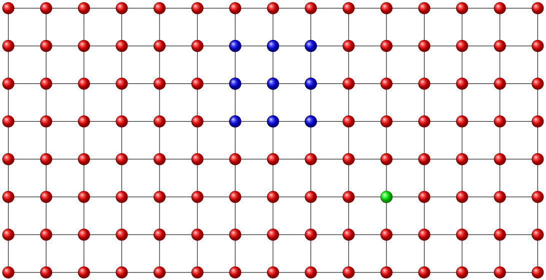

1. Prediction : $p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}$
2. Correction : $p(x_t|y_{1:t}) \propto p(y_t|x_t)p(x_t|y_{1:t-1})$

**Problème : comment faire le calcul à l'étape 1 ?**

# Motivations : modèles d'Ising

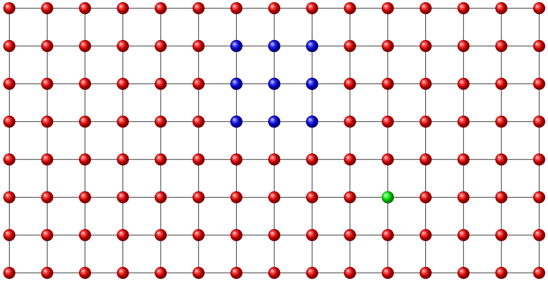


# Motivations : modèles d'Ising



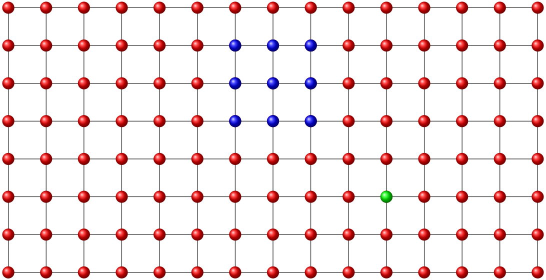
- Énergie : $E = \sum_i \sum_{j \text{ voisin de } i} \psi(i,j) + \sum_i \phi(i)$

# Motivations : modèles d'Ising



- Énergie : $E = \sum_i \sum_{j \text{ voisin de } i} \psi(i,j) + \sum_i \phi(i)$
- Probabilité d'une configuration : $P = \frac{1}{Z} e^{-\beta E}$

# Motivations : modèles d'Ising



- Énergie : $E = \sum_i \sum_{j \text{ voisin de } i} \psi(i,j) + \sum_i \phi(i)$
- Probabilité d'une configuration : $P = \frac{1}{Z} e^{-\beta E}$

**Problèmes : calculer $Z$ et calculer $E$ ?**

- Applications : magnétisme, gaz, neuroscience, nouveaux modèles probabilistes...

# Échantillonnage en statistique

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

# Échantillonnage en statistique

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x) dx$$

# Échantillonnage en statistique

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x) dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

# Échantillonnage en statistique

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x)dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret :

# Échantillonnage en statistique

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x)dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x)dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) :

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x) dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) : $\int L \times P$ car *posterior* $= \frac{L \times P}{\int L \times P}$

# Échantillonnage en statistique

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x)dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) : $\int L \times P$ car *posterior* $= \frac{L \times P}{\int L \times P}$
- Marginaliser une distribution jointe : $P(x_2) =$

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x)dx \approx \frac{1}{N}\sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) : $\int L \times P$ car *posterior* $= \frac{L \times P}{\int L \times P}$
- Marginaliser une distribution jointe : $P(x_2) = \int P(x_1, x_2)dx_1$

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x) dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) : $\int L \times P$ car *posterior* $= \frac{L \times P}{\int L \times P}$
- Marginaliser une distribution jointe : $P(x_2) = \int P(x_1, x_2) dx_1$
- Statistiques sur une distribution :

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x)dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) : $\int L \times P$ car *posterior* $= \frac{L \times P}{\int L \times P}$
- Marginaliser une distribution jointe : $P(x_2) = \int P(x_1, x_2)dx_1$
- Statistiques sur une distribution :
  - Moyenne de $P$ :

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x) dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) : $\int L \times P$ car *posterior* $= \frac{L \times P}{\int L \times P}$
- Marginaliser une distribution jointe : $P(x_2) = \int P(x_1, x_2) dx_1$
- Statistiques sur une distribution :
    - Moyenne de $P$ : $f(x) = x$

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x)dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) : $\int L \times P$ car *posterior* $= \frac{L \times P}{\int L \times P}$
- Marginaliser une distribution jointe : $P(x_2) = \int P(x_1, x_2)dx_1$
- Statistiques sur une distribution :
  - Moyenne de $P$ : $f(x) = x$
  - Moment d'ordre 2 de $P$ :

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x)dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) : $\int L \times P$ car *posterior* $= \frac{L \times P}{\int L \times P}$
- Marginaliser une distribution jointe : $P(x_2) = \int P(x_1, x_2)dx_1$
- Statistiques sur une distribution :
    - Moyenne de $P$ : $f(x) = x$
    - Moment d'ordre 2 de $P$ : $f(x) = x^2$

# Échantillonnage en statistique

L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x)dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) : $\int L \times P$ car *posterior* $= \frac{L \times P}{\int L \times P}$
- Marginaliser une distribution jointe : $P(x_2) = \int P(x_1, x_2)dx_1$
- Statistiques sur une distribution :
  - Moyenne de $P$ : $f(x) = x$
  - Moment d'ordre 2 de $P$ : $f(x) = x^2$
  - $P(A)$ :

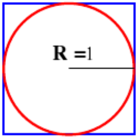
L'échatillonnage (méthodes de Monte Carlo) proposent une simulation stochastique pour le calcul d'intégrales (ou d'équations différentielles).

$$E_P(f) = \int f(x) \cdot P(x) dx \approx \frac{1}{N} \sum_{i \leq N} f(X_i), \forall i, X_i \text{ iid, suivant } P$$

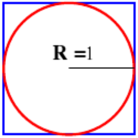
Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques :

- À partir de *posterior* $\propto L \times P$ , calculer la constante de normalisation (Z) : $\int L \times P$ car *posterior* $= \frac{L \times P}{\int L \times P}$
- Marginaliser une distribution jointe : $P(x_2) = \int P(x_1, x_2) dx_1$
- Statistiques sur une distribution :
  - Moyenne de $P$ : $f(x) = x$
  - Moment d'ordre 2 de $P$ : $f(x) = x^2$
  - $P(A)$ :$f(X) = \mathbf{1}_A$

# Calcul (un peu tiré par les cheveux) de $\pi$ par échantillonnage

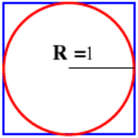


# Calcul (un peu tiré par les cheveux) de $\pi$ par échantillonnage



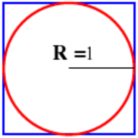

$\pi$

# Calcul (un peu tiré par les cheveux) de $\pi$ par échantillonnage



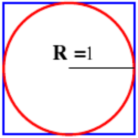

$$\pi = \int_{\bigcirc} dx$$

# Calcul ($_{\text{un peu tiré par les cheveux}}$) de $\pi$ par échantillonnage




$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

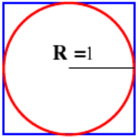
où $p$ distribution uniforme sur $\square$.




$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

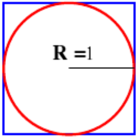
Méthode :




$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

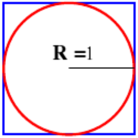
Méthode : on jette des cailloux dans le carré.




$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

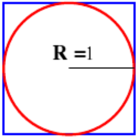
Méthode : on jette des cailloux dans le carré.
Hypothèse :




$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x) dx \propto \int_{\square} p(x).1_{\bigcirc}(x) dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

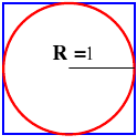



$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} =$$

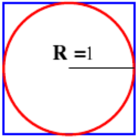



$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2 \cdot R)^2} = \frac{\pi}{4}$$



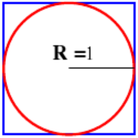

$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2\cdot R)^2} = \frac{\pi}{4}$$

6 jets dans le cercle sur 10 jets en tout



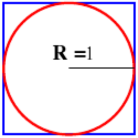

$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2 \cdot R)^2} = \frac{\pi}{4}$$

6 jets dans le cercle sur 10 jets en tout $\Rightarrow$ $\hat{\pi}_{10} = 2.4$



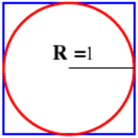

$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2 \cdot R)^2} = \frac{\pi}{4}$$

6 jets dans le cercle sur 10 jets en tout $\Rightarrow \hat{\pi}_{10} = 2.4$
89 jets dans le cercle sur 100 jets en tout



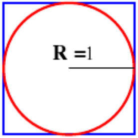

$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2 \cdot R)^2} = \frac{\pi}{4}$$

6 jets dans le cercle sur 10 jets en tout $\Rightarrow\ \hat{\pi}_{10} = 2.4$
89 jets dans le cercle sur 100 jets en tout $\Rightarrow\ \hat{\pi}_{100} = 3.57$



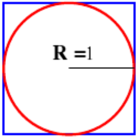

$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2 \cdot R)^2} = \frac{\pi}{4}$$

6 jets dans le cercle sur 10 jets en tout $\Rightarrow\ \hat{\pi}_{10} = 2.4$
89 jets dans le cercle sur 100 jets en tout $\Rightarrow\ \hat{\pi}_{100} = 3.57$
750 jets dans le cercle sur 1000 jets en tout



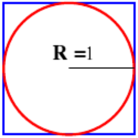

$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2\cdot R)^2} = \frac{\pi}{4}$$

6 jets dans le cercle sur 10 jets en tout $\Rightarrow\ \hat{\pi}_{10} = 2.4$
89 jets dans le cercle sur 100 jets en tout $\Rightarrow\ \hat{\pi}_{100} = 3.57$
750 jets dans le cercle sur 1000 jets en tout $\Rightarrow\ \hat{\pi}_{1000} = 3$



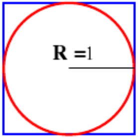

$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2 \cdot R)^2} = \frac{\pi}{4}$$

6 jets dans le cercle sur 10 jets en tout $\Rightarrow \hat{\pi}_{10} = 2.4$
89 jets dans le cercle sur 100 jets en tout $\Rightarrow \hat{\pi}_{100} = 3.57$
750 jets dans le cercle sur 1000 jets en tout $\Rightarrow \hat{\pi}_{1000} = 3$
7852 jets dans le cercle sur 10000 jets en tout



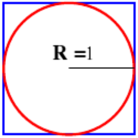

$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2 \cdot R)^2} = \frac{\pi}{4}$$

6 jets dans le cercle sur 10 jets en tout $\Rightarrow \hat{\pi}_{10} = 2.4$
89 jets dans le cercle sur 100 jets en tout $\Rightarrow \hat{\pi}_{100} = 3.57$
750 jets dans le cercle sur 1000 jets en tout $\Rightarrow \hat{\pi}_{1000} = 3$
7852 jets dans le cercle sur 10000 jets en tout $\Rightarrow \hat{\pi}_{10000} = 3.1408$

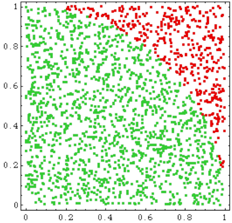



$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

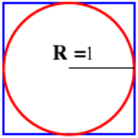
Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2\cdot R)^2} = \frac{\pi}{4}$$




6 jets dans le cercle sur 10 jets en tout $\Rightarrow$ $\hat{\pi}_{10} = 2.4$
89 jets dans le cercle sur 100 jets en tout $\Rightarrow$ $\hat{\pi}_{100} = 3.57$
750 jets dans le cercle sur 1000 jets en tout $\Rightarrow$ $\hat{\pi}_{1000} = 3$
7852 jets dans le cercle sur 10000 jets en tout $\Rightarrow$ $\hat{\pi}_{10000} = 3.1408$

# Calcul (un peu tiré par les cheveux) de $\pi$ par échantillonnage

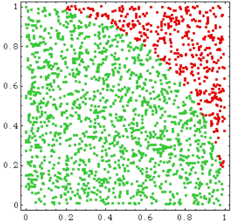



$$\pi = \int_{\bigcirc} dx = \int_{\square} 1_{\bigcirc}(x)dx \propto \int_{\square} p(x).1_{\bigcirc}(x)dx$$

où $p$ distribution uniforme sur $\square$.

Méthode : on jette des cailloux dans le carré.
Hypothèse : les jets suivent une distribution uniformes
d'où

$$\frac{1}{N}\sum_{i\leq N} 1_{\bigcirc}(X_i) = \frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2 \cdot R)^2} = \frac{\pi}{4}$$




6 jets dans le cercle sur 10 jets en tout $\Rightarrow$ $\hat{\pi}_{10} = 2.4$
89 jets dans le cercle sur 100 jets en tout $\Rightarrow$ $\hat{\pi}_{100} = 3.57$
750 jets dans le cercle sur 1000 jets en tout $\Rightarrow$ $\hat{\pi}_{1000} = 3$
7852 jets dans le cercle sur 10000 jets en tout $\Rightarrow$ $\hat{\pi}_{10000} = 3.1408$

# Plan du cours n°9

1. Échantillonnage d'une loi discrète
2. Rejection sampling
3. MCMC : Metropolis-Hastings
4. MCMC : échantillonneur de Gibbs

# Échantillonnage avec une distribution discrète

- *Problème :* échantillonner selon :

distribution $\overset{\infty}{\pi}(X) =$

| $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|
| 0,2 | 0,1 | 0,3 | 0,4 |

- *Problème :* échantillonner selon :

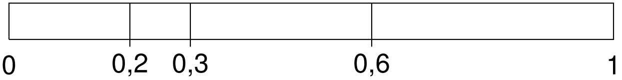
distribution $\overset{\infty}{\pi}(X) =$

| $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|
| 0,2 | 0,1 | 0,3 | 0,4 |

- *Solution :*
  1. Calculer la cumulative :

$F(X_i) = \sum_{Y \leq X_i} \overset{\infty}{\pi}(Y) =$

| | | | |
|---|---|---|---|
| 0,2 | 0,3 | 0,6 | 1 |



# Échantillonnage avec une distribution discrète

- *Problème :* échantillonner selon :

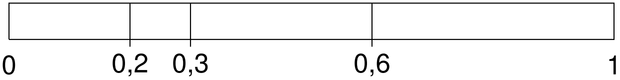
distribution $\overset{\infty}{\pi}(X) =$

| $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|
| 0,2 | 0,1 | 0,3 | 0,4 |

- *Solution :*
  1. Calculer la cumulative :

$F(X_i) = \sum_{Y \leq X_i} \overset{\infty}{\pi}(Y) =$

| | | | |
|---|---|---|---|
| 0,2 | 0,3 | 0,6 | 1 |




  2. Tirer un nombre $z$ selon une distribution uniforme sur [0,1[

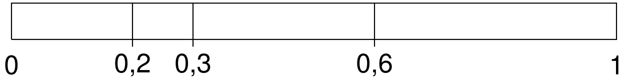
# Échantillonnage avec une distribution discrète

- *Problème :* échantillonner selon :

$$\text{distribution } \overset{\infty}{\pi}(X) = \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \\ \hline \multicolumn{1}{|c|}{0,2} & \multicolumn{1}{c|}{0,1} & \multicolumn{1}{c|}{0,3} & \multicolumn{1}{c|}{0,4} \\ \hline \end{array}$$

- *Solution :*
  1. Calculer la cumulative :

$$F(X_i) = \sum_{Y \leq X_i} \overset{\infty}{\pi}(Y) = \begin{array}{|c|c|c|c|} \hline 0,2 & 0,3 & 0,6 & 1 \\ \hline \end{array}$$




  2. Tirer un nombre $z$ selon une distribution uniforme sur $[0,1[$
  3. Soit $i$ tel que $F(X_{i-1}) \leq z < F(X_i)$ (en posant $X_0 = 0$)

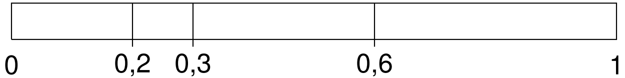
# Échantillonnage avec une distribution discrète

- *Problème :* échantillonner selon :

$$\text{distribution } \overset{\infty}{\pi}(X) = \begin{array}{|c|c|c|c|} \multicolumn{1}{c}{X_1} & \multicolumn{1}{c}{X_2} & \multicolumn{1}{c}{X_3} & \multicolumn{1}{c}{X_4} \\ \hline 0{,}2 & 0{,}1 & 0{,}3 & 0{,}4 \\ \hline \end{array}$$

- *Solution :*
  1. Calculer la cumulative :

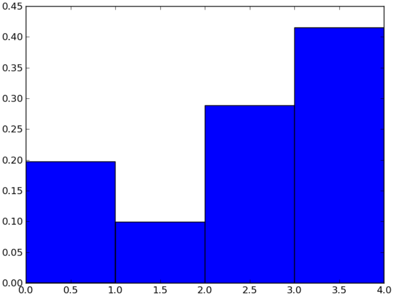
$$F(X_i) = \sum_{Y \leq X_i} \overset{\infty}{\pi}(Y) = \begin{array}{|c|c|c|c|} \hline 0{,}2 & 0{,}3 & 0{,}6 & 1 \\ \hline \end{array}$$




  2. Tirer un nombre $z$ selon une distribution uniforme sur [0,1[
  3. Soit $i$ tel que $F(X_{i-1}) \leq z < F(X_i)$ (en posant $X_0 = 0$)
  4. Renvoyer $X_i$

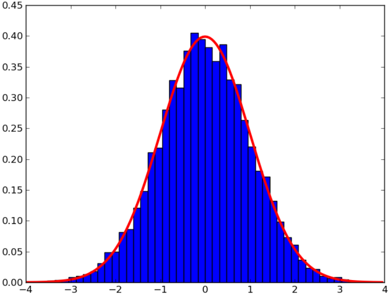
# Échantillonnage avec une distribution discrète

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| $\overset{\infty}{\pi}(X) =$ | 0,2 | 0,1 | 0,3 | 0,4 |



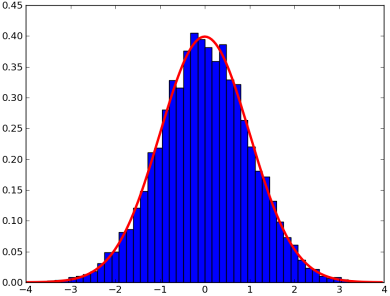
## Échantillonnage d'une loi normale

- Faire la cumulative de la fonction de densité (cf. table)



## Échantillonnage d'une loi normale

- Faire la cumulative de la fonction de densité (cf. table)




Il existe des algos dédiés performants (Ziggurat, Box-Muller)

# Distributions complexes : méthode d'inversion

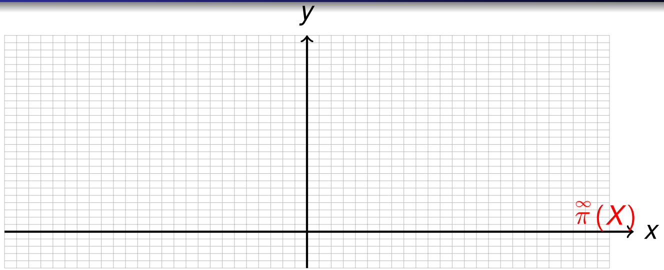
- Variable aléatoire continue $x$ de densité $f$ et fonction de répartition $F$
- $F(x) = P(X \leq x) = \int_{y=-\infty}^{x} f(y)dy$
  $\Rightarrow$ Inverser la fonction de répartition : $x = F^{-1}(u)$

$$\boxed{\text{Échantillonnage de } x \sim f}$$

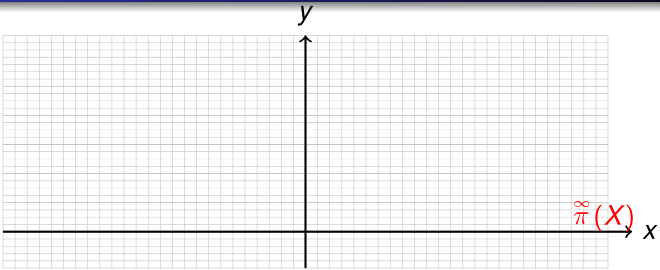
1. Tirer $u \sim U([0,1])$ - loi uniforme sur $[0,1]$
2. Appliquer $x = F^{-1}(u)$

$\Rightarrow$ x distribué selon f (preuve en TD)

- Nécessite la forme analytique de l'inverse $F^{-1}$

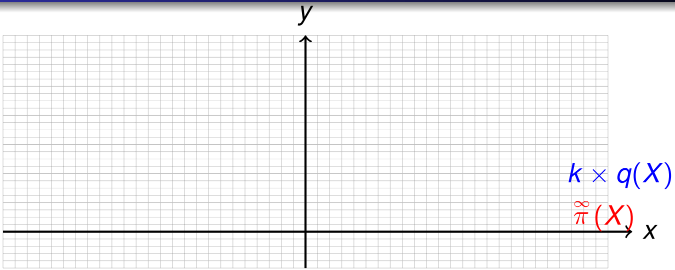

y
$\overset{\infty}{\pi}(X)$
x

# Distributions complexes : Rejection Sampling



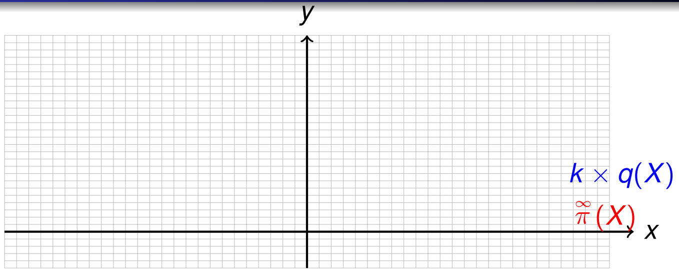

**Hypothèses :**

- $\overset{\infty}{\pi}(\cdot)$ difficile à échantillonner
- Mais pour tout $x \in X$, $\overset{\infty}{\pi}(x)$ facile à calculer

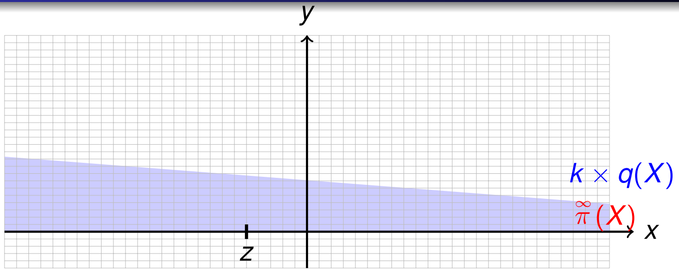



**Hypothèses :**

- $\overset{\infty}{\pi}(\cdot)$ difficile à échantillonner
- Mais pour tout $x \in X$, $\overset{\infty}{\pi}(x)$ facile à calculer
- $q(\cdot)$ facile à échantillonner
- il existe $k \in \mathbb{R}$ tel que $\overset{\infty}{\pi}(x) \leq k \times q(x)$ pour tout $x \in X$

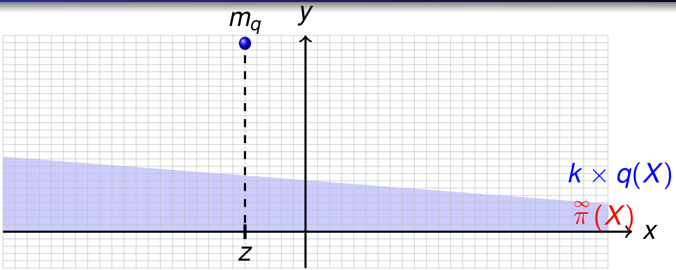



Algorithme « rejection sampling » :



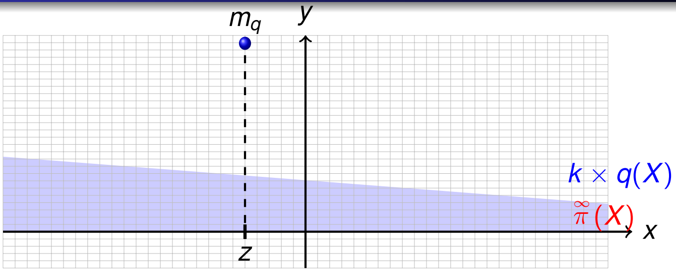

**Algorithme « rejection sampling » :**

1. Tirer un nombre $z$ selon $q(\cdot)$




**Algorithme « rejection sampling » :**

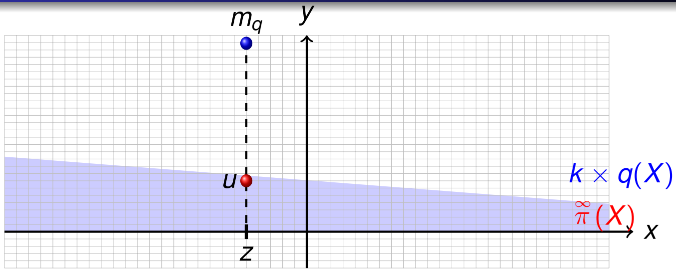
1. Tirer un nombre $z$ selon $q(\cdot)$
2. Calculer $m_q = k \times q(z)$




**Algorithme « rejection sampling » :**

1. Tirer un nombre $z$ selon $q(\cdot)$
2. Calculer $m_q = k \times q(z)$

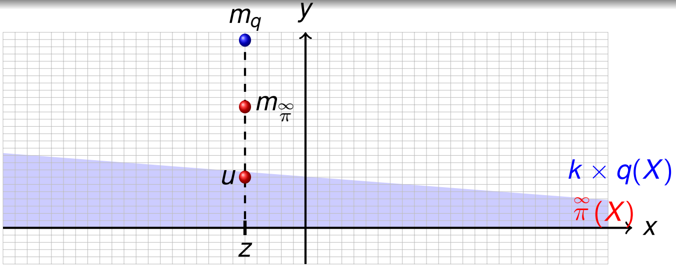
# Distributions complexes : Rejection Sampling




**Algorithme « rejection sampling » :**

1. Tirer un nombre $z$ selon $q(\cdot)$
2. Calculer $m_q = k \times q(z)$
3. Tirer un nombre $u$ selon une loi uniforme sur $[0, m_q[$

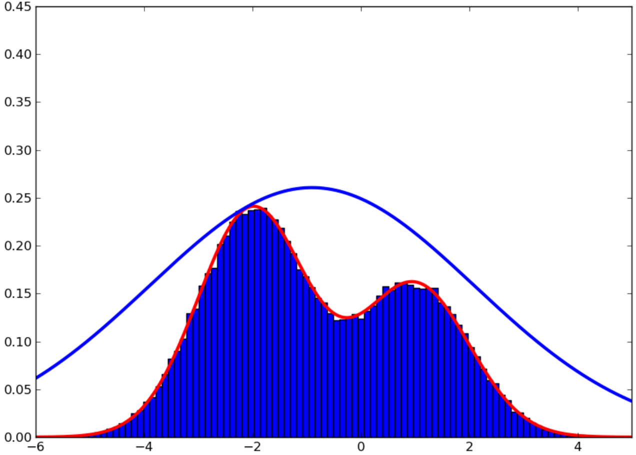
# Distributions complexes : Rejection Sampling




**Algorithme « rejection sampling » :**

1. Tirer un nombre $z$ selon $q(\cdot)$
2. Calculer $m_q = k \times q(z)$
3. Tirer un nombre $u$ selon une loi uniforme sur $[0, m_q[$
3. Accepter $z$ si $u \leq \overset{\infty}{\pi}(z) = m_{\overset{\infty}{\pi}}$

- $\overset{\infty}{\pi}(z)$ : mélange de Gaussiennes
- $q(z)$ Gaussienne, $k = 2$



# Rejection Sampling : preuve

- $P(X \leq x | X \text{ accepté}) = \frac{P(X \leq x, X \text{ accepté})}{P(X \text{ accepté})}$
- Calcul du taux d'acceptation :

  $$P(\text{acceptation}) = \int_{-\infty}^{+\infty} q(z) \times \frac{m_{\frac{\infty}{\pi}}(z)}{m_q(z)} dz$$

- $P(X \leq x | X \text{ accepté}) = \frac{P(X \leq x, X \text{ accepté})}{P(X \text{ accepté})}$
- Calcul du taux d'acceptation :

$$P(\text{acceptation}) = \int_{-\infty}^{+\infty} q(z) \times \frac{m_{\overset{\infty}{\pi}}(z)}{m_q(z)} dz$$

$$= \int_{-\infty}^{+\infty} q(z) \times \frac{\overset{\infty}{\pi}(z)}{k \times q(z)} dz$$

# Rejection Sampling : preuve

- $P(X \leq x | X \text{ accepté}) = \frac{P(X \leq x, X \text{ accepté})}{P(X \text{ accepté})}$
- Calcul du taux d'acceptation :

$$P(\text{acceptation}) = \int_{-\infty}^{+\infty} q(z) \times \frac{m_{\overset{\infty}{\pi}}(z)}{m_q(z)} dz$$

$$= \int_{-\infty}^{+\infty} q(z) \times \frac{\overset{\infty}{\pi}(z)}{k \times q(z)} dz$$

$$= \frac{1}{k} \int_{-\infty}^{+\infty} \overset{\infty}{\pi}(z) dz = \frac{1}{k}$$

- $P(X \leq x, X \text{ accepté}) = \int_{-\infty}^{x} \frac{\overset{\infty}{\pi}(z)}{k \times q(z)} q(z) = \frac{F(x)}{k}$

# Rejection Sampling

**Avantage : fonction de partition inconnue**

- $\overset{\infty}{\pi}(x) = \frac{1}{Z_p} p(x)$
- Seul $p(x)$ connu

## Avantage : fonction de partition inconnue

- $\overset{\infty}{\pi}(x) = \dfrac{1}{Z_p} p(x)$
- Seul $p(x)$ connu
- Nouvelle règle : $k \times q(x) \geq p(x)$ pour tout $x$
- Rejection sampling $\Longrightarrow$ échantillon $\langle z_1, \ldots, z_n \rangle \sim \hat{\overset{\infty}{\pi}}(\cdot)$

## Avantage : fonction de partition inconnue

- $\overset{\infty}{\pi}(x) = \dfrac{1}{Z_p} p(x)$
- Seul $p(x)$ connu
- Nouvelle règle : $k \times q(x) \geq p(x)$ pour tout $x$
- Rejection sampling $\Longrightarrow$ échantillon $\langle z_1, \ldots, z_n \rangle \sim \hat{\overset{\infty}{\pi}}(\cdot)$
- $\hat{\overset{\infty}{\pi}}(z) \propto q(z)$

# Rejection Sampling

**Avantage : fonction de partition inconnue**

- $\overset{\infty}{\pi}(x) = \dfrac{1}{Z_p} p(x)$
- Seul $p(x)$ connu
- Nouvelle règle : $k \times q(x) \geq p(x)$ pour tout $x$
- Rejection sampling $\Longrightarrow$ échantillon $\langle z_1, \ldots, z_n \rangle \sim \hat{\overset{\infty}{\pi}}(\cdot)$
- $\hat{\overset{\infty}{\pi}}(z) \propto q(z) \times \dfrac{p(z)}{k \times q(z)}$

# Rejection Sampling

## Avantage : fonction de partition inconnue

- $\overset{\infty}{\pi}(x) = \dfrac{1}{Z_p} p(x)$
- Seul $p(x)$ connu
- Nouvelle règle : $k \times q(x) \geq p(x)$ pour tout $x$
- Rejection sampling $\Longrightarrow$ échantillon $\langle z_1, \ldots, z_n \rangle \sim \overset{\hat{\infty}}{\pi}(\cdot)$
- $\overset{\hat{\infty}}{\pi}(z) \propto q(z) \times \dfrac{p(z)}{k \times q(z)}$
- $\overset{\hat{\infty}}{\pi}(z) \propto \dfrac{p(z)}{k} \propto p(z) \propto \overset{\infty}{\pi}(z)$

$\Longrightarrow$ on peut échantillonner sans connaître la fonction de partition

Calcul du taux d'acceptation :

$$P(\text{acceptation}) = \int q(z) \times \frac{m_{\overset{\infty}{\pi}}(z)}{m_q(z)} dz$$

**Calcul du taux d'acceptation :**

$$P(\text{acceptation}) = \int q(z) \times \frac{m_{\overset{\infty}{\pi}}(z)}{m_q(z)} dz$$

$$= \int q(z) \times \frac{\overset{\infty}{\pi}(z)}{k \times q(z)} dz$$

**Calcul du taux d'acceptation :**

$$
\begin{aligned}
P(\text{acceptation}) &= \int q(z) \times \frac{m_{\overset{\infty}{\pi}}(z)}{m_q(z)} dz \\
&= \int q(z) \times \frac{\overset{\infty}{\pi}(z)}{k \times q(z)} dz \\
&= \frac{1}{k} \int \overset{\infty}{\pi}(z) dz = \frac{1}{k}
\end{aligned}
$$

## Calcul du taux d'acceptation :

$$P(\text{acceptation}) = \int q(z) \times \frac{m_{\overset{\infty}{\pi}}(z)}{m_q(z)} dz$$

$$= \int q(z) \times \frac{\overset{\infty}{\pi}(z)}{k \times q(z)} dz$$

$$= \frac{1}{k} \int \overset{\infty}{\pi}(z) dz = \frac{1}{k}$$

Exemple précédent : $k \approx 2 \Longrightarrow$ seulement 1 $z$ sur 2 accepté !

# Rejection Sampling

## Calcul du taux d'acceptation :

$$P(\text{acceptation}) = \int q(z) \times \frac{m_{\overset{\infty}{\pi}}(z)}{m_q(z)} dz$$

$$= \int q(z) \times \frac{\overset{\infty}{\pi}(z)}{k \times q(z)} dz$$

$$= \frac{1}{k} \int \overset{\infty}{\pi}(z) dz = \frac{1}{k}$$

⚠ Exemple précédent : $k \approx 2 \Longrightarrow$ seulement 1 $z$ sur 2 accepté !

⚠ $k$ augmente exponentiellement avec la dimension de $\overset{\infty}{\pi}(\cdot)$ !

# Principe de l'échantillonnage par MCMC

**MCMC : Markov Chain Monte Carlo**

- **But :** échantillonner selon une loi $\overset{\infty}{\pi}(\cdot)$

### MCMC : Markov Chain Monte Carlo

- But : échantillonner selon une loi $\overset{\infty}{\pi}(\cdot)$
- Principe : construire une suite $(X_i)$ de variables aléatoires tirées selon des lois $\overset{\infty}{\pi}_i(\cdot)$ tendant vers $\overset{\infty}{\pi}(\cdot)$

# Principe de l'échantillonnage par MCMC

**MCMC : Markov Chain Monte Carlo**

- But : échantillonner selon une loi $\overset{\infty}{\pi}(\cdot)$
- Principe : construire une suite $(X_i)$ de variables aléatoires tirées selon des lois $\overset{\infty}{\pi}_i(\cdot)$ tendant vers $\overset{\infty}{\pi}(\cdot)$

  et sélectionner un échantillon $\langle x_i, \dots, x_{m+i} \rangle$
  ou sous-échantillonner : $\langle x_{\sigma(i)}, \dots, x_{\sigma(m+i)} \rangle \Longrightarrow \approx$ i.i.d.

# Principe de l'échantillonnage par MCMC

MCMC : Markov Chain Monte Carlo

- But : échantillonner selon une loi $\overset{\infty}{\pi}(\cdot)$
- Principe : construire une suite $(X_i)$ de variables aléatoires tirées selon des lois $\overset{\infty}{\pi}_i(\cdot)$ tendant vers $\overset{\infty}{\pi}(\cdot)$

  et sélectionner un échantillon $\langle x_i, \ldots, x_{m+i} \rangle$
  ou sous-échantillonner : $\langle x_{\sigma(i)}, \ldots, x_{\sigma(m+i)} \rangle \Longrightarrow \approx$ i.i.d.
- Solution : construire une chaîne de Markov de loi stationnaire $\overset{\infty}{\pi}(\cdot)$

# Loi stationnaire

- soit $P(X_{t+1}|X_t)$ la probabilité de transition (chaîne homogène)

# Loi stationnaire

- soit $P(X_{t+1}|X_t)$ la probabilité de transition (chaîne homogène)

*Loi stationnaire* $\overset{\infty}{\pi}(\cdot)$

$$\overset{\infty}{\pi}(x) = \int_y P(x|y) \overset{\infty}{\pi}(y) dy$$

# Loi stationnaire

- soit $P(X_{t+1}|X_t)$ la probabilité de transition (chaîne homogène)

*Loi stationnaire $\overset{\infty}{\pi}(\cdot)$*

$$\overset{\infty}{\pi}(x) = \int_y P(x|y)\, \overset{\infty}{\pi}(y) dy$$

ici, on connaît $\overset{\infty}{\pi}(\cdot)$ et on cherche $P(\cdot|\cdot)$

- soit $P(X_{t+1}|X_t)$ la probabilité de transition (chaîne homogène)

*Loi stationnaire $\overset{\infty}{\pi}(\cdot)$*

$$\overset{\infty}{\pi}(x) = \int_y P(x|y)\,\overset{\infty}{\pi}(y)dy$$

ici, on connaît $\overset{\infty}{\pi}(\cdot)$ et on cherche $P(\cdot|\cdot)$

**Problème :** sous quelles conditions $P(\cdot|\cdot)$ existe-t-elle ?

Ergodicité ?

# Une condition possible

**Réversibilité**

$$\overset{\infty}{\pi}(x)P(y|x) = \overset{\infty}{\pi}(y)P(x|y), \ \forall x, y$$

⚠ propriété également connue sous le nom de « detailed balance »

# Une condition possible

*Réversibilité*

$$\overset{\infty}{\pi}(x)P(y|x) = \overset{\infty}{\pi}(y)P(x|y), \ \forall x, y$$

propriété également connue sous le nom de « detailed balance »

conséquence :

$$\int_y P(x|y)\,\overset{\infty}{\pi}(y)dy = \int_y P(y|x)\,\overset{\infty}{\pi}(x)dy$$

# Une condition possible

*Réversibilité*

$$\overset{\infty}{\pi}(x)P(y|x) = \overset{\infty}{\pi}(y)P(x|y),\ \forall x, y$$

propriété également connue sous le nom de « detailed balance »

conséquence :

$$\begin{aligned}\int_y P(x|y)\,\overset{\infty}{\pi}(y)dy &= \int_y P(y|x)\,\overset{\infty}{\pi}(x)dy \\ &= \overset{\infty}{\pi}(x)\int_y P(y|x)dy\end{aligned}$$

*Réversibilité*

$$\overset{\infty}{\pi}(x)P(y|x) = \overset{\infty}{\pi}(y)P(x|y), \ \forall x, y$$

propriété également connue sous le nom de « detailed balance »

conséquence :

$$\begin{aligned}\int_y P(x|y)\,\overset{\infty}{\pi}(y)dy &= \int_y P(y|x)\,\overset{\infty}{\pi}(x)dy \\ &= \overset{\infty}{\pi}(x)\int_y P(y|x)dy \\ &= \overset{\infty}{\pi}(x)\end{aligned}$$

$\Longrightarrow \overset{\infty}{\pi}(\cdot)$ loi stationnaire !

# Garantir la réversibilité

- En général, $\overset{\infty}{\pi}(x)P(y|x) \neq \overset{\infty}{\pi}(y)P(x|y)$

# Garantir la réversibilité

- En général, $\overset{\infty}{\pi}(x)P(y|x) \neq \overset{\infty}{\pi}(y)P(x|y)$

*Intérprétation de $\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$*

Le processus markovien va évoluer plus souvent de $x$ vers $y$ que de $y$ vers $x \Longrightarrow$ non réversible.

# Garantir la réversibilité

- En général, $\overset{\infty}{\pi}(x)P(y|x) \neq \overset{\infty}{\pi}(y)P(x|y)$

*Intérprétation de $\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$*

Le processus markovien va évoluer plus souvent de $x$ vers $y$ que de $y$ vers $x$ $\Longrightarrow$ non réversible.

Correction : diminuer $P(y|x)$ ou augmenter $P(x|y)$

# Garantir la réversibilité

- En général, $\overset{\infty}{\pi}(x)P(y|x) \neq \overset{\infty}{\pi}(y)P(x|y)$

*Intérprétation de $\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$*

Le processus markovien va évoluer plus souvent de $x$ vers $y$ que de $y$ vers $x$ $\Longrightarrow$ non réversible.

Correction : diminuer $P(y|x)$ ou augmenter $P(x|y)$

$\Longrightarrow$ créer deux nombres $\alpha(x, y)$ et $\alpha(y, x)$ tels que :

$$\overset{\infty}{\pi}(x)P(y|x)\alpha(x, y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y, x)$$

# Garantir la réversibilité

- En général, $\overset{\infty}{\pi}(x)P(y|x) \neq \overset{\infty}{\pi}(y)P(x|y)$

*Intérprétation de $\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$*

Le processus markovien va évoluer plus souvent de $x$ vers $y$ que de $y$ vers $x$ $\Longrightarrow$ non réversible.

Correction : diminuer $P(y|x)$ ou augmenter $P(x|y)$

$\Longrightarrow$ créer deux nombres $\alpha(x, y)$ et $\alpha(y, x)$ tels que :

$$\overset{\infty}{\pi}(x)P(y|x)\alpha(x, y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y, x)$$

on veut que $P(y|x)\alpha(x, y)$ soit une probabilité de transition !

# Garantir la réversibilité

- En général, $\overset{\infty}{\pi}(x)P(y|x) \neq \overset{\infty}{\pi}(y)P(x|y)$

*Intérprétation de $\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$*

Le processus markovien va évoluer plus souvent de $x$ vers $y$ que de $y$ vers $x \Longrightarrow$ non réversible.

Correction : diminuer $P(y|x)$ ou augmenter $P(x|y)$

$\Longrightarrow$ créer deux nombres $\alpha(x, y)$ et $\alpha(y, x)$ tels que :

$$\overset{\infty}{\pi}(x)P(y|x)\alpha(x, y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y, x)$$

on veut que $P(y|x)\alpha(x, y)$ soit une probabilité de transition !

Remarque : $y = x \Longrightarrow \overset{\infty}{\pi}(x)P(x|x)\alpha(x, x) = \overset{\infty}{\pi}(x)P(x|x)\alpha(x, x)$
pour tout $\alpha(x, x)$

# Garantir la réversibilité

- En général, $\overset{\infty}{\pi}(x)P(y|x) \neq \overset{\infty}{\pi}(y)P(x|y)$

*Intérprétation de $\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$*

Le processus markovien va évoluer plus souvent de $x$ vers $y$ que de $y$ vers $x \Longrightarrow$ non réversible.

Correction : diminuer $P(y|x)$ ou augmenter $P(x|y)$

$\Longrightarrow$ créer deux nombres $\alpha(x, y)$ et $\alpha(y, x)$ tels que :

$$\overset{\infty}{\pi}(x)P(y|x)\alpha(x, y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y, x)$$

on veut que $P(y|x)\alpha(x, y)$ soit une probabilité de transition !

Remarque : $y = x \Longrightarrow \overset{\infty}{\pi}(x)P(x|x)\alpha(x, x) = \overset{\infty}{\pi}(x)P(x|x)\alpha(x, x)$
pour tout $\alpha(x, x)$

Si $P(x|x)\alpha(x, x) = 1 - \displaystyle\int_{y \neq x} P(y|x)\alpha(x, y)dy$, on a bien une proba !

# Garantir la réversibilité

$$P(x|x)\alpha(x,x) = 1 - \int_{y\neq x} P(y|x)\alpha(x,y)dy$$

Pour assurer que $P(x|x)\alpha(x,x) \geq 0$, on impose $\alpha(x,y) \leq 1$

# Garantir la réversibilité

$$P(x|x)\alpha(x,x) = 1 - \int_{y\neq x} P(y|x)\alpha(x,y)dy$$

Pour assurer que $P(x|x)\alpha(x,x) \geq 0$, on impose $\alpha(x,y) \leq 1$

$$\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$$
$$\overset{\infty}{\pi}(x)P(y|x)\alpha(x,y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y,x)$$

# Garantir la réversibilité

$$P(x|x)\alpha(x,x) = 1 - \int_{y \neq x} P(y|x)\alpha(x,y)dy$$

Pour assurer que $P(x|x)\alpha(x,x) \geq 0$, on impose $\alpha(x,y) \leq 1$

$$\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$$
$$\overset{\infty}{\pi}(x)P(y|x)\alpha(x,y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y,x)$$

$\Longrightarrow$ pour augmenter $P(x|y)$, on fixe $\alpha(y,x) = 1$

# Garantir la réversibilité

$$P(x|x)\alpha(x,x) = 1 - \int_{y \neq x} P(y|x)\alpha(x,y)dy$$

Pour assurer que $P(x|x)\alpha(x,x) \geq 0$, on impose $\alpha(x,y) \leq 1$

$$\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$$
$$\overset{\infty}{\pi}(x)P(y|x)\alpha(x,y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y,x)$$

$\Longrightarrow$ pour augmenter $P(x|y)$, on fixe $\alpha(y,x) = 1$

$\Longrightarrow \overset{\infty}{\pi}(x)P(y|x)\alpha(x,y) = \overset{\infty}{\pi}(y)P(x|y)$

# Garantir la réversibilité

$$P(x|x)\alpha(x,x) = 1 - \int_{y\neq x} P(y|x)\alpha(x,y)dy$$

Pour assurer que $P(x|x)\alpha(x,x) \geq 0$, on impose $\alpha(x,y) \leq 1$

$$\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$$
$$\overset{\infty}{\pi}(x)P(y|x)\alpha(x,y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y,x)$$

$\Longrightarrow$ pour augmenter $P(x|y)$, on fixe $\alpha(y,x) = 1$

$\Longrightarrow \overset{\infty}{\pi}(x)P(y|x)\alpha(x,y) = \overset{\infty}{\pi}(y)P(x|y)$

$\Longrightarrow \alpha(x,y) = \dfrac{\overset{\infty}{\pi}(y)P(x|y)}{\overset{\infty}{\pi}(x)P(y|x)}$

# Garantir la réversibilité

**Résumé :**

- Si $\overset{\infty}{\pi}(x)P(y|x) > \overset{\infty}{\pi}(y)P(x|y)$ :

  Fixer $\alpha(x,y) = \dfrac{\overset{\infty}{\pi}(y)P(x|y)}{\overset{\infty}{\pi}(x)P(y|x)}$ et $\alpha(y,x) = 1$

- Par symétrie, si $\overset{\infty}{\pi}(x)P(y|x) < \overset{\infty}{\pi}(y)P(x|y)$ :

  Fixer $\alpha(x,y) = 1$ et $\alpha(y,x) = \dfrac{\overset{\infty}{\pi}(x)P(y|x)}{\overset{\infty}{\pi}(y)P(x|y)}$

# Garantir la réversibilité

*Intérprétation de $\alpha$ : probabilité de mouvement*

$\alpha(x, y) =$ la probabilité de réaliser la transition de $x$ vers $y$

$\Longrightarrow$ à l'étape $t$, on a 2 choix :

- transiter de $x$ vers un $y$ avec la probabilité $P(y|x)\alpha(x, y)$
- ne pas réaliser de transition

*Intérprétation de $\alpha$ : probabilité de mouvement*

$\alpha(x, y) =$ la probabilité de réaliser la transition de $x$ vers $y$

$\Longrightarrow$ à l'étape $t$, on a 2 choix :

- transiter de $x$ vers un $y$ avec la probabilité $P(y|x)\alpha(x, y)$
- ne pas réaliser de transition

*Résumé*

si $\alpha(x, y) = \min\left\{1, \dfrac{\overset{\infty}{\pi}(y)P(x|y)}{\overset{\infty}{\pi}(x)P(y|x)}\right\}$ alors :

$\overset{\infty}{\pi}(x)P(y|x)\alpha(x, y) = \overset{\infty}{\pi}(y)P(x|y)\alpha(y, x)$

$\Longrightarrow$ réversibilité $\Longrightarrow$ $\overset{\infty}{\pi}(\cdot)$ distribution stationnaire

# Algorithme de Metropolis-Hastings

*Metropolis-Hastings*

Algorithme pour générer $x_{t+1}$ à partir de $x_t$ :

1. tirer $z$ selon la distribution $P(\cdot|x_t)$
2. calculer $\alpha(x_t, z) = \min\left\{1, \dfrac{\overset{\infty}{\pi}(z)P(x_t|z)}{\overset{\infty}{\pi}(x_t)P(z|x_t)}\right\}$
3. tirer un nombre $u$ selon une loi uniforme sur $[0, 1[$
4. renvoyer $x_{t+1} = \begin{cases} z & \text{si } u \leq \alpha(x_t, z) \\ x_t & \text{sinon} \end{cases}$

Références :


- N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller et E. Teller (1953) “Equations of State Calculations by Fast Computing Machines”. Journal of Chemical Physics, 21 (6), pp. 1087–1092
- W.K. Hastings (1970) “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”. Biometrika, 57 (1), pp. 97–109

# Choix de $P(\cdot|x_t)$

- $P(\cdot|x_t)$ doit être simple à échantillonner

# Choix de $P(\cdot|x_t)$

- $P(\cdot|x_t)$ doit être simple à échantillonner
- 1ère possibilité [Metropolis *et al.* (1953), Müller (1993)]

  $P(z|x_t) = q(z - x_t)$ avec $q(\cdot)$ densité multivariée

  autrement dit $z = x_t + y$ avec $y \sim q(\cdot)$

  ⚠ $q(\cdot)$ indépendante de $x_t$ !

$\Longrightarrow$ random walk chain

# Choix de $P(\cdot|x_t)$

- $P(\cdot|x_t)$ doit être simple à échantillonner
- 1ère possibilité [Metropolis *et al.* (1953), Müller (1993)]

  $P(z|x_t) = q(z - x_t)$ avec $q(\cdot)$ densité multivariée

  autrement dit $z = x_t + y$ avec $y \sim q(\cdot)$

  ⚠ $q(\cdot)$ indépendante de $x_t$ !

$\Longrightarrow$ random walk chain

- choix possible de $q(\cdot)$ : loi normale

# Choix de $P(\cdot|x_t)$

- $P(\cdot|x_t)$ doit être simple à échantillonner
- 1ère possibilité [Metropolis *et al.* (1953), Müller (1993)]

  $P(z|x_t) = q(z - x_t)$ avec $q(\cdot)$ densité multivariée

  autrement dit $z = x_t + y$ avec $y \sim q(\cdot)$

  $q(\cdot)$ indépendante de $x_t$ !

$\Longrightarrow$ random walk chain

- choix possible de $q(\cdot)$ : loi normale
- si $q$ est symétrique : $q(y) = q(-y)$ et

$$\alpha(x_t, z) = \min\left\{1, \frac{\overset{\infty}{\pi}(z)P(x_t|z)}{\overset{\infty}{\pi}(x_t)P(z|x_t)}\right\} = \min\left\{1, \frac{\overset{\infty}{\pi}(z)}{\overset{\infty}{\pi}(x_t)}\right\}$$

# Choix de $P(\cdot|x_t)$

- 2ème possibilité [Hastings (1970)]

  $P(z|x) = q(z)$ avec $q(\cdot)$ densité multivariée

$\Longrightarrow$ independent chain

$\Longrightarrow$ généralisation de rejection sampling

# Choix de $P(\cdot|x_t)$

- 2ème possibilité [Hastings (1970)]

  $P(z|x) = q(z)$ avec $q(\cdot)$ densité multivariée

$\Longrightarrow$ independent chain

$\Longrightarrow$ généralisation de rejection sampling

- 3ème possibilité : l'algorithme Langevin [Roberts et Rosenthal (1998)]

  $$z = x_t + \frac{\sigma^2}{2}\nabla \log(\overset{\infty}{\pi}(x_t)) + \sigma y \text{ avec } y \sim q(\cdot)$$

  $\sigma$ : facteur d'échelle

# Choix de $P(\cdot|x_t)$

- 2ème possibilité [Hastings (1970)]

  $P(z|x) = q(z)$ avec $q(\cdot)$ densité multivariée

**$\Longrightarrow$ independent chain**

$\Longrightarrow$ généralisation de rejection sampling

- 3ème possibilité : l'algorithme Langevin [Roberts et Rosenthal (1998)]

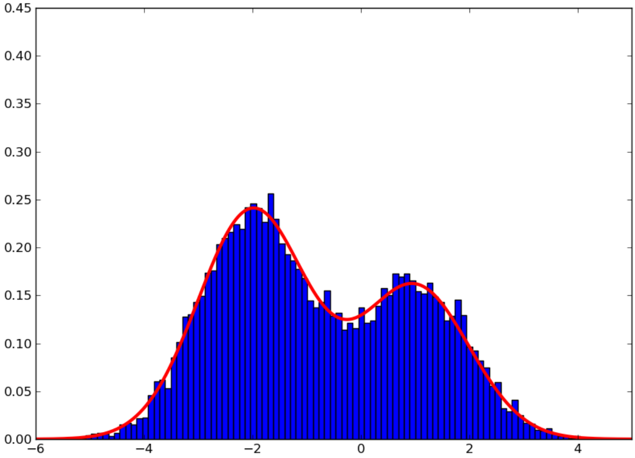
  $$z = x_t + \frac{\sigma^2}{2}\nabla \log(\overset{\infty}{\pi}(x_t)) + \sigma y \text{ avec } y \sim q(\cdot)$$

  $\sigma$ : facteur d'échelle

⚠ Il existe plein d'autres possibilités…

# Illustration de Metropolis-Hastings

Random walk avec une loi normale centrée réduite



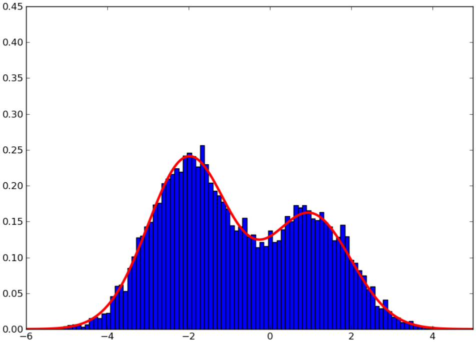
# Choix de l'étalement/variance de $P(\cdot|x_t)$

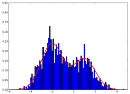
⚠ important pour la vitesse de convergence

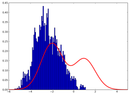
*Influence de l'étalement*

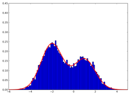
- Taux d'acceptation
- Région couverte par la chaîne de Markov

$\sigma = 1$

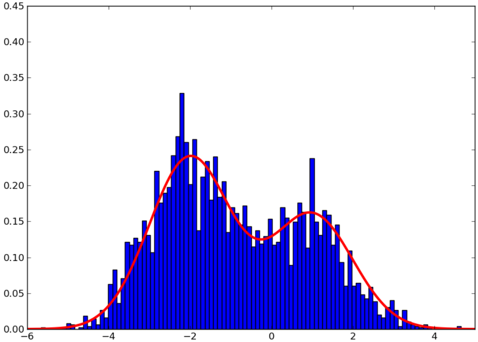


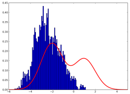





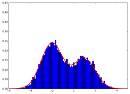


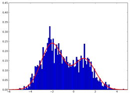
$\sigma = 10$



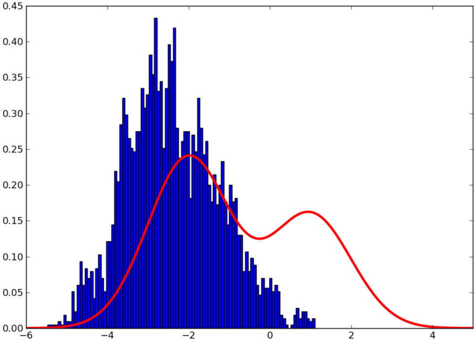



# Choix de l'étalement/variance de $P(\cdot|x_t)$





$\sigma = 0,1$



# Choix de l'étalement/variance de $P(\cdot|x_t)$

*Roberts, Gelman, Gilks (1994)*

- cadre : random walk
- $\overset{\infty}{\pi}$ et $P(\cdot|\cdot)$ : lois normales mono-dimensionnelles

  affiner l'étalement de $P(\cdot|x_t)$ pour obtenir un taux d'acceptation $\approx 0,45$
- $\overset{\infty}{\pi}$ et $P(\cdot|\cdot)$ : lois normales $n$-dimensionnelles

  affiner l'étalement de $P(\cdot|x_t)$ pour obtenir un taux d'acceptation $\approx 0,23$ lorsque $n$ tend vers $+\infty$

*Müller (1993)*

Random walk $\Longrightarrow$ taux d'acceptation $\approx$ 0,5.

# Initialisation et burn in

**Initialisation :**

Partir de n'importe quelle valeur $x_0$

# Initialisation et burn in

**Initialisation :**

Partir de n'importe quelle valeur $x_0$

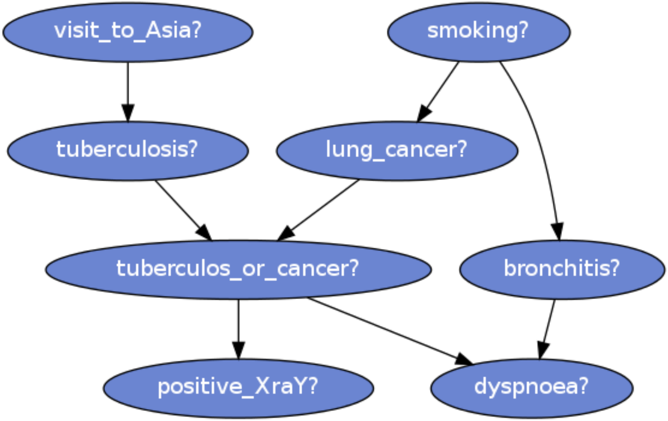
au départ l'échantillon ne suit pas $\overset{\infty}{\pi}(\cdot)$

$\Longrightarrow$ burn in nécessaire :

**Ne conserver dans l'échantillon que les $x_t$ pour $t > t_0$**

En général, $t_0$ est de l'ordre de quelques milliers

# Metropolis-Hastings et les réseaux bayésiens ?




$x_t$ = vecteur à 8 variables !

# Metropolis-Hastings par bloc

- supposons que $x_t = (x_t^1, x_t^2)$

**Précédemment :**

- stationnarité : $\overset{\infty}{\pi}(x_{t+1}) = \int_{x_t} P(x_{t+1}|x_t) \overset{\infty}{\pi}(x_t) dx_t$

# Metropolis-Hastings par bloc

- supposons que $x_t = (x_t^1, x_t^2)$

**Précédemment :**

- stationnarité : $\overset{\infty}{\pi}(x_{t+1}) = \int_{x_t} P(x_{t+1}|x_t) \overset{\infty}{\pi}(x_t) dx_t$

**Maintenant :**

- $P(x_{t+1}|x_t) = P(x_{t+1}^1, x_{t+1}^2 | x_t^1, x_t^2)$
- stationnarité :
$$\overset{\infty}{\pi}(x_{t+1}^1, x_{t+1}^2) = \int_{x_t^1} \int_{x_t^2} P(x_{t+1}^1, x_{t+1}^2 | x_t^1, x_t^2) \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$

- supposons que $x_t = (x_t^1, x_t^2)$

**Précédemment :**

- stationnarité : $\overset{\infty}{\pi}(x_{t+1}) = \int_{x_t} P(x_{t+1}|x_t)\, \overset{\infty}{\pi}(x_t) dx_t$

**Maintenant :**

- $P(x_{t+1}|x_t) = P(x_{t+1}^1, x_{t+1}^2 | x_t^1, x_t^2)$
- stationnarité :
$$\overset{\infty}{\pi}(x_{t+1}^1, x_{t+1}^2) = \int_{x_t^1} \int_{x_t^2} P(x_{t+1}^1, x_{t+1}^2 | x_t^1, x_t^2)\, \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$
- Or $P(x_{t+1}|x_t) = P(x_{t+1}^2 | x_{t+1}^1, x_t^1, x_t^2) \times P(x_{t+1}^1 | x_t^1, x_t^2)$

# Metropolis-Hastings par bloc

- supposons que $x_t = (x_t^1, x_t^2)$

**Précédemment :**

- stationnarité : $\overset{\infty}{\pi}(x_{t+1}) = \int_{x_t} P(x_{t+1}|x_t)\, \overset{\infty}{\pi}(x_t) dx_t$

**Maintenant :**

- $P(x_{t+1}|x_t) = P(x_{t+1}^1, x_{t+1}^2 | x_t^1, x_t^2)$
- stationnarité :
$$\overset{\infty}{\pi}(x_{t+1}^1, x_{t+1}^2) = \int_{x_t^1} \int_{x_t^2} P(x_{t+1}^1, x_{t+1}^2 | x_t^1, x_t^2)\, \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$
- Or $P(x_{t+1}|x_t) = P(x_{t+1}^2 | x_{t+1}^1, x_t^1, x_t^2) \times P(x_{t+1}^1 | x_t^1, x_t^2)$
$$= P(x_{t+1}^2 | x_{t+1}^1, x_t^2) \times P(x_{t+1}^1 | x_t^1, x_t^2) \quad \text{(prop. Markov)}$$

# Metropolis-Hastings par bloc : stationnarité

$$\overset{\infty}{\pi}(x_{t+1}^2, x_{t+1}^2) = \int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2) \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$

*Rappel : stationnarité pour 1 variable*

$$\overset{\infty}{\pi}(x_{t+1}) = \int_{x_t} P(x_{t+1}|x_t) \overset{\infty}{\pi}(x_t) dx_t$$

*Stationnarité par bloc*

Généralisation en rajoutant toutes les variables sauf celle en $x_{t+1}^i$ à droite des signes de conditionnement :

- $\overset{\infty}{\pi}(x_{t+1}^1|y^2) = \int_{x_t^1} P(x_{t+1}^1|x_t^1, y^2) \overset{\infty}{\pi}(x_t^1|y^2) dx_t^1$ pour tout $y^2$
- $\overset{\infty}{\pi}(x_{t+1}^2|y^1) = \int_{x_t^2} P(x_{t+1}^2|x_t^2, y^1) \overset{\infty}{\pi}(x_t^2|y^1) dx_t^2$ pour tout $y^1$

# Metropolis-Hastings par bloc : stationnarité

conséquences de la stationnarité par bloc :

$$\int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$

**conséquences de la stationnarité par bloc :**

$$\int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2) dx_t^1 dx_t^2$$

conséquences de la stationnarité par bloc :

$$\int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\overset{\infty}{\pi}(x_t^1, x_t^2)dx_t^1 dx_t^2$$

$$= \int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\overset{\infty}{\pi}(x_t^1|x_t^2)\overset{\infty}{\pi}(x_t^2)dx_t^1 dx_t^2$$

$$= \int_{x_t^2}\int_{x_t^1} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\overset{\infty}{\pi}(x_t^1|x_t^2)\overset{\infty}{\pi}(x_t^2)dx_t^1 dx_t^2$$

# Metropolis-Hastings par bloc : stationnarité

conséquences de la stationnarité par bloc :

$$\int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1, x_t^2)dx_t^1 dx_t^2$$

$$= \int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2)dx_t^1 dx_t^2$$

$$= \int_{x_t^2}\int_{x_t^1} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2)dx_t^1 dx_t^2$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2)\left[\int_{x_t^1} P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2)dx_t^1\right]\overset{\infty}{\pi}(x_t^2)dx_t^2$$

# Metropolis-Hastings par bloc : stationnarité

conséquences de la stationnarité par bloc :

$$\int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^2}\int_{x_t^1} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \left[\int_{x_t^1} P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2) dx_t^1\right] \overset{\infty}{\pi}(x_t^2) dx_t^2$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2)\ \overset{\infty}{\pi}(x_{t+1}^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2) dx_t^2 \qquad \text{(stationnarité par bloc)}$$

# Metropolis-Hastings par bloc : stationnarité

**conséquences de la stationnarité par bloc :**

$$\int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^2}\int_{x_t^1} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \left[\int_{x_t^1} P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2) dx_t^1\right] \overset{\infty}{\pi}(x_t^2) dx_t^2$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2)\ \overset{\infty}{\pi}(x_{t+1}^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2) dx_t^2 \qquad \text{(stationnarité par bloc)}$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2)\ \overset{\infty}{\pi}(x_t^2|x_{t+1}^1)\ \overset{\infty}{\pi}(x_{t+1}^1) dx_t^2 \qquad \text{(formule de Bayes)}$$

conséquences de la stationnarité par bloc :

$$\int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1, x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^1}\int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^2}\int_{x_t^1} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \times P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2) dx_t^1 dx_t^2$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2) \left[\int_{x_t^1} P(x_{t+1}^1|x_t^1, x_t^2)\ \overset{\infty}{\pi}(x_t^1|x_t^2) dx_t^1\right] \overset{\infty}{\pi}(x_t^2) dx_t^2$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2)\ \overset{\infty}{\pi}(x_{t+1}^1|x_t^2)\ \overset{\infty}{\pi}(x_t^2) dx_t^2 \qquad \text{(stationnarité par bloc)}$$

$$= \int_{x_t^2} P(x_{t+1}^2|x_{t+1}^1, x_t^2)\ \overset{\infty}{\pi}(x_t^2|x_{t+1}^1)\ \overset{\infty}{\pi}(x_{t+1}^1) dx_t^2 \qquad \text{(formule de Bayes)}$$

$$= \overset{\infty}{\pi}(x_{t+1}^2|x_{t+1}^1)\ \overset{\infty}{\pi}(x_{t+1}^1) = \overset{\infty}{\pi}(x_{t+1}^1, x_{t+1}^2) \qquad \text{(stationnarité par bloc)}$$

# Metropolis-Hastings par bloc

*Conclusion du transparent précedent*

Stationnarité par bloc $\Longrightarrow$ Stationnarité de la loi jointe

# Metropolis-Hastings par bloc

*Conclusion du transparent précedent*

Stationnarité par bloc $\Longrightarrow$ Stationnarité de la loi jointe

*Metropolis-Hastings par bloc*

Algorithme pour générer $x_{t+1} = (x_{t+1}^1, \dots, x_{t+1}^n)$ à partir de $x_t$ :

1. choisir une permutation $\sigma : \{1, \dots, n\} \mapsto \{1, \dots, n\}$
2. pour tout $i \in \{1, \dots, n\}$ faire :
   a. Posons $y = (x_{t+1}^{\sigma(1)}, \dots, x_{t+1}^{\sigma(i-1)}, x_t^{\sigma(i+1)}, \dots, x_t^{\sigma(n)})$
   b. tirer $z^{\sigma(i)}$ selon la distribution $P(\cdot | x_t^{\sigma(i)}, y)$
   c. calculer $\alpha(x_t^{\sigma(i)}, z^{\sigma(i)} | y) = \min \left\{ 1, \dfrac{\overset{\infty}{\pi}(z^{\sigma(i)}|y) P(x_t^{\sigma(i)} | z^{\sigma(i)}, y)}{\overset{\infty}{\pi}(x_t^{\sigma(i)}|y) P(z^{\sigma(i)} | x_t^{\sigma(i)}, y)} \right\}$
   d. tirer un nombre $u$ selon une loi uniforme sur $[0, 1[$
   e. $x_{t+1}^{\sigma(i)} = \begin{cases} z^{\sigma(i)} & \text{si } u \leq \alpha(x_t^{\sigma(i)}, z^{\sigma(i)} | y) \\ x_t^{\sigma(i)} & \text{sinon} \end{cases}$

### *Échantillonneur de Gibbs*

- Metropolis-Hastings par bloc
- Choix de la proba de transition : $P(z^{\sigma(i)}|x_t^{\sigma(i)}, y) =\overset{\infty}{\pi}(z^{\sigma(i)}|y)$

*Échantillonneur de Gibbs*

- Metropolis-Hastings par bloc
- Choix de la proba de transition : $P(z^{\sigma(i)}|x_t^{\sigma(i)}, y) = \overset{\infty}{\pi}(z^{\sigma(i)}|y)$

Conséquence :

$$\alpha(x_t^{\sigma(i)}, z^{\sigma(i)}|y) = \min\left\{1, \frac{\overset{\infty}{\pi}(z^{\sigma(i)}|y)P(x_t^{\sigma(i)}|z^{\sigma(i)}, y)}{\overset{\infty}{\pi}(x_t^{\sigma(i)}|y)P(z^{\sigma(i)}|x_t^{\sigma(i)}, y)}\right\} = 1$$

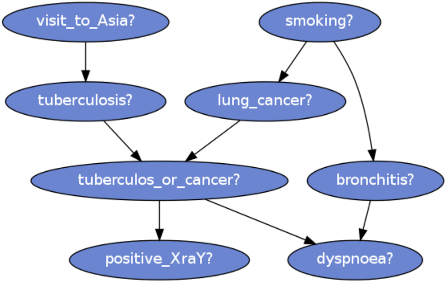
$\Longrightarrow z^{\sigma(i)}$ est toujours accepté

### Algorithme

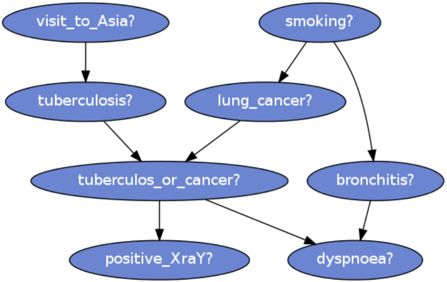
Algorithme pour générer $x_{t+1} = (x^1_{t+1}, \ldots, x^n_{t+1})$ à partir de $x_t$ :

1. choisir une permutation $\sigma : \{1, \ldots, n\} \mapsto \{1, \ldots, n\}$
2. pour tout $i \in \{1, \ldots, n\}$ faire :
   a. posons $y = (x^{\sigma(1)}_{t+1}, \ldots, x^{\sigma(i-1)}_{t+1}, x^{\sigma(i+1)}_t, \ldots, x^{\sigma(n)}_t)$
   b. tirer $x^{\sigma(i)}_{t+1}$ selon la distribution $\overset{\infty}{\pi}(\cdot|y)$

# Asia et l'échantillonneur de Gibbs

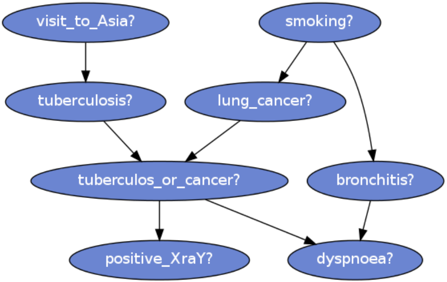


# Asia et l'échantillonneur de Gibbs

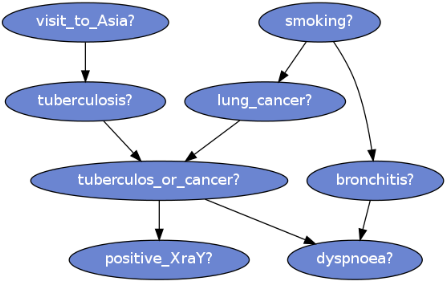



- $\overset{\infty}{\pi}(\cdot)$ connu : c'est la distribution jointe du réseau bayésien

# Asia et l’échantillonneur de Gibbs




- $\overset{\infty}{\pi}(\cdot)$ connu : c’est la distribution jointe du réseau bayésien
- nouvelle valeur de $B$ = “bronchitis” ?



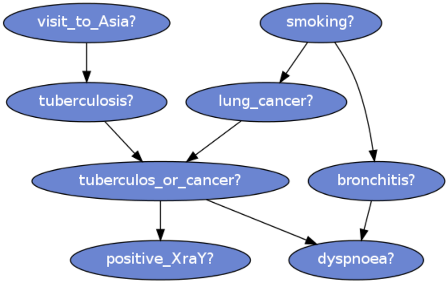

- $\overset{\infty}{\pi}(\cdot)$ connu : c'est la distribution jointe du réseau bayésien
- nouvelle valeur de $B$ = "bronchitis" ?

$\Longrightarrow$ échantillonner selon $P(B|y)$, avec $y$ les valeurs de toutes les autres variables

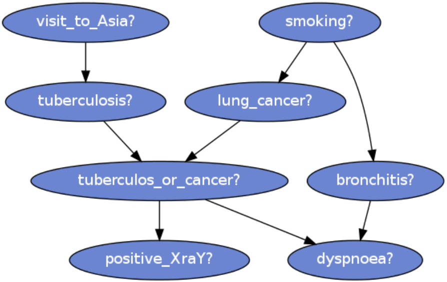
- Si on note par les initiales minuscules les valeurs observées :

$$P(B|y) = P(B|vta, s, t, lc, toc, px, d)$$

# Asia et l'échantillonneur de Gibbs

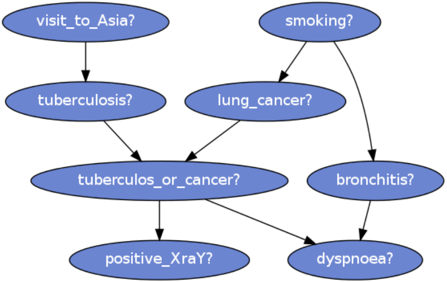


# Asia et l'échantillonneur de Gibbs




$$P(B|y) = \frac{P(B, y)}{P(y)} \propto P(B, y) = \text{vecteur de taille } |\text{bronchitis}|$$

$$P(B, y) = P(B, vta, s, t, lc, toc, px, d)$$

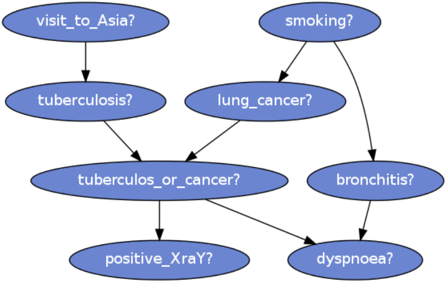



$$P(B|y) = \frac{P(B,y)}{P(y)} \propto P(B,y) = \text{vecteur de taille } |\text{bronchitis}|$$

$$P(B,y) = P(B, vta, s, t, lc, toc, px, d)$$

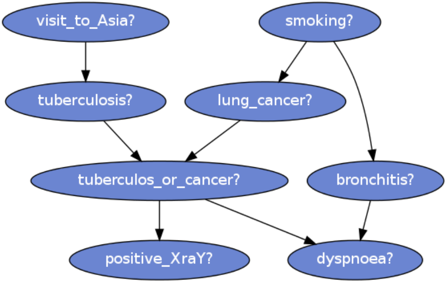
$$= P(vta)P(s)P(t|s)P(lc|s)P(toc|t,lc)P(B|s)P(px|toc)P(d|toc,B)$$

# Asia et l'échantillonneur de Gibbs




$$P(B|y) = \frac{P(B,y)}{P(y)} \propto P(B,y) = \text{vecteur de taille } |\text{bronchitis}|$$

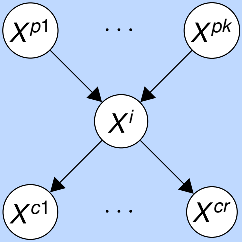
$$\begin{aligned} P(B,y) &= P(B, vta, s, t, lc, toc, px, d) \\ &= \color{red}{P(vta)P(s)P(t|s)P(lc|s)P(toc|t,lc)}\color{black}{P(B|s)}\color{red}{P(px|toc)}\color{black}{P(d|toc,B)} \end{aligned}$$



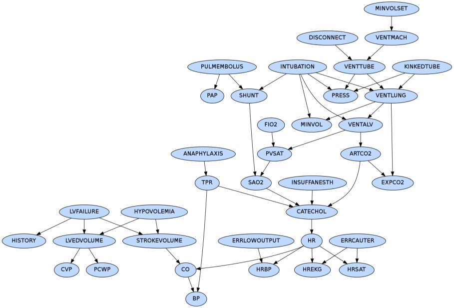

$$P(B|y) = \frac{P(B,y)}{P(y)} \propto P(B,y) = \text{vecteur de taille } |\text{bronchitis}|$$

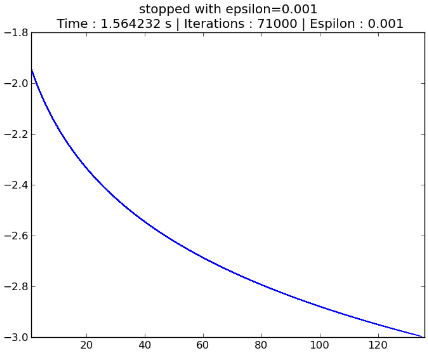
$$\begin{aligned} P(B,y) &= P(B, vta, s, t, lc, toc, px, d) \\ &= P(vta)P(s)P(t|s)P(lc|s)P(toc|t,lc)P(B|s)P(px|toc)P(d|toc,B) \\ &\propto P(B|s)P(d|toc,B) \end{aligned}$$

*Règle pour déterminer la valeur de $x^i_{t+1}$*




1. extraire le vecteur $V = P(X^i | x^{p1}, \dots, x^{pk})$
2. extraire les vecteurs $V_i = P(x^{ci} | pa(X^{ci}), X^i)$
   où $pa(X^{ci})$ = valeur des parents de $X^{ci}$ sauf $X^i$
3. calculer la ditribution $\Pi = V \times \prod_{i=1}^{r} V_i$ (produits tensoriels : terme à terme) et la normaliser
3. tirer $x^i_{t+1}$ selon $\Pi$ (distribution discrète)




- Ordonnées : $\log_{10}\left(\sum_i \text{distance Kullback Leibler}(P(X_t^i), P(X_{t+1}^i))\right)$

$$-3 \Longrightarrow \sum_i \text{dist KL} = 10^{-3}$$