Chaine de Markov :

Une chaîne de Markov est un processus $\{X_t : t \in \mathbb{N}\}$ tel que :

$$P(X_{t+1} \in A \mid X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) =$$
$$P(X_{t+1} \in A \mid X_t = x_t)$$

On dira que la chaîne est **homogène** si

$$P(X_{t+1} \in A \mid X_t = x) \qquad (1)$$

ne dépend pas de t.

Exmp : La marche aléatoire $\{X_t : t \in \mathbb{N}\}$ définie par
$X_0 \sim \nu$ et $X_{t+1} = X_t + \varepsilon_{t+1}$, $t \in \mathbb{N}$ avec
$\varepsilon_1, \varepsilon_2, \dots \overset{iid}{\sim} N(0, \sigma^2)$.

Def : Un noyau de transition sur $(\Omega, \mathcal{A})$ est une application $K : \Omega \times \mathcal{A} \longrightarrow \mathbb{R}$ telle que

* Pour tout $A \in \mathcal{A}$, $x \longmapsto K(x, A)$ est mesurable
* Pour tout $x \in \Omega$, $A \longmapsto K(x, A)$ est une mesure de probabilité sur $(\Omega, \mathcal{A})$.

Rmq : On dira alors que la chaîne homogène (1) a pour noyau de transition P où
$P(x, A) = P(X_t \in A \mid X_{t-1} = x)$. $x \in \Omega$, $A \in \mathcal{A}$.

Exmp: Pour la marche aléatoire normale, on a

$$P(x, A) = \int_A \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-x)^2}{2\sigma^2}\right\} dy .$$

Rmq: On utilisera souvent l'abus de notation $P(x,y)$ pour $x,y \in \Omega$. De sorte que

$$P(x, A) = \int_A P(x,y)\, dy .$$

Intuition: $P(x,A)$ est la probabilité d'atterir en A sachant que je suis en x.

Notations: Soit $\{X_t : t \in \mathbb{N}\}$ une chaîne de Markov homogène de transition P et loi initiale ν. On notera:

* P_ν la loi de $\{X_k : k \in \mathbb{N}\}$
* νP^k la loi de X_k pour tout $k \in \mathbb{N}$, i.e,

$$\forall A \in \mathcal{A}, \ \nu P^k(A) = P(X_k \in A) = \int \nu(dx_0) P(x_0, dx_1) \dots P(x_{k-1}, A)$$

* $P^k(x, A) = P(X_t \in A \mid X_0 = x)$

Objectif MCMC : Simuler selon une loi de probabilité donnée π.

Moyen : Construire une chaîne de Markov qui va "approcher" π au sens de la variation totale, ie,

$$\|\nu P^t - \pi\|_{VT} := \sup_{A \in \mathcal{A}} |\nu P^t(A) - \pi(A)| \longrightarrow 0, \quad t \to \infty.$$

Def : * Un noyau de transition P est π-invariant si

$$\int P(x, A)\,\pi(dx) = \pi(A), \quad \forall A \in \mathcal{A}$$

* ————— π-irréductible si pour tout $x \in \Omega$ et $A \in \mathcal{A}$ tq $\pi(A) > 0$, $\exists k \in \mathbb{N}$ tel que $P^k(x, A) > 0$.

* ————— est périodique de période $p \geq 2$ s'il existe une partition $\Omega_1, \dots, \Omega_p$ de Ω t.q. pour tout $x \in \Omega_i$, $P(x, \Omega_{i+1[p]}) = 1$, pour tout $i \in \mathbb{N}$.

Rmq : Si la période vaut $p = 1$, on dira que la chaîne est apériodique.

Intuition :

* L'irréductibilité assure que partant de n'importe quel point $x \in \Omega$ on peut arriver n'importe où dans Ω via la transition P – le temps peut être long cela dit...

* L'apériodicité permet éviter la présence de cycles lorsque la chaîne parcours un long trajet. Evite les versions aléatoires de $a_n = (-1)^n$, $n \in \mathbb{N}$.

Thm : Si $\{X_t : t \in \mathbb{N}\}$ est une chaîne irréductible de transition P et de loi invariante π, alors

$$\frac{1}{T} \sum_{t=1}^{T} g(X_t) \longrightarrow \int g(x)\pi(dx), \quad T \to \infty$$

où g est une fonction π-intégrable et ceci est vrai pour tout X_0 π-p.p..

Thm : Si de plus la chaîne est apériodique alors

$$\| P^k(X_0, \cdot) - \pi(\cdot) \|_{TV} \longrightarrow 0, \quad k \to \infty$$

pour tout X_0 π-p.p.

Harris récurrence :

Les conditions pour les 2 thm précédents sont assez faciles à vérifier sachant ~~de~~ la transi° P.

Pbm : Bien plus difficile de s'assurer que X_0 est dans une région où la densité π est > 0.

Def : Une ~~noyau de transition~~ chaine de transi° P est Harris récurrente si pour tout $x \in \Omega$

$$P(X_k \in A \text{ infiniment souvent} \mid X_0 = x) = 1.$$

Thm : Les 2 thm précédents sont vrais pour tout $X_0 \in \Omega$ si la chaine est Harris récurrente.

Soyons concret !!!

Les résultats précédents ne servent à rien si l'on ne sait pas construire une chaine de Markov vérifiant ces conditions.

C'est l'objectif des méthodes MCMC, produire des méthodes génériques pour obtenir de telles chaines de Markov.

Métropolis - Hastings :

L'idée est d'utiliser un noyau de transition π-réversible, i.e,

$$\pi(x)\,P(x,y) = \pi(y)\,P(y,x) \qquad \forall x,y\in\Omega.$$

Rmq: la π-réversibilité est facile à obtenir et implique la π-invariance puisque

$$\pi P(B) = \int_\Omega P(x,B)\,\pi(dx) \overset{\text{inv}}{=} \int_B P(x,\Omega)\,\pi(dx) = \pi(B).$$

Algo: Métropolis - Hastings

In: N longeur chaine, ~~noyau transition~~ noyau de proposition $q(\cdot,\cdot)$ et état initial x_0, loi cible π.

Out: Une chaine de Markov $\{X_t : t\in\mathbb{N}\}$ de loi invariante π.

0) $t \leftarrow 0$

1) Générer candidat $X_* \sim p(x_t,\cdot)$

2) Accepter X_*, i.e, $X_{t+1} \leftarrow X_*$ avec proba

$$\alpha(x_t,x_*) = \min\left\{1,\ \frac{\pi(X_*)\,q(X_*,X_t)}{\pi(X_t)\,q(X_t,X_*)}\right\}$$

sinon rester en X_t, i.e, $X_{t+1} \leftarrow X_t$

3) Revenir en 1 et incrémenter t, i.e, $t \leftarrow t+1$

⚠ Noyau de proposi° ≠ noyau de transi°

En effet le noyau de transi° de la chaîne est

$$P(x, dy) = q(x,y)\,\alpha(x,y) + \left\{1 - \int q(x,t)\,\alpha(x,t)\,dt\right\} \mathbb{1}_{\{x=y\}}$$

Et c'est bien π-réversible comme conséquence de

$$\begin{aligned}\pi(x)\,q(x,y)\,\alpha(x,y) &= \min\left\{\pi(x)\,q(x,y),\ \pi(y)\,q(y,x)\right\} \\ &= \pi(y)\,q(y,x)\,\min\left\{\frac{\pi(x)\,q(x,y)}{\pi(y)\,q(y,x)},\ 1\right\} \\ &= \pi(y)\,q(y,x)\,\alpha(y,x)\end{aligned}$$

Quelques noyaux de transition :

- M.-H. indépendant $q(x,y) = q(y)$ → pas très utile
- marche aléatoire $q(x,y) = q(y-x)$, q selon une loi symétrique
- marche aléatoire sur le log, $\frac{q(y,x)}{q(x,y)} = \frac{y}{x}$ → utile si on se balade sur $\mathbb{R}^+$

Exmp: Faire un M.-H. pour la loi cible $\mathcal{N}(0,1)$ et où le noyau de proposition est celui d'une marche aléatoire avec innovaº gaussiennes.

↳ à voir : impact valeur initiale, impact taille des innovaº

Exmp: Faire un M.-H pour la loi cible $LN(0,1)$ avec une transiº de $\mathcal{N}(0,1)$ et d'une marche aléatoire sur le log.

↳ à voir : impact choix loi de proposition.

Représenter $\{(X_t, X_{t+1}) : t \in \mathbb{N}\}$ pour parler de la dépendance. Introduire l'ACF. Parler de l'amincissement d'une chaîne ("thinning").

Echantillonneur de Gibbs :

Cas particulier de M.-H.

Bonne option pour simuler selon une loi cible multivariée, i.e, vivant dans $\mathbb{R}^d$, $d \geq 2$.

Exmp : Simuler selon une $\mathcal{N}(0, \Sigma)$.

Algo :

$X_0 = (X_{1,0}, \dots, X_{d,0})$

In : Etat initial, N longueur de la chaîne, loi cible π.

Out : Chaîne de Markov de loi stationnaire π.

0) $t \leftarrow 0$

1) Pour $j = 1, \dots, d$ faire

2) $\quad X_{j,t} \sim \pi(\cdot \mid X_{-j,t})$ où $X_{-j,t} = \{X_{i,t} : i \neq j\}$

3) Poser $X_{t+1} \leftarrow X_t$

4) Faire $t \leftarrow t+1$ et revenir à (1).

Rmq : Le taux d'accepta° de candidat est tjs $=$ à 1 !!!

Le noyaux de transition pour la j-ème composante est

$$P_j(x, y) = \pi(y_j \mid x_{-j}) \, \mathbb{1}_{\{x_{-j} = y_{-j}\}}, \quad j = 1, \dots, d.$$

Exmp : Échantillonneur de Gibbs pour la loi cible $\mathcal{N}(0, \Sigma)$ où $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $-1 < \rho < 1$.

Rappel : $$f(x,y) = \frac{1}{(2\pi)^{d/2}} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x\ y)\Sigma^{-1}\binom{x}{y}\right\}$$

$$= \frac{1}{2\pi}(1-\rho^2)^{-1/2} \exp\left(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right)$$

↳ à voir calcul de $f(x|y)$ et de coup $f(y|x)$

Modèles graphiques:

Pourquoi?? Supposer que les v.a. sont très ind. est clairement trop restrictif en pratique.
Essayer de considérer n'importe quelle forme de dépendance est trop ambitieux.

$\Longrightarrow$ compromis

Principe: S'inspirer de l'hypothèse de dépendance Markovienne mais la "mettre" entre les variables.

Idée: Représenter les v.a. au sein d'un graphe pour lequel les noeuds représentent les v.a. et les arêtes la dépendance.

Soit $\mathcal{J} = \{1, \dots, n\}$ un ensemble fini de noeuds représentant les variables aléatoires $Y_1, \dots, Y_n$.

Notation: Pour $A \subset \mathcal{J}$, $Y_A = \{Y_i : i \in A\}$, $Y_{-A} = \{Y_i : i \notin A\}$ et $Y_i = Y_{\{i\}}$ et $Y_{-i} = Y_{-\{i\}}$.

Def: On appelle système de voisinage la "collection" d'ensembles $\mathcal{N} = \{N_j : j \in \mathcal{J}\}$ où les ensembles N_j est le voisinage du noeud j.

Def: N_j est le voisinage du noeud j si $N_j \subset \mathcal{Y}$ t.q.

$$j \notin N_j \quad \text{et} \quad i \in N_j \iff j \in N_i$$

$\Longrightarrow$ on appelera alors $(\mathcal{Y}, N)$ un <u>modèle graphique</u>.

Def: On appelle <u>graphe acyclique orienté</u> un modèle graphique représentant une structure de dépendance hiérarchique.

La dépendance de Y_1 par rapport à Y_2 est alors représentée par une flèche partant du <u>noeud parent</u> Y_2 et pointant vers le <u>noeud descendant</u> Y_1.

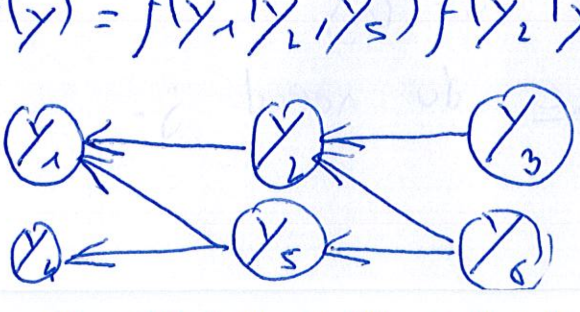
On dira que Y_1 est un descendant de Y_3 s'il y a une "chaine" de flèches liant Y_3 à Y_1.

Rmq: * C'est un graphe orienté car nous avons des flèches et plus des arcs.

* C'est un graphe acyclique car on ne peut pas en suivant les flèches partir de Y_1 et revenir en Y_1.

Exmp: $f(y) = f(y_1|y_2,y_5)\, f(y_2|y_3,y_6)\, f(y_3)\, f(y_4|y_5)\, f(y_5|y_6)\, f(y_6)$

$\Longrightarrow$



Ppté (admise) : Pour tout DAG on a

$$f(y) = \prod_{j \in \gamma} f(y_j \mid \text{parents de } y_j)$$

En conséquence

$$f(y_j \mid y_{-j}) = \frac{f(y)}{\int f(y)\, dy_j} \propto \prod_{i \in \gamma} f(y_i \mid \text{parents de } y_i)$$

$$\propto f(y_j \mid \text{parents de } y_j) \prod_{\{i : y_i \text{ fils de } y_j\}} f(y_i \mid \text{parents de } y_i)$$

Exmp : Ligue 1 2014/2015

$N_{ij}^{(d)}$: nb buts équipe i jouant à domicile contre équipe j.

$N_{ij}^{(e)}$: ——————— l'extérieur ———

Modèle : $N_{ij}^{(d)} \sim \text{Poisson}(\lambda_{ij}^{(d)})$

$N_{ij}^{(e)} \sim \text{Poisson}(\lambda_{ij}^{(e)})$

avec $\lambda_{ij}^{(d)} = \underbrace{\beta_0^{(d)}}_{\text{effet domicile}} + \underbrace{\beta_i}_{\text{perf. éq. i}} - \underbrace{\beta_j}_{\text{perf. éq. j}}$

$\lambda_{ij}^{(e)} = \beta_0^{(e)} + \beta_j - \beta_i$

$\beta_1, \dots, \beta_{20} \overset{iid}{\sim} N(\underbrace{\mu}_{\text{perf. moy. ligue 1}}, \underbrace{\sigma^2}_{\text{variabilité ligue 1}})$

Loi a priori

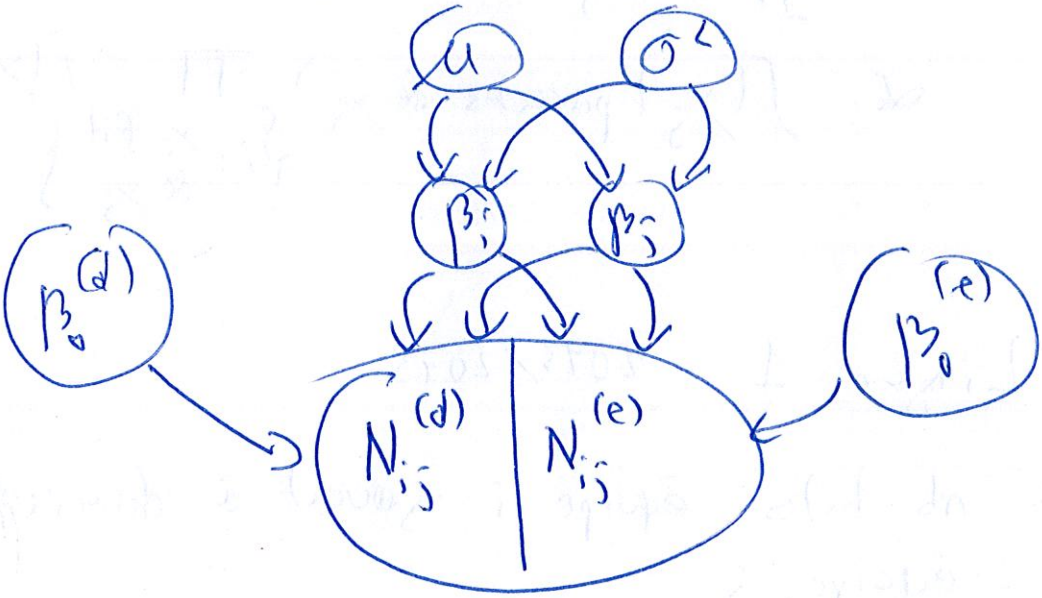
$\beta_0^{(d)}, \beta_0^{(e)} \sim N(0,16) \leftarrow$ très plat

$\mu \sim N(0,16) \leftarrow$ très plat.

$\sigma^2 \sim \mathcal{U}(0,9) \leftarrow$

Inversé !

On a donc le DAG suivant



$$\pi(\beta_i \mid \ldots) \propto \prod_{j\neq i} \pi(N_{ij}^{(d)} \mid \beta_0^{(d)}, \beta_i, \beta_j) \times \prod_{j\neq i} \pi(N_{ij}^{(e)} \mid \beta_0^{(e)}, \beta_i, \beta_j) \times \pi(\beta_i \mid \mu, \sigma^2)$$

$$\pi(\beta_0^{(d)} \mid \ldots) \propto \prod_{i\neq j} \pi(N_{ij}^{(d)} \mid \beta_i, \beta_j, \beta_0^{(d)})\, \pi(\beta_0^{(d)})$$

$$\pi(\beta_0^{(e)} \mid \ldots) \propto \prod_{i\neq j} \pi(N_{ij}^{(e)} \mid \beta_i, \beta_j, \beta_0^{(e)})\, \pi(\beta_0^{(e)})$$

$$\pi(\mu \mid \ldots) \propto \prod_{i=1}^{20} \pi(\beta_i \mid \mu, \sigma^2)\, \pi(\mu)$$

$$\pi(\sigma^2 \mid \ldots) \propto \prod_{i=1}^{20} \pi(\beta_i \mid \mu, \sigma^2)\, \pi(\sigma^2)$$