NOTES DE COURS DU M2 ISN :
CHAÎNES DE MARKOV ET APPLICATIONS

ENSEIGNANT : ADRIEN HARDY

SCRIBE : BENJAMIN KASPRZAK

1 Quelques rappels sur les chaînes de Markov

1.1 Notations habituelles

Soient $(E, \mathcal{T})$ un espace mesurable discret. $(X_n)_{n\in\mathbb{N}} : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{T})^{\otimes\mathbb{N}}$ mesurable est une chaîne de Markov homogène si elle est caractérisée par la loi μ de X_0 et une "matrice" de transition $P = [P_{x,y}]_{x,y\in E}$ telle que :

$$\forall n \geq 1,\ \forall x_0, \ldots, x_n \in E,\ \mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) = \mu_{x_0} P_{x_0,x_1} \ldots P_{x_{n-1},x_n}$$

$$\begin{aligned}\mu \in M_1(E) &= \{\text{mesures de proba sur } E\} \\ &= \{(\mu_x)_{x\in E}\,,\ \mu_x \geq 0 \text{ et } \sum_{x\in E} \mu_x = 1\}\end{aligned}$$

On identifie μ à un vecteur ligne $(\ldots \mu_x \ldots)_{x\in E}$.
De plus, $g : E \to \mathbb{R}$ bornée, on l'identifie à un vecteur colonne $(\ldots g(x) \ldots)^T_{x\in E}$. On a alors :

$$\mathbb{P}(X_n = x) = (\mu P^n)_x$$

$$\mathbb{E}[g(X_n)] = \mu P^n g \in \mathbb{R}$$

Remarque : .

$$\begin{aligned}\mathbb{P}(X_n \in A | X_0, \ldots, X_{n-1}) &= \mathbb{P}(X_n \in A | X_{n-1}) \\ P_{x,y} &= \mathbb{P}(X_n = y | X_{n-1} = x)\end{aligned}$$

1.2 Mesures invariantes

Définition : .
On dit que π est une mesure invariante pour P si, et seulement si, $\pi P = \pi$.
Si π est une probabilité invariante, alors $X_0 \simeq \pi \Rightarrow X_n \simeq \pi,\ \forall n \geq 1$.
On dit que la chaîne est irréductible si :

$$\forall x, y \in E,\ \exists m, n \geq 0,\ (P^n)_{x,y} > 0 \text{ et } (P^m)_{y,x} > 0$$

On dit que la chaîne est réversible par rapport à une mesure μ si :

$$\mu_x P_{x,y} = \mu_y P_{y,x} \quad \forall x, y \in E$$

On dit que la chaîne est récurrente positive si :

$$\mathbb{E}_x[\tau_x] < \infty \quad \forall x \in E$$

Théorème : .

Si E est fini, alors P a toujours une probabilité invariante, qui sera unique si et seulement si la chaîne est de Markov associée est irréductible.

Remarque : .

La marche aléatoire symétrique sur $\mathbb{Z}$ n'a pas de probabilité invariante.

Proposition : .

Si P est réversible par rapport à une mesure μ sur E, alors μ est invariante.

Preuve :

Supposons que P soit réversible. On a :

$$\begin{aligned}
(\mu P)_x &= \sum_{y \in E} \mu_y P_{y,x} \\
&= \underbrace{\left(\sum_{y \in E} \mu_x P_{x,y} \right)}_{=1} \mu_x \text{ par réversibilité} \\
&= \mu_x \quad \forall x \in E
\end{aligned}$$

D'où le résultat attendu. ⋆

Théorème : .

Si la chaîne est irréductible et récurrente positive, alors on a une unique probabilité invariante π donnée par :

$$\pi_x = \frac{1}{\mathbb{E}_x(\tau_x)} \quad tq \quad \tau_x = \inf\{n \geq 1 \quad tq \quad X_n = x\}$$

On note $\mathbb{P}_\mu$ et $\mathbb{E}_\mu$ pour souligner la dépendance en la loi μ de X_0 et $\mathbb{P}_x := \mathbb{P}_{\delta_x}$ et $\mathbb{E}_x := \mathbb{E}_{\delta_x}$.

Remarque : .

Si la chaîne est irréductible et admet une probabilité invariante, alors elle est récurrente positive.

Calcul de la mesure invariante ?

Supposons que E fini, $\pi = (\dots \pi_x \dots)_{x \in E}$. On doit résoudre le système linéaire suivant :

$$\begin{cases} \pi P &= \pi, \quad \text{tq} \quad \pi_x \geq 0 \\ \pi \mathbf{1} &= 1 \end{cases} \iff \begin{cases} P^T \pi^T &= \; = \pi^T \Leftrightarrow A\pi^T = b \\ \mathbf{1}^T \pi^T &= \; = 1 \end{cases} \quad \text{tq}$$

$$A = \begin{pmatrix} & P^T - I & \\ 1 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{d+1 \times d} \quad ; \quad b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^{d+1}$$

Théorème : . *(de Rouché-Capelli)*

Le système $Ax = b$ admet au moins une solution si, et seulement si, $rang(A) = rang(A|b)$.

Dans ce cas, l'espace des solutions est de dimension $d - rang(A)$.

Preuve :

Admise. ⋆

Ici, comme E est fini, si on suppose que la chaîne est irréductible, on sait que $Ax = b$ a une unique solution π^T. Ainsi, $rang(A) = d$, les colonnes de A forment donc une famille libre et on conclut que $A^T A$ est inversible. Ainsi :

$$\begin{aligned} Ax = b &\Leftrightarrow A^T Ax = A^T b \\ &\Leftrightarrow x = \left(A^T A\right)^{-1} A^T b \end{aligned}$$

En conclusion, on peut résoudre le système de départ et obtenir $\pi = x^T$.

Remarque : .

Si $|E|$ est grand, cette méthode est difficile numériquement.

Théorème : . *(ergodique)*

Si $(X_n)_{n\in\mathbb{N}}$ est irréductible et récurrente positive, alors :

$$\forall f : E \to \mathbb{R} \text{ bornée : } \frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow[n\to+\infty]{ps} \int f d\pi$$

Preuve :

Admise. ⋆

Exemple : .

Si $f(y) = \mathbf{1}_{y=x}$, alors pour tout $x \in E$:

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i=x} \xrightarrow[n\to+\infty]{ps} \pi_x$$

En passant à l'espérance $\mathbb{E}_\mu$, pour $\mu \in M_1(E)$ quelconque, on obtient :

$$\frac{1}{n}\sum_{i=1}^{n} \left(\mu P^i\right)_x \xrightarrow[n\to+\infty]{} \pi_x$$

Définition : .

La période d'un état $x \in E$ est :

$$PGCD\left(\{n \geq 1 \text{ tq } P^n_{x,x} > 0\}\right)$$

Dans le cas où P est irréductible, la période est bien définie et ne dépend pas de x.

On dit que P est apériodique si la période de tous ses états est 1.

Théorème : .

Si P est irréductible récurrente positive et apériodique, alors :

$$\forall \mu \in M_1(E),\ \left(\mu P^n\right)_x \underset{n\to+\infty}{\longrightarrow} \pi_x$$

Preuve :

Admise. ⋆

Définition : . (Condition de Doeblin)

On dit que P satisfait $D(\beta, n_0)$, où $0 < \beta < 1$ et $n_0 \geq 1$ si :

$$\exists \nu \in M_1(E) \text{ tq } \forall x, y \in E,\ \left(P^{n_0}\right)_{x,y} \geq \beta \nu_y$$

Remarque : *.*

Si P^{n_0} a une colonne j strictement positive, alors P satisfait $D(\beta, n_0)$ où β est l'inf sur la j-ème colonne et $\nu = \mathbf{1}_j$.

Théorème : .

Si P satisfait la condition de Doeblin $D(\beta, n_0)$, alors P est irréductible, récurrente positive, apériodique avec :

$$\forall n \geq 1,\ \forall \mu \in M_1(E),\ \sum_{x\in E} \left|\left(\mu P^n\right)_x - \pi_x\right| \leq 2(1-\beta)^{\lfloor \frac{n}{n_0} \rfloor}$$

Preuve : *(Cas $n_0 = 1$)*

Par hypothèse, on peut écrire $P_{x,y} = \beta\nu_y + (1-\beta) S_{x,y}$ avec $S_{x,y} = \frac{1}{1-\beta}\left(P_{x,y} - \beta\nu_y\right)$ est une matrice de transition Markovienne. Ainsi, $\forall \mu^1, \mu^2 \in M_1(E)$, on a :

$$\begin{aligned}
\left(\left(\mu^1 - \mu^2\right) P\right)_y &= \sum_{x\in E} \left(\mu^1_x - \mu^2_x\right) P_{x,y} \\
&= \sum_{x\in E} \left(\mu^1_x - \mu^2_x\right) \beta\nu_y + \sum_{x\in E} \left(\mu^1_x - \mu^2_x\right)(1-\beta) S_{x,y} \\
&= \beta\nu_y \underbrace{\sum_{x\in E} \left(\mu^1_x - \mu^2_x\right)}_{=0} + \sum_{x\in E} \left(\mu^1_x - \mu^2_x\right)(1-\beta) S_{x,y} \\
&= \sum_{x\in E} \left(\mu^1_x - \mu^2_x\right)(1-\beta) S_{x,y}
\end{aligned}$$

$$\begin{aligned}
\|\left(\mu^1 - \mu^2\right) P\|_{l^1} &= \sum_{y\in E} \left|\left(\left(\mu^1 - \mu^2\right) P\right)_y\right| \\
&\leq (1-\beta) \sum_{x\in E} \left|\mu^1_x - \mu^2_x\right| \underbrace{\sum_{y\in E} S_{x,y}}_{=1} \quad \textit{par Fubini-Tonelli pour la permutation} \\
&\leq (1-\beta) \sum_{x\in E} \left|\mu^1_x - \mu^2_x\right|
\end{aligned}$$

Donc $F : \mu \to \mu P$ est une $(1-\beta)$-contraction de $l^1(E)$ (Banach).

Par le théorème du point fixe de Banach, on déduit que :

- *Il existe un unique point fixe π ;*
- $||F^{(n)}\mu - \pi||_{l^1} \leq \frac{(1-\beta)^n}{\beta} \underbrace{||F\mu - \mu||}_{\leq 2} \quad \forall \mu \in M_1(E)$.

Ainsi, on $||\mu P^n - \pi||_{l^1} \leq 2\frac{(1-\beta)^n}{\beta}$. On obtient un résultat plus faible que ce qui était annoncé. ⋆

Remarque : .

$$M_1(E) \subset l^1(E) = \left\{ (\mu_x)_{x\in E} \ \text{ tq } \ ||\mu||_{l^1} = \sum_{x\in E} |\mu_x| < \infty \right\} \text{ complet}$$

Exemple : . (Page Rank)

On définit : $P = sG_{Markov} + (1-s)\frac{\mathbf{1}}{N}$. On obtient :

$$||\mu P^n - \pi|| \leq 2s^n, \ n \geq 1$$

- $\underline{s = 1 :}$ $P = G_{Markov}$, π non unique, pas de garantie de convergence ;
- $\underline{s = 0 :}$ $P = \frac{\mathbf{1}}{N}$, $\pi = \left(\frac{1}{N}, \dots, \frac{1}{N}\right)^T$, et :

$$(\mu P)_j = \sum_{i=1}^{N} \mu_i P_{i,j} = \frac{1}{N}$$

L'algorithme de Google propose $s = 0.85$.

2 Simulation de mesures de Gibbs

2.1 Définitions et Motivations

Définition : .

Soient E un ensemble fini, $H : E \to \mathbb{R}$ "énergie" ou "Hamiltonien", $\beta > 0$ la température inverse.

On définit la mesure de Gibbs comme la probabilité sur E :

$$\mu_\beta(x) = \frac{e^{-\beta H(x)}}{Z_\beta} \ \text{ tq } \ Z_\beta = \sum_{x\in E} e^{-\beta H(x)} \text{ fonction de partition}$$

Définition : .

L'entropie de Shannon d'une mesure μ est la quantité :

$$Ent(\mu) = -\sum_{x\in E} \mu(x) \log \mu(x) \geq 0$$

Motivation physique : (Principe d'entropie maximale)

Si on veut modéliser un système complexe, d'énergie moyenne H fixée, par une distribution de probabilité, il faut prendre celle d'entropie maximale.

Exemple : .

Pour $0 \leq p \leq 1$, considérons μ_p sur $\{0, 1\}$ telle que : $\mu_p = p\delta_1 + (1-p)\delta_0$.

On trouve que $Ent\left(\mu_p\right) = -p\log(p) - (1-p)\log(1-p)$.

Cette fonction est bien positive, nulle en 0 et 1 par prolongement par continuité, et maximale en $p = \frac{1}{2}$.

Proposition : .

Si $\mu \in M_1(E)$ avec $\mathbb{E}_\mu[H] = \mathbb{E}_{\mu_\beta}[H]$ pour un $\beta > 0$, alors $Ent\left(\mu_\beta\right) \geq Ent(\mu)$, avec égalité pour $\mu = \mu_\beta$.

Preuve :

On fait le calcul :

$$
\begin{aligned}
Ent\left(\mu_\beta\right) &= -\sum_{x\in E} \mu_\beta(x)\log\mu_\beta(x) \\
&= \sum_{x\in E} \mu_\beta(x)\left(\beta H(x) + \log(Z_\beta)\right) \\
&= \sum_{x\in E} \mu(x)\left(\beta H(x) + \log(Z_\beta)\right) \\
&= -\sum_{x\in E} \mu(x)\log\mu_\beta(x)
\end{aligned}
$$

$$
\begin{aligned}
\Longrightarrow Ent\left(\mu_\beta\right) - Ent(\mu) &= -\sum_{x\in E} \mu(x)\log\mu_\beta(x) + \mu(x)\log\mu(x) \\
&= \sum_{x\in E} \frac{\mu(x)}{\mu_\beta(x)}\log\left(\frac{\mu(x)}{\mu_\beta(x)} \times \mu_\beta(x)\right) \\
&= \mathbb{E}_{\mu_\beta}[Y\log(Y)] \qquad \text{tq} \;\; Y(x) = \frac{\mu(x)}{\mu_{\beta(x)}} \\
&\geq E_{\mu_\beta}[Y]\log\mathbb{E}_{\mu_\beta}[Y] \quad \text{par Jensen car } x \mapsto x\log(x) \text{ strictement convexe} \\
&\geq \left(\sum_{x\in E}\mu(x)\right)\log\left(\sum_{x\in E}\mu(x)\right) \\
&\geq 0 \quad \text{car} \sum_{x\in E}\mu(x) = 1
\end{aligned}
$$

Et on a égalité si, et seulement si, $Y = 1$, c'est à dire $\mu = \mu_\beta$.

D'où le résultat attendu. ⋆

Remarque : .

Si m fixé et $\mathbb{E}_\mu[H] = m$, alors il existe $\beta = \beta_m$ tq $\mathbb{E}_{\mu_\beta}[H] = m$.

Remarque : .

Le cas $\beta = 0$ implique que μ_β est la mesure uniforme sur E.

Et le cas $\beta \to \infty$?

Posons $h = \min\limits_{x\in E} H(x)$, et $\mathcal{M} = \{x \in E \ \text{ tq } \ H(x) = h\}$.

$$\begin{aligned}
\mu_\beta(x) &= \frac{e^{-\beta H(x)}}{\sum\limits_{x\in E} e^{-\beta H(x)}} \\
&= \frac{e^{-\beta(H(x)-h)}}{\sum\limits_{x\in E} e^{-\beta(H(x)-h)}} \\
&= \frac{e^{-\beta(H(x)-h)}}{|\mathcal{M}| + \sum\limits_{\substack{x\in E\\ x\notin\mathcal{M}}} e^{-\beta\left(\underbrace{H(x)-h}_{>0}\right)}} \\
&\underset{\beta\to\infty}{\longrightarrow} \frac{1}{|\mathcal{M}|}\mathbf{1}_{x\in\mathcal{M}}
\end{aligned}$$

En particulier, $\mathbb{E}_{\mu_\beta}[H] \underset{\beta\to\infty}{\longrightarrow} \min\limits_{x\in E} H(x)$.

Exemple : . (Modèle d'Ising)

Il s'agit d'un toy-model pour les transitions de phases ferromagnétique. On définit :

- $\Lambda_N = \{1,\dots,N\}^2 \ \text{ tq } \ x \in \Lambda_N \to \sigma_x \in \{\pm 1\}$;
- $E = \{-1,1\}^{\Lambda_N} = \left\{\sigma = (\sigma_x)_{x\in\Lambda_N} \ \text{ tq } \ \sigma_x \in \{\pm1\}\right\}$, avec $|E| = 2^{N^2}$;
- $x \simeq y \Leftrightarrow |x-y| = 1$ avec condition au bord périodique ;
- $H(\sigma) = -\sum\limits_{x\simeq y}\sigma_x\sigma_y = -\sum\limits_{x\in\Lambda_N}\sigma_x \sum\limits_{\substack{y\in\Lambda_N\\ y\simeq x}}\sigma_y$.

$\forall y \in \Lambda_N$, $s_y \in \{\pm1\}$ fixés, on a :

$$\begin{aligned}
\mathbb{P}\left(\sigma_x = s_x \mid \sigma_y = s_y, \ \forall y \neq x\right) &= \frac{\mathbb{P}\left(\sigma_x = s_x, \ \sigma_y = s_y, \ \forall y\neq x\right)}{\mathbb{P}\left(\sigma_y = s_y, \ \forall y \neq x\right)} \\
&= \frac{e^{\beta\sigma_x \sum\limits_{y|y\simeq x}\sigma_y}}{e^{\beta\sum\limits_{y|y\simeq x}\sigma_y} + e^{-\beta\sum\limits_{y|y\simeq x}\sigma_y}}
\end{aligned}$$

Cette quantité est maximale lorsque σ_x et σ_y voisins sont de même signe.

2.2 Simulation pratique

Comment simuler $X \simeq \mu_\beta$? On rappelle la méthode du rejet :

Soient E quelconque, λ mesure de référence sur E, X de densité f par rapport à λ.

Considérons $\mathcal{F} = \{(x,y) \in E\times\mathbb{R}_+, \ 0 \le y \le f(x)\}$.

Si $(X,Y) \simeq Unif$ sur $\mathcal{F}$, alors $X \simeq f d\lambda$.

En effet :

$$\mathbb{P}(X\in A) = \int_A\int_0^{f(x)} dy d\lambda(x) = \int_A f(x)d\lambda(x)$$

Supposons qu'on ait à disposition une densité g sur E (par rapport à λ) telle que :

- On sait simuler $\widetilde{X} \simeq gd\lambda$;
- Il existe $M > 0$, $f(x) \leq Mg(x)$.

La méthode de rejet consiste à :

- On prend $\widetilde{X}$ de loi $gd\lambda$;
- On prend Y uniforme sur $\left[0, Mg\left(\widetilde{X}\right)\right]$;
- On prend $X := \widetilde{X}$ si $0 \leq Y \leq f\left(\widetilde{X}\right)$.

Proposition : .

La variable X définie précédemment a pour loi $fd\lambda$.

Preuve :

On peut calculer :

$$\begin{aligned}
\mathbb{P}(X \in A) &= \mathbb{P}\left(\widetilde{X} \in A \mid 0 \leq Y \leq f\left(\widetilde{X}\right)\right) \\
&= \frac{\mathbb{P}\left(\widetilde{X} \in A \text{ et } 0 \leq Y \leq f\left(\widetilde{X}\right)\right)}{\mathbb{P}\left(Y \leq f\left(\widetilde{X}\right)\right)} \\
&= \frac{\int_A \left(\int_0^{f(\tilde{x})} \frac{dy}{Mg(\tilde{x})}\right) g(\tilde{x}) d\lambda(\tilde{x})}{\int_E \left(\int_0^{f(\tilde{x})} \frac{dy}{Mg(\tilde{x})}\right) g(\tilde{x}) d\lambda(\tilde{x})} \\
&= \frac{\frac{1}{M}\int_A f(\tilde{x}) d\lambda(\tilde{x})}{\frac{1}{M}\int_E f d\lambda} \\
&= \int_A f d\lambda
\end{aligned}$$

D'où le résultat attendu. ⋆

Remarque : *.*

On a montré que la probabilité d'acceptation $\mathbb{P}\left(Y \leq f\left(\widetilde{X}\right)\right)$ vaut $\frac{1}{M}$.

Une deuxième méthode de simulation est l'algorithme de Métropolis-Hastings.

On a une mesure de la forme :

$$\mu_\beta(\sigma) = \frac{1}{Z_\beta} e^{-\beta H(\sigma)} \quad ; \quad Z_\beta = \sum_{\sigma \in \{\pm 1\}^{\Lambda_N}} e^{-\beta H(\sigma)}$$

$$\Lambda_N = \{1, \ldots, N\} \times \{1, \ldots, N\}$$

Le problème est que Z_β n'est pas calculable. De manière générale, on considère :

- $\mu(x) = \frac{1}{Z}\eta(x)$ tq $Z = \sum\limits_{x \in E} \eta(x)$;
- $\eta(x) > 0$ pour tout $x \in E$ que l'on sait calculer.

On part d'une chaîne de Markov homogène de matrice de transition Q telle que :

- $\frac{Q_{y,x}}{Q_{x,y}} > 0, \quad \forall x, y \in E$;
- La chaîne (des "propositions") est irréductible ;

On construit le processus $(X_n)_{n\in\mathbb{N}}$ de la façon suivante :

- X_0 de loi arbitraire ;
- Sachant $X_0 = x_0, \ldots, X_n = x_n$, on tire y_n de loi $Q_{X_n,\cdot}$ (Loi de la chaîne sachant qu'on est en X_n).
- On accepte y_n, c'est à dire que l'on pose $x_{n+1} = y_n$ avec probabilité :

$$\rho_{x_n,y_n} \text{ avec } \rho_{x,y} = \min\left(1, \frac{\eta(y)Q_{y,x}}{\eta(x)Q_{x,y}}\right)$$

- Sinon, on pose $x_{n+1} = x_n$.

Théorème : .

La mesure μ est l'unique mesure de probabilité invariante de la chaîne de Markov $(X_n)_{n\in\mathbb{N}}$.

Preuve :

$(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov irréductible car Q l'est, E fini et $\rho_{x,y} \neq 0$.
De plus, puisque E est fini, la chaîne est récurrente positive et il existe une unique probabilité invariante. Montrons ensuite que $(X_n)_{n\in\mathbb{N}}$ est réversible par rapport à μ.

***Remarque :** .*

$$\rho_{x,y} = \min\left(1, \frac{\mu(y)Q_{y,x}}{\mu(x)Q_{x,y}}\right) \text{ car } \mu(x) = \frac{\eta(x)}{Z}$$

La matrice de transition de $(X_n)_{n\in\mathbb{N}}$ est définie comme :

$$\begin{aligned} P_{x,y} &= Q_{x,y}\rho_{x,y}, \quad \textit{si } x \neq y \\ P_{x,x} &= 1 - \sum_{y\neq x} P_{x,y} \textit{ car matrice de transition} \end{aligned}$$

$$\begin{aligned} \forall x \neq y,\ \mu(x)P_{x,y} &= \mu(x)Q_{x,y}\rho_{x,y} \\ &= \mu(x)Q_{x,y}\min\left(1, \frac{\mu(y)Q_{y,x}}{\mu(x)Q_{x,y}}\right) \\ &= \min\left(\mu(x)Q_{x,y}, \mu_y Q_{y,x}\right) \\ &= \mu(y)Q_{y,x}\underbrace{\min\left(\frac{\mu(x)Q_{x,y}}{\mu(y)Q_{y,x}}, 1\right)}_{=\rho_{y,x}} \\ &= \mu(y)P_{y,x} \end{aligned}$$

Puisque le cas $x = y$ est trivial, $(X_n)_{n\in\mathbb{N}}$ est bien réversible. ⋆

Remarque : .

$$\rho_{x,y} = \alpha\left(\frac{\mu_y Q_{y,x}}{\mu_x Q_{x,y}}\right) \text{ avec } \alpha(u) = \min(1,u)$$

On pourrait en réalité prendre n'importe quelle fonction α satisfaisant $\alpha(u) = u\alpha\left(\frac{1}{u}\right)$, comme $\alpha(u) = u/(1+u)$.

Si on retourne au modèle Gibbs, en prenant $Q_{x,y} = Q_{y,x}$, on a :

$$\begin{aligned}\rho_{x,y} = 1 &\iff \min\left(1, e^{-\beta(H(y)-H(x))}\right) = 1\\ &\iff H(y) \le H(x)\end{aligned}$$

3 Échantillonnage de Gibbs

L'idée générale est de simuler $X = \left(X^1, \dots, X^d\right)$, où $d \ge 1$, et on simule coordonné par coordonnée via les lois conditionnelles.

3.1 Cas $d = 2$

Soient (X,Y) sur $E_1 \times E_2$ discret donnée par $\mathbb{P}((X,Y) = (x,y)) = f(x,y)$.

- $f_X(x) = \sum_{y\in E_2} f(x,y)$;
- $f_Y(y) = \sum_{x\in E_1} f(x,y)$;
- $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$;
- $f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$.

Algorithme : Initialisation (x_0, y_0). (Gibbs sampler)
Étant donné (x_n, y_n), on tire :

$$\begin{cases} x_{n+1} & \simeq & f_{X|Y}(\cdot \mid y_n) \\ y_{n+1} & \simeq & f_{Y|X}(\cdot \mid x_{n+1}) \end{cases}$$

Théorème : .
Supposons $f_X(x) > 0$ et $f_Y(y) > 0$, $\forall x,y \in E_1 \times E_2$.
Alors, $f(x,y)$ est une densité de probabilité invariante par rapport à l'algorithme précédent et $(X_n, Y_n) \xrightarrow[n\to+\infty]{loi} f$.
Preuve :

L'idée de la preuve est la suivante :

$$
\begin{aligned}
(fP)_{x',y'} &= \sum_{x,y\in E_1,E_2} f(x,y)\mathbb{P}_{(x,y),(x',y')} \\
&= \sum_{y\in E_2}\left(\underbrace{\sum_{x\in E} f(x,y)}_{=f_Y(y)}\right) f_{X|Y}(x'|y) f_{Y|X}(x'|y') \\
&= \sum_{y\in E_2} f(x',y) f_{Y|X}(y'|x') \\
&= f_{Y|X}(y'|x')\left(\underbrace{\sum_{y\in E_2} f(x',y)}_{=f_X(x)}\right) \\
&= f(x',y')
\end{aligned}
$$

Pour terminer la preuve, il suffit de montrer que la chaîne est irréductible (admis), et f est une unique proba invariante. ⋆

Remarque : .
(X_n, Y_n) est une chaîne de Markov sur $E_1 \times E_2$.

$$
\begin{aligned}
\mathbb{P}_{(x,y),(x',y')} &= \mathbb{P}\left(X_{n+1} = x', Y_{n+1} = y' \mid X_n = x,\ Y_n = y\right) \\
&= f_{X|Y}(x'|y) f_{Y|X}(y'|x')
\end{aligned}
$$

3.2 Cas général

Algorithme : Initialisation $\left(x_0^{(1)}, \ldots, x_0^{(d)}\right)$.
Étant donné $\left(x_n^{(1)}, \ldots, x_n^{(d)}\right)$, on tire :

$$
\begin{cases}
x_{n+1}^{(1)} & \simeq \quad f_{X^{(1)}|X^{(2)},\ldots,X^{(d)}}\left(\cdot|x_n^{(2)}, \ldots, x_n^{(d)}\right) \\
x_{n+1}^{(2)} & \simeq \quad f_{X^{(2)}|X^{(1)},X^{(3)},\ldots,X^{(d)}}\left(\cdot|x_{n+1}^{(1)}, x_n^{(3)} \ldots, x_n^{(d)}\right) \\
& \vdots \\
x_{n+1}^{(d)} & \simeq \quad f_{X^{(d)}|X^{(1)},\ldots,X^{(d-1)}}\left(\cdot|x_{n+1}^{(1)}, \ldots, x_{n+1}^{(d-1)}\right)
\end{cases}
$$

On obtient alors les mêmes résultats que pour le cas $d = 2$ avec les mêmes preuves.

3.3 Application au modèle d'Ising

On peut calculer :

$$
\begin{aligned}
\mathbb{P}\left(\sigma_x = 1 \mid \sigma_y = \widehat{\sigma_y},\ y \neq x\right) &= \frac{\mathbb{P}\left(\sigma_x = 1,\ \sigma_y = \widehat{\sigma_y},\ y \neq x\right)}{\mathbb{P}\left(\sigma_y = \widehat{\sigma_y},\ y \neq x\right)} \\
&= \frac{\frac{1}{Z} exp\left(\beta \sum\limits_{v \in V_x} \widehat{\sigma_v} + \beta \sum\limits_{\substack{u \simeq v \\ u \neq x}} \widehat{\sigma_u}\widehat{\sigma_v}\right)}{\mathbb{P}\left(\sigma_x = 1,\ \sigma_y = \widehat{\sigma_y},\ y \neq x\right) + \mathbb{P}\left(\sigma_x = -1,\ \sigma_y = \widehat{\sigma_y},\ y \neq x\right)} \\
&= \frac{\frac{1}{Z} exp\left(\beta \sum\limits_{v \in V_x} \widehat{\sigma_v} + \beta \sum\limits_{\substack{u \simeq v \\ u \neq x}} \widehat{\sigma_u}\widehat{\sigma_v}\right)}{\frac{1}{Z} exp\left(\beta \sum\limits_{v \in V_x} \widehat{\sigma_v} + \beta \sum\limits_{\substack{u \simeq v \\ u \neq x}} \widehat{\sigma_u}\widehat{\sigma_v}\right) + \frac{1}{Z} exp\left(-\beta \sum\limits_{v \in V_x} \widehat{\sigma_v} + \beta \sum\limits_{\substack{u \simeq v \\ u \neq x}} \widehat{\sigma_u}\widehat{\sigma_v}\right)} \\
&= \frac{exp\left(\beta \sum\limits_{v \in V_x} \widehat{\sigma_v}\right)}{exp\left(\beta \sum\limits_{v \in V_x} \widehat{\sigma_v}\right) + exp\left(-\beta \sum\limits_{v \in V_x} \widehat{\sigma_v}\right)} \\
&= \frac{1}{1 + exp\left(-2\beta \sum\limits_{v \in V_x} \widehat{\sigma_v}\right)} := q\left(x \mid \widehat{\sigma_y}\ \text{ tq }\ y \in V_x\right)
\end{aligned}
$$

Dans le modèle d'Ising, on pose :

$$
\begin{aligned}
N &= \{(i,j) \in \Lambda_N \ \text{ tq }\ i = j \text{ mod } 2\} \\
B &= \{(i,j) \in \Lambda_N \ \text{ tq }\ i = j + 1 \text{ mod } 2\} \\
\Lambda_N &= N \cup B
\end{aligned}
$$

Ainsi, de la même façon que dans le calcul précédent, on obtient :

$$
\mathbb{P}\left(\sigma_N = \widehat{\sigma_N} \mid \sigma_B = \widehat{\sigma_B}\right) = \prod_{x \in N} q\left(x \mid \widehat{\sigma_y},\ y \simeq x\right)
$$

On obtient donc que :

$$
\mathbb{P}\left(\sigma_B = \widehat{\sigma_B} \mid \sigma_N = \widehat{\sigma_N}\right) = \prod_{x \in B} \mathbb{P}\left(\sigma_x = \widehat{\sigma_x} \mid \sigma_N = \widehat{\sigma_N}\right)
$$

4 Recuit simulé

4.1 Idée de la méthode

On souhaite minimiser l'énergie $H : E \to \mathbb{R}$, où E fini, et :

$$
\begin{aligned}
\mathcal{M} &= \{x \in E \ \text{ tq }\ H(x) = H_{min}\} \\
&= arg \min_{x \in E} H(x)
\end{aligned}
$$

On écrit $\mu_\beta(x) = \frac{1}{Z_\beta} exp\left(-\beta H(x)\right)$, où $\beta = \frac{1}{\text{Température}} > 0$ et on souhaite simuler $x \simeq \mu_\beta$ avec β grand. Pour simuler μ_β, on dispose de Métropolis-Hastings pour un β fixé.
Il faut choisir correctement β.

Rappel.
Si $Q_{x,y} = Q_{y,x}$, alors, en partant de x, on accepte $y \simeq Q_x$ avec probabilité:

$$\min\left(1, \frac{\mu_\beta(y)}{\mu_\beta(y)}\right) = \min\left(1, e^{-\beta(H(y)-H(x))}\right)$$

- Si $H(y) \leq H(x)$, on accepte toujours.
- Si $H(y) > H(x)$, on accepte avec probabilité $e^{-\beta(H(y)-H(x))}$.

L'idée donc est de prendre $\beta = \beta_n$. On obtient une chaîne de Markov non homogène :

$$P_{x,y}^{(n)} = \begin{cases} Q_{x,y} \min\left(1, exp\left(-\beta_n \left(H(y) - H(x)\right)\right)\right) \text{ si } x \neq y \\ 1 - \sum\limits_{y \neq x} P_{x,y}^{(n)} \text{ si } x = y \end{cases}$$

Théorème : . *(d'Hajek)*
Pour X_n une chaîne de Markov de matrice de transition $P^{(n)}$, il existe $H^* > 0$ telle que $\lim\limits_{n \to +\infty} \mathbb{P}\left(X_n \in \mathcal{M}\right) = 1$ si, et seulement si, $\beta_n \to \infty$ et $\sum\limits_{n=1}^{\infty} e^{-\beta_n H^*} = \infty$.
Preuve :

Admis. ⋆

Le choix de β_n est appelé un schéma de refroidissement.

Exemple : .
Dans le théorème, on peut prendre par exemple $\beta_n = c\log(n)$, $c > 0$ avec $c \leq \frac{1}{H^*}$ pour satisfaire la seconde condition.

Exemple : .
Le choix $\beta_n = k$, quand $n \in \left[e^{(k-1)h},\ e^{kh}\right]$ fonctionne aussi si $h \geq H^*$ pour satisfaire la seconde condition.

Considérons l'ensemble des chemin de x à $\mathcal{M}$ comme :

$$\Gamma_x = \left\{(x_0, \dots, x_m) \mid x_0 = x, x_m \in \mathcal{M}, x_0, \dots, x_m \in E,\ Q_{x_i,x_{i+1}} > 0\ \forall i\right\}$$

La constante H^* s'exprime comme :

$$H^* = \max_{x \notin \mathcal{M}} \min_{(x_0,\dots,xm) \in \Gamma_x} \max_{i \in [\![1,m]\!]} H(x_i) - \min H > 0$$

4.2 Application au voyageur de commerce

On dispose de $v_1, \ldots, v_n$ villes et :

- On part de v_1 et on revient en v_1 ;
- On passe par toutes les villes ;
- On minimise la distance parcourue.

$$\begin{aligned} E &= \{\text{permutations à } n \text{ éléments}\} \\ H(\sigma) &= \sum_{i=1}^{n} d\left(X_i, X_{\sigma(i)}\right) \\ \mu_\beta(\sigma) &= \frac{1}{Z_\beta} e^{-\beta H(\sigma)} \end{aligned}$$

Reste à choisir une permutation sur les villes. On peut proposer de tirer j, k tq $j < k$, et de retourner tout le sous-cycle $\left(\sigma_j, \sigma_{j+1}, \ldots, \sigma_k\right)$, ce qui est bien mieux que de simplement inter-changer σ_j et σ_k. Ainsi, on obtient :

$$H(\widehat{\sigma}) - H(\sigma) = d(x_{\sigma_{j-1}}, x_{\sigma_k}) - d(x_{\sigma_{j-1}}, x_{\sigma_j}) + d(x_{\sigma_j}, x_{\sigma_{k+1}}) - d(x_{\sigma_k}, x_{\sigma_{k+1}})$$

5 Échantillonnage et intégration numérique dans le cas continu

5.1 Motivation

Un méthode de Monte-Carlo est un algorithme d'échantillonnage, ou d'intégration numérique, afin d'estimer $\int f d\mu$ par $\sum_{i=1}^{n} w_i f(x_i)$ quand x_i sont aléatoires.
Si μ est discrète (μ sur E discret), on dispose de méthodes dites MCMC (Markov-Chain Monte-Carlo) où, pour X_n une chaîne de Markov :

- (Échantillonnage) $\mathbb{P}(X_n = x) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}(X = x)$, $\quad \forall x \in E$;
- (Intégration Numérique) $\frac{1}{n}\sum_{i=1}^{n} f(X_i) \underset{n\to+\infty}{\overset{ps}{\longrightarrow}} \int f d\mu + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ par le TLC, sous des conditions raisonnables.

Que dire maintenant si μ est continue/à densité sur $\mathbb{R}^d$?

5.2 Échantillonnage dans le cas continu

On part du principe qu'on sait simuler une uniforme sur $[0,1]$.

- Dans le cas où la fonction de répartition F_X est connue, alors on peut tirer X selon la formule $F^{-1}(U)$, où $U \simeq Unif([0,1])$.

- On peut utiliser des astuces de calcul peuvent nous donner des méthodes de tirage aléatoire. Par exemple, pour tirer une $\mathcal{N}(0,1)$, un changement en cordonnées polaires nous donne que :

$$X \simeq \mathcal{N}(0,1) = \varepsilon(\frac{1}{2}) \times \cos\left(2\pi \times Unif\left([0,1]\right)\right)$$

- On peut utiliser une méthode de rejet. Supposant que μ a une densité $f(x)$, alors :
  - Densité $g(x)$ avec laquelle on sait échantillonner ;
  - $f(x) \leq Mg(x), \quad \forall x, \ M$ connue ;
  - On tire $Y \simeq g$, $U \simeq Unif\left([0, Mg(Y)]\right)$ et on accepte si $U \leq f(Y)$. La probabilité d'acceptation est $\frac{1}{M}$, le temps d'attente moyen est donc de M.

On peut maintenant s'intéresser aux méthodes MCMC.

Définition : .

Un chaîne de Markov sur un espace quelconque E mesurable, muni d'une tribu $\mathcal{T}$, est caractérisée par une mesure de probabilité μ sur E (loi de X_0) et d'un noyau de transition $P(x, dy)$ vérifiant :

- $\forall x \in E, \quad A \mapsto P(x, A)$ est une proba sur E ;
- $\forall A \in \mathcal{T}, \quad x \mapsto P(x, A)$ est mesurable ;

Ainsi, on a :

$$\begin{aligned} \mathbb{P}(X_0 \in A) &= \int_A d\mu = \mu(A) \\ \mathbb{P}(X_{n+1} \in A \mid X_{n-1} = x) = P(x, A) \end{aligned}$$

Notation : .

Si $g : E \to \mathbb{R}$ est mesurable et ν est une probabilité sur E, alors :

$$\begin{aligned} \nu P(A) &= \int P(x, A)\, d\nu(x) \\ Pg(x) &= = \int g(y) P(x, dy) \\ \nu Pg &= \int g(y) P(x, dy)\, d\nu(x) \end{aligned}$$

Si P, Q sont deux noyaux de transition, alors :

$$PQ(x, A) = \int Q(y, A)\, P(x, dy)$$

La loi de X_n est donc μP^n et une mesure π sur E est invariante si, et seulement si, $\pi = \pi P$.

Remarque : .

Si on revient au cas discret, on a :

$$\int_A d\nu(x) = \sum_{x \in A} \nu(x)$$

Remarque : .

Les notions d'irréductibilité et de récurrence sont plus compliquées...

5.3 Métropolis-Hastings dans le cas continu

Supposons $E = \mathbb{R}^d$, la mesure cible est π.

Problème : échantillonner selon $\pi(x)$

- Noyau de transition $Q(x,dy) = Q(x,y)dy$ (hypothèse) ;
- Algorithme : Partant d'un x_0 initial, $\forall n \geq 1$ on tire $y_{n+1} \simeq Q(x_{n+1},\cdot)$. On prend :

$$x_{n+1} = \begin{cases} y_{n+1} & \text{avec proba } \rho(x_n, y_{n+1}) \\ x_n & \text{avec proba } 1-\rho(x_n, y_{n+1}) \end{cases} \quad \text{tq} \quad \rho(x,y) = \min\left(1, \frac{\pi(y)Q(x,y)}{\pi(x)Q(y,x)}\right)$$

On suppose que $Q(x,y) > 0$, $\forall x \in \mathbb{R}^d$ et pour π-presque tout $y \in \mathbb{R}^d$.

On obtient que π est l'unique probabilité invariante de la chaîne de Markov induite par cet algorithme.

Cas particulier : Si $Q(x,y) = Q(y)$ et $\|\frac{\pi}{Q}\|_\infty < \infty$, alors :

$$\exists M > 0,\ \exists 0 < r < 1 \ \text{ tq } \ \forall x \in E,\ \forall n \geq 1$$

$$\|P^n(x,\cdot) - \pi\|_{TV} = \sup_{A\subset\mathcal{T}} \left|P^n(x,A) - \pi(A)\right| \leq Mr^n$$

En pratique, on prend souvent $Q(x,\cdot) =$ loi de $x+Y$ avec Y indépendant de x et on sait dire des choses sur l'erreur.

5.4 Intégration numérique dans le cas continu

L'objectif est de calculer $\int_{\mathbb{R}^d} g(x)\pi(x)dx \simeq \sum_{i=1}^n w_i g(X_i)$, où π est la densité.

- Analyse numérique (maillage ou grille) : ne fonctionne pas quand g grand ("curse of dimensionnality") ;

- Monte-Carlo : Si $X_1,\dots,X_n$ iid de loi π (ex : π compliqué mais $X_1,\dots,X_n$ obtenus par rejet), alors :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n g(X_i) - \int g d\pi\right) \underset{n\to+\infty}{\longrightarrow} \mathcal{N}(0,\sigma^2)$$

$$\text{tq} \quad \sigma^2 = Var(g(X_1)) = \int g^2 d\pi - \left(\int g d\pi\right)^2$$

Conclusion : $\frac{1}{n}\sum_{i=1}^n g(X_i) \underset{n\to\infty}{\simeq} \int_{\mathbb{R}^d} g d\pi + \frac{\sqrt{Var(g(X_1))}}{\sqrt{n}}$.
Problème : σ est exponentielle en la dimension d.

Remarque : .

Si $X_1, \dots, X_n$ iid de densité $q > 0$, alors :

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i)\frac{\pi(X_i)}{q(X_i)} \underset{n\to+\infty}{\longrightarrow} \int g(x)\pi(x)dx + \sqrt{Var\left(\dots\right)}\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$$

- MCMC : $(X_n) \leftarrow$ Métropolis-Hastings (ou échantillonnage de Gibbs)

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \underset{n\to+\infty}{\simeq} \int f d\pi + \frac{\sigma}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)$$

De plus, σ est ici polynomial en d.

6 Application des méthodes MCMC à l'approche bayésienne

6.1 Introduction à l'approche bayésienne

Exemple : .

On dispose d'une urne avec 2 boules, dont une noire. L'autre boule peut-être noire ou blanche.
Deux modèles possibles : NN ou NB.
Probabilité à priori des modèles : $\mathbb{P}\left(NN\right) = \frac{1}{2} = \mathbb{P}\left(NB\right)$.

Ensuite, on fait des observations. Imaginons que l'on tire une boule noire. Comment est-ce que cela change mon à priori ?
Calcul de vraisemblance : $\mathbb{P}\left(N\ |NN\right) = 1$ et $\mathbb{P}\left(N\ |NB\right) = \frac{1}{2}$.

On calcule ensuite les probabilités à postériori :

$$\begin{aligned}
\mathbb{P}\left(NN\ |N\right) = 1 &= \frac{\mathbb{P}\left(N\ |NN\right)\mathbb{P}\left(NN\right)}{\mathbb{P}\left(N\right)}\\
&= \frac{\mathbb{P}\left(N\ |NN\right)\mathbb{P}\left(NN\right)}{\mathbb{P}\left(N\ |NN\right)\mathbb{P}\left(NN\right)+\mathbb{P}\left(N\ |NB\right)\mathbb{P}\left(NB\right)}\\
&= \frac{2}{3}\\
\mathbb{P}\left(NB \mid N\right) = \frac{1}{3}
\end{aligned}$$

On aurait donc tendance à choisir le modèle NN, ce qui parait assez logique.

Dans le cadre général, on dispose de modèles paramétrés par $\theta \in \Theta$. On se donne une loi à priori sur Θ : $p(\theta)$, puis on observe des données x.
On peut donc calculer la vraisemblance $p\left(x|\theta\right)$ et on obtient la loi à postériori sur θ :

$$p\left(\theta|x\right) = \frac{p\left(x|\theta\right)p(\theta)}{\int_{\Theta} p\left(x|\theta\right)p(\theta)d\theta}$$

Dans cette notation, $d\theta$ est une mesure de référence sur Θ.

Par exemple : Lebesgue si $\Theta \subset \mathbb{R}^d$ "continu", Comptage si Θ est discret.

Estimation de θ ? En théorie décisionnelle, on se fixe une fonction de coût L sur $\Theta \times \Theta$. L'erreur commise est donc $\mathbb{E}_\theta L\left(\widehat{\theta}, \theta\right)$. On choisit alors $\widehat{\theta}$ minimisant cette erreur.

En considérant la fonction de perte $0-1$, on trouve $\widehat{\theta}$ le maximum à postériori :

$$arg \max_{\theta \in \Theta} p\left(\theta | x\right)$$

6.2 Exemple avec la régression logistique

On considère des individus i, d'information $X^{(i)} \in \mathbb{R}^d$, et $Y_i \in \{0,1\}$ binaire.

L'objectif est de modéliser $\mathbb{P}\left(Y=1|X\right) \in [0,1]$.

En général, on considère un modèle de la forme :

$$\mathbb{P}\left(Y=1|X\right) = \sigma\left(\beta_0 + X_1\beta_1 + \ldots + \beta_d X_d\right) \quad \text{tq} \quad \sigma(x) = \frac{1}{1+e^{-x}}$$

L'approche fréquentiste est la suivante :

Densité de Y : $p(y|x) = \mathbb{P}\left(Y=1|x\right)^y \left(1 - \mathbb{P}\left(Y=1|x\right)\right)^{1-y}$.

Maintenant, on sait que l'on dispose des données $\left(X^{(i)}, y_i\right)_{1\leq i\leq n}$ et $\mathbf{X} = \left[X^{(i)}\right]$. La vraisemblance s'écrit donc :

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n p\left(y_i | X^{(i)}\right) \\ &= \prod_{i=1}^n \sigma\left(\beta_0 + X^{(i)}\beta\right)^{y_i}\left(1 - \sigma\left(\beta_0 + X^{(i)}\beta\right)\right)^{1-y_i} \end{aligned}$$

En cherchant le maximum de vraisemblance, puisque L est concave alors $\widehat{\beta}$ est unique et on fait :

$$\widehat{\beta} = arg \max_{\beta \in \mathbb{R}^{d+1}} L(\beta)$$

L'approche bayésienne est la suivante :

On dispose d'un prior (proba à priori sur β) : $p(\beta)$. (Exemple : $\mathcal{N}(0,1)$, $Unif\left([-a,a]\right)$.

On calcule un postérior (proba à postériori sur β) :

$$\begin{aligned} p\left(\beta | X, y\right) &= C \times p\left(y | \beta X\right) p(\beta) \\ &= C \prod_{i=1}^n \sigma\left(\beta_0 + X^{(i)}\beta\right)^{y_i}\left(1 - \sigma\left(\beta_0 + X^{(i)}\beta\right)\right)^{1-y_i} p(\beta) \\ &= \eta(\beta) \quad \text{tq} \quad C = \frac{1}{\int \eta(\beta) d\beta} \end{aligned}$$

7 Chaînes de Markov en temps continu

L'objectif ici est de définir une chaîne de Markov $(X_t)_{t\in\mathbb{R}^+}$.

Pour simplifier, supposons que $X_t \in E$, avec E discret (fini ou dénombrable).

7.1 Processus de Poisson

Premier exemple : processus de comptage d'un phénomène aléatoire (ex : arrivée d'un patient dans un hôpital, composant électrique qui devient défectueux dans un système, ...).
Initialisation $X_0 = 0$, $T_0 = 0$.

$$\begin{aligned} T_n &= S_1 + \ldots + S_n \text{ tq } S_i \text{ iid } \varepsilon(\lambda),\ \lambda > 0 \\ X_t &= \sup\{n \in \mathbb{N} \mid T_n \leq t\} \\ &= \sum_{n\in\mathbb{N}} \mathbf{1}_{T_n \leq t} \end{aligned}$$

Remarque : *.*

$$\{X_t \geq n\} = \{T_n \leq t\}$$

$$\{X_t \geq n\} = \{T_n \leq t < T_{n+1}\}$$

$$\{X_s < n \leq X_t\} = \{s < T_n \leq t\}$$

Rappel.
Si $X_1, \ldots, X_n \simeq \varepsilon(\lambda)$, alors $\sum_{i=1}^{n} X_i \simeq \Gamma(n, \lambda)$.

Lemme : **.**

$$\forall t > 0, \quad X_t \overset{loi}{=} \textit{Poisson } (\lambda t)$$

Preuve :

Faisons le calcul :

$$\begin{aligned} \mathbb{P}(X_t = n) &= \mathbb{P}(T_n \leq t < T_n + S_{n+1}) \quad \textit{tq } T_n \textit{ et } S_{n+1} \textit{ indépendantes} \\ &= \int_{\mathbb{R}_+^2} \mathbf{1}_{x \leq t < x+y} d\mathbb{P}_{T_n}(x) \lambda e^{-\lambda y} dy \\ &= \int_{\mathbb{R}_+^2} \mathbf{1}_{x \leq t < x+y} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} dx \lambda e^{-\lambda y} dy \quad \textit{par le rappel précédent} \\ &= \int_0^t \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \left(\underbrace{\int_{t-x}^{\infty} \lambda e^{-\lambda x} dx}_{= e^{-\lambda(t-x)}} \right) dx \\ &= \lambda^n \frac{e^{-\lambda t}}{(n-1)!} \int_0^t x^{n-1} dx \\ &= \frac{(\lambda t)^n}{n!} e^{-\lambda t} \end{aligned}$$

D'où le résultat attendu. ⋆

Remarque : *.*
On a $\frac{\mathbb{E}(X_t)}{t} = \lambda$ est le nombre moyen d'événements entre 0 et t. C'est l'intensité du processus.

On peut montrer que $(X_t)_{tin\mathbb{R}^+}$ a ses accroissements indépendants et stationnaires :

Proposition : .

$\forall m \geq 1$, $\forall 0 \leq t_0 < t_1 < \ldots < t_m$, les variable $X_{t_m - t_{m-1}}, \ldots, X_{t_1} - X_{t_0}$ sont indépendantes et $X_t - X_s \overset{loi}{=} Poisson\ (\lambda(t-s))$.

Preuve :

Admise. ⋆

Définition : .

Un processus de Poisson sur $(X_t)_{t\in\mathbb{R}_+}$ est un processus à valeur dans $\mathbb{N}$ dont les accroissements sont indépendants et stationnaires.

Théorème : .

Si $(X_t)_{t\in\mathbb{R}^+}$ est un processus de Poisson, alors il existe une intensité λ telle que :

$$X_t = \sum_{n\in\mathbb{N}} \mathbf{1}_{S_1+\ldots+S_n \leq t} \quad tq \ \ S_i \ iid \ de \ loi \ \varepsilon(\lambda)$$

Preuve :

Admise. ⋆

7.2 Chaîne de Markov en temps continu

Définition : .

Une chaîne de Markov en temps continu (CMTC) sur E est un processus $(X_t)_{t\in\mathbb{R}_+}$ tel que :

(1) $\forall w \in \Omega$, $t \mapsto X_t(w)$ est constante par morceaux, continu à droite limite à gauche, avec un nombre de discontinuité fini par intervalle compact de $\mathbb{R}_+$;

(2) (Propriété de Markov) $\forall 0 \leq s < t$, la loi conditionnelle de X_t sachant $\{X_u,\ 0 \leq u \leq s\}$ ne dépend que de X_s.

On dit que $(X_t)_{t\in\mathbb{R}_+}$ est homogène si $\mathbb{P}(X_t = y \mid X_s = x)$ ne dépend que de $t-s$. Cette probabilité peut être notée $P_{xy}(t-s)$. On dit que $P(t) = [P_{xy}(t)]_{x,y\in E}$ est un semi-groupe de transition (axiome d'inversibilité pas vérifié pour être un groupe) de X_t, avec $P(0) = Id$.

Notons $\mu(t)$ la loi de X_t.

$$\mathbb{P}\left(X_{t_0} = x_0, \ldots, X_{t_m} = x_m\right) = \mu(0)_{x_0} P_{x_0 x_1}(t_1 - t_0) \ldots P_{x_{m-1}, x_m}(t_m - t_{m-1})$$

Proposition : .

(1) $$\mu(t) = \mu(0)P(t)$$

(2) $$\mathbb{E}[g(X_t)] = \mu(0)P(t)g$$

Preuve :

1. *Il suffit de sommer sur* $x_0, \ldots, x_{m-1}$.
2. *Faisons le calcul :*

$$
\begin{aligned}
P_{xy}(t+s) &= \mathbb{P}\left(X_{t+s} = y \mid X_0 = x\right) \\
&= \frac{\mathbb{P}\left(X_{t+s} = y,\ X_0 = x\right)}{\mu(0)_x} \\
&= \sum_{z\in E} \frac{\mathbb{P}\left(X_{t+s} = y,\ X_s = z,\ X_0 = x\right)}{\mu(0)_x} \\
&= \sum_{z\in E} \frac{\mathbb{P}\left(X_{t+s} = y \mid X_s = z,\ X_0 = x\right)\mathbb{P}\left(X_s = z,\ X_0 = x\right)}{\mu(0)_x} \\
&= \sum_{z\in E} \frac{\mathbb{P}\left(X_{t+s} = y \mid X_s = z\right)\mathbb{P}\left(X_s = z,\ X_0 = x\right)}{\mu(0)_x} \\
&= \sum_{z\in E} \underbrace{\mathbb{P}\left(X_{t+s} = y \mid X_s = z\right)}_{=P_{zy}(t)} \underbrace{\mathbb{P}\left(X_s = z \mid X_0 = x\right)}_{P_{xz}(s)} \\
&= \left(P(s)P(t)\right)_{xy}
\end{aligned}
$$

D'où le résultat attendu. ⋆

Exemple : . (Processus de Poisson)

Montrons qu'il satisfait la propriété de Markov :

$$
\begin{aligned}
&\mathbb{P}\left(X_t = y \mid X_s = x, X_{t_m} = x_m, \ldots, X_{t_0} = x_0\right) \\
&= \frac{\mathbb{P}\left(X_t = y, X_s = x, X_{t_m} = x_m, \ldots, X_{t_0} = x_0\right)}{\mathbb{P}\left(X_s = x, X_{t_m} = x_m, \ldots, X_{t_0} = x_0\right)} \\
&= \frac{\mathbb{P}\left(X_t - X_s = y - x, X_s - X_{t_m} = x - x_m, \ldots, X_{t_1} - X_{t_0} = x_1 - x_0, X_{t_0} = x_0\right)}{\mathbb{P}\left(X_s = x, X_{t_m} = x_m, \ldots, X_{t_0} = x_0\right)} \\
&= \frac{\mathbb{P}\left(X_t - X_s = y - x\right)\mathbb{P}\left(X_s = x, X_{t_m} = x_m, \ldots, X_{t_0} = x_0\right)}{\mathbb{P}\left(X_s = x, X_{t_m} = x_m, \ldots, X_{t_0} = x_0\right)} \\
&= \mathbb{P}\left(X_t - X_s = y - x\right)
\end{aligned}
$$

De même, on a :

$$
\begin{aligned}
\mathbb{P}\left(X_t = y \mid X_s = x\right) &= \frac{\mathbb{P}\left(X_t = y,\ X_s = x\right)}{\mathbb{P}\left(X_s = x\right)} \\
&= \mathbb{P}\left(X_t - X_s = y - x\right)
\end{aligned}
$$

Moralité, on obtient que X_t est une CMTC homogène de semi-groupe :

$$
P_{xy}(t-s) = \mathbb{P}\left(X_t - X_s = y - x\right) = \frac{\left(\lambda\left(t-s\right)\right)^{y-x}}{(y-x)!} e^{-\lambda(t-s)}
$$

La matrice du processus de Poisson est donc :

$$
P_{xy}(t) = \frac{(\lambda t)^{y-x}}{(y-x)!} e^{-\lambda t} \mathbf{1}_{y \geq x}
$$

7.3 Générateur infinitésimal

Définition : .

On dit que $[Q_{xy}]_{x,y\in E}$ est un générateur infinitésimal si :

- $Q_{xy} \geq 0$ si $x \neq y \in E$;
- $Q_{xx} \leq 0, \quad \forall x \in E$;
- $\sum_{y\in E} Q_{xy} = 0$.

Théorème : .

Si $P(t)$ est le semi-groupe de transition d'une CMTC, alors il existe un générateur infinitésimal Q tel que :

$$\begin{aligned} \lim_{t\to 0} \frac{P(t)-I}{t} &= Q := P'(0) \\ \Longrightarrow P_{xy}(t) &= \begin{cases} tQ_{xy} + o(t) & \text{si } x \neq y \\ 1 + tQ_{xx} + o(t) & \text{sinon} \end{cases} \end{aligned}$$

De plus, conditionnellement à $X_0 = x_0$, l'instant de premier saut $T_1 = \inf\{t > 0,\ X_t \neq x_0\}$ est indépendant de la valeur au premier saut $Z_1 := X_{T_1}$ et :

$$\begin{cases} T_1 \mid X_0 = x_0 & \overset{loi}{=} \varepsilon\left(-Q_{x_0,x_0}\right) \\ \mathbb{P}\left(Z_1 = y \mid X_0 = x_0\right) & = \frac{-Q_{xy}}{Q_{x_0,x_0}} \mathbf{1}_{x_0\neq y} \end{cases}$$

Preuve :

Admise. ⋆

Corollaire : .

Soient $T_1, T_2, \ldots$ les temps successifs de changement d'état de $(X_t)_{t\in\mathbb{R}_+}$ CMTC de générateur infinitésimal Q et $Z_1, Z_2, \ldots$ la valeur associée définies comme :

- $T_0 = 0,\ X_0 \simeq \mu(0)$;
- $T_1 = \inf\{t > 0 \mid X_t \neq X_0\},\ Z_1 = X_{T_1}$;
- $\forall n \geq 1,\ T_{n+1} = \inf\left\{t > 0 \mid X_{T_1+\ldots+T_n+t} \neq Z_n\right\},\ Z_{n+1} = X_{T_{n+1}}$.

$(Z_n)_{n\in\mathbb{N}}$ est appelée la chaîne de Markov "incluse" dans $(X_t)_{t\in\mathbb{R}_+}$. Alors :

(1) La loi de $T_{n+1} - T_n \mid Z_n = x$ est $\varepsilon\left(-Q_{x,x}\right)$ et indépendante de $T_1, \ldots, T_n$.

(2) $\mathbb{P}\left(Z_{n+1} = y \mid Z_n = x\right) = P_{x,y} := \begin{cases} -\frac{Q_{x,y}}{Q_{x,x}} & \text{si } x \neq y \\ 0 & \text{sinon} \end{cases}$

Preuve :

Admise. ⋆

Remarque : .

Si E est fini, $P(t) = e^{tQ} \quad \left(= \sum_{n=0}^{\infty} \frac{1}{n!} (tQ)^n \right)$.

Exemple : . (Application au processus de Poisson)

Calculons Q, dans le cas du processus de Poisson, c'est à dire avec :

$$P_{x,y}(t) = \frac{(\lambda t)^{y-x}}{(y-x)!} e^{-\lambda t} \mathbf{1}_{y \geq x}$$

- Si $y < x$, $P_{x,y}(t) = 0 \Rightarrow Q_{x,y} = 0$;
- Si $y = x$, $P_{x,x}(t) = 1 - e^{-\lambda t} + o(t) \Rightarrow Q_{x,x} = -\lambda$;
- Si $y = x+1$, $P_{x,x+1}(t) = \lambda t + o(t) \Rightarrow Q_{x,x+1} = \lambda$;
- Si $y > x+1$, $P_{x,y} = o(t) \Rightarrow Q_{x,y}(t) = 0$.

Ainsi, on conclut que le générateur infinitésimal du processus de Poisson est :

$$Q = \lambda \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \dots & \dots & 0 & -1 \end{pmatrix}$$

Ainsi, dans le cas du processus de Poisson, on conclut que :

$$\begin{aligned} T_{n+1} - T_n &\simeq \varepsilon(\lambda) \\ \mathbb{P}(Z_{n+1} = y \mid Z_n = x) &= \begin{cases} 1 & \text{si } y = x+1 \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

En résumé, pour simuler un CMTC $(X_t)_{t \in \mathbb{R}_+}$, on tire X_0 de loi μ, $T_0 = 0$, $V_1, \dots, V_n$ iid de loi $\varepsilon(1)$. Pour $n \in [\![1, N]\!]$, on a :

$$\begin{cases} T_n = \frac{V_1}{\lambda(Z_0)} + \dots + \frac{V_n}{\lambda(Z_{n-1})} \\ \mathbb{P}(Z_n = y \mid Z_{n-1} = x) = P_{x,y} = \begin{cases} -\frac{Q_{x,y}}{Q_{x,x}} = \frac{Q_{x,y}}{\lambda(x)} & \text{si } x \neq y \\ 0 & \text{sinon} \end{cases} \end{cases}$$

On peut se demander, à l'aide des Z_t, comment reconstruire X_t :

$$X_t = \sum_{n=0}^{\infty} Z_n \mathbf{1}_{[T_n, T_{n+1}]}(t), \quad \forall t > 0$$

7.4 Mesure invariante

Définition : .

On dit que $(X_t)_{t \in \mathbb{R}_+}$ CMTC est irréductible, récurrente, récurrente positive, si $(Z_n)_{n \in \mathbb{N}}$ l'est.

Définition : .

On dit qu'une mesure de probabilité sur E π est invariante si :

$$\pi = \pi P(t), \quad \forall t > 0$$

Proposition : .

$$\boxed{\pi \text{ probabilité invariante} \Longleftrightarrow \pi Q = Q}$$

Preuve :

Admise. ⋆

Théorème : .

Si $(X_t)_{t\in\mathbb{R}_+}$ est un CMTC irréductible et récurrente positive, alors elle admet une unique probabilité invariante π et :

- $\forall x \in E,\ \mu(t)_x \underset{t\to+\infty}{\longrightarrow} \pi_x$;
- $\forall f \in L^1(\pi),\ \frac{1}{t}\int_0^t f(X_s)\,ds \underset{t\to+\infty}{\longrightarrow} \sum_{x\in E} f(x)\pi_x$.

Preuve :

Admise. ⋆

Remarque : .

La deuxième propriété du théorème précédent signifie que la moyenne temporelle est la moyenne spatiale.

Remarque : .

$$\frac{1}{t}\int_0^t f(X_s)\,ds = \frac{T_1}{t}Z_0 + \sum_{k=1}^{N_t-1}\frac{(T_{k+1}-T_k)}{t}Z_k + \frac{(t-T_{N_t})}{t}Z_{N_t} \qquad \text{tq } N_t := \sup\{n \geq 0 \mid T_n \leq t\}$$

7.5 Application aux files d'attente

On s'intéresse par exemple aux clients/patients à un guichet ou aux appels téléphoniques/requêtes sur un serveur.

Notation : . (de Kendal)

Pour une file d'attente, on note A/B/C/D/E/F avec :

- A : Loi des arrivées dans la file ;
- B : Loi des temps de service ;
- C : Nombre de guichets ;
- D : Taille de la salle d'attente (Défaut : ∞) ;
- E : Population totale (Défaut : ∞) ;
- F : Modalités d'opération, par exemple FIFO / FILO (Défaut : FIFO)

On peut dire que A,B = M si "Markovien" ou "Memoryless" où l'inter-arrivées et les temps de traitement sont de loi exponentielle de paramètre respectivement λ et μ. Sinon, on note G dans le cas général. On étudie : M/M/K avec :

- Inter-arrivées $\simeq \varepsilon(\lambda)$ indépendantes ;
- Temps de traitement $\simeq \varepsilon(\mu)$ indépendantes ;
- K guichets.

La taille de la file d'attente au temps t est appelée X_t à valeurs dans $\mathbb{N}$, c'est à dire le nombre de personnes dans la file, ainsi que ceux au guichet.

Si $X_t = 0$, alors $T_{n+1} \simeq \varepsilon(\lambda)$ et $Z_{n+1} = Z_n + 1$, avec $n = \sup\{k \mid T_k \leq t\}$.
Si $X_t = k$, alors il y a $m = \min(k, K)$ personnes au guichet et donc m horloges $\simeq \varepsilon(\mu)$ indépendantes et une horloge $\varepsilon(\lambda)$ en compétition.

Exercice.
Si $X_1, \ldots, X_n$ sont des variables indépendantes de loi $\varepsilon(\lambda_1), \ldots, \varepsilon(\lambda_n)$ respectivement, alors :

$$\begin{aligned} Y &:= \min(X_1, \ldots, X_n) \simeq \varepsilon(\lambda_1 + \ldots + \lambda_n) \\ \mathbb{P}(Y = X_j) &= \frac{\lambda_j}{\lambda_1 + \ldots + \lambda_n} \end{aligned}$$

Ainsi, au prochain temps de saut, la loi est $\min(\varepsilon(\lambda), \varepsilon(\mu), \ldots, \varepsilon(\mu)) = \varepsilon(\lambda + m\mu)$ et :

$$\begin{cases} \mathbb{P}(Z_{n+1} = k+1 \mid Z_n = k) &= \frac{\lambda}{\lambda + m\mu} \\ \mathbb{P}(Z_{n+1} = k+1 \mid Z_n = k) &= \frac{m\mu}{\lambda + m\mu} \end{cases}$$

$$\Longrightarrow \begin{cases} T_{n+1} \mid Z_n = k & \overset{loi}{=} \varepsilon\Bigg(\underbrace{\lambda + \min(k, K)}_{=\lambda(k)}\Bigg) \\ \mathbb{P}(Z_{n+1} = k+1 \mid Z_n = k) & = P_{n,n+1} \text{ connu} \end{cases}$$

Comment retrouver Q ? Puisque $\frac{Q_{x,y}}{-Q_{x,x}} = P_{x,y}\mathbf{1}_{x \neq y}$ et $\lambda(x) = -Q_{x,y}$, alors :

$$\begin{cases} Q_{x,x} &= -\lambda(x) \\ Q_{x,y} &= \lambda(x) P_{x,y} \mathbf{1}_{x \neq y} \end{cases}$$