Distance de variation et couplage de chaînes de Markov.

Autre manière d'étudier la vitesse de convergence d'une CM.
Tt au long de l'exposé, les CM sont irréductibles et apériodique.

1. Distance de variation (totale)

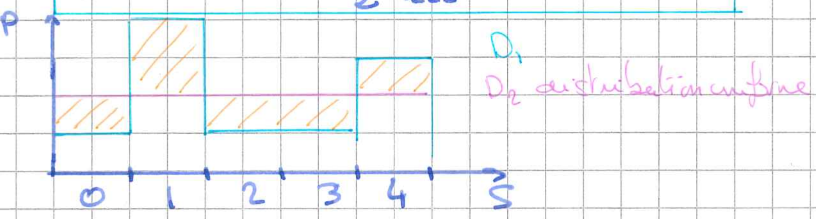
Ex: mélange de n cartes
à chaque étape, une carte est choisie indépendamment et uniformément, puis placée en haut de la pile de cartes.
La probabilité stationnaire est la probabilité uniforme sur tous les ordres de cartes :
$\pi_x = \frac{1}{n} \sum_{y \in N_x} \pi_y$ [N_x successeurs de x ds la CM]
$|N_x| = n \quad \forall x$.

Combien d'étapes faut-il pour approcher "suffisamment" la distribution uniforme ?
↳ distance de variation.

Déf: Soit D_1, D_2 deux distributions sur un espace au plus dénombrable S.
La distance de variation (totale) de D_1 et D_2 est

$$\boxed{\|D_1 - D_2\|_{(TV)} = \frac{1}{2} \sum_{x \in S} |D_1(x) - D_2(x)|}$$

ex:



$\|D_1 - D_2\| = 0{,}6/2 = 0{,}3$

Rq: $0 \leq \|D_1 - D_2\| \leq 1$ → supports disjoints
↳ égalité.

Lemme. définition équivalente : $\forall A \subseteq S$, $D_i(A) = \sum_{x \in A} D_i(x)$
Alors $\boxed{\|D_1 - D_2\| = \max_{A \subseteq S} |D_1(A) - D_2(A)|}$.

Dém. S^+ les élts tels que $D_1(x) \geq D_2(x)$ → $\max_{A \subseteq S} D_1(A) - D_2(A) = D_1(S^+) - D_2(S^+)$
S^- $D_1(x) < D_2(x)$ $\max_{A \subseteq S} D_2(A) - D_1(A) = D_2(S^-) - D_1(S^-)$

On $\overbrace{|D_1(S^+) - D_2(S^+)| = |D_2(S^-) - D_1(S^-)|}^{=1}$

Donc $|D_1(S^+) - D_2(S^+)| + |D_1(S^-) - D_2(S^-)| = \sum_x |D_1(x) - D_2(x)| = 2\|D_1 - D_2\|$.

Ex: Si D_2 uniforme et D_1 : uniforme parmi celles qui ont l'as de pique au dessus
$\|D_1 - D_2\| \geq 1 - \frac{1}{52} = \frac{51}{52}$.

2. Temps de mélange

Déf: Soit π la distribution stationnaire d'une CM sur un espace d'états S.
p_x^t représente la distribution obtenue en partant de l'état x après t étapes.

$\Delta_x(t) = \|p_x^t - \pi\|$; $\Delta(t) = \max_{x \in S} \Delta_x(t)$

$\tau_x(\varepsilon) = \min\{t \mid \Delta_x(t) \leq \varepsilon\}$ $\quad \boxed{\tau(\varepsilon) = \max_{x \in S} \tau_x(\varepsilon)}$

On dit que $\tau(\varepsilon)$ est le temps de mélange de la CM.
Une chaîne mélange rapidement si $\tau(\varepsilon)$ est polynomial en $\log(1/\varepsilon)$ et en la taille du problème.

3. Couplage d'une CM.

Technique générale pour borner le temps de mélange d'une CM.

Déf: Un couplage d'une CM $(M_t)_{t\in\mathbb{N}}$ sur un espace d'états S est une CM $Z_t=(X_t,Y_t)$ sur $S\times S$ telle que

$$\begin{cases} P(X_{t+1}=x' \mid Z_t=(x,y)) = P(M_{t+1}=x' \mid M_t=x) \\ P(Y_{t+1}=y' \mid Z_t=(x,y)) = P(M_{t+1}=y' \mid M_t=y) \end{cases}$$

$\Rightarrow$ (X_t) et (Y_t) se comportent exactement comme (M_t) en terme de transition.
Mais (X_t) et (Y_t) ne sont pas nécessairement indépendants.

On s'intéresse aux couplages qui - mènent à ce que (X_t) et (Y_t) se rejoignent ($\exists t / X_t=Y_t$)
- une fois cet état atteint, sont égaux.

Th: Soit $Z_t=(X_t,Z_t)$ un couplage d'une CM M sur S.
Supposons qu'il existe T tel que $\forall x,y\in S \quad P(X_T\neq Y_T \mid X_0=x, Y_0=y)\leq\varepsilon$
Alors $\tau(\varepsilon)\leq T$.

Dém: On choisit le couplage où $Y_0=\pi$ et $X_0=x\in S$. Soit $A\subseteq S$.

$$\begin{aligned} P(X_T\in A) &\geq P(X_T=Y_T \cap Y_T\in A) \\ &\geq 1-(P(X_T\neq Y_T \cup Y_T\notin A)) \\ &\geq 1-P(X_T\neq Y_T)-P(Y_T\notin A) \qquad [P(X_T\neq Y_T)\leq\varepsilon \text{ ici si } Y_0\sim\pi]. \\ &\geq P(Y_T\in A)-\varepsilon = \pi(A)-\varepsilon \end{aligned}$$

De même, $P(X_T\notin A)\geq P(Y_T\notin A)-\varepsilon$ et $P(X_T\in A)\leq P(Y_T\in A)+\varepsilon=\pi(A)+\varepsilon$

Donc $\max_A |p_x^t(A)-\pi(A)|\leq\varepsilon$.

Ex: mélange de cartes.
- essai 1: choisir j (position) uniformément et mettre la j° carte au dessus.
Mais alors la 1° carte du paquet pour X_t et Y_t va être différente.
$\longrightarrow$ ce n'est pas un couplage efficace.

- essai 2 $\begin{cases} X_{t+1}: \text{choisir une position } j \text{ uniforme entre } 1 \text{ et } n. \text{ La valeur de cette carte est } c. \\ Y_{t+1}: \text{mettre la carte de valeur } c \text{ au dessus du paquet.} \end{cases}$

On a bien un couplage: c est bien choisie uniformément.

$\rightarrow$ une fois la carte choisie et placée au dessus du paquet, cette carte garde la même position dans les deux tas.
Il suffit donc d'avoir choisi toutes les cartes au - une fois pour avoir la même distribution $\rightarrow$ collectionneur de coupons.

$\hookrightarrow$ $\left(1-\frac{1}{n}\right)^{n\ln n+cn}\leq e^{-(\ln n+c)} = \frac{e^{-c}}{n}$ $\leftarrow$ une carte spécifique n'a pas été choisie.

$\hookrightarrow$ proba qu'une carte n'ait jamais été choisie: $\leq e^{-c}$
P :

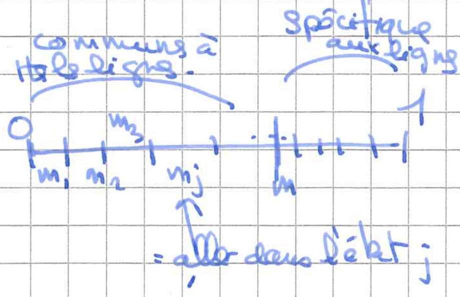
Après un nombre de pas $n \ln n + n \ln(1/\varepsilon)$ on a $p \le e^{-\ln 1/\varepsilon} = e^{\ln \varepsilon} = \varepsilon$.

$\tau(\varepsilon) \le n \ln n/\varepsilon$ → mélange rapide.

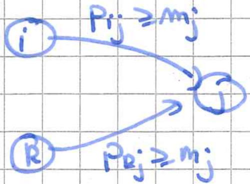
4. Convergence géométrique :

Th : Soit P la matrice de transition d'une CM finie, …
$m_j = \min_i P_{ij}$ et $m = \sum_j m_j$. Alors

$$\boxed{\|p_x^t - \pi\| \le (1-m)^t}$$



Dém : $\forall j$



→ fonction de transition $X_{t+1} = f(X_t, U_{t+1})$

→ si $U_{t+1} \in [0, m]$ → couplage.

La probabilité qu'il n'y ait pas de couplage en t étapes est $(1-m)^t$.

Pb si $\forall$ colonne $\exists$ une entrée nulle → $\exists T$ $P^T_{ij} > 0$ $\forall i,j$ (irred, ap).