Structures et algorithmes aléatoires

Notes début 2013 !

Anne Bouillard, anne.bouillard@ens.fr

On s"intérese dans ce cours à des probabilités discrètes, ie dans des espaces au plus dénombrables.

Applications à l'informatique :

- réseaux de communication
- étude de la distribution de probabilité des entrées d'un algo
- ...

Références bibliographiques :

- Mitzenmacker et Urfal *probability and computing : randomized algorithms and probabilitistic analysis* (2005)
- Pierre Brémaud *initiaion aux probabilités* (2009)
- Pierre Brémaud *markov chains, gibbs fields, monte carlo simulation and queues* (1999)

2013-09-27

Algorithme déterministe : la sortie dépend uniquement de l'entrée (on a une application $I \overset{\text{ALGO}}{\mapsto} f(I)$).

Algorithme probabiliste : la sortie dépend également d'un certain nombre de bits aléatoires r, on a alors une fonction du type $I, r \mapsto f(I, r)$.

Il y a deux types d'algorithmes aléatoires :

- On a une réponse presque toujours correcte, en un temps raisonnable la plupart du temps.
- On a une réponse toujours rapide, correcte la plupart du temps.

Deux exemples.

- *Vérification de l'égalité de deux polynomes.*

  Prenons $F(x) = (x+1)(x-2)(x+3)(x-4)(x+5)(x-6)$ et $G(x) = x^6 - 7x^3 + 25$.

  Pour un algorithme déterministe : développer les polynomes et comparer les coefficients ; complexité $O(d^2)$ pour le développement (avec une méthode naive).

  Algorithme probabiliste : soit $r \in \{1, ..., 100\,d\}$, évaluer $F(r)$ et $G(r)$, Si $F(r) \neq G(r)$ on sait que les polynomes sont différents. Sinon, on prétend qu'ils sont égaux. La probabilité d'une erreur est facilement calculable, c'est $1 / 100$ (puisque l'égalité sur d points implique leur égalité partout, pour deux polynomes distincts il y a moins de d points d'égalité).

- *Canal de communication.*

  On a des entrées qui sont mises dans une file, en attente d'etre traitées pour produire une sortie. On note $X(n)$ le nombre de clients dans la file juste après le départ du n-ème client. On note a_n le nombre de clients arrivant pendant le service du n-ème client.

  $X(0) = 0$

  $X(n+1) = \max(X(n) - 1, 0) + a_n$

$(X(n))_{n\in\mathbb{N}}$ est un *processus stochastique*. Sous certaines hypothèses, c'est une *chaîne de Markov*.

1 I. Évènements et probabilités

Ex. Lancé d'un dé.

« Obtenir 3 » : évènement de probabilité $1/6$.

« Obtenir un chiffre pair » : évènement de probabilité $1/2$.

Ex. On se donne un cercle et on tire une corde au hasard. Quelle est la probabilité que cette corde soit plus longue que le côté d'un triangle équilatéral inscrit dans ce cercle ? C'est $1/3$, en effet on fixe le premier point, on trace le triangle équilatéral passant par ce point et on constate directement que pour le second point, la corde sera plus longue seulement s'il est dans l'arc opposé au premier point.

1.1 a. Tribus et événements

Ω décrit toutes les possibilités d'une expérience. ($\Omega = \{1,2,3,4,5,6\}$ pour un dé).

Les éléments de Ω, ω, sont les épreuves ou les réalisations.

On appelle événement un sous-ensemble de Ω. Par exemple, on peut définir $\{2,4,6\}$ l'évènement « obtenir un chiffre pair ».

Déf. Une tribu sur Ω est unne famille de sous-ensembles de Ω, $\mathcal{F}$, telle que

- $\Omega \in \mathcal{F}$
- Si $A \in \mathcal{F}$, alors $\Omega \backslash A \in \mathcal{F}$
- Si $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$, alors $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

Tribu grossière : $\{\varnothing, \Omega\}$. Tribu fine : $\mathcal{P}(\Omega)$.

Dans la plupart des cas, on travaillera sur la tribu fine.

Prop. Les tribus sont stable par intersection dénombrable (complémentaire, toussa)

1.2 b. Espace de probabilités, axiome des probabilités

Déf. Soit Ω un espace d'épreuves et $\mathcal{F}$ une tribu sur Ω. Une probabilité sur (Ω, F) est une fonction $P : \Omega \mapsto [0,1]$ telle que :

- $P(\Omega) = 1$
- Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements de $\mathcal{F}$ deux à deux disjoints, alors $P\big(\bigcup_{n\in\mathbb{N}} A_n\big) = \sum_{n\in\mathbb{N}} P(A_n)$.

Si $P(A) = 1$ alors A est presque sûr. Si $P(A) = 0$ alors A est presque impossible.

Ex. $\Omega = \{0,1\}^{\mathbb{N}}$, $\omega = \omega_0 \cdots \omega_n$, avec $\omega_i \in \{0,1\}$.

$A = \{\omega \,|\, \omega_i \in A_i, i \leqslant k\}$ avec $A_i \subseteq \{0,1\}$.

Les événements de type A ne sont pas stables par union dénombrable.

Notes 2014. 2014-09-24.

Bibliographie : voir page web du cours. Première partie du cours bien basée sur *Probability and Computing. Randomized Algorithms and Probabilistic Analysis.*

Ω : espace des épreuves ou des réalisations.

Théorème. ***Continuité séquentielle.*** *(calcul de limites de probabilités)*

Soient $(A_n)_{n\in\mathbb{N}} \in \mathfrak{F}^n$ *une suite d'évenements tels que* $\forall n, A_n \subseteq A_{n+1}$*, alors :*

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

Remarque. Mutuelle indépendance $\neq$ indépendance 2 à 2 (c'est plus fort !)

Théorème. *Soit* $(A_n)_{n\in\mathbb{N}}$ *une famille d'évenements mutuellement indépendats, alors* $(B_n)_{n\in\mathbb{N}}$ *aussi, où* $\forall n, B_n = A_n$ *ou* A_n^c*.*

Théorème. *Soit* $(C_n) \in \mathcal{F}^{\mathbb{N}}$ *une famille dénombrable d'évenements mutuellement indépendants. Alors* $\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} C_n\right) = \prod_{n\in\mathbb{N}} \mathbb{P}(C_n)$*.*

Démonstration. *Poser* $B_n = \bigcap_{k=0}^{n} C_k$ *; c'est une suite décroissante. On utilise la continuité séquentielle. (la suite* $\prod_{k\leqslant n} \mathbb{P}(C_k)$ *converge car décroissante positive ; c'est une hypothèse de l'énoncé du théorème).* □

Proposition. *A et B deux évènements.*

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

2014-10-01

Fonctions de VA. Une fonction de VA est une VA. Si $X : \Omega \to E$ et $f : E \to F$, alors $f(X) : \Omega \to F$ est une nouvelle VA (il faut vérifier que ses préimages sont bien mesurables, ce qui est clair par les propriétés des ensembles qui sont ici au plus dénombrables).

Exemples de lois aléatoires.

- Loi de bernoulli : $X \sim \mathrm{Ber}(p)$ si $P(X=1) = p = 1 - P(X=0)$

  Si A est un évènement $(A \in \mathcal{F})$, alors $\mathbf{1}_A \sim \mathrm{Ber}(P(A))$

- Loi binomiale : $X \sim \mathrm{Bin}(n,p)$ si $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

  $X = \sum_{i=1}^{n} X_i$, où $X_i \sim \mathrm{Ber}(p)$ sont des va iid.

- Loi géométrique : $X \sim \mathrm{Geo}(p)$ si $P(X=n) = (1-p)^{n-1}p$ pour tout $n \in \mathbb{N}$

  (probabilité que le premier face intervienne au n^e lancer;

  La loi géométrique est sans mémoire, ie $\forall k \geqslant 0, \forall n \geqslant 1, P(X = n+k | X > k) = P(X = n)$.

- Loi uniforme sur $[0,1]$ (ce n'est pas une loi discrète!) $\forall x \in [0,1], P(X \leqslant x) = x$.
- Loi de Poisson de paramètre λ : $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$

Espérance. (voir poly). Linéarité de l'espérance : requiert l'espérance finie des deux fonctions.

Espérance conditionnelle. Soit X une VA et A un évènement ;

$$\mathbb{E}[X\,|A] \;=\; \sum_{x \in E} x\, P(X = x\,|A)$$

Lemme. *Soient X, Y des VA telles que $\mathbb{E}[X]$ existe.*

$$\mathbb{E}[X] \;=\; \sum_{y \in E} P(Y=y) \cdot \mathbb{E}[X\,|Y=y]$$

Exemples d'espérances.

- $X \sim \mathrm{Ber}(p)$, $\mathbb{E}[X] = p$
- $X \sim \mathrm{Bin}(n,p)$, $\mathbb{E}[X] = n\,p$
- $X \sim \mathrm{Geo}(p)$, $\mathbb{E}[X] = p \sum_{i=1}^{\infty} i\,(1-p)^{i-1} = \frac{1}{p}$

Proposition. *Si X est une VA réelle à valeurs entières, alors :*

$$\mathbb{E}[X] \;=\; \sum_{i \geqslant 1} P(X \geqslant i)$$

Variance. Soit X, Y des VA réelles. On définit :

- Le k-ème moment de X : $\mathbb{E}[X^k]$
- La variance de X : $V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- La covariance de X et Y : $\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])]$
- L'écart–type : $\sigma(X) = \sqrt{V(X)}$

Prop. Si X et Y sont indépendantes, alors $\mathbb{E}[X\,Y] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

Exemples.

- $\mathrm{Var}(\mathrm{Ber}(p)) = p - p^2$
- $\mathrm{Var}(\mathrm{Bin}(n,p)) = n\,p\,(1-p)$
- $\mathrm{Var}(\mathrm{Geo}(p))$ se calcule en isolant le cas où on a le résultat dès la première tentative, et en utilisant la propriété de sans mémoire.

`2014-10-08.`

Sur les génératrices, rappel : $\mathbb{P}(X=n) = \frac{g_X^{(n)}(0)}{n!}$. (vu en prépa)

Exemples de génératrices.

- $X \sim \mathrm{Ber}(p)$: $g_X(s) = (1-p) + s\,p$
- $X \sim \mathrm{Bin}(n,p)$: $g_X(s) = ((1-p) + s\,p)^n$
- $X \sim \mathrm{Geo}(p)$: $g_X(s) = \sum_{n \geqslant 1} s^n (1-p)^{n-1} p = s\,p \sum_{n \geqslant 0} s(1-p)^n = \frac{p\,s}{1-(1-p)s}$.

Propriété. $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = g_X''(1) - g_X'(1)^2 + g_X'(1)$

Plus généralement, $\mathbb{E}[X^k]$ peut s'écrire en fonction de $g_X'(1), ..., g_X^{(k)}(1)$.

Lien entre fonction génératrice et distribution. Soient X et Y deux V.A. de fontions génératrices g_X et g_Y. Si $\exists S > 0$ tq $\forall s \in [0, S]$, $g_X(s) = g_Y(s)$, alors X et Y ont même distribution.

Borne de Chernoff dans l'autre sens. Soient $X_1, ..., X_n$ des VA indépendantes, $X_i \sim \mathrm{Ber}(p_i)$ avec $X = \sum X_i$, $\mu = \mathbb{E}[X]$. Alors $\forall S \in]0,1[$, :

1. $$\mathbb{P}(X \leqslant (1-\delta)\mu) \;\leqslant\; \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$$
2. $$\mathbb{P}(X \leqslant (1-\delta)\mu) \;\leqslant\; e^{\frac{-\mu\delta^2}{2}}$$

Application au tri rapide. Comment évaluer la probabilité que le tri s'effectue en $\beta\, n \log n$? On montre que cette proba est $\geqslant 1 - \frac{1}{n}$. On décrit une exécution du tri comme un arbre où chaque noeud correspond au choix d'un pivot, et a deux sous-arbres qui correspond aux sous-tris effectués. Si un noeud est un tri sur s éléments, alors un de ses fils est dit « bon noeud » si il doit trier au plus $s/2$ éléments. On s'intéresse aux longueurs des branches de l'arbres. Soit b une branche, alors :

1. Il existe au plus $\alpha \log n$ bon noeuds sur b avec $\alpha = \frac{1}{\log 2} \simeq 1,5$

   En effet, supposons qu'il y ait k bons noeuds sur la branche, alors $1 \leqslant \left(\frac{1}{2}\right)^k n$.

2. $\mathbb{P}(|b| \geqslant \beta \log n) \leqslant \frac{1}{n^2}$

   Soit $X_i \sim \mathrm{Ber}(p_i)$ la probabilité que le i-ème noeud de la branche soit un bon noeud. On pose $Y_i \sim \mathrm{Ber}\left(\frac{1}{2}\right)$, telle que $X_i \geqslant Y_i$ p.s. et que les Y_i soient mutuellement indépendantes.

   $\mathbb{P}(|b| \geqslant m) \leqslant \mathbb{P}(\sum_{i=1}^m X_i \leqslant \alpha \log n) \leqslant \mathbb{P}(\sum_{i=1}^m Y_i \leqslant \alpha \log n)$

   On applique la borne de Chernoff...safédégrocalcul

3. $\mathbb{P}(\max_b |b| \geqslant \beta \log n) \leqslant \frac{1}{n}$
4. $\mathbb{P}(\mathrm{QS}(T) \geqslant \beta \log n) \leqslant \frac{1}{n}$

Application : estimation d'un paramètre.

- p : intensité d'une mutation (inconnue)
- n : population testée (échantillons mutuellement indépendants).

Question : évaluer p.

On pose X le nombre de mutations sur la population. $X = \tilde{p}\, n$.

Déf. Un intervalle de confiance de $(1-\gamma)$ pour un paramètre p est un intervalle $[\tilde{p}-\delta, \tilde{p}+\delta]$ tel que $\mathbb{P}(p \in [\tilde{p}-\delta, \tilde{p}+\delta]) \geqslant 1-\gamma$. On veut δ et γ les plus petits possibles.

$X \sim \mathrm{Bin}(n,p)$. $\mathbb{E}[X] = n\,p$, et $X = n\,\tilde{p}$

Deux cas pour $p \notin [\tilde{p}-\delta, \tilde{p}+\delta]$:

- $p < \tilde{p}-\delta$ et $X = n\,\tilde{p} > n(p-\delta) = \mathbb{E}[X]\Big(1+\frac{\delta}{p}\Big)$
- $p > \tilde{p}+\delta$ et $X = n\,\tilde{p} < n(p-\delta) = \mathbb{E}[X]\Big(1-\frac{\delta}{p}\Big)$

$P(x \notin [\tilde{p}-\delta, \tilde{p}+\delta]) = P\Big(X < n\,p\,\Big(1-\frac{\delta}{p}\Big)\Big) + P\Big(X > n\,p\,\Big(1+\frac{\delta}{p}\Big)\Big) \leqslant \ldots.$ (borne de Chernoff).

2014-10-15.

Limite de la loi binomiale.

Si $X_n \sim \mathrm{Bin}(n, p(n))$ avec $n\,p(n) \to \lambda$, alors $\lim_{n\to\infty} P(X_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}$. La distribution limite est la loi de Poisson : $X \sim \mathrm{Poi}(\lambda)$ si $P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$ $\forall k \in \mathbb{N}$.

Preuve : Soit g_n la série génératrice de X_n : $g_n(s) = (1-p(s)\,(1-s))^n$. Si (g_n) converge simplement vers g une série génératrice pour un intervalle $[0,r]$ avec $r>0$, alors la distribution de X_n tend vers la distribution p définie par g.

Ici on vérifie bien que $g_n(s) \xrightarrow[n\to\infty]{} e^{-\lambda(1-s)} = e^{-\lambda}\sum_{k\in\mathbb{N}} \frac{(\lambda s)^k}{k!}$, qui est une série génératrice pour X telle que $P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$, c'est-à-dire $X \sim \mathrm{Poi}(\lambda)$.

Propriétés de la loi de Poisson. On se donne $X \sim \mathrm{Poi}(\lambda)$.

- $g_X(s) = e^{-\lambda(1-s)}$
- $\mathbb{E}[X] = g_X'(1) = \lambda$
- $\mathrm{Var}(X) = g_X''(1) + g_X'(1) - g_X'(1)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

Lemme. Soient $X_1 \sim \mathrm{Poi}(\lambda_1)$ et $X_2 \sim \mathrm{Poi}(\lambda_2)$, alors $X_1 + X_2 \sim \mathrm{Poi}(\lambda_1+\lambda_2)$. (il suffit de regarder les fonctions génératrices)

Modèle possonien et modèle réel : corrolaire. Si un évènement a une probabilité au plus p dans le modèle poissonien, alors il a une probabilité au plus $e\sqrt{m}$ dans le modèle réel.

Puissance de deux choix. On a n balles et n urnes. Pour chaque balle on tire au hasard uniformément d urnes et on place la bale dans l'une, la moins pleine. Après affectation des n balles (dans n urnes ?) la charge maximale d'une urne est au plus $\frac{\ln(\ln n)}{\ln d} + O(1)$ avec probabilité $1 - o\big(\frac{1}{n}\big)$. Preuve :

Propriété. $m \in \mathbb{N}, q \leqslant 1$. On a $P(\mathrm{Bin}(m,q) \geqslant 2\,m\,q) \leqslant e^{-mq/3}$. (borne de Chernoff, avec $\delta = 1$).

Ensuite, on pose :

- B_i = nombre d'urnes de charge au moins i
- $(\beta_i)_{i\geqslant 4}$ définie par $\beta_4 = \frac{n}{4}$ et $\beta_{i+1} = 2\frac{\beta_i^d}{n^{d-1}}$
- $p_i = \frac{\beta_i^d}{n^d}$

- $\mathcal{E}_i ='' B_i \leqslant \beta_i$
- $i^* = \min\left\{i \in \mathbb{N} | p_i < \frac{6 \ln n}{n}\right\}$

Lemme. $i^* = \frac{\ln(\ln n)}{\ln d} + O(1)$

2014-10-22

La méthode probabiliste. But : prouver l'existence d'objets satisfaisant une propriété P.

Argument de comptage. On a une collection d'objets (au plus dénombrable) $(a_i)_{i \in I}$ et P une propriété, on veut montrer qu'il existe un i tel que $P(a_i)$. Idée : définir X une va sur les (a_i), et montrer que $\mathbb{P}(X \text{ satisfait } P) > 0$.

Méthode du premier moment. On utilise $\mathbb{P}(X \geqslant \mathbb{E}[X]) > 0$ et $\mathbb{P}(X \leqslant \mathbb{E}[X]) > 0$, ou bien $\mathbb{P}(X \neq 0) \leqslant \mathbb{E}[X]$ pour une VA sur $\mathbb{N}$. Exemples...

Méthode du second moment. Si X VA sur $\mathbb{N}$, alors $\mathbb{P}(X = 0) \leqslant \frac{\operatorname{Var} X}{\mathbb{E}[X]^2}$.

2014-11-12.

(voir d'abord: graphes d'Erdös-Renyi)

Graphe aléatoire et composante géante.

On se donne un graphe aléatoire avec $p = \frac{c}{n}$ la probabilité d'avoir une arête entre deux noeuds.

Si u est un sommet, C_u est la composante connexe de u. On note C_1, C_2, ... les composantes connexes du graphe de la plus grande à la plus petite. On note $|C|$ le nombre de sommets dans la composante C.

Théorème. Selon la valeur de c, ona les comportements suivants :

i. Régime sous critique : $c < 1$. Il existe a une constente (dépendant de c) telle que :

$$\lim_{n \to \infty} P(|C_1| \leqslant a \ln n) \;=\; 1$$

ii. Régimie critique : $c = 1$. Il existe $k > 0$ tel que $\forall a > 0$,

$$\lim_{n \to \infty} P\big(|C_1| \geqslant a\, n^{2/3}\big) \;\leqslant\; \frac{k}{a^2}$$

iii. Régime critique : $c > 1$. Soit p_e l'unique solution dans $[0, 1[$ de $x = e^{-c(1-x)}$. Il existe une constante a' telle que $\forall \delta > 0$,

$$\lim_{n \to \infty} P\left(\left|\frac{|C_1|}{n} - (1 - p_e)\right| \leqslant \delta \wedge |C_2| \leqslant a' \ln n\right) \;=\; 1$$

1. Application : modèle d'épidémie. (Reed-Frost). On a n individus, et à $t = 0$ on a un individu infecté. À l'étape t, un individu infecté peut infecter n'importe quel autre individu avec probabilité $p = \frac{c}{n}$ pour un certain c, indépendamment des autres infections. Si un individu est infecté à l'étape t, il est retiré de la population à $t + 1$ (immunisé). Quelle est l'espérance de la taille de la population qui a été infectée à une étape ?

- Si $c < 1$, on a $\mathbb{E}[|C|] \leqslant O(\ln n)$.

- Si $c>1$, soit le sommet original était dans C_1 la composante géante, soit pas.

  $P(|C| \approx (1-p_e)\,n) = 1-p_e$ et $P(|C| \leqslant a \ln n) = p_e$.

2. Analyse d'une composante connexe par un processus de branchement. On imite un parcours en largeur. Les sommets sont soit vivants (dans la file), soit neutres (pas encore découverts), soit morts (déjà traîtés).

$t=0$: u est vivant, les autres sont neutres

À l'étape t, on choisit w dans la file, on le retire de la file, on ajoute ses voisins neutres dans la file (deviennent vivant), et w devient mort. Quand la file est vide, les sommets de $C(u)$ correspondent aux sommets morts. On note $L(t)$, $N(t)$, $D(t)$ respectivement le nombre de sommets vivants, neutres, morts à l'étape t. On note $Z(t)$ le nombre de sommets réajoutés à la file à l'étape t.

$$\begin{aligned} D(t) &= t \\ L(t) &= L(t-1) + Z(t-1) \\ &= n - t - N(t) \\ N(t) &= n - t - L(t) \\ &= N(t-1) - Z(t) \\ Z(t) &\sim \mathrm{Bin}(N(t-1), p) \\ &\sim \mathrm{Bin}(n-t+1-L(t-1), p) \end{aligned}$$

Autre présentation des processus de branchement de Galton-Watson : Z sont des VA sur $\mathbb{N}$ iid.

- À l'étape $t=0$, on a juste la racine, numérotée 1.
- À l'étape 1, on a joute les fils de la racine, que l'on numérote 2 à $1+Z(1)$
- Etc... à l'étape t, on ajoute les Z_t fils du sommet t, numérotés de $2+\sum_{i=1}^{t-1} Z_i$ à $1+\sum_{i=1}^{t} Z_i$.

Si on note $Y(t)$ le nombre de noeuds de la rangée t (ie vivants à l'étape t), on a : $Y_0=1$ et pour $t>0$, $Y_t = Y_{t-1} + Z_t - 1$. Le processus s'arrête quand $Y_t=0$ pour la première fois (on peut définir Y_t même après l'arrêt du processus).

Comparaison entre processus de branchements. Soit $Z \sim \mathrm{Bin}(n,p)$. On note $T^{\mathrm{bin}}_{n,p}$ la taille du processus de branchement pour un arbre binomial de Galton-Watson, et $T^{\mathrm{gr}}_{n,p}$ la taille du processus de branchement pour une composante connexe.

Lemme. $P(T^{\mathrm{bin}}_{n-k,p} \geqslant k) \leqslant P(T^{\mathrm{gr}}_{n,p} \geqslant k) \leqslant P(T^{\mathrm{bin}}_{n,p} \geqslant k)$.

3. Régime sous-critique. Preuve : on fait des calculs à partir de ce lemme. $k = a \ln n$, $a = \frac{4c}{(1-c)^2}$, .;;

4. Régume sur-critique. Il y a trois étapes pour la preuve :

- Montrer qu'on a des petites et des grandes composantes
- Montrer qu'on a une seule grande composante
- Montrer que la taille de cette composante est $\sim (1-p_e)\,n$.

4.a. Il y a des petites et des grandes composantes uniquement. On note $k^- = a' \ln n$, et $k^+ = n^{2/3}$.

Lemme. Pour tout sommet v, avec forte proba, soit i) le processus de branchement depuis v s'arrête en moins de k^- étapes, soit ii) $\forall k$ entre k^- et k^+, il y a au moins $(k-1)k/2$ sommets vivants. Un mauvais sommet ne satisfait aucune de ces deux propriétés.

2014-11-19.

Chaines de Markov. Introduites en 1906, appliquées en 1913 à l'étude de l'alternance consonne-voyelle. Applications courantes : physique statistique, théorie de l'information et compression, réseaux de communication, bio-informatique, combinatoire, ...

Exemple : atelier de réparation. Au jour n, Z_{n+1} machines tombent en panne. Au jour $n+1$, les machines en panne arrivent à l'atelier, et une machine déjà présente est réparée. X_n = nombre de machines à l'atelire le jour n.

$$X_{n+1} \;=\; \max(X_n - 1, 0) + Z_{n+1}$$

Autre exemple qui donne la même chose : une file d'attente. (se résoud avec des séries génératrices).

Trois représentations. Matricielle, fonction, graphique (zoli dessin). Ex: faire un zoli dessin pour une marche aléatoire de proba p dans une dimension : ça fait une ligne.

Graphes. (poly p. 17) Définitions:

- j accesible depuis i si il existe un chemin de i vers j (rappel : les arêtes présentes sont de probabilité strictement positives, tout est cohérent) ; on note cela $i \rightsquigarrow j$.
- i et j communiquent si $i \rightsquigarrow j$ et $j \rightsquigarrow i$ (convention : i communique avec lui-même).
- on définit des classes de communication, qui correspondent aux composantes connexes du graphe.
- si une CMH a une seule classe de communication, elle est dite irréductible.

Mesure invariante. Soit μ une mesure non nulle $(\mu(i), i \in E)$, μ est une mesure invariante si $\mu = \mu P$, $\forall i \geqslant 0, \mu(i) \geqslant 0$ et $\forall i \geqslant 0, \mu(i) = \sum_{j \in E} \mu(j)\, P_{j,i}$.

Distribution stationnaire. Soit π une distribution de probabilités. π est stationnaire si $\pi P = \pi$.

Exemple de CM qui:

1. N'a pas de mesure invariante : une chaine infinie sur laquelle on ne peut qu'avancer (avec probabilité 1).
2. A une mesure invariante mais pas de distribution stationnaire : ?
3. Qui a plusieurs distributions stationnaires : l'identité (toute distribution est stationnaire...)

Exemple. Prendre :

$$P \;=\; \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

Distribution stationnaire : $\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$.

Comment calculer une distribution stationnaire.

- Résoudre un système linéaire à beaucoup trop de variables
- Utiliser les fonctions génératrices (ex: atelier de réparation)

- Chaines réversibles : théorème important : une distribution strictement positive π est stationnaire ssi la matrice Q inversant le temps par π ($\pi(i)p_{ij} = \pi(j)q_{ji}$) est stochastique. La chaine de Markov est dite réversible si $q_{ij} = p_{ij}$.

Équations d'équilibre global. On partitionne le système en deux ; dans une distribution stationnaire ce qui va de A dans B est la même quantité que ce qui va de B dans A (pour faire court).

2014-11-26

II.[3] Classification des états

Rappel: mesure invariante : $\nu = \nu P$; probabilité stationnaire : $\pi = \pi P$

1 Propriété de Markov fort, temps d'arrêt

$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = P(X_{n+1} = j | X_n = i)$

1.1 Temps d'arrêt

Soient τ une VA sur $\mathbb{N} \cup \{+\infty\}$, $(X_n)_{n\in\mathbb{N}}$ un processus stochastique sur E au plus dénombrable. On dit que τ est un temps d'arrêt pour $\{X_n\}_{n\geqslant 0}$ si $\forall m \in \mathbb{N}$, $\{\tau = m\}$ est une fonction de $X_0, ...,$ X_m et pas de X_{m+1}, et si $\forall m \in \mathbb{N}, \exists B \subseteq E^{m+1}$ tel que $\{\tau = m\} = \{\omega | (X_0(\omega), ..., X_m(\omega)) \in B\}$

Exemples.

- Temps d'attente de $i \in E$: on note $S_i = \min\{n \geqslant 0 | X_n = i\}$.

  $\{S_i = m\} = \{X_0 \neq i, ..., X_{m-1} \neq i, X_m = i\}$

- Temps de retour en $i \in E$: $T_i = \min\{n \geqslant 1 | X_n = i\}$ (différent pour le cas $X_0 = i$)

- Temps de retour successifs en 0 : $0 \in E$. $\tau_1 = T_0$, $\tau_{k+1} = \inf\{n \geqslant \tau_k + 1 | X_n = 0\}$.

  $\{\tau_k = m\} = \left\{\sum_{n=1}^{m-1} \mathbf{1}_{\{X_n=0\}} = k - 1 \wedge X_m = 0\right\}$

- Dernier passage en i : pas un temps d'arrêt, car il faut regarder toutes les valeurs prises par la suite.

Processus avant le temps d'arrêt : $\{X_{n\wedge\tau}\}_{n\geqslant 0}$; processus après le temps d'arrêt : $\{X_{n+\tau}\}_{n\geqslant 0}$.

1.2 Propriété de Markov fort

Théorème. Soit $\{X_n\}_{n\geqslant 0}$ une CMH sur E de matrice de transition P, et τ un temps d'arrêt pour $\{X_n\}_{n\geqslant 0}$ presque-sûrement fini. Alors $\forall i \in E$,

- conditionellement à $\{X_\tau = i\}$ les processus $\{X_{n+\tau}\}_{n\geqslant 0}$ et $\{X_{n\wedge\tau}\} \geqslant 0$ sont indépendants
- conditionnellement à $\{X_\tau = i\}$, $\{X_{n+\tau}\}_{n\geqslant 0}$ est une CMH de matrice de transition P.

Dém. Pas de notes. (à un moment on somme sur toutes les possibilités pour τ)

Notation. On note $P_i(A) = P(A | X_0 = i)$ et $\mathbb{E}_i[A] = \mathbb{E}[A | X_0 = i]$.

3. la numérotation ne fait aucun sens et est arbitraire

2 États récurrents, états transitoires

Temps de retour successif en $0 \in E$: supposons $P_0(T_0 < \infty) = 1$. On a alors : $P_0(\tau_2 < \infty) = 1$, et $\forall k$, $P_0(\tau_k < \infty) = 1$, et $(\tau_k - \tau_{k-1})_{k \geqslant 2}$ est une suite de VA IID (par Markov fort).

Notons $f_{j,i} = P_j(T_i < \infty)$ et $N_i = \sum_{n \geqslant 1} \mathbf{1}_{\{X_n = i\}}$

Théorème. $\forall r \geqslant 1, P_j(N_i = r) = f_{j,i}\, f_{i,i}^{r-1}(1 - f_{i,i})$ et $P_j(N_i = 0) = 1 - f_{j,i}$

Dém. Par récurrence montrer que $P_j(N_i \geqslant r+1) = f_{j,i}\, f_{i,i}^{r}$.

On a $P_i(N_i = r) = f_{i,i}^{r}(1 - f_{i,i})$.

Classification.

État i récurrent	$P_i(T_i < \infty) = 1$	$\mathbb{E}_i[N_i] = \infty$	$P_i(N_i = \infty) = 1$
État i transitoire	$P_i(T_i < \infty) < 1$	$\mathbb{E}_i[N_i] < \infty$	$P_i(N_i < \infty) = 1$

Déf.

- Un état i est dit récurrent si $P_i(T_i < \infty) = 1$, et transitoire si $P_i(T_i < \infty) < 1$
- Un état récurrent est dit récurrent nul si $\mathbb{E}_i[T_i] = \infty$, et récurrent positif si $\mathbb{E}_i[T_i] < \infty$

Critère de la matrice potentiel. On définit $G = \sum_{n \in \mathbb{N}} P^n$ (P est la matrice de transition).

$g_{i,j} = \sum_{n \in \mathbb{N}} p_{i,j}(n) = \sum_{n=0}^{\infty} P_i(X_n = j) = \sum_{n=0}^{\infty} \mathbb{E}_i(\mathbf{1}_{\{X_n = j\}}) = \mathbb{E}_i(\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = j\}}) = \mathbb{E}_i(N_j)$.

Prop. $g_{i,i} = \infty \Leftrightarrow i$ récurrent

Exemple. Marche aléatoire sur $\mathbb{Z}$: $\forall i \in \mathbb{Z}$, $p_{i,i+1} = p$ et $p_{i,i-1} = 1 - p$.

$p_{0,0}(2n+1) = 0$ et $p_{0,0}(2n) = \binom{2n}{n} p^n (1-p)^n \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$

Si $p = \frac{1}{2}$: $\sum \frac{1}{\sqrt{\pi n}}$ diverge, donc 0 est récurrent

Si $p \neq \frac{1}{2}$, $\alpha = 4p(1-p) < 1$, donc $\sum \frac{\alpha^n}{\sqrt{\pi n}}$ converge et 0 est transitoire.

Théorème. Si i et j communiquent, alors ils sont tous les deux récurrents ou tous les deux transitoires.

Dém. $\sum p_{i,i}(n)$ converge $\Leftrightarrow \sum p_{j,j}(n)$ converge (écrire les relations avec les communications).

3 Existence et unicité d'une mesure invariante

Théorème. Pour toute CMH $\{X_n\}_{n \geqslant 0}$ irréductible et récurrente possède une unique mesure invariante, à facteur multiplicatif près.

Si on note $(x_i)_{i \in E}$ une telle mesure :

- $\forall i$, $0 < x_i < \infty$ et $x_i = \mathbb{E}_0\big[\sum_{n=1}^{T_0} \mathbf{1}_{\{X_n = i\}}\big]$

  ie : $x_0 = 1$ et $x_i =$ nombre moyen de passages en i entre deux passages en 0
- $\sum_{i \in E} x_i = \sum_{i \in E} \mathbb{E}_0\big[\sum_{n=1}^{T_0} \mathbf{1}_{\{X_n = i\}}\big] = \mathbb{E}_0\big[\sum_{n=1}^{T_0} 1\big] = \mathbb{E}_0[T_0]$

Dém.

Notons P la matrice de transition de $\{X_n\}_{n \in \mathbb{N}}$. On suppose qu'il existe ν mesure invariante telle que $\exists i$ tq $0 < \nu(i) < \infty$. Alors $\forall j, 0 < \nu(j) < \infty$. En effet, i et j communiquent (l'écrire, puis calculs).

Notons $x_i = \mathbb{E}_0\big[\sum_{n=1}^{T_0} \mathbf{1}_{\{X_n=i\}}\big]$ et montrons que c'est une mesure invariante :

$$\begin{aligned} x_i &= \mathbb{E}_0\left[\sum_{n=1}^{T_0} \mathbf{1}_{\{X_n=i\}}\right] \\ &= \sum_{n\geqslant 1} \mathbb{E}_0\big[\mathbf{1}_{X_n=i}\mathbf{1}_{n\leqslant T_0}\big] \\ &= \sum_{n\geqslant 1} P_0(X_n=i, n\leqslant T_0) \\ &= \sum_{n\geqslant 1} \underbrace{P_0(X_1\neq 0, ..., X_{n-1}\neq 0, X_n=i)}_{{}_0p_{0,i}(n)} \end{aligned}$$

${}_0p_{0,i}(1) = p_{0,i}$ et $\forall n \geqslant 1, {}_0p_{0,i}(n+1) = \sum_{j\neq 0} {}_0p_{0,j}(n)\, p_{j,i}$

$$\begin{aligned} x_i &= {}_0p_{0,i}(1) + \sum_{n\geqslant 2} {}_0p_{0,i}(n) \\ &= x_0\, p_{0,i} + \sum_{n\geqslant 1}\sum_{j\neq 0} {}_0p_{0,j}(n)p_{j,i} \\ &= x_0 p_{0,i} + \sum_{j\neq 0}\sum_{n\geqslant 1} {}_0p_{0,j}(n)\, p_{j,i} \\ &= \sum_{j\in E} x_j p_{j,i} \end{aligned}$$

Unicité. On prend y mesure invariante avec $y_0 = 1$. On veut $x = y$.

Processus retourné : on pose $Q = (q_{i,j})$ avec $q_{j,i} = \frac{y_i}{y_j} p_{i,j}$. Propriétés de Q :

- Q est une matrice stochastique
- $\forall n, q_{j,i}(n) = \frac{y_i}{y_j}\, p_{i,j}(n)$
- Q est irréductible
- Q estréurrente $\Leftrightarrow$ P est récurrente (cf matrice potentiel)

On note X' le processus retourné, et on note $g_{i,0}(n) = P(X'_0 = i, X'_1 \neq 0, ..., X'_{n-1} \neq 0, X'_n = 0)$

$$\begin{aligned} g_{i,0}(1) &= g_{i,0} \\ y_i\, g_{i,0}(n+1) &= \sum_{j\neq 0} (y_i q_{i,j}) g_{j,0}(n) \\ &= \sum_{j\neq 0} (y_j p_{j,i}) g_{j0}(n) \\ &= \sum_{j\neq 0} (y_j q_{j,0}(n)) p_{j,i} \\ y_i\, g_{i,0}(1) &= y_i q_{i,0} \\ &= y_0 p_{0,i} \\ &= p_{0,i} \end{aligned}$$

D'où :

$$\begin{aligned} \forall n, {}_0p_{0,i}(n) &= y_i g_{i,0}(n) \\ \sum_{n\geqslant 1} y_i g_{i,0}(n) &= y_i \underbrace{\sum_{n\geqslant 1} g_{i,0}(n)}_{=1} \\ &= y_i \end{aligned}$$

Nous avons donc une condition suffisante pour l'existence d'une mesure invariante (attention : elle n'est pas nécessaire ! Des chaines transitoires peuvent avoir des mesures invariantes.)

Théorème. Une CMH irréductible et récurrente est récurrente positive ssi elle possède une distribution stationnaire.

En effet, $\sum x_i = \mathbb{E}_0[T_0] < \infty \Leftrightarrow$ la chaine est récurrente positive.

4 Existence d'une distribution stationnaire

Théorème. Une CMH irréductible est récurrente positive ssi elle possède une distribution stationnaire.

Preuve. Seul cas à considérer : chaîne transitoire avec π invariante. Supposons qu'on a ça. $\pi = \pi P = \pi P^n$. $\pi_i = \sum_j \pi_j p_{j,i}(n)$. $\sum_{n \geqslant 0} p_{j,i}(n)$ converge, donc $p_{j,i}(n) \xrightarrow[n\to\infty]{} 0$.

$\pi_i = \lim_{n\to\infty} \sum_j \pi_j p_{j,i}(n) = \sum_j \pi(j) \lim_{n\to\infty} p_{j,i}(n) = 0$. Donc π est la distrubution nulle, etc etc.

Référence bibliographique. Se trouve sur internet : Haggström, *Finite Markov Chains*, 2001 (titre inexact, contient également *algorithms*).

Récap. Si $\{X_n\}$ CMH irréductible :

- chaine récurrente $\Rightarrow \mu(i) = \mathbb{E}_0\big(\sum_{n=1}^{T_0} \mathbf{1}_{\{X_n = i\}}\big)$
- chaine récurrente positive $\Leftrightarrow \pi(i) = \frac{1}{\mathbb{E}_i(T_i)}$

2014-12-03.

III. Comportement asymptotique et ergodicité

Théorème. Soit $\{X_n\}$ une CMH irréductible et récurrente sur E, soit $f : E \to \mathbb{R}$ telle que :

$$\sum_{i \in E} |f(i)| \mu(i) \ < \ \infty$$

Alors : presque-sûrement et pour toute loi initiale,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{\nu(n)} \sum_{k=1}^{n} f(X_k) \ &= \ \sum_{i \in E} f(i)\mu(i) \\ \text{où} \quad \nu(n) \ &= \ \sum_{k=1}^{n} \mathbf{1}_{\{X_k = 0\}} \end{aligned}$$

Comprendre : « moyenne temporelle = moyenne spatiale ».

Corollaire. Si la chaine $\{X_n\}$ est récurrente positive et irréductible, alors, pour toute loi initiale et presque surement :

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) \ = \ \sum_{i \in E} f(i)\pi(i)$$

Où π est la distribution stationnaire.

Dém.

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \;=\; \underbrace{\left(\frac{1}{\nu(n)}\sum_{k=1}^{n} f(X_k)\right)}_{\xrightarrow[n\to\infty]{}\sum f(i)\mu(i)}\frac{\nu(n)}{n}$$

On applique le théorème pour $f=1$, on a bien la condition $\sum_{i\in E}\mu(i)=\mathbb{E}_0[T_0]<\infty$, ce qui donne :

$$\begin{aligned}\frac{n}{\nu(n)} \quad &= \quad \frac{1}{\nu(n)}\sum_{i=1}^{n} 1 \\ &\xrightarrow[n\to\infty]{} \sum \mu(i) = \mathbb{E}_0[T_0]\end{aligned}$$

Donc :

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n} f(X_k) \;&\to\; \frac{\sum_{i\in E} f(i)\mu(i)}{\sum_{i\in E}\mu(i)} \\ &\to\; \sum_{i\in E} f(i)\pi(i)\end{aligned}$$

Théorème. (loi forte des grands nombres) Soit $\{U_n\}_{n\geqslant 1}$ IID telle que $\mathbb{E}[|U_1|]<\infty$, et soit $S_n=\sum_{k=1}^{n} U_k$. Alors presque-sûrement :

$$\lim_{n\to\infty}\frac{S_n}{n} \;=\; \mathbb{E}[U_1]$$

Dém. On admet la loi forte des grands nombres (cf cours d'intégration), et on prouve le théorème ergodique.

On suppose d'abord $f\geqslant 0$. On note $\tau_1,\tau_2,...,\tau_p$ les temps de retours en 0. On note :

$$Y_p \;=\; \sum_{k=\tau_p+1}^{\tau_{p+1}} f(X_k)$$

Les (Y_p) sont IID : on peut appliquer la loi des grands nombres.

$$\begin{aligned}\mathbb{E}[Y_i] \;&=\; \mathbb{E}\left[\sum_{k=\tau_1+1}^{\tau_2} f(X_k)\right] \\ &=\; \mathbb{E}_0\left[\sum_{k=1}^{\tau_0} f(X_k)\right] \\ &=\; \mathbb{E}_0\left[\sum_{k=1}^{T_0}\left(\sum_{i\in E} f(i)\mathbf{1}_{\{X_k=i\}}\right)\right] \\ &=\; \sum_{i\in E} f(i)\,\mathbb{E}_0\left[\sum_{k=1}^{T_0}\mathbf{1}_{\{X_k=i\}}\right] \\ &=\; \sum_{i\in E} f(i)\,\mu(i)\end{aligned}$$

Presque-sûrement par la loi forte des grands nombres :

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{p=1}^{n} Y_p \;&=\; \mathbb{E}[Y_i] \\ &=\; \sum_{i\in E} f(i)\mu(i)\end{aligned}$$

On a l'encadrement :

$$\tau_{\nu(n)} \leqslant n < \tau_{\nu(n)+1}$$

D'où :

$$\sum_{k=1}^{\tau_{\nu(n)}} f(X_k) \leqslant \sum_{k=1}^{n} f(X_k) \leqslant \sum_{k=1}^{\tau_{\nu(n)+1}} f(X_k)$$

$$\underbrace{\sum_{k=1}^{T_0} f(X_k)}_{\text{p.s. fini}} + \sum_{p=1}^{\nu(n)} Y_p \leqslant \sum_{k=1}^{n} f(X_k) \leqslant \sum_{k=1}^{T_0} f(X_k) + \sum_{p=1}^{\nu(n)+1} Y_p$$

$$\underbrace{\frac{1}{\nu(n)}\left(\sum_{k=1}^{T_0} f(X_k) + \sum_{p=1}^{\nu(n)} Y_p\right)}_{\xrightarrow[n\to\infty]{} \sum_{i\in E} f(i)\,\mu(i)} \leqslant \frac{1}{\nu(n)} \sum_{k=1}^{n} f(X_k) \leqslant \underbrace{\frac{1}{\nu(n)}\left(\sum_{k=1}^{T_0} f(X_k) + \sum_{p=1}^{\nu(n)+1} Y_p\right)}_{\text{idem}}$$

Pour f quelconque, on utilise la décomposition $f = f_+ + f_-$.

Application. Automate de comptage. $\Sigma = \{0, 1\}$. Chaque lettre est 0 avec proba. $1/2$, 1 avec proba $1/2$, et indépendamment du reste du mot. Quelle est la fréquence d'apparition du motif 010 (sans compter les chevauchements) ? On définit l'automate à quatre états qui reconait ce motif, et on compte les passages par l'état acceptant. On nomme les états $\{\varepsilon, 0, 01, 010\}$, l'état acceptant est 010. Matrice de transition :

$$P = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

On a :

$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k=010\}} \xrightarrow[n\to\infty]{} \pi_{010}$$

On cherche $\pi = \pi P$. Solution:

$$\begin{aligned} \pi &= \alpha(3, 4, 2, 1) \\ &= \left(\frac{3}{10}, \frac{4}{10}, \frac{2}{10}, \frac{1}{10}\right) \end{aligned}$$

La solution est donc : $\pi_{010} = 1/10$.

Théorème de Kolmogorov. Si $\{X_n\}$ est une CMH irréductible, récurrente positive et apériodique, aors :

$$\forall i, j \in E, \quad \lim_{n\to\infty} p_{i,\,j}(n) = \pi(j)$$

où π est l'unique distribution stationnaire de (X_n).

Déf. On appelle *ergodique* une CMH irréductible, récurrente positive et apériodique.

(cf TD pour la preuve)

IV. Simulation de chaînes de Markov

1 Simulation de variables aléatoires et d'une quantité

Estimation de π. (Méthode de Monte-Carlo). On trace un cercle dans un carré de côté 2 : l'aire du cercle est π. On fait tomber des points uniformément dans le carré ; la probabilité de tomber dans le cercle est proportionnelle à l'aire du cercle : c'est $\pi/4$.

Algorithme

$W \leftarrow 0$
pour $i = 1...m$
  Tirer $X_i, Y_i \sim \text{Unif}[0,1]$ indépendants
  Si $X_i^2 + Y_i^2 \leqslant 1$, alors $W \leftarrow W + 1$
renvoyer $4\,W/m$

On a bien $\mathbb{E}[W] = \frac{m\pi}{4}$

Borne de Chernoff :

$$P\Big(\Big|W - \frac{m\,\pi}{4}\Big| \geqslant \varepsilon \frac{m\,\pi}{4}\Big) \;\leqslant\; 2\exp\Big(-\frac{m\,\pi\,\varepsilon^2}{12}\Big)$$

Déf. Un algorithme probabiliste renvoit une (ε, δ)-approximation pour la valeur v si la sortie X de l'algorithme satisfait :

$$P(|X - v| \leqslant \varepsilon\, v) \;\geqslant\; 1 - \delta$$

Dans notre cas, $2\exp\Big(-\frac{m\,\pi\,\varepsilon^2}{12}\Big) \leqslant \delta \Leftrightarrow \ln 2 - \frac{m\,\pi\,\varepsilon^2}{12} \leqslant \ln\delta \Leftrightarrow m \geqslant \frac{12}{\pi\,\varepsilon^2} \ln\frac{2}{\delta}$

Déf. On a un schéma d'approximation stochastique polynomial (FPRAS : fully polynomial random approximation scheme) pour un problème si pour tout paramètre x du problème, et $\forall \delta, \varepsilon > 0$, l'algorithme renvoie une (ε, δ)-approximation de x en temps polynomial en la taille de x, $\frac{1}{\varepsilon}$ et $\ln\frac{1}{\delta}$.

On verra d'autres schémas d'approximation plus tard.

2 Échantillonnage selon une distribution discrète

Soit π une distribution sur $\mathbb{N}$ ou $\{0, ..., n\}$. On cherche à fabriquer un algorithme qui renvoie i avec probabilité $\pi(i)$.

Méthode 1. Tirer $X \sim$ Unif[0, 1], et prendre $i = \max\Big\{k | \sum_{p=0}^{k-1} \pi(p) \leqslant X\Big\}$. Cela peut être implémenté avec une boucle.

Méthode 2. Construire un arbre binaire proba-équilibré pour π (ça ressemble beaucoup à un arbre de Huffmann).

Technique de l'alias. Dans un espace fini de taille n : on divise notre espace de proba en n cases, chacune divisée au maximum en deux lors d'un pré-traitement. On tire uniformément une case ; dans chaque case on tire uniformément entre $[0, 1]$ un point qui tombe d'un côté où de l'autre, ce qui donne le résultat.

Algorithme

Pré-traîtement :
  Poser :
    $q_i = n\,\pi(i)$,
    $s_j = 0$,

$G = \{i | q_i \geqslant 1\}$,
$H = \{i | q_i < 1\}$.
Tant que $H \neq 0$:
Soit $j \in H$ et $k \in H$
$s_j \leftarrow 1 - q_j$
$q_k \leftarrow q_k + q_j - 1$
$H \leftarrow H \setminus \{j\}$
Si $q_k < 1$ alors :
$G \leftarrow G - k$
$H \leftarrow H + k$
Algo de tirage :
Tirer $X \sim \text{Unif}\{1, ..., n\}$, puis $Y \sim \text{Unif}[0, 1]$.

3 Méthodes MCMC[6], échantillonneurs de Gibbs

Exemple. Indépendants d'un graphe. Soit $G = (S, A)$ un graphe non orienté, $S = \{1, ..., n\}$ et $A \subseteq \mathcal{P}_2(S)$. $I \subseteq S$ est un indépendant si $\forall u, v \in I, \{u, v\} \notin A$.

But : générer un indépendant de façon uniforme, ie selon la loi de proba $\pi(I) = \frac{1}{Z}$, où Z est le nombre d'indépendants.

Méthode naïve. On échantillonne $X \subseteq S$ de manière uniforme. Si X est un indépendant on retourne X, sinon on recommence.

Chaine de Markov. On construit une CM dont la distribution stationnaire est la distribution uniforme sur les indépendants. On simule ensuite cette CM sur un nombre n suffisamment grand d'étapes. L'échantillon X_n est alors presque uniforme.

3.1 Échantillonneur de Gibs

Distribution à forme produit. L'espace d'états est de la forme $E \subseteq Q^k$, π_i est une distribution sur Q pour $1 \leqslant i \leqslant k$. On a une distrubtion de la forme :

$$\pi(q_1, ..., q_k) = \frac{1}{Z}\pi_1(q_1)\pi_2(q_2)\cdots\pi_k(q_k)$$

Ex. Ensembles indépendants : $G = (S, A)$ un graphe non orienté, $k = |S|$, $Q = \{0, 1\}$. $E =$ l'ensemble des indépendats du graphe. Un état $(q_1, ..., q_k)$ correspondant à un indépendant I a $q_s = 1$ si $s \in I$, et $q_s = 0$ sinon. On prend $\pi_s(q_s) = \frac{1}{2}$.

$$\begin{aligned}\pi(q_1, ..., q_k) &= \frac{1}{Z}\left(\frac{1}{2}\right)^k \\ &= \pi(I)\end{aligned}$$

C'est-à-dire que les indépendants sont tous de probabilité égale.

Pour favoriser les indépendants de grande taille, on voudrait une probabilité de la forme :

$$\pi(I) = \frac{\lambda^{|I|}}{Z}$$

Pour cela, prendre :

$$\pi_s(q_s) = \frac{\lambda^{q_s}}{1 + \lambda}$$

6. **Monte Carlo Markov Chains**

Échantillonneur de Gibbs. On se donne $X_n = (q_1, ..., q_k)$, et on veut produire le résultat X_{n+1} tel que la chaine de Markov $\{X_n\}$ ait bien comme distribution stationnaire une distribution π de la forme du paragraphe précédent. Idée : ne transitionner qu'entre des états dont seule une composante varie à chaque pas.

- Tirer $u \in \{1, ..., k\}$ selon une distribution d
- Tirer $q \in Q$ selon la distribution π_u (indépendamment de d)
- Si $x = (q_1, ..., q_{u-1}, q, q_{u+1}, ..., q_k) \in E$, alors $X_{n+1} = x$. Sinon on garde $X_{n+1} = X_n$.

Il faut vérifier que la distribution stationnaire de cet algorithme correspond bien à π.

Soient $x = (q_1, ..., q_k), y = (q_1, ..., q_{u-1}, q, q_{u+1}, ..., q_k)$ qui diffèrent sur exactement une composante.

$$\begin{aligned}
\pi(x)\, p_{x,y} &= \frac{1}{Z} \prod_{i=1}^{k} \pi_i(q_i)\, p_{x,y} \\
&= \left(\frac{1}{Z} \prod_{i \neq u} \pi_i(q_i) \right) \pi_u(q_u)\, p_{x,y} \\
&= \left(\frac{1}{Z} \prod_{i \neq u} \pi_i(q_i) \right) \pi_u(q_u)\, d(u) \pi_u(q) \\
&= \pi(y)\, p_{y,x}
\end{aligned}$$

3.2 Échantiloneur pour les indépendants d'un graphe

- Tirer u uniformément
- Si pour tout voisin v de u, $v \notin I$, alors avec proba $1/2$ on a $I' = I \cup \{u\}$, et avec proba $1/2$, on a $I' = I \setminus \{u\}$.
- Si il existe un voisin v de u tel que $v \notin I$, on a $I' = I$.

Montrons que cette chaine est ergodique, ie apériodique irréductible récurrente positive (sur un eespace d'états fini le caractère récurrent positif découle du caractère apériodique irréductible).

- Pour chaque état I, $p_{I,I} > 0$, donc la chaine est bien apériodique
- De $\varnothing$ on peut atteindre n'importe quel indépendant I, et réciproquement, donc la chaine est irréductible.

3.3 Simulation MCMC pour des chaines ergodiques

On part d'une représentation fonctionnelle de $\{X_n\}$ une CMH : $X_{n+1} = f(X_n, U_{n+1})$, où $\{U_i\}$ IID et indépendantes avec X_0.

Principe : on part de $x \in E$ arbitraire, puis on calcule $X_1 = f(x, U_1)$, $X_2 = f(X_1, U_2)$, et ainsi de suite jusqu'à $X_r = f(X_{r-1}, U_r)$. X_r est un échantillon de la CMH. Pour r assez grand, X_r est un bon échantillon de notre distribution stationnaire. Si on veut plusieurs échantillons, on peut prendre X_r, $X_{2r}, X_{3r}, ...$ Puisque r est assez grand on peut considérer que X_r et X_{2r} sont presque indépendants.

Par exemple sur un graphe, si on veut estimer la taille moyenne d'un indépendant :

$$\begin{aligned}
\sum_{I \text{ independant}} |I| \left(\mathbb{P}(X_r = I) - \pi(I) \right) &\leqslant \sum_{I} |I| \left| \mathbb{P}(X_r = I) - \pi(I) \right| \\
&\leqslant |S| \sum_{I} \left| \mathbb{P}(X_r = I) - \pi(I) \right|
\end{aligned}$$

4 Distance de variation et couplage de chaînes de Markov

4.1 Distance de variation

Déf. Soient μ_1, μ_2 deux distributions de proba sur E. On définit la distance de variation totale :

$$\|\mu_1 - \mu_2\|_{\mathrm{DV}} = \frac{1}{2}\sum_{x\in E} |\mu_1(x) - \mu_2(x)|$$

Remarques.

- $0 \leqslant \|\mu_1 - \mu_2\|_{\mathrm{DV}} \leqslant 1$.
- $\|\mu_1 - \mu_2\|_{\mathrm{DV}} = 0 \Leftrightarrow \mu_1 = \mu_2$
- $\|\mu_1 - \mu_2\|_{\mathrm{DV}} = 1 \Leftrightarrow \mu_1$ et μ_2 ont des supports disjoints.

Définition équivalente.

$$\|\mu_1 - \mu_2\|_{\mathrm{DV}} = \max_{A\subseteq E} |\mu_1(A) - \mu_2(A)|$$

Démonstration.

$$\begin{aligned}
E^+ &= \{x \in E | \mu_1(x) \geqslant \mu_2(x)\} \\
\mu_1(E^+) - \mu_2(E^+) &= \sum_{x\in E^+} \mu_1(x) - \mu_2(x) \\
E^- &= \{x \in E | \mu_1(x) < \mu_2(x)\} \\
\mu_2(E^-) - \mu_1(E^-) &= \sum_{x\in E^-} \mu_2(x) - \mu_1(x) \\
\mu_1(E^+) + \mu_1(E^-) &= \mu_2(E^+) + \mu_2(E^-) \\
&= 1 \\
\mu_1(E^+) - \mu_2(E^+) &= \mu_2(E^-) - \mu_1(E^-) \\
\frac{1}{2}\sum_x |\mu_1(x) - \mu_2(x)| &= \sum_{x\in E^+} \mu_1(x) - \mu_2(x) + \sum_{x\in E^-} \mu_2(x) - \mu_1(x) \\
&= |\mu_1(E^+) - \mu_2(E^+)|
\end{aligned}$$

Exemple. Mélange de cartes. On a un jeu de 52 cartes. À chaque étape on choisit une carte uniformément et on la place au sommet du paquet. Distribution stationnaire ? Chaque état a 52 prédécesseurs (lui compris). La distribution stationnaire est la distribution uniforme.

On pose maintenant : μ_1 la distribution uniforme, μ_2 la distribution uniforme parmi les paquets qui ont l'as de pique au sommet. Notons A cet ensemble de paquets. $\mu_2(A) = 1$, et $\mu_1(A) = \frac{1}{52}$.

$$\begin{aligned}
\|\mu_1 - \mu_2\|_{\mathrm{DV}} &= 1 - \frac{1}{52} \\
&= \frac{51}{52}
\end{aligned}$$

4.2 Temps de mélange

Soit π la distribution stationnaire de $\{X_n\}$ une CMH ergodique. On note p_x^n la distribution de (X_n) après n pas de calcul, partant de l'état $X_0 = x$. $(p_x^n(y) = p(n)_{x,y})$

$$\begin{aligned}
\Delta_x(n) &= \|p_x^n - \pi\|_{\mathrm{DV}} \\
\Delta(n) &= \max_{x\in E} \Delta_x(n) \\
\tau_x(\varepsilon) &= \min\{n | \Delta_x(n) \leqslant \varepsilon\} \\
\tau(\varepsilon) &= \max_{x\in E} \tau_x(\varepsilon)
\end{aligned}$$

τ est appellé *temps de mélange*.

On dit qu'une CMH est rapidemant mélangeante si $\tau(\varepsilon)$ estp olynomial en la taille du problème et en $\ln \frac{1}{\varepsilon}$.

4.3 Couplage d'une CM

Déf. Un couplage d'une CMH $\{M_n\}_{n\in\mathbb{N}}$ sur E est une chaine $Z_n = (X_n, Y_n)$ sur $E \times E$ telle que:

$$\begin{aligned} \mathbb{P}(X_{n+1} = x' \,|\, Z_n = (x,y)) &= \mathbb{P}(M_{n+1} = x' | M_n = x) \\ \mathbb{P}(Y_{n+1} = y' | Y_n = y) &= \mathbb{P}(M_{n+1} = y' | M_n = y) \end{aligned}$$

C'est-à-dire que $\{X_n\}$ et $\{Y_n\}$ se comportent comme $\{M_n\}$, mais ne sont pas nécessairement indépendantes.

On s'intéresse aux couplages tels que :

- $\exists n \,|\, X_n = Y_n$
- Si $X_n = Y_n$, alors $X_{n+1} = Y_{n+1}$.

Idée. $M_{n+1} = f(M_n, U_{n+1})$, avec (U_i) IID indépendantes avec M_0. On choisit le même U_n pour X_n et Y_n.

Théorème. Soit $Z_n = (X_n, Y_n)$ un couplage d'une CM sur E. Supposons qu'il existe T tel que $\forall x$, $y \in E$, $\mathbb{P}(X_T \neq Y_T | X_0 = x, Y_0 = y) \leqslant \varepsilon$. Alors $\tau(\varepsilon) \leqslant T$.

Dém.

$Y_0 = \pi$, $X_0 = x$.

$$\begin{aligned} \mathbb{P}(Y_T \neq X_T) &= \sum_{y\in E} \mathbb{P}(X_T \neq Y_T | X_0 = x, Y_0 = y)\, \mathbb{P}(Y_0 = y) \\ &\leqslant \sum_{y\in E} \varepsilon\, \mathbb{P}(Y_0 = y) \\ &\leqslant \varepsilon \end{aligned}$$

Soit $A \subseteq E$. On veut borner $|\mathbb{P}(X_T \in A) - \underbrace{\mathbb{P}(Y_T \in A)}_{=\pi(A)}|$.

$$\begin{aligned} \mathbb{P}(X_T \in A) &\geqslant \mathbb{P}(X_T = Y_T \text{ et } Y_T \in A) \\ &\geqslant 1 - \mathbb{P}(X_T \neq Y_T) - \mathbb{P}(Y_T \notin A) \\ &\geqslant \mathbb{P}(Y_T \in A) - \varepsilon \\ \underbrace{\mathbb{P}(Y_T \in A)}_{\pi(A)} - \mathbb{P}(X_T \in A) &\leqslant \varepsilon \\ |\mathbb{P}(X_T \in A) - \mathbb{P}(Y_T \in A)| &\leqslant \varepsilon \end{aligned}$$

Exemple. Mélange de cartes.

1. Évolution indépendante de deux paquets : presque-spurement $\exists n$ tel que $X_n = Y_n$, mais n potentiellement beaucoup trop grand.
2. Si on choisit les mêmes modifications dans les deux paquets, ils ne convergent jamais !
3. Solution efficace : choisir une carte uniformément dans le premier tas, et prendre la carte avec la même valeur (exemple : $R\spadesuit$) dans le deuxième tas pour les mettre au-dessus. Les deux tas seront égaux lorsque chacune des carte aura été choisie une fois.

On retrouve le problème du collectionneur de coupons : (n = nombre de cartes)

$$\begin{aligned}
\mathbb{P}(X_{n\ln n+cn} \neq Y_{n\ln n+cn}) &\leqslant n\left(1-\frac{1}{n}\right)^{n\ln n+cn} \\
&\leqslant n\,e^{-(\ln n+c)} \\
&\leqslant e^{-c} \\
&\leqslant \varepsilon \quad \text{si } c=\ln\frac{1}{\varepsilon} \\
\tau(\varepsilon) &\leqslant n\ln\frac{n}{\varepsilon}
\end{aligned}$$

4.4 Application : ensembles indépendants pour un graphe

On veut échantillonner un indépendant de taille k fixée.

Chaine de Markov.

- choisir $v \in X_n$ uniformément et $w \in S$ uniformément
- Si $(X_n \cup \{w\})\backslash\{v\}$ est un indépendant de taille k, alors $X_{n+1} \leftarrow (X_n \cup \{w\})\backslash\{v\}$, sinon $X_{n+1} \leftarrow X_n$.

Il faut montrer que la distribution stationnaire est bien une distribution uniforme. Notons I un indépendant de taille k et $I' = (I \cup \{w\})\backslash\{v\}$.

$$\begin{aligned}
p_{I,I'} &= \frac{1}{k} \times \frac{1}{|S|} \\
\pi(I)\, p_{I,I'} &= \pi(I')\, p_{I',I} \\
\pi(I) &= \pi(I')
\end{aligned}$$

Chaine ergodique ? Apériodique car $p_{I,I} > 0$. Irréductible sous la condition $k \leqslant \frac{|S|}{3(\Delta+1)}$, où Δ est le degré maximal d'un sommet.

Couplage.

- On choisit v uniformément dans X_n.
- Si $v \in Y_n$, on choisit ce v pour y_n ($v' = v$). Sinon on choisit v' uniformément dans $Y_n - X_n$.
- On choisit w uniformément dans S, et on pose :

$$\begin{aligned}
X_{n+1} &= X_n \cup \{w\}\backslash\{v\} \text{ si c'est bien un indépendant de taille } k \\
Y_{n+1} &= Y_n \cup \{w\}\backslash\{v'\} \text{ si c'est bien un indépendant de taille } k
\end{aligned}$$

`2014-12-17.`

Chaines ergodiques. $p_{i,j}(n) \xrightarrow[n\to\infty]{} \pi(j)$, $\forall i, j \in E$

Simulation MCMC. Générer un état « presque » selon la distribution π. $X_{n+1} = f(X_n, U_{n+1})$: on génère les U_n IID. Deux inconvénients :

- Presque selon la distribution stationnaire mais pas tout à fait ;
- Il faut calculer le temps de mélange de la chaîne pour savoir environ quand s'arrêter.

V. Algorithme de Propp et Wilson

Simulation parfaite (couplage depuis le passé). Ne fonctionne que sur un espace d'états E fini.

Plutôt que de regarder $X_0, X_1, ..., X_\infty$, on étudie un X_0 en faisant comme si on avait commencé à $X_{-\infty}$.

Représentation fonctionnelle : $f: E \times F \to E$.

On prend $N_1 < N_2 < \cdots < N_n < \cdots$ une suite strictement croissante dans $\mathbb{N}$.

On prend $U_{-1}, ..., U_{-n}, ...$ une suite de VA IID dans F (selon la distribution de la représentation fonctionnelle).

Soit $x \in E$ et $N \in \mathbb{N}$, $\varphi_N(x) = f(\cdots f(f(x, U_{-N}), U_{-N+1}) \cdots, U_{-1})$

Algorithme de Propp et Wilson.

$m \leftarrow 0$

$E' \leftarrow E$

tant que $|E'| > 1$:

$E' \leftarrow \bigcup_{x \in E} \{\varphi_{N_m}(x)\}$

$m \leftarrow m + 1$

renvoyer y l'unique élément de E'

Il reste à montrer que y est distribué selon la distribution stationnaire de $\{X_n\}$.

Lemme. L'algorithme termine avec probabilité 0 ou 1 (selon la fonction f).

Preuve. S'il existe u_1, ..., u_n et y tel que $\forall x \in E$, $f(\cdots f(f(x, u_1), u_2) \cdots, u_n) = y$, et $\forall i$, $\mathbb{P}(U_{-i} = u_i) = 1$, alors $\mathbb{P}(U_{-1} = u_{-1}, ..., U_{-n} = u_{-n}) > 0$. On découpe la suite U_i en blocs de taille n ; la probabilité que l'un de ces blocs soit $(u_1, ..., u_n)$ est 1 (lemme de Borel-Cantelli). C'est-à-dire que l'algorithme termine avec probabilité 1. Sinon, l'algorithme termine avec probabilité 0 de façon évidente.

Remarque. Pour une même chaine de Markov, il peut y avoir des représentations (ie des fonctions f et des lois pour U_n) telles que l'algorithme termine très rapidement et d'autres pour lesquelles ça ne termine pas.

Lemme 2. Si l'algorithme termine vec probabilité 1, alors il termine en un temps d'espérance finie.

Preuve. Soit T le temps tel que $\left|\bigcup_{x \in E} \{\varphi_T(x)\}\right| = 1$.

$$\mathbb{E}[T] \leqslant \frac{n}{\mathbb{P}(U_{-1} = u_1, ..., U_{-n} = u_n)}$$

Théorème. Soit $\{X_n\}$ une CMH ergodique sur un espace fini E, et (N_n) une suite strictement croissante. Si l'algorithme termine avec proba 1 et si Y est sa sortie, alors $\mathbb{P}(Y = i) = \pi(i), \forall i \in E$, où π est la distribution stationnaire de $\{X_n\}$.

Preuve. On veut montrer que $\forall \varepsilon > 0$, $|\mathbb{P}(Y = i) - \pi(i)| \leqslant \varepsilon$. L'algorithme termine en un temps d'espérance finie, donc il existe M tel que $\mathbb{P}(\text{l'algo termine en } M \text{ itérations}) \geqslant 1 - \varepsilon$. Soit $\left\{\tilde{X}(n)\right\}$ copie de $\{X_n\}$ telle que $\tilde{X}(0) \sim \pi$ et $\tilde{X}(n)$ sont obtenues à partir de la même suite $U_{-N_M}, ..., U_{-1}$. Soit $\tilde{Y}$ l'état obtenu par l'algorithme avec cette chaine de Markov ; la distribution de $\tilde{Y}$ est π.

$$\begin{aligned} |\mathbb{P}(Y = i) - \pi(i)| &= \left|\mathbb{P}(Y = i) - \mathbb{P}\left(\tilde{Y} = i\right)\right| \\ &\leqslant \mathbb{P}\left(Y \neq \tilde{Y}\right) \\ &\leqslant \varepsilon \end{aligned}$$

Donc l'algorithme est correct.

Remarques.

- Choix de (U_n) : si T est la première date à laquelle toutes les trajectoires se rejoignent, si on choisit $N_m = m$, nombre de calculs $= \sum_{m=1}^{T} m = \frac{T(T+1)}{2}$

  Si on choisit $N_m = 2^m$, on fait moins de $4T$ étapes.

- Pourquoi aller en arrière plutôt qu'en avant ?

- Pourquoi réutiliser les U_{-i} ?

- Que faire quand on a trop d'états?

Chaînes monotones. E est un ensemble (partiellement) ordonné, avec $\top$ et $\bot$ états maximal et minimal ; $f : E \times F \to E$, vérifiant :

$$x \preccurlyeq y \;\Rightarrow\; \forall u \in F, f(x,u) \preccurlyeq f(y,u)$$

Dans le cas d'une chaîne monotone, il nous suffit de faire la simulation pour $\bot$ et $\top$!

Exemple de chaine monotone : file d'attente, proba p d'arrivée, proba q de départ, jamais une arrivée et un départ en même temps, $p+q=1$, capacité maximale c et minimale 0 (dans ces cas une arrivée/un départ n'a aucun effet).

Chaînes bornantes. $\forall n$, $Y_n \supseteq \bigcup_{x \in E} \{f(\cdots\ f(x,\ u_1)\ \cdots,\ u_n)\}$. Y_n CMH pas nécessairement ergodique.

Exemple. Mélange de cartes.

VI. Critères de récurrence positive

On se place dans le cas où $|E| = \infty$ (dénombrable).

Lemme. On se donne $\{X_n\}$ une CMH irréductible et F un sous-ensemble fini de E. Si $\forall j \in F$, $\mathbb{E}_j[T_F] < \infty$, alors $\{X_n\}$ est récurrente positive. (avec $T_F = \inf\{n > 0 \,|\, X_n \in F\}$)

Dem. Supposons que $X_0 \in F$. On note : τ_1, ..., τ_n, ... les temps de retour successifs en F. $\tau_{n+1} = \inf\{k > \tau_n \,|\, X_k \in F\}$. τ_n sont presque-sûrement finis. Posons $Y_n = X_{\tau_n}$. $\{Y_n\}$ est une CMH sur F. On sait que X_n est irréductible, donc Y_n l'est aussi. Y_n est irréductible sur un espace fini, donc récurrente positive.

Soit $\tilde{T_i}$ le temps de retour en i pour $\{Y_n\}$.

$$\begin{aligned}
T_i &= \sum_{k=0}^{\infty} (\tau_{k+1} - \tau_k)\, \mathbf{1}_{k < \tilde{T_i}} \\
\mathbb{E}_i[T_i] &= \sum_{k=0}^{\infty} E_i[(\tau_{k+1} - \tau_k)\mathbf{1}_{k < \tilde{T_i}}] \\
&\vdots \\
\mathbb{E}_i[T_i] &= \sum_k \sum_{l \in F} \mathbb{E}_l[T_F]\mathbb{P}\big(X_{\tau_k} = l; k < \tilde{T_i}\big) \\
&\leqslant \Big(\max_{l \in F} E_l[T_F]\Big) \sum_k \mathbb{P}_i\big(k < \tilde{T_i}\big) \\
&< \infty
\end{aligned}$$

On a donc un état de X_n récurrent positif, donc X_n est récurrente positive (puisqu'irréductible).

Théorème. (Foster) Soit $\{X_n\}$ une CMH irréductible sur E de matrice de transition P. S'il existe $h: E \to \mathbb{R}^+$, $F \subseteq E$ fini et $\varepsilon > 0$ tels que :

$$\begin{aligned} \forall i \in F, \quad & \sum_{k \in E} p_{i,k}\, h(k) \;<\; \infty \\ \forall i \notin F, \quad & \sum_{k \in E} p_{i,k}\, h(k) \;\leqslant\; h(i) - \varepsilon \end{aligned}$$

C'est-à-dire :

$$\begin{aligned} &\forall i \in F, \quad \mathbb{E}_i(h(X_1)) < \infty \\ &\forall i \notin F, \quad \mathbb{E}_i(h(X_1) - h(X_0)) \leqslant -\varepsilon \end{aligned}$$

Alors $\{X_n\}$ est récurrente positive.

Preuve. Soit T le temps de retour en F. On note $Y_n = h(X_n)\,\mathbf{1}_{n<T}$.

$\forall i \notin F, \forall n \geqslant 0$,

$$\begin{aligned} \mathbb{E}_i[Y_n | X_0, ..., X_n] &= \mathbb{E}_i[h(X_{n+1})\mathbf{1}_{n+1<T} | X_0, ..., X_n] \\ &\leqslant \mathbb{E}_i[h(X_{n+1})\mathbf{1}_{n<T} | X_0, ..., X_n] \\ &\leqslant \mathbf{1}_{n<T}\, \mathbb{E}_i[h(X_{n+1}) | X_n] \\ &\leqslant \mathbf{1}_{n<T}(h(X_i) - \varepsilon) \qquad \text{p.s.} \\ &... \end{aligned}$$

Aparté : espérance conditionnelle.

$\mathbb{E}[X|Y]$ peut être vu comme une variable aléatoire $f(Y)$ qui ne dépend que de Y : $f(y) = \mathbb{E}[X|Y=y]$.

Prop. Si $X \perp\!\!\!\perp Y$, alors $\mathbb{E}[X|Y] = \mathbb{E}[X]$.

Prop. $\mathbb{E}[\mathbb{E}[X|Y]] = \sum_y \mathbb{E}[X|Y=y]\,\mathbb{P}(Y=y) = \sum_{x,y} x\,\mathbb{P}(X=x)\mathbb{P}(Y=y) = \sum_x x\,\mathbb{P}(X=x) = \mathbb{E}[X]$.

Prop. $\mathbb{E}[g(Y)f(X,Y)|Y] = g(Y)\mathbb{E}(f(X,Y)|Y)$

Ex. $X_{n+1} = [(X_n - 1) \vee 0] + A_{n+1}$ avec A_{n+1} IID. Si $\mathbb{E}[A_{n+1}] < 1$ alors X_n récurrente positive. En effet le théorème s'applique avec $h = \mathrm{id}$, $F = \{0\}$, $\varepsilon = 1 - \mathbb{E}[A_1] > 0$.

Théorème. Soit $\{X_n\}$ une CMH irréductible sur E. On suppose qu'il existe h: $E \to \mathbb{R}^+$ à incréments bornés et F un ensemble fini tels que :

$$\begin{aligned} & \exists j_0 \notin F \text{ tq} \quad h(j_0) \;>\; \max_{i \in F} h(i) \\ \forall i \notin F, \quad & E_i[h(X_1) - h(X_0)] \;\geqslant\; 0 \end{aligned}$$

Alors $\{X_n\}$ ne eput être récurrente positive.

Dem. On suppose $\{X_n\}$ récurrente positive. $\forall j \notin F, \mathbb{E}_j[T_F] < \infty$.

$$\begin{aligned} h(X_{T_F}) &= h(X_{T_F})\mathbf{1}_{\{T_F<\infty\}} \\ &= h(X_0) + \sum_{n=0}^{\infty} (h(X_{n+1}) - h(X_n))\mathbf{1}_{T_F>n} \\ \mathbb{E}_j[h(X_{T_F})] &= h(j) + \sum_{n=0}^{\infty} \mathbb{E}_j[(h(X_{n+1}) - h(X_n))\mathbf{1}_{T_F>n}] \\ &> h(j) \end{aligned}$$

$\forall j \notin F$, $\exists i \in F$ tel que $h(i) \geqslant h(j)$, contradictoire avec les hypothèses.

Ex. $X_{n+1} = \max(X_n - 1, 0) + A_{n+1}$ avec $\{A_n\}$ IID. On prend $h = \mathrm{id}$, $F = \{0\}$. Si $\mathbb{E}[A] < \infty$ incrément borné, $\mathbb{E}[h(X_1) - h(X_0)] = i + \mathbb{E}[A_i] - 1 - i = \mathbb{E}[A_i] - 1 \geqslant 0$ si $\mathbb{E}[A_i] \geqslant 1$.

Application

Protocole de communication par canal à accès multiple. (Aloha) Si un message est transmis, ok. Si plusieurs, tous sont brouillés et à recommencer.

Protocole Aloha discret. Le temps est divisé en slots $k \in \mathbb{N}$. Les transmissions et retransmissions commencent en début de slot et si la transmission est possible, elle se fait pendant ce slot. Les nouveaux messages (première tentative de transmission) essayent de passer dès leur arrivée ; les messages retardés (ayant subi au moins une collision) tentent de passer avec probabilité ν fixée (assez faible) et indépendamment les uns des autres.

Soit A_n le nombre de nouveaux messages au slot n. On suppose que (A_n) est IID, on note $\lambda = \mathbb{E}[A_n]$ que l'on appelle *intensité du traffic*. On suppose $\lambda < \infty$. On note X_n le nombre de messages retardés au slot n. On note $a_i = \mathbb{P}(A_1 = i)$.

X_n CMH ? Irréductible ? On veut montrer que $\forall \lambda$, X_n n'est pas récurrente positive. (le protocole est instable). X_n est bien une CMH ; irréductible sous condition que $\mathbb{P}(A_1 = 0) > 0$ et $\mathbb{P}(A_1 > 1) > 0$.

Matrice de transition de X_n : introduisons $b_i(k)$ la proba que i messages tentent de retransmettre si k sont retardés. $b_i(k) = \binom{k}{i} \nu^i (1-\nu)^{k-i}$.

$$p_{k,l} = \begin{cases} a_0\, b_1(k) & l = k-1 \\ a_1\, b_0(k) + a_0\,(1 - b_1(k)) & l = k \\ a_1(1 - b_0(k)) & l = k+1 \\ a_{l-k} & l > k+1 \end{cases}$$