

Introduction aux chaînes de Markov

Un **processus stochastique** en temps discret est une suite

$\mathcal{W} = (X_0, X_1, X_2, \dots, X_k, \dots)$ de variables aléatoires.

Ce processus \mathcal{W} sera dit **sans mémoire** si les variables qui le composent sont *indépendantes et identiquement distribuées (i.i.d.)*.

Il est clair qu'un tel modèle de génération de données aléatoires (texte, image...) est bien éloigné de la réalité !

Comment modéliser de façon plus satisfaisante un texte aléatoire, une image aléatoire, ou encore d'autres types de sources ??

Définition:

- Soit $\mathcal{W} = (X_0, X_1, X_2, \dots)$ un processus constitué de variables à valeurs dans un même ensemble fini ou dénombrable \mathcal{V} .

On dira que \mathcal{W} constitue une **chaîne de Markov** si les variables qui le composent satisfont à la propriété suivante:

quel que soit l'instant $n \geq 0$ et les valeurs $v_0, \dots, v_n, v_{n+1} \in \mathcal{V}^{n+2}$, si $\mathbb{P}\{X_0 = v_0, X_1 = v_1, \dots, X_n = v_n\} > 0$ alors

$$\begin{aligned} \mathbb{P}\{X_n = v_n, X_{n-1} = v_{n-1}, \dots, X_0 = v_0\} \{X_{n+1} = v_{n+1}\} \\ = \mathbb{P}\{X_n = v_n\} \{X_{n+1} = v_{n+1}\} \end{aligned}$$

- En Français: dans une chaîne de Markov, **conditionnellement au présent**, le passé et le futur sont indépendants!

$$\mathbb{P}\{X_n = v_n\} \{(X_{n+1} = v_{n+1}) \cap (X_{n-1} = v_{n-1}, \dots, X_0 = v_0)\}$$

$$\mathbb{P}\{X_n = v_n\} \{X_{n+1} = v_{n+1}\} \cdot \mathbb{P}\{X_{n-1} = v_{n-1}, \dots, X_0 = v_0\}$$

La valeur du processus à un instant n ne dépend que de sa valeur à l'instant $n-1$

- On pourra aussi décrire cette forme très naturelle d'interdépendance comme suit :

dans une chaîne de Markov, le futur ne dépend du passé qu'à travers le dernier instant du passé !

Premiers exemples de Chaînes de Markov:

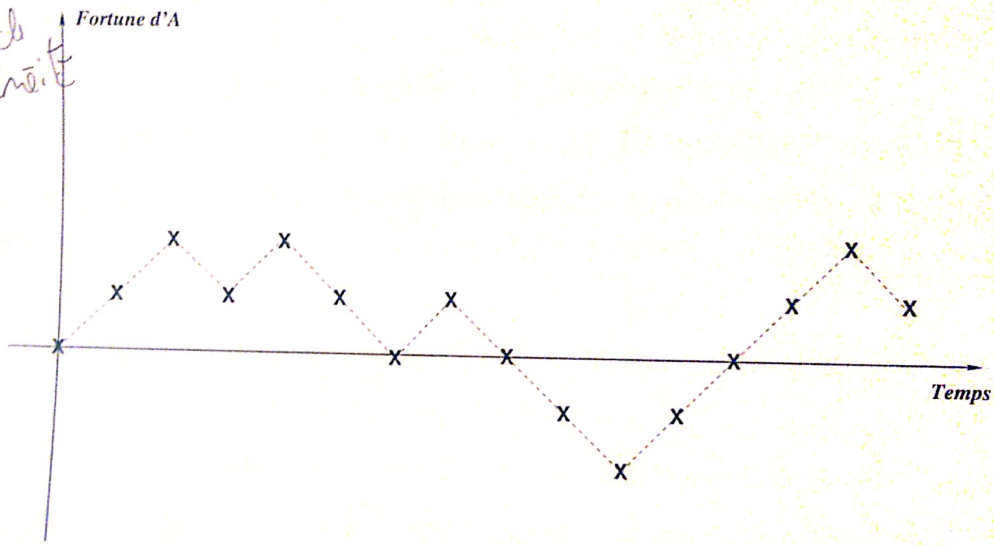
1^o) Alice et Bob jouent à l'aide d'une pièce en pariant sur "Pile" ou "Face" lors de lancers successifs; à chaque fois que "Pile" sort, Alice gagne 1 Euro, tandis que dans le cas contraire c'est Bob qui gagne 1 Euro.

La suite $\mathcal{G} = (G_1, G_2, \dots, G_k, \dots)$ des gains d'Alice constitue une chaîne de Markov; si R_l est la variable de Bernoulli prenant la valeur +1 lorsque le $l^{\text{ème}}$ lancer donne "Pile", -1 sinon, on a

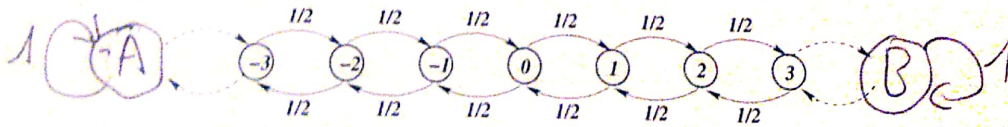
$$G_{k+1} = \sum_{l=1}^{k+1} R_l = G_k + R_{k+1}$$

Une réalisation donnée de $\mathcal{G} = (G_k)_{k \geq 1}$ pourra être représentée par

On change pas de pièce: homogénéité



On pourra par ailleurs représenter la loi de cette chaîne au moyen de son **graphe de transitions**:

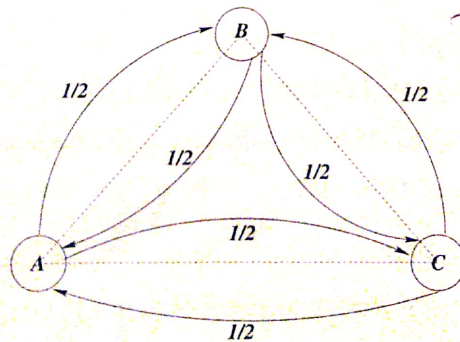


si la pièce est rigée on aurait p en bas et 1-p dans l'autre sens
 $P(G_{m+1} = v+1 | G_m = v) = p, q = P(G_{m+1} = v-1 | G_m = v)$

Premiers exemples de Chaînes de Markov (suite) :

2°) Un scarabée se promène le long des côtés d'un triangle équilatéral $T = (ABC)$ dans le plan. à chaque fois qu'il atteint un sommet de T , il décide de se diriger ensuite vers l'un des deux sommets restants; cette décision est prise de manière aléatoire et équiprobable, et le parcours du côté correspondant lui prend toujours 5s.

La suite $(X_0, X_1, X_2, \dots, X_k, \dots)$ des positions du scarabées (relevées toutes les 5s) est manifestement une chaîne de Markov, dont l'évolution pourra représentée graphiquement comme suit :



Triangle

$$P(X_{n+1} = B | X_n = A) = P(X_{n+1} = C | X_n = A) = \frac{1}{2}$$

Case

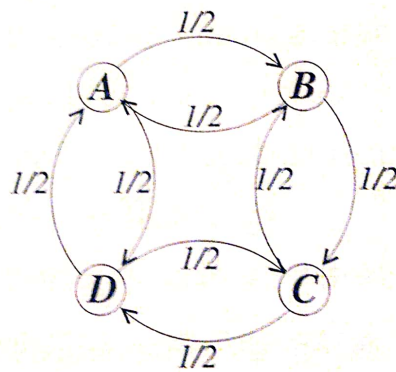
$$P(X_{n+1} = A | X_n = A) = 0$$

$$P(X_{n+1} = B | X_n = A) = \frac{1}{2}$$

$$P(X_{n+1} = C | X_n = A) = \frac{1}{2}$$

Premiers exemples de Chaînes de Markov (suite et fin) :

3^o) Si notre scarabée se promène le long des côtés d'un carré ($ABCD$) dans le plan, la suite $(X_0, X_1, X_2, \dots, X_k, \dots)$ des positions du scarabées (relevées à intervalles réguliers) est manifestement une chaîne de Markov, dont l'évolution pourra représentée graphiquement comme suit :



$$\begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$

Définition:

- La chaîne de Markov $\mathcal{W} = (X_0, X_1, X_2, \dots, X_k, \dots)$ sera dite **homogène** lorsque, pour toutes valeurs possibles $v, v' \in \mathcal{V}$, la prob. cond. $\mathbb{P}\{X_{k+1} = v' | X_k = v\}$ ne dépend pas de l'instant k considéré mais *seulement des valeurs* v, v' .
- Lorsque \mathcal{V} est fini ($|\mathcal{V}| = m$), en numérotant les éléments de \mathcal{V} on pourra poser

$$\forall (i; j) \in \{1, \dots, m\}^2, \quad \mathbb{P}\{X_{k+1} = v_j | X_k = v_i\} = p_{i,j}$$

Ce faisant, on obtient une *matrice stochastique* P appelée **matrice des transitions** de la chaîne \mathcal{W} . Dans l'ex. du scarabée sur un triangle:

colonne : état d'arrivée
 ligne : état de départ
 Matrice stochastique

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

chaque ligne vaut 1, c'est un état de proba
 ce qui sert d'indice a sa somme égale à 1.

Pour spécifier entièrement la loi d'une telle chaîne, il nous faut encore décrire le comportement de la première variable X_0 . En posant $\mathbb{P}\{X_0 = v_i\} = \pi_i$ ($i = 1, 2, \dots, m$), on obtient un *vecteur-ligne*

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_m)$$

La **multiplication à gauche** de la matrice de transitions P par $\boldsymbol{\pi}$ permet de passer de la loi de X_0 à celle de X_1 :

$$\text{Loi}(X_1) = \boldsymbol{\pi} \times P = \boldsymbol{\pi}^{(1)}$$

Plus généralement, pour tout instant $n \geq 1$:

$$\text{Loi}(X_n) = \boldsymbol{\pi}^{(n)} = \boldsymbol{\pi} \times P^n$$

Dans l'ex. du scarabée, en supposant qu'il se trouve en A à l'instant initial:

$$\text{Loi}(X_0) = (1; 0; 0), \text{Loi}(X_1) = (0; \frac{1}{2}; \frac{1}{2}), \text{Loi}(X_2) = (\frac{1}{2}; \frac{1}{4}; \frac{1}{4}), \dots$$

$$\text{Si loi } X_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \boldsymbol{\pi} \quad \text{loi } (X_1) = \begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \boldsymbol{\pi} \times P = \boldsymbol{\pi}^{(1)}$$

Déf. : la matrice P sera dite *stochastique* si elle est carrée à entrées ≥ 0 avec des sommes par lignes toutes égales à 1.

Théorème (Perron-Frobenius):

Si $P = (p_{i,j})_{1 \leq i,j \leq m}$ est une matrice stochastique de format $m \times m$, alors

1°) 1 est valeur propre de P (pour les multiplications à gauche tout comme pour les multiplications à droite).

2°) Il existe $\pi^{inv.}$, vecteur propre à gauche de P pour la valeur 1, tel que

$$\pi^{inv.} \times P = \pi^{inv.}, \quad \pi^{inv.}(1), \pi^{inv.}(2), \dots, \pi^{inv.}(m) \geq 0 \quad \text{et} \quad \sum_{i=1}^m \pi^{inv.}(i) = 1.$$

Qd est ce que une chaîne de Markov admet plusieurs mesures invariables?
 Si le graphe n'est pas fortement
 Si il est connexe il y a 1 et 1 seul mesure invariable

Dans beaucoup des exemples que nous rencontrerons, la chaîne étudiée est **ergodique** ; $\pi^{inv.}$ est alors unique et s'obtient à partir d'une distribution initiale ($\pi^{(0)} = \text{Loi}(X_0)$) quelconque en passant à la limite:

$$\pi^{inv.} = \lim_{n \rightarrow +\infty} \pi^{(0)} \times P^n = \lim_{n \rightarrow +\infty} \text{Loi}(X_n) = \pi^{éq.}$$

Ainsi obtient-on dans l'exemple du scarabée sur $T = (ABC)$:

$$\pi^{inv.} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \pi^{éq.}$$

En revanche, si notre scarabée évolue sur un carré $(ABCD)$, on constatera que $\pi^{inv.} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ sans qu'il y ait lieu de parler d'une convergence systématique de $\text{Loi}(X_n)$ vers $\pi^{inv.}$!

Pour chacune de ces marches polygonales du scarabée, la mesure invariante est uniforme ; mais la convergence de $\text{Loi}(X_n)$ vers $\pi^{inv.}$ a cours dans le contexte triangulaire, tandis qu'elle est aisément mise en défaut dans le cas du carré ... Pourquoi ??

Autres exemples (Modèles de sources):

- Relever les fréquences d'apparitions de couples (v_i, v_j) de caractères successifs (digrammes) dans la langue considérée (en utilisant un grand nombre de données textuelles) puis construire une chaîne de Markov conformément à ces fréquences.
- Relever les fréquences d'apparitions de triplets (v_i, v_j, v_k) (resp. quadruplets, etc...) de caractères successifs dans cette langue, puis construire une chaîne de Markov conformément à ces fréquences.
- Relever les fréquences d'apparitions de couples (w_i, w_j) de mots successifs dans le lexique de la langue considérée, puis construire une chaîne de Markov conformément à ces fréquences.
- Relever les fréquences d'apparitions de triplets (w_i, w_j, w_k) (resp. quadruplets, etc...) de mots successifs dans le lexique de cette langue, puis construire une chaîne de Markov...

- Echantillon obtenu à partir d'une approx. markovienne d'ordre 1 des caractères latins de la langue anglaise

OUCTIE IN ARE AMYST TE TUSE SOBE CTUSE

- Echantillon obtenu à partir d'une approx. markovienne d'ordre 2 des caractères latins de la langue anglaise

HE AREAT BEIS HEDE THAT WISHBOUT SEED DAY OFTE
AND HE IS FOR THAT MINUMB LOOTS WILL AND GIRLS

- Echantillon obtenu à partir d'une approx. markovienne d'ordre 1 des caractères latins apparaissant sous la plume de Cicéron

IENEC FES VIMONILLITUM M ST ER PEM ENIM PTAUL

- Echantillon obtenu à partir d'une approx. markovienne d'ordre 2 des caractères latins apparaissant sous la plume de Cicéron

SENECTOR VCI QUAEMODOMIS SE NON FRATURDIGNAVIT
SINE VELIUS

Convergence des Chaînes de Markov

Soit $(X_n)_{n \geq 0}$ une Chaîne de Markov *homogène* et à *espace d'états fini*.

↪ QUESTION FONDAMENTALE :

"Y-a-t-il des conditions portant sur les *transitions* de cette Chaîne de Markov et permettant d'affirmer que la loi de X_n converge au fil du temps ($n \rightarrow \infty$) vers une certaine loi limite π ?"

Du simple fait que l'espace d'états $E = \{e_1, e_2, \dots, e_N\}$ est fini, on sait déjà (cf Thm de Perron-Frobenius) qu'il existe *au moins* une mesure invariante $\pi = (\pi_1, \dots, \pi_N)$ pour notre chaîne homogène $(X_n)_{n \geq 0}$.

↪ Question intermédiaire naturelle :

Sous quelle(s) condition(s) peut-on affirmer que la chaîne $(X_n)_{n \geq 0}$ admet *exactement une mesure invariante* π ?

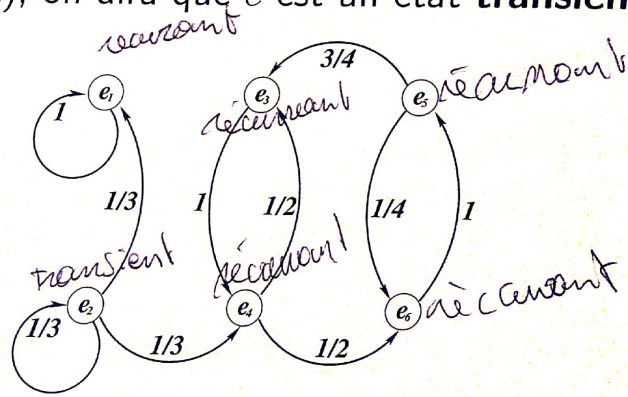
Déf.:

un état $e \in E$ sera dit **récurrent** s'il est t.q.

$$\mathbb{P} \{ \exists n \geq 1 : X_n = e | X_0 = e \} = 1,$$

autrement dit si l'on est \mathbb{P} -presque sûr de retrouver ultérieurement l'état e en démarrant la chaîne $(X_n)_{n \geq 0}$ en cet état.

Si par contre $\mathbb{P} \{ \exists n \geq 1 : X_n = e | X_0 = e \} < 1$ (il y a une prob. non nulle pour que l'on ne retrouve jamais l'état e en démarrant la chaîne $(X_n)_{n \geq 0}$ en cet état), on dira que e est un état **transient** de la chaîne.



Dans une ch de M. avec un nombre fini d'états, un état fini a qu'un nombre fini d'états récurrents.

Des états communiquants ont du même type/moder communiquant. Si membres d'une classe fermée comme $\{e_3, e_4, e_5, e_5\}$.

Dans l'exemple précédent, on voit que l'état e_2 est transient, tandis que tous les autres états sont récurrents.

Il nous faut maintenant une nouvelle

Déf.:

- L'état e_j sera dit **atteignable depuis** e_i s'il existe un entier n t.q.

$$p_{i,j}^{(n)} = \mathbb{P}\{X_n = e_j | X_0 = e_i\} > 0$$

On écrira alors $e_i \rightarrow e_j$; si e_i est aussi atteignable depuis e_j ($e_i \leftrightarrow e_j$) on dira que ces états **communiquent** (l'un avec l'autre).

- Une partie non vide C de l'espace E sera dite
 - **fermée** si

$$p_{i,j} = 0, \quad \forall e_i \in C, \forall e_j \in (E \setminus C)$$

- **irréductible** si

$$e_i \leftrightarrow e_j, \quad \forall e_i, e_j \in C$$

Proposition 1:

Il y a une unique partition de l'espace d'états E de la forme

$$E = T \cup C_1 \cup C_2 \cup \dots \cup C_p,$$

où T désigne l'ensemble des états transients de la chaîne (X_n) , tandis que les composantes C_1, C_2, \dots, C_p sont des sous-ensembles non-vides, fermés et irréductibles, constitués d'états récurrents.

Dans l'exemple donné précédemment, on obtient

$$T = \{e_2\}, \quad p = 2 \text{ et } C_1 = \{e_1\}, C_2 = \{e_3, e_4, e_5, e_6\}$$

NB:

on pourrait bien sûr avoir $T = \emptyset$, par contre $T = E$ n'est pas envisageable! On a toujours $p \geq 1$, il y a toujours au moins une classe (non-vide) d'états récurrents.

Un état transient ne peut pas recevoir de masse sous une mesure invariante

Plusieurs mesures invariantes \rightarrow infinie

Proposition 2:

La chaîne $(X_n)_{n \geq 0}$ (sur l'espace d'états fini E) admet une **unique mesure invariante** π ssi $p = 1$, c'ad ssi il y a une unique classe irréductible $C_1 \subset E$ d'états récurrents pour cette chaîne.

Sinon, dès lors que $p \geq 2$, il y a pour chacune des classes récurrentes C_i ($1 \leq i \leq p$) une unique mesure invariante $\pi^{[i]}$ t.q. *ex: les moles placées dans leur moules dans la classe qui leur corresp*

$$\pi^{[i]}(C_i) = 1 \text{ et } \pi^{[i]}(E \setminus C_i) = 0,$$

et toute mesure invariante de la chaîne $(X_n)_{n \geq 0}$ peut être décomposée sous la forme

$$\pi[\alpha_1, \alpha_2, \dots, \alpha_p] = \sum_{i=1}^p \alpha_i \pi^{[i]},$$

Mesure univ. extrême

avec $\alpha_1, \alpha_2, \dots, \alpha_p \geq 0$ et $\sum_{i=1}^p \alpha_i = 1$.

~~20/10/03~~

Handwritten notes:

$$\begin{cases} \pi_5 = \frac{2}{3} \pi_3 \\ \pi_6 = \pi_5 \\ \pi_6 = \frac{2}{3} \pi_3 \end{cases}$$

$$\pi_3 + \pi_3 + \pi_5 + \pi_6 = 1$$

Handwritten equations:

$$\begin{aligned} \frac{1}{2} \pi_4 + \frac{3}{4} \pi_5 &= \pi_3 \\ \pi_3 &= \pi_4 \\ \pi_6 &= \pi_5 \\ \frac{1}{2} \pi_4 + \frac{1}{4} \pi_5 &= \pi_6 \end{aligned}$$

Handwritten notes:

$\pi^{[1]} = (100000)$

$\pi^{[2]}(E_2) = 1$

Mec Q la sous matrice de la chaîne

Dans l'exemple donné précédemment, on obtient aisément

$$\pi[1] = (1, 0, 0, 0, 0, 0), \quad \pi[2] = (0, 0, 3/10, 3/10, 1/5, 1/5),$$

et les distributions invariantes de la chaîne sont donc toutes les distributions de la forme

$$\pi = (1 - \alpha, 0, 3\alpha/10, 3\alpha/10, 1\alpha/5, 1\alpha/5)$$

pour α variant dans $[0; 1]$.

Revenons maintenant à la question initiale : *sous quelles conditions peut-on affirmer que la loi de X_n converge vers une distribution fixée π ?*

Manifestement : une telle distribution-limite π doit être invariante!

Si la chaîne (X_n) n'a qu'une seule classe récurrente, il y a donc une seule distribution-limite π envisageable, mais un phénomène pourrait encore empêcher la convergence en loi de X_n vers π :

l'éventuelle **périodicité** de la chaîne!

Déf.:

- Soit e_l un état récurrent de la chaîne (X_n) .
On appelle **période** de cet état et l'on note $Pér(e_l)$ le PGCD de tous les entiers n pour lesquels

$$p_{l,l}^{(n)} > 0$$

- Deux états récurrents e_k et e_l situés dans la même classe récurrente C_i doivent avoir la même période! Cette période commune à tous les états de C_i est appelée période de C_i , notée $Pér(C_i)$.
- Une classe récurrente C sera dite **apériodique** si $Pér(C) = 1$.
- Dans le cas d'une chaîne (X_n) irréductible ($T = \emptyset, p = 1, C_1 = E$), on parlera de chaîne **irréductible apériodique** lorsque l'unique classe $C = E$ est apériodique.

Dans une classe, tous les états récurrents ont la même période

transitions
 ↗

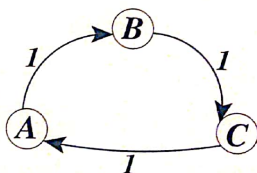
N.B. :

- La chaîne (X_n) est en fait **irréductible apériodique** si et seulement s'il existe un entier $n_0 \geq 1$ t.q.

$$\forall e_i, e_j \in E, \quad p_{i,j}^{(n_0)} = (P^{n_0})_{i,j} = \mathbb{P}\{X_{n_0} = j | X_0 = i\} > 0$$

Un nombre fixe de pas me permet d'aller de nimporte quel état à nimporte quel état

- Le fait que la périodicité de la chaîne (X_n) puisse constituer un obstacle empêchant $Loi(X_n)$ de converger vers la mes. invariante π se voit bien sur l'exemple suivant (véritablement trivial!) :



Unique invariante : $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$

Théorème de Döblin:

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov sur l'espace d'états fini $E = \{e_1, \dots, e_N\}$.

Si $(X_n)_{n \geq 0}$ est irréductible et apériodique, elle admet une unique mesure invariante $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ et l'on a

$$\forall (i, j) \in \{1, 2, \dots, N\}^2, \quad p_{i,j}^{(n)} \xrightarrow[n \rightarrow \infty]{} \pi_j$$

En outre, cette convergence est exponentiellement rapide, au sens où

$$\exists \kappa, K > 0, \quad \forall n \geq 1, \quad \max_{1 \leq j \leq N} \max_{1 \leq i \leq N} |p_{i,j}^{(n)} - \pi_j| \leq K \cdot e^{-n\kappa}$$

NB:

Théorème fondamental prouvé dans les années 1930 par le jeune mathématicien franco-allemand *Wolfgang Döblin*, au moyen d'une méthode de couplage qui connut un grand succès dans de nombreux autres contextes de recherche en Probabilités.

Ultimes remarques : 1°) La conclusion du Thm vaudrait tout aussi bien pour une chaîne (à espace d'états fini) *quasi-irréductible* pour laquelle $p = 1$ (toujours une seule classe récurrente) mais $T \neq \emptyset$.
2°) En présence de *plusieurs classes récurrentes* C_1, C_2, \dots, C_p , il s'agira de traiter séparément chaque classe récurrente et de trouver une mesure invariante extrémale pour chacune d'entre elles : pour $i \in \llbracket 1; p \rrbracket$, il y a une unique mesure invariante $\pi[i]$ satisfaisant

$$\pi[i](C_i) = 1 \text{ et } \pi[i](C_j) = 0 \text{ si } j \neq i.$$

Toute mesure inv. de la chaîne $(X_n)_{n \geq 0}$ peut être décomposée sous la forme

$$\pi = \pi[\alpha_1, \alpha_2, \dots, \alpha_p] = \sum_{i=1}^p \alpha_i \pi[i],$$

avec $\alpha_1, \alpha_2, \dots, \alpha_p \geq 0$ et $\sum_{i=1}^p \alpha_i = 1$.

Cependant, si la chaîne démarre de façon déterministe depuis un état situé dans la classe C_i et que C_i est apériodique, on constatera que le vecteur $Loi(X_n)$ converge à vitesse exponentielle vers $\pi[i]$.

Tests d'Hypothèses et Distance du χ^2

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Introduction

- **3 types d'hypothèses :**

1. Les observations sont-elles conformes à un modèle probabiliste donné ? – "*Hypothèse de Conformité*"
2. Les observations correspondent-elles à 2 variables indépendantes ? – "*Hypothèse d'Indépendance*"
3. Les observations correspondent-elles à 2 échantillons provenant d'une seule et même population homogène ? –

"Hypothèse d'Homogénéité"

Exemples introductifs

- **Test d'une hypothèse d'Indépendance :**

La même entreprise essaye de déterminer s'il y a un quelconque rapport entre l'âge d'un client et sa propension à acheter un certain produit.

Une enquête est donc menée auprès de 100 clients, qui se voient poser un certain nombre de questions. Cette enquête permet en particulier de produire le tableau ci-dessous :

		"Jeune"	"Plus si Jeune"
	Nombre de Clients	y	\bar{y}
"Achète"	b	29	11
"N'achète pas"	\bar{b}	51	9

Exemples introductifs

- **Question principale :**

Doit-on considérer que l'âge d'un client et sa propension à acheter ont un quelconque rapport ?

Ou plutôt que ces deux variables sont *sans rapport* ??

Hypothèse de base :

H_0 : "L'âge est sans rapport avec la propension à acheter"

Ou, plus mathématiquement :

H_0 : "Les var. X_1 et X_2 sont indépendantes"

Il s'agit donc pour l'entreprise de tester une hypothèse H_0 d'indépendance des variables X_1 et X_2 .

Exemples introductifs

- **Test d'une hypothèse de Conformité :**

Une entreprise vendant des produits cosmétiques se demande si la couleur de l'emballage d'un certain produit a une quelconque influence sur l'appréciation qu'en ont ses clients. Elle décide donc de distribuer à 200 clients 4 paquets contenant ce même produit sous 4 emballages de couleurs différentes (Rouge, Jaune, Vert, Bleu).

Les clients ne savent pas que ces 4 produits sont identiques ; après un certain temps, il leur est demandé d'identifier celui des 4 produits qui leur semble être le meilleur :

Couleur Fav.	Rouge	Jaune	Bleu	Vert
Nombre Clients	51	74	30	45

Exemples introductifs

- **Question principale :**

Doit-on considérer que la couleur de l'emballage a une quelconque *influence* sur l'appréciation de ce produit ?

Ou plutôt que nos clients sont *indifférents* à ce produit ??

Hypothèse de base :

H_0 : "Les clients sont indifférents à la couleur d'emballage"

Mathématiquement :

H_0 : "La var. X de couleur préférée suit une loi uniforme"

Il s'agit donc pour l'entreprise de tester une hypothèse **H_0** de conformité ("Goodness of Fit") à une distribution uniforme sur 4 valeurs (R, J, B, V).

En résumé :

- Peut-on considérer que les *résultats observés* sont proches des *résultats théoriques* correspondant à l'hypothèse H_0 , les différences entre observations et résultats théoriques étant dues seulement au hasard (aux "*fluctuations d'échantillonnage*") ?
- Devrait-on au contraire considérer que les résultats observés sont tellement éloignés des résultats théoriques qu'un phénomène contraire à H_0 doit avoir cours (auquel cas il convient de rejeter H_0) ?

Méthode du χ^2

- Calculer les résultats théoriques
- Reporter ces résultats théoriques dans la table des résultats observés
- Evaluer les *différences* entre chaque résultat, et *élever ces différences au carré*
- *Diviser* chacun de ces carrés de différence par le *résultat théorique*, puis *additionner* :

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O : résultat expérimental
E : résultat théorique selon H_0

Méthode du χ^2

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Pratiquer un tel calcul revient à évaluer une certaine "Distance" séparant le *tableau des valeurs observées* du "tableau des valeurs théoriques".

Idée-Clef :

S'il s'avère que le *tableau des valeurs observées* est "très éloigné" du *tableau des valeurs théoriques* – ce qui signifie que $\chi^2 \geq x$ pour un certain seuil x – considérons que l'hypothèse H_0 ne peut pas être acceptée, qu'elle doit être rejetée.

Sinon, H_0 peut être considérée comme acceptable.

Détermination du Seuil x

2 TYPES d'ERREURS ...		
	POSSIBILITES	
DECISION	H_0 Vraie	H_0 Fausse
Ne Pas Rejeter H_0	Pas d'Erreur ($1 - \alpha$)	Erreur de Type II (β)
Rejeter H_0	Erreur de Type I (α)	Pas d'Erreur ($1 - \beta$)

On déterminera le seuil x après avoir fixé α (le plus couramment : $\alpha=0,1$ ou $\alpha=0,05$ ou $\alpha=0,01$).

Détermination du Seuil x

Après avoir choisi α (e.g. $\alpha=0,1$ ou $0,05$ ou $0,01$), on fixe x en sorte que

V désignant ici une variable du χ^2 à d degrés de liberté où :

- pour un test de conformité : $d = Nb. (cellules) - 1$
- plus généralement : $d = Nb. (cellules) - Nb. (par. est) - 1$
- pour un test d'indépendance :

$d = \{ Nb. (lignes) - 1 \} \times \{ Nb. (colonnes) - 1 \}$
 (plus généralement : $d = Nb. (cellules) - Nb. (par. est) - 1$)

- pour un test d'indépendance :

$$d = Nb. (cellules) - Nb. (par. est) - 1$$

Table of " χ^2 -thresholds" (quantiles)

Degrees of freedom	Choices of α			
	0.1	0.05	0.01	0.001
d=1	x=2.71	3.84	6.64	10.83
d=2	4.60	5.99	9.21	13.82
d=3	6.25	7.82	11.34	16.27
d=4	7.78	9.49	13.28	18.46

Ex.: for a χ^2 -r. v. V having $d=1$ degree of freedom,
 $P\{V \geq 3,84\} = 0,05$

Back to our examples

- Testing “*Goodness of Fit*”:

Taking H_0 into account, original table may be completed as follows:

Colour	Red	White	Blue	Green
Obs.Results/Exp.	51/50	74/50	30/50	45/50

Thus:

$$\begin{aligned}\chi^2 &= \sum \frac{(O - E)^2}{E} \\ &= \frac{1^2}{50} + \frac{(24)^2}{50} + \frac{(-20)^2}{50} + \frac{(-5)^2}{50} \cong 20,04\end{aligned}$$

Back to our examples

- Testing “*Goodness of Fit*”:

On the other hand, since the original table

Colour	Red	White	Blue	Green
Obs.Results	51	74	30	45

only has 4 cells of results, the χ^2 -r. v. V we shall refer to will have $d=4-1=3$ “degrees of freedom”.

For such a r.v. V : $P\{V \geq 7,82\} = 0,05$

Hence: choosing (*Type I Error*) risk $\alpha = 5\%$, one may reject hypothesis H_0 (since $\chi^2 \geq 7,82$)

(Actually: one may even choose $\alpha = 0,1\%$, and still reject H_0 !).

Back to our examples

- Testing “Independence”:

First complete original table as follows:

Number of Clients	y	\bar{y}	SUMS
b	29	11	40
\bar{b}	51	9	60
SUMS	80	20	100

Forcing independence to appear yields the following “expected results”:

$$\text{Nr. } \{(b; y) - \text{clients}\} = 0,4 \times 0,8 \times 100 = 32$$

$$\text{Nr. } \{(b; \bar{y}) - \text{clients}\} = 0,4 \times 0,2 \times 100 = 8$$

$$\text{Nr. } \{(\bar{b}; y) - \text{clients}\} = 0,6 \times 0,8 \times 100 = 48$$

$$\text{Nr. } \{(\bar{b}; \bar{y}) - \text{clients}\} = 0,6 \times 0,2 \times 100 = 12$$

Back to our examples

- Testing “Independence”:

One should thus compare 2 tables as follows:

Number of Clients	y	\bar{y}	SUMS
b	29/32	11/8	40
\bar{b}	51/48	9/12	60
SUMS	80	20	100

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$= \frac{(-3)^2}{32} + \frac{(3)^2}{8} + \frac{(3)^2}{48} + \frac{(-3)^2}{12} \cong 2,32$$

Back to our examples

- Testing “*Independence*”:

Since the original table

Number of Clients	y	\bar{y}
b	29	11
\bar{b}	51	9

has 2 rows and 2 columns of result-cells, the χ^2 -r. v. V we shall refer to will have $d=(2-1)\times(2-1)=1$ “degrees of freedom”.

For such a r.v. V : $P\{V \geq 3,84\} = 0,05$

Hence: choosing (*Type I Error*) risk $\alpha = 5\%$,
one cannot reject hypothesis H_0
(we have found $\chi^2 = 2,32 < 3,84$).

2 exercises

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Exercise 1

Wishing to compare the efficiency of 3 fertilizers (A, B and C) regarding flowering, a Chemist uses each of them on a sample of 200 plants and obtains the following results:

	<i>Fertilizer A</i>	<i>Fertilizer B</i>	<i>Fertilizer C</i>
Flowering	37	70	63
No Flowering	13	10	7

In view of these results, taking $\alpha = 5\%$ (tolerated risk), may one assert that « *Flowering is independent of the Fertilizer* »? What should our Chemist conclude?

Exercise 2

In 2014, the nation-wide « Baccalauréat » examination took place under 3 different versions according to the following distribution:

« *General* » (47% of the candidates)

« *Technological* » (20% of the candidates)

« *Professional* » (33% of the candidates)

Some 250 candidates are chosen at random in a certain examination center; it turns out that 131 of them have prepared the « *General* » exam, whereas 37 have prepared the « *Technological* » exam and 82 of them the « *Professional* » exam.

Choosing $\alpha = 5\%$ (tolerated risk), may one assert that this sample of 250 candidates is representative of the entire population?

And for $\alpha = 1\%$?

A further example

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Worked example 2:

A blood-drive is carried out on a large university campus and the blood-type data are collected. Local health authorities have the distribution of blood types in that area recorded as:

O = 47%, A = 26%, B = 19%, AB = 8%.

The blood drive received 1200 donors, and their blood types were recorded as thus:

O = 501, A = 315, B = 249, AB = 135.

- What is the value of χ^2 ?
- How many degrees of freedom are there?
- What is the critical value at $p=0.05$?

Worked example 2:

A blood-drive is carried out on a large university campus and the blood-type data are collected. Local health authorities have the distribution of blood types in that area recorded as:
 O = 47%, A = 26%, B = 19%, AB = 8%. probabilities

The blood drive received (1200 donors.) = sample size

Step 1: Calculate E

Blood type	Probability	x1200	Expected Frequency
O	0.47	x1200 =	564
A	0.26	x1200 =	312
B	0.19	x1200 =	228
AB	0.08	x1200 =	96

Calculating χ^2

	Collected data		Calculating χ^2		$\frac{(O-E)^2}{E}$
	Observed (O)	Expected (E)	(O - E)	(O - E) ²	
O	501	564	-63	3969	7.0372
A	315	312	3	9	0.0288
B	249	228	21	441	1.9342
AB	135	96	39	1521	15.844
Total:	1200	1200	Total (χ^2) =		24.844

$$\chi^2 = 24.844$$

$$c.v = 7.815$$

$$24.844 > 7.815$$

∴ we reject the null hypothesis



the collected blood types DO NOT fit the data provided by the local health authority

WHY MIGHT THIS BE?

Critical Values table

D.F. (n-1)	p (certainty) (how sure do you want to be?)		
	0.1 (90%)	0.05 (95%)	0.01 (99%)
1	2.70554	3.84146	6.6349
2	4.60517	5.99146	9.21034
3	6.25139	7.81473	11.34487
4	7.77944	9.48773	13.2767
5	9.23636	11.0705	15.08627
6	10.64464	12.59159	16.81189
7	12.01704	14.06714	18.47531
8	13.36157	15.50731	20.09024
9	14.68366	16.91898	21.66599
10	15.98718	18.30704	23.20925

Possible answer:

A blood-drive is carried out on a large university campus and the blood-type data are collected. Local health authorities have the distribution of blood types in that area recorded as:
O = 47%, A = 26%, B = 19%, AB = 8%.

The blood drive received 1200 donors, and their blood types were recorded as thus:
O = 501, A = 315, B = 249, AB = 135.

Local health authority collected data of people native to the area, whereas the student population may include a lot of immigrants with different blood types. Therefore, the LHA's data are not representative of the student population.

There may have been an extrinsic factor increasing the turnout of people with blood types A or B (such as advertising or an appeal for those blood types).

D.F. (n-1)	p (certainty) (How sure do you want to be?)		
	0.1 (90%)	0.05 (95%)	0.01 (99%)
1	2.70554	3.84146	6.6349
2	4.60517	5.99146	9.21034
3	6.25139	7.81473	11.34487
4	7.77944	9.48773	13.2767
5	9.23636	11.0705	15.08627
6	10.64464	12.59159	16.81189
7	12.01704	14.06714	18.47531
8	13.36157	15.50731	20.09024
9	14.68366	16.91898	21.66599
10	15.98718	18.30704	23.20929
11	17.27501	19.67514	24.72497
12	18.54935	21.02607	26.21697
13	19.81193	22.36203	27.68825
14	21.06414	23.68479	29.14124
15	22.30713	24.99579	30.57791
16	23.54183	26.29623	31.99993
17	24.76904	27.58711	33.40866
18	25.98942	28.8693	34.80531
19	27.20357	30.14353	36.19087
20	28.41198	31.41043	37.56623
21	29.61509	32.67057	38.93217
22	30.81328	33.92444	40.28936
23	32.0069	35.17246	41.6384
24	33.19624	36.41503	42.97982
25	34.38159	37.65248	44.3141
26	35.56317	38.88514	45.64168
27	36.74122	40.11327	46.96294
28	37.91592	41.33714	48.27824
29	39.08747	42.55697	49.58788
30	40.25602	43.77297	50.89218

Resources:

Field Studies Council

<http://www.theseashore.org.uk/theseashore/Stats%20for%20twits/Chi-Squared.html>

BMJ (British Medical Journal)

<http://www.bmj.com/collections/statsbk/8.dll>

Extract from

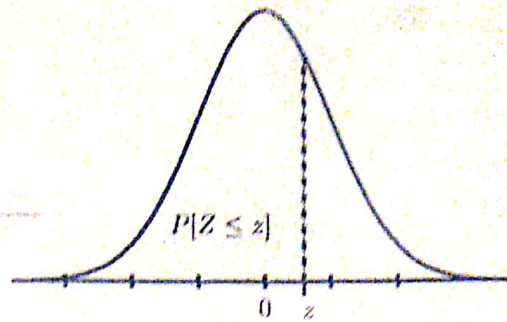
$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

For more help and animations, please visit:

<http://sciencevideos.wordpress.com>

Table N

Aire sous la courbe normale à gauche de z , c'est à dire $P[Z \leq z]$, où $Z \sim N(0, 1)$.



	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09

Chaînes de Markov - Séance Scilab (semaine du Lu. 8 janvier 2018)

I. Modélisation et calcul du "PAGE RANK" (label de Google)

Dans cette partie, nous allons décrire une version simplifiée de la chaîne de Markov utilisée par Google pour calculer l'indice de popularité, ou "Page Rank" - noté PR dans la suite - de chaque page web. L'indice PR est un réel situé entre 0 et 1, et la somme des indices PR de toutes les pages répertoriées vaut 1. Plus l'indice PR d'une page est grand et plus cette page apparaît tôt dans la liste des pages affichées par le moteur de recherche à l'appel d'un mot lié.

On numérote de 1 à K l'ensemble des pages répertoriées. Pour $1 \leq i, j \leq K$, la notation $i \rightarrow j$ signifie que "la page i contient un lien vers la page j ", tandis que $i \nrightarrow j$ signifie que "la page i ne contient pas de lien vers la page j ". On désigne par N_i le nombre de pages vers lesquelles il existe un lien depuis la page i . On impose la convention $i \rightarrow i$, ce qui implique $N_i \geq 1$.

On considère la chaîne de Markov $(T_n)_{n \geq 0}$ à valeurs dans $\{1, \dots, K\}$ qui décrit la position à chaque instant n d'un "random web surfer". Les probabilités de transition sont données par la matrice $P = (P_{ij})_{1 \leq i, j \leq K}$ telle que

$$P_{ij} = \mathbb{P}_{(T_n=i)}(T_{n+1} = j) = \begin{cases} \frac{1-\alpha}{K} & \text{si } i \nrightarrow j \\ \frac{1-\alpha}{K} + \frac{\alpha}{N_i} & \text{si } i \rightarrow j \end{cases}$$

où α est un coefficient entre 0 et 1 qui décrit l'attitude du web surfer (si $\alpha \approx 0$, le web surfer a fortement tendance à sauter d'une page à l'autre sans jamais suivre aucun lien, tandis que si $\alpha \approx 1$, le web surfer suit presque exclusivement les liens existants).

Dans ces conditions, on parvient à démontrer le théorème suivant :

Théorème : Il existe un vecteur $r = (r(1), r(2), \dots, r(K))$ tel que

- $\forall i, r(i) \geq 0$ et $r(1) + \dots + r(K) = 1$ (les $(r(i))_{1 \leq i \leq K}$ décrivent une loi de probabilité)
- $\forall \pi^{(0)} = (\mathbb{P}(T_0 = 1), \mathbb{P}(T_0 = 2), \dots, \mathbb{P}(T_0 = K))$,

$$(\mathbb{P}(T_n = 1), \mathbb{P}(T_n = 2), \dots, \mathbb{P}(T_n = K)) = \pi^{(0)} \times (P)^n \xrightarrow{n \rightarrow +\infty} r$$

(en temps long, asymptotiquement, le web surfer se trouve en i avec probabilité $r(i)$)

- $\forall i$, en notant $S_n(i)$ le nombre de passages en i du web surfer au cours de n étapes,

$$S_n(i)/n \xrightarrow{n \rightarrow +\infty} r(i)$$

(asymptotiquement, la proportion de temps passée en i converge vers $r(i)$)

- $\forall i$, en notant $\tau(i)$ (prononcer "tau") le temps de premier retour en i du web surfer, en sorte que $\tau(i) = \inf\{n > 0; T_n = i\}$, on a

$$\mathbb{E}_{(T_0=i)}(\tau(i)) = 1/r(i)$$

(le temps moyen de retour en i vaut $1/r(i)$)

- r est vecteur propre à gauche de P associé à la valeur propre 1.

Le Page Rank de la page i est donné par $r(i)$ dans un tel modèle, en prenant $\alpha = 0,85$.

II. Sujet du TP. Quelques conseils: utilisez l'aide Scilab, respectez les notations et les noms des variables fournies et pensez à commenter vos instructions. Cette démarche de qualité sera appréciée d'un éventuel lecteur, et vous aidera aussi !

1. Description du réseau de pages

- Choisir une valeur pour K et une valeur pour α (il faudra pouvoir modifier ces valeurs).
- Remplir une matrice appelée *LIEN*, de taille $K \times K$, décrivant les liens entre les pages:

$$LIEN(i, j) = 1 \text{ si } i \rightarrow j; 0 \text{ sinon}$$

La méthode de remplissage est laissée à votre choix (de préférence aléatoire), faites preuve d'imagination! Mais les contraintes suivantes doivent être satisfaites:

- par convention, une page est toujours liée à elle-même,
- vous devez avoir une page "spéciale", différente de la page 1, vers laquelle pointent toutes les autres pages. Retenez le numéro de cette page et commentez votre programme en conséquence.
- Calculer le vecteur $NBRE = (NBRE(1), NBRE(2), \dots, NBRE(K))$ qui fournit les quantités N_1, N_2, \dots, N_K du modèle.
- Déterminer la matrice de transition P du modèle, calculée à partir de *LIEN* et *NBRE*.

2. Calcul du Page Rank

- Vérifier que 1 est la première valeur propre de tP fournie par Scilab.
- Récupérer un vecteur propre de tP associé à la valeur propre 1 et vérifier que toutes ses coordonnées sont de même signe. (*Un tel vecteur a pour transposé un vecteur propre à gauche de P pour la valeur propre 1*).
- Calculer r à partir du vecteur propre obtenu (relire le théorème) en vous rappelant que $r(i) \geq 0$ pour tout i et que $r(1) + \dots + r(K) = 1$.
- Calculer le Page Rank de la page 1 et de votre page "spéciale".

3. Simulation du random web surfer

Vous noterez $T(1), T(2), \dots$ les positions successives du web surfer.

- Choisir une valeur pour $T(1)$ et calculer $T(2), T(3), \dots, T(10)$.
- Dans la suite, on va s'intéresser à la page 1 et à la page "spéciale" numérotée s ici.
- Initialiser le web surfer sur la page 1 et calculer -sans afficher- pour une valeur de N assez grande que vous fixerez,
 - les N premiers termes de la suite $(T(n))_n$,
 - les N premiers termes des suites $(S_n(1)/n)_n$ et $(S_n(s)/n)_n$.
 - Tracer sur un graphique $n \mapsto S_n(1)/n$ et $n \mapsto S_n(s)/n$. Intuire les valeurs limites et les écrire en commentaire dans votre programme. Comparer avec les résultats obtenus à la question 2.
 - *facultatif*: Calculer les valeurs de $\tau(1)$ et de $\tau(s)$ ainsi que $1/\tau(1)$ et $1/\tau(s)$. Commenter en vous reportant au théorème.

Indications pour Scilab:

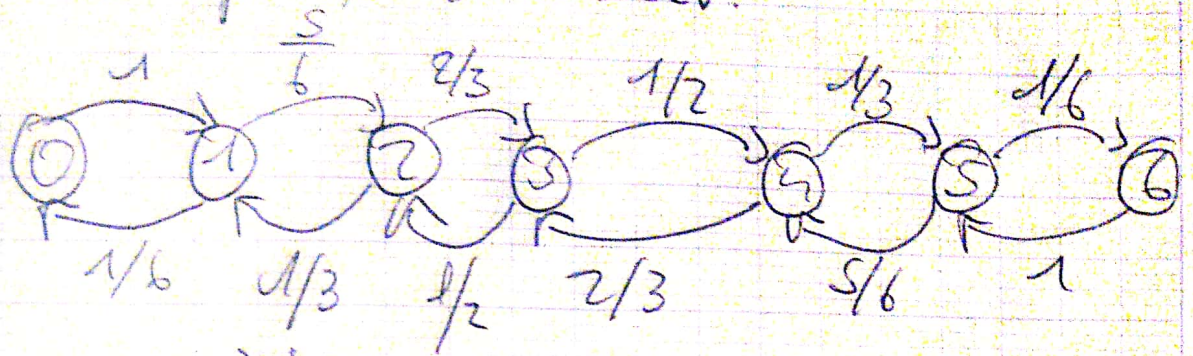
- Si MAT désigne une matrice carrée, l'instruction $[Q, D] = spec(MAT)$ fournit une matrice diagonale D et une matrice de changement de base Q qui permettent de diagonaliser la matrice MAT sous la forme $MAT = Q D Q^{-1}$. En général, les valeurs propres sont des nombres complexes et la matrice Q est aussi à coefficients complexes. Cependant, si une valeur propre est réelle, alors les coordonnées des vecteurs propres associés le sont aussi.
- Pour choisir une valeur au hasard entre 1 et K selon une loi de probabilité donnée par le vecteur ligne $p = (p(1), \dots, p(K))$, utiliser la fonction hasard définie ci-dessous.

```

function x=hasard(p)
  N=length(p); z=rand(); x=1; q(1)=p(1);
  for i=1:N-1,
    if z>q(i) then x=x+1; end
    q(i+1)=q(i)+p(i+1);
  end
endfunction
  
```

source: U.S. Patent 6,285,999 et http://en.wikipedia.org/wiki/Markov_chain

Exemple mesure rev.



$\pi = \bar{\pi}$ (inv) , unique mesure invariante, est elle reversible

$$\text{DBE: } \pi_0 \times 1 = \pi_1 \times \frac{1}{6} \rightsquigarrow \pi_1 = 6\pi_0$$

$$\pi_1 \times \frac{5}{6} = \pi_2 \times \frac{1}{3} \rightsquigarrow \pi_2 = \frac{15}{6}\pi_1 = 15\pi_0$$

$$\pi_2 \times \frac{2}{3} = \pi_3 \times \frac{1}{2} \rightsquigarrow \pi_3 = 20\pi_0$$

$$\rightsquigarrow \pi_4 = 15\pi_0$$

$$\rightsquigarrow \pi_5 = 6\pi_0$$

$$\rightsquigarrow \pi_6 = \pi_0$$

On peut résoudre bcp + rapidement

$$\pi_0 + \pi_0 + 6\pi_0 + 15\pi_0 + 20\pi_0 + 15\pi_0 + 6\pi_0 = 1$$

$$\Leftrightarrow 64\pi_0 = 1$$

$$\pi_0 = \frac{1}{64}$$

$$\Rightarrow \pi \left(\frac{1}{64}, \frac{6}{64}, \frac{15}{64}, \frac{20}{64}, \frac{15}{64}, \frac{6}{64}, \frac{1}{64} \right)$$