Table 1

Bibliographie

[**1]** **J. Bass** : '*Eléments de Calcul des Probabilités*' Masson, 1974

[2] **K. L. Chung, F. Aitshalia** : ' *Elementary probability Theory*' Springer, 2003

[3] **P. Deheuvels** : '*La probabilité, le hasard et la certitude*' Que sais-je, PUF, 1982.

[4] **W. Feller** : ' *An Introduction to Probability Theory and its Applications* ' Tome 1, Wiley, 1966

[5] **D. Foata, A. Fuchs** : '*Calcul des Probabilités*' Dunod, 1998

[6] **B. V. Gnedenko** : '*The Theory of probability*' CRC Press, 1998

[7] **C. Graham** : '*Chaînes de Markov*' Dunod, 2008

[8] **A. Jacquard** : '*Les probabilités*' Que sais-je, PUF, 1974.

[9] **Johnson, Kemp, Kotz** : '*Univariate Discrete Distributions*' Wiley, 2005.

[10] **P. Meunier** : '*Probabilités discrètes*' Cepad, 2015

[11] **H. Ventzell** : '*Théorie des probabilités*' Editions Mir, 1973, 1987

I - Modèles discrets

1.1. Evénements et probabilités

a) Evénements

Les événements peuvent être représentés par des assertions, en langage courant ou de manière plus formelle, représentant des cas finis simples (une pièce tombe sur Pile), ou des situations moins immédiates (un thermomètre indique une température de 20° C, ou la météo a prévu 3 jours de pluie). Ils sont modélisés dans le modèle de Kolmogorov comme des ensembles, parties d'un univers Ω, ensemble de tous les possibles. Cet univers peut être explicité pour représenter des situations assez simples comme des jeux de hasard, ou être implicite dans la plupart des cas. Les deux langages sont utilisés pour exprimer des événements courants, avec une correspondance entre conjonction et disjonction d'une part, intersection et réunion de l'autre.

L'ensemble des événements peut être l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω, ou un sous ensemble $\mathcal{A}$. Cet ensemble $\mathcal{A}$ sera stable pour la réunion, l'intersection (finie ou dénombrable) et pour la complémentation, et contiendra toujours Ω (et donc aussi l'ensemble vide). On dira que $\mathcal{A}$ est une **tribu** de Ω. On dira donc aussi la tribu des événements.

b) Mesure de probabilité

Les évènements sont les objets dont on peut exprimer la probabilité : P est une application de $\mathcal{A}$ dans $\mathbb{R}$ ou dans [0, 1] qui est dite **mesurable** : si A et B sont des événements incompatibles, ou si ce sont des ensembles disjoints,
$P(A \cup B) = P(A) + P(B)$, et plus généralement si $(A_n)_{n \geq 0}$ est une suite d'événements deux à deux disjoints,

$$P\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n \geq 0} P(A_n)$$

Cette propriété est dite de mesurabilité (dénombrable), et l'application P est dite **mesure de probabilité**. C'est la même propriété que pour la mesure des longueurs ou des aires du plan, et on peut prendre cette dernière comme représentation géométrique de la mesure de probabilité des événements.

La principale différence avec la mesure des longueurs est que la mesure de probabilité est finie, et plus précisément $P(\Omega) = 1$, et pour tout événement A, $P(A) \leq 1$

On déduit de la mesurabilité de P que :

i) $P(A^c) = 1 - P(A)$

ii) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

iii) Si $A \subset B$, $P(B - A) = P(B) - P(A) \geq 0$, et $P(A) \leq P(B)$

iv) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$
(Formule du crible, généralisable à un nombre fini quelconque)

v) Si A est un ensemble fini : $A = \{\omega_1, \omega_2, ..., \omega_m\}$, $P(A) = P(\{\omega_1\}) + P(\{\omega_2\}) + ... + P(\{\omega_m\})$

En particulier quand on cherche à calculer en pratique P (A) on cherchera l'expression la plus simple de P (A) et $P(A^c)$
Et si on connait P (A), P (B) et $P(A \cup B)$, il suffira d'écrire $P(A \cap B) = P(A) + P(B) - P(A \cup B)$.

Par exemple, si on lance un dé à 6 faces, on peut choisir comme univers $\Omega = \{1, 2, 3, 4, 5, 6\}$. Cet univers suffit à décrire le jeu ou l'expérience consistant en un seul lancer de dé. On peut évaluer les probabilités suivantes :

P (face différente de 6) = $P(\{1, 2, 3, 4, 5\}) = 1 - P(6) = 5/6$, en écrivant $P(\omega) = P(\{\omega\})$

P (face paire) = $P(\{2, 4, 6\}) = P(2) + P(4) + P(6) = 1/6 + 1/6 + 1/6 = 3/6 = 1/2$

P (face paire ou le numéro de face est un nombre premier) = $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1/2 + 1/2 - 1/6$

Les événements de probabilité nulle seront négligés, et correspondent de fait à des situations qui ne peuvent arriver : Si $P(A) = 0$, A ne peut être observé en pratique. Si A correspond au résultat d'une expérience qu'on peut itérer, on pourra également répéter cette expérience autant de fois qu'on le souhaite sans jamais observer ce résultat.

Pour les questions pratiques, il en sera de même d'événements de très faible probabilité. même si la probabilité de cet événement peut paraitre très faible, si les joueurs de Loto achètent des billets c'est bien dans l'espoir de gagner un jour le gros lot. Qu'en serait-il si cette probabilité était de 10^{-50}, quel que soit le montant des lots ? Il est à parier que peu de personnes jouent.

c) Evénements indépendants et loi conditionnelle

Deux événements A et B sont indépendants si l'occurrence de l'un n'influe pas sur la probabilité de l'autre, ou encore si $P(A/B) = P(A)$, ou de manière équivalente si $P(B/A) = P(B)$.

Si on note P_B la mesure (*) de probabilité sachant B : pour tout événement A, $P_B(A) = P(A/B)$. L'univers, ensemble de tous les possibles, est alors réduit à l'ensemble B, et tout événement A ne peut être considéré que conjointement à B, soit si $P(B) \neq 0$, $P(A/B) = P(A \cap B)/ P(B)$. Par suite, quels que soient les événements A et B,

<u>Définition 1.1</u> A et B sont indépendants si $P(A \cap B) = P(A) \,.\, P(B)$.

Ainsi si on lance deux fois un dé à six faces, avec des lancers indépendants, la probabilité d'avoir deux fois un 6 est de $1/6 \,.\, 1/6 = 1/36$

* <u>Propriété 1.1</u> P_B est une mesure.

<u>Démonstration</u> : $P_B(\Omega) = P(B) / P(B) = 1$.

Si $A \cap C = \varnothing$, : $P_B(A \cup C) = P[(A \cup C) \cap B] / P(B) = P[(A \cap B) \cup (C \cap B)] / P(B)$

et comme $(A \cap B) \cap (C \cap B) = A \cap C \cap B = \varnothing$,

$$P_B(A \cup C) = P(A \cap B) / P(B) + P(C \cap B) / P(B) = P(A/B) + P(C/B) = P_B(A) + P_B(C)$$

Cette propriété d'additivité se généralise immédiatement au cas dénombrable. □

<u>Remarque</u> : Si A et B sont indépendants, $P(A \cup B) = P(A) + P(B) - P(A) \,.\, P(B)$

d) Modèle équiréparti

Si $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ est un ensemble fini d'ordre n, la mesure de probabilité est caractérisée par $(P(\omega_k))_{1 \leq k \leq n}$

Si Toutes les éventualités élémentaires ω_k ont la même probabilité $P(\omega_k) = 1/n$, alors pour tout événement A de cardinalité $|A| = m$, $P(A) = m / n = |A| / |\Omega|$.

Nous verrons au paragraphe 3 avec le modèle hypergéométrique un cas particulier important de ce modèle équiréparti.

1.2. Jeu de Pile ou Face et loi binomiale

a) Modèle de Bernoulli, variable aléatoire et loi de probabilité

Un jeu de Pile ou Face peut être identifié à n'importe quelle situation ou phénomène aléatoire où un résultat peut prendre deux formes distinctes, comme l'observation d'une face paire ou impaire d'un dé, ou une valeur binaire aléatoire. Si on veut représenter Pile et Face par des nombres la façon la plus naturelle - ou du moins la plus simple - est de le faire par 0 et 1. Ce schéma prend le nom de **Bernoulli**, du nom d'un fameux mathématicien suisse. Le jeu, ou l'expérience aléatoire, est caractérisé par la probabilité de Pile, ou de 1. Pour le cas d'une pièce équilibrée, cette probabilité est $p = 1/2$. Il est clair que la probabilité de Face est alors $q = 1 - p$.

Dans le modèle de Bernoulli général B (p), cette valeur est un paramètre a priori quelconque de [0, 1].

Ainsi, le jeu de Pile ou Face peut être caractérisé par une variable X qui peut prendre ces deux valeurs, 0 et 1. Cette variable prend le nom de **variable aléatoire**. On peut représenter l'expérience par un petit tableau

$X = 1$	p
$X = 0$	q

Ce tableau résume les valeurs possibles et le probabilités de ces valeurs, c'est à dire la loi de probabilité de la variable aléatoire X.

Remarque : Bien que ce soit à ce stade non nécessaire, on peut donner un sens mathématique à la notion de variable aléatoire :

Définition 1.2 Une variable aléatoire réelle est une application de Ω dans $\mathbb{R}$ telle que pour tout ouvert U de $\mathbb{R}$, $E = X^{-1}(U)$ est un événement de $\mathcal{A}$. Cette variable aléatoire X est dite discrète si son ensemble image est fini ou dénombrable. La loi de probabilité de la variable aléatoire discrète X est l'ensemble des couples (x, p) de $\mathbb{R} \times [0, 1]$ tels que x est un élément de l'image de X, et $p = \mathrm{P}(X = x) > 0$.

La condition sur E permet de donner un sens aux probabilités $\mathrm{P}(X \in B) = \mathrm{P}(X^{-1}(B))$, où B est un ouvert, ou plus généralement un borélien* de $\mathbb{R}$.

La loi de probabilité d'une variable aléatoire discrète finie peut être présentée par un tableau :

$X =$	x_1	p_1
	x_2	p_2
	...	
	x_n	p_n

On notera que la somme de ces probabilités est toujours égale à 1 !

Pour une loi de probabilité infinie il s'agira d'un tableau également infini, ou simplement de la donnée des valeurs possibles x_k et des probabilités génériques $p_k = \mathrm{P}(X = x_k)$

Dans ce cas $\sum_{k=1}^{+\infty} p_k$ est une série convergente de somme 1.

* Les boréliens de $\mathbb{R}$ sont les parties qui contiennent les intervalles, et tous les sous ensembles obtenus par réunions, intersections (finies ou dénombrables) et par complémentations successives. Ce sont les parties assez 'régulières' pour pouvoir en définir la mesure.

b) Variables aléatoires indépendantes

Définition 1.3 Deux variables aléatoires X et Y sont indépendantes si pour deux parties (boréliens) quelconques B_X et B_Y de $\mathbb{R}$, les événements $A_X = X^{-1}(B_X) = (X \in B_X)$ et $A_Y = Y^{-1}(B_Y) = (Y \in B_Y)$ sont indépendants.

Cela signifie que X et Y prennent leurs valeurs indépendamment l'une de l'autre.

Exemple : suite de tirages de Bernoulli indépendants : $X_1, X_2, ..., X_n$. Si $p = 1/2$, la probabilité d'avoir deux Pile sur les deux premiers tirages est 1/4.

c) Loi binomiale et applications

Dénombrements

Permutations : Le nombre de permutations (ou de bijections) d'un ensemble à n éléments (dans lui-même) est exactement n ! : n places pour le premier élément, puis $n - 1$ pour placer le deuxième, ...

Arrangements : Le nombre d'arrangements de k éléments parmi n est $A_n^k = n(n-1)(n-2) ... (n-k+1)$

Combinaisons : Le nombre de combinaisons de k éléments parmi n est $C_n^k = A_n^k / k! = \dfrac{n!}{(n-k)!\,k!}$

Les combinaisons successives peuvent se calculer par produits, mais aussi par sommes d'après la relation :
$C_{n+1}^{k+1} = C_n^{k+1} + C_n^k$ (qui permet d'écrire le triangle de Pascal).

Pour les grandes valeurs de n (pour $n > 50$) on peut approcher $n!$ à l'aide le la formule de Stirling :

$$n! = \sqrt{2\pi n}\ \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12\,n} + o\left(\frac{1}{n}\right)\right) \sim \sqrt{2\pi n}\ \left(\frac{n}{e}\right)^n$$

Remarque : On peut pour éviter des overflow de calcul employer le logarithme (décimal ou népérien) pour calculer séparément exposant et mantisse de n!

Loi binomiale

Si on compte le nombre de Pile sur n tirages Pile ou Face, ou le nombre de 1 sur une suite de n variables de Bernoulli indépendantes B (p), on obtient une variable aléatoire suivant une **loi binomiale** (en abrégé B (n, p)).

Si on note $X_1, X_2, ..., X_n$ les variables aléatoires de Bernoulli ($P(X_k = 1) = p$), et si Y désigne le nombre de 1, on aura, pour $0 \leq k \leq n$, $P(Y = k) = C_n^k\ p^k\, q^{n-k}$

En effet si on obtient k Pile (1) sur n tirages on doit avoir également $n - k$ Face (0), et comme on ne tient pas compte de l'ordre il y a C_n^k façons de ranger les k Pile sur les n tirages

On peut également noter la relation algébrique entre ces variables aléatoires : $Y = X_1 + X_2 + ... + X_n$

Comme on l'a déjà noté ce calcul s'applique à tous les schémas de Bernoulli itérés.

* Sur $n = 100$ tirages, la probabilité de 50 Pile est : $P(Y = 50) = C_{100}^{50}\ \left(\frac{1}{2}\right)^{50}\left(\frac{1}{2}\right)^{50} = C_{100}^{50}\ \left(\frac{1}{2}\right)^{100}$

On peut faire un calcul exact, ou approché à l'aide de la formule de Stirling : si $n = 2k$,

$$P(Y = k) \sim \frac{\sqrt{4\pi}\ \left(\frac{2k}{e}\right)^{2k}}{\left(\sqrt{2\pi k}\ \left(\frac{k}{e}\right)^{k}\right)^2}\left(\frac{1}{2}\right)^{2k} = \frac{\sqrt{4\pi k}\ \left(\frac{2k}{e}\right)^{2k}}{2\pi k\ \left(\frac{k}{e}\right)^{2k} 2^{2k}} = \frac{1}{\sqrt{\pi k}}$$

donc ici, $P(Y = 50) \approx 0.0798$

* Si on lance 6 dés à 6 faces, quelle est la probabilité d'avoir exactement deux 6 ? Ici $p = 1/6$, et donc

$$P(Y=2) = C_6^2 \left(\tfrac{1}{6}\right)^2 \left(\tfrac{5}{6}\right)^4 \approx 0.201$$

* Sur un fichier binaire équiprobable de 1 ko, quelle est la probabilité qu'il y ait autant de 0 que de 1 ?

Ici $n = 8192$, $p = 1/2$ et $P(Y = 4096) \approx 8.8\ 10^{-3}$

1.3. Modèle hypergéométrique et jeux (cartes, Loto, ...)

Si on joue au Loto de la française des jeux, 6 boules sont tirées dans une urne qui en contient 49. L'ordre du tirage n'est pas pris en compte, et toutes les combinaisons de 6 numéros par les 49 possibles sont équiprobables. On est donc dans le cadre d'un modèle équiréparti, et pour tout événement A, $P(A) = \frac{|A|}{C_{49}^6}$

Pour on nombre N de bons numéros : Probabilité de gagner le gros lot $P(N = 6) = \frac{1}{C_{49}^6} \approx 7.15\ 10^{-8}$

$$P(N=5) = \frac{C_6^5\, C_{43}^1}{C_{49}^6} \approx 1.85\ 10^{-5}$$

En effet il y a exactement C_6^5 possibilités pour choisir les 5 bons numéros parmi les 6, et C_{43}^1 façons de choisir le sixième parmi les 43 restants, ce qui donne pour l'événement considéré exactement $C_6^5\, C_{43}^1$ combinaisons.

De même, $P(N=4) = \frac{C_6^4\, C_{43}^2}{C_{49}^6} \approx 9.868\ 10^{-3}$

et $P(N=3) = \frac{C_6^3\, C_{43}^3}{C_{49}^6} \approx 1.765\ 10^{-2}$

Bien sûr si on change de jeu en conservant la même approche, le calcul sera le même.

Si on tire 5 cartes sur un jeu de 52 cartes, quelle est la probabilité d'avoir exactement trois as ?

Si on note encore N le nombre d'as, $P(N=3) = \frac{C_4^3\, C_{48}^2}{C_{52}^5} \approx 1.736\ 10^{-3}$

Le problème général traité ici peut s'énoncer de la manière suivante : Une population M se décompose en deux groupes, de nombres respectifs M_1 et M_2. On tire au hasard k individus dans la population, et on écrit la probabilité p d'avoir sur ces k individus $N = r$ individus du premier groupe :

$$p = P(N=r) = \frac{C_{M_1}^r\, C_{M_2}^{k-r}}{C_M^k}$$

Il s'agit d'un modèle hypergéométrique, et la variable aléatoire N suit la loi hypergéométrique $H(M, M_1, k)$

1.4. Formule de Bayes et calcul de probabilités

a) Enoncé et démonstration

On calcule la probabilité d'un événement A, à partir deus seules probabilités connues, qui sont des probabilités conditionnelles.

On considère une suite d'événements $B_1, B_2, ..., B_m$ formant une partition de l'univers Ω :

$B_1 \cup B_2 \cup ... \cup B_m = \Omega$ et pour tout couple $(j, k), j \neq k$, $B_j \cap B_k = \emptyset$.

Cela signifie que l'un et l'un seulement des événements B_k est vérifié

La relation suivante prend le nom de Formule de *Bayes* :

Proposition 1.2 $P(A) = P(A / B_1) . P(B_1) + P(A / B_2) . P(B_2) + ... + P(A / B_m) . P(B_m)$

Démonstration : $P(A) = P(A \cap \Omega) = P[A \cap (B_1 \cup B_2 \cup ... \cup B_m)] = P[(A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_m)]$

et comme ces ensembles sont disjoints, $P(A) = P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_m)$ (1)

Pour des événements A et B tels que $P(B) \neq 0$, $P(A / B) = \frac{P(A \cap B)}{P(B)}$, donc $P(A \cap B) = P(A / B) . P(B)$ (2)

Si $P(B) = 0$, la relation précédente reste valide à condition de pouvoir définir $P(A / B)$

En appliquant (2) aux ensembles B_k et en réécrivant (1) on obtient la relation cherchée. □

b) Exemple : tirage itérés

Cette formule peut être appliquée à tous les jeux en deux étapes, et à des problèmes particuliers, comme à l'étude des chaines de Markov (par. 8)

Exemple : On considère 2 urnes dont on connait la composition, avec des boules rouges, bleues ou jaunes :

Urne U1 : 3 boutes Rouges; 4 Bleues et 2 Jaunes

Urne U1 : 5 boutes Rouges; 2 Bleues et 3 Jaunes

On tire 2 boules dans U1 pour les placer dans U2, puis on tire une boule dans U2.

Quelle est la probabilité que cette boule soit rouge ? L'exercice est laissé au lecteur

1.5. Moyennes

a) Espérance mathématique d'une loi discrète finie

Soit une variable aléatoire X pouvant prendre les valeurs $x_1, x_2, ..., x_n$ avec les probabilités respectives $p_1, p_2, ..., p_n$

Si on s'intéresse à la valeur moyenne de X, celle -ci dépend des valeurs x_k, mais elle doit aussi tenir compte des probabilités p_k, pour qu'une valeur très probable influe davantage sur cette moyenne qu'une valeur moins probable. L'espérance mathématique de X est justement la somme des valeurs, pondérée par les probabilités :

Définition 1.4 Si X est une variable aléatoire discrète, prenant les n valeurs $x_1, x_2, ..., x_n$ avec les probabilités $p_1, p_2, ..., p_n$, l'espérance Mathématique (ou espérance) de X est $E(X) = x_1 p_1 + x_2 p_2 + ... + x_n p_n$

Proposition 1.3 Si X est définie comme ci-dessus, et si $Y = \varphi(X)$, $E(Y) = \varphi(x_1) p_1 + \varphi(x_2) p_2 + ... + \varphi(x_n) p_n$

Pour une variable aléatoire Presque Surement constante, soit de loi de probabilité définie par $P(X = c) = 1$, $E(X) = c$.

Si X est une variable aléatoire équidistribuée, avec $P(X = x_k) = \frac{1}{n}$, $1 \leq k \leq n$, $E(X) = \frac{1}{n} \sum_{k=1}^{n} x_k$

Pour une variable de Bernoulli $X : B(p)$, $E(X) = p$

Pour une pièce équilibrée, $E(X) = 1/2$; Pour un dé à 6 faces, $E(X) = 1/6 (1 + 2 + 3 + 4 + 5 + 6) = 1/6 . 6 . 7/2 = 7/2$

b) Variance et écart-type

La variance d'une variable aléatoire X caractérise un écart moyen à la moyenne, ou une forme de dispersion de ses valeurs. La définition mathématique que nous allons donner ici pour variance et écart type se justifie par ses applications !

Définition 1.5 Si X est une variable aléatoire finie, sa variance est $V(X) = E([X - E(X)]^2) = E(X^2) - [E(X)]^2$
L'écart type de X est $\sigma_X = \sqrt{V(X)}$

Pour une variable $B(p)$, $V(X) = p\,q = p(1-p)$, et pour le jeu de Pile ou Face, $\sigma_X = 1/2$

Pour un dé à 6 faces,

$V(X) = 1/6\,(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - (7/2)^2 = 1/6\,.\,6\,.\,7\,.\,13/6 - (7/2)^2 = 91/6 - 49/4 = 35/12$

c) Cas d'un modèle dénombrable; loi géométrique et loi de Poisson

Si X est une variable aléatoire discrète infinie, ses moyennes, comme l'espérance ou la variance, sont définies comme des séries qui peuvent converger ou diverger. Ainsi l'espérance mathématique n'existera pas toujours.

Si la série $\sum_{k=1}^{+\infty} x_k\, p_k$ est absolument convergente, on pourra écrire, comme dans le langage des intégrales,

$E(|X|) < \infty$, ou encore $X \in L^1(\Omega)$, et dans ce cas l'espérance de X, $E(X) = \sum_{k=1}^{+\infty} x_k\, p_k$, est définie.

Si la série $\sum_{k=1}^{+\infty} x_k^2\, p_k$ est convergente, $E(X^2) < \infty$, ou encore $X \in L^2(\Omega)$, et dans ce cas la variance de X est définie.

En effet,

Théorème 1.4 $L^1(\Omega) \subset L^2(\Omega)$

Exemples :

1. Loi géométrique. Si la variable aléatoire Z désigne l'ordre du premier 1 dans une suite de tirages de Bernoulli $B(p)$ indépendants, les valeurs de Z sont les entiers positifs, et pour $k \in N^*$, $p_k = P(Z = k) = p\,q^{k-1}$

Z est dite variable géométrique de paramètre p, et sa loi est écrite en abrégé $G(p)$.

Espérance : $E(Z) = \sum_{k=1}^{+\infty} k\, p_k = p \sum_{k=1}^{+\infty} k\, q^{k-1} = p\, S'(q)$, où $S(x) = \sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}$, pour $-1 < x < 1$

Si $q = 1$, L'apparition d'un 0 est certaine à chaque tirage, et la loi $G(p)$ n'est pas définie. On suppose donc $q \neq 1$.

Par suite, $E(Z) = p\left(\frac{1}{1-q}\right)^2 = \frac{1}{p}$

Ainsi pour une suite de jeux de Pile ou Face, avec $p = 1/2$, $E(Z) = 2$.

On peut calculer la variance de Z. $E(Z^2) = \sum_{k=1}^{+\infty} k^2\, p\, q^{k-1} = p \sum_{k=1}^{+\infty} k(k-1)\, q^{k-1} + E(Z)$

$\sum_{k=1}^{+\infty} k(k-1)\, q^{k-1} = q \sum_{k=2}^{+\infty} k(k-1)\, q^{k-2} = q\, S''(q) = 2q\left(\frac{1}{1-q}\right)^3$

Par suite, $E(Z^2) = \frac{2q}{p^2} + \frac{1}{p} = \frac{q+1}{p^2}$, et $V(Z) = \frac{q+1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$

2. Loi de Poisson

La loi de Poisson a de nombreux usages en théorie des probabilités, et dans certaines applications (files d'attente , ...)

Loi $\mathcal{P}(\lambda)$ (de paramètre $\lambda > 0$) : Pour $k \in N$, $P(X = k) = e^{-\lambda}\,.\,\frac{\lambda^k}{k!}$

Espérance : $E(X) = \sum_{k=0}^{+\infty} k\, e^{-\lambda}\,.\,\frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{+\infty} k.\,\frac{\lambda^k}{k!} = e^{-\lambda}.\sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} = \lambda\,.\,e^{-\lambda}.\sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!}$

$$= \lambda\,.\,e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = \lambda$$

Variance : $V(X) = \sum_{k=0}^{+\infty} k^2\, e^{-\lambda}\,.\,\frac{\lambda^k}{k!} - \lambda^2 = \lambda$

3. Dans les exemples précédents l'espérance et la variance existent. Si $P(X = k) = \frac{6}{\pi^2 k^2}$, $k > 0$, il s'agit bien d'une loi de probabilité (infinie) car la somme des probabilités est égale à 1, et son espérance mathématique n'existe pas.

d) Propriétés de l'espérance mathématique

Les propriétés générales sont valides pour des lois discrètes (finies ou infinies) mais le seront également pour des lois continues ; Lorsque les espérances considérées existent :

i) Linéarité : $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$, et par suite

ii) pour tout entier n pour toute suite de n variables aléatoires, $E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$

iii) Si $X \geq 0$, $E(X) \geq 0$, et si $E(X^2) = 0$, $X = 0$ P.S.

iv) Si X et Y sont des variables aléatoires indépendantes, $E(X\,Y) = E(X)\,E(Y)$

v) Si X et Y sont des variables aléatoires indépendantes, $V(X + Y) = V(X) + V(Y)$

vi) $V(\alpha X) = \alpha^2\, V(X)$

<u>exercice</u> : Démontrer les propriétés i) à vi)

On peut également considérer $L^2(\Omega)$ comme un espace vectoriel normé, pour la norme euclidienne associée au produit scalaire : $\langle X, Y \rangle = E(X\,Y)$

On aura alors $\|X\|_2 = \sqrt{E(X^2)}$

L'inégalité de Cauchy Schwarz s'écrit $|E(X\,Y)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$, ou encore $E(|X\,Y|) \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$

Ce qui montre que si X et Y sont dans $L^2(\Omega)$, $E(X\,Y)$ existe.

Par ailleurs, si $X \in L^2(\Omega)$, et si on pose $Y = 1$, $E(|X\,Y|) = E(|X|) \leq \sqrt{E(X^2)} < \propto$, et donc $X \in L^1(\Omega)$, ce qui démontre le théorème 1.4

e) Calculs et applications (complexité d'algorithmes)

Pour étudier la complexité d'un problème ou d'un algorithme qui se comporte de manière aléatoire on peut en calculer la valeur moyenne, et éventuellement observer la dispersion par rapport à cette moyenne. L'aspect sommatoire de l'espérance (et de la variance dans le cas indépendant) permet alors de décomposer des problèmes complexes en problèmes plus simples.

Si on prend l'exemple type des algorithmes de tri, on peut étudier 3 classes de tri que sont le tri par insertion (présumé 'lent'), le tri rapide et le tri fusion. Le comportement du tri rapide ne dépend pas des entrées, et à peu de choses près (ou si la taille n du tableau est une puissance de 2) on peut indiquer exactement le nombre de comparaisons en fonction de n. C'est également le cas des procédures de traitement de matrices pleines.

Par contre le tri par insertion et le tri fusion dépendent des données d'entrée, et on peut essayer d'en déterminer la complexité moyenne, représentée par le nombre de comparaisons, et/ou le nombre d'échanges. On pourrait pour les procédures récursives ajouter le nombre d'appels des fonctions.

De fait la seule hypothèse qui permettra ces calculs est que les données du tableau jouent des rôles parfaitement symétriques, ou que tous les ordres sont a priori équiprobables.

Si on prend le cas du tri par insertion, Le nombre total C de comparaisons se décompose et la somme du nombre C_1 de comparaisons pour trier le premier élément du tableau, du nombre C_2 de comparaisons pour trier le deuxième élément du tableau, de C_3 pour trier le 3ème élément, jusqu'à C_n pour trier le dernier élément.

Si C peut être considéré comme une variable aléatoire qu'on souhaite évaluer par le calcul de E (C), alors comme $C = C_1 + C_2 + ... + C_n$, $E(C) = E(C_1) + E(C_2) + ... + E(C_n)$

Et il suffit pour cela de connaitre la loi de probabilité des variables aléatoires C_k, ce qui est plus simple que de connaitre celle de C.

On peut indiquer les nombres de comparaison minimum et maximum :

$C_{\min}$ correspond où le tableau est déjà trié, et à chaque étape une seule comparaison suffit : $C_{\min} = n - 1$

$C_{\max}$ correspond où le tableau est trié en sens inverse. Il faut comparer chaque élément à tous ceux qui le précèdent : $C_{\max} = 1 + 2 + ... + n - 1 = \frac{n(n-1)}{2}$

Pour écrire la loi de C il faudrait connaître toutes les probabilités, pour C variant de $C_{\min}$ à $C_{\max}$

Loi de probabilité de C_k :

$C_k = 1$ $\quad p_1 = 1/k$

$C_k = 2$ $\quad p_1 = 1/k$

...

$C_k = k - 2$ $\quad p_{k-2} = 1/k$

$C_k = k - 1$ $\quad p_{k-1} = 2/k$

En effet, si le rang r dans le tableau trié du k-ème élément (T_k) du tableau est strictement supérieur à 2, il faudra exactement $k - r + 1$ comparaisons, avec donc une probabilité de $1/k$, mais si ce rang est 1 ou 2 il faudra comparer T_k à tous les éléments qui le précèdent, soit $k - 1$ comparaisons, nombre pour lequel on aura donc la probabilité $2/k$.

$$E(C_k) = 1 \cdot \frac{1}{k} + 2 \cdot \frac{1}{k} + ... + (k-2) \cdot \frac{1}{k} + (k-1) \cdot \frac{2}{k} = \frac{1}{k}(1 + 2 + ... + k - 1) + \frac{k-1}{k} = \frac{k-1}{2} + \frac{k-1}{k} = \frac{k}{2} + \frac{1}{2} - \frac{1}{k}$$

Soit $E(C) = \sum_{k=2}^{n} \left(\frac{k}{2} + \frac{1}{2} - \frac{1}{k}\right) = \sum_{k=1}^{n} \left(\frac{k}{2} + \frac{1}{2} - \frac{1}{k}\right) = \frac{n(n+1)}{4} + \frac{n}{2} - H_n = \frac{n(n+3)}{4} - H_n$

En notant $H_n = 1 + 1/2 + ... + 1/n$ (on rappelle que $H_n \sim \log(n)$)

Il s'agit d'une valeur exacte, valide pour de petits tableaux. Si on ne traite que de grands tableaux on aura :

$$E(C) \sim \frac{n^2}{4}$$

Mais c'est plutôt à éviter, car il ne s'agit pas d'un tri rapide ...

On pourrait également calculer la variance de C, pour avoir une idée plus précise de la dispersion autour de sa moyenne (valeurs souvent assez proches de la moyenne, ou variations importantes, entre $C_{\min}$ et $C_{\max}$)

Annexe Table des espérance et variance des lois discrètes usuelles

Loi de Bernoulli X : B (p) $\quad E(X) = p$, $V(X) = p\,q$

Loi Binomiale Y : B (n, p) $\quad E(Y) = n\,p$, $V(Y) = n\,p\,q$

Loi géométrique Q : G (p) $\quad E(Q) = \frac{1}{p}$, $V(Q) = \frac{q}{p^2}$

Loi de Poisson Z : $\mathcal{P}(\lambda)$ $\quad E(Z) = \lambda$, $V(Z) = \lambda$

Remarque : Pour la loi binomiale, comme elle peut s'écrire comme somme de variables de Bernoulli indépendantes, il suffit de sommer les espérances et les variances de ces variables.

1.6. Suite de Variables Aléatoires, convergence et Loi des grands nombres

a) Convergence en loi et en probabilité

Si on considère une suite de variables aléatoires, soit une suite de fonctions, on peut envisager plusieurs notions de convergence, comme la convergence simple (qui sera de fait la convergence Presque Sûre), une convergence en moyenne ou une convergence en mesure (ici la convergence en probabilité). Le convergence uniforme n'est pas utilisée pour des variables aléatoires, mais une notion de convergence importante est la convergence en loi.

Convergence en probabilité

Soit une suite de variables aléatoires (pas nécessairement discrètes) $X_1, X_2, ...$

Définition 1.6 La suite (X_n) converge en probabilité vers une variable aléatoire X (on notera $X_n \overset{P}{\rightarrow} X$) si pour tout $\varepsilon > 0$, la suite de nombres réels positifs $P(|X_n - X| > \varepsilon)$ tend vers 0.

Cela signifie que pour n grand, X_n a une probabilité élevée d'être proche de X.

Convergence en Loi

Dans ce cas seule la loi de probabilité est considérée. On devra spécifier ici que les lois sont discrètes, car dans le cas général il faudra faire appel à la fonction de répartition.

Soit une suite de variables aléatoires discrètes $X_1, X_2, ...$

Définition 1.6 La suite (X_n) converge en loi vers une variable aléatoire discrète X (on notera $X_n \overset{Loi}{\rightarrow} X$) si pour tout $x \in \mathbf{R}$, il existe une suite (x_n) convergeant vers x telle que la suite $P(X_n = x_n)$ converge vers $P(X = x)$.

Remarque : Si les variables aléatoires X_n et X prennent les mêmes valeurs x_k, $X_n \overset{Loi}{\rightarrow} X$ si et seulement si pour tout k, $P(X_n = k) \rightarrow P(X_n = k)$

b) Inégalité de Bienaymé - Tchebychev

Soit une variable aléatoire discrète X de carré intégrable ($X \in L^2(\Omega)$),

Théorème 1.5 Pour tout nombre réel positif c, $P(|X| \geq c) \leq \dfrac{E(X^2)}{c^2}$

La démonstration est laissée en exercice.

Exemple Si X suit la loi $G(\frac{1}{2})$, $E(X) = 2$, $E(X^2) = 6$, et $V(X) = 2$

$P(|X-2| \geq c) \leq \dfrac{V(X)}{c^2}$, et pour $c = 4$, $P(|X-2| \geq c) = P(X \geq 6) \leq \dfrac{2}{4^2} = 0.125$

Mais dans le cas de la loi géométrique $G(p)$ on peut calculer exactement cette valeur en écrivant la somme de la série géométrique : $P(X \geq k) = q^{k-1}$

Soit pour $k = 6$, $P(X \geq 6) = 0.0325$

L'inégalité peut souvent comme ici être assez loin de la valeur exacte. On va voir dans le paragraphe suivant une application importante, que nous retrouverons dans un contexte plus général au chapitre II.

c) Loi de grands nombres : théorie et pratique

On considère ici une suite (X_n) de variables aléatoires de même loi de probabilité de carré intégrable. On note μ l'espérance commune de ces variables aléatoires, et σ leur écart type.

Si on note S_n la moyenne arithmétique des n premières variable :

$$S_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Théorème 1.6 $\quad S_n \overset{P}{\to} \mu$

Démonstration : Soit $\varepsilon > 0$, $P(|S_n - \mu| > \varepsilon) \leq \dfrac{E(S_n - \mu)^2)}{\varepsilon^2}$

$$S_n - \mu = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)$$

Comme $E(S_n - \mu) = 0$, $E[(S_n - \mu)^2] = V(S_n - \mu) = \frac{1}{n^2}\sum_{k=1}^{n} V(X_k - \mu) = \frac{1}{n^2}\sum_{k=1}^{n} V(X_k) = \frac{n\,\sigma^2}{n^2} = \frac{\sigma^2}{n}$

Par suite, $P(|S_n - \mu| > \varepsilon) \leq \dfrac{\sigma^2}{n\,\varepsilon^2} \to 0$ □

La loi des grands nombres signifie que les moyennes des différentes itérations d'un même phénomène aléatoire finiront toujours par se comporter comme sa moyenne statistique.
Si on prend la loi de Bernoulli, ou le jeu de Pile ou Face, elle signifie que la fréquence relative de Pile tendra toujours cers 1/2, quelle que soit la séquence initiale. Bien sûr cela ne signifie pas qu'il y aura autant de Pile que de Face.

Si on revient sur l'analyse d'algorithmes, elle permet de justifier l'emploi de simulations de type Monte Carlo : au lieu de faire le calcul de l'espérance comme on l'a fait on peut programmer le tri par insertion, l'appliquer à des tableaux aléatoires indépendants de même taille n, et mettre un compteur des comparaisons. Si on le fait un grand nombre de fois et qu'on fait la moyenne on aura une valeur approchée de $E(C)$. Evidemment on ne peut faire cela pour tout n, mais à partir de certaines valeurs on peut interpoler une courbe qui permet d'étudier le comportement asymptotique de l'algorithme.

1.7. Loi binomiale et loi de Poisson; fonction génératrice et calculs pratiques

a) Convergence vers une loi de Poisson

On peut citer un exemple élémentaire de convergence en loi, qui relie la loi binomiale et la loi de Poisson.

Théorème 1.7 Si Y_n est une suite de variables aléatoires binomiales $B(n, p_n)$ telles que la suite de nombres réels positifs $n \,.\, p_n$ converge vers λ, la suite Y_n converge en loi vers une variable aléatoire Z suivant la loi de Poisson $\mathcal{P}(\lambda)$.

Démonstration : Soit un entier k, et on suppose dans ce qui suit que $n \geq k$

D'après la définition $P(Y_n = k) = C_n^k\ p_n^k\,(1-p_n)^{n-k} = \frac{C_n^k}{n^k}\,.\,(n\,p_n)^k\,(1-p_n)^{n-k} = \frac{A_n^k}{n^k\,k!}\,.\,(n\,p_n)^k\,(1-p_n)^{n-k}$

Ici k est fixé, et n tend vers l'infini, et $\frac{A_n^k}{n^k} = \frac{n-1}{n}\,.\,\frac{n-2}{n}\,...\,\frac{n-k+1}{n} \to 1$

De plus comme $n\,p_n \to \lambda$, $n\,p_n \sim \lambda$, et $p_n \sim \frac{\lambda}{n} = O\left(\frac{1}{n}\right)$,

$$P(Y_n = k) \sim \frac{A_n^k}{n^k}\,.\,\frac{\lambda^k}{k!}\,.\,(1-p_n)^{n-k} \sim \frac{\lambda^k}{k!}\,.\,(1-p_n)^{n-k}$$

Par ailleurs, $\log[(1-p_n)^{n-k}] = (n-k)(-p_n + o(p_n)) = (n-k)(-p_n + o\left(\frac{1}{n}\right)) = -n\,p_n + k\,p_n + o(1) \to -\lambda$

Donc $(1-p_n)^{n-k} \to e^{-\lambda}$, et $P(Y_n = k)\ {}^k \to e^{-\lambda}\,.\ \frac{\lambda^k}{k!} = P(Z = k)$ □

b) Evaluation de coefficients binomiaux et lois approchées

Le théorème 1.7 peut avoir une utilisation pratique pour évaluer une probabilité du type $P(Y = k)$ ou $P(Y \in B)$, où Y est une variable binomiale $B(n, p)$, pour n 'grand' (supérieur à 100) et p 'petit' de telle sorte que np soit d'un ordre de grandeur de 1. Dans ce cas on pourra considérer que $Y = Y_n$ est le terme d'ordre n d'une suite de variables binomiales convergeant vers une variable aléatoire de Poisson Z suivant la loi $\mathcal{P}(\lambda)$, avec $\lambda = n\,p$.

On pourra alors écrire $P(Y = k) = C_n^k\ p^k (1-p)^{n-k} \approx e^{-\lambda}\,.\ \frac{\lambda^k}{k!}$

Pour de faibles valeurs de k cela évite de calculer A_n^k, et d'évaluer simplement $k!$ (à condition de disposer d'un calculateur qui évalue les exponentielles).

Par contre si k est grand, par exemple de l'ordre de $n/2$ (si k est de l'ordre de n, $C_n^k = C_n^{n-k}$) il faudra de toute manière un outil pour calculer k !, plus éventuellement $n!$ et $(n-k)!$

On a déjà vu que ces calculs étaient possibles en utilisant le logarithme pour éviter des dépassements machine. On peut approcher les factorielles par la formule de Stirling (si n et k sont supérieurs à 50), ou écrire $\log(n!) = \log(2) + \log(3) + ... + \log(n)$. L'avantage de la formule de Stirling est de nécessiter une simple calculatrice électronique.

Prenons un exemple. Pour calculer $m = 1000!$, on peut utiliser sa calculatrice directement, ou programmer le calcul des factorielles sur un outil informatique quelconque, et entrer la valeur 1000. Il y a peu de chances que cela donne quelque chose de très utile ...

Maintenant écrivons $\log_{10}(m) = \frac{\log(m)}{\log(10)} = \frac{1}{\log(10)}\,.\,[\frac{\log(2\pi n)}{2} + n(\log(n) - 1)] \approx 2567.6046$

et $10^{\,0.6046} \approx 4.02$, donc m sera écrit en notation scientifique : $m = 1000! \approx 4.02\ 10^{2567}$

c) Calculs pour un modèle de Bernoulli itéré

On a déjà remarqué que la loi binomiale et la loi géométrique pouvaient être associées à une suite de jeux de Pile ou Face indépendants, ou de manière plus générale à des variables aléatoires de Bernoulli indépendantes de même paramètre p.

1. Prenons un premier exemple : sur un ouvrage imprimé de $n = 400$ pages il peut y avoir des erreurs typographiques, qui peuvent être supposées indépendantes

Si on suppose une faible probabilité d'erreurs, disons $p = 0.01$ par page, on peut poser des questions du genre : quelle est la probabilité qu'il y ait au plus 6 erreurs en tout, ou qu'il n'y en ait pas dans le premier chapitre qui comporte 100 pages.

Un calcul approché peut utiliser l'hypothèse d'au plus une erreur par page. En effet d'après l'indépendance il y aura une probabilité $\frac{p^2}{2} = 5\ 10^{-5}$ de 2 erreurs dans une page, ou $\frac{p^3}{6} = 1.6\ 10^{-7}$ de 3 erreurs, ... *

Dans ce cas on pourra considérer un modèle de Bernoulli $B(p)$, et le nombre d'erreurs suivrait une loi binomiale $Y : B(400, p)$, et l'ordre de la première erreur une loi géométrique $Q : G(p)$.

$P(Y \le 6) = C_{400}^0\, q^{400} + C_{400}^1\, p\, q^{399} + ... + C_{400}^6\, p^6\, q^{394} \approx 0.890$

Si on utilise la loi de Poisson Z de paramètre $\lambda = 4$, $P(Z \le 6) = e^{-4}\left(1 + 4 + \frac{4^2}{2!} + \frac{4^3}{3!} + \frac{4^4}{4!} + \frac{4^5}{5!} + \frac{4^6}{6!}\right)$

$= e^{-4}\left(1 + 4 + 8 + \frac{32}{3} + \frac{32}{3} + \frac{128}{15} + \frac{256}{45}\right) = \frac{2185}{45}\ e^{-4} = \frac{437}{9}\ e^{-4} \approx 0.889$

La deuxième probabilité s'écrit P (Q > 100) = q^{100} = 0.366

* Il faut préciser de quoi il est question ici. Une erreur de typo étant possible à chaque caractère, les éléments de base sont donc les caractères et non les pages. S'il y a N caractères par page, et si la probabilité d'une erreur par caractère est p_0, la probabilité d'existence d'erreurs sur une page est $1-(1-p_0)^N \approx N p_0 \approx p$

La probabilité de deux erreurs est $C_N^2 p_0^2(1-p_0)^{N-2} \approx \frac{p^2}{2}$

La probabilité de trois erreurs est $C_N^3 p_0^3(1-p_0)^{N-3} \approx \frac{p^3}{6}$

2. Pour un modèle d'élections à deux candidats A et B, dans une population homogène, on peut considérer qu'en l'absence totale d'information le vote d'un électeur suit une Loi de Bernoulli, avec une probabilité p de voter pour le candidat A (X = 1) , et la probabilité $q = 1 - p$ de voter pour le candidat B (X = 0). Pour N suffrages exprimés, supposés indépendants, le nombre Y de votes A suit donc une loi binomiale B (N, p).

On peut observer l'erreur' commise par un sondage sur N personnes à l'aide de l'écart type du pourcentage $\frac{Y}{N}$ de votes A. $V\left(\frac{Y}{N}\right) = \frac{V(Y)}{N^2} = \frac{Npq}{N^2} = \frac{Npq}{N^2} = \frac{pq}{N} \approx \frac{1}{4N}$ si p est voisin de 0.5. L'écart type est donc $\sigma = \frac{1}{2\sqrt{N}}$

Pour N = 1000 personnes, cela donne $1.58\ 10^{-2}$

Cela peut paraitre élevé, mais on peut assurer en outre que c'est une valeur encore optimiste. On sait qu'en réalité la population d'a rien d'homogène, et que si l'échantillon retenu de représente par finement cette inhomogénéité, il faudra ajouter à l'erreur aléatoire précédente un biais lié à la représentativité de l'échantillon (les hommes ne votent pas comme les femmes, ni les jeunes comme les vieux, les travailleurs précaires comme les PDG, ...)

d) Fonction génératrice

Les fonctions génératrices permettent d'étudier les lois de probabilités discrètes à partir d'une écriture de série formelle. Leur étude est liée à celle des séries entières. Si une variable aléatoire X prend ses valeurs dans $\mathbb{N}$, et si P ($X = k$) = p_k, on définit la série G_X par $G_X(x) = \sum_{k=0}^{+\infty} p_k x^k$, qui converge pour $|x| \leq 1$. G_X est la **fonction génératrice** de la variable aléatoire X. Elle caractérise la loi de probabilité de X : si X et Y sont deux variables aléatoires à valeurs dans $\mathbb{N}$, telles que $G_X = G_Y$, X et Y ont la même loi de probabilité.

Remarque : Le fait que X prenne des valeurs entières est limitatif a priori, mais la plupart des lois discrètes usuelles (Bernoulli, binomiale, géométrique, Poisson, équidistribuées) ont cette propriété.

La F. G. permet de calculer plus directement la moyenne ou les moments de la loi, ou de déterminer la loi elle-même.

On aura en particulier si G_X est dérivable, comme $G_X'(x) = \sum_{k=1}^{+\infty} k\, p_k\, x^{k-1}$, E (X) = $G_X'(1)$

et comme $G_X''(x) = \sum_{k=2}^{+\infty} k(k-1)\, p_k\, x^{k-2}$, $G_X''(1) = \sum_{k=2}^{+\infty} k(k-1)\, p_k) = \sum_{k=1}^{+\infty} k(k-1)\, p_k$ = E (X^2) – E (X)

Donc E (X^2) = E (X) + $G_X''(1) = G_X'(1) + G_X''(1)$, d'où V ($X$)

On peut également étudier la somme de variables aléatoires à valeurs entières indépendantes, à partir du

Théorème 1.8 Si X et Y sont deux variables aléatoires indépendantes, $G_{X+Y} = G_X\, G_Y$

Démonstration : On note pour tout entier k, P ($X = k$) = p_k et P ($Y = k$) = q_k

On peut écrire $P(X+Y=k) = \sum_{j=0}^{k} P(X=j \text{ et } Y=k-j) = \sum_{j=0}^{k} P(X=j)\,.\,P(Y=k-j) = \sum_{j=0}^{k} p_j\, q_{k-j}$

et donc $G_{X+Y}(x) = \sum_{k=0}^{+\infty} P(X+Y=k)\, x^k = \sum_{k=0}^{+\infty} \sum_{j=0}^{k} p_j\, q_{k-j}\, x^k = \sum_{j=0}^{+\infty} p_j\, x^j \sum_{k=j}^{+\infty} q_{k-j}\, x^{k-j}$

$= \sum_{j=0}^{+\infty} p_j\, x^j \sum_{l=0}^{+\infty} q_l\, x^l = \sum_{j=0}^{+\infty} p_j\, x^j\, G_Y(x) = G_X(x)\, G_Y(x)$ □

Exemples et applications :

1. Loi binomiale Y : B (n, p) $G_Y(x) = \sum_{k=0}^{n} C_n^k\ p^k q^{n-k} x^k = \sum_{k=0}^{n} C_n^k\ (p\,x)^k q^{n-k} = (q + p\,x)^n$

On peut retrouver directement les moments de Y :

$G_Y'(x) = np\,(q + p\,x)^{n-1}$, et E $(Y) = G_Y'(1) = np\,(q + p)^{n-1} = 1$

Si on somme deux variables aléatoires binomiales indépendantes de même paramètre p, soit X : B(n, p) et Y :B (m, p) $G_{X+Y}(x) = (q + p\,x)^{n+m}$ et $X + Y$ suit une loi binomiale B$(n + m, p)$

2. Loi géométrique Q : G (p). $G_Q(x) = \sum_{k=1}^{+\infty} p\,q^{k-1} x^k = p\,x \sum_{k=1}^{+\infty} (q\,x)^{k-1} = p\,x \sum_{l=0}^{+\infty} (q\,x)^l = p\,x\,\frac{1}{1-qx}$

3. Loi de Poisson Z : $\mathcal{P}(\lambda)$. $G_Z(x) = \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k\,!} x^k = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda x)^k}{k\,!} = e^{-\lambda} e^{\lambda x} = e^{\lambda(x-1)}$
$G_Z'(x) = \lambda\ e^{\lambda(x-1)}$, soit $G_Z'(1) = \lambda$ = E (Z), et
$G_Z''(x) = \lambda^2\ e^{\lambda(x-1)}$, d'où V $(Z) = G_Z'(1) + G_Z''(1) - G_Z'(1)^2 = \lambda$

Ces calculs sont certainement plus simples que ceux faits au paragraphe 5 c), mais on peut également montrer de manière immédiate que :

Théorème1.8 La somme de deux variables aléatoires de Poisson indépendantes, de paramètres respectifs λ et μ, est une variable de Poisson de paramètre $\lambda + \mu$.

Démonstration : La démonstration est, avec les fonctions génératrices, plus courte que l'énoncé. Si on note Z : $\mathcal{P}(\lambda)$ et U : $\mathcal{P}(\mu)$, $G_{Z+U}(x) = e^{\lambda(x-1)}\ e^{\mu(x-1)} = e^{(\lambda+\mu)(x-1)}$, ce qui caractérise une loi de Poisson $\mathcal{P}(\lambda+\mu)$ □

On dit que la classe des lois de Poisson, ou que La loi de Poisson est **stable**.

e) Variation de constantes (suites); paradoxe des anniversaires

Un certain nombre de calculs de probabilités dépendent d'un paramètre entier, avec des évènements qu'on peut noter A_n, et il arrive fréquemment qu'on soit amenés à faire varier ce paramètre n, et considérer une suite p_n = P (A_n).

En particulier si on met en lumière une relation de récurrence sur les p_n, et qu'on peut calculer les solutions de cette suite récurrente on pourra répondre à la question initiale. On peut citer ici deux exemples. On en verra d'autres, notamment au paragraphe 9.

paradoxe des anniversaires

Cet exercice est classique comme situation où l'intuition peut jouer certains vilains tours, mais aussi en raison de ses implications, notamment en cryptographie.

On considère n personnes d'un groupe, et on demande si au moins deux d'entre elles sont nées le même jour de l'année. On va essayer pour répondre à cette question *a priori* de calculer la probabilité de cet événement.

Si on note $J_1, J_2, ..., J_n$ les jours anniversaires des n individus, ces valeurs peuvent être représentées par des entiers, de 1 à N (N = 365 ?). On peut considérer qu'il s'agit de variables aléatoires, a priori indépendantes. Leur loi de probabilité est inconnue - à moins d'avoir des statistiques précises sur la répartition des naissances. En l'absence d'une telle information, la seule hypothèse raisonnable est celle d'une équirépartition.

Dans ce cas, P (A_n) = P $(\exists\, k \neq l, 1 \leq k \leq n, 1 \leq l \leq n, J_k = J_l) = 1 -$ P $(\forall k \neq l, 1 \leq k \leq n, 1 \leq l \leq n, J_k \neq J_l) = 1 -$ P (B_n)

on note P $(B_n) = v_n$. S'il y a $n + 1$ individus, on peut écrire $B_{n+1} = B_n \cap C_n$, $C_n = (\forall k, 1 \leq k \leq n, , J_k \neq J_{n+1})$
Par ailleurs on pourra toujours écrire v_{n+1} = P (B_{n+1}) = P $(B_n \cap C_n)$ = P (C_n / B_n) . P (B_n),

et P $(C_n / B_n) = \frac{N-n}{N}$. On a donc une relation de récurrence sur v_n : $v_{n+1} = \frac{N-n}{N}.\ v_n$
La valeur initiale est $v_1 = 1$.

On pourra alors écrire en itérant la récursion : $v_n = \frac{N-(n-1)}{N} . v_{n-1}$

$$v_n = \frac{N-(n-1)}{N} . \frac{N-(n-2)}{N} v_{n-2} = \frac{N-(n-1)}{N} . \frac{N-(n-2)}{N} \dots . \frac{N-1}{N} v_1 = \frac{N-(n-1)}{N} . \frac{N-(n-2)}{N} \dots . \frac{N-1}{N} . \frac{N}{N}$$

Soit encore : $v_n = \frac{A_N^n}{N^n}$, et $P(A_n) = 1 - P(B_n) = 1 - v_n = 1 - \frac{A_N^n}{N^n}$.

Suites de 1 consécutifs (ou détections consécutives dans un système)

On considère une suite de n variables aléatoires de Bernoulli B (p) indépendantes, $X_1, X_2, ..., X_n$, et on demande la probabilité qu'il existe au cours de la suite au moins une suite d'au moins r fois 1 consécutifs. Cet événement peut être noté A, ou encore A_n, et on pose $u_n = P(A_n)$.

La suite est une suite croissante d'ensembles. Si on pose $B_n = A_n - A_{n-1}$, on aura donc $P(B_n) = P(A_n) - P(A_{n-1})$

L'évènement B_n peut être interprété en fonction des valeurs de la suite :

Si une suite de k 1 consécutifs existe à l'ordre n mais pas à l'ordre $n-1$, c'est que les k dernières valeurs sont des 1, précédées par un 0, et qu'il n'existe pas d'autre suite de k 1 consécutifs auparavant, soit :
$B_n = (X_n = 1) \cap (X_{n-1} = 1) \cap ... \cap (X_{n-k+1} = 1) \cap (X_{n-k} = 0) \cap A_{n-k-1}{}^c$
Ce qui donne : $P(B_n) = u_n - u_{n-1} = p^k q (1 - u_{n-k-1})$

ou encore $\boldsymbol{u_n = u_{n-1} + p^k q (1 - u_{n-k-1}), n > k+1}$,

avec les valeurs initiales : $u_1 = u_2 = ... = u_{k-1} = 0$, $u_k = p^k$, $u_{k+1} = p^{k+1} + 2q p^k$,

Remarque : on pourrait poser $u_0 = 0$, et dance ce cas on retrouve la valeur de u_{k+1} par la formule de récurrence, pour $n > k$
Exemples : si $p = 0.5$ et $n = 100$, alors pour $k = 5, 6, 7$ on obtient respectivement $u_{100} = 0.81, 0.55$, et 0.32

Si $p = 0.4$ et $n = 100$, alors pour $k = 5, 6, 7$ on obtient respectivement $u_{100} = 0.46, 0.21$, et 0.09

Si $p = 0.3$, et $n = 38$, alors pour $k = 2, 3, 4$ on aura respectivement $u_{38} = 0.94, 0.52$, et 0.19

1.8. Chaines de Markov

a) Modèle et Formule de calculs

On considère une suite de variables aléatoires discrètes (X_k) dont les valeurs sont dans un même ensemble fini $E = \{e_1, e_2, ..., e_N\}$. E est appelé ensemble des états. On peut représenter les états par leurs indices, et poser $e_i = i$. On suppose que les probabilités $p_{ij}^{(n)} = P(X_n = i / X_{n-1} = j)$ existent pout tout couple d'indice (i, j) de 1, 2, ..., N.

La suite (X_k) est appelée **chaine de Markov**.

Si les **probabilités de transition** $p_{ij}^{(n)} = p_{ij}$ ne dépendent pas de n, la chaine de Markov est dite **stationnaire**.

L'étude d'une chaine de Markov consiste en la recherche des lois des variables aléatoires, caractérisées par les probabilités $P(X_n = i)_{1 \leq i \leq N}$.

On peut écrire $p^{(n)}{}_i = P(X_n = i)$ avec la formule de Bayes : $P(X_n = i) = \sum_{j=1}^{N} P(X_n = i / X_{n-1} = j) . P(X_{n-1} = j)$

Soit $p^{(n)}{}_i = \sum_{j=1}^{N} p_{ij}^{(n)} . P(X_{n-1} = j) = \sum_{j=1}^{N} p_{ij}^{(n)} . p^{(n-1)}{}_j$

Si on définit le vecteur colonne $p^{(n)}$ de composantes $p^{(n)}{}_i$, et la matrice (de transition) $P^{(n)}$ de composantes $p_{ij}^{(n)}$

on obtient la relation matricielle $p^{(n)} = P^{(n)} p^{(n-1)}$

On en déduit l'expression du vecteur $p^{(n)} = P^{(n)} P^{(n-1)} p^{(n-2)} = P^{(n)} P^{(n-1)} P^{(n-2)} ... P^1 p^{(0)}$

Pour une chaine stationnaire, de matrice de transition d'états P, $p^{(n)} = P^n p^{(0)}$

b) Calculs pratiques

Il est fréquent de connaitre l'état initial, et dans ce cas on connait aussi le vecteur $p^{(0)}$.

Si la chaine de Markov est non stationnaire la difficulté principale est de connaitre, et d'écrire les différences matrices $P^{(k)}$. Le calcul de $p^{(n)}$ se fait avec n produits matrice - vecteur, soit pour des matrices pleines $n \cdot N^2$ multiplications.
Pour une chaine de Markov stationnaire, il y a 3 calculs potentiellement réalisables

1. Produits matrice - vecteur. C'est comme pour le cas non stationnaire.

2. Si n est grand par rapport au nombre N d'états, et si on veut calculer une valeur particulière de $p^{(n)}$, on peut calculer les puissances de P avant de calculer le produit avec $p^{(0)}$. On aura de l'ordre de $\log_2(n) \cdot N^3$ multiplications, à un facteur 2 près.

3. Les deux méthodes précédentes consistent en des calculs numériques. Dans certains cas on peut calculer directement l'expression de P^n en fonction de n, et on aura donc une écriture algébrique de $p^{(n)}$, comme fonction de n.

C'est possible en particulier si la matrice P est diagonalisable. Si D est la matrice diagonale formée des valeurs propres, et si Q est la matrice de passage à une base diagonale, on aura $P = Q\,D\,Q^{-1}$, et donc $P^n = Q\,D^n\,Q^{-1}$

Mais en pratique cela sera faisable si on peut calculer explicitement les valeurs propres, ce qui n'est pas possible le plus souvent, sauf pour $N = 2$ ou $N = 3$ (on sait que pour $N \geq 5$ la recherche explicite des zéros du polynôme caractéristique est impossible en général).

c) Applications : durée de vie d'équipements, redondance

On va citer trois petits problèmes qui peuvent être abordés à l'aide de chaines de Markov.

1. La charge utile d'un satellite de communication comporte deux équipements identiques. Les deux sont utilisés, mais en cas de panne un seul peut être utilisé en mode secours. Pour des raisons claires il n'y aura pas de réparation.

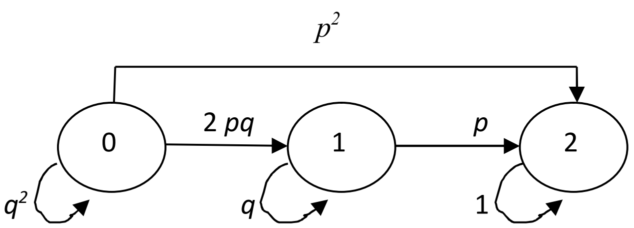
On va supposer que les pannes sont totalement indépendantes, et ne dépendent pas de l'âge des équipements, avec une probabilité $p = 1 - q$ de panne sur une période de référence (de 24 heures). Ce problème peut être représenté par une chaine de Markov, avec 3 états possibles (0, 1, ou 2 équipements en panne).

Si il y a 0 équipement en panne, Le jour suivant il y aura toujours 0 équipement en panne si aucun des deux ne tombe en panne, soit avec la probabilité q^2. Un unique équipement sera en panne avec probabilité $2\,pq$, et les 2 seront en panne avec probabilité p^2 (il s'agit ici d'une loi binomiale B $(2, p)$)

S'il y a un équipement en panne, celui restant en état de marche peut tomber en panne avec probabilité p, ou rester en fonction avec probabilité q.

Si les deux équipements sont en panne, à moins d'envoyer un technicien dans l'espace, cela va perdurer.

On peut faire un schéma représentant les états et les transitions :



La matrice de transition P peut s'écrire conformément à ce schéma : $P = \begin{bmatrix} q^2 & 0 & 0 \\ 2pq & & 0 \\ p^2 & p & 1 \end{bmatrix}$.

Par ailleurs on peut supposer que les équipements sont en fonction lors du lancement du satellite, soit donc

$$p^{(0)} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

La matrice P étant triangulaire, ses valeurs propres sont ses éléments diagonaux, q^2 , q et 1, qui sont distincts, et P est diagonalisable. Il reste à déterminer la matrice de passage, à calculer son inverse, et à effectuer 3 produits matrice-vecteur, ce qui est de l'algèbre linéaire basique.

On peut noter que pour toutes les chaines sans retour ou causales, la forme triangulaire de la matrice de transition d'états assurera la diagonalisation de la matrice, et la connaissances de ses valeurs propres.

2. Si un flot de communications peut utiliser deux bus parallèles, avec à chaque étape une probabilité p ($0 < p < 1$) de changer de bus, on a un système symétrique pouvant être représenté par un petit automate, et modélisé par une chaine de Markov. La matrice de transition est : $P = \begin{bmatrix} q & p \\ p & q \end{bmatrix}$.

Son polynôme caractéristique s'écrit : $\Psi_P(X) = \begin{vmatrix} q-X & p \\ p & q-X \end{vmatrix} = (q-X)^2 - p^2 = (q-X-p)\ (q-X+p)$
$= (q-p-X)\,(1-X)$

et les valeurs propres sont $q-p$ et 1.

On peut déterminer des vecteurs propres associés à ces valeurs propres :

$* \; P\begin{pmatrix} x \\ y \end{pmatrix} = (q-p)\begin{pmatrix} x \\ y \end{pmatrix} \Leftrightarrow \begin{Bmatrix} qx + py = (q-p)\,x \\ px + qy = (q-p)\,y \end{Bmatrix}$

Ces deux équations sont équivalentes (l'espace propre est une droite, d'équation $y = -x$, et de vecteur directeur $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$)

$* \; P\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \Leftrightarrow \begin{Bmatrix} qx + py = x \\ px + qy = y \end{Bmatrix} \Leftrightarrow p\,y = (1-q)\,x = p\,x \Leftrightarrow y = x$ (vecteur directeur $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$)

On peut donc prendre $Q = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$, avec $Q^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

Si le bus utilisé au départ est le premier, $p^{(n)} = Q\ D^n\ Q^{-1}\,p^{(0)} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} . \begin{bmatrix} (q-p)^n & 0 \\ 0 & 1 \end{bmatrix} . \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} . \begin{bmatrix} (q-p)^n & 0 \\ 0 & 1 \end{bmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} . \begin{pmatrix} (q-p)^n \\ 1 \end{pmatrix} = \frac{1}{2} . \begin{pmatrix} 1 + (q-p)^n \\ 1 - (q-p)^n \end{pmatrix}$

et comme $-1 < q - p = 1 - 2\,p < 1$, $(q-p)^n$ tend vers 0.

On obtient à la limite (de fait après quelques itérations) une répartition symétrique, avec une chance sur 2 d'utiliser chacun des deux bus.

3. On reprend le problème des suites de 1 consécutifs, qu'on peut représenter par un modèle markovien : Si on note j ($0 \le j \le k-1$) le nombre de 1 consécutifs précédant l'étape n (incluse), auxquels on ajoure un état j absorbant, on a

donc $k + 1$ états, et la matrice de transition s'écrit : $P = \begin{bmatrix} q & q & \cdots & q & 0 \\ p & 0 & \cdots & 0 & 0 \\ 0 & p & & 0 & 0 \\ 0 & 0 & \cdots & p & 1 \end{bmatrix}$

Cette matrice est structurée, et on peut la diagonaliser, mais ce n'est pas immédiat. Dans ce cas la solution proposée au paragraphe 1.7 e parait sans doute plus directe.

1.9. Marches aléatoires

a) Définition et marche bornée

Une **marche aléatoire** est une suite de variables aléatoires (S_n), où $S_n = S_0 + X_1 + X_2 + ... + X_n$ est la somme de n variables de Bernoulli indépendantes, telles que P $(X_k = \alpha) = p$ et P $(X_k = \beta) = q = 1 - p$, où

S_0 peut être une valeur constante, ou plus généralement une variable aléatoire

α et β sont des nombres réels. Dans la plupart des situations, on aura $\beta = 1$, et $\alpha = 0$ (modèle B (p)) ou $\alpha = -1$

La marche aléatoire (S_n) est définie par la relation de récurrence : $S_n = S_{n-1} + X_n$

Une marche aléatoire **bornée** dans $I = [a, b]$ est définie par la relation de récurrence :

si $S_{n-1} = a$ ou $S_{n-1} = b$, $S_n = S_{n-1}$, sinon $S_n = S_{n-1} + X_n$

<u>Remarque</u> : Si Y est une variable de Bernoulli avec $\alpha = -1$ et $\beta = 1$, $X = \frac{Y+1}{2}$ suit la loi B (p), et on aura E $(Y) = p - q$

Les marches aléatoires permettent de modéliser ///

La marche aléatoire peut être représentée comme une ligne brisée, où n est en abscisse et S_n en ordonnée, et la courbe ainsi générée, pour des réalisations des termes X_k , en est une trajectoire. Si on observe ces trajectoires avec des points d' abscisse plus rapprochés elles se rapprochent de courbes régulières d'un type particulier.

b) Jeu de roulette et ruine des joueurs; applications

On considère une marche aléatoire bornée, avec $\alpha = -1$ et $\beta = 1$, et $a = 0$.

On note $p = $ P $(X_k = 1)$ et $q = $ P $(X_k = -1)$

On suppose que $S_0 = k$. On souhaite répondre à la question suivante : quelles sont les probabilités que la marche converge vers 0, ou qu'elle atteigne la valeur stationnaire b.

Ce problème peut modéliser un jeu de casino, dans lequel le joueur mise la même mise 1 à chaque jeu, et possède une fortune k, avec l'objectif d'atteindre une somme b avant d'arrêter de jouer. La convergence vers 0 peut représenter la 'ruine' du joueur.

Ici encore nous allons établir une relation de récurrence, en prenant k comme indice. Et comme nous l'avons déjà fait, c'est la formule de Bayes que nous mettons en oeuvre. On note q_k la probabilité de ruine en partant de la valeur initiale k, que l'on peut encore noter P $(R / S_0 = k) = $ P $(R / S_0 = k$ et $X_1 = -1) \cdot q$ + P $(R / S_0 = k$ et $X_1 = 1) \cdot p$

Soit $q_k = q\, q_{k-1} + p\, q_{k+1}$, $0 < k <$ b, et les conditions extrêmes : $q_0 = 1$ et $q_b = 0$.

Il s'agit d'une suite récurrente linéaire d'ordre 2. L'équation caractéristique s'écrit $p X^2 - X + q = 0$ de discriminant $1 - 4pq = 1 - 4p(1-p) = 4p^2 - 4p + 1 = (2p-1)^2$.

* Si $p = \frac{1}{2}$, il y a une racine double, $x_0 = \frac{1}{2p} = 1$, et $q_k = \alpha_1 + k\, \alpha_2$, d'où $\alpha_1 = 1$ et $\alpha_2 = \frac{-1}{b}$

On obtient donc $q_k = 1 - \frac{k}{b}$

* Si $p \neq \frac{1}{2}$, les solutions de l'équation caractéristique sont $x_1 = 1$ et $x_2 = q / p$, et $q_k = \alpha_1 + \alpha_2\, x_2^k$.

α_1 et α_2 vérifient le système : $\alpha_1 + \alpha_2 = 1$

$\alpha_1 + \alpha_2\, x_2^b = 0$

Et on obtient $\quad q_k = \dfrac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^b}{1 - \left(\frac{q}{p}\right)^b}$

On peut vérifier que cette fonction est décroissantes en fonction de k (ou de k/b).

Si on modélise le jeu de type roulette au casino, où le joueur qui dispose d'une fortune F_0, mise la même mise m autant de fois que nécessaire pour obtenir une fortune F_1, ou d'être ruiné : on peut choisir comme unité monétaire la valeur m, et dans ce cas on prendra $k = F_0/m$ et $b = F_1/m$.

Le résultat précédent signifie que l'on a intérêt à choisir pour k la plus grande valeur possible, soit k = 1, ou m = F0 : on joue toute sa fortune

<u>exemple</u> : à la roulette française $p = 18/37$, et $q/p = 19/18$

Si on part d'une fortune de 100 euros, en espérant 400 euros

* avec une mise de 1 euro, la probabilité de ruine est $p_R = 1$ (ici, $k = 100$ et $b = 400$)
* avec une mise de 4 euros, la probabilité de ruine est $p_R = 0.987$ ($k = 25$ et $b = 100$)
* avec une mise de 20 euros, la probabilité de ruine est $p_R = 0.84$ ($k = 5$ et $b = 20$)
* avec une mise de 100 euros, la probabilité de ruine est $p_R = 0.77$ ($k = 1$ et $b = 4$)

Pour la roulette américaine, $p = 18/38 = 9/19$, et $q/p = 10/9$

Si on part d'une fortune de 100 dollars, en espérant 400 dollars (respectivement 600 dollars)

* avec une mise de 1 dollar, la probabilité de ruine est $p_R = 1$ (respectivement $p_R = 1$)
* avec une mise de 4 dollars, la probabilité de ruine est $p_R = 0.9997$ (respectivement $p_R = 1$)
* avec une mise de 20 dollars, la probabilité de ruine est $p_R = 0.89$ (respectivement $p_R = 0.9693$)
* avec une mise de 100 dollars, la probabilité de ruine est $p_R = 0.788$ (respectivement $p_R = 0.874$)