Mathématiques appliquées aux probabilités

Cours de Philippe CHASSAING
Retranscrit par Hélène ACHIMOFF

Chapitre Introductif

I/ Probabilités, univers, évènements

Définition 1 : L'univers, souvent noté Ω, est l'ensemble des éventualités

Définition 2 : On appelle éventualité tout élément de l'univers Ω.

(*définitions qui tournent en rond, en effet…*)

Exemples :

- Un lancer de dés
$\Omega = \{1,2,3,4,5,6\}$
Nombre d'éléments de l'univers Ω : il se note $\#\Omega$ et vaut ici $\#\Omega = 6$

- Deux lancers de dés successifs :
$\Omega = \{1,2,3,4,5,6\} \text{ x } \{1,2,3,4,5,6\} = \{1,2,3,4,5,6\}^2 = \{l_1,l_2 \ / \ 1 \leq l_i \leq 6\}$
Ω est le couple formé par le résultat du premier lancer suivi du résultat du second.

- 20 lancers de dés successifs :
$\Omega = \{1,2,3,4,5,6\}^{20} = \{l_1,l_2, \dots ,l_{20}\} \ / \ 1 \leq l_i \leq 6 \ \forall \ i \in \{1,2,\dots,20\} \}$

- rangement d'une liste en désordre, de longueur N :
Ω = ensemble des permutations de $\{1,2,3,4,5,6]$ et $\#\Omega = N!$

- suites des identifiants de 10^M paquets de données :
Un identifiant : N bits → identifiant $\in \{0,1\}^N$
Une suite $\in \Omega = (\{0,1\}^N)^{(10\text{^}M)}$
$\# \ \Omega = (2^N)^{(10\text{^}M)} = 2^{N\text{x}10\text{^}M}$
→ Problème des occurrences distinctes

- Tirer un nombre entre 0 et 1, au hasard, à l'aide de l'instruction RAND() ou ALEA() …
$\Omega = [0,1]$

- Tirer n nombres au hasard à l'aide d'appels successifs à l'instruction RAND() (ou autres…)
$\Omega = [0,1]^n$
$n=3 \Rightarrow \Omega$ est le cube unité

Dans les premiers exemples Ω est fini, dans les deux derniers exemples Ω est infini, et même non dénombrable, ce qui induira des méthodes différentes pour calculer les probabilités, en utilisant des intégrales sur A plutôt que des sommes sur A.

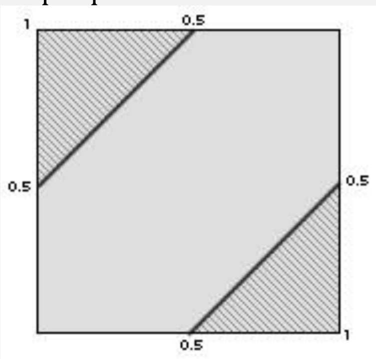
Définition 3 : Un évènement A est une partie de l'univers Ω ($A \subset \Omega$, $A \in \mathcal{P}(\Omega)$)
(on note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω)

Exemples

- Soit A = "la somme des deux premiers lancers est paire"
C'est à la fois une affirmation (vraie ou fausse) (du moins a posteriori) et une partie de $\Omega=\{1,2,3,4,5,6\}^2$
On a A = {(1,1),(1,3),(1,5),(2,2),(2,4),(3,1),(3,3),(3,5), … }
On voit que #A = 18 alors que # $\Omega = 6^2 = 36$, on a envie d'accorder à A la valeur $\frac{18}{36} = \frac{1}{2}$ plutôt que la valeur 0 ou 1, a priori.

- 2 personnes a et b arrivent à deux instants X et Y tirés au hasard entre 0 et 1, en heures. Le premier arrivé attend l'autre.
B = "l'attente est supérieure à ½ heure" est un évènement de $\Omega = [0,1]^2$
On a B="{X ≥ Y + ½ } ou {Y ≥ X + ½ }" = {(x,y) ∈ Ω / x ≥ y + ½ ou y ≥ x + ½ }

Graphiquement :



On a envie de dire que $\mathbb{P}(B) = \frac{\text{aire de B}}{\text{aire de }\Omega} = \frac{1/4}{1} = 1/4$

$\mathbb{P}(B_t) = \mathbb{P}(\text{attente} \geq t) = (1\text{-}t)^2$ = aire d'une zone hachurée similaire à la précédente, mais de côtés (1-t)
$\mathbb{P}(B_t) = ½ = \mathbb{P}(B_t^c)$ si $(1\text{-}t)^2 = ½$ ou si $t = 1 - \frac{1}{\sqrt{2}} = 0.292$ h = 17' 34''
Donc l'attente a une valeur typiquement aux alentours de sa médiane.
(Ndld : Non, je n'ai toujours pas compris même après réflexion intense…)

<u>Définition 4</u> : Une <u>probabilité</u> $\mathbb{P}$ (ou plutôt une <u>mesure de probabilité</u> $\mathbb{P}$) est une application

$$\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$$

Possédant les deux propriétés suivantes :

- $\mathbb{P}(\Omega) = 1$ et $\mathbb{P}(\emptyset)=0$
- Si $A_1, A_2, \dots, A_n, \dots$ est une suite d'évènements <u>disjoints</u> ($A_i \subset \Omega$, $A_i \cap A_j = \emptyset$ si $i \neq j$), alors $\mathbb{P}(\bigcup_{i\geq 1} Ai) = \sum_{i\geq 1} \mathbb{P}(Ai)$

<u>Propriétés</u> : En conséquence des deux précédentes propriétés, on a :

1 – $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

2 – $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

3 – $\mathbb{P}(A^c) = 1\text{-}\mathbb{P}(A)$

4 – $\{A_i \cap A_j = \emptyset \ (\forall\, i \neq j,\ 1 \leq i, j \leq n)\} \Rightarrow \mathbb{P}(\bigcup_{i=1}^{n} Ai) = \sum_{i=1}^{n} \mathbb{P}(Ai)$

5 – $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$ si $B \subset A$

6 – $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

7 – $\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3)$

8 – Généralisation à $A_1, \ldots, A_4$ $\quad (4 - 6 + 4 - 1 = 1)$

$A_1, \ldots, A_n$ $\quad (\sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} = 1)$

→ Formule de Poincarré

9 – Si $\forall\, n \geq 1$ on a $A_n \subset A_{n+1}$ alors $\mathbb{P}(\bigcup_{n\geq 1} A_n) = \lim_n \mathbb{P}(A_n)$

10 – Si $\forall\, n \geq 1$ on a $B_n \supset B_{n+1}$ alors $\mathbb{P}(\bigcap_{n\geq 1} B_n) = \lim_n \mathbb{P}(B_n)$

Exemple d'algorithme randomisé plus rapide que l'algorithme standard

Trions une liste $y = (y_1, \ldots, y_n) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$

Où σ est une permutation telle que $x_1 < x_2 < \ldots < x_n$ et $\{x_1, \ldots, x_n\} = \{y_1, \ldots, y_n\}$

A l'aide de comparaisons deux à deux où un oracle retourne une réponse binaire "x>y" ou "y>x"

Si $y = \{x_4, x_6, x_7, x_2, x_5, x_3, x_1\}$

On compare tout à x_4 ($x_4 > x_1$, etc...)

Cela sépare la liste en deux parties $\{x_2, x_3, x_1\}$ et $\{x_6, x_7, x_5\}$ — coût : 6 comparaisons

| _ | _ | _ | x_4 | _ | _ | _ |

$\{x_2, x_3, x_1\}$

On compare tout à x_2

Cela sépare la liste en deux parties $\{x_1\}$ et $\{x_3\}$ — coût : 2 comparaisons

| _ | x_2 | _ | x_4 | _ | _ | _ |

$\{x_6, x_7, x_5\}$

On compare tout à x_6

Cela sépare la liste en deux parties $\{x_5\}$ et $\{x_7\}$ — coût : 2 comparaisons

| _ | x_2 | _ | x_4 | _ | x_6 | _ |

Total : 10 comparaisons

| x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 |

Si on avait eu $y = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$

Cela nous aurait fait 6 comparaisons, puis 5, puis 4, etc... soit au total 21 comparaisons.

Exemples de mesures de probabilités :

voir feuilles 6bis et 6ter des notes de cours

II / Probabilités conditionnelles

Définition : Sur un espace probabilisé $(\Omega , \mathbb{P})$ pour A et B $\in \mathcal{P}(\Omega)$, tels que $\mathbb{P}(B)>0$, on définit la probabilité conditionnelle de A sachant B, notée $\mathbb{P}(A/B)$, de la manière suivante :

$$\mathbb{P}(A/B) = \frac{\mathbb{P}(\boldsymbol{A} \cap \boldsymbol{B})}{\mathbb{P}(\boldsymbol{B})}$$

Commentaire : On revoit le pronostic $\mathbb{P}(A)$ en fonction d'une information nouvelle, à savoir que l'évènement B s'est produit : le résultat est le nouveau pronostic $\mathbb{P}(A/B)$

Propriétés :

1 – L'application $\mathbb{P}(./B)$: $\begin{array}{ccc} A & \to & \mathbb{P}(A/B) \\ \mathcal{P}(B \text{ ou } \Omega) & \to & [0,1] \end{array}$

est une probabilité sur l'univers B (sur l'univers Ω également) et en possède toutes les propriétés. Par exemple, si $(C_i)_{i\geq 1}$ ou $(C_i)_{1\leq i\leq n}$ est une famille d'évènements disjoints, alors $\mathbb{P}(\bigcup_i Ci/B) = \sum_i \mathbb{P}(Ci/B)$

2 – $\mathbb{P}(A\cap B) = \mathbb{P}(B).\mathbb{P}(A/B)$

3 – $\mathbb{P}(A_1\cap A_2\cap \dots\cap A_n) = \mathbb{P}(A_1).\mathbb{P}(A_2/A_1).\mathbb{P}(A_3/A_1\cap A_2). \dots .\mathbb{P}(A_n/A_1\cap \dots\cap A_{n-1})$
→ chain rule

Application :

Une urne contient n-1 boules blanche et une boule noire.
On retire les boules de l'urne une par une, sans remise, jusqu'à épuisement de toutes les boules.
$\mathbb{P}(N_k) = \mathbb{P}(B_1).\mathbb{P}(B_2/B_1). \dots .\mathbb{P}(B_{k-1}/ B_1\cap B_2\cap \dots \cap B_{k-2}).\mathbb{P}(N_k/B_1\cap \dots\cap B_{k-1})$
Où B_i = "la i-ième boule tirée de l'urne est blanche"
N_i = "la i-ième boule tirée de l'urne est noire"

On voit facilement que

$\mathbb{P}(B_1) = \frac{n-1}{n}$
…
$\mathbb{P}(B_l / B_1\cap \dots\cap B_{l-1}) = \frac{n-l}{n-l+1}$
…
$\mathbb{P}(N_k / B_1\cap \dots\cap B_{k-1}) = \frac{1}{n-k+1}$

Donc $\mathbb{P}(N_k) = \frac{n-1}{n}\frac{n-2}{n-1}\dots\frac{n-k+1}{n-k+2}\frac{1}{n-k+1} = \frac{1}{n}$

Qu'on peut trouver plus directement, évidemment…

4 – Formule des probabilités totales

Etant donnés $(B_i)_{i\geq 1}$ (ou $1\leq i\leq n$) une famille de parties disjointes telles que $A \subset \bigcup_i Bi$ (et que $\mathbb{P}(B_i) > 0$), alors $\mathbb{P}(A) = \sum_i \mathbb{P}(\boldsymbol{Bi})\mathbb{P}(\boldsymbol{A}/\boldsymbol{Bi})$

Application :

Une boite contient N boules dont N_1 sont blanches et N_2 sont noires. Calculer la probabilité que la n+1 ième boule tirée de la boite (sans remise) soit noire.

A = "La n+1 ième boule tirée est noire"
B_k = "Parmi les n première boules tirées, k sont noires et n-k sont blanches"

Alors $\bigcup_{k=0}^{n} Bk = \Omega \supset A$ et $k \neq l \Rightarrow B_k \cap B_l = \emptyset$

Donc $P_{n+1} = \mathbb{P}(A) = \sum_{k=0}^{n} \mathbb{P}(Bk)\mathbb{P}(A/Bk)$

Clairement, on a $\mathbb{P}(A / B_k) = \frac{N2-k}{N-n} = \frac{N2-k}{N1+N2-n}$

Moins clairement $\mathbb{P}(B_k) = \frac{\binom{N1}{n-k}\binom{N2}{k}}{\binom{N}{n}}$

Remarque :

$\sum_{k=0}^{n} \mathbb{P}(Bk) = 1 = \sum_{k=0}^{n} \frac{\binom{N1}{n-k}\binom{N2}{k}}{\binom{N}{n}}$

… à finir

5 – Formule de Bayes

Si $\mathbf{A} \subset \bigcup_i \boldsymbol{B_i}$ et si $\{\mathbf{i \neq j}\} \Rightarrow \{\mathbf{B_i \cap B_j = \emptyset}\}$ alors $\mathbb{P}(\mathbf{B_i/A}) = \frac{\mathbb{P}(\mathbf{B_i})\mathbb{P}(\mathbf{A/B_i})}{\sum_j \mathbb{P}(\mathbf{B_j})\mathbb{P}(\mathbf{A/B_j})}$

III / Indépendance

Définition 1 : On dit que deux évènements A et B sont indépendants (on note cela A ⫫ B) si et seulement si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ **(*)**

Remarque :
Si $\mathbb{P}(B) > 0$, **(*)** est équivalent à $\mathbb{P}(A) = \mathbb{P}(A/B)$
Si $\mathbb{P}(A) > 0$, **(*)** est équivalent à $\mathbb{P}(B) = \mathbb{P}(B/A)$
($\mathbb{P}(A) = \mathbb{P}(A/B)$ signifie "la nouvelle information B ne change pas mon pronostic sur A")

Définition 2 : On dit que les évènements $(A_i)_{1\leq i\leq n}$ sont indépendants, et on note A_1 ⫫ A_2 ⫫ … ⫫ A_n si l'une des conditions suivantes (équivalentes) est remplie :

a) $\forall (\varepsilon_1,\dots,\varepsilon_n) \in \{0,1\}^n \quad \mathbb{P}(A_1^{\varepsilon 1} \cap \dots \cap A_n^{\varepsilon n}) = \mathbb{P}(A_1^{\varepsilon 1}) \dots \mathbb{P}(A_n^{\varepsilon n})$ où $\begin{cases} \varepsilon i = 0 \Rightarrow Ai \\ \varepsilon i = 1 \Rightarrow \overline{Ai} \end{cases}$

b) $\forall$ k tel que $2 \leq k \leq n$, $\forall\, i_1,\dots,i_k$ tels que $1 \leq i_1 < i_2 < \dots < i_k \leq n$
$\mathbb{P}(A_{i1} \cap \dots \cap A_{ik}) = \mathbb{P}(A_{i1})\dots\mathbb{P}(A_{ik})$

Contre exemples
Voir notes manuscrites

Chapitre 2 : Variables Aléatoires

I / Généralités

Exemples

1) R_1 rang auquel se classera finalement le premier pivot dans le Quicksort classique (Le paquet inférieur possède R_1-1 éléments et le paquet supérieur possède n-R_1 éléments) $\mathbb{P}(R_1 = k) = 1/n$, $1 \leq k \leq n$, si l'élément pivot est tiré au hasard. On dit que R_1 suit la loi uniforme sur [[1,n]]. On a vu que $\mathbb{P}(\frac{n}{4} \leq R_1 \leq \frac{3n}{4}) = ½ + O(1/n)$

2) Supposons qu'on tire trois éléments au hasard dans la liste, sans remise. Posons $R=(R_1,R_2,R_3)$ le vecteur de leurs rangs respectifs. C'est une variable aléatoire à valeurs dans $[[1,n]]^3$.

   On dit que R est un <u>vecteur aléatoire</u>.

   On a $\mathbb{P}(R=(k_1,k_2,k_3)) = \frac{1}{n(n-1)(n-2)}$ si $(k_1,k_2,k_3) \in [[1,n]]^3$ et si les k_i sont tous $\neq$

   $= 0$ sinon

   C'est encore une loi uniforme, sur un sous ensemble de $[[1,n]]^3$. Posons M=médiane(R_1,R_2,R_3). Une question intéressante (pour l'étude de Quicksort avec médiane) est de calculer $\mathbb{P}(M=k)$, $1 \leq k \leq n$, puis, par exemple, $\mathbb{P}(\frac{n}{4} \leq M \leq \frac{3n}{4})$, de manière à vérifier que $\mathbb{P}(\frac{n}{4} \leq R_1 \leq \frac{3n}{4}) = \lambda n + O(1/n)$ avec $\lambda >$ ½.

   Ainsi on obtient une indication en faveur de l'accélération de Quicksort en cas d'utilisation de la médiane.

3) Soit le problème des anniversaires pour N étudiants et m (~365) jours dans l'année. (N boules et m boites). Soit V le nombre de boîtes vides, à la fin (ou le nombre de jours sans anniversaire). Connaître la loi de V permet de vérifier que, en un certain sens, N~-m.log(V/m), ce qui permet d'estime N de manière économe en temps de calcul et en espace mémoire.

Exemples 4 et 5 à écrire, disponibles sur les notes de cours

II / Définitions Générales

<u>Définition 1</u> : Etant donné un espace de probabilité $(\Omega,\mathbb{P})$ et un ensemble E, on appelle variable aléatoire (notée v.a.) à valeurs dans E n'importe quelle application

$$X : \Omega \rightarrow E$$
$$\omega \rightarrow X(\omega)$$

Si $E \subseteq \mathbb{Z}$ ou $\mathbb{N}$ on dit que X est une variable aléatoire à valeurs entières.

Si $E \subseteq \mathbb{R}$ on dit que X est une variable aléatoire réelle (v.a.r.)

Si $E \subseteq \mathbb{R}^d$ ou $E \subseteq \mathbb{Z}^d$ on dit que X est un vecteur aléatoire.

Définition 2 : La "loi de X", notée $\mathbb{P}_X$, est une probabilité sur $\mathcal{P}(E)$ définie de la manière suivante : pour $A \subset E$, $\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega \,/\, X(\omega) \in A\}) = \mathbb{P}(X^{-1}(A))$

$X^{-1}(A)$ est appelée "l'image réciproque de A par X"

Remarque : on note aussi souvent $\mathbb{P}_X(A)=\mathbb{P}(X \in A)$, "$X \in A$" est une affirmation (à valeur booléenne, vraie ou fausse) qu'on confond, comme souvent en probabilités, avec l'ensemble $X^{-1}(A)= \{\omega \in \Omega \,/\, X(\omega) \in A\} = \mathbb{P}(X^{-1}(A))$

Propriété

$\mathbb{P}_X$ est une probabilité sur l'univers E, c'est-à-dire que $\mathbb{P}_X$ a les propriétés :

- $\mathbb{P}_X(E)=1$, $\mathbb{P}_X(\emptyset)=0$
- $\forall\ (A_k)_{k\geq 1}$ tel que $A_k \subset E$ et ($k \neq l \Rightarrow A_k \cap A_l = \emptyset$) on a $\mathbb{P}_X(\bigcup_{k\geq 1} A_k) = \sum_{k\geq 1} \mathbb{P}_X(A_k)$

Il en écoule que $\mathbb{P}_X$ a toutes les propriétés d'une probabilité.

Exemples de lois
Voir polycopiés

Définition 3

Si $X = (\Omega , \mathbb{P}) \rightarrow \mathbb{R}$ est une variable aléatoire réelle, alors on appelle **espérance de X** et on note $\mathbb{E}(X)$ ou bien $\int_\Omega X.d\mathbb{P}$ la valeur suivante :

$$\mathbb{E} = \int_\Omega X.d\mathbb{P} = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\})$$ si Ω est fini ou dénombrable.

Si $\Omega = \mathbb{R}$ ou $\mathbb{R}^d$ et $\mathbb{P}$ défini par une fonction f (la densité de $\mathbb{P}$) [telle que $f\geq 0$, $\int_\Omega f(\omega).d\omega = 1$] de la manière suivante :

$$\mathbb{P}(A) = \int_\Omega \mathbb{1}_A(\omega).f(\omega).d\omega$$
$$= \int \dots \int_{\Omega=\mathbb{R}^\wedge d} \mathbb{1}_A(\omega_1,\dots, \omega_d).f(\omega_1,\dots, \omega_d).d\omega_1\dots d\omega_d$$

Alors $\mathbb{E}(X) = \int_\Omega X.d\mathbb{P} = \int_\Omega X(\omega).f(\omega).d\omega$

$$= \int \dots \int_{\Omega=\mathbb{R}^\wedge d} X_A(\omega_1,\dots, \omega_d).f(\omega_1,\dots, \omega_d).d\omega_1\dots d\omega_d$$

Rappel
$\mathbb{1}_A$ fonction indicatrice de A :
Si $x \in A$, $\mathbb{1}_A(x) = 1$, sinon $\mathbb{1}_A(x) = 0$

En particulier, si Z est une v.a. $(\Omega,\mathbb{P})\to E$ et si $\phi : E \to \mathbb{R}$ est une fonction,
$X = \phi_o Z$ définie de $(\Omega,\mathbb{P})\to\mathbb{R}$ par $X(\omega) = \phi(Z(\omega))$ est une v.a.r.

donc $\mathbb{E}(\phi(Z)) = \int_\Omega \phi(Z(\omega)).\mathbb{P}(\{\omega\})$ dans le premier cas

$= \int \dots \int_\Omega \phi(Z(\omega)).f(\omega).d\omega$ dans le deuxième cas

En conséquence si X et Y sont deux v.a.r définies de $(\Omega,\mathbb{P})\to\mathbb{R}$, X + Y est encore une v.a.r. et on a

$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ (*Linérarité de l'espérance*)

De la même manière, pour $\lambda,\mu \in \mathbb{R}$, on a $\mathbb{E}(\lambda X + \mu Y) = \lambda\, \mathbb{E}(X) + \mu\, \mathbb{E}(Y)$

$$\mathbb{E}\left(\sum_{i=1}^{n} \lambda_i X_i\right) = \sum_{i=1}^{n} \lambda_i \mathbb{E}(X_i)$$

Lorsque X, Y, Z sont définis sur $\Omega = \mathbb{R}^d$, $\mathbb{P}$ tel que $\mathbb{P}(A) = \int_\Omega \mathbb{1}_A(\omega).f(\omega).d\omega$, avec Z défini par Z=X+Y, c'est-à-dire $Z(\omega) = X(\omega)+Y(\omega)$, alors :

$$\begin{aligned}\mathbb{E}(X+Y) = \mathbb{E}(Z) &= \int_\Omega Z(\omega).f(\omega).d\omega \\ &= \int_\Omega (X(\omega)+Y(\omega)).f(\omega).d\omega \\ &= \int_\Omega X(\omega).f(\omega).d\omega + \int_\Omega Y(\omega).f(\omega).d\omega \\ &= \mathbb{E}(X) + \mathbb{E}(Y)\end{aligned}$$

<u>Croissance de l'espérance</u> : Si X et Y sont deux v.a.r. définies toutes deux sur le même espace $(\Omega,\mathbb{P})$ et telles que $X\leq Y$ (c'est-à-dire $\forall\ \omega \in \Omega$, $X(\omega)\leq Y(\omega)$), alors $\mathbb{E}(X)\leq\mathbb{E}(Y)$

<u>Interprétation de l'espérance</u> : $\mathbb{E}(X)$ est un nombre réel, une constante particulière : $\mathbb{E}(X)$ est la meilleure prévision (non aléatoire) qu'on puisse faire sur la valeur aléatoire $X(\omega)$.
Cette précision est erronée, mais c'est la prévision qui minimise l'erreur, en un sens.
En particulier, si $X(\omega) = a\ \forall\ \omega \in \Omega$ (où a est un nombre réel fixé), alors

$$\begin{aligned}\mathbb{E}(X) = \int_\Omega X.d\mathbb{P} &= \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}) \\ &= a\left(\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})\right) \\ &= a.\mathbb{P}\left(\cup_{\omega\in\Omega}\ \mathbb{P}(\{\omega\})\right) \\ &= a.\mathbb{P}(\Omega) = a\end{aligned}$$

Autrement dit, si $X(\omega)=a\ \forall\ \omega \in \Omega$, alors X est facile à prévoir, car X n'est pas aléatoire. Dans ce cas la meilleure prévision possible de X par une constante est bien sûr a. Il est donc heureux que l'on ait $\mathbb{E}(X) = a$ dans ce cas particulier.

Nota : Si $\Omega = \mathbb{R}^d$ et $\mathbb{P}$ défini par $\mathbb{P}(A) = \int_\Omega \mathbb{1}_A(\omega).f(\omega).d\omega$ avec $\begin{cases} f \geq 0 \\ \int_\Omega f(\omega).d\omega = 1 \end{cases}$,

et si $X(\omega)=a \; \forall \; \omega \in \Omega$

alors $\mathbb{E}(X)= \int_\Omega X(\omega).f(\omega).d\omega = \int_\Omega a.f(\omega).d\omega = a.\int_\Omega f(\omega).d\omega = a \times 1 = a$

> **Exemples**
>
> À voir sur les notes de cours

Remarque : L'espérance de X est la meilleure prévision possible de la variable aléatoire X à l'aide d'une constante, au sens suivant : l'erreur qu'on commet lorsqu'on approche X par la constante $a \in \mathbb{R}$ est $|X-a|$ donc la moyenne de cette erreur au carré est $\psi(a)=\mathbb{E}[|X-a|^2]$

Propriété : le minimum de $a \rightarrow \psi(a)$ est atteint pour le choix $a_0=\mathbb{E}(X)$

En effet, $\psi(a)=\mathbb{E}[(X-a)^2] = \mathbb{E}[(X-\mathbb{E}(X)+\mathbb{E}(X)-a)^2]$

$$= \mathbb{E}[(X-\mathbb{E}(X))^2 + 2\,(X-\mathbb{E}(X))\underbrace{(\mathbb{E}(X)-a)}_{(cte)} + \underbrace{(\mathbb{E}(X)-a)}_{(cte)}{}^2]$$

$$= \mathbb{E}[(X-\mathbb{E}(X))^2] + 2.(\mathbb{E}(X)-a).\mathbb{E}[(X-\mathbb{E}(X)] + (\mathbb{E}(X)-a)^2$$

$$= \mathbb{E}[(X-\mathbb{E}(X))^2] + \cancel{2.(\mathbb{E}(X)-a).(\mathbb{E}(X)-\mathbb{E}(X))} + (\mathbb{E}(X)-a)^2$$

D'où $\psi(\mathbb{E}(X))=\mathbb{E}[(X-\mathbb{E}(X))^2] \leq \psi(a) \; \forall \; a \in \mathbb{R}$

Définition

La grandeur $\mathbb{E}[(X-\mathbb{E}(X))^2] = \mathbb{E}(X^2)-\mathbb{E}(X)^2$ est appelée la **variance de X** et est notée **Var(X)**. Elle mesure le degré d'imprévisibilité de la variable aléatoire X : c'est l'erreur commise lorsqu'on approche X par son espérance $\mathbb{E}(X)$, c'est-à-dire par la constante qui est la plus proche de X. La variance est toujours positive ou nulle. Si Var(X)=0, alors X est constante.

On peut voir $\mathbb{E}(X)$ comme la projection orthogonale de X sur la droite des constantes, Var(X) comme le carré de la distance de X à cette droite.

La relation $\psi(a) = \mathbb{E}[(X-a)^2] = Var(X) + (a-\mathbb{E}(X))^2$ s'interprète alors comme une application du théorème de Pythagore.

III / Variables aléatoires discrètes

a) Généralités

Définition : Une v.a. $X=\Omega \rightarrow E$ est dite discrète si son image $X(\Omega)$ est finie ou dénombrable.

<u>Propriété</u> : La loi de X, $\mathbb{P}_X$, est alors déterminée par la donnée des nombres $\mathbb{P}_X(\{x\})=\mathbb{P}(X=x)$ pour $x \in X(\Omega)$.

<u>Rappel</u>

$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega \ / \ X(\omega) \in A\}) = \mathbb{P}(X \in A)$

$\mathbb{P}_X(\{x\}) = \mathbb{P}(X^{-1}(\{x\})) = \mathbb{P}(\{\omega \in \Omega \ / \ X(\omega) \in \{x\}\}) = \mathbb{P}(X=x)$

Nota : $\{\forall\, x\ \mathbb{P}(X=x) \geq 0\}$ et $\{\sum_{x\in X(\Omega)} \mathbb{P}(X=x)=1\}$

Exemples de lois discrètes

X suit la :

- Loi uniforme sur E fini, par exemple $E=\{e_1,\dots,e_n\}$

Alors $\mathbb{P}(x=e_i) = 1/n,\ 1\leq i\leq n$

Si $E = \{1,2,\dots,6\}x\{1,2,\dots,6\}$, $X(\omega) =(X_1(\omega),X_2(\omega))$ est le résultat des deux premiers lancers d'un dé.

$\mathbb{P}(X=(i,j))=1/36,\ 1\leq i,j\leq 6$

$\mathbb{P}(X_1=X_2)=\mathbb{P}(X\in A)$ où A est la première bissectrice.

$\mathbb{P}(X\in A) = \sum_{x\in A\cap X(\Omega)} \mathbb{P}(X = x) = \sum_{i=1}^{6} \mathbb{P}(X = (i,i)) = 6 * 1/36 = 1/6$

- $E = \{1,2,\dots,n,\dots\} = \mathbb{N}^*$

X désigne le numéro du premier lancer où apparait un numéro pair.

$\mathbb{P}(X=k) = (1/2)^k$ pour $k\geq 1$

$= 0$ sinon

$\mathbb{P}(X\leq\infty) = \sum_{k\geq 1} \mathbb{P}(X = x) = ½ + ¼ + \dots \ = ½\,(1+ ½ + ¼ + \dots) = ½ \frac{1}{1-\frac{1}{2}} = 1$

Car $1 + x + x^2 + \dots = \frac{1}{1-x}$ si $|x|\leq 1$

$\mathbb{P}(X<\infty)=1 \Rightarrow$ on finit toujours par voir apparaître un numéro pair (pas très surprenant)

- $E=(\mathbb{N}^*)^2$

$X = (X_1,X_2)$ = numéro des lancers où apparaissent les deux premiers numéros pairs.

$X(\Omega) = \{(n,m) \in E \ / \ n\geq 1,\ m\geq n+1\}$

$\mathbb{P}(X=(n,m)) = (½)^k$ si $n\geq 1$ et $m\geq n+1$

$= 0$ sinon

$\mathbb{P}(X_2=2X_1) = \mathbb{P}(X\in A)$ avec $A = \{(x,y) \ / \ y=2x\}$

$= \dots = \sum_{n\geq 1}(\frac{1}{2})^{2n} = ¼ * \frac{1}{1-\frac{1}{4}} = 1/3$

b) Espérances des v.a.r. discrètes

On suppose ici que $X(\Omega) \subset \mathbb{R}$
$X(\Omega)$ est fini ou dénombrable.

On a alors :

THÉORÈME : $\mathbb{E}(X) = \sum_{x\in X(\Omega)} x\mathbb{P}(X=x)$

Conséquence : Si X est une v.a. discrète (non nécessairement à valeur réelle) et ϕ une application de $X(\Omega)$ dans $\mathbb{R}$, alors $\mathbb{E}(\phi(X)) = \sum_{x\in X(\Omega)} \phi(x)\mathbb{P}(X=x)$

Ces formules restent valables lorsque Ω est infini non dénombrable.

Cas particulier important : $\phi(x) = x^2$

$\mathrm{Var}(X) = \left(\sum_{x\in X(\Omega)} x^2\mathbb{P}(X=x)\right) - \left(\sum_{x\in X(\Omega)} x\mathbb{P}(X=x)\right)^2$

Dans les domaines de sommation ci-dessus on peut remplacer $X(\Omega)$ par un ensemble $B \supsetneq X(\Omega)$ (on ajoute donc des termes "$y\ \mathbb{P}(X=y)$" ou "$\phi(y)\mathbb{P}(X=y)$", tels que $y \notin X(\Omega)$ donc $\{\omega\in\Omega \ / \ X(\omega)=y\}=\emptyset$ et par conséquent $\mathbb{P}(X=y)=0$)

Exemple :

Si X désigne le numéro du premier lancer pari, alors $\mathbb{P}(X=k)=(1/2)k$, $k\geq 1$. Intuitivement, les lancers pairs et impairs sont aussi fréquents d'où une fréquence ½, et un temps d'attente moyen $\mathbb{E}(X)=2$

Le calcul donne $\mathbb{E}(X)=\sum_{k\geq 1} k\left({}^{1}/_{2}\right)k = \phi(1/2)$, avec $\phi = \sum_{k\geq 1} kxk$

On remarque alors que $\phi(x)/x = \sum_{k\geq 1} k.\, xk\text{-}1 = \frac{\partial\theta}{\partial x}(x)$

avec $\theta(x) = \sum_{k\geq 1} xk = 1 + x + x^2 + \ldots = \frac{1}{1-x}$ si $|x|\leq 1$

donc $\phi(x) = x\frac{\partial\theta}{\partial x}(x) = x\frac{1}{(1-x)^2}$ et donc $\phi(1/2) = 2$

c) Lois jointes de v.a.r. discrètes

Si $X_1, \ldots, X_k : \Omega \to E_k$ sont des variables aléatoires discrètes, alors $X(\omega)=(X_1(\omega),X_2(\omega),\ldots,X_k(\omega))$ définit une application $X : \Omega\to E_1\times\ldots\times E_k$ et on a $X(\Omega) \subset X_1(\Omega) \times \ldots \times X_k(\Omega)$ (les $X_i(\Omega)$ étant dénombrables ou finis) donc $X(\Omega)$ est fini ou dénombrable et X est une variable aléatoire discrète.

Définition : On appelle loi joinde de X_1, X_2,…,X_k la loi de la v.a. $X = (X_1,X_2,…,X_k)$ définie de $\Omega \to E_1 \times … \times E_k$. Cette loi jointe est décrite par la donnée de

$$\mathbb{P}(X=(x_1,…,x_k)) = \mathbb{P}(X_1=x_1 \text{ et } … \text{ et } X_k=x_k)\ \forall\, x_1 \in E_1, …, x_k \in E_k.$$

Définition : On dit que les X_i sont indépendants **ssi** $\forall\, x_1 \in E_1, …, x_k \in E_k$,

$$\mathbb{P}(X_1=x_1 \text{ et } … \text{ et } X_k=x_k) = \prod_{i=1}^{k} \mathbb{P}(X_i=x_i)$$

Exemple :

- La hauteur H et la couleur C d'une carte sont indépendantes. En effet,

$\mathbb{P}(H=h \text{ et } C=c) = 1/52 = 1/13 * ¼ = \mathbb{P}(H=h)\mathbb{P}(C=c)$,

$\forall\, h \in \{2,3,4,…,9,10,V,D,R,A\}$ et $\forall \{\blacklozenge,\heartsuit,\clubsuit,\spadesuit\}$

- Les numéros des deux premiers tirages pairs X_1 et X_2 (cf exemples sur les lois discrètes) ne sont pas indépendants car $\mathbb{P}(X_1=3)> 0$, $\mathbb{P}(X_2=2)>0$ et $\mathbb{P}(X_1=3 \text{ et } X_2=2)=0 \neq \mathbb{P}(X_1=3)\mathbb{P}(X_2=2)$

c) Variable de Bernoulli

On dit que X est une variable de Bernoulli si et seulement si $X(\Omega) \subset \{0,1\}$ ("X vaut 0 ou 1")

Si $\mathbb{P}(X=1)=p$, alors $\mathbb{P}(X=0)=1-\mathbb{P}(X=1)=1-p$ (avec $0 \leq p \leq 1$)

On dit alors que X suit la loi de Bernoulli de paramètre p

Exemple

Si $X(\omega) = \mathbb{1}_A(\omega)\ \forall\, \omega$ alors X suit une loi de Bernoulli de paramètre $\mathbb{P}(A)$.

Les fonctions indicatrices suivent les lois de Bernoulli

On a **$\mathbb{E}(X)=p$**

$Var(X)=p(1-p)$

Nota : $p \to p-p^2 = p(1-p)$ atteint sont maximum pour $p=0.5$

Les variables de Bernoulli les moins prévisibles sont celles de paramètre 0.5.

Exemples :

(voir première partie sur la page 10 du cours manuscrit, la seconde partie est celle sur la loi géométrique)

C_{bis}) La loi géométrique G(p)

On dit qu'une variable X suit **la loi géométrique de paramètre p** si

$$\mathbb{P}(X=k) = (1-p)^{k-1}p \text{ pour } k\geq 1$$
$$= 0 \text{ sinon}$$

On vérifie alors que $g_X(s) = \mathbb{E}(s^X) = \sum_{k\geq 1} \mathbb{P}(X=k)s^k = \frac{ps}{1-(1-p)s} = \frac{p}{1-p}\left(\frac{1}{1-(1-p)s} - 1\right)$

$g_X'(s) = \frac{p}{1-p}\frac{1-p}{(1-(1-p)s)^2}$

$g_X'(1) = \boxed{\mathbb{E}(X) = 1/p}$

$g_X''(1) = \mathbb{E}(X(X-1)) = \frac{2(1-p)}{p^2}$

D'où $\mathrm{Var}(X) = g_X''(1) + \mathbb{E}(X) - \mathbb{E}(X)^2 = \boxed{\frac{1-p}{p^2} = \mathrm{Var}(X)}$

$\mathbb{P}(X>k) = (1-p)^k$

Nota : La loi géométrique est la loi du nombre de tentatives nécessaires pour obtenir un succès, lorsque ces tentatives sont indépendantes et que la probabilité d'obtenir un succès est p à chaque tentative.

Supposons qu'un algorithme grossier mais rapide résolve un problème de taille n, en n^α pas, et que la solution soit satisfaisante avec proba $1/n^2$ (si elle n'est pas satisfaisante on recommence jusqu'à obtenir un succès). Soit X le nombre de tentatives jusqu'au succès, alors $X \sim G(1/n^2)$ et $\mathbb{E}(X)=n^2$, d'où un algorithme qui résout le problème "parfaitement" en $n^\alpha.\mathbb{E}(X) = n^{\alpha+2}$ pas.

IV / Variance, covariance et concentration

a) Propriétés générales

Définition : la covariance de deux variables aléatoires réelles X et Y est notée Cov(X,Y). Elle est définie par

$\mathrm{Cov}(X,Y)=\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

$\quad = \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y)))$

Propriétés :

- C'est une forme bilinéaire symétrique :

  $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$

  $\mathrm{Cov}(\sum_i \lambda_i X_i,Y) = \sum_i \lambda_i \mathrm{Cov}(X_i,Y)$

$\mathrm{Cov}(\sum_i \lambda_i X_i, \sum_j \mu_j Y_j) = \sum_{i,j} \lambda_i \mu_j \mathrm{Cov}(X_i, Y_j)$

- La forme quadratique associée est la variance : Var(X)=Cov(X,Y)

  En particulier $\mathrm{Var}(\lambda X) = \lambda^2 \mathrm{Var}(X)$

  $\mathbf{Var}(\sum_i \lambda_i X_i) = \sum_i^n \lambda_i^2 \mathbf{Var}(X_i) + 2\sum_{1\leq i,j\leq n} \mathbf{Cov}(X_i, X_j)$
- C'est une forme positive : $\mathrm{Var}(X) \geq 0 \ \forall X$

  → inégalité de Cauchy Schwarz $|\mathrm{Cov}(X,Y)|^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)$
- Les constantes appartiennent au noyau de Cov (sont des vecteurs isotropes)

  $\mathrm{Cov}(a,X)=0 \quad \forall a \in \mathbb{R}$ et $\forall$ X variable aléatoire réelle

  $\mathrm{Cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y)$

  $\mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)$
- $\frac{Cov(X,Y)}{\sqrt{Var(X) Var(Y)}} = \varsigma_{X,Y}$ coefficient de corrélation de Pearson satisfait $|\varsigma_{X,Y}| \leq 1$

  $\varsigma_{X,Y} = 1 \Rightarrow \exists\, a,b \in \mathbb{R}_+^* \times \mathbb{R}$ tel que $Y = aX + b$

  $\varsigma_{X,Y} = -1 \Rightarrow \exists\, a,b \in \mathbb{R}_-^* \times \mathbb{R}$ tel que $Y = aX + b$

  → parallèle avec le produit scalaire

Exemple

N le nombre de lettres dans la bonne enveloppe

$N = X_1 + \ldots + X_n$ avec $X_i = \mathbb{1}_{\text{"la i-ème lettre dans la bonne enveloppe"}}$

$\mathrm{Var}(X_i) = p(1-p)$ pour $p=1/n$

$= \frac{n-1}{n^2}$

$\mathrm{Cov}(X_i,X_j) = \mathbb{E}(X_iX_j) - 1/n^2$ $\qquad \mathbb{P}(X_iX_j=1)=\frac{1}{n(n-1)}$

$= \frac{1}{n(n-1)} - \frac{1}{n^2}$

$= \frac{1}{n^2(n-1)}$

$\mathrm{Var}\, N = \sum Var(X_i) + 2\sum Cov(X_i,X_j) = n.\frac{n-1}{n^2} + 2.\frac{n(n-1)}{2}\,\frac{n-1}{n^2} = 1$

b) variance et concentration

Inégalité de Markov : Soit X une variable aléatoire supérieure ou égale à 0 et soit a un réel strictement positif. Alors $\mathbb{P}(X \geq a) \leq {}^{\mathbb{E}(X)}/_{a}$

Inégalité de BienAymé Tchebysheff : Soit Y une v.a.r. de variance finie (>0) et a>0.

Alors $\mathbb{P}(|Y - \mathbb{E}(Y)| \geq a.\sqrt{\mathrm{Var}(Y)}) \leq \frac{1}{a^2}$

c) Indépendance et Variance

THEOREME : Si $X_1,\dots,X_d$ sont des v.a. indépendantes ($X_i : \Omega \to E_i$) et $\phi_i : E_i \to \mathbb{R}$, alors

$$\boxed{\mathbb{E}(\phi_1(X_1)\ \phi_2(X_2)\ \dots \phi_d(X_d)) = \prod_{i=1}^{d} \mathbb{E}(\phi_i(X_i))}$$

En particulier :

Corollaire 1 : Si $X_1,\dots,X_d$ sont des v.a.r indépendantes, alors

$$\mathbb{E}(X_1 X_2 \dots X_d) = \prod_{i=1}^{d} \mathbb{E}(X_i)$$

Corollaire 2 : Si X et Y sont indépendantes ($X \coprod Y$), $Cov(X,Y) = 0$

Corollaire 3 : Si $X_1 \coprod \dots \coprod X_n$,

$Var(\sum_i^n \lambda_i X_i) = \sum_i^n \lambda_i^2 Var(X_i)$

$Var(\sum_i^n X_i) = \sum_i^n Var(X_i)$

Corollaire 4 : Si $X_1,\dots,X_n$ sont indépendantes et de même loi (donc de même espérance $m=\mathbb{E}(X1)=\dots= \mathbb{E}(X_n)$ et de même variance $\sigma^2=Var(X_1)=\dots=Var(X_n)$) alors on aura :

$\mathbb{P}(X_1 + \dots + X_n \in\ ]nm-a\sigma\sqrt{n}, nm+a\sigma\sqrt{n}[\ ) \geq 1 - 1/a^2$

Inégalité de Hoeffding

On suppose que les v.a.r. X_i sont indépendantes et satisfont $\mathbb{P}(a_i \leq X_i \leq b_i)=1$, $1 \leq i \leq n$.

On pose $S_n = X_1 + \dots + X_n$

Alors $\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq e^{-2t^2/\sum_1^n (bi-ai)^2}$

$\mathbb{P}(S_n - \mathbb{E}(S_n) \leq -t) \leq e^{-2t^2/\sum_1^n (bi-ai)^2}$

$\mathbb{P}(|S_n - \mathbb{E}(S_n)| \geq t) \leq 2e^{-2t^2/\sum_1^n (bi-ai)^2}$

Exemple
$Xi \sim Bernoulli(\frac{1}{2})$ $ai = 0$, $bi=1$, $\sum_1^n (bi - ai)^2=n$
$\mathbb{E}(Sn)=n/2$ et $Var(Sn) = n/4$
B.T : $\mathbb{P}(\lvert Sn - \mathbb{E}(Sn)\rvert \geq a\sqrt{n}) \leq \frac{1}{4a^2}$
Hoeffding : $\mathbb{P}(\lvert Sn - \mathbb{E}(Sn)\rvert \geq a\sqrt{n}) \leq 2\ e^{-2a^2}$

Démonstration de l'inégalité sur les notes de cours

V / Lois binomiale et de Poisson

$\Omega = 000100111100000100\ldots = \omega_1\, \omega_2 \ldots$ $X_i(\omega) = \omega_i \in \{0,1\}$

Combien de 1 dans un mot de longueur n ? dans $\omega_1\, \omega_2 \ldots \omega_n$?

Définition – Loi binomiale : Si X vérifie

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{si } 0 \leq k \leq n$$
$$= 0 \quad \text{sinon}$$

On dit que X suit la **loi binomiale de paramètres n et p** (ou que X suit la loi B(n,p))

Alors $\mathbb{E}(X) = np$

$\mathrm{Var}(X) = np(1-p)$

Supposons qu'on accélère le temps : au lieu de jouer une fois par unité de temps → p succès par unité de temps, on joue n fois par unité de temps mais avec toujours p succès par unité de temps → probabilité de succès p/n à chaque tentative

Définition – Loi de Poisson : Si X vérifie

$$\mathbb{P}(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{pour } k \geq 0$$
$$= 0 \quad \text{sinon}$$

On dit que X suit la **loi de Poisson de paramètre λ**. ($\lambda > 0$)

On a $\mathbb{E}(X) = \lambda$

$\mathrm{Var}(X) = \lambda$

Loi binomiale négative

Où se trouve le k-ième "1" ?

La position X_k du k-ième "1" suit **la loi binomiale négative de paramètre k et p** définie par

$$\mathbb{P}(X_k=l) = \binom{l-1}{k-1} (1-p)^{l-k} p^k \quad \text{si } l \geq k$$
$$= 0 \quad \text{sinon}$$

La loi binomiale négative de paramètre 1 et p n'est autre que **la loi de Pascal** (ou loi géométrique) de paramètre p.

On a $\mathbb{E}(X_k)=\frac{k}{p}$

$\text{Var}(X) = \frac{k(1-p)}{p^2}$

Loi Hypergéométrique $H(N_1,N_2,n)$

C'est la loi du nombre de 1 au bout de n tirages sans remise dans une urne contenant N_1 boules numérotées "1" et N_2 boules numérotées "0". X suit la loi hypergéométrique ssi :

$$\mathbb{P}(X=k) = \begin{cases} \dfrac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N_1+N_2}{n}} & \text{si } 0\leq k\leq n \\ = 0 & \text{sinon} \end{cases}$$

Si $p = \frac{N_1}{N_1+N_2}$ et si $n \ll N_1 + N_2$ alors $\mathbb{P}(X=k) \simeq \binom{n}{k} p^k(1-p)^{n-k}$

(Sans remise ≅ avec remise, si $n \ll N_1 + N_2$)

On a $\mathbb{E}(X)=np$

$\text{Var}(X) = np(1-p)\frac{N_1+N_2-n}{N_1+N_2-1}$

VI / Fonctions génératrices

Définition : Soit X une v.a. à valeurs dans $\mathbb{N}$, on appelle fonction (ou série) génératrice de X la somme de la série entière $\mathbb{P}(X=k)s^k$, lorsqu'elle converge. On la note $G_X(s)$

$$G_X(s)= \sum_{k\geq 0} \mathbb{P}(X = k)s^k= \mathbb{E}(s^X)$$

Propriétés :

- La série converge absolument dès que $|s|\leq 1$ (il est possible qu'elle converge aussi pour $|s|>1$) . En particulier, $G_X(1)=1$
- {G_X est dérivable à gauche en 1} si et seulement si $\{\sum k\mathbb{P}(X=k)<+\infty\}$ et $G_X'(1)=\sum k\mathbb{P}(X=k)=\mathbb{E}(X)$
- {G_X est deux fois dérivable à gauche en 1} si et seulement si $\{\sum k^2\mathbb{P}(X=k)<+\infty\}$ et $G_X''(1)=\sum k(k-1)\mathbb{P}(X=k)=\mathbb{E}(X^2)-\mathbb{E}(X)$

  En particulier $\text{Var}(X)=G_X''(1)+G_X'(1)-G_X'(1)^2$
- Si $G_X=G_Y$ alors $\forall k$ $\mathbb{P}(X=k)=\mathbb{P}(Y=k)$, X et Y ont même loi.
- Si $X_1 \amalg X_2 \amalg \ldots \amalg X_n$, alors $G_{X_1+X_2+\ldots+X_n}(s)=\prod_{i=1}^{n} G_{X_i}(s)$

Exemples et démonstrations
Voir notes de cours (encore !)

VII / Lois jointes de variables discrètes

1) Définitions

* Si X et Y sont deux variables discrètes, **la loi jointe de X et Y** est la loi de la variable discrète Z=(X,Y). Cette loi jointe est décrite par la donnée de la famille de nombres

$$\{\mathbb{P}(Z=(x,y))\}_{(x,y)\in X(\Omega)\times Y(\Omega)} = \{\mathbb{P}(X=x \text{ et } Y=y)\}_{x\in X(\Omega),\, y\in Y(\Omega)}$$

* De même la loi jointe de k variables aléatoires $X_1,\dots,X_k$ discrètes est la loi de $Z=(X_1,\dots,X_k)$

Elle est décrite par $\mathbb{P}(Z=(x_1,\dots,x_k))=\mathbb{P}(X_1=x_1 \text{ et } \dots \text{ et } X_k=x_k)$ pour $x_i \in X_i(\Omega)$, avec $1\leq i\leq k$

On note souvent $\mathbb{P}(X=x \text{ et } Y=y)=P_{x,y}$ et $\mathbb{P}(X_1=x_1 \text{ et } \dots \text{ et } X_k=x_k)=PP_{x_1,\dots,x_k}$

2) Calcul de la loi de X et de Y connaissant la loi jointe de (X,Y)

$\mathbb{P}(X=x) = \mathbb{P}\left(\dot{\bigcup}_{y\in Y(\Omega)}(X=x \text{ et } Y=y)\right) = \sum_{y\in Y(\Omega)} \mathbb{P}(X = x \text{ et } Y = y)$

$$\boxed{\mathbb{P}(X=x) = \sum_{y\in Y(\Omega)} P_{x,y}}$$

On note souvent $P_{x,\cdot} = \sum_{y\in Y(\Omega)} P_{x,y}$ donc $\mathbb{P}(X=x) = P_{x,\cdot}$

On dit que $(P_{x,\cdot})_{x\in X(\Omega)}$ ou $(\mathbb{P}(X=x))_{x\in X(\Omega)}$ est une <u>marginale</u> (la première marginale) de $(P_{x,\cdot})_{x,y\in X(\Omega)\times Y(\Omega)}$

De même $\mathbb{P}(Y=y) = \sum_{x\in X(\Omega)} \mathbb{P}(X = x \text{ et } Y = y)$

$= \sum_{x\in X(\Omega)} P_{x,y}$

$= P_{\cdot,y}$

En général

$\mathbb{P}(X_i=u) = \sum \mathbb{P}(X_1 = x_1 \text{ et } \dots \text{ et } X_{i-1} = x_{i-1} \text{ et } X_i = u \text{ et } X_{i+1} = x_{i+1} \text{ et } \dots \text{ et } X_k = x_k)$

Avec $(x_1, \dots, x_k) \in X_1(\Omega)\times\dots\times X_k(\Omega)$

Exemple (important ?)
Voir notes de cours… (p6 cours du 24/10)

Rappels

$X \sim$ loi multinomiale de paramètre d, n, $(P_1,\ldots,P_d)$

Si $X = (X_1,\ldots,X_d)$ vérifie

$$\mathbb{P}(X=(k_1,\ldots,k_d)) = \frac{n!}{k_1!k_2!\ldots k_d!} {P_1}^{k_1} \ldots {P_d}^{k_d} \quad \text{si } k_1+\ldots+k_d = n$$

$$= 0 \quad \text{sinon}$$

En particulier, si Y suit la loi Bin(n,p), alors (Y,n-Y) suit la loi multinomiale(2,n,p,1-p)

$$\mathbb{P}((Y,n-Y)=(k_1,k_2)) = \frac{n!}{k_1!k_2!} p^{k_1} (1-p)^{k_2} \quad \text{si } k_1 + k_2 = n$$

$$= \binom{n}{k_1} p^{k_1} (1-p)^{k_2} \quad \text{si } k_2 = n - k_1$$

$$= \mathbb{P}(Y=k_1)$$

$$(a_1 + \ldots + a_d)^n = \sum \binom{n}{k_1 \ldots k_d} {a_1}^{k_1} \ldots {a_d}^{k_d}$$