Table des matières

I Cadre théorique

I.1 Ensembles dénombrables

I.1.1 Définition
Soit E un ensemble quelconque.

1. On dit que E est dénombrable ssi il existe une application $\varphi : \mathbb{N} \to E$ bijective. En d'autre termes, on peut écrire $E = \{x_0, x_1, \dots\}$ sans oublier un seul élément.
2. On dit que E est au plus dénombrable ssi E est fini ou dénombrable.

I.1.2 Fini ou dénombrable

Les ensembles finis ou dénombrables sont exactement les ensembles pour lesquels on peut numéroter les éléments, ou encore les décrire sous la forme $\{x_n \ n \in \mathbb{N}\}$ (quitte à prendre une infinité de fois la même valeur pour x_n dans le cas des ensembles finis).

I.1.3 Théorème
1. $\mathbb{N}\backslash\{0\}$ est dénombrable.
2. $\mathbb{Z}$ est dénombrable.
3. $\mathbb{N} \times \mathbb{N}$ et $\mathbb{Z} \times \mathbb{Z}$ sont dénombrables.
4. Si E et F sont dénombrables alors $E \times F$ est dénombrable.

Preuve.
1. Encore heureux ! $\varphi : \left\{ \begin{array}{ccc} \mathbb{N} & \to & \mathbb{N}\backslash\{0\} \\ n & \mapsto & n+1 \end{array} \right.$ est une bijection convenable.
2. Exhibons une bijection de $\mathbb{Z}$ dans $\mathbb{N}$. On pose
$$\varphi : \left\{ \begin{array}{ccc} \mathbb{Z} & \to & \mathbb{N} \\ k & \mapsto & \begin{cases} 2n \text{ si } n \geqslant 0 \\ 2 \times (-n) + 1 \text{ si } n < 0 \end{cases} \end{array} \right. .$$
φ est une bijection. Pour le prouver on peut soit examiner l'injectivité et la surjectivité, soit exhiber sa réciproque.
3. cf 4.
4. Notons $E = \{e_i | \ i \in \mathbb{N}\}$ et $F = \{f_i | \ i \in \mathbb{N}\}$. L'idée ici et d'énumérer tous les éléments de $E \times F$ “par diagonale” : on représente E sur l'axe des abscisses, F sur l'axe des ordonnées (un élément de chaque sur chaque entier, e_0, f_0 situés en 0).
On énumère les éléments de $E \times F$ de la manière suivante : pour chaque $k \in \mathbb{N}$, on part de (e_k, f_0) (graphiquement sur l'axe des abscisses), puis on considère $(e_{k-1}, f_1), (e_{k-2}, f_2) \dots (e_0, f_k)$.
Plus précisément, si $(i, j) \in \mathbb{N}^2$, on note $k = i + j$ et alors on a déjà rempli k diagonales donc numéroté $\sum_{p=1}^{k} p = \frac{k(k+1)}{2}$ éléments, et (e_i, f_j) est l'élément numéro $\frac{k(k+1)}{2} + j$ (on vient de créer la bijection...) ■

I.1.4 Remarque

On doit pouvoir prouver que tout ensemble inclus dans un ensemble dénombrable est fini ou dénombrable. Ainsi $\mathbb{Q}$ doit être dénombrable, mais ce n'est pas au programme.

I.1.5 Coin culture

$\mathbb{R}$ n'est pas dénombrable, $\mathcal{P}(\mathbb{N})$ non plus. Il semble alors évident que $\mathbb{C}, \mathbb{R}^I, \mathcal{C}^\infty(I,\mathbb{R})$ ne sont pas dénombrables (pour le dernier, considérer le sous ensemble des fonctions constantes...).

I.1.6 Objectif

On souhaite étendre la notion de variable aléatoire à ces variables à valeurs dans un ensemble dénombrable (le cas fini est traité en 1ère année). Un des buts est de pouvoir modéliser le genre de situation suivante :

On joue à pile ou face jusqu'à ce que la pièce tombe sur pile. Quel est le nombre moyen de coup ? Le problème pour l'instant est qu'on ne peut pas borner à priori le nombre de coups à jouer et donc la variable aléatoire dont la valeur est ce nombre de coup est a priori à valeurs dans $\mathbb{N}\backslash\{0\}$.

I.2 Espaces probabilisable

I.2.1 Notation

Si les A_i sont des ensembles pour $i \in \mathbb{N}$, on note $\bigcup_{i\in\mathbb{N}} A_i = \{x|\ \exists i \in \mathbb{N}\ x \in A_i\}$ la réunion de ces ensembles et $\bigcap_{i\in\mathbb{N}} A_i = \{x|\ \forall i \in \mathbb{N}\ x \in A_i\}$ leur intersection.

I.2.2 Définition

Soit Ω un ensemble que l'on appellera univers. Une **tribu** sur Ω est un sous ensemble T de $\mathcal{P}(\Omega)$ (les éléments de T sont des sous ensembles de Ω) qui vérifie les 3 conditions :

1. $\Omega \in T$
2. $\forall A \in T\ A^C = \overline{A} = \Omega\backslash A \in T$.
3. Si $(A_n) \in T^{\mathbb{N}}$ alors $\bigcup_{n\in\mathbb{N}} A_n \in T$.

Les éléments de T (qui sont des ensembles, rappelons le) sont des **événements**. Le couple (Ω, T) est un **espace probabilisable**.

I.2.3 En pratique

Ω représente l'ensemble des issues possibles d'une expérience aléatoire et un événement un ensemble de résultat possibles. Pour reprendre notre jeu de pile ou face, on peut prendre $\Omega = \mathbb{N}\backslash\{0\}$ et un événement peut être "le jeu s'arrête en un nombre pair de coup" qui est l'ensemble $\{2n|\ n \in \mathbb{N}\backslash\{0\}\}$.

Bien souvent, Ω n'est pas précisé et sa connaissance n'est pas indispensable au bon déroulé de l'exercice. On supposera dans ce cas qu'une bonne tribu est choisie.

I.2.4 Proposition
Soit (Ω, T) un espace probabilisable.
1. $\emptyset \in T$.
2. Si $(A_n)_{n\in\mathbb{N}} \in T^{\mathbb{N}}$ $\bigcap_{n\in\mathbb{N}} A_n \in T$. De plus, $\bigcap_{n\in\mathbb{N}} A_n = \overline{\bigcup_{n\in\mathbb{N}} A_n^c}$

Preuve.
Exo ! ∎

I.3 Espace probabilisé

I.3.1 Définition

Soit Ω un ensemble et T une tribu sur Ω. Une **probabilité** sur Ω est une fonction $\mathbb{P}$ qui associe à chaque événement A une probabilité $\mathbb{P}(A) \in [0,1]$ avec les contraintes suivantes :
1. $\mathbb{P}(\Omega) = 1$
2. Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements incompatibles deux à deux (ie disjoints deux à deux), alors
$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n) \text{ : propriété de } \sigma\text{-additivité}$$

En particulier, toute série de la forme précédente doit converger vers un nombre dans $[0,1]$.

Le triplet $(\Omega, T, \mathbb{P})$ est appelé un **espace probabilisé**. Dans la suite du cours, nous utiliserons ces notations.

I.3.2 Cas fini

Si on considère une suite $(A_n)_{n\in\mathbb{N}}$ d'événements telle que $A_n = \emptyset$ pour tout $n \geqslant 2$, on retrouve la définition de 1ère année. Avec un nombre fini de A_n non vide, la σ-additivité est une propriété démontrée en 1ère année.

I.3.3 Mais pourquoi des tribus ?

Dans le cas où Ω est fini ou dénombrable, on pourra prendre $T = \mathcal{P}(\Omega)$ sans problème. Les choses se corsent singulièrement si on prend Ω non dénombrable.

Par exemple, on prouve (un “on” qui est bien en dehors du cadre de ce cours), qu'on ne peut pas poser $\Omega = [0,1], T = \mathcal{P}(\Omega)$ et la probabilité uniforme naturelle qui vérifie $\mathbb{P}([a,b]) = b-a$.

I.3.4 Définition
Soit A un événement.

1. Si $A \neq \emptyset$ et $\mathbb{P}(A) = 0$ on dit que A est **négligeable**.
2. Si $A \neq \Omega$ et $\mathbb{P}(A) = 1$ on dit que A est presque sûr.

I.4 Propriétés des probabilités

I.4.1 Proposition (Adaptation de la 1ère année)
Soit $(\Omega, T, \mathbb{P})$ un espace probabilisé. Soient A, B deux événements et $(A_n)_{n\in\mathbb{N}}$ une suite d'événements.

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(A^c) = \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
3. Si $A \subset B$ alors $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.
4. $\mathbb{P}(A\cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A\cap B)$ et $\mathbb{P}(A\cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A\cup B)$.
5. $\mathbb{P}\left(\bigcup\limits_{k=0}^{N} A_k\right) \leqslant \sum\limits_{k=0}^{N} \mathbb{P}(A_k)$ pour tout $N \in \mathbb{N}$.

Preuve.

1. $\Omega = \Omega \sqcup \emptyset$
2. $\Omega = A \sqcup A^c$
3. $B = A \sqcup (B\backslash A)$
4. $A\cup B = A \sqcup (B\backslash A)$
5. Par récurrence, en partant du cas $n=2$ prouvé par le point précédent. ■

I.4.2 Théorème
Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements.

1. Si $(A_n)_{n\in\mathbb{N}}$ est croissante au sens de l'inclusion ($\forall n \in \mathbb{N}\ A_n \subset A_{n+1}$) alors
$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mathbb{P}(A_n)$$
2. Si $(A_n)_{n\in\mathbb{N}}$ est décroissante au sens de l'inclusion ($\forall n \in \mathbb{N}\ A_{n+1} \subset A_n$) alors
$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mathbb{P}(A_n)$$

Le résultat important est l'existence de ces limites.

Preuve.
Le complémentaire d'une réunion étant l'intersection des complémentaires, nous prouvons seulement le premier point.

Supposons donc $(A_n)_{n\in\mathbb{N}}$ croissante au sens de l'inclusion. Alors la suite $(\mathbb{P}(A_n))_{n\in\mathbb{N}}$ est croissante et majorée par 1 donc converge.

Pour $n \in \mathbb{N}\backslash\{0\}$, posons $B_n = A_n \backslash A_{n-1}$ et $B_0 = A_0$. Alors les B_i sont disjoints deux à deux (pour imager, B_n est ce qu'il manquait à A_{n-1} pour devenir l'ensemble A_n qui est “plus grand”).

De plus, $\bigcup\limits_{k=0}^{n} B_k = A_n = \bigcup\limits_{k=0}^{n} A_k$. Ainsi $\bigcup\limits_{n\in\mathbb{N}} A_n = \bigcup\limits_{n\in\mathbb{N}} B_n$ et donc $\mathbb{P}\left(\bigcup\limits_{n\in\mathbb{N}} A_n\right) = \sum\limits_{n=0}^{+\infty} \mathbb{P}(B_n) = \mathbb{P}(B_0) + \sum\limits_{n=1}^{+\infty} (\mathbb{P}(A_n) - \mathbb{P}(A_{n-1}))$.

La série est télescopique et converge vers $\lim\limits_{n\to+\infty} \mathbb{P}(A_n) - \mathbb{P}(A_0)$, ce qui conclut la preuve. ■

I.4.3 Proposition (Sous-additivité)
Soit $(A_n)_{n\in\mathbb{N}} \in T^{\mathbb{N}}$ telle que $\sum \mathbb{P}(A_n)$ converge. Alors $\mathbb{P}\left(\bigcup\limits_{n\in\mathbb{N}} A_n\right) \leqslant \sum\limits_{n=0}^{+\infty} \mathbb{P}(A_n)$.

Preuve.
Pour $n \in \mathbb{N}$, on pose $B_n = \bigcup_{k=0}^{n} A_k$ qui est une suite croissante d'événements. Alors $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$ et d'après le théorème précédent, $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty}(\mathbb{P}(B_n))$.

Ainsi $\left(\mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right)\right)_{n\in\mathbb{N}}$ possède une limite et on peut passer à la limite l'inégalité 5 de la proposition I.4.1 (Rappel : l'hypothèse du passage à la limite des inégalité est seulement l'existence des limites). ∎

I.4.4 Evénements négligeables

Si tous les A_n sont négligeables, alors leur réunion l'est aussi.

II Calcul de probabilités

II.1 Systèmes complets

II.1.1 Proposition (Définition d'une probabilité)
Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements incompatibles deux à deux, alors

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n) \in [0,1]$$

et toute série de la forme précédente (série de probabilités d'événements deux à deux incompatibles) est une série à termes positifs convergente, de limite (somme de la série) dans $[0,1]$

II.1.2 Définition
Notons, pour $n \in \mathbb{N}$, A_n un événement (donc A_n est un sous-ensemble de Ω).

On dit que $(A_n)_{n\in\mathbb{N}}$ est un système complet d'événements ssi $\forall (i,j) \in \mathbb{N}^2$ $i \neq j \Rightarrow A_i \cap A_j = \emptyset$ (disjoints 2 à 2) et $\bigcup_{n\in\mathbb{N}} A_n = \Omega$.

II.1.3 Proposition
Si $(A_n)_{n\in\mathbb{N}}$ est un système complet d'événements, alors $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = 1$

Preuve.
C'est une conséquence immédiate de la définition. ∎

II.1.4 Définition
Notons, pour $n \in \mathbb{N}$, A_n un événement (donc A_n est un sous-ensemble de Ω).

On dit que $(A_n)_{n\in\mathbb{N}}$ est un système quasi-complet d'événements ssi

$$\forall (i,j) \in \mathbb{N}^2 \ i \neq j \Rightarrow A_i \cap A_j = \emptyset$$

(les A_n sont disjoints 2 à 2) et

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = 1.$$

En d'autre termes, la probabilité qu'aucun des A_n ne soit réalisé est nulle, et l'un exactement des A_n est réalisé presque sûrement (avec une probabilité 1, un et un seul des A_n est réalisé).

II.1.5 Exemple
Reprenons l'exemple du jeu de pile ou face. On pose A_n l'événement : le jeu s'arrête au $n+1$ième lancé, ie on a obtenu n fois face avant d'obtenir pile. On pose en plus A_{-1} l'événement : le jeu ne s'arrête pas.

Alors $(A_n)_{n\geqslant -1}$ est un système complet d'événements. Essayons de construire un probabilité raisonnable.

On doit avoir $\sum_{n=-1}^{+\infty} \mathbb{P}(A_n) = 1$. Il semble raisonnable de poser $\mathbb{P}(A_n) = \frac{1}{2^{n+1}}$ (quelle hypothèse faisons nous sur chaque lancé, sur la pièce ?)

Alors $\mathbb{P}(A_{-1}) = 0$ est la seule possibilité (calculer la somme des probabilités imposées), ce qui semble raisonnable.

Les événements $(A_n)_{n\in\mathbb{N}}$ forment un système quasi-complet d'événements et on peut exclure l'événement A_{-1} de notre système.

II.1.6 Proposition (Système complet d'événements)
Si $(A_n)_{n\in\mathbb{N}}$ est un système (quasi)-complet d'événements et B est un événement, alors

1. $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = 1$
2. $\mathbb{P}(B) = \sum_{n=0}^{+\infty} \mathbb{P}(B \cap A_n)$

On retrouve le cours de première année en prenant un système complet fini (tous les A_n sont vides sauf les quelques premiers).

Preuve.
B est la réunion disjointe des $B \cap A_n$ et éventuellement d'un ensemble de probabilité nulle. ∎

II.2 Probabilités conditionnelles

II.2.1 Définition-Proposition

Soit B un événement tel que $\mathbb{P}(B) > 0$.

1. Pour un événement A, la probabilité de A sachant B est $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
2. L'application $\mathbb{P}_B : \left\{ \begin{array}{lcl} T & \to & [0,1] \\ A & \mapsto & \mathbb{P}_B(A) = \mathbb{P}(A|B) \end{array} \right.$ est une probabilité. C'est la probabilité conditionnelle sachant B.

Preuve.
On doit prouver que $\mathbb{P}_B$ est une probabilité sur l'espace probabilisable (Ω, T).

— $\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(B \cap \Omega)}{\mathbb{P}(B)} = 1$ car $B \cap \Omega = B$.

— Soit (A_n) une suite d'événements incompatibles 2 à 2. Évaluons $\mathbb{P}_B\left(\bigcup_{n \in \mathbb{N}} A_n\right)$.

On a déjà

$$\left(\bigcup_{n \in \mathbb{N}} A_n\right) \cap B = \bigcup_{n \in \mathbb{N}} A_n \cap B$$

De plus, les événements $(A_n \cap B)$ sont 2 à 2 incompatibles et donc

$$\begin{aligned} \mathbb{P}_B\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \frac{1}{\mathbb{P}(B)} \mathbb{P}\left(\left(\bigcup_{n \in \mathbb{N}} A_n\right) \cap B\right) \\ &= \frac{1}{\mathbb{P}(B)} \sum_{n=0}^{+\infty} \mathbb{P}(A_n \cap B) \\ &= \sum_{n=0}^{+\infty} \mathbb{P}_B(A_n) \end{aligned}$$

où on a utilisé le fait que $\mathbb{P}$ est une probabilité et la liéarité de la somme de série convergente. ∎

II.2.2 Interprétation

Tout se passe comme si on prennait pour univers l'événement B. Ou encore que l'on suppose que B est réalisé pour évaluer la probabilité que d'autres événements se réalisent.

Il ne faut pas confondre cette notion avec l'intersection qui représente le fait que deux événements se réalisent en même temps.

Pour illustrer ceci, voici un exemple. On considère que l'on lance un dé équilibré à 6 faces.

— La probabilité d'obtenir un nombre pair et un 6 est de $\frac{1}{6}$.

— La probabilité d'obtenir un 6 sachant qu'on a obtenu un nombre pair est $\frac{1}{3}$

II.2.3 Proposition (Formule des probabilités composées)

1. Pour A, B des événements, si $\mathbb{P}(B) > 0$ alors $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. Rappelons en plus que $\mathbb{P}(A \cup B) = 1 - \mathbb{P}(\overline{A} \cap \overline{B})$
2. Si $A_1, \ldots, A_n$ sont des événements tels que $\mathbb{P}\left(\bigcap_{k=1}^{n-1} A_k\right) \neq 0$ alors

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \mathbb{P}(A_1) \times \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_3|A_1 \cap A_2) \times \cdots \times \mathbb{P}\left(A_n | \bigcap_{k=0}^{n-1} A_k\right)$$

II.2.4 Exemple

Un savant fou choisi dans la salle de classe des cobaye parmi les élèves. Aucune chance de s'en sortir. Quelle est la probabilité pour qu'il choisisse successivement un garçon, une fille puis un garçon ?

On cherche $\mathbb{P}(G_1 \cap F_2 \cap G_3) = \mathbb{P}(G_1)\mathbb{P}(F_2|G_1)\mathbb{P}(G_3|G_1 \cap F_2)$.

II.2.5 Proposition (Probabilité totales)

Il s'agit de traduire les propriétés des probabilités vis-à-vis de l'intersection en termes de probabilités conditionnelles.

1. Soit A un événement ni négligeable ni presque sûr ($\mathbb{P}(A) \in]0,1[$). Alors $A, \overline{A}$ forment un système complet d'événements et pour tout événement B,

$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\overline{A})\mathbb{P}(\overline{A}).$$

2. Pour $(A_n)_{n \in \mathbb{N}}$ un système complet d'événements (y compris fini) et B un événement

$$\mathbb{P}(B) = \sum_{n=0}^{+\infty} \mathbb{P}(B|A_n)\mathbb{P}(A_n)$$

où l'on convient que $\mathbb{P}(B|A_n)\mathbb{P}(A_n) = 0$ si $\mathbb{P}(A_n) = 0$.

Preuve.
Immédiat. ∎

II.2.6 Exemple
Un site internet a une audience séparée en deux types : les respectueux qui représentent 90% des inscrits et les trolls 10%. Les premiers ont un probabilité de 0.1 de participer à une discussion houleuse sur une journée, les second 0.7.

Un nouvel utilisateur s'inscrit. Avec quelle probabilité participe-t-il à une discussion houleuse dès le premier jour ? Dans les deux premiers jours ?

Notons T l'événement "le nouvel arrivant est un troll" et D l'événement "il participe à une discussion houleuse".

$\mathbb{P}(D) = \mathbb{P}(T) \times \mathbb{P}(D|T) + \mathbb{P}(\overline{T}) \times \mathbb{P}(D|\overline{T}) = \frac{1}{10} \times \frac{7}{10} + \frac{9}{10} \times \frac{1}{10} = \frac{4}{25}$.

Ainsi $\mathbb{P}(\overline{D}) = \frac{21}{25}$ et la réponse à la deuxième question est

$$1 - \left(\frac{21}{25}\right)^2 = 1 - \frac{441}{625} = \frac{184}{665} \approx 0.3$$

On fait ici l'hypothèse que les journées sont indépendantes.

Pour calculer la probabilité d'une réunion (discussion houleuse le jour 1 OU le jour 2, on calcule plutôt la probabilité de l'événement contraire qui est une intersection.)

II.2.7 Proposition (Formule de Bayes)
Soient A, B deux événements non négligeables ($\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$). Alors

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

En pratique, on calcule souvent le dénominateur par la formule des probabilités totales.

II.2.8 Exemple
Malfaçon ou triche organisée ? Toujours est-il que sur les 100 dés à 6 faces produits aujourd'hui 25 on une probabilité de 1/2 de tomber sur 6...

On choisi un de ces dés et on le lance. Il tombe sur 6. Avec quelle probabilité est-il pipé ?

Notons S l'événement "le dé tombe sur 6" et T l'événement "le dé choisi est pipé". On cherche $\mathbb{P}(T|S)$.

On connaît $\mathbb{P}(S|T) = \frac{1}{2}$, $\mathbb{P}(T) = \frac{1}{4}$. Il nous manque

$$\mathbb{P}(S) = \mathbb{P}(T)\mathbb{P}(S|T) + \mathbb{P}(\overline{T})\mathbb{P}(S|\overline{T}) = \frac{1}{4}\frac{1}{1} + \frac{3}{4}\frac{1}{6} = \frac{1}{4}.$$

Ainsi $\mathbb{P}(T|S) = \frac{\mathbb{P}(T)\mathbb{P}(S|T)}{\mathbb{P}(S)} = \frac{1}{2}$.

II.2.9 Cas d'application

Généralement l'énoncé donne $\mathbb{P}(A|B)$ et $\mathbb{P}(B)$. Il faut calculer $\mathbb{P}(A)$ par la formule des probabilités totales.

II.3 Evénements indépendants

II.3.1 Définition
Soient A, B deux événements

On dit que A et B sont indépendants ssi $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

II.3.2 Lien avec les probabilités conditionnelles

Si on suppose $P(B) > 0$, la condition A et B sont indépendants devient $P(A) = \mathbb{P}(A|B)$.

La réalisation de A ne dépend pas de celle (préalable) de B.

II.3.3 Définition
Soient $A_1, \dots, A_n$ des événements. On dit qu'ils sont indépendants ssi

$$\forall I \subset [\![1, n]\!] \mathbb{P}(\bigcap_{i\in I} A_i) = \prod_{i\in I} \mathbb{P}(A_i)$$

II.3.4 Attention

Trois événements indépendants 2 à 2 ne sont pas forcément indépendants.

II.3.5 Proposition
Si (A, B) sont indépendants, il en est de même de $(A^c, B), (A, B^c), (A^c, B^c)$. On peut généraliser ce résultat à $n > 2$ événements indépendants (et mettre des complémentaires ou non où bon nous semble).

II.3.6 En pratique

L'énoncé supposera très souvent que certains événements sont indépendants. On pourra alors très facile calculer des probabilités d'intersection et de réunion (1 - $\mathbb{P}$) en passant au complémentaire, grâce à la remarque précédente.

III Variables aléatoires

III.1 Lois

III.1.1 Définition

1. Une variable aléatoire **discrète** est une fonction $X : \Omega \to \mathbb{R}$ où $X(\Omega)$ (l'ensemble des valeurs de X) est dénombrable ou fini.

   On impose en plus que l'ensemble $X^{-1}(\{x\})$ est un événement (un élément de notre tribu) pour tout $x \in X(\Omega)$.
2. Si H est un ensemble de valeurs de X (on a donc $H \subset X(\Omega)$), on note $(X \in H)$ l'événement "X prend l'une des valeurs dans H", c'est à dire l'ensemble $X^{-1}(H)$ (image réciproque par la fonction X de l'ensemble H).
3. Si x est l'une des valeurs que peut prendre X (ie. $x \in X(\Omega)$), on note $(X = x)$ l'événement $X^{-1}(\{x\})$, c'est à dire l'événement "X prend la valeur x"

Le première chose à préciser sur une variable aléatoire (en théorie comme en pratique) est l'ensemble de ses valeurs

III.1.2 Théorème
Soit X une variable aléatoire discrète sur Ω. Notons $\{x_n |\ n \in \mathbb{N}\} = X(\Omega)$ l'ensemble de ses valeurs. Alors $((X = x_n))_{n\in\mathbb{N}}$ est un système complet d'événements.

Preuve.

1. Si $i \neq j$, et $\omega \in \Omega$ $\omega \in (X = x_i) \cap (X = x_j)$ signifie que $X(\omega) = x_i$ et $X(\omega) = x_j$ ce qui est absurde.
2. Si $\omega \in \Omega$, notons $x_n = X(\omega)$. Alors $\omega \in (X = x_n)$.

Ces deux points prouvent que les événements $(X = x_n)$ sont deux à deux incompatibles et que leur réunion est Ω. ∎

III.1.3 Exemple
Revenons à notre pile ou face. Cette fois la pièce est truquée et tombe sur pile avec un probabilité $p \in]0, 1[$. On note X le numéro du lancé ou le jeu se fini. Calculer pour $n \in\ N\backslash\{0\}$, $\mathbb{P}(X = n)$ ainsi que leur somme.

III.1.4 Définition
Soit X une variable aléatoire discrète. La loi de X est l'application

$$\mathbb{P}_X : \left\{ \begin{array}{rcl} X(\Omega) & \to & [0,1] \\ x & \mapsto & \mathbb{P}(X = x) \end{array} \right.$$

Avec les notations du théorème précédent, il s'agit de donner, $\mathbb{P}(X = x_n)$, pour tout $n \in \mathbb{N}$.

III.1.5 Somme

D'après le théorème précédent, la loi de X vérifie $\sum_{n=0}^{+\infty} \mathbb{P}(X = x_n) = 1$.

Réciproquement, on admet que si $(p_n)_{n\in\mathbb{N}}$ est une suite de réels positifs telle que $\sum_{n=0}^{+\infty} p_n = 1$, alors on peut trouver une probabilité $\mathbb{P}$ sur Ω et une variable aléatoire X telles que $\mathbb{P}(X = x_n) = p_n$.

III.2 Loi usuelles

III.2.1 Répétitions

Considérons une répétition illimité de la même expérience aléatoire (par exemple on lance deux dés), et on s'intéresse à un résultat précis que l'on nomme succès (on considère donc une répétition d'expérience "de Bernoulli") qui apparaît avec une probabilité $p \in]0, 1[$. On suppose les expériences mutuellement indépendantes.

On note X le rang du premier succès. Donner la loi de X.

III.2.2 Définition
On dit qu'une variable aléatoire X suit la loi géométrique de paramètre $p \in]0, 1[$ (on note $X \hookrightarrow \mathcal{G}(p)$) ssi

$$\forall k \in \mathbb{N}\backslash\{0\}\ \mathbb{P}(X = k) = p(1 - p)^{k-1}.$$

— L'ensemble des valeurs de X est $X(\Omega) = \mathbb{N}\backslash\{0\}$.

— On peut interpréter X comme donnant le rang d'apparition du premier succès lors de répétitions INDÉPENDANTES d'une expérience de probabilité de succès p.

III.2.3 Proposition
Soit X une variable aléatoire, $X \hookrightarrow \mathcal{G}(p)$ pour un $p \in]0,1[$. Soient $n, k \in \mathbb{N}\backslash\{0\}$. $\mathbb{P}(X > n+k | X > n) = \mathbb{P}(X > k)$.

On dit que la loi géométrique est sans mémoire.

Preuve.
On a au niveau des événements, $(X > n+k) \cap (X > n) = (X > n+k)$. Ainsi $\mathbb{P}(X > n+k | X > n) = \frac{\mathbb{P}(X>n+k)}{\mathbb{P}(X>n)}$.

Or $\mathbb{P}(X > n+k) = \sum\limits_{i=0}^{+\infty} \mathbb{P}(X = n+k+1+i) = \sum\limits_{i=0}^{+\infty} p(1-p)^{n+k+i} = p(1-p)^{n+k} \frac{1}{1-(1-p)} = (1-p)^{n+k}$.

De même $\mathbb{P}(X > n) = (1-p)^n$ et $\mathbb{P}(X > k) = (1-p)^k$. Ce qui conclut. ∎

Explication Le fait de savoir que les n premières expériences ont échoués ne présage en rien du nombre d'échec ou de succès à venir.

III.2.4 Exemple
Soit $\lambda \in \mathbb{R}^*$

Trouver $\alpha \in \mathbb{R}$ tel que la série $\sum\limits_{n\in\mathbb{N}} \frac{\alpha\lambda^n}{n!}$ converge et que sa somme soit 1.

C'est un simple calcul de somme d'une série exponentielle, on trouve $\alpha = e^{-\lambda}$.

III.2.5 Définition
Soit $\lambda \in]0, +\infty[$. On dit qu'une variable aléatoire X à valeurs dans $\mathbb{N}$ suit la loi de Poisson de paramètre λ (noté $X \hookrightarrow \mathcal{P}(\lambda)$) ssi

$$\forall k \in \mathbb{N}\ \mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

III.2.6 Cas d'utilisation

On peut utiliser cette loi pour approximer une loi binomiale de paramètre faible (on le verra plus loin), ou pour modéliser une expérience ou les valeurs de X ont de fortes chances d'être faible. Nous verrons l'interprétation de λ plus loin dans le chapitre.

III.3 Loi conjointe, indépendance

III.3.1 Définition
Soient X, Y deux variables aléatoires discrètes sur Ω. On note $\{x_n |\ n \in \mathbb{N}\}$ et $\{y_m |\ m \in \mathbb{N}\}$ les valeurs possibles de X et Y respectivement.

1. (X, Y) est une variable aléatoire $\omega \mapsto (X(\omega), Y(\omega))$.
2. La **loi conjointe** du couple (X, Y) est la loi décrite par la donnée de $\mathbb{P}(X = x_n, Y = y_m)$ pour toutes les valeurs de n et m.
   Il s'agit de la loi de la variable discrète $Z = (X, Y)$.
3. Les lois marginales de la loi conjointe de (X, Y) sont les lois de X et Y.
4. Pour $n_0 \in \mathbb{N}$ fixé tel que $\mathbb{P}(X = x_{n_0}) \neq 0$, la loi conditionnelle de Y sachant $(X = x_{n_0})$ est la loi donnée par $\mathbb{P}(Y = y_m | X = x_{n_0})$

Preuve.
Il faut a priori justifier que $Z = (X, Y)$ est une variable aléatoire.

Soit $z \in Z(\Omega)$, $z = (x, y)$ où $x \in X(\Omega)$, $y \in Y(\Omega)$. Nous devons montrer que $A = Z^{-1}(\{z\})$ est un événement.

Or $\omega \in A \iff Z(\omega) = z \iff Z(\omega) = (x, y) \iff X(\omega) = x$ et $Y(\omega) = y$. Ainsi $A = X^{-1}(\{x\}) \cap Y^{-1}(\{y\})$ est une intersection d'événements donc est un événements (car X, Y sont supposés être des variables aléatoires). ∎

III.3.2 Calcul de lois

Notons pour $(n, m) \in \mathbb{N}^2$, $p_{n,m} = \mathbb{P}(X = x_n, Y = y_m)$. On suppose donc la loi conjointe connue.

1. Pour un $n \in \mathbb{N}$ fixé, $\sum\limits_{m=0}^{+\infty} p_{n,m} = \mathbb{P}(X = x_n)$ car $((Y = y_m))_{m\in\mathbb{N}}$ forment un système complet d'événements. (On retrouve la première loi marginale par somme).
   De même, pour un $m \in \mathbb{N}$ fixé, $\sum\limits_{n=0}^{+\infty} p_{n,m} = \mathbb{P}(Y = y_m)$.
2. Alors $\sum\limits_{n=0}^{+\infty}\sum\limits_{m=0}^{+\infty} p_{n,m} = \sum\limits_{m=0}^{+\infty}\sum\limits_{n=0}^{+\infty} p_{n,m} = 1$.
3. Pour n, m fixés, $\mathbb{P}(Y = y_m | X = x_n) = \frac{\mathbb{P}(Y=y_m, X=x_n)}{\mathbb{P}(X=x_n)} = \frac{p_{n,m}}{\sum\limits_{k=0}^{+\infty} p_{n,k}}$.

III.3.3 Exemple
On lance une pièce qui tombe sur pile avec une probabilité $p \in]0,1[$ (et donc face avec une probabilité $q = 1 - p \in]0,1[$). On note X le rang d'apparition du premier pile et Y le rang du second.

Donner la loi conjointe. Soient $n, m \in \mathbb{N}\backslash\{0\}$.

— Si $n \geqslant m$ alors $\mathbb{P}(X = n, Y = m) = 0$.

— Si $1 \leqslant n < m$, $\mathbb{P}(X = n, Y = m) = \mathbb{P}(Y = m | X = n)\mathbb{P}(X = n) = p(1-p)^{m-n-1} \times p(1-p)^{n-1} = p^2(1-p)^{m-2}$.

Vérifions que la somme vaut 1.

$$\sum_{m=1}^{+\infty}\sum_{n=1}^{+\infty}\mathbb{P}(X=n,Y=m)=\sum_{m=2}^{+\infty}\sum_{n=1}^{m-1}p(1-p)^{m-n-1}\times p(1-p)^{n-1}$$
$$=\sum_{m=2}^{+\infty}(m-1)p^2(1-p)^{m-2}=p^2\sum_{m=1}^{+\infty}m(1-p)^{m-1}$$

On re connaît la dérivée de $f : x\mapsto \frac{1}{1-x}=\sum_{n=0}^{+\infty}x^n$ évalué en $x=1-p$.

Ainsi $\sum_{m=2}^{+\infty}(m-1)(1-p)^{m-2}=\frac{1}{(1-(1-p))^2}=\frac{1}{p^2}$ et tout est bien qui fini bien.

III.3.4 Retrouver une loi marginale

Ceci se fait toujours par calcul de somme. Reprenons l'exemple précédent. on connaît la loi de X qui est géométrique de paramètre p. Donnons la loi de Y.

— Y est à valeurs dans $\mathbb{N}\backslash\{0,1\}=\{2,3,4,\dots\}$.

— Soit $m\in\mathbb{N}$, $m\geqslant 2$. On sait que $((X=n))_{nin\mathbb{N}\backslash\{0\}}$ est un système complet d'événements (on a considéré toutes les valeurs envisageables pour X). Ainsi

$$\mathbb{P}(Y=m)=\sum_{n=1}^{+\infty}\mathbb{P}((Y=m)\text{ et }(X=n))=\sum_{n=1}^{+\infty}\mathbb{P}(Y=m,X=n)$$

Or, pour les n tels que $n\geqslant m$, $\mathbb{P}(Y=m,X=n)=0$ et donc

$$\mathbb{P}(Y=m)=\sum_{n=1}^{m-1}\mathbb{P}(Y=m,X=n)=\sum_{n=1}^{m-1}p^2(1-p)^{m-2}=(m-1)p^2(1-p)^{m-2}$$

Autre méthode : on peut arriver à ce résultat par dénombrement.

En effet pour que le 2ème succès arrive au rang m, il faut que la succession des m premières expériences soit de la forme $(?,\dots,?,S)$ où S représente un succès et que parmi les $m-1$ premières expériences, il y ait exactement un succès.

Il y a exactement $m-1$ possibilité pour écrire les premières expériences sous cette forme : le 1er succès est au rang 1, ou (exclusif) le premier succès est au rang 2 ou ...

Par indépendance des expériences élémentaires, chaque manière d'écrire les m expérience se produit avec une probabilité $p^2(1-p)^{m-2}$ (2 succès et $m-2$ échecs). De plus, on a souligné le fait que la réunion est disjointe (les cas sont disjoints) et on obtient donc la probabilité totale par somme.

III.3.5 Définition
Soient X,Y deux variables aléatoires discrètes sur Ω. on dit qu'elles sont indépendantes ssi

$$\forall x,y\in X(\Omega)\times Y(\Omega)\mathbb{P}(X=x,Y=y)=\mathbb{P}(X=x)\mathbb{P}(Y=y)$$

ie ssi les événements $(X=x)$ et $(Y=y)$ sont deux à deux indépendants pour toutes les valeurs possibles de x et y.

Lorsque X et Y sont indépendantes, on note $X\perp\!\!\!\perp Y$

III.3.6 Exemple
Les variables aléatoires de l'exemple III.3.3 ne sont pas indépendantes. On a par exemple $\mathbb{P}(X=2,Y=2)=0\neq\mathbb{P}(X=2)\mathbb{P}(Y=2)$

III.3.7 Somme de deux lois de Poisson

A savoir refaire

Soient X,Y deux variables $X\hookrightarrow\mathcal{P}(\lambda)$ et $Y\hookrightarrow\mathcal{P}(\mu)$, indépendantes. On note $Z=X+Y$. Calculer la loi de Z.

— Premièrement, Z est à valeurs dans $\mathbb{N}$, comme X et Y.

— Pour $n\in\mathbb{N}$,

$$\mathbb{P}(Z=n)=\sum_{k=0}^{n}\mathbb{P}(X=k,Y=n-k)=\sum_{k=0}^{n}e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}=e^{-\lambda-\mu}\frac{1}{n!}(\lambda+\mu)^n.$$

Ainsi $Z\hookrightarrow\mathcal{P}(\lambda+\mu)$.

III.3.8 Proposition
Soient X,Y deux variables indépendantes et $G\subset X(\Omega),H\subset Y(\Omega)$. Alors

$$\mathbb{P}((X,Y)\in G\times H)=\mathbb{P}(X\in G)\times\mathbb{P}(Y\in H).$$

Extension des notions de premières années On admet que les résultats suivants sont encore vrais pour des variables aléatoires discrètes.

III.3.9 Proposition
Si X, Y sont des variables aléatoires discrètes indépendantes et si on peut calculer $f(X)$ et $g(Y)$ pour des fonctions f et g alors $f(X)$ et $g(Y)$ sont des variables aléatoires et sont indépendantes.

On peut résumer par : $X \perp\!\!\!\perp Y \Rightarrow f(X) \perp\!\!\!\perp f(Y)$.

Preuve.

— Montrons d'abord que $H = f(X) = f \circ X$ est une variable aléatoire. Soit $y \in H(\Omega)$. Montrons que $H^{-1}(\{y\})$ est un événement.

Posons $A = f^{-1}(\{y\}) \subset X(\Omega)$ (car la composition est licite).

Alors $H^{-1}(\{y\}) = X^{-1}(A) = (X \in A)$ qui est bien un événement (par réunion). ■

III.3.10 Définition
Soit $(X_n)_{n\in\mathbb{N}}$ des variables aléatoires discrètes sur Ω. On dit qu'elles sont indépendantes ssi pour $k \in \mathbb{N}\backslash\{0\}$ et $i_1, \ldots, i_k \in \mathbb{N}$,

$$\forall x_1 \in X_{i_1}(\Omega), \ldots, x_k \in X_{i_k}(\Omega)\ \mathbb{P}(X_{i_1} = x_1 \text{ et } \ldots \text{ et } X_{i_k} = x_k) = \prod_{j=1}^{k} \mathbb{P}(X_{i_j} = x_i).$$

Autrement dit, on peut calculer toute probabilité d'intersection finie par produit.

III.3.11 Avec des événements

La proposition III.3.8 s'étend au cas de n variables indépendantes.

III.3.12 A retenir

Comme pour les événements, on supposera souvent dans l'énoncé que des variables aléatoires sont indépendantes. On pourra alors calculer des probabilités d'intersection (et) par produit.

IV Moments d'une variable aléatoire

Dans cette partie, toutes les variables aléatoires sont supposées à valeurs réelles.

IV.1 Espérance, variance

Explication La notion d'espérance s'étend de manière naturelle aux variables aléatoires discrète. Par contre l'existence de l'espérance n'est pas garantie a priori, vu qu'il s'agit de la convergence d'une série numérique.

IV.1.1 Définition
Soit X une variable aléatoire discrète à valeurs dans $X(\Omega) = \{x_n |\ n \in \mathbb{N}\} \subset \mathbb{R}$. On dit que X est d'espérance finie ssi $\sum x_n \mathbb{P}(X = x_n)$ converge **absolument**.

Dans ce cas, on appelle **espérance de** X et on note $E(X)$ le réel $\sum_{n=0}^{+\infty} x_n \mathbb{P}(X = X_n)$.

Explication Comme pour le théorème sur le produit de Cauchy il nous faut ici supposer la convergence absolue. La raison est hors de notre programme : la valeur de la somme ne dépend pas de l'ordre dans lequel on calcule celle-ci. En particulier ici, on peut numéroter les éléments de $X(\Omega)$ comme bon nous semble sans changer l'espérance (encore heureux !). En pratique, nos variables aléatoires sont très souvent à valeurs dans $\mathbb{N}$ et l'ordre de sommation est naturel (mais pas imposé).

IV.1.2 Exemple
On peut définir une loi à valeurs dans $\mathbb{N}\backslash\{0\}$ par $\forall n \in \mathbb{N}\backslash\{0\}\ \mathbb{P}(X = n) = \frac{6}{\pi^2 n^2}$ car la série converge et est de limite 1.

Dans ce cas X n'est pas d'espérance finie.

IV.1.3 Proposition

1. Si $X \hookrightarrow \mathcal{G}(p)$ pour $p \in]0, 1[$ alors X est d'espérance finie et $E(X) = \frac{1}{p}$.
2. Si $X \hookrightarrow \mathcal{P}(\lambda)$ pour $\lambda > 0$ alors X est d'espérance finie et $E(X) = \lambda$.

Preuve.
Il s'agit ici d'exemples de calculs d'espérance.

1. La série numérique $\sum_{n\geqslant 1} np(1-p)^{n-1}$ converge par comparaison de séries à termes positifs (car $n(1-p)^{n-1} = o_{+\infty}(\frac{1}{n^2})$) donc converge absolument. De plus, $\sum_{n=1}^{+\infty} np(1-p)^{n-1} = p\frac{1}{(1-(1-p)^2)} = \frac{1}{p}$ en dérivant terme à terme la série géométrique et en évaluant en $1 - p \in] - 1, 1[$.
   Interprétation : Si la probabilité de succès élémentaire est p, on s'attend à faire en moyenne $\frac{1}{p}$ essais avant d'obtenir un succès.

2. La série $\sum_{n\geqslant 0} ne^{-\lambda}\frac{\lambda^n}{n!} = \lambda e^{-\lambda}\sum_{n\geqslant 1}\frac{\lambda^{n-1}}{(n-1)}$ converge absolument par produit par une constante d'une série exponentielle et $E(X)=\lambda$.
   Ainsi le paramètre λ d'une loi de Poisson s'interprète comme son espérance (ou intuitivement, valeur moyenne). ∎

IV.1.4 Proposition (Propriétés de l'espérance)
Soient X, Y deux variables aléatoires discrètes réelles sur Ω, **d'espérances finies**.

1. Linéarité : soient $\lambda,\mu\in\mathbb{R}$. $\lambda X+\mu Y$ est d'espérance finie et $E(\lambda X+\mu Y) = \lambda E(X)+\mu E(Y)$.
2. Positivité : si $X\geqslant 0$ alors $E(X)\geqslant 0$.
3. Croissance : si $\forall\omega\in\Omega\ X(\omega)\leqslant Y(\omega)$ (que l'on note $X\leqslant Y$) alors $E(X)\leqslant E(Y)$.
4. Si X et Y sont **indépendantes** alors XY est d'espérance finie et $E(XY) = E(X)E(Y)$.

Preuve.

1. Admis. on peut appliquer le théorème de transfert à la VA (X,Y) et la fonction $(x,y)\mapsto x+y$ et utiliser le cours sur les séries absolument convergentes.
2. Evident : une série à termes positifs convergente a une somme positive.
3. Conséquence directe (et classique !) des deux propriétés précédentes.
4. Admis. ∎

IV.1.5 Exemple

1. Ceci est tout à fait en accord avec notre calcul de la loi d'une somme de deux lois de Poisson indépendantes.
2. Si X est d'espérance finie, alors $Y=X-E(X)$ est une variable **centrée** ie d'espérance nulle.

IV.1.6 Théorème (Théorème de transfert)
Soit X une variable aléatoire discrète et f une fonction définie sur $X(\Omega)=\{x_n|\ n\in\mathbb{N}\}$ à valeurs réelles. $f(X)$ est d'espérance finie ssi $\sum f(x_n)\mathbb{P}(X=x_n)$ est absolument convergente.

Alors $E(f(X)) = \sum_{n=0}^{+\infty} f(x_n)\mathbb{P}(X=x_n)$. Ainsi l'espérance de $f(X)$ est déterminée par la loi de X.

Preuve.
Admis. ∎

IV.2 Variance

IV.2.1 Définition-Proposition
Soit X une variable aléatoire discrète. Si X^2 est d'espérance finie alors X aussi. Dans ce cas :

1. on appelle **variance** de X le nombre réel positif $V(X) = E\left((X-E(X))^2\right) = E(X^2)-E(X)^2$.
2. on appelle **écart-type** de X le nombre réel positif $\sigma(X)=\sqrt{V(X)}$.

Si $\sigma(X)=1$, on dit que X est réduite.

Preuve.
On suppose X^2 d'espérance finie ie $\sum_{n\in\mathbb{N}} x_n^2\mathbb{P}(X=x_n)$ converge absolument. Montrons que $\sum_{n\in\mathbb{N}}|x_n|\mathbb{P}(X=x_n)$ converge.

Remarquons que pour $n\in\mathbb{N}$, $|x_n|\mathbb{P}(X=x_n)\leqslant\begin{cases}|x_n|^2\mathbb{P}(X=x_n) \text{ si } |x_n|\geqslant 1\\ \mathbb{P}(X=x_n)\textbf{sinon}\end{cases}$.

Dans tous les cas $|x_n|\mathbb{P}(X=x_n)\leqslant(|x_n|^2+1)\mathbb{P}(X=x_n)$ qui est la somme de deux TG de séries convergentes et positives. Par comparaison de séries à termes positifs, X est d'espérance finie. ∎

IV.2.2 Proposition

1. Si $X\hookrightarrow\mathcal{G}(p)$ pour $p\in]0,1[$ alors X est de variance finie et $V(X)=\frac{1}{p^2}-\frac{1}{p}$.
2. Si $X\hookrightarrow\mathcal{P}(\lambda)$ pour $\lambda>0$ alors X est de variance finie et $V(X)=\lambda$.

Preuve.
Il s'agit ici d'exemples de calculs de variance.

1. Soit $X \hookrightarrow \mathcal{G}(p)$ pour un $p \in]0,1[$.
   La série $\sum\limits_{n\geqslant 1} n^2 p(1-p)^{n-1}$ converge par comparaison de séries à termes positifs car $n^2(1-p)^{n-1} = o_{+\infty} n^2$. Donc X possède une variance
   Posons $f : x \mapsto \frac{1}{1-x} = \sum\limits_{n=0}^{+\infty} x^n$ (sur $]-1,1[$) et $x \in]-1,1[$.
   Alors $f'(x) = \sum\limits_{n=1}^{+\infty} n x^{n-1} = \frac{1}{(1-x)^2}$ et $f''(x) = \sum\limits_{n=2}^{+\infty} n(n-1)x^{n-2} = \frac{2}{(1-x)^3}$.
   Ainsi $xf''(x) = \sum\limits_{n=1}^{+\infty} n(n-1)x^{n-1}$ (le premier terme est nul) et $\sum\limits_{n=1}^{+\infty} n^2 x^{n-1} = f'(x) + xf''(x) = \frac{1}{(1-x)^2} + \frac{2x}{(1-x)^3}$.
   On a donc $E(X^2) = p(\frac{1}{p^2} + \frac{2(1-p)}{p^3}) = 2\frac{1}{p^2} - \frac{1}{p}$.
   Finalement $V(X) = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}$
2. Soit $X \hookrightarrow \mathcal{P}(\lambda)$ pour un $\lambda > 0$.
   La série $\sum\limits_{n\geqslant 0} n^2 e^{-\lambda}\frac{\lambda^n}{n!}$ converge (d'Alembert, ou produit par une puissance dans une série entière, ce qui ne modifie pas le rayon de convergence).
   De plus, $\sum\limits_{n=0}^{+\infty} n^2 e^{-\lambda}\frac{\lambda^n}{n!} = \lambda e^{-\lambda} \sum\limits_{n\geqslant 0} (n+1)\frac{\lambda^n}{n!} = \lambda e^{\lambda}(\lambda e^{\lambda} + e^{\lambda})$.
   Ainsi $V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$. ∎

IV.2.3 Proposition
Soit X une variable aléatoire réelle de variance finie et $a, b \in \mathbb{R}$.

Alors $aX + b$ est de variance finie et $V(aX+b) = a^2 V(X)$.

Preuve.
Montrons que $aX+b$ est de variance finie c'est à dire que $(aX+b)^2$ est d'espérance finie.

Or $(aX+b)^2 = a^2X^2 + 2abX + ab$ est bien d'espérance finie par combinaison linéaire.

On utilise $V(aX+b) = E((aX+b)^2) - (E(aX+b))^2 = E(a^2X^2 + 2abX + b^2) - (aE(X)+b)^2 = a^2E(X^2) - a^2(E(X))^2 = a^2V(X)$. ∎

IV.3 Covariance

IV.3.1 Définition-Proposition
Soient X, Y deux variables aléatoires discrètes. Si X et Y admettent un moment d'ordre 2 (ie. admettent une variance finie) alors la variable aléatoire $(X - E(X))(Y - E(Y))$ est d'espérance finie.

Dans ce cas on appelle covariance de X et Y le réel

$$\mathrm{cov}(X,Y) = E((X - E(X))(Y - E(Y)))$$

.

Preuve.
Il s'agit de montrer que XY est d'espérance finie (les autres VA le sont facilement quand on développe).

Or $|XY| \leqslant X^2 + Y^2$. Notons $Z = (X,Y)$, $f : (x,y) \mapsto xy$ et $g : (x,y) \mapsto x^2 + y^2$. On sait que $g(Z)$ est d'espérance finie d'après le théorème de transfert. En notant $Z(\Omega) = \{z_n ; n \in \mathbb{N}\}$, la série $\sum g(z_n)\mathbb{P}(Z = z_n)$ converge et donc, par comparaison de série à termes positifs, $\sum f(z_n)\mathbb{P}(Z = z_n)$ converge absolument car $|f(z_n)| \leqslant g(z_n)$.

IV.3.2 Remarque
On a $\mathrm{cov}(X,X) = V(X)$.

IV.3.3 Proposition
Dans les conditions de la définition précédente :
1. $\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y)$.
2. **Si** X et Y sont indépendantes **alors** $\mathrm{cov}(X,Y) = 0$.
3. la covariance est bilinéaire et symétrique.

Preuve.
Simple utilisation de la linéarité de l'espérance, en plus de la propriété $E(a) = a$ quand a est une constante.

Le deuxième point est une conséquence directe d'un théorème du chapitre sur les probabilités.

La symétrie est évidente, la bilinéarité conséquence simple de la linéarité de l'espérance. Notons qu'une combinaison linéaire de VA de variances finies est encore de variance finie (c'est une conséquence de la partie proposition de la définition de la covariance). ∎

IV.3.4 Proposition
Soit X, Y deux variables aléatoires admettant une variance finie. Alors $X + Y$ est de variance finie et

$$V(X + Y) = V(X) + V(Y) + 2\operatorname{cov}(X, Y)$$

Preuve.
En effet, $V(X+Y) = E((X+Y)^2) - (E(X+Y))^2 = E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y^2)$. ∎

IV.3.5 Exemple
Ainsi pour des variables indépendantes, $V(X + Y) = V(X) + V(Y)$ et plus généralement la variance d'une somme de variables mutuellement indépendantes est la somme des variances.

Rappelons une application importante, posons $(X_i)_{i \in [\![1,n]\!]}$ des variables aléatoires mutuellement indépendante de même loi de Bernoulli de paramètre $p \in]0, 1[$.

Alors $S = \sum_{i=1}^{n} X_i$ suit une loi binomiale de paramètres n et p.

Or $V(X_1) = E(X_1^2) - E(X_1)^2 = p - p^2 = p(1 - p)$ et donc $V(S) = np(1 - p)$.

Finissons le rappel par $E(S) = nE(X_0) = np$ par linéarité de l'espérance.

IV.3.6 Théorème (Cauchy-Schwartz)
On a $|\operatorname{cov}(X, Y)| \leqslant \sqrt{V(X)V(Y)}$

Preuve.
Soit $t \in \mathbb{R}$. Alors $V(X + tY) = \cdots = t^2 V(Y) + 2t\operatorname{cov}(X, Y) + V(X)$ qui est de degré 2 (si $V(Y) \neq 0$) et positif. Le discriminant est donc négatif. ∎

IV.3.7 Définition
Soient X, Y deux variables aléatoires de variance finie et non nulle. Le coefficient de corrélation de X et Y est

$$\operatorname{cor}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\operatorname{cov}(X, Y)}{\sigma(X)\sigma(Y)} \in [-1, 1]$$

IV.3.8 Interprétation

L'étude du cas d'égalité dans l'inégalité de Cauchy-Schwartz précédente, permet de montrer que $\operatorname{cor}(X, Y) = \pm 1$ ssi $Y = aX + b$ pour a, b des réels. De plus, si X et Y sont indépendantes, $\operatorname{cor}(X, Y) = 0$. On peut “donc” interpréter ce coefficient comme une mesure du lien (autrement appelé corrélation) qui existe entre X et Y.

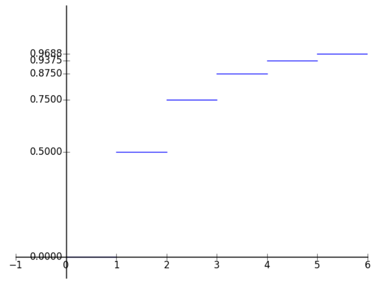
V Fonctions et probabilités

V.1 Fonction de répartition

V.1.1 Définition
Soit X une variable aléatoire discrète. On appelle fonction de répartition de X et on note F_X la fonction

$$F_X : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \mathbb{P}(X \leqslant x) \end{array} \right.$$

V.1.2 Exemple
Tracer un partie de la courbe représentative de F_X où X suit une loi géométrique de paramètre $\frac{1}{2}$.



V.1.3 Remarque

Imaginons que l'on connaisse la fonction de répartition F_X d'une VA X mais pas sa loi. Notons $X(\Omega) = \{x_n |\ n \in \mathbb{N}\}$ l'ensemble des valeurs de X où on a ordonnée les x_n, ie la suite (x_n) est croissante.

Alors $\mathbb{P}(X = x_0) = F_X(x_0)$ et pour tout $n \neq 0$, $\mathbb{P}(X = x_n) = F_X(x_n) - F_X(x_{n-1})$

V.1.4 Proposition
Avec les notations de la définition, on a :

1. F_X est croissante sur $\mathbb{R}$.
2. $\lim\limits_{x\to-\infty} F_X(x) = 0$.
3. $\lim\limits_{x\to+\infty} F_X(x) = 1$.

Preuve.
Il s'agit d'utiliser les propriétés de $\mathbb{P}$ suivantes : croissance, limite de la probabilité d'une suite décroissante d'événements et limite de la probabilité d'une suite croissante d'événements. ∎

V.1.5 Exemple
En pratique, il est parfois plus pratique de calculer des probabilités de la forme $\mathbb{P}(X \leqslant n)$, ce qui revient à calculer la fonction de répartition sans le dire.

Soient par exemple X, Y deux VA indépendantes de loi géométrique de paramètre $p \in]0,1[$ et $Z = \min(X,Y)$. Calculer $1 - F_Z$

Pour $n \in \mathbb{N}\backslash\{0\}$, $\mathbb{P}(Z > n) = \mathbb{P}((X > n) \cap (Y > n)) = \mathbb{P}(X > n)\mathbb{P}(Y > n) = ((1-p)^2)^n$.

On en déduit facilement que $Z \hookrightarrow \mathcal{G}(2p - p^2)$.

V.1.6 Proposition
Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. X est d'espérance finie ssi $\sum \mathbb{P}(X \geqslant n)$ converge et dans ce cas

$$E(X) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geqslant n)$$

Preuve.
Remarquons d'abord que pour $n \in \mathbb{N}$ on a $\mathbb{P}(X = n) = \mathbb{P}(X \geqslant n) - \mathbb{P}(X \geqslant n+1)$.

Soit maintenant $N \in \mathbb{N}$. On a

$$\begin{aligned}
\sum_{n=0}^{N} n\mathbb{P}(X=n) = \sum_{n=1}^{N} n\mathbb{P}(X=n) &= \sum_{n=1}^{N} n(\mathbb{P}(X \geqslant n) - \mathbb{P}(X \geqslant n+1)) \\
&= \sum_{n=1}^{N} n\mathbb{P}(X \geqslant n) - \sum_{n=1}^{N} n\mathbb{P}(X \geqslant n+1) \\
&= \sum_{n=1}^{N} n\mathbb{P}(X \geqslant n) - \sum_{n=2}^{N+1} (n-1)\mathbb{P}(X \geqslant n) \\
&= \sum_{n=1}^{N} \mathbb{P}(X \geqslant n) - N\mathbb{P}(X \geqslant N+1) \text{ en ajoutant un terme nul}
\end{aligned}$$

De plus, $\mathbb{P}(X \geqslant N+1) = \sum\limits_{n=N+1}^{+\infty} \mathbb{P}(X = n)$.

Ainsi

$$N\mathbb{P}(X \geqslant N+1) = \sum_{n=N+1} N\mathbb{P}(X=n)$$

— Supposons que X est d'espérance finie.
  On a alors

$$0 \leqslant N\mathbb{P}(X \geqslant N+1) \leqslant \sum_{n=N+1}^{+\infty} n\mathbb{P}(X=n)$$

  et le majorant trouvé est un reste de série convergente donc tend vers 0 lorsque $N \to +\infty$. Par encadrement puis par somme $\sum \mathbb{P}(X \geqslant n)$ converge et

$$E(X) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geqslant n)$$

— Réciproquement, supposons que $\sum \mathbb{P}(X \geqslant n)$ converge.
  Alors, par majoration de série à termes positifs, X est d'espérance finie et le calcul précédent conclut. ∎

V.2 Fonction génératrice

V.2.1 Une série entière

Soit X une VA à valeurs dans $\mathbb{N}$ (son ensemble de valeurs est un sous-ensemble de $\mathbb{N}$).

Considérons la série entière $\sum \mathbb{P}(X = n)t^n$. Comme cette série converge absolument pour $t = \pm 1$, son rayon de convergence vaut au moins 1.

V.2.2 Définition

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. La fonction génératrice (ou série génératrice) de X est la fonction

$$G_X : t \mapsto \sum_{n=0}^{+\infty} \mathbb{P}(X = n)t^n$$

G_X est définie au moins sur le segment $[-1, 1]$, $\mathcal{C}^\infty$ sur $]-1, 1[$ et $G_X(1) = 1$.

V.2.3 Remarque

Par unicité des coefficients d'une série entière de rayon de convergence non nul, la loi de X est entièrement déterminée par la fonction G_X.

V.2.4 Valeurs manquantes

On convient de poser $\mathbb{P}(X = n) = 0$ pour tous les n qui ne sont pas des valeurs de X. En particulier, pour une variable aléatoire sur un univers fini G_X est polynomiale !

V.2.5 Exemple

Calculons les fonctions génératrices pour les lois usuelles.

1. Soit $X \hookrightarrow \mathcal{B}(p)$ (Bernoulli, $p \in]0, 1[$).
   Alors $G_X : t \mapsto (1-p)t^0 + pt^1 = 1 - p + pt$.
2. Soit $X \hookrightarrow \mathcal{B}(n, p)$ (binomiale, $n \in \mathbb{N}\backslash\{0\}$, $p \in]0, 1[$).
   Alors $G_X : t \mapsto \sum\limits_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} t^k = (1 - p + pt)^n$.
3. Soit $X \hookrightarrow \mathcal{G}(p)$ (où $p \in]0, 1[$).
   La série considérée est $\sum\limits_{n \geqslant 1} p(1-p)^{n-1}t^n = pt \sum\limits_{n \geqslant 0} ((1-p)t)^n$.
   Cette série géométrique converge ssi $|(1-p)t| < 1$ et donc
   $$\forall t \in \left]-\frac{1}{1-p}, \frac{1}{1-p}\right[\ G_X(t) = \frac{pt}{1-(1-p)t}.$$
   Remarquons que le rayon de convergence de la série est $\frac{1}{1-p} > 1$.
4. Soit $X \hookrightarrow \mathcal{P}(\lambda)$ $(\lambda > 0)$.
   Pour $t \in]-1, 1[$, $G_X(t) = \sum\limits_{n=0}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} t^n = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}$.

V.2.6 Exercice

Déterminer la fonction génératrice d'une loi uniforme sur $[\![1, n]\!]$.

V.2.7 Interprétation en tant qu'espérance

Pour une variable aléatoire X à valeurs dans $\mathbb{N}$ (on a $x_0 = 0, x_1 = 1\ldots$ quitte à considérer trop de valeurs pour X qui donneront des probabilités nulles) et $t \in [-1, 1]$ on a

$$G_X(t) = E(t^X)$$

V.2.8 Théorème

Si X et Y sont deux variables aléatoires à valeurs dans $\mathbb{N}$ et **indépendantes**, notons R_X et R_Y les rayons de convergence de G_X et G_Y respectivement. Posons également $r = \min(R_X, R_Y)$

Alors G_{X+Y} est de rayon $R \geqslant r$ et

$$\forall t \in]-r, r[\ G_{X+Y}(t) = G_X(t)G_Y(t)$$

Preuve.
La série produit (de Cauchy) $G_X G_Y$ est de rayon $R \geqslant \min(R_X, R_Y)$ et pour $t \in]-r, r[$

$$G_X(t)G_Y(t) = \sum_{n=0}^{+\infty} c_n t^n$$

où $c_n = \sum\limits_{k=0}^{n} \mathbb{P}(X = k)\mathbb{P}(Y = n-k) = \sum\limits_{k=0}^{n} \mathbb{P}((X = k) \cap (Y = n-k)) = \mathbb{P}(X + Y = n)$.

Deuxième méthode, qui s'applique si on connaît les propriétés de l'espérance.

Pour $t \in]0, r[$, on pose $f_t : x \mapsto t^x$. Alors $f(X)$ et $f(Y)$ sont indépendantes et donc $E(t^X)E(t^Y) = E(t^X t^Y) = E(t^{X+Y})$. Ainsi $t \leqslant R$ et on a bien $G_{X+Y}(t) = G_X(t)G_Y(t)$. ∎

V.2.9 Exemple

On peut utiliser ce théorème pour calculer la loi d'une somme de variables indépendantes.

1. Soient $\lambda, \mu > 0$ et $X \hookrightarrow \mathcal{P}(\lambda), YX \hookrightarrow \mathcal{P}(\mu)$.
   Alors, pour tout $t \in \mathbb{R}$, $G_X(t)G_X(Y) = e^{\lambda(t-1)}e^{\mu(t-1)} = e^{(\lambda+\mu)(t-1)} = G_{X+Y}(t)$.
   Ainsi $X + Y \hookrightarrow \mathcal{P}(\lambda + \mu)$ (car la fonction génératrice détermine la loi).
2. Lançons deux dés équilibrés à 6 faces et notons X, Y les résultats obtenus pour le premier et le second dé respectivement.
   Donner la loi de $X + Y$ (la somme des deux dés).

Ici les lois prennent un nombre fini de valeurs et donc les fonctions génératrices sont polynomiales.

Pour $t \in \mathbb{R}$, $G_X(t) = G_Y(t) = \frac{1}{6} t \sum_{k=0}^{5} t^k$. De plus X et Y sont indépendantes.

Ainsi $G_{X+Y}(t) = \frac{t^2}{36} \left(\sum_{k=0}^{5} t^k \right)^2 = \frac{t^2}{36}(1+2t+3t^2+4t^3+5t^4+6t^5+5t^6+4t^7+3t^8+2t^9+t^{10})$. On obtient

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(X=k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

V.2.10 Théorème
Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ et G_X sa fonction génératrice.

1. X est d'espérance finie ssi G_X est dérivable en 1 et alors $E(X) = G'_X(1)$.
2. X est de variance finie ssi G_X est deux fois dérivable en 1 et alors

$$V(X) = G''_X(1) + G'_X(1) - \left(G'_X(1)\right)^2.$$

V.2.11 Retrouver les formules

Tout d'abord, on a $E(X) = \sum_{n=1}^{+\infty} n\mathbb{P}(X=n)$ et $E(X^2) = \sum_{n=1}^{+\infty} n^2\mathbb{P}(X=n)$ (théorème de transfert).

De plus, on supposant la dérivabilité terme à terme, $G'_X(t) = \sum_{n=1}^{+\infty} n\mathbb{P}(X=n)t^{n-1}$ donc on a bien $G'_X(1) = E(X)$.

De plus, $G''_X(t) = \sum_{n=2}^{+\infty} n(n-1)\mathbb{P}(X=n)t^{n-1}$ donc $G''_X(1) = E(X^2) - E(X)$.

V.2.12 Exemple
Retrouvons l'espérance et la variance des lois géométriques et de Poisson.

1. Soit $X \hookrightarrow \mathcal{G}(p)$ (où $p \in]0,1[$).
   G_X des DSE sur $]-\frac{1}{1-p}, \frac{1}{1-p}[$ donc est deux fois dérivable en 1.
   De plus, $G_X(t) = \frac{pt}{1-(1-p)t}$ donc $G'_X(t) = \frac{p(1-(1-p)t)+pt\times(1-p)}{(1-(1-p)t)^2} = \frac{p}{(1-(1-p)t)^2}$ donc $E(X) = \frac{p}{p^2} = \frac{1}{p}$.
   De même $G''(t) = p \times (-2) \times (-(1-p)) \times (1-(1-p)t)^{-3}$ donc $G''(1) = \frac{2p(1-p)}{p^3} = \frac{2(1-p)}{p^2} = E(X^2) - E(X)$.
   Ainsi $V(X) = E(X^2) - (E(X))^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$.
2. Soit $X \hookrightarrow \mathcal{P}(\lambda)$ ($\lambda > 0$).
   Cette fois G_X est DSE sur $\mathbb{R}$ donc dérivable deux fois en 1.
   De plus $G'_X(t) = \lambda e^{\lambda(t-1)}$ et $G''_X(1) = \lambda^2 e^{\lambda(t-1)}$.
   Ainsi $E(X) = G'_X(1) = \lambda$ et $V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

VI Etude asymptotique

VI.1 Interprétation de la loi de Poisson

VI.1.1 Proposition
Soit $\lambda > 0$.

On considère une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires telles que $X_n \hookrightarrow \mathcal{B}(n, p_n)$ où $p_n \underset{+\infty}{\sim} \frac{\lambda}{n}$.

Pour $k \in \mathbb{N}$ fixé, on a $\lim_{n\to+\infty} \mathbb{P}(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

Preuve.
On cherche à estimer la limite de $\binom{n}{k} p_n^k (1-p_n)^{n-k}$.

Or $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{1}{k!}\prod_{i=0}^{k-1}(n-i) \underset{n\to+\infty}{\sim} \frac{1}{k!}n^k$. (par produit d'un nombre fixé d'équivalents)

De plus, $p_n^k \underset{+\infty}{\sim} \frac{\lambda^k}{n^k}$ (encore une fois, k est fixé).

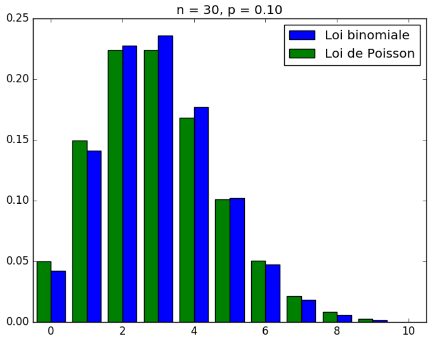
De plus, $(1-p_n)^{n-k} \underset{+\infty}{\sim} (1-p_n)^n$ car $(1-p_n)^{-k} \underset{n\to+\infty}{\to} 1$.

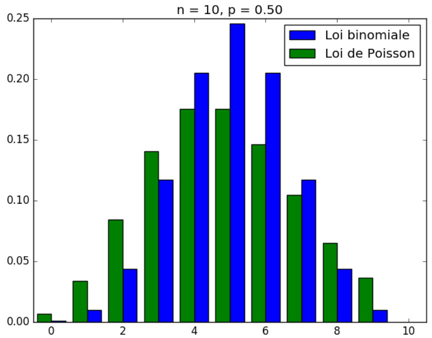
Comme $(1-p_n)^n = e^{n\ln(1-p_n)} = e^{n(-p_n + o_{+\infty}(p_n))} = e^{-\lambda + o_{+\infty}(1)}$ (avec 2 $o(1)$ obtenus en remplaçant p_n par son équivalent dans le o). Ainsi $(1-p_n)^n \underset{+\infty}{\to} e^{-\lambda} \neq 0$ (et donc on peut transformer cette limite en équivalent).

Il n'y a plus qu'à effectuer le produit de nos équivalents.. ∎

VI.1.2 En pratique

On peut utiliser une loi de Poisson pour approximer une loi binomiale de paramètre (n, p) dans le cas où $\lambda = np$ n'est "pas trop grand".





VI.2 Loi des grands nombres

VI.2.1 Théorème (Inégalité de Markov)
Soit X une variable aléatoire réelle à valeurs positives, d'espérance finie.

$$\forall a > 0 \ \mathbb{P}(X \geqslant a) \leqslant \frac{E(X)}{a}.$$

Preuve.
Notons $X(\Omega) = \{x_n ; n \in \mathbb{N}\}$
On a $a\mathbb{P}(X \geqslant a) = a \sum_{x_n \geqslant a} \mathbb{P}(X = x_n) \leqslant \sum_{x_n \geqslant a} a\mathbb{P}(X = x_n)$. Comme de plus les valeurs de X sont positives $a\mathbb{P}(X \geqslant a) \leqslant \sum_{n=0}^{+\infty} x_n \mathbb{P}(X = x_n) = E(X)$ et il suffit de diviser par $a > 0$. ∎

Explication L'idée "grossière" derrière ce théorème est que si l'espérance (la valeur moyenne) de X vaut m, alors X ne prend pas des valeurs trop grande par rapport à m, ou alors avec une probabilité très faible.

VI.2.2 Théorème (Inégalité de Bienaymé-Tchebychev)
Soit X une variable aléatoire de variance finie.

$$\forall \varepsilon > 0 \ \mathbb{P}(|X - E(X)| > \varepsilon) \leqslant \frac{V(X)}{\varepsilon^2}.$$

Preuve.
$(|X - E(X)| > \varepsilon) = ((X - E(X))^2 < \varepsilon^2)$. Il ne reste plus qu'à appliquer l'inégalité de Markov à $(X - E(X))^2$ qui est d'espérance finie et positive. ∎

Explication On quantifie cette fois l'écart entre X et sa "moyenne". La variance apparaît naturellement.

VI.2.3 Exemple
On pose S_n la moyenne arithmétique de n variables de loi de Bernoulli indépendantes de paramètre p. $S = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Exemple pratique : on dépouille une urne contenant n bulletins dans une élection à deux candidats. $(X_i = 1)$ est l'événement : le i-ème bulletin est pour le candidat A. Ici l'indépendance des variables n'est sûrement pas respecté dans la pratique. Tant pis, poursuivons.

Le but est d'estimer p, la proportion de votant ayant choisi le candidat A. Cette probabilité (théorique) est inconnue au moment de l'expérience.

Alors $E(S) = p$ et $V(S) = \frac{p(1-p)}{n}$.

S représente la proportion votes après n dépouillements indépendants. Alors $\mathbb{P}(|S - p| > \varepsilon) \leqslant \frac{p(p-1)}{n\varepsilon^2}$.

On veut $\mathbb{P}(|S - p| > \varepsilon) \leqslant 5\%$. Comment choisir ε ? Il faut $\frac{p(p-1)}{n\varepsilon^2} \leqslant \frac{5}{100}$ soit encore $\varepsilon^2 \leqslant 20\frac{p(1-p)}{n}$.

Or $p(1-p) \leqslant \frac{1}{4}$ (étape obligatoire, on ne connaît pas encore p). On a donc $\varepsilon^2 \leqslant \frac{5}{n}$.

Ainsi, si on veut une approximation de p à 1% près, on prend $\frac{1}{100} \leqslant \sqrt{5}\frac{1}{\sqrt{n}}$ soit encore $n \geqslant 5.10^4$.

Attention, on à juste le résultat : la probabilité pour que la fréquence théorique s'écarte de plus de 1% de la fréquence observée est $\leqslant \frac{5}{100}$

VI.2.4 Théorème (Loi faible des grands nombres)
Soit $(X_n)_{n \geqslant 1}$ une suite de variables aléatoires deux à deux indépendantes et de même loi, admettant un moment d'ordre 2.

On pose, pour $n \geqslant 1$, $S_n = \sum_{k=1}^{n} X_k$ et on note $m = E(X_1)$ l'espérance commune aux X_k.

$$\forall \eta > 0 \ \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geqslant \eta\right) \underset{n\to+\infty}{\to} 0.$$

Pour un $\eta > 0$ fixé, la limite est nulle.

Explication Ce théorème est la formalisation mathématique d'une idée naturelle.

Je répète n fois la même expérience aléatoire de Bernoulli (paramètre p) sans connaître a priori le paramètre p (on cherche à estimer une fréquence de manière empirique, par exemple pour réaliser un sondage...)

Alors la fréquence moyenne de succès converge vers le paramètre théorique p.

Résumé sur les lois usuelles

Nom	Notation	Valeurs	Loi	Fonctions génératrice	Espérance	Variance
Bernoulli	$\mathcal{B}(p)$	$\{0,1\}$	$\mathbb{P}(X=1)=p, \mathbb{P}(X=0)=1-p$	$G_X(t)=1-p+pt, t\in\mathbb{R}$	p	$p(1-p)$
Binomiale	$\mathcal{B}(n,p)$	$[\![0,n]\!]$	$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	$G_X(t)=(1-p+pt)^n, t\in\mathbb{R}$	np	$np(1-p)$
Géométrique	$\mathcal{G}(p)$	$\mathbb{N}\backslash\{0\}$	$\mathbb{P}(X=k)=p(1-p)^{k-1}$	$G_X(t)=\frac{pt}{1-(1-p)t}, t\in]-\frac{1}{1-p},\frac{1}{1-p}[$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson	$\mathcal{P}(\lambda), \lambda>0$	$\mathbb{N}$	$\mathbb{P}(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$	$G_X(t)=e^{\lambda(t-1)}, t\in\mathbb{R}$	λ	λ

Index