Université de Toulon - ESPE de Nice
MEEF Mathématiques 2019-2020
Thierry Champion

Probabilités

L'objet des probabilités est l'étude des expériences aléatoires.

Définition. *Une **expérience aléatoire** est une expérience qui conduit à des éventualités (ou résultats) bien définies à l'avance mais de manière imprévisible.*

Exemple. Le jeu de pile ou face consiste à lancer une pièce de monnaie et regarder quel côté (parmi "pile" et "face") est visible une fois qu'elle est tombée. C'est une expérience aléatoire : il y a deux résultats possibles bien définis à l'avance, mais on ne peut pas prévoir quel sera le résultat effectivement obtenu à l'issue d'une expérience.

1 Probabilité - Espace probabilisé

Dans toute la suite, Ω est un ensemble non vide, et $\mathcal{P}(\Omega)$ est l'ensemble des parties de Ω.

Pour chacune des définitions et notions suivantes, on précisera (lorsque cela a un intérêt) le cas où Ω est fini ou dénombrable.

1.1 Vocabulaire

- Ω est l'**univers** ou **univers des possibles**.
- Toute partie A de Ω est appelée **évènement**.
- Pour tout élément ω de Ω, le singleton $\{\omega\}$ est appelé **évènement élémentaire**.
- L'évènement $A \cup B$ se lit "A ou B", l'évènement $A \cap B$ se lit "A et B".
- $\emptyset$ est l'**évènement impossible**, et Ω est l'**évènement certain**.
- Pour A et B de $\mathcal{P}(\Omega)$, si $A \cap B = \emptyset$, ils sont dits **évènements incompatibles**.
- Pour $A \in \mathcal{P}(\Omega)$, son complémentaire dans Ω est noté $\overline{A}$ et est l'**évènement contraire** de A.

1.2 Tribu

Définition. *On appelle **tribu** (ou $\sigma-$**algèbre**) sur Ω un sous-ensemble $\mathcal{T}$ de $\mathcal{P}(\Omega)$ satisfaisant*
1. *Ω et $\emptyset$ appartiennent à $\mathcal{T}$,*
2. *si $A \in \mathcal{T}$, alors son contraire $\overline{A}$ appartient à $\mathcal{T}$,*
3. *si $(A_n)_{n\in\mathbb{N}}$ est une famille dénombrable d'éléments de $\mathcal{T}$, alors $\displaystyle\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{T}$.*

Définition. *Un espace probabilisable est un couple $(\Omega, \mathcal{T})$ où $\mathcal{T}$ est une tribu sur Ω.*

Remarque. *Dans le cas où Ω est fini ou dénombrable*, la tribu naturelle sur Ω est $\mathcal{P}(\Omega)$. Dans le cas général, $\mathcal{P}(\Omega)$ et $\{\emptyset, \Omega\}$ sont toujours des tribus sur Ω. La tribu Borélienne sur $\mathbb{R}$ n'est pas au programme.

1.1. Propriété - Stabilité par intersection.

Si $\mathcal{T}$ est une tribu sur Ω et si $(A_n)_{n\in\mathbb{N}}$ est une collection dénombrable d'éléments de $\mathcal{T}$, alors $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{T}$.

1.3 Probabilité

Définition. *Soit $(\Omega, \mathcal{T})$ un espace probabilisable. Une* ***probabilité*** *P sur $(\Omega, \mathcal{T})$ est une application $P : \mathcal{T} \to [0, 1]$ satisfaisant*

1. *normalisation : $P(\Omega) = 1$,*
2. *σ-additivité : si $(A_n)_{n\in\mathbb{N}}$ est une suite d'évènements de $\mathcal{T}$ deux à deux incompatibles alors $P\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n=0}^{\infty} P(A_n)$.*

Définition. *Un* ***espace probabilisé*** *est un triplet $(\Omega, \mathcal{T}, P)$ où P est une probabilité sur l'espace probabilisable $(\Omega, \mathcal{T})$.*

Remarque. *Dans le cas où Ω est fini*, il est courant de munir Ω de la tribu $\mathcal{P}(\Omega)$. Une probabilité P sur $\mathcal{P}(\Omega)$ est alors une application $P : \mathcal{P}(\Omega) \to [0, 1]$ qui vérifie :

1. normalisation : $P(\Omega) = 1$,
2. additivité : si A et B sont deux évènements incompatibles alors $P(A \cup B) = P(A) + P(B)$.

De plus dans ce cas, si Ω est de cardinal n et $\Omega = \{\omega_1, \dots, \omega_n\}$, alors une probabilité P sur $(\Omega, \mathcal{P}(\Omega))$ est totalement définie par la donnée des nombres $P(\{\omega_i\})$ puisqu'alors l'additivité permet d'écrire

$$\forall A \in \mathcal{P}(\Omega), \qquad P(A) = \sum_{\omega\in A} P(\{\omega\}).$$

Exemple. Soit Ω un ensemble fini de cardinal n avec $\Omega = \{\omega_1, \dots, \omega_n\}$. On dit qu'il y a équiprobabilité sur Ω lorsqu'on considère l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ tel que

$$P(\{\omega_1\}) = P(\{\omega_2\}) = \ldots = P(\{\omega_n\}) = \frac{1}{n}.$$

Dans ce cas pour tout évènement A on a $P(A) = \frac{\#A}{\#\Omega}$.

Remarque. *Dans le cas où Ω est dénombrable*, il est courant de munir Ω de la tribu $\mathcal{P}(\Omega)$. De plus si $\Omega = \{\omega_n : n \in \mathbb{N}\}$, alors une probabilité P est totalement définie par la donnée des nombres $P(\{\omega_n\})$ pour tout $n \in \mathbb{N}$ puisqu'alors la σ-additivité permet d'écrire

$$\forall A \in \mathcal{P}(\Omega), \qquad P(A) = \sum_{\omega\in A} P(\{\omega\}).$$

La somme ci-dessus est bien définie car il s'agit d'une somme de nombres positifs sur un ensemble au plus dénombrable, et que cette somme est majorée par 1 (pour le cas $A = \Omega$).

Remarque. *Dans le cas où Ω est fini ou dénombrable*, l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ est souvent simplement noté (Ω, P).

1.2. Propriété - Généralités.

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, alors :
- $P(\emptyset) = 0$,
- pour tout évènement $A \in \mathcal{T}$ on a $P(A) + P(\overline{A}) = 1$,
- si $A, B \in \mathcal{T}$ sont incompatibles alors $P(A \cup B) = P(A) + P(B)$,
- formule du crible : si $A, B \in \mathcal{T}$ alors $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,
- formule du crible (générale) ou formule de Poincaré : soit $A_1, A_2, \ldots, A_n$ des événements de $\mathcal{T}$ alors

$$P(A_1 \cup A_2 \cup ... \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{\{i_1, ..., i_k\} \subset [1, n] \\ \mathrm{Card}(\{i_1, ..., i_k\}) = k}} P(A_{i_1} \cap ... \cap A_{i_k}).$$

- inégalité de Boole : si $(A_n)_{n \in \mathbb{N}}$ est une suite d'évènements de $\mathcal{T}$ alors

$$P\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) \leq \sum_{n=0}^{\infty} P(A_n).$$

- si A et B sont deux évènements tels que $A \subset B$ alors $P(A) \leq P(B)$,
- si $(A_n)_{n \in \mathbb{N}}$ est une suite croissante d'évènements de $\mathcal{T}$ alors

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

Définition. *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé. Si $A \in \mathcal{T}$ est tel que $A \neq \Omega$ et $P(A) = 1$ (respectivement $A \neq \emptyset$ et $P(A) = 0$), on dit que A est un **évènement presque sûr** ou **presque certain** (respectivement A est un **évènement presque impossible** ou **négligeable**).*

1.4 Indépendance - Probabilité conditionnelle

Dans toute cette section, $(\Omega, \mathcal{T}, P)$ est un espace probabilisé.

Définition. *Deux évènements A et B de $\mathcal{T}$ sont **indépendants** si $P(A \cap B) = P(A)\, P(B)$.*

Remarque. Si deux évènements A et B de $\mathcal{T}$ ne sont pas indépendants, on dit que "A et B ne sont pas indépendants" (le terme "dépendant" ne s'emploie pas dans ce cas là).

Définition. *Soit $(A_i)_{i \in I}$ une famille au plus dénombrable d'évènements de $\mathcal{T}$. Les évènements A_i sont **indépendants deux à deux** si pour tout $(i, j) \in I^2$ les évènements A_i et A_j sont indépendants. Les évènements A_i sont **mutuellement indépendants** si*

$$P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$$

pour tout $J \subset I$.

Remarque. Des évènements mutuellement indépendants sont indépendants deux à deux. En revanche des évènements peuvent être indépendants deux à deux sans être nécessairement mutuellement indépendants.

Définition. *Soient A et B deux évènements de $\mathcal{T}$. On suppose que $P(A) \neq 0$. La **probabilité de B sachant A** (ou **probabilité conditionnelle de B sachant A**) est le nombre noté $P(B|A)$ ou $P_A(B)$ donné par $P(B|A) = \frac{P(A \cap B)}{P(A)}$.*

1.3. Propriété - Probabilité conditionnelle.

Soit A un évènement de $\mathcal{T}$ tel que $P(A) \neq 0$. Alors l'application P_A définie sur $\mathcal{T}$ par $P_A : B \longmapsto P(B|A)$ est une probabilité sur $(\Omega, \mathcal{T})$.

1.4. Proposition - Caractérisation de l'indépendance.

Si A et B sont deux évènement de $\mathcal{T}$ de probabilité non nulle, alors

$$A \text{ et } B \text{ sont indépendants} \iff P(B|A) = P(B) \iff P(A|B) = P(A).$$

1.5. Proposition - Formule des probabilités totales.

Soit $(A_1, ..., A_m)$ un **système complets d'évènements** (c'est-à-dire une partition finie de Ω formée d'éléments de $\mathcal{T}$ tous non vides) telle que $P(A_i) \neq 0$ pour tout $i \in \{1, \ldots, m\}$. Alors pour tout évènement B de $\mathcal{T}$ on a

$$P(B) = \sum_{i=1}^{n} P(B|A_i)\, P(A_i).$$

1.6. Proposition - Formule de Bayes.

Soit $(A_1, ..., A_m)$ une partition finie de Ω telle que A_i est un évènement de $\mathcal{T}$ et $P(A_i) \neq 0$ pour tout $i \in \{1, \ldots, m\}$. Alors pour tout évènement B de $\mathcal{T}$ tel que $P(B) \neq 0$ on a

$$\forall j \in \{1, \ldots, m\}, \qquad P(A_j|B) = \frac{P(B|A_j)\, P(A_j)}{P(B)} = \frac{P(B|A_j)\, P(A_j)}{\displaystyle\sum_{i=1}^{n} P(B|A_i)\, P(A_i)}.$$

2 Variables aléatoires

Dans toute cette partie, $(\Omega, \mathcal{T}, P)$ est un espace probabilisé.

2.1 Généralités

Définition. *Une application $X : \Omega \to \mathbb{R}$ est une **variable aléatoire réelle sur** $(\Omega, \mathcal{T})$ si $X^{-1}(\,]-\infty, a]) \in \mathcal{T}$ pour tout $a \in \mathbb{R}$.*

2.1. Propriété - Caractérisation des variables aléatoires.

Une application $X : \Omega \to \mathbb{R}$ est une variable aléatoire réelle sur $(\Omega, \mathcal{T})$ si et seulement si $X^{-1}(I) \in \mathcal{T}$ pour tout intervalle I de $\mathbb{R}$.

Définition. *Soit $d \geq 1$. Une application $X : \Omega \to \mathbb{R}^d$ donnée par*

$$\forall \omega \in \Omega, \qquad X(\omega) = (X_1(\omega), \ldots, X_d(\omega))$$

*est une **variable (ou un vecteur) aléatoire à valeurs dans $\mathbb{R}^d$ sur** $(\Omega, \mathcal{T})$ si X_i est une variable aléatoire réelle pour tout $i \in \{1, \ldots, d\}$.*

Remarque. Dans le cas où Ω est fini ou dénombrable et $\mathcal{T} = \mathcal{P}(\Omega)$, toute application $X : \Omega \to \mathbb{R}$ est une variable aléatoire réelle, et toute application $Y : \Omega \to \mathbb{R}^d$ est une variable aléatoire à valeurs dans $\mathbb{R}^d$.

2.2. Propriété - Algèbre des variables aléatoires [admis].

La somme et le produit de deux variables aléatoires réelles sur $(\Omega, \mathcal{T})$ sont des variables aléatoires réelles sur $(\Omega, \mathcal{T})$.
La somme de deux variables aléatoires à valeurs dans $\mathbb{R}^d$ sur $(\Omega, \mathcal{T})$ est une variable aléatoire à valeurs dans $\mathbb{R}^d$.

Notation. *Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle sur $(\Omega, \mathcal{T})$. Si $E \subset \mathbb{R}$, on note généralement $(X = E)$ ou $X = E$ l'ensemble $X^{-1}(E)$. De plus, si $a < b \in \mathbb{R}$, alors*

- *on note $(X = a)$ l'évènement $X^{-1}(\{a\})$,*
- *on note $(X < a)$ l'évènement $X^{-1}(\,]-\infty, a[\,)$,*
- *on note $(X \leq a)$ l'évènement $X^{-1}(\,]-\infty, a]\,)$,*
- *on note $(a \leq X \leq b)$ l'évènement $X^{-1}(\,[a, b]\,)$...*

On verra que l'étude des propriétés d'une variable aléatoire est étroitement liée à sa loi : on définira la loi en fonction des hypothèses faites sur la variable aléatoire (discrète ou à densité). Dans tous les cas, la loi d'une variable aléatoire est déterminée par sa fonction de répartition, qui est donc essentielle dans l'étude des propriétés d'une variable aléatoire.

Définition. *Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{T}, P)$, la **fonction de répartition de** X est l'application $F_X : x \longmapsto P(X \leq x)$, définie sur $\mathbb{R}$ et à valeurs dans $[0, 1]$.*

2.2 Variables aléatoires réelles discrètes

Définition. *Une variable aléatoire réelle X sur $(\Omega, \mathcal{T})$ est **discrète** si $X(\Omega)$ est fini ou dénombrable.*

Remarque. Si X est une variable aléatoire discrète sur $(\Omega, \mathcal{T})$ et $g : X(\Omega) \to \mathbb{R}$, alors $g \circ X$ est une variable aléatoire discrète sur $(\Omega, \mathcal{T})$.

Définition. *Soit X une variable aléatoire réelle discrète sur $(\Omega, \mathcal{T})$, la **loi de probabilité de** X est l'application $E \longmapsto P(X \in E)$, définie sur $\mathcal{P}(\mathbb{R})$ et à valeurs dans $[0, 1]$.*

2.3. Propriété - Caractérisation de la loi de probabilité.

Soit X une variable aléatoire réelle discrète sur $(\Omega, \mathcal{T})$, alors la loi de probabilité de X est une probabilité sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$.
De plus, la loi de probabilité de X est complètement déterminée par la donnée de $P(X = x)$ pour tout $x \in X(\Omega)$.

2.4. Propriété - Régularité de la fonction de répartition.

Soit X une variable aléatoire réelle discrète sur $(\Omega, \mathcal{T})$, alors la fonction de répartition de X est une fonction continue à droite et admettant une limite à gauche en tout point de $\mathbb{R}$ (on dit que c'est une fonction *cadlag*).

Remarque. Si X est une variable aléatoire réelle discrète sur $(\Omega, \mathcal{T})$, alors les points de discontinuité de sa fonction de répartition F_X sont exactement les valeurs prises par X sur Ω. En particulier, pour toute valeur $x \in X(\Omega)$ on a

$$P(X = x) \;=\; F_X(x) - F_X(x^-) \;=\; F_X(x) - \lim_{\substack{t \to x \\ t < x}} F_X(t)$$

où $F_X(x^-)$ est la limite à gauche de F_X en x. Ainsi, la fonction de répartition F_X permet de retrouver la loi de X (voir la Propriété 2.3).

Définition. *Soit X une variable aléatoire réelle discrète sur $(\Omega, \mathcal{T})$, X* ***admet une espérance*** *si la série $\displaystyle\sum_{x \in X(\Omega)} x\, P(X = x)$ converge absolument. Dans ce cas,* ***l'espérance de*** *X (ou la* moyenne de *X) est le nombre réel noté $\mathbb{E}(X)$ donné par $\displaystyle\mathbb{E}(X) = \sum_{x \in X(\Omega)} x\, P(X = x)$.*

2.5. Propriété.

L'espérance est linéaire : si X et Y sont des v.a. discrètes $(\Omega, \mathcal{T})$ admettant une espérance, alors

$$\forall \alpha, \beta \in \mathbb{R}, \qquad \mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)\,.$$

Si X une variable aléatoire réelle discrète bornée sur $(\Omega, \mathcal{T})$, alors X admet une espérance et

$$\inf\{X(\omega)\,:\, \omega \in \Omega\} \leq \mathbb{E}(X) \leq \sup\{X(\omega)\,:\, \omega \in \Omega\}.$$

Si Ω est fini ou dénombrable, alors toute variable aléatoire réelle X sur $(\Omega, \mathcal{T})$ est discrète et on a $\displaystyle\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\, P(\{\omega\})$ si cette série converge absolument.

Exemple. Soit X une v.a.r. constante sur $(\Omega, \mathcal{T})$, c'est-à-dire $X(\Omega) = \{x\}$ pour un certain réel x, alors $\mathbb{E}(X) = x$.

Définition. *Soit X une variable aléatoire réelle discrète discrète sur $(\Omega, \mathcal{T})$ admettant une espérance, on dit que X admet une variance si la v.ar. $(X - \mathbb{E}(X))^2$ admet une espérance. Dans ce cas, la* ***variance de*** *X est le nombre réel noté $V(X)$ donné par $V(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$. L'****écart-type de*** *X est alors le nombre noté $\sigma(X)$ et donné par $\sigma(X) = \sqrt{V(X)}$.*

2.6. Propriété.
Si X est une v.a.r. sur $(\Omega, \mathcal{T})$ admettant une variance et $a, b \in \mathbb{R}$ alors

$$V(aX+b) = a^2\, V(X).$$

Soit X une variable aléatoire réelle discrète sur $(\Omega, \mathcal{T})$, alors on peut appliquer la formule de Koenig :

$$V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Enfin, si X est une variable aléatoire réelle discrète bornée sur $(\Omega, \mathcal{T})$, alors

$$\sigma(X) \leq \sup(|X(\omega) - \mathbb{E}(X)| \,:\, \omega \in \Omega\}.$$

Définition. *Soit X une variable aléatoire réelle discrète sur $(\Omega, \mathcal{T})$ et $k \in \mathbb{N}^*$, alors le* ***moment d'ordre** k **de** X est le nombre $m_k(X) = \mathbb{E}(X^k)$ et le* ***moment centré d'ordre** k **de** X est le nombre $\mu_k(X) = \mathbb{E}([X - \mathbb{E}(X)]^k)$ (lorsque ces deux nombres sont bien définis).*

2.3 Variables aléatoires réelles discrètes - Lois au programme

Loi uniforme. La loi uniforme modélise une expérience aléatoire ayant un nombre fini d'issues qui sont équiprobables.
Une variable X aléatoire réelle sur $(\Omega, \mathcal{T})$ suit la loi uniforme si $X(\Omega)$ est un ensemble fini de cardinal $n \geq 1$ et $P(X = x) = \frac{1}{n}$ pour tout $x \in X(\Omega)$.

Loi de Bernoulli $\mathcal{B}(p)$. Soit $p \in [0,1]$. La loi de Bernoulli de paramètre p modélise une expérience ayant deux issues possibles : le succès (avec une probabilité p) et l'échec (avec une probabilité $1-p$).
Une variable X aléatoire réelle sur $(\Omega, \mathcal{T})$ suit la loi de Bernoulli de paramètre p (notée $\mathcal{B}(p)$) si $X(\Omega) = \{0,1\}$, $P(X=1) = p$ et $P(X=0) = 1-p$. Dans ce cas, $\mathbb{E}(X) = p$ et $V(X) = p\,(1-p)$.

Loi de binomiale $\mathcal{B}(n,p)$. Soit $n \geq 1$ et $p \in [0,1]$. La loi de binomiale de paramètres n et p modélise la répétion successive de n épreuves de Bernoulli de paramètre p, dont on compte le nombre de succès.
Une variable X aléatoire réelle sur $(\Omega, \mathcal{T})$ suit la loi binomiale de paramètres n et p (notée $\mathcal{B}(n,p)$) si $X(\Omega) = \{0, 1, \ldots, n\}$ et

$$\forall k \in \{0, 1, \ldots, n\}, \qquad P(X = k) = \left(\begin{array}{c} n \\ k \end{array}\right) p^k\,(1-p)^{n-k}.$$

Dans ce cas, $\mathbb{E}(X) = n\,p$ et $V(X) = n\,p\,(1-p)$.

Loi hypergéométrique $H(N,p,n)$. Soit $M \geq n \geq 1$ et $p \in [0,1]$. La loi hypergéométrique de paramètres N, n et p modélise l'expérience qui consiste à choisir un échantillon de n individus dans une population d'effectif total N, population dans laquelle la fréquence d'un certain caractère C est p (donc l'entier $N\,p$ est le nombre d'individus ayant ce caractère dans la population) (on peut aussi le formuler en disant que chaque individu a la probabilité p d'avoir le caractère C) : on compte alors le nombre d'éléments de l'échantillon choisi

ayant ce caractère C.
Une variable X aléatoire réelle sur $(\Omega, \mathcal{T})$ suit la loi hypergéométrique $H(N,p,n)$ de paramètres N, n et p (notée $H(N,p,n)$) si $X(\Omega) = \{0,1,\ldots,n\}$ et

$$\forall k \in \{0,1,\ldots,n\}, \qquad P(X=k) = \frac{\begin{pmatrix} N\,p \\ k \end{pmatrix} \begin{pmatrix} N\,(1-p) \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}.$$

Dans ce cas, $\mathbb{E}(X) = n\,p$ et $V(X) = n\,p\,(1-p)\frac{N-n}{N-1}$.

Loi géométrique. Soit $p \in [0,1]$. La loi géométrique de paramètre p modélise la répétition successive et indépendante d'expériences suivant la loi de Bernoulli de paramètre p, et ce jusqu'à l'obtention d'un succès : on s'intéresse alors au nombre d'expériences qu'il a fallu effectuer pour obtenir le premier succès.
Une variable X aléatoire réelle sur $(\Omega, \mathcal{T})$ suit la loi géométrique de paramètre p si $X(\Omega) = \mathbb{N}^*$ et $P(X=k) = (1-p)^{k-1}\,p$ pour tout $k \in \mathbb{N}^*$. Dans ce cas, $\mathbb{E}(X) = \frac{1}{p}$ et $V(X) = \frac{1-p}{p^2}$.

Loi de Poisson $\mathcal{P}(\lambda)$. Soit $\lambda > 0$. Une variable X aléatoire réelle sur $(\Omega, \mathcal{T})$ suit la loi de Poisson de paramètre λ (ou suit $\mathcal{P}(\lambda)$) si $X(\Omega) = \mathbb{N}$ et $P(X=k) = \frac{\lambda^k}{k!}\,e^{-\lambda}$ pour tout $k \in \mathbb{N}$. Dans ce cas, $\mathbb{E}(X) = V(X) = \lambda$.

2.4 Vecteurs aléatoires (à valeurs dans $\mathbb{R}^d$) discrets

2.4.1 Généralités

Définition. *Soit $d \geq 1$. Un vecteur aléatoire discret sur $(\Omega, \mathcal{T})$ à valeurs dans $\mathbb{R}^d$ est une variable aléatoire $X : \Omega \to \mathbb{R}^d$ donnée par*

$$\forall \omega \in \Omega, \qquad X(\omega) = (X_1(\omega), \ldots, X_d(\omega)),$$

dont la coordonnée X_i est une variable aléatoire réelle discrète sur $(\Omega, \mathcal{T})$ pour tout $i \in \{1,\ldots,d\}$.

Définition. *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$, la **loi de probabilité de** X (ou **loi conjointe des v.a.** X_i) est l'application $E \longmapsto P(X \in E)$, définie sur $\mathcal{P}(\mathbb{R}^d)$ et à valeurs dans $[0,1]$.*

2.7. Propriété - Caractérisation de la loi de probabilité.

Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$, alors la loi de probabilité de X est une probabilité sur $(\mathbb{R}^d, \mathcal{P}(\mathbb{R}^d))$.
De plus, la loi de probabilité de X est complètement déterminée par la donnée de $P(X=x)$ pour tout $x \in X(\Omega) = X_1(\Omega) \times \ldots \times X_d(\Omega)$.

On se penche maintenant sur le cas $d = 2$.

Définition. *Soit $X = (X_1, X_2)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$, la loi de X_1 (respectivement X_2) sur $(\Omega, \mathcal{T}, P)$ est appelée **loi marginale de** X_1 (resp. X_2).*

Remarque. La loi marginale de X_1 s'obtient facilement à partir de la loi conjointe de (X_1, X_2) en calculant chaque probabilité $P(X_1 = x)$ par la formule :

$$P(X_1 = x) = \sum_{y \in X_2(\Omega)} P((X_1, X_2) = (x,y)) = \sum_{y \in X_2(\Omega)} P((X_1 = x) \cap (X_2 = y)).$$

On présente en général la loi conjointe et les lois marginales de X_1 et X_2 dans un tableau.

Définition. *Soit $X = (X_1, X_2)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$, et soit $x_2 \in \mathbb{R}$ tel que $P(X_2 = x_2) \neq 0$. La* ***loi conditionnelle de*** *X_1* ***sachant*** *$X_2 = x_2$ est la loi de X_1 dans l'espace probabilisé $(\Omega, \mathcal{T}, P_{(X_2=x_2)})$, c'est-à-dire l'application $E \longmapsto P(X_1 \in E | X_2 = x_2)$, définie sur $\mathcal{P}(\mathbb{R})$ et à valeurs dans $[0, 1]$.*

Remarque. La connaissance des lois conditionnelles de X_1 sachant $X_2 = x_2$ pour toutes les valeurs x_2 de X_2 telles que $P(X_2 = x_2) \neq 0$ détermine la loi de X_1 grâce à la formule des probabilités totales.

2.4.2 Indépendance

Définition. *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$, les variables aléatoires $X_1, \ldots, X_d$ sont dites* ***(mutuellement) indépendantes*** *si pour toute valeur $x = (x_1, \ldots, x_d)$ de X les évènements $(X_1 = x_1), \ldots, (X_d = x_d)$ sont mutuellement indépendants.*

Remarque. En particulier, si $X = (X_1, X_2)$ est un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$ alors X_1 et X_2 sont indépendantes si et seulement si pour tout $x_1 \in X_1(\Omega)$ et $x_2 \in X_2(\Omega)$ on a $P((X_1 = x_1) \cap (X_2 = x_2)) = P(X_1 = x_1)\, P(X_2 = x_2)$.

2.8. Propriété - Caractérisation de l'indépendance.

Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$, alors les variables aléatoires $X_1, \ldots, X_d$ sont mutuellement indépendantes si et seulement si pour tous intervalles $I_1, \ldots, I_d$ de $\mathbb{R}$ les évènements $(X_1 \in I_1), \ldots, (X_d \in I_d)$ sont mutuellement indépendants.

2.9. Proposition - [Admis].

Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$, si les variables aléatoires $X_1, \ldots, X_d$ sont mutuellement indépendantes alors pour tout $k \in \{1, d-1\}$ et pour toutes fonctions $f : \mathbb{R}^k \to \mathbb{R}$ et $g : \mathbb{R}^{d-k} \to \mathbb{R}$ les variables aléatoires discrètes $\omega \longmapsto f(X_1(\omega), \ldots, X_k(\omega))$ et $\omega \longmapsto g(X_{k+1}(\omega), \ldots, X_d(\omega))$ sont indépendantes.

2.4.3 Somme de variables aléatoires

On rappelle que l'espérance est une forme linéaire sur le $\mathbb{R}$-espace vectoriel des variables aléatoires réelles discrètes sur $(\Omega, \mathcal{T})$. On étudie dans ce qui suit l'additivité de la variance.

Définition. *Soit $X = (X_1, X_2)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$ tel que X_1 et X_2 ont une variance, la* ***covariance*** *des variables aléatoires X_1 et X_2 est le nombre réel noté $cov(X_1, X_2)$ donné par*

$$cov(X_1, X_2) \;=\; \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))\,(X_2 - \mathbb{E}(X_2))\right].$$

2.10. Propriété.

Soit $X = (X_1, X_2)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$ tel que X_1 et X_2 ont une variance, alors

$$cov(X_1, X_2) \;=\; \mathbb{E}(X_1\, X_2) - \mathbb{E}(X_1)\,\mathbb{E}(X_2).$$

De plus, l'inégalité de Cauchy-Schwartz donne

$$|cov(X_1, X_2)| \;\leq\; \sqrt{V(X_1)\,V(X_2)}.$$

2.11. Propriété - Variance d'une somme de deux v.a..

Soit $X = (X_1, X_2)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$, alors

$$V(X_1 + X_2) \;=\; V(X_1) + V(X_2) + 2\,cov(X_1, X_2).$$

De plus, si les variables aléatoires X_1 et X_2 sont indépendantes alors

$$cov(X_1, X_2) \;=\; 0 \qquad et \qquad V(X_1 + X_2) \;=\; V(X_1) + V(X_2).$$

Dans le cas où on somme plus de deux variables aléatoires discrètes, on peut utiliser le résultat suivant.

2.12. Proposition - Variance d'une somme de v.a. indépendantes.

Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$ tel que les variables aléatoires $X_1, \ldots, X_d$ soient mutuellement indépendantes, alors

$$V(X_1 + \ldots + X_d) \;=\; V(X_1) + \ldots + V(X_d).$$

On réalise en général la somme de variables aléatoires mutuellement indépendantes suivant certaines loi spécifiques, en voici deux exemples au programme.

2.13. Proposition - Stabilité de la loi binomiale par la somme.

Soit $X = (X_1, X_2)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$ tel que X_1 suit la loi binomiale de paramètres n et p, X_2 suit la loi binomiale de paramètres k et p, et les variables aléatoires X_1 et X_2 sont indépendantes. Alors $X_1 + X_2$ suit la loi binomiale de paramètres $n + k$ et p.

2.14. Proposition - Stabilité de la loi de Poisson par la somme.

Soit $X = (X_1, X_2)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$ tel que X_1 suit la loi de Poisson de paramètre $\lambda > 0$, X_2 suit la loi de Poisson de paramètre $\mu > 0$, et les variables aléatoires X_1 et X_2 sont indépendantes. Alors $X_1 + X_2$ suit la loi de Poisson de paramètre $\lambda + \mu$.

2.4.4 Corrélation linéaire

Soit $X = (X_1, X_2)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$, alors la régression linéaire de X_2 par rapport à X_1 est la droite dont l'équation est donnée par

$$y = a\,x + b \quad avec \quad a = \frac{cov(X_1, X_2)}{V(X_1)} \quad et \quad b = \mathbb{E}(X_2) - a\,\mathbb{E}(X_1).$$

On dit que les variables aléatoires X_1 et X_2 sont **bien corrélées** (ou que l'ajustement de X_2 sur X_1 est bon) si l'approximation $X_2 \simeq a\,X_1 + b$ est de bonne qualité. Dans le cas contraire, on dit que ces variables aléatoires sont décorrélées.

Définition. *Soit $X = (X_1, X_2)$ un vecteur aléatoire discret sur $(\Omega, \mathcal{T}, P)$, alors le* ***coefficient de corrélation linéaire des variables aléatoires*** *X_1* ***et*** *X_2 est le nombre réel $\rho(X_1, X_2)$ donné par*

$$\rho(X_1, X_2) \;=\; \frac{cov(X_1, X_2)}{\sqrt{V(X_1)\,V(X_2)}} \;=\; \frac{cov(X_1, X_2)}{\sigma(X_1)\,\sigma(X_2)}.$$

La valeur absolue $|\rho(X_1, X_2)|$ du coefficient de corrélation linéaire des variables aléatoires X_1 et X_2 donne une mesure de leur corrélation : plus ce nombre est proche de 1 (respectivement proche de 0), plus (resp. moins) les variables aléatoires X_1 et X_2 sont corrélées. En effet, on peut vérifier que ce nombre vaut exactement 1 dans le cas où on a exactement $X_2 = a\,X_1 + b$
Le signe de $\rho(X_1, X_2)$ renseigne aussi sur le sens de variation de X_2 par rapport à X_1 (dans le cas où ces variables sont corrélées).

2.5 Variables aléatoires à densité

2.5.1 Généralités

Définition. *Une variable aléatoire réelle X sur $(\Omega, \mathcal{T})$* ***admet une densité*** *f_X si sa fonction de répartition F_X est donnée par*

$$\forall t \in \mathbb{R}, \qquad F_X(t) \;=\; \int_{-\infty}^{t} f_X(s)ds.$$

Définition. *Une* ***densité*** *est une fonction $f : \mathbb{R} \to \mathbb{R}$ qui est* positive*,* continue par morceaux avec un nombre fini de discontinuités*, et* intégrable sur $\mathbb{R}$ *avec $\displaystyle\int_{\mathbb{R}} f(s)ds = 1$.*

2.15. Proposition - Caractérisation des v.a. à densité.

Une variable aléatoire réelle X sur $(\Omega, \mathcal{T})$ admet une densité f_X si et seulement si sa fonctoin de répartition F_X est **continue sur** $\mathbb{R}$ et C^1 par morecaux sur $\mathbb{R}$ (avec un nombre fini de points où elle n'est pas dérivable). Dans ce cas on a $f_X(x) = F'_X(x)$ en tout point $x \in \mathbb{R}$ auquel F_X est dérivable.

2.16. Proposition - Loi d'une v.a. à densité.

Si une variable aléatoire réelle X sur $(\Omega, \mathcal{T})$ admet une densité f alors pour tout intervalle I de $\mathbb{R}$ on a

$$P(X \in I) \;=\; \int_I f(t)dt,$$

et de plus $P(X = x) = \int_x^x f(t)dt = 0$ pour tout réel $x \in \mathbb{R}$.

Remarque. Si une variable aléatoire réelle X sur $(\Omega, \mathcal{T})$ admet une densité f_X alors les valeurs prises par X sont les nombres réels x appartenant à la fermeture $\overline{\{t\,:\,f_X(t)>0\}}$ de l'ensemble $\{t\,:\,f_X(t)>0\}$ (cette fermeture est appelée *support de f*).

Définition. *Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{T})$ qui admet une densité f_X. On dit que X **admet une espérance** si $t \longmapsto t\, f_X(t)$ est absolument intégrable sur $\mathbb{R}$; dans ce cas, **l'espérance de** X (ou la* moyenne de X*) est le nombre réel noté $\mathbb{E}(X)$ donné par*

$$\mathbb{E}(X) = \int_{\mathbb{R}} t\, f_X(t)dt.$$

2.17. Propriété.

Si la variable aléatoire réelle X sur $(\Omega, \mathcal{T})$ admet la densité f, et si $g : \mathbb{R} \to \mathbb{R}$ est continue par morceaux et est telle que $t \longmapsto g(t)\, f_X(t)$ est absolument intégrable sur $\mathbb{R}$, alors $g \circ X$ admet une espérance donnée par

$$\mathbb{E}(g \circ X) = \int_{\mathbb{R}} g(t)\, f_X(t)dt.$$

Définition. *Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{T})$ qui admet une densité f_X. On dit que X **admet une variance** si X admet une espérance et si $t \longmapsto t^2\, f_X(t)$ est intégrable sur $\mathbb{R}$; dans ce cas, **la variance de** X est le nombre réel noté $V(X)$ donné par*

$$V(X) = \int_{\mathbb{R}} (t - \mathbb{E}(X))^2\, f_X(t)dt.$$

2.18. Propriété.

Si la variable aléatoire réelle X sur $(\Omega, \mathcal{T})$ admet la densité f_X, et si X admet une variance, alors on peut appliquer la **formule de Koenig** :

$$V(X) \ = \ \mathbb{E}(X^2) - \mathbb{E}(X)^2 \ = \ \int_{\mathbb{R}} t^2\, f_X(t)dt - \left(\int_{\mathbb{R}} t\, f_X(t)dt\right)^2.$$

2.5.2 Vecteur aléatoire à densité

Définition. *Un vecteur aléatoire (X_1, X_2) à valeurs dans $\mathbb{R}^2$ **admet une densité** f si sa fonction de répartition F est donnée par*

$$\forall a, b \in \mathbb{R}, \qquad F((a,b)) \ = \ P[(X_1 \leq a) \cap (X_2 \leq b)] \ = \ \int_{-\infty}^{a} \int_{-\infty}^{b} f(x_1, x_2) dx_2 dx_1$$

où f est le produit d'une fonction continue et positive sur $\mathbb{R}^2$ par la fonction indicatrice d'un ensemble "géométriquement simple", et si f est intégrable sur $\mathbb{R}^2$ avec $\displaystyle\int_{\mathbb{R}^2} f(x_1, x_2)dx_1dx_2 = 1$.

2.19. Proposition - Loi d'un vecteur à densité.

Si un vecteur aléatoire (X_1, X_2) à valeurs dans $\mathbb{R}^2$ admet une densité f alors pour tout intervalle I_1, I_2 de $\mathbb{R}$ on a

$$P[(X_1 \in I_1) \cap (X_2 \in I_2)] \;=\; \int_{I_1} \int_{I_2} f(x_1, x_2) dx_2 dx_1.$$

Remarque. Puisque f est le produit d'une fonction continue et positive sur $\mathbb{R}^2$ par la fonction indicatrice d'un ensemble E "géométriquement simple", on doit pouvoir écrire pour tout élément (a, b) dans l'intérieur de E que

$$P[(X_1 \in [a - \varepsilon, a + \varepsilon]) \cap (X_2 \in [b - \varepsilon, b + \varepsilon])] \;\simeq\; (2\,\varepsilon)^2\, f(a, b),$$

ce qui permet de calculer f.

2.20. Propriété - Lois marginales d'un vecteur à densité.

Si un vecteur aléatoire (X_1, X_2) à valeurs dans $\mathbb{R}^2$ admet une densité f, alors la fonction de répartition de X_1 est donnée par

$$\forall t \in \mathbb{R}, \qquad F_{X_1}(t) \;=\; \int_{-\infty}^{t} \int_{\mathbb{R}} f(x_1, x_2) dx_2 dx_1.$$

et en particulier X_1 a pour densité

$$f_{X_1} : x \longmapsto \int_{\mathbb{R}} f(x, x_2) dx_2.$$

Définition. *Soit (X_1, X_2) un vecteur aléatoire à valeurs dans $\mathbb{R}^2$ qui admet une densité f. Les variables aléatoires X_1 et X_2 sont **indépendantes** si pour tout point $(x_1, x_2) \in \mathbb{R}^2$ en lequel f est continue on peut écrire $f(x_1, x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2)$.*

2.21. Propriété - Variance d'une somme de deux v.a. .

Soit $X = (X_1, X_2)$ un vecteur aléatoire sur $(\Omega, \mathcal{T}, P)$ admettant la densité f. On suppose que X_1 et X_2 ont une variance, alors

$$V(X_1 + X_2) \;=\; V(X_1) + V(X_2) + 2\, cov(X_1, X_2)$$

où la covariance $cov(X_1, X_2)$ de X_1 et X_2 est donnée par

$$cov(X_1, X_2) \;=\; \mathbb{E}(X_1\, X_2) - \mathbb{E}(X_1)\, \mathbb{E}(X_2)$$

et

$$\mathbb{E}(X_1\, X_2) \;=\; \int_{\mathbb{R}^2} x_1\, x_2\, f(x_1, x_2) dx_1 dx_2.$$

En particulier, si X_1 et X_2 sont indépendantes alors $V(X_1 + X_2) = V(X_1) + V(X_2)$.

2.6 Variables aléatoires à densité - Lois au programme

Exemple (Loi uniforme)**.** La loi uniforme sur un segment $[a,b]$ de $\mathbb{R}$ (pour deux réels a et b tels que $a<b$) modélise le choix "au hasard" d'un nombre réel entre a et b.
Une variable aléatoire réelle X sur $(\Omega,\mathcal{T})$ suit la loi uniforme sur $[a,b]$ si sa densité est donnée par

$$f_X(t) \;=\; \begin{cases} \frac{1}{b-a} & \text{si } t\in[a,b], \\ 0 & \text{sinon.} \end{cases}$$

Dans ce cas, X a pour espérance $\mathbb{E}(X)=\frac{a+b}{2}$ et pour variance $V(X)=\frac{(b-a)^2}{12}$.

Exemple (Loi exponentielle)**.** Soit $\lambda>0$. La loi exponentielle de paramètre λ modélise la durée de vie d'un phénomène qui a la propriété suivante : pour tout $h>0$, la probabilité que le phénomène "meure" entre t et $t+h$ ne dépend pas de l'instant t. En termes probabilistes, pour une v.a. X qui suite cette loi, cela s'écrit :

$$\forall t,h\geq 0, \qquad P_{(X\geq t)}(X\leq t+h)=P_{(X\geq 0)}(X\leq h)\,.$$

Une variable aléatoire réelle X sur $(\Omega,\mathcal{T})$ suit la loi exponentielle de paramètre λ si sa densité est donnée par

$$\forall t\in\mathbb{R}, \qquad f_X(t) \;=\; \begin{cases} 0 & \text{si } t<0, \\ \lambda\, e^{-\lambda t} & \text{si } t\geq 0. \end{cases}$$

Dans ce cas, X a pour espérance $\mathbb{E}(X)=\frac{1}{\lambda}$ et pour variance $V(X)=\frac{1}{\lambda^2}$.

Exemple (Loi normale (ou de Laplace-Gauss))**.** Soit $\mu\in\mathbb{R}$ et $\sigma>0$. Une variable aléatoire réelle X sur $(\Omega,\mathcal{T})$ suit la loi normale de paramètres μ et σ (ou suit $\mathcal{N}(\mu,\sigma)$) si sa densité est donnée par

$$\forall t\in\mathbb{R}, \qquad f_X(t) \;=\; \frac{1}{\sqrt{2\,\pi\,\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}.$$

Dans ce cas, X a pour espérance $\mathbb{E}(X)=\mu$ et pour variance $V(X)=\sigma^2$.
Une variable aléatoire réelle X sur $(\Omega,\mathcal{T})$ suit la loi normale centrée réduite suit X suit $\mathcal{N}(0,1)$.
Si une variable aléatoire réelle X sur $(\Omega,\mathcal{T})$ suit la loi $\mathcal{N}(\mu,\sigma)$ alors $a\,X+b$ suit la loi $\mathcal{N}(a\,\mu+b,|a|\,\sigma)$. En particulier, $\frac{X-\mu}{\sigma}$ suit la loi normale centrée réduite.

3 Convergence des variables aléatoires

Les résultats de cette partie sont valables aussi bien pour les variables aléatoires réelles discrètes que pour les variables aléatoires réelles à densité.

3.1. Proposition - Inégalité de Markov.

Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{T}, P)$. On suppose que X a une espérance, notée $\mathbb{E}(X)$. On suppose que X est à valeurs positives. Alors pour tout $\alpha > 0$ on a

$$P(X \geq \alpha) \leq \frac{\mathbb{E}(X)}{\alpha}.$$

Remarque. Dans le cas où X est une variable aléatoire réelle ayant pour densité f, le fait d'être à valeurs positives signifie que f est nulle sur $]-\infty, 0[$.

3.2. Proposition - Inégalité de Bienaymé-Tchebychev.

Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{T}, P)$. On suppose que X a une espérance, notée $\mathbb{E}(X)$. Alors pour tout $\alpha > 0$ on a

$$P(|X - \mathbb{E}(X)| \geq \alpha) \leq \frac{1}{\alpha^2} \mathbb{E}\left[\Big(X - \mathbb{E}(X)\Big)^2\right].$$

Définition. *Une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires réelles sur $(\Omega, \mathcal{T}, P)$* ***converge en probabilité*** *vers une variable aléatoire réelle X si pour tout $\varepsilon > 0$ on a*

$$\lim_{n\to+\infty} P(|X_n - X| < \varepsilon) = 1.$$

3.3. Théorème - Loi faible des grands nombres.

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles sur $(\Omega, \mathcal{T}, P)$.
On suppose que les X_n sont mutuellement indépendantes et de même loi. On suppose également qu'elles admettent une espérance, notée μ, et une variance, notée σ^2.
Alors, la suite $(\overline{X}_n)_{n\in\mathbb{N}}$ des moyennes $\overline{X}_n := \frac{1}{n}(X_1 + ... + X_n)$ converge en probabilité vers μ, c'est-à-dire

$$\forall \varepsilon > 0, \qquad \lim_{n\to+\infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1.$$

Les trois résultats siuvants sont des résultats d'approximation, qui pourraient s'exprimer en terme de convergence en loi (notion hors programme)

3.4. Propriété - Loi hypergéométrique et loi binomiale.

Soit $p \in]0,1[$ et $n \in \mathbb{N}^*$. On suppose que la variable aléatoire X suit la loi binomiale de paramètres n et p. Pour tout $N \geq n$ on considère une variable aléatoire Y_N qui suit la loi hypergéométrique de paramètres N, p et n. Alors pour tout $k \in \{0,\ldots,n\}$ on a

$$\lim_{N\to+\infty} P(Y_N = k) \;=\; P(X = k).$$

Dans la pratique, on utilise l'approximation $P(Y_N = k) \simeq P(X = k)$ dès que $N \geq 10\,n$.

Remarque. Intuitivement, cela vient du fait que tirer n individus parmi N avec remise est plus ou moins équivalent à tirer n individus parmi N sans remise lorsque N est grand, car la probabilité de tirer au moins deux fois le même individu (au cours de n expériences) dans une grande population est faible.

3.5. Propriété - Loi binomiale et loi de Poisson.

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles. On suppose que pour tout $n \geq 1$, la variable aléatoire X_n suit la loi binomiale de paramètres n et p_n, et que

$$\lim_{n\to+\infty} n\,p_n \;=\; \lambda.$$

Alors pour tout $k \in \mathbb{N}$ on a

$$\lim_{n\to+\infty} P(X_n = k) \;=\; \frac{\lambda^k}{k!}\,e^{-\lambda}.$$

Dans la pratique, on utilise l'approximation $P(X_n = k) \simeq \frac{\lambda^k}{k!}\,e^{-\lambda}$ avec $\lambda = n\,p_n$ dès que $n \geq 30$, et $n\,p_n \leq 5$.

3.6. Propriété - Loi binomiale et loi normale.

Soit $p \in]0,1[$ et soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles telle que pour tout $n \geq 1$ la variable aléatoire X_n suit la loi binomiale de paramètres n et p.

Pour tout $n \geq 1$ on pose $Z_n := \frac{X_n - n\,p}{\sqrt{n\,p\,(1-p)}}$. Alors pour tout intervalle I de $\mathbb{R}$ on a

$$\lim_{n\to+\infty} P\left(Z_n \in I\right) \;=\; \frac{1}{\sqrt{2\,\pi}} \int_I e^{-\frac{x^2}{2}}\,dx.$$

Dans la pratique, on utilise cette approximation dès que $n \geq 30$, $n\,p \geq 10$ et $n\,(1-p) \geq 10$.

On énonce enfin un résultat fondamental en probabilités, dont la Propriété 3.6 ci-dessus est un cas particulier.

3.7. Théorème - Le théorème Central Limite [Admis].

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles sur $(\Omega, \mathcal{T}, P)$.
On suppose que les X_n sont mutuellement indépendantes et de même loi. On suppose également qu'elles admettent une espérance, notée μ, et une variance, notée σ^2.
Alors, la suite $(Z_n)_{n\in\mathbb{N}}$ donnée par

$$Z_n := \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$$

vérifie

$$\lim_{n\to+\infty} P(Z_n \in I) \; = \; \frac{1}{\sqrt{2\,\pi}} \int_I e^{-\frac{x^2}{2}}\, dx$$

pour tout intervalle I de $\mathbb{R}$.