Table des matières

Première partie
Probabilités

1 Espaces probabilisés dénombrables

Soit $\Omega = (\omega_n)_{n\geqslant 1}$ un ensemble dénombrable, et soit $f : \Omega \to \mathbb{R}^+$ une fonction.

$$\lim_{n\to+\infty} \sum_{\omega\in F} f(\omega) = \sup\{\sum_{\omega\in F} f(\omega); F \subset \Omega \; et \; F \; fini\} = \sum_{\omega\in\Omega} f(\omega)$$

Définition :
On appelle probablité sur Ω une fonction P définie sur $\mathcal{P}(\Omega)$ et à valeur dans [0,1] tel que :

1. $P(\Omega) = 1$
2. Pour toute famille dénombrable $\mathcal{A}$ de partie de Ω 2 à 2 incompatibles, on a :
$$P(\bigcup_{A\in\mathcal{A}} A) = \sum_{A\in\mathcal{A}} P(A)$$

Proposition :
Soit $\mathbb{P}$ une fonction définie sur $\mathcal{P}(\Omega)$ à valeur dans $\mathbb{R}^+$. $\mathbb{P}$ vérifie la σ-additivité si et seulement si :

1. $P(A \cup B) = P(A) + P(B)$ pour tout A et B incompatibles
2. Pour toute suite $(A_n)_{n\geqslant 1}$ croissante de parties de Ω,
$$\mathbb{P}(\bigcup_{n=1}^{+\infty} A_n) = \lim_{n\to+\infty} \mathbb{P}(A_n)$$

Demonstration :
$\sigma \Rightarrow 1$: évident $\sigma \Rightarrow 2$: Soit $(A_n)_{n\geqslant 1}$ croissante. Posons $B_{k+1} = A_{k+1} \backslash A_k$ $\bigcup_{n\geqslant 1} A_n = \bigcup_{n\geqslant 1} B_n$ et les $(A_n)_{n\geqslant 1}$ sont 2 à 2 incompatibles. D'après la propriété de σ-additivité,

$$\begin{aligned} P(\bigcup_{n\geqslant 1} A_n) &= P(\bigcup_{n\geqslant 1} B_n) \\ &= \lim_{n\to+\infty} \sum_{p=0}^{n} P(B_n) \\ &= \lim_{n\to+\infty} P(A_0) + P(A_n) - P(A_0) \\ &= \lim_{n\to+\infty} P(A_n) \end{aligned}$$

1 et 2 $\Rightarrow \sigma$: Soit $(A_n)_{n\geqslant 1}$ 2 à 2 incomptaibles. Posons $B_n = \bigcup_{k=1}^{n} A_k$ et $\bigcup_{n\geqslant 1} A_n = \bigcup_{n\geqslant 1} B_n$.

$$\begin{aligned} P(\bigcup_{n\geqslant 1} A_n) &= P(\bigcup_{n\geqslant 1} B_n) = \lim_{n\to+\infty} P(B_n) = \lim_{n\to+\infty} P(\bigcup_{k=1}^{n} A_k) \\ &= \lim_{n\to+\infty} \sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{+\infty} P(A_k) \end{aligned}$$

Définition :
On appelle espace dénombrable probabilisé un couple $(\Omega, \mathbb{P})$ où Ω est un ensemble dénombrable non vide et $\mathbb{P}$ une mesure de probabilité sur Ω.

Proposition :
Les proproétés 1 à 7 vues dans le cas fini restent vraies. De plus, pour toute famille $\mathcal{A}$ dénombrable d'évènement, on a

$$\mathbb{P}(\bigcup_{A\in\mathcal{A}} A) \leqslant \sum_{A\in\mathcal{A}} \mathbb{P}(A)$$

Remarque : La notion de densité discrète se généralise au cas où Ω est infini dénombrable.

Lois usuelles sur $\mathbb{N}$

Loi géométrique Pour tout $p \in]0,1[$, la fonction f définie pour tout $k \in \mathbb{N}^*$ par

$$f(k) = (1-p)^{k-1}p$$

est une densité de probabilité sur $\mathbb{N}^*$ On appelle la probabilité sur $\mathbb{N}^*$ associée, la loi géométrique de paramètre p et on la note $\mathfrak{G}(p)$.

Loi binomiale négative Pour tout $p \in]0,1[$ et $r \in \mathbb{N}^*$, la fonction f définie pour tout $k \in \mathbb{N}^*$ par

$$f(k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r$$

est une densité de probabilité sur $\mathbb{N}^*$ On appelle la probabilité sur $\mathbb{N}^*$ associée, la loi binomiale négative de paramètre r et p et on la note $\mathfrak{BN}(r,p)$.

Remarque : $\mathfrak{BN}(1,p) = \mathfrak{G}(p)$ et f(k)=0 si r>k. Ici, on ne s'intéresse pas au 1$^{\text{er}}$ succès mais au r-ième succès.

Loi de Poisson Pour tout λ>0, la fonction f définie pour tout $k \in \mathbb{N}$ par

$$f(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

est une densité de probabilité sur $\mathbb{N}$ On appelle la probabilité sur $\mathbb{N}$ associée, la loi géométrique de paramètre λ et on la note $\mathfrak{P}(\lambda)$.

Théorème :
Soit $(p_n)_{n\geqslant 1}$ une suite d'éléments de [0,1] tel que

$$\lim_{n\to+\infty} np_n = \lambda$$

Alors

$$\lim_{n\to+\infty} \binom{n}{k} p_n^k(1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}$$

Demonstration :
$np_n = \lambda u_n$ avec $\lim_{n\to+\infty} u_n = 1$. Soit $k \in \mathbb{N}$ fixé.

$$\binom{n}{k} p_n^k(1-p_n)^{n-k} = \frac{n!}{k!(n-k)!} \times \frac{1}{n^k}(\lambda u_n)^k \left(1-\lambda\frac{u_n}{n}\right)^{n-k}$$

Or, $\left(1-\frac{x}{n}\right)^n \to e^x$ et $\frac{n!}{(n-k)!}\frac{1}{n^k} \to 1$ d'où le résultat.

En pratique, on conviendra que l'approximation d'une binomiale par une loi de Poisson de paramètre np est correcte pour $n \geqslant 50$, $n \leqslant 0,01$ et $np \leqslant 10$.

2 Espaces probabilisés généraux

Définition :
Soit Ω un ensemble quelconque non vide. Une tribu $\mathfrak{F}$ sur Ω est un ensemble de parties de Ω vérifiant :

1. $\Omega \in \mathfrak{F}$
2. $\mathfrak{F}$ stable par complémentaire
3. $\mathfrak{F}$ stable par union dénombrable

Le couple $(\Omega, \mathfrak{F})$ est appelé espace probabilisable. Un élément de $\mathfrak{F}$ est appelé évenement.

Définition :
On appelle tribu borélienne de $\mathbb{R}^d$ et on note $\mathfrak{B}(\mathbb{R}^d)$ la tribu engendrée par l'ensemble des ouverts de $\mathbb{R}^d$, c'est-à-dire la plus petite tribu qui contient l'ensemble des ouverts de $\mathbb{R}^d$.

Définition :
On appelle mesure de probabilité ou loi de probabilité sur l'espace mesurable (ou probabilisable) (Ω, F) une application définie sur $\mathcal{F}$ et à valurs dans $\mathbb{R}^+$ tel que :

1. $\mathbb{P}(A) \in [0,1] \forall A \in \mathcal{F}$
2. $\mathbb{P}(\Omega) = 1$
3. Pour toute famille $\mathcal{A}$ dénombrable d'éléments deux à deux incompatibles de $\mathcal{F}$, on a

$$\mathbb{P}(\bigcup_{A \in \mathcal{A}} A) = \sum_{A \in \mathcal{A}} \mathbb{P}(A)$$

Définition :
On appelle densité de probabilité (continue) sur $\mathbb{R}$ (ou sur $\mathbb{R}^d$) une fonction f continue par morceaux à valeurs positives et tel que

$$\int_{\mathbb{R}} f(x)dx = 1$$

(ou $\int_{\mathbb{R}^d} f(x)dx = 1$).

Théorème :
Si f est une densité de probabilité sur $\mathbb{R}$ (ou $\mathbb{R}^d$) alors il existe une unique mesure de probabilité $\mathbb{P}$ sur $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ (ou sur $(\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d))$ tel que pour tout intervalle]a,b[(ou tout cylindre $]a_1, b_1[\times ... \times]a_n, b_n[$) on ait :

$$\mathbb{P}(]a,b[) = \int_a^b f(x)dx \ ou \ \mathbb{P}(]a_1, b_1[\times ... \times]a_n, b_n[) = \int_{a_1}^{b_1} ... \int_{a_n}^{b_n} f(x)dx$$

Exemples de lois continues

Loi normale $\mathcal{N}$(0,1) X$\hookrightarrow \mathcal{N}(0,1)$

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \\ F_X(x) &= \int_{-\infty}^{+\infty} f_X(t)dt \\ \mathbb{P}(a \leq X \leq b) &= \int_a^b f_X(t)dt \end{aligned}$$

Loi normale $\mathcal{N}(\mu, \sigma^2)$ X$\hookrightarrow \mathcal{N}(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Loi exponentielle $\mathcal{E}(\lambda)$ X$\hookrightarrow \mathcal{E}(\lambda)$

$$f_X(x) = \lambda e^{-\lambda x}\ 1_{\mathbb{R}_+^*}(x)$$

Loi uniforme sur l'intervalle]a,b[X$\hookrightarrow \mathcal{U}(]a,b[)$

$$f_X(x) = \frac{1}{b-a}\ 1_{]a,b[}(x)$$

Loi uniforme sur un borélien $\mathfrak{B}$ de $\mathbb{R}^2$ X$\hookrightarrow \mathcal{U}(B)$ avec $B \in \mathfrak{B}(\mathbb{R}^2)$ fixé
$\forall A \in \mathfrak{B}(\mathbb{R}^2)$, $\mathbb{P}(X \in A) = \frac{\lambda_2(A\cap B)}{\lambda_2(B)}$

$$f_X(x) = \frac{1}{\lambda_2(B)}\ 1_B(x)$$

Autres lois : Gamma, Bêta, de Student, du χ^2, de Fisher...

Définition :
On appelle v.a. réelle (ou d-dimensionnelle) à valeur dans un borélien E une application mesurable X définie sur Ω à valeurs dans E, ie

$$\forall A \in \mathfrak{B}(E)\{x \in A\} = X^{-1}(A) \in \mathcal{F}$$

Si E=$\mathbb{R}$, on dit que X est une v.a. réelle
Si E=$\mathbb{R}^d$, on dit que X est un vecteur aléatoire réel

Proposition :

1. La somme et le produit d'un nombre fini de v.a.r. est une v.a.r.
2. Si $(X_n)_{n\geq 1}$ est une suite de v.a.r. tel que $X_n(\omega) \xrightarrow[n\to+\infty]{} X(\omega)$ pour tout $\omega \in \Omega$ alors l'application X est une v.a.r

Définition :
Soit X une v.a.r. à valeur dans $E \in \mathfrak{B}(\mathbb{R})$. L'application Q définie pour tout $B \in \mathfrak{B}(\mathbb{R})$ par

$$Q(B) = \mathbb{P}(X \in B)$$

est une mesure de probabilité. On l'appelle loi de probabilité sur E de la v.a. X.

Loi exponentielle Soit λ>0. On dit qu'une var X suit la loi exponentielle de paramètre λ>0 si

$$\forall 0 < a < b, \mathbb{P}(X \in [a,b]) = \int_a^b \lambda e^{-\lambda x}$$

On note alors $X \hookrightarrow \mathcal{E}(\lambda)$

Remarque : $\mathbb{P}(X \in]0,+\infty[) = 1$ et donc nécessairement, $\mathbb{P}(X \leq 0) = 0$.

Proposition :
Si $X \hookrightarrow \mathcal{E}(\lambda)$ avec λ >0, alors $aX \hookrightarrow \mathcal{E}(\frac{\lambda}{a})$, $\forall a > 0$. En particulier, $\lambda X \hookrightarrow \mathcal{E}(1)$

Demonstration :
La fonction de répartition de $\mathcal{E}(\lambda)$ est

$$F_X(x) = \left(1 - e^{-\lambda x}\right) 1_{\mathbb{R}_+^*}(x)$$

Posons Y=aX.

$$\begin{aligned} F_Y(x) &= \mathbb{P}(Y \leq x) \\ &= \mathbb{P}\left(X \leq \frac{x}{a}\right) \\ &= F_X\left(\frac{x}{a}\right) \\ &= \left(1 - e^{-\frac{\lambda}{a}x}\right) 1_{\mathbb{R}_+^*}(x) \end{aligned}$$

Il s'agit de la fonction de répartition de la loi $\mathcal{E}(\frac{\lambda}{a})$

Proposition :
Pour rappel, [X]=E(X) (partie entière de X).
Si $X \hookrightarrow \mathcal{E}(\theta)$, alors $[X] + 1 \hookrightarrow \mathcal{G}(1 - e^{-\theta})$

Demonstration :
On pose Z=[X]+1. On a $Z(\Omega) = \mathbb{N}^*$.
Soit $k \in \mathbb{N}^*$.

$$\begin{aligned} \mathbb{P}(Z = k) &= \mathbb{P}([X] = k-1) \\ &= \mathbb{P}(k-1 \leq X < k) \\ &= \int_{k-1}^{k} \theta e^{-\theta x} dx \\ &= \left[-e^{-\theta x}\right]_{k-1}^{k} \\ &= e^{-(k-1)\theta} - e^{-k\theta} \\ &= e^{-\theta(k-1)}\left(1 - e^{-\theta}\right) \end{aligned}$$

En posant $p = 1 - e^{-\theta}$, $\mathbb{P}(Z = k) = (1-p)^{k-1}p$.
Donc $[X] + 1 \hookrightarrow \mathcal{G}(1 - e^{-\theta})$.

Proposition :
De la même manière, on peut montrer que :
Soit $X \hookrightarrow \mathcal{E}(\theta)$ et $(N_n)_{n\geq 0} \in \left(\mathbb{R}_+^*\right)^{\mathbb{N}}$ tel que $N_n \xrightarrow[n\to+\infty]{} +\infty$.

1. Pour tout $n \geq 0$, $[N_n X] \hookrightarrow \mathcal{G}\left(1 - e^{-\frac{\theta}{N_n}}\right)$
2. $\lim_{n\to+\infty} \frac{[N_n X]}{N_n} = X$ de façon presque sûr (ie $\mathbb{P}\left(\left\{\omega \in \Omega | \lim_{n\to+\infty} \frac{[N_n X](\omega)}{N_n} = X(\omega)\right\}\right) = 1$)

Théorème :
Soit X une v.a. à valeurs dans $\mathbb{R}^+$. Les propriétés suivantes sont équivalentes :

1. X est une v.a. de loi exponentielle
2. Pour tout réel t>0, $\mathbb{P}(X > t) \neq 0$ et

$$\forall s > 0, \mathbb{P}(X > t + s | X > t) = P(X > s)$$

(Propriété sans mémoire)

Demonstration (de la première implication) :
t>0 et s >0

$$\begin{aligned}
\mathbb{P}(X > t + S | X > t) &= \frac{\mathbb{P}\left((X > t+s) \cap (X > t)\right)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > t)} \\
&= \frac{1 - F_X(s+t)}{1 - F_X(x)} \\
&= \frac{e^{-\theta(s+t)}}{e^{-}\theta t} = e^{-\theta s} \\
&= 1 - F_X(s) \\
&= \mathbb{P}(X > s)
\end{aligned}$$

Théorème :
Soit X une v.a. à valeurs dans $\mathbb{N}^*$. Il t a équivalence entre :

1. X est une v.a. de loi géométrique
2. $\forall n \in \mathbb{N}$, $\mathbb{P}(X > n) \neq 0$, et :

$$\forall k \in \mathbb{N}, \mathbb{P}(X > n + k | X > n) = \mathbb{P}(X > k)$$

3 Notion d'indépendance

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

Définition :

1. On dit que deux évenements A et B sont indépendants s'ils vérifient $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
2. Soit $(A_i)_{i \in I}$ une famille quelconque d'évenements. On dit que les $(A_i)_{i \in I}$ sont mutuellement indépendants si pour tout ensemble fini d'indices distincts $\{i_1, ..., i_n\} \subset I$, on a

$$\mathbb{P}(A_{i_1} \cap ... \cap A_{i_k}) = \mathbb{P}(A_{i_1})...\mathbb{P}(A_{i_k})$$

Soient $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ deux espaces probabilisés. Considérons l'espace probabilisable $(\Omega, \mathcal{F})$ où $\Omega = \Omega_1 \times \Omega_2$ et $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ (produit tensoriel) est la tribu engendrée par les parties de Ω de la forme $A_1 \times A_2$ avec $A_1 \in \mathcal{F}_1$ et $A_2 \in \mathcal{F}_2$.

Proposition :
$\exists ! \mathbb{P}$ sur $(\Omega, \mathcal{F}) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ tel que :

$$\forall A_1 \in \mathcal{F}_1, \forall A_2 \in \mathcal{F}_2, \mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$$

$\mathbb{P}$ s'appelle le produit tensoriel de $\mathbb{P}_1$ et $\mathbb{P}_2$ et on note $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$.

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On note X et Y deux variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ tel que X soit à valeur dans un espace mesurable (E,$\mathcal{E}$) et Y à valeur dans un espace mesurable (F,$\mathcal{F}$).

Définition :
On dit que X et Y sont indépendantes (et on note X$\perp\!\!\!\perp$Y) si pour tout $A \in \mathcal{E}$ et tout $B \in \mathcal{F}$ les évenements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants.
De même, soit $(X_i)_{i \in I}$ une famille de v.a. dans les espaces mesurables $((E_i, \mathcal{E}_i))_{i \in I}$. On dit que les $(X_i)_{i \in I}$ sont indépendantes entre elles si pour toute famille $(B_i)_{i \in I}$ d'éléments de $(\mathcal{E}_i)_{i \in I}$, les éléments $\{X_i \in B_i\}_{i \in I}$ sont mutuellement indépendants.

Proposition :

1. X$\perp\!\!\!\perp$Y $\Rightarrow$ f(X)$\perp\!\!\!\perp$g(Y) pour tout f et g mesurables
2. Si la loi du couple (X,Y) admet une densité, alors

$$X \perp\!\!\!\perp Y \Leftrightarrow f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$$

4 Fonction de répartition

4.1 Définition générale

Définition :

- Pour toute mesure de probabilité μ sur $(\mathbb{R}, B_{\mathbb{R}})$, on apelle fonction de répartition de μ, notée F_μ, et définie pour tout $x \in \mathbb{R}$ par :
$$F_\mu(x) = \mu(]-\infty, x])$$
- Soit X une v.a. réelle définie sur $(\Omega, \mathcal{F}, \mathbb{P})$. On parlera alors de fonction de répartition de la v.a. X au lieu de la fonction de répartition de la loi μ_X de X. On notera $F_\mu = F_X$ et on aura donc :
$$F_X(x) = \mu_X(]-\infty, x]) = \mathbb{P}(X \leq x)$$
- Plus généralement, si X=$(X_1, ..., X_d)$ est un vecteur aléatoire de $\mathbb{R}$, alors la fonction de répartition F_X de X est définie pour tout x=$(x_1, ..., x_d) \in \mathbb{R}^d$ par
$$F_X(x) = \mathbb{P}(X \leq x)\mathbb{P}(X_1 < x_1, ..., X_d < x_d)$$

Théorème (admis) :
Deux mesures de probabilité sur $(\mathbb{R}, B_{\mathbb{R}})$ sont égales ssi elles ont même fonction de répartition.

Proposition (de la f.d.r) :
Soit X une v.a. définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et soit F_X sa f.d.r.

- F_X est croissante tel que $\lim_{x\to+\infty} F_X(x) = 1$ et $\lim_{x\to-\infty} F_X(x) = 0$
- F_X est "Cadlag" (continue à droite et limité à gauche) et pour tout $\alpha \in \mathbb{R}$ on a :

$$F_x(\alpha^-) = \lim_{\substack{x\to\alpha \\ x<\alpha}} F_X(x) = \mathbb{P}(X < \alpha)$$

-

$$\forall \alpha \in \mathbb{R}, \mathbb{P}(X = \alpha) = F_X(\alpha) - F_X(\alpha^-)$$

Demonstration (de la dernière propriété) :

$$\begin{aligned}
F_X(\alpha) - F_X(\alpha^-) &= \lim_{n\to+\infty} \mathbb{P}(X \leq \alpha) - \mathbb{P}(X \leq \alpha - \frac{1}{n}) \\
&= \lim_{n\to+\infty} \mathbb{P}(\alpha - \frac{1}{n} < X \leq \alpha) \\
&= \lim_{n\to+\infty} \mathbb{P}(X \in]\alpha - \frac{1}{n}, \alpha]) \\
&= \mathbb{P}(X \in \bigcap_n \alpha - \frac{1}{n}, \alpha]) \\
&= \mathbb{P}(X = a)
\end{aligned}$$

Proposition :
Soit X une variable aléatoire réelle de densité f_X. La f.d.r. F_X de X vérifie :

- $\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^{x} f_X(t)dt$
- F_X continue sur $\mathbb{R}$
- Si f_X est continue en $x_0 \in \mathbb{R}$ alors F_X est dérivable et $F'(x_0) = f(x_0)$

Proposition :
On suppose que la f.d.r. F_X de la var X est $\mathcal{C}^1$ par morceaux au sens suivant :

- F_X continue sur $\mathbb{R}$ sauf éventuellement en un nombre fini de points $a_1 < a_2 < ... < a_n$.
- Sur chacun des des intervalles $]-\infty, a_1[,]a_n, +\infty[$ et $]a_i, a_{i+1}[$ pour tout $1 \leq i \leq n-1$, la dérivée f de F_X est continue.

Alors X a pour densité f

4.2 Espérance et variance d'une var

Définition :
Soit X une var définie sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On appelle espérance de la v.a. X et on note E(X) l'intégrale (au sens de Lebesgue)

$$\int_\Omega X(\omega) d\mathbb{P}(\omega)$$

lorsque celle-ci est bien définie.
Si E(|X|) est un nombre fini, on dit que la v.a. X est intégrable.

Remarque :
Posons $X^+ = \sup(X, 0)$ et $X^- = \inf((-X), 0)$. ($X^+ \geq 0$ et $X^- \geq 0$ p.s.)
Les intégrales $\int_\Omega X^+ d\mathbb{P}$ et $\int_\Omega X^- d\mathbb{P}$ ont toujours un sens. Comme $X = X^+ + X^-$ on pose

$$\int_\Omega X d\mathbb{P} = \int_\Omega X^+ d\mathbb{P} + \int_\Omega X^- d\mathbb{P}$$

Théorème :

1. Si $X(\Omega)$ dénombrable, alors :

$$E(X) = \sum_{k \in X(\Omega)} k\mathbb{P}(X = k)$$

et pour toute fonction g définie sur $X(\Omega)$ à valeur dans $\mathbb{R}$:

$$E(g(X)) = \sum_{k \in X(\Omega)} g(k)\mathbb{P}(X = k)$$

2. Si la loi de X admet une densité f_X alors

$$E(X) = \int_{\mathbb{R}} x f_X(x) dx$$

et plus généralement, si g est une fonctions mesurable sur $\mathbb{R}$ à valeur dans $\mathbb{R}$ alors :

$$E(g(X)) = \int_{\mathbb{R}} g(x) f_X(x) dx$$

Proposition :
Soient X et Y deux v.a.r.

1. $\forall a, b \in \mathbb{R}, E(aX + bY) = aE(X) + bE(Y)$
2. Si $X \geq 0$ p.s. alors $E(X) \geq 0$
3. Si $X \geq Y$ p.s. alors $E(X) \geq E(Y)$
4. $|E(X)| \leq E(|X|)$ et plus généralement, si ϕ est une fonction convexe, alors

$$\phi(E(X)) \leq E(\phi(X))$$

(inégalité de Jensen)

Théorème (de la convergence dominée de Lebesgue) :
Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. qui converge p.s. vers une v.a.r. X.
S'il existe une v.a. Y intégrable tel que $|X_n| \leq Y$ p.s. $\forall n \geq 1$ alors

$$E(X) = \lim_{n \to +\infty} E(X_n)$$

Proposition (Inégalité de Markov) :
Soit X est une v.a.r. définie sur $(\Omega, \mathcal{F}, \mathbb{P})$
Si $X \geq 0$ p.s. alors

$$\forall \lambda > 0, \mathbb{P}(X > \lambda) \leq \frac{E(X)}{\lambda}$$

Demonstration :

$$\forall A \in \mathcal{F},\ \mathbb{P}(A) = \int_A d\mathbb{P} = \int 1_A d\mathbb{P} = E(1_A)$$

$$\begin{aligned}
\mathbb{P}(X \leq \lambda) &= E(1_{\{X \leq \lambda\}}) = \int_\Omega 1_{\{X \leq \lambda\}} d\mathbb{P} \\
&\leq \int_\Omega \frac{X}{\lambda} 1_{\{X \leq \lambda\}} d\mathbb{P} \\
&\leq \int_\Omega \frac{X}{\lambda} d\mathbb{P} (car\ X \geq 0\ p.s.) \\
&\leq \frac{1}{\lambda} \int_\Omega X d\mathbb{P} \\
&\leq \frac{E(X)}{\lambda}
\end{aligned}$$

Définition :
Pour toute v.a.r. X, on appelle variance de X le nombre (s'il existe)

$$V(X) = E\left((X - E(X))^2\right) = E\left(X^2\right) - (E(X))^2$$

Proposition (Inégalité de Bienaymé-Tcheychev) :

$$\forall \lambda > 0, \mathbb{P}(|X - E(X)| > \lambda) \leq \frac{V(X)}{\lambda^2}$$

Demonstration :

$$\begin{aligned} \mathbb{P}(|X - E(X)| > \lambda) &= \mathbb{P}\left((X - E(X))^2 > \lambda^2\right) \\ &\leq \frac{E\left((X - E(X))^2\right)}{\lambda^2} \\ &\leq \frac{V(X)}{\lambda^2} \end{aligned}$$

Proposition : – $V(X) \geq 0$ (car $E(X)^2 \leq E(X^2)$)

- Si la loi de X dmet une densité d_x alors

$$V(X) = \int_{\mathbb{R}} x^2 f_X(x)dx - \left(\int_{\mathbb{R}} x f_X(x)dx\right)^2$$

- $\forall (a, b) \in \mathbb{R}^2,\ V(aX + b) = a^2 V(X)$
- Considérons la fonction

$$\begin{aligned} g : \mathbb{R} &\to \mathbb{R}^+ \\ a &\mapsto E((X - a)^2) \end{aligned}$$

  alors

$$\operatorname*{argmin}_{a \in \mathbb{R}} g(a) = E(X) \ et \ \min_{a \in \mathbb{R}} g(a) = V(X)$$

- $X \perp\!\!\!\perp Y \Rightarrow V(X + Y) = V(X) + V(Y)$

Exemples à connaître :

- $X \hookrightarrow \mathcal{B}(p)$ alors E(X)=p et V(X)=p(1-p)
- $X \hookrightarrow \mathcal{B}(p, n)$ alors E(X)=np et V(X)=np(1-p)
- $X \hookrightarrow \mathcal{G}(p)$ alors E(X)=$\frac{1}{p}$ et V(X)=$\frac{1-p}{p^2}$
- $X \hookrightarrow \mathcal{P}(\lambda)$ alors E(X)=V(X)=λ
- $X \hookrightarrow \mathcal{U}(]a, b[)$ alors E(X)=$\frac{a+b}{2}$ et V(X)=$\frac{(b-a)^2}{12}$
- $X \hookrightarrow \mathcal{E}(\theta)$ alors E(X)=$\frac{1}{\theta}$ et V(X)=$\frac{1}{\theta^2}$
- $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$ alors E(X)=μ et V(X)=σ^2

5 Convergence d'une variable aléatoire

5.1 Convergence en probabilité et presque sûr

Définition :
Soit $(Y_n)_{n\geq 1}$ une suite de v.a.r. et soit Y une v.a.r.

1. On dit que $(Y_n)_{n\geq 1}$ converge en probabilité vers Y si :

$$\forall \varepsilon > 0, \mathbb{P}(|Y_n - Y| \geq \epsilon) \xrightarrow[n\to+\infty]{} 0$$

   On note

$$Y_n \xrightarrow[n\to+\infty]{\mathbb{P}} Y$$

2. On dit que $(Y_n)_{n\geq 1}$ converge presque-sûrement vers Y si :

$$\mathbb{P}(\{\omega \in \Omega | \lim_{n\to+\infty} Y_n(\omega) = Y(\omega)\}) = 1$$

   On note

$$Y_n \xrightarrow[n\to+\infty]{p.s.} Y$$

Proposition :
La convergence p.s. entraîne la convergence en probabilité.

Demonstration :
A reprendre

Proposition :
Soient Y et $(Y_n)_{n\geq 1}$ des v.a.r. telles que

$$\forall \varepsilon > 0, \sum_{n\geq 1} \mathbb{P}(|Y_n - Y| > \varepsilon) < +\infty$$

alors

$$Y_n \xrightarrow[n\to+\infty]{p.s.} Y$$

Demonstration :
Posons $B_{n,\varepsilon} = \{|Y_n - Y| > \varepsilon\}$ et $A_\varepsilon = \overline{\lim}_{n\to+\infty} B_{n,\varepsilon}$
D'après le lemme de Borel-Cantelli :

$$\mathbb{P}(A_\varepsilon) = 0 \ \forall \varepsilon > 0$$

Or, $A_\varepsilon = \bigcap_{k\geq 1} \bigcup_{k\geq n} B_{k,\varepsilon}$ et $\overline{A_\varepsilon} = \bigcup_{k\geq 1} \bigcap_{k\geq n} \overline{B_{k,\varepsilon}}$
On a $\mathbb{P}(\overline{A_\varepsilon}) = 1 \ \forall \varepsilon > 0$. Posons $E = \bigcap_{s\in\mathbb{N}^*} \overline{A_{\frac{1}{s}}}$

$$\mathbb{P}(\overline{E}) = \mathbb{P}(\bigcap_{s\in\mathbb{N}^*} A_{\frac{1}{s}}) \leq \sum_{s\in\mathbb{N}^*} \mathbb{P}(A_{\frac{1}{s}}) = 0$$

D'où $\mathbb{P}(E) = 1$

$$\begin{aligned}
\omega \in E & \Leftrightarrow \forall s \in \mathbb{N}^*, \omega \in \overline{A_{\frac{1}{s}}} \\
& \Leftrightarrow \forall s \in \mathbb{N}^*, \exists n > 1, \forall k \geq n, \omega \in B_{k,\frac{1}{s}} \\
& \Leftrightarrow \forall s \in \mathbb{N}^*, \exists n > 1, \forall k \geq n, |Y_k(\omega) - Y(\omega)| \leq \frac{1}{s} \\
& \Leftrightarrow \forall \varepsilon > 0, \exists n > 1, \forall k \geq n, |Y_k(\omega) - Y(\omega)| \leq \varepsilon \\
& \Leftrightarrow Y_k \xrightarrow[k\to+\infty]{} Y
\end{aligned}$$

5.2 Covariance de deux variables aléatoires réelles

Définition :
Soient X et Y deux v.a.r.
La covariance de X et Y, notée cov(X,Y) est définie par :

$$\begin{aligned}\text{cov(X,Y)} &= E((X - E(X)(Y - E(Y)) \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

Si cov(X,Y)=0, on dit que X et Y sont non corrélées.

Théorème (Inégalité de Cauchy-Schwarz) :
Soient X et Y deux v.a.r. On a :

$$|\text{cov(X,Y)}| \leq \sqrt{V(X)}\sqrt{V(Y)}$$

et

$$\text{cov}(X,Y)^2 = V(X)V(Y) \Leftrightarrow \exists(\alpha,\beta,\gamma) \neq (0,0,0); \alpha X + \beta Y = \gamma \; p.s.$$

Demonstration :
Considérons le prolynôme P défini pour tout $\lambda \in \mathbb{R}$ par :

$$P(\lambda) = V(X + \lambda Y) = V(X) + \lambda^2 V(Y) + 2\lambda \text{cov}(X,Y) \geq 0$$

Par conséquent :

$$\begin{aligned}\Delta &= 4\text{cov}^2(X,Y) - 4V(X)V(Y) \leq 0 \\ &\Leftrightarrow \text{cov}^2(X,Y) \leq V(X)V(Y) \\ &\Leftrightarrow |\text{cov}(X,Y)| \leq \sqrt{V(X)}\sqrt{V(Y)}\end{aligned}$$

Supposons : $\exists(\alpha,\beta,\gamma) \neq (0,0,0); \alpha X + \beta Y = \gamma$ p.s. On peut supposer $\alpha \neq 0$

$$\begin{aligned}X &= \frac{\gamma}{\alpha} - \frac{\beta}{\alpha}Y \\ \text{Cov}(X,Y) &= \text{Cov}(\frac{\gamma}{\alpha}, Y) - \frac{\beta}{\alpha}\text{cov}(Y,Y) \\ &= 0 - \frac{\beta}{\alpha}V(Y) \\ \Rightarrow \text{Cov}^2(X,Y) &= \frac{\beta^2}{\alpha^2}V(Y) = V\left(\frac{\beta}{\alpha}Y\right)V(Y) \\ &= V\left(\frac{\gamma}{\alpha} - \frac{\beta}{\alpha}Y\right)V(Y) \\ &= V(X)V(Y)\end{aligned}$$

Réciproquement, si $\text{cov}(X,Y)^2 = V(X)V(Y)$ alors $\Delta = 0$ et P admet une racine réelle (double) λ_0

$$\begin{aligned}& P(\lambda_0) = V(X + \lambda_0 Y) = 0 \\ \Leftrightarrow \; & E\left(((X + \lambda_0 Y) - E((X + \lambda_0 Y))^2\right) = 0 \\ \Leftrightarrow \; & X + \lambda_0 Y = E((X + \lambda_0 Y)\end{aligned}$$

Autrement dit, $X + \lambda_0 Y = c$ p.s.

Lemme :
Si $X \geq 0$ p.s. tel que E(X)=0 alors X=0 p.s.

Demonstration :
D'après l'inégalité de Markov :

$$\begin{aligned}
& \forall \varepsilon > 0, 0 \leq \mathbb{P}(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon} = 0 \\
\Rightarrow \quad & \forall n \in \mathbb{N}^*, \mathbb{P}\left(X \geq \frac{1}{n}\right) = 0 \\
\Rightarrow \quad & \mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} \left\{X \geq \frac{1}{n}\right\}\right) \leq \sum_{n \geq 1} \mathbb{P}\left(X \geq \frac{1}{n}\right) = 0 \\
\Rightarrow \quad & \mathbb{P}\left(\bigcap_{n \in \mathbb{N}^*} \left\{X \leq \frac{1}{n}\right\}\right) = 1 = \mathbb{P}(X = 0) \\
\Rightarrow \quad & X = 0 \ p.s.
\end{aligned}$$

Proposition :
1. cov(X,Y)=cov(Y,X)
2. cov(aX+bY,Z)=a cov(X,Z)+b cov(Y,Z), $\forall (a,b) \in \mathbb{R}^2$
3. cov(X,X)=V(X)
4. V(X+Y)=V(X)+V(Y)+2cov(X,Y)
5. X$\perp\!\!\!\perp$Y $\Rightarrow$ cov(X,Y)=0 (Réciproque fausse)

5.3 Les différentes lois des grands nombres

Théorème (Loi faible des grands nombres) :
Soit $(X_k)_{k \geq 1}$ une suite de v.a.r. de même loi tel que $E(X_1^2) < +\infty$ et deux à deux non corrélées. Alors

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow[n \to +\infty]{\mathbb{P}} E(X_1)$$

Demonstration :
Posons $S_n = \sum_{k=1}^{n} X_k$.

$$\frac{S_n}{n} \xrightarrow[n \to +\infty]{\mathbb{P}} E(X_1) \Leftrightarrow \forall \varepsilon > 0, \mathbb{P}\left(\left|\frac{S_n}{n} - E(X_1)\right| > \varepsilon\right) \xrightarrow[n \to +\infty]{} 0$$

Soit $\varepsilon > 0$ fixé. On a

$$E(S_n) = \sum_{k=1}^{n} E(X_k) = nE(X_1)$$

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n}{n} - E(X_1)\right| > \varepsilon\right) &= \mathbb{P}\left(|S_n - nE(X_1)| > n\varepsilon\right) \\
&\leq \frac{V(S_n)}{(n\varepsilon)^2}
\end{aligned}$$

Or,

$$\begin{aligned}
V(S_n) &= V(\sum_{k=1}^{n} X_k) \\
&= \sum_{k=1}^{n} V(X_k) \ (car\ X_k\ 2 \text{ à } 2\ non\ correlées) \\
&= nV(X_1) \ (car\ X_k\ identiquement\ distribuées)
\end{aligned}$$

$$\Rightarrow \mathbb{P}\left(\left|\frac{S_n}{n} - E(X_1)\right| > \varepsilon\right) \leq \frac{V(X_1)}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0$$

avec $V(X_1)$ fini car $E(X_1^2) < +\infty$. D'où

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{\mathbb{P}} E(X_1)$$

Théorème (loi forte des grands nombres - admis) :
Soit $(X_k)_{k\geq 1}$ une suite de v.a.r. i.i.d.

$$E(|X_1|) < +\infty \Rightarrow \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{p.s.} E(X_1)$$

5.4 Convergence en loi

Définition :
Soient Y et $(Y_n)_{n\geq 1}$ des v.a.r.
On dit que la suite $(Y_n)_{n\geq 1}$ converge en loi vers la v.a. Y si :

$$F_{Y_n}(x) \xrightarrow[n\to+\infty]{} F_Y(x)$$

pour tout point de continuité x de F_Y (avec F_X f.d.r. de la v.a. X)
On note alors :

$$Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y$$

Remarque : – Convergence p.s. $\Rightarrow$ Convergence en proba $\Rightarrow$ Convergence en loi
– La convergence en loi n'est pas stable pour la somme des variables aléatoires :

$$Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y \ et \ Z_n \xrightarrow[n\to+\infty]{\mathcal{L}} Z \not\Rightarrow Y_n + Z_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y + Z$$

Théorème (Central limit) :
Soit $(X_k)_{k\geq 1}$ une suie de v.a.r., iid et de carré intégrable (ie $E(X_1^2) < +\infty$). On a alors :

$$W_n \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu) \xrightarrow[n\to+\infty]{\mathcal{L}} N$$

où $\mu = E(X_1)$, $\sigma^2 = V(X_1) > 0$ et $N \hookrightarrow \mathcal{N}(0,1)$

On a donc :

$$\forall x \in \mathbb{R}, \mathbb{P}(W_n \leq x) \xrightarrow[n\to+\infty]{} \mathbb{P}(N \leq x) = F_N(x)$$

avec

$$F_n(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-t^2}{2}}\, dt$$

Ce qui équivaut à :

$$\forall a < b, \mathbb{P}(a \leq W_n \leq b) \xrightarrow[n\to+\infty]{} \sqrt{2\pi}\int_{a}^{b} e^{\frac{-t^2}{2}}\, dt$$

6 Vecteurs aléatoires

Définition :
Soit X=$^t(X_1, ..., X_d)$ un vecteur aléatoire de dimension $d \in \mathbb{N}$. On appelle espérance de X le vecteur

$$E(X) =^t (E(X_1), ..., E(X_d))$$

Proposition :

- Si X=$^t(X_1,...,X_d)$ et Y=$(Y_1,...,Y_d)$ sont deux vecteurs aléatoires de dimension d, alors
$$E(X+Y)=E(X)+E(Y)$$
- Si M est une matrice de nombres réels et si X est un vecteur aléatoire tel que MX soit bien défini, alors $E(MX)=M\times E(X)$
- Si $\phi:\mathbb{R}^d\to\mathbb{R}$ est convexe alors $\phi(E(X))\leq E(\phi(X))$ (Inégalité de Jensen)

Définition :
Soit X=$^t(X_1,...,X_d)$ tel que X_i soit de carré intégrable pour tout $1\leq i\leq d$. On appelle matrice de covariance (ou matrice de dispersion) du vecteur aléatoire X, la matrice :

$$\begin{aligned}V(X) &= E((X-E(X))^t(X-E(X)) \\ &= (cov(X_i,X_j))_{1\leq i,j\leq d}\end{aligned}$$

Proposition :

1. Si b=$(b_1,...,b_d)$ est un vecteur constant et si X=$(X_1,...,X_d)$ est un vecteur aléatoire alors $V(X+b)=V(X)$
2. Si M est une matrices de nombres réels tel que MX soit bien défini, alors
$$\begin{aligned}V(MX) &= E((MX-E(MX))^t(MX-E(MX))) \\ &= E(M(X-E(X))^t(M(X-E(X)))) \\ &= E(M(X-E(X))^t(X-E(X))^tM) \\ &= MV(X)^tM\end{aligned}$$
3. V(X) est une matrice symétrique semi-définie positive.
Autrement dit, pour tout vecteur Y non nul, on a $^tYV(X)Y\geq 0$, ou encore, toutes les valeurs propres de V(X) sont positives ou nulles.

Définition (Vecteurs alétoires gaussiens) :
On dit que le vecteur alétoire X=$^t(X_1,...,X_d)$ de $\mathbb{R}^d$ est gaussien si pour toute application linéaire $u:\mathbb{R}^d\to\mathbb{R}$, la variable aléatoire u(X) est une v.a. réelle gaussienne.

Remarque :
$\forall 1\leq i\leq d$, considérons l'application linéaire

$$\begin{aligned}u_i:\mathbb{R}^d &\to \mathbb{R} \\ (x_1,...,x_d) &\mapsto x_i\end{aligned}$$

Ainsi, $X_i=u_i(X)$, avec X vecteur aléatoire gaussien. Donc, par définition, X_i est une v.a. gaussienne sur $\mathbb{R}$.
La réciproque est fausse.

Proposition :
La loi d'un vecteur aléatoire gaussien X=$^t(X_1,..,X_d)$ est caractérisée par son vecteur espérance

$$m=^t(E(X_1),...,E(X_d))$$

et sa matrice de dispersion
$$\Gamma=(cov(X_i,X_j))_{1\leq i,j\leq d}$$
La loi de X est notée $\mathcal{N}_j(m,\Gamma)$

Proposition :
Soient X et Y deux v.a. réelles indépendantes dont les lois admettent des densités de probabilité f_X et f_Y respectivement.
Alors la loi de Z=X+Y admet également une densité de probabilité f_Z définie pour tout $x\in\mathbb{R}$ par

$$f_Z(x)=(f_X*f_Y)(x)=\int_{\mathbb{R}}f_X(t)f_Y(x-t)dt$$

Demonstration :
$X \perp\!\!\!\perp Y$, Z=X+Y.
Montrons que la loi de Z admet une densité de probabilité f_Z définie pour tout $x \in \mathbb{R}$ par

$$f_Z(x) = (f_X * f_Y)(x)$$

Tout d'abord, comme X et Y sont indépendantes, la loi du couple (X,Y) admet une densité de probabilité $f_{(X,Y)}$ définie pour tout $(x, y) \in \mathbb{R}^2$ par

$$f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$$

Soit $h : \mathbb{R} \to \mathbb{R}$ mesurable.

$$\begin{aligned}
E(h(Z)) &= E(h(\phi(X, Y))) \ ou \ \phi(X, Y) = X + Y \\
&= E(\tilde{h}(X, Y)) \ ou \ \tilde{h} = h \circ \phi \\
&= \int\int_{\mathbb{R}^2} \tilde{h}(x, y) f_{(X,Y)}(x, y) dxdy \\
&= \int\int_{\mathbb{R}^2} h(x + y) f_X(x) f_Y(y) dxdy \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x + y) f_X(x) dx \right) f_Y(y) dy
\end{aligned}$$

Posons $u = x + y \Rightarrow du = dx$.

$$\begin{aligned}
E(h(Z)) &= \int\int_{\mathbb{R}^2} h(u) f_X(u - y) f_Y(y) dudy \\
&= \int_{\mathbb{R}} h(u) \left(\int_{\mathbb{R}} f_X(u - y) f_Y(y) dy \right) du
\end{aligned}$$

On a donc :

$$\begin{aligned}
f_Z(u) &= \int_{\mathbb{R}} f_X(u - y) f_Y(y) dy \\
&= (f_X * f_Y)(u)
\end{aligned}$$

Corollaire :
La somme de deux v.a. gaussiennes indépendantes est encore une v.a ; gaussienne.

Demonstration :
Soient $X \hookrightarrow \mathcal{N}(0, 1)$ et $Y \hookrightarrow \mathcal{N}(0, 1)$ tel que $X \perp\!\!\!\perp Y$
Montrons que $Z = X + Y \hookrightarrow \mathcal{N}(0, 2)$

$$
\begin{aligned}
f_Z(x) &= \int_{\mathbb{R}} f_X(t) f_Y(x-t)dt \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dt \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\frac{t^2}{2}} e^{-\frac{1}{2}(x^2-2xt+t^2)} dt \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-(\frac{x^2}{2}-xt+t^2)} dt \\
&= \frac{1}{2\pi} e^{-\frac{x^2}{4}} \int_{\mathbb{R}} e^{-(\frac{x^2}{4}-xt+t^2)} dt \\
&= \frac{1}{2\pi} e^{-\frac{x^2}{4}} \int_{\mathbb{R}} e^{-(\frac{x}{2}-t)^2} dt \\
&= \frac{1}{2\pi} e^{-\frac{x^2}{4}} \underbrace{\int_{\mathbb{R}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt}_{=\sqrt{2\pi}\times\sigma} \qquad ou\ \mu = \frac{x}{2},\ \sigma^2 = \frac{1}{2} \\
&= \frac{1}{2\pi} e^{-\frac{x^2}{4}} \times \sqrt{2\pi}\frac{1}{\sqrt{2}} \\
&= \frac{1}{2\sqrt{\pi}} e^{-\frac{x^2}{4}}
\end{aligned}
$$

D'où $Z \hookrightarrow \mathcal{N}(0,2)$

Théorème (admis) :
Soit $X = (X_1, ..., X_d)$ i, vecteur aléatoire gaussien de moyenne $m = (m_1, ..., m_d)$ et de matrice de covariance Γ (on note $X \hookrightarrow \mathcal{N}_d(m, \Gamma)$)
Si Γ est inversible alors la loi de X admet une densité de probabilité f_X définie pour tout $x = (x_1, ..., x_d) \in \mathbb{R}^d$ par :

$$
f_X(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{|det(\Gamma)|}} \exp\left(-\frac{1}{2} < x-m, \Gamma^{-1}(x-m) >\right)
$$

Remarque :
Si d=1, on retourve la densité de la loi $\mathcal{N}(m, \sigma^2)$

7 Fonctions caractéristiques

Définition :

- On appelle fonction caractéristique d'une v.a.r. X l'application ϕ_X définie sur $\mathbb{R}$ et à valeur dans le disque unité fermé du plan complexe par :

$$
\forall t \in \mathbb{R}, \phi_X(t) = E(e^{itX}) = E(\cos(tX)) + iE(\sin(tX))
$$

- Si la v.a. est à valeur dans $\mathbb{R}^d$ ($d \in \mathbb{N}^*$), la fonction caractéristique (de la loi) de X, notée encore ϕ_X, est l'application définie sur $\mathbb{R}^d$ et à valeurs dans le disque unité fermé du plan complexe par :

$$
\forall t \in \mathbb{R}, \phi_X(t) = E(e^{i<t,X>}) = E(\cos(< t, X >)) + iE(\sin(< t, X >))
$$

Remarque : 1. Si $X(\Omega) \subset \mathbb{Z}$ alors

$$\forall t \in \mathbb{R}, \phi_X(t) = \sum_{k \in \mathbb{Z}} \mathbb{P}(X = k) e^{itk}$$

2. Si la loi de X admet une densité de probabilité f_X sur $\mathbb{R}^d$ ($d \in \mathbb{N}^*$) alors

$$\forall t \in \mathbb{R}^d, \phi_X(t) = \int_{\mathbb{R}} e^{i<t,X>} f_X(t) dt$$

Il s'agit de la transformée de Fourier de f_X

3. Si $X \hookrightarrow \mathcal{N}(0,1)$, alors $\forall t \in \mathbb{R}, \phi_X(t) = e^{-\frac{t^2}{2}}$
(Réciproque vraie)

Propriétés :

1. $\forall a, b \in \mathbb{R}$, $\forall t \in \mathbb{R}$, $\phi_{aX+b}(t) = e^{itb} \phi_X(at)$
2. Si $X(\Omega) \subset \mathbb{R}^d$, alors pour toute matrice A réelle $n \times d$ et tout matrice B réelle $n \times 1$:

$$\phi_{AX+B}(t) = e^{i<t,B>} \phi_X({}^t At), \ \forall t \in \mathbb{R}^d$$

3. $\phi_X(0) = 1$ et $\phi_{-X}(t) = \phi_X(-t) = \overline{\phi_X(t)}$
4. Si X est une v.a.r. intégrable alors ϕ_X est de classe $\mathcal{C}^1$ et :

$$\forall t \in \mathbb{R}, \phi'_X(t) = iE(Xe^{itX})$$

En particulier, si t=0, on obtient $\phi'_X(t) = iE(X)$

Plus généralement, si X est p-intégrable (ie $E(|X|^p) < \infty$) avec $p \in \mathbb{N}^*$, alors ϕ_X est de classe $\mathcal{C}^p$ et pour tout $t \in \mathbb{R}$

$$\phi_X^{(p)}(t) = i^p E(X^p e^{itX})$$

En particulier, $\phi_X^{(p)}(0) = i^p E(X^p)$

Si $X(\Omega) \subset \mathbb{R}^d$ et si X est d-intégrable, ϕ_X est de clsse $\mathcal{C}^\alpha$ et pour tout $p = (p_1, ..., p_d) \in \mathbb{N}^d;\ p_1 + ... + p_d \leq \alpha$, on a :

$$\frac{\partial^{p_1 + ... + p_d}}{\partial t_1^{p_1} ... \partial t_d^{p_d}} \phi_X(0) = i^{p_1 + ... + p_d} E(X_1^{p_1} ... X_d^{p_d})$$

5. Si X et Y sont deux v.a.r. indépendantes :

$$\phi_{X+Y}(t) = \phi_X(t) \phi_Y(t)$$

Théorème :
Deux v.a.r. (ou d-dimensionnelle) ont même loi si et seulement si elles ont même fonction caractéristique.

Théorème :
Si X est une v.a. d-dimensionnelle et si

$$\int_{\mathbb{R}^d} |\phi_X(t)| dt_1 ... dt_d < \infty$$

alors X admet une densité de probabilité f_X continue sur $\mathbb{R}^d$ définie pour tout $x \in \mathbb{R}^d$ par :

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_X(t) e^{-i<t,X>} dt$$

Transformée de Laplace d'une v.a.r. positive :
Lorsque X est positive p.s., on peut utiliser la transformée de Laplace.

Définition :
Soit X une v.a.r. positive p.s. On appelle transformée de Laplace de la loi de X la fonction :

$$\begin{array}{rcl} L_X : \mathbb{R}^+ & \to &]0,1] \\ \lambda & \mapsto & E(e^{-\lambda X}) \end{array}$$

8 Conditionnement d'une variable aléatoire, espérance conditionnelle

Soient X et Y deux v.a. réelles définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

8.1 Conditionnement d'une v.a. par rapport à une autre

Théorème (de Doob) :
Il existe une application :

$$\begin{array}{rcl} q : \mathcal{B}(\mathbb{R}) \times \mathbb{R} & \to & [0,1] \\ (B, x) & \mapsto & q(B, x) \end{array}$$

vérifiant :

1. Pour tout $B \in \mathcal{B}(\mathbb{R})$, l'application $q(B, \bullet)$ est mesurable.
2. $\forall x \in \mathbb{R}$, l'application $q(\bullet, x)$ est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$
3. Pour tout $A \in \mathcal{B}(\mathbb{R}^2)$ on a :

$$\begin{array}{rcl} \mu_{(X,Y)}(A) & = & E(1_A(X,Y)) \\ & = & \displaystyle\iint_{\mathbb{R}^2} 1_A(x,y) q(dy, x) d\mu_X(x) \end{array}$$

avec $\mu_{(X,Y)}$ et μ_X les lois de probabilité du vecteur aléatoire (X,Y) et de la variable aléatoire X respectivement.

Définition :
L'application q est appellée "loi conditionnelle de Y sachant X"

Conséquence :
Pour toute fonction $\mu_{(X,Y)}$-intégrable $f : \mathbb{R}^2 \to \mathbb{R}$ (ie $\iint_{\mathbb{R}^2} |f(x,y)| d\mu_{(X,Y)}(x,y) < \infty$) on a :

$$E(f(X,Y)) = \iint_{\mathbb{R}^2} f(x,y) q(dy, x) d\mu_X(x)$$

Remarque :
On montre que q du théorème ci-dessus est unique dans le sens suivant :
Si $\tilde{q}$ est une autre loi conditionnelle de Y sachant X alors il existe un borelien μ_X-négligeable N de $\mathbb{R}$ (ie : $N \in \mathcal{B}(\mathbb{R})$ et $\mu_X(N) = 0$) tel que :

$$\forall x \in \mathbb{R} \backslash N,\ \forall B \in \mathcal{B}(\mathbb{R}),\ q(B, x) = \tilde{q}(B, x)$$

Définition :
Pour tout $x \in \mathbb{R}$ on note

$$q(\bullet, x) = \mathbb{P}(\bullet | X = x)$$

et $q(\bullet, x)$ est appelée "loi conditionnelle de Y sachant X".

Attention : Pour tout $B \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}(B|X = x) = q(B, \bullet)$ est classe d'équivalence par l'égalité μ_X-p.s. de fonctions mesuablres de $\mathbb{R}$ dans [0,1].

Par conséquent, pour un $x \in \mathbb{R}$ particulier tel que $\mu_X(\{x\}) = 0$ (ie $\mathbb{P}(X = x) = 0$), l'expression $\mathbb{P}(B|X = x)$ n'a pas de sens.

On détermine en général q par identification à l'aide des points 1, 2 et 3 du théorème de Doob. Cependant, il y a au moins 3 cas où l'on a un résultat explcite :

1er cas : Si μ_X est discrète alors pour tout $x \in \mathbb{R}$ tel que $\mathbb{P}(X = x) > 0$ alors :

$$q(B, x) = \mathbb{P}(B|X = x) = \frac{\mathbb{P}(B \cap \{X = x\})}{\mathbb{P}(X = x)}$$

2eme cas : Si le couple de v.a. (X,Y) admet une densité de probabilité $f_{(X,Y)}$ alors, pour μ_X-presque tout $x \in \mathbb{R}$ tel que $f_X(x) \neq 0$:

$$\forall B \in \mathcal{B}(\mathbb{R}), q(B, x) = \int_B \frac{f_{(X,Y)}(x, y)}{f_X(x)} dy$$

3eme cas : Si les v.a. X et Y sont indépendantes, alors, pour μ_X-presque tout $x \in \mathbb{R}$, on a $q(\bullet, x) = \mu_Y$
(ie la loi conditionnelle de Y sachant $X = x$ ne dépend pas de x)

8.2 Espérance conditionnelle de Y sachant X

On suppose que Y est intégrable. On montre à l'aide du théorème de Radon-Nikodyn qu'il existe une unique classe d'équivalence pour l'égalité $\mathbb{P}$-p.s. de la v.a. Z à valeur dans $\mathbb{R}$ vérifiant :

1. Z est $\sigma(X)$-mesurable
   ie : $\forall A \in \mathcal{B}(\mathbb{R}),\ Z^{-1}(A) \in \sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}))$
2. $\forall A \in \sigma(X)$:
$$\int_A Z d\mathbb{P} = \int_A Y d\mathbb{P}$$

Définition :
La classe d'équivalence de v.a. Z ainsi définie est appelée "espérance conditionnelle de Y sachant X" (ou encore "epsérance conditionnelle de Y sachant la tribu $\sigma(X)$").
Elle est notée E(Y|X).

On détermine E(Y|X) par identification à l'aide de 1 et 2.
D'autre part, on peut vérifier que si q est la loi conditionnelle de Y sachant X, alors l'application :

$$\begin{array}{ccc} \Omega & \to & \mathbb{R} \\ \omega & \mapsto & \displaystyle\int_{\mathbb{R}} y\ q(dy, X(\omega)) \end{array}$$

est une version de E(Y|X).

Plus généralement, pour toute application mesurable f tel que $f(Y)$ soit intégrable $\omega \mapsto \int_{\mathbb{R}} f(y)\ q(dy, X(\omega))$ est une version de E(f(Y)|X)

Remarque :
E(Y|X) est une fonction $\sigma(X)$-mesurable. D'après un (autre) théorème de Doob, il existe une fonction $\phi : \mathbb{R} \to \mathbb{R}$ mesurable tel que :

$$E(X|Y) = \phi(X)$$

ϕ est notée $\phi(x) = E(Y|X = x)$ mais il faut faire attention au fait que ϕ n'est définie que modulo l'égalité μ_X-ps.
On a alors pour tout $h : \mathbb{R} \to \mathbb{R}$ mesurable :

$$E(h(Y)|X = x) = \int_{\mathbb{R}} h(y)\ q(dy, x)$$

pour μ_X-presque tout $x \in \mathbb{R}$

Remarque :
Pour tout $B \in \mathcal{B}(\mathbb{R})$:

$$E(1_B(Y)|X = x) = q(B, x) = \mathbb{P}(Y \in B|X = x)$$

pour μ_X-presque tout $x \in \mathbb{R}$