

FORMULAIRE du cours PROBABILITÉS 1

I) Généralités du calcul des probabilités : Soit Ω un ensemble (univers des possibles)

Espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$: \mathcal{F} est une tribu sur Ω c'est à dire $\mathcal{F} \subset \mathcal{P}(\Omega)$ vérifie

$$1) \Omega \in \mathcal{F}, \quad 2) (A_n)_{n \in \mathbb{N}^*} \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{+\infty} A_n \in \mathcal{F}, \quad 3) A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F},$$

et $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ (la probabilité) est telle que :

$$1) \mathbb{P}(\Omega) = 1, \quad 2) \mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n) \text{ pour tous } (A_n) \in \mathcal{F}, \text{ tels que } A_i \cap A_j = \emptyset \text{ si } i \neq j.$$

Limite sup et Limite inf : Si $(A_n)_{n \in \mathbb{N}^*} \in \mathcal{F}$, on note

$$1) \liminf A_n = \bigcup_{n=1}^{+\infty} \left(\bigcap_{m=n}^{+\infty} A_m \right), \quad 2) \limsup A_n = \bigcap_{n=1}^{+\infty} \left(\bigcup_{m=n}^{+\infty} A_m \right). \text{ On a alors :}$$

$$1) \liminf A_n \subset \limsup A_n, \quad 2) (\liminf A_n)^c = \limsup A_n^c, \quad 3) (\limsup A_n)^c = \liminf A_n^c.$$

Continuité de \mathbb{P} : Pour toute suite monotone $(A_n) \in \mathcal{F}$, on a $\mathbb{P}(\lim A_n) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n)$

(où $\lim A_n = \bigcup_n A_n$ si $A_n \nearrow$ et $\lim A_n = \bigcap_n A_n$ si $A_n \searrow$).

Probabilité conditionnelle : Si $B \in \mathcal{F}$ et $\mathbb{P}(B) > 0$, la quantité $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ est la probabilité conditionnelle de A sachant B .

Formule de l'intersection : Si $A_1, A_2, \dots, A_k \in \mathcal{F}$ et $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{k-1}) > 0$,

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_k) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_k|A_1 \cap \dots \cap A_{k-1}).$$

Formule de la probabilité totale : Soit (A_n) un système complet d'événements. Alors :

$$\forall A \in \mathcal{F}, \mathbb{P}(A) = \sum_n \mathbb{P}(A|A_n) \mathbb{P}(A_n).$$

Formule de Bayes : $\mathbb{P}(A_n|A) = \frac{\mathbb{P}(A|A_n) \mathbb{P}(A_n)}{\sum_k \mathbb{P}(A|A_k) \mathbb{P}(A_k)}$ ($A \in \mathcal{F}$) si (A_n) est un système complet.

Événements indépendants : Ce sont des événements A_1, \dots, A_n tels que $\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \dots \mathbb{P}(A_n)$.

Espace probabilisé produit : Si $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ ($1 \leq i \leq n$) sont des espaces probabilisés, leur espace produit est $(\Omega, \mathcal{F}, \mathbb{P})$ où $\Omega = \Omega_1 \times \dots \times \Omega_n$, $\mathcal{F} = \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n$ et $\mathbb{P} = \mathbb{P}_1 \otimes \dots \otimes \mathbb{P}_n$ où \mathcal{F} est engendrée par les rectangles $A_1 \times \dots \times A_n$, $A_i \in \mathcal{F}_i$, et \mathbb{P} est telle que $\mathbb{P}(A_1 \times \dots \times A_n) = \prod_{i=1}^n \mathbb{P}_i(A_i)$.

II) Variables aléatoires discrètes : Ce sont les applications $X : \Omega \rightarrow \mathbb{R}$ telles que $X(\Omega) = \{x_k; k \in D\}$ est fini ou dénombrable ($D = \{1, \dots, N\}$ ou $D = \mathbb{N}$) et vérifiant la condition de mesurabilité : $\forall k \in D, [X = x_k] \in \mathcal{F}$.

Distribution de probabilité : C'est la suite des nombres $p_k = \mathbb{P}(X = x_k)$ ($k \in D$). Ce sont des nombres $0 \leq p_k \leq 1$ tels que $\sum_{k \in D} p_k = 1$.

Moments : La v.a. X a un moment d'ordre n ($n \in \mathbb{N}^*$) si $\sum_{k \in D} p_k |x_k|^n < +\infty$ et le moment d'ordre n est $\mathbb{E}(X^n) = \sum_{k \in D} p_k x_k^n$ (c'est l'espérance si $n = 1$). Si $\mathbb{E}(X^2)$ existe, la variance est $Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ et $\sigma_X = \sqrt{Var(X)}$ est l'écart type.

Fonction génératrice : Si X est à valeurs dans \mathbb{N} , c'est la série entière $G_X(t) = \sum_{k=0}^{\infty} p_k t^k$. Si X et Y sont indépendantes, $G_{X+Y}(t) = G_X(t)G_Y(t)$. Si X a un moment d'ordre 2, G_X est deux fois dérivable, $\mathbb{E}(X) = G'_X(1)$ et $Var(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$.

v.a. binomiale $\mathcal{B}(n, p)$: Toute v.a. X telle que $X(\Omega) = \{0, 1, \dots, n\}$ et pour tout $0 \leq k \leq n$ $\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}$, (où $n \geq 1$ et $p \in [0, 1]$ sont fixés). Si $n = 1$, on dit que X est une v.a. de Bernoulli. On a $\mathbb{E}(X) = np$ et $Var(X) = np(1-p)$.

v.a. de Poisson de paramètre $\lambda > 0$: Toute v.a. X telle que $X(\Omega) = \mathbb{N}$ et $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ ($k \in \mathbb{N}$). On a $\mathbb{E}(X) = Var(X) = \lambda$.

v.a. de Pascal de paramètre $p \in]0, 1[$: Toute v.a. telle que $X(\Omega) = \mathbb{N}^*$ et $\mathbb{P}(X = k) = p(1-p)^{k-1}$ (k entier ≥ 1) (on l'appelle aussi v.a. géométrique, ou v.a. instant du premier succès).

III) Variables aléatoires quelconques, lois de probabilité : Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

Vecteur aléatoire (v.a.) (ou variable aléatoire si $d = 1$) : C'est une application mesurable $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B}_d)$ où \mathcal{B}_d est la tribu borélienne de \mathbb{R}^d , c'est à dire : $\forall B \in \mathcal{B}_d, [X \in B] = \{\omega \in \Omega ; X(\omega) \in B\} \in \mathcal{F}$ (condition de mesurabilité).

Condition suffisante de mesurabilité : Soit \mathcal{C} une partie génératrice de \mathcal{B}_d . Si $X : \Omega \rightarrow \mathbb{R}^d$ est telle que $\forall C \in \mathcal{C}, [X \in C] \in \mathcal{F}$, alors X est un v.a.

Transformée déterministe d'un(e) v.a. : Si $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^p, \mathcal{B}_p)$ est un v.a. de \mathbb{R}^p et $f : (\mathbb{R}^p, \mathcal{B}_p) \rightarrow (\mathbb{R}^q, \mathcal{B}_q)$ est mesurable, alors $f \circ X = f(X)$ est un vecteur aléatoire de \mathbb{R}^q .

Limite d'une suite de v.a. : Si (X_n) est une suite de v.a. telle que $\forall \omega \in \Omega, \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \in \mathbb{R}$ existe, alors X est une v.a..

v.a. indicatrice et v.a. simple : Soit $A \in \mathcal{F}$. La v.a. indicatrice de A est donnée par $\mathbf{1}_A(\omega) = 0$ si $\omega \notin A$ et $= 1$ si $\omega \in A$. Toute v.a. $X = \sum_{finie} \alpha_i \mathbf{1}_{A_i}$ ($\alpha_i \in \mathbb{R}$ et $A_i \in \mathcal{F}$) est dite v.a. simple.

Théorème de structure : Toute v.a. $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}_1)$ (resp. toute v.a. ≥ 0) est limite simple d'une suite de v.a. simples (resp. de v.a. simples ≥ 0).

Loi de probabilité d'un v.a. X de \mathbb{R}^d : C'est la mesure de probabilité μ_X sur \mathcal{B}_d définie par : $\forall B \in \mathcal{B}_d, \mu_X(B) = \mathbb{P}(X \in B)$.

Fonction de répartition d'une v.a. X : C'est la fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ telle que : $\forall t \in \mathbb{R}, F(t) = \mathbb{P}(X \leq t)$. Elle est telle que : 1) $0 \leq F(t) \leq 1$; 2) F est croissante (au sens large), continue à droite en chaque $t \in \mathbb{R}$; 3) $\lim_{t \rightarrow -\infty} F(t) = 0$ et $\lim_{t \rightarrow +\infty} F(t) = 1$.

Densité d'un v.a. : Le v.a. X de \mathbb{R}^d a une densité f si $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ est intégrable avec $\int_{\mathbb{R}^d} f(x_1, \dots, x_d) dx_1 \dots dx_d = 1$ et si $\forall B \in \mathcal{B}_d, \mu_X(B) = \int_B f(x_1, \dots, x_d) dx_1 \dots dx_d$.

Densité normale $N(0, 1)$: C'est la fonction $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ ($x \in \mathbb{R}$).

Densité normale $N(m, \sigma^2)$: C'est la fonction $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right)$ ($x \in \mathbb{R}$).

Densité uniforme sur $[a, b]$: C'est la fonction $f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$ ($x \in \mathbb{R}$).

Densité exponentielle de paramètre $\lambda > 0$: C'est la fonction $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$.

Densités marginales : Si (X, Y) est un couple de v.a. ayant une densité $f(x, y)$ sur \mathbb{R}^2 , les densités marginales sont données par $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$ et $f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$.

v.a. indépendantes : Les v.a. X_1, \dots, X_k sont dites (mutuellement) indépendantes si $\forall B_1, \dots, B_k \in \mathcal{B}_1, \mathbb{P}(X_1 \in B_1, \dots, X_k \in B_k) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_k \in B_k)$.

CNS d'indépendance : Des v.a. X et Y de lois respectives μ_X et μ_Y , sont indépendantes si et seulement si $\mu_{(X,Y)} = \mu_X \otimes \mu_Y$ (où $\mu_{(X,Y)}$ est la loi dans \mathbb{R}^2 du vecteur (X, Y)). Si (X, Y) a une densité $f(x, y)$ ceci équivaut à : $f(x, y) = f_X(x) f_Y(y)$ (produit des densités marginales).

IV) Espérance, variance et moments des v.a. : Soit (Ω, \mathcal{F}, P) un espace probabilisé.

Espace \mathcal{L}^1 : Une v.a. X est dans \mathcal{L}^1 si $\int_{\Omega} |X| dP < +\infty$. On pose alors $\mathbb{E}(X) = \int_{\Omega} X dP$ et on l'appelle l'espérance mathématique de X .

Formule du transfert : Soit X un v.a. de \mathbb{R}^d et μ_X sa loi de probabilité sur \mathcal{B}_d . Soit $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ borélienne. Alors $\varphi \circ X \in \mathcal{L}^1(\Omega, \mathcal{F}, P) \Leftrightarrow \varphi \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}_d, \mu_X)$ et on a $\mathbb{E}(\varphi \circ X) = \int_{\Omega} \varphi \circ X dP = \int_{\mathbb{R}^d} \varphi(x) d\mu_X(x)$. Ainsi si X est réelle et dans \mathcal{L}^1 on a : $\mathbb{E}(X) = \int_{\mathbb{R}} x d\mu_X(x)$.

Espérance d'un produit : Si X_1, \dots, X_n sont dans \mathcal{L}^1 et sont indépendantes : $\mathbb{E}(\prod_{i=1}^n X_i) = \prod_{i=1}^n \mathbb{E}(X_i)$.

Espace \mathcal{L}^2 et variance : Une v.a. X est dans \mathcal{L}^2 si $\mathbb{E}(X^2) < +\infty$. On a $\mathcal{L}^2 \subset \mathcal{L}^1$. $\mathbb{E}(X^2)$ est le moment d'ordre 2 de X . La variance de X est le nombre $Var X = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ et l'écart type est $\sigma_X = \sqrt{Var X}$.

Covariance : Pour X et Y dans \mathcal{L}^2 c'est le nombre $cov(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$. Si X et Y sont indépendantes alors $cov(X, Y) = 0$ (la réciproque est fautive).

Variance d'une somme : Soient X_1, \dots, X_n des v.a. de \mathcal{L}^2 deux à deux de covariance nulle. Alors $Var(X_1 + \dots + X_n) = Var X_1 + \dots + Var X_n$.

Inégalité de Markov : Si $X \in \mathcal{L}^1$, alors $\forall a > 0$, on a $\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}$.

Inégalité de Bienaymé-Tchebychev : Si $X \in \mathcal{L}^2, \forall a > 0, \mathbb{P}\left(\left|\frac{X - \mathbb{E}(X)}{\sigma_X}\right| \geq a\right) \leq \frac{1}{a^2}$.

V) Convergence des v.a., loi forte des grands nombres et théorème limite central

Convergence p.s. et en probabilité : La suite de v.a. (X_k) converge p.s. (resp. en probabilité) vers la v.a. X si $\mathbb{P}([\lim_{k \rightarrow \infty} X_k = X]) = 1$ (resp. $\forall \delta > 0, \lim_{n \rightarrow +\infty} P(|X_n - X| \geq \delta) = 0$). La convergence p.s. implique la convergence en probabilité.

CS de convergence p.s. : Si $\forall \varepsilon > 0, \sum_{k=1}^{\infty} P(|X_k| > \varepsilon) < +\infty$ alors $X_n \rightarrow 0$ p.s. En particulier si $(X_n) \in \mathcal{L}^p$ et si $\sum_{n=1}^{+\infty} E(|X_n|^p) < +\infty$ alors $X_n \rightarrow 0$ p.s.

Lemme de Borel Cantelli : Si $(A_n) \in \mathcal{F}$ et $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < +\infty$. Alors $\mathbb{P}(\limsup A_n) = 0$.

Loi forte des grands nombres : Si (X_k) est une suite de v.a. de \mathcal{L}^2 , de même espérance m , de variances bornées et 2 à 2 de covariance nulle alors $\frac{1}{n} \sum_{k=1}^n X_k \rightarrow m$ p.s. ($n \rightarrow +\infty$).

Loi forte des grands nombres de Kolmogorov : Si (X_k) est une suite de v.a. de L^1 indépendantes et de même loi alors $\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n X_k = m$ p.s. ($m = \mathbb{E}(X_1)$).

La convergence en loi : La suite de v.a. (X_n) (de loi μ_n) converge en loi vers la v.a. X (de loi μ) si $\forall f \in C_b(\mathbb{R}), \lim_{n \rightarrow +\infty} \langle \mu_n, f \rangle = \langle \mu, f \rangle$. Si F_n (resp. F) est la fonction de répartition de X_n (resp. X) ceci équivaut à : $\lim_{n \rightarrow +\infty} F_n(t) = F(t)$ en tout point de continuité t de F .

Théorème limite central : Si $X_n (n \geq 1)$ sont des v.a. de \mathcal{L}^2 , indépendantes et de même loi d'espérance m et de variance σ^2 , alors $\lim_{n \rightarrow +\infty} \frac{1}{\sigma\sqrt{n}} ((\sum_{i=1}^n X_i) - nm) = N(0, 1)$ en loi.

VI) Chaînes de Markov : Soit (X_n) une chaîne de Markov homogène à espace d'états fini $E = \{1, 2, \dots, r\}$ et $P = (p_{ij})$ la matrice des transitions.

Chaînes de Markov absorbantes : On renumérote les états en commençant par les $r - s$ états absorbants puis par les s états transients pour écrire P . Soit Q la sous matrice $s \times s$ des transitions entre les états transients et R la matrice $s \times (r - s)$ des transitions des états transients vers les états absorbants. La matrice $N = (I_s - Q)^{-1}$ (matrice fondamentale) a pour coefficient n_{ij} le nombre moyen de passages en j partant de i et $n(i) := \sum_j n_{ij}$ est la durée de vie moyenne de la chaîne partant de i . Le coefficient a_{ij} de $A = NR$ est la probabilité partant de i d'être absorbé en j .

Chaînes de Markov ergodiques régulières : Elles sont telles qu'il existe un entier n tel que P^n ait tous ses coefficients > 0 . Il existe un vecteur ligne de probabilité α à coefficients > 0 (loi stationnaire) tel que pour tout vecteur de probabilité $c, cP^n \rightarrow \alpha$ ($n \rightarrow +\infty$) et c est l'unique vecteur de probabilité tel que $\alpha P = \alpha$. De plus la i -ième coordonnée de α est égale à l'inverse du temps moyen de retour en l'état i .

Chaînes de Markov ergodiques périodiques : Elles ont une seule classe d'éléments communicants mais pour tout état i , les éléments de $A_{ii} = \{n \in \mathbb{N}; p_{ii}^{(n)} > 0\}$ sont des multiples d'un même nombre $d \geq 2$ (la période). Les états sont partitionnés en d classes cycliques telles qu'on passe avec probabilité 1 d'une classe à l'autre et on revient au bout de d étapes dans la classe de l'état initial. Il existe une loi stationnaire α telle que $\alpha P = \alpha$ et pour tout vecteur de probabilité c, cP^n converge au sens de Cesaro vers α .

Fonctions harmoniques : Si $f : E \rightarrow \mathbb{R}$, soit $\bar{f}(i) = \sum_{j \in E} p_{ij} f(j)$. Alors f est harmonique (resp. sur(-sous)harmonique) sur $A \subset E$ si pour tout $i \in A, f(i) = \bar{f}(i)$ (resp. $f(i) \geq \bar{f}(i)$, resp. $f(i) \leq \bar{f}(i)$). Dans une chaîne absorbante, soit ℓ un état absorbant fixé, la fonction $a(i) := a_{i\ell}$ est harmonique sur les états transients, la fonction $n(i)$ est surharmonique (car $\forall i$ transient, $n(i) = 1 + \bar{n}(i)$). Enfin pour j transient fixé, la fonction $m(i) := n_{ij}$ vérifie $m(j) = 1 + \bar{m}(j)$ et $m(i) = \bar{m}(i)$ si $i \neq j$ est transient (pour les notations voir ci-dessus le résumé sur les chaînes de Markov absorbantes).