

SIC 3101 - Probabilités

Rappels, précisions pour la suite

Notations

— En général, sauf précision supplémentaire, on utilisera les notations usuelles du cours ou celles précédemment employées.

Par exemple, P désignera toujours une probabilité sur un univers Ω , A,B,C... des événements (parties de Ω), $\mathscr A$ une tribu sur Ω , n un entier de $\mathbb N$ ou $\mathbb N^*$, X,Y des variables aléatoires...

— On notera A^c le complémentaire d'un événement A de l'univers Ω .

Premières définitions

Incompatibilité : $^1A \cap B = \emptyset$

Indépendance : ${}^{2}P(A \cap B) = P(A)P(B)$

Indépendance mutuelle : $(A_i)_{i \in I}$ est une famille d'événements mutuellement indépendants ssi :

$$\forall J_{fini} \subset I, P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$$

Remarque. Les notions d'indépendance mutuelle et d'indépendance deux à deux sont différentes. L'indépendance "tout court" désigne souvent par abus de langage l'indépendance mutuelle, attention.

Négligeabilité : P(A) = 0

Événement presque sûr : P(A) = 1Probabilité uniforme sur Ω fini :

$$\begin{array}{ccc} P(\Omega) & \to & [0,1] \\ A & \mapsto & \frac{cardA}{card\Omega} \end{array}$$

Dénombrabilité : Un ensemble E est dénombrable ssi il existe une bijection entre E et N.

Probabilités conditionnelles

Définition : Soit un événement B tel que P(B)>0.

On définit, pour tout événement A, la probabilité ³ conditionnelle de A sachant B par :

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Théorème de Bayes (première forme) : Il s'agit d'une simple réécriture du dernier résultat, exprimant P(A|B) en fonction de P(B|A) afin d'inverser le conditionnement.

Mais en pratique, c'est la 2ème forme du théorème qui est utilisée.

^{1.} Dans ce cas, on a de plus $P(A \cup B) = P(A) + P(B)$, soit une propriété sur la somme. (Si A et B sont compatibles, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$).

^{2.} L'indépendance implique donc des relations sur le produit. Ne pas confondre avec l'incompatibilité.

^{3.} On démontre en effet que cela définit une application qui est une probabilité dans le sens que l'on donnera ensuite.



Inversion du conditionnement - Probabilités totales

Formule des probabilités totales : Pour $(C_i)_{i=1}^n$ un système complet d'événements 4 (non vides) :

$$\forall i \in [1, n], P(A) = \sum_{i=1}^{n} P(A \cap C_i) = \sum_{i=1}^{n} P(C_i) P(A|C_i)$$

Théorème de Bayes (forme pratique) : Avec le système complet précédent :

$$\forall i \in [1, n], P(C_i|A) = \frac{P(C_i)P(A|C_i)}{\sum_{j=1}^{n} P(C_j)P(A|C_j)}$$

Remarque. Dans les exos sur le conditionnement, le choix du système complet est souvent déterminant. S'il n'est pas amené naturellement par l'énoncé ou les réponses précédentes, penser au système $\{B, B^c\}$, suffisant dans les cas les plus simples.

Eléments de base en probabilités

Définition (Tribu). Par proposition, une tribu \mathscr{A} sur Ω peut être définie comme une partie de $P(\Omega)$ (famille de parties de Ω) telle que :

- $-\Omega \in \mathscr{A} \text{ ou } \varnothing \in \mathscr{A}$
- $-\forall A \in \mathcal{A}, A^c \in \mathcal{A}$
- $(A_n) \in \mathscr{A}^{\mathbb{N}}, \bigcap_{n \in \mathbb{N}} A_n \in \mathscr{A} \text{ ou } : \forall (A_n) \in \mathscr{A}^{\mathbb{N}}, \bigcup_{n \in \mathbb{N}} A_n \in \mathscr{A}$

Remarque. $P(\Omega)$ est une toujours une tribu, et en probabilités finies ou dénombrables, l'espace probabilisé étudié sera toujours défini à partir de $P(\Omega)$, d'où son importance.

Définition (Espace probabilisable, probabilisé). À partir de Ω , on définit :

- Espace probabilisable : Couple (Ω, \mathscr{A}) avec \mathscr{A} tribu sur Ω
- Espace probabilisé : Triplet (Ω, \mathcal{A}, P) avec P probabilité sur \mathcal{A}

Définition (Probabilité). Par proposition, $P: \mathscr{A} \to \mathbb{R}^+$ est une probabilité sur \mathscr{A} ssi:

- $-\forall A \in \mathcal{A}, P(A) \in [0,1]$
- $-P(\Omega)=1$
- L'une des propriétés équivalentes suivantes est vérifiée :
 - $\cdot \ \forall A,B \in \mathscr{A} : A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$
 - $\forall (A_n) \in \mathscr{A}^{\mathbb{N}} : (\forall i, j \in \mathbb{N}, A_i \cap A_j = \varnothing) \Rightarrow P(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$

Proposition (Continuité monotone). Soit $(A_n) \in \mathscr{A}^{\mathbb{N}}$:

^{4.} C'est une partition de $\Omega: \bigcup_{i=1}^{n} C_i = \Omega$ et $\forall i, j \in [\![1,n]\!]: i \neq j \Rightarrow C_i \cap C_j = \emptyset$



Variables aléatoires scalaires et multidimensionnelles

Variables aléatoires discrètes

Définition (Variable aléatoire). C'est une application $X:(\Omega,\mathscr{A})\to (E,\varepsilon)$ mesurable.

Une variable aléatoire peut être réelle (si $E = \mathbb{R}$), complexe, vectorielle, multidimensionnelle ou multivariée, ou encore une suite de variables aléatoires.

Notation (Image réciproque).

$$\forall A \in E, [X \in A] = X^{-1}(A)$$

Définition (Loi de probabilité P_X d'une variable aléatoire X). C'est la mesure de probabilité suivante :

$$\begin{array}{ccc} P_X: & \varepsilon & \to & [0,1] \\ & A & \mapsto & P([X \in A]) \end{array}$$

Remarque. On note plutôt $P(X \in A)$ pour alleger.

Aussi:
$$P_X(x) = P_X(x) = P(X = x) = p_x = P(X \in \{x\})$$

Notation (Pour des variables suivant la même loi). Si X et Y sont des v.a. telles que $P_X = P_Y$, $alors^5$ on notera $X \sim Y$

Voici une liste de lois de probabilité à connaître absolument, et qui peuvent potentiellement servir dans tous les exos. Notamment, le fait que la somme d'une loi de probabilité vaut 1 simplifie énormément de calculs a priori difficiles.

Loi uniforme discrète : Si n = cardE :

$$\forall x \in E, P(X = x) = \frac{1}{n}$$

Loi de Poisson de paramètre λ : (notée parfois $\mathscr{P}(\lambda)$)

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

Loi binomiale de paramètres n et p : (notée $\mathcal{B}(n,p)$)

$$\forall k \in [0, n], P(X = k) = \binom{n}{k} p^k (1 - p)^{n - k}$$

Loi de Bernoulli de paramètre p : Cette loi est définie par ⁶

$$\begin{cases} P(X=0) &= 1-p \\ P(X=1) &= p \\ P(X=x) &= 0 \end{cases}$$
 sinon

^{5.} On ne peux pas en déduire que X=Y, erreur à ne pas commettre.

^{6.} En résumé : $\forall x \in \mathbb{R}, P(X = x) = p^x (1 - p)^{1 - x} \mathbb{1}_{\{0,1\}}(x)$.

Cela doit vous rappeler la loi binomiale.



Remarque (Exo à maîtriser). Si $(X_i)_{i=1}^n$ est une famille de v.a. de Bernoulli de paramètre p, indépendantes p et identiquement distibuées p p alors :

$$\sum_{k=1}^{n} X_k \sim \mathscr{B}(n, p)$$

Loi géométrique de paramètre p : (notée $\mathscr{G}(p)$) 10

$$\forall k \in \mathbb{N}^*, P(X = k) = p(1 - p)^{k-1}$$

Approximation de Poisson : Soit (X_n) une suite de v.a. telle que

 $\forall n \in \mathbb{N}, X_n \sim \mathscr{B}(n, p_n), \text{ où } (np_n) \text{ converge vers } \lambda \in \mathbb{R}.$

Alors (X_n) converge en loi vers la loi de Poisson de paramètre λ :

$$X_n \xrightarrow[n \to \infty]{Loi} \mathscr{P}(\lambda)$$

i.e.

$$\forall k \in \mathbb{N}, P(X = k) \xrightarrow[n \to \infty]{} \frac{\lambda^k e^{-\lambda}}{k!}$$

Variables aléatoires continues = Variables aléatoires à densités

Définition (Fonction de répartition).

$$F_X: \quad \mathbb{R} \quad \to \quad \begin{bmatrix} 0,1 \end{bmatrix} \\ \quad x \quad \mapsto \quad P(X \le x)$$

Proposition. Par définition, F_X est croissante.

De plus,
$$\lim_{x \to -\infty} F_X(x) = 0$$
 et $\lim_{x \to +\infty} F_X(x) = 1$

^{7.} Deux v.a. X_i et X_j sont dites indépendantes ssi

 $[\]forall (k, k'), (i, j), P(X_i = k \cap X_j = k') = P(X_i = k)P(X_j = k')$

^{8.} Deux v.a. X_i et X_j sont dites identiquement distribuées ssi

 $[\]forall k, (i, j), P(X_i = k) = P(X_j = k)$

^{9.} L'abréviation i.i.d est souvent employée pour désigner les v.a. indépendantes et identiquement distribuées.

^{10.} Il s'agit de la seule loi de probabilité discrète sans mémoire, sachant que la seule loi de probabilité continue sans mémoire est la loi exponentielle.



Définition (Densité de probabilité). Si X est μ -mesurable, la densité de probabilité f_X est l'application intégrable telle que : $\forall x \in \mathbb{R}, F_X(x) = \int\limits_{-\infty}^x f_X(t) \mu(dt)$

$$Donc: \forall x \in \mathbb{R}, f_X(x) = F_X'(x)$$

Voici une liste de densités de probabilité à connaître **absolument**, et qui peuvent potentiellement servir dans tous les exos. Notamment, le fait que **l'intégrale d'une densité de probabilité sur** \mathbb{R} vaut 1 simplifie énormément de calculs a priori difficiles.

Loi uniforme continue : $X \sim \mathcal{U}([a,b])$

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

Loi exponentielle de paramètre λ : $X \sim \mathcal{E}(\lambda)$

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

Définition (Densité conditionnelle).

$$\forall x, y \in \mathbb{R}, f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Proposition (Changement de variables aléatoires).

$$f_{R,\Theta}(r,\theta) = \left| \det \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} \right| f_{X,Y}(x,y)$$

$$f_{R,\Theta}(r,\theta) = \frac{1}{\left| \det \begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{pmatrix} \right|} f_{X,Y}(x,y)$$

Espérance

Définition.

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbb{R}} x dP_X(x) = \int_{\mathbb{R}} x f_X(x) dx$$

Remarque (Linéarité).

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

Proposition (Théorème du transfert).

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

Remarque (L'espérance d'une indicatrice est une probabilité).

$$\mathbb{E}(1_A) = P(1_A = 1) = P(A)$$



Variance 11

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

Définition (Covariance).
$$Cov(X,Y) = {}^{12} \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Important:
$$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$$

Remarque. $Si \ X \ et \ Y \ sont \ indépendantes : Cov(X,Y) = 0, \ alors :$

$$V(X+Y) = V(X) + V(Y)$$

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

Les réciproques sont fausses.

Mode

Définition (Mode). Pour une variable aléatoire X,

Si X est discrète : C'est la valeur de k pour laquelle la loi de probabilité est la plus grande.

Si X est continue : C'est la valeur de x pour laquelle la densité de probabilité est maximale.

Médiane

Définition (Médiane). C'est la valeur m pour laquelle :

$$P(X \le m) = P(X \ge m) = \frac{1}{2}$$

Caractéristiques de la loi normale : $X \sim \mathcal{N}(\mu, \sigma^2)$

Une v.a. suivante une loi normale est appelée une gaussienne.

Densité:

$$\forall x \in \mathbb{R}, f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

Moyenne:

$$\mathbb{E}(X) = \mu = moyenne = mode = m\'ediane$$

Variance:

$$\sigma^2 = V(X)$$

Loi normale centrée $X \sim \mathcal{N}(0, \sigma^2)$:

$$\mathbb{E}(X) = 0$$

Loi normale réduite $X \sim \mathcal{N}(\mu, 1)$:

$$V(X) = 1$$

Loi normale centrée réduite :

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Exemple important, tombe fréquemment.

^{11.} La variance V(X) est le carré de l'écart-type $\sigma.$

^{12.} On a aussi : $Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$



Courbe de la densité : C'est une gaussienne (cloche).

Symétrie : On a une symétrie par rapport à l'axe $x = \mu$:

$$\forall x \in \mathbb{R}, F_X(\mu - x) = F_X(\mu + x)$$

i.e. (aire ¹³ sous la courbe):

$$\mathscr{A}_{x \ge \mu} = \mathscr{A}_{x \le \mu} = \frac{1}{2}$$

Forme : Plus σ est grand, plus la gaussienne est aplatie. En effet :

$$\int_{\mu-\sigma}^{\mu+\sigma} f_X(x)dx = 0.7, \int_{\mu-2\sigma}^{\mu+2\sigma} f_X(x)dx = 0.95, \int_{\mu-3\sigma}^{\mu+3\sigma} f_X(x)dx = 0.997$$

Remarque. Si $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ sont **indépendantes**, alors :

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Vecteurs gaussiens

Définition. Soit $(X_i)_{i=1}^n$ une famille de v.a.

 $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \text{ est un vecteur gaussien ssi toute combinaison linéaire des } X_i \text{ est une v.a. gaussienne}^{14}.$

Remarque. Dans ce cas, on note $X \sim \mathcal{N}(M, \Sigma)$, avec :

$$M = \mathbb{E}(X) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{pmatrix}$$

 $\Sigma = C_X = Var(X) = \text{matrice de variance-covariance} = \left(\mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^T] \right)_{1 \leq i,j \leq n}$ avec, $\forall i, j \in [1, n] : i \neq j \Rightarrow Var(X)_{i,j} = Cov(X_i, X_j)$

Proposition. Si X est gaussien, alors les v.a. marginales X_i sont gaussiennes. Réciproque fausse.

Proposition. Si les v.a. marginales X_i sont gaussiennes **et indépendantes**, alors X est gaussien et Σ est diagonale.

Proposition. Si X et Y sont des vecteurs gaussiens tels que $\mathbb{E}(X) = \mathbb{E}(Y)$ et $\Sigma_X = \Sigma_Y$, alors $X \sim Y$.

Proposition. Si X est gaussien et que Σ est définie positive ¹⁵, alors :

$$\forall x \in \mathbb{R}^n, f_{M,\Sigma}(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma}} \exp \left[\frac{-(x-M)^T \Sigma^{-1} (x-M)}{2} \right]$$

Proposition. Si X est gaussien, alors pour tout $i,j \in [1,n]$, X_i et X_j sont non corrélées ¹⁶.

^{13.} Car pour toute densité de probabilité, $\int_{\mathbb{R}} f_X(x) dx = 1$

^{14.} i.e. suivant une loi normale

^{15.} A est définie positive ssi : $\forall x \in \mathbb{R}^n : x^t A x \geq 0$, avec égalité ssi x = 0.

^{16.} i.e. $Cov(X_i, X_j) = 0$



Inégalités importantes ¹⁷

Proposition (Inégalité de Bienaymé-Tchebichev).

$$\forall \lambda \in \mathbb{R}, P(X \ge \lambda) \le \frac{\mathbb{E}(X)}{\lambda}$$

Proposition (Inégalité de Markov).

$$\forall \lambda \in \mathbb{R}, P(|X - \mathbb{E}(X)| \ge \lambda) \le \frac{V(X)}{\lambda^2}$$

Espérance conditionnelle

Définition. L'espérance conditionnelle de X sachant Y est notée $\mathbb{E}(X|Y) = \mathbb{E}_Y(X)$ et est une variable aléatoire fonction de Y.

Méthode (Cas discret). Pour $n \in \mathbb{Z}$, calcul de $\mathbb{E}_{Y=n}(X) = \sum_{k \in \mathbb{Z}} k P_{Y=n}(X=k)$

- Bayes:
$$P_{Y=n}(X=k) = \frac{P(Y=n,X=k)}{P(Y=n)} = \frac{P(Y=n \cap X=k)}{P(Y=n)}$$
- Puis, de cette expression pour tout $n \in \mathbb{Z}$, déduire une expression "en remplaçant n par Y"

Méthode (Cas continue ¹⁸). Pour $y \in \mathbb{R}$, calcul de $\mathbb{E}_{Y=y}(X) = \int_{\mathbb{R}} x f_{X|Y=y}(x) dx$

- Bayes:
$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_{Y}(y)}$$

 $- Bayes: f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ $- Puis, de cette expression pour tout <math>y \in \mathbb{R}$, déduire une expression "en remplaçant y par Y"

Convergence des suites de variables aléatoires

Soit (X_i) une suite de v.a. et X une v.a. (limite potentielle de (X_i)).

Convergence en loi

Proposition. On note ¹⁹: $X_n \xrightarrow[n \to \infty]{Loi} X$. Obtenue si:

$$\forall x \in \mathbb{R}, F_n(x) \xrightarrow[n \to \infty]{} F(x)$$

· Si pour les fonctions de répartitions : $\forall x \in \mathbb{R}, F_n(x) \xrightarrow[n \to \infty]{} F(x)$ · Si pour toute $f : \mathbb{R} \to \mathbb{R}, C^0$, bornée : $\mathbb{E}[f(X_n)] \xrightarrow[n \to \infty]{} \mathbb{E}[f(X)]$

Convergence en probabilité

Définition.

$$\forall \varepsilon > 0, P(||X_n - X|| \le \varepsilon) \xrightarrow[n \to \infty]{} 1$$

Équivalent à :
$$\forall \varepsilon > 0, P(||X_n - X|| > \varepsilon) \xrightarrow[n \to \infty]{P} 0.$$
 On note $X_n \xrightarrow[n \to \infty]{P} X$

^{17.} X,Y sont toujours des v.a., discrètes ou continues selon le contexte.

^{18.} Soyez convaincus de la parfaite équivalence des deux démarches.

La formule de Bayes doit vous évoquer tout de suite le conditionnement.

^{19.} CF page 4. En gros, Cela signifie que la convergence s'applique aux lois de probabilités

^{20.} C'est même un ssi.



Convergence presque sûrement

Définition.

$$P(\lim_{n\to\infty} X_n = X) = 1$$

Équivaut à une convergence simple sauf sur un ensemble de mesure nulle. On note $X_n \xrightarrow[n \to \infty]{P.S.} X$

Remarque. $Cv \ p.s. \Rightarrow Cv \ en \ proba \Rightarrow Cv \ en \ loi.$

Loi des grands nombres

Proposition (Loi faible des grands nombres). Soit (X_i) une suite de v.a.i.i. d^{21} :

$$\left(\mathbb{E}(X_n^2) < \infty\right) \Rightarrow \begin{cases} \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P.} \mathbb{E}(X_1) \\ \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{L^2} \mathbb{E}(X_1) \end{cases}$$

Proposition (Loi forte des grands nombres). Soit (X_i) une suite de v.a.i.i.d :

$$(\mathbb{E}(|X_1|) < \infty) \Rightarrow \left(\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \to \infty]{p.s.} \mathbb{E}(X_1)\right)$$

 $\mathbb{E}(|X_1|) < \infty$ signifie en fait que X_1 est intégrable.

Théorème centrale limite

Proposition. Soit (X_i) une suite de v.a.i.i.d telle que :

$$\begin{cases} \exists m \in \mathbb{R}, & \forall i \in \mathbb{N}, & \mathbb{E}(X_i) = m \\ \exists \sigma \in \mathbb{R}, & \forall i \in \mathbb{N}, & V(X_i) = \sigma^2 \end{cases}$$

Alors:

$$\frac{\sum_{i=1}^{n} (X_i - \mathbb{E}(X_i))}{\sigma \sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - nm}{\sigma \sqrt{n}} \xrightarrow[n \to \infty]{Loi} \mathcal{N}(0, 1)$$

Simulation sur un ordinateur

L'ordinateur ne "maîtrise" rien d'autre qu'une loi de probabilité uniforme sur [0,1].

Si on pose
$$Y = F_X^{-1}(U)$$
, avec :
$$\begin{cases} U & \sim \mathscr{U}_{[0,1]} \\ X & v.a. \end{cases}$$
 Alors : $Y \sim X$

^{21.} Variables Aléatoires Indépendantes et Identiquement Distribuées. Voir notes en page 4 pour rappel.