

Cours de probabilités

Terminale S

Pour aller plus loin ...

Paul Milan

Table des matières

1	Espace probabilisé	2
1.1	Cas où l'univers est fini	2
1.2	Cas où l'univers est infini	3
2	Probabilité conditionnelle	6
2.1	Définition	6
2.2	Formule de Bayes	7
3	Indépendance	8
4	Variable aléatoire	9
4.1	Fonction de répartition :	9
4.2	Espérance	11
4.3	Fonction d'une variable aléatoire réelle	12
4.4	Variance et écart type	13
5	Couple de variables aléatoires réelles	15
5.1	Loi d'un couple de variables	15
5.2	Somme de variables aléatoires	17
6	Lois usuelles	22
6.1	Lois normales	22
6.2	Lois binomiales $B(n, p)$	25
6.3	Lois de Poisson	29
6.4	Convergence vers une loi normale	32
A	Table de la loi normale centrée réduite	35
B	Table de la loi binomiale	36
C	Table de la loi de Poisson	40

Avant propos

Face à une épreuve aléatoire - lancement d'un dé, mesure du poids d'un nouveau né, etc. - il arrive assez souvent qu'avant toute réalisation de l'épreuve on ait certaines connaissances concernant les résultats auxquels on doit s'attendre :

Pour les deux exemples cités, on connaît l'ensemble de tous les résultats possibles et on sait, pour le lancer de dé, que chacune des 6 faces a les mêmes chances de sortir pourvu que le dé ne soit pas pipé, et, pour le poids du nouveau né, qu'on a plus de chances de tomber sur un poids voisin de 3 kg que sur un poids supérieur à 4 kg ou inférieur à 2 kg. Pour pouvoir modéliser ces connaissances on associe à une épreuve aléatoire un espace probabilisé.

1 Espace probabilisé

Un espace probabilisé se compose de trois éléments :

- Ω l'ensemble fondamental (univers), qui est l'ensemble de tous les résultats possibles e , appelés aussi événements élémentaires . Pour le lancer de dé, $\Omega = \{1, 2, 3, 4, 5, 6\}$.
Pour le poids des nouveaux nés exprimé en grammes, on peut considérer que $\Omega = [1500, 5500]$.
- \mathcal{P} : Une famille de parties de Ω . Chaque élément A de \mathcal{P} est donc une réunion d'événements élémentaires : on l'appelle un événement composé.
Par exemple : le résultat du lancer de dé est pair signifie que la réalisation e de l'épreuve appartient à l'événement composé $A = \{2, 4, 6\}$. On l'écrit :
 $e \in A \Leftrightarrow$ le résultat est pair.
- P une probabilité, fonction définie sur \mathcal{P} et à valeurs dans $[0, 1]$.
Si le dé n'est pas pipé, chacune des 6 faces du dé a la même probabilité de sortir, c'est-à-dire que P associe à chaque numéro de 1 à 6 la probabilité $\frac{1}{6}$, et à tout ensemble de k numéros la probabilité $\frac{k}{6}$.

Donnons maintenant une définition précise du triplet (Ω, \mathcal{P}, P) qu'est l'espace probabilisé, en commençant par le cas le plus simple où le nombre des résultats possibles est fini.

1.1 Cas où Ω est fini

Puisque Ω a un nombre fini d'éléments, notons $\Omega = \{e_1, e_2, \dots, e_k\}$.

P associe à chaque élément e_i de Ω une probabilité $P(e_i) = p_i$, positive ou nulle. On demande à P de vérifier la même propriété d'additivité que les fréquences, c'est-à-dire que pour tout sous ensemble A de Ω :

$$P(A) = \sum_{e_i \in A} P(e_i) \quad A \subset \Omega$$

et en particulier, la somme de toutes les fréquences vaut 1.

$$P(\Omega) = \sum_{i=1}^k P(e_i) = 1$$

On a ainsi défini P pour tout sous-ensemble de Ω , et \mathcal{P} est la famille de toutes les parties de Ω .

On remarque que si \bar{A} est le complémentaire de A dans Ω , d'après (1) et (2) : $P(\bar{A}) = 1 - P(A)$

Exemple :

- Pour le lancer d'un dé non pipé, P attribue à chacun des numéros de 1 à 6 la probabilité $\frac{1}{6}$ et à tout événement composé de k événements élémentaires la probabilité $\frac{k}{6}$.

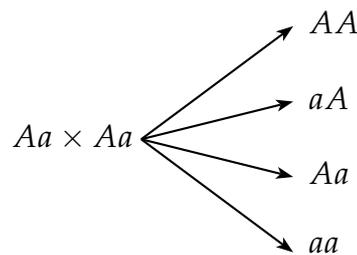
Le résultat est pair, $e \in \{2, 4, 6\}$ a donc pour probabilité $\frac{1}{2}$.

- On croise deux individus hétérozygotes Aa , A est le caractère dominant et a le caractère récessif (par exemple, la couleur des yeux, noirs pour A et bleus pour a). D'après la loi de Mendel, chacun des parents donne au hasard (c'est-à-dire avec la même probabilité) l'un ou l'autre de ses gènes.

Si l'on considère le génotype de l'individu issu du croisement, quel est l'espace probabilisé correspondant ? Et si c'est le phénotype qui nous intéresse, quel est l'espace probabilisé correspondant ?

En ce qui concerne le génotype, l'ensemble fondamental est : $\Omega = \{AA, Aa, aa\}$

Et puisque chacun des parents donne A ou a avec la probabilité $\frac{1}{2}$ et cela de façon indépendante, il y a 4 possibilités :



chacune de probabilité $\frac{1}{4}$, et comme le génotype Aa comprend les deux cas Aa et aA , on a finalement :

$$P(AA) = P(aa) = \frac{1}{4} \quad \text{et} \quad P(Aa) = \frac{1}{2}$$

Pour le phénotype, il n'y a que deux événements élémentaires A (yeux noirs) et a (yeux bleus). Or

$$A = \{AA, Aa, aA\} \quad \text{et} \quad a = \{aa\} \quad \text{donc} \quad P(A) = \frac{3}{4} \quad \text{et} \quad P(a) = \frac{1}{4}$$

1.2 Cas où Ω est infini

- a) Lorsque Ω est dénombrable, c'est-à-dire que $\Omega = \{e_i, i \in \mathbb{N}\}$, la seule différence avec le cas fini est que les $p_i = P(e_i)$ sont en quantité dénombrable, et que P doit être dénombrablement additive :

$$\text{Si } A = \bigcup_{i \in I} e_i \quad I \text{ dénombrable : } P(A) = \sum_{i \in I} P(e_i) \quad \text{et} \quad \sum_{i=0}^{+\infty} p_i = 1$$

\mathcal{P} est à nouveau la famille de toutes les parties de Ω .

Exemple :

Soit $\Omega = \mathbb{N}$ et $P(i) = \frac{\lambda^i}{i!} e^{-\lambda}$ avec $\lambda > 0$ fixé.

On appelle cette probabilité - ou loi - la loi de Poisson.

On vérifie que :

$$\sum_{i=0}^{+\infty} p_i = e^{-\lambda} + \dots + \frac{\lambda^i}{i!} e^{-\lambda} + \dots = e^{-\lambda} \left(1 + \dots + \frac{\lambda^i}{i!} + \dots \right) = 1$$

car la somme de la série qui figure dans la parenthèse vaut e^λ .

- b) Le cas où Ω n'est pas dénombrable sera pour nous presque toujours le cas où Ω est l'ensemble de tous les nombres réels \mathbb{R} , ou un intervalle de \mathbb{R} , ou le plan \mathbb{R}^2 (tout ou partie)

Contrairement aux cas fini et dénombrable, la probabilité P ne peut plus être définie par sa valeur pour chacun des événements élémentaires e :

P est définie directement sur une famille de parties \mathcal{P} , appelée tribu, et qui, dans le cas réel, comprend essentiellement les intervalles $[a; b[$.

\mathcal{P} doit avoir les propriétés suivantes :

- Ω appartient à \mathcal{P} .
- Si $A_i \in \mathcal{P}$ pour tout i de $I \in \mathbb{N}$, alors $\bigcup_{i \in I} A_i \in \mathcal{P}$
- Si $A \in \mathcal{P}$ alors $\bar{A} \in \mathcal{P}$

autrement dit, \mathcal{P} doit être fermée pour la réunion dénombrable et la complémentation.

P doit vérifier $P(\Omega) = 1$ et être dénombrablement additive :

Si $A = \bigcup_{i \in I} A_i$ pour tout i de I et $A_i \cap A_j = \emptyset$ pour tout $i \neq j$ et I dénombrable

On doit avoir :

$$P(A) = \sum_{i \in I} P(A_i)$$

Toujours dans le cas d'une variable réelle, la probabilité est le plus souvent définie grâce à une densité de probabilité, fonction positive f définie sur \mathbb{R} et telle que :

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

Alors on définit P par :

$$P(A) = \int_A f(x) dx \quad A \in \mathcal{P}$$

Il est clair que P ainsi définie sur \mathcal{P} a bien la propriété d'additivité dénombrable (ou σ -additivité).

Exemple :

On vérifiera que les fonctions suivantes définissent bien des probabilités sur \mathbb{R} :

$$\begin{cases} f(x) = \lambda e^{-\lambda x} & x > 0 \\ f(0) = 0 \end{cases} \quad \begin{array}{l} \text{Loi exponentielle de paramètre } \lambda > 0 \\ \text{loi sans mémoire des composants électroniques} \end{array}$$

$$\begin{cases} f(x) = \frac{1}{b-a} & x \in [a; b] \\ f(x) = 0 & \text{sinon} \end{cases} \quad \text{Loi uniforme sur } [a; b].$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \begin{array}{l} \text{où } \mu \text{ et } \sigma \text{ sont deux paramètres réels,} \\ \mu \text{ quelconque et } \sigma \text{ positif.} \end{array}$$

La probabilité correspondante est appelée la **loi normale** ou gaussienne $\mathcal{N}(\mu, \sigma)$.

On pourra poser $y = \frac{x-\mu}{\sigma}$ puis utiliser la valeur de l'intégrale $\int e^{-\frac{1}{2}y^2} dy$ qui vaut $\sqrt{2\pi}$

2 Probabilité conditionnelle

2.1 Définition

Dans la pratique, on est souvent conduit à évaluer la probabilité d'un événement A sachant qu'un autre événement B est réalisé. Par exemple, il est intuitif que la probabilité pour que le résultat d'un lancer de dé soit un 2 est de $\frac{1}{3}$ si l'on sait déjà que le résultat est pair, alors qu'elle vaut $\frac{1}{6}$ si l'on ne sait rien. De même, lors de l'établissement d'un diagnostic, la probabilité pour que le patient soit atteint d'une maladie donnée se modifie au fur et à mesure que les symptômes et les résultats des examens successifs sont connus.

On appelle probabilité de A sachant B , que l'on note $P_B(A)$, le rapport :

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

$P(B)$ est supposé différent de 0.

De manière générale, la probabilité pour que deux événements A et B se produisent simultanément,

$$P(A \cap B) = P_B(A) \times P(B) = P_A(B) \times P(A)$$

Il reste à vérifier que l'application $A \rightarrow P_B(A)$, où B est fixé et A parcourt \mathcal{P} est effectivement une probabilité, c'est-à-dire qu'elle vérifie :

$$P_B(\Omega) = 1 \quad \text{l'additivité : } P_B(\cup_{i \in I} A_i) = \sum_{i \in I} P_B(A_i)$$

dès que les A_i sont tous dans \mathcal{P} d'intersection vide et en quantité au plus dénombrable.

Pour la première, c'est clair car $P_B(\Omega) = P(\Omega \cap B) = P(B) = 1$. Et pour la seconde, il suffit de le vérifier lorsque I a deux éléments :

$$P_B(A_1 \cup A_2) = \frac{(A_1 \cup A_2) \cap B}{P(B)} = \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} = P_B(A_1) + P_B(A_2)$$

Remarque :

Si l'on sait déjà que B est réalisé, on ne s'intéresse qu'aux expériences qui réalisent B , et on restreint donc l'ensemble fondamental Ω à $B \subset \Omega$. On pourra ainsi définir un nouvel espace probabilisé ayant B pour ensemble fondamental, $\mathcal{P}_B = \{A \cap B ; A \in \mathcal{P}\}$ pour tribu et P_B qui est définie sur \mathcal{P}_B , incluse dans \mathcal{P} , pour probabilité.

Si, à la suite d'une première expérience, on sait que B est réalisé, on pourra prendre comme nouvel espace probabilisé

$$(B, \mathcal{P}_B, P_B)$$

Exemple :

Parmi les patients qui consultent un médecin la probabilité d'avoir la maladie M est $P(M) = 0,2$. Un symptôme S permet de détecter à coup sûr la présence de la maladie M , mais il n'apparaît pas chez tous ceux qui sont atteints de M . On sait que 10 % des patients qui se présentent à la consultation chez ce médecin présentent à la fois la maladie et le symptôme. Quelle est la probabilité pour un patient atteint de M de présenter le symptôme S , lorsqu'il se présente à la consultation chez ce médecin ?

L'énoncé implique que :

$$P(M) = 0,2$$

$$P_S(M) = 1 \quad \text{Le symptôme } S \text{ permet de détecter à coup sûr } M$$

$$P(M \cap S) = 0,1$$

Par définition de la probabilité conditionnelle

$$P_M(S) = \frac{P(M \cap S)}{P(M)} = \frac{0,10}{0,20} = 0,5$$

2.2 Formule de Bayes

Cette formule est aussi appelée « théorème de la probabilité des causes », car elle permet de renverser un conditionnement. On l'obtient en remarquant que la probabilité d'une intersection $A \cap B$ peut s'écrire soit en conditionnant A par B , soit en conditionnant B par A :

$$P(A \cap B) = P_B(A) \times P(B) = P_A(B) \times P(A)$$

Comme par ailleurs

$$P(B) = P(B \cap A) + P(B \cap \bar{A}) = P_A(B) \times P(A) + P_{\bar{A}}(B) \times P(\bar{A})$$

On obtient la formule :

$$P_B(A) = \frac{P_A(B) \times P(A)}{P_A(B) \times P(A) + P_{\bar{A}}(B) \times P(\bar{A})}$$

Valable dès que $P(B)$ est différent de 0.

Exemple :

Une maladie M se présente sous deux formes M_1 et M_2 avec les probabilités respectives $P(M_1) = 0,2$ et $P(M_2) = 0,8$, et le symptôme S apparaît dans 80 % des cas de M_1 et dans 10 % des cas de M_2 .

Quelle est la probabilité pour un patient atteint de M qui présente le symptôme S , d'être atteint de M_1 ?

$$P_S(M_1) = \frac{P_{M_1}(S) \times P(M_1)}{P_{M_1}(S) \times P(M_1) + P_{M_2}(S) \times P(M_2)} = \frac{0,8 \times 0,2}{0,8 \times 0,2 + 0,1 \times 0,8} = \frac{2}{3}$$

3 Indépendance

A et B sont deux événements relatifs à un même espace probabilisé.

On dit que deux événements A et B sont indépendants si, et seulement si :

$$P(A \cap B) = P(A) \times P(B)$$

Cette définition est donc symétrique par rapport aux deux événements qu'elle concerne. D'après la définition des probabilités conditionnelles, elle entraîne que :

$$P_B(A) = P(A) \quad (1) \quad \text{et} \quad P_A(B) = P(B) \quad (2)$$

autrement dit le fait que l'un des deux événements soit réalisé ne modifie pas la probabilité de l'autre. L'une ou l'autre des deux équations (1) et (2) aurait pu être choisie comme définissant l'indépendance de A et B. Cela aurait eu l'avantage d'être plus intuitif, mais l'inconvénient de masquer la symétrie des rôles des événements A et B.

Exemple :

Au cours de l'épreuve d'un lancer de dé à 6 faces, non pipé, on considère les trois événements :

A : le numéro sorti est strictement plus grand que 4

B : le numéro sorti est impair

C : le numéro sorti est strictement supérieur à 3

Montrer que A et B sont indépendants et que \bar{B} et C ne le sont pas. On dit qu'ils sont liés.

$$P(A \cap B) = \frac{1}{6} \quad \text{et} \quad P(A) \times P(B) = \frac{2}{6} \times \frac{3}{6} = \frac{1}{6} = P(A \cap B)$$

$$P(\bar{B} \cap C) = \frac{1}{3} \quad \text{et} \quad P(\bar{B}) \times P(C) = \frac{3}{6} \times \frac{3}{6} = \frac{1}{4} \neq P(\bar{B} \cap C)$$

4 Variable aléatoire

Nous avons vu qu'on pouvait à tout événement élémentaire e d'un espace probabilisé associer un nombre réel. Si Ω est fini ou même dénombrable, la variable X ainsi définie est une variable aléatoire réelle, notée v.a.r. .

Si Ω est un intervalle de \mathbb{R} , ou \mathbb{R} tout entier, X est encore une v.a.r. pourvu que les ensembles de la forme $\{e : X(e) \leq x\}$ fassent partie de \mathcal{P} pour tout x réel : (on demande cette propriété pour que la probabilité pour que X soit inférieure à une valeur donnée soit définie, et P est définie sur \mathcal{P}).

4.1 Fonction de répartition :

La probabilité pour qu'une variable X soit inférieure ou égale à une valeur donnée x est appelée la fonction de répartition de X au point x :

$$F(x) = P(X \leq x)$$

Elle permet de calculer facilement la probabilité pour que X soit comprise entre deux valeurs a et b :

$$P(a < X \leq b) = F(b) - F(a)$$

La **fonction de répartition** de X , ou de la loi de X est l'équivalent, pour la population toute entière, de la fonction de répartition empirique, ou fonction cumulative observée associée à un échantillon.

Lorsque X est discrète, sa fonction de répartition est égale à :

$$F(x) = \sum_{x_i \leq x} P(X = x_i) = \sum_{x_i \leq x} P_i$$

Lorsqu'elle est continue de densité de probabilité égale à f :

$$F(x) = \int_{-\infty}^x f(t) dt$$

4.1.1 Propriétés de la fonction de répartition

Les propriétés de F sont les mêmes que celles des fonctions de répartition observées, que nous avons vues au premier chapitre :

F est croissante : Si $x_2 > x_1$, $F(x_2) - F(x_1)$ est égal, par définition de F , à la probabilité pour que X appartienne à l'intervalle $]x_1; x_2]$ quantité qui est nécessairement positive ou nulle. Donc F est croissante.

F croît de 0 à 1 : En effet lorsque x tend vers $-\infty$, $P(X \leq x)$ tend vers 0, donc $F(x)$ aussi. De même, lorsque x tend vers $+\infty$, $P(X \leq x)$ tend vers 1, donc $F(x)$ aussi.

F est continue à droite en tout point : En effet soit x un réel quelconque et h un réel positif. Quand h tend vers 0, ce qu'on note $h \rightarrow 0^+$ car h est positif, $F(x+h) - F(x) = P(X \in]x; x+h])$ et $]x; x+h]$ converge vers l'ensemble vide \emptyset de probabilité nulle. Donc

$$F(x+h) - F(x) \xrightarrow{h \rightarrow 0^+} 0$$

Par contre, il peut y avoir des discontinuités à gauche car :

$$F(x) - F(x - h) = P(X \in]x - h, x])$$

et $]x - h, x] \xrightarrow{h \rightarrow 0^+} \{x\}$ et il suffit donc que $P(X = x)$ soit différent de 0 pour qu'il y ait un saut de F :

$$F(x) - F(x - h) \xrightarrow{h \rightarrow 0^+} P(X = x)$$

L'allure de la fonction de répartition d'une variable discrète est une succession de sauts :

En chaque point x_i , F saute de p_i (fig.1).

Par contre la fonction de répartition d'une variable dont la loi est définie par une densité est continue (fig.2).

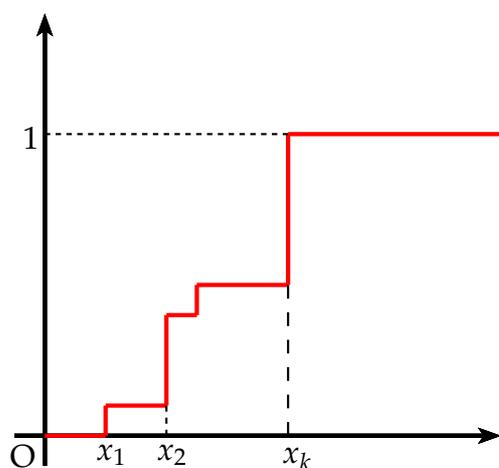


fig 1 : variable discrète

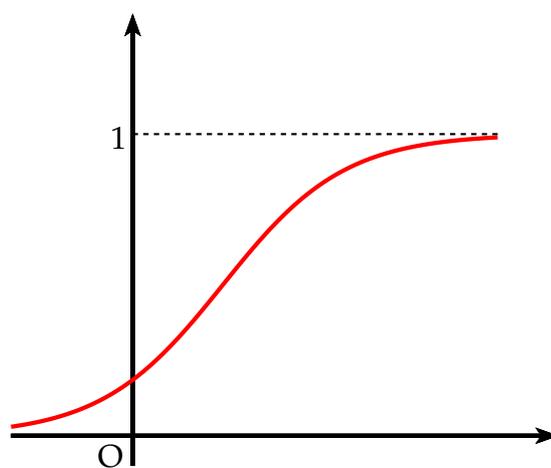
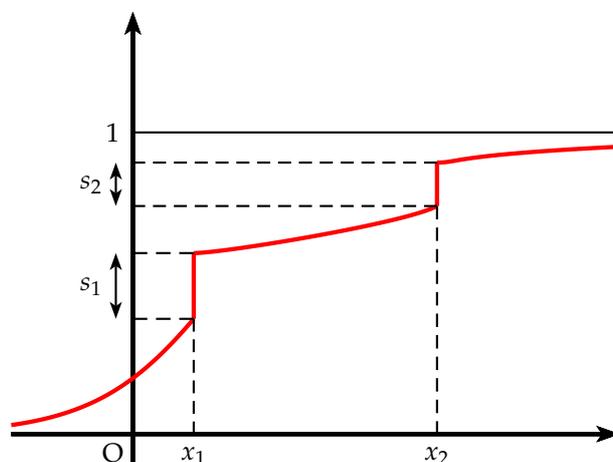


fig.2 : variable à densité

Remarque :

Au lieu d'être purement discontinue (fig.1) ou purement continue (fig.2), une variable aléatoire réelle X peut être mixte et avoir une fonction de répartition continue par morceaux et présentant un certain nombre de sauts (figure ci-dessous)



La variable dont la fonction de répartition est représentée ci-dessus est telle que $P(X = x) = 0$ si x est différent de x_1 et x_2 , et $P(X = x_1) = s_1$ et $P(X = x_2) = s_2$.

4.2 Espérance

L'espérance, ou moyenne, d'une variable aléatoire réelle X est notée $E(X)$. Si X est discrète et vaut x_i avec la probabilité p_i , variant de 1 à k

$$E(X) = \sum_{i=1}^k p_i x_i$$

Si X est continue et admet f comme densité de probabilité

$$E(X) = \int_{-\infty}^{+\infty} t f(t) dt$$

Exemple :

- 1) X est une variable de Bernoulli, qui vaut 1 avec la probabilité p et 0 avec la probabilité $q = 1 - p$

$$E(X) = p$$

X peut représenter le succès (1) ou l'échec (0) d'une intervention chirurgicale dont la proportion de réussite est p .

- 2) Y est une variable de Poisson, de paramètre λ , c'est-à-dire que Y prend toutes les valeurs entières de 0 à l'infini, avec les probabilités :

$$P(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k \in \mathbb{N}$$

$$E(Y) = \sum_{k=0}^{+\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}$$

Or on peut démontrer que la fonction e^λ peut aussi s'écrire comme la somme de la série :

$$e^\lambda = 1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^n}{n!} + \dots$$

Donc

$$E(Y) = \lambda$$

Y peut représenter par exemple le nombre de bactéries dans une suspension.

- 3) Z est une variable normale dont la loi est définie par sa densité :

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$$E(Z) = \int_{-\infty}^{+\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

Cette intégrale existe puisque la fonction à intégrer tend suffisamment vite lorsque z tend vers $\pm\infty$.

Comme la fonction à intégrer est impaire, l'intégrale est nulle : $E(Z) = 0$.

4.3 Fonction d'une variable aléatoire réelle

A partir d'une variable aléatoire X , on peut définir une nouvelle variable $Y = \varphi(X)$. Pourvu que φ soit suffisamment régulière, Y est encore une variable aléatoire réelle.

La loi de Y est appelée la loi image de celle de X par φ .

Exemple :

X est la variable aléatoire qui désigne le poids d'un nouveau né. On ne s'intéresse en fait qu'aux trois classes de poids :

- faible : inférieur ou égal à 3 kg
- moyen : entre 3 et 3,5 deuxième borne comprise
- élevé : supérieur à 3,5.

On définit ainsi une nouvelle variable Y , fonction de X , qui peut valoir par exemple :

- -1 pour les poids faibles,
- 0 pour les moyens et
- $+1$ pour les poids élevés.

Il est facile de déduire la loi de Y de celle de X :

- $P(Y = -1) = P(X \leq 3)$
- $P(Y = 0) = P(3 < X \leq 3,5)$
- $P(Y = +1) = P(X > 3,5)$

Mais si l'on s'intéresse seulement à la moyenne, ou à la variance de Y , il n'est pas nécessaire de calculer sa loi car :

Dans le cas discret : $E(Y) = \sum_{i=1}^k \varphi(x_i)P(X = x_i)$

Dans le cas continu : $E(Y) = \int \varphi(x)f(x)dx$

4.3.1 Changement d'origine et d'unité

Un cas particulier important est celui de la fonction affine, $Y = aX + b$ où a et b sont des constantes qui correspondent respectivement à un changement d'unité a et d'origine b . Alors :

$$E(aX + b) = aE(X) + b$$

Démonstration : Dans le cas d'une variable continue :

$$\begin{aligned} E(aX + b) &= \int (ax + b)f(x) dx = a \int xf(x) dx + b \int f(x) dx \\ &= aE(x) + b \end{aligned}$$

La démonstration serait tout à fait analogue dans le cas discret.

4.4 Variance et écart type

La variance d'une variable aléatoire réelle X est notée $V(X)$. C'est, par définition, l'espérance de $[X - E(X)]^2$:

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

Les deux écritures de la variance sont équivalentes (cela tient à la linéarité de l'espérance démontrée plus haut)

Exemple :

1) Pour la variable de Bernoulli X : $P(X = 1) = p$, $P(X = 0) = q = 1 - p$

$$\begin{aligned} V(X) &= E[(X - p)^2] = (1 - p)^2 p + (0 - p)^2 q = p(1 - 2p + p^2) + p^2(1 - p) \\ &= p - p^2 = pq \end{aligned}$$

2) Pour la variable de Poisson Y de paramètre λ

$$\begin{aligned} V(Y) &= E[(Y - \lambda)^2] = E[Y(Y - 1) + (1 - 2\lambda)Y + \lambda^2] \\ &= \sum_{k=0}^{+\infty} [k(k - 1) + (1 - 2\lambda)k + \lambda^2] \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda^2 e^{-\lambda} \sum_{k \geq 2} \frac{\lambda^{k-2}}{(k-2)!} + \lambda(1 - 2\lambda) e^{-\lambda} \sum_{k \geq 1} \frac{\lambda^{k-1}}{(k-1)!} + \lambda^2 e^{-\lambda} \sum_{k \geq 0} \frac{\lambda^k}{k!} \\ &= \lambda^2 + \lambda - 2\lambda^2 + \lambda^2 = \lambda \end{aligned}$$

3) Pour la variable normale Z de densité $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$

$$V(Z) = \int_{-\infty}^{+\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

Pour calculer cette intégrale, on utilise une intégration par parties, en posant $u = z$ et $dv = \frac{z}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$ et la valeur, connue, de l'intégrale $\int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$ qui vaut $\sqrt{2\pi}$. On trouve

$$V(Z) = 1$$

L'écart type d'une variable aléatoire réelle X est noté $\sigma(X)$. C'est la racine positive de la variance de X .

Les écarts types des trois variables précédentes sont donc :

- $\sigma(X) = \sqrt{pq}$ pour la variable de Bernoulli de paramètre p ,
- $\sigma(Y) = \sqrt{\lambda}$ pour la variable de Poisson de paramètre λ
- $\sigma(Z) = 1$ pour la variable normale.

4.4.1 Changement d'origine et d'unité

Un cas particulier important est celui de la fonction affine, $Y = aX + b$ où a et b sont des constantes qui correspondent respectivement à un changement d'unité a et d'origine b . Alors :

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

Démonstration :

$$\begin{aligned} V(aX + b) &= \int [ax + b - aE(X) - b]^2 f(x) dx \\ &= a^2 \int [x - E(X)]^2 f(x) dx = a^2V(X) \end{aligned}$$

4.4.2 Variable centrée réduite

En particulier, la variable $\frac{X - E(X)}{\sigma(X)}$ a pour espérance 0 et pour variance 1. Elle est appelée la **variable centrée réduite** associée à X .

Exemple :

Supposons qu'on mesure la taille des individus d'une population en centimètres ; cela donne une variable aléatoire X , d'espérance $E(X) = 170$ et de variance $V(X) = 16$, par exemple.

Si maintenant cette même mesure est faite avec comme nouvelle unité le mètre, la nouvelle variable qui en résulte est

$$X' = 10^{-2}X$$

La constante $a = 10^{-2}$ caractérise le changement d'unité.

La nouvelle unité u' est telle que $u' = \frac{1}{a}u$. Et on a bien :

$$\begin{aligned} E(X') &= aE(X) = 1,70 \\ V(X') &= a^2V(X) = 16 \times 10^{-4} \\ \sigma(X') &= a\sigma(X) = 4 \times 10^{-2} \end{aligned}$$

La variable $Y = \frac{X - 170}{4}$ est centrée ($E(Y) = 0$) et réduite ($V(Y) = 1$).

5 Couple de variables aléatoires réelles

Il arrive souvent qu'on examine simultanément plusieurs variables aléatoires sur un même individu.

Par exemple, pour établir un diagnostic, on fait l'inventaire sur le patient des symptômes cliniques apparents, puis on effectue plusieurs mesures (pouls, tension,...) et dosages (numération globulaire, taux d'urée dans le sang,...).

C'est souvent dans la mesure où la loi de cet ensemble de variables dans la population atteinte d'une maladie donnée est connue et différente de ce qu'elle est dans la population saine qu'un premier diagnostic peut être établi.

Nous n'étudierons ici que le cas de deux variables simultanées.

5.1 Loi d'un couple de variables

5.1.1 Loi jointe

Lorsque chacune des deux variables peut prendre un nombre fini de valeurs, k pour la variable X et k' pour la variable Y , la loi jointe du couple peut être présentée sous la forme d'un tableau :

X \ Y	y_1		y_j		$y_{k'}$	Total
x_1	p_{11}					p_1
x_i			p_{ij}			p_i
x_k					$p_{kk'}$	p_k
Total	q_1		q_j		$q_{k'}$	1

A l'intersection de la ligne i et de la colonne j figure la probabilité p_{ij} pour que X et Y vaille respectivement x_i et y_j :

$$p_{ij} = P(X = x_i, Y = y_j)$$

5.1.2 Lois marginales

Les lois de chacune des deux variables prises séparément peuvent se déduire de la loi jointe du couple :

- la loi de X apparaît dans la marge de droite : $P(X = x_i) = p_i$
- la loi de Y apparaît dans celle du bas : $P(Y = y_j) = q_j$.

Pour cette raison, les lois de X et Y seuls sont appelées lois marginales de X et de Y .

$$p_i = \sum_{j=1}^{k'} p_{ij} \quad \text{somme des éléments de la } i^{\text{e}} \text{ ligne}$$

$$q_j = \sum_{i=1}^k p_{ij} \quad \text{somme des éléments de la } j^{\text{e}} \text{ colonne}$$

Exemple :

La population est celle des malades atteints d'une maladie M et X et Y sont deux dosages dont les valeurs sont réparties en trois classes, faible, moyenne et forte :

- A, B et C pour X X taux d'urée
- I, II et III pour Y Y taux de cholestérol

La loi jointe du couple (X, Y) dans cette population est donnée par le tableau suivant :

X \ Y	I	II	III
A	0,00	0,01	0,03
B	0,02	0,11	0,25
C	0,05	0,22	0,31

De ce tableau, on peut déduire la probabilité pour que, dans cette population le dosage Y ait la valeur (III) : elle vaut 0,59, somme des éléments de la troisième colonne.

Remarque : Variables continues

Lorsque X et Y sont des variables continues, dont les valeurs n'ont pas été divisées en un nombre fini de classes comme dans l'exemple précédent, la loi jointe du couple (X, Y) est donnée par une densité de probabilité f fonction des deux variables x et y .

La probabilité pour que X soit compris entre deux valeurs x_i et x_{i+1} et Y entre deux valeurs y_j et y_{j+1} , simultanément vaut :

$$P(x_i < X \leq x_{i+1}, y_j < Y \leq y_{j+1}) = \int_{x_i}^{x_{i+1}} \left(\int_{y_j}^{y_{j+1}} f(x, y) dy \right) dx$$

En particulier la loi de X seul, ou loi marginale de X, a pour densité la fonction :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

De même Y a pour densité la fonction $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$

5.1.3 Lois conditionnelles

La loi de X conditionnellement à $Y = y$ est la loi de X dans la sous population telle que $Y = y$:

$$P_{Y=y_j}(X = x_i) = \frac{p_{ij}}{q_j}$$

Exemple :

En se référant à l'exemple précédent :

$$P_{Y=I}(X = A) = 0$$

$$P_{Y=I}(X = B) = \frac{2}{7}$$

$$P_{Y=I}(X = C) = \frac{5}{7}$$

5.1.4 Indépendance

X et Y sont deux variables ayant respectivement k et k' valeurs possibles. Si toutes les lois conditionnelles $\mathcal{L}_{Y=y}(X)$ sont identiques quand y varie, autrement dit si la loi de X est la même dans toutes les sous populations correspondant à une valeur particulière de Y , X et Y sont indépendantes. Il résulte que l'indépendance de X et Y s'exprime de la manière suivante :

On dit que les variables X et Y sont indépendantes si tous les événements $\{X = x_i\}$ et $\{Y = y_j\}$ sont indépendants, c'est-à-dire si :

$$P(X = x_i, Y = y_j) = P(X = x_i) \times P(Y = y_j) \quad i = 1, 2, \dots, k \quad j = 1, 2, \dots, k'$$

Dans le cas contraire, on dit que X et Y sont liées.

Par exemple, les dosages de l'exemple ci-dessus ne sont pas indépendants. Pour le prouver, il suffit de trouver deux valeurs x_i et y_j telles que :

$$P(X = x_i, Y = y_j) \neq P(X = x_i) \times P(Y = y_j)$$

C'est le cas par exemple pour C et III :

$$P(X = C, Y = III) = 0,31$$

$$P(X = C) \times P(Y = III) = 0,58 \times 0,59 = 0,3422$$

La probabilité pour que les deux dosages donnent des valeurs élevées n'est pas égale au produit des probabilités pour que l'un et l'autre aient séparément des valeurs élevées.

Remarque : Variables continues :

Lorsque la loi du couple (X, Y) est définie par sa densité $f(x, y)$ dans le plan xOy , les lois marginales étant f_X pour X et f_Y pour Y , l'indépendance se traduit par

$$f(x, y) = f_X(x) \times f_Y(y) \quad \text{pour tout } (x, y)$$

5.2 Somme de variables aléatoires

Si l'on a défini sur un même espace probabilisé un couple de variables aléatoires (X, Y) , on peut obtenir une nouvelle variable aléatoire en considérant une fonction $\varphi(X, Y)$ de ces deux variables (pourvu bien sûr que φ soit une "bonne" fonction, mais ce sera toujours le cas pour nous).

Par exemple si on relève la pression artérielle d'un patient, on obtient un couple de valeurs, et la quantité intéressante est bien souvent l'écart entre ces deux valeurs : $\varphi(X, Y) = Y - X$.

La loi de la nouvelle variable $W = \varphi(X, Y)$ est appelée la loi image de celle de (X, Y) par la fonction φ . On peut calculer cette loi, mais ce n'est pas nécessaire lorsqu'on veut simplement connaître certaines quantités relatives à la nouvelle

variable, comme par exemple son espérance, sa variance et son écart type. En effet

$$E(W) = E[\varphi(X, Y)] = \sum_{i,j} \varphi(x_i, y_j) \times P(X = x_i, Y = y_j)$$

$$V(W) = E[W - E(W)]^2 = \sum_{i,j} (\varphi(x_i, y_j) - E(W))^2 \times P(X = x_i, Y = y_j)$$

Nous allons regarder en détail l'opération la plus simple et la plus fréquente : la somme, ou différence, de variables aléatoires

On a les relations suivantes :

$$E(X + Y) = E(X) + E(Y)$$

$$V(X + Y) = V(X) + V(Y) + 2\text{cov}(X, Y) \quad \text{avec}$$

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

Si X et Y sont indépendantes : $E(XY) = E(X) \times E(Y)$ la réciproque est fautive

Ce résultat est valable que les variables X et Y soient discrètes ou continues.

Démonstration : Dans le cas discret, supposons que p_{ij} désigne la probabilité pour que $X = x_i$ et $Y = y_j$, X et Y prenant respectivement k et k' valeurs. Alors

$$\begin{aligned} E(X + Y) &= \sum_{i=1}^k \sum_{j=1}^{k'} (x_i + y_j) p_{ij} \\ &= \sum_{i=1}^k x_i \left(\sum_{j=1}^{k'} p_{ij} \right) + \sum_{j=1}^{k'} y_j \left(\sum_{i=1}^k p_{ij} \right) \\ &= \sum_{i=1}^k x_i p_i + \sum_{j=1}^{k'} y_j q_j \\ &= E(X) + E(Y) \end{aligned}$$

Dans le cas continu où la loi du couple (X, Y) est définie par sa densité $f(x, y)$ la démonstration est similaire, les signes de sommation \sum étant remplacés par ceux d'intégration \int :

$$\begin{aligned} E(X + Y) &= \iint (x + y) f(x, y) \, dx dy = \int x \left(\int f(x, y) \, dx dy \right) + \int y \left(\int f(x, y) \, dx dy \right) \\ &= \int x f_X(x) \, dx + \int y f_Y(y) \, dy = E(X) + E(Y) \end{aligned}$$

On dit que l'espérance est linéaire

$$V(X + Y) = E \left([(X + Y) - E(X + Y)]^2 \right) \quad \text{par définition de la variance}$$

$$= E \left[(X - E(X))^2 + (Y - E(Y))^2 + 2(X - E(X))(Y - E(Y)) \right]$$

par linéarité de l'espérance

$$= E \left[(X - E(X))^2 \right] + E \left[(Y - E(Y))^2 \right] + 2E \left[(X - E(X))(Y - E(Y)) \right]$$

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

Remarque :

On voit que, contrairement à l'espérance, la variance d'une somme fait intervenir un terme complémentaire, qu'on appelle la covariance de X et Y et qui s'écrit $\text{cov}(X, Y)$:

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

car la deuxième écriture de la covariance se déduit de la première par linéarité de l'espérance : $(X - E(X))(Y - E(Y)) = XY - YE(X) - XE(Y) + E(X)E(Y)$

et $E[YE(X)] = E(X)E(Y)$ puisque $E(X)$ est une constante.

De même, $E[XE(Y)] = E(Y)E(X)$, ce qui donne le résultat.

On peut se demander dans quels cas ce terme complémentaire est nul :

1) C'est vrai lorsque X et Y sont indépendantes :

En effet, prouvons qu'alors $E(XY) = E(X)E(Y)$, en faisant le calcul dans le cas continu (le calcul est tout à fait analogue dans le cas discret) :

$$E(XY) = \iint xyf(x, y) \, dx \, dy$$

Comme X et Y sont indépendantes $f(x, y) = f_X(x) \cdot f_Y(y)$

$$E(XY) = \iint xyf_X(x)f_Y(y) \, dy = \int xf_X \, dx \times \int yf_Y(y) \, dy = E(X) \times E(Y)$$

2) Mais il se peut que la covariance de X et Y soit nulle alors que les deux variables X et Y sont liées.

Exemple : Soit le couple (X, Y) de variables aléatoires ayant la loi suivante :

Y \ X	-1	0	1
0	0	$\frac{1}{3}$	0
1	$\frac{1}{3}$	0	$\frac{1}{3}$

$$q_0 = \frac{1}{3}$$

$$q_1 = \frac{2}{3}$$

$$p_{-1} = \frac{1}{3} \quad p_0 = \frac{1}{3} \quad p_1 = \frac{1}{3}$$

Les variables X et Y sont liées : $p_{-10} = 0$ et $p_{-1}q_0 = \frac{1}{9}$

Pourtant, on peut vérifier facilement que leur covariance est nulle, car :

$$E(XY) = 0 \quad \text{et} \quad E(X)E(Y) = 0$$

Le résultat à retenir est donc le suivant

On a l'implication suivante :

$$X \text{ et } Y \text{ indépendantes} \Rightarrow \text{cov}(X, Y) = 0 \text{ et donc } V(X + Y) = V(X) + V(Y)$$

La réciproque est fautive.

5.2.1 Échantillon

Soit X une v.a. réelle de moyenne μ et de variance σ^2 , sur laquelle on décide de faire n observations indépendantes, X_1, \dots, X_n . Une telle suite X_1, \dots, X_n est appelée un échantillon de taille n ou n -échantillon.

On définit une nouvelle variable aléatoire notée $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ appelée moyenne de l'échantillon.

D'après la linéarité de l'espérance, qui s'étend au cas de l'addition de plus de deux variables

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$$

De même, les variables X_i étant indépendantes :

$$V(\bar{X}) = \frac{1}{n^2} \times \sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n}$$

On voit que \bar{X} a la même moyenne μ que X mais elle est moins dispersée que X autour de cette moyenne car sa variance décroît comme l'inverse $1/n$ de la taille de l'échantillon.

Nous verrons, par la suite, que lorsque la taille n de l'échantillon croît indéfiniment, la probabilité pour que \bar{X} diffère de μ de plus qu'une petite quantité ϵ (fixée arbitrairement à l'avance), tend vers 0.

C'est ce qu'on appelle la **loi des grands nombres**.

Exemple : Soit X une variable aléatoire de Bernoulli ;

$p(X = 1) = p$, $P(X = 0) = q$, et \bar{X} la moyenne d'un échantillon de taille n , c'est-à-dire par exemple la fréquence relative du succès à terme – d'une opération du cœur.

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

$$\text{Alors } E(\bar{X}) = E(X) = p \quad \text{et} \quad V(\bar{X}) = \frac{1}{n} V(X) = \frac{pq}{n}$$

5.2.2 Droite de régression

Considérons un couple de variables (X, Y) . Dans la sous population telle que $X = x$, Y est une variable aléatoire dont la moyenne dépend de x .

Cette moyenne est appelée moyenne de Y conditionnellement à $X = x$ et notée :

$$E(Y/X = x)$$

Courbe de régression

Si à chaque valeur de x on fait correspondre la moyenne de Y lorsque $X = x$, on obtient une courbe $y = g(x)$ appelée courbe de régression de Y par rapport à x

$$g(x) = E(Y/X = x)$$

Cette courbe s'obtient en calculant, pour chaque valeur de x la moyenne de la loi conditionnelle $\mathcal{L}(Y/X = x)$. On peut l'utiliser pour prévoir y lorsqu'on connaît

x ; g a la propriété de rendre minimum l'écart entre y et sa prévision $g(x)$, lorsque cet écart est mesuré en **moyenne quadratique**

$$E \left([Y - h(x)]^2 \right) \text{ est minimum pour } h = g$$

Droite de régression

Si l'on recherche maintenant une bonne prévision de y , non pas par une fonction quelconque h de x mais par une fonction linéaire de x , soit $y = ax + b$, on cherche de même à rendre minimum

$$E \left([Y - (aX + b)]^2 \right) = E$$

En calculant les valeurs de a et b qui rendent cette expression minimum, on obtient la droite de régression de Y par rapport à x , meilleur prévision linéaire de Y quand on connaît X .

En tenant compte de la linéarité de l'espérance, cette expression vaut :

$$E = E(Y^2) - 2aE(XY) - 2bE(Y) + a^2E(X^2) + 2abE(X) + b^2$$

Pour trouver les valeurs de a et b , on doit rechercher le minimum soit :

$$\frac{\partial E}{\partial a} = -2E(XY) + 2aE(X^2) + 2bE(X) = 0$$

$$\frac{\partial E}{\partial b} = -2E(Y) + 2aE(X) + 2b = 0$$

Ces deux équations du premier degré en a et b ont pour solution

$$a = \frac{E(XY) - E(X)E(Y)}{E(X^2) - E^2(X)} \quad b = E(Y) - aE(X)$$

qui correspondent effectivement à un minimum de E .

L'équation de la droite de régression de Y par rapport à X est donc :

$$y - E(Y) = \frac{\text{cov}(X,Y)}{\sigma^2(X)} (x - E(X))$$

qui s'écrit aussi, en notant ρ le coefficient de corrélation, $\rho(X, Y) = \frac{\text{cov}(X,Y)}{\sigma(X)\sigma(Y)}$

$$\frac{y - E(Y)}{\sigma(Y)} = \rho \frac{x - E(X)}{\sigma(X)}$$

6 Lois usuelles

Nous allons voir, dans ce chapitre, trois des types de loi les plus importants :

- D'abord les **lois normales** dont le rôle de pivot est dû à ce que le résultat de l'addition d'un grand nombre de variables aléatoires indépendantes est une variable qui suit presque toujours une loi très proche d'une loi normale.
- Ensuite les **lois binomiales** qui sont les lois du nombre de succès – ou d'échecs – lors d'une expérience répétée.
- Et enfin les **lois de Poisson**, qui sont une approximation des précédentes lorsque le nombre d'expériences est grand avec une faible probabilité de voir arriver l'événement auquel on s'intéresse, qui peut être par exemple de type accidentel.

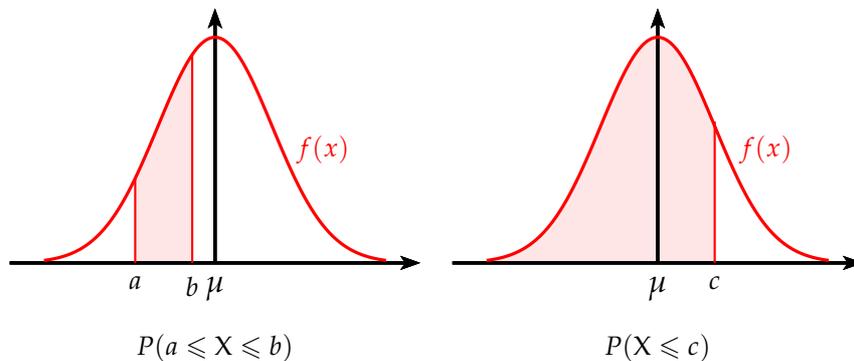
6.1 Lois normales

Si μ et σ sont deux nombres, dont le second est positif, on dit que la variable aléatoire X suit la loi normale de paramètres μ et σ^2 , notée $\mathcal{N}(\mu, \sigma^2)$, si X est une variable continue dont la densité de probabilité vaut en tout point x réel :

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

La probabilité pour que X appartienne à un intervalle $I = [a; b]$ vaut donc :

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$



La probabilité pour que X soit inférieur ou égal à une valeur c est donnée par la valeur, au point c , de la fonction de répartition F de la variable X .

$$F(c) = \int_{-\infty}^c f(x) dx$$

6.1.1 Espérance et Variance

Si une variable aléatoire X suit la loi normale $\mathcal{N}(\mu, \sigma^2)$, l'espérance et la variance de X sont respectivement μ et σ^2

Démonstration :

$$E(X) = \int_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

En faisant le changement de variable ci-dessous, on obtient :

$$z = \frac{x-\mu}{\sigma} \quad \text{et} \quad dz = \frac{1}{\sigma} dx$$

$$E(X) = \int_{-\infty}^{+\infty} \frac{(\sigma z + \mu)}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$E(X) = \int_{-\infty}^{+\infty} \sigma \frac{z}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$E(X) = \mu$$

Ces deux intégrales sont convergentes, la première est nulle, car c'est l'intégrale d'une fonction impaire, et la seconde vaut μ car :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1$$

$$V(X) = E[(x-\mu)^2] = \int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

avec le même changement de variable

$$V(X) = \sigma^2 \int_{-\infty}^{+\infty} \frac{z^2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \sigma^2 I$$

l'intégrale I se calcule par partie, en choisissant :

$$u = \frac{z}{\sqrt{2\pi}} \quad du = \frac{dz}{\sqrt{2\pi}}$$

$$dv = z e^{-\frac{z^2}{2}} dz \quad v = -e^{-\frac{z^2}{2}}$$

$$I = \left[-\frac{z}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \frac{-1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

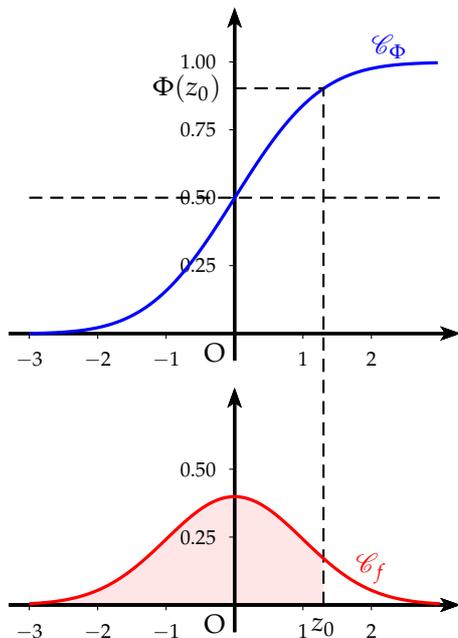
$$I = 1$$

car le premier terme vaut 0 et le second 1

Si X suit la loi normale $\mathcal{N}(\mu, \sigma^2)$, la variable $Z = \frac{X-\mu}{\sigma}$ suit la loi normale centrée réduite $\mathcal{N}(0, 1)$ de densité : $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

Il suffit donc de connaître la fonction de répartition de la loi normale réduite $\mathcal{N}(0, 1)$ pour pouvoir calculer toute probabilité relative à une variable normale $\mathcal{N}(\mu, \sigma)$.

Cette fonction de répartition intervenant très fréquemment, on lui donne un nom particulier : on l'appellera Φ .



En haut, la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0, 1)$

On remarque que :

$$\forall z, \Phi(-z) = 1 - \Phi(z)$$

La relation entre les deux figures

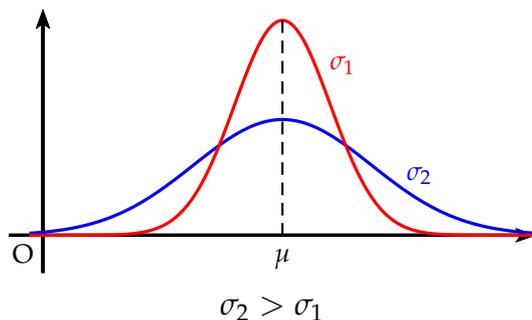
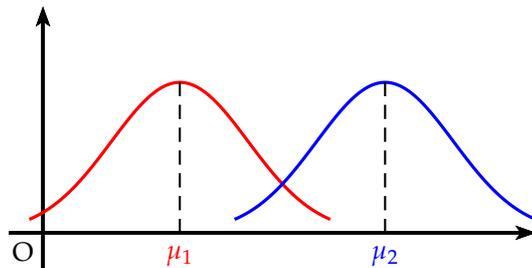
$$\Phi(z_0) = \int_{-\infty}^{z_0} f(z) dz$$

6.1.2 Comparaison de lois normales

La droite $x = \mu$ est axe de symétrie pour la courbe de la densité de la loi normale $\mathcal{N}(\mu, \sigma^2)$,

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

- Deux courbes de densités normales de même variance σ^2 , et de moyenne $\mu_2 > \mu_1$ sont deux courbes se déduisant l'une de l'autre par la translation $\mu_2 - \mu_1$ le long de l'axe Ox
- De deux courbes de densités normales de même moyenne μ et de variances différentes $\sigma_2^2 > \sigma_1^2$, la plus aplatie est celle qui correspond à la plus grande variance.



6.1.3 Somme de variables normales indépendantes

Si X et Y sont deux variables aléatoires normales indépendantes avec $\mathcal{L}(X) = \mathcal{N}(\mu, \sigma^2)$ et $\mathcal{L}(Y) = \mathcal{N}(\mu', \sigma'^2)$, la variable somme $X + Y$ est normale :

$$\left. \begin{array}{l} X \text{ et } Y \text{ indépendantes} \\ \mathcal{L}(X) = \mathcal{N}(\mu, \sigma^2) \\ \mathcal{L}(Y) = \mathcal{N}(\mu', \sigma'^2) \end{array} \right\} \Rightarrow \mathcal{L}(X + Y) = \mathcal{N}(\mu + \mu', \sigma^2 + \sigma'^2)$$

6.2 Lois binomiales $B(n, p)$

Soit une population P dont une certaine proportion p possède un caractère donné. Par exemple, p est la proportion, parmi les sujets âgés de plus de cinquante ans, de ceux qui ont une tension maximale supérieure à 18.

Si de cette population on extrait n sujets au hasard, constituant ce qu'on appelle un échantillon de taille n , on observera sur cet échantillon une proportion p_0 , généralement différente de p , d'individus présentant le caractère considéré. Si l'on répète l'expérience avec un nouvel n - échantillon, on observera une fréquence p' différente en général à la fois de p_0 et de p .

On se propose ici d'étudier la variabilité de la proportion observée p , sur un échantillon de taille n , connaissant la proportion exacte p dans l'ensemble de la population.

La résolution de ce problème permet ensuite de résoudre celui qui se pose en réalité dans la pratique (et c'est alors le point de vue statistique), où l'on ne connaît pas la proportion p dans l'ensemble de la population, mais où l'on a pu observer des proportions relatives à des groupes extraits de cette population. Le problème est alors : comment évaluer p à l'aide des proportions observées ? (cf l'exemple initial, où p n'est évidemment pas connu).

6.2.1 Schéma de l'urne

La population P peut toujours, dans le cas étudié ici d'un caractère à deux classes, être représentée par une urne contenant une proportion :

- p de boules noires (N)
- $q = 1 - p$ de boules blanches (B).

On suppose, d'une part qu'à chaque tirage chacune des boules de l'urne a la même probabilité de sortir, d'autre part que la proportion p de boules noires ne change pas d'un tirage au suivant.

Cette deuxième condition implique qu'après chaque tirage la boule sortie soit remise dans l'urne. Toutefois, si la taille n de l'échantillon est petite par rapport au nombre total de boules dans l'urne, on néglige en général de prendre cette précaution, car le tirage de n boules sans remise ne modifie pas sensiblement la composition de l'urne.

6.2.2 Question

Quelle est, avec ces hypothèses, la probabilité pour que, sur les n boules tirées, il y en ait exactement k noires ? Chaque événement élémentaire peut être représenté par une suite de n caractères dont chacun est B (boule blanche) ou N (noire) : B N N B N B... N, N figurant par exemple k fois et par suite B, $n - k$ fois. Soit S la variable aléatoire qui à chaque événement élémentaire, fait correspondre le nombre de boules noires : B N N B N B ... N \xrightarrow{S} k

Comme les n tirages sont indépendants, et que pour chacun des tirages, $P(N) = p$ et $P(B) = q$, la probabilité d'un tel événement élémentaire est le produit :

$$q \times p \times p \times q \times p \times \cdots \times p = p^k q^{n-k}.$$

Donc, tout événement élémentaire contenant exactement k boules noires, aura la probabilité : $p^k q^{n-k}$.

or il y a $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ distincts, le nombre de combinaisons de n objets pris k par k .

Par suite de l'additivité de la probabilité

$$P(S = k) = \binom{n}{k} p^k q^{n-k} \quad k \in \{0, 1, 2, \dots, n\}$$

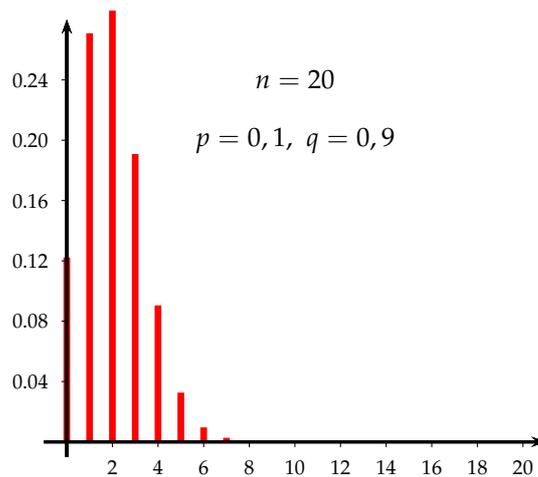
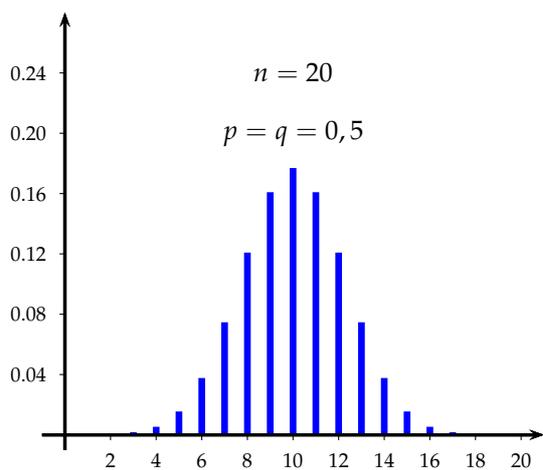
La probabilité pour que, sur les n boules tirées, il y en ait exactement k noires est :

$$P(S = k) = \binom{n}{k} p^k q^{n-k} \quad k \in \{0, 1, 2, \dots, n\}$$

Pourvu que pq soit différent de 0 (le contraire signifiant que toutes les boules de l'urne sont d'une même couleur, blanche si $p = 0$, noire si $q = 0$), la variable aléatoire S peut prendre les $n + 1$ valeurs : $0, 1, 2, \dots, n$.

Cette loi qui dépend des deux paramètres p , composition de l'urne et n , taille de l'échantillon, est appelée la loi binomiale $\mathcal{B}(n, p)$.

Exemples :



6.2.3 Moyenne et variance

Un calcul direct de la moyenne et de la variance de S , grâce aux formules de définition $\mu = \sum_{k=0}^n P(S = k)$ et $\sigma^2 = \sum_{k=0}^n (k - \mu)^2 P(S = k)$ serait pénible.

On a intérêt à considérer S comme la somme de n variables aléatoires de Bernoulli indépendantes :

$$\begin{array}{r}
 B \xrightarrow{X_1} 0 \\
 N \xrightarrow{X_2} 1 \\
 \dots\dots\dots \\
 N \xrightarrow{X_n} 1 \\
 \hline
 S = X_1 + X_2 + \dots + X_n
 \end{array}$$

L'espérance de S est donc égale à la somme des espérances des X_i , et, à cause de l'indépendance des X_i , la variance de S est égale à la somme des variances des X_i .

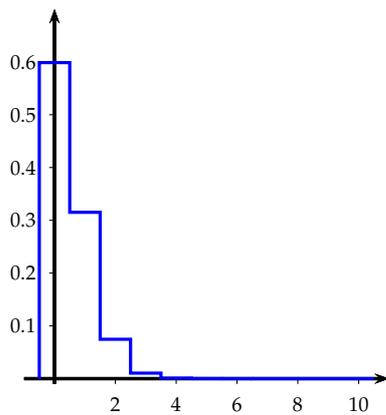
Comme $E(X_i) = p$ et $V(X_i) = pq$ pour tout i ,

$$E(S) = np \quad V(S) = npq$$

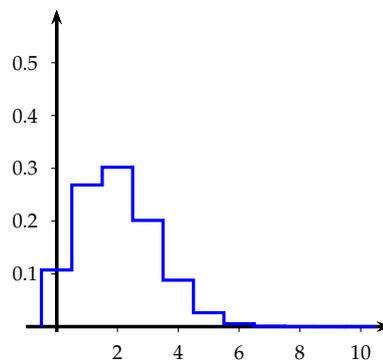
Soit S^* , la **variable aléatoire centrée réduite** qui correspond à S : (p_0 proportion observable)

$$S^* = \frac{S - E(S)}{\sqrt{V(S)}} = \frac{S - np}{\sqrt{npq}} = \frac{p_0 - p}{\sqrt{\frac{pq}{n}}} \quad \text{où } p_0 = \frac{S}{n}$$

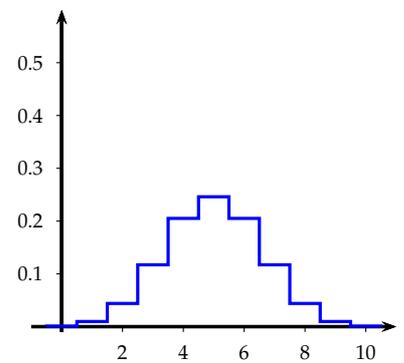
Le figure ci-dessous représente l'histogramme de S^* pour des valeurs croissantes de n (10, 30, 90) et trois valeurs différentes de p (0,05 ; 0,2 ; 0,5). Sur cette figure pour chaque valeur fixée de p , la forme de l'histogramme parait se rapprocher de plus en plus, au fur et à mesure que la taille de l'échantillon augmente, de la courbe en cloche, qui est la densité de la loi normale centrée réduite $\mathcal{N}(0,1)$



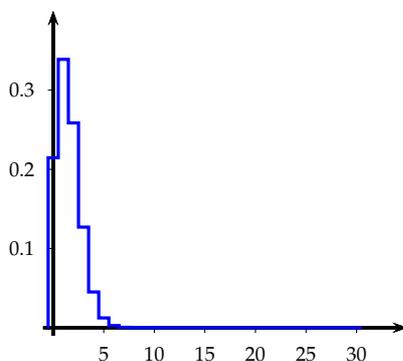
$n = 10$ et $p = 0.05$



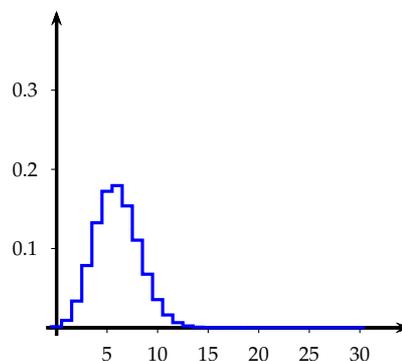
$n = 10$ et $p = 0.2$



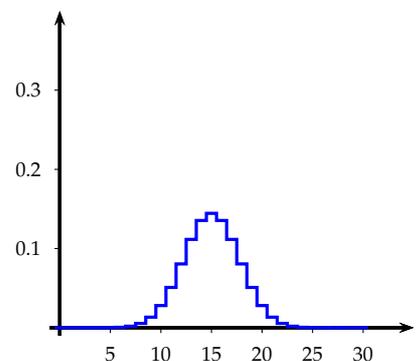
$n = 10$ et $p = 0.5$



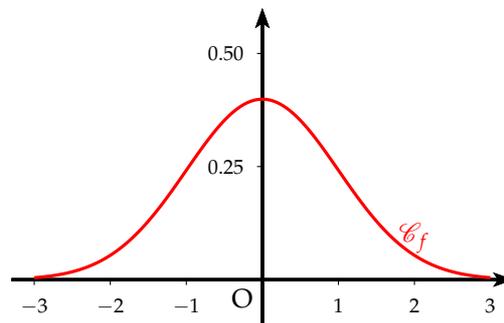
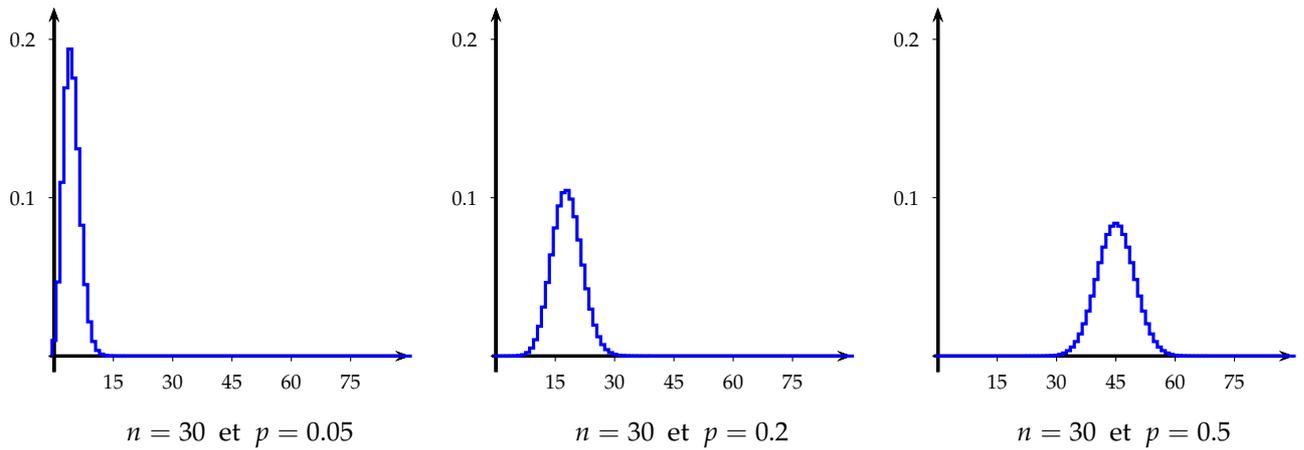
$n = 30$ et $p = 0.05$



$n = 30$ et $p = 0.2$



$n = 30$ et $p = 0.5$



Densité de la loi normale centrée réduite

On peut remarquer que cette tendance vers la loi normale est beaucoup plus lente pour $p = 0,05$ que pour $p = 0,5$. De façon générale, la tendance vers la loi normale est d'autant plus lente que p est plus éloigné de $0,5$. Si p est voisin de 0 ou de 1 , n doit être très grand pour que l'approximation soit acceptable. On peut considérer que l'approximation est assez bonne pourvu que np et nq soient simultanément supérieurs à 5 .

6.2.4 Approximation normale des lois binomiales

Soit S est une variable binomiale $\mathcal{B}(n, p)$, avec $np \leq 5$ et $nq \leq 5$

alors la variable $\frac{S - np}{\sqrt{npq}}$ a une fonction de répartition proche de celle de la loi normale centrée réduite $\mathcal{N}(0, 1)$

$$P\left(\frac{S - np}{\sqrt{npq}} \leq z\right) \simeq \Phi(z)$$

Remarque : Il se peut par contre que n soit grand et cependant p trop petit pour qu'on soit dans les conditions de l'approximation normale :

Cela se produit par exemple lorsque qu'on considère le nombre d'accidents provoqués par un vaccin, le nombre de suicides dans une grande ville, pour une période donnée.

On peut à ce moment là approcher la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson de paramètre np .

6.2.5 Approximation de Poisson des lois binomiales

Soit S une variable binomiale $\mathcal{B}(n, p)$, $p < 0,1$ et n assez grand pour que $1 \leq np \leq 10$, la loi de S ne dépend plus que de ce produit np , On a alors l'approximation suivante :

$$P(S = k) \simeq \frac{(np)^k}{k!} e^{-np} \quad k = 0, 1, 2, \dots, n$$

Remarque : La loi de S est pratiquement la même que l'on ait fait 10 observations d'un phénomène de probabilité $\frac{1}{10}$ ou 100 observations d'un phénomène de probabilité $\frac{1}{100}$.

Lorsque np est compris entre 5 et 10, il semble que l'on ait droit aux deux approximations : celle de Poisson est d'autant meilleure, et donc préférable, que p est plus proche de 0.

Démonstration :

$$\begin{aligned} P(S = k) &= \frac{n(n-1) \dots (n-k+1)}{k!} p^k (1-p)^{n-k} \\ &= \frac{(np)^k}{k!} \times \frac{n(n-1) \dots (n-k+1)}{n \times n \times \dots \times n} \times (1-p)^n \frac{1}{(1-p)^k} \end{aligned}$$

Lorsque n est grand et k petit par rapport à n , le second facteur qui vaut $1 \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times \dots \times \left(1 - \frac{k-1}{n}\right)$ est proche de 1. Il en est de même du quatrième et dernier facteur, le logarithme du troisième vaut $n \ln(1-p) \simeq -np$ puisque p est petit, et donc $(1-p)^n \simeq e^{-np}$, ce qui achève de prouver que :

$$P(S = k) \simeq \frac{(np)^k}{k!} e^{-np}$$

6.3 Lois de Poisson

Une variable aléatoire X suit la loi de Poisson de paramètre $\lambda > 0$, si elle peut prendre toutes les valeurs entières, 0 compris : $0, 1, 2, \dots, n, \dots$. La probabilité pour qu'elle prenne la valeur k est :

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \dots$$

On vérifie que $\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1$. En effet, on sait que pour tout u , e^u est égal à la somme de la série de terme général u :

$$e^u = 1 + u + \frac{u^2}{2!} + \dots + \frac{u^n}{n!} \dots$$

Dans ces conditions :
$$\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \left(\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \right) e^{-\lambda} = e^{\lambda} \times e^{-\lambda} = 1$$

6.3.1 Conditions d'application

La loi de Poisson intervient chaque fois que, sur un échantillon de grande taille, on étudie le nombre de fois que se produit un événement de faible probabilité. Par exemple, c'est le cas dans :

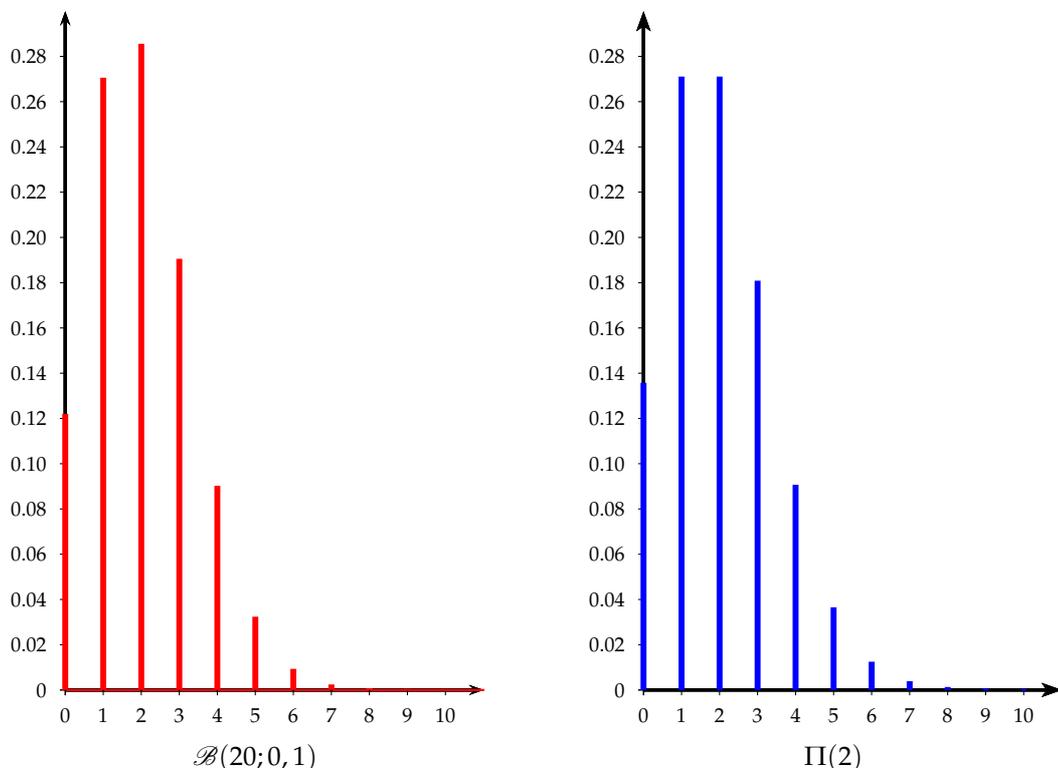
- la numération globulaire, l'étude numérique des cultures microbiennes.
- l'occupation des lignes téléphoniques et les bouchons de la circulation routière.
- les accidents mortels, les maladies exceptionnelles.
- le contrôle de fabrication à très faible proportion de rebut.

6.3.2 Allure de la courbe représentant la loi de probabilité

a) Pour les petites valeurs de X , la probabilité décroît très vite quand k augmente. Par exemple :

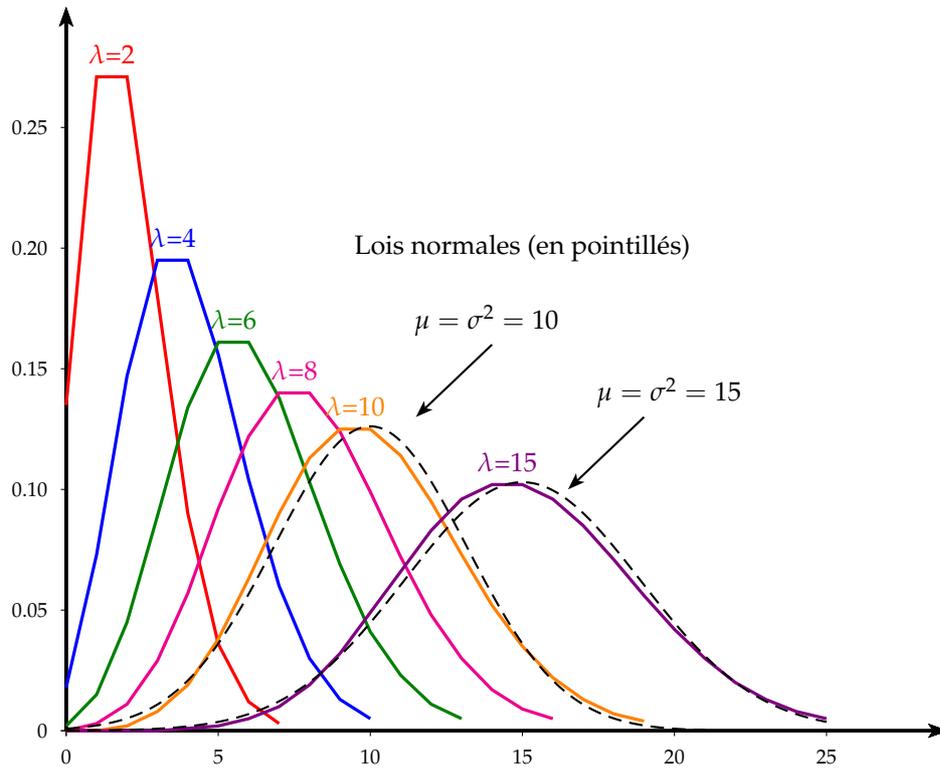
- pour $\lambda = 1$ $P(X = 6) = 0,0005$ et $P(X = 7) = 0,0001$
- pour $\lambda = 2$ $P(X = 6) = 0,0120$ et $P(X = 7) = 0,0034$

Comparons les 2 figures ci-après : les deux lois qu'elles représentent – binomiale $\mathcal{B}(n = 20, p = 0,1)$ et Poisson ($\Pi(\lambda = 2)$) – sont très voisines, comme on s'y attendait puisque $np = \lambda = 2$ et p est petit.



b) Pour les grandes valeurs de λ :

La loi de Poisson, dissymétrique avec étalement vers la droite pour les petites valeurs de λ ($\lambda \leq 6$), devient presque symétrique et "en cloche" à partir de $\lambda = 15$.

Évolution de la loi de Poisson quand le paramètre λ augmente

A partir de la valeur $\lambda = 15$, la loi de Poisson peut être assimilée, avec une erreur en général négligeable, à la loi normale vers laquelle elle tend quand λ augmente indéfiniment.

6.3.3 Moyenne et variance

Nous reprenons ici le calcul qui a déjà été fait au paragraphe 1.2
Par définition de l'espérance

$$E(X) = \sum_{k=0}^{\infty} k P(X = k) = \sum_{k=1}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

On rappelle le résultat : $e^{\lambda} = 1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^n}{n!} + \dots$

Dans le calcul de la variance, pour simplifier le calcul on considère $E[X(X-1)]$ au lieu de $E[(X-\lambda)^2]$:

$$E[X(X-1)] = \sum_{k=0}^{\infty} \frac{k(k-1)}{k!} \lambda^k e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 = E(X^2) - E(X)$$

Donc $E(X^2) = \lambda^2 + \lambda$ et $V(X) = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

La moyenne et la variance d'une loi de Poisson de paramètre λ sont toutes deux égales au paramètre.

$$E(X) = V(X) = \lambda$$

Exemple : Durant la semaine qui suit le suicide d'une actrice de cinéma, un journaliste constate qu'il y a eu dans sa ville 12 suicides, alors qu'en moyenne il n'y en a ordinairement que 8. Il attribue les morts supplémentaires à l'effet de suggestion produit par le suicide de l'actrice. Son explication est-elle convaincante ?

Soit X le nombre de suicides commis dans la ville durant une semaine donnée. Considérant que sur l'ensemble des habitants de la ville, le fait de se suicider est un phénomène relativement rare, de probabilité p au plus égale à quelques centièmes, on peut appliquer l'approximation de Poisson et prendre pour loi de X la loi de Poisson de moyenne – donc de paramètre – $\lambda = 8$.

Avec une calculatrice ou sur une table donnant la loi de Poisson de paramètre $\lambda = 8$, on trouve :

$$P(X \geq 12) = 1 - P(X \leq 11) = 1 - 0,8881 = 0,1119 \simeq \frac{1}{9}$$

Autrement dit, il y aurait 12 suicides ou davantage, environ une semaine sur 9. Le nombre observé pourrait donc bien être dû au hasard, sans qu'il soit besoin d'invoquer un effet de suggestion !

6.4 Convergence vers une loi normale

X étant une variable aléatoire de moyenne μ et de variance σ^2 dont on a observé n réalisations indépendantes $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, on est souvent amené à calculer la moyenne des observations :

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad \text{de réalisation} \quad \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

L'espérance de \bar{X} est égale à celle de X , mais la variance de \bar{X} est n fois plus petite que celle de X :

$$E(\bar{X}) = \mu \quad \text{et} \quad V(\bar{X}) = \frac{\sigma^2}{n}$$

Théorème central limite : Lorsque n tend vers l'infini, la variable réduite $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ a une loi qui tend vers la loi normale centrée réduite $\mathcal{N}(0, 1)$:

$$P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq t\right) \simeq \Phi(t) \quad t \in \mathbb{R}$$

Remarque importante : la valeur de n à partir de laquelle l'approximation peut être considérée comme valable dépend de la loi commune des observations X_i :

- Pas de problème si les X_i sont normales : \bar{X} est alors exactement normale, et cela quel que soit n .
- Si les variables X_i sont des variables de Bernoulli (à deux valeurs 0 et 1), l'approximation est valable dès que np et nq sont au moins égaux à 5.
- Pour les variables continues usuellement rencontrées en biologie et en médecine, il suffit généralement que n soit supérieur ou égal à 30.

6.4.1 Utilisation pratique du théorème de convergence

Toute moyenne observée peut être assimilée à la réalisation d'une variable normale de mêmes espérance et variance qu'elle (pourvu, bien sûr que le nombre des observations soit suffisamment grand) :

$$\bar{X} \simeq E(\bar{X}) + \sigma(\bar{X}) \times Z$$

où Z est une variable de loi $\mathcal{N}(0, 1)$. Si $E(X) = \mu$ et $V(X) = \sigma^2$, cela s'écrit

$$\bar{X} \simeq \mu + \frac{\sigma}{\sqrt{n}} Z$$

En particulier si les X sont de Bernoulli et donc \bar{X} une proportion observée, qu'on notera alors p_{obs}

$$p_{\text{obs}} \simeq p + \sqrt{\frac{pq}{n}} Z$$

Car alors $P(X = 1) = p$, $P(X = 0) = 1 - p = q$ et $E(X) = p$, $V(X) = pq$

Exemple : Intervalle de pari (de fluctuation) au risque α

Étant donnée une variable aléatoire Y dont on connaît la loi P , on peut calculer le **risque** α que l'on prend lorsqu'on parie, avant d'en observer une réalisation, que cette réalisation va tomber dans un intervalle I . En identifiant ce risque à la probabilité de se tromper, on a :

$$\alpha = P(Y \notin I) \quad I \text{ est un intervalle de pari au risque } \alpha \text{ pour } Y.$$

Lorsque Y est une moyenne observée sur un échantillon de taille n d'une variable X , il n'est pas nécessaire de connaître la loi de X mais seulement son espérance $E(X) = \mu$ et sa variance $V(X) = \sigma^2$ pour donner un intervalle de pari, de risque choisi α , pour $Y = \bar{X}$ dès que n est assez grand.

Supposons par exemple que le poids X des nouveau-nés ait pour espérance 3,500 kg et pour écart type $\sigma = 0,2$ kg dans la population française. Si l'on veut parier sur le poids moyen de 100 nouveau-nés pris au hasard en prenant le risque de ne se tromper qu'une fois sur cent, on peut choisir comme intervalle de pari

$$I = \left[\mu - 2,576 \frac{\sigma}{\sqrt{n}} ; \mu + 2,576 \frac{\sigma}{\sqrt{n}} \right] = [3,5 - 2,576 \times 0,02 ; 3,5 + 2,576 \times 0,02] \\ \simeq [3,45 ; 3,55]$$

On rappelle que :

$$P(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha \quad \Leftrightarrow \quad u_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$
$$P(-u_\alpha \leq Z \leq u_\alpha) = 0,99 \quad \Leftrightarrow \quad \Phi^{-1}(0,995) \simeq 2,576$$

