L'essentiel pour comprendre les TDs

Hadrien Bigo-Balland

Probabilités : les bases

Définition

Un **événement aléatoire** est quelque chose qui peut se produire. On cherche à savoir quelle est la chance qu'il se produise.

> *Par exemple, "mon bus arrive dans 2 min", "le médicament va marcher sur ma femme" ou "je vais avoir une calvitie avant 30 ans" sont des événements aléatoires.*

Les matheux sont flemmards, pour éviter d'écrire beaucoup de mots, on note souvent un événement par une lettre. L'événement "mon bus arrivera dans 2 min" devient "A" par exemple.

Définition

La **probabilité d'un événement** est un nombre entre 0 et 1 qui dit à quel point il a de chance de se produire.

> *Probabilité de 0 : l'événement ne se passe jamais, de 1 : l'événement se passe presque sûrement. De 0.5 : 50% de chance que l'événement se produise.*

Définition

L'**événement contraire** d'un événement s'appelle le complémentaire, noté avec une barre au dessus, $\overline{A}$.

> *Le complémentaire de "mon bus arrivera dans 2 min" est "mon bus n'arrivera pas dans 2 min".*

⚠ Il faut faire attention à bien écrire le contraire de l'événement. Le complémentaire de l'événement "je vais avoir une calvitie avant 30 ans" n'est pas "je ne vais pas perdre de cheveux avant 30 ans", mais bien "je n'aurais pas de calvitie avant 30 ans".

> *Si un événement est très probable, on s'attend à que son complémentaire ne le soit pas, et vice versa. Par exemple, "gagner au loto" est un événement très peu probable, donc "perdre au loto" est un événement très probable.*

En fait, on a

$$P(\overline{A}) = 1 - P(A)$$

Méthode

Comment calculer la probabilité d'un événement, dans le cas où je connais tous les événements possibles et qu'ils ont la même chance de se produire ?

$$P(\text{événement}) = \frac{\text{nombre de cas favorables à l'événement}}{\text{nombre total}}$$

> *Si j'ai un sachet avec 2 bonbons à la menthe, 3 à l'orange et 5 au citron, et que j'ai autant de chance de piocher chaque bonbon, la probabilité que je pioche un bon bon à la menthe est nombre de bonbons à la menthe / nombre total de bonbons =* $2/(2+3+5)$.

Intersection d'événements

Il arrive que j'ai plusieurs événements et je veuille étudier la probabilité qu'ils arrivent en même temps.

> *Quelle est la probabilité que je gagne au Loto ET à l'Euromillion le même jour ?*

Définition

C'est ce qu'on appelle l'**intersection de deux événements** :

$$A \cap B.$$

La probabilité $P(A \cap B)$ est appelée **probabilité conjointe** des deux événements A et B.

Décomposer un événement : la formule des probabilités totales

Parfois, pour calculer la probabilité d'un événement, il est utile de le découper selon plusieurs cas possibles, qui forment une partition. Cela permet de simplifier le calcul en considérant des situations plus simples, puis en les additionnant.

> *Quelle est la probabilité que je sois en retard, sachant que je peux venir en bus, en vélo ou à pied ?*

Chaque jour, je prends un seul moyen de transport :

- B_1 : je prends le bus,
- B_2 : je viens à vélo,
- B_3 : je viens à pied.

Ces événements sont disjoints deux à deux et couvrent toutes les possibilités : ils forment une partition. Le retard R peut donc survenir de différentes façons :

- je suis en retard et j'ai pris le bus : $R \cap B_1$,
- je suis en retard et je suis venu à vélo : $R \cap B_2$,
- je suis en retard et je suis venu à pied : $R \cap B_3$.

Puisque ces cas sont exclusifs, la probabilité d'être en retard peut s'écrire comme une somme :

$$P(R) = P(R \cap B_1) + P(R \cap B_2) + P(R \cap B_3).$$

Supposons que l'on dispose directement des probabilités conjointes suivantes :

- $P(R \cap B_1) = 0{,}20$,
- $P(R \cap B_2) = 0{,}03$,
- $P(R \cap B_3) = 0{,}01$.

Alors, on peut simplement écrire :

$$\begin{aligned} P(R) &= P(R \cap B_1) + P(R \cap B_2) + P(R \cap B_3) \\ &= 0{,}20 + 0{,}03 + 0{,}01 = 0{,}24. \end{aligned}$$

> *On additionne ici toutes les façons possibles d'être en retard, selon le moyen de transport utilisé. Cette approche fonctionne dès qu'on peut découper un événement global en plusieurs cas exclusifs plus simples.*

Théorème

Formule des probabilités totales — version conjointe.
Soit A un événement, et soit $(B_1, B_2, \dots, B_n)$ une partition de l'univers, c'est-à-dire une famille d'événements deux à deux disjoints tels que :

- $B_i \cap B_j = \emptyset$ pour tout $i \neq j$,
- $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$.

Alors, la probabilité de A peut se décomposer comme :

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n).$$

Union d'événements

Maintenant, pour reprendre l'exemple des jeux d'argent, on peut aussi se dire : en vrai, si je gagne déjà un des deux jeux, c'est déjà pas mal.

> *Quelle est la probabilité que je gagne AU MOINS un des deux jeux de hasard (voir les deux) ?*

Définition

C'est ce qu'on appelle **l'union de deux événements** :

$$A \cup B$$

Dans les cas les plus faciles, on peut calculer la probabilité de l'union de deux événements et la probabilité de l'intersection de deux événements avec du bon sens. Par exemple, si on prend un jeu de 52 cartes, et qu'on se demande

> *Quelle est la probabilité de tirer une reine ou une carte rouge ?*

On peut y arriver sans formule, juste en comptant.

> *Il y a 4 reines dans le jeu, et 26 cartes rouges (13 cœurs + 13 carreaux).*

Mais attention ! Les reines rouges (la reine de cœur et la reine de carreau) sont comptées deux fois si on additionne bêtement 4 + 26. Donc il faut faire : Nombre total de cartes qui sont soit des reines, soit rouges = 4 (reines) + 26 (rouges) - 2 (reines rouges) = 28. Donc la probabilité = 28 / 52. Et si maintenant on se demande :

> *Quelle est la probabilité de tirer une carte qui est à la fois une reine ET rouge ?*

C'est simple, il y'en a 2 (reine de cœur et reine de carreau), donc la probabilité = 2 / 52.

⚠️ Attention à bien enlever les éléments en double (donc faire $-P(A \cap B)$) quand vous calculez la probabilité de l'union $P(A \cup B)$ en additionnant simplement $P(A)$ et $P(B)$. Dans notre exemple on enlevait bien les deux reines rouges. Voir la propriété qui suit.

Propriété

Voici la **formule pour l'union en général** :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Événements indépendants et incompatibles

Définition

Deux événements sont dits **indépendants** si la réalisation de l'un n'apporte aucune information sur la réalisation de l'autre.

> *Les événements "j'ai révisé" et "je réussi le contrôle" ne sont pas indépendants.*

Quand deux événements sont indépendants, on peut utiliser la formule $P(A \cap B) = P(A) \times P(B)$ (on va voir pourquoi après).

Définition

Enfin, deux événements sont **incompatibles** si ils ne se produisent jamais en même temps :

$$P(A \cap B) = 0$$

> *Au jeu de pile ou face, les événements "Je suis tombé sur pile au premier lancer" et "Je suis tombé sur face au premier lancer" sont incompatibles. Soit on tombe sur pile, soit sur face.*

Probabilités conditionnelles

Définition

La **probabilité conditionnelle** d'un événement aléatoire B "sachant un événement aléatoire A" est la probabilité que B se réalise sachant que A s'est réalisé avec certitude. On la note

$$P_A(B)$$

et elle est définie par la formule

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

> *Imaginons que Jules se soit inscrit sur Tinder et notont les événements A : Jules reçoit une réponse après avoir envoyé un message et B : Jules a matché avec quelqu'un. En général, Jules match dans 10% des cas et quand il envoie un message après, il reçoit une réponse dans 30% des cas. Donc $P(B) = 0.1$ et $P(A) = 0.3$. Vu qu'on ne peut pas envoyer de message sans avoir match avant, la probabilité de B sachant A vaut 1 : $P_A(B) = 1$.*

On a vu précédemment que si deux événements sont indépendants, $P(A \cap B) = P(A) \times P(B)$. En fait, la formule qui est vraie même si les deux événements ne sont pas indépendants est

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

En gros, la probabilité que A et B se produisent, $P(A \cap B)$, est égale à la probabilité que A se produise, $P(A)$, fois la probabilité que B se produise sachant que A s'est produit.

> *Considérons une boîte contenant :*
>
> - *3 chocolats noirs,*
> - *2 chocolats au lait.*
>
> *On tire successivement deux chocolats sans remise. Définissons les événements :*
>
> - *A : "Le premier chocolat tiré est au lait",*
> - *B : "Le deuxième chocolat tiré est au lait".*

Ces événements ne sont clairement pas indépendants.

On cherche à calculer $P(A \cap B)$, c'est-à-dire la probabilité de tirer deux chocolats au lait. Commençons par calculer $P(A)$. Il y a 5 chocolats au total, dont 2 au lait. Donc

$$P(A) = \frac{2}{5}.$$

Maintenant, pour la probabilité conditionnelle $P_A(B)$. Si le premier chocolat tiré est au lait, il ne reste plus qu'1 chocolat au lait sur 4 chocolats. Alors :

$$P_A(B) = \frac{1}{4}$$

En appliquant la formule,

$$P(A \cap B) = \mathbb{P}(A) \times P_A(B) = \frac{2}{5} \times \frac{1}{4} = \frac{1}{10}.$$

Si on calculait la probabilité $P(A \cap B)$ à la main avec la formule $P(\text{événement}) = \frac{\text{nombre de cas favorables à l'événement}}{\text{nombre total}}$, on trouverait aussi $\frac{1}{10}$.

Formule des probabilités totales (version conditionnelle)

On a vu que si un événement A peut survenir via plusieurs "chemins" exclusifs $B_1, ..., B_n$, alors :

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n).$$

Mais grâce à la formule suivante :

$$P(A \cap B_i) = P_{B_i}(A)P(B_i),$$

on peut écrire la version conditionnelle de la formule des probabilités totales :

$$P(A) = P_{B_1}(A) + P_{B_2}(A) + \cdots + P_{B_n}(A).$$

Définition

C'est ce qu'on appelle la **formule des probabilités totales en version conditionnelle.**

Reprenons l'exemple du retard, avec les probabilités suivantes :

- $P(B_1) = 0{,}5$, $P(B_2) = 0{,}3$, $P(B_3) = 0{,}2$,
- $P_{B_1}(R) = 0{,}4$, $P_{B_2}(R) = 0{,}1$, $P_{B_3}(R) = 0{,}05$.

Alors :

$$\begin{aligned} P(R) &= P_{B_1}(R)P(B_1) + P_{B_2}(R)P(B_2) + P_{B_3}(R)P(B_3) \\ &= 0{,}4 \times 0{,}5 + 0{,}1 \times 0{,}3 + 0{,}05 \times 0{,}2 \\ &= 0{,}24. \end{aligned}$$

> *C'est une approche très puissante quand on connaît bien les cas particuliers.*

Indépendance avec les probabilités conditionnelles

Propriété

Deux événements aléatoires A et B sont indépendants si

$$P_A(B) = P(B)$$

ou encore

$$P_B(A) = P(A).$$

> *Cette formule signifie : que A se produise avant ou non, B a la même chance de se produire. En gros, la réalisation de A n'a pas d'impact sur la réalisation de B, ce qui est exactement l'indépendance. À partir de cette formule, on retrouve que A et B sont indépendants si $P(A \cap B) = P(A) \times P(B)$.*

Les tableaux de probabilité (tableaux de contingence)

Définition

Un **tableau de probabilité** (ou **tableau de contingence**) est un outil permettant d'organiser et de visualiser les probabilités associées à deux variables aléatoires, généralement croisées selon deux critères (comme l'âge et l'avis dans le TD).

> *Les tableaux sont particulièrement utiles pour calculer des probabilités marginales, conditionnelles ou totales, et pour appliquer le théorème de Bayes.*

Comment lire un tableau de probabilité

	Événement A	$\overline{A}$	**Total**
Événement B	$P(A \cap B)$	$P(\overline{A} \cap B)$	$P(B)$
$\overline{B}$	$P(A \cap \overline{B})$	$P(\overline{A} \cap \overline{B})$	$P(\overline{B})$
Total	$P(A)$	$P(\overline{A})$	1

Propriété

Chaque case d'un tableau croisé indique la **probabilité conjointe** de deux événements. En additionnant les cases sur une ligne ou une colonne, on obtient les **probabilités marginales** (c'est la formule des probabilités totales) :

$$P(A) = \sum_i P(A \cap B_i) \quad ; \quad P(B) = \sum_j P(A_j \cap B)$$

> *Les probabilités conditionnelles peuvent être calculées en divisant une probabilité conjointe par une probabilité marginale, c'est à dire en divisant une case par le total sur une ligne ou colonne :*
>
> $$P_A(B) = \frac{P(A \cap B)}{P(A)} \quad ; \quad P_B(A) = \frac{P(A \cap B)}{P(B)}$$

⚠ Assurez-vous que la somme de toutes les probabilités du tableau est égale à 1. Sinon, il y a une erreur.

Exemple concret : avis sur le concours de l'internat

	Pour le concours (I)	**Contre le concours** ($\overline{I}$)
Moins de 22 ans (A)	7.8%	32.2%
Plus de 22 ans ($\overline{A}$)	18.2%	41.8%

Interprétation et calculs

- $P(A \cap I) = 0.078$
- $P(A \cap \overline{I}) = 0.322$
- $P(\overline{A} \cap I) = 0.182$
- $P(\overline{A} \cap \overline{I}) = 0.418$

$$P(I) = 0.078 + 0.182 = 0.26 \quad ; \quad P(A) = 0.078 + 0.322 = 0.40$$

Quelle est la probabilité qu'un étudiant soit pour le concours sachant qu'il a moins de 22 ans ? On applique :

$$P_A(I) = \frac{P(A \cap I)}{P(A)} = \frac{0.078}{0.40} = 0.195 = 19.5\%$$

Méthode

Pour utiliser un tableau de probabilité :

1. Identifiez clairement les variables croisées
2. Complétez le tableau avec les probabilités conjointes
3. Additionnez lignes ou colonnes pour obtenir les marginales
4. Utilisez les formules pour obtenir des probabilités conditionnelles

Les arbres de probabilité

Définition

Un **arbre de probabilité** est une représentation graphique qui permet de visualiser les différentes possibilités lors d'expériences aléatoires successives, ainsi que leurs probabilités.

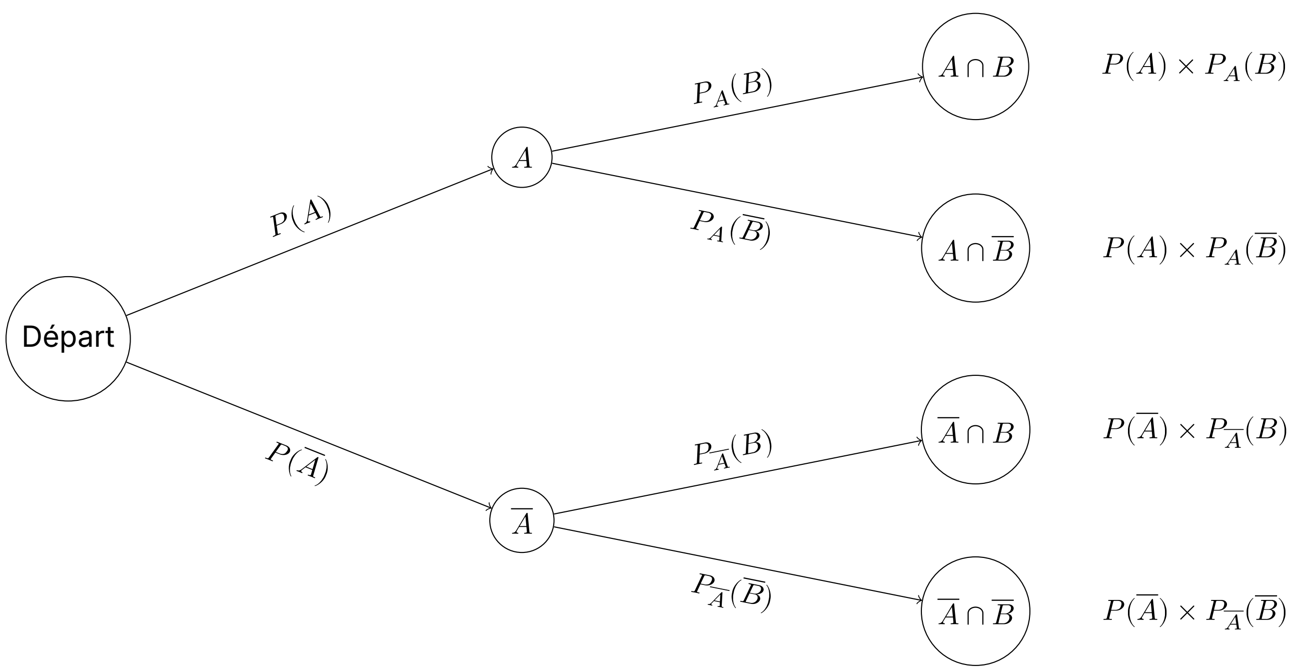
> *Les arbres sont particulièrement utiles pour résoudre des problèmes impliquant des probabilités conditionnelles et pour calculer la probabilité d'une séquence d'événements.*

Comment construire un arbre de probabilité

Méthode

Pour construire un arbre de probabilité :

1. Partez d'un point initial (la racine)
2. Pour chaque étape de l'expérience, créez des branches représentant toutes les possibilités
3. Étiquetez chaque branche avec la probabilité conditionnelle correspondante
4. À la fin de chaque chemin, indiquez l'événement final correspondant



Comment lire un arbre de probabilité

Propriété

Pour calculer la probabilité d'un événement final dans un arbre de probabilité :

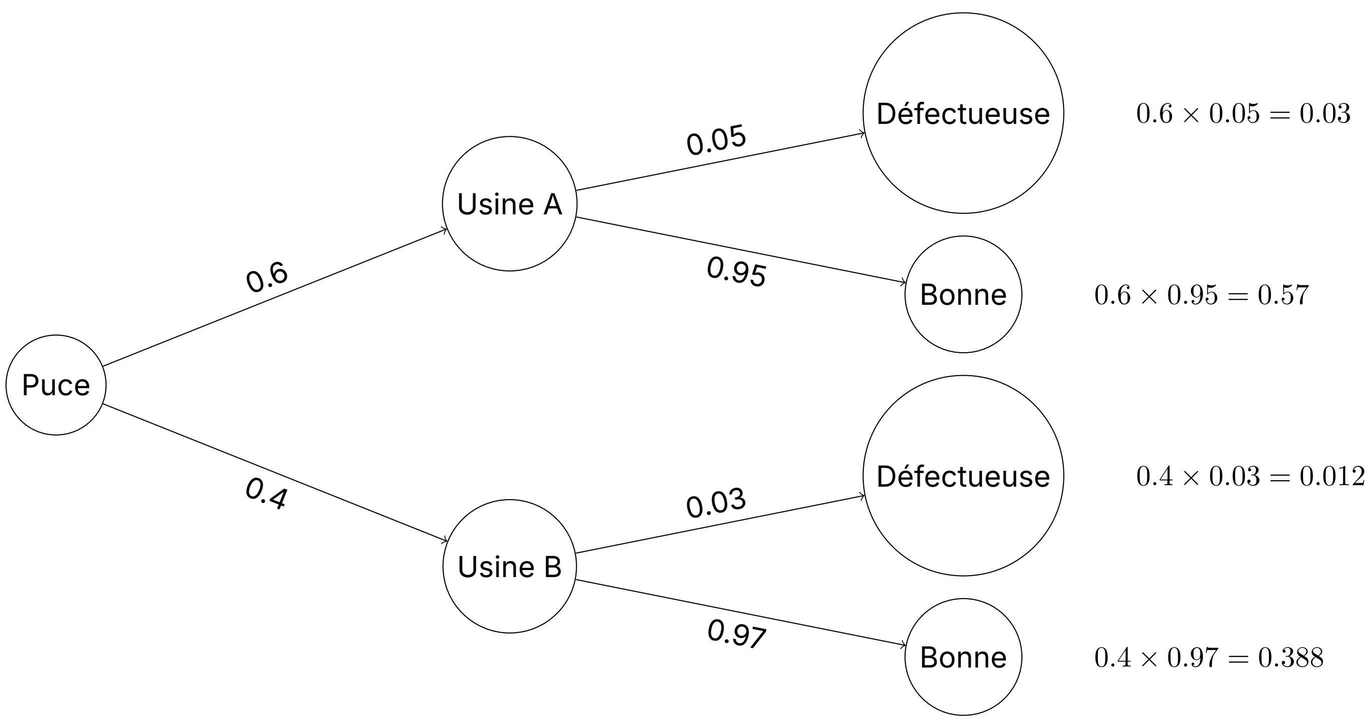
$$\boxed{\text{Probabilité d'un chemin} = \text{Produit des probabilités le long du chemin}}$$

> *Pour trouver la probabilité d'un événement qui peut se produire de plusieurs façons différentes, additionnez les probabilités des chemins correspondants.*

⚠ Il est essentiel que les probabilités sur les branches partant d'un même nœud somment à 1. Si ce n'est pas le cas, l'arbre est incorrect.

Exemple concret

> *Une entreprise fabrique des puces électroniques dans deux usines. L'usine A produit 60% des puces, et l'usine B produit les 40% restants. Dans l'usine A, 5% des puces sont défectueuses, tandis que dans l'usine B, 3% des puces sont défectueuses. Quelle est la probabilité qu'une puce choisie au hasard soit défectueuse?*



Pour trouver la probabilité qu'une puce soit défectueuse, nous additionnons les probabilités des chemins menant à une puce défectueuse:

P(puce défectueuse) = 0.03 + 0.012 = 0.042 = 4.2%

Méthode

Pour résoudre un problème avec un arbre de probabilité :

1. Identifiez clairement les événements et les étapes
2. Dessinez l'arbre en partant de la racine
3. Notez les probabilités sur chaque branche
4. Multipliez les probabilités le long de chaque chemin
5. Additionnez les probabilités des chemins correspondant à l'événement recherché

Propriété

L'arbre de probabilité est particulièrement utile pour appliquer le théorème de Bayes, car il permet de visualiser et calculer facilement les probabilités conditionnelles inverses.

Théorème de Bayes

Théorème

Le **théorème de Bayes** est une formule qui permet de calculer la probabilité conditionnelle inverse : si on connaît $P_A(B)$, le théorème nous donne $P_B(A)$ avec la formule :

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

> *Le théorème de Bayes nous aide à "renverser" la condition. Par exemple, si on connaît la probabilité qu'un test médical soit positif sachant qu'on est malade, on peut calculer la probabilité d'être malade sachant que le test est positif.*

⚠ Attention, la formule de Bayes peut aussi s'écrire :

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

C'est la même chose car on sait que $P(A \cap B) = P(A) \times P_A(B)$.

Exemple d’application

> *Un nouveau test de dépistage d’une maladie est correct dans 95% des cas chez les personnes malades (sensibilité), et dans 90% des cas chez les personnes non malades (spécificité). Si cette maladie touche 1% de la population, quelle est la probabilité qu’une personne soit réellement malade si son test est positif?*

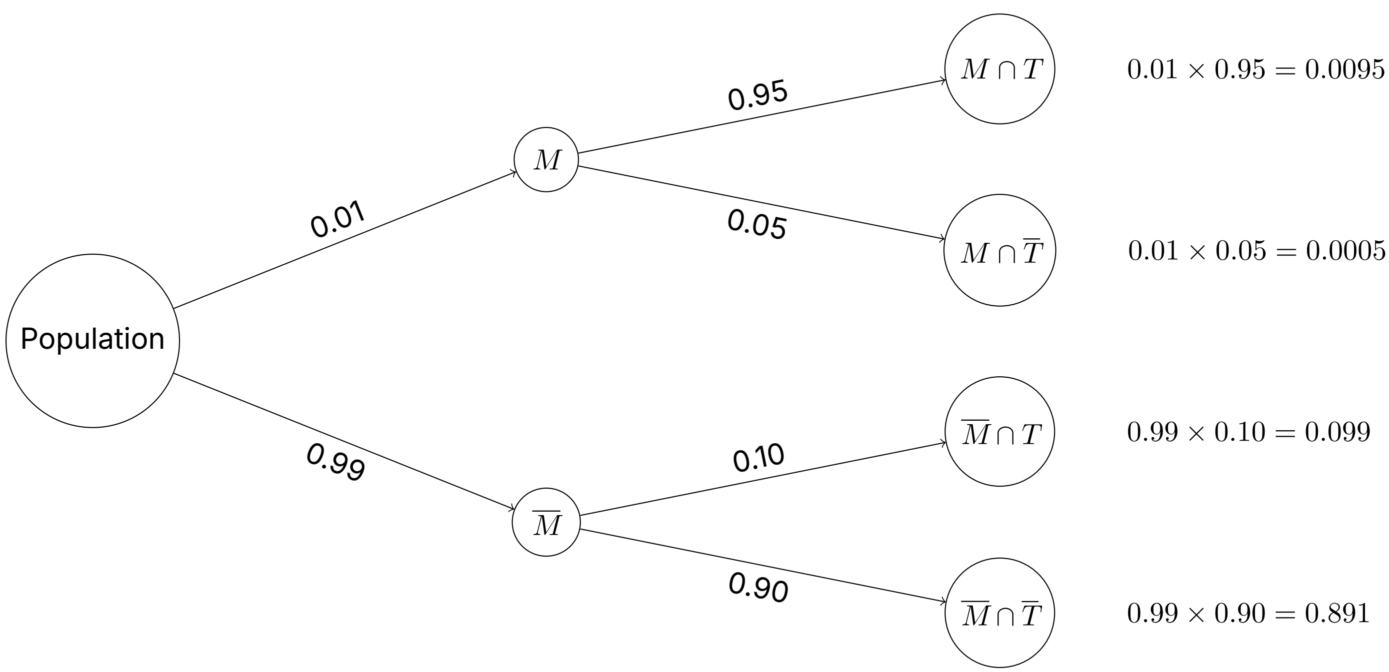
Notons :

- M : la personne est malade
- T : le test est positif

Nous connaissons :

- $P_M(T) = 0.95$
- $P_{\overline{M}}(T) = 0.10$
- $P(M) = 0.01$

Nous cherchons $P_T(M)$.



Appliquons la formule de Bayes :

$$
\begin{aligned}
P_T(M) &= \frac{P(M) \times P_M(T)}{P(T)} \\
&= \frac{P(M) \times P_M(T)}{P(M) \times P_M(T) + P(\overline{M}) \times P_{\overline{M}}(T)} \\
&= \frac{0.01 \times 0.95}{0.01 \times 0.95 + 0.99 \times 0.10} \\
&= \frac{0.0095}{0.0095 + 0.099} \\
&= \frac{0.0095}{0.1085} \\
&\approx 0.0875 \approx 8.8\%
\end{aligned}
$$

Seulement 8.8% des personnes ayant un test positif ont réellement la maladie! Ce résultat peut sembler étonnant, mais s'explique par la faible prévalence de la maladie dans la population. C'est ce qu'on appelle le "paradoxe de la base rate" ou "paradoxe des probabilités conditionnelles".

Lois de probabilité discrètes

Définition

Une **loi de probabilité discrète** est une loi qui décrit un phénomène aléatoire dont les issues possibles sont des entiers (0, 1, 2, etc.), et pour lesquels on peut calculer la probabilité exacte de chaque issue.

Loi de Bernoulli

Définition

La loi de Bernoulli modélise une **expérience à deux issues** (succès/échec, oui/non, pile/face...). On note $X \sim \mathcal{B}(p)$ si X suit une loi de Bernoulli de paramètre p (probabilité de succès).

$$P(X = 1) = p \quad ; \quad P(X = 0) = 1 - p$$

Exemple : Je lance un dé et je dis "succès" si j'obtiens un 6. Alors $X = 1$ si j'obtiens 6, $X = 0$ sinon. La variable X suit une loi de Bernoulli avec $p = \frac{1}{6}$.

Propriété

Espérance et variance :

$$\mathbb{E}(X) = p \quad ; \quad \text{Var}(X) = p(1-p)$$

Loi Binomiale

Définition

La loi binomiale modélise la **répétition de n expériences indépendantes de Bernoulli**. On note $X \sim \mathcal{B}(n,p)$.

Propriété

Fonction de masse :

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{pour } k = 0, 1, ..., n$$

> *Exemple : Je lance 10 fois une pièce. X est le nombre de fois où j'obtiens pile. Alors $X \sim \mathcal{B}(10, 0.5)$.*

Propriété

Espérance et variance :

$$\mathbb{E}(X) = np \quad ; \quad \text{Var}(X) = np(1-p)$$

Loi de Poisson

Définition

La loi de Poisson modélise le **nombre d'événements rares** sur un intervalle de temps ou d'espace. On note $X \sim \mathcal{P}(\lambda)$, où λ est le nombre moyen d'événements.

Propriété

Fonction de masse :

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

Exemple : Le nombre d'accidents de voiture à une intersection en un mois suit souvent une loi de Poisson. Si on observe en moyenne 2 accidents par mois, alors $X \sim \mathcal{P}(2)$.

⚠ Attention à bien faire les conversions, si on vous demande la probabilité d'avoir trois accidents en un année, il faut convertir le paramètre de la poisson en années, donc au lieu de prendre $\lambda = 2$ mois, on prendra $\lambda = 2/12 = 1/6 \approx 0.167$ années. On aura ainsi tous nos résultats en années.

Propriété

Espérance et variance :

$$\mathbb{E}(X) = \lambda \quad ; \quad \text{Var}(X) = \lambda$$

La loi de Poisson est une limite de la loi binomiale lorsque n est grand, p est petit, et $np = \lambda$ reste constant.

Lois de probabilité continues

Définition

Une **loi de probabilité continue** est une loi qui modélise une variable aléatoire prenant une infinité de valeurs réelles. Contrairement aux lois discrètes, on ne peut pas dire "$P(X = a)$", car cela vaut toujours zéro : on calcule des probabilités sur des intervalles.

Loi uniforme continue

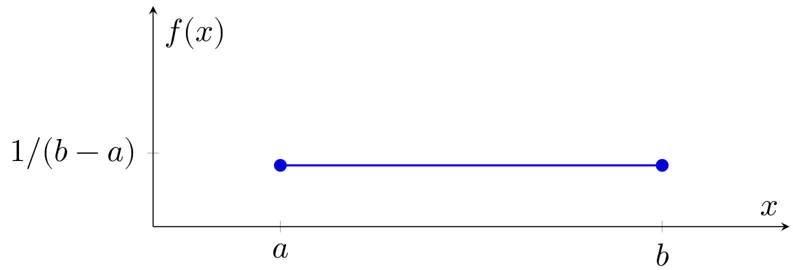
Définition

La loi uniforme continue modélise une variable aléatoire qui a la même probabilité d'être n'importe où sur un intervalle $[a, b]$. On note $X \sim \mathcal{U}(a, b)$.

Propriété

Densité de probabilité :

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



Exemple : Si un train peut arriver n'importe quand entre 8h00 et 8h10, alors l'heure d'arrivée X suit une loi uniforme sur $[0, 10]$.

Propriété

Espérance et variance :

$$\mathbb{E}(X) = \frac{a+b}{2} \quad ; \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

Loi normale (ou loi gaussienne)

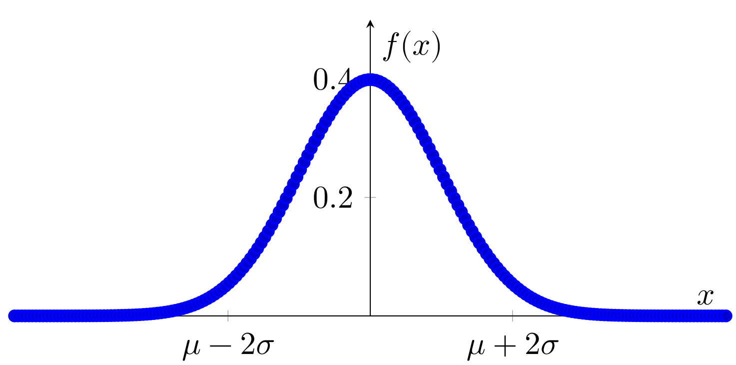
Définition

La loi normale est la loi la plus importante en probabilités. Elle modélise les phénomènes naturels qui fluctuent autour d'une moyenne. On note $X \sim \mathcal{N}(\mu, \sigma^2)$, où μ est la moyenne et σ l'écart-type.

Propriété

Densité de probabilité :

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



> *Exemple : La taille des adultes dans une population suit en général une loi normale. Si la taille moyenne est 170 cm et l'écart-type est 10 cm, alors* $X \sim \mathcal{N}(170, 10^2)$.

Propriété

Espérance et variance :

$$\mathbb{E}(X) = \mu \quad ; \quad \text{Var}(X) = \sigma^2$$

⚠ On ne peut pas calculer les probabilités de la loi normale avec une formule simple : on utilise une **table de valeurs** ou une calculatrice.

Méthodes : faire un test statistique

Avant de plonger dans les méthodes, rappelons brièvement ce qu'est un test statistique.

Définition

Un test statistique permet de prendre une décision sur une hypothèse concernant une ou plusieurs variables, à partir des données observées.

> *Il sert à répondre à des questions du type : "Est-ce que ce que j'observe est suffisamment improbable sous une hypothèse donnée pour que je la remette en question ?".*

En pratique, on formule une hypothèse nulle (H_0) que l'on cherche à tester. Puis on regarde si les données sont compatibles avec cette hypothèse, ou si elles s'en écartent suffisamment pour la rejeter en faveur d'une hypothèse alternative (H_1). Le seuil de décision est souvent fixé à 5% : on accepte de se tromper dans 1 cas sur 20.

Quand utiliser chaque test ? Chaque test statistique est adapté à une situation particulière, selon le type de données (qualitatives ou quantitatives), le nombre d'échantillons, ou ce qu'on cherche à comparer (moyennes, proportions, dépendances...).

Test du chi-2 d'indépendance

Ce test est utilisé pour vérifier s'il existe une relation entre deux variables qualitatives.

> *Par exemple : est-ce que le genre influence les préférences musicales ? Est-ce que le type de contrat est lié au secteur d'activité ?*

Il s'applique lorsque les données sont présentées sous forme de tableau de probabilités (tableau de contingence).

Méthode

Test du χ^2 d'indépendance (variables qualitatives)

1. **Hypothèses :**
   - H_0 : les deux variables sont indépendantes.
   - H_1 : les deux variables sont liées (dépendantes).
2. **On choisit un niveau de confiance**, souvent 95%, donc $\alpha = 0{,}05$.
   Cela signifie qu'on accepte de se tromper dans 5% des cas où on rejette H_0.
3. **Calcul des effectifs attendus** :
$$E_{ij} = \frac{(\text{Total ligne}_i) \times (\text{Total colonne}_j)}{\text{Total général}}$$
4. **Calcul de la statistique du test** :
$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$
5. **Calcul des degrés de liberté** :
$$\text{ddl} = (\text{nombre de lignes} - 1) \times (\text{nombre de colonnes} - 1)$$
6. **Décision** :
   - On regarde la valeur critique dans une table de χ^2 pour $\alpha = 0{,}05$ et les ddl trouvés.
   - Si $\chi^2_{\text{calculé}} > \chi^2_{\text{critique}}$, on **rejette** H_0 : les variables ne sont pas indépendantes.
   - Sinon, on **ne rejette pas** H_0 : on n'a pas de preuve qu'elles soient liées.

Test de Student

Ce test permet de comparer la moyenne d'un échantillon à une valeur de référence (souvent une norme ou une hypothèse théorique).

> *Par exemple : la moyenne de satisfaction des clients est-elle différente de 7 sur 10 ? Le poids moyen des colis dépasse-t-il 2 kg ?*

Il s'applique aux variables quantitatives lorsque l'on a un seul échantillon.

Méthode

Test de Student (moyenne d'un échantillon)

1. **Hypothèses :**
   - H_0 : la moyenne est égale à la valeur attendue ($\mu = \mu_0$).
   - H_1 : la moyenne est différente de la valeur attendue ($\mu \neq \mu_0$).
2. **On choisit un niveau de confiance**, souvent 95%, donc $\alpha = 0{,}05$.
   Cela signifie qu'on accepte de se tromper dans 5% des cas où on rejette H_0.
3. **Calcul de la statistique** :
   $$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
   avec $\bar{x}$: moyenne de l'échantillon, s : écart-type, n : taille de l'échantillon.
4. **Degrés de liberté** :
   $$\text{ddl} = n - 1$$
5. **Décision** :
   - On regarde la valeur critique dans une table de Student pour $\alpha = 0{,}05$ et les ddl trouvés.
   - Si $|t| > t_{\alpha/2,\text{ddl}}$, on **rejette** H_0 : la moyenne est significativement différente.
   - Sinon, on **ne rejette pas** H_0 : on n'a pas assez d'éléments pour dire qu'elle est différente.