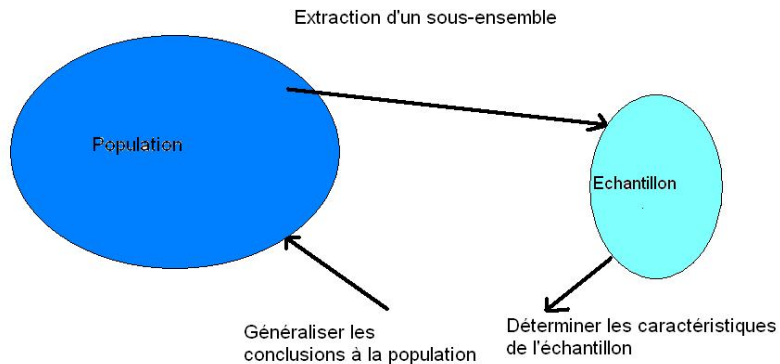


La théorie des probabilités

La démarche



La Théorie des probabilités :

- permet de modéliser des phénomènes aléatoires et d'y effectuer des calculs théoriques
- concerne les **populations** : on ne peut donc pas faire de mesures.

La Statistique :

- concerne les **échantillons**, le monde réel, la pratique,
- on fait des mesures (observations) sur des individus,
- repose sur la modélisation probabiliste des observations.

Un exemple concret

Un fabricant d'ampoules souhaite vérifier la qualité des ampoules électriques produites dans sa chaîne de montage.

Pour cela, il propose donc d'évaluer la durée moyenne de bon fonctionnement d'une ampoule.

Comment faire ? on ne peut pas tester toutes les ampoules produites par la chaîne de montage !

On tire un échantillon au hasard

On réalise l'expérience, on effectue des mesures, on calcule la durée moyenne de bon fonctionnement des ampoules de l'échantillon

On approxime la durée moyenne de bon fonctionnement des ampoules de la population entière par la durée moyenne de bon fonctionnement des ampoules de l'échantillon

Un exemple concret (suite)

Comment savoir si ce qu'on vient de faire est licite ? Quelle est la qualité de l'approximation ? Il faut étudier la théorie des probabilités et la statistique !

Si on tire un autre échantillon, il y a de fortes chances que l'on n'obtienne pas les mêmes résultats.

Ces fluctuations (ou erreurs d'échantillonnage) sont dues à la variabilité. Cela signifie que des objets semblables en apparence peuvent présenter des différences lorsqu'on effectue des mesures.

Les différents aspects de la Statistique

Observer ne suffit pas !

Statistique descriptive :

- Résumer les mesures sur un échantillon (moyenne, variance,...)
- Représenter les mesures (histogramme, distribution)

Statistique inférentielle :

- Généraliser les propriétés d'un échantillon à une population en prenant en compte les fluctuations d'échantillonnage
- il faut modéliser les observations (par des variables aléatoires) : on fait appel à la théorie des probabilités

Tests d'hypothèses :

- Contrôler la validité d'un modèle
- Comparer un échantillon à une référence

Statistique décisionnelle :

- Savoir prendre une décision alors que les résultats sont exprimés en termes de probabilités (i.e. de pourcentage de chances, de risques)

Expérience aléatoire, évènements

Une expérience est **aléatoire** si on ne peut pas prévoir à l'avance son résultat, et si, répétée dans des conditions identiques, elle peut donner lieu à des résultats différents. Lorsqu'on effectue une expérience, les valeurs obtenues s'appellent des **réalisations** ou des observations.

Univers (noté Ω) : ensemble de tous les résultats possibles d'une expérience. Il peut être :

- fini, par ex $\{x_1, \dots, x_k\}$
- infini dénombrable : on peut indiquer, numéroter ses éléments jusqu'à l'infini, par ex $\{x_1, x_2, \dots, x_n, \dots\}$
- infini non-dénombrable : ceci signifie qu'il n'est pas possible de décrire l'ensemble sous la forme d'une liste numérotée $\{x_1, x_2, \dots, x_k, \dots\}$, par ex l'intervalle $[0, 1]$ est un ensemble infini non-dénombrable.

Evènement élémentaire : un des éléments de Ω lorsqu'on peut les énumérer.

Evènement : sous-ensemble de Ω .

Rappel : opérations sur les ensembles

Soient A et B deux évènements d'un ensemble fondamental Ω

$\{A \text{ ou } B\} = A \cup B =$ réunion de A et B

$\{A \text{ et } B\} = A \cap B =$ intersection de A et B

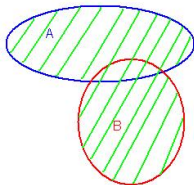
complémentaire de A dans $\Omega = \bar{A} = \Omega - A$

$\emptyset =$ évènement impossible

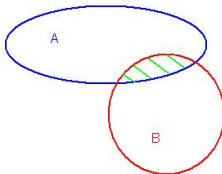
$\Omega =$ évènement certain

A et B sont incompatibles ou disjoints lorsque $A \cap B = \emptyset$

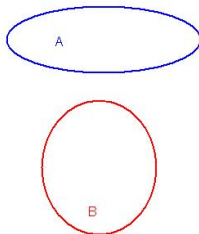
surface hachurée en vert =
réunion de A et B



surface hachurée en vert = intersection de
 A et B



A et B sont disjoints



Règle de calcul des probabilités

Une probabilité est une **fonction** notée $\mathbb{P}$ qui attribue à tout évènement A une valeur $\mathbb{P}(A)$ désignant la probabilité que A se réalise.

Une probabilité possède les propriétés suivantes :

- $0 \leq \mathbb{P}(A) \leq 1$ pour tout évènement A
- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- en général, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- mais si A et B sont disjoints, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
- en général, on a $\mathbb{P}(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathbb{P}(A_i)$
- mais si les A_i sont 2 à 2 disjoints, $\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$

Probabilités sur les univers finis

Soit Ω est un univers **fini**, on peut alors l'écrire sous la forme $\Omega = \{\omega_1, \dots, \omega_n\}$. On note $\text{card}(\Omega)$ le nombre d'éléments de Ω qui représente le nombre de cas possibles à l'issue de l'expérience aléatoire.

Lorsque Ω est un univers **fini**, il est parfois approprié de supposer que la probabilité associée à chaque événement élémentaire est identique i.e. $\mathbb{P}[\omega_i] = 1/\text{card}(\Omega)$ pour $i = 1, \dots, n$. On dit alors qu'il y a **équiprobabilité**.

L'hypothèse d'équiprobabilité implique que la probabilité d'un événement A s'obtient en calculant le rapport du nombre de cas possibles correspondant à l'évènement A (noté $\text{card}(A)$) sur le nombre de cas possibles, soit :

$$\mathbb{P}[A] = \frac{\text{card}(A)}{\text{card}(\Omega)}$$

Probabilités conditionnelles : introduction

Considérons une expérience réalisée sur une certaine population et un évènement A qui a une probabilité $\mathbb{P}[A]$ de se réaliser,

par ex : A = présence d'une maladie M .

Que devient $\mathbb{P}[A]$ si on se restreint à une sous-population ?

par ex : sous-population = les individus présentant un signe S .

On introduit un évènement B conditionnant, qui définit la sous-population,

par ex : B = présenter le signe S .

$\mathbb{P}[B]$ ne doit pas être nul.

Probabilités conditionnelles : définition

La probabilité que l'évènement A se réalise sachant que l'évènement B a eu lieu (=probabilité de A parmi la sous-population caractérisée par B) est définie par :

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

De même, la probabilité que l'évènement B se réalise sachant que l'évènement A a eu lieu est définie par :

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}.$$

Ne PAS confondre $\mathbb{P}[A|B]$ = probabilité que A se réalise sachant qu'on a observé B avec $\mathbb{P}[A \cap B]$ =probabilité que A et B se réalisent simultanément !!

Probabilités conditionnelles : règles de calcul

Soit B un évènement **fixé**. La fonction $A \rightarrow \mathbb{P}[A|B]$ est une vraie probabilité i.e. les règles de calcul avec les probabilités conditionnelles sont les mêmes qu'avec les probabilités classiques.

$$0 \leq \mathbb{P}(A|B) \leq 1 \text{ pour tout évènement } A$$

$$\mathbb{P}(\Omega|B) = 1$$

$$\mathbb{P}(\emptyset|B) = 0$$

$$\mathbb{P}(\bar{A}|B) = 1 - \mathbb{P}(A|B)$$

$$A_1 \subset A_2 \Rightarrow \mathbb{P}(A_1|B) \leq \mathbb{P}(A_2|B)$$

en général,

$$\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) - \mathbb{P}(A_1 \cap A_2|B)$$

mais si A_1 et A_2 sont disjoints,

$$\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B)$$

$$\text{en général, on a } \mathbb{P}(\cup_{i=1}^n A_i|B) \leq \sum_{i=1}^n \mathbb{P}(A_i|B)$$

mais si les A_i sont 2 à 2 disjoints,

$$\mathbb{P}(\cup_{i=1}^n A_i|B) = \sum_{i=1}^n \mathbb{P}(A_i|B)$$

Indépendance de 2 évènements

1^{ère} définition avec formule : A et B sont indépendants si

$$\mathbb{P}[A|B] = \mathbb{P}[A] \quad \text{ou/et} \quad \mathbb{P}[B|A] = \mathbb{P}[B].$$

2^{ème} définition avec formule : A et B sont indépendants si

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \times \mathbb{P}[B].$$

Cette formule est symétrique en A et B, on en déduit que si $\mathbb{P}[A|B] = \mathbb{P}[A]$, alors on a aussi $\mathbb{P}[B|A] = \mathbb{P}[B]$.

Indépendance et incompatibilité

Ne pas confondre des événements incompatibles et des événements indépendants...

ex : considérons A ="l'enfant à naître est un garçon" et B ="l'enfant à naître est une fille". Les événements A et B sont incompatibles. Mais ils ne sont pas indépendants!!! En effet,

$$\mathbb{P}[A \cap B] = 0 \neq \mathbb{P}[A] \times \mathbb{P}[B] = 0.5 \times 0.5 = 0.25$$

Système complet

Une famille d'évènements $\{A_1, A_2, \dots, A_n\}$ forme un système complet d'évènements si

- $A_1 \neq \emptyset, A_2 \neq \emptyset, \dots, A_n \neq \emptyset$
- $A_1 \cap A_2 = \emptyset, A_1 \cap A_3 = \emptyset, \dots, A_{n-1} \cap A_n = \emptyset$ (i.e. les évènements du système complet sont deux à deux disjoints)
- $\Omega = A_1 \cup A_2 \cup \dots \cup A_n$

Par exemple, soit l'univers $\Omega = \{1, 2, 3, 4, 5, 6\}$. On peut définir un système complet d'évènements $\{A_1, A_2, A_3\}$ avec, par exemple, $A_1 = \{1, 2\}$, $A_2 = \{3, 5\}$ et $A_3 = \{4, 6\}$. On a bien $A_1 \neq \emptyset, A_2 \neq \emptyset, A_3 \neq \emptyset, A_1 \cap A_2 = \emptyset, A_1 \cap A_3 = \emptyset, A_2 \cap A_3 = \emptyset$ et $\Omega = A_1 \cup A_2 \cup A_3$.

Lorsque A est un évènement de probabilité non nulle, $\{A, \bar{A}\}$ forme un système complet d'évènements.

Pour tout évènement B , on a

$$\mathbb{P}[B] = \sum_{i=1}^n \mathbb{P}[B \cap A_i] = \sum_{i=1}^n \mathbb{P}[B|A_i]\mathbb{P}[A_i]$$

et en particulier

$$\mathbb{P}[B] = \mathbb{P}[B \cap A] + \mathbb{P}[B \cap \bar{A}] = \mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|\bar{A}]\mathbb{P}[\bar{A}]$$

Quelques formules utiles

Formule de Bayes : lorsque A et B sont deux évènements de probabilité non-nulle, on peut écrire :

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

Formule des probabilités totales : Soit $\{A_1, \dots, A_n\}$ un système complet d'évènements. Pour $k = 1, \dots, n$, on peut écrire :

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{j=1}^n \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

ce qui dans le cas d'un système complet d'évènements donné par $\{A, \bar{A}\}$ revient à :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\bar{A})\mathbb{P}(\bar{A})}$$

Définition d'une variable aléatoire

Une **variable aléatoire** X est le procédé qui relie l'expérience aléatoire à un nombre. On note D_X l'ensemble des valeurs que X peut prendre après réalisation de l'expérience : D_X s'appelle le domaine de définition de X .

A chaque fois que l'on reproduit l'expérience, on obtient une réalisation de X que l'on note x : x est un nombre alors que X est une fonction !!!

Soit l'expérience "tirer une pièce parmi une production" et soit X la variable aléatoire représentant la longueur de la pièce tirée.

L'ingénieur d'usine effectue une 1ère fois cette expérience, il obtient la réalisation $x_1 = 10.2\text{cm}$. Il recommence une 2ème fois l'expérience et obtient la réalisation $x_2 = 9.9\text{cm}$, etc...

Soit l'expérience "jeter un dé" et soit X la variable aléatoire représentant la valeur inscrite sur la face supérieure. Un joueur effectue une 1ère fois cette expérience, il obtient la réalisation $x_1 = 4$. Il recommence une 2ème fois l'expérience et obtient la réalisation $x_2 = 3$, etc...

Les différents types de variables

On distingue :

les variables aléatoires **continues** : toute valeur d'un intervalle de $\mathbb{R}$ est acceptable, ex : taille, poids, volume, temps écoulé...

les variables aléatoires **discrètes** : elles prennent un nombre dénombrable (fini ou infini) de valeurs, par ex : nombre de pièces défectueuses dans la production journalière d'une usine, nombre de clients arrivant à un guichet en une journée, variable aléatoire binaire codant pour "succès" ou "échec"...

Distribution de probabilité et fonction de répartition

La **loi de probabilité** aussi appelée **distribution de probabilité** et notée f_X d'une variable aléatoire X a pour but de décrire quelles sont les valeurs possibles prises par la variable et avec quelle probabilité ces différentes valeurs sont prises.

Une variable aléatoire est entièrement caractérisée par sa distribution de probabilité f_X ou de manière équivalente par sa **fonction de répartition** notée F_X .

La **théorie des probabilités** vise à évaluer le comportement des variables aléatoires (espérance, variance, probabilités de dépassement d'un seuil, comportement de sommes,...) étant donné la distribution de probabilité F_X .

La **statistique** fournit des méthodes pour résoudre le problème inverse dit d'inférence statistique : caractériser F_X au vu des observations des variables.

Fonction de répartition d'une variable aléatoire

Soit X une variable aléatoire et soit x un nombre.

Considérons l'évènement $\{X \leq x\}$ = ensemble des résultats d'expérience dont le codage est inférieur ou égal à x .

$\mathbb{P}[X \leq x]$ est un nombre qui dépend de la valeur de x

On définit F_X la fonction de répartition de X par

$$F_X(x) = \mathbb{P}[X \leq x].$$

Pour tout x , on a $0 \leq F_X(x) \leq 1$ avec $\lim_{x \rightarrow +\infty} F_X(x) = 1$ et $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

On note $F_X(x^-) = \mathbb{P}[X < x] = \lim_{y \rightarrow x, y < x} F_X(y)$.

$\mathbb{P}[X > x] = 1 - \mathbb{P}[X \leq x]$ i.e. $\mathbb{P}[X > x] = 1 - F_X(x)$

$\mathbb{P}[X \leq x] - \mathbb{P}[X < x] = \mathbb{P}[X = x]$ i.e. $F_X(x) - F_X(x^-) = \mathbb{P}[X = x]$

Pour $a < b$, on a :

- $F_X(b) - F_X(a) = \mathbb{P}[a < X \leq b]$
- $F_X(b) - F_X(a^-) = \mathbb{P}[a \leq X \leq b]$
- $F_X(b^-) - F_X(a) = \mathbb{P}[a < X < b]$
- $F_X(b^-) - F_X(a^-) = \mathbb{P}[a \leq X < b]$

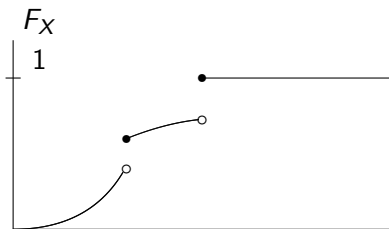
Quelques propriétés de la fonction de répartition

F_X est croissante ($x \leq y \Rightarrow F_X(x) \leq F_X(y)$).

F_X est continue à droite ce qui signifie que F_X est continue sauf éventuellement en un nombre dénombrable de points isolés $(a_i)_{i=1,\dots}$ en lesquels $F_X(a_i) \neq F_X(a_i^-)$.

Si F_X est discontinue en un point b alors on a $\mathbb{P}[X = b] = F_X(b) - F_X(b^-) > 0$.

Si F_X est continue en un point a alors on a $F_X(a) = F_X(a^-)$ et $\mathbb{P}[X = a] = 0$.



Fonction de répartition d'une variable discrète

La fonction de répartition d'une variable discrète au point x correspond à l'accumulation des probabilités des valeurs inférieures ou égales à x :

$$F_X(x) = \sum_{x_i \leq x, x_i \in D_X} \mathbb{P}[X = x_i]$$

Ainsi, F_X est une fonction en escalier, continue à droite.

Pour $a < b$, on a :

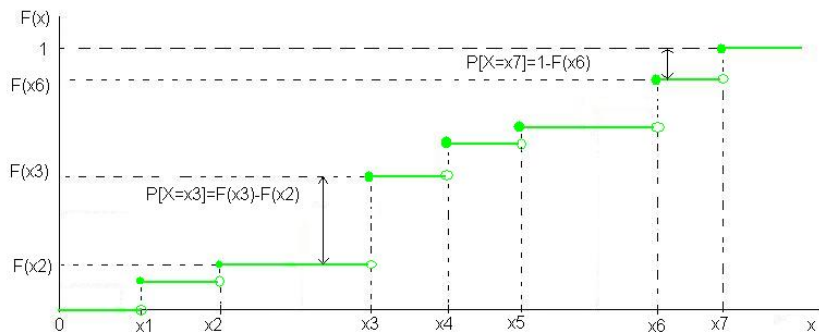
$$F_X(b) - F_X(a) = \sum_{a < x_i \leq b, x_i \in D_X} \mathbb{P}[X = x_i]$$

$$F_X(b) - F_X(a^-) = \sum_{a \leq x_i \leq b, x_i \in D_X} \mathbb{P}[X = x_i]$$

$$F_X(b^-) - F_X(a) = \sum_{a < x_i < b, x_i \in D_X} \mathbb{P}[X = x_i]$$

$$F_X(b^-) - F_X(a^-) = \sum_{a \leq x_i < b, x_i \in D_X} \mathbb{P}[X = x_i]$$

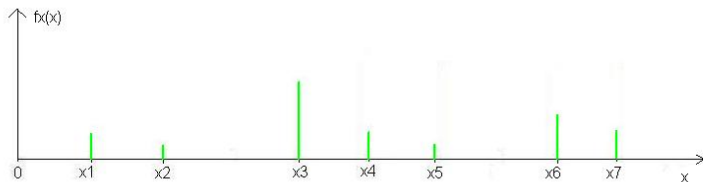
Allure de la fonction de répartition d'une variable discrète



Distribution de probabilité d'une variable discrète

La distribution de probabilité d'une variable aléatoire discrète est la donnée des nombres $p_i = \mathbb{P}[X = x_i]$ pour chacune des valeurs possibles x_i pour X . Ces nombres sont compris entre 0 et 1 et leur somme vaut 1. Cela revient à définir une fonction f_X telle que

$$f_X(x) = \begin{cases} \mathbb{P}[X = x] & \text{si } x \in D_X \\ 0 & \text{si } x \notin D_X \end{cases}$$



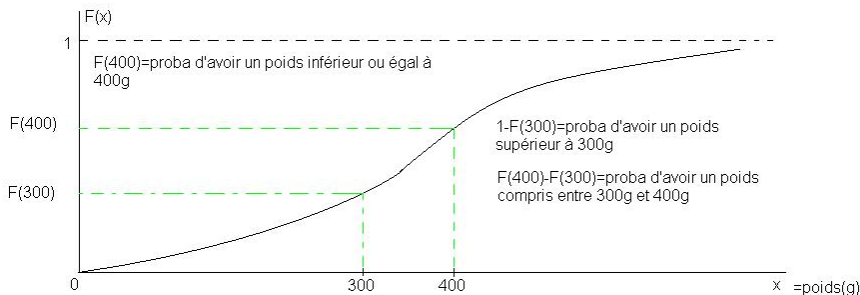
Fonction de répartition d'une variable continue

La fonction de répartition d'une variable aléatoire continue est une fonction continue, dérivable presque-partout.

En tout point x , on a $\mathbb{P}[X = x] = 0$ et $\mathbb{P}[X < x] = \mathbb{P}[X \leq x]$.

Pour tout $a < b$, on a :

$$\mathbb{P}[a \leq X \leq b] = \mathbb{P}[a \leq X < b] = \mathbb{P}[a < X \leq b] = \mathbb{P}[a < X < b]$$



Distribution de probabilité d'une variable aléatoire continue

La distribution de probabilité d'une variable aléatoire continue (aussi appelée **densité**) est donnée par :

$$f_X(x) = \begin{cases} \frac{dF_X(x)}{dx} & \text{pour } x \in D_X \text{ tel que } F_X \text{ est dérivable} \\ 0 & \text{pour } x \notin D_X \text{ et lorsque } F_X \text{ n'est pas dérivable} \end{cases}$$

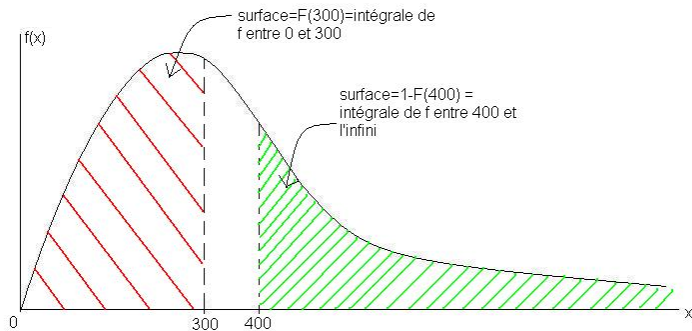
$$f_X(x)dx = \mathbb{P}[x \leq X \leq x + dx] \approx \mathbb{P}[X = x]$$

$$F_X(x) = \int_{-\infty}^x f_X(u)du$$

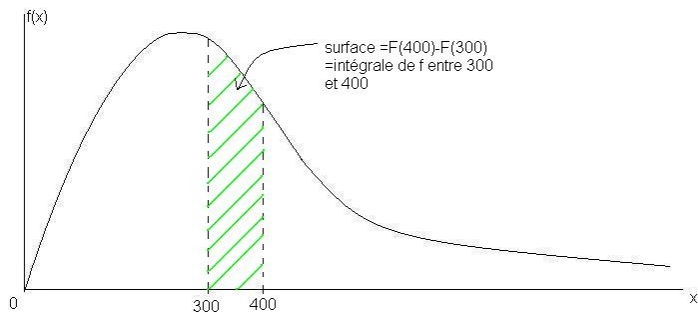
$$f_X \text{ est positive (car } F_X \nearrow) \text{ et } \int_{-\infty}^{\infty} f_X(u)du = 1$$

$$\mathbb{P}[a < X \leq b] = F_X(b) - F_X(a) = \int_a^b f_X(u)du$$

Densité d'une variable aléatoire continue (suite)



Densité d'une variable aléatoire continue (suite 2)



Variables aléatoires indépendantes

Deux variable X_1 et X_2 sont **indépendantes** lorsque le fait de connaître la valeur obtenue par X_1 n'apporte aucune information sur la valeur qui sera prise par X_2 et réciproquement.
ex : X_1 ="poids d'une souris" et X_2 ="couleur du pelage" sont indépendantes alors que X_1 et Y_1 ="taille d'une souris" ne le sont vraisemblablement pas.

Caractérisation de l'indépendance pour un couple de variables **discrets** : X et Y sont indépendants lorsque pour tout couple de valeurs (x_i, y_j) pris par (X, Y) , on a

$$\mathbb{P}[X = x_i, Y = y_j] = \mathbb{P}[X = x_i]\mathbb{P}[Y = y_j]$$

Caractérisation de l'indépendance pour un couple de variables **continues** : X et Y sont indépendants lorsqu'on a

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{pour tout } (x, y)$$

où $f_{X,Y}$ représente la densité jointe du couple (X, Y) , f_X représente la densité de X et f_Y représente la densité de Y .

Caractéristiques de position et de dispersion

Caractéristiques de **position** :

- l'**espérance** est un nombre qui représente la valeur moyenne prise par X .
- la **médiane** (lorsqu'elle est définie) est la valeur telle que X a autant de chance de se réaliser au-dessus qu'en dessous.
- un **mode** correspond à une valeur ayant une probabilité maximale de se réaliser (il existe des distributions n'ayant aucun mode).
- moyenne=médiane=mode pour les distributions symétriques i.e. telles que les valeurs prises par X sont également réparties autour d'une valeur centrale.

Caractéristiques de **dispersion** :

- La **variance** exprime à quel point les valeurs prises par X sont dispersées autour de la moyenne. Une grande variance indique une grande dispersion. A l'inverse, une variance nulle révèle que X est en fait non-aléatoire.
- L'**écart-type** fournit la même information.
- Les **quantiles** (lorsqu'ils sont définis) permettent de fournir l'intervalle dans lequel X se réalise avec 95% de chances par ex.

Espérance mathématique

Soit X une variable aléatoire. On note $\mathbb{E}[X]$ l'**espérance** de X . C'est un nombre qui représente la valeur moyenne prise par X .

Si X est discrète, on calcule $\mathbb{E}[X]$ par la formule :

$$\mathbb{E}[X] = \sum_{x_i \in D_X} x_i \mathbb{P}[X = x_i].$$

Si X est continue, on calcule $\mathbb{E}[X]$ par la formule :

$$\mathbb{E}[X] = \int x f_X(x) dx.$$

On dit qu'une variable est **centrée** lorsque $\mathbb{E}[X] = 0$.

On a toujours $\mathbb{E}[aX] = a\mathbb{E}[X]$, $\mathbb{E}[a + X] = a + \mathbb{E}[X]$ et $\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$.

On calcule $\mathbb{E}[X^2]$ par la formule :

- $\mathbb{E}[X^2] = \sum_{x_i \in D_X} x_i^2 \mathbb{P}[X = x_i]$ si X est discrète,
- $\mathbb{E}[X^2] = \int x^2 f_X(x) dx$ si X est continue.

Variance mathématique, écart-type mathématique

La **variance** d'une variable X est le nombre **positif** défini par :

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

Si X est **discrète**, on calcule $\text{Var}(X)$ par l'une des 2 formules :

$$\text{Var}(X) = \sum (x_i - \mathbb{E}[X])^2 \mathbb{P}[X = x_i] = \sum x_i^2 \mathbb{P}[X = x_i] - \mathbb{E}[X]^2.$$

Si X est **continue**, on calcule $\text{Var}(X)$ par l'une des 2 formules :

$$\text{Var}(X) = \int (x - \mathbb{E}[X])^2 f(x) dx = \int x^2 f(x) dx - \mathbb{E}[X]^2.$$

Une variable aléatoire est dite **réduite** lorsque $\text{Var}(X) = 1$.

On a toujours $\text{Var}(X + a) = \text{Var}(X)$ et $\text{Var}(aX) = a^2 \text{Var}(X)$ mais la relation $\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)$ n'est vraie que lorsque X_1 et X_2 sont indépendantes!!!

L'écart-type est défini par : $\sigma(X) = \sqrt{\text{Var}(X)}$.

Médiane, quantiles

La **médiane** (si elle existe) de la variable X est la valeur u qui vérifie la relation $F_X(u) = 1/2$. Par conséquent, la médiane d'une variable aléatoire X est la valeur u telle que $\mathbb{P}[X \leq u] = \mathbb{P}[X > u] = 1/2$.

Le **quantile** (ou **fractile**) d'ordre $1 - \alpha$ (si il existe) de la variable X est la valeur u qui vérifie la relation $F_X(u) = 1 - \alpha$. Le nombre α représente le risque que X dépasse la valeur u .

L'intervalle interpercentile (si il existe) de la variable X est la donnée des valeurs u et v qui vérifient la relation $\mathbb{P}[u \leq X \leq v] = 1 - \alpha$. Le nombre α représente le risque que X sorte de l'intervalle $[u, v]$.

L'intervalle interquartile (si il existe) de la variable X est la donnée des valeurs u et v qui vérifient la relation $\mathbb{P}[u \leq X \leq v] = 0.5$.

On dit que u est un mode de la variable aléatoire X si $f_X(u)$ est un maximum (local ou global).

- Si X est une variable continue, u doit donc vérifier $f'_X(u) = 0$ et $f''_X(u) < 0$.
- Si X est une variable discrète définie sur $\{x_1, x_2, \dots\}$, alors $u = \max\{f_X(x_1), f_X(x_2), \dots\}$.

Les distributions de probabilité qui n'ont qu'un seul mode sont dites unimodales et celles qui possèdent plusieurs modes sont dites multimodales.

Il existe des distributions de probabilité qui n'ont aucun mode.

Caractéristiques de forme : skewness, kurtosis

Le **coefficient d'asymétrie** (skewness) ν_1 est défini par :

$$\nu_1 = \frac{\mathbb{E}[(X - \mathbb{E}[X])^3]}{\text{Var}(X)^{3/2}}$$

$\nu_1 = 0$ pour les distributions symétriques, $\nu_1 > 0$ lorsque les valeurs prises par X sont très étalées sur la droite, $\nu_1 < 0$ lorsque les valeurs prises par X sont très étalées sur la gauche.

Le **coefficient d'aplatissement** (kurtosis) ν_2 est défini par :

$$\nu_2 = \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{\text{Var}(X)^2}$$

La valeur de référence est $\nu_2 = 3$, c'est celle pour la loi gaussienne standard (cf plus tard). Pour $\nu_2 > 3$, la courbe est aigüe. Pour $\nu_2 < 3$, la courbe est aplatie. La valeur de ν_2 est intéressante seulement pour des distributions peu asymétriques.

Distributions usuelles continues

Des familles de lois de proba usuelles continues sont

- lois normales
- lois exponentielles

Ces lois sont paramétrées. Cela signifie que la famille de lois donne la forme générale mais qu'à l'intérieur de ces familles chaque loi dépend de un ou plusieurs nombres appelés **paramètres**.

Loi normale

On dit qu'une variable X suit une loi normale de paramètres m et σ^2 , notée $\mathcal{N}(m, \sigma^2)$, lorsqu'elle prend ses valeurs dans $\mathbb{R}$ avec la densité suivante pour $x \in \mathbb{R}$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

La densité est symétrique par rapport à la droite verticale d'abscisse $x = m$.

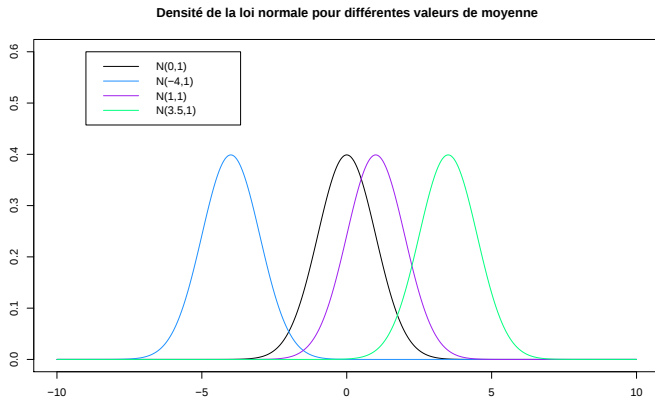
Une variable X de loi normale $\mathcal{N}(m, \sigma^2)$ représente une variable qui oscille de façon symétrique autour de sa moyenne.

$\mathbb{E}[X] = m = \text{médiane} = \text{mode}$ et $\text{var}(X) = \sigma^2$.

Si $X_1, \dots, X_n$ est une suite de variables indépendantes de loi normales telle que $X_i \simeq \mathcal{N}(m_i, \sigma_i^2)$, alors $\sum_{i=1}^n X_i$ suit une loi $\mathcal{N}(m_1 + \dots + m_n, \sigma_1^2 + \dots + \sigma_n^2)$.

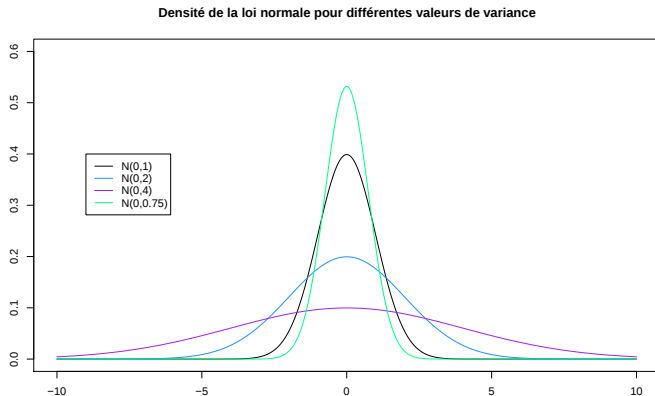
Loi normale : influence de la moyenne

L'allure de la courbe se conserve si on change de moyenne. Il s'agit d'un simple décalage.



Loi normale : influence de la variance

La courbe s'aplatit lorsque la variance augmente, elle se resserre si la variance diminue, le maximum s'ajuste pour que la surface reste 1, le maximum peut dépasser 1.



Loi normale et transformation

D'une manière générale, si X suit une loi $\mathcal{N}(m, \sigma^2)$, alors $aX + b$ suit une loi $\mathcal{N}(am + b, a^2\sigma^2)$.

On ne peut déterminer la fonction de répartition de X de loi $\mathcal{N}(m, \sigma^2)$ que par approximations numériques (par ordinateur).

Un cas important est la loi $\mathcal{N}(0, 1)$ appelée **loi gaussienne standard** ou **loi normale centrée réduite** qui est tabulée.

Pour se ramener à la loi $\mathcal{N}(0, 1)$ à partir d'une variable X de loi $\mathcal{N}(m, \sigma^2)$, on utilise la variable $Y = \frac{X-m}{\sigma}$: Y suit une loi $\mathcal{N}(0, 1)$ et s'appelle la variable centrée réduite associée à X .

Lois exponentielles

On dit qu'une variable X suit une loi exponentielle de paramètre λ , notée $\mathcal{E}(\lambda)$, lorsqu'elle prend ses valeurs dans $\mathbb{R}^+ = [0, +\infty[$ avec la densité et la fonction de répartition suivantes pour $x \in \mathbb{R}$:

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

$$F(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

$$\mathbb{E}[X] = \frac{1}{\lambda} \text{ et } \text{var}(X) = \frac{1}{\lambda^2}.$$

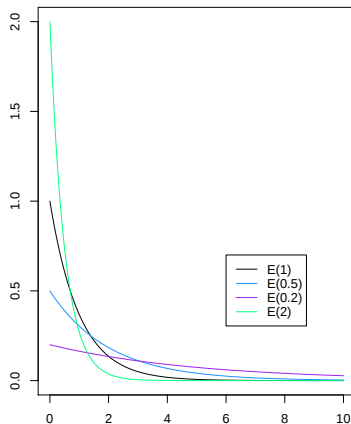
La loi exponentielle est une loi sans mémoire :

$$\mathbb{P}[T > t + s | T > s] = \mathbb{P}[T > t].$$

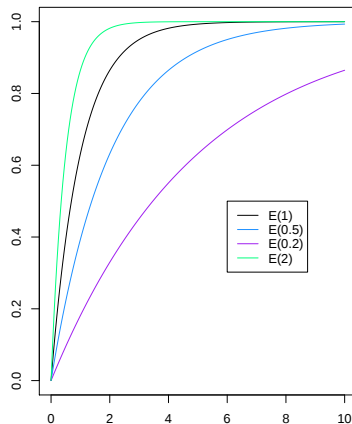
Elle ne convient donc pas pour modéliser des objets soumis à une usure non-négligable.

Lois exponentielles (suite)

Densité de la loi Exp(1)



Fonction de répartition de la loi Exp(1)



Distributions usuelles discrètes

Les familles de lois de proba usuelles discrètes sont

- lois de Bernoulli
- lois binômiales
- lois hypergéométriques
- lois de Poisson

Ces lois sont paramétrées. Cela signifie que la famille de lois donne la forme générale mais qu'à l'intérieur de ces familles chaque loi dépend de un ou plusieurs nombres appelés **paramètres**.

Lois de Bernoulli

On dit qu'une variable X suit une loi de Bernoulli de paramètre p , notée $\mathcal{B}(p)$, lorsqu'elle prend les valeurs 0 avec probabilité $(1 - p)$ et 1 avec probabilité p : $\mathbb{P}[X = 1] = p$ et $\mathbb{P}[X = 0] = 1 - p$.

$\mathbb{E}[X] = p$ et $\text{var}(X) = p(1 - p)$.

Définir une variable de Bernoulli revient à coder par 1 la réalisation d'un évènement et par 0 sa non-réalisation, autrement dit, $X = 1$ si l'évènement est réalisé et $X = 0$ sinon.

La loi de Bernoulli est utilisée pour coder des caractéristiques à 2 modalités selon un schéma succès/échec, par ex : défectueux/non-défectueux, conforme/non-conforme, ce qui donne : $X = 1$ si la pièce est conforme et $X = 0$ sinon. Un autre exemple est $X = 1$ si X est inférieur à un seuil et $X = 0$ si X dépasse ce seuil.

Lois binômiales

On dit qu'une variable X suit une loi binômiale de paramètres n et p , notée $\mathcal{B}(n, p)$, lorsqu'elle prend ses valeurs parmi $\{0, \dots, n\}$ avec les probabilités suivantes pour $k \in \{0, \dots, n\}$:

$$\mathbb{P}[X = k] = C_n^k p^k (1 - p)^{n-k}.$$

$$\mathbb{E}[X] = np \text{ et } \text{var}(X) = np(1 - p).$$

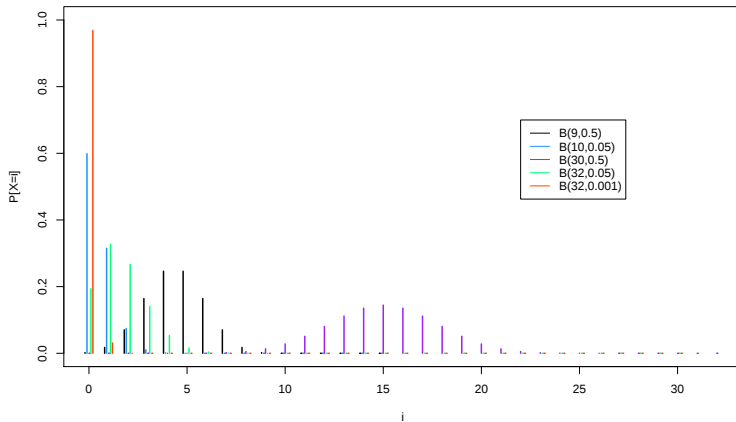
X représente le nombre de fois où un évènement se réalise en n répétitions indépendantes de l'expérience. Par ex, lorsque qu'on prélève au hasard n factures de façon indépendante, X représente le nombre de factures erronées. La loi binômiale sert aussi à compter le nombre de succès lors d'un tirage avec remise de n éléments.

Si $Y_1, \dots, Y_n$ est une suite de variable indépendantes de loi de Bernoulli de même paramètre p , $\mathcal{B}(p)$, alors $\sum_{i=1}^n Y_i$ suit une loi $\mathcal{B}(n, p)$.

Si X_1 et X_2 sont deux variables indépendantes de lois respectives $\mathcal{B}(n_1, p)$ et $\mathcal{B}(n_2, p)$ (lois de même paramètre p !), alors $X_1 + X_2$ suit une loi $\mathcal{B}(n_1 + n_2, p)$.

Lois binômiales

Lois de proba pour la loi binomiale



Lois de Poisson

On dit qu'une variable X suit une loi de Poisson de paramètre λ , $\mathcal{P}(\lambda)$, lorsqu'elle prend ses valeurs dans $\mathbb{N} = \{0, 1, \dots, n, \dots\}$ (infinité de valeurs possibles) avec les probabilités suivantes pour $k \in \mathbb{N}$:

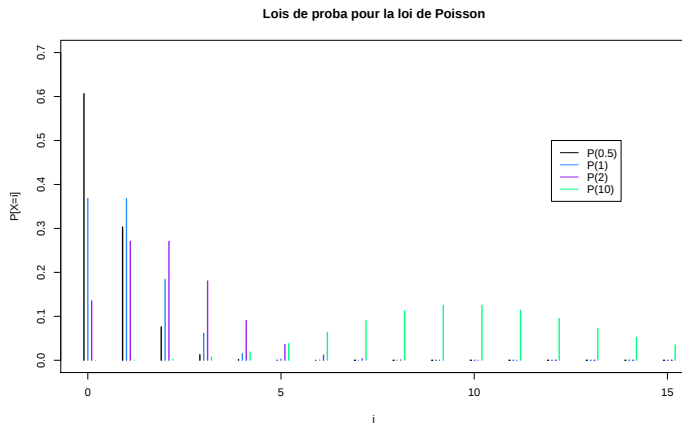
$$\mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}.$$

$\mathbb{E}[X] = \lambda$ et $\text{var}(X) = \lambda$.

X représente le résultat d'un comptage effectué sur une durée fixée, par ex : X compte le nombre d'arrivées de clients à un guichet de banque en une semaine, le nombre de véhicules passant par un péage donné en une journée, le nombre de pannes en un an...

Si $X_1, \dots, X_n$ est une suite de variables indépendantes de loi de Poisson telle que $X_i \sim \mathcal{P}(\lambda_i)$, alors $\sum_{i=1}^n X_i$ suit une loi $\mathcal{P}(\lambda_1 + \dots \lambda_n)$.

Lois de Poisson (suite)



Approximations de lois

Pour $n = 5, 10, 15, 20, 30, 40, 50$ et $p = 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5$, la loi binômiale est tabulée.

Pour n et p satisfaisant $np > 5$ et $n(1 - p) > 5$, on approxime la loi binômiale $\mathcal{B}(n, p)$ par la loi normale $\mathcal{N}(np, np(1 - p))$.

Pour n grand ($n \geq 30$) et p petit ($p < 0.1$) et $np \leq 5$, on approxime la loi binômiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(np)$.

Pour n grand ($n \geq 30$) et p petit ($p < 0.1$) et $np \geq 10$, on approxime la loi binômiale $\mathcal{B}(n, p)$ par la loi normale $\mathcal{N}(np, np)$.

Pour λ grand ($\lambda \geq 10$), on approxime la loi de Poisson $\mathcal{P}(\lambda)$ par la loi normale $\mathcal{N}(\lambda, \lambda)$.

Lois hypergéométriques

Une variable X suit une loi hypergéométrique de paramètres N , n et p , notée $\mathcal{H}(N, n, p)$, lorsqu'elle prend ses valeurs parmi $\{\max(0, n - N(1 - p)), \dots, \min(Np, n)\}$ avec les probabilités suivantes pour $k \in \{\max(0, n - N(1 - p)), \dots, \min(Np, n)\}$:

$$\mathbb{P}[X = k] = \frac{C_{Np}^k C_{N(1-p)}^{n-k}}{C_N^n}.$$

$$\mathbb{E}[X] = np \text{ et } \text{var}(X) = np(1 - p) \frac{N-n}{N-1}.$$

X représente le nombre de fois où un évènement se réalise en n tirages sans remise parmi N éléments, par ex : X représente le nombre de boulons non-conformes obtenus au cours d'un tirage sans remise de n boulons parmi N .

En pratique, on n'hésite pas à substituer la loi $\mathcal{B}(n, p)$ à la loi $\mathcal{H}(N, n, p)$ dès que $n/N \leq 0.1$ (en fait, lorsque $N \rightarrow \infty$, les tirages avec remise ou sans remise sont pratiquement équivalents).