Brochure IREM

n°101_1

Novembre 2023

Des ressources mathématiques pour les enseignants en classe de terminale

Fascicule 1 pour la classe de terminale mathématiques complémentaires

Monique Chappet-Paries
Jacqueline Penninxck
Françoise Pilorge
Aline Robert

ISSN : 0993-6947

INSTITUT DE RECHERCHE SUR L'ENSEIGNEMENT DES MATHEMATIQUES
UNIVERSITE DE PARIS

Imprimé par l'IREM de Paris – Université Paris Cité

Exemplaire **téléchargeable** sur notre site dans la section Publication
https://irem.u-paris.fr/

Coordonnées de l'IREM

Pour venir à l'IREM (il est possible de consulter et d'acheter les publications sur place):
Université Paris Cité, Bâtiment Sophie-Germain,
8 place Aurélie Nemours (sur l'avenue de France), huitième étage,
75013 Paris 13ème arrondissement
(métro 14 -Bibliothèque François Mitterrand ou tramway ligne T3a – Avenue de france)

Nous Contacter

Pour téléphoner: 01 57 27 91 93

Pour écrire à l'IREM concernant les publications:

par voie postale:
Locufier Nadine
IREM de Paris – Case 7018
Université Paris Cité
75205 Paris cedex 13

par voie électronique:
nlocufier@irem.univ-paris-diderot.fr

La liste des publications de l'IREM est mise à jour sur notre site web :

https://irem.u-paris.fr/ressources-en-ligne-de-lirem-de-paris-documents-videos-liens

Pour rester informé:

inscription à la liste de diffusion de l'IREM de Paris également sur le site de l'IREM

Des ressources mathématiques pour les enseignants en classe de terminale

Fascicule 1 : pour la classe de terminale mathématiques complémentaires

Monique Chappet-Paries
Jacqueline Penninxck
Françoise Pilorge
Aline Robert

Des ressources mathématiques pour les enseignants en classe de terminale

Auteures : Monique Chappet-Paries, Jacqueline Penninxck, Françoise Pilorge, Aline Robert

Introduction

A l'origine des ressources présentées ici, une équipe d'auteur.e.s, proches de l'IREM, a produit en 2019, et pour les nouveaux programmes de cette année-là, deux manuels de 2de et 1re spécialité, édités chez Belin – c'est la collection Métamaths. Les livres du professeur associés à ces manuels présentent des analyses didactiques de certains chapitres, comportant notamment des commentaires méta. Ces manuels sont accompagnés de 6 vidéos pour les professeurs qui se trouvent sur le site de l'IREM de Paris à l'onglet : vidéos, sous-onglet : vidéos Manuels... Mais la collection s'est arrêtée là.

Les difficultés liées aux confinements, renforçant celles de la mise en œuvre simultanée des nouveaux programmes de seconde et de première, aggravées par l'éclatement du aux spécialités en première et terminale et à l'option « Mathématiques complémentaires » en terminale ont incité une petite partie de l'équipe à continuer à produire, hors édition scolaire, les quelques ressources qui constituent cette brochure, sur des thèmes choisis pour leur difficulté éventuelle.

Il y en a 3 pour le programme de mathématiques complémentaires et 3 pour les spécialités. Ces documents ont pour ambition de compléter ce que les enseignants peuvent consulter avant leur préparation, sur des chapitres « nouveaux », avec des retours sur les programmes. Deux ressources transversales peuvent être utilisées plus largement, notamment au lycée, et terminent cette offre[1].

En ce qui concerne les mathématiques complémentaires précisément, nous avons tenté de jouer le jeu des programmes en partant, dans les documents produits, des thèmes indiqués et non du découpage en chapitres.

Nous devons souligner que ces ressources n'ont rien à voir avec des chapitres de manuels, ne serait-ce qu'en raison de leur longueur et parce qu'il y a peu d'exercices. Ces documents, écrits pour les professeurs, diffèrent aussi de ressources IREM habituelles, en particulier parce qu'ils n'ont pas été expérimentés dans des classes, à une exception près (Inférence bayésienne) qui n'a pas été complètement analysée, même si cela a déjà donné lieu à quelques enrichissements – les enseignants n'ayant pas tout expérimenté, vu la taille du document. Cependant nous nous inscrivons dans la lignée des deux manuels cités en ce qui concerne la réflexion qui précède la production de ressources, intégrant les difficultés éventuelles des élèves et les liens entre les connaissances nouvelles et anciennes. De nombreux commentaires méta émaillent ces textes.

Pour chaque thème, deux fichiers complets sont mis à disposition, l'un en pdf et l'autre en doc, un lien vers ces fichiers est indiqué **dans la table des matières ci-jointe**. Ainsi, un collègue peut utiliser une partie de document doc, voire l'adapter pour s'en servir en classe (pour un exercice, une démonstration, etc.).

Cela nous semble d'autant plus important que les documents sont longs et trop complets pour la classe.

---

[1] F. Héraut a assuré la relecture de l'ensemble.

LISTE DES RESSOURCES

La numérotation est double, l'une pour l'ensemble du document (en bas, au milieu) indiquée dans la table des matières ci-dessous et l'autre pour chaque ressource (en bas à droite)

Fascicule 1 : pour la classe de terminale mathématiques complémentaires

Liens pour télécharger les documents en docx ou en pdf

Correlation docx
http://docs.irem.univ-paris-diderot.fr/up/B101/Bayes_Maths_comp.docx

Correlation pdf
http://docs.irem.univ-paris-diderot.fr/up/B101/Correlation.pdf

Bayes docx
http://docs.irem.univ-paris-diderot.fr/up/B101/Bayes_Maths_comp.docx

Bayes pdf
http://docs.irem.univ-paris-diderot.fr/up/B101/Bayes_Maths_comp.pdf

Temps_d_attente docx
http://docs.irem.univ-paris-diderot.fr/up/B101/Temps_d_attente.docx

Temps_d_attente pdf
http://docs.irem.univ-paris-diderot.fr/up/B101/Temps_d_attente.pdf

Fascicule 2 : pour la classe de terminale enseignement de spécialité

Deux ressources plus transversales

Des calculs d'aires élémentaires, rappelés au début, on passe aux calculs plus compliqués (approximations), puis on aborde les aires comme surfaces sous une courbe associées aux intégrales. C'est donc une ressource adaptable à différents publics

Cette ressource est indépendante des programmes en tant que tels mais peut être proposée dans différentes classes et à différents niveaux tant pour son intérêt historique que pour les problèmes qu'elle permet d'aborder. Plusieurs domaines sont en jeu (géométrie plane, nombres complexes, analyse, arithmétique, algèbre linéaire).

Liens pour télécharger les documents en docx ou en pdf

Combinatoire_et_denomb docx
http://docs.irem.univ-paris-diderot.fr/up/B101/Combinatoire_et_denomb.docx

Combinatoire_et_denomb
pdf http://docs.irem.univ-paris-diderot.fr/up/B101/Combinatoire_et_denomb.pdf

Concentration-Loi_GN docx
http://docs.irem.univ-paris-diderot.fr/up/B101/Concentration-Loi_GN.docx

Concentration-Loi_GN pdf
http://docs.irem.univ-paris-diderot.fr/up/B101/Concentration-Loi_GN.pdf

Vecteurs_droites_plans docx
http://docs.irem.univ-paris-diderot.fr/up/B101/Vecteurs_droites_plans.docx

Vecteurs_droites_plans pdf
http://docs.irem.univ-paris-diderot.fr/up/B101/Vecteurs_droites_plans.pdf

Calculs d_aires docx
http://docs.irem.univ-paris-diderot.fr/up/B101/Calculs_d_aires.docx

Calculs d_aires pdf
http://docs.irem.univ-paris-diderot.fr/up/B101/Calculs_d_aires.pdf

Nombre_d_or docx
http://docs.irem.univ-paris-diderot.fr/up/B101/Nombre_d_or.docx

Nombre_d_or pdf
http://docs.irem.univ-paris-diderot.fr/up/B101/Nombre_d_or.pdf

CORRELATION ET CAUSALITE

Statistiques à une variable, à deux variables

Le document peut être téléchargé ici :

http://docs.irem.univ-paris-diderot.fr/up/B101/Correlation.docx

http://docs.irem.univ-paris-diderot.fr/up/B101/Correlation.pdf

Introduction

La notion de causalité apparait dans beaucoup de champs, scientifiques ou non, mais pas directement en mathématiques. Causalité ou coïncidence ou raisons d'être, hasard ou incertitude ou déterminisme, relations fonctionnelles multiples, autant de notions proches qui ont un rapport avec la recherche d'explications des phénomènes, voire leur prévision, et qui occupent beaucoup les scientifiques.

Ce qui apparait en mathématiques en revanche ce sont les notions de corrélation, covariance, coefficient de corrélation, droite de régression… autant de caractères numériques, attachés en général à deux séries statistiques, qui livrent une information numérique sur une dépendance éventuelle, le mot dépendance étant pris dans un sens commun, comme une forme de relation. Mais toute interprétation de cette dépendance éventuelle, notamment causale, qu'elle soit temporelle ou non, sort du champ des mathématiques. En témoignent des exemples simples dont nous donnons quelques-uns ci-dessous, ou encore l'usage, normal en langue naturelle mais souvent abusif en logique formelle (et implicative), des mots donc, alors etc…

La recherche des causes des phénomènes est bien entendu une question ancienne, qui a donné lieu au cours de l'Histoire à des développements renouvelés tant philosophiques que physiques, biologiques, géographiques, et en Sciences Humaines et Sociales. Pour dire les choses schématiquement, une première tendance « naturelle » serait de trouver à un phénomène une cause, unique, qui provoque le phénomène dans un enchainement temporel : on augmente I et V augmente dans un circuit avec seulement une résistance (V = RI). Pourtant I n'est pas « la cause » de V. Si deux séries statistiques présentent une forte corrélation – alors les caractères mesurés auraient un lien causal, quitte à introduire le temps pour mieux le voir… Ce raisonnement causal linéaire mécanique est souvent erroné, voire s'établit comme un obstacle, vu sa prégnance, à un raisonnement plus complexe et cependant nécessaire.

A l'heure actuelle, alors qu'on voit les systèmes complexes et la complexité gagner du terrain en sciences, on trouve dans les champs scientifiques, soit des relations fonctionnelles à plusieurs variables pour rendre compte des causalités, soit des associations probabilistes, mettant en jeu différentes échelles, notamment macroscopiques et microscopiques. On évoque des chaines causales complexes et l'importance des médiations (phénomènes intermédiaires). En fait dans un système complexe il y a des propriétés émergentes qui ne sont pas prédictibles à partir des propriétés composantes, ou qui interviennent à un moment donné. On note ainsi une grande diversité des explications causales, non réductibles les unes aux autres, et certains vont jusqu'à dire qu'il est plus important de définir en creux les contraintes (bornes) et espaces qui restent possibles que directement les causes, hors de portée.

Il n'empêche … qu'il est très intéressant de pouvoir « comparer » deux séries statistiques, pour livrer au travail des non-mathématiciens impliqués des dépendances avec leur degré et leur variabilité…

Dans ce qui suit, après avoir donné quelques exemples pour entrer au cœur des questions, et fait retravailler rapidement les notions antérieures qui serviront, nous abordons les séries statistiques doubles et les définitions utilisées, puis nous développons les notions de coefficient de corrélation et de droite de régression, pour terminer par un retour à des problèmes possibles sur causalité et corrélation et enfin quelques exercices d'application.

Cette ressource s'adresse aux enseignants, charge à eux d'y choisir ce qui pourra intéresser leurs élèves. Le programme correspondant est joint en annexe ainsi que des références, des notes historiques et des compléments mathématiques.

Table des matières

Annexe 6 : Fonctions de deux variables - recherche d'extremums

Annexe 7 : Gauss – Legendre : la méthode des moindres carrés.

I- Corrélation et causalité

1) Un exercice pour mobiliser ses connaissances : une série statistique à une variable avec calcul de moyenne et d'écart-type. (Belin 2de)

Dans une usine, deux machines A et B remplissent des paquets de pâtes de 500 g. Pour vérifier le réglage de chaque machine, on prélève un lot de 150 paquets que l'on pèse un à un ; on obtient les résultats suivants :

Masse (en g)	Effectifs machine A	Effectifs machine B
491	4	4
494	7	5
496	10	10
497	10	9
498	19	17
499	26	25
500	31	35
501	18	25
502	16	10
503	9	10

1) Déterminer la moyenne et l'écart-type des masses arrondis à 10^{-3} près pour la machine A puis pour la machine B.

2) Comparer la production des deux machines.

3) Une machine ne nécessite aucun réglage si les trois conditions sont vérifiées simultanément :

- La moyenne m des masses est comprise entre 499 g et 501 g.
- L'écart-type σ est inférieur à 3 g.
- Au moins 90% des masses x_i des paquets de pâtes vérifient

  $|500 - x_i| \leq 2\ \sigma$.

Une des deux machines nécessite-t-elle un réglage ?

Remarque : dans cet exercice, on étudie deux séries statistiques à une seule variable.

2) Un exercice pour découvrir une série statistique à deux variables.

Contrôle de qualité : niveau BTS (grpt A), tiré de Sigma (p.187)

Une machine produit automatiquement des pièces cylindriques. Réglée initialement pour un diamètre de 8 mm, elle se dérègle en cours d'utilisation. Le but de l'exercice est de déterminer le nombre de pièces que l'on pourra produire avant que le diamètre n'atteigne 8,1 mm.

Afin de contrôler la fabrication et de procéder aux réglages éventuellement nécessaires, on mesure le diamètre de la dernière pièce dans chaque série de dix pièces produites. On note x_i le numéro de la pièce et y_i son diamètre. Les résultats obtenus sont les suivants :

Numéro de la pièce : x_i	10	20	30	40	50	60	70	80	90	100
Diamètre (mm) de la pièce : y_i	8,03	8,00	8,01	8,01	8,02	8,03	8,03	8,04	8,05	8,06

On a cette fois une série statistique double (ou à deux variables) et on cherche s'il existe un lien entre les deux variables.
1) Représenter les points $M_i(x_i\ ;\ y_i)$ associés à la série statistique précédente dans le plan muni d'un repère orthogonal. On prendra pour origine le point de coordonnées (0 ; 8) pour unité, 1 cm pour dix pièces en abscisses et 1 cm pour 0,01 mm en ordonnées.
Que constate-t-on ? Quel renseignement pourrait-on en tirer ?

Cet ensemble de points est appelé le nuage de points associé à la série statistique.

2) On appelle **point moyen du nuage** le point G ayant pour coordonnées $(\bar{x}\ ;\ \bar{y})$ où $\bar{x}$ est la moyenne des x_i et $\bar{y}$ la moyenne des y_i .

Calculer les coordonnées du point G et le placer sur la figure.

3) a. Calculer les coordonnées du point moyen G1 associé aux points du nuage ayant les cinq plus petites abscisses et les coordonnées du point moyen G2 associé aux cinq autres points du nuage. Placer ces deux points sur le graphique. Que constate-t-on ?

b. On considère la droite (G1G2) que l'on prend comme **droite d'ajustement** de la situation. La tracer.

c. Déterminer une équation de (G1G2). Démontrer que G appartient à cette droite.
Cette droite est appelée « Droite de Mayer »

4. Les pièces produites doivent avoir un diamètre de 8 mm, avec une tolérance de 0,1 mm
Déterminer le nombre de pièces que l'on pourra produire avant que le diamètre n'atteigne 8,1 mm

En déduire une méthode pour aborder ce type de situation.

3) Les notions de corrélation et de causalité – une notion mathématique et l'autre pas

Dans la plupart des cours de statistiques, lorsqu'on étudie des séries statistiques à deux variables quantitatives dont on connait des valeurs relevées mais incomplètes, on met en garde sur le fait qu'il ne faut pas confondre corrélation et causalité. On introduit des outils mathématiques comme indicateurs du degré de corrélation entre les deux variables, mais comment savoir si une des deux variables est la ou une cause de l'autre ? Pour distinguer les deux notions causalité/corrélation nous donnons d'abord quelques exemples.

Le mot corrélation est associé à une relation entre variables statistiques dont on connait un certain nombre de valeurs, dans laquelle la connaissance de la valeur d'une variable informe sur la valeur probable d'une autre variable. Par exemple, quand il fait plus chaud, les achats de crème glacée sont plus élevés ; dans ce cas on dit que la température et les ventes de crème glacée sont positivement corrélées. En revanche, si les achats de boissons chaudes diminuent quand il fait plus chaud, nous considérons que la température et les ventes de boissons chaudes sont négativement corrélées. Cependant il y a des situations où les liens ne sont pas évidents, où on s'interroge sur leur « force », où on aurait besoin d'interpréter les résultats pour établir des hypothèses sur des causalités éventuelles. En particulier pour prédire ce qui n'a pas encore été « mesuré », voire n'est pas mesurable.

Il y a deux types de problèmes : établir un lien entre deux séries de valeurs obtenues ensemble, à partir de valeurs incomplètes (cf. variable statistique double) et interpréter ce lien.
Une des interprétations est l'établissement d'une causalité éventuelle.
Donnons d'abord une illustration du premier piège à éviter

a) Effet cigogne : corrélation ou causalité ?

Une erreur de raisonnement courante consiste à dire, pour deux variables X et Y dont on connait un certain nombre de valeurs : « X et Y sont corrélées, donc X cause Y ». On confond alors corrélation et causalité car en réalité, il se pourrait aussi que Y cause X, ou bien que X et Y aient une cause commune Z, ou encore que X et Y soient accidentellement liées mais n'aient aucun lien de causalité.

Par exemple, dans les communes qui abritent des cigognes, le taux de natalité est plus élevé que dans l'ensemble du pays. Conclusion : les cigognes apportent les bébés ! Voici une explication plus probable : les cigognes nichent de préférence dans les villages plutôt que dans les grandes agglomérations, et il se trouve que la natalité est plus forte en milieu rural que dans les villes. C'est d'ailleurs plus vrai en Alsace, sans que la natalité y soit plus élevée !

Voilà pourquoi l'on nomme parfois « **effet cigogne** » cette tendance à confondre corrélation et causalité.

b) Exemples de lien entre deux variables statistiques et interprétation.

Nous donnons ci-dessous 4 exemples différents de deux étapes, lien et interprétation, à partir de séries de mesures associées à deux variables statistiques.

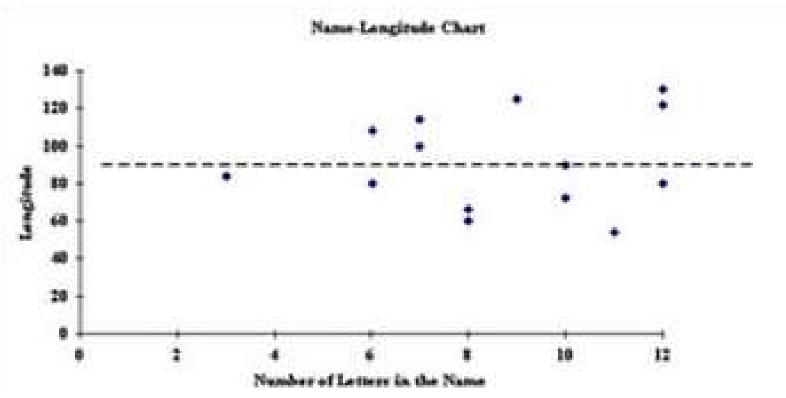
Exemple 1 : Un exemple classique : nombre de prix Nobel obtenus et quantité de chocolat consommée par pays. Corrélation positive mais pas de causalité.

Les deux variables varient dans le même sens (lorsqu'une variable augmente, l'autre augmente également ; ou lorsqu'une variable diminue, l'autre diminue également).
Dans l'exemple 1 ci-dessus, on observe bien que plus la consommation de chocolat s'accroît, et plus le nombre de prix Nobel obtenus augmente. **C'est donc bien une corrélation positive** (la pente de la droite est positive).

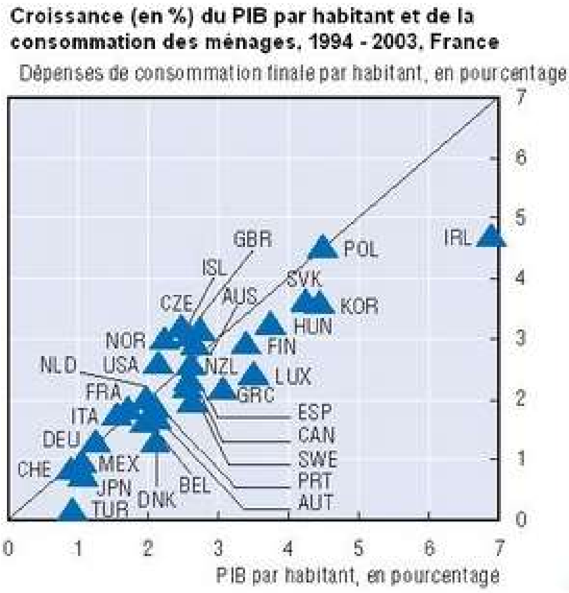
En revanche, **s'il y bien corrélation, il n'y a pas causalité** : consommer du chocolat ne rend pas plus efficace dans une discipline donnée (ce qui se traduirait, très schématiquement, par un plus grand nombre de prix Nobel). **C'est l'existence d'une variable "cachée"** (une troisième variable qui influence la variable X et la variable Y) qui peut conduire à une corrélation qui suggère une causalité. Cette variable, **la richesse d'un pays mesuré par son revenu par habitant**, est tout **à la fois la cause de la consommation de chocolat** (variable X : le chocolat est un produit de luxe, cher, d'autant plus accessible que les revenus sont élevés) **et la cause du nombre de prix Nobel obtenus** (variable Y : un pays riche peut consacrer plus de moyens à son système éducatif pour le rendre plus performant qu'un pays pauvre

Exemple 2 : Lien entre la longitude d'une ville et le nombre de lettres constituant son nom. Une absence de corrélation.

Lorsqu'une variable varie l'autre ne varie pas ou alors sans lien avec la première (on dit que les variables sont indépendantes l'une de l'autre). Les différents points sont particulièrement dispersés sur le graphique, et même si le nuage de points est allongé aucune droite ne semble approcher la série de points. Dans cet exemple, il n'y a aucun lien entre les deux variables. On dira qu'il n'y a pas de corrélation.



Exemple 3 : PIB par habitant et consommation des ménages.
Corrélation mais absence de causalité entre les deux variables (rôle d'une variable cachée)



Dans l'exemple ci-dessus, on observe **une corrélation très nette entre la croissance des dépenses de consommation des ménages et la croissance du PIB par habitant.**

Mais quelle est la causalité ?
En effet cette corrélation peut traduire tout aussi bien :
1) l'impact causal de la croissance des dépenses de consommation sur la croissance du PIB ("plus de dépenses de consommation, donc plus de débouchés, donc plus de production")

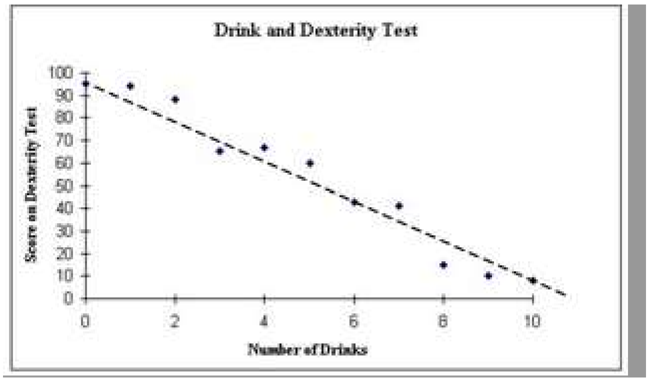
ou

2) l'impact causal de la croissance du PIB sur la croissance des dépenses de consommation ("plus de PIB, donc plus de richesses créées, donc plus de revenus, donc plus de dépenses de consommation")

ou

3) l'impact causal des deux à la fois

Exemple 4 : Score à un test de dextérité en fonction du nombre de verres d'alcool bus par un individu. Corrélation négative et causalité.



9

Les deux variables varient en sens inverse (lorsqu'une variable augmente, l'autre diminue). Dans l'exemple 4 ci-dessus, plus le nombre de verres consommés augmente, plus la dextérité diminue. **C'est donc bien une corrélation négative** (pente de la droite négative).

On peut interpréter en termes de causalité, l'intérêt est de mesurer l'importance de la relation.

La consommation d'alcool augmentant, les réflexes s'émoussent, l'individu est de moins en moins habile, il perd en dextérité. C'est bien l'alcool qui est responsable de cet état. **Il y a donc à la fois corrélation et causalité.**

<u>Conclusion</u> : Attention, les observations seulement statistiques sont parfois trompeuses. Il faut donc se montrer très prudent, faire preuve de bon sens ou de connaissances extra-mathématiques et bien prendre le temps de réfléchir ....

Le mot corrélation est associé à une relation statistique, entre variables dont on connait un certain nombre de valeurs, dans laquelle la connaissance de la valeur d'une variable informe sur la valeur probable d'une autre variable. Cependant il y a des situations où les liens ne sont pas évidents, où on s'interroge sur leur « force », où on aurait besoin d'interpréter les résultats pour établir des hypothèses sur des causalités éventuelles. En particulier pour prédire ce qui n'a pas encore été « mesuré », voire n'est pas mesurable.

Le plus souvent, c'est à partir d'une **représentation graphique** que l'on **met en évidence l'existence, ou l'absence, d'une corrélation entre 2 variables ainsi que la nature de cette corrélation.** Cette représentation graphique prend (souvent, mais pas seulement) la forme d'un **nuage de points** dont les coordonnées sont les valeurs prises par la variable X (sur l'axe des abscisses) et par la variable Y (sur l'axe des ordonnées).

Dans certains cas **le nuage de points obtenu peut être approché par une droite**, par exemple la droite dite de Mayer comme on l'a vu dans l'exercice d'introduction, ou encore la droite dite des moindres carrés (ou droite de régression ou encore d'ajustement affine), dont **le profil et la pente renseignent sur la présence et le sens de la corrélation** : **plus les différents points sont proches ("alignés") de la droite et plus la corrélation est forte**. La pente de la droite indique si la corrélation est positive ou négative.

II - Corrélation

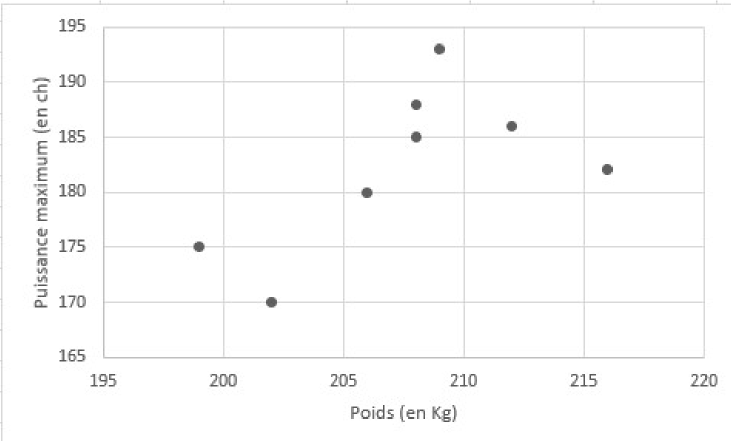
Dans une première partie il s'agit de montrer sur différents exemples comment la forme du nuage de points, associé aux données numériques dont on dispose pour les deux variables statistiques en jeu, peut conduire à chercher un lien entre ces deux variables.
Dans la deuxième partie on présente un premier indicateur d'une liaison possible entre deux variables, la covariance, et ce qu'elle permet de renseigner : à la fois comparer

simultanément les deux variables à leur moyenne respective et permettre de savoir si elles varient dans le même sens ou non. Mais, si par exemple elles varient dans le même sens, est-ce que cette relation est forte ou non ?
Dans la troisième partie et dans le cas où la relation est linéaire (nuage de points proche d''une droite), on introduit un autre indicateur d'un lien éventuel entre deux variables statistiques, le coefficient de corrélation linéaire et on en donne une interprétation en termes de force, dans le cas où ce coefficient est proche de 1 ou -1.

1) Exemples pour comprendre

a) Puissances de motos

On a relevé le poids (en kg) et la puissance maximum (en chevaux) de quelques motos de compétition et on se demande si le poids est en relation avec la puissance. On a placé les points de coordonnées (poids ; puissance) dans un repère et obtenu le nuage de points ci-contre.

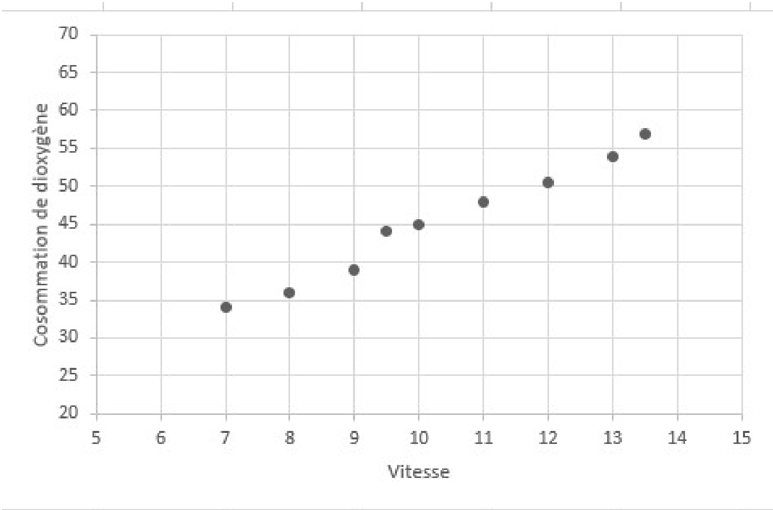


La forme de ce nuage de points ne se rapproche d'aucune droite ni d'aucune courbe connue. Il semble qu'on ne puisse pas trouver de corrélation entre poids et puissance.

b) Test d'effort

On a mesuré chaque minute la consommation de dioxygène (en mL.min^{-1}.kg^{-1}) d'un coureur lorsque sa vitesse (en km.h^{-1}) augmente afin de savoir s'il existe un lien entre les deux. Le nuage de points (vitesse ; consommation de O^2) est représenté ci-contre :

On constate que les points sont à peu près alignés, la forme du nuage évoque une droite. Il semble y avoir une relation de type affine entre les deux variables.



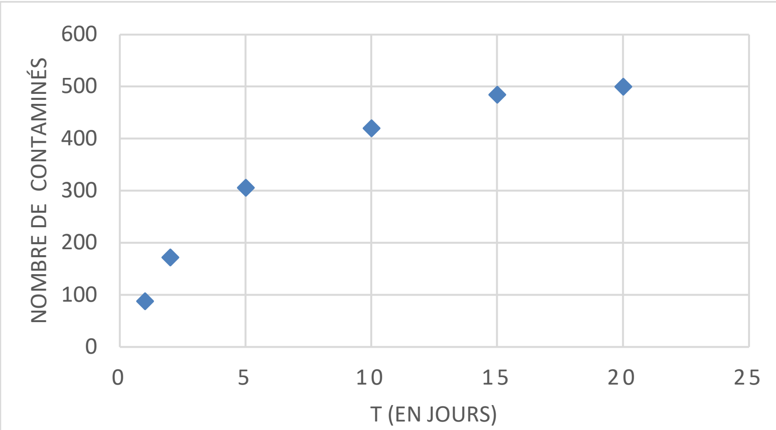
11

c) Progression d'une épidémie

Pour étudier la progression d'une épidémie, on procède à une enquête sur un échantillon de personnes. Le tableau ci-dessous indique le nombre N(t) d'individus ayant été contaminés à la date t exprimée en jours.

T (en jours)	1	2	5	10	15	20
N(t)	88	172	306	420	485	500

Pour préciser l'évolution de l'épidémie en fonction du nombre de jours on a placé les points de coordonnées (t, N(t)) dans le graphique ci-dessous. A partir de ce nuage de points, on se demande si on pourrait prévoir le nombre de personnes contaminées au bout de 30 jours.



Des nuages de points différents qui conduisent à un questionnement

Dans les trois situations ci-dessus, on étudie simultanément deux variables quantitatives, X et Y. On dispose de **relevés statistiques** concernant X et Y sous forme de n couples de valeurs $(x_i\ ;\ y_i)$ qui forment une **série statistique double** (à deux variables ou bivariée).
On souhaite :

- déterminer s'il existe une relation (liaison) entre X et Y ; et si oui
- caractériser la forme de cette relation ;
- quantifier son intensité.

Ce lien statistique, s'il existe, s'appelle corrélation entre les 2 variables X et Y.

Si on peut exprimer la relation entre X et Y sous une forme fonctionnelle, et si l'intensité de la relation est forte, dans un sens que l'on précisera plus loin, on peut alors faire des interpolations et/ou des extrapolations pour évaluer une *valeur probable* de Y pour une valeur donnée de X.
X est appelée la variable explicative et Y la variable expliquée. On peut aussi se demander si Y est causé par X.

La recherche d'une relation éventuelle entre X et Y commence le plus souvent par une étude exploratoire. L'observation de la forme du nuage permet de conjecturer l'existence ou l'absence d'une corrélation entre les deux variables ainsi que la nature de cette corrélation. Les statisticiens se sont dotés d'indicateurs pour étudier une éventuelle relation entre les deux variables X et Y.

Cette approche est différente de celle que l'on rencontre lorsque l'on dispose de deux séries de données d'une **même** variable statistique quantitative qu'il s'agit de comparer. Pour cela on peut résumer les deux séries de valeurs par des indicateurs tels que moyenne et écart-type, voire médiane et quartiles et interpréter ces indicateurs (voir par exemple l'exercice « Mobiliser »)

2) La covariance : un premier indicateur d'une liaison entre deux variables statistiques

a) Introduction et définition

Pour une série statistique quantitative à une seule variable X, l'écart-type, par l'intermédiaire de la variance, permet d'évaluer la dispersion des valeurs de la série par rapport à la moyenne de ces valeurs.

<u>Rappels</u> : Soit x_i les n valeurs de la série statistique. La moyenne est donnée par $\bar{x} = \frac{1}{n}\sum_1^n x_i$
et la variance, V, est égale à $\frac{1}{n}\sum_1^n (x_i - \bar{x})^2 = \frac{1}{n}[\sum_1^n (x_i)^2] - \bar{x}^2$. La valeur de la variance est donc positive.
L'écart-type, noté souvent s ou σ, est égal à la racine carrée de la variance : $\sigma = \sqrt{V}$
Il s'exprime avec la même unité que les valeurs x_i de la variable *X.* C'est un indicateur de la dispersion des valeurs x_i par rapport à la moyenne $\bar{x}$.

Dans le cas d'une série statistique double (à deux variables dite aussi, bivariée), par analogie avec la variance, on introduit la covariance.
Avec les notations ci-dessus, par définition la covariance de X et de Y, notée $Cov(X\,;Y)$ est donnée par $Cov(X;Y) = \frac{1}{n}\sum_1^n (x_i - \bar{x})(y_i - \bar{y})$

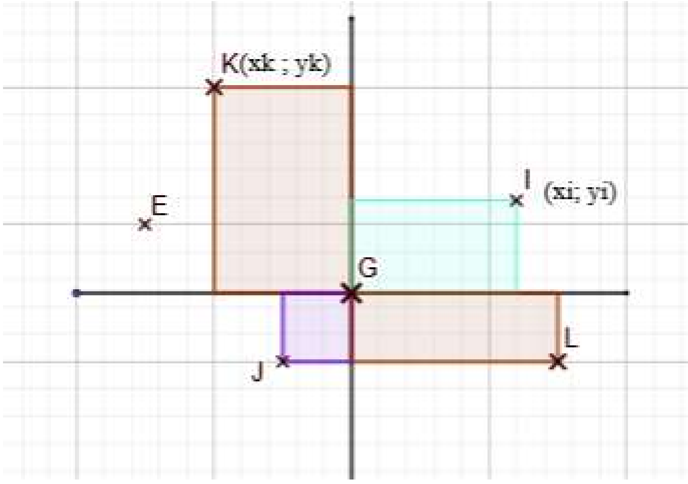
b) Ce que renseigne la covariance

La covariance mesure la tendance des deux variables à être simultanément au-dessus ou en dessous de leurs moyennes respectives. Elle permet de savoir si, lorsque les valeurs de X sont en moyenne supérieures à $\bar{x}$, celles de Y ont tendance à être supérieures (ou inférieures) à $\bar{y}$.

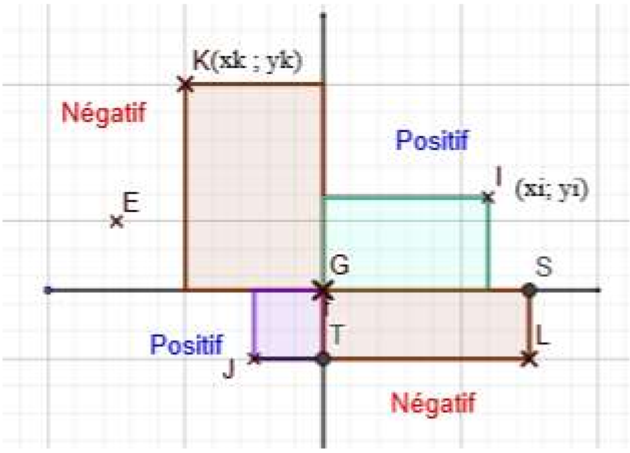
Le signe de la covariance indique si X et Y ont tendance à varier dans le même sens. En effet, la covariance Cov(x, y) aura tendance à être positive si, pour de nombreux $i \in \{1, \ldots, n\}$, on a $x_i > \bar{x}$ et $y_i > \bar{y}$ ou bien si $x_i < \bar{x}$ et $y_i < \bar{y}$.

Autrement dit
- Cov(X, Y) > 0 si les variables X et Y ont tendance à varier dans le même sens,
- Cov(X, Y) < 0 si les variables X et Y ont tendance à varier en sens inverse.

Pour illustrer ceci on peut définir des quadrants en prenant le point moyen $G(\bar{x}\,;\bar{y})$ pour origine.



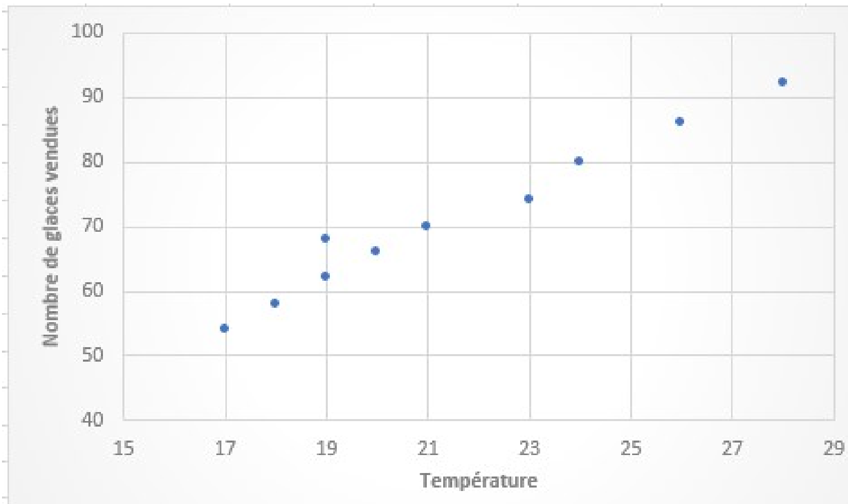
$(x_K - \bar{x})(y_K - \bar{y})$ est négatif puisque $x_K < \bar{x}$ et $y_K > \bar{y}$. L'aire du rectangle délimité par G et K, comme sur la figure ci-dessus, est donc égale à $-(x_K - \bar{x})(y_K - \bar{y})$
$(x_I - \bar{x})(y_I - \bar{y})$ est positif puisque $x_I > \bar{x}$ et $y_I > \bar{y}$. L'aire du rectangle délimité par G et I est donc égale à $(x_I - \bar{x})(y_I - \bar{y})$
On attribue donc un signe à chacun des quatre quadrants et un point qui appartient à une zone contribue d'autant plus au signe de la zone qu'il est plus ou moins éloigné du point moyen et des limites de la zone.



Remarques :
1- La covariance a une unité qui est le produit des unités de X et de Y)
2- La covariance est un nombre qui peut prendre n'importe quelle valeur, autrement dit $Cov(X,Y)\epsilon\mathbb{R}$ (voir exemples ci-dessous). La valeur absolue de la covariance n'a pas de signification particulière.

Exemple 1 : on a relevé pendant dix jours la température (en degrés Celsius) et le nombre de crèmes glacées vendues (en centaines). On a représenté le nuage de points et le calcul de la covariance est donné ci-dessous :

Température x_i	Nombre de crèmes glacées y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
17	54	-4,5	-17	76,5
18	58	-3,5	-13	45,5
19	62	-2,5	-9	22,5
19	68	-2,5	-3	7,5
20	66	-1,5	-5	7,5
21	70	-0,5	-1	0,5
23	74	1,5	3	4,5
24	80	2,5	9	22,5
26	86	4,5	15	67,5
28	92	6,5	21	136,5
$\bar{x} = 21{,}5$	$\bar{y} = 71$			$Covariance = 39{,}1$

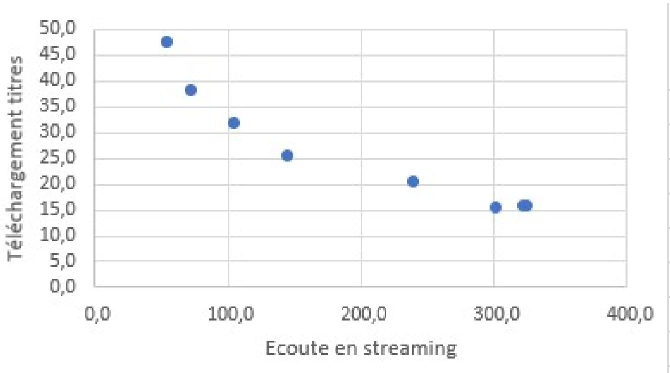


Remarque : on verra plus loin que l'on peut obtenir directement ce résultat à la calculatrice ou avec un tableur.

15

Exemple 2 : le tableau ci-dessous donne l'évolution de 2013 à 2020 du chiffre d'affaires (en millions d'euros) sur deux marchés différents de vente de musique.

Année	Écoute en flux (en streaming)	Téléchargement de titres
2013	54,0	47,6
2014	72,0	38,2
2015	104,0	31,9
2016	144,0	25,6
2017	239,0	20,6
2018	301,0	15,5
2019	322,0	16
2020	325,0	15,8



Calcul de la covariance

Année	Écoute en flux (en streaming) x_i	Téléchargement de titres y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
2013	54,0	47,6	-141,1	21,2	-2991,850
2014	72,0	38,2	-123,1	11,8	-1452,875
2015	104,0	31,9	-91,1	5,5	-501,188
2016	144,0	25,6	-51,1	-0,8	40,900
2017	239,0	20,6	43,9	-5,8	-254,475
2018	301,0	15,5	105,9	-10,9	-1154,038
2019	322,0	16	126,9	-10,4	-1319,500
2020	325,0	15,8	129,9	-10,6	-1376,675
	$\bar{x}$= 195,1	$\bar{y}$= 26,4			Cov = -1126,213

c) Une autre expression de la covariance

La plupart du temps les valeurs calculées des moyennes sont des valeurs approchées. Par suite, les erreurs sur chacune des n valeurs $x_i - \bar{x}$ et $y_i - \bar{y}$ se cumulent.

Mais lorsqu'on développe l'expression $\frac{1}{n}\sum_1^n (x_i - \bar{x})(y_i - \bar{y})$ on obtient

$$Cov(X;Y) = \frac{1}{n}\left(\sum_1^n x_i y_i - \sum_1^n x_i \bar{y} - \sum_1^n \bar{x} y_i + \sum_1^n \bar{x}\bar{y}\right)$$
$$= \frac{1}{n}\left(\sum_1^n x_i y_i\right) - \left(\frac{1}{n}\sum_1^n x_i\right) \times \bar{y} - \bar{x}\left(\frac{1}{n}\sum_1^n y_i\right) + \frac{1}{n} \times n\bar{x}\bar{y}$$
$$= \frac{1}{n}\left(\sum_1^n x_i y_i\right) - \bar{x} \times \bar{y} - \bar{x} \times \bar{y} + \bar{x}\bar{y}$$

D'où la formule $\boldsymbol{Cov(X;Y) = \frac{1}{n}[\sum_1^n x_i y_i] - \overline{x}\overline{y}}$

On peut remarquer l'analogie avec la formule « développée » de la variance pour une seule variable statistique.

3) Le coefficient de corrélation linéaire

a) Introduction et définition

La covariance nous indique si, *en moyenne*, les variables statistiques X et Y varient dans le même sens ou non mais ne nous donne pas de renseignement sur la « force » de la corrélation entre X et Y.

Le coefficient de corrélation de Pearson (ou de Bravais-Pearson) mesure la liaison linéaire existant entre deux variables quantitatives aléatoires. Il a pour but de déterminer la « force » d'une relation *linéaire* entre X et Y.

Karl Pearson (1857-1936), mathématicien britannique est un des fondateurs de la statistique moderne appliquée à la biomédecine.

Définition :

Le coefficient de corrélation de Pearson est le quotient de la covariance de *X* et de Y par le produit des écarts type de *X* et de *Y*. On le note souvent r ou ρ.

Soit $\sigma(X)$ et $\sigma(Y)$ les écarts-type de *X* et *Y*, alors $r = \frac{Cov(X,Y)}{\sigma(X)\times\sigma(Y)}$

Remarques :

1 - On sait que les variables statistiques et les écarts-type correspondants s'expriment dans la même unité. Par suite le coefficient de corrélation est un nombre sans dimension (sans unité). Ce qui permet de comparer des coefficients de corrélation pour des séries de variables statistiques doubles (ou bivariées), ce qui n'est pas le cas pour les covariances.

2 - $Cov(X,Y)$ et r sont de même signe. En effet, un écart-type est un nombre positif.

Propriété : le coefficient de corrélation est compris entre -1 et 1

$$-1 \leq r \leq 1$$

Cette propriété peut être admise.
Elle requiert les propriétés de la variance et de la covariance (1), (2) et (3) qui sont assez faciles à établir.

(1) $V(aX) = a^2V(X)$

En effet $V(aX) = \frac{1}{n}\sum_1^n(ax_i - a\bar{x})^2 = \frac{1}{n}\sum_1^n a^2(x_i - \bar{x})^2 = a^2\frac{1}{n}\sum_1^n(x_i - \bar{x})^2 = a^2V(X)$

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

Par définition $V(X+Y) = \frac{1}{n}\sum_1^n[(x_i + y_i) - (\overline{x+y})]^2 = \frac{1}{n}(\sum_1^n[(x_i - \bar{x}) + (y_i - \bar{y})]^2)$

V(X+Y)=$\frac{1}{n}\left(\sum_1^n[(x_i-\bar{x})^2+2(x_i-\bar{x})(y_i-\bar{y})+(y_i-\bar{y})^2]\right)$

V(X+Y)= $\frac{1}{n}\sum_1^n(x_i-\bar{x})^2+2\frac{1}{n}\sum_1^n(x_i-\bar{x})(y_i-\bar{y})+\frac{1}{n}\sum_1^n(y_i-\bar{y})^2$

On reconnait les variances de X et de Y et la covariance de (X,Y), d'où

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

(3) $Cov(X,aY)=aCov(X,Y)$. *En effet* $Cov(X,aY)=\frac{1}{n}\sum_1^n[(x_i-\bar{x})(ay_i-a\bar{y})]$

$$Cov(X,aY)=a\frac{1}{n}\sum_1^n[(x_i-\bar{x})(y_i-\bar{y})]=aCov(X,Y)$$

<u>Démonstration de la propriété : $-1\leq r\leq 1$</u>

D'après les propriétés (1), (2) et (3) on a$\mathrm{V(X+tY)=V(X)+2tCov(X,Y)+t^2V(Y)}$

La variance de $X+tY$ est un polynôme du second degré en t, positif ou nul quel que soit t puisqu'une variance est positive ou nulle. Le coefficient de t^2 est strictement positif.

Donc $\mathrm{V(X)+2tCov(X,Y)+t^2V(Y)}$ a au plus une racine d'où son discriminant est négatif ou nul. Or $\Delta=4[Cov(X,Y)]^2-4V(Y)V(X)$, d'où $4[Cov(X,Y)]^2-4V(Y)V(X)\leq 0$

Donc $\frac{[Cov(X,Y)]^2}{V(X)\times V(Y)}\leq 1$. On en déduit $-1\leq\frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}\leq 1$ ou encore $-1\leq r\leq 1$

b) Interprétation du coefficient de corrélation linéaire

On peut démontrer que si tous les points de coordonnées $(x_i\,;y_i)$ des données statistiques sont alignés alors le coefficient de corrélation est égal à -1 ou à 1, et réciproquement.
Mais ce n'est en général pas le cas.

<u>**i)** On considère que plus le coefficient de corrélation linéaire est proche de -1 ou de 1, plus les points sont proches d'une droite et plus la liaison entre les variables X et Y est forte</u>. Cependant l'interprétation d'un coefficient de corrélation dépend du contexte et des objectifs. Une corrélation de 0,9 peut être considérée comme faible si l'on vérifie une loi physique en utilisant des instruments de qualité, mais peut être considérée comme très élevée dans les sciences sociales.

C'est le cas de l'exemple 1 du paragraphe 1) b). (Température/crèmes glacées)

Pour calculer le coefficient de corrélation, il suffit de compléter le tableau par le calcul des $(x_i-\bar{x})^2$ et $(y_i-\bar{y})^2$ afin de déterminer les écarts-type $\sigma(X)$ et $\sigma(Y)$.

Température x_i	Nombre de crèmes glacées y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
17	54	-4,5	-17	76,5	20,25	289
18	58	-3,5	-13	45,5	12,25	169
19	62	-2,5	-9	22,5	6,25	81
19	68	-2,5	-3	7,5	6,25	9
20	66	-1,5	-5	7,5	2,25	25
21	70	-0,5	-1	0,5	0,25	1
23	74	1,5	3	4,5	2,25	9
24	80	2,5	9	22,5	6,25	81
26	86	4,5	15	67,5	20,25	225
28	92	6,5	21	136,5	42,25	441
$\bar{x} = 21{,}5$	$\bar{y} = 71$			$Covariance = 39{,}1$	*Variance de X* *11,85*	*Variance de Y* *133*

On a donc $r = \frac{39{,}1}{\sqrt{11{,}85} \times \sqrt{133}} \approx 0{,}98$. Ce coefficient indique une forte corrélation et est compatible avec le nuage de points associé (voir page 15)

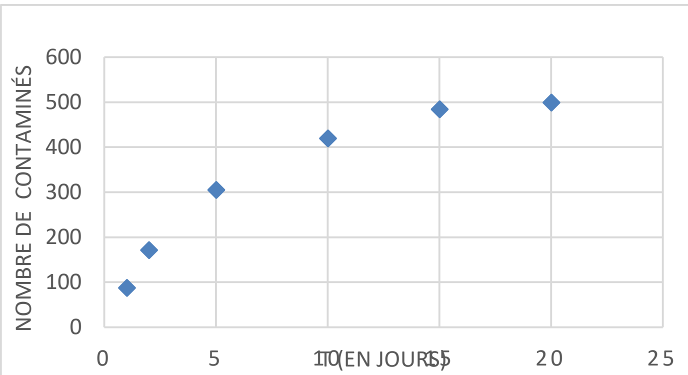
<u>Remarque</u> : ce coefficient est donné directement avec une calculatrice ou un tableur.

ii) Un coefficient de corrélation proche de 0 indique qu'il n'y a pas de relation affine entre les deux variables qui peuvent cependant être fortement corrélées de façon non affine(ceci sera précisé dans le paragraphe III).

Par exemple, en reprenant les données de l'enquête sur un échantillon de personnes (exemple c) pour comprendre page 12) où N(t) est le nombre d'individus ayant été contaminés à la date t exprimée en jours on trouve que le coefficient de corrélation r est environ égal à 0,117.

T (en jours)	1	2	5	10	15	20
N(t)	88	172	306	420	485	500

Là encore la valeur de r est compatible avec la forme du nuage de points. Cependant, l'utilisation du coefficient de corrélation linéaire n'est pas pertinente.

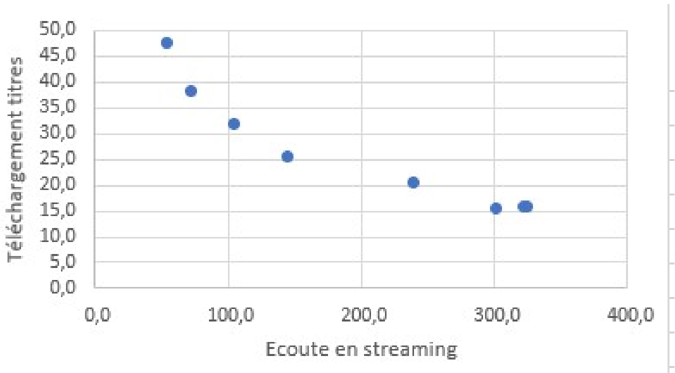


19

iii) Un coefficient de corrélation peut être considéré comme satisfaisant ($|r|$ compris entre 0,6 et 1) bien que le nuage de points ne suggère pas des points proches d'une droite. Dans ce cas c'est que r n'est pas un indicateur pertinent. C'est le cas de l'exemple 2 dont les données sont rappelées ci-dessous. Le coefficient de corrélation est environ égal à -0,94

Écoute en flux (en streaming)	54,0	72,0	104,0	144,0	239,0	301,0	322,0	325,0
Téléchargement de titres	47,6	38,2	31,9	25,6	20,6	15,5	16	15,8

La corrélation est forte, pourtant le nuage de points suggère plutôt une liaison non affine.



En résumé, le sens de la relation est indiqué par le signe du coefficient de corrélation$r$alors que l'intensité de la relation est donnée par la valeur absolue de r. Mais l'utilisation d'un coefficient de corrélation n'est pas toujours pertinent comme on l'a vu dans l'exemple précédent. Cependant si le nuage de points suggère une proximité avec une courbe de fonction connue, on peut y recourir après un changement de variable. On propose quelques exemples en exercices.

Par ailleurs, même si deux variables sont fortement corrélées, l'évolution de l'une n'est pas nécessairement le résultat de l'évolution de l'autre, ce qui s'interpréterait comme une relation de causalité établissant qu'un changement dans cette variable produit un changement dans l'autre variable.
Alors qu'une corrélation qualifie seulement le fait que deux choses évoluent ensemble, une causalité implique un mécanisme expliquant l'association : la causalité est donc un concept plus restrictif. La corrélation ne signifie pas qu'il existe une relation de cause à effet entre les variables.

III – Régression affine simple

A partir d'un exemple simple, on explique d'abord ce qui est cherché dans ce paragraphe pour préciser certaines corrélations, en prenant en compte une certaine dispersion des points. On donne ainsi sur cet exemple plusieurs manières de construire une droite qui ajuste au mieux les points représentant les données, en faisant varier cet « au mieux ».

Le paragraphe suivant est dédié à la recherche de la droite de régression par la méthode des moindres carrés dans le cas général, illustré cependant sur des exemples par différents modes de calcul (à la main, sur tableur, sur calculatrices).

Le paragraphe 3 présente le coefficient de détermination qui indique en quelque sorte l'importance de l'intervention de la variable X dans la corrélation avec la variable Y (en pourcentage, donc de manière « moyenne »). C'est le carré du coefficient de corrélation dans le cas où la régression est affine.

Le dernier paragraphe indique, à partir d'exemples, une manière pratique de se ramener au cas d'une régression affine, plus facile à étudier, quand c'est possible, par un changement de variable.

1) Un exemple pour comprendre : position du problème

Superficie	Prix
26	165
30	170
35	170
38	245
40	240
50	320
55	290

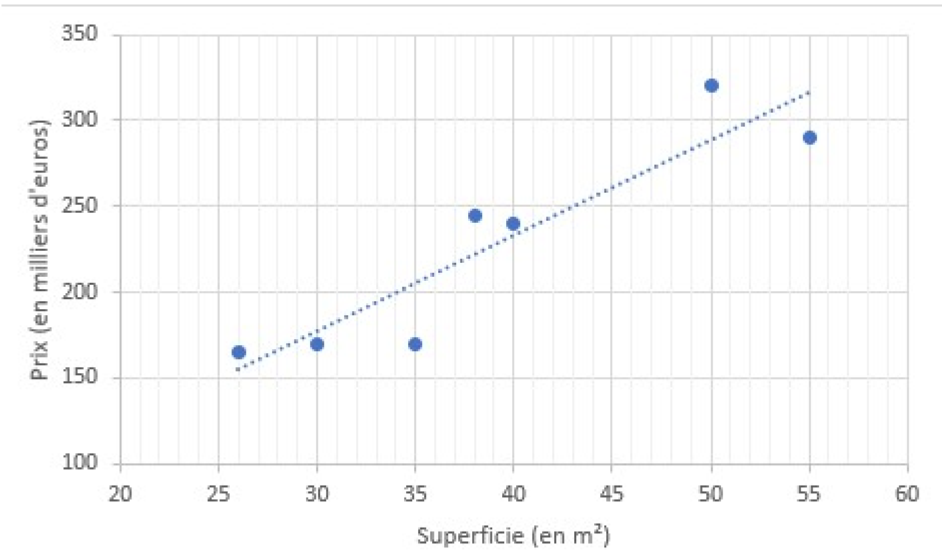
Sur un site de ventes de logements, on a relevé les superficies en m² (X) et les prix de ventes en milliers d'euros (Y) de quelques logements.

En vue d'acheter un logement, on veut étudier le lien entre ces deux variables : quel prix paraitrait raisonnable en fonction de la superficie ?

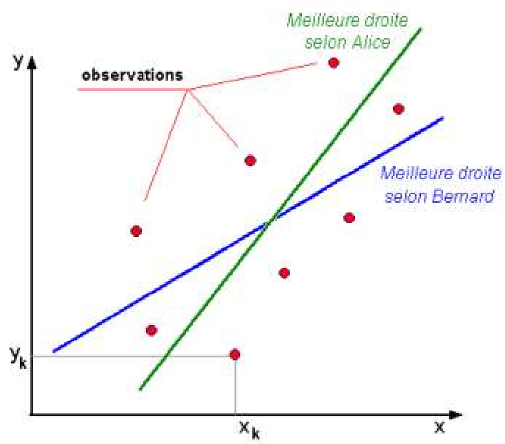
a) Première approche

Une première approche simple consiste à représenter ces données par un nuage de points : on indique la superficie en abscisse et le prix en ordonnée (voir ci-dessous). Selon la forme de ce nuage, on peut chercher la droite qui **traduit au mieux la manière dont y varie par rapport à x**, ce qui donnera une relation affine de la forme $y = ax + b$ (a, b $\in \mathbb{R}$)

On a tracé ci-dessous une droite qui pourrait « s'ajuster » au nuage de points, mais on pourrait en choisir une autre qui paraitrait plus appropriée.



Si on se contente de tracer à main levée la droite qui "passe au mieux" près des points représentatifs, différentes personnes vont obtenir des résultats différents. Ainsi, par exemple, sur le schéma ci-contre, la meilleure droite pour Alice n'est pas la meilleure pour Bernard ! Il faut donc choisir une méthode pour déterminer la droite qui passe au plus près de l'ensemble des points.



b) Une première méthode pour avoir une même droite pour tous : la méthode de Mayer

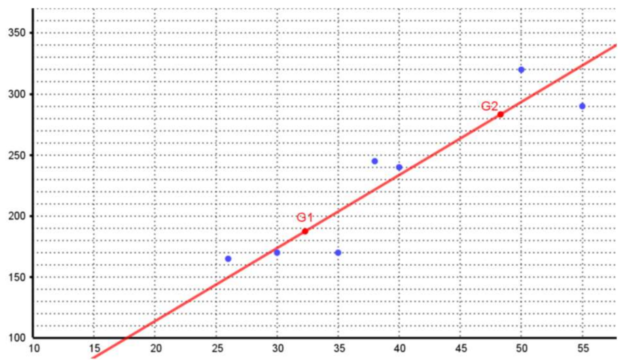
Cette méthode consiste à :

- fractionner le nuage en deux nuages dont les effectifs sont égaux ou diffèrent de 1 ;
- calculer les coordonnées des points moyens G1 et G2 des deux parties du nuage. - Déterminer l'équation de la droite (G1G2).

Dans la série précédente on peut considérer les quatre premiers points pour déterminer G1 et les trois derniers pour G2.

(on a déjà rencontré cette méthode dans l'exercice d'introduction au I 2.).

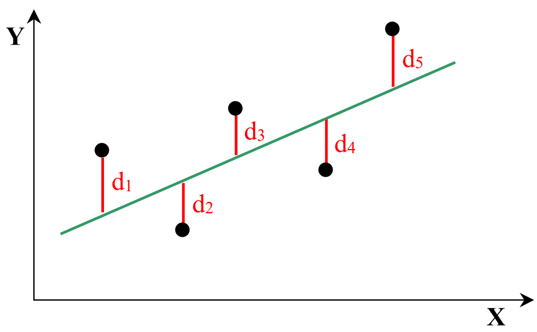
On trouve G1(32,3 ;187,5) et G2(48,3 ;283,3) après arrondi au dixième donc l'équation réduite de la droite (G1G2) après arrondis y=6x-6,3
On peut vérifier que le point moyen de coordonnées $(\bar{x};\bar{y})$ appartient à la droite (G1G2)



Avec cette méthode, on peut estimer le prix d'un logement de 45 m² à 263720 €.

c) Une deuxième méthode : la droite qui minimise (qui rend minimum) la somme des écarts

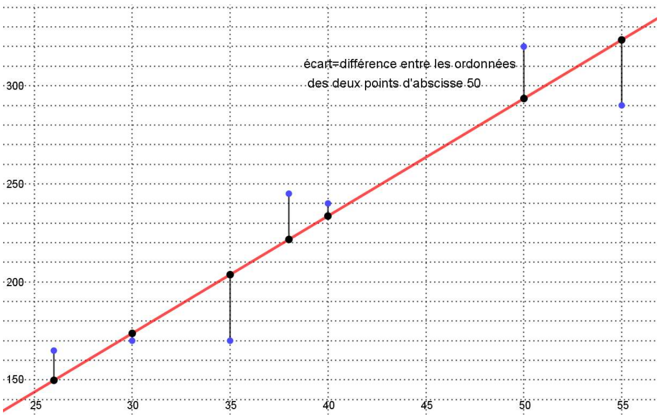
L'idée a été de chercher une droite qui minimise la somme des écarts entre les points obtenus par le relevé statistique et ceux de la droite (D) que l'on cherche (en rouge sur la figure). Un calcul naturel consisterait à prendre les valeurs absolues des différences entre les ordonnées des points du nuage et les ordonnées des points de même abscisse situés sur la droite (D). Mais cette méthode ne se prête pas facilement à des calculs algébriques simples.



d) Une troisième méthode : la méthode des moindres carrés

Comme on l'a vu ci-dessus, les écarts n'ont pas tous le même signe donc on va chercher à minimiser la somme des carrés des écarts (un peu comme pour la variance). On cherche donc deux réels a et b tels que la somme des carrés des écarts à la droite d'équation $y = ax + b$ soit la plus petite possible.

Reprenons l'exemple : on note $x_1, x_2, \dots x_7$ les superficies et $y_1, y_2, \dots y_7$ les prix correspondants.



Le premier écart entre les deux points d'abscisse 26 s'écrit $165-(26a+b)$ donc son carré est $[165-(26a+b)]^2 = 165^2 - 2\times 165\times(26a+b)+(26a+b)^2$
En développant encore, on obtient
$165^2 - 2\times 165\times 26a - 2\times 165b + 26^2a^2 + 2\times 26ab + b^2$
On fait la même chose pour les 7 écarts et on ajoute les résultats, ce qui donne après factorisation :

$$S = (165^2+\cdots+290^2) - 2\times(165\times 26+\cdots+290\times 55)a - 2\times(165+\cdots+290)b + (26^2+\cdots 55^2)a^2 + 2\times(26+\cdots+55)ab + 7b^2$$

Ou encore $S = 363050 - 2\times 66200a - 2\times 1600b + 11370a^2 + 2\times 274ab + 7b^2$

C'est une expression du second degré en a et en b dont on cherche le minimum.
Ce qui suit est justifié par la recherche d'extremums éventuels d'une fonction de deux variables (voir annexe 5). On procède en deux temps :

- On considère S comme un trinôme du second degré de la variable $\boldsymbol{a}$:

$S(a) = 11370\,\boldsymbol{a}^2 - 2\boldsymbol{a}(66200-274b) + 7b^2 - 132400b + 363050$

Or on sait que le minimum de $f(a) = \alpha a^2 + \beta a + \gamma$ est atteint pour $a = -\frac{\beta}{2\alpha}$ si $\alpha > 0$
Donc ici pour $a = \frac{66200-274}{11370}$ ce qui équivaut à $11370a + 274b - 66200 = 0$ (E)

- On considère S comme un trinôme du second degré de la variable $\boldsymbol{b}$:

$S(b) = 7\boldsymbol{b}^2 - \boldsymbol{2b}(1600-274a) + 274a^2 - 132400a + 363050$

On admet que S admet un seul minimum qui est atteint pour $b = \frac{1600-274a}{7}$

Il ne reste plus qu'à remplacer *b* dans (E) et on obtient

$11370a + 274\times\frac{1600-274a}{7} - 66200 = 0$ on résout cette équation du 1er degré en a et on trouve $a = \frac{25000}{4514} = \frac{12500}{2257}$. On en déduit $b = \frac{26600}{2257}$.

On admet ici que ces valeurs de a et b minimisent la somme S et que ce minimum est unique.
24

Mais le recours aux propriétés des fonctions à deux variables confirmerait le résultat (voir annexe 6).

La droite cherchée a donc pour équation $y = \frac{12500}{2257}x + \frac{26600}{2257}$ c'est-à-dire, en arrondissant les valeurs de a et b au centième, $y = 5{,}54x + 11{,}78$

Avec cette méthode, le prix d'un appartement de 45 m² devrait être voisin de 261,08 milliers d'euros soit 261080 €.

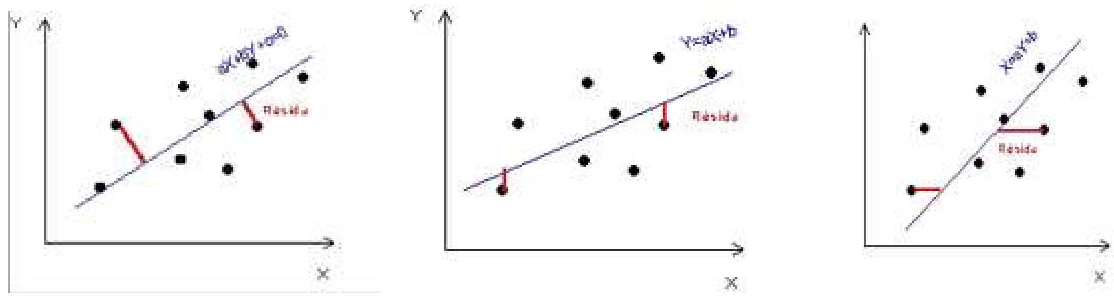
La méthode des moindres carrés a été initiée par deux grands savants, Adrien Marie Legendre (1752-1833) et Carl Friedrich Gauss (1777-1855). Ils ont travaillé indépendamment l'un de l'autre, ce qui a provoqué une controverse entre eux concernant la paternité de la méthode

(voir Annexe 6 et https://publimath.univ-irem.fr/numerisation/DI/IDI17004/IDI17004.pdf).

Plus précisément, il s'agit de minimiser la somme des carrés des d_i c'est-à-dire de minimiser $\sum(d_i)^2$. Cette méthode se prête facilement à un traitement algébrique et fournit une solution unique.

Remarque : On pourrait choisir des écarts parallèlement à l'axe des abscisses ou entre les points du nuage et leur projection orthogonale sur la droite cherchée, mais dans ce dernier cas, les calculs sont plus compliqués.

Trois façons de chercher à minimiser la somme des carrés des écarts



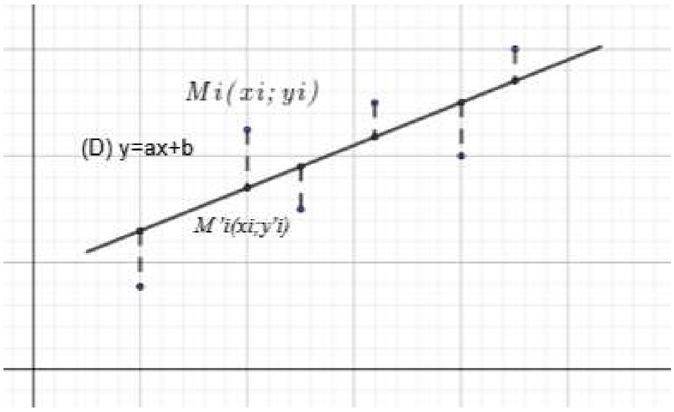
2) Equation de la droite de régression linéaire (ou d'ajustement affine) par la méthode des moindres carrés

a) Cas général

Comme on l'a vu précédemment, les calculs sont compliqués et fastidieux. On va donc établir une formule pour les simplifier et permettre d'utiliser la calculatrice ou le tableur, qui fournissent directement les résultats.

On considère un relevé de deux variables statistiques X et Y de même effectif n. On a donc n couples $(x_i; y_i)$ qui sont les coordonnées des n points M_i nuage de points qui représente cette série de deux variables. X est ici la variable explicative et Y la variable expliquée (dans l'exemple d'introduction la variable explicative est la superficie et la variable expliquée est le prix).

On note (D) la droite d'ajustement par la méthode des moindres carrés définie comme ci-contre, d'équation $y = ax + b$.
Soit $M_i'(x_i; y_i')$ le point de (D) de même abscisse que M_i.
Il s'agit donc de minimiser $S = \sum_1^n (y_i - y_i')^2$
Or $y_i' = ax_i + b$, donc $S = \sum_1^n [y_i - (ax_i + b)]^2$
Les valeurs de a et b cherchées correspondent aux valeurs pour lesquelles S est minimum.



Propriété :

L'équation réduite de la droite d'ajustement affine (ou de régression) est

$$y = a(x - \bar{x}) + \bar{y} \text{ où } a = \frac{cov(X,Y)}{V(X)}$$

Remarques :

- Cette droite passe par le point moyen $G(\bar{x}; \bar{y})$ du nuage de points
- Le coefficient *a* s'exprime aussi à l'aide du coefficient de corrélation : $a = r \times \frac{\sigma_Y}{\sigma_X}$.

En effet $r \times \frac{\sigma_Y}{\sigma_X} = \frac{Cov(X,Y)}{\sigma_X \times \sigma_Y} \times \frac{\sigma_Y}{\sigma_X} = \frac{Cov(X,Y)}{\sigma_X^2} = \frac{cov(X,Y)}{V(X)} = a$

Démonstration en quatre étapes

Etape 1 : On explicite l'écriture de S en développant $S = \sum_1^n [y_i - (ax_i + b)]^2$
On obtient $S = \sum_1^n [y_i^2 - 2y_i(ax_i + b) + (ax_i + b)^2]$
ou encore $S = \sum_1^n y_i^2 - 2a \sum_1^n y_i x_i - 2b \sum_1^n y_i + a^2 \sum_1^n x_i^2 + 2ab \sum_1^n x_i + nb^2$

Les sommes $\sum_1^n y_i^2$; $\sum_1^n y_i x_i$; $\sum_1^n y_i$; $\sum_1^n x_i^2$; $\sum_1^n x_i$ sont des nombres que l'on peut calculer puisque l'on connait les n valeurs de x_i et de y_i. Les inconnues à déterminer sont *a* et *b*.

Dans la suite, pour simplifier l'écriture on n'indique plus les bornes « de 1 à n » de ces sommes.

Etape 2 : Pour minimiser S on peut regarder cette somme comme un trinôme du second degré de la variable *a* :

$S = a^2 \sum x_i^2 - 2a(\sum x_i y_i - b \sum x_i) - 2b \sum y_i + nb^2 + \sum y_i^2$

On sait qu'un trinôme du second degré de la forme $\alpha x^2 + \beta x + \gamma$, où $\alpha > 0$, atteint son minimum quand $x = -\frac{\alpha}{2\beta}$, d'où, en revenant à la variable a, S est minimum pour

$$a = \frac{2(\sum x_i y_i - b\sum x_i)}{2\sum x_i^2} = \frac{(\sum x_i y_i - b\sum x_i)}{\sum x_i^2}$$ ou encore $a\sum x_i^2 \; - \sum x_i y_i + b\sum x_i = 0$ **(1)**

Etape 3 : On peut maintenant regarder S comme un trinôme du second degré de variable b et le minimiser (en coefficient de b^2 est positif)

$S = nb^2 - 2b(\sum y_i - a\sum x_i) + a^2\sum x_i^2 \; - 2a\sum x_i y_i + \sum y_i^2$ qui est minimum quand

$$b = \frac{2(\sum y_i - a\sum x_i)}{2n} = \frac{(\sum y_i - a\sum x_i)}{n} = \frac{\sum y_i}{n} - a\frac{\sum x_i}{n} = \bar{y} - a\bar{x}$$

$b = \bar{y} - a\bar{x}$ **(2)** en notant $\bar{x}$ et $\bar{y}$ les moyennes des variables X et Y.

Etape 4 : On résout le système d'inconnues a et b formé par les égalités (1) et (2). On reprend alors l'expression (1) où l'on remplace *b* par son expression en fonction de *a*

On trouve $a\sum x_i^2 \; - \sum x_i y_i + (\bar{y} - a\bar{x})\sum x_i = 0$ d'où l'on tire

$$a(\sum x_i^2 \; - \bar{x}\sum x_i) = \sum x_i y_i - \bar{y}\sum x_i.$$

Or $\sum x_i = n\bar{x}$ donc $a = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum x_i^2 - n\bar{x}^2}$

Au numérateur de *a* on reconnait la covariance de X et Y et au dénominateur la variance de X dans leurs formes développées (voir le paragraphe II).

Finalement $a = \frac{cov(X,Y)}{V(X)}$.

On admet que les valeurs de a et b trouvées correspondent au seul minimum de la somme S, ce qui pourrait être démontré à l'aide des propriétés des fonctions à deux variables (voir annexe 6)

La droite de régression par la méthode des moindres carrés a pour équation :

$$\boldsymbol{y = a(x - \bar{x}) + \bar{y}} \;\; \textbf{où } \boldsymbol{a = \frac{cov(X,Y)}{V(X)}}$$

On peut vérifier que cette droite passe par le point moyen $G(\bar{x};\bar{y})$ du nuage de points.

Remarque :
Dans une régression on cherche à expliquer la variabilité d'une variable, disons y, par une autre variable, disons x, et on cherche une équation de la forme $y = ax + b$ qui passe au mieux près du nuage de points.
Même s'il y a des liens mathématiques entre les deux, établir une corrélation ou une régression ne revient pas du tout à faire la même chose.
Pour s'en convaincre, si on calcule la corrélation de y avec x et de x avec y on obtient le même résultat (cela résulte de sa définition, $r = \frac{Cov(X;Y)}{\sigma_X \sigma_Y}$).
En revanche si on cherche la régression affine de y en fonction de x (ou de y en x) et celle de x en fonction de y alors les valeurs des coefficients directeurs des droites d'ajustement sont différents.

b) En pratique (à la main, avec un tableur ou une calculatrice)

Trouver la droite de régression par la méthode des moindres carrés : différents outils.

Exemple

Au cours des épreuves de physique et chimie d'un concours, 10 candidats ont obtenu les notes suivantes :

On se demande si une relation existe entre ces deux séries de notes.

Candidat n°	1	2	3	4	5	6	7	8	9	10
Note de Physique : x	8	4	9	14	10	7	12	15	13	6
Note de chimie : y	10	5	8,5	13	12	10	15	10	11	9

<u>i) Faire les calculs à la main</u>

Equation de la droite d'ajustement de y en x

$$y = a(x - \bar{x}) + \bar{y} \text{ où } a = \frac{cov(X,Y)}{V(x)} = \frac{\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n(\bar{x})^2}$$

On établit le tableau suivant

x_i	x_i^2	y_i	y_i^2	$x_i y_i$
8	64	10	100	80
4	16	5	25	20
9	81	8,5	72,25	76,5
14	196	13	169	182
10	100	12	144	120
7	49	10	100	70
12	144	15	225	180
15	225	10	100	150
13	169	11	121	143
6	36	9	81	54
$\sum x_i$	$\sum x_i^2$	$\sum y_i$	$\sum y_i^2$	$\sum x_i y_i$
98	1080	103,5	1137,25	1075,5

$\bar{x}$ = 9,8 ; $\bar{y}$= 10,35 ; $a = \frac{1075{,}5-1014{,}3}{1080-10\times 9{,}8^2} = \frac{61{,}2}{119{,}6} \approx 0{,}5117$

L'équation réduite de la droite d'ajustement de y en x est $y = 0{,}5117(x - 9{,}8) + 10{,}35$ ou encore $y = 0{,}5117x + 5{,}3353$

Equation de la droite d'ajustement de x en y :

Il faut pour cela refaire le même tableau en appelant x' les notes de chimie et y' les notes de physique.

En revanche si on échangeait seulement x et y dans l'équation précédente on ne minimiserait pas les distances verticales entre les points du nuage et la droite de régression mais les distances horizontales.

x'_i	${x'_i}^2$	y'_i	${y'_i}^2$	$x'_i y'_i$
10	100	8	64	80
5	25	4	16	20
8,5	72,25	9	81	76,5
13	169	14	196	182
12	144	10	100	120
10	100	7	49	70
15	225	12	144	180
10	100	15	225	150
11	121	13	169	143
9	81	6	36	54
$\sum x'_i$	$\sum {x'_i}^2$	$\sum y'_i$	$\sum {y'_i}^2$	$\sum x'_i y'_i$
103,5	1137,25	98	1080	1075,5

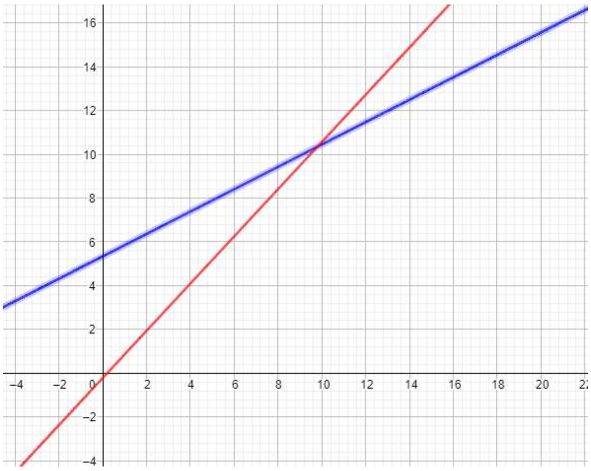
Comme précédemment $y' = a'(x' - \overline{x'}) + \overline{y'}$ où $a' = \frac{cov(X',Y')}{V(x')} = \frac{\sum x'_i y'_i - n\overline{x'}\overline{y'}}{\sum x'^2_i - n(\overline{x'})^2}$

$$a' = \frac{1075{,}5-101\ \ {,}3}{1137{,}25-1\ \ \times 10{,}35^2} = \frac{61{,}2}{66{,}025} \approx 0{,}9269$$

$y' = 0{,}9269(x' - 10{,}35) + 9{,}8$ d'où $y' = 0{,}9269x' + 0{,}2066$

Pour représenter les deux droites dans le même repère il faut exprimer x’ en fonction de y’

$x' = \frac{1}{0,9269}\,y' - \frac{0,2066}{0,9269}$ ou encore, puisque $x' = y$ et que $y' = x$, on a $y = 1,079\,x - 0,2229$

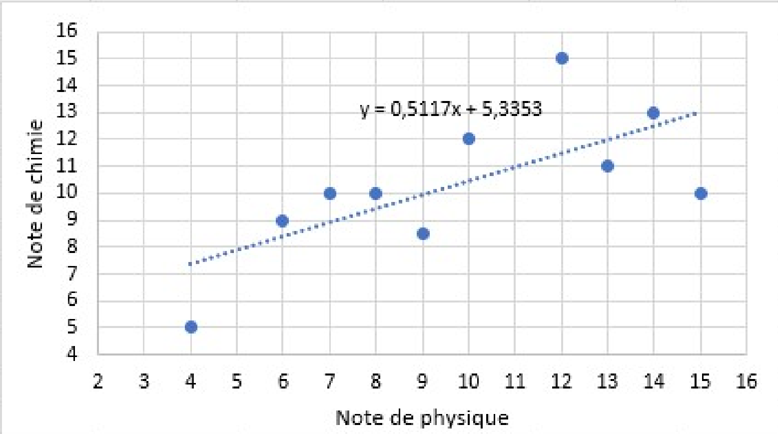


On remarque que les deux droites passent par le point moyen G(9,8 ; 10,35).

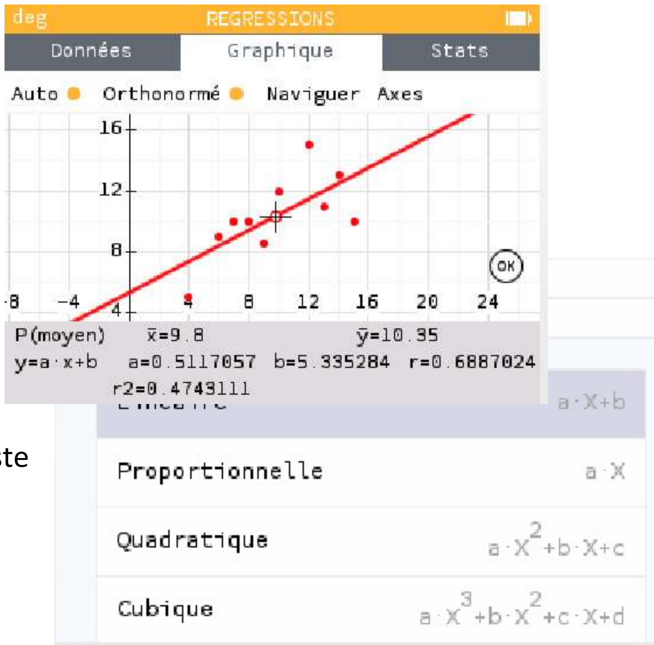
ii) Utiliser un tableur.

Il suffit de saisir les valeurs de x_i et y_i pour obtenir le nuage de points, la droite d’ajustement affine et son équation réduite, comme le montre la copie d’écran ci-dessous :

	A	B
1	Note de Physique	Note de chimie
2	8	10
3	4	5
4	9	8,5
5	14	13
6	10	12
7	7	10
8	12	15
9	15	10
10	13	11
11	6	9
12		



iii) Utiliser une calculatrice (ici, la calculatrice NUMWORKS)
Avec le programme **« Régressions »** on entre le tableau de données, puis on peut demander le nuage de points et la régression linéaire (on devrait plutôt dire affine).



On peut aussi demander un ajustement suivant différentes courbes (en position Graphique, appuyer plusieurs fois sur EXE). La liste affichée ci-contre n'est pas exhaustive, on peut demander une régression logarithmique, exponentielle, polynômiale, etc.)

On peut demander des valeurs obtenues à partir de la courbe d'ajustement. Ainsi,
« Prédiction sachant X » indique la valeur de Y correspondant à une valeur de X que l'on s'est donnée.

Remarque : si, à l'aide d'une calculatrice ou d'un tableur, on demande les résultats statistiques intermédiaires (moyennes, écarts-type, etc.), il faut vérifier que les écarts-types indiqués sont ceux de la série statistique donnée et non ceux d'une population dont la série statistique serait considérée comme un échantillon.

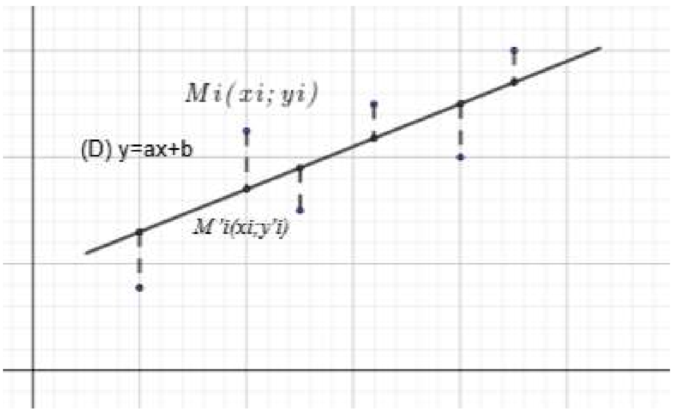
3) Coefficient de détermination

a) Exemple pour comprendre

On reprend l'exemple des logements : la superficie est la variable explicative X et le prix des logements, la variable expliquée Y. On a vu qu'un ajustement linéaire est raisonnable et que l'on peut expliquer en partie le prix par la superficie du logement. Mais on peut imaginer que d'autres éléments entrent en jeu dans le prix d'un logement (situation, environnement, état du logement, etc.) On peut se demander dans quelle proportion la superficie du logement intervient. Le coefficient de détermination, noté R^2, est un indicateur de cette proportion. Il est exprimé en pourcentage.
Dans le cas d'une régression linéaire simple, comme celles que nous étudions dans ce thème, R^2 est le carré du coefficient de corrélation r.
Dans notre exemple, on obtient directement avec une calculatrice ou avec un tableur, $r = 0{,}9187$ et $R^2 = 0{,}8439$, ce qui signifie que 84,39 % du prix est expliqué par la superficie. Bien sûr, il faudrait avoir plus de données, savoir si les statistiques dont on a disposé ont été relevées dans le même quartier, etc.

b) Introduction et définition

La variation de la variable Y est obtenue en considérant les différences entre les valeurs observées y_i et leur moyenne $\bar{y}$.



$y'_i - \overline{y'}$ est la variation expliquée par le modèle (ajustement affine), tandis que $y_i - \bar{y}$ est la variation non expliquée par le modèle.

(y'_i est l'ordonnée du point de la droite d'ajustement d'abscisse x_i)
On a : $\overline{y'} = \bar{y}$.
En effet $\overline{y'} = \frac{1}{n}\sum_1^n y'_i = \frac{1}{n}\sum_1^n (ax_i + b) = \frac{a}{n}[(\sum_1^n x_i) + nb] = \frac{a}{n} \times (n\bar{x} + nb) = a\bar{x} + b$ et en remplaçant b par sa valeur $\bar{y} - a\bar{x}$ on a bien $\overline{y'} = \bar{y}$
On a : $y_i - \bar{y} = y'_i - \bar{y} + y_i - y'_i$ d'où $y_i - \bar{y} = y'_i - \boldsymbol{\overline{y'}} + y_i - y'_i$

On peut établir la formule de décomposition de la variance de Y. On note

- SCT (Somme des Carrés Totale) traduit la variation totale de Y.
- SCE (Somme des Carrés Expliquée) traduit la variation expliquée par le modèle.
- SCR (Somme des Carrés Résiduelle) traduit la variation inexpliquée par le modèle.

SCT	SCE	SCR
$\sum_{i=1}^{n}(y_i - \bar{y})^2$	$\sum_{i=1}^{n}(y'_i - \bar{y})^2$	$\sum_{i=1}^{n}(y_i - y'_i)^2$

On appelle coefficient de détermination la quantité suivante : $R^2 = \frac{SCE}{SCT}$ qui exprime la part de la variation expliquée par le modèle par rapport à la variation totale, ou encore, en divisant chaque terme par l'effectif des points du nuage de points, $R^2 = \frac{Var(Y')}{Var(Y)}$.

On peut démontrer dans le cas d'une régression affine simple que $R^2 = r^2$ où r est le coefficient de corrélation linéaire (voir en annexe 4).

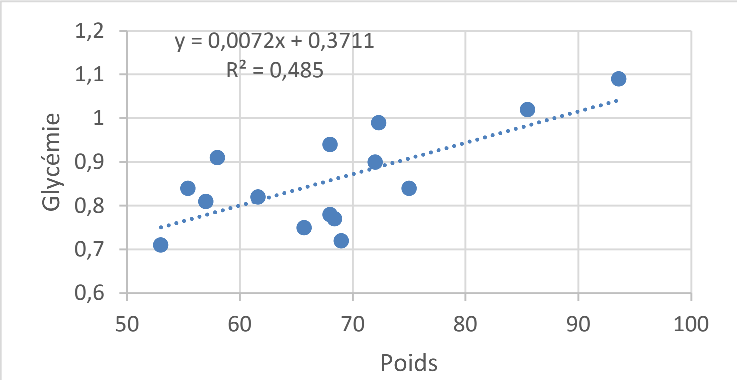
Ce coefficient est compris entre 0 et 1. En effet $0 \leq SCE \leq SCT$.

c) Exercice d'application

On cherche à déterminer quelle part du poids d'un individu influence son taux de glycémie. On a relevé pour 15 individus, leur poids (ou masse) en kg et leur taux de glycémie en g/l

Individus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Poids	68	55,4	69	61,6	65,7	68,4	75	53	85,5	93,6	72,3	68	57	58	72
Glycémie	0,94	0,84	0,72	0,82	0,75	0,77	0,84	0,71	1,02	1,09	0,99	0,78	0,81	0,91	0,9

A l'aide d'un tableur, on a représenté le nuage de points et obtenu l'équation réduite de la droite de régression ainsi que le coefficient de détermination. Il est égal à 0,485 ce qui signifie que seulement 48,5 % de la glycémie est expliquée par le poids de ces 15 personnes.



4) Régression avec un changement de variable

Lorsqu'un ajustement affine n'est pas adapté (le nuage de points ne fait pas penser à une droite), on peut procéder à un changement de variable : c'est le cas lorsque le nuage de points évoque une courbe connue, exponentielle ou logarithmique ou autre. On présente un exemple ci-dessous ; d'autres cas seront étudiés en exercices dans les parties IV et V. Le changement de variable à réaliser sera toujours indiqué dans l'énoncé.

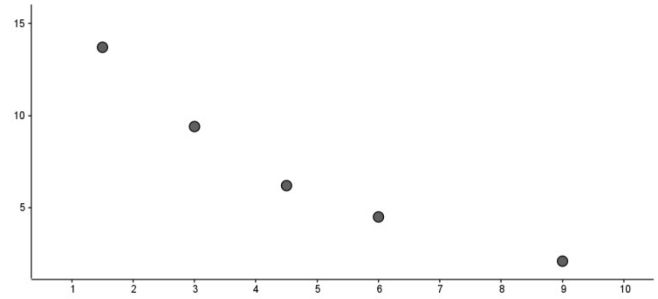
Un exemple pour comprendre

On injecte à l'instant $t = 0$ une substance médicamenteuse dans le sang d'un patient et on mesure pour différentes valeurs de t (en heures) la concentration C (en $\mu g/L$). On veut savoir comment évolue cette concentration et au bout de combien de temps elle sera inférieure à 1 $\mu g/L$ (il faudra alors faire une nouvelle injection).

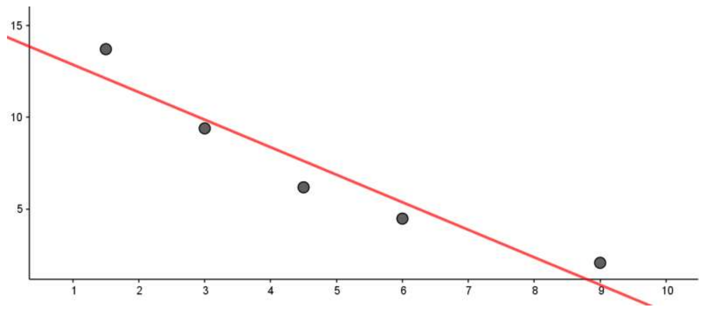
On a obtenu les résultats suivants

Temps (en h)	1,5	3	4,5	6	9
Concentration	13,7	9,4	6,2	4,5	2,1

puis à l'aide d'un tableur, on a représenté le nuage de points ci-contre.
On a calculé le coefficient de corrélation linéaire de C en t : il est égal à environ $-0{,}956$ (arrondi au millième)



Le nuage de points et le coefficient de corrélation linéaire, proche de 1, suggèrent un ajustement affine. En utilisant le tableur, on a $y = -1{,}49t + 14{,}35$ qui est l'équation de la droite d'ajustement ou de régression de y en x (en rouge).



En utilisant ce modèle, on trouve que la concentration sera inférieure à 1 $\mu g/L$ au bout de 9h environ : on a résolu l'inéquation

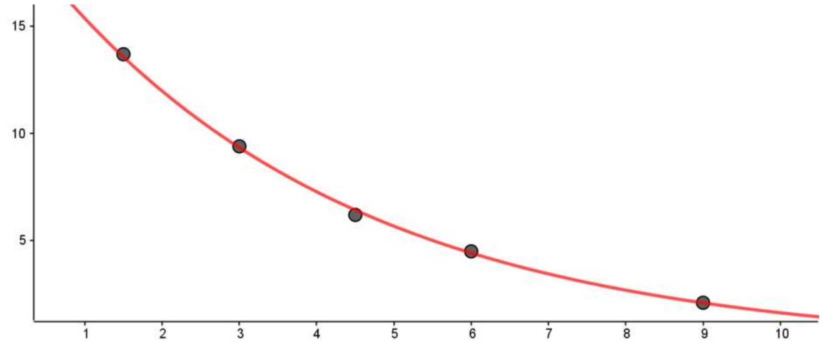
$-1{,}49t + 14{,}35 < 1$, ce qui donne $t > 8{,}96$.

On constate que cet ajustement n'est pas satisfaisant : il ne permet pas de répondre à la question (au bout de 9h, la concentration n'est pas inférieure à 1)

On va donc chercher une courbe qui approche plus précisément les points du nuage. Pour cela, deux possibilités :

1) Le tableur propose d'autres modèles d'ajustement ; parmi eux le modèle exponentiel semble bien adapté : la courbe en rouge proposée par le tableur a pour équation

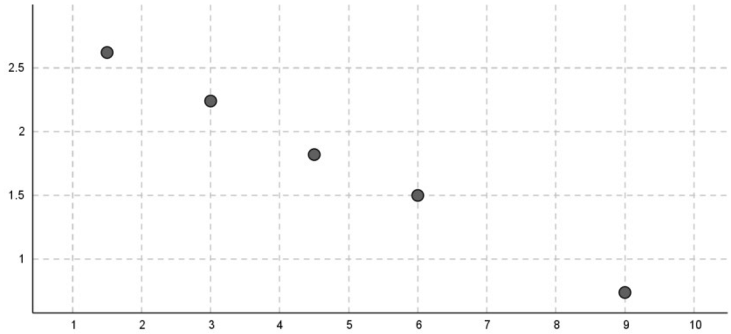
$y = 19{,}72e^{-0{,}25t}$ (les coefficients sont arrondis au centième).



Avec ce modèle, on cherche quand la concentration sera inférieure à 1 en résolvant l'inéquation $19{,}72e^{-0{,}25t} < 1$ ce qui donne $t > 11{,}9$ environ, soit 11h et 54 minutes.

2) On peut également effectuer un changement de variable. On pose z=ln(C(t)) puisque l'allure du nuage suggère un modèle exponentiel, on calcule les valeurs de z= ln(C(t)) et on représente le nouveau nuage de points comme ci-dessous :

Ce nouveau nuage montre qu'un ajustement affine sera cette fois bien adapté, ce qui est confirmé par la valeur du coefficient de corrélation linéaire de z en t, très voisin de -1 (il vaut environ -0,9996).



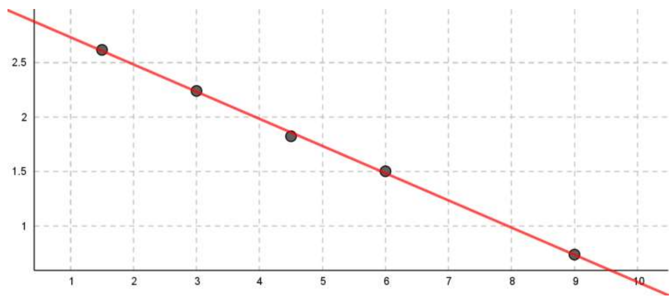
La droite de régression proposée par le tableur a pour équation $z = -0{,}25t + 2{,}98$.

D'où $ln(C(t)) = -0{,}25t + 2{,}98$

On en déduit le modèle approximatif :

$C(t) = e^{-0{,}25} \times e^{2{,}98} = 19{,}69e^{-0{,}25t}$

Il correspond à peu près au modèle trouvé précédemment (pas exactement à cause des arrondis).



On peut répondre à notre question en résolvant, comme ci-dessus l'inéquation $C(t) < 1$; on obtient quasiment le même résultat (11h55min)

IV – Retour au thème : quelques problèmes possibles

1) Évolution de la température et des émissions de gaz à effet de serre dans le cadre du réchauffement climatique.

Emission de CO2 et réchauffement climatique : corrélation/causalité ?

a) Comprendre le problème

En 1988, l'ONU forme le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) pour synthétiser les études scientifiques sur le climat. Dans son quatrième rapport datant de 2007, auquel ont participé plus de 2 500 scientifiques de 130 pays, le GIEC affirme que le réchauffement climatique depuis 1950 est « très probablement » dû à l'augmentation des gaz à effet de serre d'origine anthropique (liés aux activités humaines). Les conclusions du GIEC ont été approuvées par plus de quarante sociétés scientifiques et académies des sciences, y compris l'ensemble des académies nationales des sciences des grands pays industrialisés. Le degré de certitude est passé à « extrêmement probable » dans le cinquième rapport de 2014.

Plusieurs scénarii ont été étudiés par le GIEC (Groupe d'experts Intergouvernemental sur l'Evolution du Climat) et tous vont dans le même sens :

- réchauffement global de 1,4 à 5,8 °C entre 1990 et 2100,
- élévation du niveau de la mer (dilatation thermique des océans) de 9 à 88 cm entre 1990 et 2100
- fonte des glaciers de montagne, fragilité des pôles Nord et Sud, bouleversement du cycle de l'eau, dérèglement des saisons, extinction de certaines espèces, …
- accentuation de caractéristiques climatiques (précipitation, sécheresse, …), dérèglement brutal et imprévisible des variations climatiques naturelles, …
- augmentation des maladies transmissibles par les parasites (paludisme, fièvre jaune…) en raison d'une progression des zones climatiques favorables à leur reproduction, …

Des incertitudes subsistent sur l'ampleur et la géographie du réchauffement futur, du fait de la précision des modèles, de l'imprévisibilité du volcanisme, mais aussi des comportements étatiques et individuels (présents et futurs) variables. Les enjeux socioéconomiques, politiques, sanitaires, environnementaux, voire géopolitiques ou moraux, étant majeurs, ils suscitent des débats nombreux, à l'échelle internationale, ainsi que des controverses. Néanmoins, depuis 2000, un consensus émerge sur le fait que les effets du réchauffement se font déjà sentir de manière significative, qu'ils devraient s'accroître à moyen et long terme et qu'ils seraient irréversibles sauf actions concertées, locales aussi bien que planétaires.

L'effet de serre

La température moyenne à la surface du globe est de 15°C. Sans l'effet de serre cette température moyenne serait de - 18 °C et la vie serait impossible.

Notre atmosphère est principalement composée d'azote et d'oxygène : des gaz qui laissent passer les rayonnements, visibles et infrarouges, du soleil. L'énergie solaire est absorbée par la surface de la terre, convertie en chaleur et une partie est réémise sous forme de rayons infrarouges.
Dans notre atmosphère, sont aussi naturellement présents ce qu'on appelle des gaz à effet de serre (vapeur d'eau, gaz carbonique, méthane, protoxyde d'azote, …) en quantité très réduite (moins de 1 %). Ces derniers jouent un rôle déterminant dans l'équilibre de l'atmosphère. Ils absorbent une partie des rayonnements infrarouges puis les réémettent, ce qui permet de réchauffer la basse atmosphère et la surface de la terre.

L'effet de serre est donc un phénomène naturel comparable à celui que produit la vitre d'une serre. Les gaz à effet de serre présents dans l'atmosphère, jouent le rôle de cette vitre qui piège la chaleur du soleil et l'emprisonne. Mais l'équilibre de ce phénomène naturel, complexe et variable, est fragile. Les activités humaines produisent aujourd'hui d'importantes quantités de gaz à effet de serre (dioxyde de carbone, méthane, gaz fluorés…) et amplifient le phénomène naturel d'effet de serre avec pour principale conséquence... : le réchauffement climatique. Outre l'industrie (fabrication de ciment, d'aluminium, de composants d'ordinateurs, procédés chimiques, …), ou l'agriculture (élevages de ruminants, culture du riz, fabrication d'engrais, ….), la combustion des énergies fossiles, l'utilisation de climatiseurs, … produisent des gaz à effet de serre.

Schéma du Giec

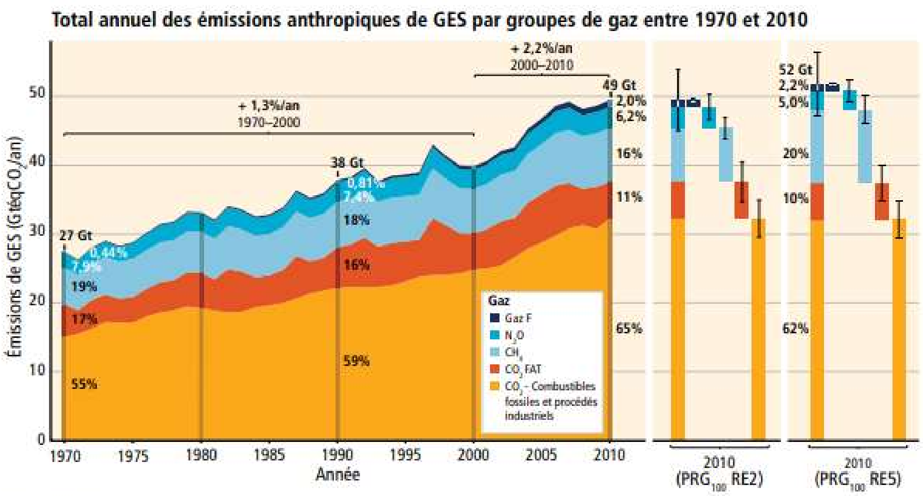


Figure RID.2 | Total annuel des émissions anthropiques de gaz à effet de serre (GES) (gigatonne d'équivalent CO_2 par an, GtéqCO_2/an), pour la période 1970–2010 et par gaz: CO_2 issu de l'usage des combustibles fossiles et des procédés industriels; CO_2 issu de la foresterie et d'autres affectations des terres (FAT); méthane (CH_4); oxyde nitreux (N_2O); gaz fluorés (gaz F) réglementés en vertu du protocole de Kyoto. À droite sont représentées côte à côte les émissions de 2010 en équivalent CO_2, suivant des pondérations établies à partir d'une part des valeurs du deuxième Rapport d'évaluation du GIEC (RE2) et d'autre part des valeurs du cinquième Rapport d'évaluation du GIEC (RE5). Sauf indication contraire, les émissions en équivalent CO_2 dont il est question dans le *Rapport de synthèse* regroupent tous les gaz réglementés en vertu du protocole de Kyoto (CO_2, CH_4, N_2O et gaz F); elles sont établies à partir des valeurs du potentiel de réchauffement global calculées à un horizon de cent ans (PRG_{100}) tirées du RE2 (voir le glossaire). En se servant des valeurs du PRG_{100} les plus récentes, à savoir celles du RE5 (boîtes à moustaches à droite), on obtient un total annuel des émissions de GES plus élevé (52 GtéqCO_2/an) qu'il faut attribuer à une contribution accrue du méthane, ce qui ne modifie pas beaucoup la tendance à long terme. *[figure 1.6, encadré 3.2]*

Conséquences de l'effet de serre :

Depuis le début de l'ère industrielle, l'homme a rejeté dans l'atmosphère des gaz (gaz carbonique, méthane, oxydes d'azote, etc.) qui augmentent artificiellement l'effet de serre. Si cet ajout à l'effet de serre naturel est faible (environ +1 %), il est amplifié par la vapeur d'eau et a ainsi contribué à l'augmentation de la température moyenne de notre planète d'environ 0,5 °C observée dans la seconde moitié du vingtième siècle. A très long terme, des perturbations importantes pourront également intervenir dans les courants marins et les glaces polaires, avec des conséquences sur la répartition du réchauffement climatique selon les régions du globe.

ÉVOLUTION DU NIVEAU MOYEN DES MERS DU GLOBE DEPUIS 1993

Source : E.U. Copernicus Marine Service Information

Les dernières années ont donné quelques aperçus des risques que feraient courir le changement climatique au continent européen : même s'il n'est généralement pas possible d'attribuer tel ou tel événement météorologique extrême (tempête, inondation, vague de chaleur…) au dérèglement climatique, les faits observés matérialisent fidèlement les résultats du Groupe d'Experts Intergouvernemental sur l'Evolution du Climat (GIEC). Certains effets du dérèglement climatique sont d'ailleurs déjà visibles en France : élévation de 0,9 °C en un siècle de la température moyenne annuelle et retrait des glaciers.

Comme à l'échelle mondiale, l'évolution des températures moyennes annuelles en France métropolitaine témoigne d'un réchauffement net depuis 1900. Ce réchauffement a connu un rythme variable, avec une augmentation particulièrement marquée depuis les années 1980. La température moyenne annuelle de 13,4 °C en 2017 a dépassé la normale (référence 1961-1990) de 1,6 °C, plaçant l'année 2017 au cinquième rang des années les plus chaudes.

Compte tenu de la durée de vie très longue des gaz à effet de serre dans l'atmosphère, le phénomène de réchauffement climatique ne peut être stoppé à court terme. Pour exemple, une molécule de CO_2 émise aujourd'hui dans l'atmosphère alimentera le phénomène d'effet de serre et de réchauffement climatique pendant au moins 1 siècle. Il est donc urgent d'agir afin de stabiliser puis de diminuer durablement les émissions de gaz à effet de serre puisque l'avenir de la planète et des générations futures dépend des décisions qui sont prises aujourd'hui.

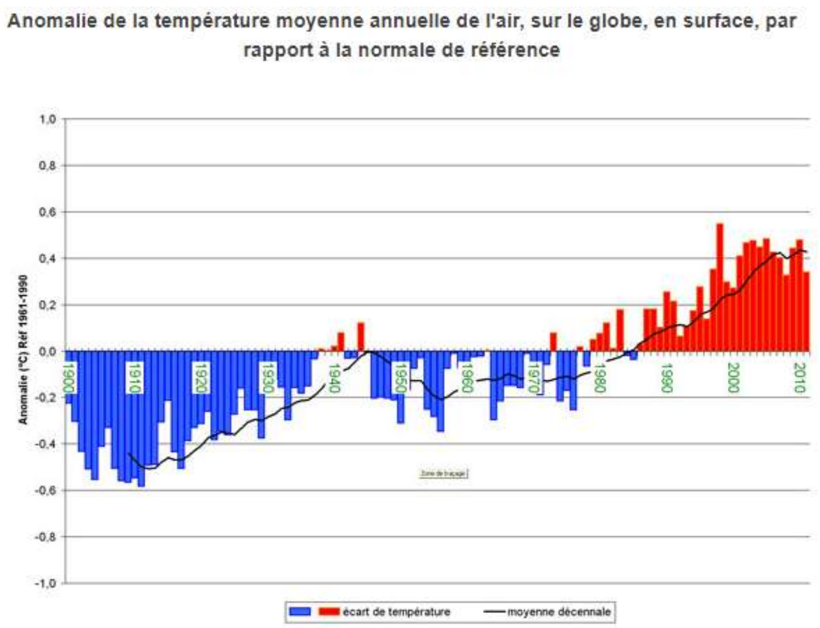
Quelques définitions

Équivalent CO_2 : méthode de mesure des émissions de gaz à effet de serre qui prend en compte le pouvoir de réchauffement de chaque gaz relativement à celui du CO_2 .

Gaz à effet de serre (GES) : les principaux gaz responsables de l'effet de serre, dont les émissions sont encadrées par le protocole de Kyoto, sont le dioxyde de carbone (CO_2), le méthane (CH_4), l'oxyde nitreux (N_2O) et les gaz fluorés (HFC, PFC et SF_6). Les émissions de ces gaz sont pondérées par leurs potentiels de réchauffement global (PRG) et exprimées en équivalents CO2 pour donner un total d'émissions en équivalents CO2 .

UTCATF : utilisation des terres, leur changement d'affectation et la forêt. C'est une catégorie utilisée dans les inventaires d'émissions de gaz à effet de serre qui couvre les émissions et les absorptions de ces gaz liées à l'utilisation des terres, leur changement d'affectation et à la forêt.

Quelques données pour justifier un lien (éventuel) entre émissions de gaz à effet de serre et réchauffement climatique



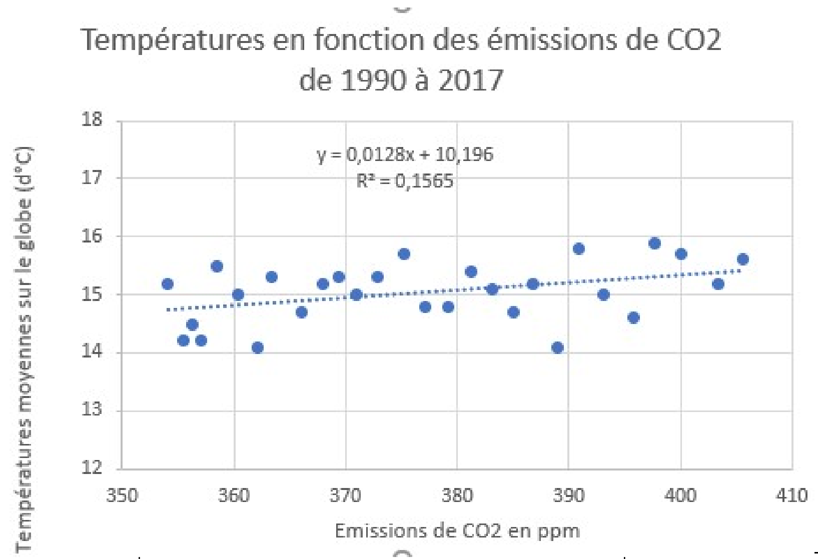
(Données du Climatic Research Unit, University of East Anglia. Le zéro correspond à la moyenne de l'indicateur sur la période 1961-1990, soit 14,0°C).

Le tableau ci-dessous donne les émissions de CO2 dans le monde au regard des températures moyennes sur le globe.

Il s'agit de mesurer la contribution éventuelle des émissions de CO_2 au réchauffement climatique.

La mesure des émissions de CO_2 est, dans ce tableau, exprimée en ppm (1 ppm correspond à un litre de CO2 dans 10^6 litres d'air).

Années	Emissions de CO_2 dans le monde : Mole fraction (ppm)	Températures moyennes sur le globe en °C
1990	354,1	15,2
1991	355,4	14,2
1992	356,2	14,5
1993	357	14,2
1994	358,5	15,5
1995	360,3	15
1996	362,1	14,1
1997	363,3	15,3
1998	366	14,7
1999	368	15,2
2000	369,4	15,3
2001	370,9	15
2002	372,9	15,3
2003	375,3	15,7
2004	377,1	14,8
2005	379,2	14,8
2006	381,3	15,4
2007	383,1	15,1
2008	385,1	14,7
2009	386,7	15,2
2010	388,9	14,1
2011	390,9	15,8
2012	393,1	15
2013	395,7	14,6



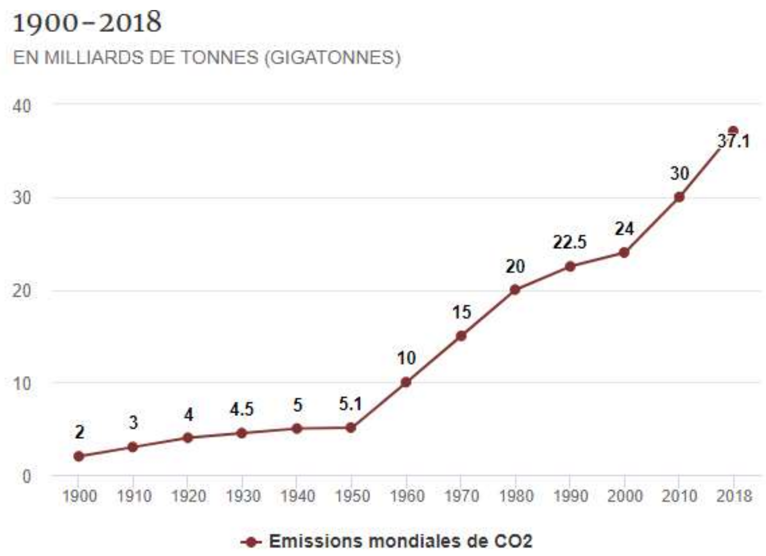
2014	397,7	15,9
2015	400	15,7
2016	403,3	15,2
2017	405,5	15,6

Le nuage de points associé à cette série statistique est présenté ci-dessous.
Le coefficient de corrélation linéaire est égal à 0,40

b) Exercice d'application

La courbe ci-dessous indique la masse de CO_2 en milliards tonnes pour la période 1900 - 2018 relevée tous les 10 ans.

1) A partir du graphique, quelles observations peut-on faire concernant la température moyenne du globe ?



2) Les températures moyennes sur le globe sont données dans le tableau ci-dessous

Années	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	2018
températures	14,3	13,6	14	14,4	13,1	14,2	14	13,9	13,4	15,2	15,3	14,1	16,1

Tracer le nuage de points associé à la série de deux variables (X ; Y) où X est la masse mondiale de CO_2 en milliards de tonnes et y la température moyenne sur le globe.

3) Est-il légitime de chercher la droite d'ajustement affine correspondant ?

4) Si oui donner son équation.

5) Si les émissions de CO2 sont de 45 Gigatonnes en 2026 quelle sera la température moyenne du globe ?

6) Quelle est la part des émissions de CO2 qui participe au réchauffement climatique ?

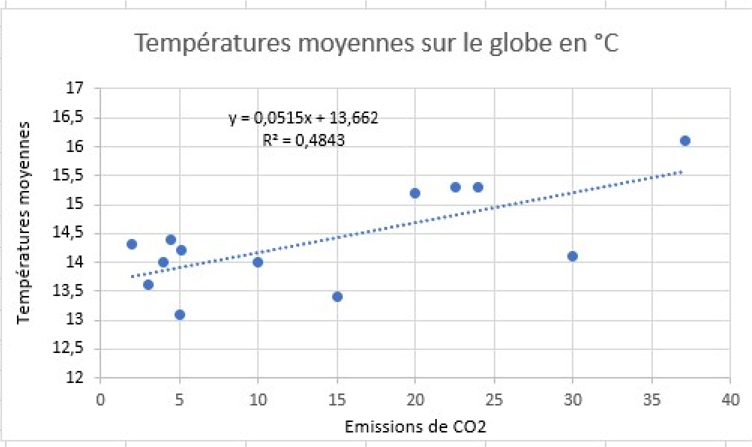
Indications

2)

Années	Emissions de CO_2 dans le monde en milliards de tonnes	Températures moyennes sur le globe en °C

1900	2	14,3
1910	3	13,6
1920	4	14
1930	4,5	14,4
1940	5	13,1
1950	5,1	14,2
1960	10	14
1970	15	13,4
1980	20	15,2
1990	22,5	15,3
2000	24	15,3
2010	30	14,1
2018	37,1	16,1

Nuage de points



Le nuage de points est allongé. Il paraît légitime de chercher l'équation de la droite d'ajustement affine de Y en X.

Coefficient de corrélation 0,70.
Equation de la droite d'ajustement affine : y = 0,052 x+13,656
Si les émissions de CO2 sont de 45 gigatonnes en 2026 la température moyenne du globe sera d'environ 16°C.
Coefficient de détermination 0,4843. Les émissions de CO2 participent pour 48% à l'augmentation de la température moyenne dans le globe.

2) Evolution d'une population : deux exercices.

Exercice 1 : A propos d'éléphants : recherche d'une corrélation.

On se demande s'il existe un lien entre le nombre d'éléphants en Afrique centrale et le nombre de braconnages.

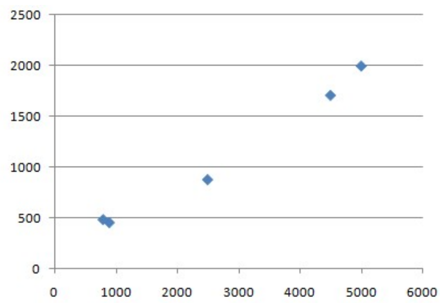
On dispose des données suivantes :

Population d'éléphants en Afrique Centrale

Années	Nombre X d’éléphants en Afrique centrale	Nombre Y de braconnages
2002	5000	2000
2004	4500	1710
2006	2500	875
2008	900	450
2010	800	480

1) Déterminer les pourcentages de braconnages pour chaque année considérée.
2) Représenter le nuage de points de coordonnées (x_i ; y_i).
3) Déterminer l’équation de la droite des moindres carrés de Y en X.
4) Calculer le coefficient de corrélation et conclure.

<u>Indications 2)</u>



3) Equation de la droite des moindres carrés de Y en X :

$y = \frac{cov(x,y)}{var(x)}(x - \bar{x}) + \bar{y}$; $cov(x,y)$= *4134300-2740×1103 =1112080* ; $var(x)$= *3082400*

4) $y = \frac{1112080}{3082400}(x - 2740) + 1103$; y *= 0,361*x $+114{,}45$

5) Coefficient de corrélation de Y et X : $\frac{cov(x,y)}{\sigma_x \ \sigma_y} = \frac{1112080}{\sqrt{3082400}\sqrt{407916}} \approx$*0,992*

Il existe une forte corrélation, cependant le braconnage n’est pas la seule cause du déclin de la population d’éléphants. (On peut invoquer aussi la déforestation par exemple.)

Exercice 2 : Population en extinction, un modèle exponentiel.

Une étude de l’évolution d’une population animale en extinction conduit à penser que le nombre d’individus N de cette population varie avec le temps suivant une loi du type : N(t) =

ae^{-kt} où a et k sont des constantes strictement positives. On veut déterminer expérimentalement la valeur de la constante k. Pour cela, on observe pendant 8 mois un échantillon composé initialement de 200 individus et on note à la fin de chaque mois le nombre de survivants. Les résultats sont les suivants :

t_i en mois	1	2	3	4	5	6	7	8
Survivants après t_i mois N_i	180	154	140	120	112	97	84	76

1) Déduire de ces résultats une valeur approchée de *k* au dixième près (t est exprimé en mois).

2) Calculer $\ln(N_i)$ pour chaque valeur de t_i puis représenter le nuage de points de coordonnées (t_i ; $\ln(N_{(i)})$)

3) Déterminer la droite de régression linéaire de ce nuage de points par la méthode des moindres carrés.

4) D'après ce modèle, au bout de combien de temps n'y aura-t-il plus aucun survivant ?

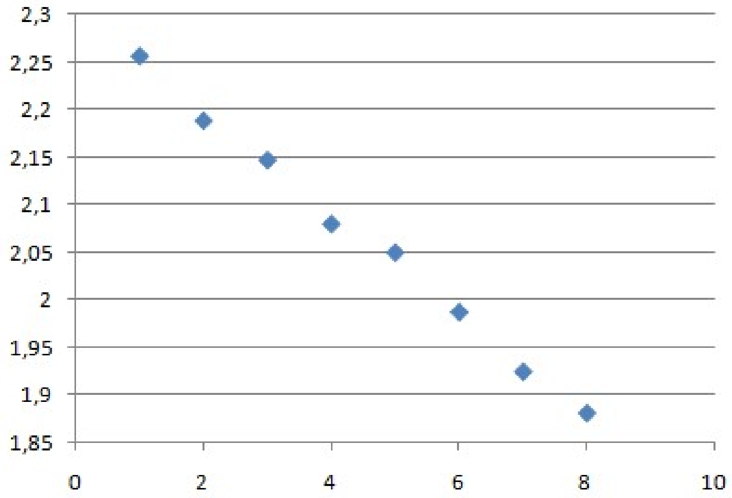
Indications

1) Pour t=0 on sait que ae^{-kt}*= 200. On en déduit que a=200.*

On a plusieurs possibilités pour déterminer une valeur approchée de k. Par exemple, pour t=1, $200e^{-k} = 180$ *d'où k= -* $\ln 0{,}9 \approx 0{,}105$ *et pour t=5,* $200e^{-5k}=112$

ce qui donne $k \approx \frac{\ln(0{,}56)}{-5}$ *ou* $k \approx 0{,}116$

t_i en mois	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Survivants après t_i mois N_i	*180*	*154*	*140*	*120*	*112*	*97*	*84*	*76*
$Ln(N_i) = y_i$ (valeurs arrondies au dix-millième)	*2,2553*	*2,1461*	*2,1461*	*2,0792*	*2,0492*	*1,9868*	*1,9243*	*1,8808*



2) Equation de la droite de régression $y=\frac{cov(t,y)}{var(t)}(t-\bar{t})+\bar{y}$

$\frac{cov(t,y)}{var(t)}=\frac{-0{,}2778}{5{,}25}\approx -0{,}053$; $y=-0{,}053\ (t-4{,}5)+2{,}0636$; $y=-0{,}053\,t+2{,}3021$

3) On cherche pour quelle valeur de t on a y *nulle. On trouve t* ≈ 43 *mois*

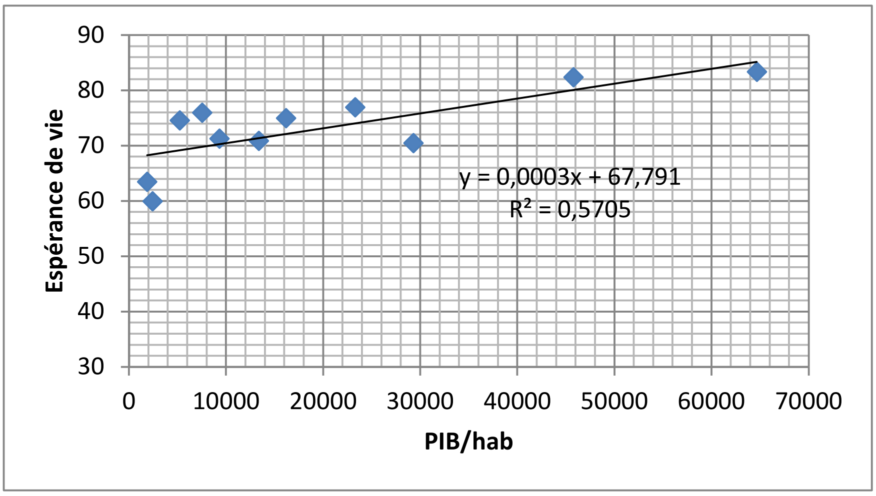
3) Corrélation et/ou causalité ? Deux exercices pour interpréter des données.

Exercice 1 : PIB et espérance de vie, existe-t-il une corrélation ?

Pays	Suisse	France	Russie	Uruguay	Brésil	Egypte	Ukraine	Vietnam	Honduras	Bénin	Haïti
PIB/hab	64649	45775	29267	23274	16154	13366	9283	7510	5212	2415	1864
Espérance de vie à la naissance	83,4	82,4	70,5	77	75	70,9	71,3	76	74,6	60	63,5

Le tableau ci-dessus indique pour dix pays du monde, le PIB par habitant (en dollars internationaux, année 2018) et l'espérance de vie à la naissance (données de 2015).

1) On donne ci-dessous un ajustement linéaire fourni par le tableur : peut-on dire qu'il existe une corrélation entre ces deux grandeurs ? Argumenter.



2) En utilisant cet ajustement, calculer l'espérance de vie à la naissance au Qatar dont le PIB par habitant est égal à 130475$. Commenter le résultat.

<u>Indications</u>

1) On constate que les points sont relativement proches de la droite (sauf pour le Bénin); le coefficient de corrélation linéaire est égal à 0,755, ce qui signifie qu'il ne s'agit pas d'une très forte corrélation. Il semble logique de penser que plus le pib/hab est élevé, plus le pays est riche donc plus le système de santé permet d'avoir une espérance de vie élevée. Cependant il existe souvent de très grandes disparités dans la population et dans le système de santé qui ne profite pas toujours à tous…

2) On trouve presque 107 ans, ce qui est impossible. De très grandes richesses pour une toute petite partie de la population donnent un PIB élevé mais ne permettent pas d'élever l'espérance de vie (qui de toute façon est limitée)

Exercice 2 : Etudier une corrélation, rechercher une causalité.

Dans le tableau ci-dessous, X_i désigne le taux d'alphabétisation des femmes de plus de 15 ans (en pourcentage) et Y_i désigne le taux de mortalité infantile (pour mille nouveau-nés) dans 10 pays du monde en 2016.

1) Représenter sur calculatrice ou tableur le nuage de points associé à cette série statistique double.

Dans les questions suivantes, le détail des calculs n'est pas demandé. Les résultats seront arrondis au centième.

2) a) Déterminer le coefficient de corrélation linéaire de la série (X_i, Y_i) pour i variant de 1 à 10 et donner l'équation de la droite D de régression linéaire de Y en X.

b) Si on applique le modèle précédent à un pays où le taux d'alphabétisation des femmes est de 77%, quel taux de mortalité infantile le calcul donne-t-il ?

3) L'un des points du nuage est particulièrement éloigné de la droite D : quel est le numéro du pays concerné ?

4) Reprendre la question 2) en éliminant les données relatives au pays repéré à la question 3. Que peut-on constater ?

5) Peut-on dire que l'analphabétisme des femmes est la cause de la mortalité infantile ? Expliquer.

<u>Indications</u>

2) a) $r \approx -0{,}95$ *et D a pour équation* $y = -1{,}01x + 103{,}58$

Numéro du pays (i)	1	2	3	4	5	6	7	8	9	10
X_i	11	24,2	27,3	32,8	41,1	55,1	65,5	70,5	85,6	99,7
Y_i	91,3	70,4	97,6	67,4	58,4	34,5	44,5	30,6	12,90	7,70

b) on remplace x par 77 dans l'équation ci-dessus et on trouve $y \approx 25{,}81$

3) le pays 3

4) a) Si on élimine le pays 3, $r \approx -0{,}98$. *Il est plus voisin de -1 que dans la question 2 : l'ajustement linéaire sera de meilleure qualité. D a pour équation* $y = -0{,}93x + 96{,}60$

b) pour x=77, on obtient $y \approx 24{,}99$

La mortalité infantile calculée est inférieure à celle trouvée au 2. Ce modèle donne un résultat plus fiable.

5) Commentaire d'après http://grasland.script.univ-paris-diderot.fr/STAT98/stat98_6/stat98_6.htm

Même si on observe une corrélation assez importante entre l'analphabétisme des femmes et la mortalité infantile, on ne peut pas dire que l'analphabétisme des femmes est la cause de la mortalité infantile. Autrement dit, il n'y a pas de causalité directe.

En effet, ce n'est pas parce-que les femmes ne savent pas lire qu'elles ne peuvent pas soigner leur enfant. Et ce n'est pas non plus parce-que les enfants meurent en bas-âge que les femmes ne savent pas lire ! La découverte d'une corrélation ne permet pas de conclure à l'existence d'une relation de cause à effet mais montre qu'on a affaire à un « système causal » au sein duquel sont placés les deux caractères étudiés. L'analphabétisme des femmes peut n'être qu'un facteur parmi d'autres (malnutrition, climat, sous médicalisation…etc.) qui influencent la mortalité infantile.

4) Loi de désintégration radioactive : deux exercices avec changement de variable.

Exercice 1 : Choisir un bon ajustement (d'après bac)

Une société souhaite exploiter un nouveau détecteur qui permet de mesurer la désintégration de noyaux radioactifs. Pour tester ce détecteur, la société l'utilise pour déterminer le nombre de noyaux radioactifs présents dans un échantillon radioactif à des instants donnés. Voici les résultats des relevés réalisés au cours des heures qui ont suivi le début du test.

Nombre d'heures écoulées depuis le début du test t_i	0	2	4	6	8	10
Nombre de noyaux détectés dans l'échantillon (en milliards) N_i	500	440	395	362	316	279

1. On souhaite évaluer le nombre de noyaux présents dans l'échantillon 24 heures après le début du test : quel résultat obtient-on si on a réalisé un ajustement affine ? Est-ce satisfaisant ? On pourra utiliser la calculatrice.

2. Afin d'améliorer l'ajustement, on effectue un changement de variable en posant $y_i = \ln(N_i)$

a. Compléter le tableau et représenter le nuage de points de coordonnées $(t_i; y_i)$ sur la calculatrice. On arrondira les valeurs de y_i à 10^{-3}.

b. À l'aide de la calculatrice, donner une équation de la droite D d'ajustement de y en t par la méthode des moindres carrés sous la forme $y = at + b$, où les coefficients a et b seront arrondis à 10^{-3}.

3. a. On choisit la droite D obtenue à la question 2.b comme modèle d'ajustement du nuage de points $M_i(t_i; y_i)$. Exprimer alors en fonction de t le nombre de noyaux, en milliards, détectés dans l'échantillon au bout de t heures écoulées depuis le début du test.

b. La loi de désintégration assure que la fonction *f*, qui à tout réel *t* positif ou nul, associe le nombre de noyaux, en milliards, présents dans l'échantillon *t* heures après le début du test, est définie par $f(t) = 500e^{-0,06}$. Le test réalisé doit-il conduire la société à exploiter le nouveau détecteur?

4. On admet que la fonction *f* définie pour tout réel *t* positif par $f(t) = 500e^{-0,06t}$ est une bonne approximation du nombre de noyaux, en milliards, présents dans l'échantillon radioactif t heures après le début du test.

a. En utilisant la fonction *f*, évaluer le nombre de noyaux présents dans l'échantillon 24 heures après le début du test. On arrondira au million.

b. Au bout de combien d'heures environ la moitié des noyaux présents dans l'échantillon au début du test aura-t-elle disparu? On justifiera la réponse par un calcul et on arrondira à l'heure.

Indications

1) Avec un ajustement affine, on obtient un résultat négatif, ce qui ne convient pas.

2) a.

t_i	0	2	4	6	8	10
$y_i = \ln(N_i)$	6,215	6,087	5,979	5,892	5,756	5,631

La droite D a pour équation $y = -0{,}057t + 6{,}212$

3) a) Avec ce modèle, $\ln(N) = -0{,}057t + 6{,}212$ *ce qui équivaut à*

$N = e^{-0{,}057t+6{,}212} = e^{-0{,}057t} \times e^{6{,}212} \approx 499e^{-0{,}057t}$

b) Le test conduit à une fonction qui est proche de la fonction f donc on peut utiliser ce dispositif.

4) a) On calcule $f(24) \approx 118{,}464$ *donc environ 118,464 milliards de noyaux*

b) On doit résoudre $f(t) \leq 250$ *soit* $500e^{-0{,}06t} \leq 250$ *c'est-à-dire* $e^{-0{,}06t} \leq \frac{1}{2}$

qui équivaut à $-0{,}06t \leq \ln\left(\frac{1}{2}\right)$*. Ainsi* $t \geq -\frac{\ln\left(\frac{1}{2}\right)}{0{,}06}$ *ou* $t \geq 11{,}6$

La moitié des noyaux aura disparu au bout de 12h.

Exercice 2 : Choisir un bon ajustement (une variante)

Suite à un accident nucléaire sur un site, on a consigné dans le tableau suivant, heure par heure, les résultats fournis par un appareil de mesure de la radioactivité. Les N_i sont des nombres entiers représentant le nombre de particules recueillies par l'appareil pendant une seconde. Le 1er relevé a eu lieu au temps $t = 0$ juste après l'accident.

t_i en heures	0	1	2	3	4	5	6
N_i	170	102	63	39	24	16	9

1) Etudier si un ajustement affine semble adapté. Expliquer.

2) On décide d'effectuer un changement de variable en posant $z_i = \ln(N_i - 2)$ pour tout i variant de 0 à 6 (ln désigne le logarithme népérien).

a) Déterminer les valeurs de z_i arrondies au millième et représenter le nuage de points de coordonnées $(t_i; z_i)$ sur l'écran de la calculatrice.

b) Préciser le coefficient de corrélation linéaire et donner une équation de la droite de régression de z en t (on utilisera la calculatrice et on arrondira les coefficients au millième)

3) Donner l'expression de N en fonction de t déduite de l'ajustement obtenu à la question 2.

4) On considère que le risque de contamination est modéré lorsque le nombre de particules recueillies par l'appareil en une seconde devient inférieur à 3. En supposant que l'expression obtenue à la question 3 reste valable, déterminer au bout de combien de temps le risque de contamination sera modéré.

Indications

1) Un ajustement affine n'est pas très adapté

2)

t_i en heures	0	1	2	3	4	5	6
N_i	170	102	63	39	24	16	9
z_i	5,124	4,605	4,111	3,611	3,091	2,639	1,946

Les points sont proches de la droite de régression d'équation $z = -0{,}517t + 5{,}142$ *et* $r = 0{,}999$

On a $\ln(N - 2) = -0{,}517t + 5{,}142$

On en déduit $N - 2 = e^{-0{,}517t+\ \ {,}142}$ *d'où* $N = e^{-0{,}517t+\ \ {,}142} + 2$

$N < 3$ *équivaut à* $e^{-0{,}517t\quad {,}142} < 1$

soit $-0{,}517t + 5{,}142 < 0$ *ce qui donne* $t > 9{,}948.$

C'est donc au bout de 10h que le risque deviendra modéré.

V Autres exercices

Exercice 3 : Parfum (M2 Pro Ingénierie Mathématique Année 2011-2012 Université d'Angers)

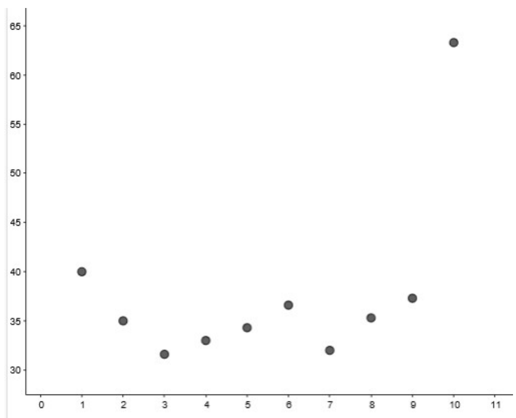
Un « nez » note la qualité de 10 parfums. La qualité est une variable notée de 1 à 10 et les prix des parfums en euros sont présentés dans le tableau ci-dessous.

Parfum	Qualité	Prix
1	10	63,3
2	1	40
3	2	35
4	5	34,3
5	4	33
6	3	31,6
7	6	36,6
8	7	32
9	9	37,3
10	8	35,3

Les prix dépendent-ils de la qualité du parfum ? Justifier la réponse.

Indications

On appelle X la variable « qualité » et Y la variable prix puis on construit le nuage de points de coordonnées (x_i ; y_i). Celui-ci est allongé donc on peut se demander s'il existe une corrélation entre X et Y.



On calcule successivement les moyennes $\bar{x}$ et $\bar{y}$ des valeurs de X et Y puis leur variance et leur covariance pour en déduire l'équation de la droite de régression de Y en X.

On trouve $\bar{x}$= 5,5 ; $\bar{y}$= 37,84 ; $\frac{1}{n}\sum_{i=1}^{10} x_i y_i - \bar{x}\bar{y} =$ 220,3-5,5 ×37,84 =12 ,18 ; Var(X)= 8,25 ; Var(Y)=77,8224 ; σ_x =2,872 ; σ_y =8,8217

Cov(X,Y) = $\frac{12,18}{8,25}$ ≈1,476

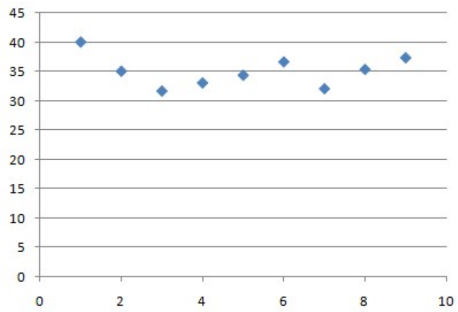
Equation de la droite de régression de Y en X : $y = 1{,}476(x - 5{,}5) + 37{,}84$ *;*

$y = 1{,}476\,x + 29{,}722$

Coefficient de corrélation :

r= $\frac{Cov(X,Y)}{\sigma_x \sigma_y} \approx$ *0,058*

<u>Remarque</u>



Si on néglige le couple (10 ; 63,3) qui semble extérieur au nuage global, on trouve comme équation de la droite de régression : ($\bar{x}$*= 5 ;* $\bar{y}$*= 35,01 ;* $\frac{1}{n}\sum_{i=1}^{10} x_i y_i - \bar{x}\bar{y} =$ *174,44 -175,06 =-0,62 ; Var(X)=6,67 ; Var(Y)=20,39 ;* σ_x *=2,58 ;* σ_y *=4,52 ; cov(X,Y)=0,093)*

$$y = -0{,}093\,x + 35{,}475$$

Coefficient de corrélation : r= - 0,05334,

On peut en conclure dans tous les cas que la qualité du parfum n'intervient pas dans le prix.

Exercice 4 : Poids des pères, poids des fils

Une étude statistique porte sur les poids respectifs en kg des pères p_i et ceux de leur fils ainé f_i pour i variant de 1 à 12.

Voici les résultats numériques obtenus : $\sum_{i=1}^{12} p_i$ = 800 ; $\sum_{i=1}^{12} {p_i}^2$ = 53418 ; $\sum_{i=1}^{12} p_i f_i$ = 54107 ;

$\sum_{i=1}^{12} f_i$ = 811 ; $\sum_{i=1}^{12} {f_i}^2$ = 54849.

1) Déterminer l'équation de la droite des moindres carrés du poids des fils en fonction du poids des pères.

2) Déterminer l'équation de la droite des moindres carrés du poids des pères en fonction du poids des fils.

3) En quel point se coupent ces deux droites ? Justifier.

4) Que vaut le produit des pentes des deux droites ?

<u>Indications</u>

1) On utilise les résultats donnés pour calculer

$\bar{p} = 66{,}67, \bar{f} = 67{,}58, Var(p) = 6{,}61, Var(f) = 3{,}69,\ \sigma_p = 2{,}57, \sigma_f = 1{,}92$

On en déduit Cov (p, f)=3,36

Equation de la droite des moindres carrés du poids des fils en fonction du poids des pères :

$f = \frac{3{,}36}{6{,}61}(p - 66{,}67) + 67{,}58$ *soit* f *=0,508* p *+33,71*

2) Equation de la droite des moindres carrés du poids des pères en fonction du poids des fils :

Pour cela on note f' *le poids des pères et* p' *celui des fils. L'équation de la droite de régression du poids des pères en fonction du poids des fils est*

$f' = \frac{3{,}36}{3{,}69}(p' - 67{,}58) + 66{,}67$ *soit* $f' = 0{,}911p' + 5{,}105$ *or f'= poids du père et p'= poids du fils. Donc* $p = 0{,}911f + 5{,}105$ *d'où* $f = \frac{1}{0{,}911}\ p - 5{,}604$*.*

3) Les deux droites se coupent au point moyen G ($\bar{p}$; $\bar{f}$).

4) Comme les coefficients directeurs des deux droites sont respectivement $\frac{cov(p,f)}{var(p)}$ *et* $\frac{1)}{\frac{cov(p,f)}{var(f)}}$ *le produit des deux est égal à* $\frac{var(f)}{var(p)}$*.*

Exercice 5 (IUT de Toulouse 3 A Département GEA PONSAN)

On étudie la capacité d'oxygénation d'un milieu dans un fermenteur. Soit C la concentration de dioxygène dans ce milieu à l'instant t. Soit CM la concentration maximale en dioxygène que l'on peut obtenir dans ce milieu. C et CM s'expriment en g/l. Un appareil mesure, toutes les trente secondes, le rapport, pourcentage de saturation en dioxygène. Les résultats sont rassemblés dans le tableau ci-dessous :

Temps ti (en secondes)	0	30	60	90	120	150	180	210
xi = C/CM	0,180	0,392	0,569	0,690	0,784	0,838	0,879	0,908

1) a. Représenter le nuage de points de coordonnées (ti ; xi), ainsi que la droite des moindres carrés de x en t.

b. Préciser le coefficient de corrélation linéaire entre x et t.

2) a. On pose y = ln(1 − x). Calculer les valeurs de y_i arrondies au dix millième et représenter le nuage de points de coordonnées (ti ; yi), ainsi que la droite des moindres carrés de y en t.

b. Préciser le coefficient de corrélation linéaire entre y et t.

3) En s'appuyant sur les questions précédentes, déterminer la formule la mieux adaptée pour exprimer x en fonction de t

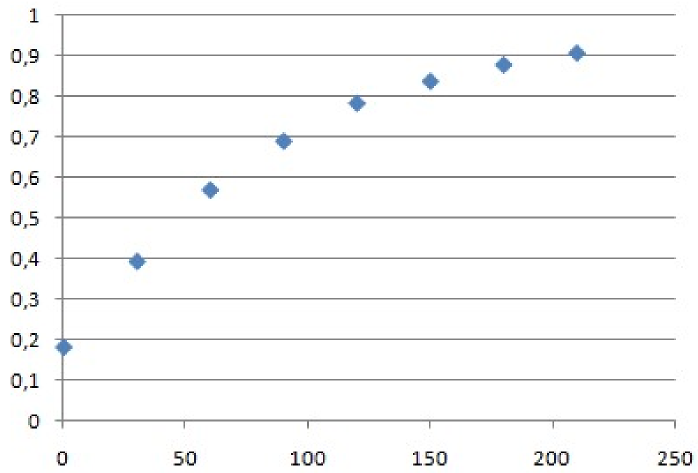
4) Grâce à la formule obtenue au 3), déterminer x lorsque t=220, t=240, t=250.

5) Toujours en utilisant la formule du 3), déterminer le temps t au bout duquel x=0,99.

Indications

1)

2)



Covariance entre x et t : 15,81 ; Coefficient de corrélation entre x et t : $\frac{cov(x,t)}{\sigma_x\ \sigma_t} = \frac{15,81}{\sqrt{4725}\sqrt{0,0585}}$ *=0,951*

2)

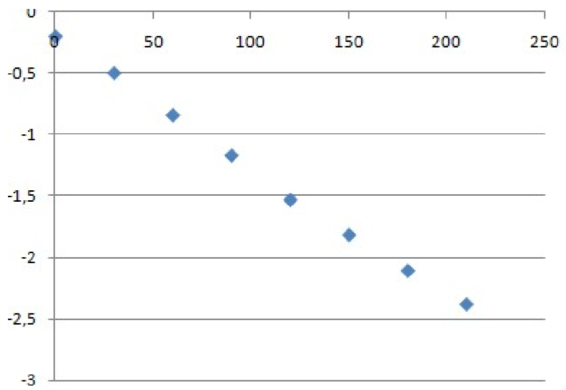
Temps ti (en secondes)	*0*	*30*	*60*	*90*	*120*	*150*	*180*	*210*
x_i *= C/CM*	*0,180*	*0,392*	*0,569*	*0,690*	*0,784*	*0,838*	*0,879*	*0,908*
y_i *=ln(1-xi)*	*-0,1985*	*-0,4976*	*-0,8416*	*-1,1712*	*-1,5325*	*-1,8202*	*-2,112*	*-2,386*

Cov(t, y) ≈ - 50,016 ; $\sigma_t \approx 73{,}485$; $\sigma_y \approx 0{,}778$

Equation de la droite de régression de y en t : y=-0,009(t-105)-1,320 ou encore

$$y = -0{,}009\ t - 0{,}375$$

Coefficient de corrélation entre y et t : - 0,5299



3) La formule qui semble la meilleure est celle trouvée dans le 2) car les points du nuage obtenu sont pratiquement alignés et le coefficient de corrélation est très proche de -1.

4) En utilisant la formule trouvée dans le 2), lorsque t=220 on a y=-2,355 donc ln(1-x)= - 2,355

On en déduit que 1-x= e $^{-2{,}355}$ *d'où x=1-e* $^{-2{,}355}$ $\approx$*0,905.*

On raisonne de façon analogue pour les autres résultats demandés.

5) Si x=0,99 alors y= ln(1-x)= - 4,605 et t= $\frac{4{,}605 \quad ,375}{0{,}009} = 470{,}019$

Exercice 6

On veut déterminer si la tension artérielle Y est corrélée à l'âge X d'une personne. Après mesures et calculs on a obtenu les résultats suivants : $\bar{X}$= 35 et $\bar{Y}$ = 13.5 ; Var(X) = 4 ; Var(Y) = 64 et Cov(X, Y) = 10. Calculer le coefficient de corrélation linéaire entre X et Y puis conclure.

Exercice 7 : https://miashs-www.u-ga.fr/prevert/MathSHS/SHS1/Stat1/TD/TD10.htm

Lors d'une certaine expérience, des sujets sont soumis à deux stimuli. On observe leur temps de réaction mesuré en secondes. Les résultats sont récapitulés dans le tableau ci-dessous.

Sujets	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1635
Stimulus 1	35	29	30	31	33	32	33	34	32	33	32	30	28	34	33	33
Stimulus 2	35	36	34	34	36	37	33	35	35	35	35	33	36	37	34	35

On désigne par X la variable temps de réaction au stimulus 1 et par Y la variable temps de réaction au stimulus 2.

1) Tracer le nuage de points des individus et placer le point moyen.

2) Calculer la covariance de X et Y.

3) Calculer le coefficient de corrélation linéaire entre X et Y. Que peut-on en conclure ?

Exercice 8 : deux juges et des notes

Le tableau de données ci-dessous représente les notes attribuées par deux juges à 10 sportifs lors d'une compétition.

Sportif	1	2	3	4	5	6	7	8	9	10
Juge 1	72	65	80	84	74	68	82	68	76	71
Juge 2	76	63	84	87	79	69	81	71	73	77

1) Représenter le nuage des 10 sportifs.

2) Calculer la moyenne et la variance des notes des deux juges.

3) Calculer le coefficient de corrélation linéaire entre les deux notes.

4) Un onzième sportif a obtenu 85 avec le juge 1. Quelle note est-il susceptible d'obtenir avec le juge 2 ?

Exercice 9 : Deux sortes d'ajustement : affine et exponentiel

Pour étudier la progression d'une épidémie, on procède à une enquête sur un échantillon de 1000 personnes. Le tableau ci-dessous indique le nombre N(t) d'individus ayant été contaminés à la date t exprimée en jours.

t	1	2	5	10	15	20
N(t)	88	172	306	420	485	500

On considère qu'après 20 jours, l'épidémie est terminée c'est-à-dire que le nombre total de personnes ayant été contaminées ne varie plus.

1) a. Dans un plan muni d'un repère orthogonal (O, $\vec{i}$; $\vec{j}$), placer les points de coordonnées

(t ; N(t)). On prendra 0,5 cm pour 1 jour en abscisse et 1 cm pour 50 individus en ordonnée.

b. Donner la valeur arrondie à $\mathbf{10^{-2}}$du coefficient de corrélation linéaire de la série statistique double donnée dans le tableau. Un ajustement linéaire est-il envisageable ?

c. Déterminer une équation de la droite de régression de N en t et la tracer. Les coefficients seront arrondis à l'unité.

2) On considère la fonction définie par $f(t) = 500(1 - e^{-0,2t})$

a) Recopier et compléter le tableau ci-dessous. (Les résultats seront arrondis à l'unité)

t	1	2	5	10	15	20	30	40
f(t)								

b) Tracer la courbe de f dans le repère précédent.

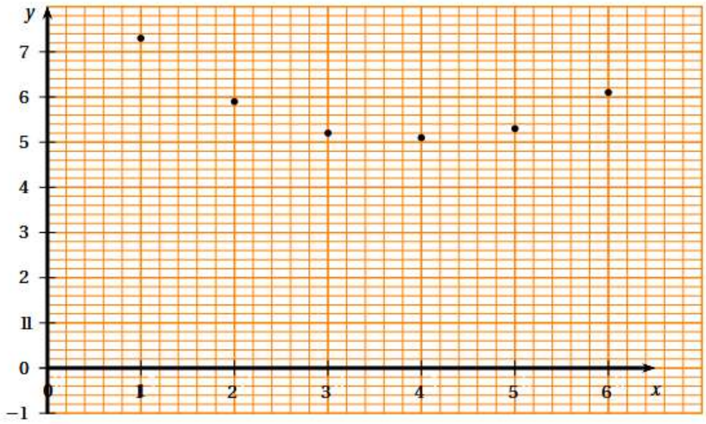
c) On considère la droite du 1) et la courbe précédente. Déterminer graphiquement laquelle des deux ajuste le mieux le nuage de points. L'utiliser pour indiquer la date à laquelle le quart de la population étudiée a déjà été atteint.

Exercice 10 Avec une fonction du second degré

Le tableau ci-dessous donne, pour chaque année de 2011 à 2016, le nombre de visiteurs d'un site touristique en milliers.

année	2011	2012	2013	2014	2015	2016
Nombre de visiteurs en milliers	7,3	5,9	5,2	5,1	5,3	6,1

On note x_i le rang de l'année à partir de 2011 (2011 est l'année de rang 1) et y_i le nombre de visiteurs (en milliers) lors de l'année x_i. Le nuage de points associé à la série statistique $(x_i;\ y_i)$ est représenté ci-contre.



1) Le coefficient de corrélation linéaire entre x et y est environ égal à $-0{,}5$.
Est-ce en accord avec le graphique ? Justifier la réponse. Un ajustement affine est-il adapté ?

2) Afin de réaliser un ajustement à l'aide d'une parabole, on effectue le changement de variable

$$t_i = (x_i - 4)^2.$$

Présenter dans un tableau la nouvelle série $(t_i; y_i)$ et donner son coefficient de corrélation linéaire à 10^{-3}près à l'aide de la calculatrice. Qu'en déduisez-vous ?

3) Donner l'équation de la droite de régression de y en t fournie par la calculatrice. On arrondira les résultats à 10^{-2} près.

4) Le site a accueilli 8980 visiteurs en 2018 ; cette valeur correspond-elle à ce qui était prévu par l'ajustement réalisé dans la question 3 ?

5) A l'aide de l'équation obtenue à la question 3, donner un ajustement de la forme $y = ax^2 + bx + c$ et retrouver par calcul le nombre de visiteurs attendus en 2018.

<u>Indications</u>

1) le coefficient est négatif car on a plutôt une baisse du nombre de visiteurs ; sa valeur absolue 0 ,5 indique que les points ne sont pas du tout alignés. Un ajustement linéaire ne serait pas adapté.

2)

x_i	1	2	3	4	5	6
t_i	9	4	1	0	1	4
y_i	7,3	5,9	5,2	5,1	5,3	6,1

On trouve $r = 0{,}995$; l'ajustement linéaire est cette fois pertinent.

3) La calculatrice donne $y = 0{,}25t + 5{,}03$

4) 2018 est l'année de rang 8 donc $t = (8-4)^2 = 16$ d'où$y = 0{,}25 \times 16 + 5{,}03 = 9{,}03$. Donc 9030 visiteurs étaient prévus en 2018 et 8980 sont venus. Les deux valeurs sont proches, la tendance n'a pas changé, l'ajustement était encore valable pour 2018.

5) On remplace t par $(x-4)^2$ dans l'équation de la droite ; on obtient
$y = 0{,}25(x-4)^2 + 5{,}03 = 0{,}25x^2 - 2x + 9{,}03$.
Pour $x = 8$, on retrouve évidemment $y = 9{,}03$ c'est-à-dire 9030 visiteurs.

ANNEXES

Annexe 1 : Descriptif du programme

À travers l'étude de séries statistiques à deux variables, l'objectif de ce thème est d'amener l'élève à évaluer une corrélation entre deux phénomènes, à développer une réflexion critique sur le lien entre deux phénomènes corrélés, et finalement à distinguer corrélation et causalité.

C'est aussi l'occasion de travailler sur la droite de régression, et de faire percevoir le sens de l'expression « moindres carrés ».

Des ajustements affines ou s'y ramenant à l'aide d'un changement de variable permettent des interpolations et des extrapolations, sur lesquelles l'élève porte un regard critique.

Ce thème d'étude a d'innombrables applications en sciences expérimentales ou en sciences sociales. La corrélation entre deux variables peut être une première approche vers une loi déterministe ou non. Quand une des variables est le temps, le problème de l'extrapolation prend souvent une importance particulière, comme le montre l'exemple du changement climatique.

Problèmes possibles

– Établissement de la loi d'Ohm.
– Loi de désintégration radioactive.
– Évolution de la température et des émissions de gaz à effet de serre dans le cadre du réchauffement climatique.
– Loi de Moore.

Contenus associés

– Fonctions usuelles.
– Représentations graphiques.
– Minimum d'une fonction trinôme.
– Séries statistiques à deux variables.

Histoire

Au XVIIIe siècle, sous l'influence d'hommes politiques et d'économistes, les publications de données sur la démographie, les maladies, les impôts, etc., se multiplient considérablement, consacrant la naissance de la statistique en tant qu'instrument mathématique d'observation sociale.

Avec Bayes, on assiste aux débuts de la statistique inférentielle.

Au début du XIXe siècle, la modélisation des erreurs de mesure va devenir centrale pour faire de la statistique une science à part entière.

Lagrange et Laplace développent une approche probabiliste de la théorie des erreurs. **Gauss (1809, 1821), après Legendre (1805)**, **imagine une méthode des moindres carrés** qu'il applique avec succès à la prédiction de la position d'un astéroïde. Il y propose de comprendre l'écart-type comme une « erreur moyenne à craindre ».

L'introduction de méthodes statistiques en sociologie est l'œuvre du mathématicien et astronome belge **Quételet** dans les années 1830. Il réfléchit à la distribution de données autour de la moyenne, ce qui sera approfondi notamment par l'Anglais Galton. De son côté, Pearson s'intéresse à la corrélation entre variables quantitatives, à la base de la régression linéaire. Au XXe siècle, Student et Fisher développent la biométrie et précisent la différence entre le domaine des probabilités et celui d'une statistique devenue mathématique. Aujourd'hui, les statistiques jouent un rôle essentiel dans les algorithmes de l'intelligence artificielle et de l'apprentissage machine.

Statistique à deux variables quantitatives

L'étude de séries statistiques à deux variables permet de conjecturer des relations, affines ou exponentielles par exemple, entre deux quantités physiques, biologiques ou autres.

Elle apparaît ainsi naturellement dans plusieurs thèmes d'étude. Elle s'appuie notamment sur les études de fonctions classiques et les représentations graphiques.

Contenus

– Nuage de points. Point moyen.
– Ajustement affine. Droite des moindres carrés. Coefficient de corrélation.
– Ajustement se ramenant par changement de variable à un ajustement affine.
– Application des ajustements à des interpolations ou extrapolations.

Capacités

– Représenter un nuage de points.
– Calculer les coordonnées d'un point moyen.
– Déterminer une droite de régression, à l'aide de la calculatrice, d'un logiciel ou par calcul.
– Dans le cadre d'une résolution de problème, utiliser un ajustement pour interpoler, extrapoler.

Démonstration possible– Droite des moindres carrés.

Annexe 2 : Références

https://www.quebecscience.qc.ca/14-17-ans/encyclo/correlation-coincidence-ou-causalite/

https://econofides.ac-versailles.fr/terminale-ses-2019/text/toolbox-16.html

https://www.bcpst.eu/2019/10/16/correlation-et-causalite/

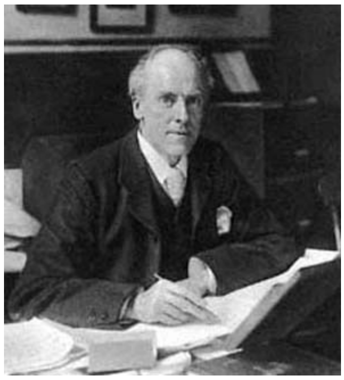
https://www.citeco.fr/faites-parler-les-donn%C3%A9es-causalit%C3%A9s-et-corr%C3%A9lations

https://www.statsoft.fr/concepts-statistiques/statistiques-elementaires/correlations.php

http://www.biostat.ulg.ac.be/pages/Site_r/corr_pearson.html

Annexe 3 : Pearson et la covariance- quelques repères historiques

L'étude conjointe de la variation de deux mesures (par exemple, de la taille du père avec celle de son fils; ou de la longueur du bras avec celle de la jambe d'un même homme) fait la gloire de l'anglais Galton (1822-1911). Les statisticiens se sont posé le problème de la comparaison des enfants aux parents. La théorie de Darwin a inspiré de grands enthousiasmes et même des passions durables qui soutinrent l'extraordinaire labeur de l'école biométrique issue de F. Galton.
Quand Galton entreprit de comparer les caractères des enfants à ceux des parents, les documents anthropométriques manquaient totalement pour cela. Il commença par l'étude des petits pois : il mesura des graines, les sema et mesura les graines de sa récolte. Il compara le diamètre j des graines et celui i des pois récoltés et trouva une moyenne $\bar{\imath}$ approximativement fonction linéaire des j à savoir $\bar{\imath}(j)$ - $\bar{\imath}$ = r (j-$\bar{j}$)
Ayant trouvé pour les diamètres des pois r = 0,33, Galton souligna d'abord que $\bar{\imath}(j)$ - $\bar{\imath}$ des pois de sa récolte, n'est que le tiers de l'écart j-$\bar{j}$ des semences: il y a donc eu d'une génération à la suivante retour vers la moyenne.
En 1877, devant la Royal Institution of Great Britain, Galton parle de réversion, on parlera ensuite de régression. La lettre r en est restée pour désigner le coefficient de corrélation
C'est le Physicien, astronome, géologue, géodésiste, probabiliste Auguste Bravais (1811-1863) qui le premier prononce le mot de corrélation.
A la suite de ses travaux et après la lecture de l'ouvrage de Galton, K. Pearson (1857-1936) et R. Weldon (1860-1906), en devinrent des disciples et collaborèrent s'appliquant à fournir par la biométrie des preuves expérimentales rigoureuses de la doctrine de l'évolution et perfectionnant l'outil statistique pour élaborer de telles données et en critiquer l'interprétation.



Karl Pearson

Annexe 4 : Une interprétation du coefficient de corrélation

Le **coefficient de Bravais-Pearson**, dit coefficient de corrélation linéaire, a pour expression $r_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y}$

où r_{xy} représente le cosinus des variables x et y centrées sur leur moyenne respective.

Interprétation géométrique du coefficient de corrélation

Les deux séries de valeurs X(x1, x2, …, xn) et Y(y1, y2, …, yn) peuvent être considérées comme des vecteurs dans un espace à n dimensions. Remplaçons-les par des vecteurs centrés : $\vec{\boldsymbol{X}}(\boldsymbol{x_1}\text{-}\bar{\boldsymbol{x}}, \boldsymbol{x_2}-\bar{\boldsymbol{x}}, \dots, \boldsymbol{x_n}-\bar{\boldsymbol{x}})$ et $\vec{\boldsymbol{Y}}(\boldsymbol{y_1}\text{-}\bar{\boldsymbol{y}}, \boldsymbol{y_2}\text{-}\bar{\boldsymbol{y}}, \dots, \boldsymbol{y_n}-\bar{\boldsymbol{y}})$

Le cosinus de l'angle α entre ces vecteurs est donné par la formule suivante (produit scalaire normé) : $\cos(\alpha) = \frac{\vec{X}.\vec{Y}}{\|\vec{X}\|\times\|\vec{Y}\|}$

$$\cos(\alpha) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\times\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{\frac{1}{n}cov(X,Y)}{\sqrt{\frac{1}{n}}\sigma_X\times\sqrt{\frac{1}{n}}\sigma_Y} = r$$

Donc $\cos(\alpha) = r$, ce qui explique que r est toujours compris entre -1 et 1.

Le coefficient de corrélation n'est autre que le cosinus de l'angle α entre les deux vecteurs centrés.

Si r = 1, l'angle α = 0, les deux vecteurs sont colinéaires (parallèles).
Si r = 0, l'angle α = 90°, les deux vecteurs sont orthogonaux.
Si r = -1, l'angle α vaut 180°, les deux vecteurs sont colinéaires de sens opposé.

Annexe 5 : Relation entre les coefficients de corrélation linéaire $\boldsymbol{r}$ et de détermination $\boldsymbol{R^2}$ dans le cas d'une régression affine simple : $R^2 = r^2$

<u>Démonstration</u> (*pour faciliter la lecture on simplifie $\sum_i^n$ … en $\sum$ …)*

Par définition $R^2 = \frac{\sum(y_i'-\bar{y})^2}{\sum(y_i-\bar{y})^2}$ est aussi égal à $1 - \frac{\sum(y_i-y_i')^2}{\sum(y_i-\bar{y})^2} = \frac{nV(Y)-\sum(y_i-y_i')^2}{nV(Y)}$

On développe $(y_i - y_i')^2 = [y_i - (ax_i + b)]^2$en remplaçant a et b par leurs expression en fonction de r, de σ_X et de σ_Y.
On obtient :

$$(y_i - y_i')^2 = \left[y_i - r\times\frac{\sigma_Y}{\sigma_X}\times x_i - \left(\bar{y} - r\times\frac{\sigma_Y}{\sigma_X}\times\bar{x}\right)\right]^2 = [(y_i-\bar{y}) - r\times\frac{\sigma_Y}{\sigma_X}\times(x_i-\bar{x})]^2$$

$$(y_i - y_i')^2 = (y_i-\bar{y})^2 - 2\,r\times\frac{\sigma_Y}{\sigma_X}\times(x_i-\bar{x})(y_i-\bar{y}) + r^2\times\frac{V_Y}{V_X}\times(x_i-\bar{x})^2$$

En sommant de 1 à n on a :

$$\sum(y_i - y_i')^2 = nV(Y) - 2\,nr\times\frac{\sigma_Y}{\sigma_X}\times r\times\sigma_X\times\sigma_Y + r^2\times\frac{V_Y}{V_X}\times nV(X)$$

$$\sum(y_i - y_i')^2 = nV(Y) - 2\,nr^2V(Y) + nr^2V(Y) = nV(Y)(1-r^2)$$

D'où $R^2 = \frac{nV(Y)-(nV(Y)(1-r^2)}{nV(Y)} = 1-(1-r^2) = r^2$. On a bien $R^2 = r^2$

Annexe 6 : Fonctions de deux variables - recherche d'extremums

f est une fonction de classe C^2 sur un ouvert U de deux variables x et y.

Pour une fonction f de deux variables x et y, on utilise les notations de Monge.

En un point $(x_0;\ y_0)$, on note $p = \frac{\partial f}{\partial x}(x_0, y_0) \quad q = \frac{\partial f}{\partial y}(x_0, y_0)$

$r = \frac{\partial^2 f}{\partial x^2}(x_0, y_0) \quad s = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) = \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) \quad t = \frac{\partial^2 f}{\partial y^2}(x_0, y_0)$

Théorème

Soit f une fonction de classe C^2 sur un ouvert U et soit (x_0, y_0) un point de U tel que $p = q = 0$.

- Si $rt - s^2 > 0, f$ admet un extremum en ce point ;

  un minimum si $r > 0,$ un maximum $si\ r < 0.$

- Si $rt - s^2 < 0$, il n'y a pas d'extremum en ce point.
- Si $rt - s^2 = 0$, on ne peut pas conclure par cette méthode ; il faut alors étudier le signe de $f(x, y) - f(x_0, y_0)$

Conclusion :

Pour chercher les extremums éventuels d'une fonction f de deux variables sur un ouvert U, il faut déterminer les points critiques de f ce qui revient à résoudre le système de deux équations à deux inconnues $\begin{cases} \frac{\partial f}{\partial x}(x, y) = 0 \\ \frac{\partial f}{\partial y}(x, y) = 0 \end{cases}$ puis à calculer, pour chacun des points trouvés, $rt - s^2$ et conclure à l'aide du théorème ci-dessus.

Exemple : recherche des extremums éventuels de la fonction f définie par

$f(x, y) = x^3 + 3xy^2 - 15x - 12y$

$\frac{\partial f}{\partial x}(x, y) = 3x^2 + 3y^2 - 15$; $\frac{\partial f}{\partial y}(x, y) = 6xy - 12$;

$$r = \frac{\partial^2 f}{\partial x^2}(x, y) = 6x\ ; \quad s = \frac{\partial^2 f}{\partial x \partial y}(x, y) = 6y\ ;\ t = \frac{\partial^2 f}{\partial y^2}(x, y) = 6x$$

On résout le système $\begin{cases} 3x^2 + 3y^2 - 15 = 0 \\ 6xy - 12 = 0 \end{cases}$ équivalent à $\begin{cases} x^2 + y^2 - 5 = 0 \\ xy = 2 \end{cases}$

On trouve $\begin{cases} x = 1 \\ y = 2 \end{cases}$ ou $\begin{cases} x = -1 \\ y = -2 \end{cases}$ ou $\begin{cases} x = 2 \\ y = 1 \end{cases}$ ou $\begin{cases} x = -2 \\ y = -1 \end{cases}$.

On a donc 4 points critiques A(1 ;2), B(-1 ; -2), C(2 ; 1), D(-2 ; -1)

Pour A et B, $rt - s^2$ =36-12² donc $rt - s^2 < 0. f$ n'admet pas d'extremum en ces points.

Pour C, $rt - s^2$ =108 et $r = 12$ donc $rt - s^2 > 0$ et $r > 0$. f admet un minimum local en C qui est égal à $f(2,1) = -28$

Pour D, $rt - s^2$ =108 et $r = -12$ donc $rt - s^2 > 0$ et $r < 0$. f admet un maximum local en D qui est égal à $f(-2,-1) = 28$

Si on reprend l'exemple du paragraphe III d)

On a $S = 11370a^2 + 548ab - 132400a - 3200b + 7b^2 + 363050$

$$\frac{\partial S}{\partial a}(a,b) = 22740a + 548b - 132400$$

$$\frac{\partial S}{\partial b}(a,b) = 548a + 14b - 3200$$

Le système $\begin{cases} 22740a + 548b - 132400 = 0 \\ 548a + 14b - 3200 = 0 \end{cases}$ admet pour seule solution ($\frac{12500}{2257}$; $\frac{26600}{2257}$)

$$r = \frac{\partial^2 S}{\partial a^2}(a,b) = 22740 \text{ ; } s = \frac{\partial^2 S}{\partial a \partial b}(a,b) = 548 \text{ ; } t = \frac{\partial^2 S}{\partial b^2}(a,b) = 14$$

$$rt - s^2 = 18056 \text{ et } r > 0 \text{ donc } S \text{ admet un minimum pour } a = \frac{12500}{2257} \text{ et } b = \frac{26600}{2257}$$

Annexe 7 : Gauss – Legendre : la méthode des moindres carrés.

https://publimath.univ-irem.fr/numerisation/DI/IDI17004/IDI17004.pdf

Les scientifiques cherchent le plus souvent à adapter un modèle mathématique aux phénomènes qu'ils étudient. Malheureusement les données expérimentales ne concordent pas toujours avec le modèle à cause de la présence de différentes erreurs lors des mesures. Après avoir catalogué ces erreurs, il s'est agi de trouver une méthode permettant d'en minimaliser les conséquences. C'est ainsi qu'est née la méthode des moindres carrés par le travail de deux grands savants.

Il semble généralement accepté qu'Adrien Marie Legendre (1752-1833) est bien l'auteur de la première publication sur la méthode des moindres carrés, méthode dont il a forgé le nom (1805). Mais, il semble aussi souvent accepté que Cari Friedrich Gauss (1777-1855) en fut l'inventeur dès 1795 comme il l'a toujours revendiqué.
Gauss et Legendre ont eu, pour établir « leur » méthode des moindres carrés, deux approches différentes. Legendre présente une méthode algébrique alors que Gauss a une approche probabiliste du problème.

Legendre utilise une méthode empirique, simple à mettre en œuvre, qui consiste à rendre minimale la somme des carrés des erreurs et qu'il justifie par la phrase :*« De tous les principes qu'on peut proposer pour cet objet, je pense qu'il n'en est pas de plus général, de plus exact, ni d'une application plus facile que celui (...) qui consiste à rendre minimum la somme des carrés des erreurs. Par ce moyen, il s'établit entre les erreurs une sorte d'équilibre qui empêchant les extrêmes de prévaloir, est très propre à faire connaître l'état du système le plus proche de la vérité. »*

Gauss, dans son traitement de la théorie des erreurs de 1809, introduit en fait le principe du maximum de vraisemblance (puisqu'il cherche la valeur de la quantité inconnue X qui rendra maximale la probabilité d'observation des mesures observées $x_1, ..., x_n$ et donc des erreurs $e_1, ...e_n$.). Après ajout du postulat de la moyenne arithmétique, considérant que si une quantité a été obtenue par plusieurs observations, faites avec le même soin dans des circonstances semblables, la moyenne arithmétique des valeurs observées sera la valeur la plus probable de cette quantité, il obtient à la fois la loi de distribution normale et le principe des moindres carrés (https://journals.openedition.org/bibnum/580).

INFÉRENCE BAYÉSIENNE

Probabilités conditionnelles, totales, formule de Bayes

Le document peut être téléchargé ici :

http://docs.irem.univ-paris-diderot.fr/up/B101/Bayes_Maths_comp.docx

http://docs.irem.univ-paris-diderot.fr/up/B101/Bayes_Maths_comp.pdf

Formule de Bayes - Formule des probabilités totales.

Cette ressource n'a pas été conçue comme un chapitre de manuel, elle est trop longue pour que tout soit utilisé en classe. Notre intention est d'offrir aux enseignants à la fois des éléments de compréhension de la formule, de sa structure et des prérequis nécessaires, et même de son émergence (cf. point d'histoire au paragraphe C). Le texte peut conduire à des utilisations variées, qui dépassent ce qui est exigible des élèves. Il comporte ainsi des commentaires didactiques et des énoncés nombreux, autorisant des choix adaptés parfois sous forme d'exemples corrigés ou d'exercices. Cependant, nous n'avons pas cherché l'originalité dans ces énoncés.

L'ensemble de la ressource a déjà été étudié par plusieurs enseignants qui ont contribué à son enrichissement en proposant des énoncés testés en classe, présentés ci-dessous sous forme d'alternative aux exemples ou exercices initiaux.

Pourquoi une ressource sur le thème « Formule de Bayes – Formule des probabilités totales » ?

Les intentions du programme

En classe terminale de la voie générale, l'enseignement optionnel de mathématiques complémentaires est destiné prioritairement aux élèves qui, **ayant suivi l'enseignement de spécialité de mathématiques en classe de première** et ne souhaitant pas poursuivre cet enseignement en classe terminale, ont cependant besoin de compléter leurs connaissances et compétences mathématiques par un enseignement adapté à leur poursuite d'études dans l'enseignement supérieur, en particulier en médecine, économie ou sciences sociales. Le programme de mathématiques complémentaires s'appuie sur le programme de spécialité de mathématiques de la classe de première qu'il réinvestit et enrichit de nouvelles connaissances et compétences mathématiques, elles-mêmes reliées à des thèmes d'étude où les notions sont mises en situation dans divers champs disciplinaires.

La formule des probabilités totales fait partie intégrante du programme de 1ère. En terminale « option complémentaire », l'enseignement est organisé autour de thèmes d'étude mettant en contexte des contenus mathématiques. L'inférence bayésienne est l'un de ces thèmes d'étude. La formule de Bayes peut aussi être abordée en Enseignement scientifique de terminale (voir ressource Inférence bayésienne).

Ce que dit le programme

Inférence bayésienne
<u>Descriptif</u>

Le raisonnement bayésien est à la base de nombreux algorithmes de décision et se retrouve dans de nombreux domaines pratiques : sport, médecine, justice … où l'on doit raisonner à partir de probabilités et d'informations incomplètes. Il s'agit ici de décrire et mettre en œuvre les principes du calcul utilisant des probabilités conditionnelles et notamment la formule de Bayes pour l'inversion des conditionnements.

La question d'intérêt est représentée par un événement A de probabilité $P(A)$, dite probabilité a priori. L'observation d'un événement B conduit à remplacer la probabilité a priori $P(A)$ par la probabilité conditionnelle $P_B(A)$, dite a posteriori.

La formule de Bayes $P_B(A) = \frac{P_A(B)\times P(A)}{P(B)}$ permet d'exprimer la probabilité a posteriori lorsque l'expression du second membre est évaluable. Elle montre la distinction essentielle entre $P_B(A)$ et $P_A(B)$. Bien comprendre cette distinction est un objectif majeur.

Problèmes possibles

- Tests binaires pour le diagnostic médical. Notion de vrais/faux positifs et négatifs, sensibilité, spécificité, valeurs prédictives positive (diagnostique) et négative, lien avec les probabilités conditionnelles. Tests de dépistage de sensibilité et de spécificité données : étude des valeurs prédictives en fonction de la proportion de malades et interprétation.
- Exemples de problèmes du type : « De quelle urne vient la boule ? ».

Contenus associés

- Probabilités conditionnelles, inversion du conditionnement, formule de Bayes.
- Étude de fonction.

Cette ressource est organisée en trois grandes parties et des annexes

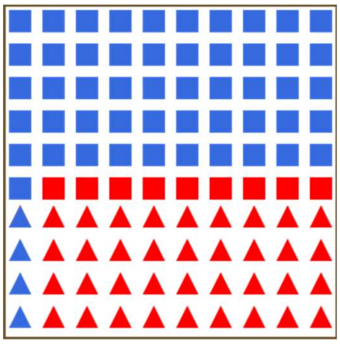
Ils sont choisis pour leur lien avec les thèmes du programme : paradoxe des tests de dépistage, comprendre et exploiter des statistiques. A partir d'observations statistiques on calcule la probabilité d'un événement hypothétique utilisant le retournement de la formule de Bayes (inférence bayésienne).

Annexe 1 : Mise au point sur la démarche suivie, les données, le vocabulaire utilisé selon le contexte et récapitulatif des termes utilisables : quand les probabilités se « mélangent » aux statistiques…
Annexe 2 : Termes utilisés dans les tests médicaux

A– Probabilité conditionnelle et formule des probabilités totales

I – Pour bien démarrer. Ces exercices permettent de mettre en œuvre quelques notions vues en $1^{\text{ère}}$ et qui seront présentées de manière générale dans le II : utilisation d'un arbre de probabilité, probabilités conditionnelles et totales.

Exercice 1 : Revoir les probabilités conditionnelles
(d'après **https://scienceetonnante.com/2012/10/08/les-probabilites-conditionnelles-bayes-level-1/**)



Une boite contient 100 jetons : certains sont carrés, d'autres triangulaires, certains sont bleus et d'autres rouges comme ci-contre.

1) Paul tire un jeton au hasard ; quelle est la probabilité

a) qu'il soit carré ?
b) qu'il soit rouge ?
c) que ce soit un carré rouge ?

Si on note R et C les événements « obtenir un jeton rouge » et « obtenir un carré », écrire les résultats trouvés en utilisant cette notation.

2) Paul tire un jeton au hasard et sans voir sa forme, il entrevoit sa couleur qui est rouge.

a) Quelle est alors la probabilité qu'il soit carré ?
On rappelle que la probabilité demandée s'appelle la probabilité **conditionnelle** de C sachant R, notée $P_R(C)$.

b) Dans la question 1), l'ensemble de référence ou univers était formé de l'ensemble des 100 jetons ; quel est l'univers dans cette question ?

c) Ecrire $P_R(C)$ comme un quotient de deux probabilités choisies parmi $P(R)$, $P(C)$, $P(R \cap C)$.

3) Paul tire un jeton au hasard et, sans le voir, il devine sa forme : il est carré.

a) Quelle est la probabilité qu'il soit rouge ? Exprimer cette probabilité comme une probabilité conditionnelle.
b) Ecrire cette probabilité comme un quotient de deux probabilités (comme à la question 3)c)

4) Peut-on trouver une relation liant $P_R(C)$, $P_C(R)$, $P(R)$ et $P(C)$?

Indications

1) a) On a 60 carrés sur les 100 jetons donc $\frac{60}{100} = 0{,}6 = P(C)$

b) $P(R) = \frac{40}{100} = 0{,}4$

c) $P(C \cap R) = \frac{9}{100} = 0{,}09$

2) On peut s'aider d' un tableau à double entrée (tableau de contingence).

a) On a 9 carrés parmi les 45 jetons rouges donc la probabilité est $\frac{9}{45} = 0{,}2$ *C'est la probabilité d'avoir un jeton carré* ***sachant qu'il*** *est rouge notée* $P_R(C)$.

b) L'ensemble de référence (ou univers) n'est plus l'ensemble des 100 jetons, il est restreint à l'ensemble des 45 jetons rouges.

Avant *de voir sa couleur, la probabilité (dite* ***a priori****) d'avoir un carré était de 0,6 ;* ***après*** *avoir vu sa couleur, cette probabilité (dite* ***a posteriori****) n'est plus que de 0,2. La probabilité d'avoir un carré a donc été modifiée par la connaissance de la couleur du jeton.* $P_R(C) = \frac{9}{45} = 0{,}2$.

Il ne faut pas confondre $P_R(C)$ *avec* $P(R \cap C)$ *égale à* $P(C \cap R)$ *qui vaut* $\frac{9}{100}$. *Cette dernière est évidemment plus petite que* $P_R(C)$ *car son dénominateur est plus grand. Les deux fractions ont ainsi le même numérateur (9 jetons carrés et rouges) mais pas le même dénominateur (tous les jetons, ou seulement les jetons rouges).*

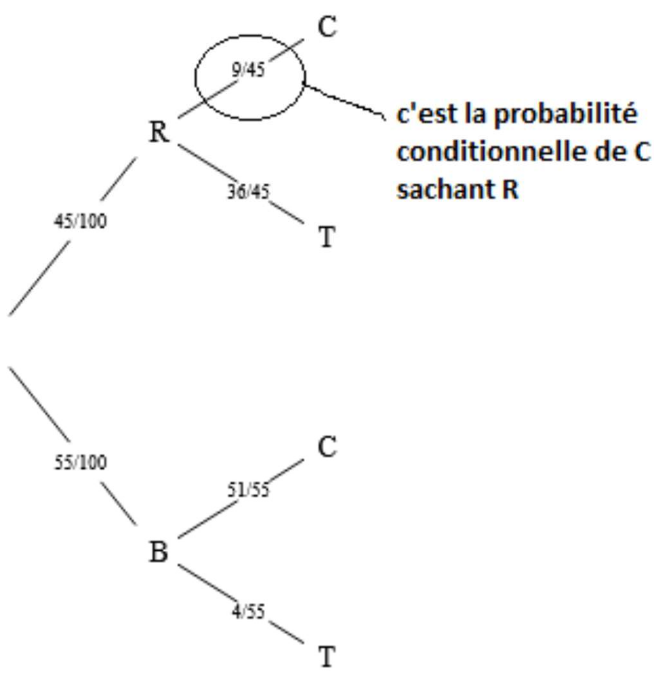
	Carrés	Triangles	Total
Bleus	51	4	55
Rouges	9	36	45
Total	60	40	100

c) $P_R(C) = \frac{9}{45} = \frac{\frac{9}{100}}{\frac{45}{100}} = \frac{P(R \cap C)}{P(R)}$ *; on retrouve la définition vue en 1ère :* $\boldsymbol{P_R(C) = \frac{P(R \cap C)}{P(R)}}$

Remarque : *On peut aussi s'aider d' un arbre :*

On place $p(R) = \frac{45}{100}$ *et on complète la branche R----C avec la probabilité de C sachant R qui est égale à* $\frac{9}{45}$.

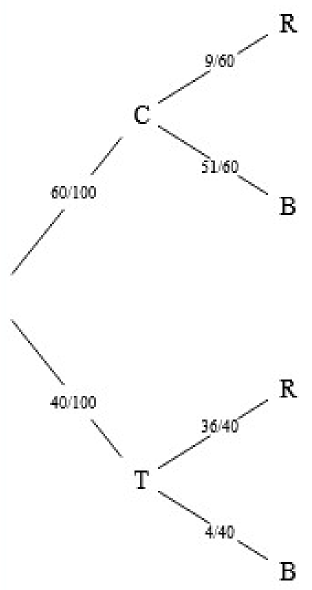
Sur la branche du haut, on multiplie la probabilité que le jeton soit rouge $P(R) = \frac{45}{100}$ *par la probabilité qu'il soit carré sachant qu'il est rouge* $P_R(C) = \frac{9}{45}$, *on retrouve la probabilité de l'intersection* $P(R \cap C) = \frac{9}{100}$.



C'est la formule $P(R\cap C) = P(R)\times P_R(C)$ *que l'on obtient immédiatement à partir de la définition ci-dessus.*
On peut compléter l'arbre en vérifiant qu'à partir d'un même nœud, la somme des probabilités est toujours égale à 1.

3) a) On cherche la probabilité que le jeton soit rouge ***sachant qu'il est carré.***
Avec le tableau, on a 9 jetons rouges (encore 9) parmi les 60 carrés donc $P_C(R) = \frac{9}{60} = 0{,}15$
Pour trouver cette valeur dans un arbre, il faut « inverser » l'arbre précédent. Au lieu de commencer par les couleurs, on commence par les formes.

$P_C(R) = \frac{P(C\cap R)}{P(C)} = \frac{\frac{9}{100}}{\frac{60}{100}} = \frac{9}{60}$ *On peut compléter l'arbre comme précédemment.*



Attention : On remarque que $P_C(R) \neq P_R(C)$ *alors que* $P(C\cap R) = P(R\cap C)$

4) $P_R(C) = \frac{P(R\cap C)}{P(R)}$ *et* $P_C(R) = \frac{P(C\cap R)}{P(C)}$ *donc*
$P(R\cap C) = P(R)\times P_R(C)$ *et* $P(C\cap R) = P(C)\times P_C(R)$
Comme $P(R\cap C) = P(C\cap R)$*, on a*

$P(R)\times P_R(C) = P(C)\times P_C(R)$ *ou encore*

$P_R(C) = \frac{P_C(R)\times P(C)}{P(R)}$ *appelée formule de Bayes (voir plus loin)*

Exercice 2 : Revoir la formule des probabilités totales
D'après le manuel de première Editions Belin, collection Métamaths (savoir-faire 3 page 283)

L'exploitant d'une forêt communale décide d'abattre des arbres : certains sont réservés à la vente aux habitants, les autres sont à destination des entreprises privées.
Parmi les arbres abattus, 30% sont des chênes, 50% sont des sapins et les autres sont des arbres d'essence secondaire.
45,9% des chênes et 80% des sapins abattus sont réservés aux habitants de la commune et les trois quarts des arbres d'essence secondaire abattus sont réservés à la vente à des entreprises privées.
Parmi les arbres abattus on en choisit un au hasard.
On note H l'événement « l'arbre est réservé aux habitants », C l'événement « l'arbre est un chêne », S, l'événement « l'arbre est un sapin », E, l'événement « l'arbre est d'essence secondaire ».
Calculer la probabilité que l'arbre abattu soit réservé aux habitants de la commune.

Indications

La probabilité P(H) que l'arbre soit réservé aux habitants est telle que
P(H)= P(H∩ C) + P(H∩ S) + P(H∩ E)
D'après la formule des probabilités conditionnelles on sait que
P(H∩ C)=P(C)×P_C(H)= 0,3× 0,459 =0,1377
De même P(H∩ S)=P(S)×P_S(H)=0,5×0,8=0,4
De l'énoncé on déduit que 20% des arbres abattus sont d'essence secondaire et qu'un quart des arbres abattus d'essence secondaire est réservé aux habitants donc P_E(H)=0,25
P(H∩ E)=P(E)×P_E(H)= 0,2× 0,25 =0,025
On en déduit que P(H)= 0,1377+0,4+0,05=0,5877

II - ***Probabilité conditionnelle et formule des probabilités totales*** (rappels du cours du programme de l'enseignement de spécialité - classe de première)

1) Probabilités conditionnelles

Définition A et B sont deux événements d'un même univers Ω et A est de probabilité non nulle.

On appelle **probabilité conditionnelle de B sachant A le réel** $\boldsymbol{P_A(B) = \frac{P(A\cap B)}{P(A)}}$

$P_A(B)$ se lit « probabilité de B sachant A ».

On définit ainsi une nouvelle probabilité sur l'univers Ω. On démontre qu'une probabilité conditionnelle est une probabilité (particulière). Autrement dit, l'application P_A qui, à tout événement B de Ω associe P_A(B) vérifie les propriétés des probabilités, notamment :

- $P_A(A) = 1$
- $0 \leq P_A(B) \leq 1$
- $P_A(B) + P_A(\bar{B}) = 1$
- *si B_1 et B_2 sont deux événements disjoints, alors* $P_A(B_1 \cup B_2) = P_A(B_1) + P_A(B_2)$.

Conséquence de la définition : probabilité d'une intersection.

Si A et B sont deux événements de probabilités non nulles, on a :
$P(A \cap B) = P(A) \times P_A(B)$ et aussi $P(A \cap B) = P(B) \times P_B(A)$

Un algorithme utile pour calculer ces probabilités

Script Python qui permet d'afficher $P(A \cap B)$ connaissant $P(A)$ et $P_A(B)$ (def p_1(a,c) :

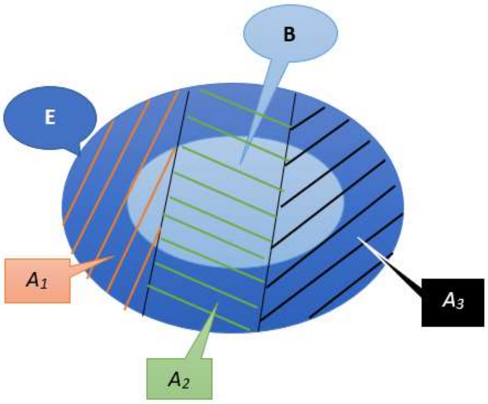
puis $P_B(A)$ connaissant P(B) (def p_2(a,b,c))

```
from math import *
print("Probabilité de A : a")
print("Probabilite de B sachant A: c")

print ("Probabilité de A inter B: p(a,c)")


def p_1(a,c):
    return a*c

print("Probabilite de B: b")
print ("Probabilité de A sachant B: p_2(a,c,b)")

def p_2(a,c,b):
    return (a*c)/b
```

2) Partition de l'univers et formule des probabilités totales

Définition

Si A_1, A_2, A_3 sont des événements deux à deux disjoints et que leur réunion est l'univers E, on dit que A_1, A_2, A_3 forment une **partition** de E.
On a alors $P(A_1) + P(A_2) + P(A_3) = 1$



En particulier, pour tout événement A, les événements A et $\overline{A}$ forment une partition de E (on note « nonA » ou $\bar{A}$ l'événement contraire de A)

Propriété

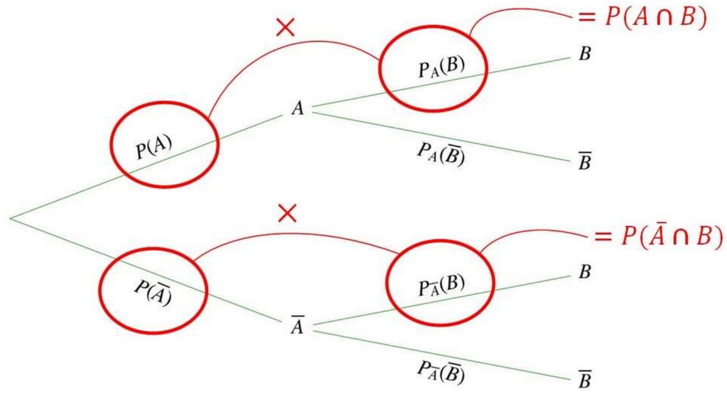
Si les événements A_1, A_2, A_3 forment une partition de l'univers E, on a pour tout événement B de E,
$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$

D'où $\boldsymbol{P(B) = P_{A_1}(B) \times P(A_1) + P_{A_2}(B) \times P(A_2) + P_{A_3}(B) \times P(A_3)}$

C'est la formule dite « des probabilités totales ».

Cas particulier très fréquent : A et B sont deux événements d'un même univers tels que A et $\overline{A}$ sont de probabilité non nulle. La formule des probabilités totales permet de calculer la probabilité d'un événement B à partir des deux probabilités conditionnelles de cet événement sachant A et sachant $\overline{A}$: $\boldsymbol{P(B) = P(A) \times P_A(B) + P(\bar{A}) \times P_{\bar{A}}(B)}$

On peut l'illustrer par l'arbre ci-dessous : $P(B) = P(A \cap B) + P(\overline{A} \cap B)$ puisque $A \cap B$ et $\overline{A} \cap B$ sont disjoints. On multiplie les probabilités rencontrées sur chaque branche de l'arbre comme sur le schéma (www.mathematiques.club) pour obtenir la formule.



On peut aussi étendre la formule à une partition de l'univers formée de 4,5…n événements (n entier supérieur ou égal à 2)

3) Exercices d'application

Exercice 1 : Probabilité d'une intersection
Une entreprise comprend 40% de cadres dont 20% parlent anglais. Quelle est la probabilité qu'un employé pris au hasard soit un cadre qui parle anglais ?

Indications : $P(C \cap A) = P_C(A) \times P(C) = 0{,}2 \times 0{,}4 = 0{,}08$

Exercice 2 : Vaccination et probabilités conditionnelles (d'après bac S).
Au cours d'une épidémie de grippe, on estime que la probabilité qu'une personne choisie au hasard dans une population donnée soit grippée est 0,25. On décide de vacciner le tiers de cette population. Parmi les grippés, on constate qu'une personne sur dix est vaccinée.

1) Comment s'appelle et se note la probabilité qu'une personne qu'on sait grippée soit vaccinée ? Quelle est la probabilité qu'une personne de la population considérée soit à la fois grippée et vaccinée ?
2) Quelle est la probabilité pour un individu vacciné de cette population de contracter la grippe ?
3) Quelle est la probabilité pour un individu non vacciné de cette population de contracter la grippe ?
4) Comparer les résultats des deux questions précédentes et conclure.

Indications

1) *On note G l'événement « la personne choisie est grippée » et V l'événement « la personne choisie est vaccinée »*
D'après l'énoncé $P_G(V)=0{,}10$ car la mention « parmi les grippés » indique qu'on se place dans l'ensemble des personnes grippées.
On sait de plus que $P(V)=\frac{1}{3}$, et $P(G)=0{,}25$.
Or $P_G(V) = \frac{P(V \cap G)}{P(G)}$ d'où $P(V \cap G) = P_G(V) \times P(G) = 0{,}10 \times 0{,}25 = 0{,}025$.
On peut s'aider d'un arbre, même s'il est incomplet.

2) *On cherche $P_V(G)$: $P_V(G) = \frac{P(V \cap G)}{P(V)} = \frac{0{,}025}{\frac{1}{3}} = 0{,}075$*

Une personne vaccinée de cette population a 7,5 chances sur 100 d'attraper la grippe.

3) *On cherche $P_{\overline{V}}(G)$: $P_{\overline{V}}(G) = \frac{P(\overline{V} \cap G)}{P(\overline{V})} = \frac{0{,}25 \times 0{,}90}{1-P(V)} = \frac{0{,}225}{\frac{2}{3}} = 0{,}3375$*

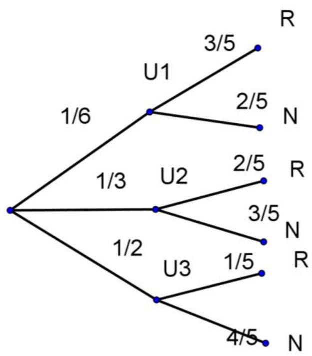
Une personne non vaccinée a environ 34 chances sur 100 d'avoir la grippe.

4) *Si on n'est pas vacciné, on a donc environ 4,5 fois plus de chances d'avoir la grippe…*

Exercice 3 : La formule des probabilités totales

On dispose d'un dé cubique équilibré (non truqué) et de trois urnes numérotées 1, 2 et 3 contenant chacune cinq boules rouges ou noires. Le jeu consiste à lancer le dé puis à tirer une boule de la façon suivante :
Si on obtient 1 on tire une boule de l'urne 1 qui contient 2 boules noires et 3 boules rouges.
Si on obtient 3 ou 5 on tire une boule de l'urne 2 qui contient 3 boules noires et 2 boules rouges.
Si on obtient 2, 4 ou 6 on tire une boule de l'urne 3 qui contient 4 boules noires et 1 boule rouge.
Déterminer la probabilité d'obtenir une boule rouge.

Indications
On désigne respectivement par U1, U2, U3 les événements « on tire une boule de l'urne 1 », « on tire une boule de l'urne 2 », « on tire une boule de l'urne 3 » et par R l'événement « la boule tirée est rouge ».
On a donc $P(U1) = \frac{1}{6}$, $P(U2) = \frac{1}{3}$, $P(U3) = \frac{1}{2}$ *et*
$P_{U1}(R) = \frac{3}{5}$, $P_{U2}(R) = \frac{2}{5}$, $P_{U3}(R) = \frac{1}{5}$



$P(R) = P(R \cap U1) + P(R \cap U2) + P(R \cap U3)$

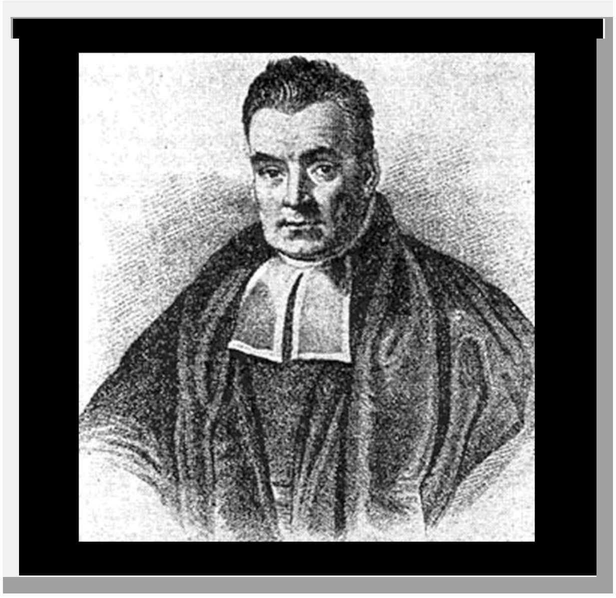
$P(R) = P_{U1}(R) \times P(U1) + P_{U2}(R) \times P(U2) + P_{U3}(R) \times P(U3)$.

$P(R) = \frac{1}{6} \times \frac{3}{5} + \frac{1}{3} \times \frac{2}{5} + \frac{1}{2} \times \frac{1}{5} = \frac{1}{3}$

B – Inférence bayésienne

1) Formule de Bayes - Inverser le conditionnement

Thomas Bayes (1702-1761/UK)



(d'après https://www.mathemathieu.fr/art/articles-maths/64-thm-bayes)

Thomas Bayes est un mathématicien britannique et pasteur de l'Église presbytérienne, connu pour avoir formulé le théorème de Bayes.
Ses découvertes en probabilités ont été résumées dans son ouvrage « Essais sur la manière de résoudre un problème dans la doctrine des risques » (Essay TowardsSolving a Problem in the Doctrine of Chances - 1763) publié à titre posthume dans les comptes-rendus de l'Académie royale de Londres (the Philosophical Transactions of the Royal Society of London). Ce résultat est très utilisé en classification automatique. Un exemple parmi d'autres est la lutte contre le spam, par la méthode dite d'inférence bayésienne.

Premier cas : A et B sont deux événements de probabilité non nulle.

On a par définition $P_A(B) = \frac{P(A \cap B)}{P(A)}$ et de même $P_B(A) = \frac{P(A \cap B)}{P(B)}$

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

D'où la formule de Bayes $\boldsymbol{P_B(A) = \frac{P_A(B) \times P(A)}{P(B)}}$

Elle permet d'obtenir $P_B(A)$ appelée probabilité a posteriori, lorsqu'on connait la probabilité dite a priori $P_A(B)$ et les probabilités de A et de B.

Cas général :
Soit B un événement de l'univers E et $A_1, A_2, \ldots A_n$ une partition de l'univers E.
On connait $P_{A_i}(B)$ pour tout i entier naturel de 1 à n et on cherche $P_B(A_i)$

$$P_B(A_i) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i) \times P_{A_i}(B)}{P(B)}$$
$$= \frac{P(A_i) \times P_{A_i}(B)}{P_{A_1}(B) \times P(A_1) + P_{A_2}(B) \times P(A_2) + \cdots P_{A_n}(B) \times P(A_n)}$$

en utilisant la formule des probabilités totales au dénominateur.

On obtient la **formule de Bayes**

$$\boldsymbol{P_B(A_i) = \frac{P(A_i) \times P_{A_i}(B)}{\sum_{i=1}^{n} P_{A_i}(B) \times P(A_i)}}$$

2) Exemples d'application

Exemple 1 : Paradoxe des tests de dépistage (d'après https://www.mathemathieu.fr/art/articles-maths/64-thm-bayes)

Une maladie est présente dans la population, dans la proportion d'une personne malade sur 10000. Un responsable d'un grand laboratoire pharmaceutique vient vanter son nouveau test de dépistage : si une personne est malade, le test est positif à 99%. Si une personne n'est pas malade, le test est positif à 0,1%.
Ces chiffres ont l'air excellent. Toutefois, ce qu'on cherche à connaître ce n'est pas vraiment les résultats présentés par le laboratoire, c'est la probabilité qu'une personne soit malade si le test est positif. La formule de Bayes permet de calculer cette probabilité.
On note M l'événement : "La personne est malade", et T l'événement : "Le test est positif".
Le but est de calculer $P_T(M)$.

On sait que $P(M) = 0{,}0001$ donc $P(\overline{M})=0{,}9999$; $P_M(T) = 0{,}99$ et $P_{\overline{M}}(T) = 0{,}001$.
La formule de Bayes s'écrit : $P(T) = P_M(T) \times P(M) + P_{\overline{M}}(T) \times P(\overline{M})$
D'où $P(T) = 0{,}99\times10^{-4}+10^{-3}\times0{,}9999=1{,}0989\times10^{-3}$

$P_T(M) = \frac{P(T\cap M)}{P(T)}$ avec $P(M \cap T)=10^{-4}\times 0{,}99$ D'où $P_T(M) \approx 0{,}09$.

Ce n'est pas satisfaisant ! Il n'y a que 9% de chances qu'une personne positive au test soit effectivement malade ! C'est tout le problème des tests de dépistage pour des maladies rares : ils doivent être excessivement performants, sous peine de donner beaucoup trop de "faux-positifs".

Un énoncé alternatif testé en classe

Paradoxe des tests de dépistage, inférence Bayésienne.

Une maladie est présente dans la population, dans la proportion d'une personne malade sur 10000.

Un responsable d'un grand laboratoire pharmaceutique vient vanter son nouveau test de dépistage : Si une personne est malade, le test est positif à 99% (on dit que la sensibilité du test est de 99 %). Si une personne n'est pas malade, le test est positif à 0,1%. Ces chiffres ont l'air excellent.

Toutefois, ce qu'on cherche à connaître ce ne sont pas vraiment les résultats présentés par le laboratoire, mais plutôt la probabilité qu'une personne soit malade si le test est positif.

La formule de Bayes permet de calculer cette probabilité. On note M l'événement : "La personne est malade", et T l'événement : "Le test est positif".

1. Calculer la probabilité pour une personne ayant un test positif, d'être malade, c'est-à dire $P_T(M)$ (appelée valeur prédictive positive). Interpréter le résultat précédent. Que penser de l'efficacité de ce nouveau test de dépistage vanté par le laboratoire ?

2. Calculer la probabilité pour une personne ayant un test négatif, de ne pas être malade, c'est-à dire $P_{\overline{T}}(\overline{M})$ (appelée valeur prédictive négative). Interpréter le résultat.

3. Quelle(s) conclusion(s) tirer des résultats trouvés ?

Un deuxième énoncé alternatif proposé en évaluation

Les résultats seront donnés sous forme décimale en arrondissant à 10^{-4}.

Dans un pays, il y a 2 % de la population contaminée par un virus.

On dispose d'un test de dépistage de ce virus qui a les propriétés suivantes :

La probabilité qu'une personne contaminée ait un test positif est de 0, 99 (sensibilité du test).

La probabilité qu'une personne non contaminée ait un test négatif est de 0, 97 (spécificité du test). On fait passer un test à une personne choisie au hasard dans cette population. On note V l'évènement « la personne est contaminée par le virus » et T l'évènement « le test est positif ». $\overline{V}$ et $\overline{T}$ désignent respectivement les évènements contraires de V et T.

1. (a) Préciser les valeurs des probabilités $P(V)$, $P_V(T)$, $P_{\overline{V}}(\overline{T})$. Traduire la situation à l'aide d'un arbre de probabilités.
(b) En déduire la probabilité de l'évènement $V \cap T$.

2. Démontrer que la probabilité que le test soit positif est 0,0492.

3. (a) Justifier, par un calcul, la phrase : « Si le test est positif, il n'y a qu'environ 40 % de « chances » que la personne soit contaminée ».
(b) Déterminer la probabilité qu'une personne ne soit pas contaminée par le virus sachant que son test est négatif.

Exemple 2 : Hospitalisations et covid

(d'après https://www.arsouyes.org/blog/2020/15_COVID19_Inference_Bayesienne/)
Le tableau suivant donne, **au 31 mars 2020**, le nombre d'hospitalisations par tranche d'âge, ainsi que leur proportion en pourcentage.

Classe d'âge	Hospitalisations	Proportion
0-14	81	0,4 %
15-44	1799	8 %
45-64	6811	30 %
65-74	5479	25 %
75+	8241	37 %
Total	22411	

Remarque : compte-tenu des arrondis, la somme des proportions n'est pas exactement 100%

On constate d'après ce tableau que

- plus d'une hospitalisation sur 3 concerne un patient de plus de 75 ans,
- près de 40% des malades hospitalisés ont moins de 65 ans,
- un peu moins d'une hospitalisation sur 10 concerne un malade de moins de 45 ans.

Même si ces phrases sont vraies, elles ne nous apportent pas l'information qu'on aimerait trouver c'est-à-dire le risque, pour nous, d'être hospitalisé. Pour le trouver, il est tentant d'inférer avec les données dont on dispose de **mauvaises conclusions** par exemple :
Après 75 ans, on a *une chance sur 3* d'être hospitalisé.

Ce qu'on cherche, c'est la probabilité d'être hospitalisé (H) sachant que notre âge est dans un intervalle (A), soit $P_A(H)$, la condition est en sens inverse par rapport au cas précédent et cela a toute son importance : **Ces deux probabilités ne sont pas égales,** on ne travaille pas avec la même population.
Si on désigne par « +75 » l'événement « avoir plus de 75 ans » et H l'événement « être hospitalisé », ce qu'on a obtenu avec le tableau était $P_H(+75)=\frac{8241}{22411}\approx 0{,}37$ et ce qu'on voudrait connaître c'est $P_{+75}(H)$.

Heureusement, on peut quand *même passer de l'une à l'autre, en utilisant l'Inférence Bayésienne.* On sait que $P_H(+75)=\frac{P(H\cap +75)}{P(H)}$ et que $P_{+75}(H)=\frac{P(H\cap +75)}{P(+75)}$.

Pour pouvoir faire les calculs corrects il manque des données que fournit le tableau suivant :

Age	Population	Hospitalisation	Incidence / 100000
0-14	11943747	81	0,7
15-44	23972387	1799	7,5
45-64	17396991	6811	39,2
65-74	7377042	5479	74,3
75+	6373536	8241	129,3
Total	67063703	22411	33,4

$P(H)=\frac{22411}{67063703}\approx 0{,}00033$ et $P(+75)=\frac{6373536}{67063703}\approx 0{,}09504$

On en déduit que $P(H\cap +75)=0{,}37\times 0{,}00033=0{,}0001221$

d'où $P_{+75}(H)=\frac{P(H\cap +75)}{P(+75)}=\frac{0{,}0001221}{0{,}09504}=0{,}00128$ ce qui signifie que la probabilité pour une personne de plus de 75 ans d'être hospitalisée est en réalité inférieure à 2 pour 1000.

Cette erreur de raisonnement est due à plusieurs biais cognitifs connus : le biais d'appariement qui nous pousse à chercher une réponse en n'utilisant seulement les éléments qu'on a sous les yeux et réciproquement l'oubli de la fréquence de base qui nous occulte ces probabilités individuelles. En ne regardant que les fréquences au sein des hôpitaux, on se trompe.

Application : calculer un autre risque d'être hospitalisé, par exemple pour les moins de 44 ans

Un énoncé alternatif testé en classe

(d'après https://www.arsouyes.org/blog/2020/15_COVID19_Inference_Bayesienne/)

Le tableau suivant donne, **au 31 mars 2020**, le nombre d'hospitalisations par tranche d'âge, ainsi que leur proportion en pourcentage.

Classe d'âge	Hospitalisations	Proportion
0-14	81	0,4 %
15-44	1799	8 %
45-64	6811	30 %
65-74	5479	25 %
75+	8241	37 %
Total	22411	

Remarque : compte-tenu des arrondis, la somme des proportions n'est pas exactement 100%

1. Compléter par vrai ou faux :

On constate d'après ce tableau que	Vrai	Faux
Plus d'une hospitalisation sur 3 concerne un patient de plus de 75 ans,		
Un peu moins d'une hospitalisation sur 10 concerne un malade de moins de 45 ans		
Près de 40% des malades hospitalisés ont moins de 65 ans		
Après 75 ans, on a « une chance sur trois » d'être hospitalisé		

On désigne par « +75 » l'événement « avoir plus de 75 ans » et H l'événement « être hospitalisé »,

2. A quelle probabilité correspond le nombre 37% du tableau précédent ?

3. Nous aimerions trouver le risque, pour une personne de plus de 75 ans, d'être hospitalisée, exprimer le sous forme de probabilité et la calculer si possible.

4. Modéliser la situation, à l'aide d'un arbre.

On dispose maintenant des informations suivantes :

Age	Population	Hospitalisation	Incidence d'hospitalisation / 100000
0-14	11943747	81	0,7
15-44	23972387	1799	7,5
45-64	17396991	6811	39,2
65-74	7377042	5479	74,3
75+	6373536	8241	129,3
Total	67063703	22411	33,4

5. En déduire la probabilité demandée à la question 3.

> Parfois nous commettons des erreurs de raisonnement est dues à plusieurs biais cognitifs connus :le biais d'appariement qui nous pousse à chercher une réponse en n'utilisant seulement les éléments qu'on a sous les yeux et réciproquement l'oubli de la fréquence de base qui nous occulte ces probabilités individuelles. En ne regardant que les fréquences au sein des hospitalisations, on se trompe.

Exercice 3 : **Une compagnie d'assurance** (questions corrigées au fur et à mesure)

Une compagnie d'assurance estime que les gens peuvent être répartis en deux classes :

- Ceux qui sont enclins aux accidents « à haut risque » soit 30% de la population
- Les autres dits « à risque modéré »

Les statistiques de cette compagnie montrent qu'un individu à haut risque a une probabilité de 0,40 d'avoir un accident en l'espace d'un an ; cette probabilité tombe à 0,20 pour les gens à risque modéré.
1) Quelle est la probabilité qu'un nouvel assuré soit victime d'un accident durant l'année qui suit la signature de son contrat ?

1) On définit les évènements :
- *B : le nouvel assuré a un accident durant l'année qui suit la signature de son contrat*
- *A_1 : le nouvel assuré est « à haut risque »*
- *A_2 : le nouvel assuré est « à risque modéré »*

En appliquant la formule des probabilités totales, on trouve
$P(B)=0,4\times0,3+0,2\times0,7=0,26$

2) Un nouveau signataire a un accident dans l'année qui suit la signature de son contrat. Quelle est la probabilité qu'il fasse partie de la classe « haut risque » ?

$P_B(A_1)=\frac{P(B\cap A1)}{P(B)}$ or $P(B \cap A1)=P_{A1}(B)\times P(A_1)= 0,4\times0,3=0,12$ d'où $P_B(A_1)=\frac{0,12}{0,26}=\frac{6}{13}$

Remarque : dans la pratique on est amené à mélanger des démarches de probabilités et de statistiques (cf. annexe)

Enoncé alternatif, donné en évaluation en classe

20 % des assurés d'une compagnie d'assurance sont jeunes conducteurs (J), les autres étant conducteurs expérimentés.

Une étude statistique indique que la probabilité qu'un assuré jeune conducteur ait un sinistre responsable (S) au cours de l'année est de 10 %, contre 6 % pour les assurés expérimentés.

a) Faire un arbre pondéré illustrant la situation.

b) Calculer $p_S(J)$, c'est-à dire la probabilité qu'une personne ayant eu un sinistre au cours de l'année soit un jeune conducteur

3) Le cas particulier des tests sur un exemple : deux tests médicaux successifs pour affiner un diagnostic

On s'intéresse à une maladie qui touche certains individus d'une population. Une étude a permis de déterminer la **prévalence** de cette maladie, c'est-à-dire la probabilité « a priori » d'être malade, estimée ici, à 1‰ ou 0,1%. Si on appelle M l'événement « être malade », $P(M) = \frac{1}{1000} = 0{,}001$.

Partie A - Un premier test

On étudie un premier test, destiné à savoir si la personne testée est malade ou non.
On a pu mesurer, en s'aidant d'un autre test, que sur cent malades, 98 avaient un résultat positif : on dit que la **sensibilité** du test est de 98%. Autrement dit, la probabilité d'avoir un test positif sachant qu'on est malade est égale à 0,98. Si on appelle t l'événement « le test est positif », $s_e = P_M(t) = 0{,}98$.
On a également évalué que sur cent personnes non malades (ou saines), 99 avaient un test négatif : on dit que la **spécificité** du test est de 99%, ce qui signifie que la probabilité d'avoir un test négatif sachant qu'on n'est pas malade est égale à 0,99.
$s_p = P_{\overline{M}}(\bar{t}) = 0{,}99$ ($\overline{M}$ est l'évènement « ne pas être malade » et $\bar{t}$ « le test est négatif »).

1) Qu'est-ce qu'on cherche? Une personne qui a eu un test positif veut connaitre ses « chances » (ou « risques ») d'être malade. Il faut donc déterminer la probabilité que cette personne ait la maladie sachant que son test est positif, appelée aussi **valeur prédictive positive** ou encore probabilité a posteriori d'être malade. Sans calcul, diriez-vous que cette probabilité est égale à : 0,98 (on peut aussi écrire 98 %) ? 70 % ? 50 % ? moins de 30 % ? moins de 10% ?

Dans les questions suivantes, on va calculer de plusieurs manières cette probabilité afin de répondre à la première question en termes de probabilité.

2) A partir d'un tableau

On suppose que la population est de 100 000 personnes. Compléter le tableau ci-contre, appelé **tableau de contingence**.

	$\overline{M}$	M	Total
$\bar{t}$			
t			
Total	99 900	100	100 000

a) En déduire la probabilité d'être malade sachant qu'on a un test positif. Donner le résultat en pourcentage arrondi au dixième. Interpréter le résultat en le comparant à la réponse à la question 1).
b) Parmi les personnes qui ont un résultat positif, quel est le pourcentage de **faux-positifs**, c'est-à-dire de personnes qui ont un résultat positif au test alors qu'elles ne sont pas malades ? Interpréter le résultat.

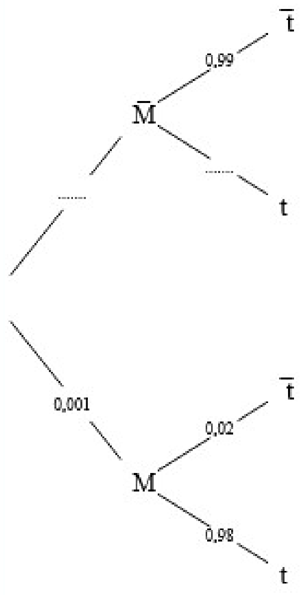
3) A partir d'un arbre

a) Compléter l'arbre ci-contre.

b) Calculer $P(M \cap t)$ et $P(t)$. En déduire $P_t(M)$.

4) En utilisant la formule de Bayes

a) Calculer $P(t)$ en utilisant la formule des probabilités totales.
b) En appliquant la formule de Bayes $P_t(M) = \frac{P_M(t) \times P(M)}{P(t)}$, retrouver la probabilité qu'un individu soit malade sachant que son test est positif.



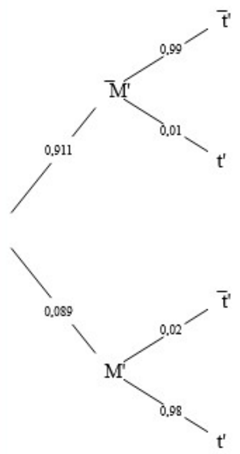
Partie B - Deuxième utilisation de la formule, pour préciser.

Comme on a vu qu'une grande partie des personnes qui ont eu un test positif ne sont en réalité pas malades, on va procéder à un deuxième test afin de mieux cibler les personnes malades.
On change donc la population initiale, en ne considérant cette fois que les individus ayant un premier test positif. On appelle M' l'événement « être malade et positif au 1er test» et t' l'événement « être positif au 2ème test ». La probabilité initiale (a priori) est donc maintenant $P(M') = P_t(M)$=0,089 ou 8,9%, c'est-à-dire la probabilité trouvée a posteriori dans le calcul précédent.
On suppose que la sensibilité et la spécificité de ce nouveau test sont les mêmes que celles du premier test, soit respectivement 98% et 99%. Donc si on appelle t' l'événement « le deuxième test est positif » on a,

$$P_{M'}(t') = 0{,}98 \text{ et } P_{\overline{M'}}(\bar{t}') = 0{,}99$$

a) Justifier l'égalité $P(\overline{M'}) = 0{,}911$

b) A l'aide de l'arbre ci-dessous et de la formule de Bayes, calculer $P_{t'}(M')$ et interpréter le résultat.



Indications

1) Comparer l'estimation donnée avec le résultat de la question 3) c). On a tendance à surestimer cette probabilité. Attention à ne pas la confondre avec la sensibilité…

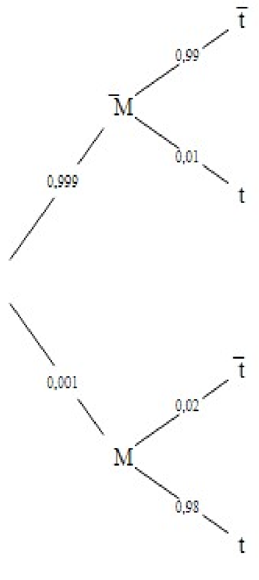
2)

	$\overline{M}$	M	*Total*
$\bar{t}$	*98 901*	*2*	*98 903*
t	*999*	*98*	*1 097*
Total	*99900*	*100*	*100 000*

a) On lit : il y a 98 malades parmi les 1097 personnes testées positives donc $P_t(M) = \frac{98}{1097} \approx 0{,}089$ *soit 8,9% ce qui est très faible, moins de 10%. Etant donné que la sensibilité du test est grande, on pouvait s'attendre à une proportion de malades bien plus importante.*

b) On a 999 faux-positifs sur 1097 testés positifs soit environ 91% : une grande majorité des personnes qui ont un test positif sont en bonne santé et vont donc s'inquiéter inutilement…. Cela s'explique par le fait que la prévalence est faible.

3) a) On complète l'arbre :



b) $P(M \cap t) = 0{,}001 \times 0{,}98 = 0{,}00098$.
De même $P(\overline{M} \cap t) = 0{,}999 \times 0{,}01 = 0{,}00999$.

D'où $P(t) = P(t \cap M) + P(t \cap \overline{M}) = 0{,}00098 + 0{,}00999$.
$P(t) = 0{,}01097$
On en déduit :
$P_t(M) = \frac{P(t \cap M)}{P(t)} = \frac{0{,}00098}{0{,}01097} \approx 0{,}089$. *On retrouve 8,9%.*

4) a) $P_{\overline{M}}(t) = 1 - P_{\overline{M}}(\overline{t}) = 0{,}01$
$P(t) = P_M(t) \times P(M) + P_{\overline{M}}(t) \times P(\overline{M})$
$P(t) = 0{,}98 \times 0{,}001 + 0{,}01 \times 0{,}999 = 0{,}01097$

b) On en déduit $P_t(M) = \frac{P_M(t) \times P(M)}{P(t)} = \frac{0{,}98 \times 0{,}001}{0{,}01097} \simeq 0{,}089$
La probabilité pour un individu d'être malade sachant que son test est positif est bien d'environ 8,9% ce qui est très faible.

Partie B
a) $P(\overline{M'}) = 1 - 0{,}089 = 0{,}911$
On peut utiliser la formule des probabilités totales pour calculer P(t')
$P(t') = P_{M'}(t') \times P(M') + P_{\overline{M'}}(t') \times P(\overline{M'}) = 0{,}089 \times 0{,}98 + 0{,}911 \times 0{,}01 = 0{,}09633$

Et la formule de Bayes nous donne ensuite $P_{t'}(M') = \frac{0{,}089 \times 0{,}98}{0{,}09633} \approx 0{,}9054$

Un individu ayant eu les deux tests positifs a donc environ 90,5 % de chances d'être malade. Il a donc plus de 9 chances sur 10 d'être malade et il est normal qu'il s'inquiète ! On avait 8,9% de malades parmi les positifs au 1^er^ test et maintenant on a 90,5% de malades parmi les positifs aux deux tests donc un résultat beaucoup plus fiable. Le deuxième test a permis de multiplier par plus de 10 la probabilité d'être malade.

Remarque : on trouvera en annexe un récapitulatif des termes utilisés dans les tests médicaux.

5) Exercices d'application

Exercice 1 : Dopage et tests.

Un groupe de 500 coureurs participe à une course cycliste au cours de laquelle aucun abandon n'est constaté. D'après des statistiques, on admet qu'en moyenne, un coureur sur 20 est dopé.

Partie 1
A l'issue de la course un premier test antidopage est pratiqué à l'aide d'une bandelette urinaire. Pour un coureur on appelle T l'événement « le coureur est positif au test urinaire» et D l'évènement « le coureur est dopé ».
Le contrôle anti-dopage n'étant pas fiable à 100 %, on sait que
– si un coureur est dopé, le contrôle est positif dans 92 % des cas (on dit que la sensibilité du test est de 92%)

– si un coureur n'est pas dopé, le contrôle est négatif dans 96 % des cas (on dit que la spécificité du test est de 96%)

1. Compléter le tableau de contingence ci-dessous :

	Nombre de coureurs dopés	Nombre de coureurs non dopés	Total
Nombre de coureurs positifs au test			
Nombre de coureurs négatifs au test			
Total			500

2. Un coureur a un contrôle positif. Quelle est la probabilité qu'il soit dopé ? Interpréter le résultat.

Partie 2
Le premier test pratiqué n'étant pas entièrement satisfaisant, on procède pour les coureurs positifs au test urinaire à un second test (une prise de sang) dont la sensibilité (probabilité d'avoir un test sanguin positif sachant que le coureur est dopé) et la spécificité (probabilité d'avoir un test sanguin négatif sachant que le coureur n'est pas dopé) sont égales respectivement à 90% et 95%.
Déterminer la probabilité pour qu'un coureur déclaré positif avec le second test soit dopé.

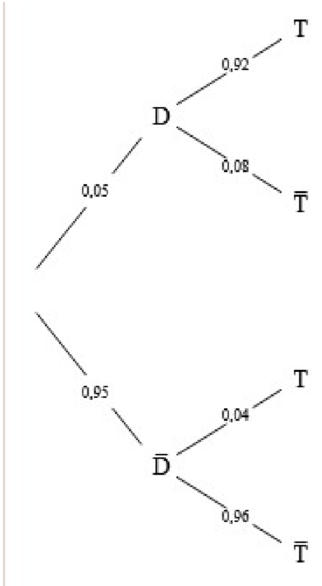
Indications
Partie 1
1. Tableau de contingence : on indique les calculs faits dans chaque case.
1 coureur sur 20 se traduit par $\frac{1}{20} = 0{,}05$ *d'où le premier calcul du nombre de coureurs dopés* $500 \times 0{,}05 = 25$

	Nombre de coureurs dopés	*Nombre de coureurs non dopés*	*Total*
Nombre de coureurs avec un test positif	*23 (25×0,92=23) (vrais positifs)*	***19*** *(42-23=19) (faux positifs)*	***42***
Nombre de coureurs avec un test négatif	*2 (25-23=2) (faux négatifs)*	*456 (458-2=456) (vrais négatifs)*	*458 (500-42=458)*
Total	*25 (500×0,05=25)*	*475 (500-25=475)*	*500 (donné)*

2. $P_T(D) = \frac{23}{42} \approx 55\%$

On peut aussi raisonner avec un arbre de probabilités.
La formule des probabilités totales donne :

$$P(T) = 0{,}05 \times 0{,}92 + 0{,}95 \times 0{,}04 = 0{,}084$$

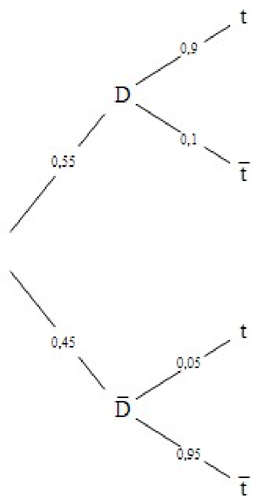


On a alors $P_T(D) = \frac{P(D\cap T)}{P(T)} = \frac{0{,}05\times 0{,}92}{0{,}05\times 0{,}92+0{,}95\times 0{,}04} \approx 0{,}55$

Ainsi, parmi tous les coureurs déclarés positifs au test, il ne sont que 55% à être dopés, autrement dit presque la moitié des coureurs ayant un test positif seront soupçonnés à tort de dopage ! Ce test n'est pas très fiable.

Partie 2

On désigne par t le test sanguin est positif et $\overline{t}$ *le test sanguin est négatif.*



On peut raisonner avec un arbre de probabilité :

Pour calculer $P_t(D)$, *on calcule d'abord*

P(t) = $0{,}55 \times 0{,}90$*+0,45×0,05=0,52.*

D'où $P_t(D) = \frac{0{,}55\times 0{,}90}{0{,}52}$ *=0,95.*

Donc si le test sanguin t est positif, le coureur a 95% de chances d'être dopé.
Avec le test urinaire, on avait 55% de dopés parmi les positifs ;
avec les deux tests successifs, on n'a que 5% de chance de se tromper si on accuse de dopage un coureur qui a eu les deux résultats positifs.

Exercice 2 Parents et enfants

Lors d'une enquête réalisée auprès d'élèves de classes de Terminale, on apprend que 60% des élèves sont des filles. De plus, 40% des filles et 30% des garçons fument.

1. On choisit un élève au hasard. Quelle est la probabilité que l'élève choisi soit :

   a) un garçon ? b) une fille qui fume ? c) un garçon qui fume ?

2. On note A l'événement « l'élève choisi fume » et F l'événement « l'élève choisi est une fille » et P(A) et P(F) leurs probabilités.

Calculer P(F) et définir les événements $\bar{F}$, A ∩ F, A ∩ $\bar{F}$.

3. Déduisez des questions précédentes, en le justifiant, que P(A) = 0,36.

4. L'enquête permet de savoir que parmi les élèves fumeurs, la moitié ont des parents qui fument et que parmi les élèves non-fumeurs, 65 % ont des parents non-fumeurs.

On note B l'événement « l'élève choisi a des parents fumeurs ».

Dans les questions qui suivent, on arrondira éventuellement les résultats au millième.

a) Calculer les probabilités P(A ∩ B) et P($\bar{A}$ ∩ B). En déduire P(B).

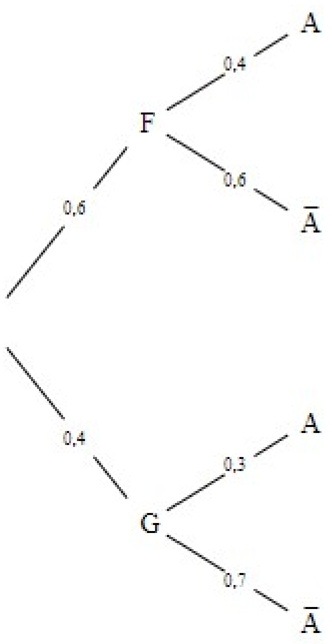
b) Calculer P_B(A), probabilité qu'un élève fume sachant qu'il a des parents fumeurs.

c) Calculer$P_{\bar{B}}$(A), probabilité qu'un élève fume sachant qu'il a des parents non-fumeurs. Que peut-on en conclure ?

<u>Indications</u>

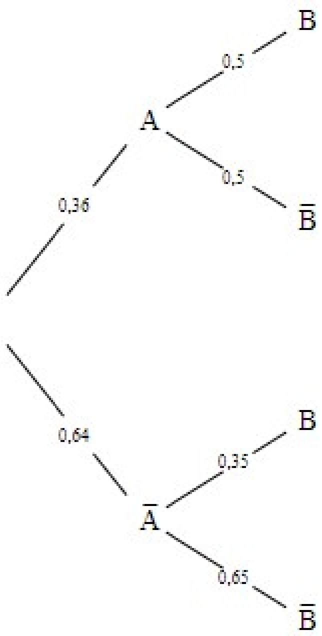
1. $P(G) = 0{,}4$ $\quad P_A(F) = 0{,}4$ $\quad P_A(G) = 0{,}3$

2. On établit l'arbre de probabilité



3. On a donc $P(A) = 0{,}6 \times 0{,}4 + 0{,}4 \times 0{,}3 = 0{,}36$

4. On établit un deuxième arbre



a) $P(A \cap B) = 0{,}36 \times 0{,}5 = 0{,}18$

$$P(\bar{A} \cap B) = 0{,}64 \times 0{,}35 = 0{,}224$$

$$P(B) = 0{,}18 + 0{,}224 = 0{,}404$$

b) $P_B(A) = \frac{P(A \cap B)}{P(B)} = \frac{0,18}{0,404} \approx 0,446$

(valeur arrondie au millième)

c) $P_{\bar{B}}(A) = \frac{P(A \cap \bar{B})}{P(\bar{B})}$; $P(\bar{B}) = 1 - P(B) = 0,596$

et $P(A \cap \bar{B}) = 0,36 \times 0,5 = 0,18$

Donc $P_{\bar{B}}(A) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{0,18}{0,596} \approx 0,302$ (valeur arrondie au millième)

On peut donc en conclure qu'un élève a moins de chances d'être fumeur si ses parents ne le sont pas.

Exercice 3 : Une association sportive (D'après bac Amérique du Nord 2012)

Dans une association sportive, un quart des femmes et un tiers des hommes adhèrent à la section tennis. On sait de plus que 30 % des membres adhèrent à la section tennis. On choisit au hasard un membre de cette association et on note :

– F l'événement « le membre choisi est une femme »,

– T l'événement « le membre choisi adhère à la section tennis ».

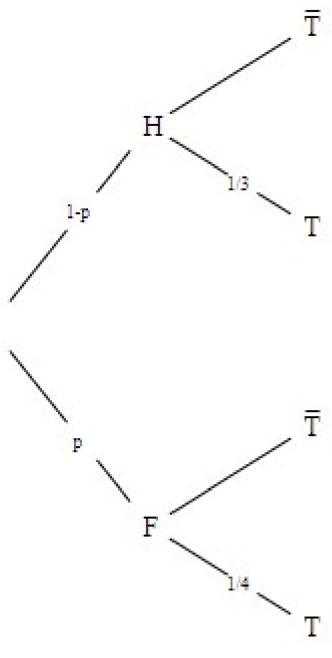
1. On note *p* la proportion de femmes dans l'association. Etablir un arbre de probabilités et montrer que la probabilité de l'événement F est égale à $\frac{2}{5}$.

2. On choisit un membre parmi les adhérents à la section tennis. Quelle est la probabilité que ce membre soit une femme ?

Indications

1) On note p la proportion de femmes de l'association et donc la probabilité de rencontrer une femme à la sortie de l'association.

D'où l'arbre de probabilité :



On sait que la proportion d'adhérents qui sont dans la section tennis est 0,30.

On a donc $\frac{1}{4} \times p + \frac{1}{3} \times (1-p) = 0,3$

En résolvant l'équation on trouve $p = 0,4 = \frac{2}{5}$

2) $P_T(F) = \frac{P_F(T) \times P(F)}{P(T)} = \frac{p \times \frac{1}{4}}{0,3}$

Or $p = 0,4 = \frac{2}{5}$ *d'où* $P_T(F) = \frac{1}{3}$

Exercice 4 (d'après Bac Métropole 2011) **Dépistage d'un virus**.

Les résultats seront donnés sous forme décimale en arrondissant à 10^{-4}.
Dans un pays, il y a 2 % de la population contaminée par un virus.
On dispose d'un test de dépistage de ce virus qui a les propriétés suivantes :
La probabilité qu'une personne contaminée ait un test positif est de 0, 99 (sensibilité du test).
La probabilité qu'une personne non contaminée ait un test négatif est de 0, 97 (spécificité du test). On fait passer un test à une personne choisie au hasard dans cette population. On note V l'évènement « la personne est contaminée par le virus » et T l'évènement « le test est positif ».
$\overline{V}$ et $\overline{T}$ désignent respectivement les évènements contraires de V et T .
1. (a) Préciser les valeurs des probabilités $P(V)$, $P_V(T)$, $P_{\overline{V}}(\overline{T})$. Traduire la situation à l'aide d'un arbre de probabilités.
(b) En déduire la probabilité de l'évènement $V \cap T$.
2. Démontrer que la probabilité que le test soit positif est 0,0492.
3. (a) Justifier, par un calcul, la phrase : « Si le test est positif, il n'y a qu'environ 40 % de « chances » que la personne soit contaminée ».
(b) Déterminer la probabilité qu'une personne ne soit pas contaminée par le virus sachant que son test est négatif.

Exercice 5 (Amérique du Sud 2013) **Malformation cardiaque**

En utilisant sa base de données, la sécurité sociale estime que la proportion de Français présentant, à la naissance, une malformation cardiaque de type anévrisme est de 10 %. L'étude a également permis de prouver que 30 % des Français présentant, à la naissance, une malformation cardiaque de type anévrisme, seront victimes d'un accident cardiaque au cours de leur vie alors que cette proportion n'atteint plus que 8 % pour ceux qui ne souffrent pas de cette malformation congénitale. On choisit au hasard une personne dans la population française et on considère les évènements :

M : « La personne présente, à la naissance, une malformation cardiaque de type anévrisme »
C : « La personne est victime d'un accident cardiaque au cours de sa vie »

1. (a) Montrer que $P(M \cap C) = 0,03$
(b) Calculer $P(C)$.

2. On choisit au hasard une victime d'un accident cardiaque. Quelle est la probabilité qu'elle présente une malformation cardiaque de type anévrisme ?

Exercice 6 (d'après Bac Asie 2013) **Boites de thé**

Dans cet exercice, les probabilités seront arrondies au centième.

Un grossiste achète des boîtes de thé vert chez deux fournisseurs. Il achète 80 % de ses boîtes chez le fournisseur A et 20 % chez le fournisseur B. 10 % des boîtes provenant du fournisseur A présentent des traces de pesticides et 20 % de celles provenant du fournisseur B présentent aussi des traces de pesticides. On prélève au hasard une boîte du stock du grossiste et on considère les évènements suivants :

- évènement A : « la boîte provient du fournisseur A » ;
- évènement B : « la boîte provient du fournisseur B » ;
- évènement S : « la boîte présente des traces de pesticides ».

1. Traduire l'énoncé sous forme d'un arbre pondéré.
2. (a) Quelle est la probabilité de l'évènement B∩S ?
(b) Justifier que la probabilité que la boîte prélevée ne présente aucune trace de pesticides est égale à 0,88.
3. On constate que la boîte prélevée présente des traces de pesticides. Quelle est la probabilité que cette boîte provienne du fournisseur B ?

Exercice 7 : (d'après Maths-France/Math Sup http ://www.maths-france.fr)

On dispose de 100 dés dont 25 sont pipés. Pour chaque dé pipé, la probabilité d'obtenir le chiffre 6 lors d'un lancer vaut $\frac{1}{2}$.

a) On tire un dé au hasard parmi les 100 dés. On lance ce dé et on obtient le chiffre 6. Quelle est la probabilité que ce dé soit pipé ?

b) Soit n ∈ N∗. On tire un dé au hasard parmi les 100 dés. On lance ce dé *n* fois et on obtient *n* fois le chiffre 6. Quelle est la probabilité p_n que ce dé soit pipé ?

c) Déterminer $\lim\limits_{n\to+\infty} p_n$. Interpréter ce résultat.

Indications

a) Notons A l'événement « le dé est pipé » et B l'événement « on obtient le chiffre 6 ».
La probabilité demandée est $P_B(A)$. Or A et $\overline{A}$ forment une partition de l'univers.
On a $P(A) = \frac{25}{100} = \frac{1}{4} \neq 0$ et $P(\bar{A}) = 1 - \frac{1}{4} = \frac{3}{4}$. Ensuite $P_A(B) = \frac{1}{2}$ et $P_{\bar{A}}(B) = \frac{1}{6}$.
Donc, $P(B) = P(A) \times P_A(B) + P(\bar{A}) \times P_{\bar{A}}(B) = \frac{1}{4} \times \frac{1}{2} + \frac{3}{4} \times \frac{1}{6} = \frac{1}{4}$ résultat non nul.
D'après la formule de Bayes, $P_B(A) = \frac{\mathrm{P}(A) \times P_A(B)}{P(A) \times P_A(B) + P(\bar{A}) \times P_{\bar{A}}(B)} = \frac{\frac{1}{4} \times \frac{1}{2}}{\frac{1}{4}} = \frac{1}{2}$.
La probabilité que ce dé soit pipé est $\frac{1}{2}$.

b) Notons A l'événement « le dé est pipé » et S l'événement « on obtient *n* fois le chiffre 6 ».
La probabilité demandée est $P_S(A)$.
$A, \overline{A}$ est un système complet d'événements.
On a toujours $P(A) = \frac{1}{4}$ (résultat non nul) et $(\bar{A}) = \frac{3}{4}$.
$P_A(S) = \frac{1}{2^n}$ et $P_{\bar{A}}(S) = \frac{1}{6^n}$.
Donc, $P(S) = P(A) \times P_A(S) + P(\bar{A}) \times P_{\bar{A}}(S) = \frac{1}{4} \times \frac{1}{2^n} + \frac{3}{4} \times \frac{1}{6^n}$ (résultat non nul).
D'après la formule de Bayes,
$$P_S(A) = \frac{\mathrm{P}(A) \times P_A(S)}{P(A) \times P_A(S) + P(\bar{A}) \times P_{\bar{A}}(S)} = \frac{\frac{1}{4} \times \frac{1}{2^n}}{\frac{1}{4} \times \frac{1}{2^n} + \frac{3}{4} \times \frac{1}{6^n}} = \frac{1}{1 + \frac{1}{3^{n-1}}}$$
La probabilité que ce dé soit pipé est $p_n = \frac{1}{1 + \frac{1}{3^{n-1}}}$.

c) $\lim\limits_{n\to+\infty} \frac{1}{3^{n-1}} = 0$ et donc $\lim\limits_{n\to+\infty} p_n = 1$. Ceci signifie que, si au bout d'un grand nombre de lancers, on a obtenu à chaque fois le 6, il est quasiment sûr que le dé est pipé.

Exercice 8 – Des billes dans une bouteille (D'après le travail de P. Carranza voir C)

Un professeur choisit dans une urne ne contenant que des billes orange ou noires (mais on ne sait pas dans quelle proportion) quatre billes et les place dans une bouteille opaque sans les montrer.

1) Quelles sont les compositions possibles des billes dans la bouteille ? Déterminer la probabilité de chacune d'elles.
2) Ensuite le professeur retourne deux fois la bouteille et montre la bille visible dans le goulot. Il obtient successivement une bille orange, puis une bille noire.
   a) On désigne par O_1 l'événement « le premier retournement montre une bille orange ». Justifier que $P(O_1)=0{,}5=\frac{1}{2}$.
   b) Quelles sont à présent les compositions possibles ? Justifier que les probabilités de chaque composition sachant que le premier retournement montre une bille orange (sachant O_1) sont : $P_{NNNN}(O_1) = 0$; $P_{ONNN}(O_1) = \frac{1}{4}$; $P_{OONN}(O_1) = \frac{2}{4}$; $P_{OOON}(O_1) = \frac{3}{4}$; $P_{0000}(O_1) = \frac{4}{4} = 1$.
   c) En utilisant la formule de Bayes démontrer les résultats suivants :
   $P_{O_1}(NNNN) = 0$; $P_{O_1}(ONNN) = \frac{1}{10}$; $P_{O_1}(OONN) = \frac{2}{10}$;
   $P_{O_1}(OOON) = \frac{3}{10}$; $P_{O_1}(0000) = \frac{4}{10}$.
   d) Lors du second retournement la bille apparue est noire. On désigne par N_2 l'événement « le second retournement montre une bille noire ».
   Quelles sont alors les combinaisons possibles ? Calculer la probabilité de chaque composition sachant N_2 après avoir calculé $P(N_2)$.

Indications

1) *Les cinq compositions possibles sont donc NNNN, ONNN, OONN, OOON, OOOO.*
   La probabilité de chaque composition est donc $\frac{1}{5}$ puisqu'on ne connait pas la composition de l'urne.
2) *a) La probabilité de O_1 est égale à 0,5 puisque a priori on a une chance sur 2 d'obtenir une boule orange.*
   b) Le premier retournement montre une boule orange donc l'événement O_1 sachant NNNN est impossible. La probabilité de l'événement O_1 sachant NNNO est égale à $\frac{1}{4}$ puisqu'il y a une boule orange sur 4 dans la combinaison. On raisonne de façon similaire pour la probabilité de O_1 sachant NNOO qui est égale à $\frac{2}{4}$, la probabilité de O_1 sachant NOOO qui est égale à $\frac{3}{4}$, la probabilité de O_1 sachant OOOO qui est égale à $\frac{4}{4}$ =1.

   c) La probabilité de l'événement NNNN sachant O_1 est $P_{O_1}(NNNN)$ =0

D'après la formule de Bayes, la probabilité de l'événement NNNO sachant O_1

($P_{O_1}(NNNO)$) est égale à $\frac{P_{NNNO}(O_1)\times P(NNNO)}{P(O_1)} = \frac{\frac{1}{4}\times\frac{1}{5}}{\frac{1}{2}} = \frac{1}{10}$

On raisonne de même pour les autres combinaisons sachant O_1.

$P_{O_1}(NNOO) = \frac{P_{NNOO}(O_1)\times P(NNOO)}{P(O_1)} = \frac{\frac{2}{4}\times\frac{1}{5}}{\frac{1}{2}} = \frac{2}{10}$; $P_{O_1}(NOOO) = \frac{P_{NOOO}(O_1)\times P(NOOO)}{P(O_1)} = \frac{\frac{3}{4}\times\frac{1}{5}}{\frac{1}{2}} = \frac{3}{10}$

$P_{O_1}(OOOO) = \frac{P_{OOOO}(O_1)\times P(OOOO)}{P(O_1)} = \frac{\frac{4}{4}\times\frac{1}{5}}{\frac{1}{2}} = \frac{4}{10}$

d) *Au deuxième retournement on a obtenu une boule noire.*
Les valeurs des probabilités des événements NNNN, NNNO, NNOO, NOOO, OOOO sachant O_1 sont celles obtenues précédemment.
On sait de plus que N2 sachant OOOO est impossible d'où $P_{OOOO}(N_2) = 0$.
De plus $P_{ONNN}(N_2) = \frac{3}{4}$; $P_{OONN}(N_2) = \frac{2}{4}$ *et* $P_{OOON}(N_2) = \frac{1}{4}$
Les calculs sont similaires aux calculs précédents.
$P_{O_1}(N_2) = \frac{3}{4}\times P_{O_1}(NNNO) + \frac{2}{4}\times P_{O_1}(NNOO) + \frac{1}{4}\times P_{O_1}(NOOO) =$
$\frac{3}{4}\times\frac{1}{10} + \frac{2}{4}\times\frac{2}{10} + \frac{1}{4}\times\frac{3}{10} = \frac{1}{4}$
On en déduit d'après la formule de Bayes les probabilités
$P_{N_2}(OOOO)=0$

$P_{N_2}(ONNN)=\frac{\frac{3}{4}\times\frac{1}{10}}{\frac{1}{4}} = \frac{3}{10}$

$P_{N_2}(OONN)=\frac{\frac{2}{4}\times\frac{2}{10}}{\frac{1}{4}} = \frac{4}{10}$

$P_{N_2}(OOON)=\frac{\frac{1}{4}\times\frac{3}{10}}{\frac{1}{40}} = \frac{3}{10}$

$P_{N_2}(NNNN)=0$

On peut remarquer que la probabilité la plus élevée est celle de la combinaison OONN et c'est peut-être ce que pourrait suggérer les deux tirages successifs.

Exercice 9 : Formule de Bayes et étude de fonction (Bac Antilles 2013)

On cherche à étudier le nombre d'étudiants connaissant la signification du sigle URSSAF. Pour cela, on les interroge en proposant un questionnaire à choix multiples. Chaque étudiant doit choisir parmi trois réponses possibles, notées A , B et C , la bonne réponse étant la A. On note ***r*** la probabilité pour qu'un étudiant connaisse la bonne réponse. Tout étudiant connaissant la bonne réponse répond A, sinon il répond au hasard (de façon équiprobable).

On interroge un étudiant au hasard. On note : A l'évènement « l'étudiant répond A », B l'évènement « l'étudiant répond B », C l'évènement « l'étudiant répond C », R l'évènement « l'étudiant connait la réponse », $\bar{R}$ l'évènement contraire de R.
a) Traduire cette situation à l'aide d'un arbre de probabilité.

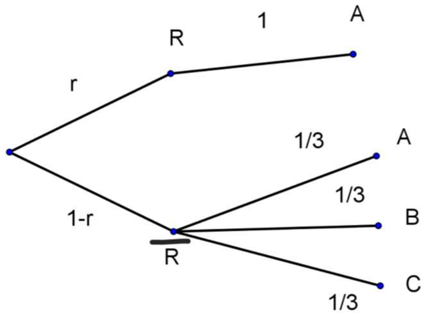
b) Montrer que la probabilité de l'évènement A est $P(A) = \frac{1}{3}(1 + 2r)$.

c) Exprimer en fonction de $\boldsymbol{r}$ la probabilité qu'une personne ayant choisi A connaisse la bonne réponse.

<u>Solution</u>

a) $P(R) = r$; $P(\bar{R}) = 1 - r$; $P_R(A) = 1$; $P_{\bar{R}}(A) = \frac{1}{3}$

b) $P(A) = P(R \cap A) + P(\bar{R} \cap A) = P_R(A) \times P(R) + P_{\bar{R}}(A) \times P(\bar{R})$



$P(A) = r + (1 - r) \times \frac{1}{3} = \frac{1}{3}(1 + 2r)$

Il s'agit de calculer $P_A(R)$

$$P_A(R) = \frac{P(A \cap R)}{P(A)} = \frac{r}{\frac{1}{3}(1+2r)} = \frac{3r}{1+2r}$$

Remarque : on sait que $r \in [0; 1]$. On vérifie qu'il en est de même pour $P_A(R)$

Partie C - Deux points de vue pour calculer des probabilités : la dualité fréquentiste/ bayésienne

Dans sa thèse de 2011 ainsi que dans son article (RDM Vol.31/2[1]), Pablo Carranza parle de « dualité » de la probabilité pour indiquer deux interprétations possibles qui coexistent depuis les origines jusqu'à nos jours.

D'une part, l'interprétation **fréquentiste :** la probabilité représente la fréquence d'apparition d'un phénomène lorsqu'on le reproduit un nombre infini de fois, comme par exemple l'apparition d'un 6 lorsqu'on lance un dé. Cette probabilité est souvent considérée comme <u>objective</u> puisque sa valeur, caractéristique de la série, est indépendante de tout observateur. La loi des grands nombres (Bernoulli 1713) constitue un énoncé fondamental pour cette interprétation : la proportion d'apparitions a pour limite une valeur fixe appelée probabilité.

D'autre part, l'interprétation **bayésienne** : la probabilité représente une mesure de certitude ou un degré de croyance portant sur une hypothèse ou sur la véracité d'une proposition. Par exemple, la vente d'avions français du type Rafale au Brésil n'étant pas confirmée à la date T, on peut supposer

---

[1] Dualité dans l'enseignement de la probabilité. Apport pour l'enseignement de la statistique.

d'après des articles de journaux qu'elle est en bonne voie et estimer sa probabilité à 0,7 ; elle est alors la traduction numérique d'un état de connaissance, ce qui met en évidence son caractère subjectif puisque la probabilité bayésienne représente une appréciation du sujet. Deux individus peuvent ainsi évaluer différemment cette probabilité et un même individu peut également modifier son évaluation s'il obtient une nouvelle information. Un énoncé fondamental pour cette approche est bien sûr le théorème de Bayes (1763) qui permet de recalculer notre mesure de certitude sur une hypothèse donnée H_i lorsque l'information D est acquise. Ainsi le passage de $P(H_i)$ à $P_D(H_i)$ est vu comme une sorte d'évolution de nos certitudes.
Il s'agit là d'une première différence entre les deux interprétations de la probabilité : dans l'interprétation fréquentiste, la valeur est fixe et caractérise la série infinie de répétitions du phénomène aléatoire ; dans l'approche bayésienne, la valeur peut varier, par exemple en fonction d'une nouvelle information disponible.
Or malgré toutes les tentatives de découpage de cette dualité, Carranza explique qu'elle se présente inévitablement dans l'enseignement et que nous devons donc en tenir compte lorsque nous enseignons les probabilités.

Carranza commence par rechercher dans l'histoire des probabilités les places des deux interprétations et constate que, dès l'aube des probabilités au 17ème siècle, on trouve cette dualité. En effet, la correspondance entre **Pascal** et Fermat (1654) autour d'un jeu de partage constitue un sujet classique en histoire de la probabilité fréquentiste. Le problème proposé à Pascal par le Chevalier de Méré consiste à partager les gains d'un jeu devant être interrompu avant la fin de la partie; dans ce problème, on peut reproduire l'expérience autant de fois que l'on veut, il s'agit donc bien d'une interprétation fréquentiste de la probabilité.
Au contraire, le célèbre **pari de Pascal** propose deux hypothèses disjointes "Dieu est" et "Dieu n'est pas". On part donc ici de propositions ou hypothèses qui ne peuvent être reproduites : on a affaire à une interprétation bayésienne.
En 1765, **Leibniz** publie un ouvrage qui porte sur les aspects philosophiques de la probabilité. Considérant que la probabilité permet de modéliser la crédibilité, il propose une nouvelle sorte de logique pour traiter l'incertain, qui s'appuie sur l'interprétation bayésienne de la probabilité.
Quant à **Jacques Bernoulli** dont le texte posthume « Ars conjectandi », publié en 1713 est l'œuvre la plus célèbre, il se place résolument du côté fréquentiste et prouve ce que l'on appelle aujourd'hui la loi faible des grands nombres[2]. Ce théorème constitue un pilier de la théorie de la probabilité.
On remarque que la dualité est déjà bien présente à cette époque puisque le terme probabilité signifie tantôt un degré de certitude, tantôt la fréquence d'apparition d'un phénomène.
C'est dans l'œuvre posthume de **Thomas Bayes** (Philosophical Transaction 1764), que l'on trouve pour la première fois l'inversion de la probabilité. Il est toutefois difficile, voire impossible de trouver les traces des interprétations de la probabilité dans le texte original de Bayes. En fait c'est Price qui joindra des exemples et qui à la fin de l'article donnera des applications à des problèmes de cause à effet, c'est-à-dire dans lesquels l'observation d'un effet servirait à confirmer l'hypothèse.
Bayes a en effet trouvé une expression permettant d'inverser le schéma du problème de Bernoulli et de traiter mathématiquement des situations où il faut expliquer un phénomène donné par la recherche

[2] p = chance de succès (inconnu), s_n= proportion de succès après n essais
la probabilité d'une série de n termes pour laquelle $|p - s_n|$ < ε tend vers 1 lorsque n tend vers l'infini. De plus pour toute erreur donnée ε , il montre comment calculer un nombre n tel que la probabilité que n s se trouve dans l'intervalle [p- ε ; p+ε] dépasse elle-même toute probabilité donnée 1− δ.

d'une cause, la plus probable. D'après Jordan « …Pour résoudre ce problème, on a recours au théorème de Bayes. Un événement peut avoir pour causes C_1, C_2, C_3 … Désignons par w_i la probabilité a priori, c'est-à-dire avant l'arrivée de l'événement, pour que la cause C_i soit en jeu ; de plus désignons par p_i la probabilité pour que l'événement se produise si la cause C_i agit. On démontre que la probabilité a posteriori, c'est-à-dire après l'arrivée de l'événement, pour que celui-ci soit dû à la cause C_i est $p_i = \frac{w_i p_i}{\sum w_s p_s}$ »

Pour Carranza, l'inversion de la probabilité n'est pas un fait anecdotique, il ne s'agit pas d'un simple échange de places car elle met plus en évidence encore les différences entre les deux interprétations de la probabilité et que c'est pour pouvoir appliquer cette inversion que l'on est amené à interpréter la probabilité comme étant un degré de certitude.

Un peu plus tard, **Pierre-Simon Laplace**, dans son essai philosophique sur la probabilité (1795) donne une définition opératoire de la probabilité.
On trouve une interprétation fréquentiste dans le chapitre sciences naturelles (grand nombre d'observations, attente à long terme) mais aussi une interprétation bayésienne dès le début de l'ouvrage : des urnes de compositions inconnues amènent à probabiliser sur une hypothèse, une estimation de la probabilité d'un événement futur par un événement passé utilisant la formule des probabilités totales.
Au siècle dernier, **John Keynes** propose une théorie de la probabilité comme une relation logique entre deux propositions (A treatise on probability Keynes, 1921). Influencé par Leibniz, il est resté très attaché à la recherche de fondements logiques d'une probabilité caractérisée comme un degré de croyance rationnelle en la vérité d'une proposition.

Pendant une bonne partie du vingtième siècle, des années trente jusqu'aux années soixante-dix, on a assisté à de grandes controverses au cours desquelles les contributions de nombreux mathématiciens ont permis aux deux écoles de se consolider, tant dans leurs aspects philosophiques que méthodologiques. Il a donc fallu attendre plus d'un siècle pour que la théorie inférentielle se consolide et construise ses objets de base, à peu près le même temps que pour le développement de la théorie fréquentiste.

Revenons sur les différences entre les deux interprétations analysées par Carranza.

En plus de la différence déjà vue sur la variabilité de la valeur (fixe pour l'approche fréquentiste et susceptible de variations pour la bayésienne), une différence essentielle entre les deux approches concerne les **types de raisonnement** mobilisés.
Pour la probabilité fréquentiste, on raisonne par **déduction** : en admettant les hypothèses du modèle, on déduit la probabilité cherchée. Par exemple, si on suppose le dé équilibré, la probabilité d'obtenir 6 sera la fréquence théorique d'apparition du 6, soit 1/6. C'est un raisonnement du type « A implique B. A est vrai. Donc B est vrai ». Dans un tel raisonnement déductif, la conclusion est indiscutable et ne laisse aucune place au doute.
Pour la probabilité bayésienne, le raisonnement mobilisé est connu sous le terme d'**abduction**. Il désigne l'action de choisir la ou les hypothèses les plus vraisemblables pour expliquer un phénomène afin d'aboutir à une conclusion qui n'est pas certaine mais qui concorde avec les observations.[3]
Par exemple, lorsqu'on énonce « Tous les hommes sont mortels. Socrate est un homme. Donc Socrate est mortel », on fait un raisonnement déductif, mais lorsqu'on fait la démarche dans l'autre sens « Tous les hommes sont mortels. Socrate est mortel. Donc *il y a des chances que* Socrate soit un

[3] https://www.toupie.org/Dictionnaire/Abduction.htm

homme », on fait cette fois un **raisonnement inductif** du type « A implique B. B est vrai. Donc *il y a des chances que* A soit vrai ».
Ce raisonnement par abduction n'aboutit pas à une vérité absolue ; il donne seulement des indices qui poussent à une conclusion. On peut d'ailleurs se tromper en faisant un raisonnement inductif comme dans le fameux exemple « *Tous les chats sont mortels. Socrate est mortel. Donc Socrate est un chat* » !

Le raisonnement par abduction se retrouve en sciences pour valider une théorie à partir des expériences, dans le domaine de la justice pour décider d'un coupable à partir de preuves, ou en médecine pour faire un diagnostic à partir de symptômes. Mais le problème, c'est qu'il laisse la place au doute. Et c'est la formule de Bayes qui nous permet de quantifier notre degré de confiance. [4]

Une troisième différence concerne les **valeurs logiques**.
Par exemple, lorsqu'on lance un dé équilibré, la probabilité fréquentiste ne peut pas prédire le résultat d'un lancer en particulier mais donne 1/6 pour la probabilité d'obtenir 6 ; toute autre valeur serait erronée.
La probabilité bayésienne est une mesure de "croyance", elle donne un degré de certitude : plus on est sûr de la véracité d'une proposition, plus la probabilité sera proche de 1 (1=vrai et 0= faux) mais toute valeur logique intermédiaire est possible.

Les **critères d'évaluation** constituent une autre différence entre les deux interprétations.
La probabilité fréquentiste est évaluée numériquement de deux façons : on détermine a priori la valeur de la fréquence d'apparition de l'événement (nombre de cas favorables sur nombre de cas possibles en cas d'équiprobabilité) ou on l'estime à partir d'échantillons.
La probabilité bayésienne permet d'évaluer les probabilités portant sur des hypothèses. Lorsqu'on n'a pas d'argument en faveur de l'une ou l'autre de ces hypothèses, on utilise le « principe de raison insuffisante», ce qui signifie que l'on attribue la même probabilité à chaque d'elles (distribution uniforme).

Enfin, Carranza note une autre différence, fondamentale d'après lui, qui est la **nature du phénomène** sur lequel porte la probabilité.
Du côté fréquentiste, on étudie un phénomène que l'on peut reproduire autant de fois que l'on veut et on obtient des réponses à des questions formulées sur les tendances à long terme. Par exemple, le point de vue fréquentiste ne donne pas le résultat du $k^{ème}$ lancer d'un dé mais indique seulement ce qui arrive lorsqu'on reproduit l'expérience un nombre infini de fois (ce qu'on a appelé la série infinie).
Du côté bayésien, au contraire, l'intérêt porté au phénomène est de caractère épistémique (c'est-à-dire qu'il porte sur l'essence des choses) et la probabilité mesure notre certitude. Carranza distingue encore deux types de phénomènes incertains : d'une part un « évènement générique », comme le lancer d'un dé, qui peut être reproduit et dont le résultat s'inscrit dans un ensemble de référence, d'autre part des situations dont la reproduction n'est pas envisageable et pour lesquelles on ne peut pas trouver d'ensemble de référence. Elles sont appelées des « hypothèses », comme par exemple dans le pari de Pascal ou la vente d'avions au Brésil ; la probabilité est alors utilisée comme un outil rationnel de prise de décision.

[4] D'après https://scienceetonnante.com/2012/10/15/linference-bayesienne-bayes-level-2/

Cependant, la lecture de nombreux auteurs sur la construction du concept de probabilité a amené l'auteur à conclure que les deux interprétations de la probabilité sont indissociables, à tel point que cette dualité observée dans l'objet épistémologique allait se reproduire dans l'objet à enseigner.

Dans une deuxième partie de sa thèse, Carranza analyse programmes et documents d'accompagnement pour l'année 2000 ainsi que des exercices tirés de plusieurs manuels scolaires. Il constate que les exercices bayésiens sont bien présents dans les manuels mais que ces exercices servent plutôt à entraîner les élèves à des techniques calculatoires. Aucune question n'est posée relativement à l'interprétation des calculs effectués de sorte que l'interprétation bayésienne de la probabilité reste « cachée » aux élèves.

La troisième partie est consacrée à trois expérimentations en classe de BTS électrotechnique (faute d'avoir pu expérimenter au lycée), l'objectif principal étant une sensibilisation à la dualité de la probabilité. Nous reprenons l'une d'entre elles ci-dessous.

La bouteille de Brousseau : une expérimentation en classe de BTS
Le professeur dispose d'une bouteille opaque dont le bouchon est transparent. Il prélève dans une urne ne contenant que des billes noires et oranges (mais on ne sait pas dans quelle proportion) quatre billes qu'il introduit dans la bouteille sans les montrer. Il demande alors aux élèves s'ils voient une méthode leur permettant d'estimer le nombre de billes de chaque couleur dans la bouteille.
Il distribue ensuite à chacun des six binômes 20 jetons identiques et une feuille présentant les cinq compositions possibles (NNNN, NNNO, NNOO, NOOO, OOOO) et un questionnaire. Il explique que les jetons sont un moyen pour représenter le degré de confiance accordé à chaque composition (« si vous croyez plus en une composition qu'en une autre, vous mettez plus de jetons là où vous croyez le plus »)

1ère mise **a priori** : on observe plusieurs répartitions ; un seul binôme sur les 6 a misé 4 jetons sur chaque composition (probabilité uniforme de 0,2 qui correspond au principe de raison insuffisante) ; d'autres ont misé moins sur les deux compositions monochromatiques, pensant qu'elles étaient moins probables.
Ensuite le professeur retourne la bouteille et montre la bille visible dans le goulot ; il propose alors aux élèves de modifier leur répartition des jetons, c'est-à-dire d'estimer une probabilité **a posteriori** pour chaque composition. Par exemple si la bille apparue est orange, les jetons placés sur NNNN seront répartis sur les autres compositions, de plusieurs façons possibles (on peut penser que, comme on a vu une orange, c'est plutôt NOOO que NNNO).
Le procédé sera reproduit dix fois ; après chaque retournement, les élèves ont la possibilité de redistribuer les 20 jetons selon leurs croyances.
A partir du 3ème retournement, un tableur est à leur disposition, les jetons étant remplacés par des valeurs numériques. Finalement, le professeur demande quelle est la composition de la bouteille la plus probable.

L'auteur insiste sur le rôle fondamental joué par les jetons qui permettent aux élèves de se concentrer sur les raisons qui les poussent à modifier leurs distributions. L'intérêt est donc moins de trouver la composition de la bouteille que de voir évoluer la mesure proposée pour chaque composition au fur et à mesure des retournements : c'est comme si les élèves indiquaient à chaque étape leur degré de confiance pour chaque hypothèse.

Or les changements de distributions se justifient par un raisonnement par abduction qui est difficile à percevoir. Après le second retournement, le professeur a dû faire une rapide institutionnalisation de ce type de raisonnement et a présenté le théorème de Bayes comme un algorithme objectivant la démarche abductive. Il a ensuite montré comment ce calcul s'effectuait automatiquement à l'aide du tableur et a demandé de reprendre le problème au début afin qu'ils comparent leurs distributions avec celles obtenues en utilisant la formule.

On s'est limité ci-dessous à 5 retournements. La conclusion est la même qu'avec 10.
Si l'on suit le principe de raison insuffisante, les cinq compositions C_0(NNNN), C_1(NNNO), C_2(NNOO), C_3(NOOO) et C_4(OOOO) ont la même probabilité

$$P(C_0) = P(C_1) = P(C_2) = P(C_3) = P(C_4) = \frac{1}{5}$$

Au premier retournement, on obtient une bille orange ; on note O l'événement « on a obtenu une bille orange au 1er retournement ».
Pour chaque composition de la bouteille, on peut déterminer la probabilité de O sachant cette composition :
$P_{C_0}(O) = 0$; $P_{C_1}(O) = \frac{1}{4}$ puisqu'il y a une seule bille orange sur les quatre ; $P_{C_2}(O) = \frac{1}{2}$; $P_{C_3}(O) = \frac{3}{4}$ et enfin $P_{C_4}(O) = 1$

La formule des probabilités totales permet de calculer la probabilité de O

$$P(O) = \sum_{i=0}^{4} P(O \cap C_i) = \sum_{i=0}^{4} P_{C_i}(O) \times P(C_i) = 0 \times \frac{1}{5} + \frac{1}{4} \times \frac{1}{5} + \frac{2}{4} \times \frac{1}{5} + \frac{3}{4} \times \frac{1}{5} + 1 \times \frac{1}{5} = \frac{1}{2}$$

On trouverait aussi que la probabilité d'avoir une bille noire au 1er retournement est égale à ½.

Grâce à la formule de Bayes, on peut à présent calculer les probabilités conditionnelles pour i variant de 0 à 4 :

$$P_O(C_i) = \frac{P(C_i \cap O)}{P(O)} = \frac{P_{C_i}(O) \times P(C_i)}{P(O)}$$

On obtient $P_O(C_0) = 0$; $P_O(C_1) = \frac{\frac{1}{4} \times \frac{1}{5}}{\frac{1}{2}} = \frac{1}{10}$; $P_O(C_2) = \frac{\frac{1}{2} \times \frac{1}{5}}{\frac{1}{2}} = \frac{1}{5}$; $P_O(C_3) = \frac{\frac{3}{4} \times \frac{1}{5}}{\frac{1}{2}} = \frac{3}{10}$

Et $P_O(C_4) = \frac{1 \times \frac{1}{5}}{\frac{1}{2}} = \frac{2}{5}$

D'où le tableau

	C_0 =NNNN	C_1=NNNO	C_2=NNOO	C_3=NOOO	C_4=OOOO
Probabilité a priori	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
Après le 1er retournement O	0	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{4}{10}$

Les élèves qui avaient mis au départ 4 jetons sur chaque composition ont bien sûr retiré les 4 jetons mis sur C_0 après le 1er retournement O mais comment les répartir sur les autres ? On peut en ajouter 1 sur chacune des 4 compositions restantes (0-5-5-5-5) ou en enlever sur C_1 qui devient « moins probable » et en ajouter sur C_4 qui devient au contraire « plus probable »…Il n'est pas aisé de deviner la répartition (0-2-4-6-8) pour les 20 jetons.

Au deuxième retournement, on obtient à nouveau une bille orange ; pour ne pas compliquer les notations, on note encore O l'événement « on a obtenu une boule orange au 2ème retournement ».
Les probabilités conditionnelles $P_O(C_i)$ sont les mêmes que précédemment mais les nouvelles valeurs de $P(C_i)$ sont celles que l'on a obtenues ci-dessus donc :
$P(O) = \sum_{i=0}^{4} P_{C_i}(O) \times P(C_i) = \frac{1}{4} \times \frac{1}{10} + \frac{2}{4} \times \frac{2}{10} + \frac{3}{4} \times \frac{3}{10} + 1 \times \frac{4}{10} = \frac{3}{4}$

La formule de Bayes donne alors :
$P_O(C_0) = 0$; $P_O(C_1) = \frac{\frac{1}{4} \times \frac{1}{10}}{\frac{3}{4}} = \frac{1}{30}$; $P_O(C_2) = \frac{\frac{2}{4} \times \frac{2}{10}}{\frac{3}{4}} = \frac{4}{30}$; $P_O(C_3) = \frac{\frac{3}{4} \times \frac{3}{10}}{\frac{3}{4}} = \frac{3}{10} = \frac{9}{30}$

Et $P_O(C_4) = \frac{1 \times \frac{4}{10}}{\frac{3}{4}} = \frac{16}{30}$

On remarque que les probabilités sont dans le même ordre qu'après le 1er retournement

$\frac{1}{30} < \frac{4}{30} < \frac{9}{30} < \frac{16}{30}$ et que la probabilité de C_3 n'a pas changé.

Au troisième retournement, on a une bille noire. Si on note N l'événement « on a obtenu une bille noire au 3ème retournement », les calculs sont analogues :

$P(N) = \sum_{i=0}^{4} P_{C_i}(N) \times P(C_i) = \frac{3}{4} \times \frac{1}{30} + \frac{2}{4} \times \frac{4}{30} + \frac{1}{4} \times \frac{9}{30} + 0 \times \frac{16}{30} = \frac{1}{6}$ puis

$P_{C_1}(N) = \frac{\frac{3}{4} \times \frac{1}{30}}{\frac{1}{6}} = \frac{3}{20}$; $P_{C_2}(N) = \frac{\frac{2}{4} \times \frac{4}{30}}{\frac{1}{6}} = \frac{4}{10}$; $P_{C_3}(N) = \frac{\frac{1}{4} \times \frac{9}{30}}{\frac{1}{6}} = \frac{9}{20}$ et $P_{C_4}(N) = 0$

Au quatrième retournement, on a encore une bille noire

On trouve $P(N) = \frac{34}{80}$; les probabilités conditionnelles sont dans le tableau ci-dessous.

Le cinquième retournement montre une bille orange. Un calcul analogue donne $P(O) = \frac{1}{2}$
Ce résultat n'est pas étonnant car après avoir vu deux fois O puis deux fois N, on se retrouve comme au début de l'expérience. On obtient finalement le tableau suivant :

	C_0 =NNNN	C_1=NNNO	C_2=NNOO	C_3=NOOO	C_4=OOOO
Probabilité a priori	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
Après le 1er retournement O	0	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{4}{10}$
Après le 2ème retournement O	0	$\frac{1}{30}$	$\frac{4}{30}$	$\frac{9}{30}$	$\frac{16}{30}$
Après le 3ème retournement N	0	$\frac{3}{20}$	$\frac{8}{20}$	$\frac{9}{20}$	0
Après le 4ème retournement N	0	$\frac{9}{34}$	$\frac{16}{34}$	$\frac{9}{34}$	0
Après le 5ème retournement O	0	$\frac{9}{68}$	$\frac{32}{68}$	$\frac{27}{68}$	0

On en conclut que la composition la plus probable après 5 retournements qui ont donné OONNO est C_2(NNOO), résultat que certains avaient deviné sans calcul !

Pourquoi s'agit-il bien d'un problème de type bayésien ?
On part d'une hypothèse que l'on ne peut pas reproduire puisqu'on ne connait pas la composition de l'urne dans laquelle on a tiré les quatre billes.
Dans ses interventions, le professeur s'attache à respecter la subjectivité de chacun (« combien tu mets ici ? ») et évite les expressions comme « la probabilité est égale à … »
Le critère d'évaluation privilégié est le principe de raison insuffisante (même mise sur chacune des répartitions) bien qu'un seul binôme l'ait choisi lors de la 1ère mise.
La variabilité des valeurs de la probabilité s'est manifestée tout au long du déroulement, cette probabilité évoluant au fur et à mesure des retournements.
Après le 1er retournement qui a montré une bille orange, les groupes ont retiré naturellement les jetons placés sur NNNN et certains ont également retiré un jeton de NNNO pour le placer sur NOOO qui est devenu plus probable d'après eux. Ainsi quelques élèves ont mis en œuvre un raisonnement par abduction en modifiant leurs croyances compte-tenu de la couleur apparue.

Tous ces arguments confirment que le problème est de type bayésien. Les éléments spécifiques à chaque interprétation ont été utilisés, soit de manière implicite par les élèves, soit explicitement par le professeur lors de l'institutionnalisation. Cependant Carranza souligne que le théorème de Bayes est un objet complexe et qu'il n'était pas prévu de créer les conditions pour faciliter sa construction. Même si, faute de temps, plusieurs notions ont été présentées de manière trop rapide, le fait de pouvoir différencier les deux interprétations a permis au professeur de disposer d'une plus grande marge de manœuvre pour proposer à sa classe d'autres types d'activités que les exercices classiques de calcul de probabilités.

Nous signalons que le problème de la bouteille peut aussi être envisagé de façon fréquentiste. C'est ce que propose Brigitte Sotura en classe de 2nde (voir manuel Belin 2nde collection metamaths) : après de nombreux retournements, les élèves doivent deviner la composition de la bouteille mais il s'agit cette fois de faire émerger le lien entre fréquence et probabilité.

Annexes

Annexe 1 : Mise au point sur la démarche suivie, les données, le vocabulaire utilisé selon le contexte et récapitulatif des termes utilisables : quand les probabilités se « mélangent » aux statistiques…

Lorsqu'on « fait » des probabilités, on étudie des phénomènes aléatoires. On part d'une population (personnes ou individus ou objets, etc.), modélisée par un ensemble d'éléments, qu'on appelle aussi univers, souvent noté Ω. Chaque élément de l'univers est une éventualité ou encore une issue. Tout sous-ensemble de Ω est un événement. Un phénomène aléatoire à étudier est associé à cet univers.

Lorsqu'on a affaire à un univers fini, et qu'on peut connaître les effectifs des éléments en jeu et que toutes les issues sont équiprobables, pour obtenir la probabilité que se réalise un événement A, on utilise la formule P(A) = $\frac{nombre\ d'issues\ constituant\ A}{nombre\ d'issues\ possibles}$

Souvent on connait certaines probabilités (déduites d'observations directes), on veut en connaître d'autres, c'est ce que permet ce qui suit.

En statistiques, on étudie des populations, des données, obtenues effectivement – soit pour en décrire la nature, la diversité, la fréquence (statistiques descriptives), soit pour tirer de données initiales, obtenues sur des échantillons, de nouvelles données encore inconnues, souvent à tester (statistiques inférentielles).

Souvent sont alors en jeu des caractères de ces populations ou de ces données, connus ou à étudier – si on se penche sur l'exemple introductif d'un point de vue statistique, on évoquera les caractères forme et couleur « carré », « rouge », « carré rouge » etc…

Le fait est qu'on est amené à mélanger les deux points de vue.
Par exemple on part de données effectives, on mesure la fréquence de la valeur d'un caractère, puis on l'assimile à la probabilité de l'événement associé à cette valeur.

Annexe 2 : Termes utilisés dans les tests médicaux

On part d'une population sur laquelle on étudie deux caractères (état de santé : malade et non malade) et (test : positif et négatif).
On peut aussi s'intéresser, d'un point de vue probabiliste, aux événements A « être malade », B « avoir un test positif », etc.

Probabilité a priori (que l'on peut estimer à partir des données) : c'est la probabilité de l'événement A. **C'est aussi la prévalence**. Très important : plus elle est grande plus ce qu'on calcule ensuite est indicatif !

Probabilité a posteriori : c'est la probabilité **à trouver** en introduisant une restriction sur les cas possibles, du type probabilité de B sachant A, les cas possibles correspondent à ceux qui présentent déjà le caractère A et pas à toute la population en jeu : penser à la probabilité d'être malade en ayant un test positif. **Une telle probabilité s'appelle conditionnelle : elle est conditionnée par la restriction des cas possibles : cf. a) et b).**

i) Tableaux de contingence (tableaux à double entrée) : ils donnent les effectifs de chaque valeur des caractères, dans chaque case du tableau.

On lit dans les cases du tableau :
Faux positifs : les personnes testées positivement et n'ayant pas la maladie testée,
Faux négatifs : les personnes testées négativement et ayant la maladie testée
Vrais positifs : les personnes testées positivement et ayant la maladie testée,
Vrais négatifs : les personnes testées négativement et n'ayant pas la maladie testée.
La somme des faux positifs et vrais positifs est égale à l'effectif des positifs.

La somme des faux négatifs et des vrais positifs est égale à l'effectif des malades

ii) **Sensibilité et spécificité d'un test**

On appelle **sensibilité du test** le pourcentage de tests positifs parmi les malades ou la probabilité correspondante et **spécificité** du test le pourcentage de tests négatifs parmi les non malades ou la probabilité correspondante.

LOIS DE PROBABILITÉS

BINOMIALE, GÉOMÉTRIQUE,

EXPONENTIELLE, UNIFORME

En liaison avec le thème « TEMPS D'ATTENTE »

Mathématiques complémentaires Temps d’attente
Préambule

Le programme de terminale de Mathématiques complémentaires s’appuie sur le programme de première de Spécialité mathématique qu’il enrichit de nouvelles connaissances. Celles-ci peuvent être soit listées dans la deuxième partie du programme (contenus) soit apparaître dans des Thèmes d’étude (première partie du programme). Dans ces thèmes peuvent intervenir des notions mathématiques déjà vues en première ou dans l’année et/ou ces nouvelles connaissances. Pour un même thème d’étude, il est parfois nécessaire d’avoir recours à des notions qui figurent dans des champs différents du descriptif des contenus, comme, par exemple, l’analyse et les probabilités, notions qui ont pu être ou non déjà rencontrées au moment où on aborde le thème.

Il y a cependant, dans ce qui est proposé par le programme, davantage qu’une succession de thèmes et de contenus nécessaires à traiter les thèmes. Il y est stipulé en effet : « L’objectif est de traiter l’ensemble des contenus et capacités attendues au travers des thèmes d’étude ».

C’est pour suivre cette ambition que nous avons voulu entrer dans le thème ci-dessous sans nécessairement étudier **au préalable** tous les contenus nouveaux du programme qui permettent de l’aborder.

La ressource « Temps d’attente » que nous présentons est dans une large mesure autonome. D’une part, en termes mathématiques, nous introduisons toutes les notions de probabilité nécessaires (et nouvelles) par des exemples complétés par un cours – définitions, propriétés, etc. Mais cette introduction des notions est motivée par leur inscription dans le thème.

Certaines démonstrations des propriétés de la loi exponentielle utilisent la notion d’intégrale ; elles peuvent cependant être admises sans dénaturer l’exploration du thème.

Par ailleurs, dans certains exercices la connaissance de la fonction *ln* pourrait faciliter la résolution d’équations ou inéquations mais ces dernières peuvent être résolues avec les connaissances de première en ayant recours à la calculatrice ou à un tableur.

Soulignons qu’on peut retrouver dans les contenus listés en deuxième partie du programme toutes les notions qui sont introduites à l’occasion du thème.

D’autre part nous avons cherché à présenter un document directement utilisable par un enseignant, quitte à supprimer des paragraphes qui feraient perdre trop de temps ou qui sembleraient trop difficiles pour une classe donnée. Le choix d’exercices proposés à la fin permet aussi de s’adapter à chaque classe, espérons-nous.

Précisons enfin que cette ressource n’a pas été l’objet d’une expérimentation précise.

Pourquoi ce thème ?

Dans le programme figurent **deux sortes de descriptifs**[1]**,** *ceux des thèmes d’études* auxquels sont associés des contenus mathématiques et *ceux des contenus mathématiques par champ* (analyse, probabilités et statistiques, etc.) Dans le descriptif du thème d’étude « Temps

[1] Les descriptifs pour le thème Temps d’attente se trouvent en annexe.

d'attente », la mention « paradoxe de l'inspection[2] » nous a interpelées. Il y est indiqué « Utilisation de la loi uniforme. Temps d'attente à un arrêt de bus, paradoxe de l'inspection », alors que dans l'introduction il est surtout fait référence à la propriété d'absence de mémoire. La recherche de références sur « paradoxe de l'inspection » nous conduit presque immanquablement au problème du temps d'attente à un arrêt d'autobus. Alors, quel lien y a -t-il entre l'utilisation de la loi uniforme mentionnée et les propriétés d'absence de mémoire ? Par ailleurs, y-a-t-il un lien entre « Durée de vie d'un atome radioactif » signalé comme problème possible et la propriété d'absence de mémoire ?
Finalement, après quelques recherches, cette simple mention « paradoxe de l'inspection » nous entraîne beaucoup plus loin et nous conduit à explorer tout le thème « Temps d'attente ».

Plan

[2] Il ne s'agit pas d'inspection « en classe » ! On trouvera une explication sur cette appellation dans la partie IV - Conclusion.

Introduction

Dans une acception courante, un temps d'attente ne semble pas relever d'une modélisation probabiliste. Spontanément on pourrait dire que la durée de vie d'un objet dépend de son utilisation ou de sa fabrication, que le temps d'attente à un arrêt d'autobus dépend de la fréquence des autobus. Alors, où interviennent les lois de probabilités sans mémoire ? Comment intervient la loi uniforme dans le temps d'attente à un arrêt d'autobus ? Y a-t-il différents « temps d'attente » ?

Pour répondre à ces questions nous avons besoin de préciser, d'une part ce que l'on entend par « temps d'attente » et, d'autre part, ce que traduit la « propriété d'absence de mémoire ».

- On parle de **temps d'attente** dans un processus aléatoire donné pour indiquer le délai avant la survenue d'un événement. Il peut être modélisé par une variable aléatoire et une loi de probabilité associée. Ce peut être par exemple, lors du lancer d'un dé, le nombre d'essais nécessaires pour obtenir un 6, c'est-à-dire le rang du premier 6 obtenu lorsqu'on ne limite pas le nombre d'essais (cas discret) ; ce peut être le temps d'utilisation avant la première panne d'un appareil ou le temps mis par une substance (molécule, médicament, etc.) pour perdre la moitié de son activité pharmacologique ou physiologique, ou encore le temps au bout duquel la moitié des noyaux radioactifs d'une source se sont désintégrés (cas continu.

Au niveau du lycée, les lois les plus utilisées concernant ces variables aléatoires sont la loi géométrique (cas discret) et la loi exponentielle (cas continu). Mais sur quoi s'appuie-t-on pour déterminer ou choisir un modèle pour la variable aléatoire ? En particulier, quel lien peut-on faire entre ce qu'on peut connaître du processus par des relevés statistiques ou expérimentaux et la recherche d'un modèle ?

La loi peut être connue parce qu'il a été établi que telle ou telle situation relève de ce modèle et il s'agit alors de répondre à des questions de prévisions ou de probabilités comme souvent dans les exercices scolaires.

- Un phénomène physique possède la **propriété d'absence de mémoire** si le fait qu'il ait duré pendant *t* heures ne change rien à sa durée (son espérance de vie) à partir du temps *t.* Dans le cas où le phénomène physique est modélisé par une variable aléatoire égale au temps d'attente ou à la durée de vie, on formalise cette idée en disant que la probabilité que le phénomène dure au moins *s* + *t* heures sachant qu'il a déjà duré *t* heures à partir de sa mise en fonction initiale sera la même que la probabilité de durer *s* heures de plus, ou encore, pour tous $s, t \in [0; +\infty[, P_{(X>t)}(X > s + t) = P(X > s)$
ou, dans le cas discret, pour tous $n, m \in \mathbb{N}$, $P_{(X>n)}(X > n + m) = P(X > m)$.

Au cours de l'étude de ce thème « temps d'attente » nous introduisons les lois de probabilité géométrique et exponentielle possédant la propriété d'absence de mémoire, puis le paradoxe de l'inspection où intervient la loi uniforme sur un intervalle. Ce faisant, nous constatons que « temps d'attente » peut prendre des sens différents.

I –Variables aléatoires discrètes–la loi géométrique, loi discrète sans mémoire

Dans cette partie on introduit les outils nécessaires à l'étude de la seule loi sans mémoire associée à une variable aléatoire discrète, la loi géométrique. On a en particulier besoin du schéma de Bernoulli. On donne ensuite l'exemple d'une loi intermédiaire, la loi géométrique tronquée, qui n'est pas sans mémoire. Cela permet d'amener la loi géométrique et la propriété caractéristique annoncée qui est établie à la fin du paragraphe. Cela donne la modélisation des temps d'attente que l'on peut associer à de telles variables aléatoires discrètes.

1) Epreuve et loi de Bernoulli, schéma de Bernoulli

Dans ce paragraphe, à partir d'un exemple de jeu de cartes, on rappelle la définition d'une expérience aléatoire particulière (à deux issues seulement) et la loi de probabilité associée ainsi que ses premières propriétés. Puis on rappelle ce qu'est un schéma de Bernoulli, notion qui sera utile dans le paragraphe suivant (loi géométrique).

a) Epreuve et loi de Bernoulli

Loi de Bernoulli

Exemple
Au bonneteau, deux cartes noires et une carte rouge sont présentées, faces cachées, sur une table.
Un joueur choisit une carte au hasard. Il gagne s'il choisit la carte rouge, il perd dans le cas contraire.
On appelle S l'événement « le joueur gagne » et $\bar{S}$ l'événement « le joueur perd ».
Quelles sont les probabilités des événements S et $\bar{S}$?

Définition
On appelle **épreuve de Bernoulli** de paramètre p ($0<p<1$) une expérience aléatoire ayant deux issues :

- l'une appelée **succès** (généralement notée S) de probabilité p,
- l'autre appelée **échec** (généralement notée $\bar{S}$) de probabilité $1-p$.

On considère la variable aléatoire X qui vaut 1 en cas de succès et 0 en cas d'échec.
Cette variable aléatoire suit la **loi de Bernoulli de paramètre *p***, définie par le tableau suivant:

$\boldsymbol{x_i}$	0	1
$\boldsymbol{p(X = x_i)}$	$1-p$	p

Dans l'exemple ci-dessus on a affaire à une loi de Bernoulli de paramètre $\boldsymbol{p} = \frac{1}{3}$
La probabilité de succès est $\boldsymbol{p(S) = p} = \frac{1}{3}$ et la probabilité d'échec $\boldsymbol{p}(\bar{S}) = \boldsymbol{1 - p} = \frac{2}{3}$

Propriétés[3]

L'espérance mathématique et la variance d'une variable aléatoire X qui suit une loi de Bernoulli de paramètre *p* sont $E(X) = p$ et $Var(X) = p(1-p)$

En effet, par définition, $E(X) = 0 \times (1-p) + 1 \times p = p$
et $Var(X) = (0-p)^2(1-p) + (1-p)^2 p \ = p(1-p)$

b) Schéma de Bernoulli

Définition

On appelle **schéma de Bernoulli** la répétition d'épreuves de Bernoulli **identiques** et **indépendantes**.
On peut associer à cette répétition d'épreuves différentes variables aléatoires, comme, par exemple, le nombre de succès à l'issue de n épreuves (loi binomiale – voir le thème « Répétition d'expériences indépendantes…) ou rang du premier succès (lois géométrique tronqué et géométrique – voir paragraphe suivant)

2) Variables aléatoires discrètes

Dans ce paragraphe, on rappelle la notion de variable aléatoire discrète finie, introduite en 1ère, et on introduit celle de variable aléatoire discrète définie sur $\mathbb{N}$.

* Si une variable aléatoire *X* prend un nombre fini *n* de valeurs (l'univers Ω est fini, $\Omega = \{x_k / k \epsilon \{0; 1; 2; \dots n\}\}$) on peut définir sa loi de probabilité en donnant $P(X = x_k)$ pour tout $k \epsilon \{0; 1; 2; \dots n\}$. Par exemple, somme des points obtenus en lançant deux dés cubiques, loi binomiale, etc.
C'est aussi le cas lorsqu'on répète *n* fois une même expérience aléatoire les expériences étant indépendantes les unes des autres et que l'on introduit la variable aléatoire égale au rang du premier succès.

* On peut ne pas mettre de limite au nombre d'expériences aléatoires et s'intéresser au rang du premier succès. Dans ce cas la variable aléatoire est définie sur $\mathbb{N}$ (et à valeurs dans $\mathbb{R}$).
* Dans les deux cas précédents, on dit que la variable aléatoire est discrète.

Dans les deux paragraphes suivants on développe deux cas de loi de probabilité associée à une variable aléatoire discrète.

3) Loi géométrique tronquée (avec mémoire)

Dans ce paragraphe, toujours à partir d'un exemple, on introduit une nouvelle variable aléatoire associée au rang de l'obtention du succès. Mais ici l'expérience s'arrête, soit au bout d'un

[3] Les définitions de l'espérance et de la variance d'une variable aléatoire ont été vues en 1ère.

nombre fini de répétitions, soit au premier succès, ce qui explique l'appellation « tronquée » et empêche d'introduire le caractère « sans mémoire ».

http://pedagogie.ac-limoges.fr/maths/IMG/pdf/temps_d_attente.pdf

Exemple

On lance un dé bien équilibré à quatre faces numérotées 1 ; 2 ; 3 et 4. Ici on s' intéresse à l'événement « obtenir 1 ou ne pas obtenir 1 ».

La probabilité d'obtenir 1 à l'issue d'un lancer est donc $\frac{1}{4}$. On répète cette expérience au maximum cinq fois mais l'on s'arrête dès que 1 est sorti. On suppose que les lancers sont indépendants les uns des autres (autrement dit, on ne triche pas !).

On s'intéresse au rang du premier lancer où l'on obtient 1.

On peut obtenir 1 au premier lancer ou au deuxième lancer, au troisième lancer, etc. ou ne pas obtenir 1 à l'issue des 5 lancers.

On note *X* la variable aléatoire égale au rang où l'on obtient 1 la première fois et à 0 si l'on n'obtient jamais 1.

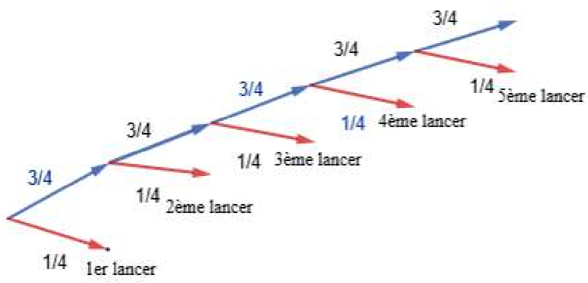
$X \in \{0; 1\,; 2\,; 3\,; 4\,; 5\,\}$

La probabilité d'obtenir 1 au premier lancer est $\frac{1}{4}$.

L'arbre ci-contre permet de représenter la situation.

La probabilité de n'obtenir 1 qu'au deuxième lancer est $\frac{3}{4} \times \frac{1}{4}$, celle de n'obtenir 1 qu'au troisième essai est $\left(\frac{3}{4}\right)^2 \times \frac{1}{4}$, etc.

Mais le 1 peut ne pas « sortir » et dans ce cas $P(X = 0) = \left(\frac{3}{4}\right)^5$



D'où la loi de probabilité :

X	1	2	3	4	5	0
P(X)	$\frac{1}{4}$	$\frac{3}{4} \times \frac{1}{4}$	$\left(\frac{3}{4}\right)^2 \times \frac{1}{4}$	$\left(\frac{3}{4}\right)^3 \times \frac{1}{4}$	$\left(\frac{3}{4}\right)^4 \times \frac{1}{4}$	$\left(\frac{3}{4}\right)^5$

L'espérance mathématique est égale à :

$$1 \times \frac{1}{4} + 2 \times \frac{3}{4} \times \frac{1}{4} + 3 \times \left(\frac{3}{4}\right)^2 \times \frac{1}{4} + 4 \times \left(\frac{3}{4}\right)^3 \times \frac{1}{4} + 5 \times \left(\frac{3}{4}\right)^4 \times \frac{1}{4} + 0 \times \left(\frac{3}{4}\right)^5$$

$$E(X) = \frac{3205}{1024} \approx 3{,}13$$

3,13 est le rang moyen de la première sortie d'un « 1 ». $E(X)$ est une moyenne, il ne faut donc pas s'étonner que le résultat ne soit pas un nombre entier.

Cas général

On se place dans un schéma de Bernoulli : une expérience est répétée n fois de façon aléatoire et indépendante chaque fois. L'issue de chaque expérience a une probabilité p $(p \neq 0)$de réussite. On définit la variable aléatoire *X* comme **le rang du premier succès** et *X = 0* si les *n* essais sont des échecs.

En généralisant l'exemple précédent, on a pour tout k entier naturel,

$$1 \leq k \leq n, \quad P(X=k) = p(1-p)^{k-1} \text{ et } P(X=0) = (1-p)^n$$

Les formules pour l'espérance et la variance ne sont pas simples à établir et n'ont pas grand intérêt pour notre propos.

Remarque 1

Qu'on n'ait pas obtenu « 1 » quatre fois de suite ou même dix fois de suite, la probabilité d'obtenir « 1 » au lancer suivant est toujours $\frac{1}{4}$. Contrairement à ce qu'on pense communément, le temps d'attente d'un « 1 » n'est pas modifié par le fait qu'il n'est pas sorti depuis longtemps. On pourrait penser que cela constitue « l'absence de mémoire ». Mais dans le cas où le nombre de lancers est limité, la loi de probabilité n'est pas sans mémoire au sens de la définition adoptée en probabilités.

En effet, on reprend l'exemple d'introduction. On peut, par exemple, calculer $P_{(X>1)}(X > 3)$ et la comparer à $P(X > 2)$. En effet, 3=2+1

$$P_{(X>1)}(X>3) = \frac{P(X>3 \; et \; X>1)}{P(X>1)} = \frac{P(X=4)+P(X=5)}{P(X \geq 2)} = \frac{\frac{1}{4} \times [\left(\frac{3}{4}\right)^3 + \left(\frac{3}{4}\right)^4]}{\frac{1}{4} \times [\left(\frac{3}{4}\right)^2 + \left(\frac{3}{4}\right)^3 + (\frac{3}{4})^4]}$$

$$= \frac{21}{55} \approx 0{,}38$$

qui est différent de $P(X > 2)$ qui est égal à

$P(X \geq 3) = P(X=3) + \; P(X=4) + P(X=5) = \frac{1}{4} \times [\left(\frac{3}{4}\right)^2 + \frac{1}{4} \times [\left(\frac{3}{4}\right)^3 + \left(\frac{3}{4}\right)^4] = \frac{333}{1024} \approx$ 0,325

Remarque 2

Alors que la loi binomiale permet de calculer la **probabilité d'obtenir *k* succès lors de la répétition *n* fois** d'une même épreuve à deux issues possibles de manière indépendantes, on calcule ici le **rang** du premier succès lors de cette répétition de *n* épreuves dans les mêmes conditions.

4) Loi géométrique (loi discrète- loi sans mémoire)

Ne pas confondre loi « hypergéométrique » et loi géométrique.

Dans ce paragraphe on commence par généraliser la loi précédente, sans imposer un arrêt des expériences répétées sauf en cas de succès. Cela permet d'introduire et de démontrer une *propriété, associée à cette nouvelle loi : la propriété d'être sans mémoire, c'est-à-dire que le passé n'a aucune « influence » sur l'avenir. On termine en démontrant la réciproque : toute loi discrète sans mémoire est géométrique.*

Dans le cas de la loi géométrique tronquée le nombre d'essais est limité. On peut aussi envisager de ne pas limiter le nombre d'essais et on est conduit à définir une variable aléatoire discrète.

Définition

X est une variable aléatoire **discrète** si elle prend un nombre fini de valeurs ou si ses valeurs sont indexées par $\mathbb{N}^*$. Autrement dit, en notant Ω l'ensemble des valeurs de la variable *X*, on a

$\Omega = \left\{x_i \epsilon \mathbb{R} /\ i \epsilon \{1; 2; \dots; n\}\right\}$ (*n* valeurs) ou $\Omega = \{x_i \epsilon \mathbb{R} /\ i \epsilon\ \mathbb{N}\}$ (une infinité de valeurs dénombrables)

a) La loi géométrique

C'est un prolongement de la loi géométrique tronquée au cas où on ne limite pas le nombre d'essais.

Définition

Dans une suite non limitée d'épreuves de Bernoulli identiques et indépendantes, on note *p* (0<p<1) la probabilité d'un succès. La loi de probabilité de la variable aléatoire *X* qui donne le rang du premier succès (ou le nombre d'expériences nécessaires pour obtenir le premier succès) s'appelle loi géométrique de paramètre *p*.

Il s'agit alors d'une variable aléatoire discrète infinie définie sur $\mathbb{N}^*$ à valeurs dans $\mathbb{R}$

Propriété

Pour tout k entier naturel, $1 \leq k$, $P(X = k) = p(1-p)^{k-1}$

et $P(X = 0) = \lim\limits_{n \to +\infty} (1-p)^n = 0$

On remarque que l'événement $X = 0$ est un événement « presque impossible »

Propriété admise

L'espérance mathématique d'une variable aléatoire qui suit une loi géométrique de paramètre p est $E(X) = \frac{1}{p}$

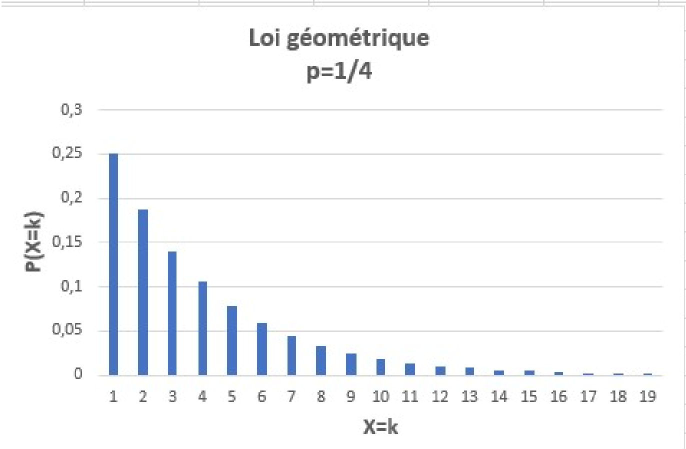
Dans l'exemple du dé à 4 faces, si on ne limite pas le nombre de lancers, la loi de X est donnée par le tableau :

X	1	2	3	4	5	6	…	k	…
P(X)	$\frac{1}{4}$	$\frac{3}{4} \times \frac{1}{4}$	$\left(\frac{3}{4}\right)^2 \times \frac{1}{4}$	$\left(\frac{3}{4}\right)^3 \times \frac{1}{4}$	$\left(\frac{3}{4}\right)^4 \times \frac{1}{4}$	$\left(\frac{3}{4}\right)^5 \times \frac{1}{4}$		$\left(\frac{3}{4}\right)^{k-1} \times \frac{1}{4}$	

On lit $P(X=6)=\left(\frac{3}{4}\right)^5\times\frac{1}{4}\approx 0{,}06$ c'est la probabilité que le 1 ne sorte qu'au 6ème lancer.

Et l'espérance de X est égale à 4, c'est-à-dire qu'en moyenne, le 1 sortira au 4ème lancer.

Représentation graphique : sur le graphique ci-dessous on a représenté la loi géométrique pour $p=\frac{1}{4}$

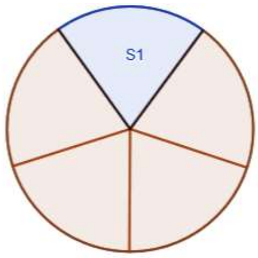


b) La loi géométrique est une loi sans mémoire (on dit aussi « sans vieillissement »)

Remarque : l'appellation « sans mémoire » semble remplacer « sans vieillissement », peut-être parce que le champ des lois sans mémoire est élargi à des situations autres que celles des objets industriels ?

Exemple

On fait tourner une roue de loterie, schématisée ci-contre (cinq secteurs égaux), sans limitation du nombre d'expériences.

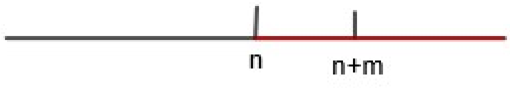


La probabilité de « tomber » sur le secteur S1 lors d'un essai est $p=\frac{1}{5}$.
Les essais sont indépendants.On sait que le succès n'a pas été réalisé jusqu'au rang 5. Alors, la probabilité que le succès soit réalisé à un rang supérieur à 8 est égale à la probabilité pour que le succès soit réalisé à un rang supérieur à 3 car 8=5+3.
Autrement dit : $P_{(X>5)}(X>8)=P(X>3)$
En effet, $P_{(X>5)}(X>8)=\frac{P(X>8\ et\ X>5)}{P(X>5)}=\frac{P(X>8)}{P(X>5)}$
Or $P(X>5)=1-P(X\leq 5)=1-[P(X=1)+P(X=2)+\cdots+P(X=5)]$
$P(X>5)=1-[\frac{1}{5}+\frac{1}{5}\times(\frac{4}{5})^1+\frac{1}{5}\times(\frac{4}{5})^2+\frac{1}{5}\times(\frac{4}{5})^3+\frac{1}{5}\times(\frac{4}{5})^4]$
Dans l'expression entre crochets on reconnait la somme des termes d'une suite géométrique de premier terme $\frac{1}{5}$ et de raison $\frac{4}{5}$. D'où $P(X>5)=1-\frac{1}{5}\left[\frac{1-(\frac{4}{5})^5}{1-\frac{4}{5}}\right]=(\frac{4}{5})^5$
De même $P(X>8)=(\frac{4}{5})^8$, d'où $P_{(X>5)}(X>8))=(\frac{4}{5})^8/(\frac{4}{5})^5=(\frac{4}{5})^3=P(X>3)$

Cas général :

La loi de probabilité du nombre d'épreuves à répéter jusqu'à l'obtention d'un premier succès dans une suite d'épreuves de Bernoulli identiques indépendantes est la même quel que soit le nombre d'échecs accumulés auparavant.
Ce qu'on traduit par : pour tous $n,m\in\mathbb{N}$, $P_{(X>n)}(X>n+m)=P(X>m)$

Démonstration : $P(X > m) = 1 - P(X \leq m) = 1 - \sum_{i=1}^{i=m} p(1-p)^{i-1}$
Or $\sum_{i=1}^{i=m} p(1-p)^{i-1} = p\sum_{i=1}^{i=m}(1-p)^{i-1} = p(\frac{1-(1-p^m)}{1-(1-p)})$
(on reconnait la somme des termes d'une suite géométrique).
Donc $P(X > m) = 1 - (1 - (1-p)^m) = (1-p)^m$
$P_{(X>n)}(X > n+m) = \frac{\boldsymbol{P(X>n+m\ et\ X>n)}}{\boldsymbol{P(X>n)}}$.
Or $X > n + m\ et\ X > n$ équivaut à $X > n + m$



Donc $P_{(X>n)}(X > n+m) = \frac{P(X>n+m)}{P(X>n)}$
D'après le calcul précédent $\frac{P(X>n+m)}{P(X>n)} = \frac{(1-p)^{n+m}}{(1-p)^n} = (1-p)^m = P(X > m)$, ce qui démontre la propriété :

$$P_{(X>n)}(X > n+m) = P(X > m)$$

c) La loi géométrique est la seule loi discrète sans mémoire

Soit X une variable aléatoire sur $\mathbb{N}^*$ suivant une loi sans mémoire donc vérifiant :
pour tous $n, m \in \mathbb{N}^*$, $P_{(X>m)}(X > m+n) = P(X > n)$. Démontrons que X suit une loi géométrique c'est-à-dire qu'on a pour tout k entier naturel non nul, $P(X = k) = p(1-p)^{k-1}$ où $p = P(X = 1)$.

En appliquant la définition lorsque n=1, on obtient $P_{(X>m)}(X > m+1) = P(X > 1)$
On sait que $P(X > 1) = 1 - P(X = 1)$.

On note p le nombre $P(X = 1)$; on obtient $P_{(X>m)}(X > m+1) = 1 - p$ ou encore, pour tout $m \in \mathbb{N}^*$, $\frac{P(X>m+1)}{P(X>m)} = 1 - p$
Ainsi la suite $(P(X > m))_{m \in \mathbb{N}^*}$ est une suite géométrique de raison $1 - p$ et de premier terme $P(X > 1) = 1 - P(X = 1) = 1 - p$.
On en déduit, pour tout $m \in \mathbb{N}^*$, $P(X > m) = (1-p)^{m-1} \times (1-p) = (1-p)^m$.
Mais on doit déterminer, pour tout m, $P(X = m)$, qui est égal à $P(X > m-1) - P(X > m)$

$$P(X = m) = (1-p)^{m-1} - (1-p)^m = (1-p)^{m-1}(1 - (1-p))$$

Finalement, pour tout $m \in \mathbb{N}^*$, $P(X = m) = p(1-p)^{m-1}$ ce qui prouve que X suit une loi géométrique de paramètre p.

Remarque : on démontre ainsi que la loi géométrique est la seule loi de probabilité discrète sans mémoire.

d) Une modélisation en temps d'attente discret
L'exercice suivant illustre une telle modélisation
Supposons que l'on lance plusieurs fois une pièce de monnaie truquée de manière indépendante.

La probabilité d'obtenir pile est deux fois grande que celle d'obtenir face. On appelle G la variable aléatoire indiquant le rang du premier lancer où la pièce donne pile. Le but de cet exercice est de déterminer la loi suivie par G.

1. Quelle est la probabilité qu'on obtienne pile au nième lancer ? d'obtenir face au nième lancer ?
2. Que signifie l'événement G = 1 en termes d'issue du premier lancer ? En déduire P(G = 1).
3. Que signifie l'événement G = 2 en termes d'issues des deux premiers lancers ? En déduire P(G = 2).
4. Plus généralement, calculer P(G = n) pour tout n ≥ 1.
5. Quelle est la loi suivie par G ?

<u>Solution</u>

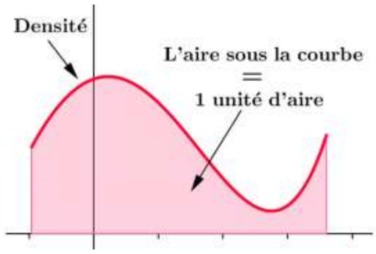
1. La probabilité d'obtenir pile au nième lancer est $\frac{2}{3}$, de même celle d'obtenir face est $\frac{1}{3}$.
2. L'événement G = 1 signifie qu'on obtient pile au premier lancer. Donc P(G=1)=$\frac{2}{3}$
3. L'événement G = 2 signifie qu'on obtient pile au deuxième lancer. Donc P(G=2)=$\frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$
4. P(G=n)=$(\frac{1}{3})^{n-1} \times \frac{2}{3}$
5. G suit une loi géométrique de paramètre p=$\frac{2}{3}$

II – Variables aléatoires continues - la loi exponentielle, loi à densité sans mémoire

Dans ce paragraphe on introduit les outils nécessaires à l'étude de la seule loi sans mémoire associée à une variable aléatoire continue, la loi exponentielle. Cette propriété caractéristique est établie à la fin du paragraphe. Cela donne la modélisation des temps d'attente que l'on peut associer à de telles variables aléatoires continues.

1) Variable aléatoire continue - Lois à densité[4]

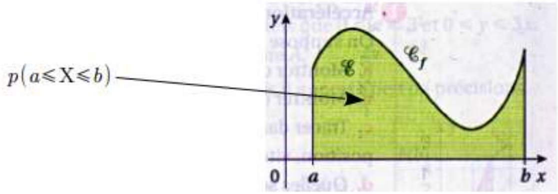
Lorsque la variable aléatoire X prend toutes ses valeurs dans un intervalle de $\mathbb{R}$, voire $\mathbb{R}$ tout entier, il n'est plus possible d'exprimer sa loi en donnant la probabilité de chaque éventualité. Dans ce cas on introduit une fonction de densité f définie et positive sur l'intervalle considéré. L'univers Ω est donc I ou $\mathbb{R}$. <u>La variable aléatoire X est dite variable aléatoire continue.</u>



L'ordonnée d'un point d'abscisse donnée de la courbe de cette fonction correspond en quelque sorte à la probabilité d'une éventualité d'une loi de probabilité d'univers fini mais, dans le cas où la fonction loi de densité est continue, on démontre que la probabilité d'une valeur précise de cette variable aléatoire est nulle.

[4] Nous donnons ici un bref aperçu de ce qu'est une loi à densité. Un tableau donnant les correspondances entre variables aléatoires discrètes et continues et les lois de probabilités est proposé en annexe 3.

On définit la probabilité pour que X appartienne à un intervalle $[a; b]$ inclus dans Ω comme égale à l'aire située entre la courbe représentative de f et l'axe des abscisses, la probabilité que X appartienne à Ω étant égale à 1 (une unité d'aire).



On introduit aussi la fonction de répartition F définie sur Ω :

Si $\Omega = [\mathrm{a}; \mathrm{b}]$, alors pour tout $x \in [a; b]$, $F(x) = \int_a^x f(t)dt$

Si $\Omega = \mathbb{R}$, alors pour tout $x \in \mathbb{R}$, $F(x) = \int_{-\infty}^x f(t)dt$

Alors pour tout intervalle $J = [\, a\ ;\ b\,]$ inclus dans Ω, $P(X \in J) = P(a \leqslant X \leqslant b)$ est égale à l'aire comprise entre la courbe représentative de f sur l'intervalle J et l'axe des abscisses.

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

La loi d'une telle variable aléatoire est dite « loi à densité »(en raison de l'introduction de la fonction densité). La loi exponentielle présentée dans le paragraphe IV en est un exemple important.

L'espérance mathématique est donnée par : $E(X) = \int_{x \in \Omega} x f(x)dx$

2) La loi exponentielle

a) Introduction à la loi exponentielle

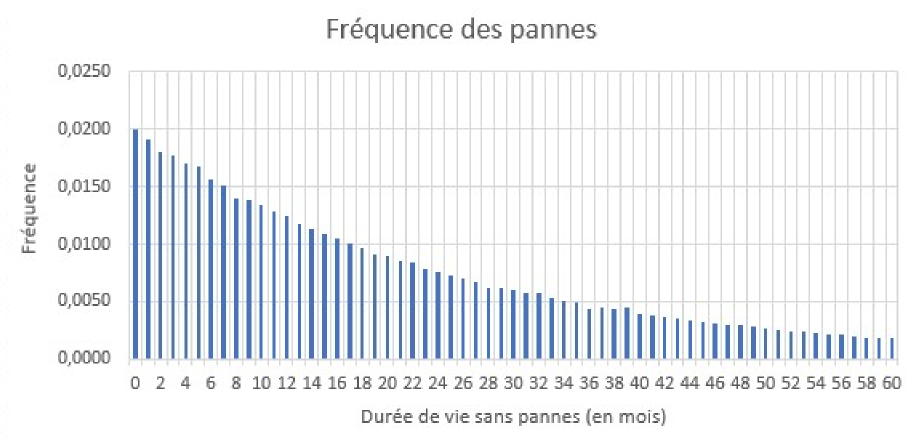
Inspiré de https://lecluseo.scenari-community.org/TS/Ch11LoisContinues_web_gen_auroraWSH.zip/co/ER_introLoiExp.html

i) Étude statistique d'une situation

Une entreprise fabriquant des téléviseurs a effectué **un suivi de la première panne des appareils qu'elle a fabriqués et vendus**.

On a réalisé ci-dessous un histogramme résumant les résultats (en abscisses la durée en mois et en ordonnées la fréquence). Les classes ont une amplitude de 1 mois.

Par exemple, 1,5% des appareils vendus ont subi leur première panne 6 mois après leur achat par le client.



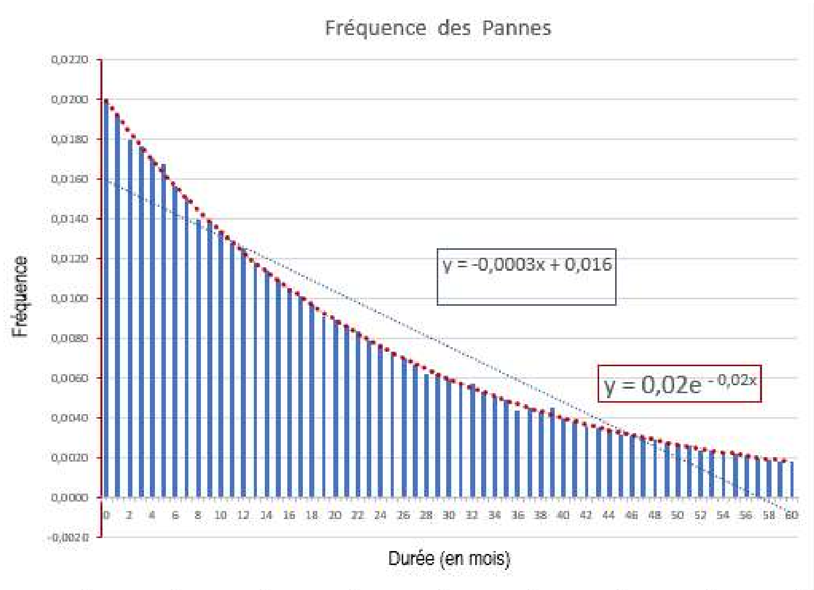
13

ii) Vers la modélisation par une loi de probabilité

On note T la variable aléatoire donnant l'instant (exprimé en mois) de la première panne d'un téléviseur. T est une variable aléatoire continue qui prend ses valeurs dans $[0; +\infty[$ (le téléviseur peut tomber en panne à n'importe quel moment à partir de sa mise en fonction).

$P(T \geq t)$ est la probabilité que la première panne d'un téléviseur survienne après plus de t mois.
En supposant que lors de la sortie le téléviseur était en état de marche, on a $P(T \geq 0) = 1$

On veut modéliser le relevé statistique ci-dessus par une loi de probabilité continue. La somme des aires des rectangles coloriés devrait correspondre à $P(0 \leq T \leq 60)$ et la courbe en rouge devrait correspondre à la densité de probabilité de la variable aléatoire T.



En utilisant le fichier Excel de relevé de données (lien Courbe de tendance.xlsx), on peut chercher une courbe de tendance (il y a différents choix). On a représenté un modèle « linéaire » et un modèle exponentiel, qui, lui, semble bien convenir.

En affichant l'équation de la courbe, la fonction densité serait donc f définie sur $[0; +\infty[$ par $f(t) = 0{,}02e^{-0{,}02\,t}$ (on a arrondi la valeur lue sur la courbe de tendance).

On montre que c'est bien une densité de probabilité sur l'intervalle $[0; +\infty[$, c'est-à-dire $\lim\limits_{x \to +\infty} \int_0^x f(t)dt = 1$

En effet la fonction définie sur $[0; +\infty[$ par $F(t) = -e^{-0{,}02}$ est une primitive de f donc $\int_0^x f(t)dt = -e^{-0{,}02x} + e^{-0{,}02\times 0} = 1 - e^{-0{,}02}$ et $\lim\limits_{x \to +\infty} e^{-0{,}02} = 0$.

On en déduit, par exemple, $P(0 \leq X \leq 10) = \int_0^{10} f(t)dt$ et $P(X > 10) = 1 - \int_0^{10} f(t)dt$

$$P(0 \leq X \leq 10) = \int_0^{10} 0{,}02e^{-0{,}02\,t}dt = (-e^{-0{,}02\times 10}) - (-e^{-0{,}02\times 0}) \approx 0{,}18$$

On peut le vérifier sur le fichier courbe de tendance en calculant la somme des fréquences de 0 à 10 (=SOMME(B2:B12))

Plus généralement : $P(0 \leq X \leq x) = \int_0^x f(t)dt = 1 - e^{-0{,}02\,x}$

Exercice

1) Déterminer la probalité que la première panne survienne au bout de 65 mois

14

2) a) Déteminer la probabilité que le téléviseur fonctionne sans panne plus de 65 mois sachant que le téléviseur n'a pas eu de panne avant le 60[ième] mois.

b) Comparer avec la probabilité que le téléviseur fonctionne encore plus de 5 mois.

Solution

1) $P(T > 65) = 1 - P(0 \leq X \leq 65) = e^{-0,02 \times 65} \approx 0,273$

2) a) $P_{(T>60)}(T > 65) = \frac{P(T>65 \text{ et } T>60)}{P(T>60)} = \frac{P(T>65)}{P(T>60)} = \frac{e^{-0,02 \times 65}}{e^{-0,02 \times 60}} \approx 0,9048$

b) $P(T > 5) = 1 - P(0 \leq X \leq 5) = e^{-0,02 \times 5} \approx 0,9048$. On trouve la même valeur approchée qu'à la question a).

Or $\frac{e^{-0,02 \times 65}}{e^{-0,02 \times 60}} = e^{-0,02 \times (65-60)} = e^{-0,02 \times 5}$ ce qui confirme que $P_{(T>60)}(T > 65) = P(T > 5)$

b) La loi exponentielle : la seule loi continue sans mémoire

i) Approche

Au paragraphe précédent, on a vu que le tableur propose une courbe de tendance exponentielle pour le relevé statistique et que, sous cette hypothèse, la loi de probabilité s'exprime effectivement à l'aide de la fonction exponentielle.
Ceci conduit à introduire pour une variable aléatoire, dans certaines conditions, la « Loi exponentielle »

> Distinguer
> « Loi exponentielle », qui est une loi de probabilité, et
> « Fonction exponentielle »

ii) Définition et propriétés

Certaines de ces propriétés peuvent être démontrées à partir de la densité de probabilité

* Une variable aléatoire *X* suit une loi de probabilité exponentielle de paramètre λ ($\lambda > 0$) si sa densité de probabilité f est définie par $\begin{cases} f(x) = \lambda e^{-\lambda x} \; si \; x \geq 0 \\ f(x) = 0 \; sinon \end{cases}$

* $P(X \leq x) = \int_0^x f(t)dt = 1 - e^{-\lambda x}$

La fonction *F* définie par $\begin{cases} F(x) = 1 - e^{-\lambda x} \; si \; x \geq 0 \\ F(x) = 0 \; sinon \end{cases}$ est appelée **fonction de répartition de X**

* $P(X > x) = 1 - P(X \leq x) = e^{-\lambda x}$

La fonction *G* définie par $G(x) = e^{-\lambda} \quad si \; x \geq 0$ est appelée fonction de survie.

* Propriété admise : $E(X) = \int_0^{+\infty} tf(t)dt = \frac{1}{\lambda}$

* Ci-dessous deux représentations de la loi exponentielle de paramètre 2 :

Fig 1 : l'aire colorée sous la courbe égale $P(X < 1,5)$

Fig 2 : l'aire sous la courbe égale $P(0,5 < X < 1,5)$

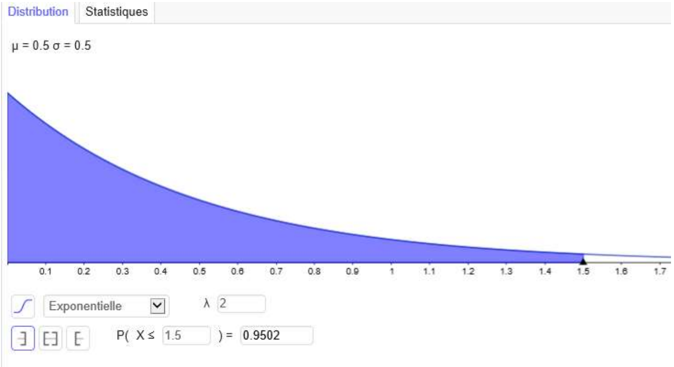


Fig 1

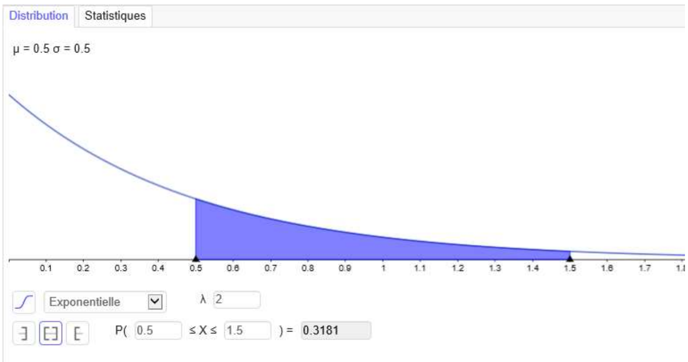


Fig 2

* La loi exponentielle est sans mémoire.
En effet, quels que soient s et t deux réels positifs

$$P_{(X>s)}(X > s + t) = \frac{P(X > t + s)}{P(X > s)} = \frac{e^{-(t+s)}}{e^{-s}} = e^{-t} = P(X > t)$$

iii) Montrons maintenant qu'une loi continue sans mémoire est une loi exponentielle.

Soit X une variable aléatoire définie sur $[0; +\infty[$ suivant une loi de probabilité sans mémoire.

On note F la fonction de répartition définie par $F(t) = P(X \leq t)$ et G la fonction définie sur $[0; +\infty[$ par $G(t) = P(X > t) = 1 - F(t)$.

Puisque X est sans mémoire, pour tous réels t et t', $P_{(X>t')}(X > t + t') = P(X > t)$

Ou encore, $\frac{P(X>t'+t)}{P(X>t')} = P(X > t)$. D'où $P(X > t' + t) = P(X > t') \times P(X > t)$

Ce qu'on peut écrire $G(t + t') = G(t) \times G(t')$. On reconnait une propriété de la fonction exponentielle étudiée en 1ère. On admet que c'est une propriété caractéristique des fonctions exponentielles $x \to e^{kx}$.

On a $G(0) = 1$ et $\lim_{t\to+\infty} G(t) = 0$
Donc on peut trouver un réel k tel que pour tout réel t positif, $G(t) = e^{kt}$ d'où $F(t) = 1 - e^{kt}$

Et, en dérivant, $f(t) = -ke^{kt}$. En posant $\lambda = -k$, on retrouve la définition d'une loi exponentielle.

c) Exemples de modélisation de temps d'attente continu

<u>Exercice 1</u>

La durée de vie d'un ordinateur portable exprimée en années est une variable aléatoire X suivant la loi exponentielle de paramètre λ = 0,125.

16

1) Déterminer la probabilité que la durée de vie de cet ordinateur portable dépasse 5 ans. (On arrondira les résultats au millième)

2) Déterminer la probabilité que la durée de vie de cet ordinateur portable soit inférieure à 3 ans. (On arrondira les résultats au millième)

Solution
1) P(X>5)= $e^{-0,125\times 5}$ ≈0,535
2) P(X<3)=1-$e^{-0,125\times 3} \approx 0,313$

Exercice 2
Le temps d'attente exprimé en minutes au guichet d'une banque est une variable aléatoire T suivant la loi exponentielle de paramètre λ. On sait que la probabilité qu'un client attende moins de 8 minutes est égale à 0,7.
1) Calculer une valeur approchée à 0,0001 de λ.
2) Calculer la probabilité qu'un client attende entre 15 et 20 minutes.

Solution
1) On sait que P(X<8)= 1-$e^{-\lambda\times 8}$ d'où 1-$e^{-\lambda\times 8} = 0,7$
$e^{-\lambda\times 8} = 0,3$ et $-8\lambda = ln 0,3$. On trouve $\lambda = \frac{\ln 0,3}{-8}$ ≈0,1505

Si les élèves n'ont pas étudié la fonction *ln* ils peuvent déterminer λ à l'aide de la calculatrice en tabulant $x \to e^{x}$

2) P(15<X<20)=$\int_{15}^{20} \lambda\, e^{-\lambda x} dx$= $\mathrm{P}(\mathrm{X} < 20) - P(X < 15) = e^{-15\lambda} - e^{-20\lambda}$ ≈0,0553

III- Temps d'attente – paradoxe du temps d'attente

Dans cette partie on reprend une notion qui, bien que portant le nom « temps d'attente » n'est pas associée (en général) à une variable aléatoire du type précédent. On explique ce qu'on peut trouver sous l'appellation « paradoxe de l'inspection », ou « paradoxe du temps d'attente à un arrêt d'autobus ». Ces paradoxes sont liés à la différence entre un temps moyen d'attente calculé sur des données globales (durée en jeu et nombre de bus en ligne par exemple), on pourrait parler d'un intervalle de temps moyen séparant les événements, et le temps d'attente effectivement« vécu » par un individu attendant le bus à un moment donné. Pour cela on associe ce dernier temps à une variable aléatoire, suivant une loi uniforme (qui n'est pas sans mémoire), alors que l'intervalle moyen du début est une simple moyenne.

1) Rappel sur la loi uniforme sur un intervalle [a ;b] (a et b réels distincts)
* Une variable *X* suit une loi uniforme sur $[a\,;\ b]$ si elle admet pour densité de probabilité la fonction *f* définie sur $[a\,;\ b]$ par $f(x) = \frac{1}{b-a}$
* La fonction de répartition est définie par, pour tout $x \in [a; b]$, $F(x) = \int_a^x f(t)dt = \frac{x-a}{b-a}$
* Pour tout intervalle $J = [c; d]$ inclus dans $[a\,;\ b]$, on a $P(c \le X \le d) = \frac{d-c}{b-a}$

* $E(X) = \frac{a+b}{2}$

<u>Exercice</u>

On cherche si la loi uniforme est une loi sans mémoire.
Soit *X* une variable aléatoire qui suit une loi uniforme sur un intervalle [a ; b].
Soit *c* un réel compris entre *a* et *b* et *t* un réel positif tel que $c + t$ appartient à l'intervalle [a ; b].
Calculer $P(X > t)$ et $P_{(X>c)}(X > c + t)$
La loi uniforme est-elle sans mémoire ?

<u>Solution</u>

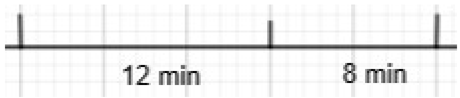
$(X > c + t)$ et $(X > c) \ = (X > c + t)$
Donc $P_{(X>c)}(X > c + t) = \frac{P(X>c+t)}{P(X>c)} = \frac{P(c+t<X<b)}{P(c<X<b)} = \frac{b-(c+t)}{b-c}$
D'autre part $P(X > t) = P(t < X < b) = \frac{b-t}{b-a}$ donc $P_{(X>c)}(X > c + t) \neq P(X > t)$, donc la loi n'est pas sans mémoire.

2) Temps d'attente à un arrêt d'autobus – paradoxe de l'inspection

Il y a une différence entre le temps moyen entre deux bus, mesuré par les ingénieurs de la ligne, et le temps moyen d'attente pour un individu qui prend le bus : c'est ce qu'on appelle le paradoxe de l'autobus (ou de l'inspection). En général le temps moyen entre deux bus ne permet pas de trouver le temps moyen d'attente d'un individu qui prend le bus si on ne connaît pas le rythme d'arrivée des bus et l'intervalle de temps pendant lequel arrive le passager ! Dans certains cas on peut cependant modéliser le rythme des bus en s'appuyant sur des statistiques.
1) Dans un premier temps considérons qu'à un arrêt d'autobus les bus arrivent régulièrement toutes les dix minutes. Nous nous rendons à cet arrêt à un instant aléatoire de la journée. Nous prenons le premier bus qui se présente. Combien de temps risque-t-on d'attendre ? En fait, nous pouvons attendre au maximum 10 minutes si nous arrivons juste après le passage d'un bus.
Mais combien de temps attend-on <u>en moyenne</u> ? Comme on a autant de chances d'arriver au début de cet intervalle qu'au milieu ou à la fin, on peut penser qu'en moyenne on attendra 5 minutes. En effet, l'attente est uniformément répartie dans cet intervalle de 10 minutes et si on considère que le temps d'attente est une variable aléatoire X qui suit la loi uniforme dans un intervalle de 10 minutes, l'attente moyenne est bien de 5 minutes : c'est l'espérance mathématique $\mathrm{E(X)}$ d'une variable aléatoire X qui suit une loi uniforme dans l'intervalle [0 ; 10], $\mathrm{E(X)} = \frac{10+0}{2} = 5$.
Et la probabilité d'attendre plus de 5 minutes est :
$P(X > 5) = P(5 < X \leq 10) = \frac{10-5}{10-0} = \frac{1}{2}$

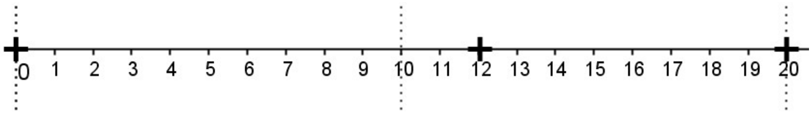
2) Considérons maintenant un arrêt de bus où l'intervalle de temps séparant l'arrivée de deux bus consécutifs est variable avec une moyenne de 10 minutes (temps moyen entre deux bus). Nous

nous rendons à un instant *complètement aléatoire de la journée à cet arrêt* et nous prenons le premier bus qui arrive. Combien de temps attend-t-on en moyenne (temps d'attente) ? En raisonnant comme précédemment, on a envie de répondre 5 minutes. **Ce raisonnement est faux dès que les bus ne sont pas strictement réguliers car, la plupart du temps, on a plus de chances d'arriver dans un grand intervalle de temps séparant deux bus que dans un petit intervalle.**
Supposons par exemple que le bus passe 1 fois sur 2 au bout de 12 minutes et 1 fois sur 2 au bout de 8 minutes (de fait, ce n'est guère réaliste, mais cela permet de raisonner).



Si l'on suppose en outre que la personne qui va prendre son bus dans cet intervalle de 20 minutes se rend à l'arrêt de manière aléatoire la probabilité pour elle d'arriver dans un intervalle de temps de 12 minutes est égale à $\frac{12}{20} = \frac{3}{5}$ et celle d'arriver dans un intervalle de 8 minutes est $\frac{8}{20} = \frac{2}{5}$.
Si on considère que le temps d'attente entre deux bus sur chacun des intervalles de temps suit une loi uniforme, alors le temps moyen d'attente lorsqu'on arrive dans un intervalle de 12 minutes est égal à 6 minutes ($\frac{12+0}{2}$) et celui dans un intervalle de 8 minutes est égal à 4 minutes.
Donc, **en moyenne**, on devrait attendre $6 \times \frac{3}{5} + 4 \times \frac{2}{5} = 5{,}2$ minutes (c'est une moyenne pondérée).

3) Peut-on retrouver ce résultat en modélisant la situation ?
Pour simplifier on va considérer que le bus passe aux instants t=0, t=12 et t= 20



et que le passager est susceptible d'arriver à tout instant dans cet intervalle [0;20], ce qu'on modélise ainsi : on appelle X la variable aléatoire qui désigne l'instant où arrive le passager exprimé en minutes et on considère que X suit une loi uniforme sur [0 ;20].
On définit la variable aléatoire Y égale au temps d'attente du passager jusqu'au passage d'un bus (en minutes).
Si $0 < X \leq 12$ alors $Y = 12 - X$ et $0 < Y \leq 12$ et
si $12 < X \leq 20$ alors $Y = 20 - X$ et $0 < Y \leq 8$.

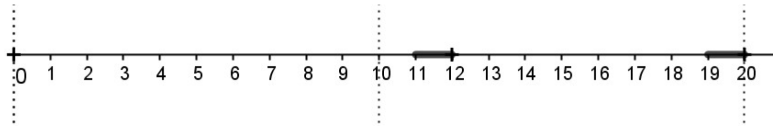
Finalement quand X varie dans l'intervalle [0 ;20] alors Y varie dans l'intervalle [0 ;12]
Puisqu'on a supposé que X suit une loi uniforme, on a $P(0 \leq X \leq 12) = \frac{12}{20}$ et $P(12 < X \leq 20) = \frac{8}{20}$
Pour simplifier l'écriture, on note A l'événement $(0 \leq X \leq 12)$ et B l'événement $(12 \leq X \leq 20)$
On cherche à déterminer la loi suivie par Y.
Pour avoir une idée, on peut calculer quelques probabilités de Y, par exemple, P(Y≤1) et P(Y>5)
On a $P(Y \leq 1) = P((Y \leq 1) \cap A) + P((Y \leq 1) \cap B)$ car A et B sont disjoints.

$$P(0 \leq Y \leq 1) = P\big((Y \leq 1) \cap A\big) + P\big((Y \leq 1) \cap B\big) = P(11 < X < 12)$$
$$+ P(19 < X < 20) = \frac{1}{20} + \frac{1}{20} = \frac{1}{10}$$

Remarques

- *On peut écrire l'égalité des événements :*
  $\big((Y \leq 1) \cap A\big) = (Y \leq 1) \cap (0 \leq X \leq 12) = (12 - X \leq 1) \cap (0 \leq X \leq 12) = (X \geq 11) \cap (0 \leq X \leq 12) = (11 \leq X \leq 12)$.
  De même :
  $$\big((Y \leq 1) \cap B\big) = (20 - X \leq 1) \cap (12 \leq X \leq 20) = (X \geq 19) \cap (12 \leq X \leq 20)$$
$= (19 \leq X \leq 20)$

On peut aussi s'aider du schéma : si on attend au plus 1 minute, cela veut dire qu'on est arrivé, soit entre les minutes 11 et 12, soit entre les minutes 19 et 20



- *On sait que* $P(Y = 1) = 0$ *donc* $P(Y < 1) = P(Y \leq 1)$. *Ainsi on utilise indifféremment des inégalités larges ou strictes.*

De même,
$$P(Y > 5) = P\big((Y > 5) \cap A\big) + P\big((Y > 5) \cap B\big) = P(0 < X < 7) + P(12 < X < 15)$$
$$= \frac{7}{20} + \frac{3}{20} = \frac{1}{2}$$

On peut définir et représenter la **fonction de répartition** de la loi suivie par Y c'est-à-dire la fonction F définie par $F(t) = P(Y \leq t)$ pour tout *t* réel élément de [0 ; 12]. On va distinguer deux cas suivant que t appartient à [0 ;8] ou à [8 ;12].

1^er^ cas : t appartient à [0 ; 8]

$P(Y \leq t) = P((Y \leq t) \cap A) + P((Y \leq t) \cap B)$

Or $(Y \leq t) \cap A = (12 - X \leq t) \cap (0 \leq X \leq 12) = (12 - t \leq X \leq 12)$

$(Y \leq t) \cap B = (20 - X \leq t) \cap (12 \leq X \leq 20) = (20 - t \leq X \leq 20)$

On obtient $P(Y \leq t) = P(12 - t \leq X \leq 12) + P(20 - t \leq X \leq 20) = \frac{t}{20} + \frac{t}{20} = \frac{t}{10}$

2ème cas : t appartient à]8 ;12]

On calcule d'abord $P(Y > t)$ pour déterminer $P(Y \leq t)$

$P(Y > t) = P((Y > t) \cap A) + P((Y > t) \cap B) = P(0 < X < 12 - t) + 0 = \frac{12-t}{20}$

On en déduit $P(Y \leq t) = 1 - P(Y > t) = \frac{8+t}{20}$

Finalement, la fonction F est définie par :

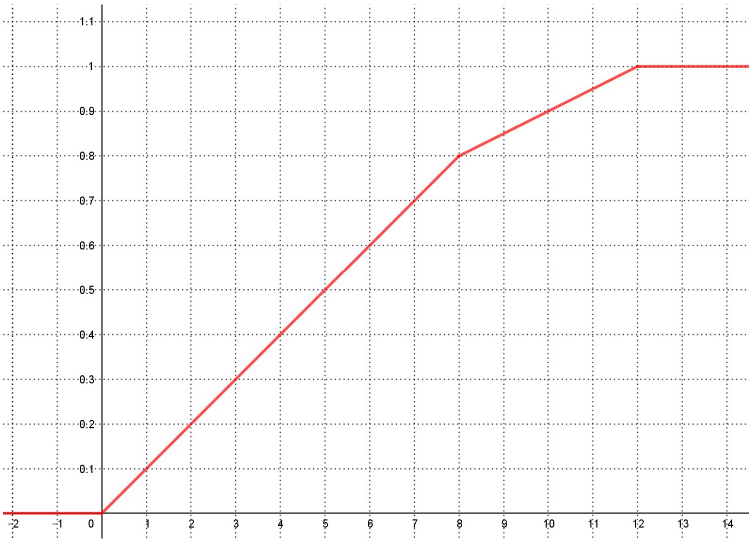
$F(t) = 0$ pour tout réel t de $]-\infty; 0]$

$F(t) = \frac{t}{10}$ pour tout t de]0 ;8]

$F(t) = \frac{8+t}{20}$ pour tout t de]8 ;12]

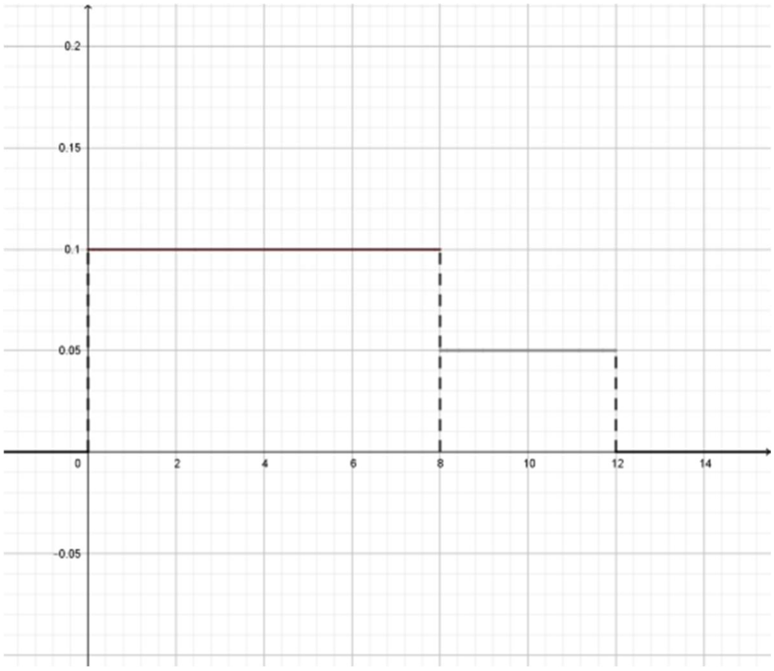
$F(t) = 1$ pour tout t supérieur à 12

D'où sa représentation graphique :



On peut lire graphiquement, par exemple la probabilité d'attendre au maximum 10 minutes $P(Y \leq 10)$ qui est égale à 0,9. (On lit l'ordonnée du point de la courbe d'abscisse 10).
On remarque que la fonction F est croissante (comme toute fonction de répartition) et affine par intervalles (ou par morceaux).

On obtient la **fonction de densité** f de la variable aléatoire Y en dérivant F.
Pour tout t de $]-\infty;0[$ ou de $]12;+\infty[$, on a $f(t) = 0$
Pour tout t de]0 ;8[,$f(t) = \frac{1}{10}$
Pour tout t de]8 ;12[,$f(t) = \frac{1}{20}$
On peut définir les images de 0 ; 8 et 12 de multiples façons, par exemple
$f(0) = 0, f(8) = \frac{1}{10}\ et\ f(12) = \frac{1}{20}$.
Les différentes propriétés des fonctions de répartition et de densité sont hors programmes et ne seront pas étudiées ici (cf Yves Ducel, conversation privée)
La fonction f est représentée ci-dessous.



L'espérance de Y est donnée par :

E(Y)= $\int_0^8 xf(x)dx + \int_8^{12} xf(x)dx$ = $\int_0^8 \frac{2}{20} xdx + \int_8^{12} \frac{1}{20} xdx = \frac{26}{5} = 5{,}2$ minutes.

Donc, **en moyenne**, le passager attend 5,2 minutes.
On retrouve bien le résultat « intuitif » du calcul de la moyenne vue comme moyenne pondérée de deux moyennes $6 \times \frac{3}{5} + 4 \times \frac{2}{5} = 5{,}2$

On peut se demander si le fait que les deux méthodes donnent le même résultat n'est pas le fait d'une coïncidence numérique : on trouvera en annexe la preuve que ce résultat est général (voir annexe 4).

4) Dans le raisonnement précédent on s'est placé dans le cas de bus qui arrivent 1 fois sur 2 au bout de 12 minutes et 1 fois sur 2 au bout de 8 minutes et où le passager arrive dans une séquence « 12 min/ 8 min ». On peut se demander si le résultat est le même dans plusieurs séquences de 12 et 8 réparties différemment, par exemple :

On aura encore, en moyenne, un temps d'attente moyen pour la personne qui arrive dans cette période d'une heure ($3 \times 12 + 3 \times 8 = 60$ minutes) de $6 \times \frac{36}{60} + 4 \times \frac{24}{60} = 5{,}2$

En fait si on ne connait pas le rythme de passage des bus on ne peut rien dire. On peut cependant s'appuyer sur des relevés statistiques et modéliser le temps d'attente.

5) Exemple (inspiré de http://accromath.uqam.ca/volume/volume-15-1/)

Le professeur Tournesol part entre 15h03 et 15h59 aujourd'hui. Pour simplifier nos calculs, on arrondit à la minute près et on considère que sa sortie suit une loi uniforme entre 15h03 et 15h59 (il part à n'importe quel moment dans cet intervalle de temps). S'il arrive à la même heure qu'un bus, ou avant, il peut prendre ce bus. Comme il y a 56 minutes dans cet intervalle, il a une probabilité égale d'arrivée à chacune des minutes. Le tableau ci-contre indique les heures de passage des autobus et les durées depuis le dernier bus précédent (colonne T). En colonne A figurent les temps moyens correspondants en considérant que l'attente dans chacun de ces intervalles de temps suit une loi uniforme (comme dans les cas précédents). La probabilité de prendre un bus arrivé k minutes après le bus précédent est de $\frac{k}{56}$=*P(b)*. C'est ainsi que le professeur Tournesol réalise que plus le temps d'attente entre deux autobus est long, plus il a de chance d'arriver à l'arrêt de bus pendant cet intervalle de temps. Avec son raisonnement initial, il prévoyait un temps moyen d'attente égal à la moitié de la durée moyenne entre deux bus, soit $\frac{6{,}22}{2}$ = 3,11 minutes ($\frac{18+2+4+1+3+6+3+11+}{9} = 6{,}22$) selon les données collectées, alors que la moyenne pondérée par la probabilité de prendre chaque bus est plutôt de 5,21 minutes

$(0{,}321 \times 9 + 0{,}036 \times 1 + 0{,}071 \times 2 + \cdots + 0{,}143 \times 4\,)$.

Tableau 1

Calcul du temps d'attente pondéré par la durée de passage entre les bus.

Heure du bus	*T*	*A*	*P(b)*
15h21	18	9	0,321
15h23	2	1	0,036
15h27	4	2	0,071
15h28	1	0,5	0,018
15h31	3	1,5	0,054
15h37	6	3	0,107
15h40	3	1,5	0,054
15h51	11	5,5	0,196
15h59	8	4	0,143
Attente moyenne pondérée :			5,214

Légende

T : temps depuis le dernier bus,

A : attente moyenne,

P(b) : probabilité de prendre ce bus.

6) Une autre modélisation

Si on assimile l'intervalle de temps entre deux bus à une variable aléatoire X qui suit la loi exponentielle de paramètre $\lambda=\frac{1}{6{,}22}$, ce qui constitue une autre modélisation, la probabilité que le bus arrive dans plus de 3,11 minutes (le temps estimé par le professeur Tournesol) est

$P(X>3{,}11)=e^{-\frac{1}{6{,}22}\times 3{,}11} \approx 0{,}61$

En utilisant ce modèle, la probabilité que le temps d'attente de Tournesol soit supérieure, par exemple, à 10 minutes est égale à $e^{-\frac{1}{6{,}22}\times 10} \approx$ *0,20* ce qui n'est pas négligeable !

Un tel modèle est-il plausible ?

On note $t_1, t_2, t_3 \ldots, t_n \ldots$ les instants aléatoires où l'on observe l'arrivée d'un bus et soient $X_1, X_2, X_3 \ldots X_n \ldots$ les durées aléatoires entre deux instants où l'on observe l'arrivée d'un bus.

On a donc $X_1 = t_1$, $X_2 = t_2 - t_1$, $X_3 = t_3 - t_2$, $X_n = t_n - t_{n-1}$…

On suppose que les variables aléatoires X_i suivent la même loi et, par suite, que quel que soit i, les $P(X_i \leq t)$ sont égales et que les arrivées des bus sont indépendantes, c'est-à-dire que les X_i sont indépendantes.

Soit $g(t) = 1 - P(X \leq t)$ la probabilité qu'aucun bus n'arrive entre les instants 0 et t. $g(t) = P(X > t)$

$g(t+s)$ est la probabilité qu'aucun bus n'arrive entre les instants 0 et t+s. Or l'événement « il n'y a pas de bus entre les instants 0 et $t+s$ » est l'intersection des événements « il n'y a pas de bus entre les instants 0 et t » et « il n'y pas de bus entre les instants t et *t+s* ». On fait l'hypothèse que

1) ces deux événements sont indépendants

2) la probabilité qu'il n'y ait pas de bus entre t et $t+s$ ne dépend pas de t et est donc égale à la probabilité qu'il n'y ait pas de bus entre 0 et 0+s soit g(s)

donc $g(t+s) = g(t) \times g(s)$

On reconnait une propriété de la fonction exponentielle étudiée en 1ère. On admet que c'est une propriété caractéristique des fonctions exponentielles $x \to e^{kx}$.

De plus $g(0) = 1$ puisque c'est la probabilité qu'il n'y ait pas d'autobus avant l'instant 0.

Sous les hypothèses précédentes, on peut donc modéliser la probabilité d'arrivée d'un bus, à un instant supérieur à un instant donné *t*, et donc le temps d'attente pour une personne qui se présente à un arrêt d'autobus, par une loi exponentielle dont le paramètre est à déterminer à partir de données statistiques.

V – Conclusion : différents « temps d'attente »

Le terme « temps d'attente » est associé à plusieurs notions différentes.
Comme on l'a évoqué dans le premier paragraphe, on peut penser à une première acception, liée au sens commun, et source d'erreurs.
Dans le cas discret, le temps d'attente d'un événement aléatoire, comme le nombre de lancers nécessaires pour obtenir un 6 avec un dé, est modélisé par une loi géométrique qui est une loi sans mémoire ; ainsi, contrairement à ce que l'on pourrait penser, ce n'est pas parce qu'un événement ne s'est pas produit depuis longtemps que le temps d'attente va être plus court !
Dans le cas continu, on a cherché à définir le temps d'attente et on a été amené à étudier ce qu'on appelle le « paradoxe de l'autobus » : c'est un cas particulier d'un problème plus général souvent appelé « paradoxe de l'inspection ». De quelle inspection s'agit-il ? Une tâche répétitive doit être effectuée par un groupe d'employés et un inspecteur se rend sur le lieu de travail à différents moments afin de connaitre le temps requis pour effectuer cette tâche. On peut encore parler de temps d'attente (avant que la tâche soit réalisée) mais, comme pour l'autobus, il faut distinguer deux temps d'attente dont on peut penser qu'ils sont égaux mais qui s'avèrent différents :

- Le temps d'attente estimé en calculant une moyenne pondérée des durées (à partir des écarts entre deux bus indiqués par la compagnie ou à partir d'une durée moyenne de la tâche prévue a priori)
- Le temps d'attente moyen mesuré, soit par un passager qui se rend aléatoirement à l'arrêt, soit par un inspecteur qui évalue à plusieurs reprises le temps mis pour effectuer la tâche.

On constate, et cela semble paradoxal, que même si les mesures sont très nombreuses, le temps d'attente moyen obtenu à partir de ces mesures est plus long que celui qu'on avait estimé par calcul (sauf si les bus ou les tâches sont parfaitement réguliers, ce qui n'est pas le cas qui nous intéresse).
En réalité le paradoxe n'est qu'apparent, il vient du fait que les mesures ne sont pas complètement aléatoires mais dépendent du moment où on les réalise. En effet, comme on l'a vu, lorsqu'on se rend à l'arrêt de bus, on a plus de chances d'arriver lors d'un temps d'attente long (plus long que la moyenne) que lors d'un temps d'attente court. De même, l'inspection a plus de chances de se dérouler pendant les traitements de tâche les plus longs (ce qui peut faire croire, à tort, que les employés ne sont pas assez productifs).
C'est ce que l'on appelle un « biais ». Cette confusion dans l'observation d'un phénomène peut avoir des conséquences importantes et on la retrouve fréquemment (par exemple en génétique, à la Bourse etc. voir http://accromath.uqam.ca). Elle montre la difficulté pour un individu de reconstituer la réalité à partir de son expérience[5].
On a pu modéliser certaines situations en introduisant la loi uniforme ou plus souvent la loi exponentielle qui est une loi sans mémoire mais il sera nécessaire d'avoir recours à d'autres outils mathématiques pour étudier complètement les phénomènes en jeu. D'autres modélisations pourraient intervenir en introduisant par exemple la loi de Poisson qui n'est pas disponible pour les élèves.

VI - Exercices : des situations où interviennent ces lois

Dans un certain nombre d'exercices la connaissance de la fonction *ln* facilite les calculs, cependant les élèves peuvent recourir à la calculatrice pour répondre aux questions.

A- Exercices où intervient la loi géométrique

Ex 1 www.gerad.ca/Sebastien.Le.Digabel/MTH2302D/5_lois_discretes.pdf

Exercice résolu

On lance un dé jusqu'à l'obtention d'un 6. Soit X le nombre de lancers nécessaires.
1. Quelle est la probabilité d'obtenir un 6 pour la première fois au deuxième lancer ?

2. Quelle est la probabilité qu'il faille plus de 10 lancers pour obtenir un 6 ?

3. Si aucun 6 n'a été obtenu lors des 8 premiers lancers, quelle est la probabilité qu'au moins deux autres lancers soient nécessaires ?

Solution

[5] https://www.blablasciences.com

On reconnait que X est une variable aléatoire qui suit une loi géométrique de paramètre p=$\frac{1}{6}$

1. Si on obtient un premier 6 au deuxième lancer c'est qu'on a obtenu un autre nombre au premier lancer. On calcule donc P(X=2). P(X=2)= $\frac{5}{6} \times \frac{1}{6} = \frac{5}{36}$
2. P(X>10)= 1-(P(X=1)+P(X=2)+…+P(X=10))=1-($\frac{1}{6} + \frac{1}{6} \times \frac{5}{6} + \cdots + \frac{1}{6} \times (\frac{5}{6})^9$)= $1 - \frac{1}{6}(1 + \frac{5}{6} + \cdots (\frac{5}{6})^9)$
   On reconnaît dans la parenthèse la somme des 10 premiers termes de la suite géométrique de raison $\frac{5}{6}$ et de premier terme 1.
   P(X>10)=1-$\frac{1}{6} \times \frac{1-(\frac{5}{6})^{10}}{1-\frac{5}{6}}$ = $(\frac{5}{6})^{10} \approx 0{,}162$ (On peut aussi utiliser directement le résultat du cours)
3. Si aucun 6 n'a été obtenu lors des 8 premiers lancers, la probabilité qu'au moins deux autres lancers soient nécessaires est égale à $P_{(X>8)}(X \geq 10) = P_{(X>8)}(X > 9)$
   La loi géométrique étant une loi sans mémoire :
   $P_{(X>8)}(X > 9) = P_{(X>8)}(X > 8 + 1) = P(X > 1) = \frac{5}{6}$

Ex 2 http://pgoutet.free.fr/td/probasL2_va_discretes.pdf

On suppose que le temps d'attente (en minutes) d'un métro suit une loi géométrique. Durant les heures de pointes du matin, le temps d'attente moyen d'un métro pour la ligne 8 est de 3 minutes tandis qu'il est de 2 min pour la ligne 9.

(a) Quels sont les paramètres des lois géométriques pour les lignes 8 et 9 ?

(b) Quelles sont les probabilités d'attendre entre 2 et 4 minutes un métro pour chacune des deux lignes ?

(c) Quelles sont les probabilités d'attendre plus de 5 minutes pour chacune des deux lignes ?

Ex 3 d'après http://pgoutet.free.fr/td/probasL2_va_discretes.pdf

En France, on estime que le groupe sanguin de 3% de la population est le groupe AB.

Dans un hôpital, l'état d'un patient de groupe AB, nécessite une transfusion.

Des personnes se présentent en très grand nombre pour donner leur sang (on suppose que les groupes sanguins des personnes sont indépendants les uns des autres et que le nombre de personnes n'est pas limité). On prélève une poche de sang à chaque donneur. Les prélèvements se succèdent toutes les 10 minutes.

1) On appelle X la variable aléatoire qui donne le rang du premier prélèvement de sang de groupe AB. Quelle loi suit X ?
2) Quelle est la probabilité d'attendre au maximum une heure avant d'avoir une poche de sang du groupe AB ?
3) Quelle est la probabilité d'attendre plus de 2 heures avant d'avoir une poche de sang du groupe AB ?
4) Sachant qu'on n'a pas eu de donneur du groupe AB pendant la première heure, quelle est la probabilité d'attendre encore plus de 2 heures ? Comparer avec le résultat de la question 2) et expliquer ce résultat.

5) Quelle est la probabilité de n'obtenir aucune poche de sang du groupe AB à la fin de la journée c'est-à-dire après 80 donneurs?
6) Au Japon, 10% de la population est du groupe AB. Reprendre les questions 1) à 5) dans ce pays et expliquer les différences de résultats obtenus.

EX 4 http://www.jybaudot.fr/Probas/loigeometrique.html

Un photographe animalier se poste là où un chat sauvage passe parfois à l'aube, de façon totalement aléatoire. L'un de ses collègues lui a affirmé qu'au cours des deux mois précédents le félin était passé six fois et il en a conclu que la probabilité d'apparition du chat un matin donné s'établissait à 0,1.
Le photographe ne peut consacrer que cinq jours à cet affût mais il partira avant s'il a pu photographier le chat. Selon les services de la météo, il pleuvra le deuxième jour. Sinon, il fera beau temps.

1. Quelle est la probabilité de photographier le chat sous la pluie ?
2. Quelle est la probabilité de photographier le chat au cours de ces cinq jours ?

Indication

La situation peut être modélisée par une loi géométrique avec une probabilité de succès égale à 0,1. En effet, les passages du chat s'apparentent à une suite d'épreuves de Bernoulli identiques et indépendantes et la variable aléatoire *X* est le nombre de jours nécessaires pour photographier le chat.

B - Exercices où intervient la loi exponentielle

Ex 5 -Durée d'attente pour entrer dans un parking souterrain (Bac 2017)

Exercice résolu

On appelle durée d'attente le temps qui s'écoule entre le moment où une voiture se présente à l'entrée du parking et le moment où elle franchit la barrière d'entrée du parking.

Le tableau suivant présente les observations faites sur une journée :

Durée d'attente en minutes	[0 ; 2[	[2 ; 4[	[4 ; 6[	[6 ; 8[
Nombre de voitures	75	19	10	5

1) Proposer une estimation de la durée d'attente moyenne d'une voiture à l'entrée du parking.
2) On décide de modéliser cette durée d'attente (exprimée en minutes) par une variable aléatoire T qui suit une loi exponentielle de paramètre λ.
   a. Justifier qu'on peut choisir λ=0,5.
   b. Une voiture se présente à l'entrée du parking. Quelle est la probabilité pour qu'elle mette moins de 2 minutes pour franchir la barrière ? (On donnera les résultats avec une précision de 10^{-4}.)
   c. Une voiture attend devant la barrière depuis une minute. Quelle est la probabilité pour qu'elle franchisse la barrière la minute suivante ?

Solution

1) La durée d'attente moyenne d'une voiture est environ égale à 1,991 min que l'on arrondit à 2 min environ.
($\frac{75\times1+19\times3+10\times5+5\times7)}{75+19+10+5}=\frac{217}{109}$)
2) a. T suit une loi exponentielle de paramètre λ. On considère que la durée moyenne d'attente est égale à l'espérance mathématique de la loi suivie par T de paramètre λ. On sait que l'espérance est égale à $\frac{1}{\lambda}$. On a donc $\frac{1}{\lambda}=2$ d'où $\lambda=0,5$.
b. On calcule P(T<2)= 1- $e^{-0,5\times2}$ ≈0,6321
c. On cherche $P_{(T>1)}(T<2)=\frac{P(1<T<2)}{P(T>1)}$ (probabilité conditionnelle)
Or T suit une loi exponentielle donc $P(1<T<2)=P(0<T<2)-P(0<T<1)$
Et en utilisant les propriétés de la loi exponentielle, on a
$\frac{P(1<T<2)}{P(T>1)}=\frac{(1-e^{-1})-(1-e^{-0,5})}{e^{-0,5}}=\frac{(e^{-0,5}-e^{-1})}{e^{-0,5}}=1-e^{-1+0,5}=1-e^{-0,5}$ ≈0,3935

<u>Remarque</u> : on pourrait aussi écrire $P(1<T<2)=\int_1^2(1-e^{-0,5t})dt$

<u>Autre méthode (Annales APMEP Bac S 2017 Liban)</u>
On cherche $P_{(T>1)}(T<2)$. T suit une loi exponentielle donc une loi de durée de vie sans vieillissement. Donc $P_{(T>1)}(T>2)=P(T>1)$ car 2=1+1.
En passant au complémentaire, on a $P_{(T>1)}(T<2)=P(T<1)=1-e^{-0,5\times1}\approx 0,3935$

Exercice 6 - Extrait d'un vrai/faux (Bac 2003)

La durée d'attente en secondes à la caisse d'un supermarché est une variable aléatoire Y qui suit la loi exponentielle de paramètre 0,01.
Déterminer pour chacune des affirmations suivantes si elle est vraie ou fausse.
a. La densité de probabilité de Y est la fonction f définie sur [0 ; +∞[par f(t)=$e^{-0,01t}$;
b. Pour tout réel t positif, on a : P(Y≤t) = 1– $e^{-0,01t}$;
c. La probabilité d'attendre moins de 3 minutes à cette caisse est égale à 0,16 au centième près ;
d. Il y a plus d'une chance sur deux que l'attente à cette caisse soit supérieure à une minute.

Exercice 7 - La compagnie d'autocars (Bac 2003)
Une entreprise d'autocars dessert une région montagneuse. En chemin, les véhicules peuvent être bloqués par des incidents extérieurs comme des chutes de pierres, des troupeaux sur la route, etc. Un autocar part de son entrepôt. On note D la variable aléatoire qui mesure la distance en kilomètres que l'autocar va parcourir jusqu'à ce qu'il survienne un incident. On admet que la variable D suit une loi exponentielle de paramètre $\lambda=\frac{1}{82}$
1) Calculer la probabilité pour que la distance parcourue sans incident soit :
a) comprise entre 50 et 100 km
b) supérieure à 300 km
2) Sachant que l'autocar a déjà parcouru 350 km sans incident, quelle est la probabilité qu'il n'en subisse pas non plus au cours des 25 prochains km ?
3) Déterminer la distance moyenne parcourue sans incident.

Exercice 8 - Durée de vie d'un composant électronique (Bac 2004)

On s'intéresse à la durée de vie, exprimée en semaines, d'un composant électronique. On note X la variable aléatoire qui mesure cette durée de vie en semaines. On suppose que X suit une loi exponentielle de paramètre λ définie sur l'intervalle [0 ; +∞[. Une étude statistique montrant qu'environ 50% d'un lot important de ces composants sont encore en état de marche au bout de 200 semaines, permet de poser P([0 ; 200[) = 0,5.

1) Vérifier que $\lambda=\frac{ln}{200}$.
2) Quelle est la probabilité qu'un de ces composants pris au hasard ait une durée de vie supérieure à 300 semaines ?
3) Déterminer la durée de vie moyenne dm de ces composants.

Si les élèves n'ont pas étudié la fonction *ln* ils peuvent déterminer λ à l'aide de la calculatrice en tabulant $x \to e^x$

Ex 9 - Bac 2017

Romane utilise deux modes de déplacement pour se déplacer entre son domicile et son lieu de travail : le vélo ou les transports en commun.

En hiver, Romane roule en vélo de nuit. Son vélo est visible grâce à une ampoule dont la durée de fonctionnement en heures peut être modélisée par une variable aléatoire, notée X, suivant une loi exponentielle de paramètre λ, réel strictement positif. La fonction de densité associée est donc la fonction f définie sur [0 ; +∞[par $f(t) = \lambda e^{-\lambda t}$.

1. Soit b un réel positif. Démontrer, à l'aide d'une intégrale, que $P(X \leq b) = 1-e^{-\lambda b}$.

2. On sait que la probabilité que l'ampoule fonctionne encore après 50 heures d'utilisation est 0,9.

a. En déduire la valeur exacte de λ.

b. Calculer la probabilité que la durée de fonctionnement de l'ampoule soit supérieure à 250 heures sachant que l'ampoule a déjà fonctionné 200 heures.

Ex 10 - Bac 2016

Une usine fabrique un composant électonique. Deux chaînes de fabrications sont utilisées. La chaîne A produit 40% des composants et la chaîne B produit le reste.
Une partie des composants fabriqués présente un défaut qui les empêche de fonctionner à la vitesse prévue par le constructeur. En sortie de la chaîne A, 20% des composants présentent ce défautnalors qu'en sortie de la chaîne B, ils ne sont que 5%.
On choisit au hasard un composant fabriqué dans cette usine.
On note :
A l'événement « le composant provient de la chaîne A »,
B l'événement « le composant provient de la chaîne B »,
S l'événement « le composant est sans défaut ».

Partie A

Montrer que la probabilité de S est P(S)=0,89

Partie B

La durée de vie, en années, d'un composant électronique fabriqué dans cette usine est une variable aléatoire T qui suit la loi exponentielle de paramètre λ.

On note f la fonction densité associée à la variable aléatoire T. (On rappelle que pour tout nombre réel x≥0, $f(x)= \lambda e^{-\lambda x}$ et que pour tout nombre réel a≥0, $P(T\leq a)= \int_0^a f(x)dx$)

1) On suppose que $P(T\leq 7)$ =0,5. Déterminer λ à 10^{-1}près.
2) Dans cette question λ = 0,099 et on arrondit les résultats des probabilités au centième.

a. On choisit au hasard un composant fabriqué dans cette usine. Déterminer la probabilité que ce composant fonctionne au moins 5 ans.

b. On choisit au hasard un composant parmi ceux qui fonctionnent encore au bout de 2 ans. Déterminer la probabilité que ce composant ait une durée de vie supérieure à 7 ans.

c. Donner l'espérance mathématique E(T) de la variable aléatoire T à l'unité près.

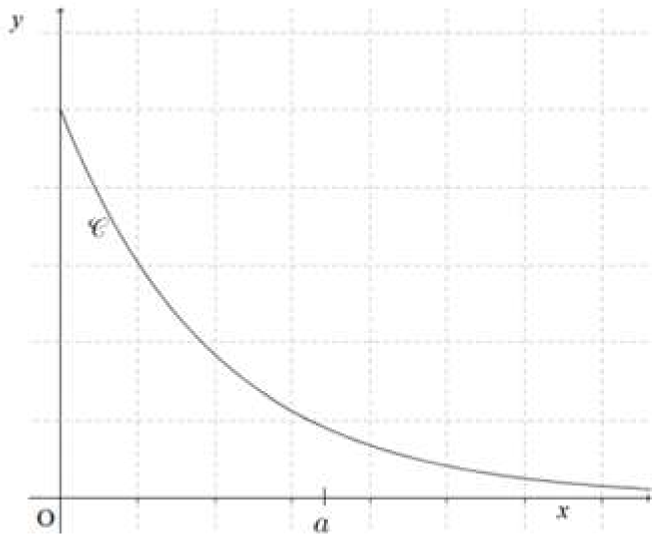
d. Interpréter ce résultat.

3) La courbe représentative de la fonction f est donnée ci-dessous.

   a. Interpréter graphiquement $P(T\leq a)$ où a>0.

   b. Montrer que pour tout nombre réel t≥0, $P(T\leq t)$=1-$e^{-\lambda}$

   c. En déduire que $\lim_{t\to+\infty} P(T\leq t) = 1$



Exercice 11 - Durée de vie d'un oscilloscope (Bac 2004)

Le laboratoire de physique d'un lycée dispose d'un parc d'oscilloscopes identiques. La durée de vie en années d'un oscilloscope est une variable aléatoire notée X qui suit la loi de durée de vie sans vieillissement ou encore loi exponentielle de paramètre λ>0.

1) Sachant que P(X>10) = 0,286 , montrer que λ = 0,125 au centième près. Dans la suite de l'exercice, on prendra λ = 0,125.
2) Calculer la probabilité qu'un oscilloscope du modèle étudié ait une durée de vie inférieure à 6 mois.
3) Sachant qu'un appareil a déjà fonctionné 8 années, quelle est la probabilité qu'il ait une durée de vie supérieure à 10 ans.

4) On considère que la durée de vie d'un oscilloscope est indépendante de celle des autres appareils. Le responsable du laboratoire décide de commander 15 oscilloscopes. Quelle est la probabilité qu'au moins un oscilloscope ait une durée de vie supérieure à 10 ans ?
5) Combien l'établissement devrait-il acheter d'oscilloscopes pour que la probabilité qu'au moins l'un d'entre eux fonctionne plus de 10 ans soit supérieure à 0,999 ?

Exercice 12 -Temps d'attente à un standard téléphonique (D'après Bac 2018)
Les quinze jours précédant la rentrée universitaire, le standard téléphonique d'une mutuelle étudiante enregistre un nombre record d'appels. Les appelants sont d'abord mis en attente et entendent une musique d'ambiance et un message préenregistré. Lors de cette première phase, le temps d'attente, exprimé en secondes, est modélisé par la variable aléatoire X qui suit la loi exponentielle de paramètre $\lambda = 0,02$ s. Les appelants sont ensuite mis en relation avec un chargé de clientèle.

1. Quelle est la durée totale moyenne du temps d'attente ?
2. Un étudiant est choisi au hasard parmi les appelants du standard téléphonique. Calculer la probabilité que l'étudiant soit mis en attente plus de 2 minutes.
3. Une étudiante, choisie au hasard parmi les appelants, attend depuis plus d'une minute d'être mise en relation avec le service clientèle. Lasse, elle raccroche et recompose le numéro. Elle espère attendre moins de trente secondes cette fois-ci. Le fait de raccrocher puis de rappeler augmente-t-il ses chances de limiter à 30 secondes l'attente supplémentaire ou bien aurait-elle mieux fait de rester en ligne ?

C - Exercices où intervient la loi uniforme

Ex 13 - Rendez-vous à la bibliothèque (exercice résolu)

Pierre étudie à la bibliothèque entre 14h30 et 16h. Jean y arrive, selon ses disponibilités, entre 13h30 et 17h30.

1) On appelle *X* la variable aléatoire « temps d'arrivée de Jean en minutes de 13h30 à 17h30». Sur quel ensemble est-elle définie ?
2) Quelle est la loi de probabilité suivie par X ?
3) Quelle est la probabilité que Jean et Pierre se rencontrent ?
4) Jean n'est toujours pas arrivé à 15h. Quelle est la probabilité qu'ils se rencontrent ?
5) Quelle est l'heure moyenne (espérée) de la rencontre ?

Solution

Pierre			14h30	15h	16h			
Jean	13h30							17h30
X	0		60	90	150			240

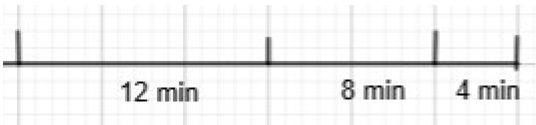
1) X est définie sur l'intervalle [0 ; 240]
2) X suit une loi uniforme.
3) P(60≤ $X \leq 150$)=$\frac{150-6}{240}$ =0,375
4) $P_{X>90}$(60≤ $X \leq 150$ =$\frac{P(90\leq X\leq 150)}{P(90\leq X\leq 240)}$=$\frac{150-}{240-90}$ =0,4

Les deux derniers résultats apparaissent clairement sur le schéma qui est une aide utile, dont il ne faut pas se priver. On lit ci-dessus 0,375=$\frac{3}{8}$ et 0,4=$\frac{2}{5}$

L'heure moyenne probable de la rencontre est égale à l'espérance mathématique de X c'est-à-dire $\frac{240-0}{2}$ = 120 ce qui donne une heure de 15h30.

Ex 14 – Temps d'attente de l'autobus

On considère un arrêt de bus où un bus arrive 1 fois sur 3 au bout de 12 min, 1 fois sur 3 au bout de 8 min et 1 fois sur 3 au bout de 4 min comme indiqué sur le schéma.



1) Quelle est la durée moyenne séparant l'arrivée de deux bus ? (on dit souvent « quel est le temps moyen » séparant l'arrivée de deux bus ?)

2) Un usager arrive à un moment quelconque de la journée à cet arrêt et prend le premier bus qui arrive. Attendra-t-il en moyenne 8 minutes ? 6 minutes ? 4 minutes ? Moins de 4 minutes ? Plus de 4 minutes ?

<u>Solution</u>

1) La moyenne des durées entre deux bus est égale à $\frac{12+8+}{3} = 8$

2) Un usager qui arrive à l'arrêt d'autobus peut penser qu'il attendra en moyenne $\frac{8}{2} = 4$ min. (voir paragraphe III -2).

Le temps moyen d'attente pour une personne qui arrive au hasard pendant cette période de 24 minutes est égal à $6 \times \frac{12}{24} + 4 \times \frac{8}{24} + 2 \times \frac{4}{24} = \frac{112}{24} \approx 4{,}7$ qui est supérieur à 4 minutes.

Ex 15 - (d'après bac 2016)

Le temps d'attente au guichet de l'agence de location, exprimé en minutes, peut être modélisé par une variable aléatoire T qui suit la loi uniforme sur l'intervalle [1 ; 20].
1. Quelle est la probabilité d'attendre plus de douze minutes ?
2. Préciser le temps d'attente moyen.

Ex 16 - Attente à un feu tricolore
Un feu tricolore reste 55 secondes au vert, 5 secondes à l'orange et 60 secondes au rouge. Un piéton ne peut traverser que lorsque le feu est rouge. A 8h00 le feu passe au rouge.
T est la variable aléatoire qui donne, en secondes, le temps écoulé de 8h00 jusqu'à l'heure d'arrivée devant le feu d'un piéton désirant traverser.
On suppose que T suit la loi uniforme sur [0 ; 300].
1) Réaliser un schéma indiquant les couleurs du feu tricolore entre 0 et 300 secondes.
2) Calculer la probabilité qu'un piéton attende moins de 10 secondes.
3) Calculer la probabilité qu'un piéton attende plus de 20 secondes.

Ex 17 - **Temps d'attente à une caisse**
Dans un supermarché, le temps d'attente X à la caisse, exprimé en minutes, suit la loi uniforme sur l'intervalle [1;11[.
1) Déterminer la fonction de densité de probabilité f de la loi de X.
2) Quelle est la probabilité que le temps d'attente soit compris entre trois et cinq minutes ?
3) Quelle est la probabilité qu'un client attende plus de huit minutes à la caisse ?

Ex 18 - **Temps d'attente à une caisse (**Bac 2017)
Un supermarché dispose de plusieurs caisses. Un client qui se présente à une caisse doit attendre un certain temps $T1$ avant d'être pris en charge par le caissier. On considère que ce temps d'attente $T1$, exprimé en minute, est une variable aléatoire qui suit la loi uniforme sur l'intervalle [0 ; 12].
1) a. Quelle est la probabilité qu'un client attende au moins 5 minutes avant d'être pris en charge ?
b. Quel est le temps moyen d'attente à une caisse ?
2) Ces caisses automatiques tombent souvent en panne. On donne les informations suivantes. Le nombre de caisses automatiques est n = 10.
La probabilité qu'une caisse automatique tombe en panne pendant une journée donnée est p = 0,1. Une panne constatée sur une caisse automatique n'influence pas les autres caisses automatiques. Soit X la variable aléatoire correspondant au nombre de caisses automatiques qui tombent en panne pendant une journée donnée.
Calculer la probabilité pour qu'aucune caisse automatique ne tombe en panne pendant une journée donnée.

Ex 19 - QCM (une seule réponse est exacte)

A l'entrée d'un port, un feu destiné aux plaisanciers est rouge pendant 2,5 min puis vert pendant 1,5 min.
A midi le feu passe au rouge. Un bateau se présente entre 12h00 et 12h08. La variable aléatoire T qui donne le temps écoulé en minutes de midi à l'heure d'arrivée du bateau devant le feu suit la loi uniforme sur [0 ; 8]. La probabilité que le bateau

a) passe sans s'arrêter est 0,25 ;
b) attende plus de 45 s est 0,4375 ;
c) passe ou n'attende pas plus de 30 s est 0,75.

D – Deux problèmes

Problème 1 - Vie et mort d'un noyau radioactif [6]

I - Modèle discret

On considère que le temps est une variable discrète : il varie de un en un (par exemple d'année en année).

Un noyau radioactif «meurt sans vieillir ». Cela signifie que si le noyau n'est pas désintégré l'année *n*, la probabilité pour qu'il se désintègre l'année *n*+1 ne dépend pas de son « âge »
On note V_n l'événement « le noyau n'est pas encore désintégré l'année *n* » c'est à dire « vit encore l'année *n* »)
On note D_n l'événement « le noyau se désintègre l'année *n* »
On note *p* la probabilité pour que le noyau se désintègre la première année ($0<p<1$). Ainsi $p(D_1) = p$
On pose $q=1-p$ d'où $p(V_1)=q$
1° Faire un arbre de probabilité correspondant aux trois premières années.
2° L'année de désintégration d'un noyau radioactif est une variable aléatoire notée N. Quelle est la loi de probabilité suivie par N ?
3° On prend p=0,1. A l'aide de la calculatrice, déterminer la durée de vie médiane d'un noyau c'est-à-dire l'année à partir de laquelle la probabilité pour que le noyau « vive encore » est inférieure à 0,5.

II - Modèle continu

La désintégration d'un noyau radioactif est un phénomène aléatoire modélisé de la façon suivante : la durée de vie (exprimée en années) d'un noyau radioactif de carbone 14 est une variable aléatoire X qui suit une loi exponentielle de paramètre $\boldsymbol{\lambda}$ où $\boldsymbol{\lambda}$ est un nombre réel strictement positif.
1° Justifier pourquoi la probabilité pour que le noyau ne soit pas encore désintégré à l'instant *t* est égal à $\boldsymbol{e^{-\lambda t}}$

[6] Voir l'annexe 6 Désintégration radioactive

2° Sachant qu'un noyau n'est pas désintégré l'année t, quelle est la probabilité pour qu'il ne soit pas désintégré l'année suivante ? Montrer que la réponse ne dépend pas de t. Interpréter ce résultat.
3° Déterminer la durée de vie moyenne d'un noyau.

<u>Problème 2 - Montage en série ou en parallèle</u>
(inspiré d'un sujet de baccalauréat)

La durée de vie (en heures) d'une pièce déterminante pour la sécurité dans le fonctionnement d'un avion a été modélisée par une variable aléatoire X dont la loi de probabilité est la loi exponentielle de paramètre 0,0001.

1° Un contrôle a lieu toutes les 1000 heures de vol pour cet élément.

a) Calculer la probabilité pour qu'un élément, pris au hasard, ait une durée de vie supérieure à 1000 heures.
b) Pour plus de sécurité, on monte n éléments de ce type en parallèle.

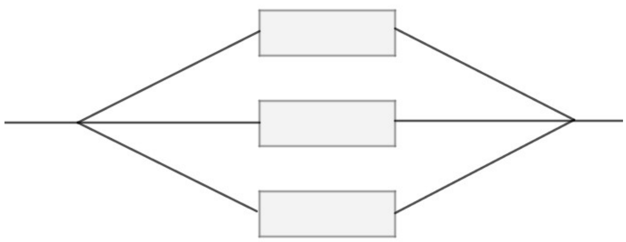


Schéma avec 3 éléments en parallèle (n=3)

Si les élèves n'ont pas étudié la fonction *ln* ils peuvent déterminer λ à l'aide de la calculatrice en tabulant $x \to e^x$

On suppose que les durées de vie des différents éléments sont indépendantes.

Il suffit alors que l'un des éléments soit encore en état de marche pour que l'ensemble fonctionne (on rappelle que lorsque deux événements A et B sont indépendants, $P(A \cap B) = P(A) \times P(B)$)

Pour n= 2 montrer que la probabilité que l'ensemble fonctionne au moins 1000 h est environ égales à 0,991

Calculer la probabilité pour n=3.

c) Calculer la valeur n à partir de laquelle cette probabilité est supérieure à 0,9999.

2° Dans un autre montage trois de ces éléments sont montés en série.

Il faut alors que les trois éléments soient encore en état de marche pour que l'ensemble fonctionne.
Quelle est la probabilité pour que l'ensemble fonctionne au moins 1000 heures ?

ANNEXES

Annexe 1

Descriptif du thème « Temps d'attente »

Certains phénomènes physiques (temps de désintégration d'un atome radioactif) ou biologiques (durée de vie de certains organismes) possèdent la propriété d'absence de mémoire. Leur modélisation mathématique repose sur l'utilisation des lois géométriques et exponentielles selon que le temps est considéré comme discret ou continu. La loi géométrique est vue soit comme la distribution du premier succès dans un schéma de Bernoulli, soit comme une loi discrète possédant la propriété d'absence de mémoire. La loi exponentielle peut être introduite à partir de la propriété d'absence de mémoire.

Problèmes possibles
– Durée de vie d'un atome radioactif. Discrétisation d'une variable aléatoire suivant une loi exponentielle.
– Exemples de modélisation par une variable aléatoire suivant une loi géométrique ou exponentielle : durée entre deux appels téléphoniques, durée de vie d'un composant électronique, période de retour de crue, etc.
– Utilisation de la loi uniforme. Temps d'attente à un arrêt de bus, paradoxe de l'inspection.

Contenus associés
– Lois à densité.
– Loi géométrique, loi exponentielle.
– Absence de mémoire, discrète ou continue.
Exemples d'algorithme
– Simulation d'une variable aléatoire de loi géométrique à partir du schéma de Bernoulli.
– Simulation d'une loi exponentielle à partir d'une loi uniforme.
– Demi-vie d'un échantillon de grande taille d'atomes radioactifs.

Descriptif des contenus de la partie du programme «Probabilités et statistiques »
Lois discrètes
Contenus :
– Loi uniforme sur $\{1,2, \dots , n\}$. Espérance.
– Épreuve de Bernoulli. Loi de Bernoulli : définition, espérance et écart type.
– Schéma de Bernoulli. Représentation par un arbre.

– Coefficients binomiaux[7] : définition (nombre de façons d'obtenir k succès dans un schéma de Bernoulli de taille n), triangle de Pascal, symétrie.
– Variable aléatoire suivant une loi binomiale $\mathcal{B}(n, p)$. Interprétation : nombre de succès dans le schéma de Bernoulli. Expression, espérance et écart type (admis). Représentation graphique.
– Loi géométrique : définition, expression, espérance (admise), représentation graphique et propriété caractéristique (loi sans mémoire)

Lois à densité
Contenus
– Notion de loi à densité à partir d'exemples. Représentation d'une probabilité comme une aire. Fonction de répartition $x \mapsto P(X \leq x)$.
– Espérance et variance d'une loi à densité, expressions sous forme d'intégrales.
– Loi uniforme sur [0,1] puis sur $[a, b]$. Fonction de densité, fonction de répartition. Espérance et variance.
– Loi exponentielle. Fonction densité, fonction de répartition. Espérance, propriété d'absence de mémoire.

Annexe 2

Correspondances entre variables aléatoires discrètes et continues et les lois de probabilités

X variable aléatoire	discrète	continue (ou à densité)
Univers	$\Omega = \{x_k / k \epsilon \{0; 1; 2; \dots n\}\}$ ou $\Omega = \{x_k / k \epsilon \mathbb{N}\}$	Ω est un intervalle I inclus dans $\mathbb{R}$ ou $\Omega = \mathbb{R}$
La loi de probabilité est	donnée, pour tout $k \epsilon \{0; 1; 2; \dots n\}$ ou pour tout $k \epsilon \mathbb{N}$, par $P(X = x_k)$ souvent notée p_k.	déterminée par la fonction densité. On introduit aussi la fonction de répartition $F(x) = P(X \leq x) = \int_a^x f(t)dt$ où f est la fonction de densité si $\Omega = [a; b]$ ou $F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt \ si \ \Omega = \mathbb{R}$
Espérance mathématique	$E(X) = \sum x_k \times p_k$	$E(X) = \int_{x \in \Omega} x f(x) dx$
Exemples	Loi binomiale, loi géométrique	Loi uniforme : fonction de densité constante Loi exponentielle (voir paragraphe IV)

[7] On trouvera une étude sur ce sujet sur le site de l'IREM de Paris – Groupe Manuels de lycée –dans la ressource pour les professeurs de Terminale **Combinatoire et dénombrement**

Annexe 3 - Paradoxe de l'inspection

Références

1) Introduction à la fatalité mathématique 23 mars 2015Posted by Jérôme Malot

2) Paradoxe de l'inspection – Introduction à la fatalité ...
www.blablasciences.com/?p=134

3) https://towardsdatascience.com/the-inspection-paradox-is-everywhere-2ef1c2e9d709

4) https://jakevdp.github.io/blog/2018/09/13/waiting-time-paradox/

5) https://allendowney.blogspot.com/2015/08/the-inspection-paradox-is-everywhere.html

6) http://accromath.uqam.ca/2020/02/tout-vient-a-point-me%cc%82me-sil-faut-attendre-plus-longtemps/

Annexe 4 - Temps d'attente de l'autobus – une généralisation du temps moyen d'attente

Cas où le bus passe 1 fois sur 2 au bout de a minutes et 1 fois sur 2 au bout de b minutes

On va généraliser la formule obtenue pour le temps d'attente en supposant que le bus passe aux instants 0, *a, a+b* .

Sur l'intervalle [0; *a*], la durée moyenne est $\frac{a}{2}$, Sur l'intervalle [*a, a+b*], la durée moyenne est $\frac{b}{2}$,

La moyenne des durées d'attente pondérée par l'amplitude des intervalles est égale à :

est $\frac{a}{2} \times \frac{a}{a+b} + \frac{b}{2} \times \frac{b}{a+b} = \frac{a^2+b^2}{2(a+b)}$

Avec les même notations que précédemment X suit une loi uniforme sur [0, a+b]

La fonction de répartition de Y

On appelle Y la durée de l'attente

On suppose b<a ; on note A l'événement $(0 < X \leq a)$ et B l'événement $(a \leq X \leq a + b)$

Soit t un réel

Si t≤0 alors $P(Y \leq t) = 0$

Si $0 < t < b$ alors $P(Y \leq t) = P(Y \leq 0) + P\big((0 < Y < t) \cap A\big) + P\big((0 < Y \leq t) \cap B\big) =$

$0 + P(0 < X < a - t) + P(a + b - t < X < a + b) = \frac{t}{a+b} + \frac{t}{a+b} = \frac{2t}{a+b}$

Si $b \leq t \leq a$ alors $P(Y > t) = P\big((Y > t) \cap A\big) + P\big((Y > t) \cap B\big) = P(0 < X < a - t) + 0 = \frac{a-t}{a+b}$

d'où P(Y≤t)= $1 - \frac{a-t}{a+b} = \frac{b+t}{a+b}$

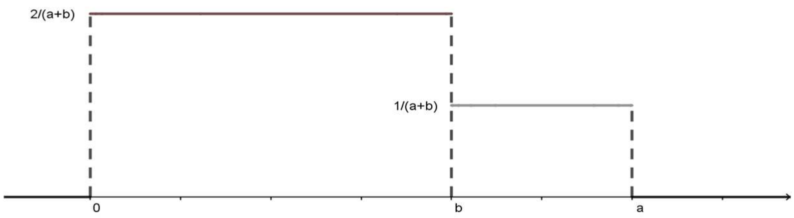
Si t>a alors P(Y≤t)=1

La Fonction de densité *f*

Si t≤0 alors f(t)=0 ;

Si 0<t<b alors f(t)=$\frac{2}{a+b}$;

Si b≤t≤a alors f(t)=$\frac{1}{a+b}$;
Si t>a on a f(t)=0
Sa représentation graphique est la suivante :



L'espérance de Y est donnée par :

E(Y)= $\int_0^b xf(x)dx + \int_b^a xf(x)dx = \int_0^b \frac{2}{a+b} xdx + \int_b^a \frac{1}{a+b} xdx = \frac{b^2}{a+b} + \frac{\frac{a^2}{2} - \frac{b^2}{2}}{a+b} = \frac{a^2+b^2}{2(a+b)}$

Remarque : le résultat final est symétrique en a et b : il reste donc valable si a<b.

Annexe - 5 - Temps d'attente de l'autobus – une deuxième généralisation

Cas où le bus passe n fois au bout de a minutes et p fois au bout de b minutes

Si on a n intervalles d'amplitude a minutes et p intervalles d'amplitude b minutes, donc en tout n+p intervalles pour une durée totale de na+pb minutes.

Le temps moyen entre deux bus est $\frac{na+p}{n+p}$ donc le temps estimé par le voyageur est la moitié $\frac{na+pb}{2(n+p)}$

La probabilité d'arriver dans un intervalle d'amplitude a est $\frac{na}{na+pb}$ et celle d'arriver dans un intervalle d'amplitude b est $\frac{pb}{na+pb}$

Donc la moyenne pondérée ou temps observé est $\frac{a}{2} \times \frac{na}{na+pb} + \frac{b}{2} \times \frac{pb}{na+} = \frac{na^2+pb^2}{2(na+\quad)}$

A-t-on toujours temps estimé par le voyageur plus petit que le temps observé ?

C'est-à-dire, a-t-on $\frac{na+pb}{2(n+p)} < \frac{na^2+pb^2}{2(na+pb)}$?

Les nombres sont tous positifs donc $\frac{na+p}{2(n+p)} < \frac{na^2+pb^2}{2(na+\quad)}$ équivaut à

$(na + pb)^2 < (n + p)(na^2 + pb^2)$ ou $2napb < npb^2 + npa^2$ ou encore $2ab < a^2 + b^2$ qui est toujours vrai donc le temps estimé par l'usager est toujours inférieur ou égal (égal si les bus sont réguliers, n=p et a=b) au temps observé. L'usager attendra en général plus longtemps que ce qu'il avait prévu.

Annexe 6 - La désintégration radioactive

https://www.cea.fr/comprendre/jeunes/Pages/s-informer-reviser/radioactivite/essentiel-sur-la-radioactivite.aspx
http://acces.ens-lyon.fr/acces/thematiques/limites/Temps/datation-isotopique/comprendre/la-decroissance-radioactive
http://res-nlp.univ-lemans.fr/NLP_C_M13_G03/co/grain3-1-4.html

Parmi les problèmes ou les exemples d'algorithmes possibles, le programme stipule

- Durée de vie d'un atome radioactif. Discrétisation d'une variable aléatoire suivant une loi exponentielle.
- Demi-vie d'un échantillon de grande taille d'atomes radioactifs.

Afin de mieux appréhender les énoncés des exercices classiques de mathématiques faisant intervenir la radioactivité, il nous paraît utile de préciser ces deux notions, durée de vie et demi-vie, ainsi que le lien entre radioactivité et datation avec le carbone 14.

Un atome est composé d'un noyau[8] dans lequel se trouvent des protons et des neutrons. Des électrons gravitent autour du noyau mais pour ce qui est de la durée de vie d'un atome radioactif ou de la demi-vie d'un échantillon de grande taille d'atomes radioactifs, les électrons n'interviennent pas.

Un élément du tableau périodique des éléments, également appelé tableau ou table de Mendeleïev, est caractérisé par son nombre de protons. Par exemple, le plomb (Pb) ou le carbone (C) ou encore l'uranium (U) ont chacun un nombre spécifique de protons. Un élément peut avoir un isotope, c'est par exemple le cas du carbone 12 (6 protons et 6 neutrons), le carbone 14 (6 protons et 8 neutrons) étant un isotope du carbone 12. Les isotopes se caractérisent par un même nombre de protons mais des nombres de neutrons différents.

La plupart des noyaux d'atomes sont stables, mais certains noyaux sont instables et se désintègrent spontanément, ils sont radioactifs. Un élément radioactif est dans un état métastable (pas tout à fait stable) et a tendance à se transformer en un élément plus stable, ce qui constitue un type de réaction nucléaire.

Les mathématiques[9] permettent de modéliser la radioactivité à l'échelle microscopique (un seul noyau instable) et à l'échelle macroscopique (une population de noyaux dont l'effectif est de l'ordre du nombre d'Avogadro, soit 6×10^{23}).

Il est impossible de prédire la date de désintégration d'un noyau instable : sa durée de vie est aléatoire et il n'est pas possible d'étudier expérimentalement le comportement d'un seul atome radioactif. On est amené à étudier celui d'un grand nombre de ces atomes (de l'ordre d'une mole) et on en déduit expérimentalement la notion de 1/2 vie d'un échantillon et en extrapolant la notion, celle d'un

8 Le terme « nucléaire » vient du mot « noyau »
9 On pourra aussi se référer au document « Les mathématiques de l'enseignement scientifique – La désintégration radioactive » du Ministère de l'Education nationale et de la jeunesse – juin 2019.
https://cache.media.eduscol.education.fr/file/les_mathematiques_de_l_ES/34/3/RA19_Lycee_G_1_MATH_Enseignement-scientifique_DesintegrationRadioactive_1149343.pdf

élément radioactif (C, U, Ce...)
A partir de ces résultats expérimentaux, ce sont les mathématiques qui interviennent. Les lois de la statistique vont permettre alors de définir la probabilité que possède un atome donné de ne pas se désintégrer avant un instant T, ce qui définit la durée de vie d'un atome radioactif.

On observe donc un grand nombre, voire un très grand nombre de particules, ce qui permet d'obtenir des résultats aussi bien sur le comportement d'un noyau radioactif que sur des échantillons de grande taille d'atomes radioactifs.

C'est à partir d'observations expérimentales, par exemple à l'aide de compteurs Geiger, où l'on mesure le nombre de réactions nucléaires par unité de temps, que l'on peut établir un modèle probabiliste en observant qu'un noyau instable ne vieillit pas (ou n'a pas de mémoire) : sa probabilité de désintégration par unité de temps reste constante au cours du temps. De plus, hormis le cas spécifique de réactions en chaine, la désintégration d'un noyau n'affecte pas celle des autres.

Ces deux conditions conduisent à un modèle probabiliste qui est la « loi exponentielle ».

En effet, on peut observer le nombre N de particules qui changent d'état pendant une courte durée ; au bout d'un temps t, on passe de N particules à N/2 particules, puis pendant la même durée on observe que la moitié de ces N/2 particules changent d'état. Il reste alors (N/2)/2 particules qui vont changer d'état pendant la même durée t, ce qui correspondant à une suite géométrique décroissante de premier terme N et de raison ½. En passant à des durée très petites, on obtient un modèle continu du nombre de noyaux présents à n'importe quelle date t. Dans ce modèle la fonction densité de probabilité est une exponentielle décroissante $t \to \lambda e^{-\lambda t}$ (λ>0) qui correspond à la loi exponentielle qui est une loi sans mémoire. Le coefficient λ dépend du type de particules radioactives, plus précisément un noyau radioactif peut être caractérisé par sa constante radioactive λ.

<u>Durée de vie d'un atome radioactif</u>

La radioactivité s'atténue progressivement au fil du temps avec la désintégration des atomes radioactifs : c'est la décroissance radioactive. Plus ou moins lent selon les noyaux, ce processus détermine la durée de vie des atomes radioactifs et leur niveau de radioactivité.

On note *X* la variable aléatoire égale au nombre de particules radioactives restant dans le noyau. Les particules sont caractérisées par la constante radioactive λ et la densité de probabilité est définie par $f(t) = \lambda e^{-\lambda t} si\ t \geq 0$. Cette fonction donne la probabilité de la durée de vie d'un atome radioactif.

L'espérance de vie moyenne est égale à $\frac{1}{\lambda}$

<u>La demi-vie d'un noyau radioactif</u>, ou période radioactive, est la durée nécessaire pour que la moitié des noyaux radioactifs initialement présents dans un échantillon macroscopique se soit désintégrée. Cela ne dépend pas de l'environnement (température, pression) mais c'est une propriété liée à l'élément radioactif. $T_{1/2}=\frac{ln}{\lambda}$

Datation par le carbone 14 (^{14}C)

Le carbone possède plusieurs formes – ou « isotopes » – parmi lesquelles le carbone 14, ou ^{14}C. Cet élément est radioactif, et sa radioactivité décroît au fil du temps à un rythme parfaitement régulier.

Tout être vivant comprend une certaine quantité de carbone. Ce carbone « vit » tant que l'être en question vit. Comme tout isotope du carbone, le carbone 14 se combine avec l'oxygène de notre atmosphère pour former alors du CO_2 (dioxyde de carbone). **Ce CO_2 est assimilé par les organismes vivants tout au long de leur vie : respiration, alimentation… En mourant, ils n'en assimilent plus**. La quantité de carbone 14 assimilé diminue alors au cours du temps de façon exponentielle tandis que celle de carbone 12 reste constante, puisque c'est un élément stable.

En observant à un instant *t* ce qu'il reste de C14 (décroissance exponentielle) dans un organisme, on peut dater depuis combien de temps il est mort.

TITRE :

Des ressources pour les enseignants mathématiques de terminale

AUTEURES:

Monique Chappet-Paries, Jacqueline Penninxck, Françoise Pilorge et Aline Robert

RESUMÉ :

Cette brochure, en deux fascicules, présente des ressources pour les mathématiques de terminale et, pour deux d'entre elles, plus larges. L'ambition est de servir à la préparation des séances. Inscrits dans les programmes (sauf les deux dernières), ces documents, réservés aux enseignants, sont longs, sans beaucoup d'exercices, et n'ont pas donné lieu à expérimentation sauf exception.

Fascicule 1 : pour la classe de terminale mathématiques complémentaires (à partir de thèmes)

- Corrélation et causalité (statistique à une variable, à deux variables)
- Inférences bayésiennes et Probabilités (conditionnelles, totales, formule de Bayes)
- Lois binomiale, géométrique, exponentielle, uniforme en liaison avec le thème « temps d'attente »

MOTS- CLÉS :

corrélation et causalité, inférence bayésienne, temps d'attente, combinatoire, Bienaymé-Tchebychev, vecteurs de l'espace, aires, nombre d'or et Fibonacci.

Éditeur: IREM de Paris
Responsable de la publication: C.Hache
IREM de Paris – Case 7018
Université de Paris
75205 Paris cedex 13
irem_de_paris@univ-paris-diderot.fr
https://irem.u-paris.fr/

Dépôt légal : 2023
ISBN : 978-2-86612-408-3