REPERES - IREM. N° 51 - avril 2003

DES LOIS DE PROBABILITE CONTINUES EN TERMINALE S, POURQUOI ET POUR QUOI FAIRE ?

Michel HENRY
Irem de Franche-Comté

I - Quelques questions posées par l'introduction de lois continues en TS

II – Expériences aléatoires et modèles probabilistes : du discret au continu

1 – Lancer d'un dé, loi des faces
2 – Paradoxe des 3 bancs : quel modèle ?
3 – Le jeu de Franc-Carreau : une loi continue « naturelle »
4 – L'aiguille de Buffon : un résultat surprenant dans un modèle ad hoc

III – Du discret au continu : attente d'un événement

1 – Attente d'un événement fortuit, hypothèses de travail
2 – Modélisation du temps d'attente (cas discret)
3 – Du discret au continu : de la loi géométrique à la loi exponentielle
4 – Hypothèses heuristiques pour un processus de Poisson : la loi exponentielle s'impose simplement
5 – De la loi binomiale à la loi de Poisson

IV – Une loi de base hors programme : la loi normale

1 – Phénomènes gaussiens, observations statistiques
2 – Contexte naturel de la loi normale (Laplace-Gauss)
3 – Le théorème limite central (TLC)
4 – Exemple historique de base : le théorème de Moivre - Laplace
5 – Intervalle de confiance pour une proportion (sondages)

V – Des lois continues, pour quoi faire en Terminale ?

1 – La notion de loi dans les classes de Première
2 – Quelques remarques sur les lois continues au programme de TS
3 – Notion de densité, prolongements des outils probabilistes pour des modèles performants

Cet article prolonge un exposé présenté à Luynes le 2 décembre 2001 dans le cadre d'un stage de statistique organisé par la régionale d'Aix-Marseille de l'APMEP. « Aix-Marseille Vert », le bulletin de la régionale, en a été publié un extrait dans son numéro 9 du 4ème trimestre 2002.

I - Quelques questions posées par l'introduction de lois continues en TS

Deux lois discrètes de base sont aux programmes des terminales S et ES : la loi de Bernoulli et la loi binomiale. Implicitement, la loi uniforme discrète (équiprobabilité sur un ensemble fini d'issues) est en place depuis la classe de première. Leurs applications à des exemples concrets permettent de mettre en évidence leur statut de modèles standards, applicables à de nombreuses situations. Elles permettent aussi d'expliciter les paramètres qui les déterminent entièrement, respectivement la probabilité p du « succès » et le couple (n, p). Cette référence à des modèles de base devrait favoriser la compréhension en termes de modèles des deux lois continues introduites ensuite en terminale S : la loi uniforme sur [0, 1] et la loi exponentielle.

Il y a là, cependant, un saut conceptuel majeur. Si une loi de probabilités a les propriétés d'une distribution de fréquences relatives à un ensemble fini d'issues, comment ce concept peut-il être étendu à un ensemble continu de valeurs possibles ?

Une variable aléatoire, au sens de la classe de première, est une application définie sur un ensemble fini représentatif des issues. Elle ne peut donc avoir comme ensemble image qu'un ensemble fini de valeurs numériques. Comment expliquer son plongement dans un ensemble numérique continu ?

La probabilité d'un événement a été définie comme la somme des probabilités des événements élémentaires le constituant. Transposées au cas continu, ces probabilités sont nécessairement toutes nulles, comme nous l'avons vu. Comment donner un sens probabiliste à la notion de densité ?

On voit que le pari d'introduire des lois continues en terminale est pour le moins audacieux. Le point de vue de la modélisation est ici incontournable. Les lois continues sont alors des outils mathématiques permettant, dans les modèles considérés, de calculer les probabilités de certains types d'événements. Ces modèles continus sont choisis et leurs paramètres sont déterminés (hypothèses de modèle) pour coller assez bien à une réalité discrète décrite (hypothèses de travail), cela à partir de considérations heuristiques et d'estimations statistiques. Les valeurs des probabilités ainsi calculées se révèlent alors assez proches des fréquences stabilisées des événements associés, obtenues par l'observation statistique.

II – Expériences aléatoires et modèles probabilistes : du discret au continu

1 – Lancer d'un dé, loi des faces

Dans l'esprit des programmes des années 2000, séparons l'expérience concrète (ou sa simulation, éventuellement répétée de nombreuses fois pour permettre d'observer les effets du hasard sur les fluctuations des fréquences d'apparition des différentes faces), du modèle probabiliste très simple, introduit en classe de première en termes de « loi de probabilité », censé la décrire théoriquement.

Nous sommes alors amenés à distinguer des « *hypothèses de travail* » et des « *hypothèses de modèle* » qui pourront en découler :

Hypothèses de travail : le dé est correctement construit dans une matière suffisamment homogène pour bien accepter les symétries en jeu, plaçant les six faces dans des conditions équivalentes, il est lancé dans des conditions reproductibles assez ouvertes pour qu'aucune prévision mécanique ne puisse être avancée pour connaître par avance la face qui se présentera, etc.

Hypothèse de modèle : Il y a six issues possibles, toutes équiprobables de probabilité 1/6. Si on désigne ces issues par les points portés par les faces du dé, la probabilité est distribuée sur cet ensemble noté Ω suivant une loi que l'on peut représenter par le tableau :

Ω = {•, ••, •••, ••••, •••••, ••••••} ,

loi sur Ω :

ω_i	•	••	•••	••••	•••••	••••••
p_i	1/6	1/6	1/6	1/6	1/6	1/6

Dans le modèle probabiliste, on peut numériser ces issues en introduisant une variable aléatoire X, définie sur Ω, à valeurs dans{1, 2, 3, 4, 5, 6} et dont la loi est donnée par les probabilités élémentaires p_i.

Cette loi est notée $\mathbb{P}_X$ et les valeurs p_i des probabilités élémentaires associées aux événements « $X = x_i$ » sont notées indifféremment :

$$p_i = \mathbb{P}(X = x_i) = \mathbb{P}_X(\{x_i\})$$

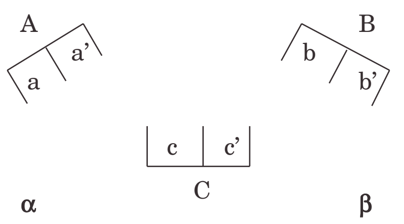
La loi de X est appelée loi uniforme discrète sur {1, 2, 3, 4, 5, 6}.

Cette distinction entre hypothèses de travail et hypothèses de modèle sera omniprésente chaque fois qu'il s'agira de comprendre le processus de modélisation qui permettra d'analyser les situations aléatoires de la réalité à partir de connaissances probabilistes théoriques. Cela devient incontournable pour comprendre le passage d'une situation aléatoire concrète, où l'ensemble des issues possibles est naturellement fini, à un modèle probabiliste continu. Elle s'avère aussi très utile pour lever certains paradoxes qui peuvent se présenter quand les situations réelles ne sont pas décrites avec suffisamment de précision, comme le montre le petit exemple bien connu qui suit.

2 – Paradoxe des 3 bancs : quel modèle ?

Voici ce problème des trois bancs :

Dans un square il y a trois bancs à deux places. α et β vont s'asseoir « au hasard ». Quelles chances ont-ils de se retrouver sur le même banc ?



La manière dont le hasard intervient dans cette situation (d'école) n'est pas précisée. La locution « au hasard », nous dit-on, suppose implicitement l'équiprobabilité des issues. Mais quel est l'ensemble des issues considéré : les bancs ou les places ? Cela n'est pas non plus précisé. Cette imprécision est d'ailleurs à la source de nombreux paradoxes du calcul des probabilités, quand plusieurs modèles peuvent être invoqués pour décrire une même situation [1]. Analysons ce problème en précisant nos hypothèses de travail et de modèle.

Hypothèses de travail 1 :

α va s'asseoir en a. Peu importe sa manière de choisir sa place. β tire un banc « au hasard » parmi A, B et C, et va s'asseoir. L'événement E est réalisé par le choix du banc A.

Hypothèses de modèle 1 :

On prend un modèle à trois éléments, $\Omega = \{A, B, C\}$ pour représenter le choix de β parmi les 3 bancs. L'hypothèse de travail suggère de répartir la probabilité de manière uniforme sur Ω. On a donc la loi :

A	B	C
1/3	1/3	1/3

L'événement E est réalisé par le choix de A : $\mathbb{P}(E) = 1/3$.

Hypothèses de travail 2 :

α va s'asseoir en a. β tire une place « au hasard » parmi les cinq restantes a', b, b', c et c', et va s'asseoir. L'événement E est réalisé par le choix de la place a'.

Hypothèses de modèle 2 :

On prend un modèle à cinq éléments, $\Omega = \{a', b, b', c, c'\}$ pour représenter le choix de β parmi les 5 places. L'hypothèse de travail suggère de répartir la probabilité de manière uniforme sur Ω. On a donc la loi :

a'	b	b'	c	c'
1/5	1/5	1/5	1/5	1/5

E est réalisé par le choix de a' : $\mathbb{P}(E) = 1/5$.

Quelle conclusion tirer ? Que l'énoncé ne décrit pas suffisamment précisément une véritable expérience aléatoire : quel choix aléatoire pour β ? Dans ces conditions deux modèles aussi pertinents l'un que l'autre peuvent être proposés. Il n'est pas surprenant qu'ils conduisent à deux résultats différents.

3 – Le jeu de Franc-Carreau : une loi continue « naturelle »

Ce jeu, répandu à la cour, a été étudié par Georges Louis Leclerc, comte de Buffon en 1733 [2]. Pour la première fois, un point de vue probabiliste s'appliquait à une situation où les issues théoriques considérées forment un ensemble continu, géométrique en l'occurrence, d'où le nom de « probabilités géométriques » que l'on donne parfois à ce type de problèmes. Nous allons voir que le point de vue de Buffon, qui semble introduire une nouvelle notion de probabilité, peut se ramener au concept élémentaire de probabilité tel qu'il se présentait à l'époque, défini alors par la formule que Laplace érigera en 1812 en premier principe :

$$\frac{\textit{nombre de cas favorables}}{\textit{nombre de cas possibles}}.$$

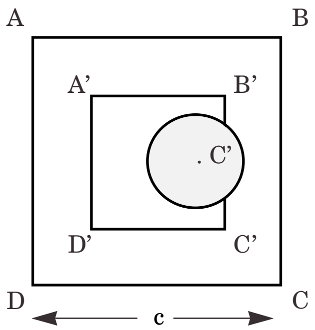
Le jeu consiste à jeter un écu (pièce de monnaie en forme de disque) sur un carrelage

1 Comme exemple historique de tels paradoxes, le paradoxe de Bertrand est omniprésent dans les manuels. Il est présenté dans le n° 13 de Repères-IREM (octobre 1993) : Paradoxes et lois de probabilités, p. 11 à 14.

2 Pour une présentation de l'étude historique de Buffon du jeu de Franc-Carreau et du fameux problème de l'aiguille, on pourra lire l'article de Frédéric Métin : « Buffon et le problème de l'aiguille » dans Si le nombre m'était conté, Comm.inter-Irem Histoire et Épistémologie, ellipses 2000, pp. 175-198

régulier. Les joueurs parient sur la position finale de l'écu : à cheval sur un ou plusieurs joints entre les carreaux ou entièrement dans un seul carreau (position de « Franc-Carreau »). Comment déterminer la probabilité de faire « Franc-Carreau » ?

Si on désigne par ABCD le carreau de côté c dans lequel le centre O de l'écu est tombé, la condition « Franc-Carreau » est géométriquement élémentaire : elle est réalisée si et seulement si le centre du disque tombe à l'intérieur du carré A'B'C'D' de côté $c - 2r$, homothétique du carré ABCD par rapport à son centre.



Buffon considère comme évident que la probabilité de « Franc-Carreau » est égale au rapport des aires des carrés A'B'C'D' et ABCD. Cela semble en effet le plus raisonnable. Cependant les élèves peuvent s'interroger sur la légitimité de ce choix et la pratique montre que certains pensent spontanément à d'autres possibilités incongrues, comme par exemple le rapport de l'aire du disque à celle du carré ABCD.

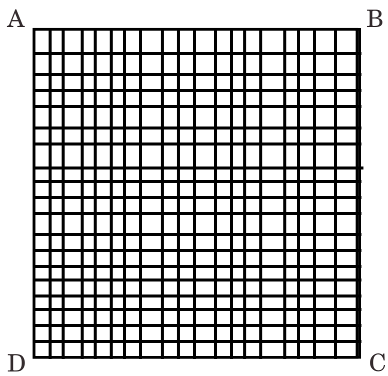
On peut clarifier cela en posant ce problème en termes de modélisation.

Hypothèse de travail : l'écu est lancé « au hasard » et tous les points à l'intérieur du carré ABCD ont « la même chance » d'être atteints par le centre O.

Hypothèse de modèle continu : L'ensemble des issues est l'ensemble Ω des points intérieurs au carré ABCD. Mais quelle loi proposer pour refléter l'hypothèse de travail ?

Remarquons que si l'on cherchait à affecter une même probabilité à chacun de ces points, celle-ci ne pourrait être strictement positive en vertu de l'axiome d'additivité, vu l'infinitude de ces points. Elle ne peut être nulle, car la probabilité de tomber dans le carré ABCD vaut 1. Cette situation probabiliste, apparemment contradictoire, ne peut être abordée élémentairement, elle suppose l'introduction de nouveaux outils dans la théorie.

On va ramener ce problème à un problème élémentaire par le biais d'un modèle approché. Pour cela, on « discrétise » le carré ABCD : On tapisse ABCD par un quadrillage de petits carrés unité u_i de côté c/1000 par exemple.



Pour fixer les idées, prenons c = 10 cm, r = 1 cm, et u_i de côté 0,01 cm. L'aire de ABCD est 100 cm^2 et celle de A'B'C'D' vaut 64 cm^2.

Hypothèse de modèle discret : on prend alors comme ensemble Ω' des issues possibles l'ensemble des petits carrés u_i. Ils sont au nombre de 10^6. On traduit l'hypothèse de travail en posant qu'ils sont équiprobables :

$$\forall\, i,\ \mathbb{P}(u_i) = 1/10^6.$$

L'événement « Franc-Carreau » (FC) est réalisé si O tombe dans l'un des u_i qui tapissent A'B'C'D' (exactement avec nos hypothèses numériques). Par définition de la probabilité d'un événement,

$$\mathbb{P}(FC) = \sum_{\{i : u_i \subset A'B'C'D'\}} \mathbb{P}(u_i) =$$

$$= [(c - 2r)\times 100]^2 \times \mathbb{P}(u_i) = 0{,}64.$$

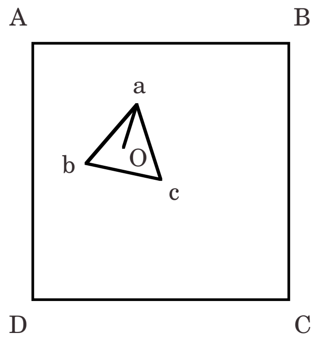
Par définition élémentaire de la mesure de l'aire d'une figure comme nombre de carrés unités qui peuvent la recouvrir, on a obtenu : $\mathbb{P}(FC) = \dfrac{\text{aire de } A'B'C'D'}{\text{aire de } ABCD}$.

Cette remarque élémentaire nous conduit à compléter et généraliser notre *hypothèse de modèle continu* :

> pour tout domaine A , inclus dans ABCD (d'aire mesurable), si E est l'événement : « le centre O est tombé dans A », on pose que la probabilité de E est proportionnelle à l'aire de A : $\mathbb{P}(E) = \dfrac{\text{aire de } A}{\text{aire de ABCD}}$.

La propriété de sous-additivité de l'aire (l'aire de la réunion de deux parties disjointes est la somme des aires de ces parties) donne à $\mathbb{P}$ les propriétés d'une loi de probabilité, c'est la **loi uniforme continue** sur le carré ABCD. On retrouve le fait nécessaire que dans ce modèle, la probabilité que O tombe sur une figure réduite à un point ou à un ensemble dénombrable de points, voire sur une ligne continue (comme par exemple les côtés du carré A'B'C'D' pour lesquels la pièce « touche » tangentiellement le bord du carreau), est nulle.

Ce point de vue des « probabilités géométriques » semble satisfaisant dans le cas très simple traité ici où les événements considérés peuvent s'énoncer en termes d'aires. Ce n'est pas toujours le cas. Prenons celui d'un jeu de franc-carreau où l'on a remplacé l'écu par une plaque en forme de triangle équilatéral.

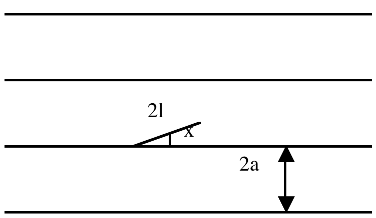


On peut prendre pour paramètres aléatoires les coordonnées x et y du centre O du triangle et l'angle polaire α du vecteur $\overrightarrow{Oa}$. Il n'y a pas d'interprétation géométrique simple de l'événement « franc-carreau ». On peut représenter cet événement dans un repère en trois dimensions comme une partie du pavé $[0, c]\times[0, c]\times[0, 2\pi]$ muni d'une loi uniforme continue. C'est ce que nous ferons pour résoudre le fameux problème de l'aiguille de Buffon.

En classe, ce problème du triangle peut être simulé sur ordinateur et la probabilité de « Franc-Carreau » peut être évaluée par une approche fréquentiste, mettant en œuvre une forme naïve (vulgarisée !) de la loi des grands nombres. Mais quelles lois introduire dans cette simulation pour les paramètres x, y et α ? En gros, quel modèle implanter dans la machine ?

4 – L'aiguille de Buffon : un résultat surprenant dans un modèle ad hoc

Sur un parquet formé de planches de largeur 2a, séparées par des rainures droites, parallèles et équidistantes, on jette une aiguille de longueur 2l, avec l < a.

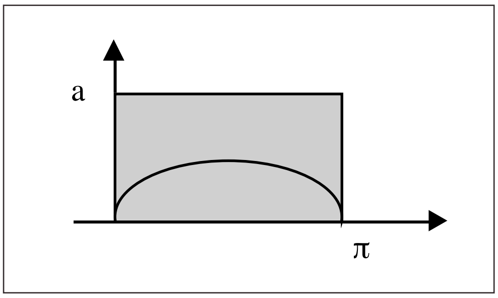


Quelle est la probabilité que l'aiguille coupe l'une des rainures (événement *A*) ?

Hypothèses de travail :

Soit X la distance du milieu de l'aiguille à la rainure la plus proche. X prend une valeur aléatoire quelconque dans [0, a]. Soit θ l'angle des droites formées par cette rainure et l'aiguille. θ prend une valeur aléatoire quelconque dans $[0, \pi]$.

Dans une représentation cartésienne du pavé $\Omega = [0, a]\times[0, \pi]$ (rectangle hachuré), l'événement *A* est représenté par la partie grisée délimitée par la courbe d'équation $x = l \sin\theta$. Dans cette représentation, on prend pour *hypothèse de modèle* probabiliste la loi uniforme sur le rectangle Ω.



On a donc :

$$\mathbb{P}(A) = \frac{\text{aire grisée}}{\text{aire hachurée}} = \frac{1}{\pi a}\int_0^{\pi} l \sin\theta \, d\theta = \frac{2l}{\pi a}.$$

Ainsi, le nombre π intervient dans l'expression d'une probabilité, éventuellement non rationnelle, ce qui peut être surprenant si l'on s'en tient à la définition laplacienne.

Dans ce problème, le modèle uniforme choisi ne s'applique pas à la situation géométrique d'origine, mais à sa représentation cartésienne.

Aurions-nous le même résultat avec une représentation en coordonnées polaires ? Quelle légitimité le modèle uniforme sur le rectangle cartésien a-t-il en l'occurrence ? Outre les lois uniformes pour X et θ, il fait intervenir une hypothèse d'indépendance entre X et θ. Nous laisserons ici cette question en l'état.

L'expérimentation numérique par simulation informatique ne permettra pas d'y répondre, car ayant implanté ces hypothèses dans le programme, il ne sera pas surprenant d'observer que la fréquence stabilisée de l'événement *A* tend à confirmer le résultat théorique.

III – Du discret au continu : attente d'un événement

1 – Attente d'un événement fortuit, hypothèses de travail

Un événement A peut survenir inopinément et se répéter fortuitement : panne d'un appareillage complexe, arrivée d'un client dans une file d'attente, accident de voiture dans la région, rupture du fil dans un métier à tisser, rayon cosmique, radioactivité etc.

On s'intéresse au temps d'attente aléatoire avant la prochaine manifestation de A. Cette situation se caractérise par un paramètre qui peut être évalué à partir d'une statistique : on peut observer que, dans des conditions analogues, A se produit en moyenne c fois dans un intervalle de temps unité (par exemple, si c = 7,2 fois par heure, c est la cadence horaire du phénomène).

Quelle loi attribuer à cette attente ? Sous quelles hypothèses de modèle ?

On ne peut pas douter de l'intérêt de ce problème. Les informations de nature probabiliste qu'un tel modèle peut apporter seront utiles pour « gérer » ces situations : mobilisation de personnels, budgétisation, contrôle de qualité …

Dans la foulée de ce qui précède, décomposons le processus de modélisation en hypothèses de travail et hypothèses de modèle.

Hypothèses de travail :

— Il n'y a pas de moments où A apparaît plus souvent : le phénomène est homogène dans le temps.

— Les « chances » de voir A se produire entre deux instants donnés t et t + Δt ne dépendent pas de ce qui s'est passé auparavant : le phénomène est sans mémoire.

— Plus Δt est petit, moins il y a de chance de voir A entre t et t + Δt. De plus A ne se produit pas deux fois presque en même temps : A est un événement « rare »[3].

Nous allons transformer ces hypothèses de nature heuristique en termes mathématiques adaptés aux outils probabilistes. Habituellement, on représente une durée par un intervalle de $\mathbf{R}^+$. Nous ne nous engageons pas ici dans une discussion sur la nature du temps et la légitimité de cette représentation. Ce continu nous servira donc de modèle pour représenter le temps d'attente de l'événement A.

Hypothèses de modèle

On considère comme ensemble des issues possibles l'ensemble continu de tous les instants où A peut se produire : $\Omega =]0, +\infty[$. Posons comme hypothèses :

1) La probabilité que A se produise au moins une fois dans un intervalle de temps $]t, t + \Delta t]$ ne dépend que de Δt (phénomène homogène et sans mémoire).

2) Soit $P(\Delta t)$ cette probabilité. On suppose que $P(\Delta t) \sim \lambda.\Delta t$ quand $\Delta t \to 0$, où $\lambda > 0$ est une constante : A est « rare ».

3) Si t_1, t_2, t_3, t_4 sont quatre instants successifs, les événements « A se produit entre t_1 et t_2 » et « A se produit entre t_3 et t_4 » sont indépendants (A est sans mémoire).

3 Même si A est relativement fréquent, ce terme est traditionnellement utilisé par les probabilistes dans ce type de situation.

Soit Z le temps aléatoire qu'il faut attendre pour observer le premier *A* à partir d'un instant initial quelconque t_0. Quelle est la loi de Z ? Nous verrons plus loin que les hypothèses proposées suffisent pour résoudre directement ce problème, modulo pour simplifier une hypothèse technique de régularité de la densité de cette loi, comme le suggère le document d'accompagnement des programmes de Terminales.

Mais, hors programme vu les calculs qui suivent, il est plus intéressant et plus porteur de sens de considérer la variable continue Z comme limite de variables discrètes, comme on l'a fait pour le jeu de franc-carreau.

2 – Modélisation du temps d'attente (cas discret)

On fait une observation par minute pour voir si *A* s'est produit. On note les instants d'observation : 1, 2, 3, … , k-1, k, … séparés par l'intervalle $\Delta t = 1$ mn. Soit [0, T] l'intervalle d'étude du phénomène pendant lequel on a observé un nombre (aléatoire) N de manifestations de *A* .

On note aussi $t_1, \ldots, t_N, \ldots$ les instants où *A* se produit, et $Z_k = t_k - t_{k-1}$ les variables d'attentes successives. On peut représenter les notations utilisées par le schéma ci-dessous que nous appellerons « schéma de Poisson ».

Soit e_k l'événement « *A* s'est produit pendant la $k^{\text{ème}}$ minute » (i.e. entre les instants k-1 et k) et e_k l'événement contraire. Les e_k et e_i ($k \neq i$) sont des évènements indépendants en vertu de la troisième hypothèse de modèle.

Pour tout k, la probabilité que *A* se produise pendant la $k^{\text{ème}}$ minute ne dépend pas de k (première hypothèse). Posons $\mathbb{P}(e_k) = P(\Delta t) = p$. On a donc $\mathbb{P}(e_k) = 1 - p$.

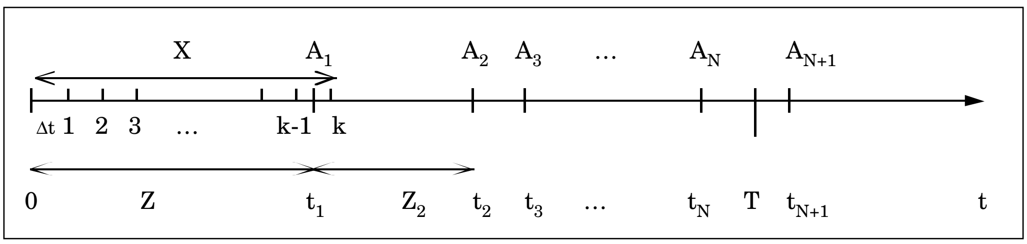
Soit X le temps d'attente discrétisé de *A* (X est un nombre aléatoire de minutes). X est à valeurs dans $\mathbf{N}^*$. Soit E_k l'événement : « A apparaît pour la première fois à la $k^{\text{ème}}$ minute ». Soit p_k la probabilité de E_k, encore notée $p_k = \mathbb{P}(E_k) = \mathbb{P}(X = k) = \mathbb{P}_X(k)$.

La loi $\mathbb{P}_X$ de X est entièrement déterminée par les probabilités élémentaires p_k . Calculons les p_k : E_k est la conjonction de e_k et des e_i pour $i < k$. Ces événements sont indépendants, d'où $p_k = \mathbb{P}(e_k).\prod \mathbb{P}(e_i) = p.(1-p)^{k-1}$.

La loi de X est donc donnée par :

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

C'est la loi géométrique *G*(p) de paramètre p.



On a bien :

$$\sum_{k=1}^{\infty} p_k = p\sum_{k=1}^{\infty}(1-p)^{k-1} = \frac{p}{1-(1-p)} = 1 .$$

La probabilité p est déterminée par la cadence c.

En effet, on peut voir que $E(X) = \frac{1}{p}$.

[Pour cela, par définition de l'espérance mathématique, il faut calculer la somme de la série $E(X) = \sum_{k=1}^{\infty} k\, p_k$.
Celle-ci est obtenue par un artifice. On écrit :

$$\sum_{k=1}^{\infty} k\, p_k = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\,\Big(\sum_{k=1}^{\infty} -(1-p)^k\Big)'$$

en considérant p comme une variable réelle, avec les précautions d'usage : la série entière est uniformément convergente puisque qu'avec p < 1 on est à l'intérieur de son intervalle de convergence.

La série géométrique obtenue livre son secret, remplaçant p dans le $\sum$ par la variable réelle x :

$$E(X) = \sum_{k=1}^{\infty} k\, p_k = p\left(-\frac{1-x}{1-(1-x)}\right)' = p\,\frac{1}{x^2} = \frac{1}{p} \quad]$$

E(X) représente le nombre moyen de minutes qu'il faut attendre pour avoir *A*… Il y a en moyenne c événements *A* par heure. Il faut donc attendre en moyenne 1/c heure ou 60/c minutes pour observer le premier *A*. On a donc avec ces unités, p = c/60, si dans notre modélisation, c est assez petit (< 60) pour justifier le fait que *A* est un événement rare.

Exemple numérique : si c = 7,2 / heure, la probabilité de ne pas avoir de A dans le premier quart d'heure est :

$$(1-p)^{15} = \left(1-\frac{7{,}2}{60}\right)^{15} = 0{,}147.$$

3 – Du discret au continu : de la loi géométrique à la loi exponentielle

On rapproche les temps d'observations, effectuées à chaque instant multiple de 1/n minute (par exemple toutes les secondes). La probabilité que *A* se produise dans un intervalle donné de durée $\Delta t = \frac{1}{n}$ est $P(\frac{1}{n}) \sim \frac{\lambda}{n}$ (car *A* est homogène et rare et par hypothèse de modèle : $P(\Delta t) \sim \lambda\Delta t$).

Soit Z la variable continue du temps d'attente de *A*, et F_Z sa fonction de répartition : $F_Z(t) = \mathbb{P}(Z \leq t)$. On a, d'après le théorème des accroissements finis, si F_Z de classe C^1 :

$$\mathbb{P}(t < Z \leq t+\varepsilon) = F_Z(t+\varepsilon) - F_Z(t) = \varepsilon\, F'_Z(\tau),$$
$$\text{avec } t < \tau < t+\varepsilon.$$

Soit Y_n le nombre de fractions de minutes à attendre avant que *A* se produise (on a $\varepsilon = \Delta t = \frac{1}{n}$).

D'après le paragraphe précédent, la loi de Y_n est une loi géométrique de paramètre $p = P(\frac{1}{n})$. On a :

$$\mathbb{P}(Y_n = k_n) = P(\tfrac{1}{n})(1-P(\tfrac{1}{n}))^{k_n-1}.$$

Supposons que *A* se produise à l'instant t (compté en minutes), compris entre $(k_n-1)\Delta t$ et $k_n\Delta t$, c'est-à-dire entre $\frac{k_n-1}{n}$ et $\frac{k_n}{n}$. Avec ces notations, on a $\frac{k_n}{n} \to t$ quand $n \to \infty$.

Si $Y_n = k_n$, on a $\frac{k_n-1}{n} < Z < \frac{k_n}{n}$ et

$\mathbb{P}\left(\frac{k_n - 1}{n} < Z \leq \frac{k_n}{n}\right) = \mathbb{P}(Y_n = k_n) = \frac{1}{n} F'_Z(\tau_n)$,

où $\frac{k_n - 1}{n} < \tau_n \leq \frac{k_n}{n}$.

On obtient donc :

$$F'_Z(\tau_n) = n.P(\frac{1}{n})(1 - P(\frac{1}{n}))^{kn-1} .$$

Quand $n \to \infty$, $\tau_n \to t$ avec $\frac{k_n}{n}$,

d'où $F'_Z(\tau_n) \to F'_Z(t)$, et dans le second membre,

$n\, P(\frac{1}{n}) \to \lambda$, et :

$$(1 - P(\frac{1}{n}))^{k_n - 1} = (1 - P(\frac{1}{n}))^{n\frac{k_n - 1}{n}} \to e^{-\lambda t},$$

car $n.\ln(1-P(\frac{1}{n})) \sim n(-P(\frac{1}{n})) \to -\lambda$.

D'où la *densité de la loi de* Z :

$$F'_Z(t) = f_Z(t) = \lambda\, e^{-\lambda t}, \text{ pour } t > 0.$$

C'est la densité de la loi exponentielle de paramètre λ.

La fonction de répartition de cette loi est

$$F_Z(t) = \mathbb{P}(Z \leq t) = \int_0^t \lambda e^{-\lambda\tau} d\tau = 1 - e^{-\lambda t},$$

car $F_Z(0) = 0$. On peut calculer son espérance mathématique :

$$E(Z) = \int_0^\infty \tau\, \lambda\, e^{-\lambda\tau} d\tau = \frac{1}{\lambda}.$$

De même que dans le cas discret, E(Z) est le temps moyen d'attente de *A*, $\frac{1}{\lambda} = \frac{60}{c}$, si Z en mesuré en minutes et si c est la cadence horaire du phénomène.

Exemple de la désintégration radioactive

Une matière fissile contient N atomes radioactifs. La désintégration de l'un ou l'autre de ces atomes (événement *A*) vérifie (en gros) les hypothèses heuristiques précédentes.

Soit $P(\Delta t)$ la probabilité d'observer une désintégration pendant un petit intervalle de temps de durée Δt, et soit λ la limite de $P(\Delta t)/\Delta t$ quand Δt tend vers 0.

Le paramètre λ caractérise cette radioactivité, c'est la constante de désintégration.

La période T de l'élément radioactif considéré est la durée pendant laquelle la moitié de la masse fissile s'est désintégrée. La fréquence des atomes non encore désintégrés dans la masse N est alors 1/2.

Un raisonnement fréquentiste permet de trouver simplement la relation entre T et λ.

Dans le cas de la radioactivité naturelle, on peut considérer que les désintégrations des atomes sont indépendantes. De plus, pour un intervalle de temps donné, chaque atome a la même probabilité d'être désintégré. Soit p cette probabilité de désintégration entre 0 et T. Un atome étant considéré, on lui associe l'expérience de Bernoulli qui consiste à voir s'il est désintégré au bout du temps T. Cet événement est donc de probabilité p. On recommence cette expérience avec les N atomes de la masse radioactive. Le théorème de Bernoulli, forme élémentaire de la loi des grands nombres, indique que la fréquence des atomes désintégrés à l'instant T tend (en probabilité) vers p quand le nombre d'épreuves tend vers l'infini. N étant très grand, on peut conclure qu'il y a une probabilité infime que p soit notablement différente de cette fréquence 1/2.

T est donc la durée au bout de laquelle un atome donné a la probabilité 1/2 d'être (ou ne pas être) désintégré.

Dans cette situation, on retrouve le passage d'un modèle discret à un modèle continu.

Soit un intervalle de temps Δt très petit. Soit k (très grand) tel que $T = k\,\Delta t$. Soit X le temps d'attente (mesuré discrètement en nombre de Δt) de la désintégration d'un atome donné. Si au bout du temps T cet atome n'est pas désintégré, on a $X > k$, et cet événement est de probabilité 1/2.

Mais $\mathbb{P}(X > k) \sim (1-\lambda\Delta t)^k$ puisque $P(\Delta t) \sim \lambda\Delta t$, et

$$(1-\lambda\Delta t)^k = \left(1-\frac{\lambda T}{k}\right)^k \xrightarrow[\Delta t \to 0]{} e^{-\lambda T}.$$

On a donc $e^{-\lambda t} = 1/2$ et on obtient $\lambda = (\ln 2)/T$.

La loi exponentielle de Z, temps d'attente de la désintégration de l'atome (modèle continu), fournit directement ce résultat :

$\mathbb{P}(Z \leq T) = 1/2 = 1 - e^{-\lambda T}$, d'où $T = (\ln 2)/\lambda$.

Par exemple T = 1580 ans pour le radium, T/ln2 = 2280 est la cadence annuelle de désintégration des atomes de radium.

4 – Hypothèses heuristiques pour un processus de Poisson : la loi exponentielle s'impose simplement

On reprend la variable Z du temps continu d'attente de l'événement *A* à partir de l'instant $t_0 = 0$ ($Z > 0$), et sa fonction de répartition $F_Z(t) = \mathbb{P}(Z \leq t)$, que l'on suppose assez régulière (de classe C^1).

Soit $g(t) = 1 - F_Z(t) = \mathbb{P}(Z > t)$. g est décroissante (car F_Z est croissante) et $g(0) = 1$.

« $Z > t$ » est l'événement « *A* ne se produit pas avant t ». $g(t + s)$ est donc la probabilité que *A* ne se produise pas avant t + s, c'est-à-dire ni avant t, ni entre t et t+s. Ces deux événements sont indépendants, en vertu de la troisième hypothèse heuristique, puisque $[0, t]$ et $]t, t+s]$ sont disjoints : *A* est sans mémoire.

Le premier événement est de probabilité g(t). Le deuxième a pour probabilité :

$$1 - \mathbb{P}(t < Z \leq t+s) = 1 - \mathbb{P}(Z \leq s) = g(s)$$

puisque cette probabilité ne dépend que de la durée s considérée (phénomène homogène).

On obtient donc pour tout t et s positifs :

$$g(t + s) = g(t).g(s),$$

relation fonctionnelle qui permet de déterminer la fonction g.

On peut le voir simplement en utilisant l'hypothèse que g est de classe C^1 (ce qui n'est pas nécessaire, la monotonie de g suffirait), régularité qui sera vérifiée a posteriori.

En effet, en dérivant cette égalité par rapport à s (à t constant), on a pour tout $t > 0$ et tout $s > 0$: $g'(t + s) = g(t).g'(s)$.

Pour $s \to 0$, g' étant supposée continue, on a à la limite : $g'(t) = g(t).g'(0)$.

En posant $g'(0) = -\lambda > 0$ et sachant que $g(0) = 1$, on obtient :

$$g(t) = e^{-\lambda t} \text{ pour } t > 0.$$

On retrouve donc la fonction de répartition de la loi de Z : $F_Z(t) = 1 - e^{-\lambda t}$, pour $t > 0$. Sa densité est obtenue par dérivation :

$$f_Z(t) = \lambda\, e^{-\lambda t}.$$

De cette loi exponentielle, on peut déduire (cela n'est pas immédiat) la loi du nombre N aléatoire des événements *A* qui se produisent entre les instants 0 et T donnés :

$$N \in \mathbf{N} \text{ et } \mathbb{P}(N = k) = \frac{(\lambda T)^k}{k\,!} e^{-\lambda T}$$

C'est la loi de Poisson de paramètre λT, notée $P(\lambda T)$. On a $E(N) = \lambda T$ et $Var(N) = \lambda T$. Cette loi peut aussi être considérée comme loi limite de la loi binomiale.

5 – De la loi binomiale à la loi de Poisson

On part des mêmes hypothèses heuristiques d'un processus de Poisson. On étudie le phénomène sur un intervalle de temps [0, t] fixé.

Soit N le nombre (aléatoire) d'apparitions de l'événement *A* dans cet intervalle de temps. Quelle est la loi de N ?

A nouveau, discrétisons le problème. Découpons $[0, \tau]$ en n intervalles de durée Δt : $\tau = n\,\Delta t$. La probabilité que *A* apparaisse dans l'un quelconque de ces intervalles est

$$p = P(\Delta t) \sim \lambda \Delta t = \frac{\lambda\tau}{n}.$$

Pour chacun de ces intervalles, l'apparition de *A* réalise une épreuve de Bernoulli de paramètre p. Ces épreuves sont indépendantes (troisième hypothèse de modèle). L'événement « N = k » est réalisé si *A* apparaît dans k quelconques de ces intervalles, pris parmi les n. N suit donc une loi binomiale *B* (n, p) :

$$\mathbb{P}(N = k) = \binom{n}{k} p^k (1-p)^{n-k} =$$

$$= \frac{1.(1-\frac{1}{n}).(...).(1-\frac{k-1}{n})}{k!} n^k p^k (1-p)^{n-k}.$$

L'entier k étant fixé, quand n tend vers l'infini le numérateur tend vers 1, np tend vers $\lambda\tau$, et comme $\ln(1-p)^{n-k} \sim (n-k).(-p)$

$$\sim (n-k)\frac{-\lambda\tau}{n} \to -\lambda\tau,$$

il vient :

$$\mathbb{P}(N = k) \to \frac{(\lambda\tau)^k}{k\,!} e^{-\lambda\tau}.$$

La loi de N tend donc vers la loi de Poisson de paramètre $\lambda\tau$.

De plus, $E(N) = np \to \lambda\tau$, espérance de cette loi, ainsi que $Var(N) = np(1-p) \to \lambda\tau$, variance de la loi de Poisson.

IV – La loi normale

1 – Phénomènes gaussiens, observations statistiques, contexte naturel de la loi normale

La loi normale (ou de Laplace-Gauss, appelée « normale » par Pearson en 1893) est la loi de certains phénomènes continus qui fluctuent autour d'une valeur moyenne μ, de manière aléatoire, résultante d'un grand nombre de causes algébriquement additives et indépendantes. C'est une illustration du théorème le plus important du calcul des probabilités, le théorème de la limite centrée (TLC, appelé aussi « théorème central limite », traduction de « the central limit theorem », dénomination introduite par G. Polya en 1920).

La dispersion des valeurs observées d'un même caractère gaussien est représentée par un écart type σ. L'observation statistique de ces phénomènes conduit à la remarque que 95 % environ des observations se situent dans l'intervalle $]\mu-2\sigma;\mu+2\sigma[$, dit « plage de normalité ». (Doc. GEPS d'accompagnement du programme de 1ère L).

La loi normale est, entre autres, la loi des erreurs de mesure d'un phénomène physique. Historiquement, c'est par l'étude du comportement de ces erreurs que Laplace (1777, 1812) puis Gauss (1821) ont mis en évidence l'expression mathématique de cette loi.

Mais tout le mérite revient à Moivre (1718), qui, en appliquant la formule de son ami Stirling, en a obtenu l'expression en cherchant des équivalents des probabilités binomiales pour améliorer le contrôle de l'écart $|F_n - p|$ entre fréquence observée et probabilité limite dans le problème de Bernoulli. (Théorème de Moivre-Laplace).

2 – Modèle probabiliste de la loi normale

Comme précédemment, pour modéliser un phénomène gaussien, nous allons distinguer sa description heuristique du modèle probabiliste que l'on peut proposer pour la représenter.

Hypothèse de travail : on étudie un caractère aléatoire continu (en dimension 1), pouvant prendre n'importe quelle valeur réelle, symétriquement autour d'une valeur centrale μ. Une étude statistique montre que pour un grand nombre d'observations, 95% d'entre elles se situent dans un intervalle centré en μ et de demi-longueur 2σ.

Pratiquement, dans beaucoup d'applications, les valeurs possibles du caractère sont nécessairement positives (mesures de grandeurs, quantités, …). Nous verrons qu'il n'y a pas de gros inconvénient à proposer un modèle continu sur tout $\mathbf{R}$, si la probabilité est surtout concentrée autour de μ. Cette propriété du modèle rendra compte du fait que la quasi-totalité des valeurs observées se concentrent autour de cette valeur centrale.

Hypothèse de modèle : On prend donc comme ensemble référentiel, $\Omega = \mathbf{R}$.

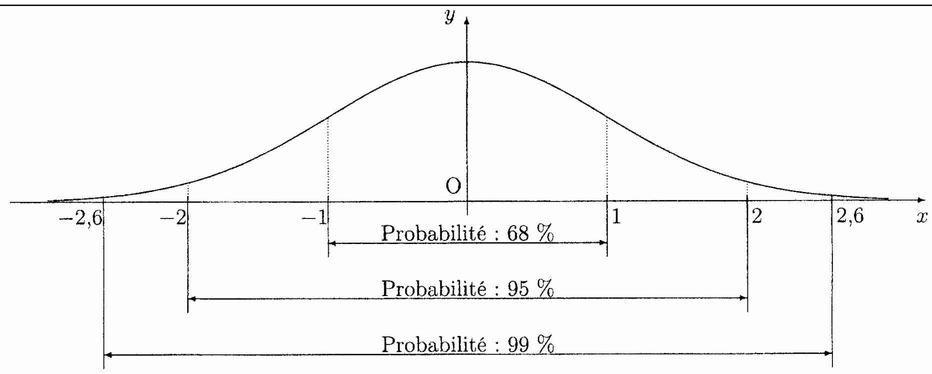
Soit X une v. a. définie sur Ω à valeurs dans $\mathbf{R}$ représentant les occurrences possibles du caractère étudié. La loi $\mathbb{P}_X$ d'une variable continue X de densité f_X est entièrement déterminée par les probabilités

$$\mathbb{P}_X(]a, b]) = \mathbb{P}(X \in]a, b]) = \int_a^b f_X(x)\,dx\,.$$

Sa fonction de répartition permet le calcul de toutes les probabilités des événements liées à X :

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt\,.$$

X est une variable normale si sa densité f_X est donnée par :



Valeurs précises : $\mathbb{P}(|U| \leq 1{,}65) = 0{,}9$; $\mathbb{P}(|U| \leq 1{,}96) = 0{,}95$; $\mathbb{P}(|U| \leq 2{,}58) = 0{,}99$

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

Cette loi est désignée par $N(\mu, \sigma)$. On ramène toutes les lois normales à un même type de base, $N(0, 1)$, en réduisant la variable X : on pose $U = \frac{X-\mu}{\sigma}$.

La densité de la loi de U est alors :

$$f_U(u) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2}.$$

La représentation graphique de f_U est la fameuse courbe en cloche dite « courbe de Gauss » (ci-dessus).

3 – Le théorème limite central (TLC)

Nous allons voir pourquoi cette loi s'impose « naturellement ». Reprenons la question historique des erreurs de mesure. Mesurer une grandeur consiste à mettre en œuvre un phénomène physique qu'un instrument de mesure est susceptible d'enregistrer sous la forme d'une donnée numérique, dont la précision dépend de la sensibilité de l'appareil et des aléas qui accompagnent inéluctablement toute procédure de mesure. Soit X la variable aléatoire qui à une telle opération fait correspondre la valeur affichée par l'appareil.

Remarque de physicien : si on recommence n fois la même mesure, on obtient n valeurs entachées des « erreurs de mesure ». En prenant leur moyenne, on obtient une valeur plus précise, ayant amélioré virtuellement la sensibilité des appareils en réduisant la dispersion des résultats. Cette propriété, connue depuis longtemps, peut être expliquée simplement par un petit raisonnement probabiliste.

En effet, si on désigne par X_i les différentes v.a. représentant les mesures successives, ces X_i sont des répliques indépendantes de la v.a. X. La valeur retenue pour la grandeur à mesurer sera donc :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

On peut faire deux remarques :

— Si on suppose que les X_i se répartissent aléatoirement autour de la valeur μ à mesurer (on suppose qu'il n'y a pas d'erreur systématique : $E(X_i) = \mu$), il en est de même pour $\overline{X}$ dont la valeur moyenne est :

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

— Mais, tenant compte de l'indépendance des X_i, on a d'après une propriété de base de la variance : $\mathrm{Var}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}.n.\sigma^2$.

La dispersion de $\overline{X}$ est représentée par l'écart type $\sigma/\sqrt{n}$. $\overline{X}$ est donc plus concentrée autour de μ que chacune des X_i, et ceci d'autant plus que n est grand : avec n = 100, on gagne un ordre de grandeur sur la sensibilité des instruments.

Un théorème (facile) dit que si X et les X_i sont des variables normales de loi $N(\mu, \sigma)$, alors $\overline{X}$ est aussi une variable normale, mais sa loi est $\overline{X} \sim N(\mu, \sigma/\sqrt{n})$.

Le théorème limite central, théorème beaucoup plus fort, dit que :

Si $(X_n)_{n\in\mathbf{N}}$ *est une suite de v.a. indépendantes et de même loi, telle que* $E(X_n) = \mu$ *et* $\mathrm{Var}(X_n) = \sigma^2$*, alors la suite des variables « moyennes réduites »* $Z_n = \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n}$ *« converge en loi » vers une variable normale* U *centrée réduite :* $U \sim N(0, 1)$.

La loi commune des X_n peut-être tout à fait quelconque, la seule condition est que son espérance et sa variance existent. Dans cet énoncé du TLC, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ et « converge en loi » veut dire que la suite des fonctions de répartition des Z_n converge simplement vers celle de U :

Pour tout z réel, $\mathbb{P}(Z_n \le z) \to \mathbb{P}(U \le z)$.

Même si on ne sait rien sur le comportement des X_i, le TLC dit que pour n assez grand (en pratique, on a une assez bonne précision dès que n > 50), on connaît approximativement la loi de $\overline{X}$: c'est presque une loi normale $N(\mu, \sigma/\sqrt{n})$.

On peut alors calculer des probabilités telles que $\mathbb{P}(|\overline{X} - \mu| < \varepsilon)$, permettant par exemple :

— de déterminer des intervalles de confiance :

Pour un niveau de confiance $1 - \alpha$ donné, trouver ε tel que $\mathbb{P}(\overline{X} \in]\mu-\varepsilon, \mu+\varepsilon[) > 1 - \alpha$. En prenant le risque α de se tromper, on peut alors affirmer que $\mu \in]\overline{X} - \varepsilon, \overline{X} + \varepsilon[$.

— ou de faire un test d'hypothèses :

Si sous une certaine hypothèse à tester H_0, la probabilité $\alpha = \mathbb{P}(|\overline{X} - \mu| > \varepsilon)$ est très petite, et si on observe $\overline{X}$ en dehors de l'intervalle $]\mu-\varepsilon, \mu+\varepsilon[$, on rejette H_0 en prenant le risque α de se tromper.

Le TLC est un outil puissant de contrôle de phénomènes qui sont des moyennes arithmétiques. La loi normale s'impose dans son énoncé. Sa démonstration fait intervenir des objets mathématiques de haut niveau accompagnant les lois continues (transformation de Fourier).

4 – Exemple historique de base : le théorème de Moivre - Laplace

Plaçons nous dans la situation du problème de Bernoulli : on répète n fois une même épreuve de Bernoulli (expérience aléatoire à 2 issues, succès de probabilité p ou échec de probabilité (1–p)).

On sait (théorème de Bernoulli) que les fréquences F_n des succès « tendent à se stabiliser » vers p quand n devient très grand. Plus précisément, $\mathbb{P}(|F_n - p| < \varepsilon) \to 1$ quand $n \to \infty$.

Mais comment contrôler cette « stabilisation », c'est à dire le degré de précision obtenu quand on évalue p par la valeur observée de la fréquence F_n ?

La technique consiste à construire un intervalle de confiance pour p, dont les bornes dépendent de la valeur observée de la fréquence des succès pour un n donné.

Modélisons cela : à chaque épreuve on associe la v.a. de Bernoulli : $X_i = 1$ si succès (de probabilité p), 0 si échec ((1–p)).

On a $E(X_i) = p$ et $Var(X_i) = p(1-p)$.

Or, $F_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, puisque ce Σ est égal au nombre de succès obtenus en n épreuves.

On a $E(F_n) = p$ et $Var(F_n) = \frac{p(1-p)}{n}$.

On peut donc appliquer le TLC à F_n, ce qui donne le théorème de Moivre-Laplace :

Si n est assez grand, la loi de F_n *est « proche » de la loi normale*

$$\mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

Pratiquement, cette approximation est acceptable quand n > 50 et si p n'est pas trop voisin de 0 ou 1.

Comme nF_n est égal au nombre de succès en n épreuves de Bernoulli, c'est une variable binomiale de loi $\mathcal{B}(n, p)$. Le théorème de Moivre-Laplace donne l'approximation en loi de la loi binomiale :

$$\mathcal{B}(n, p) \approx \mathcal{N}\left(np, \sqrt{np(1-p)}\right)$$

dès que n > 50 et p pas trop voisin de 0 ou 1.

On peut aussi énoncer cette propriété en considérant la fréquence réduite sous la forme :

$$Z_n = \frac{F_n - p}{\sqrt{p(1-p)}}\sqrt{n} \approx \mathcal{N}(0, 1).$$

La connaissance de la loi normale centrée réduite permet alors le contrôle souhaité :

$$\mathbb{P}(p \in]F_n - \varepsilon\,;\, F_n + \varepsilon[) = \mathbb{P}\left(\frac{|F_n - p|}{\sqrt{p(1-p)}}\sqrt{n} < \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \cong \mathbb{P}\left(|U| < \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 1 - \alpha$$

Le niveau de confiance $1 - \alpha$ étant fixé (par exemple 0,95), la précision ε de l'estimation est la plus petite valeur vérifiant cette dernière inégalité. Elle peut donc être obtenue à partir du fractile correspondant de la loi $N(0, 1)$. C'est la démarche de base pour déterminer les intervalles de confiance pour les sondages aléatoires.

5 – Intervalle de confiance pour une proportion (sondages)

On a donc obtenu :

$$\mathbb{P}(p \in]F_n - \varepsilon\ ;\ F_n + \varepsilon[) = \mathbb{P}(|U| < \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}) \geq 1 - \alpha\ ,$$

condition de confiance, où $U \sim N(0, 1)$.

Désignons par $u_{\alpha/2}$ le fractile d'ordre $\alpha/2$ de cette loi : $\mathbb{P}(U \leq u_{\alpha/2}) = 1 - \alpha/2$. Avec cette notation, on a $\mathbb{P}(|U| \leq u_{\alpha/2}) = 1 - \alpha$. Par exemple (lecture dans la table), pour $\alpha = 0{,}05$, on trouve $u_{\alpha/2} = 1{,}96$.

La condition de confiance est donc réalisée avec $u_{\alpha/2} = \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}$, d'où $\varepsilon = \frac{u_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}$, variant comme $\frac{1}{\sqrt{n}}$. Mais p est inconnu, c'est justement la proportion à estimer !

On peut cependant majorer $\sqrt{p(1-p)}$, car si on augmente ε, on élargit l'intervalle de confiance et la probabilité $\mathbb{P}(p \in]F_n - \varepsilon\ ;\ F_n + \varepsilon[)$ augmente, la condition de confiance est alors d'autant mieux vérifiée.

Or $p(1-p) \leq 1/4$ est acceptable si p est assez voisin de 1/2. (entre 0,3 et 0,7). De plus, si $\alpha = 0{,}05$ on a $u_{\alpha/2} = 1{,}96 < 2$, d'où $\varepsilon < \frac{1}{\sqrt{n}}$ est une majoration acceptable dans ces conditions. On obtient donc la formule simplifiée de l'intervalle de confiance proposé comme fourchette d'échantillonnage dans un thème d'études du programme de seconde :

$$\mathbb{P}(p \in]F_n - \frac{1}{\sqrt{n}}\ ;\ F_n + \frac{1}{\sqrt{n}}[) \geq 0{,}95\ .$$

De manière plus générale, on peut voir que $\frac{F_n - p}{\sqrt{F_n(1-F_n)}}\sqrt{n}$ converge aussi en loi vers U (application d'un lemme sur la convergence en loi, en remarquant que $\frac{F_n(1-F_n)}{p(1-p)} \xrightarrow{p.s.} 1$ d'après la loi forte des grands nombres). On en tire une forme plus précise de l'intervalle de confiance pour la proportion p au niveau $1 - \alpha$:

$$\mathbb{P}(p \in]F_n - \frac{u_{\alpha/2}\sqrt{F_n(1-F_n)}}{\sqrt{n}}\ ;\ F_n + \frac{u_{\alpha/2}\sqrt{F_n(1-F_n)}}{\sqrt{n}}[) \cong 1 - \alpha$$

D'où la « fourchette de sondage » théorique pour estimer p au niveau de confiance $1 - \alpha$, à partir d'un échantillon de taille n pour lequel la valeur observée de F_n est f_n :

$$]f_n - \frac{u_{\alpha/2}\sqrt{f_n(1-f_n)}}{\sqrt{n}}\ ;\ f_n + \frac{u_{\alpha/2}\sqrt{f_n(1-f_n)}}{\sqrt{n}}[$$

V – Des lois continues, pour quoi faire en Terminale ?

1 – La notion de loi dans les classes de Première

Les programmes des premières S, ES et L introduisent les probabilités d'emblée par la notion de loi (ou de distribution) de probabilités, c'est-à-dire une répartition de la certitude en un ensemble fini de valeurs associées aux événements élémentaires qui représentent les issues d'une expérience aléatoire. La notion même de probabilité n'est pas objet d'étude, le programme se limitant aux calculs portant sur des probabilités d'événements.

La définition proposée dans le document d'accompagnement est qu'« *une loi de probabilités est un objet mathématique ayant les mêmes propriétés qu'une distribution de fréquences* ». Cela signifie simplement que les éléments de cette loi (les probabilités élémentaires) sont des nombres compris entre 0 et 1 et dont la somme est égale à 1.

Mais le terme d'« objet mathématique » a son importance : il signifie clairement que la probabilité n'est pas une grandeur de la réalité, elle fait partie des notions mathématiques qui contribuent à la modélisation de cette réalité.

Ce point de vue est en rupture avec celui de l'ancien programme de Première, qui faisait de la probabilité une sorte de limite de fréquence stabilisée. Du coup elle appartenait au même paradigme, celui de la description de la réalité et non de sa modélisation. Nous nous associons résolument à ce point de vue, car il va permettre d'introduire en Terminale S les lois continues dans cette cohérence épistémologique.

Le programme précise :

« *Le lien entre loi de probabilité et distribution de fréquences sera éclairé par un énoncé vulgarisé de la loi des grands nombres* ».

Concernant la loi « faible », notamment le théorème de Bernoulli, ce lien n'est pas simple à expliciter. En effet, comme nous l'avons vu, il fait intervenir deux sortes de probabilités : la probabilité p, affectée à une issue d'une même expérience aléatoire répétée n fois, et la probabilité que la fréquence observée de cette issue ne s'écarte pas de p de plus qu'un ε. Historiquement, l'assimilation de ces deux notions de probabilité ne s'est pas faite sans difficultés. Laplace lui-même dans son *essai philosophique sur les probabilités* (1814), utilise le terme de « possibilité » à l'égard de la deuxième.

Dans l'esprit du programme, il conviendrait d'installer un « méta-modèle » pour représenter cette convergence (en probabilité) de la fréquence, qui fait l'objet de la loi faible des grands nombres. Bien évidemment, cela est exclu au niveau secondaire, ce qui explique pourquoi l'énoncé « vulgarisé » proposé par le programme est très vague :

« *Pour une expérience donnée, dans le modèle défini par une loi de probabilité P, les distributions des fréquences calculées sur des séries de taille n se rapprochent de P quand n devient grand* ».

Pour expliciter cet énoncé sans soulever cette difficulté épistémologique, le document d'accompagnement s'appuie sur la loi « forte », théorème dû à Émile Borel (1909) : dans certaines conditions, la suite des fréquences de réalisation d'une issue en n épreuves d'une même expérience aléatoire, converge (presque sûrement, i.e. avec probabilité 1) vers la probabilité de cette issue. L'ensemble des k issues considérées étant fini, il en est de même pour la suite des distributions de ces fréquences qui converge (presque sûrement dans $\mathbf{R}^k$) vers la loi de probabilité qui devra être admise pour un modèle pertinent de cette expérience.

Notons que pour tous les exemples proposés, invitant à la modélisation d'expériences de référence éventuellement simulées dans un environnement informatique, les lois de base envisagées sont toutes fondées sur l'hypothèse d'équiprobabilité. Cette hypothèse de modèle est supposée découler "naturellement" de symétries admises en hypothèse de travail (dés, pièces, urnes, cartes, chiffres au hasard équirépartis, …). On peut comprendre cette option qui permet d'échapper à la question de l'estimation statistique des probabilités élémentaires à introduire dans un modèle où l'équiprobabilité ne serait pas acceptable. Il ne peut donc être question de modéliser le lancer d'une punaise …

En Première S, un modèle de non équiprobabilité pourra être conçu comme un modèle image par une variable aléatoire dont les valeurs regroupent diverses issues d'une loi de probabilité équirépartie. Ce sera notamment le cas en Terminale de la loi binomiale.

2 – Quelques remarques sur les lois continues au programme de TS

On a vu comment introduire la loi uniforme en termes de proportionnalité, ce qui n'est pas sans difficulté pour les élèves. Dans l'exemple du jeu de franc-carreau, la réalité est décrite dans un cadre géométrique faisant naturellement intervenir des grandeurs continues.

La discrétisation en dimension 2 a permis de relier la distribution uniforme continue sur un carré, où les probabilités s'expriment en termes de rapports d'aires, à l'équiprobabilité de base. Cette démarche peut être transférée facilement à des exemples conduisant à une loi uniforme sur [0, 1], les probabilités s'exprimant en termes de rapports de longueurs (dommage que celle de [0, 1] soit égale à 1 !).

Il n'est donc nul besoin pour le modèle uniforme de recourir à des calculs d'intégrales, et la notion de densité ne semble pas s'imposer.

La loi exponentielle est désignée dans le programme par « loi de durée de vie sans vieillissement », par référence aux conditions heuristiques qui conduisent à ce modèle continu, telles que nous les avons proposées.

La discrétisation et le passage par la loi géométrique, qui permettent de comprendre pourquoi la densité de probabilités intervenant dans ce modèle est de type exponentiel, sont hors d'atteinte en terminale.

La détermination directe de la fonction de répartition de cette loi pourrait être accessible, bien que de bon niveau pour des élèves de terminale (indépendance d'événements, équation fonctionnelle et intégration d'une équation différentielle).

On peut se limiter à présenter le résultat en affirmant que la relation :

$$\mathbb{P}(a < X < b) = \int_a^b \alpha\, e^{-\alpha x}\, dx ,$$

où X est la durée de vie et α un paramètre positif caractérisant le phénomène (la cadence), est un moyen mathématique pratique pour calculer les probabilités de cette sorte d'événements dans un modèle continu assez bien adapté à ce type de situations concrètes. Les exemples sont nombreux : pannes, files d'attente, …

L'exemple de la désintégration radioactive permet une bonne illustration de cette loi, mettant en œuvre les connaissances acquises en analyse. En marge du programme (en TER ?), il permet aussi un sérieux travail de synthèse et de prolongements. D'une part il met en œuvre de manière non évidente le théorème de Bernoulli comme exemple d'application de la loi des grands nombres (non vulgarisée), pour relier la période d'un élément radioactif au paramètre de la loi exponentielle qui modélise la durée de vie d'un atome. D'autre part il permet de manipuler l'approximation en loi avec le cas de l'approximation de la loi binomiale par une loi de Poisson.

3 - Notion de densité, prolongements des outils probabilistes pour des modèles performants

Au-delà du programme de Terminale S, les exemples simples des lois uniformes et des lois exponentielles permettront de dégager la notion fondamentale de densité de probabilité. Elle donnera lieu à une utilisation intensive du calcul intégral et à un sérieux investissement des outils de l'analyse.

C'est bien sûr par sa densité que le modèle gaussien peut être introduit. Malgré son importance, il ne l'a pas (encore ?) été au niveau des Terminales S (il faut bien faire des choix, et la manipulation de l'intégrale de Gauss n'est pas aussi évidente que celle de la fonction exponentielle). La loi normale figure dans certains programmes de BTS comme exemple de base, pour ses multiples applications à la statistique inférentielle notamment : sondages, estimations, tests d'hypothèses, …

Les modèles à Ω fini ne concernent que les situations pas trop vastes, avec équiprobabilité quelque part. Le passage aux modèles continus permet une meilleure appréhension de classes de phénomènes inaccessibles aux modèles discrets. Mais il suppose le développement d'outils de calcul performants (intégrales), dans des cadres théoriques nouveaux (axiomatique de Kolmogorov, 1933), dans lesquels peuvent être obtenus des théorèmes puissants (théorème limite central, loi forte, etc.), permettant de résoudre de nouveaux problèmes.

L'introduction de lois continues en Terminale, outre la résolution de problèmes plus intéressants que les problèmes traditionnels de combinatoire à propos de jeux de hasard, a pour importance essentielle de rendre incontournable la notion de modèle. La simulation informatique de situations simples y concourt aussi.

Le bénéfice qu'en tireront les élèves sera sans doute une compréhension plus claire du statut du calcul des probabilités comme un outil mathématique des plus performants pour maîtriser les problèmes concrets issus de la complexité de la réalité, aussi bien dans les domaines des sciences de la nature que dans les sciences économiques ou dans les sciences humaines.