Lois continues en Terminale S, quelle approche ?

Michel Henry(*)

1 – Le programme et le document d'accompagnement

La notion de loi de probabilité a été introduite en Première. La définition est abstraite et la loi de base privilégiée est la loi uniforme discrète (équiprobabilité). En terminale S, le programme propose l'étude des lois discrètes de Bernoulli et binomiale et introduit deux lois continues « à densité » : la loi uniforme sur [0, 1] et la loi exponentielle appelée « loi de durée de vie sans vieillissement » pour souligner l'approche heuristique qui est préconisée.
Le document d'accompagnement explicite l'expression « Prendre un nombre au hasard dans [0, 1] » qui ne peut prendre de sens que par référence à la loi continue uniforme, ce qui montre la nécessité de cette nouvelle notion.
L'atelier avait pour objectif d'approfondir le concept de loi de probabilité introduit en première en vue de son utilisation en terminale dans le cas des lois continues. Son enjeu était de montrer en quoi la démarche de modélisation peut clarifier cet enseignement.

2 – Qu'est-ce qu'une loi de probabilité ?

La définition en première ne vaut que dans le cas discret. Sa généralisation au cas continu en terminale n'est pas élémentaire. Le programme opte pour une définition analytique via la densité de probabilité. On peut remarquer que pour les deux exemples étudiés, cette notion n'est pas nécessaire, la fonction de répartition apparaissant beaucoup plus naturelle et facilement accessible par la voie heuristique pour la loi exponentielle (cf. l'article *Des lois continues, pourquoi et pour quoi faire ?* dans Repères-IREM n° 51, Avril 2003). Mais il n'est pas inutile de revenir à la définition générale de la notion de loi de probabilité.

La réponse du programme de Première (extraits du document d'accompagnement) :
« On définira une loi de probabilité comme un objet mathématique ayant les mêmes propriétés qu'une distribution de fréquences ».
« Une loi de probabilité P sur un ensemble fini est la liste des probabilités des éléments de E ».
« Le fait de ne pouvoir simplement généraliser aux ensembles continus cette définition et la nécessité d'une définition ensembliste seront abordés en terminale ».

La réponse du programme de terminale S (document d'accompagnement) :
– « On se limite à des lois définies sur I = (a, b), … dites *à densités continues* ».

(*) IREM de Franche-Comté.

– Une loi P de densité f définie sur I, continue, positive est donnée par :

Pour tout $(c, d) \subset \mathrm{I}$, $\mathrm{P}((c, d)) = \int_c^d f(x)dx$. Ou si $d = +\infty$, $\mathrm{P}((c, d)) = \lim\limits_{t \to +\infty} \int_c^t f(x)\, dx$.

La fonction de répartition $\mathrm{F}(x) = \mathrm{P}((a, x]) = \int_a^x f(t)\, dt$ est hors programme.

La réponse universitaire

Rappelons le modèle probabiliste général (Kolmogorov, 1933). On se donne :
– Un « univers » Ω pour représenter les issues d'une expérience aléatoire,
– Une famille de parties $\mathcal{T}$ de Ω pour représenter les événements possibles,
– Une « probabilité P » (appelée aussi « distribution de probabilité »), application de $\mathcal{T}$ dans $[0, 1]$, pour représenter les aléas des issues de l'expérience, avec leurs poids respectifs.
– Le triplet $(\Omega, \mathcal{T}, \mathrm{P})$ est appelé « espace probabilisé ». P vérifie les définitions et axiomes de base d'une mesure sur Ω :

a) Pour tout $\mathrm{A} \in \mathcal{T}$, $\mathrm{P(A)} \in [0, 1]$, P(A) est « la probabilité de A ».
b) $\mathrm{P}(\varnothing) = 0$ (événement impossible) et $\mathrm{P(A^c)} = 1 - \mathrm{P(A)}$ ($\mathrm{A^c}$ événement contraire de A).
c) Si pour une suite d'événements $\{\mathrm{A}_n \in \mathcal{T} \mid n \in \mathbf{N}\}$ on a $\mathrm{A}_i \cap \mathrm{A}_j = \varnothing$ pour tout i et j distincts, alors $\mathrm{P}(\cup\, \mathrm{A}_n) = \sum \mathrm{P}(\mathrm{A}_n)$ (σ-additivité).

La notion de variable aléatoire permet le transfert de cette structure abstraite dans un ensemble numérique : $\mathrm{X} : \Omega \to \mathrm{E}$ ($\mathbf{N}$, $\mathbf{R}$, $\mathbf{R}^+$, $\mathbf{R}^d$, …), on en tire la notion de loi de la variable aléatoire X.
On désigne par $\mathcal{B}$ la famille des ensembles « mesurables » de E (boréliens, combinaisons d'intervalles) et on suppose que si $\mathrm{B} \in \mathcal{B}$ alors $\mathrm{X}^{-1}(\mathrm{B}) \in \mathcal{T}$ (X est dite « mesurable »).
On définit $\mathrm{P_X(B)} = \mathrm{P}(\mathrm{X}^{-1}(\mathrm{B}))$. $\mathrm{P_X}$ est une probabilité sur E et $(\mathrm{E}, \mathcal{B}, \mathrm{P_X})$ est l'espace probabilisé image de $(\Omega, \mathcal{T}, \mathrm{P})$ par X. $\mathrm{P_X}$ est appelée « la loi de X ».

Quels outils pour donner une loi ?

– Pour $\mathrm{E} = \mathbf{N}$, on prend $\mathcal{B} = \mathcal{P}(\mathbf{N})$: tout événement est réunion dénombrable d'événements élémentaires de la forme $\{k\}$. La « liste » des $p_k = \mathrm{P_X}(\{k\})$ suffit pour décrire la loi $\mathrm{P_X}$. Les probabilités cumulées $\sum\limits_{i=0}^{k} p_i$ aussi (fonction de répartition).

Exemple : la loi binomiale $\mathcal{B}(n, p)$.
$\Omega = \{0, 1\}^n$, $\mathcal{T} = \mathcal{P}(\Omega)$, $\mathrm{E} = [[0, n]] \subset \mathbf{N}$. Pour $k \in \mathrm{E}$, $\mathrm{A}_k = \mathrm{X}^{-1}(\{k\})$ est l'événement : « ensemble des suites de n éléments 0 ou 1 qui contiennent k chiffres 1 », avec $p_k = \mathrm{P}(\mathrm{A}_k) = \mathrm{P}(\mathrm{X}^{-1}(\{k\})) = \mathrm{P_X}(\{k\})$. On note $p_k = \mathrm{P}(\mathrm{X} = k)$, on a :

$$p_k = \binom{n}{k} p^k (1-p)^{n-k},\ \mathrm{E(X)} = \sum_{k=0}^{n} k p_k = n\,p,\ \mathrm{Var(X)} = n\,p\,(1-p).$$

– Pour $\Omega = \mathbf{R}$ …, on prend $\mathcal{B}$ = ensemble des parties « mesurables » de $\mathbf{R}$, les « boréliens » : tout événement est une partie de Ω que l'on peut obtenir par une suite d'opérations ensemblistes à partir d'intervalles. Les intervalles de la forme $]-\infty, x]$ suffisent pour les décrire. La loi P_X de X est entièrement donnée par la famille des

$$P_X(]-\infty, x]) = P(X \leq x) = F_X(x)$$

où F_X est « la fonction de répartition de la loi de X ». Si F_X est dérivable, remarquant que $F_X(-\infty) = 0$, il suffit de donner sa dérivée f_X, c'est plus simple en pratique. f_X est la « densité de la loi de X ». D'autres moyens existent (fonctions caractéristiques, liste des moments, …).

Exemple : la loi exponentielle $\mathcal{E}(\alpha)$.

$E = \mathbf{R}^+$, B = boréliens de $\mathbf{R}^+$. On pose par définition $F_X(x) = 1 - e^{-\alpha x} = P(X \leq x)$ pour $x \geq 0$.

D'où la densité de la loi de X : $f_X(x) = F'_X(x) = \alpha\, e^{-\alpha x}$ pour $x \geq 0$. La probabilité de l'événement $A = X^{-1}([a, b])$ est :

$$P(A) = P(a \leq X \leq b) = F_X(b) - F_X(a) = e^{-\alpha a} - e^{-\alpha b} = \int_a^b f_X(t)dt\ .$$

Équiprobabilité et lois uniformes

Équiprobabilité = même probabilité pour tous les événements élémentaires. Mais il faut préciser quelle population d'issues on considère. Comme $P(\Omega) = \sum p_k = 1$, Ω est un ensemble fini à n éléments et $p_k = 1/n$: équirépartition de la probabilité, c'est la « distribution uniforme de la probabilité sur Ω ». La locution « Tirage au hasard » signifie le choix au hasard d'un élément d'un ensemble fini, avec équiprobabilité (souvent omis, implicite didactique).

Dans ce contexte, soit A un événement de Ω réalisé par k issues « favorables » parmi les n issues possibles formant Ω. Nécessairement,

$$P(A) = \sum_{e_i \text{ réalisant A}} P(e_i) = \sum_{e_i \text{ réalisant A}} p_i = \frac{k}{n}.$$

On retrouve la définition de Laplace (premier principe) :

$$P(A) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Peut-on « tirer au hasard » un élément dans un ensemble infini dénombrable comme $\mathbf{N}$?

Si $p_k = 1/n$ et n très grand, p_k est proche de 0. À la limite, toutes les probabilités élémentaires sont nulles : $\forall \omega \in \Omega$, $P(\{\omega\}) = 0$ et par conséquent pour tout événement A, $P(A) = 0$!

Pour lever ce paradoxe il convient de mieux comprendre le concept de répartition uniforme pour pouvoir le généraliser.

3 – Remarques épistémologiques et didactiques

La notion de probabilité revêt un caractère dual : calcul *a priori* ou évaluation fréquentiste. Dans la pratique, on rencontre différentes déterminations d'une probabilité :

– L'équiprobabilité est postulée au vu des symétries dans le cas où l'on estime qu'aucune issue ne peut être plus attendue qu'une autre : appréciation subjective de cette équiprobabilité pour un ensemble fini d'issues.
– L'observation de la stabilisation des fréquences d'un événement permet l'estimation expérimentale d'une probabilité.
– Diverses conditions heuristiques conduisent à l'hypothèse d'une loi modèle et à l'estimation de ses paramètres desquels découlent les probabilités des événements décrits dans ce modèle.
– Évaluation par des spécialistes, loi *a priori* (approche bayésienne).
Rappelons les objectifs du programme de Seconde : Fluctuations d'échantillonnage, distributions de fréquences et simulations, … : installer les bases d'une approche expérimentale. Il n'y a pas de travail théorique sur la notion de probabilité.
Objectif du programme de Première : introduction de la notion de loi de probabilité comme modèle théorique sur un ensemble fini. Liste des $p_k \in [0, 1]$ avec $\sum p_k = 1$.
Remarquons qu'une telle famille arbitraire de p_k ne rend pas compte d'un phénomène réel qui a une logique interne. Quels liens y a-t-il entre les p_k ?
Exemples :
– loi uniforme discrète (équiprobabilité) : $p_k = 1/n$,
– loi binomiale (nombre de succès en n expériences de Bernoulli) :

$$p_k = \binom{n}{k} p^k (1-p)^{n-k},$$

– loi géométrique (attente du premier succès) : $p_k = p\,(1-p)^{k-1}$,
– loi de Poisson (nombre d'événements dans un processus sans mémoire) :

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!},$$

– loi hypergéométrique (nombre de succès en n tirages sans remise) :

$$p_k = \frac{\binom{\mathrm{M}}{k}\binom{\mathrm{N}-\mathrm{M}}{n-k}}{\binom{\mathrm{N}}{n}},$$

etc.

4 – Pour aller plus loin, la modélisation

Le choix « au hasard » d'un nombre dans [0, 1] est de probabilité nulle !
Concrètement, quel protocole expérimental pour un tel choix ? Par exemple, l'ordinateur (fonction ALEA d'EXCEL) donne 14 décimales. Il a donné en réalité un intervalle réel de longueur 10^{-14}. Il y en a 10^{14} dans [0, 1], chacun étant « pris au hasard » est de probabilité 10^{-14}. L'ordinateur travaille avec un modèle discret pour simuler la loi continue uniforme.

Modèles et réalité.

Si Ω est seulement conçu comme l'ensemble des issues observables, Ω est fini (réalité).

Toute image E de Ω par une application X est finie (modèle probabiliste). Considérer que E est un ensemble infini de réels, c'est sortir d'une conception naïve du calcul des probabilités, c'est distinguer réalité et modèle, faute de quoi on court le risque de paradoxes tels que celui de Bertrand (cf. *Paradoxes et lois de probabilité* dans Repères-IREM n° 13 d'octobre 1993).

Réalité : expériences aléatoires, protocoles, observations, statistiques des issues, fréquences, distributions des fréquences (f_i), estimation expérimentale des « chances ».

Exemple d'un générateur de hasard : la punaise avec deux issues possibles, « tête » ou « pointe ». Fréquence observée de « tête » en n lancers : f_n.

Modèle : Ensemble structuré d'objets théoriques pour refléter certains aspects de la réalité. Urne pseudo concrète de Bernoulli de paramètre p, modèle mathématique de la loi de Bernoulli : $P(\{1\}) = p$, $P(\{0\}) = 1 - p$, choix de p pour représenter la probabilité de « tête ».

Modèle approché : $p \approx f$. Combien de décimales ?

Plusieurs points de vue concernant cette probabilité p :

– p n'existe pas ! Concept abstrait, objet théorique (De Finetti, subjectiviste),

– p existe et f_n en est une mesure approchée (Alfred Renyi, objectiviste),

– l'existence de p est postulée et f_n en est une estimation (Hélène Ventsel, modélisation).

Approximation de p par des mesures expérimentales : on lance 1 000 fois la punaise pour pouvoir estimer p à 3 % près, avec un risque 0,05 de se tromper.

Modèle aussi précis que l'on veut (loi des grands nombres), si on en a le temps et les moyens.

Modèles pertinents ou pratiques

On représente une situation hypergéométrique par n tirages sans remises de boules d'une urne ayant au départ N boules dont M blanches, avec $n \leq$ M. La probabilité d'obtenir k boules blanches est

$$p_k = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

Le modèle binomial (avec remises) :

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

peut s'avérer plus pratique et suffisamment approché si n est petit devant N et M. Il y a alors un problème de contrôle de l'adéquation d'un modèle à la réalité observée.

Processus de modélisation

Pour l'analyse didactique, je distingue trois étapes dans un processus de modélisation (cf. *Autour de la modélisation en probabilités*, Presses Universitaires Franc-Comtoises, 2000, p. 149-159).

Première étape de la modélisation : l'observation d'une situation réelle et sa description en termes courants. Cette description est déjà une sorte d'abstraction et de simplification de la réalité. Cette description est pilotée par « un regard théorique » s'appuyant sur des modèles généraux pré-construits conduisant à des hypothèses de travail.
Puis il s'agit de traduire cette description en un système simplifié et structuré : c'est le niveau du « modèle pseudo-concret ».
Deuxième étape : la « mathématisation » ou « formalisation » du modèle. Cela suppose que les élèves soient capables de représenter le modèle dans la symbolique propre aux mathématiques, qu'ils sachent interpréter la question posée et les hypothèses de travail en un problème purement mathématique (hypothèses de modèle) et faire appel aux outils mathématiques adaptés pour le résoudre.
Troisième étape : elle consiste à revenir à la question posée pour traduire dans les termes du modèle pseudo-concret les résultats mathématiques obtenus, leur donner du sens pour dégager des réponses et relativiser ces réponses par rapport aux hypothèses de modèle ; il faut ensuite interpréter ces réponses pour apprécier leur validité et leur étendue dans la situation concrète.

5 – Modèles continus. Lois continues

Prenons l'exemple de la loi uniforme sur un ensemble E.
Si E est un ensemble fini, c'est l'équiprobabilité des événements élémentaires (loi uniforme discrète, définition de Laplace).
Sur un ensemble E quelconque, on suppose qu'on a défini une mesure de grandeur μ et que $\mu(E)$ est finie. Par exemple, E et μ peuvent être :
– E une partie bornée de **N** et μ le nombre des éléments (cardinal) des parties de E (loi uniforme discrète),
– E un intervalle borné (a, b) de **R** et μ la longueur des intervalles inclus dans E (mesure de Lebesgue),
– E une partie (quarrable) bornée du plan et μ l'aire de ses parties (quarrables),
– E un solide et μ le volume, etc.
Principe de la loi uniforme sur E : dans le choix « au hasard » d'un élément de E, la probabilité d'obtenir une partie A (réaliser l'événement A) est proportionnelle à la mesure de A : $P(A) = \dfrac{\mu(A)}{\mu(E)}$. Autrement dit, les parties de E de même grandeur représentent des événements équiprobables.
Exemple de la loi uniforme sur $E = [a, b] \subset \mathbf{R}$, avec la longueur : la probabilité qu'un nombre « pris au hasard dans E » appartienne à l'intervalle $[c, d] \subset [a, b]$ est :

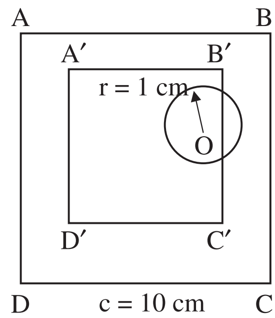
$$P([c,d]) = \frac{d-c}{b-a}.$$

Comme $P(\{x\}) \leq P([x, x + 1/n])$ pour tout n, on a $P(\{x\}) \leq 1/n$, d'où $P(\{x\}) = 0$.

Exemple du jeu de Franc-carreau

Ce jeu, en vogue à la cour de Louis XV, consiste à jeter un écu sur un carrelage. Les joueurs parient sur la position finale de l'écu : entièrement dans un carreau (Franc-

carreau), ou rencontrant des joints (cf. *Des lois continues, pourquoi et pour quoi faire ?* dans Repères-IREM n° 51, Avril 2003).
En 1733, Buffon a proposé une solution théorique, introduisant un nouveau concept de probabilité « géométrique » sur un ensemble continu : les points du carré représentant le carreau dans lequel est tombé le centre de l'écu représenté par un disque.



Hypothèse de travail : tous les points du carré ABCD ont « la même chance » d'être atteints.

Hypothèse de modèle : $\Omega = \{M \mid M \in ABCD\}$.

Quelle loi de probabilité ?

On admet le principe suivant :

Si l'événement A est : « O est tombé dans le domaine A », la probabilité de A est proportionnelle à l'aire de A.

On pose donc :

$$P(A) = \frac{\text{aire de A}}{c^2}.$$

C'est la loi uniforme continue sur le carré ABCD. Dans l'exemple,

$$P(FC) = \frac{\text{aire de A'B'C'D'}}{\text{aire de ABCD}} = 0,64.$$

Quel lien avec la « définition de Laplace » donnée en première avec l'équiprobabilité ?
Discrétisons ce modèle : On tapisse ABCD par un quadrillage de petits carrés unité u_i de côté $c/1\,000$ par exemple. Soit $\Omega = \{u_i \mid 1 \leq i \leq 10^6\}$.
Hypothèse de modèle discret : tous les u_i sont équiprobables : $\forall i$, $P(u_i) = p = 1/10^6$.
Événement « Franc-carreau » (FC) : O est tombé dans l'un des u_i qui tapissent A′B′C′D′.

$$P(FC) = \frac{\text{nombre de petits carrés dans A'B'C'D'}}{10^6} = 0,64.$$

On retrouve le modèle continu par la définition élémentaire de l'aire :

$$P(FC) = \frac{\text{aire A'B'C'D'}}{\text{aire ABCD}}.$$

Le modèle discret est approché si le quadrillage n'est pas assez fin et s'il ne recouvre pas exactement les carrés ABCD et A′B′C′D′.

L'atelier s'achève avec la présentation de l'exemple célèbre d'application d'un modèle uniforme : l'aiguille de Buffon (cf. *Des lois continues, pourquoi et pour quoi faire ?* dans Repères-IREM n° 51, Avril 2003).