L'HISTOGRAMME : UN LEVIER POUR INTRODUIRE LES LOIS A DENSITE EN TERMINALE SCIENTIFIQUE

Charlotte Derouet[1]

[1] Laboratoire de Didactique André Revuz, Université Paris Diderot – Case 7014
75205 Paris Cedex 13, charlotte.derouet@univ-paris-diderot.fr

Résumé. L'histogramme est une des représentations graphiques permettant de rendre compte de certaines données statistiques, au programme du collège dès le début du cycle 4 (classe de cinquième). Pourtant cette notion est souvent mal définie et les graphiques sont souvent erronés, à commencer dans les manuels. Dans cette communication, nous souhaitons redonner du sens à l'histogramme et notamment montrer son intérêt pour introduire la notion de fonction de densité de probabilité et donc les lois à densité en terminale scientifique. Après avoir fait un rapide état des lieux des erreurs que l'on peut trouver dans les manuels concernant l'histogramme, nous revenons sur la définition de cet objet mathématique. Ensuite, nous présentons un problème de modélisation qui a été conçu puis expérimenté dans des classes de terminale S pour introduire la notion de fonction de densité en s'appuyant sur la notion d'histogramme.

Mots-clés. Statistique, probabilités, histogramme, fonction de densité, enseignement, lycée.

Abstract. The histogram is one of graphical representations to account of statistical data, in the curriculum of middle school from the grade 7. But this notion is often bad defined and we can remark some mistakes on these graphics in textbooks. In this communication, we wish to make sense about the histogram and show his interest to introduce the notion of density function and so continuous distributions in the scientific track of the grade 12. We begin with an inventory of errors on the histogram that we can observe in textbooks, and we give a definition of this mathematical object. Then we present a modelling problem that has been designed and experimented in 12th grade of the scientific track to introduce the density function based on the histogram.

Keywords. Statistics, probability, histogram, density function, teaching, highschool.

1 L'histogramme ou la représentation graphique mal aimée

Bien que présente dans les programmes de mathématiques dès le cycle 4 (à partir de la cinquième), la notion d'histogramme est pourtant mal connue des élèves tout au long de leur scolarité secondaire mais aussi des enseignants. Il suffit d'ouvrir les manuels de collège et lycée pour rencontrer des « histogrammes » qui n'en sont pas, des définitions de la notion qui ne sont pas satisfaisantes… Une analyse de manuels de la classe de seconde faite par Derouet & Parzysz (2016) a montré les différentes erreurs que l'on peut trouver sur cette notion notamment il y a parfois des histogrammes qui ne sont rien d'autre que des diagrammes en bâtons (cf. figure 1).

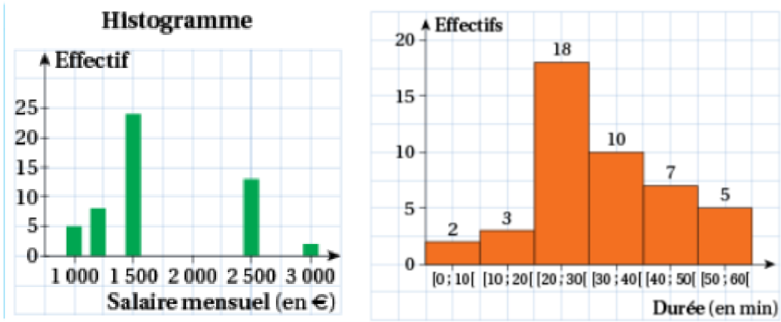


Figure 1. Extraits du manuel Odyssée seconde[1], p. 265 et p. 281

Pour le graphique de droite, cela peut paraître subtil mais que serait le graphique si les intervalles considérés n'étaient pas de même amplitude ? Le tableur qui appelle « histogramme » le diagramme en bâton est certainement une des explications à la mauvaise appréciation de cet objet. L'analyse précise des manuels nous a permis de conclure que la notion d'histogramme est très mal maîtrisée par les auteurs de manuels et donc de formuler l'hypothèse qu'il en est de même pour les élèves. Une étude de Roditi (2009) fait apparaître que les enseignants, au lycée comme au collège, ne consacrent que très peu de temps à l'histogramme (environ une heure par an) et que près de la moitié des enseignants interrogés estiment que « *l'enseignement de l'histogramme ne contribue pas à la formation mathématique des élèves* » (p. 131).

Le manque de sens ainsi que le manque d'intérêt porté à cette représentation graphique expliquent ce constat. Cependant, tout comme le précise Roditi (2009), nous pensons que l'histogramme peut être un appui important et pertinent pour introduire la notion de fonction de densité de probabilité et donc les lois à densité notamment en terminale scientifique (Derouet, 2016 ; Derouet & Parzysz, 2016). En effet, le lien entre histogramme et courbe de densité est à mettre en correspondance avec le lien entre fréquence et probabilité qui est utilisé en classe dès le collège. Il s'agit implicitement de la loi faible des grands nombres, dont la formulation vulgarisée proposée (dans le cas fini) dans le document d'accompagnement du programme de 2001 de la classe de première S (MEN, 2001) est la suivante : « *Si on choisit n éléments d'un ensemble fini E selon une loi de probabilité P, indépendamment les uns des autres, alors la distribution des fréquences est proche de la loi de probabilité P lorsque n est grand* ». En effet, si par exemple on simule plusieurs échantillons de taille suffisamment grande de variables aléatoires suivant une loi uniforme sur $[0\,;\,1]$, on peut voir que le haut des rectangles des histogrammes est « proche » de la courbe représentative d'une fonction constante sur [0 ; 1].

Encore faut-il dans un premier temps bien définir ce qu'est un histogramme et lui donner du sens pour pouvoir s'en servir pour introduire une autre notion. Tout d'abord il nous semble indispensable, avant de définir un histogramme, de définir ce qu'est la densité de fréquence. Une analogie qui nous semble adaptée pour les élèves et que nous reprenons de Brezout & Kahané (2010) est celle avec la densité de population. Nous reprenons leur exemple :

« En 2005, Monaco avait 32 543 habitants et la Japon 127 417 244 (source : Institut nationale d'études démographiques). Bien sûr, les démographes diront que ces renseignements sont très largement insuffisants pour comparer la démographie des deux pays : il faut au minimum s'intéresser aux superficies de ces deux pays et calculer pour chacun la densité de population, c'est-à-dire le nombre d'habitants au kilomètre carré. Avec une superficie de 2,02 km^2 pour Monaco et de 378000 km^2 pour le Japon, les densités sont respectivement $d_1 = \frac{32\,543}{2{,}02} = 16\,110{,}40 h/km^2$ pour Monaco et $d_2 = \frac{127\,417\,244}{378\,000} = 337 h/km^2$ pour le Japon. Autrement dit, alors que la population de Monaco est la moins importante en taille, sa densité est plus importante que celle du Japon » (p. 11).

[1] Sigward, E. (dir.). (2010). *Odyssée. Mathématiques. 2de.* Paris : Hatier.

Ainsi, une représentation pertinente des populations de Monaco et du Japon doit rendre visible cette différence de densité. Il en est de même quand l'on cherche à représenter des fréquences de classes (notamment si elles sont d'amplitudes différentes) : il n'est pas pertinent, par exemple, de représenter de la même façon la fréquence égale à 0,05 de la tranche d'âge [6 ; 11[et de la tranche d'âge [80 ; 105[. Ces deux classes d'âge ont la même fréquence mais l'amplitude de la première classe est inférieure à celle de la seconde, on pressent que la densité de la première classe est supérieure à celle de la seconde. Par analogie avec la densité de population utilisée en géographie, on calculera le quotient de la fréquence d'une classe d'âge par l'amplitude de cette classe, et on parlera de densité de fréquence. Comme perçue, la densité de fréquence de la première tranche d'âge est plus grande que la densité de fréquence de la seconde tranche d'âge. La représentation graphique, qui n'est autre que l'histogramme, doit rendre visible ce phénomène. Nous proposons les définitions suivantes :

- Dans le cas d'une variable quantitative continue, on définit la densité de fréquence d_i d'une classe de fréquence f_i et d'amplitude a_i par : $d_i = \frac{f_i}{a_i}$.
- Un histogramme de fréquences est un diagramme composé de rectangles collés dont les aires sont proportionnelles aux fréquences et dont les bases sont déterminées par les intervalles des classes.

Nous trouvons donc que l'axe des ordonnées représente la densité de fréquence, et non la fréquence comme on peut le voir très souvent dans les manuels (cf. figure 1). Bien entendu, fréquence et densité sont égales (ou tout du moins proportionnelles) lorsque les amplitudes des classes sont égales, mais tout de même.

Maintenant que les choses sont (plus) claires, nous allons présenter une proposition d'introduction de la notion de fonction de densité prenant appui sur la notion d'histogramme expérimentée dans plusieurs classes de terminale S (notamment dans le cadre de la thèse Derouet, 2016).

2 Une introduction de la notion de fonction de densité prenant appui sur l'histogramme

Si l'on regarde les activités d'introduction des manuels de terminale S du chapitre sur les lois à densité, cinq manuels proposent un passage par l'histogramme pour introduire la courbe de densité de probabilités (une analyse détaillée des activités des huit manuels de terminale S de l'édition 2012 a été faite par Derouet (2016)). Cependant beaucoup d'erreurs mathématiques sont présentes notamment à nouveau sur la notion d'histogramme, de plus aucune réelle réflexion sur la fonction (appelée ensuite fonction de densité) n'est attendue de la part des élèves (la courbe est déjà tracée). Le travail est très guidé et finalement les propriétés qui caractérisent une fonction de densité sont imposées aux élèves.

L'idée dans la proposition présentée ci-dessous est que les élèves construisent et donnent réellement du sens à la notion de fonction de densité et donc que les différentes propriétés de la fonction de densité émergent de la situation. Nous présenterons ici seulement le premier des deux problèmes d'introduction de la notion de fonction de densité conçus en collaboration avec une enseignante de terminale S. Nous donnerons aussi des éléments du déroulement effectif dans une classe de terminale S d'un lycée parisien (en 2015).

Il est à prendre en considération qu'en amont de ces séances, les élèves ont fait un devoir maison autour de la notion d'histogramme avec notamment un appui sur les définitions données ci-dessus, pour que cette notion soit réellement disponible chez les élèves. Ils ont aussi eu une première approche de la loi uniforme continue (sans qu'elle soit nommée ainsi) avec une partie « intuitive », une partie simulation pour aboutir à des histogrammes et la recherche d'une « courbe de tendance ». En revanche, aucun travail n'a été fait avant sur le calcul intégral. La séquence a en effet l'originalité d'articuler à la fois les lois à densité et le calcul intégral, mais nous n'insisterons pas sur ce point ici.

Le problème de modélisation donné aux élèves est le suivant :

Karine et Olivier décident de se retrouver au café de l'Hôtel de Ville entre 7h et 8h. Ils peuvent arriver à tout moment entre 7h et 8h. Que peut-on dire du temps d'attente du premier arrivé ?

On peut retrouver cette situation (que nous appelons le problème de la rencontre) dans le document Ressources Probabilités et statistiques de la classe de terminale (MENJVA, 2012) mais sous forme d'exercice d'application ; ici l'énoncé est beaucoup plus ouvert et a pour objectif d'introduire la notion de fonction de densité. Dans un premier temps, les élèves ont une phase de recherche individuelle. Une première mise en commun permet alors d'engager des discussions dans la classe sur la vraisemblance ou non du problème, sur le caractère aléatoire de la situation, sur le caractère discret ou continu du temps… Des considérations qualitatives peuvent alors arriver, comme par exemple : le temps d'attente est entre 0 et 1 heure, il y a plus de « chance » que le premier arrivé attende 5 minutes que 55 minutes… Progressivement des variables aléatoires X_O et X_K représentant respectivement les heures arrivées d'Olivier et Karine sont introduites par les élèves et la variable aléatoire représentant le temps d'attente X est définie par
$X = |X_O - X_K|$. Ayant rencontré dans le devoir maison des simulations et ensuite une mise sous forme d'histogrammes, les élèves proposent assez naturellement (bien que cela puisse mettre du temps) de simuler des réalisations de la variable aléatoire X et ensuite de les représenter sous forme d'histogramme. Le logiciel GeoGebra permet d'obtenir par exemple l'histogramme de la figure 2.

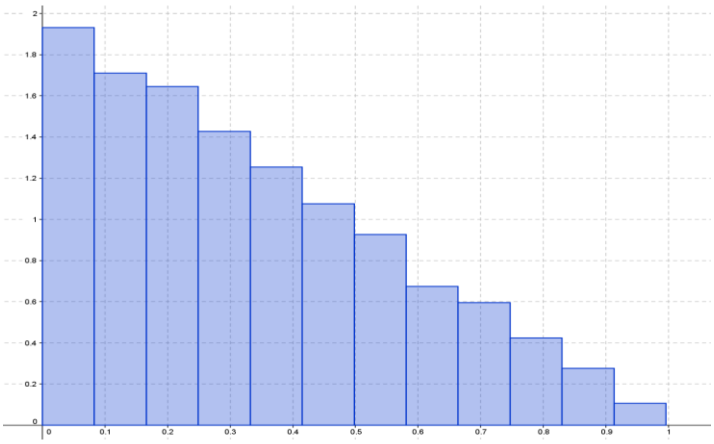


Figure 2. Histogramme d'un échantillon de taille 10 000, classe d'amplitude 5 minutes

Les élèves cherchent ensuite une « courbe de tendance » en prenant en compte l'allure de l'histogramme, courbe qui semble être un segment. Cependant, il faut une réflexion collective pour faire émerger le besoin d'imposer des contraintes pour choisir ce segment, notamment que l'aire sous la courbe fasse 1 sur $[0\,;1]$ et que $f(1) = 0$, ce qui permet d'arriver à l'expression de la fonction suivante :

$$f(x) = -2x + 2.$$

A partir de cette expression, il est possible de déterminer toutes les probabilités que l'on veut.

Cette présentation succincte du problème et de son déroulement permet de montrer un problème qui prend appui sur l'histogramme pour introduire une notion importante en terminale S pour véritablement comprendre les lois à densité, la fonction de densité. Nous souhaitons maintenant illustrer, à travers la synthèse faite en classe et des extraits de verbatim, les savoirs qui ressortent de cette situation et notamment les justifications données par les élèves sur les propriétés de la fonction de densité mettant en évidence l'importance des connaissances sur les histogrammes.

Lors de la séance suivant la recherche autour du problème de la rencontre, l'enseignante propose un temps d'institutionnalisation. La synthèse écrite au tableau, construite au fur et à mesure avec les élèves, est la suivante :

Synthèse

On note T_O *le moment d'arrivée d'Olivier*

T_K *le moment d'arrivée de Karine*

T_O *et* T_K *sont des variables aléatoires qui prennent toutes les valeurs réelles de* $[7\,;\,8]$ *que l'on peut ramener à* $[0\,;\,1]$.

On dit que ce sont des variables aléatoires ***continues*** *qui suivent le modèle de la puce.*

On note X *la variable aléatoire égale au temps d'attente du premier arrivé.*

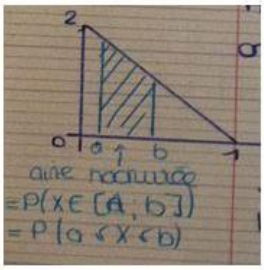
X *prend toutes les valeurs de* $[0\,;\,1]$.

Pour décrire la loi de probabilité de X, *on introduit une fonction* f *définie sur* $[0\,;\,1]$ *par* $f(x) = -2x + 2$ *telle que la courbe de* f *« lisse le haut des rectangles de l'histogramme ».*

Ainsi pour calculer la proba $P(X \in [a\,;\,b])$ *où* $[a\,;\,b] \subset [0\,;\,1]$, *on calcule l'aire sous la courbe* C_f *entre* a *et* b.

f *s'appelle la* ***fonction de densité*** *de probabilité de* X.

f *possède les propriétés suivantes :* f *est positive sur* $[0\,;\,1]$, *l'aire sous* C_f *entre* 0 *et* 1 *vaut* 1.



L'enseignante introduit elle-même le vocabulaire « fonction de densité de probabilité », cependant elle demande ensuite aux élèves s'ils peuvent justifier ce choix de vocabulaire. Voici la réponse d'un élève :

> E1 : C'est parce que sur l'axe des ordonnées, c'est la densité.

Pour justifier la positivité de la fonction de densité puis le fait que l'aire sous la courbe est égale à 1, on retrouve à nouveau une justification prenant appui sur l'histogramme

> E2 : Parce qu'en fait c'est par rapport à l'histogramme donc on peut pas avoir de rectangles en-dessous.

> E3 : Parce que ça correspond à la fréquence. L'aire correspond... […] L'aire correspond à la somme de toutes les probabilités... […] La somme des fréquences.
> P : Et donc on sait que ça fait...
> E3 : Un.

La construction de la nouvelle notion fonction de densité est ici possible grâce à la disponibilité des connaissances sur l'histogramme. Notamment, le fait que l'axe des ordonnées de l'histogramme soit véritablement défini et ait un sens pour les élèves permet de légitimer le vocabulaire « fonction de densité de probabilité », ce qui n'est jamais fait dans les manuels.

3 Conclusion

Nous espérons avoir montré l'intérêt et l'utilité de l'histogramme, si ce n'est pour lui-même comme moyen pour représenter des données statistiques (continues ou tout du moins regroupées), pour l'étude des lois à densité et notamment pour l'introduction de la fonction de densité. Ce chapitre qui est très conséquent en terminale (pas seulement en terminale S) peut dépasser la simple mémorisation de formule si un véritable travail d'introduction de la notion de fonction de densité est fait. Il semble donc important de former les enseignants sur l'histogramme pour que leur point de vue à son sujet change et qu’ils en perçoivent les enjeux notamment en probabilités.

Bibliographie

[1] Bressoud, E. & Kahané, J.-C. (2010), *Statistique descriptive*, 2è édition. Pearson France.
[2] Derouet, C. (2016). La fonction de densité au carrefour entre probabilités et analyse. Etude de la conception et de la mise en œuvre de tâches d'introduction articulant lois à densité et calcul intégral, Thèse de doctorat, Université Paris Diderot.
[3] Derouet, C., & Parzysz, B. (2016). How can histograms be useful for introducing continuous probability distributions?, *ZDM – Mathematics Education*, 48(6), 757-774.
[4] Ministère de l'Education nationale (MEN) & Direction de l'Enseignement Scolaire (2001). *Mathématiques. Classe de première des séries générales*. Paris : Centre national de documentation pédagogique.
[5] Ministère de l’Education nationale de la Jeunesse et de la Vie associative (MENJVA), & DGESCO. (2012). *Ressources pour la classe terminale générale et technologique. Probabilités et statistique.*
[6] Roditi, E. (2009). L’histogramme : à la recherche du savoir à enseigner. *Spirale Revue de recherches en éducation, 43*, 129–138.