*Cette fiche a été élaborée par des enseignantes et des enseignants des lycées et universités de l'académie de Créteil.*

**Thème** Lois de probabilité à densité

**Titre** Se familiariser avec les fonctions de densité

**Présentation**

Voici une sélection d'exercices issues du site nrich.maths.org.
Ce site anglo-saxon créé à l'initiative de l'Université de Cambridge met à disposition des élèves et des professeurs une grande variété d'exercices niveau collège à post-bac, concernant toutes les branches des mathématiques.

Pour chaque exercice, on trouve sur le site internet :

- l'énoncé (en anglais) ;
- des commentaires pour le professeur (pourquoi traiter cet exercice ? comment ?) ;
- des indices pour la résolution ;
- le plus souvent une solution.

Sont présentés dans ce document trois exercices extraits du site concernant les fonctions densité (« probability density function », abrégé pdf) des lois de probabilités à densité.
Plutôt que des calculs de probabilités, l'objectif est ici plutôt de manipuler le concept de densité pour se l'approprier tout en réinvestissant des compétences classiques d'analyse.

**Mise en place (suggestion)**

Durée : 1h30

*Première partie :*
Les élèves préparent un exercice par groupes de 3 ou 4 pendant 30 minutes environ.
Le professeur laisse chercher les élèves seuls pendant 10 minutes puis guide éventuellement les élèves si nécessaire.
Production : Chaque groupe rédige sur feuille une solution pour son exercice.

*Deuxième partie :*
Rappel des propriétés d'une fonction densité par le professeur, des fonctions densités vues dans le cours.
Compte-rendu de chaque exercice par une équipe.
Les feuilles solutions sont ramassées, photocopiées et distribuées aux élèves au cours suivant.

**Exemple 1 Espérance, Médiane et Mode d'une variable aléatoire définie par une fonction de densité.**

## Compétences travaillées :

- ☐ Savoir calculer l'espérance d'une loi à densité.
- ☐ Utiliser ses connaissances en statistiques pour imaginer/étendre les concepts de médiane et de mode dans le cas d'une loi de probabilité à densité.
- ☐ Calcul d'intégrales simples.
- ☐ Variations et extremum d'une fonction polynôme du second degré.

## Source : « PDF »

*Préliminaire : Rappeler les définitions de médiane et de mode dans le cadre d'une série statistique.*

Une variable aléatoire $X$ est définie par sa fonction densité $\rho$ définie sur l'intervalle [0; 3] par :

$$\rho(x)=\frac{2}{27}\left(6+x-x^2\right)$$

1) Calculer l'espérance de $X$.
2) Proposer une définition de la « médiane » de $X$, puis déterminer cette valeur. On pourra s'aider pour cette question d'un outil de calcul formel.
3) Proposer une définition du « mode » de $X$, puis déterminer cette valeur.

## Eléments de correction :

1) $E(X)=\int_0^3 x\rho(x)dx=\frac{2}{27}\left[3x^2+\frac{1}{3}x^3-\frac{1}{4}x^4\right]=\frac{7}{6}$

2) La médiane de $X$ est la valeur $m$ telle que $P(X\le m)=P(X\ge m)$, ou encore telle que $P(X\le m)=0,5$.

   Après calcul de l'intégrale on obtient l'équation : $4m^3-6m^2-72m+81=0$

   Cette équation admet trois solutions réelles $m_1=\frac{9}{2}$, $m_2=\frac{3}{2}\left(-1-\sqrt{3}\right)$ et $m_3=\frac{3}{2}\left(-1+\sqrt{3}\right)$.

   La seule solution dans l'intervalle [0 ;3] est : $m_3=\frac{3}{2}\left(-1+\sqrt{3}\right)\approx 1,098$.

3) Le mode d'une série discrète est la valeur de la série qui apparaît le plus souvent.
   On peut penser que pour une loi de probabilités continue le mode correspond au maximum de la densité.

   $$\rho'(x)=\frac{2}{27}(1-2x)$$

   $\rho$ admet un maximum pour $x$ = 0,5 : le mode de $X$ est 0,5.

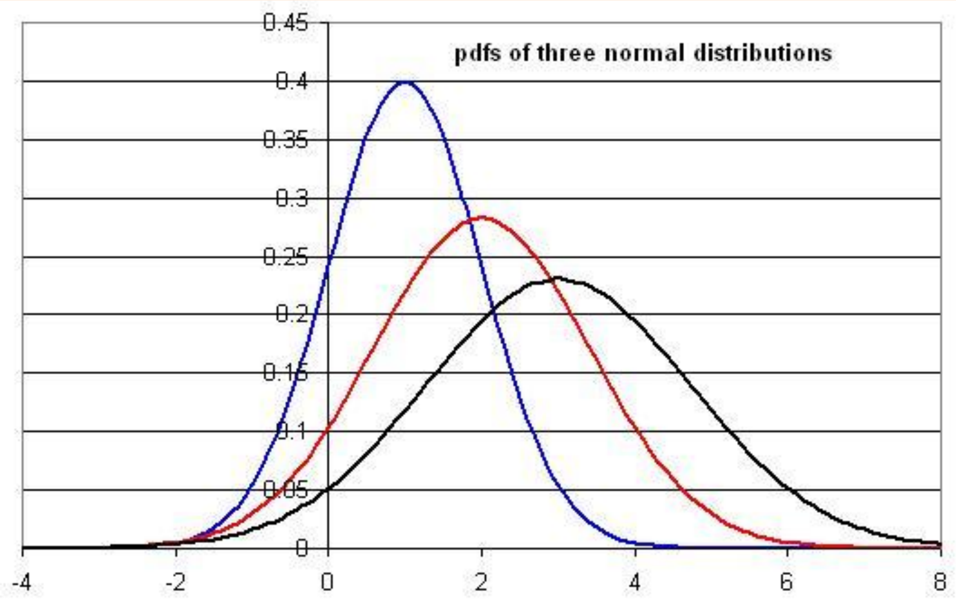
| | |
|---|---|
| **Exemple 2** | **Autour de la densité d'une loi normale** |

## Compétences travaillées :

- ☐ Connaître et utiliser la densité d'une loi normale.
- ☐ Mettre en place une approximation simple d'aire (par un triangle).
- ☐ Utilisation des cadres graphique et algébriques pour déterminer un paramètre inconnu .

**Source :** « Into the Normal Distribution »

Le graphique ci-dessous représente les densités $f$, $g$ et $h$ de trois variables aléatoires $X$, $Y$, $Z$ (respectivement) qui suivent chacune une loi normale inconnue. On sait de plus que :
- le maximum de $f$ est atteint en $x = 1$ et vaut environ 0,4,
- le maximum de $g$ est atteint en $x = 2$ et vaut environ 0,28,
- et le maximum de $h$ est atteint en $x = 3$ et vaut environ 0,23.




1) D'après le graphique on peut obtenir l'estimation $P(X < 0) \approx 0.25$.
En utilisant l'aire d'un triangle bien choisi, retrouver ce résultat.
Donner une estimation similaire pour $P(Y < 0)$ et $P(Z < 0)$.

2) a) Rappeler les propriétés de la courbe d'une fonction densité.

b) Quelle autre propriété remarquable possède la courbe d'une fonction densité dans le cas d'une loi normale ?

c) On admet dans cette question que l'expression de la densité d'une loi $(\mu, \sigma^2)$ est :

$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ (cette formule n'est *pas* à connaître pour le baccalauréat).

On sait que les espérances et variances de chacune de ces lois normales sont des nombres entiers.
Déterminer l'espérance et la variance de $X$, de $Y$ et de $Z$.

4) En déduire une valeur approchée à $10^{-3}$ près de $P(X < 0)$, $P(Y < 0)$ et $P(Z < 0)$.
Comparer avec les résultats obtenus à la question 1).

## Eléments de correction :

1) On peut approximer ces aires par des aires de triangles :
$P(X < 0) \approx 1/2 \times 2 \times 0{,}25 = 0{,}25$
$P(Y < 0) \approx 1/2 \times 2 \times 0{,}1 = 0{,}1$
$P(Z < 0) \approx 1/2 \times 2 \times 0{,}05 = 0{,}05$

2) La courbe d'une fonction densité suivant une loi normale ($\mu$, $\sigma^2$) est symétrique par rapport à la droite d'équation $x = \mu$. Par lecture graphique et sachant (d'après l'énoncé) que les espérances sont des entiers on en déduit que $\mu_X = 1$, $\mu_Y = 2$ et $\mu_Z = 3$.

   On évalue ensuite $f(x) = \dfrac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X{}^2}}$ pour $x = \mu_X$ et on résout en $\sigma_X$ pour trouver $\sigma_X{}^2 = 1$.

   De même on trouve $\sigma_Y{}^2 = 2$ et $\sigma_Z{}^2 = 3$.

3) On sait maintenant que $X \sim (1\ ;\ 1)$, $Y \sim (2\ ;\ 2)$ et $Z \sim (3\ ;\ 3)$.
   On utilise le tableur ou la calculatrice pour trouver :

   $P(X < 0) \approx 0{,}159$
   $P(Y < 0) \approx 0{,}079$
   $P(Z < 0) \approx 0{,}042$

   Comme attendu graphiquement, l'erreur commise en approximant par un triangle est la plus faible avec la courbe de *h*.

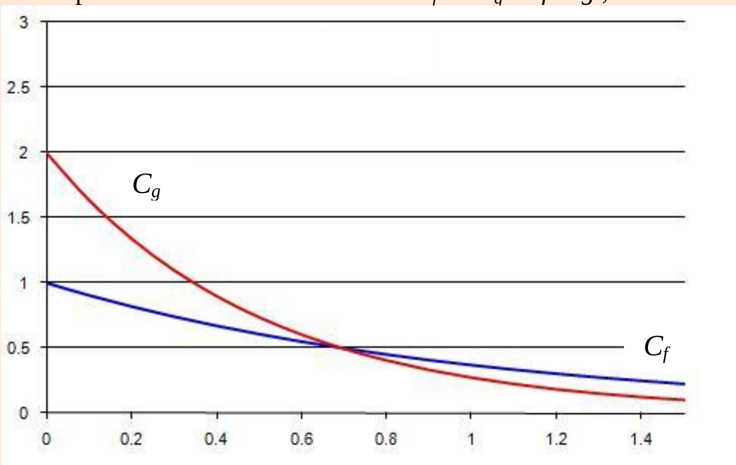
## Exemple 3 Autour de la densité d'une loi exponentielle

### Compétences travaillées :

- ☐ Connaître et utiliser la densité d'une loi exponentielle.
- ☐ Déterminer le paramètre d'une loi exponentielle à partir du graphique et de l'expression algébrique de la densité.
- ☐ Connaître l'espérance d'une loi exponentielle.
- ☐ Utiliser en parallèles les cadres graphique et algébrique.
- ☐ Mettre en place une approximation d'aire simple (par un trapèze).
- ☐ Calculer l'aire entre deux courbes à l'aide d'une intégrale.
- ☐ Déterminer les coordonnées du point d'intersection entre deux fonctions densités : résoudre une équation avec la fonction exponentielle.

### Source : « Into the Exponential Distribution »

Soient $X$ et $Y$ deux variables aléatoires suivant une loi exponentielle.
On a représenté ci-dessous les courbes $C_f$ et $C_g$ de $f$ et $g$ , les densités de $X$ et $Y$ respectivement.




**1)** A l'aide de vos connaissances et du graphique, donner l'expression algébrique de chacune de ces densités. Laquelle des deux variables aléatoires admet l'espérance la plus grande ?

**2)** En utilisant le graphique on peut obtenir l'estimation $P\,(0,5 < Y < 0,7) \approx 0,125$.
Comment a-t-on obtenu cette estimation ?
Comparer avec la valeur de $P\,(0,5 < Y < 0,7)$ obtenue par le calcul.

*Aide : Quelle est la formule donnant l'aire d'un trapèze ?*

**3)** On note α l'abscisse du point d'intersection des courbes $C_f$ et $C_g$.
Interpréter graphiquement les intégrales suivantes : $\int_0^{\alpha} g(x) - f(x)dx$ et $\int_{\alpha}^{+\infty} f(x) - g(x)dx$.
Déterminer α et calculer ces deux intégrales.

### Eléments de correction :

1) $X$ suit une loi exponentielle donc $f(x) = \lambda e^{-\lambda x}$ . Par lecture graphique on a $f(0) = 1$, d'où $\lambda = 1$ et $f(x) = e^{-x}$. De même on trouve $g(x) = 2e^{-2x}$.

2) En utilisant l'aire d'un trapèze on obtient : $P(0,5 < Y < 0,7) \approx 0,2 \times \left(\dfrac{g(0,5)+g(0,7)}{2}\right) \approx 0,2 \times \left(\dfrac{0,75+0,5}{2}\right)$,
d'où $P\,(0,5 < Y < 0,7) \approx 0,125$. D'autre part : $P(0,5 < Y < 0,7) = \int_{0,5}^{0,7} 2e^{-2x}dx = \left[-e^{-2x}\right]_{0,5}^{0,7} = e^{-1} - e^{-1,4} \approx 0,121$

3) La résolution de $f(x) = g(x)$ aboutit à α = ln2 (une factorisation est nécessaire). Les aires entre les deux courbes sur l'intervalle [0 ; α] et sur l'intervalle [α ; + ∞ [ sont égales et valent chacune 0,25 (unités d'aires).

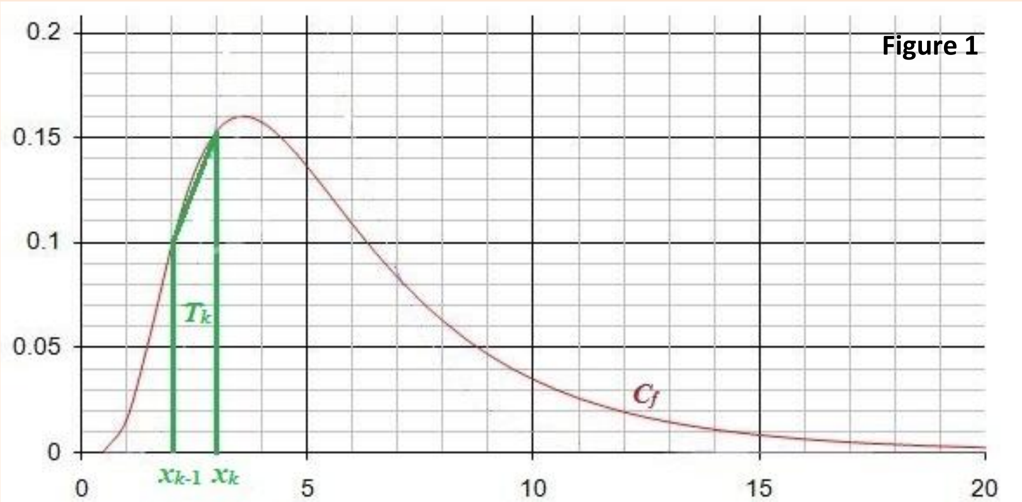
## Exemple 4 Approximation de paramètres d'une loi inconnue à l'aide de la méthode des trapèzes

Compétences travaillées :

- ☐ Mettre en place une approximation d'aire par la méthode des trapèzes pour déterminer la médiane d'une loi continue à densité dont on connaît seulement le graphique et pas l'expression algébrique.
- ☐ Utilisation du tableur.

Source : « What's your mean? »

Le graphique de la Figure 1 représente la courbe d'une fonction de densité $f$ dont on ignore l'expression algébrique.
On note $X$ la variable aléatoire et P la loi de probabilité associées.
On se propose dans cet exercice de déterminer une valeur approchée de la médiane de X (c'est-à-dire le réel $m$ tel que $P(X \leq m) = 0{,}5$) en mettant en place la méthode d'intégration numérique « des trapèzes ».




Pour $k$ entier quelconque compris entre 1 et 20, on considère le trapèze $T_k$ (on a dessiné $T_k$ pour $k = 3$ sur la Figure 1), où $x_k = k$.

a) Donner l'aire de $T_k$ en fonction de $x_{k-1}$ et $x_k$.
*Rappel :* L'aire d'un trapèze est
$$A(trapèze) = \frac{(petite\ base + grande\ base) \times hauteur}{2}$$

b) Recopier dans une page de calcul les valeurs du tableau de la Figure 2.
Quelle est la formule à écrire (à copier vers le bas) dans la cellule C3 ?
Compléter les cellules C4 à C20 en glissant-copiant cette formule.

c) Sans faire le calcul pour l'instant, qu'attend-t-on comme résultat pour la somme des aires des trapèzes $T_k$, pour $k$ allant de 1 à 20 ?
Effectuer ce calcul dans la cellule C24. Commenter.

d) Justifier que pour tout $k$ entier entre 1 et 20 on a :
$P(X \leq x_k) \approx P(X \leq x_{k-1}) + Aire(T_k)$.
Utiliser ce résultat pour compléter la colonne D.

e) Conclure pour la question posée.

| | A | B | C | D |
|---|---|---|---|---|
| 1 | $x_k$ | $f(x_k)$ | Aire du trapèze $T_k$ | Approximation de $P(X \leq x_k)$ |
| 2 | 0 | 0 | | |
| 3 | 1 | 0,015 | | |
| 4 | 2 | 0,1 | | |
| 5 | 3 | 0,15 | | |
| 6 | 4 | 0,155 | | |
| 7 | 5 | 0,135 | | |
| 8 | 6 | 0,11 | | |
| 9 | 7 | 0,085 | | |
| 10 | 8 | 0,06 | | |
| 11 | 9 | 0,045 | | |
| 12 | 10 | 0,035 | | |
| 13 | 11 | 0,025 | | |
| 14 | 12 | 0,02 | | |
| 15 | 13 | 0,015 | | |
| 16 | 14 | 0,01 | | |
| 17 | 15 | 0,009 | | |
| 18 | 16 | 0,006 | | |
| 19 | 17 | 0,005 | | |
| 20 | 18 | 0,004 | | |
| 21 | 19 | 0,002 | | |
| 22 | 20 | 0,001 | | |
| 23 | | | | |
| 24 | | | | |
| 25 | | | | |

Figure 2

## Approfondissement

Proposer une méthode permettant de déterminer une valeur approchée de l'espérance de *X*.
En complétant le fichier tableur précédent donner la valeur obtenue.

*Aide:*

*Justifier et utiliser l'approximation suivante :* $E(X) \approx \sum_{k=1}^{n} m_k P(x_k \leq X \leq x_{k+1})$ *où* $m_k = \frac{x_k + x_{k+1}}{2}$

## Remarque

Pour les élèves les plus à l'aise on peut proposer l'exercice en problème ouvert.

**Pour approfondir… Petite sélection de problèmes ouverts  (sur le même thème)**

De nombreux exercices sur le site sont des problèmes ouverts ou en tous cas des problèmes qui laissent une grande place à l'initiative de l'élève. Voici deux exemples.

« Circle pdf »
Parmi un certain nombre de courbes en forme d'arcs de cercle, lesquelles pourraient représenter une densité ?

« Normal Intersection »
Peut-on trouver les paramètres de deux lois normales de façon à ce que les courbes des densités se coupent en un point ? deux points ? aucun point ?
Même question ensuite avec les courbes des fonctions de répartition.