Construction de la fonction de densité

Licence Sciences pour la Santé

Intervenant : Mathieu Fauvernier

Rappels sur les probabilités et question

Soit $\Omega = \{pile, face\}$ l'ensemble des possibles d'une expérience aléatoire

On peut définir $P(pile)$ et $P(face)$ comme les probabilités d'obtenir pile et face respectivement (classiquement $P(pile)$=$P(face)$=0,5)

Qu'en est-il si $\Omega = \mathbb{R}$ ou un intervalle de $\mathbb{R}$?

La suite de ce cours s'appuie sur des exercices écrits par Charlotte Derouet :

$La\ fonction\ de\ densité\ au\ carrefour\ entre\ probabilités\ et\ analyse\ en\ terminale\ S.\ Etude\ de\ la\ conception\ et\ de\ la\ mise\ en\ oeuvre\ de\ tâches\ d'introduction\ articulant\ lois\ à\ densité\ et\ calcul\ intégral.\ Histoire\ et\ perspectives\ sur\ les\ mathématiques\ [math.HO].\ Université\ Paris\ Diderot (Paris\ 7)\ Sorbonne\ Paris\ Cité, 2016.\ Français.\ NNT\ :.\ tel - 01431913$

Problème de la rencontre

Karine et Oliver décident d’un rendez-vous entre 7h et 8h. Ils peuvent arriver n’importe quand entre 7h et 8h.

Que peut-on dire du temps d’attente du premier arrivé ?

Problème de la rencontre - modélisation

Imaginons que Karine arrive en premier et notons T_K le temps d'arrivé de Karine et T_O le temps d'arrivé d'Oliver.

Le temps d'attente est donné par $X = T_O - T_K$

Problème de la rencontre - modélisation

Comme Oliver peut également arrivé premier, on comprend qu’en toute généralité, on doit considérer

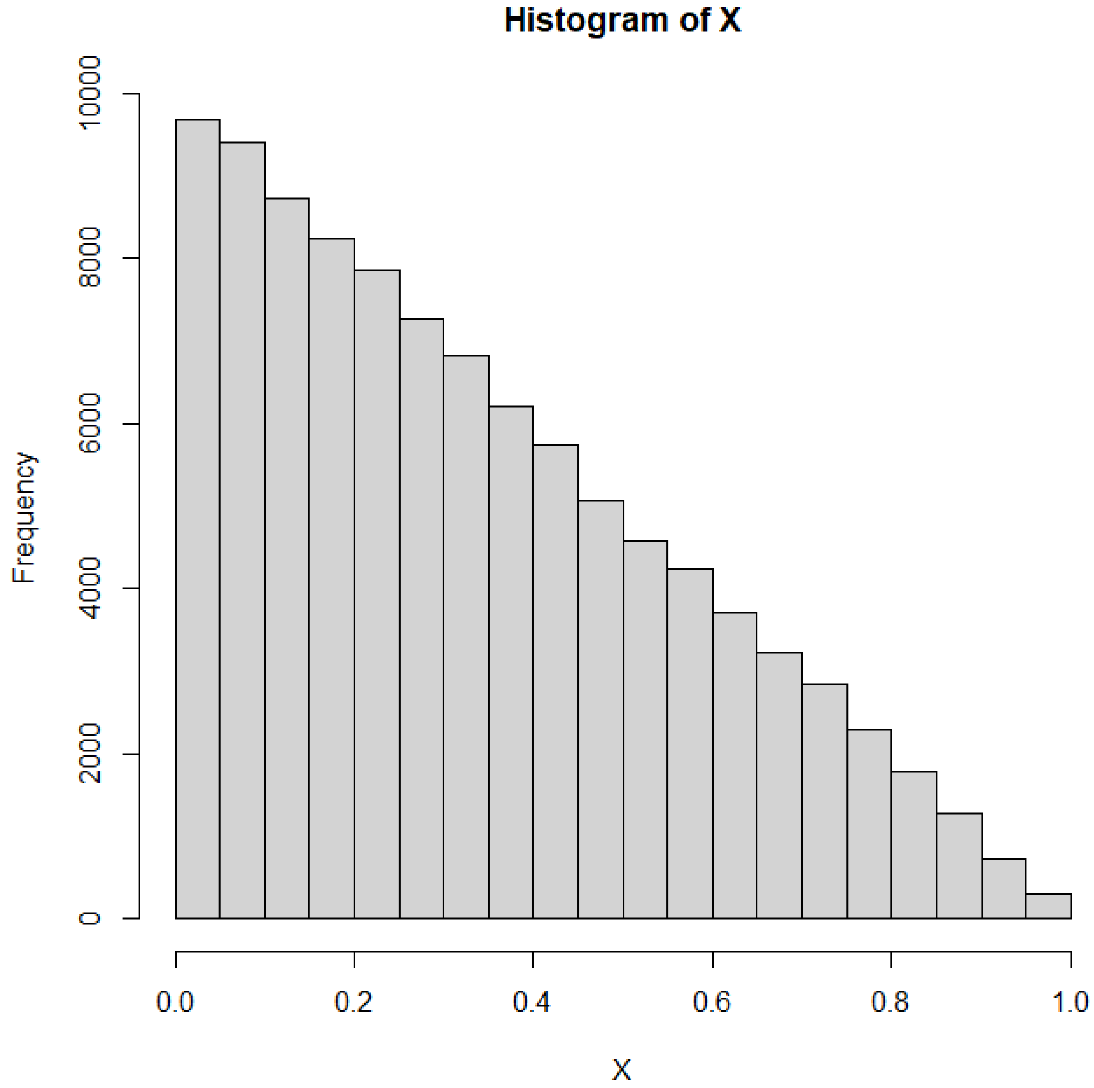
$$X = |T_O - T_K|$$

T_O et T_K sont des variables aléatoires réparties uniformément entre 7 et 8. Quant à X, il est forcément compris entre 0 et 1.

Reformulation de la question

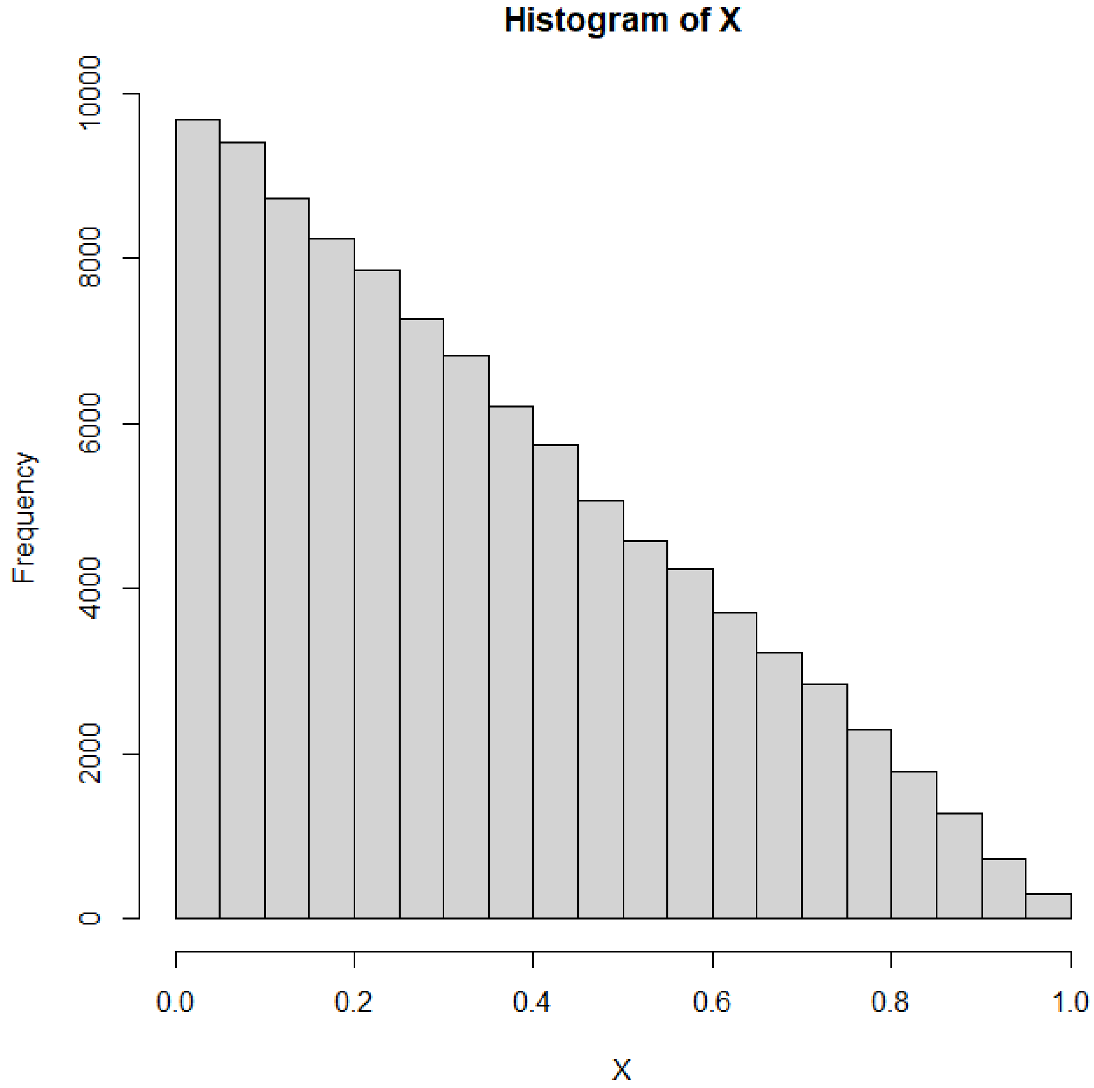
Que vaut $P(X \in [a\,; b])$? avec $[a\,; b] \subset [0; 1]$

Question intermédiaire n°1 : Quelle est la distribution de X ?

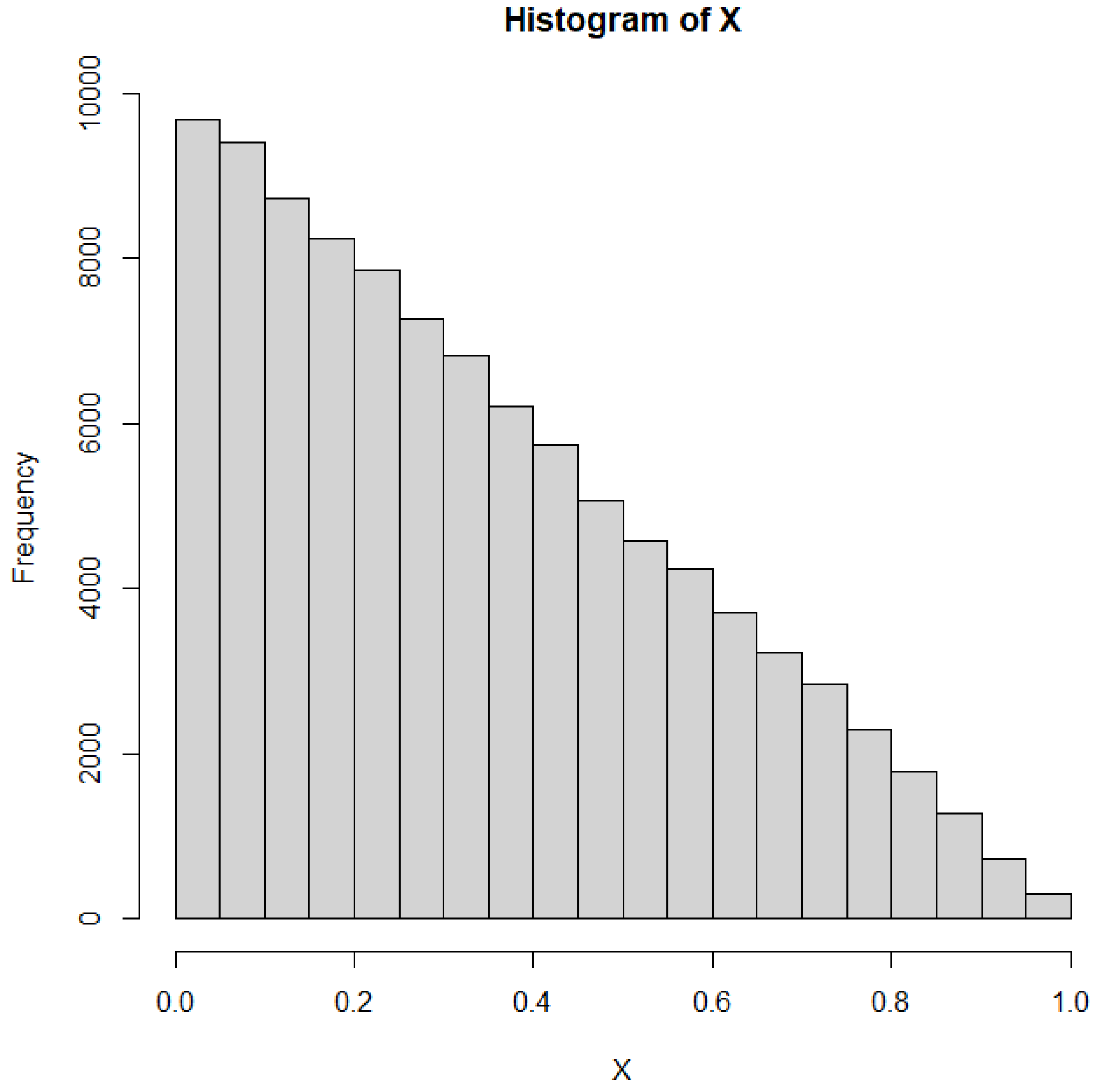


On simule n=10 000 valeurs de X et on trace l'histogramme associé

Question intermédiaire n°2 : Comment calculer la probabilité voulue à partir de l'histogramme ?

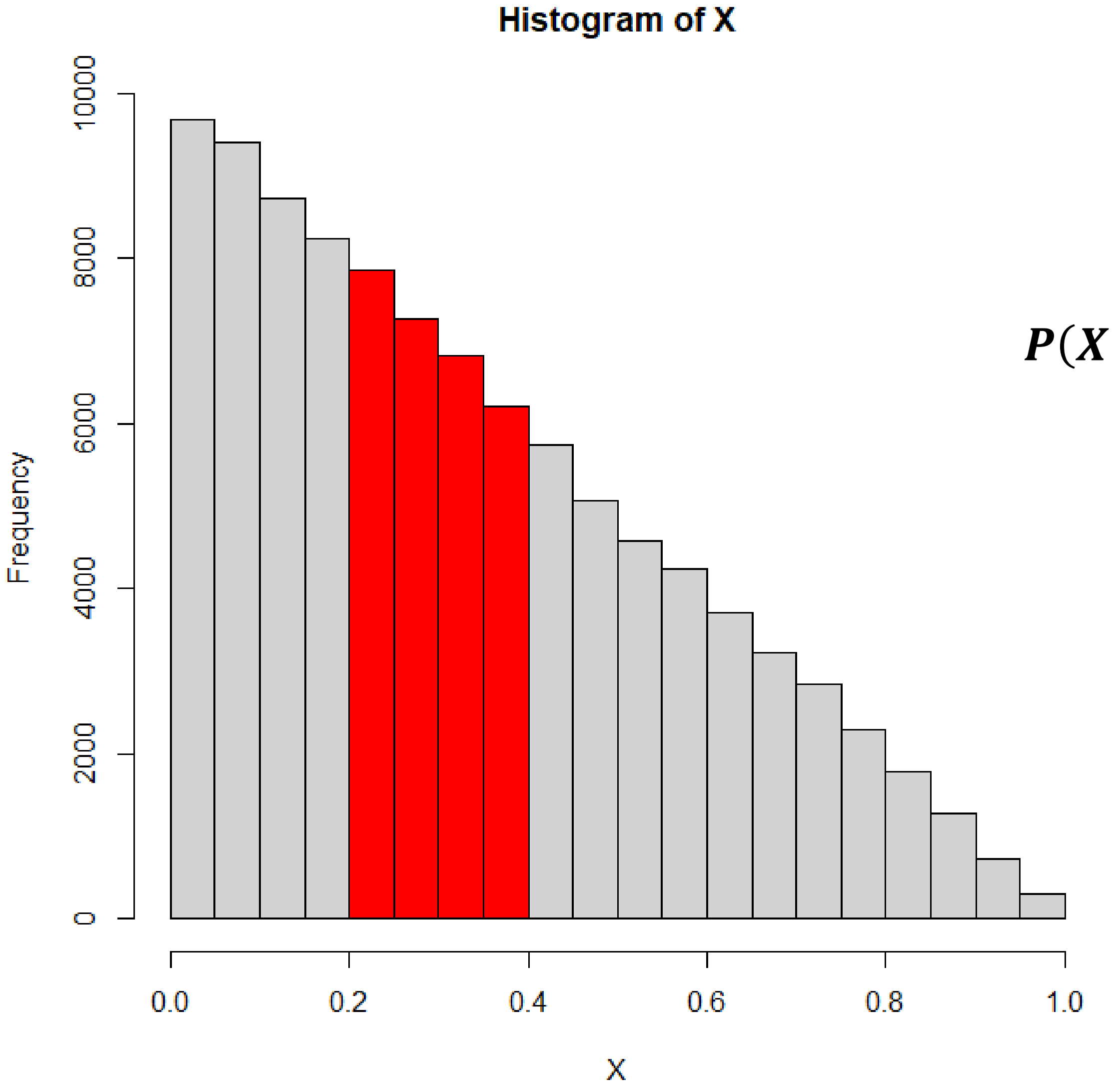


Question intermédiaire n°2 : Comment calculer la probabilité voulue à partir de l'histogramme ?



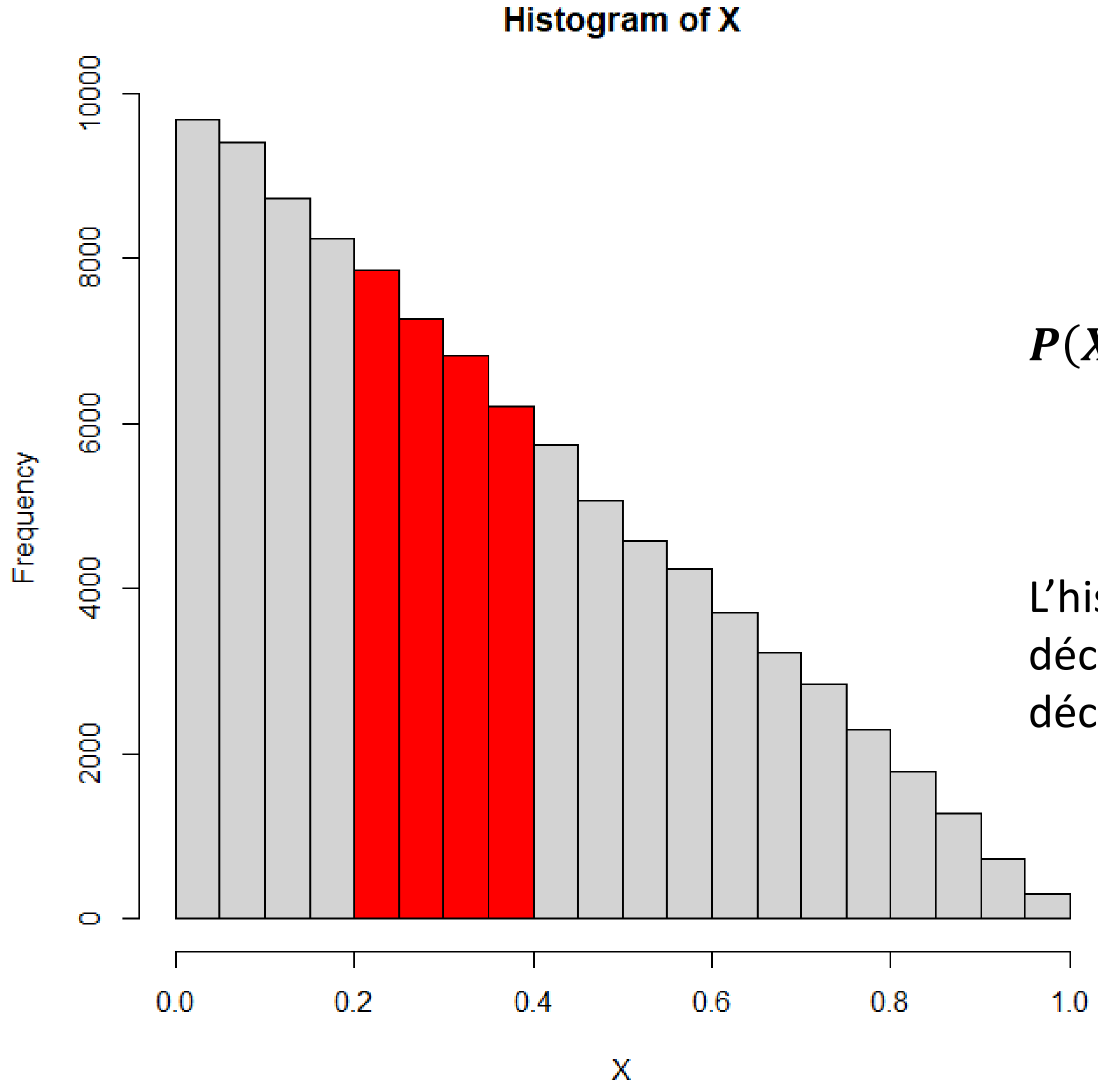
Réponse : somme des aires des rectangles !

Question intermédiaire n°2 : Comment calculer la probabilité voulue à partir de l'histogramme ?



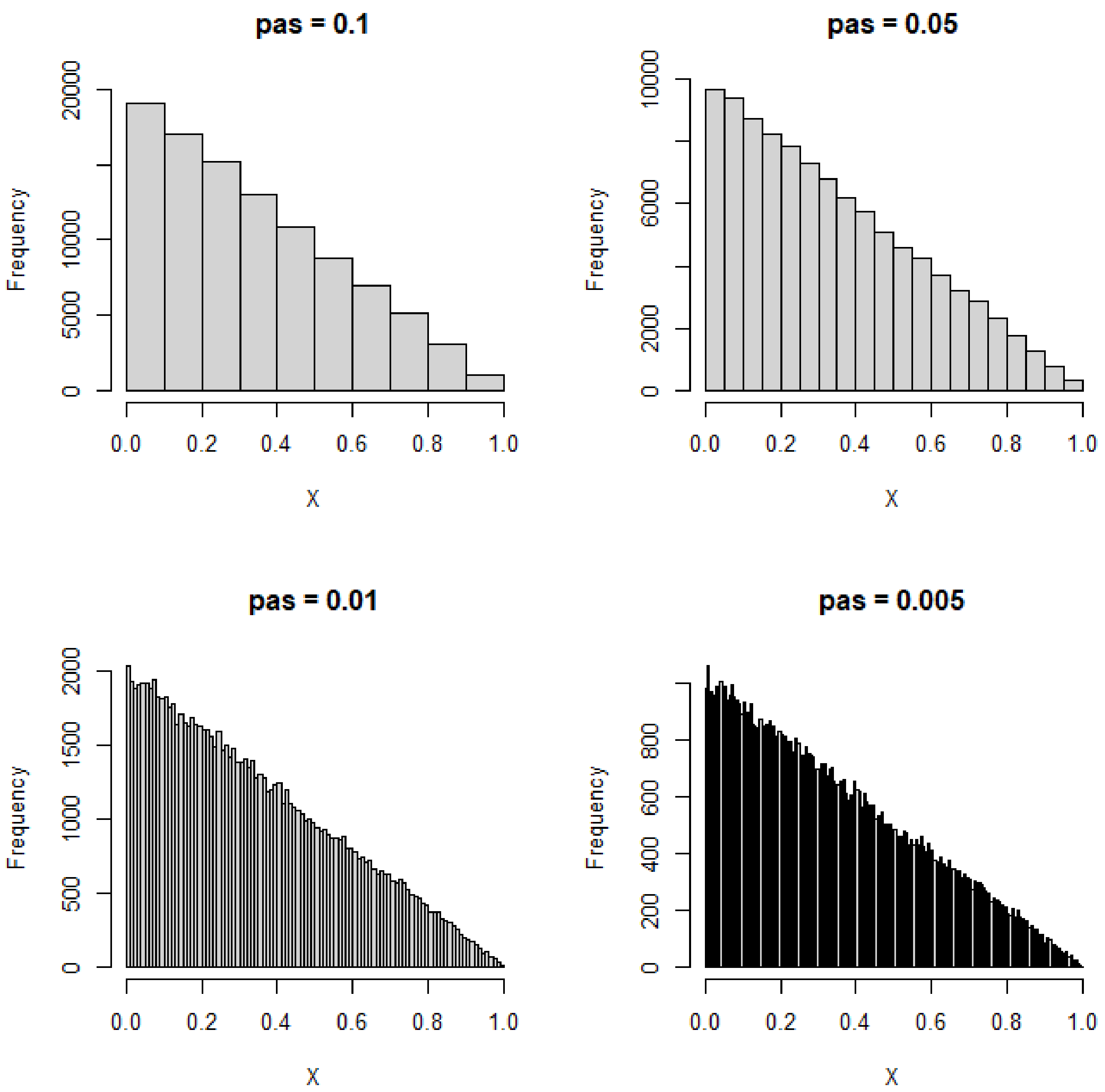
$P(X \in [0,2\,;0,4])$ = zone en rouge

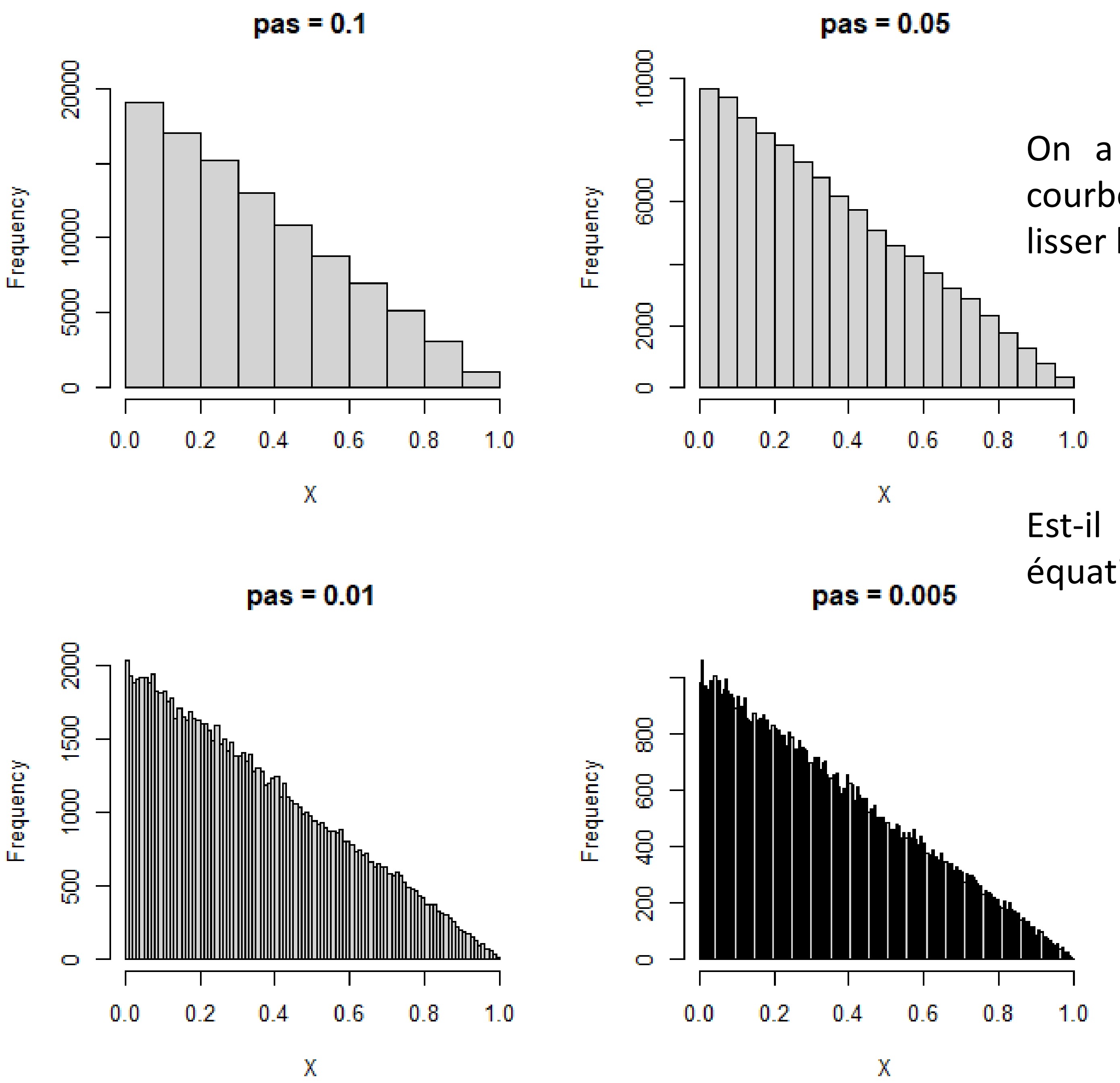
Question intermédiaire n°2 : Comment calculer la probabilité voulue à partir de l'histogramme ?



$\boldsymbol{P(X \in [0,2\,;0,4])}$ **= zone en rouge**

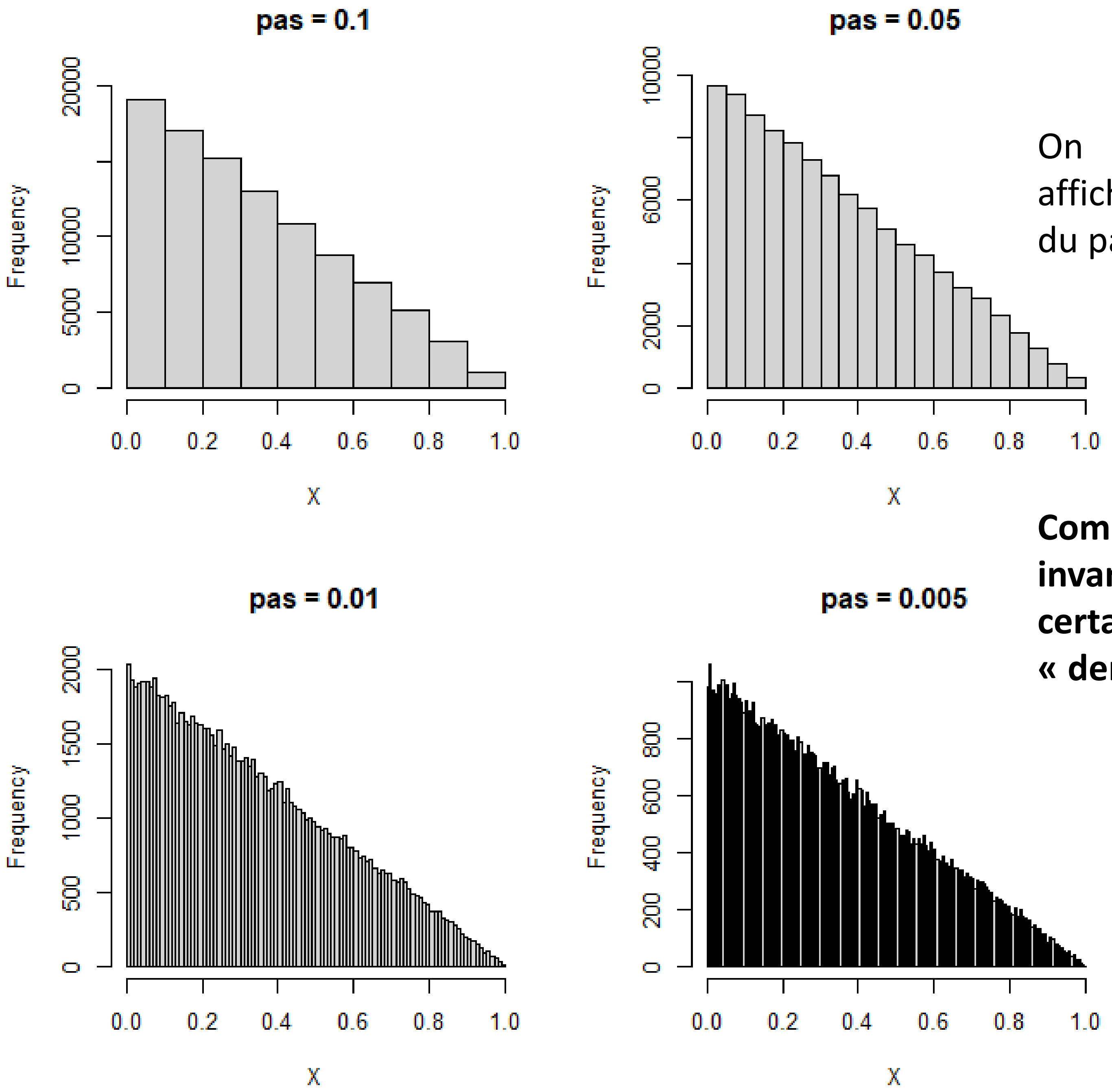
L'histogramme ci-contre utilise un découpage de pas 0,05, mais d'autres découpages sont possibles


pas = 0.1
pas = 0.05
pas = 0.01
pas = 0.005
Frequency
X
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
20000
2000
6000
500
1000
1500
200
400
600
800



On a l'impression qu'il existe une courbe théorique qui permet de lisser l'histogramme

Est-il possible de trouver son équation ?



On remarque que les fréquences affichées en ordonnées dépendent du pas, ce qui est un peu gênant.

Comment obtenir une mesure invariante pour dire à quel point certaines zones de valeurs sont « denses » en individus ?

Concept de densité

Pour rappel on a simulé n = 100 000 réalisations de $X = |T_O - T_K|$

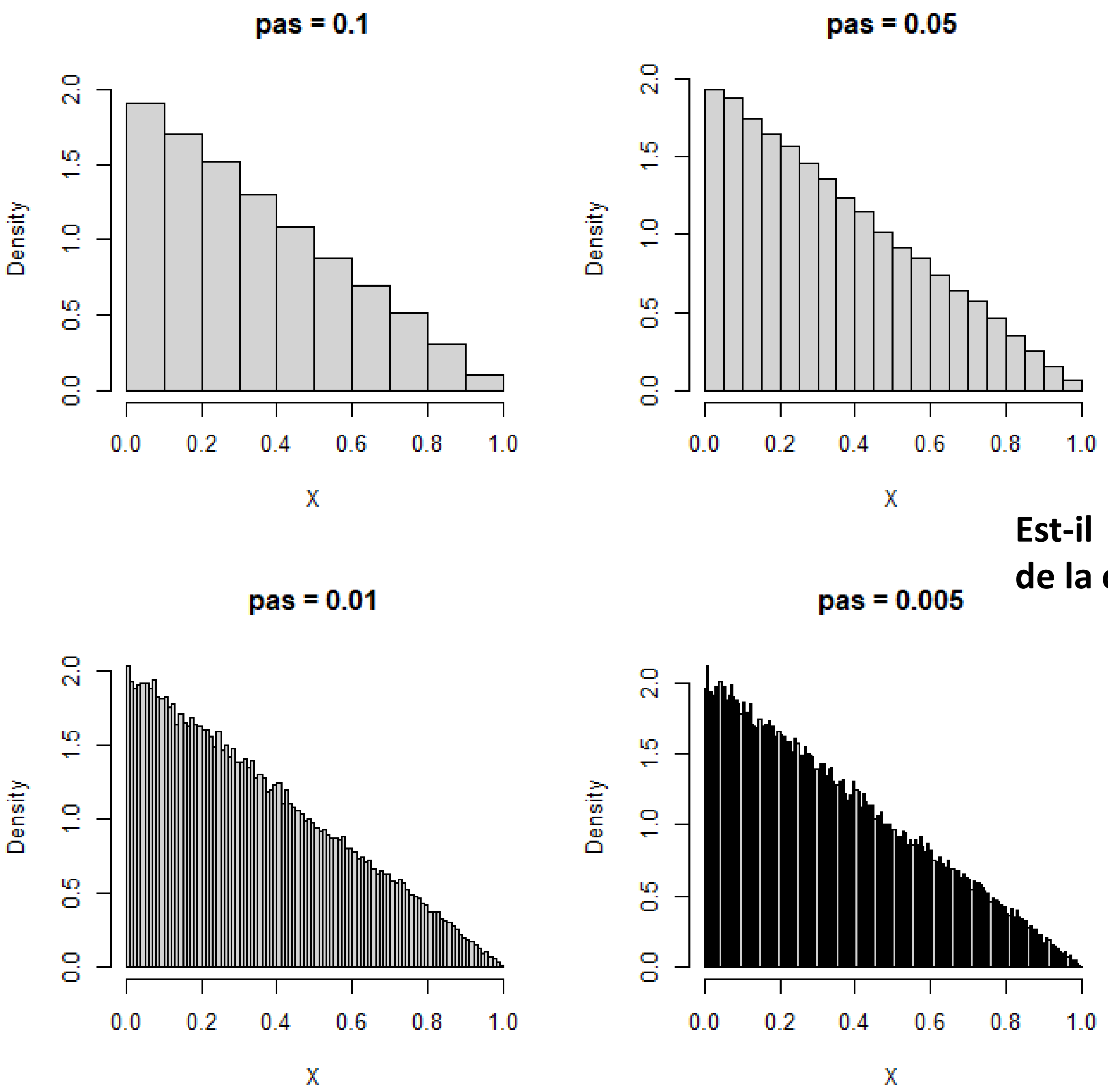
Pour une valeur de pas donnée, on découpe l'intervalle [0;1] en classes de longueur le pas en question.

Pour chaque classe, on peut calculer

- Le nombre d'individus -> **fréquence**
- Le nombre d'individus rapporté au total -> **fréquence relative**
- La fréquence relative rapportée au pas -> **densité**

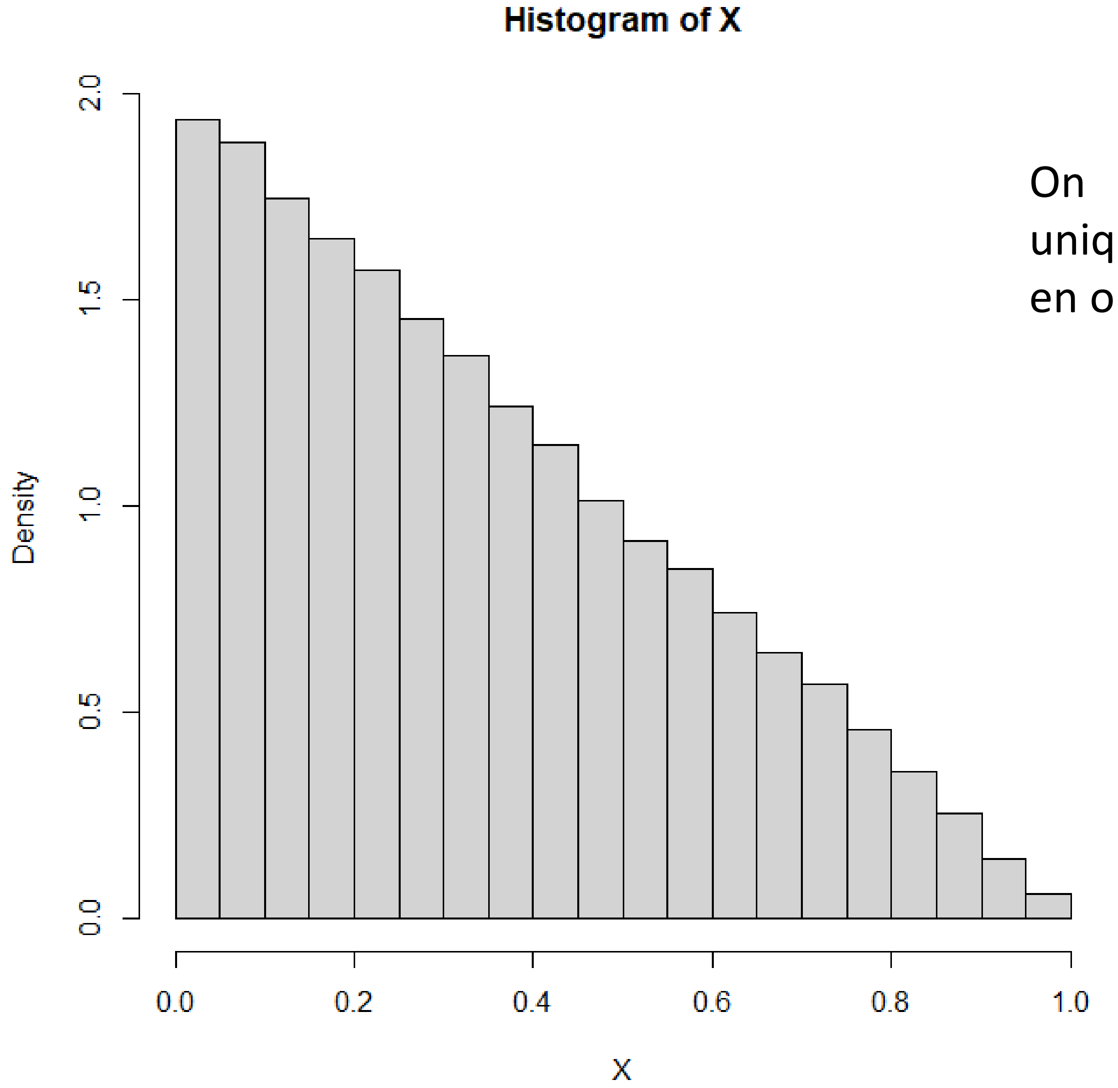
Concept de densité

Pas	Fréquence max	n	Fréquence relative max	Densité
0,1	20 000	100 000	20%	20%/0,1=**2**
0,05	10 000	100 000	10%	10%/0,05=**2**
0,01	2 000	100 000	2%	2%/0,01=**2**
0,005	1 000	100 000	1%	2%/0,005=**2**



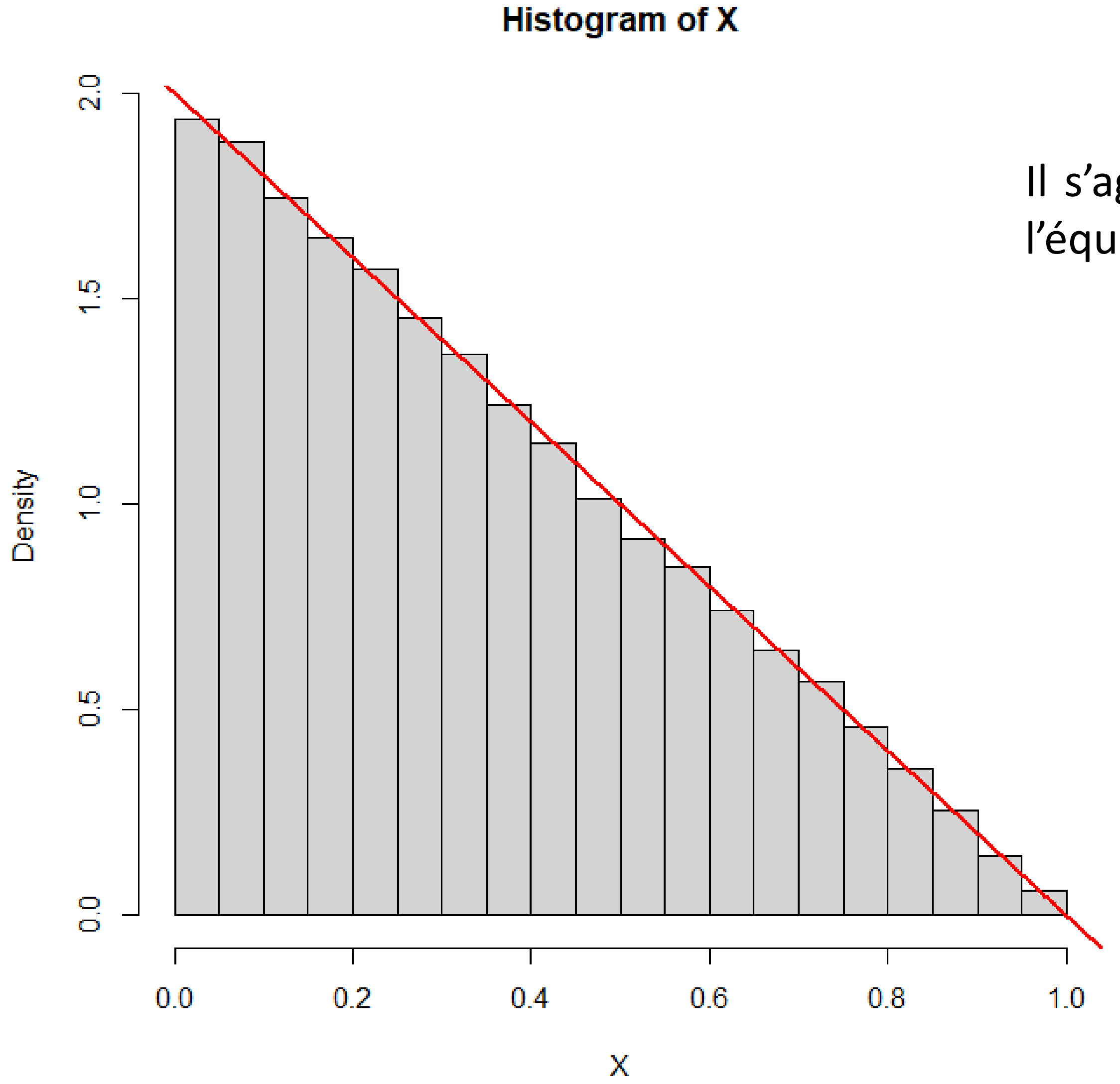
Est-il possible de trouver l'équation de la courbe qui lisse l'histogramme ?

Question intermédiaire n°3 : quelle est l'équation de la courbe qui lisse l'histogramme ?



On revient avec un découpage unique de 0,05, mais avec la densité en ordonnées

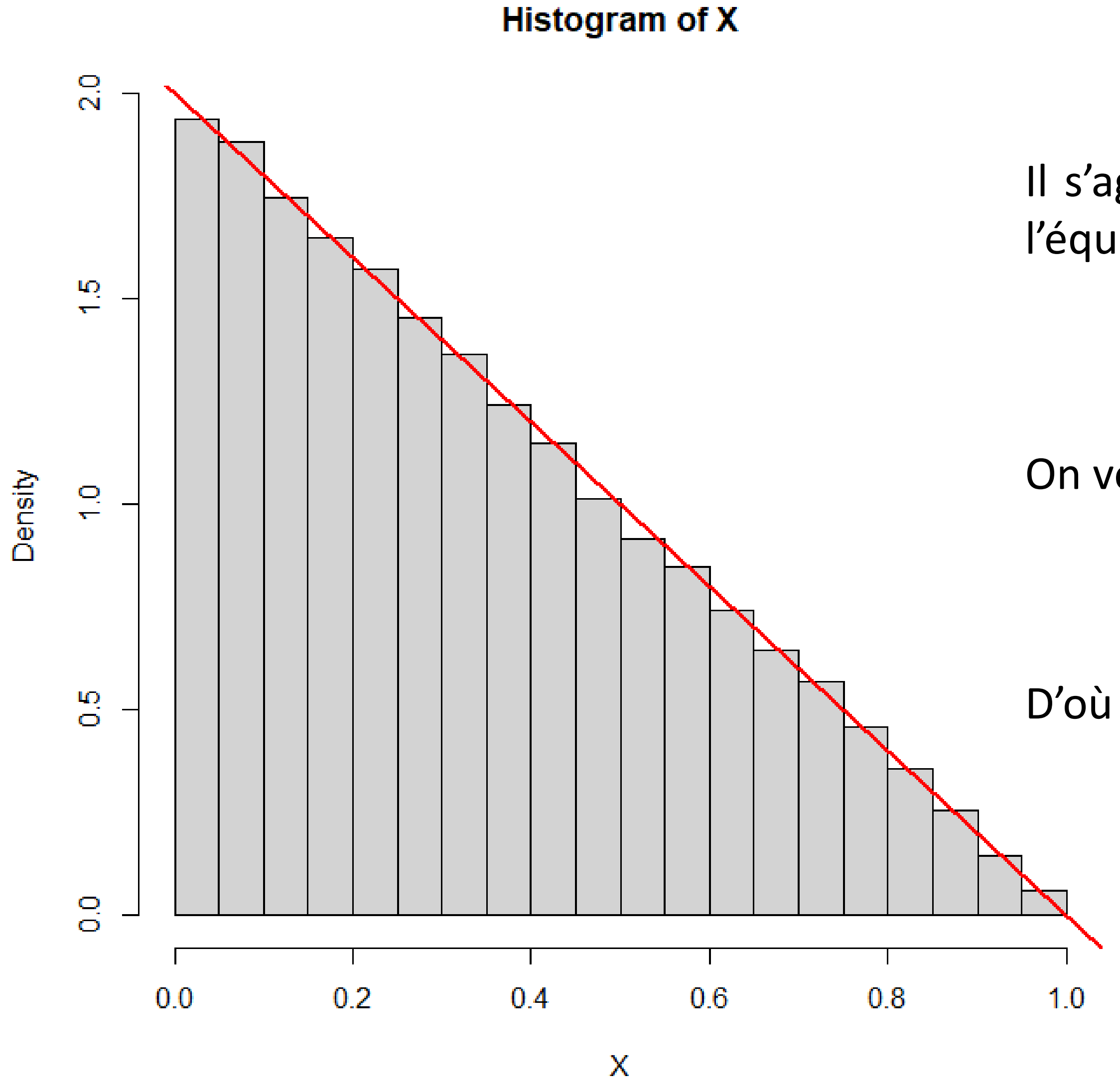
Question intermédiaire n°3 : quelle est l'équation de la courbe qui lisse l'histogramme ?



Il s'agit clairement d'une droite dont l'équation s'écrit

$$f(x) = ax + b$$

Question intermédiaire n°3 : quelle est l'équation de la courbe qui lisse l'histogramme ?



Il s'agit clairement d'une droite dont l'équation s'écrit

$$f(x) = ax + b$$

On voit que

$$f(0) = 2$$
$$f(1) = 0$$

D'où

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

Question finale : quel calcul permet de répondre à la question posée ?

La fonction f telle que

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

correspond à la densité de probabilité de $X = |T_O - T_K|$

La probabilité que $X \in [a\,;b]$ est donnée par l'aire sous la courbe représentative de f. Autrement dit

$$P(X \in [a\,;b]) = \int_a^b f(x)dx$$

Question finale : quel calcul permet de répondre à la question posée ?

Pour reprendre l’exemple donné plus haut, on a

$$P(X \in [0{,}2\,;0{,}4]) = \int_{0,2}^{0,4} (-2x+2)dx$$

$$= \int_{0,2}^{0,4} -2xdx + \int_{0,2}^{0,4} 2dx$$

$$= [-x^2]_{0,2}^{0,4} + 2 * (0{,}4 - 0{,}2)$$

$$= 0{,}2^2 - 0{,}4^2 + 0{,}4 = 0{,}04 - 0{,}16 + 0{,}4 = 0{,}28$$

Question bonus : combien vaut la quantité suivante ?

$$\int_{0}^{1} (-2x + 2)dx$$

Question bonus : combien vaut la quantité suivante ?

$$\int_0^1 (-2x + 2)dx = \int_0^1 -2xdx + \int_0^1 2dx$$

$$= [-x^2]_0^1 + 2 * (1 - 0)$$

$$= 0^2 - 1^2 + 2 = -1 + 2 = 1$$

La somme des probabilités est toujours égale à 1 !

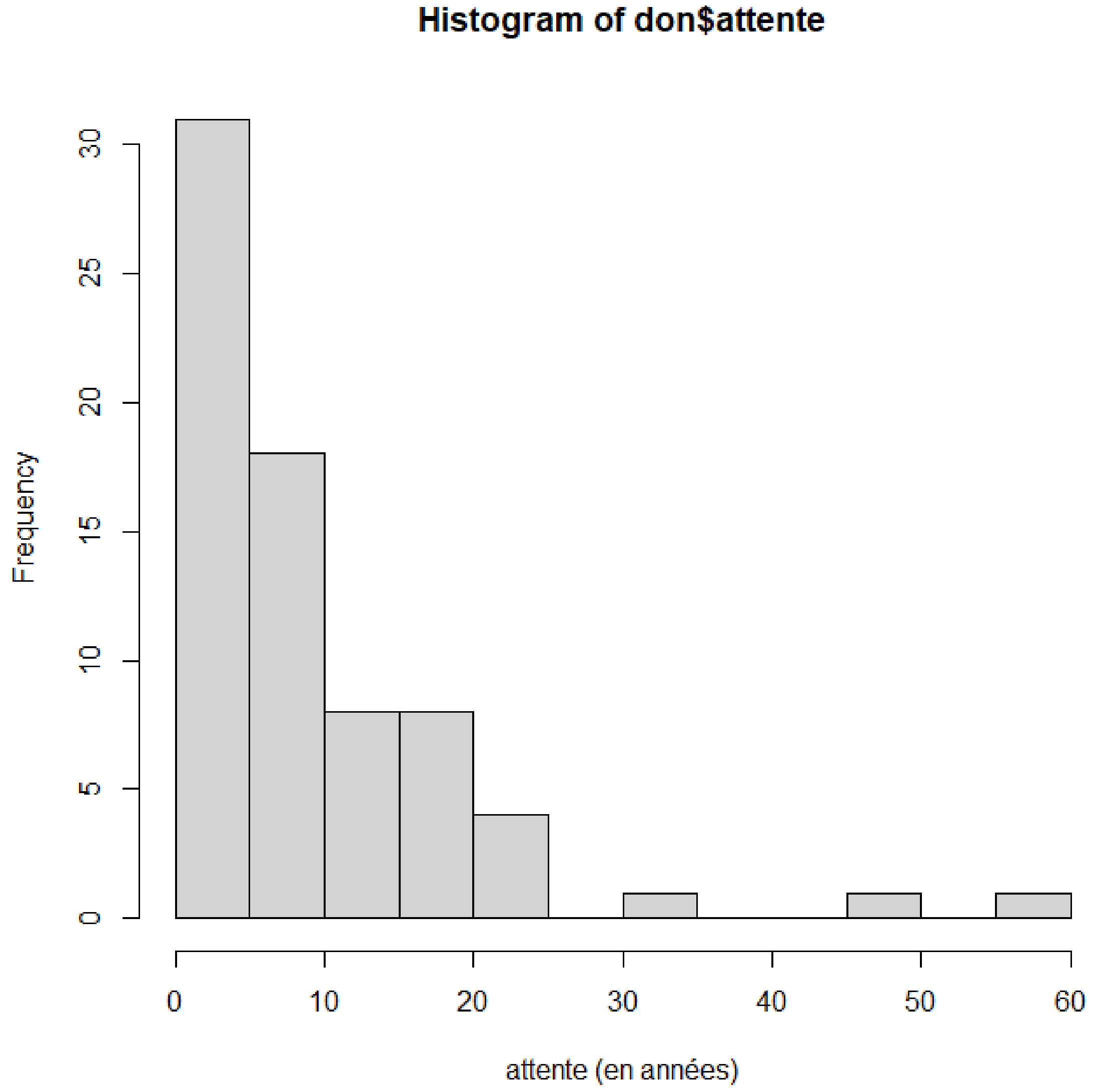
Le volcan Aso

On dispose des données d'éruption du volcan Aso (Japon) entre 1229 et 1897.

Sachant que le volcan est en éruption à la date d'aujourd'hui, quelle est la probabilité que la prochaine éruption ait lieu :

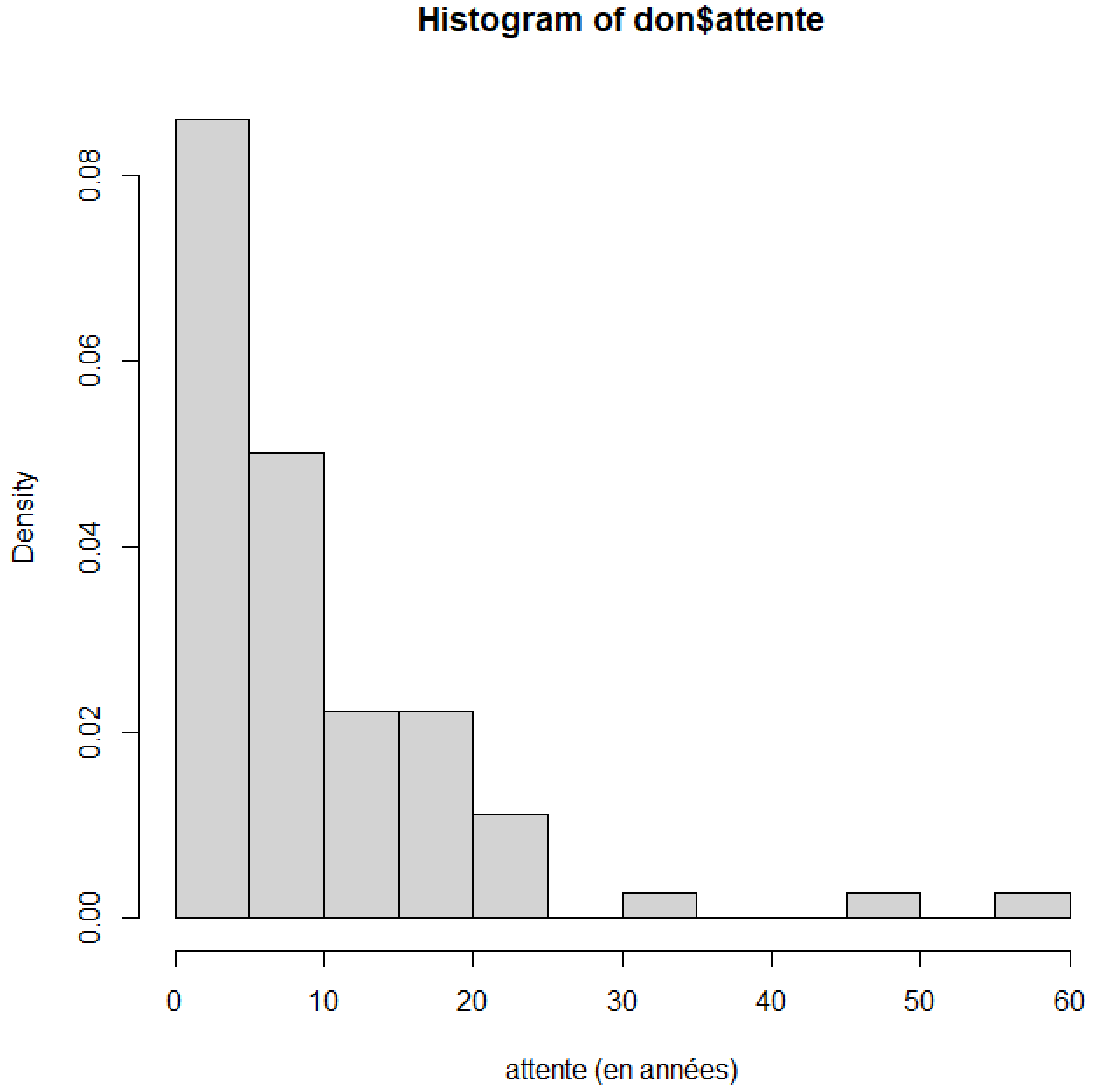
- Dans les 5 ans ?
- Au cours de l'année 2030 ?

Question intermédiaire n°1 : Quelle est la distribution du temps d’attente ?



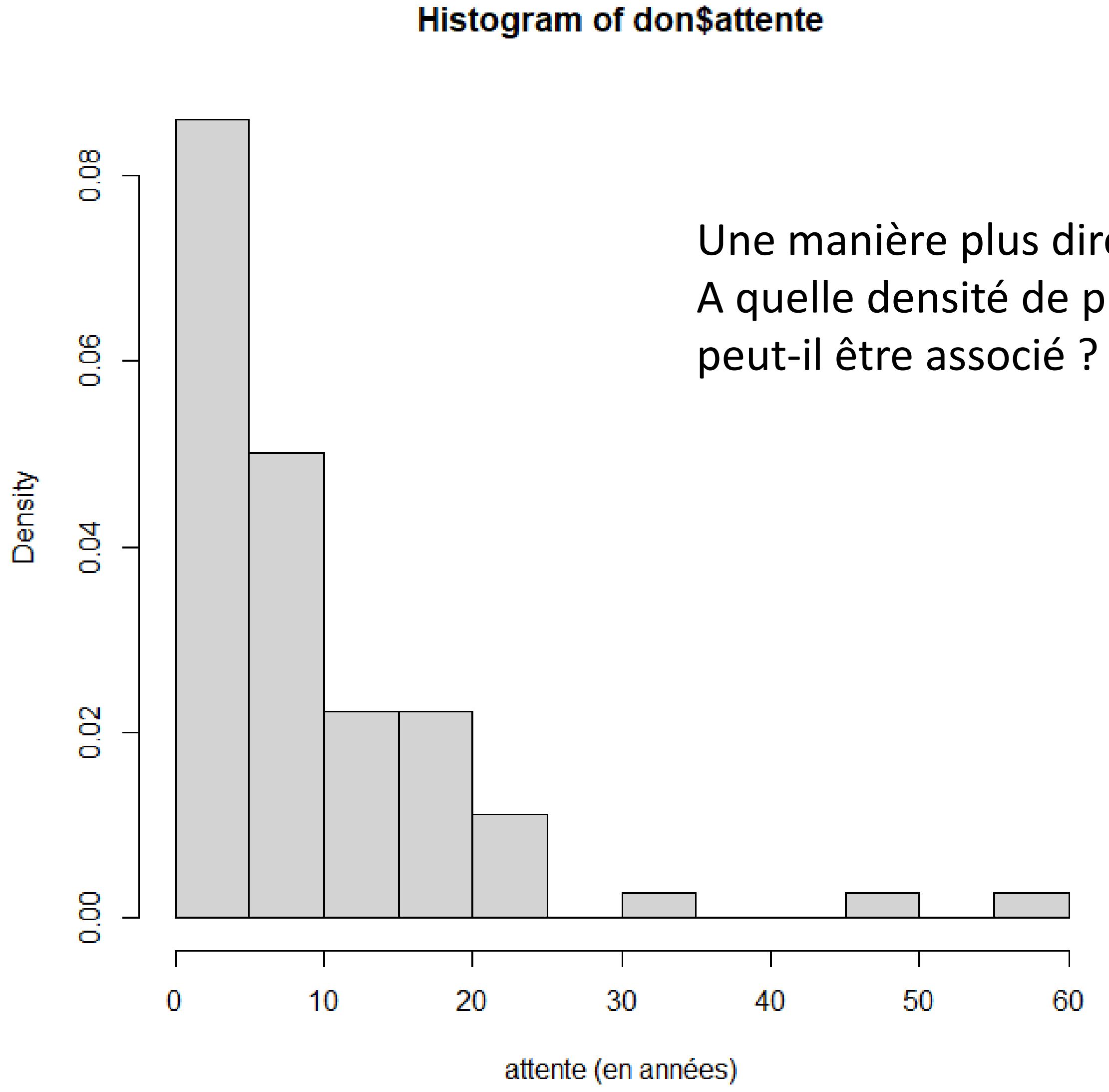
On dispose de 73 années d’éruption donc de n=72 temps d’attente entre deux éruptions

Question intermédiaire n°1 : Quelle est la distribution du temps d’attente ?



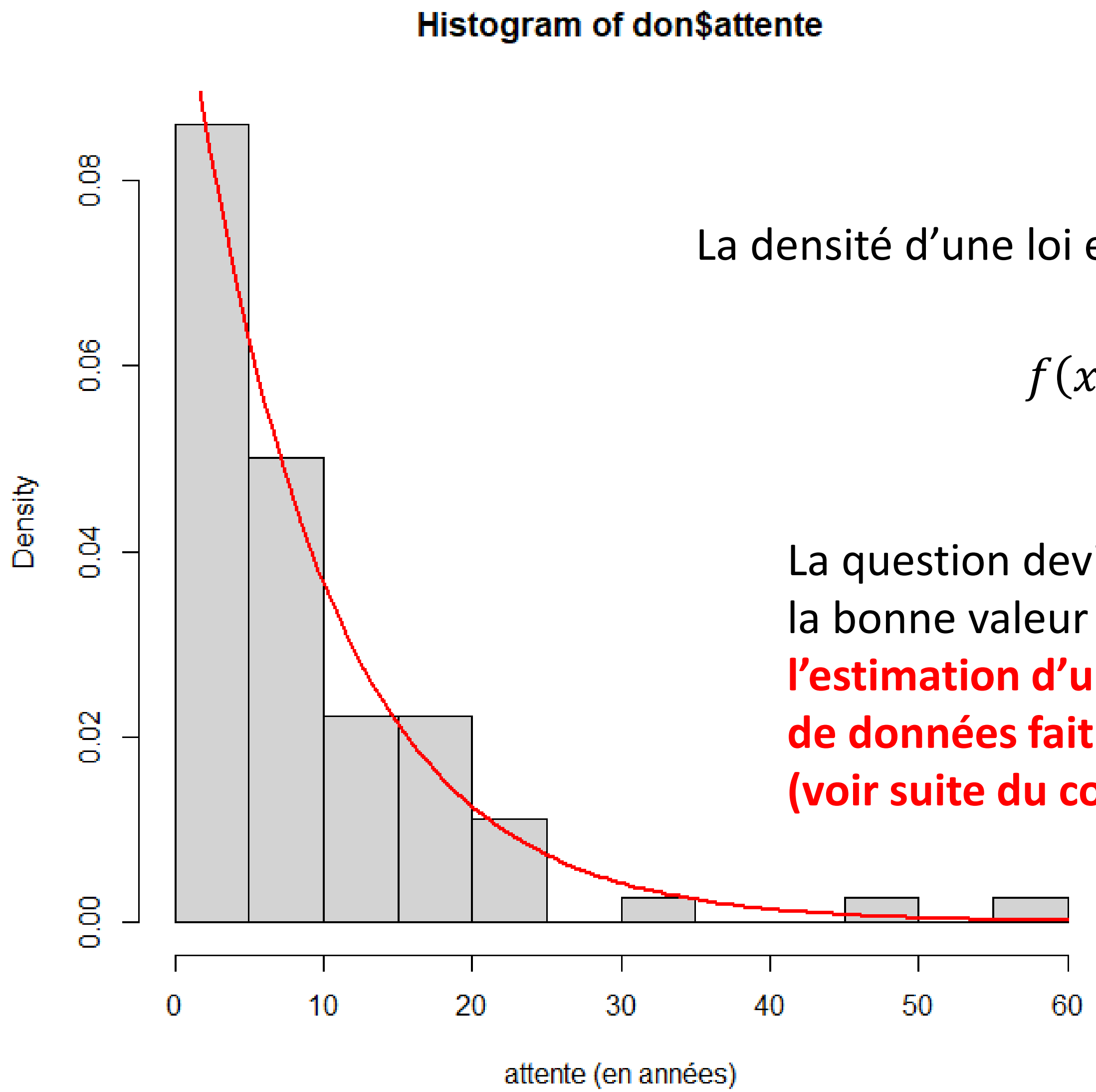
L’histogramme sous forme de densité

Question intermédiaire n°2 : Quelle est l'équation de la courbe qui lisse l'histogramme ?



Une manière plus directe de poser la question serait :
A quelle densité de probabilité l'histogramme ci-contre peut-il être associé ?

Question intermédiaire n°2 : Quelle est l'équation de la courbe qui lisse l'histogramme ?

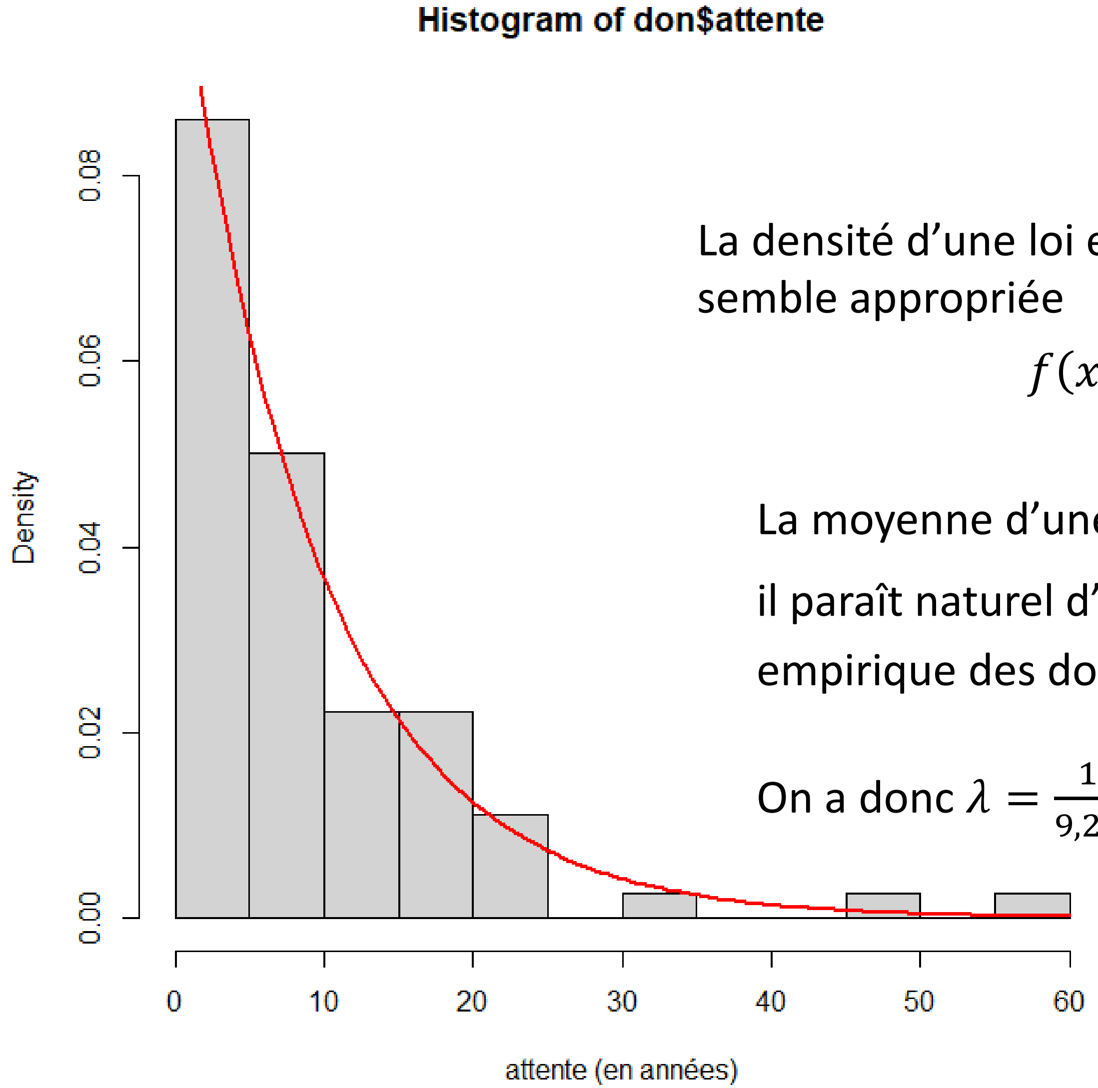


La densité d'une loi exponentielle semble appropriée

$$f(x) = \lambda \exp(-\lambda x)$$

La question devient comment trouver la bonne valeur pour λ ? **-> cette question de l'estimation d'un paramètre théorique à partir de données fait partie de la statistique inférentielle (voir suite du cours)**

Question intermédiaire n°2 : Quelle est l'équation de la courbe qui lisse l'histogramme ?



La densité d'une loi exponentielle de paramètre λ semble appropriée

$$f(x) = \lambda \exp(-\lambda x)$$

La moyenne d'une telle loi est donnée par $\frac{1}{\lambda}$. Ainsi, il paraît naturel d'estimer $\frac{1}{\lambda}$ par la moyenne empirique des données, notée $m = 9{,}28$.

On a donc $\lambda = \frac{1}{9{,}28} = 0{,}108$

Réponse aux questions

Pour rappel, le volcan est actuellement (2023) en éruption, la probabilité qu'une nouvelle éruption ait lieu dans les 5 ans est donc

$$P(X \in [0\,;5]) = \int_0^5 \lambda \exp(-\lambda x)\, dx = [-\exp(-\lambda x)]_0^5 = 1 - \exp(-5\lambda)$$

Sachant que $\lambda = 0{,}108$, on a

$$P(X \in [0\,;5]) = 0{,}42$$

Réponse aux questions

Pour rappel, le volcan est actuellement (2023) en éruption, la probabilité qu'une nouvelle éruption ait lieu au cours de l'année 2030 est donc

$$P(X \in [6{,}25\,;7{,}25]) = \int_{6,25}^{7,25} \lambda \exp(-\lambda x)\, dx = [-\exp(-\lambda x)]_{6,25}^{7,25}$$

$$= \exp(-6{,}25\lambda) - \exp(-7{,}25\lambda)$$

Sachant que $\lambda = 0{,}108$, on a

$$P(X \in [6{,}25\,;7{,}25]) = 0{,}05$$

Variables aléatoires discrètes

- Une variable aléatoire X est dite discrète si elle prend ses valeurs dans un ensemble fini ou dénombrable.

- Dans ce cas, la loi de X est donnée par les probabilités suivantes, pour tout x

$$P(X = x)$$

Variables aléatoires continues

- Une variable aléatoire X qui prend un nombre infini non dénombrable de valeurs est dite continue.

- Dans ce cas, la loi de X est déterminée par l'ensemble de probabilités

$$P(a < X < b)$$

- Pour tout $a < b$

Fonction de répartition

- La fonction de répartition d'une variable aléatoire X, notée F_X est

$$F_X(x) = P(X \leq x)$$

- Propriétés : F_X est croissante, continue à droite, et comprise entre 0 et 1 (elle tend vers 0 en -∞ et vers 1 en +∞)

Densité de probabilité

Si l'on peut écrire la fonction de répartition d'une variable aléatoire continue X de la manière suivante :

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

Alors f_X est la densité de probabilité de X

Densité de probabilité : propriétés

$$P(a < X < b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$$

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

Espérance d'une variable aléatoire

L'espérance d'une variable aléatoire correspond à sa valeur moyenne

- Cas discret

$$E[X] = \sum_{i=1}^{+\infty} x_i P(X = x_i)$$

- Cas continu

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Variance d’une variable aléatoire

La variance d’une variable aléatoire mesure la dispersion autour de la moyenne

$$\mathrm{Var}[X] = \mathrm{E}[(X - E[X])^2] = E[X^2] - E[X]^2$$

L’écart-type correspond à la racine carrée de la variance

$$\sigma[X] = \sqrt{Var[X]}$$

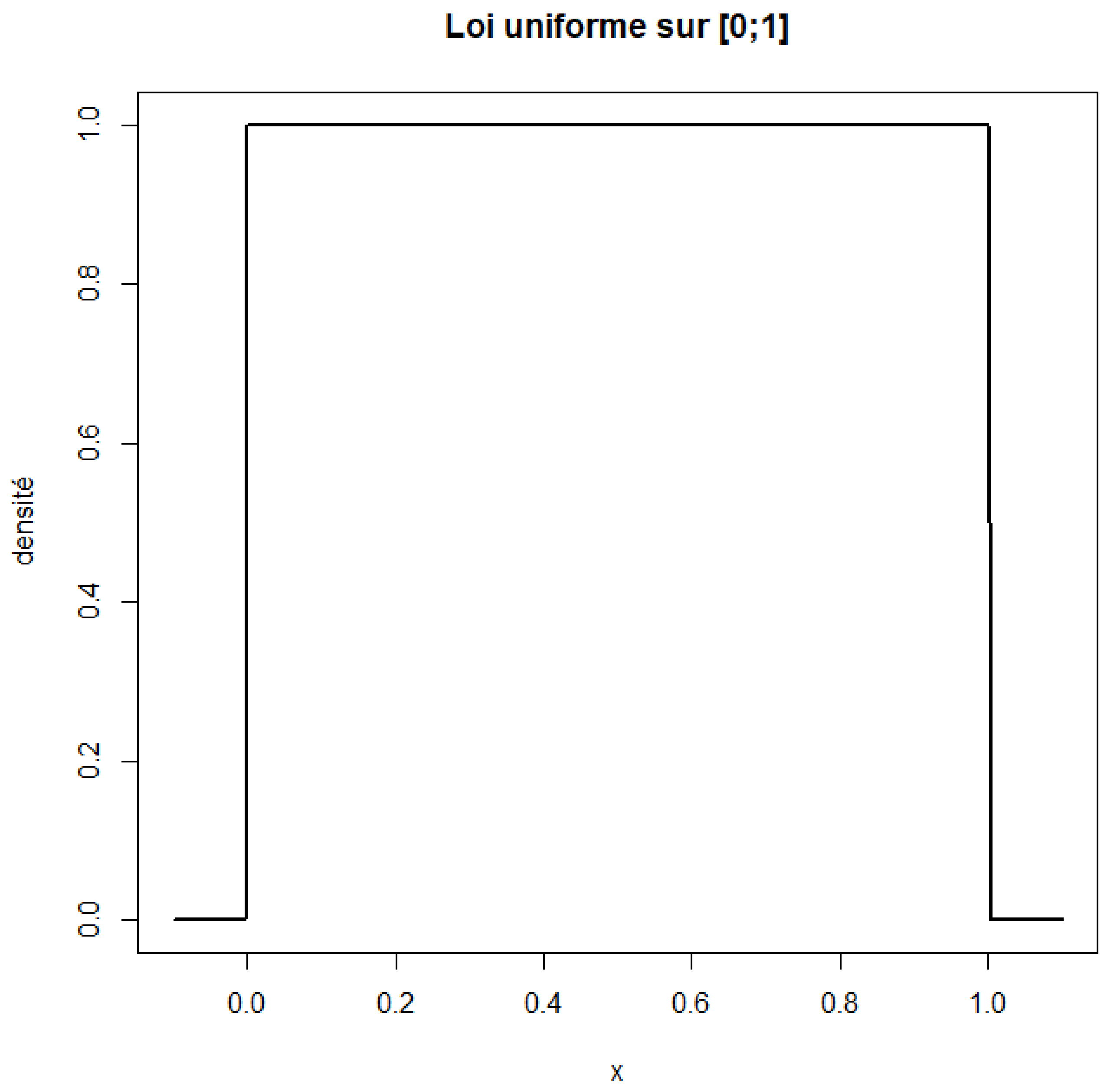
Exemple de loi continue : uniforme

Soit X une variable aléatoire sur $[a;b]$telle que

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & si\ x \in [a;b] \\ 0, & sinon \end{cases}$$

On dit que X suit une loi uniforme sur l'intervalle [a;b]

On a $\mathrm{E}[X] = \frac{\mathrm{a+b}}{2}$ et $\mathrm{Var}[X] = \frac{(b-a)^2}{12}$



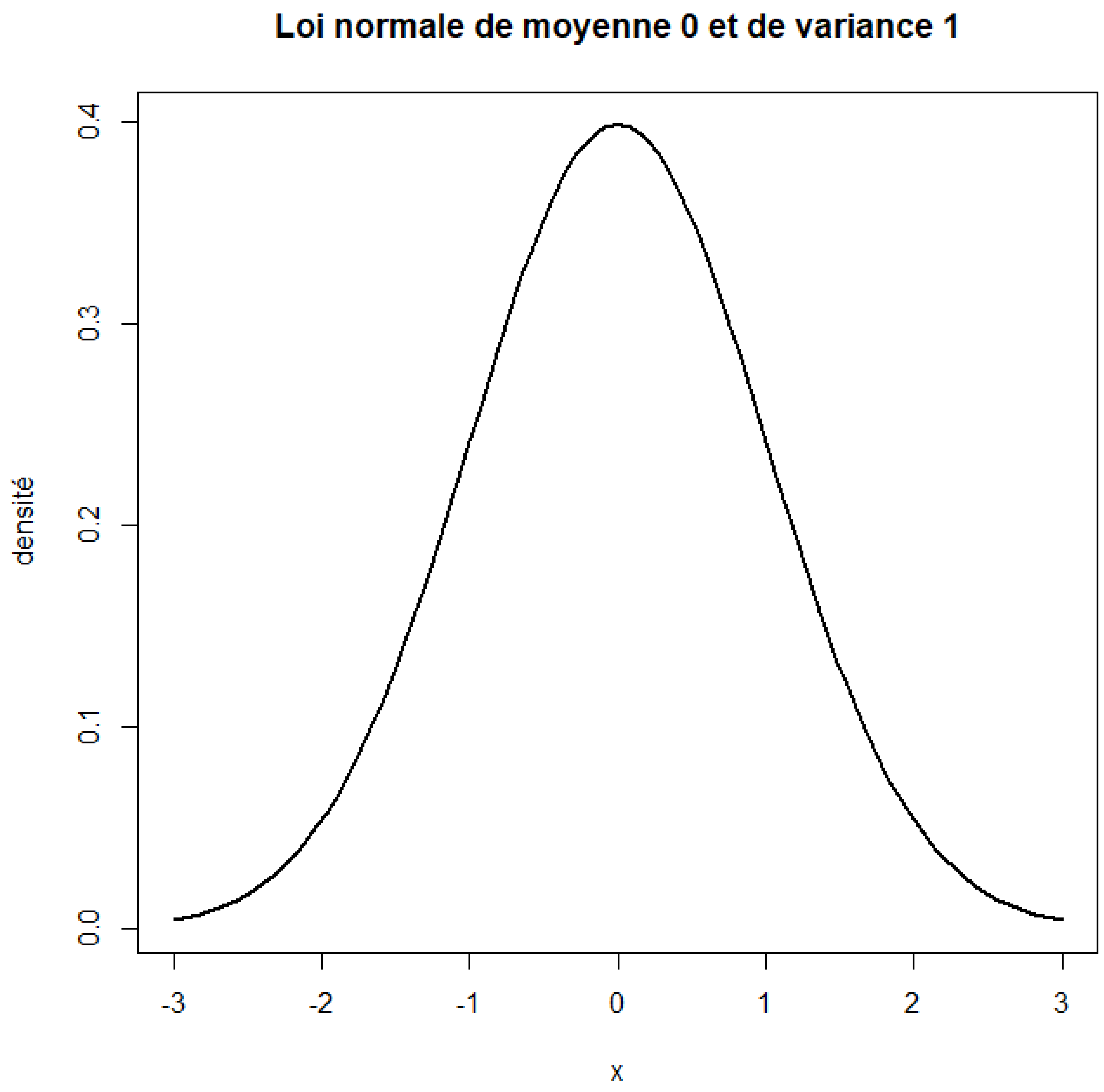
Loi uniforme sur [0;1]
densité
0.0
0.2
0.4
0.6
0.8
1.0
x

Exemple de loi continue : Normale

Soit X une variable aléatoire sur $\mathbb{R}$ telle que

$$f_X(x) = \frac{1}{\sigma^2\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

On dit que X suit une loi Normale (ou gaussienne) de moyenne μ et de variance σ^2; on note $X \sim N(\mu, \sigma^2)$


Loi normale de moyenne 0 et de variance 1
densité
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
x

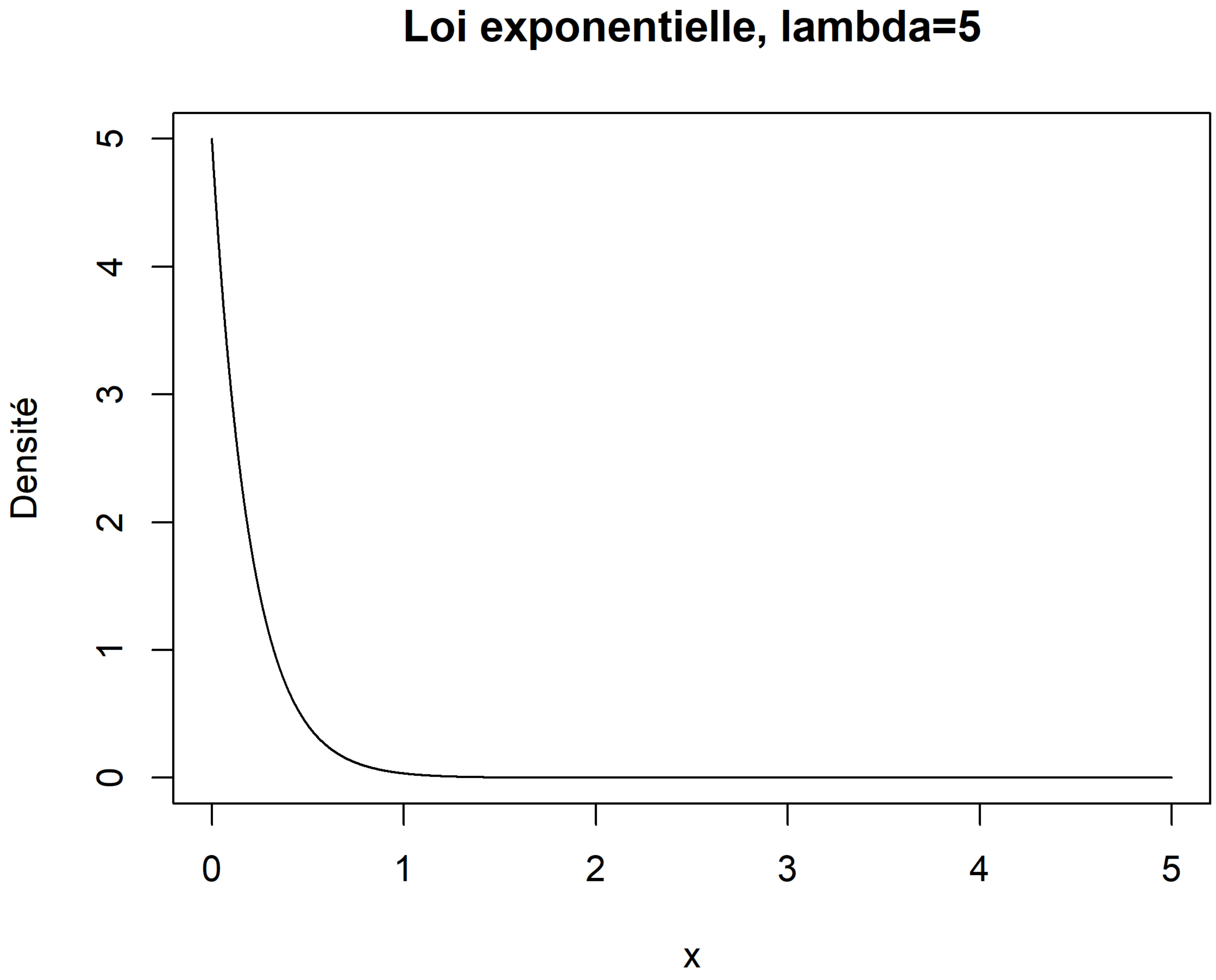
Exemple de loi continue : Exponentielle

Soit T une variable aléatoire sur $[0; \infty[$ telle que

$$f_T(t) = \lambda e^{-\lambda t}$$

On dit que T suit une exponentielle de paramètre λ ; $T \sim \mathcal{E}(\lambda)$

On a $\mathrm{E}[T] = \frac{1}{\lambda}$ et $\mathrm{Var}[T] = \frac{1}{\lambda^2}$


Loi exponentielle, lambda=5
Densité
0
1
2
3
4
5
0
1
2
3
4
5
x