Et si on s'intéressait à la moyenne des écarts ?

Véronique Cerclé(*)

Lorsqu'on présente le calcul de la variance en première, il se trouve toujours un élève pour demander pourquoi on élève les écarts au carré, ce qui donne une moyenne qu'il n'est pas facile d'interpréter, alors qu'on pourrait simplement calculer la moyenne des écarts absolus. La question resurgit dans le cadre de la loi binomiale, puis de la loi normale. Selon la classe et l'humeur, il faut bien avouer que le professeur évacue généralement plus ou moins habilement la question. Il évoque par exemple le fait que cet objet n'a pas de propriétés mathématiques intéressantes, ou fait remarquer le parallèle avec le calcul de distance dans un repère.
Nous allons nous intéresser ici à cet objet absent du cours de maths : d'abord pour des séries à valeurs discrètes, nous mettrons en parallèle les définitions des contextes discret et continu et étudierons le cas du continu.

1. Avec des valeurs discrètes

Soit une série $x_1, x_2, \ldots, x_n$ de n valeurs rangées par ordre croissant. On définit
- sa **moyenne** μ par $\mu = (x_1 + x_2 + \ldots + x_n)/n$.
- sa **médiane** m par $\mathrm{m} = x_p$ si $n = 2p + 1$ et m = un nombre de l'intervalle $[x_p , x_{p+1}]$ si $n = 2p$.

Quelles sont les propriétés de ces nombres ?

a) Somme des écarts algébriques

Résultat : la moyenne est la valeur de a pour laquelle la somme des écarts algébriques à a est nulle.
Preuve : $\mathrm{s}(a) = (x_1 - a) + \ldots + (x_n - a) = (x_1 + \ldots + x_n) - na = n\mu - na$.
$\mathrm{s}(a) = 0$ si et seulement si $a = \mu$; on a $\mathrm{s}(\mu) = 0$.

Exemple simple : avec une série de trois valeurs 10, 14, 15, la moyenne est 13, la somme des écarts à 13 est $(-3) + 1 + 2 = 0$.

b) Somme des carrés des écarts

Résultat : la moyenne est la valeur de a qui minimise la somme des carrés des écarts à a.
Autrement dit : $\mathrm{d}(a) = (x_1 - a)^2 + \ldots + (x_n - a)^2$ est minimum lorsque $a = \mu$.

Preuve : $\mathrm{d}(a) = (x_1{}^2 + \ldots + x_n{}^2) - 2\,(x_1 + \ldots + x_n)\,a + n\,a^2$
$= (x_1{}^2 + \ldots + x_n{}^2) - 2\,n\mu\,a + n\,a^2$.
Ce polynôme de degré 2 en a admet un minimum au point où sa dérivée s'annule.
$\mathrm{d}'(a) = 2na - 2n\mu$, et $\mathrm{d}'(a) = 0$ si et seulement si $a = \mu$.
Le minimum de d est donc $\mathrm{d}(\mu)$.

(*) professeur au lycée Jean Moulin de Pézenas (34).

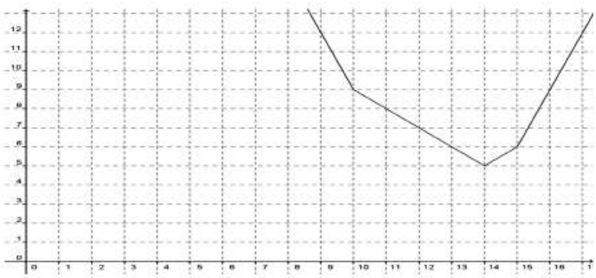
Exemple simple : Avec les mêmes valeurs (10, 14, 15) que précédemment :
$d(a) = (10 - a)^2 + (14 - a)^2 + (15 - a)^2 = 3a^2 - 78a + 521 = 3(a - 13)^2 + 14$ est minimum pour $a = 13$ qui est bien la moyenne. Ce minimum égale 14.

Définition : En divisant ce minimum $d(\mu) = (x_1 - \mu)^2 + \ldots + (x_n - \mu)^2$ par n, on obtient la **moyenne des carrés des écarts à la moyenne** : c'est la **variance**, notée V.

c) Somme des écarts absolus

Comme précédemment, je pose $D(a) = |x_1 - a| + \ldots + |x_n - a|$. La fonction D est une fonction affine par morceaux, de coefficients directeurs en ordre croissant, d'abord négatifs puis positifs, qui a donc un minimum.
Si $n = 2p + 1$ avec p entier, les coefficients directeurs de la fonction affine par morceaux sont impairs, d'abord négatifs puis positifs. Cette fonction a un minimum en x_p, qui est la médiane.

Dans notre exemple avec trois valeurs 10, 14, 15 :
$D(a) = |10 - a| + |14 - a| + |15 - a|$.



$D(a)$ est minimum pour $a = 14$ qui est bien la médiane. Le minimum de D est $D(14) = 5$.

Si $n = 2p$ avec p entier, les coefficients directeurs de la fonction affine par morceaux sont pairs, d'abord négatifs puis positifs. Cette fonction est minimale sur l'intervalle $[x_p, x_{p+1}]$ où elle est constante, toute valeur de cet intervalle convient, on reconnaît là aussi « la » médiane.

Exemple de quatre valeurs :
$D(a) = |x_1 - a| + |x_2 - a| + |x_3 - a| + |x_4 - a|$.

$$D(a) = \begin{cases} -4a + (x_1 + x_2 + x_3 + x_4) & \text{pour } a < x_1 \\ -2a + (-x_1 + x_2 + x_3 + x_4) & \text{pour } x_1 < a < x_2 \\ 0a + (-x_1 - x_2 + x_3 + x_4) & \text{pour } x_2 < a < x_3 \\ 2a + (-x_1 - x_2 - x_3 + x_4) & \text{pour } x_3 < a < x_4 \\ 4a - (x_1 + x_2 + x_3 + x_4) & \text{pour } a > x_4 \end{cases}$$

La valeur minimale de D est $(x_4 + x_3 - x_2 - x_1)$, atteinte pour n'importe quelle valeur de a comprise entre x_2 et x_3.

Ainsi la valeur de a qui minimise la somme D(a) des écarts absolus à a est « la » médiane m (ou toute médiane possible), et non la moyenne. Lorsqu'on s'intéresse aux écarts absolus, les écarts à la médiane ont une particularité plus pertinente que les écarts à la moyenne.

Définition : En divisant ce minimum $D(m) = |x_1 - m| + \ldots + |x_n - m|$ par n, on obtient la **moyenne des écarts absolus à la médiane** : c'est l'**écart absolu médian** que je noterai W par la suite.

d) De la somme à la moyenne

Plutôt qu'aux sommes des écarts, on va en réalité s'intéresser aux moyennes des écarts ; il suffit pour cela de diviser chacune des sommes précédentes par n, ce qui ne change pas les propriétés obtenues. Notons donc

Cas général	propriété	exemple simple avec trois valeurs 10, 14, 15 • moyenne μ=13 et médiane m=14
$s(a) = \Sigma(x_i-a)/n$ $d(a) = \Sigma(x_i-a)^2/n$	s s'annule en μ d est minimum en $a = \mu$, et son minimum est V	$V = d(\mu) = (3^2+1^2+2^2)/3 = 14/3 \approx 4{,}7$. Pour tout a, la moyenne $d(a)$ des carrés des écarts à a est supérieure à la variance V. Ainsi, pour a=14 par exemple, $d(14)=(4^2+0^2+1^2)/3=17/3 \approx 5{,}7$ est plus grand que V.
$D(a) = \Sigma\|x_i-a\|/n$	D est minimum en $a = m$, et son minimum est W	$W = D(m) = (4+0+1)/3 = 5/3 \approx 1{,}7$ Pour tout a, la moyenne $D(a)$ des écarts absolus à a est supérieure à W. Ainsi, pour a=13 par exemple, $D(13)=(3+1+2)/3 = 6/3 = 2$, plus grand que W.

2. Du discret au continu

Nous allons maintenant regarder nos objets (nos écarts s, d, D ainsi que les valeurs remarquables qui leur sont associées μ, m, V et W) dans trois cadres : le domaine des statistiques, le domaine des probabilités discrètes, et enfin le domaine des probabilités continues avec les lois à densité.

a) Moyenne des écarts algébriques

Cadre	Expression et valeur de la moyenne des écarts algébriques
Statistiques μ=moyenne = $(\Sigma x_i)/n$	$s(a) = \Sigma(x_i-a)/n = (\Sigma x_i)/n - a = \mu - a$
Probabilités discrètes μ = espérance = $\Sigma x_i \,.\, p_i$	$s(a) = \Sigma(x_i - a)\,.\,p_i = \Sigma x_i\,.\,p_i - a\,.\,\Sigma p_i = \Sigma x_i\,.\,p_i - a.1 = \mu - a$
Probabilités continues loi de proba de densité f μ = espérance = $\int t.f(t)\,dt$	$s(a) = \int (t-a).f(t)\,dt = \int t.f(t)\,dt - a \int f(t)\,dt = \mu - a.1$ $= \mu - a$

Dans ces trois situations, on vérifie que s s'annule en μ

b) Moyenne des carrés des écarts

Cadre	Expression de la moyenne des carrés des écarts
Statistiques	$d(a) = \Sigma(x_i - a)^2/n$
Probabilités discrètes	$d(a) = \Sigma(x_i - a)^2 \cdot p_i$
Probabilités continues loi de proba de densité f	$d(a) = \int (t-a)^2 f(t)\,dt$

On a montré qu'en statistique, d est minimum en μ, et que ce minimum est égal à la variance V : d(μ)=V. On va voir qu'il en est de même en probabilités continues .

c) Moyenne des écarts absolus

Cadre	Expression de la moyenne des écarts absolus
Statistique	$D(a) = \Sigma\|x_i - a\|/n$
Probabilités discrètes	$D(a) = \Sigma\|x_i - a\|.p_i$
Probabilités continues loi de proba de densité f	$D(a) = \int \|t-a\| f(t)\,dt$

On a montré qu'en statistique, D est minimum en la médiane m, et que ce minimum est l'écart absolu médian W : D(m) = W . Nous allons regarder ce qu'il en est dans le cadre des probabilités continues.

3. Cas du continu

On considère une loi de probabilité associée à la fonction de densité f sur I = [u , v] (les bornes de cet intervalle peuvent être éventuellement infinies).

On note F sa fonction de répartition : $\mathrm{F}(a) = \int_u^a f(t)\,\mathrm{d}t$; F est une primitive de f donc la dérivée de F est f ; par définition de F, on a F(u)=0, et par définition d'une fonction de densité, on a $\mathrm{F}(v) = \int_u^v f(t)\,\mathrm{d}t = 1$.

On note aussi $\mathrm{G}(a) = \int_u^a t f(t)\,\mathrm{d}t$; G est la primitive de $t \longrightarrow t\, f(t)$ qui vérifie G(u) = 0 ; la dérivée de G est G′(t) = $t\, f(t)$.

Enfin l'espérance est définie par $\mu = \int_u^v t f(t)\,\mathrm{d}t = \mathrm{G}(v)$.

a) Moyenne des carrés des écarts

$$\mathrm{d}(a) = \int_u^v (t-a)^2 f(t)\,\mathrm{d}t = \int_u^v t^2 f(t)\,\mathrm{d}t - 2a\int_u^v t f(t)\,\mathrm{d}t + a^2\int_u^v f(t)\,\mathrm{d}t.$$

$\mathrm{d}(a) = \int_u^v (t-a)^2 f(t)\,\mathrm{d}t = \int_u^v t^2 f(t)\,\mathrm{d}t - 2a\mu + a^2$ est minimum en $a = \mu$;

ce minimum est $\mathrm{V} = \mathrm{d}(\mu) = \int_u^v t^2 f(t)\,\mathrm{d}t - \mu^2$.

On reconnaît la formule habituelle $V(X) = E(X^2) - (E(X))^2$.

b) Moyenne des écarts absolus

$$D(a) = \int_u^v |t-a| f(t)\,dt = \int_u^a |t-a| f(t)\,dt + \int_a^v |t-a| f(t)\,dt.$$

$$D(a) = \int_u^a (a-t) f(t)\,dt + \int_a^v (t-a) f(t)\,dt.$$

$$D(a) = a \times \left(\int_u^a f(t)\,dt - \int_a^v f(t)\,dt \right) + \int_a^v t f(t)\,dt - \int_u^a t f(t)\,dt$$

$$D(a) = a \times \left(F(a) - (1 - F(a)) \right) + \mu - 2\int_u^a t f(t)\,dt.$$

$$D(a) = a \times (2F(a) - 1) + \mu - 2G(a).$$

En dérivant : $D'(a) = 2F(a) - 1 + a \times 2f(a) - 2\,a\,f(a) = 2F(a) - 1$.

$D'(a) = 0$ si et seulement si : $F(a) = 0{,}5$.

On reconnaît la définition de la **médiane** : **c'est la valeur de *a* pour laquelle F(*a*) = 0,5** (elle est unique car F est strictement croissante)

Notons m cette médiane ; **la valeur de *a* qui minimise la somme D(*a*) des écarts absolus à *a* est donc la médiane m**.

Ce minimum est : $D(m) = \int_m^v t f(t)\,dt - \int_u^m t f(t)\,dt = \mu - 2G(m)$.

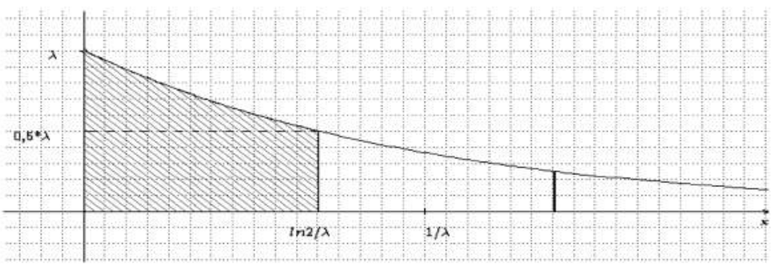
4. Calcul explicite pour les lois continues de terminale S

a) Loi exponentielle de paramètre λ

Résultats :

moyenne = espérance = $\mu = \dfrac{1}{\lambda}$ et variance $V = d(\mu) = \dfrac{1}{\lambda^2}$.

médiane $m = \dfrac{\ln 2}{\lambda}$ et $W = D(m) = \dfrac{\ln 2}{\lambda}$.



Preuves :

$f(t) = \lambda\, e^{-\lambda t}$ sur $[0;+\infty[$ donc :

$$F(a) = \int_0^a f(t)\,dt = 1 - e^{-\lambda a}.$$

$$G(a) = \int_0^a t\, f(t)\,dt = \left[\left(-t - \frac{1}{\lambda}\right)e^{-\lambda t}\right]_0^a = \frac{1}{\lambda} - \left(a + \frac{1}{\lambda}\right)e^{-\lambda a},$$

ce qui donne

• $\mu = \int_0^{+\infty} t\, f(t)\,dt = \lim_{a\to+\infty} G(a) = \frac{1}{\lambda}$.

• La variance V se calcule par intégration par parties.

• m vérifie $F(m) = 0{,}5 \Leftrightarrow e^{-\lambda m} = 0{,}5$; on retrouve bien $m = \frac{\ln 2}{\lambda}$.

• Le minimum W de D est $W = D(m) = \mu - 2\,G(m)$.

Or $\mu = \frac{1}{\lambda}$ et

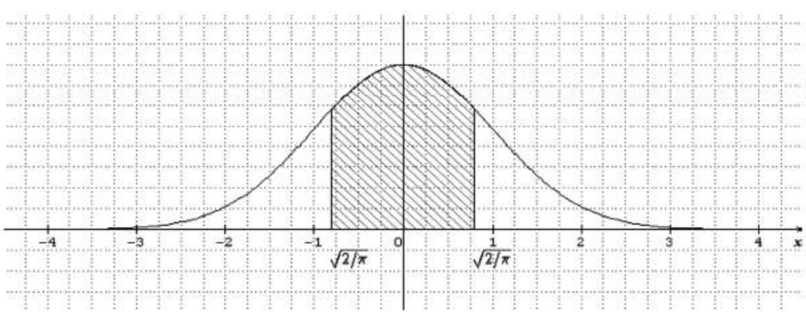
$$G(m) = \frac{1}{\lambda} - \left(m + \frac{1}{\lambda}\right)e^{-\lambda m} = \frac{1}{\lambda} - \left(m + \frac{1}{\lambda}\right)\times 0{,}5 = 0{,}5 \times \left(\frac{1}{\lambda} - m\right),$$

donc W =m.

b) loi normale centrée réduite N(0,1)

Résultats :
moyenne = espérance = $\mu = 0$ et variance $V = d(\mu) = 1$

médiane $m = 0$ et $W = D(m) = \sqrt{\frac{2}{\pi}}$.



Preuves :

$f(t)=\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-t^2/2}$ sur $]-\infty;+\infty[$ donne $\mathrm{F}(a)=\int_{-\infty}^{a} f(t)\,\mathrm{d}t$.

• $\mathrm{G}(a)=\int_{-\infty}^{a} t\,f(t)\,\mathrm{d}t=\int_{-\infty}^{a}\dfrac{1}{\sqrt{2\pi}}t\,\mathrm{e}^{-t^2/2}\,\mathrm{d}t=\left[\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-t^2/2}\right]_{-\infty}^{a}=\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-a^2/2}$,

ce qui donne $\mu=\int_{-\infty}^{+\infty} t\,f(t)\,\mathrm{d}t=\lim\limits_{a\to+\infty}\mathrm{G}(a)=0$ (évident d'ailleurs par symétrie).

• la variance est $\mathrm{V}=\mathrm{d}(\mu)=\int_{-\infty}^{+\infty} t^2 f(t)\,\mathrm{d}t-\mu^2$. Pour calculer l'intégrale, on utilise une propriété de la fonction de densité : $f'(t)=-t\times f(t)$ et on remplace alors $t^2f(t)$ par $-t\,f'(t)$. Par intégration par parties, il vient

$$\mathrm{V}=[-t\,f(t)]_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t-\mu^2=0+1-0^2=1.$$

• m vérifie F(m) = 0,5 $\Leftrightarrow$ m = 0.

• Le minimum de D est

$$\mathrm{W}=\mathrm{D}(0)=\mu-2\mathrm{G}(0)=0-2\left(-\frac{1}{\sqrt{2\pi}}\right)=\frac{1}{\sqrt{2\pi}}.$$

Conclusion

On a ainsi construit deux couples (tendance centrale ; dispersion) :

- l'étude de la **moyenne d(*a*) des carrés des écarts** conduit à la **moyenne** μ (valeur de *a* en laquelle d(*a*) est minimum), et à la **variance** V (valeur du minimum d(μ)). Le couple (moyenne, variance) est donc cohérent. Il a par ailleurs prouvé sa pertinence, par exemple pour centrer et réduire une binomiale, ce qui permet de se ramener à la loi normale standard par le théorème de Moivre-Laplace.
- l'étude de la **moyenne D(*a*) des écarts absolus** conduit à la **médiane** m (valeur de *a* en laquelle D(*a*) est minimum), et à W, **écart absolu médian**, (valeur du minimum D(m)). On obtient de ce point de vue un nouveau couple cohérent (médiane, W).

De plus la médiane comme mesure de tendance centrale est déjà associée au paramètre de mesure de dispersion qu'est l'intervalle inter-quartile, plus cohérent du point de vue du sens comme le montrent les diagrammes en boîte.

Tout ceci peut contribuer à expliquer que les écarts absolus à la moyenne, que les élèves aimeraient tant calculer, ne sont pas pertinents en mathématiques.