Notes du cours d'Équations aux Dérivées Partielles de l'ISIMA, première année
http://www.isima.fr/~leborgne

Les enseignants, les moyennes, les écarts types

Gilles LEBORGNE

16 juillet 2009

1 Résultat

But : comparer les niveaux de différentes classes.

Point de départ : dans chaque classe, suite à une correction de copies, on dispose de :

E l'ensemble des élèves de la classe, de cardinal $\mathrm{Card}E \stackrel{\text{noté}}{=} |E|$, et
$f : E \to \mathbb{R}$ fonction qui à chaque élève e associe sa note $f(e)$.

Usuellement on souhaite faire cette comparaison en ne retenant que deux valeurs pour chaque classe : la valeur moyenne M et l'écart type σ.

Question : à quoi vont servir M et σ ?

Réponse : ils servent à représenter la courbe gaussienne :

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-M)^2}{2\sigma^2}}, \tag{1}$$

ici gaussienne de moyenne M et d'écart type σ. Le facteur multiplicatif $\frac{1}{\sigma\sqrt{2\pi}}$ est là pour que la "masse" (l'aire sous la courbe) soit égale à $1 = 100\%$:

$$\int_{\mathbb{R}} g(x)\, dx = 1 = 100\%. \tag{2}$$

On peut ainsi, dans un premier temps, comparer deux classes en comparant les deux gaussiennes relative à chaque classe.

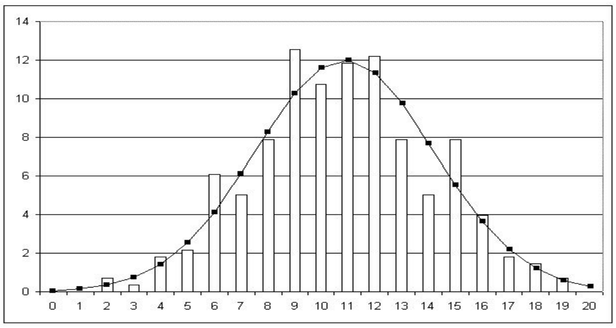

FIG. 1 – Histogramme de notes comparer à la gaussienne de même moyenne et écart type. En abscisse : l'échelle des notes de 0 à 20 ; en ordonnée : les pourcentages (exemple : à peu près 12.2% des élèves a eu 12/20). Dans cet exemple (les notes de mes élèves), la moyenne est $M \simeq 10.86$, et l'écart type est $\sigma \simeq 3.32$ qui est "assez grand" : les notes sont "assez étalées".

Représentation des notes : histogramme. Pour comparer les notes des élèves, pour chaque classe on fait un histogramme, c'est à dire on représente la fonction $h : [0, 20] \to \mathbb{N}$ en escalier t.q. :
en abscisse : l'échelle des notes, ici de 0 à 20,
en ordonnée : les valeurs $h(n) = \frac{\text{nb d'élèves ayant la note } n}{\text{nb total d'élèves}}$ = le pourcentage d'élèves ayant une même note.
Prendre le pourcentage permet d'avoir $\sum_{n=0}^{20} h(n) = 1 = 100\%$, et ainsi de pouvoir comparer cette histogramme avec la gaussienne g, cf. (2). Voir figure 1.

Remarque La “masse” de h (l'aire sous la courbe) est calculée avec h considérée comme fonction en escalier de valeur $h(n)$ sur chaque intervalle $]n-\frac12,n+\frac12[$ pour $n\in[0,20]_{\mathbb N}$:

$$\int_{0-\frac12}^{20+\frac12} h(x)\,dx = \sum_{n=0}^{20}\int_{n-\frac12}^{n+\frac12} h(x)\,dx = \sum_{n=0}^{20} h(n) = 1 = 100\%.$$

Ici on a supposé que les notes x étaient toutes des entiers : $x\in[0,20]_{\mathbb N}$.

Remarque Si le barème de notation va de demi-point en demi-point, i.e. $f(e)\in[0,\frac12,...]=\frac12[0,40]_{\mathbb N}$, la fonction en escalier h a “2 fois plus de marches”, et la masse est donnée par :

$$\int_{0-\frac14}^{20+\frac14} h(x)\,dx = \sum_{k=0}^{40}\int_{\frac k2-\frac14}^{\frac k2+\frac14} h(x)\,dx = \sum_{k=0}^{40}\frac12\,h(\frac k2) = 1.$$

Théorème central limite. Ce théorème de probabilités indique que, si on prend N classes avec N “très grand” (par exemple on prend $N=$ toutes les classes de 6ème en France), alors l'histogramme résultant, de moyenne M_F et d'écart type σ_F, “ressemblera vraiment beaucoup” à la gaussienne de moyenne M_F et écart type σ_F.

On peut ainsi comparer une classe de moyenne M et d'écart type σ avec la “classe France” en comparant les gaussiennes :

$$\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-M)^2}{2\sigma^2}} \qquad \text{vs} \qquad \frac{1}{\sigma_F\sqrt{2\pi}}\,e^{-\frac{(x-M_F)^2}{2\sigma_F^2}}.$$

2 Rappels des définitions : valeur moyenne et écart type

Définition. La moyenne (ou valeur moyenne) de la classe est :

$$M = \frac{\sum_{e\in E} f(e)}{|E|}. \tag{1}$$

Exemple. Moyenne des notes de 3 élèves, ayant 15, 10 et 12 : $M=\frac{f(1)+f(2)+f(3)}{3}=\frac{15+10+12}{3}\simeq 12.33$.

Remarque. Notation des probabilités : une fonction f est notée X (et appelée une variable aléatoire réelle), et sa valeur en $e\in E$ est notée x, soit ici $f(e)=^{\text{noté}} X(e)=^{\text{noté}} x$.

Et $\frac{1}{|E|}=^{\text{noté}} p$ est la probabilité de tirer au hasard l'élève e (sous l'hypothèse d'équiprobabilité).

La moyenne s'écrit $E(X)=\displaystyle\sum_{e\in E} X(e)\frac{1}{|E|}\overset{\text{noté}}{=}\sum_x x\,p\overset{\text{noté}}{=}\overline{X}$ (appelée l'espérance).

Définition. La variance (homogène à une aire) est l'écart quadratique moyen relativement à la moyenne :

$$V = \frac{\sum_{e\in E}(f(e)-M)^2}{|E|}. \tag{2}$$

Remarque. Notation des probabilités : la variance s'écrit $V(X)=\sum_x(x-\overline{X})^2\,p$.

Définition. L'écart type σ (homogène à une longueur) est la racine carrée de V :

$$\sigma = \sqrt{V}. \tag{3}$$

L'écart type est une mesure de la “dispersion” des notes par rapport à la moyenne : si $\sigma=0$ alors tous les élèves ont la même note M. Plus σ est grand et plus les notes sont “dispersées”; ainsi sur 100 notes, s'il y a 50 notes$=0$ et 50 notes$=20$, la moyenne est $M=10$, et l'écart type vaut $\sigma=\sqrt{\frac{1}{100}\sum_{e=1}^{100}10^2}=\sqrt{10^2}=10$ (une telle dispersion de notes est très très loin d'une gaussienne...). Et si $\sigma=0$, alors il n'y a aucune dispersion : toutes les notes sont identiques et valent M (ici c'est la limite de la gaussienne g quand $\sigma\to0$, limite appelée masse de Dirac δ_M).

Remarque. Pour la gaussienne, dans l'intervalle de notes $[M-\sigma,M+\sigma]$ on trouve $\simeq$ 68.26% des élèves : plus des deux-tiers des élèves ont leur note dans cet intervalle. En effet : $\int_{x=M-\sigma}^{M+\sigma}\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-M)^2}{2\sigma^2}}\,dx=\frac{1}{2\pi}\int_{z=-1}^{1}e^{-\frac{z^2}{2}}\,dz\simeq 68.26\%$.