

10ème Congrès Français d'Acoustique

Lyon, 12-16 Avril 2010

Acoustique, Statistique et “logique des niveaux”

Michel Maurin

INRETS-LTE, case 24, 69675 Bron cedex, maurin@inrets.fr

Les calculs qui s'appliquent sur les niveaux de bruit en acoustique peuvent être du ressort d'une modélisation théorique, mais aussi des calculs numériques pour traitements statistiques. Et là, la nature logarithmique des niveaux devrait appeler à une certaine vigilance, d'autant plus que les calculs sont à présent la plupart du temps “sous-traités” à l'informatique, laquelle considère par défaut que tous les nombres sont additifs et multiplicatifs.

Ce n'est pas le cas pour les niveaux. Il y a même incompatibilité entre les règles arithmétiques et les logarithmes, et cela nécessite que l'on recherche en conscience des arbitrages réfléchis entre la “logique des niveaux” et les règles ordinaires du calcul arithmétique. La communication fait le point sur certains arbitrages ou compromis qui sont envisageables à propos du calcul de la variance de niveaux, et de leur condition d'application pour des situations de portée locale.

1 Introduction

L'acoustique est une discipline de la physique qui travaille en priorité sur des niveaux de grandeur $L_S = 10 \log s/s_0$ [1] au lieu de grandeurs s elles-mêmes. Les raisons qui conduisent à l'introduction du logarithme peuvent être multiples et diversement situées [2], mais quoi qu'il en soit cette présence n'est pas sans conséquences sur les calculs que l'on peut mener.

En effet d'une manière générale en matière de calculs, on peut d'un côté faire la distinction entre les grandeurs mathématiques et les grandeurs physiques ou des sciences de la nature, et de l'autre la distinction entre les traitements théoriques et les traitements statistiques à portée heuristique, ce qui conduit au tableau croisé ci-contre, [3] :

développement et traitement	grandeurs	
	mathématiques	physiques
théorique	1	2
statistique	3	4

Comme le rappelle Poirier [4] “à la différence des mathématiques qui pratiquent parfois l'art pour l'art et le non-figuratif, la physique est essentiellement figurative, elle cherche la ressemblance”. L'art pour l'art (évidemment auto-cohérent) correspond à la case 1 du tableau, et l'on sait mener avec les niveaux des raisonnements et calculs sans perdre de vue leur nature logarithmique, des calculs qui correspondent au respect du *figuratif* de la case 2. Les développements de la statistique mathématique sont eux-mêmes artistiques et rigoureux (case 3), mais quand on aborde les calculs statistiques sur les grandeurs de la seconde colonne, en case 4, l'observation courante montre que les considérations figuratives auxquelles on pense assez naturellement pour la case 2 sont quelque peu perdues de vue.

Les statistiques les plus usuelles sont en effet des moyennes et des moments d'ordre deux (variance, covariance) qui sont fondés par défaut sur le fait que les

données traitées sont additives. Il s'ensuit que l'on risque être moins vigilant en présence de calculs statistiques, et ce d'autant plus qu'à présent ces derniers sont de plus en plus sous-traités à des moyens logiciels tout-venants “clés en main”, et étrangers à l'acoustique.

Bien entendu les niveaux ne sont pas additifs et ces mises en oeuvre de la statistique ne peuvent a priori respecter la représentativité des niveaux pour les phénomènes qu'ils repèrent. Il en résulte de facto un facteur d'incompatibilité entre les calculs ordinaires et banals qui souscrivent implicitement à la “logique de l'additivité”, et leur application sur les niveaux qui suivent la “logique des niveaux” [5] ; cette note est centrée sur des passerelles ou compromis que l'on peut établir entre ces deux logiques à propos des niveaux.

2 A propos de la moyenne

Dans un premier temps la moyenne soulève peu de difficultés car en acoustique on a explicitement développé la notion de niveau équivalent, ou énergétiquement équivalent, entre différents niveaux L_i , avec $L_{eq} = 10 \log\{1/n \sum 10^{L_i/10}\}$ en accord avec la case 2 ; (mais déjà un logiciel de régression linéaire entre niveau de bruit et vitesse de véhicule par exemple peut fort bien l'ignorer).

Techniquement cette définition rejoint la h -moyenne $h^{-1}\{1/n \sum_i h(y_i)\}$ de données y_i pour toute application h bijective, avec une moyenne dans l'espace des transformées et un “retour” dans l'espace d'origine avec la réciproque h^{-1} , [6]. En acoustique on utilise $h(L) = 10^{L/10}$ et $h^{-1}(x) = 10 \log x$, il en résulte que cette h -moyenne possède la propriété de la “translativité”, c'est à dire que si les niveaux L_i sont translatés d'une même quantité δ , le niveau équivalent $10 \log\{1/n \sum_i 10^{(L_i+\delta)/10}\} = 10 \log\{1/n \sum_i 10^{L_i/10}\} + \delta$ l'est également [7], au même titre que la moyenne arithmétique.

3 A propos de la variance

La question du conflit entre les deux logiques de calcul ci-dessus se pose avec davantage d'acuité pour les moments d'ordre 2. En effet la variance n'est pas une notion d'acoustique comme l'est le niveau équivalent, et le recours à une méthode statistique usuelle (case 4) peut négliger les précautions que l'on doit prendre pour respecter la représentativité du phénomène physique par les nombres et leurs opérations. On a déjà signalé que dans le cas d'une régression le niveau moyen des L_i est par défaut (et par automatisme) la moyenne arithmétique $1/n \sum L_i$.

Il faut donc rechercher un compromis pour la variance. La question est d'autant plus importante que la variance est une notion très prolixue en statistique qui se rencontre en de nombreux contextes (indice de dispersion, analyse linéaire en régression, décomposition en analyse de la variance). Mais ce peut être aussi une opportunité, car si une même approximation ne peut convenir à tout ces emplois techniques en statistique, on peut éventuellement développer des passerelles particulières, chacune avec une portée locale ; ci-dessous nous en présentons quatre.

3.1 L'étendue des niveaux

En premier s'il s'agit d'une simple information de dispersion sans vocation calculatoire, les indices statistique comme l'étendue des niveaux $[L_{\min}, L_{\max}]$ ou les quartiles de l'histogramme des L_i peuvent suffire, à condition de ne pas faire plus que des comparaisons numériques.

3.2 Une variance équivalente

Il y a en statistique la notion de transformation de variable aléatoire et de calcul des moments de la variable transformée [8, 9]. Cela permet, quand X est la variable, d'en déduire de manière approchée les premiers moments de la variable transformée $Y = f(X)$ à partir des moments de X , et c'est intéressant lorsque la variable X est additive avec des statistiques m_x et σ_x^2 qui ont un sens. A partir de $E(X) = m_x$ et $\text{var}(X) = \sigma_x^2$ on en déduit

$$E(Y) = m_y = f(m_x) + \sigma_x^2/2 f''(m_x) + \dots$$

$$\text{et } \sigma_y^2 = \sigma_x^2 f'^2(m_x) + \dots \quad (1)$$

a - Ici m_x et σ_x^2 ont un sens tandis que les statistiques m_y et σ_y^2 de la transformée n'en ont pas. Mais on note que le premier terme $f(m_x)$ du développement de $E(Y)$ est la f^1 -moyenne des Y_i avec des $f^1(Y_i)$ additifs et moyennables, et dans l'espace des Y on peut interpréter la f^1 -moyenne comme un indicateur de tendance centrale, comme une "moyenne équivalente" $m_{eqy} = f(m_x) = f(m_{f^1(y)})$ (2) conforme à celle de l'acoustique.

b - Par analogie on peut prendre comme "variance équivalente" le premier terme du développement de σ_y^2 . Cette quantité $\sigma_{eqy}^2 = \sigma_x^2 f'^2(m_x)$ (3) est calculée à partir des statistiques m_x et σ_x^2 , et avec la réciproque $h = f^{-1}$ on

note qu'elle est la variance ordinaire des $h(y_i)/h'(m_{eqy})$ puisque $m_{eqy} = f(m_x)$ et $h'[f(m_x)] f'(m_x) = 1$.

En acoustique les Y_i sont des niveaux L_i et les réciproques $x_i = 10^{L_i/10}$ sont des puissances additives. On a $f(x) = M \ln x$ avec $M = 10/\ln 10$, $f'(x) = M/x$ et $h(y) = 10^{y/10}$, et puis $\sigma_x^2 = 1/n \sum 10^{L_i/5} - (1/n \sum 10^{L_i/10})^2 = 1/n \sum 10^{L_i/5} - 10^{Leq/5} = 10^{Leq/5} \{1/n \sum 10^{(L_i-Leq)/5} - 1\}$, $f'(m_x) = M 10^{-Leq/10}$. On en déduit que $\sigma_{eqL}^2 = M^2 \{1/n \sum 10^{(L_i-Leq)/5} - 1\}$ (4) est exprimée en fonction des seuls niveaux L_i , et qu'elle est invariante par toute translation δ des niveaux pris en compte grâce à la translativité du Leq , au même titre que la variance usuelle. La variance équivalente est notamment utile pour des régressions linéaires avec des niveaux, que l'on peut appeler des régressions équivalentes [5, 10].

3.3 Une nouvelle décomposition algébrique

3.3.1 - h-moyenne et h-dispersion

i) Sur des données x_i générales on a l'inégalité $h^{-1} \{1/n \sum h(x_i)\} \geq 1/n \sum x_i$ lorsque h est convexe monotone [6], et la différence $h^{-1} \{1/n \sum h(x_i)\} - 1/n \sum x_i$ entre la h-moyenne et la moyenne arithmétique est positive ou nulle. C'est l'une des propriétés de la variance (ordinaire).

ii) On peut montrer que cette différence ne dépend des écarts arithmétiques ordinaires $\Delta x_i = \text{moy}(x_i) - x_i$ que si et seulement si l'application h est une exponentielle e^{cx} , [11]. Cela signifie que dans ce cas, et uniquement dans ce cas, la différence ci-dessus possède un second point avec la variance. Pour cette raison $h^{-1} \{1/n \sum h(x_i)\} - 1/n \sum x_i$ est appelée la "h-dispersion" et notée $h\text{-disp}(x_i)$, et en l'occurrence on a $h\text{-disp}(x_i) = 1/c \ln (1/n \sum e^{c\Delta x_i})$ (5). On veillera toutefois à éviter une confusion, en effet si la h-moyenne est définie pour toute application h bijective (convexe ou non), la h-dispersion n'est définie que pour les applications h exponentielles.

iii) On observe ici une autre propriété commune à la variance à propos du "mélange" de populations. En effet avec une population de données x_{pr} , d'indices p pour désigner une sous-population $p = 1 \dots P$, et r pour la répétition, $r = 1 \dots n_p \geq 2$, on a la décomposition algébrique classique

$$\text{variance totale} = \text{moyenne des variances}(p) + \text{variance des moyennes}(p).$$

Dans cette relation les variances(p) sont les variances internes à chaque sous-population et leur moyenne est la "variance intra" de la population totale, tandis que la variance des moyennes(p) par sous-population est la "variance inter" de la population totale.

Avec la h-dispersion d'une population totale on observe la décomposition suivante [11] :

h-dispersion totale = moyenne des h-dispersions(p) +
h-dispersion des h-moyennes(p).

De manière analogue à ci-dessus le premier terme moyenne des h-disp(p) est la “h-dispersion intra” de la population totale, quant au second terme, la h-dispersion des h-moy(p) est la “h-dispersion inter” qui caractérise la différence entre les sous-populations.

3.3.2 - Un autre compromis entre règles arithmétiques et logique des niveaux

Ces considérations algébriques et statistiques trouvent un écho immédiat en acoustique grâce au logarithme. En effet le niveau équivalent étant une h-moyenne, la différence positive h-disp(L_i) = Leq(L_i) - moy(L_i) est la h-dispersion des niveaux, et l'on voit comment cette définition peut constituer un nouveau compromis entre l'arithmétique et la logique des niveaux grâce à la présence simultanée et non artificielle des deux moyennes arithmétique et énergétique.

La décomposition ci-dessus correspond au cas d'un plan d'expériences à un facteur, avec des répétitions (en nombres inégaux éventuellement) pour chaque modalité ; en l'occurrence les modalités du facteur correspondent aux sous-populations p. La décomposition permet ainsi d'identifier la h-dispersion inter et la h-dispersion intra sur des niveaux, et de les comparer entre elles pour noter d'où provient la dispersion la plus élevée, au sein des modalités ou entre les modalités.

3.3.3 – Le cas d'un facteur bimodal

Le cas d'un facteur à deux modalités permet de faire la comparaison directe entre deux situations acoustiques, en deux sites différents par exemple, ou sur un même site avec des niveaux du genre “avant – après” ou “avec ou sans” à propos de la mise en place d'un changement de conditions, (autre mode de gestion du trafic, nouveau revêtement, ...) pour des écoulements de véhicules routiers.

Pour cela on dispose d'une première sous-population de niveaux notés L_i en nombre n₁ et d'une seconde de niveaux G_j en nombre n₂.

i) on a par définition h-disp(pop1) = h-disp(L_i) = Leq(L_i) - 1/n₁ ∑_i L_i
et h-disp(pop2) = h-disp(G_j) = Leq(G_j) - 1/n₂ ∑_j G_j ;

ii) il en résulte que h-disp intra = n₁/n₊ {Leq(L_i) - 1/n₁ ∑_i L_i} + n₂/n₊ {Leq(G_j) - 1/n₂ ∑_j G_j}
= 1/n₊ {n₁ Leq(L_i) + n₂ Leq(G_j)} - 1/n₊ [∑_i L_i + ∑_j G_j]
(6) ;

iii) h-disp inter = Leq{Leq(L_i), Leq(G_j)} - 1/n₊ [n₁ Leq(L_i) + n₂ Leq(G_j)] (7) .

iv) On peut aller au-delà et caractériser l'effet acoustique entre les deux sous-populations. En effet si l'on suppose que cet effet acoustique se manifeste essentiellement par une différence δ entre les Leq de chaque

sous-population, δ = Leq(G_j) - Leq(L_i), δ positif sans restreindre la généralité, on en déduit

h-disp inter = Leq{Leq(L_i), Leq(L_i) + δ} - Leq(L_i) - δ n₂/n₊
= 10 log[1/n₊ (n₁ 10^{Leq(L_i)/10} + n₂ 10^{Leq(L_i)/10} 10^{δ/10})] -
Leq(L_i) - δ n₂/n₊

= Leq(L_i) + 10 log(1 + n₂/n₊ (10^{δ/10} - 1)) - Leq(L_i) - δ n₂/n₊
= 10 log(1 + n₂/n₊ (10^{δ/10} - 1)) - δ n₂/n₊

ou encore M ln(1 + n₂/n₊ (e^{δ/M} - 1)) - δ n₂/n₊ (8).

La h-dispersion inter entre les deux sous-populations s'exprime à l'aide de la différence δ entre les Leq respectifs. Cela peut paraître trivial mais cela provient d'une construction de compromis qui n'est nullement triviale, et de la décomposition additive de la h-dispersion totale en ses parties inter et intra.

Dans cette relation on note la présence de l'opération arithmétique “différence” entre des Leq mais elle figure aussi dans la définition d'origine de la h-dispersion et cela fait partie du compromis.

3.4 L'introduction d'un aléa et de l'imprécision statistique

Les calculs précédents négligent la présence de l'aléa qui affecte toute mesure en pratique. Nous introduisons ici un terme pour cette dispersion qui tient compte de la logique des niveaux.

3.4.1 - Un terme d'erreur aléatoire (approché)

i) Soit s' le signal de la grandeur s que l'on mesure, le modèle classique des fluctuations de valeur revient à poser s' = s + ε avec un terme ε naturellement additif.

ii) Avec les niveaux L = 10 log s/s₀ on est conduit à considérer les niveaux perturbés L' = 10 log (s+ε_s)/s₀. Les propriétés générales du logarithme donnent ln(z+ε) = ln z + ln(1+ε/z) # ln z + ε/z pour des faibles perturbations telles que ε/z est petit par rapport à l'unité ; c'est à dire ε faible en valeur relative.

Dans le cadre de l'acoustique cela donne 10 log (s+ε_s)/s₀ = 10 log [s/s₀ (1+ε_s/s)] = 10 log s/s₀ + M ln(1+ε_s/s) # 10 log s/s₀ + M ε_s/s (9). Il en résulte que lorsque le signal est affecté par une perturbation additive ε_s qui demeure faible en valeur relative, son niveau 10 log s/s₀ (non additif par définition) est affecté par une perturbation elle-même additive sans dimension ε_L qui vérifie M ε_s # s ε_L (10), [12].

iii) Comme on travaille sur les niveaux et non pas sur les signaux, c'est l'aléa de ε_L qu'il faut modéliser, on peut prendre par exemple une variable aléatoire centrée de variance σ², comme un terme additif ordinaire.

3.4.2 - Des conséquences sur les niveaux équivalents

Avec plusieurs niveaux L_i = 10 log (s+ε_{si})/s₀ i = 1...n, ou 10 log s/s₀ + ε_{Li} avec n aléas ε_{si} et aussi ε_{Li} indépendants, on a Leq(L_i) = 10 log{1/n ∑ 10^{Li/10}} = 10 log(1/n ∑ s/s₀ + 1/n ∑ ε_{si}/s₀) = 10 log{s/s₀ (1 + 1/ns ∑ ε_{si})}

$= L + 10 \log(1 + 1/n_s \sum \varepsilon_{s_i}) \# L + M/n_s \sum \varepsilon_{s_i}$ lorsque les perturbations ε_{s_i} sont faibles en valeur relative.

On en déduit un nouveau terme additif (approché) sans dimension ε_{Leq} , et compte tenu des relations $M \varepsilon_{s_i} \# s \varepsilon_{L_i}$ il est égal à $\varepsilon_{L_i} = 1/n \sum \varepsilon_{L_i}$, centré et de variance σ^2/n .

3.4.3 - Des conséquences sur les h-dispersions, [10]

a - la h-dispersion inter d'un facteur bimodal

Avec les deux populations de niveaux $L_i = L_{pop1} + \varepsilon_{L_i}$ et $G_j = L_{pop2} + \varepsilon_{L_j}$, on a $Leq(L_i) = L_{pop1} + \varepsilon_{Leq1}$ avec un terme additif aléatoire de variance σ^2/n_1 , $Leq(G_j) = L_{pop2} + \varepsilon_{Leq2}$ avec un autre terme de variance σ^2/n_2 , (avec ici $\sigma_1 = \sigma_2 = \sigma$), et les deux termes ε_{Leq1} , ε_{Leq2} indépendants.

* Si l'on pose $L_{pop2} = L_{pop1} + \delta$ avec l'effet acoustique δ du § 3.3.3.iv, on en déduit pour la moyenne arithmétique des Leq , $1/n_+ \{n_1 Leq(L_i) + n_2 Leq(G_j)\} = L_{pop1} + \delta n_2/n_+ + 1/n_+ (n_1 \varepsilon_{Leq1} + n_2 \varepsilon_{Leq2})$ avec le terme aléatoire additif $n_1/n_+ \varepsilon_{Leq1} + n_2/n_+ \varepsilon_{Leq2}$.

* Pour la h-moyenne des deux h-moyennes on a $Leq\{Leq(L_i), Leq(G_j)\} = 10 \log\{1/n_+ (n_1 10^{L_{pop1}/10} 10^{\varepsilon_{Leq1}/10} + n_2 10^{L_{pop2}/10} 10^{\varepsilon_{Leq2}/10})\}$

$$\# L_{pop1} + 10 \log(1/n_+ (n_1 + n_2 10^{\delta/10})) + \frac{n_1 \varepsilon_{Leq1} + n_2 10^{\delta/10} \varepsilon_{Leq2}}{n_1 + n_2 10^{\delta/10}} \quad (11).$$

• Par différence la h-dispersion inter présente un terme constant égal à la quantité (8), positif croissant convexe en fonction de δ et nul à l'origine, et un terme aléatoire additif

$$n_1 \left(\frac{1}{n_1 + n_2 10^{\delta/10}} - \frac{1}{n_1 + n_2} \right) \varepsilon_{Leq1} + n_2 \left(\frac{10^{\delta/10}}{n_1 + n_2 10^{\delta/10}} - \frac{1}{n_1 + n_2} \right) \varepsilon_{Leq2}$$

$$\text{de variance } \sigma^2 \frac{n_1 n_2}{n_+} \left(\frac{10^{\delta/10} - 1}{n_1 + n_2 10^{\delta/10}} \right)^2.$$

Cette variance est nulle à l'origine, croissante et tend vers $\sigma^2 \frac{n_1 n_2}{n_+}$ quand δ tend vers $+\infty$.

Il en résulte que la présence du bruit dans les niveaux se traduit par un terme aléatoire additif de variance connue dans la h-dispersion inter. Dans ces conditions, connaissant

$$\text{l'écart-type } \sigma_\delta = \sigma \left(\frac{n_1 n_2}{n_+} \right)^{1/2} \frac{10^{\delta/10} - 1}{n_1 + n_2 10^{\delta/10}} \quad (12) \text{ qui}$$

affecte la h-dispersion inter pour un effet acoustique δ , on en déduit un "intervalle de confiance" $\delta \pm 1,96 \sigma_\delta$ au niveau de confiance 0,95, ou $\delta \pm 1,7 \sigma_\delta$ au niveau de 0,90 pour mieux caractériser l'effet en tenant compte des fluctuations aléatoires.

b - à propos de la h-dispersion intra

En revanche des calculs comparables appliqués à la h-dispersion intra donnent un terme aléatoire additif nul ; en l'occurrence les approximations dues à $\ln(z+\varepsilon) \# \ln z + \varepsilon/z$

sont ici trop "rustiques" et ne conviennent pas pour la h-dispersion intra.

4 Discussion et quelques conclusions

La recherche de compromis entre les deux logiques arithmétiques et logarithmiques que l'on ne peut manquer de croiser en acoustique est une question que l'on ne devrait éluder. Il s'agit d'un rapprochement entre des démarches à la fois légitimes et formellement inconciliables sur le plan des opérations numériques, et cela ne peut naturellement faire appel qu'à des approximations. La solution ne semble pas immédiate, l'unicité non plus, et il y a lieu de prospecter les opportunités possibles ; dans le même temps lorsque l'on en voit une se dessiner, cela peut demander un peu d'agilité technique et prendre des tournures relativement originales.

Les h-dispersions et les termes aléatoires approchés font partie des passerelles qui se présentent. La h-dispersion permet notamment de faire la part entre les effets intra et inter de deux populations comme ce que l'on fait ordinairement dans les plans d'expérience à un facteur, et l'on a ensuite la possibilité d'évaluer l'effet acoustique entre ces deux populations avec un intervalle de confiance associé.

Références

- [1] Liénard P., 1978, Décibels et indices de bruit, Masson
- [2] Maurin M., 1999, Quelques propos sur le logarithme et le décibel, Canadian acoustics, vol 27, part 4, 9-18
- [3] Maurin M., 2009, Sur des exigences de cohérence et de pertinence, le cas des traitements statistiques et celui de l'acoustique environnementale, note de travail INRETS-LTE, version janvier 2010
- [4] Poirier R., 1981, Mesure et réalité de l'objet, dans Parain-Vial J., Les difficultés de la quantification et de la mesure, Maloine S-A
- [5] Maurin M., 2003, Logarithme, niveaux, décibels et "logique des niveaux", rapport INRETS-LTE 0304, version juin 2009
- [6] Hardy G. H., Littlewood J. E., Pólya G., 1934, Inequalities, Cambridge Un. Press
- [7] Aczél J., 1987, A short course on functional equations, D. Reidel P. Company
- [8] Kendall M. G., Stuart A., 1963, The advanced theory of statistics, Ch. Griffin
- [9] Johnson N. L., Kotz S., 1970, Continuous univariate distributions, volume 1, New-York, J. Wiley and sons
- [10] Maurin M., 2008, Sur des compromis entre la logique des niveaux et les statistiques classiques, note de travail INRETS-LTE
- [11] Maurin M., 2007, Un amusement sur la dispersion des données, la h-dispersion, Journées SFdF 2007, Angers
- [12] Maurin M., 1998, Integrated versus instantaneous measurements, continued, what about precision ?, Euronoise 98, Munich