Moyenne arithmétique ou moyenne pondérée?

Claude Angers

1. Introduction

Les enseignants qui empruntent la moyenne pondérée pour évaluer les travaux de leurs étudiants sont conscients du caractère arbitraire de la sélection des poids qui déterminent dans les faits l'importance relative des travaux. Les poids assignés aux critères d'évaluation suscitent parfois des questions de la part d'étudiants qui «contestent» la pondération proposée et souhaitent en proposer une autre. Cette problématique suscite deux questions importantes: L'efficacité de la moyenne pondérée est-elle sensible aux variations des poids? Quand a-t-il une différence importante d'efficacité entre la moyenne arithmétique et la moyenne pondérée? Cet article a justement pour but de répondre à ces questions et de faciliter ainsi à l'enseignant le choix des poids à donner à ses critères d'évaluation.

Nous présentons, dans cet article, une technique statistique fort simple, qui permet de déterminer si l'emploi d'une moyenne arithmétique plutôt que d'une moyenne pondérée, et vice versa, exerce un effet sensible sur l'efficacité de l'évaluation finale. Nous y montrons que le calcul d'une moyenne pondérée comportant des variations relativement fortes des poids, plutôt que d'une moyenne arithmétique, n'augmente souvent que bien modestement la précision statistique de l'évaluation. Il s'ensuit qu'il n'est pas toujours opportun de chercher à déterminer les poids avec très grande précision. À ce propos, nous indiquons la façon de mesurer la perte d'efficacité de poids approximatifs.

2. Efficacité comparative de la moyenne arithmétique

Supposons qu'un professeur évalue un étudiant d'après n travaux d'importance inégale, au moyen de la moyenne pondérée

$$u = \sum_{i=1}^{n} p_i y_i = p_1 y_1 + p_2 y_2 + \dots + p_n y_n,$$

où y_i désigne la note de l'étudiant au i^e travail et p_i le poids de ce travail.

Si ce professeur calculait plutôt la moyenne arithmétique

$$v = \sum_{i=1}^{n} y_i/n = (y_1 + y_2 + \dots + y_n)/n,$$

celle-ci pourrait différer sensiblement de la moyenne pondérée et être même la source d'un biais prononcé. Pour que les deux moyennes, u et v, ne comportent ni l'une ni l'autre de biais statistique (ce qui ne signifie pas qu'elles donnent exactement le même résultat), il faut que les notes des travaux de l'étudiant représentent des observations aléatoires ayant la même moyenne (cette moyenne représente la valeur réelle de l'étudiant et chaque note la reflète; par exemple, un travail plus difficile coté plus sévèrement irait à l'encontre de l'hypothèse précédente; cette moyenne peut évidemment varier d'un étudiant à l'autre) et que la somme des poids soit égale à un.

Supposons maintenant que nos deux moyennes ont la propriété de n'être pas biaisées. Il nous reste à établir les conditions qui font de la moyenne pondérée le meilleur estimateur statistique de la performance de l'étudiant. On peut démontrer, en statistique, que la moyenne pondérée est le meilleur estimateur — parmi la classe infinie des estimateurs linéaires sans biais, d'après le théorème de Gauss-Markov — lorsque les cotes accordées à l'étudiant représentent des observations aléatoires indépendantes, ayant la même moyenne mais des variances (la variance est une mesure statistique de la dispersion autour de la moyenne) inversement proportionnelles à leurs poids. Grosso modo, l'indépendance des cotes signifie que les critères ne sont pas interreliés, ce qui est plutôt irréaliste dans le genre d'évaluation qui nous intéresse; l'hypothèse sur la variance veut dire que les critères estiment la valeur réelle de l'étudiant de façon beaucoup plus précise à mesure que leurs poids augmentent, ce qui est plus réaliste. Heureusement, des critères interreliés ne posent aucune difficulté au plan théorique, puisque la moyenne pondérée conserve toujours ses propriétés optimales, mais il n'est alors plus possible d'exprimer les poids idéaux (ceux qui conduisent à l'estimateur le plus efficace de la performance de l'étudiant, c'est-à-dire à l'estimateur possédant la plus petite variance) en fonction seulement de la variance des cotes.

Supposons donc que les notes de l'étudiant ne sont pas biaisées et que les poids choisis pour le calcul de u sont effectivement les meilleurs. Dans ce cas, on peut comparer l'efficacité des deux estimateurs, u et v, grâce au ratio

$$R = \frac{\text{variance de v}}{\text{variance de u}}$$

Comme u est plus efficace que v pour estimer la performance de l'étudiant, sa variance est plus petite et on a $R \geq 1$. Toute valeur de R tombant près de 1 signifie que les deux moyennes sont presque aussi efficaces l'une que l'autre, au plan statistique.

En pratique, il est facile d'étudier le ratio R à partir de la seule connaissance du poids le plus élevé, p_{max}, et du poids le moins élevé, p_{min}. En effet, à partir du ratio

$$B = p_{max}/p_{min},$$

Bloomfield et Watson (1975), Bloch et Moses (1988) ont démontré que

$$R \leq (B+1)^2/4B.$$

Ce résultat nous permet d'étudier rapidement l'inefficacité de la moyenne arithmétique. Par exemple, si les poids de quatre critères sont {0,2, 0,2, 0,3, 0,3} ou varie de 0,2 à 0,3, la moyenne arithmétique est presque aussi valable que la moyenne pondérée, puisque $B = 1{,}5$ et $R < 1{,}05$. Si ces poids varient plutôt de 0,15 à 0,35, alors $B = 2{,}34$, $R < 1{,}20$ et la moyenne arithmétique n'a pas encore perdu beaucoup de son efficacité. Par contre, si les poids en question s'étalent plutôt de 0,1 à 0,4, l'efficacité de la moyenne arithmétique pourrait s'avérer sensiblement plus faible, puisque $R < 1{,}57$. Personnellement, nous croyons qu'il y a peu de risque à choisir l'une ou l'autre moyenne $B \leq 4$ parce que la borne supérieure de R est très prudente (Bloomfield et Watson, 1975).

Le professeur adepte de la moyenne arithmétique trouvera donc avantageux, pour des raisons pédagogiques, de pondérer, par exemple, quatre travaux consécutifs {0,20, 0,25, 0,25, 0,30} (afin que le premier travail compte moins que le dernier), sans que son mode d'évaluation soit vraiment perturbé, puisque $R < 1{,}05$. Par contre, inverser les poids, c'est-à-dire {0,30, 0,25, 0,25, 0,20}, n'aurait pas de sens pédagogique, mais donnerait la même valeur de R.

3. Efficacité d'une moyenne pondérée approximative

Jusqu'à présent, nous avons opposé la moyenne pondérée à la moyenne arithmétique. Or, on peut aussi s'interroger sur l'à-propos ou l'inefficacité d'un autre choix de poids. Nous allons montrer comment traiter ce nouveau problème à l'aide d'un exemple. Supposons que les poids optimaux (en supposant toujours que les notes ne sont pas biaisées) de quatre critères soient $p = \{0{,}1,\ 0{,}1,\ 0{,}4,\ 0{,}4\}$. Si on choisit plutôt les poids $p' = \{0{,}15,\ 0{,}15,\ 0{,}35,\ 0{,}35\}$, on peut comparer l'efficacité des deux ensembles de poids en calculant $b = \{p'/p\}$, puis $B = b_{max}/b_{min}$, le calcul de R demeurant inchangé. Dans cet exemple, $b = \{1{,}5,\ 1{,}5,\ 0{,}876,\ 0{,}875\}$, $B = 1{,}72$ et $R < 1{,}08$. Par contre, en choisissant plutôt $p' = \{0{,}05,\ 0{,}05,\ 0{,}45,\ 0{,}45\}$, on obtient alors $R < 1{,}18$. Cet exemple montre bien qu'une bonne approximation des poids devrait suffire dans la plupart des cas.

4. Conclusion

Nous avons présenté au professeur soucieux de son évaluation une façon d'étudier l'effet de poids approximatifs et de déterminer s'il y a danger ou pas à utiliser une moyenne arithmétique plutôt qu'une moyenne pondérée et vice versa.

Évidemment, les résultats proposés dans cet article reposent sur un modèle statistique qui n'est sûrement pas représentatif de toutes les situations. Par exemple, le professeur qui cote très sévèrement certains types de travaux s'écarte du modèle, puisque chaque note octroyée à l'étudiant ne reflète plus sa moyenne véritable. Il est aussi bien connu que certains étudiants faibles aux examens individuels se rattrapent fort bien dans les activités de groupe. N'oublions pas, non plus, l'effet du professeur dans le déroulement d'un cours. Ce genre de situations engendre des biais dans le calcul des deux moyennes, mais ne rend pas forcément inutilisables les résultats de cet article. Ainsi, si le professeur peut partager ses critères en sous-groupes qui respectent les hypothèses à la base de la méthode pondérée, il pourra alors appliquer les résultats de cet article à chacun des sous-groupes.

5. Bibliographie

BLOCH, D. A. et MOSES, L. E. (1988), «Nonoptimally Weighted Least Squares», *The American Statistician*, Vol. 42, no 1, p. 50-53.

BLOOMFIELD, P. et WATSON, S. (1975), «The Ineffiency of Least Squares», *Biometrika*, 62, 121-128.

Claude Angers
École nationale d'administration publique