Sévérité et Significativité

Salim Lardjane [1]

[1] *LMBA UMR CNRS 6205, Université de Bretagne Sud*
Centre Yves Coppens, Bat. B, 1er ét., Campus de Tohannic
BP 573, 56017 Vannes, France
salim.lardjane@univ-ubs.fr

Résumé. L'auteur compare, dans la contexte d'une classe de tests de significativité, la notion d'évidence due à Deborah Mayo, qui est basée sur la sévérité (Mayo 1996, 2004, 2010, 2011) à la notion d'évidence bivariée de Bill Thompson (Thompson 2007), qui est basée sur la significativité. Il conclut en faveur de l'approche par la sévérité.

Mots-clés. Sévérité, Significativité, Tests de Significativité.

Abstract. The author compares, in the context of a class of significance tests, Deborah Mayo's notion of evidence, which is based on severity (Mayo 1996, 2004, 2010, 2011) to Bill Thompson's bivariate notion of evidence (Thompson 2007), which is based on significativity. He concludes in favor of the severity approach.

Keywords. Severity, Significativity, Significance tests.

1 Introduction

On se place dans ce travail dans le cadre du test d'une hypothèse simple contre une hypothèse simple, c'est-à-dire de

$$H_0 : \theta = \theta_0$$

contre

$$H_1 : \theta = \theta_1$$

où θ désigne la vraie valeur du paramètre de la loi commune d'un échantillon i.i.d. d'observations X. On suppose que celle-ci admet une densité f_θ par rapport à une mesure de référence λ (mesure de Lebesgue ou de comptage, en général).
On adopte comme statistique de test le rapport de vraisemblance

$$r(X) = f_1(X)/f_0(X),$$

où l'on a noté respectivement f_0 et f_1 la densité de X sous l'hypothèse que $\theta = \theta_0$ et $\theta = \theta_1$. On note x l'échantillon réalisé. Le degré de significativité, ou *p-value*, du test pour H_0, est défini par

$$p(H_0, H_1, x) = \int_{r(u)>r(x)} f_0(u)du.$$

On définit de même le degré de significativité du test pour H_1, par,

$$p(H_1, H_0, x) = \int_{r(u)\leq r(x)} f_1(u)du.$$

La règle de décision retenue consiste à accepter H_1 si $p(H_0, H_1, x) \leq \alpha$ et à accepter H_0 sinon, où α est un scalaire positif fixé, généralement "petit" (typiquement, $\alpha = 0.05$).

2 Sévérité

On dit que l'hypothèse H passe un test de sévérité s pour les données x (Mayo 1996, 2004, 2010, 2011) si (i) x conduit à accepter H et (ii) la probabilité d'obtenir un échantillon qui s'ajuste aussi bien ou moins bien à H que x, si H est fausse, est égale à s.
Ici, l'ajustement de x à H est quantifié par le rapport de vraisemblance, c'est-à-dire que x' s'ajuste moins bien à H_0 que x si $r(x') > r(x)$ et que x' s'ajuste moins bien à H_1 que x si $r(x') < r(x)$.
Pour tous les échantillons x *qui conduisent à accepter* H_1, on pose

$$s(H_0, H_1, x) = \int_{r(u)\leq r(x)} f_0(u)du.$$

Pour les échantillons x *qui conduisent à accepter* H_0, on pose

$$s(H_0, H_1, x) = \int_{r(u)>r(x)} f_1(u)du.$$

La fonction $s(H_0, H_1, x)$ ainsi définie est appelée *sévérité du test de H_0 contre H_1 pour les données x*. On dira que l'échantillon x est une évidence décisive en faveur de l'hypothèse acceptée si la sévérité correspondant est élevée (supérieure à $1 - \alpha$).
Notons que si x conduit à accepter H_1,

$$s(H_0, H_1, x) = 1 - p(H_0, H_1, x) > 1 - \alpha$$

et que, si x conduit à accepter H_0,

$$s(H_0, H_1, x) = 1 - p(H_1, H_0, x).$$

On retrouve, ainsi qu'il a été déjà souligné par Mayo (Mayo 1996, p. 194), que la sévérité d'un test de significativité au seuil α *conduisant à rejeter l'hypothèse nulle* est

nécessairement supérieure à $1-\alpha$.
Autrement dit, un test de significativité qui conduit à accepter l'hypothèse alternative est *automatiquement* un test sévère pour les données observées, pourvu que le seuil de significativité retenu soit faible.

3 Evidence bivariée

Bill Thompson (Thompson 2007) propose d'associer comme mesure d'évidence bivariée à un test de significativité du type précédent le couple de valeurs

$$ev(x) = (p(H_0, H_1, x), p(H_1, H_0, x)).$$

Dans le cadre de cette approche, on considère que x est une évidence décisive en faveur de H_1 si $p(H_0, H_1, x)$ est faible ($< \alpha$) et $p(H_1, H_0, x)$ est élevée ($> 1-\alpha$).
L'approche de Bill Thompson conclut donc à la nécessité de *qualifier* le résultat du test de significativité quand bien même il conduit à rejeter l'hypothèse nulle alors que dans le cadre de l'approche par la sévérité, un rejet de l'hypothèse nulle correspond automatiquement à une sévérité élevée et constitue donc une évidence décisive en faveur de l'hypothèse alternative. Les deux approches semblent donc se contredire.

4 Sévérité et évidence bivariée

Rappelons à quoi correspondent les quantités calculées dans le cas d'un rejet de l'hypothèse nulle, c'est-à-dire lorsque le résultat est peu favorable à l'hypothèse nulle ($p(H_0, H_1, x) < \alpha$).

- Approche par la sévérité : probabilité d'obtenir un résultat aussi favorable ou moins favorable à l'hypothèse alternative *sous l'hypothèse nulle*,

$$s(H_0, H_1, x) = \int_{r(u) \leq r(x)} f_0(u)du = 1 - p(H_0, H_1, x)$$

- Approche par la significativité : probabilité d'obtenir un résultat aussi favorable ou moins favorable à l'hypothèse alternative *sous l'hypothèse alternative*,

$$p(H_1, H_0, x) = \int_{r(u) \leq r(x)} f_1(u)du$$

Si la *sévérité* est élevée, cela signifie que, si l'hypothèse nulle était quand même vraie, alors le rapport de vraisemblance calculé serait atypique et donc qu'on a très peu de chances de se tromper en rejetant l'hypothèse nulle au vu de celui-ci.

Si la *significativité* du test pour l'alternative est élevé, cela signifie que le rapport de vraisemblance calculé est également atypique pour l'hypothèse alternative.
Ainsi, retenir le critère de Thompson comme mesure d'évidence revient à ne retenir comme évidence décisive pour l'hypothèse alternative que des rapports de vraisemblance à la fois atypiques sous l'hypothèse nulle et l'hypothèse alternative, ce qui, *en plus de mettre en cause la pertinence des deux hypothèses*, paraît indûment restrictif.
Notons que dans l'exemple donné dans (Thompson 2007, p. 108), Bill Thompson conclut au rejet d'une hypothèse nulle pour une évidence de $(0.05, 0.76)$; or si une significativité de 0.05 est considéré comme faible, une significativité élevée doit correspondre à une valeur de 0.95 au moins; il semble donc y avoir une contradiction, à moins d'adopter des seuils indépendants pour ce qu'on entend par significativité élevée et significativité faible.
L'approche basée sur la sévérité aboutit, quant à elle, à la conclusion donnée par Bill Thompson, sans requérir une telle indépendance.

5 Conclusion

Pour obtenir une évidence décisive en faveur de l'hypothèse alternative, il faut, dans le cadre de l'approche par la sévérité, observer un rapport de vraisemblance atypique sous l'hypothèse nulle. Dans le cadre de l'approche par l'évidence bivariée, il faut observer un rapport de vraisemblance atypique sous l'hypothèse nulle *et* sous l'hypothèse alternative. Ceci paraît être indûment restrictif et correspond à des situations où les hypothèses sont en fait toutes deux remises en cause par les valeurs observées. L'approche basée sur la sévérité paraît plus à même de correspondre à la démarche inférentielle, telle qu'elle est pratiquée quotidiennement dans le domaine de la recherche scientifique.

Bibliographie

[1] Mayo D. G. (1996). *Error and the growth of experimental knowledge*, The University of Chicago Press.

[2] Mayo D. G. (2004). *An error-statistical philosophy of evidence* in *The nature of scientific evidence* (Tapper M. L. and Lele S. R., eds), The University of Chicago Press.

[3] Mayo D. G., Spanos A. (2010). *Error and Inference*, Cambridge University Press.

[4] Mayo D. G., Spanos A. (2011). *Error Statistics* in *Philosophy of Statistics* (Bandyopadhyay P. S. and Forster M. R., eds), Elsevier.

[5] Thompson B. (2007). *The nature of statistical evidence*, Springer.