L'Année Psychologique, 1984, *84*, 77-83

NOTES

Groupe Mathématiques et Psychologie
Université René-Descartes
Sciences humaines — Sorbonne[1]

RÉINTERPRÉTATION FIDUCIO-BAYÉSIENNE DU TEST F DE L'ANALYSE DE LA VARIANCE

par Bruno LECOUTRE

Persée creative commons BY: =

SUMMARY : The Bayes-fiducial reinterpretation of the analysis of variance F*-test.*

The Bayes-fiducial reinterpretation of the usual F*-test of the analysis of variance was illustrated, with concrete examples, first for one degree of freedom comparisons, then for several degrees of freedom comparisons. This reinterpretation enables us to understand the reasons for the inadequacy of significance tests in relation to the problem of the generalisability of descriptive conclusions in experimental research.*

Key-words : Statistics, analysis of variance, Bayes-fiducial reinterpretation.

Dans toute expérimentation en psychologie, se pose le problème incontournable de la *généralisabilité* des conclusions, donc de l'utilisation de méthodes d'inférence statistique. Or les procédures usuellement employées (tests de signification) sont notoirement inadaptées ; ce fait est d'autant plus éclatant quand il s'agit, pour « valider » un modèle, de chercher à « accepter » l'hypothèse nulle. A l'inverse, les procédures fiducio-bayésiennes ou, plus généralement, bayésiennes apparaissent aptes à répondre aux besoins les plus diversifiés.

1. 12, rue Cujas, 75005 Paris.

L'objectif de la présente note est d'expliciter les raisons de l'insuffisance des tests de signification, en donnant la réinterprétation fiducio-bayésienne des tests F de l'analyse de la variance.

UN EXEMPLE A VALEUR GÉNÉRALE

Nous considérerons ici l'exemple fourni par une expérience de temps de réaction de choix, dans laquelle on étudie le temps de réaction en fonction du nombre de signaux possibles, les différents signaux étant équifréquents. Les données proviennent de l'article de Rouanet, Oléron et Régnier (1966). Cinq sujets ont effectué chacun 64 essais, pour chacune des quatre conditions expérimentales caractérisées par le nombre de signaux possibles, respectivement un, deux, quatre et huit. Cet exemple est à valeur générale, et on ne s'attardera pas sur le cas particulier choisi.

Dans un précédent article (Lecoutre, 1981), nous avons illustré, à propos de ces mêmes données, l'apport des procédures fiducio-bayésiennes pour l'*investigation des mécanismes individuels* ; nous ne reviendrons pas sur cet apport essentiel, mais le

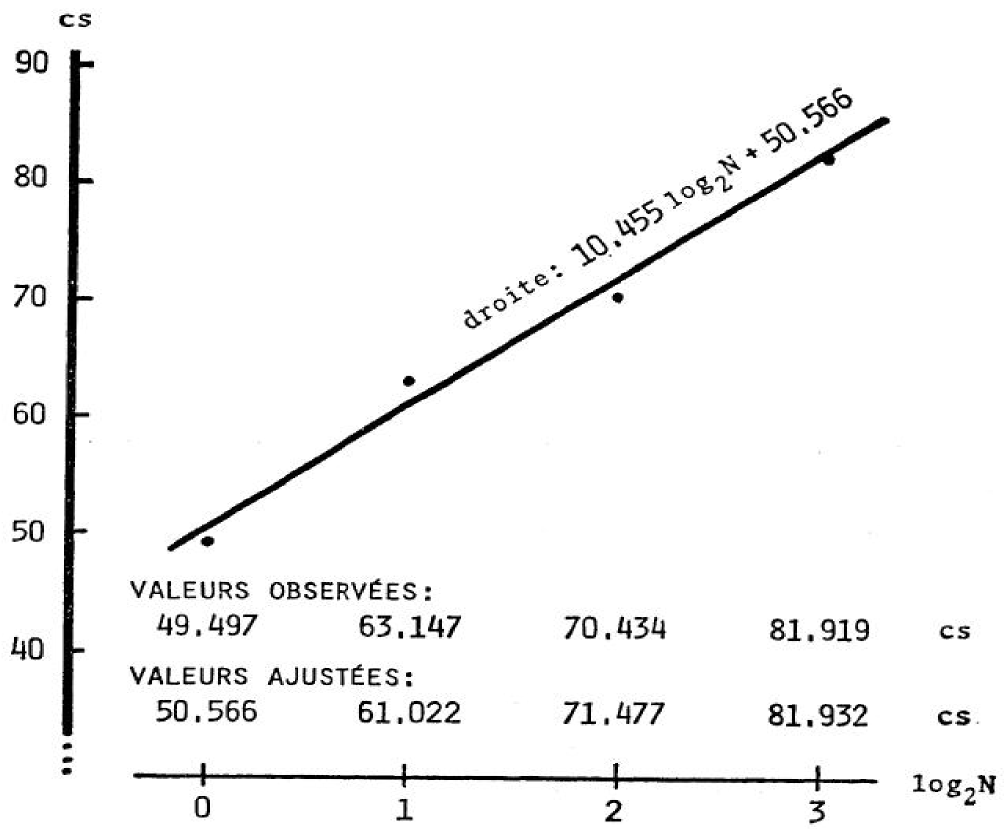

Fig. 1. — Expérience de temps de réaction de choix : données de groupe (temps exprimés en centièmes de secondes)

point précis que nous développerons ici, en complément de cet article, est la réinterprétation fiducio-bayésienne des tests F usuels. Ainsi, proposons-nous d'examiner le modèle suivant lequel le temps de réaction augmente suivant une fonction linéaire du logarithme binaire du nombre de signaux possibles. Nous pouvons distinguer dans ce modèle deux prédictions : 1) le temps de réaction augmente avec le nombre de signaux ; 2) le temps de réaction est, au moins approximativement, une fonction affine du logarithme binaire du nombre de signaux.

La figure 1 montre les données de groupe — moyennes des temps de réaction observés pour chacune des quatre conditions — ainsi que la droite ajustée par la méthode des moindres carrés, lorsqu'on porte en abscisse le logarithme binaire du nombre de signaux. Nous considérerons ici que, au niveau descriptif, les données de groupe sont en accord avec les prédictions du modèle ; il s'agit maintenant de chercher à prolonger cette conclusion au niveau inductif, par l'emploi de procédures statistiques inférentielles.

Le tableau I fournit les résultats de l'analyse de la variance usuelle, portant sur les comparaisons permettant d'examiner les

TABLEAU I. — *Résultats de l'analyse de la variance usuelle correspondant à la décomposition de la comparaison globale des quatre conditions (C) suivant les sous-comparaisons LIN C et C-LIN C*

Comparaison	d.l.	Somme des carrés	Rapport F	Seuil de signification observé
C	3	2 766,58		
LINC C	1	2 732,84	28,61	$p = 0{,}006$ (1-4 d.l.)
C-LIN C	2	33,74	1,75	$p = 0{,}23$ (2-4 d.l.)

prédictions du modèle : la comparaison LIN C à un degré de liberté et la comparaison résiduelle C-LIN C à deux degrés de liberté. Comme on le voit, la première est significative au seuil observé $p = 0{,}006$, tandis que la seconde est non significative au seuil observé $p = 0{,}23$.

Dans le premier cas, on peut conclure que l'*existence* de l'effet (augmentation du temps de réaction avec le nombre de signaux) est bien établie, mais cela ne nous renseigne pas le moins du monde sur l'*importance* de cet effet. Dans le second cas, le résultat non significatif n'est à strictement parler qu'un constat d'ignorance (qu'on ne puisse pas conclure à l'existence de l'effet n'implique sûrement pas qu'on puisse conclure à sa non-existence), et l'*absence de conclusion* à laquelle on devrait donc en toute rigueur se tenir révèle l'inadaptation complète du test de signification vis-à-vis de l'acceptabilité du modèle logarithmique.

RÉINTERPRÉTATION FIDUCIO-BAYÉSIENNE

Ces déclarations de principe constituent une mise au point utile, mais elles ne nous éclairent pas réellement sur les raisons de l'inadaptation des tests de signification. En revanche, les procédures fiducio-bayésiennes, qui prolongent de manière considérable les résultats des tests de signification, doivent permettre de bien comprendre les raisons de cette inadaptation.

1) Reconsidérons d'abord la situation de la comparaison LIN C (résultat significatif) ; nous pouvons prendre comme effet observé associé à cette comparaison à un degré de liberté la pente de la droite de régression de la figure 1, soit $d = + 10{,}455$ cs ; la valeur 10,455 représente l'augmentation moyenne observée du temps de réaction due à la multiplication par deux du nombre de signaux, quand on remplace les données par l'approximation linéaire. Cet effet observé est important ; pour prolonger inductivement cette conclusion, il s'agit d'effectuer une inférence sur l'effet moyen parent δ. La distribution fiducio-bayésienne relative à l'effet δ est une distribution du t de Student (ici à 4 degrés de liberté), centrée sur l'effet observé d, qui traduit l'information

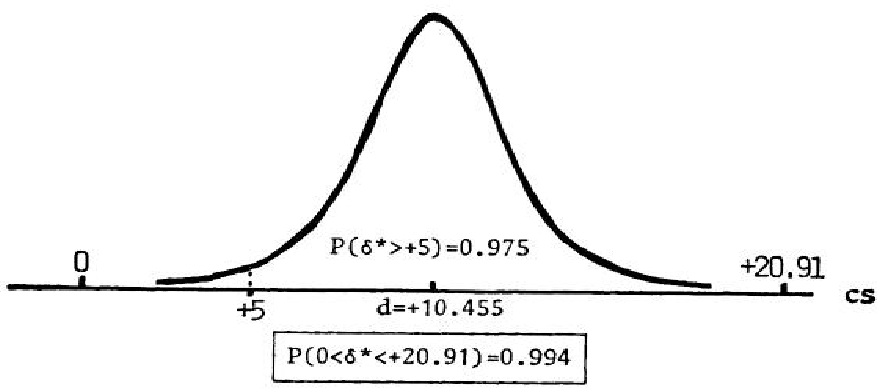

Fig. 2. — Réinterprétation fiducio-bayésienne du test F pour la comparaison à un degré de liberté LIN C : distribution fiducio-bayésienne relative à l'effet parent δ

apportée par les données sur l'effet parent δ ; cette distribution est figurée dans la figure 2.

Dans ce cas d'une comparaison à un seul degré de liberté, la réinterprétation fiducio-bayésienne du test F est que la probabilité fiducio-bayésienne que δ soit situé à l'extérieur de l'intervalle $[0, 2\,d]$, soit $[0, +20{,}91]$, est égale au seuil observé $p = 0{,}006$. Autrement dit, on a la probabilité fiducio-bayésienne complémentaire 0,994 que δ soit compris entre 0 et $+ 20{,}91$; on voit qu'un tel énoncé n'exprime certainement pas l'information *utile* apportée par les données, information manifestement en faveur d'un effet *important*, comme le montre la distribution fiducio-bayésienne : à partir de cette distribution, nous pouvons par exemple énoncer « avec la garantie fiducio-bayésienne 0,975, δ est supérieur à 5 cs ».

2) Considérons maintenant la situation de la comparaison C-LIN C (résultat non significatif). Le problème est ici plus délicat, puisque la comparaison a deux degrés de liberté. Il s'agit d'abord de définir un indicateur de l'écart entre les données et le modèle logarithmique. Pour cela, on peut considérer, pour chacune des quatre conditions, la différence entre la valeur observée et la valeur ajustée fournie par la droite ; on retient alors la moyenne quadratique de ces quatre différences, soit $l = ((1{,}069^2 + 2{,}125^2 + 1{,}043^2 + 0{,}013^2)/4)^{1/2} = 1{,}299$ cs. Mais il est bien clair que la valeur $l = 1{,}299$ ne nous renseigne que sur la *grandeur* de l'effet observé (écart moyen au modèle logarithmique) et, à elle seule, ne saurait résumer l'information apportée par les données sur l'effet observé lui-même. En particulier, si nous voulons figurer cet effet, celui-ci sera représenté par un *vecteur* dans le plan, où le point origine 0 correspond à un effet nul ; la longueur de ce vecteur sera proportionnelle à la grandeur de l'effet l définie précédemment.

Techniquement, pour les développements formels (cf. Lecoutre, 1984 ; Rouanet et Lecoutre, 1983), il convient de rapporter ce vecteur effet $\vec{d}$ à un système d'axes, ce qui revient à choisir deux sous-comparaisons à un degré de liberté de C-LIN C, ou de manière plus précise deux contrastes. Mais, d'une part ce choix est évidemment totalement arbitraire, et d'autre part, ce qui est primordial, les résultats qui suivent ne dépendent pas de ce choix particulier dont on pourra donc ici se passer.

Dans ce cas, la distribution fiducio-bayésienne relative à l'effet parent est encore une distribution du t de Student, centrée sur l'effet observé (l'extrémité du vecteur $\vec{d}$), mais ici de dimension 2. A partir de cette distribution, on peut attribuer une probabilité à toute région du plan. En particulier, la réinterprétation du test F, qui est illustrée dans la figure 3 et généralise celle du cas à un degré de liberté, est la suivante : si $\vec{\delta}$ est le vecteur représentant l'effet parent, la probabilité fiducio-bayé-

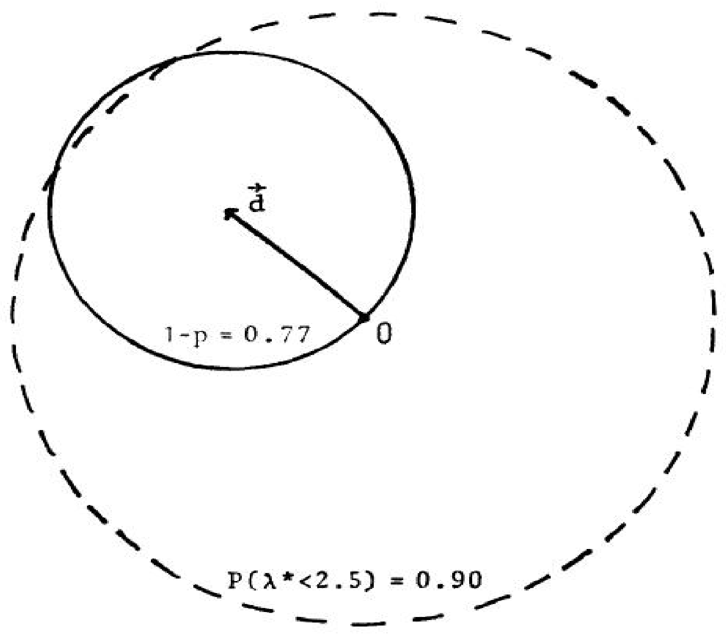

Fig. 3. — Réinterprétation fiducio-bayésienne du test *F* pour la comparaison à deux degrés de liberté C-LIN C

sienne que $\vec{\delta}$ ait son extrémité à l'extérieur du cercle centré sur l'effet observé et passant par le point origine (cercle en trait plein sur la figure 3) est égale au seuil observé $p = 0{,}23$. Autrement dit, on a la probabilité fiducio-bayésienne complémentaire 0,77 que $\vec{\delta}$ soit contenu à l'intérieur de ce cercle. Un tel énoncé ne peut *en aucun cas* être interprété en faveur d'un effet négligeable : *il n'est pas pertinent* pour cela.

Pour rechercher une conclusion utile en ce qui concerne l'acceptabilité du modèle logarithmique, il convient en fait de s'interroger sur la *grandeur* de l'effet parent associée à la comparaison C-LIN C, grandeur notée ici λ (notée $\hat{\varepsilon}$ dans Lecoutre, 1981) ; et ceci revient à attribuer une probabilité fiducio-bayésienne correspondant à un cercle centré, non pas sur l'effet observé, mais sur l'origine. On peut ainsi voir sur la figure 3 le cercle (en pointillé) qui représente l'énoncé « avec la garantie fiducio-bayésienne 0,90, λ est inférieur à 2,5 cs »[2]. Si la valeur 2,5 cs est tenue pour un écart tolérable, on pourra conclure à l'*acceptabilité* du modèle (relativement à l'analyse des données de groupe) ; si la valeur 2,5 cs est au contraire considérée trop élevée, on devra suspendre le jugement inductif, l'information expérimentale étant insuffisante pour prolonger le résultat descriptif (nous renvoyons le lecteur à l'article de 1981 pour une discussion sur la manière dont on pourrait juger du caractère « négligeable » ou « notable » de la valeur 2,5 cs, en tant qu'écart à l'hypothèse nulle $\lambda = 0$).

Bien entendu, la réinterprétation précédente se généralise aisément à des comparaisons à un nombre quelconque de degrés de liberté, les cercles étant remplacés par des sphères (pour 3 degrés de liberté), ou plus généralement par des hypersphères.

RÉSUMÉ

On illustre, à partir d'exemples concrets, la réinterprétation fiducio-bayésienne du test F *usuel de l'analyse de la variance, d'abord pour des comparaisons à un degré de liberté, puis pour des comparaisons à plusieurs degrés de liberté. Cette réinterprétation permet de comprendre les raisons de l'inadaptation des tests de signification vis-à-vis du problème de la généralisabilité des conclusions descriptives dans la recherche expérimentale.*

Mots clefs : statistique, analyse de la variance, réinterprétation fiducio-bayésienne.

RÉFÉRENCES BIBLIOGRAPHIQUES

Lecoutre (B.) — Procédures fiducio-bayésiennes pour l'investigation des mécanismes individuels en psychologie, *L'Année Psychologique*, 1981, *81*, 453-464.

Lecoutre (B.) — *L'Analyse Bayésienne des Comparaisons*, Presses Universitaires de Lille, 1984, sous presse.

Rouanet (H.), Lecoutre (B.) — Specific inference in ANOVA : from significance tests to Bayesian procedures, *British Journal of Mathematical and Statistical Psychology*, 1983, *36*, 252-268.

Rouanet (H.), Oléron (G.), Régnier (J.) — Analyse de la variance pour données appareillées et modèles de dépendance, illustration d'une démarche, *L'Année Psychologique*, 1966, *66*, 131-165.

(Accepté le 5 décembre 1983.)

2. La valeur 3,2 donnée dans l'article de 1981 pour la garantie 0,90 était en fait erronée, ainsi qu'on pouvait d'ailleurs s'en rendre compte à l'examen de la figure 4.