

NOTE THÉORIQUE

Laboratoire de Psychologie
ERA 235
Université de Paris VIII¹
Groupe de Recherche
Mathématiques et Psychologie
Université de Paris V

PROCÉDURES FIDUCIO-BAYÉSIENNES POUR L'INVESTIGATION DES MÉCANISMES INDIVIDUELS EN PSYCHOLOGIE

par Bruno LECOUTRE

SUMMARY

*Illustrates, with concrete examples, how Bayes-fiducial procedures allow the investigation of individual mechanisms by yielding inferential results, not only about the mean effect but also about individual effects. Technically, these procedures have been developed for the effect associated with a contrast and for the effect associated with a comparison (with any number of freedom) in a $S * T$ (Subjects * Treatments) design.*

L'objectif de l'analyse des données expérimentales doit être de répondre le mieux possible aux questions que se pose le chercheur. Dans la pratique de cette analyse, on constate l'existence de deux niveaux. A un premier niveau, *essentiellement descriptif*, partant des données brutes, on dérive un certain nombre de statistiques élémentaires usuelles (moyennes, écarts-types, etc.) que l'on dispose dans des tableaux ou sur des figures dont l'examen va constituer une base importante pour répondre aux questions posées. A un second niveau, on va mettre en œuvre des procédures inférentielles qui, en principe, devraient répondre aux mêmes questions, mais cette fois avec une *visée inductive* (c'est-à-dire généralisante).

Au premier niveau, l'utilisation d'une formalisation appropriée (plans $S \langle G \rangle$, $S * T$, $S \langle G \rangle * T...$, demandes d'analyse, etc.), dont on trouvera les bases dans le texte de Rouanet et Lépine (1977), fournit une aide importante ; mais la démarche apparaît à ce niveau naturelle au chercheur qui, en général, n'éprouvera pas de difficultés à interpréter les résultats obtenus vis-à-vis des questions qu'il se pose. En revanche, les procédures usuelles mises en œuvre au second niveau (traditionnellement en psychologie expérimentale les tests de signification) fournissent des résultats dont le lien avec les questions du chercheur est souvent loin d'être évident. Dans bien des cas la réponse apportée par le test de signification ne sera pertinente que si le chercheur transforme sa question en une nouvelle question dont la portée sera beaucoup moindre. Si, par exemple, il se demandait si l'effet lié à un certain facteur est important, il pourra obtenir une réponse naturelle au niveau descriptif, mais, au second niveau, il devra se contenter de la question « est-ce que cet effet existe ? » ; s'il se demandait si cet effet est important pour la plupart des individus, il devra au second niveau se contenter de la question « est-ce que l'effet moyen existe ? », etc.

Il apparaît donc nécessaire, si l'on admet le rôle essentiel de la visée inductive dans la démarche expérimentale, de rechercher et d'utiliser des procédures inférentielles qui permettent de répondre aux véritables questions posées. De ce point de vue, les méthodes bayésiennes, et en particulier les méthodes fiducio-bayésiennes qui englobent elles-mêmes les méthodes fiduciaires (cf. appendice), permettent d'ouvrir des perspectives considérables.

Nous avons, dans un précédent article, illustré le fait que ces méthodes permettent de se prononcer sur l'*importance* d'un effet, et non seulement sur son *existence* (Lecoutre et Lecoutre, 1979). Nous nous proposons ici de montrer comment elles conduisent à l'investigation des mécanismes individuels, en fournissant des résultats inférentiels sur les *effets individuels*, et non seulement sur l'*effet moyen*. Ce point a déjà été introduit par Rouanet, Lépine et Holender (1978) pour le cas particulier de la validation du modèle additif ; en fait, comme nous allons le voir, des méthodes appropriées peuvent être généralisées à un grand nombre de situations courantes.

PREMIER EXEMPLE

Prenons, à titre de premier exemple, le type d'expériences bien connu, « lecture et dénomination », dans lequel on se propose de comparer le temps mis pour lire des mots désignant des objets au temps mis pour dénommer ces objets présentés sur des figures. Les questions que l'on se pose sont les suivantes (de la plus grossière à la plus fine) :

1) « Est-ce que le temps *moyen* de dénomination est *supérieur* au temps *moyen* de lecture ? »

2) « Si oui, est-ce que le temps *moyen* de dénomination est *notablement supérieur* au temps *moyen* de lecture ? »

3) « Si la réponse est encore oui, est-ce que, dans *la plupart des cas*, le temps de dénomination est *notablement supérieur* au temps de lecture ? »

Considérons les données recueillies auprès de 16 sujets adultes ayant effectué chacun la lecture et la dénomination de 80 items (mots désignant des couleurs et pastilles colorées) ; il s'agit d'un extrait des données d'une expérience réalisée au C1 de psychologie expérimentale à l'Université de Paris V². Voyons les réponses que nous apporte, au *niveau descriptif*, une analyse très simple de ces données (voir tableau I).

TABLEAU I. — *Données de l'expérience*
« lecture et dénomination » :
temps par item en centièmes de seconde

Sujet	Lecture	Dénomination	Différence : temps de dénomination moins temps de lecture (centièmes de secondes)
1	28	41	+ 13
2	44	65	+ 21
3	44	56	+ 12
4	37	59	+ 22
5	40	60	+ 20
6	40	51	+ 11
7	34	50	+ 16
8	26	51	+ 25
9	35	69	+ 34
10	41	55	+ 14
11	33	53	+ 20
12	31	55	+ 24
13	37	62	+ 25
14	41	58	+ 17
15	38	59	+ 21
16	32	57	+ 25
Moyenne	36,3	56,3	$d = + 20,0$ $s = 6,0$

1) Le temps moyen observé pour la dénomination est effectivement plus grand que le temps moyen observé pour la lecture.

2) Le temps moyen observé pour la dénomination (56,3 cs par item) est effectivement supérieur de plus de moitié au temps moyen observé pour la lecture (36,3 cs par item). Notons $\bar{d} = 20,0$ cette différence (le symbole « \bar{d} » rappelle qu'il s'agit d'une différence moyenne).

2. L'ensemble des données est fourni par ROUANET (1979-1980), p. 100-102 ; il s'agit ici des données relatives au premier groupe de travaux dirigés.

3) La réponse à la dernière question est encore affirmative ; une première manière de l'exprimer est de remarquer que, pour chacun des seize sujets, le temps de dénomination est notablement supérieur au temps de lecture (la plus petite différence observée est de 11 cs pour le sujet 6) ; une autre manière d'exprimer cette réponse est de noter que l'écart-type-corrigé des différences observées pour chaque sujet vaut 6,0 cs, c'est-à-dire est petit par rapport à la moyenne des différences observées (20,0 cs). Notons $s = 6,0$ cet écart-type corrigé.

Examinons maintenant ces mêmes questions au *niveau inductif*. Des procédures inférentielles appropriées doivent nous permettre à ce niveau, soit de prolonger les réponses précédentes, soit de suspendre notre jugement par manque d'information expérimentale.

1) La réponse est encore affirmative ; nous pouvons conclure à l'existence d'une différence moyenne parente en utilisant le résultat du *test de signification* usuel (t de Student pour groupes appariés), très significatif : $t(15) = 13,2$.

2) La réponse est encore affirmative ; nous avons cette fois recours à une *procédure fiducio-bayésienne* qui nous permet d'énoncer : « Avec la garantie fiducio-bayésienne 0,99 la différence moyenne parente $\widehat{\delta}$ est supérieure à 16,1 cs. » Nous pouvons donc conclure que la différence moyenne parente est notable.

3) Nous avons recours à une *seconde procédure fiducio-bayésienne* qui nous permet d'énoncer : « avec la garantie fiducio-bayésienne 0,99 l'écart-type parent σ est inférieur à 10,2 cs ». 10,2 étant petit comparativement à la limite inférieure obtenue pour la différence moyenne (16,1), nous pouvons donc conclure que dans la plupart des cas l'effet est important (à titre indicatif, pour une distribution normale de moyenne 16,1 et d'écart-type 10,2, une proportion 0,90 des valeurs dépassent 3,0 et une proportion 0,80 dépassent 7,5).

D'un point de vue pratique, les procédures que nous venons de mettre en œuvre ne posent guère de difficultés. Il s'agit de traduire notre incertitude sur les valeurs des paramètres (ici la moyenne $\widehat{\delta}$ et l'écart-type σ des différences entre temps de dénomination et temps de lecture) par une distribution de probabilité sur les valeurs possibles de ces paramètres. Ces distributions sont aisément obtenues à partir de la moyenne \bar{d} et de l'écart-type corrigé s des différences observées, elles sont figurées dans la figure 1.

Les énoncés précédents sont obtenus à l'aide des tables usuelles (t de Student et khi 2). On pourrait même se montrer encore plus exigeant et, moyennant l'écriture d'un petit programme informatique, obtenir des résultats encore plus directement interprétables, notamment sur le rapport $\widehat{\delta}/\sigma$ ou sur la proportion des effets dans la population parente qui dépassent telle limite fixée.

Finalement nous pouvons obtenir aux questions 2) et 3) des réponses

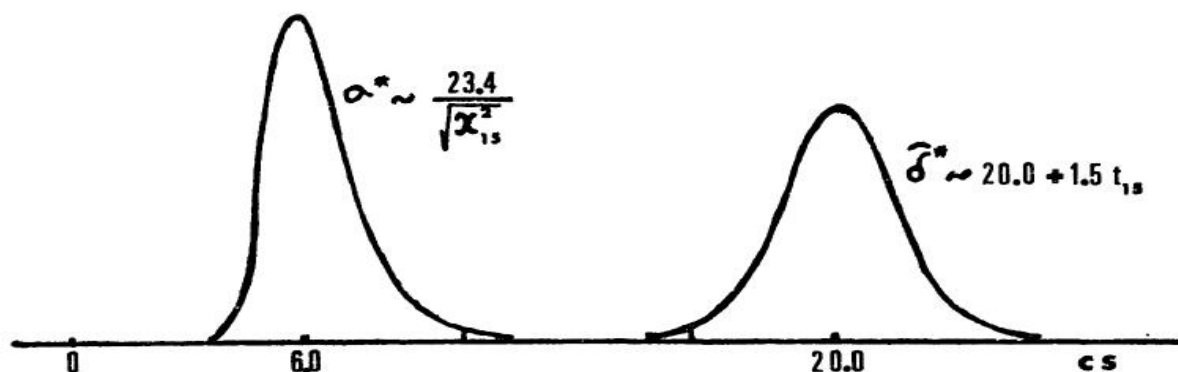


Fig. 1. — Expérience lecture et dénomination : distributions fiducio-bayésiennes relatives aux paramètres $\hat{\delta}$ et σ

que nous n'aurions pu faire que d'une manière impressionniste, avec tout le manque de rigueur et tous les risques que cela comporterait, en nous limitant au test de signification.

SECOND EXEMPLE

Nous avons volontairement choisi un premier exemple très simple pour illustrer les possibilités offertes par les méthodes fiducio-bayésiennes ; considérons maintenant un second exemple plus complexe.

Il s'agit d'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 équiprobables. Les données proviennent de l'article de Rouanet, Oléron et Régner (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. Les données relatives à ces cinq sujets, ainsi que les données de groupe obtenues en prenant pour chaque condition la moyenne des données individuelles, sont figurées dans la figure 2.

Proposons-nous d'examiner l'hypothèse suivant laquelle le temps de réaction augmente suivant une fonction linéaire du logarithme binaire du nombre de signaux possibles. A cette hypothèse, nous pouvons faire correspondre deux types de questions : 1° « Est-ce que le temps de réaction augmente d'une manière importante lorsque le nombre de signaux augmente ? » ; 2° « Est-ce que le temps de réaction est, approximativement, une fonction linéaire du nombre de signaux possibles ? ».

La figure 2 fait apparaître la droite ajustée par la méthode des moindres carrés lorsqu'on porte en abscisse le logarithme binaire du nombre de signaux. Nous pouvons prendre comme indicateur de l'effet

associé à l'augmentation du nombre de signaux la pente de cette droite ; l'effet observé d ainsi défini (pour un sujet) représente l'augmentation du temps de réaction due à la multiplication par deux du nombre de signaux, quand on remplace les données par l'approximation linéaire.

Il apparaît que l'effet observé est important pour chacun des cinq

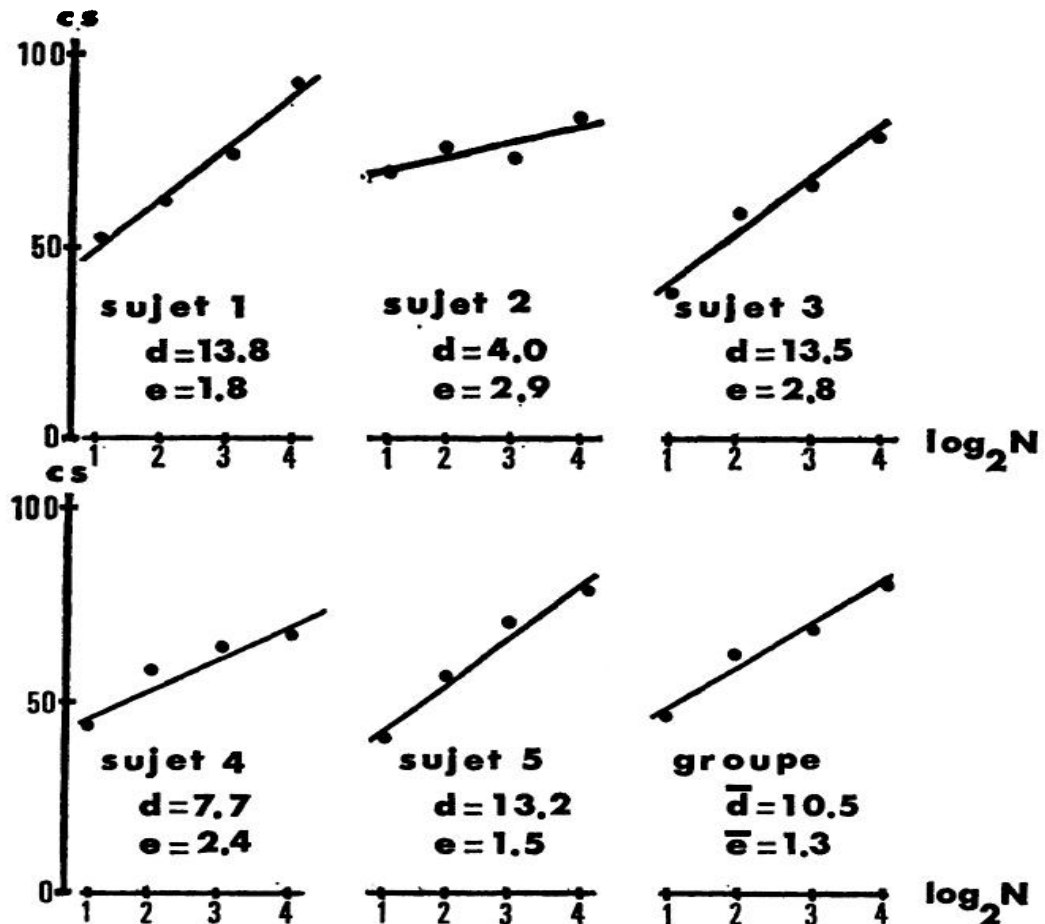


Fig. 2. — Expérience de temps de réaction de choix : figuration des données individuelles et des données de « groupe »

sujets, ce que nous pouvons encore exprimer en disant que l'effet moyen \bar{d} est important, et est en outre grand comparativement à l'écart-type corrigé des cinq pentes individuelles, le rapport \bar{d}/s valant 2,4.

Pour prolonger ces conclusions au niveau inductif, il s'agit d'effectuer une inférence sur la moyenne $\bar{\delta}$ et l'écart-type σ parents correspondants. En procédant comme dans le premier exemple, nous obtenons les distributions fiducio-bayésiennes, figurées dans la figure 3.

« Avec la garantie fiducio-bayésienne 0,90, $\bar{\delta}$ est supérieur à 7,5 cs. » 7,5 cs représentent 15 % du temps de réaction moyen observé dans la première condition (condition de base, sans incertitude sur le signal) ; nous pouvons donc considérer que la valeur 7,5 est notable. Nous

concluons par conséquent que le temps de réaction *moyen* augmente de manière importante lorsque le nombre de signaux est multiplié par deux.

Pouvons-nous conclure qu'il en est de même *dans la plupart des cas* ? L'examen de la figure 3 montre que nous ne pouvons pas ici (contrairement à l'exemple précédent) tenir σ pour petit par rapport à $\widehat{\delta}$, la distribution de σ étant très « étalée » dans la zone des valeurs probables de $\widehat{\delta}$. Ceci peut être mis en évidence plus directement par une inférence sur le rapport $\widehat{\delta}/\sigma$, qui nous permet d'énoncer « avec une garantie fiducio-bayésienne 0,90, le rapport $\widehat{\delta}/\sigma$ est supérieur à 0,7 » ; à titre de

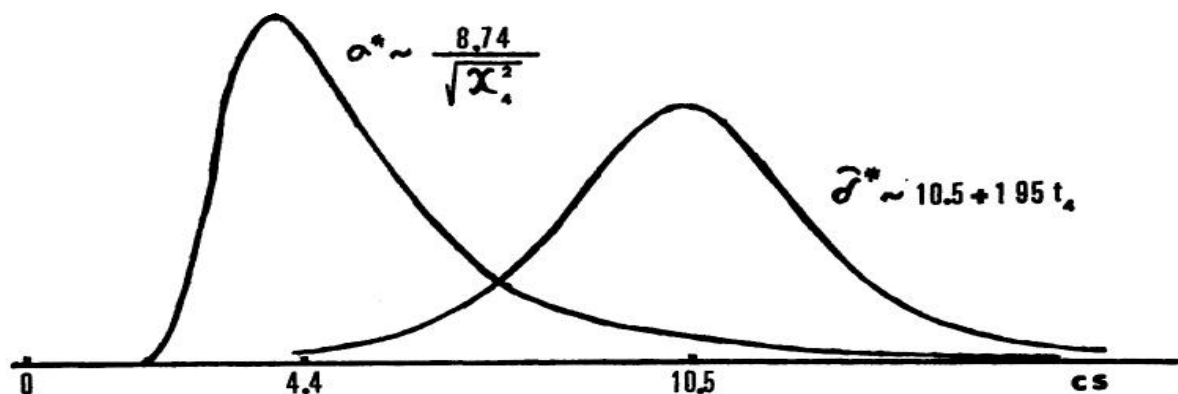


Fig. 3. — Expérience de temps de réaction de choix : distributions fiducio-bayésiennes relatives aux paramètres $\widehat{\delta}$ et σ

comparaison la même inférence effectuée pour l'exemple « lecture et dénomination » donne pour $\widehat{\delta}/\sigma$ une limite inférieure de 2,4. En dépit de la valeur élevée observée pour le rapport \bar{d}/s , l'information apportée par les données est insuffisante pour permettre de conclure ici que le temps de réaction augmente de manière importante dans la plupart des cas.

Le même type d'analyse peut encore être effectué en ce qui concerne la seconde question, c'est-à-dire l'acceptabilité de l'ajustement. Il s'agit d'abord de définir un indicateur de l'écart entre les données et le modèle logarithmique ; la solution n'est pas évidente dans ce cas car il s'agit d'une comparaison à plusieurs (en fait deux) degrés de liberté. Nous pouvons retenir la solution suivante. Si nous voulions caractériser cet écart pour une seule condition, nous prendrions naturellement, pour chaque sujet, la différence entre la valeur observée et la valeur théorique fournie par la droite, et nous serions ainsi amenés à retenir la moyenne et l'écart-type corrigé de ces différences pour chaque sujet. Si nous considérons maintenant l'ensemble des quatre conditions, il est raisonnable de retenir une moyenne de ces moyennes et de ces écarts-types corrigés obtenus pour chaque condition (voir tableau II) ; pour des

raisons qui sortiraient du cadre de cet exposé, il convient de prendre la moyenne quadratique, et nous obtenons ainsi les écarts individuels e , dont la distribution peut être caractérisée par les valeurs \bar{e} et s . Il faut insister ici sur le fait que \bar{e} n'est plus la moyenne des écarts individuels ; dans cet exemple \bar{e} est plus petit que chacun des écarts individuels, ce qui traduit bien le fait que l'ajustement est meilleur pour les données de groupe que pour les données de chaque sujet (cf. figure 2). Au niveau descriptif, le fait que \bar{e} soit petit ne peut même pas, à lui seul, être utilisé comme argument en faveur de l'acceptabilité du modèle logarithmique ; celui-ci ne peut être acceptable que si à la fois \bar{e} et s sont tenus pour négligeables. Il en sera bien entendu de même au niveau inductif, où il s'agira de montrer que les paramètres correspondants, respectivement $\hat{\varepsilon}$ et σ , peuvent être considérés comme négligeables pour pouvoir conclure à l'acceptabilité du modèle logarithmique.

TABLEAU II. — *Caractérisation de l'écart entre les données et le modèle logarithmique : tableau des différences entre la valeur observée et la valeur théorique pour chaque sujet et pour chaque condition (N désigne le nombre de signaux possibles)*

	Condition				Moyenne quadratique (en centièmes de seconde)
	N = 1	N = 2	N = 4	N = 8	
Sujet 1	+ 1,40	— 0,78	— 2,62	+ 2,01	$e = 1,84$
Sujet 2	+ 0,01	+ 2,33	— 4,68	+ 2,35	$e = 2,87$
Sujet 3	— 2,80	+ 4,75	— 1,11	— 0,84	$e = 2,85$
Sujet 4	— 2,64	+ 3,26	+ 1,40	— 2,02	$e = 2,43$
Sujet 5	— 1,31	+ 1,07	+ 1,81	— 1,56	$e = 1,46$
Moyenne	— 1,07	+ 2,13	— 1,04	— 0,01	$\bar{e} = 1,30$
Ecart-type corrigé	1,79	2,11	2,73	2,05	$s = 2,20$
Groupe	— 1,07	+ 2,13	— 1,04	— 0,01	$\bar{e} = 1,30$

Les distributions fiducio-bayésiennes relatives à $\hat{\varepsilon}$ et à σ sont figurées dans la figure 4.

Nous pouvons ici énoncer : 1) « Avec la garantie fiducio-bayésienne 0,90, $\hat{\varepsilon}$ est inférieur à 3,2 » ; 2) « Avec la garantie fiducio-bayésienne 0,90, σ est inférieur à 3,3. » Comment juger du caractère notable ou négligeable de ces limites ? Il nous paraît pertinent de considérer un autre modèle à titre de référence : nous envisagerons ici le modèle linéaire « le temps de réaction est une fonction linéaire du nombre de signaux possibles ». Cela ne signifie pas ici que nous considérons ce modèle comme un

concurrent réaliste du modèle logarithmique ; mais, précisément parce que ce modèle linéaire fournit un ajustement réellement très médiocre au niveau descriptif, il nous fournit un critère d'acceptabilité, ou plutôt de non-acceptabilité. En procédant de la même manière que pour le modèle logarithmique, nous trouvons pour le modèle linéaire les valeurs $\bar{\varepsilon} = 3,86$ et $s = 2,73$. Comparativement, les limites affectées à $\bar{\varepsilon}$ (3,2) et à σ (3,3) pour le modèle logarithmique peuvent difficilement être tenues pour négligeables, surtout en ce qui concerne σ . Les données que nous avons analysées ne nous permettent donc pas de conclure

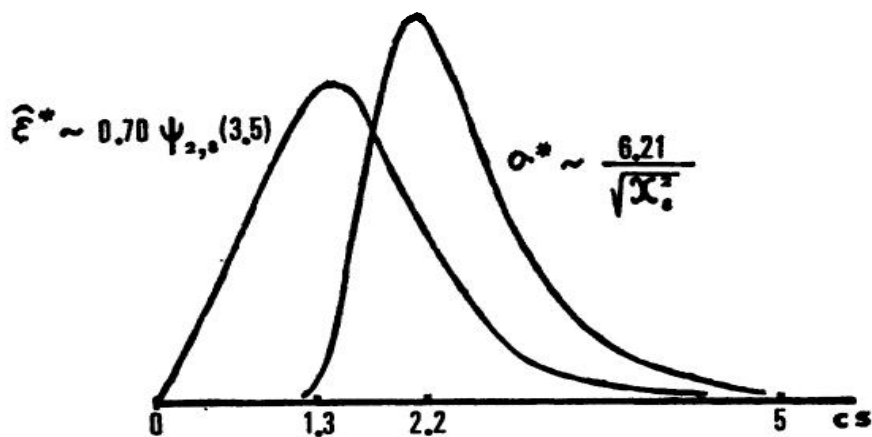


Fig. 4. — Expérience de temps de réaction de choix : distributions fiducio-bayésiennes relatives aux paramètres $\bar{\varepsilon}$ et σ

raisonnablement à l'acceptabilité du modèle logarithmique, conclusion à laquelle une analyse limitée aux données de groupe et à l'utilisation du test de signification aurait pu donner l'illusion d'aboutir.

APPENDICE

Les méthodes d'inférence permettant de traduire notre incertitude sur la valeur vraie d'un paramètre par une distribution de probabilité sur les valeurs possibles de ce paramètre se rattachent à deux grandes classes :

1) Les *méthodes bayésiennes*, qui font intervenir explicitement un élément extérieur aux données à analyser, à savoir une *distribution initiale* sur les valeurs possibles du (ou des) paramètre ; la distribution cherchée est la *distribution finale* (ou révisée), qui est déduite du modèle d'échantillonnage et de cette distribution initiale, par le théorème de Bayes ;

2) Les *méthodes fiduciales*, qui, techniquement, ne font pas intervenir d'élément extérieur aux données à analyser (hormis le modèle d'échantillonnage) ; la distribution fiduciale est obtenue par un argument direct, l'argument du « pivot » (cf. notamment Fisher, 1959), à partir du modèle d'échantillonnage ; ces méthodes supposent cependant

implicitement de se placer dans un état d'ignorance initial vis-à-vis du (ou des) paramètre.

Nous laisserons ici de côté la question de savoir si, en ce qui concerne les fondements conceptuels, les méthodes fiduciaires doivent bénéficier d'un statut autonome ou, au contraire, doivent être ramenées à la problématique bayésienne. Nous nous contenterons de remarquer que, dans les situations courantes d'analyse de données expérimentales, la distribution fiduciaire peut également être obtenue dans le cadre des méthodes bayésiennes, en choisissant une distribution initiale traduisant explicitement un état d'ignorance, une telle distribution ainsi que les procédures sous-jacentes étant souvent qualifiées de *non-informatives* (cf. notamment Jeffreys, 1961). Etant donné que, dans ces situations, procédures fiduciaires et procédures bayésiennes non-informatives³ aboutissent aux mêmes résultats, et qu'il est donc équivalent, sur le plan pratique, d'utiliser les unes ou les autres, nous les qualifions indistinctement, du point de vue technique, de *procédures fiduciaire-bayésiennes*.

Nous avons par ailleurs (Lecoutre, 1980) développé d'une manière systématique l'étude des procédures fiduciaire-bayésiennes (et également de procédures bayésiennes *au sens strict*), pour une large classe de plans d'analyse, recouvrant la plupart des situations courantes. Ces procédures englobent et généralisent les procédures précédemment proposées, soit dans le cadre fiduciaire (notamment Lépine et Rouanet, 1975 ; Lecoutre, 1978), soit dans le cadre bayésien non-informatif (en particulier Lindley, 1965 ; Box et Tiao, 1973).

En ce qui concerne le problème particulier de l'investigation des mécanismes individuels, celui-ci se posera typiquement à propos d'un plan du type $S * T$ (pour chaque sujet on a une observation pour chaque traitement, le facteur « Traitements » T pouvant être composé de différents facteurs systématiques élémentaires), ou d'un plan du type $R \langle S * T \rangle$ (pour chaque sujet on a plusieurs observations pour chaque traitement, d'où l'introduction du facteur *Répétitions* R). Les plans du type $R \langle S * T \rangle$ présentent l'intérêt supplémentaire de pouvoir faire la part entre différences interindividuelles et différences intra-individuelles ; mais leur mise en œuvre peut être difficile, voire impossible si la tâche ne peut pas être répétée pour un même sujet, et les procédures d'analyses, que nous avons développées (Lecoutre, 1980), sont relativement complexes. En revanche les plans du type $S * T$, qui peuvent se présenter, soit directement en tant que plans d'expérience, soit, souvent, en tant que plans de protocoles dérivés, conduisent à des procédures simples³.

3. Ces procédures sont résumées dans le document ERA 235, n° 143, disponible au Laboratoire de Psychologie, Université de Paris VIII, 2, rue de la Liberté, 93526 Saint-Denis, Cedex 02.

RÉSUMÉ

*Cet article montre, à partir d'exemples concrets, comment les procédures fiducio-bayésiennes permettent l'investigation des mécanismes individuels, en fournissant des résultats inférentiels, non seulement sur l'effet moyen, mais aussi sur les effets individuels. Techniquement, ces procédures sont développées pour l'effet associé à un contraste et pour l'effet associé à une comparaison (à un nombre quelconque de degrés de liberté) dans un plan du type $S * T$ (Sujets * Traitements).*

RÉFÉRENCES BIBLIOGRAPHIQUES

- BOX (G. E. P.), TIAO (G. C.) — *Bayesian inference in statistical analysis*, Wesley Publishing Company, 1973.
- FISHER (R. A.) — *Statistical methods and scientific inference*, Edinburgh and London, Oliver & Boyd, 2^e éd., 1959, (1^{re} éd. en 1956).
- JEFFREYS (H.) — *Theory of probability*, Oxford University Press, 3^e éd., 1961, (1^{re} éd. en 1939).
- LECOUTRE (B.) — Note sur le calcul de la distribution fiduciaire pour une inférence sur un contraste entre moyennes, *Cahiers de Psychologie*, 1978, 21, 279-282.
- LECOUTRE (B.) — *Extensions bayésiennes de l'analyse de la variance*, thèse de doctorat de troisième cycle Mathématiques et Applications, Université René-Descartes, Paris, 1980.
- LECOUTRE (B.), LECOUTRE (M.-P.) — A propos d'une expérience d'apprentissage perceptif incident : quelques aspects de la démarche d'analyse des données et méthodes fiduciaires, *Psychologie française*, 1979, 24, 269-278.
- LÉPINE (D.), ROUANET (H.) — Introduction aux méthodes fiduciaires : inférence sur un contraste entre moyennes, *Cahiers de Psychologie*, 1975, 18, 193-218.
- LINDLEY (D. V.) — *Introduction to probability and statistics from a bayesian viewpoint, Part 2 — Inference*, Cambridge University Press, 1965.
- ROUANET (H.) — *C1 de Psychologie générale. Documents pour l'enseignement de statistique : procédures statistiques fondamentales et initiation à l'usage de l'ordinateur*, texte photocopé, Université René-Descartes, Paris, 1979-1980.
- ROUANET (H.), LÉPINE (D.) — Introduction à l'analyse des comparaisons pour le traitement des données expérimentales, *Informatique et Sciences humaines*, 1977, 33-34, numéro spécial, 125 p.
- ROUANET (H.), LÉPINE (D.), HOLENDER (D.) — Model acceptability and the use of Bayesfiducial methods for validating models, in J. REQUIN (ed.), *Attention and Performance VII*, Hillsdale, Lawrence Erlbaum Associates, 1978, 687-701.
- ROUANET (H.), OLÉRON (G.), RÉGNIER (J.) — Analyse de variance pour données appariées et modèles de dépendance, illustration d'une démarche, *L'Année psychologique*, 1966, 66, 131-165.

(Accepté le 19 juin 1981.)