



**HAL**  
open science

## Sur quelques points de la théorie des tests

Georges Morlat

► **To cite this version:**

Georges Morlat. Sur quelques points de la théorie des tests. Annales de l'ISUP, 1959, VIII (4), pp.253-266. hal-04094678

**HAL Id: hal-04094678**

**<https://hal.science/hal-04094678>**

Submitted on 11 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## SUR QUELQUES POINTS DE LA THÉORIE DES TESTS<sup>(1)</sup>

Georges MORLAT

Il s'agit des principes logiques qui gouvernent le choix des tests, et non pas des techniques de calcul, des lois d'échantillonnage sur lesquelles se fondent les tests, ou des tests divers applicables à telle ou telle situation. Là dessus, existe une littérature extrêmement abondante, et qui s'enrichit constamment. Nous nous en tiendrons ici aux principes, ou à la "Théorie générale des Tests", sur laquelle ont été exprimés par divers auteurs des points de vue en apparence contradictoires (Fisher, Neyman, Jeffreys, etc.).

°  
° °

Faut-il parler de "tests d'hypothèses"- de "tests de signification, (des paramètres)" - ou de "tests" tout court ?

Dans son ouvrage [10], imprégné du point de vue en vogue, consistant à rattacher la statistique et quelques autres sciences à la théorie de la Décision - tendances behavioristes" ou "opérationnelles"! - Schlaifer écrit que l'expression "test d'hypothèse" n'est qu'un "résidu historique". Ce jugement s'explique par la considération exclusive de problèmes économiques, mais certains statisticiens persistent à utiliser des tests dans le domaine de la recherche scientifique, et leurs travaux s'en portent bien. Dans ce domaine, ainsi que le rappelle Fisher [4], il est déraisonnable de vouloir énoncer tous les problèmes en termes économiques, car :

a) une décision est toujours provisoire;

b) il est parfaitement irréaliste d'imaginer une longue suite de décisions semblables;

c) et surtout, les coûts des erreurs sont tout à fait dépourvus de sens, car on ne peut imaginer d'avance à quoi servira un progrès scientifique.

-----  
(1) Exposé présenté au Séminaire de Statistique, le 3 novembre 1959.

Il semble donc que la théorie des tests doive répondre à des problèmes variés. Tests d'hypothèses ou règles de décision économique ne doivent pas nécessairement obéir aux mêmes principes. Mais parce que certaines situations concrètes tiennent à la fois du choix d'une hypothèse scientifique satisfaisante et de la recherche d'un optimum économique, il ne semble pas que la théorie doive subir une scission complète : d'ailleurs, les "techniques" des tests utilisés sont souvent les mêmes.

#### Décisions statistiques, estimation et tests.

Si l'on a en vue les seuls problèmes susceptibles d'une interprétation économique, peut-on rattacher la théorie des tests à la théorie des fonctions de décision statistique, développée par Wald [6] ? Certes, il existe entre la théorie de Neyman et les travaux de Wald une parenté intellectuelle, et dans l'ouvrage cité, Wald consacre quelques pages à montrer qu'un problème classique de test peut être facilement énoncé dans le langage des fonctions de décision statistique; cependant, on ne peut déduire des principes admis par Wald, les conventions de Neyman pour définir un test optimum. Cela tient à ce que la théorie de Neyman évite, au prix de quelques conventions, de faire intervenir effectivement les coûts des erreurs. S'il vaut mieux admettre des conventions supplémentaires ou apprécier des coûts dans des conditions discutables, c'est une autre affaire, et nous y reviendrons. Bornons-nous pour l'instant à constater que la théorie des tests, telle qu'elle existe sous ses différentes formes, ne saurait être regardée comme un cas particulier de la théorie de Wald, sans perdre à la fois ses traits spécifiques et le plus clair de son efficacité.

La théorie des tests et la théorie de l'estimation doivent-elles demeurer distinctes ? On a songé aussi à les confondre sous le vocable de "décisions statistiques". S'il s'agit bien, dans les deux théories, de décider de quelque chose - et pas nécessairement d'une action concrète, mais plus souvent de la manière de conduire des calculs ultérieurs, voire simplement d'une opinion à mettre en réserve pour plus tard - il nous semble plus fécond ici d'opposer que de confondre. Car la théorie de l'estimation vise à restreindre une hypothèse, en attribuant, selon les observations, une valeur numérique précise à un paramètre resté jusqu'alors indéterminé, tandis que les tests ont pour objet de remettre en cause une hypothèse jusqu'ici admise, peut-être en l'élargissant, en y introduisant un nouveau paramètre, voire en la démolissant tout à fait, si cela est nécessaire. Les deux théories ont donc des objectifs opposés.

Bien sûr, à côté de l'estimation ponctuelle, on a développé des méthodes d'estimation par intervalles (notamment la théorie des intervalles de confiance de J. Neyman), et ces méthodes ressemblent

diablement à certains tests. Il y a tout un arsenal commun de calculs, entre les intervalles de confiance portant sur la moyenne d'une loi normale d'écart-type inconnu, et le test de Student, par exemple. Mais il faut considérer à quoi servent finalement ces méthodes. Une estimation par intervalle n'est guère satisfaisante que vue comme un stade intermédiaire, l'objectif restant de préciser entièrement le paramètre. Et lorsque le statisticien doit éclairer une décision concrète (car cela arrive, bien sûr, assez souvent) alors un intervalle sera souvent inacceptable. Comme le remarquait E. Halphen, l'ingénieur ne sera pas très content si le statisticien lui conseille de donner à un barrage une hauteur comprise entre  $h_1$  et  $h_2$ . Il doit construire un barrage, une fois pour toutes, et souhaite qu'on lui conseille une solution déterminée.

A l'inverse, on a parfois envisagé, dans la théorie des tests, le cas où l'hypothèse alternative est plus restrictive que l'hypothèse testée [8]. Nous connaissons quelques exemples concrets dans lesquels on est effectivement tenté de poser un problème de test en ces termes; mais à y regarder de plus près, il nous semble que ces problèmes-là doivent être décomposés en un problème d'estimation, puis un problème de test.

Nous poserons donc en principe que les théories de l'estimation et des tests, si elles peuvent être au départ rattachées toutes deux à des règles de décision plus générales, doivent cependant être développées de manière distincte, et que :

- l'estimation vise à restreindre une hypothèse;
- un test vise au contraire à décider s'il faut élargir une hypothèse.

#### La théorie de Neyman.

La théorie de Neyman et Pearson [5] est le plus généralement admise pour justifier les tests classiques. Elles permettent, comme on sait, d'éviter tout recours à la "probabilité des hypothèses" - et fonde les raisonnements exclusivement sur les lois conditionnelles des observations (conditionnel voulant dire ici : pour une hypothèse déterminée). On est ainsi conduit à la considération des risques d'erreur des deux espèces classiques; - dont le tableau ci-dessous résume les définitions :

$H_0$  : hypothèse nulle admise antérieurement, et mise à l'épreuve ;

$H_1$  : hypothèse alternative qui remplacera  $H_0$  si les observations la controuvent.

Hypothèse "vraie"

		$H_0$	$H_1$
Hypothèse retenue par le statisticien	$H_0$	Bon	Erreur de 2ème espèce Probabilité $\beta$
	$H_1$	Erreur de 1ère espèce Probabilité $\alpha$	Bon

Cette façon de poser le problème des tests possède le grand mérite de mettre en lumière la nécessité absolue de formuler l'hypothèse alternative - qui peut d'ailleurs comporter des paramètres indéterminés plus nombreux que l'hypothèse nulle. Il ne semble pas qu'aucune théorie puisse se passer de cette exigence, et surtout pas la théorie de Bayes, dont nous parlerons un peu plus loin. L'idée qu'une hypothèse soit "vraie" et une autre "fausse" peut paraître à certains un peu simple, mais il convient sans doute de voir là un langage abrégé et commode.

Par ailleurs, la considération exclusive des deux risques d'erreur de Neyman - qui équivaut exactement à l'exclusion des probabilités des hypothèses - est certainement raisonnable dans d'assez nombreux problèmes, mais nous ne pensons pas que toutes les situations où l'on parle habituellement de tests se ramènent à ce schéma. C'est pourquoi nous examinerons plus loin, sous le titre de "théorie de Bayes", les méthodes pronées par Jeffreys, Maurice Dumas et d'autres.

Mais, restons pour l'instant dans le cadre de la théorie de Neyman. On sait que le choix d'un test optimum présente peu d'arbitraire lorsque les hypothèses  $H_0$  et  $H_1$  sont toutes deux simples (c'est-à-dire qu'elles ne contiennent aucun paramètre indéterminé). La seule convention consiste à attribuer d'office une valeur (ou une borne supérieure) à la probabilité  $\alpha$ . Cette valeur (faible en général) est le seuil de confiance. Le meilleur test est alors celui qui minimise  $\beta$ , tout le monde sera sans doute d'accord là dessus.

On a souvent critiqué cette façon dissymétrique de traiter les deux espèces d'erreurs. La justification classique se réfère à la nature des coûts inhérents à ces erreurs. On dira par exemple qu'une erreur de première espèce peut être grave, tandis qu'une erreur de seconde espèce n'est qu'une perte monétaire banale, s'il s'agit de

tester la composition d'un produit pharmaceutique présentant des dangers pour le client lorsqu'on laisse passer un échantillon défectueux (c'est l'erreur de première espèce). A notre sens cette justification est fort particulière et ne s'applique qu'à une classe restreinte de problèmes. Mais nous pensons que le traitement dissymétrique des deux espèces d'erreurs est beaucoup mieux justifié par le fait que l'hypothèse nulle est privilégiée, correspond à une explication de structure, peut-être grossière, des phénomènes étudiés, et en tout cas avait été jusqu'ici admise, et non pas gratuitement sans doute. Au contraire, l'alternative n'est souvent qu'un pis aller. C'est une manière faible d'introduire la préférence du statisticien pour ses hypothèses - beaucoup plus faible, cela va de soi, que les probabilités a priori de la formule de Bayes.

Mais enfin ... chassez l'intuition, elle est revenue ... à pas de loup.

C'est donc dans le choix du seuil de confiance que nous la retrouvons (et alors nous ne dirons plus que ce choix est arbitraire).

(Notons au passage que même si les coûts des deux espèces d'erreur, étaient jugés égaux, il serait complètement dépourvu de sens de vouloir minimiser la somme  $\alpha + \beta$ , et que la minimation d'une combinaison pondérée de ces deux risques reviendrait très précisément à introduire les probabilités des hypothèses que cette théorie se propose d'exclure).

Cela dit, faut-il garder aveuglément la règle proposée par Neyman : fixer  $\alpha$  et minimiser  $\beta$  ? Considérons l'exemple suivant :

$H_0$  : des observations proviennent d'une loi normale (0, 1)

$H_1$  : elles proviennent d'une loi normale (1, 1)

La moyenne arithmétique  $\bar{x}$ , obéit à une loi normale d'écart-type  $\frac{1}{\sqrt{n}}$  dans les deux hypothèses, et de moyenne égale à 0 dans l'hypothèse nulle, à 1 dans l'hypothèse alternative.

Si on se donne le seuil de confiance du test, par exemple 1%, le risque d'erreur de seconde espèce sera très voisin de zéro si le nombre d'observations  $n$  est très grand. La puissance du test sera donc voisine de 1.

On peut facilement déterminer  $n$  pour que la puissance soit, par exemple, égale à  $1 - 10^{-6}$ . (il suffit pour cela qu'on ait  $n = 50$ ).

Un statisticien superficiel (s'il en existe) trouverait peut-être que

cette situation est très bonne. Mais en vérité, si l'on a trouvé avec 50 observations  $\bar{x} = 0,35$ , notre test de seuil 1% nous conduirait à rejeter  $H_0$  (qui rend l'échantillon assez improbable) au profit de  $H_1$  (qui rend l'échantillon infiniment plus improbable).

La leçon à retenir, c'est que les tests très puissants, ne sont pas satisfaisants, ou encore que le risque d'erreur de seconde espèce doit être minimum, mais pas trop petit!

Il est vrai qu'un test entre deux hypothèses simples est un problème très particulier - souvent peu réaliste. Le test d'une hypothèse simple contre une hypothèse paramétrique est beaucoup plus courant.

Dans ce cas, hormis l'existence d'un test U M P (circonstance heureuse, mais exceptionnelle) on est conduit à choisir un test en vertu d'une règle conventionnelle, choisie dans un arsenal assez décousu.

Un excellent inventaire de cet arsenal se trouve dans un mémoire de Lehman [8].

La règle la plus connue, consiste à retenir comme optimum un test défini par une région de rejet constituée par la réunion des régions critiques des tests optimaux relatifs aux diverses hypothèses simples constituant  $H_1$ , pour un rapport de densité  $\lambda = \frac{L_0}{L_1}$  constant.

C'est ce que les anglo-saxons ont coutume de nommer "likelihood ratio test", et que nous désignerons d'une manière plus abrégée par "test de type  $\lambda$ ". Cette règle permet de justifier la plupart des tests couramment utilisés, mais divers exemples ont été donnés, notamment par Stein [8] dans lesquels un test de type  $\lambda$  apparaît comme un très mauvais test.

Donnons un exemple beaucoup moins perfectionné que celui de Stein, mais peut-être plus suggestif : il s'agit d'une observation unique, susceptible de prendre 3 valeurs a, b, c. L'hypothèse  $H_0$  est simple, l'hypothèse  $H_1$  est formée des 2 hypothèses simples  $H_1'$  et  $H_1''$ . Le tableau ci-dessous donne les lois de probabilité de l'observation.

	a	b	c
$H_0$	.05	.05	.90
$H_1'$	0	.08	.92
$H_1''$	.08	0	.92

Si l'on veut tester  $H_0$  au seuil 10%, le test de type  $\lambda$  nous donne pour région critique  $W = \{a, b\}$ . Le risque d'erreur de seconde espèce pour ce test est égal à .92.

Si l'on choisissait pour région critique  $W = \{c : 1/9\}$  - c'est-à-dire si l'on effectuait un tirage au sort avec une probabilité de 1/9 de rejeter  $H_0$ , lorsqu'on a obtenu  $c$ , alors on aurait également un test de seuil 10%, pour lequel l'erreur de seconde espèce serait égale à .83. Il est donc certainement meilleur que le test de type  $\lambda$ .

Ce genre de mésaventure ne semble guère à craindre pour la plupart des tests présentés dans les traités classiques; il souligne seulement le caractère conventionnel des principes proposés par Neyman. Il serait intéressant de pouvoir joindre à ces principes quelque restriction garantissant qu'ils conduisent à un test admissible. (C'est-à-dire tel qu'il n'existe aucun test de même seuil et de puissance supérieure).

#### La théorie de Bayes.

Si l'on met à part l'idée audacieuse de l'équiprobabilité représentant l'ignorance, on peut bien dire que Bayes [1] avait résolu à sa façon les problèmes de tests, comme toutes les décisions statistiques.

La probabilité des hypothèses (qui implique une interprétation "subjectiviste" de la notion de probabilité) a été considérée durant quelques décennies comme pelée, galeuse... et dangereuse - et son éviction a permis, il faut le dire, de brillants développements de la statistique moderne [5]. [6].

Mais un mouvement en sens contraire semble maintenant s'amplifier. En vérité l'attitude "subjectiviste" n'a jamais cessé d'avoir des adeptes, et parmi les plus notables il faut citer de Finetti et Jeffreys [12]. Halphen [11] avait noté qu'on sait toujours quelque chose concernant un fait incertain, et qu'on ne peut se passer de cette connaissance imparfaite (vraisemblance). Cependant les pages de son mémoire relatives à la théorie des tests sont parmi les moins fermes.

L.J. Savage a achevé entre ses deux ouvrages [9] - celui de 1954 et celui qu'il prépare - une conversion complète. L'ouvrage de Schlaifer [10] présente d'emblée un point de vue bayésien, qu'il oppose, comme "moderne", à la théorie "classique" de Neyman. Comme symbole, signalons qu'il ne désigne pas autrement que par "basic random variables" les paramètres inconnus d'une loi de probabilité.

Les principes de la méthode de Bayes sont fort simples.



Si l'on donne des probabilités a priori aux hypothèses envisagées, la connaissance des observations permet de calculer des probabilités conditionnelles (ou a posteriori) d'après lesquelles on jugera de l'hypothèse à retenir. Il n'y aurait pratiquement pas lieu de parler de théorie des tests, n'était la recherche des probabilités a priori qui justifient les tests couramment utilisés [12] - et la confrontation avec les autres théories, seul point sur lequel nous porterons ici l'attention. Lorsqu'il s'agit de choisir entre deux hypothèses simples, relatives par exemple à la densité de probabilité d'une variable continue, on aura donc :

Hypothèses	Densité	Probabilité a priori
$H_0$	$f_0(x)$	$p_0$
$H_1$	$f_1(x)$	$p_1$

et les probabilités a posteriori seront :

$$P_0 = \frac{p_0 f_0(x)}{p_0 f_0(x) + p_1 f_1(x)}$$

$$P_1 = \frac{p_1 f_1(x)}{p_0 f_0(x) + p_1 f_1(x)}$$

La région d'acceptation (de  $H_0$ ) sera défini par  $\frac{P_0}{P_1} > k$ , soit

$$\frac{p_0}{p_1} \frac{f_0(x)}{f_1(x)} > k$$

On constate que cette condition est formellement identique à celle qui définit la région critique de Neyman, mais le seuil de signification du test dépendra, naturellement, de  $\frac{p_0}{p_1}$ . Par ailleurs, si les probabilités a priori  $p_0$  et  $p_1$  représentent intégralement les "attitudes" à l'égard des hypothèses  $H_0$  et  $H_1$ , celles-ci doivent être traitées symétriquement, il faut prendre  $k = 1$ . Mais ce n'est sans doute pas toujours le cas.

Le choix des valeurs de  $p_0$ ,  $p_1$  et  $k$  est éminemment subjectif; il ne convient pas de rechercher des règles standard pour fixer ces valeurs selon la forme analytique du problème, comme on l'a tenté parfois.

Si maintenant l'hypothèse  $H_1$  est paramétrique ( $H_0$  correspond, comme c'est généralement le cas, à une valeur particulière du paramètre) alors il faut distinguer deux types de situations.

Dans un premier type -  $H_0$  non privilégiée - la distribution a priori du paramètre  $\vartheta$  est représentée par une densité continue  $\varphi(\vartheta)$  - et la distribution a posteriori sera :

$$\varphi(\vartheta/x) = \frac{\varphi(\vartheta)f_{\theta}(x)}{\int \varphi(\vartheta)f_{\theta}(x)d\vartheta}$$

Si dans l'hypothèse  $H_0$  on suppose que  $\vartheta = \vartheta_0$ , le test consistera à examiner si la distribution a posteriori que nous venons d'écrire est suffisamment concentrée autour de  $\vartheta_0$ . Reste à définir quelque critère pour en juger.

Le second type -  $H_0$  privilégiée - représente peut-être de manière plus réaliste l'attitude mentale du statisticien vis-à-vis des hypothèses dans beaucoup de situations concrètes. On donnera une probabilité a priori finie  $p_0$  à l'hypothèse nulle, et on répartira le complément  $p_1$  suivant une courbe de densité convenable  $\varphi(\vartheta)$ .

Alors, la probabilité a posteriori de  $H_0$  sera :

$$P = \text{Prob. } (H_0/x) = \frac{p_0 f_0(x)}{p_0 f_0(x) + p_1 \int f_{\theta}(x) \varphi(\vartheta) d\vartheta}$$

et la règle pourra être de retenir l'hypothèse  $H_0$  si la valeur de  $P$  est au moins égale à 90% par exemple.

Ce schéma a été utilisé par Lindley [13] pour une confrontation instructive entre la méthode de Neyman et la méthode de Bayes. Mais il a donné lieu à peu de développements d'ordre général, en dépit du caractère réaliste qu'il nous semble bien posséder.

Revenant au premier type -  $H_0$  non privilégiée - il est intéressant de signaler que ce schéma permet de justifier dans certains cas la méthode de Fisher (probabilités fiduciaires) : la probabilité fiduciaire est la limite de la probabilité conditionnelle du paramètre (les observations étant données) lorsque sa probabilité a priori devient très "dispersée". Les termes de cette proposition seraient à préciser, et nous nous bornerons à donner, avec Schlaifer [10] l'exemple particulièrement simple de la moyenne d'une loi normale d'écart-type connu. Si un échantillon a une moyenne  $\bar{x}$ , dont l'écart-type sera noté  $\sigma$  pour alléger les formules, et si la distribution a priori de  $m$  est une loi normale ( $m_0, \sigma_0$ ), alors la distribution a posteriori de  $m$  sera une loi normale, de moyenne

$$m = \frac{\frac{m_0}{\sigma_0^2} + \frac{\bar{x}}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}} \text{ et d'cart-type } \sigma_1 = \sqrt{\frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}}}$$

On voit que si l'on suppose  $\sigma_0$  très grand, cette loi est approximativement une loi normale de moyenne  $\bar{x}$  et d'écart-type  $\sigma$  : c'est la loi de probabilité fiduciaire de Fisher.

### Le test du $\chi^2$ et la théorie de Neyman.

Le test du  $\chi^2$ , qui est la première pierre de la statistique moderne [2] a eu un destin assez curieux. D'une part, dans l'emploi le plus courant (test d'ajustement) on lui donne un sens un peu différent de ce que Karl Pearson avait en vue : le titre de son mémoire montre clairement que pour lui, la chose à mettre à l'épreuve était le tirage au hasard - "... that it can be reasonably supposed to have arisen from random sampling." - tandis qu'on admet souvent que la question traitée est plutôt : "cette urne-là ou bien une autre ?", l'hypothèse de tirages au hasard n'étant point en cause.

D'autre part, si l'on retrouve la loi de  $\chi^2$  bien souvent - et en particulier comme distribution asymptotique générale dans les tests entre hypothèses paramétriques selon Neyman - le problème traité par Karl Pearson ne semble pas avoir été revendiqué par les auteurs de la "théorie générale" : le test du  $\chi^2$  est-il "le meilleur" test d'ajustement ? Nous ne trouvons pas la réponse à cette question dans les traités classiques.

Il est cependant bien facile de s'assurer que ce test, dans son emploi courant, possède les propriétés d'optimalité asymptotique définies par Neyman. Car, dans cet emploi, si l'hypothèse testée consiste dans les probabilités  $p_1, p_2, \dots, p_k$ , l'alternative, que l'on n'explicité pas en général, sera tout autre système de probabilités  $p'_1, p'_2, \dots, p'_k$  (de somme unité).

On doit donc tester une hypothèse simple contre une hypothèse paramétrique, et l'on peut chercher le test de type  $\lambda$  (équivalent comme on sait au test contre l'hypothèse simple obtenue en estimant les paramètres par la méthode de Fisher).

Le rapport de probabilité ("likelihood ratio") est alors le maximum de

$$\prod_{i=1}^k \left( \frac{p'_i}{p_i} \right)^{x_i} \quad \text{et s'écrit} \quad \prod_{i=1}^k \left( \frac{x_i}{n p_i} \right)^{x_i}$$

La meilleure région critique est de la forme

$$\prod_{i=1}^k \left( \frac{x_i}{n p_i} \right)^{x_i} > C$$

En passant aux logarithmes, il vient :

$$\sum x_i \log \frac{x_i}{n p_i} \approx \sum \frac{(x_i - n p_i)^2}{n p_i} > \text{Cte}$$

On peut donc ainsi justifier le test du  $\chi^2$  par la méthode de Neyman.

#### L'introduction du coût des erreurs

Avant d'aborder la conclusion, à laquelle mènent ces réflexions sur la théorie des tests, revenons un instant sur la question du coût des erreurs. Beaucoup de statisticiens - surtout économistes - ont salué avec joie la parution des travaux de Wald (statistical Decision functions) moins pour le rapprochement avec la théorie des jeux et l'usage du minimax, que pour l'accent mis sur les coûts : remplacer une condition fixant un seuil de confiance, par la condition de minimiser quelque coût moyen, voilà qui semblait permettre d'échapper complètement à l'arbitraire dont souffrait - en apparence - la statistique.

Il faut alors examiner cette question : d'affecter des coûts aux erreurs que peut commettre le statisticien, cela permet-il d'échapper à toute règle conventionnelle ? Dans le cadre de la théorie de Bayes, il est facile de s'en assurer. Quelques complications formelles peuvent se présenter dans le cas (paramétrique) où l'hypothèse testée est privilégiée, mais cela est assez secondaire. Raisonnons, pour la simplicité, avec deux hypothèses  $H_0$  et  $H_1$  simples. Supposons que les erreurs seules sont coûteuses, et soient  $p_0$  et  $p_1$  les probabilités a priori des hypothèses,  $f_0(x)$  et  $f_1(x)$  les lois des observations dans chaque hypothèse ; les observations  $x$  étant connues, on sera conduit à rejeter l'hypothèse  $H_0$  si les probabilités a posteriori  $P_0$  et  $P_1$  sont telles que :

$$P_1 B > P_0 A$$

A et B étant respectivement les coûts, des erreurs de 1ère et de 2ème espèce ; c'est-à-dire que la région critique sera de la forme :

$$\frac{f_0(x)}{f_1(x)} < \frac{p_1 B}{p_0 A}$$

La région critique est donc de la même forme que celle que définit la théorie de Neyman, mais les coûts et les probabilités a priori déterminent le seuil de confiance du test, comme il est naturel.

Que peut-on dire, maintenant, dans la théorie de Neyman, si l'on introduit les coûts des erreurs, et cela seulement ? Il est facile de voir qu'on ne peut que reporter la convention des probabilités sur les

coûts - mais le choix d'un test ne saurait être ramené à une simple condition d'optimum.

Le tableau des coûts éventuels est donné ci-dessous.

		Hypothèse "vraie"	
		$H_0$	$H_1$
Hypothèse retenue	$H_0$	O	B (probabilité $\beta$ )
	$H_1$	A (probabilité $\alpha$ )	O

Mais les probabilités  $\alpha$  et  $\beta$  sont des probabilités "conditionnelles" et ne permettent pas de calculer un coût moyen global, en l'absence de probabilités des hypothèses. Que peut-on faire alors ?

On peut fixer a priori non plus  $\alpha$ , mais le coût moyen conditionnel  $A\alpha$ , et choisir le test qui rend  $B\beta$ , c'est-à-dire  $\beta$ , minimum. On retrouve la règle de Neyman, avec un seuil de confiance inversement proportionnel au coût de l'erreur de première espèce.

L'application d'un critère autre que la minimation d'un coût, tel que le minimax, ne présente ici aucun intérêt, puisqu'elle impliquerait qu'on ne tienne aucun compte des observations. Et c'est justice, car l'absence de probabilités a priori interdit qu'on affecte aux hypothèses des probabilités, ce qui peut se traduire en disant que les observations ne nous apprennent rien sur la crédibilité des hypothèses. La théorie de Neyman donne des règles de conduite, non pas des règles de croyance. Pour citer Neyman, "this may be considered unfortunate, but we cannot help it, and have simply to face the fact." [5].

#### Peut-on conclure pour une théorie des tests - Neyman ou Bayes ?

Cela dit, quelle théorie faut-il donc admettre comme bonne ?

Nous pensons qu'il ne convient pas de poser cette question. Pour nous, la théorie de Bayes et la théorie de Neyman sont bonnes toutes les deux.

Dans un problème concret, lorsque nous éprouvons le besoin de tester une hypothèse, ce que nous savons des "chances", de la "vraisemblance" (Halphen) - de cette hypothèse et de celles qui peuvent la remplacer, est souvent assez flou, et de représenter cela par des probabilités a priori peut paraître une schématisation outrancière. Mais de n'en tenir aucun compte n'est pas moins outrancier, sans doute.

A se rallier à Neyman, on bénéficie de cet avantage, que toutes les probabilités dont on parle sont susceptibles d'une interprétation comme limites de fréquences - conditionnelles, il est vrai - et si le problème concret est tel que ces notions suffisent pour éclairer la décision, alors il faut certainement s'en tenir là. C'est le cas de beaucoup de problèmes touchant aux fabrications industrielles. Mais si le désir réel du "client" est d'éprouver sa confiance dans l'hypothèse testée, alors la théorie classique des tests repose sur la fameuse erreur de troisième espèce, et conduit à traiter avec beaucoup de rigueur un problème qui n'est pas le problème posé. Car l'interprétation du seuil de signification d'un test comme une probabilité de l'hypothèse testée (et qui nous garantit contre une telle interprétation même inconsciente ?) constitue une erreur capitale, hormis les cas particuliers où la distribution fiduciaire de Fisher sera regardée comme forme limite de la loi a posteriori des hypothèses. Encore faut-il que la loi a priori implicitement admise soit directement jugée raisonnable - et alors c'est Bayes et non plus Neyman, qui est en cause.

D'un autre côté, la théorie de Bayes a l'avantage de donner la réponse à un problème de test avec le minimum d'arbitraire - si l'attribution de probabilités aux hypothèses peut être jugée elle-même pas trop arbitraire.

Les avantages et inconvénients que nous venons de rappeler ne constituent nullement des raisons de choisir l'une ou l'autre théorie, puisque celles-ci ne répondent pas à la même question. C'est du problème concret que doit dépendre le point de vue adopté - les seules règles générales étant d'éviter les diverses sources de confusion qui ont joué un certain rôle historique, et d'accorder quelque soin à l'interprétation concrète des notions et des conventions proposées par la théorie.

#### BIBLIOGRAPHIE

- 1 BAYES - An essay towards solving a problem in the doctrine of chances. Philosophical Transactions - 1763.
- 2 Karl PEARSON - On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine - 1900.
- 3 STUDENT - The probable error of a mean. Biometrika - 1908.

- 4 R.A. FISHER - Statistical methods and scientific Inference - 1956 .
- 5 J. NEYMAN - Basic Ideas and some recent Results on the Theory of testing statistical Hypotheses. Journal of the Royal statistical Society - 1942.
- 6 WALD - Statistical Decision Functions - 1950.
- 7 COCHRAN - The  $\chi^2$  Test of goodness of Fit. Annals of Mathematical Statistics - 1952.
- 8 LEHMAN - Some principles of the theory of testing Hypotheses. Annals of Mathematical statistics - 1950.
- 9 L. J. SAVAGE - The foundations of statistics - 1954.
- 10 SCHLAIFER - Probability and statistics for Business Decisions. 1959.
- 11 E. HALPHEN - La notion de vraisemblance - Essai sur les Fondements du Calcul des Probabilités et de la Statistique Mathématique. Publications I. S. U. P. - 1955.
- 12 JEFFREYS - Theory of Probability. Oxford - 2ème edition 1947 .
- 13 LINDLEY - A statistical Paradox. Biometrika - 1957.

## BIBLIOGRAPHIE

- 1 BAYES - An essay towards solving a problem in the doctrine of chances. Philosophical Transactions - 1763.
- 2 KARL PEARSON - On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine - 1900.
- 3 STUDENT - The probable error of a mean. Biometrika - 1908.