

Réflexions synthétiques sur les tests statistiques

Cet exposé suit les lignes de la thèse de doctorat de Jacques POITEVINEAU, *Méthodologie de l'analyse des données expérimentales : Étude de la pratique des tests statistiques chez les chercheurs en psychologie, approches normative, prescriptive et descriptive*, soutenue le 11 mars 1998.

1 Les grandes interprétations de la probabilité

Si la probabilité en tant qu'objet mathématique est clairement définie par un système d'axiomes, son interprétation est diverse et l'on distingue trois grandes conceptions.

1.1 La conception objectiviste ou fréquentiste

Dans cette conception qui remonte à Venn et Von Mises, la probabilité de réalisation d'un événement est définie comme la limite de sa fréquence d'apparition quand le nombre d'épreuves (répétitions) tend vers l'infini. De ce point de vue, la probabilité d'un événement singulier n'a pas de sens.

1.2 La conception logiciste

Dans ce cadre, la probabilité exprime le degré de vérité d'une proposition incertaine. La logique classique en est un cas particulier où toute proposition ne peut être que vraie ou fausse. C'est une conception normative : le calcul de la probabilité d'un énoncé ne dépend pas des opinions de celui qui effectue le calcul. Le premier exposé systématique de cette conception est dû à Keynes.

1.3 La conception subjectiviste

Elle est très proche de la conception précédente. Cette fois la probabilité est conçue comme une mesure du degré d'incertitude d'un individu rationnel à l'égard d'un énoncé. Par individu rationnel il faut entendre un individu qui appliquerait rigoureusement les lois du calcul des probabilités. Cette conception n'est donc plus normative dans le sens où chaque individu est libre d'attribuer à un événement la probabilité qu'il souhaite. Simplement, deux individus qui attribuent une même probabilité à un certain événement et qui sont ensuite confrontés aux mêmes données le concernant doivent réviser leur probabilité initiale de la même manière.

1.4 Exemple illustratif

Il n'est peut être pas inutile de prendre un exemple pour illustrer les différences de conception de la probabilité. Considérons l'expérience qui consiste à tirer au hasard une carte dans un jeu de 52 cartes, à ne pas la regarder, et à se demander si cette carte est l'as de cœur. Si cette expérience est répétée indéfiniment, la fréquence de tirage de l'as de coeur tendra vers $1/52$.

Pour le fréquentiste, c'est tout ce que peut signifier “la probabilité de tirer l'as de cœur vaut $1/52$ ”. Maintenant considérons un tirage particulier. Pour le fréquentiste, il n'est pas pertinent de se demander quelle est la probabilité de tirer l'as de coeur, ou, plus précisément, cette probabilité ne peut être que 1 ou 0 (selon que c'est ou non l'as de cœur).

Pour le logiciste, l'essence même de la probabilité est bien de caractériser l'incertitude sur ce tirage, qu'il égale à cette fréquence limite de $1/52$.

2 La théorie de Fisher

Elle s'inscrit dans la lignée des travaux de Yule et de Karl Pearson. Fisher cristallise, en quelque sorte, les notions et les pratiques jusque là en vigueur, et les développe, tout en leur conférant le statut de méthode d'inférence incontournable. C'est en 1925 que paraît la première édition de *Statistical Methods for Research Workers*, puis en 1935 celle de *The Design of Experiments* qui connaissent un succès considérable (14 éditions pour le premier ouvrage, 8 pour le second).

2.1 Mise en œuvre

Une seule hypothèse est mise à l'épreuve. Elle est appelée hypothèse nulle, dans le sens de “to be nullified”; c'est-à-dire à réfuter, et non nécessairement, comme on le trouve écrit dans certains manuels, d'une valeur de zéro pour le paramètre testé, même si c'est le cas le plus fréquent. En général l'hypothèse nulle sera la négation de l'hypothèse à laquelle le chercheur s'intéresse réellement.

Le résultat de la procédure de test est :

- soit le rejet de l'hypothèse nulle,
- soit la suspension du jugement : on ne rejette ni n'accepte l'hypothèse nulle.

Il n'y a, par conséquent, qu'une seule possibilité d'erreur : rejeter l'hypothèse nulle alors qu'elle est vraie.

2.2 Une extension du raisonnement par l'absurde

Le raisonnement sous-jacent correspond à une extension probabiliste du raisonnement par l'absurde, où, pour démontrer la fausseté d'une proposition A , on procède ainsi :

- Supposons A vraie.
- Alors, et sans utiliser d'autres propriétés et théorèmes que ceux tenus pour vrais, on en déduit quelque chose de contradictoire sur une certaine proposition B , elle-même déjà connue par ailleurs (c'est-à-dire indépendamment de A) : que B est vraie alors qu'on la sait fausse, ou inversement.
- Si donc la véracité de A entraîne quelque chose d'impossible, c'est que A est nécessairement fausse (par application de modus tollens).

seuil α

C zone critique
= zone de rej.

Cela devient, dans le cas du test de l'hypothèse nulle :

- Supposons vraie l'hypothèse nulle.
- Calculons, dans ce cas, la probabilité associée à un résultat au moins aussi extrême que celui observé. Si celle-ci est faible cela signifie qu'un événement rare, improbable, sous cette hypothèse, a été observé.
- Posons comme principe supplémentaire que ce qui se produit effectivement (ce qui est observé) n'est pas rare.
- Nous sommes alors en présence d'une contradiction, et nous sommes conduits à rejeter l'hypothèse nulle, la considérant comme fausse (par application de modus tollens).

forte probabilité que H_0
soit fausse.

stat du test

H_0 supposée vraie

$\{ \in C \Rightarrow$ probable contradiction \Rightarrow Rejet de H_0

$\{ \notin C \Rightarrow$ pas de contradiction \Rightarrow pas de décision.

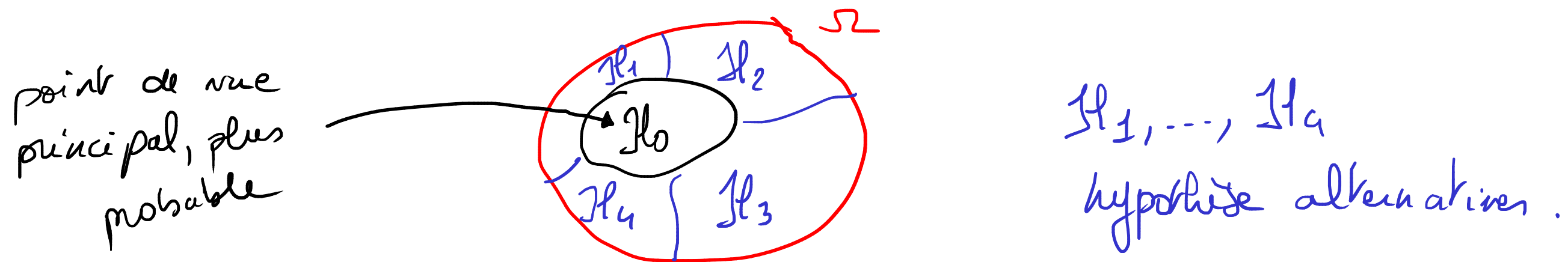
3 La théorie de NEYMAN et PEARSON

C'est en 1928 et 1933 que paraissent leurs articles fondateurs. Bien que s'inspirant, au départ, des travaux de Fisher, leur approche participe davantage d'une théorie statistique de la décision et ils ne la présentent pas, au contraire de Fisher, comme le fondement d'une nouvelle logique inductive.

Pour eux il s'agit bien moins de pouvoir conclure sur la véracité d'une hypothèse que d'adopter, à son égard, un comportement rationnel selon un critère de contrôle des erreurs de décision :

3.1 Mise en œuvre

Comme chez Fisher il s'agit de mettre à l'épreuve une hypothèse particulière, qu'ils notent H_0 . Cependant cette hypothèse est envisagée d'un point de vue plutôt opposé à celui de Fisher. En effet, alors que chez ce dernier l'hypothèse effectivement testée est la négation de l'hypothèse d'intérêt, Neyman et Pearson précisent que H_0 est l'hypothèse à laquelle on s'intéresse particulièrement, et qu'elle est souvent celle qui semble la plus probable a priori. Mais cette fois H_0 n'est plus suffisante pour construire le test. Il convient de prendre également en compte les hypothèses alternatives admissibles, notées H_i , et qu'on suppose toujours pouvoir être définies.



But : éclairer le choix entre H_0, H_1, \dots, H_4 .

Parfois H_0 contre H_1 (cas particulier) -

3.2 Procédure de test

On doit décider entre :

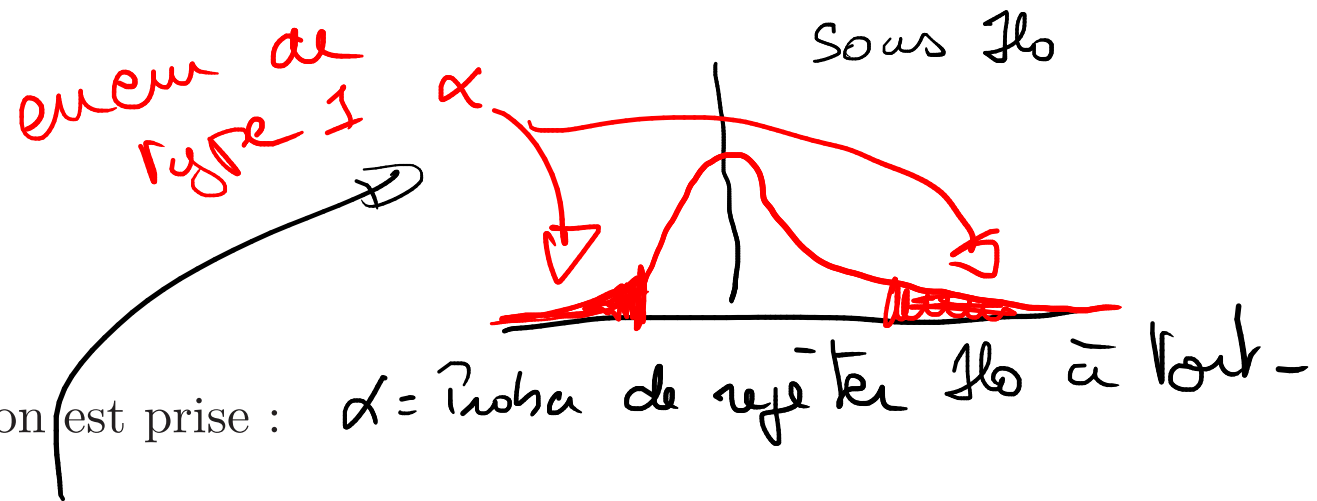
- soit rejeter H_0 , mais sans que cela soit au profit d'une hypothèse H_i particulière, sauf, bien sûr, s'il n'y a qu'une hypothèse alternative, auquel cas on l'accepte,
- **soit accepter H_0 ,**
- soit éventuellement rester dans le doute (cas où les données sont insuffisantes). Cette dernière possibilité est présentée comme un cas particulier de la deuxième, mais elle est rarement, pour ne pas dire jamais, illustrée par les auteurs qui, de ce fait, mettent l'accent sur la dichotomie acceptation/rejet de l'hypothèse H_0 .

Il peut être utile de rappeler que “rejeter” ou “accepter” signifient pour eux “choisir telle ou telle action” et non pas croire ou non (ou plus ou moins) en l'hypothèse.

3.3 Quantification des erreurs

Il y a donc deux possibilités d'erreurs, quand une décision est prise :

- L'erreur dite de première espèce ou de type I consiste à rejeter H_0 alors que H_0 est vraie. La probabilité conditionnelle (à la véracité de H_0) correspondante est appelée risque de première espèce, et est notée α le plus souvent.
- L'erreur dite de deuxième espèce ou de type II consiste à accepter H_0 alors qu'une hypothèse alternative H_i est vraie. Le risque de deuxième espèce est la probabilité conditionnelle (à la véracité de H_i) correspondante, souvent notée β_i (mais aussi parfois $1 - \beta_i$). Son complément $1 - \beta_i$ (ou bien β_i), probabilité de choisir H_i alors que H_i est vraie, est appelée la puissance du test par rapport à H_i .



Dans le cas où il existe plus d'une hypothèse alternative on définit la puissance résultante du test comme étant la probabilité, non conditionnelle cette fois, de correctement rejeter H_0 . Mais cette

puissance résultante fait intervenir, en plus des β_i propres à chaque H_i , les probabilités a priori de ces H_i et ne peut donc être déterminée, en général.

- Toujours dans ce cas, on parlera également de la fonction de puissance du test pour désigner la fonction qui associe la valeur de la puissance à la valeur du paramètre correspondant à chacune des hypothèses H_i . Cette fonction caractérise le test.

L'enjeu d'accepter H_0 devient alors clair puisque toutes les probabilités d'erreurs sont connues.

⎧ Toutes choses égales par ailleurs, les risques α et β varient en sens inverse : diminuer l'un fait augmenter l'autre. La puissance dépend du risque α , de la valeur du paramètre sous l'hypothèse alternative, et de la taille N de l'échantillon. En particulier, les autres termes étant fixés, la puissance augmente avec N .

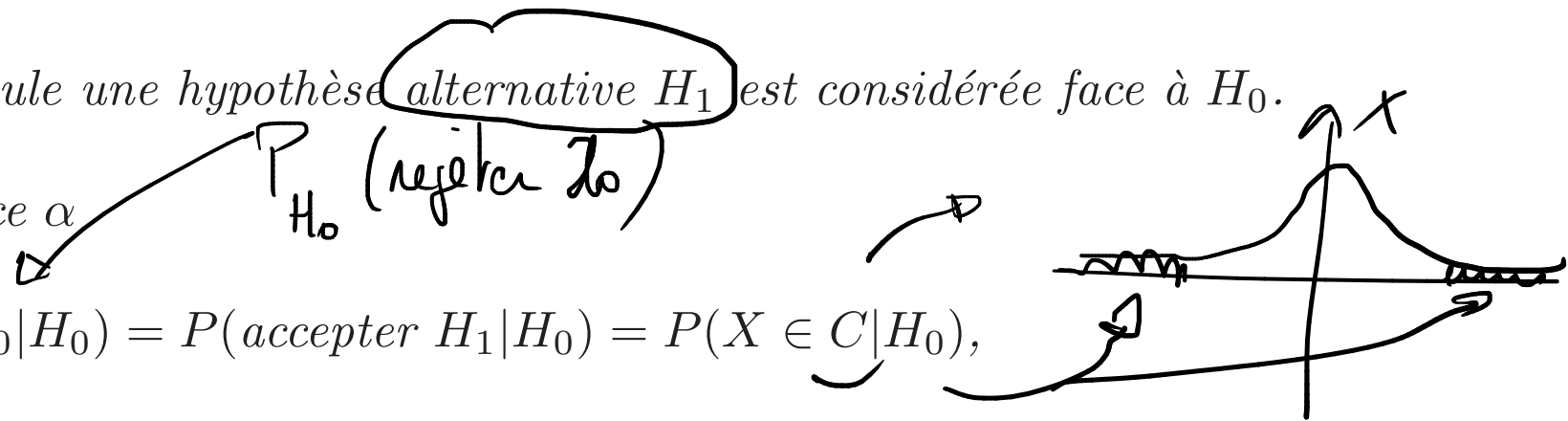
3.4 Probabilités d'erreur

Definition 3.1 Considérons le cas où seule une hypothèse alternative H_1 est considérée face à H_0 .

- On appelle risque de première espèce α

$$\alpha = P(\text{rejeter } H_0 | H_0) = P(\text{accepter } H_1 | H_0) = P(X \in C | H_0),$$

où X est la statistique du test et C la zone de rejet (région critique). La quantité α est aussi appelée niveau du test.



- *On appelle risque de deuxième espèce β*

$$\beta = P(\text{accepter } H_0 | H_1) = P(\text{rejeter } H_1 | H_1) = P(X \notin C | H_1),$$

où X est la statistique du test et C la zone de rejet (région critique). La quantité $1 - \beta$ est aussi appelée puissance du test.

Dans le cadre général, on appelle risque de deuxième espèce β_i associée à H_i , $\beta_i = P(\text{accepter } H_0 | H_i)$.

3.5 Exemple

Soit X une variable aléatoire discrète. On veut confronter les deux hypothèses suivantes

- $H_0 : X \sim \mathcal{U}([1; 6])$,
- $H_1 : X \sim \mathcal{B}(6; 0, 5)$.

Déterminer le niveau et la puissance de la région critique $C = \{0, 2, 3\}$.

$$\alpha = \mathbb{P}(\text{rejeter } H_0 | H_0) = \mathbb{P}(X \in C | \underbrace{X \sim \mathcal{U}([1, 6])}_{\text{sachant } H_0}) = \overbrace{\mathbb{P}(X=0) + \mathbb{P}(X=2) + \mathbb{P}(X=3)}^{X \sim \mathcal{U}([1, 6])} = 0 + \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$$\beta = P(\text{accepter } H_0 | H_1) = P(X \notin C | X \sim \mathcal{B}(6, 0,5)) = P(X \in \{1, 4, 5, 6\} | X \sim \mathcal{B}(6, 0,5))$$

$\hookrightarrow \{0, 2, 3\}$

$$= \frac{6}{64} + \frac{15}{64} + \frac{6}{64} + \frac{1}{64} = \frac{28}{64} = \frac{7}{16}$$

$$X \sim \mathcal{B}(6; 0,5)$$

Sous H_1 la fonction de masse de X est

x	0	1	2	3	4	5	6
$p(x)$	$\frac{1}{64}$	$\frac{6}{64}$	$\frac{15}{64}$	$\frac{20}{64}$	$\frac{15}{64}$	$\frac{6}{64}$	$\frac{1}{64}$

On obtient donc que

$$1 - \beta = \Pr(\mathbf{X} \in C | H_1) = \frac{1}{64} + \frac{15}{64} + \frac{20}{64} = \frac{36}{64} = \frac{9}{16}$$

Ainsi,

$$\beta = \frac{7}{16}$$

En utilisant ce test, on a donc 1 chance sur 3 de se tromper si H_0 est vraie et 7 chances sur 16 de se tromper si c'est H_1 qui est vraie.

4 Le cadre fréquentiste

Bien qu'il soit habituel de classer les deux approches précédentes dans le cadre d'une conception fréquentiste de la probabilité, c'est-à-dire où celle-ci est exclusivement définie comme la limite d'une fréquence, ce qualificatif s'applique surtout à la théorie de Neyman et Pearson. À ce propos, Rouanet (1997) parle de "fréquentiste modéré" pour Fisher et de "fréquentistes radicaux" pour Neyman et Pearson. En effet, α et β prennent leur sens seulement quand on envisage que le chercheur réplique à l'infini son expérience (ce qui signifie dans des conditions identiques non seulement quant aux conditions expérimentales mais aussi quant aux règles de décision) : 100 α % des fois le chercheur commettrait une erreur de type I si H_0 est vraie. D'ailleurs Neyman introduit explicitement cette conception fréquentiste de la probabilité dans ses ouvrages de 1950 et 1952.

5 Un amalgame des deux théories

Le plus souvent les chercheurs, et même les statisticiens, ne distinguent pas clairement les deux théories, mais utilisent divers amalgames. Par exemple, les chercheurs se basent le plus souvent sur le seuil observé p (Fisher), mais ils font aussi appel, occasionnellement, à la notion de risque de deuxième espèce et/ou de puissance (Neyman et Pearson). Ou bien les deux risques, de première et deuxième espèces, sont mis en avant (Neyman et Pearson), ou même la référence à Neyman et Pearson est explicite, mais le choix de H_0 se fait par négation de l'hypothèse d'intérêt, et sans considération aucune de l'importance relative des risques d'erreur, ce qui relève de l'approche fishérienne. Nous avancerons même l'opinion qu'il n'existe pas, dans la pratique, de “purs” neymaniens dans la mesure où nous n'avons jamais vu appliqué à des données réelles un test répondant au principe méthodologique de Neyman mentionné plus haut.

Enfin, nous évoquerons encore les interprétations des résultats du test (cf. la section 6.3) qui hors du cadre des théories fréquentistes.

6 Critiques et abus des tests

6.1 Bis repetita

La théorie de Neyman et Pearson conçoit la procédure de test sur le long terme, comme ces auteurs l'ont bien précisé. Donner un sens fréquentiste à la probabilité α requiert de considérer une répétition à l'infini de tests identiques (les hypothèses, α , N et la région critique doivent rester inchangés), ce qui implique que le résultat d'un test particulier, c'est-à-dire la décision prise, ne doit pas jouer sur la construction des autres tests, ne doit pas modifier les expériences futures (dans le cadre d'un même problème, bien sûr).

Dans le domaine du contrôle de qualité, auquel se réfèrent souvent Neyman et Pearson, on peut facilement imaginer la situation suivante. Dans une usine une machine produit un très grand nombre de pièces. Chaque jour un échantillon de pièces est prélevé, examiné, et l'on décide sur la base d'un test statistique d'accepter la production journalière ou de la rejeter et de réparer la machine. Réparer la machine est bien sûr coûteux pour l'entreprise (coût de la réparation auquel s'ajoute celui relatif à la perte de production), de même que la livraison de lots défectueux (entraînant remboursement ou remplacement). La même procédure est répétée chaque jour, et peu importe finalement, du point de vue de l'entreprise, qu'un jour particulier la décision soit erronée ou correcte; ce qui compte c'est qu'à la longue (sur un an, dix ans,...) les risques soient contrôlés.

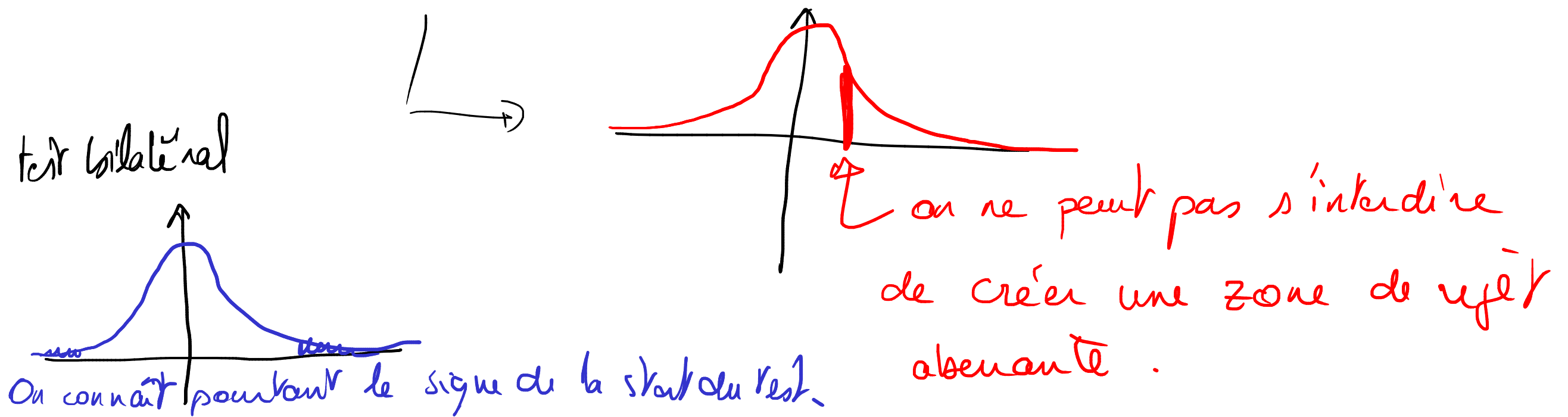
On est bien loin de la situation dans la recherche scientifique où, même si le critère de reproductibilité est fondamental, une expérience n'est pratiquement jamais répétée à l'identique un très grand nombre de fois (surtout par un même chercheur).

Même en se plaçant sur le terrain (apparemment) favorable de l'exemple précédent des lots de pièces manufacturées, un problème se pose. Si le test réalisé se révèle non significatif, l'hypothèse nulle correspondant au bon fonctionnement de la machine sera acceptée (même provisoirement), rien ne sera modifié et le processus pourra se perpétuer. Mais si l'on observe un résultat significatif, il est alors évident que cela va entraîner des modifications, à savoir la mise au rebut du lot produit et la réparation de la machine (il serait évidemment absurde de laisser la machine en l'état : si le test est réalisé c'est bien pour détecter ce cas). Or cette modification de la machine signifie, du point de vue du modèle statistique, un changement de la valeur du paramètre considéré.

6.2 De l'arbitraire

6.2.1 Zone de rejet

Dans la théorie de Fisher, puisqu'il n'existe pas d'erreur de type II, donc pas de risque β à minimiser, aucune justification formelle n'existe quant au choix de la région de rejet de l'hypothèse nulle. Dès lors, comme le remarque Rozeboom (1960), on pourrait tout aussi bien la choisir au centre de la distribution d'échantillonnage. La pratique qui consiste à choisir comme région de rejet l'extrémité de la distribution d'échantillonnage, bien que raisonnable, est donc arbitraire.



6.2.2 La valeur du seuil

Dans les deux théories le choix de la valeur du seuil de référence, qui marque la frontière entre significatif et non significatif, est aussi éminemment arbitraire ou subjectif et introduire des fonctions de coûts des erreurs ne fait que déplacer le problème (Rozeboom, 1960; Camilleri, 1962; Winer, 1971, p. 14; Skipper et al., 1967, par exemple). Cela est encore plus sensible dans la théorie de Neyman et Pearson puisque les conclusions pourront être diamétralement opposées selon que l'on se trouve en deçà ou au delà du seuil de référence (à partir de mêmes données, deux chercheurs pourraient prendre des décisions différentes sur la base de risques α différents); alors que dans le cadre fishérien, dans la mesure où un résultat non significatif n'amène qu'à suspendre le jugement, il n'y aura pas de véritable contradiction.

$p = 0,03$ \rightarrow On rejette pour $\alpha = 5\%$
 \rightarrow On ne rejette pas pour $\alpha = 2\%$.

6.3 Conditionnement Observations/Données

Le raisonnement manque de naturel, il est contre-intuitif. On calcule une probabilité correspondant à l'événement observé, conditionnellement à une hypothèse sur le paramètre d'intérêt

$$P(Donnees|Hypothese),$$

alors qu'il semblerait plus naturel, à l'inverse, comme on le fait dans le cadre bayésien, de calculer une probabilité sur les valeurs possibles du paramètre, conditionnellement à l'événement observé

$$P(Hypothese|Donnees).$$

6.4 Discontinu ou continu ?

Le passage d'un résultat significatif à un résultat non significatif est discontinu alors que la statistique de test est continue (Rozeboom, 1960). Ceci explique les discussions sur le choix de la valeur du seuil de signification α : ce choix n'aurait pas tant d'importance s'il n'y avait intériorisation, de la part des chercheurs, de la différence entre 0.05 et 0.06 comme différence entre le “vrai” et le “faux”, la “réussite” et l’“échec”. Skipper et al. (1967) évoquent même la joie ou l'horreur du chercheur selon que “son” F atteint 0.05 ou ne donne que 0.06. Ils regrettent particulièrement à ce propos que le choix du seuil se fasse, presque toujours, sans considération de la nature et du type de l'étude (contrairement à ce que préconisaient Fisher, cf. 1.1.). Ce problème est également lié à celui portant sur l'alternative décision/jugement : une décision a un caractère discontinu, tandis que l'évolution d'un niveau de confiance en une hypothèse apparaît plutôt comme continu.

6.5 Le statut de l'hypothèse de recherche

Les problèmes soulevés par les tests sont particulièrement bien illustrés par la question du statut de l'hypothèse de recherche, celle à laquelle le chercheur s'intéresse réellement, et qu'il espère confirmer : doit-on identifier l'hypothèse de recherche à l'hypothèse nulle ou à l'hypothèse alternative ?

Souvent, le choix est limité par le fait que l'hypothèse nulle doit être ponctuelle pour permettre les calculs.

6.6 La question de l'intensité de l'effet

Le test ne dit rien quant à l'intensité, l'importance de l'effet parent (cf., par exemple, O'Brien et Shapiro, 1968; Rouanet et al., 1976). Un résultat significatif n'est qu'une indication de l'existence de l'effet supposé; un résultat non significatif un constat d'ignorance selon Fisher. Aller plus loin sur la seule base du test renvoie à l'erreur d'assimiler significativité statistique et importance de l'effet. Or, presque tous les auteurs reconnaissent que la question de l'intensité de l'effet est essentielle.

6.7 Significativité statistique et significativité substantielle

La troisième erreur est de confondre la significativité statistique avec la significativité substantielle (substantive significance). C'est considérer que plus un résultat est significatif, plus il est scientifiquement intéressant, et/ou que plus l'effet correspondant dans la population parente est grand (substantive importance).

7 Quelques éléments de solutions

7.1 La mesure de l'effet

Prenons le test du χ^2 ... En statistiques, on ne peut que “quantifier la distance à l’indépendance” par la statistique du χ^2 ,

$$D_{\chi^2} = n \sum_{j=1}^J \sum_{k=1}^K \frac{(f_{jk} - f_{j\bullet} f_{\bullet k})^2}{f_{j\bullet} f_{\bullet k}}.$$

où J et K sont le nombre de modalités de chacune des deux variables considérées.

	x	
y	f_{ij}	$f_{i.}$
	$f_{.j}$	

expose.

ne bouge pas si
on augmente l'effectif
mais on garde la répartition

Si n grand, Dx^2 augmente. (q_{log} augmente avec le nombre de modalités (DDL)
mais pas avec n).

7.1.1 Coefficients ϕ et C

Les coefficients ϕ et C découlent de la statistique du χ^2 par les formules

$$C = \sqrt{\frac{D_{\chi^2}}{D_{\chi^2} + n}}, \quad \phi = \sqrt{\frac{D_{\chi^2}}{n}}.$$

En réalité ces deux coefficients sont une variante l'un de l'autre. L'avantage de C est qu'il est compris entre 0 et 1, alors que ce n'est pas le cas pour le ϕ . Plus ces indicateurs sont proche de zéro, plus il y a indépendance entre les deux variables X et Y étudiées.

7.1.2 V de Cramér

Comme pour le coefficient ϕ , plus le V de Cramér est proche de zéro, plus il y a indépendance entre les deux variables X et Y étudiées. Il vaut 1 en cas de complète dépendance.

Le coefficient V de Cramér nécessite l'utilisation de la statistique du χ^2 via la formule

$$V = \sqrt{\frac{D_{\chi}^2}{n \times \min\{l - 1; c - 1\}}},$$

où n est l'effectif total de la population, c est le nombre de colonnes (nombre de modalités de Y) et l le nombre de lignes (modalités de X).

7.1.3 Interprétation

L'interprétation des coefficients ϕ et V est empirique et dépend du domaine d'application (sciences économiques, sciences humaines, médecine...). On peut considérer le tableau suivant pour l'interprétation (tout en vérifiant les valeurs frontières d'usage dans chaque domaine).

Valeur du V de Cramér	Intensité de la relation entre les variables
inférieur à 0,10	relation nulle ou très faible
entre 0,10 et 0,20	relation faible
entre 0,20 et 0,30	relation moyenne
au dessus de 0,30	relation forte

7.2 Régression linéaire et indicateurs en part de variance expliquée

Dans le cadre des régressions linéaires, on citera bien entendu les coefficients de corrélation linéaire r , de détermination R^2 ou encore d'amélioration $a = 1 - \sqrt{1 - R^2}$ dont l'interprétation est la suivante :

- Existence d'une corrélation linéaire si $a > 0,5$,
- Corrélation linéaire faible si $a < 0,5$.

7.3 Les intervalles de confiance

7.3.1 Les mêmes critiques que les tests de signification

La correspondance avec le test (neyman-pearsonien) a pour conséquence que la plupart des critiques développées contre ce dernier s'appliquent autant à l'intervalle de confiance : manque de naturel, recours à la notion de répétition à l'infini, caractère décisionnel, rôle des valeurs non observées (en énonçant que 95% des intervalles possibles contiennent le paramètre, on fait référence à des intervalles qui ne sont pas observés), etc.

7.3.2 Des conclusions surprenantes

Il est des cas où l'intervalle de confiance amène à des conclusions qui peuvent sembler surprenantes. Par exemples, il est possible de trouver une borne négative pour un paramètre positif, telle une variance.

7.3.3 Est-ce le bon intervalle ?

L'intervalle de confiance est souvent centré sur une valeur observée d'une valeur théorique non atteignable, ce qui peut amener au décalage de cet intervalle.

7.4 L'approche Bayésienne

Exercice 7.1 Soient (Ω, \mathbb{P}) un espace probabilisé et $A, B \subseteq \Omega$ deux évènements tels que

$$\mathbb{P}(A) \neq 0, \quad \mathbb{P}(\bar{A}) \neq 0 \quad \text{et} \quad \mathbb{P}(B) \neq 0.$$

(i) Montrer la **formule des probabilités totales**:

$$\mathbb{P}(B) = \mathbb{P}(A) \mathbb{P}_A(B) + \mathbb{P}(\bar{A}) \mathbb{P}_{\bar{A}}(B).$$

(ii) Montrer la **formule de Bayes**:

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A) \mathbb{P}_A(B)}{\mathbb{P}(A) \mathbb{P}_A(B) + \mathbb{P}(\bar{A}) \mathbb{P}_{\bar{A}}(B)}.$$

$\overbrace{\mathbb{P}(A) \mathbb{P}_A(B)}^{\mathbb{P}(A \cap B)}$
 $\underbrace{\hspace{10em}}_{\mathbb{P}(B)}$

(iii) Soit une population où 90% des individus sont vaccinés, 0.5% des individus vaccinés sont malades et 15% des non vaccinés sont malades. On examine un individu qui s'avère être malade. Quelle est la probabilité qu'il ait déjà été vacciné?

$$P(V) = 0,9$$

$$P(\bar{V}) = 0,1$$

$$P_V(M) = 0,005$$

$$P_{\bar{V}}(M) = 0,15$$

$$P_M(V) = \frac{0,9 \times 0,005}{0,9 \times 0,005 + 0,1 \times 0,15} = 23\%$$

7.5 Les méta-analyses

- On n'a pas souvent accès aux données,
- Pas souvent le même protocole,
- Comment trancher sur les p-values ?
 - moyenne des p-values ? Rien de logique !
 - rejet dès qu'une p-value est trop grande ?
 - ...