M. EYTAN

# Les diverses écoles de statisticiens (d'après une conférence de M. Raiffa)

*Mathématiques et sciences humaines*, tome 4 (1963), p. 34-40
<http://www.numdam.org/item?id=MSH_1963__4__34_0>

M. EYTAN

# LES DIVERSES ECOLES DE STATISTICIENS
## (d'après une conférence de M. RAIFFA)

Nous allons indiquer, sur un exemple schématique, la démarche d'esprit suivie par certaines écoles de Statisticiens.

Soit A un attribut que peuvent posséder des individus d'une certaine population, et soit p la probabilité qu'un individu possède effectivement A. Comment s'y prendra-t-on pour calculer p ?

1) R.A. FISHER: se pose la question fondamentale "tel phénomène peut-il être dû au seul hasard ? "

Considérons par exemple la proposition

$$"p \geqq p_0 = 0{,}3 \text{ (pour fixer les idées)}".$$

On tire de la population un échantillon de n individus (n = 20 par exemple); le nombre d'individus possèdant l'attribut A est r(r = 9 par exemple). La proportion des succès est donc de 0,45. Quelle est la probabilité d'un tel évènement si effectivement $p < p_0$ ?

Or $Pr(r \geq 9 \mid n = 20,\ p = 0{,}3) \# 0{,}113$ (schéma d'urne avec remplacement); cette probabilité n'est pas très petite, l'évènement n'est pas extrêmement rare, il y a lieu de procéder à des essais supplémentaires.

2) NEYMAN-PEARSON: ils considèrent que tester une hypothèse est un problème de décision. Il faut donc se préoccuper de la probabilité d'une erreur au cas où l'hypothèse serait vraie et au cas où elle serait fausse. Nous devons donc poser l'hypothèse nulle $H_0$ : $p \leqq p_0$ contre l'hypothèse alternative $p > p_0$. On aimerait rejeter $H_0$. Pour cela on décide d'une stratégie pour accepter ou rejeter $H_0$ (avant de connaître r), de la manière suivante: on choisira un nombre-limite c tel que :

si $r \leqq c$ alors $H_0$ sera accepté;

si $r > c$ alors $H_0$ sera rejeté et l'alternative acceptée.

Ceci est une "règle (n,c)" où n est le nombre des tirages. On devra donc :

1) Evaluer les règles (n,c).

2) Choisir un c optimal, n étant donné.

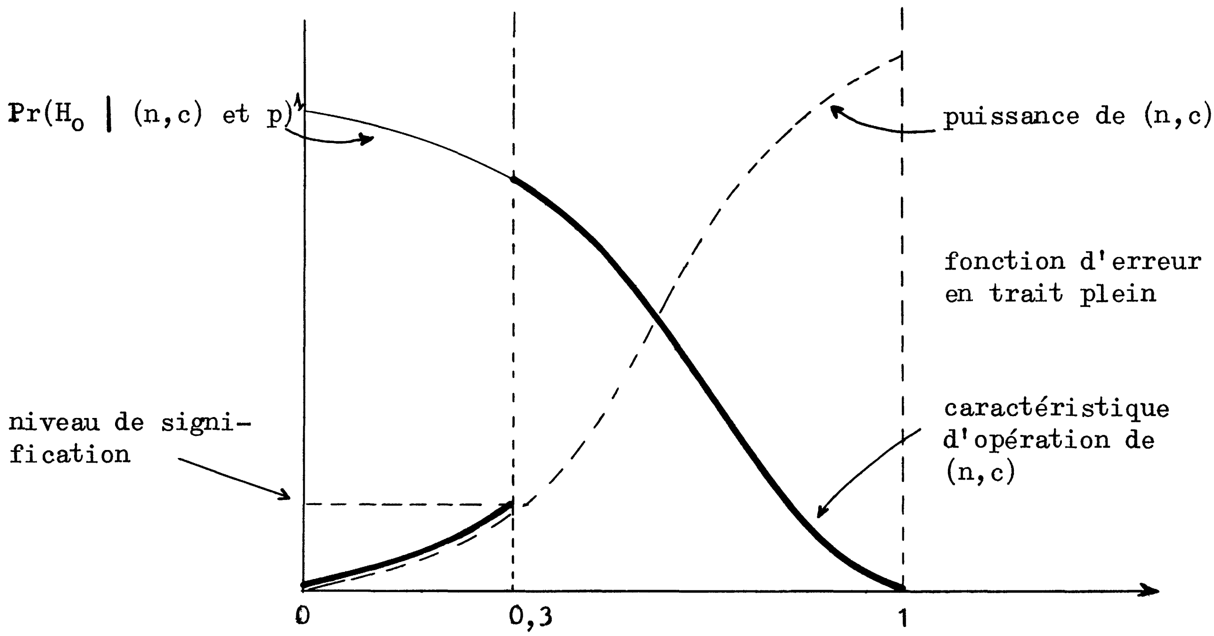
3) Choisir un n approprié.

Supposons donnée une règle (n,c), p. ex. (20,8).
La "puissance" est la fonction $Pr(\overline{H_0} \mid (n,c) \text{ et } p)$ c'est-à-dire :

$$1 - Pr(H_0 \mid (n,c) \text{ et } p).$$

La "fonction d'erreur" du test est la probabilité d'une erreur: c'est la puissance pour les $p \leqq 0,3$ et la caractéristique d'opération, c'est-à-dire la fonction $Pr(H_0 \mid (n,c)$ et $p)$ pour les $p > 0,3$.

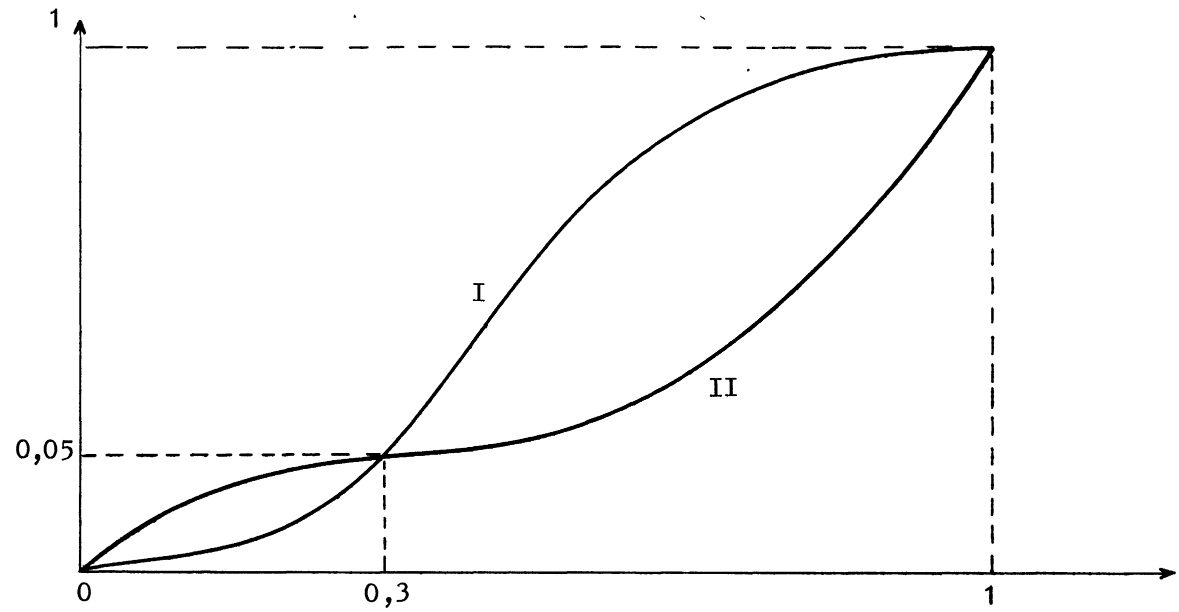
Le "niveau de signification" d'une règle (n,c) est le nombre: $Pr(H_0 \mid (n,c)$ et $p_0 = 0,3)$.



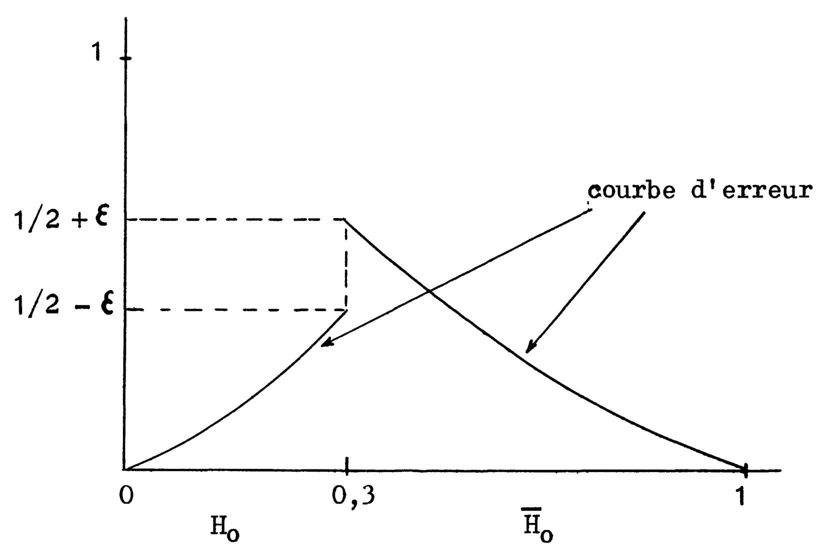

Dans le cas pratique, le choix de (n,c) devra dépendre de toute la courbe d'erreur, et le choix de c du risque encouru, nous disent Neyman et Pearson. Or ceci est difficile à appliquer, il s'agit d'un jugement personnel. En effet, fixons n et faisons varier c. Lorsque c croît, le risque d'erreur sur $H_0$ décroît, mais le risque d'une erreur sur $\overline{H}_0$ croît. C'est ce dilemme qui est à l'origine de certaines conventions qui se sont établies, à l'encontre des considérations explicites de Neyman et Pearson, et qui stipulent qu'il faut choisir c de manière que le niveau de signification soit 0,05 ou 0,01... Or une telle "recette" est inadmissible : le choix du niveau de signification doit dépendre de $H_0$ et de $\overline{H}_0$ (que nous voulons vérifier), et certainement aussi de n.

Soient par exemple (dans la figure ci-après) deux courbes de puissance, d'aspects très différents et pour lesquels on a un même niveau de signification, 0,05.

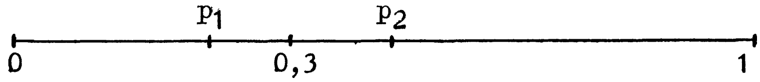
Il est manifeste que l'on ne choisira pas (n,c) de la même manière dans les deux cas, bien que le niveau de signification soit le même pour la courbe I et la courbe II.




Une autre difficulté de la procédure est que lorsqu'on a fixé n, on ne peut diminuer le risque d'erreur à la fois sur $H_0$ et sur $\overline{H}_0$. Si donc nous avons une situation où la probabilité $Pr(\overline{H}_0 \mid (n,c)$ et $p)$ est voisine de 1/2, alors que nous avons décidé d'accepter $H_0$, la situation devient intolérable (cf. fig.). On doit alors augmenter la taille de l'échantillon n. Ceci s'ajoute à la liste des problèmes du statisticien.




Par ailleurs, comment choisir n ? Neyman et Pearson proposent alors la solution suivante: soit $p_1$ un point voisin de $p_0 = 0,3$ $(p_1 < p_0)$ et $\alpha$ l'erreur tolérable en $p_1$ :




de même soit $p_2$ un autre point voisin de $p_0$ $(p_0 < p_2)$ et $\beta$ l'erreur tolérable en

$p_2$. Les nombres $p_1$, $\alpha$, $p_2$, $\beta$ dépendent des risques encourus. Puis trouver une règle (n,c) qui justement garantisse une erreur inférieure à $\alpha$ en $p_1$ et inférieure à $\beta$ en $p_2$. Si n est trop grand, recommencer tout en modifiant $\alpha$ et $\beta$. On voit qu'il s'agit encore d'un compromis.

Remarquons qu'ici il s'agit d'un problème à un seul paramètre: un problème à deux paramètres exigerait une discussion bien plus complexe, puisque les fonctions d'erreur, de puissance .... seraient représentées par des surfaces. Indiquons toutefois que la procédure recommandée dans ce cas est de vérifier en quelques points choisis l'aspect de la puissance.

Quant aux domaines où interviennent une foule de variables, la procédure décrite ci-dessus devient manifestement inapplicable. Dans ce cas l'hypothèse nulle, $H_0$ est par exemple :

$$\mu_1 = \mu_2 = \dots = \mu_m.$$

Neyman et Pearson recommandent de remplacer ceci par l'"indicateur" :

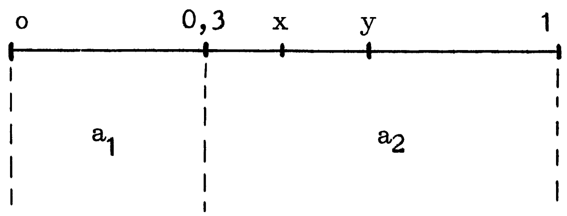
$$\delta^2 = \sum_{i=1}^{m} (\mu_i - \overline{\mu})^2$$

(d'après $H_0$, $\delta^2$ doit être nul), et de considérer la courbe de puissance de cet indicateur.

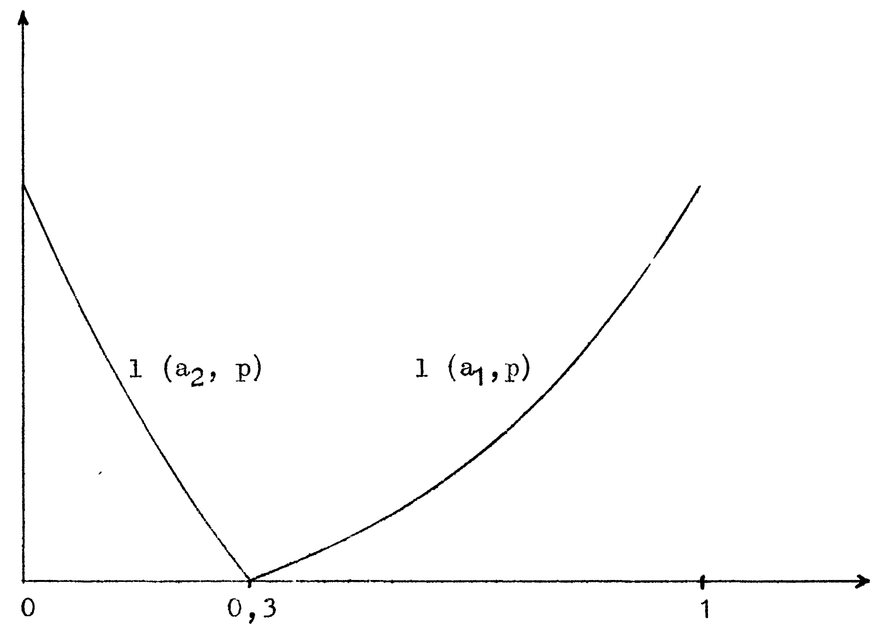
Les objections que l'on peut faire à cette Ecole sont de poids: le niveau de signification, érigé en méthode standard, n'a pas de sens. Les facteurs ne sont jamais totalement indépendants; si l'échantillon est assez grand on obtient une bonne vérification de l'hypothèse et le niveau 0,05 est atteint. Répétons que le sens du "niveau de signification" dépend de l'aspect global de la courbe.

3) **A. WALD:** introduit les concepts de décision et de perte.

Supposons que p soit plus grand que 0,3, et qu'au point x (si la vraie valeur est x, cf. fig. ci-dessous) nous convenions de prendre la décision $a_1$. Ceci est moins grave que de prendre la décision $a_1$ au point y (si la vraie valeur est y).

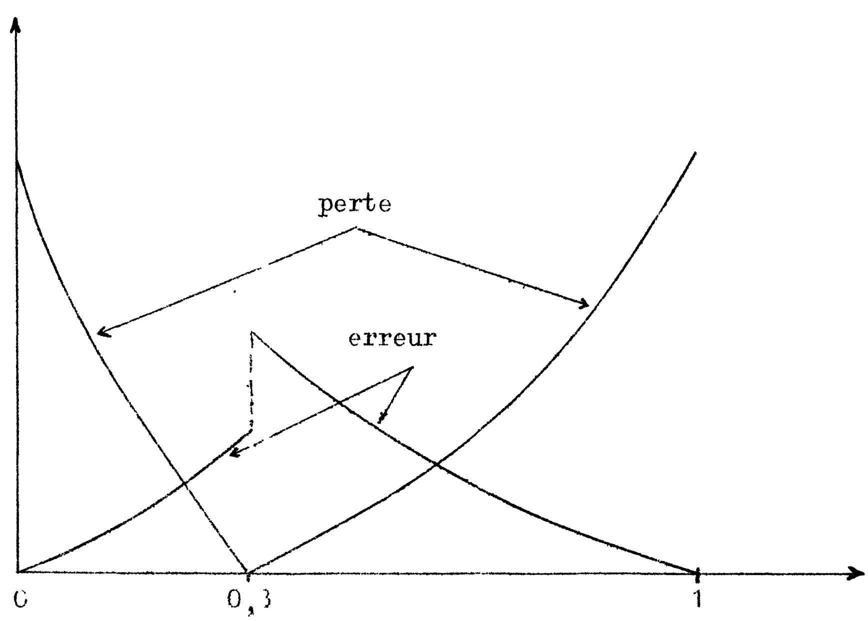



Il est donc naturel de postuler l'existence d'une fonction de "perte conditionnelle", $l(a_1,p)$, croissante avec p; de même pour la fonction de perte conditionnelle $l(a_2,p)$ de l'autre côté de la valeur $p_0 = 0,3$.

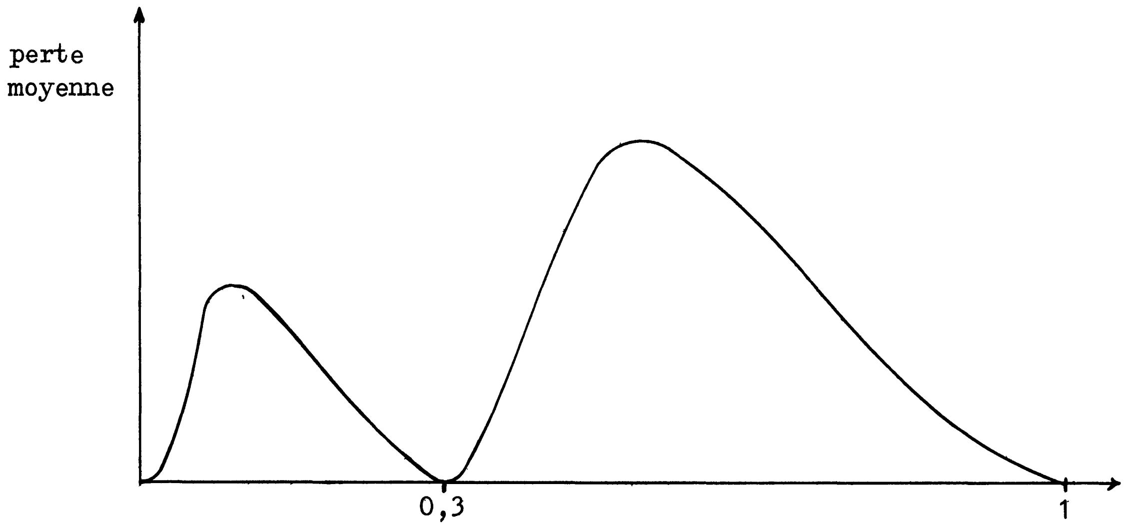



Une stratégie est le choix d'un couple (n,c). La théorie de Neymann et Pearson est valable, mais il faut considérer non seulement la probabilité d'une erreur, mais aussi la perte conditionnelle subie.

Il faut donc superposer la courbe de perte conditionnelle et la courbe d'erreur, puis calculer la perte probable en faisant le produit de la probabilité d'une erreur par la perte due à cette erreur.

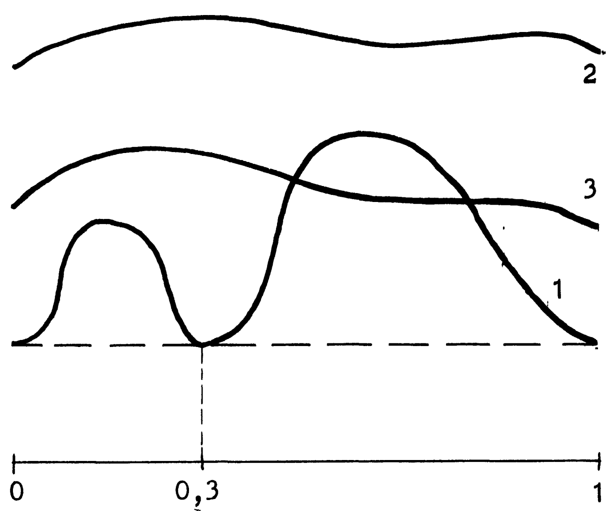



On obtient finalement la courbe de perte probable d'une règle (n,c) donnée.




Lorsque c décroît, la perte probable décroît si $p \leq p_0$, et croît si $p > p_0$. n étant fixé, comment faut-il choisir c? Ceci dépend de l'aspect global de la courbe de perte probable. Le conseil de Wald est: si vous pensez que $p \leqq p_0$, alors faites décroître c. On doit obtenir un bon contrôle de l'erreur du côté où l'on pense que se trouve p; si l'on ne sait rien, il faut se servir du critère minimax, mais ce critère n'est justifié que dans le cas d'une ignorance absolue. Reste que le minimum peut être atteint en un point sensiblement éloigné de $p_0$.

Nous pouvons diminuer la perte probable en augmentant n, mais ceci est coûteux, au fond c'est une perte. Ajoutons donc le coût de l'expérimentation à la perte probable d'où la perte globale. Supposons que cette dernière ne dépende que de n (et non pas du résultat) et considérons les trois courbes de perte schématisées ci-dessous:




1 est préférable à 2 par le principe de dominance stricte. Mais comment décider entre 1 et 3 ?

On pourrait appliquer le critère minimax, mais on est conduit à des résultats qui contredisent la plus élémentaire intuition.

4) **SAVAGE:** se propose de résoudre la difficulté rencontrée par Wald, en pondérant les p, et cherchent à rendre la somme de la perte probable et du coût de l'expérimentation, minimum.

La perte probable est donc remplacée par l'utilité (ou plutôt la désutilité, puisqu'il s'agit de pertes). Mais du point de vue mathématique, on peut effectuer la pondération dès le départ, on obtient une probabilité subjective. Puis on la modifie pour obtenir une probabilité a posteriori. En ce qui concerne le choix de n, il faut considérer la forme extensive, calculer les diverses utilités ... etc.. Cette dernière procédure est extrêmement simple.

La réalisation pratique de ces considérations est exposée dans le livre de Raiffa et Schlaiffer: Stat. Decision Theory.

Indiquons pour terminer comment généraliser au cas de trois décisions à prendre: $a_1$, $a_2$, $a_3$.

- La théorie de Neymann - Pearson se heurte à des difficultés insurmontables, on ne sait plus quelle espèce d'erreur on commet.
- La théorie de Wald n'est pas modifiée, la perte conditionnelle l étant fonction des $a_i$ et de p.
- La méthode de Savage non plus ne rencontre pas de difficultés, et de plus présente l'avantage de pouvoir utiliser les méthodes de la Théorie des Jeux.