Exemple d'application des méthodes non paramétriques et un nouveau test pour l'égalité de plusieurs probabilités

PAR

M. J. HEMELRIJK

CBRM

COLLOQUE SUR L'ANALYSE STATISTIQUE

Tenu à Bruxelles les 15, 16 et 17 décembre 1954

Exemple d'application des méthodes non paramétriques et un nouveau test pour l'égalité de plusieurs probabilités (¹)

par M. J. HEMELRIJK (Amsterdam)

SOMMAIRE

Les méthodes non paramétriques de corrélation, développées par Pitman, Kendall et d'autres, peuvent être appliquées dans beaucoup de situations pratiques. Un exemple est la mesure de la capacité d'un inspecteur (ou d'une méthode d'inspection), qui juge la qualité d'une série d'objets. La théorie existante ne dispose pas encore d'un test non paramétrique pour l'égalité de deux coefficients de corrélation dérivés de deux séries d'observations. Dans l'exemple mentionné et dans des situations analogues on peut obvier à cette difficulté au moyen d'un artifice, dont l'utilité ne se borne peut-être pas au cas traité.

Les méthodes non paramétriques donnent parfois des résultats d'une simplicité surprenante. Lié aux problèmes mentionnés un nouveau test pour l'égalité de plusieurs probabilités est discuté. Le test, développé par M^lle C. van Eeden et l'auteur est différent du test bien connu, basé sur χ^2, quant aux hypothèses alternatives. Celles-ci sont plus spécifiées pour le nouveau test et elles ont comme cas particulier une tendance des probabilités testées dans un ordre donné. Il en résulte que sous certaines conditions un test pour le problème de deux échantillons, par exemple le test de Wilcoxon, suffit pour atteindre ce but. Les résultats sont appliqués à l'exemple mentionné plus haut. Certaines possibilités pour généraliser ces méthodes sont indiquées.

SUMMARY

This paper describes an application of the rank correlation methods developed by Pitman, Kendall a.o. to the problem of measuring the quality of the judgment of an inspector (or of a system of inspection), approving and rejecting objects under inspection. This theory does not, as far as the author knows, furnish a distributionfree test for the comparison of two rank correlation coefficients, both of which are different from zero. The comparison of the qualities of judgment of two inspectors which in our example is the analogue of the comparison of two rank correlation coefficients, may, however, be effected by means of a simple artifice, which may be useful in other situations too.

The method treated is closely connected with a new distribution-free test for the hypothesis that a number of probabilities are equal,

(¹) Rapport SP 38 du Centre Mathématique, Amsterdam.

which has recently been developed by Miss C. van Eeden and the author. Of this test, the alternative hypotheses of which are more specified than those of the well-known χ^2-test, a concise description is given. The test may be used to test the equality of probabilities against the alternative hypothesis of a trend of these probabilities and under certain conditions a distributionfree two sample test proves to be sufficient to attain this end. Results with respect to its consistency are applied to the above mentioned example. The possibility of generalisation is indicated.

1. Introduction

L'application des méthodes non paramétriques est encore peu répandue, quoique dans l'industrie, par exemple, il y ait beaucoup de possibilités pour de telles applications. Le but du présent exposé est de donner un exemple simple de ces méthodes, basé sur la théorie des échantillons prélevés dans une population finie avec comme cas particuliers les tests bien connus de E. J. G. Pitman (1937) et de F. Wilcoxon (1945). Les applications de cet exemple nous mèneront à la considération d'un nouveau test pour l'égalité de plusieurs probabilités avec des hypothèses alternatives de caractère spécifié, qui a été développé récemment par M^lle C. van Eeden (1954), en collaboration avec l'auteur.

Le problème, dont une solution sera donnée, est le jugement de la capacité d'inspecteurs quant au contrôle de la qualité d'une série d'objets donnés ([2]), mais on pourra remarquer que la méthode proposée est aussi adaptable à d'autres problèmes pratiques de même espèce. On rencontre ce type de problème par exemple dans la psychologie appliquée. D'ailleurs le mot « inspecteur » ne doit pas être pris au pied de la lettre : il signifie seulement « un système quelconque d'inspection ». Nous considérerons aussi la comparaison des capacités de deux inspecteurs par rapport à une échelle, déterminée d'une manière quelconque, des qualités des objets sous inspection.

2. Des coefficients de capacité et leur distribution en cas d'incompétence complète

Soient $C_1, \dots, C_n$ un nombre (n) d'objets, dont nous supposerons d'abord les qualités $Q_1, \dots, Q_n$ connues et déterminées d'une manière quantitative. S'il est seulement possible

([2]) Ce problème m'a été posé par M. A. R. van der Burg, du « Bureau consultatif sur la Statistique appliquée » (« Adviesbureau voor Toegepaste Statistiek »), à La Haye, qui a écrit sur ce sujet un rapport intitulé *Judging the Judge*, édité par ce bureau. Ce rapport s'occupe d'autres aspects du problème que ceux traités ici.

de les ordonner par qualité croissante, leurs numéros d'ordre peuvent être envisagés comme les valeurs $Q_1, \ldots, Q_n$. Nous admettons la possibilité que certains objets aient la même qualité. En tout cas les objets peuvent être rangés selon leurs qualités, de sorte que

$$Q_1 \leqq Q_2 \leqq \cdots \leqq Q_n . \tag{1}$$

Un inspecteur, qui ne connaît pas les nombres Q_i, va maintenant juger les n objets, chaque objet étant soit approuvé soit rejeté ([3]). On souhaite dans ces conditions définir une mesure pour la capacité de l'inspecteur par rapport à l'échelle donnée (1) des qualités.

Or, ce problème n'est pas compliqué et suivant les méthodes habituellement utilisées en statistique non paramétrique, nous introduisons la valeur typique

$$V = \sum_{i=1}^{n} x_i Q_i \tag{2}$$

où x_i est une variable, qui prend la valeur 1 pour les objets C_i approuvés et 0 pour les objets rejetés. La plus grande valeur possible de V, étant donné que

$$x = \sum_{i=1}^{n} x_i \tag{3}$$

objets ont été approuvés, est

$$V_{\max} = \sum_{i=n-x+1}^{n} Q_i \tag{4}$$

et le minimum de V, dans cette même condition, est

$$V_{\min} = \sum_{i=1}^{x} Q_i . \tag{5}$$

Evidemment les grandes valeurs de V indiquent un bon accord entre le jugement de l'inspecteur et la qualité des objets, tandis que les petites valeurs de V indiquent le jugement opposé à la qualité.

Si l'on veut avoir une mesure, qui prenne des valeurs entre 0 et 1, on peut définir le coefficient

$$W_1 = \{ V - V_{\min} \} / \{ V_{\max} - V_{\min} \} . \tag{6}$$

([3]) Cette situation se réalise souvent en pratique et nous la considérons à cause de sa simplicité. Pourtant, la plupart des méthodes utilisées ici, peuvent être généralisée sans trop de difficulté.

On peut se demander maintenant quelle valeur W_1 prendrait pour un inspecteur incompétent et à partir de quelle valeur on peut dire que l'inspecteur n'est pas incompétent. Pour répondre à cette question nous recourrons à une définition statistique : un inspecteur sera dit « complètement incompétent », si son jugement des objets est indépendant de leur qualité, c'est-à-dire si, pour chaque nombre x donné d'objets approuvés tous les groupes possibles de x objets entre les n ont la même probabilité de se composer des objets approuvés.

Sous cette hypothèse H_0 d'incompétence complète les x objets approuvés peuvent être considérés comme un échantillon aléatoire prélevé dans la population des n objets, avec les qualités $Q_1, \ldots, Q_n$ et il n'est pas difficile de trouver la distribution de probabilité de $\underline{V}$ et par conséquent celle de $\underline{W}_1$ [4]. La théorie des échantillons prélevés sans remplacement dans une population finie nous fournit les moments

$$\mathcal{E}\{\underline{V}\,|\,H_0\}=x\,\mu\ ;\quad \sigma^2\{\underline{V}\,|\,H_0\}=x(n-x)(n-1)^{-1}\sigma^2 \quad (7)$$

où

$$\mu=n^{-1}\Sigma Q_i\ ;\qquad \sigma^2=n^{-1}\Sigma(Q_i-\mu)^2\,. \qquad (8)$$

Les équations (7) ont été obtenues sous la condition $\Sigma\underline{x}_i=x$ et par conséquent ces résultats, compte tenu d'un théorème de W. G. Madow (1948) sur la normalité de la distribution limite de $\underline{V}$, nous fournissent un test conditionnel de H_0. Ce théorème de Madow, qui est une généralisation d'un théorème limite de A. Wald et J. Wolfowitz (1944), dit que, sous l'hypothèse H_0, et pour n et $x\to\infty$, $\underline{V}$ est distribué selon la distribution normale, si les conditions suivantes se réalisent :

$$\sup_n |n^{-1}\Sigma(Q_i-\mu)^r| \,/\, \{n^{-1}\Sigma(Q_i-\mu)^2\}^{r/2}<\infty \qquad (9)$$

pour chaque $r>0$,

$$\varepsilon n<x<(1-\varepsilon)n \qquad (10)$$

pour une valeur quelconque de $\varepsilon>0$ et n assez grand.

Ces conditions ne sont pas très restrictives. La seconde signifie seulement que l'inspecteur ne doit pas approuver presque tous les objets mais qu'il doit en rejeter au moins une fraction positive, aussi petite soit-elle. La première condition est satisfaite, par exemple, si les Q_i sont uniformément bornés, avec un écart type positif à la limite pour $n\to\infty$,

(4) Une variable aléatoire sera désignée par un caractère souligné. Le même symbole, non souligné, indique une valeur, qui peut être prise par la variable aléatoire.

si $Q_i = i$ (dans le cas des numéros d'ordre), si les moments des Q_i ont des limites finies, etc.

La connaissance de la distribution de $\underline{V}$ nous donne un test pour l'hypothèse H_0. La région critique sera constituée des grandes valeurs de V, indiquant la concordance entre la qualité et le jugement de l'inspecteur et, si l'on veut, des petites valeurs de V, qui indiquent un faux jugement de l'inspecteur. Nous formulerons plus loin avec plus de précision les hypothèses alternatives pour lesquelles ce test est consistant.

En concordance avec l'équation (7) on peut aussi introduire, à côté de W_1, le coefficient

$$W_2 = \{ V - x\mu \} / \{ V_{\max} - x\mu \} , \tag{9}$$

dont l'espérance mathématique est égale à 0 si l'inspecteur est complètement incompétent. La plus grande valeur de W_2 est 1 et les valeurs négatives indiquent un faux jugement.

3. Le rapport avec d'autres tests

En désignant les qualités des x objets approuvés par Q_1', ..., Q_x' et les autres par Q''_1, ..., Q''_{n-x}, l'hypothèse H_0 signifie que toutes les divisions différentes de Q_1, ..., Q_n en ces deux groupes ont la même probabilité. Cela est précisément la situation utilisée par E. J. G. Pitman (1937*a*) d'après une idée de R. A. Fisher (1935) pour donner sa solution non paramétrique du problème des deux échantillons et le test indiqué dans la section précédente est alors identique au test de Pitman, les qualités des objets approuvés et rejetés constituant les deux échantillons.

Si les Q_i sont des numéros d'ordre, ou s'ils sont remplacés par leurs numéros d'ordre, on obtient le test de Wilcoxon (1945) et dans ce cas la distribution de $\underline{V}$ est connue *exactement* pour des petites valeurs de n et x. Ce test étant un cas particulier de la méthode non paramétrique de corrélation de M. G. Kendall (1948), basée sur le coefficient désigné par τ, on peut aussi dire que V est équivalent à ce coefficient τ, dérivé des n pairs de nombres (Q_i, x_i), $i = 1, \dots, n$, où les x_i sont 0 pour les objets rejetés et 1 pour les objets approuvés.

On peut envisager le problème posé d'un autre point de vue, plus pratique pour formuler des hypothèses alternatives. En désignant par $p_i (i = 1, \dots, n)$ la probabilité que l'inspecteur approuvera l'objet C_i, et supposant que les jugements des divers objets sont indépendants, H_0 peut être formulée comme suit

$$H_0' : \quad p_1 = p_2 = \dots = p_n \qquad (= p_0 \text{ par ex.}) . \tag{10}$$

La condition de l'indépendance des jugements est une condition supplémentaire, qui n'est pas nécessairement réalisée en pratique. Si l'on permet, par exemple, que l'inspecteur compare les objets en les jugeant, une certaine dépendance est inévitable. On touche ici à un problème ressemblant au problème de la « Lady tasting tea », sujet d'une discussion prolongée entre R. A. Fisher et J. Neyman (cf. par exemple J. Neyman (1950), p. 285). Nous ne traiterons pas cette question plus à fond ici, remarquons seulement que si l'on veut employer H_0', on doit réaliser cette indépendance des jugements.

Sous cette condition H_0' est équivalent à H_0. Car la probabilité que x objets spécifiés soient approuvés et les autres rejetés est, dans ce cas, égale à $p_0^x \, q_0^{n-x}$ pour chaque groupe de x objets ; alors H_0 est satisfait. D'autre part si H_0' n'est pas satisfait, ces probabilités ne seront pas égales pour tous les groupes possibles de x objets spécifiés et alors H_0 n'est pas vrai.

Il suit de ces considérations qu'il est possible de tester H_0' au moyen de $\underline{V}$ et notamment qu'on peut ramener le test de H_0' à un test de Wilcoxon en partageant les épreuves en deux échantillons, les « succès » d'une part et les « insuccès » d'autre part, les Q_1', ..., Q_x' étant les valeurs prises dans le premier échantillon et Q''_1, ..., Q''_{n-x} dans le second. C'est cette idée, trouvée d'une autre manière par A. Bernard et M^{lle} C. van Eeden, qui a amené cette dernière et l'auteur à élaborer un test général pour H_0' avec des hypothèses alternatives bien spécifiées. Une brève description de ce test sera donnée plus loin et nous en profiterons alors pour étudier les hypothèses alternatives pour lesquelles le test de la section 2 et un autre test, donné dans la section suivante, sont consistants.

4. Comparaison des capacités de deux inspecteurs

En eux-mêmes des coefficients comme V, W_1 et W_2 ne sont pas très intéressants. C'est davantage leur distribution de probabilité sous des hypothèses d'importance pratique, qu'il faut connaître en vue des applications. Très souvent il est tellement évident que l'hypothèse H_0 (ou H_0') n'est pas vraie, qu'il n'est pas du tout nécessaire de la tester. Dans ce cas il est fort souhaitable qu'on puisse comparer, au moyen d'un test, deux valeurs de V provenant de deux inspecteurs différents. C'est-à-dire, qu'il nous faut un test pour l'hypothèse que les jugements de deux inspecteurs sont équivalents contre l'hypothèse alternative, qu'un des deux inspecteurs juge les objets mieux en accord avec leur qualité que l'autre. Et ce test doit être valable si H_0 n'est pas vrai.

Or, dans la théorie non paramétrique de la corrélation on ne dispose pas encore d'un test pour l'hypothèse que les valeurs du τ de Kendall de deux populations soient égales ; de même il n'existe pas de test pour la comparaison de deux valeur de l'U de Wilcoxon. La solution du problème actuel n'est pas fournie alors par la théorie non paramétrique générale de la corrélation ; mais par un petit artifice on peut ramener ce problème au problème traité dans les sections précédentes.

Les deux inspecteurs, que nous considérons, jugent les mêmes objets $C_1, \ldots, C_n$. Certains de ces objets, disons $n - m$, seront ou bien approuvés par les deux inspecteurs ou rejetés par eux. Pour ces objets alors les deux jugements sont égaux et ils ne nous fournissent pas d'information sur une différence éventuelle des jugements. Il est alors avantageux de les éliminer et de chercher un test conditionnel, basé sur les m objets restants. Chacun d'entre eux a été approuvé par un des deux inspecteurs seulement et, en appelant les inspecteurs A et B, on peut diviser les objets en deux groupes : le groupe A^*, constitué au nombre de y qui ont été approuvés par A et rejetés par B et le groupe B^*, constitué du reste de ces m objets. Maintenant on peut envisager les probabilités, étant donné les valeurs de m et de y de tous les groupements possibles de cette sorte et on peut *définir* l'équivalence des jugements des inspecteurs A et B, par rapport aux objets jugés, par l'égalité de toutes ces probabilités. Cette définition semble acceptable : elle signifie que la distribution de ces m objets entre A et B, définissant les deux groupes A^* et B^*, est indépendante de leurs qualités.

Or, on peut tester l'hypothèse $H_0{}^*$, que les jugements de A et B sont équivalents en ce sens, par le même test que H_0. Dans ce cas, la valeur typique V [cf. (2)] est remplacée par

$$V^* = \sum_{j=1}^{m} y_j Q_j{}^* \qquad (12)$$

où $Q_j{}^*$ est la qualité du j^{me} des m objets formant les groupes A^* et B^* et où y_j prend la valeur 1 si cet objet appartient au groupe A^* et 0 s'il appartient à B^*. Les grandes et les petites valeurs de V^* forment la région critique bilatérale, les grandes valeurs indiquant que le jugement de l'inspecteur A est en meilleur accord avec l'échelle des qualités donnée que celui de B et les petites valeurs indiquant la situation inverse.

Il est intéressant d'examiner aussi l'autre point de vue mentionné plus haut, c'est-à-dire de poser la condition supplémentaire que tous les objets $C_1, \dots, C_n$ sont jugés indépendamment et d'introduire pour chaque objet C_i deux probabilités p_{A_i} et p_{B_i} d'être approuvé respectivement par les inspecteurs A et B. La probabilité que l'objet C_i soit approuvé par un des inspecteurs et rejeté par l'autre est alors égale à

$$P_i = p_{A_i} q_{B_i} + p_{B_i} q_{A_i} \quad (q_{A_i} = 1 - p_{A_i} \; ; \; q_{B_i} = 1 - p_{B_i})$$

et la probabilité conditionnelle pour C_i d'appartenir à A^*, à condition d'appartenir soit à A^* soit à B^*, est égale à

$$P[C_i \in A^* \mid C_i \in A^* \cup B^*] = p_{A_i} q_{B_i} / P_i \, . \tag{13}$$

Considérons maintenant l'hypothèse $H_0{}^{*\prime}$, qu'il y ait un p_0 avec

$$H_0{}^{*\prime} : \qquad p_0 = p_{A_i} q_{B_i} / P_i \qquad (i = 1, \dots, n) \, . \tag{14}$$

Sous cette hypothèse tous les y_j de (12) ont une probabilité p_0 de prendre la valeur 1 et ils sont distribués indépendamment. Alors, comme dans la section 3, l'hypothèse $H_0{}^*$ résulte de $H_0{}^{*\prime}$ et réciproquement, et c'est alors $H_0{}^{*\prime}$, qu'on teste avec V^*. Si l'on a pour deux inspecteurs

$$p_{A_i} = p_{B_i} \qquad (i = 1, \dots, n) \, , \tag{15}$$

$H_0{}^{*\prime}$ est satisfait et on peut alors tester (15) avec ce test. Cependant, ceci est un cas particulier, impliquant que les inspecteurs sont de la même sévérité. En général $H_0{}^{*\prime}$ est équivalent à

$$H_0{}^{*\prime} : \qquad p_{A_i} q_{B_i} / p_{B_i} q_{A_i} = p_0 / (1 - p_0) = c \qquad (i = 1, \dots, n) \tag{16}$$

mais cette hypothèse ne semble pas très naturelle. D'autre part nous ne disposons pas d'un test pour une hypothèse plus naturelle : l'équivalence des jugements de deux inspecteurs n'impliquant pas une équivalence de sévérité. Pour le moment nous devons nous contenter du test mentionné. On pourrait éviter cette difficulté en imposant aux inspecteurs le nombre des objets à approuver, mais alors on perd l'indépendance des jugements et on peut seulement se servir de $H_0{}^*$, et non de $H_0{}^{*\prime}$.

5. Le nouveau test pour l'égalité de plusieurs probabilités

Considérons maintenant le test décrit par Mlle C. van Eeden (1954) et (1955) pour l'hypothèse $H_0{}'$ [cf. (10)] avec des hypothèses alternatives spéciales. Le test est basé sur n épreuves indépendantes, $E_1, \dots, E_n$, E_i ayant une probabilité p_i de donner un « succès » ($i = 1, \ldots, n$). La forme de ce test est alors la suivante :

L'hypothèse à tester est

$$H_0' : \quad p_1 = p_2 = \ldots = p_n . \tag{10}$$

En définissant $\underline{x}_i$ comme une variable aléatoire, prenant la valeur 1 si E_i est un « succès » et 0 autrement, la valeur typique sur laquelle le test est basé, est

$$\underline{V}' = \sum_{i=1}^{n} g_i \underline{x}_i , \tag{17}$$

où les « poids » g_i sont des fonctions données de n

$$g_i = g_i(n) \qquad (i = 1, \ldots, n) , \tag{18}$$

qui ne sont pas toutes égales.

On est libre de choisir les g_i comme on veut ; l'ensemble des hypothèses alternatives pour lesquelles le test est consistant dépend de ce choix. Cependant, sans nuire à la généralité, on peut imposer, pour chaque n, les conditions suivantes

$$\sum_{i=1}^{n} g_i = 0 ; \qquad \sum_{i=1}^{n} g^2_i = 1 . \text{ [5]} \tag{19}$$

Avec

$$\underline{x} = \sum_{i=1}^{n} \underline{x}_i , \tag{3}$$

la variable aléatoire $\underline{V}'$ a sous H_0' et sous la condition $\underline{x} = x$, les moments suivants

$$\mathrm{E} \{ \underline{V}' \mid \mathrm{H}_0' ; \underline{x} = x \} = 0 , \tag{20}$$

et

$$\sigma^2 \{ \underline{V}' \mid \mathrm{H}_0' ; \underline{x} = x \} = n^{-1} (n - 1)^{-1} x (n - x) . \tag{21}$$

En outre, la distribution conditionnelle de $\underline{V}'$, sous H_0', est asymptotiquement normale si

$$\lim_{n \to \infty} \ \operatorname*{Max}_{0 < i \leq n} \ | g_i(n) | = 0 \tag{22}$$

et s'il y a un $\varepsilon > 0$ avec

$$\varepsilon < n^{-1} x(n) < 1 - \varepsilon \tag{23}$$

(5) On peut aussi prendre une autre condition, par exemple $\Sigma |g_i| = 1$. Cela change un peu les formules, sans modification importante des résultats.

pour n assez grand, où $x(n)$; $n=1, 2, \ldots$ désigne une série de valeurs de x, prises pour $n=1, 2, \ldots$ ([6]).

Alors, si on trouve une valeur de $n^{-1}x$, qui n'est pas proche de 0 ou 1, on peut approcher la distribution conditionnelle de $\underline{V}'$, sous H_0' et sous la condition $\underline{x}=x$, par une distribution normale dont les paramètres sont donnés par (20) et (21). La région critique du test est constituée par des valeurs extrêmes de $\underline{V}'$.

Quant aux hypothèses alternatives, nous nous bornerons au cas, où il y a un $\varepsilon > 0$ avec

$$\varepsilon < n^{-1} \sum_{i=1}^{n} p_i < 1 - \varepsilon \tag{24}$$

pour n assez grand, à cause de la condition (23), qui ne serait pas satisfaite autrement. Alors on a

$$\mathcal{E} \left\{ \underline{V}' \mid p_1, \ldots, p_n \right\} = \sum_{i=1}^{n} g_i p_i \tag{25}$$

et

$$\sigma^2 \left\{ \underline{V}' \mid p_1, \ldots, p_n \right\} = \sum_{i=1}^{n} g_i^2 p_i q_i \tag{26}$$

et on trouve que le test est consistant, si

$$L = \lim_{n \to \infty} \sum_{i=1}^{n} g_i p_i = \pm \infty . \tag{27}$$

Si la $\lim \Sigma g_i p_i$ existe et est finie, le test n'est pas consistant pour des niveaux de signification assez petits. En ce sens la condition (27) est aussi nécessaire et, en omettant les cas pathologiques où la $\lim \Sigma g_i p_i$ n'existe pas, le problème de la consistance est résolu, sauf à quelques détails près, non mentionnés ici.

Quelques *exemples* peuvent aider à comprendre la portée de ce test. Si l'on veut tester H_0' par rapport à une tendance des p_i avec i en ce sens que les hypothèses alternatives soient caractérisées par

$$\Delta = \sum_{i<j} (p_i - p_j) = \sum_{i=1}^{n} (n + 1 - 2\,i)\, p_i , \tag{28}$$

([6]) Cette condition est suffisante mais non nécessaire. On peut aussi formuler des conditions suffisantes un peu différentes.

on peut prendre pour les g_i

$$g_i = (n + 1 - 2\,i)/\sqrt{n\,(n^2 - 1)/3}\;, \tag{29}$$

satisfaisant (19) et (22). Alors (27) devient

$$\lim_{n \to \infty} n^{-\frac{3}{2}} \sum_{i<j} (p_i - p_j) \to \pm\,\infty\;. \tag{30}$$

Les g_i étant équidistants, le test se ramène dans ce cas au test de Wilcoxon appliqué aux i des « succès » relativement à ceux des « insuccès ».

Un autre cas important se présente, si les hypothèses alternatives forment une classe limitée par l'égalité de certains p_i. Cela arrive par exemple si l'on a un nombre fixé k de groupes d'expériences avec un p_i constant dans chaque groupe, comme l'indique le schéma suivant :

$$\left\{\begin{array}{llllll} \text{groupe :} & 1 & 2 & \dots & k & \\ \text{nombre d'expériences :} & n_1 & n_2 & \dots & n_k & (\Sigma n_\nu = n) \\ \text{nombre de succès :} & a_1 & a_2 & \dots & a_k & (\Sigma a_\nu = x) \\ \text{probabilité de succès :} & p_1' & p_2' & \dots & p_k' & \\ \text{valeur de } g_i \text{ dans } \underline{V}' : & h_1 & h_2 & \dots & h_k & \end{array}\right. \tag{31}$$

Dans ce cas on a $p_1 = p_2 = \dots = p_{n_1} = p_1'$, etc., et il est raisonnable de prendre aussi $g_1 = g_2 = \dots = g_{n_1} = h$, etc., comme l'indique la dernière ligne de (31).

La forme de $\underline{V}'$ [cf. (17)] devient maintenant

$$\underline{V}' = \sum_{\nu=1}^{k} h_\nu a_\nu \tag{32}$$

et d'après (27) le test est consistant si

$$L' = \lim_{n \to \infty} \sum_{\nu=1}^{k} n_\nu p_\nu' h_\nu = \pm\,\infty\,. \tag{33}$$

Les h_ν sont encore à choisir et maintenant il est utile de prendre des fonctions de $n_1, \dots, n_k$ (et non seulement de n)

$$h_\nu = h_\nu(n_1, \dots, n_k) \qquad (n = 1, \dots, k)\,. \tag{34}$$

Les relations (19), imposées sur les poids g_i, deviennent maintenant

$$\sum_{\nu=1}^{k} n_\nu h_\nu = 0\;; \qquad \sum_{\nu=1}^{k} n_\nu h_\nu^2 = 1\,. \tag{35}$$

La condition (22) pour la normalité asymptotique de la distribution de $\underline{V}'$ devient

$$\lim_{n \to \infty} \max_{0 < \nu \leqq n} |h_\nu| = 0 \tag{36}$$

et (23) subsiste. Quant aux hypothèses alternatives, (24) devient

$$\varepsilon < n^{-1} \sum_{\nu=1}^{k} n_\nu p_\nu' < 1 - \varepsilon \, . \tag{37}$$

Remarquons qu'en général on souhaitera que les alternatives (33) dépendent d'une manière déterminée seulement par les p_ν', mais pas des n_ν. Par exemple, si l'on veut tester $H_0'(p_1' = p_2' = \ldots = p_k')$ relativement à une augmentation ou une diminution des p_ν avec ν, caractérisée par

$$\Delta' = \sum_{\nu < \mu} (p_\nu' - p_\mu') < 0 \quad \text{ou} \quad > 0 \text{ respectivement} \, , \tag{38}$$

il suffit que L' de (33) soit $+\infty$ si $\Delta' < 0$ et $-\infty$ si $\Delta' > 0$. Car dans le premier cas on trouvera une grande valeur de V', indiquant une augmentation et dans le second une petite valeur de V', indiquant une diminution. C'est pour cette raison qu'il est nécessaire d'introduire (34) au lieu de $h_\nu = h_\nu(n)$. Si l'on avait des h_ν dépendants seulement de n on pourrait dans certains cas choisir les n_ν de telle manière, que L' deviendrait $+\infty$ ou $-\infty$ au choix, quel que soit le signe de Δ'. En effet, cela est possible s'il y a trois indices $\nu_1 < \nu_2 < \nu_3$ avec

$$(p_{\nu_2} - p_{\nu_1})\,(p_{\nu_2} - p_{\nu_3}) > 0 \, ,$$

c'est-à-dire si la tendance des p_i n'est pas tout à fait stricte.

Pour l'application il est très important d'éviter cette difficulté et on exigera, par exemple, que le test soit consistant relativement aux hypothèses satisfaisant

$$\sum_{\nu=1}^{k} w_\nu p_\nu' > 0 \text{ où } < 0 \qquad \left(\sum_{\nu=1}^{k} w_\nu = 0\right) \tag{39}$$

avec des w_ν constants. Les deux possibilités > 0 et < 0 correspondent dans ce cas à $L' = +\infty$ et à $L' = -\infty$ ou inversement. On peut atteindre ce but en choisissant les h_ν comme suit :

$$h_\nu = \varphi w_\nu n_\nu^{-1} \qquad (\nu = 1, \ldots, n), \tag{40}$$

où

$$\varphi = \varphi(n_1, \ldots, n_k) \tag{41}$$

est une fonction de $n_1, \ldots, n_k$, la même pour tous les ν, et positive pour $n(=n_1+\ldots+n_k)$ assez grand. On déduit de (35)

$$0 = \Sigma n_\nu h_\nu = \varphi \Sigma w_\nu ,$$

relation qui est satisfaite parce que $\Sigma w_\nu = 0$, et

$$1 = \Sigma n_\nu h_\nu^2 = \varphi^2 \Sigma w_\nu^2 n_\nu^{-1} ,$$

d'où

$$\varphi^2 = \left\{ \sum_{\nu=1}^{k} w_\nu^2 n_\nu^{-1} \right\}^{-1} . \tag{42}$$

Ensuite [cf. (33)]

$$L = \lim_{n\to\infty} \varphi \sum_{\nu=1}^{k} w_\nu p_\nu' \tag{43}$$

et alors la condition nécessaire et suffisante pour que (43) conduise à (39), est que

$$\lim_{n\to\infty} \varphi = \infty . \tag{44}$$

De (42) et (44) on obtient la condition, que pour chaque ν tel que $w_\nu \neq 0$:

$$\lim_{n\to\infty} n_\nu = \infty . \tag{45}$$

Considérons maintenant la condition (36). On voit qu'elle résulte de (45). Car

$$| h_\mu | = | w_\mu | \cdot \varphi n_\mu^{-1}$$

et pour chaque μ

$$\lim \varphi^2 n_\mu^{-2} = \lim \{ n_\mu^2 \Sigma w_\nu^2 n_\nu^{-1} \}^{-1} \leqq \lim \{ w_\mu^2 n_\mu \}^{-1} = 0 .$$

La condition (37) est très simple et n'a pas besoin de commentaire.

Une exposition plus complète de ce test est donnée par M^lle^ C. van Eeden (1954) et par C. van Eeden et J. Hemelrijk (1955).

6. Application au coefficient de capacité

Retournons maintenant au sujet du jugement des objets $C_1, \ldots, C_m$ et appliquons les résultats de la section 5 d'abord à la valeur typique $\underline{V}$ [cf. (2)] concernant un seul inspecteur. Pour satisfaire (19) il faut transformer les qualités Q_i en

$$Q_i' = (Q_i - \mu) / \sigma \sqrt{n} , \tag{46}$$

où μ et σ sont donnés par (8). Alors le test basé sur $\underline{V}$ est consistant, selon (27), si

$$\lim_{n \to \infty} \sum_{i=1}^{n} Q_i' p_i = \pm \infty . \tag{47}$$

Une situation intéressante se présente, lorsque les objets $C_1, \ldots, C_n$ forment un échantillon prélevé au hasard dans une population infinie d'objets. Sur cette population la qualité Q et la probabilité p d'être approuvé par l'inspecteur considéré ont une distribution combinée et les paires (Q_i, p_i) appartenant aux objets jugés sont des variables aléatoires indépendantes, chacune d'entre elles ayant cette distribution. Alors

$$\underline{\mu} = n^{-1} \sum_{i=1}^{n} \underline{Q}_i \qquad \text{et} \qquad \underline{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (\underline{Q}_i - \underline{\mu})^2$$

sont des variables aléatoires qui convergent en probabilité pour $n \to \infty$ vers leurs espérances mathématiques si un nombre suffisant des moments de la distribution de (Q, p) sont bornés. Indiquant par le symbole $\mathcal{E}$ l'espérance mathématique dans la population considérée, on a donc

$$\underline{\mu} \xrightarrow[\text{pr}]{} \mathcal{E}\underline{\mu} = \mathcal{E}\underline{Q} \quad \text{et} \quad \underline{\sigma}^2 \xrightarrow[\text{pr}]{} \mathcal{E}(\underline{Q} - \mathcal{E}\underline{Q})^2 . \tag{48}$$

De la même manière, on a

$$n^{-1} \sum_{i=1}^{n} \underline{Q}_i \underline{p}_i \xrightarrow[\text{pr}]{} \mathcal{E}\underline{Q}\underline{p} \quad \text{et} \quad n^{-1} \sum_{i=1}^{n} \underline{\mu}\underline{p}_i = \underline{\mu} n^{-1} \sum_{i=1}^{n} \underline{p}_i \xrightarrow[\text{pr}]{} \mathcal{E}\underline{Q} \cdot \mathcal{E}\underline{p} \tag{49}$$

Alors il résulte de (46), ..., (49) que

$$\underline{\sigma} n^{-\frac{1}{2}} \sum_{i=1}^{n} \underline{Q}_i' \underline{p}_i \xrightarrow[\text{pr}]{} \mathrm{Cov} \{ \underline{Q}, \underline{p} \} , \tag{50}$$

c'est-à-dire que le test est consistant, si le coefficient de corrélation de $\underline{Q}$ et $\underline{p}$ sur la population n'est pas égal à zéro. Ce résultat n'est pas étonnant, parce que le test peut, dans le cas considéré, être envisagé comme une application du test de Pitman (1937*b*) pour le coefficient de corrélation entre $\underline{Q}$ et $\underline{x}$, où $\underline{x}$ est une variable aléatoire adjointe à la population, qui, pour chaque élément, a la probabilité p, appartenant à cet élément, de prendre la valeur 1 et la probabilité $1 - p$ d'être égale à zéro [7]. En désignant l'espérance mathématique par

[7] Pitman n'a pas considéré, dans son article (1937*b*), la question de consistance ni celle de la normalité asymptotique de $\underline{V}$. D'autre part il donne des approximations de la distribution de $\underline{V}$ basées sur plus de deux moments, qu'on pourrait également utiliser ici.

rapport à x pour un élément fixe de la population par $\mathcal{E}_2$ et celle par rapport à l'échantillonnage par $\mathcal{E}_1$ on a pour $\underline{Q}$ et $\underline{p}$

$$\mathcal{E}_1 \underline{Q} = \mathcal{E}\underline{Q} \quad \text{et} \quad \mathcal{E}_1 \underline{p} = \mathcal{E}\underline{p} \quad \text{etc.},$$

et pour $\underline{x}$

$$\mathcal{E}\underline{x} = \mathcal{E}_1 \mathcal{E}_2 \underline{x} = \mathcal{E}_1 \underline{p} = \mathcal{E}\underline{p} .$$

Donc

$$\mathcal{E}\underline{Q}\underline{x} - \mathcal{E}\underline{Q}\mathcal{E}\underline{x} = \mathcal{E}_1 \{ \underline{Q}\mathcal{E}_2 \underline{x} \} - \mathcal{E}\underline{Q}\mathcal{E}\underline{p} = \mathcal{E}\underline{Q}\underline{p} - \mathcal{E}\underline{Q}\mathcal{E}\underline{p} .$$

Une corrélation de $\underline{Q}$ et $\underline{p}$ entraîne ainsi une corrélation du même signe entre $\underline{Q}$ et $\underline{x}$, soit que la dernière est plus petite parce que

$$\sigma^2\{\underline{x}\} > \sigma^2\{\underline{p}\} .$$

Si les objets $C_1, \ldots, C_n$ ne sont pas prélevés au hasard dans une population, mais choisis d'une manière systématique — ce qui peut être souhaitable dans la pratique — on peut exiger que ce choix soit exécuté de telle sorte que μ et σ aient des limites pour $n \longrightarrow \infty$ et alors le résultat est le même que le précédent : s'il y a, en limite, une corrélation positive ou négative entre les valeurs de Q et de p pour les objets choisis, le test est consistant.

Le deuxième cas, c'est-à-dire la comparaison des capacités de deux inspecteurs, est un peu plus compliqué, en raison de l'omission des paires de jugements identiques (cf. section 4). Seules les paires différentes comptent et en introduisant [cf. (14)]

$$p_i^* = p_{Ai} q_{Bi} / P_i \qquad (i = 1, \ldots, n), \tag{51}$$

on pourrait utiliser les résultats des paragraphes précédents en disant d'une manière peu précise, que le test est consistant s'il y a une corrélation entre les valeurs de Q et de p^* pour ceux des objets inspectés, qui ont reçu des jugements différents des inspecteurs A et B. Mais l'ensemble de ces objèts est évidemment un ensemble aléatoire et on peut essayer d'arriver à une formule plus complète, tenant également compte de cet aspect.

Considérons une population d'objets, sur laquelle trois variables aléatoires sont données : la qualité $\underline{Q}$ et les probabilités $\underline{p}_A$ et $\underline{p}_B$, que l'objet soit approuvé par les inspecteurs A et B respectivement. De plus ces variables vérifient les relations suivantes

$$\underline{P} = \underline{p}_A \underline{q}_B + \underline{p}_B \underline{q}_A \quad \text{et} \quad \underline{p}^* = \underline{p}_A \underline{q}_B / \underline{P} . \tag{52}$$

Si $\underline{z}$ est une variable aléatoire, qui prend la valeur 1 pour un élément jugé différemment par les deux inspecteurs (probabilité P) et qui est égale à zéro autrement, le nombre $\underline{m}$ d'éléments d'un échantillon, qui sont jugés différemment par les deux inspecteurs, est

$$\underline{m} = \sum_{i=1}^{n} \underline{z}_i \tag{53}$$

où les $\underline{z}_i$ sont distribués indépendamment selon la distribution de $\underline{z}$. La moyenne arithmétique et la variance des $\underline{Q}_j^*$ utilisés pour le test [cf. (12)] peuvent donc être notées comme suit

$$\underline{\mu}^* = \underline{m}^{-1} \sum_{i=1}^{n} \underline{z}_i \underline{Q}_i ; \qquad \underline{\sigma}^{*2} = \underline{m}^{-1} \sum_{i=1}^{n} \underline{z}_i (\underline{Q}_i - \underline{\mu}^*)^2 \tag{54}$$

et la consistance du test est déterminée maintenant par le comportement asymptotique de la quantité

$$(\underline{\sigma}^* \sqrt{\underline{m}})^{-1} \sum_{i=1}^{n} \underline{z}_i (\underline{Q}_i - \underline{\mu}^*) \underline{p}_i^* , \tag{55}$$

où les triples $(\underline{z}_i, \underline{p}_i^*, \underline{Q}_i)$ sont distribués indépendamment selon la distribution du triple $(\underline{z}, \underline{p}, \underline{Q})$ dans la population envisagée. Les $\underline{z}_i$ tiennent compte qu'une partie des objets jugés est omise parce que les jugements des deux inspecteurs sont identiques.

Sans trop de calculs nous remarquons que sous des conditions assez générales $\underline{\sigma}^*$ converge en probabilité vers une valeur bornée positive. D'ailleurs, comme plus haut, $\mathcal{E}\underline{z} = \mathcal{E}\underline{P}$ et

$$n^{-1}\underline{m} \underset{\text{pr}}{\rightarrow} \mathcal{E}\underline{P} ,$$

$$\underline{\mu}^* \underset{\text{pr}}{\rightarrow} \mathcal{E}\underline{PQ}/\mathcal{E}\underline{P} ,$$

$$n^{-1} \sum_{i=1}^{n} \underline{z}_i \underline{Q}_i \underline{p}_i^* \underset{\text{pr}}{\rightarrow} \mathcal{E}\underline{PQp}^* ,$$

$$n^{-1} \sum_{i=1}^{n} \underline{z}_i \underline{p}_i^* \underset{\text{pr}}{\rightarrow} \mathcal{E}\underline{Pp}^* ,$$

et alors (55) augmente indéfiniment avec n, si

$$\mathcal{E}\underline{P} \cdot \mathcal{E}\underline{PQp}^* - \mathcal{E}\underline{PQ} \cdot \mathcal{E}\underline{Pp}^* \tag{56}$$

est positif et (55) $\rightarrow -\infty$ si (56) est négatif.

L'interprétation de cette formule n'est pas aussi simple que dans le premier cas. L'expression (56) n'est pas équivalente à un coefficient de corrélation sur la population dont les objets ont été prélevés. Cependant il n'est pas difficile d'indiquer des hypothèses alternatives pour lesquelles le signe de (56) est connu.

Si, par exemple, p^* et P sont des fonctions de Q et si p^* augmente avec Q, le signe de (56) sera positif. Car, en désignant

$$P^* = p^*(Q) \quad \text{et} \quad \nu = \mathcal{E}\underline{PQ} / \mathcal{E}\underline{P}$$

on a dans ce cas

$$(Q-\nu)[p^*(Q)-p^*(\nu)] > 0 \quad \text{pour} \quad Q \neq \nu$$

et alors

$$\mathcal{E}(\underline{Q}-\nu)[p^*(\underline{Q})-p^*(\nu)]\underline{P} > 0 .$$

Mais le membre de gauche est égal à (56) à un facteur $\mathcal{E}\underline{P}$ près et $\mathcal{E}\underline{P}$ est positif.

Lorsque p^* et P ne sont pas des fonctions de $\underline{Q}$ mais des variables aléatoires pour chaque valeur fixe de $\underline{Q}$, il est évident que (56) soit, par exemple, positive, lorsque pour chaque P la corrélation conditionnelle de $\underline{Q}$ et $\underline{p}^*$, sous la condition $\underline{P}=P$, est positive.

Remarquons enfin que dans cette section nous nous sommes placés exclusivement au second point de vue des sections 3 et 4, c'est-à-dire au point de vue le moins général, avec la supposition de l'indépendance des jugements d'un inspecteur. Pour la situation plus générale, où cette supposition n'est pas avancée, d'autres méthodes d'investigation seront nécessaires pour trouver le domaine de consistance de ce test.

7. Remarques diverses

Il arrive souvent, que les qualités des objets à inspecter ne sont pas connues et qu'on peut seulement les déterminer d'une manière subjective. Ce qu'on peut faire alors c'est de juger les objets $C_1, \ldots, C_n$ par un nombre d'inspecteurs, un jury, et d'introduire, comme qualité Q, le nombre d'approbations qu'un objet a reçu de ce jury. Les coefficients W_1 et W_2 [cf. (6) et (9)] d'un inspecteur n'appartenant pas à ce jury sont alors des mesures de l'accord entre le jugement de cet inspecteur et celui du jury et le test pour la comparaison des capacités de deux inspecteurs décide lequel des deux inspecteurs, non membres du jury, est en meilleur accord avec le jury. Au lieu du nombre d'approbations reçu par un objet de la part du jury, on peut aussi prendre pour Q une fonction de ce nombre, ce qui ne change pas la théorie. Cela peut être utile dans certains cas. Il est possible, par exemple, qu'on ait une connaissance vague de la distribution de probabilité de certaines propriétés des objets, propriétés qui déterminent en effet la qualité mais qu'on ne peut pas observer sans détruire les objets ou sans les soumettre à des épreuves de longue durée. Le caractère non paramétrique de la méthode donne donc la possibilité d'incorporer, vaille que vaille, cette connaissance peu précise dans la mesure de la capacité d'inspecteurs. Cela peut se faire, dans le cas le plus simple, lorsqu'on connaît à peu près la forme de la fonction de répartition F de la qualité des objets $C_1, \ldots, C_n$. Si le jury est composé de m membres, le nombre w d'approbations peut prendre les valeurs $0, 1, \ldots, m$ et en supposant qu'il y ait $n_0, n_1, \ldots, n_m$ $(\Sigma n_i = n)$ objets pour lesquels $w = 0, 1, \ldots, m$ respectivement, on peut estimer les valeurs $Q_0, Q_1, \ldots, Q_m$ pour lesquelles

$$F\left(\frac{n_0}{2n}\right) = Q_0, \qquad F\left(\frac{2n_0 + n_1}{2n}\right) = Q_2 \text{ etc.},$$

et prendre ces valeurs pour les Q_i.

En ce qui concerne les généralisations des méthodes traitées, il y a plusieurs possibilités. Nous en indiquerons une seulement. Si l'on ne se borne pas à approuver ou à rejeter les objets mais si l'on admet plusieurs jugements, disons $J_1, \ldots, J_k$, on obtient le schéma suivant

		Jugements			
		J_1	J_2	$\ldots$	J_k
Objets	A_1	$\underline{x}_{11}$	$\underline{x}_{12}$	$\ldots$	$\underline{x}_{1k}$
	A_2	$\underline{x}_{21}$	$\underline{x}_{22}$	$\ldots$	$\underline{x}_{2k}$
	$\vdots$				
	A_n	$\underline{x}_{n1}$	$\underline{x}_{n2}$	$\ldots$	$\underline{x}_{nk}$

où les x_{ij} prennent la valeur 1 si le jugement sur l'objet C_i est J_j et 0 autrement. Il serait intéressant de rechercher les propriétés et les applications possibles de la valeur typique

$$\sum_{i,j} g_i h_j x_{ij} \tag{57}$$

où g_i et h_j sont des poids analogues aux g_i de la section 5. Une variable plus générale encore est obtenue en substituant des poids g_{ij} aux $g_i h_j$ dans cette formule. La généralisation des résultats donnés plus haut est aisée à imaginer.

Une autre généralisation, indiquée par M[lle] van Eeden, conduit à une modification du test de T. J. Terpstra (1952), test de l'hypothèse que plusieurs échantillons ont été prélevés dans des populations ayant la même distribution relativement à l'hypothèse alternative d'une tendance de ces populations quant à la variable aléatoire étudiée.

Bibliographie

van Eeden, C. (1954), *A test for the equality of probabilities against a class of specified alternative hypotheses, including trend*, Report S 157 (VP 3) of the Statistical Dept. of the Mathematical Centre, Amsterdam.

van Eeden, C. and Hemelrijk, J. (1955), *A test for the equality of probabilities against a class of specified alternative hypotheses, including trend* (*Proc. Kon. Ned. Ak. v. W.*, **A 58**, *Indagationes Mathematicae*, **17**, 191-193, 301-308).

Fisher, R. A. (1935), *The Design of Experiments*, Oliver and Boyd, London.

Kendall, M. G. (1948), *Rank Correlation Methods*, Griffin & Co., London.

Madow, W. G. (1948), *On the limiting distributions of estimates based on samples from finite universes* (*Ann. Math. Stat.* **19**, 535-545).

Neyman, J. (1950), *First Course in Probability and Statistics*, Holt and Co., N. Y.

Pitman, E. J. G. (1937*a*), *Significance tests which may be applied to samples from any populations*, I (*Jrn. Roy. Stat. Soc.*, Suppl. **4**, 117-130).

— (1937*b*), II. *The correlation coefficient test* (*Ibid.*, 225-232).

Terpstra, T. J. (1952), *The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking* (*Proc. Kon. Ned. Ak. v. Wet.*, **A 55**; *Indagationes Mathematicae*, **14**, 327-333).

Wald, A. and Wolfowitz, J. (1944), *Statistical tests based on permutations of the observations* (*Ann. Math. Stat.*, **15**, 358-372).

Wilcoxon, F. (1945), *Individual comparisons by ranking methods* (*Biometrics*, **1**, 80-82).