

REVUE DE STATISTIQUE APPLIQUÉE

J. RAISON

Les principaux tests non paramétriques. Quelques généralités et références bibliographiques

Revue de statistique appliquée, tome 7, n° 1 (1959), p. 83-106

http://www.numdam.org/item?id=RSA_1959__7_1_83_0

© Société française de statistique, 1959, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LES PRINCIPAUX TESTS NON PARAMÉTRIQUES

Quelques généralités et références bibliographiques

J. RAISON

Licencié en Sciences

Attaché à la Direction Générale de la Compagnie des Machines BULL

Les tests non paramétriques ont fait l'objet de nombreuses études publiées pour la plupart dans les revues spécialisées de langue anglaise. Seul l'article que Monsieur Morice a fait paraître en 1956 dans la Revue de Statistique Appliquée (Vol IV n° 4) tranche sur la très grande pauvreté de la littérature française en la matière.

Si, dans les conditions où les tests d'hypothèses classiques peuvent s'appliquer, les tests non paramétriques présentent une efficacité moindre, ils ont, par contre, par rapport à ces tests classiques de nombreux avantages et une plus grande généralité d'application. Il serait souhaitable que les tests non paramétriques fassent en France l'objet d'études plus nombreuses et d'un emploi justifié plus large.

Il peut être intéressant de procéder à un inventaire un peu poussé de la littérature qui les concerne et de donner, par une série d'articles dans la Revue de Statistique Appliquée, une suite à la publication de M. Morice. Un premier article permettra, après quelques généralités, de présenter les principaux tests non paramétriques. Sa rédaction très concise n'a pas pour but de fixer des règles d'emploi mais de définir seulement les problèmes visés par ces tests et de donner pour chacun d'eux quelques références bibliographiques. Des articles ultérieurs décriront plus particulièrement certains de ces tests ainsi que leurs conditions d'utilisation et leurs efficacités.

A. GENERALITES

1/ Au cours de la 28ème session de l'Institut International de Statistique à Rome (6-12 Septembre 1953), les Professeurs D. Van Dantzig et J. Hemelrijk présentèrent une communication intitulée "Statistical methods based on few assumptions" (Bulletin of the International Statistical Institute - Tome XXXIV, 2nd part). Ils distinguèrent dans le développement de la statistique mathématique 4 étapes :

la 1ère celle de John Graunt (1662) ou de la statistique à un paramètre. Exemple : Calcul par William Petty de la population des grandes villes européennes par application du rapport de mortalité.

la 2ème celle de Laplace (1778) et de sa loi normale ou de la statistique à deux paramètres.

la 3ème caractérisée par le fait que l'on constata que les distributions étaient rarement normales et que l'on essaya d'étudier les 3ème et 4ème moments.

Exemple : Pearson, Gram-Charlier etc. , ou statistique à plusieurs paramètres pour enfin arriver à :

la 4ème étape où l'on essaya de trouver des méthodes plus générales s'appliquant à des distributions quelconques ou statistique à 0 paramètre.

2/ Il ne faudrait pas croire, si le développement actuel des méthodes dites nonparamétriques a pris une extension considérable depuis la dernière guerre, que l'on n'avait pas essayé il y a déjà longtemps de déterminer des méthodes très générales.

- Todhunter : "A history of the mathematical theory of probability" (Cambridge and London 1865) cité par Cramer et Neyman, donne ainsi l'exemple d'application du test des signes dès 1710 par Arbuthnot.
- L'idée d'utiliser les rangs au lieu de valeurs mesurées n'est pas très neuve :
Kruskal et Wallis Jrn. Am. Stat. Ass. 47, pp. 584-621 (1952) "Use of ranks in one criterion variance analysis" remarquent en note que Laplace s'en était servi dès 1778.
- La loi de χ^2 était connue d'Helmert en 1876 sous la forme de la loi de la somme des carrés de variables normales réduites, elle fut redécouverte par K. Pearson en 1900 à propos d'un problème en apparence différent : ajustement d'un modèle théorique à une distribution observée. Yule et Fisher donnèrent en 1922 les moyens de déterminer les degrés de liberté.
- Un des plus anciens tests est sans doute également celui de Spearman (1906) "The proof and measurement of association between two things", étudié par Student en 1921 dans "An experimental détermination of the probable error of Dr. Spearman's correlation coefficient" Biometrika 13 - pp. 262-82.

3/ Définitions

On parle souvent de méthodes nonparamétriques ; dans la littérature anglo-saxonne, on rencontre les expressions :

distribution free
non parametric
rank invariant
order statistic...

Il est nécessaire de bien préciser ce que l'on entend dans les 2 premiers termes. Kendall dans son dictionnaire des termes statistiques donne la définition suivante de distribution free method :

Méthode " pour tester une hypothèse ou déterminer un intervalle de confiance qui ne dépend pas de la forme de la distribution en question".

Exemple : Intervalles de confiance pour la médiane basée sur une variation binomiale et variables pour toute distribution continue.

Dans les tests d'hypothèses cette expression signifie que le test est indépendant de la distribution pour l'hypothèse nulle.

On emploie souvent le mot non paramétrique, c'est une erreur. Il faut réserver le mot non paramétrique à la description d'hypothèses qui ne font pas d'affirmations explicites sur un paramètre.

Exemple : Non parametric tolerance limits.

Ce sont des limites de tolérance qui ne dépendent pas des paramètres de la population mère dont l'échantillon est tiré.

Il semble possible, bien que cela ne soit pas d'un usage universel, de faire une distinction entre des limites non paramétriques dans lesquelles la population mère est connue en forme, et "distribution free" (c'est-à-dire indépendante de la distribution) dans lesquelles la forme de la population mère est inconnue.

L'inégalité de Tchébycheff est ainsi indépendante de la distribution (distribution free) puisqu'elle est valable pour toute distribution, elle n'est pas strictement non paramétrique puisqu'elle fait intervenir μ et σ^2 .

4/ Avantages et inconvénients

a) L'on entend souvent parler du manque d'efficacité des méthodes non paramétriques par rapport aux méthodes classiques. Citons ici les mots de J. Wolfowitz (1949) "Nonparametric statistical inference, Proc. of the Berkeley Symposium, Berkeley and Los Angeles, pp. 93/113, rapportés par J. Hemelrijk.

... Si les formes fonctionnelles des fonctions de distribution sont connues, c'est une perte que de ne pas faire usage de cette information. Si l'on ne dispose pas de cette information, il faudra bien que le raisonnement statistique s'en passe. Dans ce dernier cas, la critique de quelques statisticiens qui prétendent que les méthodes non paramétriques ne sont pas efficaces ne tient pas, car l'efficacité au sens habituel du mot implique l'usage complet des ressources disponibles et il ne peut pas être inefficace de ne pas faire usage d'informations dont on ne dispose point...

Dans les cas où l'on peut effectuer des mesures dans des échelles où les opérations arithmétiques peuvent s'appliquer, l'on ne connaît pas toujours la forme de distribution de population et l'on est quelquefois obligé de travailler sur des échantillons de différentes populations. Les méthodes non paramétriques permettent de tirer des conclusions là où les méthodes classiques ne peuvent s'appliquer.

Dans bien des cas, on ne peut quantifier les opérations dans de telles échelles de mesures et l'on ne peut guère que faire des comparaisons de rangs : meilleur, pire, etc. Seules les méthodes non paramétriques peuvent s'appliquer dans ces cas où l'on n'a que des rangs.

Dans certains cas, il est plus simple et plus rapide d'appliquer des tests non paramétriques, la perte d'efficacité étant relativement faible et justifiant la méthode.

Pour des échantillons petits, les "statistiques"⁽¹⁾ sont en général plus faciles à calculer et l'application est plus aisée que dans le cas des méthodes classiques.

Le raisonnement employé dans les tests non paramétriques est en général plus simple que dans les tests classiques tandis que l'arrière plan mathématique est, par contre, assez difficile.

Les méthodes non paramétriques ne permettent cependant pas toujours la détermination d'intervalles de confiance et n'existent pas dans tous les cas où l'on est armé par les méthodes classiques. La détermination des rangs cause quelquefois une perte de temps et d'information.

b) Les questions qui se posent alors sont les suivantes :

Quand faut-il utiliser les tests non paramétriques ?

(1) Statistique est employé ici dans le sens de fonction dont la valeur numérique est calculée à partir des observations d'un échantillon.

Quels sont les tests à utiliser ?

De toute manière, il faut toujours penser aux frais de rassemblement et d'analyse des données avant de choisir un test.

Il faut aussi tenir compte :

- de la puissance du test, c'est-à-dire de la probabilité que l'on a de rejeter une hypothèse alternative,
- de l'efficacité de ce test, c'est-à-dire de la taille de l'échantillon qu'il est nécessaire de considérer pour être aussi puissant que le test le plus puissant,
- de l'hypothèse nulle et des hypothèses alternatives que l'on veut considérer,
- du niveau des mesures auquel on est parvenu dans la quantification des données (nominal, ordinal, d'intervalles, de rapport).

Si les hypothèses que l'on fait sur la forme de la distribution dans une population sont douteuses, tout raisonnement statistique basé sur ces hypothèses sera douteux. Ceci n'est pas le cas des méthodes non paramétriques. Dans ces méthodes, il y en a qui sont meilleures que d'autres.

Ainsi, pour les populations normales, le test des signes requiert de 4 à 50 % d'observations de plus pour être aussi puissant que le test t, le test de Wilcoxon n'en requiert que 4 % de plus.

On a effectué des comparaisons entre plusieurs tests non paramétriques pour des échantillons de différentes tailles, pour déterminer les efficacités relatives sur des distributions normales ou non. Un très grand nombre d'études a paru à ce sujet, leurs résultats et la comparaison des cas d'emploi devront compléter cette première partie sur les principaux tests non paramétriques.

Citons, parmi quelques-unes, les références suivantes :

MUNTER - Consistance de tests non paramétriques pour la comparaison d'échantillons. Acad. Roy. Bel. Bull. Cl. Sciences 1954, 40 n° 11 - pp. 1106-19.

SUNDRUM - On the relation between estimating efficiency and the power of tests. Biometrika G. B. 1954 - 41 - n° 3-4 pp. 542-4.

DIXON-TEICHROEW - Some sampling results on the power of non parametric tests against normal alternatives. Ann Math. Stat. U. S. A. - 1954 - 25 - n° 1 - pp. 175.

TEICHROEW - Empirical power functions for non parametric two sample tests for small samples. Ann. Math. Stat. U. S. A. - 1955 - 26 - n° 2 - pp. 340-4.

CHIA-KUEI-TSAO - Approximations to the power of rank tests. Ann. Math. Stat. U. S. A. 1956 - 27 - n° 2 pp. 548-9.

NOETHER - The efficiency of non parametric tests. Ann. Math. Stat. U. S. A. - 1957 - 28 - n° 2 pp. 532.

c) A côté des risques de 1er et 2ème ordre, Mosteller a proposé en 1948 un test non paramétrique pour déceler si une population, parmi k caractérisées par un paramètre de centrage, a bougé trop à droite ou trop à gauche. Il proposait de dénommer risque de 3ème ordre le risque de rejeter correctement l'hypothèse nulle pour une mauvaise raison.

5/ Références et littérature

a) Un très grand nombre d'articles ont paru à ces sujets dans les revues spécialisées. Signalons :

- l'article de M. Morice dans la Revue de Statistique Appliquée (1956) Vol IV n° 4 "Quelques tests non paramétriques" - pp. 75-107. 16 réf. C'est l'un des très rares articles français et le premier article général français où M. Morice présente les plus importants tests basés sur la somme ou la moyenne des rangs ainsi que sur le nombre et la longueur des suites homogènes, tests repris sub B.

S. WILKS "Order Statistics" Bulletin of the American Mathematical Society Vol 54 (1948) pp. 6-50.

A. MOOD indique dans son Introduction to Theory of Statistics que c'est à Wilks qu'une bonne part du crédit pris par ces méthodes doit revenir.

I. R. SAVAGE (1952) Bibliography of nonparametric statistics and related topics National Bureau of Standards - Rep. nr 1828. (72 pages) et Jrn. Am. Stat. Ass. (1953) 48 n° 264, pp. 844-906.

I. R. SAVAGE (1957) Jrn. of the Am. Stat. Ass. 52 - n° 279 pp. 331/44 (53 références) = "Non parametric statistics".

SCHEFFE (1943) Statistical inference in the non parametric case Ann. Math. Stat. 14 pp. 305-32.

WOLFOWITZ (1949) Non parametric statistical inference.

Proc. of the Berkeley Symposium on math stat and Prob. Un. of California Press - pp. 93/113 déjà cité.

MORAN

WHITFIELD

DANIELS

} (1950) Symposium on ranking methods
J. R. Stat. Society B - 12 - 153/91.

LEHMANN (1951) Consistency and unbiasedness of certain non parametric tests Ann. Math. Stat. Vol 22 - p. 165.

D'une façon générale, on trouve régulièrement des articles dans les revues suivantes :

Biometrika

Journal of the American Statistical Association

Annal of Mathematical statistics

Sankhyà

Rapports spécialisés du Centre mathématique d'Amsterdam (où D. Van Dantzig et J. Hemelrijk se sont particulièrement penchés sur les méthodes non paramétriques, et auxquels nous sommes particulièrement redevables).

b) Les ouvrages regroupant ces méthodes sont rares. Citons :

M. G. KENDALL "Rank correlation methods" - 2ème Edition - Londres Griffin and Co. 196 pages (1955) à recommander.

Sidney SIEGEL "Non parametric statistics for the behavioral sciences" Mc. Graw Hill 1956. 312 pages dont 55 pages de tables et une centaine de références. Très clair, ce livre fait penser à l'ouvrage de Dixon et Massey. "Introduction to statistical analysis". Exposé simple et correct. L'auteur ne parle pas des intervalles de confiance et des régressions non paramétriques. L'auteur est professeur à l'Université de Pennsylvanie.

D. A. S. FRASER " Non parametric methods in statistics" - Wiley (1957) 299 pages. Aux U. S. A. , Hoeffding et Lehmann entre autres, ont publié, en général dans les "Annals of Mathematical Statistics", toute une série d'articles assez difficiles. L'ouvrage de Fraser est théorique, assez difficile et résume tout ceci.

Quelques tests donnés en exercices, 80 références dont 2 seulement celles concernant Gini et Kendall, ont trait à des spécialistes non américains. L'auteur est professeur à l'Université de Toronto.

M. G. KENDALL et R. H. SUNDRUM "Distribution free methods and order properties".

c) Notons enfin le Bulletin signalétique du C.N.R.S. "Statistique mathématique".

B. PRINCIPAUX TESTS STATISTIQUES

Les principaux tests seront énumérés ci-après en résumant rapidement leurs objets, leurs références bibliographiques principales et en indiquant les tables principales existant. actuellement.

1/ Test de χ^2

Ce test bien connu est utilisé en particulier pour comparer une distribution théorique avec les résultats observés en pratique. On peut en trouver les conditions d'application dans tout ouvrage de statistique.

La condition généralement imposée est que la fréquence théorique de chaque classe soit > 5 . Cochran a montré que l'on pouvait descendre à 1 pour 20 % des classes si l'on adopte un seuil de confiance de l'ordre de 2 à 5 %.

Une application intéressante est celle des tables de contingence, en particulier des tables 2×2 . Il ne faut pas oublier d'appliquer la correction de continuité dite de Yates. Dans une table 2×2 de la forme :

$$\begin{array}{c|c} A & B \\ \hline C & D \end{array}$$

la valeur de χ^2 est égale à l'expression corrigée suivante :

$$\chi^2 = \frac{N (| AD - BC | - \frac{N}{2})^2}{(A+B) (C+D) (A+C) (B+D)} \quad \text{d. d. l.} = 1$$

Réf : COCHRAN (1950) The comparison of percentages in matched samples - Biometrika 37 - pp. 256-60.

(1952) The χ^2 test of goodness of fit
Ann. Math. Stat. 23 - pp. 315/45

(1954) Some methods for strengthening the common χ^2 test
Biometrics 10 - pp. 417/51

LEWIS et BURKE (1949) - The use and misuse of the chi square test. Psych. Bull. 46 - pp. 433-89.

A. VESSEREAU (1958) - Sur les conditions d'application du critérium χ^2 de Pearson - Revue de statistique appliquée Vol VI n° 2.

2/ Test de Fisher

Le test de probabilité exacte de Fisher est utilisé quand les scores de 2 échantillons indépendants de variables discrètes tombent dans l'une ou l'autre de deux classes s'excluant mutuellement. Le test détermine si les deux échan-

tillons diffèrent par les proportions dans lesquelles ils sont répartis dans les 2 classes.

Si l'on a le tableau ci-après :

<table style="border-collapse: collapse; margin: auto;"> <tr> <td style="border-right: 1px dashed black; padding: 2px 5px;">A</td> <td style="padding: 2px 5px;">B</td> </tr> <tr> <td style="border-right: 1px dashed black; padding: 2px 5px;">C</td> <td style="padding: 2px 5px;">D</td> </tr> <tr> <td style="border-top: 1px solid black; border-right: 1px solid black; padding: 2px 5px;">A+C</td> <td style="border-top: 1px solid black; padding: 2px 5px;">B+D</td> </tr> </table>	A	B	C	D	A+C	B+D	<table style="border-collapse: collapse; margin: auto;"> <tr> <td style="border-top: 1px solid black; padding: 2px 5px;">A + B</td> </tr> <tr> <td style="border-top: 1px solid black; padding: 2px 5px;">C + D</td> </tr> <tr> <td style="border-top: 1px solid black; padding: 2px 5px;">N</td> </tr> </table>	A + B	C + D	N
A	B									
C	D									
A+C	B+D									
A + B										
C + D										
N										

pour des totaux marginaux fixes, la probabilité exacte d'observer une combinaison particulière est donnée par la distribution hypergéométrique :

$$p = \frac{(A + B)! (C + D)! (A + C)! (B + D)!}{N! A! B! C! D!}$$

Il faut évidemment sommer les probabilités individuelles pour obtenir les seuils de confiance, ce qui n'est possible que pour des échantillons petits.

$$\Pr (A \leq A_0) = \sum_{A=0}^{A_0} p$$

Pour $N \leq 30$ et les totaux marginaux < 15 , Finney a publié une table.

"The Fisher Yates test of significance in 2 x 2 contingency tables"

Biometrika (1948) 35 - pp. 149-54 (niveaux de confiance 0,05 0,025 0,01 0,005).

Ce test est à employer dans le cas des tables 2 x 2 si les conditions de Cochran ne sont pas satisfaites ou si $N < 20$.

3/ Dans bien des cas, l'on désire vérifier l'accord entre une distribution empirique et une distribution théorique ou bien entre deux distributions empiriques.

DRION "Some distribution free tests for the difference between two empirical cumulative distribution function".

Ann. Math. Stat. U. S. A. - 1952 - 23 - n° 4 - pp. 563/74 donne une réponse partielle au problème d'André cité dans l'ouvrage de Bertrand : "Calcul des Probabilités", Paris 1907, dans le cas où, des 2 cumuls successifs des distributions empiriques, l'un se trouve toujours en dessous de l'autre.

KOLMOGOROV (1933) et SMIRNOV (1939) ont développé les études de Cramer (1928) et Von Mises (1931) en étudiant la valeur maximum des différences des fonctions de répartition :

$$D = \max [S_{n_1} (X) - S_{n_2} (X)]$$

dans le cas de deux distributions empiriques ou de :

$$D = \max [F_0 (X) - S_n (X)]$$

dans le cas d'une comparaison d'une fonction de répartition empirique $S_n (X)$ à une fonction de répartition théorique $F_0 (X)$.

Le Professeur Darmois, dans la préface qu'il écrivait au traité de statistique théorique et pratique de D. Dugué (Dunod 1958), dénommait cette loi de Kolmogorov Smirnov l'un des bijoux de la statistique.

- Comparaison à une répartition théorique.

Tables :

MASSEY (1951) The Kolmogorov-Smirnov test for goodness of fit Jrn Am. Stat. Ass. 46 - pp. 68/78
N = 1 (1) 20 (5) 35 niveaux 0, 20 0, 15 0, 10 0, 05 0, 01

BIRNBAUM (1952) Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample values. Jrn. Am. Stat. Ass. 47 - pp. 425-441.

(1952) Distribution free tests of fit for continuous distribution functions.

Am. Math. Stat. 23 n° 3 pp. 481-2.

(1953) - d° - 24 n° 1 pp. 1-8.

MILLER (1956) Table of percentage points of Kolmogorov's statistic Jrn. Am. Stat. Ass. 51 - n° 273 - pp. 111-21.

- Comparaison de 2 répartitions théoriques :

petits échantillons le problème n'a été résolu que pour $n_1 = n_2 \leq 40$:

MASSEY : The distribution of the maximum deviation between two sample cumulative step functions.

Ann. Math. Stat. (1951) 22 - pp. 125-8.

Table donnant la déviation maximum au seuil de confiance 0,05 et 0,01.

grands échantillons Table de Smirnov ($\alpha = 0,10$ à $0,001$)

Table for estimating the goodness of fit of empirical distribution Ann. math. Stat. (1948) - 19 - pp. 279/81. Zones critiques bilatérales.

GOODMAN a déterminé des zones critiques unilatérales: "Kolmogorov Smirnov tests for psychological research" - Psych. Bull. (1954) 51 pp. 160-8.

- Voir Kolmogorov (1941) Confidence limits for an unknown distribution function. Ann. Math. Stat. 12 - pp. 481-3.

4/ Tests dits "Randomization tests"

- Paires associées à la suite d'un traitement, les paires A_i, B_i présentent une différence $A_i - B_i = d_i$.

On fait, pour chacune des paires l'hypothèse que le signe de la différence est donné au hasard (H_0).

On détermine alors les probabilités de dépassement de la somme trouvée $\sum d_i$.

Réf : FISHER (1935) - The design of experiments

PITMAN (1937) - Significance tests which may be applied to samples from any populations

Suppl. J. R. Stat. Soc. 4 119-130
225-232

Biometrika 29 - pp. 322-335.

Si le nombre de paires n'est pas petit, le procédé devient rapidement impraticable et il faudra utiliser le test de symétrie de Wilcoxon.

- Cas de deux échantillons indépendants n_1 et n_2

Le test est utilisé pour déterminer si les moyennes des 2 échantillons diffèrent significativement.

Les $(n_1 + n_2)$ observations peuvent être réparties entre les 2 échantillons de $C_{n_1+n_2}^{n_1}$ façons en admettant que les deux échantillons ne diffèrent pas et que les scores sont divisés entre les deux échantillons, tout à fait au hasard. On détermine donc les probabilités de dépassement du cas actuel et l'on rejette ou accepte l'hypothèse de 0 différence.

Réf : PITMAN déjà cité.

Pour n_1 et n_2 grands et dans certaines conditions, on obtient une distribution de Student.

Voir aussi DWASS : Modified Randomization tests for non parametric hypotheses Ann. Math. Stat. U. S. A. 1957 - 28 - n° 1 - pp. 181-7.

Si n_1 et n_2 sont grands, il est nécessaire de passer au test W de Wilcoxon.

- Signalons dans ce domaine Hoeffding "The large sample power of tests based on permutations of observations" - Ann. Math. Stat. U. S. A. (1952) - 23 n° 2 - 169-92.

5/ Test de la médiane

a) Voir par exemple Mood "Introduction to the theory of statistics" pp. 388 et suivantes.

Les intervalles de confiance de la médiane peuvent être déterminés par des calculs simples de répartition binomiale.

Table pour $N = 6 (1) 65$ ($\alpha \leq 0,05$ et $\alpha \leq 0,01$)

Voir aussi :

K. R. NAIR "Table of confidence interval for the median in samples from any continuous population" (Sankhya vol 4 (1940) pp. 551-8).

CHU, HOTELLING "The moments of the sample median" - Ann. Math. Stat. U. S. A. 1954 - 25 - n° 4 - 809.

b) Si l'on a 2 échantillons $x_1 x_2 \dots x_{n_1}$

$y_1 y_2 \dots y_{n_2}$

et que l'on veuille déterminer s'ils proviennent de la même population, une solution connue à ce problème des 2 échantillons est décrite par exemple par : J. Westenberg "Significance test for the median and interquartile range in samples from continuous populations of any form - Proc. Kon. AK. v. Wet 51 - pp.252-61. En rangeant les observations dans leur ordre numérique et en comptant le nombre des observations u des observations x_1 à x_{n_1} se trouvant dans le groupe des grandes valeurs (les grandes valeurs étant séparées par des petites par le nombre r déterminé en comptant depuis la plus petite valeur par exemple la moitié plus un, des observations $n_1 + n_2$ suivant que $n_1 + n_2$ est pair ou impair, et en nous arrêtant tout de suite après le 1er groupe de valeurs égales), on a :

$$P [\underline{U} = u \mid \underline{R} = r ; H_0] = \frac{C_u^r \times C_{n-r}^{n_1-u}}{C_n^{n_1}} \quad \text{où } n = n_1 + n_2$$

c) Hemelrijk a généralisé cette méthode : thèse de Doctorat Un. d'Amsterdam 1950 et "A family of parameter free tests for symmetry with respect to a given point" Proc. Kon. Ned. AK. v. Wet 53 - pp. 945-55 et 1186-98 (Tables) par son test des 2 quantiles où il détermine les nombres u_1 et u_2 des observations x_1 à x_{n_1} se trouvant entre la plus petite observation et la r_1 ème observation d'une part, entre la plus grande observation et la $n - r_2$ ème observation d'autre part, des 2 échantillons.

d) Walsch a établi un test intéressant : "Applications of some significance tests for the median which are valid under very general circumstances" *Jrn. Am. Stat. Ass.* 1949 - 44, pp. 342-55 - Tables. Il s'agit dans un groupe de paires d'observations de déterminer après un traitement les différences d_i entre les paires et de les ranger en ordre croissant $d_1 < d_2 < \dots < d_n$. Ce que le test suppose, c'est que les populations sont symétriques et que, par conséquent, la moyenne représentant la tendance centrale est égale à la médiane.

Les tables établies pour $N = 4(1)15$ permettent de rejeter l'hypothèse H_0 en utilisant des zones critiques unilatérales ou bilatérales et en comparant les différents d_i trouvés suivant les indications que ces tables fournissent. Le test est compliqué. Consulter aussi l'article de Walsh "Some significance tests for the median which are valid under very general conditions". *Ann. Math. Stat.* 17 - 358-62.

e) Dans le cas de 2 ou k échantillons on peut déterminer la médiane commune, dichotomiser par rapport à sa valeur et effectuer un χ^2 ou dans le cas 2 x 2 un test de Fisher.

6/ Tests sur les suites (runs)

Si l'on range des éléments d'espèces différentes x et y on appelle suite toute séquence partielle constituée par un ou plusieurs éléments de même espèce. Exemple : le rangement :

xx y x yyy xx

comporte 5 suites. Le rangement peut être ainsi fait par ordre croissant ou par ordre chronologique.

a) Si l'on a deux échantillons en provenance d'une même population, les x et y seront bien mélangés et il y aura beaucoup de suites.

On comprend que l'on puisse tester l'appartenance des deux échantillons à la même population en les rangeant en une série par ordre de valeur croissante et en comptant le nombre de suites.

$$\begin{aligned} \text{si } n_1 \text{ et } n_2 > 10 \quad n_1 + n_2 &= n \\ n_1 &= n\alpha \\ n_2 &= n\beta \end{aligned}$$

si d est le nombre de suites, la quantité :

$$t = \frac{d - 2n\alpha\beta}{2\alpha\beta\sqrt{n}}$$

est distribuée normalement (0, 1). Ce test est sensible à la fois aux différences de forme et de centrage entre les deux distributions (ce point sera repris sous 2). On trouve ici un résultat fréquemment rencontré dans les méthodes non paramétriques : à savoir que les distributions des statistiques utilisées tendent fréquemment vers la normalité. Voir à ce sujet par exemple :

DWASS "On the asymptotic normality of some statistic used in non parametric tests". *Ann. Math. Stat. U. S. A.* 1955 - 26 - n° 2 - pp. 334-9.

MOOD - The distribution theory of runs - 1940 - *Ann. Math. Stat.* Vol II - pp. 367-92. - Introduction to the theory of statistics.

MOSTELLER - Note on an application of runs to control charts. *Ann. Math. Stat.* 12 (1941) pp. 228-32.

WOLFOWITZ - cités par M. Morice.
 OLMSTEAD

HALD - Statistical theory with engineering applications pp. 372.

Tables de SVED et EISENHART "Tables for testing randomness of grouping in a sequence of alternatives" - Ann. Math. Stat. 14 pp. 66-87 - $n_1 = n_2 = 2(1)20$

Ce test est quelquefois appelé le test de Wald-Wolfowitz :

"On a test whether two samples are from the same population. Ann. Math. Stat. II - 1940 - 147-62.

b) On peut également utiliser la longueur des suites comme test. $R_{x,k}$ est une suite d'éléments x de longueur k .

Olmstead a donné des tables "Distribution of sample arrangements for runs up and down. - Ann. Math. Stat. Vol 17 (1946) pp. 24-33.

On a : $E(R_{x,k}) = V(R_{x,k}) = \frac{N}{2k+1} = \lambda$ ceci suggère une distribution de Poisson :

$$P [R_{x,k} = u] = e^{-\lambda} \frac{\lambda^u}{u!}$$

c) On peut encore considérer, dans une suite chronologique de n observations successives, les différences $x_i - x_{i-1}$

soit $x_1 \ x_2 \ \dots \ x_n$ n observations (série chronologique)

on a $x_2 - x_1, x_3 - x_2, \dots, x_n - x_{n-1}$ soit $n - 1$ différences.

si S est le nombre des différences négatives.

$$\text{On a} \quad E(S) = \frac{n-1}{2}$$

$$V(S) = \frac{n+1}{12}$$

et la distribution de S devient normale si $n \geq 12$.

MOORE et WALLIS - Time series significance tests based on sign of differences
 Jrn. Am. Stat. Ass. Vol 38 (1943) pp. 153-64.

H. B. MANN - On a test of randomness based on sign of differences. Ann. Math. Stat. Vol 16 (1945) pp. 200-4.

d) On peut enfin considérer le nombre et la longueur des suites de différences de même signe

$$x_j - x_{j-i} \text{ est } > 0 \text{ ou } < 0$$

si r_i nombre de suites de longueur i

$$R_k = \sum_{i=k}^{n-1} r_i = \text{nombre de suites de longueur } \geq k$$

le nombre total de suites R est distribué normalement si $n > 20$:

$$E(R) = \frac{1}{3} (2n - 1)$$

$$V(R) = \frac{1}{90} (16n - 29)$$

Si $k \geq 5$ on a $V(R_k) \simeq E(R_k)$ ce qui autorise à envisager une loi de Poisson pour les suites de longueur $k \geq 5$ avec :

$$P(R_k = u) = e^{-\lambda} \frac{\lambda^u}{u!} \quad \text{où } \lambda = E(R_k)$$

L'utilisation du nombre de suites et de leur longueur peut permettre de tester des tendances dans la variation de x . Voir Hald déjà cité pp. 353-4.

e) Signalons le test de Ramachandran "A non parameter two samples test" Jrn. Madras. Univ. Sect B. (1953) 23 - n° 1 - 70-91, où les sommes des carrés des longueurs des suites de succès sont considérées.

7/ Test des signes

Si l'on veut comparer 2 matières ou 2 traitements dans des conditions d'utilisation différentes, on rassemblera des paires d'observations, les observations de chaque paire étant faites dans les mêmes conditions, conditions variant d'une paire à l'autre. On teste l'hypothèse H_0 que chacune des différences entre les résultats des 2 matières ou traitements a une distribution de probabilité dont la médiane est nulle, ce qui veut dire qu'il y a en moyenne autant de différences positives que de différences négatives. On ne s'occupe pas de la valeur de ces différences, mais uniquement de leur signe. On rejettera l'hypothèse H_0 quand le nombre de différences positives ou négatives diffère significativement de l'égalité.

Lorsqu'il y a des égalités, il convient de ne pas tenir compte des paires d'observations où elles se produisent.

Ce test revient à un test binomial $p = 0,5$.

Réf : W. J. DIXON et A. M. MOOD "the statistical sign test". Jrn. Am. Stat. Ass. 41 (1946) 557-66.

(Tables pour $n = 1(1)100$ $\alpha = 0,01$ 0,05 0,10 0,25)

MAC STEWART "A note on the power of the sign test" Ann. Math. Stat. Vol 12 (1941) pp. 236-8.

L'efficacité de ce test qui doit son nom à Fisher, par comparaison au test t (quand il s'applique), est de 64 % pour les échantillons importants. L'examen des courbes du pouvoir discriminant indique que le test des signes utilisant 18 paires est sensiblement équivalent au test t utilisant n paires.

Réf : Graphiques Annales des Mines Février 1957.

- Revue de Statistique Appliquée, vol V, n° 3 pp. 85-98
 - les rayures dans les tissus de popeline et le test des signes
 - Note complémentaire théorique.
- "A simplification of the sign test" WEINBERG et TRIPP Psy. Bull 54 (1957) pp. 79-80
- Signalons le "bivariate sign test" de HODGES Ann. Math. Stat. U. S. A. 1955 - 26 - n° 3 - 523-7

8/ Le test de symétrie de Wilcoxon

Le test des signes utilise l'information relative à la direction des différences entre paires. Si l'on utilise à la fois leur grandeur relative et leur direc-

tion, on a un test plus puissant. C'est ce que fit Wilcoxon en donnant plus d'importance à une paire montrant une grande différence entre les 2 conditions qu'à une paire montrant une faible différence.

Cette différence d_i est le score de chaque paire.

On range les d_i en ordre croissant par valeur absolue et on leur fait correspondre un rang de 1 à n .

Le rang est affecté du signe + pour les $d_i > 0$

" " " " " " - " " $d_i < 0$

s'il y a des ex-aequo : si $d_i = 0$, on néglige la paire
si plusieurs d_i sont égaux, on leur donne la moyenne des rangs qu'ils auraient eu.

La statistique T est égale à la somme des rangs des d_i positifs diminuée de la somme des rangs des d_i négatifs.

L'hypothèse de symétrie est rejetée si la valeur absolue de T est \geq à une valeur critique donnée par des tables (zone critique bilatérale).

Ce test peut d'ailleurs être combiné avec le test des signes.

Réf : MOOD (1954) On the asymptotic efficiency of certain non parametric two sample tests. Ann. Math. Stat. 25 - 514-22.

WILCOXON F (1945) - Individual comparison by ranking methods Biometrics Bull. 1, 80-83.

(1947) - Probability table for individual comparisons by ranking methods Biometrics Bull. 3 - 119-122

Ces deux publications contiennent quelques erreurs rectifiées dans :

(1949) - Some rapid approximation statistical procedure, Stanford Conn. American Cyanamid Co.

Rapport S 208 (M 76) et tables du Centre Mathématique d'Amsterdam (1956) - 33 pages dont 10 pages de tables et 1 page de références. ✕

Wilcoxon prend la somme des rangs positifs (cf. tables de Dixon et Massey) Il est préférable d'employer T car T sera entier et $E(T) = 0$.

$$\sigma^2 = \frac{1}{6} n(n+1)(2n+1) - \frac{1}{12} (D-n),$$

D étant une quantité dépendant du nombre d'ex-aequo.

Cette valeur T se relie d'ailleurs à la valeur W de Wilcoxon que l'on peut calculer en considérant comme 1er échantillon les valeurs positives et comme 2ème échantillon les valeurs négatives.

9/ Le test W de Wilcoxon

Ce test est assez souvent connu sous le nom de test U de Mann et Whitney ($U = \frac{1}{2} W$).

Quand on veut comparer les distributions de 2 variables on est naturellement conduit à prendre une observation de chaque et à déterminer la plus grande (échantillon $x_1 \dots x_n$ et échantillon $y_1 \dots y_m$). La statistique W est le double du nombre de paires d'observations (x_i, y_j) pour lesquelles $x_i > y_j$ augmenté du nombre de paires (x_k, y_l) pour lesquelles $x_k = y_l$.

Si les échantillons proviennent de la même population, la statistique W est distribuée normalement pour n_1 , et n_2 grands, moyennant certaines précautions quant au nombre d'ex-aequo, avec :

$$E(W) = mn$$

$$V(W) = \frac{mn(N^3 - D)}{3N(N-1)} \quad \text{où D dépend du nombre d'ex-aequo}$$

pour $m \leq n \leq 8$, Mann et Whitney ont calculé les probabilités de dépassement de : U : "On a test whether one of two random variables is stochastically larger than the other" - Ann. Math. Stat. Vol 18 - (1947) 50-60,

pour $m \leq n \leq 10$ le Centre Mathématique d'Amsterdam, Rapport MC 176 (1955) - 37 pages - a calculé ces probabilités de dépassement pour la statistique W (H. R. Vander Vaart) et donné les tables pour :

$$\begin{array}{l} m = 1(1)20 \\ n = 11(1)35 \end{array} \quad \alpha = \begin{array}{l} 0,01 \\ 0,025 \\ 0,05 \end{array} \left\{ \begin{array}{l} \text{Zone critique unilatérale} \end{array} \right.$$

Un abaque de Rijkoort permet de comprimer les tables $m \leq n \leq 10$.

Voici quelques autres références :

HALDANE The Wilcoxon and related tests of significance - Experientia Science 1956 - 12 - n° 6 - p. 205.

KRUSKAL Historical notes on the Wilcoxon unpaired two sample test - Jrn. Am. Stat. Ass. 52 (1957) n° 279 pp. 356-60.

FIX HODGES - Significance probabilities of the Wilcoxon test (tables) - Ann. Math. Stat. 1955 - n° 26 n° 2 pp. 301-312.

H. R. Vander VAART - A closed expression for certain probabilities in Wilcoxon's two samples test. - Experientia Science 1956 - 12 n° 1, pp. 14-15.

J. HEMELRIJK - Note on Wilcoxon's two sample test when ties are present - Ann. Math. Stat. (1952) 23 - n° 1 - pp. 133-5.

Au lieu de calculer la statistique W (ou U) à partir des inversions, on peut très bien en rangeant les deux échantillons dans l'ordre croissant totaliser la somme des rangs pris par les observations d'un échantillon. Cette somme est reliée très simplement à la quantité W. Colin White a tabulé cette somme :

"The use of ranks in a test of significance for comparing two treatments" Biometrics Vol 8 - n° 1 - 1952.

Signalons dans le même ordre d'idées :

CHIA KUEI TSAO - Distribution of the sum in random samples from a discrete population" - Ann. Math. Stat. U. S. A. (1956) 27 - n° 3 pp. 703-12.

à rapprocher de son article :

"Sequential rank sum tests" - Ann. Math. Stat. U. S. A. (1954) 25, n° 1, 177.

Le pouvoir discriminant du test de Wilcoxon a été étudié en particulier par Van der Vaart et son efficacité par Noether et Pitman.

10/ Le test X de Van der Waerden

"Ein neuer test für das problem der zwei stich-proben" - Math. Ann. Deuts. 126 S (93) 1953.

"Tafeln zum Vergleich zweier Stichproben mittels X test und Zeichentest"
Springer Verlag 1956 - Bl Van der Waerden et E. Nievergelt.

Pour n petit, le test de Wilcoxon a l'inconvénient qu'il y a fréquemment plusieurs arrangements avec le même nombre d'inversions et on est amené soit à prendre α trop grand, soit β trop grand.

Ψ étant la fonction inverse de la fonction de distribution normale $\Phi(z)$ on range les valeurs des deux échantillons en ordre croissant en leur affectant les rangs de 1 à n .

On calcule $X = \sum \Psi\left(\frac{r}{n+1}\right)$ pour les g observations $x_1 \dots x_g$

Quand cette somme est $\geq X_\beta$ on rejette H_0

pour n petit, la valeur est calculée,

" n grand, la distribution de X est asymptotiquement normale ; caractérisée par :

moyenne 0

variance $\sigma^2 = \frac{gh}{n-1} Q$ avec

$$Q = \frac{1}{n} \sum_{r=1}^n \Psi^2\left(\frac{r}{n+1}\right)$$

g et h étant les tailles des 2 échantillons.

NOETHER "Math Rev. 15 46 "a étudié la normalité asymptotique également valable au cas où g et h ne restent pas bornés. Tables $\Psi\left(\frac{r}{n+1}\right)$ pour $n = 6(1) - 50$. Puissance du test : pour g fixe et $n \rightarrow \infty$ le test X a le même pouvoir discriminant que le test de Student.

D'autres statistiques ont été proposées pour traiter ce problème des deux échantillons :

TERRY "Two rank order tests which are most powerful against specific parametric alternatives" - Ann. Math. Stat. 23 (1952) - 246-66.

PITMAN "Significance tests which may be applied to samples from any populations - Jrn. Roy. Stat. Soc. Suppl. 4 - 119-129.

11/ Test de Moses

Il est utilisé dans le problème des échantillons lorsqu'on pense qu'une cause va provoquer des effets en sens contraires.

On combine les scores (expérimentaux et de contrôle) en une série et l'on compte l'étendue (span) entre la plus basse et la plus haute observation du groupe de contrôle (valeurs extrêmes comprises) par utilisation des rangs.

MOSES "A two sample problem" - Psychometrika (1952) 17 - 239-47.

12/ Test de Kruskal et Wallis, Test de Terpstra

a) Test de Kruskal et Wallis

C'est une généralisation du test de Wilcoxon au problème de k échantillons indépendants. L'hypothèse H_0 est que les k échantillons proviennent de la même population ou de populations identiques en ce qui concerne leurs moyennes.

On détermine les rangs de chacune des observations pour les k échantillons de 1 à N.

On cherche à voir si la somme des rangs est si disparate qu'il n'est pas vraisemblable que les échantillons soient tirés d'une même population.

k = nombre d'échantillons,

n_j = nombre d'observations du j^{ème} échantillon.

$N = \sum n_j$

R_j = somme des rangs du j^{ème} échantillon.

Si n_j est suffisamment grand, supérieur à 5 on calcule :

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$$

et l'on démontre que cette statistique suit une loi de χ^2 à k - 1 d. d. l, si les n_j sont suffisamment grands. Pour tenir compte des valeurs ex-aequo on divisera H par $1 - \frac{\sum T}{N^3 - N}$, avec $T = t^3 - t$ pour chaque groupe de t rangs égaux.

S'il n'y a pas plus de 25 % d'ex-aequo, la probabilité associée ne varie pas de plus de 10 %.

Les probabilités exactes ont été calculées et tabulées pour k = 3, $n_j \leq 5$.

KRUSKAL et WALLIS "Use of ranks in one criterion variance analysis" - Jrn. Am. Stat. Ass. 47 - (1952) 583-621 et correction 48 p. 910.

Voir également :

KRUSKAL (1952) "A non parametric test for the several sample problem" - Ann. Math. Stat. 23 - 525-540.

P. J. RIJKOORT "A generalization of Wilcoxon's test" - Kon. Ned. Ak. v. Wet. Proc. A55 (1952).

b) Test de Terpstra, test T^2

TERPSTRA "Anonparametric k sample test and its connection with the H test". Rep. S. 92 - VP2 - MC - Amsterdam 1952.

Dans ce test l'on opère de façon analogue à celle prise pour le test H mais l'on compare de plus 2 à 2 les échantillons.

Les grandeurs U sont les statistiques de Mann et Whitney, soient :

$$\tilde{U} = U - \frac{1}{2}mn \text{ si } m \text{ et } n \text{ sont les tailles des 2 échantillons.}$$

$$T^2 = 12 \sum_{k < j} \frac{\tilde{u}_{h,j}^2}{n_h n_j} - nH \quad \text{ce qu'on peut écrire :}$$

$$T^2 = 12 \sum_{h < j} \frac{\tilde{u}_{h,j}^2}{n_h n_j} - \frac{12}{n+1} \sum_{i=1}^k \frac{\tilde{u}_i^2}{n_i}$$

où $\tilde{u}_{h,j} = R_h^{(j)} - \frac{1}{2}n_h(n_h + n_j + 1)$, $R_h^{(j)}$ = somme des rangs du h^{ème} échantillon par rapport au j^{ème}.

$$\tilde{u}_i = R_i - \frac{1}{2}n_i(n+1)$$

u_i a rapport entre le $i^{\text{ème}}$ échantillon et tous les autres ensembles.
 $u_{h,j}$ " " les $h^{\text{ème}}$ et $j^{\text{ème}}$ échantillons.

Cette statistique T^2 est, pour les n_i grands, distribuée suivant une loi de χ^2 à $\frac{k(k-1)}{2}$ d. d. l.

c) Signalons :

JONCKHEERE "A distribution free k sample test against ordered alternatives" - Biometrika GB (1954) 41 - n° 1-2 (pp. 133-45) Tables.

13/ Le coefficient de corrélation de Spearman

Comme nous l'avons vu ce coefficient est apparu il y a déjà assez longtemps.

Si l'on range n individus par rapport à 2 variables, on peut calculer le coefficient de corrélation des rangs de ces 2 variables. Ce coefficient est le coefficient de Spearman. Si l'on appelle d_i la différence des rangs de ces 2 variables pour l'individu i :

$$R_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N^3 - N} \quad N \text{ étant le nombre d'individus}$$

Voir par exemple :

KENDALL (1948) The advanced theory of statistics I - London, Griffin.

En cas d'ex-aequo, on peut, soit appliquer une correction, soit, si les ex-aequo sont nombreux, calculer directement le coefficient de Spearman.

Pour tester l'hypothèse H_0 que les deux variables ne sont pas associées, on peut calculer le coefficient R_s .

Pour les petits échantillons :

OLDS (1938) "Distribution of sums of squared rank differences for small number of individuals" Ann. Math. Stat. 2 - 122, 48.

OLDS (1949) "The 5 % significance levels for sums of squares of rank differences and a correction. Ann. Math. Stat. 20 - 117-18, donne une table pour $n = 4(1) \dots 10(2)30$, $\alpha = 0,5$ et $\alpha = 0,1$ (zone critique unilatérale).

Si $n \geq 10$ Kendall "Rank correlation methods" 1948 - pp. 47-8 indique que $t = R_s \sqrt{\frac{N-2}{1-R_s^2}}$ suit une loi de t à $N-2$ ddl.

14/ Coefficient de corrélation de Kendall

a) Ce coefficient se rapporte à des paires de nombres x_i, y_i qui sont les rangs de 2 observations d'un même objet i (rangs dans les x_i et rangs dans les y_i).

Si l'on range les observations x_i en ordre croissant et en dessous les observations y_i , ces dernières se présentent dans un certain ordre. Si l'on appelle R le nombre des paires de la 2^{ème} rangée dont l'ordre est dans le même sens que l'ordre des numéros des rangs des paires associées de la 1^{ère} rangée, R sera compris entre 0 et $1/2 n(n-1)$, s'il y a n objets.

On appelle coefficient de corrélation des rangs τ de Kendall la quantité :

$$\tau = \frac{4R}{n(n-1)} - 1$$

cette quantité varie entre -1 s'il y a opposition parfaite et +1 s'il y a accord parfait entre les 2 observations de chaque objet.

Kendall s'est également servi de la quantité S

$$S = 2 R - \frac{1}{2} n (n - 1)$$

variant symétriquement autour de 0 de $-\frac{1}{2} n (n - 1)$ à $\frac{1}{2} n (n - 1)$

S est la différence du nombre de paires de numéros de rangs de la 2ème rangée dont l'ordre des numéros est dans le même sens que l'ordre des numéros de rangs de la paire associée de la 1ère rangée, et du nombre de paires pour lesquelles ce n'est pas le cas.

Si les observations x_i et y_i ne sont pas associées, toutes les conclusions sont également possibles et S aura une distribution.

Kendall a calculé cette distribution pour $n \leq 10$.

Le Centre Mathématique d'Amsterdam (Rapport S73) a calculé la plus petite valeur de S pour laquelle $P[S \geq S] \leq \alpha$ pour $\alpha =$

0,005
0,010
0,025
0,050
0,100

et $n = 1(1) \dots 40$

La distribution de S est symétrique et tend à être normale avec moyenne 0 et variance $\sigma^2 = \frac{1}{18} n(n - 1)(2n + 5)$.

La valeur de σ est tabulée dans le même rapport S73 of the Computation Department of the Mathematical Center. L. Kaarsemaker et A. Van Wijngaarden.

b) Ce coefficient τ avait été examiné pour la 1ère fois par :

R. GREINER en 1909 "Ueber das fehlersystem der Kollektiv ma β lehre" Zeit. für Math und Phy. 57 - 121, 225-37.

puis par :

F. ESSCHER en 1924 "On a method of determining correlation from the ranks of variates". - Skand. Akt. 7 201-219.

et redécouvert par :

KENDALL en 1938 - "A new measure of rank correlation Biometrika 30 - pp. 81-93.

complété par :

"the variance of τ when both rankings containties"- Biometrika 34 (1947) 297-98.

Si une seule des rangées est une rangée d'une variable stochastique, l'autre peut être utilisée pour l'ordonner.

H. B. MANN (1945) "Nonparametric tests against trend" - Econometrica 13 245, 259 - utilise ainsi une statistique T fonction linéaire de S pour tester l'hypothèse que les variables aléatoires x_i sont distribuées indépendamment suivant la même loi de distribution.

c) S'il y a des ex-aequo (ties) dans une série, G.P. SILLITO a calculé la distribution exacte de S avec des doublets et des triplets : "The distribution of Kendall's τ coefficient of correlation in rankings containing ties" - *Biometrika* **34** (1947) 36-40.

TERPSTRA a démontré la normalité asymptotique de S sous des conditions assez générales - "The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking" - *Proc. Kon. N. Ak. v. Wet A.* **55** (1952).

TERPSTRA s'est servi de ces résultats pour établir un test de tendance.

Considérons k variables aléatoires indépendantes à fonction de distribution continue Z_1, Z_2, \dots, Z_k . Si pour chaque variable Z_j on a n_j observations indépendantes, il peut être intéressant de tester l'hypothèse H_0 de l'identité des fonctions de distribution de Z_j contre celle par exemple d'une tendance croissante ou décroissante.

TERPSTRA a utilisé la statistique S de Kendall en formant pour chaque observation une paire (x_i, y_i) déterminée de la façon suivante :

la valeur $x_i = 1$ est affectée à toutes les n_1 observations de la variable Z_1
 $x_i = k$ " " " " n_k " " de la variable Z_k

la valeur y_i correspond aux observations faites

y_1, y_2, \dots, y_{n_1} sont les n_1 observations de Z_1

$y_{n_1 + 1}, \dots, y_{n_1 + n_2}$ sont les n_2 observations de Z_2 , etc.

On obtient donc pour les $(n_1 + n_2 + \dots + n_k)$ paires x_i, y_i les valeurs :

$$\left\{ \begin{array}{l} x_1 = x_2 = \dots = 1 \\ x_{n_1 + 1} = x_{n_1 + 2} = \dots = 2 \\ \dots \\ \dots \\ x_{n_1 + n_2 + \dots + n_{k-1} + 1} = x_{n_1 + n_2 + \dots + n_{k-1} + 2} = \dots = x_{n_1 + n_2 + \dots + n_{k-1} + n_k} = k \\ \left\{ \begin{array}{l} y_1, y_2, \dots, y_{n_1} \text{ les } n_1 \text{ observations de } Z_1 \\ y_{n_1 + 1}, y_{n_1 + 2}, \dots, y_{n_1 + n_2} \text{ les } n_2 \text{ observations de } Z_2 \\ \dots \\ \dots \\ y_{n_1 + n_2 + \dots + n_{k-1} + 1}, \dots, y_{n_1 + n_2 + \dots + n_{k-1} + n_k} \text{ les } n_k \text{ observations de } Z_k \end{array} \right. \end{array} \right.$$

Les grandes ou les petites valeurs de S étant critiques pour une tendance croissante ou décroissante dans l'arrangement des variables Z_1, Z_2, Z_k .

Dans le test de Wilcoxon, on a un cas particulier du test de Terpstra avec $k = 2$ et la statistique W est une fonction linéaire de S.

d) S'il y a des ex-aequo dans les 2 séries on n'a pas de théorèmes généraux. La normalité asymptotique est vraisemblable mais elle n'a été établie pour le test de Wilcoxon que par :

W.H. KRUSKAL (1952) déjà cités.
 J. HEMELRIJK (1952)

e) Kendall cite dans le cas spécial de la table 2×2

n objets possèdent ou non une qualité A
 " " " B

$u_i = 1$ si le $i^{\text{ème}}$ objet a la qualité A

$u_i = 2$ si " " n'a pas la qualité A

$v_i = 1$ si le $i^{\text{ème}}$ objet a la qualité B
 $v_i = 2$ " " " n'a pas la qualité B.

La quantité S^2 est alors proportionnelle au χ^2 habituel d'une table 2×2 . On peut de la même façon rapprocher S des tables de contingence et des tests qui peuvent en être dérivés comme le test de la médiane de Westenberg (cité), celui de G. W. Brown et A. Mood (1948) "Homogeneity of several samples" (the American statistician 2 - 22) et ses généralisations par Mood (cité), Brown et Mood "On median test for linear hypotheses (Second Berkeley Symposium 159, 166), Blomquist "Some tests based on dichotomization", Ann. Math. Stat. 22, 362-71 et Hemelrijk (cités Statistical methods based on few assumptions).

f) Ces tests sont basés sur des fonctions linéaires des rangs des observations ; On peut également généraliser et employer des fonctions quadratiques des rangs. Le test de Wilcoxon a ainsi été généralisé en test de k échantillons par Kruskal et Wallis et indépendamment par Rijkooort et Terpstra.

15/ Une autre généralisation est celle du coefficient de corrélation de Spearman, par la méthode des m classements de Friedmann (1937) "The use of ranks to avoid the assumption of normality implicit in the analysis of variance" - Jrn ; Am. Stat. Ass. 32, 675-99.

Kendall dans les chapitres 6 et 7 de "Rank correlation methods" explique la méthode des m classements

- m juges classent n objets en leur affectant un rang de 1 à n.

La méthode consiste : à déterminer la moyenne des rangs pour un classement :

$$\frac{n + 1}{2}$$

à déterminer la moyenne des rangs pour les m classements :

$$m \left(\frac{n + 1}{2} \right),$$

à totaliser pour chaque objet les m rangs qu'il a obtenus et à en soustraire la moyenne des rangs :

$$m \left(\frac{n + 1}{2} \right),$$

à sommer les carrés des différences en une somme S.

Au cas où les objets sont équivalents S a une distribution résultant d'un pur tirage au sort.

Dans cette hypothèse H_0 , on peut calculer pour chaque valeur prise par S sa probabilité d'apparition et établir les tables.

Réf : le rapport S59 - MC - (1951) Ph. Van ELTEREN ; indique les procédés de calcul approchés.

On ne peut appliquer ce système que dans un schéma rectangulaire où toutes les cases sont remplies.

DURBIN "Incomplete blocks in ranking experiments" - British Jrn. of Psychology 4 85-90 (1951).

A. BENARD, Ph. Van ELTEREN "A generalization of the method of m rankings" -
 Proceeding A. 56 (1953) pp. 358-369.

ont généralisé cette méthode qui comporte comme cas particulier :

- la méthode des m classements,
- le test des k échantillons de Kruskal et Terpstra,
- le test de Wilcoxon,
- le test des signes.

16/ Coefficients de concordance, de consistance et d'agrément de Kendall

- le Coefficient de concordance est le rapport $W = \frac{S \text{ calculé}}{S_{\max}}$ dans la méthode des m classements, il varie de 0, hasard absolu à 1, concordance parfaite. On peut tester sa signification à l'aide du test de Fisher Snedecor. On tiendra compte d'une correction de continuité en écrivant :

$$W = \frac{S - 1}{S_{\max} + 2}$$

et on calculera le F de Snédécour $F = \frac{(m - 1)W}{1 - W}$ avec :

$(n - 1) - \frac{2}{m}$ ddl pour la plus grande estimation

$(m - 1) \left[n - 1 - \frac{2}{m} \right]$ ddl pour la plus petite.

ce qui permet de tester la signification.

Si l'on avait calculé la moyenne des coefficients de Spearman 2 à 2 on aurait eu :

$$R_{s \text{ moyen}} = \frac{m W - 1}{m - 1}$$

- le Coefficient de consistance se rapporte à l'homogénéité du jugement d'un juge :

$$\tau = 1 - \frac{24 d}{n^3 - n} \quad \text{si } n \text{ impair}$$

$$\tau = 1 - \frac{24 d}{n^3 - 4n} \quad \text{si } n \text{ pair}$$

où d est le nombre de triades circulaires cad de jugements homogènes d'un juge sur 3 objets : si A est préféré à B et B préféré à C, A doit être préféré à C.

- le coefficient d'agrément des m juges est égal à 1 si tous les juges sont d'accord et sa valeur minimum est :

$$A_{m,r} = - \frac{1}{m - 1} \quad \text{si } m \text{ est pair}$$

$$A_{m,n} = - \frac{1}{m} \quad \text{si } m \text{ est impair}$$

Il est possible de tester par un test de χ^2 une fonction linéaire du nombre total d'accords entre 2 juges.

17/ Signalons un test de décalage non paramétrique :

"A distribution free k sample slippage test" Report S 206 (VP8) of the Mathematical Center Amsterdam (Septembre 1956) R. Doornbos, H. Kesten et H. J. Prins.

On fait l'hypothèse H_0 que k variables indépendants U_1, \dots, U_k ont la même distribution continue :

- de la $i^{\text{ème}}$ population u_i , on a t_i observations indépendantes :

$$U_{ij} \quad (j = 1, 2, \dots, t_i)$$

- on désire tester H_0 contre les alternatives H_1 et H_2 pour une valeur inconnue de i (c'est-à-dire que l'une des variables indépendantes U_i est décalée à droite ou à gauche des autres u_j)

$$H_1 \quad \left\{ \begin{array}{l} P [U_i > u_j] > \frac{1}{2} \quad j \neq i \\ U_j \quad (j = 1, \dots, i-1, i+1, \dots, k) \text{ ont la même distribution} \end{array} \right.$$

$$H_2 \quad \left\{ \begin{array}{l} P [U_i > u_j] < \frac{1}{2} \quad j \neq i \\ U_j \quad (j = 1, 2, \dots, i-1, i+1, \dots, k) \text{ ont la même distribution} \end{array} \right.$$

On range alors tous les U_{ij} ($i = 1 \dots k$; $j = 1 \dots t_i$) et l'on appelle T_i la somme des rangs des observations U_{ij} ($j = 1, \dots, t_i$).

T_i est une fonction linéaire de la statistique de Wilcoxon appliquée au $i^{\text{ème}}$ échantillon et à l'ensemble des $k-1$ autres échantillons. On connaît donc sa distribution dans l'hypothèse H_0 et on peut calculer pour l'alternative $(H_1)p_i = P [T_i \geq T_i]$ et pour l'alternative $(H_2)q_i = P [T_i \leq T_i]$.

On rejettera H_0 et on décidera que la $i^{\text{ème}}$ variable est décalée à droite si :

$$\min p \leq \epsilon/k \quad (H_1)$$

On rejettera H_0 et on décidera que la $i^{\text{ème}}$ variable est décalée à gauche si :

$$\min q_i \leq \epsilon/k \quad (H_2)$$

ϵ étant le niveau de confiance.

18/ Tests de dispersion non paramétriques

a) Test de KAMAT : Ferguson College University of Poona - "A two sample distribution free test" - Biometrika GB 1956 - 43 - n° 3-4 pp. 377-85.

BARTON : "The limiting distribution of Kamat's test statistic" (même référence).

KAMAT "Further contribution to the theory of a non parametric test" - Proc. indian. Sc. Congr. 1956 - 43 ; n° 3 - pp. 16.

Les échantillons sont supposés avoir le même centrage.

On les fusionne et on forme $D_{n,m} = R_n - R_m + m$ où m et n sont les tailles des échantillons ; R_n et R_m l'étendue des rangs des échantillons.

Une table pour $m+n = 12, 16, 20$ donne les percentiles supérieurs et inférieurs 0, 5 et 2, 5 de cette statistique.

Citons :

ROSENBAUM "Tables for a non parametric test of dispersion" - Ann. Math. Stat. (1953) 24 n° 4 pp. 663-8.

b) Notons dans cet ordre d'idées :

BOX : "Non Normality and test on variances" - Biometrika 40 - 1953 - pp. 318-35.

SARHAN (Egypte) "Estimation of the mean and standard deviation of order statistics - Ann. Math. Stat. U. S. A. I - (1954) 5 n° 2 - 317-28 - Tables.

II - (1955) 26 - n° 3 - 505-11

III - (1955) 26 - n° 4 - 576-92

WILSON "A distribution free test of analysis of variances hypotheses" - Psych. Bull. 53 (1956) pp. 96-101.

ROY et MITRA "An introduction to some non parametric generalizations of analysis of variance and multivariate analysis" - Biometrika GB (1956) 43 - n° 3, 4 - 561-76.

SUKHATME "On certain two sample non parametric tests for variances" - Ann. Math. Stat. U. S. A. - 1957 (28 n° 1 - pp. 188-94 - Proc. ind. Scie. Cong. 1956 43 n° 4-7.

19/ Limites de tolérance non paramétriques

Signalons les références :

R. B. MURPHY "Non parametric tolerance limits" - Ann. of Math. Stat. Vol 19 (1948) pp. 581-9 (graphiques)

Rapport S59 TERPSTRA Centre mathématique d'Amsterdam (1951).

D. A. S. FRASER "Non parametric tolerance regions" - Ann. Math. Stat. Vol 24 (1953) n° 1 pp. 44-45.

J. W. TUKEY I "Non parametric estimation"

Ann. Math. Stat. 16 (1945) pp. 187-92

II " " " 18 (1947) pp. 529-39

III " " " 19 (1948) pp. 30 - 39.

S. S. WILKS "Determination of sample sizes for setting tolerance limits" - Ann. Math. Stat. U. S. A. - 12 - (1941) pp. 91-96 - Statistical prediction with special reference to the problem of tolerance limits" - Ann. Math. Stat. U. S. A. 13 - (1942) pp. 400-409.

20/ Analyses de régression et méthodes non paramétriques

Signalons les références :

BARTLETT (1949) "Fitting a straight line when both variables are subject to error" - Biometrics Vol 5 n° 3 - pp. 207-12.

HEMELRIJK (1949) "Construction of a confidence region for a line" - Proc. Kon. AK. v. Wet. Vol 52 - pp. 995-1005.

HOUSNER et BRENNAN (1948) "The estimation of linear trends" - Ann. Math. Stat. Vol 19 - pp. 380-93.

- NAIR et SHRIVASTAVA "On a simple methode of curve fitting" - Sankhya vol 6 - pp. 121-32.
- NAIR et BANERJEE (1942) "A note on fitting of straight lines if both variables are subject to error" - Sankhya vol 6 - pp. 331.
- THEIL (1950) "A rank invariant method of linear and polynomial regression analysis" - Proc. Kon. Ned. Ak. v. Wet. Vol 53 - pp. 386-92-521-25-1397-1412. (1951) Report S 59 of the Mathematical Center Amsterdam.
- WALD (1940) "The fitting of straight lines if both variables are subject to error" Ann. Math. Stat. Vol. 11 - pp. 284-300.
- DANIELS (1954) "A distribution free test for regression parameter" - Ann. Math. Stat. U. S. A. - 1954 - 25 - n° 3 pp. 499-513 Tables.
- HILL (1954) "The distribution of the regression coefficient in samples from a non normal population" - Biometrika GB - 1954 - 41 - n° 3-4 548-52.