Quelques tests non paramétriques pour le traitement des petits échantillons

Arnaud BRINGÉ

Introduction

Dans des cas classiques d'échantillons de taille suffisamment importante (n>30), ou de lois respectant fidèlement une loi normale, les tests de Fisher et de Student peuvent être utilisés afin de tester une différence significative de mesure entre plusieurs populations. Cependant, ces hypothèses nécessitent des règles d'application assez strictes sur l'adéquation des données à la loi normale.

Les méthodes non paramétriques requièrent, elles, peu d'hypothèses concernant la population étudiée. Elles ignorent notamment l'hypothèse de la normalité de la population. Ces tests représentent le meilleur moyen de prendre en compte de petits échantillons, pour lesquels les hypothèses pour appliquer les tests habituels ne sont pas respectées. Ainsi, en deçà de 30 observations, il peut être hasardeux d'approximer l'échantillon par une loi Normale.

On parle de tests non paramétriques lorsque l'on ne fait aucune hypothèse sur la distribution des variables, et donc aucune hypothèse sur la distribution sous-jacente des données.

Les tests non paramétriques sont basés sur l'utilisation de statistiques de rang (par exemple des notes attribuées, ou un classement entre individus). Ils peuvent s'appliquer à des caractères quantitatifs, à des grandeurs de mesure, à des rangs de classement.

Un test non paramétrique est donc un test d'hypothèse pour lequel il n'est pas nécessaire de spécifier la forme de la distribution de la population étudiée.

Il faut cependant que les observations soient indépendantes, c'est-à-dire que la sélection d'un quelconque individu dans la population en vue de former un échantillon ne doit pas influencer le choix des autres individus.

L'efficacité des tests non paramétriques n'est que légèrement inférieure à celle de leurs équivalents paramétriques quand la distribution de la population étudiée est spécifiée, par exemple la loi normale. D'après Dodge [2007], elle peut s'avérer supérieure quand la distribution de la population dévie de la loi normale.

Nous ne présenterons dans ce qui suit que quelques tests non paramétriques : Mann-Whitney et Wilcoxon dans le cas de deux échantillons, Wilcoxon dans le cas d'échantillons appariés, et Kruskal-Wallis dans le cas d'une comparaison à plus de deux échantillons. Une feuille de calcul illustrera parallèlement le calcul effectué dans l'utilisation de chacun de ces tests. Notre but n'est pas d'être exhaustif, de présenter un catalogue de ces tests assez anciens, mais de montrer un exemple de leur utilisation.

Le test de Mann-Whitney

Principe

On dispose de mesures d'une variable sur un premier échantillon A de taille *n* provenant d'une population P_1, noté (X_1, X_2,...X_n) et sur un second échantillon B de taille *p* provenant d'une population P_2 noté (Y_1, Y_2, … Y_p).

Ces échantillons vont être fusionnés et classés par ordre croissant des valeurs de cette mesure.

Pour chaque élément X_i de X, on calcule le nombre total d'éléments de Y supérieurs à cet élément de X augmenté de la moitié du nombre de fois où un élément de Y est égal à X_i.

La statistique de Mann-Whitney U_{xy} sera la somme de ce calcul sur l'ensemble des éléments de X. On calcule de la même manière U_{yx}, somme sur tous les Y_i du nombre total d'éléments de X supérieurs à cette valeur, augmentée de la moitié du nombre de fois où il y a égalité). On peut montrer que $U_{yx}=np-U_{xy}$. Les grandeurs U_{xy} et U_{yx} vont donc être calculées sur cet ensemble de (n+p) observations.

Si les deux groupes étaient complètement séparés, alors $U_{xy}=0$ et $U_{yx}=np$ (ou $U_{yx}=0$ et $U_{xy}=np$ selon le cas). Inversement, si les groupes sont parfaitement semblables, $U_{xy} = U_{yx}= np/2$

La plus petite des deux statistiques calculées (U_{xy}, U_{yx}) est comparée à une statistique U de Mann Whitney tabulée. Si $\min(U_{xy},U_{yx})<U$, alors l'hypothèse d'égalité des distributions des populations P_1 et P_2 sera rejetée.

Remarques :

- Si m,n>12, alors la loi de U est approximée par la loi normale de moyenne $\mu = \frac{m*n}{2}$ et d'écart type $\sigma = \sqrt{\frac{m*n*(m+n+1)}{12}}$. On compare alors la statistique $\frac{U-\mu}{\sigma}$ à la valeur lue dans la table de la loi normale centrée réduite.

- Les valeurs sont lues dans une table de Mann-Whitney, reproduite notamment sur le site http://math.usask.ca/~laverty/S245/Tables/wmw.pdf . On utilisera pour le test la plus petite des deux valeurs U_{xy} et U_{yx}. Le test sera significatif si cette valeur est inférieure à la valeur lue dans la table.

Le test de Wilcoxon

Principe

Le cadre d'utilisation de ce test sera le même que pour le test de Mann-Whitney, deux échantillons de taille *n* et *p*. Ces deux échantillons seront réunis selon une procédure identique.

On attribue ensuite un rang à chacune de ces valeurs dans ce vecteur de dimension (*n*+*p*). Si plusieurs observations présentent la même valeur de cette mesure, un rang moyen est assigné. Comme pour le test de Mann-Whitney, cette procédure s'applique dans le cadre de mesures, ou de variables au moins ordinales. On attribue ensuite un rang à chacune de ces valeurs dans ce vecteur de dimension (n+p).

On note R(Xi) le rang associé à l'observation X_i, et on calcule la statistique $W = \sum_{i=1}^{n} R(X_i)$

On peut montrer qu'il existe une relation liant les statistiques U de Wilcoxon et T de Mann-Whitney [Dodge, 2007] : $U = np + \frac{n(n+1)}{2} - T$

Test de Wilcoxon pour échantillons appariés

Principe

On se place ici dans un cadre différent. On dispose de deux échantillons A et B, tous deux de taille n, représentant deux valeurs d'une mesure effectuées à deux instants t_1 et t_2 sur les mêmes individus.

On calcule dans un premier temps les différences entre les valeurs prises par les deux mesures appariées pour chaque individu, puis ces valeurs sont classées par rang. On note W la somme des rangs des différences positives. On peut ensuite comparer W à des valeurs tabulées (à risque fixé) ou utiliser un logiciel pour calculer une p-value.

Sous l'hypothèse H_0, il n'y a pas de différence entre les deux groupes et donc la somme des rangs positifs ne sera pas significativement différente de la somme des rangs négatifs. Le rejet de H_0 indiquera donc que la somme des rangs positifs n'est pas égale à la somme des rangs négatifs et qu'il y a une différence entre les deux groupes comparés. La table de Wilcoxon donne la limite inférieure de T, plus faible total de rangs positifs ou négatifs.

Kruskal-Wallis

Principe

On dispose de mesures d'une variable sur *k* échantillons A_1 de taille n_1, A2 de taille n_2, … A_K de taille n_k, et on cherche à mesurer si ces mesures sur les *k* échantillons sont identiques, ou si au moins une mesure sur un échantillon est différente des autres.

Soit n la taille totale des *k* échantillons. En réitérant la procédure du test de Wilcoxon adaptée cette fois-ci à *n* échantillons, on fusionne les *n* échantillons et on applique un indice de rang à chacune des observations. De la même manière que précédemment, on applique un rang moyen si plusieurs observations ont la même valeur.

Comme indiqué dans Dodge [2007], si le nombre de rangs moyens est limité, alors on calcule la statistique suivante : $H = \left(\frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} \right) - 3(n+1)$

De plus, si le nombre de rangs moyens est important, la statistique H est divisée par un terme $1 - \frac{\sum_{i=1}^{g} (t_i^3 - t_i)}{n^3 - n}$, avec *g* nombre de groupes contenant des rangs moyens et t_i la taille de chaque groupe contenant des rangs moyens.

Tests effectués

(H_0) : Il n'y a pas de différence significative de la mesure entre les différents échantillons
(H_1) : Il existe au moins un échantillon pour lequel les valeurs de la mesure sont significativement différentes de celles des autres échantillons.

Si H est supérieur à la valeur $\chi^2_{k-1,1-\alpha}$ lue dans une table, alors l'hypothèse nulle sera rejetée.

Exemple

Sur le classeur joint figurent 2 exemples de tests de Kruskal-Wallis. Les différentes grandeurs énoncées précédemment sont calculées, la correction pour les rangs moyens est systématiquement indiquée même si elle n'a pas lieu de s'appliquer dans le cas présents, le nombre de groupes de rangs moyens étant assez faible. Le premier exemple ne permet pas de conclure à une différence entre les échantillons, le second exemple rejette l'hypothèse nulle et permet de mettre en valeur qu'au moins un échantillon est différent des autres.

Références bibliographiques

Dodge Y., *Dictionnaire statistique encyclopédique,* Springer, 2007.

Hollander M. and Wolfe A. D., *Nonparametric Statistical Methods.* New York: John Wiley & Sons, 1973.

Kruskal W.H. and Wallis W.A., Use of ranks in one criterion variance analysis, *Journal of American Statistical Association*, 47, 1952, pp. 583-621 et correction 48 p. 910.

Morice E., Quelques tests non paramétriques, *Revue de Statistique Appliquée*, RSA, tome 4, n°4, 1956, pp. 75-107.

Mann H.B., Whitney D.R., On a test whether one of two random variables is stochastically larger than the other, *Annals of mathematical statistics*, 18, 194, pp. 50-60.

Raison J., Les principaux tests non paramétriques. Quelques généralités et références biblographiques, *Revue de Statistique Appliquée*, RSA, tome 7, n°1, 1959, pp 83-106.

Wilcoxon F., Individual comparaisons by ranking methods, *Biometrics*, 1, pp. 80-84, 1945.