REVUE DE STATISTIQUE APPLIQUÉE

AMIDI ALI

Étude de la fonction de puissance du test de signe

Revue de statistique appliquée, tome 24, n° 2 (1976), p. 89-94
<http://www.numdam.org/item?id=RSA_1976__24_2_89_0>

ÉTUDE DE LA FONCTION DE PUISSANCE DU TEST DE SIGNE (1)

Amidi ALI
Université Jundi Shapur, AHWAZ-IRAN

Le présent article a pour but une recherche qui permet de calculer la puissance du test de signe. Dans la première partie nous avons rappelé le test de signe, dans la 2^e partie nous avons calculé un intervalle de confiance pour ce test et finalement nous avons presenté la puissance de ce test.

a) RAPPEL DU PROBLEME

On considère deux variables aléatoires indépendantes X et Y, et n paires d'observations indépendantes (x_i, y_i) de ces deux variables. Les observations sont faites dans les mêmes conditions pour les deux éléments de chaque paire, tandis qu'elles peuvent varier de paire à paire. On utilise le test de signe pour tester l'hypothèse H_0 que les deux variables X et Y ont la même distribution. En général, la distribution des observations d'une part peut être différente de la distribution normale, donc on ne peut pas utiliser le test de Student ni d'autre test paramétrique, d'autre part les x_i et y_i peuvent être difficiles à déterminer et il n'est pas toujours possible de les quantifier.

Le test de signe est basé sur le nombre des signes (+ ou −) des différences $x_i - y_i$ des deux observations de chacune des paires. Il consiste à tester l'hypothèse que chacune de ces différences appartient à une distribution à médiane nulle, on rejette l'hypothèse dans le cas où les nombres de signes + ou − différent significativement.

Soit U le nombre des différences $x_i - y_i$ positives, si $r = \Pr(X=Y) = 0$, on peut accepter ou rejeter l'hypothèse nulle suivant que U se rapproche ou s'éloigne de la valeur théorique n/2. Dans l'hypothèse nulle U a une répartition binomiale (n, 1/2). Il faut donc déterminer les zones critiques de sa distribution qui nous indiqueront à un seuil donné α s'il faut accepter ou rejeter l'hypothèse $p = 1/2$, donc finalement la statistique utilisée dans le test de signe est $U = \#(X_i > Y_i)$ (2) $i = 1, \ldots, n$.

---

(1) Article remis en octobre 1974, révisé en juillet 1975

(2) La notation $\#(X_i > Y_i)$ désigne le nombre de paires où $X_i > Y_i$

Mots-Clés : **Tests. Puissance de tests. Tests non paramétriques. Test de signes.**

b) L'INTERVALLE DE CONFIANCE DE U

L'évènement : on a obtenu u fois $(X_i - Y_i > 0)$ sous l'hypothèse nulle a pour probabilité $C_n^u \left(\frac{1}{2}\right)^u \left(\frac{1}{2}\right)^{n-u}$. α étant donné on peut chercher le plus grand entier L_1 tel que $2^{-n} \sum_{i=0}^{L_1-1} C_n^i \leqslant \frac{\alpha}{2}$, puis le plus petit entier L_2 tel que $2^{-n} \sum_{i=L_2+1}^{n} C_n^i \leqslant \frac{\alpha}{2}$ de sorte que la région d'acceptation sera $[L_1, L_2]$ avec un seuil $\leqslant \alpha$. Posons, par symétrie $L_2 = \frac{n}{2} + a$ et $L_1 = \frac{n}{2} - a$. Sous H_0 nous avons

$$\alpha = \left[2^{-n} \sum_{i=0}^{\frac{n}{2}-a} C_n^i \right] . 2 = 2^{1-n} \sum_{i=0}^{\frac{n}{2}-a} C_n^i$$

Connaissant α on peut calculer le nombre a et ensuite les nombres L_1 et L_2. Il faut signaler que, en raison du caractère discret de la distribution binomiale, la valeur fixé α ne peut pas être atteinte exactement.

Sur un ordinateur, nous avons calculé ces deux nombres pour n = 5 (1) 30 avec le seuil $\alpha = 0.05$.

Nous donnons les résultats dans le tableau suivant :

n	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
L_1	1	1	1	1	2	2	2	3	3	3	4	4	5	5	5	6	6	6	7	7	8	8	8	9	9	10
L_2	4	5	6	7	7	8	9	9	10	11	11	12	12	13	14	14	15	16	16	17	17	18	19	19	20	20

On sait que lorsque n est grand, la loi de U tend vers la loi normale de moyenne E (U) = n/2 et de variance Var (U) = n/4 donc $T = \frac{U - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}$ suit à peu près la loi N (0, 1).

Nous avons cherché à partir de quel n cette affirmation est réalisée, calculons les nombres L_1 et L_2 sous l'hypothèse de normalité

$$\Pr \left[\frac{n}{2} - a \leqslant U \leqslant \frac{n}{2} + a \right] = 1 - \alpha$$

donc

$$\Pr \left[- a \leqslant U - \frac{n}{2} \leqslant a \right] = 1 - \alpha$$

ou

$$\Pr \left[\frac{-a}{\frac{1}{2}\sqrt{n}} \leqslant \frac{U - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \leqslant \frac{a}{\frac{1}{2}\sqrt{n}} \right] = 1 - \alpha$$

ou $$\Pr\left[\frac{-a}{\frac{1}{2}\sqrt{n}} \leqslant T \leqslant \frac{a}{\frac{1}{2}\sqrt{n}}\right] = 1 - \alpha$$

Pour $\alpha = 0.05$ nous avons

$$2 \Pr\left[0 \leqslant T \leqslant \frac{2a}{\sqrt{n}}\right] = 0{,}95,$$ donc L_1 et L_2

sont les entiers voisins de :

$$\frac{n}{2} - 0.98\sqrt{n} \qquad \text{et} \qquad \frac{n}{2} + 0.98\sqrt{n}$$

D'où le tableau

n	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
L_1	0	1	1	1	2	2	2	3	3	3	4	4	4	5	5	6	6	6	7	7	8	8	8	9	9	10
L_2	5	5	6	7	7	8	9	9	10	11	11	12	13	13	14	14	15	16	16	17	17	18	19	19	20	20

En comparant les deux tables ci-dessus, nous voyons que pour $n \geqslant 18$ les nombres L_1 et L_2 sont les mêmes, donc nous pouvons utiliser la loi normale pour obtenir ces deux nombres dès que $n \geqslant 18$. Il est clair que pour un seuil donné α, on peut présenter également un test unilatéral sur U.

c) ETUDE DE LA FONCTION DE PUISSANCE DU TEST DE SIGNE

Nous avons vu que la variable $U = \#(X_i > Y_i)$ suit une loi binomiale. Nous allons tester $H_0 : p_0 = 1/2$ contre l'alternative $H_1 : p_1 \neq 1/2$. On sait que la puissance d'un test est la probabilité de rejeter H_0 quand H_1 est vrai. La fonction de puissance étant évidemment symétrique par rapport à $p = 1/2$, il suffit de considérer l'alternative $p > 1/2$. Appliquons la méthode de Neyman et Pearson, pour cela envisageons un m-échantillon constitué de m ensembles de n paires chacun, On a

$$P_0(u) = C_n^u \left(\frac{1}{2}\right)^u \left(1 - \frac{1}{2}\right)^{n-u} = 2^{-n} C_n^u$$

$$P_1(u) = C_n^u\, p^u\, (1-p)^{n-u}$$

Calculons les fonctions de vraisemblances correspondantes

$$f_0(\vec{u}) = 2^{-mn} \prod_{i=1}^{m} C_n^{u_i}$$

$$f_1(\vec{u}) = p^{\Sigma u_i} (1-p)^{mn - \Sigma u_i} \prod_{i=1}^{m} C^{u_i}$$

d'après la méthode choisie

$$\frac{f_1(\vec{u})}{f_0(\vec{u})} = \frac{p^{m\bar{u}}\ (1-p)^{mn-m\bar{u}}}{2^{-mn}} = 2^{mn}\ [p^{m\bar{u}}\ (1-p)^{mn-m\bar{u}}]$$

On sait que la région critique est définie par $\frac{f_1(\vec{u})}{f_0(\vec{u})} \geqslant K(\alpha), \alpha$ étant le seuil envisagé pour un test unilatéral sur U.

$$2^{m.n}\ [p^{m\bar{u}}\ (1-p)^{mn-m\bar{u}}] \geqslant K(\alpha)$$

ou

$$\bar{u} \log \frac{p}{1-p} \geqslant \frac{1}{m}\ [\log \{2^{-mn} K(\alpha)\}] - n \log (1-p)$$

ou

$$\bar{u} \geqslant \{ \frac{1}{m} \log\ [2^{-mn} K(\alpha)] - n \log (1-p) \} / \log \frac{p}{1-p}$$

ou $\bar{u} \geqslant A$ (bien entendu $\bar{u}$ est une réalisation de $\bar{U}$). On doit donc calculer la densité de probabilité de $\bar{U}$ et sa fonction de répartition.

Nous considérons les variables $U_1, U_2, \ldots U_m$ qui suivent la loi binomiale (p, n) indépendamment. On sait que la fonction génératrice de la loi binomiale est

$$A_U(s) = (q + p \cdot s)^n$$

d'autre part

$$\frac{1}{m} \sum_{i=1}^{m} U_i = \bar{U} \qquad \text{ou} \qquad m\bar{U} = \sum_{i=1}^{m} U_i$$

donc $A_{m\bar{U}}(s) = (q + p \cdot s)^{mn}$. Il en résulte que $m\bar{U}$ suit une loi binomiale (mn, p) donc

$$P[m\bar{U} = t] = C^t_{m.n}\, p^t\, q^{mn-t} \qquad \text{soit avec} \qquad \frac{t}{m} = \lambda$$

$$P\ [\bar{U} = \lambda] = C^{m\lambda}_{mn}\, p^{m\lambda}\, q^{mn-m\lambda} \qquad \text{avec} \qquad \lambda \in \{0, \frac{1}{m}, \ldots, n\}$$

A partir de la loi de $\bar{U}$ on peut calculer la fonction de puissance. Nous avons $\alpha = P_0\ [\bar{U} \geqslant A]$ ou $1 - \alpha = P_0\ [\bar{U} < A]$. Soit $Q_p(y)$ la fonction de répartition de $\bar{U}$ alors $1 - \alpha = Q_{1/2}(A)$. Pour obtenir la valeur de A il y a deux cas :

I – on peut définir une valeur unique pour A de sorte que

$$A = Q^{-1}_{1/2}(1-\alpha) : A = \frac{k}{m} \quad \text{avec} \quad Q_{1/2}\left(\frac{k}{m}\right) < 1 - \alpha < Q_{1/2}(k + 1/m)$$

II – On ne peut pas définir une valeur unique pour A. Soit $\mathfrak{A}$ l'ensemble $\mathfrak{A} = \{x : Q_{1/2}(x) = 1 - \alpha\}$ et soit x_1 la plus grande valeur de $\mathfrak{A}$. Nous prendrons alors par convention $A = x_1 = Q^{-1}_{1/2}(1-\alpha)$

Ainsi il y aura toujours une valeur de A unique. Soit

$$B_1 = \{ \vec{u} : \bar{U} \geqslant A \} = \{ \vec{u} : \bar{U} > Q_{1/2}^{-1} (1 - \alpha) \}$$

B_1 est la région de rejet

Si π représente la puissance du test

$$\pi = 1 - P_1 \{ \bar{U} < Q_{1/2}^{-1} (1 - \alpha) \} \quad \text{soit} \quad \pi = 1 - Q_p \, [Q_{1/2}^{-1} (1 - \alpha)]$$

Si nous supposons que $\mathcal{B}_{(n,p)}(u)$ représente la fonction de répartition de la loi binomiale, alors

$$\pi = 1 - \mathcal{B}_{(mn,p)} \, [\mathcal{B}^{-1}_{(mn,\frac{1}{2})} (1 - \alpha)] \quad \text{avec} \quad p > \frac{1}{2}$$

Pour nm = constante, et pour un seuil α donné, $\mathcal{B}^{-1}_{(mn,\frac{1}{2})} (1 - \alpha)$ pour tout p est constante, de sorte que les différentes valeurs de π quand p varie sont obtenues à partir de la table de la loi binomiale.

Pour nm = N et $\alpha = 0.05$ nous présentons les valeurs de π dans la table

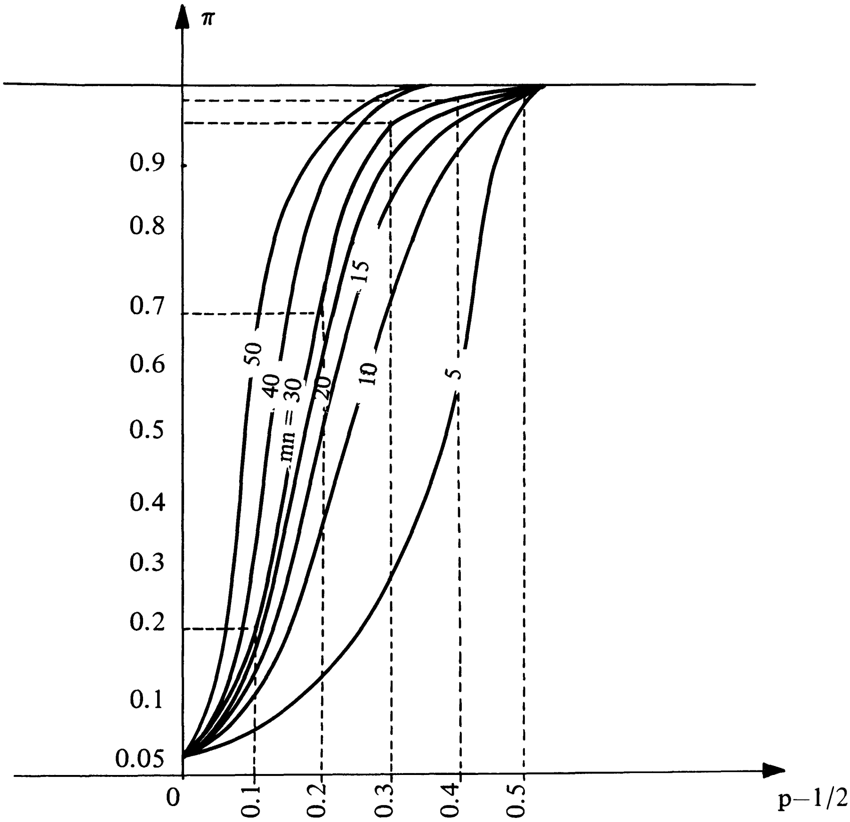
p	$p - \frac{1}{2}$	$\mathcal{B}^{-1}_{(N,\frac{1}{2})}$							π						
		N = 5	N = 10	N = 15	N = 20	N = 30	N = 40	N = 50	N= 5	N=10	N=15	N=20	N=30	N=40	N=50
0.5	0	0	2	4	6	10	14	19	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.4	– 0.1	"	"	"	"	"	"	"	0.07	0.16	0.21	0.25	0.29	0.31	0.44
0.3	– 0.2	"	"	"	"	"	"	"	0.16	0.38	0.51	0.60	0.73	0.80	0.91
0.2	– 0.3	"	"	"	"	"	"	"	0.32	0.67	0.83	0.91	0.97	0.99	0.99
0.1	– 0.4	"	"	"	"	"	"	"	0.59	0.92	0.98	0.99	0.99	1	1

les courbes correspondantes pour $p > 1/2$ sont tracées dans la page suivante.

Remarque

Dans le paragraphe c nous avons utilisé m échantillons de chacun n paires. Il est remarquable que la puissance ne dépende que du produit N = mn.

Les résultats sont donc valables pour m = 1 et n = M, et nous pouvons donner la courbe de puissance pour m = 1.



BIBLIOGRAPHIE

[1] MANN. – On a test of randomness based on signs of differences. The Ann. of Math. Stat.-Vol. 16-1945.

[2] MORICE E. – Quelques tests non parametriques *Rev. Stat. Appl.*, n° 4 1956.

[3] GOTTFRIED E. NOETHER. – Elements of Nonparametric Statistics John Wiley (1967).