ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria III: MATEMATYKA STOSOWANA XXXIV (1991)

PRZEMYSŁAW GRZEGORZEWSKI

Warszawa

Dwupróbkowe nieparametryczne testy położenia

(*Praca wpłynęła do Redakcji 30.10.1989*)

0. Wstęp. Niniejsza praca przedstawia znane w literaturze przedmiotu i stosowane w praktyce dwupróbkowe nieparametryczne testy położenia.

Starano się nadać tej pracy charakter aplikacyjny – stąd uwagi i wskazówki dotyczące konkretnych zastosowań testów.

W rozdziale trzecim omówiono wyniki badań symulacyjnych przeprowadzonych dla porównania mocy opisywanych testów (wyniki zamieszczono w aneksie).

Rozdział czwarty zawiera (zapisaną w języku PASCAL) procedurę obliczania poziomu krytycznego najpopularniejszego ze stosowanych testów – testu Wilcoxona.

W rozdziale piątym omówiono zagadnienia pokrewne.

W nawiasach kwadratowych [ ] podano odnośniki do bibliografii. Literatura tematu jest tak bogata, że zdecydowano się podać jedynie pozycje cytowane w niniejszej pracy i teksty źródłowe. Szczegółowa bibliografia zamieszczona jest np w pracach [10] i [21].

I. Sformułowanie problemu.

1. Niech $X_1, X_2, \ldots, X_m$ będzie m-elementową próbą z rozkładu o nieznanej ciągłej dystrybuancie $F(x)$.

Niech $Y_1, Y_2, \ldots, Y_n$ będzie n-elementową próbą z rozkładu o dystrybuancie $G(x) = F(x - \Delta)$.

Testujemy hipotezę

$$H : \Delta = 0$$

mówiącą, że obie próby pochodzą z populacji o tym samym rozkładzie, wo-

bec hipotezy alternatywnej

$$K : \Delta > 0,$$

odpowiadającej sytuacji, w której Y-ki są "średnio" większe od X-ów (zmienne losowe Y_j są stochastycznie większe od X_i).

2. Formalnie, rozpatrujemy model statystyczny $(\Omega, \mathcal{A}, \mathcal{P})$, gdzie Ω oznacza przestrzeń prób, $\mathcal{A}$ – σ-ciało zdarzeń losowych, a $\mathcal{P}$ – rodzinę rozkładów na $\mathcal{A}$.

W naszym wypadku $\Omega = \mathbb{R}^{m+n}$, zaś

$$\mathcal{P} = \{p_{\Delta,F} : \Delta \geq 0,\ F \in \mathbf{F}\},$$

gdzie $\mathbf{F}$ jest rodziną rozkładów o ciągłych dystrybuantach.

Rozkłady z rodziny $\mathcal{P}$ można podzielić na takie, dla których hipoteza jest prawdziwa i takie, dla których jest ona fałszywa. Oznaczmy przez H i K powstałe w ten sposób rozłączne klasy rozkładów.

Hipoteza H z punktu 1. jest równoważna stwierdzeniu, że rozkład $p_{\Delta,F}$ należy do klasy H, stąd wygodnie jest posługiwać się tym samym oznaczeniem. To samo dotyczy klasy K. Rozkłady należące do klasy K nazywać będziemy alternatywami.

Każdy test przyporządkowuje klasom H i K odpowiadające im podzbiory zbioru Ω : Ω_H i Ω_K takie, że $\Omega_H \cup \Omega_K = \Omega$, $\Omega_H \cap \Omega_K = \emptyset$. Podzbiór Ω_H nazywamy obszarem przyjęcia hipotezy H, a podzbiór Ω_K – obszarem odrzucenia hipotezy H (obszarem krytycznym). Ponadto, niech α oznacza poziom istotności testu (czyli założone górne ograniczenie na prawdopodobieństwo odrzucenia hipotezy H, gdy jest ona prawdziwa), zaś

$$T = T(X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n) \text{ – statystykę testową,}$$

czyli pewną funkcję próby, zależną od stosowanego testu taką, że: $T(\Omega) = \overline{\Omega}$, $T(\Omega_H) = \overline{\Omega}_H$, $T(\Omega_K) = \overline{\Omega}_K$ oraz $\overline{\Omega}_H \cup \overline{\Omega}_K = \overline{\Omega}$, $\overline{\Omega}_H \cap \overline{\Omega}_K = \emptyset$.

W związku z powyższymi zależnościami mówiąc o obszarze krytycznym (obszarze przyjęcia hipotezy) będziemy mieć odtąd na myśli $\overline{\Omega}_K$ ($\overline{\Omega}_H$).

Zgodnie z przyjętymi założeniami

$$P_{0,F}\{T \in \overline{\Omega}_K\} \leq \alpha \quad \forall_{F \in \mathbf{F}}$$

gdzie $P_{0,F}\{\cdot\}$ oznacza prawdopodobieństwo zajścia zdarzenia $\{\cdot\}$, gdy mamy do czynienia z rozkładem o dystrybuancie $F(x)$ i gdy wartość parametru $\Delta = 0$ (czyli gdy hipoteza H jest prawdziwa).

W rozważanych przez nas testach obszar krytyczny $\overline{\Omega}_K$ ma postać: $T \geq c$ bądź $T \leq c$, gdzie $c = c(\alpha)$ – wartość krytyczna testu zależna od przyjętego poziomu istotności.

3. Dla zweryfikowania postawionej hipotezy postępujemy następująco:

–wybieramy test (o statystyce testowej T)

–obieramy poziom istotności α, który wyznacza nam obszar krytyczny $\overline{\Omega}_K$

–obliczamy wartość T statystyki z próby

–jeżeli $T \in \overline{\Omega}$, to hipoteza zostaje odrzucona (na rzecz hipotezy alternatywnej)

–jeżeli $T \notin \overline{\Omega}$, to mówimy, że nie ma podstaw do odrzucenia hipotezy H na poziomie istotności α.

II. Przegląd testów.

Definicje i oznaczenia:

– Niech $Z_1, Z_2, \ldots, Z_N$ $(N = m+n)$ będzie uporządkowaną niemalejąco próbą złożoną z $X_1, X_2, \ldots, X_m$ i $Y_1, Y_2, \ldots, Y_n$ (przypominamy, że zdarzenie losowe $\{Z_1 < Z_2 < \ldots < Z_N\}$ ma prawdopodobieństwo 1 – por. V.2.). Niech S_1 będzie numerem najmniejszego Y-ka w próbie Z-tów, S_2 – numerem drugiego co do wielkości Y – ka, ..., S_n – numerem największego Y-ka wśród Z-tów. Liczby $S_1, S_2, \ldots, S_n$ nazywamy rangami Y-ków. Podobnie definiujemy rangi X-ów $(R_1, R_2, \ldots, R_m)$. Ponieważ każdy ze zbiorów $\{R_1, R_2, \ldots, R_m\}$, $\{S_1, S_2, \ldots, S_n\}$ wyznacza drugi, więc wystarczy rozważać tylko jedną z tych statystyk.

– Funkcja indykatorowa:

$$I(\gamma) = \begin{cases} 1 & \text{gdy relacja } \gamma \text{ zachodzi} \\ 0 & \text{wpp} \end{cases}$$

gdzie γ jest pewną relacją.

– Dystrybuanty empiryczne:

$$F_m(x) = \frac{1}{m} \sum_{i=1}^{m} I(X_i \leq x), \qquad G_n(y) = \frac{1}{n} \sum_{j=1}^{n} I(Y_j \leq y).$$

W opisie każdego testu podano statystykę, w oparciu o którą zbudowany jest ów test oraz postać obszaru krytycznego. O ile to możliwe, starano się podawać tradycyjne oznaczenia statystyk. Odnośniki bibliograficzne zamieszczone po nazwie testu wskazują na pozycje, w których można znaleźć tablice wartości krytycznych omawianego testu (wartości krytyczne testu Savage'a, jak również wartości krytyczne testu Van der Waerdena dla niektórych poziomów istotności, zostały dla potrzeb niniejszej pracy policzone za pomocą odpowiedniego programu komputerowego).

1. Test Wilcoxona (test Manna–Whitneya–Wilcoxona): [3], [5], [6], [10], [18], [26]

statystyka: $$W_S = \sum_{j=1}^{n} S_j$$

obszar krytyczny: $W_S \geq c$

Statystyka W_S nosi nazwę statystyki Wilcoxona. Znana jest również statystyka Manna–Whitneya dana wzorem:

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} I(Y_j - X_i > 0)$$

związana ze statystyką Wilcoxona zależnością:

$$U = W_S - \frac{1}{2} \cdot n \cdot (n+1).$$

2. Test Fishera–Yatesa ("Normal Scores Test"): [8], [26]

statystyka: $$T = \sum_{j=1}^{n} a(S_j),$$

gdzie $a(k) = E_\Phi(\xi_{k:N})$ jest wartością oczekiwaną k-tej statystyki pozycyjnej w N-elementowej próbie $\xi_1, \xi_2, \ldots, \xi_N$ pochodzącej ze standardowego rozkładu normalnego.

obszar krytyczny: $T \geq c$.

3. Test Van der Waerdena: [23], [26]

statystyka: $$T = \sum_{j=1}^{n} a(S_j),$$

gdzie $a(k) = \Phi^{-1}\left(\frac{k}{N+1}\right)$, Φ jest dystrybuantą standardowego rozkładu normalnego.

obszar krytyczny: $T \geq c$.

4. Test Savage'a:

statystyka: $$T = \sum_{j=1}^{n} a(S_j),$$

gdzie $a(k) = E_e(\xi_{k:N})$ jest wartością oczekiwana k-tej statystyki pozycyjnej w N-elementowej próbie $\xi_1, \xi_2, \ldots, \xi_N$ ze standardowego rozkładu wykładniczego. Wartości te można obliczyć ze wzoru:

$$a(k) = \frac{1}{N} + \frac{1}{N-1} + \ldots + \frac{1}{N-k+1}$$

obszar krytyczny: $T \geq c$.

5. Test Lemmera: [12], [13]

statystyka: $$SU = \sum_{i=1}^{m} \sum_{j=1}^{n} r(|X_i - Y_j|) I(X_i - Y_j > 0),$$

gdzie $r(|X_i - Y_j|)$ jest rangą modułu różnicy $|X_i - Y_j|$ w uporządkowanej próbie złożonej ze wszystkich różnic między X-ami i Y-ami.

obszar krytyczny: $SU \leq c$.

6. Test mediany (test Browna–Mooda): [5]

statystyka: T = liczba Y-ów większych od mediany połączonej próby X-ów i Y-ów

obszar krytyczny: $T \geq c$.

7. Test Rosenbauma (tzw. nieparametryczny test parametru położenia): [18], [19], [26]

statystyka: T = liczba Y-ów większych od największego X-sa

obszar krytyczny: $T \geq c$.

8. Test Walda–Wolfowitza (test serii): [3], [26]

statystyka: W = liczba serii w połączonej próbie X-ów i Y-ów,

gdzie serią nazywamy taki ciąg wyrazów $z_j, z_{j+1}, \ldots, z_{j+l}$ $l = 0, 1, \ldots, n-j$, że $z_{j-1} \neq z_j = z_{j+1} = \ldots = z_{j+l} \neq z_{j+l+1}$ (u nas $z_j = X$ albo $z_j = Y$ $\forall_j$)

obszar krytyczny: $W \leq c$.

9. Test Kołmogorowa–Smirnowa: [3], [5], [18], [26]

statystyka: $D_{m,n} = \max_z |G_n(z) - F_m(z)|$

W praktyce korzysta się ze wzorów:

$$D^+_{m,n} = \max_{1 \leq j \leq n} \left(\frac{j}{n} - F_m(Y_{j:n}) \right) = \max_{1 \leq i \leq m} \left(G_n(X_{i:m}) - \frac{i-1}{m} \right)$$

$$D^-_{m,n} = \max_{1 \leq j \leq n} \left(F_m(Y_{j:n}) - \frac{j-1}{n} \right) = \max_{1 \leq i \leq m} \left(\frac{i}{m} - G_n(X_{i:m}) \right)$$

$$D_{m,n} = \max\{D^+_{m,n}, D^-_{m,n}\}$$

obszar krytyczny: $D_{m,n} \geq c$.

10. Test Gniedenki–Koroliuka: [3]

Test ten przeznaczony jest wyłącznie do sytuacji, gdy obie próby są tej samej liczności, tzn. $m = n$.

statystyka: $D = \max_{1 \leq k \leq 2n} |C_k|$,

gdzie $C_k = \sum_{j=1}^{k} d_j$, $d_j = \begin{cases} 1 & \text{gdy } Z_j \in \{Y_1, Y_2, \ldots, Y_n\} \\ -1 & \text{gdy } Z_j \in \{X_1, X_2, \ldots, X_m\} \end{cases}$

obszar krytyczny: $D \geq c$.

11. Test Kuipera (test Maaga–Stephensa): [3], [15]

statystyka:

$$V_{m,n} = \frac{1}{m \cdot n}\left[\max_{1 \le i \le m}\{(n+m)i - mR_i\} + \max_{1 \le j \le n}\{(m+n)j - S_j\}\right]$$

obszar krytyczny: $V_{m,n} \ge c$.

12. Test Cramera-von Misesa: [1], [3]

statystyka:

$$W^2_{m,n} = \frac{mn}{\sqrt{m+n}} \sum_{i=1}^{m+n} d_i^2$$

gdzie $d_i = G_n(Z_i) - F_m(Z_i)$, tzn. różnica dystrybuant empirycznych w i-tym punkcie połączonej próby X-ów i Y-ów.

W praktyce korzystamy ze wzoru:

$$W^2_{m,n} = \frac{n}{m+n} \sum_{i=1}^{m} \left(\frac{R_i - 1}{n} - \frac{i - \frac{1}{2}}{m}\right)^2 + \frac{n(2m+n)}{12mn(m+n)}$$

obszar krytyczny: $W^2_{m,n} \ge c$.

13. Test Watsona: [1], [3]

statystyka:

$$U^2_{m,n} = \frac{mn}{\sqrt{m+n}} \sum_{i=1}^{m+n} (d_i - \overline{d})^2$$

gdzie $d_i = G_n(Z_i) - F_m(Z_i)$, tzn. różnica dystrybuant empirycznych w i-tym punkcie połączonej próby X-ów i Y-ów

$$\overline{d} = \sum_{i=1}^{m+n} d_i$$

W praktyce korzystamy ze wzoru:

$$U^2_{m,n} = \frac{n}{m+n} \sum_{j=1}^{m} (h_j - \overline{h})^2 + \frac{n(2m+n)}{12mn(m+n)}$$

gdzie $h_j = \frac{R_j - j}{n} - \frac{j - \frac{1}{2}}{m}$, $\overline{h} = \frac{1}{m}\sum_{j=1}^{m} h_j$

obszar krytyczny: $U^2_{m,n} \ge c$.

III. Porównania mocy testów. Wskazówki i zalecenia użytkowe.

1. Podstawową charakterystyką testu statystycznego jest jego moc (tzn. prawdopodobieństwo odrzucenia hipotezy H, gdy jest ona fałszywa). Przy ustalonym rozkładzie o dystrybuancie F i różnicy parametru położenia Δ, moc testu o statystyce testowej T dana jest wyrażeniem:

$$\Pi_F(\Delta) = P_{F,\Delta}\{T \in \overline{\Omega}_k\}.$$

Dla ustalonego rozkładu F prawdopodobieństwo to, traktowane jako funkcja Δ, będziemy nazywać funkcją mocy testu.

Pożądaną cechą testu jest to, by miał on możliwie dużą moc. Co więcej, interesująca byłaby znajomość testu jednostajnie najmocniejszego, czyli takiego, który byłby mocniejszy od innych testów przy dowolnej alternatywie. Jednakże w naszym przypadku nie istnieje test jednostajnie najmocniejszy (przy pewnej alternatywie każdy z testów jest mocniejszy od pozostałych). I tak na przykład, dla dostatecznie małych Δ, test Wilcoxona jest najmocniejszy przy alternatywie F o rozkładzie logistycznym: $F(x) = 1/(1+e^{-x})$, zaś test Fishera–Yatesa dla alternatywy o rozkładzie normalnym.

Możemy więc pytać, który z testów jest lokalnie najmocniejszy, tzn. najmocniejszy przy określonej alternatywie. Dla odpowiedzi na to pytanie konieczne jest wyznaczenie funkcji mocy poszczególnych testów. Analityczne przedstawienie funkcji mocy jest jednak w większości przypadków bardzo trudne lub wręcz niewykonalne. Istnieją pewne wyniki osiągnięte dla przypadków asymptotycznych, tzn. gdy liczności prób m, n dążą do nieskończoności. Mankamentem tych wyników jest jednak brak sensownej interpretacji, wszak mamy zawsze do czynienia ze skończonymi próbami, co więcej, często liczności prób są bardzo małe. W tym miejscu pomocne stają się metody symulacyjne.

Zamieszczone poniżej (aneks) wyniki badań mocy testów uzyskane zostały właśnie metodą symulacji dla 2000 prób losowych (dana próba każdorazowo była poddana wszystkim testom omówionym w rozdziale II). Symulacje przeprowadzono na maszynie IBM XT PC.

Badania zostały przeprowadzone dla pięciu różnych rozkładów: jednostajnego na odcinku (0,1), normalnego standardowego N(0,1), wykładniczego standardowego oraz dla dwóch rozkładów gamma: $G(2,1)$ i $G(4,1)$, gdzie $G(\alpha,\beta)$ oznacza rozkład o gęstości:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \quad , x > 0, \quad \alpha, \beta > 0.$$

Wybór tych właśnie rozkładów podyktowany został dwoma względami: po pierwsze, reprezentują one klasy rozkładów o różnych własnościach (np kwestia symetrii, wielkości nośnika itd.); po drugie, wiele spotykanych w zastosowaniach rozkładów jest tej postaci, bądź "zbliżonej" doń (a to sugeruje, że i własności będą, być może, "zbliżone").

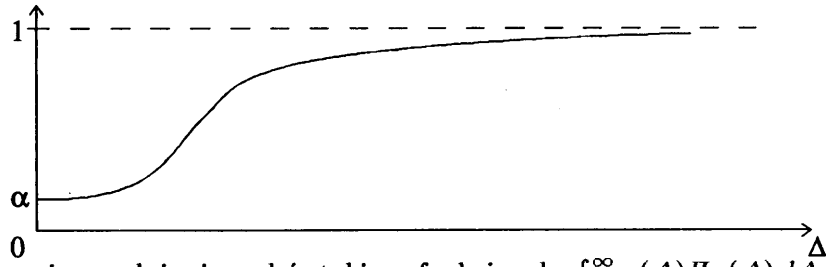
Testowanie prowadzono na poziomie istotności 0.05. Aby uzyskać rozmiar testu równy założonemu poziomowi istotności zastosowano randomizację.

Do generowania poszczególnych rozkładów zastosowano metody zamieszczone w pracach: [17], [25].

Badania przeprowadzono na małych próbkach: $m = 5$, $n = 5$; $m = 8$,

$n = 8$; $m = 4$, $n = 8$.

Zamieszczone wyniki zawierają uzyskane wartości funkcji mocy poszczególnych testów dla kilku wartości przesunięcia Δ oraz tzw. indeks mocy. Wielkość tę wprowadzamy po to, aby móc porównywać moce testów posługując się jedną tylko liczbą. Jest to więc pewien funkcjonał określony na funkcjach mocy. Funkcje mocy mają zwykle postać jak na rysunku:



Słuszny więc wydaje się wybór takiego funkcjonału $\int_o^\infty \omega(\Delta)\Pi_F(\Delta)\,d\Delta$, gdzie $\omega(\Delta)$ jest funkcją wagową, który "obciążałby" większą wagą małe wartości Δ, a mniejszą duże Δ (jest to zgodne z intuicją, gdyż bardziej wartościowy jest test, który umie rozróżnić rozkłady różniące się niewiele, co odpowiada małemu Δ).

Ponieważ obliczaliśmy funkcję mocy tylko dla kilku wartości parametru Δ, będziemy rozważać funkcjonał postać:

$$\sum_{j=1}^{v} \omega(\Delta)\Pi_F(\Delta_j)$$

gdzie v jest liczbą punktów, dla których wyznaczono funkcję mocy;

$$\forall_j\ \omega(\Delta_j) \geq 0\,;\ \sum_{j=1}^{v} \omega(\Delta_j) = 1\,.$$

W naszym wypadku zdecydowaliśmy się na wybór $v = 6$; $\omega(\Delta_1) = 0$, $\omega(\Delta_2) = 0.5$, $\omega(\Delta_3) = 0.3$, $\omega(\Delta_4) = 0.15$, $\omega(\Delta_5) = 0.05$, $\omega(\Delta_6) = 0$ (przy czym Δ_j przyjmują różne wartości dla różnych rozkładów). Wyniki badań symulacyjnych zamieszczono w aneksie.

2. Wnioski:

2.1 zdecydowanie wyróżnia się grupa pięciu najmocniejszych testów, są to: test Wilcoxona, Fishera-Yatesa, Van-der-Waerdena, Savage'a i Lemmera. Spośród nich najsłabszy jest test Savage'a.

2.2 drugą z kolei (co do mocy) grupę testów stanowią: test Rosenbauma, mediany, Cramera-von Misesa, Kołmogorowa-Smirnowa i Gniedenki-Koroliuka; przy czym, w większości przypadków, pierwsze trzy z powyższych testów są mocniejsze od dwóch ostatnich.

2.3 najsłabszymi testami są: test Walda-Wolfowitza, Watsona i Kuipera; wśród nich test Walda-Wolfowitza jest naogół najmocniejszy, a Kuipera - najsłabszy.

2.4 na uwagę zasługuje test Lemmera. Jest to najnowszy spośród przedstawionych testów; jak dotychczas ukazały się tylko dwie prace mu poświęcone [12], [13], w których jego twórca porównywał ten test z testem Wilcoxona i testem t-Studenta. Test ten, w większości badanych przypadków, okazał się testem najmocniejszym. Przyczyną tego wydaje się fakt, iż "koduje" on dość dużo informacji o próbach. Mankamentem tego testu jest to, iż w praktyce można go używać jedynie dla małych próbek (gdyż z uwagi na budowę testu konieczne jest nadanie rang różnicom X_i i Y_j, a liczba tych różnic wynosi $m \cdot n$ – rośnie więc szybko wraz ze wzrostem liczności próbek; tymczasem sortowanie jest dość czasochłonne).

2.5 niektórzy autorzy przy porównywaniu mocy testów ograniczają się jedynie do małych Δ. Zamieszczone wyniki wskazują, że przy tym kryterium oceny testów wnioski 2.1–2.3 są również prawdziwe. Warto podkreślić, że stosowany przez nas indeks mocy wydaje się być kryterium sensowniejszym, gdyż bierze pod uwagę zachowanie się funkcji mocy dla różnych wartości parametru Δ i zawiera w sobie kryterium oceny rozważane w tym punkcie (duże wagi dla małych Δ).

IV. Poziom krytyczny testu. Obliczanie poziomu krytycznego testu Wilcoxona.

1. Powszechnie stosowany sposób testowania hipotez sprowadza się do porównania obliczonej wartości statystyki z wartością krytyczną zależną od przyjętego poziomu istotności.

W zastosowaniach mamy często do czynienia z monotonicznymi rodzinami obszarów krytycznych $\{\overline{\Omega}_K^{\alpha}\}$ odpowiadających różnym poziomom istotności α (tzn. $\forall_{\alpha 1<\alpha 2}\ \overline{\Omega}_K^{\alpha 1} \subseteq \overline{\Omega}_K^{\alpha 2}$). W takim przypadku dobrze jest ustalić nie tylko, czy hipoteza zostaje przyjęta, czy odrzucona na danym poziomie istotności, lecz również określić poziom krytyczny, czyli najmniejszy poziom istotności $\overline{\alpha}$, na którym hipoteza zostałaby odrzucona przy danej obserwacji. Liczba ta daje pewien pogląd na to, w jakim stopniu obserwacje zaprzeczają hipotezie (lub potwierdzają ją) i umożliwia innym wyciągnięcie wniosku opartego na wybranym przez nich poziomie istotności.

Podanie poziomu krytycznego jest więc innym metodologicznie sposobem podejścia do problemu testowania hipotez. To, że owe podejście, mimo przedstawionych powyżej zalet ustępuje tradycyjnemu, wynika z trudności powstających przy obliczaniu poziomu krytycznego.

Poniżej przedstawiony zostanie algorytm obliczania poziomu krytycznego dla najpopularniejszego testu, jakim jest test Wilcoxona.

2. Bez straty ogólności zakładamy, że $m \leq n$ i rozważamy statystykę

$$W_R = \frac{1}{2}(m+n)(m+n+1) - W_S$$

(jak łatwo zauważyć: $W_R = \sum_{i=1}^{m} R_i$). Poziom krytyczny testu Wilcoxona wyraża się następująco:

$$\overline{\alpha} = \overline{\alpha}(v) = P_{0,F}\{W_R \leq v\},$$

gdzie v jest wynikiem obserwacji. Wielkość $\overline{\alpha}$ jest poprawnie zdefiniowana, bo prawdopodobieństwo $P_{0,F}\{\cdot\}$ w teście Wilkoxona nie zależy od rozkładu F.

Jak więc widać, potrzebna jest umiejętność obliczania dystrybuanty zmiennej losowej W_R. Szukane prawdopodobieństwo jest równe ilorazowi liczby takich permutacji X-ów i Y-ów, że $W_R \leq v$ oraz liczby $\binom{m+n}{n}$.

Najmniejszą wartością, jaką może przyjąć W_R jest $\frac{m(m+1)}{2}$ możemy zatem zapisać $W_R = \frac{m(m+1)}{2} + U$, $U \in \{0, 1, 2, \ldots, m \cdot n\}$ (łatwo zauważyć, że $U = \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i - Y_j > 0)$ jest statystyką Manna–Whitneya, dualną do przedstawionej w II.1).

Liczba sposobów otrzymania danej wartości statystyki U jest równa współczynnikowi przy t^U w rozwinięciu następującej funkcji [6]:

$$g(t) = \frac{(1-t^{m+1})(1-t^{m+2})\ldots(1-t^{m+n})}{(1-t)(1-t^2)\ldots(1-t^n)}$$

Oznaczmy tę liczbę sposobów przez $f(U)$. Zatem

$$g(t) = \prod_{i=1}^{n} \frac{(1-t^{m+i})}{(1-t^i)} = f(0) + t \cdot f(1) + \ldots + t^U \cdot f(U) + \ldots$$

Po obustronnym zlogarytmowaniu pierwszej równości otrzymamy:

$$\log g(t) = \sum_{i=1}^{n} \log(1-t^{m+i}) - \sum_{i=1}^{n} \log(1-t^i).$$

Stąd

$$\frac{g'(t)}{g(t)} = \sum_{i=1}^{n} \frac{-(m+i)t^{m+i-1}}{1-t^{m+1}} + \sum_{i=1}^{n} \frac{it^{i-1}}{1-t^i}.$$

Po odpowiednich przekształceniach dostajemy:

$$f(1) + 2 \cdot t \cdot f(2) + \ldots + U \cdot t^{U-1} \cdot f(U) + \ldots =$$
$$= \left[f(0) + t \cdot f(1) + \ldots + t^U \cdot f(U) + \ldots\right] \times$$
$$\times \left[\sum_{i=1}^{n} \sum_{j=1}^{\infty} -(m+i)t^{(m+i)j-1} + \sum_{i=1}^{n} \sum_{j=1}^{\infty} it^{ij-1}\right].$$

Niech z_i oznacza współczynnik przy t po prawej stronie równania. Porównując współczynniki przy t^{U-1} po obu stronach równania otrzymujemy:

$$f(U) = \frac{1}{U} \sum_{i=1}^{U-1} f(i) z_{U-i-1}, \qquad U = 1, 2, \ldots, m \cdot n$$

przy czym $f(0) = 1$.

Znając $f(U)$ dla każdego $U \leq v$ możemy już wyznaczyć liczbę sposobów otrzymania wartości $W_R \leq v$; jest ona równa:

$$f(0) + f(1) + \ldots + f(v).$$

A oto procedura obliczania poziomu krytycznego testu Wilcoxona zapisana w języku PASCAL:

```
FUNCTION Wilcoxon(m, n, u:integer):real;
   { m, n – wielkości próbek (1 ≤ m ≤ n ≤ 20) }
   {          u – wartość statystyki          }
function licznik(m, n, u:integer):real;
var z:array[0..400]of real;
    max,i, j, k1, k2:integer;
    s, l:real;
begin
   max:=u – 1;
   for i:=0 to max do z[i]:=0;
   for i:=1 to n do
      begin
         j:=0;
         repeat
            j:=j + 1;
            k1:=(m + i) * j – 1;
            k2:=i * j – 1;
            if k1 <=max then z[k1]:=z[k1] – m – 1;
            if k2 <=max then z[k2]:=z[k2] + i
         until ((k1 >max) and (k2 >max))
      end;
   f[0]:=1;
   for j:= 1 to u do
      begin
         s:=0;
         for i:=0 to j – 1 do s:=s + f[i] * z[j – i – 1];
         f[j]:=s/j
      end;
   l:=0;
```

```
    for i:=0 to u do l:=l + f[i];
    licznik:=l
end; { licznik }

function mianownik(m, n:integer):real;
var i:integer;
    p:real;
begin
    p:=1;
    for i:=1 to m do p:=p * (n + i)/i;
    mianownik:=p
end; { mianownik }

BEGIN
    u:=u - m * (m + 1) div 2;
    Wilcoxon:=licznik(m, n, u)/mianownik(m, n)
END. { Wilcoxon }
```

3. Uwagi:

3.1 istnieje prostsza analitycznie procedura rekurencyjna obliczania liczby sposobów otrzymania $U \leq v$ podana w pracy [16]:

$$f_{m,n}(U) = f_{m-1,n}(U - n) + f_{m,n-1}(U)$$

gdzie $f_{m,n}(-U) = 0 \quad \forall_{U>0}$, $f_{m,n}(0) = 1$, $f_{m,n}(U) = f_{n,m}(U)$, $f_{0,n}(U) = 1$; $f_{m,n}(U)$ oznacza to samo, co nasze dotychczasowe $f(U)$ przy wielkości prób m i n. Okazuje się jednak, że realizacja numeryczna tak prostej procedury ujawnia jej mankamenty – procedura jest bardzo wolna i zajmuje dużo pamięci. W przeciwieństwie do niej procedura przedstawiona w punkcie 2. jest bardzo szybka i wymaga niewiele pamięci.

3.2 jak łatwo zauważyć, przedstawiona w 2. procedura może być użyta do obliczania dystrybuanty statystyki Manna–Whitneya.

3.3 przyjęte w procedurze ograniczenia na wielkość prób spowodowane były koniecznością zadeklarowania wielkości używanych tablic. Oczywiście zmieniwszy wpierw wielkość tablic możemy stosować tę procedurę dla dowolnych prób m, n $(m \leq n)$.

V. Miscellanea.

1. Założenia przyjęte w rozdziale I. dotyczące niezależności zmiennych losowych $X_1, X_2, \ldots, X_m$; $Y_1, Y_2, \ldots, Y_n$ oraz ciągłości dystrybuant F i G opisują model, który można intuicyjnie zilustrować następująco: z danej (dużej) populacji losujemy $N = m + n$ obiektów, z których m poddawanych jest pewnemu zabiegowi A, a n – zabiegowi B. Weryfikujemy hipotezę, że te zabiegi przynoszą takie same efekty, wobec hipotezy alternatywnej, że efekty zabiegu B są "większe" od efektów zabiegu A. Ten model, z oczywistych

przyczyn, nosi w literaturze nazwę "population model".

W praktyce spotykamy się często z sytuacją, gdy mamy dane N obiektów, które losowo dzielimy na dwie grupy: m i n elementowe ($m+n=N$), poddawane następnie zabiegom A i B. Jest to tzw "randomization model".

Choć w obu przypadkach stosujemy te same testy, warto sobie zdawać sprawę z różnic pomiędzy tymi modelami (dotyczy to np. uogólniania wniosków). Dyskusję tego zagadnienia zawiera praca [10].

2. Zwróćmy uwagę na jedną z implikacji "upraszczania" założeń. Ciągłość dystrybuant F i G zapewniała nam, że z prawdopodobieństwem jeden, wszystkie obserwacje $X_1, X_2, \ldots, X_m$; $Y_1, Y_2, \ldots, Y_n$ będą różne. Tymczasem opuszczenie tego założenia, tzn. dopuszczenie możliwości powtarzania się obserwacji, wymaga odmiennego podejścia do kwestii rangowania. Szersze omówienie tego problemu zawiera praca [10]. Tutaj podamy jedynie jeden ze sposobów postępowania w takim przypadku.

Niech $Z_1, Z_2, \ldots, Z_N$ będzie, jak poprzednio, uporządkowaną niemalejąco połączoną próbą X-ów i Y-ów. Załóżmy, że pewne obserwacje powtarzają się, np.

$$Z_1 \neq Z_{i+1} = Z_{i+2} = \ldots = Z_{i+k} \neq Z_{i+k+1}\,.$$

W tym wypadku obserwacjom równym sobie nadajemy jednakowe rangi równe średniej arytmetycznej rang, które przypisanoby obserwacjom Z_{i+1}, $Z_{i+2}, \ldots, Z_{i+k}$ gdyby nie były one sobie równe. W danym przypadku (gdyby pozostałe obserwacje były różne) mielibyśmy:

$$\begin{cases} r'(Z_j) = r(Z_j) & \text{dla } j < i \text{ oraz } j > i+k \\ r'(Z_j) = \frac{r(Z_{i+1})+r(Z_{i+2})+\ldots+r(Z_{i+k})}{k} \end{cases}$$

gdzie $r(Z_j)$ oznacza rangę obserwacji Z_j (w próbie o różnych obserwacjach), równą odpowiedniemu R_l bądź S_p, zależnie od tego, czy $Z_j \in \{X_1, X_2, \ldots, X_m\}$, czy też $Z_j \in \{Y_1, Y_2, \ldots, Y_n\}$ (r' – to tzw. "midranks").

3. Przyjęte w rozdziale I. założenia o testowaniu hipotezy $H : \Delta = 0$ kontra $K : \Delta > 0$, to szczególny przypadek ogólniejszej sytuacji: $H : F = G$, $K : F > G$ (F, G – ciągłe). Odrębnym zagadnieniem będzie sytuacja: $H : F = G$, $K : F \neq G$.

Niektóre spośród testów przedstawionych w rozdziale II. były konstruowane specjalnie jako testy parametru położenia (np. test Rosenbauma, test Lemmera). Większość z nich może być jednak z powodzeniem używana w ogólniejszych przypadkach. Spodziewane efekty zależeć będą oczywiście od hipotezy alternatywnej i od budowy statystyki stanowiącej podstawę testu.

4. Jedną z często spotykanych sytuacji jest testowanie parametru skali, tzn.: $X_1, X_2, \ldots, X_m$ – i.i.d. F, $Y_1, Y_2, \ldots, Y_n$ – i.i.d. G gdzie $G(x) = F(x/\sigma)$; $H : \sigma = 1$; $K : \sigma > 1$.

Można tu posługiwać się przedstawionymi testami, lecz bardziej wska-

zane jest użycie testów "wyspecjalizowanych" w tych przypadkach. Są to: test Siegel–Tukeya [10], test Capona [2], [10], test Klotza [7], [10], test Rosenbauma parametru skali [18], [19].

5. Inną typową sytuacją jest testowanie hipotez o zmiennych losowych zgrupowanych w bloki (tzw. "blocked comparison for two treatments"). Odpowiada to np. sytuacji, gdy dane N obiektów poddajemy najpierw zabiegowi A, a potem zabiegowi B. Istnieją tu również testy "specjalistyczne": test znaków [10], test znakowanych rang [10], test Lemmera [14].

6. Naturalnym uogólnieniem modelu rozważanego w rozdziale I. jest model k-próbkowy:

$$X_1, X_2, \ldots, X_{n_1} - \text{i.i.d. } F_1$$
$$X_1, X_2, \ldots, X_{n_2} - \text{i.i.d. } F_2$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$X_1, X_2, \ldots, X_{n_k} - \text{i.i.d. } F_k$$
$$H : F_1 = F_2 = \ldots = F_k.$$

Do weryfikacji tej hipotezy służy test Kruskala–Wallisa [9], [10].

7. W przypadku prób o dużej liczności posługujemy się rozkładami przybliżonymi (np. normalnym dla testu Wilcoxona i Van der Waerdena, t-Studenta dla testu Fishera–Yatesa, χ^2 dla testu mediany).

Bibliografia

[1] E. Burr, *Small-Sample Distributions of the Two-Sample Cramer–von Mises' W^2 and Watsons's U^2*, Ann. Math. Statist. 35 (1964), 1091–1098.

[2] J. Capon, *Asymptotic Efficiency of Certain Locally Most Powerful Rank Tests*, Ann. Math. Statist. 32 (1961), 88–100.

[3] Cz. Domański, *Statystyczne testy nieparametryczne*, PWE, Warszawa, 1979.

[4] *Encyclopedia of Statistical Sciences*, Wiley, New York, 1981–1988.

[5] J. D. Gibbons, J. W. Pratt, *Concepts of Nonparametric Theory*, Springer Verlag, 1981.

[6] H. L. Harter, B. D. Owen, *Selected Tables in Mathematical Statistics*, vol. I., American Mathematical Society, 1973.

[7] J. Klotz, *Nonparametric Tests for Scale*, Ann. Math. Statist. 33 (1962):498–512.

[8] —, *On the Normal Scores Two-Sample Rank Test*, J. Am. Statist. Assoc. 59 (1964), 652–664.

[9] W. H. Kruskal, W. A. Wallis, *Use of Ranks in One-Criterion Variance Analysis*, J. Am. Statist. Assoc. 47 (1952), 583–621.

[10] E. L. Lehmann, *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisco, 1975.

[11] —, *Testowanie hipotez statystycznych*, PWN, Warszawa, 1968

[12] H. H. Lemmer, *A Modified Mann-Whitney-Wilcoxon Test for the Two-Sample Location Problem*, Research Report 86/1, Department of Statistics, Rand Africaans University, 1986.

[13] —, *A Modified Mann-Whitney-Wilcoxon Test for the Two-Sample Location Problem*, J. Statist. Comput. Simul. 27 (1987), 307–319.

[14] —, *A Test for the Median, Combining the Sign and Signed-Rank Tests*, Commun. Statist.-Simula. 16 (1987), 621–627.

[15] U. R. Maag, M. A. Stephens, *The V_{nm} Two-Sample Test*, Ann. Math. Statist. 39 (1968), 923–935.

[16] H. B. Mann, D. R. Whitney, *On a Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other*, Ann. Math. Statist. 18 (1947), 50–60.

[17] G. Marsaglia, M. D. MacLaren, T. A. Bray, *A Fast Procedure for Generating Normal Random Variables*, Comm. Assoc. Comp. Mach. 7 (1964), 4–10.

[18] D. B. Owen, *Handbook of Statistical Tables*, Addison-Wesley, 1962.

[19] S. Rosenbaum, *Tables for a Nonparametric Test of Location*, Ann. Math. Statist. 25 (1954), 146–150.

[20] —, *Tables for a Nonparametric Test of Dispersion*, Ann. Math. Statist. 24 (1953), 663-668.

[21] I. R. Savage, *Bibliography of Nonparametric Statistics*, Harvard Univ. Press, 1962.

[22] —, *Contributions to the Theory of Rank Order Statistics – the Two-Sample Case*, Ann. Math. Statist. 27 (1956), 590-615.

[23] B. L. Van der Waerden, E. Nievergelt, *Tables for Comparing Two Samples by X-Test and Sign Test*, Springer Verlag, Berlin, 1956.

[24] A. Wald, J. Wolfowitz, *On a Test Whether Two Samples are From the Same Population*, Ann. Math. Statist. 11 (1940), 147-162.

[25] R. Zieliński, *Generatory liczb losowych*, Wyd. Naukowo-Techniczne, Warszawa, 1979.

[26] R. Zieliński, W. Zieliński, *Tablice statystyczne*, wyd. II – rękopis.

Aneks. Wyniki badań symulacyjnych.

Wyniki zamieszczone w tabelach obrazują moce poszczególnych testów (pomnożone przez 1000) dla różnych wartości parametru skali Δ.

Przyjęte oznaczenia:

Wil. – test Wilcoxona
F–Y – „ Fishera–Yatesa
VdW – „ Van der Waerdena
Sav. – „ Savage'a
Lem. – „ Lemmera
Ros. – „ Rosenbauma
K–S – „ Kołmogorowa–Smirnowa
W–W – „ Walda–Wolfowitza
Med. – „ mediany
G–K – „ Gniedenki–Koroliuka
Kui. – „ Kuipera
CvM – „ Cramera–von Misesa
Wat. – „ Watsona

rozkład jednostajny

$m = 5$, $n = 5$							
test	Δ = 0.0	0.1	0.2	0.4	0.6	0.8	indeks mocy
Wil.	048	125	230	590	878	993	264
F-Y	047	119	221	585	875	993	258
VdW	047	119	221	585	875	993	258
Sav.	045	117	206	561	862	992	248
Lem.	044	122	221	626	925	999	268
Ros.	044	110	188	461	735	955	217
K-S	049	076	107	307	587	865	145
W-W	047	066	065	151	376	732	094
Med.	052	094	150	324	576	841	169
G-K	044	075	108	304	586	865	145
Kui.	041	064	057	137	346	718	087
CvM	042	078	128	402	765	970	176
Wat.	044	062	057	133	351	726	086

$m = 8$, $n = 8$							
test	Δ = 0.0	0.1	0.2	0.4	0.6	0.8	indeks mocy
Wil.	045	154	359	770	978	1000	349
F-Y	044	168	386	811	985	1000	371
VdW	043	166	382	807	984	1000	368
Sav.	046	156	359	789	979	1000	353
Lem.	045	156	379	835	992	1000	367
Ros.	048	131	282	668	935	998	297
K-S	048	077	161	476	830	991	200
W-W	055	056	110	305	627	942	138
Med.	042	113	212	488	812	975	234
G-K	043	075	161	467	830	991	197
Kui.	050	056	077	216	558	928	112
CvM	042	077	198	586	935	1000	233
Wat.	050	046	068	201	566	949	102

$m = 4$, $n = 8$							
test	Δ = 0.0	0.1	0.2	0.4	0.6	0.8	indeks mocy
Wil.	053	123	248	587	906	996	269
F-Y	048	127	263	606	915	997	279
VdW	052	126	256	598	909	997	275
Sav.	051	116	234	535	889	994	253
Lem.	048	117	243	625	948	999	273
Ros.	057	096	182	433	771	966	207
K-S	050	065	109	302	633	906	143
W-W	070	079	120	269	508	776	142
Med.	057	090	166	368	715	952	186
Kui.	054	048	062	127	347	709	079
CvM	053	074	129	405	779	979	175
Wat.	060	047	065	130	360	727	080

rozkład normalny $N(0,1)$

$m = 5,\ n = 5$							
test	Δ = 0.0	0.5	1.0	1.5	2.0	2.5	indeks mocy
Wil.	048	181	401	667	857	963	354
F-Y	047	175	393	659	854	963	347
VdW	047	175	393	659	854	963	347
Sav.	048	175	377	628	834	949	337
Lem.	050	184	423	690	886	973	367
Ros.	047	153	304	516	725	875	281
K-S	048	107	212	426	607	791	212
W-W	049	072	141	257	401	598	137
Med.	057	138	261	444	608	768	245
G-K	050	101	210	417	606	768	207
Kui.	053	074	127	240	390	590	131
CvM	051	115	263	505	738	898	250
Wat.	044	074	129	240	403	604	132

$m = 8,\ n = 8$							
test	Δ = 0.0	0.5	1.0	1.5	2.0	2.5	indeks mocy
Wil.	043	217	593	863	977	1000	465
F-Y	046	226	594	868	976	1000	470
VdW	046	225	596	868	977	1000	470
Sav.	052	205	550	832	964	997	441
Lem.	045	238	624	885	981	1000	488
Ros.	051	153	403	640	847	947	336
K-S	048	115	357	643	866	969	304
W-W	048	075	198	378	639	838	186
Med.	050	167	467	687	881	967	371
G-K	036	115	367	636	860	966	306
Kui.	049	074	193	372	645	837	183
CvM	039	116	418	731	938	995	340
Wat.	049	071	189	397	667	874	185

$m = 4,\ n = 8$							
test	Δ = 0.0	0.5	1.0	1.5	2.0	2.5	indeks mocy
Wil.	050	192	429	680	887	965	371
F-Y	049	186	432	683	892	969	370
VdW	050	187	434	687	894	967	372
Sav.	050	192	416	658	867	954	363
Lem.	050	193	442	710	915	974	382
Ros.	057	164	354	583	783	901	315
K-S	048	097	258	472	679	845	231
W-W	058	088	188	343	526	709	178
Med.	046	150	314	545	746	885	289
Kui.	046	060	131	254	421	624	128
CvM	054	108	288	546	774	914	261
Wat.	053	063	127	255	438	647	130

rozkład wykładniczy $E(0,1)$

$m = 5, n = 5$							
test	Δ = 0.0	0.2	0.5	1.0	2.0	3.0	indeks mocy
Wil.	058	124	257	535	862	963	263
F-Y	054	120	254	522	857	960	258
VdW	054	120	254	522	857	960	258
Sav.	051	103	209	433	784	924	218
Lem.	056	114	232	522	883	981	249
Ros.	054	080	151	308	667	858	165
K-S	048	074	150	395	751	911	179
W-W	042	059	127	303	650	853	146
Med.	056	087	190	389	727	899	195
G-K	049	068	154	395	754	919	177
Kui.	044	058	120	292	645	846	141
CvM	050	072	175	405	780	924	188
Wat.	044	065	131	297	652	852	149

$m = 8, n = 8$							
test	Δ = 0.0	0.2	0.5	1.0	2.0	3.0	indeks mocy
Wil.	053	154	408	754	977	998	362
F-Y	051	154	414	734	969	996	360
VdW	051	155	418	739	969	996	363
Sav.	056	121	284	542	859	952	270
Lem.	056	131	336	680	967	994	317
Ros.	052	086	154	295	625	838	165
K-S	045	084	264	646	964	997	266
W-W	052	070	196	489	867	978	211
Med.	052	115	288	587	918	992	278
G-K	046	090	252	650	962	999	266
Kui.	053	074	183	484	872	979	208
CvM	043	088	266	659	968	997	271
Wat.	055	066	192	507	887	989	211

$m = 4, n = 8$							
test	Δ = 0.0	0.2	0.5	1.0	2.0	3.0	indeks mocy
Wil.	042	128	290	533	853	938	274
F-Y	040	139	306	555	866	944	288
VdW	042	136	304	555	866	944	286
Sav.	048	112	242	449	774	907	235
Lem.	048	126	289	555	890	975	277
Ros.	048	090	176	349	713	877	186
K-S	039	090	187	413	781	912	202
W-W	055	099	188	430	758	915	209
Med.	041	088	177	358	717	878	187
Kui.	044	069	129	300	672	865	152
CvM	041	089	194	416	766	906	204
Wat.	053	070	132	306	689	876	155

rozkład gamma $G(2,1)$

$m = 5,\ n = 5$							
test	Δ = 0.0	0.5	1.0	2.0	3.0	4.0	indeks mocy
Wil.	043	142	328	672	893	959	315
F-Y	041	136	323	666	888	957	309
VdW	041	136	323	666	888	957	309
Sav.	043	121	272	583	823	926	271
Lem.	045	147	320	685	910	978	318
Ros.	044	109	196	452	702	859	216
K-S	049	068	181	480	741	888	198
W-W	050	062	122	341	589	799	148
Med.	042	113	228	482	703	863	233
G-K	049	076	180	475	744	895	201
Kui.	049	058	113	322	576	791	140
CvM	047	081	209	537	810	924	224
Wat.	051	059	122	328	585	793	145

$m = 8,\ n = 8$							
test	Δ = 0.0	0.5	1.0	2.0	3.0	4.0	indeks mocy
Wil.	050	198	473	869	982	998	421
F-Y	050	201	474	866	976	997	422
VdW	049	202	479	868	976	997	424
Sav.	050	164	358	710	897	958	341
Lem.	050	193	450	852	975	998	408
Ros.	052	105	204	478	712	852	221
K-S	050	107	280	750	954	995	298
W-W	049	080	177	541	831	952	216
Med.	056	152	337	718	921	981	331
G-K	043	102	286	745	953	995	296
Kui.	053	089	172	520	826	950	215
CvM	043	107	315	785	967	997	314
Wat.	051	076	186	559	852	961	220

$m = 4,\ n = 8$							
test	Δ = 0.0	0.5	1.0	2.0	3.0	4.0	indeks mocy
Wil.	047	160	361	705	866	948	337
F-Y	047	162	376	718	877	957	346
VdW	047	159	375	718	877	957	344
Sav.	044	142	305	617	795	915	295
Lem.	049	162	368	727	903	974	346
Ros.	042	113	233	506	727	877	239
K-S	043	090	218	535	751	902	228
W-W	061	092	207	481	711	869	216
Med.	046	117	218	489	718	876	234
Kui.	052	062	131	354	608	807	154
CvM	035	091	236	560	781	913	240
Wat.	048	055	140	372	624	816	157

rozkład gamma $G(4,1)$

$m = 5$, $n = 5$								
test	Δ =	0.0	0.5	1.2	2.0	3.0	4.5	indeks mocy
Wil.		047	102	229	431	700	906	220
F-Y		043	098	224	426	695	902	215
VdW		043	098	224	426	695	902	215
Sav.		045	095	202	366	614	852	194
Lem.		048	102	237	449	713	922	225
Ros.		044	085	152	258	476	734	151
K-S		054	067	129	235	467	744	131
W-W		042	057	082	155	310	581	092
Med.		050	088	166	294	491	714	163
G-K		049	059	130	245	469	747	129
Kui.		044	057	082	143	303	567	090
CvM		050	067	148	284	545	829	148
Wat.		045	057	083	147	301	574	091

$m = 8$, $n = 8$								
test	Δ =	0.0	0.5	1.2	2.0	3.0	4.5	indeks mocy
Wil.		046	105	280	569	848	986	264
F-Y		043	106	283	579	850	986	267
VdW		043	105	285	577	851	986	267
Sav.		043	109	247	482	735	934	238
Lem.		044	109	291	578	848	989	271
Ros.		053	094	182	306	496	774	173
K-S		057	074	142	348	650	942	165
W-W		056	058	094	196	414	780	108
Med.		047	089	205	412	667	922	201
G-K		059	072	144	336	662	942	163
Kui.		055	060	082	177	409	773	102
CvM		051	063	159	406	735	974	177
Wat.		055	054	073	192	417	814	099

$m = 4$, $n =$								
test	Δ =	0.0	0.5	1.2	2.0	3.0	4.5	indeks mocy
Wil.		047	125	247	487	727	913	246
F-Y		045	124	260	506	744	925	253
VdW		045	123	256	502	741	925	251
Sav.		045	116	220	440	655	867	223
Lem.		046	133	258	504	759	941	258
Ros.		046	104	179	348	541	781	185
K-S		046	070	132	309	505	783	146
W-W		054	070	133	261	451	717	137
Med.		050	094	170	342	519	768	176
Kui.		048	055	088	175	311	619	096
CvM		048	081	142	343	571	839	163
Wat.		049	063	088	175	315	637	100

Summary

Nonparametric tests for the two-sample location problem

Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be two independent random samples from populations with continuous distribution functions $F(x)$ and $G(x) = F(x-\Delta)$ respectively. In order to test $H : \Delta = 0$ against $K : \Delta > 0$ several tests can be used. Thirteen such nonparametric tests are presented here.

By means of a sampling study a comparison of these tests are performed. The criteria of the power and the power index was used.

Related problems are discussed as well.

An algorithm (in Pascal) for computation of the critical level of the Mann–Whitney–Wilcoxon test is given also.