

Metody nieparametryczne w statystyce

Tomasz Wójtowicz

Wydział Zarządzania

AGH Akademia Górniczo-Hutnicza im. S. Staszica w Krakowie

Definicja

Niech dla pewnej wartości n X_1, \dots, X_n będzie próbą pseudolosową z rozkładu zmiennej losowej X . Średnią

$$\bar{h} = \frac{1}{n} \sum_{i=1}^n h(X_i)$$

nazywamy estymatorem parametru $\theta = E(h(X))$ otrzymanym metodą Monte Carlo.

Metropolis N., Ulam S., (1949) The Monte Carlo Method, „Journal of the American Statistical Association”, Vol. 44, No. 247.

Przedział ufności

Za pomocą metody Monte Carlo możemy też wyznaczyć przedział ufności dla dowolnej zmiennej losowej W .

Dla zadanego współczynnika ufności $1 - \alpha$ szukamy takiego przedziału

$\left[q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}} \right]$, że

$$P\left(q_{\frac{\alpha}{2}} \leq W \leq q_{1-\frac{\alpha}{2}} \right) = 1 - \alpha.$$

W tym celu wystarczy wyznaczyć kwantyle $q_{\frac{\alpha}{2}}$ i $q_{1-\frac{\alpha}{2}}$ rzędu $\frac{\alpha}{2}$ i $1 - \frac{\alpha}{2}$ rozkładu W .

Na podstawie wylosowanej próby o długości n wyznaczamy statystyki porządkowe rzędu $[n\alpha/2]$ i $[n(1 - \alpha/2)]$ i oznaczamy je przez $\tilde{w}_{n,\alpha/2}$ i $\tilde{w}_{n,1-\alpha/2}$.

Wtedy:

$$P\left(\tilde{w}_{n,\alpha/2} \leq W \leq \tilde{w}_{n,1-\alpha/2} \right) \approx 1 - \alpha.$$

Testy permutacyjne

Testy permutacyjne

Przykład

Niech x_1, \dots, x_n będą wartościami cechy X o rozkładzie z parametrem θ_X .

Niech y_1, \dots, y_m będą wartościami cechy Y o rozkładzie z parametrem θ_Y .

Interesuje nas porównanie wartości parametrów θ_X i θ_Y , np.

$$H_0: \theta_X = \theta_Y$$

$$H_1: \theta_X \neq \theta_Y$$

Rozważanymi parametrami mogą być np.

- wartości oczekiwane,
- mediany,
- wariancje.

Testy permutacyjne

Przykład (testowanie jednorodności)

W tabeli zamieszczono wyniki pewnego egzaminu:

studenci	20.5	18.5	21	30	26	29	26	30.5
studentki	24	29.5	29	8	23.5			

Czy na tej podstawie można stwierdzić, że rozkłady wyników studentów i studentek nie różnią się istotnie?

W tym celu możemy sprawdzić np. równości wartości oczekiwanych:

$$H_0: m_1 = m_2.$$

Jednak nie są spełnione założenia, żadnej ze znanych nam wersji (lub nie jesteśmy w stanie ich zweryfikować).

Testy permutacyjne

Hipotezę:

$$H_0: m_1 = m_2$$

można rozumieć też następująco:

wynik egzaminu nie zależy od płci.

W tym przypadku mamy:

- n - liczba wyników studentów, m - liczba wyników studentek,
- $\bar{x} = 25,1875$, $\bar{y} = 22,8$.
- $\hat{\theta}_0 = \bar{x} - \bar{y} = 2,3875$

Testy permutacyjne

Sposób rozwiązania:

- połączmy obie próby (x i y) w jedną i uporządkujmy rosnąco (oznaczymy ten wektor przez v),
- wektorowi v przyporządkujemy wektor zer i jedynek r taki, że

$$r_i = \begin{cases} 1 & v_i = \textit{student} \\ 0 & v_i = \textit{studentka} \end{cases}$$

- niech V i R będą zmiennymi losowymi zdefiniowanymi analogicznie jak v i r

Testy permutacyjne

Stwierdzenie

Jeżeli prawdziwa jest hipoteza H_0 , to dla ustalonej wartości $V = v$ zmienna losowa R ma rozkład jednostajny na zbiorze wszystkich ciągów zero-jedynkowych o długości $n + m$ zawierających n jedynek.

Zdefiniujmy:

$$p_V = P_{H_0}(|\hat{\theta}| \geq |\hat{\theta}_0| \text{ dla } V = v)$$

W tym przypadku:

$$p_V = P_{H_0}(|\hat{\theta}| \geq 2,3875 \text{ dla } V = v)$$

Ta definicja jest analogiczna do definicji *p-value* (z tym, że jest to wartość warunkowa).

Testy permutacyjne

Jeżeli rozważymy wszystkie możliwe wartości $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_N$ zmiennej losowej $\hat{\theta}$ dla ustalonego wektora $V = v$, to:

$$p_V = \frac{1}{N} \cdot \#\{|\hat{\theta}_i| \geq 2,3875, i = 1, \dots, N\}$$

Dokładny test permutacyjny - jeżeli $i = 1, \dots, N$ obejmuje wszystkie możliwe przypadki.

Przybliżony test permutacyjny - jeżeli $i = 1, \dots, N$ obejmuje część losowo wybranych przypadków.

Testy permutacyjne

Przykład:

Na podstawie wyników 14 osób zbadać istotność zależności pomiędzy wynikiem z zaliczenia i wynikiem z egzaminu:

zaliczenie	55	57,5	64,4	66,9	69,3	54,7	60,6	82,5	58,6	64,7	50,7	50,7	51,1	55,3
egzamin	38,9	55,6	53,7	48,1	56,5	44,4	54,6	53,7	43,5	56,5	50	1,9	31,5	27,8

Na podstawie tych danych możemy obliczyć wartość współczynnika korelacji Pearsona pomiędzy wynikami z zaliczenia i z egzaminu:

$$r = 0,564$$

Teraz należy ocenić czy zależność ta jest istotna czy nie. Należy więc przeprowadzić test istotności współczynnika korelacji.

Testy permutacyjne

Test istotności współczynnika korelacji ma postać:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Do oceny prawdziwości hipotezy głównej możemy zastosować test permutacyjny.

Jeżeli hipoteza H_0 jest prawdziwa, to fakt, że liczby w parze (x_i, y_i) są wynikami konkretnego studenta nie ma znaczenia. Równie dobrze można wynik y_i zastąpić wynikiem dowolnego innego studenta.

Z tego wynika, że aby przeprowadzić test permutacyjny wartości pierwszej cechy (tzn. zaliczenie) pozostawiamy bez zmian, a permutujemy wartości drugiej cechy (tzn. egzamin).

Testy permutacyjne

Przykład

Na podstawie wyników 14 osób zbadać istotność zależności pomiędzy wynikiem z zaliczenia i wynikiem z egzaminu:

zaliczenie	55	57,5	64,4	66,9	69,3	54,7	60,6	82,5	58,6	64,7	50,7	50,7	51,1	55,3
egzamin	38,9	55,6	53,7	48,1	56,5	44,4	54,6	53,7	43,5	56,5	50	1,9	31,5	27,8

W tym przypadku mamy $14! = 87178291200$ permutacji.

Rozsądniej jest więc przeprowadzić przybliżony test permutacyjny i rozważyć tylko część tych permutacji (**wybranych losowo!**)