

クラスカル・ウォリス検定 (exact test)

青木繁伸

2020年3月17日

1 目的

正確な p 値を与えるクラスカル・ウォリス検定 (正確確率検定) である。データ (およびそれから導かれる分割表) によっては計算量が多くなり実用的な時間内で計算が終了できないこともあるので、そのような場合にはモンテカルロ法による近似計算を用いる必要があるかもしれない。

周辺和を固定した全ての分割表においてクラスカル・ウォリス検定の統計量と生起確率を求め、実際に観察された分割表の統計量より小さいか等しい分割表の生起確率を合計したものが p 値であるとするものである。

2 使用法

```
import sys
sys.path.append("statlib")
from exact import exact_kw_test
exact_kw_test(*x, g=None, exact=True, hybrid=False, loop=10000, verbose=True)
```

2.1 引数

<code>x</code>	分割表 (合計を含まない) もしくは群ごとに複数のデータベクトルを列挙
<code>g</code>	<code>x</code> が 1 つのデータベクトルのときは、群を表すデータベクトル
<code>exact</code>	正確検定を行うかどうか
<code>hybrid</code>	<code>True</code> にすれば、シミュレーションによる
<code>loop</code>	シミュレーションの回数
<code>verbose</code>	必要最小限のプリント出力をする

2.2 戻り値の名前

<code>"exactP"</code>	正確検定による p 値
<code>"simP"</code>	モンテカルロ法による p 値
<code>"chisq"</code>	検定統計量 (χ^2 分布にしたがう)
<code>"df"</code>	自由度
<code>"asymP"</code>	漸近検定による p 値

3 使用例

3.1 2次元配列（二重リスト）で指定する場合

```
x = [[5, 3, 2, 1], [4, 3, 5, 2], [2, 3, 1, 2]]

import sys
sys.path.append("statlib")
from exact import exact_kw_test

a = exact_kw_test(x)
```

```
Kruskal-Wallis test
chisq = 1.3248, df = 2, p value = 0.51560
exact p value = 0.52682
24871 tables are inspected.
```

10000回のシミュレーションによる p 値を求める

```
import numpy as np
np.random.seed(123) # 乱数の種（通常は不要）
a = exact_kw_test(x, hybrid=True)
```

```
Kruskal-Wallis test
chisq = 1.3248, df = 2, p value = 0.51560
simulated p value = 0.52540
based on 10000 times simulation.
```

3.2 データベクトルと群データベクトルで指定する場合

グループを表す群データベクトルは引き数の名前として $g=$ を付けなければならない。

```
data = [
    3.42, 3.84, 3.96, 3.76,
    3.17, 3.63, 3.47, 3.44, 3.39,
    3.64, 3.72, 3.91
]
group = [1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3]

import sys
sys.path.append("statlib")
from exact import exact_kw_test

a = exact_kw_test(data, g=group)
```

```
Kruskal-Wallis test
```

```
chisq = 5.5487, df = 2, p value = 0.06239
exact p value = 0.05390
27720 tables are inspected.
```

3.3 群ごとに複数のデータベクトルで指定する場合

```
a = exact_kw_test([3.42, 3.84, 3.96, 3.76],
                  [3.17, 3.63, 3.47, 3.44, 3.39],
                  [3.64, 3.72, 3.91])
```

Kruskal-Wallis test

```
chisq = 5.5487, df = 2, p value = 0.06239
exact p value = 0.05390
27720 tables are inspected.
```

以下のような分割表を与えるのと同じである。

```
a = exact_kw_test([[0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1],
                  [1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0],
                  [0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0]])
```

Kruskal-Wallis test

```
chisq = 5.5487, df = 2, p value = 0.06239
exact p value = 0.05390
27720 tables are inspected.
```