

Fisher の exact test

青木繁伸

2020年3月17日

1 目的

Fisher の正確確率検定を行う。

R には `fisher.test` 関数がある (Python から R の `fisher.test` 関数を使用する方法については、第 4 節を参照)。

ここに取り上げる `fisher` 関数は、フィッシャー流の p 値の決め方 (`fisher.test` と同じ) の他に、ピアソン流の p 値の決め方も選べる。また、計算量が多量になり実用的な時間内に計算が終了できないような場合に対応するために、モンテカルロ法に基づく近似検定も選択できる (`fisher.test` の `hybrid` オプションと同じ)。

フィッシャー流の p 値の決め方とは、周辺和を固定した全ての分割表の生起確率をもとめ、実際に観察された分割表の生起確率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。

ピアソン流の p 値の決め方とは、周辺和を固定した全ての分割表においてピアソンのカイ二乗統計量と生起確率を求め、実際に観察された分割表のピアソンのカイ二乗統計より小さいか等しい分割表の生起確率を合計したものを p 値とするものである。

本来、Fisher の exact test の目的は、独立性の検定にあると思われ、独立性からのずれを評価する統計指標は様々ある。しかるに、フィッシャー流の p 値の決め方は、基盤とする統計量を用いていない。単に生起確率が観察された分割表の生起確率より小さいことに基づくフィッシャー流の p 値の決め方は不適切であると考えるものである。

2 使用法

```
import sys
sys.path.append("statlib")
from exact import Fisher_exact_test
Fisher_exact_test(x, y=NULL, exact=TRUE, method="Fisher", hybrid=FALSE, loop=10000,
                  verbose=True)}
```

2.1 引数

<code>x</code>	分割表 (合計を含まない) もしくはベクトル
<code>y</code>	<code>x</code> がベクトルのときは、ベクトル <code>x</code> が分割表のときには無視される

<code>exact</code>	正確な p 値を求める場合、またはシミュレーションにより近似的な p 値を求めるときには <code>True</code> (デフォルト) <code>False</code> の場合にはカイ二乗分布による漸近近似検定のみを行う
<code>method</code>	p 値の決め方。"Fisher" か "Pearson" を指定する。省略した場合には "Fisher" が仮定される
<code>hybrid</code>	<code>True</code> を指定すれば、シミュレーションにより近似的な p 値を計算する
<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>"chisq"</code>	χ^2 分布を利用する漸近検定統計量
<code>"df"</code>	自由度
<code>"asymP"</code>	漸近検定による p 値
<code>"exactP"</code>	正確な p 値
<code>"simP"</code>	シミュレーションによる p 値

3 使用例

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

3.1 フィッシャーの方法による正確な p 値

```
import sys
sys.path.append("statlib")
from exact import Fisher_exact_test

a = Fisher_exact_test(x)
```

```
Fisher's Exact Test for Count Data
chisq = 3.3963, df = 6, p value = 0.75771
exact p value = 0.80911 (Fisher)
24871 tables are inspected.
```

モンテカルロ法による p 値

```
import numpy as np
np.random.seed(12345) # 指定しなくてよい
a = Fisher_exact_test(x, hybrid=True)
```

```
Fisher's Exact Test for Count Data
```

```
chisq = 3.3963, df = 6, p value = 0.75771
simulated p value = 0.81450
based on 10000 times simulation.
```

3.2 ピアソンの方法による正確な p 値

```
a = Fisher_exact_test(x, method="Pearson")
```

```
Fisher's Exact Test for Count Data
chisq = 3.3963, df = 6, p value = 0.75771
exact p value = 0.78782 (Pearson)
24871 tables are inspected.
```

モンテカルロ法による p 値

```
a = Fisher_exact_test(x, method="Pearson", hybrid=True)
```

```
Fisher's Exact Test for Count Data
chisq = 3.3963, df = 6, p value = 0.75771
simulated p value = 0.79270
based on 10000 times simulation.
```

3.3 2 変数を与える場合

```
x = [1, 2, 3, 2, 1, 2, 3, 2, 1, 2, 3, 2, 2, 3, 1]
y = ["A", "B", "C"]*5
a = Fisher_exact_test(x, y)
```

```
Fisher's Exact Test for Count Data
chisq = 1.2857, df = 4, p value = 0.86379
exact p value = 1.00000 (Fisher)
180 tables are inspected.
```

4 R の `fisher.test()` を使う方法

下準備として、`rpy2` をインストールしておく。

```
pip install rpy2
```

この後、以下のように実行する。

```
import numpy as np
import rpy2.robjects.numpy2ri
from rpy2.robjects.packages import importr
rpy2.robjects.numpy2ri.activate()
```

```
stats = importr('stats')
m = np.array([[4,4],[4,5],[10,6]]) # 対象とする分割表の定義
res = stats.fisher_test(m) # fisher.test 起動
print('pvalue: {}'.format(res[0][0])) # P 値は
print(res)
print(type(res))
print(np.array(res).shape)
print(res[0][0])
print(res[1][0])
print(res[2][0])
print(res[3][0])
```