

# 二項分布モデルでのP値

- 黒木玄
- 2025-11-26

## 目次

- ▼ [1 二項検定のP値に関する解説の例](#)
  - [1.1 P値](#)
  - [1.2 P値の計算の仕方](#)
  - [1.3 両側P値の計算練習問題](#)
  - [1.4 S値](#)
  - [1.5 相性の良さ/悪さという解釈の仕方](#)
  - [1.6 P値関数とS値関数](#)
  - [1.7 信頼区間 = 相性区間](#)
  - [1.8 報告の仕方について](#)
- [2 二項検定の複数のP値のグラフ](#)

In [1]: ▶

```
1 ENV["LINES"] = 200
2 ENV["COLUMNS"] = 200
3
4 const mincho = "ipamp"
5 const gothic = "ipagp"
6
7 using Distributions
8 using StatsPlots
9 default(fmt=:png, fontfamily=gothic,
10         titlefontsize=13, legendfontsize=13, guidefontsize=13)
11
12  $x \approx y = x < y$  ||  $x \approx y$ 
13  $x \approx y = x > y$  ||  $x \approx y$ 
14
15 """
16     central法による二項分布の両側P値
17     二項分布binにおける「k以下になる確率の2倍」と「k以上になる確率の2倍」の小さい方の値
18
19     このP値からClopper-Pearsonの信頼区間が得られる。
20 """
21 pvalue_central(bin, k) = min(2cdf(bin, k), 2ccdf(bin, k-1))
22
23 """
24 Clopper-Pearsonの信頼区間
25 """
26 function confint_clopper_pearson(n, k;  $\alpha$  = 0.05)
27     p_L = k > 0 ? quantile(Beta(k, n-k+1),  $\alpha/2$ ) : zero( $\alpha$ )
28     p_U = k < n ? quantile(Beta(k+1, n-k), 1- $\alpha/2$ ) : one( $\alpha$ )
29     [p_L, p_U]
30 end
31
32 """
33     minimum likelihood法による二項分布の両側P値
34     このP値からSterneの信頼区間が得られる。
35 """
36 function pvalue_minlike(bin, k)
37     supp = support(bin)
38     sum(pdf(bin, i) for i in supp if pdf(bin, i)  $\approx$  pdf(bin, k))
39 end
40
41 """
42     score法による二項分布の両側P値
43     このP値は二項分布の正規分布近似で定義された両側P値と同じ。
44
45     このP値はWilsonのscore信頼区間を与える。
46 """
47 function pvalue_score(bin, k)
48     (; n, p) = bin
49     phat = k/n
50     se = sqrt(p * (1 - p) / n)
51     z = (phat - p) / se
52     2ccdf(Normal(), abs(z))
53 end
54
55 """
56     Wald法による二項分布の両側P値
57     このP値も二項分布の正規分布近似で定義された両側P値だが、データ \n, k\ から推定する必要がある
58
59     このP値は教科書によく書いてある母比率に関するWaldの信頼区間を与える。
60 """
61 function pvalue_wald(bin, k)
62     (; n, p) = bin
63     phat = k/n
64     sehat = sqrt(phat * (1 - phat) / n)
65     z = (phat - p) / sehat
66     2ccdf(Normal(), abs(z))
67 end
68
69 bin = Binomial(20, 0.6)
70 ks = (7, 8, 16, 17)
71 @show pvalue_central.(bin, ks)
72 @show pvalue_minlike.(bin, ks)
73 @show pvalue_score.(bin, ks)
74 @show pvalue_wald.(bin, ks)
75 ;
```

```
pvalue_central.(bin, ks) = (0.04205785495554228, 0.1130527340685061, 0.10190390638833315, 0.031922325580016536)
pvalue_minlike.(bin, ks) = (0.03699009026777941, 0.10747832022841937, 0.07198088067193767, 0.022427038142043273)
pvalue_score.(bin, ks) = (0.02247887336612527, 0.06788915486182907, 0.06788915486182893, 0.02247887336612527)
pvalue_wald.(bin, ks) = (0.0190763221017784, 0.06788915486182907, 0.025347318677468197, 0.0017414012528258463)
```

```
In [2]: 1 ccdf(Binomial(20, 0.6), 18-1), 2ccdf(Binomial(20, 0.6), 18-1)
```

```
Out[2]: (0.003611472059128875, 0.00722294411825775)
```

```
In [3]: 1 cdf(Binomial(20, 0.6), 6), 2cdf(Binomial(20, 0.6), 6)
```

```
Out[3]: (0.00646587535203501, 0.01293175070407002)
```

## 1 二項検定のP値に関する解説の例

以下の例は中原治著『基礎から学ぶ統計学』の「1-5 二項検定」の節の「[5] STEP 4: 棄却域と有意水準」以降の説明をP値に関する説明で置き換えることを意識して書かれたものです。

その本の例題1.1では、B薬は60%の患者に効果があることが分かっていると仮定されており、A薬については無作為に選んだ患者20人中18人に効果があったというデータの値が得られていると仮定されています。以下の説明でもこの設定を使います。

### 1.1 P値

例題1.1の文脈では、特別な数値でその値の **小ささ** が次のように解釈できるようなものをデータの値から計算できると便利です：

A薬が無作為に選んだ患者20人中18人に効いたというデータの値が得られたとき、A薬が効く患者の割合がB薬と同じ60%だと信じている人は確率的にどれだけ意外だと感じるか。

以下では、この目的のために使えるP値という数値の構成の仕方についてについて、例題1.1の場合を例に説明します。

例題1.1の状況では、上の目的のために使える **片側P値** を

効果がある患者の割合が60%だと仮定したとき、無作為に選んだ患者20人中18人 **以上** に効果がある確率

と定義できます。この確率の値(片側P値の値)は0.36%とかなり小さくなります(計算の仕方は後で説明します)。そのことは、「A薬とB薬が効く患者の割合が等しい」と信じている人にとって、「A薬が無作為に選んだ患者20人中18人に効いた」というデータの値はかなり意外だと感じられることを意味しています。

さらに **両側P値** が上の片側P値の2倍で定義されます。この場合の両側P値は0.72%でやはりかなり小さい。より詳しくは、この両側P値を

「A薬が無作為に選んだ患者20人中18人に効いた」というデータの値に関する「A薬が効く患者の割合が60%である」という検定仮説の両側P値

と呼びます。検定仮説を単に仮説と呼ぶこともあります。伝統的な用語法では検定仮説は帰無仮説とも呼ばれています。帰無仮説(null hypothesis)は意味が分かり難く、英単語のnullはzeroという意味を持つせいで、差がないことを意味するゼロ仮説を主に意味するかのよう誤解されがちなので、この解説では帰無仮説という用語を避けることにします。(検定仮説(test hypothesis)という用語を初期のNeymanが使っていたので、検定仮説もある意味で十分に由緒正しい用語だと考えられる。「検定」は英語のtestの翻訳で「検定仮説」は「データの値とモデルを使ってテストする仮説」というような意味になります。)

両側P値の小ささは

「A薬が効く患者の割合が60%である」と信じている人が、「A薬が無作為に選んだ患者20人中18人に効いた」と感じる意外さ

を表しています。

以下では単にP値と言ったら、両側P値を意味するものとします。

P値は小さいほど、モデルと検定仮説の組み合わせの下でデータの値の意外さが大きいことを意味します。

**P値の報告の仕方:** P値を  $P$  と書くとき、 $P < 0.05$  のようにP値の具体的な値が不明になるような報告の仕方をしてはいけません。P値を報告する場合には  $P = 3.2\%$  のように具体的な値がわかるように報告するようにしましょう。ただし、例外的にP値が  $P = 0.000023 = 2.3 \times 10^{-5}$  のように非常に小さな値になり、近似計算の精度を超えている場合には  $P < 0.0001$  のように報告し

ても良いでしょう。 $P = 0.04205785495554228$  のように無意味に高過ぎる制度でP値を報告してもいけません。複雑な現実における統計分析では、理想的なデータの取得は不可能だったり、統計分析のために採用する統計モデルの妥当性自体が不明だったりするので、高精度なP値の値には意味がないと考えられます。一般に、統計分析の結果を報告するときには、P値などの統計分析で得られる数値では測ることができないバイアスや不確実性があることを謙虚に認めることが必要になります。この解説ではP値(P-value)のPを大文字で書くことにしたのですが、小文字の $p$ と書いても問題ありません。どちらでもよいです。

以下の注意は最初に読むときにはとばしても構いません。

**片側確率の2倍の両側P値の正確な定義と欠点:**「無作為に選んだ  $n$  人中  $k$  人に薬が効いた」というデータの値が得られたときの、片側確率の2倍で定義された「薬が効く患者の割合は $p$ である」という仮説の両側P値の正確な定義は

薬が効く確率が  $p$  であるという仮定の下で、無作為に選んだ  $n$  人中  $k$  人以上に効く確率の2倍と無作為に選んだ  $n$  人中  $k$  人以下に効く確率の小さい方の値

です。このように定義された両側P値の欠点は

- 無用に値が大きくなり過ぎるという性質を持つ。
- 1 より大きくなることもある。

P値は確率だとみなしたい値なので0以上1以下になって欲しいのですが、上の定義だと1より大きくなってしまいう場合が出て来てしまいます。その欠点をなくすために、

薬が効く確率が  $p$  であるという仮定の下で、無作為に選んだ  $n$  人中  $k$  人以上に効く確率の2倍と無作為に選んだ  $n$  人中  $k$  人以下に効く確率と 1 の中で最小値

と定義することも多いです。

**片側P値だけではなく、両側P値を考えたい理由:** A薬が無作為に選んだ患者20人中60%の12人よりもずっと大きな人数(例えば18人)に効果があった場合だけではなく、その反対側のずっと小さな人数(例えば5人)にしか効果がない場合にも、A薬が効く患者の割合が60%だと信じている人にとっては意外だと感じられるでしょう。このことから、無作為に選んだ患者20人中60%の12人からデータの値以上に離れる確率の小ささで、「A薬が効く患者の割合は60%」という仮説を信じている人から見たデータの値の意外さを評価するという考え方も自然だということに気がきます。そのような考え方で定義されるP値が両側P値です。上ではシンプルに反対側の確率はもとの片側確率に等しいだろうという大雑把な想定のもとで、片側確率の2倍で両側P値を定義しました。モデルの確率分布が左右非対称な場合には他にも無数の選択肢があるのですが、そういう問題については統計学の技術的側面についてもっと詳しくなってから考えれば十分です。

**尤度とP値の違いに注意:** 上では片側P値を「効果がある患者の割合が60%だと仮定したとき、無作為に選んだ患者20人中18人 以上 に効果がある確率」と定義し、両側P値をその2倍で定義しました。読者の中には太字になっている「以上」を削除して、「効果がある患者の割合が60%だと仮定したとき、無作為に選んだ患者20人中18人に効果がある確率」を考えたい人があります。その人は鋭い! 「以上」の語を取り去ってできる確率は **尤度(ゆうど)** と呼ばれており、統計学的に非常に基本的かつ重要な量になっています。尤度とP値を混同しないように注意して下さい。

## 1.2 P値の計算の仕方

以下では二項係数を

$$\binom{n}{k} = {}_nC_k = \frac{n!}{k!(n-k)!}$$

と表すことにします。例えば、 $0! = 1$  に注意すると、

$$\begin{aligned}\binom{20}{20} &= \frac{20!}{20!0!} = \frac{20!}{20!} = 1, \\ \binom{20}{19} &= \frac{20!}{19!1!} = \frac{20 \cdot 19 \cdot 18 \cdots 2 \cdot 1}{19 \cdot 18 \cdots 2 \cdot 1 \times 1} = 20, \\ \binom{20}{18} &= \frac{20!}{18!2!} = \frac{20 \cdot 19 \cdot 18 \cdot 17 \cdots 2 \cdot 1}{18 \cdot 17 \cdots 2 \cdot 1 \times 2} = 190.\end{aligned}$$

${}_nC_k$  は高校数学では標準的な記号ですが、専門的な文献では  $\binom{n}{k}$  の方がよく使われているように思われるので主に  $\binom{n}{k}$  を使います。どちらを使うかは好みの問題でしかないので、 $\binom{n}{k}$  という書き方が嫌いな人は  ${}_nC_k$  と書いても構いません。

片側P値と呼ばれる

効果がある患者の割合が60%だと仮定したとき、無作為に選んだ患者20人中18人 **以上** に効果がある確率

は次のように計算される。

$$\bullet (20 \text{人中} k \text{人に効果がある確率}) = \binom{20}{k} 0.6^k (1 - 0.6)^{20-k}$$

なので、関数電卓その他を使って計算すると、

$$\begin{aligned} \bullet (20 \text{人中} 20 \text{人に効果がある確率}) &= \binom{20}{20} 0.6^{20} (1 - 0.6)^0 = 0.6^{20} \approx 3.6561584400629733 \times 10^{-5} \\ \bullet (20 \text{人中} 19 \text{人に効果がある確率}) &= \binom{20}{19} 0.6^{19} (1 - 0.6)^1 = 20 \cdot 0.6^{19} \cdot 0.4 \approx 0.00048748779200839646 \\ \bullet (20 \text{人中} 18 \text{人に効果がある確率}) &= \binom{20}{18} 0.6^{18} (1 - 0.6)^2 = 190 \cdot 0.6^{18} \cdot 0.4^2 \approx 0.00048748779200839646 \end{aligned}$$

ここで  $3.6561584400629733 \times 10^{-5}$  は  $3.6561584400629733 \times 10^{-5}$  を意味する。

上の3つの値の和を計算ことによって次を得る:

$$\bullet (\text{片側P値}) = (20 \text{人中} 18 \text{人以上に効果がある確率}) \approx 0.003611472059128871$$

両側P値はこれの2倍なので、

$$\bullet (\text{両側P値}) = 0.007222944118257742$$

以上によって片側P値は約0.36%で両側P値は約0.72%になることが分かった。

```
In [4]: 1 @show p20 = binomial(20, 20) * 0.6^20 * 0.4^0
        2 @show p19 = binomial(20, 19) * 0.6^19 * 0.4^1
        3 @show p18 = binomial(20, 18) * 0.6^18 * 0.4^2
        4 @show pval_one_sided = p20 + p19 + p18
        5 @show pval_both_sided = 2pval_one_sided;

p20 = binomial(20, 20) * 0.6 ^ 20 * 0.4 ^ 0 = 3.6561584400629733e-5
p19 = binomial(20, 19) * 0.6 ^ 19 * 0.4 ^ 1 = 0.00048748779200839646
p18 = binomial(20, 18) * 0.6 ^ 18 * 0.4 ^ 2 = 0.003087422682719845
pval_one_sided = p20 + p19 + p18 = 0.003611472059128871
pval_both_sided = 2pval_one_sided = 0.007222944118257742
```

### 1.3 両側P値の計算練習問題

**問題:**「薬Aが効く患者の割合は60%である」という仮定の下で「薬Aは無作為抽出された20人中 $k$ 人に効いた」というデータの値が得られる確率 $P(k)$ を0以上20以下のすべての整数 $k$ について求め、確率の数表を作成せよ。

**解答例:** 二項分布の確率  $P(k) = \binom{20}{k} 0.6^k 0.4^{20-k}$  を  $k = 0, 1, 2, \dots, 20$  について求めればよい。その結果は次のようになる:

k	P(k)	k	P(k)
0	1.09951e-8	12	0.179706
1	3.29853e-7	13	0.165882
2	4.70041e-6	14	0.124412
3	0.0000423037	15	0.074647
4	0.000269686	16	0.0349908
5	0.00129449	17	0.0123497
6	0.00485435	18	0.00308742
7	0.0145631	19	0.000487488
8	0.0354974	20	0.0000365616
9	0.0709949		
10	0.117142		
11	0.159738		

ただし、1.09951e-8は $1.09951 \times 10^{-8}$ を意味する。以上の結果はWolframAlphaに

```
Table[k, {k, 0, 20}], Table[Binomial[20, k]*0.6^k*0.4^(20-k), {k, 0, 20}] *2
```

と入力することによって得られる([WolframAlphaで実行 \(https://www.wolframalpha.com/input?i=Table%5Bk%2C+%7Bk%2C+0%2C+20%7D%5D%2C+Table%5BBinomial%5B20%2C+k%5D\\*0.6%5Ek\\*0.4%5E%2820-k%29%2C+%7Bk%2C+0%2C+20%7D%5D+%26amp%3Bja%29\)](https://www.wolframalpha.com/input?i=Table%5Bk%2C+%7Bk%2C+0%2C+20%7D%5D%2C+Table%5BBinomial%5B20%2C+k%5D*0.6%5Ek*0.4%5E%2820-k%29%2C+%7Bk%2C+0%2C+20%7D%5D+%26amp%3Bja%29))).

Table[k, {k, 0, 20}], Table[Binomial[20, k]\*0.6^k\*0.4^(20-k), {k, 0, 20}]

自然言語

数式入力

拡張キーボード

例を見る

アップロード

ランダムな例を使う

入力

$$\{ \text{Table}[k, \{k, 0, 20\}], \text{Table}\left[\binom{20}{k} \times 0.6^k \times 0.4^{20-k}, \{k, 0, 20\}\right] \}$$

$\binom{n}{m}$  は二項係数です

次元

2 (行) × 21 (列)

行列プロット

転置

$$\begin{pmatrix} 0 & 1.09951 \times 10^{-8} \\ 1 & 3.29853 \times 10^{-7} \\ 2 & 4.70041 \times 10^{-6} \\ 3 & 0.0000423037 \\ 4 & 0.000269686 \\ 5 & 0.00129449 \\ 6 & 0.00485435 \\ 7 & 0.0145631 \\ 8 & 0.0354974 \\ 9 & 0.0709949 \\ 10 & 0.117142 \\ 11 & 0.159738 \\ 12 & 0.179706 \\ 13 & 0.165882 \\ 14 & 0.124412 \\ 15 & 0.074647 \\ 16 & 0.0349908 \\ 17 & 0.0123497 \\ 18 & 0.00308742 \\ 19 & 0.000487488 \\ 20 & 0.0000365616 \end{pmatrix}$$

In [5]: 1 Any[0:11 pdf.(Binomial(20, 0.6), 0:11) 12:23 pdf.(Binomial(20, 0.6), 12:23)]

Out[5]: 12×4 Matrix{Any}:

0	1.09951e-8	12	0.179706
1	3.29853e-7	13	0.165882
2	4.70041e-6	14	0.124412
3	4.23037e-5	15	0.074647
4	0.000269686	16	0.0349908
5	0.00129449	17	0.0123497
6	0.00485435	18	0.00308742
7	0.0145631	19	0.000487488
8	0.0354974	20	3.65616e-5
9	0.0709949	21	0.0
10	0.117142	22	0.0
11	0.159738	23	0.0

**問題:**「薬Aは無作為に抽出された患者20人中6人に効いた」というデータの値に関する「薬Aが効く患者の割合は60%である」という仮説の両側P値を求めよ。ただし、その両側P値は片側確率の2倍で定義されたものとする。その計算で上の問題で作成した確率の数表を用いて良い。

**解答例:** 求めるべき両側P値は、「薬Aが効く患者の割合は60%である」という仮定の下で「薬Aが効いたの人数は無作為抽出された20人中6人以下であった」となる確率の2倍である。ゆえに、求めたい両側P値は、上の問題で求めた確率 $P(k)$ を $k = 0, 1, 2, 3, 4, 5, 6$ について足し合わせて2倍すれば求まる。

In [6]: 1 (1.09951e-8 +  
2 3.29853e-7 +  
3 4.70041e-6 +  
4 0.0000423037 +  
5 0.000269686 +  
6 0.00129449 +  
7 0.00485435) \* 2

Out[6]: 0.012931741916200001

「薬Aは無作為に抽出された患者20人中6人に効いた」というデータの値に関する「薬Aが効く患者の割合は60%である」という仮説の両側P値は約1.3%である。

この値はかなり小さめなので、「薬Aが効く患者の割合は60%である」という仮説の下で「薬Aは無作為に抽出された患者20人中6人に効いた」というデータの値はかなり意外だとみなされる。

```
In [7]: 1 pvalue_central(Binomial(20, 0.6), 6)
```

```
Out[7]: 0.01293175070407002
```

問題: 「薬Aは無作為に抽出された患者20人中15人に効いた」というデータの値に関する「薬Aが効く患者の割合は60%である」という仮説の両側P値を求めよ。ただし、その両側P値は片側確率の2倍で定義されたものとする。その計算で上の問題で作成した確率の数表を用いて良い。

解答例:  $P(k)$  の  $k = 15, 16, 17, 18, 19, 20$  に関する和の2倍を計算すれば良い。

```
In [8]: 1 (0.074647 +  
2 0.0349908 +  
3 0.0123497 +  
4 0.00308742 +  
5 0.000487488 +  
6 0.0000365616) * 2
```

```
Out[8]: 0.25119793920000005
```

「薬Aは無作為に抽出された患者20人中15人に効いた」というデータの値に関する「薬Aが効く患者の割合は60%である」という仮説の両側P値は約25%である。

この25%のP値はかなり大きな値なので、「薬Aが効く患者の割合は60%である」という仮説の下で「薬Aは無作為に抽出された患者20人中15人に効いた」というデータの値をほとんど意外ではないと考えられる。

## 1.4 S値

以下では、「無作為に選んだ患者 $n$ 人中 $k$ 人に薬Aが効いた」というデータの値に関する「薬Aが効く患者の割合は $p$ である」という仮説の(両側)P値を以下では  $P$  と書くことにします。

P値  $P$  の値が小ささは、「薬Aが効く患者の割合は $p$ である」という仮説を信じている人が「無作為に選んだ患者 $n$ 人中 $k$ 人に薬Aが効いた」というデータの値を意外だと感じる程度を表すのです。

意外さの程度は偏りが無いコイン投げに例えると分かり易くなります。

例えば、偏りのないコイン投げで4回続けて表が出る確率は  $6.25\% \approx 6\%$  である。だから、 $P \approx 6\%$  であることは、

「薬Aが効く患者の割合は $p$ である」と信じている人が「無作為に選んだ患者 $n$ 人中 $k$ 人に薬Aが効いた」というデータの値について感じる意外さ

が

偏りが無いコイン投げで 4 回続けて表が出たのと同じ程度の意外さ

だとみなされることを意味していると考えられます。

同様に  $P \approx 3\%$  であることは、上と同じ意味での意外さが

偏りが無いコイン投げで 5 回続けて表が出たのと同じ程度の意外さ

だとみなされることを意味していると考えられます。

便宜的にコイン投げの回数として非整数も許されることにすれば、 $P \approx 5\%$  であることは、上と同じ意味での意外さが

偏りが無いコイン投げで 4.3 回続けて表が出たのと同じ程度の意外さ

だとみなされることを意味していると考えられます。実際、

$$(1/2)^{4.3} \approx 0.05.$$

より正確には  $(1/2)^{4.3} \approx 0.050765774772264724$ .

P値  $P$  に対して以上のようにして決まる偏りが無いコイン投げで連続して表が出る回数を  $S$  は次のように得られます:

$$(1/2)^S = P \quad \text{すなわち} \quad S = -\log_2 P.$$

この  $S$  をP値  $P$  に対応する **S値 (意外度, surprisal value)** と呼ばれています。

S値は曖昧な「P値の小ささ」=「意外さ」という言い方で語られる概念の定量化だと考えられます。つまり、S値が大きさは「意外さ」の大きさそのものだと解釈されます。

P値とS値の対応は以下の通り。

P値	S値(意外度)
100%	0
50%	1
25%	2
約12%	約3
約6%	約4
約3%	約5
約1.5%	約6
約0.8%	約7
約0.4%	約8
約0.2%	約9
約0.1%	約10

情報理論的にS値の単位は **ビット** だと考えられます。

P値の違いの大きさの違いは、対応するS値の側で何ビット違うかを見ると誤解し難くなります。

例えば、6% と 3% のP値はそれぞれ約 4 と 5 のS値に対応し、1ビットの違いしかありません。伝統的にP値については 5% の閾値が採用されていて、 $P < 5\%$  であるか否かによって正反対の結論を出す悪しき慣習があるのですが、6% と 3% のP値はS値で見たら1ビットの違いしかないので、 $P < 5\%$  であるか否かによって正反対の結論を出すことの不合理さもS値を理解すれば分かります。

**広く通用するP値に関する閾値は存在しない:**  $P < 5\%$  であるか否かによる「有意差あり/なし」の二分法に強く執着している人達は、P値に閾値をどうしても設けたいと感じるかもしれません。しかし、仮に閾値の設定が有効な道具として機能するとしても、ケースバイケースで適切な閾値は異なると考えられます。検定仮説が諸々の理由で非常に確からしい場合にはその否定のためのP値の閾値は非常に小さくすることが妥当だと考えられるし、逆に検定仮説が非常に疑わしい場合にはその否定のためのP値の閾値は10%のような大きめの値にしても十分に合理的になることがあり得ます。さらに統計分析の結果を用いた意思判断においても、場合ごとに許容できるリスクの大きさと種類は異なります。そういう諸々の事情を全部無視して慣用的な1つの閾値で押し通そうとすることは科学的に不合理な行為だと考えられます。

## 1.5 相性の良さ/悪さという解釈の仕方

「無作為に選んだ患者  $n$  人中  $k$  人に薬Aが効いた」というデータの値に関する「薬Aが効く患者の割合は  $p$  である」という仮説の(両側)P値  $P$  の小ささとそれに対応するS値  $S = -\log_2 P$  は

「薬Aが効く患者の割合は  $p$  である」という仮説を信じている人が「無作為に選んだ患者  $n$  人中  $k$  人に薬Aが効いた」というデータの値に感じる意外度

という意味を持っているのです。これと同じ意味で、P値の小ささ(=S値の大きさ)を

二項分布モデルとそのパラメータに関する「薬Aが効く患者の割合は  $p$  である」という検定仮説と「無作為に選んだ患者  $n$  人中  $k$  人に薬Aが効いた」というデータの値の **相性の悪さ (incompatibility)** の尺度

と解釈することにします。このとき、P値の大きさ(=S値の小ささ)は

二項分布モデルとそのパラメータに関する「薬Aが効く患者の割合は  $p$  である」という検定仮説と「無作為に選んだ患者  $n$  人中  $k$  人に薬Aが効いた」というデータの値の **相性の良さ (compatibility)** の尺度

と解釈されます。以上の言い方も便利なのでこれから使って行くことにします。

## 1.6 P値関数とS値関数

「無作為に選んだ患者  $n$  人中  $k$  人に薬Aが効いた」というデータの値に関する「薬Aが効く患者の割合は  $p$  である」という仮説の(両側)P値  $P$  は  $k, n, p$  の関数なので、それを



$\text{pvalue}(k|n, p) = (k, n, p \text{ に関する両側P値})$

と書くことにします。

データの値  $k, n$  が与えられたとき、P値  $\text{pvalue}(k|n, p)$  は  $p$  だけの関数になる。そのとき  $p$  にP値  $\text{pvalue}(k|n, p)$  を対応させる関数を **P値関数 (P-value function)** と呼ばれています。

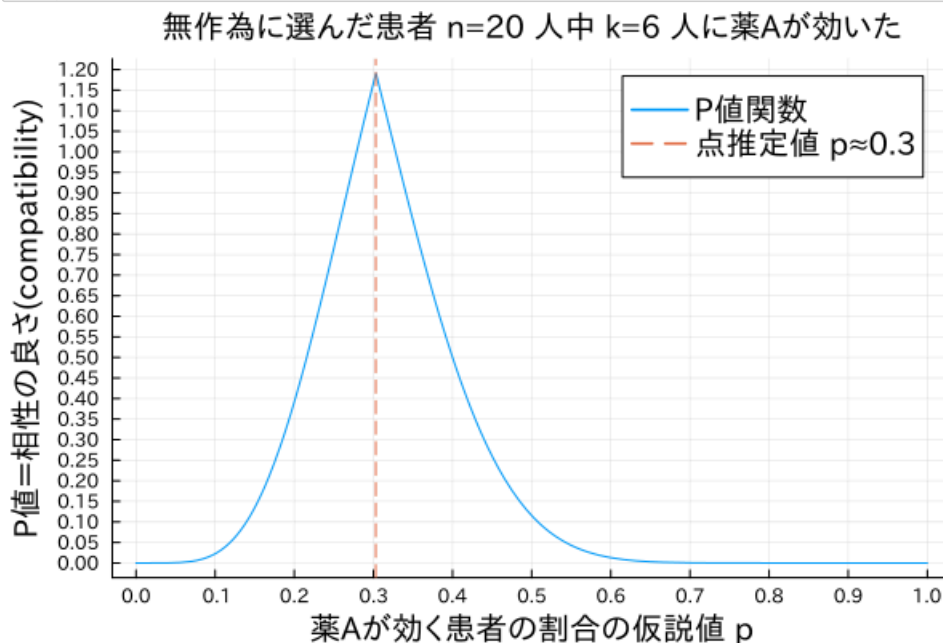
例題1.1の状況では、検定仮説が「薬Aが効く患者の割合は  $p = 0.6$  である」に固定されていました。パラメータ  $p$  の仮説値を固定せずに動かしながらP値を計算して、P値関数のグラフを描けば、パラメータ  $p$  のすべての仮説値と「無作為に選んだ患者  $n$  人中  $k$  人に薬Aが効いた」というデータの値の相性の良さ(compatibility)の様子が分かることになります。

しかし、P値関数のグラフを得るためには、コンピュータでグラフを描くスキルが必要になります。以下では、この解説の筆者が

次のセルのコードで描かれたのは「無作為に選んだ患者  $n = 20$  人中  $k = 6$  人に薬Aが効いた」というデータに対応するP値関数のグラフです。

```
In [9]: 1 n, k = 20, 6
2 plot(p → pvalue_central(Binomial(n, p), k), 0, 1; label="P値関数")
3 plot!(xtick=0:0.1:1, ytick=0:0.05:1.5, yguide="P値 = 相性の良さ(compatibility)")
4 title!("無作為に選んだ患者 n=$n 人中 k=$k 人に薬Aが効いた")
5 plot!(xguide="薬Aが効く患者の割合の仮説値 p")
6 vline!([0.303]; label="点推定値 p≈0.3", ls=:dash)
```

Out[9]:



このP値の大きさは、薬Aが効く患者の割合のすべての仮説値  $p$  と「無作為に選んだ患者  $n = 20$  人中  $k = 6$  人に薬Aが効いた」というデータの値の相性の良さが分かります。

例えば、このP値関数の最大値は  $p \approx 0.3$  で与えられるので、「無作為に選んだ患者  $n = 20$  人中  $k = 6$  人に薬Aが効いた」というデータの値と最も相性が良い検定仮説は「薬Aが効く患者の割合は  $p \approx 0.3$  である」であることが分かります。

このようにP値関数を最大化するパラメータの値  $p \approx 0.3$  は(P値関数から得られる)パラメータ  $p$  の **点推定値 (point estimate)** と呼ばれます。

P値関数が最大になる点推定値  $p \approx 0.3$  から  $p$  が離れれば離れるほどP値は小さくなるので、そのような  $p$  の仮説値とデータの値の相性は悪くなります。

例えば、「薬Aが効く患者の割合は  $p = 0.05$  である」や「薬Aが効く患者の割合は  $p = 0.65$  である」という検定仮説は、「無作為に選んだ患者  $n = 20$  人中  $k = 6$  人に薬Aが効いた」というデータの値と相性がかなり悪いことが上のグラフから分かります。

P値関数のグラフの欠点はその値が 0 に近い所の様子が見難いことです。

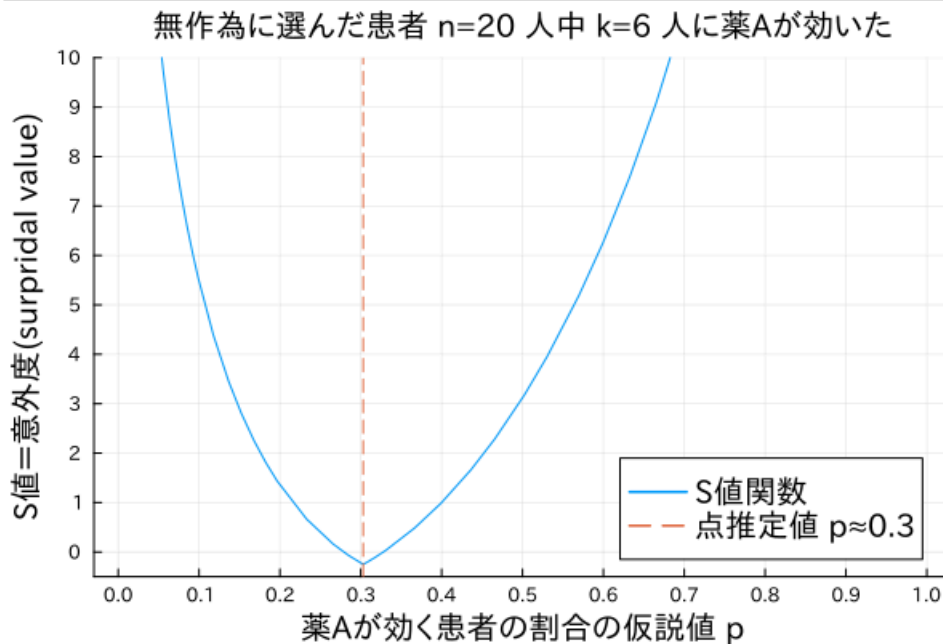
P値関数の代わりに同様に定義されたS値関数

$$p \mapsto -\log_2 \text{pvalue}(k|n, p)$$

のグラフを描くとP値が 0 に近い所の様子もよく見えるようになります。

```
In [10]: 1 n, k = 20, 6
2 plot(p → -log(2, pvalue_central(Binomial(n, p), k)), 0, 1; label="S値関数")
3 plot!(ylim=(-0.5, 10), legend=:bottomright)
4 plot!(xtick=0:0.1:1, ytick=0:10, yguide="S値 = 意外度(surpridal value)")
5 title!("無作為に選んだ患者 n=$n 人中 k=$k 人に薬Aが効いた")
6 plot!(xguide="薬Aが効く患者の割合の仮説値 p")
7 vline!([0.303]; label="点推定値 p≈0.3", ls=:dash)
```

Out[10]:



意外度が大きいことと相性が悪いことは同じ意味になり、S値 = 意外度で表される意外度は偏りのないコイン投げで何回続けて表が出たかを意味しているのです。

仮に偏りのないコイン投げで5回以上連続して表が出たときに「意外度が非常に大きい」 = 「相性が非常に悪い」ということにすると、 $p$  が 0.107 以下の場合と 0.564 以上の場合に「意外度が非常に大きい」 = 「相性が非常に悪い」ということになります。(このとき、区間 [0.107, 0.564] は 97% 信頼区間 = 相性区間になっている。信頼区間 = 相性区間については次節を参照せよ。)

ただし、ここで考えているのは「無作為に選んだ患者  $n = 20$  人中  $k = 6$  人に薬Aが効いた」というデータの値と「薬Aが効く患者の割合は  $p$  である」という仮説の相性の悪さです。

```
In [11]: 1 @show α = 0.5^5
2 @show 1 - α
3 @show confint_clopper_pearson(n, k; α);

α = 0.5 ^ 5 = 0.03125
1 - α = 0.96875
confint_clopper_pearson(n, k; α) = [0.10726583473679588, 0.564099648860263]
```

## 1.7 信頼区間 = 相性区間

前節でP値関数から点推定値がP値関数を最大化するパラメータ  $p$  の値として得られることを説明しました。

パラメータ  $p$  の区間推定もP値関数から得られます。

そのためにはP値に関する閾値  $\alpha$  を用意して、 $pvalue(k|n, p) \geq \alpha$  が否かで、パラメータ  $p$  の値はデータの値  $n, k$  と相性が良い/悪いとみなすことにします。

伝統的な用語法で閾値  $\alpha$  は有意水準と呼ばれ、 $1 - \alpha$  は信頼水準(信頼度、信頼係数)と呼ばれています。しかし、それらの伝統的用語法はP値がモデルとパラメータの値に関する仮説の組み合わせとデータの値の相性の良さ(compatibility)の尺度であるという見方と整合的ではないので、 $\alpha$  を **相性水準 (compatibility level)** と呼ぶことにします。

相性水準と呼ぶことにした閾値  $\alpha$  によってデータの値と相性が良いとみなされるパラメータ  $p$  の値全体の集合をパラメータ  $p$  の **相性区間 (compatibility interval)** と呼ぶことにします。伝統的な用語法でこれは信頼水準  $1 - \alpha$  の信頼区間と呼ばれるものと同じものになります。より詳しく述べると、我々が現在扱っているP値関数から得られる相性区間はClopper-Pearsonの信頼区間と同じものになります。

**Clopper-Pearsonの信頼区間について:** Clopper-Pearsonの信頼区間にはコンピュータでの効率的な計算に適した明示公式が存在し、その公式を使って実装すると極めて簡単にかつ効率的に計算できます。

ただし、その事実を理解するためには二項分布とベータ分布の累積分布関数間の関係やベータ分布の累積分布関数が正規化不完全ベータ関数と呼ばれる有名な特殊関数にちょうどっており、その逆関数の分位点関数もコンピュータで効率的に計算できることなどを学ぶ必要があります。数学的な事柄にも興味がある人は勉強してみると数学とコンピュータでの数値計算について沢山の教養が身に付くことでしょう。

この解説ではその点に深入りせずに、次の非常にシンプルなコードでClopper-Pearsonの信頼区間を計算できることを注意するとどめておきます。

```
using Distributions
function confint_clopper_pearson(n, k;  $\alpha$  = 0.05)
    p_L = k > 0 ? quantile(Beta(k, n-k+1),  $\alpha/2$ ) : zero( $\alpha$ )
    p_U = k < n ? quantile(Beta(k+1, n-k), 1- $\alpha/2$ ) : one( $\alpha$ )
    [p_L, p_U]
```

このコードを用いて、「無作為に選んだ患者 n=20 人中 k=6 人に薬Aが効いた」というデータの値が得られたときの、相性水準 5% の相性区間(=信頼水準 95% のClopper-Pearson信頼区間)を求めて見ましょう。

```
In [12]: 1 @show confint_clopper_pearson(20, 6;  $\alpha$  = 0.05);

confint_clopper_pearson(20, 6;  $\alpha$  = 0.05) = [0.11893159040572766, 0.5427891822762891]
```

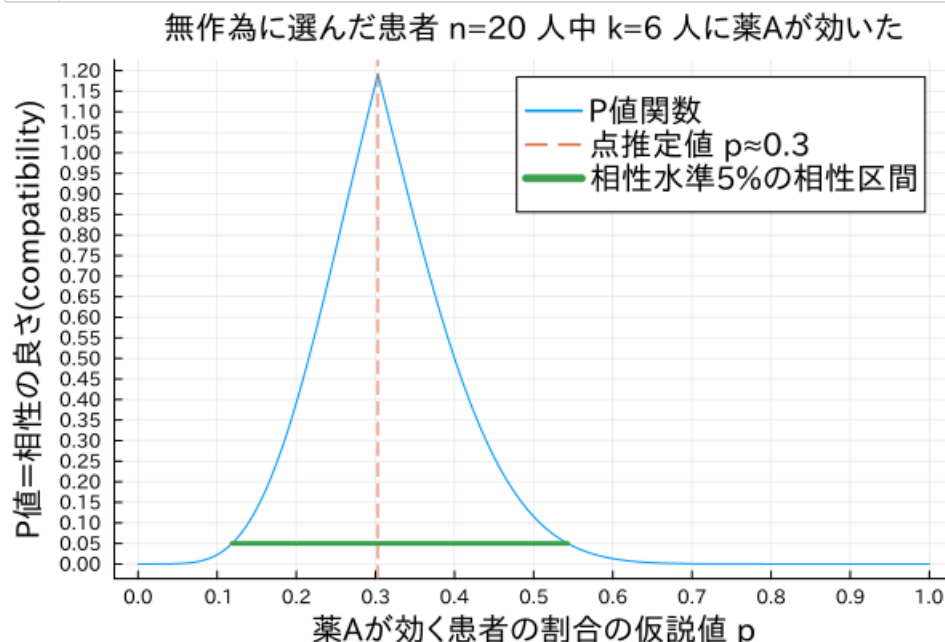
「無作為に選んだ患者 n=20 人中 k=6 人に薬Aが効いた」というデータの値が得られたときの、相性水準 5% の相性区間(=信頼水準 95% のClopper-Pearson信頼区間)はおおよそ [0.119, 0.543] になることが分かりました。

相性水準 5% の相性区間は、P値に関する 5% 以上という条件によって、「無作為に選んだ患者 n=20 人中 k=6 人に薬Aが効いた」というデータの値と相性が良いとみなされた「薬Aが効く患者の割合の仮説値  $p$ 」全体の集合になっています。

相性水準 5% の相性区間はP値関数のグラフを高さ 5% で切断する線分に等しい。次のグラフを参照せよ。

```
In [13]: 1 n, k,  $\alpha$  = 20, 6, 0.05
2 ci = confint_clopper_pearson(n, k;  $\alpha$ );
3 plot(p → pvalue_central(Binomial(n, p), k), 0, 1; label="P値関数")
4 plot!(xtick=0:0.1:1, ytick=0:0.05:1.5, yguide="P値 = 相性の良さ (compatibility)")
5 title!("無作為に選んだ患者 n=$n 人中 k=$k 人に薬Aが効いた")
6 plot!(xguide="薬Aが効く患者の割合の仮説値 p")
7 vline!([0.303]; label="点推定値 p≈0.3", ls=:dash)
8 plot!(ci, fill( $\alpha$ , 2); label="相性水準5%の相性区間", lw=3)
```

Out[13]:

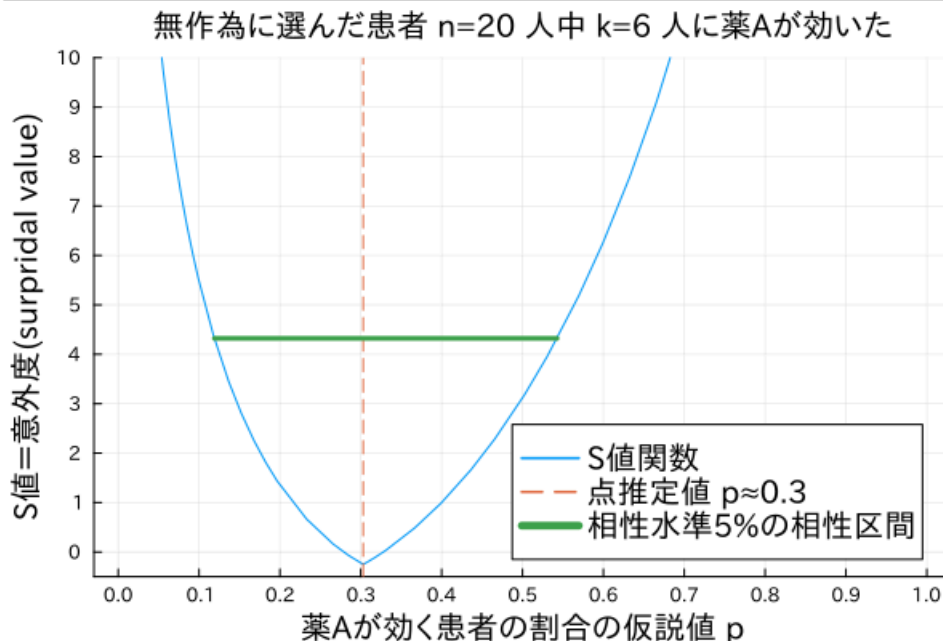


このように、P値(より正確にはP値関数)の概念から、点推定値と区間推定の両方が得られる。

S値のグラフにも相性区間を描きこむと次のようになる。相性水準 5% の相性区間は、S値関数のグラフを高さ  $-\log_2 0.05 \approx 4.3$  で切断する線分になる。

```
In [14]: 1 n, k = 20, 6
2 plot(p -> -log(2, pvalue_central(Binomial(n, p), k)), 0, 1; label="S値関数")
3 plot!(ylim=(-0.5, 10), legend=:bottomright)
4 plot!(xtick=0:0.1:1, ytick=0:10, yguide="S値 = 意外度 (surpridal value)")
5 title!("無作為に選んだ患者 n=$n 人中 k=$k 人に薬Aが効いた")
6 plot!(xguide="薬Aが効く患者の割合の仮説値 p")
7 vline!([0.303]; label="点推定値 p≈0.3", ls=:dash)
8 plot!(ci, fill(-log(2, α), 2); label="相性水準5%の相性区間", lw=3)
```

Out[14]:



## 1.8 報告の仕方について

P値関数やS値関数のグラフは、各検定仮説のP値の情報だけではなく、点推定値と区間推定値の情報も含まれます。だから、原理的にはP値関数のグラフを報告に載せれば特定の検定仮説のP値や点推定値や区間推定値を報告する必要はないと考えられます。

しかし、報告の読者の便のためには、重要だと考えられる検定仮説達のP値と点推定値と区間推定値も報告に載せた方が親切でしょう。

統計学ユーザー界には「有意差あり/なし」と報告したがる伝統的な悪習があるのですが、それを真似する必要はありません。英語の報告では *significant* の語の使用を控えことは現代的な良いスタイルだということになります。P値の具体的な値を報告していれば、読者の各々が自分の目的に適切な閾値を設定して二分的な判断を自分でできます。

「有意差あり( $P < 0.05$ )」と「有意差なし( $P \geq 0.05$ )」のように報告には書くべきではありません。報告を書く側はP値以外のあらゆる情報を加味して暫定的な結論を述べるべきです。

統計学ユーザー界には「差はゼロ」や「効果はゼロ」を意味する特殊な検定仮説(それを **ゼロ仮説** と呼ぶことにする)のP値のみを報告する悪しき伝統があります。その悪しき伝統は上で示したようにP値が推定の道具としても使えることを見逃す原因になっています。そういう悪習にも従う必要はありません。

例えば、「効果はゼロ」を意味するゼロ仮説のP値だけではなく、事前に予想していた効果の大きさやそれまでの定説の値  $a$  について「効果は  $a$  である」という検定仮説のP値も計算して報告するのは良いスタイルです。

## 2 二項検定の複数のP値のグラフ

このノートブックでは上の方で以下の4種類のP値関数を定義している。

- Clopper-Pearsonの信頼区間を与える片側確率の2倍版のP値関数 (central法)
- Sterneの信頼区間を与えるP値関数 (minlike法)
- Wilsonの信頼区間を与える二項分布の正規分布を使うP値関数 (score法)
- Waldの信頼区間を与えるP値関数 (Wald法)

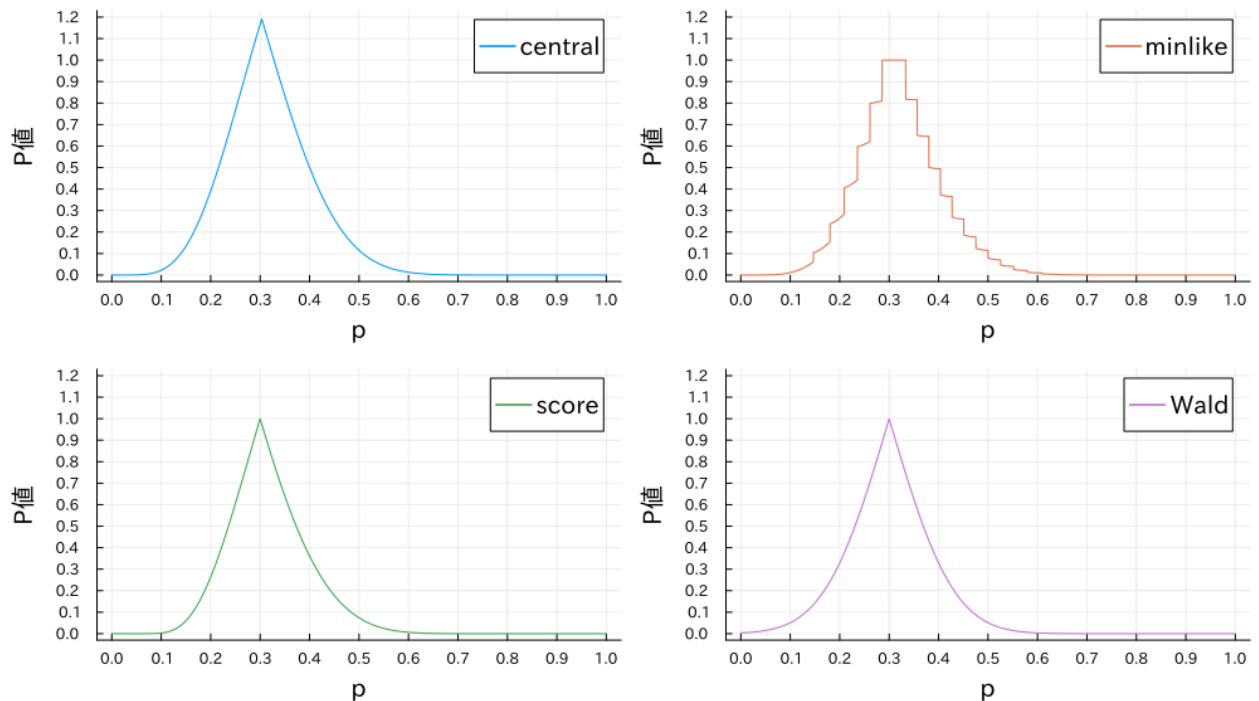
「無作為に選んだ患者 n=20 人中 k=6 人に薬Aが効いた」というデータの値に関するこれらのグラフを得外てみよう。

```

In [15]: 1 n, k = 20, 6
2 P1 = plot(p → pvalue_central(Binomial(n, p), k); label="central", c=1)
3 P2 = plot(p → pvalue_minlike(Binomial(n, p), k); label="minlike", c=2)
4 P3 = plot(p → pvalue_score(Binomial(n, p), k); label="score", c=3)
5 P4 = plot(p → pvalue_wald(Binomial(n, p), k); label="Wald", c=4)
6 plot(P1, P2, P3, P4; size=(1000, 600), layout=(2, 2))
7 plot!(xtick=0:0.1:1, ytick=0:0.1:1.2, ylim=(-0.03, 1.23))
8 plot!(xguide="p", yguide="P値")
9 plot!(plot_title="「無作為に選んだ患者 n=$n 人中 k=$k 人に薬Aが効いた」というデータの値のP
10 plot!(leftmargin=4Plots.mm, bottommargin=4Plots.mm)

```

Out[15]: 「無作為に選んだ患者  $n=20$  人中  $k=6$  人に薬Aが効いた」というデータの値のP値関数達



$n = 20, k = 6$  のような小さなデータでは上のグラフのよう4種のP値関数がかなり異なります。

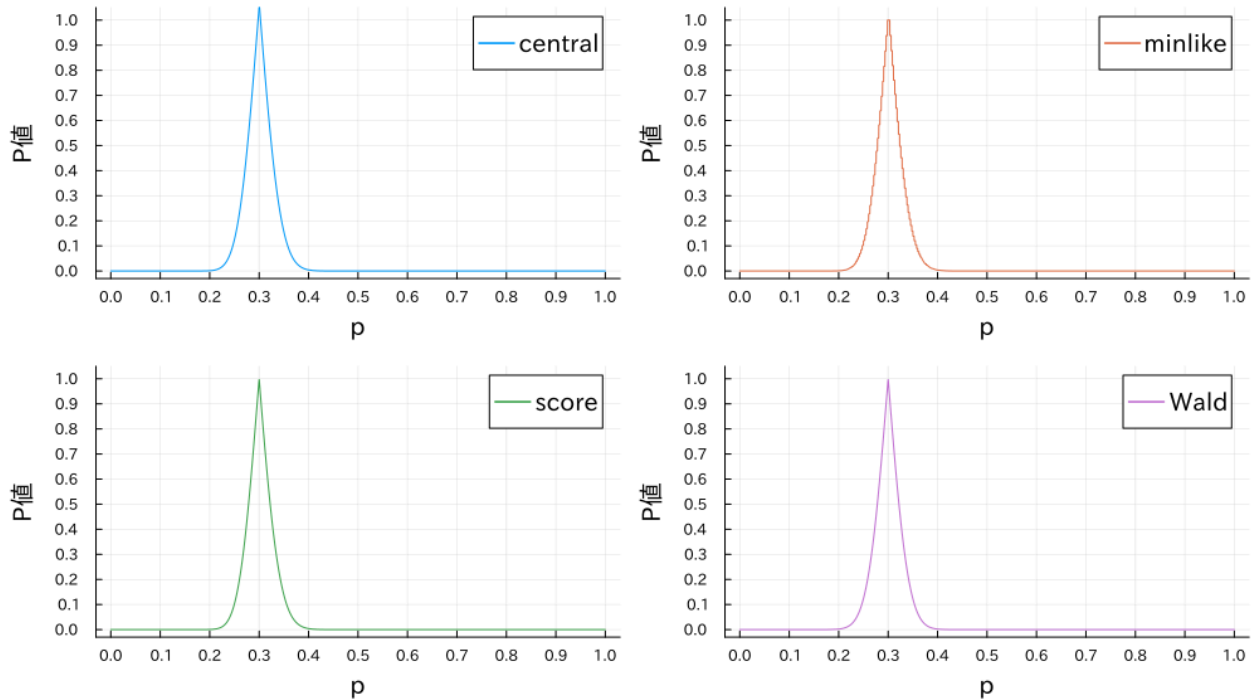
しかし、 $n = 200, k = 60$  の場合には以下のように4種のP値関数のグラフの違いは目立たなくなります。

```

In [16]: 1 n, k = 200, 60
2 P1 = plot(p → pvalue_central(Binomial(n, p), k); label="central", c=1)
3 P2 = plot(p → pvalue_minlike(Binomial(n, p), k); label="minlike", c=2)
4 P3 = plot(p → pvalue_score(Binomial(n, p), k); label="score", c=3)
5 P4 = plot(p → pvalue_wald(Binomial(n, p), k); label="Wald", c=4)
6 plot(P1, P2, P3, P4; size=(1000, 600), layout=(2, 2))
7 plot!(xtick=0:0.1:1, ytick=0:0.1:1.5, ylim=(-0.03, 1.05))
8 plot!(xguide="p", yguide="P値")
9 plot!(plot_title="「無作為に選んだ患者 n=$n 人中 k=$k 人に薬Aが効いた」というデータの値のP
10 plot!(leftmargin=4Plots.mm, bottommargin=4Plots.mm)

```

Out[16]: 「無作為に選んだ患者 n=200 人中 k=60 人に薬Aが効いた」というデータの値のP値関数達



In [ ]: 1