2009年度後期　データ解析序説　第１３回

分布についての仮説を検証するー仮説検定

前回説明した区間推定では，母平均について，１つの数で推定せずにある程度の余裕をもって推定しました．しかし，実際には母平均などのパラメータについて「ある値であるかないか」「ある値より大きい（小さい）か否か」などの**判断**が必要なことがしばしばあります．検定（仮説検定）とは，統計的推測の考え方を使って，母集団に対する判断を行うものです．その基本的考え方は，

めったに起きないことは，いま現実には起きていない
（いま起きていることは，ありふれたできごとにちがいない）

というものです．

いま，あるくじが50%の確率で当たるとします．このくじを10本ひくと，全部はずれでした．このとき，「50%の確率で当たるくじが10本続けてはずれることなど，めったにあるはずがない」と考えるのは自然なことです．そうすると，最初の仮定は間違いで「このくじの当選確率は50%よりも小さい」と考えるほうが妥当です．このような素朴な考え方を統計学を用いて述べたのが仮説検定です．

上のくじびきの問題に答えるには，「２項分布モデル」という確率分布モデルを用います．今回は，２項分布モデルにもとづいて検定の考え方を説明します．

２項分布と検定

では，冒頭でふれた状況をもう一度考えてみましょう．

> あるくじは，確率50%で当たるとします．このくじを10本ひくと，全部はずれでした．このくじは，本当に確率50%で当たるのでしょうか．当たる確率はもっと小さいのではないでしょうか？

このときは，「確率50%で当たるはずのくじが10本続けてはずれることなど，めったにあるはずがない」と考えるのは自然なことです．そうすると，最初の「確率50%で当たる」は間違いで，「このくじの当選確率は50%よりも小さい」と考えるほうが妥当です．では，次の状況はどうでしょうか．

> あるくじは，確率50%で当たるとします．このくじを20本ひくと，6本当たって14本はずれでした．このくじは，本当に確率50%で当たるのでしょうか．当たる確率はもっと小さいのではないでしょうか？

今度は少々微妙な状況です．前の状況と同じように「当たる確率は50%よりも小さい」のかもしれませんが，もしかしたら「当たる確率は本当に50%」で，当たりが少なかったのは，単に運が悪かったからなのかもしれません．

どちらが正しいのかは，わかりません．しかし，当選確率が50%か，それともそれより小さいのか，どちらの言い分がよりもっともかを，統計学を使って判断をするのが**仮説検定**（あるいは単に**検定**）という手法です．

では，仮に「当選確率は 50%である」ということを認めるとしましょう．さきほど，「当選確率 50%のとき，10 本続けてはずれることはめったにない」ということを述べましたが，では「当選確率が 50%のとき，20 本ひいて 6 本しか当たらない」ことは，どのくらい「めったにない」ことなのでしょうか．それを計算することを考えてみましょう．

計算しなければならない確率は

ここで，「そんなことはめったにない」の「そんなこと」とは，正確にはどういう意味でしょうか．それは，「20 本ひいて**ちょうど 6 本だけ**当たる」ことでは**ありません．**「20 本ひいて，5 本でも 7 本でもなく，ちょうど 6 本当たる」ことは，たしかにめったにないでしょう．しかし，われわれはそれを問題にしているのではありません．

仮に，20 本のくじをひいて 1 本も当たりが出なければ，「確率 50%で当たる」というのはきわめて疑わしいでしょう．それは，「確率 50%で当たる」はずなのに 20 本中 1 本も当たらない，という確率はきわめて小さいからです．確率がより小さいはずのできごとが起きると，そもそもの「確率 20%で当たる」という前提に対する疑いは，より強くなります．

ということは，「20 本ひいて 6 本**しか**当たらない」ことが「めったにない」，という言い回しは，言外に，5 本しか当たらないことも，4 本しか当たらないことも，. . .，1 本も当たらないことも，当然すべて「めったにない」と考えている，ということを含んでいるはずです．5 本しか当たらないことが不満な人は，4 本しか当たらないことも，当然 1 本も当たらないことも，言うまでもなく不満なのです．

つまり，計算しなければならないのは，「不満をもつような現象」が起きる確率，すなわち「20 本ひいたとき，当たり本数が 6 本以下である」確率です．

では，この確率を求めてみましょう．

「当たり」または「はずれ」が出るくじびきは，

- 1 回のくじびきで，「当たり」または「はずれ」のどちらかが必ず出る．
- 各回のくじびきで，「当たり」「はずれ」それぞれの確率はつねに一定である．
- ある回のくじびきの結果は，他の回の結果には影響を及ぼさない．

と考えられます．これを，くじびきの例から離れて一般的に表すと，

- 「当たり・はずれ」，「成功・失敗」など，2 種類の結果のどちらかが必ず起きるような実験（あるいは観測，検査）があり，
- この実験を何度も繰り返すとき，各実験での「成功」「失敗」それぞれの確率はつねに一定で，
- 1 つの実験の結果が他の実験に影響を及ぼさない（各実験が**独立**である），

となります．このような仮定ができるような実験を**ベルヌーイ試行**といいます．

ベルヌーイ試行で起こりうる 2 種類の結果を「成功」，「失敗」で表し，成功する確率が p，失敗する確率が $1-p$ とします．このとき，「n 回のベルヌーイ試行を行うとき，成功する回数」は確率変数で，これを S とします．確率変数 S のしたがう確率分布モデルを **2 項分布**といい，$B(n, p)$ という記号であらわします．確率変数 S の期待値 $E(S)$ と分散 $V(S)$ は，それぞれ $np, np(1-p)$ となることが知られています[1]．

[1] 2008 年度後期「情報統計学」第 5 回を参照してください．講義のウェブページからリンクされています．

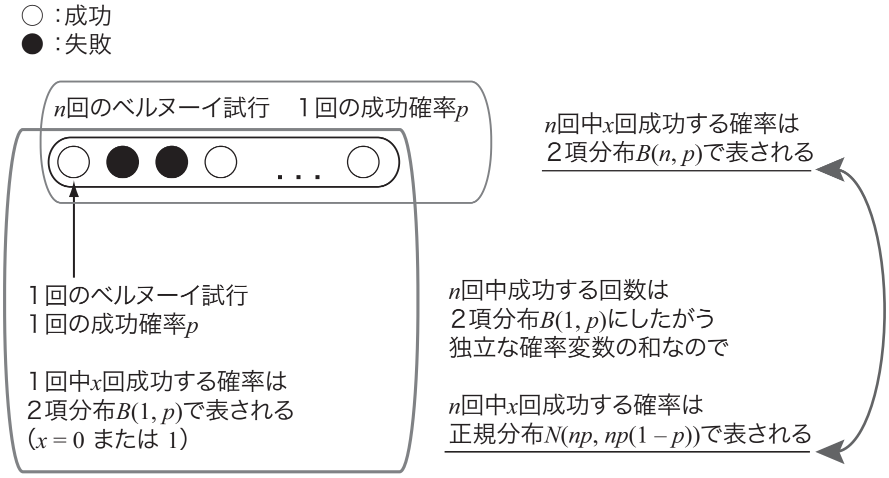

図 1: ド・モアブル＝ラプラスの定理

試行１回あたりの成功の確率が p で，それを n 回行うわけですから，「n 回試行したとき成功する回数の期待値」が np であることは直感的にわかります．また，$p = 1/2$ の時分散が最大になることは，「成功する回数」の予測が $p = 1/2$ の時一番しにくいことを示しています．

さて，２項分布に関する計算は，中心極限定理を使って２項分布を正規分布で近似することにより，正規分布の数表を使って簡単に計算をすることができます．これが**ド・モアブル＝ラプラスの定理**です．

「１回あたり確率 p で成功する，n 回のベルヌーイ試行」を別の視点から見てみましょう．ちょっと変な感じですが，「１回のベルヌーイ試行で，成功する回数」を考えて，X で表します．当然，X のとりうる値は０回または１回です．１回あたり確率 p で成功しますから，「１回のベルヌーイ試行で，成功する回数」X は２項分布 $B(1, p)$ にしたがいます．

一方，「１回あたり確率 p で成功する n 回のベルヌーイ試行で，成功する回数」を S とすると，S は２項分布 $B(n, p)$ にしたがう，すなわち，成功回数が x である確率は２項分布 $B(n, p)$ で計算できます．しかし，見方を変えれば，この S はさっきの X を n 個合計したものと考えることもできます．

n 個の X は互いに独立ですから，それらの合計である S は，n が大きいときは中心極限定理によって概ね正規分布にしたがいます．一方，S は２項分布 $B(n, p)$ にしたがうのですから，前回の説明で述べたように，S の期待値は np，分散は $np(1 - p)$ です．S の分布を，「概ね正規分布」とよぼうが，「２項分布」とよぼうが，同じ現象を別の名前でよんでいるだけですから，期待値や分散は同じです．したがって，n が大きいとき，２項分布 $B(n, p)$ は正規分布 $N(np, np(1 - p))$ で近似できることがわかります．

以上から，「当選確率 50%のくじを 20 本ひいて，当たり本数が６本以下である」確率は，$p = 0.5$, $n = 20$ としたとき，正規分布 $N(np, np(1 - p))$ にしたがう S が「$S \leq 6$」である確率となります．S は正規分布 $N(np, np(1 - p))$ にしたがうので，さらに

$$Z = \frac{S - np}{\sqrt{np(1 - p)}} \tag{1}$$

とおくと Z は標準正規分布 $N(0, 1)$ にしたがいます．また，$S = 6$ のとき

$$Z = \frac{6 - 50 \cdot 0.5}{\sqrt{50 \cdot 0.5 \cdot (1 - 0.5)}} = -1.79 \tag{2}$$

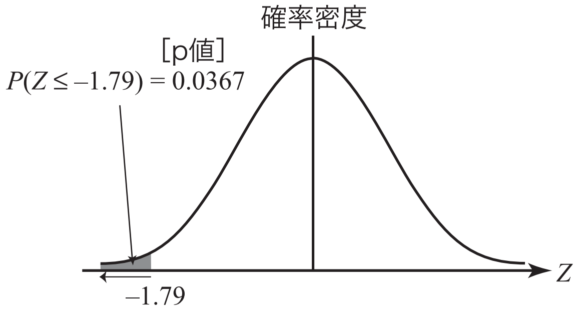

図 2: p 値

ですから，「$S \leq 6$」である確率は「$Z \leq -1.79$」となる確率となります．数表から，この確率は 0.0367 であることがわかります．

p 値と検定

以上のことから，「当選確率 50%としたとき，このくじを 20 本ひいて，当たり本数が 6 本以下である」確率は 0.0367 であることがわかりました．この値を p **値**といいます．

「当選確率 50%としたとき，このくじを 20 本ひいて，当たり本数が 6 本以下である」確率は 0.0367 です．では，「当選確率 50%としたとき，このくじを 20 本ひいて，当たり本数が 6 本以下である」ことは「めったにないこと」なのでしょうか？ それはなんとも言えません．「確率 0.0367 でしか起きない」のか，「起きる確率が 0.0367 もある」のか，それは考え方によるでしょう．

しかし，それでは話は続きません．そこで，統計学では「起きる確率がある値以下のときは，それは『めったにない』ことだ」と考えます．この値を**有意水準**といい，通常 5%(0.05) や 1%(0.01) が用いられます．

ここでは，有意水準を 5%としましょう．そうすると，0.0367 は 5%より小さいですから，「当選確率 50%としたとき，このくじを 20 本ひいて，当たり本数が 6 本以下である」ことは「めったにないこと」だ，と判断されます．そこで，

仮に，当選確率が 50%とする
→そうすると「20 本くじをひいて，当たり本数が 6 本以下である」という，有意水準よりも小さい確率でしかおきない，つまり「めったにないはず」のことが現に起きていることになる
→当選確率が 50%というのは間違いで，当選確率は本当はもっと小さい，と結論する．

という推論が行えます．この推論を**仮説検定**（あるいは単に**検定**）といいます．

最初の「当選確率が 50%とする」という仮定を**帰無仮説**といい，$H_0 : p = 0.5$ と表します．また，「当選確率が 50%というのは間違いと結論する」ことを，帰無仮説を**棄却する**といいます．さらに，帰無仮説が棄却された結果得られる「当選確率はもっと小さい」という結論を**対立仮説**といい，$H_1 : p < 0.5$ と表します．

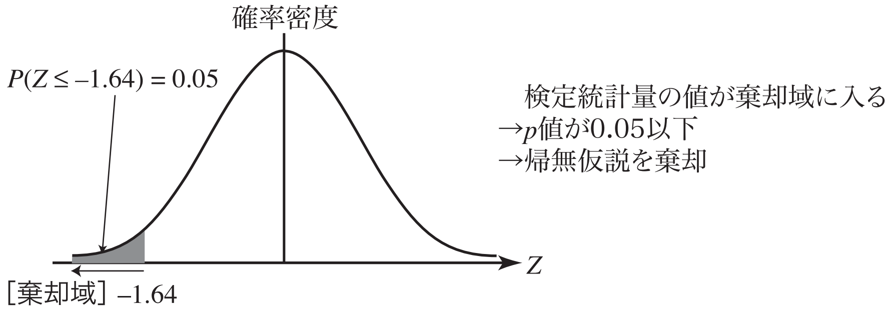

図 3: 棄却域

棄却域と検定統計量

上で，あたり本数 S が 6 以下である確率は，標準正規分布にしたがう確率変数 Z が -1.79 以下である確率と同じで，数表からこの確率，すなわち p 値は，0.0367 である，と述べました．

ところで，数表から「$Z \le -1.64$」となる確率が 5%です．ですから，Z の値が -1.64 以下であれば，p 値が 5%以下となり，このとき帰無仮説を棄却します．すなわち，いちいち p 値を計算しなくても，Z の値を求めてそれが -1.64 以下であれば，有意水準が 5%のとき帰無仮説は棄却されます．

この意味で，「$Z \le -1.64$」をこの問題における**棄却域**といい，棄却域を表すのに用いる確率変数（ここでは Z）を**検定統計量**といいます．これを用いると，上のように p 値を計算するかわりに，「当たり本数がちょうど 6 本のとき，検定統計量 Z の値が -1.79 である．一方，Z が -1.64 以下である確率は 5%である．すなわち，帰無仮説が正しいとすると $Z = -1.79$ となり，「$Z \le -1.64$」という確率 5%でしか起きないことが起きていることになるので，帰無仮説を棄却する」という言い方ができます．なお，上のように検定統計量が棄却域に入ることを「棄却域に落ちる」といいます．

棄却されないときは

ここまでに述べたように，検定では，帰無仮説は「内心では」棄却されることが期待されています．目論見通り棄却されると，「対立仮説を採択する」という結論が得られるわけです[2]．

では，今回のくじびきの例で，有意水準を 1%にしてみましょう．この場合，p 値が 1%以下ならば帰無仮説を棄却します．ところが，今回の例では p 値は 0.0367 ですから，1%よりも大きくなっています．したがって，目論見に反して，帰無仮説は棄却されません．

この場合，帰無仮説が棄却されなかったのは，「帰無仮説が正しいとき，今おきている現実得られる確率は**非常に小さいとまでは言えない**」です．したがって，**「帰無仮説が間違っているかどうかはわからない」「対立仮説が採択できるかどうかはわからない」**という結論を導かなくてはなりません．今回の例でいえば，帰無仮説が棄却されなかった場合は，「当選確率は 50%より小さいとは言えない」と答えなければなりません．つまり，「目論見ははずれた．当選確率は 50%より小さいとまで断言する自信はない」という結論になるのです．

注意しなければならないのは，あくまで，「帰無仮説が正しいとき，今おきている現実得られる確率は非常に小さいとまでは言えない」であって，「確率が大きい」のではない，ということです．したがって，帰無仮説が棄却されなかったときに，「帰無仮説が正しい」「対立仮説は間違っている」という結論

[2]帰無仮説を「無に帰す仮説」とよぶのはその意味です．

が得られるわけではありません．今回の例でも，「当選確率は 50%である」などと答えてはいけません．つまり，

帰無仮説を棄却しない　＝　~~帰無仮説を採択する~~
対立仮説を採択するべきかどうか断言できない

ということです．なお，「帰無仮説を棄却すべきなのに棄却しない」という誤りを**第２種の誤り**といいます．

有意水準の違い

ところで，最初のくじびきの例では，有意水準が 5%のときは「帰無仮説を棄却する＝ 50%の確率で当たるというのは間違い」と結論され，有意水準が 1%のときは「帰無仮説を棄却しない＝ 50%の確率で当たるというのは間違いとは言い切れない」という結論になりました．しかし，「20 本中 6 本しか当たらなかった」という現実は同じです．有意水準は勝手に決めたのに，こんなに違った結論になってもよいのでしょうか？

これについては，「検定とはそういうものだ」ということを，よく理解しなければなりません．有意水準は，検定をする人の「大胆さ・慎重さ」の程度を表しているのです．

有意水準が大きい (5%) ときは，今起きているような現実（6 本しか当たらない）が起きる確率が少々大きくても，それが 5%以下であれば「そんなことが起きるはずがない，帰無仮説は間違っている」と結論します．はっきり物をいう態度ではありますが，帰無仮説が正しい（くじの主催者が正直で，6 本しか当たらなかったのは偶然だった）ときでも「間違っている」と断言してしまう可能性があります．大胆ですが，勇み足も多い，というわけです．

有意水準が小さい (1%) ときは，今起きているような現実が起きる確率が，1%以下と相当小さくないと，「わずかでもそんなことが起きる可能性があるのだから，帰無仮説は間違っているとは言い切れない」となり，結論を出しません．慎重ですが，煮え切らない態度ということになります．

検定はどんな時にするものなのか

ところで，帰無仮説が正しい（くじの主催者が正直で，6 本しか当たらなかったのは単なる偶然だった）ときに帰無仮説を棄却してしまうという誤りを，**第１種の誤り**といいます．

つまり，

仮に帰無仮説が正しいとしても，有意水準 5%(1%) の仮説検定を何度も行うと，そのうち 5%(1%) は第１種の誤りを犯して棄却し，採択すべきでない対立仮説を採択してしまう

ことになります．

ですから，同じ現象についてなんどもデータを集めて，同じ帰無仮説について検定を繰り返し，たまに対立仮説が採択されても，直ちに「帰無仮説は間違っている」とはいえません．例えば，「血液型と性格に関係はない」という帰無仮説について何度もデータを集めて検定を行い，たまに「血液型と性格に関係がある」という結論が出ても，直ちに「やっぱり血液型と性格に関係がある」ということにはなり

ません．何度も検定を帰無仮説が間違っていない場合でも，たまに対立仮説が採択されるのは，むしろ自然なことです．血液型と性格の問題でいえば，「血液型と性格に関係がある」という結論がごくたまに出る程度であれば，「血液型と性格に関係があるとは，今のところ言えない」といっておくのが，科学的態度です．

では，検定の結論は結局何を言っているのでしょうか？ それは，今日の問題の例で言えば，

> **私は，くじの主催者が『確率 50%で当たる』と言っているのは間違いだ，と判断する．**
> **ただし，私は 100 回中 5 回はウソを言う（第 1 種の誤りを犯す）人間である．**
> **今回私がウソを言っているかどうか，それは誰にもわからない．**

と言っているのに等しいのです．

この程度のことしか言っていないのに，検定にはどういう意味があるのでしょうか？

それは，検定とは，少ない数のデータしか，しかも 1 度だけしか調べられないときに，「それだけのデータからでも十分な確信をもって述べられる疑いだけを述べる」方法ということなのです．何度も検定できるほどデータがあつまるのなら，検定を用いるのは不適切です．

くじびきの場合は，今くじをひいた数だけのデータしか集められませんから，検定を用いるのは妥当です．しかし，上で述べた血液型と性格の問題では，長い間に集められたたくさんのデータについて検定を用いるのは，検定の目的からはずれています．