２００３年度・総合科学部一日体験入学

身近な確率・身近な統計

確率とは「割合」である

確率とは，統計学を学習するうえでは，次のようなものと考えてください．

ある「できごと」が起きる確率とは，
これから何度もその「できごと」が起きる機会があるとき，
本当にその「できごと」が起きる回数の割合である．

次にその「できごと」が起きる確率とは，
長い目で見てはじめて言える「できごとの回数の割合」を，
次の１回の機会にあてはめて述べたものにすぎない．

例えば，くじ引きを十分多くの回数行なうとき，10回に3回の割合で当たりが出るとすれば，「あたりが出る確率」は 3 / 10であると考えます．しかし，次にくじを１回ひくとき，当たりが出るかどうかは何とも言えません．ただ，「これからもくじをひきつづけると，長い目で見れば10回に3回の割合で当たりが出るだろう」という数値で，次の１回の機会での当たりくじの「出やすさ」を表現しようというのが，確率の考え方です．

このような確率の考え方を，**頻度による確率の定義**といいます．つまり，確率とは「都合のいいできごとの割合」ですから，その値は0から1（0%〜100%）の範囲になります．

さいころの各目が出る確率はどれも1/6か？

高校までの教科書で確率を学ぶ時には，「さいころの各目が出る確率は，いずれも1/6である」ということを前提にしていたと思います．

しかし，頻度による確率の定義から考えれば，次にさいころをふったときにある目が出る確率は，十分に多くの回数さいころを振ってみなければわからないことになります．しかも，「十分に多くの回数」振らなければなりませんが，何回なら十分なのでしょうか？実は，何回振っても十分ではありません．上のくじ引きの例でも，1万回くじをひいて3000回当たりが出たとしても，それはあくまで「過去の実績」であって，その次に1万回くじをひいたらすべてはずれかもしれません．つまり，頻度による定義では，現実には確率を厳密に定めることはできないことになります．

では，なぜ「さいころの各目が出る確率は，いずれも1/6である」と言われているのでしょうか？それは，

１．各目が出るチャンスはみな同じ
２．ある機会に各目が出るチャンスは，他の機会の結果に影響されない

ということを皆が認めているからです．そこで「さいころには全部で６種類の目があって，いずれの目も常に同じ確率で出るから，各目が出る確率は1/6」ということになります．

高校までに習った確率の問題は，このような仮定を認めたうえで，確率すなわち「できごとの数の割合」の問題を，「（さいころの目の種類などの）できごとの種類の割合」の問題に置き換えたものです．このような確率の考え方を**ラプラスの定義**といいます．

しかし，このラプラスの定義も，よく考えるとおかしなところがあります．上で「１，２を認め

れば」と書きましたが，これが認められるかどうかは，さいころを十分な回数振ってみないとわかりません．これでは堂々めぐりです．つまり，確率の定義にはどのように考えてもあやしいところがあります．それであえてプリントの初めに．**長い目で見てはじめて言える「できごとの回数の割合」を，次の１回の機会にあてはめて述べたものにすぎない**と書いたのです．しかし，初歩の統計学では，その「あやしさ」を承知したうえで常識的に確率を理解しておけば十分です*).

問題１　ルーレットは，赤ばかり続いた後は黒が出やすいのか？

ルーレットでは，円盤に配置されている番号のうち，赤のグループに入るものと黒のグループに入るものが半々になっています．ラプラスの定義によれば，「赤と黒が出るチャンスは，おのおの毎回同じ」であることを認めるならば，１回ルーレットを回したとき，赤が出る確率も黒が出る確率も1/2ということになります**).

さて，仮に赤が10回続けて出たとしましょう．このとき，頻度による確率の定義から考えると，次は黒が確率1/2よりも出やすくないと，十分多くの回数ルーレットを回したときに赤と黒の回数の割合がそれぞれ1/2にならないような気がします．本当にそうでしょうか？

問題２　ファイナルアンサー？

A, B, C の箱があって，うちひとつだけに当たりが入っており，参加者がどれか１つを選ぶというゲームを行なっています．司会者は，どの箱に当たりが入っているかを知っています．

参加者はA の箱を選んだとします（まだ開いてはいません）．ここで，司会者は『はずれの箱をひとつ開いてあげましょう』と言って，C の箱を開けました．C の箱ははずれでした．司会者は参加者に向かって『今なら，A ではなく B を選ぶこともできます．A から B に変えますか？』と尋ねました．

当たる確率をあげるためには，選ぶ箱を A から B に変えるほうがよいでしょうか？

統計的推測とは

統計というと，データを集めて表やグラフに整理したり，平均を求めたり．．．というものを想像するのではないでしょうか．実は，それは統計学の入り口でしかありません．本当の統計学は，集めたデータをもとに，そのデータが現れる仕組みを推測したり，未来に現れるデータを予想する技術を研究する学問です．

現実世界では，

「日本男性の身長は？」「人によって違う」
「ある夫婦に次に生まれる子供は男か女か？」「生まれてみなければわからない」

という問題によく出会います．

これらの問題では，対象にしているデータが，時と場合によってばらばらになっています．人がこれを「ばらばら」と感じるのは，データが得られる仕組みを人間が完全に把握することができず，それを「神様がさいころをふって決めている」と考えているからです．このように，ある測定対象や現象から得られるデータがばらばらであることを**分布する**といい，このような分布したデータが現れる現象を**ランダム現象**といいます．統計学が扱うのは，ランダム現象によって生じた，分布しているデータです．

---

*) 現代の数学では，確率は現実の問題から離れて，集合を測る尺度（測度）のひとつとしてとらえられています．
**) 実際のルーレットには色のついていない「０」や「００」があるので，この条件を認めても赤あるいは黒が出る確率は1/2ではありません．

上の例では，「日本男性の身長」や「上の実験でできる水の量」や「次に生まれる子供の性別」は分布する，ということになります．しかし，上の問答のように「わからない」と言ってしまっては身も蓋もありません．そこで，せめて「どう分布しているか」，つまり「何cmくらいの人が多いか」とか「生まれる子供は男が多いか，女が多いか」を考えてみることにしましょう．

しかし，このような例では，すべてのデータを調べることは事実上不可能です．日本男性全員の身長を調べるには莫大な費用がかかりますし，まだ生まれていない子供の性別はわかりません．

そこで，「日本男性の一部の身長」を調べたデータや，「過去１年間に生まれた子供の性別」を調べて，日本男性の全体の身長の分布や，これから生まれてくる子供の性別の分布を推測することにします．これが，**統計的推測**です．

「くじびき」の原理

統計的推測の原理は，実は「くじびき」です．

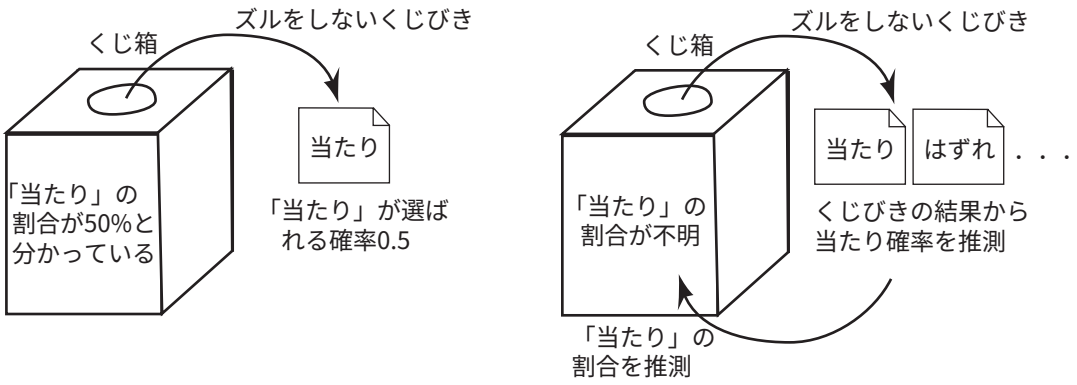
くじ箱の中に入っている当たりくじの割合が50%だとしましょう．このとき，このくじを１回ひくと，当たりをひく確率は50%です．これを逆に考えると，くじ箱の中の当たりくじの割合がわからないとするとき，当たりの確率が50%ならば，くじ箱の中も当たりくじの割合は50%です．

これと同様に，仮に「日本男性全体のうち，身長が160〜165cmの人の割合が10%」とすると，日本男性から一部を取り出すとき，160〜165cmの人が選ばれる確率は10%です．ですから，「日本男性全体のうち，身長が160〜165cmの人の割合」を調べるには，「日本男性から一部を取り出すとき，160〜165cmの人が選ばれる確率」を調べればよいわけです．

これが統計的推測の原理ですが，以下のような問題があります．

１．本当に「くじ箱の中の当たりくじの割合＝当たりの確率」か？

この関係が当たり前であるためには，このくじが「ズルをせずに」引かれる必要があります．つまり，箱の中のどのくじも平等に選ばれるチャンスがあり，しかもあるくじ引きの結果が他の回に影響しない，というものでなければなりません．ですから，日本男性の一部を調べる場合でも，これと同様に「えこひいきをせずに」選ぶ必要があります．このような選び方を**無作為抽出**といいます．



これは当たり前の関係　　これが統計的推測

図１．統計的推測の原理は「くじびき」

２．くじを引いたからといって，当たりの確率がわかるのか？

くじを１本だけ引いて結果をみても，そのくじの当たり確率はわかりません．しかし，くじを何本か引いて当たりくじの割合を見れば，当たり確率が多少は想像できます．引くくじの数が多ければ多いほど，当たり確率ははっきりわかってきます．日本男性の一部を調べる場合でも，たくさん調べれば調べるほど，「160〜165cmの人が選ばれる確率」はだんだんはっきりわかってきます．このことを**大数の法則**といいます．

３．当たり確率は本当に一定か？

いまはくじ箱の中の当たりの割合は一定だと考えていますが，もしかしたら途中で当たりの割合を増減する操作をしているかもしれません．このときは，何本かくじを引いてみても，くじ箱の中の当たりの割合はよくわからなくなってしまいます．

しかし，くじを引くほうにとっては「くじ箱の中の当たりの割合は一定」かどうかはわかりませんから，「一定である」と仮定して話を進めざるをえません．同様に，これから生まれる子供の性別の割合が一定かどうかは誰にもわかりませんが，とりあえず「一定である」とすると，これから生まれる子供の性別の割合を推測できます．このような，「概ねどんな分布であるか」についての仮定を**確率分布モデル**といいます．

問題３　おじさんはウソツキか？

いま，「くじ箱に入っているくじのうち，半分があたり」と店のおじさんが言っているくじがあるとしましょう．ところが，あなたがこのくじを10 回引いても，1 回もあたりませんでした．

おじさんは「運が悪かったねー」と言っていますが，あなたはどうも納得がいきません．「おじさんの言ってる『半分があたり』なんてうそじゃないの？」と思います．さて，おじさんかあなたか，どちらが正しいでしょうか？どうやれば，納得がいくようにこの疑問に答えることができるでしょうか？

浅野の講義・ウェブサイト

浅野の公的ウェブサイト（`http://kuva.mis.hiroshima-u.ac.jp/~asano/`）に，講義に関するページ（`http://kuva.mis.hiroshima-u.ac.jp/~asano/Kougi/`）があります．浅野の各種講義の講義録や，統計関係のリンク集，講義中の雑談の記録が載せてあります．よろしければご覧いただければ幸いです．

```
  ∧∧
 ≡・・≡
  (   )〜
```

本日は，「一日体験入学」にお越しいただきありがとうございました．
これからも広島大学総合科学部をよろしくお願いします．

```
  ∧◆∧
Ξ o−o Ξ
  (     )〜
```

浅野のメールアドレスは，`asano@mis.hiroshima-u.ac.jp`です．
ご質問等ありましたらご遠慮なくどうぞ．

解答編

問題１

赤が10回続けて出たとしても，その次の回で赤・黒が出る確率は各々1/2です．なぜかというと，「問題文にそう書いてあるから」です．

とはいうものの，やはり何かおかしいと感じるのはなぜでしょうか．そのひとつの理由は，「十分多くの回数」をたかだか20回くらいだと考えていることです．実際には，赤が10回続いたからと言って，その次にすぐ10回黒が出なければいけないわけではなく，もっともっと長い目でみて赤と黒が半々ならば「赤が出る確率も黒が出る確率も1/2」となります．10回さいころを振っても10000回振っても，具体的な回数をあげているかぎり，それは「十分な回数」ではありません．

２つ目の理由は，次の１，２を混同しているからです．

１．赤が10回続けて出れば，次は黒が出やすくなる
２．「赤が10回続けて出て，次に赤が出る」確率は小さい

「赤と黒がおのおの毎回同じ確率で出る」という仮定を認めるならば，何回赤が続けて出ても，次に赤が出る確率は1/2です．したがって，この仮定を認める限り１．は誤りです．

一方，上の仮定が正しいとしても，赤が2回続けて出る確率は1/2×1/2＝1/4，赤が3回出る確率は1/4×1/2＝1/8で，赤が10回続けて出る確率は1/2の10乗すなわち1/1024ですから，さらに赤がもう１回でる確率は1/2048ととても小さな値になります．ですから２．は正しいということになります．

しかし，このとき「赤が10回続けて出て，次に黒が出る確率」は，赤が10回続けて出る確率が1/1024ですから，さらに黒が１回でる確率もやはり1/2048となります．つまり，２．は

赤が出る確率が1/2で，「赤と黒がおのおの毎回同じ確率で出る」という仮定が正しいならば，

「赤が10回続けて出たあとで，次に赤が出る確率」も
「赤が10回続けて出たあとで，次に黒が出る確率」も
同じで，どちらも小さい

と言っているにすぎません．

問題２

```
 ∧∧
≡・・≡
 (   )～
```

「AもBも当たりが入っている確率は同じだから，選ぶ箱をAからBに変えても変えなくても，当たる確率は同じ」じゃないんですか？

Aに当たりが入ってる確率は，確かに1/3やな．でも，司会者がCの箱を開けて，当たりは入ってないのがわかったから，Cの箱に当たりが入っている確率はゼロや．それでもBに当たりが入ってる確率は1/3かいな？

```
  ∧◆∧
Ξ o-o Ξ
 (     )～
```

AからBに変えるほうが有利です．

A, B, Cのどれか１つだけに当たりが入っていて，参加者にはどの箱に中が入っているかを知る手がかりはまったくないので，最初にAを選んだ段階でAに当たりが入っている確率は1/3です．したがって，Aに当たりが入っていない確率は2/3です．

司会者がCの箱を開ける前は，「Aに当たりが入っていない」ことは「Bに当たりが入っているか，またはCに当たりが入っているかのどちらか」と同じ意味であり，その確率は2/3でした．ところが，Cに当たり

が入っていないことを司会者に教えられたので,「A に当たりが入っていない」ことは「B に当たりが入っている」ことと同じ意味になりました．したがって，この時点では，A に当たりが入っている確率は 1/3, B に当たりが入っている確率は 2/3 となりますから，B を選ぶほうが当たる確率は高いということになります．

ところで，下の考え方はどこがおかしいのか，という質問を，以前受講者から受けました．

> 最初に A を選んだ段階で，B に当たりが入っている確率は 1/3，入っていない確率は 2/3 だ．「B に当たりが入っていない」ときは「A または C に当たりが入っている」ことになるから，「A または C に当たりが入っている」確率が 2/3 だ．ところが C がはずれであることがわかったので，A に当たりが入っている確率は 2/3 である．だからそのまま A を選ぶほうがよい．

この疑問に答えるポイントは，**司会者は，参加者が選んだ A を開けることはできない**ということです．

司会者は，B か C のどちらかだけを開けることができます．また，司会者は当たりがどの箱にあるのかを知っているので，箱を開ける時には当たりが入っている箱を避けて開けます．この状態で「B を避けて C を開けた」のであれば，B に当たりが入っている確率は高まります．しかし，A については，司会者は A を開けることは絶対にないので，司会者が B を開けようが C を開けようが，A に当たりが入っているかどうかの手がかりにはなりません．ですから，「司会者が B を開けた」あるいは「C を開けた」というできごとがあっても，A に当たりが入っている確率は変わりません．ですから，「C がはずれであることがわかったので，A に当たりが入っている確率は 2/3 である」という結論にはなりません．A に当たりが入っている確率は変わらないので，B あるいは C を開けても「B または C に当たりが入っている確率は 2/3」ということは変わらないことになり，最初の解答のような推論ができます．

この問題は,「モンティ・ホールのパラドックス（逆説）」として知られているものです．インターネットで「モンティ・ホール」で検索するといろいろな解説が出てきます．講義のウェブサイトの「統計データ・ツールへのリンク」にいくつかリンクを掲載しています．

ところで，この問題と答えは「司会者は，どの箱に当たりがあるのか知っている」という前提を信用しているから成り立つ話です．もし,「司会者はどの箱に当たりがあるのかを知らず，偶然 C を開けたらはずれだった」というのなら，話は変わってきます．これに関しては，上のリンクにある塩沢由典「モンティ・ホールの問題をどう考えるか」に説明があります．しかし,「司会者は，どの箱に当たりがあるのか知っている」ことが信用できるかどうかは確かめようがなく，参加者の信念の問題です．「確率は，測るものではなく，定義するもの」なのです．

問題３

おじさんの説によれば，１回のくじ引きではあたりもはずれも確率は1/2で同じですから，「10 回ひいて１回も当たらない」確率は$(1/2)^{10}$ です．つまり，おじさんが言うように「入っているくじの半分があたり」であるという仮定をしたとき，「１０回ひいて１回も当たらない」という標本調査結果が得られる確率は$(1/2)^{10}$ ということになります．こんな小さな確率でしか起こらないことが現実に起こっていると考えるよりも，「『入っているくじの半分があたり』というおじさんの説は正しくない」と考えるほうが自然ではないでしょうか？

この考え方は，統計的推測の重要な手法の１つで，**仮説検定**といいます．