2014 年度春学期　統計学　第８回

確からしさを記述する ー 確率

これから先の講義では，データの集まりに対して，ここまでで説明した度数分布やその代表値を「推定」する方法を説明します。推定とは，データの集まり全部を調べることができないときに，その一部だけを取り出して，度数分布やその代表値を知る方法です。それを理解するには，どうしても確率の知識が必要ですので，今日は確率のお話です。

「降水確率 40%」とは，「現在と同様な天気図のパターンが現れる機会をたくさん想定すると，そのうち 40%で雨が降る」という定義になっています。しかし，明日の降水確率が 80%だから明日確実に雨が降るわけではなく，また降水確率 20%だから明日は雨が降らないわけでもありません。では，確率とは結局何を意味しているのでしょうか？

「可能性」の集合

いま，くじをひくと，当たりが出たとします。現実世界では，くじは確かに当たったのであって，それ以外の結果は現れていません。

しかし，われわれは，くじびきとはいつも当たるものではなく，いま現れている「当たり」は偶然による結果だということを知っています。「偶然による」というのは，他の可能性もあった，つまり偶然によって他の結果になるかもしれなかった，ということを意味しています。この例の場合ならば，「はずれ」が出るという可能性もあった，ということになります。このような「結果が偶然によって決まる現象」を**ランダム現象**といいます。

統計学の世界では，つねに，この「可能性の集合」を念頭において，考えを進めます。この例の場合ならば，「今は『当たり』という結果が現れたが，『はずれ』が現れる可能性もあった」と考えている，ということです。

そして，さらに「どの結果が，どのくらい現れやすいか」を考えます。これを数字で表したのが**確率**です。「現れやすさ」などというものを，どのように数字で表せばよいのでしょうか。ひとつの考え方は，下のようなものです。

> **ある結果が現れる確率とは，**
> **これからその結果が現れる可能性のある 十分多くの回数の 機会があるとき，**
> **そのうち本当にその結果が現れる回数の割合である。**
>
> **次にその結果が現れる確率とは，**
> **遠い将来までの十分多くの回数の機会を考えて初めて言える「結果の回数の割合」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 0.3 であると考えます。このように，確率とは，本来は「遠い将来までの十分多くの回数の機会」を考えたときにはじめていえる，「結果の回数の割合」です。ただ，それを「次にくじをひくと，当たる確率は 0.3」のように，次の 1 回 の機会にあてはめて述べています。

ここでいう「当たりが出る」などの「結果」を，確率論の言葉では**事象**といいます。また，事象が起きる機会，この例ならば「くじを引くこと」を**試行**といいます。また，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。確率とは「特定の結果がおきる回数の割合」ですから，その値は 0 から 1（0%から 100%）の範囲になります。

しかし，この「定義」にある「これからその結果が現れる可能性のある十分多くの回数の機会があるとき,」という言い方には，少々おかしなところがあります。

1. 「これから」といっているように，確率は「未来のできごと」について述べています。しかし，未来のことは本当はわかりません。過去の経験をもとに，未来も同じようなことが起きるだろうと期待するのは，たいていは妥当かもしれませんが，そういう想像が正しいかどうかは誰にもわかりません。
2. 「十分多くの」といっていますが，何回なら「十分多い」のでしょうか。数学でいう「十分多い」とは，「誰もが納得するほど多く，しかも納得しない人がいたらすぐに増やすことができる」という意味です。かりに，10 万回くじをひくことにして，ほとんどの人が「それは十分多い」と納得したとします。しかし，一人でも「いや，それでは十分多いとはいえない」という人がいたら，その人の求めに応じて「では 10 万 1 回に増やしましょう」というように増やせるのが「十分多い」の意味です。もちろん，現実にはそんなことはできません。

つまり，上で述べた「定義」は，確率とは何かを述べてはいますが，それが実際に測れるとは言っておらず，むしろ「実際には測れない」ことを示しているのです。

しかし逆にいえば，過去の経験を未来にも延長できると認めて，数学でいう「十分多く」ではなくても「かなり多く」と認められるくらいの試行を行えば，確率を推測することはできます。なぜならば，試行の数を何度も増やしていくと，そのうち問題にしている事象がおきる回数の割合は，その事象がおきる確率に近づいていくからです。このことを「大数の法則」といいます。この講義では，この考えにもとづいて確率を推測することで，データの集まりに対して代表値を推測することができることを，説明していきます。

確率の意味

仮に，あるくじの当たり確率がわかったとしても，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。

たとえば，プロのギャンブラーは日常的に多くの賭けをし，長い目で見た利益を考えていますから，常に確率が大きい方に賭けるほうが有利です。実際，確率論という数学の始まりは，ギャンブラーがさいころ賭博の有利不利を数学者に相談したことでした。しかし，1 回しか賭けをしない人にとっては，「確率が大きい」ことと「次の賭けで勝てる」こととは直接は結びつかないことになります。

「降水確率 80%」という表現も，このような意味でとらえる必要があります。天気予報には，長い人生で何度も接します。天気予報を信じて，「80%」のときに傘を持っていけば，長い人生の間には，うち 80%は「濡れなくてよかったな」，20%は「荷物になったな」となるので，20% のほうはがまんしても，80% の時濡れないほうをとるでしょう。

しかし，一生に 1 回しかひかないくじで，当たり確率が大きいほうに賭ける意味はあるでしょうか？「ない」とまでは言いませんが，「当たり確率が大きい」ことはあくまで「長い目で見たとき」の話でであることは，知っておく必要があります。

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，数学でいう「十分に多く」というのは，「誰も文句を言わないぐらい多く」という意味であって，何回振っても十分ではないのです。

また，さいころを１万回ふって，そのうち１の目が 1/6 の割合で出たとしても，それはあくまで「過去の実績」であって，その次に１万回さいころをふっても，１の目は１回も出ないかもしれません。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. 各目が同じ確率で出る
2. 各目が出る確率は，いつさいころを振っても同じである

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「特定の結果が現れる回数の割合」の問題を，「（さいころの目の種類などの）可能性のある結果の種類の割合」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**[1] 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「このような仮定を認めれば」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。確率は，**遠い将来までを長い目で見てはじめて言える「特定の結果が現れる回数の割合」を，次の１回の機会にあてはめて述べたものにすぎません。**また，確率の定義には「十分多くの回数さいころを投げる」という現実には実行不可能な操作や，「各目が同じ確率で出る」という真偽を確かめられない仮定が含まれています。ですから，確率は**測定するものではなく，何らかの仮定をおいて「定義する」もの**なのです。この講義で扱う統計学では，概ね常識的に確率を理解しておけば十分ですが，ここまでに述べた確率の「あやしさ」は承知しておいてもらいたいと思います[2]。

ルーレット：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[3]。ラプラスの定義によれば，「赤と黒がおのおの毎回同じ確率で出る」こと

[1]「数学的確率」ということもありますが，現代数学の確率論でいうところの確率と混同するおそれがあるので，あまり一般的ではありません。
[2]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。
[3]正確には，色のついていない「０」や「００」があります。

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。しかし，頻度による確率の定義によれば，赤が 10 回続いたならば，次は黒が出る確率が 1/2 よりも大きくないと，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。なぜこのような感じがするのでしょうか？

その１つの理由は，前にも書いたことで，10 回さいころを振っても 10000 回振っても「十分な回数」ではない，ということです。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。

２つ目の理由は，次の２つのこと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 $(1/2) \times (1/2) = 1/4$，赤が 3 回出る確率は $(1/4) \times (1/2) = (1/8)$ で，赤が 10 回続けて出る確率は $(1/2)^{10}$ 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」といっているにすぎません。

確率のパラドックス

次の問題を考えてみましょう。

> 囚人が 3 人いて，「A, B, C の 3 人のうち，2 人は明日処刑される。あとのひとりは釈放される。誰が処刑されるかは明日朝発表される」と言い渡されています。以下は，処刑前夜の囚人 A と看守の会話です。
>
> > 囚人 A「看守さんは，明日誰が処刑されるかを知っているそうですね。私が明日処刑されるかどうかを，教えてくれとは言いません。ただ，私以外の B, C のうち，どちらが処刑されるかを教えてもらえませんか？」
> > 看守「本当は言ってはいけないのだが。。。実は，C は処刑されるのだ。」
> > 囚人 A「よく教えてくださいました。これで少し安心しました。」
> > 看守「なぜだ？」
> > 囚人 A「看守さんの話を聞く前は，私が処刑される確率は 2/3 でした。ところが，看守さんに C が処刑されると教えていただきました。ということは，あとは私と B の２人にひとりが処刑されるので，私が処刑される確率は 1/2 に減りました。」

さて，なにかおかしいと思いませんか？

この問題に答えるポイントは，**「『C が処刑される』という情報は，A が処刑されるかどうかについての手がかりになるか？」**ということです。

実は，看守が同じように「C は処刑される」と答えても，その答えに至る過程によって，「A が処刑される確率」は変わってきます。それはなぜか，以下の例で考えてみましょう。

看守が「C は処刑される」と答えるに至る，以下の３通りの場合を考えます。看守は判決を知っており，その判決は「B は釈放され，A, C が処刑される」であると仮定しましょう。

1. 看守は，「A, B, C のどれについて答えるか」を確率 1/3 ずつで決めるくじをひき，C について答えることになったので，「C は処刑される」と答えた。
2. 看守は，A については何も答えないことに決めていたので，B と C のどちらについて答えるかを確率 1/2 ずつで決めるコイン投げを行ない，C について答えることになったので，「C は処刑される」と答えた。
3. （問題文と同じ例）　看守は，B が釈放されることを答えてしまうと，A が処刑されることを答えるのと同じことになってしまうので，あえて B については触れず，「C は処刑される」と答えた。

これらの３通りの例は，いったいどう違うのでしょうか。

1. の場合，看守のくじの結果は「C について答える」となりましたが，他に「A について答える」「B について答える」という結果になる可能性もありました。つまり，

- C について答える → C は処刑 → A の運命は未定
- B について答える → B は釈放 → A の処刑が確定
- A について答える → A は処刑 → A の処刑が確定

の３通りの可能性のうち，実際に起きたのはいちばん上のできごとですが，下の２通りが起きるかもしれなかったわけです。

ここで，「ある事象が起きる確率」とは，「その事象以外にすべての可能な事象を考えて，それらの事象が起きうる試行を十分に多くの回数行なうとき，その特定の事象が起きるような試行の回数の割合」であることを思い出してください。

1. の場合にも，「看守がくじをひく」ことを，仮に十分に多くの回数行なうとしましょう。このときの「起きうる事象」には，上の３通りの可能性があります。ところが，1. の文では，３通りの可能な事象のうち，いちばん上の「くじの結果，C について答えると決まった」場合しか考えていません。

ということは，「十分に多くの試行」の回数から，下の２通りの事象が起きた試行を除いたうえで，確率を計算する必要があります。全体の試行の回数が変わりますから，A が処刑される確率も，看守がくじをひく前に比べて変化します[4]。

また，2. の場合，「A については何も答えないことに決めていた」のは確かですが，この場合でも

[4] この「全体の試行の回数を変える」計算の方法は，「条件付き確率」の計算とよばれています。

- C について答える → C は処刑 → A の運命は未定
- B について答える → B は釈放 → A の処刑が確定

の２つの可能性があり，そのうち上の場合しか考えていません。したがって，上の例と同様に，やはりこのコイン投げは A が処刑される確率に影響しています。

ところが，3. の場合，つまり本文の問題文の例では，「C について答える」ことは初めから決まっており，他の可能性はありません。したがって，看守が「C について答えた」ことがわかっても，上の説明でいう「十分に多くの試行」の回数は変わりません。したがって，A が処刑される確率も変わらないことになります。つまり，この問題の答は「A が処刑される確率は，看守の言葉を聞いたとしても，依然 2/3」ということになります。

このように，実際に看守が「C が処刑される」と答えたのは同じでも，「他にどのような可能性があったのか」が，確率の計算に影響するのです。このことは，看守の答は同じでも，看守の「癖」や「心の中」が問題になる，ということを意味しています。しかし，心の中を客観的に調べることは，通常はできません。本文の問題文では，看守は，A の運命を答えてしまうことにならないように「C が処刑される」と答えたと思われますが，それが本当かどうかは囚人 A にはわかりません。ですから，この問題の答えは，「看守の言うことを信じれば」確率はこうなる，という意味でしかありません。「確率は，測るものではなく，定義するもの」なのです。

上の「看守と囚人」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のと同じものです。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。

今日の演習

プロ野球の日本シリーズの時期になると，「第 1 試合で勝ったチームが優勝する確率はいくらいくらである」などといった記事が，スポーツ新聞によく載っています。過去の日本シリーズのうち，第 1 試合で勝ったチームが優勝した回数の割合をいっているわけですが，これは確率といえるでしょうか？