

2014 年度秋学期 統計学 第9回

確からしさを記述する – 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いなかった、ということの意味をしています。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、ということなのです。

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、少々おかしいところがあります。

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」とことと「次の賭けで勝てる」とこ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のでしょうか？「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、何回なら十分なのでしょう？ 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くらい多く」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だけで、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いところがあります。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、ここまで述べた確率の「あやしき」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回で出る確率は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回で出る確率もやはり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

ここでいうパラドックスとは、「理論的に正しい推論を行うと、直観とは異なった結果になる例」という意味です。確率のパラドックスとして有名な「モンティ・ホール問題」を、ここで考えてみましょう。

モンティ・ホール氏が司会するテレビ番組にて。箱が3つあり、ひとつだけに賞品が入っている。ゲストは箱をひとつ選ぶが、まだ開けない。

モンティはどの箱に賞品があるかを知っていて、ゲストが選ばなかった箱のうち、空の箱を1つ開ける。モンティはゲストに「いまなら、さっき選んだ箱ではなく、まだ開けていないもうひとつの箱のほうを選んでかまいません。どうしますか？」という。

選ぶ箱を変えたほうが、ゲストにとって有利だろうか？

$\begin{matrix} \wedge \wedge \\ \equiv \cdot \cdot \equiv \\ () \sim \end{matrix}$ 箱が残り2つになったから、どちらを選んでも当たる確率は1/2
じゃないんですか？

選択肢が2つやからといって、確率がどちらも1/2とは限らんよ。ゲストが元々選んでた箱に当たりがある確率は、モンティ $\begin{matrix} \wedge \blacklozenge \wedge \\ \equiv \circ \circ \equiv \\ () \sim \end{matrix}$
が空箱を開けたら、変わるんやろうか？

この問題について、コラムニストのマリリン・ヴォス・サヴァント氏が自身のコラムで「当たる確率は、ゲストが選ぶ箱を変えないと1/3、変えれば2/3」と書いたところ、「それは間違っている、変えても変えなくても1/2だ」という反論が殺到したというエピソードがあります。

この問題に答えるポイントは、「**モンティが空箱を開けたことは、ゲストが最初に選んだ箱に当たりがあるかどうかを知る手がかりになるか？**」ということです。実は、モンティが同じように空箱を開けても、モンティがどのような考えでそうしたかによって、「ゲストが最初に選んだ箱に当たりがある確率」は変わってきます。それはなぜか考えてみましょう。

この問題のもっとも簡単な解答は、以下のようなものです。

箱をA,B,Cとし、ゲストが最初にAを選んだとします。

このとき、賞品がAにある確率は1/3、「BまたはC」にある確率が2/3です。モンティは、B,Cのうち「必ず空の箱を選んで開ける」ので、開けたあとも「賞品がBまたはCにある確率は2/3」であることは変わりません。だから、「BまたはCのうち開いていないほう」に賞品がある確率は2/3です。

だいたいこれで間違いはないのですが、細かいことをいうと、

モンティはB,Cのうち「必ず空の箱を選んで開ける」
→開けたあとも「賞品がBまたはCにある確率は2/3」であることは変わらない

のは本当でしょうか？

このルールでは、賞品がBにある場合、モンティは賞品の入っている箱は開けないので、モンティが開けられる箱は自動的にCに決まってしまう。だから、モンティがその行動によって、ゲストに何かの情報を与えることはできません。賞品がCにある場合も、やはり開ける箱はBに決まっていて、同様です。これらの場合は、ゲストが「モンティがB,Cのどちらを開けるか」を見ていても、賞品がどこにあるかの手がかりにはなりません。だから、「開けたあとも『賞品がBまたはCにある確率は2/3』であることは変わらない」こととなります。

しかし、賞品がAにある場合は、このルールにしたがう場合でも、モンティはB,Cのどちらを開けてもかまいません。そこで、例えば「賞品がAにあるときは、モンティはB,Cのどちらを開けてもいいのに、そのときは必ずBを開ける」という裏ルールがあるとしましょう。そうすると、モンティがBを開けると、賞品がAにある確率が大きくなります。ゲストがそれを知っていれば、選ぶ箱を変えるかどうか

かの手がかりになるでしょう⁴。

あるいは、モンティが実はこのルールにしたがっておらず、「ゲストがどれを選んだかや賞品がどこにあるかにかかわらず、A,B,Cのどれかを同じ確率でランダムに選んで開けた結果、『たまたま』ゲストが選んでいなくてかつ空の箱を開けた」のだとしましょう。その場合、「偶然選んだ箱が空だった」ことで、残りの2つに賞品がある確率が大きくなります⁵。ただ、A,B,Cのどれかを同じ確率でランダムに選んだのですから、残りのどちらの箱についても平等に、賞品がある確率が大きくなります。ですから、上で述べた解答とは違って「ゲストが選ぶ箱を変えても変えなくても、当たる確率は同じ」ことになります。

このように、実際に「モンティが空の箱を開けた」という事実は同じでも、「モンティの行動には、他にどのような可能性がどんな確率であったのか」が、確率の計算に影響するのです。このことは、モンティの「癖」や「心の中」が問題になる、ということの意味しています。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。この問題の答えは、「モンティがルールを守っていれば」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。

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。また、参考文献として、ジェイソン・ローゼンハウス著（松浦俊輔訳）「モンティ・ホール問題 テレビ番組から生まれた史上最も議論を呼んだ確率問題の紹介と解説」ISBN978-4791767526 をお勧めします。

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つているわけですが、これは確率といえるのでしょうか？

⁴ですから、この問題には、厳密には「モンティは、賞品がAにある場合は、BまたはCを同じ確率でランダムに選んで開ける」と書いておかなければなりません。

⁵条件付き確率といいます。この講義では扱いませんが、浅野が以前担当していた「解析応用」の第7回で説明していますので、興味のある人はウェブサイトで見てください。