確率を推定する

黒沢 健 *

執筆日:2023 年 5 月 18 日

1　確率論の活躍の場

スマホのソーシャルゲーム等でガチャってありますよね。強いキャラクターやカードを引くためにガチャを引いたことがある人もいるかと思います。その結果、連続してガチャを回して、お目当てのカードが出なかったときに“爆死” なんて叫ぶ言葉を聞いたことがある人もいると思います。

このようなゲームには運が関わってくるため、必ず次に当たるとか当たらないとかを確定的に知ることはできません。だからといって、当たるも八卦、当たらぬも八卦で何も考えずに博打に出るのは少々やっつけな勝負になってしまうと思います。何も知らずに、天性の勝負師の勘に頼るというのは科学的観点からすると、勝負に勝つのは難しいと言わざるを得ません。確定的な事がわからなくても、運が絡むようなこのような試行には法則があります。確率的な要素を含んだ法則を**確率法則**と呼び、これを利用して有利に勝負をしたいものです。元々、確率論はギャンブルを起源にしていると言われています。ギャンブルというと不謹慎と思われる方もいらっしゃるかもしれませんが、現代確率論は株価の予測や様々な場面で利用されています。確率法則を知り、その上で何が予測できるのか、どのようなリスクがあるのかという事を事前に知ることで有利にゲームを進めることができるのです。

必ずしもギャンブルのようなゲームだけが確率論の出番ではありません。ショッピングサイトで自分の趣味趣向にあったお勧めの商品を表示されることがありますよね。人々の属性（性別や年齢など）や過去の行動履歴からその人にマッチするであろうというレコメンデーション機能が働いているからです。そのショッピングサイトが有能であれば、確率的に各人が最も興味を持ってくれるであろう商品を表示させることが可能となります。

2　確率分布

確率の適用事例として、先程のガチャの例を考えてみます。二項分布 (Binomial distribution) という確率分布を考えてみます。二項分布とは成功と失敗がそれぞれ確率 p と $1-p$ によって定まるベルヌーイ試行と呼ばれる試行を独立に n 回繰り返し、その成功の回数 x を確率として表したものです。この確率事象に対応する確率変数を X とし、$P(X=x)$ によって x 回成功する確率を表すことにします。上記のガチャの話に合わせて考えてみると、ガチャを一回引いて当たりか外れかを考えるのは、1 回のベルヌーイ試行に該当します。ここではお目当てのカードを 1 回あたり確率 p で引き当てるということに該当し、それを 10 連、つまり $n=10$ 回連続してチャレンジして、レアなカードを何枚 (x 枚) 引いたのかを考えることになり、パラメータが n と p の二項分布に従うことになります。二項分布に従った場合、x 回レアカードを引くという確率を計算する式が

$$P(X=x) = {}_n\mathrm{C}_x p^x (1-p)^{n-x}$$

で与えられます。ここで ${}_n\mathrm{C}_x$ は二項係数と呼ばれ高校生が学習する組み合わせ論の記号となります。お目当てのカードが出る確率が 10%と知らされていたとき、“10 回引けば当たるだろう” と考え、その結果、何も当たらなかった結果に憤慨した人もいると思います。

* 東京理科大学 理学部 応用数学科

それでは、この確率を上の式に当てはめてみましょう。全く当たらない確率は、$n = 10, p = 0.1$ の状況で $X = 0$ の確率を求めればよいわけですから

$$P(X = 0) = {}_{10}\mathrm{C}_0 \times 0.1^0 \times 0.9^{10} = 0.9^{10} \approx 0.348\cdots$$

となります。爆死する確率が約 35%になるわけです。つまり、3 回に 1 回は爆死してしまうわけです。人によっては意外と思う人もいるかもしれません。

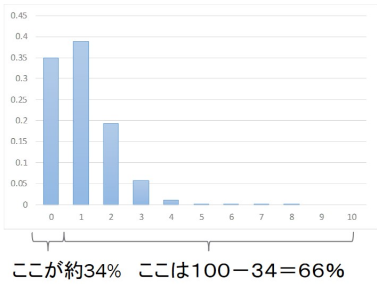

図 1: 二項分布の確率分布:$p = 0.1, n = 10$.

図 1 を見てください。これは先程考えていた二項分布の確率分布を表しています。横軸は回数 x を表していて、棒グラフの高さは各 x に対しての確率 $P(X = x)$ を表しています。皆さんは、たまにお目当てのカードが 2 枚当たるラッキーな出来事のことを忘れておりませんかね？爆死の余事象は $X \geq 1$ です。実は、爆死しないときというのは、1 枚もしくは 2 枚、もっと言えば 10 枚当たることの事象を含んでいるわけです。爆死する確率は確かに 3 回に一回は発生しますが、2 枚当たるというラッキーな確率は $P(X = 2) \approx 0.2$ はあるのです。つまり 10 連の試みを 5 回したとき、1 回は 2 枚当たっているのです。

では、ちょっと意外だと思った人の原因は期待値という言葉で解決が可能です。期待値とは平均的に確率変数が出る値を表す値です。ここではしっかり定義することは避けますが、以下のサイコロの例を考えてみると、わかりやすいと思います。サイコロの目は 1 から 6 があります。1 が出たり 5 が出たり 6 種類の値が存在しますが、それぞれの目が発生する確率は 1/6　なので、各目の値とその確率をかけ合わせて足し合わせた結果を期待値と呼び、“平均的” には

$$1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5$$

の値が出ることが期待できます。これが期待値です。上記の二項分布の場合、期待値 $E(X)$ は np で求められますので、今回の事例だと, $E(X) = 10 \times 0.1 = 1$ となり、1 枚くらいのカードが当たると “期待” できるのです。

3　確率分布のパラメータ推定

最近、ガチャ等のソーシャルゲームでは景品表示法などの規制で確率を表示する事になっているのかもしれませんが、必ずしもガチャに限らず、上記のような状況で確率 p がわかっていない場合はどうやって推定するのか？と考えるわけです。例えば、歪んだコインが与えられて、表が出るか裏が出るかを知りたい、もしくは何回か試行した時に何回表が出るのか知りたいという状況でも構いません。

古典的統計学の考え方では、このベルヌーイ分布や二項分布に設定されている確率 p をパラメータと呼び、元々知らされていない場合、神のみぞ知る真の値であり、その値を知ることができず、何回も試行を繰り返すことで統計的な意味で “近い” 推定をしようとするわけです。

確率が表示されていない場合、人柱となって、何度も何度もチャレンジして、経験や知見を重ね確率 p を知ろうとするわけです。統計学というのは、このように過去に得られた実験結果や試行の結果を用いて、真の値を推測したり、その確率分布を知ろうとする学問です。全く情報がなければ何も推測できません。つまり、このような統計的な学習データを用いて推測に使うわけです。

では、仮に5回コインを投げる状況を考え、3回表が出たという状況を考えてみます。まずは直観的に考えてみます。すると、5回中3回表が出たので $3/5 = 0.6$ という値を確率 p の推定値とするわけです。つまり5回チャレンジして、3回当たったという経験を活かし、この確率 p を推定し、次回からの行動（例えば先程の期待値の推定等）に活かすのです。

上記は直観的な推定となります。次は、統計学の**尤度**（ゆうど）**関数**を用いた方法を紹介します。尤度関数という言葉は大学で学ぶ言葉なので少々理解し難いかもしれませんが、簡単な例を紹介させていただきます。表が3回出たので、確率 p が3回かけ合わせ、裏が2回出たので、確率 $1-p$ を2回かけ合わせた関数を定義します。この "確率" みたいなものを**尤度**という専門用語を使います。この関数は p の値によって定まるため $l(p)$ で表すと、下記のように表現することができます。

$$l(p) = p \times p \times p \times (1-p) \times (1-p) = p^3(1-p)^2.$$

これは p に関する5次関数です。

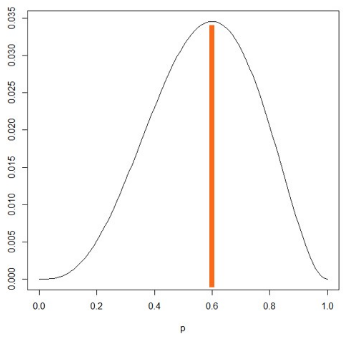

図 2: 尤度関数

図2は先程の尤度関数を横軸を p として描いたものになります。この5回中3回起きたという現象を表した関数となります。そして、この事象が起きたわけですので、これが最も大きい値を取るときの値を p の推定値と考えるのが、尤度関数を用いた統計的な推測となります。この関数を最大にする値 $\hat{p}$ を最尤推定値と呼びます。図中の縦線は、この関数の最大値を与える p の値を示すために引いており、0.6のときにこの関数が最大値を取ることがわかります。これが最尤推定値となり、直観的な推定と一致していることがわかると思います。

4 ベイズ統計学

古典的統計学という言葉を前章で使いましたが、古典的統計学とは違った考え方を用いてパラメータを推定しようと考えてみます。決して古典的な統計学に問題があるといっているわけではなく、新しい考え方の別な統計的な手法という認識で構いません。

古典的統計学においては、本来、パラメータ p は神様しか知らない未知の値という事でしたが、仮にパラメータ p が確率分布を持っていると仮定するのです。高校生までで学習してきた確率統計の考え方からすると奇妙な考え方だと思いますが、そこはあまり深くは考えず、そのようなやり方でパラメータ p の推論をする方法があると理解してくれたらと思います。

そしてベイズ統計では、情報を得る前、つまり観測を得る前に p の事前の確率分布というのものを仮定する必要があります。この確率分布を**事前分布**と呼びます。ここで考えているコイン投げの問題に相性が良い確率分布がベータ分布となります。相性が良いとは、数学的な扱いやすさの問題であり、本来、これが元々 p に対する経験に基づいて設定されたものではないことに注意しておきます。以下の式がパラメータ $\alpha, \beta > 0$ のベータ分布であり、これを事前分布 $\pi_{\text{prior}}(p)$ として設定することにします。

$$\pi_{\mathrm{prior}}(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} \qquad (0<\alpha, 0<\beta).$$

ここで $B(\alpha,\beta)$ はベータ関数と呼ばれる関数です。ここでは深く気にする必要はありません。また α,β は共に正の値で実験者自身が決めて良いパラメータとなります。

ベイズ統計ではベイズの定理と呼ばれる定理を用いて、前出の尤度関数をかけ合わせることで、**事後分布**と呼ばれるものを計算します。この計算方法については割愛しますが、n 回中、x 回表が出て、$n-x$ 回裏が出た時の尤度関数を使うと、事後分布が以下のように計算できます。

$$\pi_{\mathrm{post}}(p) = \frac{p^{\alpha+x-1}(1-p)^{n+\beta-x-1}}{B(\alpha+x, n+\beta-x)}$$

実は、この計算された事後分布もベータ分布となっており、事前分布とは異なったパラメータを持ったものになります。

では、先程の例に適用します。同様にコイン投げの問題を考え、5 回中、3 回表が出たとします。ここで、図 3 における 2 つのグラフは、事前分布のパラメータとして $\alpha=\beta=1$ を設定した時の値によって計算される事前分布と事後分布のグラフとなります。事前分布は破線で表されているもので、$\alpha=\beta=1$ を設定した時、p の値によらず一定の値を取る一様分布になります。つまり、事前に対する情報は存在していない状況になります。p を取る確率（密度）が同じ値になっていることから、どの p が起こりやすく、どの p が起こりにくいかがわからない状況にあります。一方で事後分布は曲線で表されたグラフとなっており、p の値によって高さが異なります。高い値を実現する p が最も起きやすいと考え、p の推定値として、最も高い値を与える p を推定値として選ぶことにします。その結果、今までと同様に p の値を 0.6 と推定することができます。

今回は事前分布として、確かな情報がなかったため、$\alpha=\beta=1$ として推定をしましたが、過去の知見などによって、α,β になんらかの情報があればこれを利用して推定することができます。

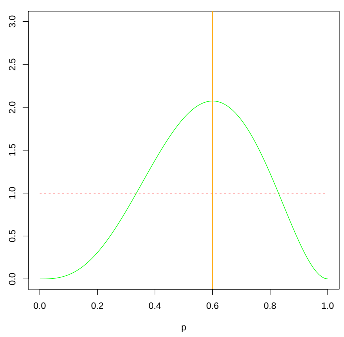

図 3: 事前分布と事後分布

5　おわりに

3, 4 章において、直観的な推定法、古典的な統計手法、ベイズ統計的手法を用いたパラメータ推定法を紹介しました。3 つの方法は結果としては同じ推定値を与えました。統計学を使わなくても同じ推論するのであればわざわざ尤度関数やベイズ統計学なんて道具を使う必要はないのでは？と思うかもしれませんが現実の世界はここで想定している状況ほど単純ではありません。複雑な分布の時でも、最尤推定等の統計的な手法を使って統計学者は真の値であるパラメータを推定しようと頑張っているのです。