

統計教育での「ベキ分布」の取り扱い

元一関工業高等専門学校
梅野 善雄

1 正規分布と中心極限定理

統計教育で頻出するいろいろな確率分布は、いずれも正規分布と関連した確率分布になっている。さらには、どのような確率分布であっても、「抽出する標本数を大きくしていくと標本平均は近似的に正規分布にしたがう」という中心極限定理がある。いろいろな検定でも、ある程度の標本数があれば正規分布にしたがっているとみなして議論が進められる。要するに、正規分布さえ抑えておけば、大体のことには対応できることになる。

しかし、中心極限定理が成り立つのは平均や分散が存在する確率分布である。それらが存在しない確率分布では中心極限定理は成立しない。正規分布や中心極限定理を強調しすぎた教育を受けていると、そのような確率分布に出会ったときに判断を間違える可能性がある。

2 いろいろな場面に現れる「ベキ分布」

統計の授業で触れられることはほとんどないが、確率密度関数がベキ関数で表される確率分布がある。確率密度関数を $f(x) = ab^a/x^{a+1}$ ($a, b > 0, x \geq b$) とすると、簡単な計算から分かるように $a \leq 1$ のときは平均が、 $a \leq 2$ のときは分散が存在しない。つまり、指数の値によっては分布を特徴づける代表値が存在しない。指数関数と比べるとベキ関数の減少の仕方は緩やかであるので、上位確率 $P(X \geq \alpha)$ が小さくてもかなり大きな値まで考慮する必要がある。また、正の定数 k に対して $f(kx) \propto f(x)$ が成り立ち自己相似性がある。座標軸を対数軸にとると、確率密度関数のグラフは右下がりの直線になるのが特徴的である。

近年、「ベキ分布」にしたがう現象が多数報告されるようになった。たとえば、「歴史は『ベキ乗則』で動く (マーク・ブキャナン, 早川書房)」によれば、次のような例が紹介されている。括弧内は参照した頁である。

- アメリカにおける、1985～1995 年の 4284 件の山火事の焼失面積と頻度 (p.150).
- アメリカの幾つかの州におけるバツタの大量発生 of 規模と頻度 (p.158).
- アイルランドとノルウェイの間にあるフェロー諸島における 1912～1969 年の「はしか」の感染者数と頻度 (p.160).
- 1984～1996 年のニューヨーク証券取引所の大企業 500 社の株価をもとにした株価変動の大きさと頻度 (p.233).

要するに、自然・生物・人間に関する事象に留まらず、経済に関することでもベキ分布が現れている。他にも、天文学のようなマクロの世界や細胞内部のミクロの世界でも、航空航路網・企業の取引関係・神経回路網などのネットワークの中にもベキ分布が現れている。

3 統計教育での取り扱い

ベキ分布は、正規分布以上にいろいろな分野のいろいろな場面で現れている。統計に関する限られた授業時間の中では、正規分布やそれに関連する事項を教えるだけで精一杯と思われるが、少なくとも「中心極限定理は万能ではないこと」、そして「平均や分散が存在しない確率分布も重要であること」を伝えて、「世の中の現象は正規分布だけでは理解できないこと」は強調しておくべきではないかと思われる。そのため、たとえば次のような課題を学生に課すことが考えられる。

平均や分散が存在しない確率分布として、確率密度関数も分かりやすいコーシー分布やベキ分布を紹介して、平均や分散が存在しないことを学生自身に確かめさせることが考えられる。特に、乱数を利用して標本平均を求めると、その値は計算のたびに大きく変動する。しかも、標本数を幾ら増やしても平均は安定しない。そのことを通して、中心極限定理が成立しないことを実感させることができる。

また、コーシー分布では標本数を幾ら大きくしても、標本平均の分布はもとのコーシー分布と全く同じ分布になる。ベキ分布の場合は、平均の分布は漸近的に同じようなベキ分布にしたがう。つまり、標本平均が計算するたびに激しく変動するように見えても、それは一定の確率分布にしたがった変動になっている。これらのことは、エクセル等の表計算ソフトを利用すれば簡単に確認することができるだろう。

世の中の現象は正規分布ばかりではないことを実感してもらうには、ベキ分布にしたがう実例を調べさせて、レポートにまとめさせるとよいかもしれない。Web 検索すると、ベキ分布が自然界の現象のみならず、経済現象や社会現象、さらにはいろいろなネットワークに関することとも関連していることに容易に気づくことができる。そのような調べ活動を通して、それらの現象は正規分布では理解できないことに気づいてくれることを期待したい。

いずれにしろ、卒業後に何らかの統計解析を行わなければならない場合のことを考えると、統計教育では下記のことは学生の記憶に留めさせておく必要があると思われる。

- 平均や分散が存在しない確率分布があること。
- そのような確率分布では中心極限定理は成立せず、標本数を幾ら大きくしても標本平均は正規分布には従わないこと。
- 検定では棄却域に入っている部分は「滅多に起きない」ものとして棄却されるが、大地震のように滅多に起きない棄却域部分こそ重要な場合があること。
- 世の中には分野を問わず「ベキ分布」にしたがう現象が多く見られ、正規分布や中心極限定理だけを念頭に置いて考えることは問題があること。

(注) 本稿は、日本数学教育学会 高専・大学部会論文誌 第 26 号第 1 巻 (2020.3) の著者による巻頭言をまとめ直したものである。

【参考】

「ベキ分布：関連リンク集」

[URL] <http://yunavi.la.coocan.jp/powerlaw.html>

[移転] <https://yunavi.lsv.jp/powerlaw.html> (2022.06)