

書 評

サンプリングって何だろう —統計を使って全体を知る方法—

廣瀬雅代，稲垣佑典，深谷肇一 著
岩波書店，2018 年

統計数理研究所
島谷 健一郎

サンプリングなど，改めて「何だろう」と考え，本を読むほどのものでないと思っている人も少なくないと思う．実際，「無作為抽出（ランダムサンプリング）」という言葉は，世論調査などマスメディアでもしばしば目にする．高校までの数学でも，赤玉と白玉の混ざった袋からランダムに玉を取り出したときの赤の個数，といった形で登場し，数学としても，この 2 項分布程度でほぼ足りているように見える．

しかし，一歩進んであれこれ素朴な疑問を呈してみると，ほどなく理解の不十分さに気づく．

例えば，世論調査というのは，国民全体の意向を知ることが目的であり，要は日本全体から「ランダム」に人をサンプリングして回答を得ればおよその見当をつけられる．ところでその場合，たとえば東京都が総人口の 1 割を占めるなら，サンプリングされる人の 1 割は東京からにするべきだろうか．それとも，「ランダム」にサンプリングするのだから，たまたま人口比より少なかったり多かったりしても，問題ないのだろうか．あるいは逆に，人口比に合わせたら「ランダム」でなくなってしまうので，そのような操作はしてはいけないのだろうか．

男女比はどうだろう．半々にするべきか．女性の方が多ければ，人口比に比例させるべきか．それとも，たまたまサンプリングされた人に女性が多くてもかまわない，むしろそうした偶然を排除してはいけないのだろうか．

年齢はどうだろう．10 代 20 代，... から等しくサンプリングするべきか．人口比に比例させるべきか．もし男女比や都道府県人口比に比例させてサンプリングするならば，各県内でそれぞれの男女比に比例させ，年齢も人口比に比例させるべきなのだろうか．

こうした疑問にどのくらい自信を持って答えられるだろう．

話は大きく飛ぶが，近年，野生のクマやイノシシが，人里や，ときには市街地にまで出没する．こうした動物は，個体数が増えたために出てくるようになったのだろうか．それには，ひとまず個体数を知る必要がある．しかし，野生動物を 1 匹残らず数え上げるのは不可能に近い．何らかの「サンプリング」調査に頼らざるを得ない．

こうした素朴な疑問に応えるべく、統計学は発展を続けている。本書は、そんな最前線に位置する3名の若手研究者が、サンプリングについて解説した書である。高校生から大人まで、広くサンプリングについて、学ぶ、あるいは考え直すのに適した解説書である。

本書の特長のひとつは、世論調査や野生動物の数、あるいは袋の中の玉の数といった現実の問題と、数学の世界とを、ひんぱんに往ったり来たりする記述にある。実は、これこそが統計思考である。現実の問題を、数学の言葉に直して数値を算出し、現場にフィードバックさせる。たいていの場合、その数値結果に満足できない。現実の複雑な問題は、容易に数学の言葉では表せないから当然の結果である。そこで、また数学の世界へ戻って再考し、それをまた現場に持ち帰る。これを繰り返す。

統計学の教科書や授業では、現実の問題は例題として数式の合間に見え隠れする体裁になっている場合が少なくない。一方本書では、現実問題を主題に置いた記述をし、その合間合間に、必要に応じて数学の言葉が登場する。本来、統計思考はこのように進む。

1章は、いわゆる赤玉と白玉で解説される2項分布の話である。玉を数える作業に従事しながら、大数の法則や中心極限定理といった数学の世界にも触れる。両者の頻繁な往来を通して、統計思考の基本を知る章である。2項分布など今さら学習する価値もない、と思われるかもしれない。しかし、多くの数学の授業では、2項分布を数学として理解することに終始し、2項分布を用いた統計思考には至っていない現実にあづかされる。

2章は、冒頭に挙げた世論調査に類した問題を題材とする。そこでは、人をランダムにサンプリングする調査に対し、サンプリングの問題と、サンプリング後にアンケートなどの回答をもらう段階の、両者について実践的な問題が続く。インターネットを使えば簡単にアンケートを実践できる今日だが、そこに統計思考が伴わない限り、偏った見解や都合のいい結果へ暴走する。

3章は、捕獲した野生動物に標識を付けて解放し、再び捕獲するという野外調査から個体数を推定する話を題材とする。当たり前のことだが、何らかの仮定を置かないと数学による推定値は導けない。仮定の妥当さやどの程度の非現実的さまでなら許容範囲か。動物の生活史や行動パターン、生息地の環境、等々を踏まえないと判断できない。

2章も3章も、冒頭に挙げたような問題には答えてくれない。それらは、調査目的や状況により、様々な複雑な統計モデルが要求されるが、そうした手法の解説は本書の主題ではないからである。主題はあくまで統計思考にある。

統計思考力は、例えば、2項分布という数学を理解する数理的思考力と、大きく異なる。本書はサンプリングという数学でなく、サンプリングという統計思考の良書である。