

## 数学月間(SGK)だより

谷 克彦

今年(2016.7.22)の数学月間懇話会は、数式の出ない一風変わった講演会になりました。数学月間は、数学同好者のためでもないし単なる数学の講習会でもありません。《数学と社会の架け橋》を目指しているのですから、今年のような数学月間の姿やテーマは、数学月間の原点であるとも言えましょう。今年の参加者は35人を数えました。

### (1) 数学者って、どんな顔をしている？

——亀井哲治郎・河野裕昭(亀書房・写真家)

始めの話題は、JIR(ジャーナリスト・イン・レジデンス)という活動の紹介です。編集者の亀井氏は、写真家の河野氏と組んで、数学者へのインタビューと写真撮影を続けて、すでに170人を超える人たち取材したといます。お二人が掛け合いで写真の説明をされ、撮影時の様子やエピソードを聞くことができました。まず、いろいろな講義風景が印象的でした。200人も入る教室に4人の学生という寂しくも贅沢な風景、オーケストラの指揮者よろしい講義ポーズなど、写真に切り取ってみると案外面白いものです。一方では、数学者の教室外の生活感のあふれる写真「耕運機を運転している姿など」もありました。



### (2) 世論調査は正しいか？

——松原 望(東京大学名誉教授, 聖学院大学)

議席事前予想(いわゆる世論調査)と開票実況中継(いわゆる「当確」打ち)の話がありました。世論調査は、ランダム・サンプリングで得たサンプル集合で行いますが、ランダム・サンプリングが本当にできているかは怪しい。性質(人口規模、地方性、産業構成など)が似ている地域を1つの層として、全国を180に層別し、これをもとに3,600人をサンプリングする方法(1963年の実施例)が紹介されました。最近では、ビッグ・データによる解析も行なわれるようになり、参院選(2016年7月)の予測では、かなり当たったことも記憶に新しい。あらゆるデータが収集され、論理的な因果関係を不問にして、予測目的で使われるのを、気持ちが悪く思うのは筆者だけではないでしょう。データ保護、調査の倫理、政治への従属、操作誘導など、心配ごとが今後に残されています。

さて、開票は田舎から開いて都市部は遅れるのだが、報道各社は一刻も早く当確を打ちたがる。しかし、自分の投票が開票されてもいないのに、当確が決まって良いものだろうか。「自分は正規分布の一点でしかない」と思い知らされるのだから、議席事前予想や当確予想が外れると快哉を叫びたくなる。過去には、当確が取り消しになった事例もあるそうだ。開票率、得票率の推移を見て、少なくとも開票率50%、あるいは、60%になってから、当確を打てば大体間違いはなく、松原氏が実際に当確判定に携わった1983年総選挙の新潟3区、田中角栄の得票率推移が例に紹介された。

### (3) がん登録の可能性

——田淵 健(都立駒込病院, 東京都がん登録室)

駒込ピペットは、感染症の避病院であったこの病院の発明(140年前)だそうだ。今年(2016年)新たにがんと診断される患者は、101万人を超える予測で、98万人(2015年)、88万人(2014年)と増加し、この3年間のがん死亡も、年間37万人程度だが増加傾向です。2014年から、多いがんのランキングや死亡率などが話題になり、がん検診も叫ばれています。ただし、過剰検診の問題もあり、がん検診を増やすことがどれほど有効なのか

はわかりません。がん罹患率の統計が整備されると、過去のがん罹患数データから、今年のがん罹患数を予測したり、次のようなことに使えます：

・助かるのか助からないのか？ どのくらいの人が助かるのか？/・同じような病気の人がどのくらいいるのか？/・この病気を治してくれる病院があるのか？/・どんな治療法があるのか？ 治療成績に違いがあるのか？

がん登録推進法は、遅ればせながら今年スタートしました(人口統計は、明治に確立している)。がん統計は、データ収集→処理登録→統計解析の流れで行い、特に、生データからがん登録を行うところが、混沌としていてとても難しい。これは数学者の仕事ですが、数学者の参入がないのは困ったことです。

その混沌を具体的にみると：・がん登録の届け出がされていない/・複数病院からダブって届けられる/・一人で多重がんをもつ、などの状態が生データにあるので、まず、同一性の判定が必要になります。死亡状況から、届け出がなかったがんを判定発見する作業も必要です。病院は電子カルテに変わった(医者が患者の顔を見なくなった)のだが、そのカルテ情報が構造化されていないので、残念ながら統計には役に立たないそうです。

病気を分類し、がんの定義を満たすリストを作る。そして、いろいろな届け出病名から、がんを特定する。例えば、「肺炎」と言っても肺がんが含まれているかもしれない。また、「心不全」は死因ではないので、死因の特定には、第1次原因、第2次原因、……、第5次まで見る必要になります。

いずれにしろ、生のデータは斯様に混沌としている。同値関係を定義したり、同値分類したりしてデータの構造化をする。これは数学者の仕事にはかなりません。

#### (4) 筆者からの追補

今年取り上げた「世論調査」と「がん登録」は、現在の社会が直面するホットなテーマです。これらに関連するその後の話題をここで追補します。

#### ■毛髪1本で乳がん検診

毛髪は約1cm/月の割合で伸びます。15cmの

毛髪ならその中に15ヶ月間の健康状態の記録が残されています。毛髪中に含まれるいくつかの元素(カルシウム、ストロンチウム、カリウム、ナトリウム、など)の分布状態(濃度の時間変化の記録)を調べて、がん検診(特に乳がん)ができるという新手法を、千川純一氏(ひょうご科学技術創造協会)が、2003年からSPring-8(放射光X線施設)で研究を始め、結果を2014年に論文発表しています。毛髪1本を抜いて送れば、がんの検診ができるとなると画期的です。この検査は人体への負荷や危険は全くありません。このがん検診法は、多くの臨床例や検体を集め、実証研究をする段階にあります。ランダム化検体でブラインド試験や、その判定のための統計学的手法、ROCカーブを用いた識別因子の決定など、まさに数学の出番です。

#### ■今年の米国大統領選の予測はずれ

米大統領選は、1票でも得票が多かった陣営がその州の選挙人を総取りするので、効率的に票獲得をすれば、少ない総獲得票数でも、選挙人数で逆転が可能です。各州の予測得票数にある不確定さが非線形に増幅され獲得選挙人数に影響するので、予測が大きく外れる傾向があります。

今回、大方の予想は、クリントンがトランプに選挙人で70人近い差の圧勝でしたが、結果は逆でした。ただし、総獲得票数は反ってクリントンの方が若干多いのです。支持率調査の全米平均値では、この状態をほぼ正しく予測したものの、州によっては大きく外れ、これが獲得選挙人数で大逆転を起しました。州によっては正しくランダム・サンプリングができなかったのです。特に、生産拠点の国外流出で、労働者や黒人層に移動混乱があるミシガン州(すたれたベルト地帯)などでは、その層のサンプルが少なく母集団の構成比が反映されなかったと思われます。また、ほとんどのメディアが、クリントン支持を意図的に流し、世論誘導をしたことと、トランプの乱暴な発言のため、トランプ支持を表明しない隠れトランプを生み、正しいサンプリングにならなかったのも原因の一つです。世論調査には数学的に批判されるべき問題がかなり存在するようです。

(たに・かつひこ/SGK 世話人)