

でたらめのお話し

—確率の効用—

逆瀬川 浩孝

前 説

「高校生のためのOR研究部会」は、文字どおり高校生にも分かるようなORの教材を作ろう、という趣旨から出発しています。高校生の段階でOR的なものの考え方に接することができれば、考えることの楽しさ、数理的思考の大切さ・有用性を理解する生徒が増え、OR関連の学科に進学を希望する人も増えるのではないかと、という期待も込められています。そのためには高校生にも分かる、共感される教材を用意しなければなりません。また、そのためには実際の高校生がどのような境遇にあるのかを知る必要があります。

というわけで、研究部会としては機会があれば現場の実状を知るためと、ORの普及を兼ねて高校生に直接話をする場を活用したいと考えています。その機会の1つが「出前授業」です。その経緯についてはこの特集に含まれる若山氏の記事に詳しく述べられていますので参照してください。筆者も今年の夏に和歌山県立日高高校の理数科3年生徒30余名を対象とした授業を経験しました。

授業の内容は「ランダム回答法」という標本調査法のトピックスを題材として確率論、乱数、シミュレーションについてまとめたものです。用意したプリント教材を使って「ランダム回答法」のやり方を一通り説明した後、実際に調査を実施して結果を求め、その確からしさをコンピュータを使って確かめる、というのが大体の流れです。実際に経験してみても改めてしっかりした教材作りの大切さを実感しました。

今の高校でORを教えることは大変にむずかしい状況にあるようです。その主な原因をあげれば、時間、ヒト、モノということができましよう。現在のカリキ

ュラムでは生徒が学ぶ内容が（盛りだくさんに）きっちり決められており、それを消化するだけで授業時間がいっぱいになってしまいます。もっと内容を減らすなどして余裕を持たせなければ、新しいことにチャレンジしようということはとても不可能です。一方、教える先生のほうでも、授業はもちろんのことそれ以外の仕事が増えていますから、よほど意欲が持てないと新しいテーマを授業に取り入れることはむずかしいでしょう。たとえ意欲があっても、その教材がほとんどありませんから、はじめから全部自分で用意しなければいけないということになります。

しかしまったくの八方ふさがりというわけでもありません。今回の出前授業を実施した日高高校では数学の先生グループが実験授業として、年に何回かカリキュラムにとられないテーマを生徒に教えるという試みが続けられているということです。そのようなグループに対して適切な教材を提供できれば、あるいは共同で開発することができれば、「高校生にORを教える」ということも可能になってくると思われます。

その際に重要なことは、当然のことですが主体が学ぶ生徒にある、ということでしょう。こう書けば分かってもらえる「はず」だ、興味を持つ「はず」だ、では通用しないので、実際にどういう反応があるかをチェックしながら作業を進めてゆくことが必要でしょう。そのためには出前授業のような共同作業が有意義なものになってくるでしょう。

以下の文章は当日テキストとして使用したプリント教材に当日の経験・反省を盛り込んだ改訂教材を要約したものです。

1. はじめに

でたらめ、言い換えれば、

ちゃらんぼらん、気まぐれ、でまかせ、良い加減、出たとこ勝負、ランダム、無作為、予想不可

さかせがわ ひろたか 早稲田大学 理工学部 経営システム工学科
〒160 新宿区大久保3-4-1

能、不可知、...

あまり良いイメージがありません。

しかし、何が起きるか分からない、というでたらめさがなかったら、世の中ずいぶん味気ないものになるでしょう。

- ・さいころがなかったら、トランプがなかったらゲームができません。
- ・テレビゲームやたまごっちでも、いつも同じパターンが繰り返されたら興味がなくなります。

でたらめを学問にしたのが確率です。

3枚のカードがあります。1枚目は表裏とも赤色、2枚目は表裏とも黒色、3枚目は表が赤色、裏が黒色になっています。さて、でたらめに1枚選んで机の上に置いたところ赤色でした。そのカードの裏が黒色である確率はいくつでしょう。

というような問題ですね。なぜこんな面倒くさいことを勉強しなければいけないのでしょうか。

- ・こんなものどこに役立つかわからない。
- ・いくら一生懸命考えても、あっているか違ってはいるか確かめようがない。
- ・いくら説明を聞いても確信が持てない。

などなど、理由はどうであれどうも不評です。確率の問題は高校生でなくとも分かりやすいとは言えないようです。何年前かにアメリカでこんな問題が議論を呼びました。プロの数学者でも正解できなかったとか.....

3枚の封筒のうち1枚だけにあなたが行きたがっているコンサートのチケットが入っています。どの封筒に入っているか当てたら差し上げましょう。まず、3枚の中から好きなものを選んでください。選ばれなかった2枚の封筒のうち少なくとも一方はチケットが入っていません。それを私が教えてあげます。そのあと、もう1回封筒を選びなおすチャンスが上がります。さあ、最初に選んだものをもう1度選ぶのと、残ったもう1つの封筒を選ぶのでは、どちらが当たるチャンスが大きいですか。

残った2枚の封筒のどちらかにチケットが入っているのだから、どっちを選んでも当たる確率は2分の1、というのが正解でしょうか。

分かりやすく説明して勉強することがあまり苦にならないような学び方を教えることはそうたやすいことではありませんから、これはこんなに役に立つのだから、

多少難しくても、一所懸命勉強しなければいけないと思ってもらえるような例をお見せするしかありません。確率が役に立つ、さあ勉強しよう、と思わせるようなことってどんなことでしょうか。

2. アンケート調査

皆さんはアンケート調査、という言葉聞いたことがあるでしょう。あるいは、社会科の勉強とか文化祭の研究発表をする目的で、自分たちでアンケート調査をしたことがあるかもしれませんね。あるいはまた、新聞社とか、お店などで、調査に協力してください、と言われた経験がある人がいるかもしれません。

新聞ではよくいろいろなテーマに関してアンケート調査を実施してその数字を公表しています。内閣の支持率が45パーセントに下がったとか、第一勧業銀行に限らずほとんどの銀行が不正な融資をしていると考えている人が80パーセントいるとか、がその例です。

さて、あなたがアンケートに答えることになったとき、どんな質問でも正直に答えることができますか。たとえば

「あなたはたばこを吸ったことがありますか」

とか

「あなたはいじめにあったことがありますか」

と聞かれても、全員が正直に答えてくれるとは思えません。他の人に答えを知られたくない、という質問があったら、答えないか、うそを答えてしまいますね。それでは何のために調査しているのか分からなくなってしまいます。アンケートは1人1人の答えを知るのが目的ではなく、集団としてどのような意見が多いかを調べるものです。1人1人の答えは調査した人には分からないけれど、全体の意見は正確に把握することができます、というような都合の良い調査のやり方はないのでしょうか。

あまり切実な問題を例に出すと、注意がそちらに集中してしまい、肝心なことが見えなくなってしまう恐れがありますので、ここでは

「あなたは、今恋人がいますか？」

という質問を例として取り上げることにしましょう。このクラスの中に恋人がいる生徒のパーセンテージを調べようというのが目的です（別にこんなことが分かってもどうということはありませんが）。

さて、この目的のためには、クラスの1人1人に上の質問に答えてもらう必要がありますね。その調査で「はい」と答えた人の数を全体の人数で割れば恋人を

持っている人の割合が計算できます。しかし、その調査結果は信用できるものでしょうか。上のように聞かれて正直に答えてくれる人はどれくらいいるでしょうか。そこで別の方法を考えることにしましょう。

もしも

「あなたは、今恋人がいますか？」

「あなたは、今恋人はいませんか？」

という2つの質問のどちらかをでたらめに選んで、どっちの質問を選んだかは言わずに、その質問に答えてください、というように聞かれたらどうでしょう。質問の答えは「はい」か「いいえ」です。調査した人はその2つの答えのどちらかを知るだけです。

「はい」と答えた人は

上の質問に答えて「はい（恋人がいます）」なのか、下の質問に答えて「はい（恋人はいません）」なのか分かりません。つまり、答えを聞いただけでは、答える人が恋人を持っているのかどうかを知ることはできません。プライバシーが守られているわけですから、直接質問した場合よりは正直に答えようとする人の割合が増えることで期待できます。このようなアンケート法をランダム回答法といいます。

「2つの質問をでたらめに選ぶ」と書きましたが、本当にでたらめに選んでいるのか、それともどうでも良いから勝手に選んだのか、調査した人には分かりません。そこで、たとえばトランプ52枚を良くきって1枚だけ選んだとき、それがハートならば上の質問、さもなければ下の質問に答える、というようにしてもらうことにします。トランプがなければさいころでも良いでしょう。以上の説明から、実際の調査の仕方をまとめてみたのが次の質問です。

ここに1組のトランプがあります。良くきってから私に見えないようにして1枚だけ取り出してください。もしそれがハートのカードならば、次の質問に「はい」か「いいえ」で答えてください。

「あなたは、今恋人がいますか？」

もしそれがハート以外のカードならば、次の質問に「はい」か「いいえ」で答えてください。

「あなたは、今恋人はいませんか？」

自分がこの調査に答えなければいけないとしたとき、「はい」と答えても「いいえ」と答えても、自分が恋人を持っているか持っていないかを調査員に知られることはない、ということが納得できますか。

全員が納得できたところで、それでは実際に調査し

てみましょう（といて調査を実施しました）。

3. ランダム回答法の理屈

今この教室にいるのは__人です。「はい」と答えた人は__人でした。（調査結果を記入して下さい）さて恋人を持っている人は何人いるでしょう。「はい」といっても恋人がいるのかいないのか分かりません。両方のケースが混ざっているのです。こんなあいまいな調査をして、その数字から何か分かることがあるのでしょうか。

もし恋人を持っている人は全体の40%と分かっているとすれば、「はい」という答えのおおよその数を見積もることができます。計算を簡単にするために、クラスに100人いるとしましょう。そうすると、恋人がいるのは40人です。「はい」と答えるのはそのうちのハートのカードを選んだ人です。ハートが選ばれる確率はだいたい4分の1と言って良いでしょう。したがって、40人の4分の1、10人が「はい」と答えるはず

です。一方恋人を持っていない生徒が「はい」と答えるためにはハート以外のカードを引く必要があります。その確率はだいたい4分の3と言って良いでしょう。したがって60人の4分の3、45人が「はい」と答えると考えて良いでしょう。

結局、「はい」と答える人はおおよそ $10+45=55$ 人いることになります。つまり全体の55%が「はい」と答えることになります。もし恋人を持っている人が全体の20%しかいないクラスで調査した場合は全体の65%が「はい」と答えることも計算できます。

この2通りの計算結果

40%ならば、「はい」の割合が55%

20%ならば、「はい」の割合が65%

は何を意味しているのでしょうか。今は恋人を持っている生徒の割合を決めてから「はい」と答える人の割合を求めましたが、逆に「はい」の割合が分かっているとき、恋人を持っている生徒の割合を求めることができるのでしょうか。上の計算からすると、「はい」の割合が60%のときは恋人を持っている生徒の割合は30%になりそうですが、それでよいのでしょうか。

そのとおりです。「はい」の数が分かると、おおよその恋人を持っている生徒の割合が計算できるのです。100人中 γ 人が恋人を持っているクラスでこの調査を実施したとき60人の生徒が「はい」と答えたとしま

ち、だいたい4分の1が「はい」と答えるでしょう。一方、恋人を持っていない人のうち、大体4分の3の生徒が「はい」と答えるはずです。結局、「はい」と答えた60人の生徒は2通りのいずれかですから、

$$60 = \frac{1}{4} \times r + \frac{3}{4} \times (100 - r)$$

という等式が成り立ちます。これは r に関する1次式ですから、解くことができます

$$r = 30$$

となりました。

一般に、「あなたは、今恋人がいますか?」という質問に答える確率を p としてみましょう。トランプの場合は $p = 0.25$ です。調査対象者を N 人とし、 m 人が「はい」と答えたものとしますと、上の式を参考にして次の式を導くことができます。

$$m = p \times r + (1 - p) \times (N - r) \iff \frac{r}{N} = \frac{m/N - (1 - p)}{2p - 1}$$

上で求めた r/N を計算する式

$$\frac{m/N - (1 - p)}{2p - 1}$$

を、恋人を持っている生徒の比率の推定値、と言います。この推定値を使えばどんな調査結果からもちどころに恋人を持っている生徒の割合を求めることができます。

4. ランダム回答法の信頼性

—コンピュータシミュレーション

ちゃんと質問に答えた人は誰もいないのに、全体の比率が分かってしまうというのは、式で説明されても、なんとなく釈然としない、という人が多いのではないのでしょうか。結果は本当に正しいのでしょうか。もし調査をやり直したとしたらどうなるのでしょうか。普通の調査ならば、調査される人が良い加減な答えかたをしない限り、何回調査をしても同じ結果が返ってきます。しかし、ランダム回答法の場合は違います。

やり直しの調査で選ぶカードは1回目の調査と同じであるという保証はありません。それどころか1回目ではハートを選んだ人のうち平均的には4分の3は2回目ではハート以外のカードを選びますから、当然「はい」と答える人の顔ぶれが違ってきますし、「はい」の回答数も違ってきます。

このクラスでもちょっと試してみましよう（といって2回目の調査を実施しました）。

2回くらいでは様子が良く分かりません。でも、こ

んなことを何回も繰り返すのはちょっと苦痛です。コンピュータを使って計算できないかしら。

コンピュータにトランプを渡しても何もできませんから、それに代わる何かを考えなければいけません。たいていのコンピュータにはEXCELという計算に便利なソフトが付いていますのでそれを動かしてみましよう。

A1からA4までのセルに「=rand()」と入力してみてください。そうすると、各セルにたとえば0.12308のように1より小さい数が表示されるはずですが、4つのセルの数字は一見でたらめですね。キーボードの上の方にあるF9というキー（再計算キーとも言います）を押してみてください。こうすると3つのセルの内容が変わります。F9を何回か押して、表示される数字のルールを見つけ、次にF9を押したら何が表示されるか当ててみてください。

多分、当たった人はいないでしょう。F9を押して次々と表示される数字に何の規則性も見出すことはできません。このようにでたらめに出てくる数字を乱数と呼んでいます。乱数を使えばトランプから1枚取り出すという実験を、トランプを使わずに実行することができます。

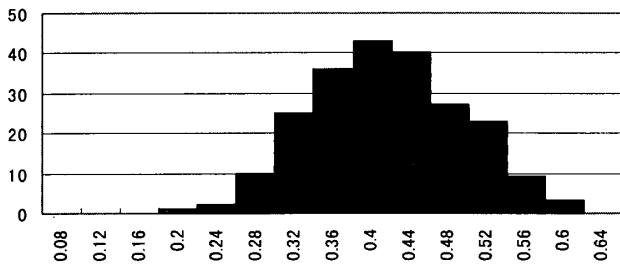
=rand()は0から1の間のでたらめな数を計算するEXCELの関数です。0.25より小さい数が表示される確率は4分の1と考えられますから、もし「rand() < 0.25」ならば「ハート」、さもなければ「それ以外」という指示をコンピュータに与えることができれば、コンピュータがトランプから1枚取り出して、それが何であったかを自分で判定することができた、ということになりませんか。

そこでセルB1に「=if(A1<0.25, “ハート”, “それ以外”）」と入力して、それをB2からB4までコピーください。B列の表示がA列の乱数の大きさによって異なっていることが分かりますね。これは4人の人がカードを取り出したとしたらどうなるかということコンピュータが「計算した」ことになりませんか。前と同じようにF9キーを何回か押すと、全体の4分の1くらいは「ハート」と表示されること、しかし毎回どれかちょうど1つのセルに「ハート」と表示されるわけではない、ということが分かります。つまり平均的に4分の1がハートといっても1回の実験で（調査で）必ず4分の1の人がハートを選ぶわけではないのです。

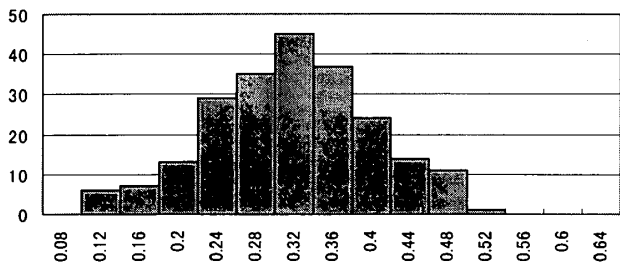
このようなやり方で、ランダム回答法の調査をコン

コンピュータに代行させることができます。

次のヒストグラムは、100人中40人が恋人を持っているクラスで調査を200回繰り返した場合の結果です。横軸の数字はそのクラスの上限です。正しい比率は0.4ですが、かなりばらついていることが分かります。しかし0.2より小さくなったり、0.6より大きくなったり、極端に違うことはほとんどない、ということが分かります。



次の図は恋人を持っている人が30人のクラスでの実験結果です。



同じようにばらついています。今度はヒストグラ

ムの中心が正しい比率の0.3になっています。

問題は、もしこの方法で調査するとしたら、1回しか調査できないことです。ですから、推定された比率がたとえば0.3であったとすると、本当の比率は0.4かもしれないし、0.3、あるいは0.2かもしれない、実際のところは分からない、ということになります。本当は比率が0.4であるのに、推定された比率は0.55ということも有りえます。しかし、0.6とか0.2とかになることはないということと言えるわけで、大雑把に全体の様子をつかむことがある程度可能であると言えます。

このように、何が起きるか分からない状況をコンピュータで「計算」させて必要な情報を得る方法をシミュレーションと言います。上の例のように、実際には1回しか調査しない場合でも、同じ状況で調査をやり直したとすればどのような回答が得られる可能性が高いか、状況が変わったら結果がどれくらい変わるものなのか、などというように実際には調査不可能なことを調べるために、コンピュータが有効に使われています。また、確率がからむ場合は上のランダム回答法のように答えを聞いても釈然としない、というような問題がありますから、そのような場合は実際に乱数を使って実験をすることによって納得してもらったり、実際にどのようなことが起きるのかを実感してもらうためにもシミュレーションが使われます。EXCELのrand()関数を使って何でも良いからシミュレーション実験を考えてみてください。