雑感　MATHコン

■　MATHコンと略称される塩野直道記念「算数・数学の自由研究」作品コンクールが昨年末，第4回の入賞者を発表した（http://www.rimse.or.jp/research/winner4th.html）．

小学生から高校生までを対象としたコンクールで，第1回の最優秀賞に選ばれた「メロスの全力を検証」（愛知県　愛知教育大学附属岡崎中学校2年　村田一真君）は，「メロスは走っていなかった」として，当時，話題を呼んだ．

■　今回，勤務先の3年生武藤拓志君が「コンプガチャの確率について」で優秀賞（日本数学検定協会賞）を受賞した．

最優秀賞や優秀賞を受賞した高校生のレポートなどを読むと，部活などで長期間にわたって研究を進め，それをまとめた大作がある一方，割と短時間で研究を行ったと思われるものもある．

武藤君のレポートは，本校が理系の3年生で行っている理数の「課題研究」（前期のみの期間履修）の中で，4月から9月まで取り組んだ結果である．実際には3ヶ月程度の期間で研究は終わり，残りの期間はレポートの執筆に当たっていたように思う．

一般に，研究では大がかりで時間がかかるものもあるが，数学の研究では，テーマが適切で，良い解法が見つかれば割と短時間で一定の成果を得ることが出来る場合がある．

本校の「課題研究」でも，理科分野は実験の準備などに相当の時間が必要で，短期間には思うような研究結果が得られていないものが多い（ただし，研究そのものには意義があることを断っておく）のと対照的である（高校生の理科分野の様々なコンクール入賞者は，ほぼ部活での研究で占められているのではないか）．

■　2年間数学の「課題研究」の指導を側面支援してきたが，このような成果が得られたことはうれしい．

「高校生の程度をやや超えた数学も使いこなし，高校生の数学の自由研究レポートとしてはオリジナリティが感じられ，高評価で，表彰されるに十分値すると考えます」との中央審査委員会による講評は，研究は評価されてこそなので，何よりありがたいことである．

■　さて，彼の研究について述べたい．

最初，コンプガチャが分からず教えを請うた．

Wikipediaによれば「コンプリートガチャとは携帯電話用などのソーシャルゲームにおけるアイテム課金の仕組みの1つ．カプセルトイ（ガチャ）のようにランダムに入手できるアイテムのうち，特定の複数アイテムをすべて揃える（コンプリートする）ことで稀少アイテムを入手できるシステムのこと．コンプガチャとも呼ばれる」とある．右上がカプセルトイである．

研究計画の中で研究内容が「等確率，レア有などの2，3種類の場合を計算する」とあって，「ああ，あれね．それって，よく見かける内容じゃないの？」というのが私の最初の印象であった．

実際にはよく見かけるのは等確率の場合で，

> n 種類のアイテムがあり，それらの出る確率がすべて $\dfrac{1}{n}$ である場合，コンプリートになる（全種類のアイテムが揃う）ために必要な試行回数の期待値は，$E_1 = n\displaystyle\sum_{k=1}^{n}\frac{1}{k}$ である．

というものである．

■　彼はこの証明から出発して，レアもの（滅多に出ないアイテム）が何種類かある場合についての期待値計算に取り組んだ．

テーマを聞いて，私も考えてみた．

2種類のアイテムA，Bがあり，Aの出る確率 p が小さい $(0<p<1/2)$ とする．この場合，コンプリートになる試行回数の期待値 E_2 については，$E_2=\displaystyle\sum_{k=1}^{\infty}(k+1)\{p^k(1-p)+(1-p)^k p\}$　で，この計算を行うと，$E_2=\dfrac{p^2-p+1}{p(1-p)}$ となる．

3種類A, B, Cで確率が $p,\ q,\ 1-p-q$ の場合を考えてみると，A，B，Cの順番に k 回でコンプリートする期待値は $\displaystyle\sum_{s=1}^{k-2}kp^s(p+q)^{k-s-2}(1-p-q)$ であることから，この形の式をA, B, Cの順列6通りについて作って加えた $E(k)$ に対して，$1-p-q=q$ すなわち $q=\dfrac{1-p}{2}$ の場合については $\displaystyle\sum_{k=3}^{\infty}E(k)$ を計算すれば良さそうである．

とは言え大変な計算で，私は数式処理ソフトなどの助けを借りて計算を行い，$E_3=\dfrac{1+6p^2-p^3}{p(1-p)(1+p)}$ を得た．

$n=4$ の場合（確率は，p，$\dfrac{1-p}{3}$，$\dfrac{1-p}{3}$，$\dfrac{1-p}{3}$ の場合）も同様にして，$E_4=\dfrac{4p^4+6p^3+79p^2+6p+4}{2p(1-p)(p+2)(2p+1)}$ となった．

とは言え，n が増えるに従って，「階乗」的に計算量が増え，この方法ではこの程度が限界である．

■　彼はこのネックを，あるアイデアを持って切り抜けた．詳しくはWeb上に公開されている彼のレポートを見ていただくしかないが，ざっくり言えば「確率 p で起こる事象が1回起こるためには平均 $\dfrac{1}{p}$ 回の試行が必要だ」というある意味当たり前のことに，結果からの条件つき確率(？：適切な表現が見あたらない)をからめて巧妙に処理している．

その結果，n 種類のアイテムがあり，レア・アイテム（出る確率 p）が1つ，残り $n-1$ 種類は等しい確率で出るコンプガチャで，コンプリートする（全 n 種類のアイテムが揃うのに必要な試行の）回数の期待値 $E(n)$ は

> $$E(n)=\sum_{s=1}^{n}f(s)g(s)\ .$$
>
> ただし，
>
> $$f(s)=\frac{(n-1)p}{(s-1)p+n-s}\times\prod_{t=1}^{s-1}\frac{(n-t)(1-p)}{(t-1)p+n-t},$$
>
> $$g(s)=\sum_{k=1}^{s}\frac{n-1}{(k-1)p+n-k}+\sum_{k=s+1}^{n}\frac{n-1}{(n-k+1)(1-p)}.$$

であることを導き出したのである．

■　一般の n の式であることがこの式の評価すべき最大の点であり，数式処理ソフトなどを用いて計算すれば，任意の n について，$E(n)$ を求めることが出来る．

実際，私はこの式を用いDeriveで計算してみた．$n=2,\ 3,\ 4$ 程度なら0.数秒で正しい式を，$n=30$ でも30秒程度で p の式（分子，分母とも p の30次式！）を出力した．

■　その後，いろいろ調べると，この問題は「クーポン・コレクターズ問題」とも呼ばれ，1954年頃にはVon Schellingによって一般的な式が得られている．文献*)に載っている式は，$\displaystyle\sum_{k=1}^{n}(-1)^{k-1}\sum_{|j|=k}\frac{1}{p_j}$　…　①となっている．Σの条件が今ひとつよく分からないが，3種類の異なるアイテムP，Q，R（出現確率はそれぞれ $p,\ q,\ r$）の場合で言えば，①式は

$E=\dfrac{1}{p}+\dfrac{1}{q}+\dfrac{1}{r}-\dfrac{1}{p+q}-\dfrac{1}{q+r}-\dfrac{1}{r+p}+\dfrac{1}{p+q+r}$ となるという．

$q=r=\dfrac{1-p}{2}$ で計算すれば，$E(3)=\dfrac{1+6p^2-p^3}{p(1-p)(1+p)}$ という結果が得られ，武藤君の導いた式と同じ式である．

①は個々の確率が異なって良いという一般性を持っていて優れた式だが，n に対して一々式を作る必要がある．

■　実は今回，彼の研究の「指導」は，ほとんど何もしていない．

レポートの査読段階できちんと読んでもらえるものにする（ここが一番肝心と考えた）ために，記述内容の配列や例示の追加などのアドバイスを行ったくらいである．

*　例えば，http://www.d.umn.edu/math/Technical%20Reports/Technical%20Reports%202007-/TR%202012/dai.pdf

2017年1月1日