チョコボールのエンゼルマークを当てる話

数学Aの課題学習として生徒が自発的に考察を進められる教材を日々探しているが、これは同僚の先生との会話から得られた話題であり、生徒への投げかけをしながら考察したものである。生徒ならばどのような考察をするか、また、教師がどのような発問をすればより効果的な教材となるかということを考えてまとめた。

チョコボールというお菓子がある。(右)
これを買うと金または銀のエンゼルマーク(くちばしの形)がおまけでついていて、
これを集めると、金なら1枚で、銀なら5枚で景品がもらえるらしい。
会社のHPで確認すると、金や銀のエンゼルマークが入っている確率は
秘密ということであるが、ある人から聞いた話では
「金が入っている確率は $\dfrac{1}{1000}$ 、銀が入っている確率は $\dfrac{1}{30}$ 」
ということである。このうわさが本当であるとして、

(問題1)金や銀のエンゼルマール 1 枚を当てるためには
何個買えばよいだろうか。

(考えられる解答)
チョコボールを n 個買うとして、金、銀のエンゼルマークが
少なくとも1枚当たる確率は、

金の場合は　$p=1-\left\{1-\left(\dfrac{1}{1000}\right)\right\}^n=1-\left(\dfrac{999}{1000}\right)^n$

銀の場合は　$p=1-\left\{1-\left(\dfrac{1}{30}\right)\right\}^n=1-\left(\dfrac{29}{30}\right)^n$　である。

これをエクセルで計算したところ次のような結果になった。

(銀の場合)
銀の場合は、30 個買えば、少なくとも1枚当たる確率は 63%以上で、89 個買えば少なくとも1枚当たる確率ははは 95%以上になる。

(金の場合)
金の場合は 1000 個買えば、少なくとも1枚当たる確率は 63%以上となり、3000 個買えば、少なくとも1枚当たる確率は 95%以上となることがわかる。

金チョコ1	n(試行回数)	1000
	p(確率)	0.001
	少なくとも1個当たる確率	0.632304575
銀チョコ	n(試行回数)	30
	p(確率)	0.033333333
	少なくとも1個当たる確率	0.638338487

これを一般的に考えてみる。

金のエンゼルマークが 1 枚当たる確率を $\dfrac{1}{k}$ とする。チョコボールを n 個買うとき、金のエンゼルマークが少なくとも1枚当たる確率は　$p=1-\left(1-\dfrac{1}{k}\right)^n$　である。

特に　$k=n$　のとき、　$p=1-\left(1-\frac{1}{n}\right)^{n}$　であるが、この値はｎがある程度以上に大きいとき、ある一定の値に近くなる（約 63%）ということに気がついた。そこでこの一定の値は何か考えてみた。

式で表すと　$\lim_{n\to\infty}\left\{1-\left(1-\frac{1}{n}\right)^{n}\right\}=1-\frac{1}{e}$　となり、これを計算すると約 63%になる。一定の値はこれであった。

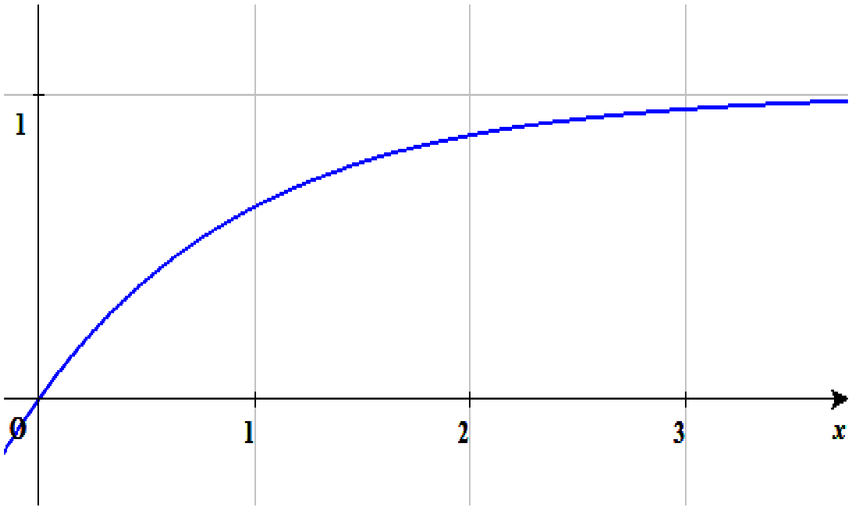
さらに一般化して、　$\lim_{n\to\infty}\left\{1-\left(1-\frac{1}{n}\right)^{kn}\right\}=1-\frac{1}{e^{k}}$　を考えることにより、次のことが言える。

（結論）

金のエンゼルマークが１枚当たる確率が　$\frac{1}{n}$　のとき、チョコボールを　kn　個買ったときに金のエンゼルマークが少なくとも１枚当たる確率は　$1-\left(1-\frac{1}{n}\right)^{kn}$　である。この確率は　$n\to\infty$　のとき収束し、その値は

$\lim_{n\to\infty}\left\{1-\left(1-\frac{1}{n}\right)^{kn}\right\}=1-\frac{1}{e^{k}}$　である。

$y=1-\frac{1}{e^{x}}$　のグラフ



このグラフから次のことがわかる。

（1）金のエンゼルマークが１枚当たる確率が　$\frac{1}{n}$　のとき、チョコボールを買う数を　n　個から　$2n$　個に増やしたとしても、金のエンゼルマークが少なくとも１枚当たる確率は２倍にはならず、

$\frac{1-\left(1-\frac{1}{n}\right)^{2n}}{1-\left(1-\frac{1}{n}\right)^{n}}=1+\left(1-\frac{1}{n}\right)^{n}\to 1+\frac{1}{e}\approx 1.36\quad (n\to\infty)$　倍となる。これは当たり前のことではあるが、消費者が陥りやすいところかもしれない。この場合、63%×1.36=85.68%であるから、２倍の出費をしさえすれば、金のエンゼルマークが少なくとも１枚当たる確率はかなり上昇する。

（2）一般にチョコボールを買う数を　n　個から　kn　個に増やすと

$$\frac{1-\left(1-\frac{1}{n}\right)^{kn}}{1-\left(1-\frac{1}{n}\right)^{n}}=1+\left(1-\frac{1}{n}\right)^{n}+\left(1-\frac{1}{n}\right)^{2n}+\left(1-\frac{1}{n}\right)^{3n}+\cdots\cdots+\left(1-\frac{1}{n}\right)^{(k-1)n}=\frac{1-\frac{1}{e^{k}}}{1-\frac{1}{e}}\quad (n\to\infty)$$ 倍になる。

今、十分大きな n について考えているとすれば、買う数を k 倍に増やしたときの金のエンゼルマークが少なくとも1枚当たる確率は関数 $y=\frac{1-\frac{1}{e^{k}}}{1-\frac{1}{e}}$ の挙動にほぼ等しいと考えることができる。

ここで $k\to\infty$ としてみると、 $\lim_{k\to\infty} y=\lim_{k\to\infty}\frac{1-\frac{1}{e^{k}}}{1-\frac{1}{e}}=\frac{1}{1-\frac{1}{e}}\approx 1.58$ となる。

つまり、無限に多く買うことができれば、約 63%×1.58 $\approx\left(1-\frac{1}{e}\right)\times\frac{1}{1-\frac{1}{e}}=100$ %であるから、確実に当たることになる。

以上のことは日常生活で言えばチョコボールの金のエンゼルマークや宝くじなど、1 回の試行で当たる確率が極めて小さいもの対して、当てたいと思っているけれどもなかなか当たらなくて、そのうちについ出費がかさんでしまったけれど、結局出費の割に当たりが少なかった、あるいは効果がなかったという経験と合致していると思われる。

（発展）
このような身近な確率で自然対数の底 e が現れるのは大変不思議であるが、これは実はポアソン分布と関係があった。ポアソン分布は、1 回の試行において起こる確率 p が非常に小さく、また試行回数 n が非常に大きい二項分布の近似として用いられる確率分布である。

今の場合、 $np=k$ （一定）であるから、ポアソン分布の公式から $P(X=0)=\frac{k^{0}}{0!}e^{-k}=\frac{1}{e^{k}}$ となる。

これを用いれば、当たる確率が $\frac{1}{n}$ のとき、 kn 個買ったときに少なくとも 1 個当たる確率は $n\to\infty$ のとき収束し、その値は $1-P(X=0)$ であると言える。

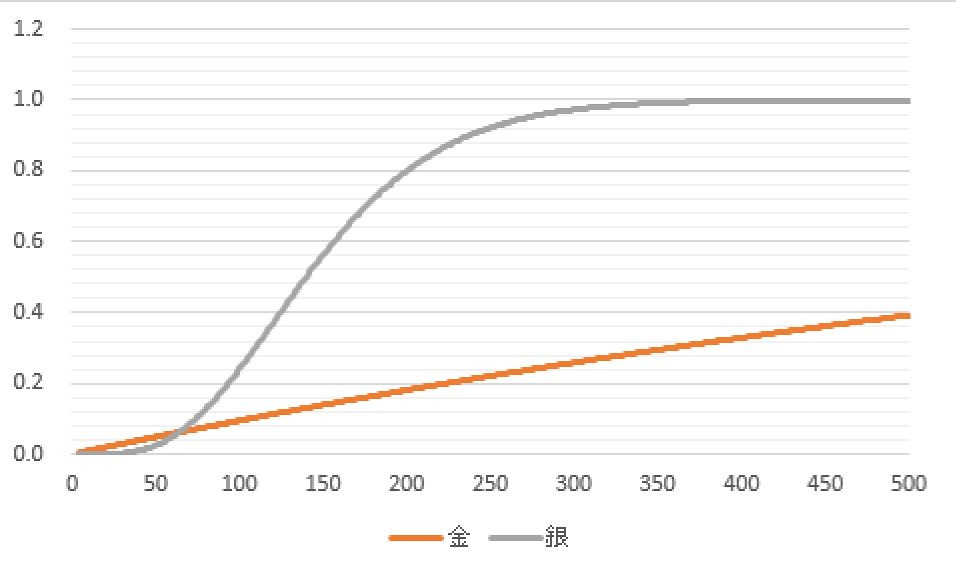
(問題2)
金のエンゼルマーク1枚、銀のエンゼルマーク5枚いずれの場合でも集まればおまけがもらえるということであるが、どちらのほうが早くおまけがもらえるようになるだろうか。

（考えられる解答1）
チョコボールを n 個買うとき、金のエンゼルマークが少なくとも1枚当たる確率 p_1 と、銀のエンゼルマークが少なくとも5枚当たる確率 p_2 を比較する。

$$p_1=1-\left(\frac{999}{1000}\right)^{n}$$

$$p_2=1-\left(\frac{29}{30}\right)^{n}-{}_{n}C_{1}\left(\frac{1}{30}\right)^{1}\left(\frac{29}{30}\right)^{n-1}-{}_{n}C_{2}\left(\frac{1}{30}\right)^{2}\left(\frac{29}{30}\right)^{n-2}-{}_{n}C_{3}\left(\frac{1}{30}\right)^{3}\left(\frac{29}{30}\right)^{n-3}-{}_{n}C_{4}\left(\frac{1}{30}\right)^{4}\left(\frac{29}{30}\right)^{n-4}$$

である。これをエクセルで計算した結果、下のグラフのようになった。（下図）



（結論）
チョコボールを買う個数が 64 個未満ならば金のエンゼルマーク 1 枚のほうが当たる確率のほうが大きい。
チョコボールを買う個数が 64 個以上ならば銀のエンゼルマーク 5 枚のほうが当たる確率のほうが大きい。

今回のようにたくさん購入することが多い場合は、銀のエンゼルマークが 5 枚集まっておまけがもらえるようになることのほうが起こりやすいということになる。

（考えられる解答2）
それぞれの場合でおまけがもらえるようになる回数の期待値を比較する。
金のエンゼルマーク 1 枚が当たったときに、それまでにチョコボールを買った個数を確率変数 X_1 とし, 銀のエンゼルマーク 5 枚が当たったときに、それまでにチョコボールを買った個数を確率変数 X_2 とする。
X_1 , X_2 の期待値はそれぞれ

$$E(X_1)=\sum_{k=1}^{\infty} k\cdot\left(1-\frac{1}{1000}\right)^{k-1}\left(\frac{1}{1000}\right)$$

$$E(X_2)=\sum_{k=5}^{\infty} k\left\{{}_{k-1}C_4\left(\frac{1}{30}\right)^4\left(1-\frac{1}{30}\right)^{k-5}\cdot\left(\frac{1}{30}\right)\right\}$$

となる。
$E(X_1)$ については（等差数列）×（等比数列）の和なので容易に計算できる。

$p=\dfrac{1}{1000}$ とおくと $\displaystyle\sum_{k=1}^{\infty} k\cdot(1-p)^{k-1}=\frac{1-(1-p)^n}{p^2}-\frac{n(1-p)^n}{p}\rightarrow\frac{1}{p^2}\ (n\rightarrow\infty)$ であるから

$$E(X_1)=p\sum_{k=1}^{\infty} k\cdot(1-p)^{k-1}=p\cdot\frac{1}{p^2}=\frac{1}{p}=1000$$

一方、 $E(X_2)$ はその意味を考えれば、銀のエンゼルマークが 1 枚が当たったときに、それまでにチョコを買った個数を確率変数 Y とすれば、 $E(X_2)=E(5Y)=5\times E(Y)=5\times30=150$ と計算できる。（負の二項分布）これより銀のエンゼルマーク 5 枚が集まっておまけをもらうことのほうが起こりやすいと言える。

（発展）
ここで現れた X_1 の確率分布や Y の確率分布は幾何分布と言われるものであり、上記の数列の和の計算から、幾何分布の期待値は $E(X_1)=\dfrac{1}{p}$ であるという事実が示されたことになる。

このようなことを考えているうちに、チョコボール28個目の購入で銀のエンゼルマークの1枚目が当たった。このことから銀のエンゼルマークが入っている確率の信憑性について考えた。

（問題3）
28個目の購入で銀のエンゼルマークの1枚目が当たった。このことから銀のエンゼルマークが入っている確率は $\frac{1}{30}$ と言ってよいか。

（考えられる解答1）
期待値から確率を求める
上の計算結果から $p\sum_{k=1}^{\infty} k\cdot(1-p)^{k-1}=p\cdot\frac{1}{p^2}=\frac{1}{p}$ であるので、これと比較すると、 $\frac{1}{p}=28$ これより $p\approx\frac{1}{28}$ となる。

（考えられる解答2）
最尤法を用いる。母集団の数は非常に大きいため、チョコを買うという試行は復元抽出であると考えてよい。
そこで n 回の復元抽出において1回銀のエンゼルマークが当たったとすると
$f(p)={}_nC_1\,p(1-p)^{n-1}=n\,p(1-p)^{n-1}$ この関数が $0<p<1$ において最大値をとるような確率 p を求める。 $f(x)=n\,x(1-x)^{n-1}$ $(0<x<1)$ とおいて微分すると、
$f'(x)=n(1-n\,x)(1-x)^{n-2}$ これより、 $f(x)$ は $x=\frac{1}{n}$ において最大値をとる。（最尤推定値）
したがって、 $f(p)$ はは $p=\frac{1}{n}$ において最大値をとる。 $n=28$ であると考えて、 $p\approx\frac{1}{28}$ である。

（結論）
いずれの考え方でも銀のエンゼルマークは 28 枚に1枚は入っていると考えられ、これは当たる確率が $\frac{1}{30}$ であるという事実に近いので、うわさは本当と考えてよいと思われる。

（発展）

最後に、p の信頼区間を考えようと思ったが、標本の数が少ないため断念した。90％信頼区間の誤差を 0.01 以内で抑えるためには $p=\frac{1}{28}$ として標本数が $\left(\frac{1.65}{0.01}\right)^2\cdot\frac{1}{28}\cdot\left(1-\frac{1}{28}\right)=937.6$ 個必要であるためである。

数学A以外に数学Bや数学Ⅲの内容も含まれてしまったが、意欲のある高校生であれば十分学習できると思われる。また他のアプローチや実践例があれば是非教えていただきたいと思います。

（参考文献）

[1] 統計学入門　　木村　等　大藪和雄　石川　浩
[2] 確率　 渡部隆一
[3] 統計学のはなし　 蓑谷千凰彦
[4] 統計解析のはなし　 石村　貞夫