赤胴鈴之助問題に関する解答

(サイアスを引用した朝日新聞夕刊コラム 2000/8/9 より)
　昔、「赤胴鈴之助」の一文字ずつがおまけについた菓子があって、すべての文字をそろえようとしたが、なかなか全部は揃わなかった。さて、平均していくつくらい買えば全部揃うのか？

k 回目ではじめて５種類とも揃うとすると、当然 $k \geqq 5$ である。そこで、その確率を p_k とすると、"平均の回数" は、"k の期待値" を計算することで求められる。つまり、

$$E = \sum_{k=5}^{\infty} k \cdot p_k$$

ここで、

$$p_k = (\frac{4}{5})^{k-1} - 4(\frac{3}{5})^{k-1} + 6(\frac{2}{5})^{k-1} - (\frac{1}{5})^{k-1}$$

である (※１)。これを用いて計算すると (※２)、

$$E = \frac{137}{12} \fallingdotseq 11.42$$

となり、平均としては１１回程度で５種類とも揃う事がわかる。

(※１)
k 回ではじめて５種類とも揃うためには、$k-1$ 回目までは４種類しか出てはいけなく、しかもその４種類とも少なくとも１回は出ていないといけない。つまり、$k-1$ 回まで４種類すべてのみを揃え、k 回目でまだ揃っていない５種類目を揃えることになり、例えば $k-1$ 回までに「赤」以外の４種類すべてのみを揃える確率を q_{k-1}、その次に赤を手に入れて５種類揃う確率を $p_{k,\,赤}$ とすると、$p_{k,\,赤} = q_{k-1} \cdot (1/5)$ となる。
　この q_{k-1} を求める。最初の $k-1$ 回の間は残りの４つのみを出す確率は、

$$(\frac{4}{5})^{k-1}$$

である。但し、これでは $k-1$ 回の間、同じもの (例えば「鈴」) が出続ける確率なども含まれるので、ここから、(1) 同じ物のみが出続ける、(2) ２種類のみが出続ける、(3) ３種類のみが出続ける、という確率を差し引かなくてはならない。各々の出方に対する (1),(2),(3) の確率をそれぞれ r_1,r_2,r_3 とすると、

$$\begin{aligned}
r_1 &= (\frac{1}{5})^{k-1} \\
r_2 &= (\frac{2}{5})^{k-1} - r_1 \cdot {}_2C_1 \\
&= (\frac{2}{5})^{k-1} - 2 \cdot (\frac{1}{5})^{k-1} \\
r_3 &= (\frac{3}{5})^{k-1} - (r_1 \cdot {}_3C_1 + r_2 \cdot {}_3C_2) \\
&= (\frac{3}{5})^{k-1} - 3 \cdot (\frac{2}{5})^{k-1} + 3 \cdot (\frac{1}{5})^{k-1}
\end{aligned}$$

と求められる。r_3 での２項目以降の引き算は、１種類、２種類しか出ない場合の確率を、その各種類の取り合わせ方を掛けて引いているものである。r_2 も同様。ここで、${}_3C_1$ 等は２項係数である。以上から、

$$\begin{aligned} q_{k-1} &= (\frac{4}{5})^{k-1} - (r_1 \cdot {}_4C_1 + r_2 \cdot {}_4C_2 + r_3 \cdot {}_4C_3) \\ &= (\frac{4}{5})^{k-1} - 4 \cdot (\frac{3}{5})^{k-1} + 6 \cdot (\frac{2}{5})^{k-1} - 4 \cdot (\frac{1}{5})^{k-1} \end{aligned}$$

となる。$p_{k,\,赤} = q_{k-1} \cdot (1/5)$ であったので、

$$p_{k,\,赤} = \frac{(\frac{4}{5})^{k-1} - 4 \cdot (\frac{3}{5})^{k-1} + 6 \cdot (\frac{2}{5})^{k-1} - 4 \cdot (\frac{1}{5})^{k-1}}{5}$$

であるが、最後に手に入れるものは５種類あるので、結局、

$$\begin{aligned} p_k &= p_{k,\,赤} \cdot 5 \\ &= (\frac{4}{5})^{k-1} - 4 \cdot (\frac{3}{5})^{k-1} + 6 \cdot (\frac{2}{5})^{k-1} - 4 \cdot (\frac{1}{5})^{k-1} \end{aligned}$$

である。

(※２)

$$\begin{aligned} E &= \sum_{k=5}^{\infty} k \cdot p_k \\ &= \sum_{k=5}^{\infty} k \cdot [(\frac{4}{5})^{k-1} - 4 \cdot (\frac{3}{5})^{k-1} + 6 \cdot (\frac{2}{5})^{k-1} - 4 \cdot (\frac{1}{5})^{k-1}] \end{aligned}$$

であるから、

$$\sum_{k=5}^{\infty} k \cdot (\frac{j}{5})^{k-1}$$

の形の和を求める必要がある。ここで、$j = 1, 2, 3, 4$ として、

$$S_{j,n} \equiv \sum_{k=5}^{n} k \cdot (\frac{j}{5})^{k-1}$$

とすると、

$$\begin{aligned} S_{j,n} &= (\frac{j}{5})^4 \cdot 5 + (\frac{j}{5})^5 \cdot 6 + \cdots + (\frac{j}{5})^{n-1} \cdot n \\ \frac{j}{5} \cdot S_{j,n} &= \qquad (\frac{j}{5})^5 \cdot 5 + \cdots + (\frac{j}{5})^{n-1} \cdot (n-1) + (\frac{j}{5})^n \cdot n \\ (1 - \frac{j}{5}) S_{j,n} &= (\frac{j}{5})^4 \cdot 5 + [(\frac{j}{5})^5 + \cdots + (\frac{j}{5})^{n-1}] - (\frac{j}{5})^n \cdot n \\ &= (\frac{j}{5})^4 \cdot 5 + \frac{(\frac{j}{5})^5 [1 - (\frac{j}{5})^{n-5}]}{1 - \frac{j}{5}} - (\frac{j}{5})^n \cdot n \\ &\longrightarrow (\frac{j}{5})^4 \cdot 5 + \frac{(\frac{j}{5})^5}{1 - \frac{j}{5}} \qquad (n \longrightarrow \infty) \end{aligned}$$

$$\therefore \lim_{n\to\infty} S_{j,n} = \frac{5(25-4j)}{(5-j)^2}(\frac{j}{5})^4$$

となる。これを使うと、

$$\begin{aligned}
E &= \sum_{k=5}^{\infty} k\cdot(\frac{4}{5})^{k-1} - 4\sum_{k=5}^{\infty} k\cdot(\frac{3}{5})^{k-1} + 6\sum_{k=5}^{\infty} k\cdot(\frac{2}{5})^{k-1} - 4\sum_{k=5}^{\infty} k\cdot(\frac{1}{5})^{k-1} \\
&= \frac{5(25-16)}{(5-4)^2}(\frac{4}{5})^4 - \frac{4\cdot5(25-12)}{(5-3)^2}(\frac{3}{5})^4 + \frac{6\cdot5(25-8)}{(5-2)^2}(\frac{2}{5})^4 \\
&\quad -\frac{4\cdot5(25-4)}{(5-1)^2}(\frac{1}{5})^4 \\
&= \frac{5(9\cdot4^4 - 13\cdot3^4 + \frac{34}{3}\cdot2^4 - \frac{21}{4}\cdot1^4)}{5^4} \\
&= \frac{2304 - 1053 + \frac{544}{3} - \frac{21}{4}}{125} \\
&= \frac{17125}{12\cdot125} \\
&= \frac{137}{12} \\
&= 11.41666\ldots
\end{aligned}$$

となり、平均として、約１１回買えば、５種類すべてが手に入ることになる事がわかる。

すべて、高校で学習する範囲の確率で計算が可能です。

ただし、５種類程度であればこのように計算ができるが、もっと大きな数になると、実際の p_k の計算が煩雑すぎて、この方法では実質的に不可能でしょう (参考３参照)。

この問題は、“クーポン収集問題” として有名な問題だそうです。

(参考１)　一般的な場合への拡張

一般に、箱の中に $1\sim n$ の n 個の異なる球があるときに、１つ取ってその番号を記録してから戻すという試行を、$1\sim n$ のすべての番号が少なくとも１度記録されるまで繰り返すとするとき、その繰り返し回数の平均を求めたい。但し、どの球を引く確率も同様に確からしいとする。

上と同じように p_k を定めると、その定義から、

$$p_k = \sum_{j=1}^{n-1} a_{j,n-1}\cdot(\frac{j}{n})^{k-1}$$

と書くことができるはずである。このとき、

$$\begin{aligned}
E_n &= \sum_{k=n}^{\infty}\{\sum_{j=1}^{n-1} a_{j,n-1}\cdot(\frac{j}{n})^{k-1}\cdot k\} \\
&= \lim_{N\to\infty}\sum_{k=n}^{N} k\{\sum_{j=1}^{n-1} a_{j,n-1}\cdot(\frac{j}{n})^{k-1}\} \\
&= \sum_{j=1}^{n-1}\{\lim_{N\to\infty}\sum_{k=n}^{N} a_{j,n-1}\cdot(\frac{j}{n})^{k-1}k\} \quad \text{(各項の収束性より)} \\
&= \sum_{j=1}^{n-1} a_{j,n-1}\{\lim_{N\to\infty}\sum_{k=n}^{N}(\frac{j}{n})^{k-1}k\}
\end{aligned}$$

ここで、

$$S_N \equiv \sum_{k=n}^{N} k(\frac{j}{n})^{k-1}$$

とすると、

$$E_n = \sum_{j=1}^{n-1} a_{j,n-1}\{\lim_{N\to\infty} S_N\}$$

であり、前半と同様の方法で S_∞ を求めることができる。その結果、

$$\begin{aligned} S_\infty &= \lim_{N\to\infty} S_N \\ &= \frac{n(n^2-nj+j)}{(n-j)^2}(\frac{j}{n})^{n-1} \end{aligned}$$

となり、

$$E_n = \sum_{j=1}^{n-1} a_{j,n-1}\frac{n(n^2-nj+j)}{(n-j)^2}(\frac{j}{n})^{n-1}$$

となる。ここで、実は、

$$a_{j,n-1} = (-1)^{n-1-j}\cdot {}_{n-1}C_j$$

である (※３)。これより、

$$E_n = \sum_{j=1}^{n-1}(-1)^{n-1-j}\cdot {}_{n-1}C_j\frac{n(n^2-nj+j)}{(n-j)^2}(\frac{j}{n})^{n-1}$$

から、E_n を数値的に計算することができる。さらに、この和に対し、

$$\begin{aligned} F_n \equiv \frac{1}{n}\cdot E_n &= \sum_{j=1}^{n}\frac{1}{j} \\ &= \int_0^1 \frac{1-(1-t)^n}{t}dt \end{aligned}$$

である (参考２を参照)。この和は収束しない。

(※３)

２より大きいある自然数 n に対し、k 回目まで ($k \geqq 1$) に、特定の t 種類しか ($t \leqq n-1,\ t \leqq k$) 引かない確率を $p_{n,k}(t)$ とすると、

$$p_{n,k}(t) = \sum_{j=1}^{t}(\frac{j}{n})^k(-1)^{t-j}{}_tC_j \quad \cdots\cdots \quad (A)$$

と書くことができる。特に、$t=n-1$ のとき、

$$p_{n,k}(n-1) = \sum_{j=1}^{n-1}(\frac{j}{n})^k(-1)^{n-1-j}{}_{n-1}C_j \quad \cdots\cdots \quad (B)$$

[証明] n,k を固定し、数学的帰納法で証明する。
(i)$t=1$ のとき、明らか。
(ii)$t \leqq T-1$ のとき、(A) 式が成り立つとすると $(1 \leqq T-1 \leqq n-2)$、$t=T$ のとき、

$$\begin{aligned}
p_{n,k}(T) &= (\frac{T}{n})^k - \sum_{s=1}^{T-1} {}_TC_s \cdot p_{n,k}(s) \\
&= (\frac{T}{n})^k - \sum_{s=1}^{T-1}\{{}_TC_s \sum_{j=1}^{s}(\frac{j}{n})^k \cdot (-1)^{s-j} \cdot {}_sC_j\} \\
&= (\frac{T}{n})^k + \sum_{s=1}^{T-1}\sum_{j=1}^{s}(\frac{j}{n})^k \cdot (-1)^{s-j+1} \cdot {}_TC_s \cdot {}_sC_j \\
&= (\frac{T}{n})^k + \sum_{j=1}^{T-1}\sum_{s=j}^{T-1}(\frac{j}{n})^k \cdot (-1)^{s-j+1} \cdot {}_TC_s \cdot {}_sC_j \\
&= (\frac{T}{n})^k + \sum_{j=1}^{T-1}(\frac{j}{n})^k\{\sum_{s=j}^{T-1}(-1)^{s-j+1} \cdot {}_TC_s \cdot {}_sC_j\} \\
&= (\frac{T}{n})^k + \sum_{j=1}^{T-1}(\frac{j}{n})^k\{\sum_{s=0}^{T-j-1}(-1)^{s+1}{}_TC_{s+j} \cdot {}_{s+j}C_j\} \\
&= (\frac{T}{n})^k + \sum_{j=1}^{T-1}(\frac{j}{n})^k\{\sum_{s=0}^{T-j-1}(-1)^{s+1}\frac{T!}{(s+j)!(T-s-j)!}\frac{(s+j)!}{j!s!}\} \\
&= (\frac{T}{n})^k + \sum_{j=1}^{T-1}(\frac{j}{n})^k\{\sum_{s=0}^{T-j-1}(-1)^{s+1}\frac{T!}{j!(T-j)!}\frac{(T-j)!}{(T-s-j)!s!}\} \\
&= (\frac{T}{n})^k + \sum_{j=1}^{T-1}(\frac{j}{n})^k\{\sum_{s=0}^{T-j-1}(-1)^{s+1}{}_{T-j}C_s \cdot {}_TC_j\} \\
&= (\frac{T}{n})^k + \sum_{j=1}^{T-1}(\frac{j}{n})^k{}_TC_j\{\sum_{s=0}^{T-j-1}(-1)^{s+1}{}_{T-j}C_s\} \\
&= (\frac{T}{n})^k + \sum_{j=1}^{T-1}(\frac{j}{n})^k{}_TC_j\{-(-1)^{T-j-1}{}_{T-j-1}C_{T-j-1}\} \\
&= (\frac{T}{n})^k + \sum_{j=1}^{T-1}\{(\frac{j}{n})^k{}_TC_j(-1)^{T-j}\} \\
&= \sum_{j=1}^{T}(\frac{j}{n})^k(-1)^{T-j}{}_TC_j
\end{aligned}$$

となり、$t=T$ のときも成立する。
以上より、命題 (A) は証明された。命題 (B) は (A) の特別な場合であり、$t=n-1$ として成立する。[証明終了]

(参考２)　別解
　参考１と同じ状況を考える。

n 種類の数を揃えるということは、1 種類ずつ数を増やしていくわけなので、$j-1$ 種類目の数を記録した後、j 種類目の数を手に入れるのに要する回数を $N(j)$ とすると、n 種類の球を手に入れるために必要な回数 N は、

$$\begin{aligned} N &= \sum_{j=1}^{n} N(j) \\ &= N(1) + N(2) + \cdots + N(n) \end{aligned}$$

で表される。

$j-1$ 種類を既に集めているとき ($j \leqq n$)、j 種類目を引くのに要する平均の回数 (つまり $N(j)$ の平均) を求める。k 回目でそれを引くとすると、その確率は

$$(1-u_j)^{k-1} \cdot u_j$$

である。但し、u_j は、1 回引いたときに今までに引いていない種類を引く確率である。すなわち、

$$u_j = \frac{n-(j-1)}{n}$$

よって、求める平均の回数 N_j は、

$$\begin{aligned} N_j = \overline{N(j)} &= \sum_{k=1}^{\infty} k \cdot [(1-u_j)^{k-1} \cdot u_j] \\ &= \sum_{k=1}^{\infty} k \cdot \{(\frac{j-1}{n})^{k-1} \frac{n-j+1}{n}\} \\ &= \sum_{k=1}^{\infty} \{k(\frac{j-1}{n})^{k-1} - k(\frac{j-1}{n})^{k}\} \\ &= \frac{1}{(1-\frac{j-1}{n})^2} - \frac{\frac{j-1}{n}}{(1-\frac{j-1}{n})^2} \\ &= \frac{\frac{n-j+1}{n}}{(\frac{n-j+1}{n})^2} \\ &= \frac{n}{n-j+1} \end{aligned}$$

さて、j 種類の数を引くそれぞれの事象は独立事象であるから、最初から n 種類目の数を記録するまでの平均の試行数 E_n、つまり N の平均は、

$$E_n = \overline{N} = \sum_{j=1}^{n} \overline{N(j)} = \sum_{j=1}^{n} N_j$$

と表される。よって、

$$\begin{aligned} E_n &= \sum_{j=1}^{n} \frac{n}{n-j+1} \\ &= n \cdot \sum_{j=1}^{n} \frac{1}{j} \end{aligned}$$

となる。最初の問題では、$n = 5$ であるので、

$$E_5 = 5 \cdot (1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}) = \frac{137}{12} = 11.4166\ldots$$

となり、当初の計算に一致する。この方がシンプルかもしれません。
　なお、

$$\sum_{j=1}^{n-1} (-1)^{n-1-j} \cdot {}_nC_j \cdot \{1 + \frac{j}{n(n-j)}\}(\frac{j}{n})^{n-1} = \sum_{j=1}^{n} \frac{1}{j}$$

$$\sum_{j=1}^{n-1} (-1)^{n-1-j} \cdot {}_nC_j \cdot (\frac{j}{n})^{n-1} = 1$$

$$\left(\sum_{j=0}^{n} (-1)^{n-1-j} \cdot {}_nC_j \cdot (\frac{j}{n})^{n-1} = 0 \right)$$

であることが解る。

(参考３)　各 n に対する数値

n	E_n			比率 (E_n/n)
1	1	$=$	1.00	1.00
2	3	$=$	3.00	1.50
3	11/2	$=$	5.50	1.83
4	25/3	≒	8.33	2.08
5	137/12	≒	11.42	2.28
6	147/10	$=$	14.70	2.45
7	363/20	$=$	18.15	2.59
8	761/35	≒	21.74	2.72
9	7129/280	≒	25.46	2.83
10	7381/252	≒	29.29	2.93
11	83711/2520	≒	33.22	3.02
12	86021/2310	≒	37.24	3.10
13	1145993/27720	≒	41.34	3.18
14	1171733/25740	≒	45.52	3.25
15	1195757/24024	≒	49.77	3.32
16	2436559/45045	≒	54.09	3.38
17	42142223/720720	≒	58.47	3.44
18	42822903/680680	≒	62.91	3.50
19	275295799/4084080	≒	67.41	3.55
20			71.95	3.60
25			95.40	3.82
30			119.85	3.99
50			224.96	4.50
100			518.74	5.19

(参考４)　１種類のみが現れにくい場合

１種類のみが現れにくい場合、その「全種類入手に必要な平均の枚数」を計算することは非常に困難である (と思う/少なくとも私には出来なかった)。そのような場合でも、シミュレーションによって求めることが出来る。10^7 回の試行の結果、例えば「鈴」の文字が他のカードに比べて $\frac{1}{5}$ しかない場合、平均して 22.97 回の購入が必要であることがわかる。全部が同じ割合のときの約２倍ですね。

以下に、カードの種類 n のうち、１枚だけが他に比べて $\frac{1}{m}$ の割合でしか入っていないときの、全種類入手に必要な平均の購入回数を示す。各々、10^7 回の試行を行ったときの平均である。$m = 1$ のときが、全種類が同じ確率のときのもの (つまり先に示したもの) である。

カードの割合を変えることによって、全種類を揃えることが容易ではない状況を簡単に作ることが出来ることがわかると思います。

	m									
n	1	2	3	4	5	6	7	8	9	10
2	3.00	3.50	4.33	5.25	6.20	7.17	8.14	9.12	10.11	11.10
3	5.50	6.42	8.00	9.78	11.64	13.54	15.46	17.41	19.36	21.30
4	8.33	9.62	11.90	14.49	17.23	20.04	22.92	25.82	28.72	31.65
5	11.42	13.03	15.98	19.36	22.97	26.68	30.46	34.30	38.18	42.06
6	14.70	16.63	20.20	24.36	28.81	33.42	38.10	42.87	47.71	52.53
7	18.15	20.38	24.56	29.49	34.74	40.22	45.83	51.54	57.30	63.09
8	21.74	24.25	29.02	34.70	40.79	47.10	53.61	60.23	66.92	73.65
9	25.46	28.23	33.58	39.98	46.85	54.07	61.46	68.98	76.62	84.30
10	29.29	32.33	38.24	45.34	53.04	61.07	69.39	77.82	86.36	94.99

補足：サイアス10月号を読んで

サイアス10月号には、寄せられた各種解答が載せられていた。幾つかをフォローする。

計算方法について

$E_n = \frac{n}{n} + \frac{n-1}{n} + \cdots + \frac{1}{n}$ に関し、私は、1～n 種類目に関してそれぞれを得るための所要回数の期待値を求めてその和を計算したが、その各期待値を求める際の方法として、もっと直感的なものがあった。すなわち、$j-1$ 種類を既に集めているとき ($j \leqq n$)、1枚引いて j 種類目を得る確率は、$u_j = \frac{n-(j-1)}{n}$ であることから、平均すれば $\frac{1}{u_j} = \frac{n}{n-(j-1)}$ 回ひくことで j 種類目を得ることができる。これから、先と同様に和を考えて、E_n を得る。

さて、実際にこの考え方での計算が正しいことを示す。一般に、確率 p である事象が起こるとき、その事象が1回起こるまでに必要な平均の試行回数は、上の考え方では $\frac{1}{p}$ であり、一方、それぞれの回数でその事象がはじめて起こる確率から求めると、

$$\sum_{n=1}^{\infty} np(1-p)^{n-1}$$

である。これを計算すると、

$$\sum_{n=1}^{\infty} np(1-p)^{n-1} = \frac{p}{1-p}\sum_{n=1}^{\infty} n(1-p)^n = \frac{p}{1-p}\cdot\frac{1-p}{p^2} = \frac{1}{p}$$

となって、確かに一致する。

1種類のみが現れにくい場合の考察

問題のケース、つまり、5種類のうち1種類のみが他に比べて $\frac{1}{5}$ の割合でしかない場合、以下のように答えが示されていた。以下、少ないものを「鈴」とする。

j 種類目 ($j = 1$～5) が「鈴」の場合に分けて考える。それぞれの場合の期待値は、1～5種類目の新しいカードが出るための期待値の和を考え、それに j 種類目で「鈴」が出る確率を掛けたものである。つまり、

$$\begin{aligned}
j=1 &:\ (1+\frac{21}{20}+\frac{21}{15}+\frac{21}{10}+\frac{21}{5})\times\frac{1}{21} \\
j=2 &:\ (1+\frac{21}{16}+\frac{21}{15}+\frac{21}{10}+\frac{21}{5})\times\frac{20}{21}\cdot\frac{1}{16} \\
j=3 &:\ (1+\frac{21}{16}+\frac{21}{11}+\frac{21}{10}+\frac{21}{5})\times\frac{20}{21}\cdot\frac{15}{16}\cdot\frac{1}{11} \\
j=4 &:\ (1+\frac{21}{16}+\frac{21}{11}+\frac{21}{6}+\frac{21}{5})\times\frac{20}{21}\cdot\frac{15}{16}\cdot\frac{10}{11}\cdot\frac{1}{6} \\
j=5 &:\ (1+\frac{21}{16}+\frac{21}{11}+\frac{21}{6}+\frac{21}{1})\times\frac{20}{21}\cdot\frac{15}{16}\cdot\frac{10}{11}\cdot\frac{5}{6}\cdot\frac{1}{1}
\end{aligned}$$

である。ここで、各期待値の計算には上で述べた“1/確率”の方法を用いた。

これを用いて計算すると、この和は

$$\frac{505}{22} = 22 + \frac{21}{22} = 22.9545454\ldots$$

となる。シミュレーションとはちょっと違う結果になってしまう。

そこで、10^6 回のシミュレーションを行って期待値を求める、という操作を 10^4 回行うと、それらの期待値の平均は 22.9614 となり、標準偏差は 0.0186 になったことから、誤差の範囲ではないかとも考えられる。

さらに、一般の場合への拡張を考える。

先の考え方を踏襲すると、1∼n の n 種類の数が書かれた球があり、そのうち 1 種類だけ (例えば「1」) が他に比べて $\frac{1}{m}$ の確率でしか存在しないとすると、j 種類目 $(1 \leqq j \leqq n)$ に「1」の球を引くとするときの期待値は、

$$\left[\sum_{i=1}^{j}\frac{(n-1)m+1}{(n-i)m+1}+\sum_{i=j+1}^{n}\frac{(n-1)m+1}{(n-i+1)m}\right]\cdot\left(\prod_{i=1}^{j-1}\frac{(n-i)m}{(n-i)m+1}\right)\frac{1}{(n-j)m+1}$$

となる。よって、期待値 $E_{n,m}$ は、

$$\begin{aligned}
& E_{n,m} \\
= & \ [(n-1)m+1] \\
& \times\sum_{j=1}^{n}\left[\frac{1}{(n-j)m+1}\left\{\sum_{i=1}^{j}\frac{1}{(n-i)m+1}+\sum_{i=j+1}^{n}\frac{1}{(n-i+1)m}\right\}\prod_{i=1}^{j-1}\frac{(n-i)m}{(n-i)m+1}\right] \\
= & \ \frac{(n-1)m+1}{m} \\
& \times\left[(m-1)\sum_{j=1}^{n}\left\{\frac{1}{\{(n-j)m+1\}(n-j+1)}\prod_{i=1}^{j-1}\frac{(n-i)m}{(n-i)m+1}\right\}+\sum_{j=1}^{n}\frac{1}{j}\right]
\end{aligned}$$

となる。これに $m=1$ を代入して整理すると、確かに E_n と同じ式が得られる。

また、m を変えて計算を行うと、

n	m = 1	2	3	4	5	6	7	8	9	10
2	3.00	3.50	4.33	5.25	6.20	7.17	8.14	9.13	10.11	11.10
3	5.50	6.42	8.00	9.78	11.63	13.54	15.46	17.41	19.37	21.33
4	8.33	9.62	11.90	14.49	17.23	20.05	22.92	25.81	28.72	31.66
5	11.42	13.03	15.97	19.36	22.95	26.67	30.46	34.29	38.16	42.06
6	14.70	16.62	20.19	24.34	28.78	33.38	38.09	42.85	47.67	52.52
7	18.15	20.36	24.53	29.43	34.70	40.18	45.78	51.47	57.23	63.03
8	21.74	24.22	28.98	34.61	40.70	47.04	53.54	60.15	66.84	73.59
9	25.46	28.19	33.51	39.87	46.76	53.96	61.36	68.98	76.50	84.19
10	29.29	32.27	38.14	45.21	52.89	60.94	69.22	77.65	86.20	94.83

という結果が得られる。

これをシミュレーションの結果と比較すると、ほぼよく似た値であることがわかるが、シミュレーションのほうが若干大きな値である (最大で約 0.3%)。$m=1$ 及び $n=2$ の値は正しい。常に大きな値が得られるということは、何らかの差異が実際にあると思われる。特に、計算値よりも大きくなることが多いのはなぜだろうか？

「サイアス」誌の考え方で正しいかどうかということであるが、問題はないであろう。一方、シミュレーションの方法を何度か見直したが、プログラムに問題はないように思われる。乱数の“乱雑さ”も確認したが、問題はなかった。

この原因は全く分からない。

```
#include <stdio.h >
#include <stdlib.h >
#include <time.h >
#include <math.h >

#define dMAXRAND 0x7fffffff
#define N_NUM 10 /* number of n */
#define M_NUM 10 /* number of m */
#define N_BGN 2 /* start number of n */
#define M_BGN 1 /* start number of m */
#define S_TIMES 100000
#define L_TIMES 10
#define N_TIMES ( S_TIMES * L_TIMES ) /* num of trial :  not to overflow */
#define REPNUM 10 /* repeat times for each (n,m) */

int main()
{
     long int val, j, ii, k, m, jj, n, base, r[N_NUM];
     unsigned long int i, sum, sum2, ulim, ulim2;
     char done[N_NUM];
     double average, sd, dsum, dsum2;

     srandom(time(NULL));
     ulim = dMAXRAND;
     ulim++;
     for( n=N_BGN ; n<=N_NUM ; n++ ) {
          for( m=M_BGN ; m<=M_NUM ; m++ ) {
               base = (n-1)*m+1;
               ulim2 = ulim - ulim % base;
               for( ii=1 ; ii<n ; ii++ ) r[ii-1] = (ulim2 / base) * m * ii;
               for( k=0 ; k<REPNUM ; k++ ) {
                    dsum = dsum2 = 0.0;
                    /* Be carefull not to overflow sum and sum2 !!  */
                    for( jj=0 ; jj<L_TIMES ; jj++ ) {
                         sum = sum2 = 0;
                         for( j=0 ; j<S_TIMES ; j++ ) {
                              for( i=0 ; i<n ; i++ ) done[i] = 0;
                              i = 0;
                              do {
                                   if( (val = random()) >= ulim2 ) continue;
                                   i++;
                                   for( ii=0 ; ii<n-1 ; ii++ ) {
                                        if( val <r[ii] ) {
                                             done[ii] = 1;
                                             break;
                                        }
                                   }
                                   if( ii == n-1 ) done[n-1] = 1;
                                   for( ii=0 ; ii<n ; ii++ ) if( !done[ii] ) break;
                                   if( ii == n ) break;
                              } while(1);
                              sum += i;
                              sum2 += i*i;
                         }
                         dsum += (double)sum;
                         dsum2 += (double)sum2;
                    }
                    average = dsum / N_TIMES;
                    sd = sqrt( dsum2 / N_TIMES - average*average );
                    fprintf(stdout,"Rep.  %d trials, n = %d, m = %d (%d), :  Ave.  = %f, SD = %f.
¥ n",N_TIMES,n,m,k,average,sd);
               }
          }
     }
     return(0);
}
```