モンモールの問題について

1　どのような問題を考えるか

今年度の東北大学前期試験数学の問題で 2 つの袋からカードを取り出していって、カードが一致していた枚数の期待値を問う問題が出題されていた．この問題は、いわゆるモンモールの問題と呼ばれる問題と関係している問題である．これについて、簡単に解説したい．カードの一致した枚数と書いたが、どのような問題かと言うと

問題 1 N 枚のカードが袋 A,B に入っている．N 回、順番にカードを取り出していき, カードが一致していた枚数を X とする．ただし, 取り出したカードは袋に戻さないとする．X の期待値は 1 である．

東北大の問題は $N=4$ の場合であった．ここで, 確率変数の期待値の定義を復習する．

定義 1 確率変数 X の取りうる値が $\{x_1,\ldots,x_n\}$ で $X=x_k$ となる確率が p_k のとき,

$$E[X]=\sum_{k=1}^{n}x_kp_k.$$

上記問題において $X=k$ $(0\leq k\leq N)$ となる確率 p_k を計算し, 期待値の定義に従って, 計算することもできるが, 一般の場合 p_k の計算はそう簡単ではない[1]．X を"ある確率変数の和"の形に表し, "**期待値の線形性** "を使うと簡単に計算できることがよく知られている．期待値の線形性とは次の内容である．

定理 1 X,Y を確率変数, a,b を実数の定数とすると

$$E[aX+bY]=aE[X]+bE[Y].$$

一方復元抽出で考えても結果は同じである．すなわち,

問題 2 N 枚のカードが袋 A,B に入っているとする．N 回, 順番にカードを取り出し、カードが一致していた枚数を X とする．ただし取り出したカードは毎回袋に戻すとする。このとき, X の期待値は 1 である．

問題 2 の復元抽出の場合 $X=k$ となる確率の計算, $\sum_{k=0}^{N}kp_k$ の計算も難しくない．しかし, 期待値の線形性を使う方が計算は簡単である．以下 X をどのような確率変数の和で表現し期待値を計算するか説明しよう．

[1] p_0(1 回もカードが一致しない確率) を求める問題がモンモールの問題である．

2　期待値の線形性を用いた解法

問題 1 の $E[X]$ の計算
　X_i を次のような確率変数とする:

$$X_i = \begin{cases} 1 & (i \text{ 回目の試行でカードが一致したとき}) \\ 0 & (i \text{ 回目の試行でカードが一致しなかったとき}) \end{cases}$$

すると $X = \sum_{i=1}^{N} X_i$ である．ここで

$$P(X_i = 1) = \frac{\{(N-1)!\}^2 N}{(N!)^2} = \frac{1}{N}.$$

したがって $E[X_i] = 1 \times \dfrac{1}{N} = \dfrac{1}{N}$ となり期待値の線形性から

$$E[X] = \sum_{i=1}^{N} E[X_i] = N \times \frac{1}{N} = 1.$$

■

問題 2 の $E[X]$ の計算
　問題 1 の証明と同じ確率変数 X_i を考える．このとき

$$P(X_i = 1) = \frac{N}{N^2} = \frac{1}{N}.$$

従って上記計算と同様にして, $E[X] = \displaystyle\sum_{i=1}^{N} E[X_i] = N \times 1/N = 1.$ ■

注意 1　今述べた 2 つの証明に現れた X_i についていずれも $X_i = 1$ となる確率は等しく $1/N$ になったが,

(i)　問題 1 では $X_1, \ldots, X_n$ は独立ではない

(ii)　問題 2 では $X_1, \ldots, X_n$ は独立

のように当然ながら結合分布は全く違う性質を持つ．実際, 問題 1 では $X_1 = \cdots = X_{n-1} = 1$ ならば $X_n = 1$ である．

練習問題 1　問題 1, 問題 2 それぞれにおいて共分散 $\mathrm{Cov}(X_i, X_j) = E[(X_i - E[X_i])(X_j - E[X_j])]$, および分散 $V[X] = E[(X - E[X])^2]$ を求めてみよ．

　次の問題はやはり一致している枚数をうまく確率変数の和で表現すると簡単に解ける問題である．

練習問題 2　袋 A,B それぞれに 1 から N までの数字が一つずつ書かれたカードが N 枚入っている．
　A,B のそれぞれから k 枚のカードを無作為に選んだとき,
　数字が一致しているカードの枚数の期待値を求めよ．

この問題では、以下の場合は簡単にわかる.

(1) $k=N$ ならすべて取り出すので, 必ず N 枚一致する. したがって $E[X]=N$.

(2) $k=1$ なら 1 枚しか一致する可能性は無いがその確率は $1/N$. 従って, $E[X]=\frac{1}{N}$.

一般の k の場合どうなるかということである.

以上, 復元抽出, 非復元抽出の場合いずれでもカードの一致する枚数の期待値は 1 であった。すると次のような疑問が湧いて来る.

発展問題 1 A, B から N 回カードを取り出すが、各試行においてある確率 p で袋に戻す, 確率 $1-p$ で戻さないとすると期待値は 1 で無くなるか？

あるいは、あらかじめ定めた回の試行のみ袋に戻すことにするとどうか.

3 モンモールの問題, $E[X]=1$ の別証明

カードを毎回袋に戻す復元抽出の場合,

$$P(X=k)={}_N\mathrm{C}_k\left(\frac{1}{N}\right)^k\left(1-\frac{1}{N}\right)^{N-k}.$$

つまり X の分布は二項分布 $B(N,1/N)$ に従うわけで, 期待値の定義に基づいて $E[X]$ を計算することも困難ではない.

練習問題 3 上記の p_k の値を用いて、定義に従って $E[X]$ を計算せよ.

非復元抽出でカードを取り出す場合, $X=k$ となる確率 (k 枚のカードが一致する確率) を求めることは, 二項分布の場合ほど簡単ではないが, 可能である. 特に一致が 1 回も起こらない確率を求める問題はモンモール[2]の問題と呼ばれ, よく知られている.

少しこれまでの設定を変えて問題を述べることにする.

問題 3 1 から n の数字が一つずつ書かれたカード n 枚をよくかき混ぜて横 1 列に左から並べる. 数字 k のカードが左から k 番目にある回数 (これを ”一致の回数” と呼ぼう) を X とするとき,
(1) $X=0$ となる確率 $q(n)$ を求めよ.
(2) $E[X]$ を求めよ.

上記の問題で $q(n)$ を求めるのがモンモールの問題である. さて, n 枚のカードを横 1 列に並べる時, 一致が 1 回も起こらないような並べ方の総数を $a(n)$ と書くことにする.

$$q(n)=\frac{a(n)}{n!}$$

である. 以下

[2] 17-18 世紀の人

(1)　$a(n)$ の間に成立する関係式から $E[X]=1$ を示す[3].

(2)　$a(n), q(n)$ を求める

ことを試みる．

N 枚のカードを横 1 列に並べたとき一致する枚数を X としたとき $p_k = P(X=k)$ を $a(n)$ を用いて表すことができる．すなわち

$$p_k = \frac{{}_N\mathrm{C}_k \cdot a(N-k)}{N!}. \tag{1}$$

また,

$$N! = \sum_{k=0}^{N} {}_N\mathrm{C}_k \cdot a(N-k). \tag{2}$$

ただし $a(0)=1$ とおく．

$E[X]$ の定義から

$$E[X] = \sum_{k=0}^{N} kp_k = \sum_{k=0}^{N} k \cdot \frac{{}_N\mathrm{C}_k \cdot a(N-k)}{N!}$$

となる．

$$N! = \sum_{k=1}^{N} k \cdot {}_N\mathrm{C}_k \cdot a(N-k) \tag{3}$$

を示せば、$E[X]=1$ が示せたことになる．関係式 (2) を用いて (3) を示すことができる．

(3) の証明 (3) の右辺を変形する．

$$\begin{aligned}
\sum_{k=1}^{N} k \cdot {}_N\mathrm{C}_k \cdot a(N-k) &= N \sum_{k=1}^{N} \frac{(N-1)!}{(k-1)!(N-k)!} a(N-k) \\
&= N \sum_{k=1}^{N} {}_{N-1}\mathrm{C}_{N-k} \cdot a(N-k) \\
&= N \sum_{l=0}^{N-1} {}_{N-1}\mathrm{C}_l \cdot a(l)
\end{aligned}$$

したがって

$$(N-1)! = \sum_{l=0}^{N-1} {}_{N-1}\mathrm{C}_{N-1-l} \cdot a(l)$$

を示せばよい．これは関係式 (2) で述べている内容である．　■

[3] この方法では, $a(n), q(n)$ の具体的な値を求めなくても $E[X]=1$ だけはわかる．しかし $E[X]=1$ がわかって無いとこのような計算をしようとは思わないかもしれない．

4　$a(n)$, $q(n)$ の求め方

次に $a(n), q(n)$ を具体的に求めてみる.

補題 1 (1) $a(1) = 0, a(2) = 1, a(3) = 2$.
(2) $a(0) = 1$ とすると次の漸化式が成立する.

$$a(n) = \sum_{k=2}^{n} {}_{n-1}\mathrm{C}_{k-1} \cdot (k-1)! \cdot a(n-k), \quad (n = 2, 3, \ldots).$$

証明 (2) n 枚のカードの並べ方で 1 回も一致が起こっていない並べ方

$$A_1, A_2, \ldots, A_n$$

を考える. $f(i) = A_i$ という $\{1, 2, \ldots, n\}$ から $\{1, 2, \ldots, n\}$ への写像を考える. ここで $f^k(1) = 1$ となる最小の自然数 k を考える. ただし f^k は f を k 回合成した写像である. 1 回も一致が起こっていないので, $k \geq 2$ であり $k \leq n$ である.

(i)　$1, f(1), \ldots, f^{k-1}(1)$ はいずれも相異なる数である

(ii)　$C = \{1, \ldots, N\} \setminus \{1, f(1), \ldots, f^{k-1}(1)\}$ とおくと f は C から C への全単射であり, かつ一致は起こっていない.

ことに注意する. (i) の $f(1), \ldots, f^{k-1}(1)$ の選び方は ${}_{n-1}\mathrm{C}_{k-1} \cdot (k-1)!$ 通りあり, (ii) の C から C への一致が起こらない並べ方は $a(n-k)$ 通りあるので $f^k(1) = 1$ となる並べ方は全部で

$${}_{n-1}\mathrm{C}_{k-1} \cdot (k-1)! \cdot a(n-k)$$

通りある. これをすべて足して $a(n)$ が求まる. ■

例 1

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 1 & 3 \end{pmatrix}$$

のようになっているとき
$f(1) = 2, f^2(1) = 4, f^3(1) = 1,\ \ f : \{3, 5\} \to \{3, 5\}$
となっており $k = 3$ の場合である.

補題 2 (1) $q(1) = 0, q(2) = 1/2, q(3) = 1/3$.
(2) $q(0) = 1$ とすると

$$q(n) - q(n-1) = -\frac{1}{n}\left(q(n-1) - q(n-2)\right) \quad n \geq 2.$$

(3) $q(n) - q(n-1) = \dfrac{(-1)^n}{n!} \quad n \geq 1$.
(4) $q(n) = \dfrac{1}{2!} - \dfrac{1}{3!} + \cdots + \dfrac{(-1)^n}{n!} \qquad n \geq 2$.

証明

$$\begin{aligned} a(n) &= \sum_{k=2}^{n} {}_{n-1}\mathrm{C}_{k-1} \cdot (k-1)! \cdot a(n-k) \\ &= \sum_{k=2}^{n} \frac{(n-1)!}{(n-k)!} a(n-k) \qquad n \geq 2 \end{aligned}$$

の両辺を $n!$ で割って

$$\frac{a(n)}{n!} = \frac{1}{n} \sum_{k=2}^{n} \frac{a(n-k)}{(n-k)!} \qquad n \geq 2.$$

つまり

$$\begin{aligned} nq(n) &= \sum_{k=0}^{n-2} q(k) \quad (q(0) = 1 \text{ とする}), \\ nq(n) - (n-1)q(n-1) &= q(n-2) \quad (n \geq 2). \end{aligned}$$

したがって

$$\begin{aligned} & q(n) - q(n-1) \\ & \quad = \left(-\frac{1}{n}\right)(q(n-1) - q(n-2)) \quad (n \geq 2), \\ & \quad = \left(-\frac{1}{n}\right)\left(-\frac{1}{n-1}\right)(q(n-2) - q(n-3)) \\ & \quad = \cdots = \left(-\frac{1}{n}\right)\cdots\left(-\frac{1}{2}\right)(q(1) - q(0)) = \frac{(-1)^n}{n!}. \end{aligned}$$

■

　この結果から

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^k}{k!} + \cdots \qquad x \in \mathbb{R}$$

なので

$$\lim_{n\to\infty} q(n) = \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!} + \cdots = \frac{1}{e}.$$

5　練習問題の解答

(1) 練習問題 1 の解答

問題 2 では X_i, X_j $(i \neq j)$ は独立なので、簡単である．例えば $V[X] = 1 - \frac{1}{N}$.

問題 1 の設定で考えよう．まず $i \neq j$ のとき

$$P(X_i = X_j = 1) = \frac{2 {}_N\mathrm{C}_2 (N-2)! \times (N-2)!}{N!^2} = \frac{1}{N(N-1)}.$$

したがって $i \neq j$ のとき

$$\mathrm{Cov}(X_i, X_j) = E[(X_i - \frac{1}{N})(X_j - \frac{1}{N})] = E[X_i X_j] - \frac{1}{N}\left(E[X_i] + E[X_j]\right) + \frac{1}{N^2} = \frac{1}{N(N-1)} - \frac{2}{N^2} + \frac{1}{N^2}$$
$$= \frac{1}{N^2(N-1)}.$$

また

$$V[X_i] = E\left[\left(X_i - \frac{1}{N}\right)^2\right] = E\left[X_i^2\right] - \frac{1}{N^2} = \frac{N-1}{N^2}$$

$$V[X] = E\left[\left(\sum_{i=1}^{N} X_i - 1\right)^2\right] = \sum_{i=1}^{N} V[X_i] + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) = \frac{N(N-1)}{N^2} + \frac{N(N-1)}{N^2(N-1)} = 1.$$

(2) 練習問題 2 の解答

確率変数 X_i, Y_i $(1 \leq i \leq N)$ を次のように定める．

$$X_i = \begin{cases} 1 & (\text{カード } i \text{ が } A \text{ から取り出された時}) \\ 0 & (\text{カード } i \text{ が } A \text{ から取り出されなかった時}) \end{cases}$$

$$Y_i = \begin{cases} 1 & (\text{カード } i \text{ が } B \text{ から取り出された時}) \\ 0 & (\text{カード } i \text{ が } B \text{ から取り出されなかった時}) \end{cases}$$

と定めると

$$X = \sum_{i=1}^{N} X_i Y_i.$$

$E[X_i] = E[Y_i] = k/N$ かつ X_i と Y_i は独立だから

$$E[X] = \sum_{i=1}^{N} E[X_i Y_i] = \sum_{i=1}^{N} E[X_i] E[Y_i] = \sum_{i=1}^{N} \left(\frac{k}{N}\right)^2 = \frac{k^2}{N}.$$

(3) 練習問題 3 の解答

$$E[X] = \sum_{k=0}^{N} k P(X = k) = \sum_{k=1}^{N} k \cdot {}_N\mathrm{C}_k \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{N-k} = \sum_{k=1}^{N} N \cdot {}_{N-1}\mathrm{C}_{k-1} \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{N-k}$$
$$= \sum_{k=0}^{N-1} {}_{N-1}\mathrm{C}_k \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{N-k-1} = \left\{\frac{1}{N} + \left(1 - \frac{1}{N}\right)\right\}^{N-1} = 1.$$