# ２０２５年度　第１回　一橋大本番レベル模試

# ソーシャル・データサイエンス・総合問題

# 解答・採点基準

**全３問　６０分　１００点満点**

## Ⅰ　（４０点）

### 【解答・採点基準】

問 1

$N=8, M=3, n=4$ のとき，$X$ のとりうる値は $X=0, 1, 2, 3$ である。ここで，$X=k\ (k=0, 1, 2, 3)$ となる確率を考える。8 個の球から 4 個の球を取り出す場合の数は ${}_8\mathrm{C}_4$ 通りであり，$k$ 個を 3 個の赤球から選び，残りの $4-k$ 個を 5 個の白球から選ぶ場合の数は ${}_3\mathrm{C}_k \cdot {}_5\mathrm{C}_{4-k}$ 通りである。したがって，確率質量関数 $P(X=k)$ は

$$P(X=k)=\frac{{}_3\mathrm{C}_k \cdot {}_5\mathrm{C}_{4-k}}{{}_8\mathrm{C}_4}=\frac{{}_3\mathrm{C}_k \cdot {}_5\mathrm{C}_{4-k}}{70}$$

と表される。よって，$X$ の期待値 $E(X)$ は

$$\begin{aligned}E(X)&=0\cdot P(X=0)+1\cdot P(X=1)\\&\quad+2\cdot P(X=2)+3\cdot P(X=3)\\&=1\cdot\frac{{}_3\mathrm{C}_1\cdot{}_5\mathrm{C}_3}{70}+2\cdot\frac{{}_3\mathrm{C}_2\cdot{}_5\mathrm{C}_2}{70}+3\cdot\frac{{}_3\mathrm{C}_3\cdot{}_5\mathrm{C}_1}{70}\\&=1\cdot\frac{3}{7}+2\cdot\frac{3}{7}+3\cdot\frac{1}{14}\\&=\frac{3}{2}\end{aligned}$$

である。

問 1 [別解 1]

$N=8, M=3, n=4$ のとき，$X$ のとりうる値は $X=0, 1, 2, 3$ である。ここで，$X=k\ (k=0, 1, 2, 3)$ となる確率を考える。8 個の球から 4 個の球を取り出す場合の数は ${}_8\mathrm{C}_4$ 通りであり，$k$ 個

**問１　６点**

**問１［別解１］　６点**

を 3 個の赤球から選び，残りの $4-k$ 個を 5 個の白球から選ぶ場合の数は ${}_3\mathrm{C}_k \cdot {}_5\mathrm{C}_{4-k}$ 通りである。よって，その確率は

$$\begin{cases} X=0 \text{ のとき } \dfrac{{}_3\mathrm{C}_0 \cdot {}_5\mathrm{C}_4}{{}_8\mathrm{C}_4}=\dfrac{5}{70} \\ X=1 \text{ のとき } \dfrac{{}_3\mathrm{C}_1 \cdot {}_5\mathrm{C}_3}{{}_8\mathrm{C}_4}=\dfrac{30}{70} \\ X=2 \text{ のとき } \dfrac{{}_3\mathrm{C}_2 \cdot {}_5\mathrm{C}_2}{{}_8\mathrm{C}_4}=\dfrac{30}{70} \\ X=3 \text{ のとき } \dfrac{{}_3\mathrm{C}_3 \cdot {}_5\mathrm{C}_1}{{}_8\mathrm{C}_4}=\dfrac{5}{70} \end{cases}$$

となる。したがって，$X$ の期待値 $E(X)$ は

$$\begin{aligned} E(X) &= 0\cdot\frac{5}{70}+1\cdot\frac{30}{70}+2\cdot\frac{30}{70}+3\cdot\frac{5}{70} \\ &= \frac{3}{2} \end{aligned}$$

である。

問 1 [別解 2]

**問 1［別解 2］ 6 点**

確率変数 $Z_i$ $(i=1, 2, \ldots, 8)$ を，

$$\begin{cases} Z_i=1 & i \text{ 番目に赤球を取り出したとき} \\ Z_i=0 & i \text{ 番目に白球を取り出したとき} \end{cases}$$

とする。このとき，袋にある全ての球を順に取り出す状況を考えると，取り出される赤球の個数の期待値は 3 個であるから

$$E(Z_1)+E(Z_2)+\cdots+E(Z_8)=3$$

である。また，赤球が取り出される順番は無作為であるから

$$E(Z_1)=E(Z_2)=\cdots=E(Z_{8N})$$

である。したがって，$i$ の値によらず

$$E(Z_i)=\frac{3}{8}$$

である。また，$X=Z_1+Z_2+Z_3+Z_4$ であるから，求める確率は

$$\begin{aligned} E(X) &= E(Z_1)+E(Z_2)+E(Z_3)+E(Z_4) \\ &= \frac{3}{8}\cdot 4=\frac{3}{2} \end{aligned}$$

となる。

問 2(1)

**問 2(1) 16 点**

1個の球を取り出すたびに袋の中の球の数が減少し，また，赤球を取り出すか白球を取り出すかによっても，残りの赤球や白球の数が変わる。したがって，$i$ 番目で赤球を引く確率は $n$ に依存しないため，$E(Y_i)$ は $n$ に依存しない。したがって，$n=N$ と拡張し，袋にある全ての球を順に取り出す状況を考えてもよい。このとき，最初から $N$ 番目までの試行を通して考えると，取り出される赤球の個数の期待値は $M$ 個であるから

$$E(Y_1)+E(Y_2)+\cdots+E(Y_N)=M$$

である。また，赤球が取り出される順番は無作為であるから

$$E(Y_1)=E(Y_2)=\cdots=E(Y_N)$$

である。したがって，$i$ の値によらず

$$E(Y_i)=\frac{M}{N}$$

である。

問 2(2)

確率変数 $A, B$ について，一般に $E(A+B)=E(A)+E(B)$ が成り立つ。また，$X$ は $n$ 個の球を取り出したときに，そこに含まれる赤球の個数であるから，(1)の $Y_i$ を用いて

$$X=\sum_{i=1}^{n}Y_i$$

と表される。したがって．期待値 $E(X)$ は

$$E(X)=E\left(\sum_{i=1}^{n}Y_i\right)=\sum_{i=1}^{n}E(Y_i)$$

と表される。また，任意の $i$ に対して $E(Y_i)=\dfrac{M}{N}$ であるから，

$$E(X)=\sum_{i=1}^{n}E(Y_i)=\sum_{i=1}^{n}\frac{M}{N}=\frac{Mn}{N}$$

となる。

問 2(2) [別解]

まずはこの分布の確率質量関数 $P(X=k)$ を求める。ただし，$k$ は整数で

$$\max(0,\ n-(N-M))\leqq k\leqq \min(n,\ M)$$

である。$N$ 個の球から $n$ 個の球を取り出す場合の数は ${}_N\mathrm{C}_n$ 通りであり，取り出した $n$ 個の球のうち赤球が $k$ 個含まれる場合

問 2(2)　12 点

問 2(2)［別解］12 点

の数は ${}_M\mathrm{C}_k \cdot {}_{N-M}\mathrm{C}_{n-k}$ 通りである。したがって，確率質量関数は

$$P(X=k)=\frac{{}_M\mathrm{C}_k \cdot {}_{N-M}\mathrm{C}_{n-k}}{{}_N\mathrm{C}_n}$$

となる。求めた確率質量関数に関して具体化して考えるべく，まず，$n \neq 0$ かつ $M \neq 0$ のときを考える。このとき，$k \neq 0$ において

$$\begin{aligned}
kP(X=k) &= k \cdot \frac{\dfrac{M!}{(M-k)! \cdot k!} \cdot {}_{N-M}\mathrm{C}_{n-k}}{\dfrac{N!}{(N-n)! \cdot n!}} \\
&= \frac{M \cdot \dfrac{(M-1)!}{(M-k)! \cdot (k-1)!} \cdot {}_{N-M}\mathrm{C}_{n-k}}{\dfrac{N}{n} \dfrac{(N-1)!}{(N-n)! \cdot (n-1)!}} \\
&= \frac{Mn}{N} \cdot \frac{{}_{M-1}\mathrm{C}_{k-1} \cdot {}_{N-M}\mathrm{C}_{n-k}}{{}_{N-1}\mathrm{C}_{n-1}}
\end{aligned}$$

とできる。ここで，$\dfrac{{}_{M-1}\mathrm{C}_{k-1} \cdot {}_{N-M}\mathrm{C}_{n-k}}{{}_{N-1}\mathrm{C}_{n-1}}$ は，赤球 $M-1$ 個と白球 $N-M$ 個からなる $N-1$ 個の球から $n-1$ 個の球を取り出し，そこに含まれる赤球が $k-1$ 個である分布の確率質量関数に一致するため，

$$\sum_{k=\max(1,\, n-(N-M))}^{\min(n,\, M)} \frac{{}_{M-1}\mathrm{C}_{k-1} \cdot {}_{N-M}\mathrm{C}_{n-k}}{{}_{N-1}\mathrm{C}_{n-1}} = 1$$

となる。また，$k=0$ のときは $kP(X=k)=0$ である。以上より，期待値 $E(X)$ は

$$\begin{aligned}
E(X) &= \sum_{k=\max(0,\, n-(N-M))}^{\min(n,\, M)} kP(X=k) \\
&= \frac{Mn}{N} \cdot \sum_{k=\max(1,\, n-(N-M))}^{\min(n,\, M)} \frac{{}_{M-1}\mathrm{C}_{k-1} \cdot {}_{N-M}\mathrm{C}_{n-k}}{{}_{N-1}\mathrm{C}_{n-1}} \\
&= \frac{Mn}{N}
\end{aligned}$$

である。次に，$n=0$ もしくは $M=0$ のとき，$X$ のとりうる値は $0$ のみであるから，$E(X)=0$ となる。したがって，$N, M, n$ の値によらず $E(X)=\dfrac{Mn}{N}$ である。

問 3

**問 3　6 点（完答）**

c，e，f

問 1

＊確率質量関数 $P(X=k)$ を求めて 2 点

＊期待値 $E(X)$ を求めて 4 点

問 1［別解 1］

＊ $X=k\ (k=0,1,2,3)$ となる確率を求めて 2 点（完答）

＊期待値 $E(X)$ を求めて 4 点

問 1［別解 2］

＊ $E(Z_i)=\dfrac{3}{8}$ を求めて 2 点

＊期待値 $E(X)$ を求めて 4 点

問 2(1)

＊ $n=N$ と拡張する方針に 4 点

＊ $E(Y_1)+E(Y_2)+\cdots+E(Y_N)=M$ を求めて 4 点

＊ $E(Y_1)=E(Y_2)=\cdots=E(Y_N)$ を述べて 4 点

＊期待値 $E(Y_i)$ を求めて 4 点

問 2(2)

＊期待値の線形性を利用する方針に 4 点

＊ $X=\sum_{i=1}^{n}Y_i$ を求めて 4 点

＊期待値 $E(X)$ を求めて 4 点

問 2(2)［別解］

＊確率質量関数 $P(X=k)$ を求めて 4 点

＊ $M\neq 0$ かつ $n\neq 0$ のとき $P(X=k)$ を求めて 4 点

＊期待値 $E(X)$ を求めて 4 点

## 2 （３０点）

### 【解答・採点基準】

問１　Tには「1」が2文字含まれていることから，Sに含まれる「A」の文字数，「B」の文字数，「C」の文字数の和は2となる。Sに含まれる文字について，この2文字以外は全て「D」である。ところで，「D」の符号語には「0」が3文字含まれることから，情報源記号が全て「D」で構成される情報源記号列の符号記号列に含まれる「0」の文字数は3の倍数となる。このことと，Tに含まれる「0」の文字数が18で3の倍数であることを考慮すると，Sを構成する，「A」，「B」，「C」のどれかを用いた2文字の符号語に含まれる「0」の文字数の和は3の倍数とならなければならない。そのようになるケースとして

(ⅰ)「B」1文字，「C」1文字

(ⅱ)「A」2文字

の2つの場合が考えられる。

(ⅰ)の場合，この2文字の符号語に含まれる「0」の文字数の合計は3であるから，Tに含まれる「0」の残り文字数は18-3=15となり，Sには「D」が5文字含まれることが分かる。

(ⅱ)の場合，この2文字の符号語に含まれる「0」の文字数の合計は0であるから，Tに含まれる「0」の残り文字数は18-0=18となり，Sには「D」が6文字含まれることが分かる。

以上より，Sとして考えられる情報源記号列の組合せは「B」1文字，「C」1文字，「D」5文字の並び替えの場合の数と，「A」2文字，「D」6文字の並び替えの場合の数の和の分だけあるから，

$$\frac{8!}{6!\cdot 2!}+\frac{7!}{5!\cdot 1!\cdot 1!}=70\text{ 通り}\ \cdots\cdots\ \text{（答）}$$

問１　12点

| | |
|---|---|
| 問 2　結論：一意に復号できない。<br>理由：「100100」という符号記号列を復号することを考えたとき，復号のパターンとして「DD」と「CBA」の 2 パターンが考えられるため。<br>問 3　どの符号語も他の符号語の接頭とならないこと。<br>【問 3 の別解】<br>「ある符号語が，他の符号語の先頭にそのまま含まれる」という状態が起こっていないこと。 | **問 2**　9 点<br><br><br><br>**問 3**　9 点 |

**問 1**

＊S に含まれる「A」の文字数と「B」の文字数と「C」の文字数の和が 2 であるという記述に 2 点

＊「B」1 文字「C」1 文字「D」5 文字を並び替えるケースに言及して 3 点

＊「A」2 文字「D」6 文字を並び替えるケースに言及して 3 点

＊計算式に 3 点

（ⅰ）のみ考慮できている場合は$\frac{7!}{5!\cdot1!\cdot1!}$に相当する式が書けていれば可

（ⅱ）のみ考慮できている場合は$\frac{8!}{6!\cdot2!}$に相当する式が書けていれば可

（ⅰ）（ⅱ）両方考慮できている場合は$\frac{7!}{5!\cdot1!\cdot1!}+\frac{8!}{6!\cdot2!}$に相当する式が書けていれば可

＊70 通りという結論に 1 点

**問 2**

＊結論に 2 点

＊理由に 7 点

## 3 （３０点）

### 【解答・採点基準】

問 1(1)

| | _ | あ | さ | っ | て |
|---|---|---|---|---|---|
| _ | 0 | 1 | 2 | 3 | 4 |
| あ | 1 | 0 | 1 | 2 | 3 |
| し | 2 | 1 | 1 | 2 | 3 |
| た | 3 | 2 | 2 | 2 | 3 |

問 1(1)　8 点

(2)

[a] 文字列 A’の i-1 文字目と文字列 B’の j-1 文字目が等しい場合

dp[i][j] =min(dp[i-1][j]+1, dp[i][j-1]+1, dp[i-1][j-1])

[b] 文字列 A’の i-1 文字目と文字列 B’の j-1 文字目が異なる場合

dp[i][j] = min(dp[i-1][j]+1, dp[i][j-1]+1, dp[i-1][j-1]+1)

または

dp[i][j] = min(dp[i-1][j], dp[i][j-1], dp[i-1][j-1])+1

問 1(2)　10 点

問 2(1)

レーベンシュタイン距離を，2 つの文字列のうち，長い方の文字列の長さで割る。

問 2(1)　6 点

問 2(2)

(1)の演算を施すと，「バニラアイスクリーム」と「チョコアイスクリーム」の距離は

$$\frac{3}{10} = 0.3$$

となり，「バニラ」と「サラダ」の距離は

$$\frac{3}{3} = 1.0$$

となることから，より値の大きい「バニラ」と「サラダ」の組み合わせが，より異なっているといえる。

問 2(2)　6 点

問 1(2)

＊[a]の式に 5 点，[b]の式に 5 点

問 2(1)

＊「長い方の」という記述に 2 点

＊「文字列の長さで割る」という記述に 4 点

問 2(2)

＊「バニラアイスクリーム」と「チョコアイスクリーム」の演算結果に 2 点

＊「バニラ」と「サラダ」の演算結果に 2 点

＊結論に 2 点