２０２４年度　第２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・解説・採点基準

全３問　６０分　１００点満点

Ⅰ　（４０点）

【解答・採点基準】

問 1

P 高校に在籍していると仮定したとき ：1.6

Q 高校に在籍していると仮定したとき ：1.2

問 2

可能性が高い高校 ：Q 高校

理由 ： 指数 d の値がより小さく，より標準正規分布の中心に近いから。

問 3

相関係数が 0 の 2 つの確率変数 X, Y の平均，標準偏差をそれぞれ $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ とおくと，図 1-2 より指数 d は

$$d = \sqrt{\frac{(X-\mu_X)^2}{\sigma_X{}^2} + \frac{(Y-\mu_Y)^2}{\sigma_Y{}^2}}$$

となる。よって，学生 T が P 高校に在籍していると仮定したときの指数 d の値は

$$\sqrt{\frac{(70-60)^2}{5^2} + \frac{(125-140)^2}{20^2}} = \frac{\sqrt{73}}{4}$$

となり，一方で，Q 高校に在籍していると仮定したときの指数 d の値は

$$\sqrt{\frac{(70-80)^2}{10^2} + \frac{(125-110)^2}{15^2}} = \sqrt{2}$$

問 1　6 点（各 3 点×2）

問 2　10 点

問 3　15 点

となる。それぞれの値の 2 乗を比較すると

$$\left(\frac{\sqrt{73}}{4}\right)^2-\sqrt{2}^2=\frac{73-32}{16}>0$$

であるので $\frac{\sqrt{73}}{4}>\sqrt{2}$ となる。よって指数 d の値の値がより小さい Q 高校に在籍している可能性が高い。

問 4

異常である店 ：B, E

理由 ：95%信頼区間で考えるとき，表 1-4 より指数 d の値の 2 乗の値が 5.99 以下であれば数値として異常がないと分かる。よって，指数 d の 2 乗の値が 5.99 より大きい店を選べばよいから。

問 4　9 点

問 2

＊ d の値を比較して 3 点

＊ d の値が小さいほうが正規分布の中心に近いことへの言及に 5 点

＊答に 2 点

問 3

＊二次元の指数 d の導出に 6 点

＊ d の値に各 3 点×2

＊大小の評価に 2 点

＊大小の評価は解答通りでなくても数学的に正しい操作であれば 2 点

＊答に 1 点

問 4

＊指数 d の値の 2 乗の値が 5.99 より大きいと異常であることへの言及に 5 点

＊答に 4 点(完答)

2 （３０点）

【解答・採点基準】

問 1　$i(\mathrm{A}_{21})=1,\ i(\mathrm{A}_{22})=2,\ i(\mathrm{A}_{23})=3$

$i(\mathrm{A}_{24})=4,\ i(\mathrm{A}_{25})=5,\ i(\mathrm{A}_{26})=5$

問 2　$i(I)$のとりうる値の範囲：

$i(I)\ddagger\ 0$

$i(I)$の大きさの表すもの：

事象 I の起こりにくさ

問 3

B さんが観測するのは，サイコロ1の目が1，サイコロ 2 の目が 6 となる事象である。よって，B さんが持つ情報の量は

$$i(\mathrm{A}_{11}\text{ かつ }\mathrm{A}_{26})=-\log_2\left(\frac{1}{2}\cdot\frac{1}{32}\right)$$

$$=-\log_2\frac{1}{64}$$

$$=6$$

である。また，C さんが観測するのは，2 つのサイコロの目の積が 6 となる事象である。サイコロの目の組み合わせは

(サイコロ1の目，サイコロ 2 の目)

$=(1,6),(2,3),(3,2),(6,1)$

の 4 通りあり，これらは互いに排反な事象である。よって，2 つのサイコロの目の積が 6 となる確率は

$$\frac{1}{2}\cdot\frac{1}{32}+\frac{1}{8}\cdot\frac{1}{8}+\frac{1}{8}\cdot\frac{1}{4}+\frac{1}{8}\cdot\frac{1}{2}=\frac{1}{8}$$

であり，C さんが持つ情報の量は

$$-\log_2\frac{1}{8}=3$$

となる。したがって，2 人が持つ情報の量を比較すると

問 1　6 点(各 1 点×6)

問 2　8 点

問 3　8 点

「Bさんの情報の量が3だけ多い」

といえる。

問4

Aさんが Bさんにふったサイコロについて伝える前と伝えた後に分けて考える。伝える前に観測されるのは，サイコロ1かサイコロ2のどちらか一方をふって3以上の目が出る事象であり，余事象を用いてこの確率を計算すると

$$\frac{1}{2}\left(1-\frac{1}{2}-\frac{1}{8}\right)+\frac{1}{2}\left(1-\frac{1}{2}-\frac{1}{4}\right)=\frac{1}{2}\cdot\frac{3}{8}+\frac{1}{2}\cdot\frac{1}{4}$$
$$=\frac{5}{16}$$

となる。また，伝えた後に観測されるのは，Aさんがふったサイコロがサイコロ1で，かつ3以上の目が出る事象であり，この確率は $\frac{3}{16}$ である。したがって，Aさんが Bさんにふったサイコロについて伝える前の情報の量は $-\log_2\frac{5}{16}$，伝えた後の情報の量は $-\log_2\frac{3}{16}$ であり，

$$-\log_2\frac{3}{16}-\left(-\log_2\frac{5}{16}\right)=-\log_2 3+\log_2 5$$
$$=-1.5850+2.3219$$
$$=0.7369$$

となるから，情報の量は

「0.7369だけ増えた」

といえる。

問4　8点

問2

＊情報量 $i(I)$ のとりうる値の範囲を求めて4点

＊情報量 $i(I)$ の大きさの表すものを説明して4点

問3

＊B さんが持つ情報の量を求めて 3 点

＊C さんが持つ情報の量を求めて 3 点

＊答に 2 点

問 4

＊A さんが B さんにふったサイコロについて伝える前の情報の量を求めて 3 点

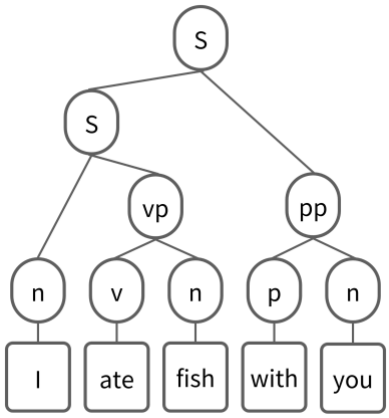
＊A さんが B さんにふったサイコロについて伝えた後の情報の量を求めて 3 点

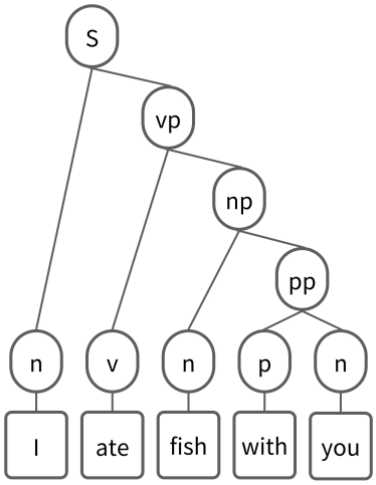
＊答に 2 点

3 （３０点）

【解答・採点基準】

問１





　どちらの解釈に基づいて生成する場合でも用いられなかった生成規則は，S→v np

問 2(i)

S→pp S

vp→pp vp

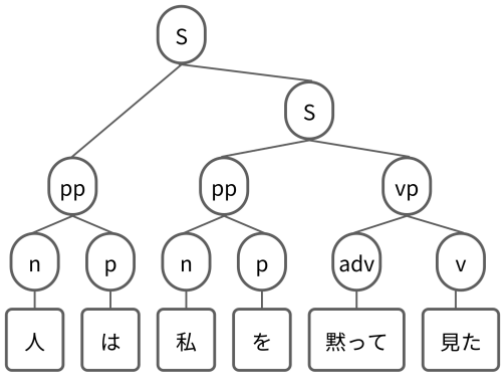
pp→pp pp

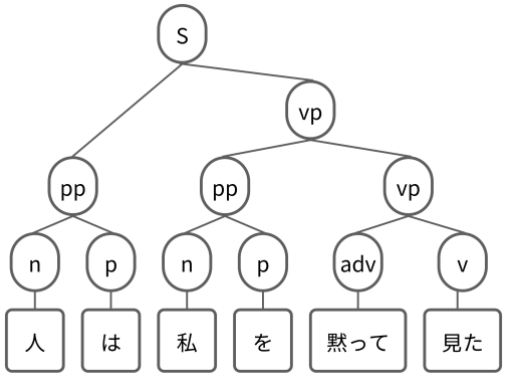
問１　６点

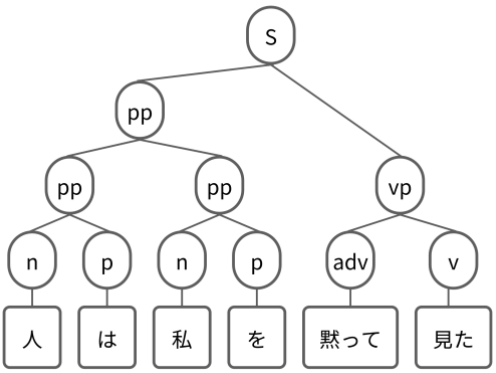
問 2(i)　８点(完答)

問 2(ii)

次の図のうち 1 つ







問 3

(解答例)　まず，規則xに登場する n 個の終端記号に対応する新たな非終端記号 A_1～A_n を導入する。つまり，規則xを「$A \to A_1\ A_2\ A_3\ \cdots\ A_n$」と書き換え(以下規則$y$とする)，形式$\alpha$の規則として新たに「$A_1 \to \alpha_1$，$A_2 \to \alpha_2$，…，$A_n \to \alpha_n$」を追加する。続いて，新たに非終端記号 B_i(i=1，2，…，n-2)を定義し，連続的に規則yを分割する。つまり，規則yを「$A \to B_1\ A_3\ A_4\ \cdots\ A_n$」と書き換

問 2(ii)　4 点

問 3　12 点

え，形式αの規則として「$B_1 \to A_1\ A_2$」を加える。以後同様に形式αでない生成規則を 2 つの規則に分割すれば，最終的に以下の形式αの生成規則群が得られる。

$$\begin{cases} A \to B_{n-2}\ A_n \\ B_1 \to A_1\ A_{2s} \\ B_i \to B_{i-1}\ A_{i+1} (i=2,\ 3,\ \cdots,\ n-2) \\ A_i \to a_i (i=1,\ 2,\ \cdots,\ n) \end{cases}$$

問 1

＊図 1 つ毎に 2 点，「S→v np」で 2 点

＊図の要素を囲む枠は不要

問 2(ii)

＊図の要素を囲む枠は不要

問 3

＊規則xにおいて終端記号を非終端記号に入れ替えるアイデアを説明して 4 点

＊規則xにおいて 3 つ以上の非終端記号列を「A→B C」の形を含んだ規則群に分割するアイデアを説明して 4 点

＊最終的に生まれる等価な規則群を示して 4 点