

# 総 問

## 総 合 問 題

令 和 5 年 度(前期)

### 注 意

1. 「解答はじめ」というまで開いてはいけない。
2. 問題は1冊（本文8ページ，白紙2枚），解答用紙は3枚である。白紙は問題冊子の中にはさみこんであるので引き抜いて下書用紙として使ってよい。
3. 全部の解答用紙に受験番号を書くこと。受験番号は次の要領で明確に記入すること。

(例) 受験番号 90001 番の場合 → 

9	0	0	0	1
---	---	---	---	---

4. 解答は解答用紙の表面のみに書くこと。裏面に書くと無効になることがある。
5. 書き損じても，代わりの用紙は交付しない。
6. 試験終了後，問題冊子と白紙は持ち帰ること。

1

標準正規分布に従う確率変数  $X$  について、数列  $x_1, \dots, x_{2022}$  を

$$P(X \leq x_i) = \frac{i}{2023}, \quad i = 1, \dots, 2022 \quad (1)$$

を満たすように定義する。

問 1 次の a ~ d のうち正しいものをすべて選び、その記号を答えなさい。

- a.  $i < j$  を満たす任意の  $i, j = 1, \dots, 2022$  について  $x_i < x_j$  が成り立つ。
- b. 任意の  $i = 1, \dots, 2022$  について  $-x_i = x_{2023-i}$  が成り立つ。
- c. 任意の  $i, j = 2, \dots, 2022$  について  $x_i - x_{i-1} = x_j - x_{j-1}$  が成り立つ。
- d. 上式(1)において、 $X$  を  $-X$  に替えたとき同様に定義される数列を  $x'_1, \dots, x'_{2022}$  とすると、任意の  $i = 1, \dots, 2022$  について  $x_i = -x'_i$  が成り立つ。

問 2 標準正規分布からの大きさ 2022 の無作為抽出されたデータを小さい順に並べ替えたものを  $z_1, \dots, z_{2022}$  とするとき、 $(x_1, z_1), \dots, (x_{2022}, z_{2022})$  の散布図はどのような形状になるか。必要に応じて図を用いて説明せよ。

問 3 平均 1, 分散 1 の正規分布からの大きさ 2022 の無作為抽出されたデータを小さい順に並べ替えたものを  $z'_1, \dots, z'_{2022}$  とするとき、 $(x_1, z'_1), \dots, (x_{2022}, z'_{2022})$  の散布図はどのような形状になるか。問 2 の散布図と比較しつつ、必要に応じて図を用いて説明せよ。

問 4 図 1 に示された散布図におけるデータ  $y_1, \dots, y_{2022}$  は、どのような分布から抽出されたと考えられるか。その分布の特徴を標準正規分布と比較して述べよ。

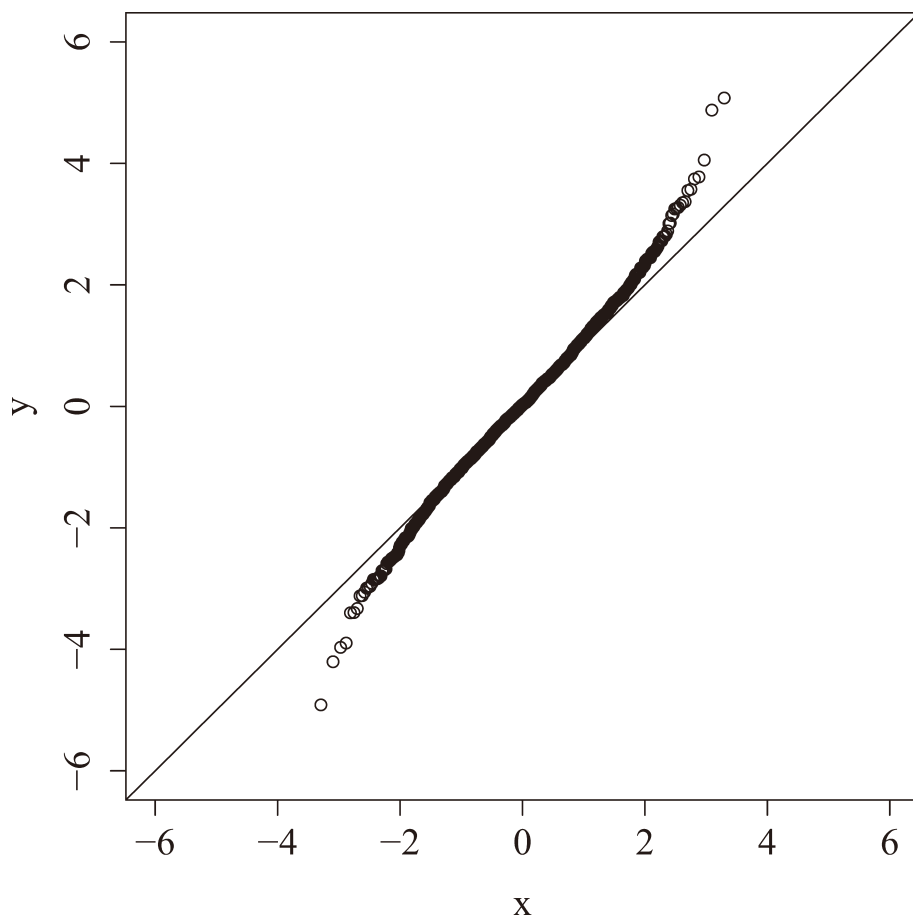


図1 :  $(x_1, y_1), \dots, (x_{2022}, y_{2022})$  の散布図

- 2 Aさんは、百科事典ウェブサイトの人物ページ閲覧数のデータをもとに、日本の芸術偉人ランキングを作成しようと考えた。各人物には、日本語ページだけでなく、英語ページと中国語ページがあり、言語別にページ閲覧数が与えられている。Aさんは以下の表1の5人についての閲覧数データを集めた。Aさんは、何らかの基準で算出した数値が高い順に各人物を順位付けようと考えている。便宜上、人物を紫式部から順に  $i = 1, \dots, 5$ , 言語を日本語から順に  $j = 1, 2, 3$  とおく。

表1：各言語ページの閲覧数データ(単位：100万回)

人物	日本語( $j = 1$ )	英語( $j = 2$ )	中国語( $j = 3$ )
紫式部( $i = 1$ )	6	1	1
清少納言( $i = 2$ )	2	2	2
松尾芭蕉( $i = 3$ )	2	3	3
葛飾北斎( $i = 4$ )	3	4	3
歌川広重( $i = 5$ )	1	1	1

- 問1 人物  $i$  についての言語  $j$  でのページ閲覧数を  $v_{ij}$  で表す。次の指標

$$R_i = \sum_{j=1}^3 v_{ij}$$

の値が大きい順に5人を並べて得られるランキングはどのようなものか説明せよ。

- 問2 指標

$$H_i = - \sum_{j=1}^3 \left( \frac{v_{ij}}{R_i} \right) \log_2 \left( \frac{v_{ij}}{R_i} \right)$$

をもとに、表で示した5人について、 $H_i$ の値が大きい順に順位を記載せよ。対数の計算は以下の値を参考にしてよい。

$$\log_2 3 = 1.5850 \quad \log_2 5 = 2.3219 \quad \log_2 7 = 2.8074$$

- 問3 指標  $H_i$ の値が大きい順に5人を並べて得られるランキングはどのようなものか説明せよ。

3 学術セミナーへの Web 登録情報を確認したところ、無効なメールアドレスが混在していることが発覚した。そこで、後述の正規表現を用いて、登録されたメールアドレスが有効かどうかを 1 件ずつ判定することにした。なお、すべてのメールアドレスは、アットマーク(@)を隔てて前半のローカル部と、後半のドメイン部から構成される。また、ドメイン部はドット(.)で区切られた複数の部分を持ち、各部分をラベルと呼ぶ。例えば、次のメールアドレス

`user@xyz.co.jp`

において、`user` がローカル部、`xyz.co.jp` がドメイン部であり、また、`xyz`、`co`、`jp` の各々がラベルである。

正規表現とは、単一の文字列で複数の文字列を表現する方法であり、特定のパターンを満たす文字列を抽出する際に用いられる。なお、ある文字列が特定の正規表現で表される文字列に該当するかどうか判定することを「文字列照合」といい、該当する場合は「照合する」という。正規表現中の英数字と一部の記号は、表記の通りの英数字・記号と照合する。その一方で、一部の記号、及び、記号と英数字の組合せは特殊な意味や機能を持つ。特殊な表現の例を表 2 に示す。特殊表現を用いた正規表現と、照合する文字列の例をそれぞれ掲載する。

表 2：正規表現の特殊文字・表現

種類	記号	機能	正規表現例	照合文字列例
特殊文字	・	任意の文字と照合	<code>x.z</code>	<code>xyz</code>
	＼	直後の特殊文字を表記通りに照合	<code>\\.</code>	<code>.</code>
特殊表現	[ ]	括弧内の文字(列)のいずれか 1 文字と照合	<code>[1@a]</code>	<code>@</code>
	[ ]	(括弧中の先頭・末尾のハイフン(-)は表記通り照合。	<code>[-abc]</code>	<code>-</code>
	[ ]	括弧中の文字間のハイフン(-)は文字範囲を表現)	<code>[a-c]</code>	<code>c</code>
	( )	括弧内の文字列をグループ化	<code>(abc)</code>	<code>abc</code>
繰り返し	+	直前の文字・グループの 1 回以上の繰り返しと照合	<code>(ab)+</code>	<code>abab</code>

従って、例えば下記の3つの文字列

**1ab@ 321ab@ 33333333ab@**

は、正規表現

**[1-3]+ab@** (2)

と照合する。翻って、下記の3つの文字列

**1a 41ab@ ab@**

は、正規表現(2)と照合しない。

下記の2つの条件を満たすメールアドレスを、ABC大学ドメインの有効なメールアドレスとして定義する。

- ・ローカル部は「1文字以上の英字」で表される。
- ・ドメイン部は「1文字以上の英数字」と「**.abc.ac.jp**」の結合で表される。

ABC大学ドメインの有効なメールアドレスのドメイン部と照合する正規表現は、下記の通りである。

**[a-zA-Z0-9]+\.\abc\.ac\.jp** (3)

問 1 ABC大学ドメインの有効なメールアドレスのローカル部と照合する正規表現を答えよ。

下記の5つの条件を満たすメールアドレスを、有効なメールアドレスとして定義する。

- ・ローカル部は、以下のいずれか、もしくはそれらの結合で構成される。
  - 1文字以上の英数字
  - 1つ以上のドット(.)
  - 1つ以上のハイフン(-)
  - 1つ以上のアンダースコア(\_)
- ・ドメイン部には複数のラベルが出現する。
- ・ドメイン部の各ラベルはドット(.)で区切られている。
- ・ドメイン部の末尾を除く各ラベルは、以下のいずれか、もしくはそれらの結合で構成される。
  - 1文字以上の英数字
  - 1つ以上のハイフン(-)
- ・ドメイン部の末尾のラベルは、1文字以上の英字から構成される。

問 2 有効なメールアドレスと照合する正規表現を答えよ。なお、必ず( )を1セットと+を3回使って解答せよ。

ある正規表現が与えられた文字列と照合するか検証するために、正規表現と対応する状態遷移図を用いる。状態遷移図は、状態(O)と、状態から状態への遷移(→)によって構成される。また、特殊な状態として、初期状態(⇒O)と受理状態(O)が存在する。

状態遷移は初期状態に文字列が入力されて始まり、その文字列が1文字ずつ、先頭から末尾に向かって読み込まれる。文字が1つ読み込まれるごとに、その文字の条件に合致した矢印に従って状態が遷移する。末尾の文字が読み込まれて最後の状態遷移が起これば、結果が出力される。そのときの状態が受理状態であれば出力は受理(正規表現と照合する)となり、その他の状態であれば出力は非受理(正規表現と照合しない)となる。

図2は、正規表現(2)と対応する状態遷移図である。ここで入力として文字列1aを考える。初期状態は(S<sub>1</sub>)であり、入力の1文字目(先頭)が1であることを受けて、状態(S<sub>2</sub>)に遷移する。次に2文字目(末尾)のaを読み込み、状態(S<sub>3</sub>)に遷移する。ここで文字列の末尾に到達したものの、状態(S<sub>3</sub>)は受理状態でないため、正規表現(2)は文字列1aと照合しないという結果が得られた。

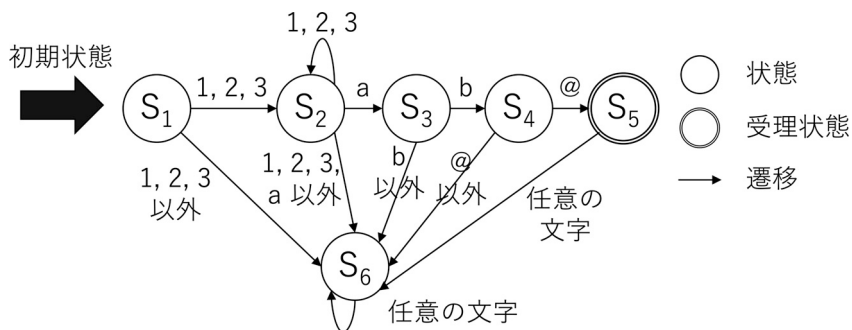


図2：正規表現(2)と対応する状態遷移図



問 3 図 3 は、問 2 で作成した正規表現と対応する状態遷移図である。空欄(ア)～(カ)に該当するものを下記の選択肢から選択せよ。なお、同一の選択肢を複数回使用してもよい。

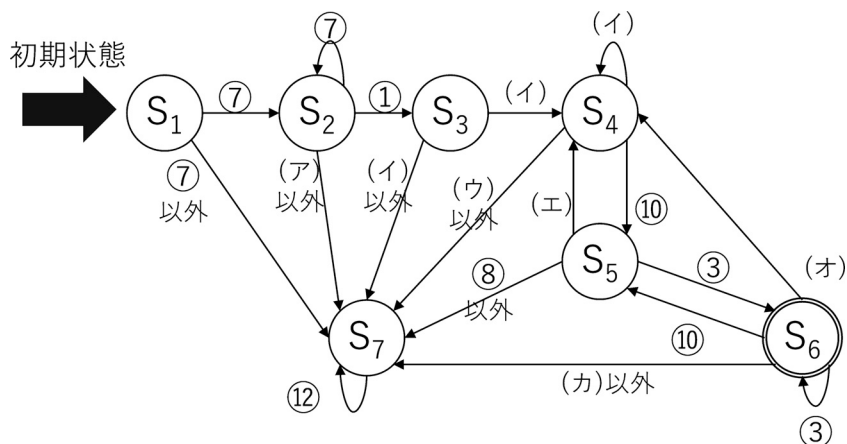


図 3 : 有効なメールアドレスを抽出する正規表現と対応する状態遷移図

- ① アットマーク (@)
- ② アンダースコア (\_)
- ③ 英字
- ④ 英数字
- ⑤ 英数字, ドット (.), ハイフン (-)
- ⑥ 英数字, ドット (.), ハイフン (-), アンダースコア (\_),  
アットマーク (@)
- ⑦ 英数字, ドット (.), ハイフン (-), アンダースコア (\_)
- ⑧ 英数字, ハイフン (-)
- ⑨ 数字, ハイフン (-)
- ⑩ ドット (.)
- ⑪ ハイフン (-)
- ⑫ 任意の文字