平成 24 年度 10 月期入学 / 平成 25 年度 4 月期入学
京都大学大学院 情報学研究科
修士課程 知能情報学専攻 入学者選抜試験問題
(専門科目)

平成 24 年 8 月 6 日 13:00〜16:00

【注意】

1. 問題冊子はこの表紙を含めて 40 枚ある.
2. 試験開始の合図があるまで中を見てはいけない.
3. 試験開始後,枚数を確認し,落丁または印刷の不鮮明なものがあれば直ちに申し出ること.
4. 問題は下記 24 題であり,一部の問題は日本語と英語の両方で出題されている.このうちいずれか **4 題を選択し**,解答しなさい.

5. 特に指定のない限り,日本語または英語で解答しなさい.
6. 解答用紙の表紙に記載されている注意事項についても留意すること.

The Japanese version of this document is the prevailing and authoritative version; the English translation below is given for reference only

October 2012 Admissions / April 2013 Admissions
Entrance Examination for Master's Program
Department of Intelligence Science and Technology
Graduate School of Informatics, Kyoto University
(Specialized Subjects)

August 6, 2012
13:00 - 16:00

NOTES

1. This is the Problem Booklet in 40 pages including this front cover.
2. Do not open the pages until you are instructed to start.
3. After start, check all pages in order and notify proctors (professors) immediately if missing pages or unclear printings are found.
4. There are 24 problems, and some problems are written in Japanese and English. The problems are classified as listed below. **Choose and solve 4 problems.**

5. Solve the problems in Japanese or English, unless otherwise specified.
6. Read carefully the notes on the front cover of the Answer Sheets as well.

普遍的な生命現象の研究にはモデル生物(model organism)が用いられる。いくつかのモデル生物については、全ゲノム配列の解析が進んでおり注目されている。

設問 1　単細胞生物(unicellular organism)及び、無脊椎動物(invertebrate)のモデル生物をそれぞれ一つ挙げ、どのような研究に用いられているか説明せよ。

設問 2　脊椎動物(vertebrate)のモデル生物を 2 つ挙げ、研究に使用する上で、その 2 つのモデル生物の有用性の違いについて述べよ。

設問 3　ゲノム情報解読の観点からは、モデル生物にはどのような特徴が求められるか簡潔に述べよ。

神経機能(neural function)を調べる方法のひとつとして、光遺伝学(optogenetics)が注目されている。

設問 1　光遺伝学を用いた解析の際、よく利用されるタンパク質(protein)の特徴について簡潔に述べよ。

設問 2　ハツカネズミ(*Mus musculus*)での光遺伝学を用いた神経機能の解析手法の例をひとつ挙げ、光遺伝学を用いないで同様の解析をした場合と比較し利点（advantage）を簡潔に説明せよ。

設問 3　光遺伝学の神経系以外への応用例をひとつ挙げ、その際に用いられるタンパク質の特徴について簡潔に述べよ。

設問　以下の語句のうち 8 つを選んで、それぞれ 4 行以上 10 行以内で説明せよ。必要に応じて図を用いてもよい。

(1) トランスクリプトーム解析 (transcriptome analysis)
(2) メタボローム解析 (metabolome analysis)
(3) ペプチドーム解析 (peptidome analysis)
(4) 減数分裂 (meiosis)
(5) コンジェニック系統 (congenic strain)
(6) 量的形質遺伝子座 (quantitative trait loci)
(7) 近交系 (inbred strain)
(8) ハーディワインベルグの法則 (Hardy-Weinberg principle)
(9) 連鎖不平衡 (linkage disequilibrium)
(10) 活動電位 (action potential)
(11) 長期増強 (long-term potentiation)
(12) 最初期遺伝子群 (immediate early genes)

以下の文章を読み、設問に答えよ。

ハツカネズミに対して、レトロウイルスを用いてゲノム DNA に GFP（green fluoresent protein）遺伝子を挿入し、変異体（mutant）を作製した。このようにして作製された変異体のうち、ある変異体 A では一方の常染色体上の遺伝子 X のエクソン（exon）に GFP が挿入されていた（遺伝子型は X/X*）。変異体 A どうしを掛け合わせたところ、得られた仔のおよそ 12.5%で内臓逆位が観察された。得られた仔のうち、野生型ホモ接合体（homozygote）（X/X）及びヘテロ接合体（X/X*）では内臓逆位は観察されず、遺伝子 X の発現は胚の左側で観察された。変異ホモ接合体（X*/X*）で胎生致死（embryonic lethal）は観察されず、正常な遺伝子 X の mRNA の発現も観察されなかった。

設問 1　変異ホモ接合体（X*/X*）では、内臓逆位はどのような頻度で起きていると予想されるか、理由とともに簡潔に述べよ。

設問 2　遺伝子挿入法を用いた突然変異体の作製法の利点（advantage）と欠点（disadvantage）について、ENU（ethylnitrosourea）を用いた変異体の作製法と比較し説明せよ。

設問 3　内臓逆位がおこる頻度に着目し、遺伝子 X の機能について仮説（hypothesis）をたてよ。

設問 4　設問 3 で得られた仮説を検証する実験計画（experimental plan）をたて、どのような結果が得られたら仮説が支持されるか説明せよ。

以下の英文を読んで設問に答えなさい。

Non-Disclosure

D. R. Zerbino, B. Paten & D. Haussler, *Science* **336**, 179 (2012) より抜粋・改変

設問 1 (A) ～ (E) に適当な単語を入れよ。

設問 2 図中のカラム 1 ～ 6 の各々について、どのような突然変異 (mutation) がどの系統 (lineage) で生じたと考えられるか、述べよ。

設問 3 アウトグループ (outgroup) の配列 (sequence) をアラインメント (alignment) に含めると何が明らかになるか、述べよ。

設問 4 アラインメントには、なぜ保存されたカラムとそうでないカラムが観察されるのか、専門語「機能的制約 (functional constraint)」を使って理由を述べよ。

$s = s_1 s_2 \ldots s_n$ を RNA 配列（RNA sequence）とする．RNA 二次構造予測問題（RNA secondary structure prediction problem）の基本版は，以下の条件をすべて満たす配列 s 上の位置の対 (i,j) の集合 B で要素数最大のもの（複数ある場合はその一つ）を見つける問題として定義される．

- すべての対 $(i,j) \in B$ において，$j > i+1$ が成立
- すべての対 $(i,j) \in B$ において，$\{s_i, s_j\} \in \{\{\mathrm{A},\mathrm{U}\},\{\mathrm{G},\mathrm{C}\}\}$ が成立
- 一つの i は高々 1 個の j と対をなし，一つの j は高々 1 個の i と対をなす
- 任意の $(i,j) \in B$ に対して，$i < i' < j < j'$ を満たす $(i',j') \in B$ が存在しない

この定義のもとでの RNA 二次構造の例を図 1 に示す．この定義のもとで以下の設問に答えよ．

設問 1 RNA 配列 ACCCGGUCGGUUUAACC に対する RNA 二次構造を求め，図 1 のように図示せよ．

設問 2 塩基数（number of bases）n の RNA 配列に対する $|B|$ の最大値を理由とともに示せ（ただし，B は RNA 二次構造）．

設問 3 以下は RNA 二次構造を求める動的計画法（dynamic programming）アルゴリズム（algorithm）の中心部分である (ただし，i,j などの適用範囲は省略)．(a) および (b) のそれぞれに入る項（term）を示せ．

$$D(i,j) = \max \begin{cases} D(i+1,j), \\ D(i,j-1), \\ \boxed{\text{(a)}} + w(s_i,s_j), \\ \max_{i<k<j}\{\, D(i,k) + \boxed{\text{(b)}} \,\}, \end{cases}$$

where

$$w(x,y) = \begin{cases} 1 & \text{if } \{x,y\} \in \{\ \{\mathrm{A},\mathrm{U}\},\ \{\mathrm{G},\mathrm{C}\}\ \}, \\ 0 & \text{otherwise.} \end{cases}$$

設問 4 RNA 二次構造予測問題の定義に以下の条件を追加した場合の RNA 二次構造を求めるアルゴリズムの概要を上にならって示し，その時間計算量（time complexity）について議論せよ．

- 任意の $(i,j) \in B$ に対して，$i < j < i' < j'$ を満たす $(i',j') \in B$ が存在しない

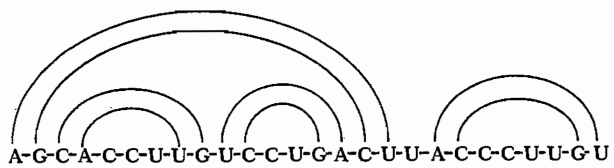


図 1: RNA 二次構造の例．$(i,j) \in B$ が円弧で結ばれている．

設問　以下の語句の中から 8 つを選んで、それぞれ 4 行以上 10 行以内で説明せよ。必要に応じて図を用いて説明してもよい。

1. ラマチャンドランプロット (Ramachandran plot)
2. プロテオーム解析 (proteome analysis)
3. UniProt
4. ホモロジーモデリング (homology modeling)
5. ハイドロパシープロット (hydropathy plot)
6. ロトカ–ヴォルテラ方程式 (Lotka-Volterra equation(s))
7. ヒストンコード (histone code)
8. 遺伝的浮動 (genetic drift)
9. 遺伝子重複 (gene duplication)
10. 遺伝的アルゴリズム (genetic algorithm)

以下の文章を読んで，設問に答えよ．

2 つの線分(line segments)がなす角の大きさを私たちが正しく知覚している(perceive)かどうかを調べるために，以下のような実験(experiment)が行われた．図 1 に示すように，線分 A と線分 B がなす角度 α が正確に知覚されているかを調べるために，線分 C と線分 B が平行になるように C の方位(orientation)を調整させた．

図 1 は，C の方位の調整誤差(error in setting)を表す．

設問 1　図 1 のデータからわかることを書け．

設問 2　図 1 の著者らは，このデータが一次視覚野(primary visual cortex)に存在する方位コラム(orientation column)の間の側抑制(lateral inhibition)によるものであると主張している．

設問 2-1　一次視覚野の方位コラムとは何か．また側抑制とは何か．

設問 2-2　この考えに基づき図 1 のようなデータが得られる理由を述べよ．

設問 3　図 2 は錯視(visual illusion)の一例である．どんな錯視であるかを定性的に(qualitatively)書け．

設問 4　この錯視が生ずる機構(mechanism)は図 1 で示された方位コラム間の側抑制によるものであると考えられている．その考えに基づき，錯視が生じる理由について説明せよ．

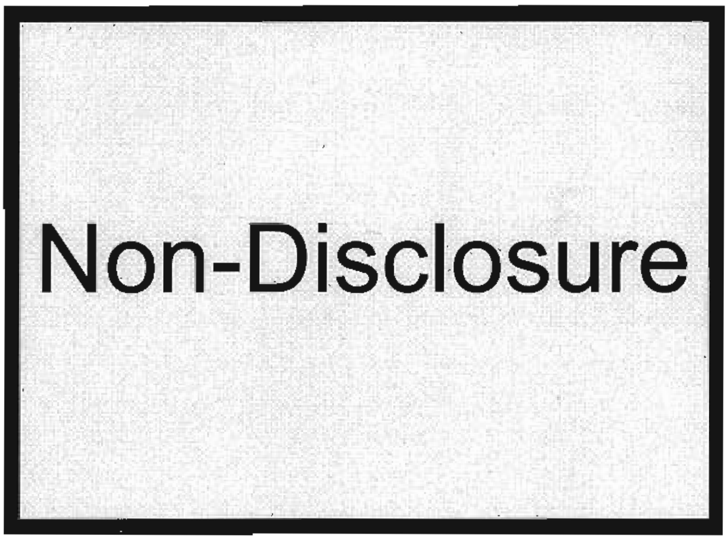


図 1

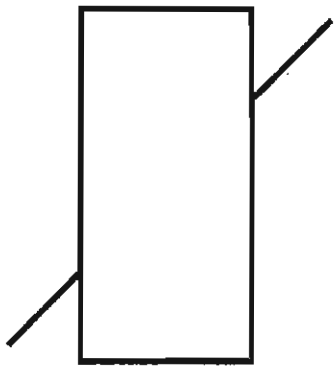


図 2

（図 1 は以下の論文の Figure 7 を改変したもの．Carpenter, R. H. S., and Blakemore, C. (1973) Interactions between orientations in human vision. *Experimental Brain Research*, **18**, 287-303.）

以下の実験についての文章を読み，設問に答えよ．

刺激(stimulus)として図 1 のような物体(object)の線画(line drawing)が 1 つずつ 100msec 呈示(present)され，実験協力者(participant)はその物体の基本レベルカテゴリ(basic-level category)での名前(例．ピアノ)を答えた．刺激は 24 種類の異なる基本レベルカテゴリからなり，図 1 のように各カテゴリについて基本レベルカテゴリでの名前が同じで形状(shape)が異なる 2 種類の物体が用いられた．さらに，それぞれについて視角(visual angle) 3.5°と 6.2°の 2 種類の大きさの刺激が作成された．実験では 24 種類の各カテゴリの物体について 1 試行ずつ，計 24 試行からなるブロックが 2 回行われた．第 1 ブロックの各試行では，各カテゴリの 2 種類の物体のいずれか 1 つが刺激として選ばれ，刺激が視角 3.5°で呈示される確率(probability)と視角 6.2°で呈示される確率はいずれも 0.5 であった．第 2 ブロックでは第 1 ブロックで呈示された各カテゴリの刺激に対して以下のいずれかの条件で刺激が呈示された．

(A) 同じ大きさ，かつ同じ形状の物体
(B) 異なる大きさ，かつ同じ形状の物体
(C) 同じ大きさ，かつ形状が異なる物体
(D) 異なる大きさ，かつ形状が異なる物体

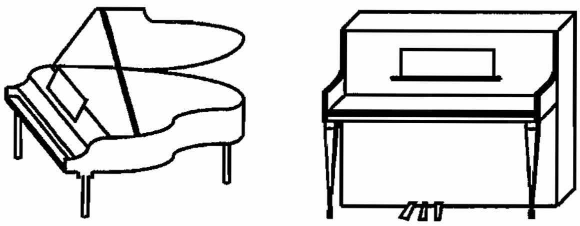


図 1. 刺激の例．下位レベル(subordinate-level)カテゴリでの名前は左から「グランドピアノ」，「アップライトピアノ」だが，実験ではいずれも「ピアノ」と答える．

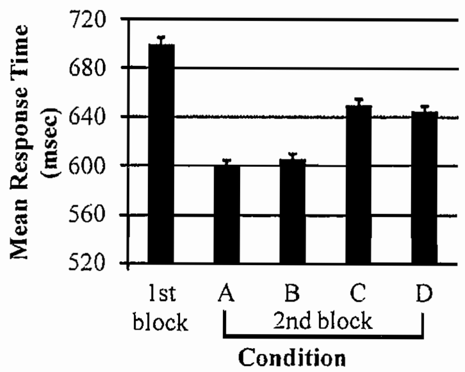


図 2.

設問 1　図 2 は正しく物体の名前を答えるまでの平均反応時間(mean response time)である．第 1，第 2 ブロック間で統計的に有意な(statistically significant)差が見られた．このように直前に呈示された刺激によって後に呈示された刺激の処理(processing)が促進(facilitate)される現象(phenomenon)を何と呼ぶか答えよ．

設問 2　第 1 ブロックより第 2 ブロックで反応時間が短くなる要因(factor)として，設問 1 の現象による効果以外に練習効果(practice effect)が挙げられる．

設問 2-1　この実験の場合の練習効果について説明せよ．

設問 2-2　この実験で新たにどのような条件を設ければ練習効果の大きさを調べることができるか述べよ．

設問 3　第 2 ブロックの各条件における平均反応時間について刺激の大きさと形状の二要因分散分析(two-way analysis of variance)を行うと，刺激の形状の主効果(main effect)が有意で，大きさの主効果，交互作用(interaction)は有意ではなかった．この結果から，物体の大きさという点で，脳内で物体がどのように表現されているか考察せよ．

ある実験の概要を記した以下の英文を読んで設問に答えよ.

Non-Disclosure

Simons & Wang (1998) *Psychological Science*, 9, 315-320. より改変

設問 1　この実験で設定された orientation-change 条件（condition）と viewpoint-change 条件の間で共通していること，および異なっていることについて述べよ.

設問 2　下線部で説明されている統制（control）はなぜ必要であるのか述べよ.

設問 3　Fig. 2 の 4 本の棒（bar）の中で，viewpoint-change 条件の結果を示しているものを 図中の番号を用いて全て挙げよ．さらに，その中でテーブル（table）を動かさなかった場合の結果を示しているのは何番であるか答えよ.

設問 4　視点依存（view-dependent）あるいは視点不変（view-invariant）という用語を用いて各条件の結果を説明せよ.

結合錯誤(illusory conjunction)に関する実験の説明を読んで，以下の設問に答えよ．

実験刺激(experimental stimuli)は色付きのアルファベット 3 文字と，その左右に黒色文字で表示された数字から構成されている(図 1)．プローブ刺激(probe stimuli)を 1 秒間呈示した後に，実験刺激を 120 ミリ秒呈示したとき，実験参加者(participants)は呈示された数字を口頭で回答するとともに，プローブ刺激と同じ色と形のアルファベットが実験刺激中に呈示されたかをボタン押しにより回答した．

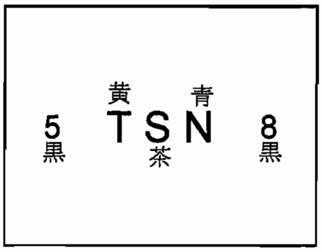


図 1 実験刺激の一例．実験に用いた文字の色は，それぞれの文字の上下に例示している．

設問 1 この実験では 3 種類のプローブ条件を用意した．ひとつは実験刺激中に呈示される色と形のアルファベットの一つを表示する一致プローブ条件(identical probe condition)，ひとつは実験刺激中に使用されない形のアルファベットを実験刺激中に使用される色で表示する特徴プローブ条件(feature probe condition)である．結合錯誤の効果を見るためのもうひとつの条件として，どのようなプローブ条件を用意すればよいか．具体例を示すとともに，その理由を述べよ．

設問 2 プローブ刺激と同一の色と形の文字が実験刺激中に呈示されたと回答した確率について，実験参加者間平均値の対応のある t 検定(paired t-test)を実施した．一致プローブ条件と特徴プローブ条件との間において統計検定(statistical test)したところ，$t(10) = 4.08$ という結果を得た．

設問 2-1 何名の実験参加者の測定結果にもとづく検定であるか．その理由とともに述べよ．

設問 2-2 両側仮説(two-tailed hypothesis)における p 値(p value)を求めよ．ただし，多重比較補正(correction for multiple comparison)は実施しなくてよい．なお自由度(degree of freedom)が 10 のとき，片側仮説(one-tailed hypothesis)の有意水準(significance level) $\alpha = 0.05$ で $t = 1.81$，$\alpha = 0.01$ で $t = 2.76$，$\alpha = 0.001$ で $t = 4.14$ として内挿補間(interpolation)してよい．

設問 2-3 設問 2-2 で求めた p 値は多重比較補正を実施していないため，正確な p 値を表しているとは言えない．その理由について説明せよ．

設問 3 実験参加者への教示(instruction)において，左右に表示された数字を確実に記憶するよう強調した．なぜ，数字の記憶を強調したと考えられるか．結合錯誤のメカニズムを特徴統合理論(feature integration theory)にもとづいて説明することにより，その理由を考察せよ．

配列 D に入った n 個のデータを降順（大きいもの順）にソートするアルゴリズムを分割統治法で設計する．この分割統治型ソートは 3 ステップから構成される．まず，n 要素の連続データを 2 つの連続データに等分する．次に，それぞれを再帰的にソートする．最後に，得られた 2 つのソート済み連続データ（ランと呼ぶ）を前から順に見ていき，併合し，新たなランを作る．最終的に得られる 1 つのランがソートの結果となる．

設問 1　配列 D の要素が `[0 7 5 2 1 3 0 4]` であるときに，上記のソートがどのように進むか図示しなさい．

設問 2　このソートを利用するある応用では，同じ値の要素があった場合にはそれらの要素間の元の順序が保存されている必要がある．このようなソートの性質は何と呼ばれるか答えなさい．

設問 3　上記の分割統治型ソート法 `Sort(`D, a, b`)` の擬似コードを書きなさい．ただし，D はソートすべき配列名，a はソート対象となる連続データの最初のインデックスであり，b は最後のインデックスである．

なお，配列 D の a 番目の要素から c 番目の要素までのランと $c+1$ 番目の要素から b 番目の要素までのランを併合する手続き `Merge(`D, a, c, b`)` （$a \leq c < b$）は自由に使用してよい．

設問 4　上記の分割統治型ソート法 `Sort(`D, 1, n`)` の計算時間を求める．今，$T(n)$ を n 個のデータをソートするのに要する計算時間とし，$M(n)$ を長さが $\lfloor n/2 \rfloor$ と $\lceil n/2 \rceil$ の 2 つのランをマージするのに要する計算時間とする．また，2 つのデータの比較は定数時間で行えるものとする．

設問 4-1) 計算時間 $M(n)$ を求めなさい．
その根拠，証明等を説明すること．

設問 4-2) $T(n)$ に対する漸化式を求めなさい．
その根拠を説明すること．

設問 4-3) 4-2) の漸化式を解き，$T(n)$ を求めなさい．
その計算過程は詳しく書くこと．

Question is translated in English in the section below; this translation is given for reference only.

We want to design a sorting algorithm based on the divide-and-conquer paradigm. The data is stored in an array of n elements and the sort order is descending, that is, from larger to smaller. It involves three steps: First, it divides the n-element sequence into two subsequences of $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ elements. Next, it sorts the two subsequences recursively using the same sorting algorithm to generate the sorted sequence called "run". Finally, it merges the runs (the two sorted subsequences) to produce the single run. The final run is the sorted result.

Question 1 Illustrate the sorting process of this sorting algorithm on the array D of which elements are `[0 7 5 2 1 3 0 4]`.

Question 2 An application of this sorting algorithm requires that the original order should be preserved for elements of the same value. What is this property of sorting called?

Question 3 Write pseudocode for the sorting algorithm `Sort(`D, a, b`)`, where D is the name of the array to be sorted, a and b are the indices of the first and the last elements of the sequence to be sorted, respectively.

You can use freely `Merge(`D, a, c, b`)` ($a \leq c < b$) that merges two runs from a-th element to c-th element and from $c+1$-th element to b-th element into one run.

Question 4 Let's analyze the running time of our sorting pseudocode, `Sort(`D, 1, n`)`. Let $T(n)$ be the running time of the sorting on n elements and $M(n)$ be the running time of the merge of two runs of the length $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$. Suppose that the comparison of two data can be executed in a constant time.

Question 4-1) Answer the running time $M(n)$ with explaining the reason or the proof.

Question 4-2) Answer the recurrence for $T(n)$ with its explanation.

Question 4-3) Solve the recurrence for $T(n)$ given by 4-2). Explain the solving process in detail.

アイテムとよばれる要素の有限集合をデータとするデータベース D がある．議論を簡単にするため，アイテムは $I = \{1, 2, \ldots, m\}$ の要素である非負整数とし，$D = \{t_1, t_2, \ldots, t_n\}$ $(t_i \subseteq I, i = 1, 2, \ldots, n)$ とする．ここで，アイテムの集合 s が D に**出現する頻度**を $f(s) = \sharp\{i \mid s \subseteq t_i\}$ と定義する．なお，集合 S の要素数を $\sharp S$ と表す．非負整数 σ が与えられたとき，アイテムの空でない集合 s で $f(s) \geq \sigma$ を満たすものすべて列挙することを考える．

設問 1 $m = 5, n = 4, D = \{\{1, 2, 4\}, \{2, 3, 4, 5\}, \{1, 2\}, \{2, 4, 5\}\}, \sigma = 2$ のとき，$f(s) \geq \sigma$ を満たすアイテムの空でない集合 s をすべて列挙せよ．

設問 2 アイテムの集合 s の頻度 $f(s)$ は次の性質 1 を満たすことを示せ．
性質 1 アイテムの集合 s_1 と s_2 に対して $s_1 \subseteq s_2$ ならば $f(s_1) \geq f(s_2)$．

設問 3 アイテムの集合 s を引数とする次の関数 f_1, f_2, f_3, f_4 の中で，性質 1 を満たすものをすべて挙げよ．理由も述べること．

1. $f_1(s)$: s に含まれるアイテムで最大のもの．
2. $f_2(s)$: s に含まれるアイテムで最小のもの．
3. $f_3(s)$: s に含まれるすべてのアイテムの最大公約数．
4. $f_4(s)$: s に含まれるすべてのアイテムの最小公倍数．

Algorithm 1 を次のように定義する．

Algorithm 1
$i := 1$; $C_1 := \{\{1\}, \{2\}, \ldots, \{m\}\}$; $L_1 := \emptyset$;
while$(C_i \neq \emptyset$ && $i \leq m)$
　　while$(C_i \neq \emptyset)$
　　　　$s :=$ choose(C_i); $C_i := C_i - \{s\}$;
　　　　if $(f(s) \geq \sigma)$
　　　　　　$L_i := L_i \cup \{s\}$;
　　　　endif
　　endwhile
　　$C_{i+1} :=$ extend(L_i); $L_{i+1} := \emptyset$; $i := i + 1$;
endwhile
output $L_1 \cup L_2 \cup \ldots \cup L_{i-1}$;

ここで，関数 choose は，空でない集合 S の中から要素を 1 個選んで返す関数である．

設問 4 関数 extend を

$$\text{extend}(L_i) = \{s \mid \text{ある } u \in L_i \text{とある } j \in I \text{ が存在して } s = u \cup \{j\} \text{ かつ } j \notin u\}$$

と定義するとき，Algorithm 1 は $f(s) \geq \sigma$ を満たす s だけを全て出力することを示せ．

設問 5 非負整数 i, σ に対して $F(i, \sigma) = \{s \subseteq I \mid \sharp s = i \text{ かつ } f(s) \geq \sigma\}$ と定義する．$F(i, \sigma)$ の任意の 2 要素 s, u に対して $\sharp(s \cup u) \geq i + 2$ であれば，$F(i + 1, \sigma) = \emptyset$ であることを示せ．

設問 6 関数 extend の定義を

$$\text{extend}(L_i) = \{s \mid \text{ある } u, v \in L_i \text{が存在して } s = u \cup v \text{ かつ } \sharp s = \sharp u + 1\}$$

に変更しても，Algorithm 1 は $f(s) \geq \sigma$ を満たす s だけを全て出力することを示せ．

Question is translated in English in the section below; this translation is given for reference only.

Assume a database D consisting of finite sets of items. In order to make our discussion simple, we assume that all items are non-negative integers in I, where $I = \{1, 2, \ldots, m\}$, and that $D = \{t_1, t_2, \ldots, t_n\}$ $(t_i \subseteq I, i = 1, 2, \ldots, n)$. We define the *frequency* of a set s appearing in D as $f(s) = \sharp\{i \mid s \subseteq t_i\}$, where $\sharp S$ denotes the number of all elements belonging to a set S. We consider how to enumerate all non-empty sets s such that $f(s) \geq \sigma$ where σ is a given non-negative integer.

Question 1 Let $m = 5$, $n = 4$, $D = \{\{1, 2, 4\}, \{2, 3, 4, 5\}, \{1, 2\}, \{2, 4, 5\}\}$, and $\sigma = 2$. Then enumerate all $s \subseteq I$ such that $f(s) \geq \sigma$.

Question 2 Show that the frequency $f(s)$ satisfies the following property:

Property 1 For any sets s_1 and s_2, $s_1 \subseteq s_2$ implies $f(s_1) \geq f(s_2)$.

Question 3 From the list of functions f_1, f_2, f_3, and f_4 defined below, choose all functions satisfying Property 1 and explain why.

1. $f_1(s)$ is the maximum of the items in s.
2. $f_2(s)$ is the minimum of the items in s.
3. $f_3(s)$ is the greatest common divisor of all items in s.
4. $f_4(s)$ is the least common multiple of all items in s.

We define Algorithm 1 as follows:

Algorithm 1

```
i := 1; C_1 := {{1}, {2}, ..., {m}}; L_1 := ∅;
while(C_i ≠ ∅ && i ≤ m)
    while(C_i ≠ ∅)
        s := choose(C_i); C_i := C_i − {s};
        if (f(s) ≥ σ)
            L_i := L_i ∪ {s};
        endif
    endwhile
    C_{i+1} := extend(L_i); L_{i+1} := ∅; i := i + 1;
endwhile
output L_1 ∪ L_2 ∪ ... ∪ L_{i−1};
```

The function "choose" takes a non-empty set S and returns one element in it according to a certain criterion.

Question 4 Let $\text{extend}(L_i) = \{s \mid s = u \cup \{j\} \text{ and } j \notin u \text{ for some } u \in L_i \text{ and } j \in I\}$. Show that Algorithm 1 outputs those and only those sets satisfying $f(s) \geq \sigma$.

Question 5 For non-negative integers i and σ, we define $F(i, \sigma) = \{s \subseteq I \mid \sharp s = i \text{ and } f(s) \geq \sigma\}$. Show that, for any pair s and u in $F(i, \sigma)$, if $\sharp(s \cup u) \geq i + 2$, then $F(i + 1, \sigma) = \emptyset$.

Question 6 We revise the definition of the function "extend" as

$$\text{extend}(L_i) = \{s \mid s = u \cup v \text{ and } \sharp s = \sharp u + 1 \text{ for some } u \text{ and } v \in L_i\}.$$

Show that, even on this definition, Algorithm 1 still outputs those and only those sets satisfying $f(s) \geq \sigma$.

設問 1 論理式の集合 Fm を以下の BNF (Backus-Naur form) で定義する．

$$\mathsf{Fm} ::= \mathsf{PV} \mid \bot \mid \mathsf{Fm} \to \mathsf{Fm}$$

ここで，PV は命題変数の集合とする．以下では A, B などで Fm の要素を表し，$\neg A$ は $A \to \bot$ を表すとする．

論理体系 ND を以下の推論規則で定義する．ここで，Γ は論理式の有限集合を表すとする．

$$\frac{}{\Gamma \cup \{A\} \vdash A} \qquad \frac{\Gamma \cup \{A\} \vdash B}{\Gamma \vdash A \to B} \qquad \frac{\Gamma \vdash A \to B \quad \Gamma \vdash A}{\Gamma \vdash B} \qquad \frac{\Gamma \vdash \bot}{\Gamma \vdash A}$$

以下の判断 1〜3 の ND における導出を与えよ．

1. $\emptyset \vdash A \to (B \to A)$　　2. $\{A \to B\} \vdash (\neg B) \to (\neg A)$　　3. $\emptyset \vdash \neg\neg((\neg\neg A) \to A)$

設問 2 $\mathbb{N}$ を 0 以上の整数の集合とし，集合 Tr を以下の BNF で定義する．

$$\mathsf{Tr} ::= \mathsf{Lf}(\mathbb{N}) \mid \mathsf{Br}\langle \mathsf{Tr}, \mathsf{Tr} \rangle$$

関数 $\mathsf{sum} : \mathsf{Tr} \to \mathbb{N}$ と $\mathsf{step} : \mathsf{Tr} \to \mathsf{Tr}$ を以下で定義する．

$$\mathsf{sum}(\mathsf{Lf}(n)) = n \qquad \mathsf{sum}(\mathsf{Br}\langle t, s \rangle) = \mathsf{sum}(t) + \mathsf{sum}(s)$$

$$\begin{aligned} \mathsf{step}(\mathsf{Lf}(n)) &= \mathsf{Lf}(n) \\ \mathsf{step}(\mathsf{Br}\langle \mathsf{Lf}(n), \mathsf{Lf}(m) \rangle) &= \mathsf{Lf}(n+m) \\ \mathsf{step}(\mathsf{Br}\langle t, s \rangle) &= \mathsf{Br}\langle \mathsf{step}(t), \mathsf{step}(s) \rangle \quad (t \text{ または } s \text{ が Lf でない場合}) \end{aligned}$$

$t \in \mathsf{Tr}$ と $n \in \mathbb{N}$ について，t に関数 step を n 回適用したものを $\mathsf{step}^n(t)$ と書く．とくに $\mathsf{step}^0(t) = t$ である．

1. $t = \mathsf{Br}\langle \mathsf{Br}\langle \mathsf{Br}\langle \mathsf{Lf}(2), \mathsf{Lf}(3) \rangle, \mathsf{Lf}(0) \rangle, \mathsf{Br}\langle \mathsf{Lf}(4), \mathsf{Lf}(1) \rangle \rangle$ とする．
   (a) $\mathsf{sum}(t)$ を求めよ．
   (b) $\mathsf{step}(t)$，$\mathsf{step}^2(t)$, $\mathsf{step}^3(t)$ を求めよ．
2. 任意の $t \in \mathsf{Tr}$ について，$\mathsf{sum}(\mathsf{step}(t)) = \mathsf{sum}(t)$ であることを示せ．
3. 関数 $\mathsf{dep} : \mathsf{Tr} \to \mathbb{N}$ を次で定義する．

$$\mathsf{dep}(\mathsf{Lf}(n)) = 0 \qquad \mathsf{dep}(\mathsf{Br}\langle t, s \rangle) = \max(\mathsf{dep}(t), \mathsf{dep}(s)) + 1$$

   (a) 任意の $t \in \mathsf{Tr}$ について，$\mathsf{dep}(t) > 0$ ならば $\mathsf{dep}(\mathsf{step}(t)) = \mathsf{dep}(t) - 1$ であることを示せ．
   (b) 任意の $t \in \mathsf{Tr}$ について，$\mathsf{step}^{\mathsf{dep}(t)}(t) = \mathsf{Lf}(\mathsf{sum}(t))$ であることを示せ．

Question is translated in English in the section below; this translation is given for reference only.

Question 1 The set of formulas Fm is defined by the following BNF (Backus-Naur form):

$$\mathsf{Fm} ::= \mathsf{PV} \mid \bot \mid \mathsf{Fm} \to \mathsf{Fm},$$

where PV is the set of propositional variables. Elements of Fm are denoted by $A, B, \ldots$. We use an abbreviation $\neg A$ to denote $A \to \bot$.

The logical system ND is defined by the following inference rules, where Γ denotes a finite set of formulas.

$$\frac{}{\Gamma \cup \{A\} \vdash A} \qquad \frac{\Gamma \cup \{A\} \vdash B}{\Gamma \vdash A \to B} \qquad \frac{\Gamma \vdash A \to B \quad \Gamma \vdash A}{\Gamma \vdash B} \qquad \frac{\Gamma \vdash \bot}{\Gamma \vdash A}$$

Give the derivations of the following judgments in ND.

1. $\emptyset \vdash A \to (B \to A)$ 2. $\{A \to B\} \vdash (\neg B) \to (\neg A)$ 3. $\emptyset \vdash \neg\neg((\neg\neg A) \to A)$

Question 2 Let $\mathbb{N}$ be the set of non-negative integers, and Tr be the set defined by the following BNF:

$$\mathsf{Tr} ::= \mathsf{Lf}(\mathbb{N}) \mid \mathsf{Br}\langle \mathsf{Tr}, \mathsf{Tr} \rangle.$$

The functions $\mathsf{sum} : \mathsf{Tr} \to \mathbb{N}$ and $\mathsf{step} : \mathsf{Tr} \to \mathsf{Tr}$ are defined as follows.

$$\mathsf{sum}(\mathsf{Lf}(n)) = n \qquad \mathsf{sum}(\mathsf{Br}\langle t, s \rangle) = \mathsf{sum}(t) + \mathsf{sum}(s)$$

$$\begin{aligned} \mathsf{step}(\mathsf{Lf}(n)) &= \mathsf{Lf}(n) \\ \mathsf{step}(\mathsf{Br}\langle \mathsf{Lf}(n), \mathsf{Lf}(m) \rangle) &= \mathsf{Lf}(n+m) \\ \mathsf{step}(\mathsf{Br}\langle t, s \rangle) &= \mathsf{Br}\langle \mathsf{step}(t), \mathsf{step}(s) \rangle \quad \text{(otherwise)} \end{aligned}$$

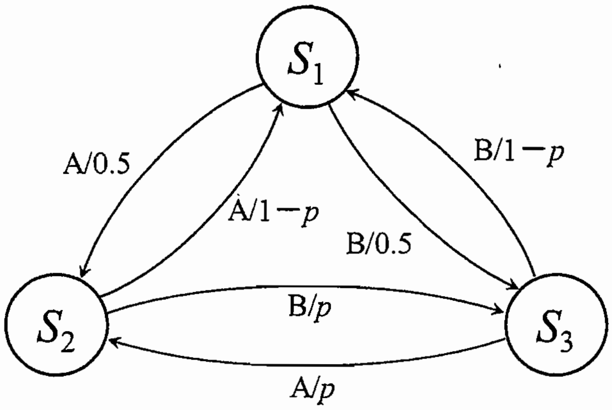
For $t \in \mathsf{Tr}$ and $n \in \mathbb{N}$, the result of n-time applications of step to t is denoted by $\mathsf{step}^n(t)$. In particular, $\mathsf{step}^0(t)$ is t.

1. Let $t = \mathsf{Br}\langle \mathsf{Br}\langle \mathsf{Br}\langle \mathsf{Lf}(2), \mathsf{Lf}(3) \rangle, \mathsf{Lf}(0) \rangle, \mathsf{Br}\langle \mathsf{Lf}(4), \mathsf{Lf}(1) \rangle \rangle$.
   (a) What is the value of $\mathsf{sum}(t)$?
   (b) What are the values of $\mathsf{step}(t)$, $\mathsf{step}^2(t)$, and $\mathsf{step}^3(t)$?
2. Prove that $\mathsf{sum}(\mathsf{step}(t)) = \mathsf{sum}(t)$ holds for any $t \in \mathsf{Tr}$.
3. Let $\mathsf{dep} : \mathsf{Tr} \to \mathbb{N}$ be the function defined by

$$\mathsf{dep}(\mathsf{Lf}(n)) = 0 \qquad \mathsf{dep}(\mathsf{Br}\langle t, s \rangle) = \max(\mathsf{dep}(t), \mathsf{dep}(s)) + 1.$$

   (a) Prove that, for any $t \in \mathsf{Tr}$, $\mathsf{dep}(t) > 0$ implies $\mathsf{dep}(\mathsf{step}(t)) = \mathsf{dep}(t) - 1$.
   (b) Prove that, for any $t \in \mathsf{Tr}$, $\mathsf{step}^{\mathsf{dep}(t)}(t) = \mathsf{Lf}(\mathsf{sum}(t))$ holds.

下図のマルコフ情報源Sについて以下の問いに答えよ．



$(0 \leq p \leq 1)$

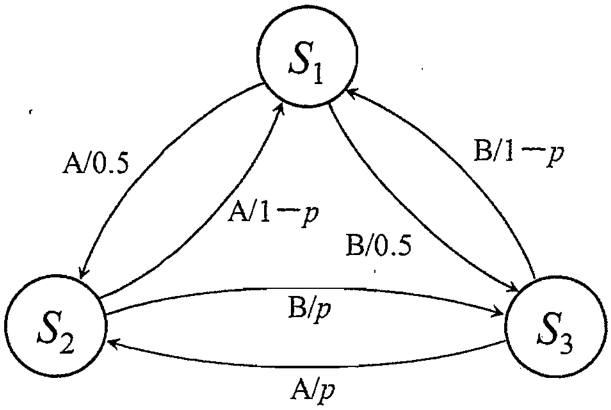
設問１　Sが正規マルコフ情報源であるための，pに関する必要十分条件を示せ．

設問２　Sが正規マルコフ情報源であるとき，1 情報源記号あたりの 1 次エントロピー$H_1(S)$とエントロピー$H(S)$の値を求めよ．ただし，$H(S) = \lim_{n \to \infty} H_n(S)$である．

設問３　Sが正規マルコフ情報源であるとき，pと$H(S)$の関係を，pを横軸，$H(S)$を縦軸とする２次元平面上に表せ．また，$H(S)$の値を最大化するpの値を示せ．

Question is translated in English in the section below; this translation is given for reference only.

Consider a Markov source S with the following transition diagram:



$(0 \leq p \leq 1)$

Question 1 Show a necessary and sufficient condition on p for S to be a regular Markov source.

Question 2 Assume that S is a regular Markov source. Compute the values of the entropy of the first order of S per alphabet $H_1(S)$ and the entropy of S per alphabet $H(S)$, where $H(S) = \lim_{n\to\infty} H_n(S)$.

Question 3 Assume that S is a regular Markov source. Sketch the rough shape of $H(S)$ as a function of p. In addition, show the value of p that maximizes $H(S)$.

判別分析法を用いて濃淡画像を２値化する閾値を求める問題を考える. 画像の画素数を N, 濃度ヒストグラムを $h(k)$ $(k=0,\ldots,m)$ と表記するとき, 以下の問いに答えよ.

設問１　$q(k)$ $(k=0, \ldots, m)$ を濃度値の生起確率とするとき, $q(k)$ を $h(k)$を用いて表せ.

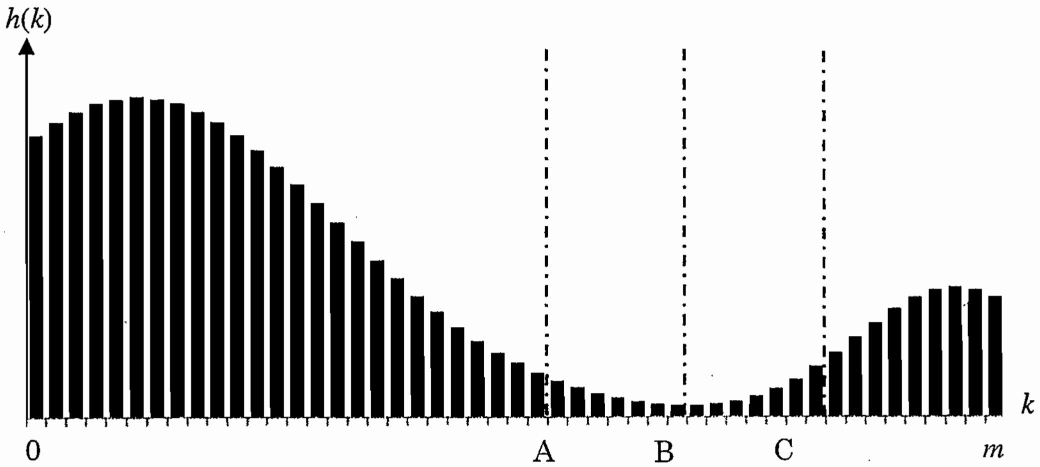
設問２　この画像の濃度値の平均 μ と分散 σ^2 を $q(k)$ を用いて表せ.

設問３　閾値 t $(0 \leqq t \leqq m\text{-}1)$ で濃度ヒストグラムを２つのクラス L $(0 \leqq k \leqq t)$ と H $(t+1 \leqq k \leqq m)$ に分割する. 各クラスの生起確率 $P_L(t)$, $P_H(t)$ を $q(k)$で表せ.

設問４　クラス L と H の濃度値の平均値 μ_L と μ_H, 及び分散 σ^2_L と σ^2_H を求めよ. また, これらを用いて画像の濃度値の平均値 μ と分散 σ^2 を表せ.

設問５　クラス L とクラス H のクラス間分散 $\sigma^2_{LH}(t)$ を求めよ.

設問６　図のようなヒストグラムに判別分析法を適用した場合, 得られる閾値はA, B, Cのいずれに近い値になるか. 理由をつけて説明せよ.



図

Master's Program	Specialized Subjects	[Computer science, Electrical and electronic engineering]	Problem Number	T−5

Question is translated in English in the section below; this translation is given for reference only.

Consider the problem to binarize a grey scale image with discriminant analysis. Let N and $h(k)$ (k=0,...,m) be the number of pixels in the image and the grey value histogram, respectively. Answer the following questions.

Question 1 Let $q(k)$ (k=0, ..., m) be the occurrence probability of the grey value k. Represent $q(k)$ by $h(k)$.

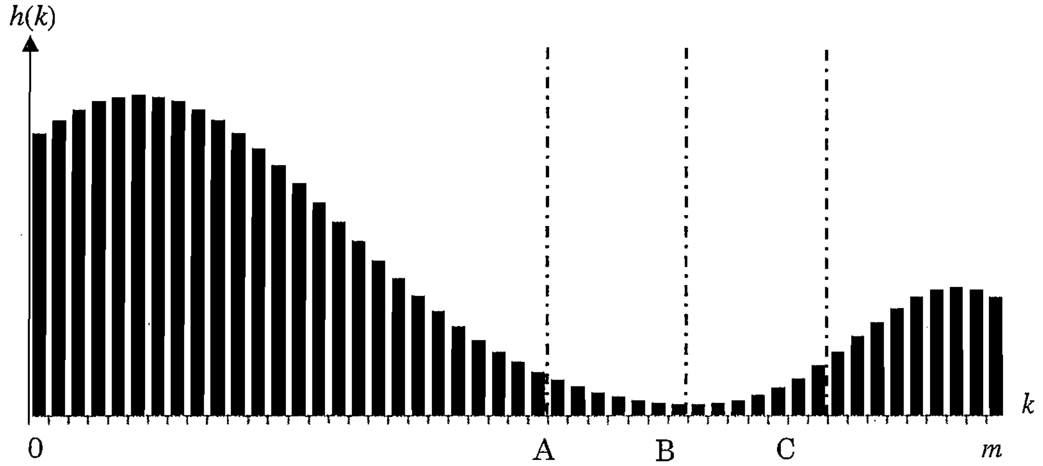
Question 2 Represent the mean μ and variance σ^2 of the grey values in the image by $q(k)$.

Question 3 Suppose we divide the grey value histogram into two classes L($0 \leqq k \leqq t$) and H ($t+1 \leqq k \leqq m$) by the threshold t ($0 \leqq t \leqq m-1$). Represent the occurrence probability of each class $P_L(t)$ and $P_H(t)$ by $q(k)$.

Question 4 Represent the means and variances of grey value of both classes L and H, which are denoted by μ_L and μ_H, and σ^2_L and σ^2_H, respectively. With these, represent the mean μ and the variance σ^2 of the entire image pixel values.

Question 5 Calculate the inter-class variance $\sigma^2_{LH}(t)$.

Question 6 If the discriminant analysis is applied to the histogram shown in the figure below, where would the threshold, closest to A, B or C, lie? Select one of them, and explain the reason.



Figure

2 次元のパターンベクトル $X=(x_1,x_2)$ を、K個のクラスに分類する問題を考える。以下の設問に答えよ。

設問1. N 個のパターンベクトル $\{X^{(1)},\ldots,X^{(N)}\}$ に対して、平均ベクトル $M=(m_1,m_2)$ と共分散行列 $\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}$ を求める式を示せ。これらの各要素に関する式を示すこと。

設問2. 各クラス i $(i=1,\ldots,K)$ 毎に X が $N(M_i,\Sigma_i)$ の正規分布に従うとして、その確率密度関数の対数で定義される識別関数 $g_i(X)$ を示せ。M_i 及び Σ_i（もしくは Σ_i^{-1}）を用いて表記する場合と、それらの要素を用いて表記する場合の両方を示すこと。

設問3. 上記の $g_i(X)$ が X に関する 1 次式になるための条件を述べよ。2 クラス(K=2)の場合について、その際のパターンの分布と決定境界を 2 次元空間に図示せよ。

設問4. 共分散行列 Σ_i が単位行列のときに、$g_i(X)$ が X と M_i との二乗距離に基づく識別関数と等価になることを示せ。

Question is translated in English in the section below; this translation is given for reference only.

Consider the problem of classifying two-dimensional patterns $X = (x_1, x_2)$ into K classes. Answer the following questions.

Question 1: Present the formulae to compute the mean vector $M = (m_1, m_2)$ and the covariance matrix $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ for given N samples $\{X^{(1)}, \ldots, X^{(N)}\}$.

The formulae should be expressed with regard to the components of the vector and matrix.

Question 2: Assume that X follows a Gaussian distribution $N(M_i, \Sigma_i)$ for each class $i\ (i = 1, \ldots, K)$. Present a discriminant function $g_i(X)$ which is defined as a logarithm of the probability density function. The formulae should be expressed with M_i and Σ_i (or Σ_i^{-1}), and also with regard to their components.

Question 3: Describe the condition for $g_i(X)$ to be a linear function of X. Depict the distribution of the patterns and the decision boundary in a two-dimensional space when there are only two classes (K=2).

Question 4: Prove that $g_i(X)$ becomes equivalent to the discriminant function based on the squared Euclidean distance between X and M_i when the covariance matrix Σ_i is an identity matrix.

下記のサンプルデータは，車の「内装」「排気量」「燃費」の属性と人気の関係を調査した結果である．このサンプルデータに対して，ID3 アルゴリズムにより決定木を生成することを考える．この時，以下の設問に答えよ．

表：サンプルデータ

	属性			クラス
	内装	排気量	燃費	人気判定
車 1	高級	大	良	高
車 2	高級	大	中	高
車 3	高級	小	良	高
車 4	普通	小	良	高
車 5	高級	小	中	低
車 6	高級	小	悪	低
車 7	普通	大	悪	低
車 8	普通	小	悪	低

※計算の際には，以下の値を利用せよ．

$$\log_2 3 = 1.58,\ \log_2 5 = 2.32,\ \log_2 6 = 2.58,\ \log_2 7 = 2.81,\ \log_2 9 = 3.17$$

設問 1　クラス分布の平均情報量（エントロピー）を求めよ．

設問 2　サンプルデータを,「内装」「排気量」「燃費」のそれぞれの属性で分類した時の情報利得を求めよ．

設問 3　ID3 アルゴリズムにより得られる決定木を示せ．

設問 4　「オッカムの剃刀」と呼ばれる原理について，その内容と決定木作成の場面における働きを説明せよ．

設問 5　ID3 アルゴリズムを改良した C4.5 アルゴリズムでは，属性による分類時に，情報利得の代わりに情報利得比を用いている．情報利得比を用いる利点について，情報利得を用いた場合と比較して述べよ．

Question is translated in English in the section below; this translation is given for reference only.

The following sample data show the relationship among the attributes, those are "Interior", "Displacement" and "Fuel Expenses", of a car and its reputation. Consider to construct a decision tree for these sample data using the ID3 algorithm. Answer the following questions.

Table: Sample Data

	Attribute			Class
	Interior	Displacement	Fuel Expenses	Reputation
Car1	HighClass	Large	Good	Good
Car2	HighClass	Large	Medium	Good
Car3	HighClass	Small	Good	Good
Car4	NormalClass	Small	Good	Good
Car5	HighClass	Small	Medium	Poor
Car6	HighClass	Small	Bad	Poor
Car7	NormalClass	Large	Bad	Poor
Car8	NormalClass	Small	Bad	Poor

Note: Use the following values in calculation.

$$\log_2 3 = 1.58,\ \log_2 5 = 2.32,\ \log_2 6 = 2.58,\ \log_2 7 = 2.81,\ \log_2 9 = 3.17$$

Question 1 Calculate the entropy of the class distribution.

Question 2 Calculate each information gain when the whole sample data is classified with the attribute "Interior", "Displacement" and "Fuel Expenses", respectively.

Question 3 Show the decision tree obtained using the ID3 algorithm.

Question 4 Explain the principle called "Occam's Razor", and its application in constructing decision trees.

Question 5 The C4.5 algorithm, which is an improved version of the ID3 algorithm, uses the gain ratio instead of the information gain to select an attribute for classification in each step.

Describe the merit of using the gain ratio compared with using the information gain.

計算機において多バイトのデータをメモリ上に配置する方式（バイトオーダ）に関して以下の問いに答えよ。C 言語処理系は負の整数の表現に 2 の補数表示を用いているものとし、short 型は 16bit 整数、long 型は 32bit 整数であるとする。

設問 1

(1) 16 進数で表現された正整数 0xcd15 を 10 進数で表せ。

(2) 16 進数で表現された 16bit 符号付き 2 進整数 0xcd15 を 10 進数で表せ。ここで負の数の表現には 2 の補数表示が用いられているものとする。

(3) 2 進数における 1 の補数表示では、負の数を、絶対値が同じ正整数の 2 進表示の 0 と 1 を反転したもので表す。1 の補数表示で表された 8bit 符号付き 2 進整数 11001011 を 10 進数で表せ。

(4) 1 の補数表示で表された 2 進数の加算で行われる循環桁上げについて説明せよ。

設問 2

(1) バイトオーダがリトルエンディアンである計算機では、多バイトの整数を主記憶に記憶する際に、データの下位のバイトを主記憶上の下位のアドレスに記録する。図 1 の C 言語プログラムをリトルエンディアンの計算機で動作させたときの(A)の printf 関数による出力を示せ。

(2) バイトオーダがビッグエンディアンである計算機では、多バイトの整数を主記憶に記憶する際に、データの下位のバイトを主記憶上の上位のアドレスに記録する。図 1 の C 言語プログラムをビッグエンディアンの計算機で動作させたときの(A)の printf 関数による出力を示せ。

設問 3　IP (Internet Protocol)などで用いられるインターネットチェックサムでは、チェックサムの対象となるバイト列が 16bit 整数になるよう対にされ、それら 16bit 整数の 1 の補数和を求めて、その和の 1 の補数をチェックサムとする。

(1) 図 1 に示す C 言語プログラムにおける関数 cksum が、ポインタ変数 addr が示すアドレスから始まるバイト列のチェックサムを計算することを説明せよ。バイト数は偶数であり変数 count に与えられるものとする。（注：C 言語における 2 項演算子 *a*>>*b* は *a* を *b* bit 右にシフトした数、また単項演算子 ~*a* は *a* の 1 の補数を表す。）

(2) この C 言語プログラムをリトルエンディアンの計算機で動作させたときの(B)の printf 関数による出力を示せ。

(3) この C 言語プログラムの(B)の printf 関数による出力が、バイトオーダがリトルエンディアンであるかビッグエンディアンであるかに依らず同じになることを説明せよ。

```c
#include <stdio.h>
unsigned short cksum ( unsigned short count, unsigned short * addr)  {
  /* Compute Internet Checksum for "count" bytes
   *        beginning at location "addr".  */
  long sum = 0;

  while( count > 1 )  {
    /*  This is the inner loop */
    sum += * addr++;
    count -= 2;
  }

   /*  Fold 32-bit sum to 16 bits */
  while (sum>>16)
    sum = (sum & 0xffff) + (sum >> 16);

  return ~sum;
}

int main (){
  const unsigned char c[8]={0x00,0x01,0xf2,0x03,0xf4,0xf5,0xf6,0xf7};
  unsigned char *a;
  unsigned long b[2];
  unsigned short d;
  int i;

  a = (unsigned char *) b;
  for (i=0; i<8; i++) a[i]=c[i];
  printf ("%08X, %08X¥n", b[0], b[1]); /* (A) */

  d = cksum( 8, (unsigned short *) b) ;
  a = (unsigned char *) &d;
  printf("%02X%02X¥n", a[0], a[1]); /* (B) */

  return 0 ;
}
```

図1

Question is translated in English in the section below; this translation is given for reference only.

Answer the following questions on the byte orders, which is how the multi-byte data are allocated on the main memory of a computer. Here, we assume that the processor of C language uses 2's complement for the representation of negative integers, a 16-bit integer for a "short" type, and a 32-bit integer for a "long" type.

Question 1

(1) Show the decimal representation of a positive integer in hexadecimal representation `0xcd15`.

(2) Show the decimal representation of a 16-bit signed binary integer whose hexadecimal representation is `0xcd15`. Here 2's complement is used for representing negative numbers.

(3) In the representation of binary numbers using 1's complements, a negative integer is represented by a bitwise inverse of a positive integer that has the same absolute value. Show the decimal representation of an 8-bit signed binary integer in 1's complement representation 11001011.

(4) Describe the end-around carry, which is performed in addition of binary numbers in 1's complement representation.

Question 2

(1) In a computer whose byte-order is little endian, the less significant bytes of a multi-byte integer are stored in memory at the lower addresses. Show the output of the `printf` function marked (A) in the C program in Figure 1 when the program is executed on a computer of little endian.

(2) In a computer whose byte-order is big endian, the less significant bytes of a multi-byte integer are stored in memory of the upper addresses. Show the output of the `printf` function marked (A) in the C program in Figure 1 when the program is executed on a computer of big endian.

Question 3 In the computation of the Internet Checksum used in IP (Internet Protocol) etc., adjacent bytes of the target byte stream to be checked are paired to form 16-bit integers, and then the 1's complement sum of these 16-bit integers is calculated. The checksum is the 1's complement of this sum.

(1) Describe how the function `cksum` calculates the Internet Checksum of a byte stream which starts at the address to which the pointer variable `addr` points. Here the length of the byte stream is an even number and given in the variable `count`. (Note that the binary operator *a*>>*b* represents *a* right shifts *b* bits, and the unary operator ~*a* represents the 1's complement of *a*.)

(2) Show the output of the function `printf` marked (B) in the C program in Figure 1.

(3) Describe that the output of the function `printf` in (B) is independent from whether the byte order of the computer is little endian or big endian.

```c
#include <stdio.h>
unsigned short cksum ( unsigned short count, unsigned short * addr)  {
  /* Compute Internet Checksum for "count" bytes
   *        beginning at location "addr".  */
  long sum = 0;

  while( count > 1 )  {
    /*  This is the inner loop */
    sum += * addr++;
    count -= 2;
  }

   /*  Fold 32-bit sum to 16 bits */
  while (sum>>16)
    sum = (sum & 0xffff) + (sum >> 16);

  return ~sum;
}

int main (){
  const unsigned char c[8]={0x00,0x01,0xf2,0x03,0xf4,0xf5,0xf6,0xf7};
  unsigned char *a;
  unsigned long b[2];
  unsigned short d;
  int i;

  a = (unsigned char *) b;
  for (i=0; i<8; i++) a[i]=c[i];
  printf ("%08X, %08X\n", b[0], b[1]); /* (A) */

  d = cksum( 8, (unsigned short *) b) ;
  a = (unsigned char *) &d;
  printf("%02X%02X\n", a[0], a[1]); /* (B) */

  return 0 ;
}
```

Figure 1.

文脈自由文法 $G = (V, T, P, S)$ がある。ここで、S は開始記号である。V と T はそれぞれ非終端記号と終端記号の集合であり、以下のように定義される。

$$
\begin{aligned}
V &= \{S, N_p, V_p, P_p, N_w, V_w, P_w, D_w\} \\
T &= \{\text{time, flies, like, an, arrow}\}
\end{aligned}
$$

さらに、P は以下の要素からなる生成規則の集合である。

$$
\begin{array}{rcl@{\qquad}rcl}
S & \to & V_p \mid N_p V_p & N_w & \to & \text{time} \mid \text{flies} \mid \text{arrow} \\
V_p & \to & V_w \mid V_p N_p \mid V_p P_p & V_w & \to & \text{time} \mid \text{flies} \mid \text{like} \\
N_p & \to & N_w \mid N_w N_w \mid D_w N_w \mid N_p P_p & P_w & \to & \text{like} \\
P_p & \to & P_w N_p & D_w & \to & \text{an}
\end{array}
$$

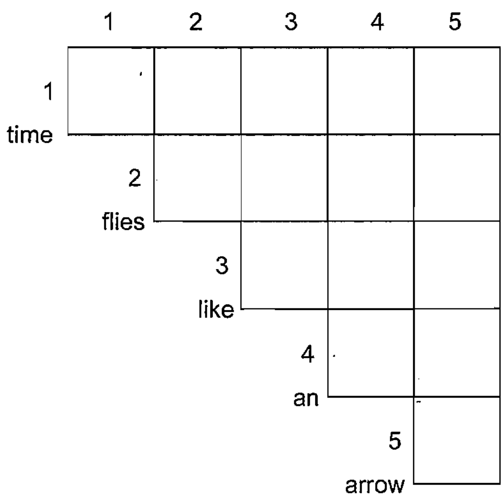
以下の設問に答えよ。

設問 1 次の英文の文法 G による全ての構文木を示し、統語的曖昧性に起因する意味の違いについて説明せよ。ただし、文頭の大文字化と文末のピリオドを無視している。

time flies like an arrow

設問 2 文法 G をチョムスキー標準形 G' に変換せよ。

設問 3 G' によって設問 1 の例文を CKY アルゴリズムを用いて構文解析することを考える。このアルゴリズムでは、下図のような表を用いる。表のセル (i, j) は i 番目から j 番目までの単語の列を覆う全ての可能な非終端記号を保持する。これを完成させよ。



Question is translated in English in the section below; this translation is given for reference only.

Consider a context free grammar $G = (V, T, P, S)$, where S is the start symbol. V and T are a set of non terminals and a set of terminals respectively, which are defined as follows:

$$V = \{S, N_p, V_p, P_p, N_w, V_w, P_w, D_w\}$$
$$T = \{\text{time, flies, like, an, arrow}\}.$$

And P is a set of production rules composed of the following elements:

$$\begin{array}{rcl@{\qquad}rcl} S & \rightarrow & V_p \mid N_p V_p & N_w & \rightarrow & \text{time} \mid \text{flies} \mid \text{arrow} \\ V_p & \rightarrow & V_w \mid V_p N_p \mid V_p P_p & V_w & \rightarrow & \text{time} \mid \text{flies} \mid \text{like} \\ N_p & \rightarrow & N_w \mid N_w N_w \mid D_w N_w \mid N_p P_p & P_w & \rightarrow & \text{like} \\ P_p & \rightarrow & P_w N_p & D_w & \rightarrow & \text{an} \end{array}$$

Answer the following questions.

Question 1 Enumerate all the possible syntactic trees of the following example sentence in English by the grammar G, and explain the difference of the meanings caused by the syntactic ambiguity. Note that the example sentence ignores the initial capitalization and the period.

time flies like an arrow

Question 2 Change the grammar G into a Chomsky normal form G'.

Question 3 According to G', consider the syntactic analysis of the example sentence in Question 1 using the CKY algorithm. This algorithm uses the table below. The cell (i, j) in this table contains all the possible non terminals covering the sequence from the i-th word to the j-th word. Fill out the table.

	1	2	3	4	5
1 time					
2 flies					
3 like					
4 an					
5 arrow					

連結な単純無向グラフ $G(V,E)$ の各辺 $e \in E$ に長さ $l(e) \geq 0$ が与えられているとき，始点 $s \in V$ から，他のすべての頂点 v への最短路は次のアルゴリズムで求められる．

アルゴリズム SHORTEST-PATH

Step 1. $d(s) \leftarrow 0$, $d(v) \leftarrow \infty$ $(v \in V - \{s\})$, $A \leftarrow \{s\}$.

Step 2. A が空なら終了．そうでなければ A の中の頂点から d の値が最小のものを 1 つ取り出し，それを v とする．(ここで，Step 3 で与えられる $prev$ を順にたどることによって v への最短路が決定する．)

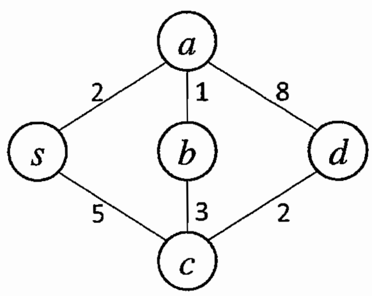
Step 3. v に隣接する各頂点 w に対して
　if $d(w) = \infty$ **then** w を A に追加し，$d(w) \leftarrow d(v) + l(v,w)$, $prev(w) \leftarrow v$.
　elseif $\boxed{\quad 1 \quad}$ **then** $\boxed{\quad 2 \quad}$.

Step 4. Step 2 に戻る．

以下の設問に答えよ．

設問 1　$\boxed{1}$, $\boxed{2}$ にあてはまる記述を示せ．

設問 2　このアルゴリズムを以下のグラフに適用し，各頂点への最短路を求めよ．計算の途中経過を示すこと．



設問 3　このアルゴリズムで得られる s から他のすべての頂点への最短路の和は，G の全域木となることを示せ．

設問 4　このアルゴリズムは $l(e) < 0$ の辺があると正しく動作しないことを，例を挙げて示せ．

設問 5　このアルゴリズムを有向グラフに対して適用できるように変更せよ．

Question is translated in English in the section below; this translation is given for reference only.

Given a connected undirected simple graph $G(V, E)$ with a length $l(e) \geq 0$ for each edge, $e \in E$, a shortest path from a designated starting node $s \in V$ to every other node v can be found by the following algorithm.

Algorithm SHORTEST-PATH

Step 1. $d(s) \leftarrow 0,\ d(v) \leftarrow \infty\ (v \in V - \{s\}),\ A \leftarrow \{s\}$.

Step 2. If A is empty, then quit. Otherwise, delete a node v from A which has the minimum d value. (A shortest path to v is determined at this time by traversing *prev* given in Step 3.)

Step 3. For each node w adjacent to v

if $d(w) = \infty$ **then** add w to A, $d(w) \leftarrow d(v) + l(v, w)$, $prev(w) \leftarrow v$.

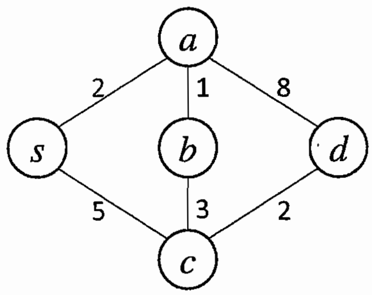
elseif $\boxed{\quad 1 \quad}$ **then** $\boxed{\quad 2 \quad}$.

Step 4. Go to Step 2.

Answer the following questions.

Question 1. Fill in the blanks $\boxed{\ 1\ }$ and $\boxed{\ 2\ }$.

Question 2. Apply the algorithm to the following graph, and detect a shortest path to every node. Show the intermediate calculations.



Question 3. Explain that the union of the shortest paths from s to other nodes detected by the algorithm is a spanning tree of G.

Question 4. Explain that the algorithm does not work properly when an edge of $l(e) < 0$ exists by showing a concrete example.

Question 5. Modify the algorithm to handle directed graphs.

デルタ関数$\delta(t)$から作られる関数を$\delta_K(t)=\sum_{n=-\infty}^{\infty}\delta(t-nK)$（$n$は整数、$K$は正の実数）としたとき、周期関数の性質を利用することによって$\delta_K(t)=\dfrac{1}{K}\sum_{n=-\infty}^{\infty}e^{jn\omega_K t}$（$j$は虚数単位、$\omega_K=2\pi/K$）と表される。以下の設問に答えなさい。

設問 1　アナログ信号$x(t)$のフーリエ変換を$X(\omega)$、$x(t)$を角周波数ω_sで標本化した信号を$x_s(t)$、そのフーリエ変換を$X_s(\omega)$としたとき、$x_s(t)$および$X_s(\omega)$をそれぞれ$x(t)$、$X(\omega)$を用いて表しなさい。

設問 2　$T=2\pi/\omega_s$とし、$x[n]=x(nT)$（nは整数）によって離散時間信号$x[n]$を作り、$x_c(t)=\sum_{n=-\infty}^{\infty}x[n]\delta(t-nT)$としたとき、$x_c(t)$のフーリエ変換$X_c(\omega)$を求めなさい。

設問 3　$X_c(\omega)$と$X_s(\omega)$を比較し、離散時間信号の持つ周波数特性の特徴について述べなさい。

設問 4　観測したアナログ信号が$x(t)=A\cos(\omega_0 t+\alpha)+B$（$A$、$\omega_0$、$\alpha$、$B$は未知の実数パラメータ）と表されることが分かっている場合、$x(t)$を標本化し、離散フーリエ変換を用いて未知の周波数ω_0を求めようとすると幾つかの問題点が生じる。どのような問題点が生じるかを述べなさい。

Question is translated in English in the section below; this translation is given for reference only.

Let $\delta_K(t) = \sum_{n=-\infty}^{\infty} \delta(t-nK)$ (n : integer, K : positive real number) denote a function defined by the delta function $\delta(t)$. Based on its periodic property, we can derive $\delta_K(t) = \frac{1}{K}\sum_{n=-\infty}^{\infty} e^{jn\omega_K t}$ (j : imaginary unit, $\omega_K = 2\pi/K$). Answer the following questions.

Question 1 Let $x(t)$ and $X(\omega)$ denote an analogue signal and its Fourier transform, respectively. Suppose $x(t)$ is sampled at a sampling angular frequency ω_s to generate $x_s(t)$. Let $X_s(\omega)$ denote the Fourier transform of $x_s(t)$. Describe $x_s(t)$ and $X_s(\omega)$ in terms of $x(t)$ and $X(\omega)$, respectively.

Question 2 Let $T = 2\pi/\omega_s$ denote the sampling interval. Suppose we generate the discrete time signal $x[n]$ by $x[n] = x(nT)$ (n : integer). Derive the Fourier transform $X_c(\omega)$ of $x_c(t) = \sum_{n=-\infty}^{\infty} x[n]\delta(t-nT)$.

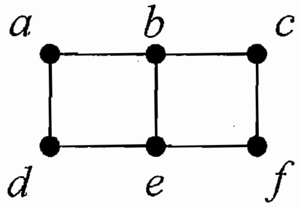
Question 3 Compare $X_c(\omega)$ with $X_s(\omega)$, and discuss the frequency characteristics of discrete time signals.

Question 4 Let $x(t) = A\cos(\omega_0 t + \alpha) + B$ (A, ω_0, α, B : unknown real parameters) denote an observed analogue signal. When we want to estimate an unknown frequency ω_0 by sampling $x(t)$ and applying the discrete Fourier transform, we face several problems. Describe the problems.

本問題では、頂点数が2以上の、連結な単純無向グラフのみを取り扱う。グラフGの頂点集合を$V(G)$で表す。また、頂点$v \in V(G)$に対して、vに隣接する頂点集合を$N(v)$で表す。

Cを$V(G)$の部分集合とする。Gの任意の枝$e=(u,v)$について、$u \in C$または$v \in C$であるとき、CはGの**頂点被覆**であるという。Gの頂点被覆の中で最小サイズのものをGの**最小頂点被覆**といい、そのサイズを$VC(G)$と書く。Dを$V(G)$の部分集合とする。任意の頂点$v \in V(G)$について、$v \in D$または、ある$u \in D$が存在して$v \in N(u)$であるとき、DはGの**支配集合**であるという。Gの支配集合の中で最小サイズのものをGの**最小支配集合**といい、そのサイズを$DS(G)$と書く。以下の問いに答えよ。

設問 1. 以下のグラフG_1に対して、(i) 最小頂点被覆、(ii) 最小支配集合をそれぞれ1つずつ求めよ。ただしa〜fは頂点の名前を表す。



設問 2. 任意のグラフGに対して$VC(G) \geqq DS(G)$であることを証明せよ。

設問 3. $VC(G_2) = DS(G_2)$であるグラフG_2を1つ示せ（解答は設問1のG_1のように図示せよ）。また、G_2に対する最小頂点被覆と最小支配集合をそれぞれ1つずつ示せ。

設問 4. $VC(G_3) > 2\,DS(G_3)$であるグラフG_3を1つ示せ（解答は設問1のG_1のように図示せよ）。また、G_3に対する最小頂点被覆と最小支配集合をそれぞれ1つずつ示せ。

設問 5. 頂点被覆を求める以下のアルゴリズムを考える。

Xを空集合とする。
グラフに枝がなくなるまで、以下を繰り返す。
　{枝を任意に1本選び、その両端の頂点（uとvとする）をXに加える。
　　uに接続する枝とvに接続する枝を全てグラフから削除する。}
Xを出力する。

以下の問いに答えよ。ただしGは入力グラフである。

(1) XがGの頂点被覆になっていることを示せ。

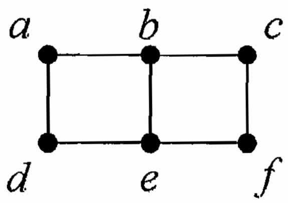
(2) $|X| \leqq 2\,VC(G)$であることを示せ。

Question is translated in English in the section below; this translation is given for reference only.

In this problem, we only consider connected, undirected, and simple graphs with at least two vertices. For a graph G, $V(G)$ denotes the set of vertices of G. For a vertex $v \in V(G)$, let $N(v)$ be the set of vertices adjacent to v.

Let C be a subset of $V(G)$. If, for any edge $e = (u, v)$ of G, $u \in C$ or $v \in C$, then C is called a **vertex cover** of G. A vertex cover of G with the minimum size is called a **minimum vertex cover** of G, and its size is denoted by $VC(G)$. Let D be a subset of $V(G)$. If, for any vertex $v \in V(G)$, $v \in D$ or $v \in N(u)$ for some $u \in D$, then D is called a **dominating set** of G. A dominating set of G with the minimum size is called a **minimum dominating set** of G, and its size is denoted by $DS(G)$.

Question 1. Find (i) a minimum vertex cover and (ii) a minimum dominating set of the following graph G_1. Here, "a" through "f" denote the names of the vertices.



Question 2. Prove that $VC(G) \geqq DS(G)$ for an arbitrary graph G.

Question 3. Give an example graph G_2 such that $VC(G_2) = DS(G_2)$ and draw it. Also, show a minimum vertex cover and a minimum dominating set of G_2.

Question 4. Give an example graph G_3 such that $VC(G_3) > 2\,DS(G_3)$ and draw it. Also, show a minimum vertex cover and a minimum dominating set of G_3.

Question 5. Consider the following algorithm for finding a vertex cover of a given graph.

Let X be the empty set.

Repeat the following process until all the edges are removed.

{Choose an arbitrary edge, then add both of its endpoints, say u and v, to X.
Remove all the edges incident to u and all the edges incident to v.}

Output X.

Answer the following questions. Here, G is an input graph.

(1) Show that X is a vertex cover of G.

(2) Show that $|X| \leqq 2\,VC(G)$.

設問1　以下の問いに答えよ．

(a) 3行3列の行列 $\boldsymbol{A} = \begin{pmatrix} 0 & 2 & -2 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix}$ を対角化せよ．計算過程も示すこと．

(b) 以下の漸化式によって定義された数列 x_n, y_n, z_n を n の式で表せ．ただし n は非負の整数，$x_0 = y_0 = z_0 = 1$ とする．

$$\begin{aligned} x_{n+1} &= 2y_n - 2z_n \\ y_{n+1} &= x_n + y_n + z_n \\ z_{n+1} &= -x_n + y_n + z_n \end{aligned}$$

設問2　実関数の導関数について以下の設問に答えよ．

(a) e^x の導関数が e^x であることを利用して，$\log x$ の導関数が $\frac{1}{x}$ であることを導け．

(b) $\operatorname{arcsinh} x$ の導関数を求めよ．ただし，$\operatorname{arcsinh} x$ は $\sinh x$ の逆関数である．なお，$\sinh x = \frac{e^x - e^{-x}}{2}$，$\cosh x = \frac{e^x + e^{-x}}{2}$ であることを用いてもよい．

設問3　連続確率分布の一つであるベータ分布は次の確率密度関数で表される．

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \quad (0 < x < 1), \qquad 0 \quad (\text{otherwise})$$

ただし，$\alpha > 0, \beta > 0$ である．$B(\alpha,\beta)$ はベータ関数であり，次のように表される．

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

ベータ分布の最頻値を求めよ．

Question is translated in English in the section below; this translation is given for reference only.

Question 1 Answer the following questions.

(a) Diagonalize the 3×3 matrix

$$A = \begin{pmatrix} 0 & 2 & -2 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix}.$$

Detail your computation.

(b) Find the explicit formulas in n of x_n, y_n and z_n defined by the recurrence relations

$$x_{n+1} = 2y_n - 2z_n,$$
$$y_{n+1} = x_n + y_n + z_n,$$
$$z_{n+1} = -x_n + y_n + z_n,$$

where n is a non-negative integer, and $x_0 = y_0 = z_0 = 1$.

Question 2 Answer the following questions about the derivative of real functions.

(a) Prove that the derivative of $\log x$ is $\frac{1}{x}$ using the fact that the derivative of e^x is e^x.

(b) Describe the derivative of $\operatorname{arcsinh} x$. Note that $\operatorname{arcsinh} x$ is the inverse function of $\sinh x$. Also, you can use the following formulas: $\sinh x = \frac{e^x - e^{-x}}{2}$, $\cosh x = \frac{e^x + e^{-x}}{2}$.

Question 3 The beta distribution, which is a continuous probability distribution, has the following probability density function:

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad (0 < x < 1), \qquad 0 \ \text{(otherwise)},$$

where the parameters α and β satisfy $\alpha > 0, \beta > 0$. $B(\alpha, \beta)$ is the beta function, which has the following form:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx.$$

Describe the mode of the beta distribution.