ゲノムリテラシー講座 演習問題解答

加藤 毅
http://www.net-machine.net/~kato/

平成 21 年 7 月 25 日

*平成 19 年度 問 21

以下に示した 6 個の数値のデータが与えられたとする．このデータにおける 3 個の基本統計量（平均値，中央値，最頻値）の値として正しい組み合わせを選択肢の中から一つ選べ．

ここで中央値はメジアン，最頻値はモードとも呼ばれる．

データ：2, 3, 8, 2, 6, 9

	平均値	中央値	最頻値
1	5	4.5	2
2	5	5	5
3	4.5	4.5	2
4	4.5	2	5

解説

平均値，中央値，最頻値は，確率変数に対する統計量であるが，ここでは，標本集合に対する標本平均値，標本中央値，標本最頻値をさしている．

サイズ n の標本集合 $\mathcal{X} = (x_1, \ldots, x_n)$ が所与のとき，その標本平均値は

$$\frac{1}{n}\sum_{i=1}^{n} x_i$$

で定義される．

標本中央値は，標本を

$$x'_1 \leq \cdots \leq x'_n$$

のように並べ替えたものを使って，

$$\begin{cases} x'_{\frac{n+1}{2}} & (n \text{ が奇数のとき}) \\ \frac{1}{2}\left(x'_{\frac{n}{2}} + x'_{\frac{n+1}{2}}\right) & (n \text{ が偶数のとき}) \end{cases}$$

で定義される．

標本最頻値は，標本集合の中で最も頻度の多かった値で定義される．

では，標本集合 $\mathcal{X} = \{2, 3, 8, 2, 6, 9\}$ に対しては，標本平均値は

$$\frac{1}{6}(2+3+8+2+6+9) = 5$$

となる．

この標本集合を昇順に整列すると 2, 2, 3, 6, 8, 9 となるので，標本中央値は

$$\frac{1}{2}(3+6) = 4.5$$

となる.

値 2 の頻度が 2 で最も多いので，標本最頻値は

$$2$$

となる.

よって，選択肢 1 が正しい.

*平成 20 年度 問 35

それぞれ，平均 0，分散 1 の標準正規分布に従う独立な確率変数 X と Y に関する記述のうち，不適切なものを選択肢の中から一つ選べ．

1 X と X^2 の相関係数は 0 である．
2 $X+Y$ の分散は 1 である．
3 $X+Y$ は正規分布に従う．
4 X と Y の相関係数は 0 である．

解説

確率密度関数 $p(X,Y)$ に従う 2 つの確率変数 X，Y に対して，2 変数関数 $f(X,Y)$ の期待値は

$$E(f(X,Y)) = \int_X \int_Y dXdY p(X,Y) f(X,Y)$$

と定義される．

この期待値を使って以下の統計量が定義されている：

$f(X)$ の分散:

$$V(f(X),g(X)) = E(f(X)^2) - (E(f(X)))^2$$

$f(X)$ と $g(X)$ の共分散:

$$\mathrm{cov}(f(X),g(X)) = E(f(X)g(X)) - E(f(X))E(g(X)),$$

$f(X)$ と $g(X)$ の相関係数:

$$\rho(f(X),g(X)) = \frac{\mathrm{cov}(f(X),g(X))}{\sqrt{V(f(X))V(g(X))}}$$

正規分布は確率分布の一つで，その確率密度関数は，2 つのパラメータ μ，σ^2 を用いて

$$p(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X-\mu)^2}{2\sigma^2}\right)$$

で表される．

この 2 つのパラメータ μ,σ^2 は特に**平均**，**分散**と呼ばれている．

平均 μ，分散 σ^2 を持つ正規分布は $\mathcal{N}(\mu,\sigma^2)$ と表される．

1 次モーメント，2 次モーメント，3 次モーメントは

$$E(X) = \mu, \qquad E(X^2) = \sigma^2 + \mu^2, \qquad E(X^3) = \mu^3 + 3\mu\sigma^2$$

で表される．$\mu = 0$，$\sigma^2 = 1$ とおいた正規分布 $\mathcal{N}(0,1)$ は**標準正規分布**と呼ばれる．

$\mathcal{N}(\mu_U,\sigma_U^2)$ に従う確率変数 U と $\mathcal{N}(\mu_V,\sigma_V^2)$ に従う確率変数 V を考え，この 2 つの確率変数は互いに独立であるとする．

任意のスカラー a，b に対して，線形結合 $aU+bV$ は正規分布 $\mathcal{N}(a\mu_U+b\mu_V, a^2\sigma_U^2+b^2\sigma_V^2)$ に従うことが知られている．

選択肢 1 X と X^2 の共分散は

$$\begin{aligned}\mathrm{cov}(X,X^2) &= E(XX^2) - E(X)E(X^2)\\ &= E(X^3) - E(X)E(X^2)\\ &= 0 - 0 = 0\end{aligned}$$

である．

よって，X と X^2 の相関係数は，

$$\rho(X,X^2) = \frac{\mathrm{cov}(X,X^2)}{\sqrt{V(X)V(X^2)}} = \frac{0}{\sqrt{V(X)V(X^2)}} = 0$$

を得る．よって，この選択肢は正しい．

選択肢 2　$X+Y$ は $\mathcal{N}(0,2)$ に従うので，分散は 2 である．よって，この選択肢には間違いがある．

選択肢 3　$X+Y$ は $\mathcal{N}(0,2)$ に従う．よって，この選択肢は正しい．

選択肢 4　X と Y が独立なとき，

$$\begin{aligned} E(XY) &= \int_X \int_Y dXdYp(X,Y)XY \\ &= \int_X \int_Y dXdYp(X)p(Y)XY \\ &= \left(\int_X dXp(X)X\right)\left(\int_Y dYp(Y)Y\right) \\ &= E(X)E(Y) \end{aligned}$$

となる．

よって，その共分散は

$$\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y) = 0$$

になる．

したがって，相関係数は

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{0}{\sqrt{V(X)V(Y)}} = 0$$

となる．

よって，この選択肢は正しい．

*平成 19 年度 問 23

１つのサンプルから２つの測定値（例，ある人の身長と体重，ある人の国語の成績と算数の成績など）が得られる場合，この２つの測定値に相関関係があるかどうかを判定する方法に相関係数 r がある．１つのサンプルが，x と y の２つの測定値をもつような集団に対して測定を行い，測定結果を散布図を用いて表現した結果を以下に示した．このとき，相関係数 r の値は，どのような範囲の値を持つと考えられるか．適当なものを選択肢の中から一つ選べ．

1　$1 < r$
2　$0 < r \leq 1$
3　$r = 0$
4　$-1 \leq r < 0$

解説

２変量の標本集合

$$\mathcal{X} = \{(x_1, y_1), \ldots, (x_\ell, y_\ell)\}$$

から，標本相関係数は

$$\rho = \frac{\sum_{i=1}^{\ell}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{\ell}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{\ell}(y_i - \bar{y})^2}}$$

で表される．

ただし，$\bar{x}$ および $\bar{y}$ は，標本平均

$$\bar{x} = \frac{1}{\ell}\sum_{i=1}^{\ell} x_i, \qquad \bar{y} = \frac{1}{\ell}\sum_{i=1}^{\ell} y_i$$

を表す．

この定義より標本相関係数は

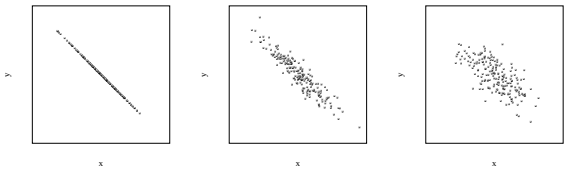
$$-1 \leq \rho \leq +1$$

を満たす．

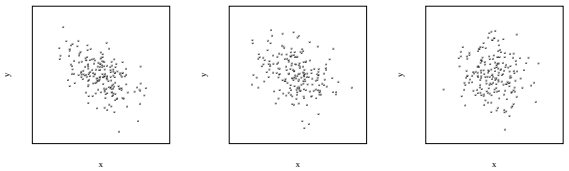
この定義に基づいて標本相関係数を計算すると，次のような傾向がある：

- x が大きくなるほど y も大きくなるとき，$\rho > 0$ になる傾向にある．
- x が大きくなるほど y も小さくなるとき，$\rho < 0$ になる傾向にある．
- x と y が無関係な分布をしているとき，$\rho = 0$ に近い値になる傾向にある．

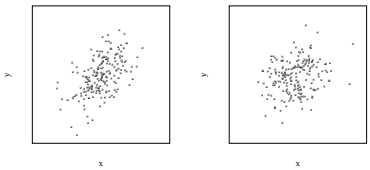
以下に例を示す：



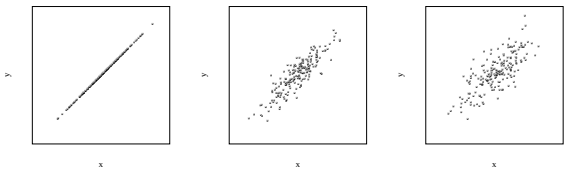
$\rho = -1$ $\rho = -0.9$ $\rho = -0.75$



$\rho = -0.5$ $\rho = -0.25$ $\rho = 0$



$\rho = +0.5$ $\rho = +0.25$



$\rho = +1$ $\rho = +0.9$ $\rho = +0.75$

よって，選択肢 4 が正しい．

*平成 19 年度 問 24

次に示した 2 つの確率変数 X, Y に関する記述において，正しくないものはどれか．一つ選べ．
ここで，それぞれの記号は，次のように用いられている．

$P(A)$	A が起きる確率,
$P(A,B)$	A と B が同時に起きる確率,
$V(A)$	A の分散,
$\sigma(A)$	A の標準偏差,
$E(A)$	A の平均値.

1 X と Y が独立のとき，X と Y が同時に起きる確率 $P(X,Y)$ は，それぞれの起きる確率 $P(X)$ と $P(Y)$ の積に等しい．
2 Y の標準偏差 $\sigma(Y)$ は，分散 $V(Y)$ の二乗に等しい．
3 Y が起こったという条件の下で X が起きる確率 (条件付確率) $P(X|Y)$ は，$\frac{P(X,Y)}{P(Y)}$ で表すことができる．
4 X の分散 $V(X)$ は，$E(X^2) - (E(X))^2$ と表すことができる．

解説

確率分布 $P(X)$ に従う確率変数 X に対して，関数 $f(X)$ の値の期待値は

$$E(f(X)) = \sum_X p(X)f(X)$$

と定義される．
この期待値を使って，様々な統計量が定義されている．
代表的なものとして，

$$\begin{aligned} X \text{ の平均 } M(X) &= E(X) \\ X \text{ の分散 } V(X) &= E(X^2) - (E(X))^2 \\ X \text{ の標準偏差 } \sigma(X) &= \sqrt{V(X)} \end{aligned}$$

がある．

選択肢 1 2つの確率変数 X と Y に対して，

$$P(X,Y) = P(X)P(Y)$$

が成り立つとき，X と Y は**統計的独立**である，という．よって，この選択肢は正しい．

選択肢 2 この選択肢には誤りがある．正しくは「Y の分散 $V(Y)$ は，標準偏差 $\sigma(Y)$ の二乗に等しい.」

選択肢 3 **ベイズの定理**より，

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

が成り立つ．よって，この選択肢は正しい．

選択肢 4 分散の定義より，この選択肢は正しい．

*平成 20 年度 問 33

コインを投げ，表が出たか，裏が出たかを調べるという試行を繰り返す．このコインの表，裏が出る確率はそれぞれ 1/2 であり，各試行は独立であるとする．この試行に関連する記述として不適切なものを選択肢の中から一つ選べ．

1 この試行を繰り返したとき，10 回続けて表が出た．このとき，その次の試行では裏が出る確率が高い．
2 このコイン投げを 100 回行ったとき，表が出る回数の期待値は 50 回である．
3 5 回の試行，表が 3 回，裏が 2 回出る確率は，

$$\frac{5!}{3!2!}\left(\frac{1}{2}\right)^5$$

と計算される．
4 この問題では，コインの表あるいは裏が出ること以外の事象（コインが立つ等）の確率は 0 と仮定されている．

解説

コイン投げはベルヌーイ試行の一つであり，その確率分布はベルヌーイ分布と呼ばれる．

ベルヌーイ分布に従う確率変数 x は，0 か 1 の値をとり，パラメータ π を使って

$$\begin{aligned} p(x) &= \begin{cases} \pi & \text{if } x = 1, \\ 1-\pi & \text{if } x = 0, \end{cases} \\ &= \pi^x (1-\pi)^{1-x} \end{aligned}$$

で表される．

この設問の場合，表が出る事象に $x = 1$ を割り当て，裏が出る事象に $x = 0$ を割り当てるとすると，$\pi = 1/2$ とおくことができる．

1 回の試行で表が出る回数の期待値は

$$\begin{aligned} E(x) &= \sum_x p(x)x \\ &= \pi \cdot 1 + (1-\pi)\cdot 0 = \pi \end{aligned}$$

と表される．

選択肢 1 各試行は独立と仮定されているので，確率は前回までの結果に依存しない．

よって，この選択肢に誤りがある．

選択肢 2 直感的にも平均 50 回表が出ると分かるが，数学的には次のように証明できる．

確率 $p(x)$ に従う確率変数 x に対して，x の期待値は，

$$E(x) = \sum_x p(x)x$$

と定義される．

今回の場合，100 回試行を行ったので，100 個のベルヌーイ分布に従う確率変数 $x_1, \ldots, x_{100}$ を考えると，表が出る回数は

$$\sum_{i=1}^{100} x_i$$

で表すことができる．

この期待値をとると，

$$
\begin{aligned}
&E\left(\sum_{i=1}^{100} x_i\right) \\
&= \sum_{i=1}^{100} E\left(x_i\right) \\
&= \sum_{i=1}^{100} \pi = 100\pi = 100\frac{1}{2} = 50
\end{aligned}
$$

を得る．

選択肢 3　ベルヌーイ試行を n 回繰り返して，ちょうど k 回成功する (表が出る) 確率は，

$$\frac{n!}{k!(n-k)!}\pi^k(1-\pi)^{n-k}$$

で表されることが知られており，特に**２項分布**と呼ばれる．

この設問では，$n = 5$，$k = 3$ の場合なので，代入すると，確かに，5 回の試行で，表が 3 回，裏が 2 回出る確率は，

$$\frac{5!}{3!2!}\left(\frac{1}{2}\right)^5$$

と表される．

選択肢 4　コインの表，裏が出る確率はそれぞれ 1/2 なので，その和は 1 である．

よってそれ以外の事象が生起する確率は 0 である．

*平成 19 年度 問 34

サンプルに対する測定結果を用いて，サンプルについての性質（例，罹患，非罹患など）の陽性もしくは陰性を予測するできる計算手法があるとする．以下の表は，独立な 100 件のサンプルについて，この手法を用いて事前に得た予測結果と，厳密な実験で確認した陽性もしくは陰性の確定結果を比較し，それぞれの場合ごとの件数を示したものである．この手法に関する記述として，適切ではないものを選択肢から一つ選べ．

		確定結果	
		陽性	陰性
予測結果	陽性	48 人	12 人
	陰性	2 人	38 人

1 確定結果が陽性であるものについて，およそ 96% の確率で正しく陽性と予測できる．
2 陽性と予測されるサンプルのうちおよそ 20% は，実際には陰性である．
3 陰性と予測されるサンプルのうちおよそ 5% は，実際には陽性である．
4 確定結果が陰性であるものについて，およそ 48% の確率で正しく陰性と予測できる．

解説

選択肢 1 確定結果が陽性である人に対して，正しく陽性で予測できた割合は

$$\frac{48}{48+2} = 0.96$$

である．よって，この選択肢は正しい．

選択肢 2 陽性と予測された人のうち，実際は陰性であった割合は

$$\frac{12}{48+12} = 0.2$$

である．よって，この選択肢は正しい．

選択肢 3 陰性と予測された人のうち，実際は陽性であった割合は

$$\frac{2}{2+38} = 0.05$$

である．よって，この選択肢は正しい．

選択肢 4 確定結果が陰性である人に対して，正しく陰性で予測できた割合は

$$\frac{38}{38+12} = 0.76$$

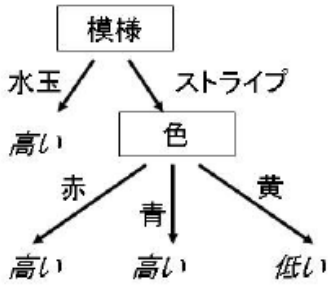
である．よって，この選択肢に誤りがある．

*平成 20 年度 問 39

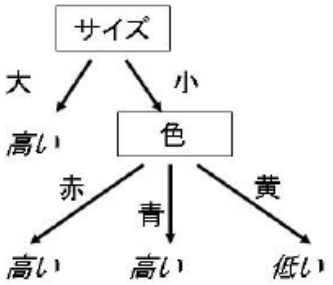
意志の決定プロセスなどをグラフを用いて表現する方法に決定木がある．ある商店において，売られている 6 種類のマグカップの売れ行きを調べた結果，次のような結果が得られた．この表をもとにして売れ行きを予測するための決定木を作成した．もっとも適切なものを選択肢の中から一つ選べ．

	マグカップの特徴 サイズ　色　模様	マグカップの 売れ行き
(a)	大　赤　ストライプ	高い
(b)	小　赤　水玉	高い
(c)	大　青　ストライプ	低い
(d)	大　青　水玉	高い
(e)	小　黄　ストライプ	低い
(f)	大　黄　水玉	低い

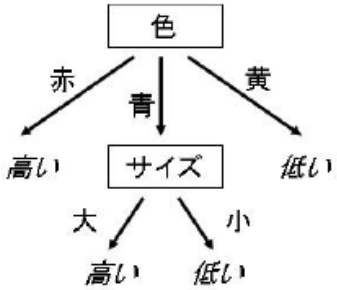
1



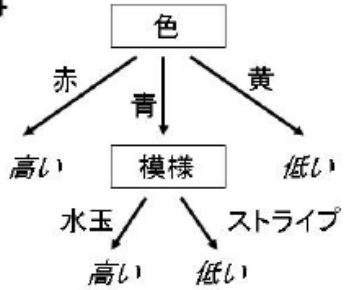
2



3



4



解説

選択肢 1

マグカップの特徴 サイズ　色　模様	マグカップの 売れ行き	
-　-　水玉	高い	(b),(d) には合致するが，(f) に矛盾する．
-　赤　ストライプ	高い	(a) に合致する．
-　青　ストライプ	高い	(c) に矛盾する．
-　黄　ストライプ	低い	(e) に合致する．

選択肢 2

マグカップの特徴 サイズ　色　模様	マグカップの 売れ行き	
大　-　-	高い	(a),(d) には合致するが，(c),(f) に矛盾する．
小　赤　-	高い	(b) に合致する．
小　青　-	高い	該当なし．
小　黄　-	低い	(e) に合致する．

選択肢 3

マグカップの特徴 サイズ 色 模様			マグカップの 売れ行き	
-	赤	-	高い	(a),(b) に合致する.
大	青	-	高い	(d) に合致するが，(c) に矛盾する.
小	青	-	高い	該当なし.
-	黄	-	低い	(e),(f) に合致する.

選択肢 4

マグカップの特徴 サイズ 色 模様			マグカップの 売れ行き	
-	赤	-	高い	(a),(b) に合致する.
-	青	水色	高い	(d) に合致する.
-	青	ストライプ	低い	(c) に合致する.
-	黄	-	低い	(e),(f) に合致する.

*平成 19 年度 問 36

次に示した説明文の中で，予測手法の性能評価の際に行われるクロスバリデーション法 (交差検定法) の説明として，不適切なものはどれか，一つ選べ．

1　クロスバリデーションは，未知データにも対応できるかを検査する目的で行われる．
2　一部のデータを学習に使わずに残しておき，テスト用に用いて予測性能を測定する．
3　leave-one-out 法は，データのうち 2 個のみをテスト用に残しておく方法である．
4　n-fold 法は，データのうち $1/n$ をテスト用に残しておく方法である．

解説

データのモデルを学習する際に，しばしば複数のモデルを考えることができる．
　交差検証法は，汎化性能が最も良いと予測されるモデルを選択するために使われる．
　汎化性能とは，未知データにどれほどフィットするかに関する能力のことをさす．
　交差検証法は基本的には次のように行う．

(1)　すでに得られているデータ集合を訓練用集合と検証用集合に分ける．
(2)　まず，あるモデルに対して，パラメータを訓練用集合にフィットさせる．
(3)　得られたパラメータに対して，検証用集合にどれほどフィットしているか測定する．

　交差検証法として，**n-fold 交差検証法**と**一つ抜き (leave-one-out) 交差検証法**がよく用いられている．

n-fold 交差検証法　データ集合を n 個のグループに分割し，$n-1$ 個のグループを訓練用集合として用い，残りの 1 個のグループを検証用集合を用いるものである．
　すべてのグループが正確に 1 回だけ検証用集合に使われるよう，これを n 回繰り返し，検証用集合に対する結果を平均して，一つの推定値を得る．

一つ抜き (leave-one-out) 交差検証法　n-fold 交差検証法におけるグループ数をサンプル数にした場合，一つ抜き交差検証法となる．
　すなわち，データ集合のうち，1 サンプルだけ検証用に用い，残り全てを訓練用に用いる．
　すべてのサンプルが正確に 1 回だけ検証用に使われるよう繰り返し，それらの結果を平均して，一つの推定値を得る．
　では，選択肢を一つずつ検証する．

選択肢 1　この記述は正しい．

選択肢 2　この記述は正しい．

選択肢 3　leave-one-out 法は，データのうち 1 個のみをテスト用に残しておく方法である．よって，この選択肢は誤りを含む．

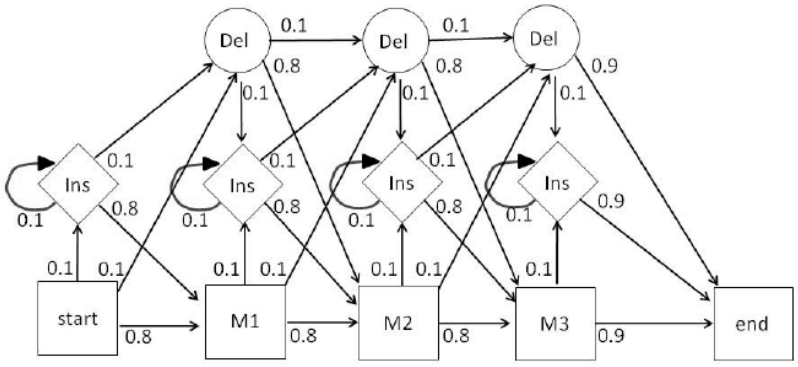
選択肢 4　この記述は正しい．

*平成 20 年度 問 40

下図のようなプロファイル HMM がある．状態間の遷移確率は図中に示されている．各状態での文字の出力確率は，M_1，M_2，M_3 の各状態では以下のようになっている．

$$\begin{array}{lllll} M_1: & P(A)=0.0, & P(T)=0.5, & P(G)=0.5, & P(C)=0.0, \\ M_2: & P(A)=1.0, & P(T)=0.0, & P(G)=0.0, & P(C)=0.0, \\ M_3: & P(A)=0.0, & P(T)=0.0, & P(G)=0.9, & P(C)=0.1, \end{array}$$

また Ins 状態では，位置によらず，$P(A)=P(T)=P(G)=P(C)=0.25$ である．この HMM から,「start $\to T \to A \to G \to A \to$ end 」という出力文字列が観測されたとする．以下の選択肢に示された状態遷移パスのうち，この出力文字列を発生させる確率がもっとも大きいものを一つ選べ．



		T	A	G	A
1	start $\to$	Ins $\to$	$M_1 \to$	$M_2 \to$	$M_3 \to$ end
2	start $\to$	$M_1 \to$	Ins $\to$	$M_2 \to$	$M_3 \to$ end
3	start $\to$	$M_1 \to$	$M_2 \to$	Ins $\to$	$M_3 \to$ end
4	start $\to$	$M_1 \to$	$M_2 \to$	$M_3 \to$	Ins $\to$ end

解説

プロファイル HMM は，マルコフ過程に基づいている．マルコフ過程において，状態系列 $s_{\text{start}}, s_1, \ldots, s_T, s_{\text{end}}$ が与えられたとき，その状態系列から観測系列 $o_1, \ldots, o_T$ が生成される確率は

$$P(o_1, \ldots, o_T \,|\, s_1, \ldots, s_T) = P(o_1 \,|\, s_1) \ldots P(o_T \,|\, s_T)$$

とあらされる．

選択肢 1　状態遷移パス start $\to$ Ins $\to M_1 \to M_2 \to M_3 \to$ end から観測系列 $T \to A \to G \to A$ を生起させる確率は

$$\begin{aligned} P(o_1 \,|\, s_1) \ldots P(o_T \,|\, s_T) &= P(T \,|\, \text{Ins}) P(A \,|\, M_1) P(G \,|\, M_2) P(A \,|\, M_3) \\ &= 0.25 \cdot 0.0 \cdot 0.0 \cdot 0.0 = 0.0 \end{aligned}$$

選択肢 2　状態遷移パス start $\to M_1 \to$ Ins $\to M_2 \to M_3 \to$ end から観測系列 $T \to A \to G \to A$ を生起させる確率は

$$\begin{aligned} P(o_1 \,|\, s_1) \ldots P(o_T \,|\, s_T) &= P(T \,|\, M_1) P(A \,|\, \text{Ins}) P(G \,|\, M_2) P(A \,|\, M_3) \\ &= 0.5 \cdot 0.25 \cdot 0.0 \cdot 0.0 = 0.0 \end{aligned}$$

選択肢 3 状態遷移パス start $\to M_1 \to M_2 \to$ Ins $\to M_3 \to$ end から観測系列 $T \to A \to G \to A$ を生起させる確率は

$$\begin{aligned} P(o_1 \,|\, s_1) \ldots P(o_T \,|\, s_T) &= P(T \,|\, M_1) P(A \,|\, M_2) P(G \,|\, \mathrm{Ins}) P(A \,|\, M_3) \\ &= 0.5 \cdot 1.0 \cdot 0.25 \cdot 0.0 = 0.0 \end{aligned}$$

選択肢 4 状態遷移パス start $\to M_1 \to M_2 \to M_3 \to$ Ins $\to$ end から観測系列 $T \to A \to G \to A$ を生起させる確率は

$$\begin{aligned} P(o_1 \,|\, s_1) \ldots P(o_T \,|\, s_T) &= P(T \,|\, M_1) P(A \,|\, M_2) P(G \,|\, M_3) P(A \,|\, \mathrm{Ins}) \\ &= 0.5 \cdot 1.0 \cdot 0.1 \cdot 0.25 = 1/80 \end{aligned}$$

となる.

よって，選択肢 4 が $T \to A \to G \to A$ を生起させる確率が最も大きい.

*平成 20 年度 問 36

次のクラスタ解析の手法の中で非階層的クラスタリングではないものを選択肢の中から一つ選べ.

1 K-平均法
2 自己組織化マップ
3 Fuzzy c-means
4 UPGMA

解説

UPGMA は階層的クラスタリングの一つで，系統樹を作る．よって，UPGMA が非階層的クラスタリングではない.

*平成 20 年度 問 21

コンピュータ上では，数値は二進数で扱われる．十進数の 2008 を二進数で表現すると「11111011000」となる．ここで，2008 を 4 で割った商である十進数の 502 は，二進数でどのように表現されるか．適切なものを選択肢の中から一つ選べ.

1 111110110
2 111011000
3 111110101
4 111010111

解説

一般に，n 桁の 2 進数で「$x_{n-1}x_{n-2}\ldots x_2x_1x_0$」と表記される数値は

$$x_{n-1}2^{n-1} + x_{n-2}2^{n-2} + \cdots + x_2 2^2 + x_1 2^1 + +x_0 2^0$$

を意味する.

この場合,

$$\begin{aligned}
&\text{「11111011000」} \\
&= 1\cdot 2^{10} + 1\cdot 2^9 + 1\cdot 2^8 + 1\cdot 2^7 + 1\cdot 2^6 + 0\cdot 2^5 + 1\cdot 2^4 + 1\cdot 2^3 + 0\cdot 2^2 + 0\cdot 2^1 + 0\cdot 2^0 \\
&= 2^{10} + 2^9 + 2^8 + 2^7 + 2^6 + 2^4 + 2^3
\end{aligned}$$

となる.

これを 4 で割ると,

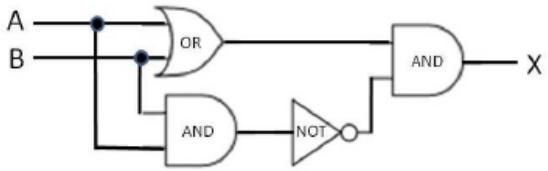
$$\begin{aligned}
&\left(2^{10} + 2^9 + 2^8 + 2^7 + 2^6 + 2^4 + 2^3\right)/4 \\
&= \left(2^{10} + 2^9 + 2^8 + 2^7 + 2^6 + 2^4 + 2^3\right)/2^2 \\
&= 2^{10-2} + 2^{9-2} + 2^{8-2} + 2^{7-2} + 2^{6-2} + 2^{4-2} + 2^{3-2} \\
&= 2^8 + 2^7 + 2^6 + 2^5 + 2^4 + 2^2 + 2^1 \\
&= \text{「111110110」}
\end{aligned}$$

となるので，選択肢 1 が正しい.

*平成 20 年度 問 22

下記の図 (ア) は，いくつかの論理素子を接続した論理回路を表現している．この回路は，A，B は入力であり，X が出力である．この回路に対する入出力の結果によって真理値表（イ）を作成した．ここで，真理値表の (a)，(b) に入る値の組み合わせとして適切なものを選択肢の中から一

つ選べ.



図（ア） 論理回路図

A	B	X
0	0	(a)
0	1	1
1	0	1
1	1	(b)

図（イ） 真理値表

1 (a) $= 0$, (b) $= 0$,
2 (a) $= 0$, (b) $= 1$,
3 (a) $= 1$, (b) $= 0$,
4 (a) $= 1$, (b) $= 1$,

解説

論理回路図より，論理式

$$X = ((A \text{ or } B) \text{ and } (\text{ not } (A \text{ and } B)))$$

を得る.

よって，(a) の値を求めるために，$A = 0$ と $B = 0$ を代入すると，

$$\begin{aligned}(\text{a}) &= ((0 \text{ or } 0) \text{ and } (\text{ not } (0 \text{ and } 0))) \\ &= (0 \text{ and } (\text{ not } 0)) \\ &= (0 \text{ and } 1) \\ &= 0\end{aligned}$$

を得る.

$A = 1$ と $B = 1$ を代入すると，

$$\begin{aligned}(\text{b}) &= ((1 \text{ or } 1) \text{ and } (\text{ not } (1 \text{ and } 1))) \\ &= (1 \text{ and } (\text{ not } 1)) \\ &= (1 \text{ and } 0) \\ &= 0\end{aligned}$$

を得る.

よって，選択肢 1 が正しい.

*平成 20 年度 問 27

データ格納方法としてキューを用いる．アルファベットの文字データをキューに出し入れする．データの操作を

A enque
B enque
C enque
deque
deque
D enque
A enque
deque
deque

のように行ったとき，最後の deque で取り出されるデータとして正しいものを選択肢の中から一つ選べ．ただし，enque はデータを追加する操作，deque はデータを取り出す操作である．

1 A
2 B
3 C
4 D

解説

キューの中身は

A enque
[A]
B enque
[B A]
C enque
[C B A]
deque
[C B] A
deque
[C] B
D enque
[D C]
A enque
[A D C]
deque
[A D] C
deque
[A] D

のようになる．
よって，最後の deque で取り出されるのは D となる．
よって，選択肢 4 が正しい．

*平成 20 年度 問 25

与えられたデータの列を一定の順序に並べ替えるソートアルゴリズムの一つに，バブルソートがある．バブルソートでは，データ列の最初の要素から最後の要素に向かって，隣接する 2 つの要素を順に比較し，2 つの要素が正しい順番に並んでいないときには交換を行うという処理を必要なだけ繰り返していく．データ間の交換がもう必要ないと判断できた場合には，そこで処理を終了するものとする．このバブルソートに関する記述として不適切なものを選択肢の中から一つ選べ．

1 バブルソートは，各種のソーティングアルゴリズムのうち交換法の一種に分類される．
2 最初の要素から最後の要素までの比較が済むと，最後の要素については最小値または最大値として確定するので，次回からは比較の範囲を 1 つ狭めてもよい．
3 データがはじめから正順に並んでいたときは計算が最も早く終了し，逆順に並んでいたときは計算が最も遅くなる．
4 理解しやすい簡便な方法ではあるが，選択法に比べて必要なメモリ容量が多くなるという欠点がある．

解説

バブルソートの算法は以下で与えられる：

```
1: function bubble_sort(A[0 ... N − 1])
2: begin
3: n := N;
4: repeat
5:    swapped := false;
6:    n := n − 1;
7:    for i := 0 to n-1 do
8:       if A[i] > A[i + 1] then
9:          swap(A[i], A[i + 1]);
10:         swapped := true;
11:      end if
12:   end for
13: until not swapped;
14: end.
```

（http://en.wikipedia.org/wiki/Bubble_sortを参考に作成）

選択肢 1 バブルソートは交換法の一種である．よって，この記述は正しい．

選択肢 2 上記にあげた算法も，次の反復では比較の範囲を 1 つ狭めるようになっている．この記述は正しい．

選択肢 3 すでに正順に並んでいることが分かった時点で算法を停止するようになっている．よって，この記述は正しい．

選択肢 4 必要なメモリ容量は，選択ソートとほぼ同じである．よって，この選択肢には誤りがある．

*平成 20 年度 問 26

ソートのアルゴリズムに関する以下の記述について間違っているものを選択肢の中から一つ選べ.

1 与えられたデータ数が n のとき，バブルソートの計算量は $O(n^2)$ である.
2 与えられたデータ数が n のとき，クイックソートの計算量は $O(n \log n)$ である.
3 左から右に数値データを昇順に並べるとき，クイックソートではまずピボット m より小さいデータを m の左に，m より大きいデータを M の右に集める．そのあと，M の左側及び右側に対してクイックソートを再帰的に適用することによりソートを完了する.
4 左から右に数値データを昇順に並べるとき，直接挿入法は，左側から昇順に整列させながら，その整列済みの範囲を左側に向かって徐々に拡大させながら整列する方法である.

解説

クイックソートの算法は以下で与えられる：

```
1: function quicksort (A : リスト)
2: begin
3: B，C は空のリストとする;
4: if A の長さ ≤ 1 then
5:    return A;
6: end if
7: ピボットなる要素 m を選び，A から m を抜く;
8: for all x in A do
9:    if x ≤ m then
10:      リスト B に x を追加する;
11:   else
12:      リスト C に x を追加する;
13:   end if
14: end for
15: B := quicksort(B);
16: C := quicksort(C);
17: return  B，m，C の順番につなげたリスト;
18: end.
```

（http://en.wikipedia.org/wiki/Quicksortを参考に作成）

選択肢 1 この記述は正しい.

選択肢 2 この記述は正しい.

選択肢 3 この記述は正しい.

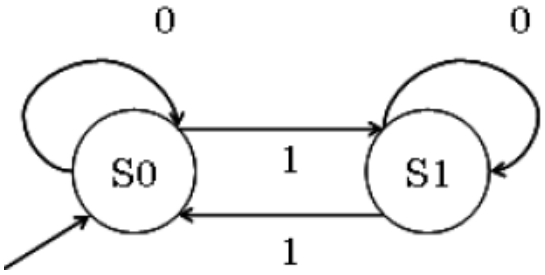
選択肢 4 直接挿入法は，その整列済みの範囲を右側に向かって徐々に拡大させながら整列する方法である．この選択肢に誤りがある.

*平成 20 年度 問 28

下図の決定性有限オートマトンは 2 つの状態 S_0 と S_1 を持つ．S_0 は初期状態である．入力数列は 1 あるいは 0 である．このオートマトンに関する記述で不適切なものを選択肢の中から一つ選べ．

表　状態遷移表

入力状態	1	0
S_0	S_1	S_0
S_1	S_0	S_1



1　入力数列が 00111 のとき，状態は S_1 となる．
2　入力数列中に 1 が偶数個あるときは，状態は S_0 となる．
3　入力数列中に 01010010101 のとき，状態は S_1 となる．
4　入力数列中に 0 が偶数個あるときは，状態は S_1 となる．

解説

選択肢 1　入力系列「00111」によって

$$S_0 \xrightarrow{0} S_0 \xrightarrow{0} S_0 \xrightarrow{1} S_1 \xrightarrow{1} S_0 \xrightarrow{1} S_1$$

と遷移する．よって選択肢 1 は正しい．

選択肢 2　この状態遷移表では，0 のときは現在の状態に留まり，1 のときに状態を変えていることが分かる．
　よって，入力数列中に 1 が偶数個あるときは，状態は S_0 となる．

選択肢 3　入力数列中に 1 が奇数個あるので，状態は S_1 となる．

選択肢 4　例えば，入力数列が 00 のとき，状態は S_0 となるので，この選択肢に誤りがある．
　よって，選択肢 4 が正しい．