統計学 A（理系）試験（2006 年 2 月実施）の解答例

$\boxed{1}$　つぎの問いに答えなさい。

(1) 5 個の標本データ $\{3,5,4,1,7\}$ について、平均と分散を計算しなさい。

(2) この結果を用いて、つぎの 2 種類のデータについて、それらの平均と分散を求めよ。

(2a) 203, 205, 204, 201, 207　　(2b) 12.3, 12.5, 12.4, 12.1, 12.7

（解） 与えられたデータ $\{3,5,4,1,7\}$ を $n=5,\ X_i, i=1,2,\cdots,n$ とおく。平均は $\overline{X}=\dfrac{3+5+4+1+7}{5}=\dfrac{20}{5}=4$ であり、分散については $V(X)=\dfrac{\sum_i X_i^2}{n}-\overline{X}^2=\dfrac{3^2+5^2+4^2+1^2+7^2}{5}-4^2=\dfrac{100}{5}-4^2=4$ と計算できる。これを用いて (2a), (2b) の計算をおこなう。それぞれデータを $\{Y_i\},\{Z_i\}$ とすれば、データの関係式として各 i で、$Y_i=200+X_i,\ Z_i=12+\dfrac{X_i}{10}$ であるから、平均と分散についての関係は, $\overline{Y}=200+\overline{X}$ $V(Y)=V(X)$ (分散はシフトによらないことに注意) $\overline{Z}=12+\overline{X}/10,\ V(Z)=V(X)/10^2$ (分散はスケール係数の 2 乗に比例する) が成り立つ。したがって $\overline{Y}=200+\overline{X}=204,\ V(Y)=V(X)=4$ また $\overline{Z}=12+\overline{X}/10=12.4$ $V(Z)=V(X)/10^2=0.04$ となる。 □

$\boxed{2}$　恒等式 $(1+x)^n(1+x)^2=(1+x)^{n+2}$ を利用し、係数比較により得られる 2 項係数の関係式を導け。

（解） 2 項係数の定義から $(1+x)^n=\sum_{i=0}^n\binom{n}{k}x^k$ これに $(1+x)^2=1+2x+x^2$ を掛けるから

$$
\begin{aligned}
& (1+x)^n(1+x)^2 \\
=\ & \left(\textstyle\sum_{i=0}^n\binom{n}{k}x^k\right)\left(1+2x+x^2\right) \\
=\ & \left(\binom{n}{0}+\binom{n}{1}x+\binom{n}{k}x^2+\cdots+\binom{n}{k}x^n+\cdots+\binom{n}{n}x^n\right)\left(1+2x+x^2\right) \\
=\ & \binom{n}{0}+\binom{n}{1}x+\binom{n}{2}x^2+\cdots+\binom{n}{k}x^k+\cdots+\binom{n}{n}x^n \\
& \quad +2\binom{n}{0}x+2\binom{n}{1}x^2+2\binom{n}{2}x^3+\cdots+2\binom{n}{k-1}x^k+\cdots+2\binom{n}{n}x^{n+1} \\
& \qquad +\binom{n}{0}x^2+\binom{n}{1}x^3+\binom{n}{2}x^4+\cdots+\binom{n}{k-2}x^k+\cdots+\binom{n}{n}x^{n+2}
\end{aligned}
$$

ここでの $x^k, k=0,1,2,\cdots,n+2$ の係数は $\binom{n}{k}+2\binom{n}{k-1}+\binom{n}{k-2}$ である。ただし $k<0, k>n$ ならば、$\binom{n}{k}=0$ に注意。一方 $(1+x)^{n+2}$ を展開することにより、x^k の係数は $\binom{n+2}{k}$ であり、両者を比較することによって

$$\binom{n}{k}+2\binom{n}{k-1}+\binom{n}{k-2}=\binom{n+2}{k}$$

ただし $k<0, k>n$ では $\binom{n}{k}=0$ とする。 □

$\boxed{3}$　確率変数 X は 正規分布 $N(10,4)$ にしたがう。つぎを正規分布表から求めなさい。

(1) $P(\,X\geq 11\,)$　　(2) $P(\,9<X\leq 12\,)$　　(3) $P(\,X<c\,)\geq 0.85$ を満たす c の値。

（解） 新たに確率変数 $Z=\dfrac{X-10}{\sqrt{4}}$ を定めると、仮定からこれは平均 0, 分散 1 の標準正規分布にしたがう。したがって、正規分布表から Z の確率計算ができる。また正規分布の場合には等号を含む不等号であっても、等号のない不等号であっても同じ確率であることに注意する。

(1) $P(X\geq 11)=P\left(\dfrac{X-10}{\sqrt{4}}\geq\dfrac{11-10}{\sqrt{4}}\right)=P(Z\geq 1/2)=1-P(Z\leq 1/2)=1-0.6915=0.3085$

(2) $P(9<X\leq 12)=P\left(\dfrac{9-10}{\sqrt{4}}<\dfrac{X-10}{\sqrt{4}}\leq\dfrac{12-10}{\sqrt{4}}\right)=P(-0.5<Z\leq 1)=P(Z\leq 1)-(1-P(Z\leq 0.5))=0.6413-1+0.6915=0.5328$

(3) 正規分布表から $P(Z\leq 1.04)=0.8508,\ P(Z\leq 1.03)=0.8485$ の関係をもちいる。 この 2 点の線形補間をすると $\dfrac{1.04-1.03}{0.8508-0.8485}(0.85-0.8485)+1.03=1.037$ よって $\dfrac{c-10}{2}=1.037$ より、$c=12.074$ となる。この値 c より大きくすると $P(X<c)$ はより大きくなるから、12.074 以上の値であればよい。 □

$\boxed{4}$ 離散型確率変数 X と Y のとり得る値と確率がつぎの表で与えられている。

X の値 \ Y の値	0	10	20
10	0.20	0.10	0.10
20	0.15	0.02	0.08
30	0.05	0.10	0.20

(1) $X+Y$ のとり得る値と確率を求めよ。

(2) $P(X+Y \leq 20)$ を求めよ。

(3) $E[X+Y]=E[X]+E[Y]$ となるか？

(解) $X+Y$ の値は表の上では

X の値 \ Y の値	0	10	20
10	10	20	30
20	20	30	40
30	30	40	50

となっているから、$X+Y$ のとり得る値は $\{10,20,30,40,50\}$ の 5 通りで

$$P(X+Y=10)=P(X=10,Y=0)=0.20,$$
$$P(X+Y=20)=P(X=20,Y=0)+P(X=10,Y=10)=0.25,$$
$$P(X+Y=30)=P(X=30,Y=0)+P(X=20,Y=10)+P(X=10,Y=20)=0.17,$$
$$P(X+Y=40)=P(X=30,Y=10)+P(X=20,Y=20)=0.18,$$
$$P(X+Y=50)=P(X=30,Y=20)=0.20$$

を得る。つまり

$X+Y$ の値	10	20	30	40	50	計
確率	0.20	0.25	0.17	0.18	0.20	1

したがって $P(X+Y \leq 20)=P(X+Y=10)+P(X+Y=20)=0.20+0.25=0.45$ である。また期待値の計算から

$$E[X+Y]=10\times 0.2+20\times 0.25+30\times 0.17+40\times 0.18+50\times 0.20=29.3$$

上の表から周辺密度も、横（行）の合計から計算でき,

X の値	10	20	30	計
確率	0.40	0.25	0.35	1

, また縦（列）の合計から

Y の値	0	10	20	計
確率	0.40	0.22	0.38	1

を得る。それぞれの期待値も $E[X]=10\times 0.4+20\times 0.25+30\times 0.35=19.5$　また $E[Y]=0\times 0.2+10\times 0.25+20\times 0.17=9.8$ と計算されて、独立ではなくても、和の期待値（平均）はそれぞれの期待値（平均）の和に等しい。 □

$\boxed{5}$ 母平均 μ の推定のため、5 個の標本データ $X_i, i=1,2,3,4,5$ から、3 つの統計量

$$\text{(i)}\ T_1=\frac{X_1+3X_3+X_5}{5} \quad \text{(ii)}\ T_2=\frac{2X_1+X_3+2X_5}{5} \quad \text{(iii)}\ T_3=\frac{X_1+X_2+X_3+X_4+X_5}{5}$$

を考える。これら推定量 $T_i, i=1,2,3$ のもつ性質を比較せよ。

(解) 母平均を μ とし、また母分散を σ^2 と表すと、各 i で $\mu=E[X_i], \sigma^2=V[X_i]=E[X_i^2]-\mu^2$ が成り立っている。3 つの推定量はいずれも不偏推定量である。なぜならいづれの期待値を計算しても $E[T_1]=\dfrac{\mu+3\mu+\mu}{5}=\mu$, $E[T_1]=\dfrac{2\mu+\mu+2\mu}{5}=\mu$, $E[T_1]=\dfrac{\mu+\mu+\mu+\mu+\mu}{5}=\mu$ となるから。つまり係数の和がすべて 1 に等しいからである。しかし分散については大きさが異なる。$V[T_1]=\dfrac{\sigma^2+3^2\sigma^2+\sigma^2}{5^2}=\dfrac{11}{25}=\sigma^2=0.44\sigma^2$ $V[T_2]=\dfrac{2^2\sigma^2+\sigma^2+2^2\sigma^2}{5^2}=\dfrac{9}{25}\sigma^2=0.36\sigma^2$ $V[T_3]=\dfrac{\sigma^2+\sigma^2+\sigma^2+\sigma^2+\sigma^2}{5^2}=\dfrac{5}{25}\sigma^2=0.2\sigma^2$ したがって T_3 が 3 つのなかでもっとも分散が小さい（より有効）推定量である。 □

$\boxed{6}$ (1) 仮説検定における「第 1 種の過誤」と「第 2 種の過誤」について、定義の説明を述べなさい。
(2) 成功と失敗の属性をもつ 2 項母集団において、成功の確率 p の検定をおこなう。帰無仮説を $p = 0.5$, 対立仮説を $p = 0.8$ とする。4 回繰り返し、検定統計量を成功の回数 X ととる。もし棄却域を $\{X \geq 3\}$ とするならば、第 1 種の過誤の確率はいくつとなるか？

(解) (1) 第 1 種の過誤とは帰無仮説 H_0 が真であるにもかかわらず、これを棄却してしまう判断の誤りをいう。一方、第 2 種の過誤とは対立仮説 H_1 が真で、帰無仮説 H_0 が偽であるとき、帰無仮説のほうを採択してしまう判断の誤りをいう。

(2) 検定の定め方から, $H_0 : p = 0.5$, $H_1 : p = 0.8$ である。

$$
\begin{aligned}
&P(\{H_0\text{を棄却}\} \,|\, \{H_0\text{が真}\}) \\
&= P(X \geq 3 \,|\, p = 0.5) \\
&= P(X = 3 \,|\, p = 0.5) + P(X = 4 \,|\, p = 0.5) \\
&= \tbinom{4}{3} 0.5^3 0.5^{4-3} + \tbinom{4}{4} 0.5^4 0.5^{4-4} \\
&= 0.25 + 0.0625 = 0.3125
\end{aligned}
$$

□