確率分布と統計的な推測　晴れの確率を求める

出典：2024 年度大学共通テスト　数学 2B　第 3 問

以下の問題を解答するにあたっては，必要に応じて正規分布表を用いてもよい。また，ここでの**晴れ**の定義は，気象庁の天気概況の「快晴」または「晴」とする。

(1) 太郎さんは，自分が住んでいる地域において，日曜日に**晴れ**となる確率を考えている。

晴れの場合は 1，**晴れ**以外の場合は 0 の値をとる確率変数を X と定義する。また，$X=1$ である確率 p とすると，その確率分布は表 1 のようになる。

表　1

X	0	1	計
確率	$1-p$	p	1

この確率変数 X の平均（期待値）を m とすると

$m=$ $\boxed{ア}$

となる。

太郎さんは，ある期間における連続した n 週の日曜日の天気を，表 1 の確率分布をもつ母集団から無作為に抽出した大きさ n の標本とみなし，それらの X を確率変数 X_1，X_2，…，X_n で表すことにした。そして，その標本平均を $\overline{X}$ を利用して，母平均 m を推定しようと考えた。実際に $n=300$ として**晴れ**の日数を調べたところ，表 2 のようになった。

表　2

天気	日数
晴れ	75
晴れ以外	225
計	300

母標準偏差を σ とすると，$n=300$ は十分に大きいので，標本平均 $\overline{X}$ は近似的に正規分布 N（m, $\boxed{イ}$）に従う。

一般に，母標準偏差σがわからないとき，標本の大きさnが大きければ，σ の代わりに標本の標準偏差 S を用いてもよいことが知られている。Sは，

$$S=\sqrt{\frac{1}{n}\left\{\left(X_1-\overline{X}\right)^2+\left(X_2-\overline{X}\right)^2+\cdots+\left(X_n-\overline{X}\right)^2\right\}}$$

$$=\sqrt{\frac{1}{n}\left(X_1^2+X_2^2+\cdots+X_n^2\right)-\boxed{\text{ウ}}}$$

で計算できる。ここで，$X_1{}^2=X_1$，$X_2{}^2=X_2$ ，…，$X_n{}^2=X_n$であることに着目し，右辺を整理すると，$S=\sqrt{\boxed{\text{エ}}}$ と表されることがわかる。

よって，表 2 より，大きさ$n=300$の標本から求められる母平均mに対する信頼度 95%の信頼区間は$\boxed{\text{オ}}$となる。

$\boxed{\text{ア}}$の解答群

⓪ p	① p^2	② $1-p$	③ $(1-p)^2$

$\boxed{\text{イ}}$の解答群

⓪ σ	① σ^2	② $\dfrac{\sigma}{n}$	③ $\dfrac{\sigma^2}{n}$	④ $\dfrac{\sigma}{\sqrt{n}}$

$\boxed{\text{ウ}}$，$\boxed{\text{エ}}$の解答群（同じものを繰り返し選んでもよい。）

⓪ $\overline{X}$	① $\left(\overline{X}\right)^2$	② $\overline{X}\left(1-\overline{X}\right)$	③ $1-\overline{X}$

$\boxed{\text{オ}}$については，最も適当なものを，次の⓪～⑤のうちから一つ選べ。

⓪ $0.201 \leqq m \leqq 0.299$	① $0.209 \leqq m \leqq 0.291$
② $0.225 \leqq m \leqq 0.250$	③ $0.225 \leqq m \leqq 0.275$
④ $0.247 \leqq m \leqq 0.253$	⑤ $0.250 \leqq m \leqq 0.275$

（2）ある期間において，「ちょうど 3 週続けて日曜日の天気が**晴れ**になること」がどのくらいの頻度で起こり得るのかを考察しよう。以下では，連続する k 週の日曜日の天気について，（1）の太郎さんが考えた確率変数のうち X_1，X_2，…，X_n を用いて調べる。ただし，k は 3 以上 300 以下の自然数とする。

X_1，X_2，…，X_k の値を順に並べたときの 0 と 1 からなる列において，「ちょうど三つ続けて 1 が現れる部分」を A とし，A の個数を確率変数 U_k で表す。例えば，$k=20$ とし，X_1，X_2，…，X_{20} の値を順に並べたとき

$$1,\ 1,\ 1,\ 1,\ 0,\ \underbrace{\underline{1,\ 1,\ 1}}_{\mathrm{A}},\ 0,\ 0,\ 1,\ 1,\ 1,\ 1,\ 1,\ 0,\ 0,\ \underbrace{\underline{1,\ 1,\ 1}}_{\mathrm{A}}$$

であったとする。この例では，下線部分は A を示しており，1 が四つ以上続く部分は A とはみなさないので，$U_{20}=2$ となる。

$k=4$ のとき，X_1，X_2，X_3，X_4 のとり得る値と，それに対応した U_4 の値を書き出すと，表３のようになる。

ここで，U_k の期待値を求めてみよう。（1）における p の値を $p=\dfrac{1}{4}$ とする。$k=4$ のとき，U_4 の値は，

$$E(U_4)=\frac{\boxed{\text{カ}}}{128}$$

となる。$k=5$ のとき，U_5 の期待値は

$$E(U_5)=\frac{\boxed{\text{キク}}}{1024}$$

となる。

表　3

X_1	X_2	X_3	X_4	U_4
0	0	0	0	0
1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
0	1	1	0	0
0	1	0	1	0
0	0	1	1	0
1	1	1	0	1
1	1	0	1	0
1	0	1	1	0
0	1	1	1	1
1	1	1	1	0

4 以上の k について，k と $E(U_k)$の関係を詳しく調べると，座標平面上の点（$4, E(U_4)$），（$5, E(U_5)$），…，（$300, E(U_{300})$）は一つの直線上にあることがわかる。この事実によって

$$E(U_{300}) = \frac{\boxed{\text{ケコ}}}{\boxed{\text{サ}}}$$

となる。

【解答解説】

（1）（ア）（2 点）

$m = 0 \times (1-p) + 1 \times p = \boxed{ア⓪\boldsymbol{p}}$

（1）（イ）（2 点）

標本平均$\overline{X}$は近似的に正規分布 N（m, $\boxed{イ③\frac{\sigma^2}{n}}$）に従う。

（1）（ウ）（エ）（完 3 点）

$$S = \sqrt{\frac{1}{n}\left\{\left(X_1 - \overline{X}\right)^2 + \left(X_2 - \overline{X}\right)^2 + \cdots + \left(X_n - \overline{X}\right)^2\right\}}$$

$$= \sqrt{\frac{1}{n}\left\{(X_1^2 + X_2^2 + \cdots + X_n^2) + 2\overline{X}(X_1 + X_2 + \cdots + X_n) + n\left(\overline{X}\right)^2\right\}}$$

ここで，$2\overline{X} \bullet \underbrace{\frac{(X_1 + X_2 + \cdots + X_n)}{n}}_{\text{平均の公式}} = 2\overline{X} \bullet \overline{X} = 2\left(\overline{X}\right)^2$だから，

$$S = \sqrt{\frac{1}{n}(X_1^2 + X_2^2 + \cdots + X_n^2) - 2\left(\overline{X}\right)^2 + \left(\overline{X}\right)^2}$$

$$= \sqrt{\frac{1}{n}(X_1^2 + X_2^2 + \cdots + X_n^2) - \underbrace{\left(\overline{\boldsymbol{X}}\right)^{\boldsymbol{2}}}_{\text{ウ①}}}$$

ここで，$X_1{}^2 = X_1$，$X_2{}^2 = X_2$ ，…，$X_n{}^2 = X_n$であることに着目し，

右辺を整理すると，$\frac{1}{n}(X_1^2 + X_2^2 + \cdots + X_n^2) = \frac{X_1 + X_2 + \cdots + X_n}{n} = \overline{X}$

なので，$S = \sqrt{\overline{X} - \left(\overline{X}\right)^2} = \sqrt{\underbrace{\overline{\boldsymbol{X}}\left(\boldsymbol{1} - \overline{\boldsymbol{X}}\right)}_{\text{エ②}}}$

（1）（オ）（3 点）

母平均 m に対する信頼度 95%の信頼区間は，

$$\overline{X}-1.96\times\frac{\sigma}{\sqrt{n}}\leq m\leq\overline{X}+1.96\times\frac{\sigma}{\sqrt{n}}$$

表 2 より，$\overline{X}=\dfrac{75\times1+225\times0}{300}=\dfrac{1}{4}$ なので，$\sigma=S=\sqrt{\overline{X}\left(1-\overline{X}\right)}=\dfrac{\sqrt{3}}{4}$

$$\overline{X}-1.96\times\frac{\sigma}{\sqrt{n}}=\frac{1}{4}-1.96\times\frac{\sqrt{3}}{4\cdot10\sqrt{3}}=0.25-\frac{1.96}{40}=0.201$$

$\overline{X}-1.96\times\frac{\sigma}{\sqrt{n}}=0.299$ なので，**オ⓪0.201≦m≦0.299**

（2）（カ）（3 点）

k＝4 のとき，U_4＝1 となるのは，（X_1, X_2, X_3, X_4）＝（1, 1, 1, 0），（0, 1, 1, 1）のときであり，それ以外では U_4＝0 である。

X＝0 となる確率は 1－p＝3/4，X＝1 となる確率は p＝1/4 なので，

求める期待値は，

$$E(U_4)=2\cdot\left\{\frac{3}{4}\cdot\left(\frac{1}{4}\right)^3\right\}\cdot1=\frac{\mathbf{3}}{128}$$

（2）（キク）（3 点）

k＝5 のとき，U_5＝1 となるのは，

（X_1, X_2, X_3, X_4, X_5）＝（1, 1, 1, 0, 0），（0, 1, 1, 1, 0），（0, 0, 1, 1, 1）

（1, 1, 1, 0, 1），（1, 0, 1, 1, 1）のときで，それ以外は U_5＝0 だから，

（1, 1, 1, 0, 0），（0, 1, 1, 1, 0），（0, 0, 1, 1, 1）だけの期待値

$$3\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3=\frac{27}{1024}$$

（1, 1, 1, 0, 1），（1, 0, 1, 1, 1）だけの期待値

$$2\cdot\frac{3}{4}\cdot\left(\frac{1}{4}\right)^4=\frac{6}{1024}$$

なので，$E(U_5)=\dfrac{27}{1024}+\dfrac{6}{1024}=\dfrac{\mathbf{33}}{1024}$

（ケ）（4 点）

4 以上の k について，k と $E(U_k)$の関係を詳しく調べると，座標平面上の点（4, $E(U_4)$），（5, $E(U_5)$），…,（300, $E(U_{300})$）は一つの直線上にあることがわかる。と書いてあるのでこれを信じて，

＋296

＋1

k	4	5	…	300
$E(U_k)$	$\dfrac{24}{1024}$	$\dfrac{33}{1024}$	…	

$+\dfrac{9}{1024}$

1 で $\dfrac{9}{1024}$ 増える

↓× 296

296 で $\dfrac{2664}{1024}$ 増える　　　$E(U_{300})=\dfrac{24}{1024}+\dfrac{2664}{1024}=\boxed{\dfrac{\mathbf{21}}{\mathbf{8}}}$