第5問　確率分布と統計的な推測

(1)　箱の中に赤球が5個，白球が15個入っている場合を考える．箱から無作為に1個の球を取り出す試行 T を1回行うとき，その球が赤球である確率は $\dfrac{5}{20}=\dfrac{1}{4}$ である．よって，確率変数 X は

$$\text{二項分布 } B\left(n,\ \frac{\boxed{1}}{\boxed{4}}\right)$$

に従う．X の期待値(平均)を $E(X)$，分散を $V(X)$ とすると

$$E(X)=n\cdot\frac{1}{4}=\frac{\boxed{1}}{\boxed{4}}n$$

$$V(X)=n\cdot\frac{1}{4}\cdot\left(1-\frac{1}{4}\right)=\frac{\boxed{3}}{\boxed{16}}n$$

である．

1回の試行 T につき，赤球，白球が取り出されるとそれぞれ5点，2点が得られるから，試行 T を n 回繰り返した後の得点の合計 Y は

$$Y=5X+2(n-X)=\boxed{3}X+\boxed{2}n$$

と表される．よって，確率変数 Y の期待値(平均)を $E(Y)$，分散を $V(Y)$ とすると

$$E(Y)=3E(X)+2n=3\cdot\frac{1}{4}n+2n=\frac{\boxed{11}}{\boxed{4}}n$$

$$V(Y)=3^2V(X)=9\cdot\frac{3}{16}n=\frac{\boxed{27}}{\boxed{16}}n$$

となる．

(2)　箱の中に赤球が4個，白球が16個入っている場合を考える．箱から無作為に1個の球を取り出す試行 T を1回行うとき，その球が赤球である確率は $\dfrac{4}{20}=\dfrac{1}{5}$ である．$n=1600$ のとき，確率変数 X は二項分布 $B\left(1600,\ \dfrac{1}{5}\right)$ に従う．

$$X\text{ の期待値(平均)は}\quad 1600\cdot\frac{1}{5}=320$$

$$X\text{ の分散は}\quad 1600\cdot\frac{1}{5}\cdot\left(1-\frac{1}{5}\right)=256$$

であり，試行回数 $n=1600$ は十分大きいと考えられるので，X は近似的に正規分布 $N(320,\ 256)$ に従う．さらに

$$Z=\frac{X-320}{\sqrt{256}}=\frac{X-\boxed{320}}{\boxed{16}}$$

二項分布

確率変数 X が二項分布 $B(n,\ p)$ に従うとき，$q=1-p$ とすると，X の期待値(平均) $E(X)$，分散 $V(X)$ は

$$E(X)=np$$
$$V(X)=npq$$

である．

期待値の性質

確率変数 X と定数 $a,\ b$ に対して

$$E(aX+b)=aE(X)+b$$

が成り立つ．

分散の性質

確率変数 X と定数 $a,\ b$ に対して

$$V(aX+b)=a^2V(X)$$

が成り立つ．

二項分布の正規分布による近似

二項分布 $B(n,\ p)$ に従う確率変数 X は，n が十分大きいとき，近似的に正規分布 $N(np,\ npq)$ に従う．ただし，$q=1-p$ である．

とおくと，Z は近似的に標準正規分布 $N(0,\ 1)$ に従う．

$300 \leqq X \leqq 336$ のとき，$-1.25 \leqq Z \leqq 1$ となるので，

$300 \leqq X \leqq 336$ となる確率は，正規分布表を参照すると

$$\begin{aligned} P(-1.25 \leqq Z \leqq 1) &= P(-1.25 \leqq Z \leqq 0) + P(0 \leqq Z \leqq 1) \\ &= P(0 \leqq Z \leqq 1.25) + P(0 \leqq Z \leqq 1) \\ &= 0.3944 + 0.3413 \\ &= 0.7357 \\ &\fallingdotseq \boxed{0}.\boxed{74} \end{aligned}$$

である．

(3) 箱の中に球が多数入っているとき，ある試行によって，赤球が取り出される確率が $\dfrac{2}{5}$ であることがわかっている．箱の中から赤球が取り出される確率を信頼度95%，信頼区間の幅が0.12以下で推定したい場合，正規分布表を参照すると

$$P(-1.96 \leqq Z \leqq 1.96) = 0.95$$

より，n が大きいとき，信頼区間の幅について

$$2 \times 1.96\sqrt{\frac{p(1-p)}{n}} \leqq 0.12$$

すなわち

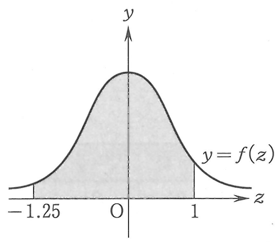
$$\boxed{3}.\boxed{92}\sqrt{\frac{p(1-p)}{n}} \leqq 0.12$$

が成り立つ．この不等式に $p = \dfrac{2}{5}$ を代入して計算すると

$$\begin{aligned} n &\geqq \frac{2}{5} \cdot \frac{3}{5} \cdot \left(\frac{392}{12}\right)^2 \\ &= \frac{2 \cdot 98^2}{3 \cdot 25} \\ &= 256.1\cdots \end{aligned}$$

より，n は $\boxed{257}$ 以上にしなければならないことがわかる．

平均0，標準偏差1の正規分布 $N(0,\ 1)$ を標準正規分布という．



標準正規分布 $N(0,\ 1)$ に従う確率変数 Z の確率密度関数は

$$f(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$$

であり，曲線 $y = f(z)$ は y 軸に関して対称であるから

$$P(-1.25 \leqq Z \leqq 0) = P(0 \leqq Z \leqq 1.25)$$

が成り立つ．

正規分布表より

$P(0 \leqq Z \leqq 1.96) = 0.4750$,

$0.95 = 0.4750 \times 2$.

母比率の推定

一般に，母集団全体の中で特性Aをもつ要素の割合を，特性Aの母比率という．これに対して，標本の中で特性Aをもつ要素の割合を，特性Aの標本比率という．

ある母集団から無作為に抽出された標本の標本比率を用いて母比率を推定する場合，標本の大きさ n が大きいとき，標本比率を p とすると，母比率に対する信頼度 k% の信頼区間の幅は

$$2 \times u\sqrt{\frac{p(1-p)}{n}}$$

と表せる．

ただし，u は正規分布表において

$$2P(0 \leqq Z \leqq u) = \frac{k}{100}$$

を満たす数である．