雑感　これって同じこと？　―超幾何分布―

■　赤球 3 個，白球 2 個が入った袋から，2 個の球を取り出すとき，取り出される赤球の個数を確率変数 X とする．

このとき，取り出し方として

(A) 同時に 2 個取り出す

(B) 元に戻さず 1 個ずつ 2 回取り出す

(C) 元に戻して 1 個ずつ 2 回取り出す

という場合を考える．

■　(C)の場合は反復試行なので，確率は

$$P(X=1)={}_2\mathrm{C}_1\left(\frac{3}{5}\right)\left(\frac{2}{5}\right)=\frac{12}{25}$$

のように求められ，X は 2 項分布 $M\left(2,\ \frac{3}{5}\right)$ に従うから，期待値（平均）$E(X)=2\cdot\frac{3}{5}=\frac{6}{5}$，分散 $V(X)=2\cdot\frac{3}{5}\left(1-\frac{3}{5}\right)=\frac{18}{25}$ のように「公式的に」計算が可能である．

■　一方，(A)では確率は ${}_n\mathrm{C}_r$ を用いて，

$$P(X=1)=\frac{{}_3\mathrm{C}_1\cdot{}_2\mathrm{C}_1}{{}_5\mathrm{C}_2}=\frac{3}{5}$$

などといった計算が，

（B）では 1 回目が赤球か白球かで分けて，乗法定理を用いた

$$P(X=1)=\frac{3}{5}\cdot\frac{2}{4}+\frac{2}{5}\cdot\frac{3}{4}=\frac{3}{5}$$

などといった計算が一般的である．

■　しかし，実際には（A），(B)の違いはなく，全く同じ確率分布が得られる．

これは，(A)の同時に 2 個取り出すと言っても「取り出す微妙な時間差がある」から，詰まるところ(B)の取り出し方と同じことだからである．

したがって，(B)の取り出し方で確率の計算が煩雑な場合，(A)の方法で確率の計算を行っても構わない．もちろん，逆も可で，(A)の取り出し方で考えづらいとき，(B)の取り出し方で樹形図などを考える方法もあり得る．

■　なお，このような確率変数 X の確率分布は，超幾何分布と呼ばれ，その期待値は次のように考えれば容易に分かる．

袋の中の 5 つの球がすべて，「$\frac{3}{5}$ が赤に $\frac{2}{5}$ が白に塗り分けられている」と考えればよく，それを 2 個取り出したときの赤の分は $2\times\frac{3}{5}=\frac{6}{5}$ で，これが赤球の個数の期待値 $E(X)$ である．

つまり，赤球が p 個，白球が q 個入った袋から，同時に m 個を取り出したときの赤球の個数 X について，その期待値(平均) $E(X)=\frac{mp}{p+q}$ である．

■　しかし，分散 $V(X)$ については，このような便利な考え方はない（多分）．

公式的には $V(X)=\frac{mpq(p+q-m)}{(p+q)^2(p+q-1)}$ である．

上のケースで公式に代入すれば，

$$V(X)=\frac{2\cdot3\cdot2(3+2-2)}{(3+2)^2(3+2-1)}=\frac{9}{25}$$ となる．

2015 年 7 月 1 日