超幾何分布 (hypergoemetric distribution)

緑川章一

2つの種類 A，B からなる総数が N コの集団があり、A が M コ、B が $N-M$ コ含まれている。いま、この集団からデタラメに n コを取り出した時、A が x コ、B が $n-x$ コであったとする。総数 N コのものから、n コの取り出し方は ${}_NC_n$ である。一方、A 種類のもの N コから x コの取り出し方は ${}_NC_x$、B 種類のもの $N-M$ から $n-x$ コの取り出し方は ${}_{N=M}C_{n-x}$ である。そこで、A 種類のものから取り出す個数 x を確率変数 X として選ぶ。すると、$X=x$ となる確率は、

$$f(x) = \frac{{}_MC_x \cdot {}_{N-M}C_{n-x}}{{}_NC_n} \tag{1}$$

となる。ここで、$0 \leqq x \leqq M$、かつ $0 \leqq n-x \leqq N-M$、すなわち、$M+n-N \leqq x \leqq n$ である。

二項展開における関係式

等式

$$(1+t)^M(1+t)^{N-M} = (1+t)^N$$

の両辺を二項展開して、t^n の係数を比較すると、

$$\sum_{x=x_L}^{x_U} {}_MC_x \cdot {}_{N-M}C_{n-x} = {}_NC_n, \text{ または、} \sum_{x=x_L}^{x_U} \frac{{}_MC_x \cdot {}_{N-M}C_{n-x}}{{}_NC_n} = 1$$

ここで、$x_L = max\{0,\ M+n-N\}$、$x_U = min\{n,\ M\}$. ゆえに、

$$\sum_{x=x_L}^{x_U} f(x) = 1$$

が成り立つ。

平均

$$\mu = \sum_x xf(x) = \sum x \frac{{}_MC_x \cdot {}_{N-M}C_{n-x}}{{}_NC_n} \tag{2}$$

ところで、

$$x\,{}_MC_x = M\,{}_{M-1}C_{x-1}, \quad {}_NC_n = \frac{N}{n}\,{}_{N-1}C_{n-1}$$

だから、$N' = N-1,\ M' = M-1,\ n' = n-1,\ x' = x-1$ とおくと、

$$\sum x \frac{{}_MC_x \cdot {}_{N-M}C_{n-x}}{{}_NC_n} = \frac{Mn}{N} \sum_{x'} \frac{{}_{M'}C_{x'} \cdot {}_{N'-M'}C_{n'-x'}}{{}_{N'}C_{n'}}$$

ところが、

$$\sum_{x'} \frac{{}_{M'}C_{x'} \cdot {}_{N'-M'}C_{n'-x'}}{{}_{N'}C_{n'}} = 1$$

だから、結局、

$$\mu = \frac{Mn}{N} \tag{3}$$

を得る。

分散

$$\begin{aligned}
\sigma^2 &= \sum_x x^2 f(x) - \mu^2 \\
&= \sum_x [x(x-1)+x] \frac{{}_M C_x \cdot {}_{N-M} C_{n-x}}{{}_N C_n} - \mu^2 \\
&= \sum_x x(x-1) \frac{{}_M C_x \cdot {}_{N-M} C_{n-x}}{{}_N C_n} + \mu - \mu^2
\end{aligned}$$

ところで、

$$x(x-1)\,{}_M C_x = M(M-1)\,{}_{M-2} C_{x-2}, \quad {}_N C_n = \frac{N(N-1)}{n(n-1)}\,{}_{N-2} C_{n-2}$$

だから、$N'' = N-2,\ M'' = M-2,\ n'' = n-2,\ x'' = x-2$ とおくと、

$$\sum_x x(x-1) \frac{{}_M C_x \cdot {}_{N-M} C_{n-x}}{{}_N C_n} = \frac{M(M-1)n(n-1)}{N(N-1)} \sum_{x''} \frac{{}_{M''} C_{x''} \cdot {}_{N''-M''} C_{n''-x''}}{{}_{N''} C_{n''}}$$

ここで、

$$\sum_{x''} \frac{{}_{M''} C_{x''} \cdot {}_{N''-M''} C_{n''-x''}}{{}_{N''} C_{n''}} = 1$$

だから、

$$\begin{aligned}
\sigma^2 &= \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{Mn}{N} - \left(\frac{Mn}{N}\right)^2 \\
&= \frac{Mn(N-M)(N-n)}{N^2(N-1)} \qquad (4)
\end{aligned}$$

ここで。$p = M/N$ とおくと、

$$\begin{aligned}
\text{平均} &: \quad \mu = np \\
\text{分散} &: \quad \sigma^2 = \frac{np(1-p)(N-n)}{N-1}
\end{aligned}$$

となる。