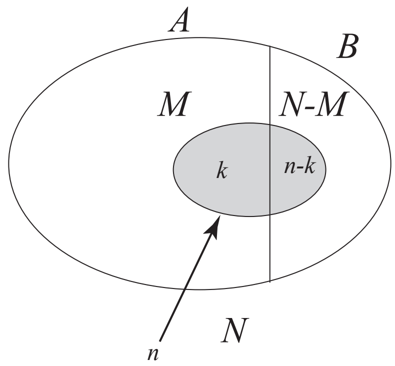
超幾何分布

緑川章一

1　分布、平均、分散

全部で N 個の物がある。その内訳は、種類 A が M 個, 種類 B が $N-M$ 個である。この中から n 個を取り出す。



その時、A が k 個である確率は、

$$\mathrm{P}(k) = \frac{{}_M C_k \, {}_{N-M} C_{n-k}}{{}_N C_n}$$

である。 ここで、k の取りうる範囲は、

(1)　$n \leqq min\{M,\ N-M\}$ のとき　　$0 \leqq k \leqq n$
(2)　$M > N-M$ かつ $N-M \leqq n \leqq M$ のとき　　$n-(N-M) \leqq k \leqq n$
(3)　$M < N-M$ かつ $M \leqq n \leqq N-M$ のとき　　$0 \leqq k \leqq M$
(4)　$n \geqq max\{M,\ N-M\}$ のとき　　$n-(N-M) \leqq k \leqq M$

すなわち、$K_{min} = max\{0,\ n-(N-M)\},\ K_{max} = min\{n,\ M\}$ とおくと、

$$K_{min} \leqq k \leqq K_{max}$$

$\displaystyle\sum_{k=K_{min}}^{K_{max}} \mathbf{P}(k) = 1$ の証明

$$(a+b)^N = (a+b)^M (a+b)^{N-M}$$

より、

$$\begin{aligned}\sum_{n=0}^{N} {}_N C_n a^n b^{N-n} &= \sum_{k=0}^{M} {}_M C_k a^k b^{M-k} \times \sum_{l=0}^{N-M} {}_{N-M}C_l a^l b^{N-M-l} \\ &= \sum_{k=0}^{M} \sum_{l=0}^{N-M} {}_M C_k \ {}_{N-M}C_l \ a^{k+l} b^{N-(k+l)} \\ &= \sum_{n=0}^{N} \sum_{k=0}^{M} \sum_{l=0}^{N-M} {}_M C_k \ {}_{N-M}C_l \ a^{k+l} b^{N-(k+l)} \ \delta_{n,k+l}\end{aligned}$$

ここで、 $\delta_{n,k+l}$ は、クロネッカーのデルタである。すなわち、

$$\delta_{n,k+l} = \begin{cases} 1 & n = k+l \text{ のとき} \\ 0 & n \neq k+l \text{ のとき} \end{cases}$$

ゆえに、

$${}_N C_n = \sum_{k=0}^{M} {}_M C_k \ {}_{N-M}C_{n-k} \ \sum_{l=0}^{N-M} \delta_{n,k+l}$$

を得る。ここで、

$$0 \leqq k \leqq M, \quad 0 \leqq l \leqq N-M, \quad k+l = n$$

である。これを (1), (2), (4) の場合に図示すると、図 1 のようになる。図から分かるように、(1) の場合に $\sum_{l=0}^{N-M} \delta_{n,k+l}$ が 1 となるのは、$0 \leqq k \leqq n$ のときのみである。同様にして, (2) の場合の k の和の範囲は、$n-(N-M) \leqq k \leqq n$ となることが分かる。(3) と (4) の場合についても同様であるので、一般に

$${}_N C_n = \sum_{k=K_{min}}^{K_{max}} {}_M C_k \ {}_{N-M}C_{n-k}$$

が成り立つことが分かる。ゆえに、

$$\sum_{k=K_{min}}^{K_{max}} \mathrm{P}(k) = \frac{1}{{}_N C_n} \sum_{k=K_{min}}^{K_{max}} {}_M C_k \, {}_{N-M}C_{n-k} = 1$$

は成り立つ。

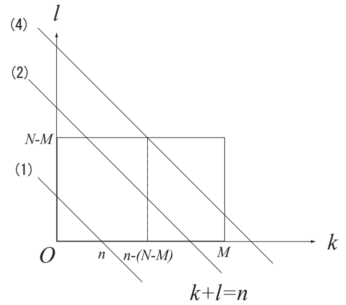


図 1: $(k,\,l)$ 平面で表した和の領域

平均

$$
\begin{aligned}
\mu &= \sum_{k=max\{0,\,n-(N-M)\}}^{min\{n,\,M\}} k\mathrm{P}(k) \\
&= \frac{1}{{}_N C_n} \sum_{k=max\{1,\,n-(N-M)\}}^{min\{n,\,M\}} k\,{}_M C_k\,{}_{N-M} C_{n-k}
\end{aligned}
$$

ここで、

$$
\begin{aligned}
k\,{}_M C_k &= k\frac{M!}{k!\,(M-k)!} \\
&= M\frac{(M-1)!}{(k-1)!\,((M-1)-(k-1))!} \\
&= M\,{}_{M-1} C_{k-1}
\end{aligned}
$$

である。そこで、

$$
k' = k-1, \qquad n' = n-1, \qquad N' = N-1, \qquad M' = M-1
$$

とおくと、

$$
\begin{aligned}
\mu &= \frac{M}{{}_N C_n} \sum_{k'=max\{0,\,n'-(N'-M')\}}^{min\{n',\,M'\}} {}_{M'} C_{k'}\ {}_{N'-M'} C_{n'-k'} \\
&= \frac{M}{{}_N C_n}\ {}_{N-1} C_{n-1} \\
&= \frac{Mn}{N}
\end{aligned}
$$

すなわち、

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

を得る。

分散

$$
\begin{aligned}
\sigma^2 &= \sum_{k=max\{0,\,n-(N-M)\}}^{min\{n,\,M\}} k^2\mathrm{P}(k) - \mu^2 \\
&= \sum_{k=max\{2,\,n-(N-M)\}}^{min\{n,\,M\}} k(k-1)\mathrm{P}(k) + \mu - \mu^2 \\
&= \frac{1}{{}_N C_n} \sum_{k=max\{2,\,n-(N-M)\}}^{min\{n,\,M\}} k(k-1)\,{}_M C_k\,{}_{N-M} C_{n-k} + \mu - \mu^2
\end{aligned}
$$

ここで、

$$
\begin{aligned}
k(k-1)\,{}_M C_k &= \frac{M!}{(k-2)!(M-k)!} \\
&= M(M-1)\ \frac{(M-2)!}{(k-2)!\,((M-2)-(k-2))!} \\
&= M(M-1)\ {}_{M-2} C_{k-2}
\end{aligned}
$$

である。そこで、

$$k'' = k-2, \qquad n'' = n-2, \qquad N'' = N-2, \qquad M'' = M-2$$

とおくと、

$$\begin{aligned}\sum_{k=max\{2,\,n-(N-M)\}}^{min\{n,\,M\}} k(k-1)\,{}_MC_k\,{}_{N-M}C_{n-k} &= M(M-1)\sum_{k''=max\{0,\,n''-(N''-M'')\}}^{min\{n'',\,M''\}} {}_{M''}C_{k''}\,{}_{N''-M''}C_{n''-k''} \\ &= M(M-1)\,{}_{N-2}C_{n-2}\end{aligned}$$

となるので、

$$\begin{aligned}\sigma^2 &= M(M-1)\frac{{}_{N-2}C_{n-2}}{{}_NC_n} + \mu - \mu^2 \\ &= \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{Mn}{N} - \frac{M^2n^2}{N^2}\end{aligned}$$

これを整理すると、

$$\sigma^2 = n\,\frac{M}{N}\,\frac{(N-M)}{N}\,\frac{(N-M)}{(N-1)}$$

を得る。

2 名前の由来

次の級数で定義される関数を超幾何関数という。

$${}_2F_1(a,b;c;x) \quad = \quad 1 + \frac{ab}{c}\frac{x}{1!} + \frac{a(a+1)b(b+1)}{c(c+1}\frac{x^2}{2!} + \frac{a(a;1)(a+2)b(b+1)(b+2)}{c(c+1)(c;2)}\frac{x^3}{3!} + \cdots$$

これは、もともとは常微分方程式

$$x(x-1)y'' + [c-(a+b+1)x]\,y' - aby = 0$$

の解として研究された。

確率分布 $\mathrm{P}(k)$ が与えられたとき、その母関数 (probability generating function) は、

$$f(x) = \mathrm{P}(0) + \mathrm{P}(1)x + \mathrm{P}(2)x^2 + \mathrm{P}(3)x^3 + \cdots$$

で与えられる。

確率分布が $\mathrm{P}(k) = \dfrac{{}_MC_k\,{}_{N-M}C_{n-k}}{{}_NC_n}$ で与えられるときの母関数は、

$$\begin{aligned}f(x) &= \frac{{}_{N-M}C_n}{{}_NC_n}\left[1 + \frac{{}_MC_1\,{}_{N-M}C_{n-1}}{{}_{N-M}C_n}x + \frac{{}_MC_2\,{}_{N-M}C_{n-2}}{{}_{N-M}C_n}x^2 + \frac{{}_MC_3\,{}_{N-M}C_{n-3}}{{}_{N-M}C_n}x^3 + \cdots\right] \\ &= \frac{{}_{N-M}C_n}{{}_NC_n}\left[1 + \frac{(-M)(-n)}{(N-M-n+1)}\frac{x}{1!} + \frac{(-M)(-M+1)(-n)(-n+1)}{(N-M-n+1)(N-M-n+2)}\frac{x^2}{2!} + \cdots\right]\end{aligned}$$

すなわち、

$$f(x) = \frac{{}_{N-M}C_n}{{}_NC_n}\,{}_2F_1(-M,-n;N-M-n+1;x)$$

を得る。