平成 13 年 6 月 8 日

超幾何分布の平均、分散、極限

新潟工科大学 情報電子工学科 竹野茂治

超幾何分布 $HG(N_1, N_0; n)$ は、確率変数 x、確率関数 $p(x)$ が

$$x \in \{0, 1, 2, \ldots, n\}, \quad p(x) = \frac{\left(\begin{array}{c} N_1 \\ x \end{array} \right) \left(\begin{array}{c} N_0 \\ n-x \end{array} \right)}{\left(\begin{array}{c} N_1 + N_0 \\ n \end{array} \right)}$$

で与えられる確率分布である。

補題 1

$$\begin{aligned} \sum_{k=0}^{n} \left(\begin{array}{c} N \\ k \end{array} \right) \left(\begin{array}{c} M \\ n-k \end{array} \right) &= \left(\begin{array}{c} N \\ 0 \end{array} \right) \left(\begin{array}{c} M \\ n \end{array} \right) + \left(\begin{array}{c} N \\ 1 \end{array} \right) \left(\begin{array}{c} M \\ n-1 \end{array} \right) + \cdots + \left(\begin{array}{c} N \\ n \end{array} \right) \left(\begin{array}{c} M \\ 0 \end{array} \right) \\ &= \left(\begin{array}{c} N+M \\ n \end{array} \right) \end{aligned}$$

証明

“母関数の方法” を用いる。二項定理により

$$(1+x)^N = \sum_{j=0}^{N} \left(\begin{array}{c} N \\ j \end{array} \right) x^j, \quad (1+x)^M = \sum_{k=0}^{M} \left(\begin{array}{c} M \\ k \end{array} \right) x^k$$

であるので、

$$\begin{aligned} (1+x)^{N+M} &= (1+x)^N (1+x)^M = \left\{ \sum_{j=0}^{N} \left(\begin{array}{c} N \\ j \end{array} \right) x^j \right\} \left\{ \sum_{k=0}^{M} \left(\begin{array}{c} M \\ k \end{array} \right) x^k \right\} \\ &= \sum_{j=0}^{N} \sum_{k=0}^{M} \left(\begin{array}{c} N \\ j \end{array} \right) \left(\begin{array}{c} M \\ k \end{array} \right) x^{j+k} = \sum_{j,k} \left(\begin{array}{c} N \\ j \end{array} \right) \left(\begin{array}{c} M \\ k \end{array} \right) x^{j+k} \\ &= \sum_{\ell=0}^{N+M} x^\ell \sum_{k+j=\ell} \left(\begin{array}{c} N \\ j \end{array} \right) \left(\begin{array}{c} M \\ k \end{array} \right) \end{aligned}$$

となるが、二項定理より

$$(1+x)^{N+M} = \sum_{\ell=0}^{N+M} \left(\begin{array}{c} N+M \\ \ell \end{array} \right) x^\ell$$

であるので、ゆえに

$$\sum_{k+j=\ell} \left(\begin{array}{c} N \\ j \end{array} \right) \left(\begin{array}{c} M \\ k \end{array} \right) = \left(\begin{array}{c} N+M \\ \ell \end{array} \right)$$

■

命題 2

超幾何分布 $HG(N_1, N_0; n)$ に対して、$E[x] = np \quad (p = N_1/(N_1 + N_0))$

証明

$$
\begin{aligned}
E[x] &= \sum_{x=0}^{n} xp(x) = \sum_{k=0}^{n} k \frac{\left(\begin{array}{c} N_1 \\ k \end{array} \right) \left(\begin{array}{c} N_0 \\ n-k \end{array} \right)}{\left(\begin{array}{c} N_1+N_0 \\ n \end{array} \right)} = \sum_{k=1}^{n} (\text{ 同左 }) \quad (k=0 \text{ のときは } 0) \\
&= \frac{1}{\left(\begin{array}{c} N_1+N_0 \\ n \end{array} \right)} \sum_{k=1}^{n} k \left(\begin{array}{c} N_1 \\ k \end{array} \right) \left(\begin{array}{c} N_0 \\ n-k \end{array} \right)
\end{aligned}
$$

となるが、ここで、

$$
\begin{aligned}
k \left(\begin{array}{c} N_1 \\ k \end{array} \right) &= k \frac{N_1(N_1-1)(N_1-2)\cdots(N_1-k+1)}{k(k-1)(k-2)\cdots 1} \\
&= N_1 \frac{(N_1-1)(N_1-2)\cdots(N_1-k+1)}{(k-1)(k-2)\cdots 1} \\
&= N_1 \left(\begin{array}{c} N_1-1 \\ k-1 \end{array} \right)
\end{aligned}
$$

なので、

$$
\begin{aligned}
E[x] &= \frac{1}{\left(\begin{array}{c} N_1+N_0 \\ n \end{array} \right)} \sum_{k=1}^{n} N_1 \left(\begin{array}{c} N_1-1 \\ k-1 \end{array} \right) \left(\begin{array}{c} N_0 \\ n-k \end{array} \right) \\
&= \frac{N_1}{\left(\begin{array}{c} N_1+N_0 \\ n \end{array} \right)} \sum_{j=0}^{n-1} \left(\begin{array}{c} N_1-1 \\ j \end{array} \right) \left(\begin{array}{c} N_0 \\ n-j-1 \end{array} \right) \quad (k-1=j) \\
&= \frac{N_1}{\left(\begin{array}{c} N_1+N_0 \\ n \end{array} \right)} \left(\begin{array}{c} N_1-1+N_0 \\ n-1 \end{array} \right) \quad (\text{補題 } 1 \text{ の } N, M, n \text{ が } N_1-1, N_0, n-1)
\end{aligned}
$$

となる。ところで、

$$
\begin{aligned}
\left(\begin{array}{c} N_1+N_0 \\ n \end{array} \right) &= \frac{(N_1+N_0)(N_1+N_0-1)\cdots(N_1+N_0-n+1)}{n(n-1)\cdots 1} \\
&= \frac{N_1+N_0}{n} \cdot \frac{(N_1+N_0-1)\cdots(N_1+N_0-n+1)}{(n-1)\cdots 1} \\
&= \frac{N_1+N_0}{n} \left(\begin{array}{c} N_1+N_0-1 \\ n-1 \end{array} \right)
\end{aligned}
$$

より、最後の式は約分されて

$$
E[x] = \frac{N_1}{\underline{\frac{N_1+N_0}{n}}} = n \frac{N_1}{N_1+N_0} = np
$$

となる。■

命題 3

超幾何分布 $HG(N_1, N_0; n)$ に対して、$V[x] = npq\dfrac{N_1+N_0-n}{N_1+N_0-1}$ $(q = 1-p = N_0/(N_1+N_0))$

証明

$$\begin{aligned} E[x(x-1)] &= \sum_{x=0}^{n} x(x-1)p(x) = \sum_{x=2}^{n} x(x-1)p(x) \quad (x=0,1 \text{ のときは } 0) \\ &= \frac{1}{\left(\begin{array}{c} N_1+N_0 \\ n \end{array}\right)} \sum_{k=2}^{n} k(k-1) \left(\begin{array}{c} N_1 \\ k \end{array}\right) \left(\begin{array}{c} N_0 \\ n-k \end{array}\right) \end{aligned}$$

であり、命題 2 の証明と同様にして

$$k(k-1)\left(\begin{array}{c} N_1 \\ k \end{array}\right) = N_1(N_1-1)\left(\begin{array}{c} N_1-2 \\ k-2 \end{array}\right)$$

がいえるので、補題 1 を使い、命題 2 の証明と同様の計算を行うと

$$\begin{aligned} E[x(x-1)] &= \frac{N_1(N_1-1)}{\left(\begin{array}{c} N_1+N_0 \\ n \end{array}\right)} \sum_{k=2}^{n} \left(\begin{array}{c} N_1-2 \\ k-2 \end{array}\right) \left(\begin{array}{c} N_0 \\ n-k \end{array}\right) \\ &= \frac{N_1(N_1-1)}{\left(\begin{array}{c} N_1+N_0 \\ n \end{array}\right)} \left(\begin{array}{c} N_1+N_0-2 \\ n-2 \end{array}\right) \\ &= N_1(N_1-1)\frac{n(n-1)}{(N_1+N_0)(N_1+N_0-1)} \end{aligned}$$

となる。一方、

$E[x(x-1)] = E[x^2-x] = E[x^2] - E[x]$ より $E[x^2] = E[x(x-1)] + E[x]$

であり、よって命題 2 より

$$\begin{aligned} V[x] &= E[x^2] - E[x]^2 = E[x(x-1)] + E[x] - E[x]^2 \\ &= N_1(N_1-1)\frac{n(n-1)}{(N_1+N_0)(N_1+N_0-1)} + np - n^2p^2 \\ &= n\frac{N_1}{N_1+N_0}\frac{(n-1)(N_1-1)}{N_1+N_0-1} + np(1-np) \\ &= np\left\{\frac{(n-1)(N_1-1)}{N_1+N_0-1} + 1 - n\frac{N_1}{N_1+N_0}\right\} \quad \left(p = \frac{N_1}{N_1+N_0}\right) \end{aligned}$$

ここで $A = N_1 + N_0$ とおくと

$$\begin{aligned} V[x] &= np\left\{\frac{(n-1)(N_1-1)}{A-1} + 1 - n\frac{N_1}{A}\right\} \\ &= np\frac{(n-1)(N_1-1)A + A(A-1) - nN_1(A-1)}{A(A-1)} \\ &= np\frac{nN_1A - nA - N_1A + A + A^2 - A - nN_1A + nN_1}{A(A-1)} \\ &= np\frac{A^2 - nA - N_1A + nN_1}{A(A-1)} = np\frac{(A-n)(A-N_1)}{A(A-1)} = np\frac{(N_1+N_0-n)N_0}{(N_1+N_0)(N_1+N_0-1)} \\ &= npq\frac{N_1+N_0-n}{N_1+N_0-1} \quad \left(q = \frac{N_0}{N_1+N_0}\right) \end{aligned}$$

■

以後、$N_1 = N,\ N_0 = M$ と書くことにする。

命題 4

超幾何分布 $HG(N, M; n)$ に対して、$p = N/(N+M)$ を固定して $N, M \to \infty$ とすると $HG(N, M; n) \to B(n, p)$

証明

$$p = \frac{N}{N+M}$$

を固定するということは、$p : q = N : M$ より

$$M = \frac{q}{p}N$$

として $N \to \infty$ するということ。このとき、

$$p(x) = \frac{\left(\begin{array}{c} N \\ x \end{array}\right)\left(\begin{array}{c} M \\ n-x \end{array}\right)}{\left(\begin{array}{c} N+M \\ n \end{array}\right)} \to \left(\begin{array}{c} n \\ x \end{array}\right) p^x q^{n-x}$$

となることを示せばよい $(0 \le x \le n)$。以後、$n - x = y$ とする。

$$\begin{aligned} p(x) &= \frac{N(N-1)\cdots(N-x+1)}{x!} \times \frac{M(M-1)\cdots(M-y+1)}{y!} \\ &\quad \times \frac{n!}{(N+M)(N+M-1)\cdots(N+M-n+1)} \\ &= \left(\begin{array}{c} n \\ x \end{array}\right) \frac{N(N-1)\cdots(N-x+1)}{(N+M)(N+M-1)\cdots(N+M-x+1)} \\ &\quad \times \frac{M(M-1)\cdots(M-y+1)}{(N+M-x)(N+M-x-1)\cdots(N+M-n+1)} \quad (x+y=n \text{ より}) \\ &= \left(\begin{array}{c} n \\ x \end{array}\right) \frac{N}{N+M} \cdot \frac{N-1}{N+M-1} \cdots \frac{N-x+1}{N+M-x+1} \end{aligned}$$

$$\times \frac{M}{N+M-x} \cdot \frac{M-1}{N+M-x-1} \cdots \frac{M-y+1}{N+M-n+1}$$

$$= \left(\begin{array}{c} n \\ x \end{array} \right) \prod_{j=0}^{x-1} \frac{N-j}{N+M-j} \prod_{k=0}^{y-1} \frac{M-k}{N+M-x-k}$$

ここで、$M = \frac{q}{p}N$ より

$$\frac{N-j}{N+M-j} = \frac{1-\frac{j}{N}}{1+\frac{M}{N}-\frac{j}{N}} = \frac{1-\frac{j}{N}}{1+\frac{q}{p}-\frac{j}{N}} \overset{N\to\infty}{\longrightarrow} \frac{1}{1+\frac{q}{p}} = \frac{p}{p+q} = p \quad (q = 1-p),$$

$$\frac{M-k}{N+M-x-k} = \frac{\frac{M}{N}-\frac{k}{N}}{1+\frac{M}{N}-\frac{x+k}{N}} = \frac{\frac{q}{p}-\frac{k}{N}}{1+\frac{q}{p}-\frac{x+k}{N}} \overset{N\to\infty}{\longrightarrow} \frac{\frac{q}{p}}{1+\frac{q}{p}} = \frac{q}{p+q} = q$$

となるので、結局

$$p(x) \overset{N\to\infty}{\longrightarrow} \left(\begin{array}{c} n \\ x \end{array} \right) p^x q^y = \left(\begin{array}{c} n \\ x \end{array} \right) p^x q^{n-x}$$

となる。■