負の二項分布 (negative binomial distribution)

緑川章一

事象 $\mathcal{H}$ は確率 p で起き、その排反事象 $\mathcal{H}^*$ は、確率 $q = 1 - p$ で起こるとしよう。独立な事象を繰り返し、事象 $\mathcal{H}$ が r 回起きるまでに $\mathcal{H}^*$ の起こった回数を x とすると、その確率 $f(x)$ は、$r + x$ 回目に起こった事象は必ず $\mathcal{H}$ であることに注意すると、

$$f_r(x) = {}_{r+x-1}C_x p^r q^x,\ x = 0,\ 1,\ 2,\ 3,\ \cdots \tag{1}$$

となる。この分布を**負の二項分布** (negative binomial distribution) と言う。

この分布が負の二項分布と言われる所以は以下の通りである。まず、二項係数は、

$$\begin{aligned} {}_{r+x-1}C_x &= \frac{(r+x-1)(r+x-2)\ldots(r+1)r}{x!} \\ &= (-1)^x \frac{(-r)(-r-1)\cdots(-r-(x-2))\,(-r-(x-1))}{x!} \\ &= (-1)^x {}_{-r}C_x \end{aligned}$$

と書けるので、(1) 式は、

$$\begin{aligned} f_r(x) &= {}_{-r}C_x p^r (-q)^x \\ &= {}_{-r}C_x \left(\frac{1}{p}\right)^{-r-x} \left(-\frac{q}{p}\right)^x \end{aligned}$$

ここで、$N = -r,\ \ P = -q/p,\ \ Q = 1/p$ とおくと、

$$f_N(x) = {}_N C_x P^x Q^{N-x}, \tag{2}$$

となる。これは、形式的には二項分布と同じ形をしている。

確率の総和

$r = 1$ のとき、$f_1(x) = pq^x\ (x = 0,\ 1,\ 2,\ \cdots)$ は、幾何分布である。その確率の総和は、

$$S_1 = \sum_{x=0}^{\infty} f_1(x) = p \sum_{x=0}^{\infty} q^x = \frac{p}{1-q} = 1$$

次に一般の r の場合における (1) 式の和について考える。まず、

$$S_r = \sum_{x=0}^{\infty} f_r(x) = \sum_{x=0}^{\infty} {}_{r+x-1}C_{r-1} p^r q^x = \sum_{x=0}^{\infty} \frac{(r+x-1)(r+x-2)\cdots(x+1)}{(r-1)!} p^r q^x$$

と書こう。すると、

$$qS_r = \sum_{x=0}^{\infty} \frac{(r+x-1)(r+x-2)\cdots(x+1)}{(r-1)!} p^r q^{x+1} = \sum_{x=0}^{\infty} \frac{(r+x-2)\cdots(x+1)x}{(r-1)!} p^r q^x$$

となる。両者の差をとると、

$$
\begin{aligned}
(1-q)S_r &= pS_r \\
&= p\sum_{x=0}^{\infty}\frac{(r+x-2)\cdots(x+1)}{(r-2)!}p^{r-1}q^x \\
&= p\sum_{x=0}^{\infty}f_{r-1}(x) \\
&= pS_{r-1}
\end{aligned}
$$

ゆえに、

$$S_r = S_{r-1} = \cdots = S_1 = 1$$

を得る。すなわち、確率の総和は１である。

平均

$$
\begin{aligned}
\mu_r &= \sum_{x=0}^{\infty} x f_r(x) \\
&= \sum_{x=1}^{\infty} x\frac{(r+x-1)(r+x-2)\dots(r+1)r}{x!}p^r q^x \\
&= \sum_{x=1}^{\infty} \frac{(r+x-1)(r+x-2)\dots(r+1)r}{(x-1)!}p^r q^x \\
&= r\frac{q}{p}\sum_{x=0}^{\infty}\frac{(r+x)(r+x-1)\dots(r+1)}{x!}p^{r+1}q^x \\
&= r\frac{q}{p}\sum_{x=0}^{\infty} f_{r+1}(x) \\
&= r\frac{q}{p}
\end{aligned}
$$

ゆえに、

$$\mu_r = r\frac{q}{p} \tag{3}$$

を得る。ここで、確率の総和が１であること、すなわち、$\sum_{x=0}^{\infty} f_{r+1}(x) = 1$ を用いた。

分散

公式

$$\sigma_r^2 = \sum_{x=0}^{\infty} x^2 f_r(x) - \mu_r^2$$

を用いて計算する。

第１項は、

$$\sum_{x=0}^{\infty} x^2 f_r(x) = \sum_{x=0}^{\infty}\{x(x-1)+x\}f_r(x) = \sum_{x=2}^{\infty} x(x-1)f_r(x) + \mu_r$$

と書き直せる。この右辺第１項は、

$$
\begin{aligned}
\sum_{x=2}^{\infty} x(x-1) f_r(x) &= \sum_{x=2}^{\infty} x(x-1)\,{}_{r+x-1}C_x p^r q^x \\
&= \sum_{x=2}^{\infty} x(x-1) \frac{(r+x-1)(r+x-2)\ldots(r+1)r}{x!} p^r q^x \\
&= (r+1)r\frac{q^2}{p^2} \sum_{x=2}^{\infty} \frac{(r+x-1)(r+x-2)\ldots(r+2)}{(x-2)!} p^{r+2} q^{x-2} \\
&= (r+1)r\frac{q^2}{p^2} \sum_{x=0}^{\infty} \frac{((r+2)+x-1)\,((r+2)+x-2)\ldots(r+2)}{x!} p^{r+2} q^{x} \\
&= (r+1)r\frac{q^2}{p^2} \sum_{x=0}^{\infty} f_{r+2}(x) \\
&= (r+1)r\frac{q^2}{p^2}
\end{aligned}
$$

ここで、$\displaystyle\sum_{x=0}^{\infty} f_{r+2}(x) = 1$ を用いた。以上をまとめて、

$$\sigma_r^2 = (r+1)r\frac{q^2}{p^2} + r\frac{q}{p} - r^2\frac{q^2}{p^2} = r\frac{q}{p^2}$$

すなわち、

$$\sigma_r^2 = r\frac{q}{p^2}$$

を得る。

負の二項分布 (まとめ１)

確率分布 $f_r(x) = {}_{r+x-1}C_x p^r q^x \;(x = 0,\ 1,\ 2,\ 3,\ \cdots)$ の平均と分散は、

$$\text{平均} \;:\; \mu_r = r\frac{q}{p} \tag{4}$$

$$\text{分散} \;:\; \sigma_r^2 = r\frac{q}{p^2} \tag{5}$$

で与えられる。

さて、二項分布 $f_N(x) = {}_N C_x P^x Q^{N-x}$ の平均と分散は、

$$\text{平均} \;:\; \mu = NP$$

$$\text{分散} \;:\; \sigma^2 = NPQ$$

で与えられる。ここで、負の二項分布のパラメータ, $N=-r,\ P=-q/p,\ Q=1/p$ を代入すると、(４), (５) 式が得られることが分かる。

負の二項分布 (まとめ 2)

負の二項分布 $f_N(x) = {}_NC_x P^x Q^{N-x}$ (ただし、$N = -r,\ P = -q/p,\ Q = 1/p$) は、分布 $f_r(x) = {}_{r+x-1}C_x p^r q^x$ $(x = 0,\ 1,\ 2,\ 3,\ \cdots)$ と同等で、その平均と分散は、

$$\text{平均} \ :\ \mu_N = NP = r\frac{q}{p} \tag{6}$$

$$\text{分散} \ :\ \sigma_N^2 = NPQ = r\frac{q}{p^2} \tag{7}$$

で与えられる。

確率が、(1) 式で与えられるときに、もう一度、試行を行うと、確率 p で事象 $\mathcal{H}$ が、確率 q でその排反事象 $\mathcal{H}^*$ が起こる。もし、事象 $\mathcal{H}$ が起きたとすると、事象 $\mathcal{H}$ は、$r+x-1$ 回の間に、あと $r-1$ 回起きれば良いので、その確率は $f_{r-1}(x)$ となる。一方、排反事象 $\mathcal{H}^*$ が起こったとすると、事象 $\mathcal{H}$ は、$r+x-1$ 回の間に、あと r 回起きなければならないので、その確率は、$f_r(x-1)$ に変わる。ゆえに、漸化式

$$f_r(x) = pf_{r-1}(x) + qf_r(x-1) \tag{8}$$

が成り立つ。