⑪ 有限母集団での假説検定

労働科学研究所　　　門　山　允

最近ある村で行った世論調査の結果を見せてもらったが，その結果は次の通りであった。

昨年四月の投票政党

		母集団
社	56人（49.2%）	967人（36.2%）
民	3（2.6%）	107（4.0）
共	1（0.9%）	146（5.5）
自	46（40.3%）	1410（52.8）
國	0（0）	6（0.2）
ソノ他	11（7.0%）	35（1.3）
計	114（100）	2671

但し，調査の方では棄権も「ソノ他」の項に入れてある。

投票率は不明であるが，母集団の大さは3000程度と思はれる。この結果は社会党支持者の比率が高すぎるやうに思うので，この調査対象が母集団からの任意標本かどうか検定して

みたい。しかし、この場合，母集団の大さに比して，標本数がかなり大きいので，ふつうのやり方がすぐつかえるかどうかわからない。そこで，次のやうな問題を考へてみた。

大さ N な母集団から、n 個の標本をとるとき、母集団に P なる割合でふくまれるものが k 個以上入ってくる確率如何。

これを解くにはまず例によって赤い球と白い球で考え，赤い球が $PN = N_1$，白い球が $qN = N_2$ ふくまれているとする。まず n 個のうち，最初の n_1 個がすべて赤で，次の $n - n_1 = n_2$ 箇がすべて白である確率を求めると，最初の一箇が赤である確率は P_1 次のが赤である確率は $(N_1-1)/(N-1)$ 以下同様にして，最初の n_1 箇分が赤である確率は

$$\frac{N_1}{N}\ \frac{N_1-1}{N-1}\ \frac{N_1-2}{N-2}\ \cdots\cdots\ \frac{N_1-(n_1-1)}{N-(n-1)}$$

である。同様にして，次の n_2 箇が白である確率は

$$\frac{N_2}{N-n_1}\ \frac{N_2-1}{N-n_1-1}\ \cdots\cdots\ \frac{N_2-(n_2-1)}{N-(n-1)}$$

よつて，最初の n_1 箇が赤で，次の n_2 箇が白である確率は

$$\frac{N_1}{N}\ \frac{N_1-1}{N-1}\ \cdots\cdots\ \frac{N_1-(n_1-1)}{N-(n_1-1)}\ \frac{N_2}{N-n_1}\ \frac{N_2-1}{N-n_1-1}\ \cdots\ \frac{N_2-(n_2-1)}{N-(n-1)} \qquad \cdots\ (1)$$

である。この形をみると，赤と白の順序が変っても，分母の方はそのまゝで，たゞ分子の順序が入れかわるだけで，全体として不変なことがわかる。よって n 箇のうち n_1 箇が赤で n_2 箇が白である確率は

$$\binom{n}{n'}\frac{N_1}{N}\ \frac{N_1-1}{N-1}\ \cdots\ \frac{N_1-(n_1-1)}{N-(n_1-1)}\ \frac{N_2}{N-n_1}\ \frac{N_2-1}{N-n_1-1}\ \cdots\ \frac{N_2-(n_2-1)}{N-(n-1)}$$

$$= \binom{n}{n_1} \frac{N_1!}{(N_1-n_1)!} \cdot \frac{N_2!}{(N_2-n_2)!} \cdot \frac{(N-n)!}{N!} \tag{2}$$

である。　よつて，求める確率は

$$F(k) = \sum_{n_1 \geqq k} \binom{n}{n_1} \binom{N}{N_1}^{-1} \binom{N-n}{N_1-n_1}$$

$$= \binom{N}{N_1}^{-1} \sum_{n_1 \geqq k} \binom{n}{n_1} \binom{N-n}{N_1-n_1} \tag{3}$$

である。これを求めれば假説検定ができるわけであるが，それは一寸簡単に求まらないので，その上限を求めることにする。

（１）を書きなおすと

$$\frac{P}{1} \; \frac{P-1/N}{1-1/N} \cdots \frac{P-\frac{n_1-1}{N}}{1-\frac{n_1-1}{N}} \; \frac{q}{1-\frac{n_1}{N}} \cdots \frac{q-\frac{n_2-1}{N}}{1-\frac{n-1}{N}} \tag{4}$$

となる。　こゝで　$\dfrac{P-\frac{r}{N}}{1-\frac{r}{N}} \leqq p$

なることは明かで，又

$$\frac{q}{1-\frac{n_1}{N}} - \frac{q-\frac{i}{N}}{1-\frac{n_1}{N}-\frac{i}{N}} = \frac{\frac{i}{N}\left(P-\frac{n_1}{N}\right)}{\left(1-\frac{n_1}{N}\right)\left(1-\frac{n_1}{N}-\frac{i}{N}\right)}$$

よつて，$n_1/N \leqq n/N = \varepsilon < P$　ならば

$$\frac{q}{1-\frac{n_1}{N}} \geqq \frac{q-i/N}{1-n_1/N-i/N}$$

故に（４）は　$\leqq P^{n_1}\left(\dfrac{q}{1-\frac{n_1}{N}}\right)^{n-n_1}$

次に

$$\frac{1}{1-x} - (1+\alpha x) = \varphi(x)$$

とすると

$$\varphi'(x) = \frac{1}{(1-x)^2} - \alpha = \frac{1-\alpha(1-x)^2}{(1-x)^2}$$

$x = n_1/N \leq n/N = \varepsilon$ ならば、 $(1-x)^2 \geq (1-\varepsilon)^2$

よって、 $\alpha = (1-\varepsilon)^{-2}$ とすると、

$$\varphi'(x) \leq 0$$

しかるに、 $\varphi(0) = 0$ なる故、 $\varphi(x) \leq 0$.

すなわち

$$\frac{1}{1-x} \leq 1 + \alpha x$$

$$\therefore \quad \frac{1}{1-n_1/N} \leq 1 + n_1 \frac{\alpha}{N} \leq \left(1+\frac{\alpha}{N}\right)^{n_1}$$

$$\therefore \quad P^{n_1}\left(\frac{q}{1-n_1/N}\right)^{n-n_1} \leq \left(1+\frac{\alpha}{N}\right)^{n_1(n-n_1)} p^{n_1} q^{n-n_1}$$

$$\leq \left(1+\frac{\alpha}{N}\right)^{\left(\frac{n}{2}\right)^2} p^{n_1} q^{n-n_1}$$

$$\therefore \quad F(k) \leq \left(1+\frac{\alpha}{N}\right)^{\left(\frac{n}{2}\right)^2} \sum_{r \geq k} \binom{n}{r} p^r q^{n-r} \qquad (5)$$

右辺の $\sum$ はふつうの二項分布であるから、これで一応 $F(k)$ の上限がわかったわけである。始の例では、大体 $N = 3000$, $n = 100$, $p = 0.36$, $k = 50$ であるから、このときは、$\varepsilon = 0.033$ $\alpha = 1.070$, $1 + \alpha/N = 1.00036$ $\left(1+\frac{\alpha}{N}\right)^{\frac{n^2}{4}} = 2.455$ である。又(5)の右辺の $\sum$ はこのとき、0.01 以下で、したがって、

$$F(k) \leq 0.05$$

よって、この標本が、母集団からの任意標本であるという假説は 5 %の危険率ですてられる。

このやうなふうにやることができるが、もっと精密な限界又は、(3)の正確な値を求める方法があれば御教示願いたい。

(終)