変形二項分布モデル

――可変な成功率を有するベルヌーイ試行――

統計数理研究所　鈴　木　義一郎

(1976年9月　受付)

Modified Binomial Distribution Model

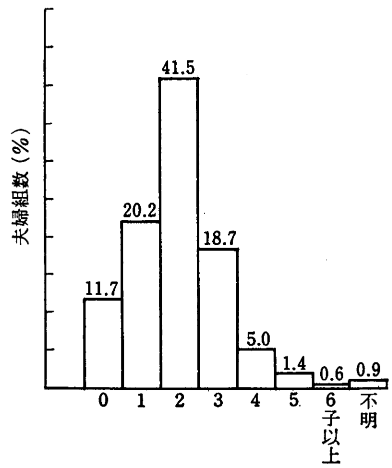
—Bernoulli trials with controllable success rate—

Giitiro Suzuki
(The Institute of Statistical Mathematics)

Examine Bernoulli trial with success rate p up to r ($<$n) times of success occur. Next changing p to $\tilde{p}$, continue n$-$r times of the new trial. The number of successes, X, is distributed as the modified binomial MB (n, r; p, $\tilde{p}$). In this note, we shall give expressions for the probability distribution (by (1)), mean (by (2)) and variance (by (3)) of the random variable X.

§1.　出生児数の分布

下のグラフは，人口問題研究所（[2]）の第6次出産力調査（標本数9355）にもとづく，出生児数毎の夫婦組数の分布状況である．これを見て気がつくのは，2子の夫婦が全体の4割強と圧倒的に多いことである．平均児数1.9人は少なすぎると感じるかもしれないが，無子夫婦が12%近くもいるためと考えられる．



さて子供が生れる，生れないを，妊娠可能となるべき試みの行為に伴う成功，失敗のように対応させたベルヌーイ試行とみなせば，各夫婦の出生児数（厳密には妊娠児数）は成功の回数ということになる．

だが，単純な二項モデルがそのままあてはまると考えるわけにはいかない．子供の数はほどほどにと考えている夫婦は，2人目位を生み了えた後は妊娠率を低くする（バス・コントロールを強化するなどして）作戦にでるのがノーマル・タイプの夫婦であろう．また，子供は何人でもよいというノバナシ・タイプから，いくら努力しても子供ができないという（妊娠率が0に近い）ウマヅメ・タイプと，大雑把に3つのタイプが混在しているものと考えられる．これら3種の割合が，それぞれどのくらいであるかを見積ること，試行回数，妊娠率といったパラメタをタイプ毎に推定すること，データとしては得られないであろう妊娠児数の分布と出生児数分布との間のカラクリを見極めること．このような問題がすべて解決されない限り，出生児数分布を説明してくれるうまい確率モデルをあてはめることはできない．

この小稿では，ノーマル・タイプの出生児数分布を記述するのに適当と考えられる，変形二項分布モデルについて考えてみる．

§2.　二項分布と変形二項分布

成功の確率が p（失敗は $q=1-p$）であるような試行を成功が r $(<n)$ 回起きるまで続け，その後は成功の確率を $\tilde{p}$ に変えて合計 n 回の試行を行う．このような変形ベルヌーイ試行での成功の回数を示す確率変数 X について考える．X の分布のことを，**変形二項分布**と呼び，記号で $MB\ (n,\ r;\ p,\ \tilde{p})$ のように表わすことにしよう．尚，各試行毎に成功率の異るものをポアソン試行と呼び，その確率分布に対しては**一般二項分布**という呼称が用いられているが，この変形二項分布はそのような一般二項分布とは異るものである．

X の確率分布が次式で与えられることは容易にわかる．

(1)
$$p_r\{X=s\}=\begin{cases} p_0(s) & 0\leqq s\leqq r-1 \\ p_r(s) & r\leqq s\leqq n \end{cases}$$

$$p_0(s)=\binom{n}{s}p^s q^{n-s}$$

$$p_r(s)=\sum_{k=r}^{n-s+r}\binom{k-1}{r-1}p^r q^{k-r}$$

（k 回目で丁度 r 回の成功）

$$\times\binom{n-k}{s-r}\tilde{p}^{s-r}\tilde{q}^{n-k-s+r}$$

（$k+1$ 回目以降 $s-r$ 回成功）

まず，(1) が確率分布となることを示してみよう．そのためには

$$\sum_{s=r}^{n}p_r\ (s)=\sum_{s=r}^{n}p_0\ (s)$$

となることを示せばよい．

$$\begin{aligned}
\sum_{s=r}^{n}p_r(s)&=\sum_{s=r}^{n}\sum_{k=r}^{n-s-r}\binom{k-1}{r-1}p^r q^{k-r}\times\binom{n-k}{s-r}\tilde{p}^{s-r}\tilde{q}^{n-k-s+r}\\
&=p^r\sum_{k=r}^{n}\binom{k-1}{r-1}q^{k-r}\sum_{s=r}^{n-k+r}\binom{n-k}{s-r}\tilde{p}^{s-r}\tilde{q}^{n-k-(s-r)}\\
&=p^r\sum_{k=r}^{n}\binom{k-1}{r-1}q^{k-r}(\tilde{p}+\tilde{q})^{n-k}\\
&=p^r\sum_{k=r}^{n}\binom{k-1}{r-1}q^{k-r}(p+q)^{n-k}\\
&=p^r\sum_{k=r}^{n}\binom{k-1}{r-1}q^{k-r}\sum_{i=0}^{n-k}\binom{n-k}{i}p^i q^{n-k-i}\\
&=\sum_{i=0}^{n-r}p^{r+i}q^{n-r-i}\sum_{k=r}^{n-i}\binom{k-1}{r-1}\binom{n-k}{i}
\end{aligned}$$

Feller [1] による次のような関係式

$$\sum_{j=0}^{m}\binom{a+j}{a}\binom{b+m-j}{b}=\binom{a+b+m+1}{a+b+1}$$

を利用すれば

$$\sum_{k=r}^{n-i}\binom{k-1}{r-1}\binom{n-k}{i}=\binom{n}{r+i}$$

従って

$$\sum_{s=r}^{n}p_r(s)=\sum_{i=0}^{n-r}\binom{n}{r+i}p^{r+i}q^{n-r-i}=\sum_{s=r}^{n}\binom{n}{s}p^sq^{n-s}=\sum_{s=r}^{n}p_0(s)$$

補助定理 1

(i) $$p\sum_{s=r}^{n}(s-r)\,p_r(s)=\tilde{p}\sum_{s=r}^{n}(s-r)\,p_0(s)$$

(ii) $$p^2\sum_{s=r}^{n}(s-r)\,(s-r-1)\,p_r(s)=\tilde{p}^2\sum_{s=r}^{n}(s-r)\,(s-r-1)\,p_0(s)$$

証　明

(i) $$\sum_{s=r}^{n}(s-r)p_r(s)=\sum_{s=r+1}^{n}(s-r)\,p_r(s)=p^r\tilde{p}\sum_{k=r}^{n-1}\binom{k-1}{r-1}(n-k)\,q^{k-r}$$

$$\times\sum_{s=r+1}^{n-k+r}\binom{n-k-1}{s-r-1}\tilde{p}^{s-r-1}\,\tilde{q}^{(n-k-1)-(s-r-1)}$$

$$=p^r\tilde{p}\sum_{k=r}^{n-1}\binom{k-1}{r-1}(n-k)\,q^{k-r}(\tilde{p}+\tilde{q})^{n-k-1}$$

$$=p^r\tilde{p}\sum_{k=r}^{n-1}\binom{k-1}{r-1}(n-k)\,q^{k-r}(p+q)^{n-k-1}$$

$$=p^r\tilde{p}\sum_{k=r}^{n-1}\binom{k-1}{r-1}(n-k)\,q^{k-r}$$

$$\times\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}p^iq^{n-k-1-i}$$

$$=p^r\tilde{p}\sum_{i=0}^{n-r-1}(i+1)\,p^iq^{n-r-1-i}$$

$$\times\sum_{k=r}^{n-i-1}\binom{k-1}{r-1}\binom{n-k}{i+1}$$

$$=p^r\tilde{p}\sum_{i=0}^{n-r-1}\binom{n}{r+i+1}(i+1)\,p^iq^{n-r-1-i}$$

$$=\frac{\tilde{p}}{p}\sum_{s=r+1}^{n}(s-r)\binom{n}{s}p^sq^{n-s}$$

(ii) $$\frac{p^2}{\tilde{p}^2}\sum_{s=r}^{n}(s-r)\,(s-r-1)\,p_r(s)$$

$$=\frac{p}{\tilde{p}^2}\sum_{s=r+2}^{n}(s-r)\,(s-r-1)\,p_r(s)$$

$$=p^{r+2}\sum_{k=r}^{n-2}\binom{k-1}{r-1}(n-k)\,(n-k-1)\,q^{k-r}$$

$$\times \sum_{s=r+2}^{n-k+r} \binom{n-k-2}{s-r-2} \tilde{p}^{s-r-2} \tilde{q}^{(n-k-2)-(s-r-2)}$$

$$= p^{r+2} \sum_{k=r}^{n-2} \binom{k-1}{r-1} (n-k)\ (n-k-1)\ q^{k-r}$$

$$\times \sum_{i=0}^{n-k-2} \binom{n-k-2}{i} p^i q^{n-k-2-i}$$

$$= p^{r+2} \sum_{i=0}^{n-r-2} (i+1)\ (i+2)\ p^i q^{n-r-2-i}$$

$$\times \sum_{k=r}^{n-i-2} \binom{k-1}{r-1} \binom{n-k}{i+2}$$

$$= \sum_{i=0}^{n-r-2} \binom{n}{r+i+2} (i+1)\ (i+2)\ p^{r+2+i} q^{n-(r+2+i)}$$

補助定理 2

(i) $$\sum_{s=r}^{n} (s-r)\ p_0(s) = np - \pi\ (r)$$

$$\pi\ (r) = \sum_{k=0}^{r-1} \sum_{i=k+1}^{n} p_0(i)$$

(ii) $$\sum_{s=r}^{n} (s-r)\ (s-r-1)\ p_0(s) = np\ (np+q) - (2r+1)\ np + 2\kappa\ (r)$$

$$\kappa\ (r) = \sum_{l=1}^{r} \pi\ (l)$$

証 明

(i) $$\sum_{r}^{n} (s-r)\ p_0(s) = \sum_{0}^{n} (s-r)\ p_0(s) + \sum_{0}^{r-1} (r-s)\ p_0(s)$$

$$= np - r + \sum_{0}^{r-1} (r-s)\ p_0(s) = np - r + \sum_{k=0}^{r-1} \sum_{i=0}^{k} p_0(i)$$

$$= np - \sum_{k=0}^{r-1} \left\{ 1 - \sum_{i=0}^{k} p_0(i) \right\} = np - \sum_{k=0}^{r-1} \sum_{i=k+1}^{n} p_0(i)$$

(ii) $$\sum_{r}^{n} (s-r)(s-r-1)\ p_0(s) = \left[\sum_{0}^{n} - \sum_{0}^{r-1} \right] (s-r)(s-r-1)\ p_0(s)$$

$$= np\ (np+q) - (2r+1)\ np + r\ (r+1) - \sum_{0}^{r-1} (r+1-s)(r-s)\ p_0(s)$$

$$= np\ (np+q) - (2r+1)\ np + 2 \sum_{l=1}^{r} \sum_{k=0}^{l-1} \left[1 - \sum_{i=0}^{k} p_0(i) \right]$$

$$= np\ (np+q) - (2r+1)\ np + 2\kappa\ (r)$$

§3. 変形二項分布のモーメント

変形二項分布の平均と分散を求めてみよう．成功率の変動比を $\gamma = \tilde{p}/p$ と置いて

$$\mu_{\langle r, \gamma \rangle} = E\ \{X\}$$

$$\sigma^2_{\langle r,\gamma\rangle}=V\{X\}$$

のように書き表すことにする．

定　理

(2) $$\mu_{\langle r,\gamma\rangle}=n\tilde{p}+(1-\gamma)\pi(r)$$

(3) $$\sigma^2_{\langle r,\gamma\rangle}=n\tilde{p}\tilde{q}+2rn\tilde{p}(1-\gamma)-(1-\gamma)(2n\tilde{p}-2r-1)\pi(r)-[(1-\gamma)\pi(r)]^2-2(1-\gamma^2)\kappa(r)$$

証　明

X の確率関数が $r-1$ までは2項分布 $B(n,\ p)$ と同一であるから

$$\mu_{\langle r,\gamma\rangle}-np=\sum_r^n s[p_r(s)-p_0(s)]=\sum_r^n(s-r)[p_r(s)-p_0(s)]$$

$$+r\sum_r^n[p_r(s)-p_0(s)]=\sum_r^n(s-r)[p_r(s)-p_0(s)]$$

$$=(\gamma-1)\sum_r^n(s-r)p_0(s)=(\gamma-1)\left[\sum_0^n-\sum_0^{r-1}\right](s-r)p_0(s)$$

$$=(\gamma-1)\left[np-r+\sum_0^{r-1}(r-s)p_0(s)\right]=(\gamma-1)\left[np-\sum_{k=0}^{r-1}\sum_{k+1}^{n}p_0(i)\right]$$

よって (2) の関係が証明された．

次に，補助定理1，2を用いれば

$$E\{X^2\}-np(np+q)=\sum_r^n s^2[p_r(s)-p_0(s)]$$

$$=\sum_r^n[(s-r)(s-r-1)+(2r+1)(s-r)+p^2]\times[p_r(s)-p_0(s)]$$

$$=(\gamma^2-1)\sum_r^n(s-r)(s-r-1)p_0(s)+(\gamma-1)(2r+1)\sum_r^n(s-r)p_0(s)$$

$$=(\gamma^2-1)[np(np+q)-(2r+1)np+2\kappa(r)]+(\gamma-1)(2r+1)[np-\pi(r)]$$

$$=(\gamma^2-1)np(np+q)\gamma(1-\gamma)(2r+1)np+2(\gamma^2-1)\kappa(r)-(\gamma-1)(2r+1)\pi(r)$$

$$=(n\tilde{p})^2+n\tilde{p}-n\tilde{p}\gamma(1-q)+2rn\tilde{p}(1-\gamma)-np(np+q)+2(\gamma^2-1)\kappa(r)-(\gamma-1)(2r+1)\pi(r)$$

関係式

$$\sigma^2_{\langle r,\gamma\rangle}=E\{X^2\}-E\{X\}^2$$

と (2) を用いて (3) の式を得る．

系1 ($r=1,\ 2$ の場合)

$$\mu_{\langle 1,\gamma\rangle}=n\tilde{p}+(1-\gamma)\sum_{i=1}^n p_0(i)=n\tilde{p}+(1-\gamma)(1-q^n)$$

$$\mu_{\langle 2,\gamma\rangle}=n\tilde{p}+(1-\gamma)\left\{\sum_{i=1}^n p_0(i)+\sum_{i=2}^n p_0(i)\right\}$$

$$=n\tilde{p}+(1-\gamma)\{2-2q^n-npq^{n-1}\}$$

$$\sigma^2_{\langle 1,\gamma\rangle}=n\tilde{p}\tilde{q}+2n\tilde{p}(1-\gamma)-(1-\gamma)(2n\tilde{p}-3)(1-q^n)-(1-\gamma)^2(1-q^n)^2-2(1-\gamma^2)(1-q^n)$$

$$= n\tilde{p}\tilde{q} - \gamma(1-\gamma) + (1-\gamma)(2n\tilde{p}+1)q^n - [(1-\gamma)q^n]^2$$

$$\sigma^2_{\langle 2,\gamma\rangle} = n\tilde{p}\tilde{q} + 4n\tilde{p}(1-\gamma) - (1-\gamma)(2n\tilde{p}-5)(2-2q^n-npq^{n-1}) - (1-\gamma)^2(2-2q^n-npq^{n-1})^2 - 2(1-\gamma^2)(3-3q^n-npq^{n-1})$$

$$= n\tilde{p}\tilde{q} - 2\gamma(1-\gamma) + 2(1-\gamma)(2n\tilde{p}+2-\gamma)q^n + (1-\gamma)(2n\tilde{p}+1-2\gamma)npq^{n-1} - [(1-\gamma)(2q^n+npq^{n-1})]^2$$

系2 $\left(\gamma=\frac{1}{2}\right.$ の場合)

$$\mu_{\langle r,1/2\rangle} = n\tilde{p} + \frac{1}{2}\pi(r)$$

$$\sigma^2_{\langle r,1/2\rangle} = n\tilde{p}(\tilde{q}+r) - \left(n\tilde{p}-r-\frac{1}{2}\right)\pi(r) - \frac{1}{4}\pi(r)^2 - \frac{3}{2}\kappa(r)$$

系3 $\left(\gamma=\frac{1}{2}\right.$ で $r=1,\ 2$ の場合)

$$\mu_{\langle 1,1/2\rangle} = n\tilde{p} + \frac{1}{2} - \frac{1}{2}(1-2\tilde{p})^n$$

$$\mu_{\langle 2,1/2\rangle} = n\tilde{p} + 1 - (n\tilde{p}+1-2\tilde{p})(1-2\tilde{p})^{n-1}$$

$$\sigma^2_{\langle 1,1/2\rangle} = n\tilde{p}(1-\tilde{p}) - \frac{1}{4} + \left(n\tilde{p}+\frac{1}{2}\right)(1-2\tilde{p})^n - \frac{1}{4}(1-2\tilde{p})^{2n}$$

$$\sigma^2_{\langle 2,1/2\rangle} = n\tilde{p}(1-\tilde{p}) - \frac{1}{2} + \frac{3}{2}(1-2\tilde{p})^n + 2(n\tilde{p}+1-2\tilde{p})n\tilde{p}(1-2\tilde{p})^{n-1} - [(n\tilde{p}+1-2\tilde{p})(1-2\tilde{p})^{n-1}]^2$$

§4. 結　　語

この小稿では扱わなかったが，このような変形二項分布に従う観測値をもとに，r, p, $\tilde{p}$ といったパラメータを推定することも興味のある問題である．さらに，モーメントの計算式が繁雑であるから，パラメータの代表的な値毎に，数表を作っておくことも有意義であろう．これらの話題に関しては，いずれ稿をあらためて言及する予定である．

最後に，査読者諸氏より有益なるコメントを頂いたことを付記し深謝の意を表わす．

参　考　文　献

[1] Feller, W (1957), An Introduction to Probability Theory and its Applications, second edition, John Wiley, p. 62.

[2] 人口問題研究所 (1972)，第6次出産力調査報告，p. 41.