工業經營學會誌 第 21卷 第 46輯 1998年 5月 127

⊠ 연구논문

# 일반화 이항분포에 관한 연구

## -A study for Generalized Binomial Distributions-

이 병 수
Lee, Byoung Soo
김 희 철
Kim, Hee Chul

## ABSTRACT

In many cases where the binomial distribution fails to apply to real world data it is because of more variability in the data than can be explained by that distribution. Several authers have proposed models that are useful in explaining extra-binomial variation. In this paper we point out a characterization of sequences of exchangeable Bernoulli variables which can be used to develop models which show more variability than the binomial. We give sufficient conditions which will yield such models and show how existing models can be continued to generate further models. A numerical example and simulation given.

## 1. 서 론

이산형 분포 중에서 이항분포의 가정은 다음과 같은 가정을 만족해야 한다. 즉 ① $n$번 시행의 열(sequence)이 존재하고 ② 각 열이 성공 혹은 실패로서 분류된다. 그리고③ 성공의 확률은 각 시행에서 동일하고 ④ 시행들은 독립성을 가지고 있다.

본 논문은 ①-③은 만족하지만 ④ 는 만족하지 않는 분포 족을 연구하고자 한다. ④ 의 독립성 가정 대신에 호환성(exchangeability)을 고려하고자 한다. 많은 연구가

들은 두갈래로 나누어지는 자료는 흔히 결과의 종속성 때문에 이항분포를 따르지 않는다는 사실을 인정하고 있다. 이러한 종속성 때문에 이항자료에서 기대하였던 결과가 보다 많은 차이를 나타내고 있다.2)

Rudolfer(1990) 추가변동이항(extra-binomial variation)모형들을 재조명하였고, 이러한 모형 중 이항분포의 모수가 사전분포(prior distribution)인 베타분포를 가지는 분포가 베타이항모형이다.

Altham(1978)은 이항분포에서 2가지 일반화 형태를 제안하였다. 이것은 베르누이(실패 0, 성공 1)변수는 시행이 반복할 때 교호작용(interaction)이 있는 대칭결합분포(symmetric joint distribution)를 가정하였다. 이러한 일반화는 교호작용에 대한 첨가(additive) 혹은 증가(multiplicative)라는 정의로서 제기하였다. Kupper & Haseman(1978)은 상관이항모형(correlated binomial model)을 제시하였다. Ng(1989)는 이항분포의 혼합(mixture)을 이용하여

* 인천대학교 전자계산학과 교수
** 동국대 통계학과 박사과정 수료
† 본 연구는 1996년 인천대학교 연구비 지원에 의해 수행되었음.

개선된 이항분포(modified binomial distribution)라고 불리는 모형의 분류를 시행하였다.
Paul(1985, 1987)은 이항분포에 대한 3가지 모수의 일반화를 시행하였고 Verducci, Mack, & De Groot(1988)들은 주변분포측면에서 비교를 이용하여 Altham과 다른 증가(multiplicative) 모형을 제기하였다. Morton(1991)은 전체적으로 2가지 추가 포아송(Poisson) 변수에 조건을 부여함으로서 추가이항변동에 대한 모형을 제시하였다.
여러 연구가들은 적합도 검정이나 이항회귀 방법을 이용한 검정을 통하여 모수의 추정에 관련된 문제들을 제시하고 있다. 예를 들면 Crowder(1985), Prentice(1986), Pack(1986), Moore(1987), & Paul, Liang 와 Self(1989)등이 있다.
본 논문은 2절에서는 일반화 이항분포의 특징을, 3절에서는 일반화이항분포의 모형을 4절에서는 베타이항분포를 서술하였고 5절에서는 모의실험 및 결과를 나열하였다.

## 2. 일반화 이항분포의 특성

표기를 일반화하기 위해 $X_1, X_2, \cdots$은 호환가능(ex- changeable)한 베르누이 확률변수의 수열(sequence)라고 하고 다음과 같이 가정하자.

$$\pi_n = P(X_1=1, X_2=1, \cdots, X_n=1),\ \pi_0 = 1 \qquad (2.1)$$

예를 들어 베타이항분포(beta-binomial distribution)는 사전분포가 베타분포를 따르는 이항분포로 표현이 가능하다. 즉 이 과정을 2-단계로서 이루어진다고 생각 할 수 있다. 1-단계는 $p$의 값은 베타분포에서 추출되고 2-단계는 이항분포에서 모수 $n$이 알고 있을 때 $p$의 값이 추출된다. 이러한 과정을 deFinetti의 정리에 연관시켜 , 호환가능이 연속적으로 이루어진다면, 결합확률은 이항분포의 혼합(mixture)으로 표현이 가능하다. 혼합된(mixing) 분포에 대한 누적분포함수를 $\zeta(p)$라고 표현한다면,

$$P(X_1=1, X_2=1, \cdots, X_k=1, X_{k+1}=0, \cdots, X_n=0) = \int_0^1 p^k(1-p)^{n-k} d\zeta(p) \qquad (2.2)$$

(참고 : Feller(1966))

(2.1)에서 정의된 $\pi_n$을 사용하기 위해 (2.2)식을 이용하면

$$\pi_n = P(X_1=1, X_2=1, \cdots, X_n=1) = \int_0^1 p^n d\zeta(p) \qquad (2.3)$$

$\pi_n$이 $[0,1]$구간 내에서 존재하는 확률이기 때문에 $\{\pi_n\}$의 열은 혼합분포 $\zeta(p)$의 적률(moment)을 제공한다. 그러나, $[0,1]$구간에 존재하는 모든 열들이 확률분포의 적률을 제공하는 것이 아니기 때문에 $\{\pi_n\}$에 조건을 부여해서 확률분포의 적률을 제공되도록 하고자 한다. 이러한 확률분포는 호환가능 베르누이 확률변수의 결합분포를 발생시키기 위해 (2.2)에 사용된 혼합분포를 이용해야 한다. 이러한 결합분포를 이용해서 $Z_n = \Sigma X_i$의 확률분포를 나타낼 수 있다.
$\{\pi_n\}$의 적당한 선택에 의해 추가이항변량(extra binomial variation)을 나타내는 $Z_n$에 대한 다양한 분포를 나타낼 수 있다.

적률 $\{\pi_n\}$의 열의 차(difference)에 대한 모든 차수들(orders)이 비음(non-negative) 한다는 조건을 만족해야만 되는 것을 보이고자 한다. $X_1, X_2, \cdots$의 베루누이 확률변수의 열을 고려하자, 즉 1 혹은 0을 가지는 변수들이다. 일반적으로 성공이 나타나는 관심의 사건들에 대한 지시변수들로 생각할 수도 있다.

열은 임의의 $n$에 대한 호환가능(exchangeable)하다면 (deFinetti, 1975), 임의의 $m$에 대하여 $X_{n_1}, X_{n_2}, \cdots, X_{n_m}$의 분포함수 $F(\cdots,\cdots,\cdots)$은 $X_{n_i}$가 어떻게 선택되든 같다. 즉 $e_i=0$, 혹은 1이 된다.

$$
\begin{aligned}
&P(X_1=e_1, X_2=e_2, \cdots, X_n=e_n)\\
&\quad =P(X_1=e_{a_1}, X_2=e_{a_2}, \cdots, X_n=e_{a_n})
\end{aligned}
$$

단, $(a_1, a_2, \cdots, a_n)$은 정수들 (1, 2, ⋯, $n$)의 어떤 순열(permutation)이다. 이것은 어떤 열에서든 $n$번 시행에서 $k$번 성공하는 열의 확률은 같음을 의미한다. 예를 들어 $X_1 \ldots X_{n-1}$의 결합분포는 $X_n$의 값을 더함으로서 $X_1, \ldots, X_n$의 결합분포를 결정할 수 있다.

(2.1)에 정의된 $\pi_n$을 사용하여 한번 성공할 확률은

$$\pi_1 = P(X_1=1) = P(X_i=1).$$

$$
\begin{aligned}
\pi_2-\pi_1^2 &= P(X_1=1, X_2=1)-\pi_1^2\\
&= E(X_1X_2)-E(X_1)E(X_2)\\
&= COV(X_1, X_2) = COV(X_i, X_j), (i \neq j)
\end{aligned}
$$

앞으로의 논의를 위해서 다음과 같은 정의를 하자.

| |
|---|
| <정의 1> |
| $a_{n,k} = P(X_1=i_1, X_2=i_2, \cdots, X_n=i_n \mid \sum i_j = k)$ |
| 즉, $a_{n,k}$는 $n$번 관찰치중 정확하게 $k$번 성공이 포함하는 열의 확률. |

$\{\pi_n\}$과 $\{a_{n,k}\}$의 내적 관계를 나타내기 위해 $n=2,3$에 대해서 설명하고자 한다.

순열불변성(permutation Invariance)의 가정 하에서 $a_{n,k}$의 값들은 $n$에 대하여 $X_1, X_2, \cdots, X_n$의 결합분포로 정의될 수 있다. 주변확률을 고려해서 $a_{n,k}$의 값들이 주어진다면 $\pi_n$의 값들을 알아낼 수 있다.

$n=2$에 대하여, 가능한 결과와 확률은 다음과 같다.

| $X_1$ | $X_2$ | $P(X_1=i, X_j=j)$ |
|---|---|---|
| 0 | 0 | $a_{2,0}$ |
| 0 | 1 | $a_{2,1}$ |
| 1 | 0 | $a_{2,1}$ |
| 1 | 1 | $a_{2,2}$ |

$$\therefore\ a_{2,2}=\pi_2,\quad a_{2,1}+a_{2,2}=\pi_1,\ a_{2,0}+2a_{2,1}+a_{2,2}=1=\pi_0$$

$n=3$에 대하여는

| $X_1$ | $X_2$ | $X_3$ | $P(X_1=i_1, X_2=i_2, X_3=i_3)$ |
|---|---|---|---|
| 0 | 0 | 0 | $a_{3,0}$ |
| 0 | 0 | 1 | $a_{3,1}$ |
| 0 | 1 | 0 | $a_{3,1}$ |
| 0 | 1 | 1 | $a_{3,2}$ |
| 1 | 0 | 0 | $a_{3,1}$ |
| 1 | 0 | 1 | $a_{3,2}$ |
| 1 | 1 | 0 | $a_{3,2}$ |
| 1 | 1 | 1 | $a_{3,3}$ |

$a_{3,k}$와 $\pi_k$의 관계는

$$\begin{aligned} a_{3,0}+3a_{3,1}+3a_{3,2}+a_{3,3}&=1=\pi_0 \\ a_{3,1}+2a_{3,2}+a_{3,3}&=\pi_1 \\ a_{3,2}+a_{3,3}&=\pi_2 \\ a_{3,3}&=\pi_3 \end{aligned}$$

행렬과 벡터를 사용하여 윗관계를 설명하면,
$\pi_n=(\pi_0, \pi_1, \cdots, \pi_n)'$, $A_n=(a_{n,0}, a_{n,1}, a_{n,2}, \cdots, a_{n,n})'$

$n=2$에 대하여　$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} A_2=\pi_2$ ,

$n=3$에 대하여　$\begin{bmatrix} 1 & 3 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} A_3=\pi_3$

일반적으로, $A_n$의 계수행렬은(matrix of coefficients)은 이항계수를 포함하는 열로 구성된다. $M_n$을 $(n+1)\times(n+1)$의 행렬로 생각하면, $M_n A_n=\pi_n$으로 나타낼 수 있다.

$$\begin{bmatrix} \binom{n}{0} & \binom{n}{1} & \cdots & \binom{n}{n-2} & \binom{n}{n-1} & \binom{n}{n} \\ 0 & \binom{n-1}{0} & \cdots & \binom{n-1}{n-3} & \binom{n-1}{n-2} & \binom{n-1}{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 1 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{n,0} \\ a_{n,1} \\ \vdots \\ a_{n,n-1} \\ a_{n,n} \end{bmatrix} = \begin{bmatrix} \pi_{.0} \\ \pi_1 \\ \vdots \\ \pi_{n-1} \\ \pi_n \end{bmatrix}$$

여기서, $a_{n,k}$값의 사용은, $n$번 시행 하에 성공의 총수인 $Z_n$의 분포를 얻기 위함이다.
즉,

$$P(Z_n=k)=\binom{n}{k}a_{n,k}$$

$\pi_k$가 주어지고, $A_n$을 풀기 위해서는 $M_n$의 역행렬을 구해야 한다. 그러나, $M_n$의 역행렬은 대각

요소가 양(positive)인 부호를 가지고, 비영(non-zero)요소를 가진다.
예를 들어, $M_3^{-1}$을 사용하여 $A_3$을 얻기 위해서는, 즉

$$\begin{bmatrix} a_{3,0} \\ a_{3,1} \\ a_{3,2} \\ a_{3,3} \end{bmatrix} = \begin{bmatrix} 1 & -3 & 3 & -1 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix}$$

$a_{3,k}$는 확률이기 때문에, 비음(non-negative)이 되어야 한다. 즉 $\pi_3 = a_{3,3} \geq 0$이 된다. 그러므로 다음 경우도 성립한다.

$\pi_k = a_{k,k} \geq 0$, $a_{3,2} = \pi_2 - \pi_3 \geq 0$, $a_{n,n-1} = \pi_{n-1} - \pi_n \geq 0$

여기서 $\pi_k$는 비감소(non-increasing), 혹은, $\pi_k$의 1차수(the first difference)는 비음(non-negative)이 되어야 한다.

다음은, $a_{3,1} = \pi_1 - 2\pi_2 + \pi_3 \geq 0$, 일반적으로

$a_{n,n-2} = \pi_{n-2} - 2\pi_{n-1} + \pi_n \geq 0 = (\pi_{n-2} - \pi_{n-1}) - (\pi_{n-1} - \pi_n) \geq 0$

이러한 조건들은 $\pi_k$의 두 번째 차수(the second di- fference)도 비음(non-negative)이 되어야 한다. deFinetti의 정리에 의하면, 비음을 가지는 $\{\pi_k\}$의 열들은 $\pi_0 = 1$과 모든 차수(all orders of difference)도 비음이 된다는 것을 베루누이 확률변수의 호환가능(exchangable)열의 결합분포에 대해서도 설명이 가능하다.

(2.3)에 대해서 이러한 열들은 $[0,1]$에서 정의된 확률분포 $\xi(p)$의 적률(moment)을 나타낼 수 있다. 이러한 분포 $\xi(p)$는 호환가능한 베루누이 확률변수열을 정의하기 위한 deFinetti의 정리에 의해 구체화된 혼합분포(mixing distribution)이다. (Feller, 1966) 이 결과를 요약하기 위하여, 호환가능한 베르누이시행(exchangeable Bernoulli trials)을 EXBERT라고 간단히 명명하면 다음과 같은 정리로 그 결과를 요약 할 수 있다.

> <정리 1> 확률열 $\pi_k, (k=0,1,\cdots)$는 $\pi_0 = 1$이고 모든 차수의차(all orders of difference)가 비음(non-negative)이면 EXBERT분포로 정의된다.

<정리 1>의 조건을 만족하는 $\{\pi_n\}$의 열이 주어지면, $a_{n,k}$의 값이 결정되고, $n$에 대한 $Z_n$의 확률분포에 대한 확실한 표현을 위해 (4)를 사용할 수 있다.
즉, $M_n$의 역행렬을 이용하여

$$P(Z_n = k) = \binom{n}{k} \sum_{j=0}^{n-k} \binom{n-k}{j} (-1)^j \pi_{k+j}$$

여기서 $\{\pi_n\}$열은 호환가능 베르누이 변수에 대한 결합분포를 발생시키는데 사용될 수 있다.

## 3. 일반화 이항분포의 모형

일반화 이항분포의 족은 $\{\pi_n\}$의 형태에 의해서 여러 가지로 나타낼 수 있다. 예를 들어, $\pi_n = p^k$ 라고 했을 때 이 열은 정리1의 조건을 만족한다. j번째 차수의 차는 $(1-p)^j p^i$의 형태가 되며, [0,1]의 구간 내에서 $p$에 대해 비음(non-negative)이 된다.
이러한 족(family)은 이항분포에 해당하고, $a_{n,k} = p^k(1-p)^{n-k}$가 된다.
또 다른 예로, $\pi_0 = 1, \pi_k = p(k \geq 1)$일 때, 역시 정리1을 만족한다. 그리고, 전체적으로 종속(dependence)의 경우에 해당한다. 즉 $a_{n,n} = p, a_{n,0} = (1-p)$. 이러한 의미는 n번 시행에서 n번 성공할 확률이 $p$이고 $n$번 성공하지 못할 확률이 $1-p$이다.
정리1을 확장해서 열들의 선형조합도 역시 <정리 1>을 만족함을 보여준다.

> <정리 2>. $\{\pi_k\}$와 $\{V_k\}$은 각각 $\pi_0 = 1, V_0 = 1$, 모든 차수 차(all difference)가 비음(non-negative)일 때, $\{\alpha\pi_k + (1-\alpha)V_k\}$, $0 \leq \alpha \leq 1$도 역시 정리1을 만족한다.

<정리 2>를 이용하여 두 모수족(two parameter families)에 대한 분포를 나타낼 수 있다. 앞의 2개의 예를 혼합하여 EXBERT분포를 나타낼 수 있다.

$$\pi_0 = 1, \pi_k = \alpha p + (1-\alpha)p^k, k \geq 1 \tag{3.1}$$

이 분포에서 두 모수 $\alpha, \pi$에 대하여, $k$=1 일 때 $\pi_1 = \alpha P + (1-\alpha)P^1 = F$ 그러므로, $\pi_1 = P(X_1 = 1)$이 되고, 모수 p는 주어진 베루누이 시행에서 성공의 확률로 해석할 수 있다. 절편 $\alpha$에 대하여, $X_1$과 $X_2$의 상관으로 생각할 수 있다.
즉, 공분산은

$$\begin{aligned} & E(X_1X_2) - E(X_1)E(X_2) \\ & = \pi_2 - (\pi_1)^2 \\ & = [\alpha p + (1-\alpha)p^2] - p^2 = \alpha p - \alpha p^2 = \alpha p(1-p) \end{aligned}$$

$X_i$의 분산은 $p(1-p)$이므로,

$$\rho_{X_1, X_2} = \frac{COV(X_1, X_2)}{\sqrt{p(1-p)p(1-p)}} = \frac{\alpha p(1-p)}{p(1-p)} = \alpha$$

즉 $\alpha$는 베르누이 변수들 사이의 상관계수가 되므로

$$\pi_k = \rho P + (1-\rho)P^k, k \geq 1.$$

$\rho = 0$일 때, 독립인 경우(표준 이항분포)이고, $\rho = 1$일 때 완전한 종속인 경우가 된다. 이 모형을 상관모형(Correlation model)으로 언급된다.
다른 두 모수에 대한 분포를 살펴보면

$$\pi_k = P^{k^a} = e^{(\ln p)k^a} \quad 0 \leq a \leq 1, 0 < p < 1, k = 1, 2, 3, \ \Pi 0 = 1$$

$a=1$일 때는 독립 혹은 이항모형을 얻을 수 있다. $a=0$일 때는 종속모형을 얻을 수 있다. 이 모형을 멱모형(power model)으로 언급된다.
Madson이 제안한 모형은(1993) $\pi_k = \beta + (1-\beta)\alpha^k$으로 놓았다. $\beta = 0$일 때, $\alpha = p$라면 이항분포가 되고, $\alpha = 0$이면 종속모형을 유도할 수 있다. (2.2)에서 $n$번 시행에 $k$번 성공확률에 대하여 $\{\pi_n\}$의 집합에 의해 모델을 정의하였다.
이러한 확률의 형태가 상관모형에 있어서는 비교적 간단하나, Madson모형은 특별한 경우이다.

$n$개의 종속베르누이 시행에서 $Z_n$은 성공의 총 횟수를 나타낼때, 상관모형을 일반화시키면 $\pi_0=1, \pi_k=\rho p+(1-\rho)p_k,\ k\geq 1$일 때

$$P(Z_n=k)=\begin{cases}\rho(1-p)+(1-\rho)(1-p)^n & \text{for } k=0\\ (1-\rho)\binom{n}{k}p^k(1-p)^{n-k} & \text{for } 1\leq k\leq n-1\\ \rho p+(1-\rho)(p)^n & \text{for } k=n\end{cases} \quad (3.2)$$

Madson모형에서, $\pi_k=\beta_1+(1-\beta_1)\alpha_1^k,\ k\geq 0$

$$P(Z_n=k)=\begin{cases}(1-\beta_1)\binom{n}{k}\alpha_1^k(1-\alpha_1)^{n-k} & \text{for } 0\leq k\leq n-1\\ \beta_1+(1-\beta_1)(\alpha_1)^n & \text{for } k=n\end{cases} \quad (3.3)$$

이러한 경우 $\rho=0, \beta_1=0$일 때는 각각 독립된 경우로 표현할 수 있다.
멱모형에서

$$P(Z_n=k)=\binom{n}{k}p(1-p),\quad k=1,2,3,\cdots$$

## 4. 베타-이항분포

$X_1 X_{2,\ldots} X_n$ 이 i. i. d 이고 Bernoulli($p$) 시행을 하고 $Z_n=\Sigma X_i$라고 하면 $p$에 대한 사전분포를 베타분포 $Beta(\alpha_2, \beta_2)$로 주어진다면 다음과 같은 분포를 유도할 수 있다. 즉

$$\begin{aligned} f(k,p) &= f(k|p)\times f(p)\\ &= \left[\binom{n}{k}p^k(1-p)^{n-k}\right]\left[\frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)}p^{\alpha_2}-1\ (1-p)^{\beta_2}-1\right]\\ &= \binom{n}{k}p^{k+\alpha_2-1}(1-p)^{n-k+\beta_2-1}\frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)} \end{aligned}$$

$$k=0,1,\ldots n,\quad 0\leq p\leq 1$$

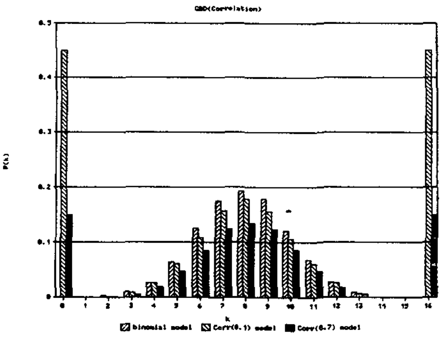
$$\begin{aligned} f(k) &= \int_0^1 f(k,p)dp\\ &= \binom{n}{k}\frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)}\ \frac{\Gamma(k+\alpha_2)\Gamma(n-k+\beta_2)}{\Gamma(n+\alpha_2+\beta_2)} \end{aligned} \quad (4.2)$$

$$\left[\because \int_0^1 \frac{\Gamma(n+\alpha_2+\beta_2)}{\Gamma(k+\alpha_2)\Gamma(n-k+\beta_2)}p^{k+\alpha_2-1}(1-p)^{n-k+\beta_2-1}\,dp=1\right]$$

즉, 식 (4.2)는 베타-이항분포이다.

## 5.모의실험 및 결과

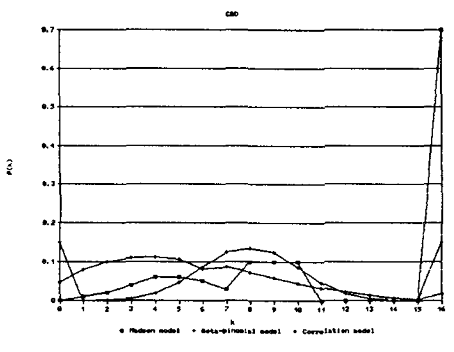
상관모형(3.2)의 경우 $\rho=0$일 때 이항분포가 되는지를 모의실험을 통하여 고찰하고 또, 상관계수가 $\rho=0.1$와 $\rho=0.7$일 경우를 각각 $n=16, k=0,1,2\cdots n$ 에 대하여 C-언어의 rand()함수를 통하여 0에서 1의 난수를 발생하여 그 시행을 20000회 시행하여 그 결과가 (그림1)에 나타내었다. 이 그림에서 $\rho=0.1$일 경우는 $k=0, k=16$ 일 때의 밀도가 이항분포의 밀도보다 큰 확률값을 나타내고 있지만 그 밖의 경우는 이항밀도가 더 높다. 그리고 상관관계가 비교적 높은 $\rho=0.7$일 때도 $\rho=0.1$ 일 경우와 유사하지만 $k=0, k=16$ 일 때의 밀도가 작은 확률값을 가지고 있고 그 밖의 경우도 이항밀도보다는 작고 $\rho=0.1$와 비교할 때 부분적으로 높은 확률을 가지는 경우도 볼 수 있다. 결과적으로 상관관계가 존재하면 추가변동이항(extra-binomial var- iation)의 성질이 있음을 확인해 주고 있는 셈이다.



(그림1) 이항분포모형과 상관모형의 비교
(Fig.1) Comparision of binomial and correlated model

상관모형의 특수한 경우인 Madson이 제시한 모형(3.3)과 모수의 사전분포가 베타분포인 베타-이항분포(4.2)를 비교 하고자 한다. 상관모형과 Madson 모형은 각각 $\rho=0, \beta_1=0$ 일 때는 독립분포가 된다.

$\rho=0.7$일 때의 상관모형과 $\alpha_1=0.3,\ \beta_2=0.7$일 경우와 Madson 모형의 밀도를 알아보고 또, $n=16,\ \alpha_2=2,\ \beta_2=4$일 경우의 베타-이항분포의 밀도를 고찰해 보았고 이 경우의 베타-이항분포표가 <표1>에서 요약되었고 비교결과는 (그림 2)에서 보여준다. 이 비교에서 Mad- son모형은 $k=16$일 때는 가장 밀도가 크고 대체적으로 상관모형보다 밀도가 작게 나타남을 볼 수 있다. 베타-이항분포는 $k$가 7에서부터 10까지는 밀도가 가장 높게 나타나지만 그 외에는 상관모형이 크게 나타나고 있다. $k$가 13에서부터 15까지는 비교 모형들이 밀도가 거의 유사하게 나타나고 있음을 보여 준다. 결과적으로 $k$에 따라서 효율적인 모형이 결정된다.



(그림 2) 일반화 분포모형의 비교
(Fig. 2) Comparision of Generl Binomial Distribution

위 그림에서 볼 수 있듯이 두갈래로 나누어지는 자료는 흔히 결과의 종속성 때문에 이항분포를 따르지 않는다는 사실을 인정할 수 있다. 독립성 가정 대신에 호환성(exchangeability)을 고려하면 즉, 종속성 때문에 이항자료에서 기대하였던 결과가 보다 많은 차이를 나타내고 있다. 결과적으로 $n$이 크고 발생횟수($k$)가 비교적 크면 비교모형이 거의 유사한 밀도를 가진다. 그러므로 앞으로의 연구는 <정리1>에 의한 모형뿐만 아니라 <정리2>에 연관된 혼합일반화 분포의 연구가 행해져야 할 것이다.

<표1> $n=16,\ \alpha_2=2,\ \beta_2=4$의 베타-이항분포표

<Table1> Beta-binomial distribution table

| $k$ | $p(k)$ | $p(K \le k)$ |
|---|---|---|
| 0 | 0.04762 | 0.04762 |
| 1 | 0.08020 | 0.12782 |
| 2 | 0.10025 | 0.22807 |
| 3 | 0.11008 | 0.33815 |
| 4 | 0.11180 | 0.44995 |
| 5 | 0.10733 | 0.55708 |
| 6 | 0.08084 | 0.65546 |
| 7 | 0.08649 | 0.74195 |
| 8 | 0.07298 | 0.81493 |
| 9 | 0.05897 | 0.87390 |
| 10 | 0.04541 | 0.91931 |
| 11 | 0.03302 | 0.95233 |
| 12 | 0.02236 | 0.97469 |
| 13 | 0.01376 | 0.98845 |
| 14 | 0.00774 | 0.99582 |
| 15 | 0.00315 | 0.99897 |
| 16 | 0.01754 | 1.00000 |

## 참 고 문 헌


[1] Altham,P.M.E. "Two Generalizations of the Binomail Distribution." Applied statics 27, 1978 pp. 62-167.

[2] Crowder, M. " Gaussian Estimation for correlated binomial data." Journal of the Royal Statistical Society Series b 47,1985 pp. 229-237.

[3] DeFinetti, B. "Theory of Probability Vol. 2." New York: Wiley 1975

[4] Feller, W. "An Introduction to Probability Theory and Its Applications, Vol. 2. 2nd ed." New York:Wiley 1966

[5] Kupper, L. L. and Haseman, J. K. " The use of a correlated binomial model for the analsis of certain toxicological experiments." Biometrics 34, 1978 pp.69-76.

[6] Madson, R, W. "Generalized Binomial Distributions", Communications Statistics-Theory and Methods, 22(11), 1993, pp. 3065-3086

[7] Moore, D. F." Modeling extraneous variance." Applied Statistics 36, 1987, pp. 8-14.

[8] Morton, R. " Analysis of extra-multinomial data derived from extra-Poisson variables conditional on their total" Biometrika 78, 1991, pp.. 1-6.

[9] Ng, T. H. "A new class of modified binomial distributions with applications to certain-toxicological experiments", Communications in Statistics", Theory and Methods 18, 1989, pp.3477-3492,

[10] Pack, S. E. " Hypothesis testing for pro portions with overdispersion.", Biometrics 42,1986, pp.967-972.

[11] Paul, S. R. "A three-parameter generalization of the binomial distribution", Commuṇications in Statistics-Theory and Methods 14, 1985, pp.497-1506.

[12] Paul, S. R. "On the beta-correlated binomial (BCB) distribution-A three-parameter generalization of the binomial distribution.", Communications in Statistics - Theory and methods 16, 1987, pp. 473-1478.

[13] Paul, S. R., Liang, K. Y. and self, S. G. "On testing departure from the binomial and multinomial assumptions", Biometrics 45, 1989 pp.231-236.

[14] Prentice, R. L. "Binary regression using an extended beta-binomial distribution, with discussion of association induced by covariate measurement errors", Journal of the American Statistical Association 81, 1986, pp. 321-327.

[15] Rudolfer, S. M. "A Markov chain model of extrabinomial variation.", Biometrika 77, 1990, pp. 255-264.

[16] Skellam, J. G. " A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials", Journal of the Royal Statistical Society Series B 10, 1948, pp. 257-261.

[17] Verducci, J. S.,Mack, M. E. and M. DeGroot, M.H "Estimating Multiple Rater Agreement", Journal of Multivariate Analysis 27, 1988 pp.512-538.