*Journal of the Korean*
*Data & Information Science Society*
*1999, Vol. 10, No. 2, pp. 279 ~ 288*

# 일반화 이항분포모형에서 시행간 종속성 규정모수의 추정량 비교 연구

문 명 상 [1]

**요 약**

통계자료분석에서 많이 다루는 이원자료(binary data)는 고전적인 이항분포모형에서 가정하는 시행간 독립성이 결여된 경우가 대부분이므로 그 자료에 고전적 이항분포이론을 그대로 적용할 경우 잘못된 분석 결과를 얻게 된다. 따라서, 최근에 이러한 가정이 타당하지 않은 경우에 대한 새로운 확률분포모형이 많이 개발되었다. 본 논문에서는 이중 한 일반화 이항분포모형을 소개하고, 그 모형에서 정의된 시행간 종속성 규정모수의 두 가지 추정량의 특성을 모의실험을 통하여 비교하여 본다.

주제어: 이원자료, 시행간 종속성 규정모수, 일반화 이항분포모형, 적률추정량, 최우추정량

## 1. 서 론

고전적인 이항분포모형(Conventional Binomial Distribution: CBD)은 많은 이산형 확률분포모형의 기초가 되는 모형으로서 통계이론에서 중요한 위치를 차지하고 있다. 그러나 많은 경우에 있어 (특히 의학, 생물통계자료에서) 실제 자료를 CBD 이론에 그대로 적용할 수 없는 것이 현실이다. CBD를 정의할 때 필요한 가정들이 만족되지 않기 때문이다. 따라서 타당하지 않은 가정들을 현실에 맞게 수정하여 만든 새로운 모형 개발의 필요성이 대두되었고, 최근에 이에 관한 활발한 연구가 진행되어 많은 연구 결과가 발표되었다(Altham(1978), Drezner & Farnum(1993), Kupper & Haseman(1978), Madsen(1993), Ng(1989), and Paul(1985, 1987)).

본 논문에서는 Drezner와 Farnum이 제시한 "일반화 이항분포모형(Generalized Binomial Distribution: GBD)"을 분석대상으로 한다. 그들은 이원자료에서 시행간 독립성이 결여된 자료에 대한 새로운 모형을 제시하고, 그 모형의 특성을 결정하는데 중요한 역할을 하는

[1]연세대학교 원주캠퍼스 문리대학 통계학과 부교수, 강원도 원주시 흥업면 매지리 234

시행간 종속성 규정모수의 추정량으로서 적률추정량을 제시하였다. 본 논문에서는 이의 대안으로서 시행간 종속성 규정모수의 최우추정량을 구해보고자 한다. 이는 수식적인 방법을 통해서는 불가능하고 numerical method를 이용해야 한다. Fortran 프로그램과 IMSL subroutine을 이용하여 GBD의 난수를 생성한 후, 생성된 자료를 이용하여 적률추정량과 최우추정량을 구해내고 이들의 평균평방오차를 비교하여 더 좋은 추정량을 결정한다.

본 논문은 5개의 절로 구성되어 있다. 2절에는 Drezner와 Farnum이 제안한 GBD와 그 모형에 포함되어 있는 종속성 규정모수의 적률추정량 도출법이 간략하게 소개되어 있다. 3절에는 종속성 규정모수의 최우추정량 도출과정이 제시되어 있고, 이들을 비교하기 위한 모의실험 과정과 결과, 그리고 이의 해석은 4절에 주어져 있다. 5절에는 간략한 결론이 제시되어 있다.

## 2. 일반화 이항분포모형

이산형 확률분포모형의 기초가 되는 고전적인 이항분포모형은 다음과 같은 기본적인 3개의 가정을 전제로 한다.

(가) 각각의 베르누이 시행 결과는 '성공' 또는 '실패', 2개만 가능하다.
(나) 각각의 베르누이 시행에서 '성공'(따라서 '실패') 사상의 확률은 동일하다.
(다) 각각의 베르누이 시행은 독립이다.

그러나 위에서 제시한 3개의 가정들이(특히 (나)와 (다)) 충족되지 않는 이원자료들을 주위에서 많이 찾아볼 수 있다. 예를 들면, 프로야구에서 한 시즌동안 어느 한 팀이 승리한 게임 수, 어느 지역에 심어진 나무들 중 일정 기간이 지난 후 살아 있는 나무의 수, 그리고 한 국회의원이 국회회기동안 회의에 참석한 일수등이 있다. 이러한 자료분석에 고전적인 이항분포이론이 그대로 적용될 경우 잘못된 분석 결과를 얻게 된다. 따라서 타당하지 않은 가정들을 현실에 맞게 수정한 새로운 모형 개발이 필요하게 되었고 최근에 이에 관한 많은 연구결과가 발표되었다. 본 논문에서는 상대적으로 다루기 용이한 "가정 (다)는 만족되지만 가정 (나)가 만족되지 않는 경우"를 다루기 보다는 더 복잡하면서도 어렵지만 현실에 보다 가까운 가정 (다)가 충족되지 않는 "각각의 베르누이 시행이 서로 종속적인 경우"의 모형을 다루고자 한다(실제로 가정 (다)가 만족되지 않으면 가정 (나)도 자동적으로 만족되지 않는 경우가 대부분이다). 이러한 경우를 다룬 모형이 지금까지 많이 개발되었는데 그 중에서도 가장 최근에 발표되었고 그전에 개발된 대부분의 모형을 특별한 예로 포함하고 있는 Drezner와 Farnum이 제시한 모형을 분석대상으로 한다. Drezner와 Farnum은 그들이 고안해낸 모형을 "일반화 이항분포모형(GBD)"이라 명명하였다. 본 절에서는 Drezner와 Farnum이 제안한 GBD와 그 모형의 정의에 포함되어 있는 종속성 규정모수의 적률추정량 도출법을 간략하게 소개한다.

Drezner와 Farnum은 고전적인 이항분포모형의 가정 (다)가 만족되지 않는 경우를 다루었다. 그들이 제시한 모형은 아래와 같다.

$$P(x,n) = P(S_n|x-1,n-1)P(x-1,n-1) + P(F_n|x,n-1)P(x,n-1). \quad (1)$$

위 모형에서 $P(x,n)$은 $n$번의 시행중 성공사상이 $x$번 나타날 확률을 나타내고, $S_n(F_n)$은 $n$번째 시행에서 성공(실패)이 얻어질 사상을 나타내는 기호이다. 각각 시행들간에 존재하는 종속성의 형태는 위 모형에서 $P(S_n|x-1,n-1)$과 $P(F_n|x,n-1)$이 어떻게 정의되느냐에 의해 결정되는데 그들이 정의한 두 값은 아래와 같다.

$$P(S_n|x-1,n-1) = (1-\theta)p + \theta\frac{x-1}{n-1}, \quad (2)$$
$$P(F_n|x,n-1) = (1-\theta)(1-p) + \theta(1-\frac{x}{n-1}).$$

위에서 $p$는 첫번째 시행의 성공확률, 그리고 $\theta$는 각각 베르누이 시행간의 종속 정도를 규정지어 주는 모수로서 Drezner와 Farnum 모형의 특성을 결정지어주는 중요한 역할을 한다. 따라서 $\theta$의 좋은 추정량을 찾아내는 것은 그들의 모형을 분석하는데 있어 필수적인 것이다. 아래에 주어진 <표 1>에 모수 $\theta$가 시행간 종속성을 어떻게 규정지어 주는지 예시되어 있다.

<표 1> $\theta$가 1~3번째 시행의 성공 확률에 미치는 영향($p$=0.90 인 경우)

| | $\theta$=0.10 일 때 | $\theta$=0.00 일 때 | $\theta$=-0.10 일 때 |
|---|---|---|---|
| $P(S_1)$ | 0.90 | 0.90 | 0.90 |
| $P(S_2\|0,1)$ | 0.81 | 0.90 | 0.99 |
| $P(S_2\|1,1)$ | 0.91 | 0.90 | 0.89 |
| $P(S_3\|0,2)$ | 0.81 | 0.90 | 0.99 |
| $P(S_3\|1,2)$ | 0.86 | 0.90 | 0.94 |
| $P(S_3\|2,2)$ | 0.91 | 0.90 | 0.89 |

<표 1>에서 볼 수 있는 바와 같이 $\theta$=0.0 일 때는 각 시행에 있어 성공 확률이 변하지 않지만(즉, 전번 시행들의 결과에 영향을 받지 않음), $\theta \neq 0.0$ 일 경우에는 각 시행에서의 성공 확률이 전번 시행들의 결과에 따라 변하게 된다. 따라서 본 논문에서는 $\theta$를 시행간 종속성 규정모수라고 정의하였다.

Drezner와 Farnum은 $\theta$의 추정방법으로서 적률법을 이용하였다. 확률변수 $X$가 식 (1)과 (2)에 주어진 일반화 이항분포를 따른다고 할 때, $X$의 평균과 분산은 아래와 같다.

$$E(X) = np, \quad Var(X) = p(1-p)\cdot\frac{n - \frac{\prod_{k=0}^{n-1}(k+2\theta)}{(n-1)!}}{(1-2\theta)}.$$

$X$의 평균이 $\theta$를 포함하고 있지 않으므로 $\theta$의 적률추정량을 구하기 위해서는 $X$의 분산을 표본분산과 등식으로 놓아 $\theta$를 추정하기 위한 방정식을 구하여야 한다. 그러나 이 방정식은 $\theta$의 $n$차 다항식으로서 $\theta$의 적률추정량을 closed form으로 구할 수 없어 그들은 점근적인 방법을 제시 하였다(Drezner & Farnum, p.3060-3061 참조).

## 3. $\theta$의 최우추정량

Drezner와 Farnum은 적률법에 의한 $\theta$의 추정량을 제시하였다. 본 절에서는 그의 대안으로 $\theta$의 최우추정량을 고려해 보고자 한다. 이의 도출은 수식적으로는 불가능한 것이고 program을 통한 numerical method에 의해서만 가능하다. 본 논문에서는 이를 다음과 같은 과정을 통하여 도출해 보고자 한다. 시행 수가 $n$인 일반화 이항분포 자료가 $m$개 있다고 가정한다.

(a) 우도함수 $P(x,n),(x=0,1,2,3,\cdots,n)$의 계산:

이는 식 (1), (2)에 주어져 있는 순환공식(recursive formula)을 이용하여 구한다. Fortran program을 이용하여 $P(\cdot,1) \sim P(\cdot,n-1)$을 순차적으로 구한 후, 이들 결과와 식 (1), (2)를 이용하면 구할 수 있다.

(b) $m$개 자료의 $x$값들을 (a)에서 구한 $P(x,n)$에 대입하여 $m$개 자료의 대수우도함수 $\sum_{i=1}^{m} lnP(x_i,n)$을 구한다.

(c) $\theta$가 취할 수 있는 값의 범위내에서 가능한 값을 변화시켜 가며 numerical method에 의해 대수우도함수를 극대화 시키는 $\theta$의 최우추정량을 구해낸다.

위의 도출과정 (c)에서 $\theta$가 취할 수 있는 값의 범위란 일반화 이항분포의 확률질량함수인 식 (1)이 $0 \le P(x,n) \le 1$ 을 만족하도록 하는 $\theta$의 범위를 의미한다. Drezner와 Farnum에 의하면 이는 아래와 같이 구하여진다.

$$\theta \ge 1 - \frac{1}{max(p, 1-p)}.$$

## 4. 모의실험

본 논문의 주된 목적은 Drezner와 Farnum의 일반화 이항분포모형에서 중요한 역할을 하는 종속성 규정모수 $\theta$에 대한 두가지 추정량 중에서 더 좋은 특성을 갖는 것을 찾고자 하는데 있다. 비교 대상인 두 가지 추정량은 Drezner와 Farnum이 제시한 적률추정량과 3절에서 제시된 numerical method를 이용한 최우추정량이다. 수식적인 비교는 불가능하므로 모의실험 결과를 이용하여 간접적인 비교를 해보고자 한다.

모의실험을 위해 가장 중요한 문제는 Drezner와 Farnum의 일반화 이항분포모형에 맞는 난수를 생성해내는 것이다. 본 논문에서는 아래에 나열한 과정을 통하여 이 모형에 해당하는 난수를 생성하고자 한다.

1단계: 주어진 첫번째 시행의 성공확률 $p$를 이용하여 IMSL의 Subroutine RNBIN으로부터 첫번째 베르누이 시행 결과를 생성한다.

2단계: 생성된 첫번째 베르누이 시행 결과와 주어진 $\theta$, 그리고 식 (2)를 이용하여 두번째 베르누이 시행의 성공확률 $P(S_2|0,1)$, 또는 $P(S_2|1,1)$를 계산한다.

3단계: 2단계에서 계산된 성공확률을 이용하여 다시 IMSL의 Subroutine RNBIN으로부터 두번째 베르누이 시행 결과를 생성한다.

4단계: 세번째 ~ $n$번째 베르누이 시행 결과를 얻기 위해 위의 2~3단계에 주어져 있는 과정을 반복한다.

5단계: 위의 4단계까지의 과정에서 얻어진 $n$개의 베르누이 시행 결과 중 "성공"에 해당하는 시행 횟수를 계산한다.

위의 1~5단계에 의하여 생성된 확률변수는 Drezner와 Farnum이 제안한 일반화 이항분포를 따르는 확률변수가 된다. 이렇게 생성된 난수를 이용하여 두 가지의 추정량을 비교하는데, 가능한 많은 경우의 모수를 포함시켜 객관적이고 일반적인 결과가 도출될 수 있도록 아래에 제시된 $p$와 $\theta$를 이용한다. 시행 수가 $n$인 일반화 이항분포자료 $m$개가 있다고 하자.

<표 2> Simulation Scheme

| n | m | p | $\theta$ |
|---|---|---|---|
| 50 | 25<br>50<br>100 | 0.1 부터 0.9 까지<br>0.1 씩 증가시키며 | 주어진 $p$ 에 대하여,<br>$\theta$=0.1, 0.2, 0.3, 0.4, 0.6, 0.8 |
| 200 | 100<br>200<br>400 | 0.1 부터 0.9 까지<br>0.1 씩 증가시키며 | 주어진 $p$ 에 대하여,<br>$\theta$=0.1, 0.2, 0.3, 0.4, 0.6, 0.8 |

위 <표 2>에 제시된 모든 $n, m, p, \theta$의 조합에 대하여 난수를 생성한 뒤 이를 이용하여 $\theta$의 두 가지 추정량을 구하였다. 위에서 언급한 바와 같이 $\theta$의 적률추정량은 Drezner와 Farnum이 제시한 점근적인 방법을 이용하여 구하였고, $\theta$의 최우추정량은 numerical method를 이용하여 구하였다. 두 추정량의 평균평방오차를 구한 후, 이를 이용하여 두 추정량의 상대효율(=적률추정량의 평균평방오차/최우추정량의 평균평방오차)을 계산하여 정리한 결과가 <표 3>와 <표 4>에 주어져 있다. <표 3>와 <표 4>에 주어져 있는 상대효율 결과로부터 아래와 같은 결론을 내릴 수 있다.

(1) 대부분의 경우에 있어 상대효율이 1보다 큰 값을 갖는다. 즉, 최우추정량의 특성이 적률추정량의 특성보다 더 좋다.

(2) $p, \theta, m$이 주어져 있을 때, 대부분의 경우에 있어($(p, \theta)$≒(0.5, 0.8), (0.6, 0.8)인 경우 제외) $n = 200$ 일 때 보다 $n = 50$ 일 때 상대효율이 더 큰 값을 갖게 된다. 즉, 다른 모수 값들이 동일하다면 시행횟수가 많을 때보다 상대적으로 시행횟수가 적을 때 최우추정량의 특성이 적률추정량의 특성보다 더욱 더 좋다.

(3) 시행횟수에 관계없이 상대효율은 $p, \theta$의 변화에 대해서 유사한 추이를 보인다. 즉, $n = 50$ 이든 $n = 200$ 이든 $n$값에 관계없이, 대부분의 경우에 있어 주어진 $p$에 대해서 $\theta$가 큰 값을 갖을수록 상대효율은 커지고, 반대로 주어진 $\theta$에 대하여는 $p$가 0 또는 1에 가까운 값을 갖을수록 상대효율은 커진다.

(4) $p, \theta, n$이 주어져 있을 때, $m$은 상대효율의 크기에 큰 영향을 미치지 않는다.

## 5. 결 론

이원자료를 분석하는데 있어 발생하는 큰 문제점은 시행간 독립성에 대한 가정이 충족되지 않는다는 것이다. 본 논문에서는 시행간 종속성을 용인하는 일반화 이항분포모형을 소개하고, 이 모형에서 중요한 역할을 하는 종속성 규정모수($= \theta$)의 적률추정량과 최우추정량을 모의실험을 통하여 비교하여 보았다. 대부분의 경우에 있어 최우추정량이 적률추정량보다 더 좋은 특성을 보여 주었고, 특히 $p$가 0 또는 1에 가까운 값을 갖고 동시에 $\theta$가 상대적으로 큰 값을 갖을수록 두 추정량간 특성의 차이가 커짐을 알 수 있었다.

## References

1. Altham, P. M. E. (1978). Two generalizations of the binomial distribution, *Applied Statistics*, 27, 162-167.

2. Drezner, Z. and Farnum, N.(1993). A generalized binomial distribution, *Communications in Statistics-Theory and Methods*, 22, 3051-3063.

3. IMSL User's Manual(1989). IMSL Inc., Houston, TX.

4. Kupper, L. L. and Haseman, J. K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments, *Biometrics*, 34, 69-76.

5. Madsen, R. (1993). Generalized binomial distributions, *Communications in Statistics-Theory and Methods*, 22, 3065-3086.

6. Ng, T. H. (1989). A new class of modified binomial distributions with applications to certain toxicological experiments, *Communications in Statistics-Theory and Methods*, 18, 3477-3492.

7. Paul, S. R. (1985). A three parameter generalization of the binomial distribution, *Communications in Statistics-Theory and Methods*, 14, 1497-1506.

8. Paul, S. R. (1987). On the beta-correlated binomial distribution-A three parameter generalization of the binomial distribution, *Communications in Statistics-Theory and Methods*, 16, 1473-1478.

<표 3> $n = 50$ 일 경우의 상대효율

| p | m | $\theta$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0.1 | 0.2 | 0.3 | 0.4 | 0.6 | 0.8 |
| 0.10 | 25 | 1.0969 | 1.2262 | 1.2213 | 1.4681 | 3.6913 | 10.8724 |
| | 50 | 1.2590 | 1.2436 | 1.3431 | 1.8473 | 3.3816 | 13.6073 |
| | 100 | 1.1491 | 1.3869 | 1.3078 | 1.9223 | 4.3214 | 7.5063 |
| 0.20 | 25 | 1.1008 | 1.1100 | 1.1533 | 1.3517 | 1.9629 | 7.5924 |
| | 50 | 0.9772 | 1.0633 | 1.1076 | 1.4198 | 1.9503 | 7.8938 |
| | 100 | 0.9769 | 1.1058 | 1.0677 | 1.4054 | 2.6987 | 4.5931 |
| 0.30 | 25 | 0.9946 | 1.0716 | 1.0618 | 1.0901 | 1.4624 | 3.3070 |
| | 50 | 1.0780 | 1.0097 | 1.0190 | 1.0777 | 1.5934 | 2.5056 |
| | 100 | 1.0369 | 1.0845 | 1.0779 | 1.1455 | 1.4400 | 2.8809 |
| 0.40 | 25 | 1.0933 | 1.1610 | 1.0685 | 1.0505 | 1.2120 | 1.6333 |
| | 50 | 1.0390 | 1.0295 | 1.0396 | 1.0584 | 1.1816 | 1.7240 |
| | 100 | 1.0064 | 1.0478 | 1.0085 | 1.0265 | 1.1609 | 1.4001 |
| 0.50 | 25 | 0.9646 | 1.0386 | 1.0654 | 1.1605 | 1.0461 | 1.1133 |
| | 50 | 1.0249 | 1.0351 | 1.0689 | 0.9712 | 1.0465 | 1.1421 |
| | 100 | 1.0312 | 1.0753 | 0.9656 | 1.0393 | 0.9795 | 1.0911 |
| 0.60 | 25 | 1.0443 | 1.0283 | 1.0608 | 0.9935 | 1.0817 | 1.4538 |
| | 50 | 1.0671 | 1.0475 | 1.0470 | 0.9997 | 1.2004 | 1.2969 |
| | 100 | 1.0405 | 1.0499 | 0.9637 | 1.0161 | 1.1617 | 1.4473 |
| 0.70 | 25 | 1.0652 | 1.0231 | 1.0808 | 1.3113 | 1.5958 | 3.2139 |
| | 50 | 1.0096 | 1.0254 | 1.0188 | 1.0389 | 1.3433 | 2.6753 |
| | 100 | 1.0167 | 1.0009 | 0.9790 | 1.1587 | 1.4573 | 2.9301 |
| 0.80 | 25 | 1.0182 | 1.1392 | 1.1332 | 1.3108 | 2.3161 | 8.3669 |
| | 50 | 1.0028 | 1.1856 | 1.0694 | 1.4726 | 2.4321 | 5.4670 |
| | 100 | 1.0523 | 1.0954 | 1.0094 | 1.2804 | 2.0537 | 4.0272 |
| 0.90 | 25 | 1.1920 | 1.1042 | 1.2703 | 1.6167 | 4.8270 | 17.6413 |
| | 50 | 1.0715 | 1.2515 | 1.2277 | 1.9625 | 3.0762 | 8.6820 |
| | 100 | 1.1396 | 1.4572 | 1.2625 | 1.7459 | 3.6682 | 9.2343 |

<표 4> $n = 200$ 일 경우의 상대효율

| p | m | $\theta$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0.1 | 0.2 | 0.3 | 0.4 | 0.6 | 0.8 |
| 0.10 | 100 | 1.0039 | 1.0313 | 1.2619 | 1.4556 | 3.3258 | 6.8288 |
| | 200 | 1.0735 | 1.0069 | 1.1358 | 1.4748 | 2.5108 | 6.3231 |
| | 400 | 1.0302 | 0.9982 | 1.1121 | 1.3330 | 2.7645 | 6.4300 |
| 0.20 | 100 | 1.0295 | 0.9702 | 1.0944 | 1.2293 | 1.6174 | 4.2050 |
| | 200 | 1.0287 | 0.9863 | 1.0467 | 1.2116 | 1.7449 | 4.7815 |
| | 400 | 1.0857 | 1.0073 | 1.1075 | 1.1871 | 1.7509 | 4.3995 |
| 0.30 | 100 | 0.9914 | 0.9387 | 1.0425 | 1.0609 | 1.3131 | 2.4977 |
| | 200 | 0.9908 | 0.9223 | 1.0201 | 1.0245 | 1.2964 | 3.1583 |
| | 400 | 0.9867 | 0.9859 | 1.0392 | 1.0291 | 1.3059 | 1.9956 |
| 0.40 | 100 | 1.0140 | 0.9419 | 1.0019 | 1.0975 | 1.0872 | 1.3979 |
| | 200 | 1.0162 | 0.9639 | 1.0258 | 1.0471 | 1.1209 | 1.5793 |
| | 400 | 0.9606 | 0.9428 | 1.0331 | 1.0237 | 1.1445 | 1.6806 |
| 0.50 | 100 | 1.0244 | 0.9469 | 1.0100 | 0.9900 | 1.0495 | 1.3334 |
| | 200 | 1.0078 | 0.9579 | 1.0020 | 0.9880 | 1.0452 | 1.3090 |
| | 400 | 0.9958 | 0.9520 | 1.0185 | 1.0023 | 1.0465 | 1.4101 |
| 0.60 | 100 | 1.0132 | 1.0269 | 0.9990 | 1.0407 | 1.0501 | 1.7494 |
| | 200 | 0.9833 | 0.9494 | 1.0038 | 1.0184 | 1.1091 | 1.3714 |
| | 400 | 1.0160 | 0.9557 | 1.0227 | 1.0223 | 1.0993 | 1.9941 |
| 0.70 | 100 | 1.0702 | 0.9787 | 1.0090 | 0.9832 | 1.2449 | 2.2726 |
| | 200 | 0.9907 | 0.9437 | 1.0503 | 1.0393 | 1.1961 | 2.2692 |
| | 400 | 0.9610 | 0.9652 | 1.0049 | 1.0395 | 1.3747 | 2.6567 |
| 0.80 | 100 | 1.0216 | 1.0026 | 1.0287 | 1.1669 | 1.6131 | 4.3155 |
| | 200 | 1.0078 | 0.9983 | 1.0801 | 1.1250 | 2.2661 | 3.4874 |
| | 400 | 0.9966 | 0.9623 | 1.0708 | 1.1301 | 1.7433 | 4.0931 |
| 0.90 | 100 | 1.0868 | 1.0815 | 1.1371 | 1.4452 | 2.6281 | 6.1653 |
| | 200 | 1.0539 | 0.9762 | 1.1638 | 1.2455 | 2.2101 | 6.3921 |
| | 400 | 0.9880 | 1.0457 | 1.1259 | 1.3478 | 2.3837 | 8.7487 |

# Comparison of Estimators of Dependence Related Parameter in Generalized Binomial Distribution

**Myung-Sang Moon** [2]

**Abstract**

In many cases where the conventional binomial distribution fails to apply to real world data, it is mainly due to the lack of independence among Bernoulli trials. Several authors have proposed models that are useful when independence assumption is not satisfied. In this paper, one proposed model is adapted, and estimators of dependence related parameter that is crucial in defining that model are considered. Simulation is performed to compare two estimators(method of moment estimator and maximum likelihood estimator) of dependence related parameter, and conclusions are made.

*Key Words and Phrases*: binary data, generalized binomial distribution(GBD), moment estimatior, maximum likelihood estiamtor.

[2]Associate Professor, Department of Statistics, Yonsei University, Wonju, Kwangwon-Do, Korea