



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

김 경 희 교수 지도  
석사학위 청구논문

이변량 영과잉 포아송 분포에서의  
추정량 비교

2019

성신여자대학교 대학원  
통 계 학 과  
허 윤 서

이변량 영과잉 포아송 분포에서의  
추정량 비교

김 경 희 교수 지도

이 논문을 석사학위논문으로 제출함

2018년 11월

성신여자대학교 대학원

통 계 학 과

허 윤 서

# 인 준 서

허윤서의 석사학위 논문으로 인준함.

2018년 11월

심사위원장.....(서명 또는 인)

심사위원.....(서명 또는 인)

심사위원.....(서명 또는 인)

성신여자대학교 대학원

## 논문개요

가산자료에 0의 값을 많이 가진다면 영과잉 가산 자료라 한다. 이러한 영과잉 가산자료를 분석하기 위해 포아송 분포를 변형한 형태인 영과잉 포아송 분포가 널리 쓰이고 있다. 만약 영과잉 가산 자료가 이변량인 경우 영과잉 포아송 분포의 확장인 이변량 영과잉 포아송 분포를 이용한다. 본 논문에서는 영과잉 가산 자료를 분석하기 위한 영과잉 포아송 분포 및 이변량 영과잉 포아송 분포를 알아보고 최우추정방법과 적률추정법을 통한 모수 추정량을 알아 보았다. 이후 모의실험 및 실증연구를 통해 최우추정량과 적률추정량의 편향(bias)과 평균제곱오차를 비교하여 더 좋은 추정량을 찾고자 하였다.

추정량 비교 결과, 모의실험 대부분의 경우 적률추정량의 편향보다 최우추정량의 편향이 작았고 적률추정량의 평균제곱오차보다 최우추정량의 평균제곱오차가 작았다. 마지막으로 한국 프로축구 리그인 K-리그 경기에서의 홈팀과 원정팀의 득점수 자료 분석을 통해 적률추정량보다 최우추정량이 실제 득점수 평균값과 근사하게 나타남을 알 수 있었다.

# 목 차

논문개요

제 1 장 서 론 .....	1
제 2 장 영과잉 포아송 분포 .....	3
2.1. 일변량 영과잉 포아송 분포 .....	3
2.1.1. 확률질량함수 .....	3
2.1.2. 적률생성함수 .....	3
2.1.3. 모수추정 .....	4
(1) 최우추정량 .....	4
(2) 적률추정량 .....	6
2.2. 이변량 영과잉 포아송 분포 .....	8
2.2.1. 확률질량함수 .....	8
2.2.2. 적률생성함수 .....	9
2.2.3. 모수추정 .....	10
(1) 최우추정량 .....	10
(2) 적률추정량 .....	11
제 3 장 추정량 비교 .....	13
3.1. 모의실험 .....	14
3.1.1. 일변량 영과잉 포아송분포 .....	14
3.1.2. 이변량 영과잉 포아송분포 .....	25
3.2. 실증분석 .....	33

제 4 장 결론 .....35

참고문헌

ABSTRACT

## 제 1장 서론

가산 자료 (Count data)는 불량품 개수, 영화 관람객 수 등과 같은 계수형 자료를 뜻한다. 가산 자료에 대해 이산형 분포를 적용한 연구는 다양한 분야에서 꾸준히 이루어져 왔다. 대표적으로 Jeong과 Choi (2014)는 포아송 회귀모형과 음이항 회귀모형을 활용하여 교통사고건수 자료를 분석하였고 Cho (2014)는 포아송 회귀모형과 음이항 회귀모형을 이용하여 대학생의 총 휴학 학기 수에 영향을 미치는 요인을 분석하였다.

가산 자료가 2변량이라면 이변량 가산자료라 하는데 주로 양 팀의 경기가 이루어지는 스포츠 분야에서 선행연구가 활발히 이루어져왔다. 예를 들어, Dixon과 Coles (1997)가 축구 게임 베팅 전략을 제시하기 위해 축구경기 자료에 대해 두 개의 팽창된 포아송 분포를 독립으로 가정하여 분석한 연구가 있다. Kalis와 Ntzoufras (2000)는 영국 프리미어리그 결과자료를 포아송 모형과 음이항 모형을 이용하여 타당성을 분석하였다. Lee (2014)는 한국 프로축구 리그인 K-리그의 골 득점을 예측하기 위해 이변량 포아송 모형을 이용하고 승점에 대한 수비력과 공격력의 영향력을 연구하였으며 Heo와 Kim (2016)은 한국 프로농구 경기에서 양 팀의 능력모수를 추정하기 위해 이변량 포아송 모형을 적용하였다 .

가산 자료에 0의 값이 과잉 관측 되면 이를 영 과잉 (Zero-Inflated) 가산 자료라 하며, 영 과잉 가산자료를 분석하기 위해 대표적으로 포아송 분포를 변형한 형태인 영과잉 포아송 분포가 널리 쓰이고 있다. 영과잉 가산자료에 영과잉 포아송 분포를 적용한 선행연구로는 Lambert (1992)가 영과잉 포아송 회귀모형을 적용하여 0의 값에 관한 확률  $p$ 와  $\lambda$ 의 공변량 효과를 분석한 연구가 있고 Kui 등 (2003)이 이변량 영과잉 포아송 모형을 이용하여 근골

격계 부상 건수 자료를 분석한 연구가 있다. Kim과 Lee (2008)는 포아송 회귀모형과 영과잉 포아송 회귀모형을 활용하여 연체건수 자료를 분석하였다. Cheon과 Park 등 (2008)은 텔레비전 시청자들의 채널전환(zapping) 자료를 분석하기 위해 영과잉 포아송 회귀모형 및 영과잉 음이항 회귀모형을 사용하였다. Choi와 Ko (2011)는 보험처리 건수에 0이 많은 것을 고려하여 실제 자료에 영과잉 모형을 적용하고, 영과잉 음이항 모형이 효율적임을 밝혔다. Lee와 Nam (2012)는 청소년 가출빈도 자료를 분석하기 위해 영과잉 포아송 회귀모형을 활용하였고 Cho (2015)는 포아송 회귀모형과 영과잉 포아송 회귀모형을 포함한 4개의 회귀모형을 이용하여 대학 재학 중 경험한 일자리 수 자료를 분석하였다.

영과잉 포아송 분포의 모수 추정량에는 대표적으로 최우추정량과 적률추정량이 있는데, 이에 관한 선행연구로는 Li 등 (1999)이 영과잉 포아송 분포를 다변량 영과잉 포아송 분포로 확장시키고 모수들의 최우추정량과 적률추정량을 비교한 연구가 있다. Kalis와 Ntzoufras (2003)는 Dixon과 Cloes (1997)가 제안한 모형을 확장시킨 대각확대 모형을 연구하였는데, 이들은 영과잉 포아송 모형을 포함한 대각확대 모형의 모수 추정을 위한 알고리즘을 R 프로그램의 'bivpois' 패키지에서 구현하였다 (Kalis와 Ntzoufras, 2005). Kim과 Um (2010)은 0의 빈도를 고려한 영빈도 추정량과 영과잉 포아송 분포에서의 적률추정량 및 최우추정량을 비교하였고, Beckett (2014)는 자연재해 자료에 대해 영과잉 포아송 분포를 적용하고 최우추정법과 적률 추정법을 통해 추정량을 구하였다.

본 논문에서는 이변량 영과잉 포아송 분포를 알아보고 모의실험과 실증분석을 통해 추정량을 비교하고자 한다. 2장에서는 일변량 영과잉 포아송 분포와 이변량 영과잉 포아송 분포의 확률질량함수 및 적률생성함수, 추정량을 설명하였다. 3장에서는 모의실험 및 실증분석을 통한 최우추정량과 적률추정량

을 비교 분석하였다. 이변량 영과잉 포아송 분포의 최우추정량을 구하기 위해 통계 분석 프로그램 R의 ‘bivpois’ 패키지를 이용하였다. 추정량의 비교 기준은 편향성(Bias)와 평균 제곱 오차(Mean Square Error; MSE)를 사용하였다. 실증분석을 위한 자료는 국내 축구리그인 K-리그 경기결과 자료를 이용하였다. 마지막으로 4장에서는 모의실험과 실제 자료 분석을 통한 결론을 제시하고 본 연구의 한계점 및 추후 연구에 대해 논의하였다.

## 제 2장 영과잉 포아송 분포

영과잉 포아송 분포 (Zero-Inflated Poisson Distribution; ZIP)는 영과잉 가산자료를 적용하기 위한 이산형 분포의 하나로서 포아송 분포를 변형한 형태이다. 일반적으로 일변량 영과잉 포아송 분포를 뜻하며, 영과잉 포아송 분포를 확장한 이변량 영과잉 포아송 분포, 다변량 영과잉 포아송 분포 등이 존재한다.

### 1.1. 일변량 영과잉 포아송 분포

확률변수  $Y$ 가 가산 자료(Count data)인 경우 이산형 확률 분포를 적용하며 대표적으로 포아송 분포가 있다. 이때 자료에 0이 과도하게 많은 경우를 영과잉(Zero-Inflated) 자료라 하며, 영과잉 자료를 적용하기 위해 포아송 분포를 변형한 것이 영과잉 포아송 분포이다. 영과잉 포아송 분포는  $ZIP(p, \lambda)$ 로 표기한다.

$$\begin{aligned} Y &\sim 0 \text{ with probability } p \\ &\sim Poi(\lambda) \text{ with probability } 1-p \end{aligned} \tag{2.1}$$

#### 1.1.1. 확률질량함수

$ZIP(p, \lambda)$ 를 따르는 확률변수  $Y$ 의 확률질량함수는 다음과 같다.

$$P(Y=y) = \begin{cases} p + (1-p)e^{-\lambda}, & y = 0 \\ (1-p)\frac{e^{-\lambda}\lambda^y}{y!}, & y = 1, 2, 3, \dots \end{cases}$$

#### 1.1.2. 적률생성함수

영과잉 포아송 분포의 적률생성함수를 구하면 다음과 같다.

$$\begin{aligned}
M_Y(t) &= E(e^{ty_i}) = \sum_{y_i=0}^{\infty} e^{ty_i} P(Y=y_i) = p + (1-p)e^{-\lambda} \sum_{y_i=0}^{\infty} e^{ty_i} \frac{\lambda^{y_i}}{y_i!} \\
&= p + (1-p)e^{\lambda(e^t-1)}, \quad t: \text{실수}.
\end{aligned}$$

위에서 구한 적률생성함수의 1차 적률과 2차 적률을 이용하여 평균과 분산을 구할 수 있다.

$$\begin{aligned}
\mu_Y &= E(Y) = M_Y'(t)|_{t=0} = (1-p)e^{\lambda(e^t-1)} \lambda e^t|_{t=0} = (1-p)\lambda, \\
\sigma_Y^2 &= M_Y''(t)|_{t=0} - [M_Y'(t)|_{t=0}]^2 \\
&= (1-p)e^{\lambda(e^t-1)} (\lambda e^t)^2|_{t=0} + (1-p)e^{\lambda(e^t-1)} \lambda e^t|_{t=0} - [(1-p)\lambda]^2 \\
&= (1-p)\lambda(1+p\lambda).
\end{aligned}$$

### 1.1.3. 모수추정

확률변수  $Y$ 가  $ZIP(p, \lambda)$ 를 따를 때 최우추정량(MLE), 적률추정량(MME)을 이용하여 모수  $p, \lambda$ 를 추정한다.

#### (1) 최우추정량

최우추정량이란 모수 추정 방법 중 하나로 로그우도함수를 최대화 하는 방법을 이용한 추정량이다. 최우추정량을 구하기 위한 우도함수와 로그우도함수는 다음과 같다.

$$\begin{aligned}
l(p, \lambda; y_1, y_2, \dots, y_n) &= \prod_{i=1}^n [(p + (1-p)e^{-\lambda})I(y_i = 0) + (1-p) \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} I(y_i > 0)], \\
\log l(p, \lambda; y_1, y_2, \dots, y_n) &= \sum_{i=1}^n \log [(p + (1-p)e^{-\lambda})I(y_i = 0) + (1-p) \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} I(y_i > 0)].
\end{aligned}$$

위의 로그우도함수를  $\lambda$ 와  $p$ 에 대해 편미분 한 결과는 다음과 같다. 이때  $y=0$ 인 경우의 수를  $n_0$ 로 정의한다.

$$\begin{aligned}
\frac{\partial \log l}{\partial \lambda} &= \sum_{i=1}^n \frac{-(1-p)e^{-\lambda}I(y_i=0) + \frac{(1-p)}{y_i!}(y_i\lambda^{y_i-1} - \lambda^{y_i})e^{-\lambda}I(y_i>0)}{(p+(1-p)e^{-\lambda})I(y_i=0) + (1-p)\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}I(y_i>0)}, \\
&= \sum_{i=1}^n \left[ I(y_i=0) \frac{-(1-p)e^{-\lambda}}{p+(1-p)e^{-\lambda}} + I(y_i>0) \frac{\frac{(1-p)}{y_i!}(y_i\lambda^{y_i-1} - \lambda^{y_i})e^{-\lambda}}{(1-p)\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}} \right] \\
&= \sum_{i=1}^n \left[ I(y_i=0) \frac{-(1-p)e^{-\lambda}}{p+(1-p)e^{-\lambda}} + I(y_i>0) \left( \frac{y_i}{\lambda} - 1 \right) \right] \\
&= n_0 \frac{-(1-p)e^{-\lambda}}{p+(1-p)e^{-\lambda}} + \frac{\sum_{i=1}^n}{\lambda} - (n-n_0) \dots\dots\dots (A)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \log l}{\partial p} &= \sum_{i=1}^n \frac{(1-e^{-\lambda})I(y_i=0) - \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}I(y_i>0)}{(p+(1-p)e^{-\lambda})I(y_i=0) + (1-p)\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}I(y_i>0)} \\
&= \sum_{i=1}^n \left[ I(y_i=0) \frac{1-e^{-\lambda}}{(p+(1-p)e^{-\lambda})} - I(y_i>0) \frac{\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}}{(1-p)\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}} \right] \\
&= \sum_{i=1}^n \left[ I(y_i=0) \frac{1-e^{-\lambda}}{p+(1-p)e^{-\lambda}} - I(y_i>0) \frac{1}{1-p} \right] \\
&= n_0 \frac{1-e^{-\lambda}}{p+(1-p)e^{-\lambda}} - \frac{n-n_0}{1-p} \dots\dots\dots (B)
\end{aligned}$$

(A), (B)를 각각 0이라 하고 연립방정식으로  $\lambda$ 와  $p$ 에 대한 추정량을 구할 수 있다.

$$1 - \frac{n_0}{n} = \frac{(1 - e^{-\widehat{\lambda}_{ml}})\bar{y}}{\widehat{\lambda}_{ml}},$$

$$\widehat{p}_{ml} = 1 - \frac{\bar{y}}{\widehat{\lambda}_{ml}}.$$

위에서 구한  $\hat{\lambda}, \hat{p}$ 이 로그우도함수를 최대로 하는지 확인하기 위해 헤시안 행렬(Hessian Matrix)를 구하면 다음과 같다.

$$H = \begin{bmatrix} \frac{\partial^2 \log l}{\partial \lambda^2} & \frac{\partial \log l}{\partial \lambda \partial p} \\ \frac{\partial \log l}{\partial p \partial \lambda} & \frac{\partial^2 \log l}{\partial p^2} \end{bmatrix} = \begin{bmatrix} n_0 \frac{p(1-p)e^{-\lambda}}{(p+(1-p)e^{-\lambda})^2} - \frac{\sum_{i=1}^n y_i}{\lambda^2} & \frac{n_0 e^{-\lambda}}{(p+(1-p)e^{-\lambda})^2} \\ \frac{n_0 e^{-\lambda}}{(p+(1-p)e^{-\lambda})^2} & \frac{-(1-e^{-\lambda})^2}{(p+(1-p)e^{-\lambda})^2} - \frac{(n-n_0)}{(1-p)^2} \end{bmatrix} \Bigg|_{\lambda=\hat{\lambda}, p=\hat{p}}$$

$$= \begin{bmatrix} n(1-\hat{p}) \left\{ \frac{\hat{p} \left(1 - \frac{n}{n_0} \hat{p}\right)}{1-\hat{p}} - \frac{1}{\hat{\lambda}} \right\} & n \left(1 - \frac{n}{n_0} \hat{p}\right) \\ \frac{n \left(1 - \frac{n}{n_0} \hat{p}\right)}{1-\hat{p}} & \frac{n \left(1 - \frac{n}{n_0}\right)}{(1-\hat{p})^2} \end{bmatrix}$$

여기에서,  $\hat{\lambda}, \hat{p}$ 에 대해 우도함수를 최대화하기 위한 두 조건

①

$$\frac{\partial^2 \log l}{\partial \lambda^2} \Bigg|_{\lambda=\hat{\lambda}, p=\hat{p}} = \frac{\partial^2 \log l}{\partial \lambda^2} \Bigg|_{\lambda=\hat{\lambda}, p=\hat{p}} - \left( \frac{\partial \log l}{\partial \lambda \partial p} \right)^2 \Bigg|_{\lambda=\hat{\lambda}, p=\hat{p}}$$

$$= \frac{n^2}{1-\hat{p}} \left[ \frac{n}{n_0} \hat{p} - 1 + \frac{\left(\frac{n}{n_0} - 1\right)}{\hat{\lambda}} \right] > \frac{n^2}{1-\hat{p}} \left[ \frac{n}{n_0} \hat{p} - 1 + \left(\frac{n}{n_0} - 1\right) \left( \frac{\frac{n_0-p}{n}}{1-\frac{n_0}{n}} \right) \right] = 0 \quad \left( \because \hat{\lambda} < \frac{1-\frac{n_0}{n}}{\frac{n_0}{n}-p} \right)$$

$$\textcircled{2} \quad \frac{\partial^2 \log l}{\partial \lambda^2} \Bigg|_{\lambda=\hat{\lambda}, p=\hat{p}}$$

$$= n \left\{ \hat{p} \left(1 - \frac{n}{n_0} \hat{p}\right) - (1-\hat{p}) \frac{1}{\hat{\lambda}} \right\} < n \left\{ \hat{p} \left(1 - \frac{n}{n_0} \hat{p}\right) - (1-\hat{p}) \frac{\hat{p} - \frac{n_0}{n}}{1 - \frac{n_0}{n}} \right\} = 0$$

을 만족하므로  $\hat{\lambda}, \hat{p}$ 은 최대우도추정량이라 할 수 있다.

(2) 적률추정량

적률추정량이란 모수를 추정할 때 모집단에서의 적률과 표본들의 적률을 같게 놓는 방법을 이용하는 추정량이다.

$$\text{모집단의 1차 적률: } E(Y) = (1-p)\lambda, \quad \text{표본의 1차 적률: } \frac{\sum_{i=1}^n y_i}{n},$$

$$\text{모집단의 2차 적률: } E(Y^2) = (1-p)(\lambda + \lambda^2), \quad \text{표본의 2차 적률: } \frac{\sum_{i=1}^n y_i^2}{n}.$$

모집단의 적률과 표본의 적률을 같다고 하고 연립방정식으로 풀면 모수  $\lambda$ 를 추정할 수 있다.

$$\text{1차 적률: } E(Y) = (1-p)\lambda \approx \frac{\sum_{i=1}^n y_i}{n}$$

$$\text{2차 적률: } E(Y^2) = (1-p)(\lambda + \lambda^2) \approx \frac{\sum_{i=1}^n y_i^2}{n},$$

$$\frac{\lambda}{\lambda + \lambda^2} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n y_i^2} \Leftrightarrow \widehat{\lambda_{mm}} = \frac{\sum_{i=1}^n y_i^2 - \sum_{i=1}^n y_i}{\sum_{i=1}^n y_i},$$

위에서 구한  $\lambda$ 의 적률 추정량을 1차 적률에 대입하여 모수  $p$ 의 적률 추정량을 구한다.

$$(1-p)\widehat{\lambda_{mm}} = \frac{\sum_{i=1}^n y_i}{n} \Leftrightarrow \widehat{p_{mm}} = 1 - \frac{\bar{y}}{\widehat{\lambda_{mm}}}.$$

## 1.2. 이변량 영과잉 포아송 분포

이변량 포아송(Bivariate Poisson; BP) 분포는  $\lambda_{10}, \lambda_{20}, \lambda_{00}$ 를 모수로 가지

며  $BP(\lambda_{10}, \lambda_{20}, \lambda_{00})$  로 표기한다. Marshall과 Olkin (1985)에 의하면  $(X_1, X_2)$  가  $BP(\lambda_{10}, \lambda_{20}, \lambda_{00})$  를 따를 때  $X_1 = U_1 + Z$ ,  $X_2 = U_2 + Z$  라고 할 수 있고  $U_1, U_2, Z$  는 각각  $\lambda_{10}, \lambda_{20}, \lambda_{00}$  을 평균으로 하는 포아송 분포를 따른다. 이변량 영과잉 포아송 분포는 일변량 영과잉 포아송 분포의 확장으로 이변량 포아송 분포를 변형한 형태이다. 확률변수  $Y_1, Y_2$ 가 영이 과도하게 많은 영과잉 자료라면 이변량 영과잉 포아송 분포를 따르고  $BZIP(p_0, p_1, p_2, \lambda_{10}, \lambda_{20}, \lambda_{00})$  로 표기한다.

$$\begin{aligned}
(Y_1, Y_2) &\sim (0,0) \text{ with probability } p_0 \\
&\sim (Poi(\lambda_1), 0) \text{ with probability } p_1 \\
&\sim (0, Poi(\lambda_2)) \text{ with probability } p_2 \\
&\sim BP(\lambda_{10}, \lambda_{20}, \lambda_{00}) \text{ with probability } p_{12}
\end{aligned} \tag{2.1}$$

여기에서  $p_0 + p_1 + p_2 + p_{12} = 1$  이고  $\lambda_1 = \lambda_{10} + \lambda_{00}$ ,  $\lambda_2 = \lambda_{20} + \lambda_{00}$ ,  $\lambda = \lambda_{10} + \lambda_{20} + \lambda_{00}$  이다.

### 1.2.1. 확률질량함수

$BZIP(p_0, p_1, p_2, \lambda_{10}, \lambda_{20}, \lambda_{00})$ 의 확률질량함수는 다음과 같이 쓴다.

$$P(Y_1 = 0, Y_2 = 0) = p_0 + p_1 \exp(-\lambda_1) + p_2 \exp(-\lambda_2) + p_{12} \exp(-\lambda),$$

$$P(Y_1 = y_1, Y_2 = 0) = \frac{p_1 \lambda_1^{y_1} \exp(-\lambda_1) + p_{12} \lambda_{10}^{y_1} \exp(-\lambda)}{y_1!},$$

$$P(Y_1 = 0, Y_2 = y_2) = \frac{p_2 \lambda_2^{y_2} \exp(-\lambda_2) + p_{12} \lambda_{20}^{y_2} \exp(-\lambda)}{y_2!},$$

$$P(Y_1 = y_1, Y_2 = y_2) = p_{12} \sum_{j=0}^{\min(y_1, y_2)} \frac{\lambda_{10}^{y_1-j} \lambda_{20}^{y_2-j} \lambda_{00}^j}{(y_1-j)! (y_2-j)! j!} \times \exp(-\lambda),$$

$$y_1, y_2 = 1, 2, \dots$$

### 1.2.2. 적률생성함수

이변량 확률변수  $(Y_1, Y_2)$  의 적률생성함수를 구하면 다음과 같다.

$$\begin{aligned}
 M_{Y_1, Y_2}(t_1, t_2) &= E[\exp(t_1 Y_1 + t_2 Y_2)] = p_0 + p_1 E[\exp(t_1 X_1)] + p_2 E[\exp(t_2 X_2)] \\
 &\quad + p_{12} E[\exp(t_1 U_1 + t_2 U_2 + Z(t_1 + t_2))] \\
 &= p_0 + p_1 \exp[\lambda_1 (e^{t_1} - 1)] + p_2 \exp[\lambda_2 (e^{t_2} - 1)] \\
 &\quad + p_{12} \exp[\lambda_{10} (e^{t_1} - 1) + \lambda_{20} (e^{t_2} - 1) + \lambda_{00} (e^{t_1 + t_2} - 1)].
 \end{aligned}$$

적률생성함수를 이용하여 평균과 분산, 공분산을 구할 수 있다.

$$\begin{aligned}
 \mu_{Y_1} = E(Y_1) &= \left. \frac{\partial M_{Y_1, Y_2}(t_1, t_2)}{\partial t_1} \right|_{t_1, t_2 = 0} \\
 &= p_1 \exp\{\lambda_1 (e^{t_1} - 1)\} \lambda_1 e^{t_1} \Big|_{t_1, t_2 = 0} \\
 &\quad + p_{12} \exp\{\lambda_{10} (e^{t_1} - 1) + \lambda_{20} (e^{t_2} - 1) + \lambda_{00} (e^{t_1 + t_2} - 1)\} (\lambda_{10} e^{t_1} + \lambda_{00} e^{t_1 + t_2}) \Big|_{t_1, t_2 = 0} \\
 &= (p_1 + p_{12}) \lambda_1,
 \end{aligned}$$

$$\begin{aligned}
 \mu_{Y_2} = E(Y_2) &= \left. \frac{\partial M_{Y_1, Y_2}(t_1, t_2)}{\partial t_2} \right|_{t_1, t_2 = 0} \\
 &= p_2 \exp\{\lambda_2 (e^{t_2} - 1)\} \lambda_2 e^{t_2} \Big|_{t_1, t_2 = 0} \\
 &\quad + p_{12} \exp\{\lambda_{10} (e^{t_1} - 1) + \lambda_{20} (e^{t_2} - 1) + \lambda_{00} (e^{t_1 + t_2} - 1)\} (\lambda_{20} e^{t_2} + \lambda_{00} e^{t_1 + t_2}) \Big|_{t_1, t_2 = 0} \\
 &= (p_2 + p_{12}) \lambda_2,
 \end{aligned}$$

$$\begin{aligned}
 \sigma_{Y_1}^2 &= \left. \frac{\partial^2 M_{Y_1, Y_2}(t_1, t_2)}{\partial t_1^2} \right|_{t_1, t_2 = 0} - \left[ \left. \frac{\partial M_{Y_1, Y_2}(t_1, t_2)}{\partial t_1} \right] \right|_{t_1, t_2 = 0}^2 \\
 &= p_1 \exp\{\lambda_1 (e^{t_1} - 1)\} (\lambda_1^2 e^{2t_1} + \lambda_1 e^{t_1}) \Big|_{t_1, t_2 = 0} \\
 &\quad + p_{12} \exp\{\lambda_{10} (e^{t_1} - 1) + \lambda_{20} (e^{t_2} - 1) + \lambda_{00} (e^{t_1 + t_2} - 1)\} \\
 &\quad \times (\lambda_{10} + \lambda_{00} e^{t_2}) e^{t_1} [(\lambda_{10} + \lambda_{00} e^{t_2}) e^{t_1} + 1] \Big|_{t_1, t_2 = 0} - [(p_1 + p_{12}) \lambda_1]^2 \\
 &= (p_1 + p_{12}) \lambda_1 \{ \lambda_1 + 1 - (p_1 + p_{12}) \lambda_1 \}
 \end{aligned}$$

$$\begin{aligned}
\sigma_{Y_2}^2 &= \frac{\partial^2 M_{Y_1, Y_2}(t_1, t_2)}{\partial t_1^2} \Big|_{t_1, t_2=0} - \left[ \frac{\partial M_{Y_1, Y_2}(t_1, t_2)}{\partial t_2} \right]^2 \Big|_{t_1, t_2=0} \\
&= p_2 \exp\{\lambda_2(e^{t_2} - 1)\} (\lambda_2^2 e^{2t_2} + \lambda_2 e^{t_2}) \Big|_{t_1, t_2=0} \\
&\quad + p_{12} \exp\{\lambda_{10}(e^{t_1} - 1) + \lambda_{20}(e^{t_2} - 1) + \lambda_{00}(e^{t_1+t_2} - 1)\} \\
&\quad \times (\lambda_{20} + \lambda_{00} e^{t_1}) e^{t_1} [(\lambda_{20} + \lambda_{00} e^{t_1}) e^{t_2} + 1] \Big|_{t_1, t_2=0} - [(p_2 + p_{12}) \lambda_2]^2 \\
&= (p_2 + p_{12}) \lambda_2 \{\lambda_2 + 1 - (p_2 + p_{12}) \lambda_2\} \\
E(Y_1 Y_2) &= \frac{\partial M_{Y_1, Y_2}(t_1, t_2)}{\partial t_1 \partial t_2} \Big|_{t_1, t_2=0} \\
&= p_{12} \exp\{\lambda_{10}(e^{t_1} - 1) + \lambda_{20}(e^{t_2} - 1) + \lambda_{00}(e^{t_1+t_2} - 1)\} \Big|_{t_1, t_2=0} \\
&\quad \times \{(\lambda_{10} e^{t_1} + \lambda_{00} e^{t_1+t_2})(\lambda_{20} e^{t_2} + \lambda_{00} e^{t_1+t_2}) + \lambda_{00} e^{t_1+t_2}\} \Big|_{t_1, t_2=0} \\
&\quad - (p_1 + p_{12}) \lambda_1 (p_2 + p_{12}) \lambda \\
&= (\lambda_1 \lambda_2 + \lambda_{00}) p_{12} \\
\sigma_{Y_1, Y_2} &= E(Y_1 Y_2) - E(Y_1) E(Y_2) \\
&= (\lambda_1 \lambda_2 + \lambda_{00}) p_{12} - (p_1 + p_{12})(p_2 + p_{12}) \lambda_1 \lambda_2.
\end{aligned}$$

위와 같은 방법으로  $E(Y_1^2 Y_2 + Y_1 Y_2^2)$ 를 구하면 다음과 같다.

$$\begin{aligned}
E(Y_1^2 Y_2) &= p_{12} \lambda_1 \{\lambda_1 \lambda_2 + \lambda_{00}\} + p_{12} \{\lambda_1 \lambda_2 + \lambda_1 \lambda_{00} + \lambda_{00}\} \\
E(Y_1 Y_2^2) &= p_{12} \lambda_2 \{\lambda_1 \lambda_2 + \lambda_{00}\} + p_{12} \{\lambda_1 \lambda_2 + \lambda_2 \lambda_{00} + \lambda_{00}\} \\
E(Y_1^2 Y_2 + Y_1 Y_2^2) &= p_{12} [\lambda_1 \lambda_2 (\lambda_1 + \lambda_2 + 2) + 2 \lambda_{00} (\lambda_1 + \lambda_2 + 1)]
\end{aligned}$$

### 1.2.3. 모수추정

본 논문에서는 Karlis와 Ntzoufras가 제안한  $BZIP(p_0, \lambda_{10}, \lambda_{20}, \lambda_{00})$  모형을 적용하여 모수를 추정하기 위해  $p_1 = p_2 = 0$  으로 가정하였다.

#### (1) 최우추정량

$$f_{BP} = P(Y_1 = y_1, Y_2 = y_2) = \sum_{j=0}^{\min(y_1, y_2)} \frac{\lambda_{10}^{y_1-j} \lambda_{20}^{y_2-j} \lambda_{00}^j}{(y_1-j)!(y_2-j)!j!} \times \exp(-\lambda),$$

$$y_1, y_2 = 0, 1, 2, \dots$$

*MLE of*  $(p_0, \lambda_{10}, \lambda_{20}, \lambda_{00})$

$$\begin{aligned} &= \arg \max l(p_0, \lambda_{10}, \lambda_{20}, \lambda_{00}) \\ &= \arg \max \log \prod_{i=1}^n f_{BZIP}(y_{1i}, y_{2i}; p_0, \lambda_{10}, \lambda_{20}, \lambda_{00}) \\ &= \arg \max \log \prod_{i=1}^n [p_0 + (1-p_0)\{f_{BP}(0, 0; \lambda_{10}, \lambda_{20}, \lambda_{00})\}^{1-a_i} \{f_{BP}(y_{1i}, y_{2i}; \lambda_{10}, \lambda_{20}, \lambda_{00})\}^{a_i}] \end{aligned}$$

여기에서  $(y_{1i}, y_{2i}) = (0, 0)$ 이면  $a_i = 0$ 이고 이외의 경우엔  $a_i = 1$ 이다.

(2) 적률추정량

$$\begin{aligned} \text{모집단의 1차 적률 : } E(Y_1) &= (p_1 + p_{12})\lambda_1, & \text{표본의 1차 적률 : } \frac{\sum_{i=1}^n y_{1i}}{n} &= \overline{Y}_1, \\ \text{모집단의 1차 적률 : } E(Y_2) &= (p_2 + p_{12})\lambda_2, & \text{표본의 1차 적률 : } \frac{\sum_{i=1}^n y_{2i}}{n} &= \overline{Y}_2, \\ \text{모집단의 2차 적률 : } E(Y_1^2) &= (p_1 + p_{12})(\lambda_1 + \lambda_1^2), & \text{표본의 2차 적률 : } \frac{\sum_{i=1}^n y_{1i}^2}{n} &= \overline{Y}_1^2, \\ \text{모집단의 2차 적률 : } E(Y_2^2) &= (p_2 + p_{12})(\lambda_2 + \lambda_2^2), & \text{표본의 2차 적률 : } \frac{\sum_{i=1}^n y_{2i}^2}{n} &= \overline{Y}_2^2. \end{aligned}$$

모집단의 적률과 표본의 적률을 같다고 하면 다음과 같이 쓸 수 있다.

$$1차 적률 : E(Y_1) = (p_1 + p_{12})\lambda_1 \approx \frac{\sum_{i=1}^n y_{1i}}{n} = \overline{Y_1},$$

$$E(Y_2) = (p_2 + p_{12})\lambda_2 \approx \frac{\sum_{i=1}^n y_{2i}}{n} = \overline{Y_2}$$

$$2차 적률 : E(Y_1^2) = (p_1 + p_{12})(\lambda_1 + \lambda_1^2) \approx \frac{\sum_{i=1}^n y_{1i}^2}{n} = \overline{Y_1^2},$$

$$E(Y_2^2) = (p_2 + p_{12})(\lambda_2 + \lambda_2^2) \approx \frac{\sum_{i=1}^n y_{2i}^2}{n} = \overline{Y_2^2}.$$

위의 식을 이용하여 연립방정식으로 모수  $\lambda_i (i=1,2)$ 를 추정한다.

$$\tilde{\lambda}_i = \frac{\overline{Y_i^2}}{\overline{Y_i}} - 1, \quad i = 1, 2.$$

$\lambda_{00}$ 의 적률추정량은  $E(Y_1 Y_2)$  과  $E(Y_1^2 Y_2)$ 을 이용한다.

$$E(Y_1 Y_2) = (\lambda_1 \lambda_2 + \lambda_{00})p_{12} \approx \frac{\sum_{i=1}^n y_{1i} y_{2i}}{n} = \overline{Y_1 Y_2},$$

$$E(Y_1^2 Y_2) = p_{12} \lambda_1 \{\lambda_1 \lambda_2 + \lambda_{00}\} + p_{12} \{\lambda_1 \lambda_2 + \lambda_1 \lambda_{00} + \lambda_{00}\} \approx \frac{\sum_{i=1}^n y_{1i}^2 y_{2i}}{n} = \overline{Y_1^2 Y_2},$$

위의 식을 이용하여 연립방정식으로 구한  $\lambda_{00}$ 의 적률추정량은 다음과 같다.

$$\tilde{\lambda}_{00} = \frac{\tilde{\lambda}_1 \tilde{\lambda}_2 \{ \overline{Y_1^2 Y_2} - \overline{Y_1 Y_2} (\tilde{\lambda}_1 + 1) \}}{\overline{Y_1 Y_2} (2\tilde{\lambda}_1 + 1) - \overline{Y_1^2 Y_2}},$$

이제 앞에서 구한  $\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_{00}$ 을 이용하여  $\tilde{p}_1, \tilde{p}_2, \tilde{p}_{12}, \tilde{p}_0$ 을 구한다.

$$\tilde{p}_i = \frac{\overline{Y_i}}{\tilde{\lambda}_i} - \tilde{p}_{12} \quad (i = 1, 2),$$

$$\tilde{p}_{12} = \frac{\overline{Y_1 Y_2}}{\tilde{\lambda}_1 \tilde{\lambda}_2 + \tilde{\lambda}_{00}},$$

$$\tilde{p}_0 = 1 - \tilde{p}_1 - \tilde{p}_2 - \tilde{p}_{12}.$$

이때  $p_1 = p_2 = 0$ 을 가정하였으므로  $\tilde{p}_1 = \tilde{p}_2 = 0$ 이고  $\tilde{p}_0 = 1 - \tilde{p}_{12}$  이다.

## 제 3장 추정량 비교

본 절에서는 이변량 영과잉 포아송 분포의 모수를 추정하기 위하여 2.2.3. 절의 최우추정방법과 적률추정방법을 이용한 추정량을 비교하고자 모의실험과 실증분석을 실시하였다.

### 3.1. 모의실험

#### 3.1.1. 일변량 영과잉 포아송 분포

(1.1)에서의 최우추정량과 적률추정량 중 더 좋은 추정량을 알아보기 위해  $n = 50, 100, 500$ 인 표본으로 하여 모의실험을  $m (= 1000)$ 번 반복 하였으며  $\lambda$ 는 1부터 5까지 1씩 증가시키고  $p$ 는 0.2부터 0.8까지 0.2씩 증가시켰다. 그림 1~그림12를 통해 모의실험으로 얻어진 추정량들의 분포를 살펴보았을 때 그림1에서  $n$ 과  $p$ 가 작고  $\lambda$ 가 작은 경우에 음수로 나타나는 경우를 발견할 수 있다. 이 경우를 제외한 대부분의 경우에 정규분포 형태를 보이는 것을 알 수 있고 대체적으로  $n$ 이 커질수록 최우추정량의 산포가 적률추정량의 산포보다 작아지는 것을 확인할 수 있다.

표1~표4를 통해 음수로 추정된 경우의 수를 자세히 살펴보면,  $m$ 번의 반복 시행 중 특히  $p$ 가 클 때에 비해  $p$ 가 작을 때 최우추정량과 적률추정량이 음수로 추정되는 경우가 더 많이 발생하였다. 수렴을 하지 않는 문제 등을 포함하여  $-\text{Inf}$  로 나타나는 경우는 주로  $n$ 과  $\lambda$ 가 작으며  $p$ 가 큰 경우에 발생하였다.

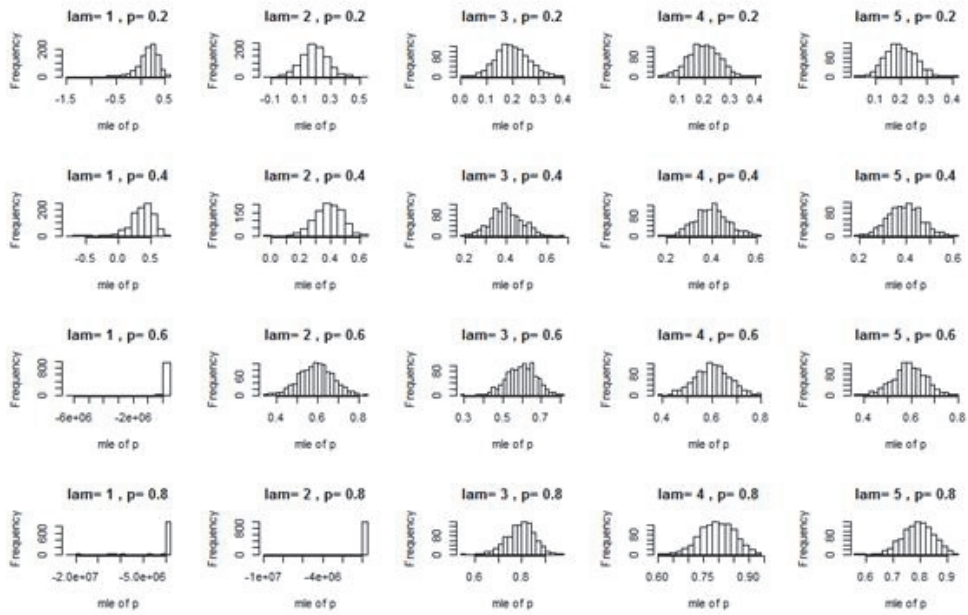


그림 1. 모의실험 결과를 이용한 히스토그램 :  $\hat{p}_{ml}$  ( $n=50$ )

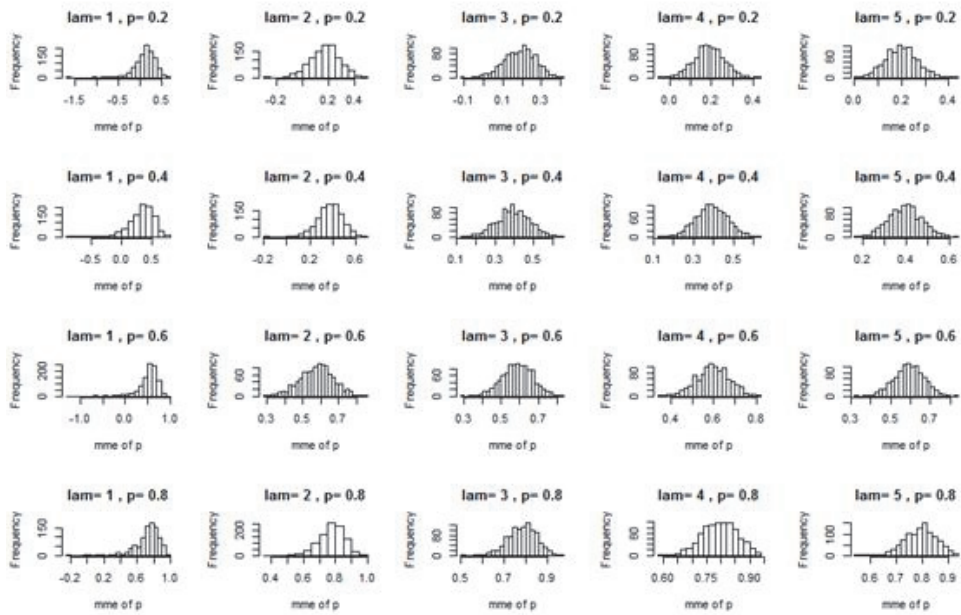


그림 2. 모의실험 결과를 이용한 히스토그램 :  $\hat{p}_{mm}$  ( $n=50$ )

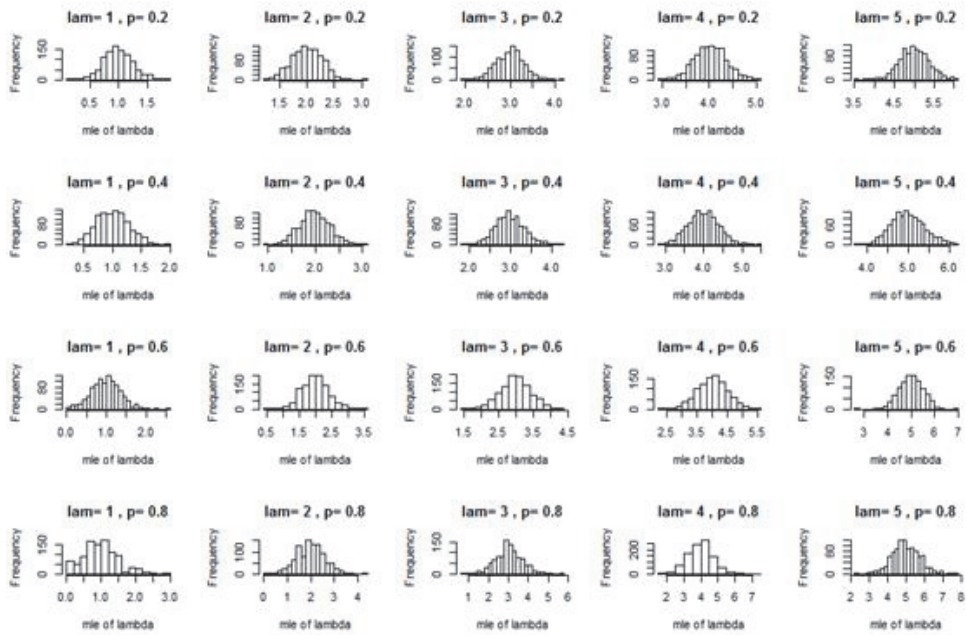


그림 3. 모의실험 결과를 이용한 히스토그램 :  $\hat{\lambda}_{ml}$  ( $n=50$ )

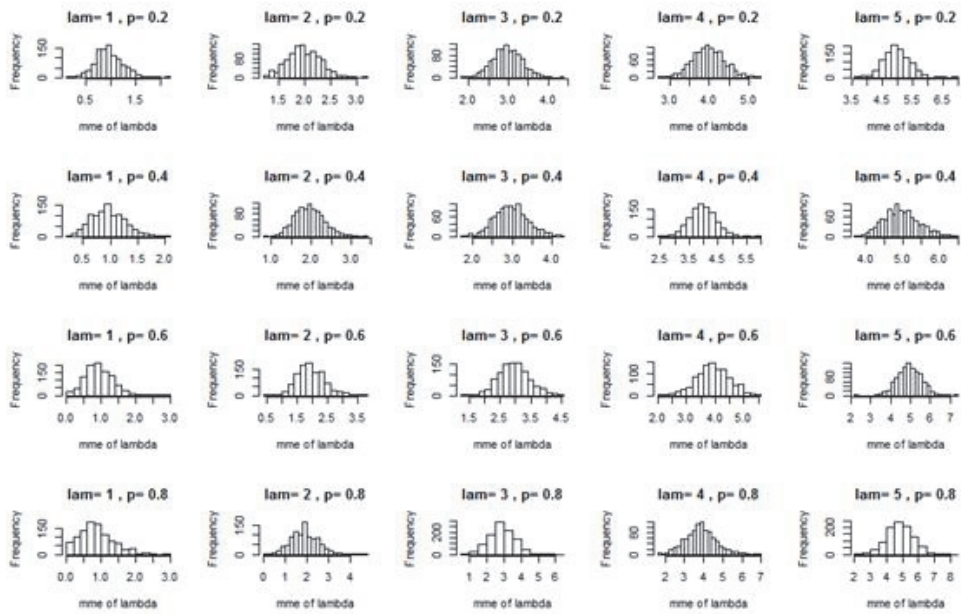


그림 4. 모의실험 결과를 이용한 히스토그램 :  $\hat{\lambda}_{mm}$  ( $n=50$ )

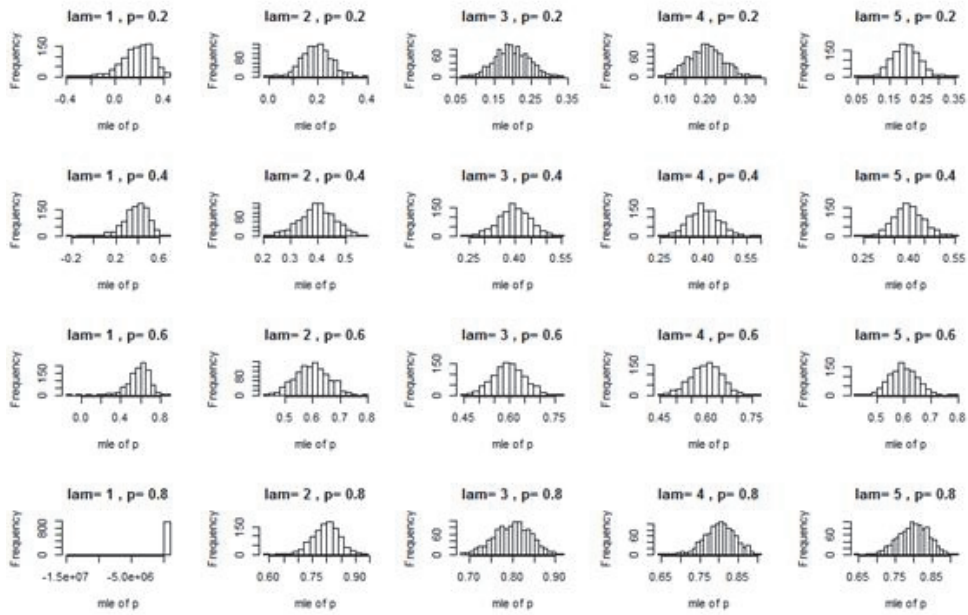


그림 5. 모의실험 결과를 이용한 히스토그램 :  $\hat{p}_{ml}$  ( $n=100$ )

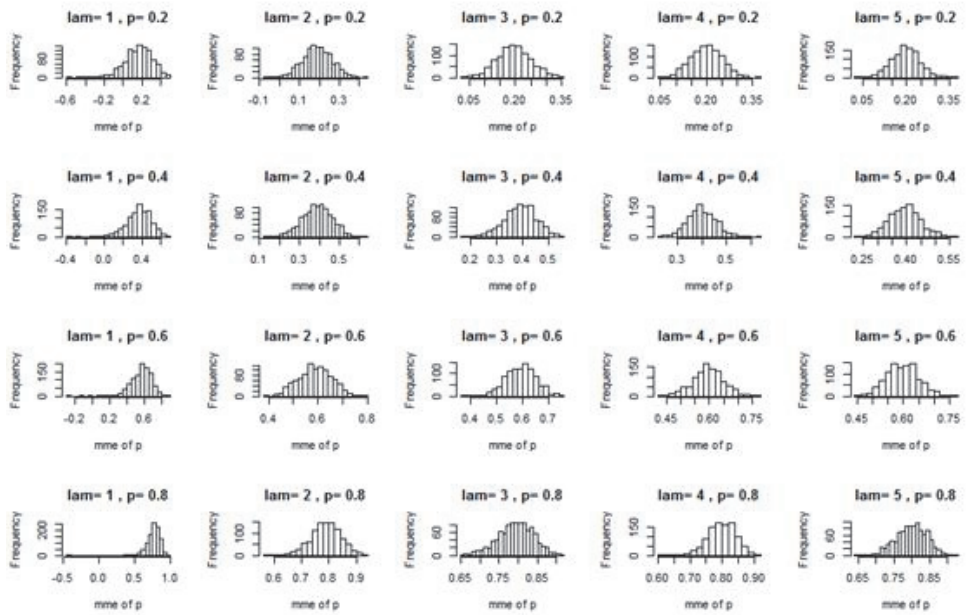


그림 6. 모의실험 결과를 이용한 히스토그램 :  $\hat{p}_{mm}$  ( $n=100$ )

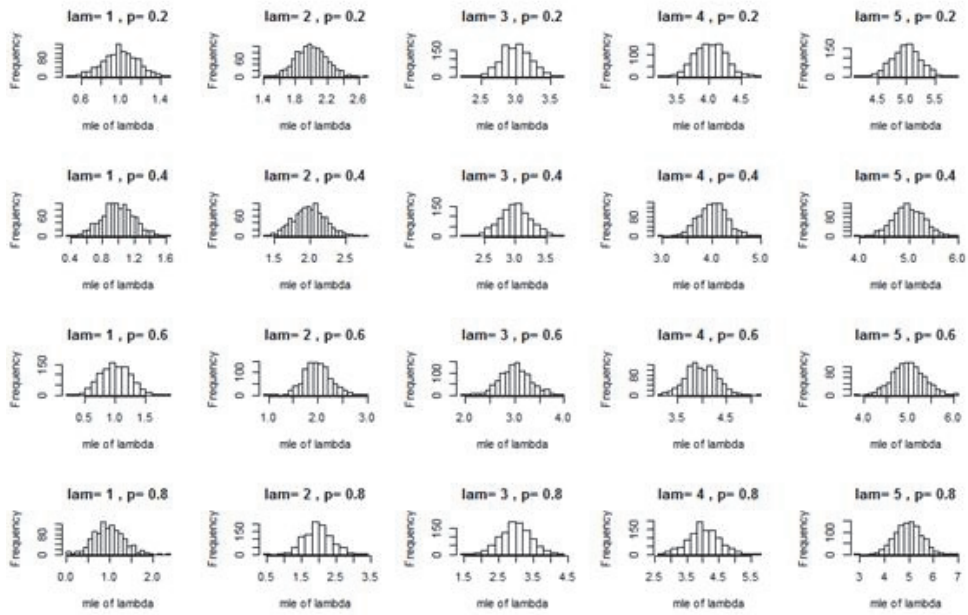


그림 7. 모의실험 결과를 이용한 히스토그램 :  $\hat{\lambda}_{ml}$  ( $n=100$ )

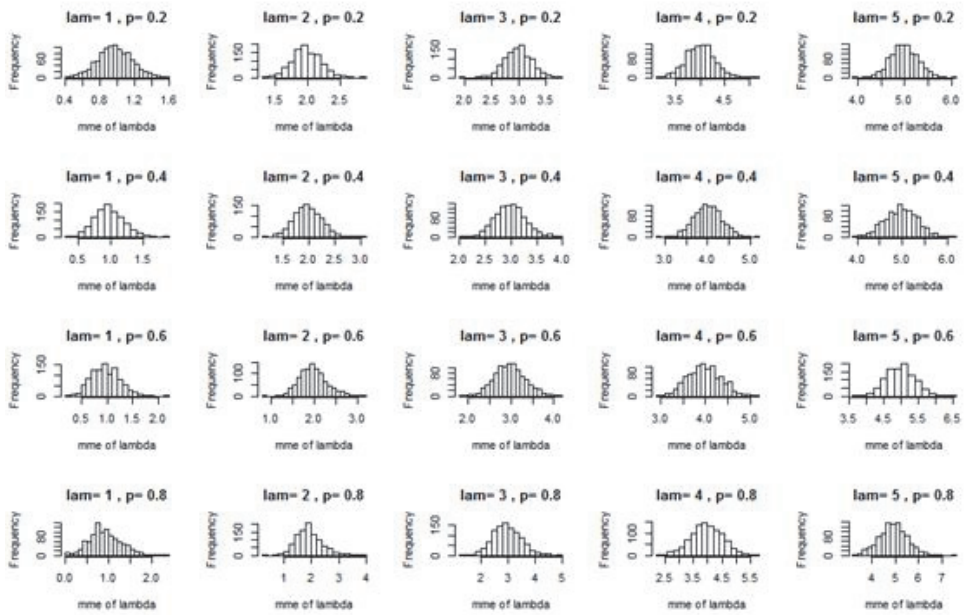


그림 8. 모의실험 결과를 이용한 히스토그램 :  $\hat{\lambda}_{mm}$  ( $n=100$ )

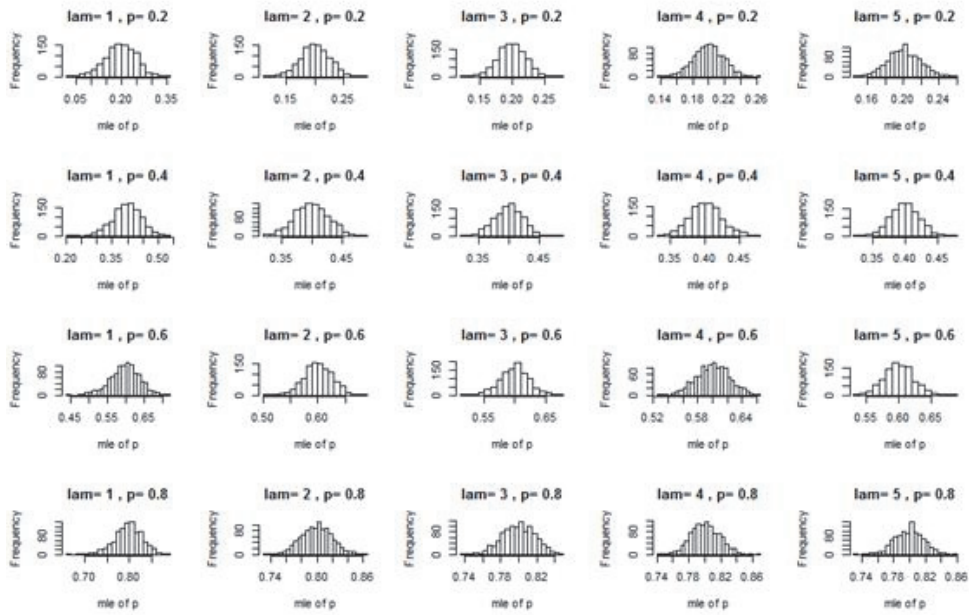


그림 9. 모의실험 결과를 이용한 히스토그램 :  $\hat{p}_{ml}$  ( $n=500$ )

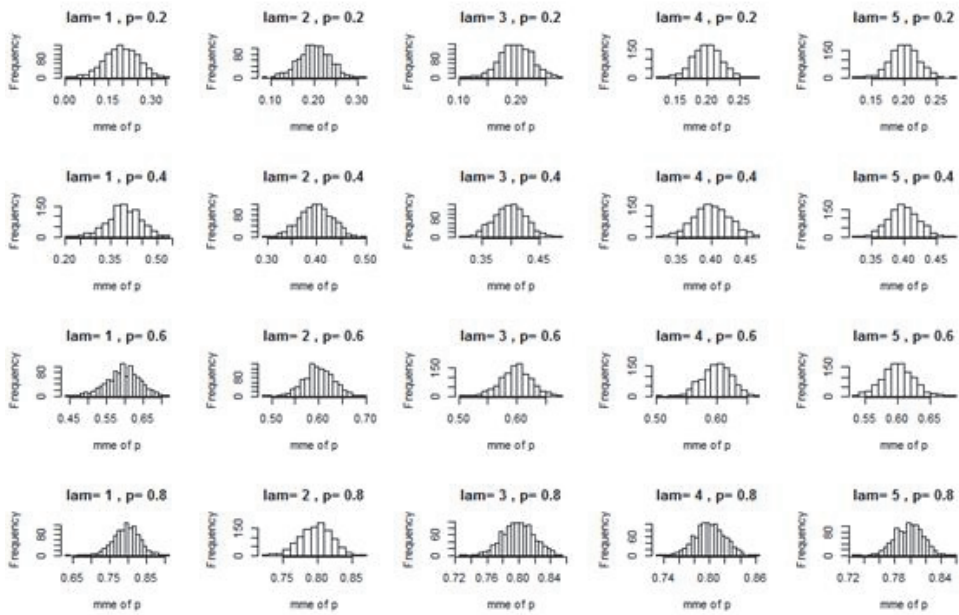


그림 10. 모의실험 결과를 이용한 히스토그램 :  $\hat{p}_{mm}$  ( $n=500$ )

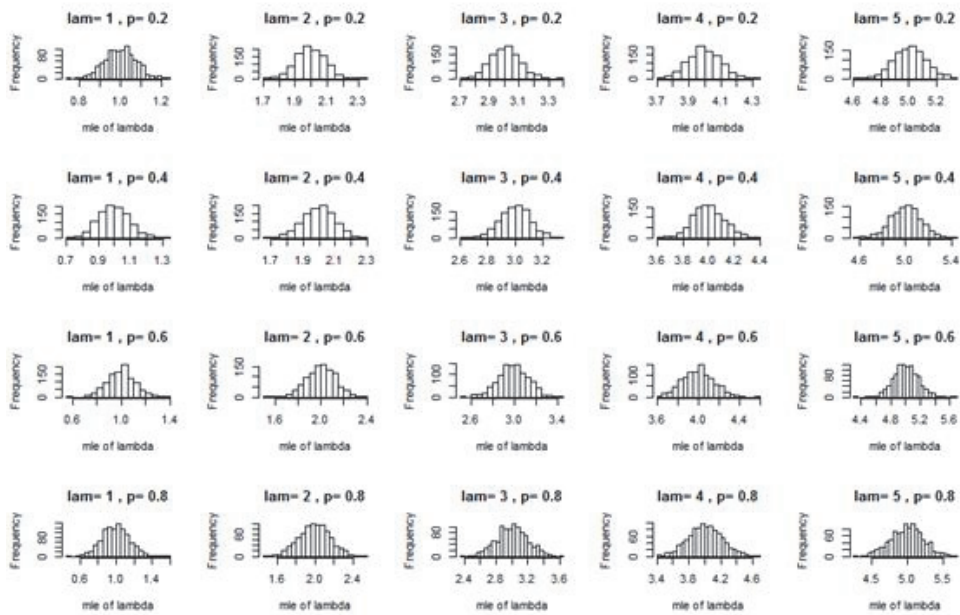


그림 11. 모의실험 결과를 이용한 히스토그램 :  $\hat{\lambda}_{ml}$  ( $n=500$ )

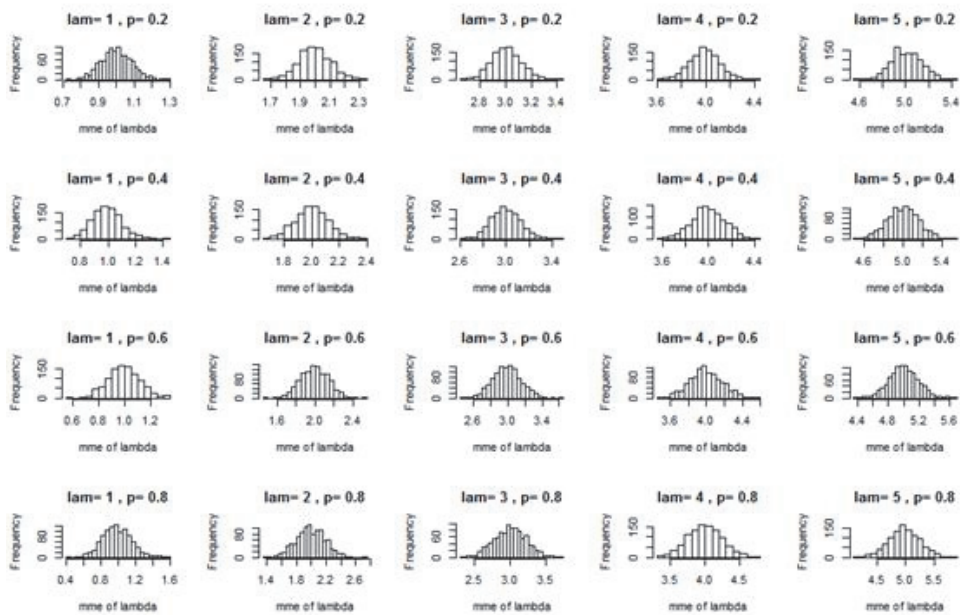


그림 12. 모의실험 결과를 이용한 히스토그램 :  $\hat{\lambda}_{mm}$  ( $n=500$ )

표 1. 음수로 추정되는  $\hat{p}_{ml}$  의 수

$n$	$p$	$\lambda$				
		1	2	3	4	5
50	0.2	165	14	0	0	0
	0.4	31	1	0	0	0
	0.6	22	0	0	0	0
	0.8	3	0	0	0	0
100	0.2	78	1	0	0	0
	0.4	8	0	0	0	0
	0.6	3	0	0	0	0
	0.8	3	0	0	0	0
500	0.2	0	0	0	0	0
	0.4	0	0	0	0	0
	0.6	0	0	0	0	0
	0.8	0	0	0	0	0

표 2. -Inf로 추정되는  $\hat{p}_{ml}$  의 수

$n$	$p$	$\lambda$				
		1	2	3	4	5
50	0.2	0	0	0	0	0
	0.4	0	0	0	0	0
	0.6	5	0	0	0	0
	0.8	66	1	0	0	0
100	0.2	0	0	0	0	0
	0.4	0	0	0	0	0
	0.6	0	0	0	0	0
	0.8	10	0	0	0	0
500	0.2	0	0	0	0	0
	0.4	0	0	0	0	0
	0.6	0	0	0	0	0
	0.8	0	0	0	0	0

표 3. 음수로 추정되는  $\hat{p}_{mm}$  의 수

$n$	$p$	$\lambda$				
		1	2	3	4	5
50	0.2	237	57	11	5	0
	0.4	53	1	0	0	0
	0.6	36	0	0	0	0
	0.8	63	1	0	0	0
100	0.2	130	9	0	0	0
	0.4	12	0	0	0	0
	0.6	3	0	0	0	0
	0.8	3	0	0	0	0
500	0.2	0	0	0	0	0
	0.4	0	0	0	0	0
	0.6	0	0	0	0	0
	0.8	0	0	0	0	0

표 4. -Inf로 추정되는  $\hat{p}_{mm}$  의 수

$n$	$p$	$\lambda$				
		1	2	3	4	5
50	0.2	0	0	0	0	0
	0.4	0	0	0	0	0
	0.6	5	0	0	0	0
	0.8	66	1	0	0	0
100	0.2	0	0	0	0	0
	0.4	0	0	0	0	0
	0.6	0	0	0	0	0
	0.8	10	0	0	0	0
500	0.2	0	0	0	0	0
	0.4	0	0	0	0	0
	0.6	0	0	0	0	0
	0.8	0	0	0	0	0

$m(=1000)$ 번 반복한  $p$ 의 추정량 중  $j$ 번째 반복해서 얻어진 추정량  $\hat{p}_j (j=1,2,\dots,1000)$  이  $p$ 의 모수공간  $[0, 1]$ 에 포함되지 않는 경우를 제외하고 편차(bias)와 평균제곱오차 (MSE)를 비교하였다. 모수  $\theta (p$  또는  $\lambda)$ 의 추정량에 대한 편차와 평균제곱오차는 다음과 같이 정의하였다.

$$bias = \left| \frac{1}{m} \sum_{j=1}^m \hat{\theta}_j - \theta \right|,$$

$$MSE = \frac{1}{m-1} \sum_{j=1}^m (\hat{\theta}_j - \bar{\theta})^2 + bias^2,$$

여기에서  $\bar{\theta} = \frac{1}{m} \sum_{j=1}^m \hat{\theta}_j$  이며, bias가 0인 경우, 즉 불편추정량의 경우 MSE가 곧 분산이므로 가장 좋은 추정량으로 볼 수 있다. bias가 0보다 큰 경우, 즉 편향성이 있는 경우에는 MSE가 가장 작은 추정량이 제일 좋은 추정량이다. 모의실험 결과는 표4, 표5에서 확인할 수 있다.  $n$ 이 작고  $\lambda$ 가 작은 경우에  $p$ 가 커질수록  $p$ 의 bias가 증가하는 것을 알 수 있다.  $\lambda$ 가 클 때에 비해  $\lambda$ 이 작을 경우  $p$ 의 최우추정량과 적률추정량의 bias값이 작아지는데, 이는  $n$ 이 작을 때 차이를 쉽게 확인할 수 있다.  $n$ 이 커질수록 bias와 MSE는 작아지는 경향이 있으며, 몇 가지의 경우를 제외하고 최우추정량의 bias가 적률추정량의 bias보다 작다. 마찬가지로 대부분의 경우에 최우추정량의 MSE가 적률추정량의 MSE에 비해 작기 때문에 최우추정량이 적률추정량보다 더 좋은 추정량이라고 할 수 있다.

표 5. 모의실험 결과:  $\hat{p}$ 의 bias,  $\hat{\lambda}$ 의 bias ( $n=50,100,500$ )

$n$	추정량	$p \setminus \lambda$	$\hat{p}$ 의 bias					$\hat{\lambda}$ 의 bias				
			1	2	3	4	5	1	2	3	4	5
50	MLE	0.2	0.020	0.006	0.002	0.003	0.002	0.008	0.002	0.012	0.015	0.007
		0.4	0.017	0.008	0.000	0.003	0.002	0.001	0.014	0.009	0.008	0.017
		0.6	0.030	0.008	0.002	0.004	0.002	0.008	0.006	0.008	0.010	0.009
		0.8	0.034	0.008	0.004	0.001	0.000	0.024	0.005	0.001	0.029	0.020
	MME	0.2	0.022	0.010	0.008	0.012	0.002	0.029	0.021	0.029	0.023	0.025
		0.4	0.034	0.022	0.007	0.009	0.006	0.026	0.042	0.024	0.036	0.039
		0.6	0.055	0.023	0.011	0.004	0.007	0.047	0.056	0.050	0.064	0.049
		0.8	0.061	0.022	0.015	0.006	0.005	0.114	0.068	0.109	0.140	0.127
100	MLE	0.2	0.003	0.005	0.002	0.002	0.002	0.011	0.005	0.001	0.008	0.003
		0.4	0.015	0.004	0.006	0.001	0.001	0.008	0.008	0.004	0.013	0.007
		0.6	0.020	0.003	0.002	0.002	0.001	0.001	0.003	0.004	0.005	0.005
		0.8	0.019	0.003	0.001	0.003	0.002	0.027	0.024	0.006	0.020	0.021
	MME	0.2	0.002	0.009	0.005	0.000	0.003	0.023	0.013	0.007	0.004	0.008
		0.4	0.028	0.011	0.009	0.002	0.005	0.023	0.019	0.012	0.000	0.027
		0.6	0.035	0.010	0.008	0.004	0.003	0.019	0.023	0.033	0.018	0.022
		0.8	0.031	0.011	0.005	0.001	0.001	0.061	0.068	0.041	0.030	0.080
500	MLE	0.2	0.003	0.000	0.001	0.001	0.001	0.000	0.002	0.005	0.004	0.007
		0.4	0.003	0.001	0.001	0.001	0.001	0.004	0.00	0.002	0.001	0.003
		0.6	0.003	0.001	0.000	0.001	0.001	0.003	0.001	0.011	0.001	0.003
		0.8	0.003	0.001	0.000	0.001	0.001	0.004	0.001	0.004	0.004	0.004
	MME	0.2	0.005	0.000	0.001	0.001	0.001	0.001	0.001	0.005	0.007	0.008
		0.4	0.007	0.002	0.001	0.000	0.001	0.009	0.001	0.004	0.003	0.001
		0.6	0.005	0.000	0.001	0.000	0.000	0.005	0.004	0.011	0.005	0.001
		0.8	0.006	0.002	0.001	0.000	0.001	0.013	0.012	0.002	0.003	0.019

표 6. 모의실험 결과:  $\hat{p}$ 의 MSE,  $\hat{\lambda}$ 의 MSE ( $n=50,100,500$ )

$n$	추정량 $p \setminus \lambda$	$\hat{p}$ 의 MSE					$\hat{\lambda}$ 의 MSE					
		1	2	3	4	5	1	2	3	4	5	
50	MLE	0.2	0.016	0.007	0.004	0.004	0.003	0.063	0.071	0.092	0.116	0.130
		0.4	0.022	0.008	0.006	0.005	0.005	0.081	0.102	0.122	0.156	0.162
		0.6	0.023	0.006	0.006	0.005	0.005	0.133	0.156	0.198	0.240	0.258
		0.8	0.016	0.005	0.003	0.003	0.003	0.277	0.322	0.429	0.502	0.542
	MME	0.2	0.018	0.009	0.006	0.005	0.004	0.073	0.095	0.135	0.154	0.169
		0.4	0.026	0.012	0.007	0.006	0.006	0.096	0.136	0.176	0.212	0.225
		0.6	0.030	0.009	0.007	0.005	0.006	0.157	0.208	0.266	0.321	0.345
		0.8	0.022	0.007	0.004	0.004	0.004	0.273	0.445	0.555	0.674	0.733
100	MLE	0.2	0.010	0.003	0.002	0.002	0.002	0.031	0.036	0.045	0.058	0.061
		0.4	0.011	0.004	0.003	0.003	0.002	0.039	0.049	0.062	0.078	0.090
		0.6	0.011	0.003	0.003	0.003	0.003	0.064	0.074	0.099	0.099	0.123
		0.8	0.007	0.002	0.002	0.002	0.002	0.123	0.160	0.190	0.226	0.285
	MME	0.2	0.012	0.005	0.003	0.003	0.002	0.039	0.049	0.060	0.083	0.084
		0.4	0.015	0.006	0.004	0.003	0.003	0.047	0.070	0.083	0.103	0.126
		0.6	0.014	0.005	0.003	0.003	0.003	0.079	0.102	0.131	0.138	0.169
		0.8	0.009	0.003	0.002	0.002	0.002	0.145	0.226	0.263	0.298	0.378
500	MLE	0.2	0.002	0.001	0.000	0.000	0.000	0.006	0.008	0.009	0.010	0.014
		0.4	0.002	0.001	0.001	0.001	0.000	0.009	0.010	0.011	0.015	0.017
		0.6	0.002	0.001	0.001	0.000	0.001	0.012	0.015	0.019	0.023	0.028
		0.8	0.001	0.000	0.000	0.000	0.000	0.023	0.029	0.038	0.045	0.053
	MME	0.2	0.003	0.001	0.001	0.001	0.000	0.007	0.011	0.013	0.015	0.018
		0.4	0.003	0.001	0.001	0.001	0.001	0.011	0.014	0.016	0.020	0.024
		0.6	0.002	0.001	0.001	0.001	0.001	0.015	0.021	0.027	0.032	0.038
		0.8	0.001	0.001	0.000	0.000	0.000	0.028	0.039	0.053	0.061	0.072

### 3.1.2. 이변량 영과잉 포아송 분포

최우추정량과 적률추정량 중 더 좋은 추정량을 알아보기 위해 (2.1)에서  $n = 500$ 인  $Y_1, Y_2$ 를  $m (= 1000)$ 번 추출하여 모든 경우에 대해 최우추정량과 적률추정량을 구하는 모의실험을 진행하였다. 이때  $\lambda_{10}$ 과  $\lambda_{00}$ 는 1부터 3까지 1씩 증가시키고  $p_0$ 는 0.2부터 0.8까지 0.2씩 증가시켰으며  $\lambda_{20}$ 은 1로 고정하였다.  $\lambda_{10}, \lambda_{20}, \lambda_{00}, p_0$ 의 편차(bias)와 평균제곱오차 (MSE)를 비교하였고 모수  $\theta(p, \lambda)$ 의 추정량에 대한 편차와 평균제곱오차는 3.1절의 bias 및 MSE와 동일하게 정의하였다.

표 9~표14는 모의실험 결과이다. 대부분의 경우에서  $n$ 이 커지면  $\lambda_{10}$  최우추정량과  $p_0$  최우추정량의 bias는 감소하였다.  $\lambda_{00}$ 이 커질수록  $\lambda_{10}$  적률추정량의 bias는 증가하였고  $p_0$  적률추정량의 bias는 감소하였다. 대체적으로  $p_0$ 가 작을 때에 비해  $p_0$ 가 큰 경우에  $\lambda_{10}, \lambda_{20}, \lambda_{00}$ 의 최우추정량과 적률추정량의 bias가 더 큰 것을 확인 할 수 있었다.

MSE의 경우,  $n$ 이 커지면  $p_0$ 의 최우추정량의 MSE와 적률추정량의 MSE는 작아지는 것을 확인 할 수 있다. 또한  $p_0$ 가 커질수록  $p_0$ 의 최우추정량의 MSE와 적률추정량의 MSE 차이가 작아진다 (그림13 참고).  $n$ 이 커질수록  $\lambda_{10}$  적률추정량의 MSE는 작아지며,  $p_0$ 가 크고  $\lambda_{00}$ 가 증가할수록  $\lambda_{10}$  적률추정량의 MSE는 증가하는 것을 확인 할 수 있고 최우추정량의 MSE와 적률추정량의 MSE 차이도 증가하는 것을 알 수 있다(그림14 참고). 결과적으로 대부분의 경우에서 적률추정량의 bias보다 최우추정량의 bias가 작고, 적률추정량의 MSE보다 최우추정량의 MSE가 작은 것으로 보아 적률추정량보다 최우추정량이 더 좋은 추정량임을 알 수 있다.

표 7. 모의실험 결과 추정량의 bias( $n=100$ )

parameter			bias of MLE				bias of MME			
$\lambda_{10}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$
1	1	0.2	0.009	0.010	0.005	0.003	0.028	0.036	0.017	0.007
		0.4	0.010	0.010	0.004	0.003	0.028	0.032	0.013	0.025
		0.6	0.031	0.025	0.017	0.002	0.059	0.061	0.034	0.025
		0.8	0.024	0.002	0.007	0.001	0.014	0.011	0.047	0.027
	2	0.2	0.008	0.009	0.003	0.001	0.003	0.007	0.015	0.011
		0.4	0.009	0.009	0.002	0.002	0.003	0.002	0.019	0.015
		0.6	0.031	0.025	0.019	0.001	0.020	0.019	0.004	0.013
		0.8	0.028	0.006	0.018	0.000	0.065	0.096	0.128	0.014
	3	0.2	0.007	0.008	0.004	0.000	0.009	0.003	0.018	0.008
		0.4	0.007	0.007	0.004	0.001	0.016	0.015	0.030	0.009
		0.6	0.029	0.023	0.020	0.000	0.010	0.005	0.013	0.008
		0.8	0.028	0.005	0.023	0.001	0.129	0.165	0.193	0.009
2	1	0.2	0.008	0.004	0.002	0.001	0.023	0.032	0.013	0.011
		0.4	0.013	0.005	0.003	0.001	0.022	0.024	0.005	0.018
		0.6	0.043	0.026	0.021	0.001	0.059	0.055	0.027	0.017
		0.8	0.047	0.017	0.029	0.000	0.002	0.025	0.061	0.019
	2	0.2	0.007	0.004	0.001	0.000	0.001	0.008	0.014	0.010
		0.4	0.013	0.006	0.003	0.001	0.001	0.002	0.019	0.012
		0.6	0.041	0.026	0.024	0.000	0.035	0.028	0.004	0.010
		0.8	0.040	0.014	0.033	0.001	0.069	0.102	0.134	0.011
	3	0.2	0.007	0.004	0.003	0.001	0.002	0.002	0.014	0.006
		0.4	0.012	0.005	0.004	0.000	0.008	0.011	0.025	0.007
		0.6	0.042	0.027	0.026	0.000	0.029	0.017	0.001	0.006
		0.8	0.048	0.022	0.043	0.001	0.132	0.170	0.198	0.007
3	1	0.2	0.013	0.009	0.009	0.000	0.018	0.026	0.007	0.011
		0.4	0.024	0.014	0.014	0.001	0.023	0.024	0.005	0.016
		0.6	0.055	0.034	0.032	0.000	0.062	0.055	0.028	0.014
		0.8	0.064	0.033	0.051	0.001	0.005	0.030	0.066	0.016
	2	0.2	0.013	0.009	0.008	0.001	0.002	0.007	0.015	0.009
		0.4	0.024	0.016	0.015	0.000	0.006	0.008	0.013	0.010
		0.6	0.057	0.037	0.037	0.000	0.046	0.036	0.012	0.008
		0.8	0.046	0.018	0.042	0.001	0.070	0.101	0.133	0.009
	3	0.2	0.014	0.011	0.010	0.001	0.001	0.005	0.011	0.006
		0.4	0.025	0.017	0.017	0.000	0.003	0.000	0.015	0.006
		0.6	0.058	0.039	0.039	0.000	0.043	0.029	0.011	0.005
		0.8	0.059	0.033	0.056	0.001	0.129	0.165	0.193	0.006

표 8. 모의실험 결과 추정량의 MSE( $n=100$ )

parameter			MSE of MLE				MSE of MME			
$\lambda_{10}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$
1	1	0.2	0.038	0.039	0.032	0.002	0.292	0.270	0.349	0.012
		0.4	0.051	0.051	0.042	0.003	0.404	0.360	0.476	0.013
		0.6	0.075	0.078	0.064	0.003	0.660	0.604	0.765	0.010
		0.8	0.167	0.168	0.145	0.002	1.036	0.933	1.161	0.006
	2	0.2	0.035	0.036	0.044	0.002	0.485	0.461	0.609	0.007
		0.4	0.048	0.048	0.058	0.002	0.680	0.644	0.855	0.008
		0.6	0.071	0.073	0.090	0.003	1.089	1.035	1.353	0.006
		0.8	0.151	0.155	0.196	0.002	1.767	1.674	2.197	0.003
	3	0.2	0.034	0.036	0.056	0.002	0.752	0.736	0.969	0.005
		0.4	0.047	0.048	0.073	0.002	1.020	0.998	1.315	0.005
		0.6	0.069	0.073	0.116	0.003	1.670	1.629	2.108	0.004
		0.8	0.147	0.151	0.246	0.002	2.742	2.647	3.493	0.003
2	1	0.2	0.075	0.062	0.056	0.002	0.288	0.249	0.320	0.09
		0.4	0.098	0.080	0.073	0.003	0.353	0.298	0.396	0.009
		0.6	0.154	0.122	0.109	0.003	0.559	0.479	0.617	0.007
		0.8	0.311	0.253	0.227	0.002	0.899	0.761	0.964	0.004
	2	0.2	0.074	0.062	0.069	0.002	0.496	0.452	0.594	0.006
		0.4	0.097	0.080	0.089	0.002	0.650	0.600	0.797	0.006
		0.6	0.150	0.122	0.135	0.003	1.032	0.949	1.251	0.005
		0.8	0.315	0.269	0.311	0.002	1.620	1.480	1.977	0.003
	3	0.2	0.073	0.063	0.082	0.002	0.784	0.747	0.975	0.004
		0.4	0.096	0.082	0.107	0.002	1.020	0.981	1.288	0.005
		0.6	0.150	0.122	0.161	0.003	1.642	1.565	2.034	0.004
		0.8	0.304	0.260	0.353	0.002	2.655	2.510	3.329	0.002
3	1	0.2	0.113	0.087	0.084	0.002	0.299	0.246	0.312	0.007
		0.4	0.147	0.113	0.108	0.002	0.375	0.302	0.392	0.008
		0.6	0.228	0.169	0.160	0.003	0.586	0.479	0.608	0.006
		0.8	0.461	0.349	0.325	0.002	0.962	0.787	0.971	0.004
	2	0.2	0.117	0.093	0.103	0.002	0.512	0.452	0.591	0.005
		0.4	0.151	0.119	0.131	0.002	0.674	0.603	0.793	0.005
		0.6	0.229	0.177	0.193	0.003	1.059	0.943	1.242	0.004
		0.8	0.517	0.423	0.468	0.002	1.746	1.568	2.055	0.003
	3	0.2	0.118	0.096	0.118	0.002	0.807	0.753	0.978	0.004
		0.4	0.155	0.125	0.152	0.002	1.050	0.988	1.288	0.004
		0.6	0.237	0.184	0.224	0.003	1.655	1.542	2.007	0.004
		0.8	0.513	0.421	0.516	0.002	2.815	2.628	3.439	0.002

표 9. 모의실험 결과 추정량의 bias( $n=200$ )

parameter			bias of MLE				bias of MME			
$\lambda_{10}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$
1	1	0.2	0.001	0.006	0.001	0.001	0.025	0.025	0.026	0.003
		0.4	0.003	0.006	0.001	0.001	0.025	0.029	0.024	0.009
		0.6	0.005	0.009	0.004	0.000	0.029	0.026	0.018	0.011
		0.8	0.011	0.014	0.014	0.001	0.050	0.051	0.036	0.012
	2	0.2	0.003	0.004	0.003	0.000	0.013	0.016	0.016	0.004
		0.4	0.005	0.004	0.001	0.000	0.006	0.013	0.007	0.005
		0.6	0.002	0.007	0.005	0.001	0.004	0.000	0.010	0.006
		0.8	0.010	0.013	0.019	0.000	0.032	0.028	0.041	0.007
	3	0.2	0.003	0.004	0.006	0.000	0.001	0.003	0.005	0.003
		0.4	0.004	0.006	0.000	0.000	0.005	0.001	0.005	0.003
		0.6	0.004	0.008	0.007	0.001	0.026	0.024	0.032	0.003
		0.8	0.012	0.015	0.023	0.001	0.054	0.051	0.059	0.004
2	1	0.2	0.004	0.003	0.001	0.000	0.021	0.021	0.022	0.003
		0.4	0.006	0.004	0.000	0.000	0.016	0.021	0.015	0.006
		0.6	0.002	0.008	0.007	0.001	0.022	0.021	0.013	0.007
		0.8	0.018	0.022	0.027	0.000	0.051	0.052	0.037	0.008
	2	0.2	0.006	0.002	0.003	0.001	0.013	0.016	0.017	0.002
		0.4	0.006	0.004	0.001	0.000	0.004	0.012	0.005	0.004
		0.6	0.003	0.009	0.010	0.002	0.002	0.004	0.007	0.004
		0.8	0.018	0.023	0.031	0.001	0.008	0.003	0.016	0.005
	3	0.2	0.005	0.002	0.006	0.001	0.002	0.006	0.008	0.002
		0.4	0.005	0.005	0.002	0.000	0.006	0.001	0.004	0.002
		0.6	0.006	0.012	0.013	0.002	0.024	0.018	0.027	0.002
		0.8	0.023	0.028	0.039	0.001	0.033	0.029	0.037	0.003
3	1	0.2	0.002	0.006	0.004	0.000	0.022	0.020	0.021	0.002
		0.4	0.002	0.009	0.004	0.000	0.017	0.020	0.015	0.005
		0.6	0.012	0.018	0.017	0.001	0.028	0.022	0.014	0.006
		0.8	0.034	0.036	0.042	0.000	0.065	0.058	0.044	0.007
	2	0.2	0.004	0.005	0.005	0.001	0.017	0.019	0.020	0.002
		0.4	0.000	0.011	0.007	0.000	0.009	0.016	0.009	0.003
		0.6	0.017	0.023	0.024	0.002	0.008	0.009	0.001	0.003
		0.8	0.040	0.044	0.051	0.001	0.015	0.013	0.000	0.004
	3	0.2	0.003	0.005	0.008	0.001	0.008	0.011	0.013	0.001
		0.4	0.000	0.011	0.008	0.000	0.000	0.007	0.001	0.002
		0.6	0.020	0.026	0.028	0.002	0.011	0.009	0.018	0.002
		0.8	0.046	0.050	0.061	0.001	0.008	0.009	0.017	0.003

표 10. 모의실험 결과 추정량의 MSE( $n=200$ )

parameter			MSE of MLE				MSE of MME			
$\lambda_{10}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$
1	1	0.2	0.019	0.019	0.016	0.001	0.146	0.144	0.172	0.007
		0.4	0.026	0.026	0.022	0.001	0.205	0.198	0.239	0.006
		0.6	0.037	0.036	0.032	0.001	0.301	0.292	0.364	0.005
		0.8	0.083	0.080	0.071	0.001	0.693	0.621	0.787	0.003
	2	0.2	0.018	0.018	0.023	0.001	0.255	0.253	0.318	0.004
		0.4	0.024	0.024	0.032	0.001	0.357	0.349	0.441	0.004
		0.6	0.034	0.033	0.047	0.001	0.571	0.560	0.714	0.003
		0.8	0.075	0.072	0.100	0.001	1.085	1.010	1.328	0.002
	3	0.2	0.017	0.017	0.030	0.001	0.375	0.369	0.460	0.003
		0.4	0.023	0.023	0.041	0.001	0.532	0.524	0.653	0.002
		0.6	0.033	0.032	0.060	0.001	0.839	0.828	1.042	0.002
		0.8	0.072	0.069	0.128	0.001	1.650	1.597	2.042	0.001
2	1	0.2	0.037	0.029	0.027	0.001	0.149	0.134	0.160	0.005
		0.4	0.050	0.040	0.038	0.001	0.206	0.182	0.221	0.004
		0.6	0.076	0.057	0.056	0.001	0.298	0.264	0.331	0.003
		0.8	0.165	0.127	0.125	0.001	0.689	0.575	0.738	0.002
	2	0.2	0.037	0.029	0.035	0.001	0.261	0.245	0.307	0.003
		0.4	0.049	0.041	0.049	0.001	0.366	0.339	0.428	0.003
		0.6	0.074	0.059	0.073	0.001	0.562	0.526	0.675	0.002
		0.8	0.164	0.131	0.163	0.001	1.172	1.046	1.368	0.001
	3	0.2	0.036	0.029	0.042	0.001	0.388	0.368	0.456	0.002
		0.4	0.048	0.040	0.058	0.001	0.554	0.527	0.655	0.002
		0.6	0.073	0.058	0.086	0.001	0.856	0.820	1.029	0.002
		0.8	0.157	0.127	0.188	0.001	1.773	1.667	2.120	0.001
3	1	0.2	0.053	0.038	0.037	0.001	0.157	0.135	0.160	0.004
		0.4	0.072	0.053	0.053	0.001	0.218	0.183	0.221	0.004
		0.6	0.110	0.077	0.079	0.001	0.312	0.258	0.323	0.003
		0.8	0.245	0.176	0.177	0.001	0.762	0.608	0.771	0.002
	2	0.2	0.056	0.041	0.047	0.001	0.271	0.248	0.309	0.003
		0.4	0.075	0.058	0.066	0.001	0.375	0.338	0.425	0.002
		0.6	0.115	0.085	0.100	0.001	0.570	0.514	0.658	0.002
		0.8	0.259	0.197	0.233	0.001	1.262	1.095	1.415	0.001
	3	0.2	0.057	0.043	0.054	0.001	0.399	0.373	0.461	0.002
		0.4	0.076	0.059	0.075	0.001	0.569	0.530	0.657	0.002
		0.6	0.117	0.087	0.114	0.001	0.878	0.818	1.025	0.002
		0.8	0.256	0.198	0.261	0.001	1.888	1.733	2.189	0.001

표 11. 모의실험 결과 추정량의 bias( $n=500$ )

parameter			bias of MLE				bias of MME			
$\lambda_{10}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$
1	1	0.2	0.004	0.001	0.000	0.000	0.004	0.007	0.005	0.004
		0.4	0.001	0.003	0.003	0.001	0.009	0.009	0.010	0.005
		0.6	0.006	0.008	0.007	0.000	0.022	0.016	0.018	0.003
		0.8	0.005	0.010	0.004	0.000	0.023	0.018	0.015	0.004
	2	0.2	0.004	0.001	0.001	0.000	0.001	0.003	0.001	0.003
		0.4	0.003	0.001	0.003	0.001	0.003	0.002	0.003	0.003
		0.6	0.004	0.006	0.008	0.000	0.017	0.012	0.013	0.002
		0.8	0.002	0.007	0.008	0.000	0.006	0.001	0.001	0.002
	3	0.2	0.004	0.001	0.003	0.001	0.005	0.003	0.004	0.002
		0.4	0.003	0.001	0.007	0.000	0.005	0.005	0.003	0.003
		0.6	0.003	0.005	0.011	0.001	0.006	0.002	0.005	0.001
		0.8	0.001	0.006	0.010	0.001	0.009	0.013	0.013	0.001
2	1	0.2	0.002	0.004	0.004	0.000	0.002	0.006	0.004	0.003
		0.4	0.001	0.005	0.007	0.000	0.007	0.007	0.008	0.003
		0.6	0.008	0.011	0.012	0.000	0.017	0.012	0.014	0.002
		0.8	0.006	0.010	0.008	0.000	0.020	0.013	0.010	0.002
	2	0.2	0.004	0.003	0.003	0.001	0.002	0.001	0.001	0.002
		0.4	0.001	0.004	0.007	0.000	0.001	0.001	0.001	0.003
		0.6	0.007	0.011	0.013	0.001	0.012	0.008	0.009	0.001
		0.8	0.005	0.009	0.012	0.001	0.003	0.004	0.006	0.002
	3	0.2	0.004	0.003	0.006	0.001	0.007	0.003	0.005	0.002
		0.4	0.002	0.003	0.010	0.000	0.005	0.004	0.002	0.002
		0.6	0.007	0.011	0.017	0.001	0.004	0.002	0.004	0.001
		0.8	0.005	0.010	0.014	0.001	0.006	0.011	0.012	0.001
3	1	0.2	0.001	0.007	0.006	0.000	0.002	0.004	0.002	0.002
		0.4	0.002	0.008	0.009	0.001	0.003	0.005	0.006	0.003
		0.6	0.010	0.014	0.015	0.000	0.015	0.010	0.012	0.002
		0.8	0.007	0.012	0.010	0.000	0.014	0.009	0.006	0.002
	2	0.2	0.005	0.003	0.003	0.000	0.006	0.001	0.004	0.002
		0.4	0.001	0.006	0.009	0.001	0.003	0.001	0.001	0.002
		0.6	0.009	0.013	0.016	0.001	0.009	0.006	0.007	0.001
		0.8	0.005	0.010	0.012	0.001	0.002	0.007	0.009	0.001
	3	0.2	0.006	0.003	0.005	0.001	0.011	0.006	0.007	0.001
		0.4	0.002	0.004	0.011	0.000	0.008	0.006	0.004	0.002
		0.6	0.010	0.014	0.019	0.001	0.003	0.001	0.003	0.001
		0.8	0.005	0.012	0.015	0.001	0.008	0.011	0.012	0.001

표 12. 모의실험 결과 추정량의 MSE( $n=500$ )

parameter			MSE of MLE				MSE of MME			
$\lambda_{10}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$	$\lambda_{10}$	$\lambda_{20}$	$\lambda_{00}$	$p_0$
1	1	0.2	0.008	0.007	0.007	0.000	0.056	0.052	0.065	0.003
		0.4	0.011	0.010	0.009	0.001	0.077	0.070	0.088	0.003
		0.6	0.016	0.015	0.013	0.001	0.116	0.106	0.132	0.002
		0.8	0.032	0.033	0.029	0.000	0.233	0.217	0.266	0.001
	2	0.2	0.008	0.007	0.009	0.000	0.106	0.098	0.128	0.002
		0.4	0.010	0.010	0.013	0.001	0.138	0.129	0.166	0.001
		0.6	0.015	0.014	0.018	0.001	0.203	0.190	0.244	0.001
		0.8	0.030	0.030	0.040	0.000	0.408	0.389	0.501	0.001
	3	0.2	0.007	0.007	0.012	0.000	0.164	0.156	0.199	0.001
		0.4	0.010	0.009	0.016	0.001	0.210	0.201	0.256	0.001
		0.6	0.014	0.014	0.023	0.001	0.303	0.291	0.371	0.001
		0.8	0.029	0.030	0.050	0.000	0.604	0.585	0.754	0.001
2	1	0.2	0.015	0.012	0.011	0.000	0.056	0.049	0.059	0.002
		0.4	0.021	0.016	0.015	0.001	0.077	0.065	0.080	0.002
		0.6	0.030	0.024	0.022	0.001	0.113	0.096	0.119	0.001
		0.8	0.062	0.050	0.048	0.000	0.226	0.197	0.240	0.001
	2	0.2	0.015	0.012	0.014	0.000	0.106	0.095	0.121	0.001
		0.4	0.020	0.016	0.019	0.001	0.141	0.126	0.160	0.001
		0.6	0.029	0.024	0.027	0.001	0.205	0.186	0.238	0.001
		0.8	0.061	0.050	0.060	0.000	0.411	0.376	0.484	0.001
	3	0.2	0.015	0.012	0.017	0.000	0.166	0.153	0.194	0.001
		0.4	0.020	0.016	0.023	0.001	0.216	0.199	0.252	0.001
		0.6	0.029	0.024	0.033	0.001	0.311	0.290	0.370	0.001
		0.8	0.060	0.050	0.072	0.000	0.621	0.583	0.748	0.000
3	1	0.2	0.022	0.016	0.016	0.000	0.061	0.050	0.060	0.002
		0.4	0.029	0.022	0.022	0.001	0.082	0.067	0.081	0.001
		0.6	0.043	0.032	0.031	0.001	0.120	0.100	0.122	0.001
		0.8	0.088	0.066	0.066	0.000	0.243	0.206	0.248	0.001
	2	0.2	0.023	0.018	0.020	0.000	0.112	0.097	0.122	0.001
		0.4	0.031	0.024	0.027	0.001	0.147	0.128	0.161	0.001
		0.6	0.045	0.035	0.038	0.001	0.216	0.192	0.242	0.001
		0.8	0.091	0.071	0.082	0.000	0.432	0.389	0.495	0.001
	3	0.2	0.024	0.018	0.024	0.000	0.173	0.55	0.195	0.001
		0.4	0.032	0.025	0.032	0.001	0.225	0.204	0.255	0.001
		0.6	0.046	0.036	0.044	0.001	0.329	0.301	0.378	0.001
		0.8	0.093	0.073	0.095	0.000	0.654	0.606	0.768	0.000

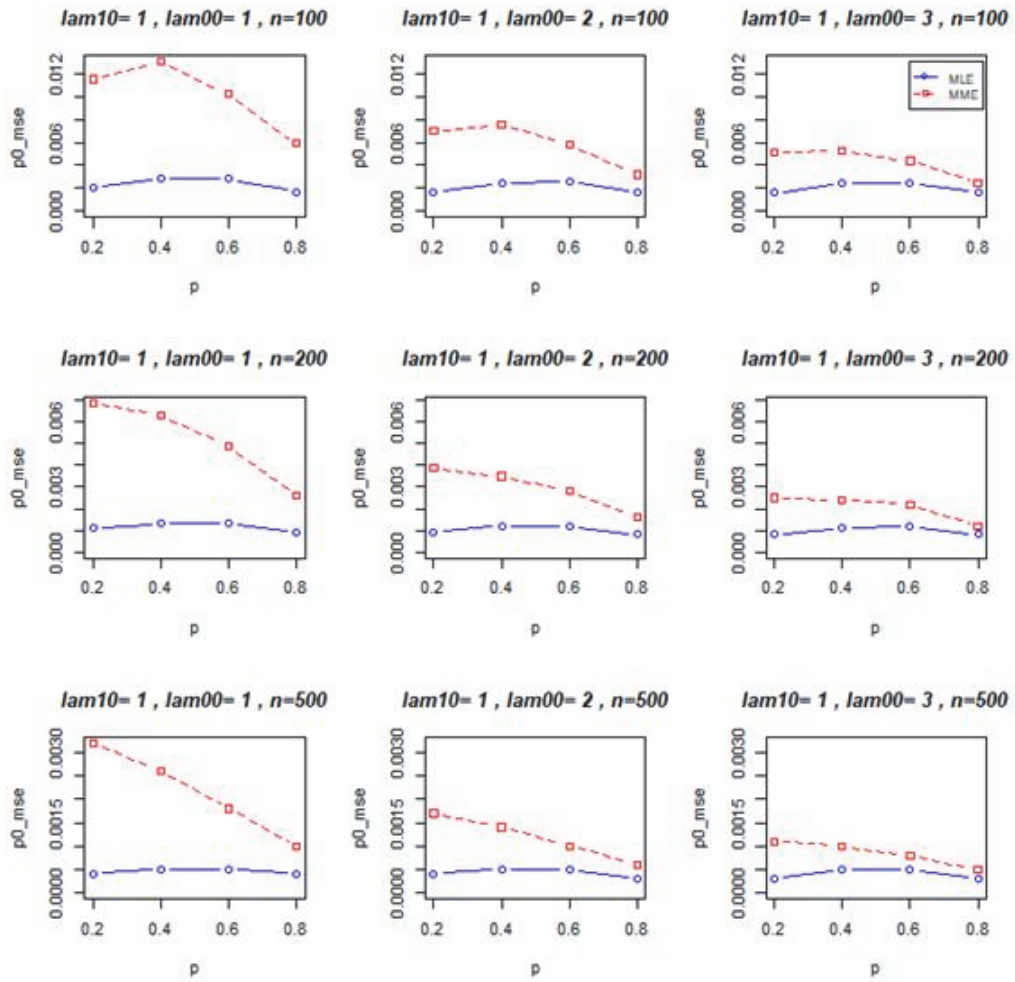


그림 13. MLE와 MME 비교(1)

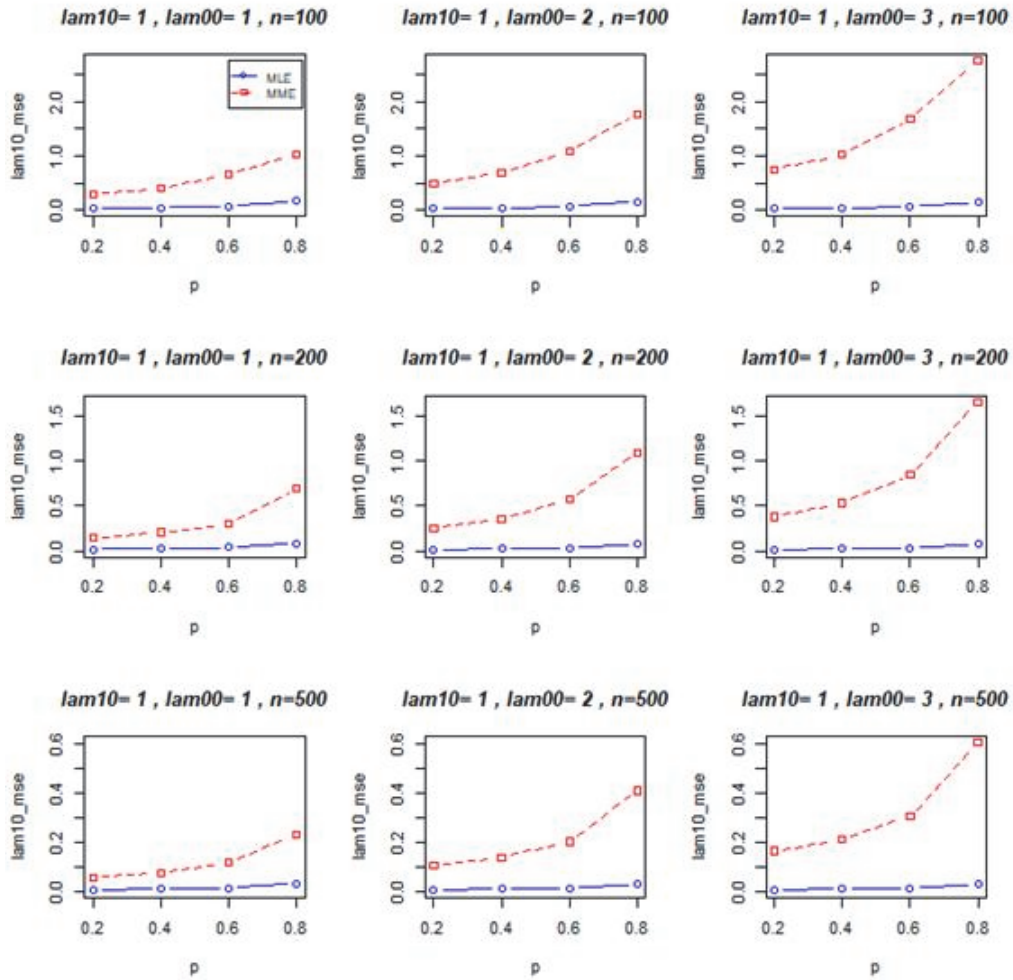


그림 14. MLE와 MME 비교(2)

### 3.2. 실증분석

실증 분석을 위해 사용한 자료는 한국 프로축구 리그인 K-리그에서의 2018년 경기 결과 자료이다(출처: 한국프로축구연맹). 2018년 3월 1일부터 2018년 8월 30일까지 22개의 팀이 참가하여 총 280번의 경기를 치렀으며, 각 경기가 실시된 날짜와 홈팀과 원정팀 이름, 양 팀의 전반전 및

후반전 득점수와 총 득점수 정보를 포함하고 있다.

경기 결과를 보면 전체 경기 중 홈팀과 원정팀의 동점인 경우는 약 26.4%이고 홈팀이 승리한 경우는 약 37.5%, 원정팀이 승리한 경우는 약 36.1%이다. 동점의 경우에서 0 대 0의 경우는 약 37.8%, 1대 1의 경우는 43.2%. 2대 2의 경우는 약 12.1%를 차지하였다. 또한 홈팀의 전반전 득점수가 0인 경우는 60%, 원정팀 전반전 득점수가 0인 경우는 61.8%이고 홈팀의 후반전 득점수가 0인 경우는 48.2%, 원정팀의 후반전 득점수가 0인 경우는 46.4%로 나타났다. 본 논문에서는 이러한 점을 고려하여  $Y_1$ ,  $Y_2$  를 홈팀 및 원정팀의 전체 득점수/ 전반전 득점수/ 후반전 득점수로 나누어 실증 분석을 진행하였다.

홈팀의 경기시간 내 전체 평균 득점수는 1.31이고 원정팀의 경기시간 내 평균 득점수는 1.25이다. 홈팀의 전반전 득점수는 0.53, 후반전 득점수는 0.78이고 원정팀의 전반전 득점수는 0.48, 후반전 득점수는 0.77로 양 팀 모두 전반전의 평균 득점수보다 후반전 평균 득점수가 높은 것을 알 수 있다.

최우추정량을 얻기 위해 R프로그램의 bivpois패키지(Karlis, 2005)를 이용하였고 적률추정량은 2.2.2의 계산 과정을 따랐다.

홈팀의 평균 득점수를  $\lambda_1(=\lambda_{10}+\lambda_{00})$ , 원정팀의 평균 득점수를  $\lambda_2(=\lambda_{20}+\lambda_{00})$ 라 할 수 있는데, 모수추정결과  $Y_1$ ,  $Y_2$  를 홈팀 및 원정팀의 전반전 득점수로 놓았을 때, 최우 추정법을 이용한 경우  $\lambda_{10}$ 은 0.519,  $\lambda_{00}$ 은 0.017로 추정되었고  $\lambda_1$ 의 최우추정량은 0.536이다.  $\lambda_{20}$ 은 0.465로 추정되었으므로  $\lambda_2$ 의 최우추정량은 0.482이다. 적률 추정법을 이용한 경우  $\lambda_{10}$ 은 0.530,  $\lambda_{20}$ 은 0.432 으로 추정되고  $\lambda_{00}$ 은 0.061로 추정되어  $\lambda_1$ 의 적률추정량은 0.591,  $\lambda_2$ 의 적률추정량은 0.493이다. 실제로 홈팀의 평균 전반전 득점수는 0.53, 원정팀의 평균 전반전 득점수는 0.48로 최우추정량이 실제 평균 득점수와 비슷하게 추정되었다.  $Y_1$ ,  $Y_2$  를 홈팀 및 원정팀의 후반전 득점수로 놓았을 때,  $\lambda_{00}$ 의 최우추

정량과 적률추정량이 모두 0이기 때문에  $\lambda_1$ 의 최우정량은  $\lambda_{10}$ 의 최우추정량과 같은 0.777,  $\lambda_2$ 의 최우추정량은  $\lambda_{20}$ 의 최우추정량과 같은 0.770이고  $\lambda_1$ 의 적률추정량은 0.838,  $\lambda_2$ 의 적률추정량은 0.743이다. 실제로 홈팀의 평균 후반전 득점수는 0.78 0.77로 최우추정량이 실제 평균 후반전 득점수와 비슷하게 추정되었다. 마지막으로  $Y_1, Y_2$ 를 홈팀 및 원정팀의 전체 득점수로 한다면, 최우추정량의 경우  $\lambda_{10}$ 은 1.300,  $\lambda_{20}$ 은 1.233,  $\lambda_{00}$ 은 0.047로  $\lambda_1$ 의 최우추정량은 1.347,  $\lambda_2$ 의 최우추정량은 1.280이다. 적률추정량의 경우  $\lambda_{10}$ 은 1.370,  $\lambda_{20}$ 은 1.075,  $\lambda_{00}$ 은 0.140로 추정되어  $\lambda_1$ 의 적률추정량은 1.510,  $\lambda_2$ 의 적률추정량은 1.510이다. 홈팀 및 원정팀의 실제 평균 전체 득점수는 1.31 1.25로 전체 득점수의 경우에도 최우추정량이 적률추정량보다 실제 평균 득점수와 더 근사한 값으로 추정하였다.

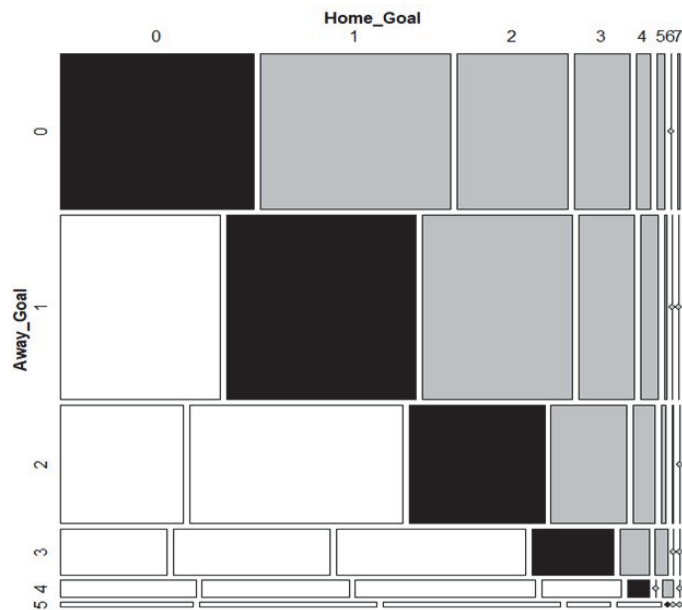


그림 15. 홈팀과 원정팀 득점

표 13. 분석에 사용한 변수 및 요약통계량

변수	Summary					
	Mean	Min	1Q	Median	3Q	Max
홈팀의 전반전 득점수	0.53	0.00	0.00	0.00	1.00	4.00
원정팀의 전반전 득점수	0.48	0.00	0.00	0.00	1.00	3.00
홈팀의 후반전 득점수	0.78	0.00	0.00	1.00	1.00	4.00
원정팀의 후반전 득점수	0.77	0.00	0.00	0.00	1.00	4.00
홈팀의 전체 득점수	1.31	0.00	0.00	1.00	2.00	7.00
원정팀의 전체 득점수	1.25	0.00	0.00	1.00	2.00	5.00

표 14. 실증분석 결과

$Y_1, Y_2$	모수	MLE	MME
$Y_1$ =홈팀 전반전 득점수 $Y_2$ =원정팀 전반전 득점수	$\lambda_{10}$	0.519	0.530
	$\lambda_{20}$	0.465	0.432
	$\lambda_{00}$	0.017	0.061
	$p_0$	0.008	0.207
$Y_1$ =홈팀 후반전 득점수 $Y_2$ =원정팀 후반전 득점수	$\lambda_{10}$	0.777	0.838
	$\lambda_{20}$	0.770	0.743
	$\lambda_{00}$	0.000	0.000
	$p_0$	0.002	0.085
$Y_1$ =홈팀 전체 득점수 $Y_2$ =원정팀 전체 득점수	$\lambda_{10}$	1.300	1.370
	$\lambda_{20}$	1.233	1.075
	$\lambda_{00}$	0.047	0.140
	$p_0$	0.026	0.125

## 제 4장 결론

본 논문에서는 영과잉 포아송 분포 및 이변량 영과잉 포아송 분포에서의 모수 추정량을 비교해보았다. 모의실험은 특정 확률로 0의 값을 가지고 이외의 확률로 포아송 분포 또는 이변량 포아송 분포를 따르는 자료를 생성하여 영과잉 포아송 분포와 이변량 영과잉 포아송 분포에 적합하였다. 영이 발생할 확률을 포함한 여러 모수를 다양하게 변화시키며 실행하였고 bias와 평균제곱오차를 통해 더 좋은 추정량을 찾고자 하였다.

일변량 영과잉 포아송 분포의 모의실험 시행 결과 대부분의 경우에 최우추정량의 bias가 적률추정량의 bias보다 작고, 최우추정량의 MSE가 적률추정량의 MSE에 비해 작다.  $n$ 이 작고  $\lambda$ 가 작으면  $p$ 가 커질수록  $p$ 의 bias가 증가하는 것을 알 수 있으며  $\lambda$ 이 작을 경우에는  $\lambda$ 가 클 때에 비해  $p$ 의 최우추정량과 적률추정량의 bias값이 작아진다.

이변량 영과잉 포아송 분포의 모의실험의 경우  $\lambda_{20}$ 를 1로 고정하여 시행하였다. 시행 결과 최우추정량의 bias 및 MSE가 적률추정량의 bias 및 MSE보다 작았다.  $p_0$ 의 최우추정량의 MSE는  $n$ 이 증가할수록 감소하였으며,  $p_0$ 의 적률추정량의 MSE는 대체적으로  $p_0$ 가 증가할수록 감소하였고  $n$ 이 증가할수록 감소하였다.  $p_0$ 가 증가할수록  $p_0$ 의 적률추정량의 MSE와 최우추정량의 MSE 차이가 적었다. 또한  $p_0$ 가 증가하고  $\lambda_{00}$ 이 증가할수록  $\lambda_{10}$ 의 적률추정량의 MSE는 증가하였지만 최우추정량의 MSE는  $n$ 이나 다른 모수의 변화에 따라 큰 변화가 없었다.

마지막으로 K리그 경기에서의 득점수 자료를 이변량 포아송 분포를 적용한 실증분석 결과, 전반전 득점수, 후반전 득점수, 전체 득점수 모두 평균 득점수와 비교했을 때 최우추정량이 적률추정량보다 근사한 값을 나타냈다.

즉 영과잉 포아송 분포와 이변량 영과잉 포아송 분포에서 가장 좋은 추정량

은 최우추정량이라는 결과를 나타낸다.

그러나 전체 자료의 수가 많지 않고 0에 대한 확률인  $p_0$ 가 굉장히 클 경우, 다른 모수를 추정하기 위한 자료의 값이 적어져 정확한 모수 추정이 어려울 수 있다는 한계점이 있다. 때문에 영과잉 포아송 분포를 적용하기 위해서는 자료의 수가 충분한 것이 중요하다고 생각되어진다. 또한 추후 연구에서 이변량 영과잉 포아송 분포 이외에도 이변량 영과잉 음이항 분포를 적용함으로써 두 모형의 모수 추정량 비교하는 연구를 진행할 수 있을 것이다.

## 참 고 문 헌

- [1] Beckett, S., Jee, J., Ncube, T., Pompilus, S., Wasington, Q., Singh, A. & Pal, N. (2015). Zero-inflated Poisson (ZIP) distribution: parameter estimation and applications to model data from natural calamities. *Involve*, 7(6), 751–767.
- [2] Cheon, S. Y., Park, N. S., Oh, S. S. & Song, S. H. (2008). A Study on the Number of TV Audience's Channel Zapping. *Journal of the Korean Data Analysis Society*, 10(1), 235–246. (in Korean).
- [3] Cho, J. S. (2014). A Study on Determinant Factor for Number of Stop-out. *Journal of the Korean Data Analysis Society*, 16(1), 201–210. (in Korean).
- [4] Cho, J. S. (2015). Determinant of the Number of Job Experience and Duration to First Job in College. *Journal of the Korean Data Analysis Society*, 17(1), 183–194. (in Korean).
- [5] Choi, J. H., Ko, I. M. & Cheon, S. Y. (2011). *The Korean Journal of Applied Statistics*, 24(3), 485–494. (in Korean).
- [6] Chun, H. J. (2016). The factors of insurance solicitor's turnovers of life insurance using Poisson regression. *Journal of the Korean Data & Information Science Society*, 27(5), 1337–1347. (in Korean).
- [7] Dixon, M. J., Coles, S. C. (1997) Modelling association football scores and inefficiencies in the football betting market. *Applied*

- Statistics*, 46, 265–280.
- [8] Heo, C. K., Kim, S. J. & Lee, W. K. (2016). The Estimation of Scoring Abilities Parameter of KBL Team Using Bivariate Poisson Model. *The Korea Journal of Sports Science*, 25(2), 1085–1095. (in Korean).
- [9] Heo, Y. S., Kim, K. H. (2018). Analysis of K-league data using bivariate Poisson and diagonal inflated model. *Journal of the Korean Data & Information Science Society*, 29(6), 1643–1653. (in Korean).
- [10] Jeong, J. P., Choi, J. H. (2014). Poisson Regression and Negative Binomial regression model fit for traffic accidents. *Journal of the Korean Data Analysis Society*, 16(1), 165–172. (in Korean).
- [11] Karlis, D., Ntzoufras, I. (2000). On modelling soccer data. *Student*, 3, 229–245.
- [12] Karlis, D., Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. *The Statistician*, 52, 281–393.
- [13] Karlis, D., Ntzoufras, I. (2005). Bivariate Poisson and Diagonal Inflated Bivariate Poisson regression models in R. *Journal of Statistical Software*, 14, 1–36.
- [14] Kim, J. Y., Lee, S. K. (2008). A Case Study on the Credit Scoring Model with Zero-Inflated Poisson Regression. *Journal of the Korean Data Analysis Society*, 10(6), 3255–3265. (in Korean).
- [15] Kim, K. M., Um, H. J. (2010). Comparisons of the estimators for

- the Zero-Inflated Poisson distribution. *Journal of The Korean Data Analysis Society*, 12(2), 1113–1124. (in Korean).
- [16] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34, 1-14.
- [17] Lee, J. T. (2014). Prediction of K-league soccer scores using bivariate Poisson distributions. *Journal of the Korean Data & Information Science Society*, 25(6), 1221–1229. (in Korean).

# ABSTRACT

## Comparisons of the Estimators for the Bivariate Zero-Inflated Poisson Distribution

YUNSEO HEO

Department of Statistics

The Graduate School

Sungshin Women's University

If there is a lot of zero in the count data, it is called the zero - inflated count data. In order to analyze this zero -inflated count data, a zero - inflated Poisson distribution, which is a variant of the Poisson distribution, is widely used. If the zero-inflated count data is bivariate, then we use a bivariate zero-inflated Poisson distribution that is an extension of the zero-inflated Poisson distribution.

In this paper, we study the zero-inflated Poisson distribution and bivariate zero-inflated Poisson distribution for the analysis of zero -inflated data and examine the parameter estimator using the maximum likelihood estimation and the moment method estimation. From the results of simulation and empirical studies, we tried to find a better estimator by bias and the mean square error.