IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 22, 2020, accepted October 31, 2020, date of publication November 10, 2020,
date of current version December 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3037213

Odds Ratio Estimation for Small Count in Zero-Inflated Poisson

KOBKUN RAWEESAWAT[1] AND KATECHAN JAMPACHAISRI[2]
[1]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand
[2]Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand

Corresponding author: Katechan Jampachaisri (katechanj@nu.ac.th)

This work was supported by the King Mongkut's University of Technology North Bangkok under Grant KMUTNB-61-NEW-010.

ABSTRACT The odds ratio estimation when observed frequencies are very small usually causes difficulty in calculation. In this paper, we proposed the estimator of odds ratio for small count using Empirical Bayes (EB) in Zero-inflated Poisson distribution, where the hyper-parameters can be estimated via the posterior marginal distribution function. We compare the proposed estimator of odds ratio based on EB in Zero-inflated Poisson distribution with moments method estimator (MME) and modified maximum likelihood estimator (MMLE) using the Estimated Relative Error (ERE) as criterion of comparison. The result of a simulated study indicates that the EB estimator is more efficient than MME and MMLE. For application, the EB odds ratio estimation is implemented in AIDS-related data which the response was the self-reported number of times that respondents having a risky sexual partners, classified by gender. The estimation based EB also yields consistent result as those in simulation, resulting in smallest ERE when compared to MME and MMLE.

INDEX TERMS Empirical Bayes, zero-inflated Poisson distribution, moments method estimation, modified maximum likelihood estimator.

I. INTRODUCTION

The odds ratio is a measurement of the magnitude association between two binary data. Binary data occur very often in clinical research and epidemiological studies referred to as success or failure. Pamela [1] indicated that the odds ratio was one of analytic measures that had frequently appeared in the physical therapy literature. The result of OR expressed concerns over ability to interpret study finding which required understanding about the strengths and weaknesses of data, design and analyses. A number of subjects in each group falling in each category can be summarized in a two-way contingency table. Total number of subjects in each group 1 and group 2 are denoted as n_1 and n_2 respectively, which assumed to be fixed. Let X_1 and X_2 be the number of observations in group 1 and group 2. Let π_1 and π_2 be probabilities of success in group 1 and group 2, respectively. The probability of success can be defined by a number of success or positive count divided by a number of subjects in that group. The odds of success in group 1 is defined to be $\mathrm{odds}_1 = \pi_1/(1-\pi_1)$, similarly for group 2. The usual maximum likelihood estimator (MLE) of odds ratio is defined as

$$\widehat{\mathrm{OR}}_{\mathrm{MLE}} = \frac{\mathrm{odds}_1}{\mathrm{odds}_2} = \frac{\pi_1\big/(1-\pi_1)}{\pi_2\big/(1-\pi_2)}. \tag{1}$$

From this equation, we can see that the odds ratio can be 0 or ∞($\widehat{\mathrm{OR}}_{\mathrm{MLE}} = 0$ if the numerator is 0, and $\widehat{\mathrm{OR}}_{\mathrm{MLE}} = \infty$ if the denominator is 0). If there is a 0 in both the numerator and denominator, then $\widehat{\mathrm{OR}}_{\mathrm{MLE}}$ is undefined. Usually the count data with excessive zero can arise, particularly in studies involving rare events. Zero-inflated Poisson has been used in situations where excess number of zero observations are generated. Cohen [2] and Lambert [3] considered a Zero-inflated Poisson model to handle a dataset which had too many zero observations. Anger and Biswas [4] studied a zero-inflated generalized Poisson model using Bayesian analysis and discussed some appropriate choices of priors and posteriors. John *et al.* [5] considered the application of Empirical Bayes to high consequence of low frequency events and the result revealed that Empirical Bayes was the choice for estimation procedures. Nanjundan and Naika [6] discussed about parameter estimation methods in the Zero-inflated Poisson by comparison of the moments method estimator (MME) with maximum likelihood estimator (MLE). Lu *et al.* [7] proposed the method to analyze count data with excess zero

The associate editor coordinating the review of this manuscript and approving it for publication was Yanbo Chen.

using Bayesian analysis approach in combination with Gibbs sampler and M-H algorithm. Unhapipat *et al.* [8] performed Bayesian predictive inference under the Zero-inflated Poisson model with various types of prior distribution of which the Empirical Bayes method yielded the best overall performance.

In this paper, we focus on rare events which observed count data have a large number of zeros causing difficulty in finding odds ratio estimation. A Zero-inflated Poisson distribution (ZIP) has a more flexible distribution for a dataset with many zeros, so call rare events. It consists of two components: a degenerating distribution for zero count and a Poisson distribution for positive count. The probability mass function of ZIP is expressed as

$$P(X=x)=\begin{cases} p+(1-p)\exp(-\lambda) & \text{if } x=0 \\ (1-p)\dfrac{\exp(-\lambda)\lambda^{x}}{x!} & \text{if } x\neq 0, \end{cases} \tag{2}$$

where $0 \leq p \leq 1$ and $\lambda \geq 0$. The parameter p indicates inflation of zero and the parameter λ is the expected value of Poisson distribution. The new estimation method of odds ratio based on EB in Zero-inflated Poisson distribution is proposed and its result is compared with the moments method estimation (MME) and modified maximum likelihood estimation (MMLE).

II. THE EMPIRICAL BAYES METHOD FOR ODDS RATIO ESTIMATION

Let X_1 and X_2 be random variables distributed as Zero-inflated Poisson, $X_j \sim \mathrm{ZIP}\left(p_j, \lambda_j\right)$, $j = 1, 2$ where p_j and λ_j denote unknown parameters. For convenience, subscript j is omitted for the derivation.

$$P(X\,|p,\lambda)=\begin{cases} p+(1-p)\exp(-\lambda) & \text{if } x=0 \\ (1-p)\dfrac{\exp(-\lambda)\lambda^{x}}{x!} & \text{if } x\neq 0. \end{cases} \tag{3}$$

Assume that the prior distributions for p and λ are $\mathrm{Beta}(a_1, b_1)$ and $\mathrm{Gamma}(a_2, b_2)$ respectively and they are also assumed to be independent, given as

$$f(p)=\frac{1}{B(a_1,b_1)}p^{(a_1-1)}(1-p)^{(b_1-1)},\quad 0<p<1 \tag{4}$$

and

$$f(\lambda)=\frac{b_2^{a_2}}{\Gamma(a_2)}\lambda^{(a_2-1)}e^{-b_2\lambda},\quad \lambda>0 \tag{5}$$

where a_1, b_1, a_2, and b_2 denote unknown hyper-parameters.

Let Y and $n-Y$ be the number of X_i's taking the value zero and non-zero, respectively.

The likelihood function of ZIP (p, λ) can be written as

$$L(p,\lambda\,|\boldsymbol{X})=\prod_{\substack{i=1\\X_i=0}}^{n}\left\{p+(1-p)e^{-\lambda}\right\}\prod_{\substack{i=1\\X_i\neq 0}}^{n}\left\{(1-p)\frac{e^{-\lambda}\lambda^{X_i}}{X_i!}\right\}$$

$$=\left\{p+(1-p)e^{-\lambda}\right\}^{Y}\left(\prod_{\substack{i=1\\X_i\neq 0}}^{n}\frac{(1-p)e^{-\lambda}\lambda^{X_i}}{X_i!}\right)$$

$$=\left\{p+(1-p)e^{-\lambda}\right\}^{Y}(1-p)^{n-Y}\cdot e^{-(n-Y)\lambda}\lambda^{n-Y}\frac{1}{\prod_{\substack{i=1\\X_i\neq 0}}^{n}X_i!}.$$

The joint pdf of (X, p, λ) is

$$\begin{aligned} f(X,p,\lambda) &= L(p,\lambda\,|\boldsymbol{X})f(p)f(\lambda) \\ &= \left[\left(p+(1-p)e^{-\lambda}\right)^{Y}\left(\frac{(1-p)^{n-Y}e^{-(n-Y)\lambda}\lambda^{n-Y}}{\prod_{\substack{i=1\\X_i\neq 0}}^{n}X_i!}\right)\right] \\ &\quad\cdot\left[\frac{1}{B(a_1,b_1)}p^{(a_1-1)}(1-p)^{(b_1-1)}\right] \\ &\quad\cdot\left[\frac{b_2^{a_2}}{\Gamma(a_2)}\lambda^{(a_2-1)}e^{-b_2\lambda}\right]. \end{aligned}$$

The estimation of hyper-parameters can be obtained from the posterior marginal distribution function as follow,

$$\begin{aligned} m(X) &= \int_0^{\infty}\int_0^{1}\frac{b_2^{a_2}}{B(a_1,b_1)\Gamma(a_2)}\cdot\frac{1}{\prod_{\substack{i=1\\X_i\neq 0}}^{n}X_i!} \\ &\quad\cdot\sum_{k=0}^{Y}\binom{Y}{k}p^{(k+a_1-1)}(1-p)^{(n-k+b_1-1)} \\ &\quad\cdot e^{-(n-k+b_2)\lambda}\lambda^{(n-Y+a_2-1)}dpd\lambda \\ &= \frac{b_2^{a_2}}{B(a_1,b_1)\Gamma(a_2)}\cdot\frac{1}{\prod_{\substack{i=1\\X_i\neq 0}}^{n}X_i!} \\ &\quad\cdot\sum_{k=0}^{Y}\binom{Y}{k}B(k+a_1,n-k+b_1) \\ &\quad\cdot\frac{\Gamma(n-Y+a_2)}{(n-k+b_2)^{(n\bar{X}+a_2)}}. \end{aligned} \tag{6}$$

The parameters a_1, b_1, a_2, and b_2 are initially estimated by the method of moments [9], [10], obtained as

$$a_1=\bar{x}\left(\frac{\bar{x}(1-\bar{x})}{s^2}-1\right), \tag{7}$$

$$b_1=(1-\bar{x})\left(\frac{\bar{x}(1-\bar{x})}{s^2}-1\right), \tag{8}$$

$$a_2=\frac{\bar{x}^2}{s^2}, \tag{9}$$

$$b_2=\frac{\bar{x}}{s^2}, \tag{10}$$

where $\bar{x}$ and s^2 are the sample mean and variance, respectively.

The joint posterior distribution function of p and λ for a given dataset X is

$$f(p,\lambda|X) = \frac{\frac{b_2^{a_2}}{B(a_1,b_1)\Gamma(a_2)} \cdot \frac{1}{\prod\limits_{\substack{i=1\\X_i\neq 0}}^{n} X_i!}}{\frac{b_2^{a_2}}{B(a_1,b_1)\Gamma(a_2)} \cdot \frac{1}{\prod\limits_{\substack{i=1\\X_i\neq 0}}^{n} X_i!}} \cdot \frac{\sum\limits_{k=0}^{Y}\binom{Y}{k} p^{(k+a_1-1)}(1-p)^{(n-k+b_1-1)}}{\sum\limits_{k=0}^{Y}\binom{Y}{k} B(k+a_1, n-k+b_1)} \cdot \frac{e^{-(n-k+b_2)\lambda}\lambda^{(n-Y+a_2-1)}}{\frac{\Gamma(n-Y+a_2)}{(n-k+b_2)^{(n-Y+a_2)}}}. \tag{11}$$

The marginal posterior pdf of p and λ can be derived as follows

$$\begin{aligned} f(p|X) &= \int_0^\infty f(p,\lambda|X)d\lambda \\ &= \frac{\sum\limits_{k=0}^{Y}\binom{Y}{k} p^{(k+a_1-1)}(1-p)^{(n-k+b_1-1)}}{\sum\limits_{k=0}^{Y}\binom{Y}{k} B(k+a_1, n-k+b_1)} \cdot \frac{\frac{\Gamma(n-Y+a_2)}{(n-k+b_2)^{(n-Y+a_2)}}}{\frac{\Gamma(n-Y+a_2)}{(n-k+b_2)^{(n-Y+a_2)}}} \end{aligned} \tag{12}$$

and

$$\begin{aligned} f(\lambda|X) &= \int_0^\infty f(p,\lambda|X)dp \\ &= \frac{\sum\limits_{k=0}^{Y}\binom{Y}{k} B(k+a_1, n-k+b_1)}{\sum\limits_{k=0}^{Y}\binom{Y}{k} B(k+a_1, n-k+b_1)} \cdot \frac{e^{-(n-k+b_2)\lambda}\lambda^{(n\bar{X}+a_2-1)}}{\frac{\Gamma(n\bar{X}+a_2)}{(n-k+b_2)^{(n\bar{X}+a_2)}}}. \end{aligned} \tag{13}$$

Thus, the estimator of π_1 is calculated as

$$\hat{\pi}_1 = (1-\hat{p}_1)\left(1-e^{-\hat{\lambda}_1}\right), \tag{14}$$

and the estimator of π_2 is

$$\hat{\pi}_2 = (1-\hat{p}_2)\left(1-e^{-\hat{\lambda}_2}\right). \tag{15}$$

The EB for odds ratio estimation can be obtained as follow

$$\widehat{OR}_{EB} = \frac{\hat{\pi}_1/(1-\hat{\pi}_1)}{\hat{\pi}_2/(1-\hat{\pi}_2)}. \tag{16}$$

III. MOMENTS METHOD ESTIMATION FOR ODDS RATIO ESTIMATION

The two parameters in Zero-inflated Poisson distribution, p and λ can be estimated by the method of moments [11], [12] as

$$\hat{p} = \frac{s^2-\bar{x}}{s^2+\bar{x}^2-\bar{x}}, \tag{17}$$

and

$$\hat{\lambda} = \frac{s^2+\bar{x}^2-\bar{x}}{\bar{x}} \tag{18}$$

where $\bar{x}$ and s^2 are the sample mean and variance, respectively.

However, the MME of parameter p undesirable property with negative value. When $\bar{x} > s^2$, $\hat{p}$ can become negative, while the actual parameter p is always between 0 and 1.

Beckett *et al.* [13] modified the MME by truncating $\hat{p}$ at zero and $\hat{\lambda}$ at $\bar{x}$ when $\bar{x} > s^2$.

Let $\hat{\pi}_1'$ and $\hat{\pi}_2'$ be estimators of π_1 and π_2 respectively, where

$$\hat{\pi}_1' = (1-\hat{p}_1)\left(1-e^{-\hat{\lambda}_1}\right), \tag{19}$$

and

$$\hat{\pi}_2' = (1-\hat{p}_2)\left(1-e^{-\hat{\lambda}_2}\right). \tag{20}$$

Then, the MME of odds ratio is obtained as

$$\widehat{OR}_{MME} = \frac{\hat{\pi}_1'/(1-\hat{\pi}_1')}{\hat{\pi}_2'/(1-\hat{\pi}_2')}. \tag{21}$$

IV. MODIFIED MAXIMUM LIKELIHOOD ESTIMATOR FOR ODDS RATIO ESTIMATION

The modified maximum likelihood estimator (MMLE) is presented by Haldane [14] and Gart and Zweifel [15] to solve the problem of zero cell counts in denominator and numerator of odds ratio estimation. They also suggested to add a correction term 0.5 to each cell, when having zero counts in both groups, which gives the modified maximum likelihood estimator (MMLE) as

$$\widehat{OR}_{MMLE} = \frac{(Y_1+0.5)(n_2-Y_2+0.5)}{(Y_2+0.5)(n_1-Y_1+0.5)}. \tag{22}$$

V. SIMULATION STUDY FOR EB, MME, AND MMLE METHOD

Simulation study with randomly generated data using R program (version 4.0.2) [16] is performed to assess performance of the proposed method in comparison with MME and MMLE method **s**. Data in both groups are generated as independent zero inflated Poisson distribution with the inflation of zero (p_1, p_2) are (0.5, 0.5) and (0.7, 0.5) and the expected value λ_i, where i = 1,2 are 0.1, 0.3, 0.5 and 1.0 for sample sizes $(n_1, n_2) = (10, 10)$ and (10, 30). Each situation is repeated 5,000 times after 1,000 burn-ins. The efficiency

TABLE 1. The estimated value of odds ratio for $p_1 = 0.5$, $p_2 = 0.5$ and $(n_1, n_2) = (10, 10)$.

λ_1	λ_2	OR_{MLE}	$\widehat{OR}_{EB}$	$\widehat{OR}_{MME}$	$\widehat{OR}_{MMLE}$
0.1	0.1	1.0000	0.8874	0.9932	0.9967
0.1	0.3	0.3356	0.3289	0.6555	0.8213
0.1	0.5	0.2040	0.2815	0.4700	0.6941
0.1	1.0	0.1081	0.1212	0.2469	0.4808
0.3	0.1	2.9802	2.9620	1.4928	1.2097
0.3	0.3	1.0000	1.0979	0.9853	0.9968
0.3	0.5	0.6079	0.9396	0.7065	0.8424
0.3	1.0	0.3222	0.4046	0.3711	0.5835
0.5	0.1	4.9025	3.2175	2.1147	1.4362
0.5	0.3	1.6450	1.1926	1.3957	1.1835
0.5	0.5	1.0000	1.0207	1.0007	1.0002
0.5	1.0	0.5300	0.4395	0.5257	0.6928
1.0	0.1	9.2500	5.5437	4.7873	2.0904
1.0	0.3	3.1039	2.0550	3.1597	1.7225
1.0	0.5	1.8868	1.7586	2.2656	1.4557
1.0	1.0	1.0000	0.7572	1.1900	1.0083

TABLE 2. The estimated value of odds ratio for $p_1 = 0.7$, $p_2 = 0.5$ and $(n_1, n_2) = (10, 10)$.

λ_1	λ_2	OR_{MLE}	$\widehat{OR}_{EB}$	$\widehat{OR}_{MME}$	$\widehat{OR}_{MMLE}$
0.1	0.1	0.6000	0.4501	0.8905	0.9550
0.1	0.3	0.2203	0.1423	0.5806	0.7849
0.1	0.5	0.1451	0.1048	0.4920	0.6592
0.1	1.0	0.0903	0.0491	0.1836	0.4183
0.3	0.1	1.6341	1.8789	1.1814	1.0837
0.3	0.3	0.6000	0.5942	0.7703	0.8906
0.3	0.5	0.3952	0.4373	0.6526	0.7480
0.3	1.0	0.2460	0.2048	0.2435	0.4747
0.5	0.1	2.4808	2.6254	1.4753	1.2196
0.5	0.3	0.9109	0.8303	0.9619	1.0023
0.5	0.5	0.6000	0.6111	0.8150	0.8418
0.5	1.0	0.3735	0.2862	0.3041	0.5342
1.0	0.1	3.9855	5.5501	2.1640	1.5844
1.0	0.3	1.4633	1.7552	1.4109	1.3021
1.0	0.5	0.9639	1.2918	1.1955	1.0936
1.0	1.0	0.6000	0.6050	0.4461	0.6940

of estimators is evaluated using the percentage of Estimated Relative Error (ERE), defined as

$$ERE = \left[\frac{|OR - \widehat{OR}_i|}{OR} \right] \times 100, \tag{23}$$

where OR denotes the usual maximum likelihood estimator of odds ratio and $\widehat{OR}_i$, where $i = 1, 2, 3$ denote the estimates

TABLE 3. The estimated value of odds ratio for $p_1 = 0.5$, $p_2 = 0.5$ and $(n_1, n_2) = (10, 30)$.

λ_1	λ_2	OR_{MLE}	$\widehat{OR}_{EB}$	$\widehat{OR}_{MME}$	$\widehat{OR}_{MMLE}$
0.1	0.1	1.0000	0.9104	2.0919	2.9369
0.1	0.3	0.3356	0.3413	0.8685	2.4571
0.1	0.5	0.2040	0.1660	0.5552	2.1240
0.1	1.0	0.1081	0.0928	0.2834	1.5572
0.3	0.1	2.9802	3.0389	3.1443	3.5644
0.3	0.3	1.0000	1.1393	1.3054	2.9822
0.3	0.5	0.6079	0.5540	0.8346	2.5779
0.3	1.0	0.3222	0.3097	0.4260	1.8899
0.5	0.1	4.9025	3.3010	4.4540	2.9369
0.5	0.3	1.6450	1.2376	1.8492	2.4572
0.5	0.5	1.0000	0.6017	1.1822	2.1241
0.5	1.0	0.5300	0.3364	0.6034	1.5572
1.0	0.1	9.2500	5.6876	10.0834	6.1595
1.0	0.3	3.1039	2.1323	4.1863	5.1533
1.0	0.5	1.8868	1.0368	2.6764	4.4547
1.0	1.0	1.0000	0.5797	1.3661	3.2658

TABLE 4. The estimated value of odds ratio for $p_1 = 0.7$, $p_2 = 0.5$ and $(n_1, n_2) = (10, 30)$.

λ_1	λ_2	OR_{MLE}	$\widehat{OR}_{EB}$	$\widehat{OR}_{MME}$	$\widehat{OR}_{MMLE}$
0.1	0.1	0.6000	0.5218	1.8164	2.8002
0.1	0.3	0.2203	0.1904	0.7733	2.3482
0.1	0.5	0.1451	0.1165	0.4865	2.0258
0.1	1.0	0.0903	0.0583	0.2484	1.4823
0.3	0.1	1.6341	2.1784	2.4097	3.1774
0.3	0.3	0.6000	0.7950	1.0258	2.6645
0.3	0.5	0.3952	0.4863	0.6454	2.2987
0.3	1.0	0.2460	0.2434	0.3295	1.6820
0.5	0.1	2.4808	3.0438	3.0092	3.5760
0.5	0.3	0.9109	1.1108	1.2811	2.9987
0.5	0.5	0.6000	0.6794	0.8059	2.5870
0.5	1.0	0.3735	0.3401	0.4114	1.8930
1.0	0.1	3.9855	6.4347	4.4140	4.6456
1.0	0.3	1.4633	2.3483	1.8791	3.8957
1.0	0.5	0.9639	1.4363	1.1822	3.3608
1.0	1.0	0.6000	0.7190	0.6035	2.4592

of odds ratio using EB, MME, and MMLE, respectively. The simulation results of odds ratio estimation for sample sizes $(n_1, n_2) = (10,10)$ and $(10,30)$ are given in Table 1-4 and their performances are compared using ERE as illustrated in Table 5-8. The ERE of three estimation methods also display in Fig.1-2 for the case $(n_1, n_2) = (10,30)$, similar to the case $(n_1, n_2) = (10,10)$ (but not shown here). Based

TABLE 5. The percentage of the Estimated Relative Error of OR estimation for $p_1 = 0.5$, $p_2 = 0.5$ and $(n_1, n_2) = (10, 10)$.

λ_1	λ_2	ERE_{EB}	ERE_{MME}	ERE_{MMLE}
0.1	0.1	11.2600	0.6800	0.3300*
0.1	0.3	1.9818*	95.3509	144.7623
0.1	0.5	38.0041*	130.4154	240.2794
0.1	1.0	12.1105*	128.3836	344.7421
0.3	0.1	0.6101*	49.9091	59.4085
0.3	0.3	9.7900	1.4700	0.3200*
0.3	0.5	54.5660	16.2206*	38.5764
0.3	1.0	25.5819	15.1840*	81.1098
0.5	0.1	34.3696*	56.8645	70.7045
0.5	0.3	27.5023	15.1560*	28.0555
0.5	0.5	2.0700	0.0700	0.0200*
0.5	1.0	17.0743	0.8099*	30.7188
1.0	0.1	40.0684*	48.2457	77.4012
1.0	0.3	33.7920	1.7993*	44.5044
1.0	0.5	6.7955*	20.0751	22.8490
1.0	1.0	24.2800	19.0000	0.8300*

Note: *is the smallest ERE in each row.

TABLE 6. The percentage of the Estimated Relative Error of OR estimation for $p_1 = 0.7$, $p_2 = 0.5$ and $(n_1, n_2) = (10, 10)$.

λ_1	λ_2	ERE_{EB}	ERE_{MME}	ERE_{MMLE}
0.1	0.1	24.9896*	48.4212	59.1690
0.1	0.3	35.3912*	163.5634	256.2832
0.1	0.5	27.8118*	239.0252	354.2475
0.1	1.0	45.6881*	103.2184	363.1366
0.3	0.1	14.9796*	27.7071	33.6860
0.3	0.3	0.9646*	28.3764	48.4370
0.3	0.5	10.6536*	65.1323	89.2515
0.3	1.0	16.7481	1.0165*	92.9549
0.5	0.1	5.8273*	40.5313	50.8392
0.5	0.3	8.8477	5.6033*	10.0413
0.5	0.5	1.8456*	35.8390	40.2985
0.5	1.0	23.3749	18.5754*	43.0440
1.0	0.1	39.2568*	45.7032	60.2464
1.0	0.3	19.9461	3.5807*	11.0157
1.0	0.5	34.0174	24.0254	13.4516*
1.0	1.0	0.8300*	25.6567	15.6717

Note: *is the smallest ERE in each row.

TABLE 7. The percentage of the Estimated Relative Error of OR estimation for $p_1 = 0.5$, $p_2 = 0.5$ and $(n_1, n_2) = (10, 30)$.

λ_1	λ_2	ERE_{EB}	ERE_{MME}	ERE_{MMLE}
0.1	0.1	8.9600*	109.1900	193.6900
0.1	0.3	1.7136*	158.8288	632.2604
0.1	0.5	18.6193*	172.1843	941.2813
0.1	1.0	14.1596*	162.1462	1340.4168
0.3	0.1	1.9703*	5.5070	19.6034
0.3	0.3	13.9300*	30.5400	198.2200
0.3	0.5	8.8660*	37.2933	324.0694
0.3	1.0	3.8737*	32.2241	486.5971
0.5	0.1	32.6664	9.1476*	40.0933
0.5	0.3	24.7668	12.4121*	49.3722
0.5	0.5	39.8300	18.2200*	112.4100
0.5	1.0	36.5274	13.8507*	193.8154
1.0	0.1	38.5127	9.0092*	33.4111
1.0	0.3	31.3015*	34.8743	66.0292
1.0	0.5	45.0504	41.8472*	136.0958
1.0	1.0	42.0300	36.6100*	226.5800

Note: *is the smallest ERE in each row.

TABLE 8. The percentage of the Estimated Relative Error of OR estimation for $p_1 = 0.7$, $p_2 = 0.5$ and $(n_1, n_2) = (10, 30)$.

λ_1	λ_2	ERE_{EB}	ERE_{MME}	ERE_{MMLE}
0.1	0.1	13.0338*	202.7387	366.7034
0.1	0.3	13.5597*	251.0080	965.9031
0.1	0.5	19.7357*	235.2438	1296.0069
0.1	1.0	35.4491*	174.9503	1541.0726
0.3	0.1	33.3060*	47.4579	94.4409
0.3	0.3	32.4998*	70.9689	344.0834
0.3	0.5	23.0330*	63.2904	481.6133
0.3	1.0	1.0532*	33.9227	583.7141
0.5	0.1	22.6949	21.2999*	44.1456
0.5	0.3	21.9529*	40.6402	229.2140
0.5	0.5	13.2396*	34.3239	331.1696
0.5	1.0	8.9294*	10.1658	406.8604
1.0	0.1	61.4527	10.7507*	16.5626
1.0	0.3	60.4763	28.4091*	166.2171
1.0	0.5	49.0106	22.6420*	248.6629
1.0	1.0	19.8387	0.5849*	309.8698

Note: *is the smallest ERE in each row.

on the percentage of ERE, the proposed estimator mostly outperform MME and MMLE. For all cases under study, it is also found that EB method yields the smallest ERE accounted for 59.38% of the time while MME and MMLE methods respectively provide smallest ERE only 32.81% and 7.81% of the time. The results indicated that EB method to incorporating prior knowledge about parameters combined with information in the observed data to produce the posterior distribution are preferable to the odds ratio estimation for small cell count.

TABLE 9. AIDS-related data.

Y	Gender	
	Male	Female
0	102	103
1	5	6
2	8	4
3	2	2
4	1	-
5	4	1
6	1	-
7	-	-
10	-	-
12	-	-
15	1	-
20	-	-
30	-	-
37	1	-
50	-	50

TABLE 10. True odds ratio and their estimates using EB, MMLE and MME methods and their corresponding ERE.

Study		Methods			
		MLE	EB	MME	MMLE
AIDS-related	OR	1.6589	1.6480	2.9099	1.6365
	ERE		0.0066	0.7541	0.0135

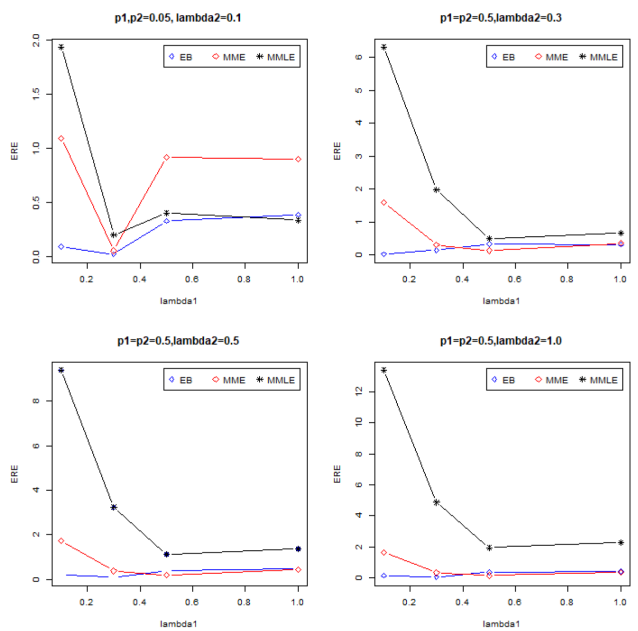

FIGURE 1. A comparison of the percentage of ERE of odds ratio estimation using EB, MME and MMLE when $p_1 = p_2 = 0.5$ and $(n_1, n_2) = (10, 30)$.

VI. ILLUSTRATIVE EXAMPLE USING REAL DATA

Our example is taken from the set of AIDS-related data discussed in Heilbron [17]. The response variable in this study was the self-reported number of times that the respondents had anal intercourse with opposite sex partners during the study period classified by to dichotomous explanatory variables; gender (male, female) and having a risky main sexual partner. Table 9 gives the number of times (Y) and gender. The estimate of odds ratio between male and female respectively are

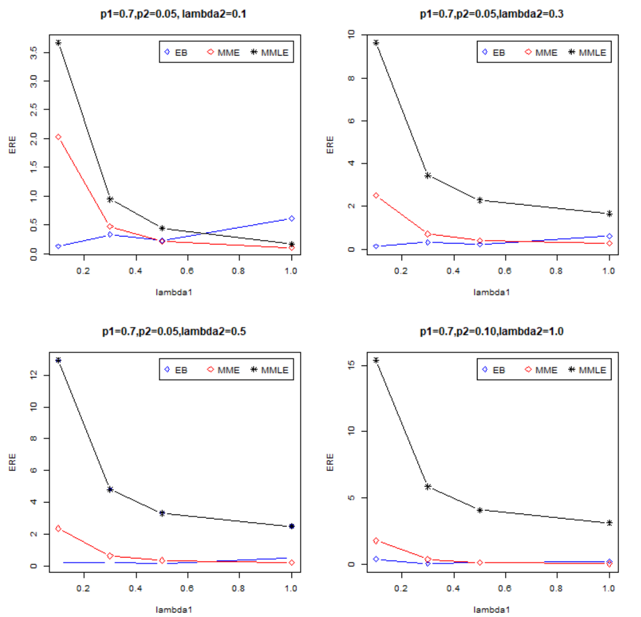

FIGURE 2. A comparison of the percentage of ERE of odds ratio estimation using EB, MME and MMLE when $p_1 = 0.7$, $p_2 = 0.5$ and $(n_1, n_2) = (10, 30)$.

$$n_1 = 125, \ \hat{p}_1 = 0.816, \ \hat{\lambda}_1 = 0.870$$

and

$$n_2 = 117, \ \hat{p}_2 = 0.880, \ \hat{\lambda}_2 = 0.640.$$

The EB method yields the ERE with 0.0066 while those using MME and MMLE methods result in ERE with 0.7541 and 0.0135, respectively as shown in Table 10. The results in this example revealed that we can estimate the odds ratio without disturbing the original data based on the EB method when observed count data have a large number of zeros.

VII. CONCLUSION

This paper presents the odds ratio estimation for small count in Zero-inflated Poisson distribution. The results obtained from both simulated data and actual data indicate that the proposed method perform rather well. The EB estimator of odds ratio is more efficient than the MMLE and MME. In addition, more robust estimation methods have been proposed by several authors, as illustrated by Chen *et al.* [18] which probably draw attention for future research.

REFERENCES

[1] P. K. Levangie, "Application and interpretation of simple odds ratios in physical therapy-related research," *J. Orthopaedic Sports Phys. Therapy*, vol. 31, no. 9, pp. 496–503, Sep. 2001.

[2] A. C. Cohen, "Estimation in mixtures of discrete distributions," in *Proc. Int. Symp. Discrete Distrib.*, Montreal, QC, Canada, 1963, pp. 373–378.

[3] D. Lambert, "Zero Inflated Poisson Regression with an application to defects in manufacturing," *Technometrics*, vol. 34, pp. 1–14, Dec. 1992.

[4] J.-F. Angers and A. Biswas, "A Bayesian analysis of zero-inflated generalized Poisson model," *Comput. Statist. Data Anal.*, vol. 42, nos. 1–2, pp. 37–46, Feb. 2003.

[5] J. Quigley, T. Bedford, and L. Walls, "Estimating rate of occurrence of rare events with empirical Bayes: A railway application," *Rel. Eng. Syst. Saf.*, vol. 92, no. 5, pp. 619–627, May 2007.

[6] G. Nanjundan and T. R. Naika, "Asymptotic comparison of method of moments estimators and maximum likelihood estimators of parameters in zero-inflated Poisson model," *Appl. Math.*, vol. 03, no. 06, pp. 610–616, 2012.

[7] L. Lu, Y. Fu, P. Chu, and X. Zhang, "A Bayesian analysis of zero-inflated count data: An application to youth fitness survey," in *Proc. 10th Int. Conf. Comput. Intell. Secur.*, Nov. 2014, pp. 699–703.

[8] S. Unhapipat, M. M. Tiensuwan, and N. Pal, "Bayesian Predictive Inference for Zero-Inflated Poisson (ZIP) Distribution with Applications," *Amer. J. Math. Manage. Sci.*, vol. 37, pp. 1–14, 2017.

[9] C. Owen, "Parameter estimation for the beta distribution," M.S. Thesis, Dept. Statist., Brigham Young Univ., Provo, UT, USA, 2008.

[10] T.-Y. Hwang and P.-H. Huang, "On new moment estimation of parameters of the gamma distribution using its characterization," *Ann. Inst. Statist. Math.*, vol. 54, no. 4, pp. 840–847, 2002.

[11] N. Johnson, S. Kotz, and A. W. Kemp, *Univariate Discrete Distribution*, 2nd ed. New York, NY, USA: Macmillan, 1992.

[12] C.-E. Lee and S. Kim, "Applicability of zero-inflated models to fit the torrential rainfall count data with extra zeros in South Korea," *Water*, vol. 9, no. 2, p. 123, Feb. 2017.

[13] S. Beckett, J. Jee, T. Ncube, S. Pompilus, Q. Washington, A. Singh, and N. Pal, "Zero-inflated Poisson (ZIP) distribution: Parameter estimation and applications to model data from natural calamities," *Involve J. Math.*, vol. 7, no. 6, pp. 751–767, Oct. 2014.

[14] B. J. B. S. Haldane, "The estimation and significance of the logarithm of a ratio of frequencies," *Ann. Hum. Genet.*, vol. 20, no. 4, pp. 309–311, May 1956.

[15] J. J. Gart and J. R. Zweifel, "On the bias of various estimators of the logit and its variance with application to quantal bioassay," *Biometrika*, vol. 54, nos. 1–2, pp. 181–187, 1967.

[16] R. Core. (2014). *A language and Environment for Statistical Computing*. [Online]. Available: http://www.R-project.org/

[17] D. C. Heilbron, "Zero-altered and other regression models for count data with added zeros," *Biometrical J.*, vol. 36, no. 5, pp. 531–547, 1994.

[18] Y. Chen, J. Ma, P. Zhang, F. Liu, and S. Mei, "Robust state estimator based on maximum exponential absolute value," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1537–1544, Jul. 2017.