

## Research Article

# Odds Ratios Estimation of Rare Event in Binomial Distribution

**Kobkun Raweesawat,<sup>1</sup> Yupaporn Areepong,<sup>1</sup>  
 Katechan Jampachaisri,<sup>2</sup> and Saowanit Sukparungsee<sup>1</sup>**

<sup>1</sup>*Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand*

<sup>2</sup>*Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand*

Correspondence should be addressed to Saowanit Sukparungsee; saowanit.s@sci.kmutnb.ac.th

Received 5 July 2016; Revised 12 September 2016; Accepted 15 September 2016

Academic Editor: Shein-chung Chow

Copyright © 2016 Kobkun Raweesawat et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We introduce the new estimator of odds ratios in rare events using Empirical Bayes method in two independent binomial distributions. We compare the proposed estimates of odds ratios with two estimators, modified maximum likelihood estimator (MMLE) and modified median unbiased estimator (MMUE), using the Estimated Relative Error (ERE) as a criterion of comparison. It is found that the new estimator is more efficient when compared to the other methods.

## 1. Introduction

The odds ratio is a measure of association between two independent groups on a categorical response with two possible outcomes, success and failure. The two independent groups can be two treatment groups or treatment and control groups. The odds ratio is widely used in many fields of medical and social science research. It is most commonly used in epidemiology to express the results of some clinical trials, such as in case-control studies.

A number of subjects in each group falling in each category can be summarized in a two-way contingency table. Total numbers of subjects in group 1 and group 2 are  $n_1$  and  $n_2$ , which are assumed to be fixed. Numbers of successes in group 1 and group 2 are  $Y_1$  and  $Y_2$ , which are considered as independent binomial random variables. Let  $\pi_1$  and  $\pi_2$  be probabilities of success in group 1 and group 2, respectively. The odds of success in group 1 are defined to be  $\text{odds}_1 = \pi_1/(1-\pi_1)$ , similar to group 2. The usual maximum likelihood estimator of odds ratio is defined as

$$\begin{aligned} \widehat{\text{OR}}_{\text{MLE}} &= \frac{\text{odds}_1}{\text{odds}_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = \frac{Y_1/(n_1 - Y_1)}{Y_2/(n_2 - Y_2)} \\ &= \frac{Y_1(n_2 - Y_2)}{Y_2(n_1 - Y_1)}. \end{aligned} \quad (1)$$

Odds ratio is nonnegative real value. When successes are similar in both groups, the odds ratio is equal to 1, meaning that groups are independent of response. When the odds of a positive response are higher in group 1 than in group 2, the odds ratio is greater than 1 and vice versa for the value less than 1. The father of odds ratio from 1 in a given direction represents stronger association. In addition, its sampling distribution is highly skewed. Sample natural logarithm of odds ratio, which is less skewed, is often utilized for inference. However, odds ratio can be zero (if zero cell count appears in numerator of (1)) or infinity (if zero cell count is in denominator of (1)) or undefined (if there are zero cell counts in both the numerator and denominator of (1)). Haldane [1] and Gart and Zweifel [2] suggested to add a correction term 0.5 to each cell, when having zero cell count, which gives the modified maximum likelihood estimator (MMLE) as

$$\widehat{\text{OR}}_{\text{MMLE}} = \frac{(Y_1 + 0.5)(n_2 - Y_2 + 0.5)}{(Y_2 + 0.5)(n_1 - Y_1 + 0.5)}. \quad (2)$$

Even though  $\widehat{\text{OR}}_{\text{MMLE}}$  still laid between 0 and infinity, some investigators discouraged adding 0.5 to each cell because of the appearance of adding “fake data”; see Bishop et al. [3] and Agresti and Yang [4]. Among controversy, several

similar alternatives to this modified maximum likelihood estimator have been proposed. Hirji et al. [5] proposed the median unbiased estimator (MUE) of the odds ratio, obtained from the conditional noncentral hypergeometric distribution. However, the median unbiased estimator of the odds ratio still caused a problem when  $Y_1 = n_1$  and  $Y_2 = n_2$  or  $Y_1 = Y_2 = 0$ , and then the MUE was undefined. Parzen et al. [6] proposed an estimator of the odds ratio based on MUE called the modified median unbiased estimator (MMUE) of which the estimated probability of success was always in the interval  $(0, 1)$ , even if there were 0 or  $n$  successes in each group. Consequently, the estimated odds ratio always laid between 0 and infinity. Additionally, this method performed well with respect to bias in small sample and was an alternative to adding “fake data.”

In this paper, we focus on “rare events” which occasionally observed zero or small counts of interesting events which happened within a given time period or a given sample such as natural disasters or some diseases. As aforementioned, rare events caused difficulty in estimation of odds ratio due to the occurrence of zeros or small observed counts in numerator or in denominator or in both, resulting in the large standard error and therefore less precise confidence interval. Only a rough estimate of the odds ratio is thus obtained. Researches involving association between categorical variables in contingency table have long been studied, using both classical and Bayesian approaches. Good [7] studied association factor, at early stage, in large contingency table with small entries, assuming log-normal and Pearson type III distribution. The author also mentioned that these assumptions may be less accurate but easy to handle. Fisher [8] estimated the odds ratio based on hypergeometric distribution utilizing exact method in a  $2 \times 2$  table. Thomas and Gart [9] constructed a table for 95% confidence limits of differences and ratio of two proportions, including odds ratio and one-tailed  $p$  value for Fisher-Irwin Exact test in various types of  $2 \times 2$  table. Altham [10] studied association and exact  $p$  value in a  $2 \times 2$  contingency table based on the cumulative posterior probabilities which was not easy to extract. Nurminen and Mutanen [11] proposed Bayesian approach for the estimation of difference between two proportions, risk ratio and odds ratio, using independent beta prior and provided integral expressions for the cumulative posterior distribution. They also applied the proposed method to real data regarding malignant lymphoma and colon cancer cases exposed to phenoxo acids and chlorophenols in agriculture. Nouri et al. [12] presented the estimation of the odds ratio in  $2 \times 2 \times J$  tables when exposure was misclassified. They compared the matrix and inverse matrix methods to the MLE method using simulation study and found that the inverse matrix method having a closed form was more efficient than the matrix method.

As previously mentioned, the estimates of association measure in two-way contingency table can be carried out based on classical and Bayesian approaches. The exact distribution using classical approach is, however, rather difficult for mathematical tractability. In Bayesian approach, where prior belief is incorporated into derivation of posterior density, the hyperparameters, characterizing the prior density, are

often unknown to researchers and need to be assessed irrespective of current data. However, controversy still exists. Alternatively, the estimation of hyperparameters is plausibly carried out with the notion of Empirical Bayes method using current data to estimate the unknown hyperparameters, contrary to Bayesian approach. As a consequence, we focus on the utilization of Empirical Bayes method to estimate the odds ratio in a two-way contingency table, focusing on small proportions of success. Our purposed estimation tends to outperform the traditional estimator, MMLE, and MMUE without interference in the original data.

The rest of this paper is organized in the following sequence. In the next section, we discuss the median unbiased estimator. The third section describes the odds ratio estimation using EB method. The fourth section illustrates simulated results, and the efficiency of EB is compared with MMLE and MUE. The fifth section displays the application of our method to real data. Our conclusion is drawn in the final section.

## 2. The Modified Median Unbiased Estimator of Odds Ratio

Parzen et al. [6] suggested the modified median unbiased estimator (MMUE) in two independent binomial distributions. Let  $\tilde{p}$  be the estimator of success probability which satisfies

$$\begin{aligned} P(\tilde{p} \leq p) &\geq 0.5, \\ P(\tilde{p} \geq p) &\geq 0.5. \end{aligned} \quad (3)$$

To obtain  $\tilde{p}_t$ , they use the binomial distribution,  $Y_t \sim B(n_t, p_t)$ , where  $Y_t$  denotes random variable representing success in the  $t$ th group ( $t = 1, 2$ ). Let  $y_t$  be the observed value of  $Y_t$ .

$$P(Y_t = y_t | p_t) = \binom{n_t}{y_t} p_t^{y_t} (1 - p_t)^{n_t - y_t}. \quad (4)$$

The MMUE can be computed from the distribution of sufficient statistics for binomial data.

Compute the values of  $p_t^L$  and  $p_t^U$  to be those value of  $p_t$  for which

$$\begin{aligned} P(Y_t \geq y_t | p_t = \tilde{p}_t^L) &\geq 0.5, \\ P(Y_t \leq y_t | p_t = \tilde{p}_t^U) &\geq 0.5, \end{aligned} \quad (5)$$

where  $p_t^L$  and  $p_t^U$  are the smallest and largest values of  $p_t$ , respectively. Then, the MMUE is defined as

$$\tilde{p}_t = \frac{(\tilde{p}_t^L + \tilde{p}_t^U)}{2}. \quad (6)$$

When  $0 < Y_t \leq n_t$ , we can find values of  $\tilde{p}_t^L$  and  $\tilde{p}_t^U$  which satisfy

$$P(Y_t \geq y_t | p_t = \tilde{p}_t^L) = P(Y_t \leq y_t | p_t = \tilde{p}_t^U) = 0.5. \quad (7)$$

Then, solve from

$$\begin{aligned} 0.5 &= P(Y_t \geq y_t | p_t = \tilde{p}_t^L) \\ &= \sum_{i=y_t}^{n_t} \binom{n_t}{i} (\tilde{p}_t^L)^i (1 - \tilde{p}_t^L)^{n_t-i}, \end{aligned} \quad (8)$$

and solve  $\tilde{p}_t^U$  from

$$\begin{aligned} 0.5 &= P(Y_t \leq y_t | p_t = \tilde{p}_t^U) \\ &= \sum_{i=0}^{y_t} \binom{n_t}{i} (\tilde{p}_t^U)^i (1 - \tilde{p}_t^U)^{n_t-i}. \end{aligned} \quad (9)$$

The values of  $\tilde{p}_t^L$  and  $\tilde{p}_t^U$  can then actually be obtained by using the relationship between the cumulative beta distribution and the cumulative binomial distribution function as follows (Daly [13] and Johnson et al. [14]).

Let  $U \sim \text{Beta}(\alpha, \beta)$ :

$$\begin{aligned} F(p_t, \alpha, \beta) &= \int_0^{p_t} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} U^{\alpha-1} (1-U)^{\beta-1} dU \\ &= \sum_{i=\alpha}^{\alpha+\beta-1} \binom{\alpha+\beta-1}{i} p_t^i (1-p_t)^{(\alpha+\beta-1)-i} \\ &= \sum_{i=\alpha}^{n_t} \binom{n_t}{i} p_t^i (1-p_t)^{n_t-i}. \end{aligned} \quad (10)$$

We need to find  $\tilde{p}_t^L$  and  $\tilde{p}_t^U$  such that

$$\begin{aligned} F(\tilde{p}_t^L | \alpha = y_t, \beta = n_t - y_t + 1) &= 0.5, \\ F(\tilde{p}_t^U | \alpha = y_t + 1, \beta = n_t - y_t + 2) &= 0.5. \end{aligned} \quad (11)$$

In particular,

$$\begin{aligned} \tilde{p}_t^L &= F^{-1}(0.5 | \alpha = y_t, \beta = n_t - y_t + 1), \\ \tilde{p}_t^U &= F^{-1}(0.5 | \alpha = y_t + 1, \beta = n_t - y_t + 2), \end{aligned} \quad (12)$$

where  $F^{-1}(Q | \alpha, \beta)$  is the  $Q$ th quantile of the beta-distribution with parameters  $\alpha$  and  $\beta$ .

TABLE 1: The estimated values of odds ratio for  $(n_1, n_2) = (10, 10)$ .

$(p_1, p_2)$	OR	$\widehat{\text{OR}}_{\text{EB}}$	$\widehat{\text{OR}}_{\text{MMLE}}$	$\widehat{\text{OR}}_{\text{MMUE}}$
(0.01, 0.01)	1.0000	1.3665	1.1514	1.2043
(0.01, 0.03)	0.3266	0.3931	1.0029	1.0377
(0.01, 0.05)	0.1919	0.2248	0.8723	0.8935
(0.01, 0.10)	0.0909	0.1040	0.6219	0.6184
(0.01, 0.15)	0.0572	0.0650	0.4481	0.4307
(0.03, 0.01)	3.0619	3.8746	1.6008	1.7848
(0.03, 0.03)	1.0000	1.1119	1.3933	1.5383
(0.03, 0.05)	0.5876	0.6363	1.2128	1.3244
(0.03, 0.10)	0.2784	0.2942	0.8640	0.9156
(0.03, 0.15)	0.1753	0.1838	0.6227	0.6378
(0.05, 0.01)	5.2105	6.5657	2.0724	2.4094
(0.05, 0.03)	1.7018	1.8851	1.8036	2.0763
(0.05, 0.05)	1.0000	1.0787	1.5693	1.7868
(0.05, 0.10)	0.4737	0.4989	1.1181	1.2356
(0.05, 0.15)	0.2982	0.3116	0.8059	0.8609
(0.10, 0.01)	11.0000	13.7434	3.3472	4.1489
(0.10, 0.03)	3.5926	3.9471	2.9135	3.5759
(0.10, 0.05)	2.1111	2.2585	2.5352	3.0777
(0.10, 0.10)	1.0000	1.0445	1.8068	2.1288
(0.10, 0.15)	0.6296	0.6523	1.3027	1.4839
(0.15, 0.01)	17.4706	21.7299	4.7827	6.1533
(0.15, 0.03)	5.7059	6.2349	4.1625	5.3026
(0.15, 0.05)	3.3529	3.5678	3.6225	4.5648
(0.15, 0.10)	1.5882	1.6498	2.5812	3.1568
(0.15, 0.15)	1.0000	1.0303	1.8602	2.1990

Now suppose  $y_t = 0$ , and then

$$\begin{aligned} P(Y_t \geq y_t | p_t = \tilde{p}_t^L) &= P(Y_t \geq 0 | p_t = \tilde{p}_t^L) = 1, \\ \sum_{i=0}^{n_t} \binom{n_t}{i} (\tilde{p}_t^L)^i (1 - \tilde{p}_t^L)^{n_t-i} &= 1. \end{aligned} \quad (13)$$

Any value of  $\tilde{p}_t^L$  in the interval  $[0, 1]$  satisfies

$$P(Y_t \geq 0 | p_t = \tilde{p}_t^L) \geq 0.5, \quad (14)$$

where  $\tilde{p}_t^L = 0$  is the smallest possible value of  $\tilde{p}_t^L$ .

Similarly, when  $y_t = 0$ ,  $\tilde{p}_t^U$  satisfies

$$\begin{aligned} P(Y_t \leq y_t | p_t = \tilde{p}_t^U) &= P(Y_t = 0 | p_t = \tilde{p}_t^U) = 0.5, \\ \binom{n_t}{0} (\tilde{p}_t^U)^0 (1 - \tilde{p}_t^U)^{n_t-0} &= 0.5, \end{aligned} \quad (15)$$

$$\tilde{p}_t^U = 1 - 0.5^{(1/n_t)}.$$

Consequently,  $y_t = 0$ ;  $\tilde{p}_t$  equals

$$\tilde{p}_t = \frac{(\tilde{p}_t^L + \tilde{p}_t^U)}{2} = \frac{1 - 0.5^{(1/n)}}{2}. \quad (16)$$

Similarly, when  $\tilde{p}_t^U = 1$  is the largest possible value of  $\tilde{p}_t^U$ , then  $\tilde{p}_t^L$  satisfies

$$\binom{n}{n} (\tilde{p}_t^L)^n (1 - \tilde{p}_t^L)^{n-n} = 0.5, \quad (17)$$

$$\tilde{p}_t^L = (0.5)^{1/n}$$

when  $y_t = n$  and  $\tilde{p}_t = (\tilde{p}_t^L + \tilde{p}_t^U)/2$ .

Then, the MMUE of odds ratio estimation is defined as

$$\widehat{\text{OR}}_{\text{MMUE}} = \frac{\tilde{p}_1 / (1 - \tilde{p}_1)}{\tilde{p}_2 / (1 - \tilde{p}_2)}, \quad (18)$$

where  $\tilde{p}_1$  and  $\tilde{p}_2$  denote success probability estimators in groups 1 and 2, respectively.

### 3. Proposed Estimation of Odds Ratio

In this section, we proposed a new method for odds ratio estimation using Empirical Bayes method in two independent

binomial distributions. Let  $Y_1$  and  $Y_2$  be random variables, distributed as binomial with equal and unequal sample sizes and unknown probability,  $Y_1 \sim \text{Bin}(n_1, p_1)$  and  $Y_2 \sim \text{Bin}(n_2, p_2)$ , where  $n_1, n_2$  and  $p_1, p_2$  denote two sample sizes and two unknown success probabilities. Adopt information priors on  $p_i$ ,  $p_i \sim \text{beta}(\alpha_i, \beta_i)$ ,  $i = 1, 2$ , where  $\alpha_i$  and  $\beta_i$  denote unknown hyperparameters. The estimation of hyperparameters can be obtained from the posterior marginal distribution function as follows:

$$m(\mathbf{y} | \alpha, \beta) = \int_0^1 f(\mathbf{y} | p) \pi(p) dp \quad (19)$$

$$= \binom{n}{y} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y + \alpha)\Gamma(n - y + \beta)}{\Gamma(n + \alpha + \beta)}.$$

Consequently, the posterior marginal distribution function of  $y$  is the beta-binomial distribution (BBD).

Then, both hyperparameters in each group can be estimated using maximum likelihood method. The likelihood function of posterior marginal distribution function is then written as

$$L(\mathbf{y} | \alpha, \beta) = \prod_{i=1}^n \binom{n}{y_i} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y_i + \alpha)\Gamma(n - y_i + \beta)}{\Gamma(n + \alpha + \beta)} \quad (20)$$

$$= \prod_{i=1}^n \binom{n}{y_i} \frac{(\alpha + y_i - 1)(\alpha + y_i - 2) \cdots \alpha (\beta + n - y_i - 1)(\beta + n - y_i - 2) \cdots \beta}{(\alpha + \beta + n - 1)(\alpha + \beta + n - 2) \cdots (\alpha + \beta)}.$$

Applying Newton-Raphson method to solve a nonlinear equation, the  $(r + 1)$ th maximum likelihood estimator of hyperparameters ( $r = 1, 2, \dots$ ) can be obtained from

$$\begin{bmatrix} \hat{\alpha}^{(r+1)} \\ \hat{\beta}^{(r+1)} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}^{(r)} \\ \hat{\beta}^{(r)} \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 L^{(r)}}{\partial \alpha^2} & \frac{\partial^2 L^{(r)}}{\partial \alpha \partial \beta} \\ \frac{\partial^2 L^{(r)}}{\partial \beta \partial \alpha} & \frac{\partial^2 L^{(r)}}{\partial \beta^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial L^{(r)}}{\partial \alpha} \\ \frac{\partial L^{(r)}}{\partial \beta} \end{bmatrix}, \quad (21)$$

where

$$\frac{\partial^2 L^{(r)}}{\partial \alpha^2} = -\sum_{i=1}^n \left[ \frac{1}{(\alpha^{(r)} + y_i - 1)^2} + \frac{1}{(\alpha^{(r)} + y_i - 2)^2} \right. \\ \left. + \cdots + \frac{1}{(\alpha^{(r)})^2} \right] + \sum_{i=1}^n \left[ \frac{1}{(\alpha^{(r)} + \beta^{(r)} + n - 1)^2} \right. \\ \left. + \frac{1}{(\alpha^{(r)} + \beta^{(r)} + n - 2)^2} + \cdots + \frac{1}{(\alpha^{(r)} + \beta^{(r)})^2} \right],$$

$$\frac{\partial^2 L^{(r)}}{\partial \alpha \partial \beta} = \sum_{i=1}^n \left[ \frac{1}{(\alpha^{(r)} + \beta^{(r)} + n - 1)^2} \right.$$

$$\left. + \frac{1}{(\alpha^{(r)} + \beta^{(r)} + n - 2)^2} + \cdots + \frac{1}{(\alpha^{(r)} + \beta^{(r)})^2} \right],$$

$$\frac{\partial^2 L^{(r)}}{\partial \beta^2} = \sum_{i=1}^n \left[ \frac{1}{(\beta^{(r)} + n + y_i - 1)^2} \right. \\ \left. + \frac{1}{(\beta^{(r)} + n + y_i - 2)^2} + \cdots + \frac{1}{(\beta^{(r)})^2} \right] \\ + \sum_{i=1}^n \left[ \frac{1}{(\alpha^{(r)} + \beta^{(r)} + n - 1)^2} \right. \\ \left. + \frac{1}{(\alpha^{(r)} + \beta^{(r)} + n - 2)^2} + \cdots + \frac{1}{(\alpha^{(r)} + \beta^{(r)})^2} \right],$$

$$\frac{\partial L^{(r)}}{\partial \alpha} = \sum_{i=1}^n \left[ \frac{1}{\alpha^{(r)} + y_i - 1} + \frac{1}{\alpha^{(r)} + y_i - 2} + \cdots \right. \\ \left. + \frac{1}{\alpha^{(r)}} \right] - \sum_{i=1}^n \left[ \frac{1}{\alpha^{(r)} + \beta^{(r)} + n - 1} \right. \\ \left. + \frac{1}{\alpha^{(r)} + \beta^{(r)} + n - 2} + \cdots + \frac{1}{\alpha^{(r)} + \beta^{(r)}} \right],$$

TABLE 2: The estimated values of odds ratio for  $(n_1, n_2) = (10, 30)$ .

$(p_1, p_2)$	OR	$\widehat{OR}_{EB}$	$\widehat{OR}_{MMLE}$	$\widehat{OR}_{MMUE}$
(0.01, 0.01)	1.0000	1.3656	2.9352	2.7925
(0.01, 0.03)	0.3266	0.4023	2.0063	1.8490
(0.01, 0.05)	0.1919	0.2279	1.4146	1.2576
(0.01, 0.10)	0.0909	0.4043	0.6802	0.5486
(0.01, 0.15)	0.0572	0.0652	0.3981	0.2971
(0.03, 0.01)	3.0619	3.8640	4.0816	4.1395
(0.03, 0.03)	1.0000	1.1385	2.7872	2.7376
(0.03, 0.05)	0.5876	0.6452	1.9663	1.8632
(0.03, 0.10)	0.2784	1.1466	0.9457	0.8131
(0.03, 0.15)	0.1753	0.1845	0.5536	0.4406
(0.05, 0.01)	5.2105	6.5508	5.2833	5.5870
(0.05, 0.03)	1.7018	1.9307	3.6077	3.6950
(0.05, 0.05)	1.0000	1.0940	2.5446	2.5144
(0.05, 0.10)	0.4737	1.9430	1.2236	1.0969
(0.05, 0.15)	0.2982	0.3128	0.7163	0.5944
(0.10, 0.01)	11.0000	13.7159	8.5346	9.6223
(0.10, 0.03)	3.5926	4.0427	5.8289	6.3651
(0.10, 0.05)	2.1111	2.2907	4.1132	4.3339
(0.10, 0.10)	1.0000	1.0327	1.6556	1.5221
(0.10, 0.15)	0.6296	0.6549	1.1578	1.0247
(0.15, 0.01)	17.4706	21.6662	12.1932	14.2687
(0.15, 0.03)	5.7059	6.3850	8.3278	9.4395
(0.15, 0.05)	3.3529	3.6181	5.8732	6.4225
(0.15, 0.10)	1.5882	6.4257	2.8237	2.8019
(0.15, 0.15)	1.0000	1.0345	1.6529	1.5181

TABLE 3: The estimated values of odds ratio for  $(n_1, n_2) = (10, 50)$ .

$(p_1, p_2)$	OR	$\widehat{OR}_{EB}$	$\widehat{OR}_{MMLE}$	$\widehat{OR}_{MMUE}$
(0.01, 0.01)	1.0000	1.5096	4.2803	3.9580
(0.01, 0.03)	0.3266	0.4210	2.3790	2.0799
(0.01, 0.05)	0.1919	0.6975	1.4513	1.1915
(0.01, 0.10)	0.0909	0.1057	0.6167	0.4475
(0.01, 0.15)	0.0572	0.0656	0.3656	0.2515
(0.03, 0.01)	3.0619	4.2760	5.9515	5.8667
(0.03, 0.03)	1.0000	1.1926	3.3063	3.0810
(0.03, 0.05)	0.5876	1.9756	2.0173	1.7653
(0.03, 0.10)	0.2784	0.2992	0.8578	0.6636
(0.03, 0.15)	0.1753	0.1856	0.5083	0.3728
(0.05, 0.01)	5.2105	7.2486	7.7012	7.9149
(0.05, 0.03)	1.7018	2.0224	4.2796	4.1586
(0.05, 0.05)	1.0000	3.3494	2.6114	2.3833
(0.05, 0.10)	0.4737	0.5073	1.1100	0.8954
(0.05, 0.15)	0.2982	0.3147	0.6579	0.5031
(0.10, 0.01)	11.0000	15.1780	12.4415	13.6337
(0.10, 0.03)	3.5926	4.2347	6.9165	7.1666
(0.10, 0.05)	2.1111	7.0132	4.2213	4.1087
(0.10, 0.10)	1.0000	1.0621	1.7938	1.5433
(0.10, 0.15)	0.6296	0.6589	1.0631	0.8669
(0.15, 0.01)	17.4706	23.9776	17.7763	20.2195
(0.15, 0.03)	5.7059	6.6884	9.8777	10.6225
(0.15, 0.05)	3.3529	11.0773	6.0256	6.0860
(0.15, 0.10)	1.5882	1.6775	2.5614	2.2869
(0.15, 0.15)	1.0000	1.0408	1.5181	1.2848

$$\begin{aligned} \frac{\partial L^{(r)}}{\partial \beta} &= \sum_{i=1}^n \left[ \frac{1}{\beta^{(r)} + n + y_i - 1} + \frac{1}{\beta^{(r)} + n + y_i - 2} + \dots \right. \\ &+ \left. \frac{1}{\beta^{(r)}} \right] - \sum_{i=1}^n \left[ \frac{1}{\alpha^{(r)} + \beta^{(r)} + n - 1} \right. \\ &+ \left. \frac{1}{\alpha^{(r)} + \beta^{(r)} + n - 2} + \dots + \frac{1}{\alpha^{(r)} + \beta^{(r)}} \right], \end{aligned} \tag{22}$$

where the moment estimators of hyperparameters in beta-binomial distribution are used as initial values; see Minka [15].

The posterior distribution function of  $p$  is thus calculated, yielding

$$\begin{aligned} \pi(p | y, \alpha, \beta) &= p^{(y+\alpha-1)} (1-p)^{(n-y+\beta-1)} \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}. \end{aligned} \tag{23}$$

Substituting the estimators of  $\alpha$  and  $\beta$ , we obtain

$$p | y, \alpha, \beta \sim \text{Beta}(y + \widehat{\alpha}, n - y + \widehat{\beta}). \tag{24}$$

Let  $p'_1$  and  $p'_2$  be estimators of  $p_1$  and  $p_2$ , respectively, where

$$\begin{aligned} p'_1 &= \frac{y_1 + \widehat{\alpha}_1}{n_1 + \widehat{\alpha}_1 + \widehat{\beta}_1}, \\ p'_2 &= \frac{y_2 + \widehat{\alpha}_2}{n_2 + \widehat{\alpha}_2 + \widehat{\beta}_2}. \end{aligned} \tag{25}$$

Thus, the EB estimator of odds ratio can be obtained as follows:

$$\widehat{OR}_{EB} = \frac{p'_1 / (1 - p'_1)}{p'_2 / (1 - p'_2)}, \tag{26}$$

where  $p'_1$  and  $p'_2$  denote success probability estimators in groups 1 and 2, respectively.

#### 4. Simulation Study for MMLE, MMUE, and EB Method

Simulation studies have been carried out using R program (version 3.2.0) [16] to assess the efficiency of the EB method in comparison with two existing methods. Binomial data are generated with equal and unequal sample sizes:  $(n_1, n_2) = (10, 10), (10, 30)$  with  $(10, 50)$  probabilities of success in group 1:  $p_1 = 0.01, 0.03, 0.05, 0.1$ , and  $0.15$ . For each value

TABLE 4: The percentages of the Estimated Relative Error of odds ratio estimation for  $(n_1, n_2) = (10, 10)$ .

$(p_1, p_2)$	ERE <sub>EB</sub>	ERE <sub>MMLE</sub>	ERE <sub>MMUE</sub>
(0.01, 0.01)	36.6535	15.1385	20.4316
(0.01, 0.03)	20.3505	206.7910	217.7388
(0.01, 0.05)	17.1281	354.5204	365.6116
(0.01, 0.10)	14.3630	584.0771	580.2372
(0.01, 0.15)	13.4914	682.8616	652.4367
(0.03, 0.01)	26.5436	47.7187	41.7082
(0.03, 0.03)	11.1895	39.3332	53.8303
(0.03, 0.05)	8.2767	106.3838	125.3783
(0.03, 0.10)	5.6857	210.3971	228.9552
(0.03, 0.15)	4.8556	255.2977	263.9451
(0.05, 0.01)	26.0092	60.2273	53.7592
(0.05, 0.03)	10.7763	5.9852	22.0098
(0.05, 0.05)	7.8661	56.9330	78.6812
(0.05, 0.10)	5.3167	136.0498	160.8427
(0.05, 0.15)	4.4721	170.2257	188.6564
(0.10, 0.01)	24.9404	69.5705	62.2831
(0.10, 0.03)	9.8685	18.9032	0.4659
(0.10, 0.05)	6.9818	20.0893	45.7877
(0.10, 0.10)	4.4467	80.6791	112.8820
(0.10, 0.15)	3.6016	106.9073	135.6854
(0.15, 0.01)	24.3801	72.6244	64.7792
(0.15, 0.03)	9.2720	27.0495	7.0675
(0.15, 0.05)	6.4085	8.0392	36.1422
(0.15, 0.10)	3.8773	62.5224	98.7588
(0.15, 0.15)	3.0338	86.0169	119.9015

TABLE 5: The percentages of the Estimated Relative Error of odds ratio estimation for  $(n_1, n_2) = (10, 30)$ .

$(p_1, p_2)$	ERE <sub>EB</sub>	ERE <sub>MMLE</sub>	ERE <sub>MMUE</sub>
(0.01, 0.01)	36.5558	193.5166	179.2467
(0.01, 0.03)	23.1721	514.2925	466.1248
(0.01, 0.05)	18.7270	637.0581	555.2608
(0.01, 0.10)	344.7753	648.1943	503.4379
(0.01, 0.15)	13.8933	595.4561	419.0998
(0.03, 0.01)	26.1981	33.3063	35.1969
(0.03, 0.03)	13.8541	178.7227	173.7624
(0.03, 0.05)	9.7978	234.6089	217.0772
(0.03, 0.10)	311.9138	239.7364	192.0982
(0.03, 0.15)	5.2578	215.8678	151.3923
(0.05, 0.01)	25.7219	1.3963	7.2250
(0.05, 0.03)	13.4545	112.0004	117.1259
(0.05, 0.05)	9.3998	154.4641	151.4381
(0.05, 0.10)	310.1875	158.3124	131.5756
(0.05, 0.15)	4.8844	140.1677	99.2831
(0.10, 0.01)	24.6897	22.4126	12.5246
(0.10, 0.03)	12.5282	62.2482	77.1730
(0.10, 0.05)	8.5055	94.8364	105.2921
(0.10, 0.10)	3.2693	65.5584	52.2104
(0.10, 0.15)	4.0198	83.8863	62.7515
(0.15, 0.01)	24.0150	30.2071	18.3274
(0.15, 0.03)	11.9013	45.9518	65.4338
(0.15, 0.05)	7.9089	75.1670	91.5471
(0.15, 0.10)	304.5808	77.7868	76.4163
(0.15, 0.15)	3.4464	65.2920	51.8093

of  $p_1, p_2$  is varied to 0.01, 0.03, 0.05, 0.1, and 0.15. Each situation is repeated 5,000 times after removing the first 1,000 iterations (1,000 burn-ins). The efficiency of proposed estimator is evaluated using Estimated Relative Error (ERE), defined as

$$ERE = \left[ \frac{|\text{OR} - \widehat{\text{OR}}_i|}{\text{OR}} \right], \tag{27}$$

where OR denotes the usual maximum likelihood estimator of odds ratio and  $\widehat{\text{OR}}_i$  denotes the estimate of odds ratio using EB, MMLE, and MMUE ( $i = 1, 2, 3, \dots$ ), respectively.

The simulation results with odds ratio estimates for sample sizes  $(n_1, n_2) = (10, 10), (10, 30)$  and  $(10, 50)$  are given in Tables 1–3. The performance of estimation uses ERE given in Tables 4–6 and compares this result with graph in Figure 1; the other case provides similar results. It is found that the odds ratio estimation using EB method mostly yields smallest ERE with 78.67%, while those using MMLE and MMUE methods result in smallest ERE with only 6.67% and 14.66%, respectively.

### 5. Illustrative Examples Using Real Data

Our first example is taken from the studies of Good [7] and Hardell [17]. As shown in Table 7, subjects with malignant

lymphoma and colon cancer cases and controls who are shortly exposed to phenoxy acids in agriculture and forestry were observed, including the true odds ratios and their estimates using EB, MMLE, and MMUE. For outcome in which  $(Y_1, Y_2) = (8, 9)$  out of  $(n_1, n_2) = (24, 53)$  for cases and control, respectively, the estimate of the odds ratio using EB method yields the least ERE with 0.5523, while those using MMLE and MMUE methods result in ERE with 1.2805 and 4.1483, respectively.

The second example is taken from the study of Perondi et al. [18], as shown in Table 7, which compared high-dose epinephrine and standard-dose epinephrine in children with cardiac arrest with 34 children in each treatment, including the true odds ratios and their estimates using EB, MMLE, and MMUE. For outcome measure was survival at 24 hours in which  $(Y_1, Y_2) = (1, 7)$  out of  $(n_1, n_2) = (34, 34)$  for high dose and standard dose, respectively. The estimate of the odds ratio using EB method yields the least ERE with 5.2097, while those using MMUE and MMLE methods result in ERE with 15.5305 and 40.4643, respectively.

### 6. Conclusion

Based on simulated study for odds ratio estimation in rare events with two independent binomial data, the result indicates that the proposed method performs rather well.

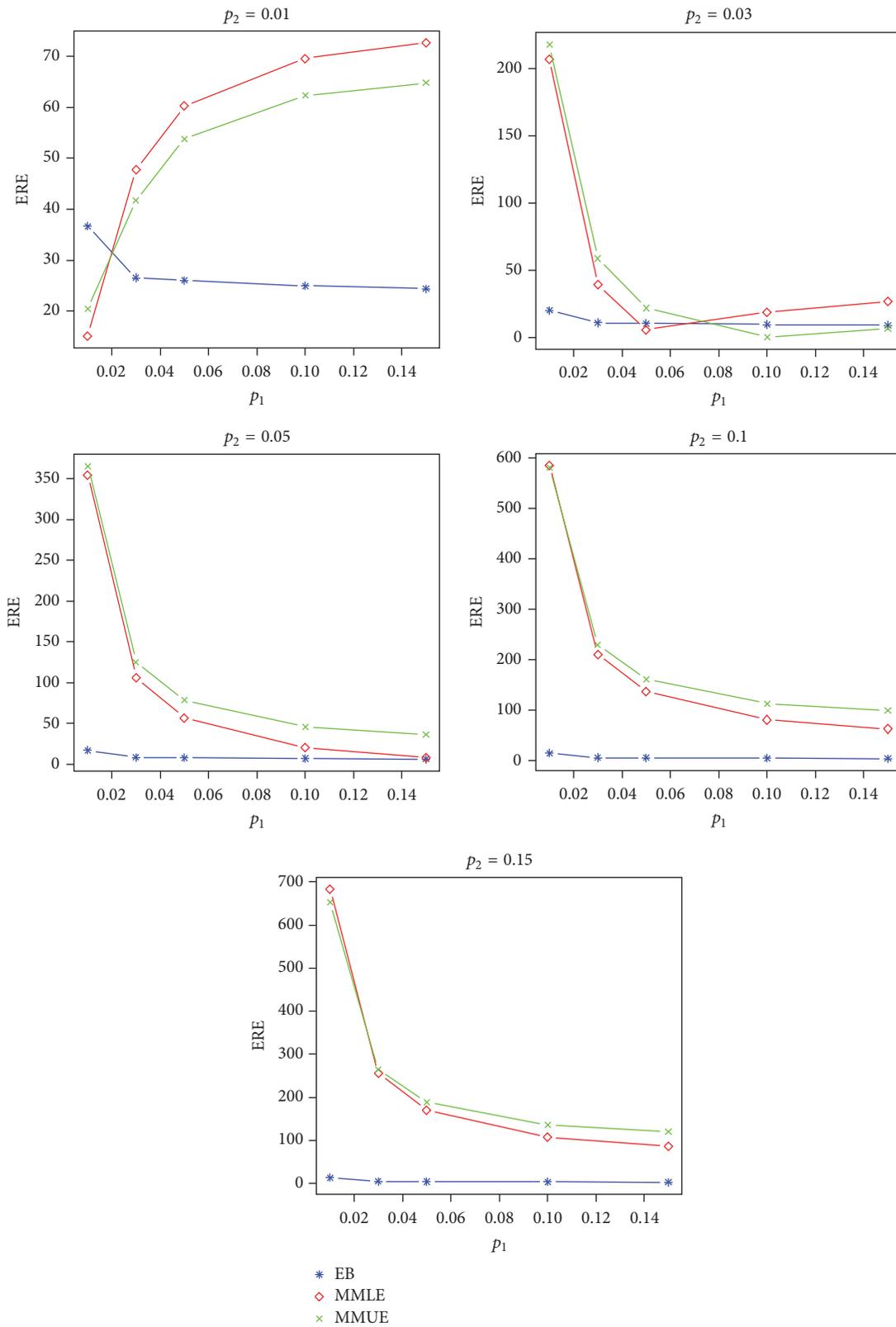


FIGURE 1: The percentages of ERE for odds ratio estimation using EB, MMLE, and MMUE when  $(n_1, n_2) = (10, 10)$ .

TABLE 6: The percentages of the Estimated Relative Error of odds ratio estimation for  $(n_1, n_2) = (10, 50)$ .

$(p_1, p_2)$	ERE <sub>EB</sub>	ERE <sub>MMLE</sub>	ERE <sub>MMUE</sub>
(0.01, 0.01)	50.9628	328.0343	295.7966
(0.01, 0.03)	28.9076	628.4030	536.8268
(0.01, 0.05)	263.4419	656.1777	520.8249
(0.01, 0.10)	16.2619	578.3828	392.2232
(0.01, 0.15)	14.5797	538.7736	339.4512
(0.03, 0.01)	39.6535	94.3767	91.6069
(0.03, 0.03)	19.2635	230.6324	208.1003
(0.03, 0.05)	236.1914	243.2889	200.4045
(0.03, 0.10)	7.4747	208.1561	138.4178
(0.03, 0.15)	5.8990	190.0357	112.7091
(0.05, 0.01)	39.1153	47.8002	51.9027
(0.05, 0.03)	18.8401	151.4810	144.3693
(0.05, 0.05)	234.9386	161.1356	138.3292
(0.05, 0.10)	7.0955	134.3384	89.0341
(0.05, 0.15)	5.5239	120.5896	68.6865
(0.10, 0.01)	37.9815	13.1041	23.9423
(0.10, 0.03)	17.8718	92.5224	99.4826
(0.10, 0.05)	232.2048	99.9578	94.6234
(0.10, 0.10)	6.2116	79.3838	54.3286
(0.10, 0.15)	4.6546	68.8420	37.6836
(0.15, 0.01)	37.2457	1.7496	15.7345
(0.15, 0.03)	17.2201	73.1152	86.1673
(0.15, 0.05)	230.3755	79.7100	81.5120
(0.15, 0.10)	5.6227	61.2754	43.9902
(0.15, 0.15)	4.0763	51.8150	28.4826

TABLE 7: True odds ratios and their estimates using EB, MMLE, and MMUE, with corresponding percentages of ERE.

		Methods			
		True	EB	MMLE	MMUE
1st example	OR	2.4444	2.4309	2.4131	2.3430
	ERE	—	0.5523	1.2805	4.1483
2nd example	OR	0.1169	0.1230	0.1642	0.1350
	ERE	—	5.2097	40.4643	15.5305

The EB estimator of odds ratio is also more efficient than the other two estimators, MMLE and MMUE. In addition, our purposed estimator is an alternative method for odds ratio estimation to the MMLE method without disturbing the original data.

**Competing Interests**

The authors declare that they have no competing interests.

**Acknowledgments**

The authors are grateful to the Graduate College, King Mongkut’s University of Technology North Bangkok, for the financial support.

**References**

- [1] J. B. S. Haldane, “The estimation and significance of the logarithm of a ratio of frequencies,” *Annals of Human Genetics*, vol. 20, no. 4, pp. 309–311, 1956.
- [2] J. J. Gart and J. R. Zweifel, “On the bias of various estimators of the logit and its variance with application of quantal bioassay,” *Biometrika*, vol. 54, no. 1-2, pp. 181–187, 1967.
- [3] Y. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, Springer, New York, NY, USA, 2007.
- [4] A. Agresti and M.-C. Yang, “An empirical investigation of some effects of sparseness in contingency tables,” *Computational Statistics and Data Analysis*, vol. 5, no. 1, pp. 9–21, 1987.
- [5] K. F. Hirji, A. A. Tsiatis, and C. R. Mehta, “Median unbiased estimation for binary data,” *The American Statistician*, vol. 43, no. 1, pp. 7–11, 1989.
- [6] M. Parzen, S. Lipsitz, J. Ibrahim, and N. Klar, “An estimate of the odds ratio that always exists,” *Journal of Computational and Graphical Statistics*, vol. 11, no. 2, pp. 420–436, 2002.
- [7] I. J. Good, “On the estimations of small frequencies in contingency tables,” *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 18, pp. 113–124, 1956.
- [8] R. A. Fisher, “The logic of inductive inference,” *Journal of the Royal Statistical Society*, vol. 98, no. 1, pp. 39–82, 1935.
- [9] D. G. Thomas and J. J. Gart, “A table of exact confidence limits for differences and ratios of two proportions and their odds ratios,” *Journal of the American Statistical Association*, vol. 72, no. 357, pp. 73–76, 1977.
- [10] P. M. E. Altham, “Exact Bayesian analysis of a  $2 \times 2$  contingency table, and Fisher’s ‘exact’ significance test,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 31, pp. 261–269, 1969.
- [11] M. Nurminen and P. Mutanen, “Exact Bayesian analysis of two proportions,” *Scandinavian Journal of Statistics*, vol. 14, no. 1, pp. 67–77, 1987.
- [12] B. Nouri, N. Zare, and S. M. T. Ayatollahi, “Validation study methods for estimating odds ratio in  $2 \times 2 \times J$  tables when exposure is misclassified,” *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 170120, 8 pages, 2013.
- [13] L. Daly, “Simple SAS macros for the calculation of exact binomial and Poisson confidence limits,” *Computers in Biology and Medicine*, vol. 22, no. 5, pp. 351–361, 1992.
- [14] N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate Discrete Distribution*, John Wiley & Sons, New Jersey, NJ, USA, 2005.
- [15] T. P. Minka, “Estimating a Dirichlet distribution,” Tech. Rep., The MIT Press, London, UK, 2000.
- [16] The R Development Core Team, An introduction to R, Vienna, 2010, <http://R-project.org>.
- [17] L. Hardell, “Relation of soft-tissue sarcoma, malignant lymphoma and colon cancer to phenoxy acids, chlorophenols and other agents,” *Scandinavian Journal of Work, Environment and Health*, vol. 7, no. 2, pp. 119–130, 1981.
- [18] M. B. M. Perondi, A. G. Reis, E. F. Paiva, V. M. Nadkarni, and R. A. Berg, “A comparison of high-dose and standard-dose epinephrine in children with cardiac arrest,” *The New England Journal of Medicine*, vol. 350, no. 17, pp. 1722–1730, 2004.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

