11-1-2013

Comparison of Three Calculation Methods for a Bayesian Inference of $P(\pi_1 > \pi_2)$

Yohei Kawasaki
National Center for Global Health and Medicine, Tokyo, Japan, yk.sep10@yahoo.co.jp

Asanao Shimokawa
Tokyo University of Science, Tokyo, Japan

Etsuo Miyaoka
Tokyo University of Science, Tokyo, Japan

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Recommended Citation
Kawasaki, Yohei; Shimokawa, Asanao; and Miyaoka, Etsuo (2013) "Comparison of Three Calculation Methods for a Bayesian Inference of $P(\pi_1 > \pi_2)$," Journal of Modern Applied Statistical Methods: Vol. 12: Iss. 2, Article 15.
Available at: http://digitalcommons.wayne.edu/jmasm/vol12/iss2/15

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized administrator of DigitalCommons@WayneState.
Comparison of Three Calculation Methods for a Bayesian Inference of $P(\pi_1 > \pi_2)$

Yohei Kawasaki  
National Center for Global Health and Medicine  
Tokyo, Japan

Asanao Shimokawa  
Tokyo University of Science  
Tokyo, Japan

Etsuo Miyaoka  
Tokyo University of Science  
Tokyo, Japan

In Bayesian inference, some researchers have examined the difference of binomial proportions using $\theta = P(\pi_1 > \pi_2 - \Delta_0|X_1, X_2)$, where $X_i$ denote binomial random variable with parameter $\pi_i$. An approximate method and the MCMC method are compared with an exact method for $\theta$, and results of actual clinical trials using $\theta$ are presented.

Keywords: Binominal proportions; Bayesian inference; MCMC method; hypergeometric series.

Introduction

Statistical inference concerning the difference between two independent binomial proportions is often discussed from the frequency rather than the Bayesian viewpoint. Some researchers have examined significant differences in binomial proportions using the index, $\theta = P(\pi_1 > \pi_2 - \Delta_0|X_1, X_2)$, which indicates the difference in the posterior density for two independent binomial proportions that are assumed to be random variables.

Originally, this index can be shown in the framework of frequency theory to be, $P(Y_1 > Y_2)$, where $Y_1$ and $Y_2$ are random variables. The inference for $P(Y_1 > Y_2)$ can be observed in various fields. In engineering, it is used in the `stress strength model' to evaluate the reliability of an industrial component (see for instance Kotz, et al. (2003)). In clinical research, it is used as an index for the comparison of two groups given different treatments. In addition, this probability corresponds to the area under the receiver operating characteristic (ROC) curve.

Dr. Kawasaki is a Senior Biostatistician at the National Center for Global Health and Medicine. Email at yk_sep10@yahoo.co.jp. Asanao Shimokawa is a graduate student. Dr. Miyaoka is a professor in the mathematics department.
In medicine, it is used as an index for evaluating the validity of a diagnostic method. Indeed, innumerable studies have been conducted for \( P(Y_1 > Y_2) \) in the framework of frequency theory (See for instance Sen (1960, 1967)). As for research papers on this index, Shirahata (1993), Zhou (2008) and Kawasaki and Miyaoka (2010) have published actively in recent years.

Conversely, there have been a number of studies to apply a construction of \( P(Y_1 > Y_2) \) to the Bayesian framework. Basu (1996) concisely showed the use of the Bayesian approach with respect to hypothesis testing. Berry (1995) using superior binomial proportions, presented a detailed comparison between two binomial proportions assumed to be random variables and presented some interesting examples. Zaslavsky (2009, 2010) applied \( \pi \) to a one-side hypothesis based on a one-sample situation. Kawasaki and Miyaoka (2012) showed an exact expression for \( \pi \), and applied \( \pi \) to a one-side hypothesis based on a two-sample situation.

There are some pending issues with the above-mentioned method. An approximate method and exact method of \( \pi \) were adopted only while using a conjugate prior. The drawback of the approximate method is that it occasionally leads to a rough result in a small sample. The drawback of the exact method is that it is slightly complicated. In addition, the exact method requires extensive computing time with a large sample size. Hence, a Markov Chain Monte Carlo (MCMC) method is proposed for \( \pi \) as a solution to these problems.

**Methodology**

Let \( X_1 \) and \( X_2 \) denote binomial random variables for \( n_1 \) and \( n_2 \) trials with parameters \( \pi_1 \) and \( \pi_2 \), respectively. The conjugate prior density for \( \pi_i \) is a beta distribution with parameters \( \alpha_i \) and \( \beta_i \), where \( \alpha_i > 0, \beta_i > 0 \), and \( i = 1, 2 \). The proposed posterior density for \( \pi_i \) is

\[
g_i(\pi_i|X_i) = \frac{1}{B(a_i,b_i)} \pi_i^{a_i-1} (1-\pi_i)^{b_i-1},
\]

where \( a_i = \alpha_i + x_i \), \( b_i = n_i - x_i + \beta_i \), and \( B(a,b) \) is the proposed beta function. Let \( \pi_{i,\pi \text{ post}} \) denote the binomial proportion following the posterior density.
THREE METHODS FOR BAYESIAN INFESSION OF $P(\pi_1 > \pi_2)$

Approximate method for $\theta$

$\theta$ can be calculated via an approximation using the standard normal table. Assume that $a_i$ and $b_i$ of the posterior density are large. It is necessary to determine a Z-test statistic. The expected difference in the posterior density and the variance in this difference can be expressed as:

$$E(\pi_{1,\text{post}} - \pi_{2,\text{post}}) = \mu_{1,\text{post}} - \mu_{2,\text{post}} \dagger,$$

$$V(\pi_{1,\text{post}} - \pi_{2,\text{post}}) = \frac{\mu_{1,\text{post}}(1 - \mu_{1,\text{post}})}{a_i + b_i + 1} + \frac{\mu_{2,\text{post}}(1 - \mu_{2,\text{post}})}{a_i + b_i + 1} \dagger,$$

where $\mu_{i,\text{post}} = a_i / (a_i + b_i)$ denotes the posterior mean of $\pi_i$. The $Z_\theta$-test statistic,

$$Z_\theta = \frac{(\pi_{1,\text{post}} - \pi_{2,\text{post}}) - E(\pi_{1,\text{post}} - \pi_{2,\text{post}})}{\sqrt{V(\pi_{1,\text{post}} - \pi_{2,\text{post}})}} \dagger,$$

is approximately distributed as the standard normal distribution. Therefore, the approximate probability of $\theta$ is given by

$$\theta = P(\pi_1 > \pi_2|X_1, X_2) \approx 1 - \Phi \left( \frac{- (\mu_{1,\text{post}} - \mu_{2,\text{post}})}{\sqrt{\frac{\mu_{1,\text{post}}(1 - \mu_{1,\text{post}})}{a_i + b_i + 1} + \frac{\mu_{2,\text{post}}(1 - \mu_{2,\text{post}})}{a_i + b_i + 1}}} \right)$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution. Thus, the approximate probability can easily be calculated.

Exact method for $\theta$

Kawasaki and Miyaoka (2012) derived the exact expression for $\theta$ using the posterior density. The exact expression for $\theta$ is
\[ \theta = P(\pi_1 > \pi_2 | X_1, X_2) \]
\[ = \frac{B(a_1 + a_2, b_1)}{a_2 B(a_1, b_1) B(a_2, b_2)} \sum_{t=0}^{\infty} \left( \frac{B(k_3)}{l_3} \right)^t \left( \frac{B(k_2)}{l_2} \right)^t \left( \frac{B(k_1)}{l_1} \right)^t \frac{1}{t!}, t + k_1 + k_2 + k_3 < l_1 + l_2 \]  

where

\[ \sum_{t=0}^{\infty} \left( \frac{B(k_3)}{l_3} \right)^t \left( \frac{B(k_2)}{l_2} \right)^t \left( \frac{B(k_1)}{l_1} \right)^t \frac{1}{t!}, t + k_1 + k_2 + k_3 < l_1 + l_2 \]

is the hypergeometric series, and \((k)_t\) is the Pochhammer symbol.

**MCMC method for \( \theta \)**

A computational procedure for \( \theta \) using the MCMC method is now introduced. The MCMC method is a means of sampling from a posterior density. A random-walk Metropolis-Hasting algorithm was used as the MCMC Method. Given that the samples come from two independent populations, the posterior joint distribution of \( \pi_1 \) and \( \pi_2 \) is a product of its marginal distributions. For this reason, one can obtain samples from the posterior distribution of \( \pi_1 - \pi_2 \) by simulating \( k \) values from the posterior distribution of \( \pi_1 \) and \( \pi_2 \) using MCMC procedure of SAS, e.g., \( \pi_{1, \text{post}}^{}, \pi_{2, \text{post}}^{}, \cdots, \pi_{1, \text{post}}^k \) and \( \pi_{2, \text{post}}^{}, \pi_{2, \text{post}}^2, \cdots, \pi_{2, \text{post}}^k \), respectively. Then, by computing
\[ \pi_{1, \text{post}}^1 - \pi_{2, \text{post}}^1, \pi_{1, \text{post}}^2 - \pi_{2, \text{post}}^2, \cdots, \pi_{1, \text{post}}^k - \pi_{2, \text{post}}^k, \]
the simulated values from the posterior distribution of \( \pi_{1, \text{post}} - \pi_{2, \text{post}} \) are obtained. The posterior samples obtained by the MCMC method after the burn-in period are \( \delta_1, \delta_2, \cdots, \delta_k \). Let \( \Delta_1, \Delta_2, \cdots, \Delta_k \) be independent identically distributed random variables with distribution function \( F \). The posterior samples is the observed value of \( \Delta_1, \Delta_2, \cdots, \Delta_k \). Note the fact that \( \theta = P(\pi_{1, \text{post}} > \pi_{2, \text{post}}^1) = P(\pi_{1, \text{post}} - \pi_{2, \text{post}} > 0) \). Thus, \( \theta \) can be expressed as,

\[ \theta = P(\pi_1 > \pi_2 | X_1, X_2) \]
\[ = P(\pi_1 - \pi_2 > 0 | X_1, X_2) \approx 1 - \hat{F}(0) \]
THREE METHODS FOR BAYESIAN INFERENCE OF $P(\pi_1 > \pi_2)$

where

$$\hat{F}_k(s) = \frac{1}{k} \sum_{i=1}^{k} I(A \leq s)$$ \hspace{1cm} (9)$$

and

$$I(\Delta_i \leq s) = \begin{cases} 
1 & \text{if } \Delta_i \leq s \\
0 & \text{if } \Delta_i > s 
\end{cases}$$ \hspace{1cm} (10)$$

is the empirical distribution function.

Results

Comparison of three methods

Now the probabilities of the three methods for $\theta$ are compared. The difference between the sample proportions (horizontal axis) were plotted against the difference between the probabilities of the MCMC and exact methods (vertical axis), as shown in Figures 1, 3, and 5. Similarly, the difference between the sample proportions (horizontal axis) were plotted against the difference between the probabilities of the approximate and exact methods (vertical axis), as shown in Figures 2, 4, and 6. In Figures 1, and 2 consider small sample sizes, i.e., $n_1 = n_2 = 5, 10, 15, 20$. Conversely, in Figures 3 and 4 consider large sample sizes, i.e., $n_1 = n_2 = 60, 70, 80, 90$. Figures 5 and 6 consider groups of different sample sizes, that is, $n_1 = 15, n_2 = 5$; $n_1 = 15, n_2 = 10$; $n_1 = 15, n_2 = 20$; and $n_1 = 15, n_2 = 20$. The following were confirmed from the results.

First, the relationship between the difference in the probabilities and the difference in the sample proportions is described. In Figure 1(d) and Figure 3(d), the probability of the MCMC method is more or less equal to that of the exact method when the difference between the sample proportions is 0.8. On the other hand, the difference between the probabilities of the MCMC and exact methods is around 0.01 when the difference between the sample proportions is 0.05. Overall, when the difference between the sample proportions is large, the probabilities of the MCMC and exact methods are roughly equal. In contrast, when the difference between the sample proportions is small, the probability of the MCMC method is
different from that of the exact method. This general pattern is similar for the difference in the probabilities of the approximation and exact methods.

Next, the relationship between the sample size and the difference in the probabilities is described. In Figure 2(a), the difference between the probabilities of the approximate and exact methods is around 0.013 when the difference between the sample proportions is 0.2. For a slightly larger sample size (Figure 2(d)), the difference between the probabilities of the approximate and exact methods is around 0.006 for the same difference between the sample proportions. In addition, there is virtually no difference between the probabilities of the approximate and exact methods when the sample size is further increased, as shown in Figure 4(d). Thus, the sample size influences the accuracy of the probability of the approximate method. It also shows the difference in the probabilities of the MCMC and exact methods. In Figure 1(a), the difference between the probabilities of the MCMC and exact methods is around 0.006 when the difference between the sample proportions is 0.2. For a slightly larger sample size (Figure 2(d)), the difference between the probabilities of the MCMC and exact methods is around 0.005 for the same difference between the sample proportions. Thus, the accuracy of the probability of the MCMC method always remains high even when the sample sizes are small.

Finally, the difference between the probabilities when groups of different sample sizes are considered is investigated. In Figure 2(d), the difference between the probabilities of the approximate and exact methods is around 0.006 when the difference between the sample proportions is 0.2. On the other hand, in Figure 6(d), the difference between the probabilities of the approximate and exact methods is around 0.012 for the same difference between the sample proportions. In both the cases, the total sample size \((n_1 + n_2)\) is the same. However, the difference between the probabilities of the approximate and exact methods is slightly greater in the case of groups with different sample sizes. It is also shown the case of the MCMC method. In Figure 1(d), the difference between the probabilities of the MCMC and exact methods is around 0.005 when the difference between the sample proportions is 0.2. On the other hand, in Figure 5(d), the difference between the probability of the MCMC and exact methods is around 0.005 for the same difference between the sample proportions. Therefore, the difference between the probabilities of the MCMC and exact methods is the same regardless of whether the sample sizes are equal or different.
THREE METHODS FOR BAYESIAN INFERENCE OF $P(\pi_1 > \pi_2)$

**Figure 1**: Comparison of the Exact and MCMC Method when sample sizes are small. (vertical axis: Differences of $\theta$ in Exact and MCMC method. Prior distribution is Beta(1,1). horizontal axis: Differences of two sample proportions.

**Figure 2**: Comparison of the Exact and Approximate method when sample sizes are small. (vertical axis: Differences of $\theta$ in Exact and Approximation method. horizontal axis: Differences of two sample proportions.
Figure 3: Comparison of the Exact and MCMC Method when sample sizes are large. 
(Vertical axis: Differences of $\theta$ in Exact and MCMC method. Prior distribution is Beta(1,1). 
Horizontal axis: Differences of two sample proportions.

Figure 4: Comparison of the Exact and Approximate method when sample sizes are large. 
(Vertical axis: Differences of $\theta$ in Exact and Approximation method. horizontal axis: Differences of two sample proportions.
THREE METHODS FOR BAYESIAN INFERENCE OF \( P(\pi_1 > \pi_2) \)

Figure 5: Comparison of the Exact and MCMC Method when sample sizes are unbalanced. (vertical axis: Differences of \( \theta \) in Exact and MCMC method. Prior distribution is Beta(1,1). horizontal axis: Differences of two sample proportions.

Figure 6: Comparison of the Exact and Approximate method when sample sizes are unbalanced. (vertical axis: Differences of \( \theta \) in Exact and Approximation method. horizontal axis: Differences of two sample proportions.

264
Example

Next the utility of $\theta$ is illustrated by applying it to the results of clinical trials. A non-informative prior was assumed. Table 1 lists the results of a double-blind, randomized, 41-center study that compares the efficacy of TJN-318 cream with that of Bifonazole (BFZ) cream in the treatment of patients suffering from cutaneous mycosis (TJN-318 Solution Study Group (1992)). The main purpose of this clinical trial was to show that TJN-318 cream is more effective than BFZ cream in the treatment of cutaneous mycosis. The primary end point of this clinical trial is a binary variable. In other words, the patient either recovers or does not recover. In short, the alternative hypothesis is $\pi > \pi_2$. In general, the frequentist approach can be adopted to verify the purpose of the clinical trial via the calculation of a $p$-value. The $p$-value was calculated using the Z-test statistic for the purpose of reference,

$$Z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

(11)

where $\hat{\pi}_i = x_i/n_i$ and $\hat{\pi} = (x_1 + x_2)/(n_1 + n_2)$. The values of $\theta$ are listed in the rightmost column of Table 1. Consequently, a non-informative prior was adopted, that is, $\alpha_i = \beta_i = 1$ and $i = 1, 2$. Clearly, $\theta$ increases when the $p$-value is low, and $\theta \approx 1$ when the null hypothesis is rejected. Moreover, $\theta \approx 1 - \hat{\pi}$.

Next, the results of a double-blind, randomized, phase-3 clinical trial that compares the efficacies of follitropin alpha (hereafter, the study drug) and human menopausal gonadotropin (hereafter, the control drug) in the treatment of patients suffering from no-ovulation-cycle syndrome (from the assessment report of PMDA (2009)) was employed. Table 2 lists the resulting ovulation rate, that is, the primary end point. The Z-test affords a $p$-value of 0.764, which suggests no significant differences. Using the non-informative prior, the approximate probability of $\theta$ is obtained as 0.238, whereas the exact probability and the MCMC probability is obtained as 0.237.
THREE METHODS FOR BAYESIAN INFERENCE OF $P(\pi_1 > \pi_2)$

**Table 1**: The result of primary end point in clinical trial for TJN-318 cream vs Bifonazole cream.

<table>
<thead>
<tr>
<th>Disease Name</th>
<th>Drug Name</th>
<th>Cure</th>
<th>Non-Cure</th>
<th>p-value</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Approximate</td>
<td>Exact</td>
</tr>
<tr>
<td>Tinea Pedis</td>
<td>TJN-318</td>
<td>110</td>
<td>27</td>
<td>0.264</td>
<td>0.734</td>
</tr>
<tr>
<td></td>
<td>BFZ</td>
<td>96</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tinea Corporis</td>
<td>TJN-318</td>
<td>70</td>
<td>13</td>
<td>0.417</td>
<td>0.581</td>
</tr>
<tr>
<td></td>
<td>BFZ</td>
<td>69</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Candidal Intergio</td>
<td>TJN-318</td>
<td>39</td>
<td>4</td>
<td>0.472</td>
<td>0.531</td>
</tr>
<tr>
<td></td>
<td>BFZ</td>
<td>37</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Candidal Interdigital</td>
<td>TJN-318</td>
<td>25</td>
<td>2</td>
<td>0.021</td>
<td>0.978</td>
</tr>
<tr>
<td></td>
<td>BFZ</td>
<td>23</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pteryasis Versicolor</td>
<td>TJN-318</td>
<td>59</td>
<td>2</td>
<td>0.236</td>
<td>0.749</td>
</tr>
<tr>
<td></td>
<td>BFZ</td>
<td>46</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**: The result of primary end point in clinical trial for follitropin alpha vs human menopausal gonadotropin.

<table>
<thead>
<tr>
<th>Drug Name</th>
<th>Cure</th>
<th>Non-cure</th>
<th>Total</th>
<th>Ovulation Ratio</th>
<th>p-value</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Approximate</td>
<td>Exact</td>
</tr>
<tr>
<td>Study</td>
<td>102</td>
<td>27</td>
<td>129</td>
<td>79.1%</td>
<td>0.764</td>
<td>0.238</td>
</tr>
<tr>
<td>Control</td>
<td>109</td>
<td>23</td>
<td>132</td>
<td>82.6%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Conclusion**

Three methods for the index $\theta = P(\pi_1 > \pi_2 | X_1, \uparrow X_2)$ were presented to determine the probability that the binomial proportion for a study drug is superior to that for a control drug. In particular, a new procedure was described based on the MCMC method. The probabilities of these three methods were compared to test the relative effectiveness of each.

The expression for the exact method was presented, which includes a hypergeometric series. It is speculated that this series causes the decrease in calculation efficiency when the sample size is very large. In addition, hypergeometric series are not built into SAS, which is a statistical software
program frequently used in pharmaceutical development. Therefore, if SAS is used, a calculation program for hypergeometric series must be developed.

It is easy to calculate the probability for using the approximation method. This is an advantage when the approximate probability is used. Conversely, when the difference in the sample proportions is small and the sample sizes are unbalanced, the accuracy the approximation method is poor. That is, the accuracy of the probability of the approximation method depends on the sample size.

This study showed that the accuracy of the MCMC method was greater than that of the approximation method. Moreover, the probability of the MCMC method can be easily calculated using SAS. In addition, it is possible to use the non-conjugate prior for the prior distribution in the MCMC method. The authors consider this as one of the advantages of the MCMC method.

References


