A Credible interval for the proportion difference

Youhei Kawasaki\textsuperscript{1} and Etsuo Miyaoka\textsuperscript{2}

\textsuperscript{1}Graduate School of Science, Tokyo University of Science
\textsuperscript{2}Department of Mathematics, Tokyo University of Science

1 Introduction

The statistical inference for the difference between two independent binominal proportions has been frequently discussed in papers. This interest arises from the fact that the actual confidence level does not correspond to the nominal confidence level. The approximate confidence interval is constructed by using the asymptotical argument. Therefore, the finite sample size, particularly small sample sizes, creates difficulties. Many authors have shown that the actual confidence level is lower than the nominal confidence level.

On the other hand, the Bayesian approach also has been applied to statistical inference of the binominal proportion. Agresti and Caffo (2000) and Pan (2002) attempted the improvement of the Wald interval by using the Bayes estimator. Agresti and Min (2005) studied the frequentist performance of Bayesian intervals for comparing the proportion of two independent binominal samples and found that the Jeffreys prior performance is as good as the score interval.

However, these works have not used an accurate posterior pdf for the difference between two independent binominal proportions. We show the expression of the accurate posterior pdf in this study. We calculate an HPD credible interval by using this expression. In addition, we calculate the approximate credible interval, and compare both the HPD credible interval and the approximate credible interval.

2 Credible Interval

Let $X_1$ and $X_2$ denote a binominal random variable for $n_1$ trials and $n_2$ trials and parameter $\pi_1$ and $\pi_2$, respectively. The conjugate prior density for $\pi_i$ ($i = 1, 2$) is the beta distribution with parameters $\alpha_i$ and $\beta_i$, where parameters $\alpha_i > 0$ and $\beta_i > 0$. The posterior pdf for $\pi_i$ is given by

$$g_i(\pi_{i,\text{post}}) = \frac{1}{B(a_i, b_i)} \pi_i^{a_i-1}(1 - \pi_{i,\text{post}})^{b_i-1},$$

where $a_i = \alpha_i + x_i$, $b_i = n_i - x_i + \beta_i$ and $B(a, b)$ denotes the beta function. Let $\pi_{i,\text{post}}$ denote the binominal proportion in the posterior density.

2.1 Approximate credible interval

An approximate $100(1 - \alpha)\%$ credible interval for $\delta = \pi_{1,\text{post}} - \pi_{2,\text{post}}$ is

$$\mu_{1,\text{post}} - \mu_{2,\text{post}} \pm z_\alpha/2 \sqrt{\nu_{1,\text{post}} + \nu_{2,\text{post}}},$$

where $z_\alpha$ is the $100(1 - \alpha)$th percentile point of the standard normal density function, $\mu_{i,\text{post}} = a_i/(a_i + b_i)$ denotes the posterior mean of $\pi_i$, and

$$\nu_{i,\text{post}} = \frac{\mu_{i,\text{post}}(1 - \mu_{i,\text{post}})}{a_i + b_i + 1}$$
which is the posterior variance of \( \pi_i \).

### 2.2 Highest posterior density credible interval

We do the inference about \( \delta \) by using the exact posterior distribution, which we can obtain from the posterior distributions. Letting \( f_\delta(\delta) \) denote the posterior distribution of \( \delta \), we define a \( 100(1 - \alpha)\% \) as any interval \((c_l, c_u)\), where

\[
\int_{c_l}^{c_u} f_\delta(\delta) d\delta = 1 - \alpha.
\]

If for any \( \delta_1 \in (c_l, c_u) \) and \( \delta_2 \notin (c_l, c_u) \), \( f_\delta(\delta_1) \geq f_\delta(\delta_2) \), then we call \((c_l, c_u)\) a \( 100(1 - \alpha)\% \) HPD credible interval. We will show the exact posterior pdf \( f_\delta(\delta) \) for the difference between two independent binominal proportions in current day.

### 3 Some Numerical Results

In this section, we calculate the HPD credible interval and the approximate credible interval. Then we give some comparisons with the credible interval and the confidence interval.

We show the HPD credible interval and the approximate credible interval. We took up some actual examples in Table 1. The first example is a vaccine study (Chan and Zhang (1999)). The second example was from an influenza vaccine study reported by Fries et al. (1993). Moreover, we calculate the Exact confidence interval and the Newcombe Hybrid confidence interval as references. In Tables 1, the HPD credible interval always has a narrow interval length, compared with the approximate credible interval. The widths in the HPD credible interval and the Newcombe Hybrid interval appear to be almost the same. In general, the Newcombe Hybrid interval is recommended. Hence, we believe that the HPD credible interval will also be recommended.

<table>
<thead>
<tr>
<th>Sampling</th>
<th>Interval</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 = 18, x_1 = 17 )</td>
<td>HPD credible</td>
<td>0.042</td>
<td>0.543</td>
</tr>
<tr>
<td></td>
<td>Approximate credible</td>
<td>0.048</td>
<td>0.552</td>
</tr>
<tr>
<td></td>
<td>Newcombe Hybrid</td>
<td>0.059</td>
<td>0.563</td>
</tr>
<tr>
<td></td>
<td>Exact</td>
<td>-0.019</td>
<td>0.629</td>
</tr>
<tr>
<td>( n_1 = 10, x_1 = 12 )</td>
<td>HPD credible</td>
<td>-0.013</td>
<td>0.595</td>
</tr>
<tr>
<td></td>
<td>Approximate credible</td>
<td>-0.017</td>
<td>0.605</td>
</tr>
<tr>
<td></td>
<td>Newcombe Hybrid</td>
<td>-0.009</td>
<td>0.587</td>
</tr>
<tr>
<td></td>
<td>Exact</td>
<td>-0.056</td>
<td>0.654</td>
</tr>
</tbody>
</table>

### 4 Conclusion

In this study, we calculate the HPD credible interval by using the exact expression. The HPD credible interval and the approximate credible interval always have narrow interval length compared with the exact confidence interval. We effortlessly calculate the HPD credible interval by using the expression. Therefore, the biostatistician facilitates statistical inference by employing the HPD credible interval, thus furthering medical progress.