Best Power-divergence Confidence Interval for a Binomial Proportion

Shaobo Jin
Title: Best Power-divergence Confidence Interval for a Binomial Proportion
Author: Shaobo Jin
E-mail: shaobo.jin@statistics.uu.se
Best Power-divergence Confidence Interval for a Binomial Proportion

Shaobo Jin*

Abstract

The confidence interval for a binomial proportion based on the power-divergence family is considered in this paper. The properties of the confidence intervals are studied in detail. Several choices of the prefixed parameter $\lambda$ are also studied. Numerical results indicate that aligning the mean coverage probability to the nominal value may not be a suitable criterion to choose a $\lambda$ for the power-divergence family. Maximizing the confidence coefficient is a good alternative which is better than some of the recommended competitors in the literature. We can also control $\lambda$ and the significance level $\alpha$ simultaneously to have unbiased intervals in the long run. Edgeworth expansions for the coverage probability and expected length are also derived.

Keywords: confidence interval estimation, mean coverage probability, confidence coefficient, coverage adjustment, Edgeworth expansion

*Department of Statistics, Uppsala University, Box 513, SE-75120, Sweden
Email: shaobo.jin@statistik.uu.se
1 Introduction

The problem of confidence interval estimation of a binomial proportion is one of the most classic but unsolved problems. Assume \( X \sim Bin(n, p) \), where \( n \) is the sample size and \( p \) is the proportion that we are interested in. Because of the discreteness of the binomial random variable, the resulting confidence interval is quite erratic and the coverage is unsatisfactory for some values of \( p \). Statisticians have proposed many methods to solve this problem. The most famous one is the Wald interval \( \hat{P} \pm \kappa \left[ \hat{P}(1 - \hat{P})/n \right]^{1/2} \), where \( \hat{P} = X/n \) and \( \kappa \) is the 100(1 – \( \alpha/2 \)) percentile of a standard normal distribution. The Wald interval is based on the normal approximation and it is easy to apply, hence it is widely accepted. However, Brown et al. (2001) found out that the Wald interval is extremely unsatisfactory. See the papers that they cited when discussing the Wald interval as well.

In order to cope with these problems many alternatives are available in the literature. Among those the most famous methods are the Wilson interval (Wilson, 1927), the Clopper-Pearson interval (Clopper and Pearson, 1934), the Agresti-Coull interval (Agresti and Coull, 1998) and the Bayesian posterior interval with Jeffreys prior (We will call it Jeffreys interval for short through this paper). Plenty of new methods are still coming out, such as Reiczigel (2003); Geyer and Meeden (2005) and Yu (2012). In order to assess different methods, it is common to treat \( p \) as a random variable with a density function \( f(p) \), then the mean coverage probability is defined as

\[
MCP = E_p[C_n(p)] = \int_0^1 C_n(p)f(p)dp, \quad (1)
\]
where
\[ C_n(p) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} I\{L(x)<p<U(x)\} \] (2)
is the coverage probability if the true proportion is \( p \) and \( L(x) \) and \( U(x) \) are the lower limit and the upper limit of the confidence interval respectively when we observe \( x \). The corresponding mean length is
\[ MLENGTH = \int_{0}^{1} E[U(x) - L(x)] f(p) dp. \] (3)
where
\[ E[U(x) - L(x)] = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} [U(x) - L(x)] \] (4)
is the expected length. Besides, the confidence coefficient is an important criterion which represents the most dangerous case that we may get. It is defined as
\[ CC = \inf_{p \in (0,1)} C_n(p). \] (5)

Brown et al. (2001) recommended either the Wilson interval or the Jeffreys interval for \( n \leq 40 \) and the Wilson interval, the Jeffreys interval and the Agresti-Coull interval otherwise. Recommended by other authors as well, the Wilson interval has been used as a benchmark for this problem, especially when \( p \) is near 0.5. However, after comparing 20 methods with explicit solutions, Pires and Amado (2008) recommended the Agresti-Coull interval. The Agresti-Coull interval in their paper is different from Brown et al. (2001). The expression in Brown et al. (2001) is used in this paper.
It is well-known that the Wilson interval has small confidence coefficient when the true $p$ is small. However, the over-estimation in the sense of mean coverage probability of the Wilson interval is stated in Newcombe and Nurminen (2011). We would like to find a method to overcome the drawbacks of the Wilson interval. We mainly focus on a family of divergence measure, the power-divergence (PD) family, proposed by Cressie and Read (1984). It contains the Pearson $\chi^2$ statistic and the log-likelihood statistic as special cases. The PD measure has been used to model multinomial proportions, see Medak and Cressie (1991) and Hou et al. (2003). However, there are no systematic studies of the PD intervals for a binomial proportion. We would like to investigate the PD interval for a binomial proportion in this paper.

The paper is organized as follows. In section 2, the PD family is introduced and properties regarding the confidence interval are studied. Numerical results are given in section 3. In section 4, a method to have unbiased confidence intervals are proposed. Edgeworth expansions for the coverage probability and expected length are provided in section 5. A conclusion ends the paper.

2 Power-divergence Family

Cressie and Read (1984) proposed the PD statistic

$$D(\lambda, d) = \frac{2n}{\lambda(\lambda + 1)} \left( \sum_{i=1}^{d} \frac{\hat{P}_i^{\lambda+1}}{P_i^\lambda} - 1 \right); \ -\infty < \lambda < \infty,$$

(6)

where $\hat{P}_i = X_i/n$ for $i = 1, 2, \ldots, d$, to model the multinomial distribution with $d$ categories. When $\lambda = 0, -1$, it is defined by continuity. When $d = 2$ it
is just the goodness-of-fit statistic for the binomial variable. Further, it is a convex function in $\hat{P}_i/p_i - 1$ and convex in $p_i$. By choosing the values of $\lambda$, we can obtain many famous goodness-of-fit statistics for the multinomial distribution. For instance when $\lambda = 1$, it is the Pearson $\chi^2$ statistic and when $\lambda = 0$ it is the likelihood ratio statistic. For other values of $\lambda$, see Read and Cressie (1988) and Medak and Cressie (1991). All the members of the PD family have the same asymptotic $\chi^2$ distribution with $d-1$ degrees of freedom for a fixed $d$ and $\lambda$ (Read and Cressie, 1988; Basu and Sarkar, 1994).

2.1 Power-divergence Confidence Interval

The confidence region derived from the PD family is

$$
\frac{2n}{\lambda(\lambda + 1)} \left( \sum_{i=1}^{d} \frac{\hat{P}_i^{\lambda+1}}{p_i^\lambda} - 1 \right) < \chi^2_{1-\alpha}(d-1),
$$

where $\chi^2_{1-\alpha}(d-1)$ is the $1-\alpha$ quantile of a $\chi^2$ distribution with $d-1$ degrees of freedom. In general, inequality (7) is a confidence region, which is however an interval when $d = 2$. Especially, when $\lambda = 1$, it is just the Wilson interval

$$
p = \frac{\chi^2_{1-\alpha}(1) + 2X \pm \sqrt{\chi^2_{1-\alpha}(1)[\chi^2_{1-\alpha}(1) + 4X(n - X)/n]}}{2[n + \chi^2_{1-\alpha}(1)]}.
$$

However, equation (7) cannot guarantee the existence of two solutions within $[0, 1]$. For some negative values of $\lambda$, we can only find one solution. For example, when $\lambda = -2$, $n = 50$, $x = 1, 2, 3$ and $\alpha = 0.05$ we only have one solution. The following proposition provides sufficient conditions of the existence of two solutions.
Proposition 1. In the binomial case consider

\[ D(\lambda, 2) - \chi^2_{1-\alpha} = 0. \]  \hspace{1cm} (8)

1. When \( \lambda \in [0, \infty) \), if \( \hat{p} \in (0, 1) \), it has two solutions within \([0, 1]\).

2. When \( \lambda \in (-1, 0) \) and \( \hat{p} \in (0, 1) \), if \( \frac{2n}{\lambda(\lambda+1)} [(1 - \frac{1}{n})^{\lambda+1} - 1] > \chi^2_{1-\alpha}(1) \), it has two solutions within \([0, 1]\).

3. When \( \lambda = -1 \) and \( \hat{p} \in (0, 1) \), if \( -2n \log(1 - \frac{1}{n}) > \chi^2_{1-\alpha}(1) \), it has two solutions within \([0, 1]\).

4. When \( \lambda < -1 \), and \( \hat{p} \in (0, 1) \), if \( \frac{2n}{\lambda(\lambda+1)} [(1 - \frac{1}{n})^{\lambda+1} - 1] > \chi^2_{1-\alpha}(1) \), it has two solutions within \([0, 1]\).

Proposition 1 guarantees that we can always find the lower limit and upper limit when \( \lambda \in [0, \infty) \) and \( \hat{p} \in (0, 1) \). However, we need to define \( L(0) = 0 \) and \( U(n) = 1 \). This is because when \( \hat{p} = 0 \), \( \lim_{p \to 0} d(p) = 0 \), so only one solution can be found within \([0, 1]\) (the same for \( \hat{p} = 1 \)). This modification is applied to all \( \lambda \in R \). For other values of \( \lambda \), the existence of solutions depends on some inequalities. Especially, when \( \lambda \leq \max\{-2/\chi^2_{1-\alpha}(1), -1\} \) the inequality is always fulfilled. For those \( \lambda \in (-1, 0) \) without two solutions, we let \( L = 0 \) if the lower limit is missing and \( U = 1 \) if the upper limit is missing. For \( \lambda < -1 \), these conditions are rarely satisfied. This partly explains the statement made by Read and Cressie (1988) that a reasonable \( \lambda \) lies in \((-1, 2]\). Therefore we only consider \( \lambda > -1 \) in the paper.

In order to use the PD interval we need to decide the value of \( \lambda \). Read and Cressie (1988) studied the choice of \( \lambda \) for hypothesis testing purpose and showed that \( \lambda = 2/3 \) was a nice compromise between the Pearson \( \chi^2 \) test and
the likelihood ratio test. Medak and Cressie (1991) recommended $\lambda = 1/2, 2/3$ to construct confidence regions for multinomial distributions with three categories. In these two studies, the hypothesis testing and confidence regions are performed under some pre-fixed $\lambda$ regardless the sample sizes. Hou et al. (2003) chose the $\lambda$ such that the coverage probability was at least $1 - \alpha$ and achieved the smallest length. However, since we do not know the true value of $p$, the coverage probability is unknown. In the simulation study, they used the sample proportions as the 'true' cell probability to continue the study. This can be viewed as the best $\lambda$ for the observed $x$.

2.2 Properties of the Power-divergence Confidence Interval

The mean coverage probability is a function of $\lambda$ for the PD interval, so a more appropriate notion is $MCP(\lambda)$. We start with a lemma considering the properties of the limits. All the proofs are placed in the appendix.

Lemma 1. For fixed $x$ and $n$, $L$ and $U$ are continuous and differentiable in $\lambda > -1$.

By the aid of Lemma 1, we can establish the following theorem.

Theorem 1. When $\lambda > -1$, $MCP(\lambda)$ is a continuous function of $\lambda$ for all $n$ and $\alpha$, and it is differentiable with respect to $\lambda$.

Newcombe and Nurminen (2011) argued that the $MCP$ should be $1 - \alpha$ which means unbiasedness in the long run. Our goal is to prove the existence of $\lambda$ such that $MCP(\lambda) = 1 - \alpha$. We have already showed in Theorem 1.
that $MCP(\lambda)$ is continuous in $\lambda$. If we can show that for certain choices of $\lambda$, $MCP(\lambda)$ can be less than $1 - \alpha$ and for the others, $MCP(\lambda)$ is greater than $1 - \alpha$, then we meet our goal because of the continuity property. The former is shown in the following lemma.

**Lemma 2.** For any $\alpha$ and $n$, we can always find a $\lambda$ such that $MCP(\lambda) < 1 - \alpha$.

Unfortunately, we do not manage to prove the existence of $\lambda$ where $MCP(\lambda) > 1 - \alpha$ in the general case. Based on these results, we cannot establish the statement that for any $\alpha$ and $n$, there is a $\lambda$ such that $MCP(\lambda) = 1 - \alpha$. Numerical results in the next section provide some evidences supporting our proposals.

**Proposition 2.** If equation (8) have two solutions, the PD interval satisfies $L(x_1) < L(x_2)$ and $U(x_1) < U(x_2)$ if $x_1 < x_2$.

The proof of Proposition 2 also implies that $L(x)$ and $U(x)$ are strictly decreasing functions of $n$ for fixed $x$, $\lambda$, and $\alpha$. Blyth and Still (1983) stated four properties which a good confidence interval estimation should satisfy. From now on we can state that the power-divergence interval fulfills these four properties, namely interval-valued property, equivariant property, monotonicity in $x$ and monotonicity in $n$.

## 3 Numerical Results

The performance of the PD interval is investigated in this part. Two significance levels are used, $\alpha = 0.01$ and 0.05. Sample sizes from $n = 5$ to $n = 200$ are considered. The proportion $p$ is assumed to be uniformly distributed.
Figure 1: Mean coverage probabilities of the power-divergence interval for
\( n = 5 \) (black), \( n = 10 \) (blue), \( n = 50 \) (darkgreen), \( n = 100 \) (magenta) and
\( n = 200 \) (red) with two nominal coverage probabilities 0.95 (left) and 0.99
(right).

3.1 Existence of \( \lambda_{MCP} \)

We have not proven the existence of \( \lambda_{MCP} \) mathematically. However, some
numerical evidences can be seen from Figure 1. In Figure 1, \( MCP(\lambda) \) is
illustrated for five sample sizes with two nominal coverage probabilities. It
is easy to see that we can always find two \( \lambda_{MCP} \)'s. In the later analysis,
the PD(\( \lambda \)) interval refers to a PD interval with a prefixed parameter \( \lambda \) to
emphasize the effect of \( \lambda \), e.g. the PD interval when \( \lambda = \lambda_{MCP} \) is called the
PD(\( \lambda_{MCP} \)) interval. The PD(\( \lambda_{MCP} \))-S interval refers to the smaller \( \lambda_{MCP} \)
while the PD(\( \lambda_{MCP} \))-L interval refers to the larger \( \lambda_{MCP} \). As an example,
some values of \( \lambda_{MCP} \) are tabulated in Table 1 when \( \alpha = 0.05 \).
Table 1: The values of $\lambda_{MCP}$ when $\alpha = 0.05$ for different sample sizes under the uniform prior. Rounding up to 4 decimals.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\lambda$</th>
<th>$n$</th>
<th>$\lambda$</th>
<th>$n$</th>
<th>$\lambda$</th>
<th>$n$</th>
<th>$\lambda$</th>
<th>$n$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.0476</td>
<td>25</td>
<td>0.1890</td>
<td>45</td>
<td>0.2150</td>
<td>65</td>
<td>0.2278</td>
<td>85</td>
<td>0.2359</td>
</tr>
<tr>
<td></td>
<td>1.3266</td>
<td></td>
<td>1.4856</td>
<td></td>
<td>1.5469</td>
<td></td>
<td>1.5847</td>
<td></td>
<td>1.6116</td>
</tr>
<tr>
<td>10</td>
<td>0.1266</td>
<td>30</td>
<td>0.1979</td>
<td>50</td>
<td>0.2189</td>
<td>70</td>
<td>0.2302</td>
<td>90</td>
<td>0.2375</td>
</tr>
<tr>
<td></td>
<td>1.3917</td>
<td></td>
<td>1.5047</td>
<td></td>
<td>1.5578</td>
<td></td>
<td>1.5922</td>
<td></td>
<td>1.6173</td>
</tr>
<tr>
<td>15</td>
<td>0.1585</td>
<td>35</td>
<td>0.2048</td>
<td>55</td>
<td>0.2223</td>
<td>75</td>
<td>0.2323</td>
<td>95</td>
<td>0.2390</td>
</tr>
<tr>
<td></td>
<td>1.4326</td>
<td></td>
<td>1.5208</td>
<td></td>
<td>1.5676</td>
<td></td>
<td>1.5991</td>
<td></td>
<td>1.6226</td>
</tr>
<tr>
<td>20</td>
<td>0.1768</td>
<td>40</td>
<td>0.2104</td>
<td>60</td>
<td>0.2252</td>
<td>80</td>
<td>0.2342</td>
<td>100</td>
<td>0.2404</td>
</tr>
<tr>
<td></td>
<td>1.4623</td>
<td></td>
<td>1.5347</td>
<td></td>
<td>1.5765</td>
<td></td>
<td>1.6056</td>
<td></td>
<td>1.6276</td>
</tr>
</tbody>
</table>

3.2 Performance of the Interval

The PD($\lambda_{MCP}$) interval is compared with the Wilson interval, the Agresti-Coull interval and the Jeffreys interval in this section. The coverage probability, length, mean coverage probability, mean length, confidence coefficient and mean absolute error are investigated, where the mean absolute error is defined as

$$MAE = \int_0^1 |C_n(p) - (1 - \alpha)|f(p)dp.$$  \hspace{1cm} (9)

We can see from Figure 2 and 3 that the PD($\lambda_{MCP}$)-L interval is similar as the Wilson interval. This is because the larger value of $\lambda_{MCP}$ is close to or larger than 1, which is the Wilson interval. When the positive $\lambda_{MCP}$ is less than 1, the coverage properties are much better than the Wilson interval,
which are close to the Jeffreys interval. The Agresti-Coull interval has larger coverage probabilities in general.

When it comes to lengths, we observe from Figure 4 that the PD-S interval and the Jeffreys interval are the two shortest intervals when we have a extreme true $p$. As $p$ approaching 0.5, they become the longest. The Agresti-Coull interval is the widest under extreme $p$’s and becomes more and more close to the Wilson interval as $p$ goes to 0.5.

Figure 5 shows the mean coverage probability, the mean length and the mean absolute error. The Wilson interval overestimates the mean coverage probability for a 95% confidence interval and it underestimates that at the 99% level. The PD(\(\lambda_{MCP}\))-S interval and the PD(\(\lambda_{MCP}\))-L interval have the same mean coverage probability hence they overlap with each other. Among the other methods, the Jeffreys interval is the closest to the nominal value. Severe overestimation is always a problem for the Agresti-Coull interval. It is inconsistent when a 99% confidence interval is required. The Wilson interval, the PD(\(\lambda_{MCP}\)) intervals and the Jeffrey’s interval share similar mean lengths (Figure 5b). The Agresti-Coull interval is always the longest. When it comes to the mean absolute error (Figure 5c), the Wilson interval has the smallest mean absolute error when $\alpha = 0.05$. The PD(\(\lambda_{MCP}\))-L interval has a similar bias as the Wilson interval and attains the smallest bias when $\alpha = 0.01$.

Exact confidence coefficients are shown in Table 2. The method in [Wang 2007, 2009] is used to calculate the confidence coefficients. Exact confidence coefficient means that the values are true values, not estimates. The PD(\(\lambda_{MCP}\))-L interval is unsatisfactory when $\alpha = 0.05$ which has a decreasing confidence coefficient. The PD(\(\lambda_{MCP}\))-S interval has higher confidence
Figure 2: Coverage probabilities of the Wilson interval and PD($\lambda_{MCP}$) intervals when the sample size is $n = 25$ with the nominal probabilities 0.95 (left) and 0.99 (right).
Figure 3: Coverage probabilities of the Agresti-Coull interval and the Jeffreys interval when the sample size is $n = 25$ with the nominal probabilities 0.95 (left) and 0.99 (right).
Figure 4: Lengths of the Wilson interval (black), the Agresti-Coull interval (blue), the Jeffreys interval (darkgreen), the PD($\lambda_{MCP}$)-S interval (magenta) and the PD($\lambda_{MCP}$)-L interval (red) when the sample size is $n = 25$ with the nominal probabilities 0.95 (left) and 0.99 (right).

coefficients than the other PD intervals, close to the Jeffreys interval.

**Remark 1.** The PD interval is sensitive to the choice of the priors. Under a prior other than the uniform prior, e.g. Jeffreys prior $\text{Beta}(1/2, 1/2)$, it is likely that $\lambda_{MCP}$ is negative. A negative $\lambda$ leads to short confidence intervals, however the mean absolute error is large. Such a $\lambda$ is less than $-0.5$. If an asymmetric prior is assumed, e.g. $\text{Beta}(1/4, 20)$ to model the defective percentage, the confidence coefficients of the PD($\lambda_{MCP}$)-S interval become extremely unsatisfactory. This is due to the fact that $\lambda_{MCP}$ is negative under the $\text{Beta}(1/4, 20)$ prior. It seems a good choice of $\lambda$ is between 0 and 1, or perhaps an even shorter support $[0, 0.5]$. 
Figure 5: Comparisons of the Wilson interval (black), the Agresti-Coull interval (blue), the Jeffreys interval (darkgreen), the PD($\lambda_{MCP}$)-S interval (magenta) and the PD($\lambda_{MCP}$)-L interval (red) with the nominal probabilities 0.95 (left) and 0.99 (right).
Table 2: Exact confidence coefficients of the Wilson interval, the $PD(\lambda_{MCP})$ intervals, the Agresti-Coull interval and the Jeffreys interval.

<table>
<thead>
<tr>
<th>Interval</th>
<th>$n = 5$</th>
<th>$n = 10$</th>
<th>$n = 15$</th>
<th>$n = 20$</th>
<th>$n = 25$</th>
<th>$n = 50$</th>
<th>$n = 100$</th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilson</td>
<td>0.8315</td>
<td>0.8350</td>
<td>0.8360</td>
<td>0.8366</td>
<td>0.8369</td>
<td>0.8375</td>
<td>0.8379</td>
<td>0.8380</td>
</tr>
<tr>
<td>PD($\lambda_{MCP}$)-S</td>
<td>0.8207</td>
<td>0.8538</td>
<td>0.8458</td>
<td>0.8786</td>
<td>0.8784</td>
<td>0.8772</td>
<td>0.8762</td>
<td>0.8757</td>
</tr>
<tr>
<td>PD($\lambda_{MCP}$)-L</td>
<td>0.8043</td>
<td>0.8029</td>
<td>0.8011</td>
<td>0.7994</td>
<td>0.7981</td>
<td>0.7936</td>
<td>0.7891</td>
<td>0.7848</td>
</tr>
<tr>
<td>Agresti-Coull</td>
<td>0.8940</td>
<td>0.9239</td>
<td>0.9316</td>
<td>0.9292</td>
<td>0.9304</td>
<td>0.9345</td>
<td>0.9380</td>
<td>0.9408</td>
</tr>
<tr>
<td>Jeffreys</td>
<td>0.8924</td>
<td>0.8681</td>
<td>0.8961</td>
<td>0.8934</td>
<td>0.8905</td>
<td>0.8838</td>
<td>0.8801</td>
<td>0.8781</td>
</tr>
<tr>
<td>Wilson</td>
<td>0.8859</td>
<td>0.8876</td>
<td>0.8882</td>
<td>0.8884</td>
<td>0.8886</td>
<td>0.8889</td>
<td>0.8891</td>
<td>0.8891</td>
</tr>
<tr>
<td>PD($\lambda_{MCP}$)-S</td>
<td>0.9350</td>
<td>0.9579</td>
<td>0.9639</td>
<td>0.9687</td>
<td>0.9686</td>
<td>0.9669</td>
<td>0.9727</td>
<td>0.9728</td>
</tr>
<tr>
<td>PD($\lambda_{MCP}$)-L</td>
<td>0.9168</td>
<td>0.9171</td>
<td>0.9167</td>
<td>0.9164</td>
<td>0.916</td>
<td>0.9149</td>
<td>0.9137</td>
<td>0.9127</td>
</tr>
<tr>
<td>Agresti-Coull</td>
<td>0.9446</td>
<td>0.9681</td>
<td>0.9767</td>
<td>0.9807</td>
<td>0.9831</td>
<td>0.9872</td>
<td>0.9875</td>
<td>0.9880</td>
</tr>
<tr>
<td>Jeffreys</td>
<td>0.9629</td>
<td>0.9639</td>
<td>0.9642</td>
<td>0.9643</td>
<td>0.9644</td>
<td>0.9646</td>
<td>0.9647</td>
<td>0.9647</td>
</tr>
</tbody>
</table>
3.3 Relax the Nominal Coverage

In this part, we allow some discrepancies from the nominal mean coverage probability and try to see what we can gain. This is the conservativeness in the sense of mean coverage and it is not necessarily conservative for all \( p \). Four methods of choosing \( \lambda \) are considered:

1. Conservative \( \lambda \) with the largest \( \text{MCP}/\text{MLENGTH} \), denoted by \( \lambda_{ME} \).

2. Conservative \( \lambda \) with the shortest mean length, denoted by \( \lambda_{SL} \).

3. Most conservative \( \lambda \), denoted by \( \lambda_{MC} \).

4. Conservative \( \lambda \) with the largest confidence coefficient, denoted by \( \lambda_{CC} \).

The routines of analysis are the same as what we did in Section 3.2, hence we only briefly present the important results here. The \( \lambda_{ME} \) and \( \lambda_{SL} \) coincide with \( \lambda_{MCP} \) often or at least around \( \lambda_{MCP} \). Thus they are (nearly) unbiased in the long run. For these methods, confidence coefficients are all small and mean absolute errors are large.

**Remark 2.** \( \lambda_{MC} \) leads to too conservative methods under the asymmetric prior Beta\((1/4, 20)\) prior.

Since the usage of power-divergence intervals is hesitated because of the confidence coefficient, it is natural to look into the PD(\( \lambda_{CC} \)) interval. The influence of priors are not as strong as other selection methods for the PD(\( \lambda_{CC} \)) interval. The mean coverage probability and the mean length are quite similar as the Jeffreys interval. The similarity can also be observed from the
figure of mean absolute errors. Further, the confidence coefficient is better than the other PD intervals, which is close to but even better than the Jeffreys interval. However, as we increase the sample size, the confidence coefficient doesn’t improve much or even decreases sometimes.

4 Control $\lambda$ and $\alpha$ Simultaneously

Up to now, we only focus on the choice of $\lambda$ for a fixed $\alpha$. Aligning the $MCP(\lambda)$ to the nominal value does not always result in a good confidence interval. Some bias can greatly improve the interval estimation. Thulin (2013) modified the Clopper-Pearson interval by choosing another significance level $\alpha'$ such that $MCP(\alpha') = 1 - \alpha$. This idea can be applied to the PD intervals controlling $\lambda$ and $\alpha$ simultaneously such that for a given $n$, $MCP(\lambda, \alpha') = 1 - \alpha$ and the confidence coefficient is maximized.

**Theorem 2.** Within the power-divergence family, we can always find an $\alpha'$ as a function of $n$ and $\lambda$ such that $MCP(\lambda, \alpha') = 1 - \alpha$.

Theorem 2 allows us to manipulate the curve of $MCP(\lambda, \alpha)$ to the nominal value. In practice we choose 300 $\lambda$'s uniformly in $(-1, 2]$ and for each $\lambda$ we try to find a $\alpha'$ aligning $MCP(\lambda, \alpha')$ to $1 - \alpha$ at a fixed sample size. Then we compare the confidence coefficients to select the best pair $(\lambda, \alpha')$. However, in practice when $\lambda$ is close to $-1$ or too large, we cannot find such $\alpha'$ because the numerical methods are unreliable for too small values of $\alpha'$. These combinations are deleted from the analysis.

In Figure 6a the mean lengths can be seen. The power-divergence intervals are always among the shortest methods. Figure 6b shows the small
Figure 6: Comparisons of the Wilson interval (black), the Agresti-Coull interval (blue), the Jeffreys interval (darkgreen), the PD($\lambda_{MCP}$) interval (magenta) when $\lambda$ and $\alpha$ are controlled simultaneously with the nominal probabilities 0.95 (left) and 0.99 (right).

Table 3: Exact confidence coefficients of the PD intervals with the technique of Thulin (2013) for $1 - \alpha = 0.95$ and 0.99.

<table>
<thead>
<tr>
<th>$1 - \alpha$</th>
<th>n = 5</th>
<th>n = 10</th>
<th>n = 15</th>
<th>n = 20</th>
<th>n = 25</th>
<th>n = 50</th>
<th>n = 100</th>
<th>n = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>0.8575</td>
<td>0.8697</td>
<td>0.8970</td>
<td>0.8956</td>
<td>0.8937</td>
<td>0.8888</td>
<td>0.8998</td>
<td>0.9037</td>
</tr>
<tr>
<td>0.99</td>
<td>0.9603</td>
<td>0.9630</td>
<td>0.9684</td>
<td>0.9722</td>
<td>0.9715</td>
<td>0.9738</td>
<td>0.9731</td>
<td>0.9728</td>
</tr>
</tbody>
</table>
biases when $\alpha = 0.01$ comparing with the Wilson interval. When $\alpha = 0.05$, the Wilson interval possesses the smallest bias. In general, the Jeffreys interval always has small bias. The PD interval has high confidence coefficients (Table 3), only smaller than the Agresti-Coull interval. Reasonable $\lambda$'s are often between 0.1 and 0.5.

**Remark 3.** Under the asymptotic prior Beta$(1/4, 20)$, the PD intervals are dangerous if a 95% confidence interval is required. When $n$ is small, unreasonable pairs of $(\lambda, \alpha')$ occur with $\lambda$ larger than 1 and $\alpha'$ in $(0.15, 0.35)$.

## 5 Some Expansions

In this section, we investigate the Edgeworth expansions of the coverage probability and the expected length for a fixed $p$. The expansions can help us to understand the numerical results in the previous section. Define $q = 1 - p$ and

$$g(p, z) = np + z\sqrt{npq} - \lfloor np + z\sqrt{npq} \rfloor,$$

where $\lfloor np + z\sqrt{npq} \rfloor$ is the largest integer which is smaller than $np + z\sqrt{npq}$.

Further let

$$Q_{21}(l, u) = 1 - g(p, l) - g(p, u),$$

$$Q_{22}(l, u) = \left[-\frac{1}{2} g^2(p, u) + \frac{1}{2} g(p, u) - \frac{1}{2} g^2(p, l) + \frac{1}{2} g(p, l) - \frac{1}{6}\right].$$
5.1 Expansion of the Coverage Probability

Theorem 3. For a fixed \( p \in (0, 1) \), \( \alpha \in (0, 1) \) assume equation [8] have two solutions, the coverage probability of the power-divergence interval is

\[
P(p \in (L_{PD}, U_{PD})) = 1 - \alpha + [g(p, l_{PD}) - g(p, u_{PD})] \phi(k(npq)^{-1/2})
+ \{(4pq - 1)\lambda^2\kappa^5 + [\lambda(\lambda - 1)(2 - 11pq) + \lambda(7 - 22pq)]\kappa^3
+ (6pq - 6)\kappa\} \phi(k)(36npq)^{-1} + \frac{1}{6}(1 - 2p)(\lambda\kappa^2 - 3)Q_{21}(l_{PD}, u_{PD})
+ Q_{22}(l_{PD}, u_{PD})\} \kappa\phi(k)(npq)^{-1} + O(n^{-3/2}),
\]

(10)

where \( l_{PD} \) (\( u_{PD} \)) is the lower (upper) limit for \( z = \sqrt{n}(\hat{p} - p)/\sqrt{pq} \) induced from \( L_{PD}(x) < p < U_{PD}(x) \) which is defined in equation [17] in the appendix.

In equation (10), the term \( 1 - \alpha \) is the nominal value. The terms including \( g \), \( Q_{21} \) and \( Q_{22} \) are oscillation parts which cause jumps in the coverage probability. The bias is \( O(n^{-1}) \). In the bias term, there is a polynomial of \( \kappa \) involved. The term \( \kappa^5 \) is always negative and is a monotone function in \( |\lambda| \). The term \( \kappa \) is also negative and is independent of \( \lambda \). The remaining \( \kappa^3 \) term, as a decreasing function in \( \lambda \in [0, 1] \), is always positive. Thus a good choice of \( \lambda \in [0, 1] \) leads to a smaller bias. When \( \lambda = 0 \), only the negative bias term \( (6pq - 6)\kappa\phi(k)(36npq)^{-1} \) remains.
5.2 Expansion of the Expected Length

Theorem 4. For a fixed $\alpha \in (0,1)$ and $p \in (0,1)$ assume equation (8) have two solutions, the expansion of the length is given by

$$E(U_{PD} - L_{PD}) = 2\kappa(pq)^{1/2}n^{-1/2} - \frac{1}{4}\kappa(pq)^{-1/2}n^{-3/2}$$

$$+ \frac{1}{36}(\lambda + 2)\kappa^3[2\lambda + 1 - (11\lambda + 13)pq]n^{-3/2} + O(n^{-2}). \quad (11)$$

Further if we assume $p$ is Beta$(a,b)$ distributed, then if $a > \frac{1}{2}$ and $b > \frac{1}{2}$

$$\int_0^1 E(U_{PD} - L_{PD})f(p; a, b)dp = 2\kappa \frac{B(a + \frac{1}{2}, b + \frac{1}{2})}{B(a, b)}n^{-1/2}$$

$$+ \frac{1}{36}(\lambda + 2)\kappa^3n^{-3/2} \left[ (2\lambda + 1) \frac{B(a - \frac{1}{2}, b - \frac{1}{2})}{B(a, b)} \right. \right.$$  

$$\left. - (11\lambda + 13) \frac{B(a + \frac{1}{2}, b + \frac{1}{2})}{B(a, b)} \right]$$

$$- \frac{1}{4}\kappa \frac{B(a - \frac{1}{2}, b - \frac{1}{2})}{B(a, b)}n^{-3/2} + O(n^{-2}). \quad (12)$$

The prefixed parameter $\lambda$ enters the mean length only through the term $O(n^{-3/2})$. Note that $B(a - \frac{1}{2}, b - \frac{1}{2}) > B(a + \frac{1}{2}, b + \frac{1}{2})$, then when $\lambda$ is negative with a large value, the term involving $\lambda$ in equation (12) is negative. This gives rise to a short interval. If $2B(a - \frac{1}{2}, b - \frac{1}{2}) > 11B(a + \frac{1}{2}, b + \frac{1}{2})$ then the second term in equation (12) is an increasing function of $\lambda > -\frac{1}{2}$. For certain choice of $\lambda$, the $O(n^{-3/2})$ term which involves $\lambda$ disappears. Especially when $a = b = 1$, $2B(a - \frac{1}{2}, b - \frac{1}{2}) > 11B(a + \frac{1}{2}, b + \frac{1}{2})$ and $\lambda = 1$ the $O(n^{-3/2})$ disappears.
6 Conclusion and Discussion

In this paper, we studied the PD interval for a binomial proportion in detail. Properties of the PD intervals are studied. We also discussed several different criteria of choosing a prefixed parameter $\lambda$. Expansions for the coverage probability and the expected length are also derived for the power-divergence family to help us understand the coverage properties. No matter what $\lambda$ we choose, the downward spike is inevitable. We can only alleviate the unwelcome feature. If we align the mean coverage probability to the nominal value with a fixed $\alpha$, it will make things worse. Hence within the PD family, such equating is not a good idea. We should allow some bias to get better alternatives. Especially, $\lambda_{CC}$, which leads to the conservative interval with largest confidence coefficients, is a good choice. It is slightly conservative with a higher confidence coefficient than the other members (including the Wilson interval). When $\alpha = 0.05$, such values are most likely to be less than 0.5 based on our numerical study up to sample size 200. When $n > 50$ the downward trend of $\lambda$ is clear, which is 0.3314 when $n = 200$. When $\alpha = 0.01$, it is most likely between 0.1 and 0.2. Thus in contradiction with the recommendation in Medak and Cressie (1991), the $\lambda$ induced by $n$ has nice properties when $\lambda \in [0.1, 0.5]$ roughly. The exact value will be decided by the specific sample size that we have. Besides, the PD($\lambda_{CC}$) interval is close to the Jeffreys interval under the uniform prior when $\alpha = 0.01$. This may provide another motivation to the Jeffreys interval from a frequentist point of view. If we control $\lambda$ and $\alpha$ simultaneously such that $MCP(\lambda, \alpha') = 1-\alpha$, we can also have nice intervals with short lengths, decent mean absolute biases and high confidence coefficients. A reasonable $\lambda$ still falls in $[0.1, 0.5]$ most
likely.

Based on the results, we can make some suggestions to practitioners who want to construct a confidence interval for a binomial proportion. If the coverage probability is the most important, then the PD($\lambda_{CC}$) interval can serve as a good tool. The absolute bias is relatively small with a decent length. If practitioners want an unbiased interval in the long run, they should control $\lambda$ and $\alpha$ together as in Section 4. The resulting confidence intervals are unbiased in terms of the mean coverage probability. But if there is a strong believe in small values of $p$ and a 95% confidence interval is wanted, then this method is dangerous to use. Alternatives should be considered. Keep in mind that a good $\lambda$ is between 0 and 1 and most likely between 0.1 and 0.5.

As a member of power-divergence family, the Wilson interval with $\lambda = 1$ performs nice when $\alpha = 0.05$ under many criteria except the low confidence coefficient. However when $\alpha = 0.01$, it loses preference. The likelihood ratio interval with $\lambda = 0$ is also a member of the power-divergence family. They represent the score approach and the likelihood approach respectively. Our numerical study shows that neither the Wilson interval nor the likelihood ratio interval is our best choice. None of the six selection methods in the paper indicate the usage of the Wilson interval or the likelihood ratio interval. Optimal $\lambda$’s are in general closer to $\lambda = 0$ which is the likelihood ratio interval.

One obvious drawback of our work is that we have not provided confirmatory mathematical proofs regarding the existence of $\lambda_{MCP}$ in this paper. Most of our studies rely on the assumption that such $\lambda_{MCP}$ exists. Although
we provide some numerical results to support this, numerical instability may still cause some problems.

A Mathematical Appendix

Proof of Proposition 1 In the binomial case,

\[ D(\lambda, 2) - \chi^2_{1-\alpha} = \frac{2n}{\lambda(\lambda + 1)} \left[ \frac{\hat{p}^{\lambda+1}}{p^\lambda} + \frac{(1 - \hat{p})^{\lambda+1}}{(1 - p)^\lambda} - 1 \right] - \chi^2_{1-\alpha}(1) = 0. \quad (13) \]

The notation \(d(p)\) is used in the proof to emphasize that \(p\) is an argument. For all \(\lambda\), \(d(\hat{p}) - \chi^2_{1-\alpha}(1) = -\chi^2_{1-\alpha}(1)\) is negative.

1. If \(\hat{p} \in (0, 1)\), \(D(\lambda, 2) - \chi^2_{1-\alpha}(1)\) is continuous in \(p\) and is a convex function in \(p\). Note that \(d(p) - \chi^2_{1-\alpha}(1)\) converges to \(\infty\) as \(p \to 0\) or 1.

Hence, we have exactly two solutions for all possible \(x = 1, \ldots, n - 1\).

2. When \(\lambda \in (-1, 0)\),

\[
\lim_{p \to 0} d(p) - \chi^2_{1-\alpha}(1) = \frac{2n}{\lambda(\lambda + 1)} \left[ (1 - \frac{1}{n})^{\lambda+1} - 1 \right] - \chi^2_{1-\alpha}(1),
\]

which is an increasing function of \(\hat{p}\). Since the minimum of \(\hat{p}\) is \(1/n\), so the minimum of \(\lim_{p \to 0} d(p) - \chi^2_{1-\alpha}(1)\) is

\[
\frac{2n}{\lambda(\lambda + 1)} \left[ (1 - \frac{1}{n})^{\lambda+1} - 1 \right] > \chi^2_{1-\alpha}(1), \quad (14)
\]

The equation \(d(p) = \chi^2_{1-\alpha}(1)\) has one solution in \([0, \hat{p}]\) for all \(\hat{p} \in (0, 1)\) if equation \((14)\) is larger than 0. Similarly,

\[
\lim_{p \to 1} d(p) - \chi^2_{1-\alpha}(1) = \frac{2n}{\lambda(\lambda + 1)} \left[ \hat{p}^{\lambda+1} - 1 \right] - \chi^2_{1-\alpha}(1),
\]
which is a decreasing function of $\hat{p}$. And consider $\hat{p} = 1 - 1/n$, then the function has the same minimum as we obtained before. Therefore we can find another solution in $(\hat{p}, 1]$.

3. When $\lambda = -1$,

$$
\lim_{p \to 0} d(p) - \chi^2_{1-\alpha}(1) = -2n \log(1 - \hat{p}) - \chi^2_{1-\alpha}(1),
$$

$$
\lim_{p \to 1} d(p) - \chi^2_{1-\alpha}(1) = -2n \log(\hat{p}) - \chi^2_{1-\alpha}(1).
$$

Both limits have the minimum value $-2n \log(1 - \frac{1}{n}) - \chi^2_{1-\alpha}(1)$. Hence if

$$
-2n \log(1 - \frac{1}{n}) > \chi^2_{1-\alpha}(1)
$$

we can always find two solutions.

4. The proof is similar as (2).

\[\square\]

**Proof of Lemma 1** If equation (8) has two solutions, the continuity and the differentiability directly follow from the implicit function theorem, see [Apostol (1974)] and [Rudin (1976)] for details. Otherwise, the implicit function theorem identifies the lower limit or the upper limit and the other one is 0 or 1, which are also continuous and differentiable.

\[\square\]

**Proof of Theorem 1** It suffices to prove that

$$
\int_0^1 p^x (1 - p)^{n-x} I_{\{L(x) < p < U(x)\}} f(p) dp = \int_{L(x)}^{U(x)} p^x (1 - p)^{n-x} f(p) dp
$$

is continuous. Note that both $L$ and $U$ are continuous and differentiable functions in $\lambda$ (Lemma 1), thus equation (15) is also continuous and differentiable in $\lambda$. Hence $MCP(\lambda)$ is continuous and differentiable.

\[\square\]
Proof of Lemma 2. For a fixed $p$ let $\lambda \to \infty$,

$$D(\lambda, 2) = \frac{2n}{\lambda(\lambda + 1)} \left[ \hat{p}^{\lambda+1} p^\lambda + \frac{(1 - \hat{p})^{\lambda+1}}{(1 - p)^\lambda} - 1 \right] \to \infty,$$

since either $\hat{p}/p$ or $(1 - \hat{p})/(1 - p)$ will be larger than 1 unless $\hat{p} = p$. However we have to solve $D(\lambda, 2) = \chi_1^2(1 - \alpha)$ to obtain the limits of the confidence interval. Hence $p$ should be sufficiently close to $\hat{p}$ for both $L$ and $U$ under a large $\lambda$ to keep the ratio $\hat{p}/p$ or $(1 - \hat{p})/(1 - p)$ sufficiently close 1 such that $D(\lambda, 2)$ is not too large. By doing this, we decrease the upper limit and increase the lower limit in the integral $\int_L^U p^x (1 - p)^{n-x} f(p) dp$ leading to smaller values of $MCP(\lambda)$ at last. This implies that we can always find a sufficiently large positive $\lambda$ such that $MCP(\lambda) < 1 - \alpha$.

Proof of Proposition 2. We treat $x$ as a continuous argument. Then the proposition follows directly follows from the implicit function theorem by considering $\partial L/\partial \hat{p}$ and $\partial U/\partial \hat{p}$ respectively.

Proof of Theorem 2. For a given $\lambda$, as $\alpha \to 0$, $L(x) \to 0$ and $U(x) \to 1$, hence

$$MCP(\lambda) \to \sum_{x=0}^{n} \binom{n}{x} \int_0^1 p^x (1 - p)^{n-x} f(p) dp = 1.$$

As $\alpha \to 1$, $L(x) \to \hat{p}$ and $U(x) \to \hat{p}$, then $MCP(\lambda, \alpha) \to 0$. By the aid of Theorem 1, we can draw the conclusion that for any fixed $\lambda > -1$, there is always one $\alpha'$ such that $MCP(\lambda) = 1 - \alpha$.

Proof of Theorem 3. Let $z = \sqrt{n}(\hat{p} - p)/\sqrt{pq}$, then $\hat{p} = p + n^{-1/2}(pq)^{1/2}z$. 

27
And equation \((7)\) with \(d = 2\) is equivalent to

\[
v(z) = \frac{2n}{\lambda(\lambda + 1)} \left\{ p \left[ 1 + n^{-1/2} \left( \frac{q}{p} \right)^{1/2} z \right]^\lambda + 1 
+ q \left[ 1 - n^{-1/2} \left( \frac{p}{q} \right)^{1/2} \right]^\lambda - 1 \right\} - \kappa^2.
\]

The Taylor expansion of \((1 + t)^{\lambda+1}\), when \(\lambda > -1\) and \(\lambda \neq 0, 1\), is

\[
(1 + t)^{\lambda+1} = 1 + (\lambda + 1)t + \frac{1}{2} \lambda(\lambda + 1)t^2 + \frac{1}{6} \lambda(\lambda - 1)(\lambda + 1)t^3 
+ \frac{1}{24} \lambda(\lambda - 1)(\lambda - 2)(\lambda + 1)t^4 + O(t^5).
\]

Therefore

\[
v(z) = z^2 + \frac{1}{3} (\lambda - 1)(1 - 2p)(pq)^{-1/2}n^{-1/2}z^3 
+ \frac{1}{12} (\lambda - 1)(\lambda - 2)(1 - 3pq)(pq)^{-1}n^{-1}z^4 - \kappa^2 + O(n^{-3/2}). \tag{16}
\]

Some algebra shows that \(v(z)\) is a convex function in \(z\). Hence we can find at most two solutions for \(v(z) = 0\). Let \(z = \pm \kappa + b_1 n^{-1/2} + b_2 n^{-1}\) and insert it to \(v(z) = 0\). Both the term \(n^{-1/2}\) and the term \(n^{-1}\) have to be 0. This implies

\[
b_1 = -\frac{1}{6} (\lambda - 1)(1 - 2p)(pq)^{-1/2} \kappa^2; \]

\[
b_2 = \pm \frac{1}{72} (2\lambda - 11\lambda pq + 1 + 2pq)(\lambda - 1)(pq)^{-1} \kappa^3.
\]

Hence the roots of \(v(z) = 0\) can be expressed as

\[
(l_{PD}, u_{PD}) = \pm \kappa - \frac{1}{6} (\lambda - 1)(1 - 2p)(pq)^{-1/2} \kappa^2 n^{-1/2} 
\pm \frac{1}{72} (2\lambda - 11\lambda pq + 1 + 2pq)(\lambda - 1)(pq)^{-1} \kappa^3 n^{-1}. \tag{17}
\]
The two-term Edgeworth expansion for a discrete distribution leads to

\[ P(Z \leq z) = 1 - \frac{\alpha}{2} + \left[ \frac{1}{2} - g(p, z) \right] \phi(k)(npq)^{-1/2} \]

\[ + \left[ \frac{1}{6}(\lambda - 1)(1 - 2p)\kappa^2 + \frac{1}{6}(1 - 2p)(1 - \kappa^2) \right] \phi(\kappa)(npq)^{-1/2} \]

\[ + \left\{ \frac{1}{72}(2\lambda - 11\lambda p + 11\lambda p^2 + 1 + 2p - 2p^2)(\lambda - 1)(pq)^{-1}\kappa^3 \right\} \phi(\kappa)n \]

\[ + \left\{ \frac{1}{36}(\lambda - 1)(1 - 2p)^2(pq)^{-1}\kappa^3(\kappa^2 - 3) \right\} \phi(\kappa)n^{-1} \]

with \( z = l_{PD}, u_{PD} \). See Esseen (1945) and Kolassa (2006) for Edgeworth series for discrete distributions. The coverage probability can be expressed as \( P(Z \leq u_{PD}) - P(Z \leq l_{PD}) \) and some algebra leads to equation (10). For other cases of \( \lambda \)'s, Brown et al. (2002) already showed the case when \( \lambda = 0, 1 \) which can be fitted into equation (10).

**Proof of Theorem 4** First consider \( \lambda \neq 0 \), and let \( t = p/\hat{p} - 1 \), equation (7) with \( d=2 \) is equivalent to

\[ \frac{2n}{\lambda(\lambda + 1)} \left\{ \hat{p} \left( \frac{1}{1 + t} \right)^\lambda + \hat{q} \left( \frac{1}{1 - \frac{\hat{p}}{\hat{q}}t} \right)^\lambda - 1 \right\} = \kappa^2. \]

Consider \( t = b_1n^{-1/2} + b_2n^{-1} + b_3n^{-3/2} + \epsilon_1 \), where \( \epsilon_1 = \sum_{i=4}^{\infty} b_in^{-i/2} \) for some \( b_i \). Keep in mind that \( b_i \) are functions of \( \hat{p} \) and \( \hat{q} \) for all \( i \). The Taylor
expansion of $[1/(1 + t)]^\lambda$ is
\[
\left(\frac{1}{1 + t}\right)^\lambda = 1 - \lambda t + \frac{1}{2} \lambda(\lambda + 1)t^2 - \frac{1}{6} \lambda(\lambda + 1)(\lambda + 2)t^3 \\
+ \frac{1}{24} \lambda(\lambda + 1)(\lambda + 2)(\lambda + 3)t^4 + O(t^5).
\]
Hence equation (18) can be reformulated as a polynomial of $t$. After some algebra we have
\[
\frac{1}{2} \kappa^2 n^{-1} = \frac{\hat{p}}{2}(b_1^2 n^{-1} + b_2^2 n^{-2} + 2b_1b_2n^{-3/2} + 2b_1b_3n^{-2}) \\
- \frac{\hat{p}}{6}(\lambda + 2)(b_1^3 n^{-3/2} + 3b_2^2b_2n^{-2}) + \frac{\hat{p}}{24}(\lambda + 2)(\lambda + 3)b_1^4n^{-2} \\
+ \frac{\hat{q}}{2} \left(\frac{\hat{q}}{\hat{p}}\right)^2 (b_1^2 n^{-1} + b_2^2 n^{-2} + 2b_1b_2n^{-3/2} + 2b_1b_3n^{-2}) \\
+ \frac{\hat{q}}{6} \left(\frac{\hat{q}}{\hat{p}}\right)^3 (\lambda + 2)(b_1^3 n^{-3/2} + 3b_2^2b_2n^{-2}) \\
+ \frac{\hat{q}}{24} \left(\frac{\hat{q}}{\hat{p}}\right)^4 (\lambda + 2)(\lambda + 3)b_1^4n^{-2} + \epsilon(n^{-5/2}).
\]
The coefficients of the terms $n^{-1}$, $n^{-3/2}$ and $n^{-2}$ should be 0, thus we have
\[
b_1 = \pm \kappa (\hat{q}/\hat{p})^{1/2}, \ b_2 = \frac{1}{6}(\lambda + 2)(1 - 2\hat{p})\kappa^2/\hat{p}
\]and
\[
b_3 = \pm \frac{1}{72}(\lambda + 2)\hat{p}^{-3/2}\hat{q}^{-1/2}\kappa^3(2\lambda - 11\lambda\hat{p} + 11\lambda\hat{p}^2 + 1 - 13\hat{p}\hat{q}).
\]
Since equation (18) has two roots, such solutions can be expressed as
\[
t_{1,2} = \pm \kappa \left(\frac{\hat{q}}{\hat{p}}\right)^{1/2} n^{-1/2} + \frac{1}{6}(\lambda + 2)\left(1 - 2\frac{\hat{p}}{\hat{p}}\right)\kappa^2 n^{-1} \\
\pm \frac{1}{72}(\lambda + 2)\hat{p}^{-3/2}\hat{q}^{-1/2}\kappa^3(2\lambda - 11\lambda\hat{p} + 11\lambda\hat{p}^2 + 1 - 13\hat{p}\hat{q})n^{-3/2} + \epsilon_\pm,
\]
where $\epsilon_\pm$, consisting of $\epsilon_+$ and $\epsilon_-$, have the means $E\epsilon_\pm = O(n^{-2})$. Hence,
\[
E(U_{PD} - L_{PD}) = 2\kappa E(\hat{p}\hat{q})^{1/2}n^{-1/2} + \frac{1}{36}(\lambda + 2)\kappa^3(2\lambda + 1)n^{-3/2}E(\hat{p}\hat{q})^{-1/2} \\
- \frac{1}{36}(\lambda + 2)\kappa^3(11\lambda + 13)n^{-3/2}E(\hat{p}\hat{q})^{1/2} + O(n^{-2}). \quad (19)
\]
Let \( z = \sqrt{n}(\hat{p} - p)/\sqrt{pq} \), then \((\hat{p}\hat{q})^{1/2}\) and \((\hat{p}\hat{q})^{-1/2}\) are equivalent to

\[
[p + (pq)^{1/2}n^{-1/2}z]^{1/2}[q - (pq)^{1/2}n^{-1/2}z]^{1/2}
\]

and

\[
[p + (pq)^{1/2}n^{-1/2}z]^{-1/2}[q - (pq)^{1/2}n^{-1/2}z]^{-1/2}
\]

respectively. The Taylor expansion of the form \((x + yz)^{1/2}\) enables us to have

\[
(\hat{p}\hat{q})^{1/2} = (pq)^{1/2} - \frac{1}{2}p n^{-1/2}z - \frac{1}{8}p^{3/2}q^{-1/2}n^{-1}z^2 + \frac{1}{2}q n^{-1/2}z
- \frac{1}{4}(pq)^{1/2}n^{-1}z^2 - \frac{1}{8}p^{-1/2}q^{3/2}n^{-1}z^2 + O(n^{-3/2}).
\]

Similarly, the Taylor expansion of \((x + yz)^{-1/2}\) indicates

\[
(\hat{p}\hat{q})^{-1/2} = (pq)^{-1/2} + O(n^{-1/2}).
\]

Therefore

\[
E(\hat{p}\hat{q})^{1/2} = (pq)^{1/2} - \frac{1}{8}(pq)^{-1/2}n^{-1} + O(n^{-3/2}),
\]

\[
E(\hat{p}\hat{q})^{-1/2} = (pq)^{-1/2} + O(n^{-1/2}).
\]

Subsequently, equation (19) can be simplified to

\[
E(U_{PD} - L_{PD}) = 2\kappa(pq)^{1/2}n^{-1/2} - \frac{1}{4}\kappa(pq)^{-1/2}n^{-3/2}
+ \frac{1}{36}(\lambda + 2)(pq)^{-1/2}\kappa^3[2\lambda + 1 - (11\lambda + 13)pq]n^{-3/2} + O(n^{-2}).
\]

When \( \lambda = 0 \), which is the likelihood ratio interval, Brown et al. (2002) has already derived the result as

\[
E(U_{PD} - L_{PD})
= 2\kappa(pq)^{1/2}n^{-1/2} - \frac{1}{4}\kappa(pq)^{-1/2}n^{-3/2} + \frac{1}{18}(pq)^{-1/2}\kappa^3[1 - 13pq]n^{-3/2} + O(n^{-2}).
\]
However, it fits the result of equation (11) by specifying $\lambda = 0$, hence we proved the first part of the theorem.

For the second part of the theorem, we only need to insert equation (11) into $\int_0^1 E(U_{PD} - L_{PD}) f(p; a, b) dp$. Some simple algebra show the result. □

References


