Computing Exact Confidence Coefficients of Simultaneous Confidence Intervals for Multinomial Proportions and their Functions

Shaobo Jin
Title: Computing Exact Confidence Coefficients of Simultaneous Confidence Intervals for Multinomial Proportions and their Functions

Author: Shaobo Jin
E-mail: Shaobo.Jin@statistics.uu.se
Computing Exact Confidence Coefficients of Simultaneous Confidence Intervals for Multinomial Proportions and Their Functions

Shaobo Jin*

May 28, 2013

Abstract

There are many methods to calculate simultaneous confidence intervals for multinomial proportions with \( d \) categories. They are asymptotic \( 1 - \alpha \) simultaneous confidence intervals. Wang (J. Multivar. Anal. (2008) 99 896-911) proposed a method to find the exact confidence coefficient for a fixed sample size \( n \) instead of using simulation results. However, her procedure is only applicable to \( d-1 \) intervals instead of \( d \) intervals. Most of the methods for such intervals focus on \( d \) intervals simultaneously. In this paper, we generalized her idea to find the exact confidence coefficient for \( d \) simultaneous confidence intervals for multinomial proportions. Our method has to consider lots of new points in order to correctly identify the point which attains the confidence coefficient. Generalizations to the SCIs for the contrasts and ratios are briefly discussed.

Keywords: Simultaneous confidence intervals, Multinomial proportions, Confidence coefficient

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

The multinomial distribution plays an important role in statistical practice. Contingency tables in the survey analysis or experimental designs can be regarded as multinomial distributions. In an election with more than two candidates, the counts of votes also construct a multinomial distribution. The problem of estimating the cell probabilities arises, for example, when the media wants to predict the result of an election. Two generic methods are the point estimation and the interval estimation.

For a multinomial distribution, \( X = (X_1, X_2, \cdots, X_d) \), the probability mass function is

\[
P(X_1 = x_1, X_2 = x_2, \cdots, X_d = x_d) = \frac{n!}{x_1!x_2! \cdots x_d!} p_1^{x_1} p_2^{x_2} \cdots p_d^{x_d},
\]

with \( n = \sum_{i=1}^{d} x_i \). The parameter space is defined as

\[
\Theta = \left\{ p = (p_1, \cdots, p_d); \sum_{i=1}^{d} p_i = 1, p_i > 0, i = 1, \cdots, d \right\}.
\]

For some \( \alpha \in (0, 1) \), we want to construct simultaneous confidence intervals (SCIs) for \( p_1, p_2, \cdots, p_d \), denoted as \( I_1, I_2, \cdots, I_d \), such that the joint coverage probability is larger than or equal to \( 1 - \alpha \). Such methods are fruitful in the literature. To our best knowledge, Quesenberry and Hurst (1964) proposed the first SCIs,

\[
p_i = \frac{a + 2X_i \pm \sqrt{a(a + 4X_i(n - X_i)/n)}}{2(n + a)}, \quad i = 1, \cdots, d,
\]

where \( a = \chi^2_{1-\alpha}(d - 1) \). Goodman (1965) suggested that this estimation can be improved by replacing \( a = \chi^2_{1-\alpha}(d - 1) \) by \( b = \chi^2_{1-\alpha/d}(1) \) using an argument based on the Bonferroni inequality. This technique is supposed to obtain shorter interval estimations. Hou et al. (2003) studied a family of simultaneous confidence intervals based on the power-divergence type statistics proposed by Cressie and Read (1984).
For each proportion $p_i$, the lower bound and the upper bound are the minimum and maximum solutions of the function

$$(1 - \hat{P}_i)^{\lambda+1}P_i^\lambda + \hat{P}_i^{\lambda+1}(1 - P_i)^\lambda = A_\lambda P_i^\lambda(1 - P_i)^\lambda; \quad i = 1, 2, \cdots, d,$$  

(4)

where $A_\lambda = \lambda(\lambda + 1)\chi^2_{1-\alpha}(d-1)/(2n) + 1$ and $\hat{P}_i = X_i/n$. However, there are usually no explicit forms of the lower and upper limits. The SCIs mentioned above are all obtained from the $\chi^2$ approximation.

There are also some SCIs which are not based on the $\chi^2$ approximations. Goodman (1965) also introduced simpler simultaneous confidence intervals given by

$$\frac{X_i}{n} - \sqrt{\frac{aX_i(1 - X_i)}{n}} < p_i < \frac{X_i}{n} + \sqrt{\frac{aX_i(1 - X_i)}{n}}, \quad i = 1, 2, \cdots, d.$$  

(5)

Sison and Glaz (1995) proposed two intervals. One was based on the approximation for multinomial proportions introduced by Levin (1981). The other method was based on an inequality for a multinomial distribution. The proposed simultaneous confidence intervals can be formulated as

$$\hat{p}_i - \frac{c}{n} < p_i < \hat{p}_i + \frac{c + 2\gamma}{n}, \quad i = 1, \cdots, d,$$  

(6)

for some $c, r$ and $\hat{p}_i = X_i/n, \quad i = 1, 2, \cdots, d$. Fitzpatrick and Scott (1987) studied the conventional simultaneous confidence interval $|\hat{p}_i - p_i| < 1/\sqrt{n}$. By aiding of the Bonferroni’s inequality, they proposed quick simultaneous intervals, given by

$$\hat{p}_i - \frac{\Phi^{-1}(1 - \frac{a}{2})}{2\sqrt{n}} < p_i < \hat{p}_i + \frac{\Phi^{-1}(1 - \frac{a}{2})}{2\sqrt{n}},$$  

(7)

where $\Phi^{-1}$ was the inverse of the distribution function of a standard normal distribution.

One common measure to assess the properties of SCIs is the confidence coefficient.
which is defined as the infimum of the coverage probability over the parameter space $\Theta$. This represents the most dangerous case which of course we would like to control. Fitzpatrick and Scott (1987) gave a lower bound of the confidence coefficient of their quick SCIs. Wang (2008) was the first one who considered the method to calculate the exact confidence coefficients of SCIs for multinomial proportions instead of using simulations. The parameter space was defined as

$$\Omega = \left\{(p_1, \cdots, p_{d-1}); \sum_{i=1}^{d-1} p_i \leq 1, 0 \leq p_i \leq 1, i = 1, \cdots, d-1 \right\},$$

in Wang (2008), which was slightly different from $\Theta$ by allowing the appearance of parameters with value 0. And only the coverage probability of SCIs for $p_1, p_2, \cdots, p_{d-1}$ were used (the form of coverage probability will be given in the next section). However, all the SCIs mentioned above are constructed for $d$ parameters simultaneously. So the method developed for the $d - 1$ SCIs will only give a upper bound to the coverage coefficients sometimes if it is used to assess the methods in the literature. In Wang (2008), she argued that in multinomial distribution there were only $d - 1$ parameters, so it was natural to consider the $d - 1$ SCIs. She used the binomial distribution as an example. But the reason to consider $d$ SCIs instead of $d - 1$ SCIs is also obvious. In binomial case, we have two parameters and one restriction about the parameters. If we construct a confidence interval for one parameter $p$, say $(L, U)$, we can easily obtain the confidence interval for another parameter $q = 1 - p$ as $(1 - U, 1 - L)$ which is equivariant in the sense of Blyth and Still (1983). However this is not the case in multinomial cases. In the multinomial problems, if we have the SCIs for $p_1, \cdots, p_{d-1}$, there is no easy way to obtain the confidence interval for the last parameter $p_d = 1 - \sum_{i=1}^{d-1} p_i$. We can not use one minus the sum of the upper (lower) bound as the lower (upper) bound for $p_d$ because most likely both of them will be negative. So it is necessary to develop a method to find the exact confidence coefficients for the $d$ SCIs.
In this paper, we extend the work of Wang (2008) to obtain the exact confidence coefficients of $d$ SCIs which is mostly focused on in the literature. The paper is organized as follows. In Section 2, we introduce some main results from Wang (2008) and present some results under our setting. The idea is generalized to other types of SCIs in Section 3. In Section 4, some numerical results are illustrated. A discussion ends the paper.

2 General Methods

For the $i$-th cell probability $p_i$, the lower bound and upper bound are denoted as $L_i(X)$ and $U_i(X)$. They can be reformulated as functions of $\hat{P}_i$, for $i = 1, \cdots, d$ which are the MLE of the population proportions. Assumption 1 contains some keys assumptions in the paper.

**Assumption 1.**

1. $L_i(X)$ have the same form for all $i$ and $U_i(X)$ have the same form for all $i$ as well.

2. For any fixed $n$, $L_i(X)$ and $U_i(X)$ only depend on $x_i$, not $x_j$ where $j \neq i$.

Hence we can write $L_i(X) = L(X_i)$ and $U_i(X) = U(X_i)$.

3. $L(X_i)$ and $U(X_i)$ are increasing functions of $\hat{P}_i$.

4. When $X_i = 0$, then $L(X_i) \leq 0 \leq U(X_i)$.

5. For any fixed $p$ in the parameter space, there exists an $x_0$ such that $p \in I(x_0)$ and $P_p(X = x_0) > 0$.

The Assumption 1(2)-(5) are just the Assumption 1 in Wang (2008).
2.1 $d$-1 SCIs

For a fixed $p$ in the parameter space, the coverage probability for $p_1, p_2, \ldots, p_{d-1}$ is defined as

$$CP_{d-1}(p_1, \ldots, p_{d-1}) = \sum_{x_1=0}^{n} \cdots \sum_{x_{d-1}=0}^{n} \left[ \frac{n!}{x_1!x_2!\cdots x_{d-1}!} p_1^{x_1}p_2^{x_2}\cdots p_{d-1}^{x_{d-1}} \prod_{i=1}^{d-1} I(p_i \in I_i) \right].$$

Collect all the values of $L(X)$ and $U(X)$ by increasing order into a set $v = \{v_i\}$ which has at most $2n + 2$ elements. Define the set $v' = \{v_i \notin \{0, 1\}\}$, hence $v' \subseteq v = \{v_i\}$. Let $g = \#v$ and $g' = \#v'$. The parameter space can be divided into several subsets following Wang (2008)'s method. For each $p_i$, $i = 1, \ldots, d$, there exists $v_j$ and $v_{j+1}$ such that $v_j \leq p_i \leq v_{j+1}$, so each subset can be formulated as

$$((v_{i_1}, v_{i_1}'), \ldots, (v_{i_{d-1}}, v_{i_{d-1}}')) \cap \Omega,$$

where $v_{i_j} \in v$ and $i_j \in \{i_j, i_j+1\}$. If $i_j' = i_j$, then the corresponding interval $(v_{i_j}, v_{i_j}')$ is defined as the hyperplane $p_j = v_{i_j}$. Figure 1(a) gives an example of the subsets when $n = 5$ and $d = 3$. In fact, the parameter space is separated by the hyperplanes $p_i = v_j'$, $i = 1, \ldots, d-1$, $j = 1, \ldots, g'$. Let $s$ be the number of total subsets and $\omega_t$ collects all $(x_1, \ldots, x_d)$ such that for all $(p_1, \ldots, p_d)$ in the subset $\tau_t$ with the form (10), we have $p_i \in (L(x_i), U(x_i))$ for all $i = 1, \ldots, d-1$.

**Lemma 1.** If the SCIs satisfy the Assumption 1 (1)-(3), within each subset (10) without any other elements of $v = \{v_i\}$ in the interior of the subset, the coverage probability (9) can be expressed as

$$CP_{d-1}(p_1, \ldots, p_d) = \sum_{x_1=a_1}^{b_1} \sum_{x_2=a_2}^{b_2(x_1)} \cdots \sum_{x_{d-1}=a_{d-1}}^{b_{d-1}(x_{1}, \ldots, x_{d-2})} \frac{n!}{x_1!x_2!\cdots x_{d-1}!} p_1^{x_1}p_2^{x_2}\cdots p_d^{x_d}. \quad (11)$$

for some constants $a_1, \ldots, a_{d-1}, b_1$ and some functions $b_2(x_1), \ldots, b_{d-1}(x_1, \ldots, x_{d-2})$. (Lemma 3, Wang (2008)).
Figure 1: An example of partitioning the parameter space using the lower limits and upper limits of the SCIs when \( n = 5 \) and \( d = 3 \).

Rewrite the formula (11) as \( CP_{d-1}(p_{d-1}|p_1, \ldots, p_{d-2}) \) representing the conditional function given \( p_1, \ldots, p_{d-2} \). Then we have the following lemma.

**Lemma 2.** If \( b_i(x_1, \ldots, x_{i-1}) > a_i \), for some \( i = 1, \ldots, d-2 \), then

1. \( CP_{d-1}(p_{d-1}|p_1, \ldots, p_{d-2}) \) is a unimodal function or an increasing function of \( p_{d-1} \) if \( a_{d-1} > 0 \) and \( x_1 + \cdots + b_{d-1}(x_1, \ldots, x_{d-2}) < n \).

2. \( CP_{d-1}(p_{d-1}|p_1, \ldots, p_{d-2}) \) is a decreasing function of \( p_{d-1} \) if \( a_{d-1} = 0 \).

3. \( CP_{d-1}(p_{d-1}|p_1, \ldots, p_{d-2}) \) is an increasing function of \( p_{d-1} \) if

\[
x_1 + \cdots + b_{d-1}(x_1, \ldots, x_{d-2}) = n.
\]

*(Lemma 2, Wang (2008)).*
2.2 d SCIs

The coverage probability of $d$ SCIs is

$$CP_d(p_1, \cdots, p_d) = \sum_{x_1=0}^{n} \cdots \sum_{x_d=0}^{n} \left[ \frac{n!}{x_1! \cdots x_d!} p_1^{x_1} p_2^{x_2} \cdots p_d^{x_d} \prod_{i=1}^{d} I_{\{p_i \in \Theta_i\}} \right].$$

(12)

Here we need a new method to separate the parameter space. Consider the following subset

$$\left( (v_{i_1}, v'_{i_1}) \times \cdots \times (v_{i_d}, v'_{i_d}) \right) \cap \Theta,$$

(13)

where $v_{i_j} \in v$ and $i'_j \in \{i_j, i_j + 1\}$. If $i'_j = i_j$, then the corresponding open interval $(v_{i_j}, v'_{i_j})$ is defined as the hyperplane $p_j = v_{i_j}$. Figure 1(b) gives an example of the subsets when $n = 5$ and $d = 3$. In fact, the parameter space is separated by the hyperplanes $p_i = v'_{i_j}$, $i = 1, \cdots, d$, $j = 1, \cdots, g'$. So each subset (10) defined before is partitioned into several smaller sets. Let $s'$ be the total number of subsets and $\omega'_t$ collects all $(x_1, \cdots, x_d)$ such that for all $(p_1, \cdots, p_d)$ in the subset $\tau'_t$ with the form of (13), we have $p_i \in (L(x_i), U(x_i))$ for all $i = 1, \cdots, d$. If we consider the coverage probability (12) and the subset (13), we do not have a nice result as equation (11). The best we can do is to deal with

$$CP_d(p_1, \cdots, p_d) = \sum_{x \in \omega'_t} \left[ \frac{n!}{x_1! \cdots x_d!} p_1^{x_1} p_2^{x_2} \cdots p_d^{x_d} \right]$$

$$= \sum_{x_1=a_1}^{b_1} \sum_{x_2=a_2(x_1)}^{b_2(x_1)} \cdots \sum_{x_d-1=a_{d-1}(x_1, \cdots, x_{d-2})}^{b_{d-1}(x_1, \cdots, x_{d-2})} \left[ \frac{n!}{x_1! \cdots x_d!} p_1^{x_1} p_2^{x_2} \cdots p_d^{x_d} \right].$$

(14)

Assume we have a subset

$$\tau'_t = ((v'_{i_1}, v'_{i_1+1}) \times \cdots \times (v'_{i_d}, v'_{i_d+1})) \cap \Theta,$$
with no hyperplane involved, the corresponding set for $X$ is $\omega'_i$. First, consider the case that we decrease $p_1$ from $p_1 = v'_{i_1} + \varepsilon$ to $p_1 = v'_{i_1}$ while $p_j$, $j = 2, \cdots, d$, are still in $\tau'_i$. When $v'_{i_1}$ is $L(x)$ for some $x$, $v'_{i_1} + \varepsilon \in (L(x), U(x))$ but $v'_{i_1} \notin (L(x), U(x))$. We lose the observation $X_1 = x$ from $\omega'_i$. When $v'_{i_1}$ is $U(x)$ for some $x$, $v'_{i_1} + \varepsilon \notin (L(x), U(x))$ and $v'_{i_1} \notin (L(x), U(x))$ still cannot cover $p_1 = v'_{i_1}$. $\omega'_i$ stays the same. Second, consider the case that we increase $p_1$ from $p_1 = v'_{i_1+1} - \varepsilon$ to $p_1 = v'_{i_1+1}$ while $p_j$, $j = 2, \cdots, d$, still in $\tau'_i$. When $v'_{i_1+1}$ is $L(x)$ for some $x$, $(L(x), U(x))$ still cannot cover $p_1 = v'_{i_1}$. $\omega'_i$ stays the same. When $v'_{i_1+1}$ is $U(x)$ for some $x$, $(L(x), U(x))$ cannot cover $p_1 = v'_{i_1}$. We lose the observation $X_1 = x$ from $\omega'_i$. Thus within each subset, the values that $X$ can take are fixed but it may or may not vary from one subset to the other.

Since Equation (14) has $d - 1$ free arguments, we can study $CP_d(p_1, \cdots, p_d)$ conditioning on $d - 2$ free arguments. Define two functions first

$$B_t(p_i) = \sum_{x_1=a_1}^{b_1} \sum_{x_2=a_2(x_1)}^{b_2(x_1)} \cdots \sum_{x_{d-1}=a_{d-1}(x_1, \cdots, x_{d-2})}^{b_{d-1}(x_1, \cdots, x_{d-2})} \frac{n!}{x_1!x_2!\cdots x_d!} p_1^{x_1} \cdots p_{d-1}^{x_{d-1}} x_d^{x_1-\cdots-x_{d-1}},$$

$$C_t(p_i) = \sum_{x_1=a_1}^{b_1} \sum_{x_2=a_2(x_1)}^{b_2(x_1)} \cdots \sum_{x_{d-1}=a_{d-1}(x_1, \cdots, x_{d-2})}^{b_{d-1}(x_1, \cdots, x_{d-2})} \frac{n!}{x_1!x_2!\cdots x_d!} p_1^{x_1} \cdots p_{d-1}^{x_{d-1}} x_d^{x_1-\cdots-x_{d-1}-1},$$

where the subscript $t$ means $x \in \omega'_i$ and the argument $p_i$ means that $p_k$, $k \neq i$ and $d$ are fixed. First we study $CP_d(p_1, \cdots, p_d)$ condition on $p_1, \cdots, p_{d-2}$, denoted by $CP_d(p_{d-1}|p_1, \cdots, p_{d-2})$. Equation (14), $B_t(p_{d-1})$ and $C_t(p_{d-1})$ are continuous and differentiable functions in $p_{d-1}$. Although they are restricted in one or some specific subsets $\tau'_i$, we temporarily extend the domains to $\left(0, 1 - \sum_{j=1}^{d-2} p_j\right)$.

**Lemma 3.** If the SCIs satisfying the Assumption 1 (1)-(3), within each subset $\{v'_i\}$ in the interior of the subset, then

1. If $a_{d-1}(x_1, \cdots, x_{d-2}) = b_{d-1}(x_1, \cdots, x_{d-2}) = 0$ for all $x_1, \cdots, x_{d-2}$ and $x_1 +$
\[ \cdots + b_{d-1}(x_1, \cdots, x_{d-2}) < n \text{ for some } x_1, \cdots, x_{d-2}, \text{ then } CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \]

is a decreasing function in \( p_{d-1} \).

2. If \( a_{d-1}(x_1, \cdots, x_{d-2}) = b_{d-1}(x_1, \cdots, x_{d-2}) \neq 0 \) and, for all \( x_1, \cdots, x_{d-2}, x_1 + \cdots + b_{d-1}(x_1, \cdots, x_{d-2}) = n \), \( CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \) is an increasing function in \( p_{d-1} \).

3. If \( a_{d-1}(x_1, \cdots, x_{d-2}) = b_{d-1}(x_1, \cdots, x_{d-2}) = 0 \) and, for all \( x_1, \cdots, x_{d-2}, x_1 + \cdots + b_{d-1}(x_1, \cdots, x_{d-2}) = n \), \( CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \) is a constant function in \( p_{d-1} \).

Further if \( B_t(p_{d-1}) - C_t(p_{d-1}) = 0 \) has no more than two solutions, then

4. \( CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \) is an unimodal function for other values of \( a_{d-1} \) and \( b_{d-1} \).

This lemma is the extension of Lemma\(^2\) from \( d-1 \) SCIs to \( d \) SCIs. We impose one assumption about \( B_t(p_{d-1}) \) and \( C_t(p_{d-1}) \) to have unimodality. Such an assumption does not appear in [Wang (2008)](Wang_2008). When she showed the unimodality property, the argument that two convex functions have at most two intersections are used. If that is the case, we do not need the assumption regarding \( B_t(p_{d-1}) \) and \( C_t(p_{d-1}) \) anymore. However, in general this argument is not true and we have to check this assumption. It can be checked numerically. If \( B_t(p_{d-1}) - C_t(p_{d-1}) = 0 \) has more than two solutions, \( CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \) may have several local minima. We need to pay more attention to these local minima. However, \( CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \) is actually defined on some subset \( \omega'_t \) instead of \( \left(0, 1 - \sum_{j=1}^{d-2} p_j \right) \). If in the subset \( \omega'_t \), \( CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \) is unimodal or monotone, we don’t need to worry about these local minima.

Another remark regarding Lemma\(^3\) is that we can switch the fixed cell probabilities. For example, if we consider \( CP_d(p_1|p_2, \cdots, p_{d-1}) \) where \( p_1 \) is the free parameter and \( p_2, \cdots, p_{d-1} \) are fixed, similar results as Lemma\(^3\) can be obtained easily.
The main idea is that if we fixed \( d - 2 \) cell probabilities, the (conditional) coverage probability is a decreasing/increasing/constant/unimodal function under different conditions.

Using Lemma 3, we can study the coverage probability within each subset (13). If the coverage probability is a decreasing/increasing/constant/unimodal function within a subset, the minimum is taken on the boundaries. This establishes Theorem 1 following.

**Theorem 1.** Assume Assumption 1 is fulfilled, and let

\[
E_0 = \{(\pi_1, \pi_2, \cdots, \pi_d); \pi_i = 0 \text{ for some } i = 1, \cdots, d, \text{ and } \pi_j \in \mathcal{V}' \text{ for } j \neq i\}
\]

\[
E_1 = \{(\pi_1, \pi_2, \cdots, \pi_d); \pi_i = v_j, i = 2, \cdots, d, j = 1, \cdots, g, \sum_{m=2}^{d} \pi_m < 1\} \cap \Theta,
\]

\[
E_k = \{(\pi_1, \pi_2, \cdots, \pi_d); \pi_i = v_j, i \neq k, j = 1, \cdots, g, \sum_{m \neq k}^{d} \pi_m < 1\} \cap \Theta,
\]

\( k = 2, \cdots, d - 1, \)

\[
E_d = \{(\pi_1, \pi_2, \cdots, \pi_d); \pi_i = v_j, i = 1, \cdots, d - 1, j = 1, \cdots, g, \sum_{m=1}^{d-1} \pi_m < 1\} \cap \Theta,
\]

\[
M_{1,t} = \{(\pi_1, \pi_2, \cdots, \pi_d); B_t(\pi_i) = C_t(\pi_i), i = 2, \cdots, d, \sum_{m=2}^{d} \pi_m < 1\} \cap \tau'_t,
\]

\[
M_{k,t} = \{(\pi_1, \pi_2, \cdots, \pi_d); B_t(\pi_i) = C_t(\pi_i), i \neq k, \sum_{m \neq k}^{d} \pi_m < 1\} \cap \tau'_t,
\]

\( k = 2, \cdots, d - 1, \)

\[
M_{d,t} = \{(\pi_1, \pi_2, \cdots, \pi_d); B_t(\pi_i) = C_t(\pi_i), i = 1, \cdots, d - 1, \sum_{m=1}^{d-1} \pi_m < 1\} \cap \tau'_t,
\]

\( t = 1, \cdots, s'. \)

Then
1. If $L(x) = 0$ if and only if $x = 0$, the confidence coefficient is obtained at some of the points in

$$E_1 \cup \cdots \cup E_d \cup \left( \bigcup_{t=1}^{s'} (M_{1,t} \cup \cdots \cup M_{d,t}) \right).$$

2. If $L(x) \leq 0$ for some nonzero $x$, the confidence coefficient is obtained at some of the points in

$$E_0 \cup E_1 \cup \cdots \cup E_d \cup \left( \bigcup_{t=1}^{s'} (M_{1,t} \cup \cdots \cup M_{d,t}) \right).$$

In Theorem 1, $E_0$ is the collection of boundary points, $E_1 \cup \cdots \cup E_d$ is the collection of the cross points and $\bigcup_{t=1}^{s'} (M_{1,t} \cup \cdots \cup M_{d,t})$ is the collection of local minima. If $L(0) < 0$ and $L(x) = 0$ for a nonzero $x$,

$$CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) > CP_d(0|p_1, \cdots, p_{d-2}),$$

if $p_{d-1}$ is sufficiently close to 0. The value itself is smaller than the right limit. Thus $CP_0(p_{d-1}|p_1, \cdots, p_{d-2})$ is a minimum value locally.

We fail to prove the nonexistence of the local minima which makes the theorem no so easy to use. The coverage probability (14) is a sum of unimodal (or monotone) functions. But the sum of unimodal functions are not guaranteed to be unimodal, so we have to consider the possibility of local minima as well. Lemma 3 says that when $B_t(p_{d-1}) - C_t(p_{d-1}) = 0$ has no more than two solutions, $CP_d(p_{d-1}|p_1, \cdots, p_{d-2})$ is a unimodal function under some conditions. In some subsets, $B_t(p_{d-1}) - C_t(p_{d-1}) = 0$ has more than two solutions. This happens when $x_i$ and $x_j$ are exchangeable in the sense that if

$$(x_1, \cdots, x_i, \cdots, x_j, \cdots, x_d) \in \omega_i$$
in equation (14) then

\[(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_d) \in \omega'_t\]
as well. For instance, if \(n = 5, d = 3\) and \(\omega'_t = \{(4,0,1), (4,1,0), (5,0,0)\}\), then \(B = C\) for all fixed \(p_1 \in (U(3), U(4))\) and \(p_2 \in (L(1), 1 - 2L(1))\). And within this subset

\[CP_3(p_1, p_2, p_3) = \sum_{x \in \omega'_t} \left[ \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \right] = 5p_1^4 - 4p_1^5,\]

which is a constant when \(p_1\) is fixed. Actually, the constant coverage can happen in many subsets. However, if we exclude such constant coverage case, we argue that it is unlikely that we face the problem of local minima. We are dealing with the sum of

\[\frac{n!}{x_1! x_2! \cdots x_d!} p_1^{x_1} p_2^{x_2} \cdots p_d^{x_d}\]

over some \(\omega'_t\), the possible values of \(x_{d-1}\) in \(\omega'_t\) are 'continous' in the sense that all the values between \(a_{d-1}(x_1, \ldots, x_{d-2})\) and \(b_{d-1}(x_1, \ldots, x_{d-2})\) belong to \(\omega'_t\) if \(x_1, \ldots, x_{d-2}\) are fixed. Thus

\[\sum_{x_{d-1}=a_{d-1}(x_1, \ldots, x_{d-2})}^{b_{d-1}(x_1, \ldots, x_{d-2})} \left[ \frac{n!}{x_1! x_2! \cdots x_d!} p_1^{x_1} p_2^{x_2} \cdots p_d^{x_d} \right]\]
is unimodal if \(a_{d-1}(x_1, \ldots, x_{d-2}) > 0\) and \(x_1 + \cdots + b_{d-1}(x_1, \ldots, x_{d-2}) < n\) when we fix \(p_1, \ldots, p_{d-2}\). The mode takes place at

\[p_{d-1} = \frac{\left( \frac{b_{d-1}! (n-x_1-\cdots-x_{d-2}-a_{d-1}-1)!}{a_{d-1}! (n-x_1-\cdots-x_{d-2}-a_{d-1}-1)!} \right)^{1/6}}{1 + 1} \quad \text{if} \quad (1 - p_1 - \cdots - p_{d-2}).\]
Hence unless we have 'discrete’ \( x_i \) in \( \omega'_t \), \( CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \) is unimodal or monotone or a constant over \( \omega'_t \).

### 3 Simultaneous Confidence Intervals for Functions of Proportions

Theorem \[1\] is only valid to identify the exact confidence coefficient of SCIs for multinomial proportions. It cannot be directly applied to other types of SCIs. In practice, we are also interested in the SCIs for the contrasts and the ratios. From the proof of Theorem \[1\], we can see that the theorem can actually be modified to some other cases as well. As long as the coverage probability can be expressed as \[14\], the idea of Theorem \[1\] can identify the point attaining the coverage coefficient. We only need to examine the intersections, local minima and the boundary points. The only difference is the partition of the parameter space.

#### 3.1 Contrast

Consider the SCIs for the \( d(d - 1)/2 \) differences \( p_i - p_j \) with \( i < j \). The confidence interval for \( p_j - p_i \) can be obtained by inverting the one for \( p_i - p_j \). \[15\]

Gold (1963) proposed the SCIs of the form

\[ p_i - p_j = \frac{X_i - X_j}{n} \pm \sqrt{\frac{a}{n} \left( \frac{X_i + X_j}{n} - \left( \frac{X_i - X_j}{n} \right)^2 \right)}, \tag{15} \]

where \( a \) was defined in the introduction. Goodman (1965) generalized the idea to the SCIs for the all possible linear combinations of \( p_i \). If we are interested in the
SCIs for $\sum \delta_i p_i$ for some constants $\delta_1, \cdots, \delta_d$, then the SCIs are

$$\sum \delta_i p_i = \sum \delta_i \frac{X_i}{n} \pm \frac{a}{n} \left( \sum \delta_i^2 \frac{X_i}{n} - \left( \sum \delta_i \frac{X_i}{n} \right)^2 \right).$$

(16)

We can also derive the SCIs from the confidence regions which are called Pearson SCIs in this paper. For example if we have a confidence region $CR_\alpha$, and we want a confidence interval for $p_i - p_j$, we can consider the line $p_i - p_j = z$. The lower bound and upper bound are the values of $z$ such that $p_i - p_j = z$ and $CR_\alpha$ have only one intersection respectively. Figure 2(a) shows the partition of the parameter space based on the SCIs for the contrasts from the Pearson $\chi^2$ confidence regions.

Figure 2: An example of the partition of the parameter space based on the SCIs derived from the Pearson $\chi^2$ confidence regions. $n = 5$, $d = 3$ and $\alpha = 0.05$. 
3.2 Ratio

We are interested in the SCIs for the ratio \( p_i/p_j \) or equivalently the natural logarithm of \( p_i/p_j \). Goodman (1965) provided a method to construct SCIs for \( p_i/p_j \) as

\[
X_i/X_j/\exp\left\{\sqrt{a\left(1/X_i + 1/X_j\right)}\right\} \leq \frac{p_i}{p_j} \leq X_i/X_j\exp\left\{\sqrt{a\left(1/X_i + 1/X_j\right)}\right\}.
\]

However, these SCIs face the problem of 0 observations. SCIs for the ratio can also be derived from confidence regions. Assume we have a confidence region \( CR_\alpha \), and we want simultaneous confidence intervals for \( p_i/p_j \). We can consider the line \( p_i/p_j = z \), the lower bound and upper bound are the values of \( z \) such that \( p_i/p_j = z \) and \( CR_\alpha \) have only one intersection respectively or when \( z = 0 \) or \( \infty \). Especially when \( z = 0 \) or \( \infty \), the line \( p_i/p_j = z \) coincides with the coordinate axis. Figure 2(b) shows the partition of the parameter space based on the SCIs for the ratios from the Pearson \( \chi^2 \) confidence regions.

4 Numerical Examples

In this part, we first study the confidence coefficient of the Q-H SCIs (3) assuming \( n = 5 \) and \( d = 3 \). We set \( \alpha = 0.05 \) throughout this section. It is easy to verify that the Q-H SCIs (3) satisfies Assumption 1 and the condition in Theorem 1(1). According to Theorem 1(1), the confidence coefficient will be taken at the dot points shown in Figure 3 and all possible local minima within each subset. But no local minimum is found. The confidence coefficient is 0.7660488, which is attained at

\[
(p_1, p_2, p_3) = (L(1), L(1), 1 - L(1) - L(1)) = (0.0260, 0.0260, 0.9480)
\]

\[
(p_1, p_2, p_3) = (L(1), 1 - L(1) - L(1), L(1)) = (0.0260, 0.9480, 0.0260)
\]

\[
(p_1, p_2, p_3) = (1 - L(1) - L(1), L(1), L(1)) = (0.9480, 0.0260, 0.0260).
\]
The confidence coefficient is invariant under permutations of \( p_1, p_2 \) and \( p_3 \) if we consider \( d \) SCIs. Thus from now on, we only present one element within each invariant class.

![Figure 3: Finding exact confidence coefficient of Q-H SCIs when \( n=5 \) and \( d=3 \).](image)

We tabulate the exact confidence coefficients of Q-H SCIs (3) assuming \( d=3 \) for different sample sizes \( n \) in Table 1. Only limited sample sizes are used here just for illustration purposes. We find that for Q-H SCIs (3), the exact confidence coefficient considering \( d-1 \) SCIs and \( d \) SCIs are exactly the same. The points that we find for \( d-1 \)-SCIs’ case are also included in \( d \)-SCIs’ case. All the confidence coefficients are achieved at the points when two proportions equal \( L(1) \). No local minimum is found.

The Q-H SCIs are special cases of the power-divergence SCIs (4). Such SCIs are in fact functions of \( \lambda \). Hou et al. (2003) proposed that we should choose the \( \lambda \) which minimize the volume of the SCIs while the coverage probability was at least \( 1 - \alpha \).
Table 1: The exact confidence coefficients for Q-H SCIs \( \alpha = 0.05 \) when \( d=3 \) and different \( n \). For both \( d-1 \) SCIs and \( d \) SCIs, there are several points achieving the confidence coefficient as coverage probabilities. We only list one example in the table. The other points can be obtained by permutating the coordinates.

<table>
<thead>
<tr>
<th>( n )</th>
<th>Confidence coefficient ( (p_1, p_2, p_3) )</th>
<th>( d-1 ) SCIs</th>
<th>( d ) SCIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.7660 ( (L(1),L(1),1-L(1)-L(1))=(L(1),L(1),1-L(1)-L(1))= ) ( (0.0260,0.0260,0.9480) )</td>
<td>( (0.0260,0.0260,0.9480) )</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.7709 ( (L(1),L(1),1-L(1)-L(1))=(L(1),L(1),1-L(1)-L(1))= ) ( (0.0128,0.0128,0.9744) )</td>
<td>( (0.0128,0.0128,0.9744) )</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.7724 ( (L(1),L(1),1-L(1)-L(1))=(L(1),L(1),1-L(1)-L(1))= ) ( (0.0085,0.0085,0.9830) )</td>
<td>( (0.0085,0.0085,0.9830) )</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.7732 ( (L(1),L(1),1-L(1)-L(1))=(L(1),L(1),1-L(1)-L(1))= ) ( (0.0064,0.0064,0.9872) )</td>
<td>( (0.0064,0.0064,0.9872) )</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.7736 ( (L(1),L(1),1-L(1)-L(1))=(L(1),L(1),1-L(1)-L(1))= ) ( (0.0051,0.0051,0.9898) )</td>
<td>( (0.0051,0.0051,0.9898) )</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.7740 ( (L(1),L(1),1-L(1)-L(1))=(L(1),L(1),1-L(1)-L(1))= ) ( (0.0043,0.0043,0.9914) )</td>
<td>( (0.0043,0.0043,0.9914) )</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.7746 ( (L(1),L(1),1-L(1)-L(1))=(L(1),L(1),1-L(1)-L(1))= ) ( (0.0025,0.0025,0.9950) )</td>
<td>( (0.0025,0.0025,0.9950) )</td>
<td></td>
</tr>
</tbody>
</table>

But no systematic study has been done in the literature about the choice of \( \lambda \) in the case of confidence interval estimation. [Read and Cressie (1988)] recommended the use of \( \lambda = \frac{2}{3} \) as a compromise of the Pearson \( \chi^2 \) and the likelihood ratio if we used the power-divergence statistic as a test statistic. [Medak and Cressie (1991)] studied the confidence region based on the power-divergence statistic, they concluded that \( \lambda = \frac{2}{3} \) gave an improvement in small samples. So we studied the SCIs by fixing \( \lambda = \frac{2}{3} \). The requirements for Theorem 1(1) are fulfilled. From Table 2 we can see that all the confidence coefficients are larger than those in Table 1. All the points are still the cross points.

Another example is the F-S SCIs \( \alpha \). It is an example where \( L(x) < 0 \) for some \( x \). Theorem 1(2) is applied. From Table 3 we can see that the confidence coefficients of the \( d \) SCIs are all smaller than those of the \( d-1 \) SCIs when \( d=3 \). As the sample size increases, the points which achieve the confidence coefficients as coverage probabilities become different. And for the \( d-1 \) SCIs, under some sample sizes, the
Table 2: The exact confidence coefficients for power-divergence SCIs with $\lambda = 2/3$ when $d=3$ and different $n$. $\alpha = 0.05$. For both $d-1$ SCIs and $d$ SCIs, there are several points achieving the confidence coefficient as coverage probabilities. We only list one example in the table. The other points can be obtained by permutating the coordinates.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Confidence coefficient</th>
<th>$(p_1, p_2, p_3)$</th>
<th>$d - 1$ SCIs</th>
<th>$d$ SCIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.8256</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0188,0.0188,0.9624)$</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0188,0.0188,0.9624)$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.8296</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0093,0.0093,0.9814)$</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0093,0.0093,0.9814)$</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.8308</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0061,0.0061,0.9878)$</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0061,0.0061,0.9878)$</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.8314</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0046,0.0046,0.9908)$</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0046,0.0046,0.9908)$</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.8318</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0037,0.0037,0.9926)$</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0037,0.0037,0.9926)$</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.8320</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0031,0.0031,0.9938)$</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0031,0.0031,0.9938)$</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.8325</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0018,0.0018,0.9964)$</td>
<td>$(L(1),L(1),1-L(1)-L(1)) = (0.0018,0.0018,0.9964)$</td>
<td></td>
</tr>
</tbody>
</table>
points achieving the confidence coefficient are not necessary to have the equality $p_1 = p_2$ which is always the case in Table 1 and Table 2. Again, all the confidence coefficients are attained at cross points and no local minimum exists. Further, the non-consistency of the confidence coefficient can be observed here. Even we increase the sample size, the confidence coefficient decreases sometimes.

Table 3: The exact confidence coefficients for F-S SCIs [7] when $d=3$ and different $n$. $\alpha = 0.05$. For both $d-1$ SCIs and $d$ SCIs, there are several points achieving the confidence coefficient as coverage probabilities. We only list one example in the table. The other points can be obtained by permutating the coordinates.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$d-1$ SCIs</th>
<th>$d$ SCIs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(p_1,p_2,p_3)$</td>
<td>$(p_1,p_2,p_3)$</td>
</tr>
<tr>
<td>5</td>
<td>0.8783</td>
<td>0.8556</td>
</tr>
<tr>
<td></td>
<td>(L(4),L(4),1-L(4)-L(4))= (L(4),1-L(4)-L(4))=</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.3617, 0.3617, 0.2765) (0.3617, 0.2765)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.8971</td>
<td>0.8831</td>
</tr>
<tr>
<td></td>
<td>(L(7),L(7),1-L(7)-L(7))= (L(6),L(6),1-L(6)-L(7))=</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.3901, 0.3901, 0.21980) (0.2901, 0.3901, 0.3198)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.8914</td>
<td>0.8731</td>
</tr>
<tr>
<td></td>
<td>(L(10),L(10),1-L(10)-L(10))= (L(9),L(9),1-L(9)-L(9))=</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4136, 0.4136, 0.1727) (0.3470, 0.31970, 0.3333)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.8898</td>
<td>0.8817</td>
</tr>
<tr>
<td></td>
<td>(L(12),L(13),1-L(12)-L(13))= (U(3),U(3),1-U(3)-U(3))=</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.3809, 0.4390, 0.1883) (0.3691, 0.3691, 0.2617)</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.8917</td>
<td>0.8693</td>
</tr>
<tr>
<td></td>
<td>(L(15),L(15),1-L(15)-L(15))= (U(4),L(4),1-L(4)-U(4))=</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4040, 0.4040, 0.1920) (0.3560, 0.3240, 0.3200)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.8970</td>
<td>0.8878</td>
</tr>
<tr>
<td></td>
<td>(L(17),L(18),1-L(17)-L(18))= (L(16),L(16),1-L(16)-L(16))=</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.3877, 0.4211, 0.1912) (0.3544, 0.3544, 0.2912)</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.8995</td>
<td>0.8878</td>
</tr>
<tr>
<td></td>
<td>(L(27),L(28),1-L(27)-L(28))= (L(23),L(23),1-L(23)-L(23))=</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4014, 0.4214, 0.1772) (0.3214, 0.3386, 0.3400)</td>
<td></td>
</tr>
</tbody>
</table>

At last we study the SCIs for the contrasts. The Gold SCIs [15] have problems when we observe $x_i = x_j = 0$ or $x_i = n$ and $x_j = 0$. In the former case, both lower bound and upper bound is 0. In the latter case, both the lower bound and upper bound is 1. Thus it is natural to expect a low confidence coefficient. From Table 4 we see that the confidence coefficients are zero even though asymptotic 95% SCIs are desired. As a comparison, we studied the confidence coefficients of the Pearson SCIs. They are superior over the Gold SCIs [15] and therefore are preferred to the Gold SCIs [15].
Table 4: The exact confidence coefficients for Gold SCIs \(^{(15)}\) for contrasts and SCIs for contrasts derived from Pearson \(\chi^2\) confidence regions when \(d=3\) and different \(n\). \(\alpha = 0.05\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>Gold SCIs (^{(15)})</th>
<th>Pearson SCIs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Confidence coefficient</td>
<td>((p_1, p_2, p_3))</td>
</tr>
<tr>
<td>5</td>
<td>0.8385</td>
<td>(0.0173, 0.0173, 0.9654)</td>
</tr>
<tr>
<td>10</td>
<td>0.8414</td>
<td>(0.0086, 0.0086, 0.9828)</td>
</tr>
<tr>
<td>15</td>
<td>0.8423</td>
<td>(0.0057, 0.0057, 0.9886)</td>
</tr>
<tr>
<td>20</td>
<td>0.8427</td>
<td>(0.0043, 0.0043, 0.9914)</td>
</tr>
<tr>
<td>25</td>
<td>0.8430</td>
<td>(0.0034, 0.0034, 0.9932)</td>
</tr>
<tr>
<td>30</td>
<td>0.8432</td>
<td>(0.0028, 0.0028, 0.9944)</td>
</tr>
</tbody>
</table>

5 Conclusions

In this paper, we extend the work by [Wang (2008)](#) by considering the confidence coefficient of \(d\) SCIs instead of \(d-1\) SCIs. The confidence coefficients obtained using our method are still true values, not estimates. If the method in [Wang (2008)](#) is used to find the confidence coefficient, the result is only an upper bound. In order to identify the correct confidence coefficient, extra points should be taken into consideration. However, for certain power-divergence SCIs both methods give the same results. The examples within the power-divergence family show that the confidence coefficients are often attained at \(p_i = L(1)\), \(i = 1, \ldots, d - 1\). This is not always the case, see [Jin (2013)](#) for the binomial case. The method can be generalized to SCIs for difference in the proportions \(p_i - p_j\) or ratios \(p_i/p_j\). We briefly discussed these in this paper.

Simulation studies suggest that the confidence coefficients are always achieved at the cross points even though we have to deal with the existence of possible local minima within some subsets. Within some subsets, the coverage probability is constant. If we exclude this case from the local minima problem, it is still an open question to prove the non-existence of local minima. If we only look at the confidence coefficient, the F-S SCIs \(^{(7)}\) for the multinomial proportions are preferred to the power-divergence SCIs with \(\lambda = 2/3, 1\) since they have higher confidence.
coefficients.

References


A Mathematical Appendix

Proof of Lemma 3. 1. If \( a_{d-1}(x_1, \ldots, x_{d-2}) = 0 \) and \( b_{d-1}(x_1, \ldots, x_{d-2}) = 0 \) for all \( x_1, \ldots, x_{d-2} \)

\[
CP_d(p_{d-1}|p_1, \ldots, p_d) = \sum_{x_1=a_1}^{b_1} \sum_{x_2=a_2(x_1)}^{b_2(x_1)} \cdots \sum_{x_{d-2}=a_{d-2}(x_1, \ldots, x_{d-3})}^{b_{d-2}(x_1, \ldots, x_{d-2})} \left[ \frac{n!}{x_1! \ldots x_d!} P_1^{x_1} \ldots P_{d-2}^{x_{d-2}} P_d^{x_d} \right].
\]

Consider the first derivative with respect to \( p_{d-1} \)

\[
\frac{\partial}{\partial p_{d-1}} CP_d(p_{d-1}|p_1, \ldots, p_d) = \sum_{x_1=a_1}^{b_1} \sum_{x_2=a_2(x_1)}^{b_2(x_1)} \cdots \sum_{x_{d-2}=a_{d-2}(x_1, \ldots, x_{d-3})}^{b_{d-2}(x_1, \ldots, x_{d-2})} \left[ \frac{n!}{x_1! \ldots x_d!} P_1^{x_1} \ldots P_{d-2}^{x_{d-2}} P_d^{x_d} \left( -x_d x_d p_d^{x_d-1} \right) \right],
\]
since \( x_1 + \cdots + b_{d-1}(x_1, \ldots, x_{d-2}) < n \) for some \( x_1, \ldots, x_{d-2} \). Thus
\[
\frac{\partial}{\partial p_{d-1}} CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) < 0,
\]
and subsequently \( CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) \) is a decreasing function in \( p_{d-1} \).

2. If \( a_{d-1}(x_1, \ldots, x_{d-2}) = b_{d-1}(x_1, \ldots, x_{d-2}) \neq 0 \) and, for all \( x_1, \ldots, x_{d-2}, x_1 + \cdots + b_{d-1}(x_1, \ldots, x_{d-2}) = n \), then \( x_d \) is always 0.

\[
CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) = \sum_{x_1=a_1}^{b_1} \sum_{x_2=a_2(x_1)}^{b_2(x_1)} \cdots \sum_{x_{d-1}=a_{d-1}(x_1, \ldots, x_{d-2})}^{b_{d-1}(x_1, \cdots, x_{d-2})} \left[ \frac{n!}{x_1!x_2! \cdots x_d!} p_1^{x_1} p_2^{x_2} \cdots p_{d-2}^{x_{d-2}} (x_{d-1}) p_{d-1}^{x_{d-1} - 1} \right].
\]

Consider the first derivative with respect to \( p_{d-1} \)
\[
\frac{\partial}{\partial p_{d-1}} CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) = \sum_{x_1=a_1}^{b_1} \sum_{x_2=a_2(x_1)}^{b_2(x_1)} \cdots \sum_{x_{d-2}=a_{d-2}(x_1, \ldots, x_{d-2})}^{b_{d-2}(x_1, \cdots, x_{d-2})} \left[ \frac{n!}{x_1!x_2! \cdots x_d!} p_1^{x_1} p_2^{x_2} \cdots p_{d-2}^{x_{d-2}} (x_{d-1}) p_{d-1}^{x_{d-1} - 1} \right],
\]
which is positive. Subsequently \( CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) \) is a increasing function in \( p_{d-1} \).

3. If \( a_{d-1}(x_1, \ldots, x_{d-2}) = b_{d-1}(x_1, \ldots, x_{d-2}) = 0 \) and, for all \( x_1, \ldots, x_{d-2}, x_1 + \cdots + b_{d-1}(x_1, \ldots, x_{d-2}) = n \)

\[
CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) = \sum_{x_1=a_1}^{b_1} \sum_{x_2=a_2(x_1)}^{b_2(x_1)} \cdots \sum_{x_{d-2}=a_{d-2}(x_1, \ldots, x_{d-2})}^{b_{d-2}(x_1, \cdots, x_{d-2})} \left[ \frac{n!}{x_1!x_2! \cdots x_d!} p_1^{x_1} p_2^{x_2} \cdots p_{d-2}^{x_{d-2}} \right].
\]

Thus no \( p_{d-1} \) is involved in the function any more. \( CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) \) is fixed when \( p_1, \ldots, p_{d-2} \) are fixed.
4. For other values of $a_{d-1}$ and $b_{d-1}$

\[
\frac{\partial}{\partial p_{d-1}} CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) = \frac{\partial}{\partial p_{d-1}} \sum_{x_1=a_{d-1}}^{b_{d-1}(x_1)} \cdots \sum_{x_{d-1}=a_{d-1}(x_1)}^{b_{d-1}(x_1, \ldots, x_{d-2})} \left[ \frac{n!}{x_1! x_2! \cdots x_{d-1}!} p_1^{x_1} p_2^{x_2} \cdots p_{d-2}^{x_{d-2}} \right] \\
= \sum_{x_1=a_{d-1}}^{b_{d-1}(x_1)} \sum_{x_2=a_{d-1}(x_1)}^{b_{d-1}(x_1, \ldots, x_{d-2})} \cdots \sum_{x_{d-1}=a_{d-1}(x_1, \ldots, x_{d-2})}^{b_{d-1}(x_1, \ldots, x_{d-2})} \left[ \frac{n!}{x_1! x_2! \cdots x_{d-1}!} p_1^{x_1} p_2^{x_2} \cdots p_{d-2}^{x_{d-2}} \right] \left( x_{d-1} p_{d-1}^{-x_{d-1}} - x_d p_{d-1}^{-x_{d-1}} + x_d p_{d-1}^{x_d-1-x_{d-1}-1} \right).
\]

We can do this even when $x_{d-1}$ or $x_d$ is 0, since in that case $x_{d-1} p_{d-1}^{-x_{d-1}} = 0$ and $x_d p_{d-1}^{-x_{d-1}} = 0$. Thus \( \frac{\partial}{\partial p_{d-1}} CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) = \frac{v_0}{p_{d-1}} (B_t(p_{d-1}) - C_t(p_{d-1})) \) and \( \frac{\partial}{\partial p_{d-1}} CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) = 0 \) if and only if $B_t(p_{d-1}) - C_t(p_{d-1}) = 0$. Note that $p_{d-1}^{-x_{d-1}} p_1^{x_1} \cdots p_{d-1}^{x_{d-1}-x_{d-1}}$ is an increasing function in $p_{d-1}$, thus $B$ and $C$ are increasing functions in $p_{d-1}$.

\[
\frac{\partial}{\partial p_{d-1}} B = \sum_{x_1=a_{d-1}}^{b_{d-1}(x_1)} \sum_{x_2=a_{d-1}(x_1)}^{b_{d-1}(x_1, \ldots, x_{d-2})} \cdots \sum_{x_{d-1}=a_{d-1}(x_1, \ldots, x_{d-2})}^{b_{d-1}(x_1, \ldots, x_{d-2})} \left[ \frac{n!}{x_1! x_2! \cdots x_{d-1}!} p_1^{x_1} p_2^{x_2} \cdots p_{d-2}^{x_{d-2}} x_{d-1} \right] \\
= \left( x_{d-1} p_{d-1}^{-x_{d-1}} - x_1 \cdots p_{d-1}^{x_{d-1}-x_{d-1}-1} + (x_1 + \cdots + x_{d-1}) p_{d-1}^{-x_{d-1}} \right) p_{d-1}^{-x_{d-1}} p_d^{x_d-1-x_{d-1}-1},
\]

\[
\frac{\partial}{\partial p_{d-1}} C = \sum_{x_1=a_{d-1}}^{b_{d-1}(x_1)} \sum_{x_2=a_{d-1}(x_1)}^{b_{d-1}(x_1, \ldots, x_{d-2})} \cdots \sum_{x_{d-1}=a_{d-1}(x_1, \ldots, x_{d-2})}^{b_{d-1}(x_1, \ldots, x_{d-2})} \left[ \frac{n!}{x_1! x_2! \cdots x_{d-1}!} p_1^{x_1} p_2^{x_2} \cdots p_{d-2}^{x_{d-2}} x_d \right] \\
= \left( (x_{d-1} + 1) p_{d-1}^{-x_{d-1}} - x_1 \cdots p_{d-1}^{x_{d-1}-x_{d-1}-1} + (x_1 + \cdots + x_{d-1} + 1) p_{d-1}^{-x_{d-1}} p_d^{x_d-1-x_{d-1}-2},
\right)
\]

which are both positive and increasing in $p_{d-1}$. Thus \( \frac{\partial^2}{\partial p_{d-1}^2} B > 0 \) and \( \frac{\partial^2}{\partial p_{d-1}^2} C > 0. \)
This means $B$ and $C$ are strictly increasing convex functions.

If $B - C = 0$ has no solution, then $CP_d(p_{d-1}|p_1, \cdots, p_{d-2})$ is either an increasing function or a decreasing function on the whole support. However

$$
\lim_{p_{d-1} \to 0} CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) = 0,
$$

$$
\lim_{p_{d-1} \to 1 - \sum_{i=1}^{d-2} p_i} CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) = 0.
$$

which is a contradiction.

If $B - C = 0$ has exactly one solution, say $p_{d-1} = \tilde{p}_{d-1}$, then $CP_d(p_{d-1}|p_1, \cdots, p_{d-2})$ is a unimodal function attaining the maximum at $p_{d-1} = \tilde{p}_{d-1}$.

If $B - C = 0$ has two solutions, $CP_d(p_{d-1}|p_1, \cdots, p_{d-2})$ is still a unimodal function. Because at the boundary

$$
\lim_{p_{d-1} \to 0} CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) = 0,
$$

$$
\lim_{p_{d-1} \to 1 - \sum_{i=1}^{d-2} p_i} CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) = 0,
$$

so one of the solution is not a maximum or minimum point.

\[ \square \]

**Proof of Theorem** \(^7\) The proof consists of two parts. In the first part, we assume that $L(0) = 0$ and $U(n) = 1$ which implies $L(x) \in (0, 1)$ and $U(x) \in (0, 1)$ for all $x \neq 0, n$. The second part considers a more general case, where $L(x) < 0$ for some nonzero $x$. Subsequently, $U(x) > 1$ for some $x \neq n$. We start with the first part.

(1). Without loss of generality, we assume we have $p_{d-1} \in (0, v_1)$ first. So unless we observe $x_{d-1} = 0$ and subsequently $a_{d-1} = 0$ and $b_{d-1}(x_1, \cdots, x_{d-2}) = 0$ for all $x_1, \cdots, x_{d-2}$, the SCIs cannot cover $p_{d-1}$. From Lemma \(^3\) we know that within those subsets $CP_d(p_{d-1}|p_1, \cdots, p_{d-2})$ is a decreasing function of $p_{d-1}$ if $p_d \notin (0, v_1)$. This means the infimum of $CP_d(p_{d-1}|p_1, \cdots, p_{d-2})$ within this subset is attained at $p_{d-1} \to v_1$. Assumption 1(5) means we shall have $U(x) > L(x + 1)$ for all
\( x = 0, \cdots, n - 1 \). Together with Assumption 1(3), we know that the infimum is actually attained at another subset nearby where \( p_{d-1} = v_1 \) and \( p_1, \cdots, p_{d-2} \) stay the same. This applies to all \( CP_d(p_1|p_1, \cdots, p_{i-1}, p_{i+1}, \cdots, p_{d-1}) \). Note that we can treat any \( p_j \) as a linear combination of the others. For example, if we let \( p_1 \) be a linear combination of \( p_2, \cdots, p_d \) and consider \( CP_d(p_d|p_2, \cdots, p_{d-1}) \), then the infimum occurs at \( p_d = v_1 \). If \( p_d \in (0, v_1) \) as well, then Lemma 3 shows \( CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \) is a constant function in \( p_{d-1} \). In this case, there should be some \( p_i \notin (0, v_1) \). We choose such \( p_i \) to be the linear dependent argument and fix \( p_1, \cdots, p_{i-1}, p_{i+1}, \cdots, p_{d-2}, p_d \), so \( CP_d(p_{d-1}|p_1, \cdots, p_{i-1}, p_{i+1}, \cdots, p_{d-2}, p_d) \) is a decreasing function in \( p_{d-1} \). This means the infimum is attained at \( p_{d-1} \rightarrow v_1 \). Thus we do not need to consider the subsets where some endpoints are 0.

Second, we consider the subsets where no endpoints are 0 and \( v'_{ij} \neq v'_{ij} \) for all \( j = 1, \cdots, d \). We can divide those subsets into two kinds: with local minima and without local minima.

Case I, assume \( CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \) has no local minimum in the subsets. Given \( p_1, \cdots, p_{d-2} \), \( CP_d(p_{d-1}|p_1, \cdots, p_{d-2}) \) is unimodal or monotone. So the infimum occurs when \( p_{d-1} \) goes to the infimum or supremum that it can take. Here because we introduced \((v'_{id}, v'_{id+1})\) in \([13]\), \( p_{d-1} \) may not attain \( v'_{id-1} \) or \( v'_{id+1} \) for all values of \( p_1, \cdots, p_{d-2} \). Thus when \( p_{d-1} \rightarrow v'_{id-1} \) or \( v'_{id+1} \) or on the hyperplane \( p_d \rightarrow v_{id} \) or \( p_d \rightarrow v_{id-1} \), the infimum is attained. By induction, the infimum of \( CP_d(p_j|p_1, \cdots, p_{j-1}, p_{j+1}, \cdots, p_{d-1}) \) is attained when \( p_j \rightarrow v'_{ij} \) or \( v'_{ij+1} \) or on the hyperplane \( p_d \rightarrow v_{id} \) or \( p_d \rightarrow v_{id-1} \) if \( p_1, \cdots, p_{j-1}, p_{j+1}, \cdots, p_{d-1} \) are fixed for \( j = 1, \cdots, d - 1 \). And further let \( p_1 \) be a linear combination of \( p_2, \cdots, p_d \), then the infimum occurs at \( p_2 \rightarrow v'_{i_2} \) or \( v'_{i_2+1} \) or \( p_d \rightarrow v'_{id} \) or \( v'_{id+1} \) or on the hyperplane \( p_1 \rightarrow v_{i_1} \) or \( p_1 \rightarrow v_{i_1-1} \). Above all, the infimum occurs at when \( p_j \rightarrow v'_{ij} \) or \( v'_{ij+1} \) for \( j \neq k \) where \( p_k \) is a linear combination of the others \( k = 1, \cdots, d \).

Case II, assume there is a local minimum in the subset, perhaps many local minima, Given \( p_1, \cdots, p_{d-2} \), the infimum occurs when \( p_{d-1} \) goes to the infimum or
supremum that it can take or when \( B(p_{d-1}) = C(p_{d-1}) \). By induction and the symmetric property, the infimum occurs when \( p_j \to v'_{i_j} \) or \( v'_{i_j+1} \) or when \( B(p_j) = C(p_j) \) for \( j \neq k \) where \( p_k \) is a linear combination of the others \( k = 1, \ldots, d \).

Use the arguments below Equation (14), we know '→' can be replaced by '='. Combine all the arguments above, we conclude that the confidence coefficient is obtained at some of the points in

\[
E_1 \cup \cdots \cup E_d \cup \left( \bigcup_{t=1}^{s'} (M_{1,t} \cup \cdots \cup M_{d,t}) \right).
\]

(2). Assume \( L(x) < 0 \) for some nonzero \( x \). For \( p_{d-1} \in (0, v_1) \), there could be multiple \( x \) where the resulting confidence interval covers \( p_{d-1} \). When it happens, the conditional coverage probability \( CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) \) is not a decreasing function in \( (0, v_1) \) any more. So we have to take the possible unimodality into account. Therefore, the infimum of \( CP_d(p_{d-1}|p_1, \ldots, p_{d-2}) \) is attained at either \( p_{d-1} = v_1 \) or \( p_{d-1} \to 0 \). Although zero parameters are not allowed in the parameter space \( \Theta \), it is still valid to calculate the coverage probability in case of zero cell probabilities. Such points construct \( E_0 \). The remaining part of the proof is the same as the first part, so the confidence coefficient is attained at

\[
E_0 \cup E_1 \cup \cdots \cup E_d \cup \left( \bigcup_{t=1}^{s'} (M_{1,t} \cup \cdots \cup M_{d,t}) \right).
\]

\( \square \)