# A family of simultaneous confidence intervals for multinomial proportions 

Chia-Ding Hou ${ }^{\text {a }}$, Jengtung Chiang ${ }^{\text {b }}$, John Jen Tai ${ }^{\mathrm{c}, *}$<br>${ }^{\text {a }}$ Department of Statistics, Fu Jen Catholic University, Taipei Hsien, 242, Taiwan, ROC<br>${ }^{\mathrm{b}}$ Department of Statistics, National Chengchi University, Taipei, Taiwan, ROC<br>${ }^{\text {c }}$ Division of Biostatistics, Institute of Epidemiology, National Taiwan University, 1 Jen-Ai Road, Section<br>1, Taipei 100, Taiwan, ROC

Received 1 July 2001; received in revised form 1 April 2002


#### Abstract

A family of simultaneous confidence intervals (SCIs) for multinomial proportions is proposed by inverting the power-divergence statistics and the best SCIs in the family is determined by Monte-Carlo technique. Numerical comparisons of this method with the other alternatives are presented. Simulation results indicate that the new procedure is preferable to all its competitors in most cases.


(C) 2003 Elsevier Science B.V. All rights reserved.

Keywords: Coverage probability; Monte-Carlo method; Power-divergence statistic; Simultaneous confidence intervals; Sparse data

## 1. Introduction

Simultaneous confidence intervals for multinomial proportions are very useful in many aspects of statistical applications such as quality control (Quesenberry and Hurst, 1964; Goodman, 1965), simulation studies (Hurtubise, 1969; Angers, 1984), opinion polling (Fitzpatrick and Scott, 1987), crime studies (Sison and Glaz, 1995), and cytogenetics studies (Hou et al. 1999, 2001). The issue on constructing simultaneous confidence intervals for multinomial proportions, hence, has been studied extensively in the literature. Let $\mathbf{N}=\left(N_{1}, N_{2}, \ldots, N_{k}\right)$ be a vector of observed cell frequencies in

[^0]a sample of size $n=\sum_{i=1}^{k} N_{i}$ from a multinomial distribution with the vector of cell probabilities $\mathbf{P}=\left(P_{1}, P_{2}, \ldots, P_{k}\right)$. We are interested in constructing a set of simultaneous confidence intervals
$$
S(\mathbf{N})=\left\{\mathbf{P} \mid P_{i} \in S_{i}, i=1,2, \ldots, k\right\},
$$
where $S(\mathbf{N})$ satisfies
\[

$$
\begin{equation*}
P_{r}[\mathbf{P} \in S(\mathbf{N})]=P_{r}\left[\bigcap_{i=1}^{k}\left(P_{i} \in S_{i}\right)\right] \geqslant 1-\alpha \tag{1}
\end{equation*}
$$

\]

for a specified significance level $\alpha$.
Quesenberry and Hurst (1964), to our knowledge, was the first to discuss the issue. They derived the following large-sample simultaneous confidence intervals based on the chi-square approximation of the Pearson's statistic:

$$
S_{1}(\mathbf{N})=\left\{\mathbf{P} \left\lvert\, P_{i} \in \frac{A+2 N_{i} \mp\left\{A\left[A+4 N_{i}\left(n-N_{i}\right) / n\right]\right\}^{1 / 2}}{2(n+A)}\right., i=1,2, \ldots k\right\},
$$

where $A$ is the upper $100 \times \alpha$ th percentile of the chi-square distribution with $k-1$ degrees of freedom. Following the work of Quesenberry and Hurst (1964), Goodman (1965) proposed a set of shorter-length intervals:

$$
S_{2}(\mathbf{N})=\left\{\mathbf{P} \left\lvert\, P_{i} \in \frac{B+2 N_{i} \mp\left\{B\left[B+4 N_{i}\left(n-N_{i}\right) / n\right]\right\}^{1 / 2}}{2(n+B)}\right., i=1,2, \ldots k\right\},
$$

where $B$ is the upper $100 \times(\alpha / k)$ th percentile of the chi-square distribution with one degree of freedom. In addition, Bailey (1980) utilized an angular transformation and a square root transformation to binomial variates and proposed two procedures as follows:

$$
\begin{aligned}
& S_{3}(\mathbf{N})=\left\{\mathbf{P} \left\lvert\, P_{i} \in\left\{\sin \left[\sin ^{-1}\left(\sqrt{\frac{N_{i}+\frac{3}{8}}{n+\frac{3}{4}}}\right) \mp \sqrt{\frac{B}{4 n+2}}\right]\right\}^{2}\right., i=1,2, \ldots k\right\}, \\
& \left.S_{4}(\mathbf{N})=\left\{\mathbf{P} \left\lvert\, P_{i} \in\left\{\sqrt{\frac{N_{i}+\frac{3}{8}}{n+\frac{1}{8}}} \mp \sqrt{C\left(C+1-\frac{N_{i}+\frac{3}{8}}{n+\frac{1}{8}}\right.}\right)\right.\right\} /(C+1)^{2}, i=1,2, \ldots k\right\},
\end{aligned}
$$

where $C=B /(4 n)$. Fitzpatrick and $\operatorname{Scott}$ (1987) applied the normal approximation to derive the following adjusted binomial intervals:

$$
S_{5}(\mathbf{N})=\left\{\mathbf{P} \left\lvert\, P_{i} \in \frac{N_{i}}{n} \mp \frac{D}{2 \sqrt{n}}\right., i=1,2, \ldots k\right\},
$$

where $D$ is the upper $100 \times(\alpha / 4)$ th percentile of the standard normal distribution. More recently, Sison and Glaz (1995) used two approximations for the rectangular multinomial probabilities to derive two "computer-based" procedures. Their procedures have no closed form and must be solved using a computer. To obtain their results, let $V_{i}$ and $Y_{i}, i=1,2, \ldots, k$, be independent Poisson random variables with mean $n_{i}$ and its truncation to $\left[n_{i}-\tau, n_{i}+\tau\right]$, respectively, where $\tau$ is some constant. Let $N_{1}^{*}, N_{2}^{*}, \ldots, N_{k}^{*}$
be the cell frequencies in a sample of $n$ observations from a multinomial distribution with cell probabilities ( $N_{1} / n, \ldots, N_{k} / n$ ). Define

$$
\begin{aligned}
& \mu_{i}=E\left(Y_{i}\right), \\
& \sigma_{i}^{2}=V\left(Y_{i}\right), \\
& \mu_{(r)}=E\left[Y_{i}\left(Y_{i}-1\right) \ldots\left(Y_{i}-r+1\right)\right], \\
& \mu_{r, i}=E\left(Y_{i}-\mu_{i}\right)^{r}, \\
& \gamma_{1}=\frac{\frac{1}{k} \sum_{i=1}^{k} \mu_{3, i}}{\sqrt{k}\left(\frac{1}{k} \sum_{i=1}^{k} \sigma_{i}^{2}\right)^{\frac{3}{2}}}, \\
& \gamma_{2}=\frac{\frac{1}{k} \sum_{i=1}^{k} \mu_{4, i}-3 \sigma_{i}^{4}}{\sqrt{k}\left(\frac{1}{k} \sum_{i=1}^{k} \sigma_{i}^{2}\right)^{2}}, \\
& f_{e}(x)=\left(\frac{1}{\sqrt{2 \pi}} \mathrm{e}^{\frac{-x^{2}}{2}}\right)\left\{1+\frac{\gamma_{1}}{6}\left(x^{3}-3 x\right)+\frac{\gamma_{2}}{24}\left(x^{4}-6 x^{2}+3\right)\right. \\
& \left.\quad+\frac{\gamma_{1}^{2}}{72}\left(x^{6}-15 x^{4}+45 x^{2}-15\right)\right\}, \\
& v(\tau)=\frac{n!}{n^{n} \mathrm{e}^{-n}}\left\{\prod_{i=1}^{k} P_{r}\left[n_{i}-\tau \leqslant V_{i} \leqslant n_{i}+\tau\right]\right\} f_{e}\left(\frac{n-\sum_{i=1}^{k} \mu_{i}}{\sqrt{\sum_{i=1}^{k} \sigma_{i}^{2}}}\right) \frac{1}{\sqrt{\sum_{i=1}^{k} \sigma_{i}^{2}}}
\end{aligned}
$$

and

$$
\omega(\xi)=\alpha_{1, m} \prod_{j=m+1}^{k}\left(\frac{\alpha_{j-m+1, j}}{\alpha_{j-m+1, j-1}}\right),
$$

where

$$
\alpha_{i, j}=P\left[n_{i}-\xi \leqslant N_{i}^{*} \leqslant n_{i}+\xi, \ldots, n_{j}-\xi \leqslant N_{j}^{*} \leqslant n_{j}+\xi\right] \quad \text { for } 1 \leqslant i<j \leqslant k
$$

Using the algorithm that approximates multinomial probabilities given by Levin (1981), their first procedure was derived as follows:

$$
S_{6}(\mathbf{N})=\left\{\mathbf{P} \left\lvert\, \frac{N_{i}}{n}-\frac{\tau}{n} \leqslant P_{i} \leqslant \frac{N_{i}}{n}+\frac{\tau+2 \gamma}{n}\right., i=1,2, \ldots, k\right\},
$$

where the integer $\tau$ satisfies the condition $v(\tau)<1-\alpha<v(\tau+1)$, and $\gamma=(1-\alpha)-$ $v(\tau) / v(\tau+1)-v(\tau)$. They also utilized the negative dependence structure inherent in the multinomial distribution and some related probability inequalities proposed by Glaz and Johnson (1984) to derive the second type of SCIs as follows:

$$
S_{7}(\mathbf{N})=\left\{\mathbf{P} \left\lvert\, \frac{N_{i}}{n}-\frac{\xi}{n} \leqslant P_{i} \leqslant \frac{N_{i}}{n}+\frac{\xi+2 \delta}{n}\right., i=1,2, \ldots, k\right\},
$$

where the constant $\xi$ is chosen such that $\omega(\xi)<1-\alpha<\omega(\xi+1)$ and $\delta=(1-\alpha)-$ $\omega(\xi) / \omega(\xi+1)-\omega(\xi)$. According to simulation studies, they concluded that $S_{6}(\mathbf{N})$ and $S_{7}(\mathbf{N})$ performed equally well, and since $S_{7}(\mathbf{N})$ is very computationally time consuming, they thought that $S_{7}(\mathbf{N})$ is impractical from a computational viewpoint and recommended $S_{6}(\mathbf{N})$ alone.

In Section 2, a new procedure for constructing simultaneous confidence intervals for multinomial proportions is proposed. In Section 3, numerical comparisons are presented to see its superb performance over those previously discussed in the literature. In Section 4 possible improvements on the new procedure are discussed.

## 2. A power-divergence-type simultaneous confidence intervals

### 2.1. A family of simultaneous confidence intervals for multinomial proportions

In this section a family of simultaneous confidence intervals for the vector of parameters of a multinomial distribution will be constructed by inverting the power-divergence statistics introduced by Cressie and Read (1984). Since Pearson's chi-square statistic is one member of the power-divergence statistics, it is obvious that the simultaneous confidence intervals proposed by Quesenberry and Hurst (1964) and Goodman (1965) are special cases of this family.

Cressie and Read (1984) proposed the following power-divergence family of statistics:

$$
\begin{align*}
I(\mathbf{P}, \lambda) & =\frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{k} N_{i}\left[\left(\frac{N_{i}}{n P_{i}}\right)^{\lambda}-1\right] \\
& =\frac{2 n}{\lambda(\lambda+1)} \sum_{i=1}^{k} \hat{P}_{i}\left[\left(\frac{\hat{P}_{i}}{P_{i}}\right)^{\lambda}-1\right], \quad-\infty<\lambda<\infty \tag{2}
\end{align*}
$$

where $\left(\hat{P}_{1}, \ldots, \hat{P}_{k}\right)=\left(N_{1} / n, \ldots, N_{K} / n\right)$ and $\lambda$ is the family parameter. For $\lambda=0$ and $-1, I(\mathbf{P}, \lambda)$ is defined using continuity. It is well known that the power-divergence statistic provides an important link with many famous statistics, such as the Pearson's chi-square statistic $(\lambda=1)$, the Neyman-modified chi-square statistic $(\lambda=-2)$, the likelihood ratio statistic $(\lambda \rightarrow 0)$, and the Freeman-Tukey statistic $(\lambda=-0.5)$. This link provides a mechanism to derive more general results about the behavior of these statistics in both large and small samples, it also hints that some new statistics that are highly competitive may be constructed. Cressie and Read (1988) further indicated that each member of the power-divergence family of statistics has the same asymptotic chi-square distribution with $k-1$ degrees of freedom. Moreover, each of them behaves algebraically as well as statistically like the Pearson's chi-square statistic and can be interpreted, geometrically, as a sum of weighted squared differences.

Suppose the approximation is good, then

$$
P_{r}\left[I(\mathbf{P}, \lambda) \leqslant \chi_{\alpha}^{2}(k-1)\right] \approx 1-\alpha, \quad-\infty<\lambda<\infty, \quad 0<\alpha<1 .
$$

For fixed $n, k, \alpha$, and $\lambda$, the inequality

$$
\begin{equation*}
I(\mathbf{P}, \lambda) \leqslant \chi_{\alpha}^{2}(k-1) \tag{3}
\end{equation*}
$$

defines a region in the parameter space, and if the equal sign is taken the equation gives the bounding surface of the region. The intersection of the hyperplane represented by

$$
\begin{equation*}
\sum_{i=1}^{k} P_{i}=1 \tag{4}
\end{equation*}
$$

with this surface is a $(k-1)$-dimensional hypercurve. Applying the technique of Lagrange multiplier, the minimum and maximum values on the curve can be found by solving the following equation:

$$
\begin{equation*}
\left(1-\hat{P}_{i}\right)^{\lambda+1} P_{i}^{\lambda}+\hat{P}_{i}^{\lambda+1}\left(1-P_{i}\right)^{\lambda}=A_{\lambda} P_{i}^{\lambda}\left(1-P_{i}\right)^{\lambda} \tag{5}
\end{equation*}
$$

for each $P_{i}(i=1,2, \ldots, k)$, where

$$
A_{\lambda}=\frac{\lambda(\lambda+1)}{2 n} \chi_{\alpha}^{2}(k-1)+1 .
$$

(A proof for this statement is given in Appendix A.) Following Eq. (5), we can construct a set of simultaneous confidence intervals as follows:

$$
S(\mathbf{N}, \lambda)=\left\{\mathbf{P} \mid P_{i} \in\left[P_{i \lambda L}, P_{i \lambda U}\right], i=1,2, \ldots, k\right\}, \quad-\infty<\lambda<\infty,
$$

where $P_{i \lambda L}$ and $P_{i \lambda U}$ are the minimum and maximum solutions for Eq. (5). Using the similar trick Goodman (1965) suggested, we replace $A_{\lambda}$ in (5) with $B_{\lambda}$, where

$$
B_{\lambda}=\frac{\lambda(\lambda+1)}{2 n} \chi_{\alpha / k}^{2}(1)+1
$$

and obtain a set of shorter-length simultaneous confidence intervals:

$$
S_{\mathbf{P D}}(\mathbf{N}, \lambda)=\left\{\mathbf{P} \mid P_{i} \in\left[P_{i \lambda L}^{4}, P_{i \lambda U}^{4}\right], i=1,2, \ldots, k\right\}, \quad-\infty<\lambda<\infty,
$$

where $P_{i \lambda L}^{4}, P_{i \lambda U}^{4}$ are the minimum and maximum solutions for Eq. (5) with $A_{\lambda}$ substituted by $B_{\lambda}$. When the interval endpoints are not explicit functions of the data, numerical methods such as Newton-Raphson's method can be used to solve the problem.

In particular, when $\lambda=1$, we have

$$
A_{\lambda}=A_{1}=\frac{1}{n} \chi_{\alpha}^{2}(k-1)+1
$$

and Eq. (5) becomes

$$
\left(1-\hat{P}_{i}\right)^{2} P_{i}+\hat{P}_{i}^{2}\left(1-P_{i}\right)=A_{1} P_{i}\left(1-P_{i}\right), \quad i=1,2, \ldots, k
$$

It follows that

$$
A_{1} P_{i}^{2}+\left(1-2 \hat{P}_{i}-A_{1}\right) P_{i}+\hat{P}_{i}^{2}=0, \quad i=1,2, \ldots, k .
$$

Hence

$$
P_{i}=\frac{A+2 N_{i} \mp\left\{A\left[A+4 N_{i}\left(n-N_{i}\right) / n\right]\right\}^{1 / 2}}{2(n+A)}, \quad i=1,2, \ldots, k .
$$

We thus have

$$
S(\mathbf{N}, 1)=S_{1}(\mathbf{N})
$$

Similarly, we can show that

$$
S_{\mathbf{P D}}(\mathbf{N}, 1)=S_{2}(\mathbf{N}) .
$$

Hence, the set of simultaneous confidence intervals $S_{1}(\mathbf{N})$ proposed by Quesenberry and Hurst (1964) is a special case of the family of simultaneous confidence intervals

$$
\begin{equation*}
F_{1}=\{S(\mathbf{N}, \lambda), \lambda \in R\} \tag{6}
\end{equation*}
$$

and the set of simultaneous confidence intervals $S_{2}(\mathbf{N})$ introduced by Goodman (1965) is a member of the family of shorter-length simultaneous confidence intervals

$$
\begin{equation*}
F_{2}=\left\{S_{\mathbf{P D}}(\mathbf{N}, \lambda), \lambda \in R\right\} . \tag{7}
\end{equation*}
$$

### 2.2. The best power-divergence simultaneous confidence intervals

The result obtained in Section 2.1 assumed that $\lambda$ is a fixed number. In this subsection, the optimal choice of $\lambda$ for the family of SCIs will be discussed.

Based on some simulation results, Cressie and Read (1988, p. 63) concluded that in almost all cases a reasonable choice of $\lambda$ would lie in the range $(-1,2]$. Therefore, it appears reasonable that we also restrict our attention here to the following subfamily:

$$
F=\left\{S_{\mathbf{P D}}(\mathbf{N}, \lambda), \lambda \in(-1,2]\right\} .
$$

Let $V(\lambda)$ denote the volume of the rectangular region of $S_{\mathbf{P D}}(\mathbf{N}, \lambda)$ and $\Lambda$ be

$$
\Lambda=\left\{\lambda \mid P_{r}\left[\mathbf{P} \in S_{\mathbf{P D}}(\mathbf{N}, \lambda)\right] \geqslant 1-\alpha, \lambda \in(-1,2]\right\} .
$$

Thus, for any given data set $\mathbf{N}$ and $\alpha$, finding the best power divergence SCIs among the class $F$ is equivalent to seeking a point $\lambda^{*}$ in $\Lambda$ that minimizes $V(\lambda)$, or at least satisfactorily closes to the minimum of $V(\lambda)$. That is, we are looking for $\lambda^{*}$ so that

$$
\operatorname{Min}_{\lambda \in \Lambda} V(\lambda)=V\left(\lambda^{*}\right) .
$$

Hence, in determining the best power-divergence SCIs among the subfamily $F$ for any given data set, $\Lambda$ needs to be defined first, the optimal value $\lambda^{*}$ can then be pursued. Unfortunately, elements in $\Lambda$ cannot be expressed explicitly, we, instead, set up $\Lambda$ based on Monte-Carlo simulated samples. Since the corresponding SCIs are derived in terms of the power-divergence statistics, they may be called "the best power-divergence SCIs".

## 3. Simulation studies

We perform a series of simulation studies to compare the proposed methods with those previously discussed in literature. We will consider four data sets for discussion. The first two data sets (Tables 1 and 2) were examined in Sison and Glaz (1995).
Table 1
Comparison of eight $95 \%$ simultaneous confidence interval procedures for the data set with $n=467, N_{1}=56, N_{2}=72, N_{3}=73, N_{4}=59, N_{5}=62, N_{6}=87$,
$N_{7}=58$

| $i$ | Quesenberry-Hurst$S_{1}(\mathbf{N})$ | Goodman$S_{2}(\mathbf{N})$ | Bailey |  | Fitzpatrick-Scott$S_{5}(\mathbf{N})$ | Sison-Glaz |  | $S_{\text {PD }}(\mathbf{N}, 0.21)$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | $S_{3}(\mathbf{N})$ | $S_{4}(\mathbf{N})$ |  | $S_{6}(\mathbf{N})$ | $S_{7}(\mathbf{N})$ |  |
| 1 - | 0.0760 .184 | 0.0850 .166 | 0.0830 .164 | 0.0830 .163 | 0.0680 .172 | 0.0790 .165 | 0.0790 .165 | 0.0840 .165 |
| 2 | 0.1040 .223 | 0.1150 .204 | 0.1130 .202 | 0.1120 .202 | 0.1020 .206 | 0.1140 .199 | 0.1140 .199 | 0.1130 .203 |
| $3 \sim$ | 0.1060 .225 | 0.1160 .207 | 0.1140 .205 | 0.1140 .204 | 0.1050 .208 | 0.1160 .201 | 0.1160 .201 | 0.1150 .205 |
| 4 | 0.0810 .191 | 0.0910 .174 | 0.0890 .171 | 0.0880 .171 | 0.0750 .178 | 0.0860 .171 | 0.0860 .171 | 0.0890 .172 |
| 5 | 0.0870 .198 | 0.0960 .181 | 0.0940 .178 | 0.0940 .178 | 0.0810 .185 | 0.0920 .177 | 0.0920 .177 | 0.0950 .179 |
| 6 | 0.1310 .258 | 0.1430 .239 | 0.1410 .238 | 0.1400 .237 | 0.1340 .238 | 0.1460 .231 | 0.1460 .231 | 0.1420 .238 |
| 7 | 0.0800 .188 | 0.0890 .171 | 0.0870 .169 | 0.0860 .168 | 0.0720 .176 | 0.0840 .169 | 0.0840 .169 | 0.0870 .170 |
| Coverage rate (the exact $p$-value) | $\begin{aligned} & 0.9969 \\ & (1.0000) \end{aligned}$ | $\begin{aligned} & 0.9483 \\ & (0.2238) \end{aligned}$ | $\begin{aligned} & 0.9443^{* *} \\ & \left(5.315 \times 10^{-3}\right) \end{aligned}$ | $\begin{aligned} & 0.9462^{*} \\ & \left(4.389 \times 10^{-2}\right) \end{aligned}$ | $\begin{aligned} & 0.9887 \\ & (1.0000) \end{aligned}$ | $\begin{aligned} & 0.9120^{* * * *} \\ & (0.0000) \end{aligned}$ | -------- | $\begin{aligned} & 0.9492 \\ & (0.3634) \end{aligned}$ |
| Volume | $2.553 \times 10^{-7}$ | $3.669 \times 10^{-8}$ | $3.621 \times 10^{-8}$ | $3.576 \times 10^{-8}$ | $1.291 \times 10^{-7}$ | $3.290 \times 10^{-8}$ | $3.267 \times 10^{-8}$ | $3.614 \times 10^{-8}$ |

$\quad *,^{* *},^{* * * *}$ represent significance at $0.05,0.01,0.005,0.001$ significance level, for the problem of testing " $H_{0}$ : coverage probability $\geqslant 0.95^{\prime \prime}$,
respectively.
Table 2


In addition, in order to see their performance under situations with sparse data, an artificial data set (Table 3) as well as a real data set (Table 4) are studied. The former is created based on the data set in Table 1, the latter is about chromosomal fragile sites discussed in Hou et al. $(1999,2001)$. The procedure is as follows:
(a) For any data set $\left(N_{1}, \ldots N_{k}\right)$, estimate $P_{j}^{\prime} s$ by $N_{j} / n$ and use them as if they were the "true" population proportions.
(b) Simulate 10000 multinomial samples from this multinomial population. The computer program was coded in FORTRAN V. The simulated multinomial samples were generated by the IMSL subroutine RNMTN. All simulation studies were performed on a 686 Pentium II 300 personal computer.
(c) Obtain the SCI intervals for each sample. Comparisons are then made in terms of the coverage probability and the rectangular volume, where the coverage probability is defined to be the proportion of the 10000 estimated SCIs that cover the "true" multinomial vector ( $N_{1} / n, \ldots, N_{K} / n$ ).
For each of the eight different confidence regions $S_{1}(\mathbf{N}), S_{2}(\mathbf{N}), \ldots, S_{7}(\mathbf{N}), S_{\mathbf{P D}}\left(\mathbf{N}, \lambda^{*}\right)$, we present the intervals for $P_{i}, i=1,2, \ldots k$, the coverage rate, the exact $p$-values related to the problem of testing " $H_{0}$ : coverage probability $\geqslant 1-\alpha$ ", and the volume of the "rectangular" region. In determining the best power-divergence SCIs $S_{\mathbf{P D}}\left(\mathbf{N}, \lambda^{*}\right)$, the optimal choice $\lambda^{*}$ is obtained such that the corresponding coverage probability of the power-divergence SCIs is greater than or equal to the nominal level by statistical testing, and the corresponding volume is the smallest. We carried out the former test based on 50 Monte-Carlo simulated samples (Although we chose to use, conservatively, 50 samples here, we found, based on our extensive simulation studies, we gain essentially nothing by using more than 20 samples).

In Sison and Glaz (1995), they simulated 10000 multinomial samples, and for each sample they calculated the vector $\left(N_{1} / n, \ldots, N_{K} / n\right)$, rather than the SCIs, to obtain the proportion for those vectors which were located in the original SCIs calculated in terms of the observed data $\left(N_{1}, \ldots N_{k}\right)$, and took it to estimate coverage probability. Thus, the simulation results presented here are a little bit different from those provided in Sison and Glaz (1995). Nevertheless, the correctness of the results is our own responsibility. Since procedure $S_{7}(\mathbf{N})$ is very time consuming, the coverage probabilities for procedure $S_{7}(\mathbf{N})$ are not presented.

Recall that the simultaneous confidence intervals are targeted to achieve a coverage probability greater than or equal to a predetermined nominal level (i.e., to satisfy inequality (1)). In Table 1, it can be found that all intervals are conservative except $S_{3}(\mathbf{N}), S_{4}(\mathbf{N}), S_{6}(\mathbf{N})$ (at 0.05 significance level). Among the intervals that are conservative, $S_{\mathbf{P D}}\left(\mathbf{N}, \lambda^{*}\right)$ has the smallest volume. According to Tables 2 and 3, it can be seen that the intervals $S_{1}(\mathbf{N}), S_{5}(\mathbf{N})$, and $S_{\mathrm{PD}}\left(\mathbf{N}, \lambda^{*}\right)$ are conservative. Among the intervals that are conservative, $S_{\mathbf{P D}}\left(\mathbf{N}, \lambda^{*}\right)$ again has the smallest volume. In Table 4, we can see that the coverage probability for $S_{2}(\mathbf{N}), S_{3}(\mathbf{N})$, and $S_{4}(\mathbf{N})$ do not attain the nominal level, while the rest are conservative. Among those that are conservative, $S_{\mathbf{P D}}\left(\mathbf{N}, \lambda^{*}\right)$ once again has the best performance in terms of volume. To sum up, as shown in Tables $1-4, S_{\mathbf{P D}}\left(\mathbf{N}, \lambda^{*}\right)$ has the smallest volume overall and also achieves a coverage closest to 0.95 . Hence, it appears that $S_{\mathbf{P D}}\left(\mathbf{N}, \lambda^{*}\right)$ is the optimal choice. Similar results are also obtained in our extensive simulation studies. Note that
Table 3
Comparison of eight $95 \%$ simultaneous confidence interval procedures for the data set with $n=467, N_{1}=1, N_{2}=1, N_{3}=1, N_{4}=108, N_{5}=111, N_{6}=136$,
$N_{7}=109$

|  | Quesenberry-Hurst$S_{1}(\mathbf{N})$ | Goodman$S_{2}(\mathbf{N})$ | Bailey |  | Fitzpatrick-Scott$S_{5}(\mathbf{N})$ | Sison-Glaz |  | $S_{\text {PD }}(\mathbf{N}, 0.07)$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | $S_{3}(\mathbf{N})$ | $S_{4}(\mathbf{N})$ |  | $S_{6}(\mathbf{N})$ | $S_{7}(\mathbf{N})$ |  |
| 1 | 0.0000 .030 | 0.0000 .019 | 0.0000 .014 | 0.0000 .013 | 0.0000 .054 | 0.0000 .051 | 0.0000 .052 | 0.0000 .014 |
| 2 | 0.0000 .030 | 0.0000 .019 | 0.0000 .014 | 0.0000 .013 | 0.0000 .054 | 0.0000 .051 | 0.0000 .052 | 0.0000 .014 |
| 3 | 0.0000 .030 | 0.0000 .019 | 0.0000 .014 | 0.0000 .013 | 0.0000 .054 | 0.0000 .051 | 0.0000 .052 | 0.0000 .014 |
| 4 | 0.1700 .307 | 0.1830 .288 | 0.1810 .286 | 0.1810 .285 | 0.1790 .283 | 0.1860 .280 | 0.1840 .281 | 0.1820 .286 |
| 5 | 0.1750 .314 | 0.1890 .294 | 0.1870 .293 | 0.1870 .292 | 0.1860 .290 | 0.1930 .287 | 0.1910 .287 | 0.1880 .293 |
| 6 | 0.2230 .371 | 0.2380 .351 | 0.2370 .350 | 0.2360 .349 | 0.2390 .343 | 0.2460 .340 | 0.2440 .341 | 0.2370 .350 |
| 7 | 0.1720 .309 | 0.1850 .290 | 0.1830 .288 | 0.1830 .288 | 0.1820 .285 | 0.1880 .282 | 0.1860 .283 | 0.1840 .289 |
| Coverage rate (the exact $p$-value) | $\begin{aligned} & 0.9790 \\ & (1.0000) \end{aligned}$ | $\begin{aligned} & 0.9169^{* * * *} \\ & (0.0000) \end{aligned}$ | $\begin{aligned} & 0.2435^{* * * *} \\ & (0.0000) \end{aligned}$ | $\begin{aligned} & 0.2436^{* * * *} \\ & (0.0000) \end{aligned}$ | $\begin{aligned} & 0.9761 \\ & (1.0000) \end{aligned}$ | $\begin{aligned} & 0.9261^{* * * *} \\ & \left(5.475 \times 10^{-25}\right) \end{aligned}$ |  | $\begin{aligned} & 0.9622 \\ & (1.0000) \end{aligned}$ |
| Volume | $1.060 \times 10^{-8}$ | $8.919 \times 10^{-10}$ | $3.182 \times 10^{-10}$ | $3.143 \times 10^{-10}$ | $1.823 \times 10^{-8}$ | $1.034 \times 10^{-8}$ | $1.214 \times 10^{-8}$ | $3.671 \times 10^{-10}$ |

${ }^{*},{ }^{* *},{ }^{* * *},{ }^{* * * *}$ represent significance at $0.05,0.01,0.005,0.001$ significance level, for the problem of testing " $H_{0}$ : coverage probability $\geqslant 0.95^{\prime \prime}$, respectively.
Table 4
Comparison of eight $95 \%$ simultaneous confidence interval Procedures for the data set with $n=3524, N_{1}=\cdots=N_{41}=1, N_{42}=\cdots=N_{70}=2, N_{71}=\cdots=N_{87}=3$, $N_{88}=\cdots=N_{103}=4, N_{104}=\cdots=N_{110}=5, N_{111}=\cdots=N_{120}=6, N_{121}=\cdots=N_{124}=7, N_{125}=\cdots=N_{131}=8, N_{132}=\cdots=N_{137}=9, N_{138}=N_{139}=10$, $N_{140}=\cdots=N_{143}=11, N_{144}=12, N_{145}=13, N_{146}=N_{147}=14, N_{148}=N_{149}=15, N_{150}=\cdots=N_{154}=16, N_{155}=\cdots=N_{157}=18, N_{158}=19, N_{159}=\cdots=N_{161}=20$, $N_{162}=21, N_{163}=N_{164}=23, N_{165}=24, N_{166}=\cdots=N_{168}=29, N_{169}=30, N_{170}=31, N_{171}=39, N_{172}=42, N_{173}=43, N_{174}=48, N_{175}=\cdots=N_{176}=51$, $N_{177}=55, N_{178}=56, N_{179}=58, N_{180}=74, N_{181}=84, N_{182}=95, N_{183}=286, N_{184}=382, N_{185}=1114$

| $i$ | Quesenberry-Hurst$S_{1}(\mathbf{N})$ | Goodman$S_{2}(\mathbf{N})$ | Bailey |  | Fitzpatrick-Scott$S_{5}(\mathbf{N})$ | Sison-Glaz |  | $S_{\text {PD }}(\mathbf{N}, 0.01)$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | $S_{3}(\mathbf{N})$ | $S_{4}(\mathbf{N})$ |  | $S_{6}(\mathbf{N})$ | $S_{7}(\mathbf{N})$ |  |
| 1,2, $\ldots, 41$ | 0.0000 .059 | 0.0000 .004 | 0.0000 .003 | 0.0000 .003 | 0.0000 .019 | 0.0000 .016 | 0.0000 .016 | 0.0000 .003 |
| 42, .., 70 | 0.0000 .059 | 0.0000 .005 | 0.0000 .003 | 0.0000 .003 | 0.0000 .019 | 0.0000 .016 | 0.0000 .016 | 0.0000 .004 |
| 71, ..., 87 | 0.0000 .060 | 0.0000 .005 | 0.0000 .004 | 0.0000 .004 | 0.0000 .020 | 0.0000 .016 | 0.0000 .016 | 0.0000 .004 |
| 88, .., 103 | 0.0000 .060 | 0.0000 .006 | 0.0000 .004 | 0.0000 .004 | 0.0000 .020 | 0.0000 .017 | 0.0000 .017 | 0.0000 .005 |
| 104, .., 110 | 0.0000 .061 | 0.0000 .006 | 0.0000 .005 | 0.0000 .005 | 0.0000 .020 | 0.0000 .017 | 0.0000 .017 | 0.0000 .005 |
| - | - | - | - | - | - | - | - | - |
| $\bullet$ | $\bullet$ | $\bullet$ | - | $\bullet$ | - | $\bullet$ | - | $\bullet$ |
| - | $\bullet$ | $\bullet$ | $\bullet$ | $\bullet$ | - | $\bullet$ | - | $\bullet$ |
| $\bullet$ | $\bullet$ | - | - | - | - | - | - |  |
| - | - |  |  |  |  |  |  |  |
| 183 | 0.0350 .176 | 0.0660 .100 | 0.0650 .099 | 0.0650 .099 | 0.0620 .100 | 0.0660 .097 | 0.0660 .097 | 0.0660 .099 |
| 184 | 0.0530 .209 | 0.0910 .129 | 0.0900 .128 | 0.0900 .128 | 0.0900 .127 | 0.0930 .124 | 0.0930 .124 | 0.0900 .129 |
| 185 | 0.2140 .439 | 0.2880 .345 | 0.2880 .345 | 0.2880 .345 | 0.2970 .335 | 0.3010 .332 | 0.3010 .332 | 0.2880 .345 |
| Coverage rate (the exact $p$-value) | $\begin{aligned} & 1.0000 \\ & (1.0000) \end{aligned}$ | $\begin{aligned} & 0.7691^{* * * *} \\ & (0.0000) \end{aligned}$ | $\begin{aligned} & 0.0000^{* * * *} \\ & (0.0000) \end{aligned}$ | $\begin{aligned} & 0.0000^{* * * *} \\ & (0.0000) \end{aligned}$ | $\begin{aligned} & 0.9838 \\ & (1.0000) \end{aligned}$ | $\begin{aligned} & 0.9464 \\ & \left(5.289 \times 10^{-2}\right) \end{aligned}$ |  | $\begin{aligned} & 0.9697 \\ & (1.0000) \end{aligned}$ |
| Volume | $1.319 \times 10^{-222}$ | $3.331 \times 10^{-407}$ | $3.062 \times 10^{-430}$ | $2.730 \times 10^{-430}$ | $1.802 \times 10^{-309}$ | $1.554 \times 10^{-323}$ | $9.881 \times 10^{-324}$ | $5.302 \times 10^{-425}$ |

${ }^{*},{ }^{* *},{ }^{* * *},{ }^{* * * *}$ represent significance at $0.05,0.01,0.005,0.001$ significance level, for the problem of testing " $H_{0}$ : coverage probability $\geqslant 0.95$ ", respectively.
achieving the specified coverage probability is generally regarded as the primary concern for good simultaneous confidence intervals. We can see from Tables $1-4$ that the intervals based Goodman (1965) and Bailey (1980) have poor coverage probabilities. Hence, although in some cases $S_{3}(\mathbf{N})$ and $S_{4}(\mathbf{N})$ based on Bailey (1980) seem to have the smallest volumes, their optimality is superficial.

## 4. Conclusions and future work

This new procedure proposed in this article is derived using the power-divergence statistics proposed by Cressie and Read (1984), and Monte-Carlo technique is applied to find the optimal solution. Based on the numerical results, the new procedure apparently outperforms all the other alternatives in terms of accuracy, where the accuracy is measured by the volume of the confidence region corresponding to the nominal coverage probability and the probability of coverage it achieves. The result derived here can be extended for several multinomial distributions.

Some might suspect that procedure $S_{\mathbf{P D}}\left(\mathbf{N}, \lambda^{*}\right)$ could be fairly time consuming. However, for a fixed $\lambda$, it took only 0.018 and 0.404 s CPU time to solve the confidence bounds for data sets considered in Tables 1 and 4, respectively. Consider a worse scenario in which we choose 30 grid points among $\{\lambda \mid \lambda \in(-1,2]\}$ and use 50 Monte-Carlo samples to search for the optimal value of $\lambda$, it will only take about 0.45 and 10.1 min CPU time to find the best power-divergence SCIs for data sets given in Tables 1 and 4, respectively. From numerical results presented in Section 4 it is obvious that procedure $S_{\mathbf{P D}}\left(\mathbf{N}, \lambda^{*}\right)$ is more accurate than all the others. This accuracy should suffice for most applications, although we have to admit that the costs in some situations may be still expensive. Therefore, a quicker numerical method in searching for the optimal value $\lambda^{*}$ is called for.

The chi-square approximation for the power-divergence family of statistics $I(\mathbf{P}, \lambda)$ is satisfactory as the sample size $n$ approaches infinity. For finite (or small) $n$, the exact distribution of $I(\mathbf{P}, \lambda)$ is not known. How to determine the appropriate critical value is still an open problem. However, Cressie and Read (1988, p. 95) indicated that we can view the parameter $\lambda$ as a transformation index on the individual cell frequencies. Note that the exact distribution of $I(\mathbf{P}, \lambda)$ varies as $\lambda$ varies from $-\infty$ to $\infty$. Hence, for any given positive value $c$, it is possible to simulate a value $\lambda^{*}$ such that the $100 \times(1-\alpha)$ th percentile of the distribution of $I\left(\mathbf{P}, \lambda^{*}\right)$, say $I_{\lambda^{*}}^{1-\alpha}$, is equal or close to $c$ (i.e., it is possible to find a $\lambda^{*}$ such that $P\left[I\left(\mathbf{P}, \lambda^{*}\right) \leqslant c\right] \approx 1-\alpha$, for any given positive value c). On the other hand, for any fixed $\lambda$, it is also possible to approximate the exact $100 \times(1-\alpha)$ percentile of the distribution of $I(\mathbf{P}, \lambda)$ by using Monte-Carlo simulation. Thus, for any fixed $n, k, \alpha, \lambda$ and $c(c>0)$, instead of using (3) to construct a set of SCIs, we can use the inequality $I(\mathbf{P}, \lambda) \leqslant c$ to play the key role.

By the same arguments as in Section 2.1 the minimum and maximum values can be found for each $P_{i}$ and a set of simultaneous confidence intervals can be stated as follows:

$$
S(\mathbf{N}, \lambda, c)=\left\{\mathbf{P} \mid P_{i} \in\left[P_{i \lambda c L}, P_{i \lambda c u}\right], i=1,2, \ldots, k\right\}, \quad-\infty<\lambda<\infty, c>0,
$$

where $P_{i \lambda c L}$ and $P_{i \lambda c U}$ are the minimum and maximum solutions of the following equations:

$$
\left(1-\hat{P}_{i}\right)^{\lambda+1} P_{i}^{\lambda}+\hat{P}_{i}^{\lambda+1}\left(1-P_{i}\right)^{\lambda}=C_{\lambda} P_{i}^{\lambda}\left(1-P_{i}\right)^{\lambda}, \quad i=1,2, \ldots, k,
$$

where

$$
C_{\lambda}=\frac{\lambda(\lambda+1)}{2 n} c+1 .
$$

Hence, for any given positive value $c$, we can have a set of simultaneous confidence intervals:

$$
F(c)=\{S(\mathbf{N}, \lambda, c), \lambda \in R\}
$$

and therefore a class of family of simultaneous confidence intervals:

$$
U=\{F(c), c>0\} .
$$

Obviously, $F_{1}=F\left(\chi_{\alpha}^{2}(k-1)\right) \in U$, and $F_{2}=F\left(\chi_{\alpha / K}^{2}(1)\right)$, where $F_{1}$ and $F_{2}$ are defined as in Eqs. (6) and (7).

Let $\Omega$ be defined as
$\Omega=\left\{(\lambda, c) \mid c \geqslant I_{\lambda}^{1-\alpha}\right.$, where $I_{\lambda}^{1-\alpha}$ is the exact $100 \times(1-\alpha)$ th percentile of the distribution of $I(\mathbf{P}, \lambda)\}$.
Note that for any significance level $\alpha \in(0,1)$,

$$
I_{\lambda}^{1-\alpha} \leqslant c, \quad \forall(\lambda, c) \in \Omega .
$$

We thus have

$$
\left\{I(\mathbf{P}, \lambda) \leqslant I_{\lambda}^{1-\alpha}\right\} \subset\{I(\mathbf{P}, \lambda) \leqslant c\}, \quad \forall(\lambda, c) \in \Omega
$$

and

$$
\begin{aligned}
P_{r}[I(\mathbf{P}, \lambda) \leqslant c] & \geqslant P_{r}\left[I(\mathbf{P}, \lambda) \leqslant I_{\lambda}^{1-\alpha}\right] \\
& =1-\alpha, \quad \forall(\lambda, c) \in \Omega
\end{aligned}
$$

Hence, every member belonging to the following class:

$$
\{S(\mathbf{N}, \lambda, c),(\lambda, c) \in \Omega\}
$$

is conservative (i.e., their coverage probabilities are not less than $1-\alpha$ ).
From our preliminary simulation results it appears that the smaller the positive value $c$ is, the smaller volume the power-divergence SCIs has. This result needs to be examined further. Nevertheless, it seems positive that we can write a computer program to search for the optimal pair of values $(\lambda, c)$ based on the Monte-Carlo simulated multinomial samples to improve the accuracy of the power-divergence SCIs.

## Acknowledgements

This research was supported by grant NSC 90-2118-M-030-001 (CDH) and NSC 90-2118-M-002-009 (JJT) from the National Science Council, Taiwan, ROC.

## Appendix A. Derivation for Eq. (5)

Note that

$$
\begin{align*}
I(\mathbf{P}, \lambda) & =\frac{2 n}{\lambda(\lambda+1)} \sum_{j=1}^{k} \hat{P}_{j}\left[\left(\frac{\hat{P}_{j}}{P_{j}}\right)^{\lambda}-1\right] \\
& =\frac{2 n}{\lambda(\lambda+1)}\left(\sum_{j=1}^{k} \frac{\hat{P}_{j}^{\lambda+1}}{P_{j}^{\lambda}}-1\right) . \tag{A.1}
\end{align*}
$$

Since

$$
I(\mathbf{P}, \lambda) \leqslant \chi_{\alpha}^{2}(k-1),
$$

we have

$$
\frac{2 n}{\lambda(\lambda+1)}\left(\sum_{j=1}^{k} \frac{\hat{P}_{j}^{\lambda+1}}{P_{j}^{\lambda}}-1\right) \leqslant \chi_{\alpha}^{2}(k-1) .
$$

It follows that

$$
\begin{equation*}
\sum_{j=1}^{k} \frac{\hat{P}_{j}^{\lambda+1}}{P_{j}^{\lambda}} \leqslant \frac{\lambda(\lambda+1)}{2 n} \chi_{\alpha}^{2}(k-1)+1=A_{\lambda} . \tag{A.2}
\end{equation*}
$$

For fixed $n, k, \alpha$, and $\lambda$, the inequality (A.2) defines a region in the parameter space, and if the equal sign is taken, i.e.,

$$
\begin{equation*}
\sum_{j=1}^{k} \frac{\hat{P}_{j}^{\lambda+1}}{P_{j}^{\lambda}}=A_{\lambda}, \tag{A.3}
\end{equation*}
$$

the equation gives the bounding surface of the region. The intersection of this surface with the hyperplane represented by Eq. (4) is a ( $k-1$ )-dimensional hypercurve. Maximum and minimum values can now be found for each $P_{i}(i=1,2, \ldots, k)$ on this curve.

From Eq. (A.3), we have

$$
\begin{equation*}
P_{i}=\left(\frac{\hat{P}_{i}^{\lambda+1}}{A_{\lambda}-\sum_{j \neq i}\left(\hat{P}_{j}^{\lambda+1} / P_{j}^{\lambda}\right)}\right)^{\frac{1}{\lambda}} \quad i=1,2, \ldots, k \tag{A.4}
\end{equation*}
$$

Thus the function to be maximized and minimized is

$$
Q=P_{i}-t\left(\sum_{j=1}^{k} P_{j}-1\right)
$$

$$
\begin{aligned}
& =(1-t) P_{i}-t\left(\sum_{j \neq i} P_{j}-1\right) \\
& =(1-t)\left(\frac{\hat{P}_{i}^{\lambda+1}}{A_{\lambda}-\sum_{j \neq i}\left(\hat{P}_{j}^{\lambda+1} / P_{j}^{\lambda}\right)}\right)^{\frac{1}{\lambda}}-t\left(\sum_{j \neq i} P_{j}-1\right)
\end{aligned}
$$

where $t$ is the Lagrange multiplier. Differentiating $Q$ with respect to $P_{m}$ and $P_{l}(m \neq i$, $l \neq i, m \neq l)$ and setting these expressions equal to zero gives

$$
\begin{aligned}
\frac{t \hat{P}_{i}^{\lambda+1}}{(t-1) P_{i}^{\lambda+1}} & =\frac{\hat{P}_{m}^{\lambda+1}}{P_{m}^{\lambda+1}}, \quad i, m, l=1,2, \ldots, k \\
\frac{t \hat{P}_{i}^{\lambda+1}}{(t-1) P_{i}^{\lambda+1}} & =\frac{\hat{P}_{l}^{\lambda+1}}{P_{l}^{\lambda+1}}, \quad i \neq m, i \neq l, m \neq l
\end{aligned}
$$

Therefore,

$$
\frac{\hat{P}_{m}^{\lambda+1}}{P_{m}^{\lambda+1}}=\frac{\hat{P}_{l}^{\lambda+1}}{P_{l}^{\lambda+1}} \quad \forall m \neq l ; m, l=1,2, \ldots, k(m \neq i, l \neq i)
$$

and

$$
\begin{equation*}
P_{m}=P_{l} \frac{\hat{P}_{m}}{\hat{P}_{l}} \quad \forall m \neq l ; m, l=1,2, \ldots, k(m \neq i, l \neq i) \tag{A.5}
\end{equation*}
$$

follows. All we have to do then is to solve these $k-2$ equations in (A.5) with constraints (4) and (A.3) simultaneously. According to Eq. (A.3),

$$
\sum_{m=1}^{k} \frac{\hat{P}_{m}^{\lambda+1}}{P_{m}^{\lambda}}=A_{\lambda}
$$

Note that the left-hand side can be further expressed as follows:

$$
\begin{aligned}
\sum_{m=1}^{k} \frac{\hat{P}_{m}^{\lambda+1}}{P_{m}^{\lambda}} & =\sum_{\substack{m \neq i \\
m \neq l}} \frac{\hat{P}_{m}^{\lambda+1}}{P_{m}^{\lambda}}+\frac{\hat{P}_{i}^{\lambda+1}}{P_{i}^{\lambda}}+\frac{\hat{P}_{l}^{\lambda+1}}{P_{l}^{\lambda}} \\
& =\sum_{\substack{m \neq i \\
m \neq l}} \frac{\hat{P}_{m}^{\lambda+1}}{\left(P_{l} \hat{P}_{m} / \hat{P}_{l}\right)^{\lambda}}+\frac{\hat{P}_{i}^{\lambda+1}}{P_{i}^{\lambda}}+\frac{\hat{P}_{l}^{\lambda+1}}{P_{l}^{\lambda}} \\
& =\sum_{\substack{m \neq i \\
m \neq l}} \frac{\hat{P}_{m} \hat{P}_{l}^{\lambda}}{P_{l}^{\lambda}}+\frac{\hat{P}_{i}^{\lambda+1}}{P_{i}^{\lambda}}+\frac{\hat{P}_{l}^{\lambda+1}}{P_{l}^{\lambda}}
\end{aligned}
$$

$$
\begin{aligned}
& =\frac{\hat{P}_{l}^{\lambda}}{P_{l}^{\lambda}}\left(\sum_{\substack{m \neq i \\
m \neq l}} \hat{P}_{m}+\hat{P}_{l}\right)+\frac{\hat{P}_{i}^{\lambda+1}}{P_{i}^{\lambda}} \\
& =\frac{\hat{P}_{l}^{\lambda}}{P_{l}^{\lambda}} \sum_{m \neq i} \hat{P}_{m}+\frac{\hat{P}_{i}^{\lambda+1}}{P_{i}^{\lambda}} \\
& =\frac{\hat{P}_{l}^{\lambda}}{P_{l}^{\lambda}}\left(1-\hat{P}_{i}\right)+\frac{\hat{P}_{i}^{\lambda+1}}{P_{i}^{\lambda}}, \quad l, i=1,2, \ldots, k ; l \neq i,
\end{aligned}
$$

where the second equality is derived using (A.5), and the last equality holds since $\sum_{m=1}^{k} \hat{P}_{m}=1$. Hence,

$$
\begin{equation*}
\frac{\hat{P}_{l}^{\lambda}}{P_{l}^{\lambda}}\left(1-\hat{P}_{i}\right)+\frac{\hat{P}_{i}^{\lambda+1}}{P_{i}^{\lambda}}=A_{\lambda}, \quad l, i=1,2, \ldots, k ; l \neq i . \tag{A.6}
\end{equation*}
$$

Note that

$$
\begin{aligned}
1 & =\sum_{\substack{m=1}}^{k} P_{i} \\
& =\sum_{\substack{m \neq i \\
m \neq l}} P_{m}+P_{i}+P_{l} \\
& =\sum_{\substack{m \neq i \\
m \neq l}} P_{l} \frac{\hat{P}_{m}}{\hat{P}_{l}}+P_{i}+P_{l} \\
& =P_{l}\left(\sum_{\substack{m \neq i \\
m \neq l}} \frac{\hat{P}_{m}+\hat{P}_{l}}{\hat{P}_{l}}\right)+P_{i} \\
& =P_{l} \frac{\sum_{m \neq i} \hat{P}_{m}}{\hat{P}_{l}}+P_{i} \\
& =P_{l} \frac{\left(1-\hat{P}_{i}\right)}{\hat{P}_{l}}+P_{i} .
\end{aligned}
$$

Hence

$$
\begin{equation*}
P_{l}=\hat{P}_{l}\left(\frac{1-P_{i}}{1-\hat{P}_{i}}\right) \quad \text { for } l, i=1,2, \ldots, k ; l \neq i . \tag{A.7}
\end{equation*}
$$

Solving Eqs. (A.6) and (A.7) simultaneously, we obtain

$$
\frac{\hat{P}_{l}^{\lambda}\left(1-\hat{P}_{i}\right)}{\left(\hat{P}_{l}\left(\frac{1-P_{i}}{1-P_{i}}\right)\right)^{\lambda}}+\frac{\hat{P}_{i}^{\lambda+1}}{P_{i}^{\lambda}}=A_{\lambda}
$$

and hence

$$
\left(1-\hat{P}_{i}\right)^{\lambda+1} P_{i}^{\lambda}+\hat{P}_{i}^{\lambda+1}\left(1-P_{i}\right)^{\lambda}=A_{\lambda} P_{i}^{\lambda}\left(1-P_{i}\right)^{\lambda} \quad i=1,2, \ldots, k .
$$

## References

Angers, C., 1984. Large sample sizes for the estimation of multinomial frequencies from simulation studies. Simulation, 175-178.
Bailey, B.J.R., 1980. Large sample simultaneous confidence intervals for the multinomial probabilities based on transformations of the cell frequencies. Technometrics 22, 583-589.
Cressie, N., Read, T.R.C., 1984. Multinomial goodness-of-fit tests. J. Roy. Statist. Soc. Ser. B 46, 440-464.
Cressie, N., Read, T.R.C., 1988. Goodness-of-Fit Statistics for Discrete Multivariate Data. Springer, New York.
Fitzpatrick, S., Scott, A., 1987. Quick simultaneous confidence intervals for multinomial proportions. J. Amer. Statist. Assoc. 82, 875-878.
Glaz, J., Johnson, B., 1984. Probability for multivariate distributions with dependence structures. J. Amer. Statist. Assoc. 79, 436-441.
Goodman, L.A., 1965. On simultaneous confidence intervals for multinomial proportions. Technometrics 7, 247-254.
Hou, C.-D., Chiang, J., Tai, J.J., 1999. Testing the nonrandomness of chromosomal breakpoints using highest observed breakages. Hum. Genet. 104, 350-355.
Hou, C.-D., Chiang, J., Tai, J.J., 2001. Identifying chromosomal fragile sites from a hierarchical-clustering point of view. Biometrics 57, 435-440.
Hurtubise, R., 1969. Sample sizes and confidence intervals associated with a Monte-Carlo simulation model possessing a multinomial output. Simulation 12, 71-77.
Levin, B., 1981. A representation for multinomial cumulative distribution functions. Ann. Stat. 9, 1123-1126.
Quesenberry, C.P., Hurst, D.C., 1964. Simultaneous confidence intervals for multinomial proportions. Technometrics 6, 191-195.
Sison, C.P., Glaz, J., 1995. Simultaneous confidence intervals and sample size determination for multinomial proportions. J. Amer. Stat. Assoc. 90, 366-369.


[^0]:    * Corresponding author. Tel.: +886-02-23123456; ext: 8340; fax: +886-02-23511955.

    E-mail addresses: stat0002@mails.fju.edu.tw (C.-D. Hou), chiangj@nccu.edu.tw (J. Chiang), jjtai@ ha.mc.ntu.edu.tw (J.J. Tai).

