CrossMark

Testing homogeneity of proportions from sparse binomial data with a large number of groups

Junyong Park[1]

Abstract In this paper, we consider testing the homogeneity for proportions in independent binomial distributions, especially when data are sparse for large number of groups. We provide broad aspects of our proposed tests such as theoretical studies, simulations and real data application. We present the asymptotic null distributions and asymptotic powers for our proposed tests and compare their performance with existing tests. Our simulation studies show that none of tests dominate the others; however, our proposed test and a few tests are expected to control given sizes and obtain significant powers. We also present a real example regarding safety concerns associated with Avandia (rosiglitazone) in Nissen and Wolski (New Engl J Med 356:2457–2471, 2007).

Keywords Asymptotic distribution · Homogeneity of proportions · Sparse data

1 Introduction

An important step in statistical meta-analysis is to carry out appropriate tests of homogeneity of the relevant *effect sizes* before pooling of evidence or information across studies. While the familiar Cochran (1954) Chi-square goodness-of-fit test is widely used in this context, it turns out that this test may perform poorly in terms of *not* maintaining Type I error rate in many problems. In particular, this is indeed a serious drawback of Cochran's test for testing the homogeneity of several proportions in case of sparse data. A recent meta-analysis (Nissen and Wolski 2007), addressing

✉ Junyong Park
junpark@umbc.edu

[1] Department of Mathematics and Statistics, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA

the cardiovascular safety concerns associated with (rosiglitazone), has received wide attention (Cai et al. 2010; Tian et al. 2009; Shuster et al. 2007; Shuster 2010; Stijnen et al. 2010). Two difficulties seem to appear in this study: first, study sizes (N) are highly unequal, especially in control arm, with over 95% of the studies having sizes below 400 and two studies having sizes over 2500; second, event rate is extremely low, especially for death end point, with the maximum death rate in the treatment arm being 2%, while in control arm, over 80% of the studies have zero events. The original meta-analysis (Nissen and Wolski 2007) was performed under fixed effects framework, as the diagnostic test based on Cochran's Chi-square test failed to reject homogeneity. However, with two large studies dominating the combined result, people agree random effects analysis is the superior choice over fixed effects (Shuster et al. 2007). Moreover, the results for the fixed and random effect analyses are discordant. While different fixed effect and random effect approaches are proposed, the problem of testing for homogeneity of effect sizes is less familiar and often not properly addressed. This is precisely the object of this paper, namely a thorough discussion of tests of homogeneity of proportions in case of sparse data situations. Recently, there are some studies on testing the equality of means when the number of groups increases with fixed sample sizes in either ANOVA (analysis of variance) or MANOVA (multivariate analysis of variance). For example, see Bathke and Harrar (2008), Bathke and Lankowski (2005) and Boos and Brownie (1995). Those studies have limitation in asymptotic results since they assume all samples sizes are equal, i.e., balanced design. On the other hand, we actually emphasize the case that sample sizes are highly unbalanced and present more fluent asymptotic results for a variety cases including unbalanced cases and small values of proportions in binomial distributions.

In this paper, we first point out that the classical Chi-square test may fail in controlling a size when the number of groups is high and data are sparse. We modify the classical Chi-square test with providing asymptotic results. Moreover, we propose two new tests for homogeneity of proportions when there are many groups with sparse count data. Throughout this study, we present some theoretical conditions under which our proposed tests achieve the asymptotic normality, while most of existing tests do not have rigorous investigation of asymptotic properties.

A formulation of the testing problem for proportions is provided in Sect. 2 along with a review of the literature and suggestion for *new* tests. The necessary asymptotic theory to ease the application of the suggested test is developed. Results of simulation studies are reported in Sect. 3, and an application to the Nissen and Wolski (2007) data set is made in Sect. 4. Concluding remark is presented in Sect. 5.

2 Testing the homogeneity of proportions with sparse data

In this section, we present a modification of a classical test which is Cochran's test and also propose two types of new tests. Throughout this paper, our theoretical studies are based on triangular array which is commonly used in asymptotic theories in high dimension. See Park and Ghosh (2007) and Park (2009) for triangular array in binary data and Greenshtein and Ritov (2004) for more general cases. More specifically, let $\Theta^{(k)} = \{(\pi_1^{(1)}, \pi_2^{(2)}, \ldots, \pi_k^{(k)}) : 0 < \pi_i^{(k)} < 1 \text{ for } 1 \leq i \leq k\}$ be the parameter space

in which $\pi_i^{(k)}$s are allowed to be varying depending on k as k increases. Additionally, sample sizes $(n_1^{(1)}, \ldots, n_k^{(k)})$ also change depending on k. However, for notational simplicity, we suppress superscript k from $\pi_i^{(k)}$ and $n_i^{(k)}$. The triangular array provides more flexible situations, for example all increasing sample sizes and all decreasing π_is. On the other hand, the asymptotic results in Bathke and Lankowski (2005) and Boos and Brownie (1995) are based on increasing k, but all sample sizes and π_is are fixed. This set up provides somewhat limited results, while we present the asymptotic results on the triangular array. Our results will include the asymptotic power functions of proposed tests, while existing studies do not provide them.

2.1 Modification of Cochran's test

Suppose that there are k independent populations and the ith population has $X_i \sim$ Binomial(n_i, π_i). Denote the total sample size and the weighted average of π_i's by $N = \sum_{i=1}^{k} n_i$ and $\bar{\pi} = \frac{1}{N} \sum_{i=1}^{k} n_i \pi_i$, respectively. We are interested in testing the homogeneity of π_i's from different groups,

$$H_0 : \pi_1 = \pi_2 = \cdots = \pi_k \equiv \pi (\text{unknown}). \tag{1}$$

To test the above hypothesis in (1), one familiar procedure is Cochran's Chi-square test in Cochran (1954), namely T_S:

$$T_S = \sum_{i=1}^{k} \frac{(X_i - n_i \hat{\bar{\pi}})^2}{n_i \hat{\bar{\pi}}(1 - \hat{\bar{\pi}})} \tag{2}$$

where $\hat{\pi} = \frac{\sum_{i=1}^{k} X_i}{\sum_{i=1}^{k} n_i}$. T_S uses an approximate Chi-square distribution with degrees of freedom $(k - 1)$ under H_0. The H_0 is rejected when $T_S > \chi_{1-\alpha,k-1}^2$ where $\chi_{1-\alpha,k-1}^2$ is the $1 - \alpha$ quantile of Chi-square distribution with degrees of freedom $(k - 1)$. In particular, when k is large, $\frac{T_S - k}{\sqrt{2k}}$ is approximated by a standard normal distribution under H_0. Although Cochran's test for homogeneity is widely used, the approximation to the χ^2 distribution of T_S or normal approximation may be poor when the sample sizes within the groups are small or when some counts in one of the two categories are low. This is partly because the test statistic becomes noticeably discontinuous and partly because its moments beyond the first may be rather different from those of χ^2.

We demonstrate that the asymptotic Chi-square approximation to T_S or normal approximation based on $\frac{T_S - k}{\sqrt{2k}}$ may be very poor when k is large or π_is are small compared to n_is. We provide the following theorem and propose a modified approximation to T_S which is expected to provide more accurate approximation. Let us define

$$T = \frac{T_S - E(T_S)}{\sqrt{\mathcal{B}_k}} \tag{3}$$

where $\mathcal{T}_S = \sum_{i=1}^k \frac{(X_i - n_i \bar{\pi})^2}{n_i \bar{\pi}(1 - \bar{\pi})}$, $\mathcal{B}_k \equiv \mathrm{Var}(\mathcal{T}_S) = \sum_{i=1}^k \mathrm{Var}\left(\frac{(X_i - n_i \bar{\pi})^2}{n_i \bar{\pi}(1 - \bar{\pi})}\right) \equiv \sum_{i=1}^k B_i$
and

$$\mathrm{Var}\left(\frac{X_i - n_i \bar{\pi}}{n_i \bar{\pi}(1 - \bar{\pi})}\right) = B_i = \frac{2\pi_i^2(1 - \pi_i)^2}{\bar{\pi}^2(1 - \bar{\pi})^2} + \frac{\pi_i(1 - \pi_i)(1 - 6\pi_i(1 - \pi_i))}{n_i \bar{\pi}^2(1 - \bar{\pi})^2}$$
$$+ \frac{3\pi_i(1 - \pi_i)(1 - 2\pi_i)(\pi_i - \bar{\pi})}{\bar{\pi}^2(1 - \bar{\pi})^2}$$
$$+ \frac{4n_i \pi_i(1 - \pi_i)(\pi_i - \bar{\pi})^2}{\bar{\pi}^2(1 - \bar{\pi})^2},$$

$$E(\mathcal{T}_S) = \sum_{i=1}^k \left(\frac{n_i(\pi_i - \bar{\pi})^2}{\bar{\pi}(1 - \bar{\pi})} + \frac{\pi_i(1 - \pi_i)}{\bar{\pi}(1 - \bar{\pi})}\right).$$

Note that $\mathcal{T}_S$ is not a statistic since it still includes the unknown parameter $\bar{\pi} = \sum_{i=1}^k \frac{n_i \pi_i}{N}$. It will be shown later that $\bar{\pi}$ can be replaced by $\hat{\bar{\pi}} = \frac{1}{N} \sum_{i=1}^k n_i \hat{\pi}_i$ under H_0 since $\hat{\bar{\pi}}$ has the ratio consistency ($\frac{\hat{\bar{\pi}}}{\bar{\pi}} \to 1$ in probability) under some mild conditions. Define

$$\mathcal{B}_{0k} = \sum_{i=1}^k B_{0i} = \sum_{i=1}^k \left(2 - \frac{6}{n_i} + \frac{1}{n_i \bar{\pi}(1 - \bar{\pi})}\right)$$

and

$$T_0 = \frac{\mathcal{T}_S - k}{\sqrt{\mathcal{B}_{0k}}} \tag{4}$$

which is the T defined in (3) under H_0 since $E(\mathcal{T}_S) = k$ and $\mathcal{B}_k = \mathcal{B}_{0k}$ under H_0. The following theorem shows the asymptotic properties of T_0 in (4).

Theorem 1 *For $\theta_i = \pi_i(1 - \pi_i)$ and $\bar{\theta} = \bar{\pi}(1 - \bar{\pi})$, if $\frac{\sum_{i=1}^k \left(\theta_i^4 + \frac{\theta_i}{n_i}\right)}{(\bar{\pi}(1 - \bar{\pi}))^4 \mathcal{B}_k^2} \to 0$ and $\frac{\sum_{i=1}^k n_i^2 \theta_i(\pi_i - \bar{\pi})^4(\theta_i + \frac{1}{n_i})}{(\bar{\pi}(1 - \bar{\pi}))^4 \mathcal{B}_k^2} \to 0$ as $k \to \infty$, then we have*

$$P(T_0 > z_{1-\alpha}) - \bar{\Phi}\left(\frac{z_{1-\alpha}}{\sigma_k} - \mu_k\right) \to 0$$

where $\mu_k = \frac{E(\mathcal{T}_S) - k}{\sqrt{\mathcal{B}_k}}$, $\sigma_k^2 = \frac{\mathcal{B}_k}{\mathcal{B}_{0k}}$ and $\bar{\Phi}(z) = 1 - \Phi(z) = P(Z \geq z)$ for a standard normal distribution Z.

Proof See "Appendix".   □

We propose to use a test which rejects the H_0 if

$$T_\chi \equiv \frac{\mathcal{T}_S - k}{\sqrt{\hat{\mathcal{B}}_{0k}}} > z_{1-\alpha} \tag{5}$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of a standard normal distribution, $\hat{\mathcal{B}}_{0k} = \sum_{i=1}^{k}\left(2 - \frac{6}{n_i} + \frac{1}{n_i\hat{\bar{\pi}}(1-\hat{\bar{\pi}})}\right)$ and $\hat{\bar{\pi}} = \frac{\sum_{i=1}^{k} n_i\hat{\pi}_i}{N}$.

Using Theorem 1, we obtain the following results which states that our proposed modification of Cochran's test in (5) is the asymptotically size α test, while $\frac{T_S-k}{\sqrt{2k}}$ may fail in controlling a size α under some conditions.

Corollary 1 *Under H_0 and the conditions in Theorem* 1, *T_χ in (5) is asymptotically size α test. A normal approximation to $\frac{T_S-k}{\sqrt{2k}}$ is not asymptotically size α test unless $\frac{\mathcal{B}_{0k}}{2k} \to 1$.*

Proof We first show that $\hat{\bar{\pi}}/\bar{\pi} \to 1$ in probability. Under H_0, $\pi_i \equiv \pi$, we have $\sum_{i=1}^{k} n_i\hat{\pi}_i \sim Binomial(N, \pi)$. Using $\sum_{i=1}^{k} n_i\pi_i = N\pi \to \infty$ under H_0, we have

$$E\left(\frac{\hat{\bar{\pi}}}{\bar{\pi}} - 1\right)^2 = \frac{1-\pi}{N\pi} \le \frac{1}{N\pi} \to 0$$

leading to $\hat{\bar{\pi}}/\bar{\pi} \to 1$ in probability. From this, we have $\frac{\hat{\mathcal{B}}_{0k}}{\mathcal{B}_{0k}} \to 1$ in probability under H_0. Furthermore, under H_0, since we have $\frac{\mathcal{B}_{0k}}{\mathcal{B}_k} = 1$ and $E(T_S) = k$, we obtain $T_\chi - T = (\sqrt{\frac{\mathcal{B}_{0k}}{\hat{\mathcal{B}}_{0k}}} - 1)T = o_p(1)O_p(1) = o_p(1)$ which means T_χ and T are asymptotically equivalent under the H_0. Since $P_{H_0}(T > z_{1-\alpha}) - \bar{\Phi}(z_{1-\alpha}) \to 0$, we have $P_{H_0}(T_\chi > z_{1-\alpha}) - \alpha \to 0$ which means T_χ is the asymptotically size α test. On the other hand, it is obvious that $\frac{T_S-k}{\sqrt{2k}}$ does not have an asymptotic standard normality unless $\mathcal{B}_{0k}/(2k) \to 1$ since $\frac{T_S-k}{\sqrt{2k}} = \sqrt{\frac{\hat{\mathcal{B}}_{0k}}{2k}}T_\chi$ under the H_0. □

Under H_0, since $\mathcal{B}_{0k} = 2k + (\frac{1}{\pi(1-\pi)} - 6)\sum_{i=1}^{k}\frac{1}{n_i}$, we expect $\frac{\mathcal{B}_{0k}}{2k}$ to converge to 1 when $(\frac{1}{\pi(1-\pi)} - 6)\sum_{i=1}^{k}\frac{1}{n_i} = o(k)$ where $\pi_i = \bar{\pi} \equiv \pi$ under H_0. This may happen when π is bounded away from 0 and 1 and n_is are large. If all n_is are bounded by some constant, say C, and $|\frac{1}{\pi(1-\pi)} - 6| \ge \delta > 0$ (this can happen when $\pi < \epsilon_1$ or $\pi > 1 - \epsilon_2$ for some $\epsilon_1 > 0$ and $\epsilon_2 > 0$), then $\frac{\mathcal{B}_k}{2k}$ does not converge to 1. Even for n_is are large, if $\pi \to 0$ fast enough, then $\frac{\mathcal{B}_{0k}}{2k}$ does not converge to 1. For example, if $\pi = 1/k$ and $n_i = k$ as $k \to \infty$, then $\frac{\mathcal{B}_{0k}}{2k} \to 3/2$ which leads to $\frac{T_S-k}{\sqrt{2k}} \to N(0, \frac{3}{2})$ in distribution. This implies that $P(\frac{T_S-k}{\sqrt{2k}} > z_{1-\alpha}) \to 1 - \Phi(\sqrt{\frac{2}{3}}z_{1-\alpha}) > \alpha$, so the test obtains a larger asymptotic size than a given nominal level. To summarize, if either π is small or n_is are small, we may not expect an accurate approximation to $\frac{T_S-k}{\sqrt{2k}}$ based on normal approximation, so the sparse binary data with small n_is and a large number of groups (k) need to be handled more carefully.

2.2 New tests

In addition to the modified Cochran's test T_χ, we also propose new tests designed for sparse data when k is large. Similar to the asymptotic normality of T_χ, it will be justified that our proposed tests have the asymptotic normality when $k \to \infty$ although n_is are not required to increase. Toward this end, we proceed as follows. Let $||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}}^2 = \sum_{i=1}^{k} n_i (\pi_i - \bar{\pi})^2$ which is weighted l_2 distance from $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ to $\bar{\boldsymbol{\pi}} = (\bar{\pi}, \bar{\pi}, \ldots, \bar{\pi})$ where $\mathbf{n} = (n_1, \ldots, n_k)$. The proposed test is based on measuring the $||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}}^2$. Since this is unknown, one needs to estimate the $||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}}^2$. One typical estimator is a plug-in estimator such as $||\hat{\boldsymbol{\pi}} - \hat{\bar{\boldsymbol{\pi}}}||_{\mathbf{n}}$; however, this estimator may have a significant bias. To illustrate this, note that

$$
E||\hat{\boldsymbol{\pi}} - \hat{\bar{\boldsymbol{\pi}}}||_{\mathbf{n}}^2 = \sum_{i=1}^{k} \pi_i (1 - \pi_i) + \sum_{i=1}^{k} \frac{n_i \pi_i (1 - \pi_i)}{N} - \frac{2}{N} \sum_{i=1}^{k} n_i \pi_i (1 - \pi_i)
$$

$$
+ \sum_{i=1}^{k} n_i (\pi_i - \bar{\pi})^2
$$

$$
= \sum_{i=1}^{k} c_i \pi_i (1 - \pi_i) + ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}}^2
$$

where $c_i = (1 - \frac{n_i}{N})$. This shows that $||\hat{\boldsymbol{\pi}} - \hat{\bar{\boldsymbol{\pi}}}||_{\mathbf{n}}^2$ is an overestimate of $||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}}^2$ by $\sum_{i=1}^{k} c_i \pi_i (1 - \pi_i)$ which needs to be corrected. Using $E\left[\frac{n_i}{n_i - 1} \hat{\pi}_i (1 - \hat{\pi}_i)\right] = \pi_i (1 - \pi_i)$ for $\hat{\pi}_i = \frac{x_i}{n_i}$, we define $d_i = \frac{n_i c_i}{n_i - 1}$ and

$$
T = \sum_{i=1}^{n} n_i (\hat{\pi}_i - \hat{\bar{\pi}})^2 - \sum_{i=1}^{k} d_i \hat{\pi}_i (1 - \hat{\pi}_i) \equiv ||\hat{\boldsymbol{\pi}} - \hat{\bar{\boldsymbol{\pi}}}||_{\mathbf{n}}^2 - \sum_{i=1}^{k} d_i \hat{\pi}_i (1 - \hat{\pi}_i) \quad (6)
$$

which is an unbiased estimator of $||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}}^2$. This implies $E(T) = ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}}^2 \geq 0$ and "=" holds only when H_0 is true. Therefore, it is natural to consider large values of T as an evidence supporting H_1, and we thus propose a one-sided (upper) rejection region based on T for testing H_0. Our proposed test statistics are based on T of which the asymptotic distribution is normal distribution under some conditions.

We derive the asymptotic normality of a standardized version of T under some regularity conditions. Let us decompose T into two components, say T_1 and T_2:

$$
T = \sum_{i=1}^{k} n_i (\hat{\pi}_i - \pi_i + \pi_i - \bar{\pi} + \bar{\pi} - \hat{\bar{\pi}})^2 - \sum_{i=1}^{k} d_i \hat{\pi}_i (1 - \hat{\pi}_i)
$$

$$
= \underbrace{\sum_{i=1}^{k} \left\{ n_i (\hat{\pi}_i - \pi_i)^2 - d_i \hat{\pi}_i (1 - \hat{\pi}_i) + 2 n_i (\hat{\pi}_i - \pi_i)(\pi_i - \bar{\pi}) + n_i (\pi_i - \bar{\pi})^2 \right\}}_{T_1}
$$

$$
(7)
$$

$$-\underbrace{N(\hat{\bar{\pi}} - \bar{\pi})^2}_{T2} \tag{8}$$

where $T_1 \equiv \sum_{i=1}^k T_{1i}$ for $T_{1i} = n_i(\hat{\pi}_i - \pi_i)^2 - d_i\hat{\pi}_i(1 - \hat{\pi}_i) + 2n_i(\hat{\pi}_i - \pi_i)(\pi_i - \bar{\pi}) + n_i(\pi_i - \bar{\pi})^2$. To prove the asymptotic normality of the proposed test, we need some preliminary results stated below in Lemmas 1, 2 and 3 and show the ratio consistency of proposed estimators of Var(T_1) in Lemma 5.

Lemma 1 *Let $\theta_i = \pi_i(1 - \pi_i)$. When $X_i \sim Binomial(n_i, \pi_i)$ and $\hat{\pi}_i = \frac{X_i}{n_i}$, we have*

$$E[(\hat{\pi}_i - \pi_i)^3] = \frac{(1 - 2\pi_i)\theta_i}{n_i^2}, \quad E[(\hat{\pi}_i - \pi_i)^4] = \frac{3\theta_i^2}{n_i^2} + \frac{(1 - 6\theta_i)\theta_i}{n_i^3}$$

$$E[\hat{\pi}_i(1 - \hat{\pi}_i)] = \frac{n_i - 1}{n_i}\theta_i,$$

$$\pi_i^l = E\left[\frac{n_i^l}{\prod_{j=0}^{l-1}(n_i - j)} \prod_{j=0}^{l-1}\left(\hat{\pi}_i - \frac{j}{n_i}\right)\right], \quad for\ n_i \geq l\ and\ l = 1, 2, 3, 4.$$

Proof The first three results are easily derived by some computations. For the last result, note that when $X_i \sim$ Binomial(n_i, π_i), $E[X_i(X_i - 1) \cdots (X_i - l + 1)] = n_i(n_i - 1) \cdots (n_i - l + 1)\pi_i^l$. Let $X = \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij} \sim$ Binomial(N, π), then we have the above unbiased estimators under H_0 using $\hat{\pi} = \frac{X}{N} = \frac{1}{N}\sum_{i=1}^k n_i\hat{\pi}_i$. □

We now derive the asymptotic null distribution of $\frac{T_1}{\sqrt{\text{Var}(T_1)}}$ and propose an unbiased estimator of Var(T_1) which has the ratio consistency property. We first compute Var(T_1) and then propose an estimator $\widehat{\text{Var}(T_1)}$.

Lemma 2 *The variance of T_1, Var(T_1), is*

$$Var(T_1) = \sum_{i=1}^k \mathcal{A}_{1i}\theta_i^2 + \sum_{i=1}^k \mathcal{A}_{2i}\theta_i$$

$$+ 4\sum_{i=1}^k n_i(\pi_i - \bar{\pi})^2\theta_i + \frac{4}{N}\sum_{i=1}^k n_i(\pi_i - \bar{\pi})(1 - 2\pi_i)\theta_i \tag{9}$$

where $\mathcal{A}_{1i} = \left(2 - \frac{6}{n_i} - \frac{d_i^2}{n_i} + \frac{8d_i^2}{n_i^2} - \frac{6d_i^2}{n_i^3} + 12d_i\frac{n_i-1}{n_i^2}\right)$ and $\mathcal{A}_{2i} = \frac{n_i}{N^2}$ for $d_i = \frac{n_i}{n_i-1}\left(1 - \frac{n_i}{N}\right)$.

Proof See "Appendix".   □

Under the H_0 ($\pi_i = \pi$ for all $1 \leq i \leq k$), the third and fourth terms including $\pi_i - \bar{\pi}$ in (9) are 0, and therefore, we obtain the Var(T_1) under H_0 as follows;

$$\mathrm{Var}_{H_0}(T_1) \equiv \mathcal{V}_1 = \sum_{i=1}^{k} \left\{ \mathcal{A}_{1i}\theta_i^2 + \mathcal{A}_{2i}\theta_i \right\} \tag{10}$$

$$= \mathcal{V}_{1*} = (\pi(1-\pi))^2 \sum_{i=1}^{k} \mathcal{A}_{1i} + \pi(1-\pi) \sum_{i=1}^{k} \mathcal{A}_{2i}. \tag{11}$$

$\mathcal{V}_1$ in (10) and $\mathcal{V}_{1*}$ in (11) are equivalent under the H_0; however, the estimators may be different depending on whether θ_is are estimated individually from x_i or the common value π is estimated in $\mathcal{V}_{1*}$ by the pooled estimator $\hat{\pi}$. We shall consider these two approaches for estimating $\mathcal{V}_1$ and $\mathcal{V}_{1*}$.

First, we demonstrate the estimator for $\mathcal{V}_1$ in (10). $\mathcal{V}_{1i} \equiv \mathcal{A}_{1i}\theta_i^2 + \mathcal{A}_{2i}\theta_i$ is a fourth degree polynomial in π_i, in other words, $\mathcal{V}_{1i} = a_{1i}\pi_i + a_{2i}\pi_i^2 + a_{3i}\pi_i^3 + a_{14}\pi_i^4$ where a_{ij}'s depend only on N and n_i. As an estimator of $\mathcal{V}_1 = \sum_{i=1}^{k}(a_{1i}\pi_i + a_{2i}\pi_i^2 + a_{3i}\pi_i^3 + a_{4i}\pi_i^4)$, we consider unbiased estimators of π_i, π_i^2, π_i^3 and π_i^4. Let $\eta_{li} = \pi_i^l$, $l = 1, 2, 3, 4$, then unbiased estimators of η_{li}, say $\hat{\eta}_{li}$, are obtained directly from Lemma 1, leading to the first estimator of $\mathcal{V}_1$, as

$$\hat{\mathcal{V}}_1 = \sum_{i=1}^{k} \sum_{l=1}^{4} a_{li}\hat{\eta}_{li} \tag{12}$$

where $\hat{\eta}_{li} = \frac{n_i^l}{\prod_{j=1}^{l-1}(n_i-j)} \prod_{j=0}^{l-1} \left(\hat{\pi}_i - \frac{j}{n_i} \right)$ for $l = 1, 2, 3, 4$ from Lemma 1 and

$$a_{1i} = \mathcal{A}_{2i}, \quad a_{2i} = \mathcal{A}_{1i} - \mathcal{A}_{2i}, \quad a_{3i} = -2\mathcal{A}_{1i}, \quad a_{4i} = \mathcal{A}_{1i}.$$

The second estimator is based on estimating $\mathcal{V}_{1*}$ in (11). Since all $\pi_i = \pi$ under H_0, we can write $\mathcal{V}_{1*} = \sum_{i=1}^{k} \sum_{l=1}^{4} a_{li}\pi_i^l = \sum_{i=1}^{k} \sum_{l=1}^{4} a_{li}\pi^l$ and use an unbiased estimator of π^l using $\sum_{i=1}^{k} x_i \sim Binomial(N, \pi)$ from Lemma 1. This leads to the estimator of $\mathcal{V}_{1*}$ under H_0 which is

$$\hat{\mathcal{V}}_{1*} = \sum_{i=1}^{k} \sum_{l=1}^{4} a_{li}\hat{\eta}_l. \tag{13}$$

where $\hat{\eta}_l = \frac{N^l}{\prod_{j=0}^{l-1}(N-j)} \prod_{j=0}^{l-1} \left(\hat{\pi} - \frac{j}{N} \right)$ and $\hat{\pi} = \frac{1}{N} \sum_{i=1}^{k} n_i\hat{\pi}_i$, as used earlier.

Remark 1 Note that $\hat{\mathcal{V}}_1$ is an unbiased estimator of $\mathcal{V}_1$ regardless of H_0 and H_1. On the other hand, $\hat{\mathcal{V}}_{1*}$ is an unbiased estimator of $\mathcal{V}_{1*}$ only under the H_0 since we use the binomial distribution of the pooled data $\sum_{i=1}^{k} x_i$ and use Lemma 1.

For sequences of $a_n(> 0)$ and $b_n(> 0)$, let us define $a_n \asymp b_n$ if $0 < \liminf \frac{a_n}{b_n} \leq \limsup \frac{a_n}{b_n} < \infty$. The following lemmas will be used in the asymptotic normality of the proposed test.

Lemma 3 *Suppose $n_i \geq 2$ for $1 \leq i \leq k$. Then,*

1. *we have $\mathcal{V}_1 \asymp \sum_{i=1}^{k} \theta_i^2 + \frac{1}{N^2} \sum_{i=1}^{k} n_i \theta_i$. In particular, if $0 < c \leq \pi_i \leq 1 - c < 1$ for all i and some constant c, we have $\mathcal{V}_1 \asymp k$.*
2. *we have*

$$\sum_{i=1}^{k} \mathcal{A}_{1i} \theta_i^2 \leq Var(T_1) \leq K \left(\mathcal{V}_1 + ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}\theta}^2 \right) \tag{14}$$

for some constant $K > 0$ where $||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}\theta}^2 = \sum_{i=1}^{k} n_i (\pi_i - \bar{\pi})^2 \theta_i$. If $|\pi_i - \bar{\pi}| \geq \frac{1+\epsilon}{N}$ for some $\epsilon > 0$ and $1 \leq i \leq k$, we have

$$Var(T_1) \asymp \mathcal{V}_1 + ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}\theta}^2. \tag{15}$$

Proof See "Appendix".   □

We provide another lemma which plays a crucial role in the proof of the main result. As mentioned, we have two types of variances such as $\mathcal{V}_1$ in (10) and $\mathcal{V}_{1*}$ in (11) and their estimators $\widehat{\mathcal{V}}_1$ and $\widehat{\mathcal{V}}_{1*}$. For T_1 in (8), we consider two types of standard deviations based on $Var(T_1)$ and $Var(T_1)_*$.

The following lemma provides upper bounds of $n^4 E(\hat{\pi} - \pi)^8$ and $E(\hat{\pi}(1 - \hat{\pi}))^4$ which are needed in our proof for our mail results.

Lemma 4 *If $X \sim Binomial(n, \pi)$, $\hat{\pi} = \frac{X}{n}$ and $\hat{\eta}_l$ is the unbiased estimator of π^l defined in Lemma 1, then we have, for $\theta \equiv \pi(1 - \pi)$,*

$$n^4 E(\hat{\pi} - \pi)^8 \leq C \min \left\{ \theta^4, \frac{\theta}{n} \right\}$$

$$E(\hat{\pi}(1 - \hat{\pi}))^4 \leq C' \min \left\{ \theta^4, \frac{\theta}{n^3} \right\}$$

$$E\hat{\pi}^l = \pi^l + O \left(\frac{\pi}{n^{l-1}} + \frac{\pi^{l-1}}{n} \right) \quad for \ l \geq 2 \tag{16}$$

$$E(\hat{\pi}^l - \pi^l)^2 = O \left(\frac{\pi^{2l-1}}{n} + \frac{\pi}{n^{2l-1}} \right) \quad for \ l \geq 2$$

$$E(\hat{\eta}_l - \pi^l)^2 = O \left(\frac{\pi^{2l-1}}{n} + \frac{\pi}{n^{2l-1}} \right) \quad for \ l \geq 2 \tag{17}$$

where C and C' are universal constants which do not depend on π and n.

Proof See "Appendix".   □

Remark 2 It should be noted that the bounds in Lemma 4 depend on the behavior of $\theta = \pi(1 - \pi)$ and the sample size n in binomial distribution. In the classical asymptotic theory for a fixed value of π, if π is bounded away from 0 and 1 and n is large, then θ^4 dominates $\frac{\theta}{n}$ (or $\frac{\theta}{n^3}$). However, n is not large and π is close to 0 or 1, then $\frac{\theta}{n}$ (or $\frac{\theta}{n^3}$) is a tighter bound of $n^4 E(\hat{\pi} - \pi)^8$ (or $E(\hat{\pi}(1 - \hat{\pi}))^4$) than θ^4.

The following lemma shows that $\hat{\mathcal{V}}_1$ and $\hat{\mathcal{V}}_{1*}$ have the ratio consistency under some conditions.

Lemma 5 *For* $\tilde{\theta} = \bar{\pi}(1 - \bar{\pi}), \bar{\pi} = \frac{1}{N} \sum_{i=1}^{k} n_i \pi_i$ *and* $\pi_i \leq \delta < 1$ *for some* $0 < \delta < 1$, *we have the followings;*

1. *if* $\dfrac{\sum_{i=1}^{k} \left(\frac{\theta_i^3}{n_i} + \frac{\theta_i}{n_i^3} \right)}{\left(\sum_{i=1}^{k} \left(\theta_i^2 + \frac{1}{N^2} \frac{\theta_i}{n_i} \right) \right)^2} \to 0$ *as* $k \to 0$, $\dfrac{\hat{\mathcal{V}}_1}{\mathcal{V}_1} \to 1$ *in probability.*

2. *if* $\dfrac{(\tilde{\theta})^3 \sum_{i=1}^{k} \frac{1}{n_i} + \tilde{\theta} \sum_{i=1}^{k} \frac{1}{n_i^3}}{\left(k(\tilde{\theta})^2 + \frac{\tilde{\theta}}{N^2} \sum_{i=1}^{k} \frac{1}{n_i} \right)^2} \to 0$, $\dfrac{\hat{\mathcal{V}}_{1*}}{\mathcal{V}_{1*}} \to 1$ *in probability.*

Proof See "Appendix". □

Remark 3 Lemma 5 includes the condition $\pi_i \leq \delta < 1$ which avoids dense case that the majority of observations are 1. Since our study focuses on sparse case, it is realistic to exclude π_is which are very close to 1. When data are dense, the homogeneity test of π_i can be done through testing $\pi_i^* \equiv 1 - \pi$ and $x_{ij}^* = 1 - x_{ij}$.

Remark 4 As an estimator of π_i^l or π^l for $l = 1, 2, 3, 4$, we used unbiased estimators of them. Instead of unbiased estimators, we may consider simply MLE, $(\hat{\pi}_i)^l$ or $(\hat{\pi})^l$ for $l = 1, 2, 3, 4$. For the first type estimator $\hat{\mathcal{V}}_1$, when sample sizes n_i are not large, unbiased estimators and MLE are different. Especially, if all n_is are small and k is large, then such small differences are accumulated so the behavior of estimators for variance is expected to be significantly different. This will be demonstrated in our simulation studies. On the other hand, for $\hat{\mathcal{V}}_{1*}$, unbiased estimators and MLEs for $(\pi)^l$ under H_0 behave almost same way even for small n_i since the total sample size $N = \sum_{i=1}^{k} n_i$ is large due to large k. The estimator of $\mathcal{V}_1$ based on $\hat{\pi}_i$, namely $\hat{\mathcal{V}}_1^{mle}$ has the larger variance

$$E \left(\hat{\mathcal{V}}_1^{mle} - \mathcal{V}_1 \right)^2 \asymp \sum_{i=1}^{k} \left(\frac{\theta_i^3}{n_i} + \frac{\theta_i}{n_i^3} \right) + \sum_{i \neq j} \frac{\theta_i \theta_j}{n_i n_j}$$

while $E(\hat{\mathcal{V}}_1 - \mathcal{V}_1)^2 \asymp \sum_{i=1}^{k} \left(\frac{\theta_i^3}{n_i} + \frac{\theta_i}{n_i^3} \right)$. Similarly, we can also define $\hat{\mathcal{V}}_{1*}^{mle}$ based on the $\hat{\pi} = \frac{\sum_{i=1}^{k} x_i}{N}$. Even with the given condition $\sum_{i=1}^{k} \left(\frac{\theta_i^3}{n_i} + \frac{\theta_i}{n_i^3} \right) / (\sum_{i=1}^{k} \theta_i^2 + \frac{1}{N^2} \sum_{i=1}^{k} \frac{\theta_i}{n_i})^2 = o(1)$, $\hat{\mathcal{V}}_1^{mle}$ may not be a ratio-consistent estimator due to the additional variation from biased estimation of π_i^l for $l = 2, 3, 4$. We present simulation studies comparing tests with $\hat{\mathcal{V}}_1$ and $\hat{\mathcal{V}}_1^{mle}$ later.

In Lemma 5, we present ratio consistency of $\hat{\mathcal{V}}_1$ and $\hat{\mathcal{V}}_{1*}$ under some conditions. Both conditions avoid too small π_is compared to n_is among k groups. It is obvious that the conditions are satisfied if all π_is are uniformly bounded away from 0 and 1. In general, however, the conditions allow small π_is which may converge to zero at some rate satisfying presented conditions on θ_is in lemmas and theorems.

Under H_0, we have two different estimators, $\hat{\mathcal{V}}_1$ and $\hat{\mathcal{V}}_{1*}$ and their corresponding test statistics, namely T_{new1} and T_{new2}, respectively:

$$T_{\text{new1}} = \frac{T}{\sqrt{\hat{\mathcal{V}}_1}}, \quad T_{\text{new2}} = \frac{T}{\sqrt{\hat{\mathcal{V}}_{1*}}}.$$

The following theorem shows that the proposed tests, T_{new1} and T_{new2}, are asymptotically size α tests.

Theorem 2 *Under $H_0 : \pi_i \equiv \pi$ for all $1 \leq i \leq k$, if the condition in Lemma 5 holds and $\frac{\sum_{i=1}^{k} \frac{1}{n_i}}{k\theta^3} \to 0$ for $\theta = \pi(1-\pi)$ under H_0, then $T_{\text{new1}} \to N(0, 1)$ in distribution and $T_{\text{new2}} \to N(0, 1)$ in distribution as $k \to \infty$.*

Proof See "Appendix". $\square$

Remark 5 The condition in Lemma 5 under the H_0 is $\dfrac{\theta^3 \sum_{i=1}^{k} \frac{1}{n_i} + \theta \sum_{i=1}^{k} \frac{1}{n_i^3}}{\left(k\theta^2 + \frac{\theta}{N^2} \sum_{i=1}^{k} \frac{1}{n_i}\right)^2} = o(1)$. This condition includes a variety of situations such as small values of π as well as small sample sizes. Furthermore, inhomogeneous sample sizes are also included. For example, when the sample sizes are bounded, we have $\sum_{i=1}^{k} \frac{1}{n_i} \asymp k$ and $\sum_{i=1}^{k} \frac{1}{n_i^3} \asymp k$ leading to $\dfrac{\theta^3 \sum_{i=1}^{k} \frac{1}{n_i} + \theta \sum_{i=1}^{k} \frac{1}{n_i^3}}{\left(k\theta^2 + \frac{\theta}{N^2} \sum_{i=1}^{k} \frac{1}{n_i}\right)^2} \leq \frac{1}{k\theta^3}$ which converges to 0 when $k\theta^3 \to \infty$. This happens when $\pi = k^{\epsilon - 1/3}$ for $0 < \epsilon < 1/3$ which is allowed to converge to 0. Another case is that sample sizes are highly unbalanced. For example, we have $n_i \asymp i^\alpha$ for $\alpha > 1$ which implies $\sum_{i=1}^{\infty} \frac{1}{n_i} < \infty$ and $\sum_{k=1}^{\infty} \frac{1}{n_i^3} < \infty$. Therefore, the condition is $\dfrac{\theta^3 \sum_{i=1}^{k} \frac{1}{n_i} + \theta \sum_{i=1}^{k} \frac{1}{n_i^3}}{\left(k\theta^2 + \theta \sum_{i=1}^{k} \frac{1}{n_i}\right)^2} \asymp \frac{\theta^3 + \theta}{(k\theta^2 + \theta)^2} \leq \frac{\theta^3 + \theta}{k^2 \theta^4} = \frac{1}{k^2 \theta} + \frac{1}{k^2 \theta^3} \to 0$ if $\pi \asymp k^\epsilon$ for $-\frac{2}{3} < \epsilon < 0$. In this case, the sample size n_i diverges as $i \to \infty$, so sample sizes are highly unbalanced. For the asymptotic normality, additional condition $\sum_{i=1}^{k} \frac{1}{n_i}/(k\theta^3) \to 0$ in Theorem 2 is satisfied for $-\frac{1}{3} < \epsilon < 0$.

From Theorem 2, we reject the H_0 if

$$T_{\text{new1}} (\text{or } T_{\text{new2}}) > z_{1-\alpha}$$

where $z_{1-\alpha}$ is $(1 - \alpha)$ quantile of a standard normal distribution. As explained in Sect. 2.2, note that the rejection region is one-sided since we have $E(T) \geq 0$, implying that large values of tests support the alternative hypothesis.

Although they have the same asymptotic null distribution, their power functions are different due to the different behaviors of $\hat{\mathcal{V}}_1$ and $\hat{\mathcal{V}}_{1*}$ under H_1. In general, it is not necessary to have the asymptotic normality under the H_1; however, to compare the powers analytically, one may expect asymptotic power functions to be more specific.

The following lemma states the asymptotic normality of $T/\sqrt{\text{Var}(T_1)}$ where $\text{Var}(T_1)$ is in (9) in Lemma 2. In the following asymptotic results, it is worth mentioning that we put some conditions on θ_is so that they do not approach to 0 too fast.

Theorem 3 *If* (i) $|\pi_i - \bar{\pi}| \geq \frac{1+\epsilon}{N}$ *for* $1 \leq i \leq k$, (ii) $\dfrac{\sum_{i=1}^{k}(\theta_i^4 + \frac{\theta_i}{n_i})}{\left(\sum_{i=1}^{k}\theta_i^4 + \frac{1}{N^2}\sum_{i=1}^{k}\frac{\theta_i}{N}\right)^2} \to 0$ *and*

(iii) $\dfrac{\max_i(\pi_i - \bar{\pi})^2(n_i\theta_i + 1)}{\mathcal{V}_1 + ||\boldsymbol{\pi} - \bar{\pi}||_{\theta\mathbf{n}}^2} \to 0$ *where* $||\boldsymbol{\pi} - \bar{\pi}||_{\theta\mathbf{n}}^2 = \sum_{i=1}^{k}n_i(\pi_i - \bar{\pi})^2\theta_i$, *then*

$$\frac{T - \sum_{i=1}^{k}n_i(\pi_i - \bar{\pi})^2}{\sqrt{\text{Var}(T_1)}} \to N(0, 1) \quad \text{in distribution}$$

where $\text{Var}(T_1)$ *is defined in* (9).

Proof See "Appendix".   □

Using Theorem 3, we obtain the asymptotic power of the proposed tests. We state this in the following corollary.

Corollary 2 *Under the assumptions in Lemma* 5 *and Theorem* 3, *the powers of* T_{new1} *and* T_{new2} *are*

$$P(T_{\text{new1}} > z_{1-\alpha}) - \bar{\Phi}\left(\frac{\sqrt{\mathcal{V}_1}}{\sqrt{\text{Var}(T_1)}}z_{1-\alpha} - \frac{\sum_{i=1}^{k}n_i(\pi_i - \bar{\pi})^2}{\sqrt{\text{Var}(T_1)}}\right) \to 0$$

and

$$P(T_{\text{new2}} > z_{1-\alpha}) - \bar{\Phi}\left(\frac{\sqrt{\mathcal{V}_{1*}}}{\sqrt{\text{Var}(T_1)}}z_{1-\alpha} - \frac{\sum_{i=1}^{k}n_i(\pi_i - \bar{\pi})^2}{\sqrt{\text{Var}(T_1)}}\right) \to 0$$

where $\bar{\Phi}(x) = 1 - \Phi(x) = P(Z > x)$ *for a standard normal random variable Z and* $\text{Var}(T_1)$ *defined in* (9).

2.3 Comparison of powers

In the previous section, we present the asymptotic power of tests, T_{new1} and T_{new2}. Currently, it does not look straightforward to tell one test is uniformly better than the others. However, one may consider some specific scenario and compare different tests under those scenario which may help to understand the properties of tests in a better way. Asymptotic powers depend on the configurations of $(\pi_i's)$, $(n_i's)$ and k. It is not possible to consider all configurations; however, what we want to show through simulations is that neither of T_{new1} and T_{new2} dominates the other.

Let $\beta(T)$ be the asymptotic power of a test statistic $\lim_{k\to\infty} P(T > z_{1-\alpha})$ where T is one of T_χ, T_{new1} and T_{new2}.

Theorem 4 *1. If sample sizes $n_1 = \cdots = n_k \equiv n$ and $\max_{1 \le i \le k} \pi_i < \frac{1}{2} - \frac{1}{\sqrt{3}}$, then*

$$\lim_{k\to\infty} (\beta(T_{\mathrm{new2}}) - \beta(T_{\mathrm{new1}})) \ge 0.$$

If $n_i = n$ for all $1 \le i \le k$ and $n\bar{\pi}(1 - \bar{\pi}) \to \infty$, then

$$\lim_{k\to\infty} (\beta(T_{\mathrm{new2}}) - \beta(T_\chi)) \ge 0.$$

2. *Suppose $\pi_i = \pi = k^{-\gamma}$ for $1 \le i \le k - 1$ and $\pi_k = k^{-\gamma} + \delta$ for $0 < \gamma < 1$ as well as $n_i = n$ for $1 \le i \le k - 1$, and $n_k = [nk^\alpha]$ for $0 < \alpha < 1$ where $[x]$ is the greatest integer which does not exceed x. Then, if $n \to \infty$,*
   (a) *for $\{(\alpha, \gamma) : 0 < \alpha < 1, 0 < \gamma < 1, 0 < \alpha + \gamma < 1, 0 < \gamma \le \frac{1}{2}\}$, then $\lim_k (\beta(T_{\mathrm{new1}}) - \beta(T_{\mathrm{new2}})) = 0$.*
   (b) *for $\{(\alpha, \gamma) : 0 < \alpha < 1, 0 < \gamma < 1, \alpha + \gamma > 1, \alpha > \frac{1}{2}\}$, then $\lim_k (\beta(T_{\mathrm{new1}}) - \beta(T_{\mathrm{new2}})) > 0$.*
3. *Suppose $\pi_1 = k^{-\gamma} + \delta$ and $n_1 = n \to \infty$ and $\pi_i = k^{-\gamma}$ and $n_i = [nk^\alpha]$ for $2 \le i \le n$. For $0 < \gamma < 1$ and $0 < \alpha < 1$, if $0 < \gamma < 1/2$ and $k^{1-\alpha-\gamma} = o(n)$, then*

$$\lim_{k\to\infty} (\beta(T_{\mathrm{new2}}) - \beta(T_{\mathrm{new1}})) > 0. \tag{18}$$

Proof See "Appendix". $\qquad\qquad\square$

From Theorem 4, we conjecture that T_{new2} has better powers than others when sample sizes are homogeneous or similar to each other. For inhomogeneous sample sizes, T_{new1} and T_{new2} have different performances from the cases of 2 and 3 in Theorem 4. We show numerical studies reflecting these cases later.

Although we compare the powers of the proposed tests under some local alternative, it is interesting to see different scenarios and compare powers. Instead of an analytical approach, we present numerical studies as follows: Since the asymptotic powers of T_{new1} and T_{new2} depend on the behavior of $\mathcal{V}_1$ and $\mathcal{V}_{1*}$, we compare those two variances under a variety of situations. If $\mathcal{V}_{1*} > \mathcal{V}_1$, then T_{new1} is more powerful than T_{new2}; otherwise, we have an opposite result. Although we compared the powers of tests in this paper in Theorem 4, there are numerous additional situations which are not covered analytically. We provide some additional situations from numerical studies here. We take $k = 100$, and we generate sample sizes $n_i \sim \{20, 21, \ldots, 200\}$ uniformly. The left panel is for $\pi_i \sim U(0.01, 0.2)$ and the left panel is for $\pi \sim U(0.01, 0.5)$ where $U(a, b)$ is the uniform distribution in (a, b). We consider $1,000$ different configurations of $(n_i, \pi_i)_{1 \le i \le 100}$ for each panel. We see that $\mathrm{Var}(T_1)$ and $\mathrm{Var}(T_1)_*$ have different behaviors when π_is are generated different ways. If π_is are widely spread out, then $\mathrm{Var}(T_1)_*$ is larger, otherwise $\mathrm{Var}(T_1)$ seems to be larger from our simulations (Fig. 1).
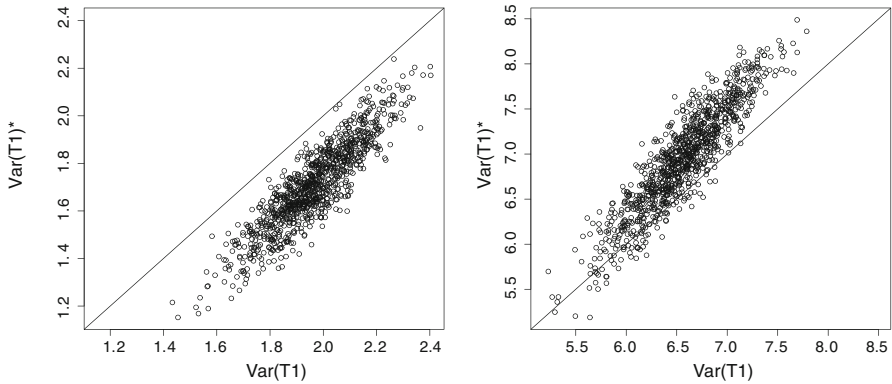
Fig. 1 Comparison of $\mathrm{Var}(T_1)$ and $\mathrm{Var}(T_1)_*$

We present simulation studies comparing the performance of T_{new1}, T_{new2} and existing tests. They have different performances depending on different situations.

3 Simulations

In this section, we present simulation studies to compare our proposed tests with existing procedures.

We first adopt the following simulation setup and evaluate our proposed tests. Let us define

$$\boldsymbol{n}_8 = 20(2, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8)$$
$$\boldsymbol{n}_{40} = 20(\boldsymbol{n}_1^*, \boldsymbol{n}_2^*, \dots, \boldsymbol{n}_8^*) = 20(2, \dots, 2, 2^2, \dots, 2^2, \dots, 2^8, \dots, 2^8)$$

where $\boldsymbol{n}_m^* = (2^m, 2^m, \dots, 2^m)$ is a 8 dimensional vector. We consider the following simulations (Tables 1, 2, 3, 4, 5, 6).

Setup 1 $\pi_i = 0.001$ for $1 \leq i \leq k-1$ and $\pi_k = 0.001 + \delta$ for $k = 8$ and $\boldsymbol{n}_8$
Setup 2 $\pi_i = 0.001 + \delta$ for $k = 1$ and $\pi_i = 0.001$ for $2 \leq i \leq k$ for $k = 8$ and $\boldsymbol{n}_8$
Setup 3 $\pi_1 = 0.001 + \delta$ and $\pi_i = 0.001$ for $2 \leq i \leq 8$, $k = 8$, $n_i = 2560$ for $1 \leq i \leq 8$
Setup 4 $\pi_i = 0.001$ for $1 \leq i \leq k-1$ and $\pi_k = 0.001 + \delta$ for $k = 40$ and $\boldsymbol{n}_{40}$
Setup 5 $\pi_i = 0.001 + \delta$ for $k = 1$ and $\pi_i = 0.001$ for $2 \leq i \leq k$ for $k = 40$ and $\boldsymbol{n}_{40}$
Setup 6 $\pi_i = 0.001 + \delta$ for $i = 1$ and $\pi_i = 0.001$ for $2 \leq i \leq k$. $n_i = 2560$ for $1 \leq i \leq 40$

As test statistics, we use T_{new1}, M_1, T_{new2}, M_2, TS, modTS and PW. Here, as discussed in Remark 4, M_1 uses $\hat{\mathcal{V}}_1^{mle}$ as an estimator of $\mathcal{V}_1$ in T_{new1} and M_2 uses $\hat{\mathcal{V}}_{1*}^{mle}$ for $\mathcal{V}_{1*}$ in T_{new2}. TS represents the test in (2) and modTS represents the test in (10). Chi represents Chi-square test based on $T_S > \chi^2_{k-1,1-\alpha}$ where $\chi^2_{k-1,1-\alpha}$ is the $(1-\alpha)$ quantile of Chi-square distribution with degrees of freedom $k-1$. PW is the test in Potthoff and Whittinghill (1966), and BL represents the test in Bathke and Lankowski

Table 1 Powers under Setup 1

δ	T_{new1}	$M1$	T_{new2}	$M2$	TS	modTS	Chi	PW
0.000	0.009	0.005	0.029	0.022	0.114	0.048	0.103	0.006
0.001	0.070	0.052	0.023	0.022	0.066	0.029	0.060	0.000
0.002	0.249	0.184	0.092	0.091	0.053	0.025	0.048	0.001
0.003	0.490	0.375	0.253	0.251	0.057	0.022	0.046	0.022
0.004	0.688	0.562	0.455	0.449	0.112	0.032	0.082	0.085
0.005	0.838	0.717	0.648	0.642	0.217	0.073	0.169	0.217
0.006	0.925	0.831	0.803	0.797	0.391	0.170	0.315	0.397
0.007	0.966	0.895	0.897	0.893	0.561	0.312	0.490	0.588
0.008	0.987	0.936	0.953	0.950	0.717	0.487	0.649	0.723
0.009	0.995	0.964	0.979	0.977	0.835	0.651	0.786	0.841

The cases of $\delta = 0$ represent Type I errors of tests. M_1 uses $\hat{\mathcal{V}}_1^{mle}$ as an estimator of $\mathcal{V}_1$ in T_{new1} and M_2 uses $\hat{\mathcal{V}}_{1*}^{mle}$ for $\mathcal{V}_{1*}$ in T_{new2}. TS represents the test in (2), and modTS represents the test in (10). Chi represents Chi-square test. PW is the test in Potthoff and Whittinghill (1966)

Table 2 Powers under Setup 2

δ	T_{new1}	$M1$	T_{new2}	$M2$	TS	modTS	Chi	PW
0.00	0.009	0.006	0.029	0.023	0.110	0.048	0.097	0.005
0.01	0.009	0.005	0.043	0.038	0.130	0.065	0.117	0.009
0.02	0.014	0.004	0.091	0.087	0.149	0.085	0.138	0.010
0.03	0.027	0.007	0.140	0.137	0.168	0.107	0.155	0.011
0.04	0.054	0.011	0.213	0.209	0.182	0.121	0.171	0.018
0.05	0.083	0.020	0.284	0.282	0.191	0.136	0.181	0.027
0.06	0.122	0.033	0.359	0.357	0.216	0.157	0.206	0.034
0.07	0.168	0.053	0.432	0.430	0.236	0.178	0.226	0.045
0.08	0.214	0.073	0.495	0.494	0.249	0.195	0.238	0.069
0.09	0.274	0.103	0.566	0.565	0.260	0.202	0.248	0.092

The cases of $\delta = 0$ represent Type I errors of tests

(2005). Note that BL is available only when sample sizes are all equal. For calculation of size and power of each test, we simulate 10,000 samples and compute empirical size and power based on 10,000 p values.

From the above scenario, we consider inhomogeneous sample sizes (Setups 1, 2, 4 and 5) and homogeneous sample sizes (Setups 3 and 6). Furthermore, when sample sizes are inhomogeneous, two cases are considered: one is the case that different π_i occurs for a study with large sample (Setups 1 and 4) and the other for a study with small sample (Setups 2 and 5). Setups 1–6 consider the cases that only one study has a different probability $(0.001 + \delta)$ and all the others have the same probability

Table 3 Powers under Setup 3

δ	T_{new1}	$M1$	T_{new2}	$M2$	TS	modTS	Chi	PW	BL
0.0000	0.036	0.023	0.054	0.060	0.040	0.034	0.030	0.018	0.065
0.0005	0.057	0.041	0.080	0.085	0.057	0.050	0.045	0.032	0.099
0.0010	0.123	0.099	0.152	0.158	0.119	0.106	0.095	0.078	0.186
0.0015	0.244	0.207	0.283	0.291	0.229	0.209	0.193	0.175	0.315
0.0020	0.388	0.345	0.430	0.436	0.379	0.358	0.341	0.309	0.459
0.0025	0.545	0.498	0.580	0.585	0.537	0.513	0.492	0.461	0.614
0.0030	0.669	0.631	0.696	0.700	0.671	0.649	0.632	0.598	0.738
0.0035	0.789	0.760	0.813	0.815	0.790	0.775	0.756	0.726	0.839
0.0040	0.863	0.842	0.880	0.882	0.863	0.853	0.840	0.816	0.900
0.0045	0.922	0.909	0.932	0.933	0.919	0.913	0.903	0.893	0.945

The cases of $\delta = 0$ represent Type I errors of tests. BL represents the test in Bathke and Lankowski (2005)

Table 4 Powers under Setup 4

δ	T_{new1}	$M1$	T_{new2}	$M2$	TS	modTS	Chi	PW
0.000	0.022	0.003	0.042	0.042	0.196	0.060	0.186	0.016
0.001	0.080	0.018	0.067	0.069	0.160	0.048	0.151	0.088
0.002	0.285	0.090	0.202	0.204	0.199	0.057	0.189	0.311
0.003	0.562	0.242	0.445	0.448	0.296	0.099	0.282	0.603
0.004	0.787	0.441	0.690	0.694	0.462	0.185	0.442	0.829
0.005	0.919	0.623	0.864	0.866	0.659	0.355	0.640	0.939
0.006	0.971	0.765	0.946	0.947	0.804	0.542	0.791	0.983
0.007	0.991	0.857	0.982	0.982	0.913	0.723	0.906	0.995
0.008	0.998	0.928	0.995	0.995	0.964	0.855	0.960	0.999
0.009	1.000	0.963	0.999	0.999	0.989	0.934	0.987	0.999

The cases of $\delta = 0$ represent Type I errors of tests

(0.001). On the other hand, we may consider the following cases which represent all probabilities are different from each other (Tables 7, 8).

Setup 7 $\pi_i = 0.001(1 + \epsilon_i)$, $k = 40$, $n_i = 2560$ for $1 \leq i \leq 40$ where ϵ_is are equally spaced grid in $[-\delta, \delta]$.

Setup 8 $\pi_i = 0.01(1 + \epsilon_i)$, $k = 40$, $\boldsymbol{n}_{40}^*$ where ϵ_is are equally spaced grid in $[-\delta, \delta]$.

From our simulations, we first see that T_{new1} obtains more powers than M_1, while T_{new2} and M_2 obtain almost similar powers. The performance of T_{new1} and T_{new2} is different depending on different situations. When sample sizes are homogeneous (Setups 3, 6 and 7), T_{new2} obtains slightly more power than T_{news} as shown in 1 in Theorem 4. On the other hand, when sample sizes are inhomogeneous, T_{new1} seems to have more advantage for the cases that different probability occurs for large sample sizes, while T_{new2} seems to obtain better powers for the opposite case. Overall, the performances of T_{new1} and T_{new2} are different depending on situations. Cochran's test

Table 5 Powers under Setup 5

δ	$T_{\text{new}1}$	$M1$	$T_{\text{new}2}$	$M2$	TS	modTS	Chi	PW
0.000	0.021	0.004	0.047	0.045	0.186	0.059	0.179	0.021
0.002	0.028	0.005	0.102	0.101	0.216	0.081	0.207	0.017
0.004	0.059	0.015	0.221	0.221	0.252	0.103	0.243	0.021
0.006	0.130	0.040	0.371	0.370	0.280	0.118	0.270	0.028
0.008	0.232	0.097	0.507	0.506	0.313	0.144	0.305	0.045
0.010	0.335	0.158	0.626	0.626	0.339	0.156	0.331	0.061
0.012	0.454	0.252	0.730	0.729	0.364	0.175	0.356	0.091
0.014	0.553	0.339	0.800	0.800	0.383	0.189	0.373	0.126

The cases of $\delta = 0$ represent Type I errors of tests

Table 6 Powers under Setup 6

δ	$T_{\text{new}1}$	$M1$	$T_{\text{new}2}$	$M2$	TS	modTS	Chi	PW	BL
0.000	0.049	0.029	0.058	0.059	0.048	0.038	0.041	0.032	0.061
0.001	0.093	0.065	0.107	0.108	0.093	0.079	0.083	0.067	0.114
0.002	0.273	0.222	0.297	0.299	0.271	0.240	0.249	0.236	0.300
0.003	0.535	0.479	0.560	0.562	0.535	0.504	0.512	0.511	0.568
0.004	0.776	0.736	0.793	0.795	0.777	0.756	0.761	0.739	0.803
0.005	0.902	0.884	0.910	0.911	0.911	0.901	0.903	0.891	0.921
0.006	0.966	0.957	0.969	0.969	0.966	0.961	0.963	0.966	0.974

The cases of $\delta = 0$ represent Type I errors of tests

Table 7 Powers under Setup 7

δ	$T_{\text{new}1}$	$M1$	$T_{\text{new}2}$	$M2$	TS	modTS	Chi	PW	BL
0	0.044	0.027	0.053	0.053	0.046	0.036	0.039	0.031	0.066
0.25	0.080	0.052	0.096	0.094	0.084	0.069	0.072	0.061	0.102
0.50	0.240	0.182	0.271	0.268	0.229	0.195	0.205	0.200	0.280
0.75	0.596	0.513	0.633	0.630	0.601	0.553	0.569	0.541	0.645
1.00	0.927	0.889	0.941	0.940	0.930	0.911	0.917	0.904	0.945

The cases of $\delta = 0$ represent Type I errors of tests

seems to fail in controlling a given size; however, the modified TS achieves reasonable empirical sizes. When sample sizes are homogeneous, the modified TS has comparable powers; however, for inhomogeneous sample sizes, the modified TS has significantly small powers compare to $T_{\text{new}1}$ and $T_{\text{new}2}$ for Setup 8 (Tables 8).

As suggested by a reviewer, we consider the following two more numerical studies when k is extremely large (Tables 9, 10).

Setup 9 $\pi_i = 0.01(1 + \epsilon_i)$, $k = 2000$, $n_i = 100$ for $1 \le i \le 2,000$ where ϵ_is are equally spaced grid in $[-\delta, \delta]$.

Table 8 Powers under Setup 8

δ	T_{new1}	$M1$	T_{new2}	$M2$	TS	modTS	Chi	PW
0.00	0.047	0.025	0.059	0.059	0.073	0.051	0.073	0.030
0.25	0.123	0.079	0.089	0.089	0.026	0.017	0.026	0.039
0.50	0.487	0.409	0.353	0.353	0.061	0.044	0.061	0.088
0.75	0.893	0.858	0.793	0.792	0.265	0.222	0.265	0.179
1.00	0.996	0.994	0.985	0.985	0.721	0.673	0.721	0.355

The cases of $\delta = 0$ represent Type I errors of tests

Table 9 Powers under Setup 9

δ	T_{new1}	$M1$	T_{new2}	$M2$	TS	modTS	Chi	PW	BL
0.0	0.0507	0.0118	0.0563	0.0563	0.0544	0.0234	0.0544	0.0499	0.0548
0.2	0.1119	0.0334	0.1197	0.1197	0.1171	0.0659	0.1171	0.1110	0.1178
0.4	0.5031	0.2796	0.5205	0.5204	0.5142	0.3868	0.5142	0.5014	0.5157
0.6	0.9709	0.9012	0.9730	0.9730	0.9727	0.9425	0.9727	0.9706	0.9728
0.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The cases of $\delta = 0$ represent Type I errors of tests

Setup 10 $\pi_i = 0.01(1 + \epsilon_i)$, $k = 2,000$, $\boldsymbol{n} = (\boldsymbol{n}_{1,250}, \boldsymbol{n}_{2,250}, \ldots, \boldsymbol{n}_{8,250})$ where $\boldsymbol{n}_{m,250} = (2^m, 2^m, \ldots, 2^m)$ is a 250 dimensional vector with all components 2^m and ϵ_is are equally spaced grid in $[-\delta, \delta]$.

Setup 9 is the case of a extremely large number of groups with small sample sizes. As mentioned in introduction, we focus on sparse count data in the sense that π_is are small, so we take $\pi_i = 0.01$ and homogeneous sample sizes $n_i = 100$ so that we have $E(X_i) = n_i \pi_i$ which represents very sparse data in each group. For the number of groups, we use $k = 2000$ which is much larger than $n_i = 100$. Table 9 shows sizes and powers of all tests, and we see that all tests have similar performances when sample sizes are homogeneous. On the other hand, for the case that sample sizes are highly unbalanced which is the case of Setup 10, Table 10 shows that our proposed tests control the nominal level of size and obtain increasing patter of powers, while tests based on Chi-square statistics fail in controlling the nominal level of size and obtaining powers. In particular, those Chi-square-based tests have decreasing patterns of powers even though the effect sizes (δ in this case) increase. PW controls the size and has increasing pattern of powers; however, the powers of PW are much smaller than those of our proposed tests. All codes will be available upon request.

4 Real examples

In this section, we provide real examples for testing the homogeneity of binomial proportions from a large number of independent groups.

Table 10 Powers under Setup 10

δ	T_{new1}	$M1$	T_{new2}	$M2$	TS	modTS	Chi	PW
0.0	0.034	0.000	0.055	0.055	0.182	0.046	0.182	0.043
0.2	0.055	0.000	0.050	0.050	0.002	0.000	0.002	0.055
0.4	0.164	0.003	0.101	0.101	0.000	0.000	0.000	0.081
0.6	0.458	0.042	0.278	0.278	0.000	0.000	0.000	0.122
0.8	0.840	0.360	0.650	0.650	0.000	0.000	0.000	0.200
1.0	0.985	0.875	0.933	0.933	0.000	0.000	0.000	0.309

The cases of $\delta = 0$ represent Type I errors of tests

Table 11 p values for homogeneity tests

	T_{new1}	$M1$	T_{new2}	$M2$	TS	modTS	Chi	PW
MI/Rosig	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DCV/Rosig	0.063	0.133	0.003	0.004	0.000	0.004	0.002	0.059
MI/Cont	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DCV/Cont	0.107	0.242	0.079	0.084	0.609	0.406	0.584	0.229

Rosig rosiglitazone group, *Cont* control group, *MI* myocardial infarction, *DCV* death from cardiovascular

We apply our proposed tests and existing tests to the rosiglitazone data in Nissen and Wolski (2007). The data set includes the 42 studies and consists of study size (N), number of myocardial infarctions (MI) and number of deaths (D) for rosiglitazone (treatment) and the corresponding results under control arm for each study.

We consider testing (1) for the proportions of myocardial infarctions and death rate (D) from cardiovascular causes. There are four situations, (i) MI/rosiglitazone, (ii) death from cardiovascular (DCV)/rosiglitazone, (iii) MI/control and (iv) death from cardiovascular(DCV)/control. Table 4 shows the p values for different situations and different test statistics. In case of MI/rosiglitazone and MI/control, all tests have 0 p value. On the other hand, for the other two cases, some tests have different results. For DCV/Rosiglitazone, T_{new2}, TS and modTS have small p values, while T_{new1} and PW have slightly larger p values. For DCV/Control, T_{new1} and T_{new2} have much small p values (0.107 and 0.079) compared to T_S, modTS, Chi and PW (0.609, 0.406, 0.584 and 0.229, respectively) (Table 11).

5 Concluding remarks

In this paper, we considered testing homogeneity of binomial proportions from a large number of independent studies. In particular, we focused on the sparse data and heterogeneous sample sizes which may affect the identification of null distributions. We proposed new tests and showed their asymptotic results under some regular conditions. We provided simulations and real data examples, which show that our proposed tests are convincing in case of sparse and a large number of studies. This is a convincing

result since our proposed test is most reliable in controlling a given size from our simulations, so small p values from our proposed test are strong evidence against the null hypotheses.

Appendix

A Proof of Theorem 1

We use the Lyapunov's condition for the asymptotic normality of $\frac{T_S - E(T_S)}{\sqrt{B_k}}$. Let $T_{Si} = \frac{(X_i - n_i \bar{\pi})^2}{n_i \bar{\pi}(1-\bar{\pi})}$, then we define $D_i = T_{Si} - E(T_{Si}) = \frac{(X_i - n_i \bar{\pi})^2}{n_i \bar{\pi}(1-\bar{\pi})} - \frac{n_i(\pi_i - \bar{\pi})^2}{n_i \bar{\pi}(1-\bar{\pi})} - \frac{\pi_i(1-\pi_i)}{n_i \bar{\pi}(1-\bar{\pi})} = \frac{1}{n_i \bar{\pi}(1-\bar{\pi})}((X_i - n_i \pi_i)^2 + 2n_i(X_i - n_i \pi_i)(\pi_i - \bar{\pi}) - n_i \pi_i(1-\pi_i))$. We show that the Lyapunov's condition is satisfied, $\frac{\sum_{i=1}^k E(D_i^4)}{B_k^2} \to 0$. We see that

$$
\begin{aligned}
\frac{\sum_{i=1}^k E(D_i^4)}{B_k^2} &\leq \frac{1}{B_k^2} \sum_{i=1}^k \frac{n_i^4 E(\hat{\pi}_i - \pi_i)^8 + 2^4 n_i^4(\pi_i - \bar{\pi})^4 n_i^4 E(\hat{\pi}_i - \pi_i)^4 + n_i^4 \pi_i(1-\pi_i)^4}{n_i^4(\bar{\pi}(1-\bar{\pi}))^4} \\
&= \frac{1}{(\bar{\pi}(1-\bar{\pi}))^4 B_k^2} \sum_{i=1}^k \left[\left(\theta_i^4 + \frac{\theta_i}{n_i} \right) + n_i^2(\pi_i - \bar{\pi})^4(3\theta_i^2 + \frac{(1-6\theta_i)\theta_i}{n_i}) + \theta_i^4 \right] \\
&\leq \frac{\sum_{i=1}^k \left(2\theta_i^4 + \frac{\theta_i}{n_i} \right)}{(\bar{\pi}(1-\bar{\pi}))^4 B_k^2} + \frac{3 \sum_{i=1}^k n_i^2 \theta_i(\pi_i - \bar{\pi})^4(\theta_i + \frac{1}{n_i})}{(\bar{\pi}(1-\bar{\pi}))^4 B_k^2} \\
&\to 0
\end{aligned}
$$

from the given conditions. Therefore, we have the asymptotic normality of $\frac{T_S - E(T_S)}{\sqrt{B_k}} \to N(0,1)$ in distribution. Furthermore, we also have the asymptotic normality of

$$
T_0 = \frac{T_S - k}{\sqrt{B_{0k}}} = \sqrt{\frac{B_k}{B_{0k}}} \frac{T_S - k}{\sqrt{B_k}} + \frac{k - E(T_S)}{\sqrt{B_{0k}}} = \sigma_k \frac{T_S - k}{\sqrt{B_k}} + \mu_k
$$

which leads to $P(T_0 \geq z_{1-\alpha}) = P(\sigma_k \frac{T_S - k}{\sqrt{B_k}} + \mu_k \geq z_{1-\alpha}) = P(\frac{T_S - k}{\sqrt{B_k}} \geq \frac{z_{1-\alpha}}{\sigma_k} - \mu_k)$. Using $\frac{T_S - k}{\sqrt{B_k}} \to N(0,1)$ in distribution, we have $P(T_0 \geq z_{1-\alpha}) - \bar{\Phi}(\frac{z_{1-\alpha}}{\sigma_k} - \mu_k) \to 0$. $\square$

B Proof of Lemma 2

Since T_{1i} and T_{1j} for $i \neq j$ are independent, we have $\mathcal{V}_1 \equiv \text{Var}(T_1) = \sum_{i=1}^k \text{Var}(T_{1i})$ where

$$
\begin{aligned}
\text{Var}(T_{1i}) &= n_i^2 \text{Var}[(\hat{\pi}_i - \pi_i)^2] + d_i^2 \text{Var}[\hat{\pi}_i(1 - \hat{\pi}_i)] + 4n_i^2(\pi_i - \bar{\pi})^2 \text{Var}[(\hat{\pi}_i - \pi_i)] \\
&\quad - 2n_i d_i \text{Cov}((\hat{\pi}_i - \pi_i)^2, \hat{\pi}_i(1 - \hat{\pi}_i)) \\
&\quad + 2\text{Cov}(n_i(\hat{\pi}_i - \pi_i)^2, 2n_i(\hat{\pi}_i - \pi_i)(\pi_i - \bar{\pi})) \\
&\quad - 2Cov(2n_i(\hat{\pi}_i - \pi_i)(\pi_i - \bar{\pi}), d_i \hat{\pi}_i(1 - \hat{\pi}_i)).
\end{aligned}
$$

Using the following results

$$\mathrm{Var}[(\hat{\pi}_i - \pi_i)^2] = E[(\hat{\pi}_i - \pi_i)^4] - (E[(\hat{\pi}_i - \pi_i)^2])^2$$

$$= \frac{2\theta_i^2}{n_i^2} + \frac{(1 - 6\theta_i)\theta_i}{n_i^3}$$

$$\mathrm{Var}[\hat{\pi}_i(1 - \hat{\pi}_i)] = \frac{(1 - \theta_i)\theta_i}{n_i} - \frac{2\theta_i(1 - 4\theta_i)}{n_i^2} + \frac{(1 - 6\theta_i)\theta_i}{n_i^3}$$

$$\mathrm{Cov}((\hat{\pi}_i - \pi_i)^2, \hat{\pi}_i(1 - \hat{\pi}_i)) = \frac{n_i - 1}{n_i^3}\theta_i(1 - 6\theta_i)$$

$$\mathrm{Cov}((\hat{\pi}_i - \pi_i)^2, \hat{\pi}_i - \pi_i) = E(\hat{\pi}_i - \pi_i)^3 = \frac{(1 - 2\pi_i)\theta_i}{n_i^2}$$

$$\mathrm{Cov}((\hat{\pi}_i - \pi_i), \hat{\pi}_i(1 - \hat{\pi}_i)) = \frac{(1 - 2\pi_i)\theta_i}{n_i}\left(1 - \frac{1}{n_i}\right),$$

we derive

$$\mathrm{Var}(T_1) = \sum_{i=1}^{k}\left\{\theta_i^2\left(2 - \frac{6}{n_i} - \frac{d_i^2}{n_i} + \frac{8d_i^2}{n_i^2} - \frac{6d_i^2}{n_i^3} + 12d_i\frac{n_i - 1}{n_i^2}\right)\right.$$

$$\left. + \theta_i\left(\frac{1}{n_i} + \frac{d_i^2}{n_i} - \frac{2d_i^2}{n_i^2} + \frac{d_i^2}{n_i^3} - 2d_i\frac{n_i - 1}{n_i^2}\right)\right\}$$

$$+ 4\sum_{i=1}^{k}n_i(\pi_i - \bar{\pi})^2\theta_i + \frac{4}{N}\sum_{i=1}^{k}n_i(\pi_i - \bar{\pi})(1 - 2\pi_i)\theta_i$$

$$= \sum_{i=1}^{k}\mathcal{A}_{1i}\theta_i^2 + \sum_{i=1}^{k}\mathcal{A}_{2i}\theta_i + 4\sum_{i=1}^{k}n_i(\pi_i - \bar{\pi})^2\theta_i$$

$$+ \frac{4}{N}\sum_{i=1}^{k}n_i(\pi_i - \bar{\pi})(1 - 2\pi_i)\theta_i$$

where $\mathcal{A}_{1i} = \left(2 - \frac{6}{n_i} - \frac{d_i^2}{n_i} + \frac{8d_i^2}{n_i^2} - \frac{6d_i^2}{n_i^3} + 12d_i\frac{n_i-1}{n_i^2}\right)$ and $\mathcal{A}_{2i} = \left(\frac{1}{n_i} + \frac{d_i^2}{n_i} - \frac{2d_i^2}{n_i^2}\right.$ $\left. + \frac{d_i^2}{n_i^3} - 2d_i\frac{n_i-1}{n_i^2}\right) = \frac{n_i}{N^2}$ from $d_i = \frac{n_i}{n_i-1}\left(1 - \frac{n_i}{N}\right)$. $\qquad\square$

C Proof of Lemma 3

1. Using $d_i = \frac{n_i}{n_i-1}(1 - \frac{n_i}{N}) < 2$, we can derive $\mathcal{A}_{1i}$ is uniformly bounded since

$\mathcal{A}_{1i} = 2 - \frac{6}{n_i} - \frac{6d_i^2}{n_i} + \frac{8d_i^2}{n_i^2} - \frac{6d_i^2}{n_i^3} + 12\frac{n_i}{n_i-1}\frac{n_i-1}{n_i^2}(1 - \frac{n_i}{N}) = 2 + \frac{6}{n_i} - \frac{12}{N} + \frac{d_i^2}{n_i}(-1 +$
$\frac{8}{n_i} - \frac{6}{n_i^2})(1 - \frac{n_i}{N})$. Let $x = \frac{1}{n_i} \le \frac{1}{2}$, then $f(x) = (-1 + \frac{8}{n_i} - \frac{6}{n_i^2}) = -6(x - \frac{2}{3})^2 + \frac{7}{9}$
which has the value $-1 < f(x) \le \frac{3}{2}$. Therefore, we have $2 + \frac{6}{n_i} - \frac{12}{N} + \frac{6}{n_i} \ge$

$\mathcal{A}_{1i} \geq 2 + \frac{6}{n_i} - \frac{12}{N} - \frac{4}{n_i}$. Using $n_i \geq 2$ and $N \to \infty$ as $k \to \infty$, lower and upper bounds are uniformly bounded away from 0 and ∞ for all i. Therefore, we have $\mathcal{A}_{1i} \asymp 1$ and $\mathcal{A}_{2i} = \frac{n_i}{N^2}$ leading to $\mathcal{V}_1 = \sum_{i=1}^{k} \mathcal{A}_{1i}\theta_i^2 + \sum_{i=1}^{k} \mathcal{A}_{2i}\theta_i \asymp \sum_{i=1}^{k} \theta_i^2 + \frac{1}{N^2}\sum_{i=1}^{k} n_i\theta_i$.

2. Let $\mathcal{G}_n = 4\sum_{i=1}^{k} n_i(\pi_i - \bar{\pi})^2\theta_i + 4\frac{1}{N}\sum_{i=1}^{k} n_i(\pi_i - \bar{\pi})(1-2\pi_i)\theta_i = 4\sum_{i=1}^{k} \theta_i G_i$ where $G_i = n_i(\pi_i - \bar{\pi})^2 + \frac{n_i}{N}(\pi_i - \bar{\pi})(1-2\pi_i)$. If we define $\mathcal{B} = \{i : |\pi_i - \bar{\pi}| \geq \frac{(1+\epsilon)}{N}\}$ for some $\epsilon > 0$, then we decompose

$$\mathcal{V}_1 = \underbrace{\sum_{i \in \mathcal{B}}(\mathcal{A}_{1i}\theta_i^2 + \mathcal{A}_{2i}\theta_i)}_{\mathcal{F}_1} + \underbrace{\sum_{i \in \mathcal{B}^c}(\mathcal{A}_{1i}\theta_i^2 + \mathcal{A}_{2i}\theta_i)}_{\mathcal{F}_2} \tag{19}$$

$$\mathcal{G}_n = 4\underbrace{\sum_{i \in \mathcal{B}}\theta_i G_i}_{\mathcal{G}_{n1}} + 4\underbrace{\sum_{i \in \mathcal{B}^c}\theta_i G_i}_{\mathcal{G}_{n2}} \equiv 4\mathcal{G}_{n1} + 4\mathcal{G}_{n2}. \tag{20}$$

For $i \in \mathcal{B}$, we have $\frac{n_i}{N}|(1-2\pi_i)(\pi_i - \bar{\pi})\theta_i| \leq \frac{n_i}{(1+\epsilon)}(\pi_i - \bar{\pi})^2\theta_i$ which implies

$$\frac{4\epsilon}{1+\epsilon}\sum_{i \in \mathcal{B}} n_i(\pi_i - \bar{\pi})^2\theta_i \leq 4\mathcal{G}_{n1} \leq \frac{4(2+\epsilon)}{1+\epsilon}\sum_{i \in \mathcal{B}} n_i(\pi_i - \bar{\pi})^2\theta_i.$$

This leads to $4\mathcal{G}_{n1} \asymp \sum_{i \in \mathcal{B}} n_i(\pi_i - \bar{\pi})^2\theta_i$ and

$$\mathcal{F}_1 + 4\mathcal{G}_{n1} \asymp \mathcal{F}_1 + \sum_{i \in \mathcal{B}} n_i(\pi_i - \bar{\pi})^2\theta_i. \tag{21}$$

For $\mathcal{B}^c = \{i \,|\, |\pi_i - \bar{\pi}| < \frac{(1+\epsilon)}{N}\}$, we first show $\mathcal{F}_2 + 4\mathcal{G}_{n2} \geq \sum_{i \in \mathcal{B}^c} \mathcal{A}_{1i}\theta_i^2$. For $i \in \mathcal{B}^c$ and $x = \pi_i - \bar{\pi}$, we have $G_i = n_i(x + \frac{1}{2N}(1-2\pi_i))^2 - \frac{(1-2\pi_i)^2 n_i}{4N^2} \geq -\frac{(1-2\pi_i)^2 n_i}{4N^2}$ leading to

$$\mathcal{F}_2 + 4\mathcal{G}_{n2} \geq \sum_{i \in \mathcal{B}^c} \mathcal{A}_{1i}\theta_i^2 + \frac{1}{N^2}\sum_{i \in \mathcal{B}^c} n_i\theta_i\left(1 - (1-2\pi)^2\right)$$

$$= \sum_{i \in \mathcal{B}^c} \mathcal{A}_{1i}\theta_i^2 + \frac{4}{N^2}\sum_{i \in \mathcal{B}^c} n_i\theta_i^2 = \sum_{i \in \mathcal{B}^c} \mathcal{A}_{1i}\theta_i^2 + 4\sum_{i \in \mathcal{B}^c} \mathcal{A}_{2i}\theta_i^2$$

$$> \sum_{i \in \mathcal{B}^c} \mathcal{A}_{1i}\theta_i^2. \tag{22}$$

The upper bound of $4\mathcal{G}_{n_2}$ is

$$4\mathcal{G}_{n_2} \leq \frac{4(1+\epsilon)}{N^2}\sum_{i \in \mathcal{B}^c} n_i\theta_i = 4(1+\epsilon)\sum_{k} \mathcal{A}_{2i}\theta_i$$

resulting in

$$\mathcal{F}_2 + 4\mathcal{G}_{n2} \leq 4(1+\epsilon) \sum_{i \in \mathcal{B}^c} \mathcal{A}_{1i}\theta_i^2 + 4(1+\epsilon) \sum_{i \in \mathcal{B}^c} \mathcal{A}_{2i}\theta_i + 4 \sum_{i \in \mathcal{B}^c} n_i(\pi_i - \bar{\pi})^2 \theta_i$$

$$< 4(1+\epsilon)(\mathcal{F}_2 + \sum_{i \in \mathcal{B}^c} n_i(\pi_i - \bar{\pi})^2 \theta_i). \tag{23}$$

Combining (22) and (23), we have

$$\sum_{i \in \mathcal{B}^c} \mathcal{A}_{1i}\theta_i^2 < \mathcal{F}_2 + \mathcal{G}_{n2} < 4(1+\epsilon)(\mathcal{F}_2 + \sum_{i \in \mathcal{B}^c} n_i(\pi_i - \bar{\pi})^2 \theta_i). \tag{24}$$

From (21) and (24), we conclude, for $K = 4(1+\epsilon)$,

$$\sum_{i=1}^{k} \mathcal{A}_{1i}\theta_i^2 < \mathrm{Var}(T_1) \leq K(\nu_1 + ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\mathbf{n}\theta}^2).$$

In particular, if $\mathcal{B}^c$ is an empty set, then we have $\mathrm{Var}(T) = \mathcal{F}_1 + 4\mathcal{G}_{n1}$, therefore (21) implies (15). □

D Proof of Lemma 4

Let $X = \sum_{i=1}^{n} X_i$ where X_is are iid Bernoulli(π). In expansion of $(X - n\pi)$, each term has the form of $(X_{i_1} - \pi)^{m_1}(X_{i_2} - \pi)^{m_2} \cdots (X_{i_k} - pi)^{m_k}$ for $1 \leq i_1, \ldots, i_k \leq n$ and $m_1 + \cdots + m_k = n$, so if there exists at least one $m_k = 1$, then expectation of the term is zero. We only need to consider the terms without $(X_{i_j} - \pi)$, so we finally have

$$
\begin{aligned}
E(X - n\pi)^8 &= E\left(\sum_{i=1}^{n}(X_i - \pi)\right)^8 \\
&= \binom{n}{1} E(X_1 - \pi)^8 \\
&\quad + 2\binom{8}{6,2}\binom{n}{2} E(X_1 - \pi)^6 E(X_1 - \pi)^2 \\
&\quad + 2\binom{n}{2}\binom{8}{5,3} E(X_1 - \pi)^5 E(X_1 - \pi)^3 \\
&\quad + \binom{n}{2}\binom{8}{4,4}[E(X_1 - \pi)^4]^2 \\
&\quad + \frac{3!}{2!}\binom{n}{3}\binom{8}{4,2,2} E(X_1 - \pi)^4[E(X_1 - \pi)^2]^2 \\
&\quad + \frac{3!}{2!}\binom{n}{3}\binom{8}{3,3,2}[E(X_1 - \pi)^3]^2 E(X_1 - \pi)^2 \\
&\quad + \binom{n}{4}\binom{8}{2,2,2,2}[E(X_1 - \pi)^2]^4.
\end{aligned}
$$

We have $E(X_1 - \pi)^m = \sum_{i=0}^m \binom{m}{i} E(X_1^i)(-\pi)^{m-i} = (-\pi)^m + \sum_{i=1}^m \binom{m}{i} E(X_1^i)$ $(-\pi)^{m-i}$ and using $E(X_1^i) = E(X_i) = \pi$ for $i \geq 1$, we obtain $E(X_1 - \pi)^m = (-\pi)^m + \pi \sum_{i=1}^m \binom{m}{i}(-\pi)^{m-i} = (-\pi)^m - \pi(-\pi)^m + \pi \sum_{i=0}^m \binom{m}{i}(-\pi)^m = (1 - \pi)(-\pi)^m + \pi(1 - \pi)^m = \pi(1 - \pi)((-1)^m \pi^{m-1} + (1 - \pi)^{m-1}) \leq \pi(1 - \pi)$ for $m \geq 2$. Since all coefficients in the expansion of $E(\sum_{i=1}^n (X_i - \pi))$ are fixed constants, for some universal constant $C > 0$, we have

$$E(X - n\pi)^8 \leq C \max(n\pi(1 - \pi), (n\pi(1 - \pi))^2, (n\pi(1 - \pi))^3, (n\pi(1 - \pi))^4)$$
$$= C \max\{n\pi(1 - \pi), (n\pi(1 - \pi))^4\}.$$

since maximum is obtained at either $n\pi(1 - \pi)$ or $(n\pi(1 - \pi))^4$ depending on $n\pi(1 - \pi) \leq 1$ or $n\pi(1 - \pi) > 1$.

For the second equation, we first consider the moment of $E(\hat{\pi}^4)$ and $E((1 - \hat{\pi})^4)$. The latter one is easily obtained from the first one by changing the distribution from $B(n, \pi)$ to $B(n, 1 - \pi)$. We first obtain

$$E\hat{\pi}^4 = \pi^4 + \frac{6\pi^2\theta}{n} + \frac{4\pi(1 - 2\pi)\theta}{n^2} + \frac{3\theta^2}{n^2} + \frac{(1 - 6\theta)\theta}{n^3}$$

$$\leq \pi^4 + \frac{6\pi^3}{n} + \frac{7\pi^2}{n^2} + \frac{\pi}{n^3}$$

$$\leq 7\left(\pi^4 + \frac{\pi^3}{n} + \frac{\pi^2}{n^2} + \frac{\pi}{n^3}\right)$$

$$= 28 \max\left(\pi^4, \frac{\pi}{n^3}\right)$$

where the last equality holds due to the fact that the maximum is obtained at either π^4 or $\frac{\pi}{n^3}$ depending on $\pi \geq \frac{1}{n}$ or $\pi < \frac{1}{n}$. Similarly, the following inequality is obtained

$$E(1 - \hat{\pi})^4 \leq 28 \max\left((1 - \pi)^4, \frac{1 - \pi}{n^3}\right).$$

Using $E\hat{\pi}^4(1 - \hat{\pi})^4 \leq \min(E\hat{\pi}^4, E(1 - \hat{\pi})^4)$, we have

$$E\hat{\pi}^4(1 - \hat{\pi})^4 \leq \min(E\hat{\pi}^4, E(1 - \hat{\pi})^4)$$

$$\leq 28 \min\left\{\max\left(\pi^4, \frac{\pi}{n^3}\right), \max\left((1 - \pi)^4, \frac{1 - \pi}{n^3}\right)\right\}$$

$$= \max\left(\pi^4, \frac{\pi}{n^3}\right) \quad \text{if } \pi \leq \frac{1}{2}$$

$$\max\left((1 - \pi)^4, \frac{1 - \pi}{n^3}\right) \quad \text{if } \pi > \frac{1}{2}.$$

If $\pi \leq \frac{1}{2}$, $\pi \geq 2\pi(1 - \pi) = 2\theta$; if $\pi > \frac{1}{2}$, $1 - \pi \leq 2\theta$. So the last equality is

$$E\hat{\pi}^4(1 - \hat{\pi})^4 \leq C' \max\left(\theta^4, \frac{\theta}{n^3}\right)$$

for some universal constant C'.

We use the following relationship: for some constants $b_m, m = 1, \ldots, l - 1$

$$X^l = \prod_{j=1}^{l}(X - j + 1) + \sum_{m=1}^{l-1} b_m \prod_{j=1}^{m}(X - j + 1).$$

For example we have $x^3 = x(x-1)(x-2)+3x(x-1)+x$. Using $E\prod_{j=1}^{l}(X-j+1) = \prod_{j=1}^{l}(n - j + 1)\pi^l$,

$$E\hat{\pi}^l = \frac{1}{n^l} E \prod_{j=1}^{l}(X - j + 1) + \frac{1}{n^l} \sum_{m=1}^{l-1} b_m E\left(\prod_{j=1}^{m}(X - j + 1)\right)$$

$$= \pi^l + O\left(\frac{\pi^l}{n}\right) + O\left(\sum_{m=1}^{l-1} \frac{\pi^m}{n^{l-m}}\right)$$

$$= \pi^l + O\left(\frac{\pi^{l-1}}{n} + \frac{\pi}{n^{l-1}}\right).$$

Using this, we can derive

$$E(\hat{\pi}^l - \pi^l)^2 = E\hat{\pi}^{2l} - 2\pi^l E\hat{\pi}^l + \pi^{2l} = O\left(\frac{\pi^{2l-1}}{n} + \frac{\pi}{n^{2l-1}}\right).$$

$$E(\hat{\eta}_l - \pi^l)^2 = E(\hat{\eta}_l - \hat{\pi}_l + \hat{\pi}_l - \pi_l)^2 \leq 2^2 E(\hat{\eta}_l - \hat{\pi}^l)^2 + 2^2 E(\hat{\pi}^l - \pi^l)^2.$$

Since $\hat{\eta}_l - \hat{\pi}^l = \hat{\pi}^l O(\frac{1}{n}) + \sum_{i=1}^{l-1} \hat{\pi}^{l-i} O(\frac{1}{n^i})$, we have $E(\hat{\eta}_l - \hat{\pi}^l)^2 \leq \left\{ E(\hat{\pi}^{2l}) O(\frac{1}{n^2}) + \sum_{i=1}^{l-1} \hat{\pi}^{2l-2i} O(\frac{1}{n^{2i}}) \right\}$. Using $E\hat{\pi}^{2l} = \pi^{2l} + O\left(\frac{\pi}{n^{2l-1}} + \frac{\pi^{2l-1}}{n}\right)$ from (4), we obtain

$$E(\hat{\eta}_l - \hat{\pi}^l)^2 = O\left(\frac{1}{n^2}\right)\left(\pi^{2l} + O\left(\frac{\pi}{n^{2l-1}} + \frac{\pi^{2l-1}}{n}\right)\right)$$

$$+ \sum_{i=1}^{l-1}\left(\pi^{2l-2i} + O\left(\frac{\pi}{n^{2l-2i-1}} + \frac{\pi^{2l-2i-1}}{n}\right)\right) O\left(\frac{1}{n^{2i}}\right)$$

$$= O\left(\sum_{i=1}^{l-1} \frac{\pi^{2(l-i)}}{n^{2i}} + \frac{\pi}{n^{2l-1}}\right). \tag{25}$$

We can show $\frac{\pi^{2(l-i)}}{n^i} \leq \frac{\pi}{n^{2l-1}} + \frac{\pi^{2l-1}}{n}$ for $2 \leq i \leq l-1$ since $\frac{\pi^{2(l-i)}}{n^i} \leq \frac{\pi^{2l-1}}{n}$ for $\pi \geq \frac{1}{n}$ and $\frac{\pi^{2(l-i)}}{n^i} \leq \frac{\pi}{n^{2l-1}}$ for $\pi < \frac{1}{n}$. Using this, we have $(25) \leq O\left(\frac{\pi}{n^{2l-1}} + \frac{\pi^{2l-1}}{n}\right)$ which proves $E(\hat{\eta}_l - \hat{\pi}^l)^2 = O\left(\frac{\pi}{n^{2l-1}} + \frac{\pi^{2l-1}}{n}\right)$. $\qquad \square$

E Proof of Lemma 5

For the ratio consistency of $\hat{\mathcal{V}}_1$, it is enough to show $\frac{E[(\hat{\mathcal{V}}_1 - \mathcal{V}_1)^2]}{(\mathcal{V}_1)^2} \to 0$ as $k \to \infty$. Since $\hat{\mathcal{V}}_1$ is an unbiased estimator of $\mathcal{V}_1$,

$$
\begin{aligned}
\text{Var}(\hat{\mathcal{V}}_1) &= E[(\hat{\mathcal{V}}_1 - \mathcal{V}_1)^2] \\
&= \sum_{i=1}^{k} \sum_{l=1}^{4} a_{li}^2 E[(\hat{\eta}_{li} - \eta_{li})^2] + \sum_{i \neq i'} \sum_{l \neq l'} a_{li} a_{l'i'} E[(\hat{\eta}_{li} - \eta_{li})(\hat{\eta}_{l'i'} - \eta_{l'i'})] \\
&= \sum_{i=1}^{k} \sum_{l=1}^{4} a_{li}^2 E[(\hat{\eta}_{li} - \eta_{li})^2]
\end{aligned}
$$

where the last equality follows since $E[(\hat{\eta}_{li} - \eta_{li})(\hat{\eta}_{l'i'} - \eta_{l'i'})] = E[(\hat{\eta}_{li} - \eta_{li})]E[(\hat{\eta}_{l'i'} - \eta_{l'i'})] = 0$ because $\hat{\eta}_{li}$ and $\hat{\eta}_{l'i'}$ are independent for $i \neq i'$ and both are unbiased estimators. Since $\mathcal{V}_1$ depends on $\theta_i = \pi_i(1 - \pi_i)$, we have the same result if we change π_i to $1 - \pi_i$; in other words, $\text{Var}(\hat{\mathcal{V}}_1) = \sum_{i=1}^{k} \sum_{l=1}^{4} a_{li}^2 (\eta_{li} - \eta_{li})^2 = \sum_{i=1}^{k} \sum_{l=1}^{4} a_{li}^2 (\hat{\eta}_{li}^* - \eta_{li}^*)^2$ where $\eta_{li}^* = (1 - \pi_i)^l$ and $\hat{\eta}_{li}^*$ is the corresponding unbiased estimator. For $\pi \leq 1/2$, we use $\mathcal{V}_1 = \sum_{i=1}^{k} \sum_{l=1}^{4} a_{li} \pi_i^l$ and obtain $\text{Var}(\hat{\mathcal{V}}_1) = O(\sum_{i=1}^{k} (\frac{\pi_i^3}{n_i} + \frac{\pi_i}{n_i^3}))$ from Lemma 4. Since $\pi_i \leq \delta < 1$, we have $\text{Var}(\hat{\mathcal{V}}_1) = O(\sum_{i=1}^{k} (\frac{\pi_i^3}{n_i} + \frac{\pi_i}{n_i^3})) = O(\sum_{i=1}^{k} (\frac{\theta_i^3}{n_i} + \frac{\theta_i}{n_i^3}))$. From Lemma 3 and the given condition, we obtain

$$
\frac{\text{Var}(\hat{\mathcal{V}}_1)}{\mathcal{V}_1^2} = O\left(\frac{\sum_{i=1}^{k} \left(\frac{\theta_i^3}{n_i} + \frac{\theta_i}{n_i^3} \right)}{\left(\sum_{i=1}^{k} \left(\theta_i^2 + \frac{1}{N^2} \frac{\theta_i}{n_i} \right) \right)^2} \right) = o(1).
$$

Similarly, we can show, for some constant C',

$$
\frac{\text{Var}(\hat{\mathcal{V}}_{1*})}{(\mathcal{V}_{1*})^2} = O\left(\frac{(\tilde{\theta})^3 \sum_{i=1}^{k} \frac{1}{n_i} + \tilde{\theta} \sum_{i=1}^{k} \frac{1}{n_i^3}}{\left(k(\tilde{\theta})^2 + \frac{\tilde{\theta}}{N^2} \sum_{i=1}^{k} \frac{1}{n_i} \right)^2} \right) = o(1).
$$

$\qquad \square$

F Proof of Theorem 2

Since the condition in Lemma 5 holds, $\hat{\mathcal{V}}_1$ and $\hat{\mathcal{V}}_{1*}$ are the ratio-consistent estimators of $\mathcal{V}_1 = \mathcal{V}_{1*}$ under the H_0. From $\frac{T}{\sqrt{\mathcal{V}_1}} = \frac{T_1 - T_2}{\sqrt{\mathcal{V}_1}}$, we only need to show (i) $\frac{T_1}{\sqrt{\mathcal{V}_1}} \to N(0, 1)$ in distribution and (ii) $\frac{T_2}{\sqrt{\mathcal{V}_1}} \to 0$ in probability. To prove (i), we show the Lyapunov's condition (see Billingsley 1995) for the asymptotic normality is satisfied. In other words, under H_0, we need to show $\frac{\sum_{i=1}^{k} E(T_{1i}^4)}{\mathrm{Var}(T_1)^2} \to 0$. Under H_0, we have $T_{1i} = n_i(\hat{\pi}_i - \pi_i)^2 - d_i \hat{\pi}_i(1 - \hat{\pi}_i)$ with $E(T_{1i}) = 0$; therefore, the Lyapunov's condition is $\sum_{i=1}^{k} E(T_{1i}^4)/\mathrm{Var}(T_1)^2 \to 0$. Using Lemma 4, we have $\sum_{i=1}^{k} E(T_{1i}^4) \leq 2^4(\sum_{i=1}^{k} n_i^4 E(\hat{\pi}_i - \pi_i)^4 + d_i^4 E(\hat{\pi}_i(1 - \hat{\pi}_i))^4) = O(\sum_{i=1}^{k}(\theta_i^4 + \frac{\theta_i}{n_i})) + O(\sum_{i=1}^{k}(\theta_i^4 + \frac{\theta_i}{n_i}^3)) = O((k\theta^4 + \frac{\theta}{\sum_{i=1}^{k} \frac{1}{n_i}}))$ since all $\theta_i = \theta$ under H_0. Combining this with the result 1 in Lemma 3, we have $\frac{\sum_{i=1}^{k} E(T_{1i}^4)}{\mathrm{Var}(T_1)^2} = \frac{O(k\theta^4 + \theta\sum_{i=1}^{k}\frac{1}{n_i})}{(k\theta^2 + \frac{\theta}{N^2}\sum_{i=1}^{k}\frac{1}{n_i})^2} \leq \frac{k\theta^4 + \theta\sum_{i=1}^{k}\frac{1}{n_i}}{k\theta^4} = \frac{1}{k} + \frac{\sum_{i=1}^{k}\frac{1}{n_i}}{k\theta^3} \to 0$ as $k \to \infty$ from the given condition $\frac{\sum_{i=1}^{k}\frac{1}{n_i}}{k\theta^3} \to 0$ which shows $\frac{T_1}{\sqrt{\mathcal{V}_1}} \to N(0, 1)$ in distribution.

Furthermore, from Lemma 3 under the H_0, we have $\mathcal{V}_1 \asymp k\theta^2 + \theta\sum_{i=1}^{k}\frac{1}{n_i}$; therefore, we obtain $E\left(\frac{T_2}{\sqrt{\mathcal{V}_1}}\right) = \frac{E(N(\hat{\bar{\pi}} - \bar{\pi})^2)}{\sqrt{\mathcal{V}_1}} \asymp \frac{\theta}{\sqrt{k\theta^2 + \frac{\theta}{N^2}\sum_{i=1}^{k}\frac{1}{n_i}}} \leq \frac{1}{\sqrt{k}} \to 0$ which leads to $\frac{T_2}{\sqrt{\mathcal{V}_1}} \to 0$ in probability. Combining the asymptotic normality of $\frac{T}{\sqrt{\mathcal{V}_1}}$ with the ratio consistency of $\hat{\mathcal{V}}_1$ and $\hat{\mathcal{V}}_{1*}$, we have the asymptotic normality of $T_{\mathrm{new}1}$ and $T_{\mathrm{new}2}$ under the H_0. □

G Proof of Theorem 3

Since $T = T_1 - T_2$ from (8), we only need to show the following:

(I) $\frac{T_1 - \sum_{i=1}^{k} n_i(\pi_i - \bar{\pi})^2}{\sqrt{\mathrm{Var}(T_1)}} \to N(0, 1)$ in distribution

(II) $\frac{T_2}{\sqrt{\mathrm{Var}(T_1)}} \to 0$ in probability.

For (I), we use the Lyapunov's condition for the asymptotic normality of T_1. We show $\frac{\sum_{i=1}^{k} E(T_{1i} - n_i(\pi_i - \bar{\pi})^2)^4}{\mathrm{Var}(T_1)^2} \to 0$ where $G_i = T_{1i} - n_i(\pi_i - \bar{\pi})^2 = n_i(\hat{\pi}_i - \pi_i)^2 - d_i \hat{\pi}_i(1 - \hat{\pi}_i) + 2n_i(\hat{\pi}_i - \pi_i)(\pi_i - \bar{\pi})$. Using $\sum_{i=1}^{k} E(G_i^4) \leq \sum_{i=1}^{k} \left(n_i^4 E((\hat{\pi}_i - \pi_i)^8) + d_i^4 E((\hat{\pi}_i(1 - \hat{\pi}_i))^4) + 2^4 n_i^4 E(\hat{\pi}_i - \pi_i)^4(\pi_i - \bar{\pi})^4 \right)$. From Lemma 4, we have $n_i^4 E((\hat{\pi}_i - \pi_i)^8) \leq O\left(\theta_i^4 + \frac{\theta_i}{n_i}\right)$, $d_i^4 E((\hat{\pi}_i(1 - \hat{\pi}_i))^4) \leq 2^4\left(\frac{3\theta_i^2}{n_i^2} + \frac{(1 - 6\theta_i)\theta_i}{n_i^3}\right) \leq O(\frac{\theta_i^2}{n_i^2} + \frac{\theta_i}{n_i^3})$ where $O(\cdot)$ is uniform in $1 \leq i \leq k$. Using the result in Lemma 1, we have $2^4 \sum_{i=1}^{k} n_i^4 E(\hat{\pi}_i - \pi_i)^4 \sum_{i=1}^{k}(\pi_i - \bar{\pi})^4 \leq 2^4 \sum_{i=1}^{k} n_i^4(\pi_i - \bar{\pi})^4 \left(\frac{3\theta_i^2}{n_i^2} + \frac{(1 - 6\theta_i)\theta_i}{n_i^3}\right) \leq \max_{1 \leq i \leq k}\left\{n_i(\pi_i - \bar{\pi})^2\left(\theta_i + \frac{1}{n_i}\right)\right\}$

$\sum_{i=1}^{k} n_i (\pi_i - \bar{\pi})^2 \theta_i = \max_{1 \leq i \leq k} \left\{ n_i (\pi_i - \bar{\pi})^2 \left(\theta_i + \frac{1}{n_i} \right) \right\} ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\theta\mathbf{n}}^2$. Therefore, we have

$$\frac{\sum_{i=1}^{k} E(G_i^4)}{\text{Var}(T_1)^2} \leq \frac{\sum_{i=1}^{k} \left(\theta_i^4 + \frac{\theta_i}{n_i} \right) + \max_{1 \leq i \leq k} \left\{ n_i (\pi_i - \bar{\pi})^2 \left(\theta_i + \frac{1}{n_i} \right) \right\} ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\theta\mathbf{n}}^2}{\left(v_1 + ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\theta\mathbf{n}}^2 \right)^2}$$

(26)

$$= \frac{\sum_{i=1}^{k} \left(\theta_i^3 + \frac{\theta_i}{n_i} \right)}{\left(\sum_{i=1}^{k} \left(\theta_i^2 + \frac{\theta_i}{n_i} \right) \right)^2} + \frac{\max_{1 \leq i \leq k} \left\{ n_i (\pi_i - \bar{\pi})^2 \left(\theta_i + \frac{1}{n_i} \right) \right\}}{v + ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\theta\mathbf{n}}^2} \to 0$$

(27)

from the given conditions.

The negligibility of $T_2 = N(\hat{\bar{\pi}} - \bar{\pi})^2$ can be proven by noting that $\frac{N E(\hat{\bar{\pi}} - \bar{\pi})^2}{\sqrt{\text{Var}(T_1)}} = \frac{\bar{\theta}}{\sqrt{\text{Var}(T_1)}} = \frac{1}{N} \frac{\sum_{i=1}^{k} n_i \theta_i}{\sqrt{\text{Var}(T_1)}} \asymp \frac{\max_i \theta_i \sum_{i=1}^{k} n_i}{N \sqrt{\mathcal{V}_1 + ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\theta\mathbf{n}}^2}}$ by (15) from the condition (i). This leads to $\left(\frac{\max_i \theta_i^2}{\mathcal{V}_1 + ||\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}||_{\theta\mathbf{n}}^2} \right)^{1/2} \to 0$ from the condition (ii), so we have $\frac{N(\hat{\bar{\pi}} - \bar{\pi})^2}{\sqrt{\text{Var}(T_1)}} \to 0$ in probability. Combining (I) and (II), we conclude $\frac{T - \sum_{i=1}^{k} n_i (\pi_i - \bar{\pi})^2}{\sqrt{\text{Var}(T_1)}} \to N(0, 1)$ in distribution. □

H Proof of Theorem 4

1. Proof of 1 : We prove $\beta(T_{\text{new2}}) \geq \beta(T_{\text{new1}})$. For this, we only need to show that $\mathcal{V}_1 \geq \mathcal{V}_{1*}$ from Corollary 2. Let $f(x) = 2x^2(1 - x)^2 + \frac{x(1-x)}{n}$, then $f(x)$ is convex for $0 < x < \frac{1}{2} - \frac{1}{\sqrt{3}} \sqrt{1 + \frac{1}{n}}$ since $f''(x) > 0$ for $0 < x < \frac{1}{2} - \frac{1}{\sqrt{3}} \sqrt{1 + \frac{1}{n}}$. Furthermore, $\mathcal{V}_1 = \sum_{i=1}^{k} f(\pi_i)$ and $\mathcal{V}_{1*} = k f(\bar{\pi})$ for $\bar{\pi} = \frac{1}{N} \sum_{i=1}^{k} n_i \pi_i$. From the convexity of f, if $n_i = n$ for all $1 \leq i \leq k$, we have $\frac{1}{k} \mathcal{V}_1 = \frac{1}{k} \sum_{i=1}^{k} f(\pi_i) \geq f(\bar{\pi}) = \frac{1}{k} \mathcal{V}_{1*}$. Therefore, $\mathcal{V}_1 \geq \mathcal{V}_{1*}$ which leads to $\lim_{k \to \infty} (\beta(T_{\text{new2}}) - \beta(T_{\text{new1}})) \geq 0$ for the given $0 < \pi_i < \frac{1}{2} - \frac{1}{\sqrt{3}}$ for all i.

Under the given condition, $\hat{\mathcal{B}}_{0k} = 2k(1 + o_p(1))$ and

$$T_{\text{new2}} = \frac{\sum_{i=1}^{k} n_i (\hat{\pi}_i - \hat{\bar{\pi}})^2 - \sum_{i=1}^{k} \hat{\pi}_i (1 - \hat{\pi}_i)}{\sqrt{2k \hat{\bar{\pi}} (1 - \hat{\bar{\pi}})}} (1 + o_p(1))$$

$$T_\chi = \frac{\sum_{i=1}^{k} n_i (\hat{\pi}_i - \hat{\bar{\pi}})^2 - k \hat{\bar{\pi}} (1 - \hat{\bar{\pi}})}{\sqrt{2k \hat{\bar{\pi}} (1 - \hat{\bar{\pi}})}} (1 + o_p(1))$$

which leads to

$$T_{\text{new2}} - T_\chi = \frac{k\hat{\bar{\pi}}(1 - \hat{\bar{\pi}}) - \sum_{i=1}^{k}\hat{\pi}_i(1 - \hat{\pi}_i)}{\sqrt{2k\hat{\bar{\pi}}(1 - \hat{\bar{\pi}})}}(1 + o_p(1)).$$

Using $k\hat{\bar{\pi}}(1 - \hat{\bar{\pi}}) \geq \sum_{i=1}^{k}\hat{\pi}_i(1 - \hat{\pi}_i)$, $\lim_{k\to\infty} P(T_{\text{new2}} - T_\chi \geq 0) \to 1$ which leads to $\lim_{k\to\infty}(\beta(T_{\text{new2}}) - \beta(T_\chi)) \geq 0$.

2. Proof of 2: Note that $\mathcal{A}_{1i} = 2(1 + o(1))$ and $\mathcal{A}_{2i} = 4(1 + o(1))$ where $o(1)$ is uniform in i. Using $\bar{\pi} = (k^{-\gamma} + \delta k^{\alpha-1})(1 + O(k^{-1}))$ and $\tilde{\theta} = \bar{\pi}(1 + o(1))$, we obtain

$$\mathcal{V}_1 = \left(2\sum_{i=1}^{k}\theta_i^2 + 4\sum_{i=1}^{k}\frac{\theta_i}{n_i}\right)(1 + o(1))$$

$$= \left(2(k-1)k^{-2\gamma} + 2(k^{-\gamma} + \delta)^2 + \frac{(k-1)k^{-\gamma}}{n} + \frac{k^{-\gamma} + \delta}{nk^\alpha}\right)$$

$$= \left(2k^{1-2\gamma} + 2\delta^2 + \frac{4k^{1-\gamma}}{n}\right)(1 + o(1))$$

$$\mathcal{V}_{1*} = 2k(k^{-\gamma} + \delta k^{\alpha-1})^2(1 + O(k^{-1})) + 4\tilde{\theta}\sum_{i=1}^{k}\frac{1}{n_i}$$

$$= 2k^{1-2\gamma} + 4\delta k^{\alpha-\gamma} + 2\delta^2 k^{2\alpha-1}$$
$$\quad + 4(k^{-\gamma} + \delta k^{\alpha-1})\left(\frac{k-1}{n} + \frac{1}{nk^\alpha}\right)(1 + o(1))$$

$$= 2k^{1-2\gamma} + 4\delta k^{\alpha-\gamma} + 2\delta^2 k^{2\alpha-1} + 4\frac{k^{1-\gamma} + \delta k^\alpha}{n}(1 + o(1))$$

so

$$\frac{\mathcal{V}_{1*} - \mathcal{V}_1}{\mathcal{V}_1} = \frac{(2\delta k^{\alpha-\gamma} + \delta^2(k^{2\alpha-1} - 1)) + 2\frac{\delta k^\alpha}{n}(1 + o(1))}{k^{1-2\gamma} + 2\delta^2 + 2\frac{k^{1-\gamma}}{n}(1 + o(1))}. \tag{28}$$

(a) if $\alpha + \gamma < 1$ and $\alpha \geq \frac{1}{2}$, then $k^{\alpha-\gamma} = o(k^{2\alpha-1})$, therefore (28) $= \frac{\delta^2 k^{2\alpha-1}I(\alpha\neq\frac{1}{2})+2\frac{k^\alpha}{n}}{k^{1-2\gamma}+\delta^2+2\frac{k^{1-\gamma}}{n}} \to 0$ where $I(\cdot)$ is an indicator function.

(b) if $\alpha + \gamma < 1$, $\alpha < \frac{1}{2}$ and $\alpha \geq \gamma$, then (28) $= \frac{2\delta k^{\alpha-\gamma}-\delta^2+2\frac{k^\alpha}{n}}{k^{1-2\gamma}+\delta^2+2\frac{k^{1-\gamma}}{n}} \to 0$.

(c) if $\alpha + \gamma < 1$, $\alpha < \frac{1}{2}$, $\gamma \leq \frac{1}{2}$ and $\alpha < \gamma$, then (28)$= \frac{-\delta^2+2\frac{k^\alpha}{n}}{k^{1-2\gamma}+\delta^2+2\frac{k^{1-\gamma}}{n}} \to 0$.

(d) if $\alpha + \gamma < 1$, $\alpha < \frac{1}{2}$ and $\gamma > \frac{1}{2}$, then there are two cases depending on the behavior of n. When $\frac{k^{1-\gamma}}{n} \to 0$, then (28) $\to \frac{-\delta^2}{\delta^2} = -1$. When $\frac{k^{1-\gamma}}{n} \to \infty$, (28) $= \frac{\frac{k^\alpha}{n}}{\frac{k^{1-\gamma}}{n}}(1 + o(1)) = k^{\alpha+\gamma-1} \to 0$.

(e) if $\alpha + \gamma > 1$, $\alpha > \frac{1}{2}$ and $\gamma < \frac{1}{2}$, then (28) $= \frac{\delta^2 k^{2\alpha-1} + 2\frac{k^\alpha}{n}}{k^{1-2\gamma} + 2\frac{k^{1-\gamma}}{n}}(1 + o(1)) \to \infty$.

(f) if $\alpha + \gamma > 1$, $\alpha > \frac{1}{2}$ and $\gamma \geq \frac{1}{2}$, then (28) $= \frac{k^{\alpha-\gamma} + \delta^2 k^{2\alpha-1} + 2\frac{k^\alpha}{n}}{I(\gamma=\frac{1}{2}) + \delta^2 + 2\frac{k^{1-\gamma}}{n}}(1 + o(1)) \to \infty$.

(g) if $\alpha + \gamma > 1$, $\alpha < \frac{1}{2}$ and $\gamma > \frac{1}{2}$, then $\alpha < \gamma$ and (28) $= \frac{-\delta^2 + 2\frac{k^\alpha}{n}}{\delta^2 + 2\frac{k^{1-\gamma}}{n}}(1 + o(1)) =$

$\frac{-\delta^2 + \frac{k^\alpha}{n}}{\delta^2 + 2\frac{k^{1-\gamma}}{n}}(1 + o(1))$. There are two situations depending n. When $\frac{k^\alpha}{n} \to \infty$,

(28) $= \frac{-\delta^2 + \frac{k^\alpha}{n}}{2\delta^2 + \frac{k^{1-\gamma}}{n}}(1 + o(1)) \to \infty$. When $\frac{k^\alpha}{n} \to 0$, we have $\frac{k^{1-\gamma}}{n} \to 0$, so we

derive (28) $= \frac{-\delta^2}{\delta^2}(1 + o(1)) \to -1$.

In $(a) \cup (b) \cup (c) = \{(\alpha, \gamma) : 0 < \alpha < 1, 0 < \gamma < 1, 0 < \alpha + \gamma < 1, 0 < \gamma \leq \frac{1}{2}\}$, we have $\lim_n \frac{\mathcal{V}_{1*}}{\mathcal{V}_1} = 1$ leading to $\lim_n (\beta(T_{\text{new1}}) - \beta(T_{\text{new1}})) = 0$. In $(e) \cup (f) = \{(\alpha, \gamma) : 0 < \alpha < 1, 0 < \gamma < 1, \alpha + \gamma > 1, 1 > \alpha > \frac{1}{2}\}$, we have $\lim \frac{\mathcal{V}_{1*}}{\mathcal{V}} > 1$ which leads to $\lim_n (\beta(T_{\text{new1}}) - \beta(T_{\text{new2}})) > 0$.

In (e) and (g), the performances are different depending on the sample sizes.

3. We first have

$$\mathcal{V}_1 = 2(k^{-\gamma} + \delta)^2 + 2(k-1)k^{-2\gamma} + \frac{4(\delta + k^{-\gamma})}{n} + 4(k-1)\frac{k^{-\gamma}}{nk^\alpha}$$

$$= \left(2\delta^2 + 2k^{1-2\gamma} + \frac{4k^{1-\gamma-\alpha}}{n}\right)(1 + o(1)).$$

Since $\tilde{\theta} = \bar{\pi}(1 - \bar{\pi}) = \frac{\delta + k^{\alpha-\gamma+1}}{k^{\alpha+1}}(1 + o(1)) = k^{-\gamma}(1 + o(1))$ from $0 < \alpha < 1$ and $0 < \gamma < 1$,

$$\mathcal{V}_1^* = 2k^{1-2\gamma} + \frac{4k^{-\gamma}}{n} + \frac{(k-1)k^{-\gamma}}{nk^\alpha}$$

$$= \left(2k^{1-2\gamma} + \frac{4k^{1-\gamma-\alpha}}{n}\right)(1 + o(1)).$$

If $1 - 2\gamma < 0$ and $k^{1-\gamma-\alpha} = o(n)$, then $\mathcal{V}_1 = \delta^2(1 + o(1))$ and $\mathcal{V}_1^* = o(1)$, we have $\frac{\mathcal{V}_1}{\mathcal{V}_1^*} \to \infty$ which leads to $\beta(T_{\text{new2}}) - \beta(T_{\text{new1}}) > 0$. $\qquad\square$

References

Bathke, A. C., Harrar, S. W. (2008). Nonparametric methods in multivariate factorial designs for large number of factor levels. *Journal of Statistical Planning and Inference*, *138*(3), 588–610.

Bathke, A., Lankowski, D. (2005). Rank procedures for a large number of treatments. *Journal of Statistical Planning and Inference*, *133*(2), 223–238.

Billingsley, P. (1995). *Probability and Measure* (3rd ed.). Hoboken: Wiley.

Boos, D. D., Brownie, C. (1995). ANOVA and rank tests when the number of treatments is large. *Statistics and Probability Letters*, *23*, 183–191.

Cai, T., Parast, L., Ryan, L. (2010). Meta-analysis for rare events. *Statistics in Medicine*, *29*(20), 2078–2089.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, *10*, 417–451.

Greenshtein, E., Ritov, R. (2004). Persistence in high-dimensional linear predictor selection and the virtue of over parametrization. *Bernoulli*, *10*, 971–988.

Nissen, S. E., Wolski, K. (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine*, *356*(24), 2457–2471.

Park, J. (2009). Independent rule in classification of multivariate binary data. *Journal of Multivariate Analysis*, *100*, 2270–2286.

Park, J., Ghosh, J. K. (2007). Persistence of the plug-in rule in classification of high dimensional multivariate binary data. *Journal of Statistical Planning and Inference*, *137*, 3687–3705.

Potthoff, R. F., Whittinghill, M. (1966). Testing for homogeneity: I. The binomial and multinomial distributions. *Biometrika*, *53*, 167–182.

Shuster, J. J. (2010). Empirical versus natural weighting in random effects meta analysis. *Statistics in Medicine*, *29*, 1259–1265.

Shuster, J. J., Jones, L. S., Salmon, D. A. (2007). Fixed vs random effects meta-analysis in rare event studies: The rosiglitazone link with myocardial infarction and cardiac death. *Statistics in Medicine*, *26*, 4375–4385.

Stijnen, T., Hamza, Taye H., Zdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine*, *29*, 3046–3067.

Tian, L., Cai, T., Pfeffer, M. A., Piankov, N., Cremieux, P. Y., Wei, L. J. (2009). Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2×2 tables with all available data but without artificial continuity correction. *Biostatistics*, *10*, 275–281.