

Hypothesis Testing About Proportions in Two Finite Populations

K. KRISHNAMOORTHY and Jessica THOMSON

The problem of hypothesis testing about proportions in two finite populations is addressed. The usual test based on the normal approximation (Z test) and a test based on estimated p values (E test) are considered. The exact properties of the tests are evaluated numerically. Numerical studies indicate that the E test is very satisfactory even for small samples and can be recommended for practical use. Power computation for a given level and sample size is also outlined for both tests. The testing methods are illustrated using an example.

KEY WORDS: Binomial distribution; Exact power; Exact size; Hypergeometric distribution; Sample size calculation; Standardized difference.

1. INTRODUCTION

There has been a continuous interest in developing small sample exact inferential procedures for discrete distributions. The growing interest in this area is partly because modern computing technology allows researchers to study the exact properties of their new methods. In particular, binomial distribution, the one-sample case as well as the two-sample case, received considerable attention in the past few decades (see Chan and Zhang 1999; Berger 1996; Storer and Kim 1990 and the references therein). Classical as well as some numerically intensive methods have been evaluated for their merit in terms of size and power. However, only limited results are available for making inferences about proportions in finite populations. There are articles that address the problem of interval estimation of finite population proportion for the one-sample case (see Katz 1953; Burstein 1975; Buonaccorsi 1987; and the most recent article by Wendell and Schmee 2001). However, not much work has been done on inferential procedures for the two-sample case in finite populations. Li and Chen (1999) considered the usual test based on normal approximation, and provided an approximation to determine the sample sizes for a given level and power.

Recall that in infinite population case we deal with a binomial distribution, and in the finite population case we deal with a hypergeometric distribution. Even though these two distributions are closely related, results based on binomial distributions are applicable to hypergeometric distributions only when the lot sizes are really large. For instance, in the one-sample case, Burstein (1975) noted that the differences between the binomial-

based confidence limits with finite population correction and the ones based on the exact hypergeometric probabilities are appreciable if the population size is less than 5,000. In general, binomial-based results are not accurate for making inferences about proportions in finite populations when the populations are not vastly larger than the samples. Thus, when binomial-based methods are applied to finite populations their validity should be carefully evaluated. The hypergeometric-based method is appropriate in some practical situations where one wants to compare two shipments of items for the proportions of defective items, to evaluate two production processes with respect to defective rates, and to compare proportions of households without health insurance in two counties. In such situations population sizes are not usually large enough to be considered as infinite. Li and Chen (1999) gave three practical examples, and we give an illustrative example in Section 6.

We shall now formulate the problem. Consider a finite population of N units of which M units possess a specified characteristic of interest. Furthermore, let X denote the number of units with the characteristic in a random sample of n units drawn (without replacement) from the population. Then, it is well-known that X follows a hypergeometric distribution with (in traditional terminologies) lot size N , the number of defective units M , and the sample size n . We simply write $X \sim H(n, M, N)$. The probability mass function of X will be denoted by

$$h(x; n, M, N) = P(X = x) = \frac{\binom{n}{x} \binom{N-M}{n-x}}{\binom{N}{n}}, \quad L \leq x \leq U, \quad (1)$$

where

$$L = \max\{0, M - N + n\} \quad \text{and} \quad U = \min\{n, M\}. \quad (2)$$

Let X_1 and X_2 be independent hypergeometric random variables with $X_1 \sim H(n_1, M_1, N_1)$ and $X_2 \sim H(n_2, M_2, N_2)$. Further, let k_1 and k_2 denote, respectively, the observed values of X_1 and X_2 and let $p_1 = M_1/N_1$ and $p_2 = M_2/N_2$. We are interested in testing

$$H_0 : p_1 = p_2 \quad \text{vs.} \quad H_a : p_1 \neq p_2, \quad (3)$$

based on (k_1, n_1, N_1) and (k_2, n_2, N_2) .

The following section considers two test procedures. The first procedure is based on the normal approximation to the distribution of the standardized difference of the sample proportions, and is referred to as the Z test. The second procedure is obtained by suitably modifying the binomial method due to Storer and Kim (1990). Their approach is based on the estimated p values of the standardized difference. For this reason and for convenience, we refer the test based on Storer and Kim's method as the E

K. Krishnamoorthy is Professor, and Jessica Thomson is Graduate Student, Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504-1010 (E-mail: krishna@louisiana.edu). We thank two reviewers for their useful comments and suggestions.

test. We also show that the E test is essentially equivalent to the one based on the parametric bootstrap approach (see Remark 2). Section 3 outlines a numerical method for computing the exact sizes (for a given p_1 and p_2 such that $p_1 - p_2 = 0$, and values of the lot sizes and sample sizes, the actual probability of rejecting H_0 in (3) is called the size of the test) and powers (the probability of rejecting H_0 when $p_1 \neq p_2$) of the proposed tests. Using this numerical method, Section 4 evaluates the sizes and powers of these two tests for various parameter configurations. We found from our numerical results that the E test is almost exact because its size seldom exceeds the nominal level. On the other hand, the Z test exceeds the nominal level by considerable amounts if the lot sizes are small and/or the population proportions are small. The sample sizes required by the Z test and the E test to attain a given power are compared in Section 5. The methods considered in this article are illustrated using an example in Section 6. In general, the E test is preferable to the Z test for practical use because it controls the Type I error rate much better than does the Z test. Some concluding remarks are given in Section 7.

2. PROPOSED TESTS

Consider the pivot statistic

$$Z_{X_1, X_2} = \frac{X_1/n_1 - X_2/n_2}{\sqrt{V_{X_1, X_2}}}, \quad (4)$$

where V_{X_1, X_2} is the variance estimate of $X_1/n_1 - X_2/n_2$ under $H_0 : p_1 = p_2$, and is given by

$$V_{X_1, X_2} = \left(\frac{N_1 - n_1}{n_1(N_1 - 1)} + \frac{N_2 - n_2}{n_2(N_2 - 1)} \right) \times \left(\frac{X_1 + X_2}{n_1 + n_2} \right) \left(1 - \frac{X_1 + X_2}{n_1 + n_2} \right). \quad (5)$$

Let k_1 and k_2 denote, respectively, the observed values of X_1 and X_2 , and let Z_{k_1, k_2} denote the observed value of Z_{X_1, X_2} . The value of Z_{k_1, k_2} can be computed using (4) and (5) with (X_1, X_2) replaced by (k_1, k_2) .

2.1 Z Test

It follows from the central limit theorem that $Z_{X_1, X_2} \sim N(0, 1)$ approximately when H_0 holds, where $N(0, 1)$ denotes the standard normal distribution. The Z test rejects H_0 in (3) whenever the p value

$$P(|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}| | H_0) = 2(1 - \Phi(|Z_{k_1, k_2}|)), \quad (6)$$

where $|\cdot|$ denotes the absolute value and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal random variable, is less than the nominal level α . One-sided tests can be carried out similarly. For example, when $H_a : p_1 > p_2$ the null hypothesis will be rejected whenever $P(Z_{X_1, X_2} \geq Z_{k_1, k_2}) = 1 - \Phi(Z_{k_1, k_2}) < \alpha$. This Z test has size approximately equal to α when sample sizes are large.

2.2 E Test

Using an indicator function $I(\cdot)$, the true p value for testing hypotheses in (3) can be expressed as

$$\begin{aligned} P(|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}| | H_0) &= E_{X_1, X_2} (I(|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}|) | H_0) \\ &= \sum_{x_1=L_1}^{U_1} \sum_{x_2=L_2}^{U_2} h(x_1; n_1, N_1 p, N_1) h(x_2; n_2, N_2 p, N_2) \\ &\quad \times I(|Z_{x_1, x_2}| \geq |Z_{k_1, k_2}|), \end{aligned} \quad (7)$$

where E_{X_1, X_2} denotes the expectation under the joint distribution of (X_1, X_2) , and p is the unknown common proportion under H_0 in (3). The p value in (7) cannot be computed because it depends on the unknown nuisance parameter p . An option suggested by Suissa and Shuster (1985) is to reject the H_0 in (3) whenever the maximum of (7) over $0 < p < 1$ is less than α . This test is exact in the sense that its size never exceeds the nominal level. However, the test is computationally involved, and enormous computation is needed for power calculation (see Section 3). Another approach, due to Liddell (1978) and later followed by Storer and Kim (1990), is to reject the null hypothesis if an estimated p value is less than α . If we use this latter approach, then a natural estimate of (7) is given by

$$\begin{aligned} P(|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}| | H_0) &= \sum_{x_1=L_{x_1}}^{U_{x_1}} \sum_{x_2=L_{x_2}}^{U_{x_2}} h(x_1; n_1, \hat{M}_1, N_1) h(x_2; n_2, \hat{M}_2, N_2) \\ &\quad \times I(|Z_{x_1, x_2}| \geq |Z_{k_1, k_2}|), \end{aligned} \quad (8)$$

where $\hat{p} = (k_1 + k_2)/(n_1 + n_2)$, $\hat{M}_i = [N_i \hat{p}]$, $L_{x_i} = \max\{0, \hat{M}_i - N_i + n_i\}$, $U_{x_i} = \min\{n_i, \hat{M}_i\}$, $i = 1, 2$, and $[x]$ denotes the integer part of x . Both terms Z_{x_1, x_2} and Z_{k_1, k_2} are defined to be zero if their numerators are zeros. The test rule is to reject H_0 in (3) whenever (8) is less than α .

Remark 1. For the binomial case, Liddell (1978) suggested using $Z_{x_1, x_2} = x_1/n_1 - x_2/n_2$ and $Z_{k_1, k_2} = k_1/n_1 - k_2/n_2$ in (8), and Storer and Kim (1990) considered the standardized difference in sample proportions similar to the one in (4). We here consider only Storer and Kim's pivot because for the binomial case these authors showed that the results based on the standardized difference are, in general, superior to the those based on just the difference between the sample proportions.

Remark 2. The proposed E test is essentially equal to the one based on the parametric bootstrap approach (see Efron 1982, p. 29). In a parametric bootstrap approach the p value of a test is obtained by using the distribution of the Monte Carlo samples generated from $H(n_1, \hat{M}_1, N_1)$ and $H(n_2, \hat{M}_2, N_2)$. In particular, in parametric bootstrap approach, the p value is estimated by the proportion of simulated samples for which $Z_{X_1, X_2} \geq Z_{k_1, k_2}$ where as in (8) we use the exact sampling distribution of Z_{X_1, X_2} based on $H(n_1, \hat{M}_1, N_1)$ and $H(n_2, \hat{M}_2, N_2)$. Therefore, the present approach is equivalent to the parametric bootstrap approach applied in an exact manner. Furthermore, it is known that the results of a pivot based parametric bootstrap approach (percentile t) are in general preferred to those based on non-pivot based approach (see Hall 1992, p. 141). This perhaps explains as to why Kim and Storer's (1990) pivot-based binomial test is superior to the one based on just the difference $X_1/n_1 - X_2/n_2$. See Remark 1. For an application of the PB method for construct-

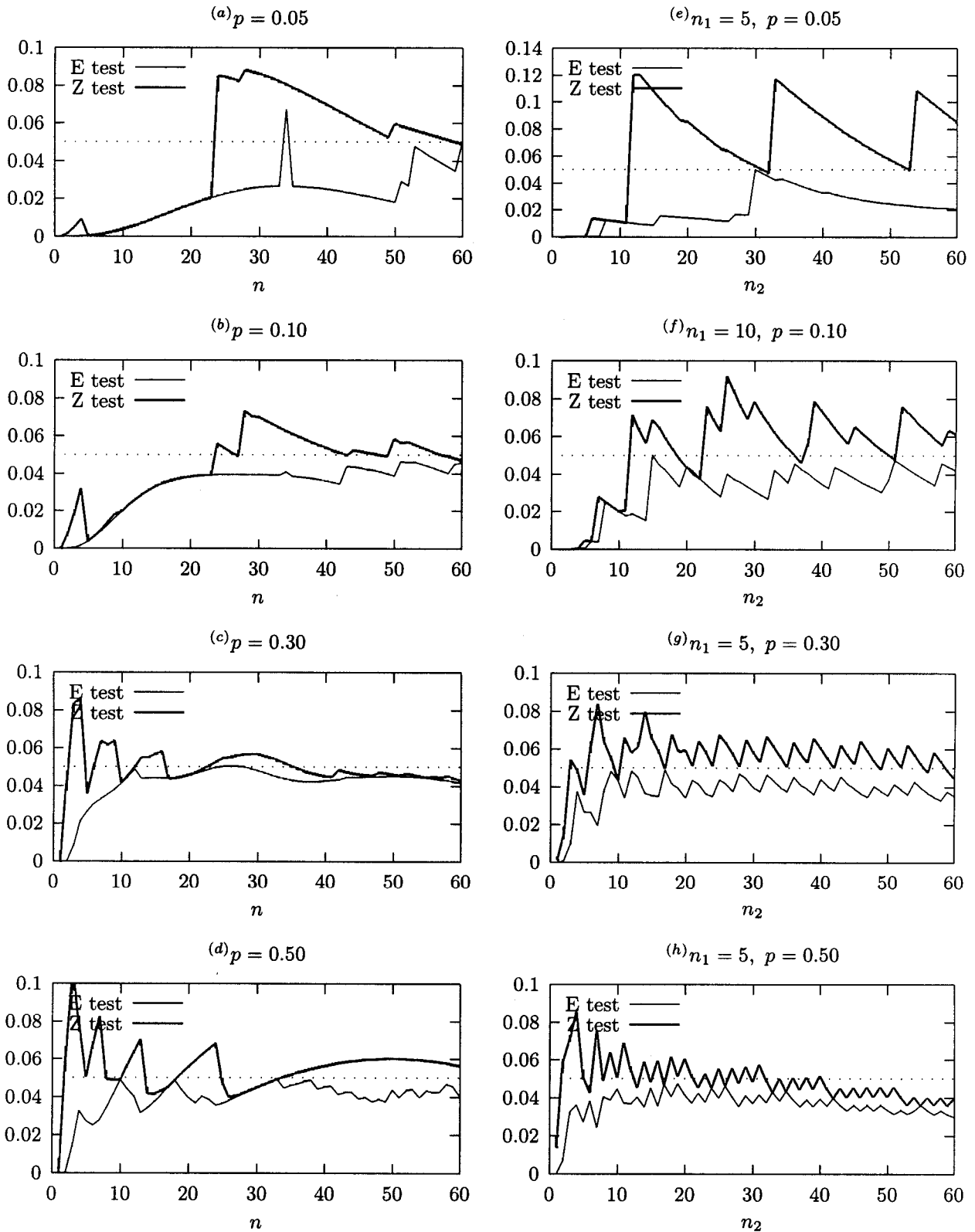


Figure 1. Exact Size of the Tests as a Function of the Sample Sizes at the Nominal Level $\alpha = 0.05$; $N_1 = N_2 = 100$.

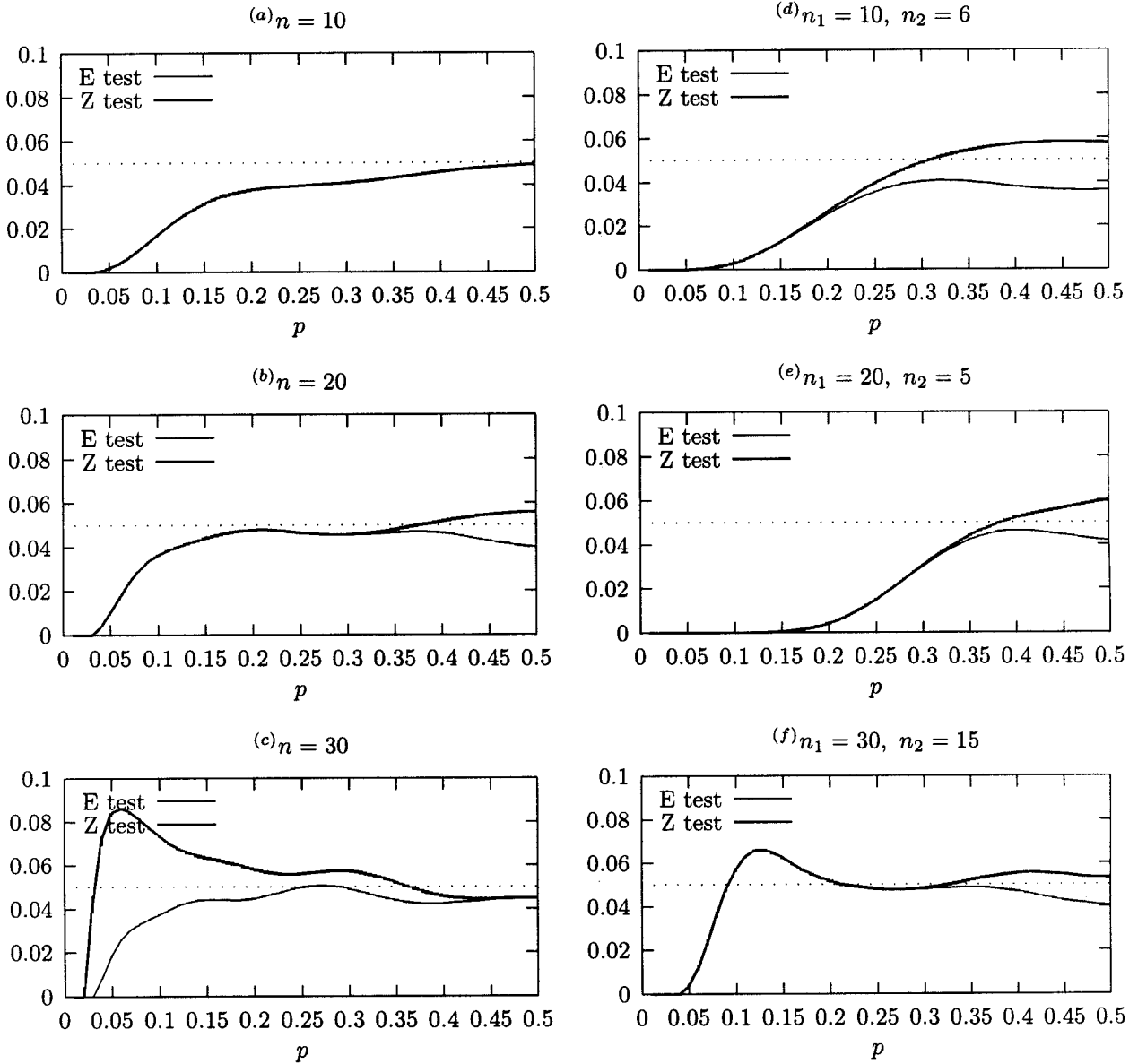


Figure 2. Exact Size of the Tests as a Function of $p = M_1/N_1 = M_2/N_2$ at the Nominal Level $\alpha = 0.05$; $N_1 = N_2 = 100$.

ing confidence intervals for the mean of a lognormal distribution, see Angus (1994).

3. POWER CALCULATION

The exact powers of the proposed tests can be computed using hypergeometric probabilities. For a given p_i, n_i and N_i , letting $M_i = [N_i p_i]$, $i = 1, 2$, where $[x]$ denotes the integer part of x , the power of the Z test when $H_a : p_1 \neq p_2$ can be expressed as

$$\sum_{k_1=L_1}^{U_1} \sum_{k_2=L_2}^{U_2} h(k_1; n_1, M_1, N_1) \times h(k_2; n_2, M_2, N_2) I(|Z_{k_1, k_2}| > z_{1-\alpha/2}), \quad (9)$$

where $L_i = \max\{0, M_i - N_i + n_i\}$, $U_i = \min\{n_i, M_i\}$, $i = 1, 2$, and z_p denotes the 100 p th percentile of the standard normal distribution. For the definition of power function, see Casella and Berger (2002, p. 383). Letting $\hat{M}_i = [N_i \hat{p}]$, $i = 1, 2$ the power

of the E test when $H_a : p_1 \neq p_2$ can be written as

$$\sum_{k_1=L_1}^{U_1} \sum_{k_2=L_2}^{U_2} h(k_1; n_1, M_1, N_1) h(k_2; n_2, M_2, N_2) \times I \left(\sum_{x_1=L_{x_1}}^{U_{x_1}} \sum_{x_2=L_{x_2}}^{U_{x_2}} h(x_1; n_1, \hat{M}_1, N_1) \times h(x_2; n_2, \hat{M}_2, N_2) I(|Z_{x_1, x_2}| \geq |Z_{k_1, k_2}|) \leq \alpha \right), \quad (10)$$

where L_i and U_i are defined below display (9), and \hat{M}_i, L_{x_i} , and U_{x_i} are defined below display (8). Recall that the power of a test is the probability of rejecting the null hypothesis when $p_1 \neq p_2$. Therefore, when $p_1 = p_2$, the expressions (9) and (10) give the sizes of the tests. For smaller sample sizes, the power can be computed in a straightforward manner by evaluating each sum from L to U . However, for moderately large samples (e.g., $n_1 \geq 20$ and $n_2 \geq 20$) this approach is very time consuming.

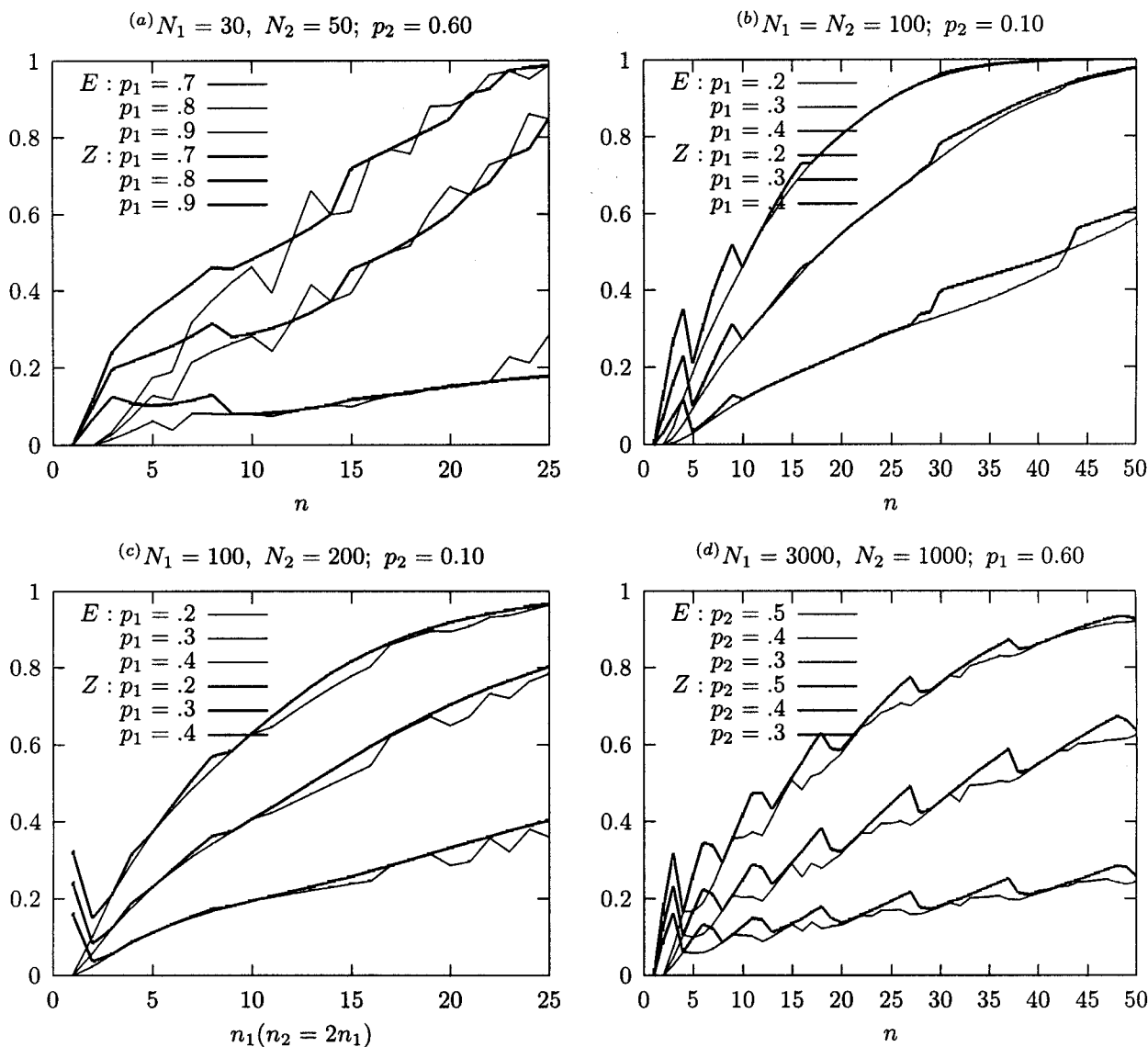


Figure 3. Actual Power of the Tests E and Z as a Function of the Sample Sizes at the Nominal Level $\alpha = 0.05$.

A computationally faster method is to evaluate each sum in (10) by first evaluating the term at the mode of the hypergeometric distribution, and then computing other terms using forward and backward recurrence relations for hypergeometric probabilities. For each sum, the evaluation process may be stopped once the hypergeometric probability becomes very small, say, 10^{-7} . For ease of reference, we note that for $X \sim H(n, M, N)$, the mode of X is given by $[(n + 1)(M + 1)/(N + 2)]$,

$$P(X = k + 1) = \frac{(n - k)(M - k)}{(k + 1)(N - M - n + k + 1)} P(X = k),$$

$$k = L, \dots, U - 1,$$

and

$$P(X = k - 1) = \frac{k(N - M - n + k)}{(n - k + 1)(M - k + 1)} P(X = k),$$

$$k = L + 1, \dots, U,$$

where L and U are given in (2). Fortran programs based on this method for computing p values and powers of the E test, and DOS executable files derived from these programs will be posted at the Web site <http://lib.stat.cmu.edu/TAS/>.

4. SIZE AND POWER STUDIES

The Z test and the E test must be evaluated for their validity because they are approximate tests. We computed the sizes and powers of both tests and plotted them in Figures 1–3. Figures 1(a–h) give the size of the tests as a function of the sample sizes when lot sizes $N_1 = N_2 = 100$. The sizes of the tests are given in Figures 1(a–d) for equal sample sizes and in Figures 1(e–h) for unequal sample sizes at the nominal level 0.05. We observe from these figures that the sizes of the E test exceeded the nominal level at only one place (Figure 1(a)) whereas the Z test exceeds the nominal level often by a considerable amount. For smaller values of p , the size of the Z test goes up to 0.12 (Figure 1(e)) whereas the the maximum size of the E test goes up to 0.07 (Figure 1(a)). In Figures 2(a–f) the sizes are plotted against the common proportion p while the sample sizes are fixed. Here, both tests exhibit almost similar properties except in Figure 2(c). The sizes of the E test never exceeded the nominal level 0.05 for the parameter configurations and sample sizes considered in Figures 2(a–f). We also plotted sizes of the tests for unequal lot sizes (they are not reported here; see Thomson 2002). The patterns that we observed in unequal lot sizes are similar to

Table 1a. Comparison of Sample Sizes for Specified Powers ($H_a : p_1 > p_2; n_1 = n_2 = n; N_1 = N_2 = N$)

Level	Power	p_1	p_2	N	Sample size n		
					Z test	Li-Chen*	E test
0.01	0.90	0.20	0.10	100	77(.010)	77	79(.008)
				500	198(.010)	199	199(.010)
				1000	246(.010)	249	247(.010)
				5000	305(.010)	310	305(.010)
0.05	0.80	0.60	0.50	100	76(.064)	76	77(.053)
				500	189(.048)	190	190(.049)
				1000	234(.048)	234	234(.048)
				5000	293(.051)	288	293(.050)
0.05	0.95	0.05	0.01	100	79(.022)	80	79(.022)
				500	201(.051)	220	215(.046)
				1000	261(.052)	282	266(.048)
				5000	321(.051)	363	331(.049)
0.01	0.95	0.010	0.001	500	328(.009)	405	251(.029)
				1000	667(.013)	681	668(.008)
				5000	1301(.010)	1493 ¹	1311(.010)
				5000	1301(.010)	1493 ¹	1311(.010)
0.10	0.90 ¹	0.005	0.001	500	342(.089)	386	342(.089)
				1000	719(.142)	628	719(.093)
				5000	1506(.110)	1261	1648(.093)
				5000	1506(.110)	1261	1648(.093)

*Sample sizes based on Li and Chen's (1999) approximate method for the Z test.

¹ The reported numbers in Li and Chen's article are in error.

those of equal lot sizes. It is clear that the E test performs better than the Z test in controlling Type I error rates in all the cases. The powers of the tests are given in Figures 3(a-d) for lot sizes

varying from 30 to 3,000. We see from these figures that the Z test has slightly more power than the E test. However, we cannot conclude that the Z test is more powerful than the E test,

Table 1b. Comparison of Sample Sizes Required for a Given Level and Power. ($H_a : p_1 \neq p_2; n_1 = n_2 = n; N_1 = N_2 = N; \alpha = 0.05$)

p_1	p_2	Power	Sample size n			p_1	p_2	Power	Sample Size n		
			N	Z test	E test				N	Z test	E test
.05	.25	.80	100	31(.053)	34(.051)*	.15	.45	.80	100	26(.062)	27(.040)
			500	43(.039)	44(.040)				500	33(.054)	34(.046)
			1000	44(.044)	44(.037)				1000	34(.050)	35(.042)
		5000	44(.047)	46(.041) ¹	5000			35(.052)	36(.046) ¹		
		.90	100	38(.052)	39(.047)			100	32(.057)	32(.037)	
			500	55(.046)	55(.046)			500	43(.047)	44(.042)	
	1000		56(.052)	59(.040)	1000	45(.049)	45(.048)				
	.05	.35	.80	5000	60(.045)	61(.045) ¹	5000	46(.054)	47(.045) ¹		
				100	19(.031)	21(.004)	100	18(.076)	19(.033)		
				500	24(.015)	24(.015)	500	20(.045)	22(.047)		
			.90	1000	24(.017)	25(.018)	1000	21(.048)	22(.049)		
				5000	24(.018)	27(.022) ¹	5000	21(.050)	22(.036) ¹		
100				25(.045)	27(.048)	100	22(.070)	22(.036)			
.05	.45	.80	500	30(.024)	32(.027)	.25	.55	.80	500	26(.050)	27(.046)
			1000	32(.029)	33(.028)				1000	27(.053)	28(.038)
			5000	32(.031)	33(.030) ¹				5000	28(.052)	29(.040) ¹
		.90	100	13(.015)	15(.001)			100	30(.054)	31(.042)	
			500	16(.005)	16(.005)			500	38(.051)	39(.050)	
			1000	16(.006)	16(.007)			1000	40(.051)	40(.044)	
	.05	.45	.80	5000	16(.006)	17(.006) ¹	5000	40(.049)	41(.050) ¹		
				100	17(.026)	18(.005)	100	36(.057)	36(.048)		
				500	21(.011)	21(.011)	500	50(.053)	50(.046)		
			.90	1000	21(.012)	21(.012)	1000	52(.052)	52(.048)		
				5000	21(.013)	21(.013) ¹	5000	53(.051)	54(.048) ²		

* The numbers in parentheses are the sizes of the tests;

¹ the sample sizes are in agreement with those for the binomial test reported in Table 2 of Storer and Kim (1990);

² the sample size is one unit less than that reported in Storer and Kim's paper.

because the Type I error rates of the Z test are in general higher than the nominal level. On the basis of controlling Type I error rate, the E test is certainly preferable to the Z test.

5. COMPARISON OF SAMPLE SIZES

We computed the sample sizes required for a given level and power for both tests in two setups. The first setup is chosen so that we can compare our exact method of calculating sample sizes for the Z test with the approximate method given by Li and Chen (1999). For this purpose, we took the configurations of Table I of Li and Chen (1999) and computed the sample sizes. These sample sizes are given in Table 1a. In the second setup, we want to demonstrate the advantage of using our finite population approach over the binomial approach due to Storer and Kim (1990). For this purpose, we chose the configurations of Table 2 of Storer and Kim's article, and computed sample sizes for both Z test and E test. These results are given in Table 1b.

In Table 1a, comparing the exact sample sizes for the Z test with the sample sizes computed using Li and Chen's (1999) approach, we see that Li and Chen's approximate method is satisfactory only when the proportions are large. For other cases, the differences between the exact and approximate sample sizes are quite significant. There is not much difference between the sample sizes required for the Z test and the E test when the proportions p_1 and p_2 are large, however, the differences are appreciable if the proportions are small. We also observe that the sizes of the Z test exceed the nominal level especially when the lot sizes are small.

Comparison of the sample sizes in Table 1b with those in Table 2 of Storer and Kim (1990) indicates that the sample sizes are the same for the binomial and hypergeometric distributions provided both lot sizes are at least 5,000. Storer and Kim's table values are not reported here because the sample sizes required for the E test and those required for the binomial test are the same when $N_1 = N_2 = 5000$ except for one case (see the footnote of Table 1b). For example, when $p_1 = 0.05$, $p_2 = 0.25$, $N_1 = N_2 = 100$, a sample of size 39 is required from each population for the E test (hypergeometric based), whereas the binomial based test requires a sample of size 61 from each population (see the sample size when $N_1 = N_2 = 5000$). This clearly indicates the advantage of our finite population method over the infinite population method. We further observe that the sizes of E test are always close to or smaller than the nominal level of 0.05 for all the cases considered whereas the sizes of the Z test exceeds the nominal level in many situations.

6. AN EXAMPLE

Suppose that the quality control engineer of a canning company wishes to determine if the number of nonacceptable cans differs between two canning machines. A can is determined to be nonacceptable (for sale) if the content of the can weighs less than 95% of the labeled weight. Denote p_1 as the proportion of nonacceptable cans for the first machine and p_2 as the proportion for the second machine. Two pallets produced from each machine, each containing $N_1 = N_2 = 250$ cans, are arbitrarily selected. The engineer, based on his own experience, suspects that $p_1 = 0.06$ and $p_2 = 0.02$. Based on these speculated pro-

portions, it is desired to determine the sample size required from each pallet for testing $H_0 : p_1 = p_2$ versus $H_a : p_1 \neq p_2$ at the level of 0.05 with power 0.70. The E test of this article requires that $n_1 = n_2 = 137$ (with size = 0.045 and power = 0.7034); the binomial test (ignores the sizes of the lots) due to Storer and Kim (1990) requires that $n_1 = n_2 = 284$ (size = 0.049 and power = 0.7008); the Z test requires $n_1 = n_2 = 136$ (with size = 0.060 and power = 0.7074). The results are in agreement with our conclusions based on the numerical studies in Section 5. They are also intuitive because the binomial test not only ignores the sizes of the population but also the sampling distribution due to the selection of items without replacement, and as a result it needs more samples (in this case more than the population size) than those required by the other two tests. Also, note that one unit reduction in sample size for the Z test is at the cost of inflated Type I error rate.

Suppose inspection of sample of 137 cans from machine 1 revealed 8 nonacceptable cans, and a sample of 110 cans from machine 2 revealed 3 nonacceptable cans. Thus, we have $N_1 = N_2 = 250$, $n_1 = n_2 = 110$, $k_1 = 8$ and $k_2 = 3$. For the two-tail test in (3), the E test produced a p value of 0.0365. The p value based on the binomial test is 0.1378. The Z statistic is 2.284, and the p value based on the standard normal distribution is 0.0224. Thus, the E test and Z test reject the null hypothesis at the level of significance 0.05, whereas the binomial test does not reject the null hypothesis.

7. CONCLUSION

This article evaluated the exact properties of two tests for testing the difference between proportions in two finite populations. Since there are six sample size-parameter configurations (i.e., $p_1, n_1, N_1, p_2, n_2, N_2$), it is practically impossible to include all configurations in numerical studies. However, we computed the sizes and powers of both Z test and E test for many arbitrarily chosen sample size-parameter configurations. The findings are similar to those in the graphs given in Figures 1–3. The approximate E test is almost exact because the sizes of the test seldom exceed the nominal level by a significant amount. Furthermore, we provided the necessary programs for computing the p values and powers of the E test at the Web page <http://lib.stat.cmu.edu/TAS/>, and so there are no technical difficulties in application of the E test for practical use. An advantage of our program is that it can be used to test the difference between two binomial proportions (if one decided to use the test T_{AU} in Storer and Kim 1990) by simply inputting large values (say, 100,000) for both lot sizes.

[Received October 2001. Revised April 2002.]

REFERENCES

- Angus, J. E. (1994), "Bootstrap One-Sided Confidence Intervals for the Log-normal Mean," *The Statistician*, 43, 395–401.
- Berger, R. L. (1996), "More Powerful Tests From Confidence Interval p Values," *The American Statistician*, 50, 314–318.
- Buonaccorsi, J. P. (1987), "A Note on Confidence Intervals for Proportions in Finite Populations," *The American Statistician*, 41, 215–218.
- Burstein, H. (1975), "Finite Population Correction for Binomial Confidence Limits," *Journal of the American Statistical Association*, 70, 67–69.
- Casella, G., and Berger, R. L. (2002), *Statistical Inference*, Pacific Grove, CA: Duxbury.

- Chan, I. S. F., and Zhang, Z. (1999), "Test-Based Exact Confidence Intervals for the Difference of Two Binomial Proportions," *Biometrics*, 55, 1202–1209.
- Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia, PA: SIAM.
- Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.
- Katz, L. (1953), "Confidence Intervals for the Number Showing a Certain Characteristic in a Population when Sampling is Without Replacement," *Journal of the American Statistical Association*, 48, 256–261.
- Li, S., and Chen, H. J. (1999), "A Note on the Determination of Sample Sizes for Hypergeometric Distributions," *Communications in Statistics—Theory and Methods*, 28, 1749–1757.
- Liddell, D. (1978), "Practical Tests of 2×2 Contingency Tables," *The Statistician*, 25, 295–304.
- Storer, B. E., and Kim, C. (1990), "Exact Properties of Some Exact Test Statistics Comparing Two Binomial Proportions," *Journal of the American Statistical Association*, 85, 146–155.
- Suissa, S., and Shuster, J. J. (1985), "Exact Unconditional Sample Sizes for the 2×2 Binomial Trials," *Journal of the Royal Statistical Society, Ser. A*, 148, 317–327.
- Thomson, J. L. (2002), "Inferential Procedures for Some Discrete Distributions," PhD Dissertation, Department of Mathematics, University of Louisiana at Lafayette.
- Wendell, J. P., and Schmee, J. (2001), "Likelihood Confidence Intervals for Proportions in Finite Populations," *The American Statistician*, 55, 55–61.