Communications in Statistics—Simulation and Computation®, 35: 149–160, 2006

ISSN: 0361-0918 print/1532-4141 online
DOI: 10.1080/03610910500415993

Taylor & Francis
Taylor & Francis Group

Multivariate Analysis

Exact Size and Power Properties of Five Tests for Multinomial Proportions

YONG CAI[1] AND K. KRISHNAMOORTHY[2]

[1]NOP World, Bala Cynwyd, Pennsylvania, USA
[2]Department of Mathematics, University of Louisiana at Lafayette, Lafayette, Louisiana, USA

In this article, exact properties of the chi-square test, likelihood ratio test (LRT), Hoel's test (1938), Nass' test (1959), and an exact test, for testing the multinomial proportions, are investigated numerically. Numerical studies show that the size behavior of the chi-square test for the multinomial distribution with three or more cells is erratic only when the probabilities are at the boundaries. Otherwise the chi-square test performs very satisfactorily. Hoel's test offers a little improvement over the χ^2-test. The χ^2-test is, in general, superior to the LRT, and inferior to the exact test with respect to size properties. The exact test exhibits very good size properties. Nass' test offers improvement over the χ^2-test. Power studies indicate that the tests are equally powerful provided the sizes of the tests are close to each other for a given parameter configuration. Some recommendations regarding the choice of a test for practical applications are given, and the tests are illustrated using a biological example.

Keywords Binomial distribution; Chi-square approximation; Confidence interval; Coverage probability.

Mathematics Subject Classification Primary 62H15; Secondary 62H17.

1. Introduction

Consider an experiment involving n independent trials, and the outcome of each trial can be classified into one of the m mutually exclusive and exhaustive categories. Let X_i denote the number of outcomes in the ith category, $i = 1, \ldots, m$, out of the n trials. The distribution of the vector of counts $(X_1, \ldots, X_m)$ is called the multinomial distribution with probability vector $(p_1, \ldots, p_m)$, where p_i is the probability that the outcome of a trial will fall into the ith category, $i = 1, \ldots, m$. This is a generalization of the binomial distribution to more than two categories. For example, if the outcomes for the drivers in accidents can be classified as "uninjured," "injury not

Received March 11, 2005; Accepted June 27, 2005
Address correspondence to K. Krishnamoorthy, Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA; E-mail: krishna@louisiana.edu

requiring hospitalization," "injury requiring hospitalization," or "fatality," then the distribution of the counts in these categories would be multinomial (Agresti, 1996). The probability distribution function of $(X_1, \ldots, X_m)$ is given by

$$f(x_1, \ldots, x_m \mid n, p_1, \ldots, p_m) = P(X_1 = x_1, \ldots, X_m = x_m \mid n, p_1, \ldots, p_m)$$
$$= \frac{n!}{x_1!x_2! \ldots x_m!} p_1^{x_1} \cdots p_m^{x_m}, \quad 0 < p_i < 1, \ \sum_{i=1}^{m} p_i = 1. \quad (1)$$

The problem of interest here is to test the hypotheses

$$H_0 : p_1 = p_{10}, \ldots, p_{m-1} = p_{m-1,0} \quad \text{vs.} \quad H_a : p_i \neq p_{i0} \text{ for some } i, \quad (2)$$

where p_{i0}'s are specified probabilities, and $\sum_{i=1}^{m} p_{i0} = 1$. This problem of testing the goodness of fit of a postulated multinomial distribution is a classical problem that remains of current practical interest. This test is routinely used in the analysis of one-way classification where the observed frequencies (the values of X_i) are compared with the theoretical frequencies (np_{i0}). The popular chi-square statistic

$$Q_x = \sum_{i=1}^{m} \frac{(X_i - np_{i0})^2}{np_{i0}} \quad (3)$$

was introduced by Pearson (1900), which can also be written as multivariate Hotelling T^2 statistic given by

$$n(\hat{\mathbf{p}}_x - \mathbf{p}_0)'\Sigma^{-1}(\hat{\mathbf{p}}_x - \mathbf{p}_0), \quad (4)$$

where $\hat{\mathbf{p}}_x = (\hat{p}_{1x}, \ldots, \hat{p}_{m-1,x}) = (x_1/n, \ldots, x_{m-1}/n)$, $\mathbf{p}_0 = (p_{10}, \ldots, p_{m-1,0})$ and the matrix $\Sigma = (\sigma_{ij})_{m-1 \times m-1}$ with $\sigma_{ii} = p_{i0}(1 - p_{i0})$, $i = 1, \ldots, m-1$, and $\sigma_{ij} = -p_{i0}p_{j0}$, $i \neq j$ (see Kendall and Stuart, 1969). Pearson (1900) showed that Q_x is approximately distributed as the chi-square random variable with degrees of freedom $m - 1$, χ^2_{m-1}. As pointed out in Johnson et al. (1997), the accuracy of the approximation has been numerically investigated by several authors. In particular, it was concluded that the accuracy of the χ^2-test increases as $\min\{np_{10}, \ldots, np_{m0}\}$ increases and decreases with increasing m. Some authors have suggested improved approximation to the distribution of the chi-square statistic. For example, see Hoel (1938) and Nass (1959). Nevertheless, Pearson's chi-square test is still popular and commonly used in applications because of its simplicity.

Brown et al. (2001) studied the exact properties of the standard confidence interval (the z-interval) for the binomial case, and pointed out that the accuracy of the intervals depends on the choices of both n and p. They presented a series of examples where the behavior of the coverage probability of the standard interval is chaotic even if the conditions $np \geq 5$ and $n(1 - p) \geq 5$ are met. In view of their article, it is of interest to see the exact size properties of the χ^2-test for the binomial case. It should be noted that, for the binomial case, the usual z-test (one based on $z = (X_i/n - p_{10})/\sqrt{p_{10}(1 - p_{10})/n}$) and the χ^2-test are the same (that is, $z^2 = Q_x$ in (3)). Exact sizes of the χ^2-test are computed using (8) and are given in Table 1 for $p_{10} = 0.5$ and some selected values of n. In the first two rows, sample sizes for which the Type I error rates are very close to the nominal level 0.05 are given; in the last

Table 1
Sizes of the χ^2-test for the binomial case with $p_{10} = 0.5$

n	12	17	22	27	32	37	44	58	67	74	94	114
Sizes	.039	.049	.052	.052	.050	.047	.049	.048	.050	.047	.049	.049
n	16	21	31	43	50	66	75	84	91	104	137	158
Sizes	.077	.078	.071	.066	.065	.064	.064	.063	.060	.062	.060	.059

two rows, sample sizes for which the Type I error rates are larger than the nominal level are presented. It is clear from the tabulated values that the size properties of the χ^2-test is very poor, and the erratic behavior is similar to that of coverage properties of the standard confidence interval reported in Brown et. al. (2001). For example, note that the size of the test is 0.049 when $n = 17$, and is 0.063 when $n = 84$!

In view of the above size properties of the χ^2-test, it is of interest to investigate its exact size properties for $m \geq 3$. In the following section, we outline the five tests to be considered: The χ^2-test, LRT (also known as G-test), Hoel's (1938) test, Nass (1959) test, and an exact test. The exact test is based on the exact p-values that can be computed using multinomial probabilities. In the binomial case, a similar exact method (Clopper and Pearson, 1934) is known to be very conservative. Brown et al. (2001) criticized that Clopper–Pearson method is also inaccurate because it is too conservative, yielding confidence intervals that are unnecessarily wide. Therefore, it is of interest to see the performance of the exact test for the multinomial case.

In Sec. 3, we give expressions for evaluating the exact sizes and powers of a test. Using these expressions, the properties of the tests are investigated numerically. Surprisingly, we observe from our studies that the χ^2-test exhibited erratic behavior only when the probabilities are at the boundaries. In other cases, the χ^2 test performed very well. The LRT is, in general, worse than the χ^2-test. The tests due to Nass and Hoel also performed poorly (even though better than the χ^2-test) when the probabilities are at the boundaries. In other situations, Nass' test offers appreciable improvement over the χ^2-test. When the probabilities are at the boundaries, the sizes (as a function of n) of the exact test are either smaller than or very close to the nominal level. If the probabilities are not too small, then the sizes of the exact test are very close to the nominal level even for small samples. In Sec. 4, the tests are illustrated using a biological example given in Katti and Sastry (1965). Some concluding remarks are given in Sec. 5.

2. The Tests

Let $(k_1, \ldots, k_m)$ be an observed value of $(X_1, \ldots, X_m)$, and define Q_k like Q_x with x replaced by k. That is, Q_k is the observed value of Q_x based on $(k_1, \ldots, k_m)$.

2.1. *The Usual χ^2-Test*

For a given nominal level α, the χ^2-test rejects the null hypothesis in (2) whenever the p-value

$$P(\chi^2_{m-1} \geq Q_k \mid H_0) \leq \alpha.$$

2.2. *The Likelihood Ratio Test*

The likelihood ratio test is also based on the χ^2 approximation. Specifically, the likelihood ratio statistic

$$G_x = 2\sum_{i=1}^{m} X_i \ln\left(\frac{X_i}{np_{i0}}\right) \sim \chi^2_{m-1} \quad \text{asymptotically.}$$

For an observed value G_k of G_x, the LRT rejects the H_0 in (2) whenever the p-value $P(\chi^2_{m-1} \geq G_k|H_0) \leq \alpha$.

2.3. *Hoel's Test*

Hoel (1938) proposed the following approximation to the cumulative distribution function of Q_x. He showed that, for $c > 0$,

$$P(Q_x \leq c) \approx P(\chi^2_{m-1} \leq c) + \frac{1}{n}(R_1(c)S_1 + R_2(c)S_2),$$

where, letting $h = m - 1$,

$$S_1 = \frac{1}{8}\left[\sum_{i=1}^{m} p_{i0}^{-1} - (h^2 + 4h + 1)\right], \quad S_2 = \frac{1}{24}\left[5\sum_{i=1}^{m} p_{i0}^{-1} - (3h^2 + 12h + 5)\right],$$

and

$$\begin{aligned} R_1(c) &= \frac{e^{-c/2}c^{h/2}(c - h - 2)}{2 \cdot 4 \cdot \cdots \cdot (h+2)} \quad \text{and} \\ R_2(c) &= \frac{e^{-c/2}c^{h/2}[c^2 - 2(h+4)c + (h+4)(h+2)]}{2 \cdot 4 \cdot \cdots \cdot (h+4)}, \end{aligned} \tag{5}$$

for even h,

$$\begin{aligned} R_1(c) &= \frac{e^{-c/2}c^{h/2}(c - h - 2)}{1 \cdot 3 \cdot \cdots \cdot (h+2)}\sqrt{\frac{2}{\pi}} \quad \text{and} \\ R_2(c) &= \frac{e^{-c/2}c^{h/2}[c^2 - 2(h+4)c + (h+4)(h+2)]}{1 \cdot 3 \cdot \cdots \cdot (h+4)}\sqrt{\frac{2}{\pi}} \end{aligned} \tag{6}$$

for odd h. For an observed value Q_k of Q_x, the test based on the above approximation rejects the H_0 in (2) when the p-value

$$P(\chi^2_{m-1} \geq Q_k) - \frac{1}{n}(R_1(Q_k)S_1 + R_2(Q_k)S_2) \leq \alpha.$$

2.4. *Nass' Test*

Nass (1959) proposed another approximation to the distribution of Q_x and is given by

$$cQ_x \sim \chi^2_v,$$

where $c = 2E(Q_x)/\mathrm{Var}(Q_x)$ and $v = cE(Q_x)$. The constants c and v are determined so that the mean and variance of cQ_x are the same as those of χ^2_v. Toward this, we note that (see Johnson and Kotz, 1969)

$$E(Q_x) = m - 1 \quad \text{and} \quad \mathrm{Var}(Q_x) = 2(m-1) - (m^2 + 2m - 2)/n + \sum (np_{i0})^{-1}.$$

In equal probability case, that is, $p_{10} = \cdots = p_{m0} = 1/m$, Nass (1959) suggested using continuity correction for the chi-square statistic. In this case Q_x is defined as $\frac{\sum_{i=1}^m X_i^2 - 1}{(n/m)} - n$ and the above variance expression simplifies to $2(m-1)(n-1)/n$. Katti and Sastry (1965) used this test for some biological examples with small sample sizes. Notice that the Pearson's χ^2 approximation matches only the first moments of the χ^2_{m-1} and Q_x whereas the Nass' approximation matches the first two moments of the χ^2_{m-1} and cQ_x. Therefore, Nass' approximation is expected to be better than the Pearson's approximation.

The test based on the above approximate distribution rejects the null-hypothesis in (2) whenever the p-value $P(\chi^2_v \geq cQ_k) \leq \alpha$.

Remark 2.1. The above approximation is equivalent to approximating the distribution of Q_x by a gamma(a, b) distribution with shape parameter a and the scale parameter b. These parameters are to be determined so that the mean and variance of Q_x are the same as those of the gamma(a, b) distribution. This yields $a = (E(Q_x))^2/\mathrm{Var}(Q_x)$ and $b = \mathrm{Var}(Q_x)/E(Q_x)$.

2.5. *The Exact Test Based on Q_x*

The exact p-value of the χ^2-statistic Q_x can be computed using multinomial probabilities, and is given by

$$P(Q_x \geq Q_k \mid H_0) = \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} \cdots \sum_{x_{m-1}=0}^{n-x_1-\cdots-x_{m-2}} f(x_1, \ldots, x_m \mid H_0) I[Q_x \geq Q_k], \tag{7}$$

where $I[\cdot]$ is the indicator function. This test rejects the H_0 in (2) whenever the above p-value is less than or equal to α.

3. Exact Evaluation of Sizes and Powers

For given values of n, $(p_1, \ldots, p_{m-1})$ and $(p_{10}, \ldots, p_{m-1,0})$, the exact sizes and powers of the approximate test can be computed using the expression

$$\sum_{k_1=0}^{n} \sum_{k_2=0}^{n-k_1} \cdots \sum_{k_{m-1}=0}^{n-k_1-\cdots-k_{m-2}} f(k_1, \ldots, k_m \mid n, p_1, \ldots, p_m) I\left[P(\chi^2_{m-1} \geq Q_k \mid H_0) \leq \alpha\right]. \tag{8}$$

Notice that, when $p_1 = p_{10}, \ldots, p_m = p_{m0}$ (that is, H_0 is true), the above expression gives the sizes of the approximate test. The exact sizes and powers of the Nass' approximate test can be computed similarly.

The exact sizes and powers of the exact test based on Q_x can be computed using the following expression.

$$\sum_{k_1=0}^{n} \sum_{k_2=0}^{n-k_1} \cdots \sum_{k_{m-1}=0}^{n-k_1-\cdots-k_{m-2}} f(k_1, \ldots, k_m \mid n, p_1, \ldots, p_m) I[P(Q_x \geq Q_k \mid H_0) \leq \alpha], \tag{9}$$

where $P(Q_x \geq Q_k \mid H_0)$ is defined in (7).

We computed the sizes of the tests for $m = 3$, 4, and 5, and various values of n. All the intermediate values are computed using double precision, and the end

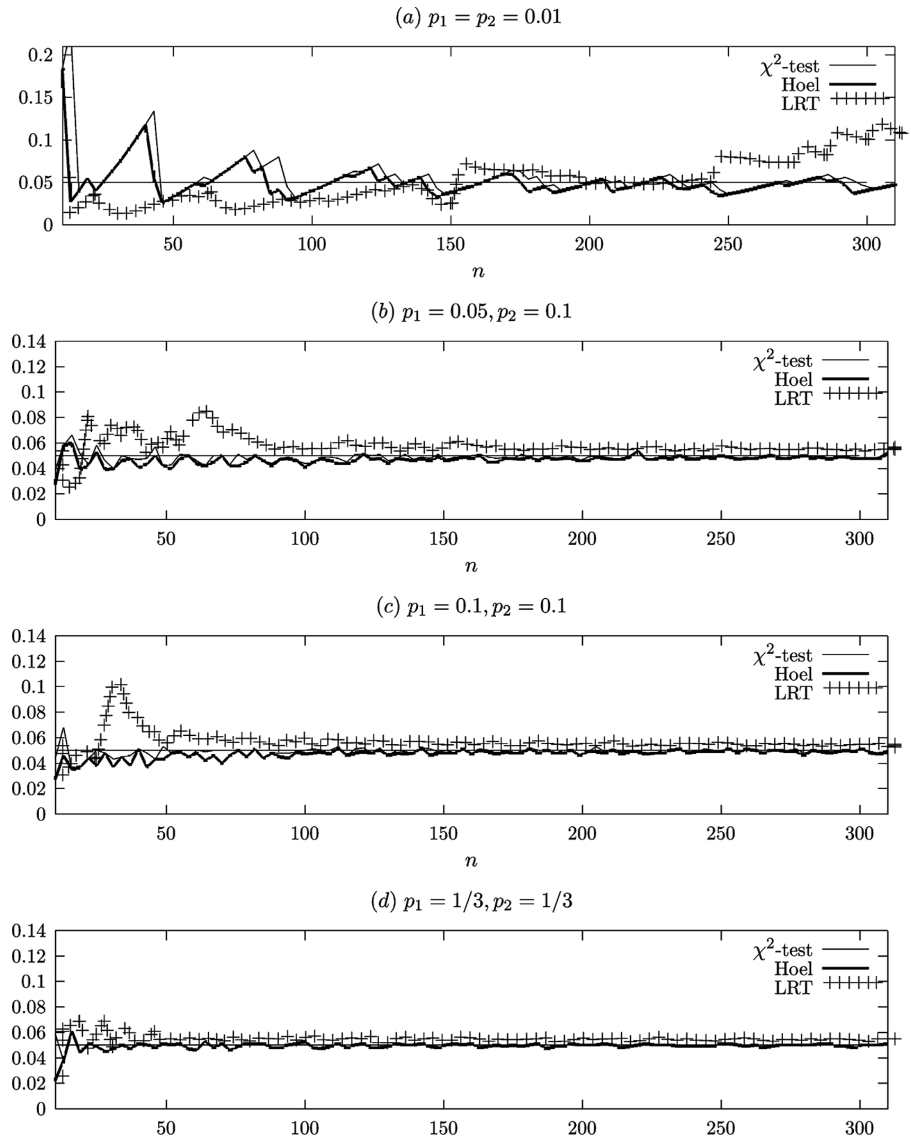

Figure 1. Sizes of the χ^2-test, Hoel's test, and LRT as functions of n; $m = 3$ and $\alpha = 0.05$.

results are rounded up to three digits. The objectives of the numerical studies are to find satisfactory approximate tests, small sample properties of the exact test and the satisfactory approximate tests, and power comparison.

In Figs. 1(a–d), the sizes of the χ^2-test, LRT, and Hoel's test are plotted as functions of $n(n = 10(3)310)$ at 5% level. We see from these four plots that the χ^2-test exhibits erratic behavior only when the probabilities are at the boundaries (see the plot for $p_1 = p_2 = .01$). Hoel's test also exhibits similar behavior but less severe. This size behavior continues to be erratic even for large sample sizes. It appears to be stable and close to 0.05 for n around 180 or larger. Hoel's test offers only slight improvement over the χ^2-test (see Figs. 1(b–d)). The LRT performs poorly in all the cases. In Fig. 1(a), where the cell probability vector is (.01, .01, .98), the LRT performs weirdly; it is conservative for sample sizes 150 or less, liberal for sample sizes more than 150, and its size exceeds 0.1 when n is around 300. Even for equal probability case (Fig. 1(d)), the LRT performs erratically for $n \leq 50$. It is clear from Fig. 1 that the LRT is inferior to the χ^2-test, and so we will not include the LRT for further comparison with other tests. As Hoel's test offers only little or no improvement over the χ^2-test, this test will also be dropped from further comparison studies.

In Fig. 2, the sizes of the χ^2-test, Nass' test, and the exact test are compared for $n = 10(1)80$. Unlike the binomial case, the accuracy of the χ^2 test increases as the sample size increases provided the cell probabilities are not too small. Even for the case ($p_1 = 0.05, p_2 = 0.1$), its sizes barely exceed 0.06. For equal probability case, its performance is very satisfactory for $n \geq 15$. The size property of the Nass test is also poor when the probabilities are small; however, it performs better than the χ^2-test in controlling the Type I error rates. For small probabilities, the exact test is

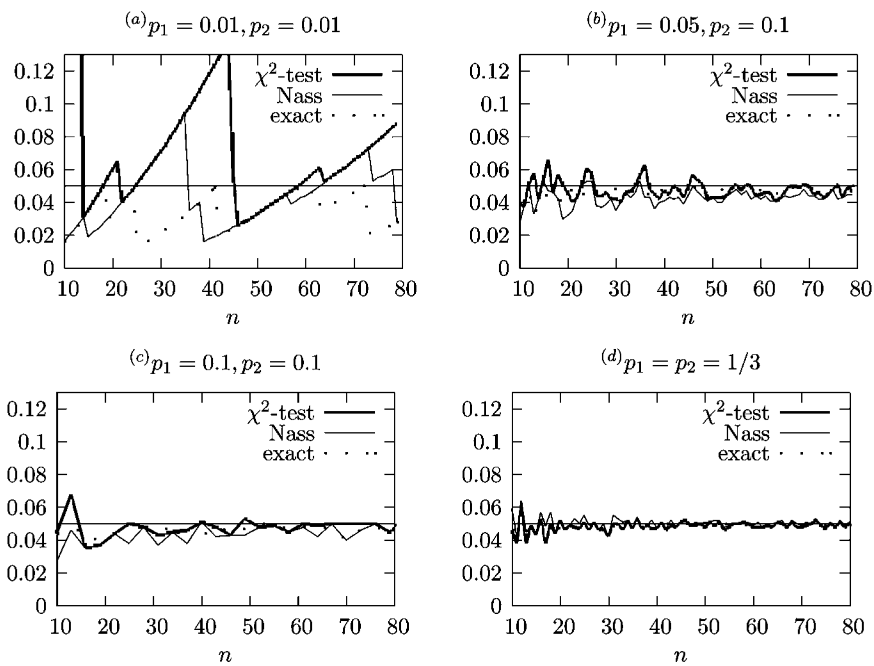

Figure 2. Sizes of the tests as functions of n; $m = 3$ and $\alpha = 0.05$.

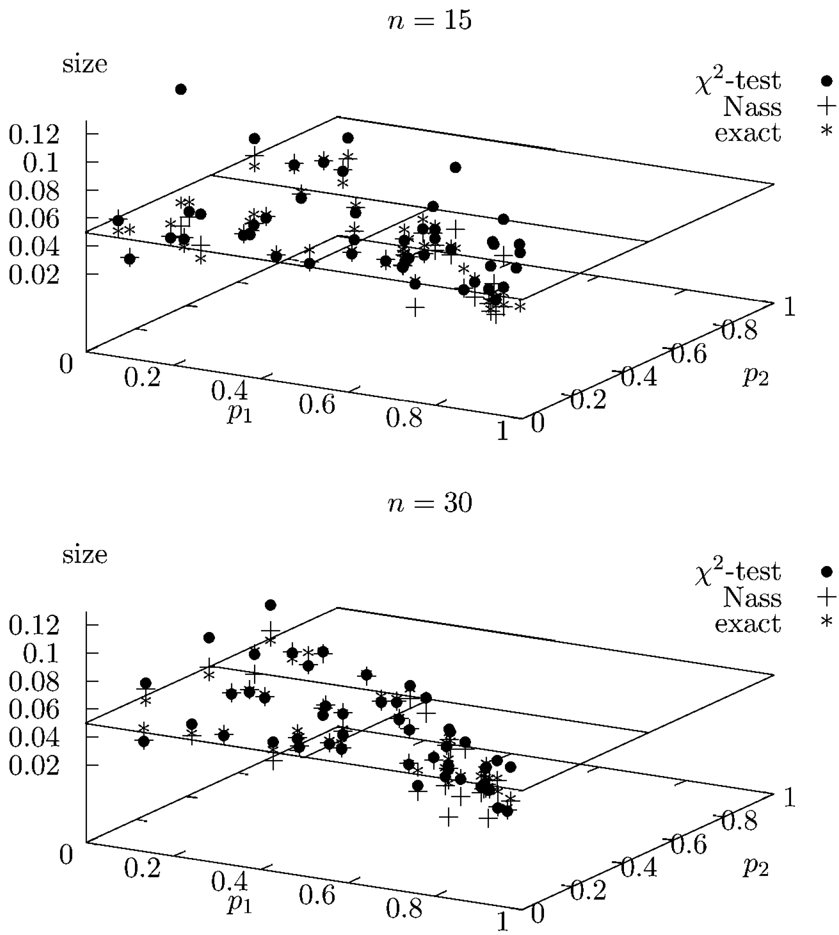

Figure 3. Sizes of the tests as functions of probabilities; $m = 3$ and $\alpha = 0.05$.

conservative for some sample sizes. The sizes of the exact test are very close to the nominal level when the probabilities are not too small.

The sizes of the χ^2-test, Nass' test, and the exact test are plotted as functions of cell probabilities in Fig. 3 for $m = 3$, $n_1 = 15$, and $n_2 = 30$. The cell probabilities were generated randomly subject to the constraint that $p_1 + p_2 + p_3 = 1$. We again observe that the χ^2-test performs poorly only at the boundaries. Nass' test is in general slightly conservative. The sizes of the exact tests are very close to the nominal level 0.05 for most of the cases considered.

For the case of $m = 4$, the sizes and powers of the tests are presented, respectively, in Tables 2 and 3. The size properties of all three tests are similar as in the case of $m = 3$. We also observe from Table 3 that the Nass' test is less powerful because it is in general conservative. The exact test and the χ^2 test are equally powerful when their sizes are equal (see the cases $(p_1, p_2, p_3) = (.4, .4, .1)$ and $(.5, .4, .05)$). Because of its inflated Type I error rates, the χ^2-test appears to be more powerful than other two tests (see the cases $(p_1, p_2, p_3) = (.01, .01, .01)$ and $(.5, .4, .05)$).

Table 2
Exact sizes of the tests when $m = 4$ and $\alpha = 0.05$

			$n = 15$			$n = 20$			$n = 25$			$n = 40$		
p_1	p_2	p_3	1	2	3	1	2	3	1	2	3	1	2	3
.01	.01	.01	.073	.031	.031	.054	.050	.050	.076	.076	.031	.041	.030	.041
.01	.05	.05	.035	.028	.035	.061	.040	.049	.054	.044	.049	.060	.042	.048
.01	.10	.20	.051	.026	.048	.052	.030	.049	.057	.041	.050	.060	.039	.050
.01	.30	.30	.049	.025	.049	.050	.030	.050	.054	.039	.049	.060	.039	.050
.20	.25	.30	.043	.047	.049	.046	.047	.050	.044	.050	.050	.048	.051	.050
.40	.40	.10	.050	.042	.050	.050	.050	.049	.044	.044	.049	.044	.044	.048
.50	.30	.10	.043	.043	.050	.047	.047	.049	.044	.044	.047	.049	.045	.050
.50	.40	.05	.055	.042	.050	.053	.043	.048	.050	.042	.050	.051	.047	.050
.50	.40	.09	.050	.029	.049	.047	.034	.050	.054	.041	.049	.058	.041	.050

1 = usual chi-square test; 2 = Nass' approximate test; 3 = exact test.

Finally, we compare the sizes of the approximate tests (χ^2-test and Nass' test) in Fig. 4 for $m = 5$ to determine the best approximate test. We now clearly see the improvement of Nass' approximation over the Pearson's χ^2 approximation. In particular, Nass' test controls the sizes very well when the probabilities are small.

Table 3
Exact powers of the tests when $m = 4$ and $\alpha = 0.05$

p_1	p_2	p_3	1	2	3	p_1	p_2	p_3	1	2	3
					$n = 15$						
$(p_{10}, p_{20}, p_{30}) = (.01, .01, .01)$						$(p_{10}, p_{20}, p_{30}) = (.4, .4, .1)$					
.01	.01	.01	.073	.031	.031	.40	.40	.10	.050	.042	.050
.01	.05	.05	.503	.342	.342	.50	.30	.10	.091	.079	.091
.01	.10	.10	.718	.576	.576	.35	.35	.10	.214	.205	.214
.10	.15	.01	.855	.737	.737	.40	.35	.20	.205	.188	.205
.20	.01	.01	.874	.840	.840	.30	.35	.10	.369	.359	.369
.01	.05	.20	.931	.879	.879	.20	.40	.10	.582	.569	.582
.01	.10	.20	.970	.932	.932	.20	.35	.10	.712	.704	.712
					$n = 25$						
$(p_{10}, p_{20}, p_{30}) = (.01, .1, .2)$						$(p_{10}, p_{20}, p_{30}) = (.5, .4, .05)$					
.01	.10	.20	.057	.041	.050	.50	.40	.05	.050	.042	.050
.05	.20	.20	.603	.531	.566	.45	.40	.10	.194	.173	.194
.05	.20	.30	.742	.663	.721	.30	.40	.10	.816	.801	.816
.10	.10	.30	.816	.788	.811	.50	.20	.05	.897	.881	.897
.10	.10	.40	.930	.902	.928	.50	.20	.10	.834	.814	.834
.20	.20	.20	.990	.987	.989	.30	.30	.05	.981	.979	.981
.21	.20	.20	.993	.990	.992	.50	.10	.05	.996	.995	.996

1 = usual chi-square test; 2 = Nass' approximate test; 3 = exact test.

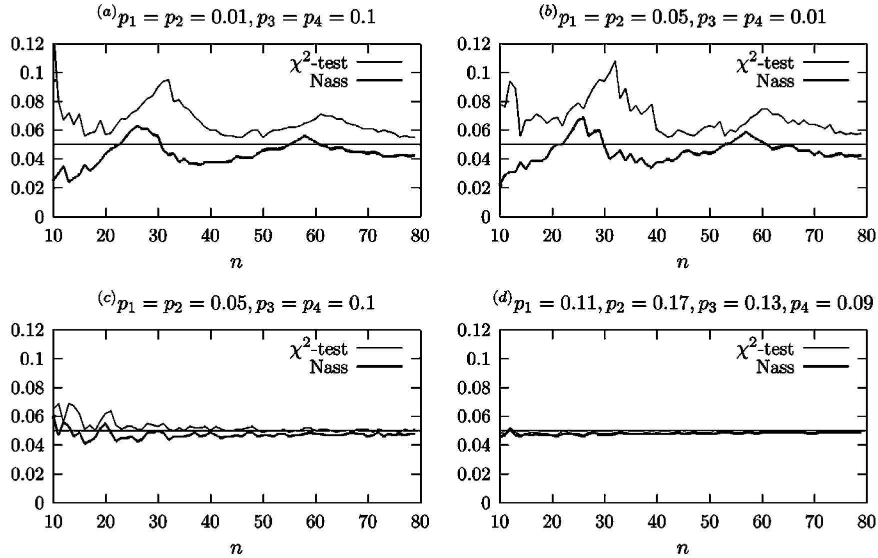

Figure 4. Sizes of the tests as Functions of n; $m = 5$ and $\alpha = 0.05$

Our overall recommendations of the tests based on their numerical accuracies and computational simplicities are as follows: (i) If one decided to use an approximate test then Nass' test is preferable to the usual χ^2-test. (ii) The exact test can be recommended for any cases; however, as it is computationally intensive, it can be used if the number cells is small, and n is not too large. (iii) Nass' test is suitable for practical applications especially when m is moderate or large.

4. An Example

We shall now illustrate the five tests in Sec. 2 using the example and the data given in Katti and Sastry (1965). The multinomial data were collected from an experiment that was conducted to understand the symbiotic relationship between crabs and host animals. The apparatus used in this experiment has a central chamber which was connected to six radial chambers symmetrically so that a crab could enter any of these radial chambers with equal ease and could not return to the central chamber. Host animals were placed first into desired radial chambers and the crabs were introduced later in the central chamber. The number of crabs entering different radial chambers were recorded after a period of time. The results are given in Table 4. In the first four experiments no host animals were used, and the cell probabilities (the probability that a crab entering a particular chamber) are expected to be 1/6. Experiments 5 through 13 consisted host animals in selected chambers.

The p-values of the usual chi-square test, LRT, Hoel's test, Nass' test with continuity correction, and the exact test are given Table 4. No continuity correction was used to compute the p-values of the exact test. The LRT is applicable only for large samples; nevertheless we applied it here for the sake of completeness and to see its performance for the small samples. The p-values for all experiments are

Table 4
p-values of the tests for 13 experiments

Expt.	n	Distr. in 6 chambers $(k_1, \ldots, k_6)$	χ^2-test	LRT	Hoel's test	Nass' test	exact
1	9	(4, 0, 1, 1, 0, 3)	.109	.065	.110	.130	.121(.122)[1]
2	13	(0, 2, 2, 2, 1, 6)	.087	.084	.087	.097	.091(.091)
3	10	(1, 5, 1, 0, 1, 2)	.101	.124	.102	.118	.099(.099)
4	16	(3, 2, 5, 2, 1, 3)	.623	.633	.621	.691	.678(.678)
5	8	(5*, 1, 0, 0, 1, 1)	.023	.042	.024	.025	.028(.029)
6	11	(2, 0, 3*, 0, 3, 3)	.315	.101	.315	.375	.387(.387)
7	16	(1, 1, 2, 1, 8*, 3)	.016	.047	.016	.016	.016(.016)
8	14	(3, 2, 5*, 1, 1, 2)	.434	.479	.433	.495	.502(.502)
9	14	(4, 2, 2, 1, 4*, 1)	.549	.549	.548	.623	.628(.624)
10	12	(5*, 2, 0, 1, 4, 0)	.051	.021	.052	.056	.060(.060)
11	10	(3, 0, 3, 3*, 0, 1)	.236	.089	.236	.284	.264(.264)
12	15	(1, 1, 8*, 0, 3*, 2)	.005	.010	.005	.005	.006(.006)
13	15	(6*, 2, 1, 5*, 1, 0)	.038	.025	.038	.040	.042(.042)

*These chambers contained one host animal.
[1]The numbers in parentheses are Monte Carlo estimates of the p-values.

computed under $H_0 : p_1 = \cdots = p_6 = 1/6$. We observe first that the p-values of the LRT are not consistent with those of other tests in many cases. Even though the p-values of the tests are considerably different for many experiments, we see that the results of all the tests are in agreement except for experiment 10, where the LRT rejects the null hypothesis while other tests do not. We also note that the p-values of the χ^2-test and Hoel's test are practically the same for all the experiments, which is in agreement with their size properties given in Sec. 3. We also computed Monte Carlo estimates (based on 100,000 runs) of the p-values, which are almost identical to those of the exact p-values. The multinomial random numbers were generated using IMSL subroutine RNMTN. For further analysis of the data regarding other objectives, we refer to Katti and Sastry (1965).

5. Some Concluding Remarks

Our numerical investigation revealed that, unlike in the binomial case, the standard method (χ^2-test) produces results that are remarkably good if the cell probabilities are not too small and the number of cells is three or more. Similarly, the exact method also performs very well, unlike its performance in the binomial case where it is too conservative even if the success probability is not small (see Cai and Krishnamoorthy, 2005). The exact method can be implemented (at least easy to code in any programming language) for any number of cells and sample size n even though it may be time consuming. For a large number of cells, we recommend the Monte Carlo method used in Sec. 4 to compute the p-value $P(Q_x \geq Q_k \mid n, p_{10}, \ldots, p_{m0})$. As noticed in Table 4, Monte Carlo method with 100,000 runs will be enough to compute the p-value with a considerable accuracy.

References

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: John Wiley.

Brown, L. D., Cai, T. T., DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.* 16:101–133.

Cai, Y., Krishnamoorthy, K. (2005). A simple improved inferential method for some discrete distributions. *Computat. Statist. Data Anal.* 48:605–621.

Clopper, C. J., Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of binomial. *Biometrika* 26:404–413.

Hoel, P. G. (1938). On the chi-square distribution for small samples. *Ann. Mathemat. Statist.* 9:158–165.

Johnson, N. L., Kotz, S. (1969). *Discrete Distributions*. New York: John Wiley.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. New York: John Wiley.

Katti, S. K., Sastry, A. N. (1965). Biological examples of small expected frequencies and the chi-square test. *Biometrics* 21:49–54.

Kendall, M. G., Stuart, A. S. (1969). *The Advanced Theory of Statistics*. New York: Hafner Publishing Company.

Nass, C. A. G. (1959). The χ^2-test for small expectations in contingency tables, with special reference to accidents and absenteeism. *Biometrika* 46:365–385.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophical Mag. 5th Ser.* 50:157–175.