Computational Statistics and Data Analysis 52 (2008) 5046–5053

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

An exact confidence set for two binomial proportions and exact unconditional confidence intervals for the difference and ratio of proportions

Jenő Reiczigel [a,*], Zsolt Abonyi-Tóth [a], Júlia Singer [b]

[a] Szent István University, Faculty of Veterinary Science, Department of Biomathematics and Informatics, Istvan u. 2, Budapest, H-1078, Hungary
[b] Baxter BioScience, Global Clinical Biostatistics Vienna, Austria

ARTICLE INFO

Article history:
Received 17 July 2007
Received in revised form 28 April 2008
Accepted 28 April 2008
Available online 5 May 2008

ABSTRACT

An exact joint confidence set is proposed for two binomial parameters estimated from independent samples. Its construction relies on inverting the minimum volume test, a two-dimensional analogue of Sterne's test for a single probability. The algorithm involves computer-intensive exact computation based on binomial probabilities. The proposed confidence set has good coverage properties and it performs much better than the likelihood-based confidence set for the same problem. Applying the principle of intersection-union tests, the method can be used to derive exact tests and confidence intervals for functions of the two binomial parameters. Based on this, new exact unconditional two-sided confidence intervals are proposed for the risk difference and risk ratio. The performance of the new intervals is comparable to that of certain well-known confidence intervals in small samples. Extension of the methods described to two hypergeometric or two Poisson variables is straightforward.

1. Introduction

There are some phenomena which are usually modeled by two probabilities because the use of a single combined measure would result in an undesirable loss of information. For example, diagnostic tests have to be characterized by their sensitivity and specificity, the risk of disease is often estimated separately for children and for adults, and so on. In these examples the two probabilities are estimated from two independent samples. We propose a method to construct exact two-dimensional joint confidence sets (CS) for the two unknown probabilities, based on two independent samples. The method relies on the inversion of the minimum volume test, a two-dimensional analogue of Sterne's test for a single probability (Sterne, 1954).

In the literature we have found no specific method for joint CS construction for two binomial proportions, neither asymptotic nor exact. Therefore, we have chosen two general methods as terms of comparison for the new CS method. One is the rectangular confidence set, which can be constructed by combining the two confidence intervals (CI) calculated from the two independent samples, namely to obtain a level $(1-\alpha)$ two-dimensional CS, a rectangular set made of two level $\sqrt{1-\alpha}$ CIs should be taken. The other CS considered for comparison is that based on the likelihood.

The proposed two-dimensional CS can be used to derive new tests and CIs for functions of the two parameters such as the difference or ratio between the two proportions, the odds ratio or linear combinations of the proportions. These are important and widely used measures in epidemiology and clinical trials. There is a variety of tests and CIs for the difference

* Corresponding author. Tel.: +36 1 478 4215; fax: +36 1 478 4217.
E-mail address: reiczigel.jeno@aotk.szie.hu (J. Reiczigel).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2008.04.032

and ratio of two probabilities, as well as for other quantities derived from the 2 × 2 table (Skipka et al., 2004; Zhou et al., 2004; Andres and Tejedor, 2004; Munk et al., 2005; Agresti and Gottard, 2007; Gupta and Tian, 2007). We illustrate how exact CIs for risk difference and risk ratio can be derived from the two-dimensional CS, and compare them with the exact procedures proposed by Agresti and Min (2001), as implemented in StatXact 7 PROCs for SAS Users (2005).

The paper is organized as follows. In Section 2 we describe the construction of the minimum volume CS, in Section 3 we make comparisons with the rectangular CS and the likelihood-based CS, in Section 4 we describe how CIs for risk difference and risk ratio can be derived, and in Section 5 we apply the new methods to real data. In Section 6 we make some comparisons and give a short discussion of the results.

2. The minimum volume confidence set

The minimum volume test is the two-dimensional analogue of the test proposed by Sterne (1954) for the one-dimensional case. The acceptance region is constructed by taking the points of the two-dimensional sample space in descending order of their null probabilities (first the one with highest probability, then the one with second highest, and so on) until the null probability of the acceptance region reaches the desired level. In case of equal probabilities all sample points with the same probability are handled together, i.e. all of them are included in or excluded from the acceptance region at the same time.

The acceptance region defined in this way has minimum volume (in two dimensions in fact minimum area). For discrete variables this means that it contains the minimal number of sample points among all possible regions having a probability greater or equal to the desired level under H_0.

For a formal description of the procedure, let us consider the simple null hypothesis $H_0 : \pi = \pi_0$ and the alternative $H_1 : \pi \neq \pi_0$, where $\pi = (\pi_1, \pi_2) \in [0, 1] \times [0, 1]$ represents the pair of unknown binomial parameters while $\pi_0 = (\pi_{01}, \pi_{02}) \in [0, 1] \times [0, 1]$ denotes the pair of hypothesized values. The test is based on two independent samples of size n_1 and n_2.

For any $t \in [0, 1]$ consider the set

$$A_{\pi_0,t} = \{(i, j) | P_{\pi_0}(i, j) > t\} \tag{1}$$

and the function

$$f_{\pi_0}(t) = P_{\pi_0}(A_{\pi_0,t}) = \sum_{(i,j)\in A_{\pi_0,t}} P_{\pi_0}(i, j), \tag{2}$$

where $P_{\pi_0}(i, j)$ is the probability of the outcome $(i, j) \in \{0, 1, \ldots, n_1\} \times \{0, 1, \ldots, n_2\}$ under the null hypothesis, i.e.

$$P_{\pi_0}(i, j) = b_{n_1,\pi_{01}}(i) \cdot b_{n_2,\pi_{02}}(j),$$

with $b_{n,\pi}(.)$ being the probability mass function of the binomial distribution with parameters n and π. Note that $f_{\pi_0}(t)$ is monotone decreasing and left continuous.

To have a minimal volume acceptance region $A^{(1-\alpha)}_{\pi_0}$ with acceptance probability $(1-\alpha)$ under H_0, we take $A^{(1-\alpha)}_{\pi_0} = A_{\pi_0,t^*}$ with $t^* = \max\{t | f_{\pi_0}(t) \geq 1 - \alpha\}$. The CS is obtained by inverting the test so that the CS consists of all parameter pairs for which the observed outcome belongs to the acceptance region at the level of interest. Formally, given the observation $(i, j) \in [0, n_1] \times [0, n_2]$, the set of parameter pairs π_0, for which $(i, j) \in A^{(1-\alpha)}_{\pi_0}$ form a confidence set at confidence level of $(1 - \alpha)$ that will be denoted by $\mathrm{CS}(i, j)$ and defined by

$$\mathrm{CS}(i, j) = \{\pi_0 \in [0, 1] \times [0, 1] | (i, j) \in A^{(1-\alpha)}_{\pi_0}\}.$$

The calculation of the CS is computationally intensive: we simply scan the whole parameter-space $[0, 1] \times [0, 1]$ with a resolution depending on the desired precision and we determine the acceptance region for each possible parameter pair, and examine whether or not the observed outcome lies within the acceptance region. Given the present speed of computers, it is feasible without any sophisticated optimization to achieve a resolution of 0.001 for sample sizes up to a few hundreds while maintaining a running time within a few minutes.

It must be noted that the resulting minimum volume CS (MVCS) may not be convex, moreover it may even contain "holes" (Fig. 1). This property is analogous to that of Sterne's interval in the one-dimensional case (Reiczigel, 2003). A typical scenario how a hole may occur is as follows. Assume that for a certain parameter pair, the observed sample (S) is the one with the smallest probability in the acceptance region. Changing the parameter pair a little, the probability of S decreases below the probability of another sample, so S drops out of the acceptance region (consequently the parameter pair drops out of the CS). Now S is the sample with the highest probability outside the acceptance region. Changing the parameter pair further, the probability of the acceptance region decreases a little, allowing inclusion of one more sample, so S comes back (thus the parameter pair comes back into the CS).

To avoid such undesirable features, we propose "filling the gaps," that is, inclusion of all parameter pairs (p_1, p_2) in the CS for which there exist some p'_1 and p''_1 so that either $p'_1 < p_1 < p''_1$ and both (p'_1, p_2) and (p''_1, p_2) belong to the CS; or similarly, if there exist some p'_2 and p''_2 so that $p'_2 < p_2 < p''_2$ and both (p_1, p'_2) and (p_1, p''_2) are in the CS. By this procedure the minimum volume property is lost, but according to our numerical investigations the increase in volume of the CS is negligible.

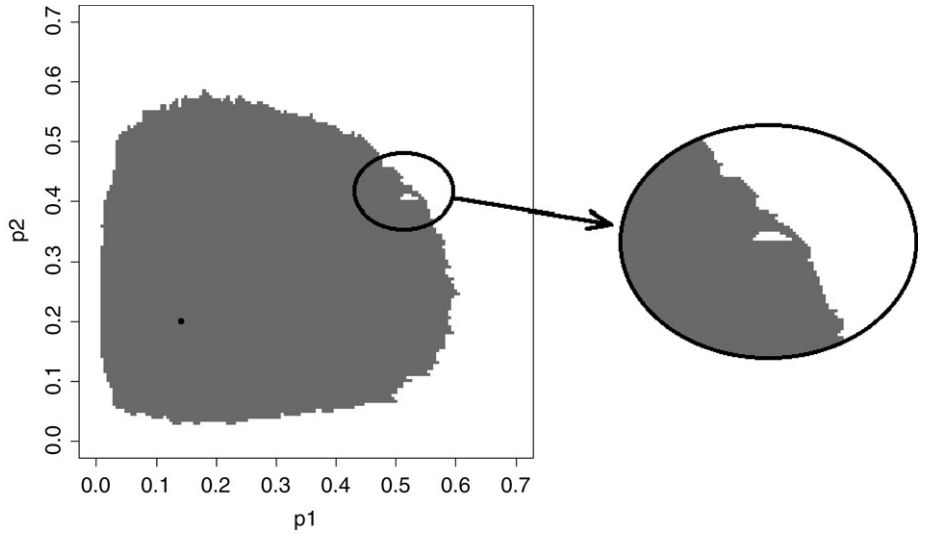

Fig. 1. An example illustrating that the minimum volume confidence set may not be convex, moreover it may contain holes (observed proportions: $p_1 = 1$ in 7; $p_2 = 2$ in 10).

Table 1

	$n_1 = n_2 = 10$	$n_1 = n_2 = 30$	$n_1 = n_2 = 100$
Level-adjusted likelihood-based CS			
Adjusted nominal level	0.9791	0.9789	0.9698
Minimum coverage	0.9511	0.9570	0.9540
Mean coverage	0.9771	0.9770	0.9686
Minimum volume CS			
Minimum coverage	0.9500	0.9500	0.9500
Mean coverage	0.9539	0.9512	0.9503

3. Comparison with the likelihood-based confidence set

As there is no other method for joint CS construction that is developed specifically for binomial parameters, we compare the MVCS method introduced in this paper with the likelihood-based CS (LBCS), which is a widely accepted general method for CS construction.

If $l_{(p_1,p_2)}$ denotes the log-likelihood function of a parameter pair $(p_1, p_2) \in [0, 1] \times [0, 1]$ and $l_{\max}$ its maximum over the unit square $[0, 1] \times [0, 1]$, then a level $(1 - \alpha)$ LBCS consists of all parameter pairs satisfying

$$2(l_{\max} - l_{(p_1,p_2)}) < \chi^2_{1-\alpha} \quad (df = 2).$$

The LBCS is simple to calculate and it is asymptotically valid, but it is not necessarily exact for finite sample sizes. Unfortunately, coverage probabilities considerably lower than the intended value may occur, even for large samples. For example, evaluating coverage at a nominal level of 0.95 for $n_1 = n_2 = 300$ for all parameter pairs (p_1, p_2) on a 0.01 grid on the parameter space $[0, 1] \times [0, 1]$, the minimum coverage was found to be 0.8836. This is far below the nominal level of 0.95. The mean coverage was 0.9481, which is fairly close to the nominal, but still too small. This is very similar to the one-dimensional likelihood-based CI (Newcombe, 1998a). The main reason for this disappointing phenomenon is the discreteness of the binomial distribution. Similar behaviour was reported for the Wilson score interval by Agresti and Coull (1998).

To force the LBCS to maintain the prescribed level, the computational method described in Reiczigel (2003) can be applied. It is based on the idea that one can vary the nominal level iteratively until the actual coverage gets as close as possible to the desired level, provided that it remains above it (discreteness may prevent it from reaching it exactly). Table 1 demonstrates, for different sample sizes, which nominal levels are needed to have 95% minimum coverage. Technically, we varied the chi-square threshold on the right hand side of the equation in steps of 0.001 to set the minimum coverage to 95%. Step size can be selected according to the required precision.

Although Table 1 shows that minimum coverage is fairly close to the nominal level, the CS obtained in this way is far from perfect because for most parameter pairs its coverage probability is much higher than the nominal, i.e. the level-adjusted LBCS is too conservative. The reason for this is the large difference between the minimal and typical coverage probabilities (ideally coverage probability should be constant over the whole parameter space).

The minimum coverage probability of the MVCS remains close to the desired level for most parameter pairs. The coverage probability of the MVCS shows considerably less fluctuation as the true parameters vary and so it tends to be closer to the

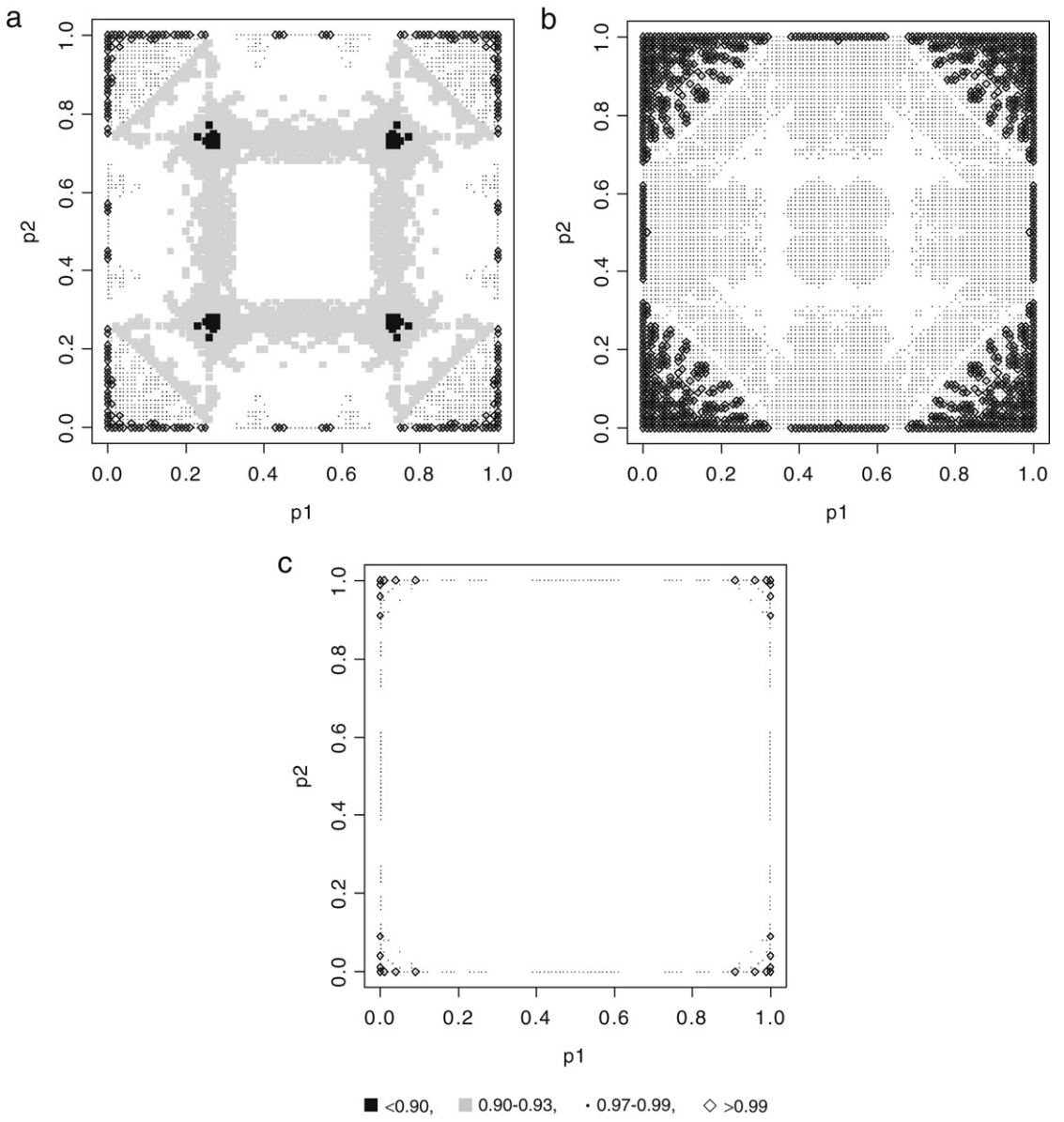

Fig. 2. Coverage probability at a nominal level of 95%, as a function of the true parameters (p_1, p_2) for $n_1 = n_2 = 10$. (a) likelihood-based CS, (b) level-adjusted likelihood-based CS, (c) minimum volume CS.

nominal level than the LBCS even when level adjustment is applied to the latter. This is the reason why the mean coverage is fairly close to the minimum coverage (Table 1). Results for other confidence levels are similar.

It is interesting to see how the pattern of coverage depends on the true parameters. The coverage plots in Fig. 2 illustrate this for $n_1 = n_2 = 10$. Here too, a 0.01-grid was used on $[0, 1] \times [0, 1]$. Fig. 3 shows the shape differences between the MVCS and LBCS for a selection of observed samples. Note that in spite of its higher minimum coverage, the MVCS has about the same volume as the LBCS. The MVCS tends to lie closer to the center of the parameter space than the LBCS. Following the terminology of Newcombe (1998a), the MVCS is relatively mesial while the LBCS is relatively distal in location. The level-adjusted LBCS is not shown because it is so much larger and would completely hide the other confidence sets.

4. New confidence intervals for the risk difference and risk ratio

Comparison of two proportions, such as the cure rates under active and control treatments or the prevalence among exposed and non-exposed subjects, is an important issue in the analysis of health data. Here we focus on confidence interval construction. Several CIs have been proposed for the difference between the two probabilities and others for the risk ratio

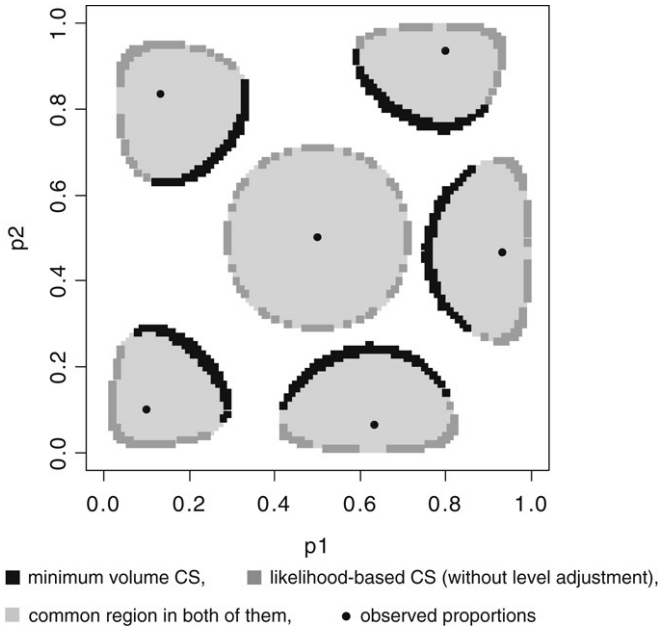

Fig. 3. Shape comparisons ($n_1 = n_2 = 30$).

and odds ratio (Newcombe, 1998b; Troendle and Frank, 2001; Agresti and Min, 2002; Dann and Koch, 2005). New exact unconditional CIs can be derived from the minimum volume test above by applying the method of intersection-union tests (Casella and Berger, 1990) and test inversion. In all three cases (risk difference, risk ratio, odds ratio) the null hypothesis is composite, so that it can be represented by a line in the unit square. According to the principle of intersection-union test, the rejection region belonging to a composite null hypothesis consists of the intersection of the rejection regions belonging to the individual simple null hypotheses. This is equivalent to considering the union of the acceptance regions as an acceptance region for the composite null hypothesis.

To see how the proposed algorithm works, consider the general composite null hypothesis $H_0 : \pi \in \Pi_0$ and the alternative $H_1 : \pi \notin \Pi_0$, where $\Pi_0 \subseteq [0, 1] \times [0, 1]$. For any $t \in [0, 1]$, define the set $A_{\Pi_0,t} = \cup_{\pi_0 \in \Pi_0} A_{\pi_0,t}$, and the function $f_{\Pi_0}(t) = \min_{\pi_0 \in \Pi_0} P_{\pi_0}(A_{\Pi_0,t})$, where $A_{\pi_0,t}$ is the set defined in (1). The function $f_{\Pi_0}(t)$, like $f_{\pi_0}(t)$ defined in (2), is monotone decreasing and left continuous. Taking $t^* = \max\{t | f_{\Pi_0}(t) \geq 1 - \alpha\}$, the set A_{Π_0,t^*} is a level $(1 - \alpha)$ acceptance region for testing the composite H_0. The CI for the difference or ratio (or other function of the parameters) follows from inverting this test.

Computationally, we first construct $A_{\pi_0,t}$ with t equal to the desired confidence level for each $\pi_0 \in \Pi_0$. Doing so, $f_{\Pi_0}(t)$ will clearly be greater or equal than the desired level. Then we reduce t in steps of 0.01 and stop if further reduction of t would bring $f_{\Pi_0}(t)$ below the desired level. Of course, the step size can be changed according to the required precision at the expense of computing time, but the presence of faster and faster computers enables a literally day-by-day increase of precision. Fig. 4 illustrates the shape of a level 0.90 two-sided acceptance region belonging to the hypothesis $H_0 : p_2/p_1 = 3$ against $H_1 : p_2/p_1 \neq 3$. The size of the test, i.e., the minimum probability of the acceptance region is 0.908, which is obtained by setting the nominal level for the individual null hypotheses to 0.77.

The test may be inverted for a CI, which means the following. Given a pair of observed proportions, the confidence limits are the minimal and maximal values of the difference, *RR*, or *OR*, for which the acceptance region contains the observed proportions. Although CIs for risk difference and ratio may also contain holes, this rarely occurs, because the acceptance region is a union of several acceptance regions, and the union will have a hole only if it appears simultaneously in each acceptance region. Note that if a hole occurs, it is filled in by the algorithm so that the CI is taken to run from one end to the other.

5. Applications

5.1. Joint exact CS for sensitivity and specificity

In a study on ultrasonographic examination for small bowel obstruction in dogs (Manczur et al., 1998), 11 of 13 dogs with mechanical ileus were diagnosed correctly (sensitivity = 85%), while obstruction was correctly excluded in 29 of 31 non-obstructive cases (specificity = 94%). A CS constructed for the two parameters will indicate the likely range of the true sensitivity and specificity.

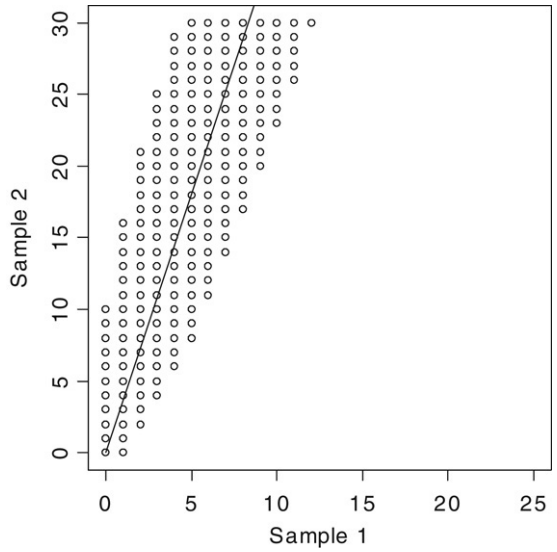

Fig. 4. Level 90% two-sided acceptance region to $H_0 : RR = 3, n_1 = 25, n_2 = 30$. The line represents the null hypothesis $RR = 3$.

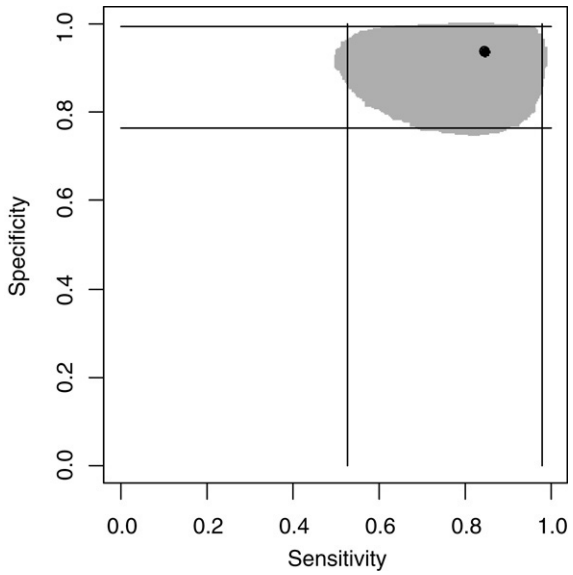

Fig. 5. Joint 95% exact confidence set (minimum volume CS) for the sensitivity and specificity of a diagnostic test (grey area). The rectangle is the 95% confidence set constructed in the usual way. The point represents the observed sensitivity and specificity.

A 95% exact joint CS resulting from the method proposed in this paper is presented in Fig. 5 (grey area). It has been constructed with a resolution of 1/500 of the parameter space. The rectangle is based on two 97.5% one-dimensional Sterne CIs (this is the usual way to construct a 95% joint CS for two parameters from two independent samples). The area of the MVCS is smaller by nearly 20% than that of the rectangular CS, and 18% smaller than that of the LBCS adjusted so that the minimum coverage reaches 95%. Without level-adjustment, the minimum coverage of the LBCS is just 89.5%, but even so, it has only 0.8% less volume than the 95% MVCS.

5.2. Exact confidence interval for the difference between prevalences

Raoul et al. (2001) studied the epidemiological status of *Echinococcus multilocularis* in foxes in France between 1996 and 1999. The prevalence is reported separately for high and low endemicity areas. Prevalence in high endemicity areas was 58.2% (39/67) and 68.9% (42/61) in the winter of 1996–97 and 1998–99, respectively, resulting in a difference in percentages of 10.7%.

We applied the proposed method to form a 95% exact unconditional CI for the true difference. The CI obtained is (−0.063, 0.269). The 95% exact unconditional CI computed by StatXact 7 is (−0.06653, 0.2726), which is 0.00713 (2.1%) longer than

our interval. In StatXact we applied the method of inverting a two-tailed test (Agresti and Min, 2001; StatXact 7 PROCs for SAS Users, 2005), which always results in a shorter CI than inverting two 1-tailed tests using the method of Chan and Zhang (1999).

5.3. Exact confidence interval for the difference between response rates

Chan (2003) re-analyses an example from Rodary et al. (1989) in which chemotherapy and radiation therapy were compared in a randomized clinical trial on nephroblastoma. The response rates were 0.9432 (83/88) and 0.9079 (69/76) in the chemotherapy and in the radiation group, respectively. Chan (2003) reports a 90% exact unconditional CI for the difference in response rates by the method of Chan and Zhang (1999) as (−0.0350, 0.1168). Our method results in a CI (−0.036, 0.112), while StatXact 7 computes a CI (−0.03861, 0.1189). Here too, inversion of a two-tailed test was chosen in StatXact. Our interval is 2.5% shorter than that of Chan and Zhang and 6% shorter than the StatXact interval.

5.4. Exact confidence interval for the risk ratio

In a study on the effects of reflux nephropathy it was investigated whether pre-existing hypertension and impaired renal function influenced the rates of preeclampsia, renal function deterioration and preterm birth (North et al., 2000). It was found that preeclampsia was increased in women with pre-existing hypertension ($8/19 = 42\%$) compared to normotensive women ($5/35 = 14\%$). This results in a point estimate of RR = 2.95. The proposed method gives a 95% exact unconditional CI for RR of (1.125, 8.030). The 95% exact unconditional CI computed by StatXact 7 by inverting a two-tailed test is (1.139, 8.336): in this example our interval is shorter by 0.292 (4.2%).

6. Discussion

There is no published method for the construction of a joint CS specific for two binomial proportions. Therefore, we compared the MVCS approach developed here with some general methods of CS construction, such as the likelihood-based CS and the rectangular CS. The MVCS has much smaller volume than the rectangular CS. Comparing it with the LBCS, we found that it shows considerably less fluctuation in coverage depending on the true parameters, so that it tends to be closer to the nominal level than the LBCS. If the LBCS is adjusted so that its minimum coverage reaches the nominal, then it has much greater volume than the MVCS. Without adjustment, the volumes are about the same, but the minimum coverage of the LBCS is subtantially below the nominal level.

Although there exist several exact unconditional tests and CIs for the risk difference (Skipka et al., 2004; Zhou et al., 2004; Andres and Tejedor, 2004; Newcombe, 1998b; Chan and Zhang, 1999; Coe and Tamhane, 1993), there are relatively few for the risk ratio (Dann and Koch, 2005; Coe and Tamhane, 1993), and for most of them computer programs are not readily available. Therefore, we compared our new intervals with those computed by StatXact 7, (incorporating the method by Agresti and Min (2001)) for both risk difference and risk ratio. This method was chosen as one of the best truly exact methods (Andres and Tejedor, 2004; Santner et al., 2007).

We made a comparison in terms of total length for small samples ($n_1 = n_2 = 10$, 15, 20, and $n_1 = 10, n_2 = 20$) at the 95% level. Total length is obtained by adding up the lengths of intervals for all possible outcomes, that is, for the whole sample space. In case of RR, total length was calculated by omitting infinite intervals: for such intervals difference in length was taken to be the difference between the lower endpoints. Total length proved to be a little less than that of the StatXact interval for both risk difference (by 0.3%, 0.6%, 0.2%, and 0.8%, respectively) and for risk ratio (by 3.6%, 1.3%, 2.7%, and 2.6%, respectively). For the risk difference the proposed CI tends to be shorter than that computed by StatXact when the observed risk difference is small, and wider when the difference is closer to ± 1. For the risk ratio, the proposed CI tends to be shorter when none of the observed probabilities is small.

R code for the MVCS, CI for difference of proportions, and CI for risk ratio is available from the first author. Running times on a notebook computer with 2GHz Intel T7200 processor, 512 MB RAM, Windows XP, and R 2.4.1 were 2, 12, 15, and 7 min for examples 5.1 to 5.4, respectively. Probably these running times can be reduced by optimizing the program code.

The application of the described methods to two hypergeometric or two Poisson variables is straightforward.

Acknowledgements

The authors are grateful to Gábor Tusnády and Zsolt Lang for many helpful discussions, and to Robert Newcombe for his valuable comments and criticism of an earlier version of the manuscript. Thanks to the anonymous reviewers whose comments also lead to several improvements. This research was supported by the Hungarian National Scientific Research Fund, grant no. OTKA T049157.

References

Agresti, A., Coull, B.A., 1998. Approximate is better than exact for interval estimation of binomial proportions. The American Statistician 52 (2), 119–126.
Agresti, A., Gottard, A., 2007. Nonconservative exact small-sample inference for discrete data. Computational Statistics and Data Analysis 51 (12), 6447–6458.

Agresti, A., Min, Y., 2001. On small-sample confidence intervals for parameters in discrete distributions. Biometrics 57 (3), 963–971.
Agresti, A., Min, Y., 2002. Unconditional small-sample confidence intervals for the odds ratio. Biostatistics 3 (3), 379–386.
Andres, A.M., Tejedor, I.H., 2004. Exact unconditional non-classical tests on the difference of two proportions. Computational Statistics and Data Analysis 45 (2), 373–388.
Casella, G., Berger, R.L., 1990. Statistical Inference. Duxbury Press, Belmont, CA.
Chan, I.S.F., 2003. Proving non-inferiority or equivalence of two treatments with dichotomous endpoints using exact methods. Statistical Methods in Medical Research 12 (1), 37–58.
Chan, I.S.F., Zhang, Z., 1999. Test-based exact confidence intervals for the difference of two binomial proportions. Biometrics 55 (4), 1202–1209.
Coe, P.R., Tamhane, A.C., 1993. Small sample confidence-intervals for the difference, ratio and odds ratio of 2 success probabilities. Communications in Statistics–Simulation and Computation 22 (4), 925–938.
Dann, R.S., Koch, G.G., 2005. Review and evaluation of methods for computing confidence intervals for the ratio of two proportions and considerations for non-inferiority clinical trials. Journal of Biopharmaceutical Statistics 15 (1), 85–107.
Gupta, R.C., Tian, S., 2007. Statistical inference for the risk ratio in 2×2 binomial trials with structural zero. Computational Statistics and Data Analysis 51 (6), 3070–3084.
Manczur, F., Voros, K., Vrabely, T., Wladar, S., Nemeth, T., Fenyves, B., 1998. Sonographic diagnosis of intestinal obstruction in the dog. Acta Veterinaria Hungarica 46 (1), 35–45.
Munk, A., Skipka, G., Stratmann, B., 2005. Testing general hypotheses under binomial sampling: The two-sample case — asymptotic theory and exact procedures. Computational Statistics and Data Analysis 49 (3), 723–739.
Newcombe, R.G., 1998a. Two-sided confidence intervals for the single proportion: Comparison of seven methods. Statistics in Medicine 17 (8), 857–872.
Newcombe, R.G., 1998b. Interval estimation for the difference between independent proportions: Comparison of eleven methods. Statistics in Medicine 17 (8), 873–890.
North, R.A., Taylor, R.S., Gunn, T.R., 2000. Pregnancy outcome in women with reflux nephropathy and the inheritance of vesico-ureteric reflux. Australian and New Zealand Journal of Obstetrics and Gynaecology 40 (3), 280–285.
Raoul, F., Deplazes, P., Nonaka, N., Piarroux, R., Vuitton, D.A., Giraudoux, P., 2001. Assessment of the epidemiological status of *Echinococcus multilocularis* in foxes in France using ELISA coprotests on fox faeces collected in the field. International Journal for Parasitology 31 (14), 1579–1588.
Reiczigel, J., 2003. Confidence intervals for the binomial parameter: Some new considerations. Statistics in Medicine 22 (4), 611–621.
Rodary, C., Com-Nougue, C., Tournade, M., 1989. How to establish equivalence between treatments: A one-sided clinical trial in paediatric oncology. Statistics in Medicine 8 (5), 593–598.
Santner, T.J., Pradhan, V., Senchaudhuri, P., Mehta, C.R., Tamhane, A., 2007. Small-sample comparisons of confidence intervals for the difference of two independent binomial proportions. Computational Statistics and Data Analysis 51 (12), 5791–5799.
Skipka, G., Munk, A., Freitag, G., 2004. Unconditional exact tests for the difference of binomial probabilities — contrasted and compared. Computational Statistics and Data Analysis 47 (4), 757–773.
StatXact 7 PROCs for SAS Users. Cytel Software Corporation. Cambridge, MA, USA. 2005.
Sterne, T.E., 1954. Some remarks on confidence or fiducial limits. Biometrika 41 (1–2), 275–278.
Troendle, J.F., Frank, J., 2001. Unbiased confidence intervals for the odds ratio of two independent binomial samples with application to case-control data. Biometrics 57 (2), 484–489.
Zhou, X.H., Tsao, M., Qin, G., 2004. New intervals for the difference between two independent binomial proportions. Journal of Statistical Planning and Inference 123 (1), 97–115.