SIMULTANEOUS TEST AND CONFIDENCE SET FOR TWO BINOMIAL PROPORTIONS

ZS. ABONYI-TÓTH., J. REICZIGEL

Szent István University, Faculty of Veterinary Science, Department of Biomathematics and Informatics

e-mail: Abonyi.Zsolt@aotk.szie.hu, Reiczigel.Jeno@aotk.szie.hu

(Received 10th Sep 2005, accepted 10th Oct 2006)

Abstract. Some phenomenons are modelled most naturally by two probabilities, because use of a single combined measure would result in an undesirable loss of information. E.g. diagnostic tests are obligatorily characterized by their sensitivity and specificity, risk of disease is often reported for two subpopulations e.g. males and females rather than for the whole population, etc. Here we present a statistical test and a related method to construct simultaneous two-dimensional confidence sets for two probabilities estimated from independent samples. We also describe a computer algorithm for these calculations.
Keywords: *diagnostic test, confidence, statistics*

Introduction

Suppose we have observed two binomial variables on two independent samples and we want to carry out a statistical test for the binomial parameters (probabilities) with H0: p1=p1(hypot) , p2=p2(hypot). Also, we want to give a simultaneous confidence set.

This problem may come up in situations when a phenomenon cannot be reasonable modelled by a single probability. This is the case e.g. if sensitivity and specificity of a diagnostic test are to be estimated, for sensitivity is calculated from a sample of people having a disease, while specificity is calculated from a sample of people not having the disease. Another example is when risk of a disease is needed separately for males and females, or urban and rural people, etc. If these two risks were combined to have a single measure (e.g. risk for the whole population), then it would be impossible to control for differences in sex ratio or urban/rural ratio between different populations or over time.

Concerning hypothesis testing, a typical question may be whether both sensitivity and specificity of a new diagnostic test exceeds those of a standard test with known sensitivity and specificity, say Se = 92% (p1(hypot)) and Sp = 87% (p2(hypot)), or whether a certain public health measure reduced the risks compared to the preceeding period of time in both parts of the population.

The solution is based on the method suggested by Sterne [5] for the one-dimensional case. The acceptance region is constructed taking the points of the two-dimensional sample-space in descending order of their probabilities (first the one with highest probability, then the one with second highest, etc.) until the probability of the acceptance region reaches the desired level, e.g. 95%. The acceptance region defined this way has a minimal area (it contains the fewest points among all possible regions at the same level). The confidence set can be constructed by inverting the test [4]. This

http://www.ecology.uni-corvinus.hu ● ISSN 1589 1623

means that the confidence set will contain all parameter pairs, for which the observed outcome is in the acceptance region at the level of interest.

The test and confidence region can be used to create exact tests and confidence intervals for functions of the two parameters like e.g. difference of proportions, relative risk, or odds ratio. Exact solutions for these problems are not always readily available [1][3].

The first idea of the computerized solution is to evaluate the whole parameter-space, i.e. [0,1]x[0,1] with a step size providing the desired precision. It needs a lot of computing time, because we have to construct acceptance regions for several pairs of parameters. Our paper describes the possible optimizations of the algorithm and introduces the first version of the program.

The problem

We have two independent samples of size n1 and n2 (each should be less than 300 in the present version of the program) with observed numbers of successes k1 and k2, respectively. What could be the probabilities p1 and p2?

The confidence interval will be created using loops. Given the step size (ss), the program calculates the acceptance region for the probabilities p1 and p2, where p1=i/ss (i=0,1,…,ss) and for every p1 value p2=j/ss (j=0,1,…,ss). Those (p1, p2) pairs are regarded to belong to the confidence region, for which the acceptance region contains the pair (k1, k2). During this process, functions of p1 and p2 are also calculated in order to create confidence intervals for them.

It's clear that (ss+1)2 acceptance regions must be computed. For each of them, we have to calculate the appropriate two-dimensional probability mass function (n1·n2 calculations), order the probabilities, and select them until we reach the desired level.

Of course it is easy to write a program to do this, but as we increase the sample sizes and the step size, the computing time will make the program inapplicable. That's why we have to optimize the program code.

Optimisation

There are several ways to optimize an algorithm and/or the corresponding program code. We used more memory to store temporary results, which were needed several times. We also simplified calculations using theoretical knowledge on the problem.

Important part of optimisation is to eliminate unrelevant points from calculations. For example we have to calculate the acceptance region (ss+1)2 times. Of course it is very important to do it as fast as possible.

With given n1 and n2, the number of possible pairs of outcomes is (n1+1)·(n2+1), that is, the size of the sample space is proportional to the square of the sample size. Calculation of the probability mass function for each pair slows down the program considerably. But we can recognize that, as the sample size increases, the acceptance region will contain just a few percent of the possible pairs. Examining the probability mass function it is not surprising, most probabilities are very close to zero. Selecting a good algorithm, we don't need to calculate the probability mass function for the whole parameter space, it's enough to calculate the highest probabilities to get the acceptance

http://www.ecology.uni-corvinus.hu ● ISSN 1589 1623

region. In this case, we need to know, where we can find higher probabilities without calculating all of them.

Another way of speeding up the program is to recognize that we don't need the whole acceptance region. If we have reached the observed k1 and k2, we can stop.

The program

Our algorithm is using the above 'tricks' to work as fast as possible. We have used the Borland Delphi 2.0 Desktop programming environment to compile a working version, but the algorithm itself is language and platform independent and can be implemented in any programming environment.

After optimization, the computing time is 20 seconds on a 2.6GHz P4 computer for n1=n2=100, stepsize=1000. The maximum value of sample sizes is 300, the maximum of step size is 100000.

The result is written into a space separated text file with a name generated from input data to make it easier to find a result later.

The output of the program consists of four windows. The log window summarizes the results: it shows the file name, input parameters and processing time. The results window displays the results from the results file. The drawing window displays a plot of the acceptance or confidence region denoting points belonging to the region by 'X' and others by '.'. It demonstrates that the confidence interval is generally not convex. It can have both holes and disjoint areas. The main reason for this is the discreteness of the binomial distribution. For practical purposes one can reasonably use the convex hull of the computed confidence set. The last window shows the confidence intervals for some functions of the parameters.

The program is able to give some partial results:

Probability mass function

For given n1, n2, p1, p2 the program calculates the two dimensional binomial probability mass function. As we have shown above, it first calculates the arrays of binomial coefficients, the powers of pi, (1-pi), then calculates the probability of each (k1, k2) pairs.

We were able to save time here by avoiding use of the built-in power function and simply multiplying the previous element in the array with the base of the power to get the next one in a cycle.

Acceptance region

For given n1, n2, p1, p2 the program calculates the acceptance region. The construction is the following: the program calculates the probability mass function, sorts the probabilities of the different (k1, k2) pairs in descending order, and adds these pairs to the region until the sum of the probabilities is under the pre-specified level. If several pairs have the same probability, all of them are added to the area together. It is necessary because the acceptance region must be well-defined and there is no other reasonable rule about which pair should be included in the acceptance region from those having the same probability.

The acceptance region defines a statistical test for H0: (p1 = p1(hypot) , p2 = p2(hypot)). If the observed (k1, k2) pair is not in the area, H0 is rejected. The p-value

http://www.ecology.uni-corvinus.hu ● ISSN 1589 1623

belonging to a certain (k1, k2) pair can be calculated stopping the process when it adds the pair to the acceptance region: then p = 1- sum of the probabilites of pairs in the area.

The algorithm doesn't calculate the probability of each sample point. First it determines a starting point: the sample point having the highest probability. It is ([n1·p1],[n2·p2]) or one of its neighbors, where [x] represents the biggest whole number, which is not greater than x. It means 9 pairs, their probabilities will be calculated and they will be added to the array of neighbors, nei. Then the following general step is repeated: The program looks for the maximum probability in nei, adds the points that have this probability to the acceptance region, deletes them from nei, and adds all their neighbors to nei. This step is repeated until the stop condition will be true: the program reaches either the pre-specified level or the observed (k1, k2) pair.

The two-dimensional binomial distribution is strictly monotonically decreasing in each direction away from its peak, so the above algorithm will really produce the acceptance region, because the biggest probability not in the area must be a neighbor of the area. The acceptance region looks like an ellipse, and nei will contain plots as the outline of the ellipse.

Confidence region

For given n1, n2, k1, k2, pv, ss the program calculates the confidence set, which consist of those (p1, p2) pairs, for which (k1, k2) is in the pv-level acceptance area belonging to (p1, p2) at a step size of ss.

Figure 1 shows the structure of the confidence region for n1=n2=20, k1=k2=1, pv=0.95, stepsize=250. The top left corner of the picture is (p1, p2)=(0, 0) and the bottom right corner is (p1, p2)=(0.292, 0.288). Holes are visible in the figure.

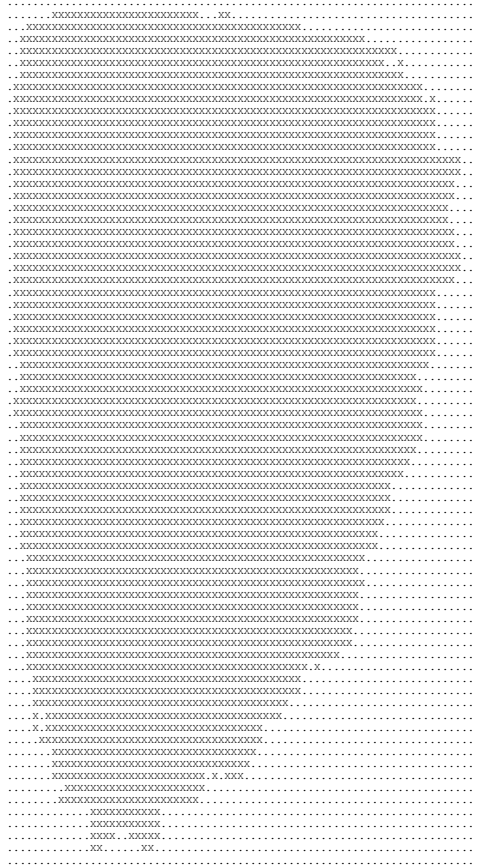

Figure 1. *Shape of the confidence set*

http://www.ecology.uni-corvinus.hu ● ISSN 1589 1623

In the general case the program has to determine the acceptance region for p1 and p2 pairs, where p1=i/ss (i=0,1,…,ss) and p2=j/ss (j=0,1,…,ss), so the previously described process should run (ss+1)2 times. It needs a lot of calculation, and most pairs won't belong to the confidence set. That's why the program first scans one row and one column only, where the area is possibly the widest. The optimal places are appriximately at k1/n1 and k2/n2. Unfortunately, the confidence set is not convex and may have holes and disjoint parts. If ki is close to zero or to ni, it is very hard to estimate, which row and column is best to scan; this part of the algorithm should still be refined.

During those scans the program stores the minimum and maximum value of pi belonging to the confidence region, called pi1 and pi2. If we were sure, these were the most extreme points of the region, it would be enough to check the [p11 , p12]x[p21 , p22] area. Unfortunately it is not true, so the rectangle has been expanded by ss/40 in every direction. This amount is enough to surely discover the whole confidence set, but the program checks a lot of irrelevant pairs as well. This part of the algorithm should be refined too.

During the calculation of the confidence set, it is possible to compute confidence intervals for p1, p2, p1-p2, p1/p2 and odds(p1)/odds(p2). The program stores their minimum and maximum values during the process, and replaces them, if it is necessarily. Finally the variables contain the global minimum and maximum of these functions of the probabilities together with the points of the parameter space where these values were taken.

Application

The method was applied to construct a two-dimensional confidence set for sensitivity and specificity of transrectal ultrasonography for pregnancy testing in ewes on 25 to 30 days of gestation [2]. Samples consisted of 34 pregnant and 50 non-pregnant ewes, in which the number of correct diagnosis was 11 and 46, respectively. This results in the estimates Se = 32.3%, Sp = 92.0%.

Figure 2 shows the structure of the confidence region for n1=34, n2=50, k1=11, k2=46, pv=0.95, stepsize=150. The top left corner of the picture is (p1, p2)=(0.14, 0.76) and the bottom right corner is (p1, p2)=(0.547, 0.987).

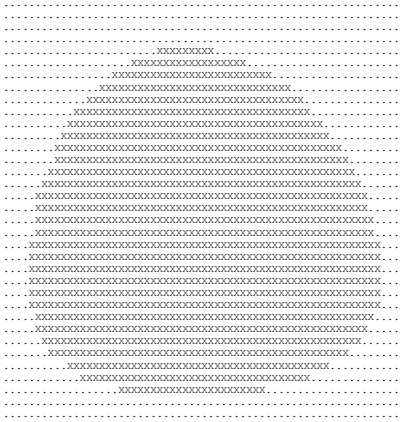

Figure 2. *Shape of the confidence set in this application*

Discussion

As we decribed, we have developed an algorithm and a computer program, which is able to make a simultaneous statistical test for the binomial parameters (probabilities)

http://www.ecology.uni-corvinus.hu ● ISSN 1589 1623

where (H0: p1 = p1(hypot) , p2 = p2(hypot)), and to give a simultaneous confidence set from observed values of two independent samples.

Based on several optimization steps, the program is now fast enough to produce useful results for sample sizes up to 300 and with a precision of 0.00001. The program is available on the first autor's home page (www.univet.hu/users/atzs).

Acknowledgment. The project is sponsored by grant OTKA T049157.

REFERENCES

[1] Agresti A (2003): Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact – Statistical Methods in Medical Research 12 (1): 3-21.

[2] Karen A, Szabados K, Reiczigel J, Beckers JF, Szenci O (2004): Accuracy of transrectal ultrasonography for determination of pregnancy in sheep: effect of fasting and handling of the animals – Theriogenology 61 (7-8): 1291-1298

[3] Newcombe R.G. (1998): Interval estimation for the difference between independent proportions: Comparison of eleven methods – Statistics in Medicine 17 (8): 873-890.

[4] Reiczigel J. (2003): Confidence intervals for the binomial parameter: some new considerations – Statistics in Medicine 22 (4): 611-621.

[5] Sterne, T. E. (1954): Some remarks on confidence or fiducial limits – Biometrika 41 (1-2): 275-278.

http://www.ecology.uni-corvinus.hu ● ISSN 1589 1623