Binomial Distribution Sample Confidence Interval Estimation for Positive and Negative Likelihood Ratio Medical Key Parameters

Sorana BOLBOACĂ [a], Lorentz JÄNTSCHI [b]

[a] *Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania, http://sorana.academicdirect.ro*
[b] *Technical University of Cluj-Napoca, Romania, http://lori.academicdirect.org*

Abstract

Likelihood Ratio medical key parameters calculated on categorical results from diagnostic tests are usually express accompanied with their confidence intervals, computed using the normal distribution approximation of binomial distribution. The approximation creates known anomalies, especially for limit cases. In order to improve the quality of estimation, four new methods (called here *RPAC*, *RPAC0*, *RPAC1*, and *RPAC2*) were developed and compared with the classical method (called here *RPWald*), using an exact probability calculation algorithm.

Computer implementations of the methods use the PHP language. We defined and implemented the functions of the four new methods and the five criterions of confidence interval assessment. The experiments run for samples sizes which vary in 14 – 34 range, 90 – 100 range ($0 < X < m$, $0 < Y < n$), as well as for random numbers for samples sizes ($4 \leq m, n \leq 1000$) and binomial variables ($1 \leq X, Y < m, n$).

The experiment run shows that the new proposed RPAC2 method obtains the best overall performance of computing confidence interval for positive and negative likelihood ratios.

Keywords

Confidence intervals; Binomial Distribution; Likelihood ratios

Introduction

Confidence intervals defines as *an estimated range of values that is likely to include an unknown population parameter, the estimated range being calculates from a given set of sample data* is used nowadays as a criterion of assessment of the trustworthiness or robustness of the finding [1]. If independent sample are take repeatedly from same population, and the confidence interval is calculated for each sample, then a certain percentage (called confidence level) of the interval will include the unknown population parameter. Confidence interval is usually computed for the percentage of 95. However, it can be produced 90%, 99%, 99.9% confidence intervals.

The main aim of a diagnostic study is to generate new knowledge which to be used in diagnostic decision process. The magnitude of the effect size of a diagnostic test can be measure in a variety of ways such as sensibility, specificity, overall accuracy, predictive values, and likelihood ratios [2,3]. Using confidence intervals associate to a diagnostic key parameter gives possibility to physicians to be more certain about the clinical value of the diagnostic test and to decide to what degree can rely on the results [4].

Likelihood ratios are alternative statistics for summarizing diagnostic accuracy which can be computed based on categorical variable, organized in a 2 by 2 contingency table [5]. The likelihood ratios, incorporate both the sensitivity and specificity of the diagnostic test providing a direct estimator of how much a test result will change the odds of having a disease [6-8].

The probability that a person with a disease to have a positive examination divided by the probability that a person without the disease to have a positive examination defines the Positive Likelihood Ratio (LR+). The probability that a person with a disease to have a negative examination divided by the probability that a person without the disease to have a negative examination defines Negative Likelihood Ratio (LR-).

The point estimation of likelihood ratios come with its confidence intervals when are reported as study results. Until now, confidence intervals of likelihood ratios calculations use the asymptotic method (called here *RPWald*) which is well known that provide too short confidence intervals [9, 10].

The aim of the paper is to introduce four new methods (called here *RPAC*, *RPAC0*, *RPAC1*, and *RPAC2*) for likelihood ratios confidence intervals estimation, and based on binomial distribution sample hypothesis to make a comprehensive study of the estimation results comparing them with also the asymptotic method (called here *RPWald*).

Materials and Methods

The normal distribution was first introduced by De Moivre in an unpublished memorandum, later published as part of [11] in the context of approximating certain binomial distribution for large sample sizes *n*. His result has extended by Laplace and is known as the *Theorem of De Moivre-Laplace*. The normal approximation of the binomial distribution is the most known method used to calculate binomial distribution based estimators.

Confidence intervals estimations for proportions using normal approximation have been commonly uses for analysis of simulation for a simple fact: the normal approximation is easiest to use in practice comparing with other distributions [12].

Our approach started with constructing of an algorithm, which use the binomial distribution hypothesis in order to calculate the exact probabilities of wrong for the choused estimator: confidence interval.

One module of the program calculates exact probabilities *X* for a sample of size *m*. The module serves for exact probabilities calculation of a two-dimensional sample (*X*, *Y*) of volumes (*m*, *n*).

Other set of algorithms implements the calculation of a set of confidence intervals formulas for Likelihood Ratio medical key parameters.

The Positive (LR+) and Negative (LR-) Likelihood Ratio medical key parameters calculations use the next formulas, where *a* = real positive (cases); *b* = false positive; *c* = false negative; and *d* = real negative:

$$LR+ = LR_{+}(a,b,c,d) = \frac{\frac{a}{a+c}}{1-\frac{d}{b+d}} = \frac{\frac{X}{m}}{1-\frac{n-Y}{n}} = \frac{X}{m}\cdot\frac{n}{Y} = LR^{+}(X,m,Y,n) \quad (1)$$

$$LR- = LR_{-}(a,b,c,d) = \frac{1-\frac{a}{a+c}}{\frac{d}{b+d}} = \frac{1-\frac{m-X}{m}}{\frac{Y}{n}} = \frac{X}{m}\cdot\frac{n}{Y} = LR^{-}(X,m,Y,n) \quad (2)$$

where:

- The proper substitutions for equation (1): $X = a$ and $Y = b$ independent binomial distribution variables; $m = a + c$ and $n = b + d$ are samples sizes;
- The proper substitutions for equation (2): $X = c$ and $Y = d$ independent binomial distribution variables; $m = a + c$ and $n = b + d$ are samples sizes;

Thus, from mathematic point of view, positive likelihood ratio, and negative likelihood ratio are of same function-type. Let us call *RP* the expression:

$$RP = RP(X,m,Y,n) = \frac{X}{m}\cdot\frac{n}{Y} \quad (3)$$

The following formula was used to compute the classical Wald type confidence interval:

$$RPWald(X,m,Y,n,z) = = RP\cdot\exp\left(\pm z\sqrt{\frac{m-X}{X\cdot m}+\frac{n-Y}{Y\cdot n}}\right) \quad (4)$$

Two Agresti-Coull correction types were applied to (4):

$$ACType2(X,m,Y,n,c_1,c_2) = RPWald(X+c_1,m+2c_1,Y+c_2,n+2c_2,z) \quad (5)$$

$$ACType1(X,m,Y,n,c) = RPWald(X+c,m+2c,Y+c,n+2c,z) \quad (6)$$

where ACType2 has two corrections (c_1 and c_2) and ACType1 has only one ($c = c_1 = c_2$).

Our proposed confidence interval estimators are (7-10):

$$RPAC(X,m,Y,n) = ACType2\left(X,m,Y,n,\frac{1}{2\sqrt{m}},\frac{1}{2\sqrt{n}}\right) \quad (7)$$

$$RPAC0(X,m,Y,n) = ACType1\left(X,m,Y,n,\sqrt{\frac{X}{m}\cdot\frac{Y}{n}\cdot\frac{1}{4}}\right) \quad (8)$$

$$RPAC1(X,m,Y,n) = ACType1\left(X,m,Y,n,\sqrt{\frac{X+1}{m}\cdot\frac{Y+1}{n}\cdot\frac{1}{4}}\right) \quad (9)$$

$$\mathrm{RPAC2}(X, m, Y, n) = \mathrm{ACType1}\left(X, m, Y, n, \sqrt{\frac{X+2}{m}\cdot\frac{Y+2}{n}\cdot\frac{1}{4}}\right) \quad (10)$$

Five criterions of confidence interval assessment methods were defined in order to be used for method comparisons:

- The average of experimental errors, *AE = Av(Err)*:

$$\mathrm{AE} = \frac{\sum_{X=1}^{m-1}\sum_{Y=1}^{n-1}\mathrm{Err}(X, Y, m, n)}{(m-1)(n-1)} \quad (11)$$

- The standard deviation of the experimental errors, *SDE = StdDev(Err)*:

$$\mathrm{SDE} = \left(\frac{\sum_{X=1}^{m-1}\sum_{Y=1}^{n-1}\left(\mathrm{Err}(X, Y, m, n) - \mathrm{AE}\right)^2}{(m-1)(n-1)-1}\right)^{1/2} \quad (12)$$

- The average of absolute difference between the experimental errors for *m*, *n* with all possible binomial variables (*1 ≤ X, Y ≤ m-1, n-1*), and the average of the experimental errors, *AADE = AvAD(Err)*:

$$\mathrm{AADE} = \frac{\sum_{X=1}^{m-1}\sum_{Y=1}^{n-1}\left|\mathrm{Err}(X, Y, m, n) - \mathrm{AE}\right|}{(m-1)(n-1)-1} \quad (13)$$

- The average of absolute difference between the experimental error for *m*, *n* with all possible binomial variables (*1 ≤ X, Y ≤ m-1, n-1*) and the imposed value, equal here with 100·α, *AADIE = AvADI(Err)*:

$$\mathrm{AADIE} = \frac{\sum_{X=1}^{m-1}\sum_{Y=1}^{n-1}\left|\mathrm{Err}(X, Y, m, n) - 100\cdot\alpha\right|}{(m-1)(n-1)} \quad (14)$$

- The deviation of experimental errors relative to the imposed significance level α, *DIE* = DevI(Err):

$$\mathrm{DIE} = \left(\frac{\sum_{X=1}^{m-1}\sum_{Y=1}^{n-1}\left(\mathrm{Err}(X, Y, m, n) - 100\cdot\alpha\right)^2}{(m-1)(n-1)}\right)^{1/2} \quad (15)$$

The Err function uses the binomial distribution hypothesis for both *X* and *Y* variables to collect all percentage probabilities that function values are outside of confidence interval.

For the *X* binomial variable, the appearance probability of the *XX* value from a sample of *m* is:

$$\mathrm{dBin}(m, X, XX) = \frac{m!}{XX!(m-XX)!}\cdot\left(\frac{X}{m}\right)^{XX}\cdot\left(1-\frac{X}{m}\right)^{m-XX} \quad (16)$$

Using (16) and supposing that the lower bound of confidence interval is given by ci8L = ci8L(X,m,Y,n) and the upper bound of confidence interval is given by ci8U = ci8U(X,m,Y,n) the Err function for the ci8 = (ci8L, ci8U) confidence interval calculation function (method) is:

$$\mathrm{Err}(X, m, Y, n) = \left(\sum_{ci8L(XX,YY,m,n)>RP(X,Y,m,n)}\mathrm{dBin}(m, X, XX)\cdot\mathrm{dBin}(n, Y, YY) + \sum_{ci8U(XX,YY,m,n)<RP(X,Y,m,n)}\mathrm{dBin}(m, X, XX)\cdot\mathrm{dBin}(n, Y, YY)\right) \Big/ \sum_{XX=1}^{m-1}\sum_{YY=1}^{n-1}\mathrm{dBin}(m, X, XX)\cdot\mathrm{dBin}(n, Y, YY) \quad (17)$$

In order to obtain a 100·(1-α) = 95% confidence interval, the experiments had run for a significance level of *α* equal with 5%. The performance of each method was assessed using the above-describe criterions (*AE*, *SDE*, *AADE*, *AADIE*, *DIE*) for samples sizes (*m*, *n*) which varies from specified ranges and different values of binomial variables (*X*, *Y*) and in 200 random sample sizes *m*, *n* (4 < *m*, *n* < 1000) and random binomial variables *X*, *Y* (*0 < X, Y < m, n*).

All described formulas (3-17) was modeled into separate algorithms and implemented in a PHP program. The output of the program produced the results.

Results

On 441 distinct pairs of samples with sizes in 14-34 range (14 ≤ *m*, *n* ≤ 34, table 1), for 110 distinct pairs in 90-100 range (table 2), for all *X* and *Y* (0 < *X* < *m*, 0 < *Y* < *n*), and for 200 random values (4 < *m*, *n* < 1000, *0 < X, Y < m, n*, see table 3) the statistical operators defined by equations (11-15) have been applied. Averages of the results are in tables (1 to 3).

Table 1. Samples sizes varying in 14 - 34 range

Method	Average of				
	AE	SDE	AADE	AADIE	DIE
RPWald	4.195	1.411	0.882	1.192	1.634
RPAC	4.220	1.262	0.874	1.132	1.485
RPAC0	4.157	1.222	0.864	1.141	1.485
RPAC1	4.166	1.226	0.870	1.140	1.484
RPAC2	4.175	1.229	0.876	1.137	1.481

Table 2. Samples sizes varying in 90 - 100 range

Method	Average of				
	AE	SDE	AADE	AADIE	DIE
RPWald	4.613	0.162	0.106	0.127	0.194
RPAC	4.641	0.148	0.096	0.119	0.178
RPAC0	4.633	0.144	0.096	0.118	0.176
RPAC1	4.635	0.144	0.095	0.117	0.176
RPAC2	4.638	0.145	0.095	0.118	0.176

Table 3. Random values

Method	AE	SDE	DIE	AADIE	AADE
RPWald	5.150	2.210	2.210	0.500	0.595
RPAC	5.041	1.264	1.262	0.383	0.402
RPAC0	5.038	1.226	1.223	0.395	0.414
RPAC1	4.972	0.836	0.834	0.330	0.316
RPAC2	4.949	0.786	0.786	0.312	0.292

Discussions

Looking at the results of the experiment for samples sizes which vary from 14 to 34 (table 1) it can be observed that the values of averages of experimental errors obtained with all methods are closed to each other, but *RPAC* method obtains the closest value to the expected value (100·α). It is observing that the *RPWald* method is the single one that obtains values greater than expected value. For *SDE* criterion the *RPWald* method obtain the greater value (1.411) showing us that the experimental errors are widely spread by each other compared with the values obtain with *RPAC0*, *RPAC1, RPAC2*, and *RPAC* methods (1.222, 1.226, 1.229, and 1.262). The *RPAC0* method obtains the less average of *AADE* while the *RPWald* obtains the greater value (0.882). The *RPAC* method, closely followed by the *RPAC2* method obtains the lowest average of *AADIE* (1.132, respectively 1.137) showing us that the experimental errors obtained with specified methods are more close to the expected value comparing with *RPAC1*, *RPAC0,* and *RPWald* methods.

The deviation of experimental errors relative to the imposed significance level α criterion of assessment can be consider the best criterion because shows us the variability of the data relative to the imposed significance level. A larger deviation of experimental errors relative to the imposed significance level reveals that the values are widely spread out relative to the expected value. The lowest deviation of experimental errors relative to the imposed significance level α is obtaining by the *RPAC2* method (1.481, table 1). The *RPAC2* method has closely followed by the *RPAC1* method (1.484), *RPAC0* and *RPAC* methods (1.485). The deviation of experimental errors relative to the imposed significance level α decrease with the increasing of sample sizes *m*, and *n* for all implemented methods and the *RPWald* method present the widely spread out experimental errors.

When the samples sizes vary from 90 to 100 (table 2), the results of the experiment are rather similar with the one for samples sizes varying from 14 to 34: the *RPAC* method obtains the average of *AE* more close to the expected value (100·α). The *RPAC0* and *RPAC1* methods obtain the lowest average of *SDE* while *RPWald* method obtains the greatest average of *SDE* showing a widely spread out of values comparing with other methods. For *AADE* criterion, the *RPAC2* and *RPAC1* obtain the same values of average, equal with 0.095 (table 2), closely followed by *RPAC* and *RPAC0* methods (0.096). The *RPAC1* method, closely followed by the *RPAC2*, *RPAC0* and *RPAC* methods obtain the lowest average of *AADIE* (1.117, 1.118, 1.118, respectively 1.119) showing us that the experimental errors obtain with specified methods are more close to the expected value comparing *RPWald* method.

The lowest deviation of experimental errors relative to the imposed significance level α has been obtained by the *RPAC0*, *RPAC1*, and *RPAC2* methods (0.176, table 2), closely followed by the *RPAC* method (0.178), showing us that the experimental errors obtain by the above describe methods are not spread out as the ones obtained with the *RPWald* method.

From the experimental results, when sample sizes vary fro 90 to 100 it can be observe that the average of *AE* increase with increasing of samples sizes but never exceed the expected value (table 2). Opposite, the average of *SDE* and respectively *DIE* decrease with increasing of samples sizes. This observation sustain that with increase of samples sizes the experimental values are closest by each other.

Looking at the results obtained from the random experiment (200 random numbers for samples sizes $4 \leq m, n \leq 1000$ and binomial variables $1 \leq X \leq m-1$, and $1 \leq Y \leq n-1$, table 3) it can be observe that *RPAC1* method (4.972), closely followed by the *RPAC2* method (4.949) obtain an average of *AE* more close to expected value. The *RPWald, RPAC,* and *RPAC0* methods exceed the expected value of averages of *AE*. For all criterions, the *RPAC2* method obtains systematically the best results, showing us that the *RPAC2* method is the best method of computing confidence interval for *RP* function-type.

The averages of statistical operators used in experiments obtained by the *RPAC*, *RPAC0*, *RPAC1*, and *RPAC2* are close to each other even if we look at the sample sizes which vary in 14 - 34 range or which

vary in 90 - 100 range. This characteristic cannot be observe if we look at the results from random samples sizes ($4 \le m, n \le 1000$) and random binomial variables ($1 \le X \le m-1$, and $1 \le Y \le n-1$). The best performances in computing confidence interval for *RP* function-type is the *RPAC2* method. The *RPAC2* method systematical obtain the lowest deviation of the average of experimental errors relative to the imposed significance level even if the samples sizes vary from 14 to 34, from 90 to 100 or are random selected samples sizes ($4 \le m, n \le 1000$) and random binomial variables ($1 \le X \le m-1$, and $1 \le Y \le n-1$).

Conclusions

All new methods of computing the confidence interval for *RP* function-type (*RPAC*, *RPAC0*, *RPAC1*, and *RPAC2*) are superior comparing with the asymptotic method (*RPWald*).

The differences between the proposed methods of computing confidence interval for *RP* function-type are situating on a scale of small to very small differences and there are situations in that one method is better than other methods. The *RPAC* method obtain almost systematic best average of *AE* for samples sizes which varying in 14 – 34 and respectively in 90 – 100 ranges. The *RPAC0* method obtain the lowest average of *SDE* for samples sizes which vary in14 – 34 range, while the *RPAC1* the best values for average of *AADE* and *AADIE* when samples sizes vary in 90 – 100 range. Systematic, the *RPAC2* method obtain the best deviation of experimental errors relative to the imposed significance level even if we looked at samples sizes which vary in 14 – 34 and respectively in 90 – 100 ranges or at random samples sizes and random binomial variables.

The best criterion of comparing the confidence interval methods is deviation relative to the imposed significance level.

Using deviation relative to the imposed significance level criterion, the *RPAC2* method is the best method of computing confidence interval for *RP* function-type in random samples and random binomial variables ($4 \le m, n \le 1000$, and $1 \le X, Y < m, n$) and overall for all $14 \le m, n \le 34$, $90 \le m, n \le 100$ and $0 < X, Y < m, n$.

Based on above conclusions, we recommend the use of *RPAC2* method for computing of the confidence interval of positive and negative likelihood ratio instead of use of *RPWald* method.

References

[1]. Huw D. What are confidence intervals? What is…. Hayward Group Publication. 2003;3:1-9.

[2]. Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. BMJ. 1994;308:1552.

[3]. Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ. 1994;309:102.

[4]. Medina LS, Zurakowski D. Measurement Variability and Confidence Intervals in Medicine: Why Should Radiologists Care?. Radiology. 2003;226:297-301.

[5]. Deeks JJ, Altman GD. Diagnostic tests 4: likelihood ratios. BMJ. 2004;329:168-169.

[6]. Sackett D, Straus ES, Richardson WS, Rosenberg W, Haynes RB. Diagnosis and screening, chapter in: Evidence-based Medicine: How to Practice and Teach EBM. 2nd ed. Edinburgh, Churchill Livingstone. 2000, pp. 67-93.

[7]. Achimaș Cadariu A. Diagnosis Test, Chapter in: Medical Research Methodology. "Iuliu Hațieganu" University of Medicine and Pharmacy Publishing House, Cluj-Napoca. 1999, pp. 29-38 (in Romanian).

[8]. Black WC, Armstrong P. Communicating the significance of radiological test results: the likelihood ratio. Am. J. Roentgenol. 1986;147:1313-8.

[9]. Drugan T, Bolboacă S, Jäntschi L, Achimaș Cadariu A. Binomial Distribution Sample Confidence Intervals Estimation 1. Sampling and Medical Key Parameters Calculation. Leonardo Electronic Journal of Practices and Technologies. 2003;3:45-74.

[10]. Hamm RM. Clinical Decision Making Spreadsheet Calculator, University of Oklahoma Health Sciences Center, available at: http://www.emory.edu/WHSC/MED/EMAC/curriculum/diagnosis/oklahomaLRs.xls

[11]. Abraham Moivre. The Doctrine of Chance: or The Method of Calculating the Probability of Events in Play. W. Pearforn, Second Edition, 1738.

[12].Pawlikowski KDC, McNickle GE. Coverage of Confidence Intervals in Sequential Steady-State Simulation. Simulation Practice and Theory. 1998;6:255-67.