Binomial Distribution Sample Confidence Intervals Estimation
4. Post Test Probability

Sorana BOLBOACĂ, Andrei ACHIMAŞ CADARIU

"Iuliu Haţieganu" University of Medicine and Pharmacy, Cluj-Napoca, Romania

sbolboaca@umfcluj.ro

Abstract

Posttest probability is one of the key parameters which can be measured and interpret in dichotomous diagnostic tests. Defined as the proportion of patients, which present particular test result and the target disorder, the posttest probability is a parameter used in assessing the efficiency of the diagnostic. As a point estimated parameter, posttest probability needs a confidence interval in order to interpreting trustworthiness or robustness of the finding. Unfortunately, for post test probability there was no confidence intervals reported in literature. The aim of this paper is to introduce six methods named *Wilson, Logit, LogitC, BayesF, Jeffreys,* and *Binomial* as methods of computing confidence intervals for posttest probability and to present theirs performances.

Computer implementations of the methods use the PHP language. The performance of each method for different sample sizes and different values of binomial variable was asses using a set of criterions. One criterion was the average of experimental errors and standard deviations. Second, the deviation relative to imposed significance level ($\alpha = 5\%$). Third, the behavior of the methods when the sample size vary from 4 to 103 and on random sample and random binomial variable in 4..1000 domain.

The results of the experiments show us that the *Binomial* method obtain the

best performances in computing the confidence intervals for posttest probability for sample size starting with 36.

Keywords

Confidence intervals estimation; binomial distribution; Posttest probability; Diagnostic test assessment

Introduction

The simplest diagnostic studies generate dichotomous results when the patients can be classifying according to the presence or absence of disease. In this case, a lot of key parameters of interest can be compute in order to measure the effect size of the diagnostic test: sensitivity, specificity, predictive positive value, negative predictive value, pre test probability, post test probability, pre test odds and so on [1, 2, 3, 4]. Whatever the parameters are, some assessment must be made of the trustworthiness or robustness of the finding [5].

The finding of a medical study can provide a point estimation of effect and nowadays is necessary that the point estimation to have a confidence intervals which to allow us to interpret correctly the point estimation of an interest parameter. Unfortunately, for post test probability there was no confidence intervals reported in literature [2, 6, 7, 8, 9, 10].

The post test probability was defined as: ″the proportion of patients with that particular test result who have the target disorder (posttest odds/[1 + post-test odds])″ [3].

The aim of this paper is to introduce six methods named *Wilson, Logit, LogitC, BayesF, Jeffreys,* and *Binomial* as methods of computing confidence intervals for posttest probability and to present theirs performances.

Materials and Methods

In a 2 by 2 contingency tables, according with the presence or absence of a disease and a positive or negative test results, the patients can be classifies in four groups. First group, patients with diseases and a positive test result (real positive cases), noted usually with *a.*

Second group, patients which do not present the disease but have a positive test results (false positive cases) usually noted with *b*. Third group, patients which present the disease but have a negative test result (false negative cases) usually noted with *c*. The fourth group, patient that do not present the disease and have a negative test result (true negative cases) usually noted with *d*.

Using following substitutions: a = *X*, b = *n-X*, c = *Y*, d = *m-Y*, (*X* and *Y* are independent binomial distribution variables of sizes *m* and *n*), the posttest probability (PTP) becomes:

$$\text{PTP} = \frac{\dfrac{X}{n-X}}{\dfrac{X}{n-X}-1} = \frac{X}{2X-n} \qquad (1)$$

Thus, we can say that from mathematical point of view the post test probability parameter is of *X/(2X-n)* function type, function named *ci3* in our program [11].

The function *ci3(X, n)* = *f(X)* = *X/(2X-n)* present an asymptotic point of variation at ±∞ at even *n* and *X* = *n*/2. The graphical representation of this function at *n* = 20 was represented in figure 1, where *X* was considered a continuous variable.

$$n := 20$$
$$X := 0,1..n$$
$$f(X) := \frac{X}{2 \cdot X - n}$$
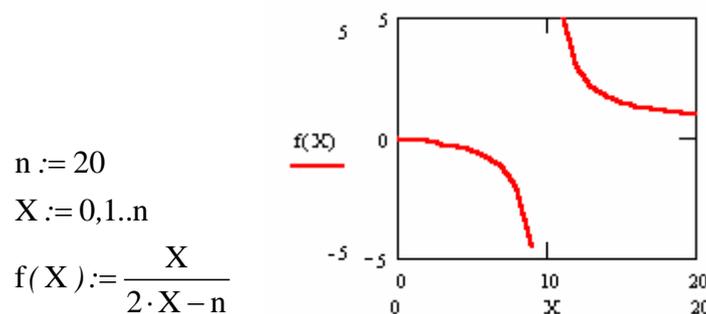
Figure 1. The dependency for continuous variation of X of the posttest probability at n = 20

Some of the methods, which obtained performance in confidence intervals estimation of a binomial proportion, were included in this study. The methods included in the experiment were *Wilson, Logit, LogitC, BayesF, Jeffreys,* and *Binomial*. The mathematical formulas and corresponding functions were presents in a previous paper [12].

In order to obtain a 100·(1-α) = 95% confidence intervals (is most frequently used confidence intervals) the experiments were runs at a significance level α = 5% (noted with *a* in our program). Corresponding to choused significance level was used its normal distribution percentile $z_{1-\alpha/2}$ = 1.96 (noted with *z* in our program). The sequence of the program is: *define("z",1.96); define("a",0.05);*

The performance of each method for different sample sizes and different values of binomial variables was asses using a set of criterions. First, was computes upper and lower boundaries for following cases:

- a given *X (15 ≤ X ≤ 19)* and a specified sample size (*n = 20*) for choused methods:

*$c_i=array("Wilson" , "Logit" , "LogitC" , "BayesF" , "Jeffreys" , "Binomial");
define("N_min",20);define("N_max",21); est_ci_er(z,a,$c_i,"RD","ci3","ci");*

- a sample size varying from 4 to 103:

*$c_i=array("Wald", "WaldC", "Wilson", "WilsonC", "ArcS", "ArcSC1", "ArcSC2",
          "ArcSC3", "AC", "Logit", "LogitC", "BS", "Bayes", "BayesF", "Jeffreys", "CP");
define("N_min",4); define("N_max",104); est_ci_er(z,a,$c_i,"RD","ci3","ci");*

The contour plots represented for samples size varying from 4 to 103 for *ci3* function and for the upper and lower confidence boundaries were obtained using a graphical module implemented in PHP [11].

Second, the average of experimental errors, the standard deviations of experimental errors, and the deviation relative to imposed significance level (*α = 5%*). The standard deviation of the experimental error (*StdDev*) was computes using the next formula:

$$StdDev(X) = \sqrt{\frac{\sum_{i=0}^{n}(X_i - M(X))^2}{n}} \qquad (2)$$

where *StdDev(X)* is standard deviation, X_i is the experimental errors for a given *i*, *M(X)* is the arithmetic mean of the experimental errors and *n* is the sample size.

If we have a sample of *n* elements with a known (or expected) mean (equal with 100α), the deviation around α = 5% (imposed significance level) is giving by:

$$Dev5(X) = \sqrt{\frac{\sum_{i=1}^{n-1}(X_i - 100\alpha)^2}{n-1}} \qquad (3)$$

The standard deviation of the experimental errors as well as the deviation around the imposed significance level was uses as quantitative descriptors of the methods performances.

The value of sample size (*n*) was chosen in the conformity with the real number from the medical studies; in our experiments, we used 10 ≤ *n* ≤ 300 domains (*n* = 20, 50, 100, 300, and 1000). The estimation of the methods was performed for the binomial value of *X* variable, which respect the rule: 0 < *X* < *n*. The sequence of the program that allowed us to compute the experimental errors for specified sample sizes and specified method were:

$c_i=array("LogitC", "BayesF", "Wilson" , "Logit" , "Jeffreys" , "Binomial");

- For *n = 20*: *define("N_min",20); define("N_max",21); est_ci_er(z,a,$c_i,"RD","ci","er");*

- For *n = 50* was modified as follows: *define("N_min",50); define("N_max",51);*

- For *n = 100* was modified as follows: *define("N_min",100); define("N_max",101);*

- For *n = 300* was modified as follows: *define("N_min",300);define("N_max",301);*

- For *n = 1000* was modified as follows: *define("N_min",1000);define("N_max",1001);*

The results were graphically represents as a dependency between error estimation and *X* variable depending on *n*. In the graphical representation, on horizontal axis were represented the *n* values depending on *X* and on the vertical axis the percentage of the experimental errors.

Third, the average of the experimental errors and standard deviations for a sample size, which vary from 4 to 103 ($4 \leq n < 104$, and $0 \leq X \leq n$) was computed using the sequence of the program:

$c_i=array("Wald","WaldC","Wilson","WilsonC","ArcS","ArcSC1","ArcSC2","ArcSC3",
        "AC", "CP", "Logit", "LogitC", "Bayes", "BayesF", "Jeffreys", "Binomial");
define("N_min",4); define ("N_max",103); est_ci_er(z,a,$c_i,"RD","ci3","er");

The contour plots of experimental errors were creates. On the contour plots were represented the error from 0% (red color) to 5% (yellow color) and $\geq 10\%$ (blue color) using 18 hue. The vertical axis represents the *n* values (from 4 to 103) and the horizontal axis the *X* values (from 0 to *n*).

Fourth, the deviations relative to the imposed significance level ($\alpha = 5\%$) when the sample size vary from 4 to 103 ($4 \leq n < 104$, and $0 \leq X \leq n$) was computed. The sequence of the program is:

$c_i=array("LogitC", "BayesF", "Wilson" , "Logit" , "Jeffreys" , "Binomial");
define("N_min",4); define ("N_max",103); est_ci_er(z,a,$c_i,"RD","ci3","va");

Fifth, the performance of sixteen methods were assesed on a 100 random numbers for binomial variable (*X*) and sampel size (*n*), when $4 \leq X < n$ and $4 \leq n \leq 1000$. The sequence of the program is:

$c_i=array("Wald","WaldC","Wilson","WilsonC","ArcS","ArcSC1","ArcSC2","ArcSC3",
        "AC", "CP", "Logit", "LogitC", "Bayes", "BayesF", "Jeffreys", "Binomial");
define("N_min", 4); define("N_max",1000); est_ci_er(z,a,$c_i,"ci3","ra");

Results

The confidence boundaries for *ci3* function-type at *n = 20* were obtained and graphically represented. The experimental results were imports in Microsoft Excel (figure 1) and in a PHP graphic module (figure 2) where the graphical representations were creates.

The lower and upper confidence intervals boundaries for a given *X (15 ≤ X ≤ 19)* at a sample size equal with 20 (*n = 20),* for each choused methods were graphical represented in the figure 1.
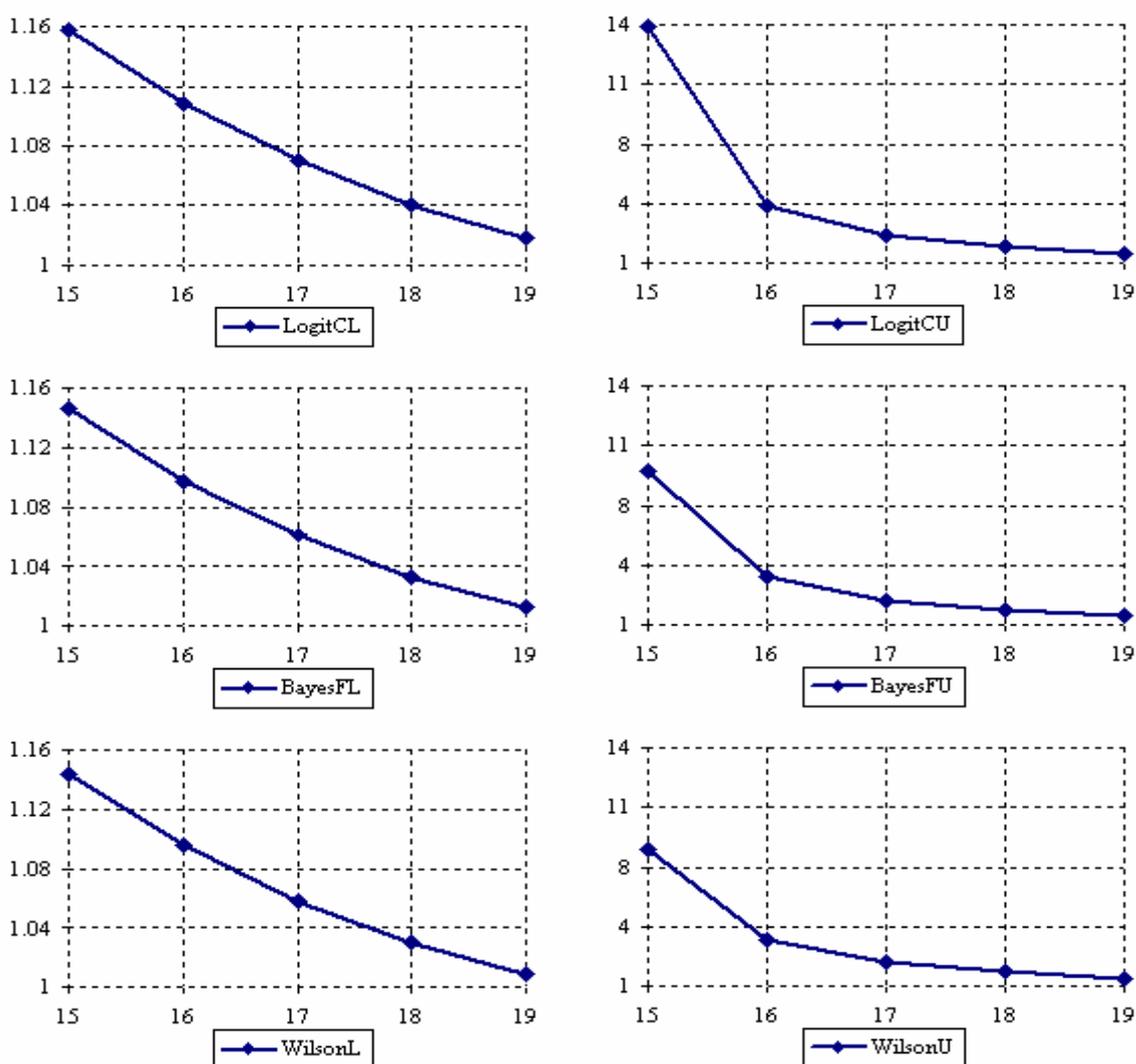

Figure 1. The lower and upper confidence boundaries for posttest probability;
*(15 ≤ X ≤ 19)) at n = 20*
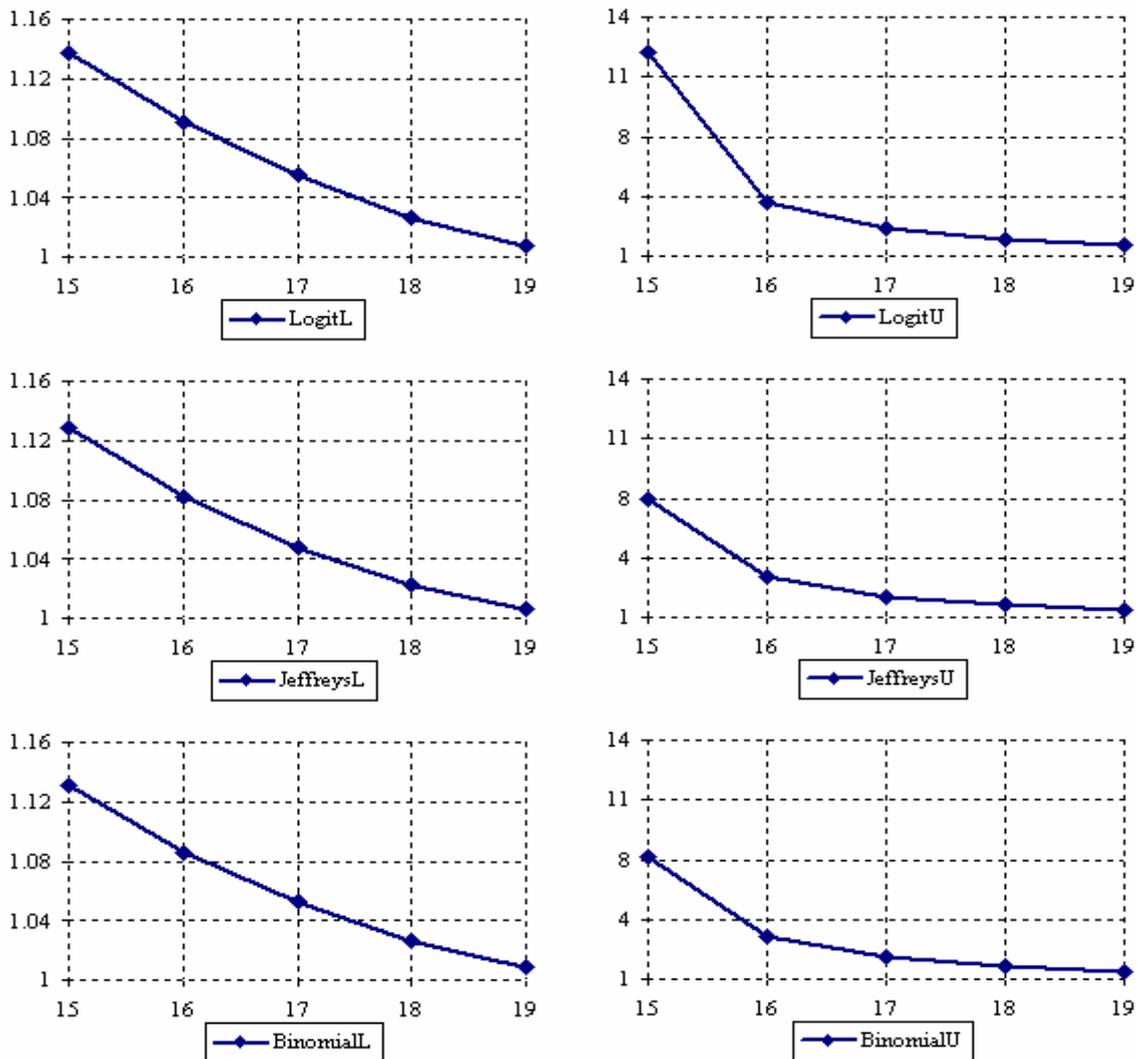
Figure 1. The lower and upper confidence boundaries for posttest probability;
(15 ≤ X ≤ 19)) at n = 20

The posttest probability and its lower and upper confidence boundaries were represents in the figure 2. With red color were represented the experimental values of -10, with yellow the experimental values of 0 and with blue the experimental values of 10. The color between the couplets colors were represented used 252 hue. The infinite values and the value for that the function was undefined were represented with white.

The percentages of experimental errors depending on X value ($0 < X < n$) for specified sample size and specified method were presented. For a sample size equal with 20, the graphical representations of the experimental errors are in figure 3.
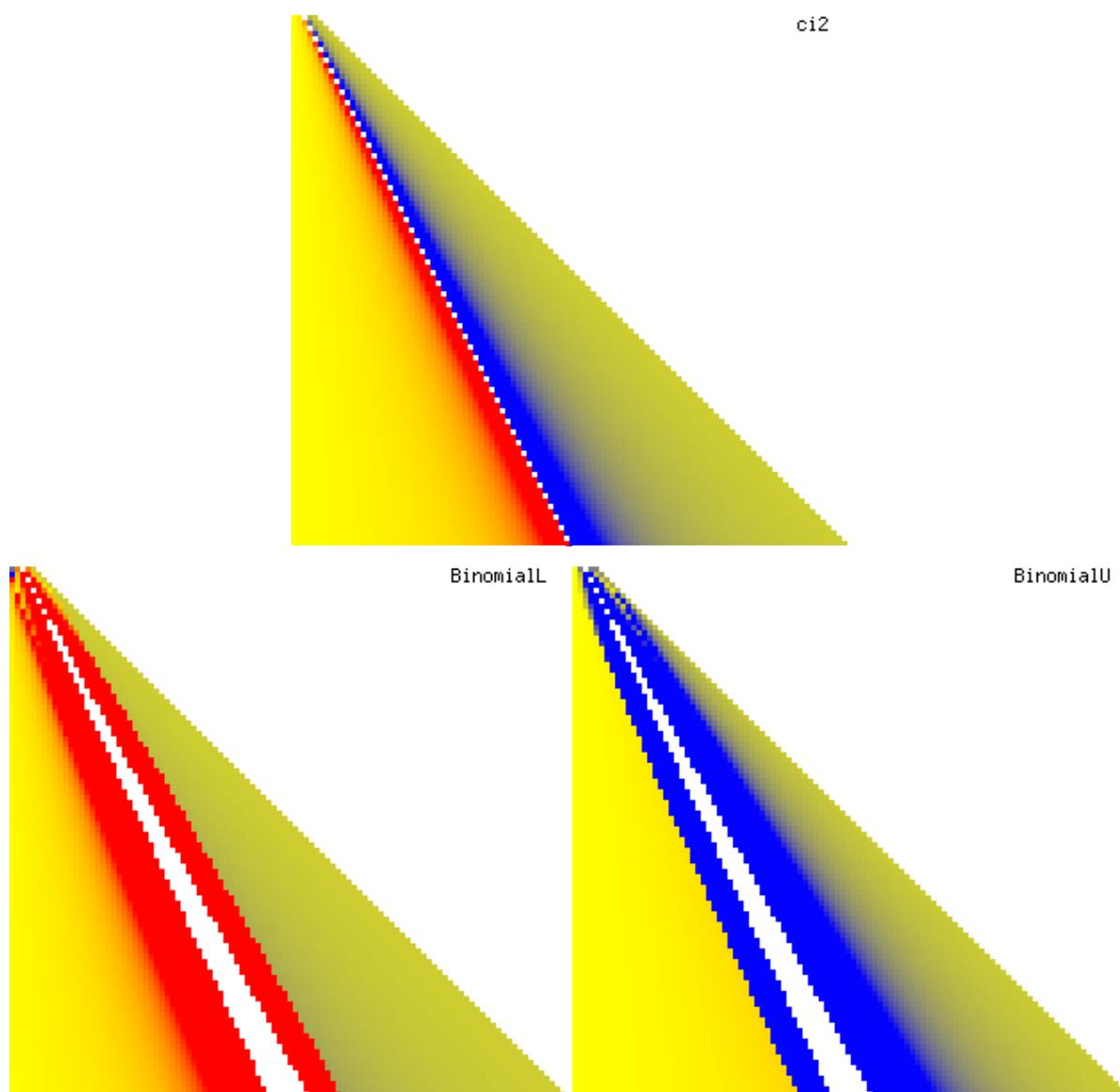
Figure 2. Level surfaces for posttest probability values and its confidence intervals boundaries computed with Binomial method
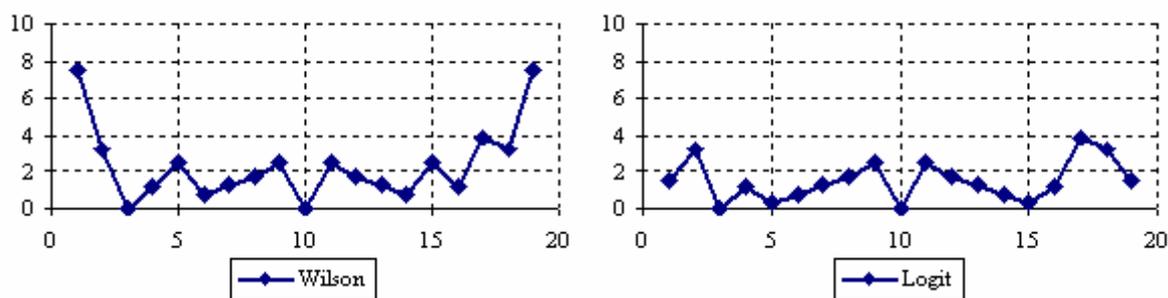

*Figure 3. The percentages of the experimental errors for posttest probability at n = 20 (0<X<n)*
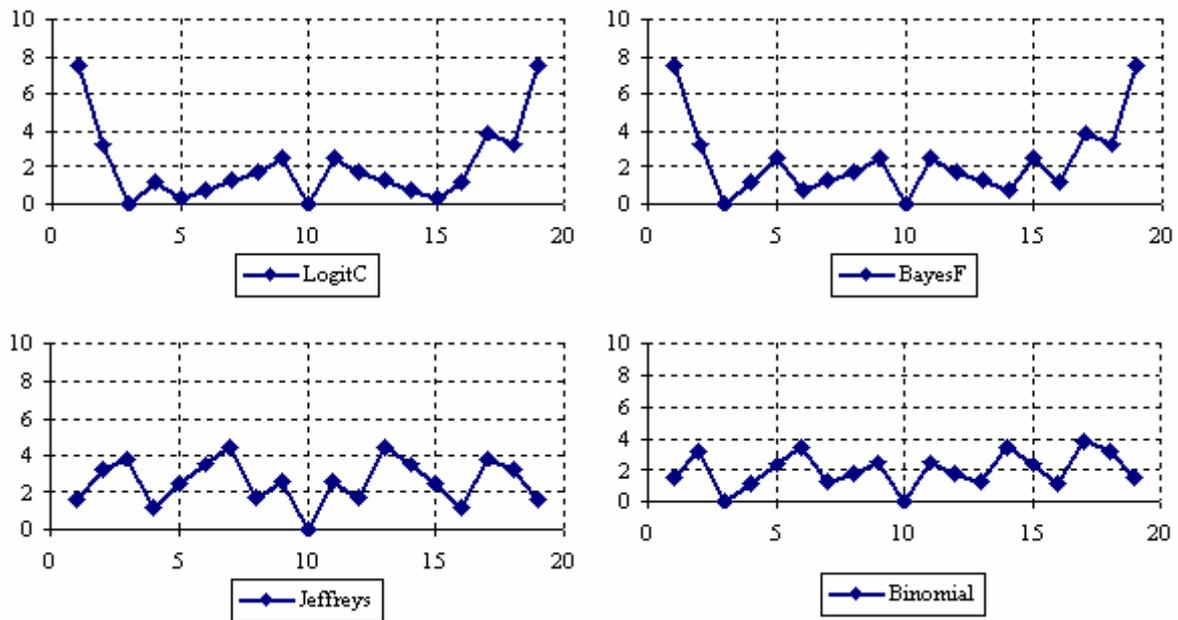
Figure 3. The percentages of the experimental errors
for posttest probability at n = 20 (0<X<n)

The percentages of the experimental errors for a sample size of 50 using specified method are in figure 4.
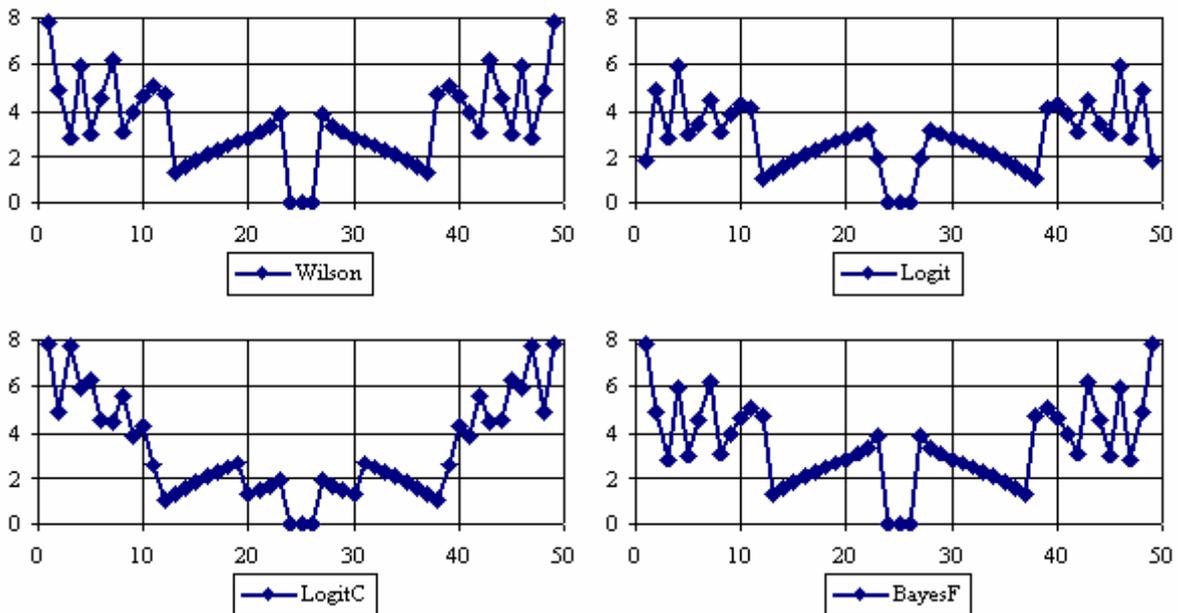
Figure 4. The percentages of the experimental errors
for posttest probability at n = 50 (0 < X < n)
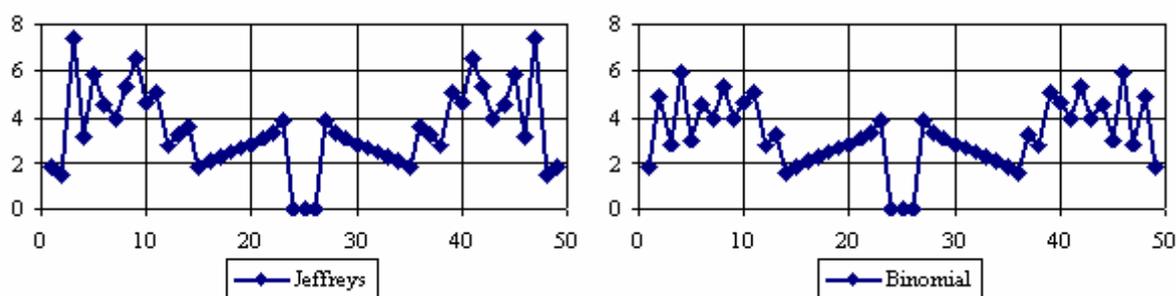
Figure 4. The percentages of the experimental errors
for posttest probability at n = 50 (0 < X < n)

The graphical representations of the percentage of the experimental errors for a sample size equal with 100 using specified method are in figure 5.
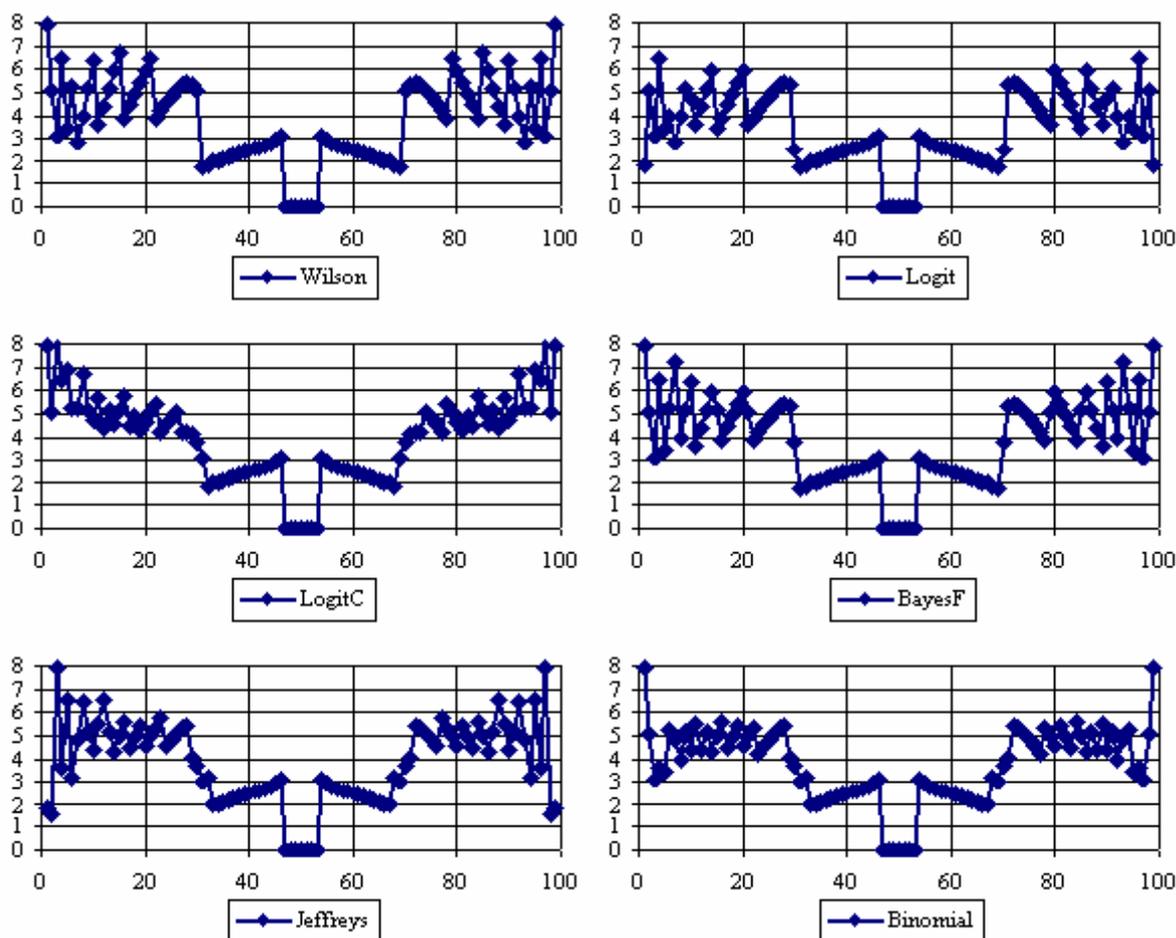
Figure 5. The percentages of the experimental errors for posttest probability
at n = 100 (0 < X < n)

The percentages of the experimental errors for a sample size of 300 using specified method are in figure 6.
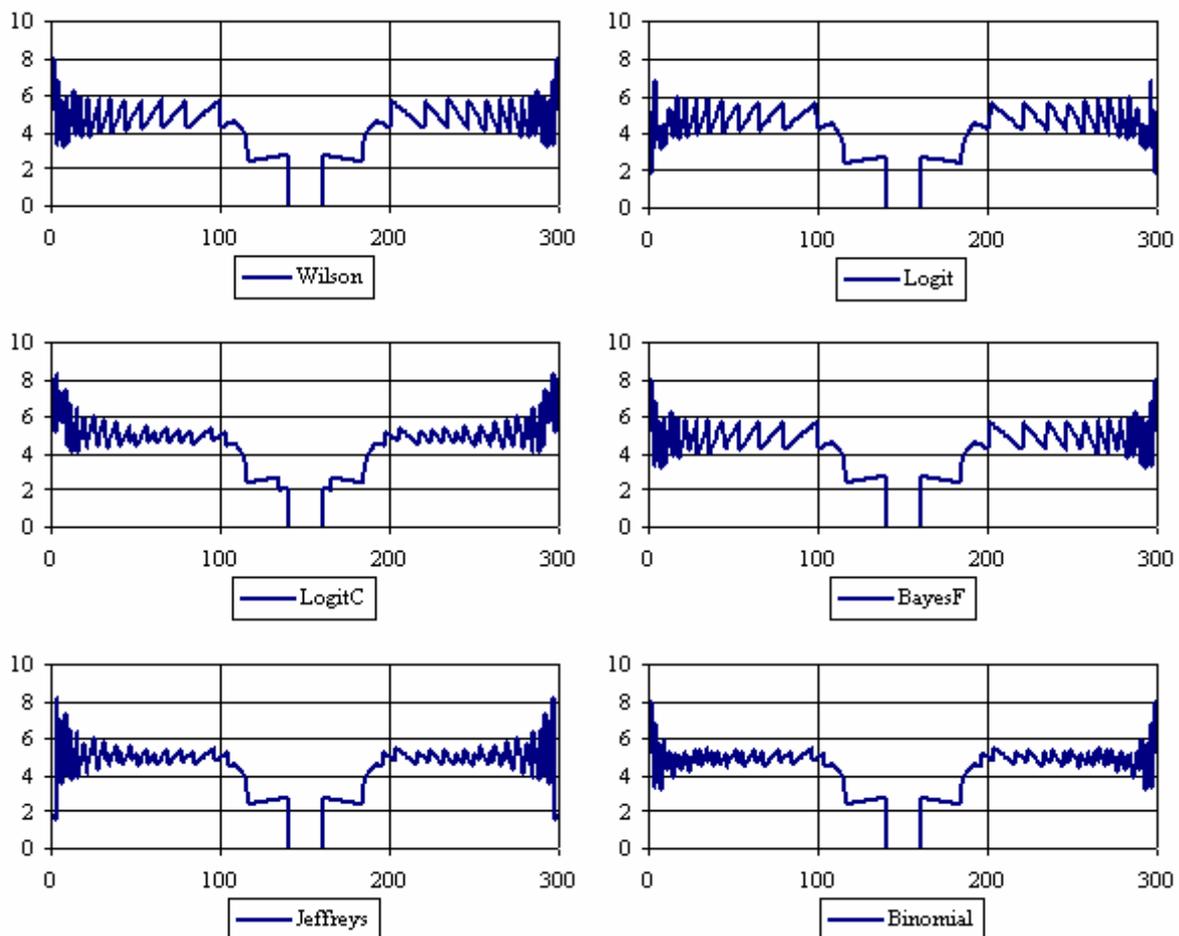
*Figure 6. The percentages of the experimental errors
for posttest probability at n = 300 (0 < X < n)*

The averages and standard deviations (into parenthesis) of experimental errors according to the sample size (*n*) and method used are in table 1.

n	LogitC	BayesF	Wilson	Binomial	Logit	Jeffreys
20	2.15 (2.19)	2.37 (2.10)	2.37 (2.10)	2.03 (1.13)	1.52 (1.11)	2.58 (1.24)
50	3.25 (2.19)	3.44 (1.79)	3.44 (1.79)	3.19 (1.44)	2.77 (1.37)	3.43 (1.76)
100	3.95 (1.88)	3.80 (1.83)	3.80 (1.84)	3.72 (1.63)	3.44 (1.59)	3.76 (1.80)
300	4.27 (1.64)	4.18 (1.54)	4.17 (1.54)	4.19 (1.50)	4.06 (1.49)	4.22 (1.54)
1000	4.53 (1.37)	4.50 (1.34)	4.48 (1.34)	4.49 (1.32)	4.43 (1.33)	4.51 (1.35)

*Table 1. The averages of the experimental errors and standard deviations (parenthesis) for
posttest probability at n = 10, 50, 100, 300 and 1000*

The averages (MErr) and standard deviations (StdDev) of the experimental errors when *n* varies from 4 to 103 were presented in table 2.

Method	Wilson	Logit	LogitC	BayesF	Jeffreys	Binomial
MErr	3.5	3.1	3.6	3.5	3.5	3.3
StdDev	2.96	3.27	3.24	3.00	2.93	2.89

Table 2. The average of the experimental errors and standard deviations
for posttest probability at $4 \leq n < 104$

Using the function (3) presented in materials and methods section of the paper, the graphical representation of frequencies of the best deviations relative to the imposed significance level ($\alpha = 5\%$) from figure 8 was obtained for $4 \leq n < 104$.
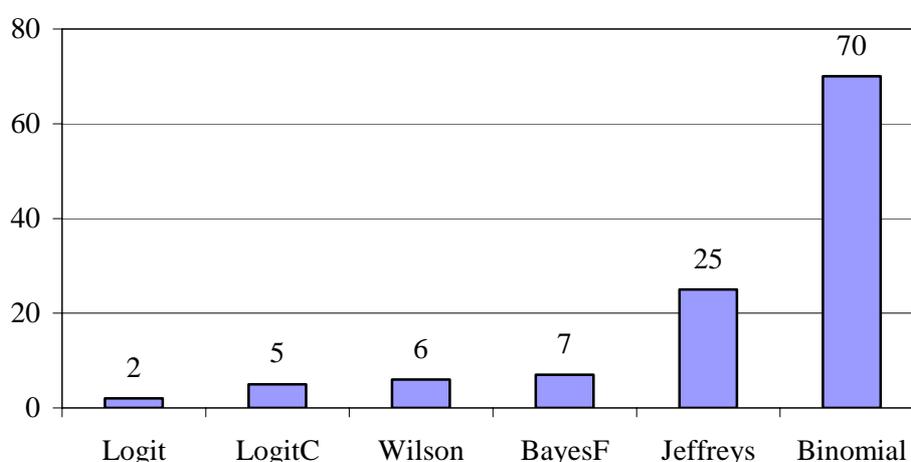
Figure 8. The frequencies of the best deviations of the experimental errors relative to
significance level ($\alpha = 5\%$) for posttest probability at $4 \leq n < 104$

The contour plots of experimental errors when $4 \leq n < 104$ were illustrated in figure 7. On the plots were represented the experimental errors from 0% (red color) to 5% (yellow color) and $\geq 10\%$ (blue color) using 18 intermediate hue. The vertical axis represent the *n* values ($4 \leq n < 104$) and the horizontal axis the *X* values (from 0 to *n-1*).
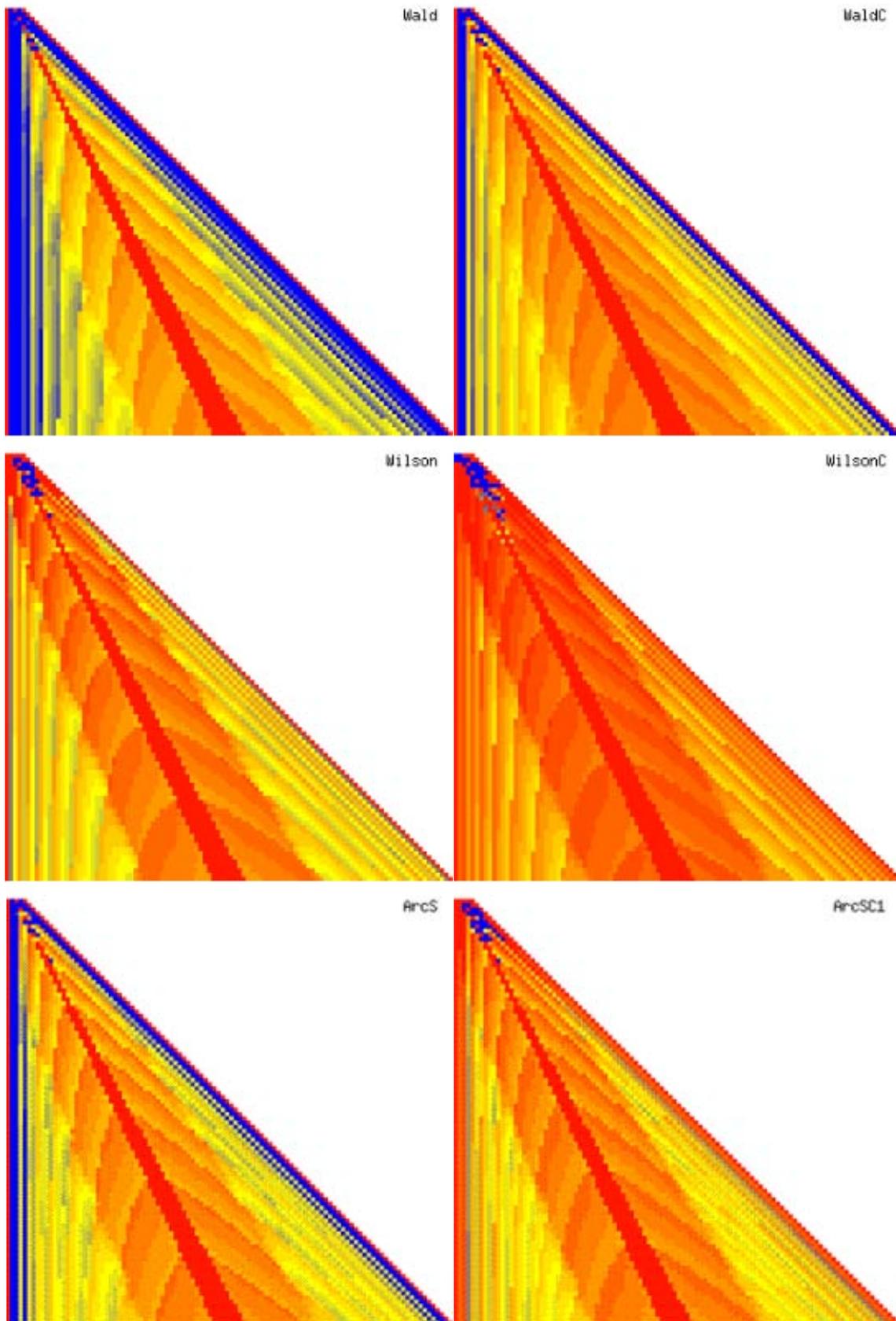
Figure 7. The contour plot of the experimental errors for ci3 function at 3 < n < 104
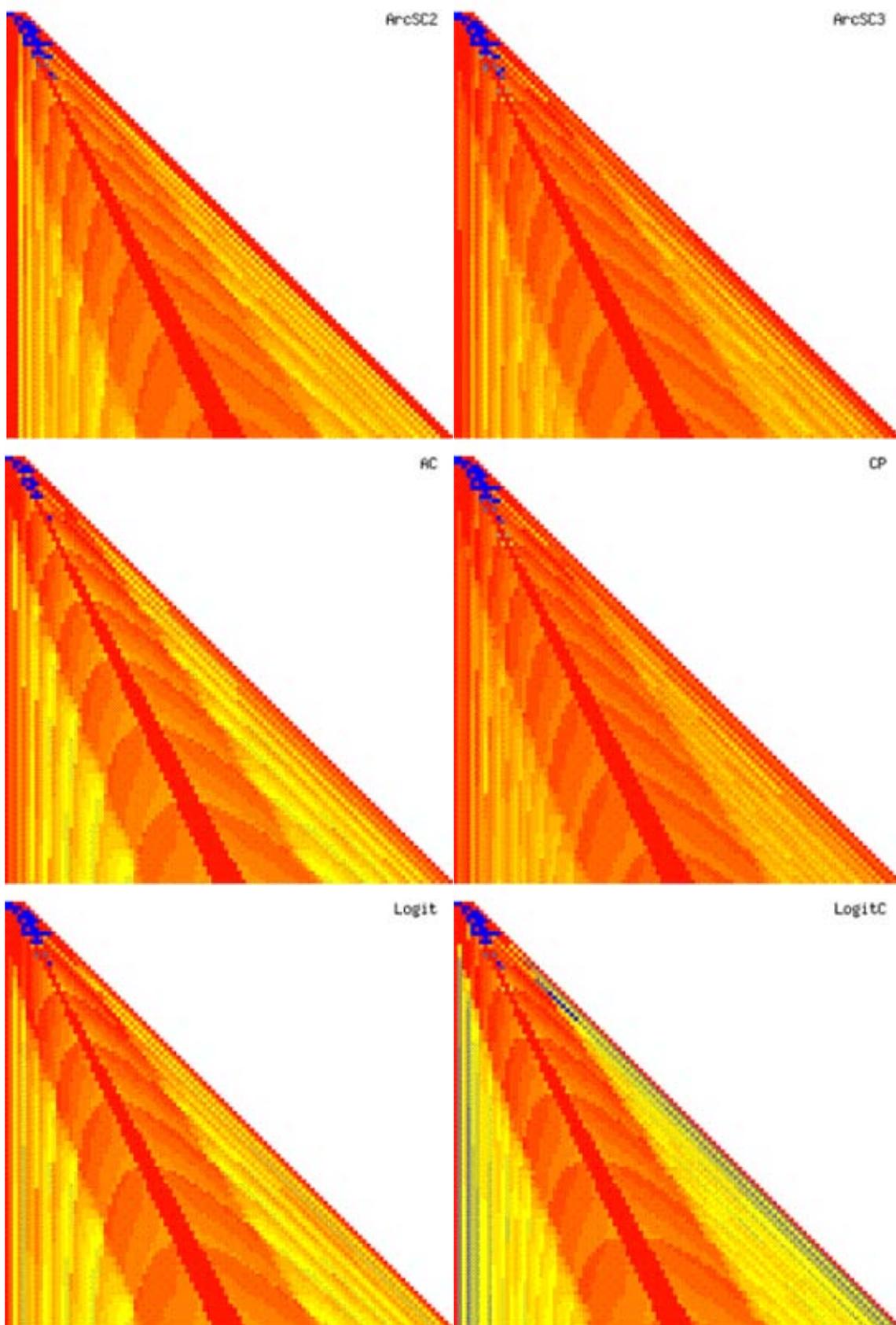
Figure 7. The contour plot of the experimental errors for ci3 function at 3 < n < 104
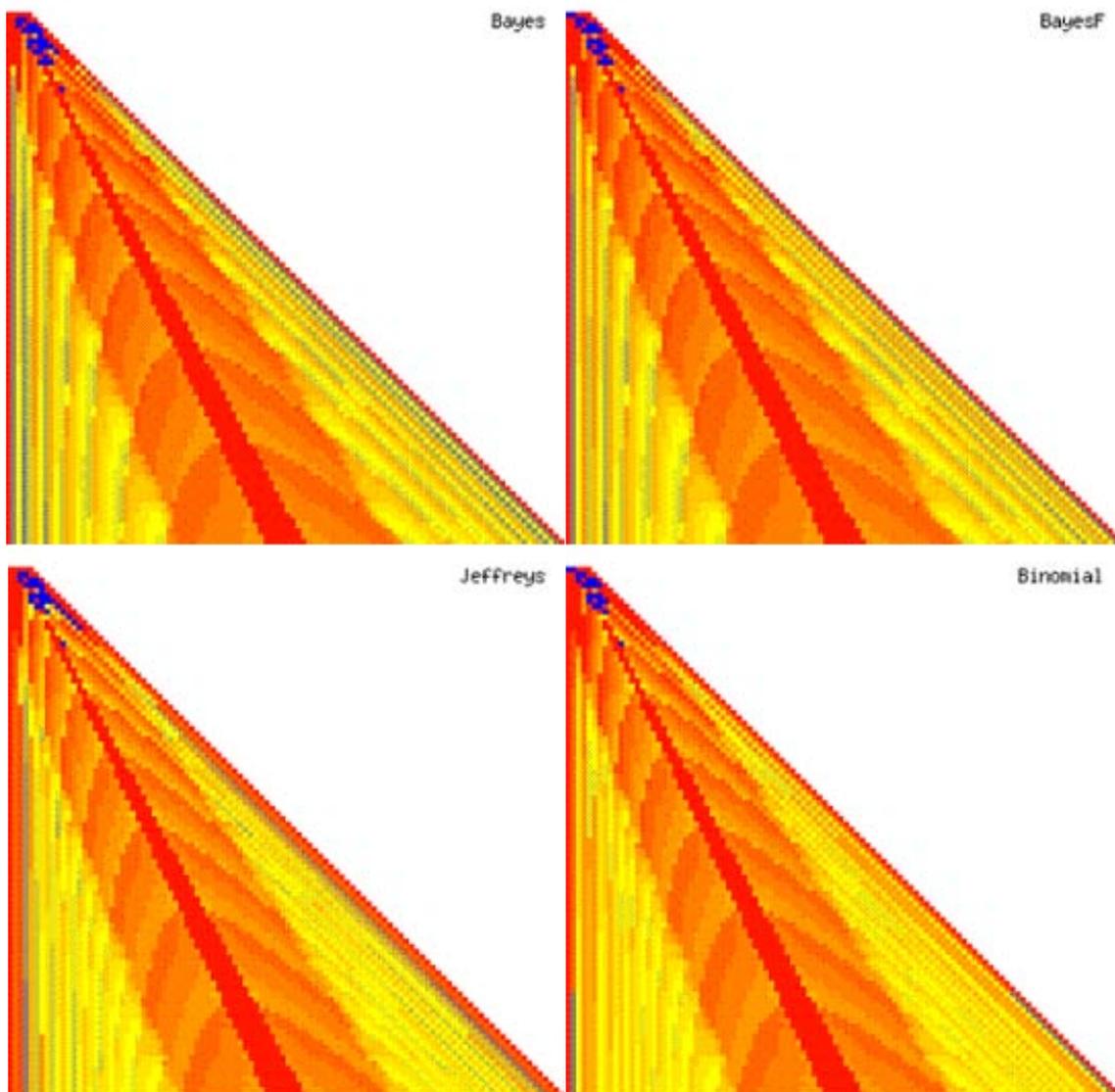
Figure 7. The contour plot of the experimental errors for ci3 function at 3 < n < 104

The results of the deviations relative to the imposed significance level for $104 \leq n \leq$ 109 are in table 3.

n	Binomial	LogitC	Jeffreys	BayesF
104	2.01	2.29	2.12	2.15
105	2.05	2.31	2.16	2.20
106	2.03	2.30	2.13	2.22
107	2.05	2.30	2.15	2.24
108	2.03	2.28	2.14	2.20
109	2.06	2.29	2.16	2.22

Table 3. The deviations relative to the significance level (α = 5%)
for ci3 function at 104 ≤ n ≤ 109

For the random values of binomial variable (*X*) and sample sizes (*n*), when the random sample size is choused from 4 to 1000, a set of calculation were performed and graphical represented in figure 10.

In the figure 10 was represents with black dots the frequencies of the experimental error for each specified method; with green line the best errors interpolation curve with a Gauss curve (*dIG(er)*). The Gauss curves of the average and standard deviation of the experimental errors (*dMV(er)*) was represents with red line. The Gauss curve of the experimental errors deviations relative to the significance level (*d5V(er)*) was represented with blue squares. The Gauss curve of the standard binomial distribution from the average of the errors equal with 100·α (*pN(er,10)*) was represented with black line.

For the random samples, the results are in tables 4 to 7. In table 4 were presented the average of the deviation of the experimental errors relative to significance level α =5% (*Dev5*), the absolute differences of the average of experimental errors relative to the imposed significance level (*/5-M/*), and standard deviations (*StdDev*).

No	Method	Dev5	Method	|5-M|	Method	StdDev
1	Binomial	1.48	WaldC	0.17	CP	1.22
2	BayesF	1.54	ArcS	0.40	WilsonC	1.26
3	Wilson	1.54	Bayes	0.50	ArcSC3	1.27
4	Bayes	1.55	Wilson	0.54	ArcSC2	1.28
5	ArcSC1	1.56	BayesF	0.55	Binomial	1.36
6	Logit	1.57	LogitC	0.57	Logit	1.36
7	Jeffreys	1.57	Binomial	0.59	AC	1.36
8	LogitC	1.57	Wald	0.60	ArcSC1	1.41
9	AC	1.61	Jeffreys	0.64	Jeffreys	1.44
10	CP	1.84	ArcSC1	0.65	BayesF	1.44
11	ArcSC2	1.88	Logit	0.78	Wilson	1.45
12	ArcSC3	1.89	AC	0.84	LogitC	1.47
13	WilsonC	1.90	ArcSC2	1.37	Bayes	1.47
14	WaldC	4.77	CP	1.37	WaldC	4.76
15	ArcS	5.00	ArcSC3	1.39	Wald	4.98
16	Wald	5.02	WilsonC	1.41	ArcS	4.98

Table 4. Methods ordered by performance according to Dev5, /5-M/ and StdDev criterions

Table 5 contains the absolute differences of the averages that result from Gaussian interpolation curve to the imposed significance level (*/5-MInt/*), the deviations that result from Gaussian interpolation curve (*DevInt*), the correlation coefficient of interpolation (*r2Int*) and the Fisher point estimator (*FInt*).

No	Method	|5-MInt|	Method	DevInt	Method	r2Int	FInt
1	Logit	0.02	WaldC	1.22	ArcSC2	0.56	24
2	LogitC	0.04	Binomial	1.23	Wald	0.58	27
3	AC	0.06	LogitC	1.27	Jeffreys	0.60	28
4	ArcSC1	0.10	ArcSC1	1.32	CP	0.60	29
5	Jeffreys	0.10	Jeffreys	1.32	ArcSC3	0.60	29
6	BayesF	0.12	ArcS	1.35	Binomial	0.60	29
7	Bayes	0.14	Wald	1.36	LogitC	0.61	29
8	Binomial	0.14	AC	1.46	ArcS	0.61	29
9	Wilson	0.15	Logit	1.46	WaldC	0.61	29
10	Wald	0.26	Bayes	1.46	ArcSC1	0.61	29
11	WaldC	0.27	BayesF	1.46	WilsonC	0.61	30
12	ArcS	0.28	ArcSC2	1.47	Bayes	0.64	34
13	WilsonC	0.63	CP	1.48	BayesF	0.65	35
14	ArcSC2	0.63	Wilson	1.49	Wilson	0.65	35
15	ArcSC3	0.63	ArcSC3	1.51	AC	0.66	37
16	CP	0.65	WilsonC	1.52	Logit	0.66	38

Table 5. The methods ordered by |5-Mint|, DevInt, r2Int and FInt criterions

The superposition of the standard binomial distribution curve and interpolation curve (*pNIG*), the superposition of standard binomial distribution curve and the experimental error distribution curve (*pNMV*), and the superposition of standard binomial distribution curve and the error distribution curve around significance level ($\alpha = 5\%$) (*pN5V*) were presented in table 6.

No	Method	pNIG	Method	pNMV	Method	pN5V
1	CP	0.83	Wald	0.49	Wald	0.49
2	ArcSC2	0.83	ArcS	0.50	ArcS	0.50
3	ArcSC3	0.84	WaldC	0.51	WaldC	0.51
4	WilsonC	0.84	CP	0.61	WilsonC	0.91
5	WaldC	0.86	WilsonC	0.61	ArcSC3	0.91
6	Binomial	0.88	ArcSC3	0.62	ArcSC2	0.92
7	LogitC	0.89	ArcSC2	0.63	CP	0.93
8	ArcS	0.90	AC	0.77	Binomial	0.97
9	Wald	0.90	Logit	0.78	Wilson	0.99
10	ArcSC1	0.91	ArcSC1	0.82	Bayes	0.99
11	Jeffreys	0.91	Jeffreys	0.83	BayesF	0.99
12	BayesF	0.95	Binomial	0.83	Jeffreys	0.99
13	Wilson	0.95	LogitC	0.85	ArcSC1	0.99
14	AC	0.96	BayesF	0.85	AC	0.99
15	Bayes	0.96	Wilson	0.85	Logit	1.00
16	Logit	0.96	Bayes	0.87	LogitC	1.00

*Table 6. The confidence intervals methods ordered by the
pIGMV, pIG5V, and pMV5V criterions*
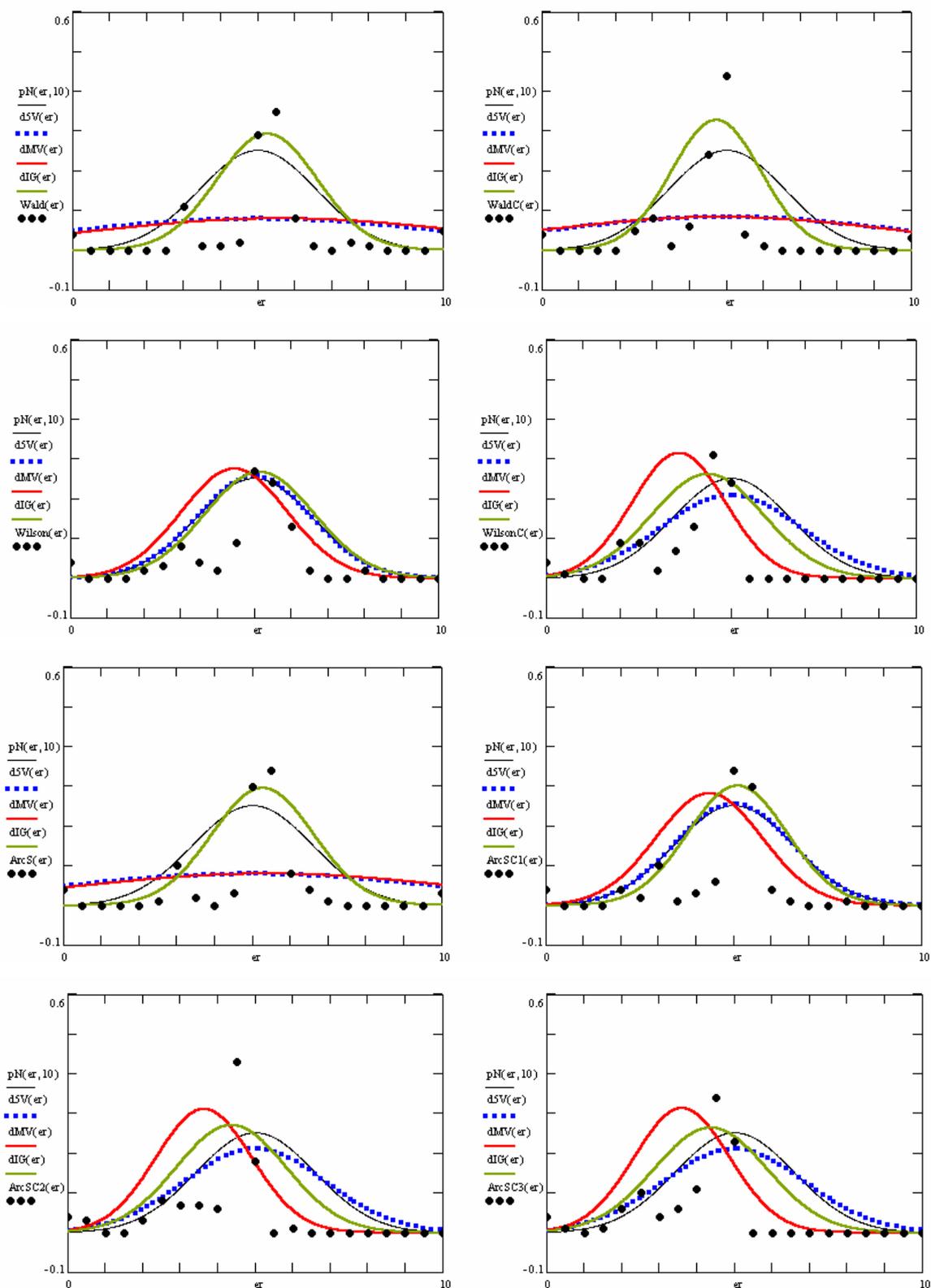
Figure 9. The pN(er, 10), d5V(er), dMV(er), dIG(er) and the frequencies of the experimental errors for each specified method for random X, n (1 ≤ X < 1000, 4 ≤ n ≤ 1000)
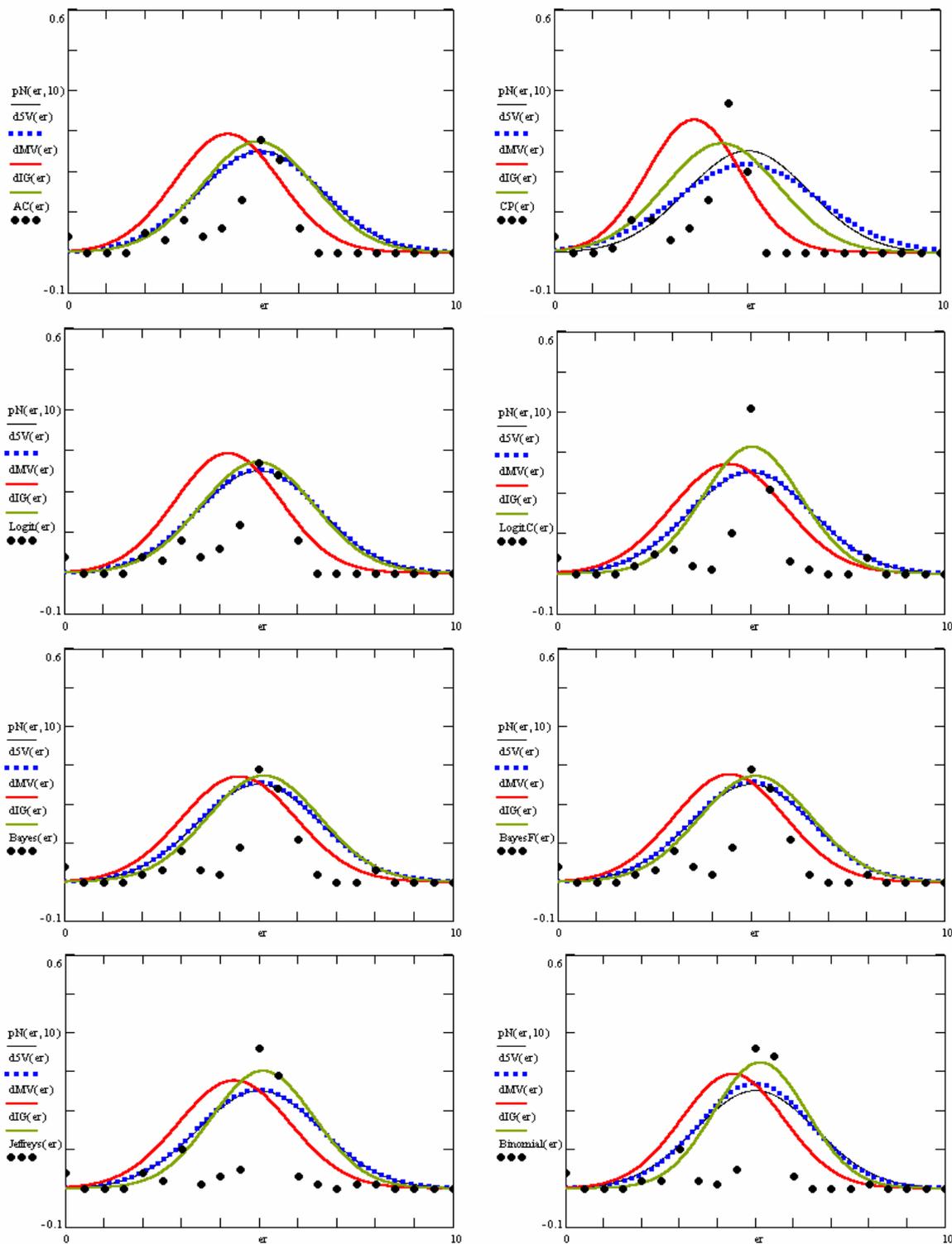
*Figure 9. The pN(er, 10), d5V(er), dMV(er), dIG(er) and the frequencies of the experimental errors for each specified method for random X, n (1 ≤ X < 1000, 4 ≤ n ≤ 1000)*

Table 7 contains the superposition of interpolation Gauss curve and the Gauss curve of error around experimental mean (*pIGMV*); the interpolation Gauss curve and the Gauss curve

of error around imposed mean (*pIG5V*); and the Gauss curve of experimental error around experimental mean and the error Gauss curve around imposed mean α = 5% (*pMV5V*).

No	Method	pIGMV	Method	pIG5V	Method	pMV5V
1	Bayes	0.83	Logit	0.97	Bayes	0.87
2	BayesF	0.82	BayesF	0.96	Wilson	0.85
3	LogitC	0.82	Wilson	0.96	BayesF	0.85
4	Wilson	0.82	Bayes	0.95	LogitC	0.85
5	Jeffreys	0.79	AC	0.95	Binomial	0.83
6	Logit	0.79	ArcSC1	0.92	Jeffreys	0.83
7	ArcSC2	0.78	Jeffreys	0.91	ArcSC1	0.82
8	ArcSC1	0.78	Binomial	0.90	Logit	0.78
9	AC	0.78	LogitC	0.90	AC	0.77
10	CP	0.78	WilsonC	0.83	WaldC	0.70
11	Binomial	0.78	ArcSC3	0.83	ArcS	0.67
12	ArcSC3	0.77	CP	0.82	Wald	0.66
13	WilsonC	0.77	ArcSC2	0.82	ArcSC2	0.64
14	Wald	0.45	Wald	0.44	ArcSC3	0.63
15	ArcS	0.44	ArcS	0.44	WilsonC	0.62
16	WaldC	0.43	WaldC	0.42	CP	0.62

Table 7. The confidence intervals ordered by the pIGMV, pIG5V, and pMV5V

Discussions

Analyzing the confidence intervals for posttest probability we can observe that for all methods of computing confidence intervals, the average of the experimental errors increases proportional with the increasing of the sample size and never exceeded the significance level (*α* = 5%).

The *Logit* and *Jeffreys* methods (both methods based on binomial distribution probability) obtains the most conservative experimental standard deviation in computing confidence intervals at a choused significance level equal with *α* = 5%. Thus, if we look after a confidence interval method, which is capable to assure the lowest standard deviation of the errors, we will choose one of these two methods.

If the standard deviation is not the ones parameter of confidence intervals assessment and we need a method which to assures an experimental error mean close to the significance level (*α* = 5%), then the *LogitC* method could be choused because obtain the best performance on the 100 ≤ *n* ≤ 300 domain. Even for *n* > 300, the best performance were obtain by the *Logit*

(average of experimental errors equal with 4.53% at *n=1000*) followed by the *BayesF* (average of experimental errors equal with 4.50% at *n=1000*).

If we looked at the experimental results presented in table 2, according to different sample sizes, the methods *Bayes*, *Binomial* and *Wilson* obtains an average of the experimental errors closed to the choused significance level ($\alpha = 5\%$) and the *Binomial* method obtains the lowest standard deviation of the experimental errors.

Deviation relative to the imposed significance level (α = 5%) is a quantitative measure of performance of methods used in computing confidence intervals. Analyzing the figure 9 some aspects were disclosure. The first observation was that the original method *Binomial* improves the *BayesF* and the *Jeffreys* methods even if, the limit the *Binomial* overlap on *BayesF* and *Jeffreys* (see the *Binomial* function):

$$\text{Binomial}(X, n) \xrightarrow{X \to 0, n} \text{BayesF}(X, n) \, , \ \text{Binomial}(X, n) \xrightarrow{2X \to n} \text{Jeffreys}(X, n)$$

The second observation is that from the binomial methods, the *Binomial* method obtains the best performance (with 70 appearances, followed by *Jeffreys* with 25 and the *BayesF* with 7 appearances).

The third observation from the experimental data (which cannot been seen from the graphical plots) was that for *n > 36* the *Binomial* method is almost systematically the method which have the lowest standard deviation, and it obtain one failure at *n = 41* where the *Jeffreys* method obtains the lowest standard deviation. Note that from $n \geq 42$, the Binomial method obtains the best performance in confidence intervals estimation for *X/(2X-n)* fraction.

For the random samples (*n*) and binomial variable (*X*), the *WaldC* method obtains the closed experimental error average to the significance level (see table 4); the lowest experimental standard deviation were obtains by the *Clopper-Pearson* method followed by the *WilsonC* method. The most closed deviation relative to the imposed significance level ($\alpha = 5\%$) was obtains by the *Binomial* method followed by the *BayesF* and *Wilson* methods.

The *Logit* method obtains the less absolute differences of the averages that results from the Gaussian interpolation curve to the imposed significance level (table 5), followed by the *LogitC* method; *WaldC* was the method with the lowest interpolation standard deviation. The *Logit* and the *Agresti-Coull* methods obtained the best correlation between theoretical curve and experimental data.

Analyzing the results presented in the table 6, it can be observes that the *Logit, BayesF* and *Agresti-Coull* methods, obtain the maximum superposition of the interpolation

curve and the curve of standard binomial distribution. The maximum superposition between the curve of standard binomial distribution and the curve of experimental errors was obtains by the *Bayes* method. The *Logit* and *LogitC* methods obtain the maximum superposition between the curve of standard binomial distribution and the curve of errors around the significance level ($\alpha = 5\%$).

The maximum superposition between the Gauss curve of interpolation and the Gauss curve of errors around experimental mean was obtains by the *Bayes* method, followed by the *BayesF, LogitC* and the *Wilson* methods (table 7).

The *Logit* method obtained the maximum superposition between the Gauss curve if interpolation and the Gauss curve of errors around significance level ($\alpha = 5\%$), followed by the method *BayesF* and the method *Wilson*. The maximum superposition between the Gauss curve of experimental errors and the Gauss curve of errors around imposed mean ($\alpha = 5\%$) were obtained by the *Bayes* method closely followed by the *Wilson, LogitC* and the *BayesF* methods.

Conclusions

The best criterion of comparing the methods of computing confidence intervals is deviation relative to the imposed significance level because in real experiments the sample sizes are not equal to each other.

The original *Binomial* method, systematically obtains the best performance for sample sizes of $36 \leq n \leq 1001$, method that could be choused if it is need a method with the lowest standard deviation of the experimental errors.

The same performances are seen when the sample size vary from 4 to 103, where the *Binomial* method obtain the lowest deviations and also when the sample size and the binomial variable are randomly choused.

Based on these conclusions, we recommend the use of *Binomial* method for computing of the confidence intervals of posttest probability.

Acknowledgements

References

1. Drugan T., Bolboacă S., Colosi H., Achimaş Cadariu A., Ţigan Ş., *Diagnostic Test Assesment* Chapter in *Statistical Inference of Medical Data*, ALMA MATER Cluj-Napoca, 2003, p. 56-65 (in Romanian).

2. Hamm R. M., *Clinical Decision Making Spreadsheet Calculator*, University of Oklahoma Health Sciences Center, available at:
www.emory.edu/WHSC/MED/EMAC/curriculum/diagnosis/oklahomaLRs.xls

3. Sackett D.L., Straus S., Richardson W.S., Rosenberg W., Haynes R.B., *Diagnosis and screening* Chapter In *Evidence-based medicine. How to practice and teach EBM,* 2nd Ed. Edinburgh, Churchill Livingstone, 2000, p. 67-93.

4. Sox C.H., *Tools for Decision Making*. Interactive Textbook on Clinical Symptom Research, available at: http://symptomresearch.nih.gov/chapter_14/Part_1/sec8/chspt1s8pg1.htm

5. Huw D., *What are confidence intervals?*, What is…, Hayward Group Publication, 2003, 3, p. 1-9 available at: http://www.evidence-based-medicine.co.uk/ebmfiles/WhatareConfInter.pdf

6. Rosner B., *Hypothesis Testing: Categorical Data*, Fundamentals of Biostatistics, Forth Edition, Duxbury Press, Belmont, 1995, p. 345-442.

7. Brown D.L., Cai T.T., DasGupta A., *Interval estimation for a binomial proportion,* Statistical Science, 2001, 16, p. 101-133.

8. Blyth C.R., Still H.A., *Binomial confidence intervals*, J. Amer. Stat. Assoc. 1983, 78, p. 108-116.

9. Michael D., Edwardes B., *The Evaluation of Confidence Sets with Application to Binomial Intervals*, Statistica Sinica, 1998, 8, p. 393-409.

10. Newcombe R.G., *Two-sided confidence intervals for the single proportion; comparison of several methods*, Statistics in Medicine, 1998, 17, p. 857–872.

11. Drugan T., Bolboacă S., Jäntschi L., Achimaş Cadariu A., *Binomial Distribution Sample Confidence Intervals Estimation 1. Sampling and Medical Key Parameters Calculation*, Leonardo Electronic Journal of Practices and Technologies, 2003, 3, p. 45-74.

12. Bolboacă S., Achimaş Cadariu A., *Binomial Distribution Sample Confidence Intervals Estimation 2. Proportion-like Medical Parameters*, Leonardo Electronic Journal of Practices and Technologies, 2003, 3, p. 75-110.