Scientific
Research
Publishing

Bootstrap Confidence Intervals for Proportions of Unequal Sized Groups Adjusted for Overdispersion

Olivia Wanjeri Mwangi, Ali Islam, Orawo Luke

Department of Mathematics, Egerton University, Njoro, Kenya
Email: owanjeri@yahoo.com, asislam54@yahoo.com, orawo2000@yahoo.com

Abstract

Group testing is a method of pooling a number of units together and performing a single test on the resulting group. It is an appealing option when few individual units are thought to be infected leading to reduced costs of testing as compared to individually testing the units. Group testing aims to identify the positive groups in all the groups tested or to estimate the proportion of positives (p) in a population. Interval estimation methods of the proportions in group testing for unequal group sizes adjusted for overdispersion have been examined. Lately improvement in statistical methods allows the construction of highly accurate confidence intervals (CIs). The aim here is to apply group testing for estimation and generate highly accurate Bootstrap confidence intervals (CIs) for the proportion of defective or positive units in particular. This study provided a comparison of several proven methods of constructing CIs for a binomial proportion after adjusting for overdispersion in group testing with groups of unequal sizes. Bootstrap resampling was applied on data simulated from binomial distribution, and confidence intervals with high coverage probabilities were produced. This data was assumed to be overdispersed and independent between groups but correlated within these groups. Interval estimation methods based on the Wald, the Logit and Complementary log-log (CLL) functions were considered. The criterion used in the comparisons is mainly the coverage probabilities attained by nominal 95% CIs, though interval width is also regarded. Bootstrapping produced CIs with high coverage probabilities for each of the three interval methods.

Keywords

Group Testing, Overdispersion, Quasi-Likelihood, Confidence Interval, Bootstrapping, Coverage Probability

1. Introduction

Group testing started or originated with [1] during World War II as a frugal method of testing blood specimens of army inductees in order to detect the presence of infection. Group testing has been applied in many areas: plant disease assessment [2], fisheries [3], quality control, drug discovery and transmission of viruses by insect vectors [4]. It has also been used to screen the population for the presence of HIV/Aids antibody [5]. [1] proposed that, rather than testing each blood specimen individually, portions each of k specimens can be pooled and the pooled specimen can be tested. If the pooled specimen is free of infection, all k inductees are passed with no further tests, otherwise the remaining portions of each of the blood specimens are to be tested individually. If the prevalence of infection is low, the expected number of tests per inductee and thus the expected cost per inductee, would be reduced. [1] assumed that tests were perfect, that is, a negative reading indicated the group contained no defective item and a positive reading indicated the presence of at least one defective item.

Group tests save resources since many units are tested without testing them individually. One of their advantages as a method of estimation is they are time efficient. In most of the studies carried out, the experimental unit/group is a litter. It has been observed that there is a tendency of littermates to respond more alike than animals from different litters, the "litter effect". This litter effect is also known as the extra-dispersion (over/under-dispersion) or the intra-litter correlation. These litters may be of equal or unequal sizes. The concern here is with methods of establishing the confidence interval for the proportion of defective units p in unequal groups with adjustment for overdispersion using bootstrap technique.

Overdispersion is the phenomenon of having greater variability than predicted by the random component of the model; this is common in the modeling of binomial distribution for group testing [6]. Overdispersion is a very common feature in applied data analysis because naturally populations are frequently heterogeneous contrary to the assumptions implicit within widely used simple parametric models. The over-dispersion parameter is common in proportions in biology, toxicology, medicine, genetics and other similar fields and is important in making inference regarding the regression parameters on the mean [7].

Maximum likelihood estimation gives a unified approach to estimation, which is well-defined, in binomial distribution and many other problems. The maximum likelihood as an estimator has been studied and seconded as an approach for using the proposed extended Beta-Binomial (BB) model to analyze over/under-dispersed proportions [8]. They estimated this parameter using maximum likelihood estimation (MLE). This concurred with the observation that the MLEs may be biased when the sample size n or the total Fisher information is small [9]. This bias is usually ignored in practice, the justification being that it is small compared with the standard errors.

Studies on point estimation in terms of bias and efficiency and the test for presence of overdispersion for both counts and proportions data have been done [10]. Interval estimation methods in which groups were of unequal sizes were carried out using interval based on the MLE, the Logit and CLL [11]. The quasi-likelihood approach to correct for overdispersion is used by [6], he examines interval estimation methods based on functions of the MLE, the Logit and CLL functions.

2. Literature Review

2.1. Bootstrapping

The bootstrap method was introduced by [12]. It is a very general resampling procedure for estimating the distributions of statistics based on independent observations. It has spread fast in statistical sciences within the last decades. The primary task of a statistician is to summarize a sample based study and generalize the finding to the parent population in a scientific manner. Of course, a summary statistic like the sample mean will fluctuate from sample to sample and a statistician would like to know the magnitude of these fluctuations around the corresponding population parameter.

To understand bootstrap, suppose it were possible to draw repeated samples (of the same size) from the population of interest, a large number of times. Then, one would get a fairly good idea about the sampling distribution of a particular statistic from the collection of its values arising from these repeated samples. But, that does not make sense as it would be too expensive and defeat the purpose of a sample study. The purpose of a sample study is to gather information cheaply in a timely fashion. The idea behind bootstrap is to use the data of a sample study at hand as a "surrogate population", for the purpose of approximating the sampling distribution of a

statistic, that is, to resample (with replacement) from the sample data at hand and create a large number of "phantom samples" known as bootstrap samples. These samples can be used to obtain more improved estimation of unknown parameter (s) of a probability model.

Recently, study has been done on the construction of bootstrap confidence intervals for the overdispersion parameter in equal proportions in Beta Binomial using Maximum Likelihood Estimator, Method of Moments Estimator and Quasi-likelihood [13]. Bootstrapping has been seen to provide improved CIs and is useful even when very little is known about the underlying distributions [14].

2.2. Estimation of Proportions

2.2.1. Confidence Intervals Based on the Maximum Quasi-Likelihood Estimate

Overdispersion causes one to underestimate the variance of parameter estimates. A quasi-likelihood approach can be employed to correct for the overdispersion phenomenon which occurs with binary data [15]. For binary data, an alternative method for handling overdispersion is to use models for which a binomial parameter itself has a beta distribution [16]. However, the MLE of the beta-binomial method has been shown to be less flexible than that of the quasi-likelihood approach, and so cannot be recommended for general use. For comparisons of CI methods in previous literature, the overdispersion phenomenon needs to be taken into account for constructing the CI of the proportion. Therefore, the objective is to use quasi-likelihood methodology to evaluate several asymptotic CI methods and construct bootstrap CI based on these methods for groups of different sizes, adjusted for overdispersion in group testing.

Suppose for $i = 1, \cdots, m$ and n_i groups of size k_i, are tested, and Y_i of these groups test positive. The binomial parameter for the distribution of Y_i, which is the probability the groups tests positive is, $1 - (1 - p)^{k_i}$, where p is the probability an individual unit tests positive. Therefore considering overdispersion, the expected value and the variance of Y_i are

$$E(Y_i) = N_i \left(1 - (1 - p)\right)^{k_i} \tag{1}$$

$$Var(Y_i) = \sigma^2 N_i \left(1 - (1 - p)^{k_i}\right)(1 - p)^{k_i} \tag{2}$$

When groups of unequal size occur and overdispersion occurs together, the variance of the quasi-score function of p is

$$i_p = \frac{1}{\sigma^2} \sum_{i=1}^{m} \frac{k_i^2 N_i (1 - p)^{k_i - 2}}{1 - (1 - p)^{k_i}} \tag{3}$$

The maximum quasi-likelihood estimate of p denoted as $\hat{p}$, is the solution to

$$\sum_{i=1}^{m} k_i N_i = \sum_{i=1}^{m} \frac{k_i Y_i}{1 - (1 - p)^{k_i}} \tag{4}$$

where

$$\hat{\sigma}^2 = \frac{1}{m-1} \sum_{i=1}^{m} \frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} = \frac{X^2}{m-1} \tag{5}$$

And

$$\hat{\mu}_i = N_i \left(1 - (1 - \hat{p})^{k_i}\right) \tag{6}$$

$V(.)$ is a variance function, and X^2 is the generalized Pearson statistic.

Under the usual limiting conditions on the i_p, the asymptotic variance of $\hat{p}$ is

$$Var(\hat{p}) = i_p^{-1} \tag{7}$$

Then the $100(1 - \alpha)\%$ Wald type CI for p is $\hat{p} \pm z_{\alpha/2} \sqrt{\dfrac{1}{i_{\hat{p}}}}$

where $Z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution.

2.2.2. Confidence Intervals Based on the Logit

Asymptotic-likelihood methods to construct the $100(1-\alpha)\%$ confidence intervals of $\hat{p}$ using the Logit transformations were presented [11].

Let $I(\theta)$ be the Fisher information that Y contains about the parameter θ. If $\theta = h(\varepsilon)$ and h is differentiable, then the information that Y contains about ε is $I(\varepsilon) = I[h(\varepsilon)] \cdot [h'(\varepsilon)]^2$

For the logit transformation

$$h(p) = \log it(p) = \log\left(\frac{p}{1-p}\right) \tag{8}$$

$$h'(p) = \frac{1}{p(1-p)} \tag{9}$$

Thus the information of logit (p) is

$$I[\log it(p)] = p^2(1-p)^2 I(p) \tag{10}$$

After adjusting for overdispersion, the $100(1-\alpha)\%$ approximate CI for *p* is

$$\log it^{-1}\left[\log it\hat{p} \pm z_\alpha \sqrt{\frac{1}{\hat{p}^2(1-\hat{p})i_{\hat{p}}}}\right] \tag{11}$$

where

$$\log it^{-1}(\theta) = \exp(\theta)/(1+\exp(\theta)) \tag{12}$$

2.2.3. Confidence Intervals Based on the Complementary Log Log (CLL)

Hepworth also presented the CLL parameter transformations.

Suppose

$$h(p) = CLL(p) = \log(-\log(1-p)) \tag{13}$$

And

$$h'(p) = \frac{-1}{(1-p)\log(1-p)} \tag{14}$$

The information of $CLL(p)$ is

$$I[CLL(p)] = (1-p)^2 \left[\log(1-p)^2 I(p)\right] \tag{15}$$

Then the $100(1-\alpha)\%$ approximate CI for *p* after allowing for overdispersion is given by

$$CLL^{-1}\left[CLL(\hat{p}) \pm z_{\alpha/2} \sqrt{\frac{1}{(1-\hat{p})^2\left[\log(1-\hat{p})\right]^2 i_{\hat{p}}}}\right] \tag{16}$$

$$\text{where} \quad CLL^{-1}(\theta) = 1 - \exp\{-\exp(\theta)\} \tag{17}$$

2.3. Choice of Group Sizes

A lot of attention has been devoted in studies to problems involving the group sizes, *k*. The question being what values should be chosen for *k*, the number of units in each group. If *k* is too large, π is close to 1 and all groups are likely to test positive, also if *k* is too small, π is closer to 0 [17]. When *k* is large group testing becomes uninformative and expensive. Therefore it is suitable to choose group sizes that produce some positives and some negatives. We therefore ensured that the sample sizes used yielded the desirable level, that is, margin of error is minimized.

2.4. Interval Width and Coverage Probability

Each Interval was examined for its Interval width and coverage probability. Confidence Intervals based on MLE ordering and Sterne's technique for a dilution assay with 64 outcomes was compared [18], those with the shorter widths were considered superior. The width of the intervals is considered an important criterion in comparing CI methods. Interval widths of the intervals were examined since the three methods; Wald, Logit and CLL do not differ a great deal with respect to coverage.

Coverage probability gives the proportion of the time that the interval contains *p*. Bootstrapping enabled us to estimate the coverage of the CIs with respect to the number of hypothetical repetitions of the entire procedure. The nominal coverage will be set at 0.95. The actual coverage of the three methods will be compared to the nominal coverage, greater coverage than the nominal is preferable and hence conservative.

3. Simulation

First, positive groups, $Y_i \sim bin(N_i, \pi_i), i = 1, 2, \cdots, m$, are simulated from the number of groups, N_i with their respective sizes k_i for $p = 0.02$. This is followed by bootstrapping Y_i's and then computing the estimate of *p*. Coverage probabilities of the three interval methods for the estimate of *p* for 500 bootstrap simulations are obtained, and the program is run repetitively in order to see how the three method's coverage probabilities vary.

The nominal coverage probability is set at 0.95.

This is then compared to estimating *p* from the same data that is the Y_i's without bootstrapping and the program run repetitively so as to calculate the coverage probability.

The number of bootstraps B's were varied for the same combination of *p*, $k_{i'}$ and N_i. Coverage probabilities are generated. This was to assess the performance of the coverage probabilities as the number of bootstrap simulations increase for each method.

Next *p* was varied for the same N_i, k_i and the same number of bootstrap simulations (500), and coverage probabilities are obtained.

Lastly, interval widths for the data are calculated for bootstrapping technique for the three interval methods and compared to those intervals got from each of the three methods without bootstrapping. This is in order to investigate whether bootstrapping gives more precise intervals.

4. Discussion

Group testing was primarily used to inspect individual members from a large population [1]. Therefore probability used in this study ranged from 0.01 to 0.2. From **Table 1** it was seen that the CLL and the Logit gave more conservative coverage probabilities consistently for different bootstrap simulations for *p* = 0.02. All the three methods differ only slightly with CLL and Logit sometimes taking the same values of coverage probability. The coverage probabilities oscillate about the nominal level with Wald being less conservative than CLL and Logit. This was then compared to coverage probabilities of the three methods for the same data combination without applying bootstrapping technique. Similarly CLL outperformed the Wald interval method and at times the Logit Interval method as seen in **Table 2**. It is clearly seen that without bootstrapping technique, CIs generated have lower coverage probabilities than CIs generated after applying bootstrapping.

In **Table 3** and **Figure 1**, the number of bootstrap simulations was varied. It was seen that as the number of

Table 1. Coverage probabilities for the three Interval methods for 500 simulations for *p* = 0.02.

Method	Wald	CLL	Logit
	0.938	0.9461	0.946
	0.954	0.948	0.948
$p = 0.02$	0.942	0.9441	0.943
	0.938	0.952	0.952
	0.95	0.958	0.958
	0.932	0.9481	0.948

Table 2. Coverage probabilities for data for 500 simulations without bootstrapping for $p = 0.02$.

Method	Wald	CLL	Logit
	0.812	0.862	0.862
	0.822	0.873	0.87
	0.832	0.912	0.91
$p = 0.02$	0.81	0.852	0.85
	0.802	0.844	0.844
	0.804	0.812	0.81

Table 3. Coverage probabilities for the three methods for varying bootstrap simulations (B) for $p = 0.02$.

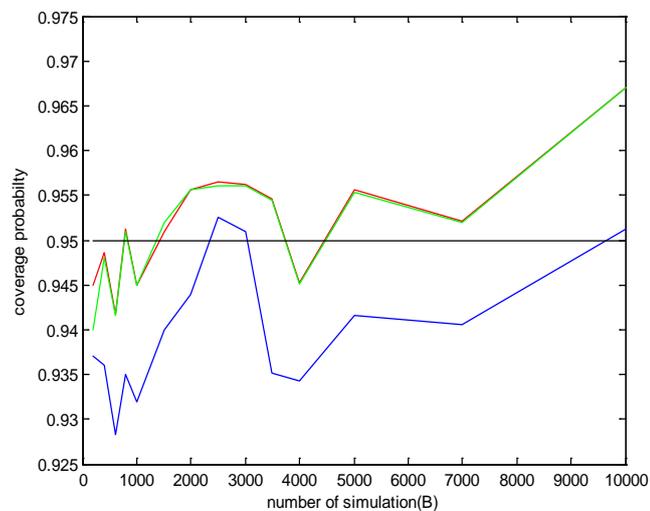| Number of Simulations (B)	Wald	CLL	Logit
200	0.937	0.945	0.94
400	0.936	0.9486	0.948
600	0.9283	0.9417	0.9416
800	0.935	0.9512	0.951
1000	0.932	0.945	0.945
1500	0.94	0.951	0.952
2000	0.944	0.9556	0.9556
2500	0.9525	0.9565	0.956
3000	0.951	0.9562	0.9561
3500	0.93514	0.95457	0.9544
4000	0.93425	0.94525	0.9451
5000	0.9416	0.9556	0.9554
7000	0.9405	0.9521	0.952
10,000	0.9512	0.967	0.967

Figure 1. The figure shows a plot of coverage probabilities for the three methods for varying bootstrap simulations for $p = 0.02$ and for a nominal coverage probability of 0.95.

bootstrap simulations increased, all the three methods became more precise, that is, their coverage probabilities hovered slightly above and below the nominal coverage probability.

Group testing works best when the prevalence rate or the proportion is sufficiently small. According to **Table 4** and **Figure 2**, we see that as the proportion increases to 0.1 the coverage probability becomes very low for all the three interval methods. Between 0.03 and 0.08, the coverage probabilities for the three methods are conservative.

From **Table 5** and **Table 6**, it is seen that the interval widths for both CLL and Logit are not differing a great deal for the same bootstrap simulations. CLL has shorter widths than logit. However, interval widths for the Wald for the three bootstrap simulations are considerably wider than those of the other two interval methods. When interval widths are compared without bootstrap simulations to those interval widths with bootstrap simulations we note that they are wider.

According to **Table 4** and **Figure 2**, all the three methods perform well as the number of bootstrap simulations increases, with the coverage probabilities of the CLL and Logit being oftentimes greater than the nominal coverage probability.
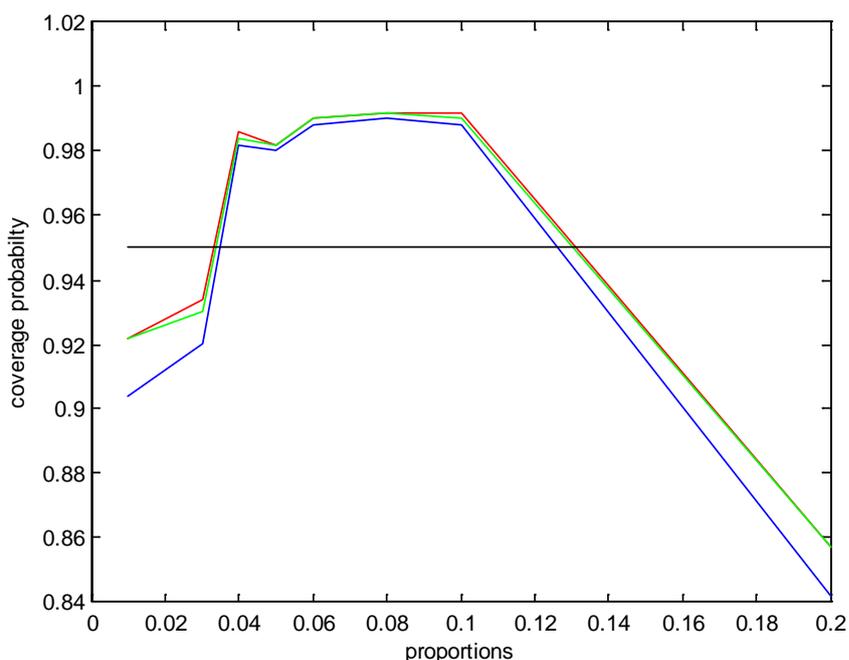

Figure 2. The figure above is a plot of coverage probabilities for the three methods for 500 bootstrap simulations for varying probability, that is, $0.01 < p < 0.2$.

Table 4. Coverage probabilities for varying p for the same 500 simulations for a nominal coverage of 0.95.

Probability	Coverage Probability		
	Wald	CLL	Logit
0.01	0.904	0.922	0.922
0.03	0.92	0.934	0.93
0.04	0.982	0.986	0.984
0.06	0.988	0.99	0.99
0.08	0.99	0.992	0.99
0.1	0.988	0.992	0.99
0.2	0.842	0.857	0.857

Table 5. Table of interval widths for the three interval methods for data without bootstrapping.

Interval Method	Y = (2,2,4,3,3)	$\hat{p} = 0.02121051$	
	Upper Confidence Limit	Lower Confidence Limit	Interval Width
Wald	0.02654614	0.01587489	0.01067125
CLL	0.02726701	0.01648786	0.01077915
Logit	0.02725691	0.01648266	0.01077426

Table 6. Table of interval widths for the three interval methods after bootstrapping.

Interval Method	Y = (2,2,4,3,3)	$\hat{p} = 0.02142545$	
	Upper Confidence Limit	Lower Confidence Limit	Interval Width
Wald	0.02673085	0.016212005	0.0106108
CLL	0.02743541	0.01672070	0.01071471
Logit	0.02742548	0.01671553	0.01070995

5. Conclusion

Bootstrap methods produce good confidence intervals by an order of magnitude upon the accuracy of standard intervals [19]. In this paper we implement bootstrapping to the three interval methods: Wald, CLL and Logit. The comparison results in this study show that bootstrapping produces more precise confidence intervals than standard intervals for all the three procedures. Likewise the coverage probabilities from the three methods are consistent with the CLL being most conservative. Wald is more liberal. As the number of bootstrap simulations increases the more conservative the coverage becomes.

References

[1] Dorfman, R. (1943) The Detection of Defective Members of Large Populations. *The Annals of Mathematical Statistics*, **14**, 436-440. http://dx.doi.org/10.1214/aoms/1177731363

[2] Fletcher, J., Russell, A. and Butler, R. (1999) Seed-Borne Cucumber Mosaic Virus in New Zealand Lentil Crops: Yield Effects and Disease Incidence. *New Zealand Journal of Crop and Horticulture Science*, **27**, 197-204. http://dx.doi.org/10.1080/01140671.1999.9514097

[3] Worlund, D. and Taylor, G. (1983) Estimation of Disease Incidence in Fish Populations. *Canadian Journal of Fisheries and Aquatic Science*, **40**, 2194-2197. http://dx.doi.org/10.1139/f83-254

[4] Swallow, W.H. (1985) Group Testing for Estimating Infection Rates and Probabilities of Disease Transmission. *Phytopathology*, **75**, 882-889. http://dx.doi.org/10.1094/phyto-75-882

[5] Kline, R.L., Bothus, T., Brookmeyer, R., Zeyer, S. and Quinn, T. (1989) Evaluation of Human Immunodeficiency Virus Seroprevalence in Population Surveys Using Pooled Sera. *Journal of Clinical Microbiology*, **27**, 1449-1452.

[6] Liu, S.-C., Chiang, K.-S., Lin, C.-H. and Deng, T.-C. (2010) Confidence Interval Procedures for Proportions Estimated by Group Testing with Groups of Unequal Size Adjusted for Overdispersion. *Journal of Applied Statistics*, **38**, 1467-1482. http://dx.doi.org/10.1080/02664763.2010.505953

[7] Paul, S. and Islam, A.S. (1998) Joint Estimation of the Mean and Dispersion Parameters in the Analysis of Proportions: A Comparison of Efficiency and Bias. *Canadian Journal of Statistics*, **26**, 83-94. http://dx.doi.org/10.2307/3315675

[8] Saha, K.K. and Paul, S.R. (2005) Bias-Corrected Maximum Likelihood Estimator of the Negative Binomial Dispersion Parameter. *Biometrics*, **61**, 179-185. http://dx.doi.org/10.1111/j.0006-341X.2005.030833.x

[9] Williams, D.A. (1975) The Analysis of Binary Responses for Toxicological Experiments Involving Reproduction and Teratogenicity. *Biometrics*, **31**, 946-952. http://dx.doi.org/10.2307/2529820

[10] Saha, K.K. (2008) Semi-Parametric Estimation for the Dispersion Parameter in the Analysis of over- or Underdispersed Count Data. *Journal of Applied Statistics*, **35**, 1383-1397. http://dx.doi.org/10.1080/02664760802382459

[11] Hepworth, G. (2005) Confidence Intervals for Proportions Estimated by Group Testing with Groups of Unequal Size. *Journal of Agricultural, Biological, and Environmental Statistics*, **10**, 478-497. http://dx.doi.org/10.1198/108571105X81698

[12] Efron, B. (1979) Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, **7**, 1-26. http://dx.doi.org/10.1214/aos/1176344552

[13] Siele, D. (2012) Interval Estimation for the Beta Binomial Dispersion Parameter. MSc Thesis, Egerton University, Njoro.

[14] Saha, K. and Sen, D. (2009) Improved Confidence Interval for the Dispersion Parameter in Count Data Using Profile Likelihood. Technical Report No. 4/09.

[15] Williams, D.A. (1982) Extra-Binomial Variation in Logistic Linear Models. *Journal of the Royal Statistical Society. Series C* (*Applied Statistics*), **31**, 144-148. http://dx.doi.org/10.2307/2347977

[16] Crowder, M.J. (1978) Beta-Binomial Anova for Proportions. *Applied Statistics*, **27**, 34-37. *Mathematical Statistics*, **14**, 436-440.

[17] Hepworth, G. (1996) Exact Confidence Intervals for Proportions Estimated by Group Testing. *Biometrics*, **52**, 1134-1146. http://dx.doi.org/10.2307/2533075

[18] Loyer, M.W. and Hamilton, M.A. (1984) Interval Estimation of the Density of Organisms Using a Serial-Dilution Experiment. *Biometrics*, **40**, 907-916. http://dx.doi.org/10.2307/2531142

[19] Efron, B. and DiCiccio, T. (1996) Bootstrap Confidence Intervals. *Statistical Science*, **11**, 189-228. http://dx.doi.org/10.1214/ss/1032280214