# Mathematical Analysis of propensity of aberration on the methods for interval estimation of the multinomial proportions

## U.Sangeetha[1], M.Subbiah[2], M.R.Srinivasan[3]

*[1](Department of Management Studies, SSN College of Engineering, Chennai)*
*[2](Department of Mathematics, L. N. Government College, Ponneri)*
*[3](Department of Statistics, University of Madras, Chennai)*

**Abstract:** *The multinomial distribution has found applications in many practical situations that deal with a discrete random variable with k possibilities (k > 2). Interval estimation for the proportion parameter in a special case k = 2, the binomial distribution has been studied extensively in literature. However for k > 2, studies have focused the performance of estimation procedure mainly on coverage probabilities and yet there are other important aspects such as propensity of aberration in the limits of confidence intervals and computational issues. The present paper makes an attempt to look beyond coverage probabilities by marshalling the existing procedures in classical and Bayesian approaches for k > 2. To alleviate the computational issues, a comprehensive R program is also made available to facilitate the implementation of the procedures in both classical and Bayesian statistical paradigms.*

**Keywords:** *Aberrations, Bayesian, Contingency Tables, Multinomial, Zero Width Intervals.*

## I.  INTRODUCTION

Probability estimation has drawn an active research attention as it has found wider applications in many areas. Statistical methods for estimating single binomial proportion or associated arithmetic forms of two proportions have discussed extensively in statistical literature (Subbiah, 2009). This includes point and interval estimation procedures and related properties such as coverage probabilities, expected length and existence of unrealistic bounds (Newcombe, 1998 and Sweeting, 2004)

The problem of estimating simultaneous confidence intervals for multinomial proportion has witnessed an active discussion in terms of theoretical, computational and application aspects. Different procedures have been discussed extensively in the literature. Quesenberry and Hurst (1964), Goodman (1965), Fitzpatrick and Scott (1987), Sison and Glaz (1995), Glaz and Sison (1999) can be considered as an earlier work. Also similar to binomial proportion problems, Wald type intervals and Wilson intervals are also available. Further boot strap methods (Jhun and Jeong, 2000), power divergence methods (Hou et al, 2003) are discussed for developing simultaneous confidence intervals. Wang (2008) has proposed exact confidence coefficient; Chafai and Concordat (2009) has discussed the problem especially for small samples.

Bayesian procedures have also been discussed in obtaining simultaneous confidence coefficients for multinomial proportion. A conjugate prior model, multinomial – Dirichlet has been discussed extensively in many literature and Gelman et al (2002) provides the details of the relations and effect of hyper parameters; Tuyl et al (2009) have shown that the uniform prior could be considered as the consensus prior for multinomial parameters and recently, Komaki (2012) has developed a prior based on a specific Dirichlet prior that would be asymptotically minimax.

Further, the performance of methods is studied based on coverage probability length of intervals and computational flexibility; small and large sample sizes; polarized cell counts are also considered in this comparison. However, aberrations are not elaborately discussed especially on small samples though direct adjustments on such limits are recommended in practice. Computational issues have also been discussed in many studies and in particular, May and Johnson (1997, 2000) have listed the computational tools for most of the available methods, however the availability of these procedures are limited to selected software so that most of the applications might rely on a specific method or a software.

This paper has two fold objectives: discussion of possible aberration more generally for widely applicable procedures in classical and Bayesian approaches. Similar discussions for other estimation problems (for example, Newcombe, 1998) only illustrate the propensity of overshoot and a more general way to understand such existence has been attempted to classify the methods.

Secondly, limited availability of procedures like Sison and Glaz in selected software that may deter its usage (http://www.rforge.net/doc/packages/NCStats/gofCI.html) and an attempt has been made to develop comprehensive computational tools for most of the methods in both classical and Bayesian methods. R programming environment has been used to provide a tool for obtaining multinomial interval estimate based on

Wald, Wilson, Quesenberry and Hurst (QH), Goodman (GM), Fitzpatrick and Scott (FS), Sison and Glaz (SG) methods in classical inference and multinomial-Dirichlet distribution method in Bayesian procedure. Numerical examples are presented to compare these methods especially on the small counts. Findings of this work have encouraged to further investigate sparse data sets in large two dimensional categorical tables that are being analyzed under multinomial sampling.

## II.     METHODS

To depict the multinomial sampling which is generally represented as a contingency table following list of notations could be useful

I = No. of rows; J = No. of columns; k = IxJ

$n_i$ = Cell counts

n = $\sum n_i$ (Size of the sample)

$\pi_i$ = Population proportion of $i^{th}$ cell where i = 1, 2,…,k

$p_i$ = Sample proportion

MLE estimation of $\pi_i = \frac{n_i}{N}$

$(p_i)_{B.}$ = Bayesian estimation of $\pi_i$, and a dot represents the prior

$A = \chi^2(\alpha, 1)$ i.e, the upper $100(1-\alpha)\%$point of the chi square distribution

$A_1 = \chi^2(\alpha, k-1)$

$A_2 = \chi^2(\alpha/k, 1)$

$Z_\alpha = 100(1-\alpha)\%$ point of the standard normal distribution

Queensberry and Hurst (1964), have proposed interval based on Pearson statistic  that has limits as $\frac{(2np_i + A_1) \mp \sqrt{A_1(A_1 + 4np_i(1-p_i))}}{2(A_1 + n)}$. Fitzpatrick and Scott (1987) have proposed confidence interval as $p_i \pm Z_{\alpha/2} \frac{1}{2\sqrt{n}}$ .   A Wald type interval based on the multivariate central limit theorem has the form $p_i \pm \sqrt{\frac{Ap_i(1-p_i)}{n}}$ and subsequently a continuity correction version  of these limits is of the form $p_i \pm \sqrt{\frac{Ap_i(1-p_i)}{n}} + \frac{1}{2n}$ .

Further Sison and Glaz (1995) have proposed a method to determine simultaneous confidence intervals for multinomial proportions; one method is based on the truncated Poisson random variable and associated central and factorial moments. In this method, cell counts are assumed to be independent Poisson random variable with parameter $\lambda_i = n\pi_i$ Since $\pi_i$ is unknown, $\lambda_i$ is estimated as $np_i = n_i$

If $A_i = \{n_i/b_i \leq n_i \leq a_i\}$ then $p[A_i,.. A_k/\sum n_i = n] = \frac{n!}{n^n e^{-n}} \prod_{i=1}^k p(b_i \leq n_i \leq a_i)p(w=n)$ where $w = \sum_{i=1}^k Y_i$ and $Y_i$ is the truncated Poisson variable in $[b_i, a_i]$

$$\therefore p[b_i \leq n_i \leq a_i] = \frac{n!}{n^n e^{-n}} \prod_{i=1}^k p(b_i \leq Z_i \leq a_i)p(w=n)$$

Also, $n_i \sim$ Poisson $[\lambda_i]$, and hence $p[b_i \leq Z_i \leq a_i] = \sum_{t=b_i}^{a_i} p(Z_i = t)$

$$= \sum_{t=0}^{a_i} p(Z_i = t) - \sum_{t=0}^{a_i} p(Z_i = t)$$

$$= F(a_i) - F(b_j - 1) \text{ where F is CPDF of Poisson random variable.}$$

Further, the central factorial moments $\mu_{(r)} = \left[\frac{F(a-r) - F(b-r-1)}{F(a) - F(b-1)}\right]$ of truncated Poisson variable $Y_i$ and the central moments (up to r=4) of $Y_i$ are calculated to obtain p(w=n). Then the limits of  $100(1-\alpha)$ % confidence limits are obtained as $LL = p_i - \frac{C}{n}$ and $UL = p_i + \frac{C}{n} + \frac{2\delta}{n}$   where C is a positive integer that satisfies the condition $\rho(C) \leq 1-\alpha \leq \rho(C+1)$ with $\rho(C) = p[p_i - C/n \leq \pi_i \leq p_i + C/n] = 1-\alpha$  and $\delta = \frac{(1-\alpha) - \rho(C)}{\rho(C+1) - \rho(C)}$.

The Bayesian approach for estimating cell probabilities for a two way contingency table under multinomial sampling could be outlined with a two-step likelihood-prior combination. If the random vector $n = (n_1, n_2, …, n_k)$ has a multinomial distribution with n trials and cell probabilities $\pi_1, \pi_2, …………. \pi_k$ then the joint pmf of $(n_1, n_2, …, n_k)$ is

$$f(n_1, n_2, …, n_k) = \frac{n!}{n_1! n_2! … n_k!} \pi_1^{n_1} \pi_2^{n_2} … \pi_k^{n_k}$$

The conjugate prior (Gelman et al, 2002) for the proportion parameters $\theta = (\pi_1, \pi_2, \ldots \ldots \ldots \pi_k)$ could be a multivariate generalization of Beta distribution known as Dirichlet $(\alpha_1, \alpha_2, \ldots \ldots \alpha_k)$ with $\alpha_j > 0$ and PDF is $p(\theta) = \frac{\Gamma(\sum \alpha_i)}{\prod \Gamma(\alpha_i)} \pi_1^{\alpha_1-1} \pi_2^{\alpha_2-1} \ldots \ldots \ldots \pi_k^{\alpha_k-1}$, $\sum \pi_j = 1$. Hence, the posterior distribution will be $\Pi(\theta/X) \alpha \pi_1^{\alpha_1+x_1-1} \ldots \ldots \ldots \ldots \pi_k^{\alpha_k+x_k-1} = $ Dirichlet $(\alpha_1+x_1, \alpha_2+x_2, \ldots \ldots \alpha_k+x_k)$ up to normalizing constant. For a non-informative prior, widely chosen values for $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ includes $\alpha_i = 1$ corresponds to the uniform prior distribution, Jeffrey's prior $\alpha_i = 1$ and $\alpha_i = 1.41$ corresponds to asymptotic prior (Komaki, 2012) distribution.

### III. CONDITIONS FOR ABERRATIONS

In most of the studies, overshoot or zero width intervals have been illustrated based on the low (zero or positive) cell counts. However, a more general scenario can be derived by considering the mathematical form of each method to indicate the possible chances of aberrations. This section includes the possibilities of cells with zero counts and the range of cell counts that may result with over shooting estimated values for the associated lower and upper limits (LL & UL) of confidence intervals.

In Qusenberry and Hurst, if $n_i = 0$ for some 'i' then LL $= \frac{(0+A_1)-\sqrt{A_1(A_1+0)}}{2(A_1+n)} = 0$ and UL $= \frac{A_1}{A_1+n}$. that yield fixed limits for the cells with zero count.

Also LL < 0 is true

$$\text{only if } A_1 + 2np_i - \sqrt{A_1\left(A_1 + 4np_i(1-p_i)\right)} < 0$$
$$\text{only if } \left(A_1 + 2np_i\right)^2 < A_1^2 + 4nA_1p_i(1-p_i)$$
$$\text{only if } A_1^2 + 4n^2p_i^2 + 4np_iA_1 < A_1^2 + 4nA_1p_i - 4nA_1p_i^2$$
$$\text{only if } 4n^2p_i^2 < -4nA_1p_i^2$$
$$\text{only if } n < -A_1 \text{ which is not possible.}$$

Hence QH never have negative lower limits.

Similarly UL > 1 is true only if $\left(2np_i + A_1\right) + \sqrt{A_1(A_1) + 4nAp_i(1-p_i)} > 2(A_1+n)$ and subsequent algebra will yield a condition $0 > n(1-p_i) + A_1(1-p_i)$ which is not possible (since n, $(1-p_i)$, $A_1$ are positive) and hence QH will never provide UL which are more than 1. Since, Goodman and Wilson differ from Qusenberry and Hurst only in the $\chi^2$ value $(A, A_1, A_2)$ similar conclusions can be made regarding these two procedures.

In Fitzpatrick and Scott method, when $n_i = 0$ corresponding LL and UL will be $\mp \frac{Z_{\alpha/2}}{2\sqrt{n}}$ that obviously yield a negative LL.

Also LL < 0 only if
$$p_i - Z_{\alpha/2} \frac{1}{2\sqrt{n}} < 0$$
$$p_i < Z_{\alpha/2} \frac{1}{2\sqrt{n}}$$
$$0 < n_i < \sqrt{n} \frac{Z_{\alpha/2}}{2}$$

This could be possible and such a case could also be attributed towards polarized cell counts with varying k. For example, if n = 36, $1 \le n_i \le 5$ will have negative lower limits for a 95% confidence interval. Further, in this case if k is small say 4 and $n_i < 5$ for at least one i (say i=1) remaining three cells must be added upto 32. A hypothetical example (1,14,12,9) would bring a negative Lower Limit for the cell i=1.

Also, in this case, UL > 1 only if $p_i + Z_{\alpha/2} \frac{1}{2\sqrt{n}} > 1$, $p_i > 1 - Z_{\alpha/2} \frac{1}{2\sqrt{n}}$ only if $n_i > n - \frac{\sqrt{n}}{2} Z_{\alpha/2}$. This case is also possible in practice and that might indicate the extremely distant values in the data set. Though FS has a symmetry property, aberration is an issue which would exist in possible cases. Hypothetical data set is (73, 1, 1, 6) will bring Upper Limit > 1 in first case and negative Lower Limit in all other 3 cases.

In Wald Type, when $n_i = 0$, both Lower Limit and Upper Limit are 0 which is an existence of zero width interval. Also, LL < 0 only if $p_i - \sqrt{\frac{A p_i(1-p_i)}{n}} < 0$

$$p_i^2 < \frac{A p_i(1-p_i)}{n}$$
$$p_i < \frac{A(1-p_i)}{n}$$
$$n^2 p_i < A(n-n_i)$$
$$nn_i < An - An_i$$

$$n_i < \frac{n}{(n+A)} A = \left(1 - \frac{A}{n+A}\right) A = (AF)A \text{ where AF is an adjusting factor}$$

$$\left(1 - \frac{A}{n+A}\right)$$

Also, UL > 1 Only if

$$p_i + \sqrt{\frac{Ap_i(1-p_i)}{n}} > 1$$

$$\frac{Ap_i(1-p_i)}{n} > (1-p_i)^2$$

$$\frac{Ap_i}{n} > 1-p_i$$

$$\frac{Ap_i}{n} > 1-p_i$$

$$An_i > n^2\left(\frac{n-n_i}{n}\right) = n(n-n_i)$$

$$n_i > \left(1 - \frac{A}{n+A}\right) n$$

However if n is large then AF→1 thereby $n_i$→n in $i^{th}$ cell. This means $n_J$→0 as j≠i and could be an illustration of polarized cell counts.

In Wald CC method, if $n_i$=0, then limits for the corresponding intervals are $\mp\frac{1}{2n}$ that yield an

obvious negative limits. Also, LL < 0 only if $p_i - \sqrt{\frac{Ap_i(1-p_i)}{n}} - \frac{1}{2n} < 0$

$$\left(p_i - \frac{1}{2n}\right)^2 < \frac{Ap_i(1-p_i)}{n}$$

$$\frac{n_i^2}{n^2} + \frac{1}{4n^2} - \frac{n_i}{n^2} < \frac{A}{n} p_i \frac{A}{n} p_i^2$$

$$\frac{n_i^2}{n^2} + \frac{1}{4n^2} - \frac{n_i}{n^2} < \frac{(n-n_i)An_i}{n^3}$$

$$n_i^2[n+A] - n_i[n+nA] + \frac{n}{4} < 0$$

$$n_i^- < n_i < n_i^+ \text{ where } n_i^{\pm} = \frac{n(n+A) \pm \sqrt{n^2A^2 + 2n^2A - nA}}{2(n+A)}$$

For, Sison and Glaz method $p_i = \frac{n_i}{n}$ $LL = \frac{n_i}{n} - \frac{C}{n}$ and hence whenever C>$n_i$ (for some i), LL could be negative in $i^{th}$ confidence interval. Similarly, $p_I + \frac{C}{n} + \frac{2\delta}{n} > 1$ only if $n_i + C + 2\delta > n$, only if $n_i > n - (C+2\delta)$ where δ depends on the distance between p(C), 1-α and p(C+1), such that δ≤1 so that -δ≥-1 ⇒$n_i > n-(C+2\delta) > (n-1)-C$ or equivalently C>(n-1)-$n_i$ For any i then UL of $i^{th}$ CI will be more than 1.

## IV. Data Analysis

To illustrate the notion of estimation and the cases in which negative lower limits occur, four data sets (referred to as I - IV) have been considered and presented in TABLE 1. The first data set have been used in Chafai and Concordet (2009) which deal with the difference in behavior of male and female students with respect to smoking habits; second one from Szyda et al (2008) that has been used to analyze DNA pooling experiments with arctic fox; a data set from Sison and Glaz (1995), Hou et al (2003) is the third set; The fourth one is a partial collection of data sets from the data repository (www.fars-nhtsa.gov/states/statesAlcohol.aspx) that deals with fatal crashes and related information from National Highway Traffic Safety Administration's (NHTSA) Fatality Analysis Reporting System (FARS) and its supplemental files. FARS has a collection of most or all fatal motor vehicle crashes that occur in United States of America. Further, alcohol related files are made up of the counts of crashes and fatalities due to three different levels of blood level consumptions (BAC equals to 0.00, between 0.01-0.07 and 0.08 or more). However, only nine tables for the year 2009 with wider range of cell counts have been included for the analysis.

The data sets are collected in such a way to describe the presence of zeros; sparse non zero counts and has wider range. Also, minimum count in I and III data sets is 3 and 56 respectively and that of IV varies from 2 to 14 over the nine tables. It has been observed that some procedures yield obvious negative lower limits whereas Wald method provides zero width intervals (ZWI) for zero counts. However, the main interest of the study lies on the aberrations that could occur for non-zero counts. In the case of data set I, uniformly three methods W, WCC, FS and SG yield negative lower limits when the cell count is 3 and such a result does not exist in the other methods. In fact QH, GM, and WS can be eliminated from such comparison as they never exhibit the characteristic of aberration and similar conclusions can be made for the Bayesian procedures too.

For data set II, Wald type intervals do not yield negative estimates for positive counts, whereas SG and FS result with negative lower limits when the cell counts are 5, 14 and 10; further, when the cell count is 16, SG has zero as lower limit but FS has a negative lower limit. In fact, the choice of c in the SG procedure will decide such values when any cell count is equal to C, lower limit is zero for the lesser cell counts it is negative. This kind of behavior warrants a deeper insight into deciding the sparse nature among positive cell counts and when cell counts are highly polarized.

Similar conclusions are visible in the case of all nine tables of data set IV but a careful observation throws the light for SG when the range of cell counts are very high as much as 176 does not result with a negative lower limit for the cell count 14 in contrast to data set II. Hence, a quick decision could not be arrived regarding the performance of SG when high polarized count exist (May and Johnson, 2000) in the given data set; but the value of C is a notably influencing number for such cases in the estimated values. On the other hand, with larger values (more than 56) present in data set III, none of the cells have aberrations though range is notably high.

Further, TABLE 2 and TABLE 3 summarize the interval estimates (LL: lower limit and UL: upper limit) for the multinomial proportions related to the data set III to facilitate a comparison of the seven classical with three Bayesian methods. Bayesian methods yield least volume when compared to all other classical methods and the estimates are quite closer to the methods due to Wilson and Wald CC. Comparisons among classical procedures have been done already in many studies such as Hou et al (2003), this work does not include similar effort among non-Bayesian methods. But an emphasis is made to note the advantage of using Bayesian direct posterior estimation from Dirichlet distribution with standard objective priors that will yield better results and involve lesser computation and never yield implausible estimates under different circumstances of cell counts and size of the I x J tables.

## V. CONCLUSION

Methods for estimating simultaneous multinomial proportions have found research attention in recent times but not specific to the unrealistic confidence intervals. For example Lee et al (2011) have pointed out the implausible limits for confidence interval as unrealistic. Though a simple practical remedy is visible such as truncating the limits suitably, a systematic method to understand such impractical estimates has not been developed at least in the case of multinomial proportions. This paper makes an attempt to derive suitable mathematical expression for the propensity of aberrations in the method for estimating multinomial proportions.

In the present work, seven classical procedures and Bayesian method with three priors are compared for the interval estimation of multinomial proportions. For the large samples and for an introductory level Wald type intervals can be suggested which are computationally simpler yet do not posses essential other properties (such as coverage probability) so as not to recommend for practical applications especially for low cell counts.

Though FS can be equally compared to Wald type intervals for its mathematical simplicity, this method is also not a favorable choice when the cell counts are small. Wilson intervals and similar forms QH and GM are providing plausible set of limits for all kind of data; but conservatism is an issue when k is comparatively large.

However, SG performs better in terms of achieving required coverage probability in all cases by the virtue of its mathematical form; yet it suffers a limitation in terms of propensity of aberration whenever the cell counts are polarized. Hence a careful judgment is needed to handle such limits which could occur even when a cell count is as large as 14. The availability of SG has also made wide spread in this work, and hence among the method considered in classical statistics, SG can be recommended with a caution about the range of cell counts. This paper provides a beforehand indication for the propensity of aberrations in confidence limits.

On the other hand, Bayesian Multinomial-Dirichlet methodology has been revisited with two well known objective priors (uniform and Jeffreys) together with a recently discussed asymptotic prior. It has been observed that Bayesian methods are quite flexible in modeling, mathematical treatment, computation and its uniform behavior over a variety of data, and never warrant corrections due to over shoot of the estimates. Further, advances in simulation techniques encourage adopting Bayesian objective methods to obtain interval estimates for multinomial proportions. Together with the existing literature on multinomial proportions, this study provides a consolidated view to select the procedures based on some criteria such as size of the contingency table (k), nature of cell counts and its range to indicate polarization, computational flexibility, propensity of aberrations, and desirable frequentist properties. However, the choice between Bayesian and classical methods need to be handled similar to other situations of data analyses.

### REFERENCES

[1]     M. Subbiah, *Bayesian inference on sparseness in 2 × 2 contingency tables and its applications*, Unpublished Ph.d Thesis., University of Madras, Chennai, 2009.

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

[3]     J.M. Sweeting, J.A. Sutton, & C.P. Lambert, What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine*, 23 (2004) 1351 –1375.
[4]     C.P. Queensberry, & D.C. Hurst, Large Sample Simultaneous Confidence Intervals for Multinational Proportions, *Technometrics,* 6 (1964) 191-195.
[5]     L.A. Goodman, On Simultaneous Confidence Intervals for Multinomial Proportions, *Technometrics*, 7 (1965) 247-254.
[6]     S. Fitzpatrick & A. Scott, Quick simultaneous confidence interval for multinomial proportions, *Journal of American Statistical Association*, 82 (1987) 875-878.
[7]     P.C. Sison & J. Glaz, Simultaneous Confidence Intervals and Sample Size Determination for Multinomial Proportions, *Journal of the American Statistical Association*, 90 (1995) 366-369.
[8]     J. Glaz & P.C. Sison, Simultaneous confidence interval for multinomial proportions,  *Journal of Statistical planning and inference*, 82 (1999) 251-262.
[9]     M. Jhun & H.C. Jeong, Applications of bootstrap methods for categorical data analysis, *Computational Statistics & Data Analysis*, 35 (2000) 83-91.
[10]    C.D. Hou, J. Chiang & J.J. Tai, A family of simultaneous confidence intervals for multinomial proportions, *Computational Statistics & Data Analysis*, 43 (2003) 29-45.
[11]    H. Wang, Exact confidence coefficients of simultaneous confidence intervals for multinomial proportions, *Journal of Multivariate Analysis,* 99 (2008) 896-911.
[12]    D. Chafai, D. Concordet, Confidence Regions for the multinomial parameter with small sample size, *Journal of American Statistical Association*, 104 (2009) 1071-1079.
[13]    A. Gelman, J.B. Carlin, H.S. Stern & D.B. Rubin, *Bayesian Data Analysis* (London: Chapman & Hall, 2002)
[14]    F. Tuyl, R. Gerlach & K. Mengersen, Posterior predictive arguments in favor of the Bayes-Laplace prior as the consensus prior for binomial and multinomial parameters, *Bayesian Analysis*, 4 (2009) 151-158.
[15]    F. Komaki, Asymptotically minimax Bayesian predictive densities for multinomial models, *Electronic Journal of Statistics*, 6 (2012) 934-957.
[16]    L.W. May & D.W. Johnson, Constructing simultaneous confidence intervals for multinomial proportions, *Computer Methods and Programs in Biomedicine*, 53 (1997) 153-162.
[17]    J. Szyda, Z. Liu, D.M. Zaton, H. Wierzbicki & A. Rzasa, Statistical aspects of genetic association testing in small samples, based on selective DNA pooling data in the arctic fox. *Journal of Applied Genetics* 49 (2008) 81-92.
[18]    Lee, P.R. Armstrong, J.A. Thomasson, R. Sui & M. Casada, Application of binomial and multinomial probability statistics to the sampling design process of a global grain tracing and recall system. *Food Control,* 22 (2009) 1085-1094.

Table 1: Illustrative data sets and their basic characteristics

| Data sets | Size of the original table | No. of Cells | Zero entries | | Range of table totals |
|---|---|---|---|---|---|
| | | | No | % | |
| I | 2 x 2 | 4 | 0 | 0 | 7 |
| II | 4 x 5 | 20 | 12 | 60 | 66 |
| III | 1 x 7 | 7 | 0 | 0 | 31 |
| IV | 1 x 3 (9 tables) | 3 | 0 | 0 | 39,64,15, 46,66,75, 176,42,74 |

Table 2: Comparison of six 95% confidence interval resulted from classical procedures for the data set III

| Cell | Wald | | Wald CC | | Wilson | | Qusenberry-Hurst | | Goodman | | Fitzpatrick-Scott | | Sison-Glaz | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LL | UL | LL | UL | LL | UL | LL | UL | LL | UL | LL | UL | LL | UL |
| 1 | 0.090 | 0.149 | 0.089 | 0.150 | 0.094 | 0.153 | 0.076 | 0.183 | 0.085 | 0.166 | 0.075 | 0.165 | 0.079 | 0.164 |
| 2 | 0.121 | 0.187 | 0.120 | 0.188 | 0.124 | 0.190 | 0.104 | 0.222 | 0.115 | 0.204 | 0.109 | 0.200 | 0.114 | 0.199 |
| 3 | 0.123 | 0.189 | 0.122 | 0.190 | 0.126 | 0.192 | 0.106 | 0.225 | 0.116 | 0.207 | 0.111 | 0.202 | 0.116 | 0.201 |
| 4 | 0.096 | 0.156 | 0.095 | 0.158 | 0.099 | 0.160 | 0.081 | 0.191 | 0.091 | 0.173 | 0.081 | 0.172 | 0.086 | 0.171 |
| 5 | 0.102 | 0.164 | 0.101 | 0.165 | 0.105 | 0.167 | 0.087 | 0.198 | 0.096 | 0.181 | 0.087 | 0.178 | 0.092 | 0.177 |
| 6 | 0.151 | 0.222 | 0.150 | 0.223 | 0.154 | 0.224 | 0.131 | 0.258 | 0.143 | 0.239 | 0.141 | 0.232 | 0.146 | 0.231 |
| 7 | 0.094 | 0.154 | 0.093 | 0.155 | 0.097 | 0.157 | 0.080 | 0.188 | 0.089 | 0.171 | 0.079 | 0.170 | 0.084 | 0.169 |
| Volume | $4.07 \times 10^{-7}$ | | $5.14 \times 10^{-9}$ | | $4.08 \times 10^{-9}$ | | $2.52 \times 10^{-7}$ | | $3.59 \times 10^{-8}$ | | $5.11 \times 10^{-8}$ | | $3.24 \times 10^{-8}$ | |

Table 3: Comparison of three 95% confidence interval resulted from Bayesian procedures for the data set III

| Cell | Bayesian-1 | | Bayesian-2 | | Bayesian-3 | |
|---|---|---|---|---|---|---|
| | LL | UL | LL | UL | LL | UL |
| 1 | 0.093 | 0.151 | 0.092 | 0.151 | 0.093 | 0.1505 |
| 2 | 0.123 | 0.188 | 0.123 | 0.188 | 0.1237 | 0.1874 |
| 3 | 0.125 | 0.191 | 0.124 | 0.191 | 0.1247 | 0.1902 |
| 4 | 0.099 | 0.159 | 0.090 | 0.158 | 0.0984 | 0.1581 |
| 5 | 0.104 | 0.164 | 0.109 | 0.166 | 0.1038 | 0.1648 |
| 6 | 0.152 | 0.223 | 0.153 | 0.223 | 0.1518 | 0.2208 |
| 7 | 0.097 | 0.156 | 0.096 | 0.156 | 0.0971 | 0.1556 |
| Volume | $3.76 \times 10^{-9}$ | | $3.96 \times 10^{-9}$ | | $3.53 \times 10^{-9}$ | |