Biostatistics Department Technical Report

BST2012-001

Bartlett's Adjustment for the Likelihood Ratio Test In Pool Screening

Charles R. Katholi, Ph.D.

April 2012

Department of Biostatistics
School of Public Health
University of Alabama at Birmingham
ckatholi@uab.edu

Bartlett's Adjustment for the Likelihood Ratio Test in Pool Screening

Charles R. Katholi, Ph.D.

April 4, 2012

We consider the likelihood ratio test for the simple hypothesis, $H_0 : p = p_0$ versus the alternative $H_a : p \neq p_0$ in the case of group testing. Assume that the group size is n and that m groups have been collected and tested. Let X_j the the Bernoulli random variable which is 0 when the group tests negative and 1 when it tests positive. Let $T = \sum_{j=1}^{m} X_j$ then T has the probability mass function,

$$P(T = t|n, m, p_0) = \binom{m}{t}[1 - (1 - p_0)^n]^t[(1 - p_0)^n]^{m-t} \tag{1}$$

The maximum Likelihood Estimator of p is well known to be,

$$\hat{p} = 1 - (1 - t/m)^{1/n} \tag{2}$$

and so the likelihood ratio statistic for the test is,

$$\lambda = \left[\frac{1 - (1 - \hat{p})^n}{1 - (1 - p_0)^n}\right]^t \left[\frac{(1 - \hat{p})^n}{(1 - p_0)^n}\right]^{m-t} \tag{3}$$

Finally, $W = -2\ln(\lambda)$ is known to be asymptotically distributed $(m \to \infty)$ as $\chi^2(1)$ in this case. When the number of groups tested, m, is small, Bartlett [1] suggested a way to improve the approximation by replacing W by $\widetilde{W} = (1 + \frac{b_1}{m})^{-1}W$, where b_1 is such that $E_{H_0}(\widetilde{W}) = 1 + \bigcirc(m^{-2})$ while $E_{H_0}(W) = 1 + \bigcirc(m^{-1})$. Barndorf-Nielsen and Cox [2] and Barndorf-Nielsen and Hall [3] among others enlarged on Bartlett's work. In making this adjustment Bartlett found a statistic whose mean was closer to the expectation of the χ^2 and in so doing hoped to improve the approximation.

1 Derivation of b_1

From equation 3 , the definition of W ,the known form of $\hat{p}$ from equation 2 and some algebra yields,

$$W = w(T) = 2Tln(T/m) + 2(m-T)ln(1-T/m) - 2Tln[1-(1-p_0)^n] \\ - 2(m-T)ln[(1-p_0)^n] \tag{4}$$

To find the necessary value of b_1, $w(T)$ will be expanded in a Taylor series in powers of $(T - E_{H_0}(T))$, $E_{H_0}(T) = m(1-(1-p_0)^n)$ and termwise expectation will be taken. The result will be an expansion for $E_{H_0}(w)$ such that,

$$E_{H_0}(w) = 1 + \frac{b_1}{m} + \frac{b_2}{m^2} + \bigcirc(m^{-3})$$

Let $g_k = \left[\frac{d^k w(T)}{dT^k}\right]_{T=m(1-(1-p_0)^n)}$ and let $t_0 = m(1-(1-p_0)^n)$ then

$$w(T) = 0 + 0(T-t_0) + \frac{1}{2}g_2(T-t_0)^2 + \frac{1}{6}g_3(T-t_0)^3 + \frac{1}{24}g_4(T-t_0)^4 + \cdots \tag{5}$$

since $g_0 = g_1 = 0$ (see Appendix A). It is easy to show at this point that $w(T)$ converges in distribution to a $\chi^2(1)$ random variable as $m \to \infty$. Note that the quadratic term in this expansion is equal to

$$\begin{aligned} \frac{1}{2}g_2(T-t_0)^2 &= \frac{1}{2}\frac{2}{m[1-(1-p_0)^n][(1-p_0)^n]}(T - m[1-(1-p_0)^n])^2 \\ &= \left(\frac{m(T/m - [1-(1-p_0)^n]}{\sqrt{m[1-(1-p_0)^n][(1-p_0)^n]}}\right)^2 \\ &= \left(\frac{(T/m - [1-(1-p_0)^n]}{\sqrt{\frac{[1-(1-p_0)^n][(1-p_0)^n]}{m}}}\right)^2 \end{aligned} \tag{6}$$

Since $T = \sum_{i=1}^{m} X_i$ and since the X_i are i.i.d random variables such that $E_{H_0}(X_i) = 1 - (1-p_0)^n$ and $Var(X_i) = [1-(1-p_0)^n][(1-p_0)^n] < \infty$ it follows from the Central Limit Theorem that

$$\frac{\sqrt{m}(T/m - [1-(1-p_0)^n]}{\sqrt{[1-(1-p_0)^n][(1-p_0)^n]}} \xrightarrow{D} Z \sim N(0,1) \text{ as } m \to \infty$$

and hence that the first term in the expansion for $w(T)$ converges in distribution to a $\chi^2(1)$ random variable. Next consider the general term for $k \geq 3$. From appendix A it follows that

$$\begin{aligned} \frac{g_k}{k!}(T-t_0)^k &= \frac{2[(1-(1-p_0)^n)^{k-1} + (-1)^k(1-p_0)^{(k-1)n}]}{k(k-1)m^{k-1}[(1-(1-p_0)^n)(1-p_0)^n]^{k-1}}(T-t_0)^k \\ &= \frac{1}{m^{k/2-1}} H(k,n,p_0) \left(\frac{\sqrt{m}[T/m - (1-(1-p_0)^n)]}{\sqrt{[1-(1-p_0)^n][(1-p_0)^n]}} \right)^k \end{aligned} \tag{7}$$

where $H(k,n,p_0) < \infty$ is a complex function independent of m. As $m \to \infty$ the quantity $(*)^k$ converges in distribution to the $k-th$ power of a standard normal random variable.The quantity $m^{-(k/2-1)} \to 0$ as $m \to \infty$ when $k \geq 3$, so by Slutsky's Theorem all the terms in the series(after the quadratic term) converge in distribution to a degenerate distribution with point mass at 0 as m tends to infinity. It is known [6] the convergence in distribution to a degenerate distribution implies convergence in probability to the same constant and so a second application of Slutsky's Theorem shows that the limiting distribution of $w(T)$ is a $\chi^2(1)$. As an aside we note as well that the first term of those which converge to zero in probability is $\bigcirc_p\left(\frac{1}{\sqrt{m}}\right)$.

Returning to the derivation of b_1, set $\tilde{g}_k = m^{k-1}g_k$ and take the expectation termwise in equation 5 to obtain,,

$$E_{H_0}(w(T)) = \frac{\tilde{g}_2\mu_2}{2m} + \frac{\tilde{g}_3\mu_3}{6m^2} + \frac{\tilde{g}_4\mu_4}{24m^3} \cdots + \frac{\tilde{g}_k\mu_k}{k!m^{k-1}} + \cdots \tag{8}$$

Next it is shown in Appendix B that the moments, μ_k can be expressed as polynomials in m with coefficients which are functions of $w = 1-(1-p_0)^n$. For example $\mu_2 = c_{21}m$, $\mu_3 = c_{31}m$, $\mu_4 = c_{41}m + c_{42}m^2$ and so forth. With this notation, equation 8 can be written as

$$E_{H_0}(w(T)) = \frac{\tilde{g}_2 c_{21} m}{2m} + \frac{\tilde{g}_3 c_{31} m}{6m^2} + \frac{\tilde{g}_4 (c_{41} m + c_{42} m^2)}{24 m^3} + \frac{\tilde{g}_5 (c_{51} m + c_{52} m^2)}{120 m^4}$$
$$+ \frac{\tilde{g}_6 (c_{61} m + c_{62} m^2 + c_{63} m^3)}{720 m^5}$$
$$+ \frac{\tilde{g}_7 (c71 m + c_{72} m^2 + c_{73} m^3)}{5040 m^6} \tag{9}$$
$$+ \frac{\tilde{g}_8 (c_{81} m + c_{82} m^2 + c_{83} m^3 + c_{84} m^4)}{40320 m^7}$$
$$+ \frac{\tilde{g}_9 (c_{91} m + c_{92} m^2 + c_{93} m^3 + c_{94} m^4)}{8! m^8} + \cdots$$

Collecting terms in powers of $1/m$ yields,

$$E_{H_0}(w(T)) = \frac{1}{2}\tilde{g}_2 c_{21} + \frac{1}{m}\left[\frac{\tilde{g}_3 c_{31}}{6} + \frac{\tilde{g}_4 c_{42}}{24}\right]$$
$$+ \frac{1}{m^2}\left[\frac{\tilde{g}_4 c_{41}}{24} + \frac{\tilde{g}_5 c_{52}}{120} + \frac{\tilde{g}_6 c_{63}}{720}\right] \tag{10}$$
$$+ \frac{1}{m^3}\left[\frac{\tilde{g}_5 c_{51}}{120} + \frac{\tilde{g}_6 c_{62}}{720} + \frac{\tilde{g}_7 c_{73}}{5040} + \frac{\tilde{g}_8 c_{84}}{40320}\right] + \bigcirc\left(\frac{1}{m^4}\right)$$

Letting $\pi = 1 - (1 - p_0)^n$, and using the moment formulas from appendix B, it is found that

$$\frac{\tilde{g}_2 \mu_2}{2} = \frac{2}{(1 - p_0)^n (1 - (1 - p_0)^n)} \frac{(1 - p_0)^n (1 - (1 - p_0)^n)}{2} = 1 \tag{11}$$

$$b_1 = \frac{\tilde{g}_3 c_{31}}{6} + \frac{\tilde{g}_4 c_{42}}{24} = \frac{[1 - (1 - p_0)^n + (1 - p_0)^{2n}]}{6[1 - (1 - p_0)^n][(1 - p_0)^n]} \tag{12}$$

$$b_2 = \frac{1}{m^2}\left[\frac{\tilde{g}_4 c_{41}}{24} + \frac{\tilde{g}_5 c_{52}}{120} + \frac{\tilde{g}_6 c_{63}}{720}\right] = \frac{[1 - 3(1 - p_0)^n + 3(1 - p_0)^{2n}]}{6[1 - (1 - p_0)^n]^2[(1 - p_0)^n]^2} \tag{13}$$

. It easily follows from equations 10 through 12 that $E_{H_0}([1 + b_1/m]^{-1} w(T))$ has the form $1 + \bigcirc\left(m^{-2}\right)$.

2 Numerical examples

The purpose of the Bartlett adjustment is to improve the χ^2 approximation to the distribution of $-2ln(\lambda)$, where λ is the likelihood ratio statistic. In examining the effectiveness of this adjustment, we recall that the $\chi^2(\nu)$ distribution is a special case of the general *Gamma*(α,β) distribution with $\alpha = \nu/2$ and $\beta = 2$. Recall as well, that for any random variable, multiplication by a constant changes the scale parameter. Thus, it would seem that for a random variable which is approximately a gamma random variable, that examination of the gamma parameters should shed some light on what the approximation accomplishes. Since for $Y \sim$ gamma(α, β), $E(Y) = \alpha\beta$ and $Var(Y) = \alpha\beta^2$, given the mean and variance, α and β can be easily calculated. For the simple hypothesis being considered here, it is possible to calculate the mean and variance of $-2ln(\lambda)$ exactly given values for the parameters m, n and p_0 all of which are known. From equation 3 in section 1, we note that $w(T)$ is a function of the sufficient statistic $T = \sum_{i=1}^{m} X_i$ which has a $Binomial(\theta, m)$ distribution where $\theta = 1 - (1 - p_0)^n$. Hence

$$E_{H_0}(w(T)) = \sum_{t=0}^{m} w(t) \binom{m}{t} [1 - (1 - p_0)^n]^t [(1 - p_0)^n]^{m-t} \tag{14}$$

In making these calculations it is important to realize that care must be taken in calculating $w(T)$ when $T = 0$ and $T = m$. In particular, it can be shown by elementary calculus that

$$\lim_{T \to 0} w(T) = -2m \ln[1 - (1 - p_0)^n] \tag{15}$$

and

$$\lim_{T \to m} w(T) = -2mn \ln(1 - p_0) \tag{16}$$

The $E_{H_0}(w(T)^2)$ can be calculated using the same approach and $Var(w(T))$ is found by the standard formula. In general, pool screening (group testing) is used when p_0 is small and so we will give calculated results for a number of values of p_0 , pool size n and number of pools m. Note that one must be careful to choose the pool sizes for a given p_0 so that it is improbable that all the pools are positive. Similarly, a large enough pool size and a sufficient number of pools is required to avoid the situation where all the pools are negative; that is n and m must be chosen so that the probability that any pool is positive is much less than one. Chiang and Reeve [4] suggest less

m	n	p_0	$*$	$E_{H_0}(*)$	α	β	$1+b_1/m$
50	25	0.01	w	1.01724	0.49858	2.04030	1.01596
			$\widetilde{w}$	1.00127	0.49858	2.00826	
100	25	0.01	w	1.00827	0.49972	2.01767	1.00798
			$\widetilde{w}$	1.00029	0.49972	2.00170	
200	25	0.01	w	1.00406	0.49994	2.00837	1.00399
			$\widetilde{w}$	1.00007	0.49994	2.00040	
50	50	0.005	w	1.01728	0.49857	2.0401	1.01599
			$\widetilde{w}$	1.00127	0.49857	2.00831	
100	50	0.005	w	1.00829	0.49972	2.01771	1.00799
			$\widetilde{w}$	1.00029	0.49972	2.00171	
200	50	0.005	w	1.00407	0.49994	2.00839	1.00400
			$\widetilde{w}$	1.00007	0.49994	2.00040	

Table 1: Effect of the Bartlett adjustment on the scale of the distribution

than 0.5. This is not a problem when $p_0 < 0.01$ but can become a problem when $p_0 >> 0.01$.

From Table 1 it is clear that in the cases presented $E_{H_0}(\widetilde{W})$ is much closer to the value 1 which would be correct for the $\chi^2(1)$ distribution, than $E_{H_0}(W)$. The coefficients of the gamma distribution calculated as moment estimates are also closer to the correct values for the $\chi^2(1)$. Because the correction factor in all cases is greater than one, it is clear that $\widetilde{W}$ is always less than W. It is unclear, however, whether this adjustment is really valuable in the case of this statistic. The actual distribution of W and hence of $\widetilde{W}$ is discrete since these are functions of the random variable T which is discrete and takes on values in the set $\Lambda = \{0, 1, 2, \cdots, m\}$. Furthermore, the probability mass function of T is given in equation 1 and is completely specified since m, n and p_0 are known. From this it follows obviously that $Pr(W = w(t)) = Pr(T = t|m, n, p_0)$. All values of W are easily calculated, even when m is large, so $Pr(w > w_C)$ can be calculated exactly. Thus asymptotic results are not required although they require much less computation. The question then becomes, *how well does the use of the adjusted asymptotic statistic, $\widetilde{W}$, or of W itself do with respect to the level of the test compared to the test based on the exact discrete distribution?*. Table 2 presents calculations intended to shed light on this question.

m	n	p_0	Desired α level	α-level of Bartlett Adjusted Asymptotic Test	Exact Test α level
25	25	0.01	0.10	0.09568	0.09568
			0.05	0.04990	0.04989
			0.01	0.00581	0.00581
50	25	0.01	0.10	0.08953	0.08953
			0.05	0.03999	0.04000
			0.01	0.01090	0.00554
100	25	0.01	0.10	0.08182	0.09182
			0.05	0.05349	0.04252
			0.01	0.01142	0.00763
200	25	0.01	0.10	0.10733	0.08837
			0.05	0.05069	0.04071
			0.01	0.00814	0.00814
25	50	0.005	0.10	0.09586	0.09587
			0.05	0.04961	0.04961
			0.01	0.00576	0.00576
50	50	0.005	0.10	0.08965	0.08965
			0.05	0.04000	0.04000
			0.01	0.01096	0.00540
100	50	0.005	0.10	0.09186	0.09186
			0.05	0.05359	0.04276
			0.01	0.01147	0.00756
200	50	0.005	0.10	0.08822	0.08823
			0.05	0.05086	0.04055
			0.01	0.00814	0.00827

Table 2: Exact α level for Bartlett Adjusted statistic versus α-level for the exact test

In Table 2 the α-level for the likelihood ratio test based on the Bartlett Adjusted test statistic $\widetilde{W}$ is presented based on direct calculations. The decision function $\psi(\widetilde{W})$ is introduced where

$$\psi(w) = \begin{cases} 1 & \text{if } w \in S \\ 0 & \text{if } w \notin S \end{cases} \tag{17}$$

and $S = \{w \,|\, w \geq \chi^2_{1-\alpha}(1)\}$. Then the level of the test is given by $\alpha = E_{H_0}(\psi(\widetilde{W}))$. For comparison purposes, the α-levels for the exact test are also given in the last column. Since these exact tests take into account the discreteness of $\widetilde{W}$ they are necessarily more conservative than desired. For the most part, the asymptotic test seems to follow the exact test pretty well.

It is instructive to consider the decision function for a randomized test as described by Lehmann [5]. The decision function ϕ is defined as

$$\phi(w) = \begin{cases} 1 & \text{if } w \in S \\ \gamma & \text{if } w = w_C \\ 0 & \text{if } w \neq w_C \text{ and } w \notin S \end{cases} \tag{18}$$

where w_C is such that $Pr(W > w_C \,|\, p = p_0) < \alpha$ and $Pr(W < w_C \,|\, p = p_0) > \alpha$, S is the set $S = \{w \,|\, w > w_C\}$ and γ is a constant such that $E_{H_0}(\phi) = Pr(W \in S) + \gamma Pr(W = w_C) = \alpha$. The constant γ is always such that $0 \leq \gamma \leq 1$. In performing a randomized test, the null hypothesis is rejected when $\phi(w) = 1$ and fails to reject when $\phi(w) = 0$. If $W = w_C$ a Bernoulli trial with success probability equal to γ is performed and if the trial is a success, then the null hypothesis is rejected. The randomized test is always of size α and the size of γ is suggestive about how often w_C is included in the rejection set; that is, how close is $Pr(W \geq w_C)$ to the desired test size α. Consider for example the case $m = 50, n = 25$ and $p_0 = 0.01$ for which $\gamma = 0.83192$. Table 3 presents the upper tail of the distribution function and probability mass function for $\widetilde{W}$ in this case. The critical

$\widetilde{w}$	$Pr(\widetilde{W} \geq \widetilde{w})$	$Pr(\widetilde{W} = \widetilde{w})$
6.23453	0.015772	0.0048705
7.15880	0.010901	0.0053651
7.82050	0.005537	0.0021564
9.57057	0.003380	0.0008799

Table 3: Upper tail of the distribution of $\widetilde{W}$ near the critical value, $\widetilde{w}_C$, when $m = 50$, $n = 25$, $p_0 = 0.01$ and $\alpha = 0.01$

value is $\widetilde{w}_C = 7.15880$. Note that $Pr(\widetilde{W} \geq 7.15880) = 0.010901$ which is very close to the desired α-level for the test. Since $\gamma = 0.83192$ the randomized test will reject 83 % of the time when w_C is observed. The critical value for the Chi Square with one degree of freedom and $\alpha = 0.01$ is 6.63489 which is less than w_C but larger than the next smallest possible value of $\widetilde{w}$ and so the asymptotic test includes w_C in the critical region.

$\widetilde{w}$	$Pr(\widetilde{W} \geq \widetilde{w})$	$Pr(\widetilde{W} = \widetilde{w})$
6.49355	0.013407	0.0023601
6.57440	0.011047	0.0029046
7.32540	0.008143	0.0015472
7.57317	0.006596	0.0017840

Table 4: Upper tail of the distribution of $\widetilde{W}$ near the critical value, $\widetilde{w}_C$, when $m = 200$, $n = 25$, $p_0 = 0.01$ and $\alpha = 0.01$

Similarly, when $m = 200, n = 25, p_0 = 0.01$ and $\alpha = 0.01$, Table 4 shows the details in the neighborhood of the critical value $w_C = 6.57440$ where again the critical value of the Chi Square is 6.63489. In this case, $\gamma = 0.63931$ which means that the Bernoulli trail part of the randomized test would reject 63 % of the time. In this case, the $\chi^2(1)$ critical value is larger than w_C so that the level of the asymptotic test is more like the level of the exact test for the discrete random variable $\widetilde{W}$.

3 Summary Conclusions

The performance of the statistic $W(T) = -2\ln(\lambda(t))$ after applying a Bartlett Adjustment in small to medium size samples has been considered. The level of the asymptotic has been compared to the level for a test based on the exact distribution of W and it has been found that, in general, the adjusted statistic $\widetilde{W}$ produces tests with significance levels comparable to what is found using an exact test, but far less computational effort. Most importantly, the size of the test never is too far from the desired level. In a few cases, the actual size is marginally larger that the desired size, but more often the level of the test approximates the level of the exact test.

References

[1] Bartlett, M.S.(1937)"Properties of sufficiency and statistical tests",*Proc. R. Soc. A*, **160**,pp 268-82.

[2] Barndorf-Nielsen, O.E. and Cox, D.R. (1984) "Bartlett Adjustments to the Likelihood Ratio Statistic and the Distribution of the Maximum Likelihood Estimator", *J.R.Statist. Soc. B*, **46**,No.3, pp 483-495.

[3] Barndorf-Nielsen, O.E. and Hall, P.,(1988) "On the level-error after Bartlett adjustment of the likelihood ratio statistic", *Biometrika*, **75**, No. 2, pp 374-378.

[4] Chiang, C. and Reeve, W. "Statistical Estimation of Virus Infection Rates in Mosquito Vector Populations", *American Journal of Hygiene*, **75**, pp 377-391, 1962.

[5] Lehmann, E.L. and Romano, J.P. (2010) *Testing Statistical Hypotheses, 3rd Edition*, Springer, New York

[6] Sen, P.K. and Singer, J.M. (2000) *Large Sample Methods in Statistics: An Introduction with Applications*, 1st CRC Edition, Chapman and Hall/CRC, New York.

Appendices

A Derivatives of $w(T)$ at $T = E_{H_0}(T)$

The derivatives of $w(T)$ with respect to T , evaluated at $E_{H_0}(T)$ are easily found using algebra software such as Maple. This leads to the coefficient values used in the expansion of section 1. In what follows, $\left[\frac{d^k w(T)}{dT^k}\right]_{T=1-(1-p_0)^n}$ will be denoted by $g_k, k = 0, 1, \cdots$.

$$g_0 = g_1 = 0 \tag{19}$$

$$g_2 = \frac{2}{m(1-p_0)^n(1-(1-p_0)^n)} \tag{20}$$

$$g_3 = \frac{2[1-2(1-p_0)^n]}{m^2[1-(1-p_0)^n]^2[(1-p_0)^n]^2} \tag{21}$$

$$g_4 = \frac{4[(1-(1-p_0)^n)^3 + (1-p_0)^{3n}]}{m^3[1-(1-p_0)^n]^3[(1-p_0)^n]^3} \tag{22}$$

$$g_5 = \frac{12[(1-(1-p_0)^n)^4 - (1-p_0)^{4n}]}{m^4[1-(1-p_0)^n]^4[(1-p_0)^n]^4} \tag{23}$$

$$g_6 = \frac{48[(1-(1-p_0)^n)^5 + (1-p_0)^{5n}]}{m^5[1-(1-p_0)^n]^5[(1-p_0)^n]^5} \tag{24}$$

In general, g_k for $k \geq 3$ is given by,

$$g_k = \frac{2(k-2)![(1-(1-p_0)^n)^{k-1} + (-1)^k(1-p_0)^{(k-1)n}]}{m^{k-1}[1-(1-p_0)^n]^{k-1}[(1-p_0)^n]^{k-1}} \tag{25}$$

B Higher moments about the mean of the Binomial Distribution

The moments about the mean of the Binomial distribution are needed to develop the expansion used in finding the expected value of $w(T)$ in section 1. To ease that derivation, the moments will be expressed as polynomials in m with coefficients which are functions of π. For a Binomial distribution, *Bin(π,m)* we know that the mean is $m\pi$. The higher moments are then

$$\mu_2 = c_{21}m = \pi(1-\pi)m \tag{26}$$

$$\mu_3 = c_{31}m = \pi(1-\pi)(1-2\pi)m \tag{27}$$

$$\begin{aligned} \mu_4 &= c_{41}m + c_{42}m^2 \\ &= [\pi(1-\pi)(6\pi^2 - 6\pi + 1)]m + [3\pi^2(1-\pi)^2]m^2 \end{aligned} \tag{28}$$

$$\mu_5 = c_{51}m + c_{52}m^2 \tag{29}$$

where

$$\begin{aligned} c_{51} &= \pi(1-\pi)(1-2\pi)(12\pi^2 - 12\pi + 1) \\ c_{52} &= 10\pi^2(1-\pi)^2(1-2\pi) \end{aligned} \tag{30}$$

Similarly,

$$\mu_6 = c_{61}m + c_{62}m^2 + c_{63}m^3 \tag{31}$$

where

$$\begin{aligned} c_{61} &= \pi(1-\pi)(120\pi^4 - 230\pi^3 + 150\pi^2 - 30\pi + 1) \\ c_{62} &= 5\pi^2(1-\pi)^2(26\pi^2 - 26\pi + 5) \\ c_{63} &= 15\pi^3(1-\pi)^3 \end{aligned} \tag{32}$$

$$\mu_7 = c_{71}m + c_{72}m * 2 + c_{73}m^3 \tag{33}$$

where

$$\begin{aligned} c_{71} &= \pi(1-\pi)(1-2\pi)(360\pi^4 - 720\pi^3 + 420\pi^2 - 60\pi + 1) \\ c_{72} &= 14\pi^2(1-\pi)^2(1-2\pi)(33\pi^2 - 33\pi + 4) \\ c_{73} &= 105\pi^3(1-\pi)^3(1-2\pi) \end{aligned} \tag{34}$$

and

$$\mu_8 = c_{81}m + c_{82}m^2 + c_{83}m^3 + c_{84}m^4 \tag{35}$$

where

$$\begin{aligned} c_{81} &= \pi(1-\pi)(5040\pi^6 - 15120\pi^5 + 16800\pi^4 - 8400\pi^3 + 1806\pi^2 - 126\pi + 1) \\ c_{82} &= 7\pi^2(1-\pi)^2(1044\pi^4 - 2088\pi^3 + 1352\pi^2 - 308\pi + 17) \\ c_{83} &= 70\pi^3(1-\pi)^3(34\pi^2 - 34\pi + 7) \\ c_{84} &= 105\pi^4(1-\pi)^4 \end{aligned} \tag{36}$$

C The Exact Distribution of the Likelihood Ratio Statistic

For the test of the simple hypothesis considered in this report, it is possible to find the exact distribution of the Likelihood Ratio Statistic or any of its transformed versions (e.g. the Bartlett Corrected Statistic). We will confine the discussion to the function $w(T)$ defined in section 1 as,

$$\begin{aligned} W = w(T) = 2Tln(T/m) + 2(m-T)ln(1-T/m) - 2Tln[1-(1-p_0)^n] \\ - 2(m-T)ln[(1-p_0)^n] \end{aligned} \tag{37}$$

T is a Binomial random variable with parameters $1-(1-p_0)^n$ and m and consequently takes values in the set $A_T = \{0, 1, \cdots, m\}$. The function, $w(T)$ is a mapping from the set A_T onto a set $B \subset \Re$ such that

$$B = \{\tilde{w} | \tilde{w} = w(T) \text{ for some } T \in A_T\}$$

The set B is discrete and contains no more than $m+1$ elements. For each $\tilde{w} \in B$ there exists a set of values of T, denoted by $S_{\tilde{w}} \subset A_T$ such that every $T \in S_{\tilde{w}}$ maps into the same value $\tilde{w}$ under the transformation $w(T)$; that is, $S_{\tilde{w}}$ is the *pre-image* of $\{\tilde{w}\}$ under the mapping. Then from elementary probability theory, $P(w(T) = \tilde{w}) = P(T \in S_{\tilde{w}})$. Since the random variable T is restricted to a set on integers, it is expected that most of the time the pre-image set for any $\tilde{w}$ will contain a single element and so $W(T)$ will be 1-1. Extensive computational work supports this conjecture. Based on this relationship, the distribution function of $\tilde{w}$ is easily calculated.