Re-Testing in Batch Testing Model Based on Quality Control Process for Proportion Estimation

Ronald Waliaula Wanyonyi[1], Olivia Wanjeri Mwangi[2], Charles Wambugu Mwangi[3]

[1]Department of Mathematics, Egerton University, Njoro, Kenya
[2]Data Science Institute, University of Delaware, Newark, DE, USA
[3]Department of Mathematics and Actuarial, Kabarak University, Kabarak, Kenya
Email: ronald.wanyonyi@egerton.ac.ke, wanjerim@udel.edu, cwmwangi@kabarak.ac.ke

Abstract

The quality of products manufactured or procured by organizations is an important aspect of their survival in the global market. The quality control processes put in place by organizations can be resource-intensive but substantial savings can be realized by using acceptance sampling in conjunction with batch testing. This paper considers the batch testing model based on the quality control process where batches that test positive are re-tested. The results show that re-testing greatly improves the efficiency over one stage batch testing based on quality control. This is observed using Asymptotic Relative Efficiency (ARE), where for values of p computed ARE > 1 implying that our estimator has a smaller variance than the one-stage batch testing. Also, it was found that the model is more efficient than the classical two-stage batch testing for relatively high values of proportion.

1. Introduction

Batch testing, also known as group testing or pool screening, is a method that combines individual items into several batches of a certain size [1]. Then the batches are tested instead of individual items. If a batch tests positive, it is assumed that at least one item in the batch is positive; otherwise it is assumed that all items in the batch are negative. Batch testing provides cost-effective method of screening a population for a given trait when the proportion of the trait in the

population is low and the tests involved are expensive and sensitive.

Batch testing, which has two main objectives; classification and estimation, has had early major contributions from [2] and [3]. It first appeared in statistics literature in the context of screening blood samples obtained from military inductees for the presence of a disease [2]. However, this method is currently applied in different fields such as safety of blood products [4], disease screening in humans [5], animals [6] or plants [7], genetics [8] and purity of seeds [9].

In implementing quality control procedures, a sample of items is taken from the produced or procured lot and pooled in batches. Each batch is tested with the appropriate test and rejected if the number of items with the trait in the batch exceeds a predetermined cut off value, *l*. Otherwise, the batch is accepted [10]. Ideally, the tests are normally assumed to be perfect but occasionally these tests are prone to misclassification errors due to dilution [11], blockers [5] and [5] or non-representative sample [12]. But misclassification is usually kept at the minimal level. Depending on the cost and benefits, an organization concerned with the production or procurement of the items may decide to re-test batches that initially test positive. Re-testing enables organization to reduce misclassification and also facilitates the investigation of dilution effect [13] and [14]. Therefore, it is in this background that we develop this article and the article is arranged as follows; Section 2 gives the literature review, that is, outlines the estimator of a proportion obtained using batch testing in quality control with perfect and imperfect tests. Section 3 illustrates the proposed batch testing model in quality control with re-testing and in Section 4; the proposed model is compared with the models proposed by [15] and [16]. Lastly, Section 5 gives brief discussion and conclusion.

2. Literature Review

The quality of products procured from or released to the market is a major concern to most if not all organizations. Thus, many of the organizations set up quality control processes to ensure quality standards of their products are met. However, quality control processes are sometimes cost-intensive, destructive and/or may take much time to get the results. To mitigate against these issues acceptance sampling is usually employed where a lot of items are either accepting or rejecting without inspecting each and every item in it. A lot is classified as positive if the number of its items with a trait of interest is greater than a cut-off value otherwise it is marked negative [10]. Further savings on testing resources can be realized if the items are tested in a batch instead of individual testing [17].

Application of batch testing in quality control is similar to threshold batch testing that was first introduced by [18], where he considered two cut-off values; lower and upper, and derived an algorithm that identifies the defective items with minimum number of tests. Other authors also studied the threshold batching testing with the view of improving on the identification algorithm and among them are [19].

Apart from identification of defectives, batch testing is also concerned with estimation of proportion of a trait in a population [20]. This was extended to quality control process by [15], in which they considered batch testing model with only the lower cut-off value with perfect tests. The maximum likelihood estimator of the proportion was derived and its properties investigated. It was found out that by introducing the cut-off value, the estimator of proportion became more efficient than existing batch testing for large values of the proportion.

An aspect of concern when using batch testing is misclassifications of items due to blockers [20] or dilution effect [5]. But it is [5] who first study estimation of a proportion by batch testing with imperfect tests. This concept of imperfect tests was later applied to batch testing in quality control process by [21].

Much of the recent studies have focused on determining more efficient estimator and optimal group sizes when the tests are not 100% perfect. For instance, [22] developed multistage adaptive batching testing model and [23], proposed a combination of experiments. These batch testing strategies produced estimators of the proportion that performed better than existing models in the presence test errors. [24] suggested a batch testing procedure that incorporates re-testing which is noted to recover lost sensitivity or specificity.

This paper considers batch testing model in quality control processes with re-testing and compares it to one stage batch testing in quality control and the usual two-stage batch testing.

2.1. Estimation of a Proportion Using Batch Testing Model Based on Quality Control

Consider a finite population with N items and each item can be classified as good or bad depending on the presence or absence of a trait and let p, be the unknown proportion of items with the trait in the population. Suppose that the population can be divided into n batches each of size k. A batch is rejected if the number of items with the trait in the batch is greater than a predefined threshold or cutoff value l. The probability of rejecting a batch is

$$\pi(p) = 1 - F(l) \tag{1}$$

where

$$F(l) = \sum_{d=0}^{l} \binom{k}{d} p^d (1-p)^{k-d}$$

Suppose X out of the n batches test positive on the test. Here X is a random variable. According to [2], X follows a binomial (n, $\pi(p)$). The Maximum Likelihood Estimator of p is obtained by solving Equation (2);

$$\sum_{d=0}^{l} \binom{k}{d} p^d (1-p)^{k-d} = 1 - \frac{x}{n} \tag{2}$$

Equation (2) has no solution in closed form except when $l = 0$, which leads to the results obtained by [3] among others as

$$\hat{p} = 1 - \left(1 - \frac{x}{n}\right)^{\frac{1}{k}} \tag{3}$$

Hence, the proposed estimator generalizes [3] results. When $l > 0$, Equation (2) can be solved iteratively. The maximum likelihood estimator of p is found to be positively biased for p and that the bias is negligible for small values of p but can be very high for large values of p [15].

The asymptotic variance is obtained from the Fisher's information given by;

$$\left\{ -E\left(\frac{\partial^2 \log L(.)}{\partial p^2} \right) \right\}^{-1} \tag{4}$$

which gives;

$$\text{var}(\hat{p}) = \frac{\pi(p)\left[1 - \pi(p)\right]p^2(1-p)^2}{n\left[E - pkF(l)\right]^2} \tag{5}$$

where

$$E = \sum_{d=0}^{l} d\binom{k}{d} p^d (1-p)^{k-d}$$

Notice that if the cut off value $l = 0$, the variance of the estimator becomes;

$$\text{var}(\hat{p}_T) = \frac{1 - (1 - p_T)^k}{nk^2(1 - p_T)^{k-2}} \tag{6}$$

as obtained by [3] and other authors. Thus, the maximum likelihood estimator of p for batch testing based on quality control is a generalization of Thompson's original estimator in the [2] model.

Utilizing the delta method $\hat{p}$ is asymptotically normally distributed [25]. That is, for fixed k and l and $n \to \infty$ we have,

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} Normal\left(0, \frac{\pi(p)\left[1 - \pi(p)\right]p^2(1-p)^2}{\left[E - pkF(l)\right]^2}\right) \tag{7}$$

2.2. Estimation of a Proportion Using Batch Testing Model Based on Quality Control with Imperfect Tests

The batch testing model based on quality control procedure with imperfect tests is applied. The probability that a batch is classified as positive is;

$$\pi_0(p) = \eta(1 - F(l)) + (1 - \phi)F(l) \tag{8}$$

where are the sensitivity and specificity of the tests assumed to be constant in the course of testing. The sensitivity of a test means the probability of correctly detecting a positive batch while specificity is the probability of correctly identifying a negative batch. If n batches of size k are tested and X_0 batches test positive, then the likelihood function is;

$$L(p/n, x_0, \eta, \phi) = \binom{n}{x_0}\left[\pi_1(p)\right]^{x_0}\left[1 - \pi_1(p)\right]^{n-x_0} \tag{9}$$

Using maximum likelihood method estimation method, the estimator of p is obtained by solving;

$$\sum_{d=0}^{l} \binom{k}{d} p^d (1-p)^{k-d} = \frac{\eta - \dfrac{x}{n}}{\eta + \phi - 1} \tag{10}$$

For $l = 0$, Equation (10) reduces to

$$\hat{p}_0 = 1 - \left\{ \frac{\eta - \dfrac{x}{n}}{\eta + \phi - 1} \right\}^{\frac{1}{k}}$$

This is a result obtained by [16]. However, Equation (10) has no solution in closed form when $l > 0$. Therefore, the equation can be solved iteratively using an R or MATLAB code that is easily developed.

The asymptotic variance is obtained from the Fisher's information given by Equation (4) to get;

$$\text{var}\left(\hat{p}_0\right) = \frac{\pi_0(p)\left(1 - \pi_0(p)\right) p^2 (1-p)^2}{n\left[\{\eta + \phi - 1\}\{E - pkF(l)\}\right]^2} \tag{11}$$

Notice that if the cut off value $l = 0$ and $\eta = \phi = 1$ then the variance of the estimator reduces to Equation (6). Thus the estimator obtained in this section generalizes both the estimator obtained in Section (2) and Thompson's estimator.

3. Estimation of a Proportion in the Proposed Model

In our proposed testing model, the finite population with N items is pooled into n batches each of size k. The batches are then tested and a batch that contains more than l items with a trait of interest is classified as positive; otherwise it is negative. Further, batches that test positive are given a re-test and a batch that has more than l items with a trait is labeled positive and testing is stopped. The probability of declaring a batch negative on initial test;

$$\pi_1(p) = (1-\eta)(1-F(l)) + \phi F(l) \tag{12}$$

The probability of declaring a batch negative on re-test after initially testing positive;

$$\pi_2(p) = \eta(1-\eta)(1-F(l)) + \phi(1-\phi)F(l) \tag{13}$$

Suppose that X_1 and X_2 are batches that test negative initially and test negative on re-test out of n batches respectively, then the likelihood function for the proposed model is;

$$L\left(p/n, \underline{x}, \eta, \phi\right) \alpha \left[\pi_1(p)\right]^{x_1} \left[\pi_2(p)\right]^{x_2} \left[1 - \pi_1(p) - \pi_2(p)\right]^{n - x_1 - x_2} \tag{14}$$

The maximum likelihood estimator of p is obtained by solving

$$\frac{\partial L(.)}{\partial q} = 0 \tag{15}$$

Equation (15) has no solution in closed form except when $l = 0$. Therefore, the

equation is solved iteratively with the aid codes developed using R or MATLAB software.

Next, we consider the calculation of the asymptotic variance of our estimator. This variance is used to calculate the asymptotic relative efficiency of this estimator with respect to other estimators. The variance is obtained by computing Equation (4) to get;

$$\text{var}\left(\hat{p}_1\right) = \frac{\pi_1(p)\pi_2(p)\left(1-\pi_1(p)-\pi_2(p)\right)p^2\left(1-p\right)^2}{n\left[\left\{E-pkF(l)\right\}\right]^2 Y} \tag{16}$$

where

$$Y = \left(\eta+\phi-1\right)^2\pi_2(p)\left(1-\pi_2(p)\right) + \left(\phi(1-\phi)-\eta(1-\eta)\right)^2\pi_1(p)\left(1-\pi_1(p)\right) \\ + 2\left(\eta+\phi-1\right)\left(\phi(1-\phi)-\eta(1-\eta)\right)\pi_1(p)\pi_2(p) \tag{17}$$

The derivation of the variance is discussed in the **Appendix**. This variance is used to determine the Wald interval of *p* as

$$\hat{p}_1 \pm z_{\alpha/2}\sqrt{\text{var}\left(\hat{p}_1\right)} \tag{18}$$

where is $\left(1-\alpha\right)$ the confidence level.

4. Model Comparison

In this section we compare our proposed model with the model outlined in Section (3) and that proposed by [16]. This is accomplished by computing Asymptotic Relative Efficiency (ARE). If the other estimators are denoted by $\hat{p}_B$ and our estimator is denoted by $\hat{p}_1$, then

$$\text{ARE} = \frac{\text{var}\left(\hat{p}_B\right)}{\text{var}\left(\hat{p}_1\right)} \tag{19}$$

Therefore, ARE > 1 implies that the proposed model is more efficient than the other models. First, we check whether re-testing improves efficiency by comparing the proposed model with the one-stage testing procedure outlined in Section (3). The sensitivity and specificity of the tests are set at 99% and 95%. When sensitivity and specificity are set to 99% the results are shown in **Figure 1** and **Table 1** below.

For $I = 0$, the comparison is between re-testing in batch testing with one-stage [2] model. Clearly, from **Table 1** and **Table 2** re-testing improves efficiency and this just confirms what other authors have noted [16] and [14]. When *I* is more than 0, we compare one stage batch testing model based on quality control and re-testing in the quality control model. It is noted that re-testing substantially improves efficiency. For instance, if $p = 0.01$, $k = 10$ and $\eta = \phi = 95\%$, re-testing improves efficiency by 18 times over one stage batch testing in quality control. This increases to about 47 times when $\eta = \phi = 99\%$.

Next, we compare the proposed model with the one developed by [16] in which batches that test negative initially are given a re-test. He demonstrated that the model was more efficient than [2] model with imperfect test for relatively higher values of *p*. The results are presented in **Table 3** and **Table 4**.
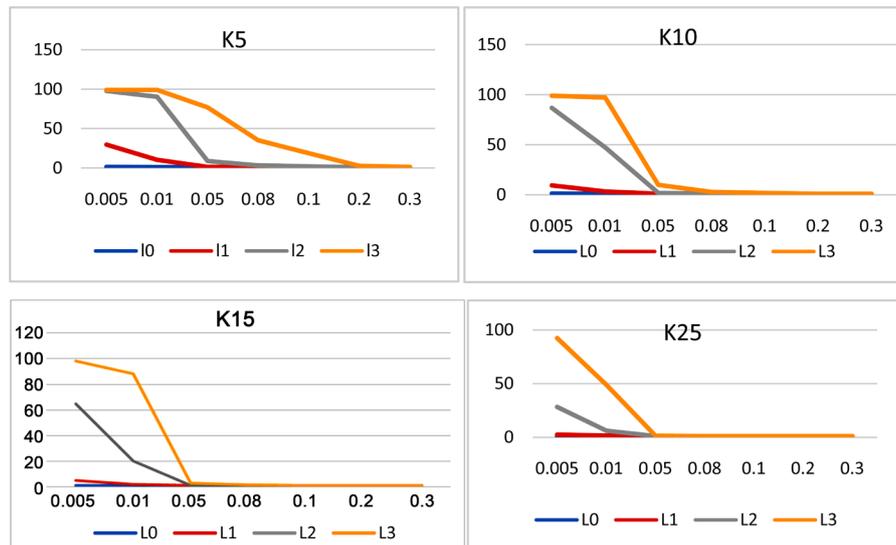
Figure 1. The figures show a plot of Asymptotic Relative Efficiency (ARE) against $\hat{p}_1$ for varying values of k.

Table 1. ARE of $\hat{p}_1$ with respect to one-stage quality control procedure with $\eta = \phi = 99\%$.

	l			
p	0	1	2	3
	$k = 5$			
0.005	1.3925	29.6041	97.8326	99.0070
0.01	1.1938	10.2325	90.3806	98.9624
0.05	1.0343	1.4308	8.9272	76.7396
0.08	1.0194	1.1738	3.1516	35.0415
0.10	1.0145	1.1127	2.1448	18.7860
0.20	1.0049	1.0282	1.1625	2.4562
0.30	1.0021	1.0113	1.0514	1.3139
	$k = 10$			
0.005	1.1943	9.3430	86.7328	98.8871
0.01	1.0946	3.2798	47.3194	97.1246
0.05	1.0150	1.1061	1.8519	9.8372
0.08	1.0078	1.0433	1.2390	2.6844
0.10	1.0055	1.0279	1.1324	1.7656
0.20	1.0013	1.0061	1.0211	1.0728
0.30	1.0004	1.0019	1.0063	1.0186
	$k = 15$			
0.005	1.1279	4.8138	64.9368	98.2320
0.01	1.0615	2.0179	20.3049	88.3148
0.05	1.0087	1.0486	1.2656	2.7860
0.08	1.0041	1.0195	1.0786	1.3549
0.10	1.0027	1.0123	1.0444	1.1698

Continued

0.20	1.0005	1.0021	1.0067	1.0185
0.30	1.0001	1.0005	1.0016	1.0043
		$k = 25$		
0.005	1.0749	2.4088	28.2591	92.4917
0.01	1.0351	1.3769	5.8627	48.8542
0.05	1.0039	1.0181	1.0687	1.2826
0.08	1.0015	1.0066	1.0210	1.0641
0.10	1.0009	1.0038	1.0117	1.0324
0.20	1.0001	1.0004	1.0012	1.0032
0.30	1.0001	1.0001	1.0002	1.0004

Table 2. ARE of $\hat{p}_1$ with respect to one-stage quality control procedure with $\eta = \phi = 95\%$.

	l			
p	0	1	2	3
		$k = 5$		
0.005	2.7786	17.5694	19.0419	19.0499
0.01	1.9232	14.3298	18.9861	19.0497
0.05	1.1720	2.9339	13.7249	18.8566
0.08	1.0988	1.8323	7.8361	17.8817
0.10	1.0744	1.5493	5.3864	16.4788
0.20	1.0269	1.1418	1.7807	6.2454
0.30	1.0127	1.0584	1.2556	2.4505
		$k = 10$		
0.005	1.9255	13.9318	18.9553	19.0492
0.01	1.4634	8.0892	18.3375	19.0369
0.05	1.0769	1.5179	4.4725	14.1599
0.08	1.0409	1.2162	2.1256	6.8122
0.10	1.0294	1.1410	1.6414	4.1806
0.20	1.0086	1.0327	1.1072	1.3591
0.30	1.0041	1.0114	1.0336	1.0948
		$k = 15$		
0.005	1.6209	10.4527	18.7026	19.0447
0.01	1.3047	5.0041	16.6937	18.9689
0.05	1.0457	1.2419	2.2422	7.0496
0.08	1.0227	1.0990	1.3868	2.6231
0.10	1.0156	1.0633	1.2216	1.8143
0.20	1.0045	1.0127	1.0356	1.0943
0.30	1.0029	1.0045	1.0098	1.0237
		$k = 25$		
0.005	1.3695	6.1213	17.4752	19.0028

Continued

0.01	1.1759	2.7141	11.5841	18.3792
0.05	1.0218	1.0919	1.3395	2.3160
0.08	1.0097	1.0352	1.1067	1.3171
0.10	1.0065	1.0212	1.0605	1.1627
0.20	1.0028	1.0040	1.0081	1.0179
0.30	1.0026	1.0027	1.0031	1.0044

Table 3. ARE of $\hat{p}_1$ with respect to Nyongesa (2011) model with $\eta = \phi = 99\%$.

	l		
p	1	2	3
	k = 5		
0.001	0.2164	5.3546e−07	2.3848e−13
0.005	3.9413	0.0008	8.6069e−09
0.01	8.5029	0.0189	9.3740e−07
0.05	36.5456	4.0149	0.0471
0.08	54.3329	10.3408	0.5451
0.10	65.2915	15.8345	1.3283
	k = 10		
0.001	1.1035	2.5420e−05	1.3884e−10
0.005	9.6112	0.0398	6.2503e−06
0.01	18.1836	0.5981	0.0007
0.05	70.0768	21.3549	3.2719
0.08	100.1277	48.0430	13.0015
0.10	118.2877	69.6712	24.0546
	k = 15		
0.001	2.3559	0.0002	3.2298e−09
0.005	14.5392	0.2588	0.0002
0.01	26.7917	2.1507	0.0159
0.05	96.8177	46.0052	12.8843
0.08	136.1110	97.7387	44.5177
0.10	160.5376	138.8346	78.5275
	k = 25		
0.001	4.9448	0.0025	1.3748e−07
0.005	23.4561	1.4957	0.0068
0.01	42.1520	6.9966	0.3764
0.05	140.3911	106.8778	53.5889
0.08	197.8769	218.4010	166.7186
0.10	236.2497	312.2300	287.0305

Table 4. ARE of $\hat{p}_1$ with respect to Nyongesa (2011) model with $\eta = \phi = 95\%$.

	L		
p	1	2	3
	k = 5		
0.001	0.0069	1.5678e−08	6.9819e−15
0.005	0.2092	1.2939e−05	1.4529e−10
0.01	0.8649	0.0003	1.2169e−08
0.05	7.7909	0.3066	0.0005
0.08	11.6831	1.4814	0.0122
0.10	13.9829	2.6849	0.0512
	k = 10		
0.001	0.0373	6.0763e−07	3.3149e−12
0.005	1.0505	0.0006	8.1271e−08
0.01	3.2221	0.0128	7.3743e−06
0.05	15.0585	3.8583	0.2162
0.08	21.1658	9.7347	1.9364
0.10	24.7741	14.3689	4.2763
	k = 15		
0.001	0.0946	4.1540e−06	6.6609e−11
0.005	2.2266	0.0044	1.8235e−06
0.01	5.4724	0.0909	0.0002
0.05	20.5304	9.2938	1.8903
0.08	28.2758	20.3172	8.6669
0.10	32.9241	28.9329	15.9529
	k = 25		
0.001	0.2941	4.307e−05	2.3228e−09
0.005	4.6059	0.0489	7.2465e−05
0.01	9.1276	0.7265	0.0065
0.05	29.1147	22.2477	10.6125
0.08	39.6495	45.4753	34.7850
0.10	45.9407	64.9692	60.7597

It is evident from the tables above the proposed model is more efficient that the model proposed by [16] when rate of defectiveness is relatively high. For example, if $l = 1$, $k = 15$, $p = 0.01$ and $\eta = \phi = 99\%$, the proposed model is 26 times more efficient than [16] model.

5. Discussion and Conclusion

Batch testing model applied in quality control where batches that test positive

initially are re-tested has been developed. The maximum likelihood estimator of p and its asymptotic variance derived. The asymptotic variance of the estimator is used to compare our model with two other models; one stage batch testing model based on quality control [15] and one proposed by [16]. This is accomplished by computing the asymptotic relative efficiency of the estimator.

The results show that our estimator generally has a small variance as compared to the estimators obtained from the other two models for relatively high values of p for given cut off values, batch sizes, sensitivity and specificity.

This work advances the field of batch testing by introducing cut off value greater than zero and further generalizing the model first introduced by [2] in which the cut off value is strictly equal to zero. This case is particularly encountered in quality control. We recommend that a model in which negative batches are re-tested is considered. If some of the negative batches test positive on re-test, then it indicates the presence of test errors.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Mwangi, O.W., Islam, A. and Luke, O. (2015) Bootstrap Confidence Intervals for Proportions of Unequal Sized Groups Adjusted for Overdispersion. *Open Journal of Statistics*, **5**, 502-510. https://doi.org/10.4236/ojs.2015.56052

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

[3] Thompson, K.H. (1962) Estimation of the Proportion of Vectors in a Natural Population of Insects. *Biometrics*, **18**, 568-578. https://doi.org/10.2307/2527902

[4] Gastwirth, J.L. and Johnson, W.O. (1994) Screening with Cost-Effective Quality Control: Potential Applications to HIV and Drug Testing. *Journal of the American Statistical Association*, **89**, 972-981. https://doi.org/10.1080/01621459.1994.10476831

[5] Tu, X., Litvak, E. and Pagano, M. (1995) On the Informativeness and Accuracy of Pooled Testing in Estimating Prevalence of a Rare Disease: Application to HIV Screening. *Biometrika*, **82**, 287-297. https://doi.org/10.1093/biomet/82.2.287

[6] McV Messam, L.L., Branscum, A.J., Collins, M.T. and Gardner, I.A. (2008) Frequentist and Bayesian Approaches to Prevalence Estimation Using Examples from Johne's Disease. *Animal Health Research Reviews*, **9**, 1-23. https://doi.org/10.1017/S1466252307001314

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

[8] Chick, S.E. (1996) Bayesian Models for Limiting Dilution Assay and Group Test Data. *Biometrics*, **52**, 1055-1062. https://doi.org/10.2307/2533066

[9] Montesinos-López, O., Montesinos-López, A., Crossa, J., Eskridge, K. and Sáenz, R.A. (2011) Erratum: Optimal Sample Size for Estimating the Proportion of Trans-

genic Plants Using the Dorfman Model with a Random Confidence Interval. *Seed Science Research*, **21**, 235-245. https://doi.org/10.1017/S0960258511000055

[10] Montgomery, D.C. (2020) Introduction to Statistical Quality Control. John Wiley & Sons, Hoboken.

[11] Hwang, F.K. (1976) Group Testing with a Dilution Effect. *Biometrika*, **63**, 671-680. https://doi.org/10.1093/biomet/63.3.671

[12] Remund, K.M., Dixon, D.A., Wright, D.L. and Holden, L.R. (2001) Statistical Considerations in Seed Purity Testing for Transgenic Traits. *Seed Science Research*, **11**, 101-119. http://europepmc.org/abstract/AGR/IND23232809

[13] Kennedy, N.L. (2004) Multistage Group Testing Procedure (Group Screening). *Communications in Statistics—Simulation and Computation*, **33**, 621-637. https://doi.org/10.1081/SAC-200033231

[14] Brookmeyer, R. (1999) Analysis of Multistage Pooling Studies of Biological Specimens for Estimating Disease Incidence and Prevalence. *Biometrics*, **55**, 608-612. https://doi.org/10.1111/j.0006-341X.1999.00608.x

[15] Wanyonyi, R.W., Nyongesa, K. and Wasike, A.A.M. (2015) Estimation of Proportion of a Trait by Batch Testing Model in a Quality Control Process. *American Journal of Theoretical and Applied Statistics*, **4**, 619-629. https://doi.org/10.11648/j.ajtas.20150406.34

[16] Kennedy, N.L. (2011) Dual Estimation of Prevalence and Disease Incidence in Pool-Testing Strategy. *Communications in Statistics—Theory and Methods*, **40**, 3218-3229. https://doi.org/10.1080/03610926.2010.493257

[17] Kline, R.L., Brothers, T.A., Brookmeyer, R., Zeger, S. and Quinn, T.C. (1989) Evaluation of Human Immunodeficiency Virus Seroprevalence in Population Surveys Using Pooled Sera. *Journal of Clinical Microbiology*, **27**, 1449-1452. https://doi.org/10.1128/JCM.27.7.1449-1452.1989

[18] Damaschke, P. (2006) Threshold Group Testing. In: Ahlswede, R., Bäumer, L., Cai, N., Aydinian, H., Blinovsky, V. and Deppe, C., Eds., *General Theory of Information Transfer and Combinatorics*, Springer, Berlin, 707-718.

[19] Chen, H.-B. and De Bonis, A. (2011) An Almost Optimal Algorithm for Generalized Threshold Group Testing with Inhibitors. *Journal of Computational Biology*, **18**, 851-864. https://doi.org/10.1089/cmb.2010.0030

[20] Xie, M., Tatsuoka, K., Sacks, J. and Young, S.S. (2001) Group Testing with Blockers and Synergism. *Journal of the American Statistical Association*, **96**, 92-102. https://doi.org/10.1198/016214501750333009

[21] Wanyonyi, R.W., Nyongesa, K.L. and Wasike, A. (2015) Estimation of Proportion of a Trait by Batch Testing with Errors in Inspection in a Quality Control Process. *International Journal of Statistics and Applications*, **5**, 268-278. http://article.sapub.org/10.5923.j.statistics.20150506.02.html https://doi.org/10.11648/j.ajtas.20150406.34

[22] Okoth, A.W., *et al.* (2017) Multi-Stage Adaptive Pool Testing Model with Test Errors, Improved Efficiency. *IOSR Journal of Mathematics*, **13**, 43-55. https://doi.org/10.9790/5728-1301024355

[23] Matiri, G., Nyongesa, K. and Islam, A. (2017) Sequentially Selecting Between Two Experiment for Optimal Estimation of a Trait with Misclassification. *American Journal of Theoretical and Applied Statistics*, **6**, 79-89. https://doi.org/10.11648/j.ajtas.20170602.12

[24] Lk, N. (2018) Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Pre-

valence and Its Extension to Multi-Stage.

[25] Lehmann, E.L. and Casella, G. (1998) Theory of Point Estimation (Springer Texts in Statistics). Second Edition, Springer, Berlin.

Appendix. Derivation of Asymptotic Variance

We consider the computation of the asymptotic variance of our estimator as presented in Equation (16). The variance is computed by solving for

$$\left\{ -E\left(\frac{\partial^2 \log L(.)}{\partial p^2} \right) \right\}^{-1}$$

But

$$\frac{\partial \log L(.)}{\partial p} = \frac{x_1}{\pi_1(p)}\pi_1'(p) + \frac{x_2}{\pi_2(p)}\pi_2'(p) + \frac{x_3}{\pi_3(p)}\pi_3'(p) \tag{20}$$

and

$$\pi_1'(p) = (\eta + \phi - 1)\left(\frac{E - pkF(l)}{p(1-p)} \right)$$

$$\pi_2'(p) = [\phi(1-\phi) - \eta(1-\eta)]\left(\frac{E - pkF(l)}{p(1-p)} \right)$$

Note that

$$\pi_3'(p) = 1 - \pi_1(p) - \pi_2(p)$$

Hence

$$\pi_3'(p) = -[\pi_1'(p) + \pi_2'(p)]$$

Getting the second derivative

$$\frac{\partial^2 L(.)}{\partial p^2} = x\left[\frac{\pi_1(p)\pi_1''(p) - (\pi_1'(p))^2}{\pi_1^2(p)} \right] + x_2\left[\frac{\pi_2(p)\pi_2''(p) + (\pi_2'(p))^2}{\pi_2^2(p)} \right]$$
$$+ \left[\frac{\pi_3(p)\pi_3''(p) - (\pi_3'(p))^2}{\pi_3^2(p)} \right] \tag{21}$$

Taking expectations of Equation (21), we obtain

$$-E\left(\frac{\partial^2 \log L(.)}{\partial p^2} \right) = n\left[\frac{(\pi_1'(p))^2}{\pi_1(p)} + \frac{(\pi_2'(p))^2}{\pi_2(p)} + \frac{(\pi_3'(p))^2}{\pi_3(p)} \right]$$
$$= \frac{n}{\pi_1(p)\pi_2(p)\pi_3(p)}\left[\pi_2(p)(1-\pi_2(p))(\pi_1'(p))^2 \right. \tag{22}$$
$$\left. + \pi_1'(p)(1-\pi_1(p))(\pi_2'(p))^2 + 2\pi_1(p)\pi_2(p)\pi_1'(p)\pi_2'(p) \right]$$

Implying that

$$\text{var}(\hat{p}_1) = \frac{\pi_1(p)\pi_2(p)(1-\pi_1(p)-\pi_2(p))p^2(1-p)^2}{n[\{E - pkF(l)\}]^2 Y} \tag{23}$$

where

$$Y = (\eta + \phi - 1)^2 \pi_2(p)(1-\pi_2(p)) + (\phi(1-\phi) - \eta(1-\eta))^2 \pi_1(p)(1-\pi_1(p))$$
$$+ 2(\eta + \phi - 1)(\phi(1-\phi) - \eta(1-\eta))\pi_1(p)\pi_2(p)$$

completing the proof as demanded in the text.