**American Journal of Theoretical and Applied Statistics**
2016; 5(3): 138-145
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20160503.19
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

# An Almost Unbiased Estimator in Group Testing with Errors in Inspection

**Langat Erick Kipyegon, Tonui Benard Cheruiyot*, Langat Reuben Cheruiyot**

Department of Mathematics & Computer Science, University of Kabianga, Kericho, Kenya

**Email address:**
ericlan64@yahoo.com (L. E. Kipyegon), tonuib@kabianga.ac.ke (T. B. Cheruiyot), rlangat@kabianga.ac.ke (L. R. Cheruiyot)
*Corresponding author

**To cite this article:**
Langat Erick Kipyegon, Tonui Benard Cheruiyot, Langat Reuben Cheruiyot. An Almost Unbiased Estimator in Group Testing with Errors in Inspection. *American Journal of Theoretical and Applied Statistics*. Vol. 5, No. 3, 2016, pp. 138-145. doi: 10.11648/j.ajtas.20160503.19

**Received**: April 26, 2016; **Accepted**: May 9, 2016; **Published**: May 25, 2016

**Abstract:** The idea of pooling samples into pools as a cost effective method of screening individuals for the presence of a disease in a large population is discussed. Group testing was designed to reduce diagnostic cost. Testing population in pools also lower misclassification errors in low prevalence population. In this study we violate the assumption of homogeneity and perfect tests by investigating estimation problem in the presence of test errors. This is accomplished through Maximum Likelihood Estimation (MLE). The purpose of this study is to determine an analytical procedure for bias reduction in estimating population prevalence using group testing procedure in presence of tests errors. Specifically, we construct an almost unbiased estimator in pool-testing strategy in presence of test errors and compute the modified MLE of the prevalence of the population. For single stage procedures, with equal group sizes, we also propose a numerical method for bias correction which produces an almost unbiased estimator with errors. The existence of bias has been shown with the help of Taylor's expansion series, for group sizes greater than one. The indicator function with errors is used in the development of the model. A modified formula for bias correction has been analytically shown to reduce the bias of a group testing model. Also, the Fisher information and asymptotic variance has been shown to exist. We use MATLAB software for simulation and verification of the model. Then various tables are drawn to illustrate how the modified bias formula behaves for different values of sensitivities and specificities.

**Keywords:** Group Testing, Maximum Likelihood Estimator, Almost Unbiased Estimator, Bias Adjuster Formula, Bias-Corrected Estimates

## 1. Introduction

Group testing, also known as pooled samples, occurs when units from a population are pooled and tested as a group for the presence of a particular attribute, such as a disease, or a defect. The problem of group testing is concerned with classifying each $N$ given units in a population into two disjoint categories which are *defectives* and *non-defectives*. The characteristic feature is that any number of units $k$ (in a group) can be tested simultaneously but the information obtained from a single test on $k$ units, without any chance of error, is either negative or positive. When the test is negative, it implies that all the $k$ units in that group are non-defective and when it is positive it implies that at least one of the units in $k$ units is defective, but it is not known which ones or how many are defective. The problem is to devise a sequential sampling scheme which minimizes the expected number of tests required to classify all the $N$ units as defective or non-defective. That is, to find $p$ the proportions of defective units in the population. The idea is to construct $n$ groups of size $k$ of, say, biological samples (e.g. blood) from a population of size $N$. The population $N$ may be from a number of individuals pooled into $n$ groups. Each group $1, 2 \ldots, n$ is tested by a single test. If the reading is negative, the group is dropped from further investigation, otherwise, sequential testing is performed on the group. The sequential testing procedure provided will enable us construct an almost unbiased estimator or propose analytical procedures that reduce biasness.

Group testing where subjects are tested in pools rather than individually has a long history of successful applications in screening of infectious diseases. Whether the aim is to diagnose individuals (classification) or estimate disease prevalence, it is

cost effective since the test is done on a group and not individuals. Group testing first appeared in the statistical literature in the context of blood testing [1] but has since been applied in many fields, including transmission of viruses by insect vectors [2], genetics [3], plant disease assessment [4] and quality control [5]. Pool testing is a two-fold procedure: The first procedure being the identification of positive individuals in a population cost effectively (see [1]). This involves testing batches of items and those that test positive, the constituent members are tested for identification of positive ones. There is abundant literature on this classification problem. For instance, [6] and [7] proposed hierarchical or multistage model based on Dorfman idea that involves subdividing positive pools into smaller pools with the purpose of reducing cost. They showed that some savings can be achieved via multistage models. The second procedure is the estimation of prevalence rate as championed by [8]. There is also an abundant literature on this problem as established by [9] and [10]. Still on estimation problem, [11] used the Maximum Likelihood Estimation (MLE) to estimate elements of drugs in a composition of elements. In multistage problem with the purpose of estimation, [12] proposed a multistage estimation model. [10] proposed confidence intervals for prevalence rate when pool testing procedures are applied. Bayesian inference on population prevalence has also been studied (see for instance [9]). Some procedures for bias reduction in group testing model without errors has been proposed, [13].

In the group testing literature, with the objective of estimating the prevalence of an attribute of interest, the MLE is the dominant procedure. If the group size is $k = 1$, the MLE has been shown to be unbiased estimator ([12] and [14]). Whereas, when the group size is $k > 1$, the MLE has been shown to be biased and this is a drawback in statistical inference in pool testing procedure as observed by [8]. A more general bias adjustment, which was not specifically derived for group testing, was described by [15]. The purpose of this study is to determine an analytical procedure for bias reduction in estimating population prevalence using group testing procedure in presence of test errors. Specifically, we construct an almost unbiased estimator in pool-testing strategy in presence of test errors and compute the modified MLE of the prevalence of the population. For single stage procedures, with equal group sizes, we also propose a numerical method for bias correction which produces an almost unbiased estimator with errors.

The rest of the paper is organized as follows. In Section 2, we give analytic construction of MLE where we discuss the MLE in group testing with errors, an almost unbiased estimator, the bias adjuster formula and bias-corrected estimates of the prevalence of the population. In Section 3 we give results and discussions while in Section 4 we give conclusion.

# 2. Analytic Construction of MLE

In this section, it is shown that the MLE for group size $k = 1$ is unbiased but for $k > 1$, it is biased. Secondly, ways of improving the MLE have been proposed using the Bias Adjuster formula.

### *2.1. MLE in Group Testing with Errors*

Suppose we have a large population of size $N$, the idea is to construct $n$ groups from this population. The population $N$ may be blood from a number of individuals then pooled into $n$ groups. The probability of classifying a group as positive in absence of errors is

$$\pi = 1 - (1 - p)^k \tag{1}$$

where $p$ is the probability of an individual being classified as positive and $k$ is the size of the pool. When the error element is introduced in (1) we obtain

$$\pi = \psi[1 - (1 - p)^k] + (1 - \phi)(1 - p)^k, \tag{2}$$

where $\psi$ and $\phi$ denotes sensitivity and specificity of the test kits. By sensitivity we mean probability of classifying a positive group as positive while specificity is the probability of classifying a non-positive group as non-positive. For the derivations of (2), see [12] and [14]. Upon using (2) the MLE of $p$ can be obtained as;

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

For $k = 1$, and upon using (3) the MLE of $p$ is unbiased, that is, $E(\hat{p}) = p$. But for $k > 1$, it has been shown to be biased. That is for $k > 1$, $E(\hat{p}) \neq p$ and this is a drawback to group testing inference. We therefore construct an analytical procedure that can help reduce bias in the subsequent sections.

### *2.2. Improved MLE to Almost Unbiased Estimator*

In this section we construct a MLE in pool testing with errors in inspection such that when $k > 1$ the proposed $\hat{p}$ is almost unbiased. To achieve this, we require Gart's formula

$$B(p) = B_G(p) + O(n_i^{-2}) \tag{4}$$

with $B_G(p) = -\dfrac{\left[\dfrac{2\partial I}{\partial p} + E\dfrac{\partial^3 l}{\partial p^3}\right]}{2I(p)^2}$

where $I$ is the Fisher information, $l$ is a log likelihood, $B_G(p)$ is the Gart's bias and $O$ is the order of the error, see [15]. We notice that the Fisher information, $I$, has been computed in pool testing literature (See, for instance [12] and [14]) and provided as;

$$I = \frac{nk^2(1-p)^{2k-2}[\psi-(1-\phi)]^2}{\pi(1-\pi)} \tag{5}$$

From (5), we have

$$\frac{\partial I}{\partial p} = \frac{nk^2(1-p)^{2k-3}(\psi+\phi-1)^2}{\pi^2(1-\pi)^2}\{2\pi - 2\pi^2 - k\pi - k\psi + 2\pi k\psi\}$$

and

$$\frac{2\partial I}{\partial p} = \frac{2nk^2(\psi+\phi-1)^2(1-p)^{2k-3}}{\pi^2(1-\pi)^2}\{2\pi + 2\pi k\psi - k\psi - k\pi - 2\pi^2\} \tag{6}$$

Detailed derivation of (6) is provided in Appendix A.
Also

$$E\left[\frac{\partial^3 l}{\partial p^3}\right] = \frac{nk^2(1-p)^{2k-3}[\psi-(1-\phi)]^2}{\pi^2(1-\pi)^2}\{3\pi k - 3\pi^2 k - 3\pi + 3\pi^2\}$$

$$-4\pi k(1-p)^k[\psi-(1-\phi)] + 2k(1-p)^k[\psi-(1-\phi)] \quad (7)$$

Technical derivation of (7) is provided in Appendix B.
Equations (6) and (7) are vital in the next sections.

### *2.3. Bias Adjuster Formula*

With equations (6) and (7) at hand and Formula (4), upon substitution we have

$$B_G(p) = -\left\{\frac{2nk^2[\psi-(1-\phi)]^2(1-p)^{2k-3}}{\pi^2(1-\pi)^2}[2\pi + 2\pi k\psi - k\psi - k\pi - 2\pi^2] + \frac{nk^2[\psi-(1-\phi)]^2(1-p)^{2k-3}}{\pi^2(1-\pi)^2}[3\pi k - 3\pi^2 k - 3\pi + 3\pi^2 - 4\pi k(1-p)^k[\psi-(1-\phi)] + 2k(1-p)^k[\psi-(1-\phi)]]\right\} \div \frac{2n^2k^4(1-p)^{4k-4}[\psi-(1-\phi)]^4}{\pi^2(1-\pi)^2} \quad (8)$$

On simplifying equation (8) above we obtain;

$$B_G(p) = \frac{(k-1)\pi(1-\pi)}{2nk^2[\psi-(1-\phi)]^2(1-p)^{1-2k}} + O(n_i^{-2}), \quad (9)$$

where $O(n_i^{-2})$ implies the order of the error.
The Gart's Bias-Corrected estimates are given by;

$$\hat{p}_V = \hat{p} - B_G(\hat{p}) \quad (10)$$

and

$$\hat{p}_H = \hat{p} - B_G(\hat{p}_H) \quad (11)$$

as suggested by [13].

We distinguish these two approaches by describing them as 'Vertical' or 'Horizontal' correction or more briefly as 'Gart-V' and 'Gart – H'. Gart–V correction has the disadvantage of not being able to handle $\hat{p} = 1$, owing to a zero denominator in $B_G(p)$. Gart– H correction, in contrast, does not require $\hat{p} = 1$, to be substituted in $B_G(p)$, and so an estimate can be found. Gart's method with Vertical correction is highly effective in reducing the bias for small $p$. With Horizontal correction, Gart's method is moderately effective, (see [13]). In our discussion in the next section, our main focus is on Vertical correction since it is highly effective.

## 3. Results and Discussion

In this section, a sample from a population is taken, split in groups and tested for some attribute of interest. The estimates of $p$, the proportions of defective units in the population, under MLE, Bias Adjuster Formula and Gart's Vertical Correction $\hat{p}_V$ are obtained. These estimates are then represented empirically by use of tables.

We considered bias as the main issue in group testing problems. We investigated the MLE for single stage procedure. The estimates in the case of the 'all positives' outcome are shown to have a large effect on bias calculations.

We base our discussions on Monte Carlo Simulation for Bias and MLE for various group sizes for given sensitivity and specificity (see Tables 1, 2, 3 and 4)

In the tables that follow, we have results for simulated MLE for various group sizes with sensitivity and specificity of $1.00, 0.99, 0.95$ and $0.90$ and prevalence rate of $p = 0.01, 0.02, 0.03, \ldots, 0.09$.

The simplest possible group testing procedure is where a single stage with equal group sizes is considered. We take a population of 200 samples split into 8 groups each of size 25 samples and tested for the prevalence of some attribute of interest. Hence $N = 200, k = 25$ and $n = 8$. From equation (3), the MLE of $p$ for different values of $\psi$ and $\phi$ yields results as tabulated in Tables 1, 2, 3 and 4.

Also from equation (9) when $\psi = \phi = 100\%$, it reduces to;

$$B_G(p) = \frac{(k-1)(1-(1-p)^k)}{2nk^2(1-p)^{k-1}} + O(n_i{}^{-2}) \quad (3)$$

Upon simulating on MATLAB, the MLE for different values of sensitivities and specificities, we obtain results provided in the following tables:

***Table 1.*** *Estimates of positive groups for* $\psi = \phi = 100\%$.

| Method | Estimates for the following numbers of positive groups ( $x$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| $MLE(\hat{p})$ | 0.000 | 0.005 | 0.011 | 0.019 | 0.027 | 0.038 | 0.054 | 0.080 | 1.000 |
| $B_G(\hat{p})$ | 0.00000 | 0.00032 | 0.00076 | 0.00145 | 0.00229 | 0.00377 | 0.00683 | 0.01554 | |
| Gart V, $\hat{p}_V = \hat{p} - B_G(\hat{p})$ | 0.00000 | 0.00468 | 0.0102 | 0.01755 | 0.02471 | 0.03423 | 0.04718 | 0.06446 | 1.000 |

***Table 2.*** *Estimates of positive groups for* $\psi = \phi = 99\%$.

| Method | Estimates for the following numbers of positive groups ( $x$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| $MLE(\hat{p})$ | 0.000 | 0.005 | 0.011 | 0.018 | 0.027 | 0.038 | 0.054 | 0.082 | 1.832 |
| $B_G(\hat{p})$ | 0.000025 | 0.000350 | 0.000799 | 0.001414 | 0.002388 | 0.003937 | 0.007200 | 0.018140 | -0.0027 |
| Gart V, $\hat{p}_V = \hat{p} - B_G(\hat{p})$ | 0.00000 | 0.00465 | 0.01020 | 0.01659 | 0.02460 | 0.03410 | 0.04680 | 0.06390 | 1.82934 |

***Table 3.*** *Estimates of positive groups for* $\psi = \phi = 95\%$.

| Method | Estimates for the following numbers of positive groups ( $x$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| $MLE(\hat{p})$ | -0.002 | 0.003 | 0.010 | 0.018 | 0.027 | 0.040 | 0.058 | 0.095 | 1.890 |
| $B_G(\hat{p})$ | 0.00001043 | 0.00034970 | 0.00090898 | 0.00169730 | 0.00283200 | 0.0051288 | 0.01040 | 0.04290 | -0.00105 |
| Gart V, $\hat{p}_V = \hat{p} - B_G(\hat{p})$ | -0.00201 | 0.00265 | 0.00909 | 0.01630 | 0.02417 | 0.03487 | 0.04760 | 0.05210 | 1.89100 |

***Table 4.*** *Estimates of Positive groups for* $\psi = \phi = 90\%$.

| Method | Estimates for the following numbers of positive groups ( $x$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| $MLE(\hat{p})$ | -0.005 | 0.001 | 0.008 | 0.017 | 0.027 | 0.042 | 0.064 | 0.129 | 1.920 |
| $B_G(\hat{p})$ | 0.0000184 | 0.000415 | 0.00103 | 0.002040 | 0.0035844 | 0.007184 | 0.01810 | 0.35729 | -0.000113 |
| Gart V, $\hat{p}_V = \hat{p} - B_G(\hat{p})$ | -0.00501 | 0.00058 | 0.00697 | 0.01496 | 0.02342 | 0.03482 | 0.04590 | -0.2283 | 1.92011 |

From the values in the tables above, the MLE of $p$ and the Bias Adjuster of $p$ increase with increase in sensitivity and specificity in the test kits. When the sensitivity and specificity are 95% and 90%, the MLE of $p$ and the correction are negative. This is when there is no positive outcome in the group test. It is also seen that the Vertical correction decreases with decrease in sensitivity and specificity of test kits. On the other hand, when all the groups test positive, $p = 1$ is almost certainly an overestimate of $p$ as it is most unusual for every unit in a population to be positive. When all the individuals in the group are positive the probability is beyond 1 and the outcome is shown to have a large effect on bias calculations. The main reason for this rare occurrence is the presence of test errors i.e. the specificity and sensitivity and not human errors during the experiment since they are assumed to have more effect.

# 4. Conclusion

From the tables in Section 3, it is observed that the bias has been considerably reduced when $\psi = \phi = 0.90$ compared to when $\psi = \phi = 1.00$, which conforms to the conventions of describing a bias of less than about 10% as acceptable. It is shown that the Vertical correction is most effective in reducing the bias. However, the correction is undefined when $\hat{p}$ =1 owing to a zero in the denominator in modified Bias Adjuster. Thus when all groups test positive, in this situation, pool testing procedure is not applicable. For instance, see [13] who simply stated that '*if all groups turn out to be positive, no sensible estimate of the infection rate can be obtained from the data'*. The derivative of the log-likelihood function of the distribution has been shown to yield the Fisher information ($I$) and the asymptotic variance. The modified Bias Adjuster Formula for bias correction has been shown to reduce the bias. This is evident in Tables 1, 2, 3 and 4 for different specificity and sensitivity.

In this study we have considered bias reduction by constructing an almost unbiased estimator in a simple group testing model. However, there exist complex group testing models in literature. For further research, it will be interesting to study the bias properties and suggest some modification to such models.

# Appendixes

***Appendix A***

In this appendix, we provide detail derivation of equation (6). First, we know that;

$$I = \frac{nk^2(\psi + \phi - 1)(1-p)^{2k-2}}{\pi(1-\pi)}$$

On simplifying the above, we get

$$I = nk^2(\psi + \phi - 1)^2(1-p)^{2k-2}\left[\frac{1}{\pi} + \frac{1}{(1-\pi)}\right]$$

The first derivative can be obtained as

$$\frac{\partial I}{\partial p} = nk^2(\psi + \phi - 1)^2 \frac{\partial}{\partial p}\left[\frac{(1-p)^{2k-2}}{\pi} + \frac{(1-p)^{2k-2}}{(1-\pi)}\right]$$

$$= nk^2(\psi + \phi - 1)^2\left[\left[\frac{-(2k-2)(1-p)^{2k-3}\pi - (1-p)^{2k-2}\pi'}{\pi^2}\right] + \left[\frac{-(2k-2)(1-p)^{2k-3}(1-\pi) - (1-p)^{2k-2}(1-\pi)'}{(1-\pi)^2}\right]\right]$$

$$= nk^2(\psi+\phi-1)^2\left\{\left[\frac{-(2k-2)(1-p)^{2k-3}}{\pi} - \frac{(1-p)^{2k-2}\pi'}{\pi^2}\right] + \left[\frac{-(2k-2)(1-p)^{2k-3}}{(1-\pi)} - \frac{(1-p)^{2k-2}(1-\pi)'}{(1-\pi)^2}\right]\right\} \quad \text{(A1)}$$

Substituting for

$$\pi' = k(1-p)^{k-1}[\psi-(1-\phi)] \; and \; (1-\pi)' = k(1-p)^{k-1}[-\psi+(1-\phi)]$$

In the above, we get

$$\frac{\partial I}{\partial p} = nk^2(\psi+\phi-1)^2\left\{\left[\frac{-(2k-2)(1-p)^{2k-3}}{\pi} - \frac{(1-p)^{2k-2}k(1-p)^{k-1}[\psi-(1-\phi)]}{\pi^2}\right] + \left[\frac{-(2k-2)(1-p)^{2k-3}}{(1-\pi)} - \frac{(1-p)^{2k-2}k(1-p)^{k-1}[-\psi+(1-\phi)]}{(1-\pi)^2}\right]\right\} \quad \text{(A2)}$$

On simplifying, yields

$$\frac{\partial I}{\partial p} = nk^2(\psi+\phi-1)^2\left\{-(2k-2)(1-p)^{2k-3}\left[\frac{1}{\pi}+\frac{1}{1-\pi}\right] - k(1-p)^{3k-3}[\psi-(\phi-1)]\left[\frac{1}{\pi^2}-\frac{1}{(1-\pi)^2}\right]\right\}$$

$$= nk^2(\psi+\phi-1)^2\left\{-(2k-2)(1-p)^{2k-3}\left[\frac{1}{\pi(1-\pi)}\right] - k(1-p)^{3k-3}[\psi-(1-\phi)]\left[\frac{(1-\pi)^2-\pi^2}{\pi^2(1-\pi)^2}\right]\right\}$$

$$= nk^2(\psi+\phi-1)^2\left\{-(2k-2)(1-p)^{2k-3}\frac{\pi(1-\pi)}{\pi^2(1-\pi)^2} - k(1-p)^{3k-3}[\psi-(1-\phi)]\left[\frac{1-2\pi+\pi^2-\pi^2}{\pi^2(1-\pi)^2}\right]\right\} \quad \text{(A3)}$$

Factoring out $\frac{1}{\pi^2(1-\pi)^2}$ and $(1-P)^{2k-3}$, we get

$$\frac{\partial I}{\partial p} = \frac{nk^2(1-p)^{2k-3}(\psi+\phi-1)^2}{\pi^2(1-\pi)^2}\{-(2k-2)\pi(1-\pi) - k(1-p)^k(1-2\pi)[\psi-(1-\phi)]\}$$

$$= \frac{nk^2(1-p)^{2k-3}(\psi+\phi-1)^2}{\pi^2(1-\pi)^2}\{(2-2k)(\pi-\pi^2) - (k-2\pi k)(1-p)^k[\psi-(1-\phi)]\}) \quad \text{(A4)}$$

Which simplifies to

$$\frac{\partial I}{\partial p} = \frac{nk^2(1-p)^{2k-3}(\psi+\phi-1)^2}{\pi^2(1-\pi)^2}\{(2-2k)(\pi-\pi^2) - (1-p)^k[\psi-(1-\phi)](k-2\pi k)\} \quad \text{(A5)}$$

But we know that $\pi = \psi-(1-p)^k[\psi-(1-\phi)]$ or $\pi-\psi = -(1-p)^k[\psi-(1-\phi)]$
Hence;

$$\frac{\partial I}{\partial p} = \frac{nk^2(1-p)^{2k-3}(\psi+\phi-1)^2}{\pi^2(1-\pi)^2}\{2\pi-2\pi^2-2k\pi+2k\pi^2+(\pi-\psi)(k-2k\pi)\}$$

$$= \frac{nk^2(1-p)^{2k-3}(\psi+\phi-1)^2}{\pi^2(1-\pi)^2}\{2\pi-2\pi^2-2k\pi+2k\pi^2+k\pi-2\pi^2k-k\psi+2\pi k\psi\} \quad \text{(A6)}$$

On simplifying, we get;

$$\frac{\partial I}{\partial p} = \frac{nk^2(1-p)^{2k-3}(\psi+\phi-1)^2}{\pi^2(1-\pi)^2}\{2\pi-2\pi^2-k\pi-k\psi+2\pi k\psi\} \quad \text{(A7)}$$

Therefore;

$$\frac{2\partial I}{\partial p} = \frac{2nk^2(\psi+\phi-1)^2(1-p)^{2k-3}}{\pi^2(1-\pi)^2}\{2\pi+2\pi k\psi-k\psi-k\pi-2\pi^2\} \quad \text{(A8)}$$

***Appendix B***

Detail derivation of equation (7) is accomplished in this appendix. Upon taking natural logs of *l (p/x),* we have

$$l(p) = \log\binom{n}{x} + x\log\pi + (n-x)\log(1-\pi)$$

On finding the derivative with respect to $p$, we have;

$$\frac{\partial\, l(p)}{\partial p} = \left[\frac{x}{\pi}\right]\pi' + \left[\frac{(n-x)}{(1-\pi)}\right](1-\pi)'$$

The second derivative becomes;

$$\frac{\partial^2 l(p)}{\partial p^2} = x\left[\frac{\pi'}{\pi}\right]' + (n-x)\left[\frac{1-\pi'}{1-\pi}\right]' = x\frac{\pi\pi''-\pi'\pi'}{\pi^2} + (n-x)\left[\frac{(1-\pi)(1-\pi)''-(1-\pi)'(1-\pi)'}{(1-\pi)^2}\right] \quad \text{(B1)}$$

Upon simplification, we have:

$$\frac{\partial^2 l(p)}{\partial p^2} = x\left[\frac{\pi''}{\pi} - \frac{\pi'\pi'}{\pi^2}\right] + (n-x)\left[\frac{(1-\pi)''}{(1-\pi)} - \frac{(1-\pi)'(1-\pi)'}{(1-\pi)^2}\right] = x\left[\frac{\pi''}{\pi}\right] - x\left[\frac{\pi'\pi'}{\pi^2}\right] + (n-x)\left[\frac{(1-\pi)''}{(1-\pi)}\right] - (n-x)\left[\frac{(1-\pi)'(1-\pi)'}{(1-\pi)^2}\right] \quad \text{(B2)}$$

The third derivative is given by;

$$\frac{\partial^3 l(p)}{\partial p^3} = x\left[\frac{\pi''}{\pi}\right]' - x\left[\frac{\pi'\pi'}{\pi^2}\right]' + (n-x)\left[\frac{(1-\pi)''}{(1-\pi)}\right]' - (n-x)\left[\frac{(1-\pi)'(1-\pi)'}{(1-\pi)^2}\right]' = x\left[\frac{\pi\pi'''-\pi'\pi''}{\pi^2}\right] - x\left[\frac{[\pi'\pi']'\pi^2-[\pi^2]'\pi'\pi'}{\pi^4}\right] +$$
$$(n-x)\frac{(1-\pi)(1-\pi)'''-(1-\pi)'(1-\pi)''}{(1-\pi)^2} - (n-x)\left[\frac{\left([(1-\pi)'(1-\pi)']'(1-\pi)^2 - (1-\pi)'(1-\pi)'[(1-\pi)^2]'\right)}{(1-\pi)^4}\right] \quad \text{(B3)}$$

To obtain equation (7), we take expectation on both sides of equation $(B3)$ to obtain;

$$E\left[\frac{\partial^3 l(p)}{\partial p^3}\right] = E\left\{x\left[\frac{\pi\pi'''-\pi'\pi''}{\pi^2}\right] - x\left[\frac{\pi^2[\pi'\pi']'-\pi'\pi'[\pi^2]'}{\pi^4}\right] + (n-x)\left[\frac{(1-\pi)(1-\pi)'''-(1-\pi)'(1-\pi)''}{(1-\pi)^2}\right] - (n-x)\left[\frac{(1-\pi)^2[(1-\pi)'(1-\pi)']'-(1-\pi)'(1-\pi)'[(1-\pi)^2]'}{(1-\pi)^4}\right]\right\} \quad \text{(B4)}$$

But $E(x) = n\pi$. Hence on substituting, we get

$$E\left[\frac{\partial^3 l(p)}{\partial p^3}\right] = n\pi\left[\frac{\pi\pi''' - \pi'\pi''}{\pi^2}\right] - n\pi\left[\frac{\pi^2[\pi'\pi']' - \pi'\pi'[\pi^2]'}{\pi^4}\right]$$
$$+n(1-\pi)\left[\frac{(1-\pi)(1-\pi)'''-(1-\pi)'(1-\pi)''}{(1-\pi)^2}\right] - n(1-\pi)\left[\frac{(1-\pi)^2[(1-\pi)'(1-\pi)']'-(1-\pi)'(1-\pi)'[(1-\pi)^2]'}{(1-\pi)^4}\right] \quad \text{(B5)}$$

Upon simplification, we get;

$$E\left[\frac{\partial^3 l(p)}{\partial p^3}\right] = n\left[\pi''' - \frac{\pi'\pi''}{\pi}\right] - n\left[\frac{[\pi'\pi']'}{\pi} - \frac{\pi'\pi'[\pi^2]'}{\pi^3}\right] + n\left[(1-\pi)''' - \frac{(1-\pi)'(1-\pi)''}{(1-\pi)}\right] - n\left[\left[\frac{(1-\pi)'(1-\pi)'}{(1-\pi)}\right]' - \frac{(1-\pi)'(1-\pi)'[(1-\pi)^2]'}{(1-\pi)^3}\right] \quad \text{(B6)}$$

where $\pi = \psi - (1-p)^k[\psi - (1-\phi)]$

$$\pi^2 = (\psi - (1-p)^k[\psi - (1-\phi)]).(\psi - (1-p)^k[\psi - (1-\phi)]) = \psi^2 - 2\psi(1-p)^k[\psi - (1-\phi)] + (1-p)^{2k}\left(\psi - (1-\phi)\right)^2 \quad \text{(B7)}$$

$$[\pi^2]' = 2\psi - 2k(1-p)^{2k-1}[\psi - (1-\phi)]^2 \quad \text{(B8)}$$

$$\pi' = k(1-p)^{k-1}[\psi - (1-\phi)] \quad \text{(B9)}$$

$$\pi'\pi' = k^2(1-p)^{2k-2}[\psi - (1-\phi)]^2 \quad \text{(B10)}$$

$$[\pi'\pi']' = (2k-2)k^2(1-p)^{2k-3}[\psi - (1-\phi)]^2(-1) \quad \text{(B11)}$$

$$\pi'' = -k(k-1)(1-p)^{k-2}[\psi - (1-\phi)] \quad \text{(B12)}$$

$$\pi''' = k(k-1)(k-2)(1-p)^{k-3}[\psi - (1-\phi)], \quad \text{(B13)}$$

and

$$(1-\pi) = 1 - \psi + (1-p)^k[\psi - (1-\phi)] \quad \text{(B14)}$$

$$(1-\pi)^2 = (1-\psi+(1-p)^k[\psi-(1-\phi)]).(1-\psi+(1-p)^k[\psi-(1-\phi)])$$
$$= 1-\psi+(1-p)^k[\psi-(1-\phi)] - \psi + \psi^2 - \psi(1-p)^k[\psi-(1-\phi)] + (1-p)^k[\psi-(1-\phi)]$$
$$- \psi(1-p)^k[\psi-(1-\phi)] + (1-p)^{2k}[\psi-(1-\phi)]^2$$

$$= 1 - 2\psi + 2(1-p)^k[\psi-(1-\phi)] + \psi^2 - 2\psi(1-p)^k[\psi-(1-\phi)] + (1-p)^{2k}[\psi-(1-\phi)]^2 \quad \text{(B15)}$$

$$[(1-\pi)^2]' = -2k(1-p)^{k-1}[\psi-(1-\phi)] + 2k\psi(1-p)^{k-1}[\psi-(1-\phi)] - 2k(1-p)^{2k-1}[\psi-(1-\phi)]^2 \quad \text{(B16)}$$

$$(1-\pi)' = -k(1-p)^{k-1}[\psi-(1-\phi)] \quad \text{(B17)}$$

$$(1-\pi)'(1-\pi)' = k^2(1-p)^{2k-2}[\psi-(1-\phi)]^2 \quad \text{(B18)}$$

$$[(1-\pi)'(1-\pi)']' = (2k-2)k^2(1-p)^{2k-3}[\psi-(1-\phi)]^2(-1) \quad \text{(B19)}$$

$$(1-\pi)'' = k(k-1)(1-p)^{k-2}[\psi-(1-\phi)] \quad \text{(B20)}$$

$$(1-\pi)''' = -k(k-1)(k-2)(1-p)^{k-3}[\psi-(1-\phi)] \quad \text{(B21)}$$

Substituting $(B7)$ although to $(B21)$ into $(B4)$, we get;

$$E\left[\frac{\partial^3 l(p)}{\partial p^3}\right] = n\left[k(k-1)(k-2)(1-p)^{k-3}[\psi-(1-\phi)] + \frac{k^2(k-1)(1-p)^{2k-3}[\psi-(1-\phi)]^2}{\pi}\right] - n\left[\frac{-k^2(2k-2)(1-p)^{2k-3}[\psi-(1-\phi)]^2}{\pi} - \frac{(k^2(1-p)^{2k-2}[\psi-(1-\phi)]^2).(2\psi k(1-p)^{k-1}[\psi-(1-\phi)])-2k(1-p)^{2k-1}[\psi-(1-\phi)]^2}{\pi^3}\right] + n\left[-k(k-1)(k-2)(1-p)^{k-3}[\psi-(1-\phi)] + \frac{k^2(k-1)(1-p)^{2k-3}[\psi-(1-\phi)]^2}{(1-\pi)}\right] - n\left[\frac{(2k-2)k^2(1-p)^{2k-3}[\psi-(1-\phi)]^2(-1)}{(1-\pi)}\right] - \left[\frac{k^2(1-p)^{2k-2}[\psi-(1-\phi)]^2.\left[-2k(1-p)^{k-1}[\psi-(1-\phi)]\right]}{(1-\pi)^3}\right] + \left[\frac{-2k\psi(1-p)^{k-1}[\psi-(1-\phi)]-2k(1-p)^{2k-1}[\psi-(1-\phi)]^2}{(1-\pi)^3}\right] \quad \text{(B22)}$$

which reduces to;

$$E\left[\frac{\partial^3 l(p)}{\partial p^3}\right] = \frac{nk^2(k-1)(1-p)^{2k-2}[\psi-(1-\phi)]^2}{\pi} + \frac{nk^2(k-1)(1-p)^{2k-3}[\psi-(1-\phi)]^2}{(1-\pi)} + \frac{nk^2(2k-2)(1-p)^{2k-3}[\psi-(1-\phi)]^2}{\pi} + \frac{nk^2(2k-2)(1-p)^{2k-3}[\psi-(1-\phi)]^2}{(1-\pi)} - \frac{(k^2(1-p)^{2k-2}[\psi-(1-\phi)]^2).\left(2k(1-p)^{k-1}[\psi-(1-\phi)]\left[\psi-(1-p)^k[\psi-(1-\phi)]\right]\right)n}{\pi^3} + \frac{\{k^2(1-p)^{2k-2}[\psi-(1-\phi)]^2\}(-1)2k(1-p)^{k-1}[\psi-(1-\phi)]\left[1-\psi+(1-p)^k[\psi-(1-\phi)]\right]n}{(1-\pi)^3} \quad \text{(B23)}$$

Hence on substituting $\pi$ and $(1-\pi)$ in $(B23)$, we get;

$$E\left[\frac{\partial^3 l(p)}{\partial p^3}\right] = nk^2(k-1)(1-p)^{2k-3}[\psi-(1-\phi)]^2\left[\frac{1}{\pi}+\frac{1}{1-\pi}\right] + nk^2(2k-2)(1-p)^{2k-3}[\psi-(1-\phi)]^2\left[\frac{1}{\pi}+\frac{1}{1-\pi}\right] + 2nk^3(1-p)^{3k-3}[\psi-(1-\phi)]^3\left[\frac{-\pi}{\pi^3}-\frac{1-\pi}{(1-\pi)^3}\right]$$

Hence;

$$E\left[\frac{\partial^3 l(p)}{\partial p^3}\right] = \frac{nk^2(1-p)^{2k-3}[\psi-(1-\phi)^2]}{\pi^2(1-\pi)^2}[\pi(1-\pi)(-k-1)+2(k-1)\pi(1-\pi)] - [2k(1-p)^k[\psi-(1-\phi)](2\pi-1)]$$
$$= \frac{nk^2(1-p)^{2k-3}[\psi-(1-\phi)^2]}{\pi^2(1-\pi)^2}\left[3(k-1)\pi(1-\pi)-(4\pi k-2k)(1-p)^k[\psi-(1-\phi)]\right]$$

Further;

$$E\left[\frac{\partial^3 l(p)}{\partial p^3}\right] = \frac{nk^2(1-p)^{2k-3}[\psi-(1-\phi)]^2}{\pi^2(1-\pi)^2}\left[3(k-1)(\pi-\pi^2)-(4\pi k-2k)(1-p)^k[\psi-(1-\phi)]\right] \quad \text{(B24)}$$

Equation $(B24)$ simplifies to;

$$E\left[\frac{\partial^3 l(p)}{\partial p^3}\right] = \frac{nk^2(1-p)^{2k-3}[\psi-(1-\phi)]^2}{\pi^2(1-\pi)^2}\{3\pi k-3\pi^2 k-3\pi+3\pi^2\}-4\pi k(1-p)^k[\psi-(1-\phi)]+2k(1-p)^k[\psi-(1-\phi)] \quad \text{(B25)}$$

# References


[1] Dorfman, R. (1943). The detection of defective members of large population. *Ann. Math. Statistics*: 14, 436-440.

[2] Walter, S. D., Hildreth, S. W and Beaty, B. J (1980). Estimation of Infection Rates in Population of Organisms using Pools of Variable Size. *Ann. J. Epidem,* 112, 124-128.

[3] Chick, S. E (1996). Bayesian Models for limiting dilution assay and group test data. *Biometrics*, 52, 1055-1062.

[4] Fletcher, J. D, Rusell. A. C and Butler. R. C (1999). Seed borne cucumber Mosaic virus in New Zealand lentil groups: yield effects and disease incidence. *New Zeal J. Crop Hort* 27, 197-204.

[5] Wanyonyi, R. W, Nyongesa, L. K, 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 Application,* 5(6), 268-278.

[6] Johnson, N. L, Kotz, S, Wu, X. (1992). Inspection Errors for Attributes in Quality Control. Chapman & Hall.

[7] Nyongesa, L. K (2004). Multistage group testing procedure (group screening). *Communication in Statistics-Simulation and Computation*, 33, 621–637.

[8] Thompson, K. H (1962). Estimation of the Proportions of Vectors in a Natural Population of Insects; *Biometrics*, 18, 568-578.

[9] Bilder, C. R and Tebbs J. M (2005). Empirical Bayes Estimation of the Disease Transmission Probability in Multiple Vector Transfer Designs. *Biometrika*. J. 47, 502-516.

[10] Hepworth, G. (2005). Confidence intervals for proportions estimated by group testing with groups of unequal sizes. *Journal of Agricultural, Biological and Environmental Statistics,* 10, 478–497.

[11] Tu, M. X, Litvak, E. and Pagano, M. (1995). On the Informative and Accuracy of Pooled Testing in Estimating Prevalence of a Rare Disease: *Application to HIV Screening. Biometrika,* 82, 287–297.

[12] Brookmeyer, R. (1999). Analysis of multistage pooling studies of Biological specimens for Estimating Disease Incidents and prevalence. *Biometrics,* 55, 608-612.

[13] Hepworth, G and Watson, R. (2009). Debiased Estimation of proportions in group testing. *Applied Statistics,* 58, 105–121.

[14] Nyongesa, L. K (2011). Dual Estimation of Prevalence and Disease Incidence in pooling strategy. *Communication in Statistics Theory and Method,* 40, 1–12.

[15] Gart, J. J. (1991). An application of score methodology: confidence intervals and tests of fit for one-hit curves. In *Handbook of Statistics* (Eds C. R. Rao and R. Chakraborty), Vol. 8, pp. 395–406. Amsterdam: Elsevier.