3-23-2012

Study on a Hierarchy Model

Suisui Che
Florida International University, sche001@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

STUDY ON A HIERARCHY MODEL

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS

by

Suisui Che

2012

To: Dean Kenneth Furton
College of Arts and Sciences

This thesis, written by Suisui Che, and entitled Study on a Hierarchy Model, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Florence George

Kai Huang

Jie Mi, Major Professor

Date of Defense: March 23, 2012

The thesis of Suisui Che is approved.

Dean Kenneth Furton
College of Arts and Sciences

Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2012

ACKNOWLEDGMENTS

I would like to gratefully and sincerely thank Dr. Jie Mi for his supervision, patience and understanding. Before I did research with him I never expected a professor could support a student so much both academically and spiritually. For everything you have done for me, Dr. Jie Mi, I thank you.

I would also like to thank the Department of Mathematics and Statistics, all the professors with who I took my graduate level courses, especially the members of my committee, Dr. Kai Huang and Dr. Florence George for their valuable accessibility. In particular, I'd would like to thank Dr. Kai Huang for his technical support and guidance in LATEX and MATLAB software.

Also, I would like to thank my close friend, Zeyi Wang. He provided lots of help when installing all the needed software and writing the programs. Without his unforgettable kindness and unwavering support during the whole period of my research I could not finish it on time.

I thank my best friends Shihua Luo, Yipin Shi, Fan Yang. I felt so lucky to have them around me when I experienced the graduate school and overseas life in U.S. It was your love and encouragement that helped me through all those hardships.

Finally, I wish to thank my parents for their love.

ABSTRACT OF THE THESIS

STUDY ON A HIERARCHY MODEL

by

Suisui Che

Florida International University, 2012

Miami, Florida

Professor Jie Mi, Major Professor

The statistical inference about the parameters of Binomial-Poisson Hierarchy Model are discussed. On the basis of the estimators of paired observations we consider the other two cases with extra observations on both the first and second layer of the model. The $MLEs$ of λ and p are derived and it is also proved that the MLE $\widehat{\lambda}$ is also the $UMVUE$ of λ. By using multivariate central limit theory and large sample theory, the asymptotic behavior of both the estimators based on extra observations on the first and second layer are obtained respectively. The performances of these estimators are compared numerically based on extensive Monte Carlo simulation. Simulation studies indicate that the performance of these estimators are more efficient than those only based on paired observations. Inference about the confidence intervals forλ and p is presented for both cases. The efficiency of the estimators are compared with condition given that same number of extra observations are provided.

Keywords: Binomial-Poisson distribution, hierarchy model, parameter estimation, UMVUE, MLE, confidence Interval.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

§ 1. Introduction

The statistical inference about the Binomial-Poisson hierarchy distribution by method of maximum likelihood is considered. The parameters of interest are rate λ and proportion p. We will try to obtain the maximum likelihood estimators of the two parameters. In some experimental situations, it is necessary to estimate a proportion using several groups of cases where the sampling is random. Therefore compound distributions will have to be considered. Inference about the parameters of this compound distribution has been studied by many authors. For instance, general method about parameter estimation by McGuire, Brindley, and Bancroft (1957), methods about $MLEs$ by Sprott (1958),limiting theorem by Hodges and Lucien (1960) and many other discussions from the other statisticians. However, all the studies in the literature used only paired data but not unpaired data, for example Ocerin and Perez (2002). Using only paired data could lead to a big loss of information and so reduce the accuracy of estimation and the power of testing hypotheses. Our research will use the additional data information for improving the estimation results and increasing the power of the tests. Because of the complexity of the data structure with the additional unpaired observations the exact sampling distributions of estimators may be not available and thus large sample theory may have to be applied in order to obtain approximate confidence intervals and and evaluate performance of estimators.

Extensive research has been conducted on the Binomial-Poisson model. The point estimator of parameters in the Binomial Poisson model can be traced back as early as the 1950's.

Sprott (1958) studied a procedure of fitting the Poisson-Binomial distribution by using the maximum likelihood method, the moments method and the sample zero frequency method. The likelihood function $L(\widehat{p}) = \sum a_k F(k) - N = 0$ served as a base of the later research. Shumway and Gurland (1960) described a much simpler maximum likelihood estimates and computed probabilities derived from the results of Sprott (1958). Finally, the maximum likelihood and recurrence relations were rewritten in terms of ratios of Poisson factorial moments in the fitting of the Poisson binomial distribution. When p_i is an ith estimate of p, we may calculate $L(\widehat{p}_i)$ and $L'(\widehat{p}_i)$, then an approximation of $\widehat{P}_{i+1}$ can be computed by using the relation $\widehat{p}_{i+1} = \widehat{p}_i - [L(\widehat{p}_i)/L'(\widehat{p}_i)]$. Hodges and Lucien (1960) raised a question about the Poisson limiting theorem of the Poisson-binomial distribution, which had been ignored for quite a long time. It drew attention to the basic assumption of the Poisson distribution that in many applications the probability p of the various trials could not be considered equally likely. In this case, the limit theorem (von Mises 1921), which required a large sample size n, a small α and a moderate λ was not restrictive enough. Hodges and Lucien (1960) presented an approximation theorem which was based on a relatively large sample size n and different probabilities p_i. The original limiting was also included in this approximation theorem as a special case.

Katti and Gurland (1962) continued research on the method of the moment estimators presented by Sprott (1958). They found the regions given by Sprott were not wide enough to include the parameter vector in most practical cases. So they discussed a new method of estimation with the estimators from minimum chi-square estimation method, which was compared with the maximum likelihood estimators and proved to be much more efficient than the method discussed by Sprott (1958). The approximate formulas are developed to evaluate the $MLEs$ of the probability p and the bound of the error can be determined as well. Johnson and Kotz (1969), and Johnson (1992) also studied the Binomial-Poisson compound distribution. Both of the discussions treated the Binomial-Poisson distribution as a discrete distribution and considered the parameter estimation methods.

Ouyang (1993) discussed Poisson-Poisson and Binomial-Poisson sampling in forestry based on the result presented by Cacoullos and Papageorgiou (1982). Ocerin and Perez (2002) restudied the numerical approximation, by using an example of an experimental design. Petri's dishes were used in the experiment to perform a bacteriological sowing, with the aim of predicting the proportion of mutations. Provided the paired data set they could obtain a numerical approximation of an estimator of the proportion p, which can be applied in any sample size. Zhu (2003) extended the study to include the Beta-Binomial-Poisson, an EM algorithm is developed to compute both the $MLEs$ and the model parameters and the corresponding stardard error. Shkedy (2005) setup a hierarchical binomial-Poisson model for the Analysis of a crossover design for correlated binary data when the number of trials is dose-dependent.

The Binomial Poisson distribution has been widely applied to studies of plant and insect populations. In this research, we study this hierarchy model but with additional data information. Using the paired observations of Ocerin and Perez(2002), we will try to derive the maximum likelihood estimators of the two parameters of interest, investigate properties such as unbiasedness and whether the estimators is the uniform minimum variances estimators of these estimators, and study the asymptotic distribution of these estimators as well. Assuming that the asymptotic normality of the estimators can be established, we then will be able to construct confidence intervals of the two parameters and to test hypothesis about these parameters. The performance of the estimators based on paired and unpaired data will be compared with that of the estimators based on paired observations merely.

§ 2. Some Preliminary Results

Following the Binomial-Poisson hierarchy model, we let $X \sim$ Poisson (λ), $Y|X = x \sim B(x, p)$. Note that in this hierarchy model Y can be zero, in this case we will define $X = 0$. From these it can be obtained that $Y \sim Poisson(\lambda p)$.

To estimate the parameters λ and p, a sample $\{(x_i, y_i), 1 \leq i \leq n\}$ is drawn from the (X,Y) population. Then the MLEs of λ and p can be derived as

$$\widehat{\lambda}^* = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \text{and} \qquad \widehat{p}^* = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} \quad , \tag{2.1}$$

Here we let $0/0 = 0$ by convention, so $\widehat{p}^*$ can be written as

$$\widehat{p}^* = \left(\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} \right) I_{(0,\infty)} \left(\Sigma_{i=1}^{n} X_i \right)$$

Clearly $n\widehat{\lambda}^* \sim Poisson(n\lambda), E(\widehat{\lambda}^*) = \lambda, Var(\widehat{\lambda}^*) = \lambda/n$, and $E\left(\widehat{p}^*\right) = p\left(1 - e^{-n\lambda}\right)$. It indicates that $\widehat{\lambda}^*$ is an unbiased estimator of λ, but $\widehat{p}^*$ is only asymptotically unbiased. The distribution of $\widehat{p}^*$ is very complicated. Sprott (1958), Ocerin and Perey (2002) studied the numerical approximation of the sampling distribution of $\widehat{p}^*$.

In the present section we will show that $\widehat{\lambda}^*$ and $\widehat{p}^*$ are asymptotically uncorrelated. To this purpose we need the following result.

Lemma 2.1 *The conditional distribution of* $\sum_{i=1}^{n} Y_i$ *given* $\sum_{i=1}^{n} X_i$ *is binomial, namely* $\sum_{i=1}^{n} Y_i | \sum_{i=1}^{n} X_i \sim B(\sum_{i=1}^{n} X_i, p)$.

Proof. It suffices to show that for any integers $0 \leq k \leq j$, it holds that

$$P(\sum_{i=1}^{n} Y_i = k | \sum_{i=1}^{n} X_i = j) = \begin{pmatrix} j \\ \\ k \end{pmatrix} p^k (1-p)^{j-k}. \tag{2.2}$$

Define $N = \{0, 1, ...\}$, $\mathbf{1} = (1, ..., 1)' \in N^n$ and $L = \{\mathbf{l} = (l_1, ...l_n)' \colon \mathbf{l} \in N^n, \mathbf{l}'\mathbf{1} = j\}$.

We have

$$\begin{aligned} P(\sum_{i=1}^{n} Y_i = k | \sum_{i=1}^{n} X_i = j) &= \frac{P(\sum_{i=1}^{n} Y_i = k, \sum_{i=1}^{n} X_i = j)}{P(\sum_{i=1}^{n} X_i = j)} \\ &= \frac{P(\sum_{i=1}^{n} Y_i = k, \sum_{i=1}^{n} X_i = j)}{\frac{(n\lambda)^j}{j!} e^{-n\lambda}} \end{aligned} \tag{2.3}$$

To obtain the numerator of the right hand side of (2.3) note that if we define $M = \{\mathbf{m} = (m_1, ...m_n)' : \mathbf{m} \in N^n, \mathbf{m}'\mathbf{1} = k\}$, then

$$\begin{aligned} & P(\sum_{i=1}^{n} Y_i = k, \sum_{i=1}^{n} X_i = j) \\ = & \sum_{\mathbf{l} \in L} P(\sum_{i=1}^{n} Y_i = k, X_i = l_i, 1 \leq i \leq n) \\ = & \sum_{\mathbf{l} \in L} \sum_{\mathbf{m} \in M} P(Y_i = m_i, X_i = l_i, 1 \leq i \leq n) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\mathbf{l}\in L}\sum_{\mathbf{m}\in M}\left(\prod_{i=1}^{n}\left(\begin{array}{c} l_i \\ \\ m_i \end{array}\right) p^{m_i}(1-p)^{l_i-m_i}\frac{\lambda^{l_i}}{l_i!}e^{-\lambda}\right) \\
&= e^{-n\lambda}\sum_{\mathbf{l}\in L}\sum_{\mathbf{m}\in M}\left(\frac{1}{\prod_{i=1}^{n} m_i!(l_i-m_i)!}\right) p^{\sum_{i=1}^{n} m_i}(1-p)^{\sum_{i=1}^{n}(l_i-m_i)}\lambda^{\sum_{i=1}^{n} l_i} \\
&= e^{-n\lambda}\sum_{\mathbf{l}\in L}\sum_{\mathbf{m}\in M}\left(\frac{1}{\prod_{i=1}^{n} m_i!(l_i-m_i)!}\right) p^{k}(1-p)^{j-k}\lambda^{j} \\
&= p^{k}(1-p)^{j-k}\lambda^{j}e^{-n\lambda}\sum_{\mathbf{l}\in L}\sum_{\mathbf{m}\in M}\left(\frac{1}{\prod_{i=1}^{n} m_i!(l_i-m_i)!}\right) \qquad (2.4)
\end{aligned}
$$

Combining (2.3) and (2.4), we obtain

$$
P(\sum_{i=1}^{n} Y_i = k | \sum_{i=1}^{n} X_i = j) = p^k(1-p)^{j-k}\frac{j!\sum_{\mathbf{m}\in M}\frac{1}{\prod_{i=1}^{n} m_i!}\left(\sum_{\mathbf{l}\in L}\frac{1}{\prod_{i=1}^{n}(l_i-m_i)!}\right)}{n^j} \qquad (2.5)
$$

Hence from (2.5) it follows that

$$
\begin{aligned}
&\frac{1}{\left(\begin{array}{c} j \\ \\ k \end{array}\right)} P\left(\sum_{i=1}^{n} Y_i = k | \sum_{i=1}^{n} X_i = j\right) \\
&= p^k(1-p)^{j-k}\frac{\sum_{\mathbf{m}\in M}\frac{k!}{\prod_{i=1}^{n} m_i!}\left(\sum_{\mathbf{l}\in L}\frac{(j-k)!}{\prod_{i=1}^{n}(l_i-m_i)}\right)}{n^j} \\
&= p^k(1-p)^{j-k}\frac{\sum_{\mathbf{m}\in M}\frac{k!}{\prod_{i=1}^{n} m_i!}n^{j-k}}{n^j} \\
&= p^k(1-p)^{j-k}\frac{n^k n^{j-k}}{n^j}
\end{aligned}
$$

$$=p^k(1-p)^{j-k}$$

which validates the desired equality in (2.2).■

Theorem 2.2 *The covariance between* $\widehat{\lambda}_n^*$ *and* $\widehat{p}_n^*$ *is*

$$Cov\left(\widehat{\lambda}_n^*, \widehat{p}_n^*\right) = \lambda p e^{-n\lambda}$$

and $\widehat{\lambda}^*$ *and* $\widehat{p}^*$ *are asymptotically uncorrelated, where the subscript* n *is used for emphasizing the dependence of the two estimators on* n.

Proof. We have

$$\begin{aligned} Cov\left(\widehat{\lambda}^*, \widehat{p}^*\right) =& E\left(\widehat{\lambda}^*\widehat{p}^*\right) - E\left(\widehat{\lambda}^*\right)E\left(\widehat{p}^*\right) \\ =& E\left(\widehat{\lambda}^*\widehat{p}^*\right) - \lambda p\left(1-e^{-n\lambda}\right). \end{aligned} \tag{2.6}$$

The mean of $\widehat{\lambda}^*\widehat{p}^*$ is

$$\begin{aligned} E\left(\widehat{\lambda}^*\widehat{p}^*\right) =& E\left(\bar{X}_n \cdot \frac{\bar{Y}_n}{\bar{X}_n} I_{(0,\infty)}\left(\bar{X}_n\right)\right) \\ =& E\left(\bar{Y}_n I_{(0,\infty)}\left(\bar{X}_n\right)\right) \\ =& \sum_{k=1}^{\infty} E\left[\bar{Y}_n I_{(0,\infty)}\left(\bar{X}_n\right) \mid \sum_{i=1}^{n} X_i = k\right] P\left(\sum_{i=1}^{n} X_i = k\right) \\ =& \sum_{k=1}^{\infty} \frac{1}{n} E\left[\sum_{i=1}^{n} Y_i | \sum_{i=1}^{n} X_i = k\right] \cdot \frac{(n\lambda)^k}{k!} e^{-n\lambda} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n} \sum_{k=1}^{\infty} kp \cdot \frac{(n\lambda)^k}{k!} \; e^{-n\lambda} && (2.7) \\
&= \frac{p}{n} \sum_{k=1}^{\infty} \frac{(n\lambda)^k}{(k-1)!} \; e^{-n\lambda} \\
&= \frac{p}{n} (n\lambda) \left(\sum_{j=0}^{\infty} \frac{(n\lambda)^j}{j!} \right) \; e^{-n\lambda} \\
&= \lambda p \cdot e^{n\lambda} e^{-n\lambda} = \lambda p && (2.8)
\end{aligned}
$$

where (2.7) follows Lemma 2.1.

Therefore, from (2.6) and (2.8) we obtain

$$
Cov\left(\widehat{\lambda}_n^*, \widehat{p}_n^*\right) = \lambda p - \lambda p \left(1 - e^{-n\lambda}\right) = \lambda p e^{-n\lambda}
$$

■

§ 3. Estimation with Additional Observations on First Layer of Hierarchy

Suppose that our data consist of observations $(x_i, y_i), 1 \leq i \leq n$, and n_1 additional observations $u_j, 1 \leq j \leq n_1$, of X. Certainly it is assumed that $\{(x_i, y_i), 1 \leq i \leq n\}$ and $\{u_j, 1 \leq j \leq n_1\}$ are independent of each other.

In this case, the likelihood function is given as

$$L(\lambda, p) = \left(\prod_{i=1}^{n} \left[\begin{pmatrix} x_i \\ y_i \end{pmatrix} p^{y_i} (1-p)^{x_i - y_i} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right] \right) \left(\prod_{j=1}^{n_1} \frac{\lambda^{u_j}}{u_j!} e^{-\lambda} \right)$$

$$= \left[\left(\prod_{i=1}^{n} y_i! \, (x_i - y_i)! \right) \left(\prod_{j=1}^{n_1} u_j! \right) \right]^{-1} p^{\sum_{i=1}^{n} y_i} (1-p)^{\sum_{i=1}^{n} (x_i - y_i)} \cdot$$

$$\cdot \, \lambda^{\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_1} u_j} e^{-(n_1 + n)\lambda}. \tag{3.1}$$

and thus the log-likelihood function is

$$\begin{aligned} \ln L(\lambda, p) = & C + \left(\sum_{i=1}^{n} y_i \right) \ln p + \left(\sum_{i=1}^{n} (x_i - y_i) \right) \ln (1-p) + \\ & + \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_1} u_j \right) \ln \lambda - (n_1 + n) \lambda \end{aligned} \tag{3.2}$$

where $C = -\left[\sum_{i=1}^{n} (\ln (y_i!) + \ln (x_i - y_i)!) + \sum_{j=1}^{n_1} \ln u_j! \right]$.

From (3.1) we have

$$\frac{\partial \ln L(\lambda, p)}{\partial \lambda} = \frac{\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_1} u_j}{\lambda} - (n_1 + n)$$

and

$$\frac{\partial \ln L(\lambda, p)}{\partial p} = \frac{\sum_{i=1}^{n} y_i}{p} - \frac{\sum_{i=1}^{n} (x_i - y_i)}{1-p}$$

Solving the equation

$$\frac{\partial \ln L(\lambda, p)}{\partial \lambda} = 0 \quad \text{or} \quad \frac{\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_1} u_j}{\lambda} = n_1 + n$$

we obtain the MLE $\widehat{\lambda}$ of λ as

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_1} u_j}{n_1 + n} \tag{3.3}$$

Similarly the MLE $\widehat{p}$ of p based on $\{(x_i, y_i), 1 \leq i \leq n\}$ and $\{u_j, 1 \leq j \leq n_1\}$ is given by

$$\widehat{p} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} \tag{3.4}$$

Therefore,the following result holds.

Theorem 3.1 *The maximum likelihood estimators of λ and p based on paired sample data $\{(x_i, y_i), 1 \leq i \leq n\}$ and additional observations $\{u_j, 1 \leq j \leq n\}$ on X are given by* (3.3) *and* (3.4) *respectively.*

It is difficult to verify whether $\widehat{p}$ is an unbiased estimator for p. However, it is easy to see that $\widehat{\lambda}$ is an unbiased estimator for λ. Actually the result below is true.

Theorem 3.2 *The MLE $\widehat{\lambda}$ of λ is an $UMVUE$ of λ.*

Proof. The mean of $\widehat{\lambda}$ is

$$E\left(\widehat{\lambda}\right) = E\left(\frac{\sum_{i=1}^{n} X_i + \sum_{j=1}^{n_1} U_j}{n_1 + n}\right) = \frac{\sum_{i=1}^{n} E\left(X_i\right) + \sum_{j=1}^{n_1} E\left(U_j\right)}{n_1 + n} = \frac{n\lambda + n_1\lambda}{n_1 + n} = \lambda$$

so $\widehat{\lambda}$ is an unbiased estimator for λ.

Further, note that from the expression (3.1) of the likelihood function, we see that

$$\mathbf{T} = \left(\sum_{i=1}^{n} X_i + \sum_{j=1}^{n_1} U_j, \sum_{i=1}^{n} Y_i, \sum_{i=1}^{n} \left(X_i - Y_i\right)\right)$$

is a sufficient statistics for parameters (λ, p). Moreover, the distribution of $(\mathbf{X}, \mathbf{Y}, \mathbf{U})$, where $\mathbf{X} = (X_1, \cdots, X_n)\,, \mathbf{Y} = (Y_1, \cdots, Y_n)$, and $\mathbf{U} = (U_1, \cdots, U_n)$, is obviously an exponential family and the parameter space $(0, \infty) \times (0, 1)$ is an open set in R^2, so $\mathbf{T}$ is also a complete statistics for (λ, p). Therefore, by Lehmann-Scheffe Theorem $\widehat{\lambda}$ is $UMVUE$ of λ. ■

We have denoted the $MLEs$ of λ and p as $\widehat{\lambda}^*$ and $\widehat{p}^*$ when only paired observations $(x_i, y_i)\,, 1 \leq i \leq n$ are available. As shown in (2.1) it was derived in the literature that

$$\widehat{\lambda}^* = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad \widehat{p}^* = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} \tag{3.5}$$

(see, for instance, Ocerin and Perey (2002)). It can be seen that $\widehat{p} = \widehat{p}^*$ and letting $n_1 = 0$ in (3.3) reduces $\widehat{\lambda}$ to $\widehat{\lambda}^*$.

In order to compare estimators $\mathbf{T}_n = \left(\widehat{\lambda}_n, \widehat{p}_n\right)'$ and $\mathbf{T}_n^* = \left(\widehat{\lambda}_n^*, \widehat{p}_n^*\right)'$ we need to find the asymptotic distribution of $\mathbf{T}_n$ and $\mathbf{T}_n^*$ where the subscript n is used for emphasizing the dependence of the two estimators on n.

Theorem 3.3 *As* $n \to \infty$, $\sqrt{n}\left(\widehat{\lambda}_n^* - \lambda,\ \widehat{p}_n^* - p\right)' \to N\left((0,0)', \sum^*\right)$ *in distribution, where* $\sum^* = \begin{pmatrix} \lambda & 0 \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix}$.

Proof. Note that

$$E(XY) = E[E(XY|X)] = E[XE(Y|X)] = E\left(pX^2\right) = p\left(\lambda^2 + \lambda\right)$$

and thus

$$Cov(X,Y) = E(XY) - E(X)E(Y) = p\left(\lambda^2 + \lambda\right) - \lambda \cdot \lambda p = \lambda p$$

Hence, by the Central Limit Theorem it follows that

$$\sqrt{n}\left(\begin{pmatrix} \bar{X}_n \\ \bar{Y}_n \end{pmatrix} - \begin{pmatrix} \lambda \\ \lambda p \end{pmatrix}\right) \to N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right) \tag{3.6}$$

in distribution as $n \to \infty$, where

$$\Sigma = (\sigma_{ij}) = \begin{pmatrix} \lambda & \lambda p \\ \lambda p & \lambda p \end{pmatrix} \tag{3.7}$$

The estimator $\mathbf{T}_n^*$ can be expressed in terms of $\bar{X}_n$ and $\bar{Y}_n$ as

$$\mathbf{T}_n^* = \begin{pmatrix} \widehat{\lambda}_n^* \\ \\ \widehat{p}_n^* \end{pmatrix} = \begin{pmatrix} \bar{X}_n \\ \\ \bar{Y}_n / \bar{X}_n \end{pmatrix} \equiv \begin{pmatrix} g_1\left(\bar{X}_n, \bar{Y}_n\right) \\ \\ g_2\left(\bar{X}_n, \bar{Y}_n\right) \end{pmatrix}$$

where, $g_1(\theta_1, \theta_2) = \theta_1$ and $g_2(\theta_1, \theta_2) = \theta_2/\theta_1$. Obviously, $g_1(\theta_1, \theta_2) = \lambda$ and $g_2(\theta_1, \theta_2) = p$ when $\theta_1 = \lambda, \theta_2 = \lambda p$. Furthermore, we have

$$\frac{\partial g_1}{\partial \theta_1} = 1, \quad \frac{\partial g_1}{\partial \theta_2} = 0; \quad \frac{\partial g_2}{\partial \theta_1} = -\frac{\theta_2}{\theta_1^2}, \quad \frac{\partial g_2}{\partial \theta_2} = \frac{1}{\theta_1}$$

where $\partial g_i / \partial \theta_j$ means $\partial g_i(\theta_1, \theta_2) / \partial \theta_j, i, j = 1, 2$.

Therefore, by the multivariate Central Limit Theorem it holds that

$$\sqrt{n}\left(\begin{pmatrix} \widehat{\lambda}_n^* \\ \\ \widehat{p}_n^* \end{pmatrix} - \begin{pmatrix} \lambda \\ \\ \lambda p \end{pmatrix}\right) \to N\left(\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}, G\Sigma G'\right) \tag{3.8}$$

where

$$G|_{\theta_1=\lambda,\theta_2=\lambda p} = (G_{ij})\,|_{\theta_1=\lambda,\theta_2=\lambda p} = \left(\frac{\partial g_i}{\partial \theta_j}\right)|_{\theta_1=\lambda,\theta_2=\lambda p}$$
$$= \begin{pmatrix} 1 & 0 \\ \\ -\frac{\theta_2}{\theta_1^2} & \frac{1}{\theta_1} \end{pmatrix}|_{\theta_1=\lambda,\theta_2=\lambda p} = \begin{pmatrix} 1 & 0 \\ \\ -\frac{p}{\lambda} & \frac{1}{\lambda} \end{pmatrix}$$

so

$$G|_{\theta_1=\lambda,\theta_2=\lambda p} = \begin{pmatrix} 1 & 0 \\ \\ -\frac{p}{\lambda} & \frac{1}{\lambda} \end{pmatrix}.$$

It is easy to obtain

$$G\Sigma G'|_{\theta_1=\lambda,\theta_2=\lambda p} = \begin{pmatrix} 1 & 0 \\ -\frac{p}{\lambda} & \frac{1}{\lambda} \end{pmatrix} \begin{pmatrix} \lambda & \lambda p \\ \lambda p & \lambda p \end{pmatrix} \begin{pmatrix} 1 & -\frac{p}{\lambda} \\ 0 & \frac{1}{\lambda} \end{pmatrix}$$

$$= \begin{pmatrix} \lambda & \lambda p \\ -p+p & -p^2+p \end{pmatrix} \begin{pmatrix} 1 & -\frac{p}{\lambda} \\ 0 & \frac{1}{\lambda} \end{pmatrix} = \begin{pmatrix} \lambda & \lambda p \\ 0 & -p^2+p \end{pmatrix} \begin{pmatrix} 1 & -\frac{p}{\lambda} \\ 0 & \frac{1}{\lambda} \end{pmatrix}$$

$$= \begin{pmatrix} \lambda & -p+p \\ 0 & \frac{1}{\lambda}\left(p-p^2\right) \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix} \tag{3.9}$$

From (3.8) and (3.9) it follows that

$$\sqrt{n}\left(\begin{pmatrix} \widehat{\lambda}_n^* \\ \widehat{p}_n^* \end{pmatrix} - \begin{pmatrix} \lambda \\ p \end{pmatrix}\right) \rightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & 0 \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix}\right)$$

in distribution as $n \rightarrow \infty$. ■

Theorem 3.4 *Suppose that there exists* $\alpha < \infty$ *such that* $n_1/n - \alpha = o\left(n^{-1/2}\right)$. *Then*

$$\sqrt{n}\left(\widehat{\lambda}_n - \lambda,\ \widehat{p}_n - p\right)' \rightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\lambda}{\alpha+1} & 0 \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix}\right)$$

in distribution as $n \to \infty$.

Proof. We can rewrite the expression of $\widehat{\lambda}_n$ and $\widehat{p}_n$ as

$$\widehat{\lambda}_n = \frac{\sum_{i=1}^n X_i + \sum_{j=1}^{n_1} U_j}{n_1 + n} = \frac{\bar{X}_n + \frac{n_1}{n}\bar{U}_{n_1}}{\frac{n_1}{n} + 1} \ , \quad \widehat{p}_n = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i} = \frac{\bar{Y}_n}{\bar{X}_n}$$

where $\bar{U}_{n_1} = \sum_{j=1}^{n_1} U_j / n_1$. Let $\widetilde{p}_n = \widehat{p}_n$ and

$$\widetilde{\lambda}_n = \frac{\bar{X}_n + \alpha \bar{U}_{n_1}}{\alpha + 1}.$$

It suffices to show the desired asymptotic normality for $\left(\widetilde{\lambda}_n, \widetilde{p}_n\right)$ due to the fact that $\sqrt{n}\left(\widehat{\lambda}_n - \widetilde{\lambda}_n\right) \to 0$ in probability as $n \to \infty$.

Define $W_n = \alpha \bar{U}_{n_1}$ then $\widetilde{\lambda}_n$ can be expressed as

$$\widetilde{\lambda}_n = \frac{\bar{X}_n + W_n}{\alpha + 1}.$$

It can be shown that

$$\sqrt{n} \begin{pmatrix} \bar{X}_n - \lambda \\ \bar{Y}_n - \lambda p \\ W_n - \alpha \lambda \end{pmatrix} \to N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma_{XYW} \right)$$

in distribution where

$$\Sigma_{XYW} = \begin{pmatrix} \lambda & \lambda p & 0 \\ \lambda p & \lambda p & 0 \\ 0 & 0 & \alpha\lambda \end{pmatrix}.$$

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$ and

$$g_1(\boldsymbol{\theta}) = \frac{\theta_1 + \theta_3}{\alpha + 1}, \quad g_2(\boldsymbol{\theta}) = \frac{\theta_2}{\theta_1}, \quad g_3(\boldsymbol{\theta}) = \theta_3$$

It is easy to see that

$$\begin{aligned} & g_1\left(\bar{X}_n, \bar{Y}_n, W_n\right) = \widetilde{\lambda}_n, \quad g_2\left(\bar{X}_n, \bar{Y}_n, W_n\right) = \widetilde{p}_n, \quad g_3\left(\bar{X}_n, \bar{Y}_n, W_n\right) = W_n \quad \text{and} \\ & g_1(\boldsymbol{\theta})\,|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\alpha\lambda} = \frac{\lambda + \alpha\lambda}{\alpha + 1} = \lambda, \\ & g_2(\boldsymbol{\theta})\,|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\alpha\lambda} = \frac{\lambda p}{\lambda} = p, \\ & g_3(\boldsymbol{\theta})\,|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\alpha\lambda} = \alpha\lambda. \end{aligned}$$

We define the 3×3 matrix $G = (G_{ij}) = (\partial g_i(\boldsymbol{\theta}) / \partial \theta_j)$. And it is easy to see that

$$G|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\alpha\lambda} = \begin{pmatrix} \frac{1}{\alpha+1} & 0 & \frac{1}{\alpha+1} \\ -\frac{\theta_2}{\theta_1^2} & \frac{1}{\theta_1} & 0 \\ 0 & 0 & 1 \end{pmatrix} |_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\alpha\lambda} = \begin{pmatrix} \frac{1}{\alpha+1} & 0 & \frac{1}{\alpha+1} \\ -\frac{p}{\lambda} & \frac{1}{\lambda} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

From the large sample theory it holds that

$$\sqrt{n}\left(\begin{pmatrix}\widetilde{\lambda}_n\\ \widetilde{p}_n\\ W_n\end{pmatrix}-\begin{pmatrix}\lambda\\ p\\ \alpha\lambda\end{pmatrix}\right)=\sqrt{n}\begin{pmatrix}g_1\left(\bar{X}_n,\bar{Y}_n,W_n\right)-g_1\left(\boldsymbol{\theta}\right)\\ g_2\left(\bar{X}_n,\bar{Y}_n,W_n\right)-g_2\left(\boldsymbol{\theta}\right)\\ g_3\left(\bar{X}_n,\bar{Y}_n,W_n\right)-g_3\left(\boldsymbol{\theta}\right)\end{pmatrix}\rightarrow N\left(\begin{pmatrix}0\\0\\0\end{pmatrix},G\Sigma G'\right)$$

due to the independence of $\{(X_i,Y_i),1\leq i\leq n\}$ and $\{U_j,1\leq j\leq n_1\}$. Straight computation yields

$$\begin{aligned}G\Sigma G'|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\alpha\lambda}&=\begin{pmatrix}\frac{1}{\alpha+1}&0&\frac{1}{\alpha+1}\\ -\frac{p}{\lambda}&\frac{1}{\lambda}&0\\ 0&0&1\end{pmatrix}\begin{pmatrix}\lambda&\lambda p&0\\ \lambda p&\lambda p&0\\ 0&0&\alpha\lambda\end{pmatrix}\begin{pmatrix}\frac{1}{\alpha+1}&-\frac{p}{\lambda}&0\\ 0&\frac{1}{\lambda}&0\\ \frac{1}{\alpha+1}&0&1\end{pmatrix}\\
&=\begin{pmatrix}\frac{\lambda}{\alpha+1}&\frac{\lambda p}{\alpha+1}&\frac{\alpha\lambda}{\alpha+1}\\ 0&p\left(1-p\right)&0\\ 0&0&\alpha\lambda\end{pmatrix}\begin{pmatrix}\frac{1}{\alpha+1}&-\frac{p}{\lambda}&0\\ 0&\frac{1}{\lambda}&0\\ \frac{1}{\alpha+1}&0&1\end{pmatrix}\\
&=\begin{pmatrix}\frac{\lambda}{\alpha+1}&0&\frac{\alpha\lambda}{\alpha+1}\\ 0&\frac{p(1-p)}{\lambda}&0\\ \frac{\alpha\lambda}{\alpha+1}&0&\alpha\lambda\end{pmatrix}\end{aligned}$$

This completes the proof. ■

Theorem 3.5 *The estimators based on* $\{(x_i, y_i), 1 \le i \le n\}$ *and* $\{u_j, 1 \le j \le n\}$ *are more efficient than* $\left(\widehat{\lambda}^*, \widehat{p}^*\right)$.

Proof. It has been established that

$$\sqrt{n}\begin{pmatrix} \widehat{\lambda}_n^* & -\lambda \\ \\ \widehat{p}_n^* & -p \end{pmatrix} \to N\left(\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & 0 \\ \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix}\right)$$

and

$$\sqrt{n}\begin{pmatrix} \widehat{\lambda}_n & -\lambda \\ \\ \widehat{p}_n & -p \end{pmatrix} \to N\left(\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}, \Sigma_u\right)$$

in distribution as $n \to \infty$, where Σ_u is

$$\Sigma_u = \begin{pmatrix} \frac{\lambda}{\alpha+1} & 0 \\ \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix}.$$

Hence the difference $\Sigma^* - \Sigma_u$ of the two asymptotic covariance matrices is

$$\begin{pmatrix} \lambda & 0 \\ \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix} - \begin{pmatrix} \frac{\lambda}{\alpha+1} & 0 \\ \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix} = \begin{pmatrix} \frac{\alpha\lambda}{\alpha+1} & 0 \\ \\ 0 & 0 \end{pmatrix}$$

and it is positive semidefinite for all $\lambda > 0, p \in (0,1)$. That is, the estimator $\left(\widehat{\lambda}_n, \widehat{p}_n\right)$ is more efficient than $\left(\widehat{\lambda}_n^*, \widehat{p}_n^*\right)$. In other words, additional observations $\{U_j, 1 \le j \le n_1\}$ provide more information and consequently improve the estimator $\left(\widehat{\lambda}^*, \widehat{p}^*\right)$.

The result of Theorem 3.4 can be applied for constructing (approximate) confidence intervals for λ and p. Obversely that

$$\widehat{\lambda}_n = \frac{\sum_{i=1}^{n} X_i + \sum_{j=1}^{n_1} U_j}{n_1 + n}$$

where the numerator $\sum_{i=1}^{n} X_i + \sum_{j=1}^{n_1} U_j \sim Poisson\left(\left(n_1 + n\right)\lambda\right)$, so there are many ways in the literature for constructing confidence interval for λ. The approximate $1 - \gamma$ confidence interval for p can be obtained as

$$\widehat{p}_n \pm z_{\gamma/2}\sqrt{\frac{\widehat{p}_n\left(1 - \widehat{p}_n\right)}{n\widehat{\lambda}_n}}$$

due to the asymptotic normality $\sqrt{n}\left(\widehat{p}_n - p\right) \to N\left(0, \frac{p(1-p)}{\lambda}\right)$ and Slutsky's Theorem.

■

§ 4. Estimation with Additional Observations on Second Layer of Hierarchy

In the present section it is assumed that in addition to the sample $\{(x_i, y_i), 1 \leq i \leq n\}$ there are n_2 extra independent observations $v_j, (1 \leq j \leq n_2)$ on Y.

The likelihood function based on observations $\{(x_i, y_i), 1 \leq i \leq n\}$ and $\{v_j, (1 \leq j \leq n_2)\}$ is

$$\begin{aligned} L(\lambda, p) = & \left(\prod_{i=1}^{n} \left[\begin{pmatrix} x_i \\ \\ y_i \end{pmatrix} p^{y_i} (1-p)^{x_i - y_i} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right] \right) \left(\prod_{j=1}^{n_2} \frac{(\lambda p)^{v_j}}{v_j!} e^{-\lambda p} \right) \\ = & C \, p^{\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j} (1-p)^{\sum_{i=1}^{n}(x_i - y_i)} \lambda^{\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j} e^{-n\lambda} e^{-n_2 \lambda p}. \end{aligned} \tag{4.1}$$

where

$$C = \frac{\prod_{i=1}^{n} \begin{pmatrix} x_i \\ \\ y_i \end{pmatrix}}{\left(\prod_{i=1}^{n} x_i!\right) \left(\prod_{j=1}^{n_2} v_j!\right)}$$

Thus the log-likelihood function is

$$\begin{aligned} \ln L(\lambda, p) = & \ln C + \left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j \right) \ln p + \left(\sum_{i=1}^{n} (x_i - y_i) \right) \ln(1-p) + \\ & + \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j \right) \ln \lambda - n\lambda - n_2 \lambda p. \end{aligned} \tag{4.2}$$

From (4.1) it can be seen that $\left(\sum_{i=1}^{n} Y_i + \sum_{j=1}^{n_2} V_j, \sum_{i=1}^{n} X_i + \sum_{j=1}^{n_2} V_j\right)$ is an sufficient and complete statistic for parameter (λ, p). Setting $\partial \ln L(\lambda, p)/\partial \lambda = 0$ and

$\partial \ln L(\lambda, p)/\partial p = 0$, we obtain the following equations for determining the $MLEs$ of λ and p:

$$\frac{\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j}{\lambda} - (n + n_2 p) = 0 \tag{4.3}$$

$$\frac{\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j}{p} - \frac{\sum_{i=1}^{n} (x_i - y_i)}{1-p} - n_2 \lambda = 0 \tag{4.4}$$

From (4.3) it follows that

$$\lambda = \frac{\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j}{n + n_2 p} \tag{4.5}$$

and from (4.4)

$$\begin{aligned} \lambda =& \frac{1}{n_2} \left\{ \frac{\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j}{p} - \frac{\sum_{i=1}^{n} (x_i - y_i)}{1-p} \right\} \\ =& \frac{1}{n_2} \cdot \frac{\left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j\right) - \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right) p}{p(1-p)} \end{aligned} \tag{4.6}$$

Equating (4.5) with (4.6) yields

$$\frac{\left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j\right) - \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right) p}{p(1-p)} = \frac{n_2 \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right)}{n + n_2 p},$$

$$n\left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j\right) + n_2 \left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j\right) p - n\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right) p - n_2 \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right) p^2$$

$$= n_2 \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right) p - n_2 \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right) p^2,$$

$$\left[n_2 \sum_{i=1}^{n} (x_i - y_i) + n\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right)\right] p = n\left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j\right),$$

$$\begin{aligned}\widehat{p} =& \frac{n\left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j\right)}{n_2 \sum_{i=1}^{n}(x_i - y_i) + n\sum_{i=1}^{n} x_i + n\sum_{j=1}^{n_2} v_j} \\ =& \frac{n\left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j\right)}{(n_2+n)\sum_{i=1}^{n} x_i - n_2\sum_{i=1}^{n} y_i + n\sum_{j=1}^{n_2} v_j}\end{aligned} \tag{4.7}$$

Substituting (4.7) into (4.5), we obtain

$$\begin{aligned}\widehat{\lambda} =& \frac{\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j}{n + n_2 \cdot \frac{n\left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j\right)}{n_2\sum_{i=1}^{n}(x_i-y_i) + n\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right)}} \\ =& \frac{\left[n_2\sum_{i=1}^{n}(x_i-y_i) + n\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right)\right]\left(\sum_{i=1}^{n} x_i + n\sum_{j=1}^{n_2} v_j\right)}{n\left[n_2\sum_{i=1}^{n}(x_i-y_i) + n\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right)\right] + n_2 n\left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{n_2} v_j\right)} \\ =& \frac{\left[n_2\sum_{i=1}^{n}(x_i-y_i) + n\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right)\right]\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right)}{n(n_2+n)\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right)} \\ =& \frac{n_2\sum_{i=1}^{n}(x_i-y_i) + n\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n_2} v_j\right)}{n(n_2+n)}\end{aligned} \tag{4.8}$$

Summarizing the above, we have the following result.

Theorem 4.1 *The MLEs of* λ *and* p *based on* $\{(x_i, y_i), 1 \leq i \leq n\}$ *and* $\{v_j, 1 \leq j \leq n_2\}$ *are given by* (4.8) *and* (4.7).

The behavior of $\widehat{p}$ given in (4.7) is hardly to be observed directly. However, $\widehat{\lambda}$ has some nice properties as shown below.

Theorem 4.2 *The MLE* $\widehat{\lambda}$ *given by* (4.8) *is the UMVUE of* λ.

Proof. The mean $E\left(\widehat{\lambda}\right)$ is

$$\begin{aligned}
E\left(\widehat{\lambda}\right) &= \frac{n_2 E\left(\sum_{i=1}^{n}\left(X_i - Y_i\right)\right) + nE\left(\sum_{i=1}^{n} X_i\right) + nE\left(\sum_{j=1}^{n_2} V_j\right)}{n\left(n_2 + n\right)} \\
&= \frac{n_2 \cdot n\left(\lambda - \lambda p\right) + n \cdot n\lambda + n \cdot n_2 \lambda p}{n\left(n_2 + n\right)} \\
&= \frac{n_2 n\lambda - n_2 n\lambda p + n^2\lambda + n_2 n\lambda p}{n\left(n_2 + n\right)} \\
&= \frac{n_2 n\lambda + n^2\lambda}{n\left(n_2 + n\right)} = \frac{n\left(n_2 + n\right)\lambda}{n\left(n_2 + n\right)} = \lambda
\end{aligned}$$

■

That is, $\widehat{\lambda}$ is an unbiased estimator for λ. Moreover, it is easy to see that $\widehat{\lambda}$ is a function of the complete sufficient statistics $\left(\sum_{i=1}^{n} X_i + \sum_{j=1}^{n_2} V_j, \sum_{i=1}^{n} Y_i + \sum_{j=1}^{n_2} V_j\right)$ and thus by Lehmann-Scheffe Theorem, $\widehat{\lambda}$ is the unique $UMVUE$ of λ based on data $\{(x_i, y_i), 1 \leq i \leq n\}$ and $\{v_j, 1 \leq j \leq n_2\}$.

Remark: The result of this theorem certainly implies $\widehat{\lambda}$ is a better estimator $\widehat{\lambda}^* = \sum_{i=1}^{n} y_i/n$ defined in the previous section. It means that the additional data set $\{v_j, 1 \leq j \leq n_2\}$ does improve the accuracy of estimation of λ. How about the performance of $\widehat{p}$? To this end we have to appeal to the limiting distribution of $\widehat{p}$ which is discussed below.

The expression of $\widehat{p}$ in (4.7) can be rewritten as

$$\begin{aligned}
\widehat{p}_n =& \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i + \frac{n_2}{n} \cdot \frac{\sum_{j=1}^{n_2} V_j}{n_2}}{\left(\frac{n_2}{n} + 1\right)\frac{\sum_{i=1}^{n} X_i}{n} - \frac{n_2}{n} \cdot \frac{\sum_{i=1}^{n} Y_i}{n} + \frac{n_2}{n} \cdot \frac{\sum_{j=1}^{n_2} V_j}{n_2}} \\
=& \frac{\bar{Y}_n + \frac{n_2}{n}\bar{V}_{n_2}}{\left(\frac{n_2}{n} + 1\right)\bar{X}_n - \frac{n_2}{n}\bar{Y}_n + \frac{n_2}{n}\bar{V}_{n_2}}
\end{aligned} \tag{4.9}$$

Similarly $\widehat{\lambda}$ in (4.8) can be rewritten as

$$\widehat{\lambda}_n = \frac{\frac{n_2}{n}\left(\bar{X}_n - \bar{Y}_n\right) + \bar{X}_n + \frac{n_2}{n}\bar{V}_{n_2}}{\frac{n_2}{n} + 1} \tag{4.10}$$

In both (4.9) and (4.10) the notations $\widehat{\lambda}_n$ and $\widehat{p}_n$ are used for emphasizing the dependence of $MLEs$ $\widehat{\lambda}$ and $\widehat{p}$ on sample size n.

Further suppose that $n_2 = n_2(n)$ and $n_2(n)/n \to \beta < \infty$ as $n \to \infty$. Under this assumption it is obvious that

$$\begin{aligned}
\lim_{n\to\infty} \widehat{\lambda}_n &= \frac{\beta(\lambda - \lambda p) + \lambda + \beta\lambda p}{\beta + 1} \\
&= \frac{\beta\lambda - \beta\lambda p + \lambda + \beta\lambda p}{\beta + 1} \\
&= \frac{\lambda(\beta + 1)}{\beta + 1} = \lambda \qquad\qquad a.s.
\end{aligned}$$

and

$$\begin{aligned}
\lim_{n\to\infty} \widehat{p}_n &= \frac{\lambda p + \beta\lambda p}{(\beta + 1)\lambda - \beta\lambda p + \beta\lambda p} \\
&= \frac{\lambda p(\beta + 1)}{(\beta + 1)\lambda} = p \qquad\qquad a.s.
\end{aligned}$$

That is, both $\widehat{\lambda}_n$ and $\widehat{p}_n$ are strongly consistent estimators for parameters λ and p respectively. Moreover, assuming $n_2(n)/n - \beta = o\left(n^{-1/2}\right)$, then in order to prove the desired asymptotic normality it suffices to show the normality for both $\widetilde{\lambda}_n$ and $\widetilde{p}_n$ as below

$$\widetilde{\lambda}_n = \frac{\beta\left(\bar{X}_n - \bar{Y}_n\right) + \bar{X}_n + \beta\bar{V}_{n_2}}{\beta+1}$$
$$= \frac{(\beta+1)\,\bar{X}_n - \beta\bar{Y}_n + \beta\bar{V}_{n_2}}{\beta+1}$$

and

$$\widetilde{p}_n = \frac{\bar{Y}_n + \beta\bar{V}_{n_2}}{(\beta+1)\,\bar{X}_n - \beta\bar{Y}_n + \beta\bar{V}_{n_2}}.$$

because $\sqrt{n}\left(\widehat{\lambda}_n - \widetilde{\lambda}_n\right) \to 0$ and $\sqrt{n}\left(\widehat{p}_n - \widetilde{p}_n\right) \to 0$ in probability as $n \to \infty$.

It has been shown in (3.6) that

$$\sqrt{n}\left(\begin{pmatrix}\bar{X}_n \\ \bar{Y}_n\end{pmatrix} - \begin{pmatrix}\lambda \\ \lambda p\end{pmatrix}\right) \to N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\lambda & \lambda p \\ \lambda p & \lambda p\end{pmatrix}\right)$$

in distribution as $n \to \infty$. Define $W_n = \beta\bar{V}_{n_2}$. Clearly

$$\sqrt{n}\left(W_n - \beta\lambda p\right) = \beta\sqrt{\frac{n}{n_2}} \cdot \sqrt{n_2}\left(\bar{V}_{n_2} - \lambda p\right) \to N\left(0, \beta\lambda p\right)$$

in distribution as $n \to \infty$. The independence of $\{(X_i, Y_i), 1 \le i \le n\}$ and $\{V_j, 1 \le j \le n_2\}$ further implies

$$\sqrt{n}\begin{pmatrix}\bar{X}_n - \lambda \\ \bar{Y}_n - \lambda p \\ W_n - \beta\lambda p\end{pmatrix} \to N\left(\begin{pmatrix}0 \\ 0 \\ 0\end{pmatrix}, \begin{pmatrix}\lambda & \lambda p & 0 \\ \lambda p & \lambda p & 0 \\ 0 & 0 & \beta\lambda p\end{pmatrix}\right)$$

in distribution as $n \to \infty$.

Notice that if we let $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ and define

$$h_1(\boldsymbol{\theta}) = \frac{(\beta+1)\theta_1 - \beta\theta_2 + \theta_3}{\beta+1}$$
$$h_2(\boldsymbol{\theta}) = \frac{\theta_2 + \theta_3}{(\beta+1)\theta_1 - \beta\theta_2 + \theta_3}$$

then

$$\widetilde{\lambda}_n = h_1\left(\bar{X}_n, \bar{Y}_n, W_n\right)$$
$$\widetilde{p}_n = h_2\left(\bar{X}_n, \bar{Y}_n, W_n\right).$$

Letting $\boldsymbol{\theta}_0 = (\lambda, \lambda p, \beta\lambda p)'$ yield

$$h_1(\boldsymbol{\theta})\,|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} = \frac{(\beta+1)\lambda - \beta\lambda p + \beta\lambda p}{\beta+1} = \frac{(\beta+1)\lambda}{\beta+1} = \lambda,$$
$$h_2(\boldsymbol{\theta})\,|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} = \frac{\lambda p + \beta\lambda p}{(\beta+1)\lambda - \beta\lambda p + \beta\lambda p} = \frac{\lambda p(\beta+1)}{(\beta+1)\lambda} = p.$$

The delta method gives

$$\sqrt{n}\left(\begin{pmatrix} \widetilde{\lambda}_n \\ \widetilde{p}_n \end{pmatrix} - \begin{pmatrix} \lambda \\ p \end{pmatrix}\right) = \sqrt{n}\begin{pmatrix} h_1\left(\bar{X}_n, \bar{Y}_n, W_n\right) - h_1(\theta_1, \theta_2, \theta_3) \\ h_2\left(\bar{X}_n, \bar{Y}_n, W_n\right) - h_2(\theta_1, \theta_2, \theta_3) \end{pmatrix}$$
$$\to N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, H\Sigma H'\right)$$

where

$$\Sigma = \begin{pmatrix} \lambda & \lambda p & 0 \\ \lambda p & \lambda p & 0 \\ 0 & 0 & \beta\lambda p \end{pmatrix}$$

and $H = (H_{ij})\,, H_{ij} = \partial h_i(\boldsymbol{\theta})/\partial\theta_j, i = 1,2; j = 1,2,3$. It is straightforward that

$$\begin{aligned}
\frac{\partial h_1(\boldsymbol{\theta})}{\partial\theta_1}|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} &= 1, \\
\frac{\partial h_1(\boldsymbol{\theta})}{\partial\theta_2}|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} &= \frac{-\beta}{\beta+1}, \\
\frac{\partial h_1(\boldsymbol{\theta})}{\partial\theta_3}|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} &= \frac{1}{\beta+1}, \\
\frac{\partial h_2(\boldsymbol{\theta})}{\partial\theta_1}|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} &= -\frac{(\beta+1)(\theta_2+\theta_3)}{[(\beta+1)\theta_1-\beta\theta_2+\theta_3]^2}|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} \\
&= -\frac{(\beta+1)(\lambda p+\beta\lambda p)}{[(\beta+1)\lambda-\beta\lambda p+\beta\lambda p]^2} = -\frac{\lambda p(\beta+1)^2}{(\beta+1)^2\lambda^2} = -\frac{p}{\lambda} \\
\frac{\partial h_2(\boldsymbol{\theta})}{\partial\theta_2}|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} &= \frac{[(\beta+1)\theta_1-\beta\theta_2+\theta_3]-(\theta_2+\theta_3)(-\beta)}{[(\beta+1)\theta_1-\beta\theta_2+\theta_3]^2}|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} \\
&= \frac{(\beta+1)\lambda+(\lambda p+\beta\lambda p)\beta}{(\beta+1)^2\lambda^2} = \frac{(\beta+1)\lambda+\beta\lambda p(\beta+1)}{(\beta+1)^2\lambda^2} \\
&= \frac{\lambda(\beta+1)(1+\beta p)}{(\beta+1)^2\lambda^2} = \frac{1+\beta p}{\lambda(\beta+1)} \\
\frac{\partial h_2(\boldsymbol{\theta})}{\partial\theta_3}|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} & \\
&= \frac{[(\beta+1)\theta_1-\beta\theta_2+\theta_3]-(\theta_2+\theta_3)}{[(\beta+1)\theta_1-\beta\theta_2+\theta_3]^2}|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} \\
&= \frac{(\beta+1)\lambda-(\lambda p+\beta\lambda p)}{(\beta+1)^2\lambda^2} = \frac{(\beta+1)\lambda-\lambda p(\beta+1)}{(\beta+1)^2\lambda^2} \\
&= \frac{\lambda(\beta+1)(1-p)}{(\beta+1)^2\lambda^2} = \frac{1-p}{\lambda(\beta+1)}
\end{aligned}$$

From the above we have

$$H\Sigma H'|_{\theta_1=\lambda,\theta_2=\lambda p,\theta_3=\beta\lambda p} = \begin{pmatrix} 1 & -\frac{\beta}{\beta+1} & \frac{1}{\beta+1} \\ -\frac{p}{\lambda} & \frac{1+\beta p}{\lambda(\beta+1)} & \frac{1-p}{\lambda(\beta+1)} \end{pmatrix} \begin{pmatrix} \lambda & \lambda p & 0 \\ \lambda p & \lambda p & 0 \\ 0 & 0 & \beta\lambda p \end{pmatrix} H'$$

$$= \begin{pmatrix} \lambda - \frac{\beta\lambda p}{\beta+1} & \lambda p\left(1 - \frac{\beta}{\beta+1}\right) & \frac{\beta\lambda p}{\beta+1} \\ -p + \frac{\lambda p(1+\beta p)}{\lambda(\beta+1)} & \lambda p\left(-\frac{p}{\lambda} + \frac{1+\beta p}{\lambda(\beta+1)}\right) & \frac{\beta\lambda p(1-p)}{\lambda(\beta+1)} \end{pmatrix} H'$$

$$= \begin{pmatrix} \frac{\lambda(\beta+1-\beta p)}{\beta+1} & \frac{\lambda p}{\beta+1} & \frac{\beta\lambda p}{\beta+1} \\ \frac{-\beta p(1-p)}{\beta+1} & \frac{p(1-p)}{\beta+1} & \frac{\beta p(1-p)}{\beta+1} \end{pmatrix} \begin{pmatrix} 1 & -\frac{p}{\lambda} \\ -\frac{\beta}{\beta+1} & \frac{1+\beta p}{\lambda(\beta+1)} \\ \frac{1}{\beta+1} & \frac{1-p}{\lambda(\beta+1)} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\lambda(\beta+1-\beta p)}{\beta+1} - \frac{\lambda\beta p}{(\beta+1)^2} + \frac{\beta\lambda p}{(\beta+1)^2} & -\frac{p(\beta+1-\beta p)}{\beta+1} + \frac{p(1+\beta p)}{(\beta+1)^2} + \frac{\beta p(1-p)}{(\beta+1)^2} \\ -\frac{\beta p(1-p)}{\beta+1} - \frac{\beta p(1-p)}{(\beta+1)^2} + \frac{\beta p(1-p)}{(\beta+1)^2} & \frac{\beta p^2(1-p)}{\lambda(\beta+1)} + \frac{p(1-p)(1+\beta p)}{\lambda(\beta+1)^2} + \frac{\beta p(1-p)^2}{\lambda(\beta+1)^2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\lambda(\beta+1-\beta p)}{\beta+1} & -\frac{\beta p(1-p)}{\beta+1} \\ -\frac{\beta p(1-p)}{\beta+1} & \frac{p(1-p)}{\lambda} \cdot \frac{1+\beta p}{1+\beta} \end{pmatrix} = \begin{pmatrix} \lambda - \frac{\beta p\lambda}{\beta+1} & \frac{-\beta p(1-p)}{\beta+1} \\ -\frac{\beta p(1-p)}{\beta+1} & \frac{p(1-p)}{\lambda} \cdot \frac{1+\beta p}{1+\beta} \end{pmatrix} \tag{4.11}$$

$$= \begin{pmatrix} \frac{\lambda(\beta+1-\beta p)}{\beta+1} & -\frac{\beta p(1-p)}{\beta+1} \\ -\frac{\beta p(1-p)}{\beta+1} & \frac{p(1-p)(1+\beta p)}{\lambda(1+\beta)} \end{pmatrix}. \tag{4.12}$$

Summarizing the above, we have shown

Theorem 4.3 *Suppose that there exists constant* $\beta < \infty$ *such that* $n_2(n)/n - \beta = o\left(n^{-1/2}\right)$, *then as* $n \to \infty$

$$\sqrt{n}\begin{pmatrix} \widehat{\lambda}_n - \lambda \\ \\ \widehat{p}_n - p \end{pmatrix} \to N\left(\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}, \Sigma_v\right)$$

in distribution, where $\Sigma_v = H\Sigma H'$ *is given by* (4.11) *or* (4.12).

Theorem 4.4 *The estimators based on* $\{(x_i, y_i), 1 \leq i \leq n\}$ *and* $\{v_j, 1 \leq j \leq n_2\}$ *are more efficient than the estimators* $\left(\widehat{\lambda}^*, \widehat{p}^*\right)$ *based on* $\{(x_i, y_i), 1 \leq i \leq n\}$.

Proof. To compare the performance of $\left(\widehat{\lambda}_n, \widehat{p}_n\right)$ with $\left(\widehat{\lambda}_n^*, \widehat{p}_n^*\right)$, recall that

$$\sqrt{n}\begin{pmatrix} \widehat{\lambda}_n^* - \lambda \\ \\ \widehat{p}_n^* - p \end{pmatrix} \to N\left(\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & 0 \\ \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix}\right)$$

in distribution so the difference $\Sigma^* - \Sigma_v = \Sigma^* - H\Sigma H'$ given as

$$\begin{pmatrix} \lambda & 0 \\ \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix} - H\Sigma H' = \begin{pmatrix} \frac{\beta p\lambda}{\beta+1} & \frac{\beta p(1-p)}{\beta+1} \\ \\ \frac{\beta p(1-p)}{\beta+1} & \frac{p(1-p)}{\lambda} \cdot \frac{\beta(1-p)}{\beta+1} \end{pmatrix}$$

is positive semidefinite. Therefore, the estimator $\left(\widehat{\lambda}_n, \widehat{p}_n\right)$ is more efficient than $\left(\widehat{\lambda}_n^*, \widehat{p}_n^*\right)$.■

Remark: To compare the performance of estimators obtained in this section and the previous section, we consider the case of $n_1 = n_2 = m$. In this case $\alpha = \beta$ and so the difference $\Sigma_u - \Sigma_v$ of the associated asymptotic covariance matrices is

$$\Delta = \Sigma_u - \Sigma_v = \begin{pmatrix} \frac{\lambda}{\alpha+1} & 0 \\ 0 & \frac{p(1-p)}{\lambda} \end{pmatrix} - \begin{pmatrix} \frac{\lambda(\alpha+1-\alpha p)}{\alpha+1} & \frac{-\alpha p(1-p)}{\alpha+1} \\ -\frac{\alpha p(1-p)}{\alpha+1} & \frac{p(1-p)(1+\alpha p)}{\lambda(\alpha+1)} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{-\alpha\lambda(1-p)}{\alpha+1} & \frac{\alpha p(1-p)}{\alpha+1} \\ \frac{\alpha p(1-p)}{\alpha+1} & \frac{p(1-p)}{\lambda}\cdot\frac{\alpha(1-p)}{\alpha+1} \end{pmatrix} = \begin{pmatrix} \frac{-\alpha\lambda(1-p)}{\alpha+1} & \frac{\alpha p(1-p)}{\alpha+1} \\ \frac{\alpha p(1-p)}{\alpha+1} & \frac{\alpha p(1-p)^2}{\lambda(\alpha+1)} \end{pmatrix}$$

which is neither positive nor negative semidefinite, so we cannot determine which estimator is more efficient. Nevertheless, note that $\Delta_{11} < 0$ and $\Delta_{22} > 0$, hence we can conclude that the estimator of λ based on $\mathbf{X}$ and $\mathbf{U}$ is more efficient than that based on $\mathbf{X}$ and $\mathbf{V}$; in the contract, the estimator of p based on $\mathbf{X}$ and $\mathbf{V}$ is more efficient than that based on $\mathbf{X}$ and $\mathbf{U}$.

§ 5. Numerical Analysis

A MATLAB simulation is carried out in order to analyze the performance of the estimators with incomplete observations on either layer. Various values of n, α and β are used with different λ and p levels. As the results show, the $MSEs$ of the estimators of both λ and p with extra observations are smaller than those of the estimators with paired observations. Therefore, based on the simulation results, we could conclude that the extra observations should not be ignored in the statistical analysis in Binomial-Poisson hierarchy model research for they provide better estimation of the parameters.

The following tables are formed according to different level of $\alpha(\beta)$ or different combinations of λ and p.

Table 1: Estimates of λ and p with Incomplete Data on First Layer

α	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_1	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
0.3	20	6	1.998545	1.998896	0.300870	0.300870
0.3	50	15	2.001888	2.002169	0.299703	0.299703
0.3	80	24	2.000652	1.999824	0.300290	0.300290
0.3	150	45	1.999625	1.999865	0.300297	0.300297
0.5	20	10	2.000105	2.001315	0.301269	0.301269
0.5	50	25	1.999794	1.999955	0.299667	0.299667
0.5	80	40	2.000571	2.000053	0.299532	0.299532
0.5	150	75	1.999463	1.998962	0.299972	0.299972
1.0	20	20	1.999760	2.001125	0.300838	0.300838
1.0	50	50	1.997352	1.998086	0.299968	0.299968
1.0	80	80	1.999533	2.000026	0.300275	0.300275
1.0	150	150	2.001493	2.001042	0.300403	0.300403

Table 2: Estimates of λ and p with Incomplete Data on First Layer

α	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_1	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
0.3	20	6	9.995710	9.998158	0.899905	0.899905
0.3	50	15	9.997616	10.001468	0.900169	0.900169
0.3	80	24	9.997430	9.996726	0.900123	0.900123
0.3	150	45	9.996972	9.998764	0.900092	0.900092
0.5	20	10	9.999800	9.996973	0.900333	0.900333
0.5	50	25	10.002402	10.002520	0.900159	0.900159
0.5	80	40	9.997187	9.998456	0.900060	0.900060
0.5	150	75	9.998813	9.999091	0.900009	0.900009
1.0	20	20	10.005995	10.001493	0.900141	0.900141
1.0	50	50	9.998370	10.000116	0.900209	0.900209
1.0	80	80	10.000754	10.002525	0.900065	0.900065
1.0	150	150	9.997407	9.998637	0.900074	0.900074

Estimates of λ vary but estimates of p remain the same because the extra observations on the first layer has no influence on estimator of p.

Table 3: Estimates of λ and p with Incomplete Data on Second Layer

β	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_2	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
0.3	20	6	1.998545	1.997479	0.300870	0.301646
0.3	50	15	2.001888	2.001674	0.299703	0.300171
0.3	80	24	2.000652	2.000081	0.300290	0.300401
0.3	150	45	1.999625	1.999348	0.300297	0.300374
0.5	20	10	2.001005	2.001588	0.301269	0.302708
0.5	50	25	1.999794	2.000085	0.299667	0.300499
0.5	80	40	2.000571	2.000864	0.299532	0.300091
0.5	150	75	1.999463	1.999510	0.299972	0.300220
1.0	20	20	1.999760	1.998073	0.300838	0.302927
1.0	50	50	1.997352	1.997291	0.299968	0.301012
1.0	80	80	1.999533	1.999021	0.300275	0.300791
1.0	150	150	2.001493	2.000700	0.300403	0.300488

Table 4: Estimates of λ and p with Incomplete Data on Second Layer

β	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_2	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
0.3	20	6	9.995710	9.997043	0.899905	0.900031
0.3	50	15	9.997616	9.995407	0.900169	0.900192
0.3	80	24	9.997430	9.996043	0.900123	0.900134
0.3	150	45	9.996972	9.996924	0.900092	0.900106
0.5	20	10	9.999800	9.997182	0.900333	0.900461
0.5	50	25	10.002402	10.003222	0.900159	0.900229
0.5	80	40	9.997187	9.998777	0.900060	0.900110
0.5	150	75	9.998813	9.999718	0.900009	0.900039
1.0	20	20	10.005995	10.001973	0.900141	0.900325
1.0	50	50	9.998370	9.998383	0.900209	0.900299
1.0	80	80	10.000754	10.002355	0.900065	0.900138
1.0	150	150	9.997407	9.998929	0.900074	0.900119

With more observations on both layers, both $\widehat{\lambda}$ and $\widehat{p}$ vary.

Table 5: MSE of $\widehat{\lambda}$ and $\widehat{p}$ with Incomplete Data on First Layer

α	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_1	$MSE(\widehat{\lambda}^*)$	$MSE(\widehat{\lambda})$	$MSE(\widehat{p}^*)$	$MSE(\widehat{p})$
0.3	20	6	0.100056	0.076018	0.005294	0.005294
0.3	50	15	0.040346	0.030935	0.002139	0.002139
0.3	80	24	0.025013	0.019162	0.001328	0.001328
0.3	150	45	0.012829	0.009874	0.000710	0.000710
0.5	20	10	0.098287	0.074661	0.005382	0.005382
0.5	50	25	0.039347	0.026298	0.002084	0.002084
0.5	80	40	0.024934	0.016653	0.001322	0.001322
0.5	150	75	0.013505	0.008844	0.000702	0.000702
1.0	20	20	0.099222	0.050017	0.005384	0.005384
1.0	50	50	0.039992	0.019863	0.002120	0.002120
1.0	80	80	0.025118	0.012340	0.001367	0.001367
1.0	150	150	0.013483	0.006693	0.000700	0.000700

Table 6: MSE of $\widehat{\lambda}$ and $\widehat{p}$ with Incomplete Data on First Layer

α	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_1	$MSE(\widehat{\lambda}^*)$	$MSE(\widehat{\lambda})$	$MSE(\widehat{p}^*)$	$MSE(\widehat{p})$
0.3	20	6	0.507474	0.388992	0.000446	0.000446
0.3	50	15	0.201608	0.156334	0.000182	0.000182
0.3	80	24	0.124477	0.094144	0.000112	0.000112
0.3	150	45	0.067481	0.051299	0.000060	0.000060
0.5	20	10	0.499605	0.335644	0.000458	0.000458
0.5	50	25	0.200849	0.133055	0.000180	0.000180
0.5	80	40	0.125845	0.083462	0.000110	0.000110
0.5	150	75	0.067047	0.044404	0.000059	0.000059
1.0	20	20	0.502968	0.249140	0.000443	0.000443
1.0	50	50	0.197279	0.098384	0.000183	0.000183
1.0	80	80	0.125660	0.062977	0.000111	0.000111
1.0	150	150	0.066654	0.033379	0.000060	0.000060

The smaller MSE of $\widehat{\lambda}$ indicates that it is a better estimator than $\widehat{\lambda}^*$.

Table 7: MSE of $\widehat{\lambda}$ and $\widehat{p}$ with Incomplete Data on Second Layer

β	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_2	$MSE(\widehat{\lambda^*})$	$MSE(\widehat{\lambda})$	$MSE(\widehat{p^*})$	$MSE(\widehat{p})$
0.3	20	6	0.100056	0.093389	0.005294	0.004506
0.3	50	15	0.040346	0.037055	0.002139	0.001789
0.3	80	24	0.025013	0.023323	0.001328	0.001102
0.3	150	45	0.012829	0.011823	0.000710	0.000591
0.5	20	10	0.098287	0.092020	0.005382	0.004474
0.5	50	25	0.039347	0.035133	0.002084	0.001587
0.5	80	40	0.024934	0.022306	0.001322	0.001004
0.5	150	75	0.013505	0.012148	0.000702	0.000546
1.0	20	20	0.099222	0.083977	0.005384	0.003578
1.0	50	50	0.039992	0.034244	0.002120	0.001360
1.0	80	80	0.025118	0.021283	0.001367	0.000866
1.0	150	150	0.013483	0.011380	0.000700	0.000451

Table 8: MSE of $\widehat{\lambda}$ and $\widehat{p}$ with Incomplete Data on Second Layer

β	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_2	$MSE(\widehat{\lambda^*})$	$MSE(\widehat{\lambda})$	$MSE(\widehat{p^*})$	$MSE(\widehat{p})$
0.3	20	6	0.507474	0.395920	0.000446	0.000434
0.3	50	15	0.201608	0.160074	0.000182	0.000177
0.3	80	24	0.124477	0.097147	0.000112	0.000109
0.3	150	45	0.067481	0.052916	0.000060	0.000059
0.5	20	10	0.499605	0.347922	0.000458	0.000445
0.5	50	25	0.200849	0.139991	0.000180	0.000174
0.5	80	40	0.125845	0.088266	0.000110	0.000107
0.5	150	75	0.067047	0.046457	0.000059	0.000057
1.0	20	20	0.502968	0.280391	0.000443	0.000418
1.0	50	50	0.197279	0.109489	0.000183	0.000173
1.0	80	80	0.125660	0.070629	0.000111	0.000104
1.0	150	150	0.066654	0.036215	0.000060	0.000057

The smaller $MSEs$ of $\widehat{\lambda}$ and $\widehat{p}$ indicate both of them are better estimators than $\widehat{\lambda^*}$ and $\widehat{p^*}$.

Table 9: Bias of $\widehat{\lambda}$ and $\widehat{p}$ with Incomplete Data on First Layer

α	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_1	Bias($\widehat{\lambda}^*$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
0.3	20	6	-0.001455	-0.001104	0.000870	0.000870
0.3	50	15	0.001888	0.002169	-0.000297	-0.000297
0.3	80	24	0.000652	-0.000176	0.000290	0.000290
0.3	150	45	-0.000375	-0.000135	0.000297	0.000297
0.5	20	10	0.001005	0.001315	0.001269	0.001269
0.5	50	25	-0.000206	-0.000045	-0.000333	-0.000333
0.5	80	40	0.000571	0.000053	-0.000468	-0.000468
0.5	150	75	-0.000537	-0.001038	-0.000028	-0.000028
1.0	20	20	-0.000240	0.001125	0.000838	0.000838
1.0	50	50	-0.002648	-0.001914	-0.000032	-0.000032
1.0	80	80	-0.000468	0.000026	0.000275	0.000275
1.0	150	150	0.001493	0.001042	0.000403	0.000403

Table 10: Bias of $\widehat{\lambda}$ and $\widehat{p}$ with Incomplete Data on First Layer

α	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_1	Bias($\widehat{\lambda}^*$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
0.3	20	6	-0.004290	-0.001842	-0.000095	-0.000095
0.3	50	15	-0.002384	0.001468	0.000169	0.000169
0.3	80	24	-0.002570	-0.003274	0.000123	0.000123
0.3	150	45	-0.003028	-0.001236	0.000092	0.000092
0.5	20	10	-0.000200	-0.003027	0.000333	0.000333
0.5	50	25	0.002402	0.002520	0.000159	0.000159
0.5	80	40	-0.002813	-0.001544	0.000060	0.000060
0.5	150	75	-0.001187	-0.000909	0.000009	0.000009
1.0	20	20	0.005995	0.001493	0.000141	0.000141
1.0	50	50	-0.001630	0.000116	0.000209	0.000209
1.0	80	80	0.000754	0.002525	0.000065	0.000065
1.0	150	150	-0.002593	-0.001363	0.000074	0.000074

The bias has a zero-centered pattern.

Table 11: Bias of $\widehat{\lambda}$ and $\widehat{p}$ with Incomplete Data on Second Layer

β	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_2	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
0.3	20	6	-0.001455	-0.002521	0.000870	0.001646
0.3	50	15	0.001888	0.001674	-0.000297	0.000171
0.3	80	24	0.000652	0.000081	0.000290	0.000401
0.3	150	45	-0.000375	-0.000652	0.000297	0.000374
0.5	20	10	0.001005	0.001588	0.001269	0.002708
0.5	50	25	-0.000206	0.000085	-0.000333	0.000499
0.5	80	40	0.000571	0.000864	-0.000468	0.000091
0.5	150	75	-0.000537	-0.000490	-0.000028	0.000220
1.0	20	20	-0.000240	-0.001928	0.000838	0.002927
1.0	50	50	-0.002648	-0.002709	-0.000032	0.001012
1.0	80	80	-0.000468	-0.000979	0.000275	0.000791
1.0	150	150	0.001493	0.000700	0.000403	0.000488

Table 12: Bias of $\widehat{\lambda}$ and $\widehat{p}$ with Incomplete Data on Second Layer

β	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_2	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
0.3	20	6	-0.004290	-0.002957	-0.000095	0.000031
0.3	50	15	-0.002384	-0.004593	0.000169	0.000192
0.3	80	24	-0.002570	-0.003957	0.000123	0.000134
0.3	150	45	-0.003028	-0.003076	0.000092	0.000106
0.5	20	10	-0.000200	-0.002818	0.000333	0.000461
0.5	50	25	0.002402	0.003222	0.000159	0.000229
0.5	80	40	-0.002813	-0.001223	0.000060	0.000110
0.5	150	75	-0.001187	-0.000282	0.000009	0.000039
1.0	20	20	0.005995	0.001973	0.000141	0.000325
1.0	50	50	-0.001630	-0.0001617	0.000209	0.000299
1.0	80	80	0.000754	0.002355	0.000065	0.000138
1.0	150	150	-0.002593	-0.001071	0.000074	0.000119

The bias has a zero-centered pattern.

Table 13: Estimates for Fixed n and Varying n_1

α	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_1	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
0.3	20	6	1.998545	1.998896	0.300870	0.300870
0.5	20	10	2.000105	2.001315	0.301269	0.301269
1.0	20	20	1.999760	2.001125	0.300838	0.300838
0.3	50	15	2.001888	2.002169	0.299703	0.299703
0.5	50	25	1.999794	1.999955	0.299667	0.299667
1.0	50	50	1.997352	1.998086	0.299968	0.299968
0.3	80	24	2.000652	1.999824	0.300290	0.300290
0.5	80	40	2.000571	2.000053	0.299532	0.299532
1.0	80	80	1.999533	2.000026	0.300275	0.300275
0.3	150	45	1.999625	1.999865	0.300297	0.300297
0.5	150	75	1.999463	1.998962	0.299972	0.299972
1.0	150	150	2.001493	2.001042	0.300403	0.300403

Table 14: Estimates for Fixed n and Varying n_1

α	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_1	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
0.3	20	6	9.995710	9.998158	0.899905	0.899905
0.5	20	10	9.999800	9.996973	0.900333	0.900333
1.0	20	20	10.005995	10.001493	0.900141	0.900141
0.3	50	15	9.997616	10.001468	0.900169	0.900169
0.5	50	25	10.002402	10.002520	0.900159	0.900159
1.0	50	50	9.998370	10.000116	0.900209	0.900209
0.3	80	24	9.997430	9.996726	0.900123	0.900123
0.5	80	40	9.997187	9.998456	0.900060	0.900060
1.0	80	80	10.000754	10.002525	0.900065	0.900065
0.3	150	45	9.996972	9.998764	0.900092	0.900092
0.5	150	75	9.998813	9.999091	0.900009	0.900009
1.0	150	150	9.997407	9.998637	0.900074	0.900074

When n is fixed, n_1 increases, the estimates of λ vary but estimates of p remain the same because the extra observations on the first layer has no influence on the estimator of p.

Table 15: Estimates for Fixed n and Varying n_2

β	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_2	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
0.3	20	6	1.998545	1.997479	0.300870	0.301646
0.5	20	10	2.001005	2.001588	0.301269	0.302708
1.0	20	20	1.999760	1.998073	0.300838	0.302927
0.3	50	15	2.001888	2.001674	0.299703	0.300171
0.5	50	25	1.999794	2.000085	0.299667	0.300499
1.0	50	50	1.997352	1.997291	0.299968	0.301012
0.3	80	24	2.000652	2.000081	0.300290	0.300401
0.5	80	40	2.000571	2.000864	0.299532	0.300091
1.0	80	80	1.999533	1.999021	0.300275	0.300791
0.3	150	45	1.999625	1.999348	0.300297	0.300374
0.5	150	75	1.999463	1.999510	0.299972	0.300220
1.0	150	150	2.001493	2.000700	0.300403	0.300488

Table 16: Estimates for Fixed n and Varying n_2

β	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_2	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
0.3	20	6	9.995710	9.997043	0.899905	0.900031
0.5	20	10	9.999800	9.997182	0.900333	0.900461
1.0	20	20	10.005995	10.001973	0.900141	0.900325
0.3	50	15	9.997616	9.995407	0.900169	0.900192
0.5	50	25	10.002402	10.003222	0.900159	0.900229
1.0	50	50	9.998370	9.998383	0.900209	0.900299
0.3	80	24	9.997430	9.996043	0.900123	0.900134
0.5	80	40	9.997187	9.998777	0.900060	0.900110
1.0	80	80	10.000754	10.002355	0.900065	0.900138
0.3	150	45	9.996972	9.996924	0.900092	0.900106
0.5	150	75	9.998813	9.999718	0.900009	0.900039
1.0	150	150	9.997407	9.998929	0.900074	0.900119

When n is fixed, n_2 increases, the estimates of both λ and p vary.

Table 17: MSE of $\widehat{\lambda}$ and $\widehat{p}$ for Fixed n and Varying n_1

α	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_1	$MSE(\widehat{\lambda^*})$	$MSE(\widehat{\lambda})$	$MSE(\widehat{p}^*)$	$MSE(\widehat{p})$
0.3	20	6	0.100056	0.076018	0.005294	0.005294
0.5	20	10	0.098287	0.074661	0.005382	0.005382
1.0	20	20	0.099222	0.050017	0.005384	0.005384
0.3	50	15	0.040346	0.030935	0.002139	0.002139
0.5	50	25	0.039347	0.026298	0.002084	0.002084
1.0	50	50	0.039992	0.019863	0.002120	0.002120
0.3	80	24	0.025013	0.019162	0.001328	0.001328
0.5	80	40	0.024934	0.016653	0.001322	0.001322
1.0	80	80	0.025118	0.012340	0.001367	0.001367
0.3	150	45	0.012829	0.009874	0.000710	0.000710
0.5	150	75	0.013505	0.008844	0.000702	0.000702
1.0	150	150	0.013483	0.006693	0.000700	0.000700

Table 18: MSE of $\widehat{\lambda}$ and $\widehat{p}$ for Fixed n and Varying n_1

α	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_1	$MSE(\widehat{\lambda^*})$	$MSE(\widehat{\lambda})$	$MSE(\widehat{p}^*)$	$MSE(\widehat{p})$
0.3	20	6	0.507474	0.388992	0.000446	0.000446
0.5	20	10	0.499605	0.335644	0.000458	0.000458
1.0	20	20	0.502968	0.249140	0.000443	0.000443
0.3	50	15	0.201608	0.156334	0.000182	0.000182
0.5	50	25	0.200849	0.133055	0.000180	0.000180
1.0	50	50	0.197279	0.098384	0.000183	0.000183
0.3	80	24	0.124477	0.094144	0.000112	0.000112
0.5	80	40	0.125845	0.083462	0.000110	0.000110
1.0	80	80	0.125660	0.062977	0.000111	0.000111
0.3	150	45	0.067481	0.051299	0.000060	0.000060
0.5	150	75	0.067047	0.044404	0.000059	0.000059
1.0	150	150	0.066654	0.033379	0.000060	0.000060

When n is fixed, n_1 increases, the $MSE(\widehat{\lambda})$ decreases but $MSE(\widehat{p})$ remains the same.

Table 19: MSE of $\widehat{\lambda}$ and $\widehat{p}$ for Fixed n and Varying n_2

β	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_2	$MSE(\widehat{\lambda}^*)$	$MSE(\widehat{\lambda})$	$MSE(\widehat{p}^*)$	$MSE(\widehat{p})$
0.3	20	6	0.100056	0.093389	0.005294	0.004506
0.5	20	10	0.098287	0.092020	0.005382	0.004474
1.0	20	20	0.099222	0.083977	0.005384	0.003578
0.3	50	15	0.040346	0.037055	0.002139	0.001789
0.5	50	25	0.039347	0.035133	0.002084	0.001587
1.0	50	50	0.039992	0.034244	0.002120	0.001360
0.3	80	24	0.025013	0.023323	0.001328	0.001102
0.5	80	40	0.024934	0.022306	0.001322	0.001004
1.0	80	80	0.025118	0.021283	0.001367	0.000866
0.3	150	45	0.012829	0.011823	0.000710	0.000591
0.5	150	75	0.013505	0.012148	0.000702	0.000546
1.0	150	150	0.013483	0.011380	0.000700	0.000451

Table 20: MSE of $\widehat{\lambda}$ and $\widehat{p}$ for Fixed n and Varying n_2

β	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_2	$MSE(\widehat{\lambda}^*)$	$MSE(\widehat{\lambda})$	$MSE(\widehat{p}^*)$	$MSE(\widehat{p})$
0.3	20	6	0.507474	0.395920	0.000446	0.000434
0.5	20	10	0.499605	0.347922	0.000458	0.000445
1.0	20	20	0.502968	0.280391	0.000443	0.000418
0.3	50	15	0.201608	0.160074	0.000182	0.000177
0.5	50	25	0.200849	0.139991	0.000180	0.000174
1.0	50	50	0.197279	0.109489	0.000183	0.000173
0.3	80	24	0.124477	0.097147	0.000112	0.000109
0.5	80	40	0.125845	0.088266	0.000110	0.000107
1.0	80	80	0.125660	0.070629	0.000111	0.000104
0.3	150	45	0.067481	0.052916	0.000060	0.000059
0.5	150	75	0.067047	0.046457	0.000059	0.000057
1.0	150	150	0.066654	0.036215	0.000060	0.000057

When n is fixed, n_2 increases, both $MSE(\widehat{\lambda})$ and $MSE(\widehat{p})$ decrease.

Table 21: Bias of $\widehat{\lambda}$ and $\widehat{p}$ for Fixed n and Varying n_1

α	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_1	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
0.3	20	6	-0.001455	-0.001104	0.000870	0.000870
0.5	20	10	0.001005	0.001315	0.001269	0.001269
1.0	20	20	-0.000240	0.001125	0.000838	0.000838
0.3	50	15	0.001888	0.002169	-0.000297	-0.000297
0.5	50	25	-0.000206	-0.000045	-0.000333	-0.000333
1.0	50	50	-0.002648	-0.001914	-0.000032	-0.000032
0.3	80	24	0.000652	-0.000176	0.000290	0.000290
0.5	80	40	0.000571	0.000053	-0.000468	-0.000468
1.0	80	80	-0.000468	0.000026	0.000275	0.000275
0.3	150	45	-0.000375	-0.000135	0.000297	0.000297
0.5	150	75	-0.000537	-0.001038	-0.000028	-0.000028
1.0	150	150	0.001493	0.001042	0.000403	0.000403

Table 22: Bias of $\widehat{\lambda}$ and $\widehat{p}$ for Fixed n and Varying n_1

α	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_1	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
0.3	20	6	-0.004290	-0.001842	-0.000095	-0.000095
0.5	20	10	-0.000200	-0.003027	0.000333	0.000333
1.0	20	20	0.005995	0.001493	0.000141	0.000141
0.3	50	15	-0.002384	0.001468	0.000169	0.000169
0.5	50	25	0.002402	0.002520	0.000159	0.000159
1.0	50	50	-0.001630	0.000116	0.000209	0.000209
0.3	80	24	-0.002570	-0.003274	0.000123	0.000123
0.5	80	40	-0.002813	-0.001544	0.000060	0.000060
1.0	80	80	0.000754	0.002525	0.000065	0.000065
0.3	150	45	-0.003028	-0.001236	0.000092	0.000092
0.5	150	75	-0.001187	-0.000909	0.000009	0.000009
1.0	150	150	-0.002593	-0.001363	0.000074	0.000074

The bias has a zero-center pattern.

Table 23: Bias of $\widehat{\lambda}$ and $\widehat{p}$ for Fixed n and Varying n_2

β	Sample Size		$\lambda = 2$		$p = 0.3$	
	n	n_2	Bias($\widehat{\lambda}^*$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
0.3	20	6	-0.001455	-0.002521	0.000870	0.001646
0.5	20	10	0.001005	0.001588	0.001269	0.002708
1.0	20	20	-0.000240	-0.001928	0.000838	0.002927
0.3	50	15	0.001888	0.001674	-0.000297	0.000171
0.5	50	25	-0.000206	0.000085	-0.000333	0.000499
1.0	50	50	-0.002648	-0.002709	-0.000032	0.001012
0.3	80	24	0.000652	0.000081	0.000290	0.000401
0.5	80	40	0.000571	0.000864	-0.000468	0.000091
1.0	80	80	-0.000468	-0.000979	0.000275	0.000791
0.3	150	45	-0.000375	-0.000652	0.000297	0.000374
0.5	150	75	-0.000537	-0.000490	-0.000028	0.000220
1.0	150	150	0.001493	0.000700	0.000403	0.000488

Table 24: Bias of $\widehat{\lambda}$ and $\widehat{p}$ for Fixed n and Varying n_2

β	Sample Size		$\lambda = 10$		$p = 0.9$	
	n	n_2	Bias($\widehat{\lambda}^*$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
0.3	20	6	-0.004290	-0.002957	-0.000095	0.000031
0.5	20	10	-0.000200	-0.002818	0.000333	0.000461
1.0	20	20	0.005995	0.001973	0.000141	0.000325
0.3	50	15	-0.002384	-0.004593	0.000169	0.000192
0.5	50	25	0.002402	0.003222	0.000159	0.000229
1.0	50	50	-0.001630	-0.0001617	0.000209	0.000299
0.3	80	24	-0.002570	-0.003957	0.000123	0.000134
0.5	80	40	-0.002813	-0.001223	0.000060	0.000110
1.0	80	80	0.000754	0.002355	0.000065	0.000138
0.3	150	45	-0.003028	-0.003076	0.000092	0.000106
0.5	150	75	-0.001187	-0.000282	0.000009	0.000039
1.0	150	150	-0.002593	-0.001071	0.000074	0.000119

The bias has a zero-center pattern.

Table 25: Estimates with Varying λ and p

$\alpha = 0.3$					
Sample Size		$\lambda = 2$		$p = 0.3$	
n	n_1	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
20	6	1.998545	1.998896	0.300870	0.300870
50	15	2.001888	2.002169	0.299703	0.299703
80	24	2.000652	1.999824	0.300290	0.300290
150	45	1.999625	1.999865	0.300297	0.300297
Sample Size		$\lambda = 10$		$p = 0.3$	
n	n_1	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
20	6	9.997700	9.994450	0.299884	0.299884
50	15	9.994126	9.999712	0.300140	0.300140
80	24	9.996091	9.993202	0.299900	0.299900
150	45	10.002533	10.000890	0.299979	0.299979
Sample Size		$\lambda = 2$		$p = 0.9$	
n	n_1	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
20	6	1.996655	1.998454	0.899588	0.899588
50	15	1.999264	1.999869	0.900561	0.900561
80	24	1.998029	1.998967	0.899956	0.899956
150	45	2.000536	2.001075	0.900007	0.900007
Sample Size		$\lambda = 10$		$p = 0.9$	
n	n_1	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
20	6	9.995710	9.998158	0.899905	0.899905
50	15	9.997616	10.001468	0.900169	0.900169
80	24	9.997430	9.996726	0.900123	0.900123
150	45	9.996972	9.998764	0.900092	0.900092

With fixed $\alpha = 0.3$ simulation is done with various values of λ and p.

Table 26: MSE with Varying λ and p

$\alpha = 0.3$					
Sample Size		$\lambda = 2$		$p = 0.3$	
n	n_1	MSE($\widehat{\lambda}^*$)	MSE($\widehat{\lambda}$)	MSE($\widehat{p}^*$)	MSE($\widehat{p}$)
20	6	0.100056	0.076018	0.005294	0.005294
50	15	0.040346	0.030935	0.002139	0.002139
80	24	0.025013	0.019162	0.001328	0.001328
150	45	0.012829	0.009874	0.000710	0.000710
Sample Size		$\lambda = 10$		$p = 0.3$	
n	n_1	MSE($\widehat{\lambda}^*$)	MSE($\widehat{\lambda}$)	MSE($\widehat{p}^*$)	MSE($\widehat{p}$)
20	6	0.498230	0.384263	0.001040	0.001040
50	15	0.200464	0.155275	0.000413	0.000413
80	24	0.126671	0.097146	0.000263	0.000263
150	45	0.067060	0.051255	0.000142	0.000142
Sample Size		$\lambda = 2$		$p = 0.9$	
n	n_1	MSE($\widehat{\lambda}^*$)	MSE($\widehat{\lambda}$)	MSE($\widehat{p}^*$)	MSE($\widehat{p}$)
20	6	0.097842	0.076381	0.002254	0.002254
50	15	0.039609	0.030555	0.000919	0.000919
80	24	0.025382	0.019500	0.000568	0.000568
150	45	0.013216	0.010288	0.000303	0.000303
Sample Size		$\lambda = 10$		$p = 0.9$	
n	n_1	MSE($\widehat{\lambda}^*$)	MSE($\widehat{\lambda}$)	MSE($\widehat{p}^*$)	MSE($\widehat{p}$)
20	6	0.507474	0.388992	0.000446	0.000446
50	15	0.201608	0.156334	0.000182	0.000182
80	24	0.124477	0.094144	0.000112	0.000112
150	45	0.067481	0.051299	0.000060	0.000060

With fixed $\alpha = 0.3$ and various values of λ and p, increasing n makes the MSE of $\widehat{\lambda}$ decreasing.

Table 27: Bias with Varying λ and p

$\alpha = 0.3$					
Sample Size		$\lambda = 2$		$p = 0.3$	
n	n_1	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
20	6	-0.001455	-0.001104	0.000870	0.000870
50	15	0.001888	0.002169	-0.000297	-0.000297
80	24	0.000652	-0.000176	0.000290	0.000290
150	45	-0.000375	-0.000135	0.000297	0.000297
Sample Size		$\lambda = 10$		$p = 0.3$	
n	n_1	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
20	6	-0.002300	-0.005550	-0.000116	-0.000116
50	15	-0.005874	-0.000288	0.000140	0.000140
80	24	-0.003909	-0.006798	-.000100	-0.000100
150	45	0.002533	0.000890	-0.000021	-0.000021
Sample Size		$\lambda = 2$		$p = 0.9$	
n	n_1	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
20	6	-0.003345	-0.001546	-0.000412	-0.000412
50	15	-0.000736	-0.000131	0.000561	0.000561
80	24	-0.001971	-0.001033	-0.000044	-0.000044
150	45	0.000536	0.001075	0.000007	0.000007
Sample Size		$\lambda = 10$		$p = 0.9$	
n	n_1	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
20	6	-0.004290	-0.001842	-0.000095	-0.000095
50	15	-0.002384	0.001468	0.000169	0.000169
80	24	-0.002570	-0.003274	0.000123	0.000123
150	45	-0.003028	-0.001236	0.000092	0.000092

With fixed $\alpha = 0.3$ and various values of λ and p, increasing n makes the bias approach to zero.

Table 28: Estimates with Varying λ and p

$\beta = 0.3$					
Sample Size		$\lambda = 2$		$p = 0.3$	
n	n_2	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
20	6	1.998545	1.997479	0.300870	0.301646
50	15	2.001888	2.001674	0.299703	0.300171
80	24	2.000652	2.000081	0.300290	0.300401
150	45	1.999625	1.999348	0.300297	0.300374
Sample Size		$\lambda = 10$		$p = 0.3$	
n	n_2	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
20	6	9.997700	9.996805	0.299884	0.300069
50	15	9.994126	9.990586	0.300140	0.299988
80	24	9.996091	9.996755	0.299900	0.300013
150	45	10.002533	10.002550	0.299979	0.300012
Sample Size		$\lambda = 2$		$p = 0.9$	
n	n_2	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
20	6	1.996655	1.996713	0.899588	0.900148
50	15	1.999264	1.999047	0.900561	0.900751
80	24	1.998029	1.999285	0.899956	0.900160
150	45	2.000536	2.000038	0.900007	0.900051
Sample Size		$\lambda = 10$		$p = 0.9$	
n	n_2	$\widehat{\lambda}^*$	$\widehat{\lambda}$	$\widehat{p}^*$	$\widehat{p}$
20	6	9.995710	9.997043	0.899905	0.900031
50	15	9.997616	9.995407	0.900169	0.900192
80	24	9.997430	9.996043	0.900123	0.900134
150	45	9.996972	9.996924	0.900092	0.900106

With fixed $\beta = 0.3$ simulation is done with various values of λ and p.

Table 29: MSE with Varying λ and p

$\beta = 0.3$					
Sample Size		$\lambda = 2$		$p = 0.3$	
n	n_2	MSE($\widehat{\lambda^*}$)	MSE($\widehat{\lambda}$)	MSE($\widehat{p^*}$)	MSE($\widehat{p}$)
20	6	0.100056	0.093389	0.005294	0.004506
50	15	0.040346	0.037055	0.002139	0.001789
80	24	0.025013	0.023323	0.001328	0.001102
150	45	0.012829	0.011823	0.000710	0.000591
Sample Size		$\lambda = 10$		$p = 0.3$	
n	n_2	MSE($\widehat{\lambda^*}$)	MSE($\widehat{\lambda}$)	MSE($\widehat{p^*}$)	MSE($\widehat{p}$)
20	6	0.498230	0.462651	0.001040	0.000871
50	15	0.200464	0.186252	0.000413	0.000348
80	24	0.126671	0.117099	0.000263	0.000218
150	45	0.067060	0.062501	0.000142	0.000120
Sample Size		$\lambda = 2$		$p = 0.9$	
n	n_2	MSE($\widehat{\lambda^*}$)	MSE($\widehat{\lambda}$)	MSE($\widehat{p^*}$)	MSE($\widehat{p}$)
20	6	0.097842	0.077782	0.002254	0.002165
50	15	0.039609	0.031557	0.000919	0.000893
80	24	0.025382	0.019888	0.000568	0.000552
150	45	0.013216	0.010469	0.000303	0.000296
Sample Size		$\lambda = 10$		$p = 0.9$	
n	n_2	MSE($\widehat{\lambda^*}$)	MSE($\widehat{\lambda}$)	MSE($\widehat{p^*}$)	MSE($\widehat{p}$)
20	6	0.507474	0.395920	0.000446	0.000434
50	15	0.201608	0.160074	0.000182	0.000177
80	24	0.124477	0.097147	0.000112	0.000109
150	45	0.067481	0.052916	0.000060	0.000059

With fixed $\beta = 0.3$ and various values of λ and p, increasing n makes the MSE of $\widehat{\lambda}$ decrease.

Table 30: Bias with Varying λ and p

$\beta = 0.3$					
Sample Size		$\lambda = 2$		$p = 0.3$	
n	n_2	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
20	6	-0.001455	-0.002521	0.000870	0.001646
50	15	0.001888	0.001674	-0.000297	0.000171
80	24	0.000652	0.000081	0.000290	0.000401
150	45	-0.000375	-0.000652	0.000297	0.000374
Sample Size		$\lambda = 10$		$p = 0.3$	
n	n_2	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
20	6	-0.002300	-0.003195	-0.000116	0.000069
50	15	-0.005874	-0.009414	0.000140	-0.000012
80	24	-0.003909	-0.003245	-.000100	0.000013
150	45	0.002533	0.002550	-0.000021	0.000012
Sample Size		$\lambda = 2$		$p = 0.9$	
n	n_2	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
20	6	-0.003345	-0.003287	-0.000412	0.000148
50	15	-0.000736	-0.000953	0.000561	0.000751
80	24	-0.001971	-0.000715	-0.000044	0.000160
150	45	0.000536	0.000038	0.000007	0.000051
Sample Size		$\lambda = 10$		$p = 0.9$	
n	n_2	Bias($\widehat{\lambda^*}$)	Bias($\widehat{\lambda}$)	Bias($\widehat{p}^*$)	Bias($\widehat{p}$)
20	6	-0.004290	-0.002957	-0.000095	0.000031
50	15	-0.002384	0.004593	0.000169	0.000192
80	24	-0.002570	-0.003957	0.000123	0.000134
150	45	-0.003028	-0.003076	0.000092	0.000106

With fixed $\beta = 0.3$ and various values of λ and p, increase n makes the bias approach to zero.

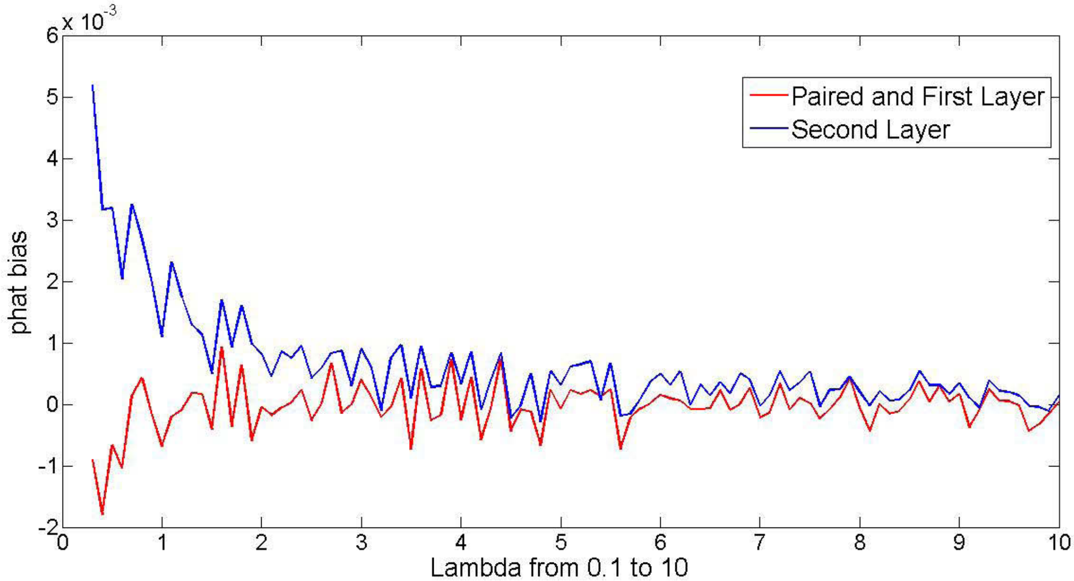

Figure 1: Bias of $\widehat{p}$ on λ: when λ increases, bias of $\widehat{p}$ decreases.

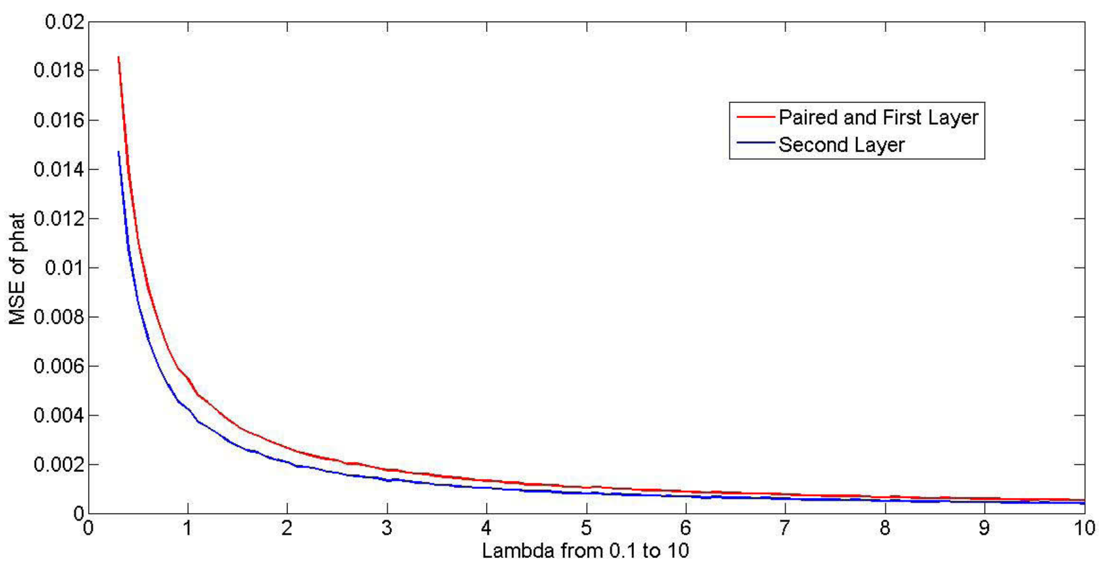

Figure 2: MSE of $\widehat{p}$ on λ: when λ increase the MSE decreases.

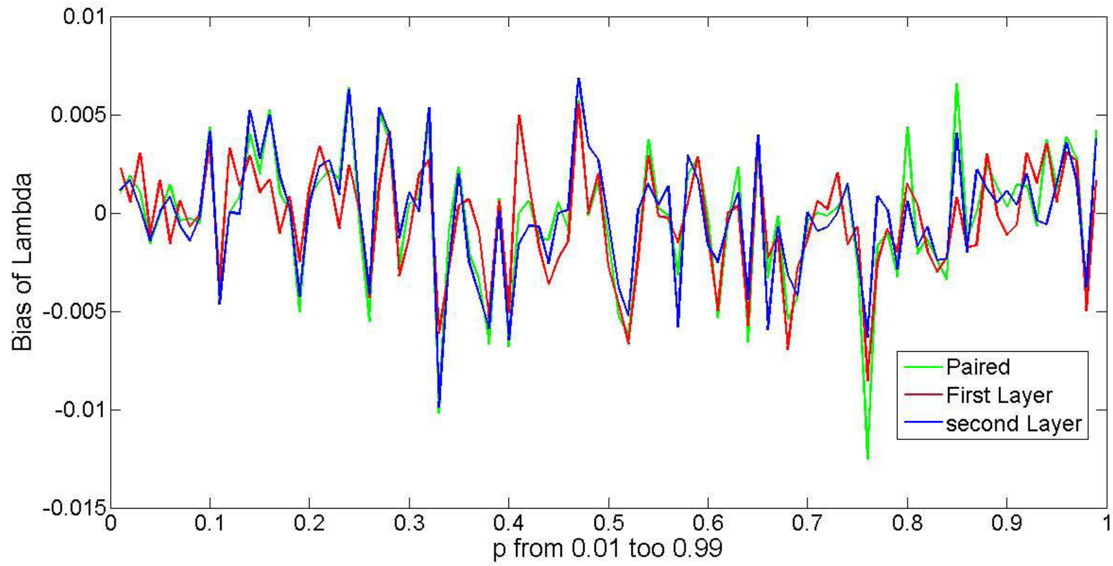

Figure 3: Bias of λ: zero-centered pattern

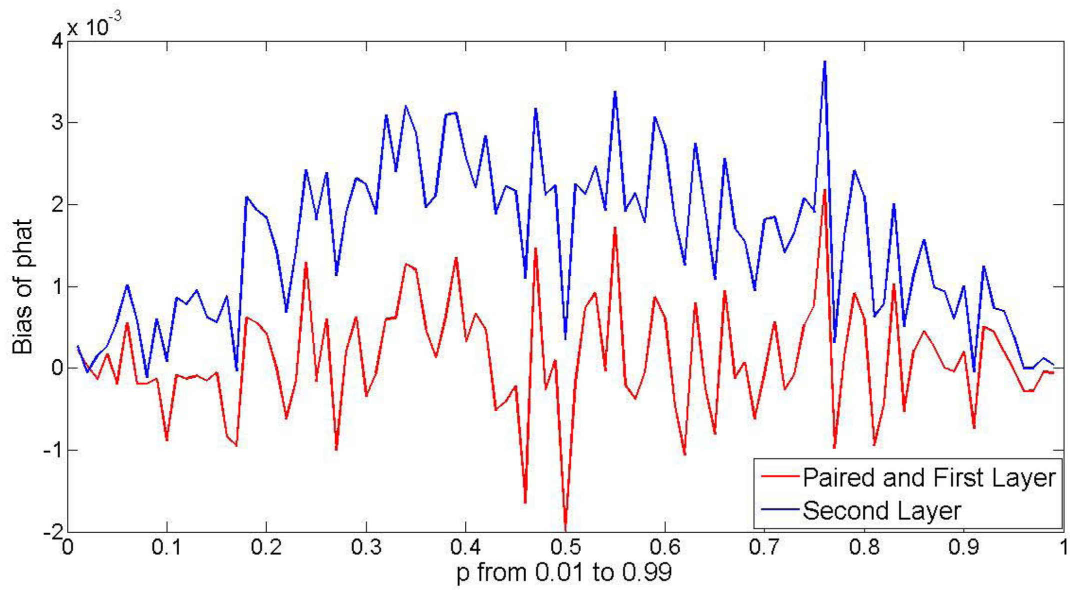

Figure 4: Bias of $\widehat{p}$ on p: zero-centered pattern

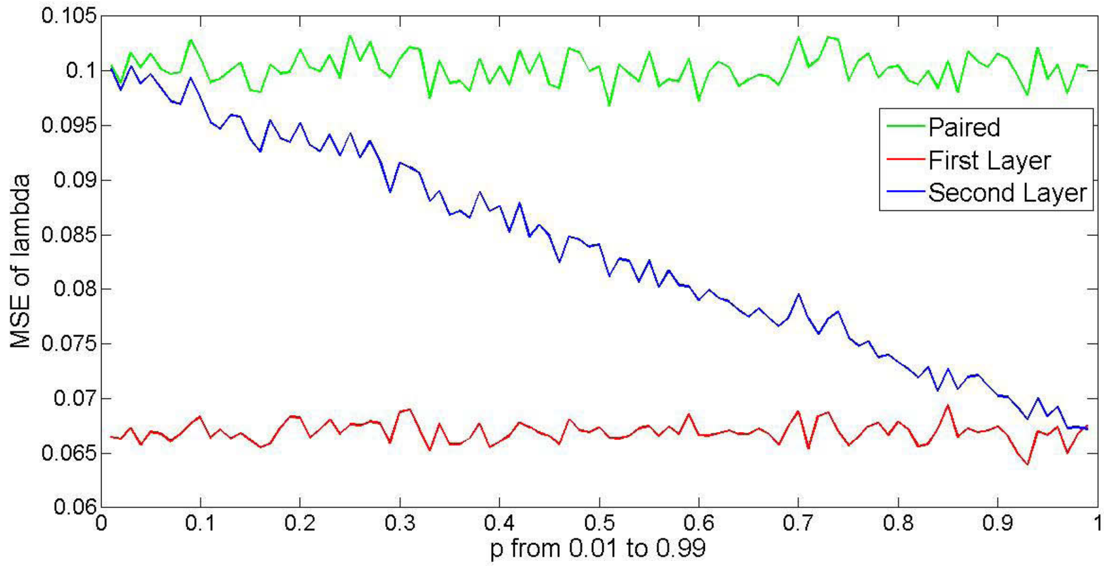

Figure 5: MSE of $\widehat{\lambda}$ on p: when p increases, the MSE of $\widehat{\lambda}$ on second layer decreases.

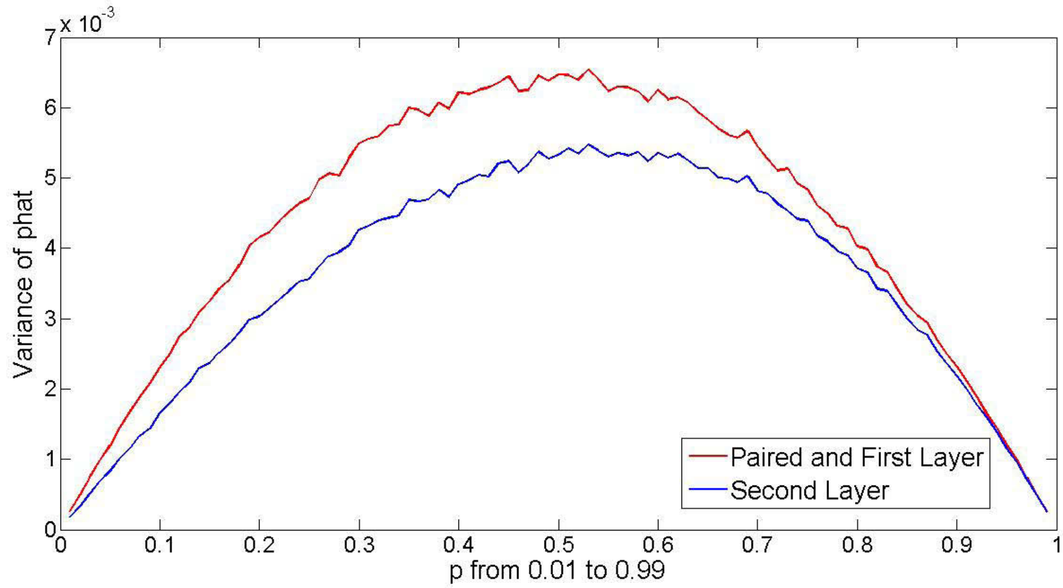

Figure 6: Variance of $\widehat{p}$ on p: when p increases, the variance of $\widehat{p}$ has an approximately parabola shape.

§ 6. Conclusion

According to the analytical and numerical results obtained above, we could make a conclusion that with more observations on either first layer or second layer of the model, the $MSEs$ of the estimators are smaller than those from the paired observations, which indicate that the estimators with incomplete data are more efficient. Therefore, in lab research it is better we keep the unpaired data in the analysis procedure and use the model established above to obtain better estimation.

REFERENCES

Cacoullos, T. and Papageorgiou, H., 1982. Bivariate Negative Binomial-Poisson and Negative Binomial-Bernoulli Models with an Application to Accident Data. Statistics and Probability: Essays in Honor of C. R. Rao, 155 - 168.

Faddy, M. J. and Smith, D. M., 2005. Modeling the Dependence between the Number of Trials and the Success Probability in Binary Trials. Biometrics, 61, 1112 - 1114.

Johnson, N. L. and Kotz, S., 1969. Discrete Distributions. Wiley-Interscience. New York.

Johnson, Norman L., Samuel Kotz, and Adrienne W. Kemp 1992. Univariate Discrete Distributions, 2nd edition, New York: John Wiley and Sons.

Jun Zhu, Jens C. Eickhoff, Mark S. Kaiser, 2003. Modeling the Dependence between Number of Trials and Success Probability in Beta-Binomial-Poisson Mixture Distributions. Biometrics, Vol. 59, No. 4, 955 - 961.

Katti, S. K. and Gurland, J., 1962. Some methods of estimation for the Poisson Binomial Distribution. Biometrics 18, 42, 51.

McGuire, J. U., Brindley, T. A., and Bancroft, T. A., 1957. The distribution of European corn borer larvae Pyrausta Nubilalis (Hbn) in field corn. Biometrics No. 13, 65-78.

Mises, R.V., 1921. Uber Die Wahrscheinlichkeit Seltener Ereignisse. Z. Angew. Math. Mech. Vol. 1, 121-124.

Ocerin, J.M.C., Perez, J.D., 2002. Point and interval estimators in a binomial-Poisson compound distribution. Statistical Papers, 43, 285 - 290.

Ouyang, Z., Schreuder, H. T., Max, T. and Williams, M. 1993. Poisson-Poisson and Binomial-Poisson Sampling in Forestry. Survey Methodology, 19, 115 - 121.

Robert Shumway, John Gurland, 1960. Fitting the Poisson Binomial distribution. Biometrics, Vol. 16, No. 4, 522 - 533.

Shkedy, Ziv, Molenberghs, Geert, Craenendonck, Hansfried Van, Steckler, Thomas and Bijnens, Luc, 2005. A Hierarchical Binomial-Poisson Model for the Analysis of a Crossover Design for Correlated Binary Data when the Number of Trials is Dose-dependent. Journal of Biopharmaceutical Statistics, 15, 225 - 239.

Sprott, D. A., 1958. The Method of Maximum Likelihood Applied to the Poisson Binomial Distribution. Biometrics, Vol. 14, No. 1, 97 - 106.