

## Preliminary Test Estimators and Phi-divergence Measures in Pooling Binomial Data

Nirian Martin  
Department of Statistics  
Carlos III University of Madrid  
28038 Getafe (Madrid), Spain  
nirian.martin@uc3m.es

Leandro Pardo  
Department of Statistics and O.R.  
Complutense University of Madrid  
28040 Madrid, Spain  
lpardo@mat.ucm.es

### Abstract

Two independent random samples are drawn from two Binomial populations with parameters  $\theta_1$  and  $\theta_2$  respectively. Ahmed (1991) considered a preliminary test estimator based on maximum likelihood estimator for estimating  $\theta_1$  when it is suspected that  $\theta_1 = \theta_2$ . In this paper we combine minimum phi-divergence estimators as well as phi-divergence test statistics in order to define preliminary phi-divergence test estimators. These new estimators are compared with the classical estimator as well as the pooled estimator.

**Keywords and phrases:** Pooling Binomial data, Minimum phi-divergence estimator, Phi-divergence statistics, Preliminary tests estimators.

### 1. Introduction

Pooling data is an old classical problem that has been studied for many authors in different context. If we have a random sample of size  $n$  from a random variable  $X$  and a random sample of size  $m$  from a random variable  $Y$  (the distributions of  $X$  and  $Y$  belong to the same family of probability distributions) how shall estimate the mean,  $E[X]$ , of the random variable  $X$ ? Shall we merely take the sample mean of the random sample  $X$ , or shall we attempt to combine the means of the two random samples in order to improve our estimation of  $E[X]$  in some sense? From a historical point of view Mosteller (1948) presented for the first time the analysis of pooling univariate normal data with variance known. Kale and Bancroft (1967) extended the study to discrete data by using proper transformations. For unknown variances the problem was studied by Han and Bancroft (1968). The problem for multivariate normal data was studied by Han and Bancroft (1970) with known covariance matrix and by Ahmed (1992) with unknown covariance matrix. Other interesting papers in pooling data are Ahmed (1993, 1997),

Ahmed et al. (1989, 1997, 1999), Mehta and Srinivasan (1971), Raghunandan (1978) and references therein.

It is advantageous to utilize a linear combination of the two sample means ponderated by the sample size of them if  $E[X]$  coincides with  $E[Y]$ , i.e., to use the restricted estimator. In many situations it is not clear if  $E[X] = E[Y]$ . In order to tackle this uncertainty we can perform a preliminary test and then choose between a restricted and an unrestricted estimator. This line of thought was first proposed by Bancroft (1944). For a wide study of preliminary test estimators in different statistical problems see Saleh (2006) and references therein.

In this paper we focus on the problem of pooling proportions of two independent random samples taken from two possibly identical binomial distributions (see Ahmed (1991)). This author considered a preliminary test based on restricted maximum likelihood estimator and classical Pearson's test statistic. In this paper instead of considering the restricted maximum likelihood estimator we shall consider the restricted minimum phi-divergence estimator and instead of Pearson's test statistics a family of phi-divergence test statistics. Therefore in this paper we introduce a family of preliminary test estimators for the problem of pooling binomial data that contains as a particular case the preliminary test estimator considered by Ahmed.

The behaviour of phi-divergence measures in the definition of preliminary test estimators can be seen in Menéndez et al (2008, 2011), Pardo and Martin (2011) and references therein.

Section 2 is devoted to introduce the family of preliminary test estimators considered in this paper and in Section 3 we obtain some asymptotic distributional results that are necessary in the next Section. Finally, Section 4 is devoted to get the asymptotic bias as well as the asymptotic mean squared errors of the family of estimators introduced in the paper.

## 2. Estimation strategies in pooling binomial data based on phi-divergence measures

Let  $X_1, \dots, X_n$  and  $Y_1, \dots, Y_m$  be two independent random samples of sizes  $n$  and  $m$  from two Bernoulli random variables of parameters  $\theta_1$  and  $\theta_2$ , respectively. The main problem in which we are interested is in the estimation of  $\theta_1$  when we suspect that  $\theta_1 = \theta_2$ . Maximum likelihood estimator (MLE) of  $\theta_1$  (unrestricted MLE) based on  $X_1, \dots, X_n$  is given by  $\hat{\theta}_1 = y_1 / n$  ( $y_1 \equiv$  number of successes associated with the random sample  $X_1, \dots, X_n$ ) and the MLE of  $\theta_1$  (restricted maximum likelihood estimator) under the assumption that  $\theta_1 = \theta_2$  is given by

$$\tilde{\theta}_1 = \frac{n\hat{\theta}_1 + m\hat{\theta}_2}{n+m} \quad (1)$$

being  $\hat{\theta}_2 = y_2 / m$  ( $y_2 \equiv$  number of successes associated with the random sample  $Y_1, \dots, Y_m$ ).

We consider the two following probability vectors:

$$\hat{\mathbf{p}} = \left( \frac{n}{n+m} \frac{y_1}{n}, \frac{n}{n+m} \frac{n-y_1}{n}, \frac{m}{n+m} \frac{y_2}{m}, \frac{m}{n+m} \frac{m-y_2}{m} \right)^T \quad (2)$$

$$\mathbf{p}(\theta_1, \theta_2) = \left( \theta_1 \frac{n}{n+m}, (1-\theta_1) \frac{n}{n+m}, \theta_2 \frac{m}{n+m}, (1-\theta_2) \frac{m}{n+m} \right)^T. \quad (3)$$

If we denote  $N = \frac{mn}{m+n}$  it is clear that

$$\sqrt{N} \left( \hat{\mathbf{p}} - \mathbf{p}(\theta_1, \theta_1) \right) \xrightarrow[N \rightarrow \infty]{L} N \left( \mathbf{0}, \sum_{\mathbf{p}(\theta_1, \theta)} \right)$$

with

$$\sum_{\mathbf{p}(\theta_1, \theta_2)} = \begin{pmatrix} \beta \sum_{\mathbf{p}(\theta_1, \theta)} & \mathbf{0} \\ \mathbf{0} & (1-\beta) \sum_{\mathbf{p}(\theta_1, \theta_2)} \end{pmatrix}$$

and

$$\sum_{\mathbf{p}(\theta_1, \theta_2)} = \begin{pmatrix} \theta_1(1-\theta_1) & -\theta_1(1-\theta_1) \\ -\theta_1(1-\theta_1) & \theta_1(1-\theta_1) \end{pmatrix}.$$

It is not difficult to see that

$$\log L(\theta_1; y_1, y_2) = k - D_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\theta_1, \theta_1)) \quad (4)$$

where  $L(\theta_1; y_1, y_2)$  is the likelihood function associated with our model when  $\theta_1 = \theta_2$ , i.e.,

$$L(\theta_1; y_1, y_2) = \binom{n}{y_1} \binom{m}{y_2} \theta_1^{y_1} (1-\theta_1)^{n-y_1} \theta_1^{y_2} (1-\theta_1)^{m-y_2}$$

and  $D_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\theta_1, \theta_1))$  is the Kullback-Leibler divergence measure between the probability vectors  $\hat{\mathbf{p}}$  and  $\mathbf{p}(\theta_1, \theta_1)$ , i.e.,

$$D_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\theta_1, \theta_1)) = \frac{y_1}{n+m} \log \frac{y_1}{n\theta_1} + \frac{y_2}{n+m} \log \frac{y_2}{m\theta_1} + \frac{n-y_1}{n+m} \log \frac{n-y_1}{n(1-\theta_1)} + \frac{m-y_2}{n+m} \log \frac{m-y_2}{m(1-\theta_1)}.$$

For more details about Kullback-Leibler divergence measure see Kullback (1985) or Pardo (2006).

Based on equation ((4)) we have

$$\tilde{\theta}_1 = \arg \max_{\theta_1 \in (0,1)} L(\theta_1; y_1, y_2) = \arg \min_{\theta_1 \in (0,1)} D_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\theta_1, \theta_1)).$$

If we replace Kullback-Leibler divergence measure by a more general divergence measure we obtain a new estimator. In the following we shall consider the family phi-divergence measures defined in the model under consideration by

$$D_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\theta_1, \theta_1)) = \theta_1 \left( \frac{n}{n+m} \phi \left( \frac{y_1}{n\theta_1} \right) + \frac{m}{n+m} \phi \left( \frac{y_2}{m\theta_1} \right) \right) + (1-\theta_1) \left( \frac{n}{n+m} \phi \left( \frac{n-y_1}{n(1-\theta_1)} \right) + \frac{m}{n+m} \phi \left( \frac{m-y_2}{m(1-\theta_1)} \right) \right) \quad (5)$$

where  $\phi \in \Phi^*$  and  $\Phi^*$  is the class of all convex functions  $\phi(x)$ ,  $x > 0$ , such that at  $x=1$ ,  $\phi(1) = \phi'(1) = 0$ ,  $\phi''(1) > 0$ . For more details about phi-divergence measures see Pardo (2006). Based on the phi-divergence measures defined in (5) we consider in this paper the restricted family of minimum phi-divergence estimators defined as

$$\tilde{\theta}_1^{\phi} = \arg \min_{\theta_1 \in (0,1)} D_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\theta_1, \theta_1)). \quad (6)$$

If we consider in (6)  $\phi(x) = x \log x - x + 1$  we obtain the restricted maximum likelihood estimator (MLE), therefore the restricted MLE can be obtained as a special case of the restricted minimum phi-divergence estimator or we can say that the restricted minimum phi-divergence estimator is a natural extension of the restricted MLE.

From a practical point of view the restricted minimum phi-divergence estimator can be obtained as a solution of the equation

$$\frac{\partial D_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\theta_1, \theta_1))}{\partial \theta_1} = 0,$$

that is, solving for  $\theta_1$  the equation

$$\frac{n}{n+m} \left\{ \phi \left( \frac{y_1}{n\theta_1} \right) - \frac{y_1}{n\theta_1} \phi' \left( \frac{y_1}{n\theta_1} \right) - \left( \phi \left( \frac{n-y_1}{n(1-\theta_1)} \right) - \frac{n-y_1}{n(1-\theta_1)} \phi' \left( \frac{n-y_1}{n(1-\theta_1)} \right) \right) \right\} + \frac{m}{n+m} \left\{ \phi \left( \frac{y_2}{m\theta_1} \right) - \frac{y_2}{m\theta_1} \phi' \left( \frac{y_2}{m\theta_1} \right) - \left( \phi \left( \frac{m-y_2}{m(1-\theta_1)} \right) - \frac{m-y_2}{m(1-\theta_1)} \phi' \left( \frac{m-y_2}{m(1-\theta_1)} \right) \right) \right\} = 0.$$

If we consider the power divergence measures introduced and studied by Cressie and Read (1984), obtained from (5) when we consider the family of functions,

$$\phi_{\lambda}(x) = \begin{cases} \frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)}, & \lambda \neq 0, -1 \\ x \log x - x + 1, & \lambda = 0 \\ -\log x + x - 1, & \lambda = -1 \end{cases}.$$

the corresponding minimum power divergence estimator is given by

$$\tilde{\theta}_1^{\phi_\lambda} \equiv \tilde{\theta}_1^\lambda = \frac{\left( \frac{y_1^{\lambda+1}}{n^\lambda} + \frac{y_2^{\lambda+1}}{m^\lambda} \right)^{1/(\lambda+1)}}{\left( \frac{y_1^{\lambda+1}}{n^\lambda} + \frac{y_2^{\lambda+1}}{m^\lambda} \right)^{1/(\lambda+1)} + \left( \frac{(n-y_1)^{\lambda+1}}{n^\lambda} + \frac{(m-y_2)^{\lambda+1}}{m^\lambda} \right)^{1/(\lambda+1)}}, \quad \lambda \neq 0, -1.$$

On the other hand the classical test statistic for testing

$$H_0 : \theta_1 = \theta_2 \quad (7)$$

is given by

$$Z_N = N(\hat{\theta}_2 - \hat{\theta}_1)^2 / \hat{\sigma}^2$$

where  $\hat{\sigma}^2 = \tilde{\theta}_1(1-\tilde{\theta}_1)$  and  $N = nm / (m+n)$ . Its asymptotic distribution is chi-squared with one degree of freedom.

If we consider the function  $\phi(x) = \frac{1}{2}(x-1)^2$ , we can see that the test statistic  $Z_N$  can be written as

$$Z_N = \frac{2n^*}{\phi''(1)} D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\tilde{\theta}_1, \tilde{\theta}_1)), \quad (n^* = m+n) \quad (8)$$

where  $\mathbf{p}(\tilde{\theta}_1, \tilde{\theta}_1)$  is obtained from (3) replacing  $\theta_1$  and  $\theta_2$  by  $\tilde{\theta}_1$  given in (1).

A first extension of ((8)) is obtained if we consider a general function  $\phi_1$  instead of the function  $\phi(x) = \frac{1}{2}(x-1)^2$ ,

$$T_{\phi_1}^{n*} = \frac{2n^*}{\phi_1''(1)} D_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\tilde{\theta}_1, \tilde{\theta}_1))$$

and a more general extension if we consider the minimum  $\phi_2$ -divergence estimator

$$\tilde{\theta}_1^{\phi_2} = \arg \min_{\theta_1 \in (0,1)} D_{\phi_2}(\hat{\mathbf{p}}, \mathbf{p}(\theta_1, \theta_1)).$$

In this case we have a new family of phi-divergence statistics defined by

$$T_{\phi_1, \phi_2}^{n*} = \frac{2n^*}{\phi_1''(1)} D_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\tilde{\theta}_1^{\phi_2}, \tilde{\theta}_1^{\phi_2})).$$

Based in Morales et al. (1995) the asymptotic distribution of  $T_{\phi_1, \phi_2}^{n*}$  is chi-squared with one degree of freedom. For more details see Pardo (2006).

It is well-known that  $\tilde{\theta}_1$  has a smaller asymptotic risk under quadratic loss than  $\hat{\theta}_1$  when  $\theta_1 = \theta_2$  holds, but as  $\theta_2$  moves away from  $\theta_1$ ,  $\tilde{\theta}_1$  may be both asymptotically biased and

inefficient, while the performance of  $\hat{\theta}_1$  remains constant over such departure. For this reason Ahmed (1991) developed an estimator which is a combination of  $\tilde{\theta}_1$  and  $\hat{\theta}_1$  which is less sensitive to the departure from  $H_0 : \theta_1 = \theta_2$  because incorporates a preliminary test on the null hypothesis  $H_0 : \theta_1 = \theta_2$  versus  $H_1 : \theta_1 \neq \theta_2$ . This estimator was termed preliminary test estimator and defined as

$$\hat{\theta}_1^* = \tilde{\theta}_1 \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Z_N) + \hat{\theta}_1 \mathbf{I}_{(\chi_{1,\alpha}^2, \infty)}(Z_N)$$

where  $Z_N$  was introduced in (8) and by  $I_A(x)$  we are denoting the indicator function taking the value 1 if  $x \in A$  and 0 if  $x \notin A$

In this paper we consider the minimum  $\phi_2$ -divergence estimator  $\tilde{\theta}_1^{\phi_2}$  instead of  $\tilde{\theta}_1$  and the statistic  $T_{\phi_1, \phi_2}^{n*}$  instead of  $Z_N$ . Then we shall consider the preliminary phi-divergence test estimator based on  $\tilde{\theta}_1^{\phi_2}$  and  $T_{\phi_1, \phi_2}^{n*}$  given by

$$\tilde{\theta}_{\phi_1, \phi_2}^{\text{Pre}} = \tilde{\theta}_1^{\phi_2} \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(T_{\phi_1, \phi_2}^{n*}) + \hat{\theta}_1 \mathbf{I}_{(\chi_{1,\alpha}^2, \infty)}(T_{\phi_1, \phi_2}^{n*}). \quad (9)$$

### 3. Asymptotic distributional results

Our problem is the estimation of  $\theta_1$  when it is suspected but one is not sure that  $\theta_1 = \theta_2$  holds. By the definition of  $\tilde{\theta}_{\phi_1, \phi_2}^{\text{Pre}}$  we must pay special attention when  $\theta_2$  is close to  $\theta_1$ . For this reason we are going to assume that  $\theta_2 = \theta_1 + \lambda N^{-1/2}$ ,  $\lambda \in \mathbb{R}$ , being  $N = nm / n^*$  and we consider the contiguous alternative hypotheses

$$H_{1,N} : \theta_2 = \theta_1 + \lambda N^{-1/2}, \lambda \in \mathbb{R} \quad (10)$$

The null hypothesis  $H_0 : \theta_1 = \theta_2$  given in (7) can be written as  $\mathbf{g}(\theta_1, \theta_2) = \theta_1 - \theta_2$ . We denote  $\mathbf{B} = (\partial \mathbf{g} / \partial \mathbf{q})_{\mathbf{q}=\mathbf{q}_1} = (1, -1)$ ,  $\theta = (\theta_1, \theta_2)^T$  and  $\theta_0 = (\theta_1, \theta_1)^T$ . By Pardo J. A. et al. (2003) (see also page 246 in Pardo 2006) we have, denoting  $\tilde{\theta}_1^{\phi_2} = (\tilde{\theta}_1^{\phi_2}, \tilde{\theta}_1^{\phi_2})^T$  that

$$\tilde{\theta}_1^{\phi_2} - \theta_0 = \mathbf{H}_{n^*}(\theta_0) \mathbf{I}_{F, n^*}(\theta_0)^{-1} \mathbf{A}_{n^*}(\theta_0)^T \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2} (\hat{\mathbf{p}} - \mathbf{p}(\theta_0)) + o_p((n^*)^{-1/2})$$

where  $\hat{\mathbf{p}}$  and  $\mathbf{p}(\theta_0)$  were defined in (2) and (3), respectively and  $\mathbf{D}_{\mathbf{p}(\theta_0)}$  represents the diagonal matrix with elements  $\mathbf{p}(\theta_0)$ . The matrices  $\mathbf{A}_{n^*}(\theta_0)$ ,  $\mathbf{H}_{n^*}(\theta_0)$  and  $\mathbf{I}_{F, n^*}(\theta_0)$  are defined respectively by

$$\mathbf{A}_{n^*}(\theta_0) = \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2} \left( \frac{\partial \mathbf{p}(\theta)}{\partial \theta} \right)_{\theta=\theta_0} = \begin{pmatrix} \theta_1^{-1/2} \left( \frac{n}{n+m} \right)^{1/2} & 0 \\ -(1-\theta_1)^{-1/2} \left( \frac{n}{n+m} \right)^{1/2} & 0 \\ 0 & \theta_1^{-1/2} \left( \frac{m}{n+m} \right)^{1/2} \\ 0 & -(1-\theta_1)^{-1/2} \left( \frac{m}{n+m} \right)^{1/2} \end{pmatrix}$$

$$\mathbf{I}_{F,n^*}(\theta_0) = \mathbf{A}_{n^*}(\theta_0)^T \mathbf{A}_{n^*}(\theta_0) = \begin{pmatrix} \frac{n}{n+m} ((1-\theta_1)\theta_1)^{-1} & 0 \\ 0 & \frac{m}{n+m} ((1-\theta_1)\theta_1)^{-1} \end{pmatrix}$$

and

$$\mathbf{H}_{n^*}(\theta_0) = \mathbf{I}_{2 \times 2} - \mathbf{I}_{F,n^*}(\theta_0)^{-1} \mathbf{B}^T \left( \mathbf{B} \mathbf{I}_{F,n^*}(\theta_0)^{-1} \mathbf{B} \right)^{-1} \mathbf{B} = \begin{pmatrix} \frac{n}{n+m} & \frac{m}{n+m} \\ \frac{n}{n+m} & \frac{m}{n+m} \end{pmatrix}, \quad (11)$$

where by  $\mathbf{I}_{2 \times 2}$  we are denoting the identity matrix of order 2.

If we denote by  $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)^T$  the MLE of  $\theta = (\theta_1, \theta_2)$  we can write

$$\hat{\theta} = \theta + \mathbf{I}_{F,n^*}(\theta)^{-1} \mathbf{A}_{n^*}(\theta)^T \mathbf{D}_{\mathbf{p}(\theta)}^{-1/2} (\hat{\mathbf{p}} - \mathbf{p}(\theta)). \quad (12)$$

Therefore by (11) and (12) we get

$$\tilde{\theta}_1^{\phi_2} - \theta_0 = \mathbf{H}_{n^*}(\theta_0) (\hat{\theta} - \theta_0) + o_p \left( (n^*)^{-1/2} \right). \quad (13)$$

In the following theorem we present the distributional asymptotic results that are necessary in the following Sections.

**Theorem 1.** *Under the contiguous alternative hypotheses*

$$H_{1,N} : \theta_2 = \theta_1 + \lambda N^{-1/2}, \lambda \in \mathbb{R}.$$

we have,

- $\sqrt{N}(\hat{\theta}_1 - \theta_1) \xrightarrow[N \rightarrow \infty]{L} \mathcal{N}(0, (1-\beta)(1-\theta_1)\theta_1)$  where  $\beta = \lim_{N \rightarrow \infty} \frac{n}{n+m}$ .
- $\sqrt{N}(\tilde{\theta}_1^{\phi_2} - \theta_1) \xrightarrow[N \rightarrow \infty]{L} \mathcal{N}(\lambda(1-\beta), \beta(1-\beta)(1-\theta_1)\theta_1)$ .
- $\sqrt{N}(\hat{\theta} - \tilde{\theta}_1^{\phi_2}) \xrightarrow[N \rightarrow \infty]{L} \mathcal{N}(\mathbf{m}, \mathbf{S})$ , where

$$\mathbf{m} = (-\lambda(1-\beta), \beta)^T \text{ and } \mathbf{S} = (1-\theta_1)\theta_1 \begin{pmatrix} (1-\beta)^2 & -(1-\beta)\beta \\ -(1-\beta)\beta & \beta^2 \end{pmatrix}.$$

d) The asymptotic distribution of the phi-divergence test statistic  $T_{n^*}^{\phi_1, \phi_2}$  is a noncentral chi-square with a degree of freedom and noncentrality parameter

$$\delta = \frac{\lambda^2}{(1-\theta_1)\theta_1}.$$

**Proof.** Under  $H_{1,N} : \theta_2 = \theta_1 + \lambda N^{-1/2}$ ,

$$\sqrt{N}(\hat{\theta}_1 - \theta_1) \xrightarrow[N \rightarrow \infty]{L} \mathcal{N}(0, (1-\beta)(1-\theta_1)\theta_1) \quad (14)$$

and

$$\sqrt{N}(\hat{\theta}_2 - \theta_1) \xrightarrow[N \rightarrow \infty]{L} \mathcal{N}(\lambda, \beta(1-\theta_1)\theta_1),$$

because

$$\sqrt{N}(\hat{\theta}_2 - \theta_1) = \sqrt{N}(\hat{\theta}_2 - \theta_2) + \lambda.$$

They are also asymptotically independent. Based on (12) on (13) we have

$$\sqrt{N}(\tilde{\theta}^{\phi_2} - \theta_1) \xrightarrow[N \rightarrow \infty]{L} \mathcal{N}(\lambda(1-\beta), \beta(1-\beta)(1-\theta_1)\theta_1). \quad (15)$$

From (12) and (14)

$$\sqrt{N}(\hat{\theta} - \tilde{\theta}^{\phi_2}) = (\mathbf{I}_{2 \times 2} - \mathbf{H}_{n^*}(\theta_0))\sqrt{N}(\hat{\theta} - \theta_0) + o_P(1)$$

and taking in account that

$$\mathbf{I}_{2 \times 2} - \mathbf{H}_{n^*}(\theta_0) = \frac{1}{n+m} \begin{pmatrix} m & -m \\ -n & n \end{pmatrix}$$

we get

$$\sqrt{N}(\hat{\theta} - \tilde{\theta}^{\phi_2}) \xrightarrow[N \rightarrow \infty]{L} \mathcal{N}\left(\begin{pmatrix} -\lambda(1-\beta) \\ \lambda\beta \end{pmatrix}, (1-\theta_1)\theta_1 \begin{pmatrix} (1-\beta)^2 & -(1-\beta)\beta \\ -(1-\beta)\beta & \beta^2 \end{pmatrix}\right).$$

It is not difficult to see that

$$\begin{aligned} T_{n^*}^{\phi_1, \phi_2} &= \frac{n+m}{\sqrt{nm}} \sqrt{N}(\hat{\theta} - \tilde{\theta}^{\phi_2})^T \mathbf{I}_{F, n^*}(\theta_1) \frac{n+m}{\sqrt{nm}} \sqrt{N}(\hat{\theta} - \tilde{\theta}^{\phi_2}) + o_P(1) \\ &= \mathbf{X}_N \mathbf{X}_N^T + o_P(1) \end{aligned}$$

where

$$\mathbf{X}_N = \frac{n+m}{\sqrt{nm}} \mathbf{I}_{F, n^*}(\theta_1)^{1/2} \sqrt{N}(\hat{\theta} - \tilde{\theta}^{\phi_2}) \xrightarrow[N \rightarrow \infty]{L} \mathcal{N}\left(\frac{\lambda}{\sigma} \begin{pmatrix} -\sqrt{1-\beta} \\ \sqrt{\beta} \end{pmatrix}, \begin{pmatrix} 1-\beta & -\sqrt{\beta(1-\beta)} \\ -\sqrt{\beta(1-\beta)} & \beta \end{pmatrix}\right).$$

The matrix

$$\begin{pmatrix} 1-\beta & -\sqrt{\beta(1-\beta)} \\ -\sqrt{\beta(1-\beta)} & \beta \end{pmatrix}$$



is idempotent of rank 1. Therefore  $\mathbf{X}_N \mathbf{X}_N^T$  is asymptotically distributed as a chi-square with a degree of freedom and noncentrality parameter

$$\mathbf{d} = \frac{\lambda}{\sigma} \left( -\sqrt{1-\beta}, \sqrt{\beta} \right) \frac{\lambda}{\sigma} \begin{pmatrix} -\sqrt{1-\beta} \\ \sqrt{\beta} \end{pmatrix} = \frac{\lambda^2}{\sigma^2}.$$

Note that  $\sigma^2 = \theta_1(1-\theta_1)$ .

#### 4 Asymptotic bias and asymptotic mean squared errors of the estimators

Let  $\hat{\theta}^*$  be an estimator of  $\theta_1$  and  $F$  the asymptotic distribution of the random variable  $\sqrt{N}(\hat{\theta}^* - \theta_1)$ . We understand by the asymptotic bias of  $\hat{\theta}^*$  the expression

$$B(\hat{\theta}^*) = \int x dF(x).$$

In the following theorem we get the asymptotic bias of  $\hat{\theta}_1$ ,  $\tilde{\theta}^{\phi_2}$  and  $\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}$ .

**Theorem 2** Under  $H_{1,N} : \theta_2 = \theta_1 + N^{-1/2}\lambda$  the asymptotic bias of  $\hat{\theta}_1$ ,  $\tilde{\theta}^{\phi_2}$  and  $\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}$  is given by:

a)  $B(\hat{\theta}_1) = 0$ .

b)  $B(\tilde{\theta}^{\phi_2}) = (1-\beta)\lambda$ .

c)  $B(\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}) = \lambda(1-\beta)G_{\chi_3^2(\delta)}(\chi_{1,\alpha}^2)$ , where  $G_{\chi_3^2(\delta)}(c)$  represents the distribution

function of a noncentral chi-square random variable with three degrees of freedom and noncentrality parameter  $\delta = \lambda^2 / \theta_1(1-\theta_1)$  evaluated at  $c$ .

**Proof.** Parts a) and b) follow by previous Theorem. We observe that  $\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}$  can be written as

$$\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}} = \hat{\theta}_1 + (\tilde{\theta}^{\phi_2} - \hat{\theta}_1) \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(T_{\phi_1, \phi_2}^{n*}).$$

Therefore,

$$\sqrt{N}(\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}} - \theta_1) = \sqrt{N}(\hat{\theta}_1 - \theta_1) + \sqrt{N}(\tilde{\theta}^{\phi_2} - \hat{\theta}_1) \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(T_{\phi_1, \phi_2}^{n*}). \quad (16)$$

Let  $Y$  be a normal random variable with mean  $\lambda / \sqrt{\theta_1(1-\theta_1)}$  and variance 1. The distribution of  $Y^2$  coincides with the asymptotic distribution of  $T_{\phi_1, \phi_2}^{n*}$ . We denote

$U_N = \sqrt{N}(\hat{\theta}_1 - \theta_1)$  and by  $U$  the random variable verifying  $U_N \xrightarrow[N \rightarrow \infty]{L} U$  where  $U$  is a normal random variable with mean zero and variance  $(1-\beta)\theta_1(1-\theta_1)$ . We can write,

$$\sqrt{N}(\tilde{\theta}^{\phi_2} - \hat{\theta}_1) = \frac{m}{m+n} \sqrt{\theta_1(1-\theta_1)} \frac{1}{\frac{m}{m+n} \sqrt{\theta_1(1-\theta_1)}} \sqrt{N}(\tilde{\theta}^{\phi_2} - \hat{\theta}_1).$$

It is clear that

$$\frac{1}{\frac{m}{m+n} \sqrt{\theta_1(1-\theta_1)}} \sqrt{N}(\tilde{\theta}^{\phi_2} - \hat{\theta}_1) \xrightarrow[N \rightarrow \infty]{L} Y.$$

Therefore,

$$\begin{aligned} B\left(\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}\right) &= E[U] + \sqrt{\theta_1(1-\theta_1)}(1-\beta) E\left[\mathbf{Y}_{(0, \chi_{1, \alpha}^2)}(Y^2)\right] \\ &= \sqrt{\theta_1(1-\theta_1)}(1-\beta) E\left[\mathbf{Y}_{(0, \chi_{1, \alpha}^2)}(Y^2)\right]. \end{aligned}$$

If  $\eta$  is a normal random variable with mean  $\mu$  and variance 1,

$$E\left[\eta \mathbf{I}_{(0, c)}(\eta^2)\right] = \mu G_{\chi_3^2}(\mu^2)(c).$$

For more details see Judge and Bock (1978) or Saleh (2006). Then,

$$B\left(\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}\right) = \sqrt{\theta_1(1-\theta_1)}(1-\beta) \frac{\lambda}{\sqrt{\theta_1(1-\theta_1)}} G_{\chi_3^2}(\delta)(\chi_{1, \alpha}^2).$$

Let  $\hat{\theta}^*$  be an estimator of  $\theta_1$  and  $F$  the asymptotic distribution of the random variable  $\sqrt{N}(\hat{\theta}^* - \theta_1)$ . We understand by the asymptotic mean squared error of  $\hat{\theta}^*$  the expression

$$MSE(\hat{\theta}^*) = \int x^2 dF(x).$$

In the following theorem we get the asymptotic mean square error of  $\hat{\theta}_1$ ,  $\tilde{\theta}^{\phi_2}$  and  $\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}$ .

**Theorem 3.** 3Under  $H_{1, N} : \theta_2 = \theta_1 + N^{-1/2} \lambda$  the asymptotic mean square error of  $\hat{\theta}_1$ ,  $\tilde{\theta}^{\phi_2}$  and  $\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}$  is given by:

- $MSE(\hat{\theta}_1) = (1-\beta)\theta_1(1-\theta_1)$
- $MSE(\tilde{\theta}^{\phi_2}) = (1-\beta)\theta_1(1-\theta_1) + (1-\beta)^2 \lambda^2$

$$\begin{aligned} \text{c) } MSE\left(\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}\right) &= (1-\beta)^2 \theta_1(1-\theta_1) + (1-\beta)^2 \lambda^2 \left\{ 2G_{\chi_3^2}(\delta)(\chi_{1,\alpha}^2) - G_{\chi_3^2}(\delta)(\chi_{1,\alpha}^2) \right\} \\ &\quad + (1-\beta)\theta_1(1-\theta_1)G_{\chi_3^2}(\delta)(\chi_{1,\alpha}^2). \end{aligned}$$

**Proof.** Parts a) and b) are immediate on the basis of (14) and (15). We denote  $W_N = \sqrt{N}(\hat{\theta}_2 - \theta_1)$  and we have

$$W_N \xrightarrow[N \rightarrow \infty]{L} W$$

where  $W$  is a normal random variable with mean  $\lambda$  and variance  $\theta_1(1-\theta_1)\beta$ . It is easy to see that the random vector  $(U, W-U)^T$  is normal with mean vector  $(0, \lambda)^T$  and variance-covariance matrix

$$\begin{pmatrix} (1-\beta)\theta_1(1-\theta_1) & -(1-\beta)\theta_1(1-\theta_1) \\ -(1-\beta)\theta_1(1-\theta_1) & \theta_1(1-\theta_1) \end{pmatrix}.$$

Based on this result we have,

$$E[U / (W-U = w-u)] = (1-\beta)(\lambda - (w-u)).$$

We denote  $V_N = \sqrt{N}(\tilde{\theta}^{\phi_2} - \hat{\theta}_1)$  and by  $V$  the random variable verifying  $V_N \xrightarrow[N \rightarrow \infty]{L} V$ , then we have,

$$MSE\left(\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}\right) = E\left[\left(U + V\mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2)\right)^2\right].$$

We know by (14) that

$$\sqrt{N}(\tilde{\theta}^{\phi_2} - \hat{\theta}_1) = \sqrt{N} \frac{m}{n+m}(\hat{\theta}_2 - \hat{\theta}_1) + o_p(1).$$

Therefore,

$$V = (1-\beta)(W-U)$$

and

$$\begin{aligned} MSE\left(\tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}}\right) &= E\left[\left(U + (1-\beta)(W-U)\mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2)\right)^2\right] \\ &= E[U^2] + (1-\beta)^2 E\left[(W-U)^2 \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2)\right] \\ &\quad + 2(1-\beta) E\left[U(W-U) \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2)\right]. \end{aligned}$$

We are going to get the expression of  $l = E\left[U(W-U) \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2)\right]$ . We have,

$$\begin{aligned}
 l &= E \left[ (W-U) \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2) E[U / (W-U)] \right] \\
 &= E \left[ (W-U)(1-\beta) \lambda \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2) \right] - (1-\beta) E \left[ (W-U)^2 \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2) \right] \\
 &= \lambda(1-\beta) E \left[ (W-U) \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2) \right] - (1-\beta) E \left[ (W-U)^2 \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2) \right].
 \end{aligned}$$

Then we get,

$$\text{MSE} \left( \tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}} \right) = E[U^2] - (1-\beta)^2 E \left[ (W-U)^2 \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2) \right] + 2\lambda(1-\beta)^2 E \left[ (W-U) \mathbf{I}_{(0, \chi_{1,\alpha}^2)}(Y^2) \right].$$

Now applying: "If  $\eta$  is a normal random variable with mean  $\mu$  and variance 1, then

$$E \left[ \eta^2 \mathbf{I}_{(0,c)}(\eta^2) \right] = G_{\chi_3^2(\mu^2)}(c) + \mu^2 G_{\chi_5^2(\mu^2)}(c)".$$

We get

$$\begin{aligned}
 \text{ECM} \left( \tilde{\theta}_{\phi_1, \phi_2}^{\text{pre}} \right) &= (1-\beta) \theta_1 (1-\theta_1) + (1-\beta)^2 \lambda^2 \left\{ 2G_{\chi_3^2(\delta)}(\chi_{1,\alpha}^2) - G_{\chi_5^2(\delta)}(\chi_{1,\alpha}^2) \right\} \\
 &\quad + (1-\beta) \theta_1 (1-\theta_1) G_{\chi_3^2(\delta)}(\chi_{1,\alpha}^2).
 \end{aligned}$$

**Acknowledgement.** This work was partially supported by Grant MTM 2009-10072.

## References

1. Ahmed, S. E. (1991). To pool or not to pool: The discrete data. *Statistics and Probability Letter*, 11, 233-237.
2. Ahmed, S. E. (1993). Pooling means under uncertain prior information with application to discrete distributions. *Statistics*, 24, 3, 265-277.
3. Ahmed, S. E. and Ullah, B. (1999). To pool or not to pool: the multivariate data. *Sankhya Ser. B*, 61, no. 2, 266-288.
4. Ahmed, S. E. and Rohatgi, V. K. (1996). Shrinkage estimation of the proportion in randomized response. *Metrika*, 43, 3, 17-30.
5. Ahmed, S. E. and Saleh, A. K. Md. E. (1989). Pooling multivariate data. *J. Statist. Comput. Simulation*, 31, 3, 149--167.
6. Ahmed, S. E. and Tomkins, R. J. (1997). On pooling means from two lognormal populations with unequal variances. *J. Appl. Statist. Sci.*, 6, no. 2-3, 125--143.
7. Ahmed, S. E., Zafaryab, K. M. and Abdul Kadir A. (1997). Pooling slopes of several regression lines. *Pakistan J. Statist.*, 13, no. 1, 79-86.
8. Bancroft, T. A. (1944). On biases in estimation due to use of preliminary tests of significance. *Ann. Math. Statist.*, 15, 190-204.
9. Cressie, N. and Pardo, L. (2002). Phi-divergence statistics, in *Encyclopedia of Environmetrics*. (A. H. ElShaarawi and W. W. Piegorich, Eds). Volume 3, 1551-1555, John Wiley & Sons, New York.

10. Csiszár I. (1963). Eine Informationtheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität Markovschen Ketten. *Publications of the Mathematical Institute of Hungarian Academy of Sciences*, Series A, 8, 84-108.
11. Ferguson T. S. (1996). *A Course in Large Sample Theory*. Texts in Statistical Science, Chapman & Hall, New York.
12. Han, Chien-Pai and Bancroft, T. A. (1968). On pooling means when variance is unknown. *J. Amer. Statist. Assoc.*, 63, 1333--1342.
13. Judge, G. G. and Bock, M. E. (1978). *The Statistical Implications of Pretest and Stein-Rule Estimators in Econometrics*. North Holland. Amsterdam.
14. Johnson, J. P. and Bancroft, T. A. and Han, Chien Pai (1977). A pooling methodology for regressions in prediction. *Biometrics*, 33, no. 1, 57-67.
15. Kale, B. K. and Bancroft, T. A. (1967). Inference for some incomplete specified models involving normal approximations to discrete data. *Biometrics*, 23, 335-348.
16. Kullback, S. (1985). In *Encyclopedia of Statistical Sciences*, Volume 4 (editors S. Kotz and N. L. Johnson), 421-425, John Wiley, New York.
17. Mahdi, T. N., Ahmed, S. E. and Ahsanullah, M. (1998). Improved prediction: pooling two identical regression lines. *J. Appl. Statist. Sci.*, 7, no. 1, 63-86.
18. Menéndez, M. L., Pardo, L. and Zografos, K. (2010). Preliminary test estimators in intraclass correlation model under unequal family sizes. *Math. Methods Statist.*, 19, no. 1, 73-87.
19. Menéndez, M. L., Pardo, L. and Pardo, M. C. (2008). Preliminary test estimators and phi-divergence measures in generalized linear models with binary data. *J. Multivariate Anal.*, 99, 10, 2265-2284.
20. Mosteller F. (1948). On pooling data. *Journal of the American Statistical Association*, 43, 231-242.
21. Mehta, J. S., Srinivasan, R. I. (1971). On pooling data. Estimation of the mean. *Ann. Inst. Statist. Math.* 23, 211-224.
22. Pardo L., *Statistical Inference Based on Divergence Measures*. Chapman & Hall/CRC, New York, 2006.
23. Pardo, L. and Martin, N. (2011). On the comparison of the pre-test and shrinkage phi-divergence test estimators for the symmetry model of categorical data. *J. Comput. Appl. Math.*, 235, no. 5, 1160--1179.
24. Raghunandanan, K. (1978). On pooling data to estimate variance. *Sankhya Ser. B* 40, no. 1-2, 94--104.
25. Saleh, A. K. Md. E. (2006). *Theory of Preliminary Test and Stein-Type Estimation with Applications*, Wiley.
26. Stein C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability I*, 197-206 (University of California Press, Berkeley, CA).