Journal of the Korean Statistical Society (2000), **29**: *2, pp 169–185*

Estimating the Population Size from a Truncated Sample †

Sung Chil Yeo[1]

ABSTRACT

Given a random sample of size N (unknown) with density $f(x|\theta)$, suppose that only n observations which lie outside a region R are recorded. On the basis of n observations, the Bayes estimators of θ and N are considered and their asymptotic expansions are developed to find the third order asymptotic properties with those of the maximum likelihood estimators and the Bayes modal estimators. The asymptotic m.s.e.'s of these estimators are expressed. An example is given to illustrate the results obtained.

Keywords: Truncated Sample; Population Size; Bayes Estimators; Asymptotic Expansions.

1. INTRODUCTION

Our view of truncated sampling is the one taken by the previous articles such as Sanathanan(1972, 1977), Dahiya and Gross(1973), Blumenthal and Marcus(1975), Blumenthal(1977, 1982), Blumenthal, Dahiya and Gross(1978), Blumenthal and Sanathanan(1980), Watson and Blumenthal(1980) and Yeo(1991). Namely, given N independent, identically distributed random variables X_1, X_2, $\cdots$, X_N with common density $f(x|\theta)$, where $x \in X$, a real valued sample space and $\theta \in \Theta$, a real valued parameter space, suppose that only the n observations in a region $\bar{R} (= X - R)$ are recorded, and the remaining $N-n$ are lost. However, we do not even record the fact that these $N-n$ observations were in R

Thus, only a random number n of N potential observations remain observable as a result of the restricting process, and for notational convenience these will be labelled X_1, X_2, $\cdots$, X_n. On the basis of this n truncated observations, we wish

†This work was supported by the research fund from Konkuk University.

[1]Department of Applied Statistics, Konkuk University, Seoul, 143-701, Korea.

to estimate both θ and N and to study the higher order asymptotic properties of the estimators for θ and N.

The previous articles mentioned above were concerned with the m.l.e.'s of θ and N and their modifications, especially Bayes modal estimators. Asymptotic expansions have been obtained and the second order asymptotic properties such as bias and mean squared error(m.s.e.) have been studied in a restricted class of Bayes modal estimators.

On the other hand, Yeo(1991) has examined the asymptotic properties of the Bayes estimators which minimize the expected loss, especially the posterior means which result from the squared error loss function. But he has only developed the asymptotic expansions of order $O(N^{-1})$ to correct the asymptotic biases and median biases of the Bayes estimators of θ and N.

In this paper, we extend the asymptotic expansions of the Bayes estimators of θ and N up to $O(N^{-\frac{3}{2}})$ and then examine their asymptotic m.s.e.'s in order to compare with the m.l.e.'s and the Bayes modal estimators. In Section 2, we briefly review the previous results about the asymptotic expansions of the m.l.e. and the Bayes modal estimator of θ for the case of complete samples. In Section 3, extending the results given in Section 2, we develope the asymptotic expansions, we give expressions for the asymptotic m.s.e.'s of these estimators. In Section 4, we present an example to illustrate the results given in Section 3.

2. COMPLETE SAMPLES

In this section, we express the asymptotic expansions of the m.l.e., the Bayes modal estimator, and the Bayes estimator of θ for the case of complete samples. We continue to use the notaions of Yeo(1991) throughout this paper.

First, consider a stochastic expansions of the form

$$\hat{\theta} = \theta + \frac{1}{\sqrt{N}}A + \frac{1}{N}B + \frac{1}{N\sqrt{N}}C + O(N^{-2}), \tag{2.1}$$

where A, B and C are polynomials in certain sums of i.i.d. random variables, which will be specified on later.

In order to express the coefficients A, B and C for $\hat{\theta}_0$, $\hat{\theta}_m$ and $\hat{\theta}_B$, we continue to take the same notations, for example, L_{ijkl}, Z_{ijkl} and etc. and the same assumptions for $f(x|\theta)$ and $\pi(\theta)$ as used in Yeo(1991) throughout this paper. Then, as shown in (2.12) of Yeo(1991), we have that $E(Z_{ijkl}) = 0$ and $Var(Z_{ijkl}) = V_{ijkl}$, where

$$V_{ijkl} = L_{2i2j2k2l} - L_{ijkl}^2. \tag{2.2}$$

As Pfanzagle(1973) indicated, under suitable regularity conditions such as interchangeability of integrals and derivatives, we see that

$$\begin{aligned} L_1 &= 0, \quad L_{01} + L_2 = 0, \quad L_{001} + 3L_{11} + L_3 = 0, \\ & L_{001} + 4L_{101} + 3L_{02} + 6L_{21} + L_4 = 0, \\ L_2' &= 2L_{11} + L_3, \quad L_{11}' = L_{21} + L_{101} + L_{02}, \quad \text{etc.} \end{aligned} \tag{2.3}$$

Now, the coefficient A and B in (2.1) for $\hat\theta_0$, $\hat\theta_m$ and $\hat\theta_B$ are as in (2.14) of Yeo(1991), and the coefficients C's for $\hat{\theta_0}$, $\hat{\theta_m}$ and $\hat{\theta_B}$ are

$$\begin{aligned} C_0 &= \tfrac{1}{L_2^3}\{Z_1 Z_{01}^2 + \tfrac{Z_1^2}{2L_2}(3Z_{01}L_{001} + Z_{001}L_2) + \tfrac{Z_1^3}{6L_2^2}(3L_{001}^2 + L_2 L_{0001})\}, \\ C_m &= C_0 + \eta_m, \\ C_B &= C_0 + \eta_m + \eta_B, \end{aligned} \tag{2.4}$$

where

$$\begin{aligned} \eta_m &= \frac{1}{L_2^3}\{\xi(L_{001}Z_1 + L_2 Z_{01}) + \xi' L_2 Z_{01}) + \xi' L_2 Z_1\} \\ &= \nu_m(\frac{L_{001}}{L_2^2}Z_1 + \frac{Z_{01}}{L_2}) + \nu_m' \frac{Z_1}{L_2}, \\ \eta_B &= \frac{1}{2L_2^2}\{(\frac{2L_{001}^2}{L_2^2} + \frac{L_{001}}{L_2})Z_1 + \frac{2L_{001}}{L_2}Z_{01} + Z_{001}\} \\ &= \nu_B\{(\frac{2L_{001}}{L_2^2} + \frac{1}{L_2})Z_1 + \frac{2Z_{01}}{L_2}\} + \frac{Z_{001}}{2L_2^2}. \end{aligned} \tag{2.5}$$

From (2.14) of Yeo(1991) and from (2.4), we find that

$$\begin{aligned} \hat\theta_m - \hat\theta_0 &= \frac{1}{N}\nu_m + \frac{1}{N\sqrt{N}}\eta_m + O(N^{-2}), \\ \hat\theta_B - \hat\theta_m &= \frac{1}{N}\nu_B + \frac{1}{N\sqrt{N}}\eta_B + O(N^{-2}), \\ \hat\theta_B - \hat\theta_0 &= \frac{1}{N}(\nu_m + \nu_B) + \frac{1}{N\sqrt{N}}(\nu_m + \nu_B) + O(N^{-2}). \end{aligned} \tag{2.6}$$

Thus, for asymptotic approximations, we may regard $\hat\theta_m$ and $\hat\theta_B$ as ajusted m.l.e's given by

$$\begin{aligned} \hat\theta_m &\approx \hat\theta_0 + \frac{1}{N}\hat\nu_m + \frac{1}{N\sqrt{N}}\hat\eta_m, \\ \hat\theta_B &\approx \hat\theta_0 + \frac{1}{N}(\hat\nu_m + \hat\nu_B) + \frac{1}{N\sqrt{N}}(\hat\eta_m + \hat\eta_B), \end{aligned} \tag{2.7}$$

where

$$\hat{\nu}_m = \nu_m(\hat{\theta}_0),\ \hat{\nu}_B = \nu_B(\hat{\theta}_0),\ \hat{\eta}_m = \eta_m(\hat{\theta}_0),\ \hat{\eta}_B = \eta_B(\hat{\theta}_0). \tag{2.8}$$

Next, from (2.1) the expansions of the first moments (bias) and second moments (mean square deviations) are given by

$$E\,(\sqrt{N}(\hat{\theta}-\theta)) = \frac{1}{\sqrt{N}}b + O(N^{-1}) \tag{2.9}$$

and

$$\begin{aligned} &E\,[(\sqrt{N}(\hat{\theta}-\theta))^2] \\ &= \frac{1}{L_2} + \frac{1}{N}[E(B^2) + 2E(AC) + 2\sqrt{N}E(AB) + O(N^{-\frac{3}{2}}), \end{aligned} \tag{2.10}$$

respectively, where $b = E(B)$ can be regarded as the asymptotic bias of $\hat{\theta}$, and the bias terms b's in (2.9) for $\hat{\theta}_0$, $\hat{\theta}_m$ and $\hat{\theta}_B$ are given by (2.22) of Yeo(1991). In evaluating (2.10), we use the relations such that

$$E\,(Z_1^2 Z_{01}) = \frac{1}{\sqrt{N}}(L_{21} + L_2^2) + O(N^{-1}) \tag{2.11}$$

which follow from the fact that the Z's are sums of i.i.d. random variables. Then, from Gusev(1976) the expressions of (2.10) for $\hat{\theta} = \hat{\theta}_0$, $\hat{\theta}_m$ and $\hat{\theta}_B$ are given by

$$\begin{aligned} &E[(\sqrt{N}(\hat{\theta}_0-\theta))^2] \\ &= \frac{1}{L_2} + \frac{1}{N}\left[\frac{1}{L_2^4}\left\{6L_{11}(L_{11}+2L_{001}) + L_{0001}(\frac{15}{4}L_{001}+L_3)\right\}\right. \\ &\quad \left. + \frac{1}{L_2^3}(2L_{21}+3L_{02}+3L_{101}+L_{0001}) - \frac{1}{L_2}\right] + O(N^{-\frac{3}{2}}), \end{aligned} \tag{2.12}$$

$$\begin{aligned} &E[(\sqrt{N}(\hat{\theta}_m-\theta))^2] \\ &= E[(\sqrt{N}(\hat{\theta}_0-\theta))^2] + \frac{1}{N}[\frac{\xi}{L_2^3}(3L_{001}+4L_{11}) + \frac{1}{L_2^2}(\xi^2+2\xi')] \\ &\quad + O(N^{-\frac{3}{2}}), \end{aligned} \tag{2.13}$$

$$\begin{aligned} &E\,[(\sqrt{N}(\hat{\theta}_B-\theta))^2] \\ &= E\,[(\sqrt{N}(\hat{\theta}_m-\theta))^2] + \frac{1}{N}[\frac{1}{L_2^4}(3L_{001}^2+3L_{11}L_{001}) \\ &\quad + \frac{1}{L_2^3}(L_{0001}+L_{101}+\xi L_{001})] + O(N^{-\frac{3}{2}}). \end{aligned} \tag{2.14}$$

The expression (2.11) may not be convenient for comparisons nor may it show the key role played by the B term in determining m.s.e.. Rao(1963) found the following relation for the m.s.e. in the multivariate case and Ghosh and Subramanyum(1974) and Efron(1975) extended it to the curved exponential family:

$$\sqrt{N}E(AB) + E(AC) = \frac{(EB)'}{L_2} \tag{2.15}$$

This allows the m.s.e. (2.10) to be expressed as

$$\begin{aligned} E[(\sqrt{N}(\hat{\theta} - \theta))^2] &= \frac{1}{L_2} + \frac{1}{N}[E(B^2) + \frac{2(EB)'}{L_2}] \\ &= \frac{1}{L_2} + \frac{1}{N}[Var(B) + (EB)^2 + \frac{2(EB)'}{L_2}]. \end{aligned} \tag{2.16}$$

The equality (2.13) can be verified formally using the terms on the right hand side of (2.11) corresponding to $\sqrt{N}E(AB) + E(AC)$ and comparing them to $\frac{(EB)'}{L_2}$ using (2.22) of Yeo(1991) for $E(B)$ and relations such as (2.3).

We now want to extend the class of estimators under study by examining for any asjusted estimator $\hat{\theta}_a$. If $\hat{\theta}$ has the expansion given by (2.1) and (2.14) of Yeo(1991) and (2.4), then $\hat{\theta}_a$ has the expansion expressed as

$$\begin{aligned} \hat{\theta}_a &= \theta + \frac{A_a}{\sqrt{N}} + \frac{B_a}{N} + \frac{C_a}{N\sqrt{N}} + O(N^{-2}) \\ &= \theta + \frac{A}{\sqrt{N}} + \frac{(B + \nu_a(\theta))}{N} + \frac{(C + Z_1\nu_a'(\theta)L_2^{-1})}{N\sqrt{N}} + O(N^{-2}), \end{aligned} \tag{2.17}$$

where $\nu_a(\theta) = \nu_m(\theta)$, $\nu_m(\theta) + \nu_B(\theta)$ for $\hat{\theta}_a = \hat{\theta}_m$, $\hat{\theta}_B$, respectively. It is clear that A_a, B_a and C_a satisfy (2.15) so that (2.16) continues to hold for $\hat{\theta}_a$. Using (2.16), the effect of ν on m.s.e. is easily obtained, namely

$$\begin{aligned} &E[((\sqrt{N}(\hat{\theta}_a - \theta))^2] - E[(\sqrt{N}(\hat{\theta} - \theta))^2] \\ &= \frac{1}{N}[2\nu(\theta)E(B) + \nu^2(\theta) + 2\nu'(\theta)L_2^{-1}] \end{aligned} \tag{2.18}$$

or alternatively,

$$E[((\sqrt{N}(\hat{\theta}_a - \theta))^2] = \frac{1}{L_2} + \frac{1}{N}[Var(B) + (\nu + E(B))^2 + \frac{2}{L_2}(\nu + E(B))']. \tag{2.19}$$

From (2.14) of Yeo(1991), (2.4) and (2.17), we see that C term is not the same for $\hat{\theta}$ and $\hat{\theta}_a$.

On the other hand, in order to distinguish among estimators having the same first order efficiency, Rao(1961, 1962, 1963) introduced a notion of second order efficiency, which we describe briefly. We assume all estimators have expansions

(1.1). Let

$$L = \frac{d}{d\theta}\log l(\mathbf{x}|\theta) = \sqrt{N}Z_1\,. \tag{2.20}$$

Rao(1962, 1963) calls an estimator first order efficient if

$$\sqrt{N}\left(\frac{L}{N} - \beta(\theta)(\hat{\theta}-\theta)\right) \xrightarrow{p} 0 \tag{2.21}$$

for some function $\beta(\theta)$. This is equivalent to $A = \frac{Z_1}{\beta(\theta)}$. In order to achieve the Cramer-Rao lower bound for asymptotic variance, it is necessary for $\beta(\theta)$ to be L_2, and we shall take it as such henceforth. Rao define the second order efficiency as

$$E_2 = \min_{\lambda} Var(U_\lambda), \tag{2.23}$$

where

$$U_\lambda = N\left(\frac{L}{N} - \beta(\theta)(\hat{\theta}-\theta) - \lambda(\theta) - \theta)^2\right). \tag{2.24}$$

When A is given by (2.14) of Yeo(1991), E_2 is given in terms of B as

$$\begin{aligned} E_2 &= L_2^2 Var(B)\left[1 - \left\{Corr(B, \frac{Z_1^2}{L_2})\right\}^2\right] \\ &= L_2^2\left[Var(B) - \frac{1}{2}\left\{E(B(\frac{Z_1^2}{L_2} - 1))\right\}^2\right]. \end{aligned} \tag{2.25}$$

We can evaluate (2.25) for the m.l.e. of θ using

$$Var(B) = L_2^{-4}(L_2V_{01} + L_{11}^2 + \frac{1}{2}L_{001}^2 + 2L_{001}L_{11}) \tag{2.26}$$

and

$$E(B(\frac{Z_1^2}{L_2} - 1)) = L_2^{-1}(2L_{11} + L_{001}) \tag{2.27}$$

giving for the m.l.e. of θ

$$E_2 = L_2^{-2}(L_2V_{01} - L_{11}^2) \tag{2.28}$$

which was given by Ghosh and Subramanyum(1974) and Efron(1975) for the curved exponential family. From (2.25) we see that E_2 is unaffected by non-stochastic change in B and is mainly useful in distinguishing among classes of estimators, but cannot distinguish within a class. Minimizing E_2 may not be the same as minimizing $Var(B)$ since a larger $Var(B)$ could be compensated for by a larger correlation between B and Z_1^2.

3. TRUNCATED SAMPLES

3.1. Asymptotic Expansions for $\hat{\theta}$

First consider the conditional stochastic expansion of $\hat{\theta}$, given n, of the form

$$\hat{\theta} = \theta + \frac{1}{\sqrt{n}}A + \frac{1}{n}B + \frac{1}{n\sqrt{n}}C + O(n^{-2}). \tag{3.1}$$

In order to develop asymtotic expansions from truncated samples, we take the same notations as used in Section 3 of Yeo(1991). By the same arguments of Yeo(1991) and from (2.4), the coefficients A and B in (3.1) for $\hat{\theta}_c$ and $\hat{\theta}_{c,m}$ are given by (3.22) of Yeo(1991), and the coefficients A and B in (3.1) for $\hat{\theta}_B$ are given by (2.28) of Yeo(1991).

From (2.4) the coefficients C in (3.1) for $\hat{\theta}_c$ and $\hat{\theta}_{c,m}$ are

$$\begin{aligned} C_c = \frac{1}{\tilde{L}_2^3}\Bigg[&\tilde{Z}_1\tilde{Z}_{01}^2 + \frac{\tilde{Z}_1^2}{2\tilde{Z}_2}\left(3\tilde{Z}_{01}\tilde{L}_{001} + \tilde{Z}_{001}\tilde{L}_2\right) \\ &+ \frac{\tilde{Z}_1^3}{6\tilde{L}_2^2}\left(3\tilde{L}_{001}^2 + \tilde{L}_2\tilde{L}_{0001}\right)\Bigg], \end{aligned} \tag{3.2}$$

$$C_{c,m} = C_c + \nu_m'\frac{\tilde{Z}_1}{\tilde{L}_2} + \nu_m\frac{\tilde{W}}{\tilde{L}_2}, \tag{3.3}$$

$$C_{u,m} = C_{c,m} + \nu_m'\frac{\tilde{Z}_1}{\tilde{L}_2} + \nu_u\frac{\tilde{W}}{\tilde{L}_2}, \tag{3.4}$$

where

$$\nu_m = \frac{\xi}{\tilde{L}_2}, \quad \tilde{W} = \tilde{Z}_{01} - \frac{\tilde{L}_{11}}{\tilde{L}_2}\tilde{Z}_1, \quad \nu_u = -\frac{U(\theta)}{2\tilde{L}_2}. \tag{3.5}$$

Next, the coefficient C in (3.1) for $\hat{\theta}_B$ is

$$C_B = C_{u,m} + \eta_B, \tag{3.6}$$

where

$$\eta_B = \frac{1}{2\tilde{L}_2^2}\left[\left(\frac{2\tilde{L}_{001}^2}{\tilde{L}_2^2} + \frac{\tilde{L}_{001}}{\tilde{L}_2}\right)\tilde{Z}_1 + \frac{2\tilde{L}_{001}}{\tilde{L}_2}\tilde{Z}_{01} + \tilde{Z}_{001}\right]. \tag{3.7}$$

On the other hand, from (2.10) and (2.12) we obtain the asymptotic m.s.e. for $\hat{\theta} = \hat{\theta}_c$ as follows:

$$\begin{aligned} &E_n[(\sqrt{n}(\hat{\theta}_c - \theta))^2] \\ &= \frac{1}{\tilde{L}_2} + \frac{1}{n}[E_n(B_c^2) + 2E_n(A_cC_c) + 2\sqrt{n}E_n(A_0B_0)] + O(n^{-\frac{3}{2}}) \\ &= \frac{1}{\tilde{L}_2} + \frac{1}{n}\left[\frac{1}{\tilde{L}_2^4}\left\{6\tilde{L}_{11}(\tilde{L}_{11} + 2\tilde{L}_{001}) + \tilde{L}_{001}\left(\frac{15}{4}\tilde{L}_{001} + \tilde{L}_3\right)\right\}\right. \\ &\qquad \left. + \frac{1}{\tilde{L}_2^3}\{2\tilde{L}_{21} + 3\tilde{L}_{02} + 3\tilde{L}_{101} + \tilde{L}_{0001}\} - \frac{1}{\tilde{L}_2}\right] \end{aligned} \tag{3.8}$$

Or, from (2.16) we have

$$E_n[(\sqrt{n}(\hat{\theta}_c - \theta))^2] = \frac{1}{\tilde{L}_2} + \frac{1}{n}\left[Var_n(B_c) + (EB_c)^2 + \frac{2(EB_c)'}{\tilde{L}_2}\right]. \tag{3.9}$$

And, from (2.13) and (2.14) the asymptotic m.s.e.'s for $\hat{\theta} = \hat{\theta}_{c,m}$, $\hat{\theta}_u$, $\hat{\theta}_{u,m}$, $\hat{\theta}_B$ are

$$E_n[(\sqrt{n}(\hat{\theta}_a - \theta))^2] = E_n[(\sqrt{n}(\hat{\theta}_c - \theta))^2] + \frac{1}{n}\left[2\nu_a E(B) + \nu_a^2 + \frac{2\nu_a'}{\tilde{L}_2}\right], \tag{3.10}$$

where

$$\begin{array}{lll} \nu_a = \nu_m & \text{for} & \hat{\theta}_a = \hat{\theta}_{c,m}, \\ \nu_a = \nu_u & \text{for} & \hat{\theta}_a = \hat{\theta}_u, \\ \nu_a = \nu_m + \nu_u & \text{for} & \hat{\theta}_a = \hat{\theta}_{u,m}, \\ \nu_a = \nu_m + \nu_u + \nu_B & \text{for} & \hat{\theta}_a = \hat{\theta}_B. \end{array} \tag{3.11}$$

Next, we consider the unconditional expansions for $\hat{\theta} = \hat{\theta}_c$, $\hat{\theta}_{c,m}$, $\hat{\theta}_u$, $\hat{\theta}_{u,m}$, or $\hat{\theta}_B$. In order to express the unconditional stochastic expansions for $\hat{\theta}$ in terms of N, namely

$$\hat{\theta} = \theta + \frac{1}{\sqrt{N}}\tilde{A} + \frac{1}{N}\tilde{B} + \frac{1}{N\sqrt{N}}\tilde{C} + O(N^{-2}), \tag{3.12}$$

we use the standardized variable Y given by

$$Y = \frac{n - Nq}{\sqrt{Npq}} \tag{3.13}$$

which is asymptotically normal. By the similar arguments to Yeo(1991), we see that

$$\begin{aligned} \tilde{A} &= \frac{A}{\sqrt{q}} = \frac{\tilde{Z}_1}{\tilde{L}_2\sqrt{q}}, \\ \tilde{B} &= \frac{1}{q}(B - \frac{1}{2}AY\sqrt{p}), \\ \tilde{C} &= \frac{1}{\sqrt{q^3}}(C - BY\sqrt{p} + \frac{3}{8}AY^2p). \end{aligned} \tag{3.14}$$

Thus, the coefficients $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ in (3.12) for $\hat{\theta} = \hat{\theta}_c$, $\hat{\theta}_{c,m}$, $\hat{\theta}_u$, $\hat{\theta}_{u,m}$, or $\hat{\theta}_B$ can be obtained by (3.14) with A, B and C given in (3.23) - (3.25) and (3.29) of Yeo(1991) and (3.2) – (3.4) and (3.7), respectively.

Next, for the unconditional moment expansions for $\hat{\theta}$, in order to evaluate asymptotic m.s.e. for $\hat{\theta} = \hat{\theta}_c$, $\hat{\theta}_{c,m}$, $\hat{\theta}_u$, $\hat{\theta}_{u,m}$, or $\hat{\theta}_B$, we find that

$$E[(\hat{\theta} - \theta)^2] = \frac{1}{Nq\tilde{L}_2} + \frac{1}{N^2q^2}\left[E(B^2) + \frac{2(EB)'}{\tilde{L}_2} + \frac{P}{\tilde{L}_2}\right]. \tag{3.15}$$

Using (3.14) and (3.15), we find a formula for the m.s.e. of $\hat{\theta}$ in terms of $\tilde{B}$ which can be useful if $\tilde{B}$ is derived directly rather than obtained from B, namely

$$\begin{aligned} &E[(\hat{\theta} - \theta)^2] \\ &= \frac{1}{Nq\tilde{L}_2} + \frac{1}{N^2q^2\tilde{L}_2}\left[\tilde{L}_2q^2E(\tilde{B}^2) + 2q(E\tilde{B})' + 2q'E(\tilde{B}) + \tfrac{3p}{4}\right]. \end{aligned} \tag{3.16}$$

3.2. Asymptotic Expansions for $\hat{N}$

We start with a conditional expansion for $\hat{N}$, given n, namely

$$\hat{N} = \frac{n}{q} - \sqrt{n}\alpha - \beta - \frac{1}{\sqrt{n}}\gamma - O(n^{-1}). \tag{3.17}$$

By similar argument to Yeo(1991), we find that

$$\begin{aligned} &\alpha = \frac{Aq'}{q^2}, \beta = \frac{1}{q^2}\left[Bq' + A^2(\frac{q''}{2} - \frac{q'^2}{q})\right] + \frac{1}{2}, \\ &\gamma = \frac{1}{q^2}\left[Cq' + 2AB(\frac{q''}{2} - \frac{q'^2}{q}) + A^3(\frac{q'''}{6} - \frac{q'q''}{q} + \frac{q'^3}{q^2})\right]. \end{aligned} \tag{3.18}$$

In (3.18), we assume that $E(\alpha) = E(A) = 0$. Then we have

$$E^n(\hat{N} - nq^{-1}) = -E^n(\beta) + O(n^{-\frac{1}{2}}). \tag{3.19}$$

And from (3.18), we have

$$E^n(\beta) = \left\{E^n(B)q' + L_2^{-1}(\frac{q''}{2} - \frac{q'^2}{q} + \frac{q^2}{2})\right\}q^{-2}, \tag{3.20}$$

where $E^n(B)$ is given by (3.31) of Yeo(1991).

In developing m.s.e. expressions, the analogue of (2.15) will be used, and we now establish such a relation. Using (3.17) and (3.18), we find

$$E^n[(\hat{N}-nq^{-1})^2]=\frac{nq'^2}{q^4\tilde{L}_2}+E^n(\beta^2+2\sqrt{n}\alpha\beta+2\alpha\gamma). \tag{3.21}$$

To simplify (3.19), we introduce a relation similar to (2.15) due to Blumenthal(1982). Let $\hat{\theta}$ have the expansion (3.1) with

$$A=A_c, B=B_c+\Delta, C=C_c+\lambda. \tag{3.22}$$

If A, B and C satisfy (2.15), then we have the following result:

$$E^n(\sqrt{n}\alpha\beta+\alpha\gamma)=\frac{q'}{q^2\tilde{L}_2}(E\beta)' \tag{3.23}$$

if and only if

$$E^n(A^2\Delta)=\frac{E^n(\Delta)}{\tilde{L}_2}. \tag{3.24}$$

Therefore, when (3.22) and (3.23) hold, we find that

$$E^n[(\hat{N}-\frac{n}{q})^2]=\frac{Nq'^2}{q^3\tilde{L}_2}+E^n(\beta^2)+\frac{2q'(E^n\beta)'}{q^2\tilde{L}_2}. \tag{3.25}$$

Let $\hat{N}_c$ be $\hat{N}$ corresponding to $\hat{\theta}_c$, and let $\hat{\theta}$ be related to $\hat{\theta}_c$ by (3.22) with (3.24) and (3.25) holding. Then the m.s.e.'s of the two $\hat{N}$'s are related by

$$\begin{aligned} &E^n[(\hat{N}-\frac{n}{q})^2]-E^n[(\hat{N}_c-\frac{n}{q})^2] \\ &=\frac{2q'}{q^4}\left[q'E^n(\Delta B_c)+\frac{q'}{2}E^n(\Delta^2)+\frac{q'}{\tilde{L}_2}(E^n\Delta)'\right. \\ &\qquad \left.+\frac{E^n(\Delta)}{2\tilde{L}_2}\left(3q''-\frac{6q'^2}{q}+\tilde{L}_2q^2\right)\right]. \end{aligned} \tag{3.26}$$

The relation

$$(E^n\Delta)'=E^n(\Delta')+E^n(\Delta S) \tag{3.27}$$

can be useful in evaluating (3.26). For an adjusted $\hat{\theta}$, $\hat{\theta}_a$, $\Delta=\nu(\theta)$, $E^n(\Delta B_c)=\nu E(B_c)$ and $(E^n\Delta)'=\nu'$ in (3.26), giving the form

$$\frac{q'^2}{q^4\tilde{L}_2^4}\left[\nu^2\tilde{L}_2^4+\frac{\nu\tilde{L}_2^3}{qq'}(3qq''-6q'^2+\tilde{L}_2q^3)+\nu\tilde{L}_2^2(\tilde{L}_{001}+2\tilde{L}_{11})+2\nu'\tilde{L}_2^3\right]. \tag{3.28}$$

Computation of the m.s.e. of $\hat{N}$ can be accomplished using either (3.26) or the specialization (3.28) along with

$$\begin{aligned}
&E^n\left[\left(\hat{N}_c - \frac{n}{q}\right)^2\right] \\
&= \frac{nq'^2}{q^4\tilde{L}_2} + \frac{1}{4} + \frac{q'^2}{q^4\tilde{L}_2^4}\left[\frac{15}{4}(\tilde{L}_{11} + \tilde{L}_3)^2 - \tilde{L}_3^2\right. \\
&\quad + \tilde{L}_2\left\{\tilde{L}_4 - 2\tilde{L}_{21} - \tilde{L}_{101} - \frac{qq'' - 2q'^2}{2qq'}(7\tilde{L}_3 + 9\tilde{L}_{11})\right\} \\
&\quad + \frac{\tilde{L}_2^2}{4q^2q'^2}\{4q^2q'q''' + 3q^2q''^2 + 36q'^2(q'^2 - qq'') + 2q^4q'(\tilde{L}_3 + 3\tilde{L}_{11})\} \\
&\quad \left. - \tilde{L}_2^3\left(1 + q - \frac{q^2q''}{2q'^2}\right)\right].
\end{aligned} \tag{3.29}$$

Since $\hat{N} - N = (\hat{N} - \frac{n}{q}) + (\frac{n}{q} - N)$, using the unbiasedness and the fact that $E^n(B)$ is independent of n, we have

$$E(\hat{N} - N) = -E^n(\beta) + O(N^{-\frac{1}{2}}), \tag{3.30}$$

Further, expanding

$$(\hat{N} - N)^2 = (\hat{N} - \frac{n}{q})^2 + (\frac{n}{q} - N)^2 + 2(\frac{n}{q} - N)(\hat{N} - \frac{n}{q}) \tag{3.31}$$

and writing $E^n(\hat{N} - \frac{n}{q})$ as $E(\beta) + \frac{1}{\sqrt{n}}E(\gamma)$, where both $E(\beta)$ and $E(\gamma)$ are independent of n, we see that the last term in (3.31) is $O(N^{-\frac{1}{2}})$. And thus,

$$\begin{aligned}
E(\hat{N} - N)^2 &= \frac{Np}{q} + EE^n[(\hat{N} - \frac{n}{q})^2] \\
&= \frac{Np}{q} + \frac{Nq'^2}{q^3\tilde{L}_2} + E^n(\beta^2) + \frac{2q'}{q^2\tilde{L}_2}(E^n\beta)' \\
&= \frac{N}{q}(p + \frac{q'^2}{q^2\tilde{L}_2}) + E(\beta^2) + \frac{2q'}{q^2\tilde{L}_2}(E\beta)'.
\end{aligned} \tag{3.32}$$

Since this differs from $E^n[(\hat{N} - \frac{n}{q})^2]$ only in the leading term which is the same for all estimators having the same α, all conditional comparisons among estimators regarding m.s.e. carry immediately to the unconditional case.

On the other hand, in order to obtain the unconditional stochastic expansion for $\hat{N}$, substituting (3.13) for n in (3.17), expanding and collecting terms, we see that

$$\hat{N} = N - \tilde{\alpha}\sqrt{N} - \tilde{\beta} - \frac{\gamma}{\sqrt{N}} - O(N^{-1}), \tag{3.33}$$

where

$$\begin{aligned}
\tilde{\alpha} &= \alpha\sqrt{q} - Y\sqrt{\tfrac{p}{q}} \\
&= \tfrac{q'}{q}\tilde{A} - Y\sqrt{\tfrac{p}{q}} \\
\tilde{\beta} &= \beta + \tfrac{1}{2}\alpha Y\sqrt{p} \\
&= \tfrac{q'}{q}\tilde{B} + \frac{q'\sqrt{pq}}{q^2}\tilde{A}Y + \frac{qq'' - 2q'^2}{2q^2}\tilde{A}^2 + \tfrac{1}{2} \\
\tilde{\gamma} &= \frac{\gamma}{\sqrt{q}} - \frac{p}{8\sqrt{q}}\alpha Y^2 \\
&= \tfrac{q'}{q}\tilde{C} + \frac{q'\sqrt{p}}{q^{\frac{3}{2}}}\tilde{B}Y + \frac{qq'' - 2q'^2}{2q^{\frac{5}{2}}}(2\sqrt{q}\tilde{A}\tilde{B} + \tilde{A}^2Y\sqrt{p}) \\
&\quad + \frac{\tilde{A}^3}{6q^3}(q^2q''' - 6qq'q'' + 6q'^3)
\end{aligned} \tag{3.34}$$

which relates the unconditional coefficients for N to the conditional coefficients for $\hat{\theta}$ given by (3.14). From (3.32) and (3.34), we can express the m.s.e. of $\hat{N}$ in terms of $\tilde{\beta}$ as

$$E[(\hat{N} - N)^2] = \frac{N}{q}(p + \frac{q'^2}{q^2\tilde{L}_2}) + E(\tilde{\beta}^2) + \frac{2q'}{q^2\tilde{L}_2}(E\tilde{\beta})' - \frac{pq'^2}{4q^4\tilde{L}_2}. \tag{3.35}$$

Denoting $\hat{N}(\hat{\theta}_a)$ as $\hat{N}_a$, and $\hat{N}(\hat{\theta})$ as $\hat{N}$, going back to (3.19) of Yeo(1991) and using a Taylor expansion for $q^{-1}(\hat{\theta}_a)$, we find that

$$\hat{N}_a = \hat{N} - \frac{\nu(\hat{\theta})q'(\hat{\theta})}{q^2(\hat{\theta})} + O(n^{-1}). \tag{3.36}$$

From (3.36) it is clear that $(\hat{N}_a - \hat{N})$ has as its stochastic limit and its expectation $\frac{\nu(\hat{\theta})q'(\theta)}{q^2(\theta)}$. The equation (3.37) gives an alternative way to obtain expressions such as (3.28) as well as showing as an obvious consequence that

$$\beta_a = \beta + \frac{\nu q'}{q^2}. \tag{3.37}$$

Further, (3.35) suggests an alternative way to asjust an $\hat{N}$, namely instead of using $\hat{N}(\hat{\theta}_a)$ we can take

$$\hat{N}_\rho = \hat{N} + \rho(\hat{\theta}) \tag{3.38}$$

for some appropriate function $\rho(\hat{\theta})$. In order to obtain a bias corrected estimator of N, say $\hat{N}^*$, we may use

$$\hat{N}^* = \hat{N}(\hat{\theta}_c) + E(\beta(\hat{\theta}_c)), \tag{3.39}$$

where $E(\beta(\hat{\theta}))$ is given in (3.20). Comparing (3.39) with (3.36), we can also obtain $\hat{N}^*$ as $\hat{N}(\hat{\theta}_a)$ with $\nu = \frac{q^2 E(\beta)}{q'}$. It is easily seen that

$$E(\hat{N}^* - N)^2 = \frac{N}{q}(p + \frac{q'^2}{q^2 \tilde{L}_2}) + Var(\beta), \tag{3.40}$$

where $Var(\beta)$ is conditional on n. Thus the m.s.e. is independent of non stochastic elements in β and in particular, in view of (3.37) remains constant for all $\hat{\theta}_a$ based on the same $\hat{\theta}$. From (3.18) we find

$$Var(\beta) = q^{-4}[q'^2 Var(B) + D^2 Var(A^2) + 2Dq' \, Cov(A^2, B)], \tag{3.41}$$

where $D = \frac{q''}{2} - \frac{q'^2}{q}$. Since A is the same for all efficient, consistent estimators, $Var(\beta)$ is minimized by finding a $\hat{\theta}$ which minimizes

$$Var(B) + (\frac{q''}{q'} - 2q'q) \, Cov(A^2, B). \tag{3.42}$$

Although (3.42) resembles Rao's E_2, in that $Cov(Z_1^2, B)$ enters, it is not the same and there is no assurance that (3.42) will be minimized by either $\hat{\theta}$ which minimizes $Var(B)$ or one with minimum E_2.

Note also that when $\hat{N}$ has an expansion (3.17) with $\alpha = \alpha_c$, $\hat{\theta}$ has an expansion (3.1) with $A = A_c$, (3.23) holds for $\hat{N}$, and $\hat{N}_\rho$ is given by (3.37), then the expansion for $\hat{N}_\rho$ has $\alpha_\rho = \alpha_c$, $\beta_\rho = \beta - \rho$, and $\gamma_\rho = \gamma - A_c\rho'$. It is easily seen that (3.23) holds also for $\hat{N}_\rho$ so that its m.s.e. is given either by (3.25) or (3.32). In particular, we see that

$$E[(\hat{N}_\rho - N)^2] = \frac{N}{q}(p + \frac{q'^2}{q^2 \tilde{L}_2}) + Var(\beta) + (E\beta - \rho)^2 + \frac{2q'}{q^2 \tilde{L}^2}(E\beta - \rho)'. \tag{3.43}$$

4. AN EXAMPLE

In order to illustrate the results given in the previous sections, we consider the problem of estimating the population size N from the truncated exponential samples examined by Blumenthal and Marcus(1975). Let the exponential density function $f(x|\theta)$ be given by

$$f(x|\theta) = \theta e^{-\theta x}, \; x \geq 0, \; \theta > 0, \tag{4.1}$$

and let $R = (t^*, \infty)$, so that observations are right censored at t^*. We assume that θ has the conjugate prior density $\pi(\theta)$ given by

$$\pi(\theta) = \frac{a^{b+1}\theta^b e^{-a\theta}}{\Gamma(b+1)},\ \theta \geq 0,\ a > 0,\ b > -1. \tag{4.2}$$

From Yeo(1991, Section 3.1), we find that

$$\begin{aligned} &p = e^{-\theta t^*}, \quad p' = -t^* e^{-\theta t^*} = -t^* p, \\ &q = 1 - p = 1 - e^{-\theta t^*}, \quad q' = -p' = t^* p, \\ &\tilde{S}(x|\theta) = \frac{1}{\theta} - x - \frac{t^* e^{-\theta t^*}}{1 - e^{-\theta t^*}}, \quad \tilde{S}'(x|\theta) = \frac{t^{*2} p}{q^2} - \frac{1}{\theta^2}, \\ &U(\theta) = -t^*, \quad \xi(\theta) = \frac{b}{\theta} - a. \end{aligned} \tag{4.3}$$

From (3.9) of Yeo(1991), $\hat{\theta}_c$ is given as the solution of

$$\frac{\bar{x}}{t^*} = \frac{1}{\hat{\theta}_c t^*} - \frac{1}{e^{-\hat{\theta}_c t^*} - 1}. \tag{4.4}$$

A table to solve (4.4) is given by Deemer and Votaw(1955). From (4.3) we find that

$$\begin{aligned} \tilde{L}_{11} &= 0, \\ \tilde{L}_2 &= \frac{q^2 - (\theta t_0)^2 p}{(\theta q)^2} = \frac{1}{(\theta q)^2}\{q^2 - (1-q)\log^2(1-q)\}, \\ \tilde{L}_3 &= \frac{1}{(\theta q)^3}\{-2q^3 + (\theta t_0)^3 p(1+p)\},\ \tilde{L}_{001} = -\tilde{L}_3. \end{aligned} \tag{4.5}$$

And from (3.12) and (3.33) with (4.5) we obtain that

$$\begin{aligned} Var(\hat{\theta}) &= Var(\tilde{A}) \\ &= \frac{q\theta^2}{q^2 - (1-q)\log^2(1-q)} \end{aligned} \tag{4.6}$$

and

$$\begin{aligned} Var(\hat{N}) &= Var(\tilde{\alpha}) \\ &= Var\left(\tilde{A}q'q^{-1} - Y\sqrt{\frac{p}{q}}\right) \\ &= \frac{q(1-q)}{q^2 - (1-q)\log^2(1-q)}. \end{aligned} \tag{4.7}$$

Denoting $Var(\hat{N})$ as σ_N^2, we see that $Var(\hat{\theta}) = \frac{\sigma_N^2\theta^2}{1-q}$. Thus, from (3.32) of Yeo(1991), the conditional asymptotic bias terms of $\hat{\theta}_c$, $\hat{\theta}_{c,m}$, $\hat{\theta}_u$, $\hat{\theta}_{u,m}$ and $\hat{\theta}_B$ are b_c, $b_{c,m} = b_c + \nu_m$, $b_u = b_c + \nu_u$, $b_{u,m} = b_c + \nu_u + \nu_m$ and $b_B = b_c + \nu_u + \nu_m + \nu_B$, respectively. And from (3.43) of Yeo(1991), the unconditional asymptotic bias terms are $\tilde{b}_c = \frac{b_c}{q}$, $\tilde{b}_{c,m} = \frac{b_{c,m}}{q}$, $\tilde{b}_u = \frac{b_u}{q}$, $\tilde{b}_{u,m} = \frac{b_{u,m}}{q}$ and $\tilde{b}_B = \frac{b_B}{q}$, respectively, where

$$\begin{aligned} b_c &= \frac{\sigma_N^4\theta}{2q(1-q)^2}\{2q^3 + (1-q)(2-q)\log^3(1-q)\}, \\ \nu_m &= \frac{\theta q\sigma_N^2}{1-q}(b - a\theta), \\ \nu_u &= -\frac{\theta\sigma_N^2 q\log(1-q)}{2(1-q)}, \\ \nu_B &= \frac{\theta\sigma_N^4}{2q(1-q)^2}\{2q^3 + (1-q)(2-q)\log^3(1-q)\}. \end{aligned} \tag{4.8}$$

On the other hand, from (3.54) of Yeo(1991) the asymptotic bias terms of $\hat{N}_c$, $\hat{N}_{c,m}$, $\hat{N}_u$, $\hat{N}_{u,m}$ and $\hat{N}_B$ are g_c, $g_{c,m}$, g_u, $g_{u,m}$ and g_B, respectively, where

$$\begin{aligned} g_c &= -\frac{\sigma^4\log(1-q)}{2q^3(1-q)}\{2q^3 + (1-q)(2-q)\log^3(1-q)\} \\ &\quad -\frac{\sigma^2\log^2(1-q)}{q}\left(\frac{1}{q} - \frac{1}{2}\right) + \frac{1}{2}, \\ \rho_m &= -\frac{\sigma^2\log(1-q)}{q}(b - a\theta), \\ \rho_u &= \frac{\sigma^2\log^2(1-q)}{2q}, \\ \rho_B &= -\sigma^4\left\{\frac{\log(1-q)}{1-q} + \frac{(2-q)\log^4(1-q)}{2q^3}\right\}. \end{aligned} \tag{4.9}$$

On the other hand, the asymptotic m.s.e's of $\hat{\theta}$'s for the conditional and unconditional cases are given by (3.8), (3.10), and (3.16), respectively, where $\tilde{L}_{11}$, $\tilde{L}_2$, and $\tilde{L}_3$ are given by (4.5) and where

$$\begin{aligned} \tilde{L}_{101} &= 0, \quad \tilde{L}_{01} = -\tilde{L}_2, \quad \tilde{L}_{001} = -\tilde{L}_3, \\ \tilde{L}_{21} &= -\tilde{L}_2^2, \quad \tilde{L}_{02} = \tilde{L}_{01}^2 = \tilde{L}_2^2, \\ \tilde{L}_{0001} &= \frac{t_0^4 p(q^2 - 6q + 6)}{q^4} - \frac{6}{\theta^4}. \end{aligned} \tag{4.10}$$

The asymptotic m.s.e.'s of $\hat{N}$'s for the conditional and unconditional cases are given by (3.32) and (3.35) respectively. And from (3.32), (3.35) and (3.43), the relationship of the asymptotic m.s.e's for $\hat{N}_c$ and $\hat{N}_a$ is given by

$$E(\hat{N}_a - N)^2 = E(\hat{N}_c - N)^2 + 2\rho_a E(\beta_c) + \rho_a^2 + \frac{2q'}{q^2\tilde{L}_2}\rho_a', \tag{4.11}$$

where $E(\beta_c) = g_c$.

Acknowledgement

This research topic was originally due to Professor Saul Blumenthal. I wish to express my gratitude to him for suggesting this topic to me. I also want to thank two anonymous referees for their careful readings and useful comments on this paper.

References

Blumenthal, S. (1977). Estimating Population Size with Truncated Sampling, *Communications in Statistics*, A **6**, 297–308.

Blumenthal, S. (1982). Stochastic Expansions for Point Estimation from Truncated Samples, *Sankhya*, A **44**, 436–451.

Blumenthal, S. and Dahiya, R. C., and Gross, A. J.(1978). Estimating the Complete Sample Size from an Incomplete Poisson Sample, *Journal of the American Statistical Association*,**73**, 182–187.

Blumenthal, S. and Marcus, R. (1975). Estimating Population Size with Exponential Failure, *Journal of the American Statistical Association*, **28**, 913–922.

Blumenthal, S. and Sanathanan, L. P. (1980). Estimation with Truncated Inverse Binomial Sampling, *Communications in Statistics*, A **9**, 997–1017.

Dahiya, R. C. and Gross, A. J. (1973). Estimating the Zero Class from a Truncated Poisson Sample, *Journal of the American Statistical Association*, **68**, 731–733.

Deemer, W. L. and Votaw, D. F. (1955). Estimation of Parameters of Truncated or Censored Exponential Distributions, *The Annals of Mathematical Statistics*, **26**, 498–504.

Efron, B. (1975), Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency), *The Annals of Statistics*, **3**, 1189–1242.

Ghosh, J. K. and Subramanyam, K. (1974). Second Order Efficiency of Maximum Likelihood Estimators, *Sankhya*, A **36**, 325–358.

Gusev, S. I. (1975). Asymptotic Expansions Associated with Some Statistical Estimators in the Smooth Case, I. Expansions of Random Variables, *Theory of Probability and Its Applications*,**20**, 470–498.

Gusev, S. I. (1976). Asymptotic Expansions Associated with Some Statistical Estimators in the Smooth Case, II. Expansions of Moments and Distributions, *Theory of Probability and Its Applications*, **21**, 14–33.

Pfanzagl, J. (1973). Asymptotic Expansions Related to Minimum Contrast Estimators, *The Annals of Statistics*, **1**, 993–1026.

Rao, C. R. (1961). Asymptotic Efficiency and Limiting Information, *Proc. Fourth Berkeley Symp. Math. Statist. Prob*, **1**, 531–545.

Rao, C. R. (1962). Efficient Estimates and Optimum Inference Procedures in Large Samples (with discussion), *Journal of the Royal Statistical Society*, B **24**, 46–72.

Rao, C. R. (1963). Criteria of Estimation in Large Samples, *Sankhya*, A **25**, 189–206.

Sanathanan, L. P. (1972). Estimating the Size of a Multinomial Population, *The Annals of Mathematical Statistics*, **43**, 142–152.

Sanathanan, L. P. (1977). Estimating the Size of a Truncated Sample, *Journal of the American Statistical Association*, **72**, 669–672.

Watson, D. and Blumenthal, S. (1980). Estimating the Size of a Truncated Sample, *Communications in Statistics*, A **9**, 1535–1550.

Yeo, S. C. (1991). On Estimating the Distributional Parameter and the Complete Sample Size from Incomplete Samples, *Journal of the Korean Statistical Society*, Vol **20**, 118–138.