CHAPTER II
Two-Stage Sequential Estimation

2.C Introduction

No discussion of two-stage sampling would be complete without mentioning Stein's (1945) two-stage sampling plan. Stein (1945) considered the problem of a two-sample test for a linear hypothesis whose power is independent of the variance. Thereafter many authors have discussed the two-stage sampling method. In particular, Birnbaum and Healy (1960) and Samuel (1966) have considered two-stage sampling techniques for estimation problems.

Birnbaum and Healy (1960) discussed the problem of estimating a real valued function \( p(\Theta) \), for which they required to find an unbiased estimator \( \hat{\Theta} \) having variance not exceeding a positive function \( B(\Theta) \). In the Birnbaum and Healy's two-stage sampling method, they used the information contained in the first sample to determine the size of the second sample and for constructions of unbiased estimates they used only second sample information. The two-stage sampling scheme of Birnbaum and Healy (1960) is given by

a) Observe \( x_1, \ldots, x_m \) and compute \( \hat{\Theta}^2 = \hat{\Theta}^1(x_1, \ldots, x_m) \) which is based on first \( m \) observations.

b) Take a second sample of \( n = \frac{\hat{\Theta}^2}{B} \) observations

\( x_{m+1}, \ldots, x_{m+n} \).
c) Estimate the mean \( \theta = E_\theta (x) \) by the mean \( t = \frac{1}{n} \sum_{i=1}^{n} x_{m+i} \) of the second sample observations only.

In a number of problems, Birnbaum and Healy's procedures for two-stage sampling lead to estimators which are unbiased; in certain problems these estimators have exactly a prescribed variance, while in other cases they have variances never exceeding but generally closed to a prescribed bound. The measure of efficiency of sequential estimate is defined (as an index of efficiency function is given) by

\[
R(\theta) = \left[ \frac{\sum_{\theta} \left( \log f(x, \theta)/\theta \right)^2}{\theta} \right]^{-1} / B(\theta)E_\theta (N')
\]

where \( E_\theta (N') \) is obtained for any given estimators satisfying the required property that

\[
\text{Var}(T) \leq B(\theta)
\]  

These methods of estimation are almost similar to the method of Stein (1945) for estimation of a normal mean. Stein (1945) obtained a sequential confidence interval for the mean of a normal distribution with known variance. The restrictions of unbiasedness which provides essential simplifications of calculations also generally entails some inefficiency from this stand-point. While Stein (1945) was able to give the exact (student's) distribution of the point estimator in his method, but the exact distributions of the estimators given by Birnbaum and Healy (1960) are not known.
On the similar lines Samuel (1966) considered the problem of unbiased estimation of the parameter $\theta$ of the Bernoulli and Poisson distributions. The variance of these estimators are required not to exceed a pre-assigned bound $B$ which is independent of the parameter $\theta$. The method of estimation discussed by Samuel's two-stage-sampling differs from the method of Birnbaum and Healy's (1960) in utilising the information contained in the preliminary sample. Samuel modified some of the Birnbaum and Healy's (1960) procedures by incorporating the first sample directly in estimating $\theta$, still achieving the required property that the variance of the estimate should not exceed the pre-assigned bound $B$.

Samuel considered three following special cases.
(a) The variance is independent of $\theta$, i.e., $\text{Var}_\theta = V$,
(b) The variance is bounded, i.e., $\text{Sup}_\theta \text{Var} = V < \infty$,
(c) $\text{Sup}_\theta \text{Var} = \infty$.

The estimator $T_m = \left( \frac{1}{m} \sum_{i=1}^{m} X_i \right)$, defined by Samuel is based on a fixed sample size, and this estimator is unbiased for the first two cases, (a) and (b), and satisfies the inequality $\text{Var}_\theta (T_m) \leq B$, for all $\theta$, provided the fixed sample size $m \geq V/B$, where $V$ is the variance and $B$ is the preassigned bound. But in the third situation (c) there exists no estimators based on a fixed sample size achieving the inequality $\text{Var}_\theta (T_m) \leq B$, and hence
sequential sampling is necessary. Sequential sampling may be necessary also in situations (a) and (b) if $V$ is unknown. If the estimators achieves the Cramer-Rao lower bound, and necessary regularity conditions are satisfied for the sequential estimators then nothing can be gained by bounded sequential sampling when one is in situation (a) with known $V$. In the second situation (b), sequential estimation may still be worth-while, in the sense that the expected sample size may be smaller than $V/B$ for some, though not for all values of $\theta$.

In the general procedure Samuel considered $X_1, X_2, \ldots, X_m$ as a first sample observations of size $m$ and $N = N(X_1, \ldots, X_m)$ as the size of the second sample $X_{m+1}, \ldots, X_{m-N}$ and defined the estimate as

$$T_N = \frac{1}{N} \sum_{i=1}^{N} \frac{X_{m+1}}{N}$$

(2.0.2)

Since $X_{m+1}$ is independent of $N(X_1, \ldots, X_m)$, $T_N$ is unbiased and

$$\text{Var}_\theta(T_N) = E_\theta \text{Var}(T_N/N) = V_\theta \cdot E_\theta (1/N)$$

therefore if $N$ satisfies

$$E_\theta (1/N) \leq B/V_\theta \quad \text{for all } \theta$$

(2.0.3)

then

$$\text{Var}_\theta(T_N) \leq B \quad \text{for all } \theta.$$
Samuel (1966) then applies this procedure to Bernoulli and Poisson distributions. In the case of Poisson distribution he modified the above estimate (2.0.2) as 
\[ T_N = \sum_{i=1}^{N} \frac{X_i}{N}, \] 
this suggests us to sample only \( N-m \) new observations \( X_{m+1}, \ldots, X_{m+N} \) in the second stage, because it results in a saving of \( m \) observations. But no longer this estimator is an unbiased estimator of \( \theta \) nor its variance satisfies (2.0.3). Hence he defined a new estimator
\[ Z = f(N) \sum_{i=1}^{m} X_i + \sum_{i=m+1}^{N} \frac{X_i}{N}. \] 
(2.0.4)
for properly defined \( f(N) \), such that \( Z \) is an unbiased estimate of \( \theta \) and satisfies the inequality \( \text{Var}(Z) \leq B \) for all \( \theta \).

The method discussed by Samuel is shown to be better in particular cases, mentioned above, in the sense of saving in number of observations required to estimate the parameter. However he has not claimed any optimal property of the method. The extension of the results of Samuel to the class of independent increment processes has been given by Maddapur and Halakatti (1983) which have been presented in section 2.1.

Suppose, a family \( \{F_t(x, \theta), \theta \in \Theta\} \) represents a stochastic models for a stochastic process
\{X(t), \ t \geq 0\} \text{ where } \theta \text{ is the real or indexing parameter, which does not involve } t, \text{ identifies a particular model and } \Theta \text{ is known parameter space. In this chapter we obtained an unbiased estimator } \hat{z}(\tau), \text{ which is a function of } X(t) \text{ at } t = \tau, \text{ of } \theta \text{ such that the variance of the estimate is not to exceed a preassigned upper bound } B \text{ which is independent of } \theta \text{ i.e.,}

\text{Var}[\hat{z}(\tau)] \leq B, \text{ for all } \theta \tag{2.0.5}

we proceed through the sequential procedure. The estimators are based on two-phase observation technique.

The section 2.1 describes two methods of estimation procedure. In the first method, the information available through the first phase observation will be used to determine the length of the period of the second phase observation; and the estimator satisfying (2.0.5) will be obtained on the information available through the second sample only. Where as in the second method of estimation, the information gained through the first sample will be used not only to determine the length of the period of the second phase observation, but is also directly used in estimating the parameter. In sections 2.1 and 2.3 respectively we illustrate the methods, by estimating the parameters of the Poisson process and gamma process and discuss the relative efficiency of the methods.
2.1 Sequential Estimation Procedures

Consider a family of probability laws indexed by a real parameter \( \theta \in (a, b) \) for a real separable continuous time processes \( \{X(t), t \geq 0\} \) with stationary independent increments. Such a family of process is called Koopman-Darmois (K-D) family of processes, if \( X(0) \sim 0 \), with probability one and if for each fixed \( X(t_0) \) form a K-D family (Bochhofer-Klofer-Sobel (1968)). The best known examples of K-D process are

(i) Wiener process with drift (ii) the Poisson process
(iii) the gamma process and (iv) the negative binomial process. For such processes \( X(T) \) is a transitive sufficient statistics at time \( T \) based on data \( \{X(t), 0 \leq t \leq T\} \).

Let \( \{X(t), t \geq 0\} \) be a Koopman-Darmois family of processes having the probability density function (p.d.f) or the probability mass function (p.m.f) of \( X(t) \) given by

\[
f_t(x, \theta) = g(x; t) [C(\theta)]^t \exp[\Omega(\theta)x], \ x \in R \quad \theta \in \Theta
\]

with \( f_0(0; \theta') = 1 \), \( f_0(x; \theta) = 0 \) for \( x \neq 0 \), where \( R \) is discrete or continuous set of points on the real line, \( C(\theta) > 0 \), \( g(x, t) > 0 \), and \( \Omega(\theta) \) is strictly monotone in \( \theta \). Denote the mean of \( X(t) \) per unit time is equal to \( \theta \) and hence the variance of \( X(t) \) per unit time will be

\[
\sigma_\theta^2 = \frac{1}{\Omega'(\theta)} \times \frac{1}{\Omega''(\theta)}
\]

where \( \Omega'(\theta) \) is the first derivative.
of $Q(\theta)$ with respect to $\theta$. For such a family the following properties hold.

**Property 2.1.1** : For fixed observational period $t$ not less than $t_0$, there exists an unbiased estimator $z(t) = x(t)/t$ of the parameter $\theta$, i.e.

$$E[z(t)] = \theta.$$ 

**Property 2.2.2** : In the first phase, suppose we observe the process for the period of length $T$, for fixed $T > t_0$, there exists $T = T(x(T), T)$ representing the length of the period of observation at the second phase.

Let $x(T)$ be the value of $X(t)$ at $t = T$ and $x(t - T)$ be the value of $X(t)$ at $t = (T - T)$. The method of estimation discussed in this chapter is a direct extension for the continuous cimo processes of the results of Birnbaum and Hsely (1960) and Samuel (1966).

2.1.1. The First Procedure

Suppose $x(T)$ denote the value of $X(t)$ at $t = T$, the end of the second phase of observation. Let

$$z(T) = \frac{x(T)}{T}.$$ 

Then the conditional expectation and variance of $Z(T)$ given $T$ fixed is
\[ E \left[ Z(T) \mid T \right] = \theta, \]
\[ \text{Var} \left[ Z(T) \mid T \right] = \left[ \frac{\Theta}{T} \right]^{-1} = \frac{\Theta^2}{T} \]

Hence \( \hat{\theta}(T) \) is an unbiased estimator of \( \theta \). The unconditional variance
\[ \text{Var} \left[ Z(T) \right] = E \text{Var} \left[ Z(T) \mid T \right] + \text{Var} \left\{ E \left[ Z(T) \mid T \right] \right\} \]
\[ = E \text{Var} \left\{ \text{Var} \left[ Z(T) \mid T \right] \right\} \]
\[ = \sigma^2 \Theta E \left( \frac{1}{T} \right) \]

(2.2)

Therefore if \( T \) satisfies,
\[ E \left( \frac{1}{T} \right) \leq \frac{B}{\sigma^2} \]

(2.3)

then (2.0.5) holds for \( \hat{\theta}(T) \). The expected length of the period of observation in this procedure is

\[ E(T') = C + E(T) \]

(2.4)

The expression for \( T^* \), for which the minimum of \( E(T') \) is attained can be obtained, by substituting the value of \( T \), determined by the equation

\[ \frac{\partial E(T)}{\partial \Theta} = -1 \]

(2.5)

The value of \( \tau \) usually depends on the value of \( \theta \) and the bound \( B \).

2.1.2 The Second Procedure

Based on the information available in the first phase we define

\[ \hat{\sigma}^2 = h(x(\tau), \tau) \]

(2.6)
Then the length of observational period at the second phase is
\[ T = \frac{h(x(T), \tau)}{B} \] (2.8)
so that,
\[ E \left( \frac{1}{T} \right) = \frac{B}{\sigma^2} \] (2.9)
which satisfies (2.3). Hence the length of the observational period at the second phase is the same for both the procedures. However, since we are interested in utilising the information given by the first phase observation, we observe the process for a further period of length \( T - \tau \).

Let \( x(T - \tau) \) be the value of \( X(t) \) at \( T - \tau \). Using \( x(T) \) and \( x(T - \tau) \) we define
\[ z(\tau) = \frac{1}{B} [1(x(\tau), \tau)] + \frac{x(T - \tau)}{T} \] (2.10)
where 1 is chosen so that
\[ E \left[ z(\tau) \right] = \theta \]
such an unbiased estimator of \( \theta \) will satisfy the condition (2.0.5) namely
\[ \text{Var} \left[ z(\tau) \right] \leq B. \]
Hence from (2.8) the expected total time for the procedure is
\[ E(T) = E \left[ h(x(\tau), \tau) \right] / B. \] (2.11)
Again substituting the value of $T$ obtained from (2.5), we get a minimum period $T^{**}$ required to have the unbiased estimate with the property (2.0.5). The relative efficiency of the second procedure as compared to the first procedure may be determined by the efficiency index,

$$R(\theta) = \frac{T^*/T^{**}}{2.12}$$

If $R(\theta) > 1$, we conclude that the second procedure is better than the first.

2.2. The Poisson Process

Consider the family of Poisson processes $\{X(t)\}_{t \geq 0}$ with stationary independent increments satisfying

$$f_t(x; \theta) = e^{-\theta t} (\theta t)^x / x! , \quad x = 0, 1, 2, \ldots, \quad \theta > 0. \tag{2.13}$$

The problem is to obtain an unbiased estimate of $\theta$ with the property satisfying (2.0.5). If $X(t)$ is observed for $0 \leq t \leq T$, then $x(T)$ is a sufficient statistic for $\theta$.

Suppose the process is observed for a period of length $T$ ($\geq T_0$) and let $x(T)$ be the value of $X(t)$ at $T$.

Let

$$\hat{\theta}_0^2 = \frac{[x(T) + 1]}{T} \tag{2.14}$$

Then

$$E(1/\hat{\theta}_0^2) = E \left[ \frac{T}{(x(T) + 1)} \right]$$

$$= \frac{1}{e} \left( 1 - e^{-T \theta} \right)$$

$$\leq \frac{1}{\lambda} = \frac{1}{\hat{\theta}_0^2}$$
which satisfies (2.7). Hence from (2.8)

\[ T = \left\lfloor x(\tau) + 1 \right\rfloor / \tau B. \] (2.15)

Hence

\[ E(1/T) \leq B/ \sigma_0^2 \text{ for all } \theta, \]

which satisfies (2.3).

Define

\[ \mathbf{z}(\tau) = B \tau l_1(x(\tau), \tau) + x(T - \tau)/T \] (2.16)

where

\[ l_1(x(\tau), \tau) = \begin{cases} 0 & \text{if } x(\tau) = 0 \\ \tau & \text{if } x(\tau) > \tau \end{cases} \] (2.17)

Then

\[ E[\mathbf{z}(\tau)] = 0, \]

\[ \text{Var}[\mathbf{z}(\tau)] = B(1 - e^{-\theta \tau}) - B^2 \tau^2 e^{-\theta \tau} \sum_{j=1}^{\infty} \frac{(\theta \tau)^j}{j!} (j^{-1}(-1)^j) \leq B \] (2.18)

Hence \( \mathbf{z}(\tau) \) satisfies (2.0.5)

Now \( E(T) = (\theta \tau + 1)/\tau B \) (2.19)

The expected period of observation for the first procedure is

\[ E(T') = \tau + (\theta \tau + 1)/\tau B \]

Using (2.5), \( E(T') \) is minimised only when \( \tau = B^{-1/2} \)

hence the minimum expected period is

\[ T^* = (\theta + 2B^{-1/2})/B \] (2.20)

Also substituting \( \tau = B^{-1/2} \) in (2.19) we get minimum expected period of the second procedure as
\( T^* = \frac{(\theta + B^{1/2})}{B} \) \hspace{1cm} (2.21)

Hence the relative efficiency index
\[
R(\theta) = \frac{(\theta + 2B^{1/2})}{(\theta + B^{1/2})}
\] \hspace{1cm} (2.22)

Thus \( R(\theta) > 1 \) for all \( 0 < \theta < \omega \),

\[
\lim_{\theta \to 0} R(\theta) = 2\quad \lim_{\theta \to \omega} R(\theta) = 1
\] \hspace{1cm} (2.23)

Hence saving in observation period will be large if \( \theta \) is small.

2.3. The Gamma Process

The p.d.f. of the gamma process is given by
\[
f_{\xi}(x; \theta) = x^{\tau-1} \exp(-x/\theta) / \theta^\tau, \quad x > 0, \quad \theta > 0.
\] \hspace{1cm} (2.24)

Based on the information through the first phase observation, let
\[
\hat{\sigma}_0^2 = x^2(\tau)/(\tau - 1)(\tau - 2), \quad (\tau > 2)
\] \hspace{1cm} (2.25)

so that
\[
E[1/\hat{\sigma}_0^2] = 1/\sigma_0^2
\]

Hence define
\[
T = x^2(\tau)/(\tau - 1)(\tau - 2)B, \quad (\tau > 2)
\] \hspace{1cm} (2.26)

which also satisfies (2.12).
Let
\[ z(\tau) = \tau (\tau - 1) B \left[ x(\tau) \right]^{-1} + x(T - \tau)/T \] (2.27)

Then
\[ E \left[ Z(\tau) \right] = 0, \]
and
\[ \text{Var} \left[ Z(\tau) \right] = B \frac{\tau B^2}{\theta^2} \left( \frac{\tau^2 - 2 \tau - 2}{(\tau - 2)(\tau - 3)(\tau - 4)} \right) \]
\[(\tau > 4).\] (2.28)

which satisfies (2.0.5). Now from (2.27)
\[ E(T) = \theta^2 \tau (1 + \tau)/(\tau - 1)(\tau - 2)B, \quad (\tau > 2) \] (2.29)

Hence
\[ E(T') = \tau + \theta^2 (\tau)(1 + \tau)/(\tau - 1)(\tau - 2)B, \]
\[(\tau > 2) \] (2.30)

However it is difficult to obtain the values of \( \tau \) for which
\( E(T') \) is minimum, since it involves the problem of getting
real positive roots of biquadratic equation in (2.5).

Nevertheless one can compute
\[ R(\phi) = E(T') / E(T) \]
\[ = 1 + \left[ \tau(\tau - 1)(\tau - 2) \right] / \alpha(1 + \tau) \]
\[ = 1 + \left[ \tau(\tau - 1)(\tau - 2) \right] / \alpha(1 + \tau) \tau \] (2.31)

where
\[ \alpha = \theta^2 / B. \] (2.32)

Since \( \alpha \geq 0 \), one can find \( R(\phi) > 1 \) for all \( \tau \geq 3 \) and
as \( \alpha \to \infty \), \( R(\phi) = 1 \). Hence we conclude that the second
procedure of estimation is better than first for all
finite values of \( \alpha \).