CHAPTER : I
INTRODUCTION AND SUMMARY

1.1 MOTIVATION

In case of fixed sample size (FSS) decision procedure, a sample of fixed size is taken and based on it a suitable decision is taken. However it might be possible to take a decision based on only first few sample observations. For example, consider a single inspection plan with sample size n = 20 and acceptance number C = 3, that is the lot is rejected if more than 3 defective items occur in a sample of 20 items, otherwise the lot is accepted. Let the effective (E) and defective (D) items in the sequence of observations be E D E D E E D E D E E D E E E D E E E D. Here the decision of rejecting the lot can be taken at the 9^{th} item itself as it conforms that the sample contains more than 3 defectives and hence it is not necessary to take further observations. Thus 10^{th} and onward observations do not contribute any additional information from the point of view of accepting or rejecting the lot. More over taking an observation involves cost and time. So for the above sequence of observations, it is desirable to stop after 9^{th} observation.

In general stop sampling as soon as either 4 defective items are observed (and rejecting the lot) or 17 effective items are observed (and accepting the lot). Such a plan is called as a curtailed inspection plan. It is to be noted that the number of observations required in the curtailed sampling scheme does not exceed 2Ø (FSS).

In a sequential procedure, observations are taken one after the other until a certain amount of information is gathered and then once stopped, an appropriate decision is taken.

There are some important problems which can not be solved by a FSS procedure. For example, in case of N (θ, σ^2) distribution, no FSS procedure is available to construct a fixed width confidence interval for θ when σ^2 is unknown and unbounded (Dantzig (194Ø)). Similarly there is no FSS procedure to obtain a fixed-width confidence interval for θ in case of U $(Ø, \theta)$ distribution $(\theta > Ø)$. However such problems can be solved by using sequential procedures.

Sequential methods was developed by Abraham Wald and G.A.Bernard. Abraham Wald (1944,1947,195Ø) developed a general theory of sequential analysis and proposed the Sequential Probability Ratio Test (SPRT). Around the same time Bernard (1944, 1946) independently developed some sequential tests for binomial proportions. Wald, Wald and Wolfwitz (1948), Arrow;

Blackwell and Girshick (1949) have made significant contribution for the developement of Sequential Decision Procedures (SDP).

1.1.1 Elements of S.D.P :

Consider a decision problem (Ω, $\mathbb{D}$, L) where Ω the parameter space, $\mathbb{D}$ a decision space and L is a loss function on $\Omega \times \mathbb{D}$. Further suppose that before a decision in $\mathbb{D}$ is chosen one can observe sequentially the values of a sequence of identically independently distributed (i.i.d) random variables (r.v.'s) X_1, X_2, ...having common distribuion function (d.f.) $F(x, \theta)$. Let C_i be the cost of observing X_i (i = 1, 2,...).

A SDP has two components, namely the stopping rule and terminal decision rule.

(a) Stopping Rule (Sampling Plan) :-

A stopping rule specifies whether a decision in $\mathbb{D}$ be chosen without any observation or whether at least one observation be taken (and be used to take a decision). If at least one observation is to be taken then for every observed values x_1, x_2,..., x_n of X_1, X_2, ..., X_n, the rule specifies whether to stop without further observations (and a decision in $\mathbb{D}$ is chosen) or whether another observation on X_{n+1} be taken.

(b) Terminal Decision Rule :-

According to the stopping rule if no observation is taken then we have to specify a decision $d_o \in \mathbb{D}$. If at least one observation is to be taken then for each possible value $x_1, \ldots, x_n$ observed under the stopping rule, specify the decision $\delta(x_1, \ldots, x_n) \in \mathbb{D}$.

1.1.2 Boundary Regions (Stopping Sets) :-

Let S denote the sample space corresponding to r.v. X_1 and $S \times S \times \ldots \times S = S^n$ be the sample space corresponding to $X_1, X_2, \ldots, X_n$ r.v's. Let S^{∞} be the sample space corresponding to infinite sequence of r.v's $X_1, X_2, \ldots$. A sampling plan in which at least one observation is to be taken can be described by a sequence of subsets $B_n \subset S^n$ (n = 1, 2,...) such that the sampling is terminated after n observations if $(x_1, x_2, \ldots, x_n) \in B_n$ and take next observation if $(x_1, x_2, \ldots, x_n) \notin B_n$. Such sets B_n are called as stopping regions. Note that B_n is a subset of $S^n - \{\bigcup_{k=1}^{n-1} B_k\}$.

1.1.3 Stopping Random Variable :-

With every stopping rule, we define a stopping r.v. N which takes the values 1, 2, ... and denotes the random number of observations taken when sampling is terminated. Let $\{N = n\}$ denote the set of points $(x_1, x_2, \ldots x_n) \in S^n$ for which sampling is terminated after n^{th} observation. Hence $\{N = 1\} = B_1$ and for

$n > 1$, $\{N = n\} = (B_1 \cup B_2 \cup \ldots \cup B_{n-1})^c \cap B_n = B_1^c \cap B_2^c \cap \ldots \cap B_{n-1}^c \cap B_n$. Note that the sets $\{N = n\}$ and $\{N \le n\} = \bigcup_{k=1}^{n} \{N = k\}$ depend only on the r.v's $X_1, X_2, \ldots, X_n$ and not on $X_{n+1}, X_{n+2}, \ldots$ Thus the sets $\{N = n\}$ and $\{N \le n\} = \bigcup_{k=1}^{n} \{N = k\}$ are subsets of S^n.

In particular for the stopping rule involving a fixed number of observations we have, $\{N = j\} = \phi$ (empty set) for $j = 1, 2, \ldots, n-1$ and $\{N = n\} = S^n$.

Example :- Consider the rule : Stop on a day whenever for the first time, the temperature on next day exceeds a certain specified level.

Let X_n be the maximum temperature on the n^{th} day, $n = 1, 2, \ldots$ For a fixed specified temperature t_o, define N as the least integer k such that $X_{k+1} > t_o$. In this example $\{N = n\}$ depends on X_{n+1} also and hence N is not a stopping r.v.

1.1.4 Closed Sequential Sampling Plan : -

If the ultimate termination of sequential procedure is guaranteed then it is called closed. That is if N is a stopping r.v. and if $P(N < \infty) = 1$ (or $P(N = \infty) = 0$) then the sequential procedure is said to be closed.

Note that such restriction is natural, otherwise with positive probability the sampling continues forever and as a consequence we may not be able to take a decision and in a situation the cost of taking observations will be infinite.

To illustrate the above concepts, in the following we shall consider three different rules associated with a sequence of i.i.d B(1, p) r.v's.

ILLUSTRATION : -

Let $X_1, X_2, \ldots, X_n$ be i.i.d B(1, p) r.v's. Let 1 and 0 respectively denote a success and a failure. Consider the following rules.

Rule :- R_1 : Stop as soon as the total is 10.

R_2 : Stop as soon as 3 successive 1's are observed.

and R_3 : Stop as soon as the number of 1's and 0's are equal.

Consider the rule R_1. Here the boundary regions are $B_1 = B_2 = \ldots = B_9 = \phi$ (empty set) and B_{10} has only one point = {1111111111}.

For $n \geq 10$, B_n consists of all sequence of 0 and 1 with 10^{th} 1 appearing at the n^{th} trial. So we have, $|B_n| = \binom{n-1}{9}$. Thus B_{11} has 10 points $\{$(01111111111), (10111111111), ..., (11111111101)$\}$, B_{12} has 55 points and so on.

Let N be a stopping r.v. corresponding to rule R_1. Then we have,

$$P(N < \infty) = \sum_{n=10}^{\infty} \binom{n-1}{9} p^{10} q^{n-10} = p^{10} (1-q)^{-10} = 1.$$

Thus the stopping rule R_1 is closed.

Consider the rule R_2. Here the stopping regions are $B_1 = B_2 = \phi$. This is a problem of filling up the places with last 3 appearances as 1's. So we have,

B_3 has only 1 point = {(111)}.

B_4 has also only 1 point = {(0111)}.

B_5 has 2 points = {(00111), (10111)}.

Similarly, B_6 has 4 points = {(000111), (010111), (100111), (110111)}.

In case of B_7, one point starting with '111' is out and so B_7 has 7 points = {(0000111), (0010111), (0100111), (0110111), (1000111), (1010111), (1100111)}.

In general for $k \geq 7$, there are 2^{k-4} points and the number of points starting with '111' in B_3, B_4, etc to be substracted are $\sum_{i=3}^{k-4} |B_i| \, 2^{k-4-i}$. So we have, $|B_k| = 2^{k-4} - \sum_{i=3}^{k-4} |B_i| \, 2^{k-4-i}$, $\forall$ $k = 7, 8, \ldots$ Thus B_8 has $16 - \sum_{i=3}^{4} |B_i| \, 2^{4-i} = 16 - \left\{ |B_3| \, 2^1 + |B_4| \, 2^0 \right\} = 16 - \{1x2 + 1x1\} = 16 - 3 = 13$ points and so on.

Further, if $P_k = P[N = k]$ then we have,

$P_3 = p^3$, $P_4 = qp^3$, $P_5 = q^2p^3 + qp^4 = qp^3$, $P_6 = q^3p^3 + 2q^2p^4 + qp^5 = q^2p^3(q+p) + qp^4(q+p) = q^2p^3 + qp^4 = qp^3(q+p) = qp^3$ and so on. In general for $k \geq 7$, let E be the event that {Ø111} occurs then we have,

P_k = P({111} has occurred for the first time at k^{th} trial)

= P({111} does not occurr in any of the first k-4 trials and E occurred at k^{th} trial)

$= P(N > k-4)\, P(N = 4)$, since by independence

$= [1 - P(N \leq k-4)]\, qp^3$

$= [1 - \sum_{r=3}^{k-4} P_r]\, qp^3$, since they are disjoint.

$$= [1 - (P_3 + P_4 + \ldots + P_{k-4})]\, qp^3. \qquad (1.1)$$

From (1.1) we have,

$$P_3 + P_4 + \ldots + P_{k-4} = 1 - P_k/qp^3 \quad \forall\ k \geq 7.$$

Thus we have,

$$P(N < \infty) = \sum_{k=1}^{\infty} P(N = k) = \lim_{k\to\infty} (P_3 + P_4 + \ldots + P_{k-4})$$

$$= 1 - (qp^3)^{-1} \lim_{k\to\infty} P_k \qquad (1.2)$$

To find this limit, we proceed as follows. From (1.1) we have,

$$P_{k+1} - P_k = \{1- (P_3 + P_4 + \ldots + P_{k-3})\}qp^3 - \{1 -$$

$$(P_3 + P_4 + \dots + P_{k-4})\} \quad qp^3$$

$$= -qp^3 P_{k-3} < 0 \ , \ \forall \ k \geq 4. \qquad (1.3)$$

Thus $\{P_k\}$ is bounded and decreasing. Hence limits exists. Let $\lim_{k \to \infty} P_k = A$ (constant), from equation (1.3) by taking limit as $k \to \infty$ we get,

$$A - A = -qp^3 A \ \Rightarrow \ A = 0 \ \Rightarrow \ \lim_{k \to \infty} P_k = 0.$$

Hence from (1.2) $P(N < \infty) = 1 - (qp^3)^{-1} \times 0 = 1$. Thus the rule R_2 is closed.

Remark (1) : From equation (1.3), we have $P_k - P_{k+1} = qp^3 P_{k-3}$, $\forall \ k = 4, 5, \dots$

$$\Rightarrow \quad \sum_{k=4}^{\infty} \{kP_k - (k+1)P_{k+1} + P_{k+1}\} = \sum_{k=4}^{\infty} qp^3 \ \{(k-3)P_{k-3} + 3P_{k-3}\}$$

$$\Rightarrow \{E(N) - 3P_3\} - \{E(N) - 3P_3 - 4P_4\} + (1 - P_3 - P_4) = qp^3\{E(N) + 3\}$$

$$\Rightarrow \quad 1 - P_3 + 3P_4 = qp^3 \ \{E(N) + 3\}$$

$$\Rightarrow \quad 1 - p^3 + 3qp^3 = 3qp^3 + qp^3 E(N)$$

$$\Rightarrow \quad E(N) = \frac{1 - p^3}{qp^3}.$$

Remark (2) : We can generalize the above rule as, "Stop as soon as a success run of length r (> 0) occurred in a sequence of Bernoulli trials". By using the similar arguments as above we can show that it is closed and the mean time is $E(N) = \frac{1 - p^r}{qp^r}$. Feller (1972 Vol. I, page 324), obtained the same

$E(N)$ by using the probability generating function of the stopping r.v. N.

Consider the rule R_3. Here boundary region B_n is a null set (ϕ) for n odd. Now for n even, we have by the ballot theorem (Feller (1972), Vol.I, page 73), the boundary B_{2k} (k = 1, 2,...) has number of points equal to

$$\frac{2}{2k-1}\binom{2k-1}{k} = \frac{1}{2k-1}\binom{2k}{k}, \ \forall\ k = 1, 2, \ldots \quad (1.4)$$

It is to be noted that the r.h.s of (1.4) is 2^{2k} times the r.h.s of expression (3.7) of page 78 of Feller (1972, Vol.I).

Thus B_2 has $2\binom{1}{1}$ = 2 points namely (1 Ø) and (Ø 1).

B_4 has $2/3\binom{3}{2}$ = 2 points namely (1 1 Ø Ø) and (Ø Ø 1 1)

B_6 has $2/5\binom{5}{3}$ = 4 points namely (1 1 1 Ø Ø Ø), (1 1 Ø 1 Ø Ø), (Ø Ø Ø 1 1 1) and (Ø Ø 1 Ø 1 1).

B_8 has $2/7\binom{7}{4}$ = 1Ø points

and so on.

Let N be a stopping r.v. corresponding to rule R_3. Since R_3 occurres only at $(2n)^{th}$ trials, R_3 is a reccurrent event of period 2 and by Feller (1972 Vol.I ; Chapter- XIII ; sections 1 - 4) we have,

$$P(N < \infty) = 1 - (1 - 4pq)^{1/2} = 1 - (1 - 4p + 4p^2)^{1/2}$$

$$= 1 - [\ (1-2p)^2\ ]^{1/2}$$

$$= 1 - |1-2p| = 1 - |q+p-2p|$$

$$= 1 - |q - p| \qquad (1.5)$$

Hence from equation (1.5) it is clear that stopping r.v. N is not proper when $p \neq q$ and is proper when $p = q$. In other words the rule R_3 is closed only when $p = q$.

1.2 CHAPTER WISE SUMMARY

This dissertation deals with sequential procedures for various models in statistical estimation. It is divided into four chapters and their contents are described in the following.

Chapter-I includes motivation for sequential procedure and describes the structure of SDP. Some interesting stopping rules are illustrated. Chapter wise summary of the dissertation is also given.

Chapter-II deals with the sequential estimation of the parameter of interest for Binomial and Multinomial models. The results of Girshick et. al (1946) and Lehmann and Stein (1950) for binomial model have been discussed in detail with illustration. Further, the results of Sato (1995) for multinomial model have been discussed. The example given by Sato (1995) have studied in detail and some possible modifications are proposed.

Chapter-III deals with the sequential estimation for Exponential model due to Chattopadhyay (2000) and Uniform due to Cooke (1973). Further we propose sequential estimation procedures for θ in case of U(θ, 2θ) distribution and these results are new in the sense that we have not seen such procedures in the existing literature. Lastly Meczarski (1985) procedure for estimating the minimum of a random variable and its modified form proposed by Schaalje et. al (2001) for Weibull distribution based on simulation study have been discussed and we have extended this for Pareto distribution. Also we have compared these methods with the sequential procedures based on the methods of moment and maximum likelihood estimators.

Chapter - IV is devoted to the simulation study of the performance of some standard procedures such as Stein's (1945) and Cooke's (1973) two-stage procedures. We have compared Stein's (1945) procedure with that of obtained by reducing the second sample size by 1 (called as modified Stein's procedure) for normal distribution. The performance of Cooke's (1973) procedure with general two-stage procedure proposed by Rattihalli and Shirke (unpublished) is carried out for uniform distribution. For this purpose, programs in C-language have been developed and the results are indicated in the form of table together with comments.

Appendices includes various C-language programs used in this dissertation.

Lastly the dissertation ends with a list of references used in this dissertation.

■