Journal of the Korean Institute of Industrial Engineers
Vol. 25, No. 1, pp 8-20, March 1999.

Design and Estimation of Multiple Acceptance Sampling Plans for Stochastically Dependent Nonstationary Processes

WonKyung Kim
Department of Industrial Engineering, Kyungnam University

확률적으로 종속적인 비평형 다단계 샘플링검사법의 설계 및 평가

김원경

In this paper, a design and estimation procedure for the stochastically dependent nonstationary multiple acceptance sampling plans is developed. At first, the rough-cut acceptance and rejection numbers are given as an initial solution from the corresponding sequential sampling plan. A Monte-Carlo algorithm is used to find the acceptance and rejection probabilities of a lot. The conditional probability formula for a sample path is found. The acceptance and rejection probabilities are found when a decision boundary is given. Several decision criteria and the design procedure to select optimal plans are suggested. The formula for measuring performance of these sampling plans is developed. Type I and II error probabilities are also estimated. As a special case, by setting the stage size as 1 in a dependent sampling plan, a sequential sampling plan satisfying type I and II error probabilities is more accurate and a smaller average sample number can be found. In a numerical example, a Polya dependent process is examined. The sampling performances are shown to compare the selection scheme and the effect of the change of the dependency factor.

1. Introduction

Multiple sampling plans can be derived from Wald's [12] sequential sampling plans because it is sometimes more practical to take a sequence of groups of items instead of sequences of items. The ordinary sequential chart may be set up; but, instead of plotting points item by item, the results are accumulated by groups. For example, the results of the first 20 items are plotted, then the results of the first 40, then the first 60, and so forth. Decisions are made as before, but at less frequent intervals in terms of the inspection process. Using such group methods affects the Operating Characteristic(OC) curve and Average Sample Number(ASN). The former is affected very much under certain conditions. The ASN is very likely to be increased, for taking groups means that a decision that might be made on the jth item inspected will have to be put off until the whole group is inspected. It may also mean that a decision will be postponed to later groups, which will further increase the amount of inspection. Group sequential plans are often truncated. That is, after a number of samples have been taken, the plan calls for acceptance or rejection. They then become multiple sampling plans.

It is true, however, that there are still no unified methods pertaining to the multiple sampling plans.

On the other hand, a sequential sampling plan can be regarded as an extension of a multiple sampling plan. In other words, sequential sampling plans can be derived from multiple sampling plans by setting stage size n as 1. But such a reverse way is only possible under the assumption that multiple sampling plans can be easily found. Unfortunately, designing multiple sampling plans is never simpler than designing sequential sampling plans. A sequential sampling plan has anadvantage in that it can give a smaller average sample number when compared to other non-sequential sampling plans. A sequential sampling plan derived from SPRT (Sequential Probability Ratio Test), however, is prone to unsatisfying the type I and II error probabilities especially when the ratio of fraction of defectives is big. Although several analytical methods[7,8,12] have been suggested, they cannot guarantee the required type I and II error probabilities sufficiently. When we design a sampling plan, we must also consider the existence of dependency. In general, if a dependency exists in a sampling plan, it is very complicated or impractical to handle the problem analytically. Even in independent processes, it is very difficult to solve this problem analytically because of computational difficulties in finding the conditional probabilities due to the complex step function shapes of the acceptance and rejection boundary lines. Moreover, the nonstationarity due to the dependency in dependent processes makes this problem more difficult. One of the approaches solving such dependency problems is the use of simulation whereby the problem of finding conditional probability can be solved by numerical methods.

There have been several studies about the dependent sampling plans. Bhat, Lal and Karunaratne[1,2] approached the single acceptance sampling problem with an augmented Markov chain matrix. Broadbent[3] proposed a work estimating the lag 1 serial correlation ρ. The process normally produces good items, but if a defect is caused, the same cause produces a run of bad items. These analytical studies are limited to special cases only, mainly Markov processes, due to the mathematical complexities of the probability structure. Moreover, most efforts have been concentrated on single sampling plans. Even for such single sampling plans, the analysis is extremely complicated. Nelson[9] has developed a method for estimating single acceptance sampling plans for general production process models by simulation. Kim[5] developed an efficient design and estimation procedure for the dependent double sampling plans when the dependent production processes can be simulated. He also developed the general methodologies of the design procedure for the sequential sampling plans when there exists dependence between serial sample data[6]. Friedlander[4] used the lattice filter method for adaptive processing when the process shows correlation between serial data. But this study is limited to time series pattern only. The purpose of this paper is to develop the optimal design procedure of the dependent multiple sampling plans when we inspect grouped items in order that they are produced in the same order that this group items are produced. In the following sections we will discuss the design problems, the nonstationary probability structure, and the measure of performance for the dependent multiple sampling plans. A Polya dependent process will be shown as a numerical example.

2. Design of Dependent Multiple Sampling Plans

In this section, we will discuss how to design multiple sampling plans from sequential sampling plans and how to find the type I and II error probabilities of multiple sampling plans. Let us consider the sequential sampling plans first. The implementation of SPRT starts from finding the acceptance and rejection numbers at each sample time. Consider a production process having a lot of size N. We assume that the production has the stochastic output process $\{X_1, X_2, \cdots, X_N\}$. If the state of the process producing the ith item is good, then $X_i = 0$, and if the process is bad, then $X_i = 1$. The inspector tests each item sequentially until a decision is made. The probability of producing a defective part varies according to previous items. For $i = 1,2,\ldots,N$, define the probability ratio,

$$l(x_1, \cdots, x_i) = p_1(x_1, \cdots, x_i)/p_0(x_1, \cdots, x_i),$$

where $p_1(x_1, \cdots, x_i)$ is the joint probability density of the first i items when the fraction of defectives is LTPD, and $p_0(x_1, \cdots, x_i)$ is the joint probability density of the first i items when the fraction of defectives is AQL. The probability ratio test is

rejected if $l(x_1,\cdots,x_i)\geq(1-\beta)/\alpha$ and accepted If $l(x_1,\cdots,x_i)\geq\beta/(1-\alpha)$. Wald[13] proved that the above fundamental inequalities remain valid for a test procedure in spite of the dependency of the successive observations, provided that the probability is one, e.g., the procedure will eventually terminate.

In general, it is difficult to find the joint probability densities $p_1(x_1,\cdots,x_i)$ and $p_o(x_1,\cdots,x_i)$ analytically if there exists dependency between the processes. However, by using simulation, we can find the dependent joint probability densities. Let us define

f_{ij}^{LTPD}: the joint probability that there are exactly j defectives among the i items when the fraction of defectives is LTPD;

f_{ij}^{AQL} : the joint probability that there are exactly j defectives among the i items when the fraction of defectives is AQL;

u_i : the number of defectives for the test to be rejected at the sample number i; and,

l_i : the number of defectives for the test to be accepted at the sample number i.

The method for finding the probabilities f_{ij}^{LTPD} and f_{ij}^{AQL} is explained in Kim[5,6] in depth. The rejection number u_i and acceptance number l_i at the ith sample number can be found from the following equations. For $i=1,2,...,N$, we get

u_i = minimum j satisfying $f_{ij}^{LTPD}/f_{ij}^{AQL}\geq(1-\beta)/\alpha$, and,

l_i = maximum j satisfying $f_{ij}^{LTPD}/f_{ij}^{AQL}\geq\beta/(1-\alpha)$.

Now let n be the stage size of a multiple sampling plan, then $S(\equiv N/n)$ is the number of stages. For the stage number $k=1,2,\ldots,S$, define the following events:

A_k = {The test is accepted at the stage k};

R_k = {The test is rejected at the stage k}; and,

G_k = {The test is neither accepted nor rejected at the kth stage, i.e., continued to the next stage $k+1$}.

Since A_k, R_k and G_k are mutually exclusive, we have

$$\Pr\{G_{k-1}\}=\Pr\{A_k\cup R_k\cup G_k\},$$
$$=\Pr\{A_k\}+\Pr\{R_k\}+\Pr\{G_k\},$$
and initially $\Pr\{G_o\}=1$.

The probabilities of acceptance $\Pr\{A_k\}$ and the probability of rejection $\Pr\{R_k\}$ at the stage k can be found by simulation or by analytical methods. Let us define the following variables to find these probabilities.

T_k : total number of defective items up to the stage k;

$C_{k,n}$: number of defective items among the n items at the stage k only;

$\Pr\{T_k=i\}\equiv p_{k,i}$: probability that total number of defective items up to the stage k is i;

$\Pr\{C_{k,n}=i\mid T_{k-1}=j\}\equiv h_{k,i,j}$: probability that the number of defectives among n items at the stage k is i under the condition that total number of defectives up to the stage k-1 is j; and,

$\Pr\{C_{k,n}\geq i\mid T_{k-1}=j\}\equiv g_{k,i,j}$: probability that the number of defectives among n items at the stage k is i or more under the condition that total number of defectives up to the stage k-1 is j.

The probability $p_{k,i}$ can be found recursively if we know the probability of the previous stage k-1. For this, let a_k and r_k be the acceptance and rejection number at the stage k respectively. Since the shape of acceptance and rejection boundary lines forms step, the probability structure is complex. So, to make notations simple, we need the following three index range interval sets:

$J_R(k-1)\equiv[a_{k-1}+1, r_{k-1}-1]$: range for rejection and continuation probability formula at the stage $k-1$;

$J_A(k-1)\equiv[a_{k-1}+1, \min(r_{k-1}-1, a_k)]$: range for accept probability formula at the stage $k-1$;

$J_G(k-1)\equiv[a_{k-1}+1, \min(r_{k-1}-1), i)]$: range for continuation formula at the stage k-1 when the total number of defectives up to the stage k is given as i.

The method of determination for the acceptance number a_k and rejection number will be discussed in the next section. Define the initial variables r_k $p_{0,0}=1$, $a_0=-1$, $r_0=1$, $T_o=0$. The probability $p_{k,i}$ for $i\in J_R(k)$ is:

$$p_{k,i}=\Pr\{T_k=i\}$$
$$=\sum_{j\in J_G(k-1)}\Pr\{T_{k-1}=j\cap C_{k,n}=i-j\}$$

$$= \sum_{j \in J_C(k-1)} \Pr\{T_{k-1}=j\} \Pr\{C_{k,n}=i-j \mid T_{k-1}=j\}$$
$$= \sum_{j \in J_C(k-1)} p_{k-1,j}\, h_{k,i-j,j}$$

The three event probabilities at each stage k are as follows:

$$\Pr\{R_k\} = \Pr\{T_k \geq r_k\} = \sum_{j \in J_R(k-1)} \Pr\{T_{k-1}=j \cap C_{k,n} \geq r_k - j\} = \sum_{j \in J_R(k-1)} \Pr\{T_{k-1}=j\} \Pr\{C_{k,n} \geq r_k - j \mid T_{k-1}=j\} = \sum_{j \in J_R(k-1)} p_{k-1,j}\, g_{k,r_k-j,j}$$

$$\Pr\{A_k\} = \Pr\{T_k \leq a_k\} = \sum_{j \in J_A(k-1)} \Pr\{T_{k-1}=j \cap C_{k,n} \leq a_k - j\} = \sum_{j \in J_A(k-1)} \Pr\{T_{k-1}=j\} \Pr\{C_{k,n} \leq a_k - j \mid T_{k-1}=j\} = \sum_{j \in J_A(k-1)} p_{k-1,j}\,(1 - g_{k,a_k-j+1,j})$$

and

$$\Pr\{G_k\} = \sum_{i \in J_R(k)} \Pr\{T_k=i\} = \sum_{i \in J_R(k)} \sum_{j \in J_C(k-1)} \Pr\{T_{k-1}=j \cap C_{k,n}=i-j\} = \sum_{i \in J_R(k)} \sum_{j \in J_C(k-1)} p_{k-1,j}\, h_{k,i-j,j}$$

In the independent Bernoulli process, since the number of defective items of the previous stage is not dependent on the past event $\{T_{k-1}=j\}$, the probability of conditional event $\{C_{k,n} \mid T_{k-1}\}$ is simply the same as that of a binomial, that is, $h_{k,i-j,j} = \binom{n}{i-j} p^{i-j}(1-p)^{n-(i-j)}$ if $i-j \leq n$, or 0 otherwise, where p is the fraction of defectives of the process. Similary,

$$g_{k,i-j,j} = \sum_{s=i-j}^{n} h_{k,s,j} = \sum_{s=i-j}^{n} \binom{n}{s} p^s (1-p)^{n-s}$$

if $i-j \leq n$, or 0 otherwise.

In the dependent process, however, it is very difficult to find the conditional probability $h_{k,i-j,j}$ and $g_{k,i-j,j}$ analytically. The fraction of defectives p in a dependent process can be affected by the number of cumulative defective items, the starting point of the inspection, the order of inspection, the last item's status and the sample size, etc. This in turn affects the rejection probability $h_{k,i-j,j}$ and $g_{k,i-j,j}$ because it depends on p. The number of defective items among the n items starting from the stage number 1 can be different from the one starting from the stage k, where $k>1$. This nonstationarity is especially an important factor in designing dependent multiple sampling plans. If a decision of acceptance or rejection is not made at the stage k, the next stage must be inspected. But the probability structure of the next stage is no more the same as that of the former stages. Therefore, from now on we will discuss the method of finding the nonstationary probabilities and the distributions using simulation.

3. Simulation Implementation for the Nonstationary Probability

Let us see the formula of the probability $p_{k,i}$ again:

$$p_{k,i} = \Pr\{T_k=i\} = \sum_{j \in J_C(k-1)} \Pr\{T_{k-1}=j \cap C_{k,n}=i-j\}$$

The probability $p_{k,i}$ can be estimated by the ratio of the number of simulation replications satisfying the event $\{T_k=i\}$ to the total replication number m. The counting proceeds sequentially from the first stage to the last stage. The joint probability in the above formula can be found by counting the simulation replications satisfying the joint event $\{T_{k-1}=j \cap C_{k,n}=i-j\}$. We can draw the accept and reject boundary lines of a multiple sampling plan on the x-y plane like those of a sequential sampling plan, which usually forms step function shapes. The probability $p_{k,i}$ is the sum of the probabilities of sample paths from point (0, 0) to point (k, i) in the plane.

The counting procedure by simulation consists of two steps. First, select all sample paths reaching the continuation range of the stage $k-1$. Next, select only the paths going to the point (k, i) from there. We count all simulation replications satisfying these two path conditions, and by dividing this count by m, the estimator of the probability is obtained. Prior to starting the process of path finding, we need to record the numbers of defectives at every stage and replication in advance. This is possible by counting the process output values according to the process state. To implement the counting, let us define the following variables. Here, index r represents the rth

simulation replication.

$U_{k,r}$: number of defectives among n items at the stage k of the rth simulation replication, and

$X_{j,r}$: process output value of the jth item which is 0 or 1 at the rth simulation replication.

We can generate $U_{k,r}$ easily:

$$U_{k,r}=\sum_{j=(k-1)n+1}^{kn} I(X_{j,r}=1).$$

Now, to select the paths efficiently, we define the following list set and variables:

$L_{k-1,j}$: a list set of which elements are simulation replication order numbers satisfying the event $\{T_{k-1}=j\}$ at the stage k-1;

$V_{k-1,j}$: size of the list set $L_{k-1,j}$, i.e., number of elements of $L_{k-1,j}$; and,

$l_{k-1,j,r}$: the rth element of the list set $L_{k-1,j}$, i.e., $l_{k-1,j,r}\in L_{k-1,j}$.

Define the initial list set and variables : $L_{0,0}=\{1,2,3,\cdots,m\}$, $l_{0,0,r}=r$ for $r=1,2,\ldots,m$, and $V_{0,0}=m$. Then the list set $L_{k-1,j}$ is represented as :

$$L_{k-1,j}=\{t \mid l_{k-1,j,r}=t, \text{ for } r=1,2,\cdots,V_{k-1,j}\}.$$

The list set $L_{k-1,j}$ contains all path information reaching the continuation range of the stage k-1. Among these $V_{k-1,j}$ simulation replications, count the replications satisfying the event $\{C_{k,n}=i-j\}$ at the stage k. Then the count is the number of joint event $\{T_{k-1}=j\cap C_{k,n}=i-j\}$. The estimated joint probability can be found as follows:

$$\widehat{\Pr}\{T_{k-1}=j\cap C_{k,n}=i-j\}=\sum_{r=1}^{V_{k-1,j}} I(U_{k,t}=i-j, \text{ for } t=l_{k-1,j,r}\in L_{k-1,j})/m$$

For the second index of U in the above formula, we used $t=l_{k-1,j,r}$, since t is one of the simulation replication orders which satisfy the path conditions reaching the continuation region of the stage k-1. This t corresponds to the event $\{T_{k-1}=j\}$. Next, $U_{k,t}=i-j$ corresponds to the event $\{C_{k,n}=i-j\}$, which satisfies selecting the sample paths going to the point (k,i). The indicator function $I(\cdot)$ adds one more count if the simulation replication order of number t satisfies the joint event. Since there can be several sample paths reaching the point (k,i) from $(k\text{-}1, j)$, the above procedure is repeated for all $j\in J_G(k-1)$. Therefore, the estimated probability $\hat{P}_{k,i}$ is:

$$\hat{P}_{k,i}=\sum_{j\in J_G(k-1)}\sum_{r=1}^{V_{k-1,j}} I(U_{k,t}=i-j, \text{ for } t=l_{k-1,j,r}\in L_{k-1,j})/m$$

The sample path information of reaching the point (k,i) just obtained must be saved to the list set $L_{k,i}$ for the next stage k. The list set $L_{k,i}$ of which element is the simulation replication order number satisfying the joint event can be newly generated as follows: For $i\in J_R(k)$,

$$L_{k,i}=\bigcup_{j\in J_G(k-1)}\{t \mid l_{k-1,j,r}=t, \text{ and } U_{k,t}=i-j, \text{ for } r=1,2,\cdots,V_{k-1,j}\}$$

Similarly, we can find the estimated probability of $\Pr\{R_k\}$and $\Pr\{A_k\}$ as follows:

$$\alpha_k=\widehat{\Pr}\{A_k\}=\sum_{j\in J_A(k-1)}\sum_{r=1}^{V_{k-1,j}} I(U_{k,t}=a_{k-j}, \text{ for } t=l_{k-1,j,r}\in L_{k-1,j})/m$$

$$\beta_k=\widehat{\Pr}\{R_k\}=\sum_{j\in J_A(k-1)}\sum_{r=1}^{V_{k-1,j}} I(U_{k,t}=r_{k-j}, \text{ for } t=l_{k-1,j,r}\in L_{k-1,j})/m$$

The above procedure is repeated twice, i.e., under the conditions that fractions of defectives p are AQL and LTPD. To denote this let's add the upper indexes AQL and LTPD. Since the multiple sampling plans must be finished before the last stage eventually, we have:

$$\sum_{k=1}^{S}[\widehat{\Pr}\{R_k^{AQL}\}+\widehat{\Pr}\{A_k^{AQL}\}]=1$$

and $$\sum_{k=1}^{S}[\widehat{\Pr}\{R_k^{LTPD}\}+\widehat{\Pr}\{A_k^{LTPD}\}]=1.$$

Now, the type I and II error probabilities can be estimated as follows:

$$\hat{\alpha}=\sum_{k=1}^{S}\widehat{\Pr}\{R_k^{AQL}\} \text{ and } \hat{\beta}=\sum_{k=1}^{S}\widehat{\Pr}\{A_k^{LTPD}\}$$

4. Searching Schemes Satisfying the α and β Conditions

In the next section we will discuss how to find the best sampling plans given p_0, p_1, α, β and stage size n.Decision of the rejection and acceptance boundaries in multiple sampling plans is not a simple problem. If we simply take the last acceptance and rejection numbers in each sequential sampling group as the rejection and acceptance numbers at each stage in a multiple sampling plan, the type I error probability will be smaller than the required one. The reason is that if a large rejection number is applied to all points in that group, the chance of rejection of the group becomes smaller. On the contrary, the type II error probability will become larger, because if a large acceptance number is applied to all points in that group, the acceptance chance of the group becomes larger. The reverse phenomena will occur if we simply take the first point of each sampling group as the rejection and acceptance number at each stage in a multiple sampling plan. Taking the middle point of each sampling group, the type I and II error probabilities can be smaller or larger according to the sampling status. Therefore, just simply taking arbitrary points is not desirable since they cannot satisfy the required type I and II error probabilities. In view of a mathematical programming problem, the multiple sampling plan is to find the rejection number r_k and acceptance number a_k for $k=1,2,\ldots, S$ by solving the below:

$$\text{Min. } func(\hat{\alpha}, \alpha, \hat{\beta}, \beta)$$

$$\text{s. t. } \hat{\alpha} = \sum_{k=1}^{S} \widehat{\Pr}\{R_k^{AQL}\} \le \alpha \ : \ \alpha \text{ condition}$$

$$\hat{\beta} = \sum_{k=1}^{S} \widehat{\Pr}\{A_k^{LTPD}\} \le \beta \ : \ \beta \text{ condition}$$

The object function $func(\hat{\alpha}, \alpha, \hat{\beta}, \beta)$ may be defined as total risk function $|\hat{\alpha}-\alpha|+|\hat{\beta}-\beta|$, or α -only risk function $|\hat{\alpha}-\alpha|$, or β-only risk function $|\hat{\beta}-\beta|$, or average sample number function ASN $(\hat{\alpha}, \alpha, \hat{\beta}, \beta)$, or average total inspection function ATI$(\hat{\alpha}, \alpha, \hat{\beta}, \beta)$ or any combination of these etc., according to the conditions of sampling environments. Solving the above nonlinear integer programming problem by mathematical programming is extremely difficult. Another approach is the use of a numerical enumeration search method. However, the simple enumeration method requires too many enumerations. For example, if the number of stages is 30 and the possible number of values for each r_k and a_k are 10, then the total enumeration count is 10^{30-2}, which exceeds the normal limit of calculation. Therefore, we need other methods to reduce the enumeration. Instead of searching r_k and a_k for all stages concurrently, searching them stage by stage can reduce enumeration counts. For all stages, initial decision boundaries r_k and a_k can be set from the group sequential sampling plan which gives upper limit of rejection numbers and lower limit of acceptance numbers.

The initial acceptance number at the stage k is taken as the first acceptance value of each group k and the initial rejection number is taken as the last value of each group k: $a_k = l_{(k-1)n+1}$ and $r_k = u_{kn}$. The initially estimated $\hat{\alpha}$ and $\hat{\beta}$ are lower than α and β because of the wider decision boundaries. By narrowing these decision boundaries, $\hat{\alpha}$ and $\hat{\beta}$ can be improved. For this, we will consider four searching schemes called nearest α search, nearest β search, central α search, and central β search.

The nearest α search starts from the first stage($k=1$). At this stage, decrease the rejection number r_k by 1 in order to increase $\hat{\alpha}$ until just before $\hat{\alpha}$ exceeds α. During the procedure, r_k must be greater than or equal to a_k+2, otherwise the sampling test cannot be continued to the next stage. Only at the last stage is $r_k = a_k+1$ allowed in order to stop the sampling test. After decreasing r_k, increase the acceptance number a_k by 1 in order to increase $\hat{\beta}$ until just before $\hat{\beta}$ exceeds β. Also the restriction $a_k+2 \le r_k$ must be kept except in the last stage. Now, since the rejection and acceptance number for the first stage has been decided, increase the stage number k by 1 in order to start the second stage search. This procedure is repeated to the last stage. One of desirable properties of good sampling plans is the ASN should be as small as possible. This searching method makes the ASN as small as possible by assigning the acceptance probability $\Pr\{A_k\}$ and rejection probability $\Pr\{R_k\}$ as large as possible to the earlier stage. The type II error probability of this searching policy is apt to be less satisfied compared to the type I error probability,

since β condition is satisfied after the α condition. The searching procedure of the nearest β search is the same as that of the nearest α search except for the searching order of type I and II error probabilities, i.e., a_k first then r_k next. The rejection numbers of the nearest α search are smaller than that of the nearest β search, since in the nearest α search, r_k is decreased to the minimum within the range of satisfying the α condition. Therefore, the two decision boundary lines of the nearest α search are positioned lower than that of the nearest β search.

In the central α search, r_k is decreased by 1 instead of decreasing to the possible minimum, and then a_k is increased by 1. The α and β conditions are checked every time when r_k decreases and a_k increases. If the α condition is satisfied, decreasing r_k is held, and if the β condition is satisfied, increasing a_k is held. The term "held" is used instead of "stop," because any change of r_k or a_k affects the type I and II error probabilities. That is, if r_k is decreased, $\hat{\alpha}$ is increased but $\hat{\beta}$ is decreased. Changes of r_k and a_k are repeated until just before $\hat{\alpha}$ and $\hat{\beta}$ exceeds α and β. Now, the searching procedure for the next stage begins and this procedure is repeated to the last stage. The decision boundary lines become narrow towarding to the central part of the initial boundary lines. Like the case of the nearest β search, the central β search is the same as the central α search except the searching order, i.e., a_k first then r_k next.

5. Measures of Performance and A Numerical Example

In this section, in order to check the performance of the multiple sampling plans, three measures, i.e., Average Sample Number(ASN), Average Total Inspection(ATI), and Average Outgoing Quality (AOQ) will be explained and a Polya sequential model will be given as a numerical example. Since we found the acceptance and rejection probabilities at each stage in the previous section, the ASN can be computed as follows.

$$\text{ASN} = \sum_{k=1}^{S} kn[\Pr\{A_k\} + \Pr\{R_k\}]$$

If the test is accepted at the stage k, then the total inspection is simply kn. If the test is rejected, then the whole lot will be inspected, hence the total inspection is the lot size N. Therefore, the ATI is:

$$\text{ATI} = \sum_{k=1}^{S} kn\Pr\{A_k\} + \sum_{k=1}^{S} N\Pr\{R_k\}.$$

To find the AOQ we need to know the average number of defective items remaining in a lot and the average number of items actually shipped after inspection. Let us define C_n be the cummulative number of defectives among the first n items, and $\gamma_{n,j} = \Pr[C_n \geq j]$ be the probability that the number of defectives in n items is j or more, and the method of finding this probability is explained in Kim[5,6]. We can obtain the expectation of C_n as follows:

$$\begin{aligned} E[C_n] &= \sum_{j=0}^{\infty} \Pr\{C_n > j\} \\ &= \sum_{j=1}^{\infty} \Pr\{C_n \geq j\} = \sum_{j=1}^{\infty} \gamma_{n,j} = \sum_{j=1}^{n} \gamma_{n,j} \end{aligned}$$

Let $[N_D]_i^j$ represent the number of defective items among the item number i through j, where $j > i$. The expected value of $[N_D]_i^j$ can be found as follows:

$$\begin{aligned} E\{[N_D]_i^j\} &= E\{[N_D]_1^j\} - E\{[N_D]_1^{i-1}\} \\ &= E[C_j] - E[C_{i-1}] \\ &= \sum_{l=1}^{j} \gamma_{j,l} - \sum_{l=1}^{i-1} \gamma_{(i-1),l} \end{aligned}$$

Let N_D be the total number of defective items remaining in a lot of size N after inspection. Also let $N_D \mid R_k$ and $N_D \mid A_k$ be the number of defective items remaining in a lot of size N under the event R_k and A_k for $k = 1, \cdots, S$. Then the expected value of N_D is :

$$\begin{aligned} E[N_D] = &\sum_{k=1}^{S} E[N_D \mid R_k]\Pr\{R_k\} \\ &+ \sum_{k=1}^{S} E[N_D \mid A_k]\Pr\{A_k\} \end{aligned}$$

The first summation term of the right hand side becomes 0 since there are no defective items in a rejected lot. Therefore, we have

$$E[N_D] = \sum_{k=1}^{S} E[N_D \mid A_k]\Pr\{A_k\}, \text{ where}$$

$$\begin{aligned} E[N_D \mid A_k] &= E\{[N_D]_{kn+1}^{N}\} \\ &= \sum_{j=1}^{N} \gamma_{N,j} - \sum_{j=1}^{kn} \gamma_{kn,j} \end{aligned}$$

Now let N_S represent the number of items actually shipped after inspection. If all defective items found are replaced with good ones, then $N_S = N$. When the defective items found are discarded and not replaced with good ones, $E[N_S]$ is calculated as follows. If the procedure is rejected at the stage k, then the actual number of shipped items is $N - r_k$, because the r_k defective items are discarded, and the probability of this event is $\Pr\{R_k\}$. Similarly, if the procedure is accepted at stage k, then the actual number of shipped items is $N - r_k$, because the r_k defective items are discarded, and the probability of this event is $\Pr\{A_k\}$. Therefore, we get the AOQ from the definition:

$$\text{AOQ} = E[N_D]/E[N_S],$$

where

$$E[N_S] = \sum_{k=1}^{S}[(N-r_k)\Pr\{R_k\} + (N-a_k)\Pr\{A_k\}]$$

Now, we will examine a Polya dependent process model. In a Polya sequence(Sarkadi and Vincze[10]) the model is highly dependent on the past history. Unlike the Markov model which depends on only the last state, the Polya sequence depends on all the previous states.

Let $\Pr(X_1 = 0) = 1 - p$, $\Pr(X_1 = 1) = p$, where p is the fraction of defectives at start of the process. The fraction of defectives at the $(n+1)$th sample is affected by the number of defective items found up to the nth inspected item, sampled number n and the dependency factor q as follows

$$\begin{aligned}&\Pr(X_{n+1} = 1 \mid X_1 = e_1, X_2 = e_2, \ldots, X_n = e_n)\\ &= \Pr\left(X_{n+1} = 1 \mid X_1 + \cdots + X_n = \sum_{i=1}^{n} e_i = s\right)\\ &= \frac{p + sq}{1 + nq}\end{aligned}$$

where $0 < p < 1$, $q > 0$, and $e_i = 0$ or 1.

If the dependency factor q is 0, the model is equivalent to the independent model and the sequence becomes a binomial sequence. Now suppose a new tool is used at the beginning of production. We want to produce a lot of 400 items and the produced n items are examined in each stage k successively. Assume with the acceptable tool wear the AQL is 0.01 with a type I error probability α of

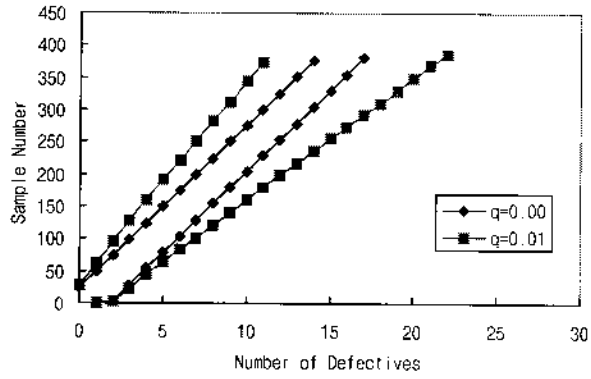

Figure 1. The boundary lines of acceptance and rejection numbers for the Polya dependent process.

Table 1. The first sample number i for the test to be rejected or accepted when the number of defectives is u_i or l_i and the dependency factor is q

u_i for reject or l_i for accept	q=0.00		q=0.01	
	reject	accept	reject	accept
0	-	24	-	26
1	1	49	1	62
2	3	74	3	96
3	28	99	23	128
4	53	124	44	159
5	78	149	63	191
6	103	175	83	222
7	128	200	102	252
8	154	225	121	283
9	179	250	141	313
10	204	275	160	344
11	229	300	179	374
12	254	325	198	
13	279	351	217	
14	305	376	236	
15	330		255	
16	355		274	
17	380		292	
18			311	
19			330	
20			349	
21			368	
22			387	

Table 2. The acceptance number a_k and r_k rejection number when sample size n is 30

q	Stage k		1	2	3	4	5	6	7	8	9	10	11	12	13	14
	kn		30	60	90	120	150	180	210	240	270	300	330	360	390	400
0.00	INI	a_k	-1	0	1	2	3	5	6	7	8	9	11	12	13	14
		r_k	2	3	4	5	7	8	9	10	11	12	13	13	14	15
	NA	a_k	0	1	1	2	3	5	6	7	8	9	11	12	13	13
		r_k	2	3	3	4	5	7	8	9	10	11	13	13	14	14
	NB	a_k	0	1	2	3	5	6	7	8	9	10	11	12	13	13
		r_k	2	3	4	5	7	8	9	10	11	12	13	13	14	14
	CA	a_k	0	1	1	2	4	5	6	7	8	9	11	12	13	13
		r_k	2	3	3	4	6	7	8	9	10	11	13	13	14	14
	CB	a_k	0	1	1	2	4	5	6	7	8	9	11	12	13	13
		r_k	2	3	3	4	6	7	8	9	10	11	13	13	14	14
0.01	INI	a_k	-1	0	0	1	2	3	4	5	6	7	8	9	10	21
		r_k	2	3	5	6	8	10	11	13	14	16	17	19	21	22
	NA	a_k	0	1	1	1	2	3	4	5	6	7	8	9	10	11
		r_k	2	3	3	3	4	5	6	7	8	9	10	11	12	12
	NB	a_k	0	1	3	4	6	8	9	11	12	14	15	17	19	20
		r_k	2	3	5	6	8	10	11	13	14	16	17	19	21	21
	CA	a_k	0	1	2	3	4	6	7	9	10	12	13	15	17	18
		r_k	2	3	4	5	6	8	9	11	12	14	15	17	19	19
	CB	a_k	0	1	2	3	5	7	8	10	11	13	14	16	18	19
		r_k	2	3	4	5	7	9	10	12	13	15	16	18	20	20

In the case when more rapid tool wear makes the product quality level poorer, assume the specified LTPD is 0.1 with type II error probability β of 0.1. Let us consider two cases, i.e., the dependency factor q is 0 and 0.01, which corresponds to independent and dependent cases respectively.

To compare the multiple sampling plans with the sequential sampling plans, the first sample numbers i for the test to be rejected or accepted when the number of defectives is u_i or l_i of the sequential sampling plans are summarized in Table 1. For an example, when q=0.01, if no defective item(l_i=0) is found until the sample number 26, or 1 defective item(l_i=1) is found between the sample number 27 and 62, etc., then the test is accepted.

The first sample number becomes smaller for fixed u_i or the number of defectives becomes larger for the same sample number as the dependency factor q deviates farther from 0. This means that if more dependency exists between the process data, then the more defectives, consequently more samples, are required for the test to be rejected. In case of acceptance of the test, however, this situation is different.

The first sample number becomes larger for fixed li or the number of defectives becomes smaller for the same sample number as the dependency factor q increases. This means that for the test to be accepted fewer defectives are needed as dependency increases. Therefore, we can conclude that if the process is more dependent, it is more difficult to finish the test early as more samples are required. <Figure 1> shows the graph of decision boundaries, i.e., acceptance and rejection lines. In the graph, the upper 2 lines are acceptance lines and the lower 2 lines are rejection lines.

Lines of the same dependency have the same symbol marks in the graph. <Figure 1> shows that the boundary line gap between the acceptance and rejection lines becomes wider if q increases. This means that the more dependent the process, the chance of test continuation at each sample number

Table 3. The estimated type I and II error probabilities, and measures of performance

Stage Size n	q	Scheme	α	β	AQL			LTPD		
					ASN	ATI	AOQ	ASN	ATI	AOQ
	0.00	SEQ	4.86	9.80	28.5	47.2	.883	17.7	363.7	.912
	0.01		6.14	10.13	36.1	59.2	.851	25.0	363.0	.928
1	0.00	INI	4.90	9.67	28.6	47.5	.882	17.8	364.1	.900
		NA	9.98	9.99	25.2	64.1	.841	12.9	362.9	.931
		NB	9.97	9.99	25.2	64.1	.841	12.9	362.9	.931
		CA	9.96	9.50	25.7	64.7	.839	13.0	364.7	.885
		CB	9.97	9.99	25.2	64.1	.841	12.9	362.9	.931
	0.10	INI	6.55	11.19	35.9	59.5	.853	25.7	360.9	.986
		NA	8.99	9.82	49.1	82.8	.793	21.3	364.7	.887
		NB	8.89	10.00	49.1	82.6	.794	21.3	364.4	.894
		CA	8.89	10.00	49.2	82.8	.793	21.3	364.4	.894
		CB	8.89	10.00	49.2	82.7	.793	21.3	364.4	.894
10	0.00	INI	2.81	2.73	49.1	59.5	.853	28.5	390.5	.239
		NA	8.34	4.86	36.3	66.6	.853	23.6	382.2	.446
		NB	4.18	9.52	32.4	48.0	.881	23.6	365.5	.868
		CA	8.28	4.95	36.1	66.3	.835	23.6	382.0	.453
		CB	4.18	9.52	32.4	48.0	.881	23.6	365.5	.868
	0.10	INI	4.92	4.94	60.4	78.2	.814	38.8	386.1	.354
		NA	9.99	8.91	34.6	71.2	.821	24.1	367.8	.809
		NB	8.80	9.98	33.6	65.9	.835	24.0	363.8	.909
		CA	8.83	9.95	33.6	66.0	.834	24.0	363.9	.907
		CB	8.80	9.98	33.6	65.9	.835	24.0	363.8	.909
30	0.00	INI	5.23	0.19	79.0	97.6	.758	37.1	399.4	.015
		NA	6.06	4.90	40.2	61.3	.848	34.8	382.0	.451
		NB	4.77	4.99	39.1	56.3	.861	34.9	381.8	.458
		CA	6.04	4.90	40.1	61.2	.848	34.8	382.0	.451
		CB	4.79	4.98	39.1	56.3	.860	34.9	381.8	.458
	0.10	INI	8.21	2.00	93.9	122.8	.707	45.8	396.6	.086
		NA	9.79	7.97	38.1	71.7	.820	36.1	371.0	.729
		NB	7.01	9.42	36.9	62.2	.844	36.3	366.6	.839
		CA	7.50	8.61	37.4	64.2	.839	36.2	369.0	.778
		CB	7.44	8.64	37.4	64.0	.840	36.2	369.0	.780

becomes higher. In other words, it is unlikely to finish the test early if the process is more dependent. Note that the independent case(q=0) has a narrower gap and the decision boundary lines are parallel. The dependent case does not have parallel slopes. Consequently, the average sample size of the dependent process necessarily becomes larger as q increases. The acceptance numbers a_k and rejection numbers r_k, at each stage k according to the dependency factor q and the search scheme when sample size n is 30, are shown in <Table 2>. In the table, INI, NA, NB, CA, and CB represent the initial plan, the nearest α, the nearest β, the central α ,and the central β search, respectively. The negative value -1 in the table means that the acceptance decision cannot be made at that stage. In the dependent

Table 4. The acceptance and rejection probabilities in percent (n=30, q=0.00)

	Stage k		1	2	3	4	5	6	7	8	9	10	11	12	13	14
	kn		30	60	90	120	150	180	210	240	270	300	330	360	390	400
INI	AQL	α_k	.00	54.3	23.9	10.3	3.9	2.2	.19	.05	.01	.01				
		β_k	3.52	1.22	.34	.14	.00	.01								
	LTPD	α_k	.00	.13	.02	.03	.01									
		β_k	81.6	14.2	3.1	.77	.10	.05	.01	.01						
NA	AQL	α_k	73.2	16.8	.00	3.13	.59	.23								
		β_k	3.52	.98	1.40	.13	.03									
	LTPD	α_k	4.35	.53	.00	.02										
		β_k	81.6	11.5	1.91	.08										
NB	AQL	α_k	73.2	16.8	4.11	.87	.24	.01	.01							
		β_k	3.52	.98	.20	.06	.00	.01								
	LTPD	α_k	4.35	.53	.10	.00	.01									
		β_k	81.6	11.5	1.6	.21	.04	.02								
CA	AQL	α_k	73.2	16.8	.00	3.13	.82	.02								
		β_k	3.52	.98	1.40	.13	.00	.01								
	LTPD	α_k	4.35	.53	.00	.02										
		β_k	81.6	11.5	1.91	.08										
CB	AQL	α_k	73.2	16.8	.00	3.13	.82	.02								
		β_k	3.52	.98	1.40	.13	.00	.01								
	LTPD	α_k	4.35	.53	.00	.02										
		β_k	81.6	11.5	1.91	.08										

case(q=0.01), the gaps between a_k and r_k of the INI plan are larger than those of the independent case(q=0.00). But the gaps after applying the search scheme are all 2, which is the smallest allowable value for the test to be continued in order to make the ASN as small as possible. In the dependent case, the NA search selected smaller a_k and r_k than the other searches; the NB search selected larger ones; and the CA and CB searches selected ones similar to each other. In the independent case, however, all four searches selected similar ones since the gap of the INI plan was not large enough.

<Table 3> shows the estimated α and β, ASN, AOQ without replacement of defectives, and ATI for the sequential and multiple sampling plans according to the search scheme when the stage size n are 1, 10 and 30. When stage size n is 1, the multiple sampling plan becomes a sequential sampling plan. Note that the estimated α and β of the four searches are more accurate than the original sequential sampling plans.

The accuracy of α and β decreases as the stage size n increases, because more stages are needed to increase accuracy. In general, if the gap between the two boundary lines is large, these error probabilities can be reduced. If the gap is large, there is more chance to narrow the gap, which means the reduction of error probabilities. Note that, when n is 1 and q is 0.00, the ASN is smaller than the initial sequential sampling plan. In this case, it is possible to make more accurate and efficient sequential sampling plans than Wald's SPRT plan by applying the search scheme. In the independent case, the ASN increases as the stage size n increases, but in

Table 5. The acceptance and rejection probabilities in percent (n=30, q=0.01)

	Stage k		1	2	3	4	5	6	7	8	9	10	11	12	13	14
	kn		30	60	90	120	150	180	210	240	270	300	330	360	390	400
INI	AQL	α_k	.00	62.3	.00	12.2	6.33	3.46	2.07	1.20	.79	.60	.44	.26	.20	1.98
		β_k	5.17	2.15	.25	.45	.07	.03	.03	.02	.03	.00	.00	.01		
	LTPD	α_k	.00	.79	.00	.06	.05	.09	.06	.05	.01	.01	.01	.01	.00	.86
		β_k	77.2	13.9	2.54	2.07	.60	.38	.45	.22	.25	.05	.18	.04	.04	.03
NA	AQL	α_k	76.7	12.1	.00	.00	1.02	.26	.08	.02						
		β_k	5.17	1.64	1.90	.95	.09	.02	.01	.01						
	LTPD	α_k	6.46	1.51												
		β_k	77.2	11.4	2.88	.47	.03									
NB	AQL	α_k	76.7	12.1	3.94	.11	.10	.01								
		β_k	5.17	1.64	.13	.06	.01									
	LTPD	α_k	6.46	1.51	1.20	.11	.11	.02	.01							
		β_k	77.2	11.4	1.27	.56	.07	.03								
CA	AQL	α_k	76.7	12.1	2.46	.73	.32	.14	.01	.00	.01					
		β_k	5.17	1.64	.42	.19	.08									
	LTPD	α_k	6.46	1.51	.50	.11	.01	.02								
		β_k	77.2	11.4	2.18	.47	.06	.02	.01							
CB	AQL	α_k	76.7	12.1	2.46	.73	.48	.06								
		β_k	5.17	1.64	.42	.19	.02									
	LTPD	α_k	6.46	1.51	.50	.11	.06									
		β_k	77.2	11.4	2.18	.47	.04	.00	.01	.00	.01					

the dependent case, this is not always true. When q is 0.01, the ASN of n is 1 is smaller than that of n is 10. In general, the ASN is affected by many factors such as stage size, search schemes, depending factors etc.

<Table 4> shows the acceptance probability α_k and rejection probability β_k of the test at each stage k according to the search scheme when n is 30 and q is 0.00. The sum $\alpha_k+\beta_k$ is the probability that the test will finish at stage k.

<Table 5> is the same as Table 4 except q is 0.01. Most search schemes reduce the number of test stages by assigning the largest probability to the earlier stages. Especially, the first stage's α_k of the search schemes at AQL has larger value compared to the INI plan, which enables reducing the test stages.

6. Conclusion

In this paper, we have developed a design procedure for nonstationary multiple sampling plans of a

dependent process. By extending the multiple sampling plans developed in this paper to the sequential sampling plans, more accurate sequential sampling plans can be acquired. In Polya dependent processes, the distribution of the number of defectives can be found by analytical method. So we do not have to use simulation. But to verify the algorithm developed in this paper, the results coming from using simulation and analytical methods are compared, and they show almost the same results. The main advantage of this study is that if the pattern of dependence of the process can be found, then we can simulate the process, and can establish optimal sampling plans.

One important point in multiple sampling plans is to decide the optimal sample size. The stage size is not given as a decision variable but as a parameter. The performance of a sampling plan varies according to the stage size. The decision may depend on what the objective function is. The type I and II error probabilities are rather inaccurate when the stage size becomes large. To solve this, we may consider the variable stage size instead of a fixed stage size. In a traditional independent multiple sampling plan, the number of stages is around seven and the stage size is fixed. But in the dependent case, the number of stages becomes larger than that of independent.

How to find the existence of dependency and what the process model is are problems when applying the multiple sampling procedure to the real world. The study for the procedure of identifying the process model and its parameters is not addressed in this paper. However, this work must be done before the actual applications. This can be established by using some commercial statistical package and by simulating the production process. For further study, the sensitivity analysis for the decision variables or parameters, and robustness for sampling methods are required.

Reference

1. Bhat, U. Narayan and Lal, Ram, *Number of Successes in Markov Trials*, Advances in Applied Probability, 20, pp. 677.
2. Bhat, U. Narayan and Lal, Ram and Mahinda Karunaratne, "A sequential inspection plan for markov dependent production processes," *IIE Transactions*, Vol. 22, No. 1, March 1990.
3. Broadbent, S. R., "The inspection of a markov process," *Journal of Royal Statistics Society*, B20, pp. 111-119, 1958.
4. Frielander, B. "Lattice filters for adaptive processing," *Proceedings of the IEEE*, Vol. 70, pp. 829-867, 1982
5. Kim, W. K., "Design and estimation of double sampling plans for the dependent production processes," *Journal of Korean Institute of Industrial Engineering*, Vol. 23, No. 2, pp. 289-305, 1997.
6. Kim, W. K., "Stochastically dependent sequential acceptance sampling plans," *Journal of Korean Society for Quality Management*, Vol. 25, No. 3, pp. 22-38, 1997.
7. Lee, J., Park, C. and Park, J., "Sequential sampling inspection plans for defectives," *Journal of Korean Society for Quality Management*, Vol. 24, No. 4, pp. 1-13, 1996.
8. Lorden, G., "On excess over the boundaries," *Annals of Mathematical Statistics*, Vol. 41, No. 2, pp. 520-727, 1970.
9. Nelson, B., "Estimating acceptance-sampling plans for dependent production processes," *IIE Transactions*, Vol. 25, No. 6, pp. 11-18, 1993.
10. Sarkadi, K. and Vincze, I., *Mathematical Methods of Statistical Quality Control*, Academic Press, New York, 1974.
11. Siegmund, D., "Boundary crossing probabilities and statistical applications," *Annals of Statistics*, Vol. 14, No. 2, pp. 361-404, 1986.
12. Wald, A., *Sequential Analysis*, John Wiley, New York, 1947.
13. Wald, A. and Wolfowitz, "Optimum character of the sequential probability ratio test," *Annals of Mathematical Statistics*, Vol. 19, pp. 326-339, 1948.

Received March 1998, accepted December 1998