# On sequential versus random sampling in statistical process control

Lance A. Matheson
*Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA*

## Introduction

Statistical process control charts are widely used to make inferences about the level of process quality; i.e., we monitor an unreliable process by taking a sample of items produced by the machine during each $h$ time units (we will refer to $h$ time units as an inspection period) and use this information to test the hypothesis that the machine remains in an acceptable in-control state. In this paper, we analyse and compare two schemes for sampling from a Poisson population: a random sample of $n$ items produced during an inspection period, and a sequential sample of the last $n$ items produced during an inspection period (it is interesting to note that the sequential sampling scheme analysed in this article is related to the concept of rational subgroups first discussed by [1]). Although random samples are often used[2,3], we will show that a sequential sample is more likely to detect a process shift and thereby results in fewer expected samples and less time to detect the shift.

Specifically, we analyse an unreliable machine which produces $\theta$ units per time unit and can shift from an acceptable in-control state to an unacceptable out-of-control state. When operating in the in-control state, the process produces defects according to a Poisson distribution at an average rate of $\Lambda_0$ defects per unit. When the process shifts to the out-of-control state, it produces defects at an average rate of $\Lambda_1$ (where $\Lambda_1 > \Lambda_0$) and requires some physical action (e.g. repair) to reset it to the in-control state. We also assume that the time between the start of the process in the in-control state and the process shift is exponentially distributed with mean $1/\lambda$.

We assume that the process is monitored by counting the number of defects in a sample of $n$ items and recording the data on a Shewhart-type c-chart (although our results would improve the performance of all types of control chart because the time to detect a shift is reduced). At the end of each inspection period we assume that we immediately test the hypothesis that the process remains in control; that is, we assume that inspection time for the sampled items is negligible. Given these assumptions, we compare two sampling schemes:

The author acknowledges the assistance of T.D. Klastorin and Kamran Moinzadeh in the development of this article.

a sample of $n$ items randomly selected from all items produced during an inspection period (such that each item has a selection probability of $1/(h\theta)$ ), and a sequential sample consisting of the last $n$ items produced during an inspection period.

The two sampling schemes differ only in the inspection period in which the process shift occurs. Samples in all other inspection periods will contain $n$ items produced at the low defect rate before the shift or the high defect rate after the shift (regardless of how the $n$ items are selected). Thus, the probability of a false alarm is the same for both sampling schemes in all inspection periods where the shift does not occur. We must therefore only consider the inspection period in which the process shift occurs in order to compare the two sampling schemes.

The case in which a random sample is used was previously studied in [4]. In the present paper, we extend the previous framework to the case in which a sequential sample is used. We will show that in the situation previously described, the probability of not detecting the shift with a random sample is always greater than or equal to the probability of not detecting the shift when a sequential sample is used. That is, sequential sampling minimizes the expected number of samples needed to detect a process shift.

## The sampling process

At the end of an inspection period consisting of $h$ time units, we use the $n$ sampled items to test the hypothesis that the process is in control. If the number of defects in the sample is less than or equal to some upper control limit (UCL), we accept the null hypothesis that the process is in control and we begin another inspection period (we assume here that the lower control limit (LCL) is zero). If the number of defects in the sample exceeds the UCL, we stop the process and search for and remove the assignable cause if one is found (there may be false alarms). Production resumes when the process is restored to the in-control state.

Consider the inspection period in which the Poisson process shifts to an out-of-control state. The expected number of samples taken after this shift occurs before an out-of-control signal is given will be denoted by $\hat{s}$ and can be shown to be[5]:

$$\hat{s} = (1 - \beta)^{-1} \tag{1}$$

where :

$$\beta = \sum_{i=0}^{\mathrm{UCL}} (n\Lambda_1)^i \frac{e^{-n\Lambda_1}}{i'} \tag{2}$$

is the probability of not detecting the shift in a given sample.

The value of $\hat{s}$ given by (1) assumes that $\beta$ is constant for all samples; i.e. it assumes that no shift occurs while a sample is being taken. In other words, it assumes that all items sampled during the inspection period in which the shift occurs are produced at the same out-of-control rate, $\Lambda_1$. For the inspection period during which the shift occurs, however, it is possible that a sample could contain some units which were produced at the lower defect rate, $\Lambda_0$, as well as items produced at the higher defect rate, $\Lambda_1$.

We will use $\beta_0$ to denote the probability that the shift is not detected by the sample taken during the inspection period in which the process shift occurs, and we will use $\beta$ to denote the corresponding probability for samples taken during succeeding inspection periods. Each sample represents an independent Bernoulli trial of the hypothesis that the process is still in control. Thus, letting $S$ denote the number of samples taken between the shift and its detection, the probability that we detect the shift on the $s$th trial while failing to detect the shift on the previous $s-1$ trials is:

$$\Pr(S = s) = \begin{cases} 1-\beta_0 & \text{for } s = 1, \\ \beta_0(1-\beta)\beta^{S-2} & \text{for } s > 1, \end{cases}$$

where $\beta$ is given by (2). The expected number of samples, $\hat{s}$, necessary to detect the shift can easily be computed ([3]) and is given by:

$$\hat{s} = 1 + \frac{\beta_0}{1-\beta}. \qquad (3)$$

*Value of $\beta_0$ for random and sequential sampling*

Recall that the time to a process shift is exponentially distributed with mean $1/\lambda$. When a random sample is used, we can define

$\Pr(i) =$ Pr[exactly the first $i$ units produced in this inspection period belong to the in-control state | the process shifts during the inspection period]

$$= \frac{(1-e^{-\lambda/\theta})e^{-i\lambda/\theta}}{1-e^{-\lambda h}} \text{ for } i = 0, 1, \dots h\theta - 1, \qquad (4)$$

$$\beta_0 = \sum_{k=0}^{n} \left\{ \sum_{i=0}^{\text{UCL}} \sum_{j=0}^{i} \Pr\begin{pmatrix} j \text{ defects from } k \text{ sampled} \\ \text{units which were} \\ \text{produced before the shift} \end{pmatrix} \Pr\begin{pmatrix} i-j \text{ defects from } n-k \\ \text{sampled units produced} \\ \text{after the shift} \end{pmatrix} \right\} \Pr[k] \quad (5)$$

where Pr[$k$] is the probability that exactly $k$ units in the sample were produced before the shift occurred. Then, for a random sample,

$$\Pr[k] = \sum_{i=k}^{\min([h\theta-1, h\theta-n+k)} \frac{\binom{i}{k}\binom{h\theta-i}{n-k}}{\binom{h\theta}{n}} \Pr(i) \quad \text{for } k = 0, 1, 2, \dots, n, \qquad (6)$$

where Pr($i$) is given by (4). Evaluating equation (5) requires adding independent Poisson distributions. Since the sum of independent Poisson distributions with parameters $\lambda_i$ follows a Poisson distribution with parameter $\Sigma\lambda_i$, (5) can be simplified to:

$$\beta_0^r = \sum_{k=0}^{n} \Pr[\text{UCL}, k\Lambda_0 + (n-k)\Lambda_1]\Pr[k] \quad (7)$$

where $\beta_0^r$ denotes the value of $\beta_0$ when a random sample is used and Pr[UCL, $x$] represents the value of the cumulative distribution function of a Poisson variable with mean equal to $x$, evaluated at UCL.

This approach can be extended to a sequential sample of the last $n$ items produced in the inspection period. In this case, the probability that $k$ units in the sample were produced before the shift is equal to the probability that $(h\theta - n + k)$ units were produced in the inspection period before the shift occurred; that is, for $k > 0$, Pr[$k$] = Pr($h\theta - n + k$) where Pr($h\theta - n + k$) is defined by (4). Applying this observation to (5), using the same argument about sums of independent Poisson distributions, and letting $\beta_0^s$ denote the probability of not detecting the shift when a sequential sample is used:

$$\beta_0^s = \Pr[\text{UCL}, n\Lambda_1] \sum_{i=0}^{h\theta - n} \Pr(i)$$
$$+ \sum_{i=h\theta-n+1}^{h\theta-1} \Pr[\text{UCL}, (i - (h\theta - n))\Lambda_0 + (h\theta - i)\Lambda_1]\Pr[i] \quad (8)$$

where Pr($i$) is given by (4). The first term in (8) defines the probability when all $n$ items in the sample are produced after the shift; the second defines the probability when the sample contains a mixture of items produced before and after the shift.

## Comparison of random and sequential sampling

Given the definitions of $\beta_0^s$ and $\beta_0^r$, we can show that a sequential sampling plan is more likely to detect a process shift than a random sample. This is proven in the following theorem which shows that $\beta_0^s \leq \beta_0^r$, for the inspection period where the process shift occurs.

*Theorem*. Given a process with production rate $\theta$ and two states in which defects have a Poisson distribution with average defect rates of $\Lambda_0$ and $\Lambda_1$, respectively, then $\beta_0^s \leq \beta_0^r$, for all samples of size $n \leq h\theta$.

*Proof*. To simplify the notation, let

$$F(k) = \Pr[\text{UCL}, k\Lambda_0 + (n - k)\Lambda_1]$$

for given values of $n$, UCL, $\Lambda_0$ and $\Lambda_1$. Based on the definition of the cumulative Poisson distribution,

$$F(0) < F(1) < F(2) < \dots < F(n-1) < F(n).$$

The definition of $\beta_0^r$ in (7) can then be rewritten as follows:

$$\beta_0^r = F(0)\Pr[0] + F(n)\Pr[n] + \sum_{k=1}^{n-1} F(k)\Pr[k]. \quad (9)$$

Similarly, the definition of $\beta_0^s$ in (8) can be rewritten as:

$$\beta_0^s = F(0)\sum_{i=0}^{h\theta-n} \Pr(i) + \sum_{i=h\theta-n+1}^{h\theta-1} F(i-h\theta+n)\Pr(i). \quad (10)$$

Without loss of generality, we can set $\lambda = 1$, let $h\theta = (n+\delta)$ for $\delta = 0, 1, 2, \dots$ , and redefine $\beta_0^s$ and $\beta_0^r$, as functions of $\delta$. (Since we cannot sample more units than produced during an inspection interval, $h\theta \geq n$; $h\theta$ and $n$ must also be integer quantities). Then, we define:

$$g(\delta) = \frac{[\beta_0^r(\delta) - \beta_0^s(\delta)]\,[1 - e^{-\lambda h}]}{1 - e^{-\lambda/\theta}};$$

clearly, $\beta_0^r \geq \beta_0^s$, if we can show that $g(\delta) \geq 0$. Substituting $h\theta = (n+\delta)$ into (9) and (10),

$$g(\delta) = F(0)\left\{\sum_{i=0}^{\delta} e^{-i/\theta}\left[\frac{\binom{n+\delta-i}{n}}{\binom{n+\delta}{n}} - 1\right]\right\} + F(1)\left\{\sum_{i=1}^{\delta+1}\left[\frac{\binom{i}{1}\binom{n+\delta-i}{n-1}}{\binom{n+\delta}{n}} e^{-i/\theta}\right] - e^{-(\delta+1)/\theta}\right\}$$

$$+ F\left\{\sum_{i=2}^{\delta+2}\left[\frac{\binom{i}{2}\binom{n+\delta-i}{n-2}}{\binom{n+\delta}{n}} e^{-i/\theta}\right] - e^{-(\delta+2)/\theta}\right\} + \dots$$

$$+ F(n-1)\left\{\sum_{i=n-1}^{n+\delta-1}\left[\frac{\binom{i}{n-1}}{\binom{n+\delta}{n}} e^{-i/\theta}\right] - e^{-(n+\delta-1)/\theta}\right\} + F(n)\sum_{i=n}^{n+\delta-1}\left[\frac{\binom{i}{n}}{\binom{n+\delta}{n}} e^{-i/\theta}\right]. \quad (11$$

If $\delta = 0$ and $h\theta = n$, it follows from (11) that $g(0) = 0$; that is, if we sample every unit produced during the inspection period, the random sampling and the sequential sampling schemes are obviously identical with equal values of $\beta_0$. However, if $\delta = 1$ such that $h\theta = n + 1$,

$$g(1) = -F(0)\left(\frac{n}{n+1}\right)e^{-i/\theta} + F(1)\left\{\sum_{i=1}^{2}\left[\frac{\binom{i}{1}\binom{n+1-i}{n-1}}{n+1}e^{-i/\theta}\right] - e^{-2/\theta}\right\}$$

$$-1 + F(2)\left\{\sum_{i=2}^{3}\left[\frac{\binom{i}{2}\binom{n+2-i}{n-2}}{n+1}e^{-i/\theta}\right] - e^{-3/\theta}\right\} + \ldots + F(n-1)\left\{\sum_{i=n-1}^{n}\left[\frac{\binom{i}{n-1}}{n+1}e^{-i/\theta}\right] - e^{-n/\theta}\right\}$$

$$+ F(n)(n+1)^{-1}e^{-n/\theta}$$

$$= F(0)\left(\frac{-n}{n+1}\right)e^{-1/\theta} + F(1)\left[\left(\frac{n}{n+1}\right)e^{-1/\theta} - \left(\frac{n-1}{n+1}\right)e^{-2/\theta}\right] + F(2)$$

$$\left[\left(\frac{n-1}{n+1}\right)e^{-2/\theta} - \left(\frac{n-2}{n+1}\right)e^{-3/\theta}\right] + F(3)\left[\left(\frac{n-2}{n+1}\right)e^{-3/\theta} - \left(\frac{n-3}{n+1}\right)e^{-4/\theta}\right]$$

$$+ \ldots + F(n-1)\left[\left(\frac{1}{n+1}\right)e^{-(n-1)/\theta} - \left(\frac{1}{n+1}\right)e^{-n/\theta}\right] + \frac{F(n)}{n+1}e^{-n/\theta}$$

$$= \left(\frac{n}{n+1}\right)e^{-1/\theta}[F(1) - F(0)] + \left(\frac{n-1}{n+1}\right)e^{-2/\theta}[F(2) - F(1)]$$

$$+ \left(\frac{n-2}{n+1}\right)e^{-3/\theta}[F(3) - F(2)] + \ldots + \left(\frac{n}{n+1}\right)e^{-n/\theta}[F(n) - F(n-1)].$$

Since $F(n) > F(n-1) > \ldots > F(1) > F(0)$ from the definition of the cumulative Poisson distribution and $e^{-n/\theta} > 0$ for any $n \geq 0$, it follows that $g(1) > 0$. Furthermore, given that:

$$e^{-(\delta+k)/\theta} \to 0 \text{ and } \sum_{i=k}^{\delta+k}\frac{\binom{i}{k}\binom{n+\delta-i}{n-k}}{\binom{n+\delta}{n}}e^{-i/\theta} \to \infty \text{ as } \delta \to \infty$$

(for any given $k > 1$), it follows that $g(\delta)$ defined in (11) will always be greater than $g(1)$ for any value of $\delta > 1$. Since $g(1) > 0$, $g(\delta) > 0$ for any value of $\delta = 2, 3, \ldots$; thus, $\beta_0^s \leq \beta_0^r$. Q.E.D.

It is interesting to note that this theorem holds for any distribution where $F(n) > F(n-1) > \ldots > F(1) > F(0)$. Thus, we conclude that a sequential sampling scheme is superior to a random sampling scheme for c charts.

**Numerical example**

An important performance measure of quality control charts is the expected number of samples, $\hat{s}$ required to detect a process shift. Since $\beta_0^s \le \beta_0^r$, it is evident from the definition of $\hat{s}$ in (3) that $\hat{s}$ will be greater when random sampling is used than when sequential sampling is used. To illustrate the magnitude of this difference and test the sensitivity of this difference to changes in parameter values, we analysed an example based on typical values reported in the literature; these parameter values are given in Table I.

| Parameter | Definition | Values |
|---|---|---|
| $\lambda$ | 1/Expected time to machine failure | 0.02 |
| $\theta$ | Production rate | 100 |
| $\Lambda_0$ | In-control defect rate | 0.02 |
| $\Lambda_1$ | Out-of-control defect rate | 0.1 |
| $h$ | Inspection interval | 3 |
| $n$ | Sample size | 20 |
| UCL | Upper control limit | 1 |
| *Results* | | |
| $\beta_0^r = 0.664$ | | |
| $\beta_0^s = 0.422$ | | |
| $\hat{s}$ (random sampling) = 2.12 samples | | |
| $\hat{s}$ (sequential sampling) = 1.71 samples | | |

**Table I.** Example parameter values and results

As indicated in Table I, we assume that the process operates at an in-control defect rate ($\Lambda_0$) of 0.02 defects per unit, and shifts to an out-of-control defect rate ($\Lambda_1$) of 0.1 defects per unit. Using the definition of $\hat{s}$ given by (3), sequential sampling results in a value $\hat{s} = 1.71$, while $\hat{s}$ increases to 2.12 samples when random sampling is used. Thus, sequential sampling requires 23.4 per cent fewer expected samples, on average, to detect a shift in the process in this example. (It is interesting to note that the value of $\hat{s}$ calculated by (1) is 1.68; this widely-used definition of $\hat{s}$ always results in a value of $\hat{s}$ which understates the true expected value.)

To investigate the sensitivity of the difference in $\hat{s}$ caused by changes in parameter values, we increased and decreased each parameter by 10 and 40 per cent (while holding all other parameter values constant). The upper control limit, UCL, being necessarily integral, was set at 1 or 2, resulting in a total of 48 cases. In all cases, we computed the percentage difference in $\hat{S}$ resulting from random sampling and sequential sampling schemes.

Our results for UCL = 1 are summarized in Table II. As indicated, the parameters with the greatest impact on the difference in $\hat{S}$ were sample size ($n$), out-of-control defect rate ($\Lambda_1$), and upper control limit (UCL). The average overall reduction in $\hat{S}$ was 23.25 per cent when UCL was equal to 1, and 16.80 per cent when UCL was equal to 2. It is interesting to note that the average per cent difference in $\hat{S}$ (denoted by per cent saved) was slightly smaller when the parameters were increased by 40 per cent; that is, the average value of per cent saved was 20.36 per cent with a 10 per cent change in parameter values, while the average value of per cent saved was 19.69 per cent when the parameters were increased or decreased by 40 per cent. In general, the results suggest that the design of a control chart is most sensitive to the choices of sample size, control limits and the accuracy with which $\Lambda_1$ is determined.

**Table II.**
Sensitivity analysis when UCL = 1

| Run | Parameter | Per cent change | $\beta_0^r$ | $\beta_0^s$ | $\beta$ | $\hat{S}_{ran}$ | $\hat{S}_{seq}$ | Per cent saved |
|---|---|---|---|---|---|---|---|---|
| 1 | $\lambda$ | ±10 | 0.664 | 0.422 | 0.406 | 2.12 | 1.71 | 23.79 |
| 2 | | ±40 | 0.664 | 0.422 | 0.406 | 2.12 | 1.71 | 23.79 |
| 3 | $\theta$ | ±10 | 0.664 | 0.422 | 0.406 | 2.12 | 1.71 | 23.77 |
| 4 | | ±40 | 0.664 | 0.425 | 0.406 | 2.12 | 1.72 | 23.40 |
| 5 | $\Lambda_0$ | ±10 | 0.664 | 0.422 | 0.406 | 2.12 | 1.71 | 23.78 |
| 6 | | ±40 | 0.663 | 0.422 | 0.406 | 2.12 | 1.71 | 23.72 |
| 7 | $\Lambda_1$ | ±10 | 0.664 | 0.424 | 0.406 | 2.14 | 1.73 | 23.61 |
| 8 | | ±40 | 0.665 | 0.461 | 0.409 | 2.55 | 2.19 | 23.76 |
| 9 | $h$ | ±10 | 0.664 | 0.422 | 0.447 | 2.12 | 1.71 | 23.76 |
| 10 | | ±40 | 0.662 | 0.419 | 0.406 | 2.15 | 1.71 | 23.99 |
| 11 | $n$ | ±10 | 0.664 | 0.425 | 0.409 | 2.14 | 1.73 | 23.59 |
| 12 | | ±40 | 0.691 | 0.480 | 0.465 | 2.60 | 2.19 | 23.74 |

## Conclusions

In this article, we have discussed the differences between random sampling and sequential sampling schemes and demonstrated that sequential sampling generally requires fewer samples to detect a process shift. The economic significance of this difference can be substantial over a long time horizon, even when the rework/repair cost per defect is relatively small.

Sequential sampling provides other benefits beyond reduced scrap/rework costs. A sequential sampling process is easier to implement; random sampling requires operators to select randomly and inspect items from the production process, while an operator just collects and inspects the last $n$ items produced using a sequential sampling scheme. Once a shift is detected, the last items produced should be inspected to ensure that they do not contain an excessive number of defects. These items require storage space and inspection time. Since sequential sampling detects shifts more quickly, fewer items will need inspection and storage space. Given that no capital expenditure is required to implement a sequential sampling process, it would appear to be a worthwhile consideration.

**References**

1. Shewhart, W.A., *Economic Control of Quality of Manufactured Product*, D. Van Nostrand, New York, NY, 1931.
2. Grant, E.L. and Leavenworth, R.S., *Statistical Quality Control*, McGraw-Hill, New York, NY, 1980.
3. Montgomery, D.C., *Introduction to Statistical Quality Control*, John Wiley & Sons, New York, NY, 1985.
4. Klastorin, T.D., Matheson, L.A. and Moinzadeh, K., "On the use of buffer inventories to minimize costs with an unreliable machine", *IIE Transactions*, Vol. 25 No. 5, 1993, pp. 50-62.
5. Duncan, A.J., "The economic design of $\bar{X}$-charts used to maintain current control of a process", *Journal of the American Statistical Association*, Vol. 51 No. 274, 1956, pp. 228-42.