

Cumulative Sum (CUSUM) Control Charts

Reference: Introduction to STATITICAL QUALITY CONTRL 5e (and 4th edition) by D. C. Montgomery (2005), Wiley, Inc.

The cumulative sum control chart, or Cusum Chart, is primarily used to detect a gradual shift (or trend) in a process and can be even more powerful in detecting sudden shifts in a manufacturing process than a Shewhart control chart. We will first discuss the one-sided procedure, which is the simpler case, followed by a 2-sided test of hypothesis $H_0: \mu = \mu_0$ versus the alternative $H_1: \mu \neq \mu_0$. In the one-sided procedure, the alternative values of process mean are denoted either by $\mu_1 < \mu_0$ (the left-tailed test), or by $\mu_2 > \mu_0$ for the right-tailed test. However, before heading long into the discussion of either procedure, we allude to fact that the statistical theory of Cusum Charts has its origin in sequential analysis. In carrying out sequential testing, generally samples of size n ($=1$, but not always) is taken at each stage of the sampling plan, the Cusum statistic is updated, and a 3-way decision is made as follows: (1) accept H_0 , (2) reject H_0 , or (3) continue the experiment by taking additional observation(s). I would venture to say that it is not common to take differing sample sizes at different stages in SPC, although this is sometimes done in acceptance sampling for FNC (specially in a 2-stage sampling plan where $n_2 = 2 \times n_1$ but then in this case the decision about a submitted lot is definitely made after stage 2 because the difference between the acceptance and rejection numbers at the 2nd stage is always equal to 1). For example, the double sampling plan $n_1 = 50$, $c_1 = 1$, $n_2 = 100$, and $c_2 = 4$ implies that a random sample of size 50 is taken from a large lot (usually $N > 1000$ units) and the lot is immediately accepted if the number of defectives, D_1 , at stage 1 is ≤ 1 and the lot is immediately rejected at stage 1 if $D_1 > 4$. If the lot quality is mediocre at stage 1, where $D_1 = 2, 3$, or 4, then the lot is given a second

chance by taking an additional sample of size $n_2 = 100$. At the 2nd stage, the lot is accepted only if the cumulative number of defectives in the 150 sampled units is 4 or less, and it is rejected at stage 2 only if $D_1 + D_2 > 4$, i.e., the rejection number $r = 5$.

When sequential sampling is applied in SPC, first the reader must be cognizant of the fact that the inspection is not always by attributes and that the decision about the mean or variance of a normal process is carried out with a statistic whose range space is on a continuous (or dense) scale. The very basic statistical idea behind testing the hypothesis $H_0: \mu = \mu_0$ versus the alternative $H_1: \mu = \mu_1 < \mu_0$ is to make use of the Neyman-Pearson Lemma by using the likelihood ratio statistic

$$\lambda_t = \frac{\prod_{i=1}^t f(x_i; \mu_1, \sigma^2) dx_i}{\prod_{i=1}^t f(x_i; \mu_0, \sigma^2) dx_i} = \frac{\prod_{i=1}^t f(x_i; \mu_1, \sigma^2)}{\prod_{i=1}^t f(x_i; \mu_0, \sigma^2)} \quad (11)$$

In equation (11) the statistic in the numerator on the RHS is related to the occurrence Pr of the sample (x_1, x_2, \dots, x_t) given that $\mu = \mu_1$ because the likelihood (or Pr) of obtaining such a sample is given by $[f(x_1; \mu_1, \sigma^2) dx_1] \times [f(x_2; \mu_1, \sigma^2) dx_2] \times \dots$

$\times [f(x_t; \mu_1, \sigma^2) dx_t] = \prod_{i=1}^t [f(x_i; \mu_1, \sigma^2) dx_i]$, where $f(x)$ represents the pdf of the

sampled population (or the underlying distribution). The quantity $\prod_{i=1}^t f(x_i; \mu_1, \sigma^2)$

in this last Pr statement is called the likelihood function under H_1 , and after the sample is drawn and the sample values x_1, x_2, \dots, x_t are known numbers (no

longer rvs), then the likelihood function $\prod_{i=1}^t f(x_i; \mu, \sigma^2)$ is only a function of the

parameters μ and σ^2 , which is denoted by $L(\mu, \sigma^2) = \prod_{i=1}^t f(x_i; \mu, \sigma^2)$. When the

likelihood function $L(\mu, \sigma^2)$ is maximized wrt (with respect to) μ and σ by setting the two first partial derivatives equal to zero, the resulting solutions for μ and σ

are called the maximum likelihood estimators (MLE) of μ and σ . In other words, the numerator of (11) gives the likelihood that μ lies in an extremely small interval around μ_1 and its denominator provides the Pr that μ lies in an infinitesimally small interval around μ_0 . Therefore, when λ_t is very large, we will reject H_0 in favor of H_1 at the stage (or at time) t ; when λ_t is too small, we will favor H_0 over H_1 , and when it is neither, then the sample point is still in the indecision region of sample space so that we have to continue on sampling to the stage $t + 1$.

Further, for notational convenience, let $L(\theta) = \ln[\underline{L}(\theta)]$, i.e., $L(\theta)$ is the natural logarithm of the likelihood function, and θ is a general vector parameter such as $\theta = [\mu \quad \sigma]'$. Below (pp. 80-81) I will show that working with the natural logarithm of $\underline{L}(\theta)$ is easier than using $\underline{L}(\theta)$ itself. Before providing such an example, we must provide some partial answer as to how large λ_t should be before we adopt $\mu = \mu_1$ and how small λ_t should be before we adopt $\mu = \mu_0$. In other words, we will reject H_0 if $\lambda_t \geq c_r$, will accept H_0 if $\lambda_t \leq c_a$, and we will continue sampling if the value of λ_t lies in the indecision interval $c_a < \lambda_t < c_r$, which depends on the specified values of α and the type II error Pr, denoted by β . Note that in general $c_r > c_a$.

One advantage that sequential testing has over the fixed-size sample tests is the fact that we can decide in advance on the sizes of type I and II errors (generally $\alpha \leq 0.05$ and $\beta \leq 0.10$) that we are willing to tolerate, which is unlike the fixed-size sampling, where α is specified a priori and β is computed for different parameter values. Suppose we fix α at 0.05 and β at 0.10 for detecting a downward shift in a process mean; then clearly the value of c_r must be proportional to $1-\beta$ (= the Pr of rejecting a false H_0). It was shown by A. Wald and G. A. Barnard (1945 and 1946) that the acceptance and rejection limits must

satisfy $\frac{\beta}{1-\alpha} \leq c_a$ and $c_r \leq \frac{1-\beta}{\alpha}$, respectively. The statistic λ_t is referred to as the

sequential Pr ratio test (SPRT) with approximate boundaries $c_a \cong \frac{\beta}{1-\alpha}$ and $c_r \cong$

$\frac{1-\beta}{\alpha}$. Note that the boundaries for the indecision interval is designed in such a

manner that the true type I and II Prs of the sequential test are at most α and β , respectively. The reader should observe that the denominator of the SPRT

statistic $\lambda_t = \prod_{i=1}^t f(x_i; \mu_1, \sigma^2) / \prod_{i=1}^t f(x_i; \mu_0, \sigma^2)$ gives the likelihood Pr that H_0

is true, and therefore, this Pr (i.e., the denominator) must be directly proportional to $1 - \alpha$ (= the Pr of accepting a true H_0) and inversely proportional to β (= the Pr of accepting a false H_0). This is how the expression for lower boundary c_a for the indecision interval $c_a < \lambda_t < c_r$ was constituted. Similarly, the numerator of λ_t must be directly proportional to $1-\beta$, which represents the power of the likelihood ratio test statistic.

To better understand how $c_r \cong \frac{1-\beta}{\alpha}$, we can argue that through stages i ($i = 1, 2, 3, \dots$) of sampling the Pr of rejecting H_0 (at stage t) if H_0 is false is given by

$$1 - \beta = \sum_{t=1}^{\infty} \int_{R_1} \dots \int_{R_t} \prod_{i=1}^t f(x_i; \mu_1, \sigma^2) dx_1 \dots dx_t$$

where R_i denotes that region of x_i which corresponds to the rejection of H_0 at the i^{th} stage, $i = 1, 2, \dots, t$. However, due to Neyman-Pearson Lemma we reject H_0

only if $\lambda_t = \frac{\prod_{i=1}^t f(x_i; \mu_1, \sigma^2)}{\prod_{i=1}^t f(x_i; \mu_0, \sigma^2)} \geq c_r$, or reject H_0 if $\prod_{i=1}^t f(x_i; \mu_1, \sigma^2) \geq c_r \prod_{i=1}^t f(x_i; \mu_0, \sigma^2)$

and thus

$$1 - \beta \geq \sum_{t=1}^{\infty} \int_{R_1} \dots \int_{R_t} \prod_{i=1}^t c_r f(x_i; \mu_0, \sigma^2) dx_1 \dots dx_t = c_r \sum_{t=1}^{\infty} \int_{R_1} \dots \int_{R_t} \prod_{i=1}^t f(x_i; \mu_0, \sigma^2) dx_1 \dots dx_t =$$

$c_r \times \Pr(\text{rejecting } H_0 | H_0 \text{ is true}) = c_r \times \alpha \rightarrow c_r \leq (1 - \beta)/\alpha$. Similar arguments can be made to illustrate that $c_a \geq \beta/(1 - \alpha)$. (This a 5-point bonus Problem)

THE TRADITIONAL ONE-SIDED CUSUM PROCEDURE

Suppose we wish to control a process at a desired process mean of μ_0 (i.e., we wish to test the null hypothesis $H_0: \mu = \mu_0$) but let the alarm sound ASA (as soon as) the process mean μ exceeds a rejectable QLEV (quality level) $\mu_2 \equiv \mu_r > \mu_0$. That is, we wish to test $H_0: \mu = \mu_0$ vs the one-sided alternative $H_1: \mu = \mu_2 > \mu_0$. It will turn out that the LOS of Cusum charts will not necessarily equal to 0.0027. Sometimes, the ideal mean μ_0 is also referred to as the acceptable quality level (AQL) and is generally denoted by μ_a . Further, it is clear to the reader that the QCH, X , in this case must be of an STB type because we are concerned only about an upward shift in the process mean.

The (upper one-sided) Cusum procedure consists of taking random samples of size $n \geq 1$, generally every hour in time-sequence, and compare the Cusum statistic S_t against the UCL_t . Further, a random sample of size $n = 1$ is very common. ASA S_t exceeds the decision point UCL_t , H_0 is rejected at time t and it is concluded that the process mean has shifted upward to $\mu_r = \mu_0 + \delta_2$, where $\delta_2 = \mu_2 - \mu_a$. To arrive at the Cusum procedure, we assume that $X \sim N(\mu, \text{known } \sigma^2)$ and apply this assumption to equation (11) and the SPR testing procedure, which requires we must continue sampling as long as

$$\frac{\beta}{1-\alpha} \leq c_a < \frac{\prod_{i=1}^t f(x_i; \mu_r, \sigma^2)}{\prod_{i=1}^t f(x_i; \mu_0, \sigma^2)} < c_r \leq \frac{1-\beta}{\alpha}$$

or approximately

$$\frac{\beta}{1-\alpha} < \frac{\prod_{i=1}^t \frac{e^{-(x_i - \mu_r)^2 / 2\sigma^2}}{\sigma\sqrt{2\pi}}}{\prod_{i=1}^t \frac{e^{-(x_i - \mu_0)^2 / 2\sigma^2}}{\sigma\sqrt{2\pi}}} < \frac{1-\beta}{\alpha}$$

$$\frac{\beta}{1-\alpha} < \frac{\prod_{i=1}^t e^{-(x_i - \mu_r)^2 / 2\sigma^2}}{\prod_{i=1}^t e^{-(x_i - \mu_0)^2 / 2\sigma^2}} < \frac{1-\beta}{\alpha}$$

$$\frac{\beta}{1-\alpha} < \frac{e^{-\sum_{i=1}^t (x_i - \mu_r)^2 / 2\sigma^2}}{e^{-\sum_{i=1}^t (x_i - \mu_0)^2 / 2\sigma^2}} < \frac{1-\beta}{\alpha}$$

$$\frac{\beta}{1-\alpha} < e^{-\sum_{i=1}^t (x_i - \mu_r)^2 / 2\sigma^2 + \sum_{i=1}^t (x_i - \mu_0)^2 / 2\sigma^2} < \frac{1-\beta}{\alpha}$$

$$\frac{\beta}{1-\alpha} < e^{(\mu_r - \mu_0) \sum_{i=1}^t x_i / \sigma^2 - t(\mu_r^2 - \mu_0^2) / 2\sigma^2} < \frac{1-\beta}{\alpha}$$

$$\ln\left(\frac{\beta}{1-\alpha}\right) < (\mu_r - \mu_0) \sum_{i=1}^t x_i / \sigma^2 - t(\mu_r^2 - \mu_0^2) / 2\sigma^2 < \ln\left(\frac{1-\beta}{\alpha}\right)$$

Dividing this last two-sided inequality first by $\delta_2 = \mu_2 - \mu_0 = \mu_r - \mu_0$ and then multiplying through by σ_x^2 results in

$$\frac{1}{\delta_2} \ln\left(\frac{\beta}{1-\alpha}\right) < \sum_{i=1}^t x_i / \sigma^2 - t(\mu_2 + \mu_0) / 2\sigma^2 < \ln\left(\frac{1-\beta}{\alpha}\right) / \delta_2$$

$$\frac{\sigma_x^2}{\delta_2} \ln\left(\frac{\beta}{1-\alpha}\right) < \sum_{i=1}^t x_i - t(\mu_2 + \mu_0) / 2 < \frac{\sigma_x^2}{\delta_2} \ln\left(\frac{1-\beta}{\alpha}\right)$$

$$\frac{\sigma_x^2}{\delta_2} \ln\left(\frac{\beta}{1-\alpha}\right) < \sum_{i=1}^t x_i - \sum_{i=1}^t \mu_0 - t(\mu_2 - \mu_0) / 2 < \frac{\sigma_x^2}{\delta_2} \ln\left(\frac{1-\beta}{\alpha}\right)$$

$$\frac{\sigma_x^2}{\delta_2} \ln\left(\frac{\beta}{1-\alpha}\right) < \sum_{i=1}^t (x_i - \mu_0) - t\delta_2 / 2 < \frac{\sigma_x^2}{\delta_2} \ln\left(\frac{1-\beta}{\alpha}\right)$$

$$\frac{\sigma_x^2}{\delta_2} \ln\left(\frac{\beta}{1-\alpha}\right) + \frac{t\delta_2}{2} < \sum_{i=1}^t (x_i - \mu_0) < \frac{\sigma_x^2}{\delta_2} \ln\left(\frac{1-\beta}{\alpha}\right) + \frac{t\delta_2}{2} \quad (12)$$

The Cusum statistic, for $n = 1$, from Eq. (12), is defined as the sum $S_t =$

$$\sum_{i=1}^t (x_i - \mu_0), \text{ where the process is concluded to be out of control at time } t \text{ iff } S_t = \sum_{i=1}^t (x_i - \mu_0) > \frac{\sigma_x^2}{\delta_2} \ln\left(\frac{1-\beta}{\alpha}\right) + \frac{t\delta_2}{2} \text{ with the amount of upward shift in the process}$$

mean equaling to δ_2 , and is assumed to be in control at μ_0 iff $S_t = \sum_{i=1}^t (x_i - \mu_0) <$

$$\frac{\sigma_x^2}{\delta_2} \ln\left(\frac{\beta}{1-\alpha}\right) + \frac{t\delta_2}{2}. \text{ Note that some authors use } S(t), \text{ or } S_t, \text{ for the cumulative}$$

sum statistic. Eq. (12) clearly shows that the UCL for the one-sided hypothesis $H_0 : \mu = \mu_0$ versus the simple alternative $H_1 : \mu = \mu_2 > \mu_0$, for sample sizes $n = 1$ at stage t of sampling, is given by

$$UCL_t = \frac{\sigma_x^2}{\delta_2} \ln\left(\frac{1-\beta}{\alpha}\right) + (\delta_2/2)t. \quad (13a)$$

The reader must bear in mind that the above SPRT conservatively maintains the same (α, β) levels at $\mu = \mu_0$ and $\mu = \mu_2$, respectively, throughout the control procedure.

If the QC engineer is using the statistic $S_t = \sum_{i=1}^t (\bar{x}_i - \mu_0)$, where equal-

sample-size $n > 1$ at each time-sequence sampling stage, then the UCL in equation (13a) must be modified to

$$UCL_t = \frac{\sigma_{\bar{x}}^2}{\delta_2} \ln\left(\frac{1-\beta}{\alpha}\right) + (\delta_2/2)t = \frac{\sigma_x^2}{n\delta_2} \ln\left(\frac{1-\beta}{\alpha}\right) + (\delta_2/2)t. \quad (13b)$$

Before going into an example, I must caution the reader to the fact that I am going a bit against the QC literature by using μ_1 and μ_2 as the values of the

downward and upward shifts in the process mean, respectively. In other words, in my notation in contradiction to QC literature $\mu_2 > \mu_0$ and $\mu_1 < \mu_0$ so that $\delta_2 = \mu_2 - \mu_0 > 0$, while $\delta_1 = \mu_1 - \mu_0 < 0$. I believe it is more natural to have μ_1 to the left of μ_0 and μ_2 to the right of μ_0 . In the case of a downward shift in the mean from μ_0 to μ_1 , equations (12 and 13) stay fully in tact, but one must replace δ_2 by $\delta_1 < 0$ and change UCL_t to LCL_t and must observe that the slope of the control line will be negative for a downward shift. Note that δ_2 is simply the maximum amount of tolerable shift in the mean to right of μ_0 , while $|\delta_1|$ is the maximum tolerable shift to the left of μ_0

Example 10. Suppose the ideal target mean for a process is $\mu_a = \mu_0 = 12$ and the rejectable quality level (RQL) is $\mu_r \equiv \mu_2 \equiv 13$ with known process variance $\sigma_x^2 = 1.80$ so that $\delta_2 = 1.0$. Further, it is desired that the traditional Cusum chart provides a type I error probability of $\alpha = 0.01$ at $\mu_0 = 12.0$ and a rejection Pr of $1 - \beta = 0.90$ at the upward shift to $\mu_2 \equiv 13.0$. Samples of size $n = 1$ are taken every hour in time-sequence with the results $X_i = 12.7, 12.3, 14.8, 11.2, 10.3, 11.0, 12.2, 10.9, 12.2, 12.7, 10.5, 11.7, 11.0, 10.8, 11.7, 10.9, 11.1, 13.8, 13.0, 11.4, 10.0, 11.2, 13.2, 10.9, 11.0, 11.7, 12.3, 11.2, 12.2, 12.0, 15.0, 14.1, 13.9, 13.5, 15.5$.

Equation (12) gives the UCL at time t as follows: $UCL_t = \frac{1.8}{1} \ln\left(\frac{0.90}{0.01}\right) + (1/2)t = 8.10 + 0.50t$, where $\delta_2 = 13 - 12 = 1.0$.

The Excel spread sheets on my website show the Cusum control chart and the corresponding Shewhart R_m –and X-charts for 35 samples each of size 1. The moving range, R_m , chart shows that the process is in a state of statistical control wrt variation (within subgroups) but the Shewhart X-chart shows that there is a sudden shift in the mean at the time $t = 35$. However, the Shewhart X-chart exhibits no trend but the Cusum-chart does show a slight downward trend in the mean starting at the 4th hour thru the 28th sample followed by a definite upward trend in the mean starting with sample 31. Further, the Traditional Cusum chart shows that the process is overall in a state of statistical control.

Exercise 20. The data on my website gives the welding joint strength in

ksi of $m = 25$ subgroups each of size $n = 4$. (a) Set up R- & \bar{X} -Charts for monitoring the process variation and process mean in order to detect sudden shifts. (b) Given that the acceptable mean quality is $\mu_0 = 28.5$ ksi, $\sigma_x = 3.40$ ksi, and rejectable mean quality level is $\mu_r = \mu_1 = 25$ ksi, set up a Traditional Cusum procedure that provides an α -level of 0.02 and a protection of $1 - \beta = 0.95$ when there is a downward shift to $\mu = 25$ ksi.

THE TWO-SIDED PROCEDURE FOR TRADITIONAL CUSUM CHARTS

A 2-sided Cusum chart monitors the mean of a process both for downward and upward gradual shifts and trends in the process mean. As before, the Cusum statistic is given by the sum $S_t = \sum_{i=1}^t (\bar{x}_i - \mu_0)$, but there are two control limits for the traditional Cusum statistic, S_t , that are given below.

$$LCL_t = \frac{\sigma_x^2}{n\delta_1} \ln\left(\frac{1-\beta}{\alpha}\right) + (\delta_1/2)t \quad (14a)$$

$$UCL_t = \frac{\sigma_x^2}{n\delta_2} \ln\left(\frac{1-\beta}{\alpha}\right) + (\delta_2/2)t, \quad (14b)$$

where $\delta_1 = \mu_1 - \mu_0 < 0$, and $\delta_2 = \mu_2 - \mu_0 > 0$. The most common of 2-sided procedure is the symmetrical case where $\delta_1 = -\delta_2$.

Example 11. To illustrate the 2-sided Traditional Cusum procedure, suppose the specifications for an incline angle, X , in making metal chips (used in the assembly of car instrument panels) is 15.00 ∓ 2.40 degrees, and it is known that the PCR = 1.00 (i.e., 6-Sigma capable, but process only at Motorola's 3-Sigma quality). Hourly random samples of size $n = 5$ has yielded sample means $\bar{x}_i = 15.29$ degrees, 16.03, 14.80, 14.72, 14.51, 13.91, 14.85, 14.61, 14.10, 14.01, 13.89, 14.02, 14.40, 13.90, and 14.08 degrees (i.e., $m = 15$ subgroups). Our objective is to set up a 2-sided traditional Cusum procedure with an AQL

(acceptable quality level) $\mu_0 = 15.00$ and RQLs (rejectable quality levels) of $\mu_1 = 14.60$ and $\mu_2 = 15.40$ that guarantees an ARL at μ_0 equaling to $ARL_0 = 100 = 1/\alpha$, and protections $1 - \beta_1 = 1 - \beta_2 = 1 - \beta = 0.90$ when H_0 is false. Then $\alpha/2 = 0.005$, $\delta_1 = 14.6 - 15 = -0.40$, $\delta_2 = 15.40 - 15 = 0.40$, and equations (14) now yield:

$$LCL_t = \frac{0.8^2}{5(-0.40)} \ln\left(\frac{0.90}{0.005}\right) - 0.20t = -1.662 - 0.20t$$

and similarly, $UCL_t = 1.66175 + 0.20t$. On the spreadsheet on my website I have tabulated the values of t , \bar{x}_i , S_t , LCL_t and UCL_t under Example11. My spreadsheet clearly confirms that the 11th sample shows a downward shift in the process mean in the amount equaling $\delta_1 = -0.40$, i.e., the process is out of control and the experimenter has to assume that the new process mean after sample 11 is $\mu_1 \cong 14.60$. Since the ideal target is $\mu_0 = 15$ degrees, corrective action has to be taken to reset the mean angle from 14.60 up to $\mu_0 = 15.00$ degrees.

The reader must bear in mind that the ARL at μ_0 for a two-sided Shewhart procedure is always $1/2$ of the corresponding one-sided procedure because the false alarm for a 2-sided procedure may occur either below the LCL or above the UCL. This implies that for a 2-sided Shewhart 3-Sigma control chart the $ARL_0 = 1/0.0027 = 370.40$, while for the corresponding 1-sided procedure the $ARL_0 = 1/0.00135 = 740.741$.

MODERN CUSUM CHARTS for ONE-SIDED SHIFTS in PROCESS MEAN

Suppose the measurement, X , is $N(\mu, \text{known } \sigma^2)$ and we wish to maintain an AQL = μ_0 but let the right alarm sound, if there is an upward shift to the RQL $\equiv \mu_2 > \mu_0$. Define the reference value $k_2 = (\mu_2 + \mu_0)/2$, i.e., exactly half-way between the AQL and RQL, and assume that we are sampling the process in time-ordered sequence on an hourly (or other interval) basis with random samples of size $n \geq 1$ resulting in \bar{x}_i ($i = 1, 2, 3, \dots, t, \dots$). For an upward shift in μ , define the Cusum statistics $C_1^+ = \bar{x}_1 - k_2 + C_0^+$, $C_2^+ = \text{Max}[0, (\bar{x}_2 - k_2) + C_1^+]$, $C_3^+ = \text{Max}[0, (\bar{x}_3 - k_2)$

$+C_2^+]$, ..., $C_t^+ = \text{Max}[0, (\bar{x}_t - k_2) + C_{t-1}^+]$. Note that this definition of Cusum statistic assures that C_t^+ is set to zero ASA C_t^+ become negative, and as a result the Cusum procedure restarts afresh. The quantity C_0^+ , referred to as Head-Start (HS), is commonly set to zero unless a fast initial response for an upward shift is needed, in which case it is generally set to either $HS = (k_2 - \mu_0)/2$, or $A/2$ (defined below). For a one-sided procedure in order to detect a downward shift to μ_1 , the Cusum statistic is defined as $C_t^- = \text{Min}[0, (\bar{x}_t - k_1) + C_{t-1}^-]$, and ASA $C_t^- > 0$ restart the Cusum procedure afresh. Note the subscript $^+$ denotes upward and $^-$ denotes downward shift in the process mean (or shift to a lower mean). Once $C_t^+ > h^+ > 0$ (= the decision limit) to be defined below, then it is ascertained that the process mean may have shifted upward to μ_2 , assignable causes must be searched for, and if none is found, then the Cusum-chart has committed a type I error. For detecting a downward shift to μ_1 , the value of C_t^- is required to be less than $h^- < 0$. Note that the value of h obtained from both K. W. Kemp's Nomograms posted on my website is always positive in all cases.

DESIGN OF a MODERN ONE-SIDED CUSUM PROCEDURE

The 1st concept in designing a Cusum one-sided procedure is the fact that modern Cusum charts rarely allow a LOS as large as 0.05; the most common LOS values are $\alpha = 0.0066667, 0.005, 0.004, 0.002, 0.0016667, 0.001349$, and 0.001 at which the false alarm (or in-control) average run lengths are $ARL_0 = 150, 200, 250, 500, 600, 741$, and 1000 , respectively. The most common desired values of average run length after an upward shift to μ_2 (or downward shift to μ_1) has occurred are $ARL_1 = L_r = 2.5, 3, 4, 5, 6, 7, 7.5, 8, 9$ and 10 . I have used the Nomogram (or Nomograph) from a paper by K. W. Kemp in Applied Statistics, 11 (1962) on page 23 of the Journal, to read the values of $A = h/\sigma_{\bar{x}} = h\sqrt{n}/\sigma$ and

$B(\mu) = \frac{|k - \mu|}{\sigma_{\bar{x}}} = \frac{|\mu - k|\sqrt{n}}{\sigma_x}$, and I am providing the Nomogram (that I copied from

page 517 of the text by A. J. Duncan, *QC & Industrial Statistics*, Fifth Edition, Irwin Publishing, 1986, ISBN:0-256-03535-0) on my website for your convenience. Please observe that $A = h/\sigma_{\bar{x}}$ gives the limit of standardized decision (or acceptance) interval, where h is referred to by A. J. Duncan, p. 516, the decision interval having the same unit as X . It seems, however, to me that Duncan meant h to be the decision limit and not an interval. For example, the Nomogram on my website shows that if $L_a = ARL_0 = 150$ (the right scale), and the value of $L_r =$ the rejectable average run length $= ARL_1 = 3$, then using a ruler to connect the two points $ARL_0 = 150$ and $L_r = 3$, we read the approximate value of $A \cong 2.05$ and $B \cong 0.87$, but if $ARL_0 = 150$ and the out-of-control $ARL_1 = 6$, then $A \cong 2.96$ and the left scale $B \cong 0.565$. However, if $L_0 = ARL_0 = 500$, $L_r = ARL_1 = 5$, then $A \cong 3.17$ and $B \cong 0.743$; further, if $L_0 = ARL_0 = 500$ and $ARL_1 = L_r = 10$, then $A \cong 4.51$ and $B \cong 0.482$; if $L_0 = ARL_0 = 370$ and $ARL_1 = L_r = 8$, then $A \cong 4.00$ and $B \cong 0.525$, etc.

Example 12. We now apply the modern Cusum procedure to the data of Example 10, where we change n from 1 to $n = 4$ for illustrative purposes, and recall that $\mu_0 = 12$, $\mu_2 = 13$, $\delta_2 = \mu_2 - \mu_0 = 1$, the reference value $k_2 = 12.5$ (or Duncan's $k = 12.5$) and $\sigma^2 = 1.80$ ($\sigma_X = 1.3416$), but we assume that the 35 values given on my website are the averages of random samples of size $n = 4$, so that $\sigma_{\bar{x}} = 0.6708$. Our objective is to design a tabular Cusum procedure that allows only an $ARL_0 = 500$ (i.e., $\alpha = 1/500 = 0.002$). Since we know the sample size $n = 4$, then the value of the

left scale $B_0 = \frac{|k - \mu_0|}{\sigma_{\bar{x}}} = \frac{|12.5 - 12|\sqrt{4}}{\sqrt{1.8}} = 0.7454$ together with $L_0 = ARL_0 = 500$ show

on the Nomogram that $A \cong 3.15 = h^+ \sqrt{n} / \sigma = h^+ \sqrt{4/1.8} = 1.4907h \rightarrow h = 2.1131$.

Therefore, the Tabular Cusum procedure, which is on my website in an Excel file, tells us that if the Cusum statistic $C_t^+ = \text{Max}[0, (\bar{x}_t - k_2) + C_{t-1}^+] > h^+ = + 2.1131$, then we must stop production and look for assignable causes, but if at any stage $C_t^+ \leq 0$ then you must let $C_t^+ = 0$ and restart the Cusum procedure afresh. As long as C_t^+ lies in the indecision, or continue-sampling, interval $(0, 2.1131)$, sampling will continue and we will have to assume that the process lacks a significant positive

trend upward. The Excel file clearly shows that for our example 12 the process is out of control at the sample number 31 and for all samples after sample number 31 because $C_{31}^+ = 2.50 > h^+ = 2.1131$, and hence the production must be stopped and assignable causes for samples 31-35 must be investigated. If found, then corrective action must be taken to adjust the mean downward toward the ideal target of $\mu_0 = 12.00$. Furthermore, both Nomograms show that the $ARL_1 \cong 5.0$, which implies that this is a powerful procedure because on the average it takes only 5 random samples each of size $n = 4$ to detect a shift from $\mu_0 = 12$ to $\mu_2 > 12.5$. Note that the Kemp Nomogram that I copied from Jerry Banks' book usually gives a bit larger ARLs. Kenett and Zacks (1998) provide an approximate formula on their page 377 for the

decision interval $h^+ = -\frac{\sigma^2 \ln(\alpha)}{n(\mu_2 - \mu_0)} = -\frac{\sigma_{\bar{x}}^2 \ln(\alpha)}{\delta_2} = \frac{\sigma_{\bar{x}}^2 \ln(\alpha^{-1})}{\delta_2}$. For the example being

discussed, this last formula gives $h^+ = -\frac{1.8 \times \ln(0.002)}{4(13 - 12)} = 2.7966$, which is always

more conservative or less powerful) than the h from Duncan's Nomogram (and perhaps not as useful). Note that as before, h^+ is directly proportional to σ^2 and inversely proportional to α , n , and $\delta_2 = \mu_2 - \mu_0$.

The reader should be able, by now, to fully understand that Modern Cusum charts are more powerful in detecting future gradual shifts than Traditional charts, because Traditional charts keep a history of process mean from the very beginning of chart construction, while modern Cusum charts set the Cusum statistic to zero if the mean quality is better than or near the ideal target. This procedure of resetting to zero will allow the experimenter to detect gradual shifts more rapidly when the process mean starts trending in the wrong direction, while ignoring the long past history. As a result of this resetting to zero, the next few future sample means right after the shift will sound the correct alarm more rapidly.

D. C. Montgomery (*Intro to Stat. QC, 4th Ed.*) provides a formula (8.5) on his page 414 (p. 394 of his 5th edition) for approximating the value of the shifted mean as $\hat{\mu} = \mu_0 + K + C_t^+ / N^+$, where his $K = |\mu_2 - \mu_0| / 2 = \delta_2 / 2$, and thus $\hat{\mu} = \mu_2$

+ C_t^+ / N^+ , $C_t^+ > 0$ and N^+ stands for the number successive times (or RL since the last zero value) that the Cusum statistic C_t^+ is positive. In my Example 12 on my website, I have provided a column of $RL = N^+$ values so that you will know what it stands for, which is equal to 1 at $t = 3$ and is equal to 5 at $t = 35$ for that example. I do not know whether $\hat{\mu} = k_2 + C_t^+ / N^+$ is Montgomery's own formula (although the formula is intuitively obvious and meaningful), because he does not provide a reference for his formula (8.5). Using Montgomery's formula (8-5), we obtain $\hat{\mu} = \mu_0 + K + C_t^+ / N^+ = 12.00 + \frac{|13-12|}{2} + 2.5 = 15$ at $t = 31$, and $\hat{\mu} = 12.5 + 9.5/5 = 14.4$ at $t = 35$. I would guess that Montgomery's formula (8.5) usually provides an estimate of the shifted mean close to the range (k_2, μ_2 or beyond). The larger sudden shifts give rise to larger $\hat{\mu}$. Further, when using C_t^- to detect a downward shift in μ from μ_0 , the value of $K = |\mu_1 - \mu_0| / 2 = -\delta_1 / 2$.

D. C. Montgomery also provides Siegmund's approximation (1985), in his formula (8-6) on page 396 of his 5th edition, for the ARL after a shift has occurred, which he gives as

$$ARL(\mu) = \frac{2\Delta b - 1 + \exp(-2\Delta b)}{2\Delta^2}, \quad (\text{Eq. 8-6 of D. C. Montgomery})$$

where $b = A| + 1.166$, $A = h / \sigma_{\bar{x}}$, and $\Delta(\mu) = \frac{\mu - \mu_0}{\sigma_{\bar{x}}} - \frac{K}{\sigma_{\bar{x}}} = \frac{\mu - k_2}{\sigma_{\bar{x}}}$ for an

upward shift, i.e., $\mu - \mu_0 > 0$, and $\Delta(\mu) = \frac{k_1 - \mu}{\sigma_{\bar{x}}}$ for a downward shift. I am fairly

sure that Sigmund's formula is valid only for a one-sided Cusum. To obtain the ARL of a two-sided Cusum (to be discussed below), obtain the ARL^- and ARL^+ from Eq. (8-6) and use the Eq. (8-7) on p. 396 D. C. Montgomery (5e), listed below, to obtain the average run length for the 2-sided ARL denoted by $ARL^{(2)}$.

$$\frac{1}{ARL^{(2)}} = \frac{1}{ARL^-} + \frac{1}{ARL^+}, \quad (\text{Eq. 8-7 on p. 396 of D. C. Montgomery})$$

where for a downward shift $\Delta(\mu) = \frac{k_1 - \mu}{\sigma_{\bar{x}}}$ and $\mu - \mu_0 < 0$. Note that Eq. (8-7) listed on p. 396 of Montgomery (5e) has been known for well over 40 years.

For our one-sided example, at $\mu = \mu_2 = 13$, $\sigma_{\bar{x}} = \sqrt{1.8/4} = 0.6708$, $A = 2.1131/0.6708 = 3.15$, $b = 3.15 + 1.166 = 4.3160$, $K = 1/2 = 0.5$, $\Delta(\mu) = \frac{13 - 12.5}{\sqrt{1.8/4}} = 0.7454$, $2b\Delta = 6.4339$ so that $ARL(\mu = 13) = \frac{6.4339 - 1 + \exp(-6.4339)}{2(0.7454)^2} = 4.8914$,

which is very close to the $L_1 = ALR_1 = 5.0$ estimated from both Nomograms. Note that I had to modify Siegmund's formula (8-6) given by D. C. Montgomery (5e, 2005) by replacing σ by $\sigma_{\bar{x}} = \sigma_x / \sqrt{n}$, because he restricts his discussions on pp. 386-400 to $n = 1$, as I have not seen the original paper by Siegmund, but I am fairly certain that the modification of replacing σ_x by the $SE(\bar{x}) = \sigma_x / \sqrt{n}$ in all the terms is correct. However, in all cases that I have tried the above modification gives very close readings of ARL to those read from both Nomograms. Below I will illustrate how to use both Kemp's Nomograms, already on my website, to guesstimate the $ARL(\mu)$ for different specified values of μ of Example 12. Note that ARL is a function of μ .

For example, if $\mu = 13.2$, then $B(13.2) = \frac{|k - \mu|}{\sigma_{\bar{x}}} = \frac{|12.5 - 13.2|}{0.6708} = 1.0435$;

connecting the Duncan's Nomogram left-scale of 1.0435 to $A = 3.15$ yields $L_1 = L_r = 4.0$. Using Siegmund's (1985) approximation, $ARL(\mu) = [2\Delta b - 1 + \exp(-2\Delta b)] / (2\Delta^2)$, we obtain $\Delta(13.2) = \frac{13.2 - 12.5}{0.6708} = 0.70/0.6708 = 1.0435$, $b = A + 1.166 = 3.15 + 1.166 = 4.3160$, $2\Delta(\mu)b = 9.008$, and $ARL(13.2) = [9.008 - 1 + \exp(-9.008)] / (2 \times 1.0435^2) = 3.68$.

If the QCH is LTB, then the concern will be only about a downward shift in the process mean, and the estimated formula for the mean will become

$\hat{\mu} = \mu_0 - K + C_t^- / N^- = k_1 + C_t^- / N^- < \mu_0 - K$, where $C_t^- < 0$. The value of $\hat{\mu}$ will

generally lie in the interval (μ_1 or below, k_1).

Exercise 21. Use the Nomogram on my website to design a one-sided tabular Cusum procedure which has an $ARL_0 = 500$ at $\mu = 10.00$, but when the process mean shifts downward to the RQL = $\mu_1 = 9.50$ (i.e., $\delta_1 = -0.50$), then $L_1 = ARL_1 = 7$, where $X \sim N(\mu, 0.525625)$. State your exact procedure. (b) Apply your procedure to the data that I have provided on my website under Ex21 and make a statistical decision. (c) Compute the ARL values if μ shifts to 9.75 and 9.4.

DESIGN OF a MODERN TWO-SIDED CUSUM PROCEDURE

The easiest way to illustrate the two-sided procedure is to provide an example, which we will refer to as the Example 13 on my website. Reconsider the content of my example 11 where $\mu_0 = 15$ degrees, $\mu_1 = 14.6$, $\mu_2 = 15.40$, $\delta = 0.40$, $\sigma_x = 2\Delta/6 = 4.8/6 = 0.80$, $\delta_1 = -0.40$ and $\delta_2 = 0.40$ degrees. We discuss only the symmetrical procedure, where the left reference value is $k_1 = (\mu_1 + \mu_0)/2 = 14.8$ degrees and the right reference value is $k_2 = (\mu_2 + \mu_0)/2 = 15.20$. Our first objective is to design a two-sided Cusum procedure such that $ARL_0 = 200$ ($\rightarrow \alpha = 1/200 = 0.005$), and $L_1 = ARL_1 = ARL(\text{at } \mu = 15 \pm 0.40 = 15 \pm \delta) = 9.0$ when there is a shift in the process mean equaling to either RQLs 14.6, or 15.40 degrees. Since the Nomograms on my website was designed for a one-sided procedure, then in order to use them for a 2-sided procedure, the specified 2-sided value of $ARL_0 = 200$ must 1st be multiplied by 2, i.e., the corresponding $ARL_0^{(1)} = 400$. This is due to the fact that for the 2-sided procedure, the corresponding average run length at μ_0 can occur both below and above μ_0 so that $ARL_0^{(1)} = 1/0.0025 = 400$. Now the

Nomogram shows that at $L_0 = 400$ and $L_1 = L_r = 9$, the value of $A = h/\sigma_{\bar{x}} = \frac{h\sqrt{n}}{0.80} \cong$

4.22 and $B_0 = \frac{|k - \mu_0|\sqrt{n}}{\sigma} \cong 0.50$. These two equations with two unknowns when

solved simultaneously yield $n = 4$ and $h = 1.688$. Therefore, the Cusum two-sided procedure consists of accumulating both C_t^- and C_t^+ for each random sample of $n = 4$ simultaneously at each stage of sampling and then deciding that μ has shifted below $k_1 = 14.80$ if $C_t^- < -1.6880$, or deciding that μ has shifted at least beyond $k_2 = 15.2$ if $C_t^+ > 1.6880$. As long as the two Cusum statistics fall in the indecision interval $(-1.6880, 1.6880)$, then sampling must continue. On my website I have provided $m = 20$ subgroup means each of size $n = 4$ and have computed Cusum statistics for illustrative purposes. Note that through 20 subgroups no definitive decision can be made about a shift in the process mean, μ , in either direction because both Cusum statistics fall within the continue-sampling interval $(-1.688, 1.688)$, and thus sampling may continue to stage 21.

If there is a shift from $\mu_0 = 15$ to $\mu = 15.20 = k_2$, the reader must first bear in mind that the value of $ARL(\text{at } \mu = 15.20)$ must lie within the interval $(L_1 = 9, L_0 =$

200). Now, by definition $A = h / \sigma_{\bar{x}} = \frac{1.688}{0.80/\sqrt{4}} = \frac{1.688}{0.40} = 4.2200$, $b = A + 1.166 =$

5.3860 , $K = 0.40/2 = 0.20$, and $\Delta = \frac{15.2 - k_2}{0.4} = 0$. However, Siegmund's formula

$ARL(\Delta) = \frac{2\Delta b - 1 + \exp(-2\Delta b)}{2\Delta^2}$ yields $ARL(\text{at } \Delta = 0) = \frac{0 - 1 + e^{-0}}{2(0)^2} = \frac{0}{0}$, which is

indeterminate. Using L'Hospital's rule and taking the derivatives of numerator

and denominator of $ARL(\Delta) = \frac{2\Delta b - 1 + \exp(-2\Delta b)}{2\Delta^2}$ twice wrt (with respect to) Δ

gives $ARL(\Delta) = \frac{0 - 0 + 4b^2 \exp(-2\Delta b)}{4}$ showing that $ARL(\text{at } \Delta = 0) = b^2 =$

$(5.3860)^2 = 29.0090$, which indeed lies within the interval $(L_1 = 9, L_0 = 200)$.

Estimating the $ARL(\text{at } \mu = 15.20)$ from Nomograms is difficult at best because the left-most scale $B(15.2) = |\mu - 15.20| / \sigma_{\bar{x}} = 0$.

Yet as another example, we estimate the ARL at $\mu = 15.50$, which will have to be less than $L_1 = 9$ because this is a larger shift in the process mean than $\mu =$

15.40. Proceeding as above, $\Delta = (15.5 - 15.2) / \sigma_{\bar{x}} = 0.4 = 0.75$, $ARL_r(\mu = 15.50) =$

$$L_1(15.5) = \frac{1.5(5.3860) - 1 + \exp(-8.0790)}{2(0.75)^2} = 6.30. \text{ In order to obtain this ARL from}$$

the Nomograms, we compute $B(15.5) = |\mu - k_2| / \sigma_{\bar{x}} = 0.30/0.4 = 0.75$ and connect $B = 0.75$ to $A = 4.22$ and the Nomogram from A. J. Duncan shows that $ARL_1 = L_r(15.5) \cong 7.1$.

In order to compare the ARL's from the Nomograms and Siegmund's formula, we consider a small shift from $\mu_0 = 15$ to $\mu = 15.05 < k_2$. Then, $B(15.05) =$

$$|15.05 - 15.20| / 0.40 = 0.375, \Delta = \frac{15.05 - 15.2}{\sigma_{\bar{x}} = 0.4} = -0.375, 2\Delta b = -0.75 \times 5.3860 =$$

$$-4.0395 \rightarrow ARL(\text{at } \Delta = -0.375) = \frac{-4.0395 - 1 + \exp(4.0395)}{2(-0.375)^2} = 184.0300. \text{ Since } B =$$

$|\Delta| = 0.375$, the Nomogram from Duncan shows that the one-sided $ARL^+ \cong 180$, but Kemp's Nomogram gives roughly $ARL^{(1)} = 260$. It must be mentioned that when the shift in μ is in the interval (k_1, k_2) , then it is worth while to consider the possibility of ARL in both directions in the Siegmund's approximation by

$$\text{computing } B^-(15.05) = (15.05 - k_1)/0.4 = (15.05 - 14.8)/0.4 = 0.625, \Delta^- = \frac{k_1 - \mu}{\sigma_{\bar{x}}} =$$

$$-0.625, 2\Delta^- b = -1.25 \times 5.3860 = -6.7325, \text{ and } ARL^- = \frac{-6.7325 - 1 + \exp(6.7325)}{2(-0.625)^2} =$$

1064.3331 (compared to Duncan's $L_a \cong 950$). Now, the corresponding 2-sided ARL

$$\text{can be computed from the following formula: } \frac{1}{L^{(2)}} = \frac{1}{L^-} + \frac{1}{L^+} \rightarrow L^{(2)} = \frac{L^- \times L^+}{L^- + L^+}$$

$$= \frac{184.03 \times 1064.3331}{184.03 + 1064.3331} = 156.901. \text{ Note that the } L_a \text{ readings from the Nomograms}$$

cannot always be adjusted for the ARL of a 2-sided procedure (except for $\mu = \mu_0$) unless both B^- and B^+ yield L_a values within the interval $100 \leq L_a \leq 1000$, and the L_r readings do not need the above adjustment. Further, for μ in the interval (k_1, k_2) the closer μ gets to the two reference points k_1 and k_2 the smaller the above

adjustment to L_a becomes. For example, at $\mu = 15.10 < k_2$, $B(15.10) = |15.10 - 15.20| / SE(\bar{x}) = 0.25$ and $A = 4.22$ yield $L_a < 100$ from Kemp's Nomogram.

However, $\Delta^+ = \frac{15.1 - 15.2}{\sigma_{\bar{x}} = 0.4} = -0.25 = -B$, $2b\Delta^+ = -2.693$, $ARL(\Delta = -0.25) =$

$$\frac{-2.6930 - 1 + \exp(2.6930)}{2(-0.25)^2} = 88.6635. \text{ As before, } B^- = (14.80 - 15.10)/0.4 = -0.75 = \Delta^-,$$

$$2\Delta^-b = -1.5 \times 5.3860 = -8.0790, \text{ and } ARL^- = \frac{-8.0790 - 1 + \exp(8.0790)}{2(-0.75)^2} = 2859.49033$$

$\rightarrow L^{(2)}(15.10) = 85.99701$ (an adjustment of roughly $88.6635 - 85.9970 = 2.6665$).

Note that the above adjustment should be done only for $k_1 \leq \mu \leq k_2$.

The reader should not confuse the two different uppercase deltas used in these notes. The regular Δ stands for the specified tolerance on either side of the ideal target m , is independent of μ , and is used only for process capability analysis. For the Ex22 below Δ is specified at 2.4 degrees. The second delta, Δ , stands for the deviation of μ from either reference points k_1 and k_2 in standardized ($\sigma_{\bar{x}}$) units, is a function of μ , and is used during process control in approximating the ARL from Siegmund's (1985) approximation. The value of Δ is a function of μ , i.e., $\Delta(\mu) = (\mu - k_2) / \sigma_{\bar{x}}$, or $\Delta(\mu) = (k_1 - \mu) / \sigma_{\bar{x}}$ for a downward shift below μ_0 . Further, when $B(\mu) = |\mu - k| / \sigma_{\bar{x}}$ is connected to A , one can approximate an ARL from the Nomograms.

Exercise 22. Use the Nomogram on my website to design a two-sided tabular Cusum procedure which has an $ARL_0 = 250$ at $\mu = 15.00$, but when the process mean shifts by $\delta = 0.40$ in either direction, then $ARL_1 = 6$, where design specification on X are $15 \mp 2.4 = 15 \mp \Delta$ degrees but the process is 8-sigma capable (i.e., Motorola's 4-sigma quality). State your exact procedure. (b) Apply your procedure to the data that I have provided on my website under Ex22 and make a statistical decision. (c) Compute L_1 at $\mu = 15.3$.