The CUSUM & EWMA Control Charts

Dr Austin R Brown

Kennesaw State University

Introduction

- To this point, we have discussing the use of Shewhart $\bar{X}$ and R/s control charts for monitoring a univariate process' location (mean) and dispersion (variation).
- Examining the $\bar{X}$ chart specifically, this chart is useful in both its ease of implementation and interpretability as well as its ability to detect large shifts in the process mean $(> 2\sigma)$.
- However, the $\bar{X}$ chart is not sensitive to small shifts in the process mean, which can be problematic in many applications.

Introduction

- So if we are interested in efficiently (as measured by ARL) detecting smaller shifts in the process mean ($\leq 2\sigma$), we need to consider alternative control charts.

- Additionally, we have also seen that the $\bar{X}$ chart is not effective when $n = 1$.

- Moreover, since we assume the individual time points are independent, we are using the information of the whole sequence of points in an inefficient manner.

Introduction

- To get around these problems, we can use control charts that incorporate information from the entire sequence of points, that are effective when $n = 1$, and that are sensitive to small shifts in the process mean.

- In this lecture, we will discuss two such control charts: the Cumulative Summation (CUSUM) chart and the Exponentially Weighted Moving Average (EWMA) chart.

The CUSUM Chart

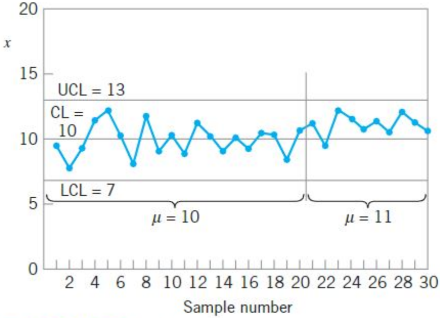

FIGURE 9.1 A Shewhart control chart for the data in Table 9.1.

The CUSUM Chart

- As we can see in this prior example, the first 20 observations come from a normal distribution with mean 10 and standard deviation 1. The following 10 come from a normal distribution with mean 11 and standard deviation 1.

- If we only rely upon the control limits, we would not stop the process to search for an assignable cause despite a shift having taken place.

- So how does the CUSUM chart solve this problem?

The CUSUM Chart

- The CUSUM chart was proposed by Page (1954). Here's the concept:
- Let's say we want to monitor the mean of a process and this process has a target value, say μ_0. To monitor the process, we still take random samples and obtain an estimate of the sample mean, say $\bar{x}_i$.
- Instead of plotting the individual value of $\bar{x}_i$ itself, we take the deviance between the target value and the sample mean, $\bar{x}_i - \mu_0$.

The CUSUM Chart

- Then, the cumulative summation component is that the plotting statistic is a function of the cumulative sum of the deviances.

$$C_i = \sum_{j=1}^{i} (\bar{x}_i - \mu_0)$$

- Here, the idea is that if the process starts to drift upwards or downwards, those cumulative differences will eventually signal to us that a shift has occurred. Notice that we are taking all $\bar{x}_i$'s into consideration instead of just the latest one!

The CUSUM Chart

- What you'll likely notice from the previous C_i is that its distribution is *non-stationary*. If you've taken time series, you'll know that this means that either the mean or the variance change with time.

- If we assume $\bar{x}_i$ is normally distributed with mean μ_0 and variance σ^2/n, then:

$$C_i = \sum_{j=1}^{i} (\bar{x}_i - \mu_0) \dot{\sim} N\left(0, \frac{i\sigma^2}{n}\right)$$

The CUSUM Chart

- This may serve as a limitation if we were to construct control limits in a Shewhart-style manner as the limits will change over time and also won't approach an asymptote.
- Consequently, there have been two proposed variants to the CUSUM chart to help solve this problem, but we will only focus on the one that's mostly used in practice: the **Tabular** CUSUM.

The CUSUM Chart

- The plotting statistic for the Tabular CUSUM is actually two statistics, an upper and a lower CUSUM:

$$C^{+} = \max[0, \bar{x}_i - (\mu_0 + K) + C^{+}_{i-1}]$$
$$C^{-} = \max[0, (\mu_0 - K) - \bar{x}_i + C^{-}_{i-1}]$$

where, $C_0^{+} = C_0^{-} = 0$ and K is referred to as the "slack" or "allowance" value.

The CUSUM Chart

- Typically, K is chosen to be the midpoint between our target value, μ_0 and an out-of-control value, μ_1, that we're interested in quickly detecting:

$$K = \frac{|\mu_1 - \mu_0|}{2}$$

- C^+ is for detecting upward shifts ($\mu_1 > mu_0$) and C^- is for detecting downward shifts ($\mu_1 < \mu_0$).

- If either C_i^+ or C_i^- exceed some control limit, say H, this signals to us that the process may be OOC.

The CUSUM Chart

- What about H? How do we decide its value?
- Typically, H is chosen to be a multiple of the standard deviation of the process, σ.
- The rule of thumb is to set $H = h\sigma$, where $h = 4$ or 5.

The CUSUM Chart

- Let's see how we can implement the CUSUM chart in R.
- In this example, suppose we are filling soda bottles. The target fill is $\mu_0 = 16$ oz
- Suppose we take a sample of 5 bottles every hour for two days (48 hours).
  - Let's define $K = 0.5$ and $H = 5$.

CUSUM Chart Design

- In the last example, we used a CUSUM control chart where $H = 5$ and $K = 0.50$. How do we choose those values?

- Generally, the rule of thumb is to define $H = h\sigma$ and $K = k\sigma$, where $h = 4$ or 5 and $k = 0.50$. So for a given known or estimated value of σ, we can obtain H and K.
  - Note, in our example we standardized our sample means using the estimated standard deviation so we were able to just use $H = 5$ and $K = 0.50$.

CUSUM Chart Design

TABLE 9.3
ARL Performance of the Tabular CUSUM with $k = \frac{1}{2}$ and $h = 4$ or $h = 5$

Shift in Mean (multiple of σ)	$h = 4$	$h = 5$
0	168	465
0.25	74.2	139
0.50	26.6	38.0
0.75	13.3	17.0
1.00	8.38	10.4
1.50	4.75	5.75
2.00	3.34	4.01
2.50	2.62	3.11
3.00	2.19	2.57
4.00	1.71	2.01

TABLE 9.4
Values of k and the Corresponding Values of h That Give $ARL_0 = 370$ for the Two-Sided Tabular CUSUM [from Hawkins (1993a)]

k	0.25	0.5	0.75	1.0	1.25	1.5
h	8.01	4.77	3.34	2.52	1.99	1.61

CUSUM Chart Usage

- So far we have seen that CUSUM can be used for monitoring sample means when $n \geq 1$.
- However, CUSUM can be used for other statistics as well, including the sample standard deviation, sample proportions, number of nonconforming items, etc.
- It can also be used for variable sample sizes by standardizing observations, like we did in the soda bottles example.

CUSUM ARL Estimation

- Unlike the Shewhart-style charts, the plotting statistics for the CUSUM chart are not independent of each other. Thus, the typical $ARL_0 = 1/\alpha$ and $ARL_1 = 1/(1-\beta)$ calculations aren't valid.
- Estimating ARL for CUSUM charts can be quite complicated. However, we do have some rough estimates that we can use in Phase I when making decisions about chart construction.
- For a one-sided CUSUM, Siegmund (1985) provided an approximation:

CUSUM ARL Estimation

$$ARL = \frac{\exp(-2\Delta b) + 2\Delta b - 1}{2\Delta^2}$$

- where $\Delta = \delta^* - k$ for the upper CUSUM, $\Delta = -\delta^* - k$ for the lower CUSUM, $b = h + 1.166$ and $\delta^* = (\mu_1 - \mu_0)/\sigma$.
- Then, if we're using a two-sided CUSUM, we can estimate the overall ARL by:

$$\frac{1}{ARL} = \frac{1}{ARL^+} + \frac{1}{ARL^-}$$

CUSUM ARL Estimation

- In general, though, the ARL for a CUSUM with $k = 0.5$ and $h = 5$ is:

Shift in Mean	ARL
0	465
0.25	139
0.50	38
0.75	17
1.00	10.4
2.00	4.01
3.00	2.57

CUSUM Limitations

- As you can probably tell, the CUSUM chart is not without limitations.
- Compared to the Shewhart $\bar{X}$ chart, the CUSUM chart is more complex to implement and interpret.
- It may also be less evident how to set the control limits, H and K.
- So how do we take the advantages of the CUSUM and combine them with the simplicity of the Shewhart $\bar{X}$ chart?

The EWMA Chart

- The Exponentially Weighted Moving Average (EWMA) chart was proposed by Roberts (1959).
- The EWMA chart is similar to the CUSUM chart in that it is designed to detect small shifts in the process mean and also utilizes information from the entire sequence of points.
- However, the EWMA chart is simpler to implement and interpret than the CUSUM chart.

The EWMA Chart

- Unlike the CUSUM, where the plotting statistic is a function of the deviations away from target, the EWMA control chart "weights" the observations to provide a sort of moving average of the observations. (Note: EWMA can be used as a method for predicting time series values)

- This may be advantageous as far as interpretability is concerned as the plotting statistic is a function of the observed values rather than the observed deviances.

The EWMA Chart

- Let's say we take a random sample ($n \geq 1$) and we calculate a sample mean (or the individual value in the case where $n = 1$) and we'll call this value x_i.

- The plotting statistic for the EWMA control chart is defined as:

$$z_i = \lambda x_i + (1 - \lambda) z_{i-1}$$

- where, $0 < \lambda \leq 1$ and $z_0 = \mu_0$. With some rearranging, we can see that z_i is the weighted average of all previously observed x_i's.

The EWMA Chart

$$z_i = \lambda \sum_{j=0}^{i-1} (1-\lambda)^j x_{i-j} + (1-\lambda)^i z_0$$

- Then, to create control limits, we need to know what the mean of z_i is as well as the variance.

The EWMA Chart

- The mean:

$$E[z_i] = E\left[\lambda \sum_{j=0}^{i-1}(1-\lambda)^j x_{i-j} + (1-\lambda)^i z_0\right]$$

$$\implies \lambda \sum_{j=0}^{i-1}(1-\lambda)^j E[x_{i-j}] + (1-\lambda)^i E[z_0]$$

$$\implies \lambda \sum_{j=0}^{i-1}(1-\lambda)^j \mu_0 + (1-\lambda)^i \mu_0 = \mu_0$$

The EWMA Chart

- The variance:

$$\sigma_{z_i}^2 = Var[z_i] = Var\left[\lambda \sum_{j=0}^{i-1} (1-\lambda)^j x_{i-j} + (1-\lambda)^i z_0\right]$$

$$\implies \lambda^2 \sum_{j=0}^{i-1} (1-\lambda)^{2j} Var[x_{i-j}]$$

$$\implies \sigma_{z_i}^2 = \sigma_{x_i}^2 \left(\frac{\lambda}{2-\lambda}\right)\left(1-(1-\lambda)^{2i}\right)$$

The EWMA Chart

- So we can construct our control limits as:

$$UCL = \mu_0 + L\sqrt{\sigma^2_{x_i}\left(\frac{\lambda}{2-\lambda}\right)(1-(1-\lambda)^{2i})}$$

$$LCL = \mu_0 - L\sqrt{\sigma^2_{x_i}\left(\frac{\lambda}{2-\lambda}\right)(1-(1-\lambda)^{2i})}$$

The EWMA Chart

- Notice as $i \to \infty$:

$$UCL = \mu_0 + L\sqrt{\sigma_{x_i}^2\left(\frac{\lambda}{2-\lambda}\right)}$$

$$LCL = \mu_0 - L\sqrt{\sigma_{x_i}^2\left(\frac{\lambda}{2-\lambda}\right)}$$

- These limits are referred to as the "steady state" limits and are usually obtained around $i = 20 - 30$. However, it is strongly recommended to use the exact limits for smaller values of i as this will avoid issues associated with ARL.

The EWMA Control Chart: Decisions for L and λ

- General guidance on choosing a value of λ is that if you want the chart to be more sensitive to smaller shifts, use a smaller value of λ and vice versa.
- We can use $L = 3$ as in a Shewhart-style chart and that works pretty well for $ARL_0 \approx 500$. More exact values were given by Lucas and Succucci (1990):

The EWMA Control Chart: Decisions for L and λ

Table 1: ARL Values for the EWMA Control Chart

Shift in Mean (multiple of σ)	$L = 3.054$ $\lambda = 0.40$	2.998 0.25	2.962 0.20	2.814 0.10	2.615 0.05
0	500	500	500	500	500
0.25	224	170	150	106	84.1
0.50	71.2	48.2	41.8	31.3	28.8
0.75	28.4	20.1	18.2	15.9	16.4
1.00	14.3	11.1	10.5	10.3	11.4
1.50	5.9	5.5	5.5	6.1	7.1
2.00	3.5	3.6	3.7	4.4	5.2
3.00	2.0	2.3	2.4	2.9	3.5

The EWMA Control Chart: Examples

- Okay, let's work through some examples. Let's say we want to monitor the amount of time customers spend on hold at a customer service support center.
- Suppose the target hold time is $\mu_0 = 10$ and $\sigma = 1$.
- Data for the most recent 30 customers placed on hold are contained in this week's Excel spreadsheet.

The EWMA Control Chart: Examples

- Now what about Phase I? As may be apparent, we don't really have distributional assumptions. We mostly are just concerned about independence of observations and having a good estimate of the process mean and standard deviation.

- In the last example, we assumed the process mean and standard deviation were known. What if they're not? Don't we also need to ensure the process is in statistical control before moving on to Phase II analysis? Yes!

- Suppose we're the Scranton branch of Dunder-Mifflin. In order for our branch to be profitable, we need to maximize sales and control our operating costs. One such operating cost is the amount of fuel our warehouse delivery trucks use. In the second sheet of this week's class data, we have the past 24 months fuel costs. Let's estimate the mean and standard deviation and see if the process is in control.

The EWMA Control Chart: Correlated Observations

- One of the main assumptions of the EWMA chart is that the observations are independent of each other (e.g., no covariance/correlation between two observations).
- What happens if this isn't reasonably met? The EWMA chart's ARL performance will suffer. Why?
- If two variables are dependent on each other, then the variance of their sum is:

$$Var[X + Y] = Var[X] + Var[Y] + 2COV[X, Y]$$

The EWMA Control Chart: Correlated Observations

- Depending on the directionality of the relationship, variance could go up or down.

- If we have positive covariance, then this means our variance is greater than what the EWMA chart assumes and our false alarm rate, α will increase. On the other hand, the chart will be incredibly sensitive to small shifts.

- If we have negative covariance, then our variance is less than what the EWMA chart assumes and our false alarm rate will become quite tiny and very insensitive to shifts.

- It is then important to be reasonably sure that our observations are not significantly correlated before using this chart.
  - Interestingly, this topic is the subject of Dr. VanBrackle's (SDSA Professor Emeritus) dissertation and one published manuscript in 1997.