Journal of the KSQC
Vol. 9, No. 2, 1981

累積合에서 出発한 累積平均에 관한 考察
A Study on Cumean – a self Starting Cusum

趙 載 岦*

ABSTRACT

A typical industrial data - monitoring scheme often requires trend detection Trend detection can be accomplished in many ways.

Common statistical methods are the sign test, the run test, and.the trend test.

Graphical methods include various smoothing schemes and the cusum.

The cusum has established itself as an efficient method of detecting changes in the mean level of a process being monitored.

The cusum requires a "target value" with which the raw data are compared. At production start - up it is often difficult to designate the target value.

This paper offers a means of initiating the cusum technique without a target value.

1. 序 論

典型的인 產業体의 데이타 추적계획은 때때로 検出動向을 要求하고 있다. 検出動向은 여러 면에서 活用되고 있으며 一般的으로 統計的 方法은 싸인検定, 런의検定, 傾向検定 등이 있다. 그래프적인 方法은 여러 平準化計劃과 累積合 計劃을 포함하고 있다.

累積合이란 알고있는 어느 工程의 平均水準 안에서 이미 알고있는 変化에 대해 効果的인 方法이다. 따라서 累積合이란 원래의 데이타와 비교되는 "目標値"를 필요로 한다. 生產이 開始되면서 目標値를 선정하는 것은 어려운 일이기 때문에 이 論文은 目標値가 없는 初期의 累積合에 대한 方法을 제공하려 한다.

品質保証에 도움이 되는 그래프적인 累積合의 使用은 잘 알려져 있다. 원래의 데이타와 예측한 目標値 사이의 差나 誤差는 累積되어 플로트 된다. 플로트된 累積合에서 기울기는 工程이 目標를 벗어났다는 것을 가르키며 또한 이것은 統計的으로 有意한 改善措置를 要請하게 된다. 累積合의 検出動向은 視覚的으로 간단한 数学的計算에 의해 용이하게 그 効果를 얻을 수 있다.

傾向에 대한 変動을 알아내기 위해서는 아주 사용하기에 간편한 싸인検定이 있다. 이 싸인検定은 파라메터가 없는 統計的 検定이다. 여기서는 目標가 데이타의 母集団分布의 中位数이고 만일 데이타의 点이 메디안보다 더 큰 점을 +로 기록하고 메디안보다 더 적은 점을 ㅡ로 기록한다. 만일 工程이 원하는 目標値나 平均에 있다면 +, ㅡ의 점은 같게 기록된다. 사실상 +나 ㅡ의 数는 이항분포를 하여 그러한 傾向이 有意한지 어떤지를 결정하기 위한 아주 용이한 検定이다.

累積合과 싸인検定에서 어떤 目標値는 주어져야 한다. 가끔 이 目標値의 부정확한 예측으로 쓸모없

* 慶熙大學校 工科大學 工業經營學科 教授

이 되고 조정된 기간동안 工程의 平均水準을 수립하도록 要求된다. 싸인檢定을 행한다는 것은 어떤 目標値를 要求하지 않는 傾向檢定이나 傾向을 発見하기 위한 "逆整理"로서 이미 알고있는 것을 使用할 수 있다.

이 論文은 傾向変動을 発見하는데 目標値의 要求를 축소할 수 있는 累積合에 修正을 提示하고 있다. 이것은 累積合보다도 더욱 計算이 간편한 것으로 자신이 출발해서 자신이 제로가 되는 것이다. 이것의 적절한 이름은 累積平均이다.

2. 累積平均 (Cumean)

먼저 두 條件下에서 累積合의 計算을 생각해보면 i) 目標値가 正確한 때 ii) 目標値가 不正確한 때로 구분할 수 있다.

i) 目標値가 正確한 때

x_i, $i=1, 2, 3, \cdots\cdots, N$이 데이타점 N의 集合이라 하고 T가 데이타를 얻게된 工程에 대한 目標値라 하자 그러면 累積合은 다음과 같이 定義된다.

$$C_R=\sum_{i=1}^{R}(x_i-T) \quad \cdots\cdots(1)$$

여기서 C_R은 R번째의 累積合의 값이다.

만일 工程이 推定할 수 있는 目標가 있다면 그것은 데이타가 다음과 같이 나타낼 수 있다.

$$x_i=T+e_i$$

여기서는 e_i는 期待値 0의 값을 갖는 確章変数이다. 그러면

$$C_R=\sum_{i=1}^{R}e_i \longrightarrow 0$$

그러나 工程이 $T+\Delta T+e_i$ 水準에서 流動한다면

$$C_R=\sum_{i=1}^{R}(\Delta T+e_i) \longrightarrow 0$$

i에 대한 C_i의 点은 ΔT의 傾斜로 위쪽으로 모이게 될 것이다.

따라서 累積合은 修正되게 된다.

目標値가 다음과 같이 修正되었다 하자

$$T_R^*=\frac{1}{R}\sum_{i=1}^{R}x_i=\bar{x}_R \quad \cdots\cdots(2)$$

따라서 T_R^*은 R번째의 流動平均이 된다.

그러면 累積平均(cumean)은 다음과 같이 定義된다.

$$C_R^*=\sum_{i=1}^{R}(x_i-T_i^*) \quad \cdots\cdots(3)$$

加重 데이타를 취하게 되면

加重 데이타가 累積平均과 관계가 있는 計劃은 용이하게 決定된다. 考察해 보면

$$C_R^*=\sum_{i=1}^{R}(x_i-T_i^*)$$

그러면

$$C_R^*=\sum_{i=1}^{R}x_i-\sum_{i=1}^{R}\frac{1}{i}\sum_{k=1}^{i}x_i \quad \cdots\cdots(4)$$

그러면

$$C_R^*=x_1\left[1-\sum_{j=2}^{R}\frac{1}{j}\right]-\bar{x}_0+\sum_{i=2}^{R}x_i\left[1-\sum_{j=i}^{R}\frac{1}{j}\right]\cdots(5)$$

여기서는 $\bar{x}_0$는 첫번째 平均값이다.

만일 우리가 $\bar{x}_0=x_1$의 값을 취한다면 우리는 C_R^*의 값을 갖는다.

$$C_R^*=-x_1\sum_{j=2}^{R}\frac{1}{j}+\sum_{i=2}^{R}x_i\left[1-\sum_{j=i}^{R}\frac{1}{j}\right] \quad \cdots\cdots(6)$$

그림 1은 여러가지 R의 값에 대한 데이타의 加重値 w_i의 타점을 보여준다.

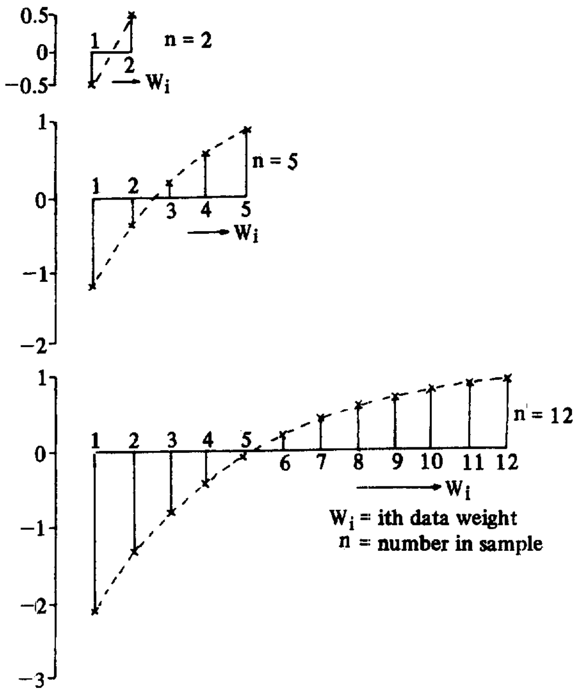

Figure 1. Examples of data weight of Cumean for various data lengths.

加重値에 대한 一般的인 方程式의 差이는

$$w_{i+1}-w_i=\frac{1}{i},\quad i=1,2,\cdots\cdots\cdots,R-1\cdots\cdots\cdots \quad (7)$$

ii) 目標値가 不正確한 때

만일 $x_i=T+e_i$(앞의 累積合을 다룬 보조부문과 같이)라 하면

$$C_R^*=\sum_{i=1}^{R}\left[(T+e_i)-\bar{x}_i\right]$$

$$=RT+\sum_{i=1}^{R}e_i-\sum_{i=1}^{R}\frac{1}{i}\sum_{j=1}^{i}(T+e_j)$$

$$=RT+\sum_{i=1}^{R}e_i-\sum_{i=1}^{R}\frac{1}{i}\left(iT+\sum_{j=1}^{i}e_j\right)$$

$$=RT+\sum_{i=1}^{R}e_i-RT-\sum_{i=1}^{R}\frac{1}{i}\sum_{j=1}^{i}e_j$$

$$C_R^*=\sum_{i=1}^{R}e_i-\sum_{i=1}^{R}\frac{1}{i}\sum_{j=1}^{R}e_j\longrightarrow 0\quad\cdots\cdots\cdots\cdots\cdots(8)$$

여기서 만약 T가 $T+\Delta T$로 대체할 수 있다면 우리는 식(8)에서 주어진 것과 똑같은 表現을 할 수 있다.

그래서 累積平均은 目標値를 갖지 않은 경우 자신이 출발하고 자신이 조정된다.

3. 累積平均(cumean)의 반복계획

여기서

$$C_R^*=\sum_{i=1}^{R}(x_i-T_i^*)$$

고로

$$C_R^*=\sum_{i=1}^{R-1}(x_i-T_i^*)+x_R-T_R^*$$

고로

$$C_R^*=C_{R-1}^*+(x_R-T_R^*)$$

또한

$$T_R^*=\frac{1}{R}\sum_{i=1}^{R}x_i$$

고로

$$T_R^*=\frac{1}{R}\left(\sum_{i=1}^{R-1}x_i+x_R\right)$$

고로

$$T_R^*=\frac{1}{R}[(R-1)T_{R-1}^*+x_R]$$

고로

$$T_R^*=\left(\frac{R-1}{R}\right)T_{R-1}^*+\frac{1}{R}x_R\quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots(9)$$

$$C_R^*=C_{R-1}^*+\left(\frac{R-1}{R}\right)(x_R-T_{R-1}^*)\quad\cdots\cdots\cdots\cdots(10)$$

식(9)와 (10)을 利用하여 간단한 반복計劃이 수립될 수 있다. 그림 2는 累積平均의 도구로서 計算式을 보여주고 있다.

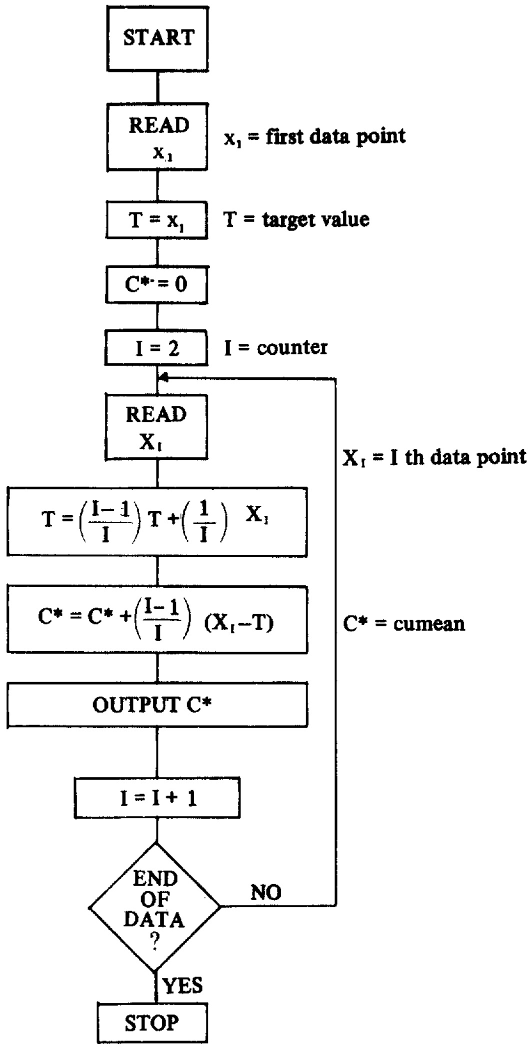

Figure 2. Cumean algorithm

따라서 $R\rightarrow\infty$로 가면

$$T_R^*\rightarrow T_{R-1}^*$$

그리고

$$C_R^*\rightarrow C_{R-1}^*+(x_R-T_R^*)$$

目標値 T_R^*이 一定한 값을 갖는 傾向이 있고 그 計劃이 累積合(cusum)과 마찬가지로 效果的이라면 위와 같이 될 수가 있다.

4. 事例

累積合(cusum)과 累積平均(cumean)의 比較를 說明한다는 것은 컴퓨터 시뮬레이션이 수행된다.

正規分布의 랜덤넘버와 集合이 산출된다.

이 分布는 평균치 0와 표준편차 1의 값을 갖는다.

이러한 30개의 숫자의 평균을 1로 바꾼다.

이 시뮬레이션은 30개의 숫자를 읽은 후에 平均값 1로 바뀐다.

이러한 양쪽의 랜덤넘버의 集合은 目標値 0와 累積平均을 갖는 累積合에 의해 처리된다.

이들 点의 타점이 그림 3에 보여지고 있다.

첫번째 30개의 랜덤넘버가 目標値 0.16의 값을 갖는 것을 볼 수 있다.

平均값의 變化段階의 効果는 양계획에서 쉽게 알아낼 수 있다. 그러나 累積平均은 이 變化의 水平点에서 그 基本을 이루게 된다.

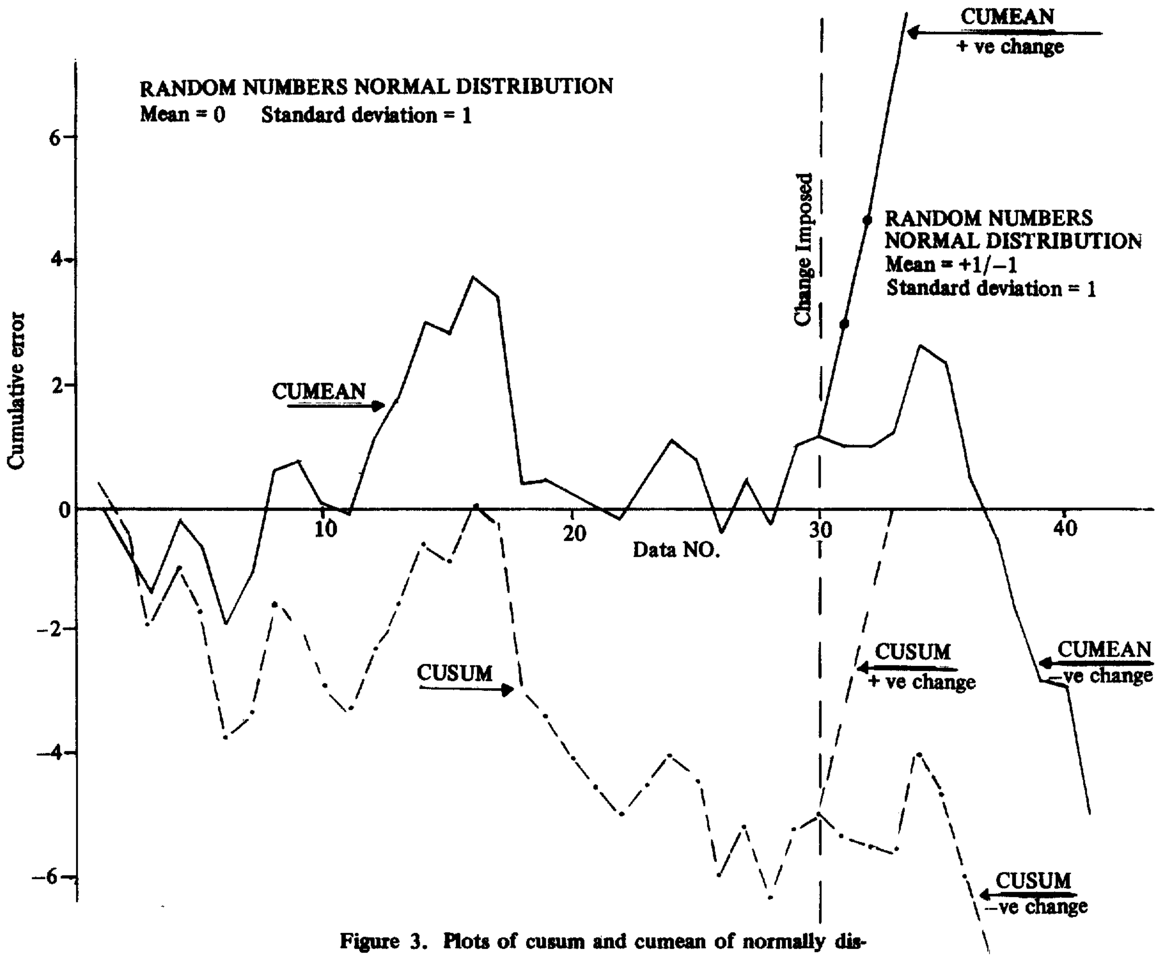

Figure 3. Plots of cusum and cumean of normally distribution pseudorandom numbers with step change in mean after 30 examples

5. 實際的인 事例

한 製造業者는 열처리시스템의 初期條件을 管理하기 위한 자동검사기를 사용하고 있다. 그 管理項目중 하나가 볼트의 測定이었다. 設計상으로 볼트의 期待값은 1.6 볼트를 나타내고 있었다. 그래서 제조자는 生産이 개시되었을 때 이 측정치를 감시하도록 요구하고 있다.

그래서 저자는 데이타 플로트계획을 수립하고 累積合과 累積平均 값을 이 技法에 使用하기로 하였다. 그림 4는 양 累積合의 플로트를 보여준다. 또

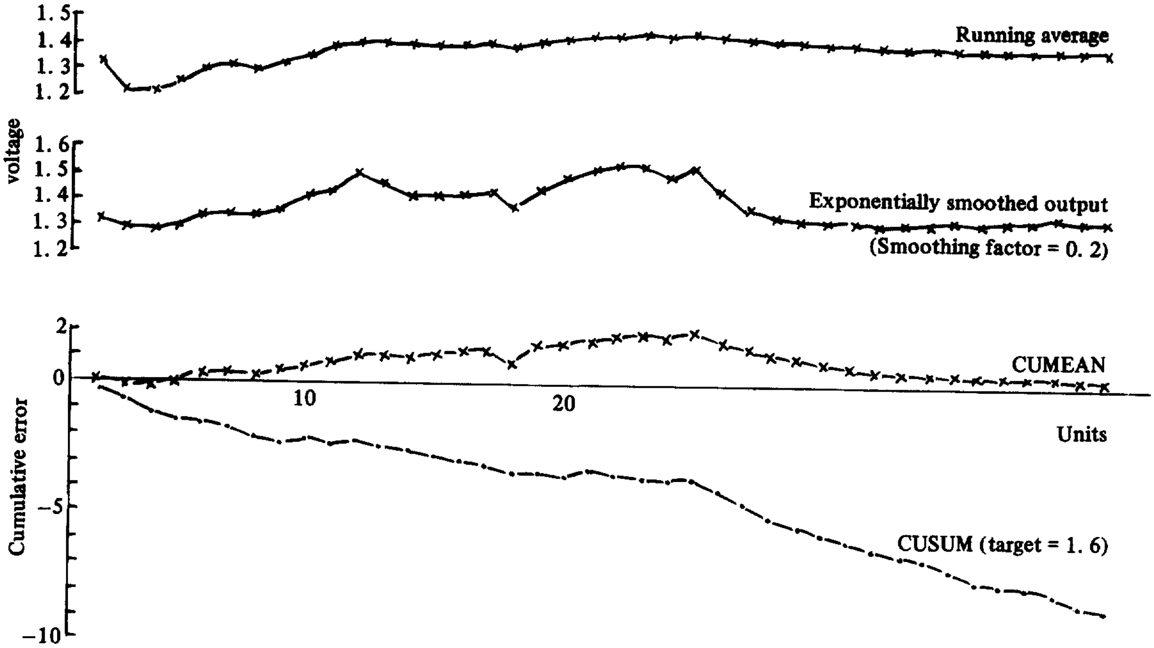

Figure 4. Cusum and Cumean plots together with running average and exponentially smoothed output of a voltage measurement of a fire system.

한 連續平均과 指數平滑法(平滑係數=0.2)의 比較를 나타낸다.

平準化 데이터의 플로트는 초기 測定値 24단위는 상승하는 傾向을 보이고 그 다음은 거의 1.35 값에 정지된 상태의 값을 보인다.

累積合은 初期 24단위는 平均水準이 目標値보다 떨어지고 있다. 이러한 현상은 어떤 傾向條件을 나타내지는 않는다. 왜냐하면 累積合은 平均 값에서 變化된 段階를 나타내기 때문이다.

따라서 累積平均 자신이 0의 값을 갖기 때문에 볼트 測定의 傾向은 累積平均을 플로트 하는 데서 觀測된다.

結　論

累積合(cusum) 대신에 累積平均(cumean)을 사용할 경우 어떤 工程에 대한 目標値 값을 推定할 수는 없으나 보다 많은 데이타의 累積合이 定해지면 目標値에 적합하도록 累積平均의 값을 처리할 수 있다.

連續된 期間동안 目標値計劃과 그 計劃을 수행할 수 있는 目標값을 수립해야 되는 경우 標準 累積合 計劃을 사용할 수도 있으나 累積平均의 값을 利用하는 것이 傾向變動을 推定하는데 效果的이다.

《參考文獻》

1. Mendenhall, Richard L. Scheaffer. *"Mathematical statistics with Applications"* 1978, pp 527 541, J. Wiley & sons Inc.
2. Ewan, W. D. *"When and how to use the cusum"*. Technometrics, 1963 Vol. 5, NO.1.
3. Hoel, P. G. *"Elementary Statistics"*, 1969. pp 254 260, J. Wiley & sons Inc.
4. Bendat, J. S. and piersol, A. G. *"Measurement and analysis of Random data,"* J. Wiley & sons Inc.
5. Myles Hollander. Douglas A. Wolfe, *"Nonparametric Statistical Methods"* 1969. pp 27 33, J. Wiley & sons Inc.