Adaptive EWMA control chart with a Dynamic Sampling Scheme

Samuel Anyaso-Samuel

11/05/2019

Introduction

Statistical process control (SPC) provides a variety of statistical tools for monitoring sequential processes to ensure such processes run stably. One ideal goal of SPC is to give signals for distributional shifts caused by disturbances in the process that is being monitored. SPC has been employed in a wide variety of fields such as mamufacturing and production lines, internet services, health care, E-commerce and so on. These processes are posed to generate lots of data and thus it becomes necessary to continually monitor the performance of the working components of these processes and raise alarms at the occurrence of any aberrant pattern.

The process readings generated from such data applications are can be described as a random sample from a parametric statistical distribution. During the working phase of the process, it is not uncommon for the parametric value(s) of the statistical distribution or even the statistical distribution which describes the data to change over time. This change may be caused by controllable variations in the working components of the process, therefore it becomes ideal to identify and correct such variations.

In the SPC literature, the sequential process can be described as either being in-control (IC) or out-of-control (OC). The process is said to be IC when there are only unavoidable randomness in the process, while the process is said to be OC when there are controllable sources of variation (such as malfunction of certain components of a manufacturing process). In addition, SPC is usually discussed in two phases - the phase I and phase II SPC. The goal of the phase-I SPC lies in estimation of the distribution of the process readings, construction of valid control limits for the control chart and obtaining clean (IC) readings from the process. This involves continuous fine tuning of the process until all controllable variations in the process are eliminated. On the other hand, the real time monitoring of the process happens in the phase-II SPC. Our discussion will be majorly based on the phase-II SPC.

The control chart is a popular tool for monitoring the performance of a sequential process. The control chart is a plot of successive readings from a quality charateristic, the chart is bounded by upper and/or lower control limit(s). The control chart gives a signal whenever their charting statistic plots beyond the control limits. A wide variety of control charts have been developed for detecting distributional shifts in sequential processes, nonetheless, the methods and discussion for this report will be based on exponentially weighted moving average (EWMA) control chart.

The IC average run length (ARL_0) and the OC average run length (ARL_1) are common metrics used to evaluate the performance of the SPC chart. The IC ARL is the expected number of IC process readings plotted on the chart before an OC signal is given while the OC ARL is the expected number of points from the occurrence of a shift up to when the control chart gives a control signal. The IC ARL is usually pre-specified and the optimal control chart is the chart that achieves the least OC ARL. Next, we present a brief description of the EWMA control chart.

Suppose, we have process readings

$$\begin{cases} X_1, X_2, ..., X_\tau \overset{\text{IID}}{\sim} N(\mu_0, \sigma^2) \\ X_{\tau+1}, X_{\tau+2}, ... \overset{\text{IID}}{\sim} N(\mu_1, \sigma^2) \end{cases} \tag{1}$$

where μ_0 and μ_1 are the IC and OC process means respectively, and $\mu_0 \neq \mu_1$, σ^2 is the constant variance of the process readings, and τ, $1 \leq \tau \leq \infty$, is an unknown change point. The charting statistic of the EWMA control chart is given as

$$\begin{cases} E_t^+ = \max\left(0, \nu(X_t - \mu_0) + (1-\nu)E_{t-1}^+\right) \\ E_t^- = \min\left(0, \nu(X_t - \mu_0) + (1-\nu)E_{t-1}^-\right) \end{cases} \tag{2}$$

and the control chart gives a signal when

$$E_t^+ > \rho \quad \text{or} \quad E_t^- < -\rho, \qquad \text{for } t \geq 1, \tag{3}$$

where $E_0 = \mu_0$, $\nu \in (0,1]$ is a weighting parameter, and ρ is chosen to reach a pre-specified IC ARL value.

The control chart described above gives a signal whenever its charting statistic plot beyond its control limits. That is, they are able to provide information about the perfromance of the process at a current time point, however, they are unable to provide information on the performance of the process in the near future. Such futuristic information are useful in making appropriate corresponding decisions. For instance, such information can be incorporated in variable sampling schemes. That is, we delay the monitoring of the next process reading if there is little evidence of a shift and monitor the next process reading immediately if there is substantial evidence of a shift. Li and Qiu (2014) suggested that the p-value will be a good quantitative measure of the future performance of the process.

The control chart described in the previous section have a common property - it has a fixed sampling scheme (FSS). A control chart with FSS is designed to monitor the process reading at every time point, i.e. the sampling interval $d(\cdot) = 1$. In constrast to this convention, Li and Qiu (2014) proposed a variable sampling scheme which they called the dynamic sampling scheme. For this scheme, the sampling interval $d(\cdot)$ is a continuous function of the p-value of the charting statistic of the control chart. The sampling interval function is approximated by a parametric function in the Box-Cox transformation family.

Methods

In this section, first, we describe a new version of the EWMA control chart called the adaptive EWMA control chart. Then, we discuss the design of the EWMA chart with p-values, also, we describe the form of the dynamic sampling scheme.

Adaptive Selection of the Weighting Parameter ν

The weighting parameter ν of the EWMA control chart given in (2) is usually pre-specified based on the shift size to be detected. However, in real-world applications, the shift size δ^*, at a given time point is usually unknown. Thus, the EWMA chart may not have optimal performance if an inappropriate value of ν is selected. To overcome the limitation of selecting a wrong value for ν, several methods have been proposed to estimate the value of δ^*, and ν adaptively. For this report, we focus on the recently proposed method of Haq, Gulzar, and Khoo (2018) which we shall call adaptive EWMA chart herein. We employ this method because amongst other methods, Haq, Gulzar, and Khoo (2018) showed that their method yields optimal performance for detecting a varierty of shift sizes.

The adaptive EWMA chart continuously estimates the unknown process mean shift using the classic EWMA statistic, and based on the value of the estimate, the weighting parameter to be used in adaptive EWMA control chart will be chosen accordingly.

First, we begin with the estimator of the shift size δ^*. Let $\psi \in (0,\ 1]$ be a smoothing constant, then an estimator of δ^* is given by

$$\hat{\delta}_t^* = \psi X_t + (1-\psi)\hat{\delta}_{t-1}^* \tag{4}$$

where $\hat{\delta}_0^* = 0$. The estimator has been shown to be unbiased for processes that are IC, but biased for processes that are OC. An unbiased estimator for when the process is both IC and OC is given by

$$\hat{\delta}_t^{**} = \frac{\hat{\delta}_t^*}{1-(1-\psi)^t} \tag{5}$$

Then, we define $\tilde{\delta}_t = |\hat{\delta}_t^{**}|$ when estimating δ^*. The resulting charting statistic for detecting upward or downward mean shift becomes

$$\begin{cases} E_0 = 0 \\ E_t = \theta(\tilde{\delta}_t)X_t + (1-\theta(\tilde{\delta}_t))E_{t-1}^*, \end{cases} \tag{6}$$

where $\theta(\tilde{\delta}_t)$ is defined by

$$\theta(\tilde{\delta}_t) = \begin{cases} 0.015 & \text{if } 0.00 < \tilde{\delta}_t \leq 0.25 \\ 0.10 & \text{if } 0.25 < \tilde{\delta}_t \leq 0.75 \\ 0.20 & \text{if } 0.75 < \tilde{\delta}_t \leq 1.00 \\ 0.25 & \text{if } 1.00 < \tilde{\delta}_t \leq 1.50 \\ 0.50 & \text{if } 1.50 < \tilde{\delta}_t \leq 2.50 \\ 0.80 & \text{if } 2.50 < \tilde{\delta}_t \leq 3.50 \\ 1.00 & \text{if } 3.50 < \tilde{\delta}_t \end{cases} \tag{7}$$

For the control chart defined in (6), notice that small values of the estimate of the weighting parameter, $\theta(\tilde{\delta}_t)$ will be chosen for small values of $\tilde{\delta}_t$ while large values of $\theta(\tilde{\delta}_t)$ will be chosen for large values of $\tilde{\delta}_t$. For the initial weighting parameter defined in (4), Haq, Gulzar, and Khoo (2018) recommend setting $\psi \geq 0.1$.

This control chart gives a signal for a mean shift whenever the charting statistic defined in (6) exceeds a pre-specified control limit, h.

The Adaptive EWMA control chart with p-values

Our discussion in this section, and in the following sections will be focused on the two-sided adaptive EWMA control chart defined in (6). The adaptive EWMA control chart does not provide information on the likelihood of a potential mean shift in the near future. In order to circumvent this limitation, we design EWMA chart with p-values. Suppose E_t^* is the observed value of the EWMA charting statistic for detecting an arbitrary mean shift at time point t, then the p-value of the charting statistic at this time point is given as

$$P_{E_t^*} = \mathbb{P}(E_t > E_t^*) \tag{8}$$

At the t-th time point, we conclude that the process is likely to be OC if

$$P_{E_t^*} < \alpha \tag{9}$$

otherwise, it is IC. If $P_{E_t^*}$ is much larger than α, this indicates that the process is likely to be IC at such time point, where as if $P_{E_t^*}$ is only marginally greater than α, the chart indicates that we should be wary of the process. On the other hand, the process should be stopped if $P_{E_t^*}$ is strictly less than α. In contrast with the classic EWMA chart and the adaptive EWMA chart, the control limit for the adaptive EWMA chart with p-values can be said to be the pre-specified significant level α.

Notice that the p-value computed for the control chart provides a quantitative measure to assess the performance of the chart at a given time point. With this measure, we can design adaptive charts which have variable sampling schemes (VSS). Recall that for control charts with fixed sampling schemes, the ARL is used evaluate the performance of the control. Now, for control charts with VSS, the ARL will can no longer be an ideal evaluation metric. In the literature, the average time to signal (ATS) and the adjusted average time to signal (AATS) are performance measures for control charts with VSS. The ATS is defined as the expected value of the time interval from the start of the Phase-II process monitoring to the time when the chart gives an OC signal. While the AATS is defined as the expected value of the time interval from the occurrence of a shift to the time when the chart gives an OC signal. As in the case of schemes with fixed sampling rates, we usually fix the IC ATS value. The control chart with a larger IC ATS will have a lower false alarm rate, and the chart with the smallest OC AATS will perform best for detecting a specific shift size.

The Dynamic Sampling Scheme

Li and Qiu (2014) proposed the dynamic sampling scheme which is an increasing function of the p-value. They constructed the sampling scheme in terms of the CUSUM control chart. Here, we describe the sampling scheme in terms of the adaptive EWMA control chart. The scheme which follows from the Box-Cox family is defined as

$$d(P_{E_t}) = \begin{cases} a + bP_t^{\lambda} & \text{if } \lambda > 0 \\ a + b\log(P_t) & \text{if } \lambda = 0, \end{cases} \tag{10}$$

where P_t is the p-value of the observed charting statistic at time point t. Using numeric experimentation in the CUSUM chart setting, Li and Qiu (2014) estimated $a = 0$, $\lambda = 2$ and b is set to achieve a given ATS_0 value. Thus

$$d(P_{E_t}) = b \cdot P_{E_t}^2 \tag{11}$$

Likewise, when estimating the parameters of Equation (10) in the EWMA chart setting, we got the same estimated values as Li and Qiu (2014).

Computing the p-value

In practice, the distribution of the charting statistic of interest is usually unknown. In order to compute the p-value, first, we employ the bootstrap method to estimate the empirical distribution of the charting statistic. In this case, let's assume that we have IC process readings from the sequential process. By the bootstrap principle, we repeatedly obtain resampled data from the available IC data, and these resampled data are then used to compute the charting statistic in the phase-II monitoring of the process. This step is repeated B times and the resulting values of the estimated charting statistic are used to estimate the empirical distribution of the charting statistic of interest. If sufficient information is available in the observed process readings, and for large B bootstrap size, we expect the empirical distribution to be a good approximation of the true IC distribution of the charting statistic. With this empirical distribution, we can then compute the p-value of an observed charting statistic at a given time point.

The Adaptive EWMA Control Chart with a Dynamic Sampling Scheme

Bringing the methods together, our proposed control chart is called the adaptive EWMA control chart with a dynamic sampling scheme. This control chart has the following properties. First, it adaptively estimates the weighting parameter of the EWMA control chart using the method proposed by Haq, Gulzar, and Khoo (2018). Secondly, the adaptive EWMA control chart is designed with p-values with a decision rule given in (9). Lastly, this chart is designed with the dynamic sampling scheme given in (11). We call the proposed control chart DyS-AEWMA herein.

With the incorporation of the dynamic sampling scheme in the design of the adaptive AEWMA control chart with p-values, we are able to skip certain observations when the process is judged to be IC. By skipping observations that are in-control, the control chart gains efficiency in reduction of monitoring run-time.

Functions

In this section, we provide detailed descriptions of the functions in the `DyEWMA` package. There are a total of 5 functions in this package.

- `arl` computes the average run length (ARL) and standard deviation of the run length for the adaptive EWMA chart. The function is written in `Rccp` for efficiency. The function takes 3 arguments, namely
  - `h` a real number; the control limit
  - `omg` a real number; smooth constant which lies between 0 and 1, *default value is 0.10*
  - `shift` a real number; shift size. If shift=0, IC ARL is returned, else OC ARL is returned
  - `chart_type` a character; either "one" or "two". Select "one" for one-sided chart and "two" for two-sided chart.

  The function returns a numeric vector of length 2, where the first element is the estimated ARL and the second element is the SD of the ARL.

- `empr_aewma` Estimates the in-control empirical distribution of the adaptive EWMA control chart. This function is written in `Rcpp` for efficiency. The function takes in 3 arguments, namely
- `x` numeric vector of the IC observations from the sequential process
- `nsimul` an integer, number of bootstrap simulations, and
- `w` an integer, the sample size needed to reach steady-state.

The function returns a numeric vector which is the estimated empirical distribution of the AEWMA control chart for the given IC data observations.

- `pVal` Computes the p-value from the empirical distribution of the adaptive EWMA control chart. For efficiency, the function is written in `Rcpp`. The function takes in 2 arguments namely
- `empr_dist` a numeric vector; the empirical distribution
- `obsStat` a real number; the observed test statistic.

The function returns the estimated p-value of the charting statistic.

- `arl_aewma` Estimates the ARL of the adaptive EWMA control chart with p-values. In this case, $d(\cdot) = 1$. The `arl_aewma` function also calls the `empr_aewma` and the `pVal` functions for the estimation procedure. This function is also written in `Rcpp` for efficiency. The function takes 3 arguments namely
- `alpha` a real number; the level of significance
- `w` an integer; the sample size needed to reach steady-state, *default value is 50*
- `nsimul` an integer; the number of replications
- `shift` a real number; shift size. If shift=0, IC ARL is returned, else OC ARL is returned

The function returns the ARL of the adaptive EWMA chart.

- `ats_arl_aewma` Estimates the ATS and the ARL of the adaptive EWMA control chart with a dynamic sampling scheme. This function is also written in `Rcpp` for efficiency. For the computation of the ATS, we use the sampling interval described in Equation (11), while for the ARL, the sampling interval is given as $d(\cdot) = 1$. Similar to the `arl_aewma` function, the `ats_arl_aewma` also calls the `pVal` and the `empr_aewma` functions for the estimation procedure. The function takes the following arguments.
- `alpha` a real number; the level of significance
- `w` an integer; the sample size needed to reach steady-state, *default value is 50*
- `nsimul` an integer; the number of replications
- `a` a real number; a parameter in the dynamic sampling interval function
- `lambda` a real number; lambda parameter in the dynamic sampling interval function
- `b` a real number; b parameter in the dynamic sampling interval function
- `shift` a real number; shift size for OC ARL/ATS computation. (=0 for IC ARL/ATS)

The function returns 2 elements which are ATS, and ARL (in the case where $d(\cdot) = 1$).

Results

In this section, we present results when the functions described in the preceeding section are applied. The package can be installed by

```
devtools::install_github("samuelanyaso/DyAEWMA", build_vignettes = TRUE)
```

If prompted by the installer to update dependencies, please skip the updates by hitting return.

First, we use the `arl` function to produce results stated in first column of Table 1 of Haq, Gulzar, and Khoo (2018). Both results are based on random simulation runs, therefore, we give allowance for a little margin of error. For ARL=100, the result shown below gives the run length profiles for the adaptive EWMA chart using different values of the smoothing constant ψ and control limits.

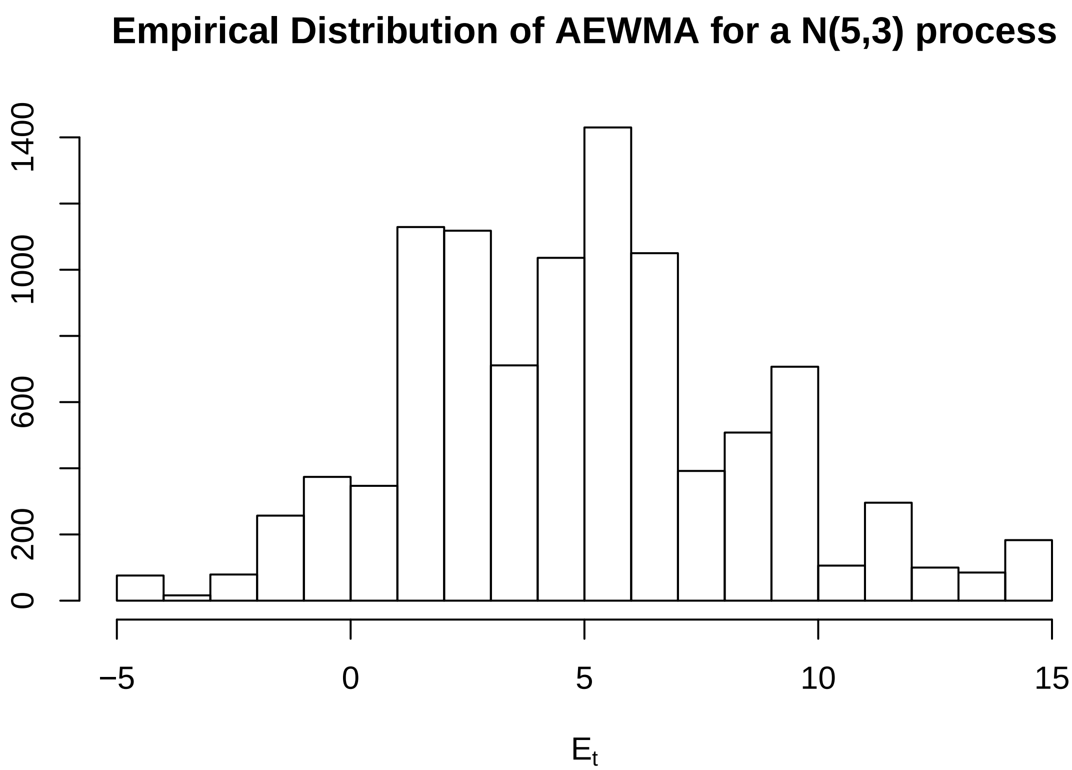
```r
h <- c(0.5762,0.6300,0.7139)
omg <- c(0.10,0.15,0.20)
delta <- c(0.00,0.25,0.50,0.75,1.00,1.50,2.00,2.50,3.00,3.50,4.00,5.00,6.00)
n <- length(delta)*2
aarl <- Vectorize(arl, vectorize.args = c("shift"))
res <- matrix(NA,nrow=n,ncol=length(omg))
for(j in seq_along(h)){
  ARL <- aarl(h=h[j],omg = omg[j],shift = delta)
  ARL <- as.vector(ARL)
  res[,j] = ARL
}
res <- as.data.frame(round(res,2))
RN <- c()
for(i in seq_along(delta)){
  RN <- c(RN,delta[i],"")
}

RN2 <- rep(c("ARL","SD"),length(delta))
result <- data.frame(delta=RN, psi=RN2)
result <- cbind(result,res)
colnames(result) <- c("delta","psi",0.10,0.15,0.20)
result
```

```
##    delta psi    0.1   0.15    0.2
## 1      0 ARL 100.68 100.33 101.28
## 2         SD 128.15 123.96 121.51
```

```
## 3   0.25 ARL  32.03  34.09  36.53
## 4          SD  35.40  37.62  40.06
## 5    0.5 ARL  12.61  13.13  13.85
## 6          SD  13.22  13.68  13.96
## 7   0.75 ARL   6.47   6.71   7.15
## 8          SD   6.56   6.65   6.79
## 9       1 ARL   3.96   4.13   4.43
## 10         SD   3.67   3.73   3.92
## 11   1.5 ARL   2.12   2.21   2.32
## 12         SD   1.55   1.63   1.78
## 13      2 ARL   1.48   1.51   1.55
## 14         SD   0.84   0.90   0.97
## 15   2.5 ARL   1.20   1.21   1.22
## 16         SD   0.50   0.53   0.55
## 17      3 ARL   1.08   1.08   1.08
## 18         SD   0.29   0.30   0.31
## 19   3.5 ARL   1.02   1.02   1.02
## 20         SD   0.16   0.16   0.16
## 21      4 ARL   1.01   1.01   1.01
## 22         SD   0.08   0.08   0.08
## 23      5 ARL   1.00   1.00   1.00
## 24         SD   0.02   0.02   0.02
## 25      6 ARL   1.00   1.00   1.00
## 26         SD   0.00   0.00   0.00
```

Next, we use the `empr_aewma` function to estimate the in-control empirical distribution of the adaptive EWMA control chart given in Equation (6). In this case, we assume that we have 100 in-control process readings from a N(5,3) process.

```
x <- rnorm(100,5,4)         # IC process readings
hist(empr_aewma(x = x,nsimul = 10000,w = 50),
     main="Empirical Distribution of AEWMA for a N(5,3) process",
     xlab=expression(E[t]),ylab = "")
```



Next, we show results when we use the `pVal` function to compute the p-value of an observed test statistic. First, let's consider the case where the distribution of the charting statistic is known. The next chunk of code compares the performance of `pVal` function with the standard functions available in `R`.

- Suppose the known distribution of the charting statistic is N(0,1)

```
z <- rnorm(100000);
obs <- rnorm(1,sd = 4)                          # Observed test statistic

# Comparing the p-value for a two tailed test,
# computed using the standard R function and pVal function
Ex1 <- 2*pnorm(abs(obs),lower.tail = F)
Ex2 <- 2*pVal(z,abs(obs))
print(round(c(Ex1, Ex2),4))
```

```
## [1] 0.0034 0.0034
```

```
# Comparing the p-value for a one tailed test (right tailed),
# computed using the standard R function and pVal function
Ex3 <- pnorm(obs,lower.tail = F)
Ex4 <- pVal(z,obs)
print(round(c(Ex3, Ex4),4))
```

```
## [1] 0.9983 0.9982
```

```
# Comparing the p-value for a one tailed test (left tailed),
# computed using the standard R function and pVal function
Ex5 <- pnorm(obs,lower.tail = T)
Ex6<- 1 - pVal(z,obs)
print(round(c(Ex5, Ex6),4))
```

```
## [1] 0.0017 0.0018
```

- Suppose the known distribution is a skewed distribution, say, χ^2(df=3)

```
z <- rchisq(100000,df=3);
obs <- rchisq(1,df = 3)                          # Observed test statistic
Ex1 <- pchisq(obs,df=3,lower.tail = F)                # standard R function
Ex2 <- pVal(z, obs)                              # two tailed test
print(round(c(Ex1, Ex2),4))
```

```
## [1] 0.2184 0.2185
```

- Now, for most real-world applications, the distribution of the charting statistic is usually unknown. In this setting, we estimate the distribution of the charting statistic using the `empr_aewma` function and then, we use the `pVal` function to compute the p-value of an observed charting statistic at a given time point. For this example, let's assume that we have available in-control process readings from a N(4,3) process.

```
 # available IC process readings
x <- rnorm(1000,mean = 4, sd=3)

# estimates the empirical distribution of the charting statistc
z <- empr_aewma(x=x, nsimul=10000, w = 50)

# an observed charting statistic at a given time point
obs <- 3.56

# the p-value
pVal(z, obs)
```

```
## [1] 0.5622
```

- For a given level of significance (the control limit),say $\alpha = 0.014$ we compute the ARL of the adaptive EWMA chart with p-values.

```
arl_aewma(alpha = 0.14,w = 50,nsimul = 1000,shift = 0)
```

```
## [1] 110.4972
```

- For a given level of significance, say $\alpha = 0.01429$ and sampling scheme parameters $a = 0,\ \lambda = 2,\ b = 2.74978$, we compute the ATS of the adaptive EWMA chart with p-values.

```
ats_arl_aewma(alpha = 0.01429,w = 50,nsimul = 1000,a = 0,lambda = 2,b = 2.74978,shift = 0)
```

```
## [1] 380.6445 370.7181
```

Conclusion

In this report, we have described the adaptive EWMA control chart with a dynamic sampling scheme. Further, we have described the functionality of the `DyAEWMA` package. The core of this package lies in estimating the ATS and/or ARL of DyS-AEWMA or AEWMA control charts. This package also contains a function for computing the p-value of an observed charting statistic based on the in-control empirically distribution of AEWMA.

As in the classical statistical process control setting, the control limit(s) to achieve a particular ARL value is usually pre-specified. This package does not estimate the control limit(s). One can easily compute the control limit(s) for the DyS-AEWMA using the algorithm shown in Page 129 of Qiu (2014). We have opted against including a function to compute the control limit, because the algorithm seem to be expensive in the case of the DyS-AEWMA. In the future, we aim to improve the efficiency for such functionality and then include it in the package.

Due to the iterative nature of the functionalities in this package, we have written the codes in `Rccp` for efficiency. Since some users will certainly favor large number of iterations when estimating the ARL and/or ATS, we aim to further improve the speed and efficiency of the functions in the future. Also, the idea behind the computation of the ARL/ATS can be found in page 127 of Qiu (2014).

Further, since the topic of SPC is a rapidly evolving research area, more updates will be continually made to this package as research evolves. In the future, we aim to incorporate dynamic sampling schemes in the design of other popular SPC charts.

REFERENCES

Haq, Abdul, Rabia Gulzar, and Michael B. C. Khoo. 2018. "An Efficient Adaptive EWMA Control Chart for Monitoring the Process Mean." *Quality and Reliability Engineering International* 34 (4): 563–71. doi:doi:10.1002/qre.2272.

Li, Z., and P. Qiu. 2014. "Statistical Process Control Using a Dynamic Sampling Scheme." *Technometrics* 56 (3): 325–35.

Qiu, P. 2014. *Introduction to Statistical Process Control*. Chapman; Hall/CRC.