Control Charts Based on Machine and Statistical Learning Techniques for Monitoring Bivariate Negative Binomial Process

Aaron T. Allen
Department of Mathematics and Statistics,
Washington State University, Pullman, WA, USA

Abstract

In this paper, we adapt several competing machine learning techniques to develop control charts for monitoring Bivariate Negative Binomial disease counts. With each chart, one can simultaneously monitor correlated multiple counts and distinguish between stable and out-of-control data. We use an average run length study to determine the best performing techniques. We apply the proposed methods on simulated data to demonstrate their usefulness over traditional simultaneous control charting.

Keywords: Average run length, bivariate negative binomial, control charts, disease monitoring, Hepatitis C, linear discriminant analysis, logistic regression, statistical process monitoring, support vector data description.

Contents

1 Introduction

In this paper, the problem of disease control specifically for Hepatitis-C will be our focus, but these applications can easily be extended to similar cases. We will be using Hepatitis-C counts from two states in Australia: New South Wales (NSW) and South Australia (SA). Our original data set consists of counts for the disease in the two states over forty months. It is largely of interest to most disease control agencies to ensure that a disease does not become out of control. Since there is natural variability in these counts, an appropriate statistical model that represents the process in stable conditions must first be determined and objective means of detecting departures from this model should be established.

This is where control chart monitoring has been in used practice for years. The objective of control charts is to establish some objective reference point, typically an upper control limit for disease counts, that allows for classification of the current status of disease occurrence. The counts will either be below the upper control limit indicating the disease is under control or above the upper control limit indicating action may be required by health authorities. These control limits in most cases are calculated according to some previous understanding

of the disease in the given environment.

The extension of this problem to the use of machine learning, in part, is due to the data set. Certainly, control charts can be created for a single variable (in our case disease counts in one Australian state); however, considering multiple variables (states) simultaneously adds some difficulty. This is especially true in the case of correlation between the variables, which, we will find, is the case here due to geographical proximity.

2 Statistical Models and Assumptions

2.1 Distributions and Probability Functions

A general form of the univariate Negative Binomial probability mass function (pmf) is

$$f(x|k,m) = \left(\frac{k}{m+k}\right)^k \frac{\Gamma(x+k)}{x!\Gamma(k)} \left(\frac{m}{m+k}\right)^x, \quad x = 0,1,2,\ldots \tag{1}$$

where the parameters are $k, m > 0$. The mean and variance are given by $\mathrm{E}(X) = m$ and $\mathrm{Var}(X) = m(1+m/k)$. Note that $E(X) < \mathrm{Var}(X)$, and the NB distribution is an appropriate model for processes with overdispersion. In contrast, the Poisson distribution is constrained to have equal mean and variance (equidispersion).

The authors in [2] present a means of modeling correlation between two Negative Binomial variables X_1 and X_2 so that each is marginally distributed Negative Binomial (NB) with pmf of the form (1). We adopt their approach to obtain random samples from the bivariate NB (BNB) distribution. Generate θ from a Gamma distribution with shape ρ and scale 1. Generate X_i from a Poisson distribution with mean $\lambda_i\theta$ for $i = 1, 2$. It can be shown that the joint probability mass function of (X_1, X_2) is given by

$$\begin{aligned} f(x_1,x_2|\lambda_1,\lambda_2,\rho) &= \frac{\Gamma(x_1+x_2+\rho)}{x!\; x_2!\; \Gamma(\rho)} \left(\frac{\lambda_1}{\lambda_1+\lambda_2+1}\right)^{x_1} \left(\frac{\lambda_2}{\lambda_1+\lambda_2+1}\right)^{x_2} \left(\frac{1}{\lambda_1+\lambda_2+1}\right)^{\rho}, \\ & x_1, x_2 = 0, 1, \ldots \end{aligned} \tag{2}$$

with parameters $\lambda_1, \lambda_2 > 0, \rho > 0$. The marginal distribution of X_i is NB with parameters $k = \rho$ and $m = \rho\lambda_i$ for $i = 1, 2$. Thus, we say that (X_1, X_2) have a BNB distribution with joint pmf given by (2). Note that $\mu_{X_i} = \mathrm{E}(X_i) = \rho\lambda_i$, $\sigma^2_{X_i} = \mathrm{Var}(X_i) = \rho\lambda_i(1+\lambda_i)$ for $i = 1, 2$, $\mathrm{Cov}(X_1, X_2) = (1/2)\rho^2\lambda_1^2\lambda_2^2\sqrt{\lambda_1(1+\lambda_1)\lambda_2(1+\lambda_2)}$, and $\mathrm{Corr}(X_1, X_2) = (1/2)\rho\lambda_1^2\lambda_2^2$.

2.2 Hepatitis C Data and BNB

As in almost all situations where one is presented with a data set, our first question should be about the underlying distribution. Since we are dealing with count data, our most likely candidates are the Bivariate Negative Binomial (BNB) and the Bivariate Poisson distributions. Using the forty month counts for our two states, quantile-quantile (probability) plots can be generated for the processes to find its approximate distribution. It can be seen in Figure 1 that a BNB distribution is an appropriate description of the disease counts for NSW and SA. It can also be noted that the data provide evidence of overdispersion, that is, $E(X_i) < \mathrm{Var}(X_i)$ for $i = 1, 2$.

2.3 Maximum Likelihood Estimation

To find the stable process values for this distribution, we will use maximum likelihood estimates (MLEs). Let $\{(x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})\}$ represent a random sample from BNB and assign $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \rho)$. The likelihood function is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_{1i}, x_{2i}|\boldsymbol{\theta})$$

where $f(x_{1i}, x_{2i}|\boldsymbol{\theta})$ is given by (2). The MLE of $\boldsymbol{\theta}$ is the set of values that maximizes the likelihood or the log likelihood.

MLEs will give a baseline for the distribution when the process is stable or under control. To determine them, we need to remove out of control points that have already been recorded within this data set. To do this, we use the following iterative sequence:

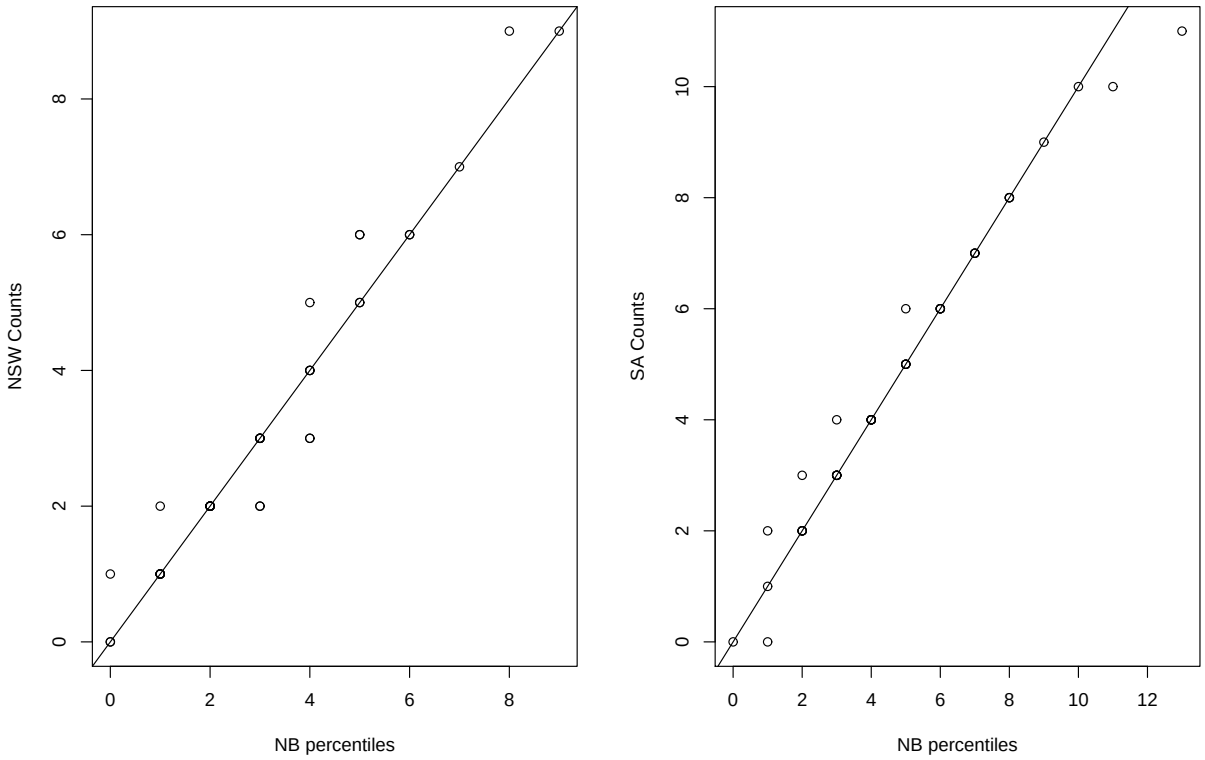

Figure 1: Quantile-Quantile plots for BNB and Bivariate Poisson

1. Calculate the MLEs for the joint probability mass function.

2. Create upper control limit charts for the two states individually, using the marginal probability mass functions.

3. Remove points classified as out of control in either of the states and return to Step 1 until all remaining points are classified as in control on both charts.

The corresponding set of MLEs based on the final reduced data set will be used as stable-process parameter values. These values will serve as the baseline of the simulations of stable data for the development of the proposed control chart methods. Before we proceed with these estimates, we must first have an understanding of control charts.

3 Control Charts

The most well-known method for distinguishing between stable and out of control counts is the control chart. We consider different approaches below. In particular, we develop control charts based on classification techniques in machine and statistical learning.

The first step in constructing a control chart is to choose a sample statistic Q to be monitored. Lower and upper control limits (LCL and UCL, respectively) for Q are determined from the stable-process distribution. An out-of-control (OOC) signal is given out if $Q < LCL$ and/or $Q > UCL$.

The number of samples till OOC is called the run length of a chart. The expected value of run length is called the average run length (ARL). The ARL is often used to assess and compare the performances of control charts. Control limits LCL and UCL are chosen so that the ARL is equal to a desired value L_0 under stable-process conditions. It is desired that if the process has shifted for the worse, then the ARL should be shorter than L_0.

3.1 Simultaneous X_1 and X_2 Charts

For a BNB process, one charting approach is to create individual X_1 and X_2 charts and implement them simultaneously. Each of these charts is referred to as a Shewhart-type control chart. Use the MLEs from a stable process to create upper and lower control limits for X_i

$$UCL_i = \mu_{X_i} + k\sigma_{X_i}, \quad LCL_i = \mu_{X_i} - k\sigma_{X_i}$$

for $i = 1, 2$ and where μ_{X_i} and σ_{X_i} are the marginal mean and variance of X_i given in Section 2.1 and k is an appropriate constant. Because we are monitoring disease counts, these LCLs are not useful, so we will not consider them further with these charts. Finding these two upper control limits will allow us to create the two control charts that we can use simultaneously.

The k value used in the upper control limit equations can be found based on a desired

stable-process ARL. For the Hepatitis C monitoring, counts are reported once a month. So we consider a stable-process ARL of 12 months. We feel that this is sufficient because we want to ensure shifts (unusual increases) in disease counts will be detected quickly and appropriate actions can be promptly taken by health officials.

Let $f_i(x_i)$ and $F_i(x_i)$ denote the marginal pmf and cumulative distribution function (cdf) of X_i for $i = 1, 2$, and let $F(x_1, x_2)$ be the joint cdf of (X_1, X_2). Thus,

$$\begin{aligned} F_i(x_i) &= \sum_{u=0}^{x_i} f_i(u|\lambda_i, \rho) \\ F(x_1, x_2) &= \sum_{u_1=0}^{x_1} \sum_{u_2=0}^{x_2} f(u_1, u_2|\boldsymbol{\theta}) \end{aligned}$$

for $i = 1, 2$. Suppose p is the probability of observing an OOC observation in either state. Then,

$$\begin{aligned} p &= P(X_1 > UCL_1 \cup X_2 > UCL_2) \\ &= 1 - P(X_1 \leq UCL_1 \cap X_2 \leq UCL_2) \\ &= 1 - F(UCL_1, UCL_2) \end{aligned}$$

by de Morgan's Law. We assessed that there is no evidence in the data of autocorrelation. Hence, we assume that Hepatitis C counts from different months are independent. In this case, the run length has a Geometric distribution, and the $ARL = 1/p$. Using this fact we can determine average run lengths (ARLs) for differing values of k as shown in Figure 2.

As the data are discrete, the ARL as a function of k is a step function. We will always choose the minimum k value that satisfies the $ARL \geq 12$. This may cause the ARL to be slightly shorter, but once again, in the situation of disease control, shorter ARL is better than having an issue of late detection. Since our response is also discrete, our UCLs will be

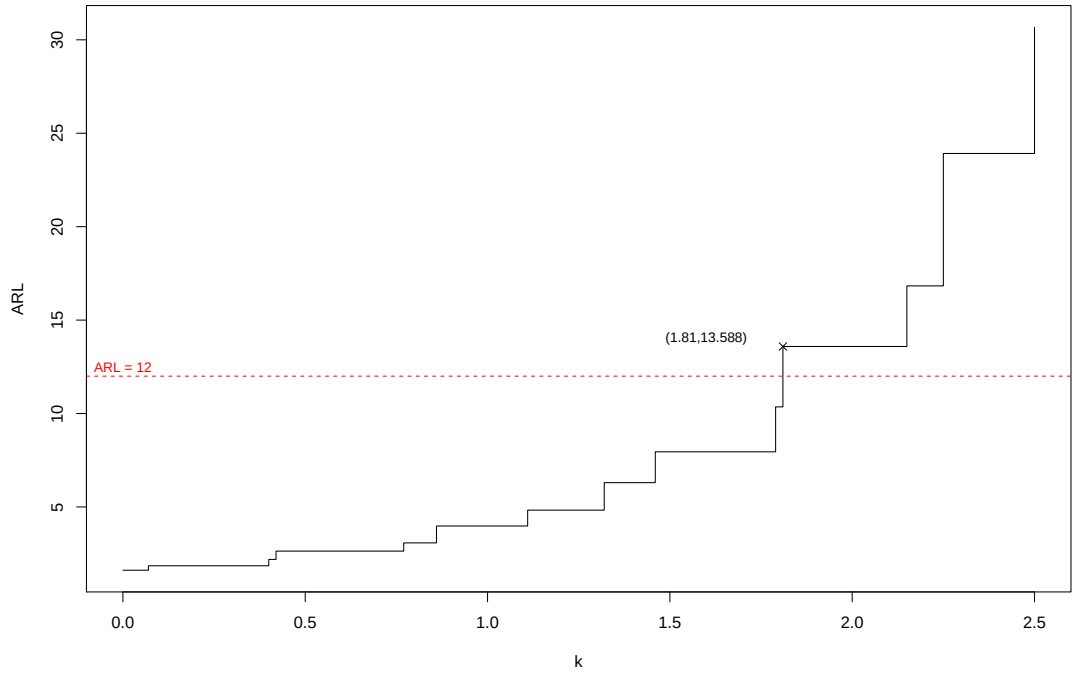

Figure 2: ARL chart for differing values of k

defined slightly differently using the floor function as follows

$$U_1 = \lfloor UCL_1 \rfloor, \quad U_2 = \lfloor UCL_2 \rfloor.$$

This also gives a more conservative ARL.

3.2 Constructing Charts and Finding MLEs

Now that we have established the idea behind X_1 and X_2 charts, we can use these in the iterative process detailed in Section 2.3 to find stable MLEs. After constructing simultaneous upper control limits and removing all OOC points, we found the following stable parameter estimates:

$$(\lambda_1, \lambda_2, \rho) = (0.4817503, 0.7339683, 6.5388904).$$

Using these as our stable parameter estimates, the following control charts can be gen-

erated. Figure 3 gives X_1 and X_2 charts containing our original 40 monthly counts. The UCLs are calculated based on the stable parameter estimates. Months 13 and 15 fell above the UCL for NSW and month 15 fell above for SA. These are considered to be the counts from our original data that are OOC.

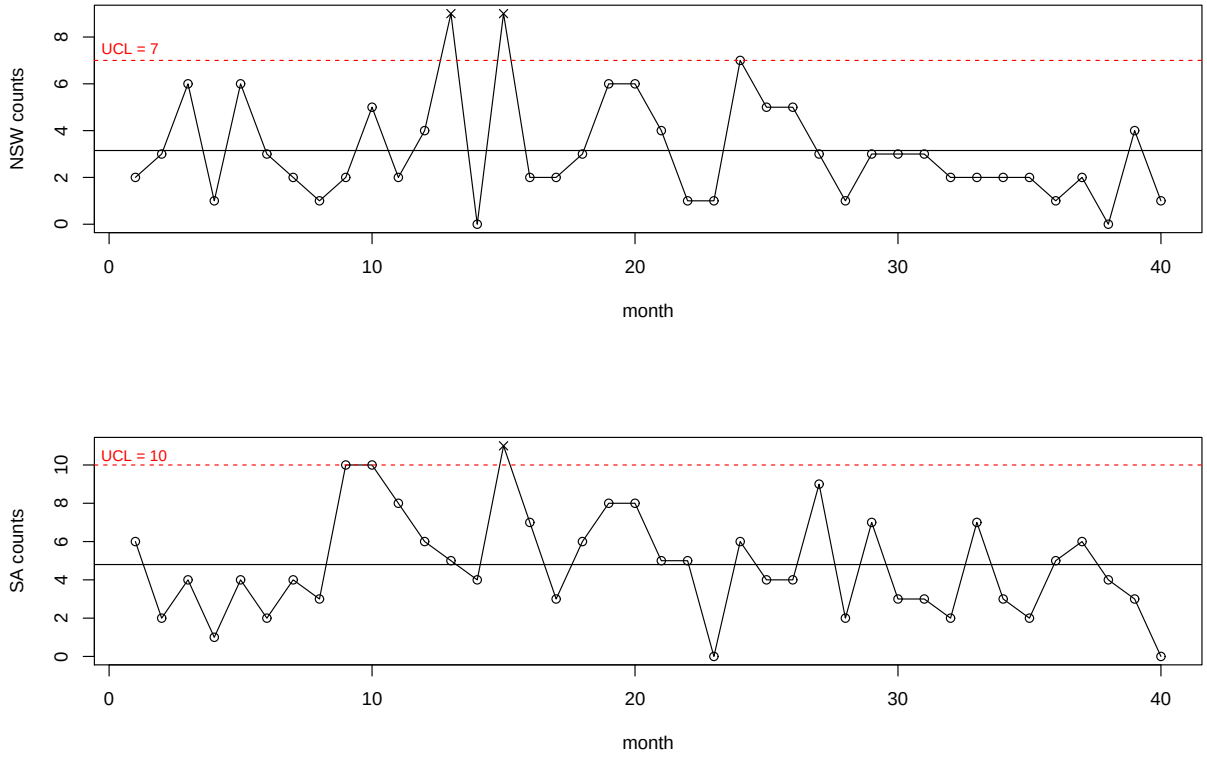

Figure 3: X_1, X_2 Charts for the Hepatitis A Data

These charts are what we will refer to as the more traditional method for simultaneously monitoring Hepatitis C counts. We could think of counts for a new month from both states as a 41st observation, plot accordingly, and signal OOC when necessary.

3.3 K-Chart for Support Vector Data Description

The first machine learning method we will examine is support vector data description (SVDD). For this method, we consider our stable data set a training sample. We will

use this sample to create a boundary that can be used to classify our data into two groups: Stable and OOC. This differs slightly from basic support vectors, as we will create a control boundary that adapts to our data rather than a hyperplane. Data is classified based on some measure of distance from this boundary. Here, we will use the kernel distance and accompanying K-chart described in [3]. This kernel distance for a standardized observation $\boldsymbol{z}$ is defined by

$$kd(\boldsymbol{z}) = \sqrt{\mathbf{K}(\boldsymbol{z},\boldsymbol{z}) - 2\sum_{i=1}^{l}\alpha_i\mathbf{K}(\boldsymbol{z},\boldsymbol{x}_i) + \sum_{i,j=1}^{l}\alpha_i\alpha_j\mathbf{K}(\boldsymbol{x}_i,\boldsymbol{x}_j)}$$

where there are l support vectors $\boldsymbol{x}_i$ with corresponding weight α_i, and

$$\mathbf{K}(x_i,x_j) = \exp\left[-(\boldsymbol{x}_i-\boldsymbol{x}_j)'\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_j)\right]$$

and Σ is the variance-covariance matrix of the process $\boldsymbol{X}_i$.

To implement this K-chart, we will need to create a UCL for these distances. We will do this in such a way that just as in the previous control charts, we have a stable ARL of $L_0 = 12$. We will simulate 200,000 kernel distances from the stable process and let $UCL = 1 - 1/12$ quantile of these distances.

Using the kernlab package in R, support vectors can be extracted. Figure 4 shows the support vector kernel-distance contours based on the stable sample. Kernel distances can be simulated for each of our original observations using these support vectors. Figure 5 shows the UCL calculated from the kernel distance simulations as well as the counts from our original data. Months 13 and 15 are considered OOC as they fall above the UCL. If we recall the results from the previous method, we will see that this K-chart has arrived at the same conclusion.

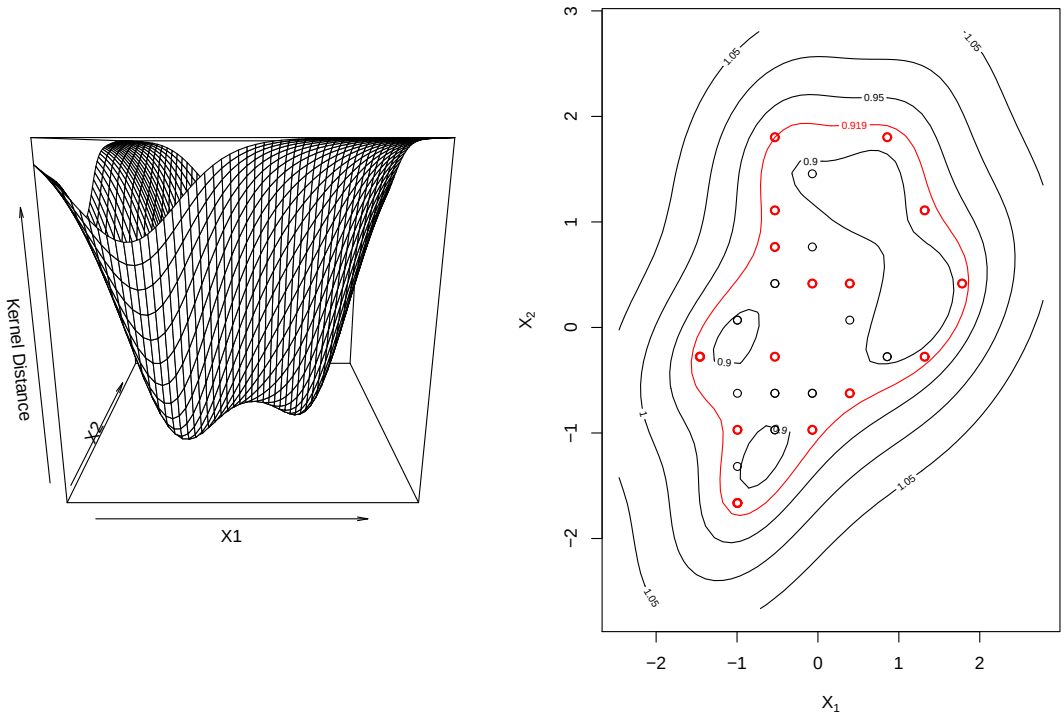

Figure 4: Kernel Distance Contour for the Hepatitis A Data

3.4 Log-Ratio Chart for Linear Discriminant Analysis

We will now turn away from the general idea of attempting to detect any shift that may cause an OOC signal to that of a specific shift. We will refer to such a deference from stability as a δ shift. Specifically, we will define a shift in the underlying parameters as follows.

$$\lambda_i^* = \delta_i \lambda_i \text{ for } i = 1, 2$$

For this work, we will not consider shifts in ρ which is directly related to the correlation.

This may not seem incredibly beneficial for disease control; however, knowledge about infection rate or other information about the disease could allow for quicker detection. We will use Linear Discriminant Analysis (LDA) as a method for detection of OOC counts.

LDA is similar to SVDD. Our goal is to classify observations as stable or OOC, but with

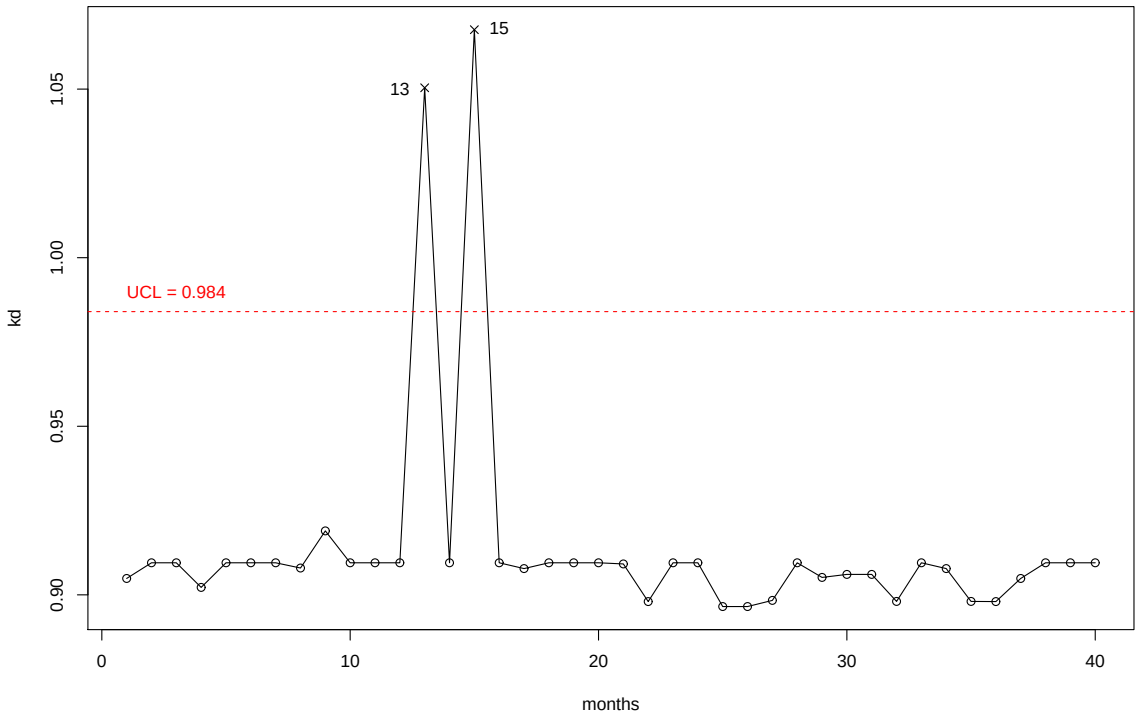

Figure 5: K-Chart for the Hepatitis A Data

LDA, we will look at a likelihood ratio statistic instead of a distance. Our assumption is that our data comes from two distributions with parameters $\boldsymbol{\theta}_0 = (\lambda_1, \lambda_2, \rho)$ (stable process) and $\boldsymbol{\theta}_1 = (\delta_1\lambda_1, \delta_2\lambda_2, \rho)$ (shifted process). Assuming some prior probabilty π_k for classes $k = 0, 1$ with $\sum_{k=0}^{1} \pi_k = 1$, an application of Bayes theorem tell us that

$$\begin{aligned} P_0 &= P(\text{Stable}|X = x) = \frac{\pi_0 f_0(x)}{\sum_{i=0}^{1} \pi_i f_i} \\ P_1 &= P(\text{OOC}|X = x) = \frac{\pi_1 f_1(x)}{\sum_{i=0}^{1} \pi_i f_i} \end{aligned}$$

where f_i for $i = 1, 2$ are the joint pmfs for the distributions with parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ respectively. For simplicity sake, we will only consider the case of $\pi_0 = \pi_1 = 1/2$. We will compare the two classes by looking at the log-ratio as described in Section 4.3 of [5] . We

see that

$$\begin{aligned} L(x) &= \log \frac{P_0}{P_1} \\ &= \log \frac{f_0(x)}{f_1(x)} + \log \frac{\pi_0}{\pi_1} \\ &= \log \frac{f_0(x)}{f_1(x)} \end{aligned}$$

It can be recognized that larger values for L will favor P_0 and are therefore more likely to be from a stable process distribution, whereas, smaller L values will favor P_1 and therefore favor an OOC signal.

Our analysis using LDA will differ from our previous two methods as we will be examining a LCL instead of a UCL. There is no strict reasoning for this as a slight alteration of the $L(x)$ calculation could result in a UCL. Again, we will pick this LCL to achieve an ARL $L_0 = 12$. This is done similarly as previously with $LCL = 1/12$ quantile of 200,000 simulated L values from the stable process $\boldsymbol{\theta}_0$.

This turns into a simple programming problem. After simulating for the LCL, LDA charts can be formulated for any given δ. We have used $\delta = \delta_1 = \delta_2 = 1.5$ and $\delta = \delta_1 = \delta_2 = 2.5$. Since we are only examining the case of equality of δ, we need only consider one of these cases. In Figure 6, the LDA chart for $\delta = 1.5$ is shown. We note that the chart for $\delta = 2.5$ is obtained form of $\delta = 1.5$ by adding an appropriate constant to all the values.

We see that just as in the previous charts, months 13 and 15 are OOC. But, using LDA, month 10 is also classified as OOC. This is no to say that any of our charts up to this point are necessarily incorrect. Rather, according to LDA, the count from month 10 is more likely from the distribution f_1 than f_0. If we return to Figure 3 for our X_1, X_2 charts, we see that for SA, the count for month 10 sat on the UCL. Based on this fact, it makes sense that the classification for this count could go either way.

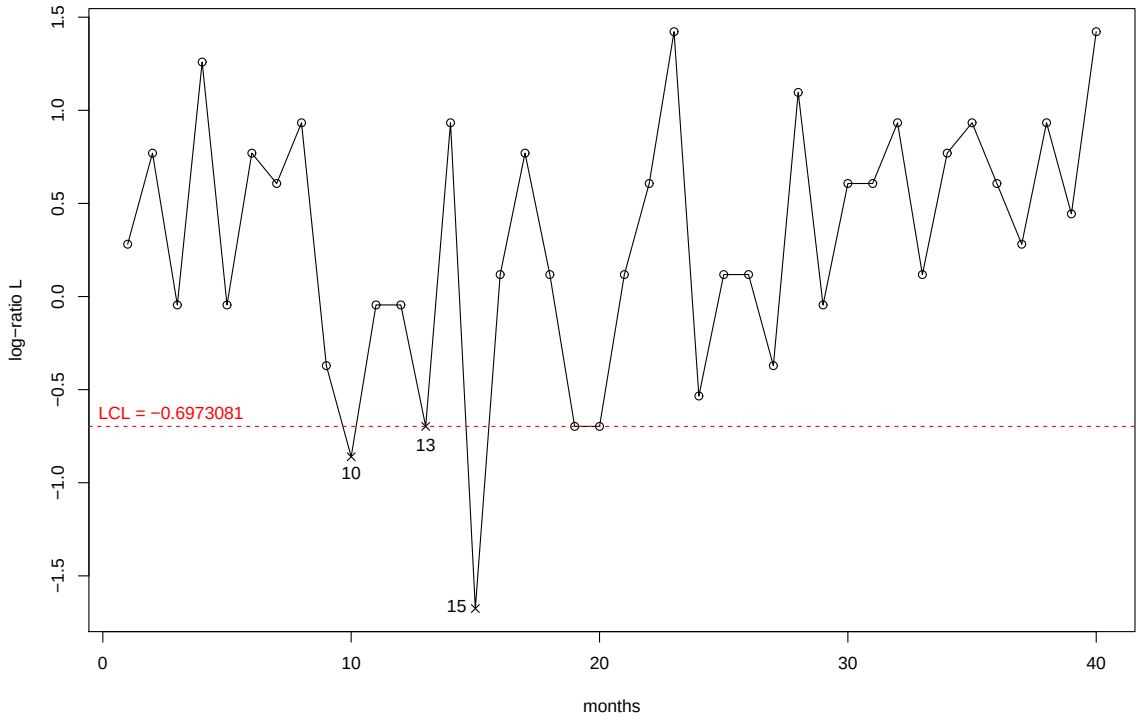

Figure 6: LDA Chart for $\delta = 1.5$

3.5 Logit Charts For Logistic Regression

We will also look at a slight alternation of LDA with Logistic Regression. We will use maximum likelihood estimation to find the estimates $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}$, and $\widehat{\beta_3}$ such that

$$\begin{aligned} L(x) &= \log \frac{P_0}{P_1} \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \end{aligned}$$

where β_0 is the intercept term, β_1 and β_2 are coefficients for the contributions from the two states individually, and β_3 is the coefficient of the interaction. Just as with LDA, we will be looking at specific δ shifts away from our stable process.

Charts will be created based on

$$P_0 = P(\text{Stable}|X = x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_1 x_2}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_1 x_2}}$$

We will analyze the distribution of P_0 and any significant decreases will imply that the probability of being OOC is greater than the probability of stable process and classification will be done accordingly. Our LCL will come from the 1/12 quantile of 200,000 simulated stable observations which will give us the desired $L_0 = 12$.

Similar to the SVDD method, we require a training sample to calculate $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2$, and $\widehat{\beta}_3$. We will generate this training sample using 15 observations generated from our stable process $\boldsymbol{\theta}_0 = (\lambda_1, \lambda_2, \rho)$ and 15 observations generated from our OOC process $\boldsymbol{\theta}_1 = (\delta_1\lambda_1, \delta_2\lambda_2, \rho)$. The benefit of Logistic over LDA would come in situations of burdensome $L(x)$ computations. We will expect this method to perform similar to LDA with slightly worse detection rates due to the approximation of $L(x)$.

Before we move on to the actually creation of these charts, we will quickly discuss the possibility of removing the interaction term and its accompanying coefficient (β_3). We will use the Akaike Information Criterion (AIC). AIC will estimate the amount of information lost in each model, consider the simplicity of the model with fewer predictors and choose accordingly. It is typical to choose the model with the smallest AIC. Using this criterion, we will soon discover that the inclusion of the interaction term in our models is not justified.

When $\delta = 1.5$, $AIC_{\beta_0,\beta_1,\beta_2,\beta_3} = 38.015$ and $AIC_{\beta_0,\beta_1,\beta_2} = 36.378$. When $\delta = 2.5$, $AIC_{\beta_0,\beta_1,\beta_2,\beta_3} = 42.811$ and $AIC_{\beta_0,\beta_1,\beta_2} = 40.813$. In the case of both values of δ, the interaction term does not significantly improve our models. For this reason, we will remove it and consider

$$P_0 = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2}} \tag{3}$$

When $\delta = 1.5$, maximum likelihood estimates based on glm in R are

$$\begin{aligned} \widehat{\beta} &= (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2) \\ &= (2.2306, -0.3801, -0.0491) \end{aligned}$$

These values of course will vary depending on the randomness of our training sample. However, using these values, we can simulate stable process observations and find the LCL for $\delta = 1.5$ using equation (3). The corresponding control chart is shown in figure 7.

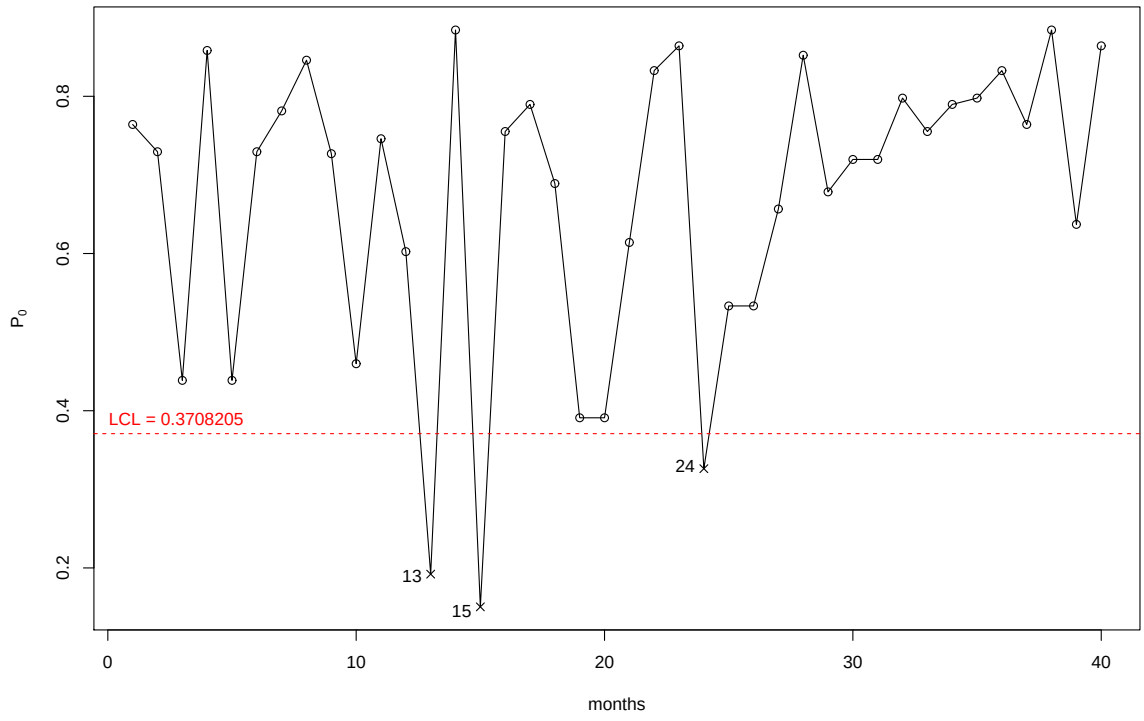

Figure 7: Logistic P_0 Chart for $\delta = 1.5$

As we would expect by now, from the original mixed data set, months 13 and 15 are OOC. Here month 24 is also OOC. This again is one of those counts that could go either way.

We can repeat this process for $\delta = 2.5$. The new control chart should be more sensitive

to these particular shifts. We will generate a new training set based on the new δ and recalculate the LCL. Here the MLE estimates are as follows.

$$\begin{aligned}\widehat{\beta} &= (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2) \\ &= (2.6463, -0.1535, -0.2250)\end{aligned}$$

Upon calculating the 1/12 quantile of the 200,000 simulated P_0 values, the chart for $\delta = 2.5$ is shown in figure 8. This chart did not classify the count for month 13 as OOC. This could be due to the training specific to larger shifts.

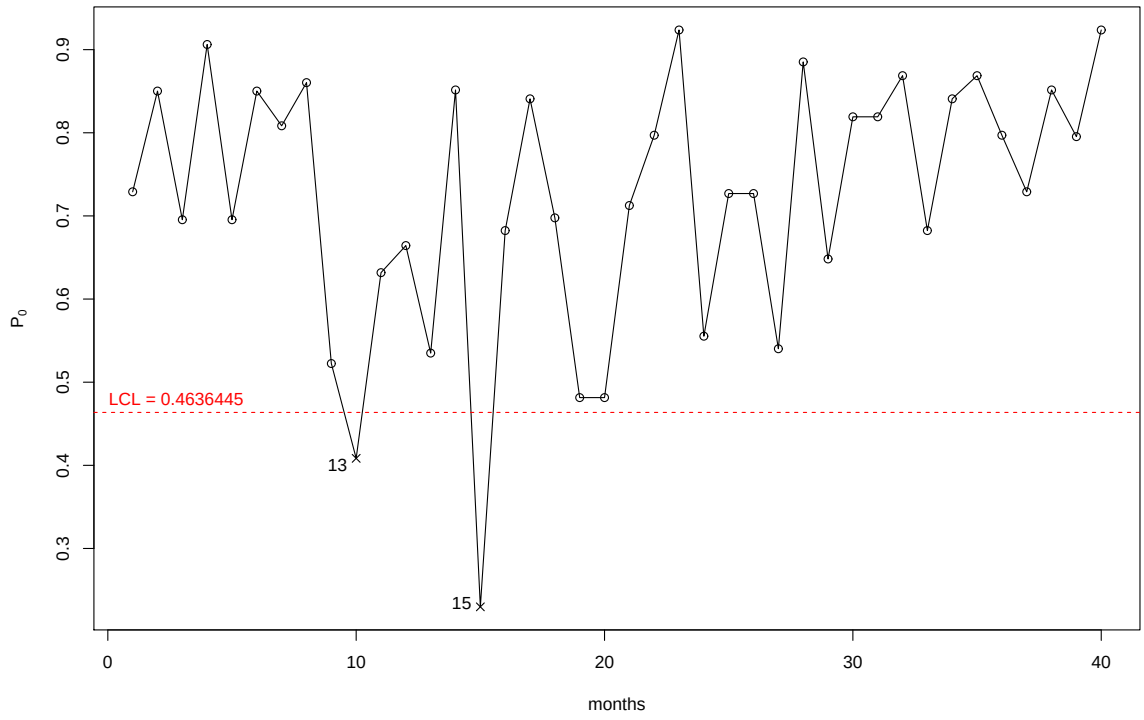

Figure 8: Logistic P_0 Chart for $\delta = 2.5$

4 ARL Study for Delta Shifts

The question that arises from the application of these three control chart method is obvious. Which will perform the best? Obviously, it is helpful to have a single control chart over simultaneous charts, especially in cases where we have observations from more than two variates. We will attempt to the answer this question based on an ARL study over differing shifts in $\delta = \delta_1 = \delta_2$. We will specifically look at $1 \leq \delta \leq 5$. For any shift $\delta \geq 5$, an immediate OOC warning is expected.

To find the ARLs for each method,

1. Fix some $1 \leq \delta \leq 5$.
2. Simulate monthly counts until an OOC count is detected and record the run length.
3. Repeat Step 2 until 2000 run lengths are recorded.
4. Take the average of these run lengths as the ARL for the given δ shift.
5. Repeat Step 1 for many δs.

For the simultaneous X_1, X_2 charts, the ARL is the true ARL found as $1/P(OOC)$. ARL comparisons are shown in Figure 9.

Clearly the methods we consider here are all quite competitive, but all except the logistic model with $\delta = 1.5$ have a slight improvement over the X_1, X_2 charts method. If we pair this advantage with that of only have to monitor one chart, we can understand their benefit.

5 Application to Simulated OOC Data

To see these models in action, we will simulate 8 OOC months from $\delta = 1.5$. We can think of this as a spike in disease happening in our 41-48th months of monitoring the disease counts. Since we know that these counts are in fact OOC, it is desirous that we see a signal immediately. We will only look at the X_1, X_2 charts, LDA chart, and SVDD chart. The

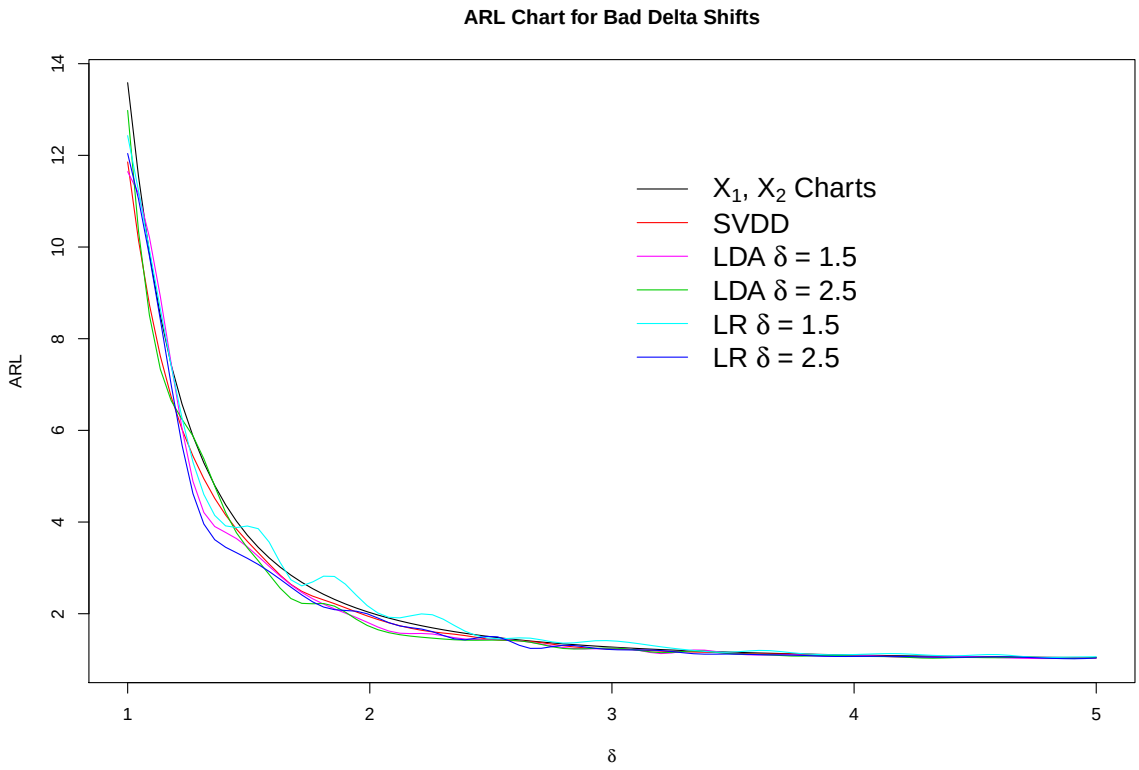

Figure 9: ARL Chart for δ Shifts

logistic regression method is not shown as it is just a small alteration of the LDA method and the approximations involved will create some variability in its ability to detect OOC counts. Figures 10-12 show these mentioned charts for the simulated data.

Based on these figures, we can conclude that it is difficult in most cases for these charts to detect shifts on this level of magnitude immediately. However, it is clear that the LDA chart was the quickest to detect in this case due to its signal at month 2 while the other charts do not signal until month 6. Of course detection would be much quicker if our data was generated with a much larger shift, but detection of smaller shifts quickly is ideal when in disease control process monitoring.

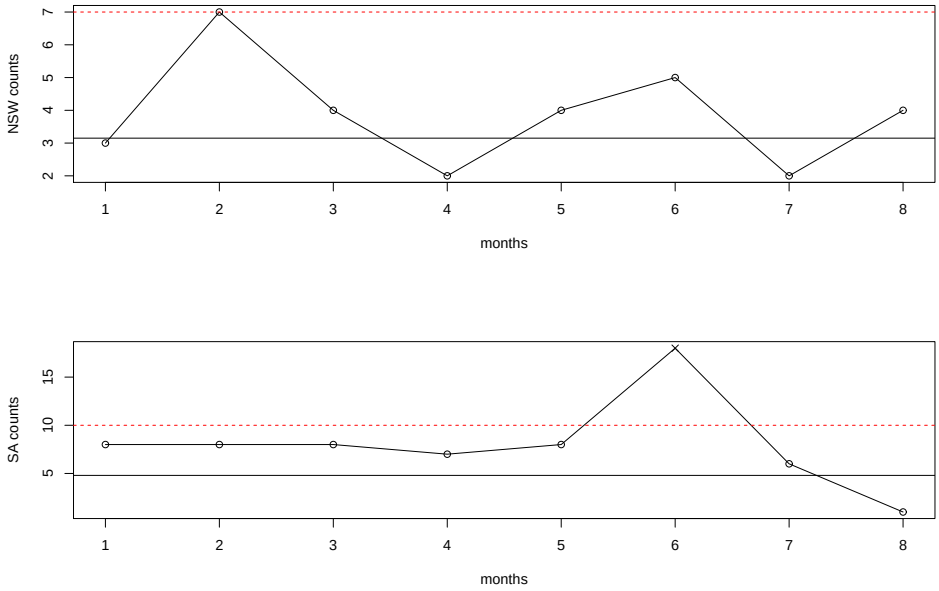

Figure 10: X_1, X_2 Charts for Simulated OOC Counts

6 Conclusion

Obviously the application of machine/statistical learning techniques will depend on the nature of the analysis being performed, but the LDA and Logistic regression charts seem to perform optimally in most cases. SVDD charts also seem to give some pretty stable results. The difficulty of computation in the LDA log-ratio could also be a reasonable situation for implementation of the Logistic approach. Although the ARL improvement over traditional control charting is not incredibly significant, the reduced complexity of analysis using a single chart cannot be understated. This is especially true in situations where the number of variables to be simultaneously monitored is larger than 2-3.

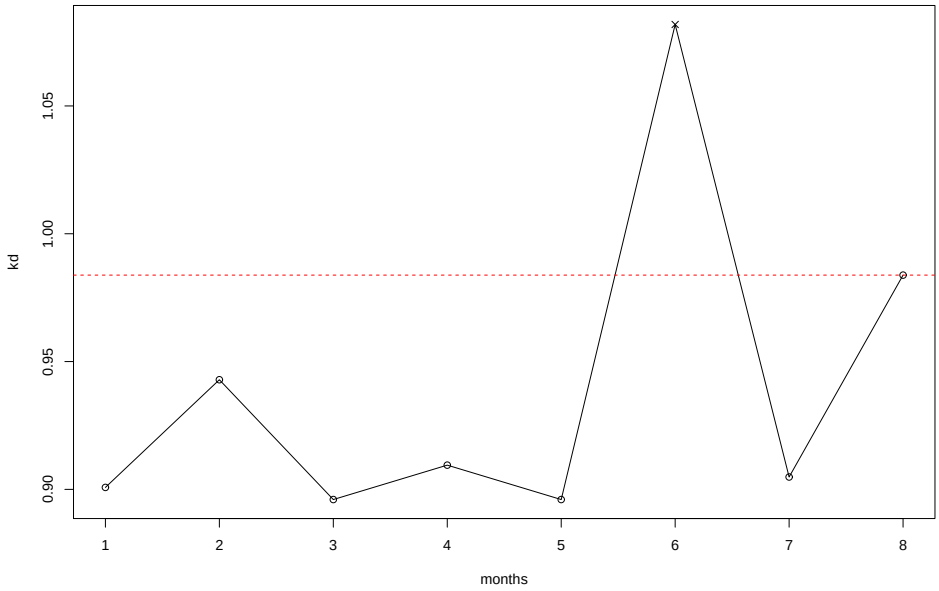

Figure 11: SVDD Chart for Simulated OOC Counts

References

[1] Arbous, A. G. and Kerrich, J. E. (1951). "Accident Statistics and the Concept of Accident Proneness". *Biometrics*, Vol. 7, No. 1951, pp. 340–342.

[2] Marshall, A. W. and Olkin, I. (1990). "Multivariate distributions generated from mixtures of convolution and product families". *Lecture Notes-Monograph Series*, Vol. 16, pp. 371–393.

[3] Sun, R. and Tsung, F. (2003). "A kernel-distance-based multivariate control chart using support vector methods". *Taylor & Francis*, Vol. 41, No. 13, pp. 2975–2989.

[4] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag: New York.

[5] Hastie, T. Tobshirani, R. and Friedman, J. P (2008). "The Elements of Statistical Learning Data Mining, Inference, and Prediction".

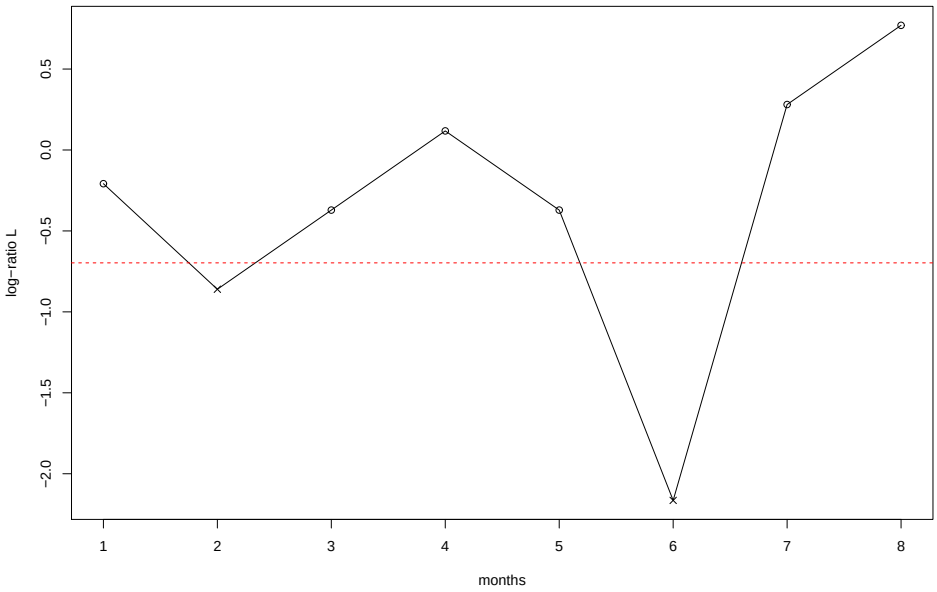

Figure 12: LDA Chart for Simulated OOC Counts