Monitoring Multivariate Correlated Negative Binomial with Machine Learning Methods

Guimiao Zhang

Department of Mathematics and Statistics, Washington State University, Pullman, WA

Abstract

In this study, the joint probability mass function of multivariate correlated negative binomial distribution has been derived along with the correlation matrix. The Shewhart method, support vector data descriptions, logistic regression, and Linear Discriminant Analysis have been adapted to develop control charts for monitoring multivariate correlated negative binomial processes. The average run length is used to compare the performances of techniques. In general, the K chart for support vector data descriptions requires longer run length to detect first out-of-control signal, whereas the Shewhart chart, LR for logistic regression, and LDA chart for linear discriminant analysis do not show significant differences. The control chart methods are applied to multivariate data on website hits.

Keywords: Average run length, control chart, linear discriminant analysis, logistic regression, multivariate negative binomial, Shewhart chart, support vector data descriptions.

Contents

1. Introduction

The number of website hits (clicks) is a measure of website traffic and refers to the number of times a visitor clicks on a webpage. Real-time trending topics often affects the website clicks. For instance, Covid-19 dramatically increases the clicks on the health-related domains, such as World Health Organization, Center for Disease Control and Prevention, and the Department of Health. To monitor the change in click patterns, the multivariate correlated negative binomial (NB) distribution is employed here to model multivariate discrete variables and account overdispersion. Two or more internet websites that cover similar subjects are influenced by trending topics simultaneously and thus using independent univariate NB distributions is insufficient for modeling. Here, a multivariate correlated NB (MVNB) is proposed as a modeling tool. Previous work on bivariate NB (BNB) derived a joint probability mass function based on Gamma and Poisson distribution assumptions. We derive a more general form of joint probability mass function for MVNB to describe 3 or more variables. For simplicity, trivariate NB (TNB) is demonstrated here. The modeling allows for a flexible correlation structure. The model parameters are estimated by maximum likelihood methods.

In a control chart for website clicks, if the number falls above or below prespecified control limits, an out of control (OOC) signal occurs suggesting a change in click patterns. Otherwise, no signal occurs which indicated that the process is stable. We adapt the Shewhart chart, a simple approach in processes monitoring, to the website hits application. With control charting, we combine several machine learning techniques such as support vector data description (SVDD), logistic regression (LR), and linear discriminant analysis (LDA). The chart performances are then compared using average run length (time to first OOC signal).

2. Statistical Models

Arbous and Kerrich (1951) expressed the probability mass function of a univariate NB as

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

where the parameters are $m, k > 0$. The mean and variance are $E(X) = m$ and $Var(X) = m(1 + \frac{m}{k})$, indicating that $E(X) < Var(X)$.

Marshall and Olkin (1990) derived a BNB model with two correlated marginally distributed univariate NB variables. They assumed that $\theta \sim$Gamma(shape $= \rho$, scale $= 1$) and $X_i|\theta \sim$Poisson$(\mu = \lambda_i\theta), i = 1, 2$. The joint probability mass function of X_1, X_2 is

$$f(x_1, x_2|\lambda_1, \lambda_2, \rho) = \frac{\Gamma(x_1+x_2+\rho)}{x_1!x_2!\Gamma(\rho)} \left(\frac{\lambda_1}{\lambda_1+\lambda_2+1}\right)^{x_1} \left(\frac{\lambda_2}{\lambda_1+\lambda_2+1}\right)^{x_2} \left(\frac{1}{\lambda_1+\lambda_2+1}\right)^{\rho}, x_1, x_2 = 0, 1, 2, \ldots.$$

It can be derived that $X_i \sim \text{NB}(k = \rho, m = \rho\lambda_i)$. Furthermore, $E(X_i) = \rho\lambda_i$ and $Var(X_i) = \rho\lambda_i(1 + \frac{\rho\lambda_i}{\rho}) = \rho\lambda_i(1 + \lambda_i)$.

Subrahmaniam and Subrahmaniam (1973) and Kopocinski (1999) discuss a MVNB model with focus on BNB.

Haussman, Hall and Griliches (1984) provided another approach in developing a model for BNB. They assumed that

$$\begin{cases} X_i|\lambda_i \sim \text{Poisson}(\lambda_i), & i = 1, 2 \\ \lambda_i \sim \text{Gamma}(\text{shape} = \gamma_i, \text{scale} = \delta) \\ \dfrac{\delta}{1+\delta} \sim \text{Beta}(a, b) \end{cases}.$$

Following this idea, Boucher, Denuit and Guillén (2008) studied time-dependent BNB. Their model for a TNB had 5 parameters.

Shi and Valdez (2014) discussed MVNB with copulas. No clear form of the joint probability mass function for this MVNB model when the number of variables is at least 3 has been developed.

Here, we follow the procedure of Marshall and Olkin (1990) and extend it to 3 or more variables. We assumed n variates $X_1, \dots, X_n$ each marginally distributed NB. Their joint probability mass function is derived as follows:

$$\begin{aligned} f(x_1, x_2, \dots, , x_n|\lambda_1, \lambda_2, , \lambda_n, \rho) &= \int_0^\infty \prod f_{X_i|\theta}(x_i|\theta) \cdot f_\theta(\theta|\rho)\, d\theta \\ &= \int_0^\infty \prod \frac{e^{-\lambda_i\theta}(\lambda_i\theta)^{x_i}}{x_i!} \cdot \frac{1}{\Gamma(\rho)} e^{-\theta}\theta^{\rho-1}\, d\theta \\ &= \frac{1}{\Gamma(\rho)} \prod \frac{{\lambda_i}^{x_i}}{x_i!} \int_0^\infty e^{-(\sum\lambda_i+1)\,\theta}\theta^{\sum x_i+\rho-1}\, d\theta \\ &= \frac{1}{\Gamma(\rho)} \prod \frac{{\lambda_i}^{x_i}}{x_i!}\, \Gamma(\sum x_i + \rho) \left(\frac{1}{\sum\lambda_i + 1}\right)^{\sum x_i+\rho} \\ &= \frac{\Gamma(\sum x_i + \rho)}{\Gamma(\rho)} \cdot \left(\frac{1}{\sum\lambda_i + 1}\right)^{\rho} \cdot \prod \left(\frac{\lambda_i}{\sum\lambda_i + 1}\right)^{x_i} \cdot \prod \frac{1}{x_i!} \end{aligned}$$

where $x_1, x_2, \dots, x_n = 0, 1, 2, \dots$. It is easy to see that Marshall and Olkin's (1990) BNB model is a special case.

The covariance and correlation between variable pairs are derived as follows:

$$Cov(X_i, X_j) = E(X_i \cdot X_j) - E(X_i)E(X_j) = E_\theta[E(X_i \cdot X_j|\theta)] - E_\theta[E(X_i|\theta)]E_\theta[E(X_j|\theta)]$$

Given θ, $X_i|\theta$ and $X_j|\theta$ are independent for $i \neq j$. Then:

$$E_\theta\left[E\left(X_iX_j|\theta\right)\right] = E_\theta\left[E(X_i|\theta)E\left(X_j|\theta\right)\right] = E_\theta\left(\lambda_i\theta\lambda_j\theta\right) = \lambda_i\lambda_j E_\theta(\theta^2) = \lambda_i\lambda_j(\rho+\rho^2),$$

$$E_\theta[E(X_i|\theta)]E_\theta[E(X_i|\theta)] = E_\theta(\lambda_i\theta)\cdot E_\theta\left(\lambda_j\theta\right) = \lambda_i E_\theta(\theta)\cdot\lambda_j E_\theta(\theta) = \lambda_i\lambda_j\rho^2.$$

Therefore, for $i \neq j$,

$$Cov(X_i, X_j) = \lambda_i\lambda_j(\rho+\rho^2) - \lambda_i\lambda_j\rho^2 = \rho\lambda_i\lambda_j,$$

$$Corr(X_i, X_j) = \frac{\sigma_{X_iX_j}}{\sigma_{X_i}\sigma_{X_j}} = \frac{\rho\lambda_i\lambda_j}{\sqrt{\lambda_i\rho(1+\lambda_i)}\cdot\sqrt{\lambda_j\rho(1+\lambda_j)}} = \sqrt{\frac{\lambda_i\lambda_j}{(1+\lambda_i)(1+\lambda_j)}}.$$

Hence, this provides some flexibility in modeling correlations between different pairs of variables from the MVNB.

Our website-hits data are collected from analytics.usa.gov over a period of 46 consecutive days. For simplicity, 3 websites are chosen for the application in this project ($n = 3$). In particular, the 3 chosen domains are from USPS, FDA and NASA. The joint probability mass function of TNB is

$$f(\boldsymbol{x}|\boldsymbol{\lambda},\rho) = \frac{\Gamma(x_1+x_2+x_3+\rho)}{\Gamma(\rho)\cdot x_1!\cdot x_2!\cdot x_3!}\cdot\frac{{\lambda_1}^{x_1}\cdot{\lambda_2}^{x_2}\cdot{\lambda_3}^{x_3}}{(\lambda_1+\lambda_2+\lambda_3+1)^{x_1+x_2+x_3+\rho}}, \qquad x_1, x_2, x_3 = 0, 1, 2, \ldots.$$

The joint cumulative distribution function of TNB is

$$F(\boldsymbol{x}|\boldsymbol{\lambda},\rho) = \sum_{all\ y_1\le x_1, y_2\le x_2, y_3\le x_3} f(\boldsymbol{y}|\boldsymbol{\lambda},\rho), \qquad x_1, x_2, x_3 = 0, 1, 2, \ldots.$$

To establish the TNB model for control charting, we need to obtain the estimates of four parameters, λ_1, λ_2, λ_3 and ρ. The maximum likelihood (ML) estimation method is adopted. The total likelihood is $L(\boldsymbol{\lambda},\rho) = \prod_{j=1}^{N} L_j(\boldsymbol{\lambda},\rho) = \prod_{j=1}^{N} f_j(\boldsymbol{x}|\boldsymbol{\lambda},\rho)$, where N is the number of observations from a random sample. No closed forms for the ML estimators are available. The nlminb function in R is used here to get ML estimates with the method of moment (MOM) estimates as the starting values. The MOM estimators are obtained using the marginal NB properties as follows:

$$\tilde{\lambda}_i = \frac{s_i^2}{\bar{x}_i} - 1, \text{ for } i = 1, 2, 3$$

$$\tilde{\rho} = \frac{1}{3}\sum_{i=1}^{3}\tilde{\rho}_i = \frac{1}{3}\sum_{i=1}^{3}\frac{\bar{x}_i}{\tilde{\lambda}_i}$$

where $\bar{x}_i$ and s_i^2 are the sample mean and sample variance for X_i observations. To avoid extremely large numbers in computations in R, the raw data on real-time clicks per

minute are divided by 5. The parameters for the stable process are denoted by $\boldsymbol{\theta_0} = (\rho, \lambda_1, \lambda_2, \lambda_3)$. For our specific data, $\boldsymbol{\theta_0} = (\rho = 6.66, \lambda_1 = 10.17, \lambda_2 = 9.69, \lambda_3 = 9.39)$ is obtained. The average clicks for X_1 is the largest among these 3 variates.

3. Control Charts

The control chart, or statistical process control chart, is a time series plot used to study how a process behaves over time. It can have a central line for the average value of the monitored statistic. To represent limits of stable-process values, the chart also has upper control limit (UCL) and a lower control limit (LCL). These lines could be determined from available historical data. Researchers can draw conclusions from a control chart about whether the process average or variation is consistent (stable) or otherwise (OOC) relative to stable-process properties. An assignable cause is determined whenever an OOC signal occurs.

The run length of a control chart is the number of sampling periods before the first OOC signal. It is a random variable that can only take on integer values. If samples are independent and identically distributed, the run length (RL) follows a geometric distribution with the parameter p, where p is the probability of OOC, i.e. $RL \sim \text{Geometric}(p),\ p = Pr(\text{OOC})$. To measure the performance, the expected value of RL, i.e., average run length (ARL) is often computed. Thus, $ARL = \frac{1}{p}$. The UCL and LCL can be chosen to attain a desired stable-process ARL which is specified by the practitioner. When the process is OOC, the best chart is the one with the smallest ARL.

3.1. Shewhart Chart

Shewhart control chart uses the mean μ_i and variance σ_i^2 of the X_i under stable process to determine the $UCL_i = \mu_i + k\sigma_i$ and $LCL_i = \mu_i - k\sigma_i$ where k is determined to satisfy a specified false-positive rate. If a single observation falls outside of (LCL_i, UCL_i), the control chart for X_i signals OOC. For the TNB process we consider here, observations come one vector (X_1, X_2, X_3) of counts at a time. We simultaneously use 3 Shewhart charts, i.e., 3 sets of (LCL, UCL), for each of the counts. If any one of X_1, X_2, X_3 falls outside its control limits, then the whole group is marked as OOC.

Because NB variables are integer-valued, the derivation using cumulative density function needs some adjustment. Let $LCL_i^* = \mu_i - k\sigma_i$. If it is an integer, let $LCL_i = LCL_i^* - 1$, otherwise, let $LCL_i = LCL_i^*$. The probability of OOC for one TNB observation is derived as follows.

If only UCL (upper-sided chart) is used:

$$Pr(\text{OOC}) = Pr\left(\bigcup_{i=1}^{3} X_i > UCL_i\right) = 1 - F(UCL_1, UCL_2, UCL_3)$$

If only LCL (lower-sided chart) is used :

$$Pr(\text{OOC}) = Pr\left(\bigcup_{i=1}^{3} X_i < LCL_i^*\right)$$
$$= \sum_{i=1}^{3} H(LCL_i) - G_{12}(LCL_1, LCL_2) - G_{23}(LCL_2, LCL_3)$$
$$- G_{13}(LCL_1, LCL_3) + F(LCL_1, LCL_2, LCL_3)$$

If both (2-sided chart) UCL and LCL are use:

$$Pr(\text{OOC}) = Pr\left(\bigcup_{i=1}^{3}(X_i < LCL_i^* \cup X_i > UCL_i)\right)$$
$$= 1 - F(UCL_1, UCL_2, UCL_3) + G_{12}(UCL_1, UCL_2) + G_{23}(UCL_2, UCL_3)$$
$$+ G_{13}(UCL_1, UCL_3) + F(LCL_1, LCL_2, LCL_3) + F(LCL_1, UCL_2, UCL_3)$$
$$+ F(UCL_1, LCL_2, UCL_3) + F(UCL_1, UCL_2, LCL_3) - F(UCL_1, LCL_2, LCL_3)$$
$$- F(LCL_1, UCL_2, LCL_3) - F(LCL_1, LCL_2, UCL_3) - \sum_{i=1}^{3} H(UCL_i)$$

Here, $H(x_i) = \sum_{all\ y_i \le x_i} h_i(y_i)$is the sum of the marginal cumulative distribution functions of the X_1, X_2, X_3, where $h_i(x_i|\lambda_i, \rho) = h_i(x_i)$ is the marginal probability mass function of X_i for $i = 1, 2, 3$. $G_{ij}(x_i, x_j) = \sum_{all\ y_i \le x_i, y_j \le x_j} g_{ij}(y_i, y_j)$ is the marginal joint cumulative distribution function of (X_i, X_j) (BNB), where $g_{ij}(x_i, x_j|\lambda_i, \lambda_j, \rho) = g_{ij}(x_i, x_j)$ is the marginal joint probability mass function of (X_i, X_j) for $i \neq j,\ i, j = 1, 2, 3.$ F is the joint cumulative distribution function of X_1, X_2, X_3.

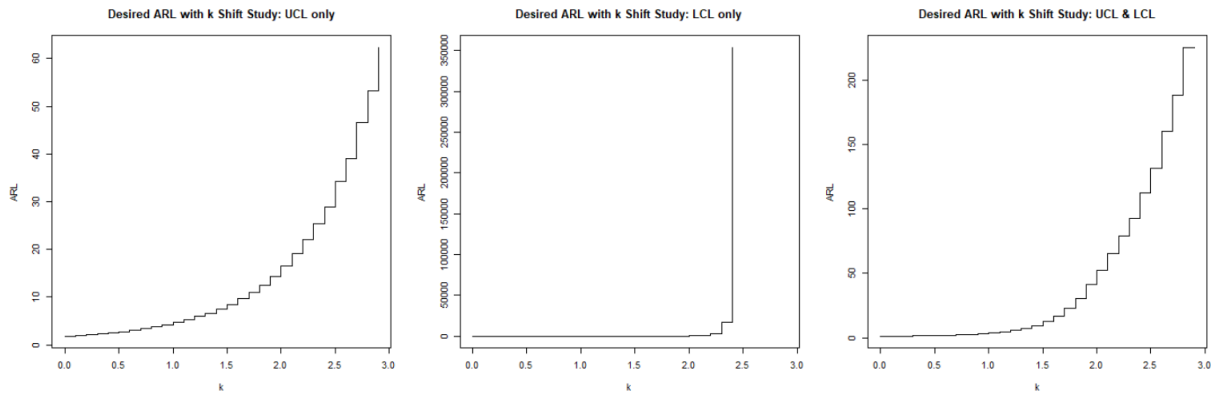

Figure 1. Desired ARL with k shift study.

Larger k shift lowers $Pr(\text{OOC})$ and thus attains larger ARL. It is consistent with Fig. 1. Classifications can be performed based on Shewhart method.

3.2. K Chart for Support Vector Data Descriptions

Sun and Tsung (2003) discussed a control chart based on SVDD. The Gaussian radial basis function $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left[-(\boldsymbol{x}_i - \boldsymbol{x}_j)'\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{x}_j)\right]$ was adopted as the kernel function, where Σ is the variance-covariance matrix of the random vector $\boldsymbol{X}$. The kernel distance (KD) for a standardized observation $\boldsymbol{Z}$ is defined as

$$KD = \sqrt{K(\boldsymbol{z},\boldsymbol{z}) - 2\sum_{i\in S}\alpha_i K(\boldsymbol{z},\boldsymbol{x}_i) + \sum_{i,j\in S, i\neq j}\alpha_i\alpha_j K(\boldsymbol{x}_i,\boldsymbol{x}_j)},$$

where S is the set of the support vectors, and α_i is the corresponding weight for $\boldsymbol{x}_i$. Each fixed value of kernel distance corresponds to an enclosed surface in high dimension. For 3 or more dimensions, that surface is difficult to visualize. Larger (smaller) KD values correspond to points outside (inside) this surface. Hence, for each new observation, an OOC signal occurs if $KD > UCL$. Hence, $Pr(\text{OOC}) = Pr\,(KD > UCL)$.

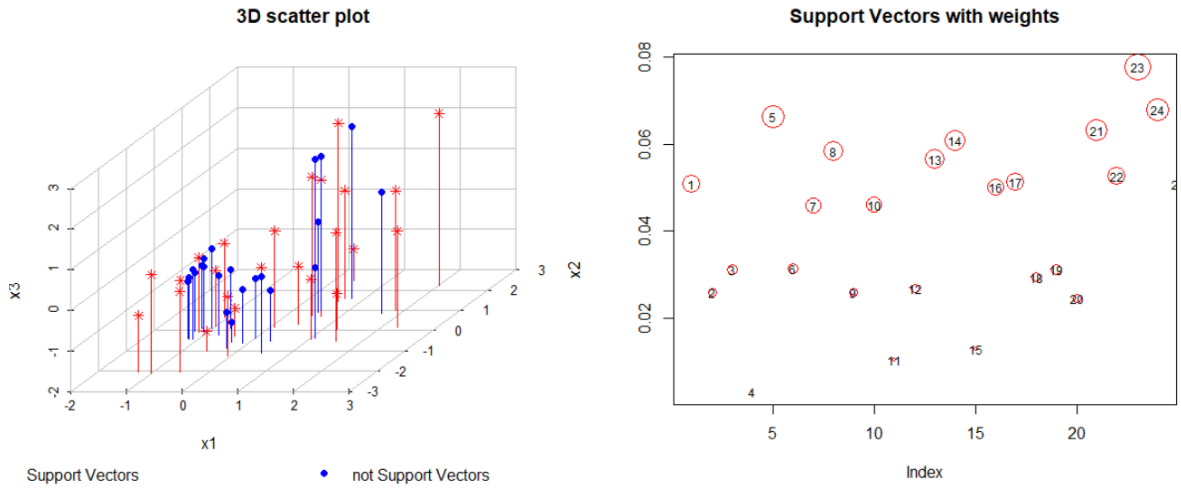

Figure 2. Support vector study for the website hits data. The sizes of red circles in the right plot indicate the relative weights.

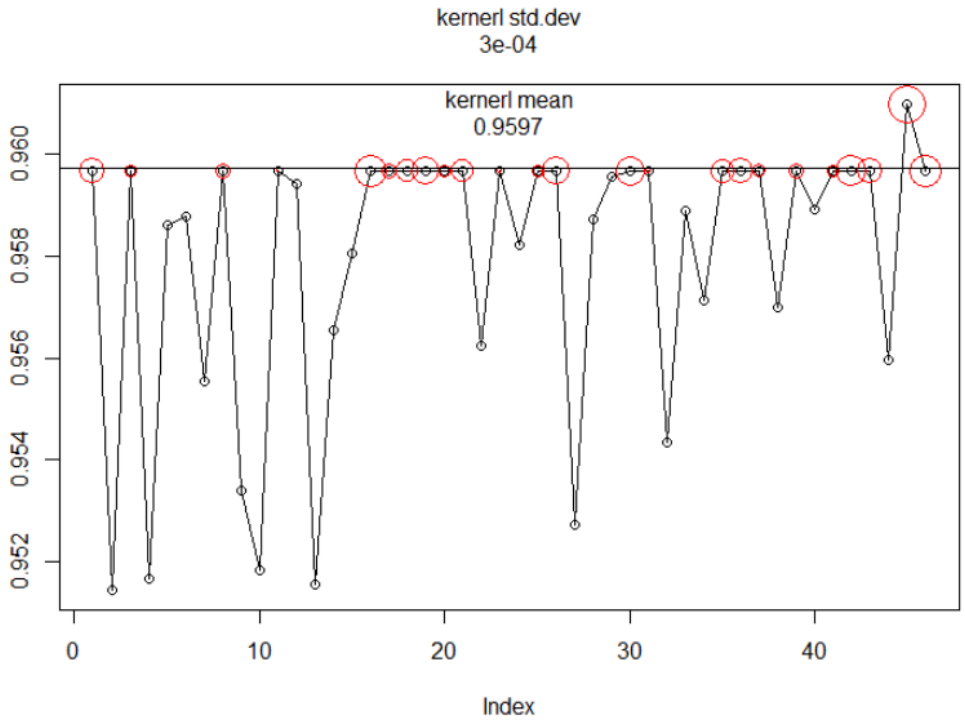

Figure 3. KDs for each observation of the collected data. Red circled are support vectors.

In this work, 24 out of 46 observations were chosen as the support vectors for the website hits data. See Figure 2. The kernel distance at the support vectors is maximum and achieves a value of 0.9597. See Figure 3. Based on the obtained support vectors and assumed stable-process parameters, KD values are obtained for two million simulated in-control data, and the UCL is chosen as the $1 - 1/ARL$ quantile. Refer to the chart as the K chart with OOC signal whenever $KD > UCL$.

3.3. LR Chart for Logistic Regression Models

As the response of our study is a binary variable (0 for in-control, 1 for OOC), logistic regression (LR) models can be established and used with control charts to detect OOC signals. The application of the model cannot be generalized in that the control chart setup depends on training samples of actual data that are known to be stable and OOC. The model can be set up as:

$$logit(P_1) = ln\left(\frac{P_1}{1-P_1}\right) = ln\left(\frac{P_1}{P_0}\right) = \boldsymbol{X\beta} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3,$$

where P_1 is the probability of OOC associated with this observation whereas P_0 is the probability of not OOC. The parameters $\boldsymbol{\beta}$ were estimated by simulation with 1000 stable data and 1000 OOC data as training sets. For simplicity, our initial models were

built on simulated OOC data by simultaneously shifting the TNB λ_i parameters by the same factor δ^*, i.e., the shifted TNB parameters are given by $\boldsymbol{\theta_1} = (\rho, \lambda_1\delta^*, \lambda_2\delta^*, \lambda_3\delta^*)$. Then the parameters estimates $\widehat{\boldsymbol{\beta}}$ are averaged through 100 replicates. Afterwards, 2 million in-control data points are simulated and P_1 for each is calculated. Larger values of P_1 suggest a shifted process. An OOC signal occurs when $P_1 > UCL$. UCL is chosen as the $1 - 1/ARL$ quantile where ARL is a prespecified desired stable-process average run length. Refer to this as a LR chart. Note that LR charts depend on δ^*, that is, these charts are designed to detect shifts described by δ^*. We study LR charts based on different δ^* values.

3.4. LDA Chart for Linear Discriminant Analysis

Hastie et. al (2009) discuss linear discriminant analysis (LDA) in detail. In their illustration, LDA is mainly applied to normal or approximately-normal data whereas we adapt the approach to TNB. LDA is essentially based on the ratio of the likelihood of under a specified OOC model to the likelihood under a specified in-control model. Assuming some prior probability π_c for classes $c = 0\,(\text{in control}), 1\,(\text{OOC})$, with $\sum_{i=0}^{1}\pi_i = 1$. According to Bayes theorem, we have the following conditional probabilities given data $\boldsymbol{X} = \boldsymbol{x}$:

$$\begin{cases} P_0 = Pr(\text{stable}|\boldsymbol{X} = \boldsymbol{x}) = \dfrac{\pi_0 f_0(\boldsymbol{x})}{\sum_{i=0}^{1}\pi_i f_i(\boldsymbol{x})} \\ P_1 = Pr(\text{OOC}|\boldsymbol{X} = \boldsymbol{x}) = \dfrac{\pi_1 f_1(\boldsymbol{x})}{\sum_{i=0}^{1}\pi_i f_i(\boldsymbol{x})} \end{cases},$$

where

$$\begin{cases} f_0(\boldsymbol{x}) = f_0(\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{\theta_0}) \\ f_1(\boldsymbol{x}) = f_1(\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{\theta_1}) \end{cases}.$$

The log-likelihood ratio, which we use to determine the classes, is defined as

$$l(\boldsymbol{x}) = ln\frac{P_1}{P_0} = ln\frac{\pi_1 f_1(\boldsymbol{x})}{\pi_0 f_0(\boldsymbol{x})} = ln\frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})} + ln\frac{\pi_1}{\pi_0}.$$

Larger log-likelihood ratio suggests a shifted process. Thus, a control chart based on it will signal OOC if $l(\boldsymbol{x}) > UCL$ for some UCL. If the prior distribution of π_c is assumed to be $\pi_0 = \pi_1 = 0.5$, the log-likelihood ratio becomes $l(\boldsymbol{x}) = ln\frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})}$. This study assumes $\pi_0 = \pi_1 = 0.5$ for demonstration. The approach requires a specified shifted process. Again, we consider shifting all the TNB λ_i by the same factor δ^*, i.e., $\boldsymbol{\theta_1} = (\rho, \lambda_1\delta^*, \lambda_2\delta^*, \lambda_3\delta^*)$. Like before, 2 million in-control data points are simulated, each time producing a value for $l(\boldsymbol{x})$. UCL is chosen as the $1 - 1/ARL$ quantile. If $l(\boldsymbol{x}) > UCL$, an OOC signal is given out. Refer to this as a LDA chart. Note that, like the

LR charts, LDA charts depend on δ^*, that is, these charts are designed to detect shifts described by δ^*. We study LDA charts based on different δ^* values.

4. Average Run Length Studies for Control Charts

ARL performance of each technique is studied under different OOC scenarios described by $\boldsymbol{\theta_1} = (\rho, \lambda_1\delta, \lambda_2\delta, \lambda_3\delta)$. This represents shifts in the process means. The table below describes the shifts δ^* needed for designing LR and LDA charts. Note that δ corresponds to the true state of process whereas δ^* describes the shifts the LR and LDA charts are designed to detect. Simultaneous Shewhart chart method and the K chart do not require specifying δ^*. We study one-sided charts with only UCL or only LCL. We use stable-process ARLs of approximately 3 days and 25-26 days.

δ^* Shift to detect	Small	Medium	Large
Increasing Means	1.05	1.25	2
Decreasing Means	0.9	0.7	0.5

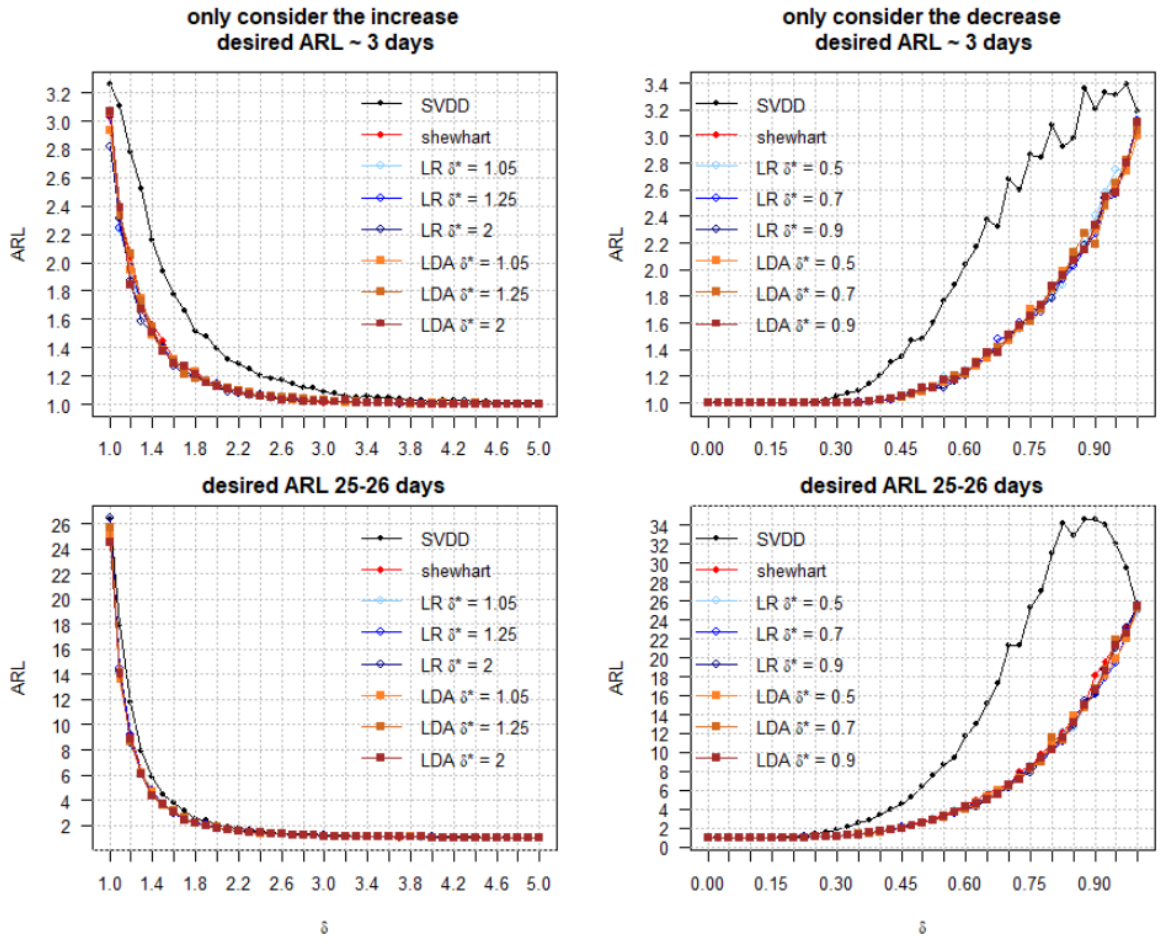

Figure 4. ARL study comparisons. This study focuses on the detection for simultaneous shifts on λ_i of all three variables by a factor of δ.

We compare chart performances when the true process is described by the parameters $\boldsymbol{\theta_1} = (\rho, \lambda_1\delta, \lambda_2\delta, \lambda_3\delta)$ for $1 \leq \delta \leq 5$. Compared with other methods, the K chart technique takes longer on average to signal OOC despite the desired in-control ARL. In addition, the K chart performs poorly if the shift coefficient δ is in the small to medium range when we decrease the process means. This chart even takes longer than the desired in-control ARL. The Shewhart, LR, and LDA chart approaches perform similarly in detecting shift. Although LR and LDA charts are based on different shift coefficients δ^*, their ARLs do not differ much.

5. Applications

Although simultaneous shifts on $\boldsymbol{\lambda}$ may happen, a shift on one out of three variates may occur more often. Here, we attempt to compare the detection performances for a single λ_i shift. Twenty OOC data points with parameters $\boldsymbol{\theta_1} = (\rho, \lambda_1\delta, \lambda_2, \lambda_3)$ are generated. Though the shifts may occur for other variates, we chose X_1 for illustration purposes. We take the first 5 points from the original data that we collected and assumed to be from a stable process. We the add the new OOC points to these 5 points. See table below. Set $\delta = 1.2$ and desired in-control $ARL = 3$. For this specific setup, the first OOC signal is expected at the 6th data point (i.e., almost as soon as when the shift occurs).

Data Index	Data
1 – 5	Collected Stable Data
6 – 25	Simulated OOC Data

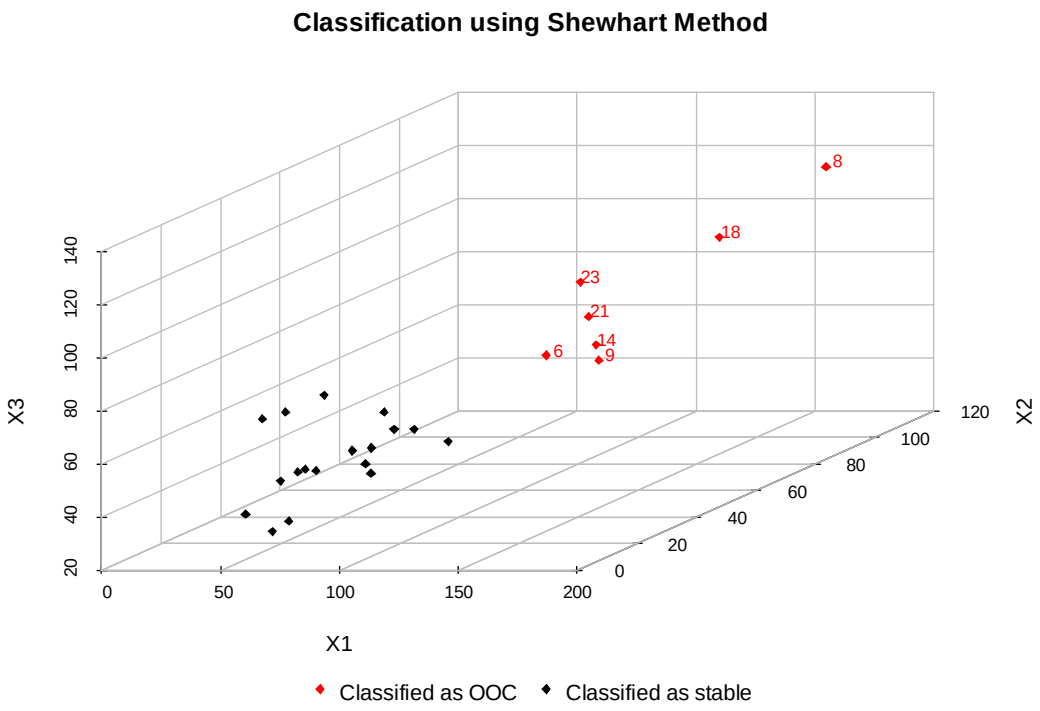

Figure 5. OOC points for the website hits data under the Shewhart chart approach. The Shewhart-based charts signal OOC the first time in period 6.

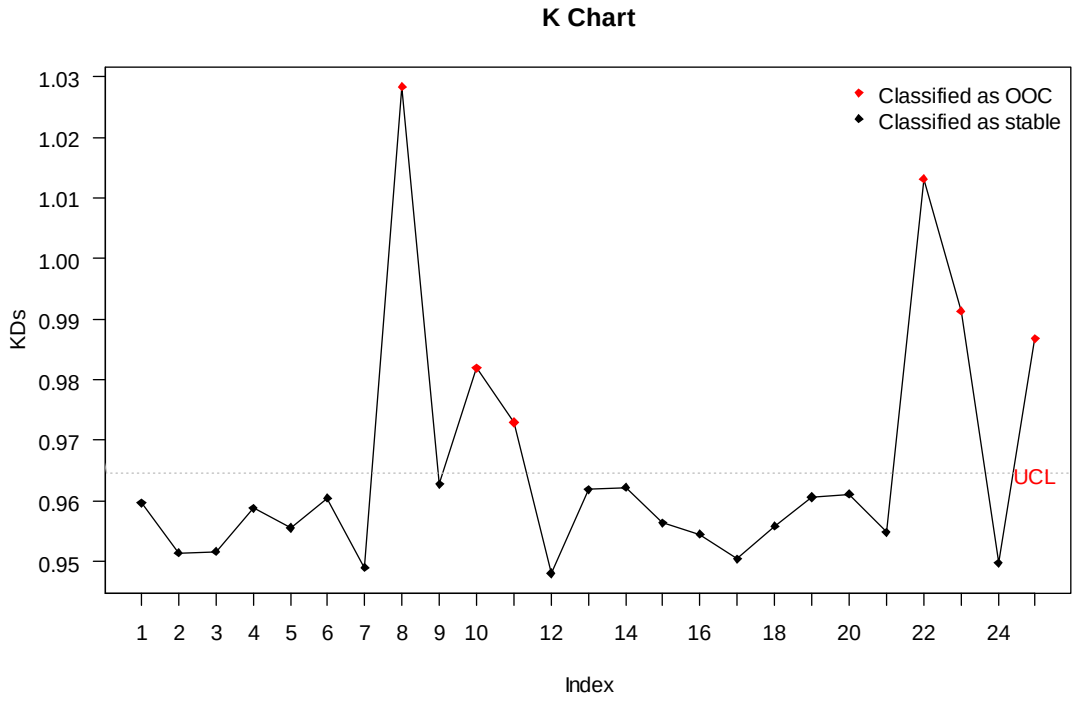

Figure 6. K chart for the website hits data. First OOC signal occurs at the 8th observation (3rd OOC point).

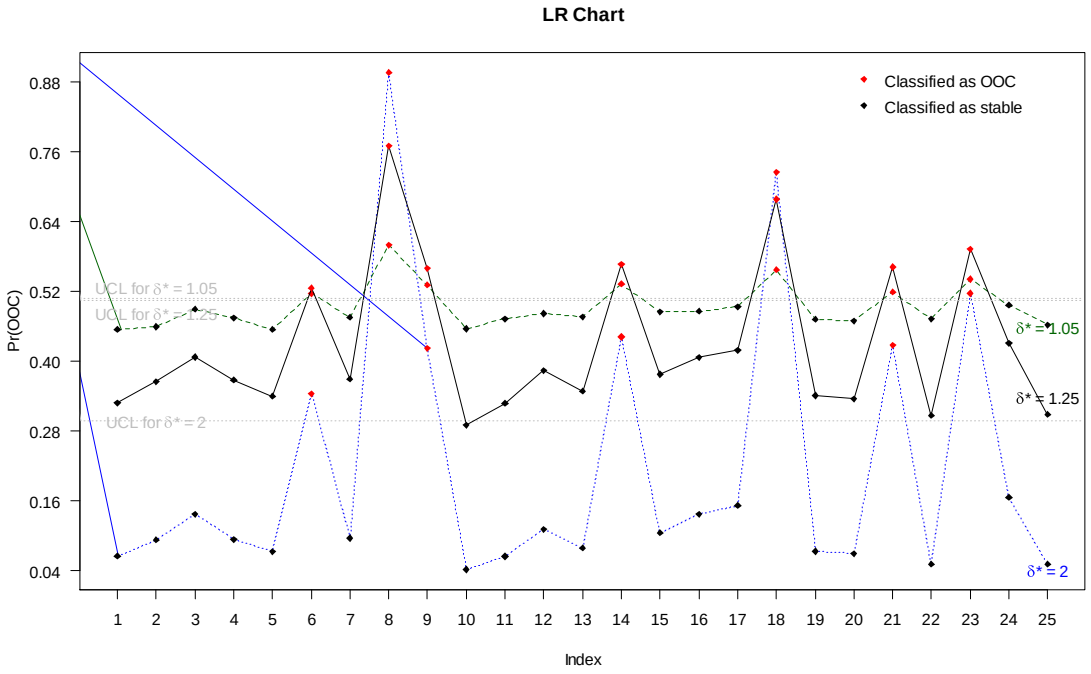

Figure 7. LR chart based on different shifts for the website hits data. LR charts based on different shift coefficients δ* models all signal on the first OOC data in the 6th period. The classification trends for the charts are similar.

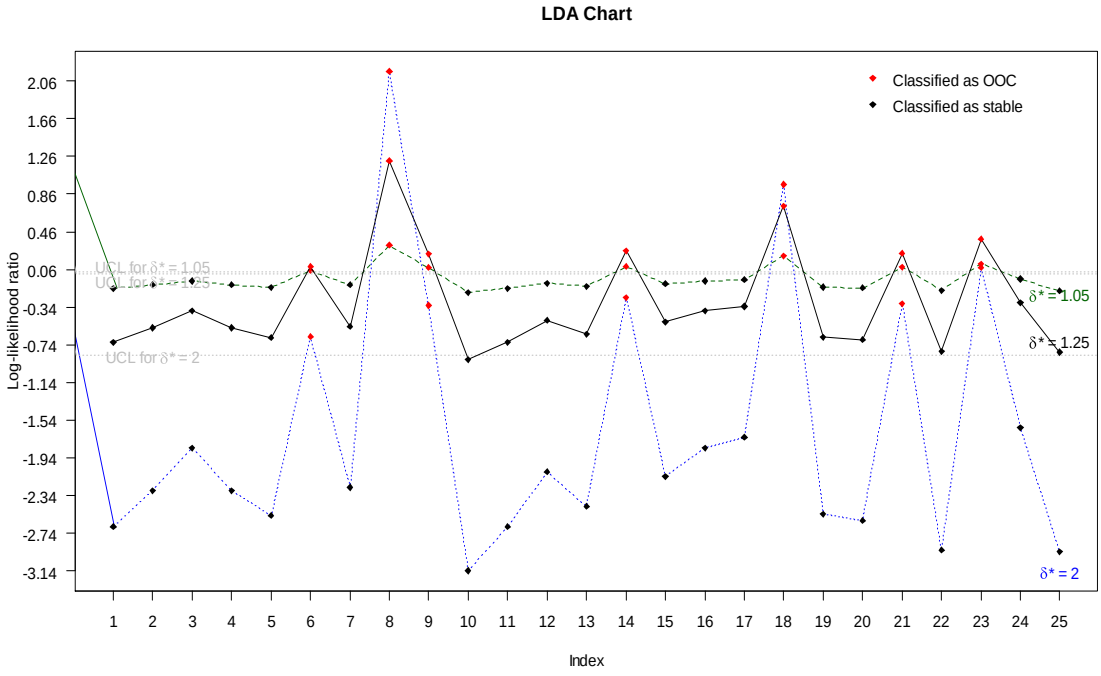

Figure 8. LDA chart for the website hits data. LDA charts based on different shift coefficients δ* all signal in period 6 when the first OOC data is obtained. The classification trends for the charts are similar.

For the application here, the K chart took 2 periods longer to signal OOC.

6. Conclusion

To detect shifts in the TNB website click process considered here, the K chart takes longer than the other charts considered here. Although the LR and LDA charts require shift specification, their performances do not seem to depend on nature of the shift.

7. References

Arbous, A.G. and Kerrich, J.E., 1951. Accident statistics and the concept of accident-proneness. *Biometrics*, *7*(4), pp.340-432.

Boucher, J.P., Denuit, M. and Guillen, M., 2008. Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. *Variance*, *2*(1), pp.135-162.

Hastie, T., Tibshirani, R. and Friedman, J., 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hausman, J.A., Hall, B.H. and Griliches, Z., 1984. Econometric models for count data with an application to the patents-R&D relationship.

Marshall, A.W. and Olkin, I., 1990. Multivariate distributions generated from mixtures of convolution and product families. *Lecture Notes-Monograph Series*, pp.371-393.

Kopociński, B., 1999. Multivariate negative binomial distributions generated by multivariate exponential distributions. *Applicationes mathematicae*, *25*(4), pp.463-472.

Shi, P. and Valdez, E.A., 2014. Multivariate negative binomial models for insurance claim counts. *Insurance: Mathematics and Economics*, *55*, pp.18-29.

Subrahmaniam, K. and Subrahmaniam, K., 1973. On the estimation of the parameters in the bivariate negative binomial distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, *35*(1), pp.131-146.

Sun, R. and Tsung, F., 2003. A kernel-distance-based multivariate control chart using support vector methods. *International Journal of Production Research*, *41*(13), pp.2975-2989.