An Exponentially Weighted Moving Average Control Chart for Zero-Truncated Poisson Processes: A Design and Analytic Framework with Fast Initial Response Feature

Robert Neil F. Leong
Frumencio F. Co
Vio Jianu C. Mojica
Mathematics Department, De La Salle University

Daniel Stanley Y. Tan
Software Technology Department, De La Salle University

Inspired by the capability of exponentially-weighted moving average (EWMA) charts to balance sensitivity and false alarm rates, we propose one for zero-truncated Poisson processes. We present a systematic design and analytic framework for implementation. Further, we add a fast initial response (FIR) feature which ideally increases sensitivity without compromising false alarm rates. The proposed charts (basic and with FIR feature) were evaluated based on both in-control average run length (ARL_0) to measure false alarm rate and out-of-control average run length (ARL_1) to measure sensitivity to detect unwanted shifts. The evaluation process used a Markov chain intensive simulation study at different settings for different weighting parameters (ω). Empirical results suggest that for both scenarios, the basic chart had: (1) exponentially increasing ARLs as a function of the chart threshold *L*; and (2) ARLs were longer for smaller ωs. Moreover, the added FIR feature has indeed improved ARL_1 within the range of *5% - 55%*, resulting to quicker shift detections at a relatively minimal loss in ARL_0. These results were also compared to Shewhart and CUSUM control charts at similar settings, and it was observed that the EWMA charts generally performed better by striking a balance between higher ARL_0 and lower ARL_1. These advantages of the EWMA charts were more pronounced when larger shifts in the parameter λ happened. Finally, a case application in monitoring hospital surgical out-of-controls is presented to demonstrate its usability in a real-world setting.

Keywords: exponentially-weighted moving average control chart, zero-truncated Poisson process, fast initial response feature, average run length, infection control

1. Introduction

Consider a process X_t indexed by some discrete sequence-defining set T. Suppose that X_t can be in any of the two states: in-control (wherein X_t follows a certain distribution F_0) and out-of-control (wherein X_t follows a certain distribution F_1 with shifted parameters). If X_t consists of counts of adverse events observed at time t (e.g., number of occurrences of a certain illness, number of non-conforming items produced by a certain manufacturing process, etc.), then the process is considered to follow some variant of the Poisson distribution. Control charting schemes for monitoring both the mean and the variance levels of the basic Poisson distribution have been widely studied in literature. Some comprehensive discussions regarding this are provided by Montgomery (2012), Borror, Champ, and Rigdon (1998), and Lucas (1985). A more recent point of interest is wherein a zero-inflated Poisson (ZIP) distribution is considered to possibly account for the overdispersion of the count distribution due to excess zeroes observed. Likewise, a rich set of literature on control charting schemes for ZIP processes can be obtained. Some of the better reads are provided by Xie, He, and Goh (2001), Katemee and Mayureesawan (2012), and Fatahi et al. (2012), which only considered a single-chart perspective, and by He, Huang, and Woodall (2011) and He, Li, and He (2014), which considered a simultaneous two-chart perspective. However, a less studied variant of the Poisson distribution in the context of control charting scheme is the zero-truncated Poisson (ZTP) distribution. A ZTP distribution arises when count data collection is done in such a way that '0' observations are discarded. That is, ZTP processes are observed whenever the realizations X_t can only take on positive integer values.

The first attempts to develop a control chart for a ZTP process were done by Chakraborty and Kakoty (1987) and Chakraborty and Singh (1990), wherein the former developed a cumulative sum (CUSUM) scheme while the latter a Shewhart scheme. More recently, Balamurali and Kalyanasundaram (2013) provided a more comprehensive discussion of the design and implementation of a CUSUM scheme for a truncated Poisson distribution. The control charting scheme can be used to detect both increases and decreases in Poisson rates. Moreover, the out-of-control average run lengths (ARLs) of such scheme were improved by implementing a fast-initial response (FIR) feature by setting the initial monitoring statistic at half the upper threshold value (Lucas and Crossier, 1982). Also, He, Li and He (2014) discussed the development of a standard ZTP-CUSUM chart in the context of monitoring positive counts for detecting increased shifts in a ZIP process.

While both Shewhart and CUSUM charts have advantages of their own (e.g. simplicity for the former and minimax optimality (Moustakides, 1986) for the latter), exponentially-weighted moving average (EWMA) charts having desirable properties of their own may be implemented. Several literature have found that EWMA charts are highly sensitive against small departures from the mean (Morton

et al., 2001; Woodall, 2006). While they perform inferiorly against CUSUM charts, the difference is not substantial in the sense that comparable performances can be achieved by using appropriately selected weighting parameters (Montgomery, 2012). Moreover, EWMA charts were found to perform as good as standard Shewhart charts with extended Westgard multirules in detecting inaccuracy and imprecision for small shifts (Neubauer, 1997). Another desirable characteristic of EWMA charts is their robustness against non-normal data (Borror, Montgomery, and Runger, 1999).

Inspired by the aforementioned desirable properties of EWMA charts, we develop in this paper an EWMA chart for monitoring a ZTP process. Specifically, design schemes for implementing such a chart will be presented and in-control and out-of-control ARL properties for different weighting parameter values (κ) and threshold values (h) will be discussed. In this regard, a brief discussion of the ZTP distribution is provided in Section 2. An EWMA chart for a ZTP process is developed in Section 3. ARL calculations via a Markov chain approach for different chart and process settings are discussed in Section 4. Introduction and evaluation of certain FIR features are discussed in Section 5. Comparisons of the proposed charts, both with and without the FIR feature, to standard Shewhart and CUSUM charts are presented in Section 6. To demonstrate its applicability in real-life scenarios, an actual case application in infection control is presented in Section 7. Then in Section 8, we give our concluding remarks.

2. Zero-Truncated Poisson Distribution

The ZTP distribution is a variant of the Poisson distribution which is truncated at zero, i.e., no '0' observations are made and its mass points consist only of positive integers. The distribution is characterized by a mean intensity parameter λ and has a probability mass function given as follows:

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!\left(1-e^{-\lambda}\right)}, x = 1,2,3,...$$

Some of the moments of the ZTP distribution are determined as follows:

$$E(X) = \frac{\lambda}{1-e^{-\lambda}}, \tag{1}$$

$$Var(X) = \frac{\lambda}{1-e^{-\lambda}}\left(1-\frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\right). \tag{2}$$

For m sampling subgroups, it can be shown that the method-of-moments and the maximum likelihood estimators of λ will be the same ($\hat{\lambda}$), which is the solution to the non-linear equation

$$\frac{\hat{\lambda}}{1-e^{-\hat{\lambda}}} = \frac{\sum x_i}{m} = \bar{x}_m, \tag{3}$$

where $\bar{x}_m$ is the sample mean of the m sampling subgroups.

An example of such process in healthcare is when one considers monitoring the number of days an appendectomy patient stays in the hospital after surgery. A typical patient and surgical procedure would only require 1.8-5.2 days of healing, with longer days typically attributed to infection. Infection, in such cases, may then be attributable to either malpractice or unsterilized surgical tools (Barrett, Hines, and Andrews, 2013). In monitoring such process, it is of particular interest to determine whether "surgical out-of-controls" are happening as observed through lengthened hospitalization of appendectomy patients. Also, the need to develop appropriate control charts for ZTP processes is if it is to be used alongside some other schemes to simultaneously and efficiently monitor ZIP processes, particularly the intensity of the process to produce non-conforming items (He, Li, and He, 2014). For more examples of the application of ZTP processes, refer to Plackett (1953) and Cohen (1972).

3. The Design of the ZTP-EWMA Control Chart

As mentioned by Fatahi et al. (2012), an EWMA control chart has high sensitivity to monitor high quality processes. As with any other control chart, it is a graphical display used for monitoring in-control and out-of-control process situations. Theoretically, it is comprised of three horizontal rules: the center line (CL: representing the mean level of the in-control process), the upper control limit (UCL), and the lower control limit (LCL) (representing the control limits of the chart wherein if the monitored statistics goes above the UCL or below the LCL, then the process is said to be in an out-of-control state). All lines are determined during Phase I when the process is assumed to be working under the in-control state.

The ZTP-EWMA control chart specifically monitors the statistic Y_t derived from a sequence $\{X_t\}$ of independent and identically-distributed ZTP random variables given as follows:

$$Y_t = (1 - \omega)\ Y_{t-1} + \omega X_t, \tag{4}$$

where $\omega \in (0,1)$ is the weighting parameter. By setting $Y_0 = E(X_0) = \frac{\lambda}{1-e^{-\lambda}}$, then following the formulation of Roberts (1959) based on asymptotic considerations (e.g., the process has been running for a very long time), the center line and the control limits may be computed as:

$$\begin{cases} UCL = E(X_t) + L\sqrt{\dfrac{\omega}{2-\omega}Var(X_t)} \\ CL = E(X_t) \\ LCL = E(X_t) - L\sqrt{\dfrac{\omega}{2-\omega}Var(X_t)} \end{cases} \tag{5}$$

where $E(X_t)$ can be taken as the expression in (1) and $Var(X_t)$ can be taken as the expression in (2). Note that whenever the computed LCL is less than zero, then it is adjusted to make it equivalent to zero. The parameter L is computationally determined such that desired in-control average run length (ARL_0) and out-of-control average run length (ARL_1) are achieved.

In the implementation of control charting procedures, phases I and II are distinguishably identified. In particular, Phase I involved the computation of reliable control limits based on a retrospective approach. In this case, observed in-control data are used to estimate the parameter λ and to empirically identify L. Then, Phase II is the prospective monitoring part wherein the developed chart in Phase I is now implemented. Following the work of Fatahi et al. (2012), we also present the following procedure to distinguish between the construction (Phase I) and the implementation (Phase II) of our proposed ZTP-EWMA control chart with real data.

Phase I:

Step 1: Observe m sampling subgroups and record the observed values X_t in each subgroup ($t = 1,2,\ldots,m$).

Step 2: Use (3) to estimate the ZTP distribution parameter λ.

Step 3: Using the observed data and the estimated $\hat{\lambda}$, check whether the collection of X_t's actually follow a ZTP distribution. Some tests which may be used are summarized in Best et al. (2007). In the case that evidences point toward the rejection of the ZTP distribution hypothesis, then increase the collected subgroups m and start again with Step 1.

Step 4: Using (4), calculate the corresponding Y_t statistic for each of the X_t values. As an initialization procedure, one may set $Y_0 = (\widehat{E(X_t)})$.

Step 5: Following (5), calculate the control limits by substituting the estimate $\hat{\lambda}$ in the expressions in (1) and (2). Note that L and ω must be identified accordingly based on desired ARL_0 considerations. A Markov chain approach to compute ARL_0 will be discussed later.

Step 6: Plot the Y_t values in the control chart based on the control limits in Step 5. If any Y_t is outside these control limits, then eliminate the corresponding Y_t (so is X_t) and return to Step 2. Otherwise, proceed to Phase II.

Phase II:

Step 7: The final control chart constructed in Phase I can then be prospectively applied on the $(m+1)^{th}$ sampling subgroup forward. If any $Y_t(\ t=m+1,\ \ m+2,...)$ is outside the control limits, then the charting procedure is stopped and in-depth investigations must be implemented to identify appropriate corrective actions to bring the process back in-control.

4. A Markov Chain Approach for ARL Computation for ZTP-EWMA Control Charts

Since the monitored statistics Y_t are not independent of one another, then the computation of ARLs must be based on the Markov chain approach first proposed by Brook and Evans (1972). Here, using the calculated control limits, the interval (LCL, UCL) is divided into S subintervals, each of length $2\delta = \dfrac{UCL - LCL}{S}$. Note that with this approach, each subinterval j ($j = 1, 2, \ldots, S$) can be seen as having its own control limits with the following form:

$$\begin{cases} UCL_j = LCL + 2j\delta \\ CL_j = LCL + (2j-1)\delta \\ LCL_j = LCL + 2(j-1)\delta \end{cases} \tag{6}$$

Each subinterval j is considered a state by itself, and the $(S+1)^{th}$ state is made to be *absorbing* and representing the out-of-control region. If the process is said to be in state j, then $Y_j = CL_j$. In this approach, the ARL is then the mean absorption time of the Markov chain. For the transition probability P_{ij} from state i to state $j \neq \mathrm{S}+1$, it can be shown that it is calculated as follows:

$$P_{ij} = ZTP_{cdf}\left(LCL + \frac{\delta}{\omega}\left(2j - (1-\omega)(2i-1)\right)\right) - ZTP_{cdf}\left(LCL + \frac{\delta}{\omega}\left(2(j-1) - (1-\omega)(2i-1)\right)\right),$$

where $ZTP_{cdf}(x)$ is the value of the cumulative distribution function of the ZTP-distributed random variable (with parameter λ) at x. Then for each $i \neq \mathrm{S}+1$, the probability of transitioning to the out-of-control state from state i is

$$P_{i(s+1)} = 1 - \sum_{j=1}^{s} P_{ij}.$$

$$Y_0 = \frac{\hat{\lambda}}{1 - e^{-\hat{\lambda}}}$$

$$j = \left\lceil \frac{E(X_t) \bullet S}{UCL} \right\rceil$$

Then, a vector $\boldsymbol{\mu}_{(S\times 1)} = [\mu_1, \mu_2, \ldots, \mu_S]'$ representing the ARLs of the process, where μ_j is the ARL of the process initialized at state j, is defined. If $\boldsymbol{Q}_{(S\times S)}$ is then defined to be the transition matrix based only on the in-control states, $\boldsymbol{I}_{(S\times S)}$ is the identity matrix of size S, and $\mathbf{1}_{(S\times 1)}$)is a column vector of ones, then Brook and Evans (1972) showed that μ is solved as

$$\boldsymbol{\mu} = (\boldsymbol{I} - \boldsymbol{Q})^{-1}\, \mathbf{1}.$$

That is, in initializing $Y_0 = \frac{\hat{\lambda}}{1 - e^{-\hat{\lambda}}}$ (i.e., midpoint initialization) and choosing S to be an odd number, then the computed ARL is the middle entry $\mu_{(S+1)/2}$ of the vector $\boldsymbol{\mu}$. However, this approach applies only to two-sided control charts. For one-sided charts (usually with only an upper limit), then for the same initialization, the computed ARL is the j^{th} entry of $\boldsymbol{\mu}$ corresponding to the subinterval j containing Y_0.

For computational demonstrations, we consider different in-control λs (i.e., 2.0, 3.5, and 5.0) commonly observed in health settings. Moreover, since in most health applications the importance is on guarding against increases in the mean levels, we only consider a one-sided (upper) control limit. The different out-of-control scenarios considered are defined by increases of 10%, 30%, and 50% in the in-control λ. Next, we present results only for some typical choices of ω (i.e., 0.10, 0.20, 0.30), following the recommendations of Fatahi et al. (2012) to keep ω small in health settings. For the computation of ARLs, we consider $S = 99$ and as an initialization procedure, we follow the most common approach by setting $Y_0 = \frac{\lambda}{1 - e^{-\lambda}}$. Note that in the results presented, the determination of ARL is based on a theoretical approach instead of the Phase I procedure. Also, using this approach, what can be computed are *initial* or *zero-state* ARLs (i.e., ARL based on the assumption that the implementation of the control chart is when the process has just begun).

For the in-control states, the ARL as a function of L are graphically shown in Figure 1 for the different in-control λs considered. Noticeably, the smaller

the weighting parameter ω, the longer the ARL. In fact, the difference is further magnified by larger values of L. Moreover, it is interesting to note the exponential growth in the ARL as L increases.

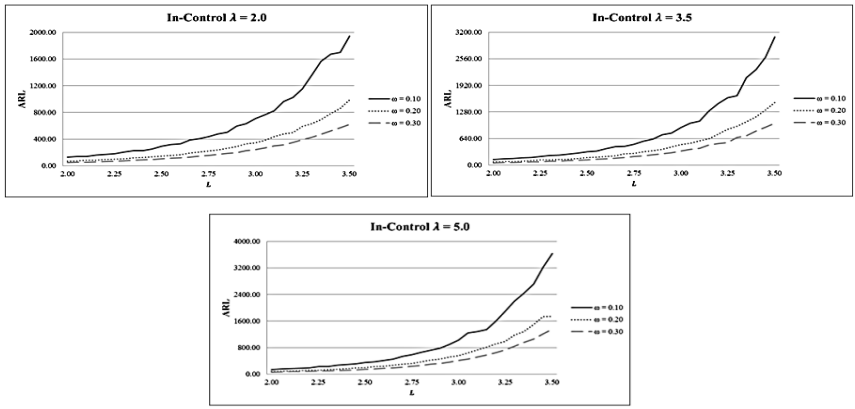

Figure 1. ARL_0 for in-control states with (A) λ=2.0, (B) λ=3.5, and (C) λ=5.0

Figures 2A-2C show the ARLs for out-of-control states with in-control λ = 2.0 (with increases of 10%, 30%, and 50%, respectively). For the out-of-control scenarios, similar observations can be made as with those seen for the in-control states. Those are: (1) ARLs increase exponentially as a function of L, and (2) smaller ω gives longer ARL. Now clearly in any control charting activity, it is desired to have longer in-control ARLs but shorter out-of-control ARLs. However, based on our computations, a choice has to be made regarding an optimal ω in the sense that longer in-control ARLs are achieved for reduced false alarms (which are typically provided by smaller ωs) and shorter out-of-control ARLs are obtained for increased sensitivity (which are typically provided by larger ωs). The choice is left to the user based on prior judgments or priorities. It can also be observed that when the out-of-control (increased) shifts are larger, ARLs are hugely reduced. Also, the differences between the out-of-control ARLs using different ωs are minimal when the increases in the in-control λ are large. These same observations hold true for in-control λs of 3.5 and 5.0, and for corresponding out-of-control increases of 10%, 30%, and 50%. Summary figures for the latter two settings of λ are available upon request.

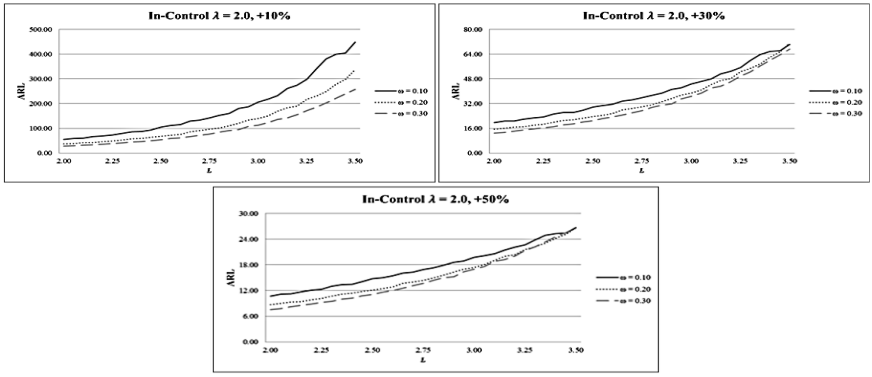

Figure 2. ARL1 for in-control λ=2.0 with increases of (A) 10%, (B) 30%, and (C) 50%

5. A FIR Feature for ZTP-EWMA Control Charts

There are certain situations when one wants to improve the sensitivity of the control chart at the start-up scheme. In this way, a FIR feature which is similar to a "head-start" is introduced to possibly increase sensitivity without risking too much false alarms. Usually, FIR features are implemented whenever: (1) one suspects a change before the monitoring scheme is started; or (2) one wants to detect a possibly ineffective control action. FIR features for EWMA charts have been widely studied in literature and are effectively summarized in Knoth (2005). In fact, FIR features may either be in any of the following forms: (1) adjusted initialization Y_0 value (Lucas and Saccucci, 1990); (2) time-varying control limits (Steiner, 1999); or (3) a combination of both (Rhoads, Montgomery, and Mastrangelo, 1996).

For this study, we consider the first approach, wherein we introduce a modified initial value for the monitoring process. Note that we only implement this for a one-sided control chart. In which case, we consider an adaptation of the work by Lucas and Saccucci (1990) in (7), i.e., the midpoint between $E(X_t)$ and UCL:

$$Y_0^* = \frac{E(X_t)+UCL}{2}. \tag{7}$$

We proceed to evaluate the performance of the control charts with FIR features using ARLs via the same Markov chain approach at the same parameter settings. In which case the corresponding ARLs are again the j^{th} entry of μ corresponding to the subinterval j containing the implemented FIR feature, i.e. $j = \left\lceil \frac{\left(E(X_t)+UCL\right)\cdot S}{2UCL} \right\rceil$.

Comparisons of ARLs and their corresponding variabilities for the same parameter settings at typical selections of L are given in Tables 1 to 4. To carry out the comparisons, both means and standard deviations are computed. For the computation of the standard deviations, factorial moments are considered. Higher s^{th} factorial moments of the run length distribution, given in the vector $\boldsymbol{\mu}^{(s)}$, are determined via the recursive relation

$$\boldsymbol{\mu}^{(s)} = s(\boldsymbol{I} - \boldsymbol{Q})^{-1}\, \boldsymbol{Q}\boldsymbol{\mu}^{(s-1)}.$$

Likewise, the j^{th} entry of $\boldsymbol{\mu}^{(s)}$ correspond to the s^{th} run length factorial moment of the scheme initialized at some value contained in the subinterval j.

Table 1 provides comparisons for the in-control cases upon implementation of the FIR feature considered in (7). For all λs considered, the decreases in ARLs are smaller (in fact, reductions are very minimal for larger values of L). However, run lengths for charts with a FIR feature have been found to be relatively more variable than the charts without a FIR feature. Nevertheless, the variability of the ARLs for the with FIR feature situations shows a declining trend as L increases.

Table 2 provide comparisons for the out-of-control situations as defined by different increases in the in-control λs upon implementation of the FIR feature considered in (7) for $\lambda = 2.0$. Note that the simulations were done for L ranging from 2.00 to 3.50, at a step of 0.10 units, but only those which are commonly used in practice are presented. Full results can be made available upon request. Expectedly, ARLs are shorter for the with FIR feature schemes. That is, the out-of-control situations are expected to be detected earlier when a FIR feature is implemented. Also, it is noteworthy that the trends in the reduction in both ARLs and their variability for similar in-control λs and associated increases are mostly similar regardless of the choice of ω. Those are: (1) the percentage reduction in ARLs decreases as L increases; (2) for the without FIR feature situations, the relative variability of the ARLs remain roughly similar or increases minimally as L increases; (3) for the with FIR feature situations on the other hand, the relative variability of the ARLs mostly decreases minimally as L increases; (4) ARLs for the without FIR feature schemes are relatively less variable than the schemes with a FIR feature implemented; and (5) the relative variabilities of the ARLs for both with and without FIR feature schemes decrease as the increases in the in-control λ are larger. The same results were obtained for the other in-control λs considered, but were not shown due to space constraints. Again, full tabular results can be made available upon request.

In summary, the implementation of the FIR feature defined in (7) has indeed reduced the ARLs, resulting to quicker out-of-control detection, but at the expense of higher false alarm rate (i.e., shorter in-control ARL) and lose of precision as evidenced by higher coefficient of variations. Nevertheless, the expenses are minimal, particularly with larger L.

Table 1. Comparisons for in-control cases upon implementation of the FIR feature							
L	Without FIR Feature			With FIR Feature			Decrease* (%)
	ARL	Std. Dev.	CV	ARL	Std. Dev.	CV	
	λ= 2.0, ω = 0.10						
2.00	128.20	126.90	0.99	110.66	125.06	1.13	13.68%
2.50	290.42	286.61	0.99	264.91	285.17	1.08	8.78%
3.00	709.22	702.91	0.99	672.48	701.85	1.04	5.18%
3.50	1945.53	1937.47	1.00	1895.68	1936.81	1.02	2.56%
	λ= 2.0, ω = 0.20						
2.00	68.57	67.53	0.98	60.21	66.80	1.11	12.19%
2.50	145.67	144.07	0.99	135.30	143.55	1.06	7.12%
3.00	345.95	343.59	0.99	332.88	343.27	1.03	3.78%
3.50	987.28	984.00	1.00	966.39	983.76	1.02	2.12%
	λ= 2.0, ω = 0.30						
2.00	47.78	47.32	0.99	43.41	46.95	1.08	9.13%
2.50	100.80	99.72	0.99	95.62	99.50	1.04	5.14%
3.00	240.57	238.90	0.99	231.27	238.69	1.03	3.87%
3.50	620.98	618.89	1.00	610.44	618.79	1.01	1.70%
	λ= 3.5, ω = 0.10						
2.00	136.21	133.97	0.98	119.13	132.35	1.11	12.53%
2.50	317.63	313.97	0.99	290.07	312.45	1.08	8.67%
3.00	901.56	894.91	0.99	862.31	893.97	1.04	4.35%
3.50	3091.03	3082.68	1.00	3043.92	3082.30	1.01	1.52%
	λ= 3.5, ω = 0.20						
2.00	74.14	72.96	0.98	66.59	72.34	1.09	10.19%
2.50	181.16	179.26	0.99	170.47	178.82	1.05	5.90%
3.00	492.29	489.35	0.99	477.16	489.07	1.02	3.07%
3.50	1511.93	1507.92	1.00	1487.73	1507.72	1.01	1.60%
	λ= 3.5, ω = 0.30						
2.00	53.83	53.02	0.98	48.51	52.62	1.08	9.88%
2.50	125.71	124.49	0.99	118.51	124.21	1.05	5.73%
3.00	339.96	337.90	0.99	330.37	337.74	1.02	2.82%
3.50	1013.11	1010.84	1.00	1000.18	1010.75	1.01	1.28%
	λ= 5.0, ω = 0.10						
2.00	135.08	132.75	0.98	117.56	131.08	1.12	12.97%
2.50	345.60	341.18	0.99	316.89	339.73	1.07	8.31%
3.00	1023.30	1016.64	0.99	986.02	1015.88	1.03	3.64%
3.50	3639.57	3631.01	1.00	3578.90	3630.50	1.01	1.67%
	λ= 5.0, ω = 0.20						
2.00	78.43	76.99	0.98	70.08	76.33	1.09	10.64%
2.50	196.92	194.46	0.99	184.94	194.00	1.05	6.08%
3.00	557.88	554.29	0.99	541.00	554.01	1.02	3.03%
3.50	1739.08	1734.81	1.00	1714.31	1734.63	1.01	1.42%
	λ= 5.0, ω = 0.30						
2.00	57.05	56.01	0.98	52.73	55.73	1.06	7.56%
2.50	141.31	139.65	0.99	134.86	139.44	1.03	4.56%
3.00	414.85	412.65	0.99	404.92	412.51	1.02	2.39%
3.50	1354.37	1351.90	1.00	1340.90	1351.83	1.01	0.99%
**decrease in ARL from without FIR feature to with FIR feature*							

Table 2. Comparisons for in-control cases for λ=2.0 with increases of 10%, 30%, and 50% upon implementation of the FIR feature							
L	Without FIR Feature			With FIR Feature			Decrease* (%)
	ARL	Std. Dev.	CV	ARL	Std. Dev.	CV	
ω = 0.10	+ 10%						
2.00	55.72	52.59	0.94	45.23	51.09	1.13	18.82%
2.50	104.72	99.13	0.95	89.98	97.83	1.09	14.08%
3.00	206.12	197.98	0.96	185.80	196.92	1.06	9.86%
3.50	448.80	438.56	0.98	422.63	437.82	1.04	5.83%
	+ 30%						
2.00	19.64	16.09	0.82	14.46	15.05	1.04	26.37%
2.50	29.73	24.22	0.81	22.72	23.19	1.02	23.59%
3.00	44.76	37.11	0.83	35.53	36.15	1.02	20.63%
3.50	70.41	60.63	0.86	59.13	59.84	1.01	16.03%
	+ 50%						
2.00	10.71	7.63	0.71	7.47	6.88	0.92	30.26%
2.50	14.75	10.24	0.69	10.44	9.44	0.90	29.20%
3.00	19.75	13.66	0.69	14.18	12.84	0.91	28.16%
3.50	26.64	18.91	0.71	19.98	18.18	0.91	24.99%
ω = 0.20	+ 10%						
2.00	36.71	35.06	0.95	30.96	34.43	1.11	15.68%
2.50	68.22	65.88	0.97	61.18	65.37	1.07	10.31%
3.00	139.71	136.50	0.98	130.96	136.14	1.04	6.26%
3.50	336.90	332.69	0.99	323.75	332.42	1.03	3.90%
	+ 30%						
2.00	15.30	13.34	0.87	11.99	12.86	1.07	21.64%
2.50	23.62	20.95	0.89	19.69	20.49	1.04	16.65%
3.00	38.79	35.21	0.91	34.00	34.83	1.02	12.36%
3.50	71.18	66.53	0.93	64.47	66.22	1.03	9.42%
	+ 50%						
2.00	8.69	6.83	0.79	6.45	6.45	1.00	25.77%
2.50	12.06	9.60	0.80	9.47	9.19	0.97	21.44%
3.00	17.37	14.11	0.81	14.26	13.73	0.96	17.88%
3.50	26.70	22.47	0.84	22.50	22.15	0.98	15.75%
ω = 0.30	+ 10%						
2.00	28.20	27.36	0.97	24.96	27.01	1.08	11.52%
2.50	53.58	52.11	0.97	49.72	51.87	1.04	7.22%
3.00	113.51	111.42	0.98	106.92	111.20	1.04	5.81%
3.50	257.78	255.22	0.99	250.32	255.10	1.02	2.90%
	+ 30%						
2.00	12.90	11.77	0.91	10.88	11.45	1.05	15.67%
2.50	20.97	19.24	0.92	18.55	18.98	1.02	11.54%
3.00	36.71	34.34	0.94	32.84	34.12	1.04	10.54%
3.50	67.31	64.40	0.96	62.94	64.24	1.02	6.49%
	+ 50%						
2.00	7.52	6.38	0.85	6.13	6.08	0.99	18.55%
2.50	11.08	9.38	0.85	9.40	9.13	0.97	15.16%
3.00	17.02	14.74	0.87	14.41	14.52	1.01	15.34%
3.50	26.74	23.92	0.89	23.82	23.74	1.00	10.95%
**decrease in ARL from without FIR feature to with FIR feature*							

6. Comparisons to Standard Shewhart and CUSUM Charts

Both proposed EWMA and FIR-EWMA charts were compared to the standard Shewhart and CUSUM (with and without FIR feature) for the same parameter settings considered. Briefly, the standard Shewhart chart for a ZTP process monitors the statistic X_t, the actual series counts, and has control limits as follows:

$$\begin{cases} UCL = E(X_t) + L\sqrt{Var(X_t)} \\ CL = E(X_t) \\ LCL = E(X_t) - L\sqrt{Var(X_t)} \end{cases},$$

where L is selected to be the desired minimum shift in the mean process level expressed in standard deviation units to be detected. The CUSUM chart monitors the following statistic:

$$Y_t = \max(0, X_t - \mathrm{k} + \mathrm{Y}_{t-1}),$$

where k is the reference value which is predetermined based on the minimum shift in the parameter that is desired to be detected. For ZTP processes, k is computed as

$$k = \frac{\lambda_1 - \lambda_0 - \ln\left(\frac{1-e^{-\lambda_0}}{1-e^{-\lambda_1}}\right)}{\ln\lambda_1 - \ln\lambda_0}.$$

The CUSUM chart is set to signal whenever $Y_t > h$, the upper control limit, which is predetermined to achieve a desired level of ARL_0 (Balamurali and Kalyanasundaram, 2013). Traditionally, CUSUM charts are only designed to monitor increases in the mean process, so no lower control limit is set. Standard application has $Y_0 = 0$; for a FIR design, $Y_0^* = \frac{h}{2}$.

For comparability with the previous simulation studies made, only one-sided (upper) control limits are considered. Also, L ranged from 2.00 to 3.50 for Shewhart charts and k for CUSUM charts is computed such that it is equivalent to detecting a shift of L-standard deviation units in the mean process level. The ARLs for standard Shewhart charts are computed as

$$ARL = \frac{1}{P(X_t > UCL \mid \lambda)'}$$

while ARLs for CUSUM charts are computed using the same Markov chain approach previously discussed for EWMA charts. Albeit, the differences are the

transition probability P_{ij} from state i to state $1 < j < S + 1$ are now computed as follows:

$$P_{ij} = (1 - ZTP_{cdf}(k)) \cdot (ZTP_{cdf}(\delta(2j - 2i + 1)) - ZTP_{cdf}\,\delta(2j - 2i - 1)),$$

the probability of transitioning to the out-of-control state from any state $i \neq S + 1$ is

$$P_{i(s+1)} = \left(1 - ZTP_{cdf}(k)\right) \cdot \left(1 - ZTP_{cdf}(h + k - \delta(2i - 1))\right),$$

and hence, for $j = 1$ from any state $i \neq S + 1$, $P_{i1} = 1 - \sum_{j=2}^{S+1} P_{ij}$.

Table 3 provides values of both ARL_0 and ARL_1 for Shewhart and CUSUM (where $h = 1$) charts for selected values of L for in-control $\lambda = 2.0$. For comparison against EWMA charts, one is referred to Table 1 for ARL_0 (i.e., in-control states) and to Table 2 for ARL_1 (i.e., out-of-control states). Complete tabulation of results for other values of L and $\lambda = 3.5$ and 5.0 can again be made available upon request.

Table 3. In-control and out-of-control ARLs for Shewhart and CUSUM charts for $\lambda = 2.0$

L	Shewhart Chart				CUSUM w/o FIR Feature				CUSUM w/ FIR Feature			
	ARL_0	ARL_1			ARL_0	ARL_1			ARL_0	ARL_1		
		+10%	+30%	+50%		+10%	+30%	+50%		+10%	+30%	+50%
2.00	16.42	12.27	7.55	5.14	96.95	58.50	25.33	13.13	95.21	56.93	24.05	12.09
2.50	52.20	35.70	18.88	11.32	270.70	146.03	52.25	23.31	264.76	141.22	48.95	20.95
3.00	190.71	119.17	53.92	28.36	300.72	164.57	59.89	26.77	298.59	162.72	58.48	25.68
3.50	190.71	119.17	53.92	28.36	840.73	425.97	136.08	54.47	833.14	419.76	131.76	51.33

Comparing the proposed EWMA charts against Shewhart, it is generally observed that ARL_0 for the former are notably longer than those of the latter. For ARL_1 on the other hand, general trends suggest that for lower values of λ, ω, and L, there are notably large differences between Shewhart and EWMA charts. For instance, at $\lambda = 2.0$, $\omega = 0.10$, and $L = 3.50$ for 50% shifts, the mean ARL_1 for the EWMA chart is 448.80 as compared to 119.17 for the Shewhart chart. However, the differences between the ARL_1 between the two charts decline as λ, ω, and L are set higher. There are even certain instances when the EWMA chart provides shorter mean ARL_1. For example (not shown in the tables), when $\lambda = 5.0$,

$\omega = 0.30$, and $L = 3.50$ for +50% shifts, mean ARL_1 for the EWMA chart is 12.02 as compared to 23.42 for the Shewhart chart. In summary, the proposed EWMA charts are generally found to be more sensitive in detecting larger shifts (+50%), albeit at a relatively small advantage as compared to lessened sensitivity against for smaller shifts (+10%) as compared to Shewhart charts.

In comparison against CUSUM charts, even with and without FIR features, we find that ARL0 for EWMA charts are still typically longer, particularly when CUSUM charts have a FIR feature. For example, at $\lambda = 2.0$, $\omega = 0.10$, and $L = 3.50$, the mean ARL0 for the FIR-EWMA chart is 1895.68 as compared to 833.14 for the FIR-CUSUM chart. In fact, FIR-EWMA charts generally have longer ARL_0 than standard CUSUM charts. This suggests that even if an EWMA is designed with a headstart feature, it is still expected to generate less false alarms than CUSUM charts without FIR. Assessing sensitivity against out-of-control situations, both proposed EWMA charts mostly have shorter ARL_1 under several settings, again making them more desirable than CUSUM counterparts. These advantages are most evident when ω for the EWMA chart is higher. For instance, at $\lambda = 2.0$, $L = 2.00$, and $\omega = 0.10$ at +10% shifts, the mean ARL_1 for the EWMA chart is 55.72 as compared to 58.50 for the CUSUM chart. However when ω is set at 0.30 while other parameters are at the same setting, the mean ARL_1 for the EWMA chart is only 28.20. This trend is also evident with larger (+30%, +50%) shifts where ARL_1 for EWMA charts are shorter than those of CUSUM counterparts. When the headstart FIR feature is implemented, the aforementioned trends are still exhibited. However, in the few cases where the choice of L is large (i.e., 3.00 or 3.50), CUSUM charts provide shorter ARL_1. For instance, at $\lambda = 2.0$, L=3.50, and ω=0.10 at +10% shifts, the mean ARL_1 for the EWMA chart is 448.80 as compared to 425.97 for the CUSUM chart. Overall, the EWMA charts can be considered advantageous in most situations as compared to the CUSUM charts by mostly having longer mean ARL_0 and shorter mean ARL_1.

7. Case Application

7.1. Dataset

The proposed EWMA control chart for a ZTP process was constructed and implemented using real clinical data. The dataset obtained contains the duration of hospital admission (recorded in number of days) of patients who have undergone either *appendectomy* (surgical removal of the appendix) or *cholecystectomy* (surgical removal of the gallbladder) from October 2010 to June 2015 in the particular hospital. A plot of the data is provided in Figure 3.

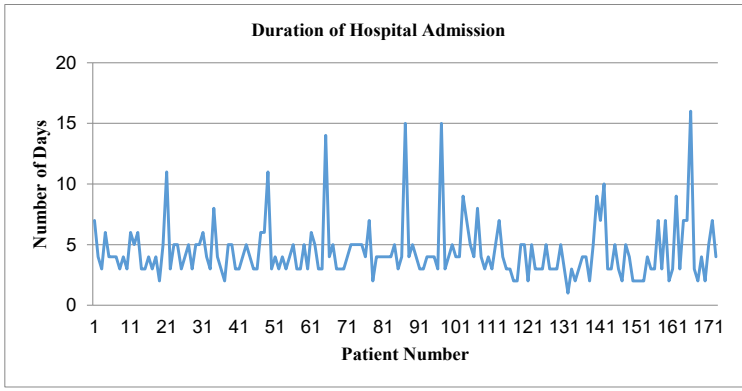

Figure 3. Time-series plot of the duration (in days) of hospital admission of patients who have undergone appendectomy or cholecystectomy from October 2010 to June 2015

7.2. Phase I procedure

For the Phase I procedure, the number of days of hospital admission for each patient from October 2010 to June 2014 were used to estimate the parameter λ assuming a ZTP distribution. It was observed that the mean duration was 4.3623 days, while the variance is 4.7875 days. Using these information, λ was estimated to be $\hat{\lambda} = 4.3033$ using (3).

To determine whether or not the assumption of a ZTP distribution for the duration of hospital admissions is valid, the dispersion test based on $\hat{U}_2^2$ from Best et al. (2007) was performed. The test statistic was computed as $\hat{U}_2^2 = 1.718$, with a *p-value* = 0.1899 based on a parametric bootstrap with 1,000 simulations. This suggests that the assumption of an underlying ZTP distribution is not violated.

Continuing with the control charting procedure, the values for the monitored statistic Y_t were computed using a weighting parameter of $\omega = 0.20$. For the central line and control limits of the chart, the formulation in (5) was followed resulting to as follows:

$$\begin{cases} \mathbf{UCL} = E(X_t) + L\sqrt{\dfrac{\omega}{2-\omega}Var(X_t)} = 6.5503 \\ \mathbf{CL} = E(X_t) \qquad\qquad\qquad\qquad = 4.3623 \end{cases}$$

Note that the lower control limit was not anymore computed since the goal of the charting procedure is to detect upward shifts in the mean process. After

plotting the Y_t values in the control chart, it was observed that no value of Y_t falls outside the control limits. As a result, no further pre-processing procedure was required, thus allowing the transition towards Phase II.

7.3. Phase II procedure

Proceeding with Phase II, the control chart constructed in Phase I was implemented on the data from July 2014 to June 2015. Figure 4 shows the plotted Y_t values, one for the basic chart and another with FIR feature, along with the control limits. Of the 35 sample points, both charts (without and without FIR) signaled an alarm on 28th observation. When investigated, the monitored statistics corresponded to a hospital stay 16 days. Noticeably however, only the FIR-enabled EWMA chart signaled an alarm on the 4th observation. For this it was determined that the monitored statistics corresponded to a hospital stay 10 days. Both observations resemble overly-prolonged admissions in relation to typical hospital stays for appendectomy and cholecystectomy surgeries. Indeed, further investigations of the patient profiles suggested that both were geriatrics and were monitored for longer durations due to some infections incurred while recovering. Clearly, this case application exhibits the sensitivity of both EWMA charts to detect anomalous increases after monitoring for a relatively long period. Also, it highlights the increased sensitivity of the proposed FIR-EWMA chart over the basic EWMA chart when anomalies happen during the early stages of Phase II monitoring.

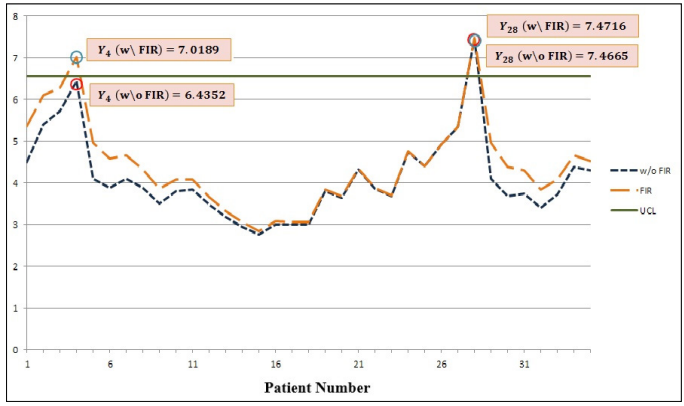

Figure 4. Visualization of the control chart implemented on data from July 2014 to June 2015

8. Conclusion

This paper has discussed the construction of an EWMA control chart for a ZTP process. The rationale behind the construction is to take advantage of the desirable properties of the EWMA charts such as its high sensitivity in monitoring high quality processes.

The ARLs were computed using a Markov chain approach. For both in-control and out-of-control states it was observed that ARLs increase exponentially as a function of L, and that smaller ω results to longer ARL. Based on the computations, the choice of optimal ω is left to the user based on prior judgments. It can also be observed that ARLs are reduced when out-of-control shifts are larger. Also, the differences between the out-of-control ARLs using different ωs are minimal when the increases in the in-control λ are large.

To improve the sensitivity of the control chart at the start-up scheme, a FIR feature was introduced. The initial value Y_0 was adjusted, and this was implemented for a one-sided control chart. Results showed that indeed, the introduced FIR feature has reduced the ARLs, resulting to a quicker out-of-control detection. However, it also resulted to a higher false alarm rate and a loss of precision. Even so, these disadvantages are minimized with larger L. The simulation results of the proposed EWMA charts were also studied against comparable settings of the traditional Shewhart, CUSUM, and FIR-enabled (i.e., with headstart) CUSUM. Extensive simulations suggested that both basic and FIR-enabled EWMA charts generally performed better by striking a balance between higher ARL_0 and lower ARL_1. Indeed, these advantages of the EWMA charts were more pronounced when larger shifts in the parameter λ happened. The viability of implementation with actual data was demonstrated in the last part using appendectomy and cholecystectomy data.

REFERENCES

BALAMURALI, S., and KALYANASUNDARAM, M., 2013, Design of cumulative sum control schemes for truncated Poisson distribution, *International Journal of Productivity and Quality Management*, 12(1), 94-119.

BARRETT, M. L., HINES, A. L., and ANDREWS, R. M., 2013, Trends in Rates of Perforated Appendix, 2001-2010. Retrieved from http://www.hcup-us.ahrq.gov/reports/statbriefs/sb159.pdf

BEST, D. J., RAYNER, J. C., and THAS, O., 2007, Goodness of fit for the zero-truncated Poisson distribution, *Journal of Statistical Computation and Simulation*, 77(7), 585-591.

BORROR, C. M., MONTGOMERY, D. C., and RUNGER, G. C., 1999, Robustness of the EWMA control chart to non-normality, *Journal of Quality Technology*, 31(3), 309-316.

BORROR, C. M., CHAMP, C. W., and RIGDON, S. E., 1998, Poisson EWMA control charts, *Journal of Quality Technology*, 30(4), 352-361.

BROOK, D., and EVANS, D. A., 1972, An approach to the probability distribution of CUSUM run length, *Biometrika*, 59(3), 539-549.

CHAKRABORTY, A. B., and KAKOTY, S. K., 1987, Cumulative sum control charts for ZTPD, *Journal of Indian Association for Productivity, Quality and Reliability*, 12, 17-25.

CHAKRABORTY, A. B., and SINGH, B. P., 1990, *Shewhart control chart for ZTPD. Proceedings of National Seminar on Quality and Reliability* (pp. 18-24). Trivandrum, India: NIQR.

COHEN, A. C. JR., 1972, Estimation in a Poisson process based on complete and truncated samples, *Technometrics*, 14(4), 841-846.

FATAHI, A. A., NOOROSSANA, R., DOKOUHAKI, P., and MOGHADDAM, B. F., 2012, Zero inflated Poisson EWMA control chart for monitoring rare health-related events, *Journal of Mechanics in Medicine and Biology*, 12(4), 1250065-1 - 1250065-14.

HE, S., HUANG, W., and WOODALL, W. H., 2012, CUSUM charts for monitoring a zero-inflated poisson process, *Quality and Reliability Engineering International*, 28(2), 181-192.

HE, S., LI, S., and HE, Z., 2014, A combination of CUSUM charts for monitoring a zero-inflated Poisson process, *Communications in Statistics-Simulation and Computation*, 43(10), 2482-2497.

KATEMEE, N., and MAYUREESAWAN, T., 2012, Control charts for zero-Inflated Poisson models, *Applied Mathematical Sciences*, 6(56), 2791-2803.

KNOTH, S., 2005, Fast initial response features for EWMA control charts. *Statistical Papers*, 46(1), 47-64.

LUCAS, J. M., 1985, Counted data CUSUMs, *Technometrics*, 27(2), 129-144.

LUCAS, J. M., and SACCUCCI, M. S., 1990, Exponentially weighted moving average control schemes: properties and enhancements, *Technometrics*, 32(1), 1-12.

MONTGOMERY, D. C., 2012, *Introduction to Statistical Quality Control* (7th ed.). New York: John Wiley and Sons, Inc.

MORTON, A. P., WHITBY , M., MCLAWS, M. L., DOBSON, A., MCELWAIN, S., LOOKE, D., et al., 2001, The application of statistical process control charts to the detection and monitoring of hospital-acquired infections, *Journal of Quality in Clinical Practice*, 21(4), 112-117.

MOUSTAKIDES, G. V., 1986, Optimal stopping times for detecting changes in distributions, *The Annals of Statistics*, 1379-1387.

NEUBAUER, A. S., 1997, The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation, *Clinical Chemistry*, 43(4), 594-601.

PLACKETT, R. L., 1953, The truncated Poisson distribution, *Biometrics*, 9(4), 485-488.

RHOADS, T. R., MONTGOMERY, D. C., & MASTRANGELO, C. M., 1996, A fast initial response scheme for the exponentially weighted moving average control chart,Quality *Engineering*, 9(2), 317-327.

ROBERTS, S. W., 1959, Control chart tests based on geometric moving averages, *Technometrics*, 1(3), 239-250.

STEINER, S. H., 1999, Exponentially weighted moving average control charts with time varying control limits and fast initial response, *Journal of Quality Technology*, 31(3), 75-86.

WOODALL, W. H., 2006, The use of control charts in health-care and public-health surveillance, *Journal of Quality Technology*, 38(2), 89-104.

XIE, M., HE, B., & GOH, T. N., 2001, Zero-inflated Poisson model in statistical process control. *Computational statistics and data analysis*, 38(2), 191-201.