Annales Univ. Sci. Budapest., Sect. Comp. 49 (2019) 129–146

MARKOV CHAIN-BASED COST-OPTIMAL CONTROL CHARTS WITH DIFFERENT SHIFT SIZE DISTRIBUTIONS

Balázs Dobi and András Zempléni
(Budapest and Debrecen, Hungary)

Communicated by András Benczúr

(Received March 31, 2019; accepted June 25, 2019)

Abstract. Control charts can be applied in a wide range of areas, this paper focuses on generalisations suitable for healthcare applications. We focus on the complete modelling of costs emerging during the process and investigate the effect of different shift size distributions on the optimal time between samplings, critical value and the resulting cost expectation and standard deviation. Namely, we show the results of optimisations for the exponential distribution and exponential-geometric mixture distributions.

1. Introduction

Control charts, as methods of statistical process control in industry have been introduced by Walter A. Shewhart in the 1920s [9]. Even though initially control charts were optimised with respect to statistical criteria, the concept of cost-efficient or cost-optimal control charts appeared not long after. One of the earliest and most important work in this field was done by Acheson J. Duncan in 1956 [4]. The concept is still very popular today as can be seen by published articles and developed software packages [10, 14].

Since the original inception of control charts, their use have been expanded considerably, such as to different fields of engineering [11], and also to areas

Key words and phrases: Control chart, cost-optimality, Markov-chain, healthcare.
2010 Mathematics Subject Classification: 62P10.
The work was supported by the project EFOP-3.6.2-16-2017-00015, which was supported by the European Union and co-financed by the European Social Fund.

further away from industrial settings. Our current research focuses on applications in healthcare settings. The authors' recent article on cost-optimal control charts showed an example for healthcare application on low-density lipoprotein monitoring [2]. Some articles in this field consider quality primarily, rather than costs [3]. For further healthcare examples see Suman et al [12]. Most of these cited papers deal with costs from an economic viewpoint, while our focus is the cost-optimal controlling of a healthcare characteristic where many costs only play a role indirectly, such as healthcare burden generated by complications.

In the above-cited paper, we developed a flexible framework which is based on the works of Zempléni et al [13]. This framework uses a Markov chain-based approach, which is similar to Duncan's cycle model as it also defines states related to the monitored process. The advantage of this approach was that it allowed generalisations for random shift sizes, random repairs and random sampling times, all of which are common in healthcare applications. Using these control charts, we were able to estimate the optimal parameters of a patient monitoring setup, which consisted of the optimal time between samplings (i.e. control visits) and critical value (i.e. medical criteria) [2].

In this paper we aim to assess the effect of different shift size distributions - distributions which model the degradation in quality - on the optimal parameters, expected cost and cost standard deviation. We demonstrate the flexibility and usefulness of the Markov chain-based framework with the comparison of a continuous and a mixed distribution. The choice of the distributions is motivated by their potential application in healthcare. We compare these distributions for different parameter setups. The implementations and results shown here were created using custom-made functions in the **R** programming language.

The rest of the paper consists of the following parts: Section 2 sets the mathematical background for the later results. Namely Subsection 2.1 briefly discusses the Markov chain-based framework and its generalisations. Subsection 2.2 deals with the implementation such as discretisation and programming. In Section 3, we show and discuss the results of different distributions and parameter setups. Section 4 concludes the paper.

2. Methods

The description of the Markov chain-based framework below is just a brief introduction and summary necessary for understanding the results presented later in the paper. For further reading and more detailed descriptions see Zempléni et al. and Dobi and Zempléni [2, 13].

2.1. The Markov chain-based framework and its generalisations

Our control charts are used to monitor a characteristic, based on a single sample element at a time. The shift in this characteristic - when it occurs - is positive. These assumptions correspond to an X-chart setup with sample size $N = 1$ and one-sided critical value K. Usually, there are three free parameters which are the focus of optimisation: the sample size N, the critical value K, and the time between samplings h. Since the sample size is fixed here, we are left with two parameters to use: K and h.

The following parameters and constants used throughout the paper are supposed to be known:

List of parameters

Notation	Meaning
μ_0	Target value, in-control expectation
σ	Process standard deviation
δ	Expected value of the exponential distribution (shift size)
q	Probability of geometric shift in the mixed distribution
ξ	Probability parameter of the geometric distribution
s	Expected number of shifts in a unit time
α, β	Parameters of the repair size beta distribution
c_s	Sampling cost
c_o	Shift-proportional out-of-control cost
c_{rb}	Base repair cost
c_{rs}	Shift-proportional repair cost

The followings are also assumed:

- The measurement error is normal with expectation 0 and known standard deviation σ. Let its cumulative distribution function (CDF) be denoted by ϕ. Thus the in-control process distribution is normal with parameters μ_0 and σ.

- The shift intensity ($1/s$ - the inverse of the expected number of shifts in a unit time) is constant.

- The shift size distribution is assumed to be known.

- The process does not repair itself, and when a repair (treatment) is carried out it does not repair the process perfectly. Furthermore, the repair itself is treated as an instantaneous event, thus all costs related to repairing should be included in the repair cost. For example if a major repair entails higher cost, then this should also be reflected in the calculation.

- The process standard deviation, the time between shifts, the shift sizes and repair effectiveness are all assumed to be independent from each other.

Using the above assumptions, the future distances from μ_0 are only dependent on the current distance. This way, one can define a Markov chain. The states of this Markov chain are defined at the sampling times and the type of the state depends on the measured value and the actual (unobservable) background process, namely whether there was a shift from the target value in the parameter. The possible difference between the two is due to the process standard deviation. This way four basic state types are defined:

- No shift - no alarm: in-control (INC)

- Shift - no alarm: out-of-control (OOC)

- No shift - alarm: false alarm (FA)

- Shift - alarm: true alarm (TA)

Our approach contains two important differences compared to the traditional models. First is the random shift size, introduced for the cost-optimal approach by Zempléni et al. [13]. The second is the random repair size, introduced by Dobi and Zempléni [2]. It is worth to note that the latter paper contains a third mayor generalisation in the form of random sampling time, which will not be discussed here, as we shall rather focus on the effect of the differences between the shift size distributions.

Let τ_i denote the random shift times on the real line and let ρ_i be the shift size at time τ_i. Assume that ρ_i follows a continuous distribution, which has a CDF with support over $(0, \infty)$, and that the shift sizes are independent from each other and from τ_i. Let the probability mass function (PMF) of the number of shifts after time t from the process' start (assumed to be in the in-control state) be denoted by ν_t. ν_t is a discrete distribution with support over $\mathbb{N}^0$. If the previous conditions are met, the resulting random process of the shifts - let us denote it by $H(t)$ - has step functions as trajectories, which are monotonically increasing between samplings. The CDF of the process values for a given time t from the start can be written the following way (assuming that there was no alarm before t):

$$Z_t(x) = \begin{cases} 0 & \text{if } x < 0, \\ \nu_t(0) + \sum_{k=1}^{\infty} \nu_t(k)\Psi_k(x) & \text{if } x \geq 0, \end{cases} \tag{2.1}$$

where $\Psi_k()$ is the CDF of the sum of k independent, identically distributed shift sizes ρ_i. The case $x = 0$ means there is no shift. The probability of zero shift size is just the probability of no shift occurring, which is $\nu_t(0)$.

Imperfect repair means that the treatment will not have perfect results on the health of the patient, or - in industrial settings - that the machines cannot be fully repaired to their original condition. In this case, the imperfectly repaired states act as out-of-control states. It is assumed that the repair cannot worsen the state of the process, but an imperfectly repaired process will still cost the same as an equally shifted out-of-control process, thus repaired and out-of-control states do not need distinction during the cost calculation. We define a random variable R to determine the proportion of the remaining distance from μ_0 after repair, and assume that it has a $Beta(\alpha, \beta)$ distribution with known parameters. Of course, this can be changed, depending on the repair process.

A process with the above assumptions and generalisations is shown on Figure 1. One can see that the process starts from an in-control state at μ_0.

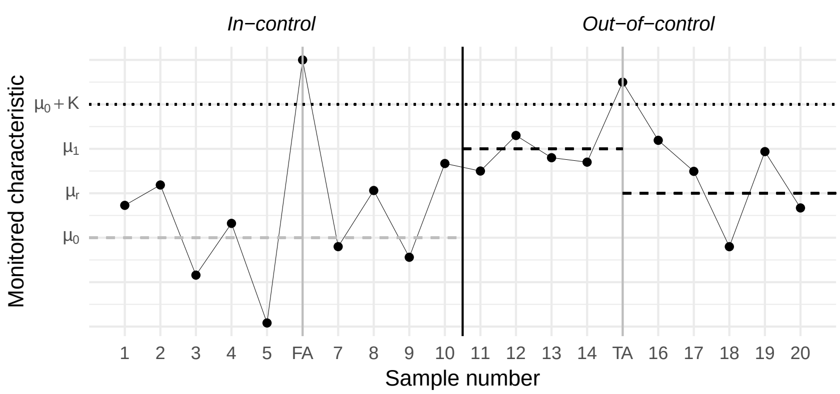

Figure 1: Definition of states
Dashed line: expected value, Black vertical line: shift in the expected value, Dotted line: critical value, FA: False alarm, TA: True alarm

Even though there is no shift, an alarm signal is still possible, which is a false alarm. After some time, there may be a shift (μ_1) in the value of the monitored characteristic (e.g. expected value), which creates out-of-control states. During this phase an alarm is called a true alarm, which induces a repair which is not necessarily perfect, thus the process stays out-of-control, but at a level (μ_r) which is closer to μ_0 than before the repair. Note that all of the states are defined at the time of samplings and that only positive shifts are possible.

The resulting process of the expected values is a monotone increasing step function between samplings and has a downward "jump" at alarm states - as the repair is assumed to be instantaneous.

The expected out-of-control operation cost can be written as the expectation of a function of the distance from the target value. At (2.1) the shift size distribution was defined for a given time t, but this time we are interested in calculating the expected cost for a whole interval. We propose the following calculation method for the above problem:

Proposition. *Let $H_{t_0,j}$ be the shifted mean-process (its unconditional distribution function Z_t was given in (2.1)) upon the condition that $H(t_0) = j$ and let $f()$ be a strictly monotonically increasing function over $\mathbb{R}_+$. The area $\mathcal{C}^f_{h,j}$ under the curve $\{t, E(f(H_{t_0,j}(t)))\}$, where $t_0 \le t \le t_0 + h$ can be written as:*

$$\mathcal{C}^f_{h,j} = \int_{t_0}^{t_0+h} \int_{f(j)}^{\infty} 1 - Z_{t-t_0}(f^{-1}(x) - j)dxdt + hf(j). \tag{2.2}$$

Proof. Let us observe that $E(f(H_{t-t_0}|H_{t_0} = j)) = \int_{f(j)}^{\infty} 1 - P(f(H_{t-t_0}) < < x|H_{t_0} = j)dx + f(j) = \int_{f(j)}^{\infty} 1 - Z_{t-t_0}(f^{-1}(x) - j)dx + f(j)$ by the monotonicity of f and since it is known that if X is a non-negative random variable, then $E(X) = \int_0^{\infty}(1 - F(x))dx$, where $F()$ is the cumulative distribution function of X. Furthermore, observe that this expected value is a monotonic function of t. So, the area under the curve is just (2.2). ■

Observe that (2.2) can be used to calculate the total cost generated by out-of-control operation over a h long sampling interval, given that the process starts at $\mu = j$.

2.2. Implementation

Exponentially distributed shift size. Let us assume first that the shift times form a homogeneous Poisson process, and the size of a single shift is exponentially distributed, independently of previous events. Using the notations of (2.1), ν_t would be the PMF of the Poisson distribution, with parameter ts - the expected number of shifts per unit time multiplied by the time elapsed. $\Psi_k()$ - the shift size CDF, for k shift events - would be now a special case of the gamma distribution, the Erlang distribution $E(k, \frac{1}{\delta})$, which is just the sum of k independent exponential variates each with mean δ.

For practical purposes we can apply the previous, general proposition to our model of Poisson-gamma shift size distribution. The connection between the distance from the target value and the resulting cost is often assumed not to be linear: often a Taguchi-type loss function is used - the loss is assumed to be proportional to the squared distance, see e.g. the book of Deming [1]. Applying this to the above proposition means $f(x) = x^2$. Since we are interested in the behaviour of the process between samplings, let $t_0 = 0$, thus:

$$\begin{aligned}\mathcal{C}_{h,j}^2 &= \int\limits_0^h \left[e^{-ts} j^2 + \left(\sum_{k=1}^{\infty} \frac{(ts)^k e^{-ts}}{k!} \cdot \int\limits_0^{\infty} (x+j)^2 \frac{(1/\delta)^k x^{k-1} e^{-x/\delta}}{(k-1)!} dx \right) \right] dt = \\ &= \int\limits_0^h e^{-ts} j^2 + \sum_{k=1}^{\infty} \frac{(ts)^k e^{-ts}}{k!} \big(k\delta^2 + (k\delta + j)^2 \big) dt = \\ &= \int\limits_0^h 2\delta^2 ts + (\delta ts + j)^2 dt = h^2 s\delta \Big(\delta + \frac{hs\delta}{3} + j \Big) + hj^2,\end{aligned}$$

where first we have used the law of total expectation - the condition being the number of shifts within the interval. If there is no shift, then the distance is not increased between samplings, this case is included by the $e^{-ts}j^2$ term before the inner integral. Note that the inner integral is just $E(X+j)^2$ for a gamma - namely an Erlang$(k, \frac{1}{\delta})$ - distributed random variable. When calculating the sum, we used the known formulas for $E(Y^2)$, $E(Y)$ and the Poisson distribution itself - where Y is a Poisson(ts) distributed random variable.

Mixture distribution as the shift size. Let us assume now that the shift times form a homogeneous Poisson process, as before, but the size of a single shift has a distribution which is a mixture of an exponential and a geometric distribution:

$$F_m(x) = qF_g(x) + (1-q)F_e(x),$$

where $q \in [0,1]$ is the mixing parameter, $F_g()$ is the CDF of the geometric distribution, $F_e()$ is the CDF of the exponential distribution, and $F_m()$ denotes the CDF of the resulting mixture distribution. Such a distribution can model processes with slow degradation, mixed with sudden jumps. A real-life example could be the effectiveness of a professional athlete, where the slow degradation can be attributed to age and sudden degradation to accidents. Another example could be simultaneously occurring chronic kidney disease and acute kidney injury [6]. Different definitions are available for the geometric distribution, the one used here has support over $\{1, 2, ...\}$, thus $F_g(x) = 1 - (1-\xi)^x$, with probability parameter ξ. A negative binomial distribution can be defined as the sum of independent geometrically distributed random variables with the same parameter. Its PMF can be written in the following way:

$$f_{nb}(x) = \binom{x-1}{x-r} \cdot \xi^r (1-\xi)^{x-r},$$

where r is the number of summed geometrically distributed variables. The support of this distribution is $x \in \{r, r+1, r+2, ...\}$. Let us assume now that n shifts occurred in a given time interval, and r of these were geometrically, while the rest were exponentially distributed. The CDF of the distribution of the sum of the shifts is then

$$F_{M|n,r}(x) = \begin{cases} F_{E_{n-r}}(x) & \text{if } r = 0, \\ \sum_{l=r}^{\lfloor x \rfloor} F_{E_{n-r}}(x-l) f_{nb}(l) & \text{otherwise,} \end{cases}$$

where $F_{E_{n-r}}$ is the distribution function of the Erlang distribution $E(n-r, \frac{1}{\delta})$. The distribution function of the sum of n variables with exponential-geometric mixed distribution is then

$$F_{M|n}(x) = \sum_{r=0}^{n} f_b(r) F_{M|n,r}(x),$$

where $f_b()$ is the PMF of the binomial distribution with parameters n and q, giving the probability of r geometrically distributed shifts out of n.

If one wants to construct the shift size distribution for a given time after start using the notations in (2.1), ν_t would be the PMF of the Poisson distribution, just as before, and $\Psi_k()$ would be $F_{M|n}()$ with $n = k$. (2.2) can also be used with this new shift size distribution, but the calculation of a closed form seems to be impossible, thus approximations will be used during application. The generality of (2.2) is highlighted by its application on mixture distributions, since there is no density function to be used to ease the calculations, like in the simpler, exponential case above.

Discretisation. For cost calculation purposes we would like to find a discrete stationary distribution which approximates the distribution of the monitored characteristic at the time of samplings. This requires the discretisation of the above defined functions, which in turn will allow us to construct a discrete time Markov chain with discrete state space.

A vector of probabilities is needed to represent the shift size PMF. Discretisation may introduce a bias: in reality, the distance from the target value can fall anywhere within a discretised interval, but without correction, the maximum of the possible values would be taken into account, which would be an overestimation of the actual shift size. We use a correction where the midpoint of the interval is used, which can still be somewhat biased, but in practice, with a fine enough discretisation, this effect is negligible.

Let Δ be the unit of the discretisation and V_d the number of states considered after discretisation. The in-control state is 0, thus the upper endpoints of the intervals are $0, \dots, (V_d - 1)\Delta$. We will define two functions for notational convenience, these will be used for correcting the discretisation bias throughout the paper:

$$\begin{aligned}\Delta_+(v) &= v\Delta + \frac{\Delta}{2}, \quad v = 0, \dots, V_d - 1,\\ \Delta_-(w) &= w\Delta - \frac{\Delta}{2}, \quad w = 1, \dots, V_d - 1.\end{aligned}$$

$v\Delta$ or $w\Delta$ would simply be the lower or upper boundary of the interval in consideration. The $\frac{\Delta}{2}$ term is added or subtracted to take the middle of the interval into account.

We can define the shift size PMF for a t long sampling interval, given that the starting state is i:

$$z_{t,i}(j) = \begin{cases} \nu_t(0) & \text{if } j = 0, i = 0,\\ \sum_{k=1}^{\infty} \nu_t(k)\big(\Psi_k(j\Delta) - \Psi_k((j-1)\Delta)\big) & \text{if } 1 \le j \le V_d - 1, i = 0,\\ \nu_t(0) + \sum_{k=1}^{\infty} \nu_t(k)\Psi_k(\Delta_+(j)) & \text{if } j = 0, i \neq 0,\\ \sum_{k=1}^{\infty} \nu_t(k)\big(\Psi_k(\Delta_+(j)) - \Psi_k(\Delta_+(j-1))\big) & \text{if } 1 \le j \le V_d - i - 1, i \neq 0. \end{cases}$$

For $j = 0$ the function is the probability of staying at the current level. The $i = 0$ case represents shifts from the healthy, in-control state. This case requires special treatment, since this value is actually given, unlike the other cases, where the value can fall anywhere within the discretised interval. Naturally, the infinite sums can only be approximated during application. This discretisation scheme works for both exponentially and mixture distributed shifts.

The discretised version of the repair size distribution can be written the following way:

$$R(l, m) = P\left(\frac{m}{l + 1/2} \le R < \frac{m+1}{l + 1/2}\right),$$

where l is the number of discretised distances closer to μ_0 than the current one. m is the index for the repair size segment we are interested in ($m \leq l$), with $m = 0$ meaning the best possible repair. The repair is assumed to move the expected value towards the target value by a random percentage, governed by $R()$. Even though discretisation is required for practical use of the framework, in reality the repair size distribution is continuous. To reflect this continuity in the background, the probability of perfect repair is set to be 0. l is set to be 0 when there is no repair, meaning $R(0, m) \equiv 1$. The $1/2$ terms are necessary for correcting the overestimation of the distances from the target value, introduced by discretisation.

Transition matrix and stationary distribution. The transition probabilities can be written using the $\phi()$ process distribution, the $z_{t,j}()$ shift size distribution and the $R()$ repair size distribution. The size of the transition matrix - let us denote it with Π - is $2V_d \times 2V_d$ since every shift size has two states: one with and one without alarm. The first V_d columns are states without alarm, the second V_d are states with alarm. Once the process leaves the healthy state it will never return, this is due to the nature of the imperfect repair we have discussed above.

The transition matrix defines a Markov chain with a discrete, finite state space with one transient, inessential class (in-control states) and one positive recurrent class (out-of-control and true alarm states). The starting distribution is assumed to be a deterministic distribution concentrated on the in-control state, which is to say that the process is assumed to always start from the target value. In finite Markov chains, the process leaves such a starting transient class with probability one. The problem of finding the stationary distribution of the Markov chain is thus reduced to finding a stationary distribution within the recurrent classes of the chain. Since there is a single positive recurrent class which is also aperiodic, we can apply the Perron–Frobenius theorem to find the stationary distribution [8]. If we consider now Π without the inessential class - let us denote it with Π' - then the stationary distribution - which is the left eigenvector of Π', normalised to sum to one - is unique and exists with strictly positive elements. Finding the stationary distribution is then reduced to solving the following equation: $\Pi'^T f_0 = f_0$, where f_0 is the left eigenvector of Π'. This amounts to solving $2V_d - 2$ equations - the number of states minus the in-control and false alarm states - for the same number of variables, so the task is easily accomplishable. The stationary distribution is then:

$$P = \frac{f_0}{\sum_{i=1}^{2V_d-2} f_{0_i}}.$$

Cost function. One can construct different cost functions to accommodate the process and monitoring setup at hand. Here, we present a general cost function which should be usable in many situations. One of the important parts of the cost function is the total cost due to faulty operation between samplings, thus we shall apply (2.2). The connection between the distance from the target value and the resulting cost will be assumed to be proportional to the squared distance. This cost will be expressed by an A^2 vector, which contains the weighted sum of the expected squared distances from the target value between samplings:

$$A_i^2 = \sum_{j=1}^{V_d-1} \mathcal{C}^2_{h,\Delta_-(j)} M_{ij}.$$

$\mathcal{C}^2_{h,\Delta_-(j)}$ is calculated using (2.2). j indicates one of the possible starting distances immediately after the sampling, and i indicates the state - shift - of the process at the current sampling. M_{ij} is the probability that $\Delta_-(j)$ will be the starting distance after the sampling, given that the current state is i. These probabilities can be written in a matrix form:

$$\begin{array}{c} \overbrace{\hphantom{R(1,0)\quad R(1,1)\quad R(2,2)\qquad\cdots}}^{\text{Distance from the target value starting from } \Delta_-(1)} \\ \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \\ 1 & 0 & 0 & \cdots \\ R(1,0) & R(1,1) & 0 & \cdots \\ R(2,0) & R(2,1) & R(2,2) & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\0\\0\\\vdots\end{matrix}}\right\}\text{Out-of-control} \\ \left.\vphantom{\begin{matrix}1\\R\\R\\\vdots\end{matrix}}\right\}\text{True alarm} \end{matrix} \end{array}$$

It can be seen, that when the process is out-of-control without alarm, the distance is not changed. The probabilities for the alarm states are calculated using the $R()$ repair size distribution.

The expected cost per unit time using the stationary distribution and A^2 is then:

$$E(C) = \frac{c_s + \sum_{i=1}^{V_d-1} \left(c_{rb} + c_{rs}\Delta_-^2(i)\right) P_{r_i} + c_o(A^2 \cdot P)}{h}.$$

The first term in the numerator is the sampling cost. The second term deals with the repair costs and the P_{r_i} true alarm probabilities. The repair cost is partitioned into a base and shift-proportional part: c_{rb} and c_{rs}, respectively. The true alarm probability is used, since it is assumed that repair occurs only if there is an alarm (and the false alarm state has probability 0 in the stationary distribution). The last term is the total cost due to faulty operation while the process is shifted.

So far only the expected cost was considered during the optimisation. In certain fields of application, the reduction of the cost standard deviation can be just as or even more important than the minimisation of the expected cost, [7]. Motivated by this, let us consider now the weighted average of the cost expectation and the cost standard deviation:

$$G = pE(C) + (1-p)sd(C).$$

Now G is the value to be minimised and p is the weight of the expected cost ($0 \leq p \leq 1$). The cost standard deviation can easily be calculated by modifying the cost function formula. All of the previous models can be used without any significant change, one simply changes the value to be minimised from $E(C)$ to G.

Implementation of the methods was done using the **R** programming language. Supplying all the necessary parameters, one can calculate the G value of the process for one time unit. It is also possible to minimise the G value by finding the control limit and the optimal time between samplings. All the other parameters are assumed to be known. The optimization step can be carried out using different tools, the results presented here were obtained with the built-in `optim()` **R** function. The optimisation procedure can be divided into three steps. First, the transition matrix needs to be constructed from the given parameters. After this, the stationary distribution of the Markov chain is computed. In the third step, the G value is calculated using the stationary distribution and the cost function. The optimisation algorithm then checks the resulting G value and iterates the previous steps with different time between sampling and/or control limit parameters until it finds the optimal ones.

3. Comparison of different distributions

Optimisation. In this section we will show cost calculation and optimisation results for different parameter setups and shift size distributions. First we ran cost calculations to create sort of a baseline, where both the expectation and the variance of the two shift size distributions were equal, and then we compared these with setups where only the expectations or the variances were equal. Table 1 contains information about the baseline and the comparison setups, as well as the resulting moments of the cost function, calculated using the stationary distribution. One can see that in the baseline model, the first two moments of the exponential and mixture shift size distribution are equal, but the higher moments are different. This also results in very similar first two moments in the stationary distributions. Only the first moments should be the same in the stationary distributions, as the calculation of the expected cost already contains squared values, as squared losses are taken into account due to the Taguchian loss function. The minor difference between the first moments

Distribution	**Data**	**Baseline**		**Comparisons**	
		Exponential $\delta = 2$	*Mixture1* $\delta = 1.5$, $\xi = 1/3$, $q = 1/3$	*Mixture2* $\delta = 1.5$, $\xi = 4/9$, $q = 2/3$	*Mixture3* $\delta = 1.5$, $\xi = 0.239$, $q = 0.1$
Shift size	$E(X)$	2	2	2	1.768
	$E(X^2)$	8	8	6.75	7.126
	$Var(X)$	4	4	2.75	4
	$E(X^3)$	48	50.5	33.563	51.966
	$E(X^4)$	384	446	222.938	602.87
Stationary $h = 1$, $K = 1$	$E(C)$	70.815	69.401	65.204	61.830
	$E(C^2)$	$2.360 \cdot 10^4$	$2.385 \cdot 10^4$	$1.748 \cdot 10^4$	$2.096 \cdot 10^4$
	$Var(C)$	$1.363 \cdot 10^2$	$1.379 \cdot 10^2$	$1.150 \cdot 10^2$	$1.309 \cdot 10^2$
	$E(C^3)$	$2.453 \cdot 10^7$	$2.632 \cdot 10^7$	$1.440 \cdot 10^7$	$2.532 \cdot 10^7$
	$E(C^4)$	$3.948 \cdot 10^{10}$	$4.445 \cdot 10^{10}$	$2.018 \cdot 10^{10}$	$4.629 \cdot 10^{10}$

Table 1: Shift size distribution moments and the resulting stationary distribution moments for $\sigma = 1$, $p = 1$, $s = 0.2$, $\alpha = 1$, $\beta = 3$, $c_s = 1$, $c_o = 20$, $c_{rb} = 60$, $c_{rs} = 10$

can be explained by computational biases introduced by e.g. approximations of infinite sums. Looking at the non-equal variance case, we can see that not even the expectations are equal if we compare the stationary distribution of the *Mixture1* and *Mixture2* columns. This is of course because of the squared loss function as mentioned above. The non-equal expectation case (*Mixture3*) shows that the resulting first and second moments in the stationary distribution are still different compared to the baseline *Mixture1*, but the higher order moments are similar. This experiment has demonstrated, that the first two moments play the most important role in the calculation of G.

Figure 2 contains optimal parameter values for the exponential distribution and the compared mixture distributions. The baseline showed near-identical results for the two distributions - the same as the solid lines on Figure 2 - thus these will not be discussed separately. Note that the results were created using a relatively high process standard deviation ($\sigma = 1$). Optimisations were run for different out-of-control costs and p values. $p = 1$ means that the cost standard deviation is not taken into account during optimisation.

Looking at the figures, one can assess that the critical value depends only weakly on the out-of-control cost, and that this dependence is affected by p. The time between samplings decreases with the increase of the out-of-control cost. The average cost and the cost standard deviation both increase with the out-of-control cost, as expected.

When the cost standard deviation is taken into account during the optimisation procedure ($p = 0.75$), then lower critical value should be used with

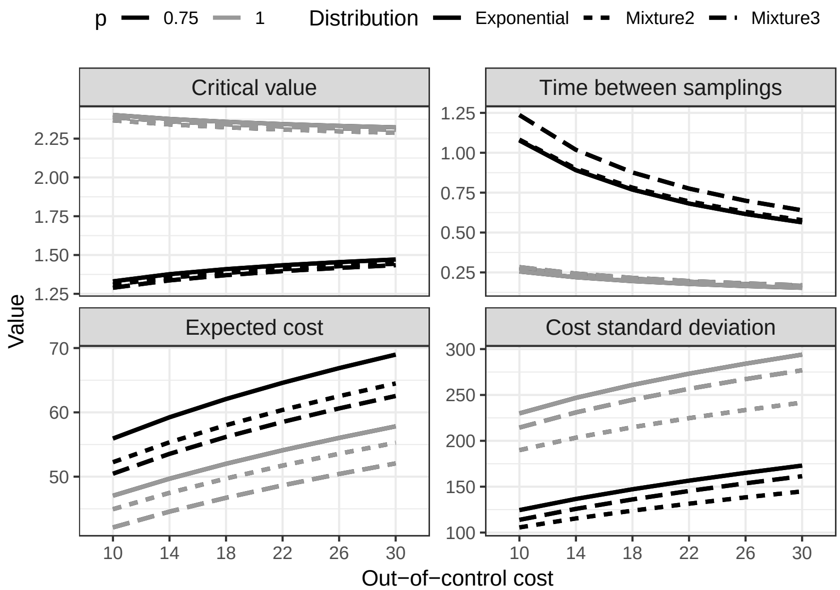

Figure 2: Optimal parameters and the resulting expected cost and cost standard deviation as function of the out-of-control cost (c_o) and weight of the cost expectation during optimisation (p) for $\sigma = 1$, $s = 0.2$, $\alpha = 1$, $\beta = 3$, $c_s = 1$, $c_{rb} = 60$, $c_{rs} = 10$

increased time between samplings. This is logical, because the increased time between sampling will lead to less frequent interventions, thus a less erratic process. Of course, at the same time we do not want to increase the expected cost, so the critical value is lowered. The cost standard deviation is decreased, as expected. What is interesting to note is that the expected costs have only mildly increased compared to the $p = 1$ setup. This is important, because it shows that by changing the parameters appropriately, the cost standard deviation can be lowered - sometimes substantially - while the expected cost is only moderately increased.

The difference between the results obtained with different distributions is virtually non-existent if we look at the critical values and the times between samplings, with one exception: a shift size distribution with lower expectation entails somewhat longer times between samplings, if the standard deviation is also taken into account as can be seen at the upper-left plot. The differences in the resulting expected costs and cost standard deviations are much more visible. A shift size distribution with lower expectation or standard deviation entails lower expected cost and cost standard deviation. Namely, *Mixture2*, which has the lowest moments starting from the order of 2, entails the lowest cost standard deviation and a lower cost expectation than the baseline. This is as expected, since second moments appear even in the calculation of the expectation. *Mixture3*, which has the lowest expectation, entails the lowest cost expectation, but a higher cost standard deviation, than *Mixture2*. It seems like - at least for these particular distributions - that only the first two moments have major effect on the optimal parameters and on the resulting expected costs and cost standard deviations. This can be quite useful in situations where only the first two moments of the shift size distribution can be reliably estimated.

Relationship of parameters. It is often helpful to look at the relationship between the parameters and the resulting expected cost and cost standard deviation. Figure 3 shows G values on a contour plot as function of the time between samplings and the critical value, using the mixture distribution. The parameters correspond to the ones used on Figure 2 for *Mixture2* with $p = 0.75$ and $c_o = 20$ (the dotted, black lines in the middle of the plots). The black dot roughly in the middle of the plot marks the optimal parameters

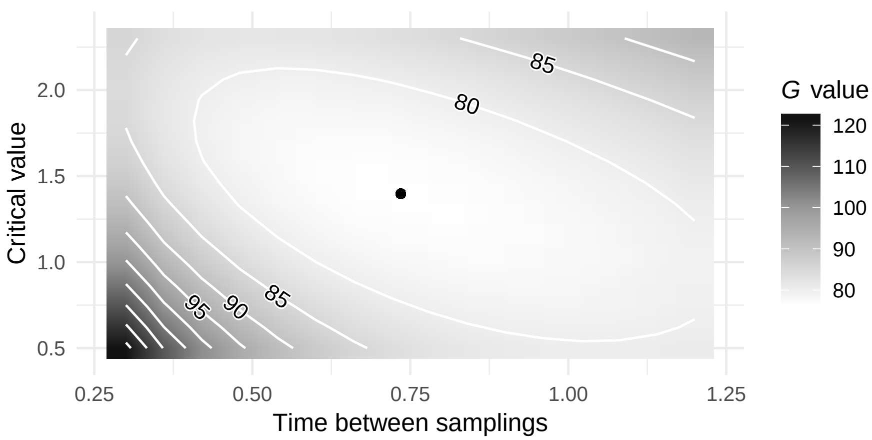

Figure 3: G value as function of the time between samplings and the critical value, exponential-geometric mixture shift size distribution with $q = 2/3$, $\delta = 1.5$, $\xi = 4/9$, $\sigma = 1$, $p = 0.75$, $s = 0.2$, $\alpha = 1$, $\beta = 3$, $c_o = 20$, $c_s = 1$, $c_{rb} = 60$, $c_{rs} = 10$

which entail the lowest possible G value. A clear elliptical shape can be seen, suggesting a close-to-linear relationship between the conditional minima: the optimal critical values as a function of the time between samplings lie on a line with negative slope.

Stationary distribution. Even though in some cases the results can be similar when using different distribution, the inspection of the stationary distributions can yield useful information. Comparisons will be made for $q = 0.9$. Figure 4 contains results about the theoretical and simulated stationary distributions, generated with the exponential and the mixture distribution. Simulations were run for 20000 sampling intervals. Simulations from the first 100 sampling intervals were discarded as a burn-in stage. Comparison to simulated results has proved to be an important tool during both result-assessment and error-finding. Note that the stationary distribution is interpreted at the time of samplings, before repair (in case of an alarm state). One may assess that the theoretical stationary distributions fit the simulated results quite well.

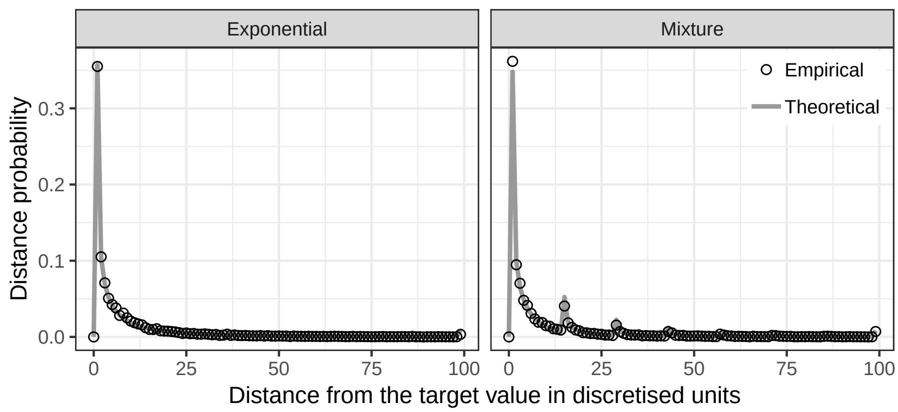

Figure 4: Theoretical and empirical stationary distributions using $h = 1$, $K = 1$, $q = 0.9$, $\delta = 1.5$ (for both distributions), $\xi = 4/9$, $\sigma = 1$, $p = 0.75$, $s = 0.2$, $\alpha = 1$, $\beta = 3$, $c_o = 20$, $c_s = 1$, $c_{rb} = 60$, $c_{rs} = 10$

The probability of the in-control state is 0, as expected, and the highest probability can be seen at the state just above the target value for both distributions. The increase in probabilities at the furthest distance taken into account is due to the finite number of states, (since the support of the distributions in reality is not finite). In the case of the exponential distribution, a near-exponential curve can be seen, as the probability rapidly decreases with the increasing distance from the target value. If we compare this to the results generated with the mixture distribution, the effect of the geometric distribution can clearly be seen, as the distribution is now multimodal. The peaks correspond to intervals around integers. Following the peaks, a near-exponential decrease can be seen in the probability. This can be attributed to the effect of exponentially distributed shifts added to the discrete shift. The shape of the repair size distribution is also such, that it promotes repairs relatively close to the target value.

4. Conclusions

This article explored the effect of different shift size distributions in the generalised Markov chain-based cost-optimal control chart model. We compared exponentially distributed shift sizes with mixture distributed ones. The mixture distribution was defined using the exponential and the geometric distributions. We showed that both distributions can be easily fitted into the Markov chain-based framework. The cost calculation between samplings was also possible with both distributions.

Our results show that the increase in the out-of-control cost has little-to-no effect on the critical value, while it decreases the time between samplings and increases both the cost expectation and cost standard deviation. The inclusion of the cost standard deviation into the optimisation algorithm decreased the critical value, increased the time between samplings, somewhat increased the expected cost and considerably decreased the cost standard deviation for all investigated cases. Our results with different shift size distributions and target functions show that - in the analysed cases - practically only the first two moments are important - although one might have suspected that higher moments would play a role as well. The relationship between the time between samplings, the critical value and the resulting expected cost and cost standard deviation can be described by elliptical shaped contours, suggesting a close-to-linear relationship between the parameters. The inspection of the stationary distributions revealed markedly different shapes, where the effect of the shift size distribution is clearly visible.

References

[1] **Deming, W.E.,** *The New Economics for Industry, Government, Education*, MIT Press, 2018.

[2] **Dobi, B. and A. Zempléni,** Markov chain-based cost-optimal control charts for health care data, *Quality and Reliability Engineering International*, **35(5)** (2019), 1379–1395.

[3] **Duclos, A., S. Touzet, P. Soardo, C. Colin, J.L. Peix, and J.C. Lifant,** Quality monitoring in thyroid surgery using the Shewhart control chart, *British Journal of Surgery*, **96(2)** (2009), 171–174.

[4] **Duncan, A.J.,** The Economic Design of X-Charts Used to Maintain Current Control of a Process, *Journal of the American Statistical Association*, **51(274)** (1956), 228–242.

[5] **Gay, D.M.,** Usage summary for selected optimization routines, *Computing Science Technical Report, AT&T Bell Laboratories, Murray Hill*, **153** (1990), 1–21.

[6] **Hsu, R.K. and C. Hsu,** The role of acute kidney injury in chronic kidney disease, *Seminars in Nephrology*, **36(4)** (2016), 283–292.

[7] **McCracken, A.K. and S. Chakraborti,** Control charts for joint monitoring of mean and variance: an overview, *Quality Technology & Quantitative Management*, **10(1)** (2013), 17–36.

[8] **Meyer, C.D.,** Matrix analysis and applied linear algebra, *SIAM*, 2000, 663–669.

[9] **Montgomery, D.C.,** *Introduction to Statistical Quality Control*, Sixth Edition, John Wiley & Sons Association, 2009.

[10] **Mortarino, C.,** Duncan's Model for $\bar{X}$-Control Charts: Sensitivity Analysis to Input Parameters, *Quality and Reliability Engineering International*, **26(1)** (2010), 17–26.

[11] **Sales, R.F., R. Vitale, S.M. de Lima, M.F. Pimentel, L. Stragevitch and A. Ferrer,** Multivariate statistical process control charts for batch monitoring of transesterification reactions for biodiesel production based on near-infrared spectroscopy, *Computers & Chemical Engineering*, **94** (2016), 343–353.

[12] **Suman, G. and D. Prajapati,** Control chart applications in healthcare: a literature review, *International Journal of Metrology and Quality Engineering*, **9(5)** (2018), Article Number 5, Number of pages 21. https://www.metrology-journal.org/

[13] **Zempléni, A., M. Véber, B. Duarte and P. Saraiva,** Control charts: A cost-optimization approach for processes with random shifts, *Applied Stochastic Models in Business and Industry*, **20(3)** (2004), 185–200.

[14] **Zhu, W. and C. Park,** edcc: An R Package for the Economic Design of the Control Chart, *Journal of Statistical Software*, **52(9)** (2013), 1–24.

B. Dobi and A. Zempléni
Department of Probability Theory and Statistics
Eötvös Loránd University
Budapest
Faculty of Informatics
University of Debrecen
Debrecen
Hungary
dobib@cs.elte.hu
zempleni@caesar.elte.hu