



ORIGINAL ARTICLE

# The integrated model of the Kolmogorov–Smirnov distribution-free statistic approach to process control and maintenance



P. Khruaesom, A. Pongpullponsak \*

Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha-utid Road, Bangmod, Toongkru, Bangkok 10140, Thailand

Received 14 November 2015; accepted 20 April 2016  
Available online 27 April 2016

## KEYWORDS

Integrated model;  
Maintenance Management;  
Distribution-free;  
Kolmogorov–Smirnov;  
Nonparametric linear  
regression;  
Genetic algorithms

**Abstract** This research is aimed to find the optimal values of four variables ( $n, h, L, k$ ) that minimize the cost of integrated system approach to process control from considering the basis of the EWMA control chart integrated model (Charongrattanasakul and Pongpullponsak, 2009, 2011) and Kolmogorov–Smirnov (KS) control chart Khruaesom and Pongpullponsak (2014). The proposed mathematical model is used to analyze the cost of the integrated model before the genetic algorithms (GA) approach is carried out in order to calculate the optimal values of four variables ( $n, h, L, k$ ) that minimize the hourly cost. Subsequently, they are subjects to the nonparametric linear regression test in order to confirm the optimal values of four variables. A comparison between four policies of integrated model KS and other models indicates that the integrated KS model has a better economic behavior when it is distribution-free. Finally, the performance or average run length ( $ARL$ ) obtained from the KS model is greater than that of the general model.

© 2016 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Since the first control chart was provided in 1924 by Dr. She-whart, the concept of the control chart has been considered in many models from past to present, and currently the quality control is widely used in various industry areas. One popular

application on control chart used in manufacturing is economic design of control charts. Montgomery (2009) stated that the economic design of control charts refers to the control charts that have been designed with respect to statistical criteria only. This usually involves selecting sample size and control limits such that the average run length ( $ARL$ ) of the chart to detect a particular shift in the quality characteristic and the  $ARL$  of the procedure when the process is in control are equal to specified values. Practically, frequency of sampling is considered from factors including production rate, expected frequency of shifts to an out-of-control state, and possible consequences of such process shifts in determination of sampling interval. In many cases, statistical criteria and practical experience have been used in setting up general guidelines for

\* Corresponding author.

E-mail address: [adisak.pon@kmutt.ac.th](mailto:adisak.pon@kmutt.ac.th) (A. Pongpullponsak).

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

the design of control charts. One thing that should be noted in designing a control chart is the choice of control chart parameter can cause economic consequences. For this reason, an economic viewpoint such as costs of sampling and testing, costs associated with investigating out-of-control signals and possibly correcting assignable causes, and costs of allowing nonconforming units to reach a consumer should be taken into account in the design of a control chart.

In recent years, considerable research has been devoted to economic models of control charts. Duncan (1956) used optimum methodology to establish design parameters including subgroup size ( $n$ ), sampling interval ( $h$ ), and control-limit width ( $\pm L$  standard deviations) for minimizing the loss cost, where the cost items include a sampling and testing cost, an increasing cost from out-of-control process, false alarm cost, and searching and repairing cost. Pongpullponsak et al. (2009) studied a chart in conjunction with an age replacement preventive maintenance policy. From the report, they also introduced an economic model established using the Shewhart method and determined the efficiency of the control chart when the data were in skewed distributions. Saniga (1977) proposed a joint economic design of and  $R$  control charts based on two assignable causes in production process. In this model, one assignable cause results in a shift of the process mean whereas the other one results in a shift of the process variance. Yu et al. (2010) studied the possibility in economic statistical design of control charts by considering only one assignable cause. In fact, multiple assignable causes such as machine problem, material deviation, human errors, etc. can occur during the production process so for this research, establishment of an economic-statistical model of control chart will be extended from consideration of single, in the original research, to multiple assignable causes for a real application. Zhou and Zhu (2008) established an economic statistical design of control chart from integrating the concepts between Statistical Process Control (SPC) and Maintenance Management (MM), which are in science and business practice, respectively. The integrated model was then used to find the optimal values of policy variables ( $n, h, L, k$ ) that minimize hourly cost, subsequently optimal product quality, little downtime, and cost reduction can be achieved by controlling variances in the process. The effects of cost parameters on the solution of the design were investigated using a numerical experiment.

Generally, the process in construction of control chart includes gathering the sampling properties of monitoring statistic, determining the chart's behavior, and comparing its performance with other existing charts. In most cases, the data are treated as normality but in some occasions, the underlying population distribution or the necessary sampling properties could not be ruled out. Some researchers have proposed reasonable alternative as a solution. For instance, Yang et al. (2011) introduced a nonparametric approach, namely a nonparametric EWMA sing control chart for dealing with such situations. The Kolmogorov–Smirnov (KS) control chart is another good example. The knowledge of sampling distribution is useful for nonparametric statistics inference because the exact sampling distributions are considerably easier to calculate compared with that for distribution-free statistic ( $D_n$ ), or Kolmogorov–Smirnov one-sample statistic (Gibbons, 1971). Pongpullponsak and Jayathavaj (2014) considered distribution-free statistics when the population from which

selected samples are not normally distributed or normality cannot be met was used. Additionally, there are several nonparametric tests that have been further applied in case of nominal or ordinal data. Bakir (2012) introduced a modified version of the two – sample Kolmogorov–Smirnov test statistic where the difference of the reference and test empirical distribution function are maximized only over the training sample values. Khruasom and Pongpullponsak (2014) developed the distribution-free or unknown distribution quality control chart based on the KS.

This research is aimed to find the optimal values of four variables ( $n, h, L, k$ ) that minimize the cost of integrated system approach to process control and maintenance model on the basis of the distribution-free/KS – control chart.

## 2. Materials and methods

### 2.1. Nomenclature

#### Cycle time ( $E [T]$ )

$T_0$	expected time searching for a false alarm
$T_P$	the expected time to identify maintenance requirement and to perform a planned maintenance
$T_A$	the expected time to determine occurrence of assignable causes
$T_R$	the expected time to identify maintenance requirement and to perform a reactive maintenance
$T_C$	the expected time to perform a compensatory maintenance
$\tau$	the mean elapse time from the last sample before the assignable cause to the occurrence of the assignable
$ARL_1$	the average run length during the in-control period
$ARL_0$	the average run length during the out-of- control period
$E$	the expected time to sample and chart one item

#### Cycle cost ( $E [C]$ )

$C_I$	the cost of quality loss per unit time (the process is in an in-control state)
$C_0$	the cost of quality loss per unit time (the process is in an out-of-control state)
$C_P$	the cost of performing planned maintenance
$C_R$	the cost of performing reactive maintenance
$C_C$	the cost of performing compensatory maintenance
$C_F$	the fixed cost of sampling
$C_V$	the variable cost of sampling
$C_f$	the cost to investigate a false alarm

The indicator variable equals 1 if production continues during planned maintenance ( $\gamma_P$ ), reactive maintenance ( $\gamma_R$ ), compensatory maintenance ( $\gamma_C$ ), validate assignable cause ( $\gamma_A$ ) or 0 otherwise

$P_i^1$	the probability that run length of control chart equals $i$ during in-control period $P_i^1 = \alpha(1 - \alpha)^{i-1}$
$P_i^0$	the probability that run length of control chart equals $i$ during out-of- control period $P_i^0 = \beta(1 - \beta)^{i-1}$

#### Optimal variable

$n$	the sampling size ( $n^*$ for optimal)
$h$	the interval between sampling ( $h^*$ for optimal)
$L$	the width of control limit in units of standard deviation ( $L^*$ for optimal)
$K$	the number of samples taken before planned maintenance ( $k^*$ for optimal)

2.2. Problem statement and assumptions

Fig. 1 shows the framework in development of integrated model. The process starts with an in-control state with a process failure mechanism that is distribution-free or follows an unknown distribution. The best-known test is the KS, the cumulative distribution function (cdf) of the sample, called the empirical distribution function (edf), may be considered, and an estimate of the population and  $D_n$  can be written as follows:

$$D_n = \sup_x |S_n(x) - F_X(x)| \tag{1}$$

For the random variable  $S_n(x)$ , which is the edf of a random sample  $X_1, X_2, \dots, X_n$ , a distribution  $F_X$  can be derived from “Nonparametric statistical Inference” by Gibbons (1971),

$$P \left[ S_n(x) = \frac{i}{n} \right] = \binom{n}{i} [F_X]^i [1 - F_X]^{n-i}, \quad i = 0, 1, \dots, n \tag{2}$$

Define the indicator random variables as below;

$$\delta_i(t) = \begin{cases} 1 & \text{if } X_i \leq t \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The  $\delta_1(t), \delta_2(t), \dots, \delta_n(t)$  constitutes a set of  $n$  independent random variables from the Bernoulli distribution with parameter  $\theta$ , where  $\theta = P[\delta_i(t) = 1] = P(X_i \leq t) = F_X(t)$ . Therefore, we obtain

$$S_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_i(x) \tag{4}$$

The random variables  $nS_n(x)$  is the sum of  $n$  independent Bernoulli random distribution, which follows the binomial distribution with parameter  $\theta = F_X(x)$ .

From Eqs. (2)–(4), we consider a sequence of  $t$  independent Bernoulli trials, where the probability of the event is  $\lambda$  and the probability of the non-event is  $1 - \lambda$ . If we consider the event to be the elimination of the player, then its absence over  $n$  trials can be described as their survival. The probability of this survival throughout  $t$  trials will be given by the binomial mass function  $B(x; t, \lambda)$  when  $x = 0$  (Pollock, 2007), which is

$$B(0, t, \lambda) \tag{5}$$

Likewise, in the case of the non-occurrence of eliminating event over a continuous finite period of time the model of survival or elimination can be converted through Bernoulli trials into similar model in which the events are distributed randomly in time. To achieve this, a Poisson’s process in continuous time should be depicted as a limiting case of a binomial process. It can be carried out by taking, as the departure point, the special case of the binomial given under Eq. (5) and considering that of each trial representing a single unit of time. Subsequently, the probability of the occurrence of the eliminating event within a single period can be denoted by  $\lambda$ . Also, it is assumed that the probability of the occurrence of two such events within the same interval is vanishingly small or zero.

Then

$$B(0; t, \lambda) = \left[ \left( 1 - \frac{\lambda}{n} \right) \right]^{n \cdot t} \approx (1 - \lambda)^t, \tag{6}$$

where the approximation is obtained by taking the first two terms of the binomial expansion.

$$\left( 1 - \frac{\lambda}{n} \right)^n = (1 - \lambda) + \frac{(n-1)}{2n} \lambda^2 - \frac{(n-1)(n-2)}{n^2 2!} \lambda^3 + \dots \tag{7}$$

In the limit, when the number of subdivisions increases indefinitely, we have

$$\lim_{n \rightarrow \infty} \left( 1 - \frac{\lambda}{n} \right)^n = e^{-\lambda} \tag{8}$$

At this point, the probability of survival in the period  $0, t$  can be described by

$$S(t) = e^{-\lambda t} \tag{9}$$

which is the continuous-time analog of Eq. (5). The probability of being eliminated during the period  $[0, t]$  or a cdf is written as

$$1 - S(t) = F(t) = \int_0^t f(s) ds = 1 - e^{-\lambda t} \tag{10}$$

Next, the corresponding density function, defined over the set of times at which elimination might occur, or a probability density function is

$$f(t) = \lambda e^{-\lambda t} \tag{11}$$

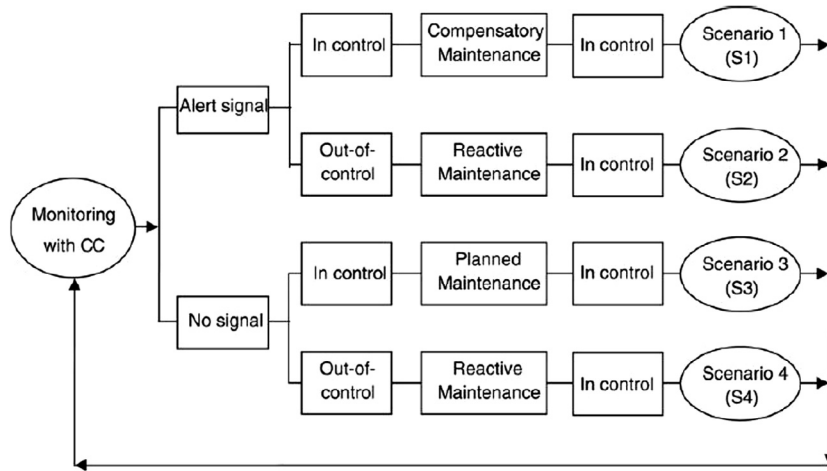


Figure 1 Four monitoring – maintenance scenarios of the integrated mode.

which is a cumulative function  $F(t)$  in respect of  $t$ . This exponential waiting time function is used to describe a duration distribution.

### 2.3. Monitoring scenarios

In Scenario 1, the process begins at “in-control” state. Inspections start after  $h$  hours of monitoring to investigate whether or not the process has shifted from an “in-control” to an “out-of-control” state. An alert signal has been added in the control chart before the scheduled time when maintenance should be carried out. But the signal will be false, if the process is still “in-control”. In this case, searching and determining false signal consumes time and incurs cost. Compensatory Maintenance will be performed.

In Scenario 2, there is an alert signal similar to Scenario 1. However, for this case the signal is valid and the process shifts to an “out-of control” state, consequently Reactive Maintenance is active.

In Scenarios 3 and 4, there is no signal in the control chart before the scheduled time. For this reason, at the  $(k + 1)^{th}$  sampling interval, appropriate maintenance is assigned. In Scenario 3, since the process is always “in-control”, Planned Maintenance is performed. But when the process shifts to an “out-of-control” state, in Scenario 4, Reactive Maintenance will be active. Since the “out-of-control” condition happens before the scheduled time, additional time and expense will be incurred to identify and solve the equipment problem.

The proposed model which consists of four different scenarios has been defined as follows (Fig. 1).

### 2.4. Economic design of integrated model

#### 2.4.1. Expected cycle time $E [T]$ of each scenario

Scenario 1 ( $S_1$ ), the process begins at an “in-control” state in which inspections start after  $h$  hours of monitoring to check whether or not the process has shifted from an “in-control” to “out-of-control” state. An alert signal is added in the control chart when maintenance should be performed before the scheduled time. But the false signal might occur sometimes when the process is still “in-control”, in this case searching and determining false signal, which consumes time and incur cost, will be carried out. Compensatory Maintenance is performed.

$$E[T|S_1] = h \sum_{i=0}^k iP_i^I(1 - F(ih)) + T_0 + T_C. \tag{12}$$

The components of the total time of scenario 1 include,

1. The interval time when process is in control, denoted by  $T_1$

$$E[T_1] = h \sum_{i=1}^k iP_i^I(1 - F(ih)) \tag{13}$$

2. The interval to search for the assignable cause, denoted by  $T_2$

$$E[T_2] = T_0 \tag{14}$$

3. The interval when the process is in Compensatory Maintenance, denoted by  $T_3$

$$E[T_3] = T_C \tag{15}$$

4. The total time of scenario 1 is

$$E[T|S_1] = E[T_1] + E[T_2] + E[T_3] \tag{16}$$

$$E[T|S_1] = h \sum_{i=1}^k iP_i^I(1 - F(ih)) + T_0 + T_C.$$

Scenario 2 ( $S_2$ ), it assumes that the process shifts to an “out-of-control” state prior to the planned maintenance, and the process failure mechanism follows a binomial distribution whereas the in-control time follows a truncated binomial distribution as Eqs. (10) and (11).

$$f(t|(k + 1)h) = \frac{f(t)}{F((k + 1)h)} = \frac{\lambda e^{-\lambda t}}{1 - e^{-(\lambda(k+1)h)}}; \theta < t \leq (k + 1)h \tag{17}$$

Then we have

$$E[T|S_2] = \int_0^{kh} tf(t|(k + 1)h)dt + hARL_0 - \tau + nE + T_A + T_R, \tag{18}$$

where  $\tau = \sum_{i=0}^k \int_{ih}^{(i+1)h} (t - ih)f(t|(k + 1)h)dt$ ,  $ARL_0$ : the average run length during out-of-control period, the  $ARL_0 = \frac{1}{\beta}$ , from Eq. (2) and application of the operating-characteristic function and average run length calculations are from Montgomery (2009), therefore  $\beta$  can be described as

$$\beta = P\{\hat{p} < UCL|p\} - P\{\hat{p} \leq LCL|p\} \tag{19}$$

$$\beta = P\{D < nUCL|p\} - P\{D \leq nLCL|p\}.$$

Since  $D$  is a binomial random variable with parameters  $n$  and  $p$ , the  $\beta$ -error defined by Eq. (19) can be obtained from the cumulative binomial (see Appendix). The components of the total time of scenario 2 include,

1. The interval when the process is in control, denoted by  $T_1$

$$E[T_1] = \int_0^{kh} tf(t|k + 1)h)dt \tag{20}$$

2. The interval when the process is out of control before the final sample of the detecting subgroup is taken, denoted by  $T_4$

$$E[T_4] = hARL_0 - \tau \tag{21}$$

3. The interval to sample, inspect, evaluate and plot the subgroup results, denoted by  $T_5$

$$E[T_5] = nE \tag{22}$$

4. The interval to search for the assignable cause, denoted by  $T_2$

$$E[T_2] = T_A \tag{23}$$

5. The interval that the process is in Reactive Maintenance, denoted by  $T_6$

$$E[T_6] = T_R \tag{24}$$

6. The total time of scenario 2 is

$$E[T|S_2] = E[T_1] + E[T_4] + E[T_5] + E[T_2] + E[T_6] \tag{25}$$

$$E[T|S_2] = \int_0^{kh} tf(t|(k + 1)h)dt + hARL_0 - \tau + nE + T_A + T_R.$$

Scenario 3 ( $S_3$ ), no signal occurs in the control chart before the scheduled time. Thus at the  $(k + 1)^{th}$  sampling interval, appropriate maintenance should be concerned. Since the process is always ‘in-control’, Planned Maintenance is performed. The components of the total time of scenario 3 include,

1. The interval when the process is in control, denoted by  $T_1$

$$E[T_1] = \int_0^{kh} tf(t|(k + 1)h)dt \tag{26}$$

2. The interval when the process is out of control and no signal is assigned in the chart, denoted by  $T_7$

$$E[T_7] = (k + 1)h - \int_0^{kh} tf(t|(k + 1)h)dt \tag{27}$$

3. The interval when the process is in Planned Maintenance, denoted by  $T_8$

$$E[T_8] = T_P \tag{28}$$

4. The total time of scenario 3 is

$$\begin{aligned} E[T|S_3] &= E[T_1] + E[T_7] + E[T_8] \\ E[T|S_3] &= \int_0^{kh} tf(t|(k + 1)h)dt(k + 1)h \\ &\quad - \int_0^{kh} tf(t|(k + 1)h)dt + T_P \\ &= (k + 1)h + T_P. \end{aligned} \tag{29}$$

Scenario 4 ( $S_4$ ), the process begins in control. When the process shifts to an ‘out-of-control’ state before the scheduled time, Reactive Maintenance will be active. This leads to additional time and expense which is used to identify and solve the equipment problem. The components of the total time scenario 4 include,

1. The interval when the process is in control, denoted by  $T_1$

$$E[T_1] = \int_0^{kh} tf(t|(k + 1)h)dt \tag{30}$$

2. The interval when the process is out of control and no signal is arranged, denoted by  $T_7$

$$E[T_7] = (k + 1)h - \int_0^{kh} tf(t|(k + 1)h)dt \tag{31}$$

3. The interval when the process is in Reactive Maintenance, denoted by  $T_6$

$$E[T_6] = T_R \tag{32}$$

4. The total time of scenario 4 is

$$\begin{aligned} E[T|S_4] &= E[T_1] + E[T_7] + E[T_6] \\ E[T|S_4] &= (k + 1)h + T_R. \end{aligned} \tag{33}$$

2.4.2. Expected cycle cost  $E [C]$  of each scenario

Similarly, the cycle cost consists of three main components including the cost of quality loss incurred while operating the process, the cost of sampling, and the cost of maintenance. The cost of quality loss includes both  $C_1$  and  $C_0$ , these two costs can be estimated using the equations given below,

1. The cost of interval when the process is in control, denoted by  $C_1$

$$E[C_1] = C_I \left[ h \sum_{i=1}^k p_i'(1 - F(ih)) + \gamma_c T_C \right] \tag{34}$$

2. The cost of sampling, inspection, evaluation and charting, denoted by  $C_2$

$$\begin{aligned} E[C_2] &= (C_F + nC_V) \sum_{i=0}^k ip_i'(1 - F(ih)) \quad \text{for scenario 1} \\ E[C_2] &= \frac{1}{h} E[T|S_2](C_F + nC_V) \quad \text{for scenario 2} \\ E[C_2] &= k(C_F + nC_V) \quad \text{for scenarios 3 to 4} \end{aligned} \tag{35}$$

3. The cost of false alarms, denoted by  $C_3$

$$E[C_3] = C_f \tag{36}$$

4. The cost of Compensatory Maintenance, denoted by  $C_4$

$$E[C_4] = C_C \tag{37}$$

5. The cost of interval occurred when the process is out of control, denoted by  $C_5$

$$E[C_5] = C_0 \left[ (k + 1)h - \int_0^{kh} tf(t|(k + 1)h)dt + \gamma_R T_R \right] \tag{38}$$

6. The cost of Reactive Maintenance, denoted by  $C_6$

$$E[C_6] = C_R \tag{39}$$

7. The cost of Planned Maintenance, denoted by  $C_7$

$$E[C_7] = C_P. \tag{40}$$

The total costs for 4 scenarios can be defined as below;

The total cost for scenario 1

$$\begin{aligned} E[C|S_1] &= E[C_1] + E[C_2] + E[C_3] + E[C_4] \\ E[C|S_1] &= C_I \left[ h \sum_{i=0}^k ip_i'(1 - F(ih)) + \gamma_c T_C \right] \\ &\quad + (C_F + nC_V) \sum_{i=0}^k ip_i'(1 - F(ih)) + C_f + C_C \end{aligned} \tag{41}$$

The total cost for scenario 2

$$\begin{aligned} E[C|S_2] &= E[C_1] + E[C_5] + E[C_2] + E[C_6] \\ E[C|S_2] &= C_I \left[ \int_0^{kh} tf(t|(k + 1)h)dt \right] \\ &\quad + C_0[(hARL_0 - \tau) + nE + \gamma_A T_A + \gamma_R T_R] \\ &\quad + \frac{1}{h} E[T|S_2](C_F + nC_V) + C_R \end{aligned} \tag{42}$$

The total cost for scenario 3

$$\begin{aligned} E[C|S_3] &= E[C_1] + E[C_2] + E[C_7] \\ E[C|S_3] &= C_I[(k + 1)h + \gamma_P T_P] + k(C_F + nC_V) + C_P \end{aligned} \tag{43}$$

The total cost for scenario 4

$$\begin{aligned} E[C|S_4] &= E[C_1] + E[C_5] + E[C_2] + E[C_6] \\ E[C|S_4] &= C_I \left[ \int_0^{kh} tf(t|(k + 1)h)dt \right] \\ &\quad + C_0[(k + 1)h - \int_0^{kh} tf(t|(k + 1)h)dt + \gamma_R T_R] \\ &\quad + k(C_F + nC_V) + C_R. \end{aligned} \tag{44}$$

2.4.3. Expected hourly cost  $E [H]$

In this section, determination of the hourly cost  $E [H]$  will be carried out. The model can be considered as a renewal-reward process; hence, the expected cost per hour  $E [H]$  can be expressed by

$$E[H] = \frac{E[C]}{E[T]} \tag{45}$$

where

$$E[T] = E[T|S_1]P(S_1) + E[T|S_2]P(S_2) + E[T|S_3]P(S_3) + E[T|S_4]P(S_4) \tag{46}$$

$$E[C] = E[C|S_1]P(S_1) + E[C|S_2]P(S_2) + E[C|S_3]P(S_3) + E[C|S_4]P(S_4). \tag{47}$$

and probability for 4 scenarios can be explained through conditional probability as below,

$$\begin{aligned} P(S_1) &= \sum_{i=1}^k P(\text{In-control and Alert Signal}) \\ &= \sum_{i=1}^k P(\text{In-control}|\text{Alert Signal})P(\text{Alert Signal}) \\ &= \sum_{i=1}^k P_i^l(1 - F(ih)). \end{aligned} \tag{48}$$

$$\begin{aligned} P(S_2) &= \sum_{i=1}^k P(\text{Out-of-control and Alert Signal}) \\ &= \sum_{i=1}^k P(\text{Out-of-control}|\text{Alert Signal})P(\text{Alert Signal}) \\ &= \sum_{i=1}^k [F(ih) - F(i-1)h] \left( 1 - \sum_{j=1}^{i-1} P_j^l \right) \sum_{l=1}^{k-i+1} P_l^0. \end{aligned} \tag{49}$$

$$\begin{aligned} P(S_3) &= \sum_{i=1}^k P(\text{In-control and No Signal}) \\ &= \sum_{i=1}^k P(\text{In-control}|\text{No Signal})P(\text{No Signal}) \\ &= (1 - F(kh)) - \sum_{i=1}^k P_i^l(1 - F(ih)). \end{aligned} \tag{50}$$

$$\begin{aligned} P(S_4) &= \sum_{i=1}^k P(\text{Out-of-control and No Signal}) \\ &= \sum_{i=1}^k P(\text{Out-of-control}|\text{No Signal})P(\text{No Signal}) \\ &= F(kh) - \sum_{i=1}^k [F(ih) - F(i-1)h] \left( 1 - \sum_{j=1}^{i-1} P_j^l \right) \sum_{l=1}^{k-i+1} P_l^0. \end{aligned} \tag{51}$$

3. Results and analysis

In this research, the numerical example and sensitivity analyses are conducted to study the effect of model parameters in the solution of economic design of the KS chart. Using the genetic algorithms (GA) with MATLAB, 7.6.0 (R2009a) software [The Math WorksTM, 2009](#), the solution procedure is carried out to obtain the optimal values of  $(n^*, h^*, L^*, k^*)$  which will be subsequently used to minimize in Eq. (45).

The GA is the stochastic and optimization search technique of natural selection and natural genetics. The GA solves problems used the approach to the process of Darwinian evolution. In recent years, many research have been devoted to the GA solves problems of economic-statistical, engineering, mathematics, production processes, etc. Current, GA models were introduced and investigated by [Holland \(1975\)](#). The solution procedures of GA (e.g., [Charongrattanasakul and Pongpullponsak, 2011](#); [Chou et al., 2006, 2008](#); [Lin et al., 2009, 2012](#); [Chen and Yeh, 2009](#); [Franco et al., 2012](#)), in this the research are briefly described below.

Step 1. *Initial Population*: The procedure starts at randomly generating 100 solutions that reach the constraint condition of individual test parameter. Meanwhile, the constraint condition represented for individual test parameter is set as below,

$$1 \leq n \leq 25, 0.1 \leq h \leq 5, 2 \leq L \leq 2.5, 20 \leq k \leq 40$$

Step 2. *Evaluation*: This step, is evaluated through the fitness function. Each solution used the expected cost per hour  $E [H]$  in Eq. (45).

Step 3. *Selection*: The selected function chooses parents (survivors) for the next generation based on their scaled values of the fitness scaling function. The four individual solutions are selected randomly and the best is chosen (For the first generation the chromosome with the lowest cost is selected to replace the highest cost chromosome).

Step 4. *Crossover*: In this step by step 3, generate new chromosomes for the next generation and a pair of parents (survivors) are selected randomly as shown in this example, parents (survivors) used for crossover operations to produce new chromosomes (or children) for the next generation. This research used crossover rate 0.8 as below,

$$D_1 = 0.8R + 0.2M, D_2 = 0.2R + 0.8M$$

where  $D_1$  is the first new chromosome,  $D_2$  is the second new chromosome, and  $R$  and  $M$  are the parent (survivors) chromosomes. If 30 parents (survivors) are randomly selected, then

Parameter	Value	Parameter	Value
$E$	0.1	$C_P$	3000
$C_P$	200	$T_A$	0.8
$C_F$	10	$T_A$	0.3
$C_F$	100	$C_I$	1
$C_C$	1000	$C_I$	10
$\lambda$	0.05	$T_0$	0.2
$C_V$	0.1	$T_C$	0.6
$C_R$	2000		

**Table 2** Eight model parameters and their level planning.

Model parameter	Level 1	Level 2
$C_I$	10	20
$C_0$	200	400
$C_0$	3000	6000
$C_R$	2000	4000
$C_C$	1000	2000
$C_F$	10	20
$C_V$	0.1	0.2
$C_f$	100	200

**Table 3** Model parameter assignment in the  $L_{16}$  orthogonal array and the corresponding solution.

Trial	Model parameter							
	$C_I$	$C_0$	$C_P$	$C_R$	$C_C$	$C_F$	$C_V$	$C_f$
1	10	200	3000	2000	1000	10	0.1	100
2	20	200	3000	2000	1000	20	0.2	200
3	10	400	3000	2000	2000	10	0.2	200
4	20	400	3000	2000	2000	20	0.1	100
5	20	200	6000	2000	1000	20	0.2	100
6	10	200	6000	4000	2000	10	0.2	100
7	20	400	6000	2000	1000	10	0.1	200
8	10	400	6000	2000	1000	10	0.2	200
9	20	200	3000	4000	1000	20	0.1	200
10	10	200	3000	4000	2000	20	0.2	100
11	20	400	3000	2000	1000	20	0.2	100
12	10	400	3000	4000	2000	10	0.1	200
13	10	200	6000	4000	2000	10	0.1	200
14	20	200	6000	4000	1000	20	0.1	100
15	10	400	6000	4000	2000	10	0.1	100
16	20	400	6000	4000	2000	20	0.2	200

Solution					
	$n$	$h$	$L$	$k$	$E [H]$
1	1.001	0.432	2.192	20.094	83.517
2	1.000	0.653	2.000	20.000	126.442
3	1.000	0.101	2.021	20.000	95.067
4	1.145	0.104	2.357	20.000	143.357
5	1.010	1.186	2.412	26.935	171.505
6	1.081	4.994	2.301	39.930	205.515
7	1.000	0.140	2.492	20.000	113.634
8	1.006	0.133	2.001	20.000	102.569
9	1.405	0.776	2.079	20.000	185.375
10	1.000	1.260	2.010	20.000	209.949
11	1.000	0.100	2.000	20.000	71.692
12	1.000	0.100	2.000	20.000	148.564
13	9.238	5.000	2.363	39.999	206.009
14	2.415	4.997	2.189	20.036	204.619
15	2.331	0.289	2.348	22.233	108.610
16	9.596	0.429	2.235	20.576	169.921

there are 60 children that will be produced. Thus, the population size increases to 90 (i.e., 30 parents (survivors) + 60 children) in this step.

Step 5. *Mutation*: Mutation function is the small change of genes in chromosomes in the population, suppose that the mutation rate is 0.1, which is also determined by  $D_2$  orthogonal array experiment.

**Table 4** Optimal values for model parameters.

Model parameter	Trial 8
$C_I$	10
$C_0$	400
$C_P$	6000
$C_R$	2000
$C_C$	1000
$C_F$	10
$C_V$	0.2
$C_f$	200

**Table 5** Optimal values for four variables and the optimal value of the total hourly costs.

Variable	Integrated model (KS)
$n^*$	1.006
$h^*$	0.133
$L^*$	2.001
$k^*$	20.000
$E [H]$	102.569

**Table 6** Nonparametric linear regression for four variables models ( $n, h, L, k$ ).

Parameter	Estimate	$T$	$P$ -values
Constant ( $n$ )	0.7492	15*	0.012
$E [H]$	0.0027		
Constant ( $h$ )	-1.0426	33*	0.012
$E [H]$	0.0122		
Constant ( $L$ )	1.9910	7.5*	0.012
$E [H]$	0.0012		
Constant ( $k$ )	19.6618	25*	0.012
$E [H]$	0.0034		

\* The Wilcoxon  $T$  statistics: reject  $H_0$  when  $T$  (calculate) >  $T$  (critical).

**Table 7** Nonparametric linear regression optimal values for four variables and the optimal value of the total hourly cost.

Variable	Integrated model (KS)
$n^*$	1.026
$h^*$	0.209
$L^*$	2.114
$k^*$	20.011
$E [H]$	102.569

In this example, we have 90 solutions and we can randomly select 9 chromosomes (i.e.,  $90 \times 0.1 = 9$ ) to mutate some parameters (or genes) in this step.

Step 6. *Stopping criteria*: Repeat Step 2 to Step 5 until the stopping criteria is found. In this example, we use “50 generations or greater than” as our stopping criteria.

**Table 8** Numerical results of four integrated models.

Variable	Integrated model (X-bar)	Integrated model (EWMA for four variables)	Integrated model (EWMA for six variables)	Integrated model (KS)
$n^*$	4.000	4.000	6.082	1.026
$h^*$	1.230	1.150	3.008	0.209
$L^*$	2.910	1.100	2.494	2.114
$k^*$	22.000	22.000	20.660	20.011
$E [H]$	158.32	153.020	194.640	102.569

\* Four and Six integrated models from [Charongrattanasakul and Pongpullponsak \(2009, 2011\)](#).

**Table 9** ARL performance of models.

ARL for the general	ARL for the KS
370	430.96

The initial values of the necessary parameters are given in [Tables 1 and 2](#), where eight independent parameters which will be tested in the sensitivity analysis and their corresponding level planning are illustrated. Accordingly, the effect of model parameters on the solution of economic design of the KS chart can be investigated by conducting numerical example and sensitivity analysis. For the sensitivity determination,  $L_{16}$  orthogonal-array experimental design shown in [Table 3](#) is used in the test.

After calculation, the best value parameters by  $L_{16}$  orthogonal array are obtained as shown in [Table 4](#). From [Table 5](#), the optimal values of the policy variables that minimize  $E[H]$  are found to be  $n^* = 1.006, h^* = 0.133, L^* = 2.001, k^* = 20.000$  and the corresponding hourly cost is  $E[H] = 102.569$ .

[Table 6](#) illustrates the output of the nonparametric linear regression applied in order to fit the regression line according to [Brown and Mood \(1951\)](#) where Wilcoxon Matched–Pair Test ([Corder and Foreman, 2009](#)) is used in hypothesis testing at 0.05 significance levels by the statistical software [SPSS 15.0](#). It is noticed that the sign of the coefficient parameter of constant is often estimated by assuming that the hourly cost  $E [H]$  is positive, which is consistent with the principle of non-parametric statistical hypothesis testing.

Using nonparametric linear regression test, the numerical results of the optimal values of the policy variables which minimize  $E[H]$  are found to be  $n^* = 1.026, h^* = 0.209, L^* = 2.114, k^* = 20.011$  and the corresponding hourly cost is  $E[H] = 102.569$  ([Table 7](#)). In [Table 8](#), the comparison between four policies of the integrated model (KS) and other models shows that the model proposed in this study has a better economic behavior when it is distribution-free.

Finally, from [Table 9](#), it can be seen that the performance for ARL of KS control chart, equals to 430.96, is greater than that of the control chart with the general model which is 370.

#### 4. Conclusion

This research proposes the method used in searching the appropriate products for the Kolmogorov–Smirnov KS control chart that is distribution-free. This control chart is suitable for detection of small changes. In some occasions, small changes in the process can lead to incredible damage. For this

reason, those factories who produce non-restricted goods using appropriate control charts would reduce or eliminate unnecessary cost. Conclusively, these four control charts shown in [Table 8](#) are suitable for using in various types of factories and products; however, selection of appropriate control chart is recommended in order to reduce unnecessary cost.

The numerical results, when sample size  $n \geq 20$  is used, are summarized in [Table 8](#). According to the optimal values of the policy variables that minimize  $E [H]$ , this provides evidence that when it is distribution-free the integrated KS model will have a better economic behavior than those previously reported models ([Charongrattanasakul and Pongpullponsak, 2009, 2011](#)). Besides, the results show that the performance for ARL of KS ([Table 9](#)), equals to 430.96, is greater than that of the ARL obtained from the general model which is equal to 370. However, the proposed model may have some defective in which investigating any weak point and improving for better performance would be an interesting issue for further study.

#### 5. Future work

For future work, it is of interest to develop integrated economic model and control chart for nonparametric multivariate which is distribution-free or follows an unknown distribution.

#### Acknowledgements

The authors would like to thank the referees for their constructive comments and suggestions. This research was financially supported by Department of Mathematics, King Mongkut’s University of Technology Thonburi.

#### Appendix A

##### Computation of ARL of the KS control chart

In Eqs. (1) and (2), application of the operating-characteristic function and ARL calculations from [Montgomery \(2009\)](#) is considered.

$$\begin{aligned} \beta &= P\{\hat{p} < UCL|p\} - P\{\hat{p} \leq LCL|p\} \\ &= P\{D < nUCL|p\} - P\{D \leq nLCL|p\}, \end{aligned} \tag{A1}$$

Since  $D$  is a binomial random variable with parameters  $n$  and  $p$ , the  $\beta$ -error defined in Eq. (A.1) can be obtained from the cumulative binomial distribution, where  $p$  is assumed to be  $P\{D_n > CL/\sqrt{n}\}$ , the edf in Eq. (2), in this case

$$\alpha = 1 - \beta. \tag{A2}$$



Then the *ARL* can be computed by Eq. (2).

$$ARL = 1/\alpha. \quad (A3)$$

Example in Eq. (A1) in calculation it is initially needed to generate the OC curve for a control chart for fraction nonconforming with parameters  $n = 20$ ,  $CL = 0.06$ ,  $LCL = 0$ , and  $UCL = 0.13$  by considering the concept from the Kolmogorov–Smirnov quality control chart. Using these parameters [3] in Eq. (A1), we obtain

$$\begin{aligned} \beta &= P\{\hat{p} < UCL|p\} - P\{\hat{p} \leq LCL|p\} \\ &= P\{\hat{p} < nUCL|p\} - P\{\hat{p} \leq nLCL|p\} \\ &= P\{D < (20)(0.13)|p\} - P\{D \leq (20)(0)|p\}. \end{aligned} \quad (A4)$$

Since  $D$  must be an integer, it is find that

$$\beta = P\{D < 2|p\} - P\{D \leq 0|p\} \quad (A5)$$

$$\begin{aligned} P(D < 2) &= \sum_{D=0}^2 \binom{20}{D} 0.0134^D (1 - 0.0134)^{20-D} \\ &\quad + \binom{20}{1} 0.0134^1 (1 - 0.0134)^{20-1} \\ &\quad + \binom{20}{2} 0.0134^2 (1 - 0.0134)^{20-2} \\ P(D < 2) &\approx 0.9977 \end{aligned} \quad (A6)$$

where  $\alpha = 1 - 0.9977$ , the  $ARL = 1/(1 - 0.9977) \approx 430.96\#$  (A7)

## References

- Bakir, S.T., 2012. A nonparametric shewhart-type quality control chart for monitoring broad changes in a process distribution. *Int. J. Qual. Stat. Reliab.*, 10
- Brown, G.W., Mood, A.M., 1951. On median tests for linear hypotheses. *Proc. Berkeley Sympos. Math. Stat. Probab.*, 159–166
- Charongrattanasakul, P., Pongpullponsak, A., 2009. Minimizing the cost of an integrated model by EWMA control chart. *Kasetsart J. (Nat. Sci.)* 43, 385–391.
- Charongrattanasakul, P., Pongpullponsak, A., 2011. Minimizing the cost of integrated systems approach to process control and maintenance model by EWMA control chart using genetic algorithm. *Expert Syst. Appl.* 38, 5178–5186.
- Chen, F.L., Yeh, C.H., 2009. Economic statistical design of non-uniform sampling scheme  $\bar{X}$  control charts under non-normality and Gamma shock using genetic algorithm. *Expert Syst. Appl.* 36, 9488–9497.
- Chou, C.Y., Chen, C.H., Chen, C.H., 2006. Economic design of variable sampling intervals  $T^2$  control charts using genetic algorithms. *Expert Syst. Appl.* 30, 233–242.
- Chou, C.Y., Cheng, J.C., Lai, W.T., 2008. Economic design of variable sampling intervals EWMA charts with sampling at fixed times using genetic algorithms. *Expert Syst. Appl.* 34, 419–426.
- Corder, G.W., Foreman, D.I., 2009. *Nonparametric Statistics for Non-Statisticians*. John Wiley & Sons.
- Duncan, A.J., 1956. The economic design of charts used to maintain current control process. *J. Am. Stat. Assoc.* 51 (274), 228–242.
- Franco, B.C., Costa, A.F.B., Machado, M.A.G., 2012. Economic-statistical design of the  $\bar{X}$  chart used to control a wandering process mean using genetic algorithm. *Expert Syst. Appl.* 39, 12961–12967.
- Gibbons, J.D., 1971. *Nonparametric Statistical Inference*. McGraw-Hill Inc.
- Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor, Univ. Michigan Press.
- Khruaesom, P., Pongpullponsak, A., 2014. Quality control chart based on the Kolmogorov-Smirnov structure. In: *Proceedings of International Conference on Applied Statistics (ICAS 2014)*, pp. 79–84 (Khon Kaen, Thailand).
- SPSS 15.0 for Windows Serial; License Number 5045778, Single License, SPSS Thailand.
- Lin, H.H., Chou, C.Y., Lai, W.T., 2009. Economic design of variable sampling intervals  $\bar{X}$  charts with A&L switching rule using genetic algorithms. *Expert Syst. Appl.* 36, 3048–3055.
- Lin, S.N., Chou, C.Y., Wang, S.L., Liu, H.R., 2012. Economic design of autoregressive moving average control chart using genetic algorithms. *Expert Syst. Appl.* 39, 1793–1798.
- Montgomery, D.C., 2009. *Introduction to Statistical Quality Control*, sixth ed. John Wiley & Sons, New York.
- Pollock, S.G., September 2007. *Topics in Econometrics. Hazard, Survival and Duration*, 1–2. Retrieved January, 2006, from Department of Economics, The University of Leicester. World Wide Web: <<http://www.le.ac.uk/users/dsgpl1/COURSES/TOPICS/hazard.pdf>>, accessed January 2015.
- Pongpullponsak, A., Jayathavaj, V., 2014. The new Hodges–Lehmann estimator control charting technique for the known process distributions. In: *Proceedings of International Conference on Applied Statistics (ICAS 2014)*, pp. 47–58 (Khon Kaen, Thailand).
- Pongpullponsak, A., Suracherkeiti, W., Panthong, C., 2009. The economic model of  $\bar{X}$  control chart using shewhart method for skewed distributions. *J. Thai Stat. Assoc.* 7 (1), 81–99.
- Saniga, E.M., 1977. Joint economically optimal design of  $\bar{X}$  and  $R$  control charts. *Manage. Sci.* 24 (4), 420–431.
- The Math WorksTM, 2009. *MATLAB 7.6.0 (R2009a)*. License Number 350306, February 12, 2009.
- Yang, S.F., Lin, J.S., Cheng, S.W., 2011. A new nonparametric EWMA sing control chart. *Expert Syst. Appl.* 38 (5), 6239–6243.
- Yu, F.J., Tsou, C.S., Huang, K.I., Wu, Z., 2010. An economic-statistical design of  $\bar{X}$  control charts with multiple assignable causes. *J. Qual.* 17 (4), 327–338.
- Zhou, W.H., Zhu, G.L., 2008. Economic design of integrated model of control chart and maintenance. *Math. Comput. Model.* 47, 1389–1395.