BAREKENG: Journal of Mathematics and Its Applications
September 2026 Volume 20 Issue 3 Page 1807-1820
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss3pp1807-1820

CONTROL LIMITS OF THE G CHART BASED ON FAST DOUBLE BOOTSTRAP WITH GENERALIZED KULLBACK-LEIBLER DIVERGENCE PARAMETER ESTIMATION

Muhammad Yahya Matdoan[1], **Muhammad Mashuri**[2*], **Muhammad Ahsan**[3]

[1,2,3]*Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember*
Kampus ITS-Sukolilo, Surabaya, 60111, Indonesia

[1]*Department of Statistics, Faculty of Science and Technology, Universitas Pattimura*
Jln. Ir. M. Putuhena, Kampus Unpatti-Poka, Ambon, 97233, Indonesia

*Corresponding author's e-mail: *m_mashuri@statistika.its.ac.id*

Article Info

Article History:

Received: 21st May 2025
Revised: 8th January 2026
Accepted: 16th March 2026
Available online: 8th April 2026

Keywords:

Control limit;
Fast double bootstrap;
g-chart;
Generalized Kulullback Leibler (GKL) divergence.

ABSTRACT

The g chart is a type of attribute control chart that is effective for monitoring processes with low defect rates. If the process parameters on the g chart are unknown, parameter estimation is performed. The most effective parameter estimation method for data contaminated with outliers is GKL divergence. This parameter estimation was developed to avoid the limitations of previous robust methods, namely the truncation method and the truncation method. However, in practice, the g chart developed from the GKL divergence parameter estimator has weaknesses, especially if there are no nonconforming items in the phase I sample, which causes a lack of sensitivity at the control limits. To overcome this problem, a bootstrap-based and double bootstrap-based control limit approach was developed. However, this approach requires high accuracy, a long time, and high computational costs. Therefore, the purpose of this study is to develop a g chart with fast double bootstrap-based control limits. The data used in this study were simulation data with contaminated and non-contaminated outliers and empirical data sourced from PT. X. regarding container weight measurements. This study found that the control limits of the g-chart based on fast double bootstrap were more sensitive than the conventional and bootstrap approaches. The results indicate that the container weighing process is still under control.

How to cite this article:

M. Y. Matdoan, M. Mashuri and M. Ahsan., "CONTROL LIMITS OF THE G CHART BASED ON FAST DOUBLE BOOTSTRAP WITH GENERALIZED KULLBACK-LEIBLER DIVERGENCE PARAMETER ESTIMATION", *BAREKENG: J. Math. & App.,* vol. 20, no. 3, pp. 1807-1820, Sep, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

The g chart is an attribute chart used to monitor the total number of occurrences. The main advantage of the g chart is its ability to monitor processes with very small or infrequent nonconforming occurrences (high quality), when compared to other attribute control charts such as the p, np, c, and u charts [1]. If the process parameters in the control chart are unknown, parameter estimation is required. The most efficient parameter estimation method for monitoring processes with data contaminated by outliers is Generalized Kullback Leibler (GKL) divergence parameter estimation. The GKL divergence estimation method was developed as a solution to the loss of information due to data truncation or cutting from previous robust parameter estimation methods, namely the truncation and memoryless property methods [2], [3], [4], [5], [6]. This method is efficient and robust, and is designed to avoid the limitations of previous methods by using an asymptotically efficient statistical minimum distance approach that is easy to apply to discrete distributions.

On the other hand, the use of control limits in the context of geometric distributions developed by Kaminsky [2], based on k times the standard deviation (kσ). This approach tends to increase the false alarm rate and causes the lower control limit (LCL) to be meaningless in geometric processes. As a solution to this problem, various researchers have developed control limit approaches based on probability limits, as done by Xie & Goh [7], Yang et al [8], L. Zhang et al [9], M. Zhang et al [10], B. J. Kim & Lee [11]. This approach allows for more accurate detection of process changes, especially in systems that follow a geometric distribution. However, this approach still has weaknesses, especially when there are no nonconforming items in the phase I sample, which causes a lack of sensitivity at the control limits [12]. To overcome this problem, it is necessary to develop bootstrap-based control limits [13]. The bootstrap method is a computational statistical technique used to estimate sample distributions when the distribution of statistical parameters is unknown [14]. This approach is carried out by repeatedly resampling from the original data set with replacement to obtain the sampling distribution of a parameter estimator [15], [16]. However, this method is not optimal for large shifts. Therefore, the double bootstrap method was developed to correct bias and improve efficiency. However, the double bootstrap approach requires high precision, a long time, and high computational costs. As a solution, a fast double bootstrap (FDB)-based control limit was developed to speed up processing time and reduce the computational load without sacrificing accuracy [14].

Based on the results of previous studies, there has been no development of a control chart using the GKL divergence method with FDB-based control limits to monitor nonconforming events. The results of this study are expected to be used to adjust control limits to be more accurate, both for non-contaminated and contaminated outliers.

2. RESEARCH METHODS

2.1 Data

The data used in this study were simulation data with contaminated and non-contaminated outlier conditions. In addition, empirical data obtained from PT. X regarding container weight measurements were used to test the developed statistical method.

2.2 g Chart

The g chart is a Shewhart-type control chart used to monitor the number of conforming cases between two consecutive nonconforming cases. This chart is one of the attribute control charts used to monitor the total number of occurrences in a process [2].

The advantage of the g chart is its ability to monitor processes with very few or rare nonconforming events (high quality). Suppose we are given data from a process divided into subgroups or samples of size n, denoted as $x_1, x_2, \dots, x_n$.. The events in this process are assumed to be independent and identically distributed following a geometric distribution under controlled or stable conditions. The statistics used to create this control chart are based on the total number of events. The random variable representing the total number of events is expressed by Eq. (1) below [3].

$$T = x_1 + x_2 + \cdots + x_n \tag{1}$$

then the mean and variance are obtained as follows.

$$E(T) = E\left(\sum_{i=1}^{n} x_i\right)$$
$$= n\left(\frac{1-p}{p} + a\right) \quad (2)$$
$$V(T) = V\left(\sum_{i=1}^{n} x_i\right)$$
$$= \frac{n(1-p)}{p^2} \quad (3)$$

This control chart can be used if the values of p and a are known, with $0 < p < 1$. The control limit of g chart can be modeled as follows.

$$\text{CL} = n\left(\frac{1-p}{p} + a\right) \quad (4)$$
$$\text{UCL} = n\left(\frac{1-p}{p} + a\right) + k\sqrt{\frac{n(1-p)}{p^2}} \quad (5)$$
$$\text{LCL} = n\left(\frac{1-p}{p} + a\right) - k\sqrt{\frac{n(1-p)}{p^2}} \quad (6)$$

According to Kaminsky *et al* [2], If the value (p) is unknown, it must be estimated using conventional parameter estimation methods, such as maximum likelihood [2] as expressed by Eq. (7) below.

$$\hat{p} = \frac{1}{\bar{\bar{x}} - a + 1} \quad (7)$$

Theoretically, suppose there are m subgroups (samples), each of size n, with the total number of events in each sample denoted as $t_1, t_2, \ldots, t_m$. Then, the average number of events per sample can be expressed as follows.

$$\bar{t} = \frac{t_1 + t_2 + \cdots + t_m}{m} \quad (8)$$

so that it is obtained

$$\bar{\bar{x}} = \frac{\bar{t}}{n} = \frac{1-\hat{p}}{\hat{p}} + a \quad (9)$$

Then, the control limit of g chart for unknown p (estimated by $\hat{p}$) is obtained, namely.

$$\text{CL} = \bar{t} \quad (10)$$
$$\text{UCL} = \bar{t} + k\sqrt{n\left(\frac{\bar{t}}{n} - a\right)\left(\frac{\bar{t}}{n} - a + 1\right)} \quad (11)$$
$$\text{LCL} = \bar{t} - k\sqrt{n\left(\frac{\bar{t}}{n} - a\right)\left(\frac{\bar{t}}{n} - a + 1\right)} \quad (12)$$

If the estimation of process parameters in the control chart is unknown, then the Maximum Likelihood Estimation method is used. However, MLE estimation has a weakness, namely that it tends to produce large biases when the sample size is small. To overcome this problem, robust parameter estimation such as truncation and generalized Kullback-Leibler divergence has been developed.

Truncation parameter estimation $(\hat{p}_t)$ is an estimation method that limits the sample space, so that only values within a certain range are considered [4]. This approach is applied to reduce the influence of outliers that can interfere with the accuracy of parameter estimation. The restriction is done by setting an upper limit (d) on the geometric distribution, so that the value of the random variable Y is always within a certain range, namely between a and d [4]. The truncation approach is formulated based on the following conditions.

$$d > 2\bar{Y} - a \quad (13)$$

with

d = upper limit of the data
$\bar{Y}$ = observed sample mean

$a =$ smallest value of the data

To determine the smallest positive value that satisfies the inequality, the upper bound value d^* is calculated as follows.

$$d^* = [2\bar{Y} - a] + 1 \tag{14}$$

If the parameter estimate $\hat{p}_t$ yields a value close to zero or negative, then the estimate will be recalculated using the value d^* to ensure that $\hat{p}_t > 0$ remains satisfied. This approach makes parameter estimation more stable and robust against outliers in the data. In the truncated geometric distribution, the Probability Mass Function (PMF) for random variable Y with success probability p and upper bound d is given by Eq. (15) below.

$$f(y) = \frac{p(1-p)^{y-a}}{1-(1-p)^{d-a+1}}, y = a, a+1 \ldots . d \tag{15}$$

In this case, the maximum likelihood estimation (MLE) of parameters for truncated geometric distributions does not have a closed-form solution, requiring a more practical alternative approach. As a solution for estimating parameter p, a modified method of moments (MM) can be used. This modification is formulated through Eq. (16) below.

$$\hat{p}_t = \frac{(a+d) - 2\bar{Y}}{(\bar{Y} - a + 1)(d - \bar{Y}) - S^2} \tag{16}$$

with

$\bar{Y}$: the sample mean, denoted by $\frac{1}{n}\sum_{i=1}^{n} Y_i$

S^2 : the sample variance, denoted by $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$

d : the upper limit obtained from $[2\bar{Y} - a] + 1$

After obtaining the parameter estimate $\hat{p}_t$ using the truncation parameter estimation, the control limits (CL, UCL, and LCL) for the g chart (total events) are set as follows [17].

$$\text{CL} = n_k\left(\frac{1-\hat{p}_t}{\hat{p}_t} + a\right) \tag{17}$$

$$\text{UCL} = n_k\left(\frac{1-\hat{p}_t}{\hat{p}_t} + a\right) + k\sqrt{\frac{n_k(1-\hat{p}_t)}{\hat{p}_t^2}} \tag{18}$$

$$\text{LCL} = n_k\left(\frac{1-\hat{p}_t}{\hat{p}_t} + a\right) - k\sqrt{\frac{n_k(1-\hat{p}_t)}{\hat{p}_t^2}} \tag{19}$$

If the LCL is smaller than the minimum value a, then the LCL value is adjusted to a, because the control limit cannot be lower than the minimum possible occurrence [4].

The GKL divergence method is one of the robust parameter estimation methods based on the statistically efficient minimum distance technique [18]. This method was then developed by Park et al [19] in the g chart to overcome information loss due to data truncation and clipping in the truncation estimation method. The main advantage of the g chart when using GKL divergence-based parameter estimation is that it performs better than other robust parameter estimation methods, both in contaminated data and non-contaminated outliers [19]. The GKL divergence concept is carried out by measuring the difference between two density functions $g(.)$ and $f(.)$, which depend on a tuning parameter $\tau \in [0,1]$. This GKL divergence is formulated as follows.

$$GKL_\tau(g,f) = \sum_{y \in S}\left[\frac{g(y)}{\bar{\tau}} log\left(\frac{g(y)}{f(y)}\right) - \left(\frac{g(y)}{\bar{\tau}} + \frac{f(y)}{\tau}\right) log\left(\tau\frac{g(y)}{f(y)} + \bar{\tau}\right)\right] \tag{20}$$

where S is the set of values (support) of the geometric random variable Y. Meanwhile, τ and $\bar{\tau} = 1 - \tau$ are scaling parameters that affect the divergence calculation.

$0 < \tau < 1$ is a tuning parameter that regulates the balance between two types of divergence, namely log likelihood and kullback-Leibler divergence.

This divergence is determined based on the following two limit cases. $\tau \to 0$,

GKL divergence acts as a log likelihood function used in MLE parameter estimation.

$$GKL_0(g,f) = \lim_{\tau \to 0} GKL_\tau(g,f) = \sum_{y \in S} g(y)\, log\left(\frac{g(y)}{f(y)}\right) \tag{21}$$

KL divergence from g to f. and $\tau \to 1$,

GKL Divergence changes to conventional KL Divergence.

$$GKL_1(g,f) = \lim_{\tau \to 1} GKL_\tau(g,f) = \sum_{y \in S} g(y)\, log\left(\frac{f(y)}{g(y)}\right) \tag{22}$$

KL divergence from f to g.

In this case, we choose g as the empirical Probability Mass Function (PMF) based on the sample $\{Y1, Y2,\ldots,Yn\}$, namely.

$$g(y) = \frac{1}{n}\sum_{i=1}^{n} \| (Y_i = y) \tag{23}$$

where $\| (Y_i = y)$ is an indicator function, equal to 1 if $Y_i = y$, and 0 otherwise.

Then the parameter p, which is the probability of success in the geometric distribution, is estimated by minimizing the GKL divergence.

$$\hat{p}_{gkl(\tau)} = \arg\min \mathrm{GKL}_\tau(g,f) \tag{24}$$

with theoretical PMF.

$$f(y) = p(1-p)^{y-a} \tag{25}$$

The purpose of this estimation is to obtain the parameter p that produces the best fit between the observed data and the geometric distribution. Furthermore, it should be noted that $GKL_0(g,f)$ is actually equivalent to the conventional log-likelihood function as in Eq. (26) below.

$$GKL_0(g,f) = \lim_{\tau \to 0} GKL_\tau(g,f) = \sum_{y \in S} g(y)\, log\left(\frac{g(y)}{f(y)}\right) \tag{26}$$

and $GKL_1(g,f)$ is equivalent to the conventional Kullback-Leibler (KL) divergence [5] as shown in Eq. (27).

$$GKL_1(g,f) = \lim_{\tau \to 1} GKL_\tau(g,f) = \sum_{y \in S} g(y)\, log\left(\frac{f(y)}{g(y)}\right) \tag{27}$$

It is known that conventional KL divergence-based estimators are robust, with a breakdown point of 50%, but their efficiency in small samples tends to be low. Conversely, estimators based on the log-likelihood function have full asymptotic efficiency [20], but lack robustness [21]. However, GKL combines the two main advantages of efficiency and robustness. As shown by Park and Basu [21], parameter estimation based on GKL has asymptotic efficiency and remains robust when $0 < \tau < 1$. Specifically, the g chart for the total number of events with n_k observations have the following control limits.

$$\mathrm{CL}(p) = n_k\left(\frac{1-p}{p} + a\right) \tag{28}$$

$$\mathrm{UCL}(p) = n_k\left(\frac{1-p}{p} + a\right) + k\sqrt{\frac{n_k(1-p)}{p^2}} \tag{29}$$

$$\mathrm{LCL}(p) = \max\left\{n_k a, n_k\left(\frac{1-p}{p} + a\right) + k\sqrt{\frac{n_k(1-p)}{p^2}}\right\} \tag{30}$$

where k = 3 is used for the American standard with a false alarm rate (FAR) of 0.27% and k = 3.09 is used for the british standard with a FAR of 0.20%. The proposed robust parameter estimation $\hat{P}_{gkl}(\tau)$ is based on the GKL divergence method and can be obtained from $\hat{p}_{gkl(\tau)} = \arg\min \mathrm{GKL}_{\tau}(g, f)$ by setting.

$$g(y) = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\|\left(X_{ij} \leq y\right) \text{ dan } f(y) = p(1-p)^{y-a}.\right. \tag{31}$$

Next, for the estimator $\hat{p}$ of the process parameter p, the control limits can be easily obtained by entering the estimate $\hat{p}$ into the control limits of the g chart.

2.3 Fast Double Bootstrap

Fast Double Bootstrap (FDB) is a method introduced by Davidson and MacKinnon [20] to improve the computational efficiency of the double bootstrap procedure without reducing its accuracy. This method is based on the assumption that the test statistics in the first and second stages of bootstrap are independent. With this assumption, the FDB procedure reduces the number of resamplings required compared to double bootstrap. This reduces computation time and costs while still producing the same level of accuracy. In the FDB method, it is assumed that $s(x_b^{**})$ is independent of $s(x_b^{*})$, so that each first-stage bootstrap sample is generated only once in the second-stage bootstrap process. After obtaining the first stage bootstrap data x_{b^*} and the test statistic value $s(x_b^{*})$, $b = 1,2,\ldots,B$. Next, for each first stage bootstrap sample, it is only generated once in the second stage bootstrap process, denoted as $x_{b^{**}}$ with $b = 1,2,\ldots,B$. From this set of second-stage bootstrap sample data, the test statistic value $s(x_b^{**})$ can be calculated for each $b = 1,2,\ldots,B$ [21]. The fast double bootstrap procedure is shown in Fig. 1 below.

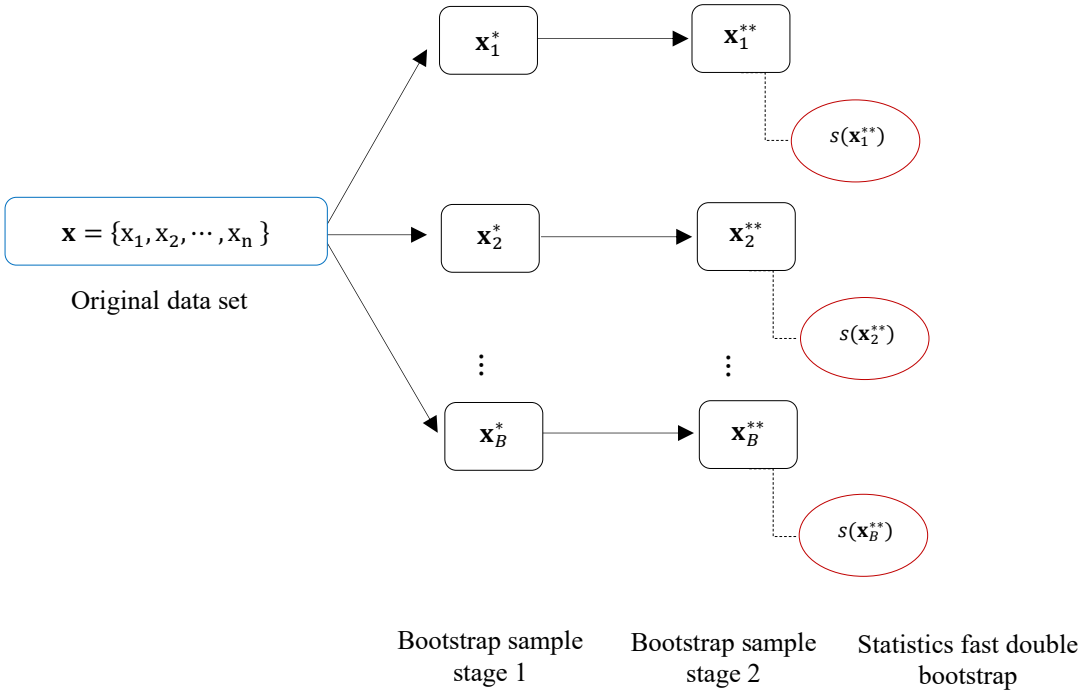

Figure 1. Fast Double Bootstrap Procedure

2.4 Research Procedures

This study develops control limits for g chart based using fast double bootstrap with Generalized Kullback-Leibler ($\hat{p}_{gkl}$) divergence parameter estimators, which are then compared with truncation parameter estimators ($\hat{p}_t$). The procedure in this study is as follows.

1. Determine the number of research subgroups.
2. Simulate data X_1 distributed geometrically with a number of subgroups ($m = 30,\ 50, 100$) with parameter p (0.01, 0.1, 0.2). as p_{clean} (non-contaminated outliers).
3. Simulate geometric distributed data X_2 as 5%, 10%, and 20% of data X_1 with parameters as p_{out}.

4. Combining simulation data X_1 and X_2 so that clean data and outlier data are mixed (outlier contamination).
5. Take a bootstrap sample of size B from the original data: $B: N_1^*, N_2^*, \ldots, N_B^*, B = 1000$..
6. Based on the results of bootstrap stage 1, then in bootstrap stage 2 (fast double bootstrap).
   a. For the first-stage bootstrap sample, perform one resampling to generate a second-stage bootstrap data set.
   b. Calculate statistics from the fast double bootstrap sample $s(\mathbf{x}_B^{**})$.
7. Based on Step 6, next calculate the generalized Kullback-Leibler divergence parameter estimate $(\hat{p}_{gkl})$ and the truncation parameter estimate $(\hat{p}_t)$.
8. Determine the value of k.
9. Set control limits (CL, UCL, and LCL) on the g chart.
10. Create a g chart based on the geometric distribution.

3. RESULTS AND DISCUSSION

This study was conducted using simulation study data to estimate Generalized Kullback-Leibler $(\hat{p}_{gkl})$ divergence parameters, then determining the control limits of the g chart based on conventional methods, bootstrap, and fast double bootstrap. Next, sensitivity testing was performed between the developed control limits and existing control limits to determine the number of events or observations that fell into the out-of-control (OOC) category. Subsequently, the developed g chart was applied to monitor container weighing services at PT. X.

3.1 Parameter Estimation

Parameter estimation in this study was conducted through a simulation study to determine the performance of g chart using the Generalized Kullback-Leibler $(\hat{p}_{gkl})$ divergence and truncation $(\hat{p}_t)$ parameter estimation methods. The results of parameter estimation are shown in Table 1.

Table 1. Parameter Estimation Results

Parameter estimation	m	p	Non-contamintaion		Outliers 5%		Outliers 10%		Outliers 20%	
			$\hat{p}$	Bias	$\hat{p}$	Bias	$\hat{p}$	Bias	$\hat{p}$	Bias
Truncation	30	0.01	0.154	0.1325	0.134	0.134	0.132	0.1225	0.124	0.1225
		0.1	0.153	0.1252	0.142	0.122	0.125	0.1152	0.112	0.1352
		0.2	0.147	0.1306	0.137	0.137	0.130	0.1306	0.137	0.1206
	50	0.01	0.134	0.1425	0.124	0.144	0.142	0.1325	0.124	0.1325
		0.1	0.142	0.1352	0.132	0.122	0.135	0.1352	0.132	0.1252
		0.2	0.147	0.1306	0.147	0.127	0.120	0.1206	0.127	0.1306
	100	0.01	0.154	0.1245	0.154	0.134	0.142	0.1325	0.134	0.1325
		0.1	0.122	0.1222	0.152	0.122	0.135	0.1352	0.122	0.1452
		0.2	0.127	0.1336	0.137	0.137	0.140	0.1206	0.117	0.1206
Generalized Kullback Leibler Divergence	30	0.01	0.144	0.1245	0.134	0.134	0.142	0.1325	0.124	0.1325
		0.1	0.122	0.1222	0.122	0.142	0.135	0.1252	0.132	0.1352
		0.2	0.127	0.1336	0.157	0.127	0.120	0.1206	0.137	0.1206
	50	0.01	0.134	0.1245	0.144	0.534	0.142	0.1325	0.124	0.1425
		0.1	0.142	0.1222	0.152	0.122	0.125	0.1252	0.132	0.1252
		0.2	0.157	0.1336	0.167	0.127	0.140	0.1306	0.147	0.1106
		0.01	**0.144**	**0.1215**	**0.124**	**0.104**	**0.112**	**0.1125**	**0.124**	**0.1125**
	100	0.1	0.132	0.1232	0.132	0.112	0.125	0.1052	0.132	0.1252
		0.2	0.127	0.1356	0.137	0.117	0.120	0.1006	0.127	0.1206

Based on Table 1, it can be seen that the parameter estimation results using the number of subgroups (m = 30, 50, 100) with parameters (p = 0.01, 0.1, 0. 2) under conditions of non-contaminated outliers and contaminated outliers (Δ = 5%, 10%, and 20%), shows that the parameter estimation $\hat{p}_{gkl}$ has a smaller bias value when compared to the parameter estimation $\hat{p}_t$ under both contaminated and non-contaminated outlier conditions. These findings also show that the larger the number of subgroups, the better the estimation of the generalized Kullback-Leibler divergence parameter $(\hat{p}_{gkl})$.

3.2 Determination of Control Limits in g Chart

After determining the best parameter estimator $\hat{p}_{gkl}$, the next step is to determine the control chart limits based on that parameter estimator. Control limits are determined to test the sensitivity of the developed g chart, using conventional, bootstrap, and fast double bootstrap methods. The results of determining the control limits of the g chart for non-contaminated outlier data conditions are presented in Table 2 below.

Table 2. Control Limits of the g Chart Based on Conventional, Bootstrap, and Fast Double Bootstrap Methods

Probability	Control limits	Non-contamination outliers					
		$m = 30$		$m = 50$		$m = 100$	
		UCL	LCL	UCL	LCL	UCL	LCL
p = 0.01	Conventional	1674	0	3657	0	5485	0
	Bootstrap	1534	0	3527	0	5450	0
	Fast Double Bootstrap	1517	0	3501	0	5430	0
p = 0.1	Conventional	248	0	374	0	317	0
	Bootstrap	224	0	361	0	312	0
	Fast Double Bootstrap	229	0	360	0	311	0
p = 0.2	Conventional	92	0	178	0	224	0
	Bootstrap	83	0	173	0	215	0
	Fast Double Bootstrap	81	0	177	0	214	0

Based on Table 2 in the simulation study results of no contamination outliers, it shows that the control limits of the g chart based fast double bootstrap are more sensitive, the number of subgroups (m = 30, 50, 100), when compared to other control limit methods. Furthermore, the results of determining control limits for 10% contamination outlier data conditions are presented in Table 3 below.

Table 3. Control Limits of the g Chart Based on Conventional, Bootstrap, and Fast Double Bootstrap Methods

Probability	Control limits	Contamination outliers 10%					
		$m = 30$		$m = 50$		$m = 100$	
		UCL	LCL	UCL	LCL	UCL	LCL
p = 0.01	Conventional	4892	0	3015	0	6598	0
	Bootstrap	4461	0	2902	0	6509	0
	Fast Double Bootstrap	4374	0	2909	0	6501	0
p = 0.1	Conventional	289	0	334	0	444	0
	Bootstrap	259	0	326	0	437	0
	Fast Double Bootstrap	256	0	323	0	437	0
p = 0.2	Conventional	60	0	150	0	247	0
	Bootstrap	54	0	145	0	242	0
	Fast Double Bootstrap	53	0	143	0	241	0

Based on Table 3, the results of the simulation study of 10% outlier contamination data show that the control limits of the g chart based fast double bootstrap are more sensitive to the number of subgroups (m = 30, 50, 100) compared to other control limit methods. Furthermore, the results of determining the control limits for the 20% outlier contamination data condition are presented in Table 4 below.

Table 4. Control Limits of the g Chart Based on Conventional, Bootstrap, and Fast Double Bootstrap Methods

Probability	Control limits	Contamination outliers 20%					
		$m = 30$		$m = 50$		$m = 100$	
		UCL	LCL	UCL	LCL	UCL	LCL
p = 0.01	Conventional	3229	0	8242	0	7222	0
	Bootstrap	2952	0	8004	0	7092	0
	Fast Double Bootstrap	2939	0	8005	0	7065	0
p = 0.1	Conventional	197	0	180	0	552	0
	Bootstrap	174	0	175	0	542	0
	Fast Double Bootstrap	162	0	174	0	529	0

Probability	Control limits	Contamination outliers 20%					
		m = 30		m = 50		m = 100	
		UCL	LCL	UCL	LCL	UCL	LCL
p = 0.2	Conventional	142	0	283	0	238	0
	Bootstrap	131	0	269	0	235	0
	Fast Double Bootstrap	128	0	268	0	235	0

Based on Table 4, the results of the simulation study of 20% outlier contamination data show that the control limits of the g chart based fast double bootstrap are more sensitive to the number of subgroups (m = 30, 50, 100) when compared to other control limit methods.

3.3 Comparison of Sensitivity in the Control Limits of the g Chart

Comparison of the sensitivity of control limits using conventional, bootstrap, and fast double bootstrap methods to determine the number of events that are out of control. The sensitivity comparison in this study was conducted using 30 and 100 subgroups, with contaminated and non-contaminated outlier data conditions. The results of the sensitivity comparison of control limits for the number of subgroups m = 30 and p = 0.01 are shown in Table 5 below.

Table 5. Comparison of the Number of Out-of-Control Points with m =30 and p = 0.01

Parameter Estimation	Control limits	Number out of control		
		Non-Contamination	Outliers 10%	Outliers 20%
Truncation	Conventional	1	1	2
	Bootstrap	1	3	3
	Fast Double Bootstrap	1	4	4
Generalized Kullback Leibler Divergence	Conventional	1	1	2
	Bootstrap	2	4	4
	Fast Double Bootstrap	2	4	5

Based on Table 5, it shows that the number of out-of-control points for non-contaminated outlier data on the g chart using conventional control limits is 1 out of control, bootstrap is 1 out of control, and fast double bootstrap is 1 out of control. Furthermore, for 10% contaminated outlier data, it shows that conventional control limits have 1 out of control, bootstrap has 3 out of control, and fast double bootstrap has 4 out of control. Furthermore, for 20% contaminated data, it shows that conventional control limits are 2 out of control, bootstrap is 3 out of control, and fast double bootstrap is 4 out of control. These findings show that fast double bootstrap-based control limits are more sensitive when compared to other control limit approaches. This is because the fast double bootstrap-based control limits produce more out-of-control events compared to other control limits, both in contaminated and non-contaminated outlier conditions. Furthermore, the results of the comparison of the sensitivity of control limits for the number of subgroups m = 100 and p = 0.01 are shown in Table 6 below.

Table 6. Comparison of the Number of Out-of-Control Points with m =100 and p = 0.01

Parameter Estimation	Control limits	Number out of control		
		Non-Contamination	Contamination 10%	Contamination 20%
Truncation	Conventional	1	1	2
	Bootstrap	1	2	3
	Fast Double Bootstrap	2	4	4
Generalized Kullback Leibler Divergence	Conventional	1	1	3
	Bootstrap	1	4	4
	Fast Double Bootstrap	2	4	5

Based on Table 6, it shows that the number of out-of-control points for non-contaminated outlier data on the g chart using conventional control limits is 3 out of control, bootstrap is 3 out of control, and fast double bootstrap is 3 out of control. Furthermore, for the 10% contaminated outlier data, it shows that the conventional control limits are 4 out of control, bootstrap is 4 out of control, and fast double bootstrap is 5 out of control. Furthermore, for 20% contaminated data, it shows that conventional control limits are 3 out of

control, bootstrap is 4 out of control, and fast double bootstrap is 5 out of control. These findings show that the fast double bootstrap-based control limits are more sensitive when compared to other control limit approaches. This is because the fast double bootstrap-based control limits produce more out-of-control events when compared to other control limits, both in contaminated and non-contaminated outlier conditions.

3.4 Empirical Study

Maritime transport plays an important role in supporting global economic growth, with around 90% of world trade carried out by sea. Every day, millions of containers move from one port to another. However, there are still major challenges related to shipping safety, especially in container transport. Recognizing the importance of this issue, the International Maritime Organization (IMO) established amendments to the 1972 International Convention for the Safety of Life at Sea (SOLAS), which is an important regulation and a key foundation for maintaining international shipping safety. This amendment requires every container to be loaded onto a ship to have a Verified Gross Mass (VGM) document, which is enforced globally.

The container terminal under PT. X plays a strategic role in both national and international logistics flows. One of the important activities at the TPS is the receiving process, where each container must undergo physical weighing to ensure the accuracy of the VGM document. Container VGM documents, which include the total weight of cargo, packaging, and containers that have been accurately verified, are crucial data in container terminal operations. VGM is not only used for service cost calculations, but also as a guarantee of shipping safety.

The implementation of VGM at TPS serves as an effort to mitigate the risk of accidents on land and at sea. Some of the main benefits of implementing VGM include improving ship safety, preventing accidents at ports, and reducing the risk of economic losses. However, in practice, there are often discrepancies between the physical weighing results at the TPS and the weight data listed on the Container Announcement Message (COPARN) document from the service user. These discrepancies not only have a financial impact, as they can affect the calculation of service costs and company profits, but also have a serious impact on shipping safety due to the risk of ship load imbalance.

Even though container terminals use high-quality measuring instruments, inaccuracies in recording container weights between TPS and Shippers in VGM documents still frequently occur. This has led to complaints from service users because it results in additional costs that must be paid for every kilogram of difference. Technical measures such as routine calibration, maintenance, and procurement of new equipment can help, but they are not entirely effective. Therefore, statistical quality control is needed to monitor the stability of the container weighing process and increase the trust of TPS service users. In this context, the control chart was chosen because it is suitable for analyzing the frequency of inaccuracies (number of incidents/differences) that follow a geometric distribution, thus complementing the limitations of the technical approach even though the measuring instruments used are of high quality.

Table 7. Number of Itmes Until a Nonconforming Item

Defect Sequence	Number of Items	Defect Sequence	Number of Items	Defect Sequence	Number of Items
1.	**229**	22.	**307**	43.	**297**
2.	**361**	23.	**430**	44.	**422**
3.	**483**	24.	**441**	45.	**644**
4.	**515**	25.	**408**	46.	**671**
5.	**443**	26.	**419**	47.	**624**
6.	**86**	27.	**152**	48.	**129**
7.	**185**	28.	**286**	49.	**208**
8.	**399**	29.	**392**	50.	**307**
9.	**362**	30.	**346**	51.	**408**
10.	**493**	31.	**447**	52.	**620**
11.	**456**	32.	**473**	53.	**650**
12.	**491**	33.	**452**	54.	**394**
13.	**109**	34.	**141**	55.	**120**
14.	**301**	35.	**345**	56.	**296**
15.	**368**	36.	**362**	57.	**443**
16.	**397**	37.	**401**	58.	**530**

Defect Sequence	Number of Items	Defect Sequence	Number of Items	Defect Sequence	Number of Items
17.	**411**	38.	**477**	59.	**431**
18.	**454**	39.	**498**	60.	**547**
19.	**430**	40.	**334**	61.	**530**
20.	**130**	41.	**62**	62.	**166**
21.	**267**	42.	**273**		

The research variable in the application of the g chart in VGM services at PT. X is to monitor the number of containers weighed until the first discrepancy is found between the weight stated in the VGM document and the weighing results carried out by the service user. This variable was chosen because it is in line with the characteristics of the geometric distribution in the g chart, where the focus of observation is on the number of container trucks monitored until a discrepancy in container weight between PT. X and the service user arises.

Sampling in this study was conducted by PT. X over a three-month period, from June to August 2025. The samples collected were the differences between the gross weight of the containers listed on the VGM documents reported by the shipping line and the results of reweighing conducted at the TPS. Every container entering the TPS must undergo a gross weight verification process. The initial data was obtained from the VGM documents submitted by the shipping line. Meanwhile, the comparative data came from the reweighing results using official weighing facilities at PT. X. The difference between the two data sets was then recorded as the research sample value to be analyzed further. The data structure used in this study presented the container weight for each observation unit. Based on the calculation results using the parameter estimator $\hat{p}_{gkl}$, the control limits for the g chart results were then determined as follows.

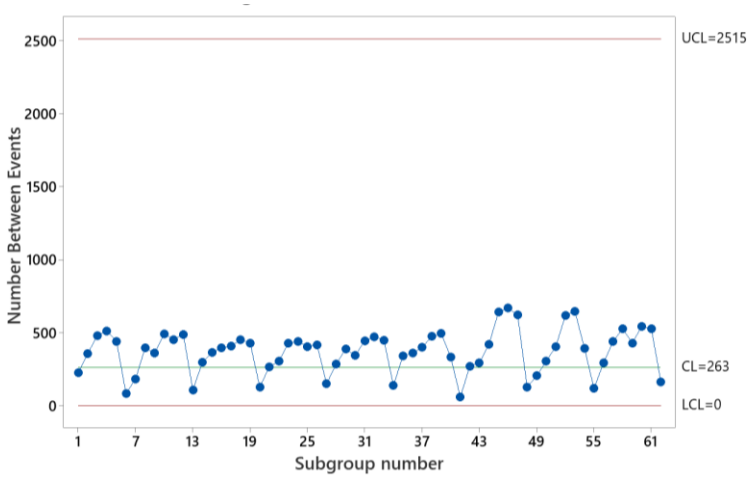

Figure 2. g Chart with Control Limits Using the Fast Double Bootstrap Method

Based on Fig. 2, it shows that container weighing by PT. X with service users is still under control. This study successfully evaluated the effectiveness of quality control in maintaining container weighing service quality standards. In general, the findings in this study indicate that the fast double bootstrap-based control limits have advantages over other control limits, both in non-contaminated and contaminated outlier conditions.

4. CONCLUSION

Estimation of generalized Kullback-Leibler divergence parameters $(\hat{p}_{gkl})$ ⟦shows better performance compared to truncation parameter estimation $(\hat{p}_t)$, both in conditions contaminated with outliers and non-contaminated with outliers. The control limits of the g chart set using the fast double bootstrap method proved to be more sensitive than the existing control limits, both in conditions with contaminated outliers and without. The results of monitoring container weighing at PT. X show that the process is still under control. Further research is recommended for the development of multivariate g charts and testing them in various sectors such as health, the environment, and so on.

Author Contributions

Muhammad Yahya Matdoan: Conceptualization, Methodology, Software Development, Visualization, and Writing of the Original Draft. Muhammad Mashuri: Conceptualization, Methodology, Validation, Supervision, and Writing – Review and Editing. Muhammad Ahsan: Conceptualization, Methodology, Validation, Supervision, and Writing – Review and Editing. All authors had approved the final version and discussed the results and contributed to the final manuscript.

Funding Statement

The authors did not receive any specific financial assistance from institutions in the public, corporate, or non-profit sectors for this study.

Acknowledgment

The authors would like to express their sincere gratitude to the Education Fund Management Institution (LPDP) and the Higher Education Financing Center (BPPT) under the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia. Special thanks also go to academic advisors, data providers, and colleagues for their insightful advice and constructive feedback.

Declarations

The authors confirm that there are no conflicts of interest related to the publication of this manuscript. Since the research did not involve human participants or animal testing, ethical approval and participation consent are not required. All authors have reviewed and approved the final manuscript and have agreed to its publication. Research data can be accessed by contacting the corresponding author, subject to reasonable request. The specific roles and responsibilities of each author are described in the Author Contributions section.

Declaration of Generative AI and AI-assisted Technologies

Generative AI tools were used solely for language refinement (grammar, spelling, and clarity). The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. The authors reviewed and approved all final text.

REFERENCES

[1] N. Chukhrova and A. Johannssen, "IMPROVED CONTROL CHARTS FOR FRACTION NON-CONFORMING BASED ON HYPERGEOMETERIKC DISTRIBUTION," *Comput Ind Eng*, vol. 128, pp. 795–806, Feb. 2019, doi: https://doi.org/10.1016/j.cie.2018.12.066

[2] F. C. Kaminsky, J. C. Benneyan, R. D. Davis, and R. J. Burke, "STATISTICAL CONTROL CHARTS BASED ON A GEOMETERIKC DISTRIBUTION," *Journal of Quality Technology*, vol. 24, no. 2, pp. 63–69, Apr. 1992, doi: https://doi.org/10.1080/00224065.1992.12015229.

[3] J. C. Benneyan, "NUMBER-BETWEEN G-TYPE STATISTICAL QUALITY CONTROL CHARTS FOR MONITORING ADVERSE EVENTS," Kluwer Academic Publishers, 2001.doi: https://doi.org/10.1023/A:1011846412909

[4] C. Park and M. Wang, "A STUDY ON THE G AND H CONTROL CHARTS," *Commun Stat Theory Methods*, vol. 52, no. 20, pp. 7334–7349, 2023, doi: https://doi.org/10.1080/03610926.2022.2044492.

[5] C. Park, M. Wang, and L. Ouyang, "NOVEL ROBUST G AND H CHARTS USING THE GENERALIZED KULLBACK–LEIBLER DIVERGENCE," *Comput Ind Eng*, vol. 176, Feb. 2023, doi: https://doi.org/10.1016/j.cie.2022.108951.

[6] C. Park and A. Basu, "THE GENERALIZED KULLBACK-LEIBLER DIVERGENCE AND ROBUST INFERENCE," *J Stat Comput Simul*, vol. 73, no. 5, pp. 311–332, May 2003, doi: https://doi.org/10.1080/0094965021000033477.

[7] V. J. Yohai, "OPTIMAL ROBUST ESTIMATES USING THE KULLBACK-LEIBLER DIVERGENCE," *Stat Probab Lett*, vol. 78, no. 13, pp. 1811–1816, Sep. 2008, doi: https://doi.org/10.1016/j.spl.2008.01.042.

[8] X. J. Zhou, D. K. J. Lin, X. Hu, and T. Jiang, "ROBUST PARAMETER DESIGN BASED ON KULLBACK-LEIBLER DIVERGENCE," *Comput Ind Eng*, vol. 135, pp. 913–921, Sep. 2019, doi: https://doi.org/10.1016/j.cie.2019.06.053.

[9] A. Bakdi, W. Bounoua, S. Mekhilef, and L. M. Halabi, "NONPARAMETRIC KULLBACK-DIVERGENCE-PCA FOR INTELLIGENT MISMATCH DETECTION AND POWER QUALITY MONITORING IN GRID-CONNECTED ROOFTOP PV," *Energy*, vol. 189, Dec. 2019, doi: https://doi.org/10.1016/j.energy.2019.116366.

[10] M. Xie and T. N. Goh, "THE USE OF PROBABILITY LIMITS FOR PROCESS CONTROL BASED ON GEOMETERIKC DISTRIBUTION."

[11] Z. Yang, M. Xie, V. Kuralmani, and K. L. Tsui, "ON THE PERFORMANCE OF GEOMETERIKC CHARTS WITH ESTIMATED CONTROL LIMITS," *Journal of Quality Technology*, vol. 34, no. 4, pp. 448–458, 2002, doi: https://doi.org/10.1080/00224065.2002.11980176.

[12] L. Zhang, K. Govindaraju, M. Bebbington, and C. D. Lai, "ON THE STATISTICAL DESIGN OF GEOMETERIKC CONTROL CHARTS," *Qual Technol Quant Manag*, vol. 1, no. 2, pp. 233–243, Jan. 2004, doi: https://doi.org/10.1080/16843703.2004.11673075.
[13] M. Zhang, Y. Peng, A. Schuh, F. M. Megahed, and W. H. Woodall, "GEOMETERIKC CHARTS WITH ESTIMATED CONTROL LIMITS," Mar. 2013. doi: https://doi.org/10.1002/qre.1304.
[14] B. J. Kim and J. Lee, "ADJUSTMENT OF CONTROL LIMITS FOR GEOMETERIKC CHARTS," *Commun Stat Appl Methods*, vol. 22, no. 5, pp. 519–530, Sep. 2015, doi: https://doi.org/10.5351/CSAM.2015.22.5.519.
[15] M. J. Zhao and A. R. Driscoll, "THE C-CHART WITH BOOTSTRAP ADJUSTED CONTROL LIMITS TO IMPROVE CONDITIONAL PERFORMANCE," *Qual Reliab Eng Int*, vol. 32, no. 8, pp. 2871–2881, Dec. 2016, doi: https://doi.org/10.1002/qre.1971.
[16] M. Kim and J. Lee, "GEOMETERIKC CHARTS WITH BOOTSTRAP-BASED CONTROL LIMITS USING THE BAYES ESTIMATOR," *Commun Stat Appl Methods*, vol. 27, no. 1, pp. 65–77, 2020, doi: https://doi.org/10.29220/CSAM.2020.27.1.065.
[17] Bradley. Efron and Robert. Tibshirani, *AN INTRODUCTION TO THE BOOTSTRAP*. Chapman & Hall/CRC, 1998.
[18] M. S. Lola, N. H. Zainuddin, M. N. A. Ramlee, and H. Sofyan, "DOUBLE BOOTSTRAP CONTROL CHART FOR MONITORING SUKUK VOLATILITY AT BURSA MALAYSIA," *J Teknol*, vol. 79, no. 6, pp. 149–157, Sep. 2017, doi: https://doi.org/10.11113/jt.v79.10410.
[19] R. Beran, "PREPIVOTING TEST STATISTICS: A BOOTSTRAP VIEW OF ASYMPTOTIC REFINEMENTS," 1988.doi: https://doi.org/10.1080/01621459.1988.10478649
[20] R. Davidson and J. G. MacKinnon, "THE POWER OF BOOTSTRAP AND ASYMPTOTIC TESTS," *J Econom*, vol. 133, no. 2, pp. 421–441, Aug. 2006, doi: https://doi.org/10.1016/j.jeconom.2005.06.002.
[21] R. Davidson and A. Monticini, "AN IMPROVED FAST DOUBLE BOOTSTRAP," 2023.