LEANDRO ALVES DA SILVA

Shewhart Control Chart to monitor the mean of a Discrete Weibull distribution

São Paulo

2024

LEANDRO ALVES DA SILVA

Shewhart Control Chart to monitor the mean of a Discrete Weibull distribution

Original version

Thesis submitted to the Polytechnic School of the University of São Paulo as requirement for the degree of Doctor of Science.

Area of Concentration: Production Engineering

Supervisor: Prof. PhD. Linda Lee Ho

São Paulo

2024

Silva, Leandro Alves da
Shewhart Control Chart to monitor the mean of a Discrete Weibull distribution / LEANDRO ALVES DA SILVA. – São Paulo, 2024-
45 p. : il. (some color.) ; 30 cm.

Supervisor: Prof. PhD. Linda Lee Ho

Thesis (Doctoral) – University of São Paulo – USP
Polytechnic School
Postgraduate Program in Production Engineering, 2024.

1. Discrete Weibull. 2. Control Charts. 3. Sum Distribution. 4. Failure Rate. 5. Average Run Length. I. Polytechnic School of the University of São Paulo II. Production Engineering.

CDU XX:XX:XXX.X

Acknowledgements

My profound gratitude to Professor PhD Linda Lee Ho, for all the support, dedication, encouragement, and guidance throughout the journey and development of this thesis, as well as her mentorship in my growth as a researcher and educator.

To Professor PhD Roberto da Costa Quinino for his steadfast cooperation, exceptional support, and insightful advice in the development of the paper that constitutes an integral part of this research.

To Professor PhD Frederico Rodrigues Borges da Cruz and Professor PhD Marcelo Bourguignon Pereira, for their valuable suggestions and constructive criticisms provided during the Qualifying Examination.

Special thanks to my wife, Claudilene, for her unyielding encouragement, support, and understanding during the time I dedicated to this thesis.

I extend my gratitude to my family and friends for their abundant words of encouragement and unwavering confidence in my ability to complete this project.

To God, always, simply for the gift of LIFE.

"It is the man who carefully advances step by step...who is bound to succeed in the greatest degree."

(Alexander Graham Bell)

Resumo

SILVA, L. A. **Shewhart Control Chart to monitor the mean of a Discrete Weibull distribution** . 2024. Thesis (Doctoral) - Polytechnic School, University of São Paulo, São Paulo, 2024.

Geralmente, o tempo de falha é modelado por uma distribuição contínua, como as distribuições Weibull ou Gamma, por exemplo. Em cenários práticos, a coleta de dados frequentemente envolve o registro de informações em termos de contagens discretas, como o número de dias, ciclos, etc. Assim, a Distribuição Weibull Discreta é utilizada para modelar tais casos. Nesta tese, propõe-se um gráfico de controle Shewhart $\overline{X}$ para monitorar a média de um processo Weibull Discreto. Embora a distribuição da soma de variáveis aleatórias Weibull Discretas não apresente uma forma fechada, sua determinação é possível por meio de um procedimento com uso de Cadeia de Markov, resultando na obtenção de limites de controle exatos. O *average run length* (ARL) é uma métrica para avaliar o desempenho do gráfico de controle, e regras suplementares de execução são incluídas para aprimorá-lo. Um exemplo numérico ilustra a aplicação do gráfico de controle proposto.

Palavras-chaves: Weibull discreta. Gráficos de controle. Distribuição de soma. Taxa de falha. ARL.

Abstract

SILVA, L. A. **Shewhart Control Chart to monitor the mean of a Discrete Weibull distribution** . 2024. Thesis (Doctoral) - Polytechnic School, University of São Paulo, São Paulo, 2024.

Usually, failure time is modeled by a continuous distribution as the Weibull or Gamma distributions, for example. In practical scenarios, data collection often involves recording information in terms of discrete counts, such as the number of days, cycles, etc. Thus, the Discrete Weibull distribution is used to model such cases. In this thesis, Shewhart $\overline{X}$ control chart is proposed to monitor the mean of a Discrete Weibull process. Although the distribution of the sum of Discrete Weibull random variables does not present a closed form, its determination is possible by a Markov Chain procedure, which leads to getting the exact control limits. The Average Run Length (ARL) is a metric for evaluating control chart performance and supplementary run rules are included to improve it. A numerical example illustrates its application.

Key-words: Discrete Weibull. Control Charts. Sum Distribution. Failure Rate. Average Run Length.

List of Figures

List of Tables

Contents

1 Introduction

In several studies focused on the analysis of failure time or the lifetime of systems, equipment, and components, the conventional approach entails the collection of continuous time measurements. Within these contexts, the Weibull distribution is extensively utilized (Abernethy, 2006).

However, in numerous instances, the documentation of these failures is conducted post-occurrence. Consequently, the management and recording of information related to these events are aptly addressed by considering discrete measures of time or counts of the number of events within a defined time interval, be it per hour, per day, per month, per year, or any other unit, particularly when the unit is frequently replaced, regardless of its age (Nakagawa, 1984).

To model such types of counts it is common the use of discrete distributions such as Poisson, Negative Binomial, and Geometric as discrete alternative for the Exponential and Gamma distributions. The utilization of discrete distributions is primarily motivated by several factors: their well-established distribution properties concerning the sum of independent and identically distributed random variables, availability of closed-form solutions, and robust goodness-of-fit in practical situations. Despite these advantages, the literature often neglects the exploration of the sum of Discrete Weibull random variables. Remarkably, the characteristic of interest demonstrates a favorable fit to this distribution when compared to other discrete alternatives. A plausible explanation for this situation is the absence of a closed-form expression for the sum of independent Weibull random variables, whether discrete or continuous.

To model such type of counts it is common the use of discrete distributions such as Poisson, negative binomial and geometric as discrete alternative for the Exponential and Gamma distributions.

Motivated by the disseminated application of the Weibull distribution, Nakagawa and Osaki (1975) introduced the initial ideas of a discrete distribution to correspond to a continuous-time Weibull distribution (which is known as type I Discrete Weibull). They (Nakagawa; Osaki, 1975) exhibit this discrete distribution, considering that in many practical cases the measurement of failure data is available as discrete data, such as number of days, cycles, revolutions, blows, shocks among others. Hereon $X \sim DW(q, \beta)$ denotes that the random variable X follows a discrete Weibull distribution of parameters q, β.

Several practical examples can be cited, such as fatigue lives for samples of Alloy T7987 measured in thousands of cycles (Meeker; Escobar; Pascual, 2022), the number of sessions required for pain relief in patients with chronic nonspecific low back pain (Silva et al., 2017), and the utilization of control charts (by individual observations of time between events) (Ali et al., 2020) for monitoring the outbreak of dengue fever (Khan et al., 2019).

Regarding the control charts, Shewhart is a pioneer in its introduction, which stands out as the primary tool in Statistical Process Control (SPC) methodology. It is extensively employed to monitor the stability of various process parameters, including the mean, variance, nonconforming fraction, number of defectives among others.

Among the Shewhart control charts, the $\overline{X}$ control chart stands out as the most widely used, attributed to its simplicity, easy implementation, and straightforward result interpretation. Numerous strategies and procedures have been proposed to improve its performance, and the incorporation of supplementary run rules emerges as a viable alternative in such scenarios.

The emergence of such an opportunity, coupled with the significance of the theme in both academic and practical contexts where studies on failure time analysis or the lifetime of systems, equipment, and components are imperative, constituted motivating factors for this research. The primary objective of this study is to address the following research questions:

- **How to develop an $\overline{X}$ control chart to monitor the stability of the process mean when $X \sim DW(q, \beta)$, taking into account that the events occur between periods with a sample size $n > 1$?**
- **Can the inclusion of supplementary rules improve the performance of an $\overline{X}$ control chart to monitor the stability of the process mean when $X \sim DW(q, \beta)$?**

The main contribution of this thesis is to present a study regards to the development of $\overline{X}$ control chart employed in monitoring the stability of $E(X)$ when $X \sim DW(q, \beta)$, taking account that the events occur between periods with a sample size $n > 1$. Additionally, the inclusion of the supplementary run rules (Klein, 2000; Khoo; Ariffin, 2006) is proposed to improve its performance. It is worth mentioning that the analyses made in this thesis regard to the type I Discrete Weibull (Nakagawa; Osaki, 1975).

Thus, this thesis is organized as follows. The introduction for the description of the problem, a literature review as well as a brief review of the Discrete Weibull is presented in Chapter 1. To build the $\overline{X}$ distribution, there is the necessity to get the distribution of $\sum_{i=1}^{n} X_i$. This is the subject of Chapter 2. The performance of the control charts is discussed in Chapter 3. Additionally, in Chapter 4 supplementary run rules are introduced. An example of the application of the chart proposed is shown in Chapter 5 and the final remarks in Chapter 6.

It is relevant to comment that the results of this thesis are partially presented in Brazilian conference (see Appendix C), as well as a paper accepted and available online in an international journal (see Appendix D).

1.1 Objective

The primary objective of this thesis is to explore the application of the Discrete Weibull distribution in statistical process control. The central focus of the study involves the development of an $\overline{X}$ control chart for monitoring the stability of $E(X)$ when $X \sim DW(q, \beta)$, considering events occurring between periods with a sample size $n > 1$. Furthermore, the study proposes the inclusion of supplementary run rules (Klein, 2000; Khoo; Ariffin, 2006) to improve the performance of the control chart. It is noteworthy that the analysis conducted in this thesis regards the type I Discrete Weibull (Nakagawa; Osaki, 1975).

1.2 Literature Review

In the literature, diverse versions of the Discrete Weibull distribution have been proposed (Almalki; Nadarajah, 2014). The first model, known as type I Discrete Weibull, was introduced by Nakagawa and Osaki (1975). They (Nakagawa; Osaki, 1975) aimed to establish a discrete distribution corresponding to a continuous-time Weibull distribution. Another version, named type II Discrete Weibull, was presented by Stein and Dattero in 1984 (Stein; Dattero, 1984). A third version, referred to as type III Discrete Weibull, was proposed by Padgett and Spurrier (Padgett; Spurrier, 1985). Other researchers have contributed to the study of the Discrete Weibull distribution. For instance, Szymkowiak and Iwi'nska (Szymkowiak; Iwińska, 2016) explored characterizations in terms of discrete aging intensity.

In a recent investigation by Valadares et al. (2023), an evaluation of repairable systems employing minimal repair was carried out. The study involved the adoption of the type I Discrete Weibull distribution instead of the Continuous Weibull distribution. The findings indicated comparable levels of complexity between both Weibull models. However, notable advantages were observed in the utilization of the discrete model. Specifically, these benefits included a reduced standard deviation of the parameter associated with the system's deterioration and a lower Akaike's information criterion.

In monitoring the quality of the lifetime of systems, control charts have been used for this aim. In Statistical Process Control (SPC), the control charts are useful and powerful tools to monitor the measurement of the quality characteristics of interest, they are employed on several areas and types of process.

These quality characteristics can have a continuous or discrete nature and the control charts can be by variables or attributes. The traditional control charts designed to monitor continuous data are the $\overline{X}$ and S^2 employed in monitoring the process mean and variance, respectively, while in case of counting data the p and np attribute control charts are commonly employed for proportion and number of defectives, respectively.

On account of current advanced technology, increasingly there are process of high quality level, high performance and low defects rate. As a result of the traditional control charts may presented problems in some scenarios. In these cases the called Time-Between-Events (TBE) control charts are used, as well as indicated for its name, this control chart monitor the time interval between events, presenting itself greater effectiveness to deal with cases of high yield production (Xie et al., 2002). Whereas the quality characteristic is discrete, it is suitable the use of TBE discrete version, named as the cumulative conforming control (CCC) charts.

For this reason to monitor the quality of the lifetime of systems other control charts have been used. They have been proposed as in Chan, Xie and Goh (2000), which used in this work the exponential distribution, afterward it was generalized by Zhang et al. (2007) using the gamma distribution, in addition Calvin (1983), Goh (1987), Xie et al. (1999), Chan, Xie and Goh (1997) made in their studies proposals of the CCC charts based on the geometric distribution. Recently Ali et al. (2020) proposed a cumulative conforming control (CCC) chart, in which was used a Discrete Weibull distribution and average run length (ARL) as one of the metrics for its performance. These control charts are most effective in the case of high-throughput production with a low defect rate, and in the current study it is intended to discuss situations where the defect rate can be higher.

1.2.1 Discrete Weibull - a brief review

Nakagawa and Osaki (1975) proposed a discrete distribution in order to correspond to a continuous-time Weibull distribution. The main motivation is that in many practical cases the measurement of failure data may be presented by integer numbers (as discrete/count data), such as the number of days, hours, cycles, blows, shocks or revolutions. In their proposal two parameters q and β are considered. Thus, they defined the discrete Weibull distribution $\{P_x\}_{x=0}^{\infty}$ as $\sum_{j=x}^{\infty} P_j(q,\beta) = (q)^{x^\beta}$ where $x = 0, 1, 2, \ldots; \beta > 0$ and $0 < q < 1$ which results a probability mass function (PMF) expressed as

$$P_x(q,\beta) = (q)^{x^\beta} - (q)^{(x+1)^\beta}, \; x = 0, 1, 2, ...\beta > 0, \quad 0 < q < 1 \tag{1.1}$$

Note that if $\beta = 1$, the equation (1.1) becomes a geometric distribution. Its cumulative distribution function (CDF) (1.2) and failure rate function (1.3) can be written respectively as:

$$F(x) = 1 - q^{(x+1)^{\beta}} \tag{1.2}$$

$$r_x(q, \beta) = \frac{P_x}{\sum\limits_{j=x}^{\infty} P_j} = 1 - (q)^{(x+1)^{\beta} - x^{\beta}} \tag{1.3}$$

Khan, Khalique and Abouammoh (1989) mention that its first and two moments are respectively expressed as:

$$E\{X\} = \sum_{x=1}^{\infty} q^{x^{\beta}} \tag{1.4}$$

$$E\{X^2\} = 2\sum_{x=1}^{\infty} x q^{(x+1)^{\beta}} + E\{X\} \tag{1.5}$$

Therefore, based on the equations (1.4) and (1.5) an expression to compute the variance of the Discrete Weibull distribution can be obtained by:

$$Var\{X\} = \left(2\sum_{x=1}^{\infty} x q^{(x+1)^{\beta}} + E\{X\}\right) - \left(\sum_{x=1}^{\infty} q^{x^{\beta}}\right)^2. \tag{1.6}$$

In this thesis it is considered the zero include in the counting, in order to analyse scenarios where there is not event observed in the sample.

In Figure 1, the plots of the PMF for the Discrete Weibull (type I), for $q = 0.7$ and $\beta = 0.5$, 1.0, 1.5, 2.0, and 2.5, are shown. This distribution is flexible and includes increasing and decreasing ratio failures, such as the continuous Weibull distribution. The dashed lines indicate the trend, whereas the markers indicate the discrete values.

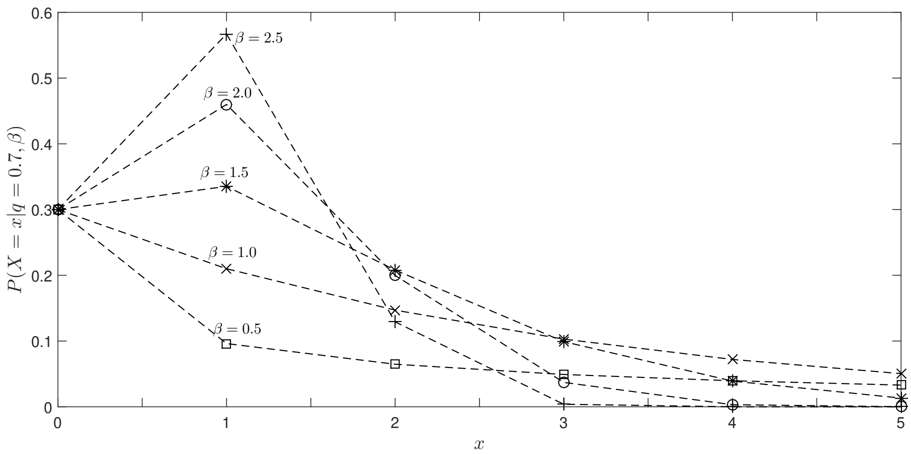


Figure 1 – Probability Function of X for $q = 0.7$ and $\beta = 0.5, 1.0, 1.5, 2.0, 2.5$. Source: Author

According to Vila, Nakano and Saulo (2019), there are two possibilities to get monotone decreasing rates in a Discrete Weibull distribution:

1 When $\beta < 1$ and $0 < q < 1$

2 And when $1 < \beta < \delta^* = \dfrac{1}{(1 - \log 2)}$ and $\dfrac{1}{2} < q \leq \delta(\beta) = \min\left(r, \dfrac{\exp(1-\beta)}{\beta}\right)$ where $r \approx 0.5006658$ is the unique root of the function $g(x) = x^{2^{\delta^*}} - 2x + 1$, with $x \in (0, 1)$.

3 Otherwise, we have cases with increasing failure rates.

As mentioned before, the first proposal about the Discrete Weibull distribution was performed by Nakagawa and Osaki (1975) and it is known as type I Discrete Weibull distribution.

This distribution coincides (and present itself similar) to geometric distribution when $\beta = 1$(Khan; Khalique; Abouammoh, 1989).

Khan, Khalique and Abouammoh (1989) and Barbiero (2016), Barbiero (2017) have contributed about the estimation of the parameters of the Discrete Weibull. Khan, Khalique and Abouammoh (1989) noted the following equivalence between the parameters of the Continuous Weibull with parameters (β, α) and the type I Discrete Weibull with distribution with parameters (β, q):

- β (in continuous Weibull) $\Leftrightarrow$ β (in type I)
- $e^{-\alpha}$ (in continuous Weibull) $\Leftrightarrow$ q (in type I Discrete Weibull)

The focus of this thesis is on type I Discrete Weibull.

2 The $\overline{X}$ control chart for Discrete Weibull distribution

In this thesis $\overline{X}$ control chart is developed to monitor the stability of $E(X)$ when $X \sim DW(q, \beta)$, taking account that the events occur between periods with a sample size $n > 1$. For this, let $X_1, X_2, \ldots X_n$, be the results of a random sample of size n of a Discrete Weibull distribution.

To build the $\overline{X}$ control chart, the determination of the distribution of $Y = \sum_{i=1}^{n} X_i$ is needed which can be obtained as

$$P(Y = y) = P\left(\sum_{i=0}^{n} X_i = y\right) = \sum_{i_1=0}^{x_1} \sum_{i_2=0}^{y-x_1} \sum_{i_3=0}^{y-x_1-x_2} \cdots \sum_{i_n=0}^{y-x_1-x_2-\ldots-x_{n-1}} \tag{2.1}$$

$$P(X_1 = i_1)P(X_2 = i_2)P(X_3 = i_3)\ldots P(X_n = i_n)$$

When X_i follows a Discrete Weibull distribution, the expression (2.1) does not present a closed form, but its obtaining can be computationally reached. Thus, a function namely **FuncDistrSum** in R is built for this aim. Details of the inputs and outputs of this function are described in the Appendix B. Despite of the practicality provided by the new function to compute the exact sum distribution of Discrete Weibull being a good alternative to obtain the sum distribution for $n > 1$, it is necessary a considerable computational capability and processing time to compute the expression (2.1), especially for large n. After several trials, it is concluded that even for sample sizes smaller than 5, large computational times (as more than 50 hours) are required. To solve this barrier, this problem is tackled in two different methods to get the distribution of the equation (2.1). The first way is by a Markov Chain approach and the second by Monte Carlo simulations and then determine also the control limits. The first one is more precise, but the second approach allows a better comprehension by the managers for the quality control. Hence the two approaches can be discussed in implementation stage of statistical process control. These are subjects of the next two sections.

2.1 Determination of the distribution of $Y = \sum_{i=1}^{n} X_i$ by Markov Chain approach

Let $\mathbf{Q}$ be the transition probability matrix of a Markov chain described by the states $E = \{0, 1, \ldots, T-1\}$ with $T = k \times n + 1$; n is the sample size and k, an integer such that satisfies $P(X \leq k) \approx 1$ (and $P(X > k) \approx 0$) with $X \sim DW(q, \beta)$. The state $E = k_1$ means that $Y_t = X_1 + X_2 + \ldots + X_t = k_1$ and the state $E = 0$ symbolizes the beginning of the process. Let $Q(i, j)$ be an element of the matrix $\mathbf{Q}$ at i-th row and j-th column which contains the conditional transition probability of the event: $Q(k_1 + 1, k_2 + 1) = P\left(Y_{t+1} = \sum_{m=1}^{t+1} X_m = k_2 | Y_t = \sum_{m=1}^{t} X_m = k_1\right)$, for $t > 1$, thus, the elements of the matrix $\mathbf{Q}$ for $t = 1$ are:

$$
\mathbf{Q} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \cdots \\ T = k \times n \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & \ldots & T = k \times n \end{array} \\
\begin{pmatrix}
P(X=0) & P(X=1) & P(X=2) & P(X=3) & P(X=4) & \ldots & 1 - \sum_{x=0}^{k \times n-1} P(X=x) \\
0 & P(X=0) & P(X=1) & P(X=2) & P(X=3) & \ldots & 1 - \sum_{x=0}^{k \times n-1} P(X=x) \\
0 & 0 & P(X=0) & P(X=1) & P(X=2) & \ldots & 1 - \sum_{x=0}^{k \times n-1} P(X=x) \\
0 & 0 & 0 & P(X=0) & P(X=1) & \ldots & 1 - \sum_{x=0}^{k \times n-1} P(X=x) \\
0 & 0 & 0 & 0 & P(X=0) & \ldots & 1 - \sum_{x=0}^{k \times n-1} P(X=x) \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & 0 & \ldots & 1
\end{pmatrix}
\end{array}
$$

and after the replacement of the probabilities in (1.1) they become

$$
\mathbf{Q} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \cdots \\ T = k \times n \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & \ldots & T = k \times n \end{array} \\
\begin{pmatrix}
1-(q)^{(1)^\beta} & (q)^{(1)^\beta}-(q)^{(2)^\beta} & (q)^{(2)^\beta}-(q)^{(3)^\beta} & (q)^{(3)^\beta}-(q)^{(4)^\beta} & (q)^{(4)^\beta}-(q)^{(5)^\beta} & \ldots & (q)^{(k \times n+1)^\beta} \\
0 & 1-(q)^{(1)^\beta} & (q)^{(1)^\beta}-(q)^{(2)^\beta} & (q)^{(2)^\beta}-(q)^{(3)^\beta} & (q)^{(3)^\beta}-(q)^{(4)^\beta} & \ldots & (q)^{(k \times n+1)^\beta} \\
0 & 0 & 1-(q)^{(1)^\beta} & (q)^{(1)^\beta}-(q)^{(2)^\beta} & (q)^{(2)^\beta}-(q)^{(3)^\beta} & \ldots & (q)^{(k \times n+1)^\beta} \\
0 & 0 & 0 & 1-(q)^{(1)^\beta} & (q)^{(1)^\beta}-(q)^{(2)^\beta} & \ldots & (q)^{(k \times n+1)^\beta} \\
0 & 0 & 0 & 0 & 1-(q)^{(1)^\beta} & \ldots & (q)^{(k \times n+1)^\beta} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & 0 & \ldots & 1
\end{pmatrix}
\end{array}
$$

Let $\mathbf{b} = (1, 0, 0, \ldots 0)$ a vector of dimension T, then $\mathbf{v} = \mathbf{b}\mathbf{Q}^n$ is the vector of the probability of $Y_n = \{0, 1, 2, \ldots, k \times n\}$ with $Y_n = \sum_{m=1}^{n} X_m$, that is, $P(Y_n = k_2)$ is the $(k_2 + 1)$-th element of the vector $\mathbf{v}$.

For a better visualization for the readers, let us consider $X \sim DW(q = 0.7, \beta = 1.5)$. If it is used $n = 5$, $k = 30$, and the program shown in Appendix A for the Markov Chain approach, it is obtained the PMF of Y presented in Figure 2. The dash lines are to indicate the trend and the markers indicate the discrete values. The choice of $k = 30$ points out that numerically it can consider $P(Y \leq k \times n) \approx 1$.

Fixed a type I error equal α, and consider that there is interest to detect increases in $E(X)$, find an integer U such that $P(Y_n > U) \approx \alpha$ then the upper control limit is $UCL_m = \dfrac{U}{n}$. In case of interest to detect decreases in $E(X)$ find an integer L such that $P(Y_n < L) \approx \alpha$ then the lower control limit is $LCL_m = \dfrac{L}{n}$. Alternatively in case of bilateral shifts, find the integers U^* and L^* such that $P(Y_n > U^*) \approx P(Y_n < L^*) \approx \alpha/2$ then the lower and upper control limits are respectively $LCL_m^* = \dfrac{L^*}{n}, UCL_m^* = \dfrac{U^*}{n}$.

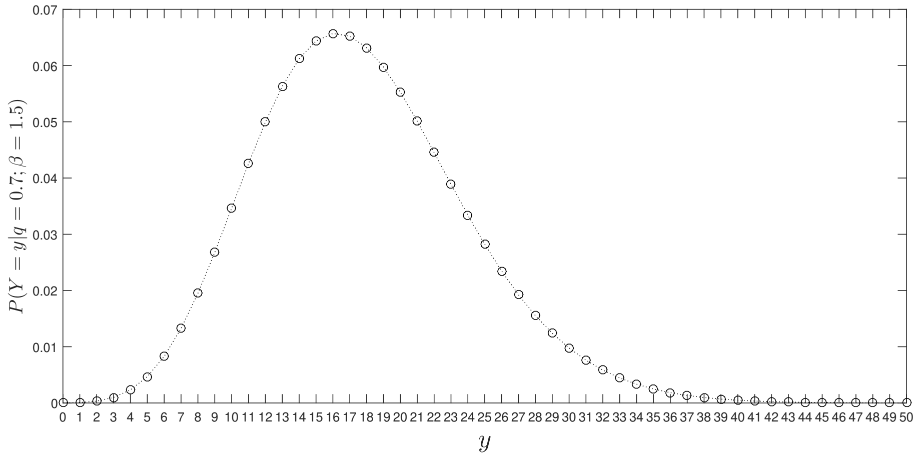


Figure 2 – Probability Function of Y for $q = 0.7$, $\beta = 1.5$ and $n = 5$. Source: Author

2.2 Determination of the distribution of $Y = \sum_{i=1}^{n} X_i$ by Monte Carlo Simulation

In this section, the algorithm used to determine the distribution of $Y = \sum_{i=1}^{n} X_i$ by Monte Carlo simulation is described. Basically it follows these steps:

- Step 1: Let $x_1, x_2, \ldots x_n$ be the observations of a random sample of size n from $X \sim DW(q, \beta)$.
- Step 2: Calculate $Y = \sum_{i=1}^{n} X_i$ and repeat Step 1 B times.
- Step 3: Fixed a type I error equal α and let consider that there is interest to detect increases in $E(X)$. Find the quantile $q_{1-\alpha}$ of the empirical distribution of Y such that $P(Y > q_{1-\alpha}) \approx \alpha$.
- Step 4: The Upper Control Limit is $UCL_s = \frac{q_{1-\alpha}}{n}$

If there is interest to detect decreases in $E(X)$, replace steps 3 and 4 by

- Step 3A: Fixed a type I error equal α find the quantile q_α of the empirical distribution of Y such that $P(Y < q_\alpha) \approx \alpha$.
- Step 4A: The Lower Control Limit is $LCL_s = \frac{q_\alpha}{n}$

Alternatively in case of bilateral shifts, replace steps 3 and 4 by

- Step 3B: Fixed a type I error equal α find the quantiles $q_{\alpha/2}$ and $q_{1-\alpha/2}$ of the empirical distribution of Y such that $P(Y < q_{\alpha/2}) \approx \alpha/2$ and $P(Y < q_{1-\alpha/2}) \approx 1 - \alpha/2$.

- Step 4B: The Lower and Upper Control Limits are respectively $LCL_s^* = \frac{q_{\alpha/2}}{n}$ and $UCL_s^* = \frac{q_{1-\alpha/2}}{n}$

In Appendix A, the program developed for this aim is included if the readers desire to run it, the simulation provides very close results to its. One may claim why do not use the asymptotic control limits as

- Step 4: $UCL_a = E(X) + z_{1-\alpha}\sqrt{\frac{Var(X)}{n}}$
- Step 4A: $LCL_a = E(X) - z_{1-\alpha}\sqrt{\frac{Var(X)}{n}}$
- Step 4B: $LCL_a^* = E(X) - z_{1-\alpha/2}\sqrt{\frac{Var(X)}{n}}$ and $UCL_a^* = E(X) + z_{1-\alpha/2}\sqrt{\frac{Var(X)}{n}}$

directly considering the normal approximations as stated by the Central Limit Theorem with $z_{1-\alpha}$ the quantile of the standard normal distribution such that $P(Z < z_{1-\alpha}) = 1 - \alpha$. The main reason for not using such approximation is it may not promote good results. This affirmation can be confirmed in Tables of the next chapter.

3 The performance of the $\overline{X}$ control chart for Discrete Weibull Sum Distribution

In this chapter the performance of $\overline{X}$ is evaluated. Among the metrics that have been adopted to assess the performance of the control charts, an alternative is the average run length (ARL), which expresses the average number of samples until an indication of an out of control condition (Montgomery, 2020).

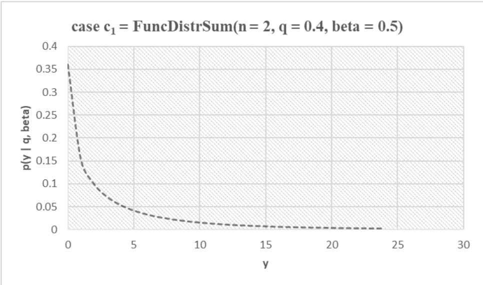
First of all, six cases ($c_1, c_2, c_3, c_4, c_5, c_6$) were chosen to analyze their features and verify their performance. The parameters and their respective mean and variance are in Table 1. Cases c_1 and c_4 provide decreasing failure rates as both q and β are lower than one; c_2 and c_3 are cases that also result decreasing failure rates even when $\beta > 1$ (as they satisfy the condition stated by Vila, Nakano and Saulo (2019) as examples refer to monotonicity properties) and finally cases c_5 and c_6 are those with increasing failure rates. Note that cases c_1 and c_4 are overdispersed, that is $E(X) < Var(X)$. Cases c_2 and c_5 are very close to being equidispersed ($E(X) \approx Var(X)$ as a Poisson distribution) and c_3 and c_6 are underdispersed. To illustrate the profile of these cases, in Figure 3, the PMFs of these six cases are shown for a sample size equal $n = 2$, that is $Y = X_1 + X_2$.

Table 1 – Parameters of the cases

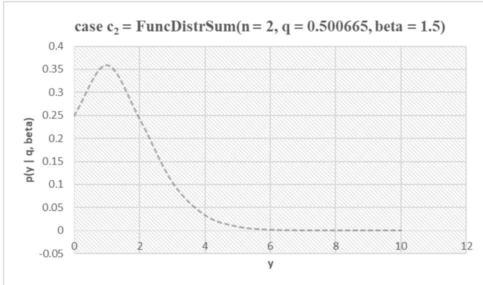
Cases	q	β	$E(X)$	$Var(X)$
c_1	0.4	0.5	2.040	27.599
c_2	0.500665	1.5	0.674	0.640
c_3	0.50005	2.5	0.520	0.289
c_4	0.5	0.5	3.787	85.325
c_5	0.51	1.455	0.711	0.711
c_6	0.75	2	1.152	0.824

Performances of the $\overline{X}$ control charts are obtained considering the sample sizes $n = \{1, 2, 3, 5, 7, 10, 30, 50, 100, 300\}$ and set to have an error type I $\alpha \approx 0.005$ to yield $ARL_0 \approx 200$. Note that as the random variable $Y = n\overline{X}$ assumes integer values in Z^+, then the control limits for $\overline{X}$ control chart may not reach exactly the desired value of ARL_0 equal 200.

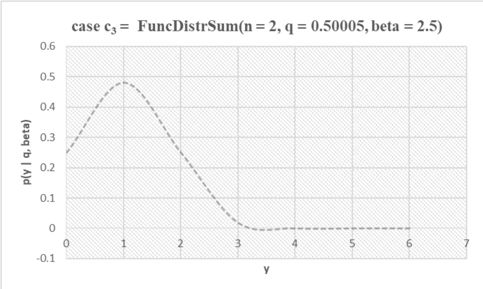
In Table 2, the values of ARL of the case c_4 obtained by two methods: Markov Chain (MC) procedure and Monte Carlo simulations (SIM) are put together for comparative purposes. It can be observed that both methods yield significantly closer values of Average Run Length (ARL). Similar results are observed for the other cases so the results which will be shown hereon are those obtained by Markov Chain as they are the exact ones.



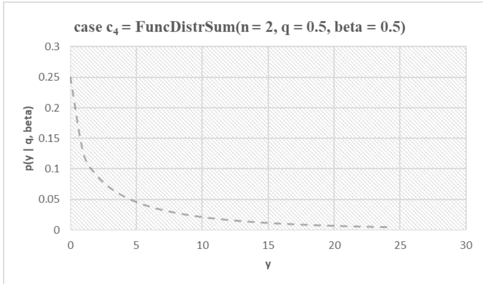
(a) The Discrete Weibull PMF for case c_1



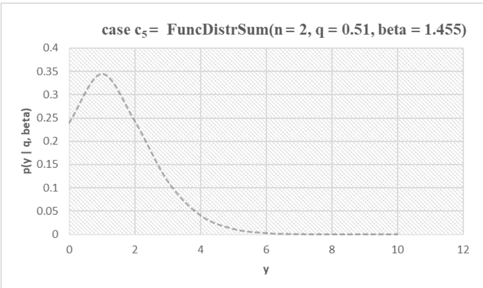
(b) The Discrete Weibull PMF for case c_2



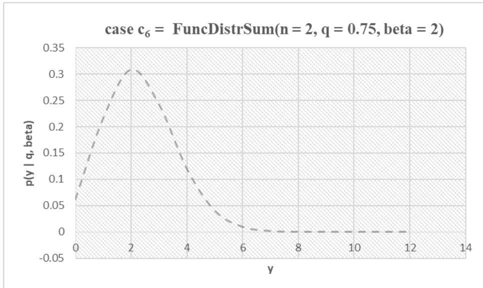
(c) The Discrete Weibull PMF for case c_3



(d) The Discrete Weibull PMF for case c_4



(e) The Discrete Weibull PMF for case c_5



(f) The Discrete Weibull PMF for case c_6

Figure 3 – PMF of the cases shown in Table 1 for a sample size $n = 2$. Source: Author

Table 2 – Case c_4: Comparing performance metric by Markov Chain and Monte Carlo Simulation - Detection for increases in the mean

						Sample size n								
q	β	E(X)	Var(X)	Method	1	2	3	5	7	10	30	50	100	300
0.5	0.5	3.787	85.325	MC	205.488	204.579	200.860	205.050	203.644	201.296	201.701	200.167	202.565	200.925
				SIM	205.237	204.575	200.860	204.069	204.474	201.487	201.890	201.078	200.088	202.593
0.6	0.5	7.259	291.093	MC	50.589	39.581	32.998	26.008	21.243	16.548	6.200	3.555	1.744	1.017
				SIM	50.682	39.680	33.040	26.053	21.210	16.550	6.201	3.555	1.737	1.017
0.8	0.5	39.666	8066.588	MC	5.551	3.417	2.541	1.786	1.446	1.208	1.001	1.000	1.000	1.000
				SIM	5.550	3.416	2.542	1.787	1.446	1.208	1.001	1.000	1.000	1.000
0.5	0.4	7.808	680.030	MC	34.513	24.576	19.692	14.904	12.093	9.521	4.139	2.694	1.575	1.022
				SIM	34.473	24.551	19.699	14.904	12.113	9.513	4.140	2.692	1.570	1.022
0.5	0.3	30.921	28869.903	MC	10.543	6.484	4.847	3.390	2.677	2.102	1.176	1.040	1.001	1.000
				SIM	10.545	6.479	4.844	3.390	2.676	2.103	1.176	1.040	1.001	1.000
0.4	0.4	3.813	163.898	MC	107.919	89.556	79.374	70.269	64.089	58.117	45.083	40.531	36.258	31.317
				SIM	107.941	89.402	79.891	70.355	63.963	58.020	45.035	40.648	35.745	31.621
0.55	0.4	11.526	1426.153	MC	21.208	14.264	11.035	7.940	6.236	4.753	1.995	1.404	1.061	1.000
				SIM	21.191	14.277	11.031	7.944	6.239	4.752	1.995	1.404	1.060	1.000

The results summarized in Table 3 are concerning to increases in the mean value, so only upper control limit is needed. For each block, the first row is the in-control parameters, followed by 2 rows with shifts only on the parameter q, specifically $q_1 = \delta + q_0$, with $\delta = \{0.1; 0.3;\}$; next two cases when only the parameter β changes, that is, $\beta_1 = \beta_0 - \gamma$, with $\gamma = 0.1; 0.2$ and finally the last two rows are the scenarios when both parameters shift (but produce an increase in the mean). And to close each block-rows, the probability and asymptotic upper control limit are included together.

Table 3 – Results of performance - Detection for increases in the mean - Markov Chain

Case		q	β	E(X)	Var(X)	Sample size n									
						1	2	3	5	7	10	30	50	100	300
c_1	$q_0 < 0.5, \beta_0 < 1$	0.4	0.5	2.040	27.599	209.107	209.678	212.842	203.720	200.525	203.854	206.259	203.038	206.423	202.028
	$q_1 > q_0$	0.5	0.5	3.787	85.325	56.923	45.520	39.339	30.107	24.755	19.955	7.837	4.486	2.118	1.045
		0.7	0.5	15.221	1235.316	8.003	5.004	3.734	2.497	1.934	1.518	1.017	1.000	1.000	1.000
	$\beta_1 < \beta_0$	0.4	0.4	3.813	163.898	42.740	31.298	25.924	19.442	15.981	13.000	6.008	3.916	2.192	1.115
		0.4	0.3	11.893	4491.546	14.008	8.756	6.634	4.585	3.603	2.812	1.411	1.140	1.012	1.000
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.35	0.4	2.642	82.626	73.874	58.190	50.691	40.703	35.310	30.758	18.515	13.865	8.997	3.844
		0.45	0.4	5.490	327.055	26.375	18.200	14.487	10.330	8.181	6.370	2.657	1.771	1.179	1.001
	UCL					33.000	22.500	18.000	13.400	11.143	9.300	5.633	4.640	3.750	2.940
	Asym UCL					15.572	11.609	9.853	8.092	7.155	6.319	4.511	3.954	3.393	2.821
c_2	$q_0 \approx 0.5, \beta_0 > 1$	0.500665	1.5	0.674	0.640	253.293	498.504	263.713	515.198	366.278	307.882	223.084	242.064	250.770	234.838
	$q_1 > q_0$	0.600665	1.5	0.929	0.957	59.011	74.025	38.463	44.292	27.534	18.392	5.233	3.107	1.589	1.009
		0.800665	1.5	1.966	2.840	5.921	4.298	2.630	1.989	1.480	1.200	1.001	1.000	1.000	1.000
	$\beta_1 < \beta_0$	0.500665	1.4	0.711	0.749	123.730	213.046	123.525	206.752	149.903	123.195	73.862	66.486	50.114	23.232
		0.500665	1.3	0.758	0.901	66.314	99.162	61.576	88.935	65.322	52.496	26.930	20.990	12.775	4.335
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.550665	1.4	0.840	0.927	63.762	88.106	49.928	64.640	43.453	31.611	11.428	7.348	3.589	1.297
		0.700665	1.3	1.561	2.531	8.643	7.010	4.334	3.452	2.422	1.785	1.043	1.003	1.000	1.000
	UCL					3.000	2.500	2.000	1.800	1.571	1.400	1.067	0.980	0.890	0.797
	Asym UCL					2.735	2.131	1.864	1.596	1.453	1.326	1.050	0.965	0.880	0.793
c_3	$q_0 \approx 0.5, \beta_0 > 1.5$	0.50005	2.5	0.520	0.289	49194.470	2422.333	1695.824	1999.965	312.327	1071.938	253.128	320.652	268.045	226.501
	$q_1 > q_0$	0.60005	2.5	0.656	0.338	2868.901	287.911	179.167	151.662	34.182	60.784	8.677	5.446	2.288	1.055
		0.80005	2.5	1.115	0.567	32.378	8.897	5.490	3.514	1.804	1.660	1.007	1.000	1.000	1.000
	$\beta_1 < \beta_0$	0.50005	2.4	0.526	0.301	15986.820	1380.735	989.874	1165.019	218.298	683.761	176.763	211.841	165.354	114.568
		0.50005	2.3	0.533	0.315	5839.776	803.025	593.263	696.408	153.343	436.943	120.598	135.891	98.477	55.898
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.55005	2.4	0.593	0.328	4223.784	485.760	327.274	322.959	70.847	157.568	28.324	21.671	9.581	2.534
		0.60005	2.2	0.699	0.415	306.810	80.264	55.630	47.584	15.350	22.255	4.267	2.753	1.430	1.004
	UCL					2.000	1.500	1.333	1.200	1.000	1.000	0.767	0.720	0.660	0.600
	Asym UCL					1.905	1.499	1.319	1.139	1.043	0.958	0.773	0.716	0.658	0.600
c_4	$q_0 = 0.5, \beta_0 < 1$	0.5	0.5	3.787	85.325	205.488	204.579	200.860	205.050	203.644	201.296	201.701	200.167	202.565	200.925
	$q_1 > q_0$	0.6	0.5	7.259	291.093	50.589	39.581	32.998	26.008	21.243	16.548	6.200	3.555	1.744	1.017
		0.8	0.5	39.666	8066.588	5.551	3.417	2.541	1.786	1.446	1.208	1.001	1.000	1.000	1.000
	$\beta_1 < \beta_0$	0.5	0.4	7.808	680.030	34.513	24.576	19.692	14.904	12.093	9.521	4.139	2.694	1.575	1.022
		0.5	0.3	30.921	28869.903	10.543	6.484	4.847	3.390	2.677	2.102	1.176	1.040	1.001	1.000
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.4	0.4	3.813	163.898	107.919	89.556	79.374	70.269	64.089	58.117	45.083	40.531	36.258	31.317
		0.55	0.4	11.526	1426.153	21.208	14.264	11.035	7.940	6.236	4.753	1.995	1.404	1.061	1.000
	UCL					58.000	39.500	31.333	23.800	19.857	16.500	10.067	8.340	6.780	5.367
	Asym UCL					27.580	20.611	17.524	14.428	12.780	11.311	8.131	7.152	6.166	5.161
c_5	$q_0 > 0.5. \beta_0 > 1$	0.51	1.455	0.711	0.711	1099.123	279.659	588.556	265.152	503.269	351.808	245.058	268.790	254.206	225.726
	$q_1 > q_0$	0.61	1.455	0.986	1.085	170.768	47.744	65.295	27.494	32.405	19.104	5.207	3.077	1.542	1.007
		0.81	1.455	2.149	3.453	8.947	3.479	2.974	1.703	1.474	1.180	1.001	1.000	1.000	1.000
	$\beta_1 < \beta_0$	0.51	1.355	0.755	0.845	388.118	127.074	227.736	112.844	182.910	128.602	73.023	65.131	44.553	19.003
		0.51	1.255	0.810	1.035	159.988	62.609	96.533	51.491	51.491	50.996	24.650	18.865	10.584	3.458
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.56	1.355	0.894	1.056	169.575	56.198	82.191	38.751	50.008	32.018	11.048	7.052	3.300	1.236
		0.61	1.255	1.156	1.709	41.496	16.254	18.237	9.211	9.278	5.981	2.029	1.409	1.051	1.000
	UCL					4.000	2.500	2.333	1.800	1.714	1.500	1.133	1.040	0.940	0.840
	Asym UCL					2.883	2.247	1.965	1.682	1.532	1.398	1.108	1.018	0.928	0.836
c_6	$q_0 > 0.5, \beta_0 > 1$	0.75	2	1.152	0.824	1328.827	556.860	536.371	230.356	498.528	231.379	289.364	293.869	227.363	232.315
	$q_1 > q_0$	0.85	2	1.698	1.401	58.147	24.472	17.867	7.723	7.929	4.078	1.559	1.149	1.004	1.000
		0.95	2	3.413	4.266	3.605	1.945	1.497	1.123	1.058	1.008	1.000	1.000	1.000	1.000
	$\beta_1 < \beta_0$	0.75	1.9	1.211	0.950	456.177	224.292	211.264	98.129	176.333	86.581	73.706	58.910	32.203	13.188
		0.75	1.8	1.279	1.113	183.408	100.641	92.627	45.593	69.641	36.186	23.016	15.710	7.329	2.404
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.8	1.9	1.455	1.221	115.501	54.391	44.196	19.791	24.651	12.137	4.781	2.800	1.421	1.005
		0.85	1.8	1.941	2.042	19.006	9.544	7.086	3.640	3.393	2.098	1.116	1.014	1.000	1.000
	UCL					4.000	3.000	2.667	2.200	2.143	1.900	1.600	1.500	1.390	1.290
	Asym UCL					3.491	2.806	2.502	2.198	2.036	1.892	1.579	1.483	1.386	1.287

Analyzing the outcomes shown in Table 3, some observations can be picked up:

- For the decreasing failure rate ($\beta_0 < 1$) and overdispersed cases c_1 and c_4: The target $ARL_0 \approx 200$ are reached even for small sizes. As expected larger shifts in mean,

Table 4 – Results of performance - Detection for decreases in the mean - Markov Chain

Cases		q	β	E(X)	Var(X)	Sample size n									
						1	2	3	5	7	10	30	50	100	300
c_1	$q_0 < 0.5, \beta_0 < 1$	0.4	0.5	2.040	27.599							209.864	229.109	211.307	200.696
	$q_1 < q_0$	0.35	0.5	1.491	15.852							43.695	29.582	13.237	3.337
		0.3	0.5	1.077	9.035							11.998	6.380	2.483	1.064
	$\beta_1 > \beta_0$	0.4	0.6	1.379	8.973							83.202	45.422	13.476	2.089
		0.4	0.8	0.867	2.369							20.658	5.432	1.304	1.000
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.3	0.4	1.811	41.365							27.587	24.337	17.827	10.369
		0.3	0.8	0.530	1.139							2.703	1.209	1.000	1.000
	LCL											0.433	0.660	0.970	1.360
	Asym LCL											-0.430	0.127	0.687	1.259
c_2	$q_0 \approx 0.5, \beta_0 > 1$	0.500665	1.5	0.674	0.640							273.768	279.848	215.690	242.543
		0.400665	1.5	0.485	0.439							13.339	6.489	2.239	1.038
	$q_1 < q_0$	0.300665	1.5	0.336	0.298							2.274	1.311	1.008	1.000
		0.500665	1.7	0.618	0.497							166.858	123.509	54.436	17.188
	$\beta_1 > \beta_0$	0.500665	1.6	0.643	0.559							210.345	180.555	101.898	53.642
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.400665	1.4	0.506	0.497							16.053	8.405	2.977	1.132
		0.400665	1.8	0.443	0.335							8.933	3.828	1.406	1.001
	LCL											0.300	0.380	0.470	0.553
	Asym LCL											0.298	0.382	0.468	0.555
c_3	$q_0 \approx 0.5, \beta_0 > 1.5$	0.50005	2.5	0.520	0.289							469.877	435.918	323.181	246.486
	$q_1 < q_0$	0.40005	2.5	0.406	0.252							24.413	11.491	3.668	1.150
		0.30005	2.5	0.301	0.213							3.596	1.779	1.060	1.000
	$\beta_1 > \beta_0$	0.50005	2.6	0.515	0.280							446.941	398.823	274.508	173.877
		0.50005	2.8	0.508	0.266							416.209	351.313	217.590	106.737
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.40005	2.4	0.408	0.257							25.024	11.937	3.853	1.177
		0.30005	2.8	0.300	0.211							3.565	1.762	1.057	1.000
	LCL											0.233	0.300	0.370	0.437
	Asym LCL											0.267	0.324	0.381	0.440
c_4	$q_0 = 0.5, \beta_0 < 1$	0.5	0.5	3.787	85.325						292.930	193.111	199.548	201.900	200.584
	$q_1 < q_0$	0.45	0.5	2.778	48.200						119.485	37.412	22.283	11.428	2.959
		0.4	0.5	2.040	27.599						53.255	10.288	5.365	2.253	1.047
	$\beta_1 > \beta_0$	0.5	0.6	2.380	23.214						255.646	45.289	20.605	6.047	1.277
		0.5	0.8	1.371	4.904						204.099	5.996	1.814	1.012	1.000
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.45	0.6	1.812	14.359						104.951	11.263	4.491	1.574	1.001
		0.4	0.7	1.054	4.151						42.352	2.142	1.142	1.000	1.000
	LCL										0.100	0.967	1.380	1.910	2.590
	Asym LCL										-3.737	-0.557	0.422	1.408	2.414
c_5	$q_0 > 0.5, \beta_0 > 1$	0.51	1.455	0.711	0.711							446.609	302.867	281.992	207.064
	$q_1 < q_0$	0.41	1.455	0.510	0.480							18.188	6.562	2.376	1.028
		0.31	1.455	0.354	0.322							2.658	1.308	1.009	1.000
	$\beta_1 > \beta_0$	0.51	1.555	0.676	0.613							334.256	185.924	121.714	40.628
		0.51	1.755	0.623	0.483							206.669	84.458	33.664	5.196
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.41	1.355	0.535	0.551							22.442	8.786	3.312	1.120
		0.31	1.755	0.330	0.261							2.177	1.154	1.001	1.000
	LCL											0.300	0.400	0.490	0.587
	Asym LCL											0.315	0.404	0.494	0.586
c_6		0.75	2	1.152	0.824					197.239	264.053	342.847	211.248	268.926	207.559
	$q_1 < q_0$	0.65	2	0.850	0.573					30.066	24.977	5.990	2.324	1.243	1.000
		0.55	2	0.646	0.430					8.554	5.808	1.377	1.027	1.000	1.000
	$\beta_1 > \beta_0$	0.75	2.1	1.102	0.724					179.782	224.303	201.635	97.834	78.685	23.220
		0.75	2.3	1.021	0.579					152.027	167.160	80.528	27.576	12.319	2.087
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.65	1.9	0.884	0.643					32.352	28.237	8.005	3.110	1.528	1.002
		0.55	2.1	0.630	0.397					8.181	5.421	1.286	1.014	1.000	1.000
	LCL									0.286	0.400	0.700	0.820	0.910	1.017
	Asym LCL									0.268	0.413	0.725	0.822	0.918	1.017

faster are the detection, reaching $ARL_1 \approx 1$ for $n > 10$. However, asymptotic and probability upper control limits are closer only for samples of $n \geq 100$.

- For the decreasing failure rate ($\beta_0 > 1$) and underdispersed cases c_2 and c_3: First it is relevant to mention that the levels of underdispersion of these two cases are quite different. The first case c_2 is almost equidispersed while the case c_3 is more underdispersed. Such differences may explain the behavior of the performance between these cases. The values of ARL_0 are slightly higher than 200 for any sample size in the case c_2 but very higher than the target value for the case c_3 mainly for sample size $n < 100$. Note the asymptotic and probability upper control limits are closer for the case c_3 even for a small sample size and true only for $n \geq 30$ in case of c_2. As expected larger shifts in mean, faster are the detection, reaching $ARL_1 \approx 1$ for $n > 30$.

- For the increasing failure rate ($\beta_0 > 1$) and underdispersed cases c_5 and c_6: Again it is relevant to distinguish the levels of underdispersion of these two cases. The first case c_5 is an equidispersed one while the case c_6 is more underdispersed. In both cases, the target value of ARL_0 is not reached for small sample sizes, very higher than 200 for $n < 30$, and even for $n \geq 30$, they are slightly higher than 200. Asymptotic and probability control limits match only for $n \geq 30$. Coherently, as larger are shifts in mean, faster are the detection, reaching $ARL_1 \approx 1$ for $n > 30$.

The results summarized in Table 4 are related to decreases in the mean value, consequently only lower control limit is presented. Like Table 3, for each block, the first row is the in-control parameters, followed by 2 rows with shifts only on the parameter q, specifically $q_1 = q_0 - \delta$, with $\delta = \{0.05; 0.1\}$; next three cases when only the parameter β changes, that is, $\beta_1 = \beta_0 + \gamma$, with $\gamma = 0.1; 0.3$ and finally the last two rows are the scenarios when both parameters shift but yield a reduction in the mean. Also the probability and asymptotic lower control limits are included together. Differently from Table 3 in which the calculated ARL_0 values are very larger than the target value of 200, here the opposite is observed, that is, the calculated ARL_0 values are very smaller than the target value of 200 mainly for small sample size $n \leq 7$ since $P(Y < 0) >> \alpha$ (thus the lower control limit is zero) yielding very low values for ARL_0, meaning very false alarms rates. In this sense, values of ARL_1 are calculated for the cases in which the calculated $ARL_0 \geq 160$, otherwise they are left blank. Analyzing the outcomes of Table 4, some interesting results may be listed:

- Increases only on β parameter (keeping q fixed) provoke reductions in the mean and the variance. If the probability distribution of the out-of-control distribution of $\overline{X}$ is drawn, probable it is embedded in the in-control distribution of $\overline{X}$. Thus, such type of shifts may be very difficult to be signed by an $\overline{X}$ control chart and consequently may yield a very poor performance metric. Observe that all values of $ARL_1 \approx 165$ when $\beta_1 = 0.6; 0.7; 0.8$ and $q = 0.4$ for a sample of ten units. Similar results are observed for other cases, values of ARL_1 closer to ARL_0 although reductions in the mean and variance. Fortunately, such problem vanishes as the sample size increases in almost all cases.

- Comparing Tables 3 and 4 it is noted that for shifts of the same magnitude either on q or on β, the signalization is faster for the increases in the mean than the decreases. Increasing (decreasing) the mean implies an increase (decrease) in the variance and the above argument in the previous bullet may explain such behavior.
- About the closeness of the asymptotic and probability control limits it can be observed that for cases c_1 and c_4 only when $n \geq 300$; c_3 for $n \geq 50$; cases c_2 and c_5, $n \geq 30$ and finally for c_6, $n \geq 7$.

One may wish to know which are the impacts if the control limits of $\overline{X}$ control chart based on Continuous Weibull (CW) are naively applied on data that follow a Discrete Weibull (DW). Considering that a (discrete) large data (as failure times of an equipment) are available in Phase I of a process monitoring. Since the practitioner may not have knowledge that such discrete random variable follows a discrete Weibull distribution, a first reasonable idea is to consider that the random variable comes from a Continuous Weibull, thus their parameters can be estimated by some method as the maximum likelihood, for example. Then these parameters are used to find the control limits of $\overline{X}$ (considering that a Continuous Weibull distribution) to be latter employed to monitor an $\overline{X}$ (but the data are originated from a Discrete Weibull). To find out these effects a complementary simulation study is conducted choosing the $\text{ARL}_0 = 200$ ($\alpha = 0.005$) as the target value and a sample size of $n = 30$ using the following steps:

Step 1: Take a random sample of $r = 10^5$ units from a Discrete Weibull of parameters q_0 and β_0

Step 2: Estimate the parameters α_0 and β_{CW_0} of a Continuous Weibull by the maximum likelihood method from the data of Step 1,

Step 3: Take a random sample of $n = 30$ units from a Continuous Weibull of parameters α_0 and β_{CW_0} of Step 2 and calculate its sample mean.

Step 4: Repeat the step 3, $k = 5 \times 10^6$ runs to get the empirical distribution of $\overline{X}_{CW}$.

Step 5: The quantile $q_{1-\alpha}^{CW}$ (q_{α}^{CW}) of the empirical distribution of $\overline{X}_{CW}$ of step 4 is the upper (lower) control limit if the interest is to detect increase (decrease) in the process mean for a fixed an error of type I equal to α.

Step 6 Take $k = 5 \times 10^6$ random samples of $n = 50$ units each from a Discrete Weibull of parameters q_1 and β_1; calculate the sample mean $\overline{X}_i^{DW}$, for $i = 1, \ldots k$; $D_i = 1$ if $\overline{X}_i^{DW} \geq q_{1-\alpha}^{CW}$ ($\overline{X}_i^{DW} \leq q_{\alpha}^{CW}$).

Step 7: $ARL^{false} = \dfrac{k}{\sum_{i=1}^{k} D_i}$ to detect increase (decrease) in the process mean.

Table 5 presents the results of this complementary simulation study. Three cases are chosen: $c_1 : (q_0 = 0.4; \beta_0 = 0.5)$ and $c_3 : (q_0 = 0.50005; \beta_0 = 2.5)$. In each case ARL_0^{false} is obtained to check if it meets to the target value of 200. Moreover, ARL_1^{false} are obtained considering shifts that yielded increase ($\mu_1 > \mu_0$) and decrease ($\mu_1 < \mu_0$) in the process mean. For comparative purposes the "true" values of ARL obtained by the Discrete Weibull is placed together (taken from Tables 3-4) at the last column. It can be noted that the target value of 200 for ARL_0 is not met in all cases if the control limits obtained under a Continuous Weibull are used equivocally, the consequences: more false alarms in case of decreasing means and true alarms postponed in case of increasing means. Moreover, the control limits for the continuous distribution are always higher than those obtained following the discrete one.

Table 5 – Impacts - the use of the control limits of Continuous Weibull on $\overline{X}$ from Discrete Weibull data

	q	β	ARL^{false}	ARL^{true}
μ_0	0.4	0.5	275.786	206.259
$\mu_1 > \mu_0$	0.4	0.4	6.403	6.008
UCL			5.896	5.633
μ_0	0.4	0.5	90.711	209.864
$\mu_1 < \mu_0$	0.4	0.6	34.351	45.422
LCL			0.556	0.433
μ_0	0.50005	2.5	719.424	253.128
$\mu_1 > \mu_0$	0.50005	2.3	293.531	120.598
UCL			0.808	0.767
μ_0	0.50005	2.5	154.283	469.877
$\mu_1 < \mu_0$	0.50005	2.4	6.172	25.024
LCL			0.302	0.233

Thus, it can be observed that there are impacts when data from a Discrete Weibull distribution are naively used as data from a Continuous Weibull distribution, since there is not match between the control limits and consequently more false alarm, according to the $ARL's$ values available in Table 5. To address the observed issues in the simulation study detailed in Table 5, it is advised to collect a sample during in-control process conditions and subject it to evaluation using the Akaike Information Criterion (AIC). This analytical approach aims to determine whether a Discrete Weibull distribution provides a superior fit for the data compared to competing distributions, including the Continuous Weibull, as well as discrete distributions such as Poisson or Negative Binomial, typically applied for count data.

4 The use of supplementary run rules to monitor the Discrete Weibull mean using the $\overline{X}$ control chart

In this chapter the supplementary rules proposed by Klein (2000) and Khoo and Ariffin (2006) are included to improve the performance of the $\overline{X}$ control chart. The first section is related to the Klein's supplementary rule and the second to Khoo's supplemntary rule.

4.1 Klein's supplementary run rule

Klein (2000) proposed an easy supplementary rule to improve $\overline{X}$. A corrective action is taken if a sequence of two values of $\overline{X}_i$ are beyond the control limits, that is, if $(\overline{X}_{i-1} > UCL) \bigcap (\overline{X}_i > UCL)$, in case to detect increases in the mean. For other results no action is taken. Such procedure may be described a set of four states

$$\mathbf{E} = \{A_1, B_1, A_2, B_2\}$$

The states A_1 is when $\overline{X}_{i-1} > UCL$ and the size of the sequence is one; B_1 is when the $\overline{X}_{i-1} < UCL$ and the size of the sequence is one. A_2 is when $\overline{X}_i > UCL$ and the size of the sequence is two; B_2 is when the $\overline{X}_i < UCL$ and the size of the sequence is two, with $p = P(\overline{X} > UCL)$. The transition matrix of one step is described by

$$\mathbf{Q} = \begin{array}{c} \\ A_1 \\ B_1 \\ A_2 \\ B_2 \end{array} \begin{array}{c} \begin{array}{cccc} A_1 & B_1 & A_2 & B_2 \end{array} \\ \begin{pmatrix} 0 & 1-p & p & 0 \\ p & 0 & 0 & 1-p \\ p & 1-p & 0 & 0 \\ p & 1-p & 0 & 0 \end{pmatrix} \end{array}.$$

The stationary distribution is $\pi = \{\pi_1, \pi_2, \pi_3, \pi_4\}$ with $\pi = \pi\mathbf{Q}$. Solving the system of equations with the restriction $\sum_{i=1}^{4} \pi_i = 1$ yields $\pi_1 = \frac{p}{1+p}$; $\pi_2 = \frac{1-p}{2-p}$; $\pi_3 = \frac{p^2}{1+p}$ and $\pi_4 = \frac{(1-p)^2}{2-p}$. The value π_3 is the long term to reach the state A_2 which requires an adjustment of the process. To get an ARL_0 around 200, the value of p must be around 0.07325. In case to detect decrease in the mean or bilateral shifts in the mean, the number of states and the transition matrix need a few adjustments.

4.2 Khoo's supplementary run rule

Khoo and Ariffin (2006) proposed an easy supplementary rule to improve $\overline{X}$. A corrective action is taken whenever X_i are beyond the control limits or if a sequence of two values are between the warming limits (say, WL) and control limits, otherwise, no action is necessary. This procedure can be described by a Markov chain with four transition states.

Whereas the focus is on detection of increases in the mean, the matrix and their states can be presented as follows:

$$\mathbf{E} = \{A_1, A_2, B, C\},$$

where the state A_1 is when $UWL < \overline{X}_{i-1} < UCL$ and the size of the sequence is one; A_2 is when $UWL < \overline{X}_i < UCL$ and the size of the sequence is two; B is when the $\overline{X} < UWL$; C is when $\overline{X} > UCL$. Whereas the transition matrix of one step is described by

$$\mathbf{Q} = \begin{array}{c} \\ A_1 \\ A_2 \\ B \\ C \end{array} \begin{array}{c} \begin{array}{cccc} A_1 & A_2 & B & C \end{array} \\ \begin{pmatrix} 0 & w_u & p & p_u \\ w_u & 0 & p & p_u \\ w_u & 0 & p & p_u \\ w_u & 0 & p & p_u \end{pmatrix} \end{array},$$

with $w_u = P(UWL < \overline{X}_{i-1} \leq UCL)$; $p = P(0 \leq \overline{X} \leq UWL)$; $p_u = P(\overline{X} > UCL)$. The process is stopped for adjustment whenever $\overline{X} > UCL$ or a sequence of two values of $\overline{X} \in]UWL; UCL]$, that is, $(UWL < \overline{X}_{i-1} \leq UCL) \bigcap (UWL < \overline{X}_i \leq UCL)$. The stationary distribution $\pi = \{\pi_1, \pi_2, \pi_3, \pi_4\}$ with $\pi = \pi\mathbf{Q}$. Solving the system of equations with the restriction $\sum_{i=1}^{4} \pi_i = 1$ yields $\pi_1 = \dfrac{w_u}{1 + w_u}$; $\pi_2 = \dfrac{w_u^2}{1 + w_u}$; $\pi_3 = p$ and $\pi_4 = p_u$. The sum of π_2 and π_4 is the long term probability to requires a stoppage for an adjustment of the process. To get an ARL_0 around 200, UWL and UCL are searched to meet $p_u + \frac{w_u^2}{1+w_u} \approx 0.005$ due to the discretiness nature of the Discrete Weibull distribution.

4.3 Performance of the $\overline{X}$ control chart for Discrete Weibull using Klein's and Khoo's supplementary run rules

In this section the performance of $\overline{X}$ is evaluated by comparing its results with those obtained when using Klein (2000) and Khoo and Ariffin (2006) supplementary rules.

Hence, three (of six cases mentioned before in Table 1) are chosen to analyse their features and verify their performance, as follows:

- case c_1 is overdispersed;

- case c_2 is very close to be equidispersed and;
- case c_3 is underdispersed.

For illustration, it is examined the performance of both the 'pure" $\overline{X}$ control chart and the procedures proposed by Klein (2000) and Khoo and Ariffin (2006). For the analyze are considered sample sizes $n = \{3, 5, 7, 10, 30\}$ and set the type I error rate $\alpha \approx 0.005$ to yield $ARL_0 \approx 200$.

The results summarized in Table 6 regard increases in the mean value, then only the upper control limit is needed. Analyzing the outcomes shown in Table 6, some observations can be point out:

- For the case c_1 the target $ARL_0 \approx 200$ is reached in all control charts, for $\overline{X}$ and for the scenarios where the Klein and Khoo procedures are used, in checking the sample sizes analyzed here. When there is a shift in the process mean (in general), Khoo's procedure presents better results compared to "pure" $\overline{X}$. Additionally, when there are small shifts in the mean with $q_1 > q_0$, Klein's also provides better results than $\overline{X}$. In other scenarios, $\overline{X}$ presents better results than Klein's procedure.
- The values of ARL_0 are higher than 200 for any sample size in the case c_2, but smaller than the target value for the case c_3. When is compared the performance of $\overline{X}$, Klein's and Khoo's procedures in scenarios where there is a shift in the mean for the cases c_2 and c_3, it can be noted that (analogously to what was seen for the case c_1) the Khoo's rules present better results than $\overline{X}$ and Klein's. However, for the c_3 case, the Klein's procedure presents better results than $\overline{X}$ in some scenarios (see $n = 5$) or even results that can compete with $\overline{X}$ (see $n = 7$, $n = 30$ with $q_1 > q_0$).
- However, for the c_3 case, the Klein's procedure presents better results than $\overline{X}$ in some scenarios (see $n = 5$) or even results that can compete with $\overline{X}$ (see $n = 7$, $n = 30$ with $q_1 > q_0$).

Table 6 – Some comparisons of the performance of the control charts: $\overline{X}$, $Klein's$ and $Khoo's$ procedures.

					Sample Size n											
					5			7			10			30		
Cases		q	β	$E(X)$	$\overline{X}$ (pure)	$Klein's$	$Khoo's$	$\overline{X}$ (pure)	$Klein's$	$Khoo's$	$\overline{X}$ (pure)	$Klein's$	$Khoo's$	$\overline{X}$ (pure)	$Klein's$	$Khoo's$
c_1	$q_0 < 0.5, \beta_0 < 1$	0.40	0.50	2.040	203.720	198.876	200.109	200.525	205.488	200.101	203.854	213.528	200.286	206.259	205.705	200.011
	$q_1 > q_0$	0.50	0.50	3.787	30.107	25.722	27.462	24.755	20.892	24.581	19.955	16.355	17.174	7.837	6.282	6.490
		0.70	0.50	15.221	2.497	3.217	2.352	1.934	2.688	1.922	1.518	2.325	1.478	1.017	2.003	1.018
	$\beta_1 < \beta_0$	0.40	0.40	3.813	19.442	28.214	18.543	15.981	22.536	15.905	13.000	17.672	12.096	6.008	7.295	5.474
		0.40	0.30	11.893	4.585	8.348	4.464	3.603	5.095	3.592	2.812	4.790	2.734	1.411	2.468	1.399
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.35	0.40	2.642	40.703	67.508	39.558	35.310	58.290	35.197	30.758	49.552	29.506	18.515	25.802	17.301
		0.45	0.40	5.490	10.330	14.071	9.746	8.181	10.878	8.133	6.370	8.239	5.873	2.657	3.510	2.483
	UCL				$UCL = 13.6$	$UCL = 5.4$	$UCL = 13.6$	$UCL = 11.14$	$UCL = 5$	$UCL = 11.14$	$UCL = 9.5$	$UCL = 4.6$	$UCL = 9.5$	$UCL = 5.7$	$UCL = 3.53$	$UCL = 5.7$
							$UWL = 8.4$			$UWL = 9.86$			$UWL = 6.1$			$UWL = 4.27$
c_2	$q_0 \approx 0.5, \beta_0 > 1$	0.500665	1.50	0.674	515.198	402.262	383.729	366.278	478.213	217.903	307.882	703.919	225.118	223.084	367.351	200.915
	$q_1 > q_0$	0.600665	1.50	0.929	44.292	32.809	30.244	27.534	27.146	16.036	18.392	23.547	12.474	5.233	5.731	4.074
		0.800665	1.50	1.966	1.989	2.772	2.111	1.480	2.392	1.403	1.200	2.165	1.183	1.001	2.000	1.001
	$\beta_1 < \beta_0$	0.500665	1.40	0.711	206.752	191.616	176.188	149.903	207.416	93.775	123.195	262.991	90.636	73.862	108.559	61.691
		0.500665	1.30	0.758	88.935	95.598	83.489	65.322	95.058	43.168	52.496	105.521	39.320	26.930	37.140	21.868
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.550665	1.40	0.840	64.640	56.281	50.829	43.453	50.634	26.679	31.611	49.064	29.995	11.428	12.893	8.641
		0.700665	1.30	1.561	3.452	4.344	3.407	2.422	3.486	2.100	1.785	2.885	1.811	1.043	2.023	1.056
	UCL				$UCL = 1.8$	$UCL = 1.2$	$UCL = 1.8$	$UCL = 1.57$	$UCL = 1.14$	$UCL = 1.57$	$UCL = 1.4$	$UCL = 1.1$	$UCL = 1.4$	$UCL = 1.1$	$UCL = 0.9$	$UCL = 1.1$
							$UWL = 1.2$			$UWL = 1.14$			$UWL = 1.1$			$UWL = 0.9$
c_3	$q_0 \approx 0.5, \beta_0 > 1.5$	0.50005	2.50	0.520	1999.965	335.615	291.705	312.327	2093.597	281.522	1071.938	1453.750	635.365	253.128	438.575	231.515
	$q_1 > q_0$	0.60005	2.50	0.656	151.662	40.380	33.776	34.182	98.617	28.809	60.784	54.992	32.328	8.677	8.911	7.190
		0.80005	2.50	1.115	3.514	3.204	2.353	1.804	3.166	1.713	1.660	2.474	1.518	1.007	2.002	1.007
	$\beta_1 < \beta_0$	0.50005	2.40	0.526	1165.019	262.844	219.111	218.298	1420.558	197.333	683.761	1004.364	421.823	176.763	302.059	160.782
		0.50005	2.30	0.533	696.408	202.724	161.753	153.343	949.234	138.816	436.943	678.923	277.998	120.598	201.858	109.025
	$q_1 \neq q_0, \beta_1 \neq \beta_0$	0.55005	2.40	0.593	322.959	86.709	71.236	70.847	288.402	61.791	157.568	176.497	89.623	28.324	32.210	23.994
		0.60005	2.20	0.699	47.584	22.264	16.972	15.350	40.402	13.310	22.255	24.178	13.805	4.267	5.159	3.733
	UCL				$UCL = 1.2$	$UCL = 0.8$	$UCL = 1.2$	$UCL = 1$	$UCL = 0.86$	$UCL = 1$	$UCL = 1$	$UCL = 0.8$	$UCL = 1$	$UCL = 0.77$	$UCL = 0.67$	$UCL = 0.77$
							$UWL = 0.8$			$UWL = 0.86$			$UWL = 0.8$			$UWL = 0.7$

5 Applications

This chapter contains the section 5.1, where it is discussed an example related to service quality. The objective is to determine whether the waiting time in a hospital emergency room is stable and in control.

The primary focus is to provide examples where the variables of interest are discrete. Both sets of control limits are determined using the procedure described in section 2.1, and their calculations are carried out using a program developed for this purpose, available as Supplementary Material A.

In this example, the variable of interest is time and, the samples are collected based on the assumption that the process is in control. If a sample is identified as out of control, additional investigation is necessary. In such cases, the out-of-control samples should be removed and replaced with new ones. This process continues iteratively until there is an indication of non-special causes.

5.1 Waiting time (in minutes) in a hospital emergency room

This example is based in Montgomery (2020). Consider the waiting time (in minutes) to be seen by a nurse or doctor in a hospital emergency room. Aiming at patient satisfaction, the hospital assumes an average waiting time of 4.57 minutes to be reasonable. Regarding that the time records are in minutes (which can be considered discrete), it was found that the Discrete Weibull distribution with parameters $q = 0.967$ and $\beta = 1.947$ represents the in-control waiting time. The objective is to detect that the average time may have increased and the hospital needs to devise strategies to get the situation in control. Thus, hourly samples of the waiting time of 5 patients are evaluated. Table 7 shows the values of the samples in 22 hours of evaluation. Adopting an ARL_0 of approximately 200, $UCL = 8.0$ is set. It can be observed that there is a sign of an increase in the average waiting time in the 5-th, 6-th and 16-th samples. Figure 4 shows the control chart.

To confirm the distribution's adequacy, waiting times for a hundred patients were collected and summarized as follows: (0)1; (1)10; (2)10; (3)15; (4)14; (5)17; (6)13; (7)5; (8)7; (9)1; (10)2; (11)3; (12)2. The number in parenthesis is the time (in minutes) followed by the frequency. That is, (0)1 means that one patient is immediately attended; (1)10, 10 patients wait for one minute to be attended, and so on. The Discrete Weibull's AIC is 472.10, while the Continuous Weibull's AIC is 483.73. A lower AIC suggests the Discrete Weibull is more suitable, as per Millar (2011).

Table 7 – Example - Waiting times in minutes

# Sample	x_1	x_2	x_3	x_4	x_5	$\overline{x}$	Decision
1	3	5	7	6	4	5	In-control
2	2	7	8	2	10	5.8	In-control
3	5	14	1	8	8	7.2	In-control
4	10	3	4	3	8	5.6	In-control
5	24	8	2	15	27	15.2	Out-of-control
6	15	4	4	13	5	8.2	Out-of-control
7	4	9	6	0	5	4.8	In-control
8	4	1	2	3	0	2	In-control
9	7	8	6	5	0	5.2	In-control
10	3	1	6	5	7	4.4	In-control
11	5	3	6	3	1	3.6	In-control
12	1	3	2	0	9	3	In-control
13	3	1	1	2	2	1.8	In-control
14	2	7	3	5	4	4.2	In-control
15	4	2	7	1	1	3	In-control
16	9	15	7	12	21	12.8	Out-of-control
17	1	3	3	3	8	3.6	In-control
18	0	6	6	9	10	6.2	In-control
19	4	10	3	3	7	5.4	In-control
20	2	9	8	6	5	6	In-control
21	3	3	4	3	6	3.8	In-control
22	2	7	1	2	8	4	In-control

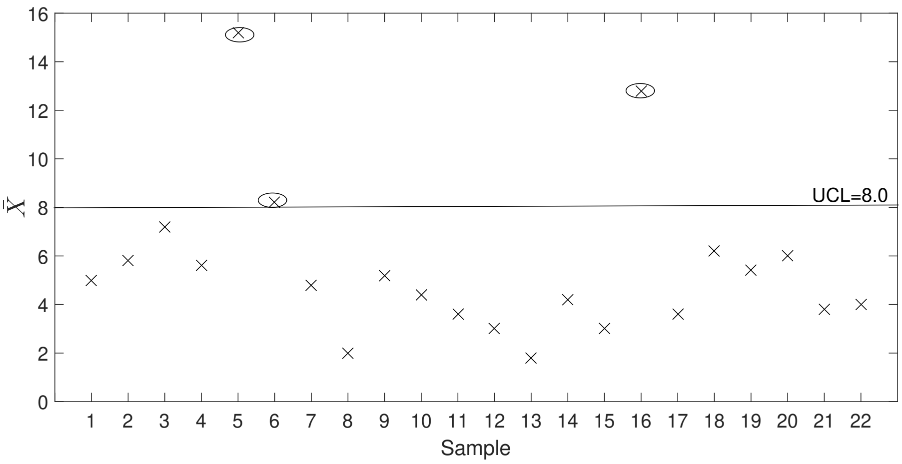


Figure 4 – Application of Control Chart using Discrete Weibull distribution for the waiting time. Source: Author

6 Final remarks

In this thesis, it is proposed a new approach to monitor a process mean when the quality characteristic X follows a Discrete Weibull distribution. Usually, the data related to time is registered in discrete terms expressed as the number of units of hours, days, months, years, or in another unit, showing the great applicability of the Discrete Weibull distribution. Similarly to the continuous Weibull distribution, the distribution of $\overline{X}$ does not present a closed form which turns difficulty the elaboration of the $\overline{X}$ control chart. Initially, it is tried to find the distribution of the $\overline{X}$ by calculating all possible cases (due to the discreteness nature of the random variable), however, this option is computationally unfeasible as the sample size n increases. Motivated by this computational trouble is presented a solution employing Markov Chain which is viable computationally and easily implemented. The results are compared with those obtained by Monte Carlo simulation and both are very close. Another finding of this thesis is the use of the Central Limit Theorem to get the approximate distribution of $\overline{X}$ is not effective as larger sample sizes are required to get similar results to those obtained with Markov Chain and Monte Carlo simulation. Significant impacts were observed through simulations conducted in this study when data from a Discrete Weibull distribution are erroneously treated as data from a Continuous Weibull distribution. These impacts are evidenced in the lack of equivalence between the control limits, leading to an increased frequency of false alarms. The discrepancy in this regard is noted in the Average Run Length (ARL) values, as illustrated in Table 5. In Chapter 4, supplementary run rules are included to improve the performance of the proposed control chart. The introduction of supplementary run rules shows promising outcomes, particularly when employing Khoo's procedure (Khoo; Ariffin, 2006). Even though Klein's supplementary run rules (Klein, 2000) yield favorable results in specific scenarios, this procedure is not presented as a viable alternative. In Chapter 5 it is exemplified the application of the control chart and it can note that its use is simple and easily implemented. Specific programs in R are developed for the readers and available in the Appendixes A and B. The results showed in Chapters 3 and 4 are partially presented in Brazilian conference (Appendix C) and a paper submitted, accepted and published online (Appendix D). As a suggestion for future works, it is suggested an expansion for bivariate processes.

Appendix

APPENDIX A – Program 1: The determination of sum of n independent identically Discrete Weibull random variables by Markov Chain

```
# These libraries are required
library(tidyverse)
library(DiscreteWeibull)
library(pracma)
library(expm)
library(beepr)
library(writexl)

tic()
clear()
n <- 50
d_q <- 0.0
d_beta <- -0.0
q <- 0.4 + d_q
beta <- 0.5 + d_beta
k <- qdweibull(0.9999, q, beta,zero = TRUE)
x <- 0:(k*n - 1)

prob <- pmap_dbl(
  .l= list(
    x=x,
    q=list(q),
    beta=list(beta),
    zero = list(TRUE)
  ),
  .f = ddweibull
)

gera_matriz <- function(prob){
  prob_last <- 1-sum(prob)

  k_n <- length(prob) + 1 # length(prob)+1
  prob_matrix <- matrix(0,nrow = k_n,ncol = k_n)
  prob_matrix[1,] <- c(prob,prob_last)
  for(i in 2:k_n){
    prob_matrix[i,] <- c(0,
                         prob_matrix[i-1,1:(k_n-2)],
                         sum(prob_matrix[i-1,-(1:(k_n-2))]))
  }

  return(prob_matrix)
}

(matrix <- gera_matriz(prob))
(Q_exp_n <- matrix %^% n)
vetor_b <- c(1,rep(0, length(prob)))
(v <- t(vetor_b) %*% Q_exp_n)
(acum_matriz <- cumsum(t(vetor_b) %*% Q_exp_n))
```

```
p_x <- seq(0, length(prob), 1)
MC_matrix_model <- data.frame(n, q, beta,
                              p_x, acum_matriz)

MC_matrix_model

writexl::write_xlsx(x = MC_matrix_model,
                    path = "MC_matrix_model.xlsx")
save(x = MC_matrix_model, file = "MC_matrix_model.RData")
beep(2) # beep warning for program ended successfully

# beep warning for program ended unsuccessfully
options(error = function(){beep(9)})

toc()
```

APPENDIX B – Program 2: Function FuncDistrSum

```
# These libraries are required
library(tidyverse)
library(DiscreteWeibull)

FuncDistrSum <- function(n, q, beta){

  n_y <- 0
  max_cum_prob <- 0
  cont <- 0

  while(isTRUE(max_cum_prob <= 0.995)==TRUE) {

    cat("searching n_y:", n_y)
    cont <- cont+1
    y <- seq(0,n_y,1)
    d <- pmap_dbl(
    .l= list(x=y,
    q=list(q),
    beta=list(beta),
    zero = list(TRUE)),
    .f = ddweibull)
    df_probs <- data.frame(y,d)

    # the grid considering "n" distributions
    tib1 <- expand.grid(as.data.frame(
    matrix(y, n_y+1, n))) %>%

      # naming y1, y2...yn
      setNames(paste0("y",1:n)) %>%

      # creating id for each row of the dataframe
      mutate(id = 1:n()) %>%

      # realocate the column "id"
      dplyr::relocate(any_of("id")) %>%
      rowwise() %>%

      # creating the column yk = sum of ys
      # through the values of the reference row "id"
      mutate(y_k = sum(c_across(-id)))

    # the grid considering "n" distributions
    tib2 <- expand.grid(as.data.frame(matrix(y, n_y+1, n))) %>%

      # naming prob1, prob2...probn
      setNames(paste0("prob",1:n)) %>%
      mutate(id = 1:n()) %>%

      # creating id for each row
      dplyr::relocate(any_of("id")) %>%

      # gathering y in groups
      gather(grupo, y, -id) %>%
      left_join(df_probs,by = "y") %>%
      dplyr::select(-y) %>%
      spread(grupo, d)%>%
      rowwise() %>%
      mutate(prob_soma = prod(c_across(-id)))
```

```
    # joing dataframes tib1 and tib2, linked by"id"

    tib_final <- left_join(tib1, tib2, by = "id")%>%
      arrange(y_k) %>%
      group_by(y_k) %>%
      summarise(events = n(), prob_soma = sum(prob_soma)) %>%
      mutate(cum_prob_1 = cumsum(prob_soma)) %>%
      mutate(cum_prob = cumsum(prob_soma)[cont]) %>%
      mutate(n = n) %>%
      mutate(q = q) %>%
      mutate(beta = beta) %>%
      dplyr::relocate(any_of("beta")) %>%
      dplyr::relocate(any_of("q")) %>%
      dplyr::relocate(any_of("n"))

    max_cum_prob <- max(tib_final$cum_prob)
    cat(", prob max:", max_cum_prob)
    cat(", y_k:", tib_final$y_k[tib_final$cum_prob_1 == max_cum_prob], "\n")
    n_y <- n_y + 1

    save(tib_final, file = "tib_final.RData")
  }

  # showing the results obtained
    results <- tib_final %>%
  # filter maximum value of interest
    filter(cum_prob_1<=0.995) %>%
    filter(cum_prob_1 == max(cum_prob_1))

  cat("results (1): prob max <= 0.995:",
       results$cum_prob_1)
  cat(", y_k:", results$y_k, "\n")
  cat("results (2): prob. to cut (target 0.995):",
      tib_final$cum_prob_1[
      tib_final$cum_prob_1 == max_cum_prob],
      ", sum of" ,n ,"distributions")
  cat(", y_k:", tib_final$y_k[tib_final$cum_prob_1 == max_cum_prob], "\n")

  return(tib_final)
}

FuncDistrSum(n = n, q = q, beta = beta)
```

APPENDIX C – LV Brazilian Symposium on Operational Research

SBPO 2023 Conference - Acceptance notification

São José dos Campos - SP

LV SBPO

CARTA DE ACEITE

Prezado(a) **Leandro Alves da Silva**

Temos o prazer de informar que o seu trabalho intitulado **THE USE OF KLEIN'S SUPPLEMENTARY RUN RULE TO MONITOR LIFETIME OF SYSTEMS** dos autores **Leandro Alves da Silva, Linda Lee Ho, Roberto Quinino** foi aceito na categoria Trabalho completo (oral) para publicação nos anais do evento **LV Simpósio Brasileiro de Pesquisa Operacional.**

LUIZ SATORU OCHI
Coordenador Científico do LV SBPO

Certification by Galoá

SBPO 2023 Conference - Certificate

São José dos Campos - SP

LV SBPO

CERTIFICADO

A SOBRAPO - Sociedade Brasileira de Pesquisa Operacional certifica que o trabalho intitulado

THE USE OF KLEIN'S SUPPLEMENTARY RUN RULE TO MONITOR LIFETIME OF SYSTEMS

de autoria de:

Leandro Alves da Silva, Linda Lee Ho, Roberto Quinino

foi apresentado na forma de *Trabalho completo (oral)* no 55º SBPO - LV Simpósio Brasileiro de Pesquisa Operacional, realizado nos dias 06 a 09 de novembro de 2023.

LUIZ SATORU OCHI

Coordenador Científico do LV SBPO

Certification by Galoá

APPENDIX D – Manuscript has been submitted, accepted and published online

Manuscript Number: JPROCONT-D-23-00470R2 - Acceptance Notification

Gmail

Leandro Alves da Silva <alvesdasilva.leandro@gmail.com>

Decision on submission to Journal of Process Control
1 mensagem

Journal of Process Control <em@editorialmanager.com> 14 de dezembro de 2023 às 13:20
Responder a: Journal of Process Control <support@elsevier.com>
Para: Leandro Alves da Silva <alvesdasilva.leandro@gmail.com>

Manuscript Number: JPROCONT-D-23-00470R2

Markov Chain approach to get control limits for a Shewhart Control Chart to monitor the mean of a Discrete Weibull distribution

Dear Mr. Alves da Silva,

Thank you for submitting your manuscript to Journal of Process Control.

I am pleased to inform you that your manuscript has been accepted for publication.

My comments, and any reviewer comments, are below.
Your accepted manuscript will now be transferred to our production department. We will create a proof which you will be asked to check, and you will also be asked to complete a number of online forms required for publication. If we need additional information from you during the production process, we will contact you directly.

We appreciate you submitting your manuscript to Journal of Process Control and hope you will consider us again for future submissions.

We encourage authors of original research papers to share the research objects – including raw data, methods, protocols, software, hardware and other outputs – associated with their paper. More information on how our open access Research Elements journals can help you do this is available at https://www.elsevier.com/authors/tools-and-resources/research-elements-journals?dgcid=ec_em_research_elements_email.

Kind regards,
Chunhui Zhao
Senior Editor

Journal of Process Control

Editor and Reviewer comments:

More information and support

FAQ: When and how will I receive the proofs of my article?
https://service.elsevier.com/app/answers/detail/a_id/6007/p/10592/supporthub/publishing/related/

You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

At Elsevier, we want to help all our authors to stay safe when publishing. Please be aware of fraudulent messages requesting money in return for the publication of your paper. If you are publishing open access with Elsevier, bear in mind that we will never request payment before the paper has been accepted. We have prepared some guidelines (https://www.elsevier.com/connect/authors-update/seven-top-tips-on-stopping-apc-scams) that you may find helpful, including a short video on Identifying fake acceptance letters (https://www.youtube.com/watch?v=o5l8thD9XtE). Please remember that you can contact Elsevier s Researcher Support team (https://service.elsevier.com/app/home/supporthub/publishing/) at any time if you have questions about your manuscript, and you can log into Editorial Manager to check the status of your manuscript (https://service.elsevier.com/app/answers/detail/a_id/29155/c/10530/supporthub/publishing/kw/status/).
#AU_JPROCONT#

To ensure this email reaches the intended recipient, please do not delete the above code

---

Manuscript has been published online (preview)

Journal of Process Control 134 (2024) 103149

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Markov Chain approach to get control limits for a Shewhart Control Chart to monitor the mean of a Discrete Weibull distribution

Leandro Alves da Silva [a,*], Linda Lee Ho [a], Roberto da Costa Quinino [b]

[a] Universidade de São Paulo, São Paulo, SP, Brazil
[b] Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

ARTICLE INFO

Keywords:
Sum distribution
Average Run Length
Count data
Monte Carlo simulation

ABSTRACT

Typically, failure time is modeled using continuous distributions such as the Weibull or Gamma distributions. In many practical scenarios, data is recorded in terms of discrete counts, such as the number of days or cycles, therefore the Discrete Weibull distribution is employed to model such cases. In this paper, we propose the use of a Shewhart $\overline{X}$ control chart to monitor the mean of a Discrete Weibull process. While the distribution of the sum of Discrete Weibull random variables does not have a closed-form expression, it can be determined through a Markov Chain procedure, which enables the calculation of precise control limits. The Average Run Length (ARL) is the metric used to assess the performance of the control chart. Two numerical examples are provided to illustrate its practical application.

1. Introduction

In studies of failure time analysis or lifetime of systems, equipment, and components, it is common to collect continuous time measurements. In such scenarios, the Weibull distribution is frequently utilized due to its strong goodness of fit in various real-world situations [1].

However, there are many instances where the recording of these failures occurs after the events have taken place. In these cases, it is appropriate to consider discrete measures of time or even counts of the number of events within defined time intervals, such as hours, days, months, or years. We often switch to a new unit of measurement after frequent use, irrespective of the unit's age [2,3]. This approach allows for effective control and recording of event information.

To model such types of counts (often involving the sum of random variables), it is common to employ discrete distributions, such as Poisson, negative binomial, and geometric distributions, as discrete alternatives to the Exponential and Gamma distributions. The primary reasons for using these discrete distributions are: their known distribution of the sum of independent and identically distributed random variables, closed-form solutions, and strong goodness of fit in real-world scenarios. However, the sum of Discrete Weibull random variables is seldom explored in the literature, even though the characteristic of interest exhibits a good fit to this distribution when compared to other discrete alternatives. One possible explanation for this situation is that the sum of independent Weibull random variables, whether discrete or continuous, does not have a closed-form expression. Nevertheless, with the increasing computational capabilities, it has become feasible to find alternatives that allow for accurate approximations of the sum of random variables following the Discrete Weibull distribution, without causing difficulties for users, using software such as R or Python.

Motivated by the questions mentioned above and the widespread use of the Weibull distribution, we assess the significance of Nakagawa and Osaki's work [4], where they introduced the initial concepts of a discrete distribution to model a continuous-time Weibull distribution (referred to as type I Discrete Weibull). This is particularly relevant in practical cases where failure data is available in discrete forms/counts (represented by integer numbers), some real applications, including population dynamics model [5], stress–strength reliability analysis, evaluation of complex systems reliability [6], semiconductor manufacturing [7] and microbial count analysis in water [8] demonstrate its applicability in different areas. Our objective is to explore the application of this distribution in statistical process control.

In the literature, various versions of the Discrete Weibull distribution can be found, including the modifications by Almalki and Nadarajah [9]. Stein and Dattero [10] introduced another form of discrete Weibull, known as type II. Another Discrete Weibull variant (type III) was proposed by Padgett and Spurrier [11]. Other researchers have conducted studies on the Discrete Weibull distribution, such as Szymkowiak and Iwińska [12], where they describe some characterizations in terms of discrete aging intensity. Additionally, we can mention Jayakumar and Sankaran [13] and their research on the Generalized

* Corresponding author.
E-mail address: leandro_silva@usp.br (L. Alves da Silva).

https://doi.org/10.1016/j.jprocont.2023.103149
Received 4 August 2023; Received in revised form 5 December 2023; Accepted 14 December 2023

References

Abernethy, R. B. The new Weibull handbook. Quinta. : CRC press, 2006. Citado na página 10.

Ali, S. et al. Cumulative conforming control chart assuming discrete Weibull distribution. IEEE Access, IEEE, v. 8, p. 10123–10133, 2020. Citado 2 vezes nas páginas 11 e 13.

Almalki, S. J.; Nadarajah, S. Modifications of the Weibull distribution: A review. Reliability Engineering & System Safety, Elsevier, v. 124, p. 32–55, 2014. Citado na página 12.

Barbiero, A. A comparison of methods for estimating parameters of the type i discrete weibull distribution. Statistics and its Interface, International Press of Boston, v. 9, n. 2, p. 203–212, 2016. Citado na página 15.

Barbiero, A. Least-squares and minimum chi-square estimation in a discrete Weibull model. Communications in Statistics-Simulation and Computation, Taylor & Francis, v. 46, n. 10, p. 8028–8048, 2017. Citado na página 15.

Calvin, T. Quality control techniques for"zero defects". IEEE Transactions on Components, Hybrids, and Manufacturing Technology, IEEE, v. 6, n. 3, p. 323–328, 1983. Citado na página 13.

Chan, L.; Xie, M.; Goh, T. Two-stage control charts for high yield processes. International Journal of Reliability, Quality and Safety Engineering, World Scientific, v. 4, n. 02, p. 149–165, 1997. Citado na página 13.

Chan, L.; Xie, M.; Goh, T. Cumulative quantity control charts for monitoring production processes. International Journal of Production Research, Taylor & Francis, v. 38, n. 2, p. 397–408, 2000. Citado na página 13.

Goh, T. A control chart for very high yield processes. Quality Assurance, v. 13, n. 1, p. 18–22, 1987. Citado na página 13.

Khan, M. A.; Khalique, A.; Abouammoh, A. On estimating parameters in a discrete Weibull distribution. IEEE Transactions on Reliability, IEEE, v. 38, n. 3, p. 348–350, 1989. Citado 2 vezes nas páginas 14 e 15.

Khan, N. et al. Historical review of dengue virus and its impact on pakistan community. 2019. `https://reliefweb.int/report/pakistan/over-25000-dengue-cases-reported-year`. Citado na página 11.

Khoo, M. B.; Ariffin, K. N. b. Two improved runs rules for the shewhart x control chart. Quality Engineering, Taylor & Francis, v. 18, n. 2, p. 173–178, 2006. Citado 6 vezes nas páginas 11, 12, 28, 29, 30 e 34.

Klein, M. Two alternatives to the shewhart x control chart. Journal of Quality Technology, Taylor & Francis, v. 32, n. 4, p. 427–431, 2000. Citado 6 vezes nas páginas 11, 12, 28, 29, 30 e 34.

Meeker, W. Q.; Escobar, L. A.; Pascual, F. G. Statistical methods for reliability data. : John Wiley & Sons, 2022. Citado na página 10.

Millar, R. B. Maximum likelihood estimation and inference: With examples in R, SAS and ADMB. New York, NY: John Wiley & Sons, 2011. ISBN 978-0-470-09482-2. Citado na página 32.

Montgomery, D. C. Introduction to Statistical Quality Control. 8th. ed. Hoboken, NJ: John Wiley & Sons, Inc., 2020. Citado 2 vezes nas páginas 20 e 32.

Nakagawa, T. A summary of discrete replacement policies. European Journal of Operational Research, Elsevier, v. 17, n. 3, p. 382–392, 1984. Citado na página 10.

Nakagawa, T.; Osaki, S. The discrete Weibull distribution. IEEE Transactions on Reliability, IEEE, v. 24, n. 5, p. 300–301, 1975. Citado 5 vezes nas páginas 10, 11, 12, 13 e 15.

Padgett, W.; Spurrier, J. D. On discrete failure models. IEEE Transactions on Reliability, IEEE, v. 34, n. 3, p. 253–256, 1985. Citado na página 12.

Silva, J. F. et al. Analysis of the time to relieving pain in patients with chronic non-specific low back pain via cox proportional hazard model. Ciência e Natura, Universidade Federal de Santa Maria-Centro de Ciências Naturais e Exatas, v. 39, n. 2, p. 233, 2017. Citado na página 11.

Stein, W. E.; Dattero, R. A new discrete Weibull distribution. IEEE Transactions on Reliability, IEEE, v. 33, n. 2, p. 196–197, 1984. Citado na página 12.

Szymkowiak, M.; Iwińska, M. Characterizations of discrete Weibull related distributions. Statistics & Probability Letters, Elsevier, v. 111, p. 41–48, 2016. Citado na página 12.

Valadares, D. G. de O. et al. Repairable system analysis using the discrete weibull distribution. IEEE Transactions on Reliability, IEEE, 2023. Citado na página 12.

Vila, R.; Nakano, E. Y.; Saulo, H. Theoretical results on the discrete Weibull distribution of Nakagawa and Osaki. Statistics, Taylor & Francis, v. 53, n. 2, p. 339–363, 2019. Citado 2 vezes nas páginas 15 e 20.

Xie, M. et al. Statistical models and control charts for high-quality processes. : Springer Science & Business Media, 2002. Citado na página 13.

Xie, M. et al. A quality monitoring and decision-making scheme for automated production processes. International Journal of Quality & Reliability Management, MCB UP Ltd, 1999. Citado na página 13.

Zhang, C. et al. A control chart for the Gamma distribution as a model of time between events. International Journal of Production Research, Taylor & Francis, v. 45, n. 23, p. 5649–5666, 2007. Citado na página 13.