STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **7**, December 2019, pp 642–652.
Published online in International Academic Press (www.IAPress.org)

Evaluating the Goodness of the Sample Coefficient of Variation via Discrete Uniform Distribution

Ioanna Papatsouma [1*], Rahim Mahmoudvand [2], Nikolaos Farmakis [3]

[1] *Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK*
[2] *Department of Statistics, Bu-Ali Sina University, Hamedan, Iran*
[3] *Department of Mathematics, Aristotle University of Thessaloniki, 54124*

Abstract Discrete uniform distribution (DUD) is one of the simplest probability models, but it is now introduced as the main tool for the evaluation of resampling techniques which are rapidly entering data analysis and discovering useful information for the researchers. In this paper we evaluate whether the sample coefficient of variation (CV) is a good estimator for the population CV, when the random variable (r.v.) follows the DUD. A method is proposed to obtain the percentage of the number of samples where the CV lies within the bounds of the corresponding population CV and this value is used as a measure of goodness. Samples both with replacement and without replacement are examined, indicating that the goodness of the sample CV estimator increases with the sample size. The overall study gives a good idea of *whether the sample CV is generally a good estimator*. A real-life data set is analyzed to demonstrate the applicability of the proposed method in practice and the results are interpreted.

Keywords Coefficient of Variation, Discrete uniform distribution, Sampling; Measure of goodness.

AMS 2010 subject classifications 62D05, 62E17

DOI: 10.19139/soic-2310-5070-798

1. Introduction

The CV of a r.v. X is given by:

$$CV = \frac{\sigma}{\mu} \tag{1}$$

where σ is the population standard deviation and $\mu \neq 0$ is the population mean of the rv X [21, 22], while the value of the CV obtained from a sample is as follows:

$$\widehat{CV} = \frac{s}{\bar{x}} \tag{2}$$

where s is the sample standard deviation and $\bar{x}$ is the sample mean.

The CV is unit-free and thus, it is one of the most widely used statistical tools in various scientific fields. It is used as a tool for quality improvement [12], as a tool for managing loan portfolio risks [17], as a tool for measuring athletic performance [25], as a tool for calculating the sample size needed for carrying out studies [29], as a tool for obtaining a distribution model [20], as an intra-observer variability assessment tool [7] and so on.

*Correspondence to: Ioanna Papatsouma (Email:i.papatsouma@imperial.ac.uk). Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 2019 International Academic Press

Discrete uniform distributions play a naturally important role in many classical problems of probability theory [1, 13]. The interest of the study of discrete random variables is that the study is theoretical and can be exhaustive because we have a small amount of data. Another advantage of the discrete r.v. is that a continuous variable can be transformed into discrete by appropriately dividing its range. The continuous life time may not necessarily always be measured on a continuous scale but may often be counted as discrete random variables [10]. In survival analysis the survival function may be a function of count r.v. that is a discrete version of the underlying continuous r.v. For example, the length of stay in an observation ward is counted by number of days or survival time of leukemia patients is counted by number of weeks. Therefore, the use of a discrete distribution is more realistic than the use of a continuous one. An extension of the study for a larger number of values can also predict the marginal behavior of the continuous r.v. when the number of values is increasing, or even tends to infinity. The importance of the CV in a DUD is that it can be used as a main evaluation tool for resampling techniques, like bootstrapping and permutation testing [9, 23], which are very powerful in all areas of data analysis, such as in fraud detection, product categorization and disease diagnosis.

We investigate the case of the discrete r.v. X, following the DUD $DU\{0, 1, \ldots, N-1\}$. It can be easily seen that the CV of X is given by:

$$CV = \sqrt{\frac{N+1}{N-1}}\frac{\sqrt{3}}{3} \tag{3}$$

From (3), we conclude that the CV is independent of the difference between the consecutive terms and that, as the population size, N, increases, the CV tends to $\sqrt{3}/3 = 0.5774$.

Let $CV(N)$ and $CV(N+1)$ denote the CV for random variables following the discrete uniform distributions $DU\{0, 1, \ldots, N-1\}$ and $DU\{0, 1, \ldots, N\}$, respectively. Thus, we have:

$$\frac{CV(N+1)}{CV(N)} = \sqrt{\frac{N^2+N-2}{N^2+N}} \tag{4}$$

and we conclude that the CV is a strictly decreasing function and has an upper limit, CV$\leqslant$1, $\forall$ N$>$2.

In the related literature, not much work is seen in the discrete case and consequently, the published work on CV and DUD is rare. [11] recently developed procedures for interval estimation and hypothesis testing for the coefficient of variation in the continuous uniform distribution, [2, 3, 19] studied the convolution powers of DUD, [5, 4] studied order statistics from a DUD, [16] obtained bounds for the population CV in DUD along with other distributions and [6] defined goodness-of-fit statistics to test fit to a DUD. In the present study, the goodness of the sample CV is ultimately attributed to the percentage of samples where their CV is lying within the bounds of the population CV: $\sqrt{3}/3 < CV \leqslant 1$ [16] or approximately in the interval (0.5774,1].

The organization of the rest of this paper is as follows. In Section 2, we describe the sampling methods used. In Section 3, we provide an efficient algorithm for computing the rate of return and we apply it for different sample sizes and population sizes. In Section 4, regression and correlation analysis were carried out between the sample size and the rate of return. In section 5, a real-data application illustrates the effectiveness of the proposed algorithm. Section 6 concludes the paper and proposes future directions.

2. Sampling Method

The sampling method selected is the random sampling (RS) [8, 15], i.e. every element of the population has the same probability of being drawn with any other element. In the RS process it is usually considered that the selected element does not return to the population, so it cannot be re-selected. This sampling is known as sampling without replacement. However, there is a case where each element is selected and after its value is recorded, it is returned to the population so that it can be re-selected. This process is known as sampling with replacement.

It has been proved that $Var\bar{X}_r > Var\bar{X}$, where $\bar{X}_r$ is the mean of the sample with replacement and $\bar{X}$ is the mean of the sample without replacement and more precisely, it is bigger by $(N-1)/(N-n)$ [8], where N

and n are the population size and the sample size, respectively. Therefore, in the case of random sampling with replacement, we have less information from a sample of size n.

The number of different samples taken during sampling without replacement is equal to the number of possible combinations of n elements from a population of size N, which is denoted by $\binom{N}{n}$, while in sampling with replacement is equal to N^n.

3. Proposed Method

The proposed method for the evaluation of the estimate of the sample $\widehat{CV}$ is described in the following steps:

1. Recording the samples as n-dimensional vectors $(x_{i1}, x_{i2}, \ldots, x_{in})$, where *i* denotes the samples number.
2. Calculating the number of samples by solving the following equation:

$$\sum_{j=0}^{N-1} B_j = n$$

   where $B_j = 0, 1, \ldots, n$ is a non-negative integer and denotes the number of times that we observe number *j* in our sample.
3. Calculating the number of permutations that can be formed with the values of each n-dimensional vector.
4. Calculating the $\widehat{CV}$ of each sample. It can be shown that:

$$\widehat{CV}^2 = \frac{n}{n-1}\left(\frac{\overline{x_i^2}}{\bar{x_i}^2} - 1\right) = \frac{n^2}{n-1}\left(\frac{B_1 + 4B_2 + \ldots + (N-1)^2 B_{N-1}}{B_1 + 2B_2 + \ldots + (N-1)B_{N-1})^2} - \frac{1}{n}\right)$$

   where $x_i = (x_{i1}, x_{i2}, \ldots, x_{in}), i = 1, \ldots, n$.
5. Calculating the percentage (%) of the number of samples where the $\widehat{CV}$ lies within the interval $(\sqrt{3}/3, 1]$. This percentage will henceforth be called *rate of return* for the sake of brevity. The probability distribution of $\widehat{CV}$ is given by:

$$P\left(\widehat{CV}^2 = \frac{n^2}{n-1}\left(\frac{B_1 + 4B_2 + \ldots + (N-1)^2 B_{N-1}}{(B_1 + 2B_2 + \ldots + (N-1)B_{N-1})^2} - \frac{1}{n}\right)\right) = \frac{n!}{B_0!B_1!\ldots B_{N-1}!}\left(\frac{1}{N}\right)^n$$

We apply the proposed method when using both sampling methods, but the following propositions are useful in the case of sampling with replacement.

Proposition 1. In the case of sampling with replacement, the proposed algorithm examines $\binom{n+N-1}{N-1}$ samples.

Proof. As mentioned in section 2, the number of possible combinations of n elements from a population of size N is equal to N^n. We will, however, prove that the proposed algorithm examines $\binom{n+N-1}{N-1}$ samples, which is too much smaller than N^n.

Assume that $(x_{i1}, x_{i2}, \ldots, x_{in})$ follow a DUD $DU\{0, \ldots, N-1\}$. Since the sample $\widehat{CV}$ does not depend on the order of the samples, we only need to find the number of ordered samples $(x_{i1}, x_{i2}, \ldots, x_{in})$ where $x_{i1} \leqslant x_{i2} \leqslant \ldots \leqslant x_{in}$. It is easy to see that the number of ordered samples $(x_{i1}, x_{i2}, \ldots, x_{in})$ is equal to the number of non-negative integer solutions of the equation $\sum_{j=0}^{N-1} B_j = n$ [26], which completes the proof of the proposition 1.

Proposition 2. In the case of sampling with replacement, the number of distinct values of $\widehat{CV}$ is $\binom{n+N-1}{N-1} - n(N-2) - 1$.

Proof. Samples $(i, 0, 0, \ldots, 0)$ for $i = 1, \ldots, N-1$ have the same $\widehat{CV}$ because we have $B_i = 1, B_0 = n-1$ and therefore:

$$\widehat{CV}^2 = \frac{n^2}{n-1}\left(\frac{i^2 B_i}{(iB_i)^2} - \frac{1}{n}\right) = \frac{n^2}{n-1}\left(1 - \frac{1}{n}\right) \tag{5}$$

which does not depend on i. Similarly, considering samples $(i, i, 0, \ldots, 0)$ for $i = 1, \ldots, N-1$, we have $B_i = 2, B_0 = n-2$ and therefore:

$$\widehat{CV}^2 = \frac{n^2}{n-1}\left(\frac{i^2 B_i}{(iB_i)^2} - \frac{1}{n}\right) = \frac{n^2}{n-1}\left(\frac{1}{2} - \frac{1}{n}\right) \tag{6}$$

We can see generally that when $B_i = m, B_0 = n-m$, for $m = 1, \ldots, n$, then:

$$\widehat{CV}^2 = \frac{n^2}{n-1}\left(\frac{i^2 B_i}{(iB_i)^2} - \frac{1}{n}\right) = \frac{n^2}{n-1}\left(\frac{1}{m} - \frac{1}{n}\right) \tag{7}$$

As a result, we can neglect $N-2$ replications of $\widehat{CV}$ for each sample with $B_i = m, B_0 = n-m, i = 1, 2, \ldots, N-1, m = 1, 2, \ldots, n$. Since $i = 1, 2, \ldots, N-1$, there are $N-1$ samples having the same $\widehat{CV}$ and thus, we can remove $N-2$ same values of $\widehat{CV}$ and since $m = 1, 2, \ldots, n$, we can remove $n(N-2)$ same values of $\widehat{CV}$ among the possible values of $\widehat{CV}$.

Finally, the case of $(0, 0, \ldots, 0)$ is also removed as it has the same $\widehat{CV}$ with samples $(i, i, \ldots, i)$. In this way, we show that the number of same $\widehat{CV}$ values is $n(N-2)+1$ which completes the proof of the proposition 2.

3.1. DU{0,1,2}

The goodness of the sample $\widehat{CV}$ estimator of the r.v. X following a DUD $DU\{0, 1, 2\}$ is evaluated, i.e. for $N = 3$. The case of samples of size 6, which are taken from sampling with replacement, is fully described below.

The equation $\sum_{j=0}^{N-1} B_j = 6$ has 28 solutions. Each of the 28 solutions is represented in Table 1 in the form of a 6-dimensional vector $(x_{i1}, x_{i2}, \ldots, x_{i6}), i = 1, \ldots, 28$, the coordinates of which can obtain the values of 0, 1 and/or 2 in all possible ways. The number of permutations that can be formed with the values given in each row of the table is denoted by *r*.

All in all, we get $N^n = 3^6 = 729$ samples of size 6. We notice that in 390 samples of them, which are marked **in bold**, the $\widehat{CV}$ lies within the interval $(\sqrt{3}/3, 1]$. Finally, the number of distinct values for $\widehat{CV}$ is 21 which is found by removing 7 duplicated values (see Proposition 2).

Table 1. Description of samples and values of CV for $N = 3$ & $n = 6$.

x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	r	CV
0	**0**	**0**	**0**	**0**	**0**	**1**	**0**
0	0	0	0	0	1	6	2.4495
0	0	0	0	0	2	6	2.4495
0	0	0	0	1	1	15	1.5492
0	0	0	0	1	2	30	1.6733
0	0	0	0	2	2	15	1.5492
0	0	0	1	1	1	20	1.0954
0	0	0	1	1	2	60	1.2247
0	0	0	1	2	2	60	1.1798
0	0	0	2	2	2	20	1.0954
0	**0**	**1**	**1**	**1**	**1**	**15**	**0.7746**
0	**0**	**1**	**1**	**1**	**2**	**60**	**0.9033**
0	**0**	**1**	**1**	**2**	**2**	**90**	**0.8944**
0	**0**	**1**	**2**	**2**	**2**	**60**	**0.8427**
0	**0**	**2**	**2**	**2**	**2**	**15**	**0.7746**
0	1	1	1	1	1	6	0.4899
0	**1**	**1**	**1**	**1**	**2**	**30**	**0.6325**
0	**1**	**1**	**1**	**2**	**2**	**60**	**0.6452**
0	**1**	**1**	**2**	**2**	**2**	**60**	**0.6124**
0	1	2	2	2	2	30	0.5578
0	2	2	2	2	2	6	0.4899
1	1	1	1	1	1	1	0
1	1	1	1	1	2	6	0.3499
1	1	1	1	2	2	15	0.3873
1	1	1	2	2	2	20	0.3651
1	1	2	2	2	2	15	0.3098
1	2	2	2	2	2	6	0.2227
2	2	2	2	2	2	1	0
			TOTAL			**729**	

We repeat the same procedure for each sample of size $n = 2, 3, 4, 5, 7, 8, 9$ and 10 which is randomly selected and replaced and record the number of total samples as well as the number and the percentage of the samples where the $\widehat{CV}$ lies within the interval $(\sqrt{3}/3, 1]$ (Table 2).

Table 2. Results of sampling with replacement for $N = 3$ & $n = 2, 3, \ldots, 10$.

Sample size	**Number of samples**	**Number of samples where** $\sqrt{3}/3 < CV \leqslant 1$	**Rate of return (%)**
2	9	1	11.11
3	27	13	48.15
4	81	33	40.74
5	243	121	49.79
6	729	390	53.5
7	2187	1170	53.5
8	6561	3585	54.64
9	19683	10795	54.84
10	59049	39871	67.52

It is noticed that the percentage of the values of the sample CV within the bounds of the population CV increases in parallel with the sample size (Figure 1).

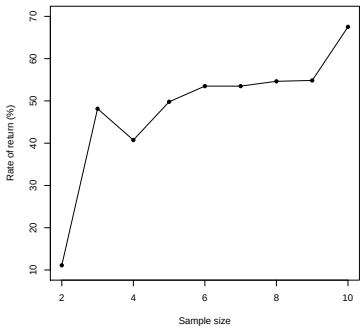

Figure 1. Rate of return (%) for $N = 3$ & $n = 2, 3, \ldots, 10$.

3.2. DU{0,1,2,3}

The goodness of the sample $\widehat{CV}$ estimator of the r.v. X following a DUD $DU\{0, 1, 2, 3\}$ is evaluated, i.e. for $N = 4$. Table 3 presents the results of sampling with replacement for samples of size $n = 2, 3, 4, 5, 6$ and 7.

Table 3. Results of sampling with replacement for $N = 4$ & $n = 2, 3, \ldots, 7$.

Sample size	**Number of samples**	**Number of samples where** $\sqrt{3}/3 < CV \leqslant 1$	**Rate of return (%)**
2	16	7	43.75
3	64	25	39
4	256	91	35.55
5	1024	526	51.37
6	4096	2137	52.17
7	16384	9479	57.86

An increasing trend of the percentage of the samples, where the $\widehat{CV}$ lies within the bounds of the population CV is obvious (Figure 2).

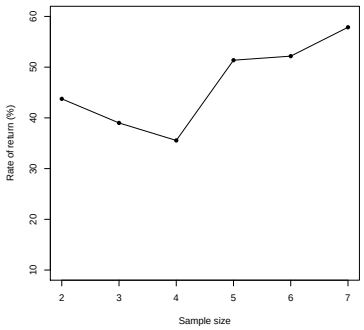

Figure 2. Rate of return (%) for $N = 4$ & $n = 2, 3, \ldots, 7$.

3.3. *DU*{0,1,2,3,4}

The goodness of the sample $\widehat{CV}$ estimator of the r.v. X following a DUD $DU\{0, 1, 2, 3, 4\}$ is evaluated, i.e. for $N = 5$. Table 4 presents the results of sampling without replacement for samples of size $n = 2, 3, 4$ and 5.

Table 4. Results of sampling without replacement for $N = 5$ & $n = 2, 3, 4, 5$.

Sample size	**Number of samples**	**Number of samples where** $\sqrt{3}/3 < CV \leqslant 1$	**Rate of return (%)**
2	10	2	20
3	10	5	50
4	5	4	80
5	1	1	100

We notice that there is only one sample of size 5, which is the population itself. In this case, the sample $\widehat{CV}$ and the population CV coincide and are equal to 0.7906.

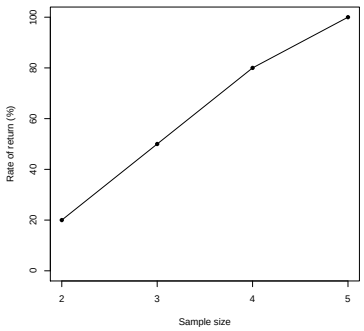

Figure 3. Rate of return (%) for $N = 5$ & $n = 2, 3, 4, 5$.

The percentage of the samples, where the $\widehat{CV}$ lies within the interval $(\sqrt{3}/3, 1]$, is still increasing and has a strong linear trend with respect to the sample size (Figure 3).

3.4. *DU*{0,1,2,3,4,5,6,7}

Finally, the goodness of the sample $\widehat{CV}$ estimator of the r.v. X following a DUD $DU\{0, 1, 2, 3, 4, 5, 6, 7\}$ is evaluated, i.e. for $N = 8$. Table 5 presents the results of sampling without replacement for samples of size

$n = 2, 3, 4, 5, 6, 7$ and 8.

Table 5. Results of sampling without replacement for $N = 8$ & $n = 2, 3, \ldots, 8$.

Sample size	Number of samples	Number of samples where $\sqrt{3}/3 < CV \leqslant 1$	Rate of return (%)
2	28	6	21.43
3	56	27	48.21
4	70	42	60
5	56	42	75
6	28	24	85.71
7	8	6	75
8	1	1	100

Similar to the case of $N = 5$, we notice that there is only one sample of size 8, which is the population itself. In this case, the sample $\widehat{CV}$ and the population CV coincide and are equal to 0.6999.

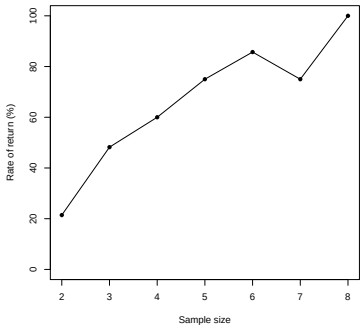

Figure 4. Rate of return (%) for $N = 8$ & $n = 2, 3, \ldots, 8$.

In general, the rate of return has an increasing trend with respect to the sample size (Figure 4).

4. Correlation between sample size and rate of return

After the steps 1 to 5 have been completed, we investigate the rate of increase of the rate of return when the sample size, n, increases by one unit.

The case of $N = 3$ and sampling with replacement is fully described, where the value of the $\widehat{CV}$ exceeds 0.8 (Table 6). In other words, at least 64% of the changes in the rate of return is explained by the change in the sample size. The value of the correlation coefficient indicates that there is a linear relationship between the rate of return and the sample size and the linear regression equation is given by:

$$\hat{y} = b_0 + b_1 \cdot n$$

where $\hat{y}$ is the estimate of the rate of return, n is the sample size and b_0, b_1 the regression coefficients. The value of the coefficient b_1 is derived from the pair sample size C rate of return (%) of Table 2 and on average is slightly higher than 2. This allows us to assume an increase in the rate of return by approximately 2% for each one-unit sample size increase. Therefore, with an initial goodness of the $\widehat{CV}$ estimator around 35%, a sample size of 35 to 40 elements is required in order to approach 100%. The value of the coefficient b_1 resulting from regression with data including samples of size $n > 10$ is marginally decreasing. Consequently, the average will fall below 2.

Table 6. Values of correlation coefficient and coefficient of determination.

	With replacement		Without replacement	
Population size, N	3	4	5	8
Correlation coefficient, r	0.8490	0.7837	0.9959	0.9361
Coefficient of determination, r^2	0.7208	0.6142	0.9918	0.8763

Table 6 above presents the values of the correlation coefficient and the coefficient of determination for both sampling cases examined. All values of the correlation coefficients are positive, therefore there is a positive linear relationship between the two variables [14], the rate of return and the sample size, whether the sample is taken with replacement or without.

The value of the correlation coefficient for $N = 5$ and sampling without replacement ($r = 0.9959$) confirms that there is almost a perfect linear relationship between the sample size and the rate of return, as seen in Figure 3. The corresponding value of the coefficient of determination confirms the fitting of the regression line to the DUD data.

5. Application

From 1st of January until 31st of December 2017, 840 patients were hospitalised in the RSA IGEA - Rehabilitation Residential Centre in Trieste, Italy. They were given the Barthel Index (BI) questionnaire [18] both at the beginning and at the end of the rehabilitation. This questionnaire contains 10 questions and assesses the degree of independence from daily activities, such as feeding, clothing and personal hygiene.

83 patients were excluded because the final data are not available and as we have been informed by the Medical Director, Dr Paolo Da CoL, these patients were discharged before ending the rehabilitation path. The sample consists of 757 patients with an average age of 84.09 $\pm$ 8.93 years old, of whom 522 (69%) are females and 235 (31%) are males.

We divided the questions into the following 3 categories based on the possible answers:

i. questions about help with personal hygiene and bathing were answered with 0 (dependent) or 1 (independent).

ii. questions about help with feeding, dressing/undressing, using the toilet and climbing stairs were answered with 0 (dependent), 1 (with help) or 2 (independent) and questions about fecal and urinary incontinence were answered with 0 (incontinent), 1 (occasionally) or 2 (continent).

iii. question about transfer from chair to bed was answered with 0 (not able), 1 (major help), 2 (minor help) or 3 (independent) and question about walking was answered with 0 (immobile), 1 (moving with wheelchair), 2 (with help of a person) or 3 (with aids for disabled).

The above answers are coded to equal levels, therefore assuming that the expected average, approximated by the sample estimator, makes sense, we calculate the coefficients of variation of all the questions before and after in each category, as well as the coefficients of variation of all the questions before and after in total (Table 7).

We first notice a decrease in the $\widehat{CV}$ in 9 of the 10 cases investigated, which indicates the improvement of the patients condition. The increase in the $\widehat{CV}$ in the case of help with bathing was expected and the reason is, as we were informed by the Medical Director, that they patients rarely stay alone in the bathroom. We can also point out that in the case of personal hygiene of patients, which is a binary variable, the final $\widehat{CV}$ has become less than 1, which means that the patients will become independent in their personal hygiene with higher probability after the end of the rehabilitation path.

From Table 7, we get the following rates of return by category of questions: 25% for $N = 2$, 50% for $N = 3$ and 25% for $N = 4$. Finally, we observe that while the initial $\widehat{CV}$ does not fall within the bounds of the population CV, the final $\widehat{CV}$ falls and coincides with the improvement of the patients condition.

We also used likelihood ratio tests (LRTs) to test whether there is statistically significant difference between the coefficients of variation before and after the rehabilitation. The results indicate that there is enough evidence to reject the null hypothesis, at 5% level of significance.

Remark: Only in the case of help with climbing stairs the value of the test statistic is not available.

Table 7. CV values for $= 2, 3$ and 4.

	Before	After
$N = 2$		
Help with personal hygiene	1.141	0.806
Help with bathing	11.188	19.416
$N = 3$		
Help with feeding	0.346	0.281
Help with using the toilet	1.052	0.683
Help with climbing stairs	4.818	1.253
Help with dressing/ undressing	1.078	0.831
Fecal incontinence	0.741	0.606
Urinary incontinence	0.892	0.736
$N = 4$		
Help with transfer from chair to bed	0.595	0.287
Help with walking	1.067	0.482
TOTAL	1.128	0.846

Table 8. Paired samples correlations and 95% CI for Before-After.

		95% CI Before-After	
$N = 2$	r	Lower Limit	Upper Limit
Help with personal hygiene	0.67	-0.201	-0.143
Help with bathing	0.406	-0.003	0.014
$N = 3$			
Help with feeding	0.784	-0.142	-0.091
Help with using the toilet	0.663	-0.496	-0.406
Help with climbing stairs	0.288	-0.61	-0.504
Help with dressing/ undressing	0.716	-0.32	-0.248
Fecal incontinence	0.834	-0.203	-0.136
Urinary incontinence	0.814	-0.182	-0.121
$N = 4$			
Help with transfer from chair to bed	0.688	-0.777	-0.667
Help with walking	0.601	-10.066	-0.921

Furthermore, paired samples t-tests were used for comparing the average responses before and after the rehabilitation and the results indicate that there is enough evidence to reject the null hypothesis, at 1% level of significance, apart from the case of help with bathing, which is not surprising as explained earlier.

Table 8 above adds the information that the average responses before and after the rehabilitation are all positively correlated and reports the 95% confidence intervals that confirm the improvement of the patients condition.

6. Conclusions and Future Directions

In this paper, we proposed a method to evaluate the goodness of the sample $\widehat{CV}$ and investigated the case of the discrete r.v. X, following the DUD when random samples are taken from the population. [16] investigated only the case of $N = 3$ and $n = 4$ when random samples are taken with replacement from the population, while we further investigated the cases of (a) $N = 3$ and $n = 2, 3, \ldots, 10$ and (b) $N = 4$ and $n = 2, 3, \ldots, 7$ and when it comes to samples without replacement, the cases of (c) $N = 5$ and $n = 2, 3, 4, 5$ and $N = 8$ and $n = 2, 3, \ldots, 8$.

The value of the sample $\widehat{CV}$ is strictly associated with the number of zeros included in each sample and thus, we stated and proved a proposition that derives the number of distinct values of the $\widehat{CV}$ in the case of sampling with replacement and we also introduced a model of low computational cost describing the relationship between the $\widehat{CV}$ and the sample. As a result, given the sample size, n, we can calculate the number of non-zero elements required in order the sample $\widehat{CV}$ to lie within the bounds of the population CV, which makes it a good estimator. As the population size, N, increases, the percentage of samples, taken both by replacement and without replacement, giving the sample $\widehat{CV}$ within the bounds of the population CV, is also increasing. The results are of interest for the scientific community, as the sample $\widehat{CV}$ can be used from now on as an evaluation measure for resampling techniques and the discrete distributions are gaining ground.

Regression analysis further describes the relationship between the sample size and the rate of return. The value of the correlation coefficient indicates that there is a stronger positive linear correlation between the sample size and the rate of return of the samples taken without replacement than the samples taken with replacement. Linear

regression models predict that samples of size $n > 40$ give high rates of samples where the sample $\widehat{CV}$ is lying within the bounds of the population CV. Numerical study showed that this further investigation can easily evaluate the variation in discrete data.

Directions for future research include using the proposed method to construct confidence intervals for the CV and examining the proposed method in (a) larger populations and (b) other families of discrete uniform distributions, like the Marshall-Olkin discrete uniform (MODU) distribution [27], the generalized DUD, known as Harris Discrete Uniform (HDU) distribution, [24] and the exponentiated Marshall-Olkin discrete uniform (E-MO-U) distribution [28].

Acknowledgement

The authors thank Dr Paolo Da Col, the Medical Director in RSA IGEA - Rehabilitation Residential Centre in Trieste, Italy for sending the data. They are also very grateful to the anonymous referees for the suggestions and comments which significantly improved the quality of the manuscript.

REFERENCES

1. N. Balakrishnan and V. B.Nevzorov, *A Primer on Statistical Distributions*, Hoboken, N.J. : Wiley, 2003.
2. H. Belbachir, *Determining the mode for convolution powers of discrete uniform distribution*, Probability in the Engineering and Informational Sciences, vol. 25, no. 4, pp. 469-475, 2011.
3. H. Belbachir,S. Bouroubi and A. Khelladi, *Connection between ordinary multinomials, Fibonacci numbers, Bell polynomials and discrete uniform distribution*, Annales Mathematicae et Informaticae 35, pp. 21-30, 2008.
4. S. Çalik and M. Güngör, *On the expected values of the sample maximum of order statistics from a discrete uniform distribution*, Applied Mathematics and Computation, vol. 157, no. 3, pp. 695-700, 2004.
5. S. Çalik and M. Güngör, *On The Moments of Order Statistics from Discrete Distributions*, Pakistan Journal of Statistics, vol. 26, no. 2, pp. 417-426, 2010.
6. V. Choulakian, R. A. Lockhart and M. A. Stephens, *CramerCvon Mises statistics for discrete distributions*, The Canadian Journal of Statistics, vol. 22, no. 1, pp. 125-137, 1994.
7. A. Christoforidis, V. Perifanis, E. Papadopoulou, M. Dimitriadou, E. Kazantzidou, E. Vlachaki and I. Tsatra, *Poor correlations between measurements of bone quality by quantitative ultrasound sonography and dual energy X-ray absorptiometry in patients with beta-thalassaemia major*, European Journal of Haematolog, vol. 82, no. 1, pp. 15-21, 2009.
8. N. Farmakis, *Introduction to Sampling*, A&P Christodoulidi Publishing Co, Thessaloniki (in Greek), 2016.
9. R. A. Fisher *The Design of Experiments*, New York: Hafner, 1935.
10. M. Gharib, B. I. Mohammed and W. E. R. Aghel, *The Exponentiated Marshall-Olkin Discrete Uniform Distribution With Application In Survival Analysis*, International Journal Of Modern Engineering Research, vol. 7, no. 8, pp. 34-48, 2017.
11. J. Hoseini and A. Mohammadi, *Estimator and Tests for Coefficient of Variation in Uniform Distribution*, Journal of Biometrics and Biostatistics, 3:149, 2012.
12. C. W. Kang, M. S. Lee, Y. J. Seong and D. M. Hawkins, *A control chart for the coefficient of variation*, Journal of Quality Technology, vol. 39, no. 2, pp. 151-158, 2007.
13. K. Krishnamoorthy, *Handbook of statistical distributions with applications (2nd edition)*, Chapman and Hall/CRC, 2015.
14. F. Kolyva-Machera and E. Bora-Senta, *Statistics: Theory and Applications (2nd edition)*, Ziti Publishing Co, Thessaloniki (in Greek), 2013.
15. P. S. Levy and S. Lemeshow, *Sampling of Populations: Methods and Applications (4th edition)*, John Wiley Sons, Inc., New York, 2008.
16. R. Mahmoudvand, H. Hassani and R. Wilson, *Is the Sample Coefficient of Variation a Good Estimator for the Population Coefficient of Variation?*, World Applied Sciences Journal, vol. 2, no. 5, pp. 519-522, 2007.
17. R. Mahmoudvand and T. Oliveira, *On the Application of Sample Coefficient of Variation for Managing Loan Portfolio Risks.* Preprint, 2018.
18. F. I. Mahoney and D. Barthel, *Functional evaluation: the Barthel Index*, Maryland State Medical Journal, 14, pp. 56-61, 1965.
19. L. Mattner and B. Roos, *Maximal probabilities of convolution powers of discrete uniform distributions*, Statistics and Probability Letters, vol. 78, no. 17, pp. 2992-2996, 2008.
20. I. Papatsouma and N. Farmakis, *Approximating Symmetric Distributions via Sampling and Coefficient of Variation*, Communications in Statistics-Theory and Methods, 2018.
21. K. Pearson, *Contributions to the mathematical theory of evolution, II: Skew variation in homogeneous material*, Philosophical Transactions of the Royal Society of London, vol. 186, pp. 343-414, 1895.
22. K. Pearson, *Mathematical contributions to the theory of evolution: Regression, heredity, and panmixia*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 187, pp. 253-318, 1896.
23. E. J. G. Pitman, *Significance tests which may be applied to samples from any population*, Royal Statistical Society Supplement 4: 119C130 and 225-232 (parts I and II), 1937.

24. C. B. Prasanth and E. Sandhya, *A Generalized Discrete Uniform Distribution*, Journal of Statistics Applications & Probability, vol. 5, no. 1, pp. 1-13, 2016.
25. D. B. Pyne, C. B. Trewin and W. G. Hopkins, *Progression and variability of competitive performance of Olympic swimmers*, Journal of Sports Sciences, vol. 22, no. 7, pp. 613-620, 2004.
26. S. M. Ross *A first course in Probability (9th edition)*, Pearson, 2012.
27. E. Sandhya and C. B. Prasanth, *Marshall-Olkin Discrete uniform distribution*, Journal of probability, Volume 2014, 10 pages, Article ID 979312, Hindawi Publishing Corporation, 2013.
28. K. Sheeja and S. Lakshmi, *The exponentiated Marshall-Olkin discrete uniform distribution with application in survival and hazard analysis on progesterone, estrogen and other various hormones*, Arya Bhatta Journal of Mathematics and Informatics, vol. 10, no. 2, pp. 413-420, 2018.
29. G. Van Belle and D. Martin, *Sample Size as a Function of Coefficient of Variation and Ratio of Means*, The American Statistician, vol. 47, no. 3, pp. 165-167, 1993.