Methods for analyse performance in distributed systems

Danil Sukhoplyuev[*], and *Alexey* Nazarov

MIREA, Russian Technology University, 119454 Moscow, Russia

Abstract. The purpose of this article is to analyze various methods for measuring central tendency in statistics, including arithmetic mean, median, winsorized mean, outlier exclusion method, Hodges-Lehmann estimator, and quantile estimation and much more. The advantages and disadvantages of each of these methods are discussed, as well as their practical applications in performance analysis in distributed systems. In particular, we focus on the importance of selecting an appropriate measure of central tendency that is robust to outliers and accurately reflects the distribution of the data. We also provide examples of how these methods can be applied to real-world datasets to gain insights into the underlying patterns and trends. Overall, this article provides a comprehensive overview of the different techniques for measuring central tendency and offers practical guidance for researchers and analysts looking to make informed decisions about perfomance analysis.

1 Introduction

Distributed systems have become increasingly important in modern computing due to their ability to provide high levels of scalability, fault-tolerance, and performance [20]. However, evaluating the performance of distributed systems is a complex task, as it involves measuring not only the individual components but also the interactions and dependencies between them. The ability to accurately evaluate the performance of distributed systems is crucial for ensuring that they meet their design goals and for identifying and addressing potential issues.

In the course of our investigation into various forms of performance testing, such as benchmarking, stress testing, and load testing, we delve into the application of monitoring tools and techniques for the real-time evaluation of distributed systems. It is evident that the methods of mathematical statistics serve as a means of descriptive analysis in this field of research.

For example, we posit an examination of the distribution of request per-second (RPS), assuming that it is characterized by normal distribution[1] – Fig.1., also known as the Gaussian distribution (1).

---

[*] Corresponding author: sukhoplyuev.d.i@edu.mirea.ru

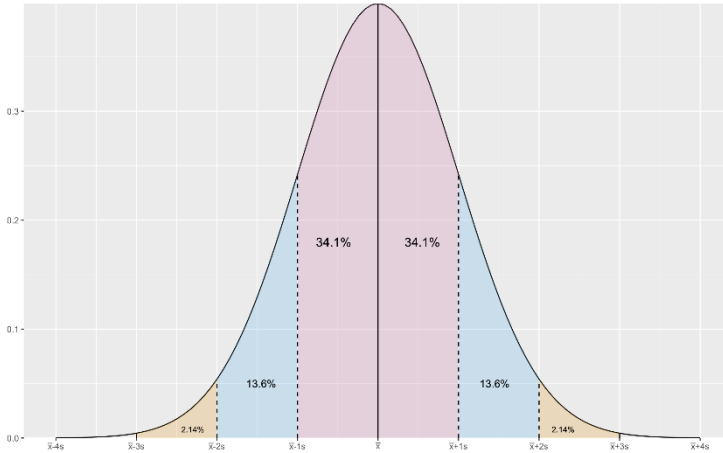$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{1}$$

Fig. 1. Normal distribution.

RPS is a performance metric used to measure the throughput of a system. RPS represents the number of requests that a system can process in a second, and it is often used as a key performance indicator for web servers, APIs, and other distributed systems that handle a high volume of requests. Measuring RPS [8] is important for ensuring that a distributed system can handle a large number of concurrent requests without experiencing performance issues or downtime. The problem is that at hand pertains that RPS encompasses various factors, such as the duration of data retrieval from the hard disk in the event of an input/output (I/O) bottleneck, the responsiveness of the random-access memory (RAM), the degree of request queue activity, among other elements. In essence, this parameter is intricate in nature and comprises a multitude of variables.

The normal distribution can be characterized by a set of two parameters - the expected value - μ and the standard deviation - σ. However, it should be noted that not all arbitrary probability distributions can be fully specified by two parameters, because they do not display the distribution form. This issue is addressed in a scholarly article [2] that explores different forms(Fig. 2.) of the normal distribution.
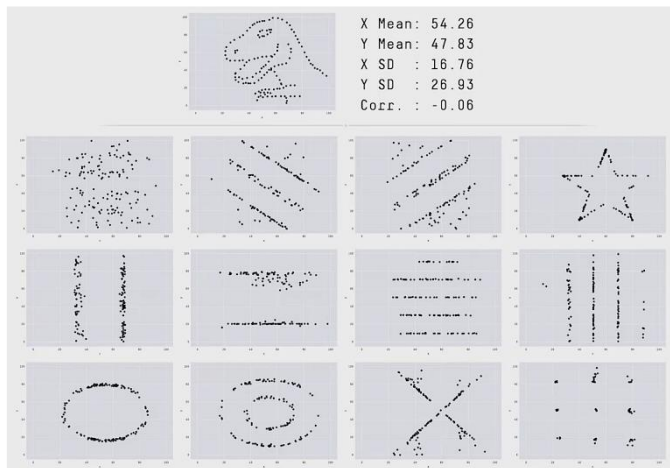
Fig. 2. Forms of normal distribution with correlation coefficients.

As a means of approbation, let us consider an API function that facilitates file creation within a distributed web application. By leveraging the widely adopted Java Spring framework, we implemented this function and with Docker Compose technology locally deployed several instances of this service in kernel mode. Following this, a load testing experiment was conducted with randomized routing and asynchronous request processing. The utilized instrumentation included the Apache JMeter application, as well as the testing platform comprising an Intel Core i9-10900KF central processing unit, 32 gigabytes of random-access memory, and a solid-state drive with a data transfer rate of 5500 megabytes per second. It is worth noting that the flow and intensity of the request remained consistent at 1 RPS throughout the entire experiment with 1000 iterations. The raw data underwent computational procedures with the assistance of the Python programming language and the NumPy library of mathematical functions. A histogram, depicted in Figure 3, was generated from the processed data.

Table 1. Load testing data during testing of the normal distribution.

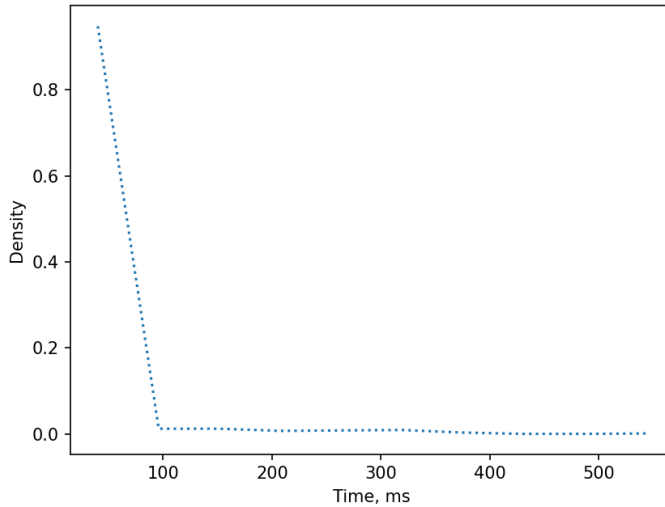| Iteration's count	Time interval, milliseconds
976	[37, 60)
19	(60, 110)
5	(110, 500+)

Fig. 3. A graph of the density of distributions during testing of the normal distribution.

In the plot depicted in Figure 3, it is observable that the principal modes remain confined to the range of 0 to 100 milliseconds, whereas minor outlines manifest subsequent to the 100-millisecond mark. Therefore, it can be hypothesized that the quantity of spikes and the heterogeneity of distributions will increase during the analysis of real production systems. Albeit lacking full clarity, this observation substantiates the conjecture that in practical settings wherein intricate functions governing system interactions, input/output operations, and intricate computations are involved and etc., the assurance of a normal distribution's manifestation is unstable.

In summary, the major issues associated with analyzing performance [9] are as follows:
-       Large variability in values - measurements are not localized in a single vicinity and may have a wide discrete range along the numerical axis;
-       Heavy-tailed and extreme outlines;
-       Multimodality;
-       Discretization effects, wherein continuous distributions acquire a discrete nature;
-       Asymmetry.

Henceforth, within the confines of this manuscript, we shall expound upon alternative techniques for appraising performance distributions that influence non-standard sectors of mathematical statistics, namely, central tendency, quantile approximations.

2 Central tendency

Central tendency [10] allows for the compression of intricate performance distributions into a single numerical value. One of the most trivial ways to implement this approach is the arithmetic mean.

2.1 Arithmetic mean

Let us consider a sample of n numerical values (2), in such an instance the arithmetic mean[11] can be computed utilizing (3).

$$x = \{x_1, x_2, ..., x_n\} \tag{2}$$

$$\overline{x} = \frac{x_1 + x_2 + ...x_n}{n} \tag{3}$$

However, if an outline include in our sample n, as in the case of (4), then the arithmetic mean loses its credibility as an indicator of central tendency.

$$x = \{1, 2, 3, 4, 5, 6, 273\}; \overline{x} = 48 \tag{4}$$

In general, the arithmetic mean is characterized by low robustness [12], indicating its vulnerability to the influence of extreme values, or outliers. The presence of a single extremely large value can significantly impact the accuracy of the result, hence prompting the consideration of an alternative measure, such as the median. While the median for samples (5) and (6) may yield identical values, the application of the Gaussian efficiency estimate (7) could reveal discrepancies in the respective measures.

$$x = \{1, 2, 3, 4, 5, 6, 7\}; \tilde{x} = 4 \tag{5}$$

$$x = \{1, 2, 3, 4, 5, 6, 273\}; \tilde{x} = 4 \tag{6}$$

$$GaussianEfficiency(T) = \frac{V(mean)}{V(T)} \tag{7}$$

The Gaussian efficiency can be defined as the ratio of the variance pertaining to the arithmetic mean estimate and the variance associated with the value of the estimate. After generating a set of samples from a normal distribution, we computed the metrics of the arithmetic mean and median values. The resulting density distribution indicators were displayed in Figure 4. Based on the obtained results, we constructed Table 2 for the Gaussian efficiency of the arithmetic mean and median.
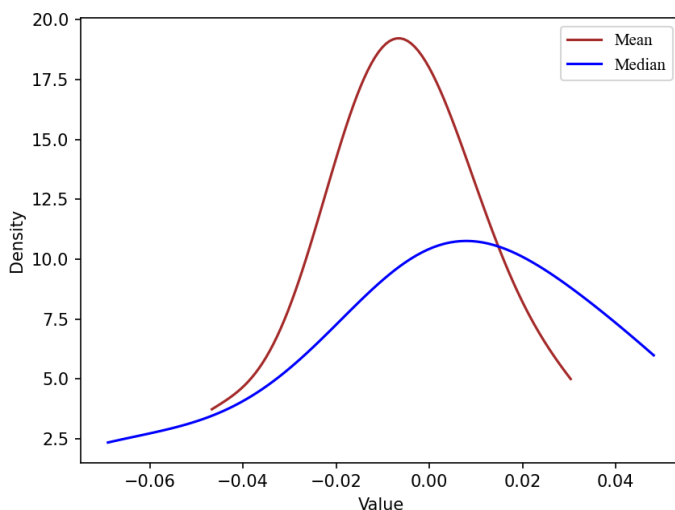

Fig. 4. A graph of the density of distributions during testing of the normal distribution.

Table 2. Evaluation of Gaussian efficiency.

	Arithmetic mean	Median
Gaussian efficiency	100%	64%

Based on the outcomes presented in Table 2, it can be contended that substituting the arithmetic mean with the median would lead to considerable dissimilarities among the experimental results.

2.2 Trimmed and winsorized mean

Let us contemplate a sample of eight sorted values, denoted as (8).

$$x = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8,\}; x_i \le x_{i+1} \tag{8}$$

In order to determine the median, it is necessary to disregard elements other than 1 and 2, which occupy the central positions. While such an approach would enhance robustness, it would result in a reduction of Gaussian efficiency in this scenario. In this instance, we eliminate the highest and lowest values and subsequently compute the trimmed arithmetic mean (9) – (10) [13].

$$x_{trimmed} = \{x_3, x_4, x_5, x_6\} \tag{9}$$

$$\overline{x_{trimmed}} = \frac{x_3 + x_4 + x_5 + x_6}{4} \tag{10}$$

Compared to the conventional arithmetic mean, the trimmed arithmetic mean offers greater robustness and Gaussian efficiency. An alternative methodology entails utilizing the winsorized mean [14], where in the minimum and maximum values are substituted (11) – (12).

$$x_{winsorized} = \{x_3, x_3, x_3, x_4, x_5, x_6, x_6, x_6\} \tag{11}$$

$$\overline{x_{winsorized}} = \frac{x_3 + x_3 + x_3 + x_4 + x_5 + x_6 + x_6 + x_6}{8} \tag{12}$$

Although both of these methods provide a foreseeable robustness measure, they are not invariably optimal since if a predetermined number of elements are trimmed or winsorized, this quantity may be either excessively small or overly large.

2.3 Exclusion of outliers

As an alternative scheme for evaluating central tendency, it may be advantageous to contemplate the prospect of excluding outliers. One of the most frequently employed methods is Tukey's criteria (13) [15].

$$[Q_{0.25} - k \cdot IQR, Q_{0.75} + k \cdot IQR] \tag{13}$$

This technique approximates a particular interval using quantiles and discards outliers that fall outside of said interval. Whilst this technique may be useful for academic evaluations of data, in practice, it is possible to mistakenly exclude regular elements that have been mislabeled as outliers, thus leading to a reduction in Gaussian efficiency. If we relax the criteria for the interval, we may begin to overlook significant outliers, which could have a detrimental impact on the robustness of the data analysis. There exist numerous methods for excluding outliers, which are discussed in detail in [3]. These techniques offer different levels of efficiency in addressing the problem.

However, defining outliers poses several challenges:
-       Outlier detection methods frequently lack stability;
-       Assessing the true performance of such methods and constructing a reliable analysis model can be challenging;
-       Outliers are an essential component of the distribution that contains valuable information for analysis.

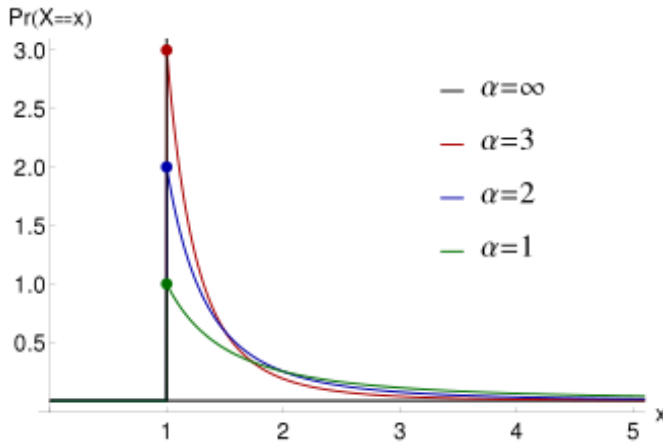Suppose that our distribution can be characterized as a Pareto distribution – Figure 5.

Fig. 5. Pareto distribution.

For this type of distribution, characterized as Paretto, the arithmetic mean will be equal to (14), which does not provide any useful information about the central tendency of this distribution. Therefore, it can be assumed that approaches based on evaluating the mean or median will not yield meaningful insights.

$$
\begin{cases}
\infty \; for \; \alpha \leq 1 \\
\dfrac{\alpha x_m}{\alpha - 1} \; for \; \alpha > 1
\end{cases}
\tag{14}
$$

2.4 Hodges-Lehmann evaluation

The formula for calculating the Hodges-Lehmann estimator is given by equation (15).

$$
HL = \underset{i<j}{median}(\frac{x_i + x_j}{2})
\tag{15}
$$

The method can be described as evaluating all pairs of numbers, calculating the arithmetic mean for each pair, and then selecting the median from the resulting arithmetic means [4].

The Hodges-Lehmann method is noted for its asymptotic Gaussian efficiency of 96%, which is practically equivalent to that of the arithmetic mean, as shown in Table 2. To evaluate the robustness, the breakdown point [18] method can be utilized, which characterizes the percentage of the sample that can be replaced with extremely large values without affecting the estimate itself. Robustness estimates according to the breakdown point for all the aforementioned methods are presented in Table 3.

Table 3. Robustness for methods of estimating the central tendency according to the breakdown point.

	Arithmetic mean	Median	Hodges-Lehmann evaluation
Breakdown point	0%	50%	29%

The arithmetic mean has a breakdown point of 0% because a single value can significantly distort the entire estimate. For the median, the breakdown point is 50%, since it is possible to arbitrarily modify half of the values in the sample without affecting the estimate. For the Hodges-Lehmann method, the breakdown point value is 29%, which is considered a reasonably good indicator of robustness. If 29% of the sample is composed of outliers, then they cannot be considered as outliers, but rather as a distinct group that requires separate analysis and description. In summary, it can be concluded that the Hodges-Lehmann method is a suitable replacement for both the arithmetic mean and median in assessing the central tendency for normal distributions and distributions of arbitrary shape.

Moreover, in order to evaluate the central tendency, the following measures can be used:
-     Modes [19];
-     Geometric mean (16);
-     Harmonic mean (17);
-     Midhinge (18) [16];
-     Trimean (19);
-     And a lot of different metrics.

$$\sqrt[n]{x_1 x_2 \ldots x_n} \tag{16}$$

$$\frac{n}{x_1^{-1} + x_2^{-1} + \ldots + x_n^{-1}} \tag{17}$$

$$\frac{(Q_{0.25}(x) + Q_{0.75}(x))}{2} \tag{18}$$

$$\frac{(Q_{0.25}(x) + 2Q_{0.50}(x) + Q_{0.75}(x))}{4} \tag{19}$$

Relevant use cases exist for each of the metrics presented. It is imperative to note that in order to select a method for assessing the central tendency, it is of utmost importance to conduct an analysis of the statistical properties of the sample at hand. By taking into account all explicit and implicit business requirements, an informed decision can be made regarding the optimal method to be employed.

3 Quantile estimates

Previously, we have examined an illustration of quantile estimation, specifically the median. This metric serves to bisect the probability density function of the distribution under investigation. In the event that a value is randomly sampled from the aforementioned distribution, it can be asserted that there is an inherent 50% likelihood of the selected value being situated within either of the two identified partitions. The subsequent prominent measures in the quantile estimation of distributions are quartiles - a set of three values that demarcate the distribution into four partitions. In addition, deciles serve as another set of quantile measures that partition the distribution into ten equivalent portions. And percentiles - which divide the distribution into one hundred equal parts [17] – Figure 6.
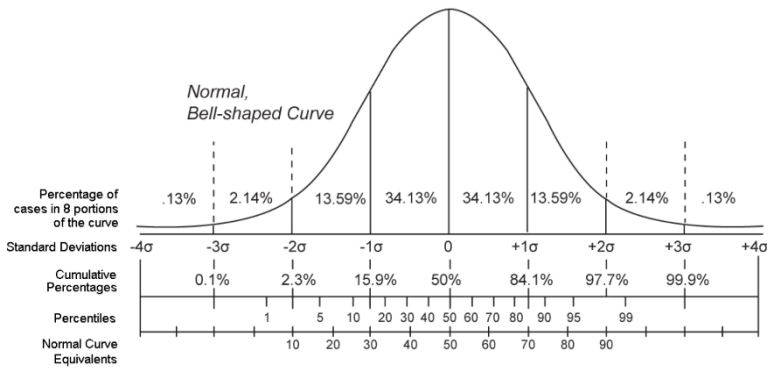

Fig. 6. Percentiles.

Quantiles serve as a broad generalization of all the previously presented definitions. A quantile of order p is a value that satisfies the property whereby a randomly sampled value from a particular distribution possesses a probability of p of being less than the value of the p -quantile, and a probability of $1-p$ of exceeding it. It is worth noting that a distinction should be made between the actual quantiles of a distribution and the quantile estimations derived from a sample thereof.

Let us contemplate a sample extracted from a normal distribution (20), and subsequently compute the quartiles via a standardized approach – Table 4.

$$x = \{-1.37, -0.19, 0.14, 0.16, 0.26, 0.39, 0.46, 0.74, 0.89\}; \qquad (20)$$

In addition to the conventional three quartiles, which coincide with the 25th, 50th, and 75th percentiles, the sample's minimum and maximum values can be included by appending the 0th and 100th percentiles, respectively. Formally, such notation would not be admissible, though it is frequently adopted nonetheless. Let us compare the computed values of these quantities with the actual quartiles of the underlying distribution, and take note of any discrepancies. It should be emphasized that the quartiles of a given distribution are fixed values and represent a characteristic feature thereof, while the quantiles derived from a sample are inherently variable across different experiments. It is important to recognize that the complete designation of these values is not merely quantiles, but rather, quantile estimates.
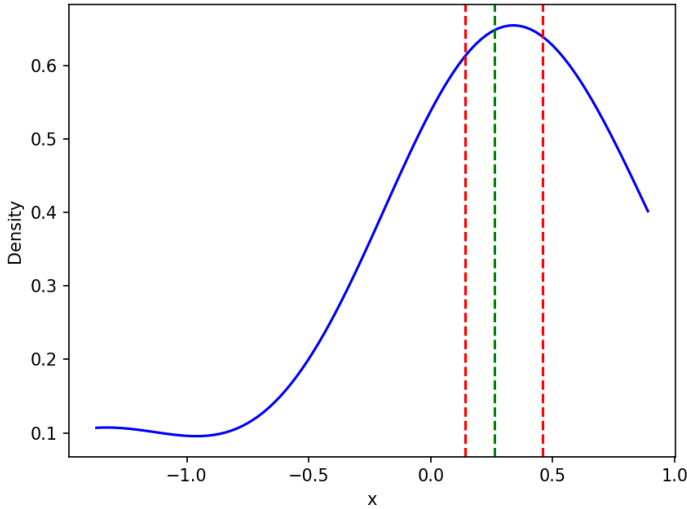
Fig. 7. Distribution of quantile numbers.

Table 4. Quantiles of distributions and quantiles of samples.

	Sample	**Distribution**
$Q_{0.00}$	−1.37	−∞
$Q_{0.25}$	0.14	−0.67
$Q_{0.50}$	0.26	0.00
$Q_{0.75}$	0.46	0.67
$Q_{1.00}$	0.89	+∞

While quantile estimates are a means of approximating the genuine quantiles of a distribution based on the available sample data, they are intrinsically limited in terms of precision due to the fact that the final sample size is generally insufficient to yield a truly accurate determination of the true values. Simultaneously, it is feasible to devise a significant number of diverse approaches to hypothesize the genuine value for this given sample. Broadly speaking, any function that maps a sample to a numerical value can be deemed as a quantile estimate. Furthermore, these functions are not fixed and may differ across various packages for statistical analysis. A comprehensive list of such functions for mapping sample values to a numerical estimate is presented in reference [5]. All formulas presented in this paper are derived from either a single ordinal statistic or a linear combination of two consecutive ordinal statistics. Here, the i ordinal statistic denotes the i element of the sample after sorting. The principal benefit of the formulas presented is their straightforwardness and efficiency, In the average case, the computation of a quantile estimate using the presented formulas has a time complexity of $O(1)$ when sample was sorted. In most contemporary statistical data analysis software, the formula that is prevalently employed is the formula numbered 7 – Figure 8.

1	$Np + 1/2$	$x_{\lceil h-1/2 \rceil}$
2	$Np + 1/2$	$(x_{\lceil h-1/2 \rceil} + x_{\lceil h+1/2 \rceil})/2$
3	Np	$x_{\lfloor h \rfloor}$
4	Np	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lceil h \rceil} - x_{\lfloor h \rfloor})$
5	$Np + 1/2$	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lceil h \rceil} - x_{\lfloor h \rfloor})$
6	$(N+1)p$	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lceil h \rceil} - x_{\lfloor h \rfloor})$
7	$(N-1)p + 1$	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lceil h \rceil} - x_{\lfloor h \rfloor})$
8	$(N+1/3)p + 1/3$	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lceil h \rceil} - x_{\lfloor h \rfloor})$
9	$(N+1/4)p + 3/8$	$x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lceil h \rceil} - x_{\lfloor h \rfloor})$

Fig. 8. Quantile estimators.

3.1 Quantile estimate of Harrell-Davis

An alternative method for estimating quantiles is the Harrell-Davis method, which was introduced in [6]. The mathematical expression for this estimate is given by equation (21).

$$Q_{HD}(p) = \sum_{i=1}^{n} W_i \cdot x_{(i)} \qquad (21)$$

This estimate is computed as a weighted sum of all ordinal statistics. The weighting coefficients are determined based on the regularized incomplete beta function (22) – (23), which is a mathematical function that arises in the context of beta distributions and has applications in various areas of statistics and probability theory.

$$W_i = I_{\frac{i}{n}}(a,b) - I_{\frac{(i-1)}{n}}(a,b) \qquad (22)$$

$$a = p(n+1), b = (1-p)(n+1) \qquad (23)$$

The primary concept underlying this estimation technique is to construct a beta distribution based on the quantile order p and the sample size n, and then divide this distribution into i segments of equal width. Here, the probability of obtaining a segment equals the probability that the p-quantile coincides with the i statistic. Consequently, the probability W is determined for each ordinal statistic, indicating the likelihood of this statistic being the desired quantile.
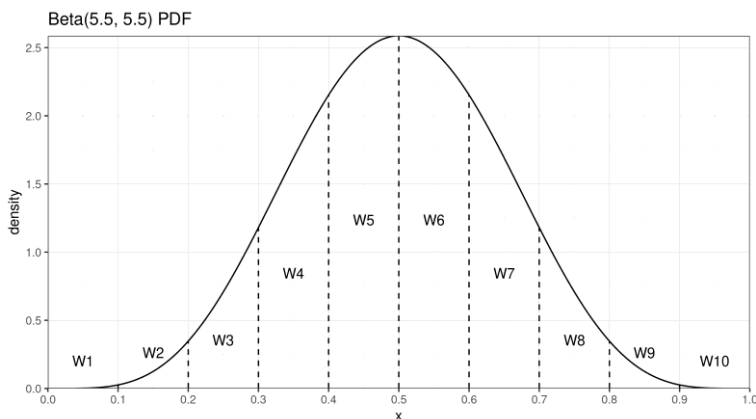
Fig. 9. Quantile estimate of Harrell-Davis.

Ultimately, a mathematical expectation is constructed for all ordinal statistics, thereby providing the ultimate estimation of quantiles. Frequently, these methodologies demonstrate efficacy in estimating the median; however, they may not always be effective in the extreme ends of distributions.

3.2 Moments of quantile estimates

Formula (24) can be utilized to calculate confidence intervals for quantile estimations. Furthermore, this formula is commonly referred to as the Maritz-Jarrett method [7].

$$C_k = \sum_{i=1}^{n} W_i \cdot x_{(i)}^k \tag{23}$$

Upon comparison with the Harrell-Davis estimate, it can be inferred that the latter is an estimation of the first moment (24). The Maritz-Jarrett method, in its turn, has demonstrated that the standard quantile error can be derived from the square root of the discrepancy between the second moment and the square of the first moment, with confidence intervals being computed based on the standard error (25).

$$Q_{HD}(p) = C_1 \tag{24}$$

$$S_{MJ}(p) = \sqrt{C_2 - C_1^2} \tag{25}$$

Nonetheless, it is pertinent to note that the selection of an appropriate quantile estimation necessitates a thorough analysis of the efficacy and robustness of various estimation techniques, as well as their suitability for specific data properties such as the median, tail elements, or exponential smoothing. Unfortunately, this aspect has not been delved into in this paper. However, if a suitable mathematical model is chosen based on the business requirements, it may be possible to investigate sliding quantile estimates in the future.

4 Conclusion

This paper focuses on a subset of statistical analyses related to performance distribution data, specifically in the evaluation of effectiveness and compliance with technical requirements for distributed systems in the context of processing external requests. The paper addressed various statistical methodologies such as central trend assessment, definitions of robustness, Gaussian efficiency, computation of censored averages, outlier exclusion, Hodges-Lehmann evaluation, and quantile estimation techniques. Additionally, the Harrell-Davis and Maritz-Jarrett methods were described in detail.Future research endeavors may include the exploration of sliding quantile estimation techniques, as well as variations in distribution densities and multimodality.

References

1. R. C. Geary, Biometrika **34(3)**, 209 (1947)
2. J. Matejka, G. Fitzmaurice, *Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing, CHI Conference on Human Factors in Computing SystemsMay* (2017), pp. 1290–1294

https://doi.org/10.1051/e3sconf/202341901029

3. H. Wang, M. J. Bah, M. Hammad, IEEE Access **7**, 107964 (2019)

4. J. L. Hodges, E. L. Lehmann, Ann. Math. Statist **34(2)**, 598 (1963)

5. R. Hyndman, F. Yanan, The American Statistician **50**, 361 (1996)

6. F. E. Harrell, C. E. Davis, Biometrika **69(3)**, 635 (1982)

7. J. S. Maritz, R. G. Jarrett, Journal of the American Statistical Association **52**, 13 (1978)

8. R. Guerraoui, J. Hamza, D. Seredinschi, M. Vukolic, *Can 100 machines agree?* (2019)

9. X. Zhou, N. Yixin, T. Hao, B. Mohit, *The curse of performance instability in analysis datasets: Consequences, source, and suggestions* (2020)

10. S. Prasad, *Measures of Central Tendencies* (2023)

11. V. A. Satopää, *Combining information from multiple forecasters: Inefficiency of central tendency* (2017)

12. D. R. Bickel, R. Frühwirth, Computational Statistics & Data Analysis **50(12)**, 3500 (2006)

13. G. Lugosi, Sh. Mendelson, *Robust multivariate mean estimation: the optimality of trimmed mean* (2021)

14. P. Orenstein, Bernoulli **28(4)**, 2862 (2022)

15. Y. She, Sh. Tang, J. Liu, *On Generalization and Computation of Tukey's Depth: Part I* (2021)

16. L. Al-Labadi, F. Forough, Z. Saberi, *A Bayesian Nonparametric Estimation of Mutual Information* (2021)

17. M. Schreiber, Scientometrics **97(3)**, 821 (2013)

18. H. P. Lopuhaä, *Highly Efficient Estimators with High Breakdown Point for Linear Models with Structured Covariance Matrices* (2023)

19. Y. Qiu, X. Wang, Journal of the American Statistical Association just-accepted **1**, 24 (2023)

20. C. Dobre, F. Pop, V. Cristea, *A simulation model for evaluating distributed systems dependability* (2012)