ISSN 2590-9770
The Art of Discrete and Applied Mathematics 3 (2020) #P2.08
https://doi.org/10.26493/2590-9770.1326.9fd
(Also available at http://adam-journal.eu)

On median and quartile sets of ordered random variables*

Iztok Banič

Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška 160, SI-2000 Maribor, Slovenia, and
Institute of Mathematics, Physics and Mechanics, Jadranska 19, SI-1000 Ljubljana, Slovenia, and
Andrej Marušič Institute, University of Primorska, Muzejski trg 2, SI-6000 Koper, Slovenia

Janez Žerovnik

Faculty of Mechanical Engineering, University of Ljubljana, Aškerčeva 6, SI-1000 Ljubljana, Slovenia, and
Institute of Mathematics, Physics and Mechanics, Jadranska 19, SI-1000 Ljubljana, Slovenia

Received 30 November 2018, accepted 13 August 2019, published online 21 August 2020

Abstract

We give new results about the set of all medians, the set of all first quartiles and the set of all third quartiles of a finite dataset. We also give new and interesting results about relationships between these sets. We also use these results to provide an elementary correctness proof of the Langford's doubling method.

Keywords: Statistics, probability, median, first quartile, third quartile, median set, first quartile set, third quartile set.

Math. Subj. Class. (2020): 62-07, 60E05, 60-08, 60A05, 62A01

1 Introduction

Quantiles play a fundamental role in statistics: they are the critical values used in hypothesis testing and interval estimation. Often they are the characteristics of distributions we usually wish to estimate. The use of quantiles as primary measure of performance has

*This work was supported in part by the Slovenian Research Agency (grants J1-8155, N1-0071, P2-0248, and J1-1693.)

E-mail addresses: iztok.banic@um.si (Iztok Banič), janez.zerovnik@fs.uni-lj.si (Janez Žerovnik)

gained prominence, particularly in microeconomic, financial and environmental analysis and others. Quartiles (i.e 0.25, 0.50, and 0,75 quantiles) are used in elementary statistics very early, c.f. for drawing box and whisker plots.

Whereas there is no dispute that the median of an ordered dataset is either the middle element or the arithmetic mean of the two middle elements (when the number of elements is even), the situation is seemingly much more complicated when quartiles are considered. There are many well-known formulas and algorithms that give certain values, claiming for these values to be medians (or quartiles) for a given statistical data (for examples see [5]). However, the trouble begins when realizing that different formulas (or algorithms) may give different values. Many authors or users of such formulas or algorithms go even further by taking the value obtained by such a formula or an algorithm to be the definition of the median or the first quartile or the third quartile of a given data. As a result, going through the literature, one may find it very difficult to find and then choose an appropriate definition (formula, algorithm) of a median or a quartile to use it for the statistical analysis of a given data. In [2, 5] provide references and comparison of several methods for computing the quartiles of a finite data set that appear in the literature and in software. While it is well known that these methods do not always give the same results, Langford writes that the *"situation is far worse than most realize "* [5]. Although the differences tend to be small, Langford further answered the question *"Why worry? The differences are small so who cares? "* with words of [1]:

> *"Before we go into any details, let us point out that the numerical differences between answers produced by the different methods are not necessarily large; indeed, they may be very small. Yet if quartiles are used, say to establish criteria for making decisions, the method of their calculation becomes of critical concern. For instance, if sales quotas are established from historical data, and salespersons in the highest quarter of the quota are to receive bonuses, while those in the lowest quarter are to be fired, establishing these boundaries is of interest to both employer and employee. In addition, computer-software users are sometimes unaware of the fact that different methods can provide different answers to their problems, and they may not know which method of calculating quartiles is actually provided by their software."*

Langford [5] also proposes a method that is consistent with the CDF (cumulative distribution function). The method is slightly more complicated than some other methods used, however it is not too much involved and there are equivalent methods that can be used in the classroom [10, 9]. Indeed, the discussion about quartiles in teaching elementary statistics is considerable, c.f. [10, 1, 4, 5, 9]. In short, some of the elementary methods are based on the idea that a quartile is a median of the lower, or the upper half of the dataset. The question arises what is the half of dataset when it has an odd number of elements. Langford naturally answers with the idea of doubling the dataset thus assuring the even number of elements, while the quantile values remain the same.

On the positive side, it seems that all methods have one thing in common: they all expect the following to hold:

1. the median to be such a value $m \in \mathbb{R}$, for which at least half of the data is less or equal to m and at least half of the data is greater or equal to m,
2. the first quartile to be such a value $q_1 \in \mathbb{R}$, for which at least quarter of the data is less or equal to q_1 and at least three quarters of the data is greater or equal to q_1,

3. the third quartile to be such a value $q_3 \in \mathbb{R}$, for which at least three quarters of the data is less or equal to q_3 and at least quarter of the data is greater or equal to q_3.

We will use this fact as a motivation to define the median set, the first quartile set, and the third quartile set of a given data.

The main contribution of this paper is the idea to redefine the median, and the quartiles, and possibly more general, the quantiles as sets (intervals) instead of the usual consideration of this notions as reals. We indicate that in this way we may avoid the dispute caused by various methods, algorithms, and even definitions of quartiles. We also show that some methods for computing the quartiles do not extend to quartile sets, and provide an elementary method that can be used to compute the quartile sets.

The rest of the paper is organized as follows. The set of all medians $M(X)$ of X is defined in Section 3, and in Section 4, the set of all first quartiles $Q_1(X)$ of X and the set of all third quartiles $Q_3(X)$ of X are defined. Main results about relationships among these sets are provided in Section 5. In Section 6, we recall some well known methods for computing of quartiles and show that one of them, the Langford's doubling method can be used to compute the quartile sets.

2 Preliminaries

Here we introduce some basic notions that we use in the paper. Suppose that we have a finite ordered m-tuple $(y_1, y_2, y_3, \ldots, y_m) \in \mathbb{R}^m$ of some data such that $y_1 < y_2 < y_3 < \ldots < y_m$, together with the m-tuple of their frequencies $(k_1, k_2, k_3, \ldots, k_m) \in \mathbb{N}^m$. This means that the datum y_i occurs k_i-times for each $i \in \{1, 2, 3, \ldots, m\}$. Let $k_1 + k_2 + k_3 + \ldots + k_m = n$. Then the random variable Y defined by

$$Y \sim \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_m \\ \frac{k_1}{n} & \frac{k_2}{n} & \frac{k_3}{n} & \ldots & \frac{k_m}{n} \end{pmatrix},$$

where $\frac{k_i}{n}$ is the probability $P(Y = y_i)$ for each $i \in \{1, 2, 3, \ldots, m\}$, represents these data.

One may represent the above data equivalently, using the random variable X in the following way

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{pmatrix},$$

where $x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$ and

$$x_1 = x_2 = x_3 = \ldots = x_{k_1} = y_1,$$

$$x_{k_1+1} = x_{k_1+2} = x_{k_1+3} = \ldots = x_{k_1+k_2} = y_2,$$

$$x_{k_1+k_2+1} = x_{k_1+k_2+2} = x_{k_1+k_2+3} = \ldots = x_{k_1+k_2+k_3} = y_3,$$

$$\vdots$$

$$x_{k_1+k_2+\ldots+k_{m-1}+1} = x_{k_1+k_2+\ldots+k_{m-1}+2} = x_{k_1+k_2+\ldots+k_{m-1}+3} = \ldots = x_n = y_m.$$

In this article, we will present data using such random variable X. We will call such a random variable X an ordered random variable.

Using this notation, we define the set of all medians $M(X)$ of X, the set of all first quartiles $Q_1(X)$ of X, and the set of all third quartiles $Q_3(X)$ of X in the following sections.

3 The median set of a random variable

We begin the section by giving the definition of a median and the median set of an ordered random variable.

Definition 3.1. Let X be an ordered random variable, given by

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix},$$

and let x be any real number. We say that x is a *median* of X, if

$$P(X \leq x) \geq \frac{1}{2} \quad \text{and} \quad P(X \geq x) \geq \frac{1}{2}.$$

We call the set

$$M(X) = \{x \in \mathbb{R} \mid x \text{ is a median of } X\}$$

the *median set* of the random variable X.

In the following proposition we give an explicit description of the median set $M(X)$ for any ordered random variable X.

Proposition 3.2. *Let X be an ordered random variable, given by*

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Then

$$M(X) = \begin{cases} \{x_k\} & \text{if } n = 2k-1 \text{ for some positive integer } k, \\ [x_k, x_{k+1}] & \text{if } n = 2k \text{ for some positive integer } k. \end{cases}$$

Proof. We consider the following two possible cases.

CASE 1: $n = 2k-1$ for some positive integer k. Since

$$P(X \leq x_k) = k \cdot \frac{1}{n} = \frac{n+1}{2} \cdot \frac{1}{n} = \frac{1}{2} + \frac{1}{2n} \geq \frac{1}{2}$$

and

$$P(X \geq x_k) = k \cdot \frac{1}{n} = \frac{n+1}{2} \cdot \frac{1}{n} = \frac{1}{2} + \frac{1}{2n} \geq \frac{1}{2},$$

it follows that $x_k \in M(X)$. Next, let $x < x_k$. Since

$$P(X \leq x) \leq P(x \leq x_{k-1}) = (k-1) \cdot \frac{1}{n} = \frac{n-1}{2} \cdot \frac{1}{n} = \frac{1}{2} - \frac{1}{2n} < \frac{1}{2},$$

therefore $x \notin M(X)$. Finally, let $x > x_k$. Since

$$P(X \geq x) \leq P(x \geq x_{k+1}) = (k-1) \cdot \frac{1}{n} = \frac{n-1}{2} \cdot \frac{1}{n} = \frac{1}{2} - \frac{1}{2n} < \frac{1}{2},$$

it follows that $x \notin M(X)$.

CASE 2: $n = 2k$ for some positive integer k and let $x \in [x_k, x_{k+1}]$. Since

$$P(X \leq x) \geq P(X \leq x_k) = k \cdot \frac{1}{n} = \frac{n}{2} \cdot \frac{1}{n} = \frac{1}{2} \geq \frac{1}{2}$$

and

$$P(X \geq x) \geq P(X \geq x_{k+1}) = k \cdot \frac{1}{n} = \frac{n}{2} \cdot \frac{1}{n} = \frac{1}{2} \geq \frac{1}{2},$$

it follows that $x \in M(X)$ for any $x \in [x_k, x_{k+1}]$. Next, let $x < x_k$. Since

$$P(X \leq x) \leq P(x \leq x_{k-1}) = (k-1) \cdot \frac{1}{n} = \frac{n-2}{2} \cdot \frac{1}{n} = \frac{1}{2} - \frac{1}{n} < \frac{1}{2},$$

therefore $x \notin M(X)$. Finally, let $x > x_{k+1}$. Since

$$P(X \geq x) \leq P(x \geq x_{k+2}) = (n-k+1) \cdot \frac{1}{n} = \frac{n-2}{2} \cdot \frac{1}{n} = \frac{1}{2} - \frac{1}{n} < \frac{1}{2},$$

therefore $x \notin M(X)$. □

Note that for any ordered random variable X,

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix},$$

the following holds:

1. the median set $M(X)$ is nonempty,
2. the median set $M(X)$ is bounded and closed in $\mathbb{R}$,
3. $\max(M(X)) = \begin{cases} x_k & \text{if } n = 2k-1 \text{ for some positive integer } k, \\ x_{k+1} & \text{if } n = 2k \text{ for some positive integer } k. \end{cases}$
4. $\min(M(X)) = \begin{cases} x_k & \text{if } n = 2k-1 \text{ for some positive integer } k, \\ x_k & \text{if } n = 2k \text{ for some positive integer } k. \end{cases}$
5. $M(X) \cap \{x_1, x_2, x_3, \ldots, x_n\} =$
$$= \begin{cases} \{x_k\} & \text{if } n = 2k-1 \text{ for some positive integer } k, \\ \{x_k, x_{k+1}\} & \text{if } n = 2k \text{ for some positive integer } k. \end{cases}$$

Clearly, the statements (1) and (2) above imply

Fact 3.3. *The median set $M(X)$ is either a singleton (one real number) or a closed interval.*

We call the maximum $\max(M(X))$ of $M(X)$ the *upper median* of X and we will always denote it by m^1; we call the minimum $\min(M(X))$ of $M(X)$ the *lower median* of X and we will always denote it by m^0. The median

$$m^{\frac{1}{2}} = \frac{\min(M(X)) + \max(M(X))}{2}$$
$$= \begin{cases} x_k & \text{if } n = 2k-1 \text{ for some positive integer } k, \\ \frac{x_k + x_{k+1}}{2} & \text{if } n = 2k \text{ for some positive integer } k \end{cases}$$

will be called the *middle median* of X or the *canonical value* of median of X.

4 The first and the third quartile sets of a random variable

We begin this section by giving the definition of a first and a third quartile as well as the first quartile and the third quartile set of an ordered random variable.

Definition 4.1. Let X be an ordered random variable, given by

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix},$$

and let x be any real number. We say that x is

1. a *first quartile* of X, if

$$P(X \leq x) \geq \frac{1}{4} \quad \text{and} \quad P(X \geq x) \geq \frac{3}{4}.$$

2. a *third quartile* of X, if

$$P(X \leq x) \geq \frac{3}{4} \quad \text{and} \quad P(X \geq x) \geq \frac{1}{4}.$$

We call the set

$$Q_1(X) = \{x \in \mathbb{R} \mid x \text{ is a first quartile of } X\}$$

the *first quartile set* of the random variable X and the set

$$Q_3(X) = \{x \in \mathbb{R} \mid x \text{ is a third quartile of } X\}$$

the *third quartile set* of the random variable X.

In the following proposition we give an explicit description of the sets $Q_1(X)$ and $Q_2(X)$ for any ordered random variable X.

Proposition 4.2. *Let X be an ordered random variable, given by*

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Then

$$Q_1(X) = \begin{cases} [x_k, x_{k+1}] & \text{if } n = 4k \text{ for some positive integer } k, \\ \{x_{k+1}\} & \text{if } n = 4k+1 \text{ for some non-negative integer } k, \\ \{x_{k+1}\} & \text{if } n = 4k+2 \text{ for some non-negative integer } k, \\ \{x_{k+1}\} & \text{if } n = 4k+3 \text{ for some non-negative integer } k \end{cases}$$

and

$$Q_3(X) = \begin{cases} [x_{3k}, x_{3k+1}] & \text{if } n = 4k \text{ for some positive integer } k, \\ \{x_{3k+1}\} & \text{if } n = 4k+1 \text{ for some non-negative integer } k, \\ \{x_{3k+2}\} & \text{if } n = 4k+2 \text{ for some non-negative integer } k, \\ \{x_{3k+3}\} & \text{if } n = 4k+3 \text{ for some non-negative integer } k. \end{cases}$$

Proof. We consider the following four possible cases.

CASE 1: $n = 4k$ for some positive integer k.

First we find all such $\ell \in \{1, 2, 3, \ldots, n\}$ that $x_\ell \in Q_1(X)$. Suppose that $\ell \in \{1, 2, 3, \ldots, n\}$ is such an integer that $x_\ell \in Q_1(X)$. Then

- $\frac{\ell}{4k} \geq \frac{1}{4}$ holds and
$$\frac{\ell}{4k} \geq \frac{1}{4} \iff \ell \geq k,$$
- $\frac{4k - \ell + 1}{4k} \geq \frac{3}{4}$ holds and
$$\frac{4k - \ell + 1}{4k} \geq \frac{3}{4} \iff \ell \leq k + 1.$$

Therefore,
$$x_\ell \in Q_1(X) \iff \ell \in \{k, k + 1\}.$$
Therefore, it can easily be seen that $Q_1(X) = [x_k, x_{k+1}]$.

Next we find all such $\ell \in \{1, 2, 3, \ldots, n\}$ that $x_\ell \in Q_3(X)$. Suppose that $\ell \in \{1, 2, 3, \ldots, n\}$ is such an integer that $x_\ell \in Q_3(X)$. Then

- $\frac{\ell}{4k} \geq \frac{3}{4}$ holds and
$$\frac{\ell}{4k} \geq \frac{3}{4} \iff \ell \geq 3k,$$
- $\frac{4k - \ell + 1}{4k} \geq \frac{1}{4}$ holds and
$$\frac{4k - \ell + 1}{4k} \geq \frac{1}{4} \iff \ell \leq 3k + 1.$$

Therefore,
$$x_\ell \in Q_3(X) \iff \ell \in \{3k, 3k + 1\}.$$
Therefore, it can easily be seen that $Q_3(X) = [x_{3k}, x_{3k+1}]$.

CASE 2: $n = 4k + 1$ for some non-negative integer k.

First we find all such $\ell \in \{1, 2, 3, \ldots, n\}$ that $x_\ell \in Q_1(X)$. Suppose that $\ell \in \{1, 2, 3, \ldots, n\}$ is such an integer that $x_\ell \in Q_1(X)$. Then

- $\frac{\ell}{4k + 1} \geq \frac{1}{4}$ holds and
$$\frac{\ell}{4k + 1} \geq \frac{1}{4} \iff \ell \geq k + \frac{1}{4},$$
- $\frac{4k + 1 - \ell + 1}{4k + 1} \geq \frac{3}{4}$ holds and
$$\frac{4k - \ell + 2}{4k + 1} \geq \frac{3}{4} \iff \ell \leq k + \frac{5}{4}.$$

Therefore,

$$x_\ell \in Q_1(X) \iff \ell = k+1.$$

Therefore, it can easily be seen that $Q_1(X) = \{x_{k+1}\}$.

Next we find all such $\ell \in \{1,2,3,\ldots,n\}$ that $x_\ell \in Q_3(X)$. Suppose that $\ell \in \{1,2,3,\ldots,n\}$ is such an integer that $x_\ell \in Q_3(X)$. Then

- $\dfrac{\ell}{4k+1} \geq \dfrac{3}{4}$ holds and

$$\frac{\ell}{4k+1} \geq \frac{3}{4} \iff \ell \geq 3k + \frac{3}{4},$$

- $\dfrac{4k+1-\ell+1}{4k+1} \geq \dfrac{1}{4}$ holds and

$$\frac{4k-\ell+2}{4k+1} \geq \frac{1}{4} \iff \ell \leq 3k + \frac{7}{4}.$$

Therefore,

$$x_\ell \in Q_3(X) \iff \ell = 3k+1.$$

Therefore, it can easily be seen that $Q_3(X) = \{x_{3k+1}\}$.

CASE 3: $n = 4k+2$ for some non-negative integer k.

First we find all such $\ell \in \{1,2,3,\ldots,n\}$ that $x_\ell \in Q_1(X)$. Suppose that $\ell \in \{1,2,3,\ldots,n\}$ is such an integer that $x_\ell \in Q_1(X)$. Then

- $\dfrac{\ell}{4k+2} \geq \dfrac{1}{4}$ holds and

$$\frac{\ell}{4k+2} \geq \frac{1}{4} \iff \ell \geq k + \frac{1}{2},$$

- $\dfrac{4k+2-\ell+1}{4k+2} \geq \dfrac{3}{4}$ holds and

$$\frac{4k-\ell+3}{4k+2} \geq \frac{3}{4} \iff \ell \leq k + \frac{3}{2}.$$

Therefore,

$$x_\ell \in Q_1(X) \iff \ell = k+1.$$

Therefore, it can easily be seen that $Q_1(X) = \{x_{k+1}\}$.

Next we find all such $\ell \in \{1,2,3,\ldots,n\}$ that $x_\ell \in Q_3(X)$. Suppose that $\ell \in \{1,2,3,\ldots,n\}$ is such an integer that $x_\ell \in Q_3(X)$. Then

- $\dfrac{\ell}{4k+2} \geq \dfrac{3}{4}$ holds and

$$\frac{\ell}{4k+2} \geq \frac{3}{4} \iff \ell \geq 3k + \frac{3}{2},$$

- $\dfrac{4k+2-\ell+1}{4k+2} \geq \dfrac{1}{4}$ holds and

$$\frac{4k-\ell+3}{4k+2} \geq \frac{1}{4} \iff \ell \leq 3k+\frac{5}{2}.$$

Therefore,

$$x_\ell \in Q_3(X) \iff \ell = 3k+2.$$

Therefore, it can easily be seen that $Q_3(X) = \{x_{3k+2}\}$.

CASE 4: $n = 4k+3$ for some non-negative integer k.

First we find all such $\ell \in \{1,2,3,\ldots,n\}$ that $x_\ell \in Q_1(X)$. Suppose that $\ell \in \{1,2,3,\ldots,n\}$ is such an integer that $x_\ell \in Q_1(X)$. Then

- $\dfrac{\ell}{4k+3} \geq \dfrac{1}{4}$ holds and

$$\frac{\ell}{4k+3} \geq \frac{1}{4} \iff \ell \geq k+\frac{3}{4},$$

- $\dfrac{4k+3-\ell+1}{4k+3} \geq \dfrac{3}{4}$ holds and

$$\frac{4k-\ell+4}{4k+3} \geq \frac{3}{4} \iff \ell \leq k+\frac{7}{4}.$$

Therefore,

$$x_\ell \in Q_1(X) \iff \ell = k+1.$$

Therefore, it can easily be seen that $Q_1(X) = \{x_{k+1}\}$.

Finally, we find all such $\ell \in \{1,2,3,\ldots,n\}$ that $x_\ell \in Q_3(X)$. Suppose that $\ell \in \{1,2,3,\ldots,n\}$ is such an integer that $x_\ell \in Q_3(X)$. Then

- $\dfrac{\ell}{4k+3} \geq \dfrac{3}{4}$ holds and

$$\frac{\ell}{4k+3} \geq \frac{3}{4} \iff \ell \geq 3k+\frac{9}{4},$$

- $\dfrac{4k+3-\ell+1}{4k+3} \geq \dfrac{1}{4}$ holds and

$$\frac{4k-\ell+4}{4k+3} \geq \frac{1}{4} \iff \ell \leq 3k+\frac{13}{4}.$$

Therefore,

$$x_\ell \in Q_3(X) \iff \ell = 3k+3.$$

Therefore, it can easily be seen that $Q_3(X) = \{x_{3k+3}\}$. □

Note that for any ordered random variable X,

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix},$$

the following holds:

1. the sets $Q_1(X)$ and $Q_3(X)$ are both nonempty,
2. the sets $Q_1(X)$ and $Q_3(X)$ are both bounded and closed in $\mathbb{R}$,
3. $\max(Q_1(X)) = \begin{cases} x_{k+1} & \text{if } n = 4k \text{ for some positive integer } k, \\ x_{k+1} & \text{if } n = 4k+1 \text{ for some non-negative integer } k, \\ x_{k+1} & \text{if } n = 4k+2 \text{ for some non-negative integer } k, \\ x_{k+1} & \text{if } n = 4k+3 \text{ for some non-negative integer } k \end{cases}$
4. $\min(Q_1(X)) = \begin{cases} x_{k} & \text{if } n = 4k \text{ for some positive integer } k, \\ x_{k+1} & \text{if } n = 4k+1 \text{ for some non-negative integer } k, \\ x_{k+1} & \text{if } n = 4k+2 \text{ for some non-negative integer } k, \\ x_{k+1} & \text{if } n = 4k+3 \text{ for some non-negative integer } k \end{cases}$
5. $\max(Q_3(X)) = \begin{cases} x_{3k+1} & \text{if } n = 4k \text{ for some positive integer } k, \\ x_{3k+1} & \text{if } n = 4k+1 \text{ for some non-negative integer } k, \\ x_{3k+2} & \text{if } n = 4k+2 \text{ for some non-negative integer } k, \\ x_{3k+3} & \text{if } n = 4k+3 \text{ for some non-negative integer } k \end{cases}$
6. $\min(Q_3(X)) = \begin{cases} x_{3k} & \text{if } n = 4k \text{ for some positive integer } k, \\ x_{3k+1} & \text{if } n = 4k+1 \text{ for some non-negative integer } k, \\ x_{3k+2} & \text{if } n = 4k+2 \text{ for some non-negative integer } k, \\ x_{3k+3} & \text{if } n = 4k+3 \text{ for some non-negative integer } k \end{cases}$
7. $Q_1(X) \cap \{x_1, x_2, x_3, \ldots, x_n\} =$
$$= \begin{cases} \{x_k, x_{k+1}\} & \text{if } n = 4k \text{ for some positive integer } k, \\ \{x_{k+1}\} & \text{if } n = 4k+1 \text{ for some non-negative integer } k, \\ \{x_{k+1}\} & \text{if } n = 4k+2 \text{ for some non-negative integer } k, \\ \{x_{k+1}\} & \text{if } n = 4k+3 \text{ for some non-negative integer } k \end{cases}$$
8. $Q_3(X) \cap \{x_1, x_2, x_3, \ldots, x_n\} =$
$$= \begin{cases} \{x_{3k}, x_{3k+1}\} & \text{if } n = 4k \text{ for some positive integer } k, \\ \{x_{3k+1}\} & \text{if } n = 4k+1 \text{ for some non-negative integer } k, \\ \{x_{3k+2}\} & \text{if } n = 4k+2 \text{ for some non-negative integer } k, \\ \{x_{3k+3}\} & \text{if } n = 4k+3 \text{ for some non-negative integer } k \end{cases}$$

Similarly as for the median, we observe that

Fact 4.3. *The quartile sets $Q_1(X)$ and $Q_3(X)$ are either singletons (one real number) or closed intervals.*

We call the maximum $\max(Q_1(X))$ and the minimum $\min(Q_1(X))$ of $Q_1(X)$ the *upper first quartile* and the *lower first quartile* of X respectively, and we will denote them by q_1^1 and q_1^0 respectively. The first quartile

$$q_1^{\frac{1}{2}} = \frac{\min(Q_1(X)) + \max(Q_1(X))}{2} = \begin{cases} \frac{x_k + x_{k+1}}{2} & \text{if } n = 4k \text{ for some positive integer } k, \\ x_{k+1} & \text{if } n = 4k+1 \text{ for some non-negative integer } k, \\ x_{k+1} & \text{if } n = 4k+2 \text{ for some non-negative integer } k, \\ x_{k+1} & \text{if } n = 4k+3 \text{ for some non-negative integer } k \end{cases}$$

will be called the *middle first quartile* of X (or, the *canonical value* of the first quartile).

We call the maximum $\max(Q_3(X))$ and the minimum $\min(Q_3(X))$ of $Q_3(X)$ the *upper third quartile* and the *lower third quartile* of X respectively, and we will always denote them by q_3^1 and q_3^0 respectively. The third quartile

$$q_3^{\frac{1}{2}} = \frac{\min(Q_3(X)) + \max(Q_3(X))}{2} = \begin{cases} \frac{x_{3k} + x_{3k+1}}{2} & \text{if } n = 4k \text{ for some positive integer } k, \\ x_{3k+1} & \text{if } n = 4k+1 \text{ for some non-negative integer } k, \\ x_{3k+2} & \text{if } n = 4k+2 \text{ for some non-negative integer } k, \\ x_{3k+3} & \text{if } n = 4k+3 \text{ for some non-negative integer } k \end{cases}$$

will be called the *middle third quartile* of X (or, the *canonical value* of the third quartile).

5 Main results

In present section we formulate and prove our main theorems. We start with the following definition.

Definition 5.1. Let X be an ordered random variable, given by

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Then $2X$ is the ordered random variable, defined by

$$2X \sim \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_{2n} \\ \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} \end{pmatrix},$$

where $y_{2i-1} = y_{2i} = x_i$ for each $i \in \{1, 2, 3, \ldots, n\}$.

The following theorem says that the set of all medians of X may be obtained by calculating the set of all medians of $2X$.

Theorem 5.2. *Let X be an ordered random variable, given by*

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Then $M(X) = M(2X)$.

Proof. Let

$$2X \sim \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_{2n} \\ \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} \end{pmatrix},$$

We look at the following two possible cases.

CASE 1: $n = 2k - 1$ for some positive integer k.
By Proposition 3.2 and by the definition of $2X$, the following holds:

$$M(2X) = [y_{2k-1}, y_{2k}] = [x_k, x_k] = \{x_k\} = M(X).$$

CASE 2: $n = 2k$ for some positive integer k.
By Proposition 3.2 and by the definition of $2X$, the following holds:

$$M(2X) = [y_{2k}, y_{2k+1}] = [x_k, x_{k+1}] = M(X).$$

□

In the following theorem, the ordered random variable $4X$ is defined to be the ordered random variable $2(2X)$. The theorem says that the set of all first (third) quartiles of X may be obtained by calculating the set of all first (third) quartiles of $4X$.

Theorem 5.3. *Let X be an ordered random variable, given by*

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Then $Q_1(X) = Q_1(4X)$ and $Q_3(X) = Q_3(4X)$.

Proof. Let

$$4X \sim \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_{4n} \\ \frac{1}{4n} & \frac{1}{4n} & \frac{1}{4n} & \cdots & \frac{1}{4n} \end{pmatrix},$$

We look at the following four possible cases.

CASE 1: $n = 4k$ for some positive integer k.
By Proposition 4.2 and by the definition of $4X$, the following holds:

$$Q_1(4X) = [y_n, y_{n+1}] = [x_k, x_{k+1}] = Q_1(X)$$

and

$$Q_3(4X) = [y_{3n}, y_{3n+1}] = [x_{3k}, x_{3k+1}] = Q_3(X).$$

CASE 2: $n = 4k + 1$ for some non-negative integer k.
By Proposition 4.2 and by the definition of $4X$, the following holds:

$$Q_1(4X) = [y_n, y_{n+1}] = [x_{k+1}, x_{k+1}] = \{x_{k+1}\} = Q_1(X)$$

and

$$Q_3(4X) = [y_{3n}, y_{3n+1}] = [x_{3k+1}, x_{3k+1}] = \{x_{3k+1}\} = Q_3(X).$$

CASE 3: $n = 4k + 2$ for some non-negative integer k.

By Proposition 4.2 and by the definition of $4X$, the following holds:

$$Q_1(4X) = [y_n, y_{n+1}] = [x_{k+1}, x_{k+1}] = \{x_{k+1}\} = Q_1(X)$$

and

$$Q_3(4X) = [y_{3n}, y_{3n+1}] = [x_{3k+2}, x_{3k+2}] = \{x_{3k+2}\} = Q_3(X).$$

CASE 4: $n = 4k + 3$ for some non-negative integer k.

By Proposition 4.2 and by the definition of $4X$, the following holds:

$$Q_1(4X) = [y_n, y_{n+1}] = [x_{k+1}, x_{k+1}] = \{x_{k+1}\} = Q_1(X)$$

and

$$Q_3(4X) = [y_{3n}, y_{3n+1}] = [x_{3k+3}, x_{3k+3}] = \{x_{3k+3}\} = Q_3(X).$$

□

In the definitions and the results that follow we try to mimic statistical methods that suggest the following well-known strategy. To find a first or a third quartile, split the data into two halves and find the medians of these halves.

Definition 5.4. Let X be an ordered random variable, given by

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_{2n} \\ \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} \end{pmatrix}.$$

Then $\frac{1}{2}X^-$ is the ordered random variable, given by

$$\frac{1}{2}X^- \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

and $\frac{1}{2}X^+$ is the ordered random variable, given by

$$\frac{1}{2}X^+ \sim \begin{pmatrix} x_{n+1} & x_{n+2} & x_{n+3} & \cdots & x_{2n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

We continue with the following theorem which gives a relationship between $M(\frac{1}{2}X^-)$ and $Q_1(X)$, and $M(\frac{1}{2}X^+)$ and $Q_3(X)$.

Theorem 5.5. *Let X be an ordered random variable, given by*

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_{2n} \\ \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} \end{pmatrix}.$$

Then $M(\frac{1}{2}X^-) = Q_1(X)$ and $M(\frac{1}{2}X^+) = Q_3(X)$.

Proof. We look at the following two possible cases.

CASE 1: $n = 2k - 1$ for some positive integer k.

By Propositions 3.2 and 4.2, and by the definition of $\frac{1}{2}X^-$ and $\frac{1}{2}X^+$, the following holds:

$$M(\frac{1}{2}X^-) = \{x_k\} = Q_1(X)$$

and

$$M(\frac{1}{2}X^+) = \{x_{n+k}\} = \{x_{3k-1}\} = Q_3(X).$$

CASE 2: $n = 2k$ for some positive integer k.

By Propositions 3.2 and 4.2, and by the definition of $\frac{1}{2}X^-$ and $\frac{1}{2}X^+$, the following holds:

$$M(\frac{1}{2}X^-) = [x_k, x_{k+1}] = Q_1(X)$$

and

$$M(\frac{1}{2}X^+) = [x_{n+k}, x_{n+k+1}] = [x_{3k}, x_{3k+1}] = Q_3(X).$$

□

Note that $\frac{1}{2}X^-$ and $\frac{1}{2}X^+$ can only be obtained if $n = 2k$ for some positive integer k. The following definition generalizes the notion of $\frac{1}{2}X^-$ and $\frac{1}{2}X^+$ to define the lower and upper parts of X in any proportion for arbitrary n.

Definition 5.6. Let X be an ordered random variable, given by

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

and let $x \in [x_1, x_n]$ be any real number. Then we define the ordered random variables L_x^c, L_x^o, U_x^c, and U_x^o by

$$L_x^c \sim \begin{cases} \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_k \\ \frac{1}{k} & \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \end{pmatrix} & \text{if } x = x_k \text{ for some } k, \\ \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_k \\ \frac{1}{k} & \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \end{pmatrix} & \text{if } x_k < x < x_{k+1} \text{ for some } k \end{cases}$$

$$L_x^o \sim \begin{cases} \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_{k-1} \\ \frac{1}{k-1} & \frac{1}{k-1} & \frac{1}{k-1} & \cdots & \frac{1}{k-1} \end{pmatrix} & \text{if } x = x_k \text{ for some } k, \\ \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_k \\ \frac{1}{k} & \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \end{pmatrix} & \text{if } x_k < x < x_{k+1} \text{ for some } k \end{cases}$$

$$U_x^c \sim \begin{cases} \begin{pmatrix} x_k & x_{k+1} & x_{k+2} & \cdots & x_n \\ \frac{1}{n-k+1} & \frac{1}{n-k+1} & \frac{1}{n-k+1} & \cdots & \frac{1}{n-k+1} \end{pmatrix} & \text{if } x = x_k \text{ for some } k, \\ \begin{pmatrix} x_{k+1} & x_{k+2} & \cdots & x_n \\ \frac{1}{n-k} & \frac{1}{n-k} & \cdots & \frac{1}{n-k} \end{pmatrix} & \text{if } x_k < x < x_{k+1} \text{ for some } k. \end{cases}$$

$$U_x^o \sim \begin{cases} \begin{pmatrix} x_{k+1} & x_{k+2} & \cdots & x_n \\ \frac{1}{n-k} & \frac{1}{n-k} & \cdots & \frac{1}{n-k} \end{pmatrix} & \text{if } x = x_k \text{ for some } k, \\ \begin{pmatrix} x_{k+1} & x_{k+2} & \cdots & x_n \\ \frac{1}{n-k} & \frac{1}{n-k} & \cdots & \frac{1}{n-k} \end{pmatrix} & \text{if } x_k < x < x_{k+1} \text{ for some } k. \end{cases}$$

The sets L_x^o, L_x^c, U_x^o, and U_x^c can respectively be called open and closed lower part, and open and closed upper parts of X relative to x.

From the definitions it directly follows:

Proposition 5.7. *Let X be an ordered random variable, given by*

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Then

1. *$L_x^c \supseteq L_x^o$, $U_x^c \supseteq U_x^o$ for any $x \in [x_1, x_n]$,*
2. *$L_x^o \cap U_x^o = \emptyset$ for any $x \in [x_1, x_n]$,*
3. *if $x = x_k \in X$ then $L_x^c \cap U_x^c = \{x\}$,*
4. *if $x \neq x_k \in X$ then $L_x^o \cup U_x^o = X$,*
5. *$L_x^c \cup U_x^c = X$ for any $x \in [x_1, x_n]$.*

Furthermore, the following theorem holds.

Theorem 5.8. *Let X be an ordered random variable, given by*

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

If $n = 2k$ for some k, then for any median $m \in M(X)$, $m \neq m^0$, $m \neq m^1$, we have
(1) $L_m^c = L_m^o = \frac{1}{2}X^-$ and $U_m^c = U_m^o = \frac{1}{2}X^+$.
(2) $M(L_m^c) = M(L_m^o) = Q_1(X)$ and $M(U_m^c) = M(U_m^o) = Q_3(X)$.

Proof. Statement (1) follows directly from the definitions. Statement (2) follows from (1) and Theorem 5.5. □

The situation is a bit more complicated for odd n. Recall that for odd number of elements $n = 2\ell + 1$, the median $m = x_{\ell+1}$ is an element of X.

Theorem 5.9. *Let n be an odd integer and X be an ordered random variable, given by*

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Then
(1) if $n = 4k+1$ then for the unique median $m = x_{2k+1}$ we have $M(L_m^c) = Q_1(X) = \{x_{k+1}\} \subseteq M(L_m^o) = [x_k, x_{k+1}]$ and $M(L_m^c) = Q_3(X) = \{x_{3k+1}\} \subseteq M(L_m^o) = [x_{k+1}, x_{k+2}]$.
(2) if $n = 4k+3$ then for the unique median $m = x_{2k+2}$ we have $M(L_m^o) = Q_1(X) = \{x_{k+1}\} \subseteq M(L_m^c) = [x_{k+1}, x_{k+2}]$ and $M(L_m^o) = Q_3(X) = \{x_{3k+3}\} \subseteq M(L_m^o) = [x_{3k+2}, x_{3k+3}]$.

Proof. The proof is straight forward. We leave it to a reader. □

Thus from Theorem 5.8 we have learned that for X with even number of elements, taking any value from the median set to divide X to obtain the lower and the upper half, and computing its median sets will provide exact values of the first and the third quartile sets.

However, by Theorem 5.9, the situation is slightly more complicated for odd n. Two cases have to be distinguished, because the quartile sets are median sets of the open halves when $n = 4k + 1$ and are medians of the closed halves when $n = 4k + 3$.

We conclude the section by stating and proving another interesting result not depending whether n is even or odd. It gives an algorithm how to obtain the first and the third quartile sets of any data by doubling the data first, and then obtaining the median sets of the first and the second halves of the obtained doubled data. The advantage of this method is the fact that it works perfectly in both cases — for any even and for any odd n.

Theorem 5.10. *Let X be an ordered random variable, given by*

$$X \sim \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Then $M(\frac{1}{2}(2X)^-) = Q_1(X)$ and $M(\frac{1}{2}(2X)^+) = Q_3(X)$.

Proof. We distinguish the following four possible cases.

CASE 1: $n = 4k$ for some positive integer k.

By Proposition 4.2, $Q_1(X) = [x_k, x_{k+1}]$ and $Q_3(X) = [x_{3k}, x_{3k+1}]$.

In this case

$$2X \sim \begin{pmatrix} x_1 & x_1 & \cdots & x_{2k} & x_{2k} & x_{2k+1} & x_{2k+1} & \cdots & x_{n-1} & x_n & x_n \\ \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} \end{pmatrix}.$$

By Proposition 3.2, one can easily get that $M(\frac{1}{2}(2X)^-) = [x_k, x_{k+1}] = Q_1(X)$ and $M(\frac{1}{2}(2X)^+) = [x_{3k}, x_{3k+1}] = Q_3(X)$.

CASE 2: $n = 4k + 1$ for some non-negative integer k.

By Proposition 4.2, $Q_1(X) = \{x_{k+1}\}$ and $Q_3(X) = \{x_{3k+1}\}$, and by Proposition 3.2, $M(X) = \{x_{2k+1}\}$.

In this case

$$2X \sim \begin{pmatrix} x_1 & x_1 & x_2 & \cdots & x_{2k} & x_{2k+1} & x_{2k+1} & x_{2k+2} & \cdots & x_n & x_n \\ \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} & \frac{1}{2n} \end{pmatrix}.$$

By Proposition 3.2, $M(\frac{1}{2}(2X)^-) = \{x_{k+1}\} = Q_1(X)$ and $M(\frac{1}{2}(2X)^+) = \{x_{3k+1}\} = Q_3(X)$.

CASE 3: $n = 4k + 2$ for some non-negative integer k.

By Proposition 4.2, $Q_1(X) = \{x_{k+1}\}$ and $Q_3(X) = \{x_{3k+2}\}$.

In this case

$$2X \sim \begin{pmatrix} x_1 & x_1 & x_2 & \cdots & x_{2k+1} & x_{2k+2} & \cdots & x_{n-1} & x_n & x_n \\ \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} \end{pmatrix}.$$

By Proposition 3.2, $M(\frac{1}{2}(2X)^-) = \{x_{k+1}\} = Q_1(X)$ and $M(\frac{1}{2}(2X)^+) = \{x_{3k+2}\} = Q_3(X)$.

CASE 4: $n = 4k + 3$ for some non-negative integer k.

By Proposition 4.2, $Q_1(X) = \{x_{k+1}\}$ and $Q_3(X) = \{x_{3k+3}\}$.

In this case

$$2X \sim \begin{pmatrix} x_1 & x_1 & x_2 & x_2 & \cdots & x_{2k+2} & x_{2k+2} & \cdots & x_n & x_n \\ \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} & \frac{1}{2n} & \cdots & \frac{1}{2n} & \frac{1}{2n} \end{pmatrix}.$$

By Proposition 3.2, $M(\frac{1}{2}(2X)^-) = \{x_{k+1}\} = Q_1(X)$ and $M(\frac{1}{2}(2X)^+) = \{x_{3k+3}\} = Q_3(X)$. □

6 On some elementary methods for computing the quartiles

The usual methods for computation of quartiles are based on the idea to split the dataset in two halves and obtain the quartiles as the medians of the halves. The obvious question arises "how to define the halves if the number of elements is odd ?". As we know it is answered differently, yielding different methods and, unfortunately, different results(!) [5]. Three methods are among the most popular, the first two being often used in elementary textbooks. The third was proposed in [5] and argued to be accessible at elementary level in [10]. All the methods below first compute the median of X and then divide X in two halves to obtain the quartiles as medians of the halves. However, when n is odd, the methods differ as follows:

- Method M1. Include the median in both halves.
- Method M2. Exclude the median in both halves.
- Method L. If $n = 4k + 1$ then include the median. If $n = 4k + 3$ then exclude the median.

Method L was suggested by Langford [5] who shows that both M1 and M2 fail to provide correct answers in some cases.

We say that a method or an algorithm for computing a first quartile of a given data is correct, if it gives a value q and $q \in Q_1(X)$. We say that a method or an algorithm for computing a third quartile of a given data is correct, if it gives a value q and $q \in Q_3(X)$.

Considering Theorem 5.9 immediately confirms that M1 and M2 are not correct. For example, for $n = 4k+3$, method M1 gives q_1 as the median of the lowest $2k+2$ elements, i.e. $\frac{1}{2}(x_{k+1} + x_{k+2})$ whereas $Q_1(X) = \{x_{k+1}\}$. Similarly, for $n = 4k + 1$, method M2 gives q_1 as the median of the lowest $2k$ elements, i.e. $\frac{1}{2}(x_k + x_{k+1})$ whereas $Q_1(X) = \{x_{k+1}\}$.

Method L however naturally extends to the general case.

Theorem 6.1. *The L method is a correct algorithm for computing the quartile sets.*

Proof. Let n be even, say $n = 2k$. Then by method L, the first quartile is the median of the set $\{x_1, x_2, \ldots, x_k\}$, and the third quartile is the median of the set $\{x_{k+1}, x_{k+2}, \ldots, x_{2k}\}$, which is correct by Theorem 5.5.

Let n be odd. If $n = 4k + 1$ then by method L, the first quartile is the median of the set $\{x_1, x_2, \ldots, x_{2k+1}\}$, and the third quartile is the median of the set $\{x_{2k+1}, x_{2k+2}, \ldots, x_{4k+1}\}$, (median included in both sets), which is correct by Theorems 5.8 and 5.9.

If $n = 4k + 3$ then by method L, the first quartile is the median of the set $\{x_1, x_2, \ldots, x_{2k+1}\}$, and the third quartile is the median of $\{x_{2k+3}, x_{2k+4}, \ldots, x_{4k+3}\}$, (median excluded from both sets), which is correct by Theorems 5.8 and 5.9. □

Another natural idea [5], equivalent to method L, can naturally be extended to a method for computing the quartile sets. Instead of asking and to answering the question whether to include or exclude the median when splitting the dataset in two halves, one can decide to give "half of the median" to each part. This can be realized by doubling the dataset and giving one copy of the median into each half. We call this the Langford's doubling method. Recall that Theorem 5.5 implies that this method works correctly for the generalized definition of quartiles.

Theorem 6.2. *The doubling method is a correct algorithm for computing the quartile sets.*

In conclusion, one may ask how some other methods for computing quartiles are related to the generalized notion of median and quartiles. For example, assuming $n = 4k$, one could ask whether a method of interest gives quartile values that are within the quartile set. This may be a good evidence that the method is sound.

Finally, we wish to note that the interval sets can be naturally associated with any quantiles, and an analogous theory may be developed.

ORCID iDs

Iztok Banič https://orcid.org/0000-0002-5097-2903
Janez Žerovnik https://orcid.org/0000-0002-6041-1106

References

[1] J. E. Freund and B. M. Perles, A new look at quartiles of ungrouped data, *The American Statistician* **41** (1987), 200–203, doi:10.1080/00031305.1987.10475479.

[2] R. Hyndman and Y. Fan, Sample quantiles in statistical packages, *The American Statistician* **50** (1996), 361–365, doi:10.1080/00031305.1996.10473566.

[3] C. Jentsch and A. Leucht, Bootstrapping sample quantiles of discrete data, 2014, https://madoc.bib.uni-mannheim.de/36588/.

[4] A. H. Joarder and M. Firozzaman, Quartiles for discrete data, *Teaching Statistics* **23**, 86–89, doi:10.1111/1467-9639.00063.

[5] E. Langford, Quartiles in elementary statistics, *Journal of Statistics Education* **14** (2006), doi:10.1080/10691898.2006.11910589.

[6] Y. Ma, M. G. Genton and E. Parzen, Asymptotic properties of sample quantiles of discrete distributions, *Ann. Inst. Statist. Math.* **63** (2011), 227–243, doi:10.1007/s10463-008-0215-z.

[7] Y. Miao, Y.-X. Chen and S.-F. Xu, Asymptotic properties of the deviation between order statistics and p-quantile, *Comm. Statist. Theory Methods* **40** (2011), 8–14, doi:10.1080/03610920903350523.

[8] J. W. Tukey, *Exploratory data analysis*, volume 2, Reading, MA, 1977.

[9] J. Žerovnik, Računanje kvartilov v elementarni statistiki, *Obzornik matematiko in fiziko* **64** (2017), 20–31, http://www.dlib.si/?URN=URN:NBN:SI:DOC-9RRUOHKH.

[10] J. Žerovnik and D. Rupnik Poklukar, Elementary methods for computation of quartiles, *Teaching Statistics* **39**, 88–91, doi:10.1111/test.12133.